

Università degli Studi di Napoli Federico II

Piano Lauree Scientifiche

PLS Virtual Summer
School for Students
PVS3

a cura di:

*Ugo Caruso, Marianna Crispino, Silvia Galano,
Alessandro Iannace, Maria Rosaria Iesce,
Marco Lapegna, Giancarlo Ragozini, Daniel Riccio, Italo Testa*



GIANNINI
EDITORE

Il presente volume è stato realizzato dal Dipartimento di Matematica e Applicazioni, nell'ambito delle attività trasversali di ateneo del Piano Lauree Scientifiche, con il contributo del Consiglio di Amministrazione dell'Università degli Studi di Napoli Federico II (DR/2020/3659 del 11/11/2020)

Giannini Editore
Via Cisterna dell'olio 6b
80134 Napoli
editore@gianninisp.it
ISBN: 978-88-6906-195-0

Copyright © 2021 del Dipartimento di Matematica e Applicazioni,
Università degli Studi di Napoli Federico II
Tutti i diritti riservati

Finito di stampare a Napoli
presso le Officine Grafiche F. Giannini & figli S.p.A.
nel mese di luglio 2021

Organizzazione della Scuola PVS3

- Ugo Caruso - *Dipartimento di Scienze Chimiche, Università Federico II di Napoli*
- Marianna Crispino - *Dipartimento di Biologia, Università Federico II di Napoli*
- Silvia Galano - *Dipartimento di Fisica "E. Pancini", Università Federico II di Napoli*
- Alessandro Iannace – *Dipartimento di Scienze della Terra, dell'Ambiente e delle Risorse (DISTAR), Università Federico II di Napoli*
- Maria Rosaria Iesce - *Dipartimento di Scienze Chimiche, Università Federico II di Napoli*
- Marco Lapegna - *Dipartimento di Matematica e Applicazioni "R. Caccioppoli", Università Federico II di Napoli*
- Giancarlo Ragozini - *Dipartimento di Scienza Politiche, Università Federico II di Napoli*
- Daniel Riccio - *Dipartimento di Ingegneria Elettrica e delle Tecnologie dell'Informazione (DIETI), Università Federico II di Napoli*
- Italo Testa – *Dipartimento di Fisica "E. Pancini", Università Federico II di Napoli*

Editori

- Maria Rosaria Iesce - *Dipartimento di Scienze Chimiche, Università Federico II di Napoli*
- Marianna Crispino – *Dipartimento di Biologia, Università Federico II di Napoli*
- Silvia Galano - *Dipartimento di Fisica "E. Pancini", Università Federico II di Napoli*

INDICE

Introduzione: La scuola PVS3	pag.	7
<i>Rossella Di Giaimo</i> Geni ed espressione genica	”	9
<i>Alessandro Iannace e Valentino Di Donato</i> Variazioni climatiche del passato: come si studiano e cosa ci insegnano per il futuro	”	21
<i>Daniel Riccio</i> Dal reale al digitale, dal problema alla soluzione	”	37
<i>Aniello Buonocore</i> Alcuni esempi di “quotidiana” probabilità	”	51
<i>Rosanna del Gaudio</i> In silico lab@home: database querying e database searching	”	63
<i>Barbara Brandolini e Daniel Riccio</i> Seminario integrato - Un mondo frattale Parte 1 (BB)	”	77
Parte 2 (DR)	”	88
<i>Lucio Palazzo, Francesco Santelli e Giancarlo Ragozini</i> Il futuro degli studenti: dal questionario all’analisi statistica dei dati testuali	”	101
<i>Giovanni Scopece e Maria Rosaria Iesce</i> Seminario integrato - Interazioni tra piante ed animali: aspetti evolutivi ed aspetti molecolari Parte 1 (GS)	”	113
Parte 2 (MRI)	”	121
<i>Umberto Scotti di Uccio e Oreste Tarallo</i> Seminario integrato - Orbitali e configurazione elettronica Parte 1 (US)	”	141
Parte 2 (OT)	”	158

<i>Rosita Diana</i>	
Celle solari polimeriche convenzionali: materiali, architetture e funzionamento	pag. 171
<i>Silvio Barra</i>	
Sensori e Segnali nell'Internet of Things (IoT)	” 187
<i>Nicola Mondillo e Roberto Esposito</i>	
Seminario integrato - I metalli critici: fonti naturali e recupero da RAEE	
Parte 1 (NM)	” 193
Parte 2 (RE)	” 201
<i>Lorenzo Manti</i>	
Un approccio scientifico al tema del rapporto tra uomo e radiazioni ionizzanti	” 211
<i>Ugo Caruso, Mariangela Succoio e Nicola Zambrano</i>	
Seminario integrato - Fluorescenza: dal fotone ai materiali e alle proteine reporter	
Parte 1 (UC)	” 219
Parte 2 (MS e NZ)	” 229
<i>Gerardino D'Errico e Roberto Di Capua</i>	
Seminario integrato -	
Parte 1 (GDE): La termodinamica come approccio scientifico alla realtà che ci circonda	” 243
Parte 2 (RDC): Il Secondo Principio della Termodinamica	” 251
<i>David Iacopini</i>	
Il futuro del sottosuolo: le sue risorse energetiche ed i suoi rischi	” 269
<i>Maria Rosaria Celentani</i>	
Matematica per la Crittografia	” 283
Biografie dei relatori	” 297

Introduzione: La scuola PVS3

La PLS Virtual Summer School for Students (PVS3) è stata organizzata congiuntamente dai referenti dei Progetti Piano Lauree Scientifiche (PLS) dell'Università di Napoli Federico II, proff. Ugo Caruso (Scienza dei Materiali), Marianna Crispino (Biologia/Biotecnologie), Alessandro Iannace (Geologia), Maria Rosaria Iesce (Chimica), Marco Lapegna (matematica), Giancarlo Ragozini (Statistica), Daniel Riccio (Informatica) e Italo Testa (Fisica). La Scuola PVS3 è rivolta agli studenti dell'ultimo biennio delle scuole secondarie di secondo grado allo scopo di approfondire le conoscenze delle materie di area scientifica, oltre gli aspetti solitamente trattati nel corso delle attività scolastiche. E' stata la prima iniziativa di questo tipo ad essere organizzata in modalità interamente on-line, ospitando gli studenti sulla piattaforma Teams dell'Università di Napoli Federico II. Si è trattato di un importante "esperimento organizzativo" in un momento in cui, a causa della pandemia COVID-19, si è reso necessario rivedere le modalità di orientamento universitario per gli studenti delle scuole secondarie. Le attività della Scuola PVS3, comprensive di lezioni e laboratori virtuali a carattere fortemente interdisciplinare, si sono svolte dal 7 all'11 settembre 2020 dalle 9.00 alle 12.00 e dalle 14.00 alle 16.00. La Scuola PVS3 ha riscosso grande successo come evidenziato dalla numerosa partecipazione: quasi 200 studenti, di età compresa tra 16 e 19 anni, con una percentuale di ragazze maggiore (61%) rispetto ai ragazzi, provenienti da scuole non solo di Napoli e provincia ma anche campane e di altre regioni. Sono stati proposti seminari monotematici relativi a tutte le aree PLS, ma anche seminari integrati tenuti da relatori afferenti a diverse aree che hanno collaborato per offrire punti di vista differenti su uno stesso tema. Ciò è stato possibile grazie al patrimonio di competenze, risorse, e collaborazioni maturato e consolidato nel corso degli anni dal PLS all'interno dell'Ateneo Federico II. In conclusione, la Scuola PVS3 ha dimostrato come sia possibile continuare ad offrire attività di altissimo livello, adattandosi alla situazione contingente, ma in continuità con le iniziative e i progetti già avviati.

I referenti PLS ringraziano tutti i relatori che hanno collaborato alla realizzazione della Scuola PVS3 e in particolare: Silvio Barra, Barbara Brandolini, Aniello Buonocore, Maria Rosaria Celentani, Rosanna del Gaudio, Gerardino D'Errico, Rosita Diana, Roberto Di Capua, Rossella Di Giaimo, Roberto Esposito, David Iacopini, Lorenzo Manti, Nicola Mondillo, Lucio Palazzo, Francesco Santelli, Giovanni Scopece, Umberto Scotti di Uccio, Oreste Tarallo, Nicola Zambrano. Un ringraziamento speciale alla Dr.ssa Silvia Galano che, con il Comitato

Organizzatore, ha svolto una insostituibile attività di coordinamento relativa a tutti gli aspetti della Scuola PVS3. Un sentito ringraziamento va anche al Consiglio di Amministrazione dell'Università degli Studi di Napoli Federico II per aver messo a disposizione il finanziamento PLS con il quale è stato stampato il presente volume. E soprattutto grazie a tutti gli studenti che hanno partecipato alla Scuola PVS3 e ai loro insegnanti che li indirizzano sempre con grande attenzione verso le iniziative PLS della Federico II.

I referenti dei progetti PLS dell'Università degli Studi di Napoli Federico II

Ugo Caruso - Scienza dei Materiali

Marianna Crispino - Biologia/Biotecnologie

Alessandro Iannace - Geologia

Maria Rosaria Iesce - Chimica

Marco Lapegna - Matematica

Giancarlo Ragozini - Statistica

Daniel Riccio - Informatica

Italo Testa – Fisica

Geni ed espressione genica

Rossella Di Giaimo

Dipartimento di Biologia, Università di Napoli Federico II

Abstract

Che cosa vuol dire esprimere un gene? Che cos'è un gene e come è fatto? Questi saranno gli argomenti di questo paragrafo. Ci focalizzeremo sul genoma umano, su quanti sono i geni e le proteine dell'uomo, sulla differenza tra individui della stessa specie e da cosa dipende. Poi discuteremo su come si può regolare l'espressione genica e infine su come si può interferire con l'espressione genica per riprogrammare una cellula a diventare un altro tipo cellulare e cambiare una sequenza di DNA per correggerla.

Iniziamo introducendo il dogma centrale della biologia molecolare che riguarda l'espressione genica e scaturisce dagli esperimenti di Watson e Crick, due scienziati cui è stato assegnato il premio Nobel in Fisiologia e Medicina nel 1962 per la scoperta della struttura del DNA che ha permesso gli studi sulla sua funzione. In questo contesto è di rigore citare il contributo determinante di un'altra importante ricercatrice, la cristallografa Rosalind Franklin che con i suoi studi sulla diffrazione ai raggi X e la sua famosa foto 51, contribuì a decifrare la struttura del DNA (Fig. 1).

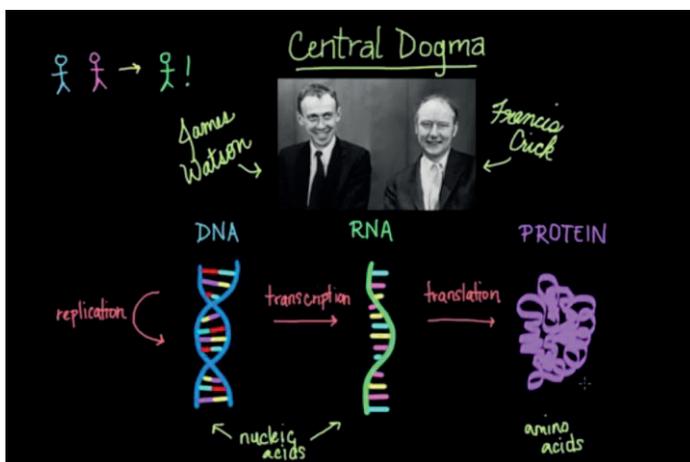


Figura 1. Immagine tratta da <https://youtu.be/KIvBn6gfRgY>

Il dogma della biologia molecolare riguarda il fluire dell'informazione genetica dal DNA alle proteine passando per l'RNA. Il DNA si replica all'interno della cellula in

modo tale che il materiale genetico della cellula madre venga ereditato dalle cellule figlie. L'espressione dei geni nel DNA, che contengono l'informazione per produrre una proteina, prevede come primo passaggio la trascrizione da DNA a RNA e un secondo passaggio la traduzione da RNA a proteina. Tutte queste molecole sono polimeri la cui unità di base nel caso degli acidi nucleici è il nucleotide e nel caso delle proteine e sono gli aminoacidi. Le proteine devono assumere una particolare struttura perché funzionino, e qualunque modifica della loro struttura determina una modifica della funzione.

L'espressione genica può essere paragonata a quando mandiamo in stampa un documento (Fig. 2): quando noi scriviamo un documento al computer, il computer rappresenta il magazzino delle informazioni quindi il computer si può paragonare al DNA che è il depositario dell'informazione genetica; quando mandiamo in stampa il nostro documento, l'impulso di stampa passa dal computer alla stampante ma il formato in cui le informazioni passano attraverso il cavo è sempre nello stesso formato cioè informazione di tipo digitale. Possiamo considerare il cavo come l'RNA. Allora il passaggio da DNA a RNA si chiama trascrizione proprio perché il linguaggio continua a essere uguale tra le due molecole, entrambe formate da nucleotidi, come nel nostro esempio parliamo di un messaggio digitale. Ma quando deve essere stampata la pagina allora il messaggio digitale arriva alla stampante e la pagina stampata rappresenta la proteina. La stampante rappresenta il ribosoma all'interno del quale avviene la sintesi proteica. Il linguaggio con cui viene stampata questa pagina, cioè il linguaggio dell'inchiostro, è ben diverso dal linguaggio digitale; infatti le proteine hanno un linguaggio diverso dagli acidi nucleici perché sono costituite da aminoacidi. Il passaggio delle informazioni da RNA a la proteina viene detto traduzione proprio perché si passa dal linguaggio degli acidi nucleici al linguaggio delle proteine.

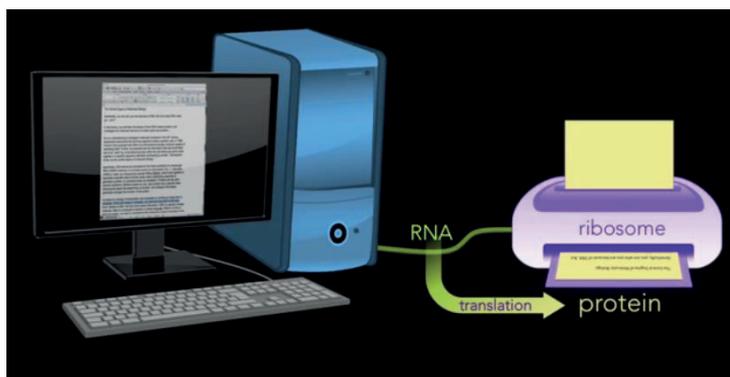


Figura 2. Immagine tratta da:

<https://www.youtube.com/watch?v=QvNdzLALvkl&feature=youtu.be>

Il DNA è un polimero i cui monomeri sono i nucleotidi; il nucleotide è costituito da tre componenti che sono lo zucchero, un gruppo fosfato e una base azotata. DNA è un acronimo che sta per “acido desossiribonucleico” perché lo zucchero del nucleotide del DNA si chiama desossiribosio. Le basi azotate del DNA sono quattro: Adenina, Citosina, Guanina e Timina. Poiché i vari nucleotidi differiscono l’uno dall’altro per la base azotata che li compone, i vari nucleotidi si indicano con l’iniziale della base azotata. ACGT sono le lettere della vita, le lettere del DNA (Fig. 3).

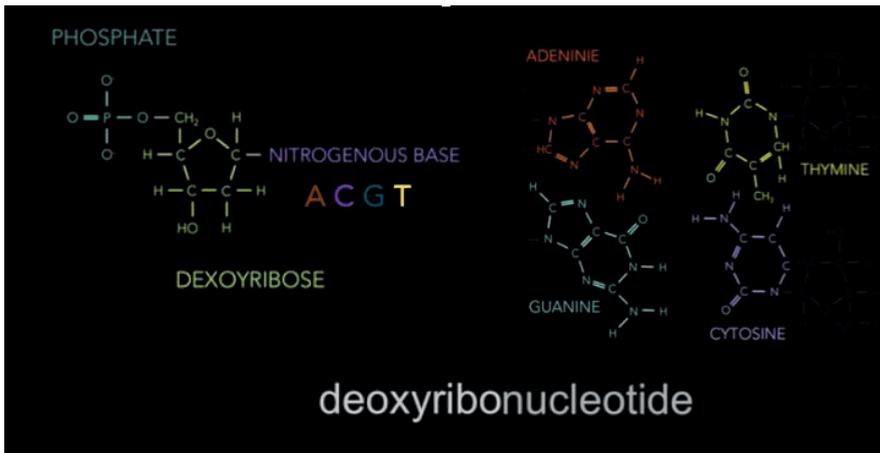


Figura 3. Immagine tratta da

<https://www.youtube.com/watch?v=QvNdzLALvki&feature=youtu.be>

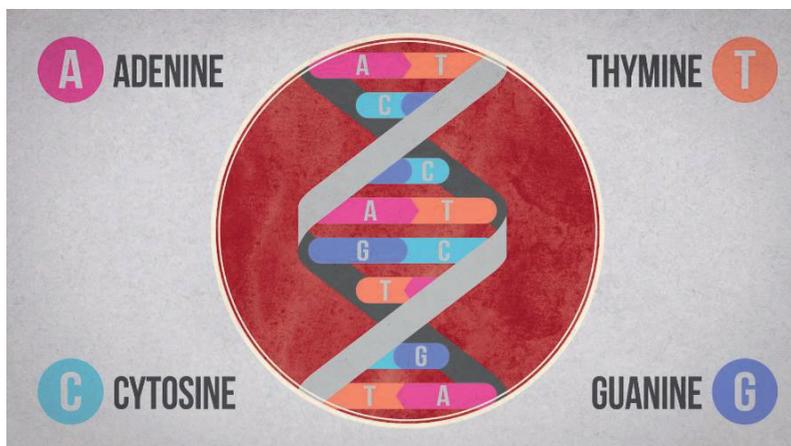


Figura 4: Immagine tratta da <https://youtu.be/MvuYATH7Y74>

Grazie alle scoperte di Watson e Crick sappiamo che il DNA è costituito da 2 filamenti di nucleotidi. I nucleotidi nei 2 filamenti sono legati tra loro in direzione

opposta, infatti si dice che i filamenti sono antiparalleli, e si associano in maniera complementare. La complementarità, cioè le regole di appaiamento tra le basi azotate dei nucleotidi presenti sui due filamenti, fu scoperta dallo scienziato Chargaff. Le regole di Chargaff prevedono che l'adenina (A) di un filamento si appai con la timina (T) sul filamento opposto e la guanina (G) si appai con la citosina (C). Quindi la struttura tridimensionale del DNA è una elica destrorsa come in Figura 4 dove i nastri rappresentano lo scheletro zucchero-fosfato all'interno del quale ci sono le coppie di basi appaiate con l'ordine di Chargaff e le coppie di basi si susseguono come gli scalini in una scala a pioli.

Qual è la dimensione del genoma umano?

La lunghezza del DNA si misura in numero di coppie di basi che si susseguono e nell'uomo il DNA è costituito da tre miliardi di coppie di basi per una lunghezza lineare di 2 metri. Come entrano questi 2m di DNA all'interno del nucleo di una cellula eucariotica che è una dimensione di 10 micron (10-6m) quindi parecchi ordini di grandezza più piccolo? Immaginiamo il DNA come il filo di un telefono che normalmente è avvolto ad elica ma è molto più lungo se lo tendete. Il DNA sfrutta lo stesso principio, si avvolge su sé stesso e si superavvolge per entrare nel nucleo. Il superavvolgimento è possibile grazie alla presenza di alcune proteine specializzate nella compattazione del DNA che sono le proteine istoniche. Le proteine istoniche sono 5: H1, H2A, H2B, H3 e H4. Queste proteine hanno una caratteristica particolare, infatti, hanno carica positiva che ben si associa con il DNA che ha una carica negativa grazie alla presenza dei gruppi fosfato nella sua struttura. Istoni e DNA si attraggono e si associano e il DNA si avvolge intorno alle proteine istoniche formando una struttura che si chiama nucleosoma che è il primo livello di compattazione del DNA nel nucleo. Immaginate questa lunga fila di nucleosomi sul DNA che si possono osservare anche al microscopio e somigliano ad una collana di perle cioè con tante palline che sono i nucleosomi collegate da filamenti di DNA. Il DNA è poi ulteriormente compattato grazie ad altre proteine a formare i cromosomi. Il DNA superavvolto e compattato nel nucleo in associazione alle proteine istoniche si chiama cromatina (Fig. 5).

Immaginiamo di entrare virtualmente all'interno di una cellula (<https://youtu.be/gG7uCskUOrA>): nel citoplasma ci sono i vari organelli e le varie strutture, tra cui il nucleo. Entriamo nel nucleo attraverso i pori nucleari e arriviamo ai cromosomi. Affinché un gene sul DNA venga espresso, i cromosomi devono essere parzialmente srotolati lì dove il gene viene espresso, perché il DNA deve essere accessibile per essere per poter essere trascritto.

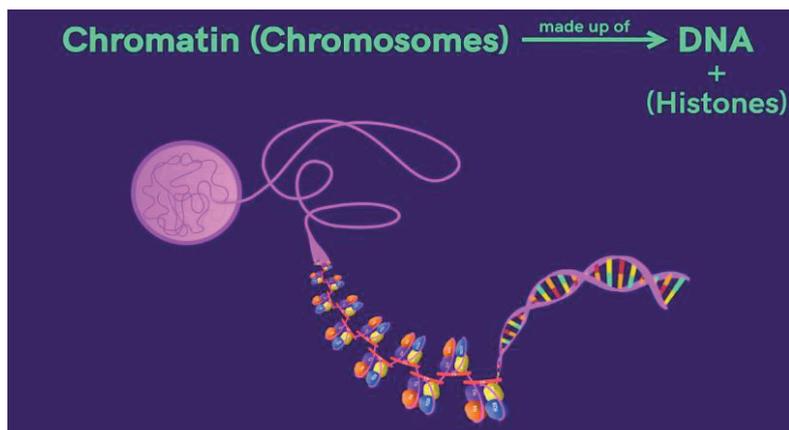


Figura 5. Immagine tratta da <https://youtu.be/4gFF1-VHHmk>

Il primo passaggio dell'espressione genica è la trascrizione ad opera di attività enzimatiche, le RNA polimerasi, che trascrivendo il DNA, sintetizzano un filamento di RNA in maniera complementare appunto seguendo la regola di appaiamento delle basi. Sulla base della sequenza di nucleotidi del DNA viene sintetizzato l'RNA messaggero. Il gene è costituito da parti che si chiamano esoni che contengono le informazioni per codificare le proteine e parti che si chiamano introni e che invece devono essere eliminati con un sistema di taglia e cucì chiamato splicing (Fig. 6)

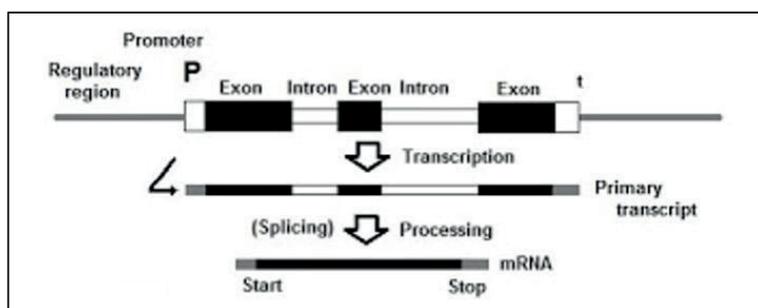


Figura 6. Esempio di struttura di un gene eucariotico per proteina

Alla fine del processo di trascrizione si ottiene una molecola di RNA messaggero maturo privo di introni. La molecola di RNA poi passa nel citoplasma dove viene associato con il ribosoma dove avviene la sintesi proteica. La traduzione avviene grazie a molecole traduttrici che si chiamano tRNA (RNA transfer) che sono in grado di leggere l'informazione dell'RNA in nucleotidi e tradurla in aminoacidi delle proteine secondo un codice genetico ben preciso. Tre nucleotidi sull'RNA costituiscono una tripletta che corrisponde ad un aminoacido. Quindi sulla base della sequenza dell'RNA e delle triplette di nucleotidi, vengono aggiunti gli aminoacidi e

si forma la proteina grazie delle attività di sintesi dei legami peptidici tra gli aminoacidi che sono all'interno del ribosoma. Successivamente la proteina acquisisce la sua struttura finale. Il progetto genoma umano è stato volto allo studio e l'analisi della sequenza del DNA e dell'informazione genetica che sta nelle nostre cellule. Sequenza del DNA che non vuol dire altro che una lunga stringa di lettere costituita da quattro lettere come vi ho detto sopra: ACGT. Il progetto genoma è stato un grosso progetto che è iniziato nel 2003 e che è durato ben vent'anni con il contributo di scienziati da tutto il mondo e anche costi elevati. L'obiettivo futuro è quello di fornire in maniera economica e accessibile a tutti la sequenza del proprio genoma. Ma capire come funziona il DNA e dove sono i geni in questa lunga stringa di 3 miliardi di lettere formata un alternarsi di sole quattro lettere, senza punteggiatura, non è semplice (Fig. 7).

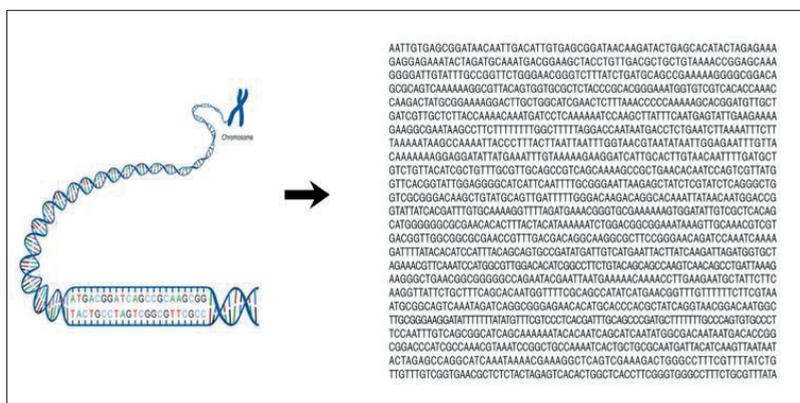


Figura 7: La sequenza di DNA è una stringa formata dall'alternarsi delle 4 lettere senza punteggiatura

Un gene eucariotico che codifica per una proteina (Fig. 6) contiene, come dicevamo prima, esoni e introni. Gli esoni contengono le informazioni per le proteine mentre gli introni devono essere rimossi. Stime recenti indicano che i geni per proteine nell'uomo sono 22.000, un numero piuttosto piccolo se consideriamo per esempio anche nel genoma dell'uva ci sono più geni, circa 30.000. Una cosa importante che bisogna considerare è che 22.000 sono il numero di geni che codificano per proteine, ma bisogna ricordare che l'RNA messaggero maturo si forma con processi complessi di splicing, un meccanismo di taglia e cuci per eliminare gli introni e ricucire gli esoni. Diversi RNA maturi si possono formare con diversi procedimenti di taglia e cuci a partire da uno stesso gene. Ne deriva che il numero di RNA messaggeri che si possono originare da questi 22000 geni è notevolmente più alto del numero di geni perché si possono anche tagliare e incollare frammenti diversi dello stesso gene per

formare più sequenze di RNA diverse. In questo modo, a partire da 22000 geni circa, si arriva a 100.000 RNA trascritti che alla fine diventano centinaia di migliaia di proteine. Inoltre va considerato che i geni non sono solamente quelli che codificano per proteine. In effetti la frazione del nostro DNA che diventerà proteine è solo circa l'1,5% del totale. Ci sono molti altri geni che codificano per RNA funzionali cioè molecole che funzionano come RNA, un esempio tra tutti è il tRNA o RNA traduttore di cui abbiamo parlato in precedenza che è importantissimo per la vita di una cellula ma non viene tradotto in proteine. Da questo deriva la definizione di gene come una sequenza di DNA che codifica per una funzione, quindi che trascrive per un RNA funzionale o per una proteina. La parte del DNA che verrà tradotta in proteine è una minima parte di tutto il genoma ma basta una variazione di un singolo nucleotide per produrre delle differenze nel fenotipo come la nostra altezza, il colore degli occhi, le lentiggini, il gusto, il colore dei capelli, colore della pelle e così via. Alcune variazioni nella sequenza possono provocare un malfunzionamento del gene e quindi una malattia come anemia, diabete, cancro e così via. Vi siete mai chiesti se individui diversi della stessa specie hanno lo stesso DNA? Il genoma umano tra individui diversi è identico al 99 % (Fig. 8).



Figura 8. Immagine tratta da https://youtu.be/jEJp7B6u_dY

Esistono moltissimi livelli di regolazione espressione genica negli eucarioti tra cui l'uomo, che partono dalla struttura del DNA fino arrivare alla proteina nella sua forma attiva; tutti questi passaggi sono regolati. Il primo livello di controllo dell'espressione di un gene è la struttura del DNA. A questo proposito vi cito un esempio: l'uomo e lo scimpanzé, pur essendo molto diversi tra loro, condividono il 98 77% del genoma che è identico tra scimpanzé e uomo ma c'è una differenza strutturale importante, infatti il cromosoma due nell'uomo non corrisponde a quello

dello scimpanzé perché il cromosoma due umano è il prodotto di fusione dei cromosomi 12 e 13 dello scimpanzé. Ecco che una variazione nella struttura e nell'organizzazione del DNA provoca una variazione di espressione genica tale per cui il risultato è ben diverso tra lo scimpanzé e l'uomo. Quando parliamo del 99% di identità tra genomi di individui umani diversi, abbiamo lo 0,1% di differenza che comunque consiste in tre milioni di nucleotidi diversi che comporta una differenza sostanziale tra individui della stessa specie. Tutte le cellule di un organismo multicellulare eucariotico hanno lo stesso DNA genomico? La risposta è sì, tipi diversi di cellule che appartengono allo stesso organismo come cellule dei nostri muscoli, delle cellule nervose, del sangue e così via, hanno lo stesso DNA. Allora cos'è che definisce il tipo cellulare a livello biomolecolare? Le proteine che le cellule esprimono; tipi cellulari diversi esprimono diverse proteine e questo dipende dalla regolazione della espressione genica e prima di tutto dall'organizzazione della cromatina che cambia nei diversi tipi cellulari. E qui entriamo nella sfera della Epigenetica. Epigenetica riguarda ciò che sta al di sopra delle informazioni genetiche c'è del susseguirsi delle quattro lettere ACGT nella stringa del DNA e che ne influenza l'espressione. Questo concetto fu introdotto per la prima volta nel 1957 da uno scienziato, Conrad Waddington. Per spiegare l'epigenetica basta immaginare una cellula rappresentata da una pallina (Fig. 9) che deve scegliere il percorso da fare per arrivare a valle, il percorso A o il percorso B. In pratica le cellule originariamente hanno tutte la stessa sequenza di DNA ma subiscono modificazioni che permettono alle diverse cellule di prendere percorsi diversi e diventare tipi cellulari differenti.

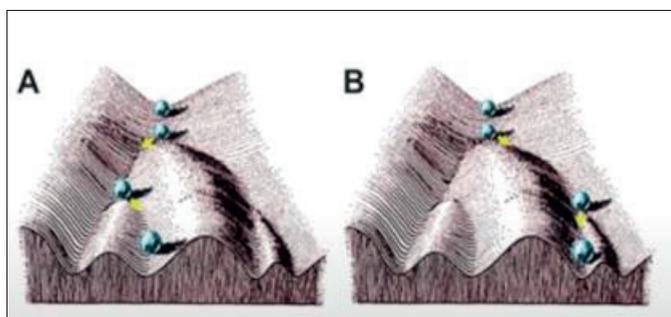


Figura 9. Modello di Conrad Waddington per rappresentare il concetto di scenario epigenetico di una cellula in differenziazione durante lo sviluppo.

Grazie ai meccanismi epigenetici, l'espressione genica può variare. Modificazioni epigenetiche possono riguardare il DNA o le proteine che partecipano alla struttura della cromatina come gli istoni. La epigenetica è importante durante lo sviluppo, infatti gli embrioni si originano da cellule che condividono tutto lo stesso DNA e poi durante lo sviluppo esprimono geni diversi e si differenziano nei 200 tipi cellulari

diversi che costituiscono il nostro organismo. L'epigenetica rappresenta una sorta di scambio sui binari sui quali viaggiano le cellule, portando l'espressione genica da una parte all'altra. L'epigenetica dipende anche molto dall'ambiente e questo spiega come mai gemelli che condividono lo stesso DNA, durante la crescita sviluppano delle caratteristiche completamente diverse che mantengono traccia delle esperienze fatte durante la vita, come per esempio una dieta equilibrata, un'attività fisica equilibrata, esposizione a sostanze chimiche e così via (https://youtu.be/_aAhcNjmvhc).

È possibile interferire artificialmente col programma di espressione genica di una cellula attraverso il così detto “reprogramming” o riprogrammazione di cellule somatiche in cellule staminali. Riprogrammare una cellula differenziata per farla tornare ad essere una cellula pluripotente in grado di diventare diversi tipi cellulari, è oggi possibile grazie agli studi di Shinya Yamanaka e di John B. Gurdon che hanno portato nel 2012 al premio Nobel per la Fisiologia e Medicina. Gli studi di Yamanaka hanno riguardato cellule della pelle del topo, fibroblasti, in cui sono state inserite quattro molecole, in particolare quattro proteine che sono fattori trascrizionali normalmente attivi nelle cellule staminali ed in grado di riaccendere il programma genico delle staminali e spegnere quello dei fibroblasti. In questo modo le cellule sono in grado di tornare indietro nel loro programma di differenziamento e diventano “pluripotenti” (Fig. 10)

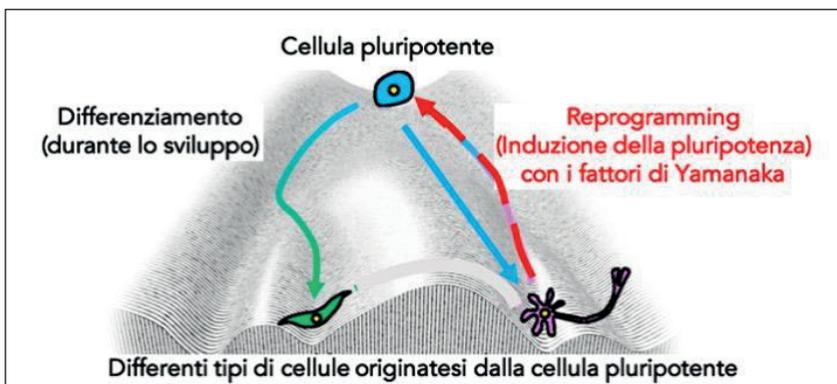


Figura 10. Concetto di riprogrammazione cellulare utilizzando il modello di Waddington (modificato da Takahashi and Yamanaka, 2013).

I ricercatori usano queste cellule staminali riprogrammate perché possono essere nuovamente differenziate in vari tipi cellulari permettendo così di studiare in laboratorio cellule diverse e di modellare varie patologie. Ad esempio, è possibile prendere cellule dal sangue e dalla pelle di persone malate di una patologia, ad esempio cellule della retina dell'occhio, inserire le molecole di Yamanaka,

trasformare queste cellule in cellule pluripotenti e indirizzarle poi al differenziamento in cellule della retina in laboratorio per studiare quale meccanismo non funziona nelle cellule del paziente (<https://youtu.be/J-PKT-xkVLO>). La tecnologia delle cellule pluripotenti riprogrammate (iPSC) negli ultimi anni sta prendendo molto piede perché a partire dalle iPSC si possono creare in laboratorio degli organoidi che assomigliano alle strutture o agli organi umani, come ad esempio organoidi di cervello. Nel mio lavoro di ricerca, mi occupo di un'epilessia umana e per far questo costruisco dei mini-cervelli (<https://youtu.be/SvnX9NJ1Zuo>) a partire da cellule di pazienti affetti da questa patologia e cerco di studiare il meccanismo che provoca la epilessia per cercare delle possibili cure (Fig. 11).

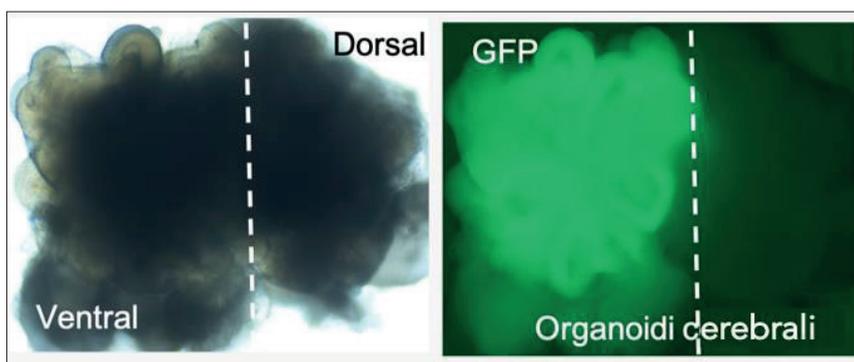


Figura 11. Immagini di organoidi cerebrali originati a partire da iPSC riprogrammate da cellule del sangue di pazienti affetti da una rara forma di epilessia (modificato da Di Matteo et al, 2020).

Un altro sistema per riprogrammare una cellula e fare variare l'espressione genica è il sistema di editing o modificazione del genoma con la tecnologia CRISPR-CAS. Il sistema CRISPR-CAS è costituito da una proteina che si chiama CAS che funziona come forbice molecolare. Per essere indirizzata sul giusto punto del DNA, questa si associa a una molecola di RNA guida che guida la forbice molecolare sul DNA in una regione precisa sulla base della complementarità di sequenza dell'RNA guida e il DNA bersaglio (Fig. 12). La molecola di RNA guida scorre sul DNA finché non trova la regione complementare sul DNA al quale si associa e a questo punto la forbice molecolare può funzionare e effettuare un taglio sul DNA nella regione specifica (<https://youtu.be/2pp17E4E-O8>). A seguito della rottura provocata dal sistema CRISPR-CAS sul DNA, la cellula cerca di correggere l'errore con dei meccanismi di controllo e introduce però spesso degli errori. Nella maggioranza dei casi, il gene a valle del punto in cui è avvenuto il taglio viene inattivato. Questo sistema estremamente potente è stato scoperto da due ricercatrici che hanno studiato

questo meccanismo usato come difesa immunitaria dai batteri contro infezioni del batteriofago che sono appunto i virus dei batteri.

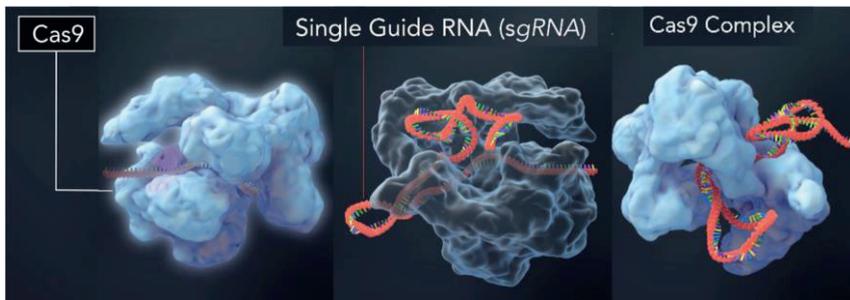


Figura 12. Complesso CRISPR-CAS tratto da <https://youtu.be/4YKFw2KZA5o>

Per questa scoperta, il Nobel per la Chimica 2020 è stato assegnato alla biochimica francese Emmanuelle Charpentier e alla chimica americana Jennifer Doudna, “per aver sviluppato un sistema di editing del genoma Crispr-Cas9 per modificare pezzi di materiale genetico con grande precisione”. Il sistema Crispr-CAS nei batteri si attiva quando un batteriofago infetta una cellula batterica. La cellula risponde all’infezione producendo molecole di RNA che vanno a riconoscere il DNA del virus e formano complessi molecolari con le proteine CAS che partecipano al processo di degradazione della molecola di DNA virale. In questo modo la cellula si difende naturalmente dall’infezione. Un altro ricercatore da menzionare è Feng Zhang del Massachusetts Institute of Technology, MIT, negli Stati Uniti che è stato il primo ad avere l’idea di utilizzare questo sistema, che normalmente funziona nei batteri, a scopi biotecnologici proprio per indurre delle variazioni del genoma nelle cellule eucariotiche. I ricercatori sono riusciti a ricostruire un sistema che simula quello dei batteri con una proteina CAS ed un RNA guida che viene inserito all'interno delle cellule e si dirige in un sito bersaglio, deciso dai ricercatori, dove opera il taglio in una maniera precisa e specifica provocando lo spegnimento del gene che si trova nella regione bersaglio. Un'altra potenzialità di questa tecnica è quella di sostituire una copia di DNA con errori, che per esempio provocano una patologia, con la copia di DNA corretta e non patologica. Le potenzialità enormi di questa tecnica sono attualmente in studio in tutto il mondo con l'obiettivo di cercare di essere d'aiuto nel controllo, nella diagnosi e nella cura di patologie genetiche.

Variazioni climatiche del passato: come si studiano e cosa ci insegnano per il futuro

Alessandro Iannace, Valentino Di Donato
*Dipartimento di Scienze della Terra, dell'Ambiente e delle Risorse,
Università di Napoli Federico II.*

Abstract

La Geologia classica ha insegnato già nell'800 che la Terra ha subito delle variazioni climatiche importanti nel corso della propria storia ed in particolare le grandi glaciazioni del Pleistocene. A lungo si è pensato che tali fenomeni fossero il risultato di variazioni di insolazione determinate da variazioni dei parametri orbitali secondo un modello matematico elaborato da Milutin Milankovitch esattamente 100 anni fa. Tuttavia, solo la nascita della moderna paleoclimatologia, per merito di Cesare Emiliani, ha consentito di dimostrare in maniera quantitativa nel 1976 questa ipotesi con lo studio delle variazioni isotopiche nei sedimenti di mare profondo. A partire dagli anni '80 l'approfondimento di questi studi ha mostrato anche che esistono cicli ed eventi di variazione climatica molto bruschi e di breve durata, mostrando che il sistema climatico è un sistema complesso controllato da molteplici variabili e da meccanismi di retroazione. Al tempo stesso lo studio delle carote di ghiaccio ha consentito di mostrare che la percentuale di CO₂ nell'atmosfera ha raggiunto negli ultimi anni valori mai sperimentati dal Pianeta nell'ultimo milione di anni.

Introduzione. La geologia classica e la ricostruzione del passato

Un cambiamento climatico che interessi l'intero pianeta, cioè che sia globale, può sembrare un'idea del tutto scontata ad un cittadino del ventunesimo secolo, a prescindere dalle distinzioni che possono essere fatte se tale cambiamento sia naturale o indotto dalle attività umane. Raramente però riflettiamo sul fatto che se oggi siamo in grado di immaginare degli scenari di cambiamenti che interessano l'intero pianeta e perché lo studio della storia del pianeta stesso ci ha mostrato che tali cambiamenti sono avvenuti nel passato.

Lo studio scientifico sistematico del pianeta terra, e soprattutto della sua storia, cioè la geologia, è iniziato alla fine del '700. Ma già nella prima metà dell'800 furono raccolte le evidenze che la Terra ha conosciuto nel recente passato un momento nel quale le calotte glaciali ricoprivano gran parte del nord Europa e del Nord America.

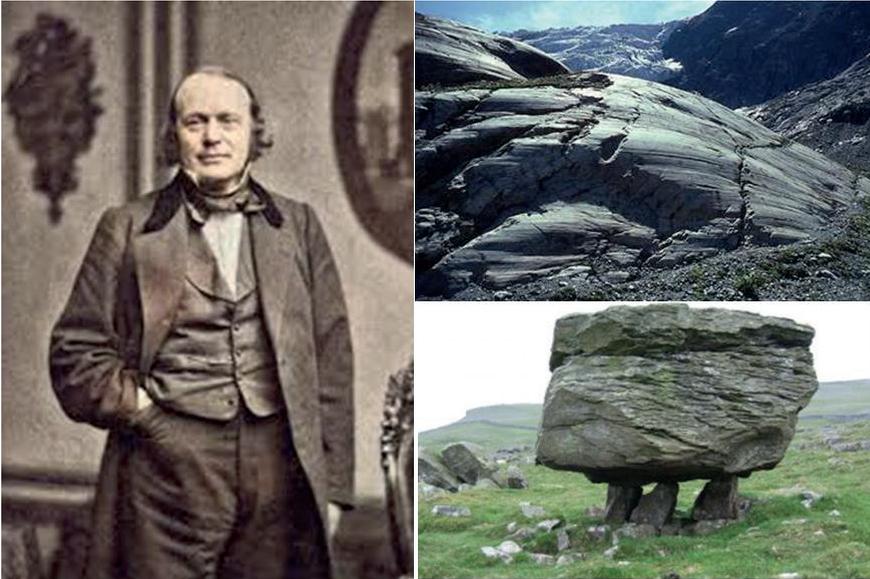


Figura 1. Louis Agassiz e, in alto a destra rocce striate dal movimento dei ghiacciai nelle Alpi e blocchi erratici nella campagna scozzese.

L'era glaciale, come l'abbiamo conosciuta fin da allora, fu una scoperta di un paleontologo svizzero trapiantato negli Stati Uniti, Louis Agassiz. Questa coraggiosa ricostruzione, che non riscontrò immediatamente il favore degli scienziati dell'epoca, era basata sulla raccolta di meticolose osservazioni effettuate nella pianura Svizzera a nord delle Alpi. Qui erano noti da tempo enormi blocchi di roccia isolati nella piatta campagna (Fig. 1), di composizione molto diversa dalle rocce circostanti, e per i quali era difficile individuare la sorgente. Montagne con la stessa composizione di tali blocchi erano lontani molti chilometri nelle aree più interne delle Alpi. Agassiz ritenne che l'unico modo di trasportare dei blocchi così giganteschi fosse quello di farli viaggiare inglobati in un ghiacciaio in modo tale che allo scioglimento di questo potessero essere depositi nella campagna ai piedi delle Alpi. Altro elemento fondamentale per tale ricostruzione fu l'osservazione che su tali massi erano presenti le caratteristiche strie che il movimento dei ghiacciai imprime nelle rocce sulle quali scorrono. Infine, nelle stesse aree era possibile osservare grandi accumuli di ciottoli in forma di grandi dorsali arcuate, le morene, del tutto simili a quelli che si trovano sul fronte dei ghiacciai attuali. Agassiz ebbe modo di effettuare osservazioni simili anche in Nord America ed in Scozia, accompagnato dai maggiori geologi dell'epoca. Il suo scenario apocalittico di una terra molto più fredda dell'attuale fu contrastato dal più influente geologo dell'epoca, Charles Lyell, il quale era riluttante ad ammettere spiegazioni di fenomeni passati che non facessero riferimento allo stato

attuale della terra. Ma lo stesso Lyell, verso la fine della sua esistenza, dovette ammettere che la ricostruzione di Agassiz corrispondeva al vero.

Nei primi anni del '900, i tedeschi Penk e Bruckner, attraverso osservazioni più sistematiche in Baviera nella valle del Danubio, perfezionarono l'ipotesi di Agassiz. Essi individuarono, studiando i depositi detritici lasciati da ghiacciai e fiumi nelle valli affluenti del Danubio, le tracce di almeno quattro successivi grandi periodi glaciali intervallati da altrettante fasi dette interglaciali. Dai nomi di tali valli essi vennero denominati Gunz, Mindel, Riss, Wurm, nomi che sono rimasti nell'uso corrente fino a qualche anno fa.

Che variazioni climatiche periodiche potessero avere avuto luogo sulla terra era in realtà un'ipotesi che i geologi avevano formulato da molto tempo. Anzi, ad esser precisi, il primo ad aver avanzato una tale ipotesi e più noto per la sua fondamentale attività di chimico che per quella di geologo. Stiamo parlando di Antoine Lavoisier, riconosciuto come il fondatore della chimica moderna, e caduto vittima della Rivoluzione francese per la sua attività di esattore delle tasse. Lavoisier fu il redattore di una carta mineralogica della Francia commissionata per la ricerca delle risorse minerarie. Nel corso di tale lavoro egli notò che spesso gli strati rocciosi presentano delle alternanze di ghiaie e di sabbie fini, che attribuì a processi di deposizione marina avvenuti rispettivamente vicino alla costa e lontano da essa. Da ciò dedusse l'idea di una ciclica variazione del livello del mare e ne ipotizzò un controllo climatico. Questa osservazione era incredibilmente in anticipo sui tempi: solo nella seconda metà dell'Ottocento diversi geologi attribuirono simili alternanze nella tipologia di strati sedimentari a variazioni cicliche del clima, per le quali non si disponeva di una causa né di una datazione.

Agli inizi del '900 venne elaborata la teoria che poi avrebbe dovuto fornire la causa di tali catastrofiche variazioni climatiche. Si tratta di un modello matematico elaborato da Milutin Milankovitch, professore all'università di Belgrado. Milankovitch decise di affrontare il tema della teoria matematica dei fenomeni termici prodotti dalla radiazione solare e, come applicazione del suo metodo, calcolò le variazioni temporali della quantità di radiazione che investe la terra nel corso del tempo. Questa è certamente la parte meglio nota dell'opera di Milankovitch, pubblicata nel 1920 e scritta quasi interamente durante la prigionia che lo scienziato dovette affrontare durante la Prima guerra mondiale nelle prigioni austro-ungariche. Il modello di Milankovitch si basa sulla conoscenza delle varie azioni che subisce il moto della Terra durante la sua rivoluzione intorno al sole. Tale variazione sono:

- ellitticità dell'orbita, con un periodo di circa 100.000 anni;
- inclinazione dell'asse terrestre, con un periodo di circa 40.000 anni;

- precessione degli equinozi, un moto complesso con un periodo di circa 21.000 anni. Secondo i calcoli di Milankovitch la composizione nel corso del tempo di tali variazioni (Fig. 2) comporterebbe una drastica e corrispondente variazione della quantità di energia solare intercettata dalla Terra alle medie latitudini. Fu immediatamente chiaro come queste variazioni potessero essere la causa dell'alternarsi di fasi glaciali ed interglaciali deducibili dalle osservazioni geologiche ma in quell'epoca non esistevano le condizioni per verificare quantitativamente tale modello.

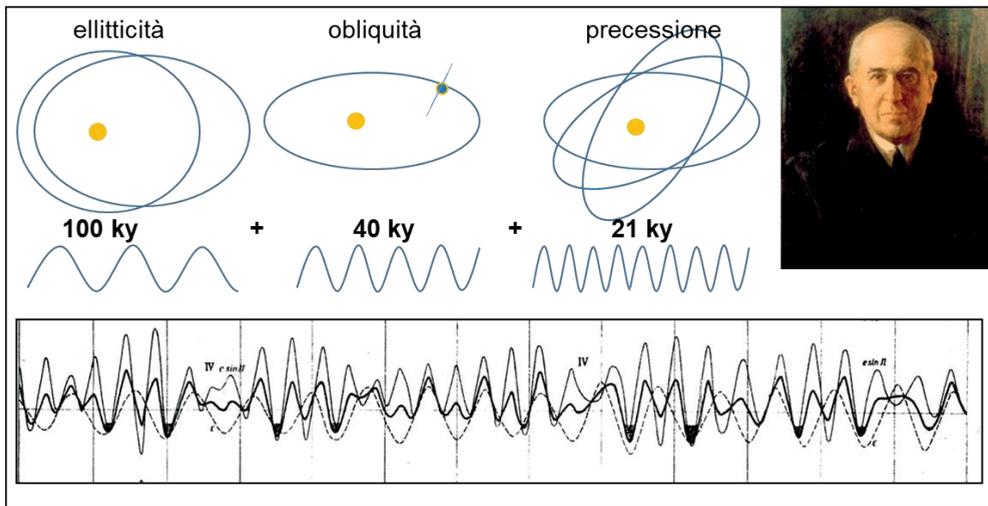


Figura 2. Dalla combinazione dei moti millenari si ottengono le variazioni della insolazione solare alle diverse latitudini nel tempo. Per Milankovitch sono cruciali i valori dell'insolazione estiva nell'emisfero boreale a circa 65°N di latitudine. Si tratta di quelle regioni, occupate da continenti, che risultano rispettivamente occupate o libere da calotte nei periodi glaciali e interglaciali. Bassi valori di insolazione estiva determinano uno scioglimento solo parziale delle nevi, innescando reazioni di feed-back legate all'aumento dell'albedo (riflettanza complessiva del Pianeta).

I sedimenti oceanici e la verifica dell'ipotesi di Milankovitch

La nascita della moderna paleoclimatologia ha dovuto però attendere gli anni '50 del '900. Questo non solo perché erano necessarie delle datazioni quantitative dei fenomeni geologici ma soprattutto perché essa è stata resa possibile dallo studio dei sedimenti marini profondi. L'esplorazione degli oceani, ed in particolare del fondo oceanico e della sua topografia, fu avviato immediatamente dopo la fine della Seconda guerra mondiale dal governo degli Stati Uniti per motivi di sicurezza

nazionale. Cospicui finanziamenti vennero affidati a geologi e geofisici di diverse istituzioni di ricerca americane sotto la guida del Office of Naval Research.

Ma la scintilla fondamentale per innescare l'esplosione di una nuova e fondamentale disciplina venne dalle ricerche effettuate, sempre per motivi militari, nei laboratori di fisica nucleare dell'università di Chicago, quelli dove aveva operato l'italiano Enrico Fermi per mettere a punto le prime pile atomiche e, ahimè, la prima bomba atomica. La preparazione dei reattori nucleari prevede l'uso della cosiddetta acqua pesante, un'acqua cioè nella quale gli atomi di idrogeno sono costituiti prevalentemente dalla specie isotopica denominata deuterio, cioè un isotopo contenente anche un neutrone rispetto all'atomo "normale" di idrogeno, il prozio, il cui nucleo è costituito da un solo protone. È proprio il neutrone del deuterio a rendere l'acqua più pesante e a presentare anche diverse altre caratteristiche fisiche diverse. Anche nel caso dell'ossigeno, che ha massa atomica prossima a 16, una piccola percentuale di atomi possiede uno o due neutroni in più e questi vanno a costituire isotopi di massa 17 e 18. Il premio Nobel della fisica Harold Urey scoprì che durante le normali reazioni chimiche, o durante processi fisici e biochimici, specie atomiche che passano da una molecola ad un'altra vengono "frazionate" durante il processo. Per esempio, il carbonio di un tessuto organico ha una composizione isotopica diversa dal carbonio presente nel carbonato di calcio di una conchiglia marina. Urey scoprì anche che in genere l'entità di questo frazionamento, in particolare quello del ^{18}O rispetto al ^{16}O , dipendono dalla temperatura: a temperature più basse in genere il processo di frazionamento è più accentuato (Fig. 3).

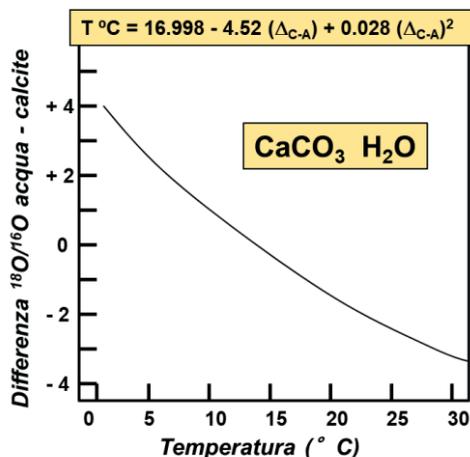


Figura 2. La composizione isotopica del carbonato che cristallizza in acqua è in relazione con quello dell'acqua stessa e con la temperatura: al diminuire della temperatura, il suo rapporto $^{18}\text{O}/^{16}\text{O}$ sarà via via più distante da quello dell'acqua

Quando una tale scoperta fisica arrivò all'orecchio dei primi geologi si pensò immediatamente di applicarla a calcolare le paleo-temperature del pianeta. Per diversi anni si sperimentò utilizzando diversi fossili, quali le ammoniti e le belemniti del Mesozoico, con l'idea che fossili depositi durante fasi più calde dovessero avere una composizione isotopica dell'ossigeno più leggera di fossili depositi durante fasi fredde. Questi tentativi però non diedero risultati di rilievo. Fu invece un giovane ricercatore dell'università di Bologna trapiantato a Chicago, Cesare Emiliani, che ebbe l'idea di misurare la composizione isotopica di conchiglie di organismi dei quali lui era uno specialista, i foraminiferi, prelevati però non in rocce molto antiche ma nei sedimenti degli oceani profondi. I foraminiferi sono organismi unicellulari, protisti, che costituiscono una significativa percentuale del plancton marino di tutti gli oceani. Emiliani studiò tre “carote” di sedimenti di diversi bacini oceanici. Con il termine carote si intende una sezione circolare dei sedimenti oceanici di qualche metro, o anche qualche centinaio di metri di lunghezza, ottenuta immergendo un carotiere perpendicolarmente al fondo dell'oceano (Fig. 4). Emiliani campionò i foraminiferi contenuti in queste carote partendo dai più recenti, cioè quelli presenti nei sedimenti superficiali, via via verso il basso fino a raggiungere quelli più antichi sulla parte inferiore della carota.

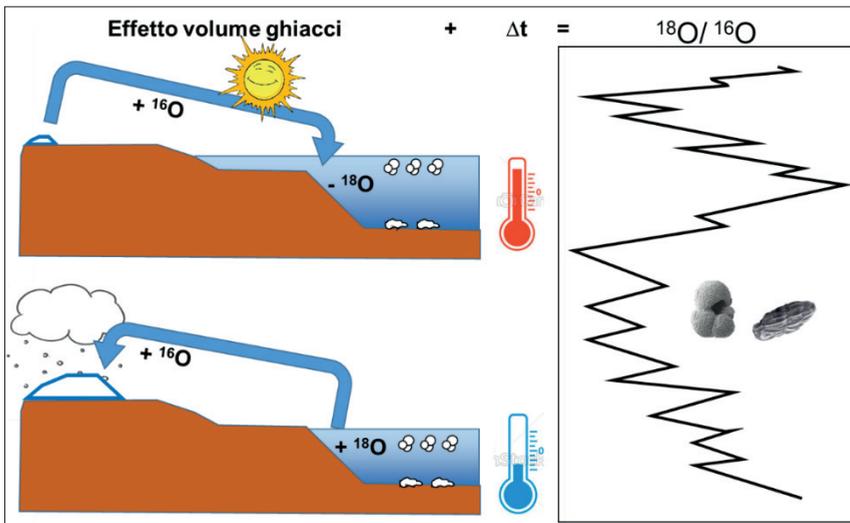


Figura 3. Il cambiamento del $^{18}\text{O}/^{16}\text{O}$ nel carbonato biogenico marino è connesso non solo a variazioni di temperatura ma anche variazione del volume dei ghiacci continentali. In effetti si ritiene che la variazione nel delta isotopico registrata tra l'ultimo periodo glaciale e l'attuale interglaciale sia riconducibili per circa il 70% al volume dei ghiacci. Il restante 30% è riconducibile alla variazione delle temperature e, regionalmente, a variazioni di salinità.

Misurando la composizione isotopica del carbonato di calcio dei gusci dei foraminiferi osservò che questa mostrava delle oscillazioni periodiche dall'alto verso il basso. Lo spessore delle carote campionate da Emiliani era relativamente modesto, solo pochi metri, ma lo scienziato italiano aveva a disposizione una datazione assoluta per almeno una di esse. Fu così in grado di dimostrare che negli ultimi 300.000 anni c'erano state almeno tre oscillazioni glaciale/interglaciale. Questo mandava in soffitta le precedenti ricostruzioni che prevedevano quattro fasi glaciali che coprivano l'intero Quaternario, per il quale ora si disponeva una datazione di circa due milioni di anni. Ma soprattutto poneva le basi della moderna paleoclimatologia, fondata sullo studio quantitativo biologico e chimico dei sedimenti di mare profondo.

Bisogna dire che il punto fondamentale dell'interpretazione di Emiliani, pubblicata nel 1955 nel *Journal of Geology*, fu contestato fin dall'inizio da molti scienziati dell'epoca. In effetti, la spiegazione delle variazioni isotopiche in termini di variazione di temperatura delle acque superficiali dell'oceano non era l'unica possibile. Fu fatto osservare che in un mondo glaciale il volume dei ghiacci terrestri intrappola una frazione significativa dell'acqua del sistema Terra, contenuta prevalentemente negli oceani. Poiché nell'acqua dei ghiacciai è concentrato in prevalenza l'isotopo 16 dell'ossigeno, aumentare a dismisura i ghiacciai significa (Fig. 5) arricchire indirettamente l'oceano di ossigeno 18 (^{18}O).

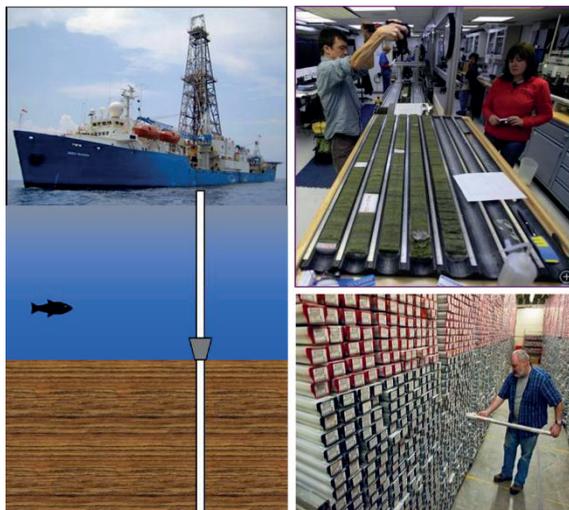


Figura 5. Le carote di sedimenti recuperate in mare profondo nel corso delle campagne oceanografiche vengono sezionate, descritte e campionate per gli studi successivi già a bordo. Dopo vengono conservate (in basso a destra il sito di stoccaggio di Brema, in Germania) in quanto costituiscono un fondamentale archivio di informazioni riguardanti l'evoluzione climatica della Terra nel corso di milioni di anni.

Pertanto, le variazioni osservate nel tempo nei gusci dei foraminiferi potevano essere spiegate semplicemente con una variazione di composizione isotopica dell'acqua di mare.

Questa diatriba non poté essere risolta per diversi anni. Bisognò aspettare il 1973 quando un ricercatore britannico, Neils Shakleton, fu in grado di misurare, grazie a uno spettrometro di nuova concezione, sia i foraminiferi platonici sia quelli bentonici contenuti in uno stesso campione di sedimento. Lo studio di Shakleton dimostrò che i foraminiferi planctonici e quelli bentonici registravano le stesse variazioni isotopiche. Poiché non è possibile che durante le variazioni glaciale /interglaciale le variazioni di temperatura delle acque superficiali possano essere state della stessa entità di quelle delle acque presenti sul fondo dell'oceano, Shakleton concluse che le variazioni osservati da Emiliani dovessero essere interpretate in gran parte come riflesso delle variazioni di composizione isotopica dell'acqua dell'oceano. E queste erano da mettere in relazione diretta col volume dei ghiacci. Lo stesso Schakleton concluse che ciò non faceva altro che rafforzare il valore paleo-climatico della interpretazione di Emiliani. Questo lavoro consacrava il metodo degli isotopi dell'ossigeno come il registratore fondamentale del volume dei ghiacci presenti in un determinato momento sul pianeta.

Ma la prima, sensazionale applicazione del nuovo metodo paleoclimatologico fu quella di poter dimostrare definitivamente l'ipotesi di Milankovic. Nel 1976 lo stesso Shakleton, insieme all'oceanografo William Hay e al matematico John Imbry, effettuò l'analisi di tre diversi parametri chimici e paleontologici in una carota oceanica. Tali parametri erano le variazioni nel corso del tempo del rapporto $^{18}\text{O}/^{16}\text{O}$ e del contenuto in foraminiferi planctonici e in radiolari, altri organismi planctonici ma aventi un guscio costituito da silice. Tale studio dimostrò, attraverso l'applicazione del metodo matematico dell'analisi di Fourier, che tutti questi parametri subivano nel corso del tempo delle variazioni cicliche nelle quali erano riconoscibili delle frequenze di circa 100.000, 40.000 e 21.000 anni, cioè esattamente le variazioni dei parametri orbitali presi in esame da Milankovic. Questo lavoro, pubblicata sulla rivista scientifica Science, deve essere considerato una pietra miliare della scienza del '900 in quanto dimostra in maniera inequivocabile quali sono i controlli fondamentali del clima terrestre.

I sedimenti oceanici e la storia climatica del Terziario

In ogni caso, Emiliani già a partire dagli anni '50 era diventato il più fervido sostenitore della necessità di lanciare un grande progetto di investigazione dei sedimenti dei fondali oceanici per ricostruire la storia della terra. Tale progetto si concretizzò nel 1968 ma solo grazie alla spinta rappresentata dalla necessità di

dimostrare la veridicità di una teoria elaborata nel corso degli anni 60, quella che sarà nota poi come teoria della tettonica a zolle o tettonica globale.

L'enorme mole di informazioni ricavate dallo studio di tali sedimenti ha consentito di definire con grande dettaglio le vicissitudini climatiche dell'intera Era Terziaria, cioè degli ultimi 65 milioni di anni (Fig. 6). Le principali informazioni ricavate da tali studi, e sulle quali c'è la concordia di tutta la comunità scientifica, possono essere così riassunte:

- 1) dall'inizio del Cenozoico ad oggi c'è stato un progressivo raffreddamento del clima interrotto soltanto da un momento più caldo tra la fine del Paleocene e l'inizio dell'Eocene (evento PETM);
- 2) il raffreddamento progressivo è stato però scandito anche da una serie di bruschi raffreddamenti, denominati *shift* isotopici, dei quali i più importanti sono avvenuti intono 35, 5, 3,4, 2,4 e 0.9 milioni di anni;
- 3) Le variazioni glaciali classiche hanno interessato gli ultimi 900.000 anni e sono consistite in 23 stadi glaciali interglaciali, con cicli che hanno avuto una durata approssimativa di circa 100.000 anni;
- 4) Prima di 900.000 anni fa le oscillazioni climatiche erano dominate da cicli con una frequenza maggiore, quella di circa 40.000 anni connessa alle variazioni di inclinazione dell'asse, ed erano meno estreme.

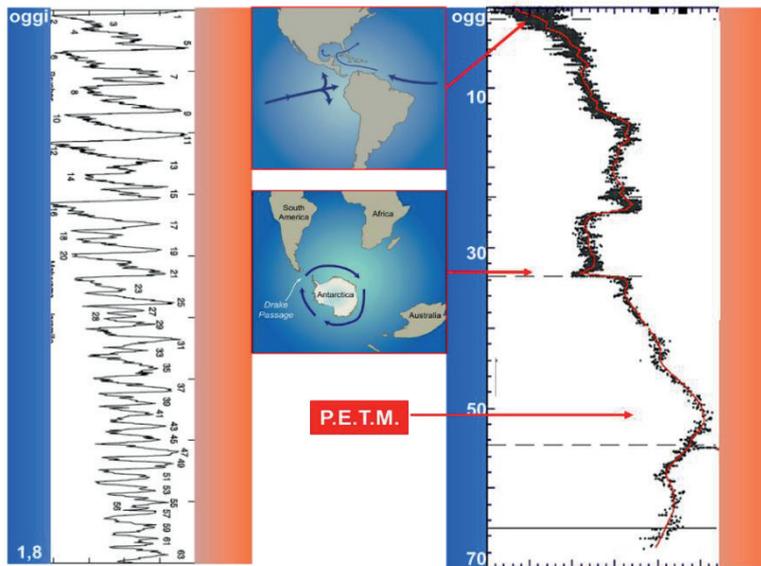


Figura 6. Sulla sinistra il dettaglio dei cicli glaciali degli ultimi 1,8 milioni di anni. Sulla destra le variazioni del Cenozoico con indicati i tre maggiori eventi climatici: il massimo termico della transizione Paleocene-Eocene, l'inizio delle glaciazioni antartiche e l'inizio delle glaciazioni artiche.

Gli eventi climatici naturali di breve durata

Un altro importantissimo campo di applicazione degli studi sugli isotopi stabili dell'Ossigeno e dell'Idrogeno è rappresentato dall'analisi delle carote di ghiaccio. Le calotte groenlandese e antartica sono state perforate con lo scopo di ricostruire l'evoluzione paleoclimatica nel corso degli ultimi cicli glaciale interglaciale. La composizione isotopica del ghiaccio è infatti in relazione con la temperatura: a temperature più basse corrispondono minori concentrazioni di ^{18}O e ^2H . Le carote di ghiaccio, anche se non consentono di andare indietro nel tempo quanto le successioni marine – la carota EPICA Dome C in Antartide raggiunge circa 800.000 anni, mentre per la Groenlandia record continui coprono gli ultimi 130.000 anni – forniscono un record paleoclimatico di estremo dettaglio.

Ad esempio, è stato appurato che durante il periodo glaciale più recente, l'area del Nord Atlantico ha subito una serie di drammatiche fluttuazioni climatiche note come Eventi Dansgaard-Oeschger, durante i quali le condizioni oceaniche e atmosferiche si sono alternate tra intervalli pienamente glaciali (i cosiddetti stadiali, cui è stata attribuita sigla GI) e intervalli relativamente temperati (i cosiddetti interstadiali - GS). Negli ultimi 115.000 anni sono stati riconosciuti 25 cicli interstadiale-stadiale. Gli stadiali sono diversi per intensità e durata, tuttavia essi sembrano caratterizzati da condizioni nel complesso più aride e da maggiore intensità dei venti, come indicato dall'incremento della concentrazione nel ghiaccio dello ione calcio, che è in relazione con il carico di polvere atmosferica.

La imprevedibilità della variabilità climatica nel corso dell'ultimo periodo glaciale è anche espressa nei cosiddetti eventi di Heinrich. Nei sedimenti fini dell'Atlantico settentrionale si rinvengono di tanto in tanto livelli caratterizzati da una elevata proporzione dei piccoli ciottoli di rocce di origine continentale (Fig. 7). La presenza di ciottoli in sedimenti fini è interpretata come dovuta allo scioglimento di grandi iceberg che incorporavano questo detrito e che si erano spinti evidentemente più a sud del "normale" durante eventi freddi. La dinamica di questi eventi, denominati con sigla HE e un numero progressivo, non è ancora del tutto chiarita. Essi ricadono all'interno di stadiali e sembrano precedere immediatamente il brusco riscaldamento che si registra al passaggio da stadiali ad interstadiali; il rilascio di grandi iceberg sarebbe in relazione alla destabilizzazione dei margini delle calotte – principalmente della calotta Laurentide che occupava gran parte del Nord America fino alle latitudini dell'attuale New York.

Tra gli stadiali, merita attenzione quello tradizionalmente denominato Dryas Recente, il cui nome deriva da *Dryas octopetala*, un fiore selvatico presente negli

ambienti montani e della tundra, e il cui polline è abbondante in alcuni livelli stratigrafici.

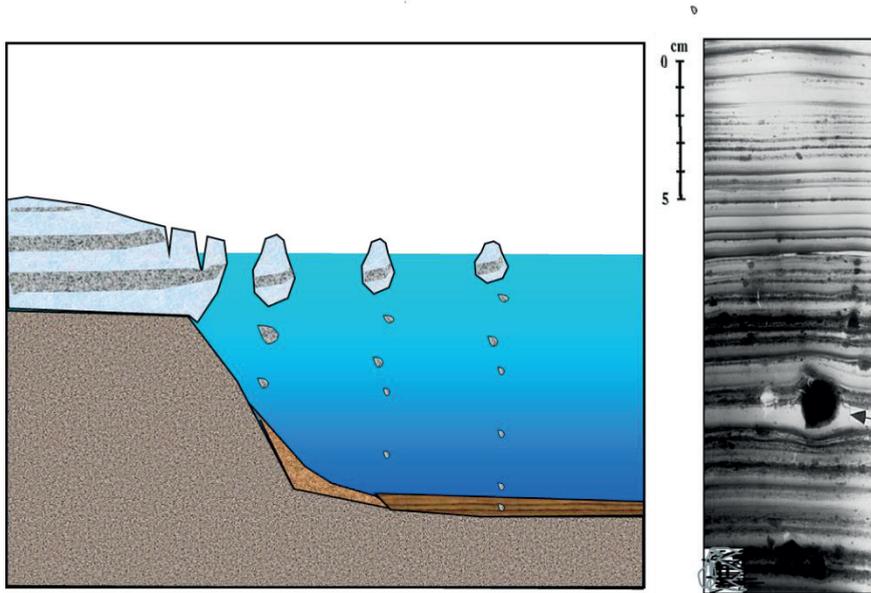


Figura 4. Gli eventi di Heinrich sono in relazione con la presenza, in depositi dei fondali dell'Oceano Atlantico, di detriti di origine glaciale rilasciati da grandi iceberg prodotti dalla destabilizzazione delle calotte continentali.

Si tratta di un intervallo freddo durato circa 1200 anni (tra circa 12.900 e 11.700 anni fa) che ha una importante espressione nelle carote della Groenlandia (dove è identificato come stadiale GS-1) e in generale nell'emisfero boreale, ivi inclusa l'area mediterranea. Questo evento interruppe in maniera brusca e anomala la deglaciazione che si era decisamente avviata a partire da ca. 15000 fa. (Fig. 8) Un aspetto particolare di questo stadiale è rappresentato dalla sua collocazione in una fase caratterizzata da un incremento dell'insolazione estiva nell'emisfero boreale, per cui le sue cause non possono essere ricercate nel *forcing* astronomico. Tra le ipotesi formulate per individuarne le cause la più accreditata prevede un improvviso e massiccio afflusso nell'Atlantico settentrionale di disgelo provenienti da un lago di origine glaciale, il lago Agassiz, collocato ai margini della calotta Laurentide in fase di ritiro. Questo avrebbe portato a una riduzione significativa o addirittura una interruzione della circolazione termoalina del Nord Atlantico, che rappresenta uno dei punti critici del sistema climatico globale. Infatti, le acque fredde del Nord Atlantico, rese più saline dalla formazione del pack, affondano dando inizio ad un nastro trasportatore che giunge fino al lontano Oceano Pacifico, contribuendo significativamente a ridistribuire il calore del Sistema Terra-

Oceano. Una ipotesi del tutto differente, e molto dibattuta, associa invece l'inizio del Dryas recente all'esplosione o all'impatto di uno o più oggetti di origine cosmica al di sopra della calotta Laurentide.

Lo studio degli isotopi stabili ricopre certamente un ruolo fondamentale nella ricerca paleoclimatica, ma esistono anche metodologie di ricostruzione basate su altri approcci. I metodi basati sullo studio di associazioni fossili consentono, sotto certe condizioni, di ottenere ricostruzioni quantitative dei parametri paleoclimatici. L'assunzione fondamentale alla base di queste tecniche è alcune biocenosi sono fortemente controllate dai parametri climatici. Questo può essere certamente affermato per organismi marini come foraminiferi planctonici e radiolari o per la vegetazione, anche se in quest'ultimo caso ci sono problematiche legate alle importanti modifiche operate dall'uomo sulla vegetazione, che potrebbe quindi non essere in equilibrio con le condizioni climatiche. Sulla base di questa assunzione si può cercare di definire una relazione matematica, ossia una funzione di trasferimento, che consenta di associare i parametri climatici alla composizione delle biocenosi.

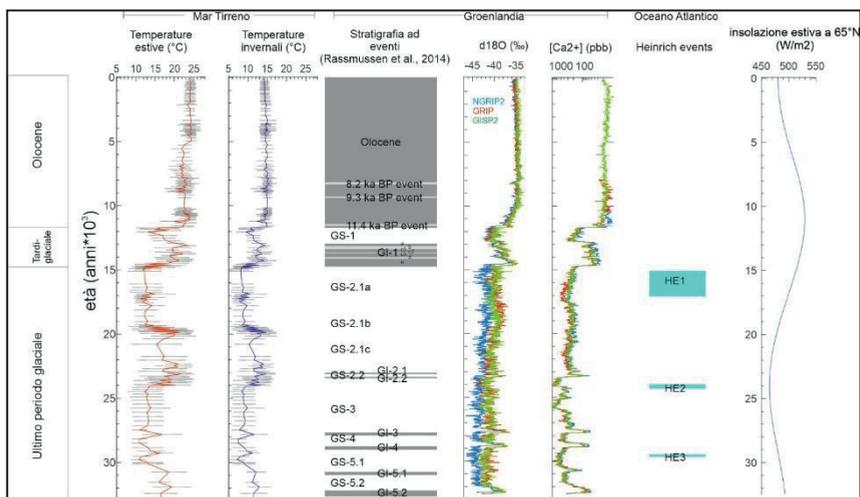


Figura 5. Ricostruzione delle SST estive e invernali del Mar Tirreno ottenute da una Carota recuperata nel Golfo di Salerno. Le barre grigie orizzontali rappresentano l'incertezza associata alla ricostruzione. La ricostruzione è messa in confronto con il record isotopico e geochemico delle carote di Ghiaccio della Groenlandia e i cicli interstadiale-stadiale di Rasmussen et al. (2014), la cronologia degli eventi di Heinrich dell'Oceano Atlantico settentrionale e con i valori della insolazione estiva a 65°N di latitudine (dati estratti da Laskar et al., 2004)

L'applicazione delle funzioni di trasferimento alle associazioni fossili consente quindi di ricostruire quantitativamente i parametri paleoclimatici del passato, tuttavia

essa richiede una seconda importante assunzione, ossia che le funzioni di trasferimento definite per il mondo di oggi siano applicabili al passato geologico. Questo assunto è d'altra parte necessario in tutte le ricostruzioni geologiche (e in generale in tutte le scienze storiche) che sempre fanno uso di leggi ricavate dal mondo attuale per interpretare fenomeni non riproducibili. In genere è attraverso l'incrocio di più indizi (*proxies*) che si cerca di verificare il grado di attendibilità di questa assunzione.

Nell'esempio a seguire è illustrata una ricostruzione delle temperature di superficie delle acque (SST) del Mar Tirreno per gli ultimi 30.000 anni ottenuta applicando una funzione di trasferimento alle associazioni a foraminiferi rinvenute in una carota di sedimenti recuperata nel Golfo di Salerno.

La ricostruzione evidenzia fluttuazioni delle SST estive che raggiunsero, nelle fasi più fredde del periodo considerato, valori fino a 10°C inferiori a quelli attuali. Un bagno nel Mar Tirreno di 16000 anni fa sarebbe stato equivalente ad un bagno nell'attuale Mare del Nord. La ricostruzione evidenzia anche come il Dryas Recente/GS1 ha impattato anche nel Mediterraneo, determinando una flessione delle SST di circa 7°.

Il dibattito sugli scenari climatici del futuro

Con lo studio del passato geologico abbiamo imparato che se è vero che la *modulazione* delle variazioni climatiche è stata controllata per tutto il Cenozoico (e probabilmente anche prima) dalle variazioni dei parametri orbitali secondo il modello di Milankovitch, il controllo di fondo da un lato, e i bruschi salti di stato dall'altro, sono determinati dalla configurazione del Sistema Terra nel suo complesso. Nel lungo periodo il fattore fondamentale è l'attività delle placche che determina la distribuzione delle terre emerse e degli oceani, la circolazione oceanica e, attraverso i cicli bio-geochimici, la percentuale di gas serra nell'atmosfera. Nel brevissimo periodo invece, numerosi fattori ed effetti di retroazione (*feedback*) possono fare in modo che, date le stesse condizioni termiche generali, il clima della terra può subire delle variazioni piuttosto brusche. Questa è una manifestazione della *complessità* del sistema, cioè di un sistema controllato da numerosissime variabili che hanno tra di loro dei sistemi di retroazione, talvolta positivi talvolta negativi, che rendono estremamente difficile la modellizzazione.

In particolare, le carote di ghiaccio, oltre alla registrazione isotopica di grande dettaglio, forniscono anche dati diretti sull'atmosfera del passato in quanto consentono di misurare la composizione di minuscole bolle intrappolate in ciascun strato di ghiaccio all'epoca della sua deposizione. Questi dati mostrano (Fig. 8) che variazioni di CO₂ hanno *accompagnato* le variazioni climatiche che sono

determinate essenzialmente, come abbiamo visto, dai parametri orbitali. Ma risulta evidente anche che la concentrazione di CO₂ nell'atmosfera è cresciuta molto rapidamente negli ultimi 150 anni, fino a valori molto più alti (più di 400 ppm) che in qualsiasi intervallo dell'ultimo milione di anni. E sappiamo anche che la CO₂ (e ancor più il metano) è anche uno dei forzanti principali del sistema climatico: il progressivo raffreddamento della Terra è stato in parte significativa determinato dalla sua progressiva diminuzione nell'atmosfera come risultato dell'attività delle placche.

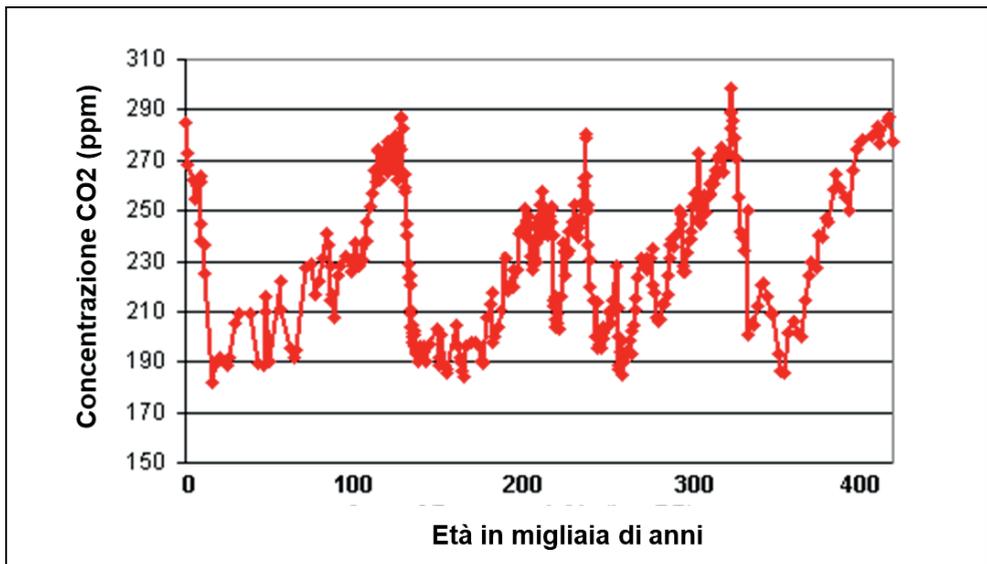


Figura 6. Variazione del contenuto atmosferico di CO₂ nel corso degli ultimi 400.000 anni ricostruita attraverso l'analisi delle bolle d'aria intrappolate nel ghiaccio della Carota Vostok recuperata in Antartide.

Tutte queste informazioni riguardano il passato come ricostruito con metodi geologici. Nel momento in cui si fanno previsioni sugli scenari climatici per i prossimi 50 o 100 anni, la "palla" passa ai climatologi, cioè coloro attingono a questi insegnamenti per delineare modelli matematici previsionali per il prossimo futuro. I numerosi lavori pubblicati da tutta la comunità scientifica sono costantemente passati in rassegna da un *Panel* di scienziati di tutto il mondo incaricato dall'ONU, il IPCC. Attualmente esiste un consenso quasi unanime tra tutti gli scienziati che è in corso un riscaldamento globale e che esso è stato innescato dal brusco aumento di CO₂ connesso alle attività umane a partire dal 1850 circa. Possiamo quindi paragonare l'aumento di CO₂ di origine antropica a uno di quegli eventi improvvisi

che alterano l'equilibrio del sistema e possono portarlo, attraverso meccanismi di feedback positivi, in un nuovo stato.

L'insegnamento per le generazioni in formazione, nella selva di opinioni discordanti che media e social media continuamente diffondono, è quello di rafforzare la propria formazione scientifica e documentarsi andando direttamente alle fonti, ed in particolare alla cospicua documentazione che l'IPCC produce e rende pubblica. Che il sistema sia complesso lo dimostra la storia geologica del Pianeta ma ciò non significa che esso sia incomprensibile e non vada studiato.

Bibliografia

- Emiliani, C. (1955). Pleistocene temperatures. *The Journal of Geology*, 63: 538-578.
- Hays, J. D., Imbrie, J., Shackleton, N. (1976). Variations in the Earth's Orbit: Pacemaker of the Ice Ages. *Science*, 194: 1121-1132.
- Heinrich, M. (1988). Origin and consequence of ice rafting in the northeast Atlantic Ocean during the past 130,000 years. *Quaternary Research*, 29:143-152.
- Laskar, J., Robutel, P., Joutel, F., Gastineau, M., Correia, A.C.M., Levrard, B. (2004). A longterm numerical solution for the insolation quantities of the Earth. *Astronomy and Astrophysics* 428, 261–285.
- Milankovitch M. (1920). *Théorie mathématique des phénomènes thermiques produits par la radiation solaires*. Gauthier-Villars, Parigi, 339 pp.
- Rasmussen S.O., Bigler M., Blockley S.P., Blunier T., Buchardt S.L., Clausen H.B., Cvijanovic I., Dahl-Jensen D., Johnsen S.J., Fischer H., Gkinis V., Guillevic M., Hoek W.Z., Lowe J.J., Pedro J.B., Popp T., Seierstad I.K., Peder Steffensen J., Svensson A.M., Vallelonga P., Vinther B.M., Walker M.J.C., Wheatley J.J., Winstrup M. (2014). A stratigraphic framework for abrupt climatic changes during the Last Glacial period based on three synchronized Greenland ice-core records: refining and extending the INTIMATE event stratigraphy, *Quaternary Science Reviews*, 106, 14-28.
- Shackleton N. (1967). Oxygen Isotope Analyses and Pleistocene Temperatures re-assessed. *Nature*, 215: 15-17.

Dal Reale al Digitale, dal Problema alla Soluzione

Daniel Riccio

*Dipartimento di Ingegneria Elettrica e delle Tecnologie dell'Informazione,
Università di Napoli Federico II*

Abstract

La filosofia della natura (*philosophia naturalis*), sin dalla sua nascita ha rivolto l'attenzione alla natura e, in una accezione più ampia, al mondo reale per astrarne gli aspetti più concreti in forme ed idee di ordine superiore, secondo la teoria platonica. Fra i suoi principali obiettivi, vi era certamente la comprensione e la spiegazione dei fenomeni naturali, nonché dei meccanismi alla loro base. Essa è considerata la precorritrice delle moderne scienze, che ne hanno certamente ereditato la peculiarità di guardare al mondo reale come alla principale fonte di spunti per la speculazione scientifica, ovvero l'individuazione di problemi ai quali il ragionamento ed il metodo scientifico cercano di dare una soluzione formalmente descrivibile e, di conseguenza, ripetibile. Si può, quindi, immaginare un circolo virtuoso, che dal mondo reale, naturale bacino da cui attingere problemi concreti, passa attraverso uno spazio delle soluzioni, in cui essi trovano una corrispondenza, la quale viene nuovamente calata nell'ambito da cui essi hanno avuto origine. Da un punto di vista squisitamente informatico, è di particolare interesse analizzare come due elementi cardini della computer science, ovvero l'uomo e la macchina, si innestano in questo ciclo. A tal fine, è possibile interpretare questo ciclo percorrendo un doppio viaggio, il primo dei quali ci accompagna dal mondo reale verso quello digitale, partendo dal dato e dall'informazione che esso porta con sé fino al bit come semplice segnale elettrico. Il secondo, invece, è di natura più astratta e ci mostra come un problema concreto generi una intuizione, come questa prende forma in uno schema ordinato e rigoroso per poi tradursi in una procedura corretta e ripetibile, al punto da poterne misurare quantitativamente l'efficacia e l'efficienza. In questo articolo, quindi, affrontiamo un doppio viaggio dal reale al digitale e dal problema alla soluzione.

Dal reale al digitale

L'informatica, alla stregua di tutte le altre discipline scientifiche, si contestualizza in una dimensione concreta e materiale dalla quale trae spunto per lo studio di problemi

e fenomeni, che astrae mediante un processo di analisi e progettazione rigoroso e formale. Il rigore ed il formalismo rappresentano requisiti fondamentali per garantire la correttezza e la valutazione di una soluzione concettuale al problema, la quale a sua volta è tradotta prima in un algoritmo e, conseguentemente, in un programma per poter essere così proiettata nel mondo digitale percorrendo il suo viaggio verso una unità di calcolo, cuore pulsante di un computer, in cui l'esecuzione fa sì che essa produca un risultato concreto chiudendo così il ciclo. Gli aspetti salienti di questo viaggio rappresentano l'oggetto di studio di questo paragrafo.

Quotidianamente, in modo pressoché automatico e meccanico, le persone trovano o applicano soluzioni note a problemi di routine attraverso l'esecuzione di procedure date per acquisite dall'esperienza o impartite da esperti di uno specifico ambito. Mentre il primo caso gioca un ruolo determinante per comprendere il processo di astrazione dal problema alla soluzione, il secondo è alla base della traduzione del problema dal reale al digitale. Infatti, un informatico svolge un compito analogo, benché per problemi di natura più complessa. Tralasciando, solo temporaneamente, il processo attraverso il quale egli formula la soluzione ad un problema, è interessante concentrarsi sul modo in cui tale soluzione venga elaborata e tradotta in una rappresentazione sufficientemente semplice da poter essere eseguita su un calcolatore, seguendo lo schema rappresentato in Figura 1.

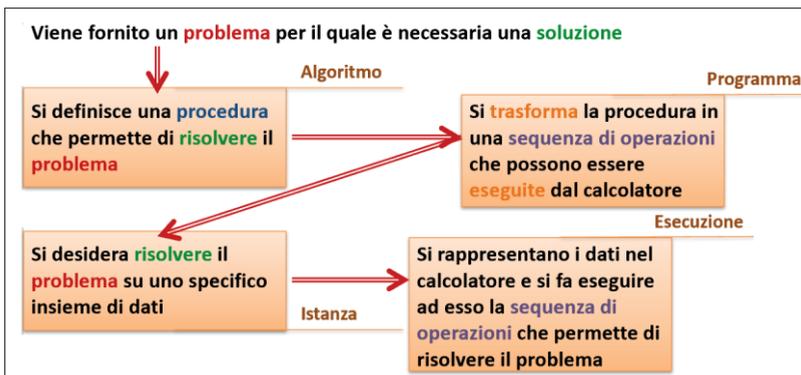


Figura 1. Il percorso problema-esecuzione.

A tale scopo, si consideri un esempio molto semplice, in cui si vuole spiegare ad un interlocutore il procedimento attraverso il quale è possibile calcolare l'area di un triangolo. Appare evidente come il nostro modo di svolgere questo compito cambi a seconda del particolare tipo di interlocutore. Il modo più conciso ed efficace di esprimere tale procedura è, senza alcun dubbio: "L'Area di un triangolo è data dal semiprodotto della sua base per l'altezza". Tuttavia, una tale spiegazione è da

considerarsi adeguata per una persona che abbia un livello di scolarizzazione pari almeno alla secondaria di primo grado.

In caso contrario, la stessa procedura verrebbe descritta in termini molto più semplici e con un maggior livello di dettaglio, al fine di garantirne la comprensione da parte dell'interlocutore, ovvero: "Dato un triangolo, la sua area è calcolata moltiplicando la base per l'altezza e dividendo il prodotto per due".

Partendo dal presupposto che un calcolatore ha una capacità di interpretare ed astrarre ben al di sotto di un interlocutore umano, appare evidente come anche questa spiegazione sia ben lontana da una formulazione della soluzione che esso sia in grado di elaborare. Questo induce un ulteriore livello di semplificazione, che ci impone di specificare la soluzione in una forma algoritmica, ovvero in una sequenza di passi che corrispondano quanto più possibile ad operazioni atomiche: "1) sia **b** la base del triangolo **T**, 2) sia **h** l'altezza del triangolo **T**, 3) Calcola **p** come **b**×**h**, 4) calcola **A** come **p**: 2". Benché estremamente semplice, anche questa forma ha in sé un grado di ambiguità troppo elevato per un elaboratore. Basti pensare a poche semplici domande: "a e b sono valori interi o reali?", "il risultato deve essere stampato?", "a e b possono essere anche nulli?". Un essere umano sarebbe in grado di gestire ciascuno di questi casi, sulla base della propria interpretazione, mentre una macchina si aspetta di non aver alcun grado di ambiguità. Questo implica, che la soluzione algoritmica necessita di essere tradotta in un programma, ovvero espressa in un linguaggio caratterizzato da una sintassi e da una semantica molto più rigide rispetto a quelle di una qualsiasi lingua naturale. I linguaggi di programmazione hanno esattamente questo scopo (Watt and Wong, 1990). Essi sono ancora sufficientemente vicini al modo di rappresentare i concetti da parte di un essere umano, ma grazie agli schemi molto rigidi che impongono non sono troppo lontani dal modo in cui un calcolatore interpreta le proprie istruzioni (vedi Fig. 2).

Un linguaggio di programmazione, generalmente, rappresenta un ottimo compromesso fra il rigore formale richiesto da una macchina e il livello di astrazione con cui un essere umano è abituato ad esprimere concetti. Tuttavia, un programma, pur essendo costituito da una sequenza di istruzioni semplici, non ha ancora una forma sufficientemente elementare da poter essere eseguito direttamente da una unità di calcolo. Richiedere ad un programmatore di scrivere un programma direttamente nella sua forma eseguibile rappresenterebbe uno sforzo eccessivo, ragion per cui si ricorre ad una soluzione software, il cui compito è del tutto analogo a quello che normalmente svolge un interprete a supporto di due o più interlocutori che non parlano la stessa lingua, ovvero traduce i concetti, nel caso specifico le istruzioni, da un linguaggio all'altro. Questo software è noto come *compilatore* (Morzenti et al., 2019). Grazie alla natura estremamente rigida e schematica dei linguaggi di

programmazione, il compilatore è in grado di analizzare un programma, scomponendolo nelle sue componenti più elementari, dette *token*. Dopo aver collezionato tutti i token presenti in un programma, e conoscendo le regole secondo cui essi possono e devono essere combinati fra loro, il compilatore procede con l'analisi sintattica del programma.

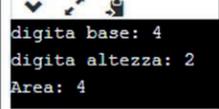
Problema: calcolare l'area di un triangolo, considerando base e altezza		
<p>Soluzione (ad alto livello) Semiprodotto della base per l'altezza</p> <p>Soluzione (elementare) Moltiplico la base per l'altezza e divido il risultato per 2.</p> <p>Soluzione (algoritmica) Sia b la base del triangolo T Sia h l'altezza del triangolo T Calcolo p come $b \times h$ Calcolo A come $p : 2$</p>	<p>Programma C</p> <pre> 1 #include<stdio.h> 2 3 int main () 4 { 5 6 int base, altezza, area; 7 8 printf ("digita base: "); 9 scanf ("%d", &base); 10 11 printf ("digita altezza: "); 12 scanf ("%d", &altezza); 13 14 area = base * altezza / 2; 15 16 printf ("Area: %d ", area); 17 18 return 0; 19 }</pre>	<p>Esecuzione</p>  <pre> digita base: 4 digita altezza: 2 Area: 4</pre>

Figura 2. Dal problema al programma

Questa prima analisi ha il solo scopo di verificare che tutte le regole imposte dal linguaggio siano state rispettate nella stesura del programma, cosicché eventuali errori possano essere immediatamente segnalati al programmatore, che provvede alla conseguente correzione. All'analisi sintattica di un programma, segue l'*analisi semantica*, attraverso la quale il compilatore verifica che gli elementi definiti nel codice, come ad esempio le variabili, siano stati correttamente utilizzati.

Un tipico esempio è dato dal controllo dei tipi delle variabili, ovvero se un elemento è stato dichiarato di tipo intero e gli viene assegnato un valore reale, ci potrebbe essere una perdita di informazioni, poiché la parte decimale del valore verrebbe troncata e, quindi, persa. Una volta garantita la correttezza sintattica e semantica di un programma, il compilatore procede con la fase di traduzione vera e propria, in cui le istruzioni di alto livello scritte dal programmatore vengono riformulate in termini di operazioni più semplici dette istruzioni macchina. Il linguaggio macchina è quanto di più vicino esista al codice realmente interpretato da un processore, che abbia ancora una forma facilmente leggibile da un essere umano. Ad esempio, l'espressione $a=b+c$, sarebbe rappresentata dall'istruzione `ADD $0xd1, $0xd2, $0xd3`, dove `ADD` indica una operazione di somma e `$0xd1`, `$0xd2` e `$0xd3` rappresentano i registri della CPU, in cui sono memorizzati i valori di a , b e c . Se nemmeno il linguaggio macchina, anche noto come *assembly*, rappresenta il codice

effettivamente eseguito da un processore, è legittimo chiedersi che aspetto abbia, allora, tale codice. Una semplice riflessione su come sia realmente fatta una CPU, conduce ad una risposta pressoché immediata a tale domanda.

I primi computer erano realizzati con serie di valvole, ovvero dispositivi in grado di lasciar passare o meno un segnale elettrico. La combinazione delle valvole accese o spente in una serie rappresentava uno specifico dato. Benché, nei processori moderni, le valvole siano state sostituite dai transistor, ovvero dispositivi estremamente più piccoli, al punto da poterne avere circa 30 miliardi in un chip di 5 nanometri, la funzione di questi ultimi è esattamente la stessa, ovvero riconducibile all'assunzione di uno stato acceso/spento. Si immagini quindi di rappresentare questi due stati con due valori distinti, ossia 0 spento e 1 acceso, se un dato è rappresentato dalla combinazione acceso/spento di una serie di transistor, esso potrà essere espresso in una forma del tipo 10001110.... In realtà, secondo lo schema che John von Neumann ha formulato per rappresentare una moderna architettura di calcolo (Von Neumann, 1993), anche le istruzioni rappresentano dei dati, in fine dei conti, e al pari di essi possono essere rappresentate in memoria esattamente nella stessa forma. Ne deriva che, il sistema di rappresentazione che più si adatta a rappresentare il codice elaborato da una CPU è quello binario, dove le due cifre di base 0 e 1, prendono il nome di *digit*. Il problema originario, fornito da un qualsivoglia contesto del mondo *reale*, viene risolto da un essere umano che intuisce una soluzione e ne fornisce una rappresentazione algoritmica, la quale poi, sulla base di un qualsiasi linguaggio di programmazione, è implementata in un programma, che a sua volta viene tradotto in codice binario da un compilatore, diventando così *digitale*.

Il codice binario, tuttavia, è solo una diversa rappresentazione dell'algoritmo e non costituisce di per sé la soluzione al problema. La sua applicazione ad un insieme di dati concreti, ovvero una sua *istanza* e la relativa *esecuzione* (vedi Figura) con i risultati che ne conseguono rappresentano l'elemento che chiude il ciclo.

In un sistema multitasking e multi-user, un programma non è una monade, ma coesiste con altri programmi anch'essi in esecuzione concorrente e in totale trasparenza, sotto forma di *processi*. Ne deriva che i diversi programmi competono per l'utilizzo delle risorse fisiche messe a disposizione dall'hardware, rendendo evidente la necessità di una sorta di arbitro che orchestri le risorse disponibili al fine di garantire l'equilibrio fra i processi e la stabilità dell'intero sistema. Questo arbitro è anch'esso un processo, che prende il nome di *sistema operativo* ed esegue al di sopra di tutti gli altri con maggiori privilegi (Silberschatz et al., 2014). Ci si potrebbe chiedere quale sia il vantaggio di avere un programma che funge da intermediario fra un programma utente e la macchina fisica. In realtà, i vantaggi sono diversi. Primo fra tutti, il sistema operativo garantisce che tutti i programmi eseguano e

portino a termine il proprio compito. In secondo luogo, solleva il programmatore dall'onere di gestire tutta una serie di aspetti legati all'esecuzione di un programma, quali la gestione diretta degli indirizzi di memoria in RAM o la comunicazione con le periferiche per il caricamento e il salvataggio/stampa dei risultati.

La Figura 3 mostra l'architettura di un sistema di calcolo ed il ruolo del sistema operativo in tale contesto. Fra le diverse mansioni del sistema operativo, si annovera quella di mandare in esecuzione i processi attraverso uno dei suoi moduli, che prende il nome di *scheduler*. In base a dei criteri specifici di ciascun sistema operativo, lo scheduler seleziona un processo (collegato ad un programma) e lo assegna ad una CPU, la quale procede con l'esecuzione sequenziale delle sue istruzioni (Harris *et al.*, 2019).

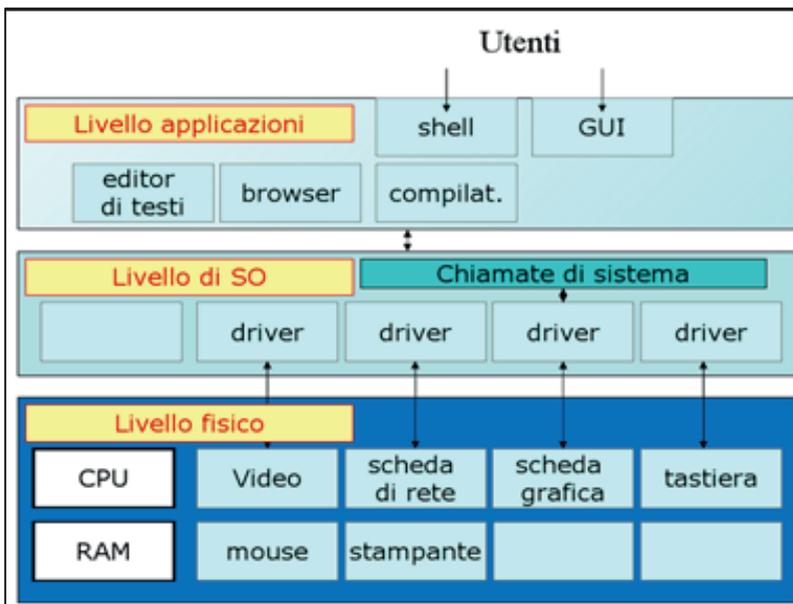


Figura 3. Architettura di un sistema operativo.

La Figura 4 mostra una rappresentazione schematica molto semplificata di una CPU. Il *program counter* (indicato con PC) è semplicemente un contatore che tiene traccia dell'istruzione corrente. L'istruzione, ovvero una sequenza di 32 bit (64 su una CPU a 64 bit) è letta dalla memoria e caricata in un registro. Successivamente, gruppi di bit sono interpretati secondo uno schema ben preciso e, a seconda del valore 1 o 0, attivano o meno segnali lungo un *bus*, ovvero una pista di connessioni che collegano le diverse componenti della CPU. Parte di questi bit permettono all'unità di controllo, che monitora il funzionamento del processore, di attivare o meno delle componenti, mentre l'unità di calcolo logico/aritmetica (detta ALU) esegue le

operazioni. Il risultato di una singola operazione è memorizzato nuovamente in un registro e successivamente trasferito alla RAM.

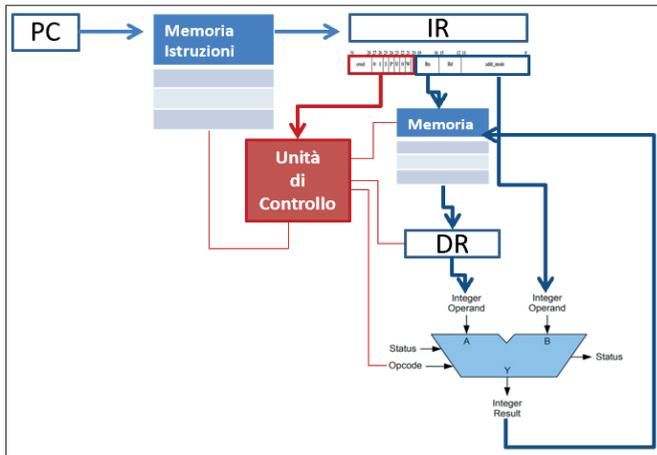


Figura 4. Rappresentazione schematica di una CPU

Human in the loop

Un importante aspetto ancora da chiarire è il ruolo che l'uomo riveste nel ciclo problema-soluzione, in quali momenti egli prende parte a questo ciclo e con quali obiettivi (vedi Fig. 5).

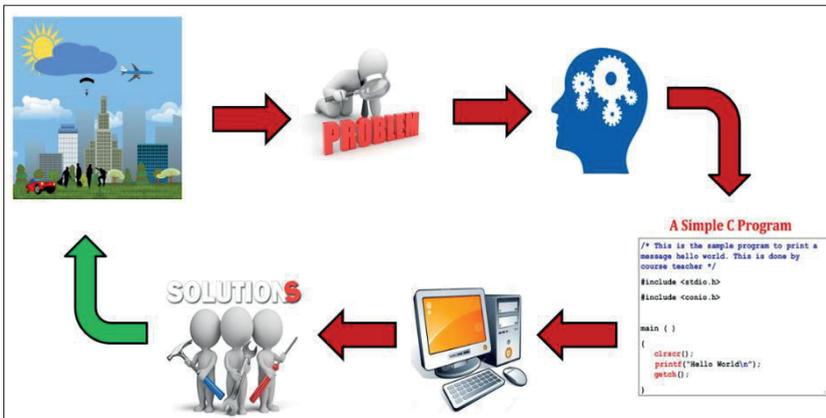


Figura 5. Il ciclo problema-soluzione.

Sulla base di quanto detto nei precedenti paragrafi, risulta evidente che l'essere umano è parte fondamentale di questo ciclo, principalmente nelle sue fasi iniziali. Infatti, una caratteristica che è ancora prerogativa esclusiva dell'uomo e di cui la macchina è completamente priva è la capacità di astrazione, ovvero la capacità di

partire da una esperienza empirica e formalizzare un modello, che descriva le regole e le modalità secondo cui un fenomeno si manifesta.

Alan Turing, uno dei padri fondatori dell'informatica moderna, sosteneva, infatti, che i calcolatori eseguono le operazioni più velocemente degli esseri umani, ma sanno eseguire solo operazioni descritte in maniera precisa e formale (Alan Turing, 1936). Essi non sono in grado di progettare procedure per risolvere problemi, a differenza degli esseri umani. Per questa ragione l'informatica, più in generale, può essere vista come la scienza avente come obiettivo la risoluzione di problemi e la cui esistenza non è affatto legata a quella dei computer. Il noto informatico olandese Edsger Wybe Dijkstra sosteneva infatti: *“Computer science is no more about computer than astronomy is about telescopes”*. Per convincersi di questa affermazione è sufficiente ripercorrerne la storia, concentrandosi su alcuni elementi fondamentali. La definizione di *algoritmo*, come sequenza finita di azioni elaborative che portano alla realizzazione di un compito, nasce a cavallo fra il 700 e l'800 d.c. con gli studi di algebra dell'iraniano Muhammad ibn Musa al-Khwarizmi, dal cui nome espresso in latino deriva il termine stesso “algoritmo”. Tuttavia, macchine che potessero anche solo lontanamente assomigliare ad un calcolatore compaiono sulla scena solo secoli dopo, quando Charles Babbage presenta la *Analytical Engine* (1824), ovvero una macchina “general-purpose”, le cui funzionalità dipendono da come è programmata. La realtà è che lo stesso Babbage non aveva intuito completamente il potenziale della macchina da lui presentata, la quale poteva essere erroneamente confusa con una delle tante macchine calcolatrici seguite alla pascalina originariamente presentata da Pascal nel 1642. Venuta a conoscenza di tale macchina e affascinata dal suo funzionamento Ada Lovelace, collaborando con Babbage, scrive il primo algoritmo per calcolare i numeri di Bernoulli con la macchina di Babbage, oggi considerato il primo software della storia. La Figura 6 mostra le immagini di alcuni studiosi eminenti che hanno contribuito alla nascita dell'informatica moderna.



Figura 6. Studiosi che hanno dato un contributo fondamentale nella nascita dell'informatica moderna. Da sinistra a destra, nell'ordine: Muhammad ibn Musa al-Khwarizmi, Charles Babbage, Ada Lovelace, John von Neumann, Alan Turing e Edsger Wybe Dijkstra.

Dal Problema alla Soluzione

In accordo alla visione di Turing una macchina è estremamente efficiente nell'applicazione di una soluzione algoritmica ad un insieme di dati, così da ridurre enormemente i tempi necessari ad ottenere l'output una volta dato l'input. In effetti, è esattamente quello che la sua macchina realizzava nella decifrazione del codice Enigma. Benché programmabile e altamente efficiente, la macchina resta priva delle capacità di astrazione, intuizione e progettazione, che restano prerogativa dell'uomo e sono alla base del *problem solving*, ovvero di quel processo alla base del ciclo problema-soluzione (Fedorenko et al., 2019).

Con il termine *problema* si intende una situazione che pone delle domande cui si devono dare risposte. Risolvere il problema vuol dire uscire da tale situazione. Un problema consta diversi elementi: i) i *dati*, che costituiscono l'istanza che si deve affrontare e generalmente indicata col termine input, ii) i *risultati*, ovvero gli elementi incogniti la cui determinazione fornisce l'output e iii) i *vincoli* che rappresentano le limitazioni cui devono soggiacere i risultati. La formulazione del problema determina la strategia risolutiva migliore da adottare per risolverlo. Da un punto di vista formale, una soluzione, e conseguentemente, un algoritmo può essere considerato come una trasformazione f che prende in input un insieme di dati x e produce dei risultati y . La risoluzione del problema si riduce, quindi, alla determinazione di una funzione tale che $y=f(x)$.

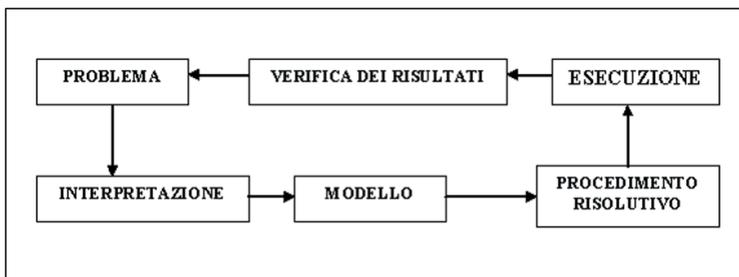


Figura 7. L'insieme dei passi alla base del *problem solving*.

Disporre di una rappresentazione formale così rigorosa di un algoritmo permette di valutarne la bontà da un punto di vista quantitativo ed assoluto. Inoltre, qualora si disponga di più soluzioni diverse ad uno stesso problema, è possibile operare una scelta sulla base di criteri oggettivi al fine di determinare la migliore fra le soluzioni possibili. In particolare, nel confronto di due algoritmi che risolvono lo stesso problema è necessario fissare il criterio rispetto al quale effettuare la valutazione. Se si considerasse il tempo di calcolo nel caso peggiore, il principale limite risiederebbe nell'avere due diverse misurazioni, ovvero la *velocità relativa*, quando si considera

l'esecuzione di entrambi gli algoritmi sulla stessa macchina e la *velocità assoluta*, quando l'esecuzione avviene su macchine diverse. In altri termini, sarebbe estremamente difficile, se non impossibile ottenere misurazioni assolute del tempo di esecuzione di un algoritmo. Un criterio alternativo, che è anche quello riconosciuto ed usato comunemente nel calcolo della complessità degli algoritmi è il tasso di crescita $T(n)$ del numero di operazioni effettuate da un algoritmo per risolvere un problema di taglia n , con n che tende a infinito. Si fa ricorso, quindi, ad una analisi asintotica (Bovet and Crescenzi, 1991).

Al fine di avere una idea più chiara, si consideri un problema estremamente comune, ovvero l'ordinamento (Karunanithi, 2014). In particolare, si supponga di voler ordinare un comune mazzo di carte. Esistono diversi metodi per effettuare questa operazione, e ciascuno adotta l'approccio a lui più congeniale. Tuttavia, è semplice verificare che non tutti i metodi hanno la stessa efficienza. Si considerino due dei più comuni approcci. In un primo metodo, si scorre l'intero mazzo di carte, si seleziona la carta con valore più piccolo e la si pone da parte. Si ripete questa operazione finché il mazzo di carte da scorrere non è vuoto. Dall'analisi dell'algoritmo descritto formalmente in Figura 8, si evince che il numero di operazioni da effettuare è dell'ordine di n^2 .

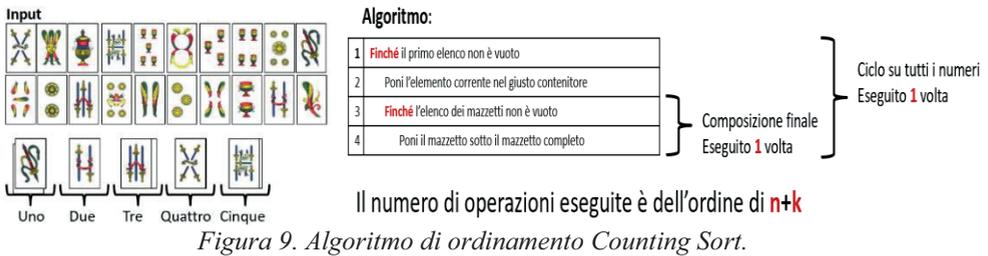


Figura 8. Algoritmo di ordinamento Selection Sort.

Si consideri un secondo approccio, in cui, piuttosto che riporre le carte selezionate in un singolo mazzo ordinato, le carte vengono suddivise in tanti mazzetti diversi a seconda del valore della carta, via via che si estraggono dal mazzo di carte di partenza. Alla fine, i diversi mazzetti ottenuti si dispongono in pila a partire da quello con le carte di valore maggiore, fino a quello con le carte di valore minimo. Analizzando l'algoritmo, formalmente descritto in Figura 9, si osserva che il numero di operazioni eseguite dall'algoritmo è dell'ordine di n .

Benché sia evidente che la seconda soluzione sia migliore della prima, ci si potrebbe chiedere quale sia il reale impatto del diverso numero di operazioni richieste dai due algoritmi sui tempi di esecuzione di un programma reale. I valori riportati in **Errore. L'origine riferimento non è stata trovata.**¹⁰ mostrano i tempi richiesti da un

programma al variare della complessità dell'algoritmo e della taglia dell'input su una macchina che richiede 1µsec per eseguire una singola operazione.



L'analisi asintotica rappresenta, quindi, un valido strumento per valutare la complessità di un algoritmo. Tuttavia, il fatto che alcuni algoritmi per la soluzione di un problema siano particolarmente inefficienti, non implica necessariamente che si tratti di un problema particolarmente complesso. In ogni caso, la complessità di un problema è legata alla complessità degli algoritmi che lo risolvono. Infatti, la complessità di un problema è definita come la complessità del miglior algoritmo noto per la sua soluzione.

complessità \ dim. input	10	10^3	10^6
costante - $O(1)$	1 µsec	1 µsec	1 µsec
logaritmica - $O(\lg n)$	3 µsec	10 µsec	20 µsec
lineare - $O(n)$	10 µsec	1 msec	1 sec
n lg n - $O(n \lg n)$	33 µsec	10 msec	20 sec
quadratica - $O(n^2)$	100 µsec	1 sec	10^{12} sec
cubica - $O(n^3)$	1 msec	10^9 sec	10^{18} sec
esponenziale - $O(2^n)$	10 msec	10^{301} sec	10^{301030} sec

11.6 gg

16.7 min

31709 anni !!

Figura 10. Tempi di esecuzione di un programma al variare della complessità dell'algoritmo che implementa, supponendo che la singola operazione richieda 1µsec.

In letteratura esistono esempi particolarmente interessanti, i quali dimostrano come problemi apparentemente semplici possano, in realtà, nascondere una complessità davvero elevata. Fra questi, figura certamente la *Torre di Hanoi* (Hinz et al., 2013). Si tratta di un gioco con regole molto semplici presentato nelle vesti di una leggenda, secondo cui all'inizio dei tempi, Brahma, divinità indiana, portò nel grande tempio di Benares, sotto la cupola d'oro che si trova al centro del mondo, tre colonnine di

diamante e sessantaquattro dischi d'oro collocati su una di queste colonnine in ordine decrescente, dal più piccolo in alto, al più grande in basso. Si trattava della sacra Torre di Brahma che vede impegnati, giorno e notte, i sacerdoti del tempio nel trasferimento della torre di dischi dalla prima alla terza colonnina. Essi non devono contravvenire alle regole precise, imposte da Brahma stesso, che richiedono di spostare soltanto un disco alla volta e che non ci sia mai un disco sopra uno più piccolo. Quando i sacerdoti avranno completato il loro lavoro e tutti i dischi saranno riordinati sulla terza colonnina, la torre e il tempio crolleranno e sarà la fine del mondo. La domanda, quindi, è: “c’è da temere l’imminente fine del mondo?”. La proprietà matematica di base è che il numero minimo di mosse necessarie per completare il gioco è $2^n - 1$, dove n è il numero di dischi. Ad esempio, avendo 3 dischi, il numero minimo di mosse è 7. Di conseguenza, secondo la leggenda, i monaci di Hanoi dovrebbero effettuare almeno 18.446.744.073.709.551.615 mosse prima che il mondo finisca, essendo $n = 64$. In altre parole, anche supponendo che i monaci facciano una mossa al secondo il mondo finirà tra 5.845.580.504 secoli.

Un problema molto più concreto, ma non meno complesso, è noto col nome di *Traveling Salesman Problem* (Problema del commesso viaggiatore) (Lin, 1965). Dato un insieme di città, e note le distanze tra ciascuna coppia di esse, si vuole trovare il tragitto di minima percorrenza che un commesso viaggiatore deve seguire per visitare tutte le città una ed una sola volta e ritornare alla città di partenza. Non esistono algoritmi efficienti per la risoluzione di questo problema. L'unico metodo di risoluzione è rappresentato dall'*enumerazione totale*, ovvero valutare tutti i possibili cammini per la scelta di quello migliore, il che implica una complessità dell'ordine di 2^n .

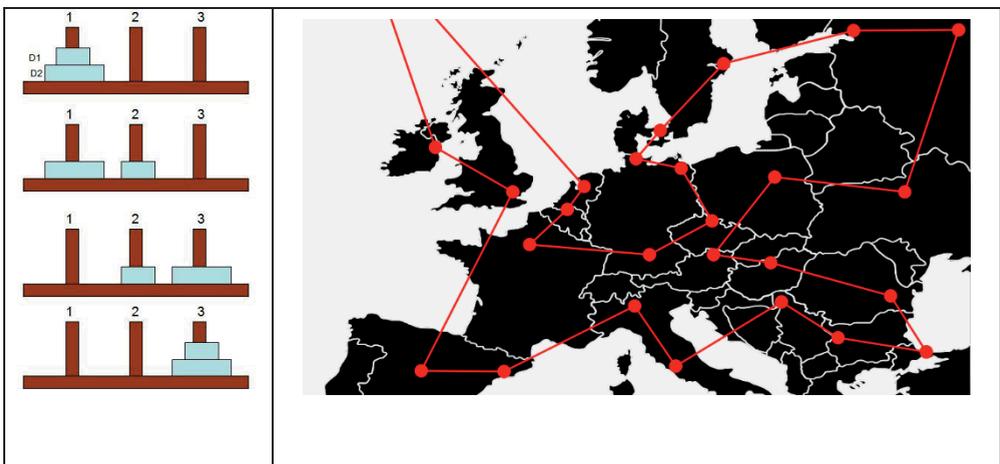


Figura 7. Esempi di problemi complessi. La Torre di Hanoi a sinistra e il commesso viaggiatore a destra.

Bibliografia

- Von Neumann, J. (1933). First Draft of a Report on the EDVAC. *IEEE Annals of the History of Computing*, 15(4), 27-75.
- Watt, D. A., Wong, S. (1990). *Programming Languages. Concepts and Paradigms* Prentice Hall.
- Morzenti, A., Breveglieri, L., Reghizzi, S. C. (2019). *Linguaggi Formali e Compilazione*. Società Editrice Esculapio.
- Harris, D. M., Harris, S. L., Prinz, P., Crawford, T. (2019). *Digital design and computer architecture*.
- Silberschatz, A., Galvin, P. B., & Gagne, G. (2014). *Operating system concepts essentials*. John Wiley & Sons, Inc.
- Alan Turing (1936). On computable numbers, with an application to the Entscheidungsproblem, *Proc. London Math. Soc.*, 42 pp. 230–265.
- Fedorenko, E., Ivanova, A., Dhamala, R., Bers, M. U. (2019). The language of programming: a cognitive perspective. *Trends in cognitive sciences*, 23(7), 525-528.
- Karunanithi, A. K. (2014). A survey, discussion and comparison of sorting algorithms. Department of Computing Science, Umea University.
- Bovet D. P., Crescenzi P., (1991). *Teoria della complessità computazionale (Vol. 7)*. Franco Angeli.
- Hinz, A. M., Klavžar, S., Milutinović, U., Petr, C. (2013). *The Tower of Hanoi-Myths and Maths (pp. I-XV)*. Basel: Birkhäuser.
- Lin, S. (1965). Computer solutions of the traveling salesman problem. *Bell System Technical Journal*, 44(10), 2245-2269.

Alcuni esempi di “quotidiana” probabilità

Aniello Buonocore

Dipartimento di Matematica e Applicazioni “Renato Caccioppoli”,

Università di Napoli Federico II

Abstract

Lo scopo di questo contributo è quello di mostrare un approccio per problemi all’insegnamento dei contenuti della probabilità presenti nelle indicazioni nazionali con particolare riferimento alla probabilità condizionata. Così, dopo aver delineato il significato dei concetti di base del calcolo delle probabilità (esperimento e suoi possibili risultati, eventi e loro probabilità) e prendendo spunto dai risultati ottenuti nel contesto del primo problema, si presenta la legge delle probabilità composte e come da essa discende la formula di Bayes. Il secondo problema, invece, fornisce un esempio di applicazione del ragionamento Bayesiano: probabilità a priori + verosimiglianza + formula di Bayes per ottenere la probabilità a posteriori desiderata. Nel terzo e ultimo problema, oltre a fare nuovamente uso della legge delle probabilità composte, si mostra un modo per “intrecciare” contenuti di diversi temi e per stimolare gli obiettivi di “processo” (argomentare e dimostrare, comunicare, rappresentare, collegare) spesso trascurati a favore dei soli obiettivi di “contenuto”.

Introduzione

La probabilità di un evento condizionata dal verificarsi di un altro evento è uno degli obiettivi specifici di apprendimento presente nel tema “dati e previsioni” delle indicazioni nazionali per i licei. La grande varietà di problemi basati su questo concetto, che sono stati proposti a più riprese e con diverse varianti (ad esempio, il paradosso dei due figli, il problema dei tre prigionieri, il problema delle due buste), può permettere all’insegnante di costruire percorsi didattici stimolanti e anche ricchi di collegamenti tra i vari temi della matematica. Per il problema delle due buste si veda, (Nalebuff, 1989). D’altra parte, nelle linee guida degli istituti professionali e tecnici tra le conoscenze si menzionano anche il ragionamento induttivo e le basi concettuali dell’inferenza. Non è azzardato affermare che tra la probabilità condizionata e il ragionamento induttivo la formula di Bayes funge da anello di congiunzione. Tale modalità di insegnamento/apprendimento può essere d’aiuto sia agli insegnanti per mitigare le difficoltà che loro stessi sovente riscontrano

nella concettualizzazione (ovvero quella discontinuità tra i processi cognitivi degli studenti e l'astrazione)¹ e sia agli stessi studenti per meglio orientarsi quando le circostanze della vita quotidiana sicuramente lo richiederanno.²

Esperimento aleatorio e operazioni con gli eventi

Un dispositivo di qualsivoglia natura che fornisce risultati dei quali non si conosce una legge per predirli si dice *aleatorio*. La messa in azione di uno di questi dispositivi viene anche denominata *esperimento aleatorio*.

Tipico esempio di esperimento aleatorio è quello del lancio di un dado (in latino *alea*) il cui risultato, detto *punteggio*, è il numero degli scavi presenti sulla faccia opposta a quella di atterraggio. Si riconosce allora che l'insieme dei suoi possibili risultati è:

$$\Omega = S_6 = \{1, 2, 3, 4, 5, 6\}. \quad ^3$$

¹ In (Benedetti, 2000) viene riportata un'esperienza didattica e a proposito del problema dei tre prigionieri si legge: *Questo genere di rompicapo sembra aver sviluppato in classe un clima che contribuisce all'implicazione degli studenti nell'attività matematica. Per di più la sua soluzione richiede, per essere compresa, un continuo cambio di registri, più rappresentazioni verbali e più rappresentazioni algebriche. Ciò arricchisce enormemente la trattazione formale del teorema di Bayes fornita dai manuali.*

² Per questo aspetto, mi piace provare a modificare un concetto espresso in (Gigerenzer, 2015). *Ai nostri figli viene insegnata l'algebra, la geometria e il calcolo infinitesimale: in altre parole, la matematica della certezza. È, invece, un po' trascurata quella dell'incertezza, ovvero il pensiero statistico. Dietro questa impostazione c'è l'idea che esercitarsi in discipline astratte migliora la qualità del pensiero e la capacità di risolvere problemi. Però, purtroppo abbiamo ancora molti medici che non comprendono le statistiche sanitarie e molti avvocati che non capiscono appieno la prova del DNA... Allora, ai nostri studenti bisogna insegnare parallelamente sia quello che è importante in matematica e sia quello che potrà essere utile per la vita quotidiana e per la professione. Ma l'aspetto che forse è maggiormente da prendere in considerazione è la modalità dell'insegnamento: quella orientata alla soluzione di problemi richiede in maniera precipua presentazioni chiare, strumenti pratici e linee guida intelligenti. Ciò è tanto più valido per la matematica dell'incerto e credo che la probabilità condizionata e il ragionamento bayesiano, sollecitando oltretutto l'insegnante a prendersi carico degli obiettivi di processo individuati dalle indicazioni nazionali, possa avere un ruolo importante a questo riguardo. Però, per realizzare cose così nuove bisogna ricercare con convinzione l'adesione degli insegnanti.*

³ Il termine *casuale* si utilizza nella particolare condizione di perfetta simmetria tra i punti campione. Nell'esempio del lancio del dado, l'esperimento è casuale se il dado è *onesto*: nessuna delle facce è preferibile rispetto all'altra.

Per un esperimento aleatorio, sia esso \mathcal{E} , si può allora dire che non è possibile predirne il risultato ma solo che questi deve appartenere ad uno specificato insieme Ω che è detto *spazio campione*. Gli elementi di Ω sono chiamati *punti campione*.

Nel caso di uno spazio campione con cardinalità finita ogni suo sottoinsieme è detto *evento*; negli altri casi sono eventi solo alcuni sottoinsiemi dello spazio campione.

Ad esempio, nell’esperimento del lancio di un dado, nel seguito indicato con \mathcal{L}_6 , il sottoinsieme $\{2, 4, 6\} \subset S_6$ è l’evento che corrisponde al presentarsi di un “punteggio *pari*”. Ulteriori esempi sono gli eventi $\{1, 2\}$ (ovvero, si presenta un “punteggio *basso*”) e $\{2\}$ (ovvero, si presenta il “punteggio *con due scavi nell’avorio*”). È utile osservare che, ad esempio, $\{1, 3, 5\} = \{2, 4, 6\}^c$ (l’evento “punteggio *dispari*” è il negato dell’evento “punteggio *pari*”) e che

$$S_6 = \{1, 3, 5\} \cup \{2, 4, 6\}$$

e quindi, come risultato di un lancio si presenta *certamente* o un “punteggio *pari*” oppure un “punteggio *dispari*”. Allo stesso modo, $S_6 = \{1,2\} \cup \{3,4\} \cup \{5,6\}$. Infine, come esempio di evento congiunto, l’evento “punteggio *pari basso*” si ottiene come $\{2,4,6\} \cap \{1,2\} = \{2\}$.

In definitiva, per un esperimento aleatorio \mathcal{E} valgono le seguenti affermazioni:

- non è possibile predirne il risultato ma solo che esso deve appartenere ad uno specificato insieme Ω che è detto spazio campione;
- ad ogni evento che si può definire sull’esperimento \mathcal{E} si associa in maniera biunivoca un sottoinsieme di Ω ;
- gli eventi si compongono con le usuali operazioni insiemistiche;
- il negato di un evento si ottiene eseguendo l’operazione di complemento sul sottoinsieme corrispondente;
- la disgiunzione logica di due eventi si ottiene eseguendo l’operazione di unione dei due sottoinsiemi corrispondenti;
- la congiunzione logica di due eventi si ottiene eseguendo l’operazione di intersezione dei due sottoinsiemi corrispondenti;
- nel caso di uno spazio campione Ω con cardinalità finita, tutti i suoi sottoinsiemi sono eventi.

Definizione della probabilità e relative proprietà comuni

Si consideri un esperimento casuale \mathcal{E} con relativo spazio campione dotato di un numero finito di punti campione perfettamente simmetrici tra loro (nel senso che nessuno di essi ha prevalenza di uscita rispetto a tutti gli altri).

In tali condizioni, la probabilità di un evento può essere calcolata *rapportando la cardinalità dell'insieme corrispondente all'evento considerato alla cardinalità dello spazio campione*. In simboli, se A è l'evento in questione, Ω lo spazio campione e $\mathbb{P}^{(c)}$ la misura della probabilità, allora

$$\mathbb{P}^{(c)}(A) = \frac{|A|}{|\Omega|} \quad 4$$

Quando non è possibile riscontrare simmetrie tra i punti campione la definizione classica non può essere utilizzata.

In questi casi, se l'esperimento casuale \mathcal{E} può essere ripetuto un numero indefinito di volte nelle medesime condizioni, si può ricorrere alla definizione *frequentista*.

Sia prefissato un intero positivo n e sia A l'evento di cui si vuole valutare la probabilità. In primo luogo, è necessario ripetere, nelle medesime condizioni, n volte l'esperimento \mathcal{E} ; ognuna di queste ripetizioni è chiamata *prova*. Poi, indicato con n_A il numero delle volte in cui A si è presentato, si considera il rapporto n_A/n come un'approssimazione della probabilità dell'evento; tale approssimazione migliora al crescere del numero n delle ripetizioni di \mathcal{E} .

Infine, si pone:

$$\mathbb{P}^{(f)}(A) = \lim_{n \rightarrow \infty} \frac{n_A}{n}$$

La coppia di apici che racchiude il segno di limite rappresenta il fatto che non si tratta del limite di una successione assegnata mediante una formula e per passare da un elemento della successione al suo successivo bisogna effettuare un'altra prova.

Si osservi che quando un esperimento è casuale e ripetibile si può applicare sia la definizione classica che quella frequentista.

Quando l'esperimento aleatorio non presenta simmetrie tra i punti campione e non è ripetibile un numero indefinito di volte nelle medesime condizioni allora si può ricorrere alla definizione soggettiva della probabilità di un evento.

Tale definizione adotta il contesto delle scommesse (ad esempio, quelle sulle corse dei cavalli) e il suo linguaggio:

la probabilità di un evento A è il prezzo $\mathbb{P}^{(s)}(A)$, compreso tra 0 e 1, che un individuo coerente ritiene equo pagare per ricevere 1 nel caso che A si verifichi.

⁴ Si tratta della definizione classica delle probabilità: essa si può applicare solo nel caso di esperimenti casuali.

Le condizioni di coerenza ed equità si sono rese necessarie per superare l’obiezione che non si può fondare una teoria sulla base di opinioni personali. Si osservi che una volta fissata la probabilità di un evento è fissata anche quella del suo negato. Quindi, ci sono sempre almeno due scommesse possibili. La condizione di coerenza afferma che non deve essere possibile individuare un numero (finito) di scommesse che possano garantire a chi ne ha fissato le probabilità una vincita (o una perdita) certa. La condizione di equità è, invece, basata sul concetto di scommessa equa: una scommessa è equa quando chi fissa la probabilità dell’evento (banco) accetta di scambiare il ruolo con lo scommettitore (giocatore) nel caso in cui questi lo dovesse richiedere.

Si indichi con $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ la famiglia degli eventi e con \mathbb{P} una qualsiasi delle misure della probabilità in precedenza descritte.

Le varie definizioni di probabilità sono del tutto differenti tra loro e, come già evidenziato, la loro applicabilità dipende dal tipo di esperimento aleatorio. Esse, però, condividono le seguenti tre proprietà:

- 1) $\forall A \in \mathcal{F}, 0 \leq \mathbb{P}(A) \leq 1$; (la probabilità di un qualsiasi evento è un numero reale compreso tra 0 e 1);
- 2) $\mathbb{P}(\Omega) = 1$; (l’evento certo ha probabilità 1);
- 3) $\forall A, B \in \mathcal{F}: A \cap B = \emptyset, \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ (*finita additività*).

Queste proprietà costituiscono la base per una trattazione di carattere generale del calcolo delle probabilità.

Un gioco tra amici

Cinque persone (*giocatori*) si accordano sulla ripetizione (*partita*) del seguente gioco. Un’urna è riempita con 5 biglie del tutto uguali tra loro tranne che per il colore: 4 sono bianche e 1 sola è nera. A turno estraggono una biglia dall’urna senza rimpiazzamento. Vince la partita la persona che estrae la biglia nera. Viene stabilito (in base alla *prepotenza*) l’ordine di estrazione vigente per ciascuna partita: G_1, G_2, G_3, G_4, G_5 .

Tenendo conto che ci si trova in condizioni di perfetta simmetria tra i punti campione, si richiede di valutare la probabilità che a vincere una partita siano, il primo, il secondo e il terzo giocatore, rispettivamente.

Si indichi con N l’evento “si estrae biglia nera” e con B l’evento “si estrae biglia bianca”. La prima richiesta ha una immediata risposta:

$$\mathbb{P}(G_1) = \mathbb{P}(N_1) = \frac{1}{5}.$$

Nella formula precedente il pedice 1 alla lettera N si riferisce alla prima estrazione.

Allo stesso modo si può impostare la risposta alla seconda richiesta:

$$\mathbb{P}(G_2) = \mathbb{P}(B_1 \cap N_2).$$

Nella formula precedente i pedici 1 e 2 alle lettere B e N si riferiscono alla prima e alla seconda estrazione, rispettivamente. Si proceda ora in maniera intuitiva seguendo l'ordine delle estrazioni: dopo l'uscita di una biglia bianca l'urna del secondo giocatore contiene 3 biglie bianche e 1 nera, per cui:

$$\mathbb{P}(G_2) = \mathbb{P}(B_1 \cap N_2) = \frac{4}{5} \cdot \frac{1}{4} = \frac{1}{5} \quad 1)$$

In maniera analoga, considerando che l'urna del terzo giocatore contiene 2 biglie bianche e 1 nera, si ha:

$$\mathbb{P}(G_3) = \mathbb{P}(B_1 \cap B_2 \cap N_3) = \frac{4}{5} \cdot \frac{3}{4} \cdot \frac{1}{3} = \frac{1}{5} = 0,2.$$

Ecco, un esempio nel quale la prepotenza non sortisce alcun beneficio!

Probabilità condizionata e formula di Bayes

È possibile ottenere una formalizzazione della formula (1). Infatti, basta fornire un pedice, ad esempio A , alla lettera \mathbb{P} per rappresentare la probabilità di un altro evento, sia esso B , quando si è acquisita l'informazione che l'evento A si è già verificato: $\mathbb{P}_A(B)$. Allora, nell'esempio in questione, si può scrivere

$$\mathbb{P}(B_1 \cap N_2) = \mathbb{P}(B_1) \mathbb{P}_{B_1}(N_2).$$

In generale, la *legge delle probabilità congiunte* asserisce che:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \mathbb{P}_A(B). \quad (2)$$

Ovviamente, dal momento che $A \cap B = B \cap A$, risulta:

$$\mathbb{P}(A \cap B) = \mathbb{P}(B \cap A) = \mathbb{P}(B) \mathbb{P}_B(A). \quad (3)$$

Uguagliando i secondi membri della (2) e della (3), se $\mathbb{P}(A) > 0$, si ottiene la famosa *formula di Bayes*:

$$\mathbb{P}_A(B) = \mathbb{P}(B) \cdot \frac{\mathbb{P}_B(A)}{\mathbb{P}(A)} \quad (4)$$

Infine, si osservi che iterando il procedimento con gli eventi pedice nel segno della misura di probabilità, si formalizza la probabilità del congiunto di tre eventi: ad esempio,

$$\mathbb{P}(B_1 \cap B_2 \cap N_3) = \mathbb{P}(B_1)\mathbb{P}_{B_1}(B_2)\mathbb{P}_{B_1 \cap B_2}(N_3).$$

Dalle leggere alle pesanti

Nel mese di dicembre del 1992 i professori Carla Rossi e Romano Scozzafava inviano ad alcuni dei maggiori quotidiani nazionali una lettera dal provocatorio titolo “Dalle leggere alle pesanti?” (Rossi e Scozzafava, 1992). Come si intuisce abbastanza facilmente essi intendono “intramettersi” nel dibattito, in quel frangente molto controverso della dialettica politico-sociale, avvertito di grande importanza da parte dell’opinione pubblica sulla punibilità e sul recupero dei tossicodipendenti. Ovviamente, la loro finalità era “solo” quella di smascherare una fallacia nel ragionamento “dalle leggere alle pesanti”:

Gli aspetti emotivi prevalgono su quelli razionali, si indaga sulle storie individuali e si interroga il consumatore di droghe per sapere, in particolare, se fumava spinelli (cioè le cosiddette droghe “leggere”, come hashish e marijuana) prima di diventare tossicodipendente. La risposta è di solito affermativa, e nei dibattiti questo fatto viene regolarmente usato per sostenere la necessità di proibire anche queste sostanze, che sarebbero la “porta d’ingresso” alla tossicodipendenza.

Ora, se anche uno volesse fare l’avvocato del diavolo per dire che i fautori del proibizionismo usassero l’argomento in maniera strumentale, è stupefacente constatare come esso facesse (e fa ancora) molta presa sugli ascoltatori delle trasmissioni televisive e sui lettori dei giornali.

L’argomentazione con la quale gli autori della lettera demoliscono il ragionamento “dalle leggere alle pesanti” è di tipo paradossale: se alle stesse persone (tossicodipendenti) intervistate fosse stata posta la domanda “ti piaceva mangiare le caramelle da bambino?” la percentuale di risposte affermative certamente sarebbe stata maggiore di quella ottenuta con la domanda sul consumo delle droghe leggere. Ma nessuno si sogna di mettere in relazione caramelle e droghe pesanti.

Alla fine della lettera gli autori forniscono una spiegazione più tecnica per la non validità del ragionamento “dalle leggere alle pesanti:

In conclusione, quindi, è fra i consumatori di spinelli che andrebbe fatta un’indagine per vedere quanti diventeranno consumatori di droghe “pesanti”, e non viceversa: ma è stata mai fatta una tale indagine?

Da parte mia, faccio osservare che il disegno sperimentale (ovvero la procedura per ottenere i dati), che è del tipo “tempo corrente” - “qualche anno prima”, contribuisce

molto a ingenerare la convinzione della validità del ragionamento “dalle leggere alle pesanti”: le persone sono indotte a credere che è agli attuali tossicodipendenti che bisogna chiedere se in passato avessero fatto uso degli spinelli. È invece molto più difficile far accettare la modalità di raccolta dei dati proposta dai professori Rossi e Scozzafava: è agli attuali fumatori di spinelli che tra qualche anno bisognerà chiedere se nel frattempo siano diventati tossicodipendenti.

Ma, tale ulteriore indagine non è necessaria, come si può evincere con il seguente ragionamento basato sulla formula di Bayes. Infatti, indicata con Ω la popolazione di riferimento (ben individuata e circoscritta prima dell’avvio dell’indagine), la tesi che i fautori del proibizionismo volevano promuovere è che la probabilità $\mathbb{P}_S(T)$ che un fumatore di spinelli (evento S) diventi in futuro un tossicodipendente (evento T) fosse molto vicina a 1 e molto più grande della probabilità di essere un tossicodipendente nella totalità della popolazione Ω . In simboli:

$$\mathbb{P}(T) \ll \mathbb{P}_S(T) \cong 1. \quad (5)$$

Invece, il loro ragionamento è stato solo quello di “produrre” il risultato dell’indagine, ovvero la probabilità $\mathbb{P}_T(S)$ al tempo corrente e affermare che tale quantità è “impressionantemente” grande (ovvero, abbastanza vicino a 1). Ma, applicando la formula di Bayes, si ottiene

$$\mathbb{P}_S(T) = \mathbb{P}(T) \cdot \frac{\mathbb{P}_T(S)}{\mathbb{P}(S)} = \mathbb{P}(T) \cdot \alpha.$$

In altri termini, la probabilità che un fumatore di spinelli diventi tossicodipendente è proporzionale alla probabilità di essere tossicodipendente nella totalità della popolazione Ω . Il fattore di proporzionalità α è il rapporto tra la probabilità che un tossicodipendente sia stato fumatore di spinelli e la probabilità di essere un fumatore di spinelli nella totalità della popolazione Ω . Inoltre, per stimare $\mathbb{P}(T)$, si consideri che il numero degli utenti in carico annualmente dai “SerD” è compreso tra 10^5 e $2 \cdot 10^5$ che su una popolazione di qualche decina di milioni di persone fornisce $\mathbb{P}(T) \cong 0,005$. In definitiva, se anche si ponesse $\alpha = 3$, la (5) non risulterebbe corretta sia nel senso del molto maggiore che nel senso del circa uguale.

Un problema di sorteggio

Una classe è composta da 25 persone: 15 femmine e 10 maschi. Si devono estrarre a sorte due rappresentanti di classe. Qual è la probabilità che siano estratti un maschio e una femmina?

Prima di pensare ad una possibile soluzione, è utile rappresentare in maniera insiemistica le informazioni descritte nell’enunciato del problema. Allora, ad esempio, possiamo indicare con $M = \{m_1, m_2, \dots, m_{10}\}$ l’insieme degli alunni e con $F = \{f_1, f_2, \dots, m_{15}\}$ l’insieme delle alunne. Ovviamente, M e F sono disgiunti e la classe è rappresentata dall’insieme $\Omega = M \cup F$. Inoltre, l’evento $E_{10,15}$ del quale bisogna calcolare la probabilità è indicato nella frase “i rappresentanti di classe estratti sono un maschio e una femmina”.

Se l’esperimento che viene utilizzato per eseguire il sorteggio è costituito da due estrazioni singole sequenziali allora risulta:

$$E_{10,15} = (M_1 \cap F_2) \cup (F_1 \cap M_2). \quad (6)$$

La (6) deve essere interpretata al seguente modo: nella prima estrazione viene scelto un alunno e nella seconda un’alunna oppure nella prima estrazione viene scelto un’alunna e nella seconda un alunno. Ovviamente, gli operandi dell’unione nella (6) sono disgiunti. La proprietà di attività e la legge delle probabilità congiunte consentono di ottenere la soluzione:

$$\begin{aligned} p(10,15) &:= \mathbb{P}(E_{10,15}) = \mathbb{P}[(M_1 \cap F_2) \cup (F_1 \cap M_2)] \\ &= \mathbb{P}(M_1 \cap F_2) + \mathbb{P}(F_1 \cap M_2) = \mathbb{P}(M_1)\mathbb{P}_{M_1}(F_2) + \mathbb{P}(F_1)\mathbb{P}_{F_1}(M_2). \end{aligned}$$

In definitiva, utilizzando i dati del problema e valutando la probabilità, per ovvie ragioni di simmetria, come rapporto tra casi favorevoli e casi possibili, si ha:

$$p(10,15) = \frac{10}{25} \cdot \frac{15}{24} + \frac{15}{25} \cdot \frac{10}{24} = 2 \frac{10}{25} \cdot \frac{15}{24} = \frac{1}{2}. \quad (7)$$

A margine della soluzione, la Figura 1 fornisce una rappresentazione grafica della differenza dello schema del calcolo della probabilità dell’evento F_2 con quello relativo al calcolo della probabilità dello stesso evento noto che nella prima estrazione è uscito il nome di un alunno:

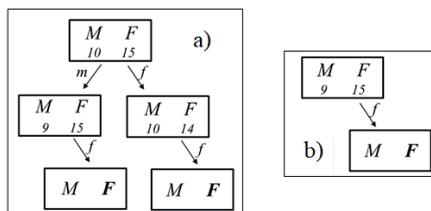


Figura 1. Parte a) - per la realizzazione dell’evento F_2 è necessario considerare le due possibili alternative dovute alla prima estrazione. Parte b) - è noto che è uscito il nome di un alunno e un nuovo spazio campionario (quello costituito da 9 alunni e 15 alunne) si è già concretizzato.

È interessante, a questo punto, la seguente domanda: perché proprio 0,5? Ovvero, come è collegato tale valore di probabilità al numero degli studenti e al numero delle studentesse? In altre parole, ci si chiede se esistono altre coppie di interi positivi per le quali la probabilità richiesta vale 0,5.

Per generalizzare il problema, si denoti con f il numero delle alunne e con $m = f - n$ il numero degli alunni; qui, n che rappresenta l'eccesso del numero di alunne rispetto al numero degli alunni. Ripercorrendo la linea della soluzione precedente, con corrispondente significato dei simboli nella (7), si ottiene la seguente equazione:

$$p(m, f) := \mathbb{P}(E_{m,f}) = \frac{1}{2} \Leftrightarrow 2 \frac{mf}{(m+f)(m+f-1)} = \frac{1}{2}.$$

Tale equazione, facendo comparire come incognita n al posto di m risulta essere equivalente a

$$f = \frac{n(n+1)}{2} \Leftrightarrow n^2 + n - 2f = 0, \quad (8)$$

che ha come unica soluzione positiva:

$$n = \frac{\sqrt{1+8f}-1}{2}. \quad (9)$$

Ovviamente, il valore di n fornito dalla (9) deve essere intero positivo e questo dipende dal valore assegnato a f . Ad esempio:

$$f = 15 \Rightarrow n = \frac{\sqrt{1+8 \cdot 15}-1}{2} = \frac{\sqrt{121}-1}{2} = \frac{10}{2} = 5 \Rightarrow m = f - n = 10,$$

ovvero i dati del problema enunciato all'inizio di questo paragrafo; invece,

$$f = 5 \Rightarrow n = \frac{\sqrt{1+8 \cdot 5}-1}{2} = \frac{\sqrt{41}-1}{2},$$

che è positivo ma non è intero.

Alcuni altri valori di f che forniscono $p(m, f) = 0,5$ si possono leggere in grassetto nella seguente Tabella 1. Si osservi che la coppia $f = 1$ e $n = 1$ conduce a $f = m - n = 0$, e quindi è da scartare. Gli altri valori di f che conducono a valori di n interi positivi costituiscono una successione $(f_n)_{n \in \mathbb{N}}$. Ad esempio,

$$f_1 = 1, f_2 = 3 \text{ e } f_6 = 21.$$

Sempre dalla Tab. 1 si può osservare che

$$f_2 = f_1 + 2, f_3 = f_2 + 3, f_4 = f_3 + 4, \dots$$

e questo porta a congetturare che la successione $(f_n)_{n \in \mathbb{N}}$ sia assegnata per ricorrenza:

$$f_1 = 1 \text{ e } n \geq 2, f_n = f_{n-1} + n.$$

<i>f</i>	<i>n</i>	<i>f</i>	<i>n</i>	<i>f</i>	<i>n</i>	<i>f</i>	<i>n</i>
1	1	16	5.178908	31	7.389867	46	9.104686
2	1.561553	17	5.35235	32	7.51561	47	9.208244
3	2	18	5.520797	33	7.63941	48	9.310708
4	2.372281	19	5.684658	34	7.761356	49	9.412114
5	2.701562	20	5.844289	35	7.881527	50	9.512492
6	3	21	6	36	8	51	9.611874
7	3.274917	22	6.152067	37	8.116844	52	9.710289
8	3.531129	23	6.300735	38	8.232125	53	9.807764
9	3.772002	24	6.446222	39	8.345903	54	9.904326
10	4	25	6.588723	40	8.458236	55	10
11	4.216991	26	6.728416	41	8.569179	56	10.09481
12	4.424429	27	6.86546	42	8.67878	57	10.18878
13	4.623475	28	7	43	8.787088	58	10.28193
14	4.815073	29	7.132169	44	8.894147	59	10.37428
15	5	30	7.262087	45	9	60	10.46586

Tabella 1. Sono evidenziati i valori di $1 \leq f \leq 60$ per i quali la (8) fornisce un valore di n intero. Ovviamente, $m = f - n$.

Per ottenere la formulazione esplicita del termine generale della successione $(f_n = f_{n-1} + n)_{n \in \mathbb{N}}$ è conveniente osservare che

$$k \geq 2, f_k = f_{k-1} + k \Leftrightarrow f_k - f_{k-1} = k.$$

Allora, per $1 \leq k \leq n$, si ottengono le n uguaglianze

$$f_1 = 1; f_2 - f_1 = 2; f_3 - f_2 = 3; \dots; f_n - f_{n-1} = n,$$

che sommate membro a membro fanno ottenere:

$$f_n = 1 + 2 + \dots + n = \frac{n(n+1)}{2} \quad 5 \quad (10)$$

La congettura è allora vera in virtù della (8). Il fatto che

$$m = f - n \Leftrightarrow f = m + n$$

consente così di affermare che anche il numero m degli alunni appartiene alla successione $(f_n)_{n \in \mathbb{N}}$.

In definitiva, le coppie di interi (m, f) per le quali risulta

$$p(m, f) = \frac{1}{2},$$

⁵ Si dice che f_n è il numero triangolare di ordine n e la (10) è la formula di Gauss

sono le coppie consecutive degli elementi della successione $(f_n)_{n \in \mathbb{N}}$:

$$\forall n \geq 2, (f_{n-1}, f_n).$$

Si riconosce, in particolare, che $(10,15) = (f_4, f_5)$.

Bibliografia

- Benedetti, C. (2000). Probabilità a scuola: un percorso di condivisione dei registri semiotici. Tesi di abilitazione all'Insegnamento nella Scuola Secondaria. SSIS-Università di Bologna.
- Gigerenzer, G. (2015). Imparare a rischiare. Come prendere decisioni giuste. Collana Scienze e Idee, Cortina Raffaello Editore, pp. XI–362.
- Nalebuff, B. (1989). Puzzles: The Other Person's Envelope is Always Greener, *Journal of Economic Perspectives*, 3 (1): 171–81, doi:10.1257/jep.3.1.171.
- Rossi, C., Scozzafava, R. (1992). Perché é falsa la teoria del passaggio da droghe leggere a droghe pesanti. Osservatorio Leggi sulla Droga, VIII rapporto. Millelire, Stampa Alternativa, pp. 30–37.

***In silico* lab@home: database querying e database searching**

Rosanna del Gaudio

Dipartimento di Biologia, Università di Napoli Federico II

Abstract

Le macromolecole biologiche informative, gli acidi nucleici e le proteine sono polimeri costituiti rispettivamente da quattro deossiribonucleotidi (DNA), ribonucleotidi (RNA), ventidue possibili amminoacidi (20 classici e 2 speciali). La successione, la sequenza e l'identità dei residui nucleotidici e/o amminoacidici determina le proprietà di queste macromolecole, che convenzionalmente per le analisi *in silico* sono rappresentate come successioni di lettere che li identificano univocamente. Le stringhe di caratteri risultanti possono essere analizzate con metodi informatici che consentono di rivelare *pattern* particolari o di effettuare confronti fra diverse sequenze. Il confronto, l'allineamento di sequenze di amminoacidi e/o nucleotidi è uno dei lavori di *routine* nella ricerca biologica a livello molecolare *in silico*, ma è anche il presupposto per analisi più complesse come le ricerche di similarità nelle banche dati biologiche, la costruzione di alberi filogenetici o l'identificazione di domini funzionali. Con l'obiettivo di implementare la didattica della biologia nell'era della **biologia sintetica** saranno proposti video, esperimenti e articoli dalla letteratura scientifica anche recente e in lingua inglese per **l'interdisciplinarietà e il training** sull'utilizzo di strumenti (*tool*) di bioinformatica (***in silico lab***) per implementare ed esemplificare il flusso e l'espressione dell'informazione genetica: dal DNA alle proteine.

Usare i database biologici per studiare l'evoluzione

La Bioinformatica è una disciplina scientifica, relativamente giovane, dedicata alla risoluzione di problemi biologici a livello molecolare con metodi informatici. In particolare, questa nuova branca della scienza, descrive i fenomeni biologici in modo numerico/statistico, fornisce modelli per l'interpretazione dei dati provenienti da esperimenti di biologia, ma genera anche nuovi strumenti matematici per l'analisi dei dati biologici (sequenze di DNA, RNA, proteine,...) organizzando le conoscenze acquisite in Banche Dati al fine di rendere i dati biologici accessibili (Fig. 1).

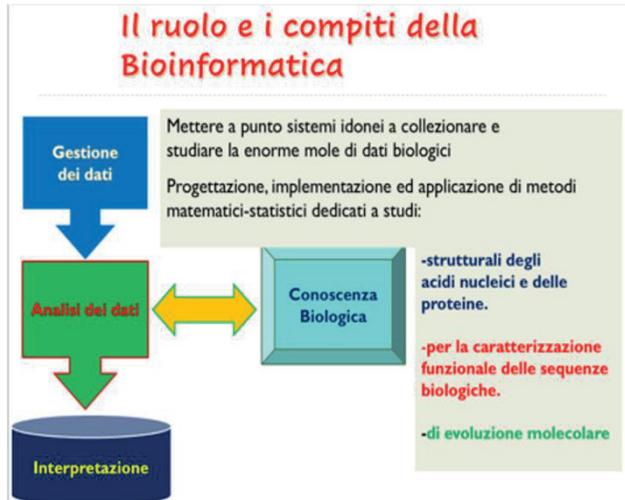


Figura 1. Analisi dei dati biologici con strumenti bioinformatici.

La Bioinformatica è uno strumento che ogni biologo moderno, nel corso della sua carriera si trova “prima o poi” a dover utilizzare. La necessità dell’approccio informatico per l’analisi e lo studio dei genomi è dovuto essenzialmente alla complessità e alla enorme mole di dati biologici. Il genoma umano, ad esempio, è costituito da 3,3 miliardi di coppie di decossi nucleotidi e, per gestire questa quantità enorme di informazioni, per confrontare dati e per analizzare strutture e sequenze, i computer si rivelano gli strumenti più adatti. Nell’ultimo ventennio sono cresciute in parallelo sia la capacità computazionale sia i dati molecolari ottenuti dalla ricerca genetica e, senza dubbio, senza l’aiuto dei computer non si sarebbe certamente potuto portare a termine il progetto del sequenziamento dell’intero genoma umano. La bioinformatica sfrutta il fatto che la catena di DNA può essere rappresentata come una lunga sequenza di lettere (A, C, G, T, i quattro deossiribonucleotidi) e le proteine sono costituite di amminoacidi (20 possibili che possono essere rappresentati con una nomenclatura a una lettera come presentato in Figura 2 ed in entrambi i casi quindi facilmente codificabile con il codice binario e immagazzinabile come dato informatico alfanumerico.

Inoltre, l’istituzione dei database bioinformatici ha messo a disposizione uno strumento molto potente e utile per gli **studi evolutivisti**.

La bioinformatica quindi consente di:

- organizzare i dati biologici in banche dati *on line* e fornisce gli strumenti per l’interrogazione di queste banche dati;
- identificare le sequenze di nucleotidi corrispondenti a geni specifici;

- confrontare le sequenze nucleotidiche e quelle amminoacidiche di diversi organismi di una stessa specie e fra organismi di specie diverse, anche molto lontane evolutivamente;
- analizzare la massa enorme di dati derivanti dalle nuove tecnologie di analisi genomica, quali lo studio del profilo di espressione di tutti i geni umani in condizioni patologiche o fisiologiche (trascrittomica) o lo studio del proteoma (vale a dire l'insieme di tutte le proteine sintetizzate da una cellula o da un tessuto).

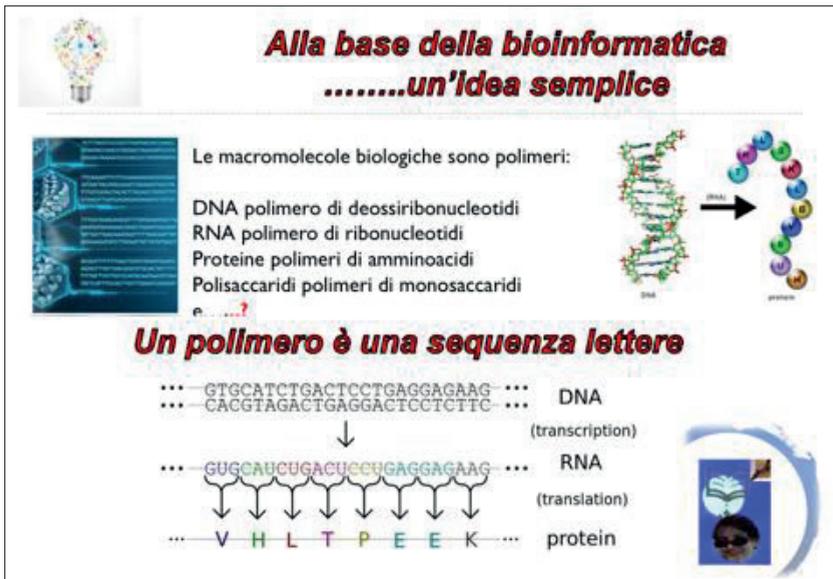


Figura 2. Le banche dati si strutturano e si integrano per favorire lo studio del flusso dell'informazione che va dal DNA alle proteine.

Si possono così formulare ipotesi sulla funzione dei geni, prevedere le strutture tridimensionali delle proteine, identificare nuovi geni associati a particolari malattie. La computer grafica, inoltre, è anche molto utile per visualizzare la struttura delle molecole dal punto di vista tridimensionale, con il *modeling* delle molecole proteiche, la farmacologia può, ad esempio, prevedere e progettare la realizzazione di nuovi farmaci più specifici ed efficaci.

Cosa è una Banca Dati biologica?

Una banca dati biologica è una raccolta di informazioni e dati che derivano dalla letteratura scientifica e dalle analisi e dagli esperimenti effettuate sia in laboratorio

(*wet lab*, *in vitro* ma anche *in vivo*) sia mediante strumenti bioinformatici (*in silico lab*).

Ogni banca dati è organizzata attorno ad un elemento centrale (nelle banche dati di sequenze nucleotidiche, ad esempio, questo elemento è rappresentato da una sequenza di DNA o di RNA). Attorno all'elemento centrale viene costruita la *entry* della banca dati, che oltre alla sequenza, nucleotidica o amminoacidica, comprende tutta una serie di informazioni supplementari, quali il nome della sequenza, la specie di appartenenza, la classificazione tassonomica, le coordinate di tutti gli elementi funzionali in essa contenuti (gene, esoni, introni, etc.) in pratica tutte le annotazioni utili per la classificazione. Generalmente quando un singolo ricercatore o un gruppo di ricerca determina una sequenza nucleotidica per poter essere utilizzata per una pubblicazione scientifica deve essere depositata nella banca dati e resa pubblica, disponibile a tutti. La procedura per la "sottomissione" alla banca dati prevede che a ciascuna sequenza depositata venga assegnato un codice per l'identificazione univoca detto *accession number* che rimane invariato anche se nel tempo la *entry sequence* viene sottoposta ad aggiornamenti.

Le informazioni contenute nelle voci di una banca dati sono scritte generalmente sotto forma di *flat file*, un file di testo che contiene le informazioni scritte in maniera sequenziale in linee identificate da un codice a sinistra che caratterizza gli attributi contenuti nella linea stessa (Fig. 3).

GenBank Flat file format

Il formato GenBank Flat file contiene un header con dei metadati:

	Locus name	Sequence length	Molecule type	Division	
LOCUS	SCU49845	5028 bp	DNA	PLN	21-JUN-1999
DEFINITION	Saccharomyces cerevisiae YCP1-beta gene, partial cds, and Axl2p (AXL2) and Rev7p (REV7) genes, complete cds.				
ACCESSION	U49845				
VERSION	U49845.1 GI:1293613				
KEYWORDS	-				
SOURCE	Saccharomyces cerevisiae (baker's yeast)				
ORGANISM	Saccharomyces cerevisiae				
	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.				
REFERENCE	1 (bases 1 to 5028)				
AUTHORS	Torpey,L.E., Gibbs,P.E., Nelson,J. and Lawrence,C.W.				
TITLE	Cloning and sequence of REV7, a gene whose function is required for DNA damage-induced mutagenesis in Saccharomyces cerevisiae				
JOURNAL	Yeast 10 (11), 1503-1509 (1994)				
PUBMED	7871890				
REFERENCE	2 (bases 1 to 5028)				
AUTHORS	Roemer,T., Madden,K., Chang,J. and Snyder,H.				
TITLE	Selection of axial growth sites in yeast requires Axl2p, a novel plasma membrane glycoprotein				
JOURNAL	Genes Dev. 10 (7), 777-793 (1996)				

Figura 3. Esempio di flat file in cui sono evidenziati alcuni dei metadati.

Il concetto di informazione è strettamente connesso a quello di dato e di struttura. Il dato è un osservabile (insieme di numeri, caratteri, simboli...). La struttura è l'organizzazione ordinata di dati che ne consente l'apprendimento. Una banca dati è l'insieme dei dati elementari, omogenei, ordinati e fruibili, in altre parole è una **collezione organizzata di dati**. Le banche dati possono essere distinte in:

- banche dati di PRIMO LIVELLO
- banche dati di SECONDO LIVELLO (o specializzate)

Le banche di acidi nucleici (DNA e RNA) sono definite di primo livello se contengono solo informazioni molto generiche associate alla sequenza, necessarie per identificarla dal punto di vista specie-funzione.

Le principali banche di acidi nucleici sono tre:

-GenBank (*USA National Center for Biology Information*)

<http://www.ncbi.nlm.nih.gov/Genbank>

-EMBL (*European Molecular Biology laboratory*)

<http://www.ebi.ac.uk/embl/>

-DDBJ (*DNA Data Bank of Japan*)

<http://www.ddbj.nig.ac.jp/>

La prima ad essere istituita è stata l'**EMBL data library**, istituita all'inizio degli anni '80 dal Laboratorio Europeo di Biologia Molecolare (EMBL) a cui ha fatto seguito la **GenBank** negli Stati Uniti (1982) e la **DDBJ** in Giappone (1986). Fra queste tre banche dati biomolecolari è stato stipulato un accordo internazionale per cui il contenuto dei dati di sequenza presenti è quasi del tutto coincidente e le informazioni vengono scambiate fra le tre banche dati quotidianamente. Qualsiasi ricercatore può depositare (attraverso un apposito sistema *on line* di invio i propri risultati sperimentali) le sequenze nucleotidiche o amminoacidiche che, dopo essere state controllate, vengono inserite nella banca dati.



Figura 4. Le banche dati.

Come già detto, le banche dati biologiche sono collezioni di informazioni riguardanti molecole presenti nei viventi, esse hanno dimensioni considerevoli e quindi vengono rese disponibili al pubblico unitamente a strumenti specializzati per svolgere ricerche e analisi al loro interno (Fig. 4). In Figura 5 sono presentate le domande fondamentali a cui si può rispondere con gli strumenti della bioinformatica.



Figura 5. Analisi dei dati biologici depositati nelle Banche Dati con strumenti bioinformatici.

Esistono due tipi di ricerca che si possono effettuare all'interno di una banca dati biologica.

- La ricerca per **parola chiave** che consente di estrarre una o più sequenze nucleotidiche e/o amminoacidiche fornendo una serie di parole chiave opportunamente combinate. La stringa di interrogazione può essere costruita dinamicamente grazie a una serie di filtri progressivi. Non può essere utilizzata in assenza di informazioni sull'obiettivo della ricerca.
- La ricerca per **similarità di sequenze** al posto delle parole chiave si utilizza una sequenza nucleotidica o amminoacidica (sonda) che serve per trovare quelle più simili depositate nella banca dati.

Indipendentemente dalla modalità di ricerca seguita è possibile raggiungere delle schede (*entries*) che contengono molti collegamenti ad altre informazioni sulla sequenza presenti in altre banche dati. In altre parole, la ricerca d'informazioni in una banca dati può essere effettuata con due diverse modalità:

“*Database querying*” oppure “*Database searching*”.

Nel caso del *Database querying*, si tratta di utilizzare delle **parole chiave** per trovare delle informazioni generiche, in modo simile a come si potrebbe fare con un qualunque motore di ricerca per trovare, ad esempio, una ricetta di cucina preferita o le opere editate da un autore e, similmente a quanto accade per questo tipo di ricerca, occorre utilizzare parole chiave opportune e il più precise possibile. Questo metodo d'interrogazione è particolarmente utile per fare ricerche inerenti interi geni, sequenze nucleotidiche o proteine d'interesse, in una certa specie oppure in un

(micro)organismo, quando si vuole trovare più informazioni possibili al riguardo presenti in letteratura scientifica.

N.C.B.I, PubMed, Entrez: la letteratura scientifica e i dati molecolari a portata di click

Il *National Center for Biotechnology Information (NCBI)* è un ente governativo statunitense che gestisce il sito web (www.ncbi.nlm.nih.gov) attraverso il quale è possibile accedere a una serie di banche di dati molecolari (genomiche, di proteine, di trascritti etc) (Fig. 6).

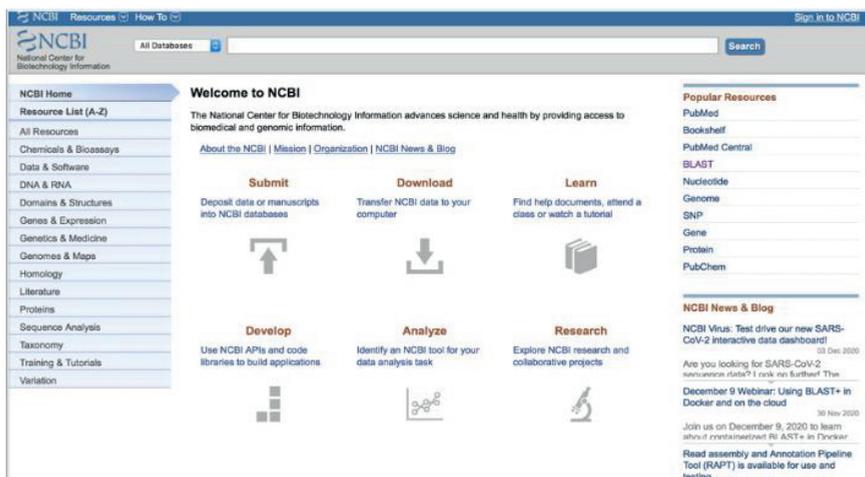


Figura 6. Analisi dei genomi con strumenti bioinformatici.

Entrez è l'interfaccia che consente di effettuare l'accesso alle banche dati. **PubMed** raccoglie le pubblicazioni in campo medico a partire dal 1966 fino ai nostri giorni e comprende anche materiale precedente (fra il 1963 e il 1965). Circa 5000 riviste internazionali in 30 lingue diverse. Utilizzando **PubMed** per la ricerca bibliografica si utilizza il sistema logico sviluppato dal matematico inglese George Boole. Questa logica consiste di operatori booleani. Per poter effettuare delle ricerche si può utilizzare il *tool All Databases* aggiungendo una parola chiave in lingua inglese (Fig. 7). Facendo *click* sul tasto *search*, si ottiene una schermata generale in cui sono presenti tutte le risposte ottenute dopo l'inserimento della parola chiave.

Data Base searching

Gli strumenti di bioinformatica utilizzati per la ricerca di omologia messi a disposizione dai *DataBase* sono **BLAST** e **FASTA**. Si tratta di programmi che

sfruttano algoritmi euristici, basati sulla ricerca di “parole” (dove ogni parola può includere i 4 nucleotidi oppure i 20 amminoacidi ed essere rappresentata da più lettere consecutive, disposte in qualsiasi modo negli acidi nucleici o nelle proteine) all’interno delle sequenze contenute in banca dati.



Figura 7. Print screen della home page in cui la ricerca per parola chiave che deve essere digitata nel campo indicato dalla freccia in colore nero dopo aver selezionato nel menu “All Databases” (nel riquadro in colore rosso)

Il principio seguito è che quando due sequenze sono correlate, esse condivideranno molte lettere identiche o simili. Pertanto, ciò che questi algoritmi fanno (in linea di massima almeno) è valutare stringhe di nucleotidi oppure amminoacidi effettuando in esse una sorta di ricerca di determinati elementi caratteristici, quindi le allineano con le sequenze identiche o più o meno simili disponibili nel database. BLAST e FASTA differiscono nel metodo statistico matematico su cui si basano: senza scendere troppo nei dettagli mi limiterò a dare un’idea generale delle potenzialità e delle applicazioni di questo tipo di confronto. BLAST si basa su punteggi di similitudine tra parole che abbiano una lunghezza di caratteri superiori a certi valori soglia, mentre FASTA utilizza l’algebra delle matrici a punti e rimanda i valori presenti sulle diagonali. L’omologia fra sequenze che mostrano una percentuale di identità superiore al 20% riflette a grandi linee il grado di parentela evolutiva: due sequenze possono definirsi omologhe quando condividono un antenato ancestrale evolutivo e derivano da questi tramite processi di duplicazione oppure speciazione. Attraverso l’algoritmo *BLAST* (*Basic Local Alignment Tool*) si possono

confrontare sia sequenze nucleotidiche sia amminoacidiche (*query sequence*) con quelle depositate nel *database* (Fig. 8).

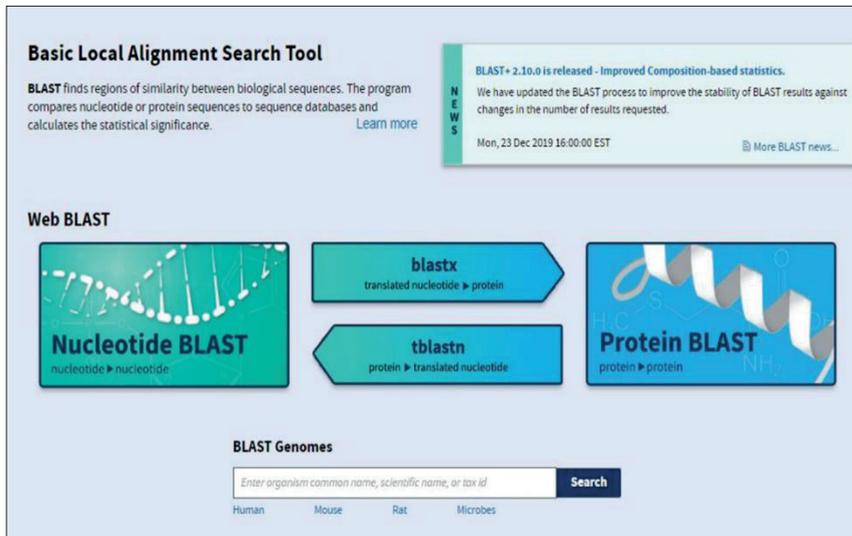


Figura 8. Print screen dei diversi strumenti (tools) dell'algoritmo BLAST con cui si possono confrontare le sequenze nucleotidiche e/o amminoacidiche e viceversa (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>)

La home page di B.L.A.S.T del N.C.B.I è organizzata in tre sezioni:

- **BLAST Assembled Genomes** - per le ricerche in specifici genomi
- **Basic BLAST** - è il portale per l'uso della maggior parte delle applicazioni BLAST (confronto di sequenze di DNA e proteine)
- **Specialized BLAST** - è il portale di accesso a una serie di applicazioni specializzate

La sezione più utilizzata: **Basic BLAST** offre cinque varianti dell'algoritmo. Per confronti di sequenze omogenee (DNA con DNA o proteine con proteine):

- **nucleotide blast** per il confronto tra sonde nucleotidiche (*query sequence*) e le sequenze depositate nei *database* nucleotidici;
- **protein blast** per il confronto tra sonde amminoacidiche (*query sequence*) e sequenze depositate nei *database* di proteine;

Le successive tre applicazioni **translated blast** sono dedicate al confronto DNA-proteine (e viceversa):

- **blastx** per la ricerca nel *database* di proteine usando una *query sequence* nucleotidica che viene tradotta in silico nelle tre cornici di lettura possibili;

- *tblastn* (in cui si ricerca nel database di nucleotidi tradotti una query di amminoacidi);
- *tblastx* per la si ricerca nel database di sequenze nucleotidiche una query nucleotidica entrambe tradotte *in silico* nelle sei cornici di lettura possibili

Per ricercare nei *databases* una sequenza nucleotidica, si può utilizzare il tool *blastn*.

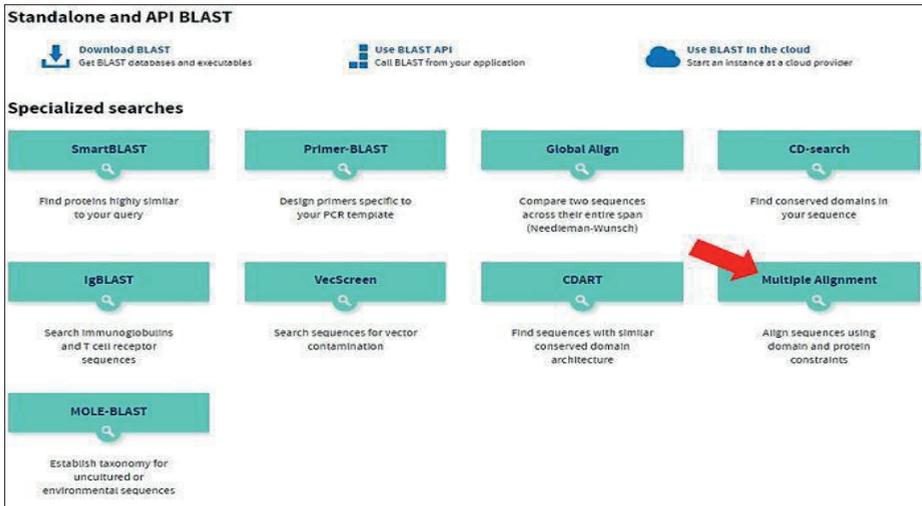


Figura 9. Print screen di BLAST in cui si evidenziano i diversi campi da compilare per effettuare il confronto fra la query sequence e il DataBase. La descrizione dettagliata è nel testo.

Si inizia con l'inserimento della sequenza nucleotidica in formato *FASTA*, o utilizzando l'*accession number* della rispettiva sequenza, nel riquadro indicato dalla freccia in colore azzurro in Figura 9. Si seleziona il *database* nel quale si vuole fare la ricerca indicato dalla freccia in colore rosso della stessa figura e anche eventualmente la specie da inserire nel campo indicato dalla freccia in colore giallo. Alla fine, prima di fare click sul tasto *BLAST*, indicato dalla freccia in colore nero, per avviare la ricerca, bisogna scegliere e selezionare i *tools* per il confronto (Fig. 9). L'algoritmo *BLAST* offre anche altre possibilità di confronto visualizzate in Figura 10. In particolare, per poter effettuare un confronto tra le proteine codificate dallo stesso gene, ma di varie specie, si utilizza l'allineamento multiplo facendo *click* come indicato dalla freccia di colore rosso

La genomica comparata rivela le regioni conservate dei genomi

Il gran numero di genomi sequenziati, consente di effettuare confronti sia a livello di singoli geni sia di genoma dando origine agli studi di **genomica comparata**.

Questi studi hanno rivelato che il genoma umano è molto simile a quello di altri organismi, infatti più della metà dei geni del moscerino della frutta, *D. melanogaster*, ha una controparte umana.

The screenshot shows the BLAST search interface with several key sections:

- Enter Query Sequence:** A text input field for accession numbers, GI IDs, or FASTA sequences. A blue arrow points to the 'Clear' button.
- Job Title:** A text input field for a descriptive title.
- Choose Search Set:**
 - Database:** A dropdown menu currently set to 'Whole-genome shotgun contigs (wgs)'. A yellow arrow points to this dropdown.
 - Limit by:** Radio buttons for 'Organism', 'BioProjectID', and 'WGS Project'. A red arrow points to the 'Organism' option.
 - Limit to Optional:** A checkbox for 'Sequences from type material'.
- Program Selection:**
 - Optimize for:** Radio buttons for 'Highly similar sequences (megablast)', 'More dissimilar sequences (discontiguous megablast)', and 'Somewhat similar sequences (blastn)'. A green arrow points to the 'Somewhat similar sequences (blastn)' option.
- BLAST Button:** A large blue button labeled 'BLAST'. A black arrow points to it.
- Search details:** A summary line: 'Search database wgs using Blastn (Optimize for somewhat similar sequences)'. A checkbox for 'Show results in a new window' is checked.

Figura 10. Print screen di altri strumenti per il confronto di BLAST. Il tool multiple alignment (freccia in colore rosso) mediante il quale si possono confrontare più di due sequenze amminoacidiche.

La genomica comparata è un'area di ricerca bioinformatica promettente anche per porsi domande sull'evoluzione quando ad esempio si analizzano i dati relativi ai genomi dei mammiferi che rivelano complessivamente una identità pari al 99% ad esempio, ma anche centinaia di geni umani (circa 300 geni) non hanno una controparte, ad esempio, nel genoma del topo. Il confronto fra i molti genomi procariotici, attualmente sequenziati, sta rivelando un livello di trasferimento genico maggiore di quanto precedentemente ipotizzato. Fra i genomi animali di cui è stato ultimato il sequenziamento c'è quello dello scimpanzé (*Pan troglodytes*) il nostro "parente" vivente evolutivamente più vicino e, i primi dati ottenuti da questa impresa scientifica confermano fra il genoma umano e quello dello scimpanzé differiscono

per l'1,23 % in termini di sostituzioni nucleotidiche, che tenendo in considerazione la lunghezza complessiva del genoma umano corrisponde a un totale di circa 4 milioni e 59 mila nucleotidi diversi. A prima vista, la differenza tra questi genomi sembra essere negli elementi trasponibili, in particolare nel genoma umano le **sequenze SINE** (*Short Interspersed Elements*) sono state tre volte più attive che nello scimpanzé, ma il genoma di quest'ultimo ha acquisito due elementi che non si trovano nel genoma umano. Le differenze dovute a inserzioni e delezioni di nucleotidi sono minori rispetto alle sostituzioni, ma costituiscono circa l'1,5 % della sequenza dell'eucromatina che contiene la frazione del genoma trascritto ed espresso.

Analogie e differenze tra i geni possono anche essere studiate proteina per proteina nelle diverse specie.

COBALT è un *tool* di *BLAST* che consente di inserire gli *accession numbers* delle proteine di diverse specie per il confronto (Fig. 11)

Figura 11. Print screen del tool COBALT. Nel riquadro identificato dalla freccia in colore azzurro è necessario inserire gli accession number relativi alle proteine delle diverse specie. La freccia in colore verde indica il pulsante su cui fare click col mouse per avviare il programma per il confronto.

Una attività che può essere seguita dagli allievi *on line* è disponibile collegandosi alla sezione di *Science Education* della videorivista internazionale Jove riportata di seguito:

<https://www.jove.com/science-education/10621/evolutionary-relationships>.

La possibilità di analizzare, infatti, nel dettaglio intere sequenze genomiche oppure amminoacidiche di specie diverse, confrontandone il grado di identità e conservazione oppure di diversificazione, permette di ricostruire l'intera storia filogenetica delle stesse e di caratterizzarle in modo ben più preciso di quanto non avvenisse in passato

Sitografia di approfondimento

Una guida preliminare al mondo delle banche dati NCBI è consultabile a questo link:

<https://a-little-book-of-r-for-bioinformatics.readthedocs.io/en/latest/src/chapter3.html>

Di seguito alcuni gli indirizzi *web* a cui si può fare riferimento:

- Il sito <http://www.genomesonline.org/gold.cgi> che riporta tutti i genomi completamente sequenziati (in continuo aggiornamento)
- Il sito <http://www.ncbi.nlm.nih.gov/> del National Center of Biotechnology Information (*NCBI*) contiene un gran numero di banche dati, strumenti e programmi di analisi bioinformatica.
- Il *browser* genomico dell'Università della California a Santa Cruz [UCSC] accessibile presso il sito <http://genome.ucsc.edu/>.
- Il sito EMBOSS (<http://www.ei.ac.uk/tolse/emboss/>) raccoglie numerosi programmi di bioinformatica.
- www.ihc-epigenomes.org, il sito del Consorzio Internazionale dell' Epigenoma Umano (IHEC, International Human Epigenomic Consortium) che colleziona informazioni sul profilo epigenetico umano
- www.hmpdacc.org, il sito del Progetto del Microbioma Umano (HMP) che colleziona dati sul microbioma umano.

Seminario integrato: Un Mondo Frattale

Parte 1

Barbara Brandolini

*Dipartimento di Matematica e Applicazioni "R. Caccioppoli",
Università di Napoli Federico II*

Abstract

Questo articolo è il sunto della conferenza tenuta durante la PLS Virtual Summer School 2020. Scopo della conferenza, e quindi dell'articolo, è stato quello di avvicinare gli studenti alla geometria frattale, facendo scoprire loro la nozione matematica di autosimilarità e l'esistenza di oggetti aventi dimensione non intera.

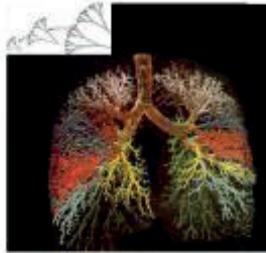
Introduzione

La geometria frattale è innanzitutto e soprattutto un nuovo linguaggio utilizzato per descrivere la complessità delle forme del mondo reale. Come scrisse Mandelbrot nel 1977 per spiegare l'incapacità della geometria euclidea di descrivere la natura (Mandelbrot, 1982): "Why is geometry often described as "cool" and "dry"? One reason lies in its inability to describe the shape of a cloud, a mountain, a coastline, or a tree. Clouds are not spheres, mountains are not cones, coastlines are not circles, and bark is not smooth, nor does lightning travel in a straight line". Allo scopo di introdurre dunque questo nuovo linguaggio, guardiamoci attorno e cerchiamo di rispondere alla seguente domanda: cosa hanno in comune una felce, un soffione ed un broccolo romano?



Certo si tratta di piante comuni e, a guardarle con un po' più di attenzione, molto frastagliate, con una forte alternanza di pieni e vuoti. Dal punto di vista matematico sono piante autosimili. Ma cosa vuol dire autosimili? Zoomiamo sulla foglia di una felce. Scopriamo che essa è formata da foglie più piccole, ciascuna delle quali

presenta in scala ridotta la stessa struttura della foglia di partenza. Perciò la felce è autosimile, cioè simile a se stessa. Possiamo notare questa stessa caratteristica anche



in una nuvola, in un fulmine, in una piuma o nella frattura su una lastra di ghiaccio. Oppure, invece di guardarci attorno, possiamo guardarci dentro e notare lo stesso tipo di irregolarità ed autosimilarità nell'albero bronchiale di un essere umano.

La domanda a questo punto sorge spontanea: perché la natura sceglie figure così complesse? Proviamo a spiegarlo ragionando sui polmoni. Sappiamo tutti che i polmoni hanno la funzione di distribuire ossigeno al corpo umano, quindi maggiore è la superficie dei bronchi maggiore è l'apporto di ossigeno al corpo umano. Del resto i polmoni devono essere contenuti all'interno del torace umano, quindi necessitano di un volume piccolo. Vedremo che, dal punto di vista matematico, i bronchi hanno superficie infinita e volume zero!

La complessità degli elementi naturali descritti e la loro autosimilarità è stata intuuta da tempo da filosofi e matematici. Ad esempio, Leibniz nel 1714 scrive nella *Monadologia*: "Ogni porzione di materia può essere concepita come un giardino pieno di piante e come uno stagno pieno di pesci; ma ogni ramo di pianta, ogni membro di animale, ogni goccia dei loro umori, è ancora un giardino simile, un simile stagno". Dobbiamo però aspettare la metà del secolo scorso ed il matematico franco-polacco Benoit Mandelbrot per l'introduzione della nozione matematica di frattale. Il termine frattale viene dal latino *fractus*, che significa interrotto o irregolare, e si presta bene ad indicare una figura geometrica F complessa che possiede, tra le altre, le seguenti proprietà che lo distinguono dalle figure della geometria euclidea:

1. F è costruito per ricorsività, cioè F non può essere descritto come luogo di punti che soddisfano semplici condizioni geometriche o analitiche, bensì tramite un algoritmo che prevede, in teoria, un numero infinito di iterazioni;

2. F è autosimile o invariante per cambiamento di scala, cioè F è unione di un numero di parti che, ingrandite di un certo fattore, riproducono tutto F . In altri termini F è unione di copie di se stesso a scale differenti;
3. F ha dimensione non intera (su questo torneremo in seguito).

Mandelbrot ha introdotto la nozione di frattale nel 1967 in un articolo scientifico pubblicato sulla rivista *Science*, dal titolo *How Long Is the Coast of Britain? Statistical Self-Similarity and Fractional Dimension* (Mandelbrot, 1967). In questo articolo viene introdotto il cosiddetto paradosso della linea costiera. La lunghezza della linea costiera dipende dalla scala nella quale viene misurata ed, empiricamente, aumenta quando si riduce l'ampiezza dell'unità di misura di riferimento. Come possiamo evincere dalla figura, essa è di circa 2400 km se misurata con unità di misura di 200 km: servono infatti 12 righelli lunghi 200 km per ricoprire la costa della Gran Bretagna. È chiaro a tutti, però, che questo valore della lunghezza della costa è ben lontano da quello vero: ci sono interi pezzi di costa che non sono stati presi in considerazione, d'altro canto abbiamo misurato lunghe porzioni di mare. Per migliorare l'approssimazione, usiamo un righello più corto, di 50 km. La lunghezza della costa questa volta è di circa 3400 km: servono in questo caso 69 righelli lunghi 50 km per ricoprire la costa della Gran Bretagna (vedi Figura 1). Dunque, la lunghezza della costa della Gran Bretagna dipende dalla scala sulla quale viene misurata, ed empiricamente essa aumenta quando riduciamo la lunghezza dell'unità di misura. Paradossalmente possiamo dedurre che la lunghezza della linea costiera della Gran Bretagna è infinita! È quindi evidente che non ha significato parlare di lunghezza della linea costiera, ma è necessario introdurre un nuovo concetto di misura ed al contempo di dimensione di un insieme.

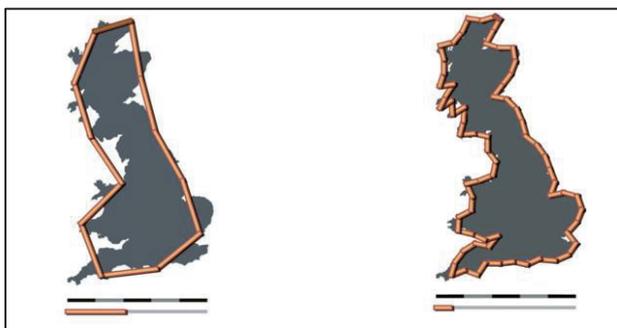


Figura 1. A sinistra: unità = 200 km, lunghezza ≈ 2400 km; a destra: unità = 50 km, lunghezza ≈ 3400 km

La dimensione (dal latino *dimensio*, “misura”) è, essenzialmente, il numero di gradi di libertà disponibili per il movimento. Ad esempio, un punto nel piano è descritto

da due parametri indipendenti (le coordinate cartesiane del punto) e così il piano è bidimensionale. Nell'uso comune, le dimensioni di un oggetto diventano le misure che ne definiscono la forma e la grandezza. In matematica non esiste una definizione di dimensione adatta a tutte le situazioni in cui vorremmo farne uso. Basti pensare che spesso i matematici lavorano in spazi di dimensione infinita (per esempio lo spazio dei polinomi)! Di conseguenza, i matematici hanno ideato molte definizioni di dimensione per i diversi tipi di spazio e corrispondenti definizioni di misura. Ad ogni modo, affinché queste definizioni siano accettabili, esse devono rispettare alcune proprietà fondamentali: la misura di un insieme, ad esempio, deve essere un numero non negativo e deve essere monotona rispetto all'inclusione tra insiemi. Inoltre, queste nuove definizioni di dimensione e misura devono restituire la dimensione e la misura che già conosciamo in casi elementari, quali possono essere un insieme di punti, un segmento, un rettangolo o un cubo.

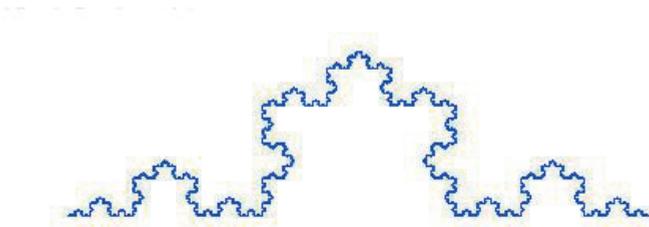


Siamo sicuramente tutti d'accordo sul fatto che un punto abbia dimensione 0, un segmento 1, un rettangolo 2, un cubo 3. Come misuriamo tali oggetti? Contiamo i punti, ottenendo un risultato strettamente positivo; calcoliamo la lunghezza di un segmento, ottenendo un risultato strettamente positivo; calcoliamo l'area di un rettangolo, ottenendo un risultato strettamente positivo; calcoliamo il volume di un cubo, ottenendo anche in questo caso un risultato strettamente positivo. Ma quale risultato otteniamo se proviamo a calcolare l'area di un segmento o la lunghezza di un rettangolo? Un segmento ha area zero, mentre un rettangolo ha lunghezza infinita (facciamo attenzione a non confondere la lunghezza del rettangolo con la lunghezza del suo bordo, cioè con il suo perimetro!). Ragionando in quest'ottica, possiamo riempire la seguente Tabella:

Dimensione	Misura che conta i punti	Lunghezza	Area	Volume
0	> 0	0	0	0
1	∞	> 0	0	0
2	∞	∞	> 0	0
3	∞	∞	∞	> 0

Osserviamo che esiste sempre un'unica misura finita, strettamente positiva. I valori delle altre misure sono zero o infinito. In quali casi la misura vale zero? In quali casi

la misura vale infinito? La misura è zero se calcolata in uno spazio avente dimensione maggiore di quella dell'insieme, vale infinito se calcolata in uno spazio avente dimensione minore di quella dell'insieme. Abbiamo visto che la lunghezza della costa della Gran Bretagna è infinita, dunque la costa ha dimensione maggiore di 1. Del resto essa ha area zero, dunque ha dimensione minore di 2. Ma allora, qual è la dimensione della costa della Gran Bretagna? Per semplicità, anziché ragionare sulla costa della Gran Bretagna, ragioniamo su un frattale più semplice, ma altrettanto frastagliato: la curva di von Koch.



Abbiamo detto che per costruire la curva di von Koch dobbiamo fornire un algoritmo. A tal fine, fissiamo un sistema di riferimento cartesiano nel piano e consideriamo il segmento AB di estremi $A(0,0)$ e $B(1,0)$, avente lunghezza unitaria. Dividiamo il segmento in tre parti uguali tramite i punti C e D ed eliminiamo il segmento centrale CD . Al suo posto, consideriamo due segmenti di pari lunghezza ruotati di 60° in senso antiorario attorno a C ed in senso orario attorno a D , rispettivamente.



Ripetiamo lo stesso ragionamento con ciascuno dei quattro segmenti ottenuti (vedi Figura 2).

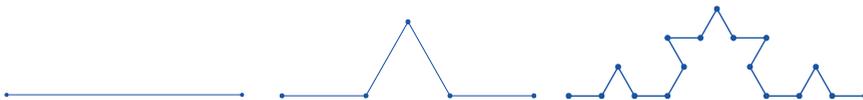


Figura 2. Passi 0, 1, 2 per la costruzione della curva di von Koch

Andiamo avanti all'infinito ed otteniamo la curva di von Koch γ , che diventa più frastagliata ogni volta che iteriamo il procedimento sopra descritto. La curva di von Koch γ è autosimile? Essa è unione di 4 copie ridotte di se stessa. Ma cosa

intendiamo per copia ridotta? Intendiamo dire che è stata applicata una trasformazione affine in cui compare un'omotetia di fattore $\frac{1}{3}$. Precisamente,

$$\gamma = T_1(\gamma) \cup T_2(\gamma) \cup T_3(\gamma) \cup T_4(\gamma),$$

dove

$$\begin{aligned} T_1 : \begin{cases} X = \frac{1}{3}x \\ Y = \frac{1}{3}y \end{cases} & \quad T_2 : \begin{cases} X = \frac{1}{6}x - \frac{\sqrt{3}}{6}y + \frac{1}{3} \\ Y = \frac{\sqrt{3}}{6}x + \frac{1}{6}y \end{cases} \\ T_3 : \begin{cases} X = -\frac{1}{6}x - \frac{\sqrt{3}}{6}y + \frac{2}{3} \\ Y = \frac{\sqrt{3}}{6}x - \frac{1}{6}y \end{cases} & \quad T_4 : \begin{cases} X = \frac{1}{3}x + \frac{2}{3} \\ Y = \frac{1}{3}y. \end{cases} \end{aligned} \tag{1}$$

Resta da discutere la dimensione d di γ . Proviamo a calcolare la lunghezza di γ per stabilire se $d = 1$. Indichiamo con L_0 la lunghezza del segmento al Passo 0: $L_0 = 1$. Indichiamo con L_1 la lunghezza della spezzata del Passo 1: $L_1 = \frac{4}{3}$. Analogamente, detta L_2 la lunghezza della spezzata al Passo 2 risulta $L_2 = \frac{16}{9}$. Andando avanti con i nostri calcoli, possiamo facilmente riempire la seguente Tabella:

n	Lunghezza di un segmento	Numero segmenti	L_n
0	$1 = \left(\frac{1}{3}\right)^0$	$1 = 4^0$	$L_0 = 1 = \left(\frac{4}{3}\right)^0$
1	$\frac{1}{3} = \left(\frac{1}{3}\right)^1$	$4 = 4^1$	$L_1 = \frac{4}{3} = \left(\frac{4}{3}\right)^1$
2	$\frac{1}{9} = \left(\frac{1}{3}\right)^2$	$16 = 4^2$	$L_2 = \frac{16}{9} = \left(\frac{4}{3}\right)^2$
3	$\frac{1}{27} = \left(\frac{1}{3}\right)^3$	$64 = 4^3$	$L_3 = \frac{64}{27} = \left(\frac{4}{3}\right)^3$
...			
n	$\left(\frac{1}{3}\right)^n$	4^n	$L_n = \frac{4^n}{3^n} = \left(\frac{4}{3}\right)^n$

È evidente che al crescere del numero di iterazioni n cresce la lunghezza L_n della spezzata e, per rendere questi conti concreti, ci basta notare che se $L_0 = 1m$, allora

$$L_{24} = 1 \text{ km}; \quad L_{128} = 1 \text{ anno luce!!!}$$

Avendo dunque lunghezza infinita, la curva di von Koch γ ha dimensione $d > 1$. Ragioniamo ora sull'area di γ per stabilire se può essere $d = 2$. Nessuno si stupisce davanti all'affermazione: un segmento ha area zero. Cosa possiamo dire a proposito di una curva "frastagliata" come quella di von Koch? Supponiamo di voler calcolare l'area della curva di von Koch ricoprendola con triangoli isosceli via via più piccoli. Partendo dal segmento AB di lunghezza 1, quanto vale la prima approssimazione A_0 dell'area? E la seconda approssimazione A_1 ? E la terza approssimazione A_2 ?

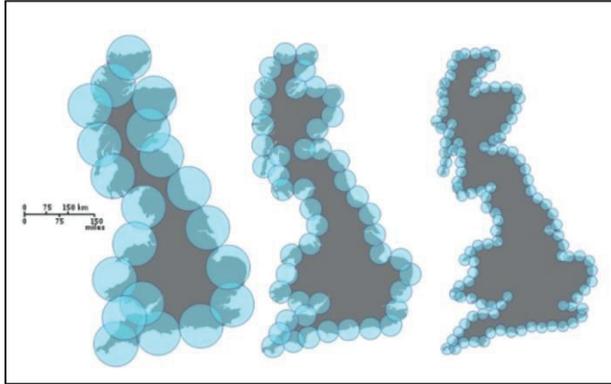


Così come abbiamo fatto per le lunghezze L_n , possiamo riempire la seguente tabella con i valori delle aree A_n .

Dai dati riportati in Tabella deduciamo immediatamente che, come nel caso di un semplice segmento, anche nel caso della curva di von Koch, l'area vale zero e, pertanto, γ ha dimensione $d < 2$. Allora, per poter misurare un frattale le nozioni di lunghezza, area e volume che abbiamo dalle scuole elementari non sono sufficienti. Abbiamo bisogno di dare una nuova definizione di misura e di dimensione. In ogni caso, qualunque sia la misura che introduciamo per misurare i frattali, essa deve lasciare inalterata la misura (quella che già conosciamo!) di oggetti "regolari" di dimensione 0, 1, 2, 3.

n	Lunghezza di un segmento	Numero segmenti	A_n
0	$\frac{\sqrt{3}}{12} = \frac{\sqrt{3}}{12} \left(\frac{1}{9}\right)^0$	$1 = 4^0$	$A_0 = \frac{\sqrt{3}}{12} \left(\frac{4}{9}\right)^0$
1	$\frac{\sqrt{3}}{12} \frac{1}{9} = \frac{\sqrt{3}}{12} \left(\frac{1}{9}\right)^1$	$4 = 4^1$	$A_1 = \frac{\sqrt{3}}{12} \frac{4}{9} = \frac{\sqrt{3}}{12} \left(\frac{4}{9}\right)^1$
2	$\frac{\sqrt{3}}{12} \left(\frac{1}{9}\right)^2$	$16 = 4^2$	$A_2 = \frac{\sqrt{3}}{12} \frac{16}{81} = \frac{\sqrt{3}}{12} \left(\frac{4}{9}\right)^2$
...			
n	$\frac{\sqrt{3}}{12} \left(\frac{1}{9}\right)^n$	4^n	$A_n = \frac{\sqrt{3}}{12} \left(\frac{4}{9}\right)^n$

Facciamoci venire un'idea riprendendo l'esempio della linea costiera della Gran Bretagna discusso precedentemente.



Ricopriamo la linea costiera con dei cerchietti aventi tutti lo stesso diametro: un'approssimazione della lunghezza della linea sarà data dalla somma delle aree dei cerchietti. Osserviamo che, se i diametri dei cerchietti sono grandi, l'approssimazione non potrà che essere pessima: i cerchi racchiudono troppo mare e troppa terra. Se però facciamo diminuire il diametro dei cerchietti, diminuirà la quantità di mare e terra da essi racchiusa visto che l'area di ciascun cerchietto diminuisce, facendo crescere la speranza di una migliore approssimazione della lunghezza della costa. Purtroppo, facendo diminuire il diametro dei cerchietti, inevitabilmente aumenta il numero di cerchietti necessari per ricoprire tutta la linea costiera. Ci sono allora due numeri in competizione: da una parte il diametro dei cerchietti, che diminuisce, dall'altra il numero di cerchietti, che aumenta. Ci incamminiamo verso la definizione di misura di Hausdorff di un insieme (per una definizione rigorosa di misura di Hausdorff rimandiamo il lettore interessato per esempio al testo di Ambrosio *et al.*, 2000).

Siano assegnati un insieme non vuoto E ed un numero reale $s > 0$. Fissato $\delta > 0$, ricopriamo E con un numero $N(s)$ di insiemi $B_1, B_2, \dots, B_{N(s)}$ (nello spazio, cioè quando $s = 3$: sfere; nel piano, cioè quando $s = 2$: cerchi; sulla retta, cioè quando $s = 1$: segmenti) di diametro δ . Calcoliamo la somma delle potenze s -esime di questi diametri e moltiplichiamo per una opportuna costante (dipendente da s e scelta in maniera tale che nei casi $s = 1, 2, 3$ ritroviamo le misure già note dalle scuole elementari):

$$C(s)N(s)\delta^s$$

Indichiamo con $H_\delta^s(E)$ il numero più grande che risulta essere più piccolo di tutte queste somme al variare degli insiemi B_i e quindi mandiamo δ a zero, cioè prendiamo ricoprimenti sempre più fini. Il risultato di questa operazione è un numero non negativo detto misura di Hausdorff s -dimensionale di E , $H^s(E)$.

A questo punto diciamo che un insieme E ha dimensione di Hausdorff d , o equivalentemente scriviamo che $\dim_H(E) = d$, se e solo se esiste $d > 0$, tale che

$$\begin{aligned} H^s(E) &= 0 && \text{per } s > d \\ H^s(E) &= \infty && \text{per } s < d, \end{aligned}$$

cioè se e solo se d è più piccolo di tutti i numeri s positivi tali che la corrispondente misura di Hausdorff di E è nulla o, equivalentemente, d è più grande di tutti numeri s positivi tali che la corrispondente misura di Hausdorff di E è infinita.

In realtà può succedere che $H^s(E) = 0$ per ogni $s > 0$; in tal caso diciamo che E ha dimensione di Hausdorff zero. Può inoltre succedere che $H^s(E) = \infty$ per ogni $s > 0$; in tal caso diciamo che E ha dimensione di Hausdorff infinita. In conclusione, qualunque sia E , risulta $0 \leq \dim_H(E) \leq \infty$.

È possibile dare altre definizioni di dimensione, per esempio la dimensione del “box-counting” o la dimensione frattale (vedi per esempio “classes.yale.edu/fractals”). Noi siamo interessati a casi in cui tutte queste nozioni coincidono tra di loro e coincidono con il risultato trovato mediante il seguente

Teorema. *Siano T_1, T_2, \dots, T_m m trasformazioni affini con rapporti di omotetia k_1, k_2, \dots, k_m minori di 1. Allora esiste un unico insieme limitato E tale che*

$$E = \bigcup_{i=1}^m T_i(E) = T_1(E) \cup T_2(E) \cup \dots \cup T_m(E)$$

la cui dimensione è l'unica soluzione della seguente equazione

$$\sum_{i=1}^m k_i^d = k_1^d + k_2^d + \dots + k_m^d \quad (2)$$

Abbiamo finalmente gli strumenti per calcolare la dimensione d della curva di von Koch. Ricordando la (1), dalla (2) ricaviamo

$$\left(\frac{1}{3}\right)^d + \left(\frac{1}{3}\right)^d + \left(\frac{1}{3}\right)^d + \left(\frac{1}{3}\right)^d = 1 \Leftrightarrow 4 \cdot \left(\frac{1}{3}\right)^d = 1 \Leftrightarrow d = \frac{\ln 4}{\ln 3}$$

Osserviamo che $1 < d < 2$ in accordo con quanto notato in precedenza: la curva di von Koch ha lunghezza infinita ed area zero.

Oltre la matematica

Chiudiamo questo articolo riflettendo sul fatto che l'idea di autosimilarità compare non solo in matematica. Per esempio, in letteratura: da *Palomar* di Italo Calvino (1983):

[...] Il signor Palomar vede spuntare un'onda in lontananza, crescere, avvicinarsi, cambiare di forma e di colore, avvolgersi su se stessa, rompersi,

svanire, rifluire. A questo punto potrebbe convincersi d'aver portato a termine l'operazione che s'era proposto e andarsene. Però isolare un'onda separandola dall'onda che immediatamente la segue e pare la sospinga e talora la raggiunge e travolge, è molto difficile; così come separarla dall'onda che la precede e che sembra trascinarsela dietro verso la riva, salvo poi magari voltarglisi contro come per fermarla. Se poi si considera ogni ondata nel senso dell'ampiezza, parallelamente alla costa, è difficile stabilire fin dove il fronte che avanza s'estende continuo e dove si separa e segmenta in onde a sé stanti, distinte per velocità, forma, forza, direzione. Insomma, non si può osservare un'onda senza tener conto degli aspetti complessi che concorrono a formarla e di quelli altrettanto complessi a cui essa dà luogo. Questi aspetti variano continuamente, per cui un'onda è sempre diversa da un'altra onda; ma è anche vero che ogni onda è uguale a un'altra onda, anche se non immediatamente contigua o successiva; insomma ci sono delle forme e delle sequenze che si ripetono, sia pur distribuite irregolarmente nello spazio e nel tempo. Siccome ciò che il signor Palomar intende fare in questo momento è semplicemente vedere un'onda, cioè cogliere tutte le sue componenti simultanee senza trascurarne nessuna, il suo sguardo si soffermerà sul movimento dell'acqua che batte sulla riva finché potrà registrare aspetti che non aveva colto prima; appena s'accorgerà che le immagini si ripetono saprà d'aver visto tutto quel che voleva vedere e potrà smettere. [...]

Rimandiamo all'articolo di Lucente (2017) per un collegamento tra i frattali e la poesia di Pier Paolo Pasolini *Poesia in forma di rosa*.

Cosa dire di frattali e pittura? Lasciamo il lettore curioso a meditare sui seguenti dipinti ed a cercare frattali oltre la matematica (vedi anche Brandolini e Tricarico, 2017).



Figura 3: A sinistra: Hokusai, *La grande onda di Kanagawa* (1830); al centro: Salvador Dalí, *Il volto della guerra* (1940-41); a destra: Jackson Pollock, *Alchemy* (1947)

Bibliografia

- Ambrosio L., Fusco N., Pallara D. (2000). Functions of bounded variation and free discontinuity problems. Oxford Mathematical Monographs. The Clarendon Press, Oxford University Press, New York.
- Brandolini B., Tricarico M. (2017). L'insostenibile leggerezza della matematica. Piume frattali e bolle di sapone. Liguori editore.
- Lucente S. (2017). Matematica in forma di poesia: viaggiando e guardando con Pier Paolo Pasolini da Parigi a Roma. Periodico di Matematiche, vol. 9, Serie XIII, anno CXXVII.
- Mandelbrot, B. B. (1967). How long is the coast of Britain? Statistical self-similarity and fractional dimension. Science 156.
- Mandelbrot B. B. (1982). The fractal geometry of nature. W. H. Freeman and Co., San Francisco, California.
- <http://classes.yale.edu/Fractals>

Seminario integrato: Un Mondo Frattale

Parte 2

Daniel Riccio

*Dipartimento di Ingegneria Elettrica e delle Tecnologie dell'Informazione,
Università di Napoli Federico II*

Abstract

Sin dall'antichità si è diffusa fra gli uomini la pratica di conservare e tramandare, in maniera compatta, qualsiasi genere di informazione, realizzando così una sorta di rudimentale processo di compressione, che permettesse una riduzione dello spazio necessario alla conservazione dei dati o del tempo impiegato nella comunicazione delle informazioni. Una prova di ciò è data dall'esistenza, presso gli antichi cinesi, di un codice compatto chiamato *Wen-Yan* usato nella trascrizione del linguaggio parlato; grazie a questo codice i caratteri ridondanti potevano essere eliminati permettendo il risparmio di un bene allora prezioso: la carta.

Per quel che concerne le immagini, la codifica frattale, detta anche codifica basata su Sistemi di Funzioni Iterate (IFS), è una famiglia di tecniche che realizzano un buon compromesso tra compressione e qualità percettiva. Inoltre, un codec che sia in grado di effettuare la codifica e la decodifica frattale delle immagini può essere integrato in applicazioni di vario genere, a seconda della tipologia di utilizzo desiderato. Infatti, le possibili applicazioni nei più disparati ambiti della elaborazione e fruizione delle immagini sono tante. La codifica di immagini e video, il recupero di immagini per contenuto, i sistemi biometrici, il trattamento di immagini GIS e satellitari rappresentano solo alcuni possibili esempi.

Immagine digitale bitmap

Ottenere una riproduzione perfetta, in tutti i suoi dettagli, di un qualsiasi aspetto del mondo reale, su di un generico calcolatore non è possibile, a causa delle limitate dimensioni della memoria di questi ultimi. Per tale motivo al fine di poter memorizzare un'immagine in un personal computer è necessario che essa venga prima sottoposta a particolari processi di campionamento e di quantizzazione, mediante dispositivi di acquisizione quali uno scanner o una fotocamera, ottenendo così una riproduzione digitale di quanto viene percepito analogicamente dall'occhio

umano. Un'immagine digitale viene poi suddivisa, mediante una griglia, in un insieme di righe e di colonne a distanza costante. I quadratini nella griglia vengono denominati *pixels* (picture elements) e sono intesi come unità costituenti l'immagine. Affinché l'immagine possa essere trasformata in una sequenza di bit è necessario ordinare i pixel che la costituiscono secondo una determinata convenzione, ossia dal basso verso l'alto e da sinistra verso destra. In questo modo l'immagine viene trasformata in un array di pixel di dimensioni $m \times n$ chiamato *bitmap*, in cui ciascun pixel assume come valore il colore medio dell'area che rappresenta. Nella Figura 1 viene mostrato tale processo di conversione.

La precisione della codifica dipende dal numero di pixel impiegati e dalla cardinalità del dominio di valori che ciascun pixel può assumere. L'immagine ottenuta rappresenta solo un'approssimazione dell'originale, tuttavia, con una risoluzione sufficientemente alta, le limitate capacità di distinguere i dettagli da parte dell'occhio umano, fanno sì che essa appaia come una riproduzione fedele dell'immagine di partenza. Dal punto di vista dello spazio occupato in memoria, si osserva che se la codifica di ciascun pixel richiede B bit, allora la codifica dell'intera immagine richiede $m \times n \times B$ bit. Con l'utilizzo di 1 singolo bit è possibile codificare esclusivamente immagini in bianco e nero. Tuttavia, immagini di questo tipo sono poco frequenti rispetto ad immagini codificate in livelli di grigio, in cui ciascun pixel è rappresentato da un numero di bit variabile fra 1 e 8, o ancor più a colori, per le quali esistono diversi sistemi di codifica, di cui i più noti sono: RGB (Red, Green, Blue) e CMYK (Cian, Magent, Yellow, Black). In alcuni casi, invece, le immagini sintetiche possono essere memorizzate come combinazione di un insieme di primitive geometriche, individuate da un codice e da un serie di parametri.

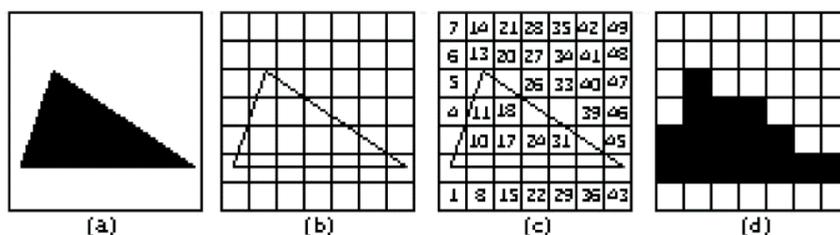


Figura 1. Fasi della conversione di un'immagine in una bitmap (a) acquisizione, (b) suddivisione in pixel, (c) numerazione dei pixel, (d) approssimazione.

Perché comprimere il multimedia

In molte applicazioni è necessario elaborare o immagazzinare nella memoria del calcolatore un numero considerevole di immagini, anche di grosse dimensioni. Si pensi ad esempio ai sistemi GIS, ai processi di telerilevamento satellitare o a

immagini istopatologiche ottenute da microscopi elettronici. Risulta immediatamente evidente che si tratta di una enorme mole di dati da gestire; basti pensare che una singola immagine bitmap di un vetrino scansionato al microscopio elettronico con risoluzione 50000×50000 pixel, in cui ciascun pixel è codificato mediante il sistema RGB, con 24 bits, occupa in memoria 480.000.000.000 bit, ossia 55,8 Gb. Una qualunque tecnica in grado di ridurre in maniera significativa le dimensioni di una immagine apporta un notevole beneficio nell'ambito di tali applicazioni. Un problema analogo si ripropone nel momento in cui un'immagine, o una sequenza di immagini (video), non deve essere immagazzinata in memoria, bensì deve essere trasmessa via internet. Anche in questo caso, risulta indispensabile un processo di riduzione della quantità di dati da gestire. Sebbene un ridimensionamento della mole di dati da elaborare o da trasmettere comporti sempre dei grossi vantaggi in termini di prestazioni, non è detto che ciò sia sempre possibile nella misura in cui lo si desidera, a meno che non si è disposti a rinunciare a parte dell'informazione stessa considerata ridondante, il che costituisce lo svantaggio fondamentale nell'utilizzo della compressione. In altri termini, sono diverse le tecniche che permettono di comprimere un'immagine. Tuttavia, è possibile classificarle tutte in due categorie fondamentali: i) *lossless* – senza perdita di informazione, ii) *lossy* – con perdita di informazione.

Alla prima categoria appartengono tutti quei formati di immagine in grado di restituire, al termine del processo di decompressione, un'immagine esattamente identica all'originale, mentre tutti i formati che non sono in grado di assicurare una reversibilità assoluta del processo di compressione fanno parte della seconda categoria. L'esistenza di algoritmi di compressione *lossless*, ossia in grado di ridurre le dimensioni di un'immagine, garantendo la conservazione di tutta l'informazione in essa contenuta, potrebbe far apparire obsoleta l'esistenza di algoritmi di compressione *lossy*; in realtà una più approfondita analisi, relativa soprattutto al *rapporto di compressione* (rapporto fra la dimensione in bytes dell'immagine originale e quella dell'immagine compressa), ossia alla capacità da parte dell'algoritmo di ridurre la quantità di informazione da memorizzare, dimostra che in alcuni casi le prestazioni degli algoritmi di compressione con perdita sono di gran lunga migliori rispetto a quelle dei compressori senza perdita. Infine, poiché la perdita di qualità dell'immagine codificata con algoritmi di compressione *lossy* costituisce il fondamentale svantaggio della compressione delle immagini, ha senso cercare di quantificare in qualche modo questo fenomeno. Sebbene constatare i benefici ottenuti dal processo di compressione, in termini di spazio in memoria risparmiato, sia molto semplice, risulta più difficile quantificare obiettivamente il corrispondente degrado dell'immagine in termini di qualità. Tuttavia, sono state

definite diverse misure di qualità, le quali però, in alcuni casi, non concordano con il giudizio soggettivo fornito dall'occhio umano; ne è un esempio l'effetto di quadrettatura caratteristico di alcune immagini codificate con alti rapporti di compressione nel formato *jpeg*. In questi casi, infatti, sebbene la qualità dell'immagine appaia scadente all'occhio umano, le misure di qualità forniscono valori relativamente elevati; ciò avviene poiché mentre l'occhio umano basa il proprio giudizio su una visione d'insieme dell'immagine, le misure di qualità rappresentano la somma di un insieme di valutazioni localizzate dell'errore.

Iterated Function Systems

Nel 1977 un matematico della IBM, Benoit B. Mandelbrot pubblica il suo libro: *The Fractal Geometry of Nature* (Mandelbrot, 1983), gettando le basi di una nuova teoria matematica che prende il nome di Geometria Frattale. Il libro porta avanti una tesi molto forte, secondo cui la geometria tradizionale con le sue rette e i suoi piani non riesce a rappresentare oggetti reali, quali alberi, nuvole e montagne; la geometria frattale ne è in grado, invece, grazie alle proprietà di contrattività e di invarianza di scala. Questa intuizione ha aperto l'orizzonte ad una vasta gamma di possibilità; difatti, gli informatici, sulla base di questi nuovi concetti matematici, hanno definito un insieme di algoritmi in grado di generare, in digitale, riproduzioni molto realistiche di oggetti reali quali coste, alberi che germogliano dal terreno o profili di catene montuose, mentre i matematici hanno ora a disposizione un nuovo insieme di entità geometriche. Nel 1981 John Hutchinson dimostra che esiste un punto di contatto fra la geometria classica e la geometria frattale, oggi nota come Teoria delle Funzioni Iterate. In seguito, Michael Barnsley, un importante ricercatore del Georgia Tech, pubblica il suo libro: *Fractals Everywhere* (Michael, 1988). Nel suo libro Barnsley presenta i concetti matematici alla base dei Sistemi di Funzioni Iterate (IFS), e dimostra un importante risultato noto come Collage Theorem, il quale afferma che un IFS può essere utilizzato per rappresentare un'immagine; con la dimostrazione del Collage Theorem si fa luce su un importante aspetto dei Sistemi di Funzioni Iterate relativo al fatto che, se, in un certo senso, la teoria frattale costituisce un ottimo strumento per la realistica riproduzione di immagini relative ad entità naturali, in senso opposto si potrebbe pensare di partire da una immagine qualsiasi e ricavare un IFS che la riproduca, o quanto meno, che generi una buona approssimazione dell'originale.

Quest'ultimo problema è noto con il nome di problema inverso ed è tuttora oggi rimasto insoluto. Tuttavia, Barnsley lascia l'accademia per fondare la Iterated System Incorporated e nel Gennaio del 1988 annuncia, mediante la rivista BYTEMagazine, il suo successo nell'utilizzare le IFS nel processo di compressione

delle immagini, riportando, come esempi, immagini dai nomi molto suggestivi quali *Black Forest*, *Monterey Coast*, *Bolivian Girl* (vedi Figura 8). In ogni caso si tratta di immagini costruite manualmente, per cui l'algoritmo di Barsnley viene bollato col nome di *Graduate Student Algorithm*, brevemente riassunto nei seguenti passi: 1) assumere uno studente, 2) fornire un'immagine allo studente, 3) fornire allo studente una stanza con una workstation, 4) chiudere a chiave la porta, 5) attendere che lo studente codifichi l'immagine, 6) aprire la porta.



Figura 8. Alcune delle immagini codificate da Michael Barsnley.

Per ironia della sorte è proprio uno studente di Barsnley, Arnaud Jacquin a trovare una buona approssimazione della soluzione del problema inverso, rendendo obsoleto il Graduate Student Algorithm (Jacquin,1990). Egli fornisce una modifica dello schema di codifica fornito da Barsnley, che prende il nome di Partitioned Iterated Function System (PIFS). Questo nuovo schema rende completamente automatica la procedura di codifica delle immagini mediante PIFS, permettendo la compressione dell'immagine e lasciando intravedere notevoli sviluppi inerenti alle tecniche di compressione basati sulla teoria dei frattali. Nonostante la scarsità di compressori/decompressori commerciali basati sui frattali si osserva comunque una larga diffusione di software gratuito, che il più delle volte, però, ha carattere esclusivamente accademico o di ricerca.

Codifica Frattale

I frattali su cui si basano le tecniche di compressione delle immagini non sono quelli definiti sul piano complesso, come l'insieme di Mandelbrot o gli insiemi di Julia, ma sono quelli basati sulla Teoria delle Funzioni Iterate. Nella maggior parte dei casi un siffatto sistema matematico viene presentato attraverso una calzante metafora, che vede i Sistemi di Funzioni Iterate come una sorta di macchina fotocopiatrice non convenzionale che prende il nome di MRCM (Multiple Reduction Copying Machine) caratterizzata da tre aspetti fondamentali:

- 1) Essa permette diversi posizionamenti dell'obbiettivo al fine di generare più copie sovrapposte dell'originale.
- 2) Ciascuna scansione riduce le dimensioni dell'immagine originale.
- 3) Essa opera a circuito chiuso, in quanto l'output di una fase di copia rappresenta l'input della fase successiva.

In altre parole, è possibile immaginare un IFS come un insieme di trasformazioni contrattive che mappano una regione rettangolare del piano in una sezione ridotta di tale regione. Nella maggior parte dei casi ci si serve di trasformazioni affini, in quanto offrono la possibilità di traslare, ridimensionare, inclinare e ruotare i punti appartenenti ad una determinata regione del piano. Un esempio è dato dal noto triangolo di Sierpinsky, illustrato nella Figura 3

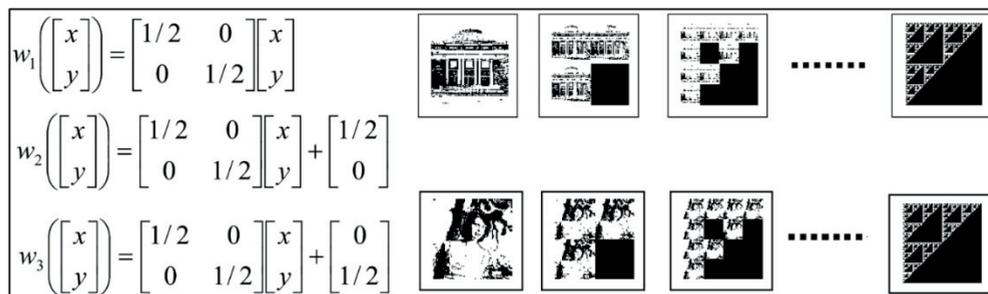


Figura 3. Il sistema di funzioni iterate che genera il triangolo di Sierpinsky.

La stupefacente proprietà di un IFS consiste nel fatto che, a prescindere dall'immagine di partenza, quando esso viene applicato iterativamente sull'immagine di input, l'immagine di output che si ottiene è sempre la stessa ed è unica; questa immagine prende il nome di punto fisso o attrattore dell'IFS. Questo risultato è dimostrato formalmente da un teorema, che prende il nome di *Contraction Theorem*. L'idea di servirsi dei frattali per attuare processi di compressione delle immagini si fonda principalmente su due osservazioni:

- 1) Molte entità del mondo reale possiedono proprietà di autosimilarità.
- 2) È possibile trovare un IFS il cui attrattore sia una buona approssimazione dell'immagine di partenza specificando un esiguo numero di trasformazioni. Dato che per rappresentare ciascuna trasformazione è sufficiente un limitato numero di parametri, un'immagine può essere rappresentata in maniera molto compatta, risparmiando così una considerevole quantità di spazio o di tempo di trasmissione dei dati, a seconda delle applicazioni. Tuttavia, gli oggetti del mondo reale non

rispettano perfettamente le leggi della geometria frattale, una felce reale può essere molto più irregolare di quella ottenuta mediante l'algoritmo di Barseley; in ogni caso al fine di poter tenere conto di queste dissimilarità è preferibile impiegare i Partitioned Iterated Function Systems.

In un PIFS le trasformazioni non mappano l'intera immagine in una sua parte, ma semplicemente modellano ampie regioni della figura di partenza in regioni di dimensioni inferiori. La denominazione scelta da Jacquin è di *domain* per la regione di input e di *range* per la regione di output, mentre la suddivisione in range dell'immagine è detta *partizionamento*. Un vincolo fondamentale consiste nel fatto che ciascun pixel dell'immagine deve appartenere esclusivamente ad un range. Un siffatto insieme di trasformazioni è ancora contrattivo e, quando viene applicato iterativamente ad una qualsiasi immagine di input, converge rapidamente al proprio punto fisso. Definire un PIFS il cui attrattore corrisponda all'immagine desiderata significa assegnare a ciascun range il domain che meglio lo approssima, secondo una trasformazione affine della quale bisogna calcolare i parametri. Se costruito in maniera opportuna il PIFS costituisce una codifica dell'immagine di dimensioni notevolmente inferiori rispetto all'originale, nonostante ne rappresenti solo una buona approssimazione. In linee generali il processo di codifica frattale (vedi Figura 4) consiste di cinque passi fondamentali:

- 1) Partizionamento dell'immagine in un insieme di range.
- 2) Suddivisione dei domain in classi.
- 3) Selezione del tipo di trasformazioni da applicare ai domain.
- 4) Scelta di una funzione di misura fra le regioni dell'immagine.
- 5) Specifica di un criterio di confronto fra range e domain.

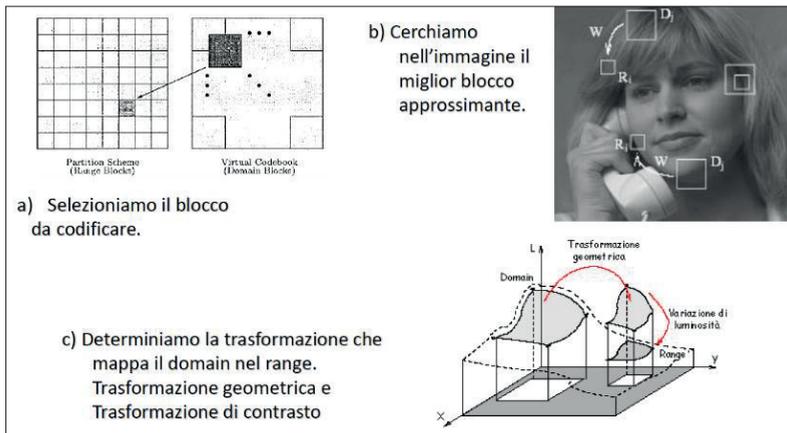


Figura 4. Rappresentazione schematica del processo di codifica frattale

Il processo di decompressione viene applicato partendo da un'immagine completamente piatta, alla quale viene iterativamente applicato l'insieme di trasformazioni affini specificato nel file compresso. Come è mostrato in Figura 5, dopo circa quattro iterazioni il metodo converge al punto fisso stabilizzandosi. L'immagine risultante non costituisce una replica esatta dell'immagine originale, ma una sua copia abbastanza fedele.

Oltre alla velocità della decodifica le tecniche di compressione basate su IFS godono della proprietà dell'indipendenza dalla risoluzione dell'immagine. In altre parole, ottenuta la codifica dell'immagine originale, la decodifica può realizzare una diminuzione della risoluzione originale dell'immagine (zoom out) oppure un aumento della risoluzione stessa (zoom in).

La possibilità di effettuare le operazioni di zoom in e zoom out scaturisce dal fatto che la codifica dell'immagine è realizzata utilizzando unicamente un pool di funzione contrattive, ognuna opportunamente parametrizzata, ma tutte definite mediante un dominio (domain) e codominio (range). In altre parole, ogni funzione contrattiva F è definita come un'applicazione che mappa un domain in un range, ossia: $F: D \rightarrow R$, dove D rappresenta l'insieme dei domains, mentre R l'insieme dei ranges. Se ciascun domain viene moltiplicato oppure diviso per 2, ed ovviamente la stessa cosa è ripetuta anche per ciascun range, allora l'immagine decodificata risulterà avere rispettivamente una risoluzione doppia (zoom in) oppure una risoluzione uguale alla metà dell'immagine originale (zoom out). Ovviamente le operazioni di zoom in e zoom out si possono ripetere per valori maggiori e minori di 2.

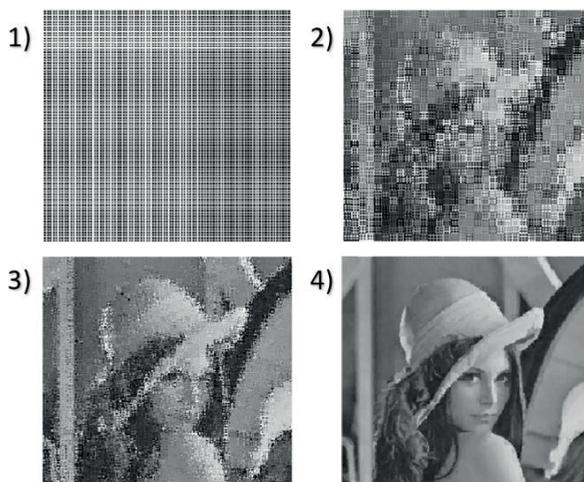


Figura 5. Il processo di decodifica frattale.

Content Based Image Retrieval

La codifica frattale sfrutta il principio di auto-somiglianza presente in un'immagine, rappresentando quest'ultima come un insieme di trasformazioni affini contrattive.

L'immagine può essere descritta in termini di sé stessa, memorizzando solo le trasformazioni necessarie per ottenerla, con un notevole risparmio, considerato che una trasformazione occupa solo poche decine di byte.

Pertanto, grazie alla sua compattezza, derivante dalla proprietà dei frattali di essere descritti con un esiguo numero di informazioni e alla sua stabilità, l'approccio frattale può essere utilizzato per ottenere un'unica firma (signature), che identifica l'immagine e ne permette il recupero. In tal modo, si ottiene un sistema di indicizzazione adatto a database di grandi dimensioni, in quanto le immagini sono manipolate in forma compressa, cioè durante la ricerca non è necessario decomprimerle.

L'efficienza di tali sistemi, contrariamente alla loro efficacia, dipende fortemente dalla velocità di compressione e dal modo in cui le caratteristiche di basso livello estratte dall'immagine sono organizzate al fine di costruire un indice.

Il sistema *Fractal Indexing with Robust Extension for Images Databases* (detto anche FIRE) (Distasiet al., 2003), è una tecnica di indicizzazione delle immagini che si basa sulla teoria dei sistemi di funzioni iterate e contrattive. Esso migliora in modo significativo in termini di efficienza e robustezza il suo predecessore FIRST (*Fractal Indexing and Retrieval SysTem for Image Databases*) (Nappi et al., 1998). Quest'ultimo, infatti, pur offrendo un buon rapporto di compressione, non risulta essere invariante a diverse trasformazioni a cui le immagini sono soggette: variazioni di luminosità, rumore, rotazioni, ecc.

Nata come una tecnica per velocizzare la compressione frattale, FIRE è stata efficacemente utilizzata per ottenere caratteristiche desiderabili (tipo l'invarianza), in un sistema di indicizzazione frattale. Sfruttando la capacità dei sistemi di funzioni iterate partizionati (PIFS) di rilevare le similarità all'interno di un'immagine, alcune informazioni estratte durante il processo di codifica vengono rielaborate in modo da formare un indice mediante il quale inserire l'immagine all'interno del database. L'indice ha una struttura gerarchica, divisa in due livelli. Durante il processo di codifica dell'immagine, FIRE estrae informazioni sulla posizione del centro di massa di ciascun range, della posizione del domain approssimante e del fattore di contrazione. La prima informazione viene utilizzata per la realizzazione dell'indice primario, mentre le seconde due per la determinazione dell'indice secondario.

I sistemi biometrici basati sui frattali

Negli ultimi anni, l'attenzione dei ricercatori si è rivolta alla gestione della sicurezza nell'ambito dell'accesso in aree riservate. In particolare, il problema dell'autenticazione è diventato un interessante campo di ricerca. Quando il soggetto non vuole essere identificato o ignora che l'autenticazione è in corso, gli approcci più affidabili, come il riconoscimento basato sull'iride o sull'impronta digitale, diventano impraticabili. Al contrario, il riconoscimento facciale può essere utilizzato con un'alta probabilità di successo

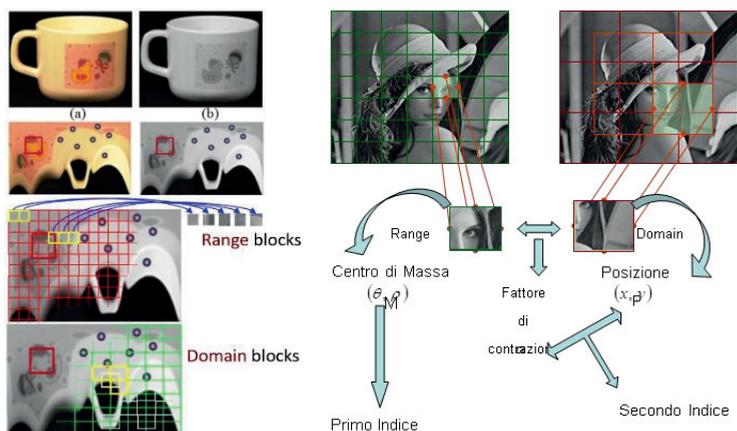


Figura 6. Firma frattale di una immagine.

Basato sulla teoria dei sistemi IFS, FARO, acronimo di *Face Recognition in Occluded Conditions* (Nappi e Riccio, 2005), è un nuovo metodo di riconoscimento facciale in presenza di occlusioni naturali e/o artificiali. Uno dei suoi vantaggi è che l'informazione usata per l'indicizzazione e il riconoscimento viene considerata a livello locale, rendendo il sistema più robusto a diverse tipologie di occlusioni. Infatti, proprio per raggiungere tale scopo, il processo di estrazione delle caratteristiche è reso locale alla regione di interesse, definita come l'unione di quattro aree principali: occhio sinistro, occhio destro, naso e bocca. Per ognuna di queste aree, viene estratto un insieme di punti e viene calcolato l'errore di approssimazione medio, così che le posizioni dei punti e gli errori di approssimazione rappresentano la firma per l'immagine facciale. La regione del volto estratta dall'immagine di input è normalizzata alla dimensione di 256×256 pixel. Nessuna operazione è stata fatta per eliminare le deformazioni dell'immagine di input, che possono influire sul tasso di riconoscimento. Le parti del volto sono indipendentemente indicizzate tramite i sistemi IFS come regioni separate di interesse, come mostrato in Figura 7. FARO è

stato valutato su una classe di immagini omogenea, ovvero volti di persone acquisiti e catalogati a scopo di ricerca. I risultati sperimentali hanno evidenziato la sua robustezza nel distinguere i volti rispetto ai cambiamenti di espressione e alle occlusioni. In particolare, anche i cambiamenti di espressione possono essere visti come occlusioni facciali, che tuttavia influiscono meno sul processo di autenticazione, rispetto ad accessori tipo sciarpe o occhiali

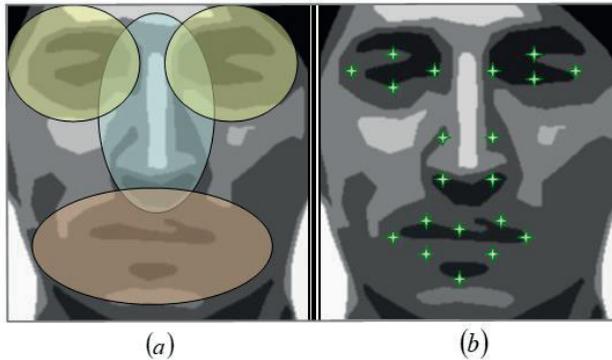


Figura 7. Regioni di interesse del volto codificate con PIFS.

Ulteriori campi di applicazione

La larga diffusione delle tecnologie informatiche, favorita dal progresso tecnologico e dal continuo calo dei prezzi ha indotto un intenso processo di trasformazione dell'informazione che dal suo essere quasi puramente testuale è passata ad una forma sempre più multimediale. Tutto ciò, da un punto di vista dell'informazione visiva si traduce in una quantità sempre crescente di immagini e sequenze video presenti in ogni documento di sorta. Contemporaneamente la continua crescita ed espansione del world wide web ha fatto sì che larga parte di questa informazione potesse essere resa disponibile ad una utenza sempre più ampia. Tuttavia, il progresso tecnologico non è ancora ad un livello sufficiente a garantire il trattamento di questi dati nella loro forma originaria, basti pensare alla trasmissione di immagini e sequenze video in rete, né del resto è accettabile che la ridondanza presente in questi stessi dati induca sprechi in termini di risorse di spazio e tempo allocabili ad ulteriori insiemi di informazioni. In questa ottica si giustifica quindi l'applicazione di tecniche di compressione che costituiscono uno strumento utile nell'aumentare le prestazioni in termini di utilizzo delle memorie di massa e di sfruttamento della banda nella trasmissione dei dati multimediali.

Anche il web così come l'ambito del trattamento dell'informazione in genere ha subito l'influsso del processo di valorizzazione della multimedialità dei dati, passando da una navigazione quasi totalmente testuale ad una veste grafica sempre

più curata e predominante rispetto ad altri aspetti. Vista dunque la massiccia presenza di informazioni visuali si accentua fortemente la necessità di strumenti che ne rendano allo stesso tempo leggera ed efficiente la rappresentazione, al fine di poter garantire una navigazione veloce ed efficace delle informazioni contenute nelle diverse pagine del web. La compressione frattale delle immagini offre diversi vantaggi in questo senso, grazie soprattutto agli alti rapporti di compressione che è in grado di offrire, ma soprattutto anche sulla base delle peculiarità del processo di decodifica, che procede per iterazioni successive, ciascuna delle quali aggiunge un grado di dettaglio via via crescente.

Il processo di decodifica frattale offre un'importante caratteristica ai fini di una efficace fruizione delle informazioni visuali contenute all'interno di una pagina web. Infatti, la decompressione di un'immagine codificata mediante sistemi PIFS procede per iterazioni successive, ciascuna delle quali aggiunge un maggiore livello di dettaglio, finché il sistema di equazioni alla base della codifica non raggiunge il punto di convergenza. Questa caratteristica offre la possibilità di avere durante la navigazione di pagine web solo una parziale, ma immediata decodifica dell'immagine, in grado di dare all'utente un'idea precisa del suo contenuto, senza che quest'ultimo debba attendere la totale decodifica, cosa che avviene solo nell'eventualità che risulti realmente interessato all'informazione contenuta nell'immagine in esame. Sulla base di questa ipotesi è anche possibile raffinare il procedimento in modo da poter utilizzare questo strumento ai fini della sicurezza, esistono lavori consolidati, che sviluppano questa idea. Infatti, è possibile associare all'immagine codificata una password, in modo tale che una parziale decodifica permetta di comprendere la natura dell'informazione in esame, ma solo con la conoscenza della password, l'utente può usufruire di tale informazione, eventualmente ai fini di un possibile acquisto.

In questi ultimi anni l'utilizzo di immagini catturate via satellite (remote sensing) ha preso sempre più piede soprattutto in ambiti applicativi quali il monitoraggio di aree protette, la prevenzione e la stima di catastrofi, ecc. Di conseguenza anche in questo ambito è fortemente sentita la necessità di ridurre la quantità di risorse occupate, soprattutto in termini di memoria di massa, da parte dei dati trattati, prevalentemente immagini multi-band. Sebbene il processo di compressione frattale di base sia progettato per immagini in toni di grigio, esso è facilmente estendibile al caso multi-band, considerando in maniera separata ciascuna banda. Infatti, una delle tre bande viene compressa con l'approccio classico, mentre la codifica delle altre due viene ricavata da questa, ottimizzando ulteriormente il rapporto di compressione.

Nell'ambito medico la necessità di tecniche di compressione efficienti ma anche efficaci è di fondamentale importanza, soprattutto dopo l'avvento di tecnologie

sempre più sofisticate che permettono acquisizioni di dati a risoluzioni molto alte o in alcuni casi in tre dimensioni per una ricostruzione tridimensionale dell'informazione originale. L'applicazione di tecniche di compressione senza perdita, sicuramente garantisce una buona affidabilità del risultato, tuttavia fornisce rapporti di compressione poco significativi che spesso non vanno oltre 3:1. Gli standard esistenti per la compressione con perdita si basano per la gran parte sulla trasformata DCT (*Discrete Cosine Transform*). Tuttavia, è stato dimostrato che tecniche di questo tipo non sono applicabili nel caso di immagini mediche. Esistono invece studi che esplorano la possibilità di applicare la compressione frattale ad immagini mediche, soprattutto nel caso di tomografie, raggiungendo rapporti di compressione di 10:1 senza perdere alcuna informazione utile ai fini della diagnosi. Ne deriva che la compressione frattale può essere applicata utilmente ad immagini tomografiche e nell'ambito della telemedicina.

Bibliografia

- Mandelbrot, B. B. (1983). *The fractal geometry of nature/Revised and enlarged edition*. whf.
- Michael Barnsley (1988). *Fractals everywhere*. Academic Press Professional, Inc., USA.
- Jacquin, A. E. (1990). *Fractal image coding based on a theory of iterated contractive image transformations*. In *Visual Communications and Image Processing'90: Fifth in a Series* (Vol. 1360, pp. 227-239). International Society for Optics and Photonics.
- R. Distasi, M. Nappi, M. Tucci, (2003). *FIRE: Fractal Indexing with Robust Extension for Images Databases*, IEEE Transactions on Image Processing, vol.12, no 3, March 2003.
- M. Nappi, G. Polese, G. Tortora (1998). *FIRST: Fractal Indexing and Retrieval SysTem for Image Databases*, Image and Vision Computing (Elsevier), vol. 16, no 14, pp. 1019-1031, 1998.
- M. Nappi, D. Riccio (2005). *FARO: Occluded FAcE Recognition by means of the IFS*. Proceedings of the Int. Conf. On Image Analysis and Recognition ICIAR 2005, vol.3656, pp. 1073-1080, Toronto, Canada, 2005

Il futuro degli studenti: dal questionario all'analisi statistica dei dati testuali¹

Lucio Palazzo, Francesco Santelli, Giancarlo Ragozini
Dipartimento di Scienze Politiche, Università di Napoli Federico II

Abstract

Il presente lavoro si incardina nell'ambito della Summer School "Piano Lauree Scientifiche" (PLS), ponendosi come obiettivo la presentazione e la discussione delle potenzialità dell'analisi di alcune tipologie dati complessi. In particolare, il focus è stato sull'analisi dei *dati testuali*, settore che ha visto negli ultimi anni un esponenziale sviluppo, anche a partire dalla diffusione e dalla rilevanza assunta dai social media e dai social networks. Discuteremo in questo lavoro l'introduzione al *text mining*, così come presentato in sede di seminario, e i principali risultati di un esperimento svolto durante tale incontro, in cui la partecipazione della platea di studenti è stata parte integrante dello studio. Verrà introdotto quindi il metodo di rilevazione con il quale è stato possibile ottenere le risposte dei singoli studenti, i principali risultati ottenuti e la discussione svolta sulla validità degli strumenti e sull'interpretazione dei risultati, questione questa tutt'altro che banale data la natura complessa del dato. L'obiettivo del lavoro è mostrare le implicazioni metodologiche e la potenzialità di analisi statistiche su dati testuali, tema diffusamente trattato nei corsi di Laurea della Federico II: Statistica per l'Impresa e la Società (triennale) e Scienze Statistiche per le Decisioni (magistrale) presso il Dipartimento di Scienze Politiche.

Introduzione

Il presente lavoro è la naturale prosecuzione e sviluppo delle riflessioni avvenute durante il seminario dell'8 settembre 2020. I principali obiettivi di tale seminario, e dunque di questo contributo, possono essere sintetizzati in tre punti: 1) introduzione ai concetti dell'analisi dei dati testuali, 2) discutere su come sia possibile ottenere una rilevazione di dati testuali in prima persona, con un esperimento effettuato in tempo reale, 3) discutere i principali risultati di tale esperimento, e validarne la

¹ Questo contributo riassume due seminari dal titolo "L'analisi statistica delle reti sociali" (Giancarlo Ragozini) e "Il linguaggio della statistica e la statistica del linguaggio" (Lucio Palazzo e Francesco Santelli)

metodologia, discutendone potenziali limiti e proponendo eventuali campi di applicazione di tali tecniche.

Il contributo è organizzato, coerentemente con i tre obiettivi, come segue: una prima parte (capitolo 1) mostrerà le definizioni e i concetti principali associati all'analisi statistica dei dati testuali, così come discussi con gli studenti. Una seconda parte (capitolo 2) mostrerà il breve questionario, così come somministrato, e l'utilizzo degli strumenti informatici associati. Una terza parte (capitolo 3) sarà dedicata alla procedura di passaggio dalle risposte aperte a dati testuali veri e propri, organizzati e strutturati rigorosamente, pronti quindi a poter essere analizzati attraverso le opportune metodologie statistiche. Infine, verranno discussi alcuni risultati dell'applicazione, e le implicazioni (e, ove presenti, anche le limitazioni) di studi quantitativi di questa natura.

Capitolo 1: cenni alla statistica applicata ai dati testuali

Le definizioni proposte di cosa sia l'analisi statistica dei dati testuali sono molteplici. D'altronde, le discipline che hanno affrontato tale discussione teorica e metodologica, ponendo l'enfasi anche su molteplici e diversi approcci, sono svariate. Tuttavia, tutte partono da una rigorosa definizione di cosa sia il dato testuale; in senso lato, tali definizioni non possono che partire dai concetti di linguaggio e testo. Il dato testuale prende sicuramente linfa da ciò che il *linguaggio* rappresenta.

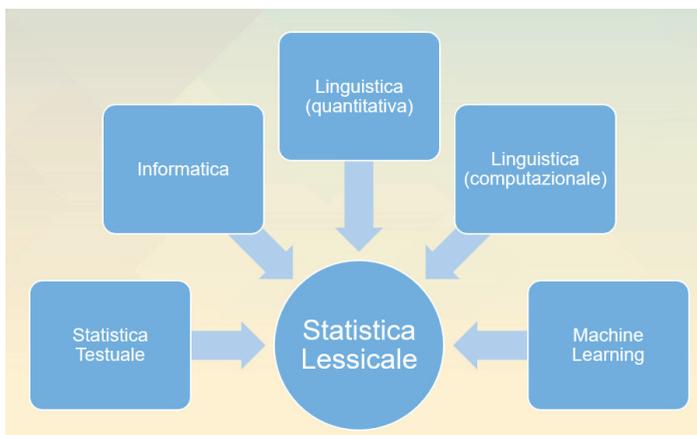


Figura 1. Interconnessioni tra diverse discipline associate alla Statistica Lessicale

“In senso generico, linguaggio è un termine che designa qualunque codice destinato a esprimere o comunicare significati. In senso più tecnico, il termine indica la facoltà, fondata biologicamente e tipica di talune specie animali (e in particolare quella umana), di collegare significati con determinati segnali fisici. Quanto all'uomo, il

linguaggio è la facoltà che gli permette di esprimere, attraverso una varietà di canali (soprattutto le lingue verbali, ma anche codici gestuali e di altro genere), i significati che è in grado di elaborare.” (Treccani, 2000)²

Il testo scritto è chiaramente una declinazione fondante di un codice linguistico. Lo step successivo, e cioè l'articolazione del passaggio dal linguaggio scritto alla statistica testuale, pone sfide metodologiche di non poco conto, data la natura stessa del codice del linguaggio che è difficilmente descrivibile mediante regole matematiche (siano esse deterministiche o probabilistiche), tanto è vero che l'analisi qualitativa dei testi riveste ancora oggi un ruolo di primo piano nella ricerca sociale. Come evidenziato da una delle massime esperte mondiali di text mining, Marti Hearst, le difficoltà le sfide dell'analisi quantitativa del testo, è quella di passare da una vasta mole di informazioni non aggregate e difficilmente classificabili a qualche interpretazione di natura statistica e, quindi, quantitativa:

“Text expresses a vast, rich range of information, but encodes this information in a form that is difficult to decipher automatically” (Marti Hearst, 1999)

A partire da tali obiettive difficoltà, l'obiettivo del text mining (analisi dei dati testuali) è dunque quello di elaborare metodi ed algoritmi per estrarre in maniera automatica informazioni dal testo, informazioni che possono poi essere utilizzate, ad esempio, per *classificare* o *raggruppare* o *descrivere* i documenti di testo. Tali documenti, com'è ovvio, possono essere delle forme e provenienti dai contesti più svariati: protocolli ufficiali, risposte aperte, tweets, post Facebook, paragrafi di libri, discorsi pubblici di politici ecc. Talvolta, occorre addirittura ipotizzare e proporre metodologie ad hoc, dato il contesto e la forma che tali documenti possono assumere. Alcuni degli obiettivi più comuni all'analisi statistica dei testi possono essere i seguenti:

- Descrivere il vocabolario (insieme di parole utilizzate) di una serie di documenti. *Esempio: Risposte aperte*
- Classificare e raggruppare testi, sulla base del lessico e del contesto utilizzato
- «Misurare» il grado di approvazione/disapprovazione di un determinato tema. *Esempio: Sentiment Analysis*
- Individuare la presenza di «neologismi» (termini nuovi)
- Valutare differenze nella tipologia di risposta tra gruppi di individui. *Esempio: uomini e donne utilizzano un lessico diverso?*

² [https://www. Treccani.it/enciclopedia/linguaggio_%28Universo-del-Corpo%29](https://www.Treccani.it/enciclopedia/linguaggio_%28Universo-del-Corpo%29)

In generale, dati considerati *strutturati* sono i dati conservati in database, organizzati secondo schemi e tabelle di rigida e consolidata formattazione. Questa è la tipologia di dati più indicata per i modelli di gestione relazionale delle informazioni. Nel caso dei dati testuali, tale struttura è tutt'altro che agevole da ottenere. Molto spesso, nel caso del text mining, si parte però da dati *non strutturati*, sono i dati conservati senza alcuno schema, solitamente testi, immagini, video e altro. In questo caso, i sistemi di gestione di dati utilizzabili sono quelli basati sul modello dell'information retrieval. L'enorme quantità di informazioni contenuta nei dati non strutturati non può essere semplicemente elaborata dai classici metodi informatici di analisi, ma è richiesto lo sviluppo di tecniche ed algoritmi capaci di scoprire ed estrarre *patterns* utili da questi dati; da qui lo sviluppo continuo di nuove tecniche di text mining. La produzione dei testi, inoltre, è fortemente associata all'evoluzione dei *mezzi* con cui tali testi vengono creati e diffusi; pertanto, lo sviluppo tecnologico si comporta come una sorta di leva, nell'amplificare la velocità con cui le tecniche statistiche di analisi dei dati testuali devono essere continuamente sviluppate e validate.

Applicare i metodi di text mining su grandi insiemi di documenti di testo è un'operazione, come detto, piuttosto complessa. Per questo motivo è necessario che i dati testuali (solitamente non strutturati) estratti da questi documenti siano convertiti in dati strutturati e che siano rappresentati adeguatamente, al fine di facilitarne la loro successiva analisi. Tipiche tecniche di trattamento iniziale di dati testuali non strutturati (*pre-trattamento*) sono:

- Tokenizzazione – è un modo di separare il testo in unità più piccole chiamate *tokens*. Tali tokens possono essere parole, caratteri, o sottoparole
- Filtraggio – eliminare parole, o gruppi di parole, non utili ai fini dell'analisi, come le *stopwords*
- Lemmatizzazione – ridurre le parole alla forma “lemma”
- Normalizzazione – rendere forme diverse, ma con stesso significato, comparabili (obiettivo vs obbiettivo)
- Stemming – ridurre tutte le parole derivanti dalla stessa parola origine, alla forma semplice di base

Esistono numerose altre procedure di pre-trattamento dati, ma queste sono tra le più diffuse ed utilizzate, soprattutto in un'ottica esplorativa e descrittiva di documenti scritti.

Si ricordi, infine, che gli sviluppi della metodologia dell'analisi statistica dei dati testuali si adatta al contesto linguistico, con sviluppi paralleli ma differenti nei diversi Paesi, seguendo la lingua principale di riferimento. In Italia, ad esempio, si

segnala la scuola derivante dai lavori di Bolasco, tra cui Statistica testuale e text mining: alcuni paradigmi applicativi (2005).

Particolarmente interessante è la combinazione della tecnica della Social Network Analysis (SNA) applicata a documenti scritti, creando la cosiddetta “semantic network” (Sowa, 1987). La teoria dei grafi presuppone una struttura di dato relazionale, nel quale sono presenti individui (nodi, in questo caso *parole*) legati o meno da relazioni (archi o link, in questo caso possibili misure legate alla similarità o co-occorrenza). In questo caso, è possibile creare una sorta di mappa in cui gruppi di parole che occorrono spesso insieme negli stessi documenti (o nelle stesse frasi) si trovano nella stessa area del grafo, con forte interconnessioni tra loro.

Come estensione della WordCloud e tenendo conto delle relazioni delle parole a due a due, è stata anche proposta una rappresentazione di tipo *bi-grams* (Bekkerman, 2004), in grado di evidenziare quali coppie di parole rivestano un ruolo cruciale nell’insieme delle risposte aperte.

Dopo aver discusso tali concetti introduttivi e possibili sviluppi di tecniche derivanti dal text mining, gli studenti sono stati coinvolti, come spiegato nel prossimo capitolo, ad un esperimento svolto durante il seminario.

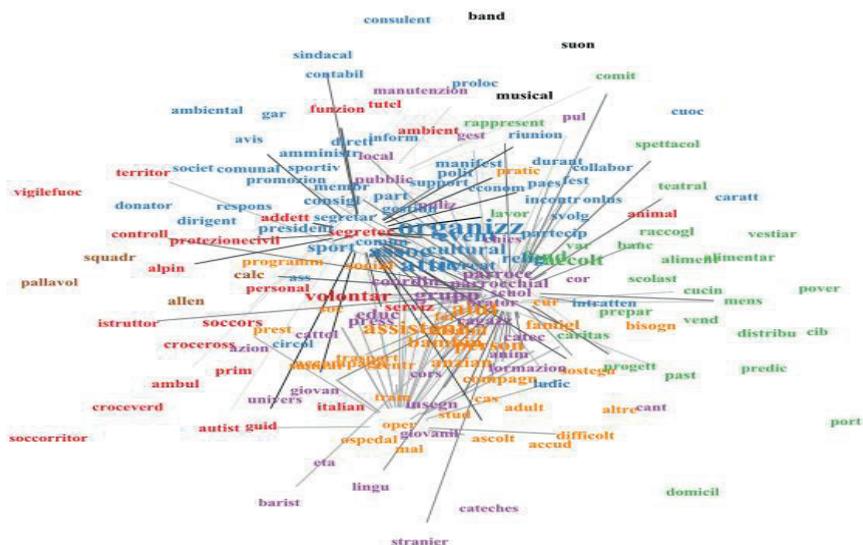


Figura 2: semantic network mostrata a scopo didattico, tratta da Santelli et al., *What volunteers do? A textual analysis of voluntary activities in the Italian context*, 2018

Capitolo 2: la rilevazione interattiva e guidata del dato testuale

In questo capitolo, verrà posta l’attenzione al coinvolgimento degli studenti nell’effettuare una rilevazione del dato testuale contestualmente al seminario. Dopo

l'introduzione al text mining così come evidenziata nel capitolo 1, è stato mostrato agli studenti come poter utilizzare un insieme di differenti strumenti informatici e computazionali per effettuare una analisi in prima persona. In tale procedura ci si è avvalsi principalmente dei seguenti strumenti: 1) La piattaforma Google moduli per creare, somministrare e rispondere al questionario 2) Microsoft Excel (o l'equivalente OpenOffice Calc) per conservare, pre-elaborare e strutturare i dati testuali 3) generatori di WordCloud (nuvole di parole) online per alcune applicazioni, in particolar modo analisi esplorative.

Per ragioni di velocità e semplicità, è stata posta la stessa sola singola domanda ad ogni studente. Attraverso un link condivisibile, è stato possibile ottenere un file unico ma con tutte le risposte adeguatamente separate. La domanda posta è stata la seguente:

“Cosa vorresti fare dopo il diploma?”



Figura 3: screenshot della piattaforma Google utilizzata per interagire con la domanda aperta

Il limite posto nella risposta aperta era di 200 caratteri. Sono state ottenute 147 risposte, segno di una ampia partecipazione da parte degli studenti.



Figura 4: screenshot della piattaforma WordCloud online utilizzata per interagire con le domande aperte

Infine, è stato mostrato come apportare alcune modifiche di pre-trattamento dati su un foglio di calcolo di tipo Excel, e sono stati utilizzati alcuni siti contenenti applicativi per creare delle WordCloud online.³

Capitolo 3: principali risultati dell'applicazione e conclusioni

Hanno partecipato al questionario un totale di 147 rispondenti. Ai ragazzi è stato quindi mostrato in maniera interattiva come passare dal dato grezzo, ottenuto partendo dal form Google Moduli, ad un dato di sintesi che evidenziasse una serie di informazioni salienti riguardanti il campione stesso.

Dopo una pulizia preliminare dei dati, le risposte sono state analizzate e catalogate in quattro classi distinte:

- Studi Scientifici: appartengono a questa classe i rispondenti che hanno specificato di voler iscriversi ad un corso di laurea di tipo scientifico.
- Studi Umanistici: gruppo di rispondenti che hanno dichiarato di voler iscriversi ad un corso di laurea di tipo umanistico.
- Scienze Mediche/Sanitarie: studenti che hanno espresso la volontà di voler prendere parte a corsi di laurea di natura medica (medicina/veterinaria) o di scienze infermieristiche.
- Altro/Non Specificato: appartengono a questa classe tutte le risposte valutate troppo vaghe, contraddittorie ed ironiche.

Per la stragrande maggioranza dei rispondenti si presenta la volontà di proseguire gli studi iscrivendosi all'Università, mentre solo due rispondenti del campione in esame hanno dichiarato di voler far parte delle forze dell'Ordine, in particolare diventare Carabinieri. In questo contesto, si nota da parte di alcuni studenti un interessamento al mestiere di "criminologo", con la ricerca di un corso di studi universitario che possa portare ad una carriera da criminologo ed all'attività investigativa.

Nella Figura 5, si può notare la distribuzione delle risposte in base alle categorie individuate.

Al netto della categoria "Altro" (44%), le discipline scientifiche sembrano attirare maggiormente gli studenti intervistati (55% delle risposte) rispetto alle discipline umanistiche, che si attestano ad un 9% del totale. Risulta interessante constatata una consistente percentuale di studenti attratti dalle discipline mediche ed infermieristiche, pari al 39%, segno dell'appeal che tali discipline destano negli studenti liceali.

³ <https://worditout.com/word-cloud/create>

Analizzando nel dettaglio le risposte, si possono inoltre notare delle caratteristiche dei rispondenti che potrebbero aiutare a comprendere nel dettaglio la natura intrinseca del fenomeno. Come si poteva già notare dalla forte percentuale di risposte appartenenti alla categoria "Altro" (mostrata nel grafico 5), si evince una forte indecisione nei confronti del futuro da parte degli studenti, contraddistinta a volte dall'utilizzo di verbi al condizionale o, nello specifico, all'utilizzo di termini quali, tra i tanti, "non so" e "non ho ancora deciso".

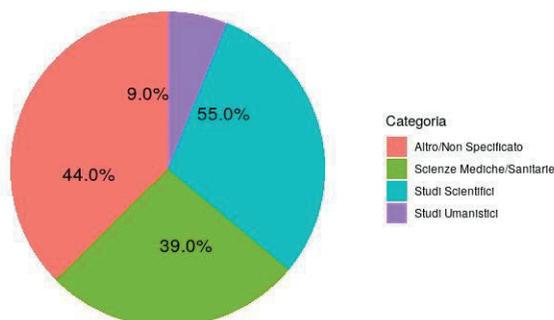


Figura 5: grafico a torta con le frequenze percentuali delle categorie di risposta

Anche il senso di incertezza è una componente che si evince con una certa frequenza attraverso frasi quali "probabilmente", "se ci riesco" e "se ne ho la possibilità". Al contrario, sono pochi i ragazzi che dichiarano una preferenza in maniera chiara e determinata il proprio percorso futuro ma, quando lo fanno, forniscono una descrizione esaustiva e molto dettagliata.

In questo contesto, alcuni rispondenti hanno evidenziato (sia in maniera diretta che indiretta) l'utilità che hanno avuto i webinar organizzati dalla Summer School PLS. Infatti, alcuni studenti hanno dichiarato apertamente di partecipare a tali corsi al fine di fugare potenziali dubbi riguardanti il tema di interesse trattato durante i vari seminari, mentre un paio hanno risposto di esser arrivati ad una scelta più decisa proprio dopo aver seguito i suddetti corsi.

In alcune risposte si può inoltre osservare una propensione alla mobilità universitaria fuori regione: non mancano infatti preferenze relative allo studio fuori regione/città di appartenenza (con i termini "non a Napoli" oppure "fuori dalla Campania" ad esempio) e addirittura alla mobilità internazionale ("Fare esperienza all'estero per poi rimanere all'estero o tornare").

Dopo aver effettuato il pre-trattamento del testo attraverso le fasi descritte in sezione 1, rimuovendo in particolare stop-words e segni di punteggiatura, siamo passati

quindi all'analisi delle parole. Sono state estrapolate dal testo 344 parole distinte, delle quali l'84% sono state utilizzate al più due volte (72% una sola volta e 12% due volte), al quale corrispondono sia termini generici ("iscrizione", "attuale", "riguardante", "incline") che più specifici ("biotecnologie", "aerospaziale", "fisioterapia", "criminologia"). Pertanto, da qui in poi focalizziamo la nostra analisi allo studio delle parole che sono state utilizzate con la maggiore frequenza.

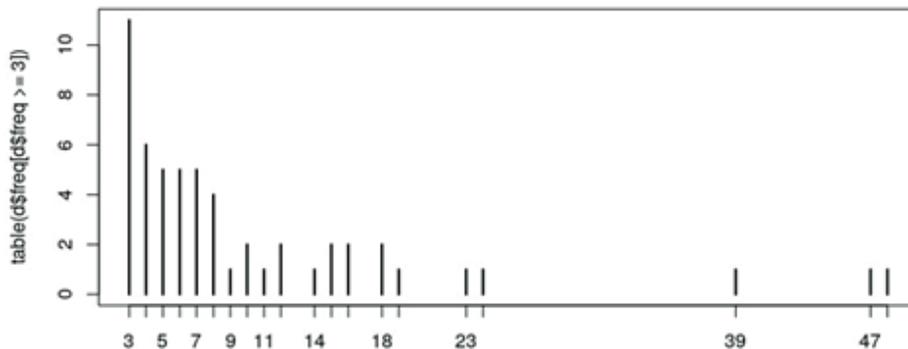


Figura 6: diagramma a barre con la distribuzione delle frequenze delle parole usate almeno 3 volte

Nel grafico riportato in Figura 6, si evince la distribuzione di frequenza delle parole utilizzate almeno tre volte all'interno del campione, che sono pari a 55. Si tratta chiaramente di una distribuzione fortemente asimmetrica a cui corrisponde un valore modale pari a 3. Sono 36 le parole che sono state utilizzate dagli studenti con una frequenza compresa tra 3 ed 8 volte, mentre 19 sono le parole a cui corrisponde una frequenza più alta, come mostrato in Tabella 1.

Tra le parole più rilevanti spiccano di certo alcuni sostantivi e verbi all'infinito relativi all'Università (come "università", "medicina", "ingegneria", "facoltà", "iscrivere", "intraprendere" e "studi") e anche alcuni verbi coniugati al condizionale ("piacerebbe", "vorrei"), che esprimono quella percezione di incertezza del futuro già osservata in precedenza.

Seguendo un'ottica di esperienza interattiva e votata al coinvolgimento attivo degli ascoltatori, si è mostrato ai ragazzi come attraverso l'utilizzo della term-document matrix è possibile costruire le WordCloud.

Per WordCloud si intende una lista "pesata" di parole, con la peculiare caratteristica di attribuire un font di dimensioni più grandi alle parole più importanti, ovvero a cui corrisponde una frequenza maggiore.

Con tali rappresentazioni grafiche si possono evidenziare alcuni aspetti fondamentali riguardanti i dati presi in considerazione.

Parola	Frequenza	Parola	Frequenza	Parola	Frequenza
universita	48	fisica	8	vita	4
vorrei	47	laurearmi	8	specializzarmi	4
medicina	39	diventare	8	bene	4
iscrivermi	24	probabilmente	8	riuscire	4
fare	23	matematica	7	trovare	4
corso	19	poi	7	cosa	4
ingegneria	18	studiare	7	veterinaria	3
dopo	18	sicuramente	7	sicura	3
studi	16	voglio	7	scienze	3
diploma	16	informatica	6	biologia	3
l'universita	15	sogno	6	chirurgia	3
frequentare	15	entrare	6	chimica	3
laurea	14	scientifico	6	studio	3
ancora	12	molto	6	idee	3
intraprendere	12	estero	5	scegliere	3
andare	11	lavoro	5	indipendente	3
piacerebbe	10	continuare	5	piu	3
facolta	10	seguire	5		
percorso	9	oppure	5		

Tabella 1. Tabella di frequenza delle parole più utilizzate, tra quelle usate almeno 3 volte

In un primo tentativo si è mostrato quali sono gli effetti potenzialmente negativi che si hanno nel costruire la WordCloud senza aver effettuato la rimozione delle stop-word. In base a ciò si è costruita la Figura 7.



Figura 7. Wordcloud senza alcuna operazione di pretrattamento dei dati

Successivamente abbiamo mostrato in Figura 8 come ottenere il risultato finale, rimuovendo in particolare stop-words e segni di punteggiatura.



Figura 8. Wordcloud dopo aver effettuato operazioni di pretrattamento dei dati

Infine, è stato mostrato ai ragazzi come rappresentare le relazioni tra coppie di parole attraverso il bigramma, intesi come grafici che mettono in relazione le coppie di parole più frequenti presenti congiuntamente in ciascuna risposta.

Dal bigram in Figura 9 si evince quindi che termini legati al mondo universitario sono spesso legati tra loro, non mancano però altre relazioni interessanti quali "trovare-lavoro" e "lavoro-carabiniere" tra tutte.

Bibliografia

- Bekkerman, R., Allan, J. (2004). *Using bigrams in text categorization* (pp. 1-10). Technical Report IR-408, Center of Intelligent Information Retrieval, UMass Amherst.
- Bolasco, S. (2005). Statistica testuale e text mining: alcuni paradigmi applicativi. *Quaderni di statistica*, 7, 17-53.
- Hearst, M. A. (1999, June). Untangling text data mining. In *Proceedings of the 37th Annual meeting of the Association for Computational Linguistics* (pp. 3-10).
- Santelli, F., Ragozini, G., Musella, M. (2018, June). What volunteers do? A textual analysis of voluntary activities in the Italian context. In *International*

Conference on the Statistical Analysis of Textual Data (pp. 265-276).
Springer, Cham.

- Sowa, J. F. (1987). Semantic networks.



Figura 9. Rappresentazione grafica dei bigrammi

Seminario Integrato: Interazioni tra piante ed animali: aspetti evolutivi ed aspetti molecolari

Giovanni Scopece⁽¹⁾, Maria Rosaria Iesce⁽²⁾

(1) *Dipartimento di Biologia, Università di Napoli Federico II*

(2) *Dipartimento di Scienze Chimiche, Università di Napoli Federico II*

Abstract

Le piante sono organismi sessili che hanno sviluppato straordinari adattamenti per affrontare e risolvere problemi complessi senza potersi muovere. In alcuni casi, questi organismi apparentemente passivi sono riusciti a cooptare gli animali, sfruttandone la capacità di muoversi, per svolgere i loro processi vitali. Questo è il caso dell'impollinazione entomofila, una interazione mutualistica in cui la pianta ottiene un efficiente trasporto del polline tra individui anche lontani in cambio di ricompense offerte agli insetti impollinatori. L'origine evolutiva di questa strategia riproduttiva risale al Cretaceo ed è avvenuta più volte in maniera indipendente. Tra i gruppi viventi, l'impollinazione entomofila caratterizza la riproduzione di alcuni gruppi di Gimnosperme e della grande maggioranza delle angiosperme. Una volta affermata, questa strategia ha determinato una parallela esplosione evolutiva nelle piante e nei principali gruppi di insetti impollinatori. Questa diversificazione è stata determinata da un profondo processo co-adattativo in cui le piante, sfruttando le capacità cognitive degli insetti hanno sviluppato spettacolari tratti floreali come forme, colori ed odori in grado di attirare gli insetti. In una famiglia vegetale, le orchidee, queste interazioni hanno raggiunto dei livelli di complessità notevoli. Queste piante furono studiate in maniera molto approfondita da Darwin, il quale gli dedicò un intero libro. Molte specie di questa famiglia attirano gli insetti con straordinarie strategie basate sull'inganno. L'inganno può essere di tipo alimentare o sessuale e consente alle piante di attirare gli insetti senza produrre alcuna ricompensa. Le ricompense, i colori e gli odori dei fiori sono dovuti a molecole organiche che, grazie alla loro struttura, svolgono funzioni fondamentali. In questo articolo saranno trattate, da un punto di vista chimico alcune delle sostanze più importanti che, nel corso dell'evoluzione, hanno consentito alle piante di stabilire interazioni con gli insetti. La stragrande maggioranza delle piante a fiore ed un'importante frazione di quelle coltivate dipende direttamente dall'azione degli impollinatori. Questo meccanismo riveste dunque una

grande rilevanza sia per la biodiversità naturale ed il mantenimento degli ecosistemi che per l'economia agricola e la sostenibilità. Nell'ultimo decennio, in varie aree geografiche si è notato un declino di alcuni gruppi di impollinatori spontanei. La comunità scientifica internazionale è dunque oggi chiamata a comprendere le cause di tale declino ed a cercare di mitigarle.

Parte 1: Aspetti evolutivi

G. Scopece

Introduzione

“Le piante sono l’anello di congiunzione tra la terra e il sole” Kliment Timiryazev

Le piante sono organismi autotrofi che, attraverso la fotosintesi, trasformano l'energia luminosa in legami chimici e composti organici. Proprio per svolgere nel modo più efficiente questa funzione, esse hanno adottato nel corso dell'evoluzione uno stile di vita sessile che ha garantito ai vegetali terrestri la possibilità di conquistare uno spazio alla luce e di raggiungere notevoli dimensioni. Allo stesso tempo però un simile stile di vita ha imposto grandi sfide a questi organismi nell'affrontare altri aspetti del loro ciclo vitale. A differenza degli animali, infatti, le piante non possono svolgere le loro funzioni o rispondere alle condizioni esterne grazie al movimento o al comportamento. Le piante dunque sono completamente esposte alle condizioni ambientali. Una pianta che vive in un deserto, ad esempio, si adatta morfologicamente e fisiologicamente a sopportare temperature estremamente elevate e lunghi periodi di aridità; un animale può rifugiarsi all'interno di una tana e poi uscire quando le temperature diminuiscono. Le piante hanno *inventato* straordinarie strategie per difendersi dai predatori, per conquistare e riprodursi con un partner lontano e per disperdere la propria progenie senza potersi muovere. Le piante sono dunque organismi che dietro una natura apparentemente passiva hanno in realtà elaborato strategie molto articolate per analizzare e risolvere problemi complessi. In alcuni casi, questi straordinari adattamenti prevedono l'utilizzo degli animali in interazioni mutualistiche che consentono alle piante, in cambio di una ricompensa, di ricevere un aiuto. Questo è il caso dell'impollinazione entomofila, un processo mediato dagli insetti che consente a piante anche distanti tra loro di far incontrare i propri gameti in modo estremamente efficiente.

Nonostante sia molto comune osservare insetti volare da un fiore all'altro, fino a pochi secoli fa si riteneva che questi fossero dei ladri occasionali di nettare che non

svolgevano alcuna funzione per le piante. La piena comprensione della funzione degli insetti nel trasporto del polline ha richiesto un lunghissimo tempo e si è chiarita solo in un'epoca relativamente recente. Questo sorprendente ritardo è legato al fatto che, per molto tempo, le piante sono state considerate come esseri intermedi tra il vivente ed il non vivente e quindi accettarne la sessualità era impossibile. Il botanico tedesco Suessenguth sosteneva che esistono due tipi di scoperta: trovare qualcosa che nessuno ha visto prima o pensare a qualcosa a cui nessuno ha pensato prima su qualcosa che tutti vedono. La storia della comprensione del meccanismo di impollinazione ed in particolare dell'impollinazione entomofila ricade in questo secondo tipo di scoperta: tutti vedevano gli insetti volare da un fiore all'altro ma nessuno aveva pensato che questo potesse avere a che fare con la sessualità e la riproduzione.

Il primo a dare un contributo alla comprensione della questione fu Linneo (1707 – 1778), il quale comprese l'importanza degli organi sessuali contenuti nei fiori e li utilizzò per classificare le piante (Linnaeus, 1758). Fino a quel momento, nonostante in qualche pratica agricola fosse conosciuto il ruolo dell'impollinazione incrociata, la natura sessuale dei fiori era stata quasi completamente ignorata. Le osservazioni di Linneo si limitarono però ad una descrizione che, sebbene molto accurata, non contribuirono in maniera significativa alla comprensione funzionale. La comprensione della natura sessuale dei fiori e degli organi in essi contenuti si deve ad altri botanici. Tra questi, Rudolf Jakob Camerarius (1665 –1721), un fisico e botanico tedesco operante alla fine del XVII secolo, il quale comprese che stami e pistillo erano necessari per consentire ad una pianta di formare un seme maturo. Basandosi su questa osservazione, egli concluse che questi potessero essere considerati gli organi sessuali delle piante. Pur considerando la maggior parte delle piante autogame, Camerarius notò che in piante unisessuali come il gelso, le piante femminili (pistillifere) più lontane dalle piante maschili (staminate) producevano frutti privi di semi. I suoi risultati furono riportati sotto forma di lettera (*De sexuplantarum epistola*; Camerarius, 1694), ed attirarono una grande attenzione stimolando i botanici contemporanei. Tuttavia, bisognerà attendere ancora mezzo secolo prima di ottenere i successivi significativi passi in avanti su questo tema arrivando all'epoca di quelli che sono universalmente riconosciuti come i fondatori della biologia dell'impollinazione, Joseph Gottlieb Koelreuter (1733 –1806) e Christian Konrad Sprengel (1750 –1816).

Koelreuter fu il primo a condurre esperimenti di ibridazione con finalità scientifiche e quindi provò definitivamente che pistillo e stami erano organi sessuali delle piante. Egli descrisse piante autogame, anemofile ed entomofile ed il ruolo di ricompensa

del nettare che fino ad allora si pensava fosse prodotto per nutrire i semi (Koelreuter,1761).

Qualche anno più tardi, senza aver mai letto il lavoro di Koelreuter, Sprengel (1793) pubblicò “*Il segreto della natura scoperto nella struttura ed impollinazione dei fiori*” un fondamentale contributo alla comprensione del processo dell’impollinazione. In questo libro Sprengel riportò osservazioni su più di 400 specie di piante. Sprengel, che aveva una formazione religiosa molto rigorosa, si convinse che i fiori fossero stati *creati* proprio per attirare gli insetti e che il nettare fosse una appunto ricompensa offerta dalla pianta in cambio della fecondazione. A conferma delle sue osservazioni, egli notò che alcune piante producevano grandi quantità di polline che era trasportato dal vento e non avendo bisogno dei servizi degli insetti non producevano nettare. Egli definì gli insetti come delle *spazzole viventi che in una relazione simbiotica svolgono il compito teologico di fecondare le piante*. Sprengel comprese anche che attraverso vari meccanismi i fiori tendevano ad evitare di auto fecondarsi smentendo una volta e per tutte l’idea che le piante fossero per la maggior parte autogame.

Nel corso del XIX secolo, il grande naturalista inglese Charles Darwin influenzò profondamente la biologia dell’impollinazione. Darwin fece numerosissime ed acutissime osservazioni sulle piante e sulle loro interazioni con gli insetti, ma il suo contributo maggiore fu di una natura più profonda (Darwin,1877). Seppur con grande acume e precisione descrittiva, Sprengel come molti suoi contemporanei ancora considerava le sue osservazioni come una specie di servizio religioso utile a descrivere e rendere omaggio a ciò che il creatore della natura aveva prodotto. Darwin invece raccolse ed integrò le osservazioni di Sprengel e le lesse con la chiave della sua teoria della selezione naturale. In questo senso, dunque, l’impollinazione entomofila diventò un argomento di grande interesse per i primi evoluzionisti in quanto straordinario esempio di co-adattamento tra piante ed animali.

Pur se con alterne vicende, lo studio dell’impollinazione entomofila è giunta fino ai giorni nostri con questo spirito. Numerosi botanici nel corso del XX secolo ed in questo primo scorcio del XXI hanno continuato a svolgere osservazioni ma acquisendo sempre più un approccio quantitativo. Attualmente, la biologia dell’impollinazione è un campo di studio e di ricerca molto vivace che coniuga l’interesse per la comprensione delle dinamiche evolutive e co-evolutive con la necessità di preservarne le caratteristiche alla luce delle attuali emergenze ambientali.

L'origine dell'impollinazione entomofila e le sue conseguenze evolutive

"Nulla ha senso in biologia se non alla luce dell'evoluzione" Theodosius Dobzhansky

Le teorie volte a ricostruire l'origine evolutiva dell'impollinazione entomofila sono ancora molto dibattute. Tuttavia, è evidente dall'osservazione dei gruppi di spermatofite viventi e fossili che questo meccanismo si è originato a partire da gruppi anemofili in cui le piante utilizzavano il vento come vettore del polline.

Cosa ha comportato per le piante l'utilizzo degli insetti come vettori del polline?

La principale spinta evolutiva che ha portato al diffondersi di questo meccanismo riproduttivo sta nel fatto che esso conferisce un enorme vantaggio in termini di efficienza nell'utilizzo delle risorse. Per avere una qualche probabilità che il proprio polline raggiunga uno stigma di un individuo conspecifico una pianta ad impollinazione entomofila ha bisogno di produrre enormi quantità di polline. Di tutto questo materiale riproduttivo, infatti, solo una percentuale bassissima arriverà a destinazione mentre la maggior parte sarà persa. L'utilizzo degli insetti come vettori del polline, invece, consente di avere un trasporto molto più preciso e quindi di produrre meno polline risparmiando grandi quantità di risorse.

Le prime interazioni tra piante ed insetti sono probabilmente state casuali, con insetti (in particolare Coleotteri) che visitavano i fiori nutrendosi del polline e, sporcandosi di questa sostanza, lo trasportavano incidentalmente di pianta in pianta. In queste prime interazioni, il trasporto mediato dagli insetti era probabilmente una piccola percentuale dell'impollinazione entomofila che ancora avveniva. Da questo punto in poi però una nuova e fortissima spinta evolutiva ha portato le piante e gli insetti ad adattarsi reciprocamente. L'impollinazione ad opera degli insetti si è originata più volte indipendentemente nel corso dell'evoluzione ed è attualmente presente in diversi gruppi di Gimnosperme (come, ad esempio, in alcune Cycadophyta e Gnetophyta) e nella stragrande maggioranza delle angiosperme. Le prime testimonianze fossili dell'impollinazione entomofila sono giunte fino a noi grazie alla fossilizzazione all'interno di gocce di ambra di insetti con polline sul proprio corpo. Questi fossili riguardano Coleotteri con tracce di polline di angiosperme risalente a circa 100 MYA (Bao *et al.*, 2019) in pieno Cretaceo.

Per più di un secolo c'è stata una grande fascinazione con la sorprendentemente rapida radiazione evolutiva delle angiosperme al punto che Darwin descrisse questa esplosione evolutiva come un *abominevole mistero*. Sicuramente una corretta datazione dell'origine delle angiosperme è fondamentale per comprendere l'origine e l'evoluzione degli impollinatori ed in particolare diapi, farfalle, falene e ditteri.

Tuttavia, a dispetto del fatto che ancora esiste un dibattito sull'origine delle angiosperme, è pacifico che l'esplosione evolutiva di questo gruppo è coincisa con l'instaurarsi dell'impollinazione entomofila. Altrettanto evidente è il fatto che parallelamente all'esplosione evolutiva delle angiosperme si è verificata una grande diversificazione dei principali gruppi di insetti impollinatori, al punto che l'epoca moderna è sostanzialmente dominata in termini di diversità di specie dai partner di questa interazione mutualistica sia nel regno animale che nel regno vegetale.

Gli attrattori primari (ricompense) e gli attrattori secondari (segnali)

Come già detto, l'impollinazione entomofila è un mutualismo in cui la pianta riceve un preciso ed efficiente trasporto del polline e l'insetto ricava una ricompensa. La ricompensa che le piante offrono ai propri visitatori è definita un attrattore primario in quanto costituisce la ragione stessa per cui gli insetti frequentano i fiori. Nei gruppi più primitivi questa ricompensa è costituita dal polline stesso. Il polline è una sostanza molto energetica e ricca di proteine e costituisce l'alimento principale dei coleotteri e di alcuni imenotteri. Nei gruppi più evoluti, le piante hanno invece separato il materiale riproduttivo dalla ricompensa producendo una sostanza zuccherina chiamata nettare (**vedi Parte 2**). Il nettare può avere contenuti di zuccheri anche molto differenti ed i vari gruppi di insetti mostrano delle chiare preferenze. Per contenere il nettare le piante hanno evoluto delle strutture chiamate *nettarii* che sono riconosciuti dagli insetti.

Il nettare è senza dubbio il tipo di ricompensa più frequentemente offerto dalle piante. Tuttavia, talvolta la ricompensa per gli insetti impollinatori non è costituita da sostanze zuccherine nutritive, ma da sostanze odorose. Questo tipo di offerta caratterizza alcune orchidee sudamericane della sottofamiglia Epidendroideae. Queste piante attirano maschi del gruppo degli Euglossini che visitano i fiori per raccogliere composti odorosi. Questi composti vengono immagazzinati in organi specializzati posti sulle zampe posteriori degli insetti. Queste sostanze prodotte dalle orchidee e raccolte dagli insetti maschi sono le stesse che essi producono naturalmente per attirare le femmine. La raccolta delle sostanze prodotte dalle orchidee, dunque, aumenta il successo riproduttivo dei maschi (Zimmermann *et al.*, 2006).

Un altro interessante tipo di ricompensa si è originato in un altro gruppo di orchidee appartenenti al genere *Serapias*. Queste piante producono dei fiori tubulari che mantengono all'interno una temperatura leggermente più alta rispetto all'ambiente esterno. Questi fiori, dunque, in cambio della visita dell'insetto e del trasporto del polline, non offrono delle sostanze ma un nido all'interno del quale riposarsi.

Come abbiamo visto, le piante producono ricompense volte a conferire un vantaggio agli insetti che visitano i propri fiori. Le ricompense però generalmente non sono direttamente visibili dagli insetti. Come fanno dunque gli insetti a capire che una pianta contiene una ricompensa? Per aiutare gli insetti, e quindi sé stesse, le piante hanno sviluppato un grande numero di adattamenti che servono ad attirare gli insetti. Questi adattamenti sono anche definiti attrattori secondari e sono rappresentati da tratti fiorali non direttamente coinvolti nella ricompensa ma che gli insetti possono associare ad essa. L'evoluzione ha dunque spinto le piante a sviluppare una grande varietà di pigmenti (principalmente carotenoidi e flavonoidi), di sostanze odorose e di strutture fiorali che hanno consentito ai fiori di assumere colori, odori e forme in grado di attrarre gli impollinatori (**vedi Parte 2**). In qualche modo, l'evoluzione dei tratti fiorali da parte delle piante ha utilizzato le capacità cognitive degli insetti. I vari gruppi di insetti, infatti, sono in grado di riconoscere solo determinati colori o odori e sono morfologicamente più adatti ad alcune forme.

Nel complesso, dunque, differenti gruppi di insetti sono adattati a differenti tipi di attrattori primari e secondari. Pertanto, i fiori adattati ad un determinato tipo di insetto impollinatore, anche se lontani evolutivamente, possono subire fenomeni di convergenza evolutiva. Questo è alla base dell'idea delle sindromi di impollinazione secondo la quale esistono veri e propri *pacchetti* di adattamenti associati ai differenti insetti (Fenster *et al.*, 2004). In qualche modo questo vuole anche dire che è possibile predire gli impollinatori di una specie semplicemente osservando un fiore. Un notevole esempio di questo è rappresentato dall'orchidea *Angraecum sesquipedale* o orchidea fantasma. Questa pianta originaria del Madagascar ha un lunghissimo nettario che contiene, alla sua estremità una soluzione zuccherina. Studiando un campione di questa specie, e con in mente la sua idea di co-adattamento, Darwin predisse l'esistenza di un lepidottero con una spiritromba lunga abbastanza. Effettivamente, molto tempo dopo, questo Lepidottero fu identificato e fu chiamato *Xanthopan morgani* subsp. *praedicta*, con l'epiteto subspecifico ad indicare la previsione di Darwin.

Relazioni asimmetriche tra piante ed insetti: il caso delle orchidee

Più studio la natura, più resto impressionato con sempre maggiore forza del fatto che le strategie e gli straordinari adattamenti lentamente acquisiti attraverso la selezione naturale (...) superano in modo incomparabile le strategie e gli adattamenti che la più fertile immaginazione del più fantasioso degli uomini potrebbe inventare avendo infinito tempo a sua disposizione. Charles R. Darwin

Questo vero e proprio inno alla natura ed *agli adattamenti lentamente acquisiti attraverso selezione naturale* fu inserito da Darwin all'interno di un libro

interamente dedicato alla famiglia vegetale delle Orchidaceae. L'interesse di Darwin per le piante è poco noto tra i non specialisti. Tuttavia, Darwin dedicò moltissime osservazioni alle piante ed in particolare alle orchidee e le considerava uno dei migliori esempi della sua teoria di evoluzione dovuta a selezione naturale. Le orchidee rappresentano dai tempi di Darwin un gruppo modello per lo studio della diversificazione in ambito vegetale e soprattutto per lo studio delle interazioni tra piante ed insetti. Questa famiglia vegetale ha subito una straordinaria radiazione evolutiva divenendo probabilmente la più ricca di specie all'interno del regno vegetale (con circa 28.000 specie). Questa eccezionale diversità è strettamente legata nelle intricate relazioni stabilite con i loro insetti impollinatori. Nel suo libro *On the various contrivances by which British and foreign orchids are fertilised by insects*, Darwin utilizzò i meccanismi di impollinazione nelle orchidee come uno straordinario esempio di processo coevolutivo che, in conseguenza di un reciproco vantaggio, portava fiori e insetti a evolvere simultaneamente. Il lavoro di Darwin ha aperto la strada a numerosi studi che hanno rivelato rapporti ancora più complessi tra questi fiori e i loro impollinatori.

Come le altre piante entomofile, la maggior parte delle orchidee attira gli impollinatori offrendo una ricompensa. Le orchidee che producono ricompensa in genere offrono nettare (vedi esempi alternativi nel precedente capitolo). Nei gruppi più primitivi della famiglia (come, ad esempio, nella sottofamiglia Apostasioideae) l'offerta consiste nel polline del quale alcuni insetti si alimentano (Garay, 1960).

Perché quella delle Orchidaceae è considerata la famiglia vegetale con i più straordinari esempi di interazioni con gli insetti impollinatori?

In realtà però la coevoluzione tra orchidee ed insetti, almeno per quella che era l'idea darwiniana, è un processo relativamente raro. Molto spesso, infatti, l'evoluzione delle orchidee è invece unilaterale in quanto i propri cambiamenti non comportano una trasformazione negli insetti. Questo è tipico delle orchidee che, non producendo una ricompensa per gli insetti impollinatori, li attirano attraverso strategie basate sull'inganno. Queste strategie si sono originate più volte indipendentemente nel corso dell'evoluzione. Attualmente si ritiene che circa un terzo delle orchidee utilizza questi meccanismi.

Per ingannare gli insetti, le orchidee producono segnali attrattori secondari che sfruttano le abilità cognitive degli insetti. I principali istinti degli insetti sono rappresentati dal nutrimento e dal sesso e le orchidee ingannevoli sfruttano proprio questi processi per convincere gli insetti a visitare i propri fiori. Il meccanismo di inganno più comune è basato sulla mimesi alimentare. In questa strategia il fiore dell'orchidea imita i segnali tipici delle piante nettariifere. Come abbiamo visto in

precedenza, le piante nettariifere hanno fiori colorati, spesso vistosi e profumati ed hanno dei nettarii dove contengono il nettare. Dunque, le orchidee con mimesi alimentare producono tutti questi segnali pur non producendo nettare. Questo gruppo di orchidee attira una grande varietà di insetti che visitano i fiori alla ricerca di cibo. In alcuni casi, queste orchidee imitano uno specifico modello. Questo è il caso di *Anacamptis israelitica* che imita in maniera specifica il suo modello *Bellevallia flexuosa* in un esempio di mimetismo Batesiano (Dafni e Ivri, 1981).

L'esempio forse più interessante di orchidee prive di nettare è rappresentato da quelle che mettono in atto un meccanismo di inganno sessuale. In questo caso le orchidee imitano segnali visivi, tattili ed olfattivi delle femmine degli insetti e sono impollinate da insetti maschi attratti sessualmente dal fiore. In queste piante, l'impollinazione si verifica quando gli impollinatori tentano di accoppiarsi con il fiore. Questo meccanismo di impollinazione si è evoluto in vari gruppi di orchidee e, nel mediterraneo, nel genere *Ophrys*. L'attrazione in queste orchidee è principalmente deputata a sostanze del tutto identiche ai feromoni emessi dalle femmine degli insetti (**vedi Parte 2**)

Parte 2: Dal macroscopico al microscopico: l'aspetto molecolare

Maria Rosaria Iesce

La percezione dei colori nelle api è strettamente legata all'impollinazione delle piante con fiori. Gli insetti, api, hanno occhi adatti a percepire alcuni colori, abilità essenziale per la ricerca di nettare e polline. I fiori presentano colori che attirano gli insetti alla ricerca di cibo. In cambio del cibo che offrono, le piante vengono impollinate.

La chimica organica offre una spiegazione molecolare di questi comportamenti. La lezione descrive pertanto alcune classi di molecole organiche coinvolte e la relazione tra la struttura e le proprietà macroscopiche osservate.

Prima di introdurre le sostanze la cui funzione cercheremo di chiarire da un punto di vista chimico, è necessario fare una piccola introduzione ai composti organici e alla loro rappresentazione, soprattutto per coloro che non conoscono o non hanno ancora incontrato questa materia nel loro percorso.

La chimica organica nasce come chimica dei composti naturali, composti che solo la natura era in grado di sintetizzare all'interno di un essere vivente attraverso un "afflato divino" e che, quindi, potevano solo essere estratti da una fonte naturale. L'osservazione che i composti organici elaborati dagli organismi viventi (animali e

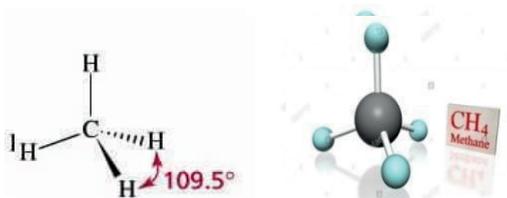
vegetali) contengono prevalentemente carbonio (simbolo C) risale al 1807 ed è da attribuirsi a Jacob Berzelius. Alla fine del 700 con la sintesi, a partire da composti inorganici, dell'urea, un composto naturale prodotto da uomini e animali, la chimica organica comincia ad occuparsi di preparare composti naturali e anche di sintetizzare nuovi prodotti e di studiarne le caratteristiche chimiche e fisiche.

Oggi, si definiscono convenzionalmente composti organici i composti del carbonio con eccezione del monossido di carbonio (CO), dell'anidride carbonica (CO₂), dell'acido carbonico (H₂CO₃) e i suoi sali, sia estratti da fonti naturali (piante, animali, uomo) che ottenuti artificialmente (sintetizzati in laboratorio). Pertanto, la chimica organica è anche detta la chimica dei composti del **carbonio**.

Cosa rende l'elemento carbonio così particolare da dovergli riservare una chimica a parte rispetto a quella di tutti gli altri elementi messi insieme (chimica inorganica)? Guardiamo la sua posizione nella Tavola Periodica.

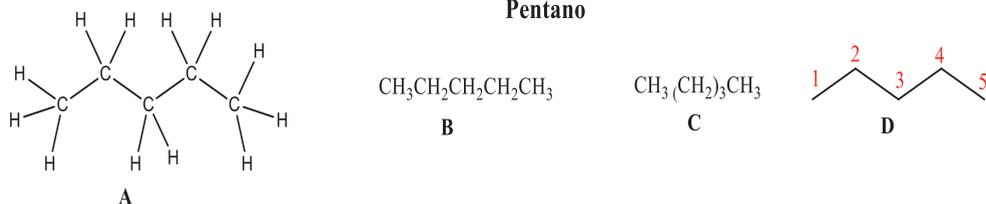
The image shows a standard periodic table of elements. The element Carbon (C) is highlighted with a red circle. An arrow points from above to the circle. The table includes element symbols, atomic numbers, and names in both Italian and English. The Lanthanide and Actinide series are shown at the bottom.

Il carbonio, simbolo C, occupa una posizione centrale nella tavola periodica, è in grado di formare forti legami covalenti con idrogeno, ossigeno, azoto, alogeni, ma anche di formare legami C-C, non solo legami semplici, ma anche doppi o tripli legami, è in grado di dare luogo a catene lineari, ramificate e cicliche. È tetravalente cioè può formare 4 legami; in fig.1 è riportata la più semplice molecola organica, il metano, in cui il C lega con legami covalenti 4 idrogeni (H) proiettati ai vertici di un tetraedro (struttura tetraedrica: piramide a base triangolare). Il metano appartiene alla classe degli idrocarburi (composti organici costituiti solo da C e H).



Il Metano

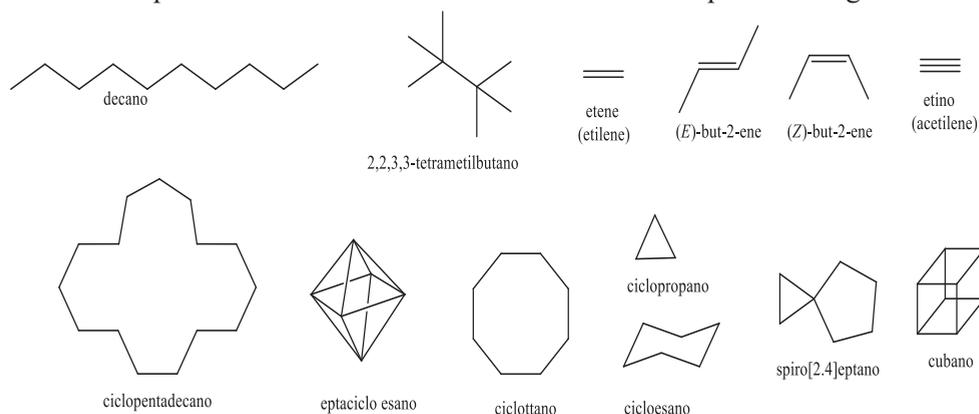
Il numero di molecole organiche naturali e non cresce costantemente grazie alle tecniche di isolamento e riconoscimento e alle procedure di sintesi sempre più all'avanguardia. Le strutture sono generalmente complesse e per lo più con molti atomi di carbonio per cui si utilizzano rappresentazioni semplificate. Di seguito riportiamo diverse rappresentazioni del pentano via via più semplificate.



Nella formula D il carbonio non è indicato e neppure gli idrogeni che si ricavano facilmente tenendo conto della valenza del carbonio: il primo (1) è legato solo ad un C (2) e quindi ha 3 idrogeni, il secondo (2) è legato a due C (1 e 3) e quindi ha due idrogeni, e così via.

La versatilità del carbonio determina una enorme varietà di forme. La nomenclatura, assegnata secondo regole internazionali, è universale ed essenziale per l'individuazione del composto organico in tutto il mondo scientifico.

Alcuni esempi della straordinaria varietà di strutture sono riportati di seguito.



Cominciamo quindi con le domande cui diamo una risposta con la chimica organica.

- a) **Che cosa offrono le piante per convincere gli insetti? Che cosa è il nettare? Che cosa è il miele? E la pappa reale? Che cosa sono gli zuccheri? Che cosa è il saccarosio? Che differenza c'è tra il miele e il saccarosio?**

Il premio che viene dato dalle piante alle api è il nettare. Il **nettare** è una soluzione acquosa zuccherina secreta dalle ghiandole dei vegetali, raggiunge la maggiore concentrazione di zuccheri in particolari condizioni climatiche di umidità che solitamente si presentano all'alba ed al tramonto, e in presenza di alternanza di sole e pioggia con basse temperature. Le api visitano le piante nettariifere soprattutto in questi orari in cui gli zuccheri sono maggiormente sviluppati. Una volta nello stomaco il nettare raccolto dall'ape subirà una prima trasformazione unendosi alla

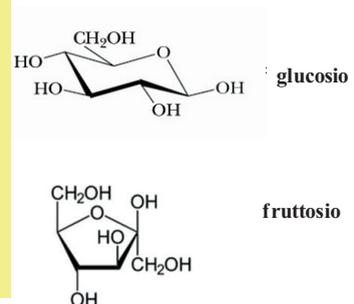
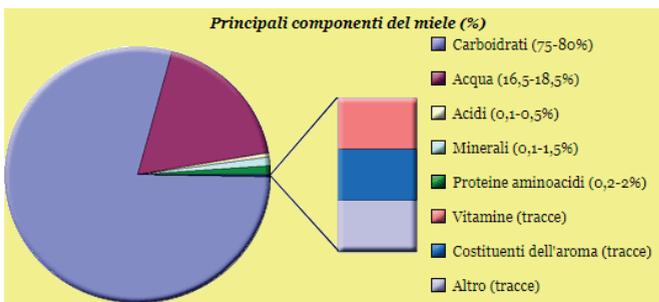


saliva e arricchendosi di acidi organici e fermenti. L'ape bottinatrice, al rientro nell'arnia, affida il suo raccolto ad un'ape interna che lo deposita nel favo di cera dove continua a trasformarsi.

Secondo la "Norma regionale europea raccomandata per il miele":

*il **miele** è il prodotto che le api domestiche producono dal nettare dei fiori o dalle secrezioni provenienti da parti vive di piante o che si trovano sulle stesse, che esse bottinano, trasformano, combinano con sostanze specifiche proprie e lasciano maturare nei favi dell'alveare.*

Come si vede in figura, il miele è costituito principalmente da carboidrati (fino all'80%, di cui **glucosio e fruttosio** rappresentano il 70% del totale), da acqua (fino al 18.5%) e altre sostanze in minore quantità (acidi organici, vitamine, proteine, etc.).

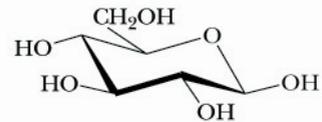
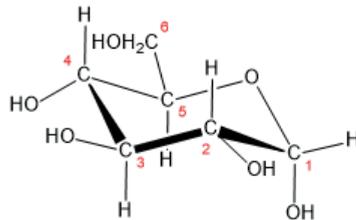


Cosa sono gli zuccheri? Che cosa sono i carboidrati?

Gli zuccheri sono composti chimici detti anche **carboidrati** o idrati di carbonio in quanto i più semplici hanno un rapporto idrogeno ossigeno simile a quello dell'acqua: $C_n(H_2O)_n$. Sono anche detti glicidi o saccaridi.

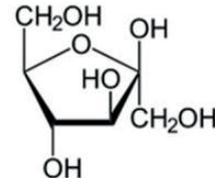
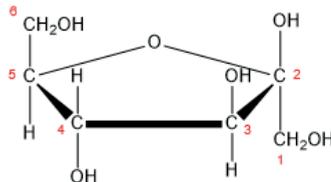
Il glucosio rientra in questo gruppo, ha formula $C_6(H_2O)_6$ ma la formula corretta è $C_6H_{12}O_6$ ed è un monosaccaride. La struttura è ciclica, un anello a 6 termini contenenti 5 atomi di carbonio e 1 atomo di ossigeno; il sesto carbonio è fuori il ciclo ed è legato a 2 idrogeni e ad un gruppo ossidrilico OH (CH_2OH). La prima rappresentazione riporta tutti gli atomi, la seconda non riporta i C del ciclo e gli idrogeni legati a questi C (forma semplificata).

glucosio
 $C_6H_{12}O_6$



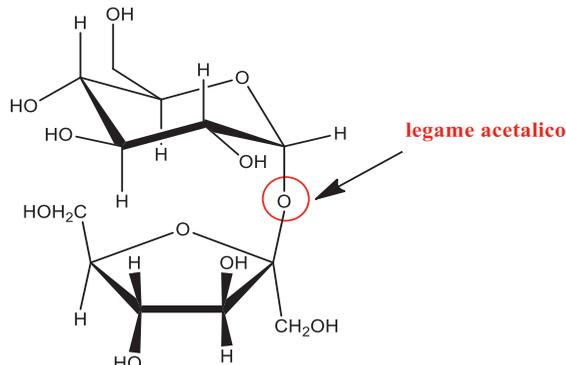
Il fruttosio è anch'esso un monosaccaride ed ha la stessa formula ($C_6H_{12}O_6$). La sua struttura ciclica è a 5 atomi con 4 carboni e 1 ossigeno nel ciclo; gli altri 2 carboni (2 gruppi CH_2OH) sono fuori il ciclo.

fruttosio
 $C_6H_{12}O_6$



Il glucosio è il **monosaccaride** più diffuso in natura, sia libero sia sotto forma di dimeri (disaccaridi) o polimeri (polisaccaridi). Insieme al fruttosio forma il saccarosio (comunemente noto come zucchero). Le due molecole sono legate da un legame acetalico.

saccarosio



Quali sono allora le differenze tra miele e zucchero (saccarosio)?

- Il miele è viscoso, contiene acqua oltre a zuccheri, vitamine, enzimi, acidi organici e sali; un contenuto elevato di glucosio determina una tendenza alla cristallizzazione, mentre una maggiore concentrazione di fruttosio, molto solubile in acqua e igroscopico, conserva il miele allo stato liquido. Il saccarosio come tutti i carboidrati è un solido cristallino a temperatura ambiente, bianco se raffinato, color ambrato se di canna.



- Il miele è costituito da zuccheri semplici. Il saccarosio è costituito da due monosaccaridi legati attraverso un legame difficile da scindere (se non in condizioni acide o ad opera di enzimi); il primo quindi è più rapidamente assimilato. Il glucosio fornisce energia di immediato utilizzo, il fruttosio viene metabolizzato a livello del fegato e costituisce una riserva energetica. Per utilizzare lo zucchero (saccarosio) deve prima verificarsi, nell'organismo, una reazione di scissione.
- Hanno simile potere nutrizionale (304 Kcal per 100 grammi di miele contro le 392 dello zucchero tradizionale e le 362 dello zucchero di canna). La scelta di uno o l'altro è del tutto personale ma in ogni caso entrambi devono essere usati con parsimonia perché è alto il valore glicemico associato. Il miele può contenere spore, che nell'intestino di un bambino di età inferiore all'anno possono germinare e determinare botulismo infantile. Di qui il divieto di dare miele ai neonati. Per gli adulti non è considerato un alimento a rischio perché l'altissima concentrazione di zuccheri impedisce la crescita del *Clostridium botulinum*.

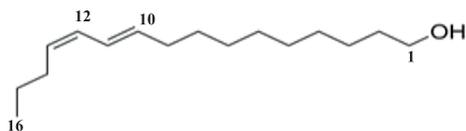
Una sostanza strettamente collegata al miele è la **pappa reale**. A differenza del miele, la pappa reale contiene circa il 70% di acqua e per il resto, oltre a zuccheri, contiene proteine e amminoacidi essenziali e un gran numero di vitamine del gruppo B, oltre ad un fattore antibiotico, per cui è consigliato per le persone debilitate. È prodotta dalle ghiandole ipofaringee delle api giovani per alimentare le api regine e gli individui neonati (per questo è detta anche "latte delle api").

b) Che cosa determina l'attrazione dell'insetto impollinatore nelle orchidee con inganno sessuale?

Le sostanze chimiche responsabili di questo fenomeno sono i **feromoni**. Queste sostanze chimiche sono liberate nell'ambiente esterno da insetti e animali e provocano cambiamenti fisiologici e di comportamento in altri individui, generalmente appartenenti alla stessa specie. Si tratta di una forma di comunicazione chimica silenziosa. L'azione dei feromoni è molto variegata: attrazione sessuale, allarme, guida nei percorsi, delimitazione territoriale. Queste sostanze, secrete in piccolissima quantità da apposite ghiandole, si disperdono nell'aria e vengono captate anche a grande distanza da particolari organi (sensili chemiorecettori) presenti in gran numero sulle antenne degli insetti.

Il primo feromone fu evidenziato nel 1953 ed è il bombykol, rilasciato dalla femmina del baco da seta, che attira i maschi.

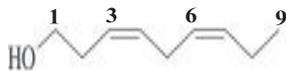
Dal punto di vista chimico si tratta di molecole semplici da 5 a 20 atomi di carbonio, con la presenza generalmente di uno o due atomi di ossigeno, e un peso molecolare tra 80 e 300.



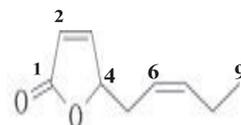
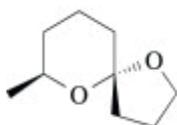
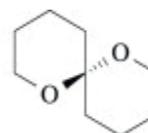
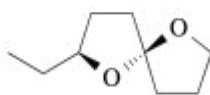
Bombykol

(10*E*,12*Z*)-esadeca-10,12-dien-1-olo

Per lo più volatili, i feromoni hanno strutture variegata: lineari, cicliche o spiro (strutture in cui un solo atomo di carbonio lega due cicli). Alcuni esempi:

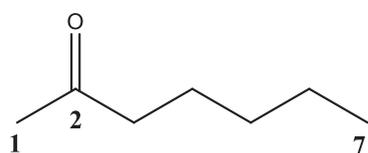
(3*Z*,6*Z*)-3,6-nonadien-1-olo

(E)-2-nonenale

(6*Z*)-2,6-nonadien-4-olide(5*R*,7*S*)-7-metil-1,6-diossaspino[4.5]decano

Nelle api, per esempio, il 2-eptanone, emesso da ghiandole mandibolari, scatena la reazione aggressiva mentre l'acido 9-idrossi-trans-2-decenoico attrae le operaie

verso la regina. I feromoni più diffusi e importanti sono tuttavia quelli sessuali emessi dalle femmine, ma talvolta anche dai maschi, per attrarre all'accoppiamento individui di sesso opposto; sono strettamente specifici pur essendo spesso assai simili tra loro dal punto di vista chimico.



2-eptanone



9-idrossi-trans-2-decenoico

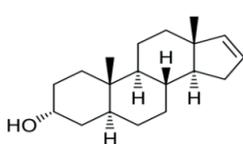
I feromoni trovano ampia applicazione nel controllo delle coltivazioni. Le trappole a feromoni sono efficaci e rispettose delle colture e dell'ambiente: la trappola permette, con l'aggiunta di un attrattivo (generalmente un feromone ad attrazione sessuale), di catturare selettivamente in base al feromone specifico per la cultura interessata, gli insetti maschi, ed impedirne quindi l'accoppiamento e la riproduzione che porterebbe alla deposizione di uova e successiva nascita delle larve causa del danno ai frutti. Le caratteristiche di un feromone sono:

- Possono essere trasmessi al buio
- Sono facili da biosentitizzare (acidi grassi)
- Durano anche parecchi giorni
- Sono trasmessi anche su lunghe distanze (km)

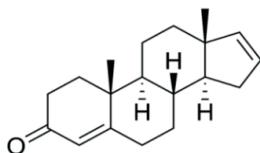
Il principale svantaggio è la lentezza della comunicazione, spesso sono miscele complesse con effetti sinergici. I più specifici sono i feromoni sessuali e su questi si basa maggiormente la chimica di sintesi dei feromoni come alternativa agli insetticidi non naturali

Esistono feromoni umani?

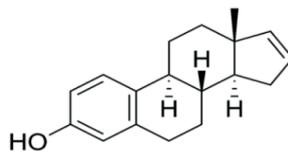
Nel 1974 il Dr George Dodd scoprì il primo feromone umano sessuale, l'androstenolo (A), dall'odore simile al muschio. Altri feromoni sono rappresentati da androstadienone (B), presente nel sudore e nello sperma maschile, e estrotetraenolo (C), presente nell'urina femminile.



A



B



C

Le strutture tetracicliche (*ciclopentanperidrofenantrene*) sono correlabili agli ormoni steroidei, molecole che svolgono molteplici funzioni dallo sviluppo dei caratteri sessuali, al metabolismo degli zuccheri, proteine.

Il primo è presente anche in altri mammiferi, in particolare i suini. Nessuno dei due B e C avrebbe influenza sulla nostra percezione del sesso opposto. Ancora non chiaro il ruolo e l'effetto.

Sebbene abbondantemente pubblicizzati, "l'effetto istantaneo" di feromoni umani, come indicato dai mass media e su alcuni siti web, in pratica non è facile da vedere, perché i feromoni vengono percepiti inconsapevolmente e anche perché nei rapporti umani intervengono altri fattori, come il contatto visivo, il parlare, il toccarsi, etc.

c) Quali sono gli adattamenti delle piante agli insetti impollinatori (i segnali)?

È la percezione dei colori nelle api che è strettamente legata all'impollinazione delle piante con fiori. A determinare il colore dei fiori sono 3 grandi gruppi di pigmenti in essi contenuti: i **carotenoidi** (tinte rosse e gialle), gli **antociani** (tinte rosse e blu) e i **flavonoidi** (gialli). Le domande cui dobbiamo rispondere sono:

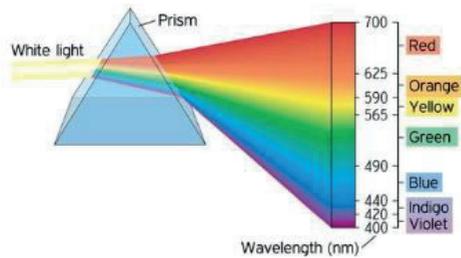
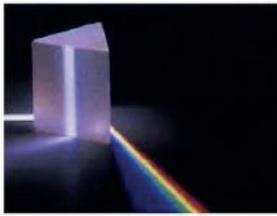
Come si genera il colore? Perché alcuni oggetti presentano colori diversi in ambienti diversi? Perché quel fiore giallo presenta colori diversi se osservato dall'insetto?

Il colore non è una proprietà della materia; richiede la presenza di 3 elementi:

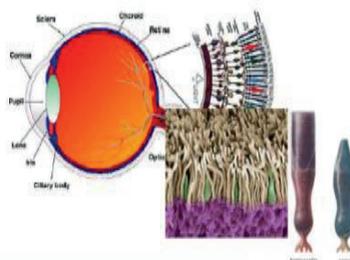
1. Una fonte di luce
2. Un oggetto che interagisce con la luce
3. Un occhio che osserva gli effetti dell'interazione tra la luce e l'oggetto



Isaac Newton fu il primo ad effettuare uno studio sistematico del colore e dimostrò facendo passare la luce solare attraverso un prisma triangolare che la luce solare (luce bianca) è composta da una miscela di tutti i colori dell'arcobaleno. Le radiazioni colorate corrispondono ai sette colori dell'iride: violetto, indaco, blu, verde, giallo, arancio e rosso.



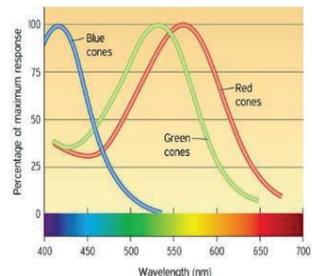
La luce solare (luce bianca) è quindi costituita da diverse radiazioni, caratterizzate da diverse lunghezze d'onda (colori) nell'intervallo 360-780 nm. La gamma delle lunghezze d'onda si chiama spettro e spazia dall'ultravioletto all'infrarosso. Solo alcune radiazioni possono essere rilevate dall'occhio umano e rientrano nella regione spettrale del visibile. Il nostro occhio è infatti in grado di rilevare solo la parte visibile dello spettro elettromagnetico, un intervallo in realtà un po' più stretto (380-760 nm), poiché la sensibilità dell'occhio umano è molto bassa agli estremi dell'intervallo. Particolari cellule presenti nella retina dell'occhio (coni e bastoncelli) rispondono alla radiazione che le colpisce ed inviano, attraverso il nervo ottico, il segnale al cervello che lo elabora e crea una impressione sensoriale, la percezione visiva, in particolare quella del colore. Ci sono tre diversi tipi di cellule (coni) nella retina che variano la loro sensibilità a diversi colori della luce: blu, verde e rosso, e questo porta alla visione tricromatica dei colori. Quando la luce entrando nell'occhio colpisce i coni sensibili alla luce rossa, questi mandano un impulso al cervello che riconosce la luce come rossa, lo stesso principio è applicabile agli altri due tipi di coni mentre per i fasci luminosi intermedi si crea un impulso nervoso combinato da coni diversi che il cervello riconosce come una miscela di colori; se ad esempio sono stimolati i coni preposti al rosso e al verde, il cervello interpreterà questo colore come giallo



I coni sono responsabili della visione diurna (detta fotopica)

- coni S (sensibili al blu 437 nm)
- coni M (sensibili al verde 533 nm)
- coni L (sensibili al rosso 564 nm)

I bastoncelli sono utili nella visione in presenza di poca luce (detta scotopica)



I coni sono responsabili della visione diurna. I bastoncelli sono utili nella visione in presenza di poca luce, in particolare sono preposti alla visione notturna e permettono una visione a scala di grigi.

Il colore dipende quindi dalle radiazioni che colpiscono il nostro occhio, quindi dall'interazione della luce con la materia che costituisce l'oggetto:

- Un oggetto appare bianco se non assorbe nessuna radiazione
- Un oggetto appare nero se assorbe tutte le radiazioni
- Un oggetto appare colorato se assorbe solo alcune radiazioni. Ad esempio un oggetto di colore blu assorbe luce rossa arancio e gialla e quindi riflette solo la luce blu (e i colori ai lati).



Le radiazioni riflesse (non assorbite dal materiale) determinano il colore dell'oggetto che è percepito dai nostri occhi (colore complementare).

Una conseguenza importante: un oggetto può riflettere solo la luce delle frequenze presenti nella luce che lo illumina. Quindi, l'aspetto di un oggetto colorato dipende anche dal tipo di luce utilizzata per illuminarlo. Infatti, i colori alla luce del giorno appaiono diversi da come appaiono quando sono illuminati da lampade artificiali.

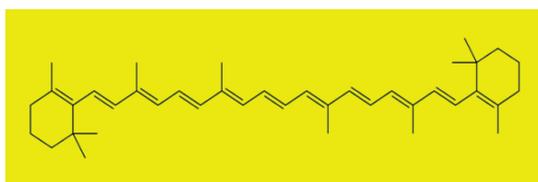
In definitiva, il colore nasce dalla luce ed è la percezione che deriva dalla composizione della luce (visibile) che colpisce la retina, un'impressione sensoriale; pertanto, l'interpretazione del colore può anche variare da individuo a individuo e, perfino, nella stessa persona che può percepire diversamente il colore in momenti diversi ed in base allo stato d'animo.



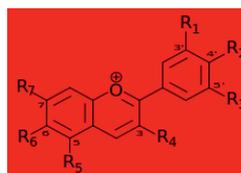
La molecola con sistemi coniugati è in grado di assorbire alcune radiazioni (e questo è misurabile con un apparecchio, spettrofotometro UV-visibile) riflettendone altre. In particolare, la dipendenza del colore dalla radiazione assorbita è riportata in Tabella. La comparsa del colore dipende dal numero di doppi legami coniugati (deve essere almeno > 7).

Lunghezza d'onda assorbita	Colore osservato
380-460	giallo
380-500	arancio
440-560	rosso
480-610	porpora
540-650	blu

Nei fiori e nelle piante le molecole organiche più diffuse che determinano colori sono i carotenoidi e i flavonoidi, in particolare gli antociani.



Carotenoidi



Antocianidine

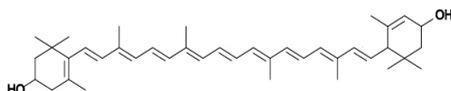
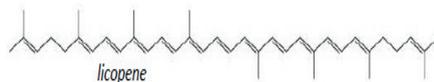
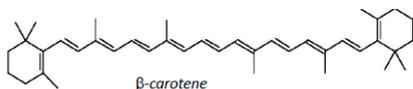
Come si vede si tratta di molecole che contengono un gran numero di doppi legami coniugati. I carotenoidi hanno una struttura molecolare tale da assorbire il blu ed il verde per riflettere il rosso e l'arancione, mentre le antocianidine assorbono la luce blu-verde e riflettono sfumature rosse fino al blu.

La sintesi delle "molecole colorate" ha luogo nei cromoplasti localizzati nei fiori e nei frutti delle piante: i flavonoidi anche nei vacuoli. Queste molecole sono presenti anche nelle foglie ma i loro colori sono coperte dal colore verde della clorofilla. In autunno, quando la clorofilla si degrada, i loro colori possono finalmente caratterizzare le foglie che assumono i bei colori dal rosso all'arancione, al giallo, tipici di questa stagione.

Carotenoidi

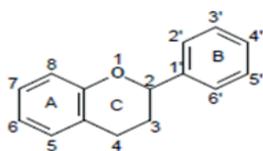
Si tratta di molecole costituite da una lunga catena di atomi di carbonio (40), talvolta terminante con anelli, e presentano un numero elevato di doppi legami coniugati. In funzione del numero di doppi legami coniugati presentano colori variabili dal giallo

pallido all'arancione fino al rosso. Il b-carotene presenta un sistema con 11 doppi legami coniugati, assorbe la radiazione a 455 nm ed è responsabile del colore tipico delle carote. Il licopene assorbe radiazione a 474 nm ed è responsabile del colore rosso del pomodoro. Le xantofille hanno la struttura chimica fondamentale del carotene contenente anche ossigeno.



Flavonoidi e antociani

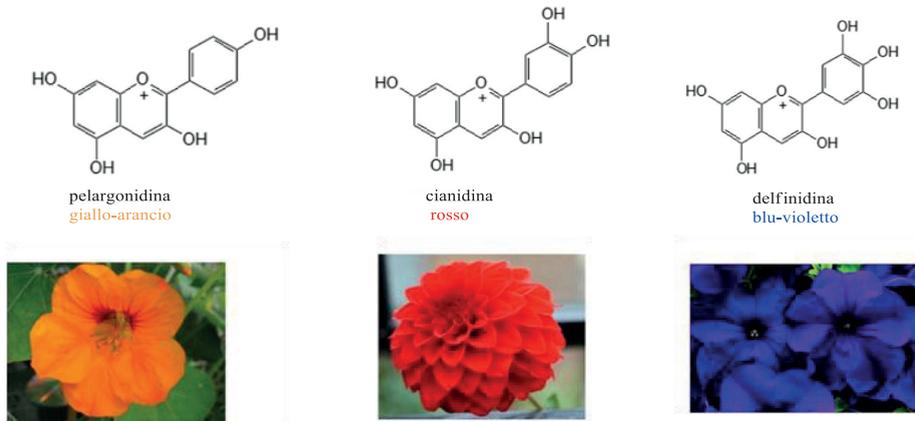
Gli antociani appartengono alla famiglia dei flavonoidi il cui nucleo base **contiene** 15 atomi di carbonio (C6-C3-C6) arrangiati in 3 anelli, indicati con le lettere A, B e C dove il ponte con 3 atomi di carbonio è ciclizzato con l'ossigeno. In particolare, l'anello benzenico (A) è condensato con un pirano (C), un ciclo a 6 termini contenente ossigeno, questo è legato a sua volta ad un altro anello benzenico. Come si vede in Figura 1, gli anelli A e B contengono ciascuno un sistema coniugato (un anello aromatico come quello del benzene) ma il nucleo base, nucleo flavanico, non è colorato. Il sistema presente negli antociani è riportato è il catione flavilio. Qui la coniugazione è estesa per cui le antocianidine sono colorate.



Gli antociani (o antocianine) si differenziano dalle antocianidine per la presenza di un gruppo glicosidico (generalmente in posizione R7 o R5 o R3). In natura esistono circa una ventina di antocianidine mentre il numero dei derivati antociani è fino a 15-20 volte maggiore. La presenza di sostituenti OH e il pH determinano tutta una variazione di colori.

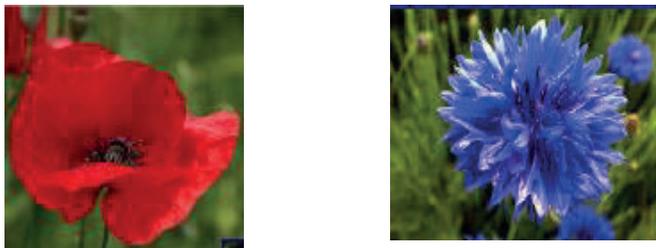
a) Dipendenza dalla presenza di gruppi OH

La figura di seguito riporta come varia il colore se aumenta il numero di sostituenti OH sull'anello aromatico B.



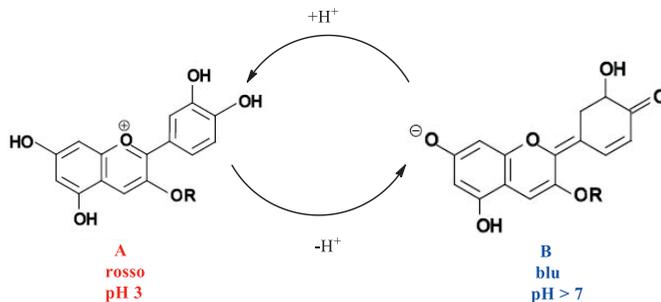
b) Dipendenza dal pH

Il papavero e il fiordaliso hanno colori completamente differenti pur contenendo la stessa molecola (cianidina).

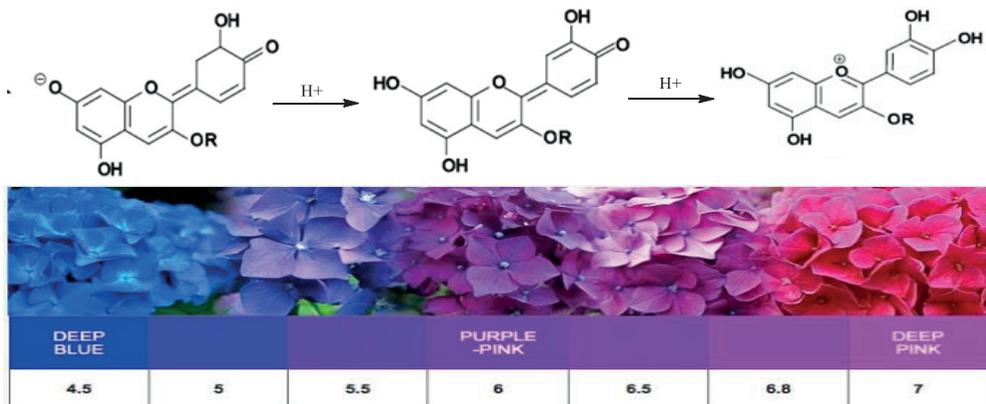


La spiegazione è legata al fatto che la cianidina presenta forme diverse in dipendenza del pH e quindi colori diversi. Ricordiamo che il pH è una grandezza fisica che indica l'acidità di una soluzione, di un liquido, etc. Un pH uguale a 7 è neutro, minore di 7 è acido e maggiore di 7 è basico. La cianidina ha all'interno della sua struttura gruppi fenolici (OH legati ad un anello aromatico) che sono gruppi acidi e facilmente ionizzabili in ambiente basico. A pH minore di 3 (ambiente acido) la cianidina ha la forma A; in ambiente basico cede protoni e presenta la forma B in cui la coniugazione si estende e questo determina un assorbimento diverso e quindi il colore blu. Nel papavero la linfa è acida, la cianidina è nella forma A e quindi il colore

è rosso; nel fiordaliso la linfa è alcalina (basica), la cianidina è nella forma deprotonata B e, quindi, il colore è blu.



Anche all'interno della stessa pianta il colore può modificarsi in dipendenza del pH, ad esempio del pH del terreno. Prendiamo l'ortensia. Questa pianta generalmente si presenta con fiori di colore blu-azzurro ma questo si verifica quando è coltivata in un terreno neutro o alcalino, inizia a produrre fiori di colore rosa quando è bagnata con acqua calcarea (acida). La molecola responsabile è ancora la cianidina che presenta forme diverse (e quindi colori diversi) in dipendenza del pH.



Per mantenere il colore blu si aggiungono delle sostanze chiamate “**azzurranti**”, che contengono **alluminio**, **ferro** o altri **microelementi** che, somministrati alla pianta, la portano a emettere fiori blu.

Considerazioni finali

Quanto illustrato rappresenta solo un piccolissimo esempio di molecole organiche e del loro ruolo. Le proprietà degli oggetti con i quali abbiamo a che fare nella nostra vita quotidiana, o meglio le caratteristiche dei materiali dei quali sono costituiti e che stanno alla base della loro funzione, derivano dalle proprietà **delle molecole** che

compongono questi stessi materiali. E' da tener presente comunque che le proprietà macroscopiche sono correlabili non solo alla natura delle molecole che li compongono (ovvero alla loro specifica identità chimica) ma anche al modo con cui queste molecole interagiscono fra loro. Spesso, sono le interazioni intermolecolari a spiegare queste grandezze macroscopiche. Di conseguenza, una stessa sostanza chimica (identiche proprietà molecolari) in un differente stato (differenti interazioni molecolari) può avere proprietà macroscopiche molto diverse.

È il concetto di molecola che ci consente di conoscere la complessità qualitativa (oltre che quantitativa) del mondo macroscopico, trasportandola in parte nel mondo microscopico. Di qui la comparsa di discipline quali biologia molecolare, medicina molecolare, elettronica molecolare e l'esigenza sempre più marcata di studi interdisciplinari che vedono competenze diverse colloquiare insieme.

La biologia molecolare studia gli esseri viventi a livello dei meccanismi molecolari alla base della loro fisiologia, concentrandosi in particolare sulle interazioni tra le macromolecole, ovvero proteine e acidi nucleici. La medicina molecolare è un ampio campo in cui vengono utilizzate tecniche fisiche, chimiche, biologiche, bioinformatiche e mediche per descrivere strutture e meccanismi molecolari, identificare errori molecolari e genetici fondamentali della malattia e sviluppare interventi molecolari per correggerli; la **medicina molecolare** è la soluzione per individuare trattamenti a misura di persona in base al suo quadro clinico, di cui l'aspetto molecolare è di grande rilevanza. La cucina molecolare affronta le trasformazioni chimiche e fisiche che avvengono negli alimenti durante la loro preparazione. L'elettronica molecolare si occupa dello studio e dell'applicazione dei *mattoni* molecolari per la fabbricazione di componenti elettronici e rappresenta una ricerca interdisciplinare che si estende alla fisica, alla chimica e alla scienza dei materiali.

Conclusioni

La stragrande maggioranza delle piante a fiore dipende direttamente dall'azione degli insetti per la propria riproduzione. L'impollinazione entomofila è pertanto considerata un servizio ecosistemico fondamentale che gli insetti svolgono per il mantenimento della comunità vegetale e di conseguenza dell'intero ecosistema. Una grande quantità di piante coltivate tra cui, solo per citarne alcune, broccoli, pomodori, mandorle, ciliegie, legumi, carote etc. adotta lo stesso meccanismo ed è stato stimato che tra i 235 ed i 577 miliardi di dollari di produzione alimentare dipende direttamente dagli impollinatori (FAO, 2016). **Perché la comunità scientifica internazionale ha preoccupazioni rispetto al futuro del processo di impollinazione entomofila?** Negli ultimi decenni in varie aree geografiche è stato

notato un preoccupante declino degli insetti impollinatori (Gallai *et al.*, 2009; Thomann *et al.*, 2013). Le cause di questo declino sono ancora poco chiare e probabilmente includono la combinazione di un utilizzo eccessivo di pesticidi, la frammentazione degli habitat, l'inquinamento dell'aria, l'inquinamento luminoso, i cambiamenti climatici. La comunità scientifica internazionale ha individuato in questo declino una delle principali minacce per la biodiversità e, dunque, una priorità dal punto di vista della ricerca scientifica. In Italia, solo recentemente, il ministero dell'ambiente ha stilato linee guida che individuano il declino degli impollinatori spontanei come la principale minaccia alla biodiversità ed ha invitato i Parchi Nazionali a mettere in campo azioni mirate al suo contrasto.

Bibliografia

- Bao, T., Wang, B., Li, J., & Dilcher, D. (2019). Pollination of Cretaceous flowers. *Proceedings of the National Academy of Sciences*, 116(49), 24707–24711.
- Bruice, P. Y. (2012) “Chimica Organica” Edises, Napoli.
- Camerarius, R.J. (1694). Epistola ad M.B. Valentini de sexuplantarum. In *Ostwalds Klassiker der exakten Naturwissenschaften*, No. 105, 1899 (Leipzig: Verlag von Wilhelm Engelmann).
- Dafni, A., & Ivri, Y. (1981). Floral mimicry between *Orchis israelitica* Baumann and *Dafni* (Orchidaceae) and *Bellevalia flexuosa* Boiss. (Liliaceae). *Oecologia*, 49(2), 229-232.
- Darwin, C. (1877). The various contrivances by which orchids are fertilised by insects. John Murray.
- FAO (2016) - News Article: Pollinators vital to our food supply under threat
- Fenster, C. B., Armbruster, W. S., Wilson, P., Dudash, M. R., & Thomson, J. D. (2004). Pollination syndromes and floral specialization. *Annu. Rev. Ecol. Evol. Syst.*, 35, 375-403.
- Gallai, N., Salles, J. M., Settele, J., & Vaissière, B. E. (2009). Economic valuation of the vulnerability of world agriculture confronted with pollinator decline. *Ecological economics*, 68(3), 810-821.
- Garay, L. A. (1960). On the origin of the Orchidaceae. *Botanical Museum Leaflets, Harvard University*, 19(3), 57-96.

- Germani R. (Università degli Studi di Perugia) dalle lezioni (“Il colore delle molecole”) nell’ambito delle attività PLS-2017
- <https://www.ukessays.com/essays/biology/exploring-the-chemistry-of-pheromones-biology-essay.php>
- Kölreuter, J. G. (1761). Vorläufige Nachricht von einigen das Geschlecht der Pflanzen Betreffenden Versuchen und Beobachtungen. Gleditsch.
- Linnaeus, C. (1758). Systema naturae (Vol. 1, No. part 1, p. 532). Laurentii Salvii: Stockholm.
- Sprengel, C. K. (1793). Das entdeckte Geheimnis der Natur. Wilhelm Engelmann, Leipzig.
- Thomann, M., Imbert, E., Devaux, C., & Cheptou, P. O. (2013). Flowering plants under global pollinator decline. Trends in plant science, 18(7), 353-359.
- Villani, G “Usiamo le molecole per spiegare l'uomo”
- Zimmermann, Y., Roubik, D. W., & Eltz, T. (2006). Species-specific attraction to pheromonal analogues in orchid bees. Behavioral Ecology and Sociobiology, 60(6), 833.

Le immagini contenute nell’articolo sono attinte da Wikipedia o da fonti riportate.

Seminario integrato. Orbitali e configurazione elettronica

Parte 1: La fisica dell'atomo di idrogeno¹

Umberto Scotti di Uccio

Dipartimento di Fisica "E. Pancini", Università di Napoli Federico II

Abstract

La meccanica quantistica spiega le proprietà della materia che non sono accessibili alla fisica classica. In questo intervento ci concentriamo sull'interpretazione quantistica della fisica atomica e, in particolare, dell'atomo più semplice, l'Idrogeno, attraverso un percorso che parte dal Principio di indeterminazione di Heisenberg e si conclude con la spettroscopia di emissione atomica.

Introduzione

Con la locuzione "Fisica Classica" si intende, di solito, l'insieme della Meccanica di Newton, della Termodinamica di Joule e Clausius e dell'Elettromagnetismo di Maxwell. Sgombriamo però il campo da un equivoco: non ha molto senso parlare di "Fisica Classica" in senso storico, dicendo per esempio "è la Fisica nota alla fine del XIX secolo". Anche se questo è un modo molto comune di raccontare le cose, resta pur sempre un modo sbagliato. In primo luogo, così si fa un torto alla Storia; non è affatto vero che la fine del XIX Secolo segni un confine di demarcazione tanto netto. Anzi, proprio nel secolo successivo sono arrivati molti contributi importanti che hanno sistemato la Fisica Classica sia sul piano formale che nelle basi fenomenologiche. In secondo luogo, si fa un torto alla Fisica; perché non fa parte di questa disciplina il ragionamento fondato su base storica. Per esempio, non ha senso dire a uno studente di Fisica o a un ricercatore: "Affronta questo problema, ma sta attento: potrai usare solo i concetti noti fino al 1900". Solo pochi storici della Fisica saprebbero rispondere correttamente a domande così particolari; meglio lasciare a loro questo tipo di approccio, e proporre agli studenti un diverso modo di ragionare. Il modo corretto di individuare la "Fisica Classica" è, dunque, un altro. La Storia non c'entra niente; la Fisica Classica è, semplicemente, un modello fisico. In generale, chiamiamo "modello fisico" un insieme di leggi, espresse in forma matematica,

¹Questo contributo riassume i due seminari tenuti dal prof. Scotti di Uccio in questo contesto e nel seminario integrato "Fluorescenza: dal fotone ai materiali e alle proteine reporter"

adatte a descrivere le relazioni tra le grandezze fisiche che osserviamo in certi fenomeni. Un modello fisico ha, quindi, due elementi: da un lato un insieme di fenomeni, che chiamiamo “ambito di validità”, dall’altro un insieme di leggi per descriverli. Delle leggi abbiamo già parlato; soffermiamoci dunque a parlare dei fenomeni, cioè a circoscrivere l’ambito.

L’ambito della Fisica Classica è un po’ sfumato, ma con un certo grado di approssimazione si può dire che questo modello si occupa con successo dei fenomeni macroscopici. Sinteticamente, la ragione è questa: le definizioni operative delle grandezze della Fisica Classica funzionano bene quando gli oggetti dello studio contengono molti atomi.

Per dare un’idea di questo limite, facciamo un esempio: la definizione di velocità media. Consideriamo la posizione x_1 di un punto materiale all’istante iniziale t_1 e la sua posizione finale x_2 all’istante t_2 ; poi calcoliamo il rapporto $v_m = \frac{x_2 - x_1}{t_2 - t_1}$.

Sembra tutto semplice e anche molto generale, ma c’è un trucco. La “posizione del punto materiale” può essere definita con una certa precisione solo in un caso particolare: il corpo deve poter essere descritto come un “punto materiale”. In realtà, ciò non è possibile se le dimensioni del corpo diventano molto piccole, per esempio se il corpo in questione è una molecola, o un atomo, o addirittura un elettrone, senza rivedere in modo molto profondo la definizione operativa di “posizione”, perché certamente non possiamo usare un righello, né qualunque altro strumento macroscopico, per effettuare la misura. D’altronde, quando proviamo a usare tecniche più sofisticate andiamo incontro a tantissimi problemi nuovi; per esempio, a un certo punto l’errore di misura non può essere ridotto in nessun modo ricorrendo a strumenti più sofisticati, ma resta caratteristico dell’oggetto che misuriamo e delle condizioni fisiche in cui lo osserviamo.

La caratteristica generale degli oggetti microscopici è questa: non possiamo trascurare la loro struttura granulare, perché contengono un numero relativamente piccolo di atomi. Un altro modo per ripetere il concetto espresso precedentemente è quindi questo: la Fisica Classica incontra serie difficoltà ogni volta che deve confrontarsi con la struttura granulare delle cose. E siccome la materia ha struttura granulare, ne consegue che la Fisica Classica non ne può spiegare le proprietà fondamentali. Questa osservazione ci riporta a una questione fondamentale: ogni teoria che voglia superare la meccanica classica nella descrizione delle proprietà della materia deve confrontarsi con la fisica degli atomi.

Ma non solo la materia, anche la radiazione ha una struttura granulare. La luce può essere descritta come un insieme di “quanti”, i fotoni, ciascuno dei quali porta una piccola quantità di energia. La storia si ripete: la fisica classica non è in grado di

descrivere le proprietà “granulari” della radiazione. Anche questa osservazione ha un’implicazione importante: ogni teoria che voglia superare la meccanica classica nella descrizione delle proprietà della radiazione deve confrontarsi con la fisica dei fotoni.

La meccanica quantistica si propone dunque questo progetto ambizioso: descrivere la fisica degli atomi e dei fotoni e il modo in cui interagiscono. Inoltre, si propone di ricondurre ogni problema di fisica della materia e della radiazione alle proprietà microscopiche, rispondendo ad esempio a domande del tipo “Perché l’Idrogeno reagisce con l’Ossigeno e forma una molecola di acqua a forma di V?”; “Perché l’acqua a bassa temperatura è solida, a temperatura ambiente è liquida e ad alta temperatura è gassosa?”; “Perché l’acqua è un dielettrico?”; e così via.

Notiamo subito che alcune di queste domande “sembrano” domande di chimica. Ma anche in questo caso conviene fare chiarezza: quando si parla delle proprietà degli atomi e del modo in cui essi interagiscono, la Fisica e la Chimica diventano un’unica scienza, che potremmo chiamare Chimica Fisica, essenzialmente fondata sulla meccanica quantistica.

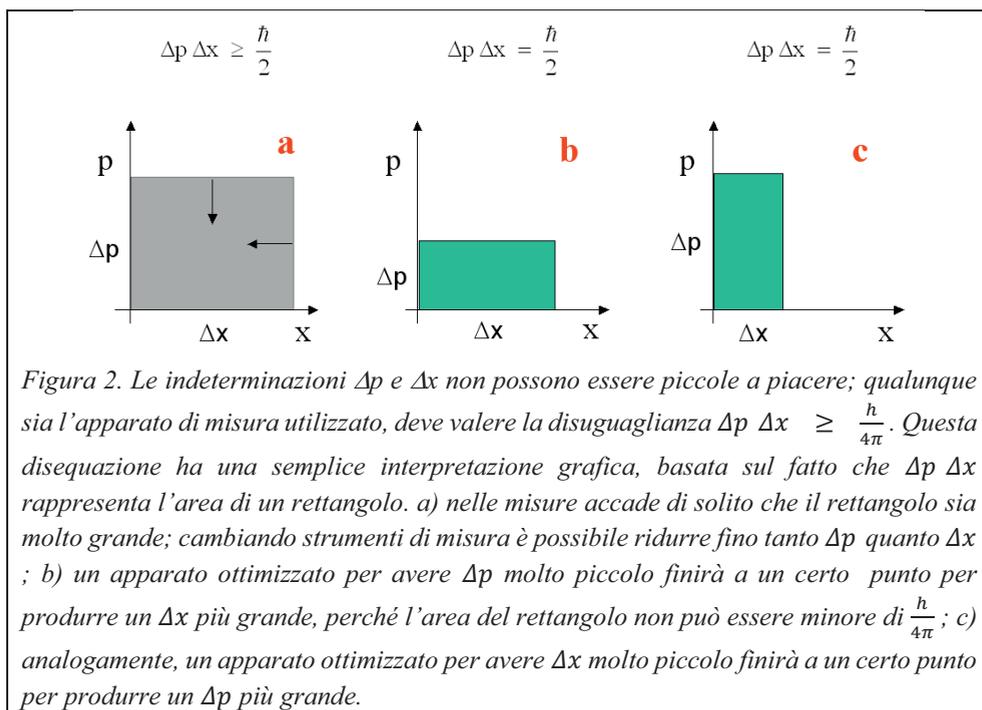
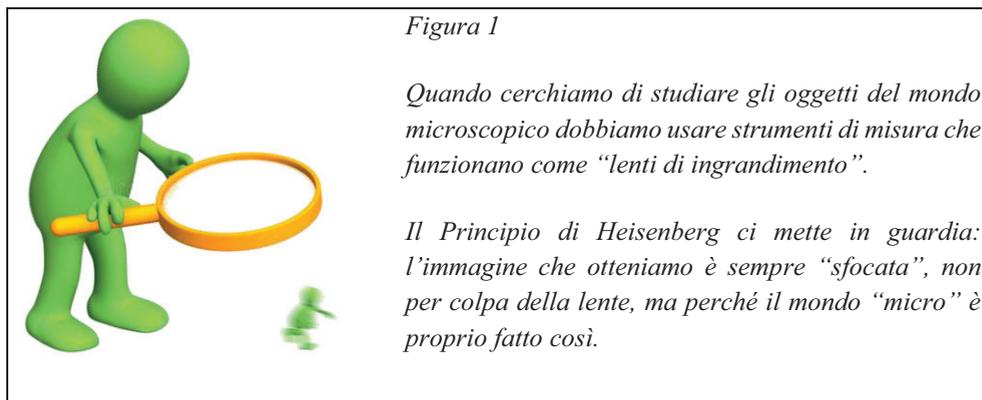
Nel seguito ci avventuriamo dunque alla scoperta delle proprietà elementari del più semplice atomo, l’Idrogeno, con l’occhio della Fisica e le conoscenze elementari che ci vengono dallo studio della Chimica. Sarà, dunque, un bel lavoro di sintesi.

Il Principio di Indeterminazione di Heisenberg

Ricordiamo che “principio” in fisica significa “affermazione basata sugli esperimenti”. Ad esempio, la Meccanica Classica si fonda sui tre principi di Newton (legge di inerzia, seconda legge di Newton, azione e reazione); la Termodinamica si fonda su tre principi (conservazione dell’energia, Il principio, irraggiungibilità dello zero assoluto); etc. Anche la Meccanica Quantistica è fondata su un principio, che prende il nome dal fisico che per primo lo propose. Lo presenteremo senza la pretesa di essere particolarmente accurati, ma con l’obiettivo di mettere in luce i concetti fondamentali.

Il Principio di indeterminazione di Heisenberg afferma che è sostanzialmente impossibile trovare insieme posizione e velocità di una particella; se miglioriamo la conoscenza dell’una, perdiamo informazione sull’altra. A livello microscopico, i concetti di traiettoria e di legge oraria perdono senso: tutto appare “sfocato” (Fig. 1). Inizialmente, Heisenberg propose questa spiegazione: se facciamo una misura, l’apparato inevitabilmente perturba la particella e ogni volta che ripetiamo la misura la perturbazione sarà diversa. Per esempio, se misuro la quantità di moto p di una particella, inevitabilmente ne cambierò il valore di Δp , e se contemporaneamente ne

misuro la posizione x modificherò anche quest'ultima di Δx . Per il principio di Heisenberg, i valori di Δp e Δx non possono essere piccoli a piacere, qualunque sforzo facciamo per migliorare l'apparato di misura (Fig. 2). Nel caso dei sistemi "micro", può facilmente accadere che il disturbo Δp sia molto significativo (cioè, che Δp sia dello stesso ordine di grandezza di p) e lo stesso vale per Δx . Il risultato è l'effetto di sfocatura.



L'interpretazione di Heisenberg è corretta, ma oggi siamo ancora più drastici nel commentare la sua stessa legge. Infatti, affermiamo che il problema non riguarda solo ciò che accade durante le misure; ci sono sempre indeterminazioni anche quando non ci sono disturbi, perché la Natura è regolata da un meccanismo intrinsecamente probabilistico. In altri termini, né con gli esperimenti, né con la teoria possiamo prevedere esattamente i fenomeni; invece, tutto ciò che possiamo misurare o calcolare è la probabilità che un fenomeno si manifesti.

Ad esempio, non potremo mai conoscere la legge oraria di un elettrone in un atomo. Il problema non è solo legato ai limiti del procedimento di misura (prima interpretazione di Heisenberg); in modo ancora più radicale possiamo affermare che non possiamo conoscere la legge oraria perché la legge oraria non esiste! Infatti, non esiste la funzione $x = x(t)$ perché l'elettrone non ha "una certa posizione x al tempo t "; ha solo "una certa probabilità di trovarsi nella posizione x al tempo t ".

Le distribuzioni di probabilità in meccanica quantistica

Come visto nel paragrafo precedente, la meccanica quantistica ha come obiettivo il calcolo della probabilità che si verifichi un certo evento. Ecco alcuni esempi:

1. Con che probabilità l'elettrone dell'atomo di Idrogeno si trova a distanza maggiore di 1 mm dal nucleo?
2. Con che probabilità lo stesso elettrone ha una velocità minore di 10^5 ms^{-1} ?
3. Con che probabilità un atomo di Idrogeno emetterà spontaneamente entro un secondo un fotone di lunghezza d'onda $\lambda = 110 \text{ nm}$?

Di fronte a questo progetto, dobbiamo innanzitutto rispondere a questa domanda: la probabilità è una grandezza fisica? Si tratta di una questione fondamentale: sappiamo infatti che la Fisica si occupa esclusivamente delle grandezze fisiche e delle relazioni che intercorrono tra queste grandezze nei fenomeni naturali.

Ovviamente, la risposta è affermativa: possiamo definire la probabilità in modo operativo. La definizione utile alla Fisica è di tipo sperimentale, la cosiddetta definizione "frequentista":

La probabilità p è il rapporto tra il numero di casi favorevoli e il numero totale di casi possibili

$$p = \frac{n}{N}$$

Per esempio, per rispondere sperimentalmente alla terza domanda, si potrebbero mettere in un recipiente N atomi di Idrogeno e poi contare il numero n di fotoni di lunghezza d'onda λ emessi in un secondo. Il numero di casi favorevoli coincide col numero di fotoni emessi; il numero di casi possibili col numero totale di atomi che

erano ipoteticamente in condizione di emettere un fotone. Quindi la probabilità cercata è data dal rapporto tra n e N .

Le prime due domande sembrano molto più complicate e, in effetti, l'approccio sperimentale è molto difficile. Tuttavia, la teoria della meccanica quantistica ci permette di rispondere utilizzando un importante strumento matematico, le distribuzioni di probabilità, che introduciamo con un esempio.

Numero di caramelle	Numero di bambini che hanno preso x caramelle
x	n
0	0
1	3
2	5
3	6
4	2
5	1

Figura 3.

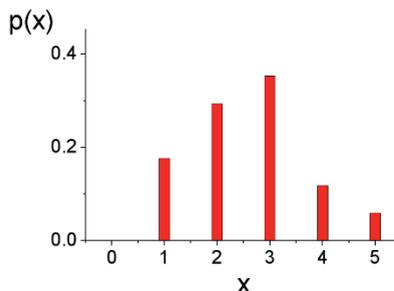
a) Per misurare le probabilità, facciamo un'indagine: contiamo quanti bambini hanno preso 0 caramelle, quanti 1 caramella, ecc. Supponiamo ad esempio di aver trovato i risultati riportati nella Tabella a fianco.

Notiamo che x indica il "numero di caramelle" e n indica il "numero di bambini che hanno preso x caramelle". Notiamo inoltre dai dati che in classe ci sono $N = 17$ bimbi: N è il "numero totale".

x	n	p
0	0	0.00
1	3	0.18
2	5	0.29
3	6	0.35
4	2	0.12
5	1	0.06

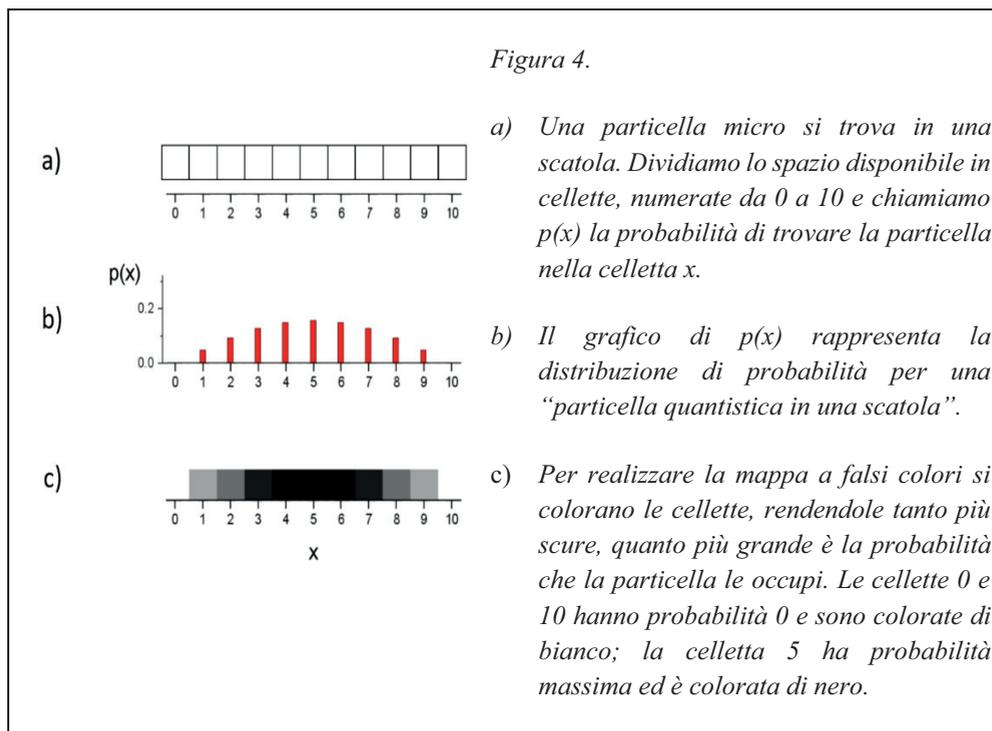
b) Possiamo ora aggiungere una colonna in cui a ogni valore di x facciamo corrispondere la probabilità $p = \frac{n}{N}$.

Troviamo così la corrispondenza $x \rightarrow p(x)$, che è la distribuzione di probabilità.



c) usualmente, $p(x)$ è rappresentata con un grafico a barre. Attenzione: non si tratta di un istogramma: l'altezza delle barre semplicemente indica il valore della funzione $p(x)$.

Immaginiamo di essere in una classe di asilo. Abbiamo lasciato un mucchietto di caramelle su un banco e i bimbi possono prendere e mettere in tasca quante caramelle desiderano. Noi non conosciamo il numero di caramelle prese da ciascun bambino; diremo che questo numero è una “variabile aleatoria”. Possiamo però chiederci: con che probabilità un bambino ha preso 3 caramelle? con che probabilità ne ha prese 4? e così via: ad ogni possibile valore della variabile aleatoria, associamo la sua probabilità. Questa associazione è una “distribuzione di probabilità”. In Figura 3 è mostrato come misurare e come rappresentare questa distribuzione di probabilità. Facciamo ora un secondo esempio. Immaginiamo che una particella micro sia contenuta in una regione di spazio stretta e lunga: si tratta del famoso problema della “particella nella scatola”. Possiamo eseguire una serie di misure consecutive per determinare la posizione della particella. A questo scopo, dividiamo la scatola in cellette numerate (Fig. 4a) e contiamo quante volte la particella ci appare in ciascuna celletta. Da queste misure possiamo ricavare la distribuzione di probabilità $p(x)$ che la particella si trovi nella posizione x (Fig. 4b). Un modo conveniente per rappresentare la distribuzione di probabilità $p(x)$ è utilizzare una mappa a falsi colori (Fig. 4c). Il principale motivo per cui usiamo queste mappe è però che funzionano molto bene anche in due dimensioni (Fig. 5).



Abbiamo precedentemente chiarito che le leggi della Meccanica Quantistica non possono prevedere i fenomeni, ma la probabilità che i fenomeni si manifestino. Ora possiamo essere più precisi:

La teoria della Meccanica Quantistica si propone di prevedere tutte le distribuzioni di probabilità che caratterizzano un sistema fisico

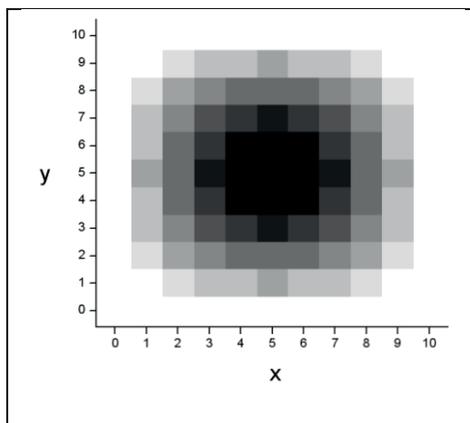


Figura 5.

In questo caso, la particella può occupare una delle cellette di una scatola quadrata. Ancora una volta l'interpretazione è semplice: la probabilità di trovare la particella in una celletta è tanto più alta, quanto più la celletta è scura. La mappa rappresenta la distribuzione di probabilità per una "particella quantistica in una scatola quadrata".

Ad esempio, con le leggi della Meccanica Quantistica è possibile calcolare la distribuzione $p(x)$ riportata in Figura 4c, oppure la distribuzione $p(x, y)$ in Figura 5. Questi calcoli furono effettuati per la prima volta da Schrödinger, utilizzando la famosa equazione che da lui prende il nome.

Riassumendo, nella seguente Tabella mettiamo a confronto cosa possiamo dire in fisica classica e ciò che possiamo dire in meccanica quantistica del moto di una particella.

Meccanica classica	Meccanica quantistica
La legge oraria ci dice dove si trova la particella in ogni istante	Non possiamo sapere dove si trova la particella in un dato istante, ma solo che probabilità c'è di trovarla in un certo punto
Dalla legge oraria ricaviamo la legge di velocità, sicché possiamo dire che quantità di moto ha la particella in ogni istante	Non possiamo sapere che quantità di moto ha la particella in un dato istante, ma solo che probabilità c'è che abbia una certa quantità di moto

Dalla legge oraria ricaviamo la traiettoria	La traiettoria non esiste
---	---------------------------

Gli orbitali atomici

Un orbitale è la distribuzione di probabilità dell'elettrone di un atomo: ci dice con che probabilità l'elettrone può essere trovato in un certo punto.

Fino a non molti anni fa, si credeva impossibile accedere sperimentalmente al mondo degli atomi per “vedere” la forma degli orbitali. Oggi abbiamo invece a disposizione varie tecniche di misura che permettono di determinare le distribuzioni di probabilità degli elettroni di un atomo. Nella Figura 6 è ridisegnato il risultato di un esperimento recente (Stodolna, 2013).

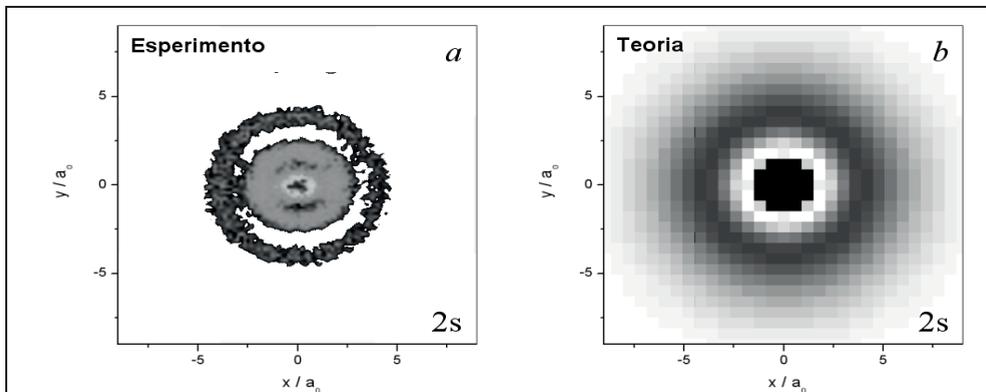


Figura 6. Le due mappe rappresentano l'orbitale 2s dell'atomo di Idrogeno:

- a) Risultato sperimentale (ridisegnato a partire dai dati sperimentali di cui in Figura 7.2);
 b) Calcolo teorico, basato sull'equazione di Schrödinger.

Le mappe devono essere interpretate in questo modo: le regioni più scure sono quelle in cui è più probabile trovare l'elettrone, quelle chiare sono quelle in cui è meno probabile trovare l'elettrone. Queste mappe sono tutto ciò che possiamo conoscere: non c'è modo di sapere “dove” si trova l'elettrone in un certo istante, ma solo con che probabilità può essere trovato in una certa posizione. Il fattore di scala $a_0 = 50 \text{ pm} = 50 \times 10^{-12} \text{ m}$ è il “raggio di Bohr”, che qui usiamo come “unità di misura”. Il raggio di Bohr è la lunghezza caratteristica del mondo degli atomi; dal più piccolo al più grande della Tavola Periodica, gli atomi hanno diametri che differiscono per meno di un ordine di grandezza dal raggio di Bohr.

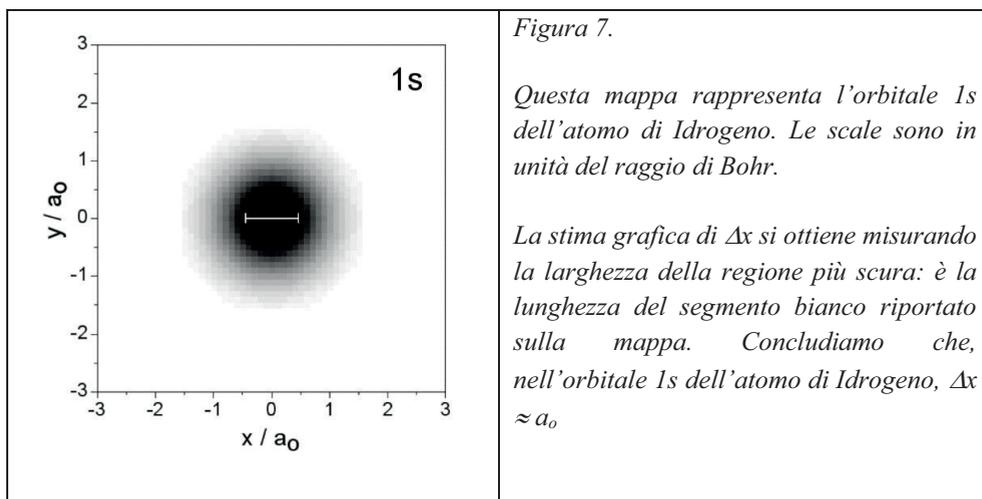
Come si giustifica in meccanica quantistica che gli elettroni non caschino prima o poi sul nucleo? In fondo, sono cariche negative e risentono di una forte attrazione

colombiana da parte del nucleo positivo. La meccanica quantistica ci dà una risposta molto semplice:

La stabilità degli atomi è garantita dal principio di indeterminazione di Heisenberg

Infatti, se un elettrone finisse sul nucleo, sapremmo esattamente dove trovarlo, in qualunque momento, e per di più sapremmo che è sempre esattamente fermo, quindi con quantità di moto $p = mv = 0$. Questa possibilità è però esclusa dal Principio di Heisenberg, perché l'incertezza sulla posizione e l'incertezza sulla quantità di moto non possono essere piccole a piacere.

L'indeterminazione Δx che compare nel Principio di Heisenberg può essere stimata graficamente dalle mappe di probabilità. Se non abbiamo la pretesa di essere troppo accurati, è davvero molto facile: basta determinare l'estensione della regione più scura della mappa. Il metodo è illustrato in Figura 7.

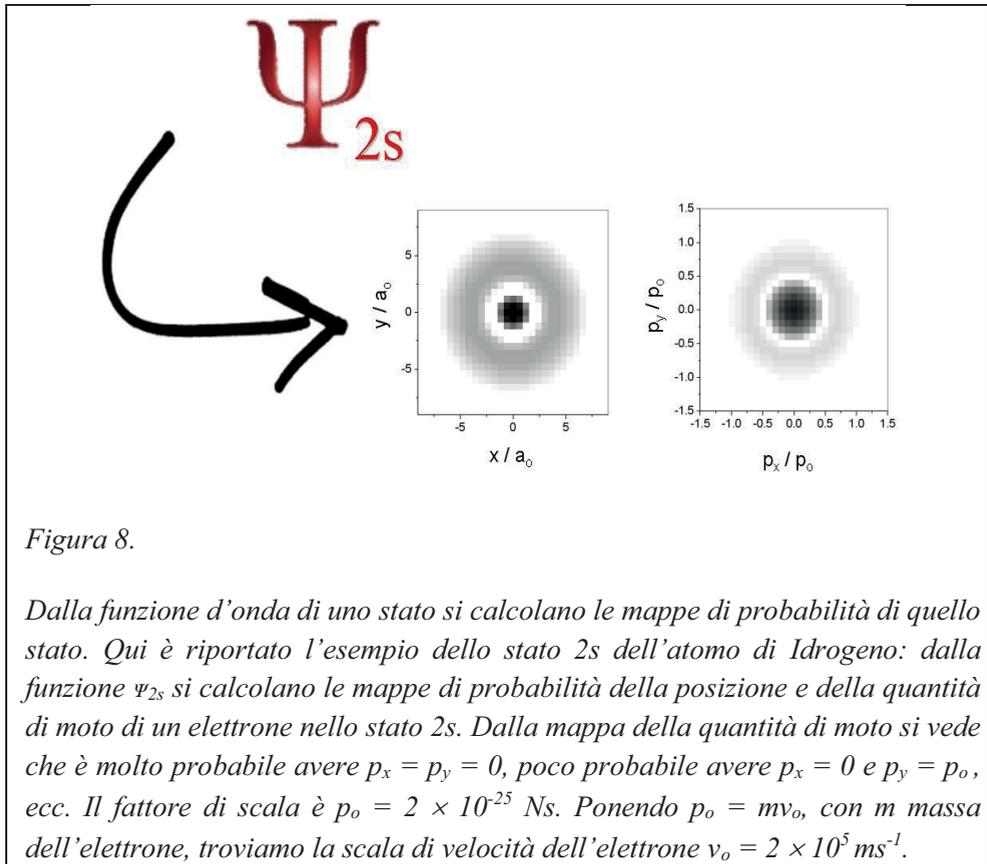


La funzione d'onda

Abbiamo affermato che la teoria della meccanica quantistica permette di calcolare tutte le mappe di probabilità; per esempio, quella della posizione, quella della quantità di moto, ecc. (Fig. 8). Si pone allora questa domanda: le mappe delle diverse grandezze fisiche sono indipendenti tra di loro?

La risposta è no: ce lo dice il Principio di Indeterminazione! Infatti, dalla mappa della posizione si può ricavare Δx , da quella della quantità di moto si può ricavare Δp , e queste due grandezze sono legate tra loro dalla diseuguaglianza di Heisenberg.

Dal punto di vista della teoria, il collegamento formale tra le due mappe è dato dalla funzione d'onda ψ . Senza entrare per ora nel dettaglio, possiamo interpretare ψ come un ente matematico che permette di calcolare le mappe di probabilità di posizione, quantità di moto e di ogni altra grandezza fisica che caratterizza il sistema.



L'atomo di Idrogeno

L'atomo di Idrogeno è uno dei sistemi quantistici più semplici: contiene un solo elettrone legato a un protone. È del tutto naturale partire da qui per approfondire lo studio della struttura elettronica degli atomi.

Quando nel paragrafo precedente abbiamo introdotto la nozione di orbitale, abbiamo già parlato di Idrogeno: per esempio, nelle Figure 6 e 7 abbiamo visto le mappe di due diversi orbitali (identificati come 1s e 2s). Sorgono spontaneamente alcune domande:

a) Quanti orbitali esistono?

- b) Come vengono indicati?
 c) Quali proprietà fisiche li contraddistinguono?

Rispondiamo ora ordinatamente a questi punti.

- a) Chiariamo innanzitutto che l'elettrone dell'atomo di Idrogeno può occupare uno tra diversi possibili orbitali, corrispondenti a stati diversi tra loro.
- b) Per “contarli”, si introducono in meccanica quantistica alcuni numeri, detti numeri quantici, seguendo le regole mostrate in Figura 9:
- il numero quantico principale n è un intero positivo: $n = 1, 2, \dots$. Come vedremo in dettaglio nel punto c), il numero quantico principale è associato all'energia meccanica dell'elettrone;
 - il numero quantico secondario ℓ è un intero compreso tra 0 e $(n-1)$;
 - il numero quantico secondario ℓ_z è un intero relativo, compreso tra $-\ell$ e ℓ ;
 - il numero di spin s_z può invece assumere solo i valori $\pm 1/2$; molto spesso è però rappresentato da una freccia rivolta verso l'alto (“spin up” \uparrow) se $s_z = 1/2$, o verso il basso (“spin down” \downarrow) se $s_z = -1/2$

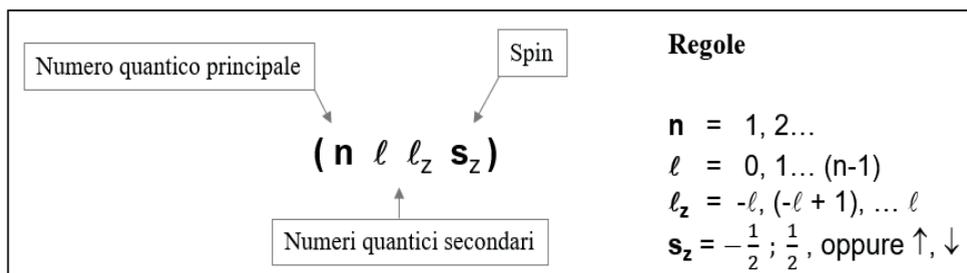


Figura 9. Regole per i numeri quantici.

Non è troppo importante soffermarsi sul significato fisico dei numeri quantici secondari e del numero di spin; in prima istanza, li possiamo usare semplicemente come etichette utili a identificare gli stati. Messi tutti insieme, formano la “targa” dello stato.

La nomenclatura degli orbitali è un po' complicata dal fatto che in chimica si usano nomi diversi per indicare le stesse cose. In particolare, si usano le lettere s, p, d, f... con particolari pedici per indicare rispettivamente $l = 0, 1, 2, 3, \dots$ e l'insieme $\{\ell_z\}$, come indicato nella Tabella 2. Nella pratica possiamo scegliere la notazione che preferiamo, perché sono del tutto equivalenti.

					Molteplicità
{1s}	n = 1	l = 0	$l_z = 0$	$s_z = \{-1/2, 1/2\}$	$1 \times 2 = 2$
{2s}	n = 2	l = 0	$l_z = 0$	$s_z = \{-1/2, 1/2\}$	$1 \times 2 = 2$
{2p_x 2p_y 2p_z}	n = 2	l = 1	$l_z = \{-1, 0, 1\}$	$s_z = \{-1/2, 1/2\}$	$3 \times 2 = 6$
{3s}	n = 3	l = 0	$l_z = 0$	$s_z = \{-1/2, 1/2\}$	$1 \times 2 = 2$
{3p_x 3p_y 3p_z}	n = 3	l = 1	$l_z = \{-1, 0, 1\}$	$s_z = \{-1/2, 1/2\}$	$3 \times 2 = 6$
{3d_{z²} 3d_{xy} 3d_{yz} 3d_{zx} 3d_{x²-y²}}	n = 3	l = 3	$l_z = \{-2, -1, 0, 1, 2\}$	$s_z = \{-1/2, 1/2\}$	$5 \times 2 = 10$

Tabella 2. Nomenclatura dei primi orbitali dell'elettrone nell'atomo di Idrogeno. Le parentesi graffe indicano un insieme di valori possibili. Nella notazione chimica lo spin non è indicato esplicitamente, sicché, ad esempio, {1s} indica sia $1s \uparrow$ che $1s \downarrow$, ecc.

c) Gli orbitali si distinguono tra di loro in base a diverse proprietà fisiche: l'energia meccanica dell'elettrone, la forma e le dimensioni degli orbitali, ecc. Qui ci concentreremo soprattutto sull'energia.

Con buona approssimazione, l'energia meccanica E dell'elettrone dipende solo dal numero quantico principale. Si trova sperimentalmente:

$$E_n = -\frac{E_0}{n^2} \quad ; \quad E_0 = 13.6 \text{ eV}$$

Allo scopo di interpretare graficamente questa relazione, in Figura 10 è mostrato il grafico dell'energia potenziale elettrostatica $U(r)$ dell'atomo di Idrogeno, con la rappresentazione dei valori di energia meccanica E_n .

Lo stato di minima energia meccanica (quello con $n = 1$) è detto stato fondamentale; quelli di energia maggiore sono detti stati eccitati. Dunque, l'elettrone non può assumere qualunque valore di energia meccanica:

I livelli permessi sono discreti e non sono equispaziati, ma si infittiscono al crescere di n

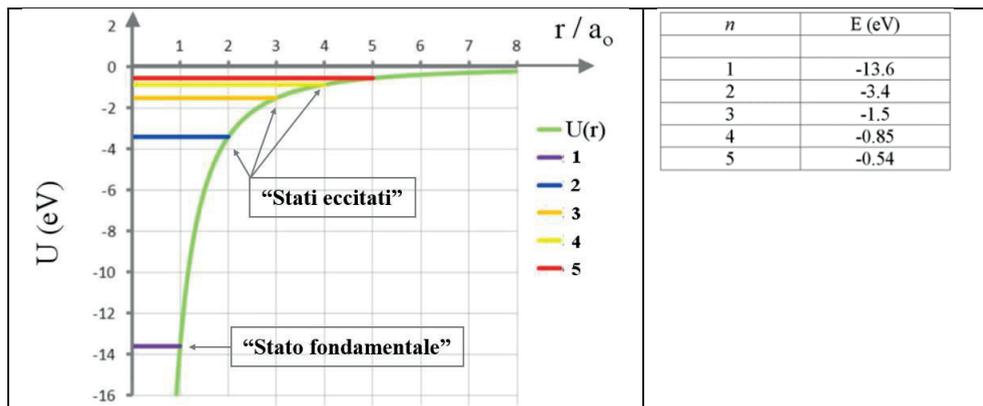


Figura 10. Grafico dell'energia potenziale elettrostatica dell'atomo di Idrogeno e rappresentazione dei valori permessi di energia meccanica. Si noti la scala verticale in elettronvolt: $1 \text{ eV} = 1.6 \times 10^{-19} \text{ J}$.

In meccanica classica, questo tipo di grafici permette di individuare “la regione permessa al moto” in base a questo ragionamento: visto che $E - U = K \geq 0$, deve essere sempre $E \geq U$; quindi, ad esempi, col valore di energia $E = E_1$ (stato fondamentale), la regione permessa sarebbe ristretta a $r \leq a_0$. Attenzione: in meccanica quantistica questo ragionamento è falso e non esiste nessuna regione proibita al moto. Piuttosto, ciò che accade è che la probabilità di trovare l'elettrone diventa più piccola man mano che ci si allontana dal nucleo.

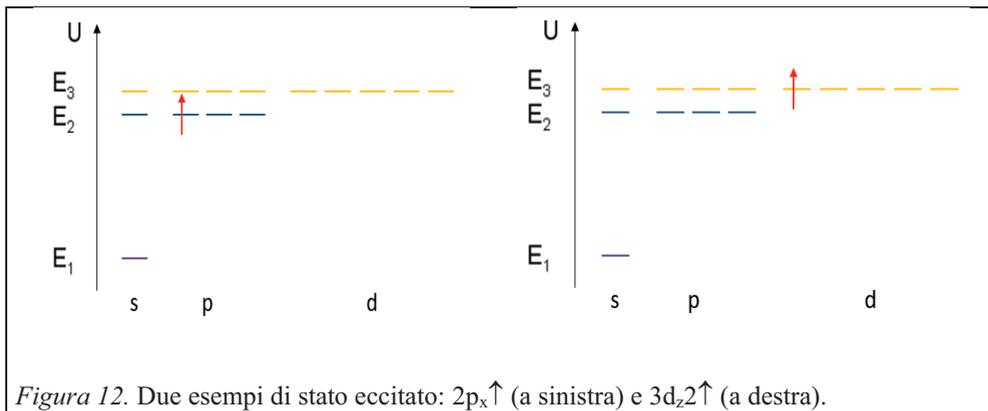
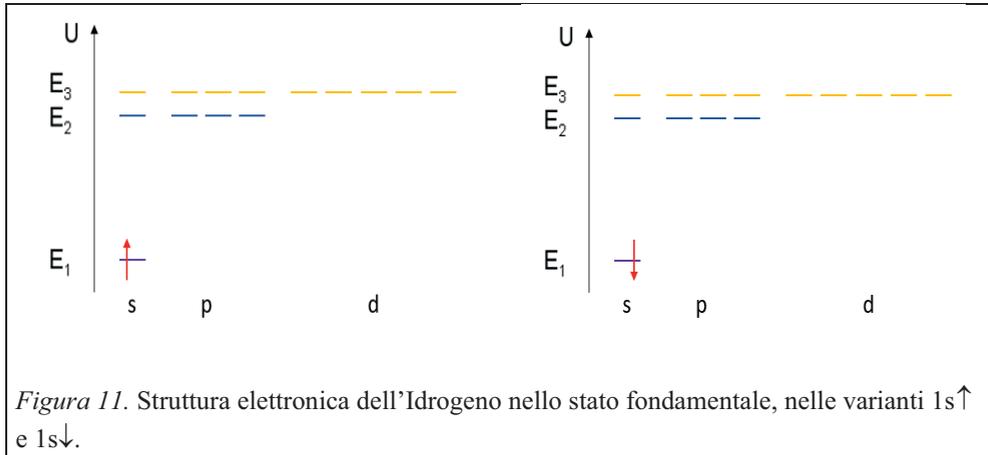
Per indicare l'orbitale effettivamente occupato dall'elettrone e la corrispondente energia meccanica si può rappresentare graficamente la configurazione elettronica dell'atomo. In Figura 11 sono mostrati gli unici due stati di minima energia; nella notazione chimica li indichiamo come $1s\uparrow$ e $1s\downarrow$. Se l'elettrone ha la minima energia possibile, si dice che occupa lo stato fondamentale: perciò nell'atomo di Idrogeno lo stato fondamentale è lo stato $1s$. In Figura 12 sono invece mostrati due possibili esempi di stati eccitati, cioè stati di energia maggiore: $2p_x\uparrow$ e $3d_z2\uparrow$.

Eccitazione e diseccitazione dell'atomo di Idrogeno

Resta ancora una domanda cui rispondere: gli stati sono stabili nel tempo? La risposta è sì e no.

Ciascuno stato è *relativamente stabile* nel tempo e per questo motivo si parla di *stati stazionari*. L'aggettivo “stazionario” indica che le mappe di probabilità associate agli stati non cambiano aspetto, cioè non si modificano gradualmente nel tempo. Tuttavia, “stazionario” non significa “statico” perché l'elettrone non è fermo.

L'avverbio "relativamente" indica che però esistono meccanismi discontinui che interrompono la quiete e permettono all'elettrone di passare da uno stato all'altro. Descriveremo due particolari processi di questo tipo: l'eccitazione e la diseccitazione dell'atomo.



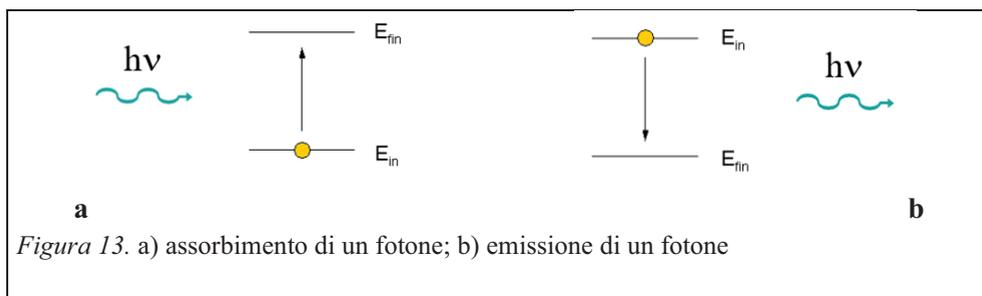
Nello stato di minima energia, l'elettrone dell'atomo di Idrogeno occupa l'orbitale $1s$. È però possibile eccitare l'atomo cedendogli energia; se ciò accade, l'elettrone può passare a un livello permesso più alto. Esistono due meccanismi principali di eccitazione: per urto o per assorbimento di un fotone:

- Un atomo può essere urtato da un altro atomo, da uno ione, da un elettrone; può essere scagliato e urtare contro una parete; ecc. Nelle condizioni giuste, dopo l'urto un atomo che era nello stato fondamentale può finire in uno stato eccitato.
- Il processo di assorbimento di un fotone è mostrato schematicamente in Figura 13a. Per la conservazione dell'energia meccanica, l'energia guadagnata

dall'elettrone passando dallo stato iniziale a quello finale è pari all'energia del fotone:

$$E_{\text{fin}} - E_{\text{in}} = h\nu$$

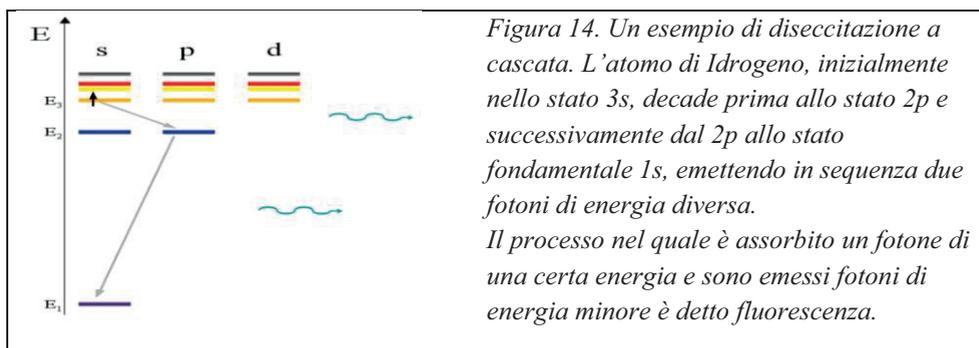
Se consideriamo i livelli di energia riportati in Figura 10, vediamo ad esempio che per eccitare l'Idrogeno dallo stato fondamentale ($n = 1$) a un primo eccitato ($n = 2$) occorre un fotone di energia $h\nu = 10.2 \text{ eV}$. Si tratta di un fotone UV.



Gli stessi meccanismi possono funzionare al contrario: un atomo può diseccitarsi cedendo energia in un urto, oppure emettendo un fotone. Il processo di emissione di un fotone è mostrato in Figura 13b. Per la conservazione dell'energia meccanica, l'energia del fotone è pari all'energia persa dall'elettrone nel passaggio dallo stato iniziale a quello finale:

$$h\nu = E_{\text{in}} - E_{\text{fin}}$$

Ad esempio, se un elettrone si diseccita passando da uno stato con $n = 3$ a uno stato con $n = 2$, viene emesso un fotone di 1.9 eV . Si tratta di un fotone rosso-arancio. La diseccitazione può anche seguire un percorso a cascata: per esempio, si può avere prima il passaggio dallo stato iniziale con $n = 3$ allo stato intermedio con $n = 2$ con l'emissione di un primo fotone, e poi da $n = 2$ a $n = 1$ con l'emissione di un secondo fotone (Fig. 14).



La spettroscopia permette di determinare la lunghezza d'onda (e, quindi, l'energia) dei fotoni emessi nella diseccitazione. Lo schema di principio è illustrato in Figura 15a.

In una misura di spettroscopia di emissione dell'Idrogeno, la sorgente è una lampada a gas (concettualmente simile alle comuni lampade al neon). La radiazione emessa è poi separata nelle varie componenti cromatiche. Newton, che inventò questa tecnica, usava a questo scopo un prisma di vetro (Fig. 8.8a); oggi si ricorre invece ai “reticoli di diffrazione”. Le righe emesse dall'Idrogeno nel visibile sono rappresentate in Figura 15b.

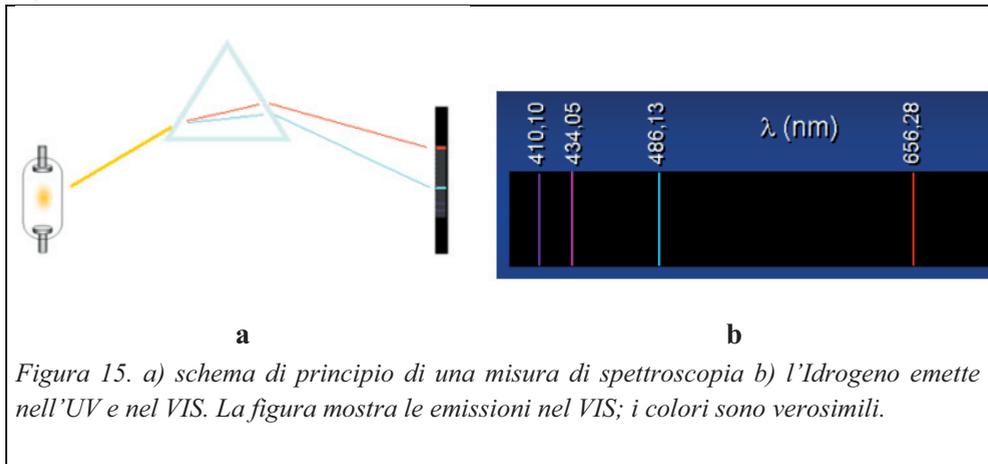


Figura 15. a) schema di principio di una misura di spettroscopia b) l'Idrogeno emette nell'UV e nel VIS. La figura mostra le emissioni nel VIS; i colori sono verosimili.

Bibliografia

- Stodolna, A. S., Rouzee, A., Lépine, F., Cohen, S., Robicheaux, F., Gijssbertsen, A., Jungmann, J. H., C. Bordas. C., Vrakking, M. J. J. (2013). Hydrogen Atoms under Magnification: Direct Observation of the Nodal Structure of Stark States, Phys. Rev. Lett. 110, 213001
- <https://physics.aps.org/featured-article-pdf/10.1103/PhysRevLett.110.213001>.

Seminario integrato. Orbitali atomici e configurazione elettronica

parte 2

Oreste Tarallo

Dipartimento di Scienze Chimiche, Università di Napoli Federico II

Introduzione

Il modello atomico attuale descrive l'**atomo** come costituito da un nucleo centrale e da elettroni che occupano la regione di spazio a questo circostante.

Nel **nucleo**, formato da neutroni (particelle elettricamente neutre), e protoni (particelle dotate di carica elettrica pari a 1.6022×10^{-19} C), si concentra quasi tutta la massa dell'atomo. Il nucleo è molto piccolo, avendo un diametro dell'ordine dei centesimi di picometro.

Gli **elettroni** sono invece “distribuiti” intorno al nucleo, in un volume proporzionalmente molto maggiore (essi si muovono infatti nello spazio che circonda il nucleo occupando un volume che è circa 10^{15} volte maggiore di quello del nucleo stesso). Possiedono una massa molto piccola (9.1094×10^{-31} kg, pari a circa 1/1836 della massa del protone) e una carica negativa pari a -1.6022×10^{-19} C (che corrisponde ad un'unità di carica elettrica).

Tra le diverse particelle che costituiscono l'atomo, gli elettroni, trovandosi nella zona più esterna, sono quelli che maggiormente ne influenzano la **reattività**, ovvero il comportamento nei confronti degli atomi circostanti nella formazione di legami.

Il loro comportamento, come quello di tutte le particelle subatomiche, è però piuttosto difficile da descrivere in quanto regolato da leggi che *sfuggono al nostro intuito*, che si basa invece su esperienze fatte con oggetti di dimensioni molto maggiori. L'elettrone, infatti, a seconda del tipo di esperimento che si conduce, si comporterà a volte come una minuscola particella di materia, altre volte come un'onda. In altre parole, l'elettrone (così come altre particelle quali i fotoni) ha una duplice natura: esso è sia un corpuscolo che un'onda. Come se non bastasse, non è neppure possibile dire che, in un certo istante, l'elettrone *stia* in un qualche posto con la stessa certezza con cui diciamo, per esempio, che una pallina da tennis *si trova* in una borsa. Tutto quello che possiamo affermare è che esiste una certa **probabilità** di trovare l'elettrone in una certa regione dello spazio che si trova attorno al nucleo, non in un punto preciso.

Orbitali atomici

Le caratteristiche ondulatorie e corpuscolari degli elettroni cui prima accennavamo (il cosiddetto **dualismo onda-particella**) e il moto degli elettroni nell'atomo possono essere descritti mediante un'equazione che prende il nome di **equazione d'onda**. La più nota equazione d'onda è l'equazione di Schrödinger,

$$H\Psi = E\Psi,$$

- dove Ψ (“psi”) viene detta **funzione d'onda** e descrive i possibili stati dell'elettrone
- E rappresenta l'**energia** collegata a ciascuna funzione d'onda
- H è un operatore matematico che definisce le condizioni meccaniche del movimento (energia cinetica e potenziale), detto hamiltoniano.

Le soluzioni dell'equazione definiscono le funzioni d'onda, o **orbitali**, che sono funzioni delle coordinate dell'elettrone, $\Psi(x, y, z)$. Gli orbitali consentono di descrivere il comportamento dell'elettrone come un'onda stazionaria, rappresentandone la distribuzione nello spazio intorno al nucleo.

Si può dimostrare che queste soluzioni esistono solo per alcuni valori ben definiti della costante E , ovvero solo per alcune ben precise energie degli elettroni negli stati definiti dalle funzioni $\Psi(x, y, z)$. Di conseguenza, l'energia dell'elettrone risulta **quantizzata**, cioè può assumere solo certi valori multipli di una quantità minima.

Risolvendo l'equazione d'onda si trova che, per specificare ciascuna funzione d'onda (quindi per descrivere ciascun orbitale), sono necessari tre **numeri quantici**: n , ℓ , e m_ℓ . Tali numeri sono correlati fra loro, nel senso che possono assumere solo certe combinazioni di valori (Tabella 1).

Numero quantico	Simbolo	Valori possibili
Principale	n	$1, 2, 3, \dots, \infty$
Secondario (o azimutale, o del momento angolare)	ℓ	$0 \leq \ell \leq n - 1$
Magnetico	m_ℓ	$-\ell \leq m \leq +\ell$

Tabella 1

Il **numero quantico principale**, n , è il numero quantico più importante in quanto il suo valore è un fattore fondamentale nel determinare l'energia di un elettrone in un atomo. Più elevato è il valore di n , maggiore sarà il livello energetico e quindi l'energia dell'elettrone e maggiore sarà la distanza media dell'elettrone dal nucleo: un elettrone che si trova in un livello con $n=2$ è più distante dal nucleo rispetto a un

elettrone con $n=1$, e così via. Il numero quantico principale può assumere tutti i valori interi da 1 fino a infinito. Tutti gli orbitali con lo stesso valore di n costituiscono un **livello** (o guscio, in inglese “*shell*”) dell’atomo.

Nel caso dell’idrogeno, ma non negli atomi polielettronici (ovvero con più di un elettrone), tutti gli orbitali con lo stesso valore di n (quindi tutti gli orbitali di un dato livello) possiedono la stessa energia (si dice perciò che sono **degeneri**).

Il **numero quantico secondario**, ℓ , può assumere invece n valori interi, che vanno da 0 a $n-1$. Esso fornisce importanti informazioni circa la forma dell’orbitale e definisce il **sottolivello** (o sottoguscio) a cui appartiene un orbitale.

Un sottolivello di un determinato livello è costituito da tutti gli orbitali con il medesimo valore di ℓ : ad esempio, il livello con $n=1$ contiene un solo sottolivello (quello con $\ell=0$), il livello con $n=2$ contiene due sottolivelli ($\ell=0$ e $\ell=1$), il livello con $n=3$ contiene tre sottolivelli ($\ell=0$, $\ell=1$ e $\ell=2$) e così via.

Comunemente, invece di indicare i sottolivelli mediante numeri si usano lettere, secondo la seguente corrispondenza:

Valore di ℓ	
0	s
1	p
2	d
3	f

Possiamo quindi dire, ad esempio, che il livello con $n=3$ comprenderà tre sottolivelli: 3s, 3p e 3d (rispettivamente caratterizzati da $\ell=0$, $\ell=1$ e $\ell=2$).

Infine, ciascun sottolivello è formato da $2\ell + 1$ orbitali, identificabili in base al **numero quantico magnetico**, m_ℓ . Per un dato valore del numero quantico secondario ℓ , il numero quantico m_ℓ può infatti assumere i valori interi da $+\ell$ a $-\ell$, zero incluso, per un totale di $2\ell+1$ valori. Per esempio, nel sottolivello p ($\ell=1$) di un qualsiasi livello n , possono esistere tre orbitali, definiti dai numeri quantici magnetici $+1$, 0 e -1 . Questi orbitali vengono comunemente indicati come p_x , p_y e p_z in quanto il numero quantico magnetico fornisce informazioni circa l’orientazione degli orbitali rispetto ad un asse prefissato.

- *Come si possono rappresentare gli orbitali?*

$\Psi(x, y, z)$, funzione d’onda soluzione dell’equazione d’onda, non ha alcun significato fisico diretto. Si può dimostrare però che $\Psi^2(x, y, z)$ rappresenta la **densità di probabilità** di trovare l’elettrone in un punto (x, y, z) dello spazio. Regioni ad alta

densità di elettroni presentano un'alta probabilità di localizzare l'elettrone, mentre il contrario vale per regioni a bassa densità di elettroni.

La funzione che dal punto di vista fisico assume un significato più interessante è però $r^2\Psi^2$, ovvero la **funzione di distribuzione radiale**. Essa rappresenta la probabilità di trovare un elettrone su una superficie sferica con centro nel nucleo con raggio pari a r . Questa funzione ha valore nullo nel nucleo e cresce fino ad arrivare ad un massimo, per poi decrescere e tendere a zero. Nel caso dell'atomo di idrogeno il massimo della funzione di distribuzione radiale si ha per $r = 0.53 \text{ \AA}$: ciò significa che l'elettrone dell'atomo di idrogeno si troverà alla distanza di 0.53 \AA dal nucleo con maggiore probabilità che a qualsiasi altra distanza (Fig. 1). Va sottolineato che in base al modello atomico che stiamo qui presentando, l'elettrone potrà trovarsi ad una qualsiasi distanza dal nucleo, ma non tutte avranno la stessa probabilità.

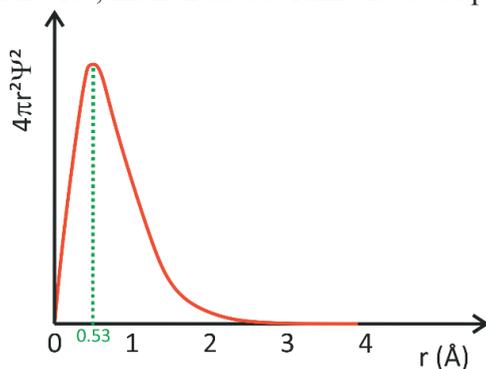


Figura 1

Per i motivi prima esposti, una maniera conveniente per rappresentare graficamente un orbitale è quella riportata in Figura 2, utilizzando cioè una rappresentazione a *nuvola elettronica*.

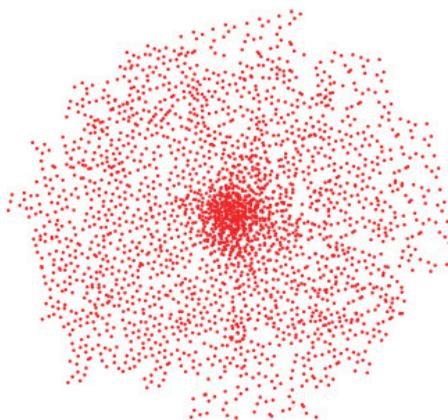


Figura 2

Possiamo immaginare di ottenere tale immagine assumendo che l'elettrone sia una piccola particella che si muove attorno al nucleo e a cui noi scattiamo una foto a intervalli regolari (per esempio ogni secondo per alcune migliaia di secondi). L'immagine che si otterrebbe dalla sovrapposizione di queste migliaia di foto scattate nel tempo risulterebbe costituita da una nuvola di punti: la densità dei puntini sarà maggiore dove è più facile trovare l'elettrone, mentre sarà minore dove la probabilità scema.

Per questo motivo è bene provare a immaginare l'elettrone che si trova in un atomo come una sorta di *nuvola dai contorni sfumati* (e non come una pallina che ruota attorno al nucleo). È chiaro quindi che le rappresentazioni degli orbitali che vediamo sui libri, vogliono suggerirci proprio tale idea: nello specifico tali rappresentazioni vanno a indicare il volume all'interno del quale si ha una probabilità definita (ad esempio >90%) di trovare l'elettrone (si vedano le Figure 3-5). Al di fuori di tali confini non è escluso che si trovi l'elettrone. È solo poco probabile. A questo proposito vale la pena osservare che la superficie che delimita queste rappresentazioni non è reale: come prima evidenziato, è sempre utile tenere a mente che il nucleo dell'atomo è circondato da una "nuvola elettronica" e non da tanti "sacchetti" dalle forme più o meno esotiche che "contengono" l'elettrone. Quando comunemente diciamo che "un elettrone *si trova* in un determinato orbitale" oppure "*occupa* un certo orbitale", intendiamo infatti dire che la distribuzione della densità elettronica (o la probabilità di localizzare l'elettrone nello spazio) è descritta dal quadrato della funzione d'onda associata a quell'orbitale.

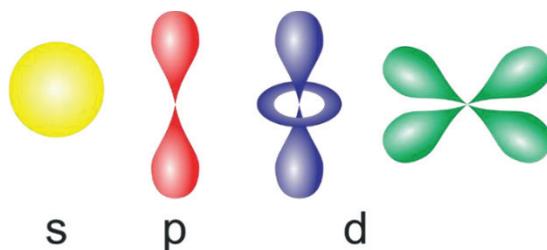


Figura 3

Come prima accennato, man mano che aumenta la loro energia (ovvero al crescere di n), gli orbitali hanno dimensioni sempre più grandi. Dalla Figura 3 è evidente però che gli orbitali si distinguono anche per **forma** (sferica, bilobata, a croce, ...).

Dall'analisi della funzione d'onda si può dimostrare infatti che a valori del numero quantico secondario $\ell=0$ corrispondono orbitali di tipo *s* in cui la distribuzione elettronica intorno al nucleo è sferica. Ci saranno allora gli orbitali 1s, 2s, 3s, ... che

corrispondono rispettivamente ad orbitali con numero quantico principale $n= 1, 2, 3, \dots$ e numero quantico secondario $\ell=0$. Tutti questi orbitali si rappresenteranno con delle sfere centrate sul nucleo, che al crescere di n saranno caratterizzate da un raggio sempre maggiore (Figura 4)., ovvero il punto di massima probabilità si trova sempre più lontano dal nucleo

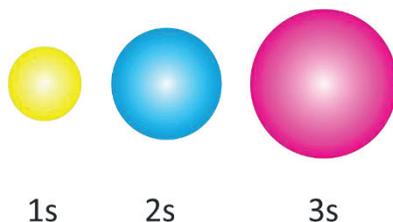


Figura 4

Gli orbitali atomici per i quali $\ell= 1$ sono detti invece orbitali p ed hanno tutti una forma bilobata, che ricorda il numero 8. Dal momento che, quando $\ell= 1$, m_ℓ può assumere i valori $+1, 0$ e -1 , per gli orbitali p sono possibili tre orientazioni spaziali dirette lungo le tre direzioni mutuamente perpendicolari (x, y, z) di una terna di assi cartesiani (Figura 5). Agli orbitali p è perciò assegnato il nome a seconda dell'asse lungo il quale sono diretti (ad esempio, nel caso degli orbitali con $n=3$ ed $\ell= 1$ avremo, in corrispondenza di $m_\ell=1, 0, -1$ i simboli $3p_x, 3p_y, 3p_z$).

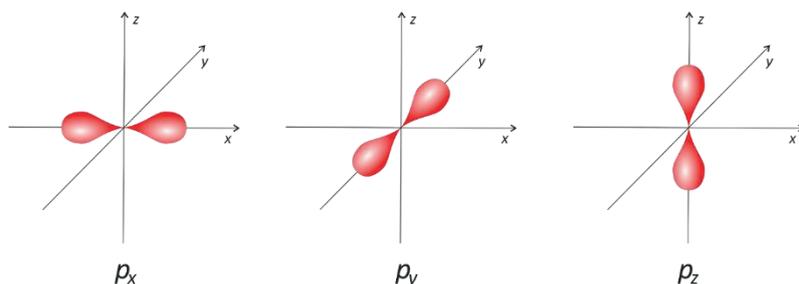


Figura 5

Tutti gli orbitali p sono caratterizzati da un piano immaginario che passa attraverso il nucleo e divide la regione di densità elettronica in due metà (in giallo in Figura 6). Tale piano è detto **piano nodale**, e rappresenta una superficie nella quale la probabilità di trovare l'elettrone è nulla. Questo significa che, al contrario degli orbitali s, per gli orbitali p non c'è alcuna probabilità di trovare l'elettrone sul nucleo. Gli orbitali d sono cinque. Ciascuno di essi presenta due superfici nodali, che delimitano quattro regioni di densità elettronica. Essi sono indicati come $d_{z^2}, d_{x^2-y^2}, d_{xy}, d_{xz}$ e d_{yz} . Gli orbitali f sono sette.

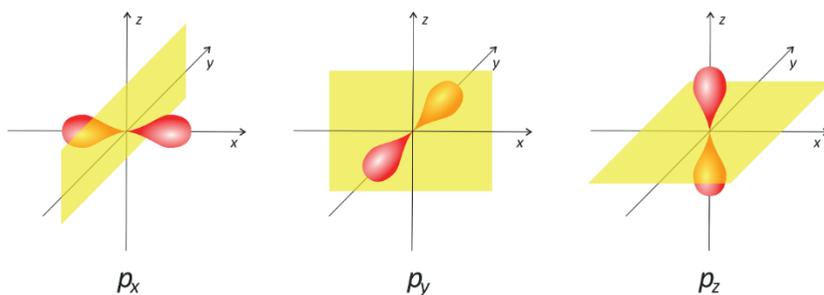


Figura 6

Non tutte le proprietà dell'elettrone possono essere descritte dai tre numeri quantici visti finora. Alcuni dettagli degli spettri atomici si spiegano ammettendo che l'elettrone si comporti come una piccola sfera che ruota su sé stessa (in inglese: *to spin*), attorno a un proprio asse, secondo due versi opposti. A questa rotazione, che genera un piccolo campo magnetico, è associato un quarto numero quantico (anch'esso quantizzato, essendo possibili solo due stati): il **numero quantico magnetico di spin**, m_s . Esso può assumere solo due valori, pari a $+1/2$ e $-1/2$, che vengono simbolicamente rappresentati con una freccia verso l'alto (\uparrow) ed una verso il basso (\downarrow), (Figura 7).

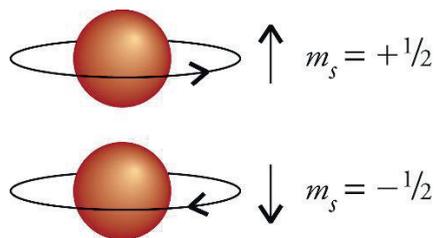


Figura 7

Configurazione elettronica

Come precedentemente accennato, gli orbitali si distinguono per l'energia che compete loro e sono raggruppabili in livelli e sottolivelli energetici. Essi possono corrispondere ciascuno a un solo elettrone oppure a due elettroni (diremo quindi che possono essere *occupati* da uno o due elettroni). È escluso che uno stesso orbitale possa essere occupato da più di due elettroni. È quindi di grande interesse andare a stabilire quale sia la *distribuzione* degli elettroni tra i vari livelli e sottolivelli, ovvero la **configurazione elettronica** di un atomo: da essa, infatti, deriveranno le proprietà chimiche dell'atomo in esame. Tale distribuzione degli elettroni negli orbitali (operazione nota anche con il termine di *aufbau*) va effettuata sistemando gli

elettroni in orbitali di energia via via crescente, a partire da quello a più bassa energia (ovvero l'orbitale 1s).

Per esempio, quando l'atomo di idrogeno si troverà nel suo stato a minore energia, **lo stato fondamentale**, il suo unico elettrone occuperà un orbitale del livello energetico più basso disponibile, cioè del livello con $n=1$ (Figura 8). D'altra parte, il livello con $n=1$ contiene un solo orbitale, di forma sferica, indicato come orbitale 1s (dove 1 indica il livello n , ed s indica $\ell=0$). Ne discende che l'elettrone che occupa un orbitale 1s è definito dai quattro numeri quantici $n=1$, $\ell=0$, $m_\ell=0$ e $m_s=+1/2$ o $-1/2$. Possiamo allora dire che la configurazione elettronica dell'idrogeno nel suo stato fondamentale, cioè la distribuzione degli elettroni tra i vari livelli, sarà $1s^1$ (si legge "uno esse uno").

Vale la pena osservare che se l'atomo viene eccitato fornendo all'elettrone energia sufficiente a raggiungere il livello con $n=2$, esso può occupare uno qualsiasi dei quattro orbitali presenti in questo livello (2s, 2p_x, 2p_y, 2p_z), poiché questi orbitali hanno tutti la stessa energia. Un apporto di energia ancora maggiore promuove l'elettrone al livello con $n=3$, dove può occupare uno qualsiasi dei nove orbitali presenti (3s, 3p_x, 3p_y, 3p_z, 3d_{z²}, 3d_{x²-y²}, 3d_{xy}, 3d_{xz} e 3d_{yz}) e così via (Figura 8). Infine, se si fornisce sufficiente energia per vincere l'attrazione del nucleo, l'elettrone viene espulso dall'atomo: a questo punto si dice che l'atomo è stato ionizzato.

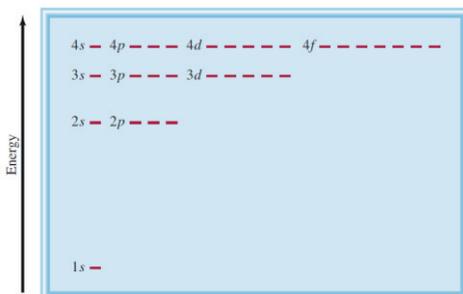


Figura 8

- Cosa succede nel caso degli atomi contenenti più di un elettrone?

In questi sistemi sorge la necessità di tenere conto delle interazioni fra elettroni poiché l'energia potenziale di un elettrone non dipenderà solo dalla sua distanza dal nucleo ma anche dalle interazioni con gli altri elettroni.

Purtroppo, l'introduzione di questi nuovi termini complica ulteriormente la già complessa equazione di Schrödinger che già con sistemi a due elettroni (quali l'atomo di elio) non è più risolvibile esattamente. Si utilizzano allora soluzioni approssimate che utilizzano i risultati ottenuti per l'atomo di idrogeno (i cosiddetti **orbitali idrogenoidi**). Gli stati dei singoli elettroni negli atomi polielettronici sono

quindi ancora rappresentati da funzioni d'onda, o orbitali, caratterizzati da numeri quantici che hanno significati analoghi a quelli definiti per l'atomo di idrogeno:

- n , numero quantico principale, definisce un livello energetico o guscio elettronico che si indica con le lettere K, L, M, N in corrispondenza a $n=1, 2, 3, 4$. Può assumere valori interi che vanno da 1 a ∞ .
- ℓ , numero quantico secondario, definisce all'interno di ogni livello, un sottolivello che viene indicato con le lettere s, p, d, f in corrispondenza di $\ell=0, 1, 2, 3$. Può assumere valori interi che vanno da 0 a $n-1$.
- m_ℓ , numero quantico magnetico, definisce l'orientazione della corrente elettronica orbitale. Può assumere valori interi che vanno da $-\ell$ a $+\ell$, zero incluso.
- m_s , numero quantico magnetico di spin, definisce l'orientazione della corrente elettronica di spin. Può assumere solo i valori $m_s=+1/2$ e $m_s=-1/2$.

Una importante differenza fra l'atomo di idrogeno e atomi polielettronici è che nel primo, dove non esistono repulsioni elettrone-elettrone, tutti gli orbitali di un dato livello, cioè con lo stesso valore di n , possiedono la stessa energia (Figura 8). Negli atomi polielettronici, invece, le repulsioni elettrone-elettrone fanno sì che, nell'ambito di un certo livello, l'energia di un orbitale dipenda anche dal numero quantico secondario: ad esempio l'energia dell'orbitale 3s sia minore di quella di un orbitale 3p, la quale a sua volta è minore di quella di un orbitale 3d e così via. Gli orbitali di un dato sottolivello, tuttavia, hanno tutti la stessa energia: ad esempio, i cinque orbitali 4d, possiedono la medesima energia. L'ordine dei livelli energetici in un atomo polielettronico risulta allora quello mostrato in Figura 9. Una maniera semplice per ricordare questo ordine energetico è descritta dalla *regola della diagonale* (o di Madelung) mostrata in Figura 10.

Stabilito l'ordine dei livelli energetici in un atomo polielettronico si può quindi procedere alla definizione della sua configurazione elettronica. Si noti che, a meno che non sia diversamente indicato, ci si riferisce comunemente ad atomi che si trovano nel loro stato fondamentale (ovvero di minima energia totale).

Come prima detto, gli elettroni occupano gli orbitali in ordine di energia crescente, a partire da quello a energia minore. Non è possibile però sistemare tutti gli elettroni dell'atomo in questo orbitale. Il **principio di esclusione di Pauli** stabilisce infatti che due elettroni in un atomo non possono avere tutti e quattro i numeri quantici uguali. Il significato fisico di questo principio è che due elettroni non possono stare nello stesso posto allo stesso tempo.

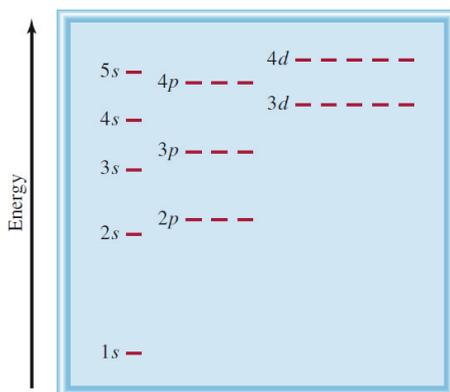


Figura 9

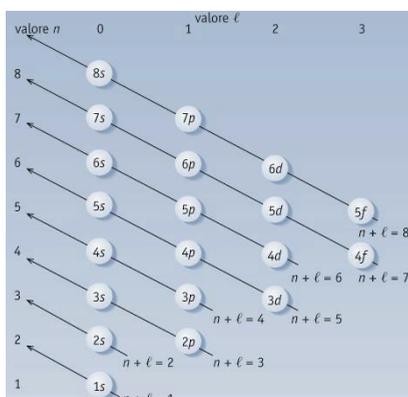


Figura 10

Ciò significa che un dato orbitale può accomodare al massimo due elettroni, e quando due elettroni occupano lo stesso orbitale i loro spin devono essere appaiati (ovvero antiparalleli, e si indicano con il simbolo $\uparrow\downarrow$). Quindi, un elettrone che occupa un orbitale 1s sarà caratterizzato dai seguenti numeri quantici: $n=1$, $\ell=0$, $m_\ell=0$ e $m_s=+1/2$. Un secondo elettrone potrà ancora occupare l'orbitale 1s, a patto che abbia spin antiparallelo al primo, e dunque: $n=1$, $\ell=0$, $m_\ell=0$ e $m_s=-1/2$. Un terzo elettrone non potrà occupare lo stesso orbitale, in quanto necessariamente avrebbe gli stessi quattro numeri quantici di uno dei due elettroni prima descritti. Dovrà quindi necessariamente occupare un orbitale con energia più elevata (o, qualora disponibile, un orbitale degenero). Nel fare questo si aggiungeranno elettroni agli orbitali partendo da quello disponibile ad energia più bassa seguendo l'ordine mostrato nella Figura 9.

Nel processo di riempimento di un sottolivello, cioè di orbitali degeneri, gli elettroni occuperanno prima il numero massimo di orbitali disponibili nel sottolivello, sistemandosi tutti con spin paralleli. Solo dopo, quando non ci saranno più orbitali degeneri vuoti disponibili, verranno appaiati in quelli semipieni. Questa regola, che prende il nome di **principio della massima molteplicità** (o regola di Hund), può essere ricondotta in prima battuta alla repulsione elettrostatica fra gli elettroni: quando due elettroni si trovano in orbitali separati saranno mediamente più lontani di quando si trovano nello stesso orbitale e quindi la loro repulsione, e quindi l'energia dell'atomo, sarà inferiore.

Una maniera graficamente accattivante di presentare le configurazioni elettroniche degli atomi è mediante l'uso dei **diagrammi a caselle**: secondo questa simbologia gli orbitali sono rappresentati da quadrati e gli elettroni da frecce il cui verso indica lo spin. Nella Figura 11 sono riportate le configurazioni elettroniche per i primi 11 elementi della tavola periodica.

H (Z=1)	$1s^1$	
He (Z=2)	$1s^2$	
Li (Z=3)	$1s^2 2s^1$	 
Be (Z=4)	$1s^2 2s^2$	 
B (Z=5)	$1s^2 2s^2 2p^1$	    
C (Z=6)	$1s^2 2s^2 2p^2$	    
N (Z=7)	$1s^2 2s^2 2p^3$	    
O (Z=8)	$1s^2 2s^2 2p^4$	    
F (Z=9)	$1s^2 2s^2 2p^5$	    
Ne (Z=10)	$1s^2 2s^2 2p^6$	    
Na (Z=11)	$1s^2 2s^2 2p^6 3s^1$	     

Figura 11

Vale la pena osservare infine che, quando si tratta di scrivere la configurazione elettronica di atomi con molti elettroni, è particolarmente comodo usare la seguente simbologia (qui riportata per semplicità nel caso dell'atomo di sodio):



Come possiamo osservare, in queste rappresentazioni il simbolo $1s^2 2s^2 2p^6$ è stato sostituito dal simbolo dell'elemento chimico che ha configurazione $1s^2 2s^2 2p^6$, il neon, [Ne], che come si può facilmente osservare, ha completato il livello elettronico con $n=2$. Questa simbologia ci spinge ad osservare che la configurazione dell'atomo di sodio (così come quella di un qualsiasi altro atomo) è descrivibile come costituita da un nocciolo interno completo detto *core* (in questo caso i livelli $n=1$ e $n=2$), ed elettroni attribuibili a un livello energetico "esterno", incompleto, detto **guscio di valenza** (in questo caso $n=3$).

Il guscio di valenza è di particolare interesse in quanto è quello che rende conto delle proprietà chimiche dell'elemento.

Celle solari polimeriche convenzionali: materiali, architetture e funzionamento

Rosita Diana

Dipartimento di Agricoltura, Università di Napoli Federico II

Abstract

Lo sviluppo di tecnologie a basso impatto ambientale rappresenta la grande sfida del futuro. La produzione di energie alternative, bio-sostenibili ed economicamente convenienti, è di fatto la maggiore aspirazione di un mondo in crescente bisogno energetico. In tale panorama si colloca la tecnologia del fotovoltaico, ed in particolare quella basata su materiali polimerici fotoattivi. Ampiamenti utilizzati nel campo dell'organica elettronica grazie alla loro elevata processabilità e modulabilità, i materiali polimerici fotoattivi verranno introdotti e discussi nel presente lavoro. In particolare, se ne riporterà una sintetica presentazione sulla base delle strutture chimiche più rappresentative e dei meccanismi fondamentali alla base della conversione fotovoltaica.

Introduzione

Attualmente più dell'80% dell'energia mondiale proviene da combustibili fossili che, oltre ad essere limitati nel tempo, hanno un notevole impatto ambientale a causa di emissioni quali anidride carbonica e gas serra, cui è imputabile un elevato contributo al riscaldamento globale.

La temperatura del globo terrestre è il risultato di un delicato equilibrio stabilito tra la radiazione proveniente dal Sole e l'energia radiata nello spazio dalla Terra stessa. Quest'ultima è fortemente influenzata dalla presenza e dalla composizione dell'atmosfera terrestre, che è costituita da una miscela di gas quali azoto, ossigeno, anidride carbonica, vapore acqueo, ozono, etc. L'anidride carbonica naturalmente presente nell'atmosfera, assorbe la radiazione nell'intervallo di lunghezze d'onda da 13 a 19 mm, mentre il vapore acqueo assorbe molto nella regione da 4 a 7 mm. Le attività umane, attraverso il rilascio di gas non legati ai cicli naturali, soprattutto anidride carbonica e clorofluorocarburi (CFC), ne aumentano a dismisura la concentrazione nell'atmosfera, al punto da alterare l'equilibrio del pianeta.

Ciò significa che le attività umane hanno ormai raggiunto dei livelli tali da intaccare direttamente l'ambiente. Gli effetti di tutto questo si stimano, su tempi non troppo lunghi, come devastanti per l'ecosistema Terra.

Un problema addizionale è legato al fatto che la disponibilità dei combustibili fossili è limitata, e dunque destinata ad esaurirsi nel tempo, mentre la domanda ed il consumo di energia continuano ad aumentare. La forte dipendenza di buona parte del mondo dai paesi produttori di combustibili fossili ne fanno levitare i prezzi, come pure per i prodotti derivati, con enormi implicazioni politico-economiche (Sharp Electronic).

In questo contesto è evidente che lo sviluppo di tecnologie a basso impatto ambientale, con ridotte emissioni e maggiore vantaggio economico, assumono notevole importanza per il futuro del pianeta. Poiché è la maggior produzione di gas serra è legata al settore dell'energia, a causa dell'impiego di combustibili fossili, tecnologie alternative come il fotovoltaico possono di fatto costituire una svolta.

La radiazione solare è sicuramente una tra le principali fonti di energia rinnovabile, in quanto eco-compatibile e ampiamente disponibile. Pertanto, ad oggi la via principale nella ricerca di nuove tecnologie sostenibili punta verso la conversione di energia solare in energia elettrica in modalità economiche ed efficienti. Tale tecnologia prende il nome di fotovoltaico.

Il mercato del fotovoltaico è basato su materiali inorganici, ed in particolare sul silicio cristallino, adoperati quali strato fotoattivo in forma di film sottile. L'indiscutibile vantaggio di una tecnologia che sfrutta film sottili come strato attivo sta nel risparmio di materiale semiconduttore utilizzato, con conseguente abbassamento dei costi.

Anche se il silicio resta in assoluto il materiale semiconduttore più studiato e sviluppato dal punto di vista tecnologico, esso presenta costi di produzione piuttosto elevati, tali da spingere a cercare altri materiali con un migliore rapporto costo/prestazioni. Pertanto, oltre al silicio cristallino, sono in uso tecnologie a film sottile basate su diversi semiconduttori inorganici, quali silicio amorfo (a-Si:H), tellurio di cadmio (CdTe) e calcopirite (CIS, CIGS). D'altro canto, sono note anche celle solari organiche, basate appunto su materiali fotoattivi organici. Queste si dividono in celle a base di piccole molecole, ossia materiali organici a basso peso molecolare (come le phthalocyanine, Pc) e celle polimeriche.

La conducibilità di materiali polimerici fu evidenziata nel 1977 da A.J. Heeger, G. McDiarmid e H. Shirakawa, ma solo nel 1985 Tang realizzò una cella solare organica, basata su uno strato attivo polimerico (Me-Ptcdi/ZnPc) con un'efficienza di conversione intorno all'1%. Ovviamente, nel mercato dell'energia la posizione competitiva di ogni tecnologia fotovoltaica è determinata da fattori fondamentali quali l'efficienza ed i costi. Anche se da allora si sono realizzati enormi passi avanti, con il raggiungimento di efficienze di conversione superiori, tali risultati non sono

ad oggi sufficienti per consentire la produzione su larga scala, se confrontati ai valori di efficienza delle celle fotovoltaiche tradizionali al silicio.

Ciononostante, le celle solari polimeriche (PSC) potrebbero essere la via per un abbassamento dei costi di produzione, grazie alla semplicità del processo di realizzazione e all'abbondanza e varietà di materiali precursori non così dispendiosi quanto i materiali inorganici. Le PSC vengono realizzate attraverso processi in soluzione a temperatura ambiente quali spin coating, screen printing e doctor blading, cosa che permette una manipolazione e processabilità relativamente facile, al contrario dei più complessi e costosi processi di evaporazione e deposizione delle celle tradizionali. Inoltre, materiali prettamente organici garantiscono economicità ed eco-sostenibilità anche nel processo di smaltimento.

Infine, le PSC possiedono il requisito peculiare di poter essere prodotte in una forma flessibile e semitrasparente, grazie alle caratteristiche intrinseche dello strato attivo. In particolare, i polimeri coniugati hanno catturato l'interesse della comunità scientifica in quanto capaci di combinare le proprietà optoelettroniche dei semiconduttori inorganici ed i vantaggi tecnologici dei materiali polimerici (Hadziioannou e van Putten, 2000).

Pertanto, c'è tuttora un vivo interesse della comunità scientifica per materiali polimerici innovativi che siano davvero competitivi nel fotovoltaico, garantendo un'efficienza di conversione adeguata e stabile nel tempo. A seguire si darà una breve panoramica dei materiali, delle architetture tipicamente in uso, e del funzionamento delle PSC.

Celle solari polimeriche

Materiali

In generale i semiconduttori organici tradizionali si possono inquadrare come semiconduttori intrinseci a gap elevato in quanto presentano gap superiore a 1.4 eV ed hanno bassa densità di portatori intrinseci a temperatura ambiente al buio (Gunes *et al.*, 2007). Per ottenere semiconduttori estrinseci si usano drogaggi fotochimici.

Le PSC si possono suddividere a seconda dei materiali utilizzati come donore e accettore nello strato attivo.

In particolare si trovano in letteratura

- celle polimero – fullerene, in cui il polimero è usato come materiale donore ed il fullerene come accettore.
- celle polimero – polimero in cui sia il donore che l'accettore sono polimeri coniugati
- celle polimero – cristalli inorganici.

Tra queste le efficienze più alte si ottengono per la tipologia polimero – fullerene. Per quel che riguarda i dispositivi polimero – fullerene, si riportano alcuni esempi di polimeri donori tradizionalmente utilizzati:

- polimeri derivati del fenilen-vinilene come il poly[2-methoxy-5-(3,7-dimethyloctyloxy)-1,4-phenylenevinylene (MDMO-PPV), mostrato in Figura 1a
- polimeri derivati del tiofene come i poly(3-alkylthiophene) (P3AT), tra i quali sono di interesse il poly(3-hexylthiophene) P3HT (Figura 1b), il poly(3-octylthiophene) P3OT e il poly(3-dodecylthiophene) P3DDT

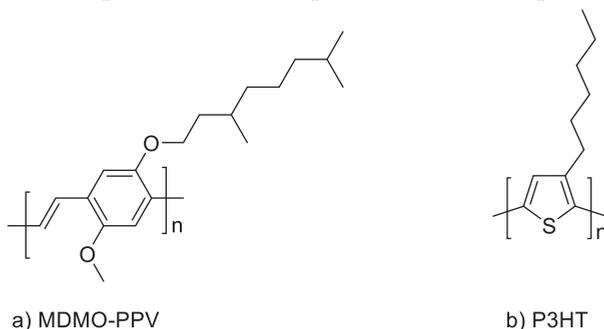


Figura 1: Esempi di polimeri usati come specie donore nelle PSC

Per gli accettori, la scelta tipicamente ricade su derivati del fullerene C60, di cui il [6,6]-phenyl-C61-butyric acid methyl ester (PCBM) è il derivato fullerenico tradizionalmente più utilizzato (Fig. 2). Tali materiali, inseriti in una matrice elettron-donor, danno luogo a reazioni di fotodrogaggio, cioè ad un trasferimento di elettroni dal donore, eccitato otticamente, all'accettore.

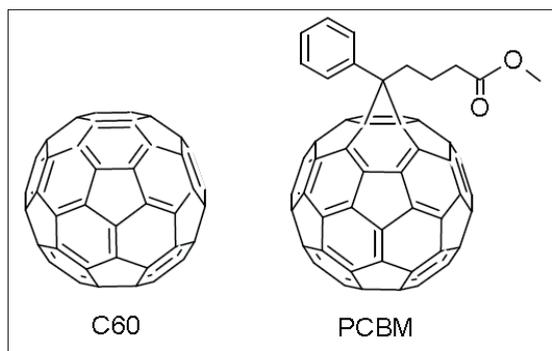


Figura 2. Struttura del fullerene (PCBM) più usato come accettore nelle PSC.

Architetture di celle polimeriche

Lo strato attivo, racchiuso tra due elettrodi e poggiato su un adeguato substrato, è il cuore di un qualunque dispositivo fotovoltaico. Nel caso delle PSC il substrato è in genere vetro o plastica, mentre gli elettrodi sono costituiti da indium-tin oxide (ITO) e da un metallo quale ad esempio alluminio. L'ITO è un ossido conduttivo trasparente (TCO) caratterizzato da un lavoro di estrazione che si allinea bene al livello HOMO della maggior parte dei polimeri coniugati e costituisce quindi l'elettrodo positivo. L'alluminio è impiegato per l'elettrodo negativo poiché presenta un lavoro di estrazione di circa 4.3 eV che si accoppia bene con il livello LUMO dell'accettore PCBM. L'elettrodo positivo è ricoperto da un sottile strato di poly(3,4-ethylene-dioxythiophene):poly(styrenesulfonate) (PEDOT:PSS). Il PEDOT è un polimero coniugato derivato dal tiofene che, miscelato con il PSS, presenta elevata trasmittanza nel visibile e buona conducibilità. Ampiamente utilizzato in dispositivi elettronici organici quali gli OLED come strato iniettore di lacune, nelle PSC il PEDOT:PSS ha il compito di migliorare la raccolta di lacune e la qualità superficiale dell'elettrodo positivo, riducendo la probabilità di cortocircuitare il dispositivo.

In generale, l'idea dell'eterogiunzione è di impiegare due materiali con differente affinità elettronica in modo che all'interfaccia i potenziali risultanti siano tali da favorire la dissociazione degli eccitoni.

L'introduzione dell'architettura bulk heterojunction (BHJ), avvenuta a metà degli anni Novanta, ha portato ad uno sviluppo decisivo della tecnologia fotovoltaica basata sui polimeri semiconduttori. In questa struttura il donore e l'accettore sono miscelati in modo da creare molteplici interfacce donore-accettore le cui dimensioni sono confrontabili con la lunghezza di diffusione degli eccitoni (Fig. 3). Ciò riduce moltissimo i processi di decadimento eccitonico, in quanto ogni eccitone fotogenerato "vede" un'interfaccia con l'accettore entro la sua lunghezza di diffusione, dove si ha una rapida dissociazione. Questo determina un aumento dell'efficienza di conversione in quanto la generazione di carica si ha in tutto lo strato attivo e non soltanto in una sua parte, come avviene nella struttura bilayer. Nella miscela donore – accettore, però, non c'è una direzione preferenziale per i campi interni, cioè elettroni e lacune creati nel volume di materiale attivo non hanno nessuna direzione preferenziale verso cui muoversi. Quindi è estremamente importante creare un'asimmetria, attraverso lavori di estrazione diversi degli elettrodi, per impedire che i portatori si muovano per diffusione. Inoltre, una volta dissociati gli eccitoni, le cariche create hanno bisogno di percorsi continui per

raggiungere gli elettrodi. In altri termini, il donore e l'accettore devono formare una rete interpenetrante su scala nanometrica.

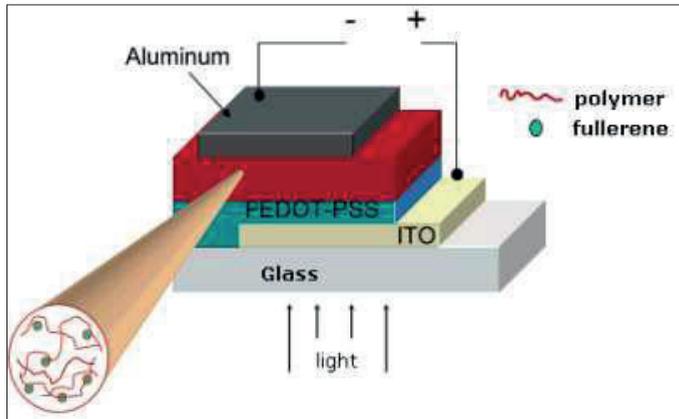


Figura 3. Architettura di una semplice cella polimerica solare dove lo strato attivo viene depositato secondo il concetto della bulk heterojunction.

Funzionamento

La distinzione più importante, dal punto di vista del funzionamento, tra celle solari inorganiche ed organiche sta nel fatto che nelle prime l'assorbimento della radiazione solare genera direttamente dei portatori liberi nel materiale attivo, mentre nelle seconde si creano coppie di portatori legati saldamente da una forza elettrostatica, noti come eccitoni. Proprio per questo i dispositivi fotovoltaici organici sono anche noti come dispositivi eccitonici. Ciò porta a differenze sostanziali dei meccanismi fisici di fotoconversione nei due tipi di celle. Il processo di conversione della radiazione solare in energia elettrica in una cella polimerica si articola nei seguenti passi:

- assorbimento dei fotoni incidenti
- formazione degli eccitoni
- diffusione degli eccitoni
- dissociazione degli eccitoni all'interfaccia donore – accettore
- trasporto dei portatori liberi
- raccolta dei portatori agli elettrodi

Osserviamo che il primo passo è legato a processi ottici mentre i restanti cinque processi sono fenomeni di natura elettrica. Inoltre, i primi due passi sono spesso

raggruppati in un unico fenomeno indicato con il nome di fotogenerazione. In generale ad ogni passo del processo di conversione si può assegnare un'efficienza in quanto solo una parte dei fotoni incidenti viene assorbita nel materiale e di questi solo una parte genera eccitoni. Inoltre, non tutti gli eccitoni fotogenerati danno luogo a portatori liberi e solo una parte di tali portatori viene raccolta e contribuisce effettivamente alla fotocorrente. Quindi ci sono molteplici fenomeni di perdita che devono essere analizzati ed opportunamente limitati per poter ottenere efficienze di conversione complessive più elevate. La Figura 4 riassume i livelli energetici ed il principio di funzionamento per una cella polimero – fullerene con strato attivo composto da una miscela di P3HT e PCBM.

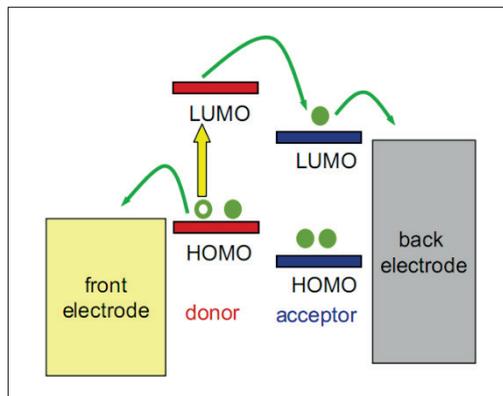
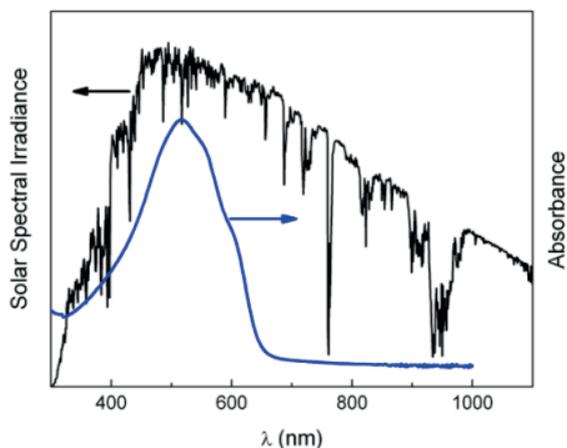


Figura 4. Schema dei livelli energetici e funzionamento di una cella polimero – fullerene

Nel seguito vengono descritti in dettaglio i vari passi del processo di conversione con riferimento ad una cella polimerica bulk-heterojunction.

1. **Assorbimento dei fotoni incidenti.** I polimeri coniugati presentano coefficienti di assorbimento molto elevati (dell'ordine di almeno 10^5 cm^{-1}) nel visibile e ciò significa che uno strato di poche centinaia di nanometri sarebbe già sufficiente per assorbire tutta la radiazione incidente alla lunghezza d'onda di picco (Blom *et al.*, 2007). Comunque, poichè la banda di assorbimento è relativamente piccola rispetto allo spettro solare, le prestazioni dei dispositivi polimerici sono ancora abbastanza basse al confronto con quelle dei dispositivi inorganici. Questo è evidenziato nella Figura 5 che mostra l'assorbanza del P3HT a confronto con l'irradianza spettrale del Sole. In generale, a causa di un gap piuttosto elevata (2.1 eV per il P3HT), l'assorbanza dei materiali polimerici risulta molto alta nel visibile e scarsa o assente nell'infrarosso (IR), il che impedisce di sfruttare al meglio la radiazione solare incidente. Questo è uno dei motivi per cui i semiconduttori inorganici come il silicio, seppur presentino un

coefficiente di assorbimento molto più basso, permettono di realizzare celle solari con un'efficienza di conversione più elevata. Proprio per questo motivo,



una delle strategie adottate per il miglioramento dell'efficienza dei dispositivi organici punta a sintetizzare polimeri a gap più basse per allargare la banda di assorbimento.

Figura 5: Spettro solare (in nero) e assorbanza del P3HT (in blu)

2. Definizione della gap. Affinché si abbia l'assorbimento della radiazione incidente, è necessario che l'energia associata alla gap del materiale sia minore o uguale all'energia dei fotoni incidenti. Nei semiconduttori inorganici quando si parla di gap ci si riferisce tipicamente alla differenza di energia tra banda di valenza e banda di conduzione in cui si trovano, rispettivamente, lacune ed elettroni liberi. Si tratta quindi di una banda proibita di tipo elettronico (Sun e Bonner, 2005). Nei semiconduttori organici come i polimeri coniugati, ma anche nei solidi amorfi, si parla più propriamente di gap ottica, in quanto è più complicato definire l'andamento delle bande a causa di uno scarso ordine periodico a lunga distanza. Inoltre, la fotoeccitazione tra gli orbitali molecolari HOMO e LUMO nei materiali organici non genera direttamente portatori liberi. La gap ottica E_{go} e quella elettronica E_{ge} nei semiconduttori organici sono legate attraverso la seguente relazione: $E_{eg} = E_{go} + E_B$, dove E_B rappresenta la minima energia necessaria per separare l'elettrone dalla lacuna, ossia l'energia necessaria a dissociare l'eccitone ed è nota come *exciton binding energy* (Knapfer, 2003). Se si definisce la banda di valenza come quella banda contenente le lacune libere e la banda di conduzione come quella che contiene gli elettroni liberi, allora per una struttura organica donore-accettore si può approssimare il livello LUMO dell'accettore alla banda di conduzione ed il

livello HOMO del donore alla banda di valenza. La gap ottica si può stimare direttamente attraverso lo spettro di assorbimento mentre la misura della gap elettronica necessita di tecniche elettrochimiche.

3. Formazione degli eccitoni. Come abbiamo più volte sottolineato, la fotoeccitazione nei materiali organici non porta direttamente alla formazione di portatori liberi bensì di coppie elettrone-lacuna, legate attraverso forze coulombiane, che prendono il nome di eccitoni. Tali eccitoni risultano quindi elettricamente neutri. Nei materiali organici, gli stati eccitati sono localizzati e non c'è transizione banda a banda, al contrario di ciò che avviene nei semiconduttori inorganici. Ciò è dovuto alle basse costanti dielettriche dei materiali organici che danno luogo ad eccitoni fortemente legati e localizzati sulle molecole dove si creano. Questo implica che gli effetti eccitonici si verifichino a temperatura ambiente, al contrario di quello che succede nei semiconduttori inorganici, dove si osservano transizioni eccitoniche solo se il materiale è altamente puro e solo a basse temperature (Bhattacharya, 1996). Sebbene in passato sia stato proposto che nei semiconduttori organici l'energia associata agli eccitoni fosse dello stesso ordine di quella termica kBT , i risultati sperimentali mostrano valori molto più alti (Pakbaz *et al.*, 1994). Uno dei metodi più affidabili per valutare tale energia è lo studio del quenching della fotoluminescenza dovuto al campo elettrico. Infatti, la dissociazione dell'eccitone richiede un trasferimento assistito dal campo proprio perché l'energia termica non è sufficiente a separare l'elettrone dalla lacuna. Inoltre, grazie a simulazioni che tengono conto, oltre all'effetto del campo elettrico, anche di perdite per decadimenti radiativi e processi non radiativi attraverso trappole, è stata calcolata un'energia di dissociazione pari a 0.4 ± 0.1 eV (Blom *et al.*, 2007), che è un valore molto più alto dei 26 meV forniti dall'energia termica a temperatura ambiente.

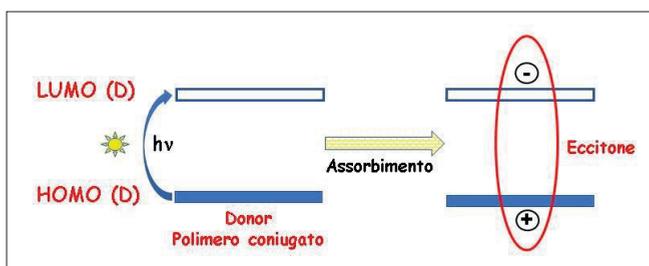


Figura 6. Assorbimento di un fotone e promozione di un elettrone del donatore dall'HOMO al LUMO.

4. Diffusione degli eccitoni. Una volta generati, gli eccitoni diffondono nel materiale, tipicamente attraverso meccanismi di hopping ossia salti tra stati energetici localizzati. La Figura 7 mostra lo schema dei livelli energetici all'interfaccia donore – accettore evidenziando i processi di formazione e diffusione degli eccitoni.

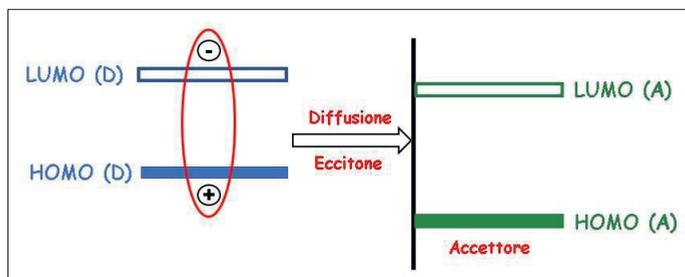


Figura 7: Formazione e diffusione degli eccitoni.

Gli eccitoni presentano, così come i portatori liberi, un tempo di vita finito e durante il loro processo diffusivo possono imbattersi in vari tipi di decadimento che ovviamente sono processi svantaggiosi per il funzionamento del dispositivo. Tali fenomeni riguardano, come accennato nel paragrafo precedente, decadimenti radiativi e intrappolamento in siti di impurità o difetti presenti nel materiale. In entrambi i casi si ha una diminuzione della quantità di eccitoni e quindi si tratta di processi di perdita. La lunghezza di diffusione degli eccitoni varia a seconda del tipo di polimero e dipende molto dalla struttura del materiale. Essa è tipicamente dell'ordine di una decina di nanometri. Ad esempio, nei polimeri derivati dal p-phenylene vinylene (PPV), come MDMO-PPV, la lunghezza di diffusione è stata stimata intorno ai 5-8 nm, ma può arrivare fino a 20 nm in altri polimeri (Markov *et al.*, 2005a; Markov *et al.*, 2005b).

D'altra parte, considerando dispositivi a singola interfaccia, come le celle polimeriche bilayer, per evitare il decadimento eccitonico si potrebbe pensare di ridurre semplicemente lo spessore dello strato attivo, ma ciò comporterebbe perdite di altro tipo. Infatti, se lo spessore dello strato attivo è minore della profondità di penetrazione della radiazione luminosa nel materiale, si ha la perdita di fotoni a causa di un ridotto assorbimento.

Poiché in teoria si vorrebbe che ogni eccitone fotogenerato fosse dissociato per dar luogo ad una coppia di portatori liberi, un modo per limitare la suddetta perdita di eccitoni è realizzare un materiale ideale privo di difetti, con fasi di donore e accettore ben ordinate e separate.

Questa è l'idea dell'architettura bulk-heterojunction in cui le fasi nanostrutturate del blend polimerico sono tali da permettere ad ogni eccitone, generato in qualsiasi posizione del materiale attivo, di diffondere e raggiungere un'interfaccia per la dissociazione in tutte le direzioni entro la sua lunghezza di diffusione. L'architettura bulk-heterojunction permette quindi di minimizzare le perdite eccitoniche aumentando le interfacce donore-accettore e, allo stesso tempo, garantisce uno spessore dello strato attivo adeguato all'assorbimento. Nel caso ideale, questa struttura garantisce la dissociazione di ogni eccitone creato entro una distanza pari alla lunghezza di diffusione.

5. Dissociazione degli eccitoni. Una volta che l'eccitone è arrivato all'interfaccia donore – accettore, deve essere dissociato perché si creino un elettrone ed una lacuna liberi di muoversi nel materiale. La dissociazione degli eccitoni in portatori liberi è quindi il passo fondamentale per la conversione dell'energia associata ai fotoni assorbiti nei dispositivi organici. Come già evidenziato in precedenza, a causa dell'elevata energia dell'eccitone nei polimeri coniugati, a temperatura ambiente l'energia termica non è sufficiente a dar luogo alla dissociazione. Quindi è necessario un campo elettrico adeguato che viene in genere fornito dall'interfaccia donore – accettore attraverso una brusca variazione di potenziale. La separazione di carica nel materiale attivo è anche detta trasferimento elettronico fotoindotto o fotodrogaggio, in quanto si tratta di una reazione di ossidoriduzione tra donore e accettore indotta dalla luce (Sun e Bonner, 2005). La Figura 8 mostra lo schema dei livelli energetici all'interfaccia donore – accettore durante il trasferimento elettronico fotoindotto.

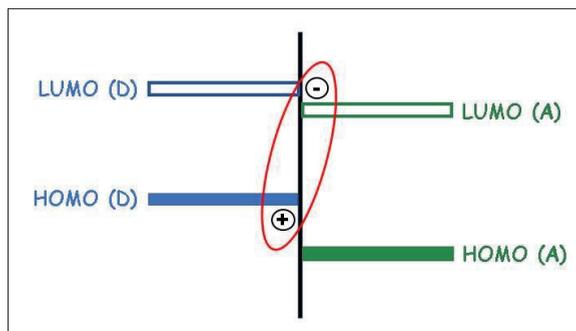


Figura 8. Trasferimento di carica dal donore all'accettore

Brabec et al. (Brabec *et al.*, 2001) hanno mostrato che il trasferimento fotoindotto in miscele polimero-fullerene costituite da MDMO-PPV:PCBM

avviene in circa 45 fs, cioè molto più rapidamente che in ogni altro processo di rilassamento. La fotoluminescenza, ad esempio, in genere avviene entro 1 ns. Proprio dallo studio del quenching della fotoluminescenza, essi hanno anche osservato che il trasferimento di carica dal MDMO-PPV al PCBM, oltre ad essere ultra-veloce, è anche efficiente. Il segnale di fotoluminescenza infatti si riduce di tre ordini di grandezza nella miscela MDMO-PPV:PCBM (1:3 wt.) rispetto a quello del solo polimero, indicando quindi il trasferimento di carica. Affinché il processo di dissociazione sia efficiente, è necessario che avvenga in un tempo molto più piccolo rispetto agli altri processi di decadimento che competono. Questi ultimi sono attribuiti principalmente alla ricombinazione dei portatori liberi dissociati in nuovi eccitoni. In particolare, si possono distinguere due tipi di ricombinazione, una in cui elettrone e lacuna generati dalla dissociazione dello stesso eccitone si riuniscono ricreando l'eccitone, e l'altra, detta anche ricombinazione bimolecolare, in cui un elettrone ed una lacuna generati dalla dissociazione di eccitoni diversi ricombinano formando un nuovo eccitone.

6. Trasporto. Dopo la dissociazione all'interfaccia, i portatori liberi devono essere trasportati agli elettrodi dove saranno successivamente raccolti, come mostrato schematicamente in Figura 9. Affinché il dispositivo fotovoltaico sia efficiente, è necessario che le cariche arrivino agli elettrodi prima che ricombinino, per poter contribuire alla corrente.

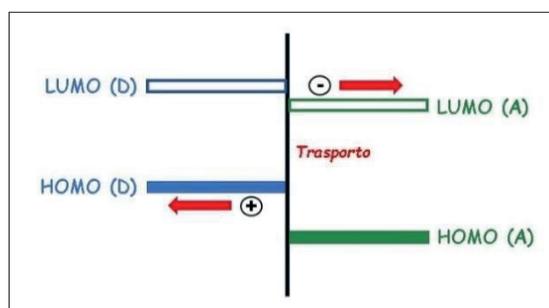


Figura 9. Trasporto verso gli elettrodi

Nelle celle polimeriche, dopo la dissociazione degli eccitoni, gli elettroni restano localizzati sulle molecole del fullerene mentre le lacune nelle catene del polimero. Successivamente questi portatori liberi devono essere trasportati agli elettrodi attraverso percorsi di percolazione nelle fasi di polimero e fullerene. Quindi il polimero trasporta le lacune verso l'anodo mentre il fullerene gli

elettroni verso il catodo. È evidente che il trasporto è molto sensibile alla morfologia del blend. Infatti, eventuali molecole di fullerene isolate, per esempio, fungono da trappole per gli elettroni e quindi rappresentano perdite di efficienza nel processo di trasporto.

I portatori di carica hanno dunque bisogno di una “forza” che li trascini verso gli elettrodi. A questa forza contribuiscono due fattori: a) i campi elettrici interni e b) i gradienti di concentrazione delle due specie di cariche.

In una qualunque giunzione p-n il trasporto verso gli elettrodi avviene grazie alla corrente di trascinamento, indotta dal campo elettrico, ed alla corrente di diffusione, indotta dal gradiente di concentrazione dei portatori stessi nel materiale. Con analisi dettagliate, che presuppongono la conoscenza dei profili di distribuzione delle cariche nel materiale, si può dimostrare che nei dispositivi a film ultra-sottile, come nel caso delle celle polimeriche, il trasporto è dominato dalla componente di trascinamento, mentre nei dispositivi più spessi, come le celle al silicio cristallino, la componente più importante della corrente è quella diffusiva (Hoppe e Sacrifitci, 2004).

7. Raccolta agli elettrodi. Come ultimo passo, i portatori di carica devono essere estratti dal dispositivo attraverso i contatti per poter dar luogo alla corrente nel circuito esterno. Tali contatti devono quindi essere di tipo ohmico, ossia devono presentare una resistenza trascurabile affinché la raccolta dei portatori avvenga in modo efficiente. In generale, la formazione di un contatto ohmico tra due materiali semiconduttori dipende dalla posizione relativa dei livelli di estrazione dei materiali stessi e deve essere tale da impedire la formazione di una regione svuotata, in modo da limitare la caduta di tensione per qualsiasi condizione di polarizzazione (Bhattacharya, 1996). In particolare, considerando l'interfaccia tra un metallo ed un generico semiconduttore e indicando con Φ_m e Φ_s i lavori di estrazione del metallo e del semiconduttore, si ha un contatto ohmico se $\Phi_s > \Phi_m$.

Per quanto riguarda i materiali organici, è stato proposto che la formazione di un perfetto contatto ohmico per gli elettroni si formi quando il livello LUMO dell'accettore è allineato con il livello di Fermi dell'elettrodo con lavoro di estrazione più basso (small work function electrode, SWFE), mentre per le lacune quando il livello HOMO del donore è allineato con il livello di Fermi dell'elettrodo con lavoro di estrazione più alto (large work function electrode, LWFE) (Yu *et al.*, 1995). Osserviamo però che anche se un elettrone o una lacuna arrivano ai rispettivi elettrodi, il fatto che vengano raccolti o meno non è un evento certo. Infatti, la probabilità è associata al meccanismo di penetrazione della barriera di potenziale

all'interfaccia con l'elettrodo ed è funzione della geometria e della qualità superficiale dell'interfaccia formata tra il materiale attivo ed il contatto. Quindi la scelta dei materiali adoperati per gli elettrodi così come il processo tecnologico di realizzazione sono aspetti altrettanto importanti per l'efficienza complessiva del dispositivo. Nelle celle polimeriche vengono spesso impiegati degli strati aggiuntivi interposti tra lo strato attivo e gli elettrodi proprio per migliorare la qualità delle interfacce. In particolare, è ormai consolidato l'uso di un sottile strato di PEDOT:PSS interposto tra ITO e blend che contribuisce a migliorare la raccolta di lacune. E, invece, meno frequente l'utilizzo di uno strato di LiF tra blend ed alluminio che permette di aumentare il FF e la V_{oc} dei dispositivi (Brabec *et al.*, 2002).

Bibliografia

- Bhattacharya, P. (1996). Semiconductor optoelectronic devices. 2nd edition, Prentice Hall
- Blom, P.W.M., Mihailetschi, V.D., Koster, L.J.A, Markov, D.E. (2007). Device physics of polymer:fullerene bulk heterojunction solar cells. *Advanced Materials*, 19, 1551-1566.
- Brabec, C. J., Zerza, G., Cerullo, G., De Silvestri, S., Luzzati, S., Hummelen, J. C., Sariciftci. N.S. (2001). Tracing photoinduced electron transfer process in conjugated polymer/fullerene bulk heterojunctions in real time. *Chemical Physics Letters* 340, 232-236.
- Brabec, C. J., Shaheen, S.E., Winder, C., Sariciftci, N.S., Denk, P. (2002). Effect of LiF/metal electrodes on the performance of plastic solar cells. *Applied Physics Letters*, 80, 1288-1290.
- Gunes, S., Neugebauer, H., Sariciftci, N. S. (2007) Conjugated Polymer-Based Organic Solar Cells. *Chemical Reviews*, 107, 1324-1338
- Hadziioannou, G., van Putten, P.F. (2000) *Semiconducting Polymers*, Wiley-VCH, Weinheim
- Hoppe, H., Sariciftci. N.S. (2005). Organic solar cells: an overview. *Journal of Materials Research*, 19, 2144-2151.
- Knupfer, M. (2003). Exciton binding energies in organic semiconductors. *Applied Physics A*, 77, 623-626.

- Markov, D. E., Amsterdam, E., Blom, P.W.M., Sieval, A.B., Hummelen, J.C. (2005). Accurate measurement of the exciton diffusion length in a conjugated polymer using a heterostructure with a side-chain cross-linked fullerene layer. *Journal of Physical Chemistry A*, 109, 5266-5274.
- Markov, D. E., Tanase, C., Blom, P. W. M., Wildeman, J. (2005). Simultaneous enhancement of charge transport and exciton diffusion in poly(p-phenylene vinylene) derivatives. *Physical Review B* 72, 045217.
- Pakbaz, K., Lee, C.H., Heeger, A.J., Hagler, T.W., McBranch, D. (1994). Nature of the primary photoexcitations in poly(arylene-vinylenes). *Synthetic Metals*, 64, 295.
- Sun, S.-S., Bonner, C. E. (2005). Optimization of Organic Solar Cells in Both Space and Energy-Time Domains, Chapter 8 in *Organic Photovoltaics*, Sun and Sariciftci Eds., Taylor and Francis.
- Yu, G., Gao, J., Hummelen, J., Wudl, F., Heeger, A. (1995). Polymer photovoltaic cells: enhanced efficiencies via a network of internal donor-acceptor heterojunctions. *Science*, 270, 1789-1791
- http://www.sharp.it/sharp/assets/internet/assets/images/Italy/pdf_Solar_LibrettoFotovoltaico_it.pdf

Sensori e Segnali nell'IoT

Silvio Barra

*Dipartimento di Ingegneria Elettrica e delle Tecnologie dell'Informazione,
Università di Napoli Federico II*

Introduzione

Si parla di Internet of Things (acronimo di IoT) o ancora di Internet delle Cose, ma forse sarebbe più corretto definirla Internet degli oggetti. Ci sono, infatti, gli oggetti intelligenti (i cosiddetti “smart objects”) alla base dell'Internet of Things; con tale terminologia non stiamo parlando soltanto di computer, smartphone e tablet, ma soprattutto degli oggetti che ci circondano all'interno delle nostre case, al lavoro, nelle città, nella vita di tutti i giorni. L'Internet of Things nasce proprio qui: dall'idea di portare nel mondo digitale gli oggetti della nostra esperienza quotidiana (vedi riferimento bibliografico). L'espressione Internet of Things è stata coniata circa 20 anni fa (si tenga in considerazione che Internet è stata inventata 50 anni fa). Sin dalla sua nascita, l'IoT si è resa altamente pervasiva nella nostra vita quotidiana, principalmente grazie a due fattori fondamentali:

- innanzitutto, l'IoT permette di fornire all'utenza un numero di servizi che lo rendono quasi indispensabile;
- in secondo luogo, può essere applicato ad una moltitudine di campi, essendo totalmente slegato dal contesto, in fase di progettazione, per poi divenire estremamente pervasivo e parte dell'ambiente stesso, in fase di esecuzione.

Dal punto di vista dell'utenza, una Internet of Things opera in maniera quasi trasparente. Il "quasi" è d'obbligo siccome l'utente deve essere sempre consapevole di "chi" fa "quale" utilizzo dei dati sensibili dell'utente stesso.

Nascita e Diffusione dell'Internet of Things

Un primo assaggio delle prime applicazioni della IoT lo abbiamo avuto con l'introduzione dei ben noti apparati RFID. Questa tecnologia che è stata ed è ancora utilizzata soprattutto per la gestione di magazzini (lettura veloce di codici a barre) si è evoluta in due branche: la prima è quella ormai nota come NFC (*Near Field Communication*) molto utilizzata per la lettura di carte contactless o per la lettura delle carte di credito dallo smartphone (Fig. 1); la seconda riguarda proprio il passaggio ai sopra citati smart objects che sono diventati pervasivi negli ambienti pubblici e privati al giorno d'oggi.



Figura 1. Sistemi RFID attivi e passivi ed il passaggio a NFC ed Internet Of Things

A differenza della tecnologia NFC, la IoT, però, va oltre gli oggetti intelligenti ed assume un pieno significato nella rete che interconnette questi oggetti. Esempi quotidiani di reti di oggetti possono essere i seguenti:

- Automobili: ad esempio i box GPS-GPRS che hanno finalità assicurative;
- Domotica: servizi cloud e altri servizi legati all'utilizzo crescente dell'Intelligenza Artificiale;
- Industrie: ambito in cui le tecnologie IoT stanno contribuendo in termini di distribuzione che in termini di automazione (Industrial-IoT);
- Città: le famose "smart cities". Pensiamo ai lampioni intelligenti delle nostre città che sono in grado di regolare la loro luminosità sulla base delle condizioni di visibilità correnti; oppure ai semafori che si sincronizza per creare un'onda verde e semplificare il passaggio di un mezzo di soccorso.

La diffusione dell'Internet of Things si è avuta maggiormente negli ultimi anni, periodo in cui il cloud è diventata una tecnologia onni presente e che ha cambiato il nostro modo di percepire Internet e l'elaborazione dei dati. In questo senso, il cloud svolge un ruolo di aggregatore, custode ed elaboratore dei dati generati dai processi e dagli Smart Objects.

Ambiti Applicativi

Di seguito verranno mostrati e descritti brevemente alcuni degli ambiti applicativi cui la IoT è applicata.

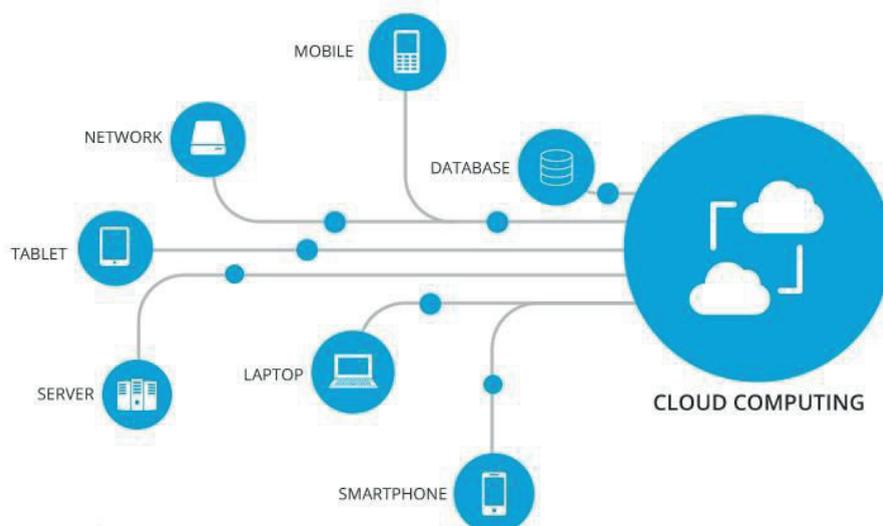


Figura 2. Cloud computing ed applicazioni (immagine tratta da <https://www.tutoreti.it/index.php/articoli/121-internet-of-things-iot-e-cloud-computing.html>)

Smart Agriculture

La Iot è applicata alla *Smart Agriculture* per il monitoraggio dei parametri microclimatici a supporto dell'agricoltura. Lo scopo è quello di migliorare la qualità dei prodotti riducendo le risorse utilizzate e di conseguenza l'impatto ambientale. Per tali opere vengono utilizzati in maniera massiccia sensori WiFi, droni, sensori ambientali e videocamere. In tale ambito, si vogliono raccogliere dati ed immagini su ogni singola area relativa ad una coltura. Vengono tracciati parametri di umidità e temperatura del suolo e delle piante, in tutte le fasi di maturazione della coltura. Tutti i dati raccolti sono inviati a sistemi cloud che analizzano lo stato di stress e di salute delle piante. Queste innovazioni consentono agli agricoltori di poter decidere in anticipo e con una precisione molto alta, come distribuire l'acqua, i pesticidi e i fertilizzanti.

La *smart agriculture* è molto diffusa in questi anni grazie a due iniziative mondiali: la sensibilizzazione per una vita a basso impatto ambientale, proattiva al problema del cambiamento climatico, e la promozione di iniziative per contribuire all'agricoltura nelle aree del mondo pi_u_povere.

Smart Car

Le Smart Car sono un altro classico esempio di IoT applicata ad un ambito piuttosto comune al giorno d'oggi. In tali ambienti, le auto ottengono l'abilità di comunicare in tempo reale non solo con il conducente, ma anche con altri veicoli o con l'infrastruttura stradale circostante allo scopo di prevenire incidenti, code o altre situazioni che possono rivelarsi spiacevoli. Vi sono diversi tipi di Smart Cars, a seconda del grado di "indipendenza" dell'automobile e del livello di pervasività all'interno dell'ambiente del conducente. I livelli maggiori di automazione li riscontriamo nelle cosiddette self driving cars (Livello di automazione 4), in cui il veicolo svolge tutte le funzioni critiche di sicurezza e di monitoraggio costante della strada, a meno di condizioni meteo estreme. Il livello 5 prevede, invece, che il conducente imposti solo la destinazione; senza alcun intervento del conducente, l'auto lo porta a destinazione. Una smart car è anche la semplice auto connessa che consente agli automobilisti, tramite uno smartphone, di ottenere le informazioni relative allo stato dell'auto, e di monitorarne i livelli. Tali smart car sono anche in grado di comunicare in maniera diretta ed autonoma con i centri di assistenza, nei casi in cui interventi tempestivi sono necessari.

Smart City

Prevedono il monitoraggio e la gestione degli elementi di una città (ad esempio mezzi per il trasporto pubblico, illuminazione pubblica e parcheggi) e dell'ambiente circostante per migliorarne vivibilità, sostenibilità e competitività. Obiettivi principali di queste enormi applicazioni sono, ad esempio:

- la riduzione dei consumi energetici (con lampioni intelligenti che si accendono in determinate condizioni di visibilità scarsa)
- ottimizzazione della raccolta dei rifiuti (alert agli addetti nei casi in cui un determinato tratto stradale necessita di particolare attenzione o un cassonetto risulta pieno - considerare, ad esempio l'applicazione di un sensore al cassonetto che ne monitora lo stato e lo comunica all'addetto delle pulizie)
- il miglioramento del trasporto pubblico (ottimizzazione dei tragitti in base al traffico, senza comunque saltare le fermate previste dalla corsa)
- riduzione del degrado urbano

Smart Home e Smart Metering

Molto spesso queste due categorie vanno di pari passo. Le Smart Home fanno riferimento alle applicazioni di domotica che permettono di trasformare l'abitazione vissuta in un ambiente intelligente, mediante l'applicazione di sensori e elettrodomestici dotati di sensoristica *ad hoc*. Si tenga in considerazione, ad esempio,

i casi in cui le lavatrici e le lavastoviglie si azionano da sole, in determinate ore della giornata in funzione del prezzo della corrente in quel particolare momento, o le persiane che si aprono autonomamente, ad esempio per permettere alle piante di prendere una sufficiente quantità di luce. Lo Smart Metering si riferisce invece alla strumentazione che si occupa di monitorare costantemente i consumi ed il carico energetico dell'ambiente. Abbiamo in questo caso bisogno di contatori intelligenti connessi a Sistemi di Raccomandazione che interagiscono con l'utente. Chiaramente lo Smart Metering si confà all'applicazione in ambiti domestici, ma non solo. Si immagina soluzioni simili applicate ad ambienti più estesi (intere aree metropolitane, quartieri), o molto più piccoli (come, ad esempio, camere che ospitano server che devono essere raffreddati o cavedi).

Industrial IoT

Si occupano dell'adozione di *Cyber Physical Systems*, connessione dei macchinari, degli operatori e dei prodotti per abilitare nuove logiche di gestione della produzione. Si tratta di applicazioni che si occupano ad esempio di migliorare ed ottimizzare le catene di montaggio e di ottimizzare il livello di automazione all'interno di un'industria. Anche le policy di gestione di magazzini rientrano in quest'ambito.

Gli Smart Objects

Gli *smart objects* (Fig. 3) permettono il verificarsi di tutte le applicazioni sopra discusse. Uno *smart Object* possiede una o più delle seguenti funzionalità:

- Identificazione: per essere riconosciuto univocamente in un ambiente
- Localizzazione: per essere localizzato all'interno di un ambiente
- Diagnosi di stato: per fornire informazioni sul suo stato corrente (Acceso/Spento/In funzione/Errore)
- Sistemi per l'interazione con l'ambiente: deve essere fornito di un sensore per permettergli di raccogliere informazioni e dati dall'ambiente
- Data Processing: non tutti gli *smart objects* hanno un sistema di elaborazione dati on-board, ma alcuni presentano un modulo di filtraggio dati
- per pulire i dati raccolti da rumore o da informazioni ridondanti
- Connessione: una scheda di rete o simile è necessaria per la trasmissione dei dati ad un server cloud, ad esempio.

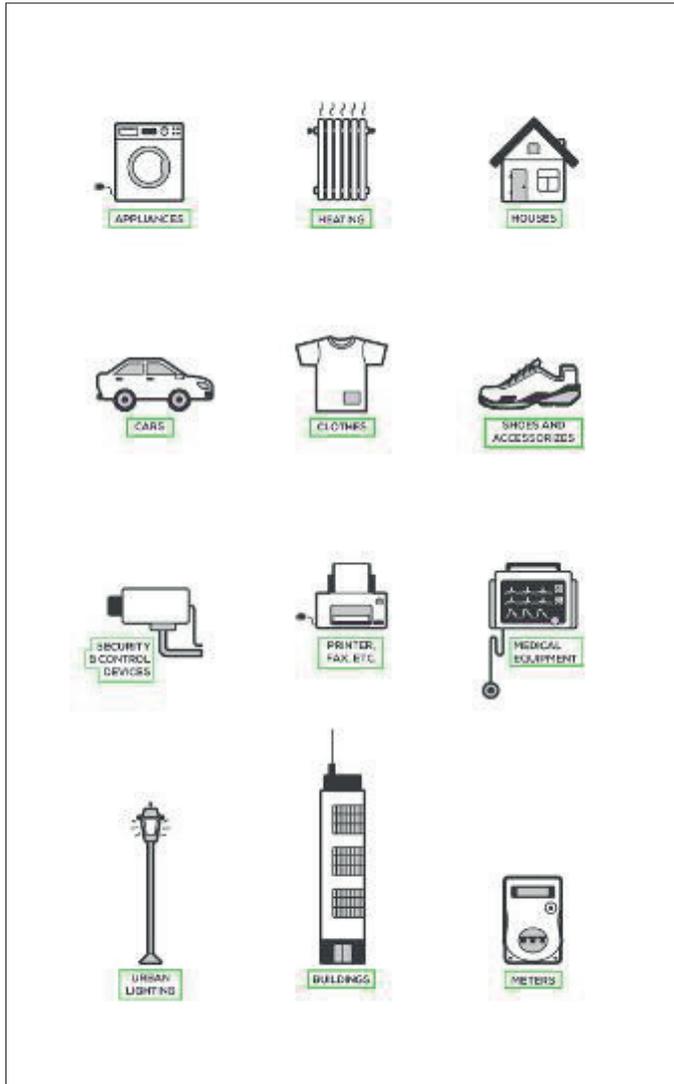


Figura 3. Esempi di Smart Objects

Bibliografia

Internet of Things (IoT): significato, esempi, ambiti applicativi e prospettive di mercato in Italia. https://blog.osservatori.net/it_it/cos-e-internet-of-things

Seminario Integrato

I Metalli Critici: fonti naturali e recupero da RAEE

Nicola Mondillo⁽¹⁾, Roberto Esposito⁽²⁾

⁽¹⁾ *Dipartimento di Scienze della Terra, dell'Ambiente e delle Risorse, Università di Napoli Federico II;*

⁽²⁾ *Dipartimento di Scienze Chimiche, Università di Napoli Federico II.*

Abstract

I cambiamenti sociali e tecnologici verificatisi nell'ultimo secolo (nell'ultimo ventennio in particolare) hanno prodotto, oltre che una forte crescita del consumo dei metalli storicamente più utilizzati, anche una forte domanda di metalli più rari (elementi delle terre rare, gallio, germanio, indio, cobalto, etc.), fortemente impiegati nell'industria high-tech

Molti di questi metalli sono recuperati dai giacimenti minerari, cioè accumuli naturali di minerali nella crosta terrestre che, per dimensione, concentrazione e localizzazione, sono sfruttabili con profitto. In questi depositi, gli elementi chimici di interesse economico raggiungono concentrazioni anche migliaia di volte superiori rispetto alla loro concentrazione media nella crosta terrestre. Ciò rende i giacimenti minerari delle importanti anomalie geologiche naturali. Comunemente si distingue tra giacimenti di metalli nobili (oro, argento, platino), di metalli di base e industriali (rame, zinco, ferro, alluminio), e di minerali (diamanti, pietre preziose, minerali industriali), che derivano sia da processi genetici profondi (endogeni), che superficiali (esogeni).

L'aumento del consumo di apparecchi elettrici ed elettronici e la necessità di abbracciare lo *Sviluppo Sostenibile* tramite un modello di *Economia Circolare* ci impone di trovare fonti di metalli critici rinnovabili. Una di queste sono i RAEE, i Rifiuti di Apparecchiature Elettriche ed Elettroniche, che possono essere un vero concentrato di questi elementi. Un recupero efficiente e sostenibile dei metalli da questi scarti deve prevedere lo sviluppo della chimica e la sua applicazione tramite la guida dei principi della *Green Chemistry*. Recenti sviluppi della ricerca in campo chimico stanno fornendo nuovi metodi sempre più efficienti e sostenibili per l'estrazione selettiva di metalli da RAEE.

Parte : I metalli critici

Nicola Mondillo

Introduzione

I cambiamenti sociali e tecnologici verificatisi nell'ultimo secolo (ma soprattutto nell'ultimo ventennio) hanno prodotto, oltre che una forte crescita del consumo dei metalli storicamente più utilizzati, anche una forte domanda di metalli più rari (elementi delle terre rare, gallio, germanio, indio, cobalto, etc.), fortemente impiegati nell'industria high-tech. Essendo presenti in tanti strumenti di uso quotidiano (es. PC, tablet, smartphone, etc.; Fig. 1), alcuni di questi metalli rari hanno un alto valore economico. Al fine di monitorarne le forniture, la Commissione Europea, sulla base dei consumi dei paesi membri e dell'evoluzione del mercato, stila dei rapporti biennali contenenti l'elenco dei metalli dei quali la nostra economia non può fare a meno. Tali metalli sono quindi classificati in base a due parametri fondamentali: l'importanza economica ed il rischio di approvvigionamento (Figura 1). L'importanza economica di un certo metallo dipende da altri fattori, come ad esempio la sua importanza nei prodotti finali (può essere sostituito da altri metalli con caratteristiche fisico-chimiche simili?), il valore economico dei settori industriali e tecnologici nei quali viene impiegato, la possibilità che si continui ad utilizzare negli anni a venire, etc. Il rischio di approvvigionamento dipende invece soprattutto dal numero di paesi che producono il metallo in questione (dai quali l'Unione Europea lo importa), e dalla stabilità politica di questi stessi paesi. I metalli che mostrano alto valore economico ed alto rischio di approvvigionamento per l'economia dell'Unione Europea sono detti “critici” (Fig. 1). Da quanto detto, è chiaro che la “criticità” di un certo metallo dipende fortemente dal paese o comunità che si considera, e dal contesto storico-sociale-temporale al quale ci riferiamo. Mantenere un costante approvvigionamento di queste risorse rappresenta una delle principali sfide del futuro per le quali è di cruciale importanza la conoscenza di discipline come la geologia e la chimica.

Fonti naturali: i giacimenti minerari

Molti metalli critici sono prelevati da fonti naturali, denominate “giacimenti minerari”. Questi sono accumuli di minerali nelle rocce della crosta terrestre che, per dimensione, concentrazione e localizzazione, sono sfruttabili con profitto. In questi depositi, gli elementi chimici di interesse economico raggiungono concentrazioni

anche migliaia di volte superiori rispetto alla loro concentrazione media nella crosta terrestre (Fig. 2).

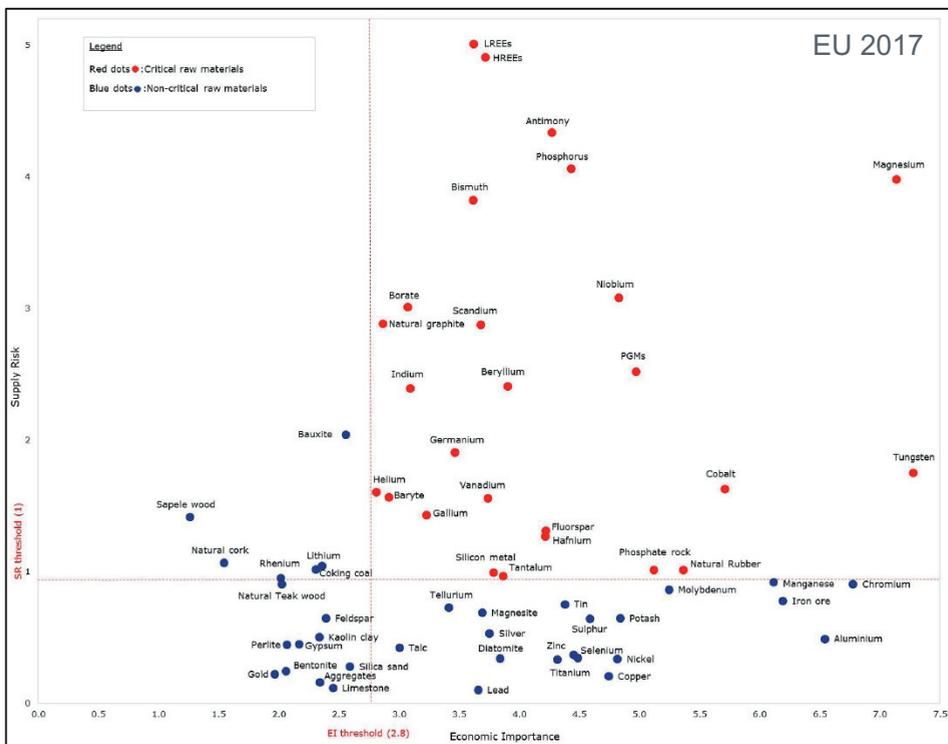
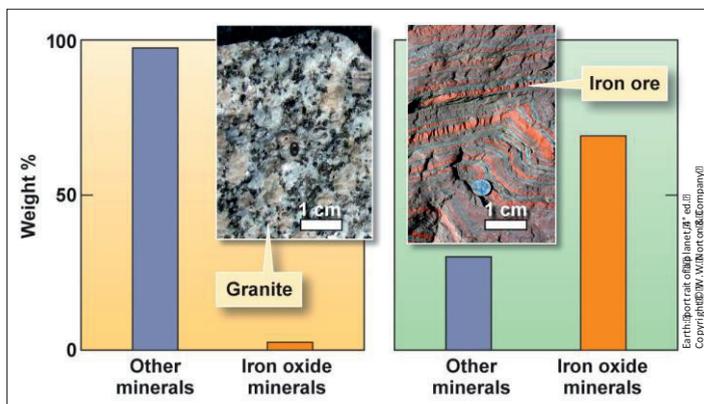


Figura 1.- (in alto) Elementi in uso in un moderno smartphone. (in basso) Distribuzione degli elementi chimici e delle materie prime di interesse economico sul diagramma di "criticità" (Importanza economica vs. rischio di approvvigionamento). Sono evidenziati in rosso gli elementi e le materie prime che, presentando alta importanza economica ed alto rischio di approvvigionamento, sono definiti "critici" (Commissione Europea 2017).

Ciò rende i giacimenti minerari delle importanti anomalie geochimiche e geologiche naturali. Comunemente si distingue tra giacimenti di metalli nobili (oro, argento, platino), di metalli di base e industriali (rame, zinco, ferro, alluminio), e di minerali (diamanti, pietre preziose, minerali industriali). Nei giacimenti minerari questi metalli hanno abbondanze, dette “tenori”, piuttosto variabili. Affinché un accumulo anomalo di metalli in una roccia sia considerato un “giacimento”, è necessario che la concentrazione del metallo nella roccia superi il “tenore minimo” per cui l’estrazione sia vantaggiosa (guadagni > costi totali). Ne consegue che al crescere del valore del metallo o del minerale di interesse, decresce il tenore minimo affinché l’estrazione sia vantaggiosa. Per questo motivo, i giacimenti di metalli preziosi hanno tenori piuttosto “bassi”: ad esempio nei giacimenti d’oro, il metallo ha una concentrazione variabile da poche decine a centinaia o migliaia di grammi per tonnellata di roccia (10-1000 g/t), mentre i metalli industriali, che hanno prezzi



piuttosto bassi, devono essere notevolmente arricchiti nelle rocce affinché possano formare dei giacimenti (Figura 2).

Figura 2. Rappresentazione grafica della differenza tra una roccia “normale (es. un granito), ed un giacimento di ferro. Il granito contiene minerali di ferro (Iron oxide minerals) in quantità troppo bassa affinché possano essere recuperati in maniera economicamente vantaggiosa. Un giacimento di ferro (schema a destra) deve contenere almeno il 60% in peso di minerali di ferro.

Da questo discorso risulta chiaro che i tenori minimi cambiano nel tempo, al variare del prezzo dei metalli sul mercato globale. Un fattore importantissimo per determinare il valore di un giacimento minerario è la tipologia di minerali che “incorporano” i metalli di interesse (che in rarissimi casi si trovano nelle rocce allo stato nativo). La maggior parte dei metalli di interesse economico sono presenti nei

giacimenti sotto forma di solfuri o ossidi metallici. I silicati (principali costituenti delle rocce della Terra) raramente formano minerali di interesse economico.

Come si formano i giacimenti minerari

I giacimenti minerari si formano, e si sono formati nel passato, in seguito a processi geologici, sia profondi (endogeni), che superficiali (esogeni). Per formare dei giacimenti minerari possono essere necessari periodi di tempo piuttosto lunghi, anche dell'ordine del milione di anni. Sono definiti endogeni (Fig- 3):

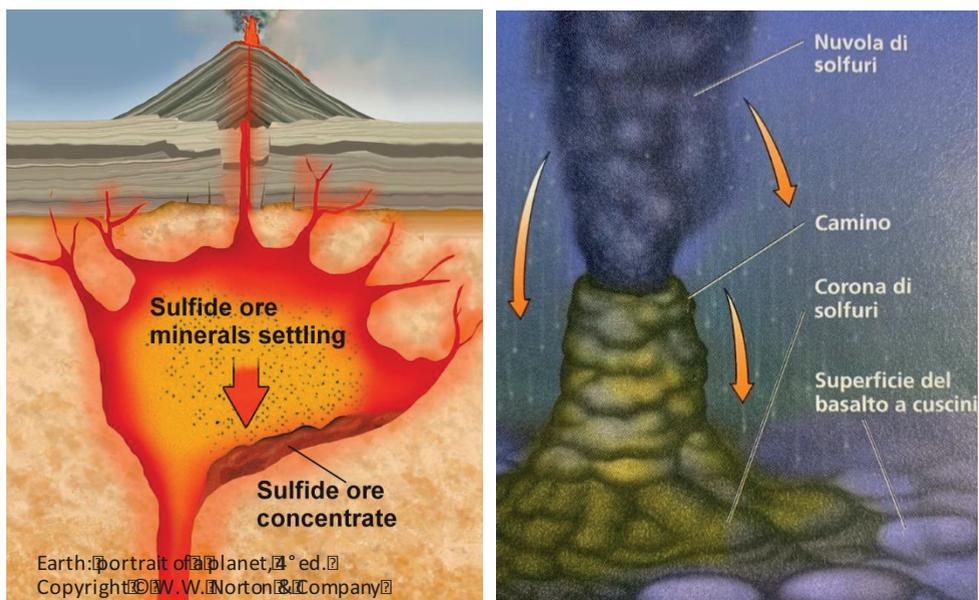


Figura 3. (a sinistra) Rappresentazione schematica di una camera magmatica che, raffreddando lentamente, permette la cristallizzazione di solfuri metallici. (a destra) Concrezioni a forma di “camino” formatesi al di sopra del fondale oceanico in corrispondenza di emissioni sottomarine di fluidi idrotermali.

- I processi magmatici: i minerali di interesse economico si formano e si concentrano durante la cristallizzazione di un magma. Esempi tipici sono i giacimenti di solfuri metallici associati a gabbri.
- I processi idrotermali: i metalli di interesse si trovano disciolti in soluzioni acquose ad alta temperatura (150°-550°C), dette fluidi idrotermali, che possono favorire la precipitazione dei minerali economici se vanno incontro a “saturazione”. La saturazione dei fluidi idrotermali può essere indotta da vari fenomeni (es. raffreddamento, decompressione, etc). I cosiddetti “fumaioli neri” rappresentano un buon esempio di questo tipo di depositi. Si tratta di accumuli

di solfuri metallici che si formano in corrispondenza dei fondali oceanici, in associazione ad emissioni sottomarine di fluidi idrotermali. Le soluzioni acquose, che si immettono nella colonna d'acqua con temperature di circa 350°C, si raffreddano repentinamente a contatto con la fredda acqua che si trova al di sopra del fondale (2°C), si saturano e precipitano i composti chimici che venivano trasportati come “sali disciolti” sotto forma di concrezioni di solfuri minerali.

I processi esogeni che permettono la formazione di giacimenti minerali sono associati all'erosione chimica e fisica di rocce e giacimenti preesistenti (Fig. 4). I ben conosciuti accumuli di “pepite d'oro” nei fiumi del Nord America derivano dall'erosione meccanica di depositi d'oro preesistenti. L'oro può concentrarsi nei sedimenti fluviali perché è stabile chimicamente (ad esempio, non “arrugginisce”) e perché ha un alto peso specifico: pesando di più rispetto ai minerali che costituiscono gli altri sedimenti si accumula nelle zone caratterizzate da “cadute” di corrente, mentre gli altri minerali più leggeri vengono dilavati (Fig. 4).

Per processi di alterazione chimica in ambiente tropicale, si formano invece i giacimenti di alluminio, comunemente detti “bauxiti”. L'alluminio, infatti, può essere recuperato in maniera efficace dalle rocce della crosta terrestre solo se presente in notevoli concentrazioni (più del 60% in peso) e sotto forma di minerali nei quali sia chimicamente legato al solo gruppo –OH. Questi minerali, detti idrossidi, si formano per alterazione chimica (idrolisi) dei più comuni silicati di alluminio (che attualmente non hanno valore economico per il recupero di questo elemento). In particolare, gli idrossidi di alluminio rappresentano i residui dell'alterazione chimica di rocce silicatiche, dalle quali le acque piovane hanno dilavato la silice (Fig. 4). Per questo motivo i giacimenti che si formano sono detti “residuali” (come le analoghe rocce sedimentarie).

Dove si formano i giacimenti minerali

I processi geologici sopra descritti non si sviluppano, e non si sono sviluppati nel passato, in maniera omogenea sulla Terra. Per questo motivo la superficie terrestre ha caratteristiche geologiche particolari e molto variabili da paese a paese. Ne consegue che particolari aree della Terra si presentino più ricche di giacimenti minerali metallici rispetto ad altre aree. Allo stesso tempo, esisteranno territori più ricchi di certi tipi di giacimenti rispetto ad altri.

Aree della crosta terrestre particolarmente adatte alla formazione dei giacimenti minerali sono i margini di placca, sia divergenti che convergenti. Ad esempio, gli accumuli di solfuri sottomarini prima descritti si formano soprattutto in corrispondenza di dorsali oceaniche. Oggigiorno questi accumuli sottomarini non

hanno un reale valore economico poiché non possono essere sfruttati con profitto. Importanti giacimenti di solfuri sottomarini possono essere però ritrovati in rocce di natura oceanica “fossili”, che sono state “trasportate” sulla terra emersa, ad esempio da processi orogenetici (gli stessi che permettono la formazione delle catene montuose).

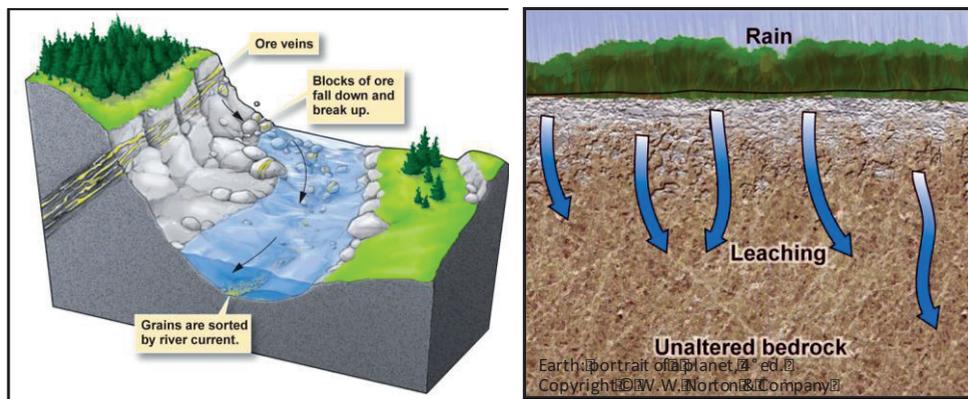


Figura 4. (a sinistra) Modello che mostra come si forma un deposito fluviale di pepite d'oro, per erosione di vene aurifere ipogeniche. (a destra) Processo di alterazione chimica di una roccia silicatica ad opera di acque piovane che penetrano nel suolo. La percolazione di acqua nel sottosuolo dilava la silice e permette l'accumulo residuale degli idrossidi di alluminio.

I più grandi giacimenti di rame della Terra si sono invece formati in corrispondenza di un margine convergente, rappresentato dalla Cordigliera delle Ande, in Cile. Essi sono il prodotto di circolazioni profonde di fluidi idrotermali innescate dal processo di subduzione, e sono strettamente associati alla formazione di catene di stratovulcani molto simili, ad esempio, al Monte Fuji (Giappone) (Fig. 5). Importanti giacimenti d'oro si ritrovano in Nord America poiché, similmente ai giacimenti di rame delle Ande, sono geneticamente associati alla formazione delle Montagne Rocciose.

Altri giacimenti si formano a grandi profondità nella crosta terrestre e possono essere ritrovati solo in aree del pianeta molto antiche, i “cratoni” (= blocchi di crosta continentale più antichi di 2 miliardi di anni). In queste aree, affiorano le “radici” della crosta continentale, che sono state portate in superficie da una perdurata erosione meccanica, che ha smantellato la maggior parte della copertura rocciosa soprastante. Un esempio di questi depositi “profondi” sono i rarissimi giacimenti di platino (Figura 5). I più importanti giacimenti di questo metallo si trovano in Sud

Africa in rocce magmatiche formatesi a vari chilometri di profondità, più di 2 miliardi di anni fa.



Figura 5. (a sinistra) Tipico aspetto di un giacimento di rame delle Ande: vene di solfuri di rame in roccia porfirica (Miniera di El Teniente, Cile). (a destra) Pannello di estrazione in una miniera di platino in Sud Africa (Bushveld).

Questa disomogenea distribuzione dei giacimenti minerari sulla crosta terrestre può generare dei veri e propri “monopoli” geologici delle risorse.

I metalli critici nei giacimenti minerari

Come detto prima, affinché un metallo sia detto “critico”, è necessario che presenti alto valore economico ed alto rischio di approvvigionamento per l’UE. I metalli critici riportati nel diagramma di Figura 1 sono per la gran parte estratti da giacimenti minerari, nei quali però, possono essere variamente distribuiti. Alcuni di essi formano giacimenti “propri”, altri invece sono recuperati da giacimenti di altri metalli tramite processi industriali molto raffinati. Formano giacimenti propri gli elementi del gruppo delle terre rare (abbreviato REE), ed il platino. Il germanio ed il gallio sono invece recuperati come metalli “secondari” da giacimenti di altri metalli, in particolare da depositi di zinco ed alluminio. Queste caratteristiche “geologiche” controllano fortemente la distribuzione “politica” di certi elementi sul pianeta, ed alcuni paesi (es. Cile, Russia, Sud Africa, Cina) sono molto ricchi di particolari metalli semplicemente perché i territori sui quali insistono hanno avuto una storia geologica favorevole alla formazione dei loro giacimenti. Per questi motivi, sono pochi i paesi che forniscono certi metalli all’unione europea (Fig. 6), e da ciò deriva la loro “criticità”. Paesi non “fertili” dal punto di vista geologico, che per vari motivi non vogliono approvvigionarsi da paesi terzi, devono necessariamente ricercare risorse metalliche alternative ai depositi naturali.

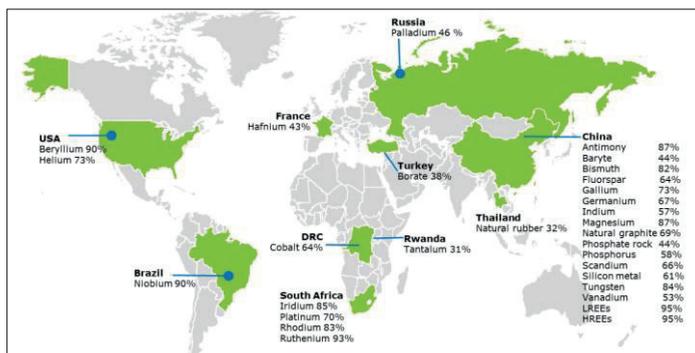


Figura 6. Planisfero politico: in verde sono rappresentati i paesi che forniscono all'UE la maggior parte dei metalli "critici" (Commissione Europea 2017).

Parte 2: Recupero da RAEE - La chimica e la Sostenibilità

(R. Esposito)

La chimica è una scienza e come tale fa parte delle conoscenze umane. Le scienze in particolare sono quelle conoscenze acquisite utilizzando un metodo: il metodo scientifico. Le discipline scientifiche la cui applicazione di questo metodo è, nella pratica fattibile, si definiscono hard sciences e la chimica è una di queste.

Purtroppo, negli anni, il termine "chimica" è stato spesso utilizzato (e lo è tuttora anche dai media nazionali) come sinonimo di tossico. Si pensi a quando si parla di sostanza chimica intesa come sostanza nociva, tossica, in contrapposizione a sostanza naturale.

La causa di ciò è da ricercarsi probabilmente nei disastri ambientali avvenuti negli anni '70/'80 del secolo scorso (si veda ad esempio il disastro di Bhopal o quello di Seveso) in cui si verificarono incidenti a carico di impianti che applicavano alcune conoscenze chimiche. Tali disastri sono avvenuti però a causa di una scarsa conoscenza della chimica e/o ad una sua cattiva applicazione. La chimica, in quanto conoscenza, è uno strumento nelle mani dell'uomo ed in quanto tale può essere utilizzato in maniera corretta oppure no. La colpa di un disastro può essere dunque attribuita ad una branca della conoscenza? In termini più eclatanti: si può dare la colpa alla fisica nucleare per l'esplosione di una bomba atomica? O piuttosto bisogna dare la colpa a chi ha progettato e pensato di utilizzare una tale arma per decimare intere città?

A tal proposito è bene sapere che fu proprio un chimico, Linus Carl Pauling, premio Nobel per la chimica del 1954 a battersi, insieme ad altri scienziati, per il disarmo nucleare durante la guerra fredda. I suoi sforzi gli valsero un secondo Nobel, per la pace stavolta, nel 1962.

Insomma, la chimica soffre di una reputazione che non merita. D'altronde anche la differenza stessa di sostanza chimica o sostanza naturale di fatto non ha senso di esistere. Tutti gli esseri viventi, noi compresi, siamo dei laboratori chimici che producono sostanze chimiche in continuazione. Una molecola d'acqua prodotta da una reazione che avviene nel nostro organismo è identica ad una qualsiasi altra molecola d'acqua presente nell'universo, sia che essa sia prodotta in un laboratorio per mano umana sia che essa sia presente in un fiume o in una cometa. Le sostanze tossiche esistono ma questo possono essere prodotte sia da una pianta che da un essere umano in un laboratorio, anche qui non vi sarà differenza.

Negli anni '90, l'ampliamento della conoscenza chimica e la possibilità che questa potesse essere utilizzata in modo da danneggiare l'uomo o l'ambiente, portò eminenti nomi della chimica mondiale, a porsi il problema etico dell'applicazione di certe conoscenze. Nacquero così i dodici principi della *Green Chemistry* (Anastas, P., Werner, J. 1992), un decalogo di principi che sono tutt'oggi una guida etica e pratica per la progettazione di reazioni ed impianti. Una guida per tutti i fruitori della chimica per far sì che l'applicazione di questa branca della conoscenza avvenga in maniera etica, ecologica e più in generale nell'ottica dello *Sviluppo Sostenibile*.



Figura 7. I dodici principi della Green Chemistry.

L'idea di *Sostenibilità* è nata nel 1969 dalla IUCN, l'Unione Internazionale per la Conservazione della Natura (*International Union for Conservation of Nature*). Questo concetto si è sviluppato nel tempo in diverse conferenze e report internazionali: la Conferenza delle Nazioni Unite sull'Ambiente Umano a Stoccolma del 1972, la Strategia mondiale per la conservazione del 1980, il Rapporto Brundtland del 1987 e la Conferenza delle Nazioni Unite sull'Ambiente e lo Sviluppo di Rio del 1992. È stato coniato esplicitamente per suggerire che era possibile raggiungere la crescita economica e l'industrializzazione senza danni ambientali. La definizione ampiamente utilizzata di sostenibilità, o più precisamente di *Sviluppo Sostenibile*, è: lo sviluppo che "soddisfa i bisogni del presente senza compromettere la capacità delle generazioni future di soddisfare i propri bisogni" (*Our Common Future*, 1987). Ciò non è lontano dal proverbio dei nativi americani: "Non ereditiamo la terra dai nostri antenati; lo prendiamo in prestito dai nostri figli". Una descrizione grafica efficace (Fig. 8), mostra come la sostenibilità si fonda su tre pilastri. Solo quando i bisogni sociali, economici e ambientali sono soddisfatti si ha uno *Sviluppo Sostenibile*.

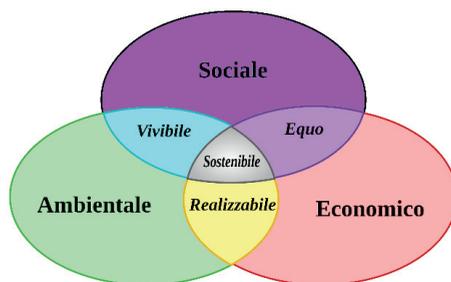


Figura 8. Diagramma di Vienn dello Sviluppo Sostenibile.

L'applicazione etica della chimica è di fondamentale importanza dal momento che questa materia ricopre un ruolo fondamentale nella vita di tutti i giorni. La maggior parte degli oggetti, degli articoli che ci circondano, hanno visto in un momento della loro produzione un'applicazione della chimica. Dai farmaci ai cosmetici, dai vestiti agli apparecchi elettronici, tutto è in qualche modo applicazione della conoscenza chimica e ancor di più, tutto è costituito di materia, quindi di sostanze chimiche ed elementi.

L'importanza della conoscenza di questa materia diviene ancora più evidente se si considera che con l'avanzamento della tecnologia, i prodotti ed i dispositivi, inclusi quelli che utilizziamo nella vita quotidiana, sono diventati così sofisticati da richiedere che le varie componenti siano costituite da un numero crescente di

sostanze ed elementi. Se ad esempio un mulino del 1700 poteva essere costruito mediante l'utilizzo di meno di una decina di elementi, tutti molto abbondanti, oggi invece, una turbina eolica richiede l'utilizzo di diverse decine di elementi differenti, laddove alcuni di essi sono anche molto rari.

Questo pone due problemi ambientali. Il primo è che le risorse del pianeta sono limitate e prima o poi, l'approvvigionamento di tali materiali non potrà tenere il passo con la domanda. Il secondo problema riguarda lo smaltimento di questi dispositivi che consiste in primo luogo al fatto in sé di dover smaltire volumi sempre maggiori di rifiuti, dato il nostro crescente utilizzo, ed in secondo luogo al fatto che questi rifiuti possono essere altamente inquinanti.

La soluzione a questi problemi può essere raggiunta solo abbracciando il concetto di *Economia Circolare*. Questo tipo di economia è nato in contrapposizione al modello di economia lineare adottato finora ed è definito come "un sistema rigenerativo in cui l'input di risorse e gli sprechi, le emissioni e la dispersione di energia sono ridotti al minimo rallentando, chiudendo e restringendo i circuiti di materie prime ed energia; ciò può essere ottenuto mediante progettazione di articoli duraturi, manutenzione, riparazione, riutilizzo e rigenerazione, ristrutturazione e riciclo" (Geissdoerfer *et al.*, 2017).



Figura 9. Rappresentazione grafica dell'Economia Circolare.

Riciclare i RAEE

Recuperare e riciclare i prodotti più o meno tecnologici è dunque una priorità. Ma non sempre riciclare è semplice. Questo è dovuto all'esistenza dell'entropia, la forza della natura che fa tendere tutto al mescolamento e al caos. L'esistenza dell'entropia, e quindi del secondo principio della termodinamica, impedisce che qualsiasi operazione di riciclo abbia un rendimento del 100%. In Figura 9 è infatti possibile

notare una freccia “residual waste” che indica lo scarto residuo. Questo non è azzerabile ma è minimizzabile sempre più, aumentando l’efficienza del processo. Per tendere alla massima efficienza è necessario perfezionare ed ottimizzare le procedure acquisendo sempre più conoscenze tramite la ricerca.

Tra gli scarti tecnologici più significativi della nostra epoca vi sono i RAEE ovvero i Rifiuti di Apparecchiature Elettriche ed Elettroniche (in inglese *WEEE - Waste Electrical and Electronic Equipment*). Questi dispositivi sono generalmente articoli molto sofisticati in cui sono presenti, in miscela, molti elementi spesso rari e non sempre semplici da separare tra loro.



Figura 10. Rifiuti di Apparecchiature Elettriche ed Elettroniche (RAEE)

Ad esempio, gli smartphone che utilizziamo tutti i giorni, sono a loro volta un assemblaggio di sofisticate componenti ognuna delle quali può contenere molti elementi. Basti pensare che degli 83 elementi stabili della tavola periodica, in uno smartphone ne sono rappresentati ben 62. Molti di questi sono metalli e alcuni di questi sono molto rari e preziosi. Alcuni di questi sono addirittura i cosiddetti metalli delle terre rare (*Rare Earth Elements - REE* da non confondere con RAEE) come il praseodimio (Pr) e il disprosio (Dy).

In aggiunta a ciò, il contenuto di metalli preziosi in alcuni dispositivi tecnologici è addirittura superiore a quello delle rocce ignee o dei minerali da cui questi sono estratti. Prendendo come esempio un metallo prezioso noto a tutti, come l’oro, si stima che nei telefoni cellulari questo può essere presente in concentrazioni dalle 1000 alle 2000 volte superiori rispetto alle rocce ignee da cui viene comunemente estratto. Circa 10 kg di telefoni cellulari, infatti, possono contenere da 1.5 g a 4.0 g di oro.

I RAEE sono quindi un concentrato di metalli preziosi e l’oro non è l’unico, anche se è spesso il metallo che influisce di più sul valore in materie prime del dispositivo stesso (Figura 11) (Hagelüken, 2007; Cui e Zhang, 2008; Tuncuk *et al.*, 2012).

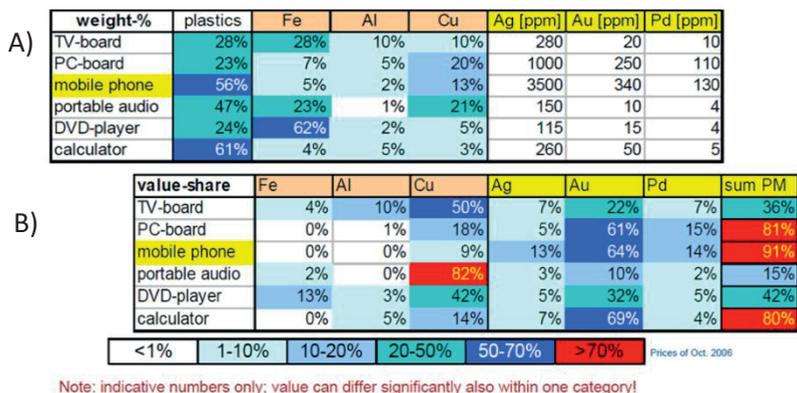


Figura 11. A) Contenuto in peso di alcuni metalli in alcuni tipi di RAEE. B) Percentuale del valore (value-share) in materie prime dovute alla presenza di un dato metallo in un dato tipo di RAEE (sum PM è la somma dei metalli preziosi) (C. Hageliiken 2007).

Riciclare i RAEE quindi, oltre ad essere vantaggioso da un punto di vista ambientale, può esserlo anche dal punto di vista economico, andando così a soddisfare due delle tre condizioni affinché possa esserci lo *Sviluppo Sostenibile*.

Per riciclare i RAEE è possibile utilizzare due tipi di tecnologie. La prima è la pirometallurgia, che prevede l'utilizzo di fornaci, temperature estremamente elevate, alti costi energetici e impianti molto costosi e sofisticati. L'utilizzo di questa tecnologia prevede che i RAEE vengano bruciati per eliminare la plastica mentre successivamente i metalli vengono per lo più separati mediante fusione. Esistono alcuni impianti molto efficienti nel recupero di metalli critici e metalli preziosi ed uno di questi è sito ad Anversa, in Belgio, ed è uno dei più importanti impianti per il riciclo di RAEE in Europa e nel mondo (<https://pmr.unicore.com/en/about-us/>). Tuttavia impianti di questo tipo richiedono investimenti elevati, dell'ordine dei miliardi di euro.

Un'altra tecnologia è invece l'idrometallurgia (Syied, 2012). Tramite questa è possibile recuperare i metalli mediante una serie di processi chimici che vanno dallo scioglimento in soluzione tramite ossidazione, al recupero tramite estrazione e infine all'ottenimento dell'elemento in forma metallica tramite riduzione. Alcune procedure prevedono una ibridazione delle due tecniche precedenti ed in particolare, generalmente si utilizza la pirometallurgia per una prima scrematura e l'idrometallurgia per un recupero selettivo di alcuni metalli.

Il riciclo di metalli da scarti tramite idrometallurgia prevede un primo step di separazione delle componenti meccaniche (parti in plastica, ferro) che può essere adoperata con tecniche moderne di automazione.

Questo prima separazione fisica è seguita dal *leaching*, che è appunto lo step che permette il passaggio dei metalli in una soluzione acquosa. Sostanzialmente il *leaching* prevede una reazione di ossidazione del metallo. Ossidare un metallo vuol dire modificare la sua natura chimica cosicché questo non sia più nella sua forma elementare ma in una forma ossidata che in quanto tale può essere disciolta in una soluzione acquosa.

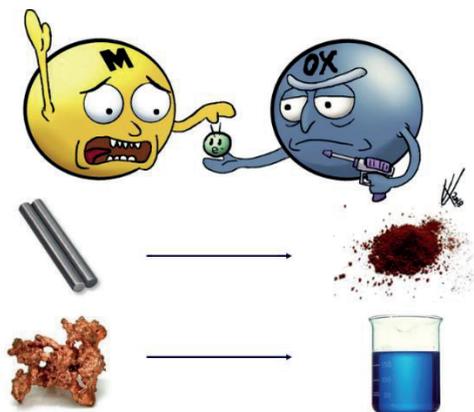


Figura 12. L'ossidazione dei metalli ed alcuni esempi.

L'estrazione di metalli tramite *leaching* è un processo di vecchia data che prevede l'utilizzo di una sostanza, detta ossidante, in grado di ossidare il metallo. Processi del secolo scorso prevedevano, però, quale agente ossidante, il cianuro. Purtroppo, questa tecnica viene utilizzata ancora oggi nei paesi più poveri per estrarre oro da RAEE con metodi dozzinali, pericolosi ed inquinanti,

Un ossidante alternativo, adottato più di recente (ma anch'esso conosciuto in antichità) è l'acqua regia. Tuttavia, l'utilizzo di questa sostanza non è privo di problematiche ambientali. La ricerca in ambito chimico ha portato alla scoperta, negli ultimi anni, di ossidanti molto più green quali tiourea, tiocianato o ipoclorito, solo per citarne alcuni (Syied, 2012).

Una volta disciolti i metalli e passati in soluzione acquosa, è possibile sequestrarli da questa e portarli in un'altra soluzione in cui il solvente, detto *solvente estraente*, è immiscibile con l'acqua. Questa tecnica è chiamata *estrazione*. Per fare ciò si utilizzano delle molecole, chiamate leganti, che agiscono come delle "pinzette" molecolari (Figura 13).

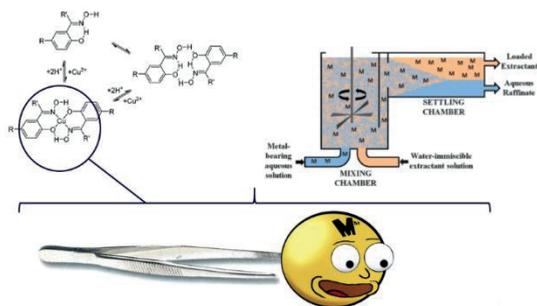


Figura 13. Separazione dei metalli mediante estrazione liquido/liquido.

Tali “pinzette” sono presenti nel *solvente estraente* che messo in contatto con la soluzione acquosa contenente i metalli, sono in grado di sequestrare selettivamente uno specifico metallo. I due liquidi vengono dunque separati (in quanto immiscibili è sufficiente interrompere l’agitazione) e la soluzione acquosa viene veicolata verso altre estrazioni per recuperare altri metalli, mentre il *solvente estraente*, arricchito del metallo selezionato viene portato verso l’ultimo step.

Il processo finale consiste nella *riduzione* (Figura 14) che è appunto l’esatto opposto della ossidazione (*leaching*) vista in precedenza. In questo step, il metallo viene riportato nella sua forma elementare mediante l’utilizzo di un *agente riducente*. Qui la scelta è molto vasta, e molti agenti riducenti di vecchia data sono tutt’ora utilizzati in quanto ampiamente disponibili e dal profilo eco-tossicologico molto vantaggioso (si veda ad esempio il solfato ferroso).

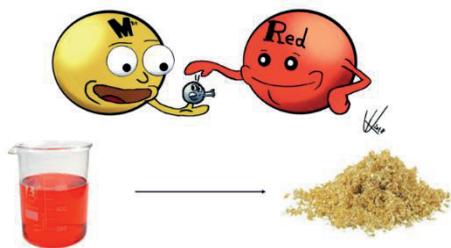


Figura 14. Riduzione del metallo.

L’efficienza e la sostenibilità dell’intero processo di riciclo dei metalli dai RAEE dipendono da ogni singolo step appena visto e dalle sostanze utilizzate nelle varie fasi. La scelta dell’ossidante, del solvente estraente e dei leganti (le “pinzette”) e della loro efficienza sono tutti fattori che impattano sul rendimento e la sostenibilità economica e ambientale del processo. Grazie alla ricerca si stanno facendo progressi in ognuno di questi campi. A titolo di esempio si segnala che, nel 2009, è stato

scoperto (Kawakita *et al.*, 2009) che una di queste pinzette molecolari in grado di sequestrare selettivamente l'oro da soluzioni acquose è presente nelle bucce d'arancia. Aderendo perfettamente al concetto di *Economia Circolare*, il gruppo di ricerca giapponese ha quindi proposto di utilizzarle nell'*estrazione* di questo metallo, aumentando la sostenibilità di questo step, in quanto facente utilizzo di una materia prima recuperabile dallo scarto della filiera alimentare.

In conclusione, il riciclo dei RAEE è una priorità dato il largo consumo che l'essere umano fa dei dispositivi elettrici ed elettronici. D'altra parte, questo è non solo eticamente ed ambientalmente corretto ma può anche essere economicamente vantaggioso. Questo rientra perfettamente nel concetto di *Sviluppo Sostenibile*.

Per aumentare la sostenibilità c'è bisogno di costante ricerca e innovazione tecnologica e queste vanno supportate e finanziate. Solo progredendo nella conoscenza della chimica (e nelle scienze in generale) riusciremo a far tendere a zero gli scarti residui delle nostre attività e a dirigerci verso un futuro più sostenibile.

Bibliografia

- Anastas, P., Warner, J. (1992). I dodici principi della Green Chemistry. <https://www.acs.org/content/acs/en/greenchemistry/principles/12-principles-of-green-chemistry.html>.
- Cui, J., Zhang, Z. (2008). Metallurgical recovery of metals from electronic waste: A review. *Journal of Hazardous Materials*, 158, 228–256 e bibliografia al suo interno.
- Geissdoerfer, M. *et al.* (2017). The Circular Economy e A new sustainability paradigm? *Journal of Cleaner Production*, 143, 757-768.
- Hagelüken, C. (2007) 6th session of OEWG Basel convention.
- Impianto Umicore, Antwerpen (Anversa), Belgio. <https://pmr.umicore.com/en/about-us/>.
- Kawakita, H. *et al.* (2009). Selective Gold Recovery Using Orange Waste. *Separation Science and Technology*, 44(12), 2797-2805.
- *Our Common Future* conosciuto anche come Brundtland report, Report of the World Commission on Environment and Development (1987). <https://archive.org/details/ourcommonfuture00worl/page/n7/mode/2up>.
- Syed, S. (2012). Recovery of gold from secondary sources—A review. *Hydrometallurgy*, 115-116, 30–51.
- Tuncuk *et al.* (2012). Aqueous metal recovery techniques from e-scrap: Hydrometallurgy in recycling. *Minerals Engineering*, 25, 28–37.

Un approccio scientifico al tema del rapporto tra uomo e radiazioni ionizzanti

Lorenzo Manti

Dipartimento di Fisica “E. Pancini”, Università di Napoli Federico II

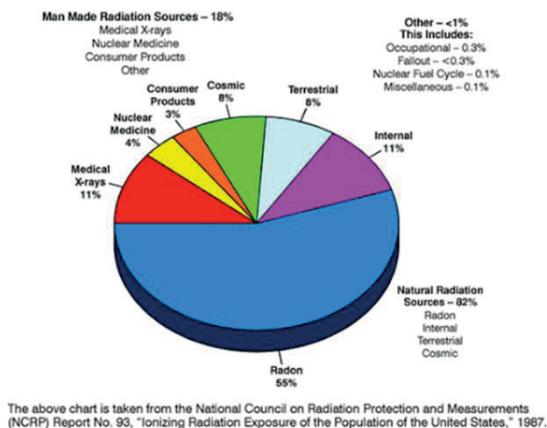
Introduzione

Il tema dell'esposizione umana alla radiazione ionizzante (RI) è emblematico dell'importanza che una corretta divulgazione assume nel trattare questioni che interessano un ampio pubblico per le loro ricadute ed implicazioni sulla salute. Infatti, la percezione che l'opinione pubblica ha dei rischi che derivano dalla RI è, di norma, negativa poiché tende a prevalere una visione “radiofobica”, in molti casi ingiustificata, che non tiene conto, razionalmente, degli indubbi benefici che l'uomo ha e continua a ricevere dall'uso delle RI nei campi della diagnostica e della terapia antitumorale. Sicuramente, la RI rappresenta uno dei più potenti e meglio studiati agenti carcinogenici ma rappresenta al contempo un formidabile strumento per la diagnostica precoce e la cura dei tumori. Questa dicotomia è alla base del rapporto spesso contraddittorio verso le RI, un rapporto che è peraltro cambiato nel tempo. In questo intervento, si è cercato di fornire una panoramica “science-based” dell'azione a livello biologico della RI, illustrando come gli eventi fisico-chimici messi in moto dall'assorbimento della radiazione nei tessuti biologici portino ad alterazioni a livello molecolare che sono poi misurabili e quantificabili, perché il passaggio da beneficio a pericolosità della RI è fortemente dipendente, tra l'altro, da caratteristiche fisiche quali la qualità della radiazione e la dose. La speranza è che il messaggio recepito sia quello non avere preconcetti verso la RI ma di saperne valutare il rapporto rischio-beneficio sulla base delle evidenze sperimentali a disposizione.

L'esposizione umana alla radiazione ed i principali effetti biologici

Essendosi la vita evoluta in un ambiente naturalmente esposto alle RI, le cellule hanno sviluppato sofisticati ed efficienti meccanismi di riparo del danno alla biomacromolecola che per definizione è il target sensibile, ossia il DNA. Ciò non perché sia l'unica molecola ad essere danneggiata dall'azione della RI ma perché ne esiste un'unica copia, a differenza di tutti gli altri organelli e delle innumerevoli proteine che presiedono alle funzioni della cellula. Come mostrato in Figura 1A, il fondo di radiazione naturale contribuisce per circa l'80% all'esposizione umana. La grandezza che quantifica la quantità di radiazione assorbita in una massa di materiale

(biologico) è la dose, definita come l'energia di 1J per 1 kg di materia e ciò rappresenta 1 Gray (Gy); a fini radioprotezionistici, la dose è espressa in termini di dose equivalente, per tener conto di vari fattori quali, e la sua unità di misura è il Sievert (Sv). Sostanzialmente Gy e Sv vengono usati in maniera interscambiabile. Mediamente, sulla Terra, ogni individuo assorbe circa 3,6 mSv (10^{-3} Sv) all'anno, cioè una dose piccolissima se paragonata a quella usata nella radioterapia dei tumori (nell'ordine dei Gy) o nei CT



BACKGROUND RADIATION IN THE WORLD

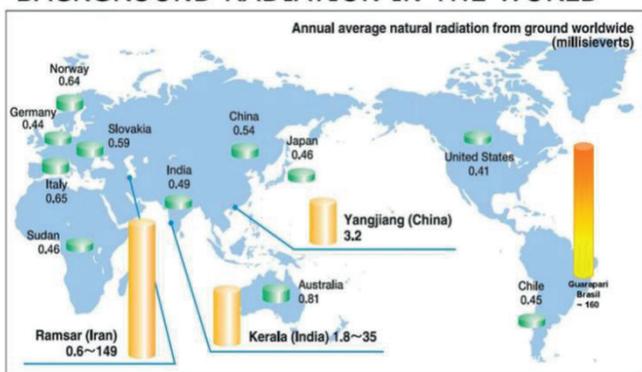


Fig. 1. in alto) Contributo di varie fonti all'esposizione umana; in basso) Esistenza di regioni dove la radiazione di fondo è estremamente più alta della media.

Scan (TAC), che sono nell'ordine dei cGy (10^{-2} Gy). Tuttavia, questo fondo di radiazione non è uniforme: esistono delle aree (Fig. 1B) dove tale esposizione è ordini di grandezza superiore (HBRA, High Background Radiation Areas). Studi di natura epidemiologica condotti sulle popolazioni delle HBRA hanno evidenziato l'assenza di una maggiore incidenza di malformazioni, di tumori e di altri effetti che

sono invece tipici dell'esposizione alle RI. Questa osservazione permette di iniziare, almeno in maniera semplice, ad inquadrare il problema: la dipendenza dalla dose dell'effetto biologico radioindotto e la natura stocastica di certi effetti, quali la carcinogenesi radioindotta. E' infatti abbastanza intuitivo accettare che maggiore è la dose ricevuta, maggiore sarà l'effetto che ne risulta: se espongo alla RI una popolazione cellulare in esperimenti in laboratorio, posso costruire una cosiddetta curva dose-risposta (Fig. 2A) scegliendo uno dei tanti effetti cui la RI dà luogo, per esempio la morte cellulare, intesa come la perdita da parte della cellula irraggiata di proliferare. Si vede come l'entità dell'effetto, ossia la frazione di cellule che "sopravvivono" diminuisca al crescere della dose. Nella stessa figura sono in realtà rappresentate due curve ad illustrare il concetto che le RI non sono tutte uguali dal punto di vista della capacità di danneggiare il DNA e quindi di produrre un effetto di una certa magnitudine.

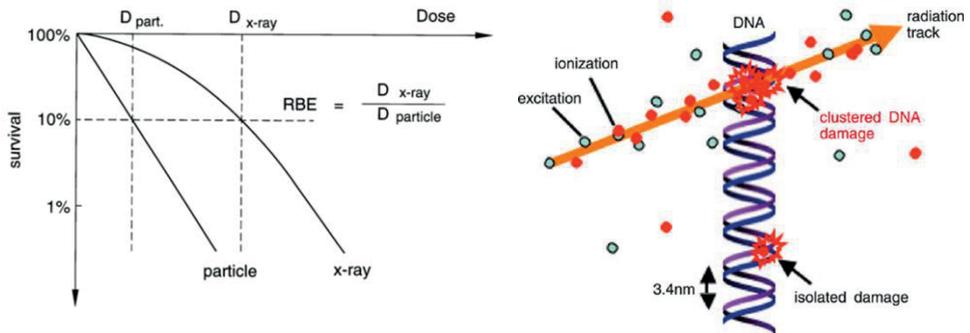


Fig. 2. A sinistra, esempio di curva dose-risposta, per quantificare la morte cellulare radioindotta. A destra, al crescere della massa della particella, la radiazione causa danni spazio-temporalmente ravvicinati (cluster) ed aumenta la sua efficacia nel danneggiare il DNA.

Ciò dipende dalle proprietà fisiche legate alla densità di ionizzazione: fotoni (raggi X e γ) sono radiazioni sparsamente ionizzanti, mentre la radiazione corpuscolata (ioni) deposita la sua energia con una maggiore densità di eventi di ionizzazione, che risultano in cluster di danni al DNA, più difficilmente riparabili dalla cellula, e quindi ad una sua maggiore efficacia a parità della dose (Fig. 2B). La morte cellulare rappresenta il goal della radioterapia, ed è quindi ciò che si vuole ottenere nelle cellule tumorali. Tuttavia, proprio a causa dei fenomeni di riparo di cui sopra, a dosi più basse, una cellula può sopravvivere nel senso che può continuare a proliferare ma può recare con sé dei danni, cioè delle alterazioni a livello della sequenza genica, sotto forma per esempio di aberrazioni cromosomiche, che dipendono dal cattivo riparo della rottura della doppia elica di DNA. Queste cellule sono quelle più

“pericolose” perché potendo proliferare e così trasmettere queste “alterazioni” alla propria progenie, possono innescare fenomeni di carcinogenesi. Il “può” è la parola chiave: la carcinogenesi radioindotta è infatti un tipico effetto stocastico, ossia per il quale, a differenza di quanto mostrato in Figura 2A, non è possibile tracciare una curva che con esattezza mi fornisca un livello di effetto per la dose ricevuta, ma al più stimare un rischio che tale evento accada. La Figura 3 raffigura il cosiddetto modello Linear-No-Threshold (LNT model) che è il paradigma della radioprotezione e sostanzialmente stabilisce che anche alla dose più piccola misurabile, il *rischio* di induzione di tumore è non nullo. Non ci dice che alla dose x corrisponde una quantità y di tumori indotta, ma solo la probabilità che per tale dose il rischio che l’individuo esposto sviluppi un tumore aumenti rispetto alla probabilità che quel tumore si sviluppi naturalmente, cioè senza esposizione. Ed è esattamente questo concetto che “spaventa”, ossia la mancanza di una certezza ed il fatto che anche dosi piccolissime, tipo quelle di una radiografia o di una TAC aumentino il rischio di sviluppare tumori. O l’esposizione alla radiazione di fondo, che come abbiamo visto è diversa da zero. Ma, allora, perché gli abitanti delle HBRA (Fig. 1B) non presentano una maggiore incidenza di tumori?

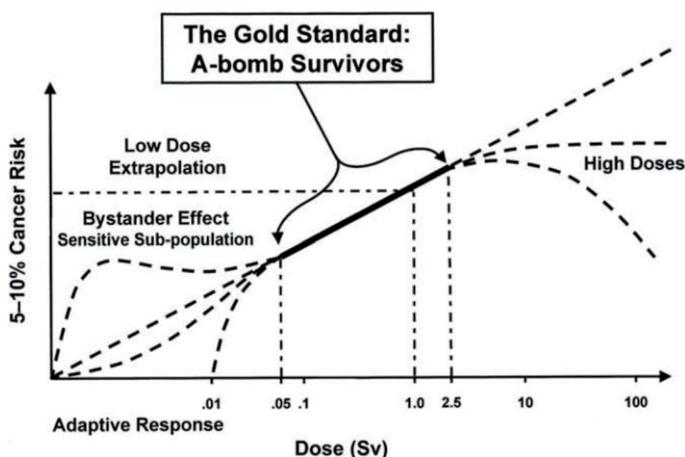


Fig.3: L’ LNT model: la linea in grassetto rappresenta i dati sull’incremento di rischio di tumore in funzione della dose ottenuti dagli studi epidemiologici sui sopravvissuti all’atomica di Hiroshima e Nagasaki: alle basse dosi, tale andamento lineare è stato estrapolato linearmente, ossia una relazione lineare è stata assunta valida anche a dosi e ratei di dose bassissimi.

L’estrapolazione lineare alle bassissime dosi risponde al principio radioprotezionistico di massima cautela: dati scientifici certi sul rischio alle dosi bassissime sono ancora incerti. Infatti, le HBRA rappresentano un esempio della

cosiddetta risposta adattativa, secondo cui una cronica esposizione alle radiazioni non sarebbe in realtà tanto nociva quanto predetto dal modello. Esistono in realtà anche altri effetti radiobiologici, quali il *bystander effect*, secondo cui cellule non direttamente colpite dalla radiazione, ossia in cui non c'è stato fisicamente deposizione di energia, esprimono danni. Questo effetto avrebbe un impatto di natura opposta sull'andamento del rischio descritto in Figura 3.

L'uso terapeutico della radiazione ionizzante

Il 50% circa dei pazienti di tumori nel mondo vengono prima o poi sottoposti ad una qualche forma di radioterapia (RT). Come accennato sopra, il problema maggiore in RT è massimizzare il rilascio della dose a carico delle cellule tumorali e minimizzare il danno ai tessuti sani che vengono inevitabilmente esposti se il tumore trattato è profondo. Questo concetto si traduce nella pratica clinica nella cosiddetta finestra terapeutica, ossia l'intervallo di dose utile affinché la probabilità di cura sia la più elevata possibile compatibilmente con un rischio di complicazioni ai tessuti sani (Fig. 4A).

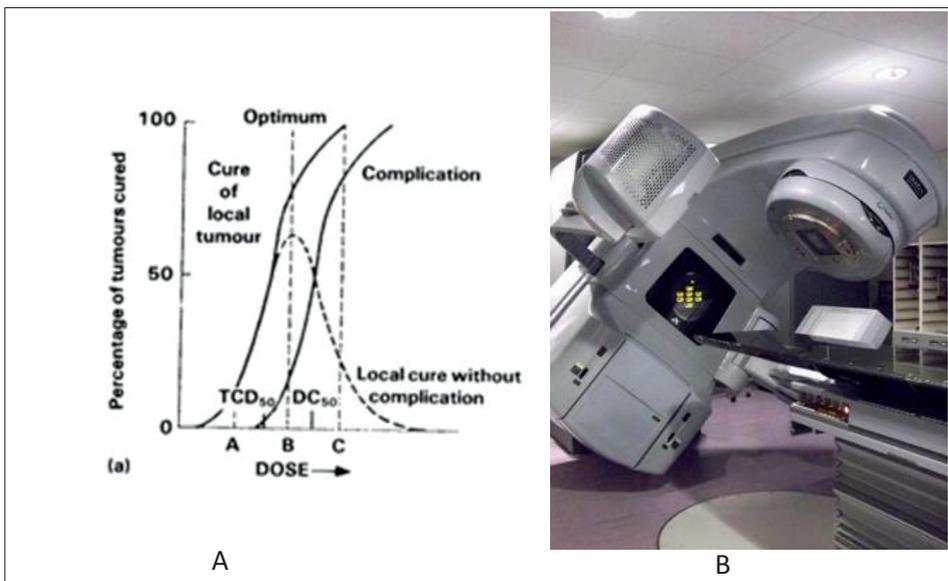
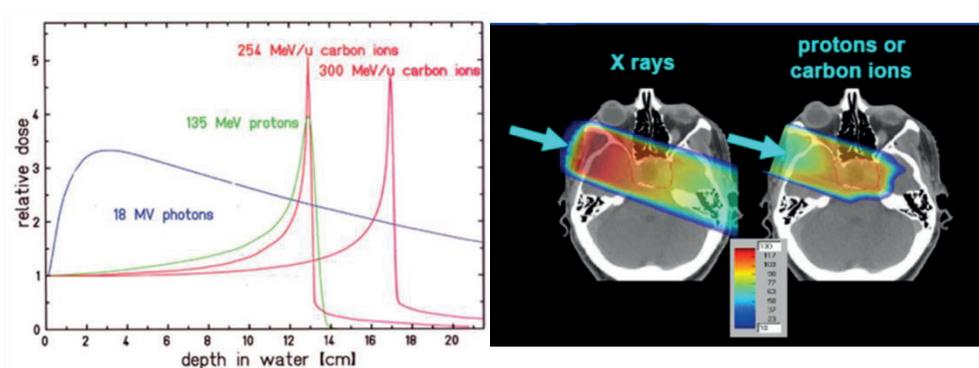


Fig. 4. A) Tipiche sigmoidi relative alla Tumour Control Probability (TCP) ed alla Normal Tissue Complication Probability (NTCP); B) un esempio di un moderno linac per RT con fotoni ad elevata energia

La RT è tipicamente realizzata con fotoni ad alta energia mediante macchine relativamente compatte (Fig. 4B) e gli sviluppi tecnologici hanno consentito enormi progressi nella capacità di conformare la dose al tumore in maniera sempre più

precisa. Tuttavia, la fisica ha fornito la possibilità di realizzare una RT in linea di principio ancora più precisa ed efficace, sfruttando le leggi che regolano la perdita di energia nella materia dei fasci di particelle cariche accelerati da macchine quali sincrotroni e ciclotroni. Questa forma di trattamento tumorale con fascie accelerate di particelle prende il nome di adroterapia. La Figura 5 illustra il principio dell'adroterapia: sono raffigurati i profili di dose dei fotoni e la cosiddetta curva di Bragg, che invece caratterizza il rilascio della dose da parte delle particelle cariche: i fotoni sono attenuati e quindi penetrano tutto il corpo, rilasciando una dose maggiore a profondità relativamente piccole; al contrario, le particelle cariche rilasciano la massima dose a fine percorso. La posizione del picco di Bragg all'interno del corpo del paziente può essere "decisa" da chi opera la macchina acceleratrice dal momento che è funzione dell'energia iniziale e quindi può essere calcolata con precisione. Inoltre, il cosiddetto profilo inverso dose-profondità della curva di Bragg chiaramente favorisce un maggior risparmio (*dose sparing*) al tessuto sano mentre l'arresto a fine percorso delle particelle permette ugualmente di non rilasciare dose a tessuti/organi sani posti "dietro" al tumore. Questo è fondamentale, per esempio, nel trattamento dei tumori pediatrici, dove il paziente curato avrà di fronte a sé una lunga aspettativa di vita e pertanto è fondamentale ridurre il rischio di danni a lungo termine derivanti dall'irraggiamento non necessari di tessuti sani

Fig. 5. A sinistra i profili di dose dei fotoni comunemente usati in RT e quelli delle particelle



(protoni e ioni carbonio,) usati in adroterapia; a destra un piano di trattamento mediante il quale il fisico medico mediante opportuni programmi computerizzati definisce i contorni di dose ai vari organi/tumori

Essere precisi nella cura dei tumori con la radiazione non basta: a volte, purtroppo, i trattamenti falliscono a causa della cosiddetta radioresistenza delle cellule tumorali. Ecco che di nuovo le proprietà radiobiologiche oltre che fisiche delle diverse qualità di radiazioni di cui abbiamo parlato prima vengono in aiuto. Come descritto in Figura

2A, al crescere della massa della particella aumenta la sua efficacia biologica perché aumenta la capacità di indurre più danni letali alla cellula. L'uso dei fasci di ioni carbonio, più pesanti dei protoni, serve proprio a questo: infatti, grazie a tali fasci, possono essere oggi curati con migliori prospettive tumori rispetto al trattamento con RT convenzionale.

Conclusioni

L'uomo deve convivere con l'esposizione alla radiazione ionizzante. In parte la Natura ha fornito alle nostre cellule efficaci mezzi di difesa dalla sua azione deleteria; d'altro lato, gli studi compiuti in vari campi hanno permesso in misura sempre crescente di comprendere i meccanismi con cui essa agisce e di sfruttarne al meglio le potenzialità benefiche. In ogni caso, è solo affidandosi al metodo scientifico, che si è in grado di giudicare in che misura ed in che ambito tali effetti.

Seminario integrato¹

Fluorescenza: dal fotone ai materiali e alle proteine reporter

Parte 1: Materiali fluorescenti

Ugo Caruso

Dipartimento di Scienze Chimiche, Università di Napoli Federico II

Abstract

I materiali fluorescenti fanno parte della vita quotidiana. Gli evidenziatori, con cui tutti noi sottolineiamo le parti importanti di uno scritto, i giubbotti di sicurezza presenti in ogni auto e finanche le palline da tennis sono tinte con vernici fluorescenti. Ma ci siamo mai chiesti a cosa è dovuto questo fenomeno? Sappiamo come si progetta un colorante fluorescente? Conosciamo applicazioni meno comuni ma, forse, più all'avanguardia. Questi sono gli obiettivi che si pone questo breve compendio, non esaustivo ma ricco di spunti di riflessione ed utile per eventuali approfondimenti.

Introduzione

I sinergici contributi di più discipline scientifiche trovano spesso applicazione nella progettazione di nuovi materiali. L'approccio multidisciplinare alla base della Scienza dei Materiali è frutto della moderna visione delle Scienze in cui la dovuta separazione specialistica delle discipline si ricongiunge in gruppi di lavoro in cui ogni studioso apporta il suo contributo originale e in cui una figura essenziale, spesso di coordinamento, è quella dello studioso della Scienza dei Materiali.

Un approccio multidisciplinare è quello legato alla progettazione di nuovi materiali fluorescenti ed il loro utilizzo in applicazioni sempre più specialistiche. L'uso di materiali fluorescenti trova oggi sempre più vasta applicazione in usi comuni, ma anche in campi quali sensori ambientali, imaging molecolare e in dispositivi elettrotici.

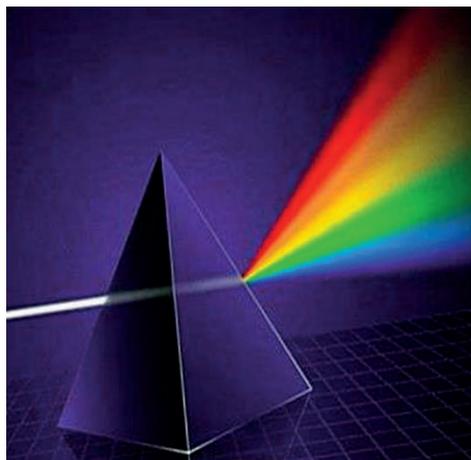
Nelle pagine seguenti tenteremo di analizzare le cause della fluorescenza, individuare la struttura molecolare di un sistema cromoforico e vederne qualche applicazione non usuale.

¹La Parte introduttiva sulle transizioni elettroniche e la fluorescenza, tenuta dal prof. Scotti di Uccio, è stata inserita nel contributo "La fisica dell'atomo di Idrogeno"

Colori e Fluorescenza

Un tratto di evidenziatore sul nostro scritto è molto più brillante di un tratto di pennarello dello stesso colore. Perché? A cosa è dovuto il fenomeno della fluorescenza, a cosa quello del colore? Rispondendo a queste domande saremo un po' più consapevoli del fenomeno della fluorescenza e delle sue possibili applicazioni.

Sin dai nostri primi studi, abbiamo imparato che un oggetto appare colorato perché, irradiato dalla luce solare, il cui spettro è un coacervo di vari colori, assorbe quasi tutti i colori restituendo, o più correttamente riflettendo, solo il colore che noi percepiamo. Andando più avanti con gli studi abbiamo imparato che la luce del sole, che ci appare bianca, è formata da radiazioni elettromagnetiche comprese tra i 380 nm e i 780 nm (ovvero tra i 428 THz e i 749 THz, esprimendola in frequenze). Lo spettrofotometro di cui tutti siamo corredati, il nostro occhio, è costruito per percepire solo questa finestra, di quella ben più ampia di radiazioni elettromagnetiche, individuando i vari tipi di onde elettromagnetiche come colori. Quindi un oggetto ci appare colorato di giallo perché assorbe tutte le onde elettromagnetiche ad eccezione di quelle con frequenza compresa in un intervallo tra i 508-526 THz (565-590 nm).



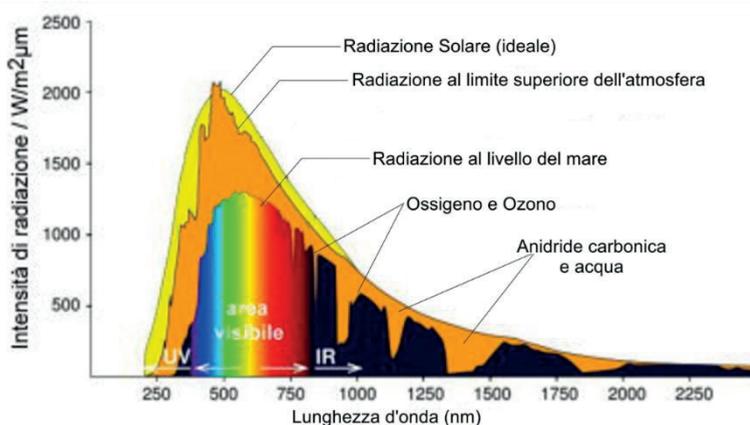
<i>Colore</i>	<i>Frequenza</i>	<i>Lunghezza d'onda</i>
Violetto	668-789 THz	380-435 nm
Indaco	631-668 THz	435-500 nm
Blu	606-631 THz	500-520 nm
Verde	526-606 THz	520-565 nm
Giallo	508-526 THz	565-590 nm
Arancio	484-508 THz	590-625 nm
Rosso	400-484 THz	625-740 nm

Ma sappiamo bene, dalla legge di Plank ($E = h\nu$) che ad ogni radiazione è associata un'energia e che, quindi, se il nostro oggetto colorato trattiene tutti i colori ad eccezione di quello che percepiamo sta assorbendo energia. Del resto, al di là del formalismo matematico, tutti noi abbiamo verificato "a pelle" tale affermazione. Il tepore che ci deriva dall'esporsi al sole in una giornata primaverile, il volante

bollente della nostra auto dopo una lunga esposizione al sole sono tutte prove sperimentali della correttezza delle asserzioni fatte.

Sarebbe lecito domandarsi che fine fa l'energia assorbita. Questa viene riceduta al mondo circostante in vari modi, il più semplice è dissipandone una parte o tutta in calore e riscaldando l'ambiente circostante, o comunque scambiandola sotto tutte le forme disponibili; si pensi, ad esempio, ai pannelli fotovoltaici, alle cellule di una pianta (fotosintesi clorofilliana), ecc. Quel che succede comunemente è che una parte dell'energia assorbita viene dissipata in calore e solo una frazione viene, poi, riceduta, in forma diversa. La diretta conseguenza di quanto affermato è che, fatta eccezione per il bianco perfetto, tutte le sostanze colorate appaiono meno luminose della luce che le investe, ovvero una parte della radiazione luminosa è stata assorbita dal nostro corpo.

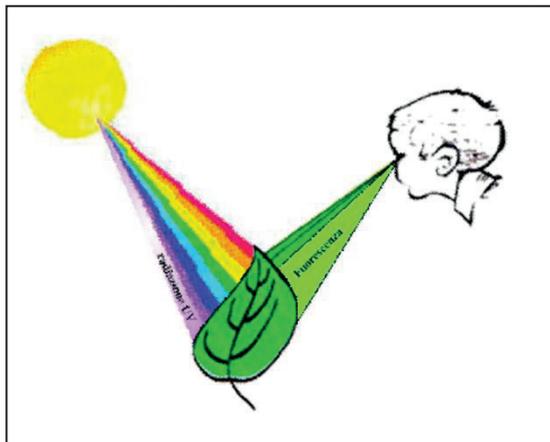
Al contrario, un'oggetto fluorescente appare di un colore più intenso della luce che lo investe. È, quasi, come se un emettesse luce propria. Per spiegare tale fenomeno dobbiamo estendere il nostro concetto di radiazione luminosa. Finora ci siamo soffermati solo su quella piccola parte dello spettro elettromagnetico emesso dal sole che il nostro occhio percepisce. Nella realtà il nostro sole emette una radiazione ben più ampia di quella legata al visibile, anche se sulla terra, per nostra fortuna, ne arriva solo una piccola parte. Buona parte viene assorbita dall'atmosfera che circonda il nostro pianeta. In particolare, la nostra atmosfera filtra i raggi fortemente energetici UVC (100-280 nm) e UVB (280-315 nm) lasciando passare solo i raggi UVA (315-400 nm) e una piccola parte di quelli UVB.



Di conseguenza i nostri oggetti, al di là della nostra percezione, continuano il gioco di assorbimenti e riflessioni delle radiazioni...anche di quelle che noi non vediamo. Il fenomeno della fluorescenza è legato a tale comportamento. Proviamo ad immaginare, come detto per gli oggetti colorati, che la nostra sostanza assorba una

radiazione a noi invisibile, per esempio quella UV. Dopo aver dissipato una parte dell'energia, riemette una radiazione con energia inferiore. Se l'energia di riemissione è tale da trovarsi nella banda del visibile, allora ci apparirà colorata. Se, invece, dovesse capitare della regione IR, allora continueremmo a non vederla.

Nel caso fortunato in cui il nostro oggetto assorba radiazioni nell'ultravioletto e le emetta nel visibile saremo in presenza del fenomeno della fluorescenza. In tal caso l'oggetto non si limiterà a riflettere la normale quantità di luce colorata, ma emetterà addirittura un po' di luce dello stesso colore (quella derivante dall'assorbimento UV),



apparendo così molto più luminoso (fino a quattro volte) rispetto alla luce incidente. Ora resta un dubbio da sedare, come fa un oggetto ad assorbire la radiazione luminosa (visibile o invisibile al nostro occhio)? In ogni sostanza colorata sono presenti dei pigmenti detti cromofori, ovvero sostanze in grado fornire colorazione ad una sostanza, molecole capaci di assorbire una radiazione nel

visibile (e nel caso dei composti fluorescenti, nel vicino UV), facendo apparire la sostanza colorata. L'assorbimento della radiazione elettromagnetica avviene ad opera degli elettroni che, passando da un orbitale molecolare di livello energetico superiore, sono in grado di assorbire energia. Tale situazione, però, è temporanea per cui, un istante dopo, gli elettroni torneranno ad orbitali molecolari più stabili, di energia inferiore, cedendo l'energia acquisita sotto forma di una radiazione, solitamente, di energia inferiore ovvero di lunghezza d'onda maggiore. Il motivo per cui l'energia della radiazione elettromagnetica uscente è minore di quella incidente è legato ai fenomeni dissipativi che avvengono all'interno della molecola o, grazie alle forze intermolecolari, con molecole limitrofe.

Per completare l'analisi, però, bisogna aggiungere che una radiazione elettromagnetica, o più genericamente qualunque effetto la cui propagazione avvenga tramite onde, assorbirà energia solo quando la frequenza dell'onda incidente, ovvero la sua energia, sia uguale al valore del salto energetico tra due orbitali molecolari (fenomeno della risonanza). Va da sé che il nostro cromoforo assorbirà una radiazione nel visibile solo se il salto energetico tra i due orbitali di frontiera sarà dello stesso valore dell'onda luminosa. Idem dicasi per le sostanze

fluorescenti, che assorbiranno i fotoni della radiazione UV, quasi esclusivamente UVA, solo se la loro energia sarà compatibile con il salto tra due orbitali molecolari. Per meglio spiegare il fenomeno bisogna introdurre il concetto di orbitali di frontiera (sistema HOMO-LUMO). Il concetto di orbitali di frontiera è stato introdotto da Kenichi Fukui (Nobel per la chimica nel 1981) negli anni Cinquanta dello scorso secolo per spiegare la reattività delle molecole organiche. L'idea, sviluppata in base alla teoria degli orbitali molecolari, si basa nell'assumere che la reattività molecolare, ma anche le proprietà spettroscopiche delle molecole, possano essere spiegate dall'analisi dei due orbitali di una molecola che si trovano al confine tra quelli occupati, in particolare quello di più alta energia (HOMO - Highest Occupied Molecular Orbital) e quelli vuoti, cioè quello vuoto di più bassa energia (LUMO - Lowest Unoccupied Molecular Orbital). In base a tale approccio è lecito pensare che l'assorbimento di una radiazione elettromagnetica avviene quando avviene una transizione elettronica di tipo HOMO-LUMO.

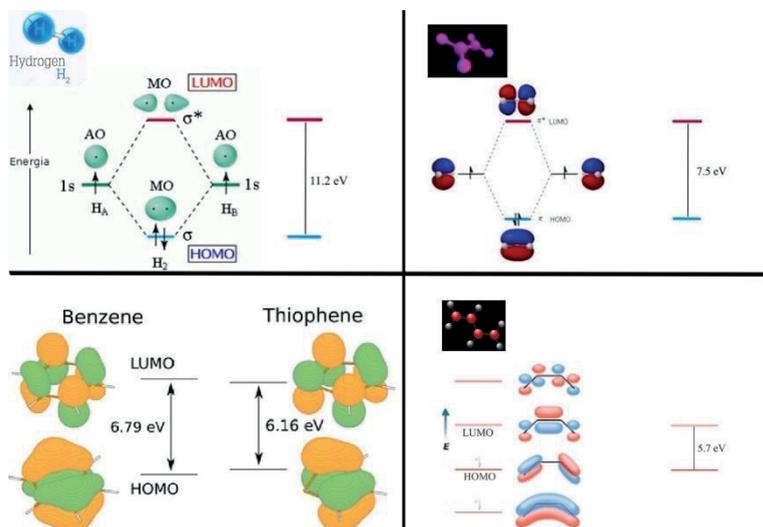
Architettura molecolare

E adesso proviamo a rispondere alla domanda delle domande. Possiamo prevedere e, nel caso, predire la struttura molecolare per un cromoforo fluorescente? Sembra chiaro che per prima cosa dobbiamo confrontare il salto energetico della transizione HOMO-LUMO con i valori energetici delle transizioni UVA. Innanzitutto, dobbiamo esprimere l'energia della radiazione UV che arriva sulla terra con dimensioni tipiche di quelle delle transizioni elettroniche. Orbene, applicando la relazione² $E = \frac{1239.84}{\lambda}$ possiamo ricavare che le energie corrispondenti alla radiazione UV con la lunghezza d'onda compresa tra i 280 nm e i 400 nm sono comprese nell'intervallo 4,5-3,0 eV. Dall'analisi dei livelli energetici HOMO-LUMO vediamo subito che per un legame singolo nella molecola di idrogeno, il salto sarebbe pari a 11.2 eV, troppo alto per una radiazione UVA. Il salto energetico HOMO-LUMO per il doppio legame C=C nell'etene è pari a 7.5 eV, mentre per il carbonio aromatico in un anello di benzene è 6.8 eV. In sistemi coniugati, quali il butadiene, il salto energetico è pari a 5.7 eV.

Questo ci porta a rilevare che un'ampia risonanza di anelli aromatici coniugati tra loro porterebbe ad individuare quelle strutture capaci di interagire con energie comparabili con quelle UVA. La presenza, poi, di eteroatomi (atomi, differenti dal carbonio, presenti in un composto ciclico organico) contribuisce ad abbassare ulteriormente l'entità della transizione energetica. C'è anche da dire che, nelle fasi

²Tale relazione deriva dalla combinazione della legge di Planck $E = h \cdot \nu$ con la definizione di velocità per un'onda elettromagnetica ove per la velocità è stata considerata quella della luce nel vuoto $c = \lambda \cdot \nu$.

condensate, giocano un ruolo di primo piano, negli scambi energetici tra le molecole e nella conseguente dissipazione, tutte quelle forze note come interazioni intermolecolari (forze di Van der Waals, legame ad idrogeno, interazioni ione dipoli, ecc.). *Last but not least*, bisogna porre molta attenzione al fenomeno del *Quenching di fluorescenza* ovvero l'attenuazione, fino alla scomparsa, della fluorescenza relazionabile o ad un'elevata concentrazione del fluoroforo e/o alla presenza di



molecole dette smorzatori di fluorescenza. In entrambi i casi giocano un ruolo essenziale gli scambi energetici intermolecolari.

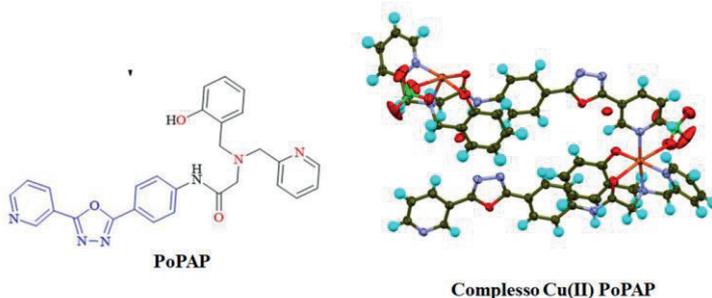
Riassumendo, per ottenere un cromoforo fluorescente dobbiamo rivolgere la nostra attenzione a molecole aromatiche, magari eterocicli, condensate e/o altamente coniugate. Per mantenere il fenomeno della fluorescenza in fasi condensate bisogna assicurare concentrazioni di cromoforo non troppo elevate ed evitare la presenza di smorzatori di fluorescenza e, qualora la fluorescenza avvenga in fase liquida, scegliere accuratamente il solvente, ricordando che le forze intermolecolari giocano un ruolo importantissimo negli scambi energetici soluto-solvente. Parametri ambientali quali, ad esempio, la temperatura possono influire sull'organizzazione molecolare e, quindi, sulle proprietà di fluorescenza del materiale.

Sonde Colorimetriche e Fluorescenti

Una recente applicazione del fenomeno della fluorescenza, e più in generale dell'impiego di sostanze cromofore, è l'uso come sonde ambientali. In due recenti sviluppi realizzati presso l'Ateneo Fridericiano, anche nell'ambito del Piano Lauree Scientifiche, sono stati studiati composti eterociclici il cui percorso di coniugazione,

e quindi la comparsa della fluorescenza, è favorito dalla presenza di metalli di transizione. Ed in effetti, tali composti vengono utilizzati come sonde ambientali per determinare la presenza di metalli pesanti nelle acque o sonde di pH *in situ*.

Attualmente la determinazione di metalli pesanti avviene grazie all'uso di tecniche di laboratorio che si basano sulla spettrometria di decadimento dei singoli elementi (ICP-MS). Tali tecniche sono molto accurate ma prevedono un complesso iter di preparazione del campione nonché un'indagine complessa e costosa eseguibile solo presso strutture laboratoriali, possibilmente accreditate. Il sistema messo a punto, invece, è utilizzabile *sul campo* e usa come sistema spettrofotometrico una comune app da smartphone.



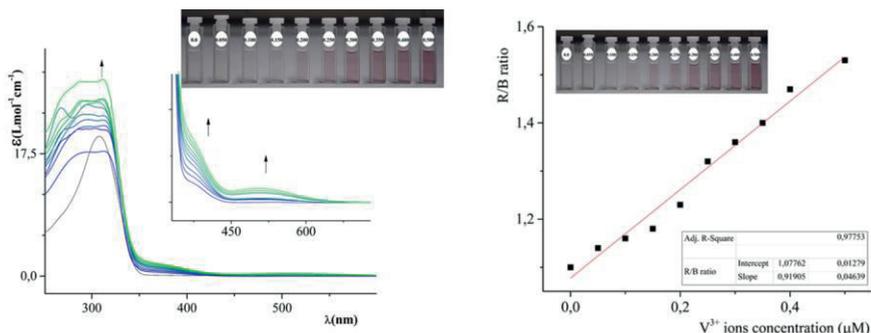
La molecola sintetizzata (PoPAP) è composta da un elevato numero di anelli aromatici, talvolta eterociclici, con scarsa coniugazione tra di loro. Il percorso di coniugazione, che abbiamo imparato abbassa l'energia di transizione HOMO-LUMO, viene ampiamente completato quando la nostra molecola viene complessata con un metallo della serie di transizione, per il suo elevato numero di elettroni d e la capacità di generare un sistema di donazione-retrodonazione³.

Questo fa sì che la molecola di cromoforo, incolore in soluzione, assuma colori differenti in base ai differenti metalli con cui viene in contatto.

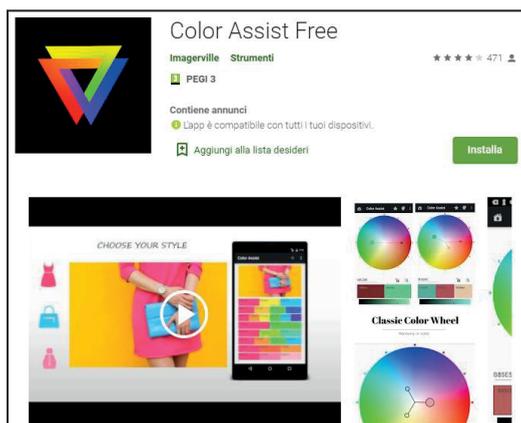


³ L'effetto di donazione-retrodonazione è una classica interazione di tipo π . In particolare, la donazione π avviene quando un orbitale p del sistema legante è pieno e l'orbitale d del metallo è vuoto. In tal caso, il legante dona i suoi elettroni al metallo instaurando un legame covalente di tipo dativo. Contemporaneamente se l'orbitale p^* del legante è vuoto e l'orbitale d del metallo è pieno, il metallo dona i suoi elettroni al legante. Tale fenomeno conferisce una maggiore stabilità al legame carbonio-metallo nella formazione del complesso.

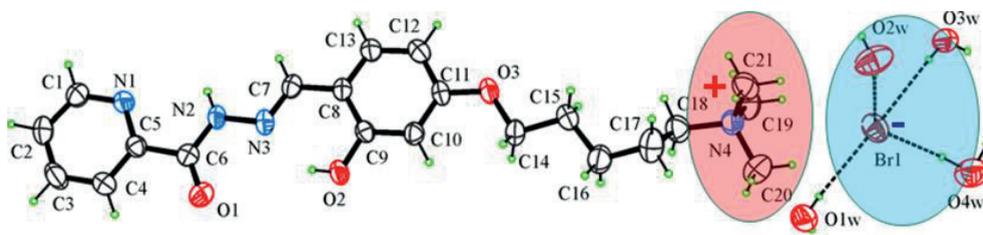
Inoltre, per la linearità della risposta in funzione della concentrazione (ovvero al raddoppio della concentrazione, raddoppia l'intensità del colore della soluzione, ovvero raddoppia l'assorbanza della soluzione) è possibile utilizzare questo sistema per individuare la concentrazione delle specie metalliche presenti in soluzione, il che rende questo sistema particolarmente interessante per la produzione di kit, da utilizzare nella zona di intervento, che permettano la determinazione istantanea di inquinamento da metallo pesante nelle acque.



Inoltre, il sistema di rilevazione è abbastanza smart potendo utilizzare un semplice smartphone per la determinazione dell'intensità di colorazione e, quindi, a seguito di una preventiva calibrazione, la concentrazione di metallo presente. Nel nostro caso, abbiamo usato una delle innumerevoli applicazioni (free) per smartphone presenti per la determinazione del colore in una foto. Le operazioni sono semplici, si versano alcune gocce di acqua da sottoporre ad analisi in una cuvette da spettrofotometria UV-Vis (ne esistono anche di usa e getta) contenente una soluzione (incolore) del nostro cromoforo. Se la soluzione si colora, nell'esempio, di viola per la presenza del Vanadio, allora potremmo determinarne la concentrazione confrontando l'intensità letta con la retta di calibrazione.



Un ulteriore esempio, sempre sviluppato nei nostri laboratori, riguarda la determinazione del pH *in situ*. Sebbene la determinazione del pH sia semplice e di uso comune e non richieda apparecchiature costose o procedure complesse, non sempre è possibile determinarne il valore ove si vorrebbe. Si provi a pensare di determinare il valore di pH in prossimità di una cellula, o addirittura sulla sua membrana, quando questa sia in un organismo vivente. Conosciamo benissimo l'importanza del pH, e del suo controllo, da parte degli organismi viventi. Nel nostro laboratorio chimico personale (il nostro corpo) c'è un complesso equilibrio acido-base, regolato da svariati tamponi, nei vari organi. Pertanto, la determinazione del pH *in situ*, e mentre la nostra macchina fisiologica continua a funzionare, è un interessante obiettivo da raggiungere. Anche in questo caso è stata sintetizzata una molecola nella cui struttura è presente una zona composta da un sistema aromatico parzialmente coniugato (a sinistra della figura) atta a recepire la radiazione UV-Vis ed un terminale "salino" (evidenziato nella parte destra della figura con le rispettive cariche ioniche) che ne assicura una perfetta solubilità in acqua.



Questo sistema, incolore e non fluorescente a $\text{pH} < 8$, risente di un delicato equilibrio acido base che ne permette il cambiamento della struttura molecolare e, conseguentemente, la comparsa di una colorazione gialla nel visibile, nonché della fluorescenza (verde) per irraggiamento con una comune lampada UV da laboratorio. Anche in questo caso una semplice determinazione dell'intensità della colorazione ci permette di risalire al valore del pH. Inoltre, mentre risulta abbastanza complicato, in un sistema colorato, determinare la presenza del giallo, è abbastanza semplice e poco influenzato dal sistema, la rilevazione della fluorescenza. Pertanto, la fluorescenza di tale sistema risulta particolarmente interessante rendendolo un'ottima sonda acido-base.



Studi di dinamica molecolare hanno, poi, consentito di determinare che il cromoforo si lega stabilmente alla membrana lipidica cellulare non alterandone lo spessore dimostrandosi, quindi, un eccellente sonda per la determinazione del pH direttamente sulla membrana cellulare.

Riassumendo potremmo dire che la fluorescenza mostrata da alcune molecole, dette cromofori o, usando una terminologia anglosassone, *dyes*, grazie alla loro interazione con la radiazione solare sono in grado di assorbire luce nell'ultravioletto (dove il nostro occhio non vede) riemettendola ad energia inferiore nel visibile, così conferendo una particolare brillantezza al colore. La struttura molecolare dei cromofori deve prevedere anelli aromatici, anche eterociclici, altamente condensati e/o coniugati. Quest'ultima ipotesi è da preferire ad un'elevata condensazione in quanto, poi, il materiale potrebbe risultare insolubile e infondibile.

Le applicazioni dei materiali fluorescenti poi sono molteplici, e tutte di interesse tecnologico e commerciale, dai più comuni evidenziatori e giubbotti di sicurezza, alle più sofisticate sonde ambientali e/o farmacologiche.

Bibliografia

- Diana R., Caruso U., Concilio S., Piotto S., Tuzi A., Panunzi B. (2018). A real-time tripodal colorimetric/fluorescence sensor for multiple target metal ions. *Dyes and Pigments*, 155:249-257. <https://doi.org/10.1016/j.dyepig.2018.03.055>
- Diana R., Panunzi B., Tuzi A., Piotto S., Concilio S., Caruso U. (2019) An Amphiphilic Pyridinoyl-hydrazone Probe for Colorimetric and Fluorescence pH Sensing. *Molecules*, 24(21):3833. <https://doi.org/10.3390/molecules24213833>

Seminario integrato⁴

Fluorescenza: dal fotone ai materiali e alle proteine reporter

Parte 2 La proteina fluorescente verde GFP, dalla ricerca di base alle applicazioni, e viceversa

Mariangela Succoio e Nicola Zambrano

Dipartimento di Medicina Molecolare e Biotecnologie Mediche,

Università di Napoli Federico II;

CEINGE Biotecnologie Avanzate S.C.aR.L., Napoli;

Associazione Culturale DiSciMuS RFC, Casoria (NA)

Abstract

Il premio Nobel rappresenta il più importante e prestigioso riconoscimento a livello internazionale, cui uno studioso possa ambire. L'istituzione del Premio si deve ad Alfred Nobel, agli inizi del '900, per premiare gli studiosi che maggiormente avessero contribuito, attraverso le loro ricerche, scoperte, opere letterarie, all'ampliamento delle conoscenze in diversi campi come la Fisica, la Chimica, la Medicina o la Fisiologia, la Letteratura e, per ultimo ma non meno importante, per l'impegno in favore della pace mondiale con il premio Nobel per la Pace.

In questo articolo ci soffermeremo sul premio Nobel per la Chimica del 2008 ottenuto *“per la scoperta e lo sviluppo della proteina verde fluorescente, GFP”* da Roger Tsien, professore presso l'Università di San Diego, California, da Martin Chalfie, della Columbia University, e da Osamu Shimomura, ricercatore presso il Marine Biological Laboratory del Massachusetts, U.S.A..

Il Premio Nobel per la scoperta della proteina fluorescente GFP può essere considerato un evidente esempio “virtuoso” di come una ricerca di base, rappresentata dalla descrizione di un fenomeno di bioluminescenza che avviene in natura, possa condurre verso importanti applicazioni nella moderna Biologia e Medicina.

Le origini del Premio Nobel

Il premio Nobel può essere senz'altro considerato il riconoscimento più prestigioso a livello mondiale. Esso fu creato e prende il nome da un ingegnere svedese, Alfred Bernhard Nobel, nato a Stoccolma nel 1833. Nel periodo di giovinezza trascorso in

⁴La Parte introduttiva sulle transizioni elettroniche e la fluorescenza, tenuta dal prof. Scotti di Uccio, è stata inserita nel contributo “La fisica dell'atomo di Idrogeno

Francia a Parigi, Alfred si appassionò alla chimica e con il passare del tempo, la sua forte passione per la materia, unita a numerosi esperimenti e non pochi incidenti (in uno dei quali morì anche un suo fratello) lo portarono all'invenzione della dinamite, brevettata nel 1867. La dinamite è un derivato della nitroglicerina che, miscelata con opportune sostanze, ne conferiscono maggiore stabilità a temperatura ambiente, e quindi una maggiore sicurezza nell'utilizzo. L'invenzione di Alfred Nobel e i guadagni derivati dal brevetto gli permisero di aprire numerose aziende e laboratori in tutto il mondo, fra cui uno dei più grandi stabilimenti in Italia, nelle vicinanze di Torino, che cercò di seguire personalmente, al punto da essere definito come *“il più ricco vagabondo d'Europa”*. La dinamite in breve tempo divenne il migliore esplosivo in commercio ampiamente utilizzato nell'industria mineraria, nelle cave, in edilizia, nelle demolizioni e, in misura minore, in campo bellico, rendendo Nobel un ricco imprenditore.

Se da un lato l'invenzione della dinamite rappresentò un valido strumento in ambito edile per facilitare la costruzione di strade, tunnel, porti, canali ecc., dall'altra parte per Alfred Nobel l'idea che la dinamite da lui inventata potesse essere uno strumento di morte, lo perseguì per tutta la sua vita. Per tale motivo, nel 1895, un anno prima della sua morte, Alfred scrisse il suo testamento, in cui inserì la sua decisione di destinare parte della sua fortuna ad un fondo da cui poter attingere per conferire una somma in denaro ai vincitori dei Premi Nobel. Nel testamento Nobel aggiunse che i premi dovessero essere assegnati alle persone più meritevoli, a prescindere dalla loro nazionalità.

I premi voluti da Alfred Nobel sono cinque: il Premio Nobel per la Fisica, per la Chimica, per la Medicina o la Fisiologia, per la Letteratura e per la Pace. A partire dal 1901, nel giorno dell'anniversario della morte di Nobel, il 10 dicembre di ogni anno avviene la consegna dei premi, a Stoccolma e a Oslo, i cui i nomi dei vincitori sono resi noti ad ottobre. Infine, esiste anche un ulteriore premio, il Premio Nobel per l'Economia ideato dalla *Banca di Svezia– Sveriges Riksbanks* – in onore di Alfred Nobel, e consegnato per la prima volta nel 1969.

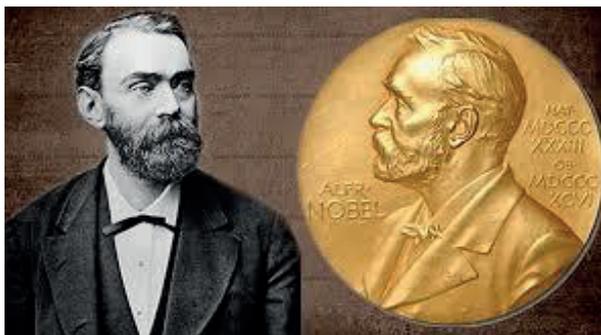


Figura 1. Alfred Nobel, l'ideatore del Premio Nobel.

Il Premio Nobel per la Chimica 2008: *Aequorea victoria*, la luce dell'oceano

Il Premio Nobel per la Chimica del 2008 fu assegnato a Roger Tsien, Martin Chalfie, e a Osamu Shimomura, “*per la scoperta e lo sviluppo della proteina verde fluorescente, GFP*”.

Fu Osamu Shimomura in Giappone che per primo si occupò di studiare dal punto di vista chimico e biochimico gli organismi bioluminescenti, sia acquatici che terrestri. In particolare, studiò la bioluminescenza prodotta da due organismi: la *Cypridina hilgedorfii*, una specie di crostaceo chiamata anche “lucciola marina” e la *Luciola lusitanica*, la comune lucciola appartenente al genere dei minuscoli insetti coleotteri. Il meccanismo di luminescenza di questi organismi si basa su una reazione di ossidazione del substrato fotogeno luciferina ad ossiluciferina, in presenza di ossigeno ad opera dell'enzima luciferasi.

Nel 1960 Shimomura dal Giappone si recò a lavorare negli Stati Uniti a Princeton per approfondire e studiare il meccanismo di bioluminescenza ritrovato in un altro organismo, una medusa. Tale meccanismo era completamente diverso rispetto agli organismi animali precedentemente studiati, in quanto non dipendeva da una reazione chimica (fenomeno definito chemiluminescenza) ma da “trasferimenti di energia” tra proteine. Oggetto di studio di Shimomura fu una medusa caratteristica dell'oceano Pacifico nord-orientale, dal nome *Aequorea victoria*, così denominata perché raccolta nella baia dell'isola Victoria in Canada. *Aequorea victoria* emette un segnale luminoso utilizzato come richiamo, o per intimidire eventuali predatori; tale segnale è generato grazie a centinaia di piccoli organi, situati sul bordo del suo ombrello (Fig. 2), ed all'azione di due proteine: una proteina legante il calcio ed un'altra proteina a quel tempo sconosciuta. Shimomura catturò migliaia di meduse, al fine di isolare entrambe le proteine e poterne comprendere il meccanismo biologico.

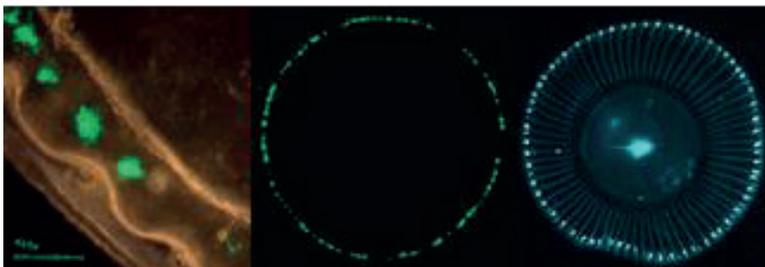


Figura 2. Gli organi circolari presenti sulla struttura della medusa *Aequorea victoria* definita ombrello sono mostrati a sinistra, l'intera medusa al buio al centro, e sotto luce visibile a destra (Zimmer,2009).

Shimomura ne comprese il meccanismo. La prima proteina, chiamata aequorina (dal nome della medusa), emette luce blu quando si lega al calcio, la seconda proteina assorbe

la luce emessa dall'aequorina ($\lambda_{\max} = 470 \text{ nm}$) e a sua volta emette una luce verde ($\lambda_{\max} = 509 \text{ nm}$) proprio per questo fu chiamata "proteina verde fluorescente" in inglese *Green Fluorescent Protein* (GFP) (Fig. 3).

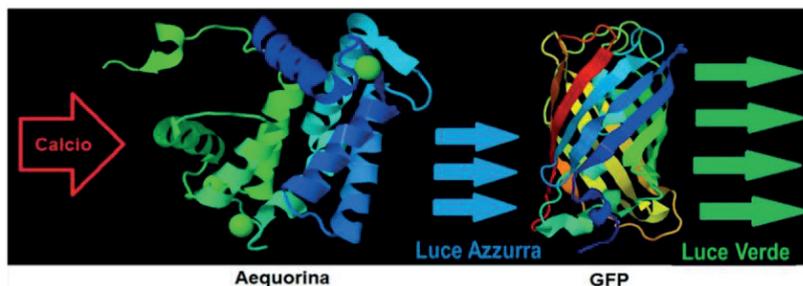


Figura 3. Il meccanismo molecolare della fluorescenza nella medusa *Aequorea Victoria*. (Immagine modificata dal sito web: <https://100cia.site/index.php/quimica/item/2106-por-que-es-importante-la-proteina-verde-fluorescente-y-por-que-deberias-conocerla>).

Le fasi di estrazione, purificazione e le proprietà della proteina aequorina furono descritte in una pubblicazione del 1962 (Shimomura *et al.*, 1962), al contrario la purificazione della proteina GFP si dimostrò molto più difficile del previsto. Fu solo nel 1971 che Shimomura e i suoi colleghi riuscirono ad isolarne grosse quantità sufficienti per iniziare a studiarla, descrivendo poi la fase di purificazione e il sistema di trasferimento di energia intermolecolare presente nell'*Aequorea victoria* in una pubblicazione del 1974 (Morise *et al.*, 1974).

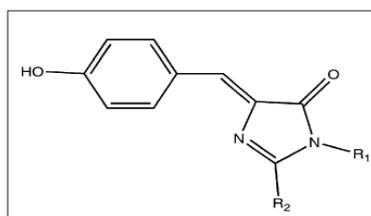


Figura 4. In Figura è rappresentata la struttura chimica del fluoroforo della proteina GFP di *Aequorea Victoria*. Il fluoroforo è la porzione proteica costituita da tre amminoacidi: Ser65 – Tyr66 – Gly67 e in grado di assorbire la luce blu ed emettere fluorescenza verde (Zimmer, 2009)

Il gruppo di ricerca di Shimomura riuscì, attraverso tecniche di spettroscopia e di sintesi, a descrivere la struttura chimica del fluoroforo della proteina GFP, cioè quel gruppo di atomi in grado di assorbire la luce blu, ed emettere fluorescenza a seguito di una transizione elettronica (Shimomura *et al.*, 1979).

Tali scoperte richiesero circa vent'anni di ricerca svolta dal laboratorio di Shimomura per comprendere appieno, da un punto di vista prettamente chimico e biochimico, il

fenomeno della bioluminescenza negli organismi animali. Di certo, Shimomura non poteva immaginare la notevole risonanza e le possibili applicazioni che la sua scoperta avrebbe avuto in futuro.

Un'intuizione da Premio Nobel

Migliaia di animali erano necessari per ottenere pochi milligrammi di proteina GFP utili per eseguirne la caratterizzazione biochimica, e questa fu la principale difficoltà che Shimomura riscontrò nei suoi studi.

Un decisivo passo in avanti fu ottenuto, grazie all'utilizzo delle moderne tecniche di biologia molecolare quali il clonaggio molecolare, che permisero la produzione di grosse quantità di proteina GFP in laboratorio. Nel 1987 Douglas Prasher, un biologo molecolare dell'Ohio State University, fu il primo ad intuire che le caratteristiche possedute dalla proteina GFP, quali le piccole dimensioni, la fluorescenza e il non necessitare di substrato per la fluorescenza, potessero essere utilizzati per ideare uno dei più inimmaginabili metodi di "cell tracking". Questa sua intuizione aprì la strada alla straordinaria rivoluzione che ha portato poi la GFP a meritare il premio Nobel. Prasher isolò da *Aequorea victoria* la sequenza di cDNA di GFP (Fig. 5) (Prasher *et al.*, 1992; Cody *et al.*, 1993) con l'intento poi di fondere, attraverso le metodologie del DNA ricombinante, la GFP a una determinata proteina da studiare e quindi, attraverso tecniche di *imaging*, poterne tracciare l'espressione e il traffico subcellulare. È infatti ben noto che, quando in un organismo è necessaria una determinata proteina, si attiva un processo di sintesi che parte dalla trascrizione, fino alla lettura dell'RNA messaggero sui ribosomi, fino alla tripletta - denominata "codone di stop" - che stabilisce il punto in cui terminare la sintesi di una proteina. Doug Prasher pensò di servirsi di tecniche biomolecolari per inserire il mini-gene della GFP in coda alla sequenza codificante di una proteina di interesse, immediatamente prima del codone di stop. In tal modo, la proteina in questione sarebbe diventata fluorescente, in quanto fusa alla GFP.

Purtroppo, Prasher incontrò non poche difficoltà tecniche, il vettore esprimente la proteina GFP da lui clonato, non esprimeva fluorescenza. La sua ipotesi era che probabilmente la proteina GFP ricombinante prodotta in laboratorio richiedeva alcune modifiche post-traduzionali caratteristiche delle cellule animali, quindi, doveva essere espressa in organismi superiori. Nel 1991 non riuscì ad ottenere ulteriori finanziamenti per la ricerca e alla fine, dovette abbandonare la scienza e il mondo accademico trovando lavoro come autista!

Il suo lavoro pionieristico fornì spunti a Martin Chalfie e Roger Y. Tsien, che avrebbero condiviso con Shimomura il Premio Nobel per la Chimica del 2008, per il lavoro che essi hanno pubblicamente riconosciuto essere sostanzialmente basato sul lavoro di Douglas

Prasher. Attraverso i loro sforzi e quelli di altri, Douglas Prasher è poi tornato al lavoro di ricerca scientifica nel giugno 2010.

```

tacacacgaataaaagataaaacaaag atg agt aaa gga gaa gaa ctt ttc act
                                M S K G E E L F T
gga gtt gtc cca att ctt gtt gaa tta gat ggc gat gtt aat ggg
G V V P I L V E L D G D V N G
caa aaa ttc tct gtc agt gga gag ggt gaa ggt gat gca aca tac
Q K F S V S G E G E G D A T Y
gga aaa ctt acc ctt aaa ttt att tgc act act ggg aag cta cct
G K L T L K F I C T T G K L P
ggt cca tgg cca aca ctt gtc act act ttc tct tat ggt gtt caa
V P W P T L V T T F S Y Q V Q
tgc ttt tca aga tac cca gat cat atg aaa cag cat gac ttt ttc
C F S R Y P D H M K Q H D F F
aag agt gcc atg ccc gaa ggt tat gta cag gaa aga act ata ttt
K S A M P E G Y V Q E R T I F
tac aaa gat gac ggg aac tac aag aca cgt gct gaa gtc aag ttt
Y K D D G N Y K T R A E V K F
gaa ggt gat acc ctt gtt aat aga atc gag tta aaa ggt att gat
E G D T L V N R I E L K G I D
ttt aaa gaa gat gga aac att ctt gga cac aaa atg gaa tac aac
F K E D G N I L G H K M E Y N
tat aac tca cat aat gta tac atc atg gca gac aaa cca aag aat
Y N S H N V Y I M A D K Q K N
gga atc aaa gtt aac ttc aaa att aga cac aac att caa gat gga
G I K V N F K I R H N I K D G
agc gtt caa tta gca gac cat tat caa caa aat act cca att ggc
S V Q L A D H Y Q Q N T P I G
gat ggc cct gtc ctt tta cca gac aac cat tac ctg tcc aca caa
D G P V L L P D N H Y L S T Q
tct gcc ctt tcc aaa gat ccc aac gaa aag aga gat cac atg atc
S A L S K D P N E K R D H M I
ctt Ctt gag ttt gta aca gct gct ggg att aca cat ggc atg gat
L L E F V T A A G I T H G M D
gaa cta tac aaa taa atgtccagac ttccaattaga cactaaagtg
E L Y K
tccgaacaat tactaaaate tcagggttcc tggttaaatt caggctgaga
tattatttat atatttatag attcaattaaa attgtatgaa taatttattg
atgttattga tagagggttat tttcttatta aacaggctac ttggagagta
ttcttaattc tatattaatt acaatttgat ttgacttget caaa

```

Figura 5. Sequenza nucleotidica e amminoacidica del cDNA della GFP riportato da Prasher nell'articolo pubblicato sulla rivista *Gene* nel 1992. Gli amminoacidi che formano i cromofori sono in grassetto e sottolineati. Nucleotidi aggiuntivi precedenti e dopo il gene GFP nel vettore utilizzato da Prasher sono in corsivo (Zimmer, 2009).

Marty Chalfie e Roger Tsien: la svolta decisiva per il Premio Nobel

“Decisi di vedere chi era il “tonto” che aveva vinto il premio Nobel quest’anno.... così aprii il mio portatile e scoprii che quel tonto ero io”. Con queste parole Marty Chalfie, docente alla Columbia University, descrisse come venne a sapere della sua assegnazione

del Premio Nobel per la Chimica del 2008 “*for the discovery and development of the green fluorescent protein, GFP*”, condiviso con Roger Tsien, e Osamu Shimomura.

La strada di Marty Chalfie per il Nobel iniziò già ai tempi del liceo, in quanto era amico di Bob Horvitz, futuro vincitore del Premio Nobel per la Medicina o la Fisiologia del 2002, insieme a Sydney Brenner e John E. Sulston, per la scoperta della morte cellulare programmata o più propriamente definita “apoptosi”. Successivamente, dopo la laurea in Chimica e un dottorato in Fisiologia presso l'Università di Harvard, Chalfie fu convinto dallo stesso Bob Horvitz a svolgere i suoi anni di post-dottorato nel laboratorio di Sydney Brenner. In quel tempo Chalfie studiava la funzione di alcuni recettori mecano-sensori nell'organismo modello *C. elegans*, un piccolo verme cilindrico appartenente al Phylum dei Nematodi, molto utilizzato per gli studi in campo biologico. *C. elegans* possiede una serie di caratteristiche che lo rendono un sistema modello ideale per la ricerca scientifica: vive facilmente su piastra, si nutre di batteri, e soprattutto è trasparente, quindi, può essere facilmente osservato utilizzando uno stereomicroscopio. Durante un seminario in cui vennero descritti gli eccellenti studi di Shimomura, Chalfie pensò che *C. elegans*, in quanto trasparente, sarebbe stato un sistema modello utilissimo per poter visualizzare proteine ricombinanti rese fluorescenti con GFP, e studiarne così la sintesi e il destino cellulare. Nel 1994, una studentessa del suo gruppo, Ghia Euskirchen, isolò tramite PCR il mini-gene GFP dal vettore che Chalfie aveva ottenuto da Prasher, e lo espresse in *E. coli*. Fu la strategia decisiva!

Ghia comprese che ciò che impediva al vettore GFP prodotto da Prasher di esprimere una proteina GFP funzionale, e quindi fluorescente, era la sequenza di DNA presente nel vettore che precedeva il gene GFP. Tale sequenza era responsabile di un ripiegamento non corretto della proteina e di un'alterata sintesi della porzione proteica contenente il fluoroforo. Inoltre, Ghia aveva esperienza con il microscopio a fluorescenza, e così fu in grado di distinguere tra l'autofluorescenza verde intrinseca dei batteri e la fluorescenza propria della proteina GFP (Chalfie *et al.*, 1994). Chalfie fu presto in grado di utilizzare promotori noti per esprimere GFP nei neuroni tattili di *C. elegans* e Tulle Hazelrigg, moglie di Martin Chalfie ed anch'essa scienziata, fornì un ulteriore e importante contributo nel campo della GFP. La studiosa riuscì a produrre la prima proteina di fusione con GFP dimostrando che avrebbe potuto sostituire funzionalmente la proteina originale, e mostrare così il suo destino in una cellula (Wang *et al.*, 1994). Uno dei più fenomenali metodi di “*cell tracking*” era stato inventato.

Ma il maggior contributo alla comprensione del funzionamento della GFP venne da Roger Tsien, professore presso l'Università di San Diego, California. La proteina GFP mostrava alcuni svantaggi, in particolare presentava un doppio picco di eccitazione attribuibile alle proprietà chimiche del cromoforo. Gli studi di Tsien condussero alla realizzazione di una proteina mutante, la GFP S65T, in cui la sostituzione della serina in

posizione 65 con una treonina, era responsabile di una proteina GFP con un solo picco di eccitazione e una luminosità sei volte più intensa. Questa proteina mutante rappresenta, a tutt'oggi, la proteina fluorescente più comunemente utilizzata nei laboratori di ricerca nota con il nome di proteina fluorescente verde o EGFP, dove E sta per "enhanced". Nel 1996 Tsien e colleghi descrissero la struttura cristallina della proteina GFP; essa è costituita da 238 amminoacidi, ha un PM di 27.000 Dalton ed è costituita da 11 foglietti beta disposti in circolo a formare una struttura denominata barile- β (β -barrel o β -can). Inoltre, sono presenti due segmenti ad alfa elica, uno alla base del barile, l'altro lungo il suo asse centrale che contiene il fluoroforo. (Ormo *et al.* 1996) (Fig. 6).

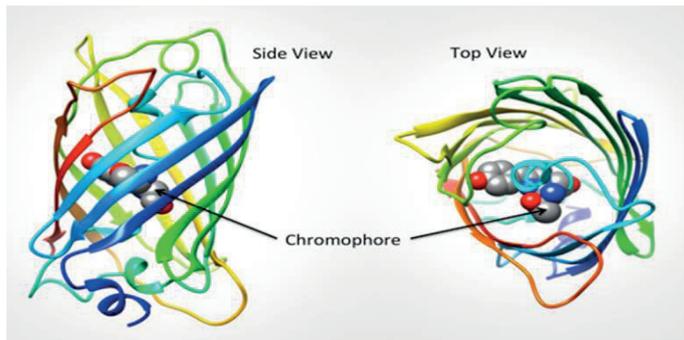


Figura 6. La struttura cristallina della proteina GFP presente sul database di strutture proteiche PDB (*Protein Data Bank*) (Immagine tratta e modificata da <https://www.rcsb.org/structure/1EMA>).

Tsien, inoltre, mise a punto nuove tecniche, produsse molte proteine mutanti, che emettono fluorescenza più rapidamente, con maggiore intensità e anche di diversi colori (Fig. 7) (Tsien *et al.*, 1998; Shaner, *et al.*, 2004).

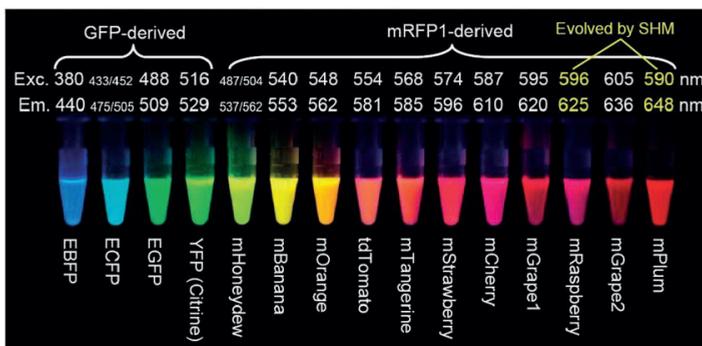


Figura 7. Le diverse versioni mutate di GFP e della Red Fluorescent Protein RFP del corallo *Discosoma*, dotate di fluorescenza potenziata. In figura sono riportate per ogni proteina fluorescente le rispettive lunghezze d'onda (λ) di eccitazione (Exc.) e di emissione (Em.). (Shaner *et al.*, 2004).

Un ulteriore passo avanti fu fatto da un altro scienziato, Sergey Lukyanov, un ricercatore russo che trovò alcune proteine simili alla GFP nei coralli del genere *Discosoma*. Nessuno ci aveva mai pensato, perché i coralli non sono bioluminescenti. Lukyanov trovò che alcune proteine avevano una fluorescenza rossa e un meccanismo simile a quello della GFP (Lukyanov *et al.*, 2000). Questa scoperta aprì la strada per la ricerca di molte altre proteine simili e a basso costo.

Esempi di possibili applicazioni:

Il premio Nobel per la chimica 2008 ha come protagonista una proteina fluorescente, la GFP, che ha trovato diverse applicazioni in molti settori della Biologia moderna. Le proteine fluorescenti permettono di marcare popolazioni cellulari di interesse e capirne la localizzazione e le funzioni. Ad esempio, mai prima della disponibilità della GFP e delle sue varianti furono ottenute immagini di cervello così belle ed informative attraverso tecniche di neuroimaging, come quelle riportate da Livet e colleghi nel 2007 sulla prestigiosissima rivista *Nature* (Fig. 8). In tale immagine, i neuroni sembrano un'opera d'arte, un quadro di Pollock! Tramite strategie genetiche sono state generate cellule nervose di cervello di topo che sono in grado di esprimere diverse tonalità di fluorescenza e visualizzare in modo univoco fino ad un centinaio di cellule diverse.

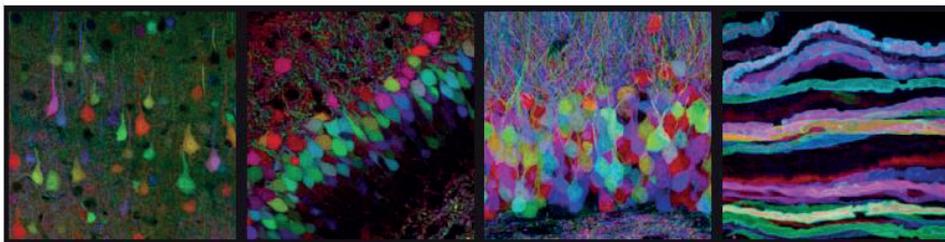


Figura 8. Il "brainbow": la mappatura dei circuiti neurali che compongono il sistema nervoso attraverso l'utilizzo di proteine fluorescenti (Livet *et al.*, 2007).

Nei topi "Brainbow" (parola che deriva dall'unione di *brain* (=cervello) e *rainbow* (=arcobaleno)), i ricercatori di Harvard attraverso l'utilizzo di diversi vettori di espressione e una strategia basata sul processo di ricombinazione genica sono riusciti a far esprimere, da un singolo promotore, una delle tre proteine fluorescenti GFP-simili (rosso, blu, verde) in numero diverso e in maniera del tutto casuale. Si ottiene una mescolanza di colori ed una tavolozza di 90 colori distinti, un risultato sicuramente straordinario rispetto alle poche tonalità possibili con le tecniche tradizionali. Tale tecnica di neuro-imaging ha permesso ai ricercatori di studiare la morfologia neurale e seguire facilmente il percorso di un neurone, capire da dove parte e dove arriva e con quali altri

neuroni viene a contatto. Sarà infatti probabile che neuroni vicini tra di loro abbiano diverso colore, rendendo così più “semplice” la mappatura dei circuiti neurali del cervello. Un altro esempio di applicazione è lo studio del ciclo cellulare. La crescita cellulare avviene attraverso un ordine sequenziale di eventi: il ciclo cellulare, infatti consiste di quattro fasi distinte, G1 (fase di gap o intervallo 1), S (fase di sintesi), G2 (fase di gap o intervallo 2) e M (mitosi). La tecnologia “Fucci” (Fluorescent Ubiquitination-based Cell Cycle Indicator) è una metodica di imaging che permette di studiare il ciclo cellulare *in vivo* in tempo reale, utilizzando proteine fluorescenti fuse a sequenze di degradazione ciclo-regolate. Le cellule geneticamente modificate sono gialle all'inizio di replicazione, verdi durante la fase G2/M e rosse durante l'intervallo G1 (Fig. 9).

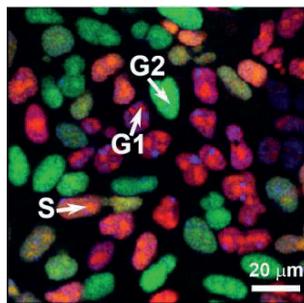


Figura 9. Cellule analizzate in laboratorio in cui è possibile riconoscere le singole fasi del ciclo cellulare utilizzando sistema Fucci (Immagine tratta da <https://www.jove.com/t/55999/advanced-confocal-microscopy-techniques-to-study-protein-protein>).

Infine, un ulteriore esempio di applicazione della GFP in biologia cellulare è rappresentato dallo studio di proteine coinvolte nel traffico intracellulare come quello tra il citoplasma e il nucleo e viceversa in cellule eucariotiche.

La membrana nucleare che circonda il nucleo è perforata da grosse strutture note come “i complessi dei pori nucleari”, costituiti da più di 50 proteine diverse chiamate *nucleoporine* con simmetria ottagonale.

I complessi dei pori nucleari sono canali acquosi che consentono un traffico bidirezionale tra il citosol e il nucleo. Le molte proteine che agiscono nel nucleo (DNA e RNA polimerasi, proteine che regolano i geni) sono importate selettivamente nel compartimento nucleare dal citosol, dove sono prodotte. Contemporaneamente, molecole come tRNA e mRNA sono sintetizzate nel compartimento nucleare e quindi esportate nel citosol. Numerose proteine, ricircolano continuamente tra nucleo e citosol, attraverso meccanismi di importazione e di esportazione. La selettività di questi processi di importazione ed esportazione delle proteine richiede la presenza di specifici segnali di

localizzazione che sono presenti sulle proteine stesse: un segnale di localizzazione nucleare (NLS) e un segnale di esportazione nucleare (NES).

Dal punto di vista sperimentale, la valutazione della funzionalità di una putativa sequenza NLS di una proteina può essere effettuata attraverso il clonaggio in un vettore esprimente la sequenza NLS della proteina unita a EGFP utilizzata come reporter.

Una volta sovraespressa in colture cellulari, la proteina ricombinante di fusione EGFP+NLS consente di sfruttare le proprietà fluorescenti EGFP per studiare la localizzazione della putativa sequenza NLS attraverso un microscopio a fluorescenza. Se effettivamente la sequenza oggetto di studio è una sequenza NLS al microscopio a fluorescenza si vedrà una fluorescenza intensa che interessa la struttura nucleare (Sasso *et al.*, 2015) (Fig.10).



Figura 10. Immagini ottenute con un microscopio a fluorescenza di cellule eucariotiche trasfettate con un vettore esprimente solo la proteina reporter EGFP (a sinistra) e di cellule trasfettate con il vettore pEGFP+NLS (a destra). L'esperimento dimostra che la sequenza NLS clonata nel vettore pEGFP è effettivamente una sequenza di localizzazione nucleare in quanto la proteina fluorescente prodotta, EGFP + NLS si localizza nel nucleo della cellula eucariotica (Immagine tratta e modificata da Sasso *et al.*, 2005)

Conclusioni

Il Premio Nobel per la Chimica 2008 ha premiato sia la ricerca di base che le sue potenzialità applicative, dimostrando quanto sia importante un equilibrio tra entrambe. Mentre l'interesse principale di Shimomura era basato sulla curiosità scientifica per la comprensione delle caratteristiche della proteina GFP e il suo ruolo nel fenomeno di bioluminescenza in *Aequorea victoria*, Chalfie e Tsien furono guidati dall'intuizione che le caratteristiche della GFP e delle sue varianti potessero rivelarsi un importante strumento

nelle scienze biomediche sperimentali. Quindi, i loro interessi furono, rispettivamente, quelli di far esprimere in organismi viventi proteine di fusione fluorescenti per studiarne localizzazione e funzioni, e di mettere a punto versioni più luminose, più fotostabili, con un ampio spettro di colori.

L'assegnazione di questo premio Nobel evidenzia, come in tanti altri casi, quanto sia importante la ricerca di base; essa può aprire le porte a diverse applicazioni biotecnologiche, e condurre a sviluppi largamente inattesi e capaci di alimentare nuove scoperte, conoscenze e applicazioni. Dall'altro lato, tale premio riconosce anche l'importanza della ricerca applicata. Non a caso, Roger Tsien concluse il suo discorso per il Nobel con la seguente citazione: "Alcune persone a volte ci hanno criticato principalmente per aver lavorato solo sulle tecniche. Mi piacerebbe attirare la loro attenzione su un vecchio proverbio cinese che dice che "se offri un pesce a un uomo, gli dai da mangiare per un giorno; ma tu insegnagli a pescare, e gli darai da mangiare per tutta la vita. Ecco perché ci piace ideare attrezzature e reti da pesca, utili per un oceano di conoscenza" (Zimmer, 2009).

Bibliografia

- Chalfie, M., Tu, Y., Euskirchen, G., Ward, W. W., Prasher, D. C. (1994). *Science*, 263, 802–805. <https://doi.org/10.1126/science.8303295>
- Cody, C. W., D. C. Prasher, W. M. Westler, F. G. Prendergast W. W. Ward. (1993). Chemical structure of the hexapeptide chromophore of the *Aequorea* green-fluorescent protein. *Biochemistry* 32, 1212–1218. <https://doi.org/10.1021/bi00056a003>
- Lukyanov, K. A., A. F. Fradkov, N. G. Gurskaya, M. V. Matz, Y. A. Labas, A. P. Savitsky, M. L. Markelov, A. G. Zaraisky, X. Zhao, Y. Fang, W. Tan S. A. Lukyanov. (2000). Natural animal coloration can be determined by a nonfluorescent green fluorescent protein homolog. *Journal of Biological Chemistry*, 275, 25879–25882. <https://doi.org/10.1074/jbc.C000338200>
- Legartová, S., Suchánková, J., Krejčí, J., Kovaříková, A., Bártová, E. (2017). Advanced Confocal Microscopy Techniques to Study Protein-protein Interactions and Kinetics at DNA Lesions. *Biology*. <https://www.jove.com/v/55999/advanced-confocal-microscopy-techniques-to-study-protein-protein>
- Livet J., et al. (2007). Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature*, 450, 56-62. <https://www.nature.com/articles/nature06293>

- Morise, H., Shimomura, O., Johnson, F. H., Winant, J. (1974). Intermolecular energy transfer in the bioluminescent system of *Aequorea*. *Biochemistry*, 13, 2656–2662. <https://doi.org/10.1021/bi00709a028>.
- Ormo, M., Cubitt, A. B., Kallio, K., Gross, L. A., Tsien, R. Y., Remington S. J. (1996). Crystal structure of the *Aequorea victoria* green fluorescent protein. *Science*, 273, 1392–1395. DOI: 10.1126/science.273.5280.1392
- Prasher, D. C., V. K. Eckenrode, W. W. Ward, F. G. Prendergast M. J. Cormier. (1992). Primary structure of the *Aequorea victoria* green-fluorescent protein. *Gene* 111, 229–233. [https://doi.org/10.1016/0378-1119\(92\)90691-H](https://doi.org/10.1016/0378-1119(92)90691-H)
- Sasso E, Vitale M, Monteleone F, Boffo FL, Santoriello M, Sarnataro D, Garbi C, Sabatella M, Crifò B, Paolella LA, Minopoli G, Winum JY, Zambrano N. (2015). Binding of carbonic anhydrase IX to 45S rDNA genes is prevented by exportin-1 in hypoxic cells. *BioMed Research International*. <https://doi.org/10.1155/2015/674920>.
- Shaner, N. C., Campbell, R. E., Steinbach, P. A., Giepmans, B. N .G., Palmer, A. E., Tsien, R. Y. (2004). Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nature Biotechnology*, 22, 1567–1572
- Shimomura, O. (1979) Structure of the chromophore of *Aequorea* green fluorescent protein. *FEBS Letters*, 104, 220–222. [https://doi.org/10.1016/0014-5793\(79\)80818-2](https://doi.org/10.1016/0014-5793(79)80818-2).
- Shimomura, O., Johnson, F. H., Saiga, Y. (1962). *Journal of Cellular and Comparative Physiology*, 59, 223–229. <https://doi.org/10.1002/jcp.1030590302>.
- Tsien, R. Y. (1998). The green fluorescent protein. *Annu. Rev. Biochem.* 67, 509–544. <https://doi.org/10.1146/annurev.biochem.67.1.509>
- Wang, S. X., T. Hazelrigg, T. (1994). Soluble, highly fluorescent variants of green fluorescent protein, *Nature*, 369 400–403. <https://link.springer.com/article/10.1023/A:1005991617182>
- Zimmer, M. (2009). GFP: from jellyfish to the Nobel prize and beyond. *Chemical Society Reviews* 2813–2968. <https://doi.org/10.1039/B904023D>.

Seminario integrato

Parte 1: La termodinamica come approccio scientifico alla realtà che ci circonda

Gerardino D'Errico

Dipartimento di Scienze Chimiche, Università di Napoli Federico II

Introduzione

La società moderna pone sfide sempre nuove, alcune delle quali imprevedibili, alla cui soluzione il mondo scientifico è chiamato a contribuire in modo determinante. Basti pensare alla pandemia da COVID-19, ma anche all'emergenza ambientale planetaria, ormai innegabile. Come è giusto, gli esperti del settore sono spesso chiamati in causa dai mass media e ciò che dicono diviene rapidamente di dominio pubblico, anche grazie alla notevole cassa di risonanza offerta dai social. Queste informazioni ci raggiungono costantemente tramite i cellulari, i computer portatili e le televisioni, determinando la nostra visione di ciò che accade nella realtà che ci circonda e le nostre opinioni politiche su come gestirla.

In effetti, nella storia dell'uomo non c'è mai stato un accesso così ampio e capillare a informazioni di carattere scientifico. Questo processo, di per sé virtuoso, presenta alcuni rischi. Infatti, le indicazioni espresse dagli scienziati vengono riassunte, schematizzate, semplificate dai giornalisti e dagli altri operatori dell'informazione, non sempre esenti da condizionamenti ideologici. Ridotte in pillole (un breve filmato o una presentazione audiovisiva, ma anche una vignetta, una battuta o altro), le informazioni scientifiche che riceviamo sono spesso parziali o manipolate, e possono essere facilmente travisate. Inoltre, informazioni false (e talvolta tendenziose) vengono diffuse utilizzando gli stessi canali. È quindi fondamentale che i singoli cittadini siano in grado di recepire correttamente i dati scientifici (e con essi gli avvertimenti, le spiegazioni, gli spunti di riflessione) che gli scienziati offrono all'intera società civile, distinguendo ciò che è provato ed attendibile da ciò che non lo è (ma, a prima vista, potrebbe sembrarlo).

In generale, diventa sempre più cogente la necessità che il cittadino comune abbia visione corretta di cosa è la scienza (con le sue immense potenzialità ed i suoi numerosi limiti) e di cosa fanno gli scienziati (sia nel pubblico sia nel privato). In altre parole, è fondamentale capire i metodi attraverso cui si costruisce quella conoscenza che ogni cittadino deve porre, in modo più o meno consapevole, alla base delle proprie scelte. Ciò che consente facilmente di distinguere se

un'informazione scientifica (o simil-tale) sia vera o falsa è proprio la verifica del metodo con cui essa è stata ottenuta.

Le diverse branche della scienza si occupano della realtà che ci circonda da punti di vista differenti. Ovviamente il metodo scientifico di stampo galileiano vale per tutti, ma esso viene declinato in modo diverso nei diversi campi.

La chimica è quella parte della scienza che guarda alla realtà che ci circonda, ed ai processi che in essa accadono, riconoscendone la natura molecolare e materica. Mantenendo sempre questo punto di vista, essa si interessa dei campi più disparati, spaziando in settori che vanno dai materiali (naturali o *high-tech*) ai principi attivi farmaceutici, dalle (bio-)plastiche ai carburanti (da fonti fossili o rinnovabili), dall'*agri-food* alle formulazioni industriali, giusto per citarne alcuni. A sua volta, la chimica fisica è quella parte della chimica che si occupa specificamente di analizzare i metodi e i modelli utilizzati dalla chimica. Usando un'espressione oggi comune, si può dire che la chimica fisica effettua un'assicurazione di qualità sui processi logici seguiti dai chimici, garantendo l'affidabilità dei risultati. Di seguito, vengono riportati i primi rudimenti dell'approccio metodologico chimico-fisico. Ovviamente, il presente contributo è tutt'altro che esaustivo. Inoltre, nel tentativo di chiarire i concetti, potrà essere sfuggita qualche imprecisione. Lo scopo di queste pagine è mostrare al lettore l'importanza di seguire coscientemente un metodo corretto di analisi della realtà che ci circonda, suscitando un interesse che potrà essere soddisfatto sui molti testi dedicati, alcuni dei quali riportati nelle fonti bibliografiche.

L'approccio chimico fisico

La scienza è un insieme organizzato di conoscenze. Essa è un edificio in continua costruzione, ampliamento e ristrutturazione. In particolare, il progredire del sapere chimico si basa su un'attività di ricerca sperimentale condotta *in vitro*, *in vivo* o *in silico* (cioè in provetta, su organismi viventi o mediante simulazioni computazionali). La pianificazione, la realizzazione e la raccolta dei risultati degli esperimenti deve essere condotta con procedimenti metodici e rigorosi. Ugualmente importante, la lettura e l'interpretazione dei risultati deve utilizzare processi logici coerentemente consequenziali. L'integrazione delle nuove conoscenze acquisite con quelle già consolidate consente di progredire nell'azione di costruzione dell'edificio della conoscenza di cui si diceva, mantenendone la solidità.

Alla base di questa costruzione troviamo necessariamente un insieme di assiomi, cioè proposizioni o principi che sono assunti come veri perché ritenuti evidenti. Essi forniscono il punto di partenza di qualsiasi teoria. È importante analizzare in modo critico questi assiomi, perché i limiti della loro validità sono i limiti di tutta la scienza

che ne deriva. Nel tempo, alcuni di questi limiti sono stati individuati e superati. Ciò non significa che tutta la conoscenza che era stata derivata in precedenza non sia più valida. Semplicemente, essa rimane valida in tutti i casi in cui i limiti degli assiomi su cui si basava non vengono superati. È il caso della quantizzazione dell'energia e della relatività generale, che hanno aperto nuovi orizzonti rispetto alla fisica classica; d'altro canto, quest'ultima rimane valida ogni qual volta in cui la quantizzazione dell'energia e della relatività generale sono ininfluenti. Basta averne coscienza e distinguere con precisione i vari casi. Solo all'interno dei limiti imposti dagli assiomi su cui si basa, la scienza perviene a una descrizione verosimile della realtà e delle leggi che regolano i fenomeni che avvengono. Inoltre, solo all'interno di questi limiti la conoscenza scientifica può essere usata con finalità predittive.

Per analizzare in maggiore dettaglio alcuni aspetti di quanto detto finora, verranno ora proposte alcune domande la cui risposta può apparire apparentemente ovvia, ma che invece nascondono più di una insidia, se ci si pone l'obiettivo di fornire una risposta chiara, univoca e, soprattutto, operativa. Con quest'ultimo termine si intende che la risposta deve rendere chi la elabora o la riceve capace di operare, cioè misurare, analizzare, pianificare, agire, prevedere.

- ***Cosa studia la scienza?***

Considerando i campi enormemente disparati in cui operano gli scienziati, appare veramente difficile dare una risposta corretta, accettabile da tutti, senza che questa diventi vaga, non operativa e, quindi, inutile.

Una risposta di carattere piuttosto generale che si può dare a questa domanda è che la scienza studia il sistema. Nell'ambito scientifico, per sistema si intende un oggetto di studio che, pur essendo costituito da parti diverse reciprocamente interconnesse e interagenti tra loro o con l'ambiente esterno, si comporta ed evolve come un tutto unico, con proprie leggi generali.

Le parti che costituiscono il sistema occupano uno spazio che, affinché la definizione sia pienamente operativa, deve essere precisamente individuato. Di conseguenza, è possibile definire il sistema come la porzione di spazio sotto indagine (con il suo contenuto).

- ***Come si individua questo sistema?***

Il sistema viene univocamente determinato definendone i confini.

I confini possono essere di due tipi:

- reali, cioè costituiti da pareti materiali;

- immaginari, cioè non materiali, ma identificabili mediante una precisa definizione matematica.

Ad esempio, se si considera un contenitore cilindrico (un beaker, ad esempio, ma anche un comune cestino da ufficio per la carta straccia) la base e le pareti laterali sono materiali. Rimane da definire il confine superiore. Una scelta che potrebbe sembrare logica è definirlo come il cerchio delimitato dal bordo superiore delle pareti laterali (ma allora, se un cestino viene riempito oltre questo limite tutto ciò che trasborda non apparterrebbe al sistema). Si potrebbero fare scelte alternative, che definiscano il sistema in modo più pratico (includendo, ad esempio, tutto ciò che rimane in equilibrio nel cestino). Non c'è quindi un modo univoco di definire i confini di un sistema. È però essenziale, prima di procedere con qualsiasi studio, che essi siano inequivocabilmente ed esplicitamente determinati.

Una volta definito il sistema, si dice che tutto lo spazio che non vi rientra costituisce l'ambiente. Risulta pratico distinguere

- l'ambiente circostante, cioè quella porzione di ambiente in relazione con il sistema (si specificherà dopo in cosa consiste questa relazione);
- il resto dell'universo, cioè la parte di ambiente non in relazione con il sistema.

- *Come si studia un sistema?*

Lo studio di un sistema deve essere effettuato con metodo. Prima di intraprendere tale studio, è essenziale che l'approccio metodologico che si intende seguire sia definito ed esplicitato con chiarezza.

Sono possibili due approcci alternativi:

- l'approccio riduzionista (o microscopico): un sistema viene suddiviso in più sottosistemi, fino a scomporlo ai minimi termini. Il comportamento degli elementi del sistema così individuati può essere studiato indipendentemente mediante leggi semplici. Il comportamento del sistema nel suo insieme dipende dalla composizione e dalla integrazione dei comportamenti dei singoli elementi. In chimica fisica, questo è l'approccio con cui la meccanica quantistica descrive la realtà, suddividendola nelle molecole e negli atomi che la compongono;
- l'approccio olistico (o macroscopico): il sistema viene studiato come un tutt'unico, definendone proprietà collettive, che consentono di analizzarne il

comportamento nel suo insieme. In chimica fisica, questo è l'approccio con cui la termodinamica chimica descrive la realtà.

A prima vista, l'approccio microscopico può sembrare preferibile, dato che consente una comprensione più profonda di ciò che si osserva a partire dalle sue origini. Tuttavia, per alcuni sistemi molto complessi esso risulta impraticabile, considerata la difficoltà di individuare i componenti del sistema o in ragione dell'elevato numero di componenti diversi, di cui è difficile comporre ed integrare il comportamento. In altri casi l'approccio sarebbe praticabile, ma risulta inutilmente pesante, e difficile da seguire senza un reale vantaggio. A tal proposito, occorre sempre aver presente lo scopo dello studio che si vuole condurre. Ad esempio, se ci si pone l'obiettivo di determinare quanto lavoro potrà ottenere da un sistema (come una macchina termica), risulta poco pratico partire dall'analisi dell'energia delle molecole che compongono il sistema.

L'approccio macroscopico, che inizialmente potrebbe sembrare superficiale, risulta invece utilizzabile anche per sistemi complessi. Se applicato correttamente produce risultati affidabili in modo semplice. Esso fornisce un punto di vista pratico ed un insieme di processi logici che si possono essere applicati in campi molto diversificati, ed in particolare nell'analisi della realtà materiale che ci circonda. Esso può quindi essere utile anche nella vita quotidiana. Di seguito si riporta qualche dettaglio in più sulla applicazione dell'approccio metodologico macroscopico.

Prima di procedere oltre, è importante sottolineare che i due diversi approcci allo studio di un sistema devono ovviamente portare a conclusioni identiche. Tuttavia, può essere difficile effettuare questo confronto. L'analisi microscopica porta spesso a risultati che riguardano sistemi piccoli, mentre l'analisi macroscopica considera insiemi molto grandi di tali piccoli sistemi. Estrapolare il comportamento studiato per un sistema piccolo a quello di un sistema grande è un'operazione che, se non fatta bene, può portare a risultati erranei. In generale, dedurre il comportamento macroscopico di un sistema da quello dei singoli elementi che lo compongono richiede l'applicazione corretta dei metodi matematici propri della statistica.

- ***Come si studia la relazione tra sistema ed ambiente?***

Nell'approccio macroscopico si analizza il comportamento del sistema nel suo insieme. Specificamente, per comportamento si intende l'insieme delle relazioni, cioè degli scambi, tra sistema ed ambiente circostante.

Il sistema può scambiare:

- materia;

- energia.

Lo scambio di energia può avvenire sotto forma di:

- calore;
- lavoro.

Il calore ed il lavoro non sono forme distinte di energia possedute da un sistema, ma modalità diverse attraverso cui il sistema scambia un unico tipo di energia (indistinta) che custodisce al suo interno, e che si può quindi chiamare energia interna.

Gli scambi avvengono attraverso i confini. Questi ultimi possono essere classificati a seconda di quali scambi consentano. I confini immaginari consentono qualsiasi scambio, mentre la seguente classificazione riguarda le pareti reali.

Relativamente allo scambio di materia, le pareti possono essere:

- permeabili, se consentono lo scambio di qualsiasi sostanza;
- impermeabili, se non consentono alcuno scambio di materia;
- semipermeabili, se consentono il passaggio solo di alcuni tipi di sostanze.

Un classico esempio di parete semipermeabile è un setto poroso, che consente il passaggio di quelle sostanze le cui molecole sono più piccole del diametro dei pori.

Relativamente allo scambio di energia sotto forma di calore, le pareti possono essere:

- diatermiche, se consentono il passaggio di energia sotto forma di calore;
- adiabatiche, se non consentono lo scambio di calore.

Relativamente allo scambio di energia sotto forma di lavoro, si deve considerare che il lavoro è definito come forza per spostamento. Le pareti consentono lo svolgimento di lavoro se possono muoversi (o deformarsi). Quindi, si possono distinguere pareti:

- mobili, che consentono che il sistema svolga lavoro o che l'ambiente svolga lavoro su di esso;
- rigide, che non consentono lo scambio di energia sotto forma di lavoro;
- vincolate.

Quest'ultimo caso richiede un ulteriore approfondimento: una parete vincolata consente lo svolgimento di lavoro solo se la forza che ne dovrebbe causare lo spostamento è applicata lungo direzione e verso determinati. Forze di uguale intensità, applicate lungo direzioni e/o versi differenti, non producono alcuno spostamento della parete e quindi non portano ad alcuno svolgimento di lavoro. Un

esempio pratico di parete vincolata è una porta, che si apre solo se viene spinta nel verso giusto (spinta e non tirata, o viceversa).

- ***Come vengono classificati i sistemi?***

I sistemi vengono classificati a seconda di quali scambi di materia e/o energia possono avere con l'ambiente. Essi possono essere:

- aperti, se hanno scambi sia di energia sia di materia;
- chiusi, se hanno scambi di energia ma non di materia;
- isolati, se non hanno scambi né di energia né di materia.

Non è possibile che un sistema scambi solo materia ma non energia, considerando che la materia in ingresso o in uscita porta necessariamente con sé la propria energia interna. Inoltre, è necessario svolgere del lavoro per fare entrare/uscire tale porzione di materia dal sistema.

- ***Come ci si accorge che il sistema sta scambiando energia o materia con l'ambiente?***

Gli scambi che il sistema ha con l'ambiente circostante comportano un aumento o una diminuzione dell'energia che il sistema possiede. Ciò porta ad una variazione di alcune sue proprietà fisiche. Le proprietà che consentono di monitorare gli scambi che un sistema ha non possono essere definite a priori su basi teoriche. Esse vengono invece individuate empiricamente.

Tali proprietà, chiamate anche coordinate, devono essere:

- definite in modo quantitativo;
- determinabili sperimentalmente.

L'insieme dei valori numerici assunti dalle coordinate di un sistema prende il nome di stato del sistema. Avendo individuato x, y, z, w come le coordinate del sistema A, il suo stato si indica con (x_A, y_A, z_A, w_A) dove x_A, y_A, z_A, w_A sono i valori assunti dalle coordinate in un determinato momento.

Ad esempio, ci si accorge che il sistema sta scambiando materia determinandone massa composizione. Quindi, massa e composizione sono coordinate del sistema. Analogamente, il sistema potrà scambiare lavoro in dipendenza dalla sua pressione (cioè, a seconda che quest'ultima sia maggiore o minore di quella esterna). Anche la pressione entra dunque nel novero delle coordinate. La temperatura è una proprietà

che viene definita proprio per monitorare lo scambio di calore del sistema con l'ambiente.

Su questi semplicissimi concetti di base (uniti a pochi altri) uomini come Sadi Carnot, William Thomson (lord Kelvin), James Prescott Joule, Rudolf Clausius e Josiah Willard Gibbs hanno costruito una scienza, la termodinamica, che è un modo di pensare piuttosto che un insieme di nozioni. È interessante osservare che essi erano interessati a problematiche pratiche più che a speculazioni accademiche. In effetti, nella termodinamica la pratica ha percorso la teoria (la teoria è nata dalla pratica).

Conclusioni

Probabilmente (e sperabilmente), molti dei concetti esposti nelle pagine precedenti sono sembrati ovvi e scontati. Proprio questa è una caratteristica dell'approccio termodinamico al mondo che ci circonda. Essendo fortemente basato sull'esperienza, esso viene spontaneo, non è una forzatura sul nostro modo di pensare.

Tuttavia, si deve riconoscere che la voglia (o la fretta) di raggiungere obiettivi prefissati e/o posizioni ideologiche preconcepite impediscono talvolta di riconoscere ciò che la realtà intorno a noi ci dice. In tutti questi casi occorre fermarsi, definire il sistema ed analizzarlo criticamente. In questo modo è possibile raggiungere una visione affidabile di ciò che ci circonda, indirizzando correttamente i nostri giudizi e le nostre scelte. Quanto detto in questo contributo, nella sua brevità e limitatezza, vuole essere uno sprone in questa direzione.

Bibliografia

Mark W. Zemansky (1970). *Calore e Termodinamica*, Zanichelli
Engel, T., Reid, P. (2010). *Physical Chemistry*, 2nd Edition, Pearson.

Seminario integrato

Parte 2: Il Secondo Principio della Termodinamica

Roberto Di Capua

*Dipartimento di Fisica “E. Pancini”, Università di Napoli “Federico II”,
e Istituto CNR-SPIN*

La conservazione dell'energia

Nei corsi (scolastici o universitari) dedicati alla Meccanica classica, l'approccio tradizionale è quello di formulare la stessa in termini di forze applicate e conseguenti accelerazioni, tramite l'enunciazione e l'applicazione dei tre principi della Dinamica del punto materiale. In particolare, il secondo di questi principi fornisce l'equazione del moto per i sistemi puntiformi: $F = ma$, dove F è la risultante (vettoriale) delle forze agenti sul punto materiale, a la conseguente accelerazione (anch'essa grandezza vettoriale), e m è la massa del punto. Nel caso di sistemi più complessi di quelli puntiformi, le equazioni “utili” possono essere meno semplici e ovvie, ma tutto sommato hanno sempre questa equazione come base (“coadiuvata” dal Terzo Principio della Dinamica, volgarmente conosciuto come principio di azione e reazione). Talvolta possono essere complicate la risoluzione e l'impostazione fisico-matematica del problema (determinare il moto note le forze e le caratteristiche del sistema), ma è innegabile che i principi di base siano estremamente semplici!

A questa formulazione si può poi affiancare quella in termini di lavoro ed energia. Al moto dei sistemi si associa una grandezza chiamata *energia cinetica*, K , e l'azione delle forze viene descritta definendo un'ulteriore grandezza fisica, il *lavoro*, L . La variazione ΔK di energia cinetica di un qualsivoglia sistema materiale è pari al lavoro compiuto da tutte le forze agenti su esso: $L = \Delta K$. Questa equazione *non* è un ulteriore principio della Dinamica: essa costituisce un teorema (*teorema dell'energia cinetica*) che si dimostra proprio a partire dall'equazione $F = ma$. Si tratta quindi di una formulazione alternativa della Dinamica, non aggiunge niente dal punto di vista dei principi.

Orbene, per talune forze in natura la variazione di energia cinetica corrisponde alla variazione di ulteriori quantità opportunamente definite a partire da ciascuna di queste forze, e conosciute col nome generico di *energia potenziale*: per ciascuna di queste forze si definisce un'opportuna forma di energia potenziale. In che senso

“opportuna”? Per queste forze accade quanto segue: nel “corso del moto” le forze compiono lavoro, varia l’energia cinetica del sistema, e varia anche l’energia potenziale associata alle forze in gioco. Si può però dimostrare che la somma di energia cinetica ed energia potenziale (*opportunamente* definita) si mantiene costante (in altre parole: ciascuna di esse varia, ma in modo da mantenere costante la somma). Questa somma di energia cinetica del sistema ed energia potenziale associata alle forze in gioco è detta *energia meccanica*, e quanto sopra si può sintetizzare affermando che, *per queste forze*, l’energia meccanica si conserva! Per questo motivo, le forze per cui ciò si verifica (altrimenti detto: per le quali è possibile definire un’energia potenziale con questa simpatica caratteristica!) sono dette *forze conservative*, e il risultato ottenuto prende il pomposo nome di *teorema di conservazione dell’energia meccanica* (una delle tante leggi di conservazione che tanto piacciono ai fisici!).

Per fare l’esempio di un esperimento semplice, e dal risultato scontato alla luce della nostra esperienza quotidiana, si pensi al lancio in verticale di un oggetto. L’oggetto sale, diminuendo il modulo della propria velocità man mano che guadagna quota; ad un certo istante la velocità si azzerava e subito dopo l’oggetto riprende a cadere. Nell’istante in cui la velocità si azzerava, quindi, l’oggetto raggiunge la sua massima quota, e subito dopo la velocità riprende ad aumentare in modulo ma in verso opposto (comincia la caduta!). Nell’ipotesi di trascurare la resistenza offerta dall’aria (si tenga a mente questa condizione nel proseguire la lettura, ancora per qualche rigo...), l’oggetto raggiunge la quota di partenza con velocità dello stesso modulo di quella di lancio (ovviamente in verso opposto). Anzi, questo è vero per ogni quota: ad ogni quota a cui passa l’oggetto, la velocità in fase di salita e quella in fase di discesa hanno lo stesso modulo. In termini di lavoro ed energia, la descrizione è tanto semplice quanto elegante. L’oggetto, durante il suo moto, è attirato dalla Terra: in termini dinamici, diremo che su di esso agisce la forza peso. La forza peso compie lavoro, che in salita fa diminuire l’energia cinetica (legata al modulo della velocità) trasformandola in energia potenziale (legata alla quota raggiunta dall’oggetto tramite un’*opportuna* definizione); nel punto di massima quota l’energia cinetica si è azzerata (velocità nulla), e si è trasformata tutta in energia potenziale (quota massima); ricadendo, il lavoro della forza peso fa ri-aumentare l’energia cinetica (aumenta il modulo della velocità) a spese dell’energia potenziale (l’oggetto perde quota). Il fatto che ad una stessa quota si abbia la stessa velocità in modulo sia in salita che in discesa è una manifestazione del fatto che l’energia meccanica, somma delle due energie cinetica e potenziale, si conservi! Se infiliamo una tavoletta sotto l’oggetto alla massima quota, impedendogli di ricadere, e la rimuoviamo dopo 10

anni, l'oggetto ricadrà dopo 10 anni esattamente con la stessa modalità: effettivamente, in un certo senso, la forza peso “si ricorda” di aver sottratto dell'energia cinetica, e l'ha immagazzinata, *conservata* appunto, sotto forma di energia potenziale (si capisce anche il senso del nome: *potenzialmente*, la forza peso può restituirci l'energia cinetica, basta darle la possibilità di compiere lavoro, rimuovendo la tavoletta!). Nel caso di forze conservative, il lavoro è un tramite per la conversione di energia meccanica da una forma all'altra.

In natura esistono tuttavia forze per le quali il discorso appena fatto non si applica. Per esse non esiste modo di definire una energia potenziale da associarvi in maniera tale da compensare le variazioni di energia cinetica determinate dal lavoro compiuto da tali forze. La resistenza dell'aria che abbiamo trascurato nell'esempio precedente è una di queste forze: se il suo ruolo non è trascurabile, nell'esperimento di cui sopra osserviamo che l'energia cinetica si trasforma solo in parte in energia potenziale, ed una parte è effettivamente persa dal punto di vista meccanico! Ce ne accorgiamo, banalmente, dal fatto che le velocità in fase di caduta hanno modulo minore delle velocità in salita alla stessa quota. Ecco perché per il risultato di conservazione era importante trascurare la resistenza dell'aria nell'esperimento precedente! La resistenza dell'aria e in generale tutte le forze “resistive” di attrito sono dette *forze non conservative*. Esse, in particolare, dal punto di vista energetico, hanno sempre come effetto quello di fare diminuire l'energia meccanica, e sono pertanto dette *forze dissipative*. Non si pensi però che tutte le forze non conservative siano dissipative (anche se nei corsi di fisica generale queste ultime ne costituiscono il prototipo): possono anche esserci forze non conservative che aumentano l'energia meccanica, ad esempio quelle che in certi ambiti potremmo chiamare *forze motrici* (si pensi ad un motore che compie lavoro “regalando” energia meccanica ad un'automobile o a parti mobili di un impianto industriale... Liberata ora la mente dalla condizione di assenza di resistenza dell'aria di poc'anzi, tenete per un po' a mente questo motore che “regala” energia meccanica).

Per riconciliare anche le forze non conservative in un quadro più ampio di conservazione (le leggi di conservazione ci piacciono troppo...) chiediamo aiuto alla Termodinamica. Questa disciplina introduce alcuni concetti nuovi rispetto alla Meccanica, e tra questi vi sono quelli di *calore* ed *energia interna*. I sistemi fisici hanno modo di scambiarsi energia tramite scambi di calore oltre che col lavoro. Il *Primo Principio della Termodinamica* (questo sì, è un principio, non si dimostra!) lega il calore Q e il lavoro L scambiati da un sistema in un qualunque processo (trasformazione termodinamica) alla conseguente variazione ΔU della sua energia interna: $Q + L = \Delta U$. Energia interna, concettualmente parlando, è una locuzione per

indicare l'energia immagazzinata da un sistema fisico (internamente, appunto) a seguito, per lo più, di processi a livello microscopico. Ad esempio, gli atomi o molecole di un sistema sono in continua vibrazione (agitazione termica), e possiedono quindi energia cinetica: la *temperatura*, altro concetto introdotto dalla Termodinamica, è legata proprio a questa energia cinetica atomica o molecolare (aumentare la temperatura di un sistema significa aumentare l'energia cinetica media delle particelle che lo costituiscono, e quindi aumentare la sua energia interna). Un'altra forma di energia interna è l'energia chimica, che invece è più assimilabile a un'energia potenziale associata a legami chimici, che possono poi modificarsi e compiere lavoro o cedere calore (oppure un particolare stato chimico può stabilirsi se si spende lavoro e se si cede calore al sistema, che potrà poi restituire l'energia modificandosi: appunto, a tutti gli effetti è un'energia potenziale nel microscopico).

Proviamo, ad esempio, ad applicare il Primo Principio della Termodinamica ad un essere umano, tenendo presente che in quanto scritto sopra si considera $Q > 0$ se assorbito dal sistema (ed è invece assunto negativo se si tratta di calore ceduto dal sistema), ed $L > 0$ se si tratta di lavoro fatto dall'esterno sul sistema (se invece si tratta di lavoro compiuto dal sistema sull'esterno, lo si definisce negativo). Nella nostra quotidianità, noi normalmente compiamo lavoro verso l'esterno (camminiamo, spostiamo oggetti...), e quindi è normalmente $L < 0$. Altresì, normalmente cediamo calore verso l'esterno (noi siamo tipicamente a temperatura maggiore dell'ambiente esterno; in effetti, lo stanziare in ambienti estremamente caldi o l'esposizione prolungata al sole o ad altre fonti di calore è tutt'altro che confortevole), e quindi è anche $Q < 0$. Se L e Q sono entrambi negativi lo sarà anche la loro somma, e quindi dal Primo Principio della Termodinamica si ha $\Delta U < 0$: la nostra energia interna diminuisce in continuazione, perché da essa attingiamo per compiere lavoro e, anche senza far niente, la stessa diminuisce comunque per il solo fatto di essere vivi (cediamo calore perché siamo "caldi"). Come si fa, allora, per ripristinare le nostre riserve di energia interna? Occorre mangiare e respirare, per ripristinare (con i nutrienti nel cibo e l'ossigeno dell'aria) i legami chimici che abbiamo "smontato" per attingere alla nostra riserva di energia interna. Il Primo Principio della Termodinamica ci insegna che per non morire dobbiamo mangiare e respirare!

Tornando alla conservazione dell'energia, con i concetti di calore ed energia interna si può completare il quadro. Se abbiamo un processo in cui intervengono solo forze meccaniche conservative, l'energia meccanica complessiva si conserva, "palleggiandosi" tra energia potenziale ed energia cinetica per il tramite del lavoro compiuto dalle forze in gioco. Se intervengono anche forze non conservative o

scambi di calore, l'energia meccanica persa o guadagnata la si ritroverà come energia interna guadagnata o persa: il bilancio si chiude sempre in pareggio! Pur senza essere scesi in alcun dettaglio sulle varie grandezze fisiche qui menzionate, ci si può rendere conto dell'efficacia di quello che è ormai quasi un motto: *nulla si crea, nulla si distrugge, tutto si trasforma* (almeno per quanto riguarda l'energia).

Anche in questo caso può valere la pena fare un esempio con un banale esperimento di vita quotidiana. Immaginiamo una pallina di gomma lasciata cadere da una certa altezza sul pavimento. Nella caduta, come abbiamo imparato, la sua energia potenziale va trasformandosi in energia cinetica, grazie al lavoro compiuto dalla forza peso. Nell'urto sul pavimento la pallina rimbalzerà e tornerà a salire. Cosa è accaduto? L'energia cinetica si azzerà nell'impatto, e si accumula energia potenziale elastica tramite la compressione della pallina urtante (anche le forze elastiche sono conservative, e vi si può associare un'energia potenziale). Questa energia potenziale elastica è ritrasformata in energia cinetica (tramite il lavoro delle forze elastiche), e questa si ritrasforma in energia potenziale associata alla forza peso mentre la pallina risale. Tuttavia, si capisce subito che l'energia meccanica non si conserva, perché il moto di salita della pallina di gomma si fermerà prima di arrivare alla quota di partenza (alla quota massima, l'unica energia meccanica è quella potenziale di quota, ed è quindi palesemente minore di quella iniziale). Nei successivi rimbalzi si perderà sempre più energia meccanica, finché non la si perderà del tutto: alla fine, infatti, ci ritroviamo con la pallina ferma (nessuna energia cinetica) sul pavimento (nessuna energia potenziale di quota), non compressa (nessuna energia potenziale elastica). Che fine ha mai fatto l'energia meccanica? Si è trasformata in energia interna! Pavimento e pallina avranno una temperatura leggermente più alta, per l'intervento, nell'esperimento, di forze non conservative (attriti interni alla pallina durante la deformazione, attriti sul pavimento, resistenza dell'aria). Vale la pena di osservare che il risultato dell'esperimento sarà sempre questo (vi sembra scontato e banale? Non sapete cosa vi aspetta dietro l'angolo... nonostante si tratti sempre di qualcosa che avete sotto gli occhi tutti i giorni).

Per tornare al motore di qualche rigo fa che “regala” energia meccanica, abbiamo appreso abbastanza per capire che non la regala affatto: trasforma in energia meccanica una qualche altra forma di energia. Per esempio, se si tratta di un motore a combustione, esso converte in energia meccanica l'energia interna che il carburante ha immagazzinato come energia chimica (cambiando lo stato chimico del carburante: il calore prodotto converte l'energia interna in energia meccanica, che potrà essere utilizzata per produrre lavoro utile agli scopi per cui il motore è stato progettato).

La conservazione non è tutto

Il quadro appena descritto appare, anche senza scendere nei formalismi e nei dettagli dei singoli fenomeni e specifici problemi, estremamente soddisfacente ed elegante. E lo è! Tuttavia, per quanto appaia ben dipinto e completo, questo quadro non coglie qualcosa di essenziale di quanto viene osservato nell'esperienza quotidiana.

Torniamo all'ultimo esperimento descritto nella sezione precedente: la pallina di gomma, dopo alcuni rimbalzi, si è fermata a terra, col pavimento e la pallina stessa un po' più caldi. E successivamente, questo incremento di temperatura della pallina stessa e del pavimento immediatamente sottostante daranno luogo a un flusso di calore, che tenderà a portare il tutto (pallina, pavimento, ambiente) all'equilibrio termico. C'è poco da commentare dal punto di vista fenomenologico: è un esperimento dall'esito scontato. Orbene, poniamoci delle semplici domande. È possibile che si verifichi il "fenomeno inverso", cioè che del calore fluisca dall'ambiente e/o dal pavimento verso la pallina, conferendole energia interna (presumibilmente in forma di aumento di temperatura) che poi si trasforma in energia cinetica e questa, successivamente, in energia potenziale corrispondente al fatto che la pallina risalga alla quota di partenza? Avete mai osservato un fenomeno del genere? Pensate di poterlo mai osservare?

La risposta alle ultime due di queste tre domande è un no deciso: chiunque di noi, probabilmente, sarebbe disposto a scommettere sull'impossibilità di osservare un tale fenomeno, cioè la risalita *spontanea* della pallina, e nessuno lo ha mai effettivamente osservato (risalita spontanea, cioè senza intervento esterno: il fenomeno esattamente inverso a quello dell'esperimento della sezione precedente, senza interventi esterni, coi flussi di calore, lavoro, energia, semplicemente ribaltati). Ma la prima domanda, "è possibile che si verifichi...?" merita una considerazione. La Fisica è pur sempre una scienza sperimentale, ed affermare con sicurezza (probabilmente anche con un sorriso di sufficienza verso chi ci ponesse una siffatta domanda) che un fenomeno non è mai stato osservato e mai sarà osservato, equivale ad affermare che il suo verificarsi sia impossibile. D'accordo. Tuttavia, se accettiamo quanto discusso nelle pagine precedenti (anche quello è frutto dell'esperienza!), non possiamo fare a meno di notare che il trasferimento di calore dall'ambiente alla pallina con conseguente risalita della stessa *non violerebbe alcuna legge di conservazione dell'energia!* Il tutto potrebbe avvenire nel pieno rispetto delle leggi della Dinamica e della Termodinamica fin qui illustrate, con completa conservazione dell'energia (appunto, sarebbe solo una sequenza di processi inversa rispetto alla sequenza che ci è sembrata scontata, nella quale si aveva passo-passo la conversione

di energia da una forma all'altra). E allora è giusto modificare la domanda da porci: *perché il processo inverso non può avvenire? Cosa glielo impedisce?*

Non è una domanda scontata. Leggi di conservazione, leggi del moto, leggi dei processi termodinamici: non c'è nulla di tutto questo che vieterebbe il verificarsi di quello che stiamo chiamando processo inverso. Semplicemente non avviene, e basta! È scontato che sia così, perché è esperienza quotidiana ovvia, ma non è scontato perché lo sia!

Proviamo allora a immaginare un altro paio di semplicissimi esperimenti: semplicissimi perché pure essi scontati come esito, ma che di nuovo si presteranno alla stessa "imbarazzante" domanda sui processi inversi.

Consideriamo due camere riempite con due gas differenti A e B (una camera piena di A, l'altra piena di B), inizialmente separate da un rubinetto chiuso. Altresì, immaginiamo di avere un bicchiere di acqua limpida in cui stiamo per far cadere una goccia di inchiostro blu scuro. Apriamo il rubinetto tra le camere, e lasciamo cadere la goccia di inchiostro nel bicchiere di acqua. Come evolveranno i due esperimenti? Se aspettiamo per un tempo sufficiente: 1) i due gas A e B (supponiamo che non reagiscano chimicamente tra loro) si misceleranno, e nell'intero volume a disposizione si avrà un aeriforme di omogenea composizione; l'inchiostro (supponiamolo miscibile in acqua e che non precipiti) si spargerà uniformemente nel bicchiere conferendo all'acqua, inizialmente limpida, una colorazione omogenea azzurra. Certo, che altri risultati potremmo mai aspettarci? Ma... perché va sempre in questo modo? Beh, d'accordo, ci sono moti microscopici di particelle, atomi e molecole, nell'uno e nell'altro esperimento, che comportano la diffusione dei gas nel primo esperimento e dell'inchiostro nel secondo esperimento; moti che seguiranno le leggi della Dinamica con totale rispetto delle leggi di conservazioni annesse. Questo non si discute (né vogliamo discuterlo: questa trattazione non intende scendere in alcun dettaglio "tecnico"). La domanda sensata è ancora: perché non può mai avvenire il contrario, magari sempre in conseguenza di questi moti microscopici e nel rispetto delle conservazioni? Ribadiamo che sul fatto che non possa verificarsi il contrario siamo certamente tutti d'accordo: non accadrà mai che, *spontaneamente*, i due gas si separino nuovamente tornando ciascuno nella propria cameretta, o che l'inchiostro si raccolga tutto di nuovo in una goccia ripulendo l'acqua. Anzi, "addolciamo" notevolmente la domanda! Potrà accadere che anche solo in qualche regione delle due camerette si abbia anche solo per qualche secondo una netta preponderanza di un gas rispetto all'altro? O che solo in qualche regione del bicchiere si abbia anche solo per qualche secondo una colorazione nettamente più

chiara o più scura del resto del bicchiere? Sarebbe stupefacente se qualcuno rispondesse sì. Ma perché?

Riassumendo, alcuni processi non avvengono (e questo lo sappiamo dall'esperienza quotidiana: gli esperimenti proposti sono scontati!), anche se il loro verificarsi non contraddirebbe alcuna legge di conservazione o del moto o di evoluzione dei sistemi (i "perché" di cui sopra, non scontati, riguardano questo aspetto!). Possiamo circoscrivere ancora meglio il senso di queste osservazioni. Alcuni fenomeni (in realtà, in un certo senso tutti) avvengono in un verso, ma non avvengono nel verso opposto, nonostante la conservazione dell'energia e il rispetto delle leggi dinamiche e termodinamiche fin qui discusse sarebbe garantito in entrambi i versi. È questo il senso dell'irreversibilità!

A quanto pare, le leggi di conservazione (e tutte le leggi, i principi e i formalismi da cui queste leggi sono derivate) non tengono conto di questa irreversibilità dei fenomeni naturali. Ecco perché lo splendido quadro dipinto in precedenza non coglie tutto quello che c'era da raffigurare! C'è ancora qualche pennellata da aggiungere.

Il Secondo Principio della Termodinamica: irreversibilità e spontaneità

Per formalizzare quella che abbiamo sintetizzato come irreversibilità dei fenomeni fisici occorre introdurre un ulteriore principio fisico. Si tratterà certamente di un principio: non sarà ricavabile da leggi o argomentazioni pregresse, giacché le considerazioni testé sviluppate ci hanno mostrato che questa irreversibilità non è in alcun modo contemplata o inquadrabile nell'ambito dei principi o delle leggi fin qui discusse. In altre parole, non lo si "dimostra": è così, e basta!

Questo principio è quello conosciuto come *Secondo Principio della Termodinamica*.

Enunceremo ora le due formulazioni classiche, storiche, del Secondo Principio della Termodinamica, e successivamente le commenteremo alla luce delle argomentazioni a-formali che stiamo sviluppando.

Il primo enunciato è noto come enunciato di Clausius: *non è possibile realizzare una trasformazione termodinamica il cui unico risultato sia il passaggio di calore da un corpo a temperatura inferiore a un corpo a temperatura superiore*. Il secondo enunciato prende il nome di enunciato di Kelvin-Planck: *non è possibile realizzare una trasformazione termodinamica il cui unico risultato sia la completa conversione in lavoro del calore assorbito da una sola sorgente*.

In entrambi questi enunciati ci sono due parole-chiave che meritano di essere sottolineate: la prima è il "non" iniziale; la seconda è l'aggettivo "unico" riferito al risultato impossibile da ottenere in una qualsivoglia trasformazione termodinamica.

Sulla prima, c'è da osservare che entrambi gli enunciati sono “in negativo”: non dicono cosa accadrà o che condizione si verifica in un certo processo, ma ci dicono cosa certamente *non* potrà mai accadere (e già questo sembra sposarsi bene con le osservazioni della sezione precedente, sul fatto che certi processi non si verificano mai nonostante non vi sia alcuna legge che lo vieti... “prima” di questo Secondo Principio della Termodinamica). Poi, si parla di *unico* risultato: non sono possibili trasformazioni (processi) “*il cui unico risultato sia...*”. Dunque, quali sono questi risultati *unici* impossibili da ottenere? Secondo l'enunciato di Clausius, è il passaggio di calore da un corpo “freddo” a uno “caldo”. *Attenzione*: questo enunciato non afferma che non sia possibile far fluire calore da un corpo “freddo” a uno “caldo”, e infatti il frigorifero lo fa continuamente. L'enunciato di Clausius afferma che *non* può essere questo l'*unico* risultato, non può avvenire senza che “intorno” sia “cambiato” qualcos'altro. Il frigorifero non contraddice l'enunciato, perché per sottrarre calore da un ambiente freddo (il vano interno dei cibi) e immetterlo in uno più caldo (la stanza della cucina) utilizza energia elettrica, che non “restituisce” (quindi è avvenuto qualcos'altro, e ce ne accorgiamo quando il gestore di energia elettrica reclama, giustamente, il pagamento dell'energia erogata). Secondo l'enunciato di Kelvin-Planck, il risultato impossibile da ottenere come *unico* è la trasformazione completamente in lavoro del calore assorbito da un'unica sorgente. *Attenzione*: questo enunciato non afferma che non sia possibile convertire del calore in lavoro, e infatti il motore di un'automobile lo fa continuamente, muovendo l'automobile grazie al calore prodotto nello scoppio della benzina. L'enunciato di Kelvin-Planck afferma che *non* può essere questo l'*unico* risultato, deve essere “accaduto” qualcos'altro. Il motore dell'automobile non contraddice l'enunciato, perché nel convertire in lavoro meccanico (l'automobile si muove) il calore prodotto a spese dell'energia interna del carburante, “sparge” calore nell'ambiente circostante (se toccate il cofano dell'automobile vi accorgete che è caldo).

A questo proposito, facciamo due osservazioni.

La prima riguarda il fatto che i due enunciati sono completamente equivalenti. La dimostrazione viene fatta semplicemente mostrando che se esistesse una “macchina” in grado di contraddire uno dei due enunciati, con la stessa si potrebbe contraddire anche l'altro.

La seconda osservazione ci interessa di più, in questa sede. Essa concerne il fatto che quanto commentato sui due enunciati del Secondo Principio della Termodinamica, in particolare nella versione di Clausius, è intimamente legato al discorso sull'irreversibilità dei fenomeni fisici con cui abbiamo sintetizzato i nostri banali esperimenti (non sembrano più tanto banali, vero?), e alla spontaneità o meno di certi

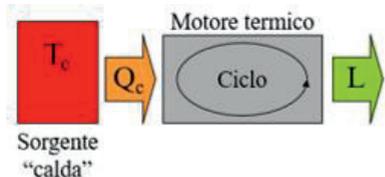
processi. Infatti, se mettiamo a contatto due corpi a temperatura diversa, il calore fluisce *spontaneamente* dal corpo “caldo” a quello “freddo”: l’enunciato di Clausius ci insegna che il processo contrario non può avvenire (irreversibilità!), e che per “riportare indietro” il calore non potete sperare che avvenga spontaneamente, dovete attaccare la spina del frigorifero! Altresì, in un esperimento come quello della pallina rimbalzante, o in una qualunque situazione che comporti attrito meccanico (come l’arresto di una ruota messa in moto con una pedalata), il lavoro meccanico si trasforma spontaneamente in calore, mentre il contrario non può avvenire spontaneamente (potete certamente risollevare la pallina, ma dovete far lavoro a spese della vostra energia interna per sollevarla). L’irreversibilità di cui parlavamo corrisponde al fatto che certe trasformazioni rispettano, è vero, tutte le leggi di conservazione in qualunque verso avvengano, ma tuttavia in un verso avvengono spontaneamente e nell’altro no; e, tutto sommato, questa è propria è proprio l’essenza contenuta nei due enunciati classici del Secondo Principio della Termodinamica.

Altre implicazioni e interpretazioni del Secondo Principio della Termodinamica

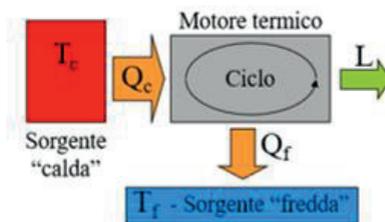
Riprendiamo uno qualunque dei semplici esperimenti finora considerati, per esempio quello della goccia di inchiostro che diffonde nel bicchiere di acqua limpida. Osservando l’evoluzione dell’esperimento vediamo che l’inchiostro è inizialmente concentrato in un piccolo volume, poi la zona colorata comincia ad allargarsi e a sbiadirsi, e continua a farlo sempre più, finché non distinguiamo più tra zona colorata e zona limpida, e ci ritroviamo con un liquido uniformemente colorato. Abbiamo già sottolineato quanto sia ovvio questo esito e questa evoluzione. Cosa pensereste, invece, vedendo un bicchiere di acqua azzurrognola sbiadire in alcune zone, col colore che progressivamente va concentrandosi, scurendosi, in una zona sempre più ristretta? Probabilmente pensereste (con ragione!) a un filmato proiettato al contrario. Ma... che vuol dire proiettato al contrario? Vuol dire che si vede prima il “futuro”, e poi il “passato”? Queste considerazioni ci offrono un’importante rivelazione: la nostra percezione dello scorrere del tempo è legata all’irreversibilità dei fenomeni naturali! Se il ritrarsi della goccia di inchiostro fosse un fenomeno parimenti verificabile come la sua diffusione, non potremmo avere la sensazione sopra descritta come di filmato proiettato al contrario. Forse non è del tutto corretto affermare che non potremmo definire un passato e un futuro, concetti legati a relazioni di causa ed effetto che potremmo far salve in fenomeni fisici indipendentemente dall’idea di irreversibilità; ma è certo che non potremmo dire che il tempo scorre solo in un verso (quello in cui l’inchiostro diffonde, il verso in cui avvengono i fenomeni spontanei), non nell’altro (quello in cui l’inchiostro dovrebbe

ritrarsi, quello in cui dovrebbero avvenire i fenomeni inversi di quelli spontanei)! La cosiddetta freccia del tempo è legata all'irreversibilità, e quindi al Secondo Principio della Termodinamica! Per vedere la cosa sotto un altro punto di vista, ricalcando quanto detto prima sulle leggi di conservazione che non privilegiano alcun verso di occorrenza dei fenomeni fisici, possiamo affermare che qualunque legge fisica dipendente dal tempo che possiamo pensare di scrivere (equazioni del moto, o di diffusione del calore), non distinguono, matematicamente parlando, i due versi del tempo, così come non distinguono versi in una coordinata spaziale. Eppure, noi possiamo muoverci e guardare a destra e a sinistra, in alto o in basso, avanti o indietro; invece, possiamo guardare solo nel passato e non nel futuro, e possiamo muoverci solo nel futuro e non nel passato: cosa rompe questa simmetria, se non c'è niente nelle equazioni della Meccanica e della Termodinamica che lo faccia dal punto di vista formale? Ecco, quindi, un altro aspetto, ovvio per quel che riguarda l'esperienza comune, che necessita del Secondo Principio della Termodinamica per essere contemplato nella nostra formulazione della fisica classica.

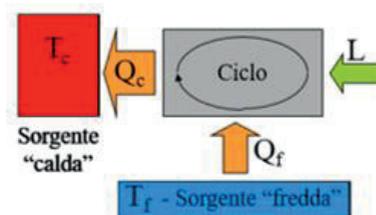
Proviamo ad essere un po' più prosaici, trattando di implicazioni tecniche, che in effetti costituiscono il senso dei due enunciati storici del Secondo Principio della Termodinamica. Un motore termico è una macchina che produce lavoro meccanico "utile" assorbendo calore da una sorgente termica (ad esempio, da carburante che brucia: sostanzialmente converte energia termica in energia meccanica; anche in questo caso non siamo interessati ad alcun dettaglio tecnico, ma solo all'idea generale e al principio). Nel pannello "a" della figura a lato è illustrato schematicamente il funzionamento di un motore termico ideale: esso preleva calore Q_c da una sorgente termica, e, attraverso un qualche processo ciclico (che lo riporti, cioè, nello stato iniziale, quello che aveva prima di assorbire calore), lo converte in lavoro L . Che il motore lavori in maniera ciclica è importante: ci farebbe piacere utilizzare il



a) Motore termico che non rispetta il Secondo Principio della Termodinamica



b) Motore termico reale

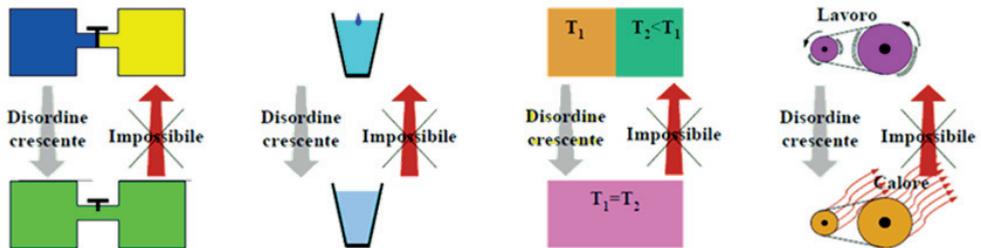


c) Frigorifero o pompa di calore

motore per molto tempo, non una volta sola, ed è quindi importante che dopo aver compiuto una conversione $Q_c \rightarrow L$ esso sia pronto a ripeterne un'altra, e un'altra ancora, e così via. Ciò posto, un motore siffatto è bello e impossibile: esso contraddirebbe infatti l'enunciato di Kelvin-Planck, perché convertirebbe il calore assorbito da una sola sorgente interamente in lavoro senza che accada null'altro (perché si riporta nello stato iniziale). La schematizzazione più semplice di un motore termico reale, allora, è quella riportata nel pannello "b": solo una parte del calore assorbito Q_c è convertito in calore L ; il resto è "buttato via" come calore, Q_f , verso una sorgente più fredda (si pensi di nuovo al motore dell'automobile: una parte del calore assorbito dallo scoppio del carburante è dissipato nell'ambiente). Ovviamente sarà $|Q_c| = |L| + |Q_f|$ (la conservazione dell'energia è sempre valida!). È naturale definire il rendimento del motore come il rapporto tra quello che "si ottiene" (L) e quello che "si spende" (Q_c). In un motore ideale, in cui $|Q_f| = 0$ e $|L| = |Q_c|$, il rendimento sarebbe del 100% ($|L|/|Q_c| = 1$). In un motore reale, invece, questo rendimento è necessariamente minore del 100% ($|L|/|Q_c| < 1$ perché $|Q_f| \neq 0$). Un'altra implicazione del Secondo Principio della Termodinamica, conseguenza direttissima dell'enunciato di Kelvin-Planck, è l'impossibilità di realizzare un motore termico ideale, avente rendimento del 100%. Il pannello "c" della figura riporta lo stesso schema del pannello "b" con "tutte le frecce al contrario". Si tratta di una macchina che, funzionando ciclicamente (stesse considerazioni di prima sul fatto di volerla utilizzare ripetutamente), estrae calore Q_f da un ambiente "freddo" e immette calore Q_c in un ambiente "caldo": un frigorifero, o una pompa di calore (come idea generale sono la stessa macchina: se siamo interessati a quanto calore sottraiamo all'ambiente freddo la chiamiamo frigorifero, se siamo interessati a quanto calore immettiamo nell'ambiente caldo la chiamiamo pompa di calore). Per realizzare questo trasferimento di calore, che non si verificherebbe spontaneamente, il frigorifero richiede che si compia lavoro dall'esterno (la famosa spina attaccata...), L . Vale sempre $|Q_c| = |L| + |Q_f|$, per la conservazione dell'energia ($|L|$ e $|Q_f|$ "entrano", $|Q_c|$ è quello che "esce", e il bilancio deve essere in pareggio perché niente altro cambia visto che la macchina funziona ciclicamente). Sarebbe bello che il frigorifero (o la pompa di calore) funzionasse senza bisogno di spenderci lavoro: un frigorifero ideale, la cui esistenza contraddirebbe però l'enunciato di Clausius, perché realizzerebbe un trasferimento di calore da una sorgente "fredda" a una "calda" senza che accada niente altro ($L = 0$, motore ciclico: non sarebbe cambiato niente altro, a parte il calore trasferito dal "freddo" al "caldo"). Il Secondo Principio della Termodinamica ci dice anche, direttamente con l'enunciato di Clausius, che non può esistere il frigorifero (o la pompa di calore) ideale.

Disordine ed entropia

Facciamo mente locale sugli esperimenti illustrati, per mostrare un altro punto di vista sotto cui guardare l'irreversibilità e il senso del verso privilegiato in cui avvengono i fenomeni fisici. La figura che segue riassume sinteticamente quanto detto sulla diffusione dei gas nelle due camere e dell'inchiostro nel bicchiere, sul passaggio di calore tra corpi a contatto, sulla trasformazione di lavoro in calore (di energia meccanica in energia interna). Ciascuno di questi "esperimenti" presenta un verso in cui il processo avviene spontaneamente, mentre il processo inverso non si verifica.



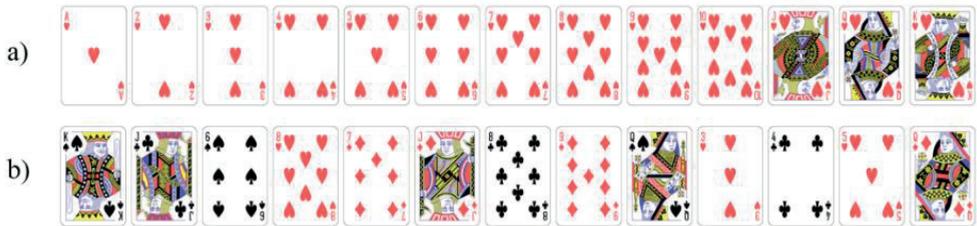
Senza scendere nei formalismi, così come stiamo facendo in tutta questa trattazione, potremmo affermare che il procedere spontaneo dei fenomeni illustrati avviene nel verso che conduce i sistemi fisici verso uno stato più... disordinato! All'inizio possiamo distinguere i due gas, e l'inchiostro dall'acqua, mentre successivamente è tutto mischiato, indistinguibile... confuso! All'inizio distinguiamo il sistema caldo da quello freddo, poi diventano (almeno per quanto riguarda la temperatura), indistinguibili. E nell'ultimo esperimento, in cui forse questa evoluzione spontanea verso il disordine è meno palese dal punto di vista intuitivo, l'energia meccanica "utile", "spendibile", che abbiamo all'inizio, è convertita in energia interna di tipo termico, la quale, a dar retta all'enunciato di Kelvin-Planck, non è altrettanto "utile". Insomma, sembra che i tanti stati in cui un sistema potrebbe trovarsi compatibilmente con la conservazione dell'energia non siano equivalenti: quello a cui corrisponde la situazione di massimo disordine è in qualche modo "privilegiato" rispetto agli altri. Vale la pena di ribadire quanto scritto sopra: il discorso sul disordine è qualitativo e intuitivo, così come lo sono i discorsi, presentati in certe trattazioni sull'argomento, sulle analogie "quotidiane" (su cassette e scrivanie che tendono a disordinarsi se non ci si mette di impegno a ricostituire l'ordine: non si riordinano *spontaneamente*). Niente di male in questo, almeno al livello di questa trattazione. È però importante specificare che ciò non significa che il concetto non possa essere formalizzato in maniera rigorosa! Con estremo rigore, infatti, si può definire una grandezza fisica

chiamata *entropia*, S . Si tratta di un termine che probabilmente tutti quelli che stanno leggendo hanno già sentito nominare, e con tutta probabilità non suona nuova nemmeno la sua associazione al grado di disordine di un sistema fisico (qualunque cosa significhi, formalmente, questa locuzione). Questa associazione, alla luce dei discorsi qualitativi sulla spontanea evoluzione verso stati disordinati, ci aiuta a cogliere il senso dell'espressione formale del Secondo Principio della Termodinamica, pur mettendo da parte tutti i formalismi, che non ci interessa sviluppare in questa sede. In effetti, i lettori più attenti avranno notato che non abbiamo scritto né fatto menzione ad alcuna espressione matematica che descriva il Secondo Principio della Termodinamica, da affiancare all'equazione $F = ma$, a quella del Primo Principio della Termodinamica, e alle varie equazioni di conservazione. Orbene, non v'è alcuna equazione che assolva a questo scopo, bensì una disequazione! E non poteva essere altrimenti, visto che il senso profondo del Secondo Principio della Termodinamica, alla luce di tutto quello che abbiamo osservato finora, è quello di una rottura di simmetria (freccia del tempo, verso delle trasformazioni spontanee) (e i lettori veramente attenti avranno notato il non casuale utilizzo del termine "espressione" invece di "equazione" per ben due volte nelle ultime righe!). I due enunciati del secondo principio, così come tutte le altre considerazioni illustrate, possono essere formalizzate matematicamente affermando che *l'entropia di un sistema isolato non può diminuire*: $\Delta S \geq 0$ (per un sistema isolato). È importante specificare "per un sistema isolato": con un intervento esterno si possono ri-separare i gas miscelati o l'inchiostro dall'acqua, o risollevarne la pallina da terra, o... Semplicemente, tutti questi processi "inversi" non avvengono spontaneamente come accade invece per quelli "diretti"! A enunciati in negativo, su cosa *non* può accadere invece che su cosa accade, corrisponde una disequazione invece di un'equazione: tutto sommato sembra coerente! E anche se non abbiamo speso una virgola sull'idea formale di entropia, la sua identificazione col disordine ci aiuta a cogliere il senso di tale risultato formale: i sistemi fisici tendono a evolvere *spontaneamente* verso stati più disordinati, e a questo livello tanto ci basta e ci soddisfa! Si tenga presente che il caso "limite" $\Delta S = 0$ rappresenta un'idealizzazione, valida solo per processi interamente reversibili; ma ogni processo reale macroscopico presenta un qualche grado di irreversibilità.

Entropia e probabilità

Il concetto di entropia può essere visto sotto un differente punto di vista. Per comprenderne l'essenza anche senza introdurre formalismi, poniamoci un paio di semplici domande. Si considerino le due seguenti "cinquine": a) 1 - 2 - 3 - 4 - 5; b)

16 - 39 - 77 - 24 - 50. Quale di queste due cinquine ha maggiore probabilità di essere estratta su una ruota del lotto? Poniamoci una domanda del tutto analoga sulle due mani di 13 carte riportate nella figura seguente:



In entrambi i casi, la risposta è: le due combinazioni sono equiprobabili, e questo non dovrebbe essere nuovo né difficile per la maggior parte di coloro che staranno leggendo. I lettori che ritengono le combinazioni “b” più probabili potranno forse cogliere in quanto scritto di seguito il comune equivoco alla base di questa errata convinzione. In effetti, senza girarci troppo attorno, per ricollegarci a quanto illustrato sopra su entropia e disordine, le combinazioni “a” appaiono estremamente “ordinate” rispetto alle “anonime” combinazioni “b”. Eppure, la sequenza di 13 carte di cuori ha la stessa probabilità di presentarsi della “anonima” sequenza riportata in riga “b”. Ciò che ci porta, istintivamente ed erroneamente, a considerare non equiprobabili le due mani (o le due cinquine) è la nostra tendenza alla “categorizzazione”: la mano “a” rappresenta 13 carte dello stesso seme, una distribuzione di tipo 13-0-0-0 tra i 4 semi; la mano “b” è invece un esempio di distribuzione 4-3-3-3 (ci sono 4 quarte di quadri, e 3 carte per ciascuno degli altri semi). Orbene, con le 52 carte del mazzo francese, vi sono circa 635 miliardi di combinazioni possibili per una mano di 13 carte (stiamo considerando equivalenti le permutazioni delle stesse 13 carte), e soltanto una di queste realizza la mano con 13 cuori! Solo 4 combinazioni (una per seme) su 635 miliardi corrispondono ad una distribuzione 13-0-0-0. Invece, la combinazione con 4 quadri e 3 carte degli altri tre semi sono circa 16.7 miliardi, e tutte le possibili 4-3-3-3 sono quasi 67 miliardi, che rappresentano circa il 10.5% delle combinazioni totali. Una probabilità del 10.5% della “categoria” 4-3-3-3 è molto più della probabilità irrisoria di 4 su 635 miliardi della 13-0-0-0 (e non è nemmeno la distribuzione più probabile: la 4-4-3-2 ha probabilità più che doppia di verificarsi). Ecco quindi spiegato l’arcano: ogni “tredicina”, anche la sequenza di 13 cuori, ha la stessa probabilità di verificarsi di tutte le altre; ma esistono solo 4 sequenze che realizzano una 13-0-0-0, mentre ce ne sono 67 miliardi che realizzano una 4-3-3-3. Ciascuna delle singole 4-3-3-3 è equiprobabile a una singola 13-0-0-0, ma sono molte di più! Quindi, è corretto

pensare che la distribuzione 4-3-3-3 sia enormemente più probabile della 13-0-0-0, nonostante due singole combinazioni da 13 carte siano sempre equiprobabili. Se giocate a bridge, una 4-3-3-3 vi uscirà di sovente; ma provate un po' a valutare voi per quanto tempo dovrete giocare per avere una significativa probabilità di vedere una sola volta una 13-0-0-0!

Su carte e numeri del lotto, la “sequenza ordinata” è definita da un particolare significato che attribuiamo noi alle singole carte o ai singoli numeri (carte di un seme, oppure numeri consecutivi), ma quanto illustrato dovrebbe illuminare sul modo alternativo di intendere l'entropia. Considerando, a titolo di esempio, l'esperimento dei due gas A e B nelle due camerette, la distribuzione di molecole con tutte le A in una camera e tutte le B nell'altra camera è equiprobabile a una qualunque distribuzione che preveda molecole omogeneamente mescolate; semplicemente, il numero di distribuzioni a cui corrisponde mescolamento omogeneo è enormemente più alto delle distribuzioni molecolari che danno gas separati! Detto con linguaggio un po' più tecnico: gli *stati microscopici* (le distribuzioni di molecole, o di carte) sono tutti equiprobabili, ma ne esistono molti di più che corrispondono allo *stato macroscopico* di gas mescolati (o di 4-3-3-3) rispetto a quelli corrispondenti a gas separati (o 13-0-0-0). L'entropia può essere, in effetti, definita anche in maniera probabilistica: essa è maggiore per gli stati macroscopici che si realizzano con più combinazioni microscopiche (rileggete l'ultimo rigo della sezione precedente: è stato piazzato non a caso un aggettivo “macroscopico”, prima di questo discorso; può darsi che al momento non ci abbiate fatto caso, ma riguardate ora il contesto in cui è stato inserito). Così ragionando, si può affermare che non è impossibile che i gas si separino spontaneamente: semplicemente, è estremamente improbabile! (attenzione: se con appena 52 carte abbiamo ottenuto una sproporzione del tipo 4 combinazioni contro 67 miliardi, le sproporzioni che si ottengono con le, diciamo, 10^{22} molecole di un gas in una cameretta da banco sono inimmaginabilmente grandi; continuate tranquillamente a considerare impossibile la separazione spontanea!)

Osservazioni e note finali

Oltre a quanto visto dal punto di vista squisitamente fisico e tecnologico (irreversibilità, spontaneità, freccia del tempo, motori e frigoriferi), il Secondo Principio della Termodinamica e la nozione di entropia hanno implicazioni in campi diversissimi.

Una reazione chimica, ad esempio, può avvenire in due versi (da prodotti a reagenti, o viceversa). Essa avviene, generalmente parlando, nel verso in cui diminuisce quella

che si chiama *energia libera* G , la cui variazione si può scrivere come: $\Delta G = \Delta H - T \cdot \Delta S$, dove T è la temperatura assoluta e ΔH può essere visto come il calore sviluppato dalla reazione (se positivo, è calore assorbito, mentre è negativo nel caso di calore ceduto). Se la reazione produce calore ($\Delta H < 0$) e produce uno stato più disordinato ($\Delta S > 0$), sicuramente è $\Delta G < 0$ e la reazione è favorita: in effetti, le reazioni sono favorite quando sono esotermiche (cessione di calore, i sistemi tendono ad abbassare la propria energia) e quando evolvono verso stati disordinati (il Secondo Principio della Termodinamica dice sempre la sua!). Se la reazione assorbe calore e produce stati più ordinati, è tutto il contrario, la reazione non avviene in quel verso ma nel verso opposto. Ma cosa accade in una situazione “mista”, in cui ΔH e ΔS hanno lo stesso segno? Se sono entrambi positivi, ad esempio, il primo termine tende a dare un $\Delta G > 0$ e quindi a sfavorire la reazione, il secondo tende invece a favorire la reazione perché darebbe $\Delta G < 0$. Cosa accade in questo caso? Dipende dal valore di T , che rende il termine $T \cdot \Delta S$ più o meno “pesante”. Un esempio? Il più familiare è di un processo fisico più che chimico, ma per dare l’idea va bene lo stesso: la fusione del ghiaccio in acqua. Per fondere, il ghiaccio deve assorbire calore, quindi in questo passaggio è $\Delta H > 0$; ma si passerebbe verso uno stato più disordinato (le molecole sono più ordinate in un solido che in un liquido), e quindi in questo passaggio è $\Delta S > 0$. Ci si rende allora conto che se T è “abbastanza alta”, allora “vince” il termine $T \cdot \Delta S$ e la reazione può avvenire, il ghiaccio può fondere; viceversa, se T è “bassa” allora prevale il termine ΔH sfavorevole e il ghiaccio non fonde. E infatti, sappiamo tutti che per la fusione del ghiaccio (o di altro solido) ci vuole una temperatura “sufficientemente” alta! Il termine entropico definisce la temperatura di fusione.

Noi stessi siamo sistemi fisici che ubbidiscono alle leggi fisiche, ivi incluso il Secondo Principio della Termodinamica. Anche noi, quindi, tendiamo ad evolvere verso stati più... disordinati, e siamo soggetti allo scorrere del tempo sempre nello stesso verso: in parole più povere, invecchiamo, e non possiamo ringiovanire! Certo, non siamo un sistema isolato, e abbiamo imparato a intervenire su noi stessi, ma a lungo andare la tendenza generale resta pur sempre quella di un aumento dell’entropia. In sostanza, se il Primo Principio della Termodinamica ci ha insegnato che dobbiamo mangiare per non morire, il Secondo Principio della Termodinamica ci insegna che moriremo ugualmente!

Come menzionato per noi stessi nell’esempio precedente, ogni sistema fisico è potenzialmente non isolato. Ogni sistema fisico ha un proprio “esterno”. A meno che... non consideriamo come sistema tutto ciò che esiste: l’Universo, qualunque

cosa sia! Più isolato di così! Allora, possiamo dire che *l'entropia dell'Universo non può che aumentare*: il Secondo Principio della Termodinamica ci insegna che l'Universo in persona sta evolvendo verso uno stato sempre più disordinato, irreversibile. Insomma, anche se imparassimo a intervenire “dall'esterno” su noi stessi in modo da diminuire la nostra entropia (ringiovanire...), l'Universo nel suo complesso non può che invecchiare. Ma qui ci fermiamo: questo discorso è davvero al di là delle intenzioni di questa breve trattazione.

Concludiamo con una nota letteraria, legata a quanto appena menzionato sul destino dell'Universo e sul concetto di entropia in generale. Più o meno tutti conoscono il nome di Isaac Asimov, ed è difficile non considerarlo il più grande scrittore di fantascienza di tutti i tempi (era anche uno scienziato). Oltre ai suoi notevoli romanzi, Asimov ci ha lasciato una serie di racconti su vari argomenti e di diversa consistenza: racconti che fanno parte di una saga o legati da un filo comune, oppure racconti sciolti, di lunghezza da una pagina a veri e propri romanzi brevi. Buona parte della critica letteraria considera “Notturmo”, di Asimov, il più bel racconto di fantascienza mai scritto. Naturalmente, si può essere d'accordo o meno su questo giudizio. Uno di quelli che non era d'accordo era lo stesso Asimov, visto che lui considerava un altro come il suo racconto più bello: *L'ultima domanda*. Non sono pochi i lettori, incluso chi sta scrivendo la presente dissertazione, a pensarla come Asimov. Orbene, *L'ultima domanda* è un racconto incentrato sul concetto di entropia e sulle implicazioni sul destino dell'Universo. Non è il caso di scrivere altro: è molto meglio lasciare ai lettori il gusto di poterlo leggere. Aggiungendo però un consiglio: leggetelo quando siete soli, magari di notte, quando nessuno può disturbarvi, per assaporarlo meglio, e vivere davvero il dilemma dell'ultima domanda. Il racconto è breve, letto di un fiato richiede un tempo dell'ordine dei 5 minuti: ma se deciderete di leggerlo assaporando parola per parola, ci metterete qualche minuto in più, ma non ve ne pentirete!

Il futuro del sottosuolo: le sue risorse energetiche ed i suoi rischi

David Iacopini

*Dipartimento di Scienze della Terra, dell'Ambiente e delle
Risorse, Università di Napoli Federico II*

Abstract

Il sottosuolo terrestre rappresenta da circa due secoli il luogo da cui si estraggono le risorse minerarie alla base di gran parte delle fonti energetiche in uso nella nostra società moderna. Il loro utilizzo ha permesso a gran parte degli stati occidentali di svilupparsi, crescere ed evolversi tecnologicamente con una rapidità senza precedenti: grazie allo sfruttamento di energie fossili e minerarie dal 1850 ad oggi l'energia prodotta è decuplicata da poche migliaia di TeraWatt/ora (TWh) a centinaia di migliaia di TWh e nell'ultimo secolo, in coppia con la finanza moderna, ha permesso una crescita della popolazione di tipo esponenziale senza precedenti, con un tasso pari al miliardo di popolazione ogni 25 anni. Tuttavia, l'utilizzo intensivo di tali energie ha innescato un rapido processo di riscaldamento superficiale terrestre globale che sta modificando irreversibilmente la composizione chimica del nostro sistema atmosferico con un impatto sempre maggiore sulla società umana e di difficile previsione. Questa nuova condizione sta ponendo importanti questioni che riguardano lo sfruttamento delle risorse minerarie e sul loro ruolo in una transizione energetica che mira a sviluppare una produzione ed un consumo più sostenibile e meno inquinante delle sue risorse energetiche. Mostriamo perciò il potenziale ruolo che la conoscenza e l'esplorazione del sottosuolo, grazie alle avanzate tecnologie di esplorazione, potrà giocare durante la transizione energetica e quale sarà il suo nuovo ruolo nel sostenere lo sviluppo di energie rinnovabili ed alternative.

Introduzione: il sottosuolo cosa ne sappiamo?

Il sottosuolo definito come il complesso degli strati del terreno che si trovano sotto la superficie terrestre, rappresenta un'area immensa, attiva e dinamica (e.g terremoti e vulcani) che da sempre ha rappresentato una sorgente energetica importante ma anche fonte di rischio per le civiltà umane. Attraverso il sottosuolo (poche centinaia di metri) superficiale come luogo di risorse idriche per l'agricoltura e per lo sviluppo dei primi agglomerati urbani, sino alle strutture più profonde (km), intese come

luogo di risorse minerarie (carbone, metalli, terre rare) e di idrocarburi, le geotecnologie hanno rappresentato un importante strumento e mezzo tecnologico per molte delle società del mondo occidentale (dalla rivoluzione industriale sino alle società contemporanee). In questo contesto sebbene la parola sottosuolo rappresenti in senso lato tutto ciò che giace sotto il suolo sino al suo nucleo terrestre (per uno spessore totale di oltre sei mila km), dal punto di vista operativo in un contesto energetico, quando parleremo di sottosuolo ci riferiremo a quella porzione dello strato più esterno della crosta terrestre esplorabile con mezzi geofisici e raggiungibile attraverso miniere e profonde trivellazioni.

Sempre dal punto di vista operativo sono ancora troppo poche le informazioni dirette disponibili che riguardano il sottosuolo: i pozzi più profondi hanno raggiunto profondità di circa poco più di 12 km sotto il suolo. Il pozzo Kola (Russia) rappresenta il pozzo più profondo a terra-12.262 m. Al Shaheen Oil field nel Qatar e il pozzo di Exxon presso Sakhalin in Russia invece hanno al momento raggiunto le lunghezze massime di perforazione rispettivamente di 12.289 m e 12.376 m (Fig 1).

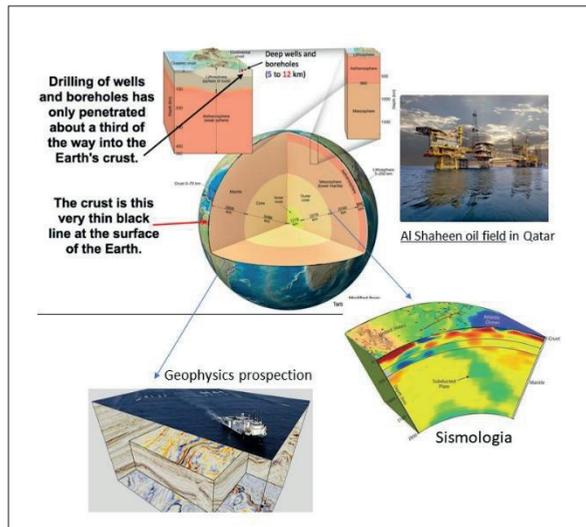


Figura 1. Struttura del sottosuolo terrestre e tecniche di esplorazione ed imaging

In pratica come mostrato in Figura 1, queste profondità corrispondono ad aver grattato la sola ‘pellicola superficiale’ della Terra (12 km a fronte dei 6000 km del raggio terrestre). Tuttavia la comprensione di parte delle strutture più profonde ci è stata permessa attraverso l’utilizzo di tecniche di cartografia geologica su catene montuose come gli Appennini, Alpi, Himalaya, Ande o attraverso catene più antiche e totalmente erose a formare cratoni di cui i continenti più antichi quali Africa,

Australia o Siberia per esempio, ne sono ricchi. Il motivo sta nel fatto che tali catene rappresentano porzioni fossili di crosta superiore ed inferiore (equivalenti a rocce oggi presenti alle nostre profondità di 60-100 km) esumata attraverso potenti meccanismi di collisione tettonica e smantellate da lenti processi di erosione.

I metodi di indagine geologica di tali catene, perciò, ci consentono di analizzare porzioni che corrispondono a quello che oggi accade a circa 60-120 km di profondità. Anche le lave vulcaniche possono darci ulteriori informazioni preziose: spesso attraverso il loro condotto vulcanico strappano frammenti di rocce di crosta profonda, chiamati xenoliti, i quali rappresentano importanti informazioni dal sottosuolo profondo (20-120 km). Se vogliamo invece esplorare, fotografare e caratterizzare cosa succede oggi al di sotto della crosta superiore, gli unici mezzi di indagine sono rappresentati da tecniche di indagine geofisica che rappresentano però mezzi di indagine indiretta. Indiretta, perché non ci permettono di toccare o campionare frammenti di roccia; ci permettono invece di ricostruire immagini e rappresentazioni visive prodotte da interazioni di onde fisiche in grado di tracciare e memorizzare proprietà fisiche: analisi termiche (attraverso l'uso di sensori termici in pozzi e pozzetti), utilizzo di proprietà di anomalie gravimetriche (densità delle rocce), utilizzo di onde elettromagnetiche elastiche ma anche di geo-neutrini (particelle prodotte per decadimento da minerali del nucleo e mantello terrestre). Tuttavia, il mezzo più potente che ci permette di mappare il sottosuolo con una precisione ancora insuperata (risoluzione da pochi metri a pochi km a seconda delle frequenze e profondità di indagine) è rappresentato dalle onde sismiche. Un po' come stessimo producendo ecografie del sottosuolo, si utilizzano le onde sismiche prodotte da potenti esplosioni (*geophysical prospecting*) o terremoti (in questo caso si parla di sismologia) che penetrano nel sottosuolo e tornano indietro per riflessione o diffrazione indietro ricche di informazioni geofisiche. Tramite l'utilizzo di metodi di analisi di tali segnali e grazie a tecniche di ricostruzione di immagini, si ricostruiscono le caratteristiche fisiche e geometriche del sottosuolo (Fig 1). Tutte queste metodologie descritte ci hanno permesso e ci permettono di costruire il modello rappresentato in Figura 1 che raffigura la struttura della Terra ben nota, e diffusa in tutti i libri di geologia e geografia fisica, ed essenzialmente composta da:

- crosta terrestre superiore ed inferiore (1-100 km) a composizione media granitica,
- mantello con la sua componente astenosferica (strato fluido-viscoso, situato al di sotto della crosta terrestre) di spessore di oltre 600 km ed a composizione ultrabasica (silicati di ferro e magnesio),
- mantello inferiore (spessore di circa 2000 km) definito da minerali ultrabasici ad alta pressione,

- nucleo esterno (spessore di circa 2200 km) di minerali di ferro e solfuro di ferro e nucleo interno (1200 km) essenzialmente modellizzato da minerali di lega ferro-nichel.

Di tutte queste strutture profonde noi ci interesseremo essenzialmente della parte che maggiormente ha influito ed interagito con la nostra civiltà, per necessità economica ed energetica e cioè della parte superiore della crosta terrestre. Infatti, la totalità delle esplorazioni geofisiche più avanzate e le loro relative rappresentazioni tridimensionali del sottosuolo a scopo estrattivo si sono concentrate nei primi 50 km della crosta terrestre.

Perciò fatto un rapido quadro generale su cosa veramente sappiamo della crosta terrestre, come l'esploriamo, è tempo di cambiare prospettiva e di concentrarci invece sull'impatto economico che il sottosuolo e le sue risorse hanno avuto nella nostra società nei recenti due secoli e sul ruolo che le tecnologie del sottosuolo potranno ancora avere in un futuro sempre più incentrato nelle energie definite rinnovabili perché direttamente dipendenti dall'irraggiamento solare (solare, eolico ma anche indirettamente biomasse)

Sottosuolo, energia e popolazione

Un aspetto che spesso sfugge a chi studia la storia economica delle nostre civiltà riguarda il ruolo ed effettivo uso delle fonti energetiche in gran parte degli imperi ma anche durante lo sviluppo degli agglomerati urbani negli ultimi 10.000 anni. Si tratta di aspetti importanti che ci danno una chiave di lettura per decifrare quelli che sono stati i meccanismi economici e scelte geopolitiche della recente storia moderna ma ci permettono anche di interpretare il motivo per cui alcune società si sono evolute rapidamente. Non c'è dubbio infatti che la tecnologia e l'accesso o controllo di risorse energetiche abbiano contribuito al predominio militare e commerciale di alcuni imperi o nazioni anziché altre. In questo breve contributo non avremo spazio per dilungarci ed affrontare in dettaglio i vari aspetti di cui sopra, ma cercheremo di analizzare che impatto ha avuto l'energia derivante dal sottosuolo nel recente sviluppo della società occidentale specialmente negli ultimi due secoli.

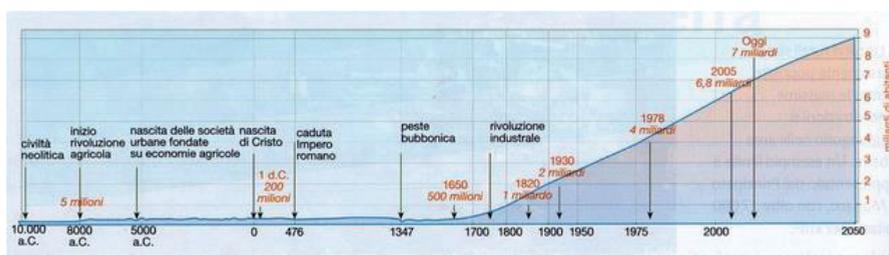


Figura 2. Stima dell'evoluzione della popolazione globale dal 10000 aC sino al 2050.

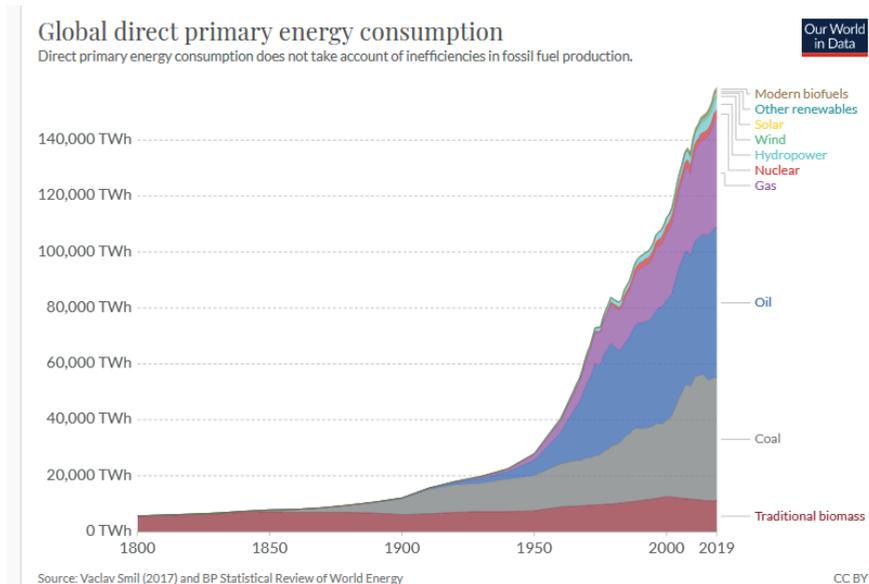


Figura 3. Distribuzione del consumo energetico primario (senza tenere conto delle eventuali inefficienze produttive delle risorse fossili) dal 1800 al 2019. $1 \text{ Wattora} = \text{Wh} = 3.6 \text{ Kjoule}$, intesa come un'unità di misura derivata dell'energia, definita come l'energia complessiva fornita qualora una potenza di un watt (W) sia mantenuta per un'ora (h). $\text{TWh} = 10^{12}$ wattora nel tempo. (Dati tratti da BP statistical Review of World Energy 2017).

Che impatto ha avuto l'energia dal sottosuolo nel recente sviluppo della società occidentale? La prima curva presente in Figura 2 mostra in maniera eloquente l'evoluzione della popolazione umana negli ultimi 12.000 anni. La curva indica chiaramente come a partire dalla rivoluzione agricola si sia passati da pochi milioni di abitanti a centinaia di milioni di abitanti durante l'impero romano (in circa 8000 anni) per crescere in maniera costante ma lenta sino alla metà del 1700 (arrivando a 500 milioni di abitanti) e successivamente subire una rapida impennata a metà 800, in piena rivoluzione industriale, arrivando in poco meno di due secoli a 7 miliardi di abitanti. Difatti, in termini di crescita demografica, globalmente, poco sembra essere successo per quasi 11.000 anni, per poi esplodere in piena rivoluzione industriale (e finanziaria) sino ad insistere ancora più rapidamente durante la rivoluzione tecnologica digitale moderna. Se confrontiamo tale figura con la Figura 3 che rappresenta il variare del consumo totale energetico in termini di TWh ($\text{TWh} = 10^{12}$ wattora) a partire da metà Ottocento, notiamo tre aspetti importanti:

- gran parte dell'energia utilizzata sino ai primi del 1800 era basata su quello che chiameremmo la tradizionale biomassa: essenzialmente forza bruta umana, animale e calore da legna
- a partire da metà 800, l'utilizzo del carbone come risorsa energetica primaria ha incrementato la produzione energetica sino a più che raddoppiarla ed ha coinciso

- con la prima rivoluzione industriale e la prima vera crescita di popolazione, oltre il miliardo in poco più di 50 anni. Il carbon fossile costituisce ancora oggi intorno al 30% della totale fonte primaria (la terza maggiore dopo petrolio e gas);
- a partire dalla metà del 900 gli idrocarburi diventano invece le maggiori risorse di energia primaria in assoluto. Le energie rinnovabili (solare, eolico, idroelettrico e nucleare) compaiono già a metà Novecento ma cominciano a diffondersi globalmente solo nella fine del XX secolo.

La Figura 4 mostra un ulteriore spaccato dei dati precedenti: a partire dal 1970 gran parte della energia primaria viene consumata per costruzioni, attività industriali e trasporti che crescono ovviamente (Fig. 1) proporzionalmente con l'aumento della popolazione (anni '70 4 miliardi, nel 2010 poco più di 7 miliardi...quasi raddoppiata in 40 anni!). Popolazione e globalizzazione crescono enormemente grazie al facile accesso a fonti energetiche primarie fossili.

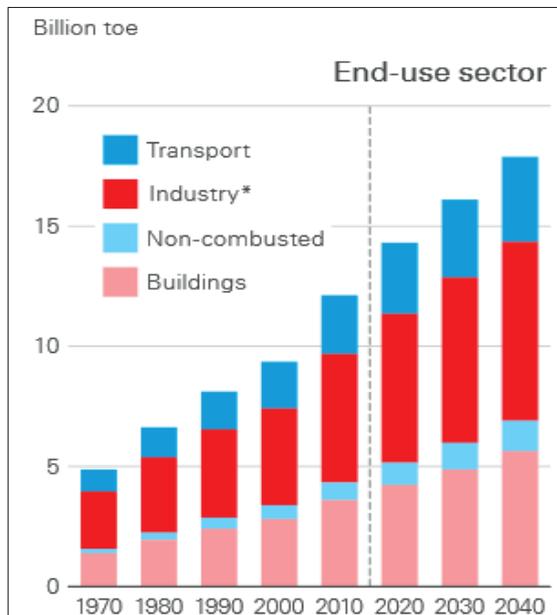


Figura 4. Utilizzo energetico per settori globali (source BP statistical review of World Energy, 2019)

Quadro storico sulle risorse energetiche

Se osserviamo ancora la Figura 3 osserviamo come la quasi totalità delle risorse energetiche che hanno contribuito a sviluppare sia la rivoluzione industriale che le successive rivoluzioni tecnologiche che ci hanno guidato nell'era digitale sono derivate dal sottosuolo e sostanzialmente si riducono a carbone ed idrocarburi.

Difatti il recente secolo veloce e digitale ha usufruito molto della grande abbondanza di tali risorse la cui efficienza nell'esploarle, estrarle e produrle ha contribuito alla veloce crescita sia in termini di popolazione che di consumo energetico, nel bene e nel male: aumento dei mezzi di trasporto, diffusione della corrente elettrica, tecnologia digitale a più di due miliardi di persone negli ultimi 30-40 anni. In Figura 5 si osserva come invece sino prima della rivoluzione industriale la totalità dell'energia dipendesse dal lavoro e dalla fatica umana ed animale e dal consumo di legna per il riscaldamento. Il carbone, nei decenni della prima e seconda rivoluzione industriale, ha piano piano moltiplicato la disponibilità energetica in Gran Bretagna (avendo gran disponibilità minerarie, cosa che il nostro paese non aveva) e sostituito l'uso delle biomasse (come la legna), fermo restando che la forza uomo non è affatto diminuita.

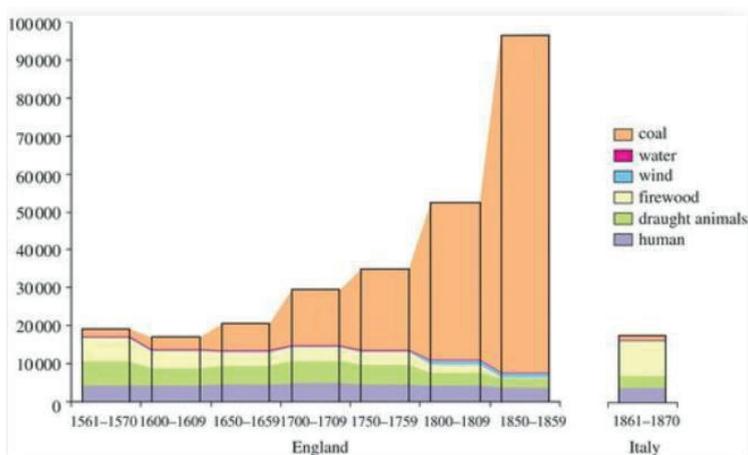


Figura 5. Distribuzione delle risorse energetiche a partire dal rinascimento sino alla rivoluzione industriale. (Dati tratti da *Energy and the English industrial Revolution*. Wrigley, 2013. *Phil Trans of the Royal Society A*)

In sintesi, si può sostenere che il recente sviluppo della società moderna sia pertanto derivato dall'accesso alle seguenti risorse fossili:

- carbon fossile
- idrocarburi

Diamoci un'occhiata veloce:

- Il carbone è il risultato della trasformazione di resti vegetali che sono stati compressi, alterati chimicamente e trasformati da calore e pressione nel corso dei tempi geologici, fra 80 e 400 Ma (megaannum è uguale a un milione (10^6) anni), perdendo ossigeno. Durante la prima rivoluzione industriale è stato il fondamento energetico che ha innescato il passaggio dalla manifattura ad

un'economia basata su macchine. La legna, che ha rappresentato per tutto il medioevo e rinascimento l'unico combustibile utilizzato per fondere minerali di ferro e riscaldare, verrà totalmente sostituita a livello industriale dal carbone. Il carbone durante la rivoluzione industriale darà origine a mezzi di trasporto emergenti, quali le locomotive e le navi a vapore, e ad al riscaldamento degli edifici. Il carbone convertito in gas (coke) verrà poi distribuito per mezzo di tubature per essere bruciato a scopo di illuminazione, riscaldamento e cucina.

- Gli idrocarburi sono invece composti organici che contengono catene più o meno complesse di atomi di carbonio e di idrogeno. Gli idrocarburi danno origine a petrolio inteso come miscela liquida di vari idrocarburi, in prevalenza alcani, acqua e varie impurità e metano, i quali rappresenteranno la fonte principale per tutto il trasporto, riscaldamento (ad affiancare e sostituire il carbon fossile) e parte dei materiali plastici ad uso industriale. La ragione del loro veloce sopravvento e sostituzione in parte del carbon fossile sta nel loro potere calorifico: il petrolio greggio ha un potere calorifico di 10.000 kcal/kg; il Gas propano Liquido di circa 11.000 kcal/kg che entrambi sono molto maggiore rispetto all' Antracite la quale rappresenta la componente più pura e vecchia e dunque energetica del carbone (8.000 kcal/kg). Inoltre, rispetto al carbone il trasporto degli idrocarburi risulta molto più semplice e meno costoso (attraverso navi, gasdotti e oleodotti).

Come mostrato nella Figura 6, al momento la disponibilità comprovata e mappata di queste due risorse energetiche risulta ancora enorme e in grado di sostenere la produzione energetica mondiale ad alti ritmi per circa altri 50 anni. Tuttavia, a dispetto di tale relativa abbondanza, tali risorse energetiche hanno purtroppo innescato un problema di ordine e importanza e impatto maggiore nel sistema terrestre ed atmosferico: il cambiamento climatico. Questo richiede ora e nel futuro molto prossimo una radicale revisione dei futuri piani di approvvigionamento energetico. Non possiamo più continuare ad utilizzare in maniera perenne le risorse fossili disponibili nel sottosuolo nonostante la loro abbondanza (Fig 7)

Risorse energetiche fossili e il problema delle emissioni antropiche di CO₂

Come mostrato nella Figura 8 a partire da circa metà dell'800 la quantità di emissioni di anidride carbonica ha cominciato a crescere esponenzialmente. I vari spike mostrano le misurazioni mensili di emissioni di CO₂ (di origini diverse), ma la curva interpolata di maggiore interesse, in arancione, invece mostra il riscaldamento indotto da attività umane. A partire della metà del 900 gli effetti antropogenici hanno radicalmente modificato l'andamento sia delle emissioni di CO₂ che del riscaldamento atmosferico globale, aprendo un nuovo fronte di preoccupazioni. Il

diagramma in Figura 9 mostra come la totalità di queste emissioni, in termini di Giga tonnellate, è risultata prodotta proporzionalmente dal carbone (attualmente in diminuzione), petrolio e gas metano che ora sono le sorgenti principali delle emissioni totali. Perciò i motori principali del recente e rapido aumento di CO₂ sono indubbiamente da associare ai combustibili fossili.

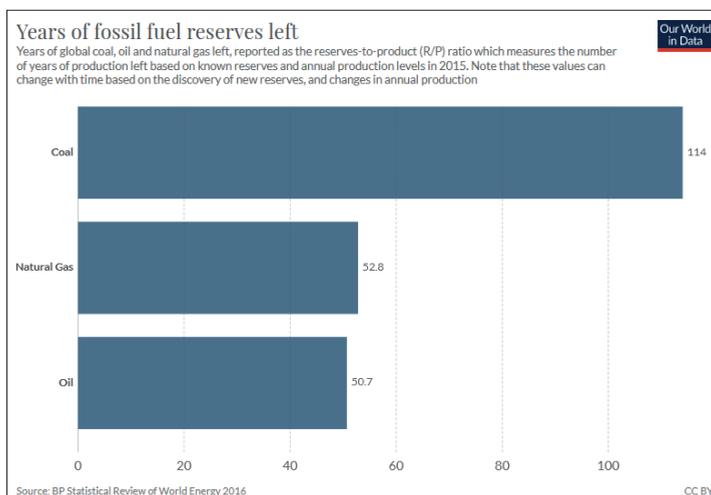


Figura 7. Disponibilità di riserve energetiche fossili attualmente disponibili. Dati ottenuti da BP statistical review of world Energy 2016

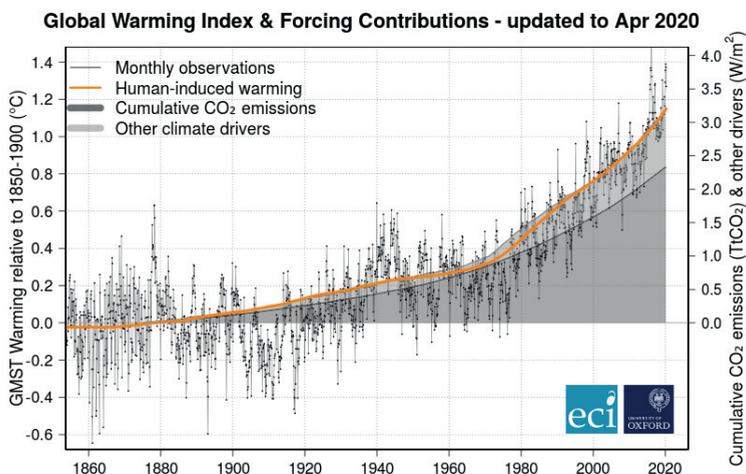


Figura 8. Andamento delle emissioni di CO₂ cumulative e andamento del global warming rispetto al 1850-1900 (riferito come 0.0). Dati Global Warming Index. Oxford.

Come mostrato dalle cifre proposte in Figura 10, in termini di *societal challenge* il problema che rimane da affrontare è perciò piuttosto ovvio: la nostra società è al

momento ancora destinata a crescere in termini sia di popolazione (le stime più conservative parlano di 10 miliardi di abitanti nel 2070) sia di scambi commerciali (a meno di ulteriori cataclismi sanitari naturali da non escludere...per esempio la diffusione del covid ha ridotto dell'8% la produzione energetica mondiale, dati *DNV Outlook 2020*) per effetto sia della globalizzazione commerciale che dell'azione circolo di informazioni.

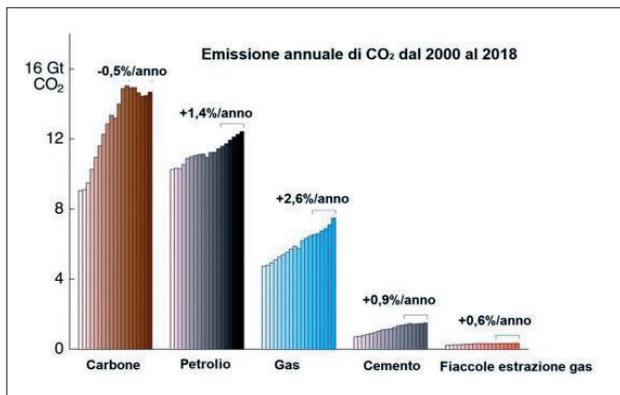


Figura 9. Emissione annuale di CO₂ dal 2000 al 2018 per diverse fonti energetiche. (*Global Carbon Project / L.Bignami*)

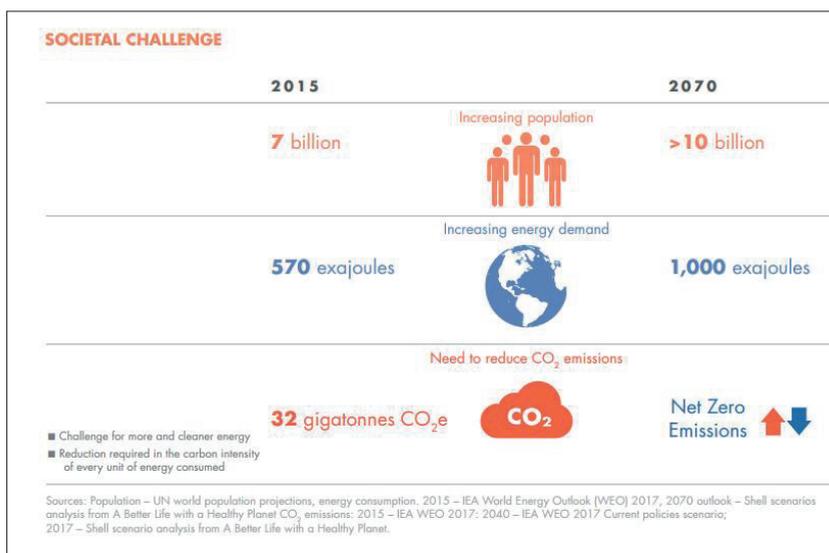


Figura 10. Stime e proiezioni di crescita di energia,popolazione ed emissioni di CO₂ (dati *IEA World Energy Outlook 2017*)

Perciò le diverse società tenderanno a convergere verso una diffusione tecnologica alla ricerca di stabilità energetica, che porterà inesorabilmente ad un aumento della richiesta di energia. Se le previsioni proposte sono corrette. se si vuole abbattere le

emissioni di origine antropogenica di CO₂ ed arrivare a zero emissioni nel 2070, se si vuole permettere a buona parte della società di continuare a crescere e svilupparsi economicamente anche tenendo conto di un raffinamento tecnologico nel versante dell'efficienza energetica (altro capitolo fondamentale, ma non abbia abbiamo qui spazio e tempo per affrontarlo) abbiamo un grosso problema: i due target a medio lungo termine (crescere tecnologicamente abbattendo le emissioni) non sono facilmente attuabili e necessariamente compatibili. È da tenere conto che nel frattempo sino al 2070 continueremo a emettere CO₂ e dunque per i prossimi 50 anni l'atmosfera continuerà a saturarsi in anidride carbonica. In questo contesto le società devono ri-orientarsi in termini di sviluppo di nuove risorse energetiche, chiamate alternative alle energie fossili, e questo però richiederà notevoli sforzi a livello di ricerca scientifica e tecnologica e non potrà avvenire in maniera brusca e tantomeno a costo zero. Per questo si parla oramai di *transizione energetica* come di una fase in cui le tecnologie alla base della distribuzione energetica adattate alle risorse fossili devono essere sostituite da tecnologie che sviluppino energie alternative. E questo ci porterà ad affrontare un altro aspetto e dunque ritornare alle analisi delle potenzialità future del sottosuolo e del suo nuovo ruolo e in questa fase di transizione.

Futuro del sottosuolo nella transizione energetica

Ci è dunque chiaro che il futuro del sottosuolo deve essere tecnologicamente ri-orientato in una direzione che è nota come la transizione energetica: a supporto di uno sviluppo di energie sostenibili, verso minori emissioni di CO₂ ma al tempo stesso permettere finanziariamente la transizione che richiederà energia a supporto degli investimenti ancora in corso. Al momento sono quattro i principali scenari e investimenti tecnologici in corso e che vedono un promettente sviluppo:

- Gas Naturale liquido a sostituire il carbon fossile. Ha il vantaggio di essere facile da trasportare, presente nel sottosuolo in quantità tali da sostenere una transizione di 20-30 anni. Purtroppo, emette CO₂ che sebbene in quantità inferiori al carbone e' sempre troppo invasivo.
- Il sottosuolo è ancora il luogo della maggiore risorsa necessaria per sviluppare le energie alternative: metalli quali litio e vanadio che si renderanno necessarie per creare batterie fondamentali per stoccare energia rinnovabile e ridistribuirla nel sistema (oltre che come vettore energetico per il trasporto).
- La Geotermia rappresenta un'opzione tecnologicamente matura in aree in cui il sottosuolo è dotato di gradienti geotermici importanti. Costituisce oggi meno dell'1% della produzione mondiale di energia ma ha potenzialità enormi se

sostenuta finanziariamente (ENEL è fra l'altro la maggiore azienda produttrice al mondo).

- Il sottosuolo come area di stoccaggio per CO₂ (Fig 11) e di idrogeno (Fig 12). Come già descritto in Figura 10, oltre agli investimenti necessari per sviluppare energie alternative sarà necessario ridurre la CO₂ giacente e continuamente immessa in atmosfera. Una strategia è data dal suo stoccaggio anche nel sottosuolo,
- L'idrogeno rappresenta invece uno dei più interessanti e potenzialmente efficienti vettori energetici: ha un potere calorifico enorme ma se utilizzato massicciamente (come energia per i trasporti pesanti), richiederà serbatoi o aree di stoccaggio in grado di contenere il gas a pressioni sino a 250 bar. L'idrogeno è però un materiale molto corrosivo.

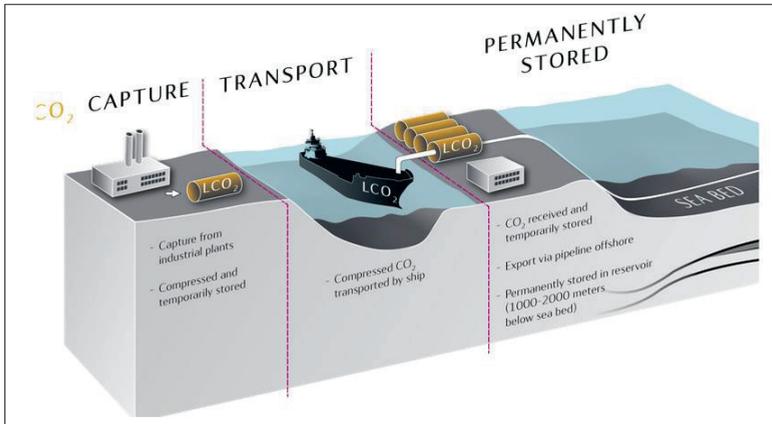


Figura 11. Classico schema di cattura trasporto e stoccaggio di CO₂ (Equinor)

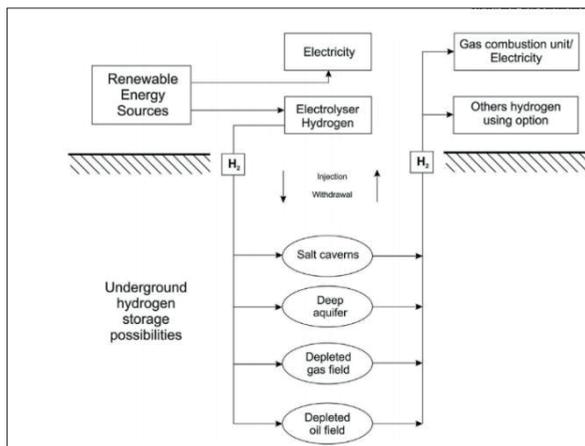


Figura 12. Classico schema di utilizzo di idrogeno.

Conclusioni

A partire dalla rivoluzione industriale il sottosuolo e le sue risorse energetiche fossili hanno avuto certamente un ruolo straordinario e fondamentale nel sostenere la crescita economica e lo sviluppo tecnologico della società moderna e tuttora rappresentano la principale risorsa energetica (80-90 % del totale). Gran parte di tali risorse fossili hanno tuttavia prodotto emissioni importanti di CO₂ che cumulate negli ultimi 50 anni nella nostra atmosfera ne hanno modificato la composizione e innescato importanti cambiamenti climatici di difficili previsioni e controllo. In tale scenario, oltre a rendersi necessario lo svilupparsi di tecnologie in grado di abbattere le emissioni di CO₂, le geotecnologie preposte all'esplorazione del sottosuolo devono trovare una loro strada nel contesto delle energie rinnovabili (solare, eolico) o alternative (elettricità, idrogeno, biomasse). Al momento le tecnologie in atto ed in fase di sperimentazione a livello globale riguardano sia il sostituire gran parte del carbon fossile attraverso una transizione a gas e geotermia, sia il preparare e monitorare aree di stoccaggio nel sottosuolo in cui anidride carbonica e potenziali vettori energetici quali l'idrogeno dovranno essere iniettati. Lo sforzo richiederà enormi investimenti tecnologici e finanziari e aprirà notevoli opportunità di innovazione e lavoro nelle geoscienze verso un futuro *low carbon*.

Matematica per la Crittografia

Maria Rosaria Celentani

*Dipartimento di Matematica e Applicazioni "R. Caccioppoli",
Università di Napoli Federico II*

Abstract

L'esigenza di mandare messaggi privati, evitando che estranei possano impadronirsene, è un'esigenza antica quanto la scrittura: le più antiche civiltà hanno lasciato testimonianza di testi volutamente cifrati. Diversi matematici in diverse epoche si sono cimentati sul versante della crittoanalisi. Al giorno d'oggi la maggior parte delle nostre azioni quotidiane sono protette dalla crittografia. Questo articolo, nato dalla conferenza tenuta nell'ambito della PLS Virtual Summer School 2020, ha come scopo introdurre in maniera informale chi legge nel complesso ed articolato mondo della crittografia e della matematica.

Introduzione

Quante volte, sin da piccoli, abbiamo utilizzato un "modo segreto" per mandare un messaggio ad un amico? Poter comunicare in maniera riservata con una persona senza correre il rischio che altri possano capire il nostro messaggio è da sempre una esigenza dell'uomo. Il termine "sin da piccoli" si adatta anche alle epoche storiche: abbiamo testimonianze di messaggi cifrati già dagli antichi Egizi. Plutarco nei suoi scritti parla della scitola lacedemonica. In che cosa consisteva? Per mandare un messaggio segreto si sceglieva un bastone (scitola) con un certo spessore: sia chi inviava il messaggio sia chi lo riceveva doveva avere un bastone avente lo stesso diametro. Intorno al bastone si avvolgeva una strisciolina e si scriveva il messaggio da inviare. Una volta scritto il messaggio quella che partiva era la sola striscia, contenente una serie di lettere apparentemente prive di significato: poteva decifrare il messaggio solo chi possedeva un bastone avente esattamente lo stesso diametro. Insomma in questo caso la chiave per criptare e poi decriptare è il bastone, la scitola.



Figura 1

Altro classico esempio è il famoso Codice di Cesare: si applicava uno slittamento (fissato) di ogni lettera: se lo slittamento fissato ad esempio era di 3 posizioni (e questa era dunque la chiave per criptare) nel messaggio cifrato alla lettera A era sostituita la lettera D, alla lettera B la lettera E, e così di seguito. Per decriptare invece si “ritornava indietro” esattamente di 3 posizioni. Più avanti ritorneremo in maniera più approfondita sul cifrario di Cesare, vedendo come questo codice sia un particolare cifrario affine. Da Plutarco ad oggi la crittografia ha svolto un ruolo fondamentale nella storia e nella società. Non c’è stata guerra in cui non ci si sia posti il problema di trasmettere messaggi in segreto. Andando avanti nel tempo l’utilizzo della crittografia ha iniziato ad interessare molti più campi. Per secoli la crittografia è stata legata a situazioni “lontane” dalla vita quotidiana: guerra, spionaggio industriale o diplomatico. Al giorno d’oggi invece ognuno di noi ne fa uso quotidianamente: ogni volta che utilizziamo un bancomat, spediamo un messaggio, usiamo la firma digitale, facciamo un acquisto online è la crittografia a proteggerci, a permettere che ciò sia possibile senza correre (si spera!) rischi.

Allora cerchiamo di fare un breve viaggio in questo affascinante mondo e cerchiamo di capire quanta e quale matematica serve per criptare in maniera sicura un messaggio.

Prima di iniziare dobbiamo capire come si passa da un messaggio ai numeri. La tecnica, che un po’ tutti abbiamo utilizzato sin da bambini e a cui si è già accennato, è di fatto proprio quella che si utilizza:

Tabella 1

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25

In realtà quella che si usa è un po’ più raffinata, poiché nei nostri messaggi noi utilizziamo oltre alle lettere, gli spazi e la punteggiatura: pertanto ognuno di questi caratteri viene sostituito dal suo codice ASCII (a due cifre) e ad una stringa di caratteri corrisponderà il numero ottenuto per giustapposizione di questi codici:

Tabella 2: codice ASCII

...		,	-	A	B	C	...	Y	Z
...	32	44	45	46	...	65	66	67	...	89	90

Ad esempio, volendo riscrivere **CAB, Y** utilizzando la Tabella 2 si trova: **676566443289**. Da qui in poi un messaggio verrà trasformato in una stringa di numeri: pertanto è con i numeri che avremo a che fare e sui numeri ragioneremo. In realtà per semplicità ci rifaremo alla sostituzione proposta nella Tabella 1. Vediamo quali sono i principali schemi di trasmissione criptata di messaggi:

- **a chiave privata** (o simmetrica, crittografia classica);
- **a chiave pubblica** (o asimmetrica). In realtà, nella pratica vengono quasi sempre usati metodi ibridi combinando questi primi due;
- **senza scambio di chiave**.

Cerchiamo di capire quanti problemi incontriamo quando vogliamo trasmettere un messaggio. Ci troviamo di fronte a due problemi molto importanti: se “chiudiamo” (criptiamo) un messaggio con una chiave, occorre che chi riceve il messaggio posseda la stessa chiave per “aprire” (decriptare) il messaggio. Pertanto, sono subito in evidenza i due problemi:

- 1. lo scambio della chiave:** durante lo scambio qualche malintenzionato può impossessarsi della chiave e dunque del messaggio;
- 2. la serratura con cui chiudiamo deve essere sicura:** deve essere difficile “forzarla”.

Poiché, lo abbiamo detto, ogni messaggio viene in maniera naturale trasformata in una stringa di numeri, per “chiuderlo” si utilizzerà una funzione (biunivoca e dunque invertibile) che trasforma la stringa numerica in maniera opportuna, per “aprirlo” occorrerà utilizzare la funzione inversa, così da ottenere il messaggio originario. Allora la “chiave” da scambiare eventualmente non è tanto la funzione con cui si cripta, ma la “chiave” per aprire, ossia la funzione inversa. Inoltre, la sicurezza della serratura è assicurata se è “difficile” (matematicamente parlando) calcolare la funzione inversa (a meno di conoscerla). Iniziamo ad esaminare l’ultimo degli schemi proposti.

Scambio di messaggi senza scambio di chiave

Questo metodo per criptare i messaggi si può paragonare al chiudere il messaggio segreto in un baule non dotato di serratura ma che abbia la possibilità di inserire dei **lucchetti esterni**: supponiamo che Andrea (A) voglia inviare un messaggio segreto a Bruna (B) e che il messaggio sia spedito in un cofanetto a cui sia possibile applicare dei lucchetti. Sia Andrea che Bruna posseggono un proprio lucchetto e la relativa chiave. Andrea scrive il messaggio, lo mette nel cofanetto e chiude il cofanetto con

il proprio lucchetto. Spedisce il cofanetto a Bruna. Quando Bruna lo riceve invece di cercare di forzare il lucchetto applica al cofanetto il suo lucchetto e rispedisce il tutto ad Andrea. Andrea a questo punto toglie il primo lucchetto e rispedisce a Bruna il cofanetto; Bruna ha il cofanetto chiuso solo dal lucchetto di cui lei possiede la chiave, può aprirlo e finalmente leggere il messaggio che Andrea le ha inviato. Durante tutti i percorsi il cofanetto ha viaggiato protetto, chiuso da almeno un lucchetto. Andrea e Bruna sono riusciti a comunicare tra loro senza scambiare la chiave ma utilizzando ciascuno una sua chiave personale. Inutile dire che i “lucchetti” e le “chiavi” sono matematici. Ecco lo schema grafico di quello che abbiamo descritto:

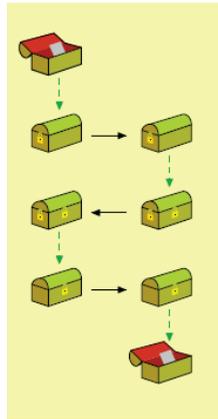


Figura 2

C'è un altro modo di procedere senza utilizzare uno scambio di chiave: è possibile concordare una chiave a distanza.

Protocollo di Diffie-Hellman Ovvero: come accordarsi su una chiave senza trasmetterla

Questo protocollo illustra, in maniera semplificata, come sia possibile accordarsi su una chiave senza trasmetterla.

- Andrea vuole utilizzare per comunicare con Bruna una chiave numerica segreta h ma non vuole correre il rischio che qualcuno intercetti la loro chiave. Pertanto sceglie di trasmettere una chiave provvisoria per cifrare un messaggio: se h la chiave numerica, fissa un intero segreto α , calcola $h^\alpha = h_1$ e trasmette a Bruna h_1 ;

[Sia ad esempio $h = 7$, $\alpha = 3$, $7^3 = 343 = h_1$]

- Bruna riceve la chiave provvisoria e la eleva ulteriormente ad un intero β che solo lei conosce: $h_1^\beta = h_2$. Trasmette ad Andrea h_2 ;

$$[\beta = 2, 343^2 = 117649]$$

- Ora Andrea può calcolare $h_2^{\alpha^{-1}} = h^{\alpha\beta\alpha^{-1}} = h^\beta$ e trasmetterlo a Bruna;

$$[\sqrt[3]{117649} = 117649^{-3} = 49]$$

- Bruna riceve h^β e dunque può calcolare $h = \sqrt[\beta]{h^\beta}$.

$$\sqrt[2]{49} = 7$$

ATTENZIONE: va sottolineato che questo protocollo funziona poiché $h^{\alpha\beta} = h^{\beta\alpha}$

Occupiamoci ora di esaminare gli altri schemi di trasmissione, certamente i più utilizzati nella pratica.

Crittografia simmetrica

La crittografia simmetrica è quella che storicamente è stata maggiormente utilizzata: lo schema è quello a cui si sono ispirati per secoli tutti coloro che avevano esigenza di mandare un messaggio segreto. La situazione è sostanzialmente diversa da quella appena illustrata: in questo caso quando Andrea vuole codificare il messaggio (chiuderlo nel cofanetto) per trasmetterlo ha a disposizione un'unica serratura e sia Andrea che Bruna devono possedere la stessa chiave affinché Bruna possa decodificare il messaggio (aprire la serratura). Pertanto questo schema utilizza una chiave privata che deve essere in qualche modo condivisa.



Figura 3

Poiché la chiave k per criptare il messaggio m altro non è che una funzione biettiva f_k che trasforma m in m' , per decriptare (aprire) il messaggio m' occorrerà la funzione inversa f_k^{-1} . Ecco lo schema matematico che rappresenta quanto detto.

è il criptosistema **RSA** (dai nomi, Rivest, Shamir e Adleman, di coloro che lo proposero nel 1977). Sempre paragonando la crittazione di un messaggio alla chiusura in cofanetto, la crittografia asimmetrica utilizza il seguente schema: il «cofanetto» contenente il messaggio è dotato di una serratura a scatto (la chiave pubblica è la serratura a scatto, la funzione C_B): tutti possono chiuderla (ossia la funzione C_B è pubblica) ma solo chi possiede la chiave privata (la chiave per aprire, la funzione $C_B^{-1} = D_B$) può riaprirla! Sembra incredibile che la chiave per criptare il messaggio sia pubblica ma in realtà l'idea importante è che solo chi già determina la chiave di apertura è in grado di tornare indietro: la funzione C_B è tale che il calcolo di $C_B(m) = m'$ sia facile da eseguire mentre sia estremamente complesso determinare la funzione inversa D_B , a meno di non avere delle informazioni aggiuntive (più avanti faremo un esempio in tal senso).

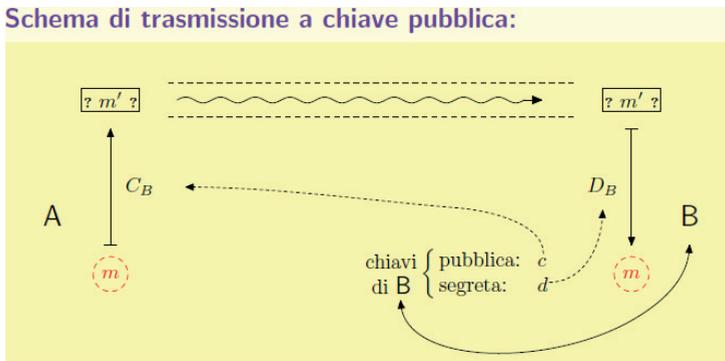


Figura 5

Qui C_B è la chiave pubblica: tutti la posseggono e sono in grado di cifrare il messaggio calcolando $C_B(m) = m'$; la chiave privata, che possiede SOLO B, è l'inversa di C_B , ed è D_B : $D_B(m') = m$

Un po' di matematica (poca!): l'aritmetica modulare

Per poter proseguire in questa breve panoramica occorre ora accennare brevemente a quali sono gli strumenti matematici che occorrono. In realtà ci sono diversi ambiti della matematica che vengono utilizzati in crittografia: l'aritmetica modulare, le curve ellittiche - costruite in piani che, a differenza di quelli della tradizionale geometria euclidea, posseggono solo un numero finito di punti- ne sono un esempio. Ci limitiamo ad un brevissimo cenno all'aritmetica modulare, anche nota come l'aritmetica dell'orologio.

Assegnato un intero $n > 1$, diremo che due interi a e b sono congrui modulo n se e solo se n divide la differenza tra a e b , vale a dire se:

$$a - b = n \cdot h$$

In tal caso scriveremo che $a \equiv b \pmod{n}$ o anche $a \equiv_n b$. E' molto facile verificare che $a \equiv_n b$ se e solo se il resto della divisione tra a e n è uguale al resto della divisione tra b e n . Se divido un numero intero per n i possibili resti saranno: $0, 1, 2, \dots, n - 1$. Si indica con

$$i = [i]_n = \{a \mid a \equiv_n i\} = \{a \mid \text{rest}(a, n) = \text{rest}(i, n)\} = \{i + n \cdot k \mid k \in \mathbb{Z}\}$$

i è detta classe modulo n . Ad esempio, se $n = 7$ e $i = 5$

$$5 = [5]_7 = \{k7 + 5 \mid k \in \mathbb{Z}\} = [12]_7 = [19]_7 = \mathbf{19} \dots$$

è la classe di 5 modulo 7. Per capire che già conosciamo di cosa si tratta (e giustificare il nome di "aritmetica dell'orologio"):

- Se considero la suddivisione oraria in 24 ore e dico: «sono le 9; tra 35 ore che ora sarà?», la risposta sarà «le ore 20». Questo vuol dire ragionare modulo 24: $[9]_{24} = [9 + 35 = 44]_{24}$

Indichiamo l'insieme delle classi modulo n nel modo che segue:

$$\mathbb{Z}_n = \{\mathbf{0}, \mathbf{1}, \dots, \mathbf{n - 1}\}$$

In tale insieme è possibile definire somma e prodotto: $\mathbf{i + j = i + j}$, $\mathbf{i \cdot j = i \cdot j}$. Facciamo un esempio; sia $n = 5$. Allora se consideriamo $\mathbb{Z}_5 = \{\mathbf{0}, \mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4}\}$ si ha, ad esempio, $\mathbf{4 + 4 = 8 = 3}$, $\mathbf{4 \cdot 3 = 12 = 2}$. Pertanto ha senso considerare l'anello $(\mathbb{Z}_n, +, \cdot)$ in cui è definita l'aritmetica modulare. Alcuni calcoli in aritmetica modulare sono particolarmente semplificati, in particolare risulta particolarmente facile il calcolo delle potenze. Ad esempio

$$4^2 \equiv_{13} 3, 4^3 \equiv_{13} 4 \cdot 4^2 \equiv_{13} 4 \cdot 3 \equiv_{13} 12 \equiv_{13} -1, 4^{10000} \equiv_{13} 4 \cdot 4^{3(3333)} \equiv_{13} 4 \cdot (-1) \equiv_{13} -4$$

Viceversa in aritmetica modulare è estremamente complesso calcolare logaritmi: spesso l'unico modo per procedere a calcolare un logaritmo in aritmetica modulare è per tentativi. Per questo motivo le funzioni esponenziali, di cui è facile il calcolo diretto ma è estremamente complesso il calcolo inverso, vengono spesso utilizzate nei protocolli crittografici. In aritmetica modulare un ruolo fondamentale è ricoperto dai seguenti risultati

Proposizione 1: *Considerati due interi non nulli a e n , esiste a' tale che $aa' \equiv_n 1$ se e solo se $M.C.D.(a, n) = 1$.*

La funzione di Gauss-Eulero: Sia $\varphi(n)$ il numero degli interi positivi minori o uguali ad n e coprimi con n . Ad esempio $\varphi(5) = 4$, $\varphi(6) = 2$. E' immediato verificare che se p è un numero primo, allora $\varphi(p) = p - 1$. Ruolo centrale in tutte le considerazioni crittografiche è svolto dal

Il Teorema di Fermat-Eulero: *Siano a e n interi positivi e coprimi. Allora:*

$$a^{\varphi(n)} \equiv_n 1$$

Purtroppo lo studio di tali argomenti, matematicamente non difficili, non può trovare spazio in questa breve nota. Per tutti l'invito ad approfondirli: scoprirete una matematica "facile" e dalle innumerevoli applicazioni.

Sfruttando la (Tabella 1) è possibile definire una corrispondenza biettiva tra $A = \{A, B, C, \dots, X, Y, W, Z\}$ e $\mathbb{Z}_{26} = \{0, 1, \dots, 25\}$: identifichiamo ogni lettera non semplicemente con il numero corrispondente ma con tutta la sua classe di resto. Ad esempio la lettera D corrisponde a **3** (ma anche a **29**, **55** ...). Una trasformazione affine su un alfabeto A di 26 lettere (identificate con le corrispondenti classi), è una trasformazione del tipo:

$$f_{\alpha,\beta}: i \in A \rightarrow \alpha i + \beta \in A \text{ (con } \alpha \text{ e } \beta \in \mathbb{Z})$$

[ad esempio, se $\alpha = 11$,

$$\beta = 5 \text{ e } i = 3 \text{ si ha: } 11 \cdot 3 + 5 = 38 = 12, \text{ dunque } f_{11,5}(D) = M]$$

Una trasformazione affine di questo tipo, se biettiva, può essere utilizzata come chiave per criptare (e siamo certi che esiste la chiave per decifrare, la funzione inversa $f_{\alpha,\beta}^{-1}$!): in tal caso parleremo di cifrario affine. Ma quando la funzione $f_{\alpha,\beta}$ è biettiva? Solo quando α è invertibile (modulo n) e dunque $M.C.D.(\alpha, n) = 1$ (cfr. proposizione 1) e la sua inversa (ossia la chiave per «aprire»), si vede facilmente, è data da:

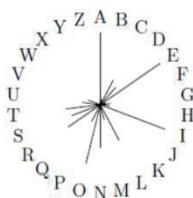
$$f_{\alpha,\beta}^{-1}: j \in A \rightarrow \alpha^{-1}j - \beta \in A$$

Quando $\alpha = 1$ tale funzione affine altro non fa che definire il famoso Codice di Cesare. Facciamo un esempio facile: $\alpha = 1, \beta = 5$

$$f_{1,5}: i \in A \rightarrow i + 5 \in A \text{ (chiave per cifrare) e } f_{1,5}^{-1}: j \in A \rightarrow j - 5 \in A \text{ (chiave per decifrare)}$$

$m = \langle \text{buongiorno a tutti} \rangle = \langle 1-20-14-13-6-8-14-17-13-14 \quad 0 \quad 19-20-19-19-8 \rangle$,
 $f_{1,5}(m) = m' = \langle \text{gztslntwst e zzyyn} \rangle = \langle 6-25-19-18-11-13-19-22-18-19 \quad 5 \quad 24-25-24-13 \rangle$, $f_{1,5}^{-1}(m') = m$

Ovviamente un simile modo di procedere protegge poco. Si può (troppo!) facilmente trovare la funzione inversa (forzare la serratura). Ad esempio, pur non conoscendo α e β si può andare ad analizzare la frequenza delle lettere nel messaggio cifrato e risalire ad essi. Ecco schematizzate le frequenze delle 26 lettere dell'alfabeto latino nella lingua italiana:



Supponiamo di essere noi gli spioni e di avere intercettato questo messaggio:

fqqtwf, qf htshqzxtsj j hmj nq hnkfwnt
in Hjxfwj sts j fifyyt f hzxytinwj n stxywn
rjxxfln xjlwjyn!

Andando ad esaminare le ricorrenze delle lettere nel testo trovo che:

f=9, h=6, i=3, j=10, k=1, l=3, m=1, n=10, q=5, r=1, s=5, t=8, w=7, x=7, y=5, z=2

Per cui si ha:



Cercando di sovrapporre i dischi trovo che $\beta = 5$ e $f^{-1}: a \in A \rightarrow a - 5 \in A$

fqqtwf, qf htshqzxtsj j hmj nq hnkfwnt in Hjxfwj sts j fifyyt f hzxytinwj n stxywn

Allora, la conclusione è che il cifrario di Cesare non è adatto a custodire i nostri

rjxxfln xjlwjyn!

messaggi segreti!

Si può generalizzare (e rendere un po' più difficoltoso) questo processo utilizzando non un semplice «slittamento costante» ma una «chiave» più lunga: stiamo parlando del metodo di Vigenère. Se come chiave utilizziamo la frase “vorrei andare in vacanza”

Tabella 3

V	O	R	R	E	I	A	N	D	A	R	E	I	N	V	A	C	A	N	Z	A
2	1	1	1	4	8	0	1	3	0	1	4	8	1	2	0	2	0	1	2	0
2	3	7	7				3			7			3	2				3	5	

Allora la prima lettera slitterà di 22 posti, la seconda di 13 e così via

m= «buongiorno a tutti» = «1-20-14-13-6-8-14-17-13-14 0 19-20-19-19-8»

C(m) → m' = «xifekqoetqo r ...»

Quando si utilizza questo metodo per criptare è comodo l'utilizzo della tavola di Vigenère:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
B	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A
C	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B
D	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C
E	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D
F	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E
G	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F
H	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G
I	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H
J	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I
K	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J
L	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K
M	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L
N	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M
O	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N
P	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Q	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
R	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
S	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
T	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
U	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
V	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
W	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
X	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
Y	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
Z	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y

Figura 6

Crittografia asimmetrica e RSA

Nella crittografia asimmetrica ogni utente del sistema crittografico ha una personale coppia di chiavi, di cui una comune a tutti (pubblica) ed una segreta (privata). Un mittente A che voglia inviare un messaggio riservato a B provvederà innanzitutto a cifrarlo, utilizzando a questo scopo la chiave pubblica, nota anche a B, il quale B potrà poi usare la sua chiave segreta per decifrarlo. Affinché il sistema possa funzionare, è necessario che chiunque sia in grado di effettuare la cifratura, ma che la decifrazione sia praticamente impossibile («difficile») da eseguirsi a meno di non conoscere la chiave segreta di B. Per realizzare un sistema del genere bisogna trovare una funzione (biunivoca e dunque invertibile) $f = C_B$ da usare per la cifratura tale che, dato un qualunque messaggio m , sia semplice il calcolo di $m' = f(m)$ (il messaggio cifrato) ma, viceversa, qualunque algoritmo che permetta di ricavare m da m' senza disporre di una informazione extra (la chiave segreta, $f^{-1} = D_B$) richieda troppe risorse e troppo tempo di esecuzione per essere effettivamente utilizzabile. Questo è lo schema su cui è fondato il crittosistema RSA che verrà ora illustrato brevemente. Esso è basato sul fatto che non sono noti efficienti metodi che permettano di calcolare i fattori primi di un numero “molto grande”: mentre, assegnati due numeri primi p e q , è “facile” il calcolo del prodotto $n = pq$, il calcolo inverso — ricavare i fattori p e q da n — può essere “molto difficile”, praticamente impossibile anche per i più veloci computer.

Ad esempio se $p = 137562311$, $q = 289877297$ allora è molto facile calcolare $n = pq = 39876190881753367$, mentre, viceversa, è molto difficoltoso conoscendo n risalire a p e q .

Per definire le sue chiavi, l'utente B

1. sceglie due primi distinti (molto grandi) p e q e ne calcola il prodotto $n = pq$;
2. sceglie poi un intero positivo c che sia minore di n e primo con $\varphi = \varphi(n) = (p - 1)(q - 1)$;
3. dall'aritmetica modulare (cfr. Proposizione 1) segue l'esistenza di un intero d tale che $cd \equiv_{\varphi} 1$;
4. la coppia (n, c) costituisce la chiave pubblica di B e permetterà di definire la chiave privata d (che solo B possiede poiché solo lui conosce φ).

La cifratura di un messaggio da Andrea per Bruna avviene come segue:

1. occorre che il numero m che rappresenta il messaggio sia minore di n : se ciò non è possibile perché il messaggio è troppo lungo, quest'ultimo andrà preliminarmente suddiviso in blocchi minori di n da trasmettere uno per volta.
2. Il mittente A calcola il resto m' di m^c nella divisione per n : questo sarà il messaggio cifrato.
3. La decifrazione avviene in modo simile: B calcola il resto di $(m')^d$ modulo n : questo resto sarà proprio m .

Il motivo è conseguenza della seguente variante del teorema di Fermat e Eulero:

Teorema 3: *se n non è divisibile per il quadrato di nessun numero intero maggiore di 1, allora*

$$a^{1+k\varphi(n)} \equiv_n a$$

per ogni intero a ed ogni intero non negativo k .

e nel fatto che risulta $\varphi = (p - 1)(q - 1) = \varphi(n)$.

Allora

$$cd \equiv_{\varphi} 1 \quad e \quad (m')^d \equiv_n m^{cd}$$

e dunque $cd = 1 + k\varphi$ e $(m')^d \equiv_n m^{cd} = m^{1+k\varphi} \equiv_n m$. Abbiamo riottenuto il messaggio originale m ! Facciamo un esempio (facile!) per capire i passaggi del crittosistema RSA. Se $p = 3$, $q = 11$, allora $n = 33$, $\varphi = 2 \cdot 10 = 20$; fissiamo $c = 3$. Supponiamo di voler trasmettere un numero (messaggio) segreto: $m = 12$: la chiave pubblica è data da $(n, c) = (33, 3)$, Bruna determina la sua chiave privata, l'intero d tale che $c \cdot d \equiv 1 \pmod{20}$: in questo facile caso $d = 7$ è la chiave privata di Bruna. Andrea cripta il messaggio: $12^3 = 1728 \equiv 8 \pmod{20}$, dunque $m' = 8$ è il messaggio criptato che Andrea invia a Bruna; Bruna, per leggere il messaggio originario, eleva m' a d :

$$(m')^d = 8^7 = 2097152 \equiv 12 \pmod{20}$$

[se eleviamo 8 a potenza, modulo 20, da $8^7 = (8^2)^3 \cdot 8$, essendo:
 $8^2 \equiv 4 \pmod{20}$, $(8^2)^3 \equiv 4^3 \equiv 4 \pmod{20}$ e dunque
 $8^7 = (8^2)^3 \cdot 8 \equiv 4 \cdot 8 = 32 \equiv 12 \pmod{20}$.
 Allora è davvero facile calcolare le potenze modulo n !]

I numeri utilizzati, ovviamente, non sono significativi, ma ci fanno capire come si procede. Quando ad esempio facciamo un acquisto in Internet utilizzando la carta di credito è proprio questo quello che succede, con numeri molto elevati.

Conclusione: questo è solo il primo sguardo su un mondo articolato ed affascinante

Lo scopo di questa breve nota è solo quello di proporre a chi legge un piccolo assaggio di come la crittografia sia un mondo articolato e di quanta matematica occorra anche per comprare un libro o spedire un semplice messaggio online. Molti e più complessi sono i metodi crittografici, tanta è la matematica che viene utilizzata: avete un meraviglioso mondo da scoprire!

Bibliografia

- Celentani M. R., Cutolo, G. Un'introduzione all'aritmetica modulare, Incontri matematici del terzo tipo, pp. 121-141, Liguori Editore (2019)
- Celentani M. R., Leone, A. Aritmetica modulare: approfondimenti, applicazioni e... un po' di mistero, Incontri matematici del terzo tipo pp. 142-163, Liguori Editore (2019)
- Cutolo. G. <http://www.dma.unina.it/cutolo/didattica/varia/coinor-cutolo.pdf>
- Singh, S. Codici & Segreti, BUR (1999)

Biografie degli autori

Barra, Silvio	Silvio Barra è Ricercatore di Informatica presso il Dipartimento di Ingegneria Elettrica e delle Tecnologie dell'Informazione dell'Università di Napoli Federico II. Gli interessi di ricerca comprendono la Pattern Recognition in Immagini, Video e Serie Temporali, con particolare attenzione volta all'analisi del mercato finanziario, al riconoscimento biometrico ed alla videosorveglianza.
Brandolini, Barbara	Barbara Brandolini è attualmente Professore Ordinario di Analisi Matematica presso l'Università degli Studi di Palermo. E' stata Professore Associato di Analisi Matematica presso l'Università degli Studi di Napoli Federico II. I suoi interessi di ricerca sono rivolti allo sviluppo di metodi analitici e geometrici in problemi di Calcolo delle Variazioni ed Equazioni Differenziali.
Buonocore, Aniello	Aniello Buonocore è Professore Associato di Probabilità e Statistica presso il Dipartimento di Matematica e Applicazioni "Renato Caccioppoli" dell'Università di Napoli Federico II. I suoi interessi di ricerca sono orientati allo studio dei processi di diffusione e delle relative applicazioni. Negli ultimi anni si è interessato alla didattica della matematica con particolare riferimento al tema "Dati e Previsioni".
Caruso, Ugo	Ugo Caruso è Professore Associato di Chimica Generale ed Inorganica presso il Dipartimento di Scienze Chimiche dell'Università di Napoli Federico II. Le sue ricerche riguardano i materiali polimerici, organometallici e organici, in particolare l'optoelettronica, e lo studio di sonde fluorescenti per applicazioni ambientali e biologiche. È referente di Ateneo per il PLS-Scienza dei Materiali.
Celentani, Maria Rosaria	Maria Rosaria Celentani è Ricercatore di Algebra presso il Dipartimento di Matematica ed Applicazioni R.Caccioppoli dell'Università di Napoli Federico II. La sua attività di ricerca si svolge nell'ambito della Teoria dei Gruppi. Si è occupata di problemi legati alla didattica e ha partecipato a diverse attività di formazione scolastica; in particolare partecipa da diversi anni al Progetto Lauree Scientifiche.

del Gaudio, Rosanna	Rosanna del Gaudio è Professore Aggregato di Biologia Molecolare presso il Dipartimento di Biologia dell'Università degli Studi di Napoli Federico II. Negli ultimi anni la sua attività di ricerca si è prevalentemente orientata su problematiche connesse all'evoluzione con lo studio della regolazione dell'espressione genica tessuto-specifica in Vertebrati e Invertebrati e, in particolare, dei meccanismi molecolari che portano alla trascrizione e riattivazione di pseudo-geni in presenza di stress ambientale. Svolge attività di ricerca didattica, è referente del Dipartimento di Biologia per i percorsi PCTO ed è delegata all'orientamento del Corso di Studi in Scienze Biologiche.
D'Errico, Gerardino	Gerardino D'Errico è Professore Ordinario di Chimica Fisica presso il Dipartimento di Scienze Chimiche dell'Università degli Studi di Napoli Federico I. E' coordinatore del Corso di Laurea in Chimica Industriale. La sua attività di ricerca riguarda i processi di auto-aggregazione in soluzione acquosa, spaziando dai sistemi di interesse biologico (le membrane lipidiche) a quelli di rilevanza industriale (le formulazioni nanostrutturate). Le sue ricerche hanno portato a tre brevetti industriali.
Diana, Rosita	Rosita Diana è Ricercatore di Chimica Generale e Inorganica presso il Dipartimento di Agraria dell'Università degli Studi di Napoli Federico II. La sua attività di ricerca riguarda tematiche quali studio e sperimentazione di processi per la realizzazione di dispositivi fotovoltaici costituiti da materiali polimerici di sintesi e/o commerciali; progettazione, sintesi ed impiego di molecole e macromolecole ad attività mirata nell'ambito dell'interazione tra radiazione elettromagnetica e sistemi materiali; sintesi di leganti per composti di coordinazione in grado di fornire una risposta ottica specifica che trovano impiego come sensori di metalli, infatti tali molecole sono capaci di rilevare la presenza di metalli in soluzioni acquose; sintesi di chelanti chemosensori per metalli anche di interesse ambientale e biologico.
Di Capua Roberto	Roberto Di Capua è Professore Associato di Fisica presso il Dipartimento di Fisica dell'Università di Napoli Federico II. La sua ricerca verte essenzialmente sullo studio delle proprietà elettroniche, spettroscopiche e strutturali di composti ossidi di interesse per l'elettronica e di materiali a base di grafene per varie applicazioni (quali la biosensoristica e l'accumulo di energia). Una

	parte dell'attività di ricerca e della relativa produzione editoriale riguarda tematiche legate all'insegnamento della fisica e alla proposizione di esperimenti per studenti.
Di Donato Valentino	Valentino di Donato è Professore Associato di Paleontologia e Paleoecologia presso il Dipartimento di Scienze della Terra, dell'Ambiente e delle Risorse dell'Università degli Studi di Napoli Federico II. Si occupa prevalentemente di analisi dei dati in ambito micropaleontologico nonché di ricostruzioni paleoclimatiche e paleoambientali in ambito quaternario e in contesti geoarcheologici. Considera parte importante della propria formazione gli studi musicali svolti presso il Conservatorio di Avellino.
Di Giaimo, Rossella	Rossella Di Giaimo è Professore Aggregato di Biologia Molecolare presso il Dipartimento di Biologia dell'Università degli Studi di Napoli Federico II. Il suo lavoro di ricerca è incentrato principalmente sulla neurobiologia molecolare interessandosi alle basi molecolari della malattia rara umana EPM1 e utilizzando organoidi cerebrali umani come sistemi modello.
Esposito, Roberto	Roberto Esposito è Ricercatore di Chimica Generale ed Inorganica presso il Dipartimento di Scienze Chimiche dell'Università degli Studi di Napoli Federico II. La sua principale attività di ricerca riguarda la conversione catalitica di biomasse e derivati in prodotti chimici e l'applicazione di questi nell'industria chimica. In particolare, tra le biomasse, l'interesse è volto agli oli vegetali e agli oli vegetali esausti quali fonti di materie prime rinnovabili e a basso costo. Altre linee di ricerca riguardano la sintesi di complessi glicconiugati a base di platino e loro applicazione quali agenti antitumorali, chimica dei metalli di transizione con focus sull'idrometallurgia, chimica metallorganica del platino. È socio fondatore di ISusChem s.r.l., startup e spinoff universitario che si occupa di ricerca e innovazione nel campo dell'applicazione di prodotti derivati da biomasse nell'industria chimica di base e cosmetica. Dal 2017 al 2019 è stato autore dei quiz di Chimica Inorganica per i Giochi della Chimica organizzati dalla Società Chimica Italiana di cui è socio dal 2015.
Iacopini, David	David Iacopini è Professore Associato di Geologia Strutturale presso il Dipartimento di Scienze della Terra, dell'Ambiente e delle Risorse dell'Università di Napoli Federico II. E' stato docente di

	Geologia e Geofisica per oltre 10 anni presso l'Università di Aberdeen. Le sue principali attività di ricerca riguardano la geologia marina, geologia dei margini passivi, strutture geologiche del sottosuolo, imaging di faglie e esplorazione di risorse del sottosuolo attraverso il globo.
Iannace, Alessandro	Alessandro Iannace è Professore Ordinario di Geologia Stratigrafica presso il Dipartimento di Scienze della Terra, dell'Ambiente e delle Risorse dell'Università di Napoli Federico II. Si occupa principalmente di stratigrafia delle successioni calcareo-dolomitiche e delle informazioni che contengono in termini sia di storia del Pianeta che di applicazioni nella ricerca di risorse. E' Coordinatore dei Corsi di Studio di Laurea Triennale e Magistrale, ed è da tempo impegnato nei rapporti con le Scuole superiori nell'ambito dei progetti PLS Federico II nella scuola. E' referente di Ateneo per il Piano Lauree Scientifiche di Geologia.
Iesce, Maria Rosaria	Maria Rosaria Iesce è Professore Ordinario di Chimica Organica presso il Dipartimento di Scienze Chimiche dell'Università degli Studi di Napoli Federico II. L'attività di ricerca è rivolta principalmente alla messa a punto di nuove metodologie "green" per la sintesi di molecole naturali e/o analoghi di interesse biologico e allo studio delle interazioni della luce con sistemi eterociclici e con molecole bioattive. E' nel Consiglio Direttivo del Centro Interuniversitario di Ricerca sulle Reazione Pericicliche e Sintesi di Sistemi Etero e Carbociclici (CIRP). E' stata referente di Ateneo per il PLS di Chimica negli anni 2010-15, lo è attualmente a partire dal 2018 e coordina il gruppo dei referenti PLS per le varie aree disciplinari dell'Ateneo Federico II.
Manti, Lorenzo	Lorenzo Manti è Professore Associato di Fisica Applicata presso il Dipartimento di Fisica, Università di Napoli Federico II. I suoi principali interessi di ricerca riguardano la radiobiologia di fasci terapeutici di particelle cariche (adroterapia) e alterazioni citogenetiche radioindotte. È stato Presidente della European Radiation Research Society, ed è attualmente Consigliere eletto per la Fisica nel Consiglio Direttivo della SIRR (Società Italiana per la Ricerca sulle Radiazioni).
Mondillo, Nicola	Nicola Mondillo è Ricercatore di Giacimenti Minerari presso il Dipartimento di Scienze della Terra dell'Ambiente e delle Risorse dell'Università di Napoli Federico II. E' Scientific

	<p>Associate presso il Natural History Museum di Londra. La sua attività di ricerca riguarda principalmente la geologia, mineralogia e geochimica di giacimenti minerali a solfuri ed ossidati di Zn e Pb, giacimenti bauxitici e lateriti a Ni, con lo scopo di individuare le caratteristiche che permettono di determinarne la genesi ed ottimizzarne la coltivazione. Conduce inoltre studi sulla distribuzione di metalli critici (Ge, Ga, REE) in mineralizzazioni ad ossidati di Zn. Altri argomenti oggetto di studio nel campo giacimentologico sono la mobilizzazione supergenica dell'oro nei depositi primari a bassa solfidazione, e la formazione di minerali di REE in vene idrotermali contenenti fluorite.</p>
Palazzo, Lucio	<p>Lucio Palazzo è assegnista di ricerca in Statistica presso il Dipartimento di Scienze Politiche dell'Università Federico II di Napoli. I suoi principali interessi di ricerca riguardano i modelli non lineari per serie temporali discrete e l'analisi di dati sportivi.</p>
Ragozini, Giancarlo	<p>Giancarlo Ragozini è Professore Ordinario di Statistica Sociale presso il Dipartimento di Scienze Politiche, dove è coordinatore dei Corsi di Laurea in Statistica per l'impresa e la società e in Scienze Statistiche per le decisioni. E' responsabile scientifico dell'Osservatorio Regionale delle Politiche Giovanili in Campania per il medesimo Dipartimento. Tra i suoi interessi di ricerca figurano l'analisi delle reti sociali, l'analisi multivariata, l'analisi della condizione giovanile e i metodi statistici per l'analisi delle politiche sociali. È referente di Ateneo per il PLS-Statistica.</p>
Riccio, Daniel	<p>Daniel Riccio è Professore Associato di Informatica presso il Dipartimento di Ingegneria Elettrica e delle Tecnologie dell'Informazione dell'Università di Napoli Federico II e Associato di Ricerca presso l'istituto ICAR-CNR di Napoli. E' Co-responsabile del Projects of Intelligent Robotics and Advanced Cognitive Systems. I suoi interessi di ricerca includono principalmente la Computer Vision, la Pattern Recognition e l'Intelligenza Artificiale, con particolare interesse all'ambito dei sistemi biometrici e del medical imaging. È membro IEEE e dello IAPR, e responsabile per la certificazione GRIN del Corso di Laurea e Laurea Magistrale in Informatica. E' referente di Ateneo per il PLS di informatica.</p>

Santelli, Francesco	Francesco Santelli è assegnista di ricerca, presso l'Università di Napoli Federico II, all'interno di un progetto incentrato sulla mobilità universitaria. Ha svolto un periodo di ricerca all'Università di Trieste sull'analisi di dati provenienti da Social Network, in collaborazione con l'azienda di analisi e ricerche di mercato SWG. Tra i suoi interessi figurano anche la bio-statistica, il Text Mining e tecniche di Unsupervised Learning come l'Analisi degli Archetipi.
Scopece, Giovanni	Giovanni Scopece è Professore Associato di Botanica generale presso il Dipartimento di Biologia dell'Università di Napoli Federico II. La sua attività di ricerca è incentrata sullo studio delle interazioni mutualistiche ed antagonistiche tra pianta ed insetto. Ha numerose collaborazioni internazionali ed è stato ospite scientifico presso le Università di Zurigo (Svizzera), di Friburgo (Svizzera), del Western Australia e di Campinas (Brazil). E' membro della Società Europea di Biologia Evolutiva (ESEB) e della Società Botanica Italiana (SBI) ed è Associate Editor per Plant Biology (Wiley).
Scotti di Uccio, Umberto	Umberto Scotti di Uccio è Professore Associato di Fisica Generale presso il Dipartimento di Fisica "E. Pancini" dell'Università di Napoli Federico II. Nell'arco della carriera si è occupato di Fisica sperimentale, con particolare interesse alla Fisica della Materia condensata e delle proprietà degli ossidi funzionali (superconduttori ad alta temperatura, ferromagneti, dielettrici, ferroelettrici) e dei dispositivi a film sottile basati sulla tecnica della crescita epitassiale di ossidi. Più di recente si è interessato alla ricerca in Didattica della Fisica, con contributi prevalentemente centrati sui temi della Meccanica Quantistica insegnati in contesti scolastici e universitari.
Succoio, Mariangela	Mariangela Succoio è Dottore di Ricerca in Biochimica e Biologia Cellulare e Molecolare e Specializzata in Patologia Clinica e Biochimica Clinica. Ha svolto attività di ricerca di base e attualmente lavora per il Centro Unico Regionale di screening neonatale della Regione Campania presso il CEINGE Biotecnologie Avanzate. Svolge da tempo attività di divulgazione scientifica e orientamento per gli studenti.

Tarallo, Oreste	Oreste Tarallo è Professore Associato di Chimica Generale e Inorganica presso il Dipartimento di Scienze Chimiche dell'Università degli Studi di Napoli Federico II. La sua attività di ricerca è prevalentemente rivolta allo studio della strutturistica chimica e delle relazioni tra la struttura molecolare e le proprietà fisiche di materiali polimerici semicristallini e di materiali ibridi organico-inorganici.
Zambrano, Nicola	Nicola Zambrano è Professore Ordinario di Biologia Molecolare nella Scuola di Medicina e Chirurgia dell'Università degli Studi di Napoli Federico II. È attualmente il Coordinatore del Corso di Laurea triennale in Biotecnologie per la Salute e Vice-Direttore del Dipartimento di Medicina Molecolare e Biotecnologie Mediche dell'Ateneo Federico II. I suoi più recenti interessi scientifici riguardano la biologia del cancro e l'approccio biotecnologico all'immunoterapia dei tumori. Svolge attività editoriali e di revisione in ambito scientifico internazionale, e intensa attività di orientamento e divulgazione scientifica nei settori delle Biotecnologie e della Biomedicina.

Il Piano Lauree Scientifiche è oggi il principale strumento di orientamento universitario operante in Italia per le discipline scientifiche ed è comunemente riconosciuto con un esempio di *best practice*, al punto da divenire un modello per programmi simili in altri ambiti disciplinari.

Dal 2005, anno della sua istituzione, i vari dipartimenti dell'Università degli Studi di Napoli Federico II di riferimento per le aree disciplinari si sono progressivamente aggregati al progetto, accogliendo ogni anno migliaia di studenti delle scuole superiori di Napoli e provincia.

Le attività di orientamento per gli studenti ruotano attorno ai Laboratori PLS e alla Scuola Estiva per Studenti, in cui gli studenti si confrontano con tematiche e metodologie proprie delle discipline.

In questo volume sono raccolte le sintesi delle lezioni e gli esercizi presentati agli incontri della Scuola Estiva per Studenti del 2020.

