ISTITUTO LOMBARDO
ACCADEMIA di SCIENZE e LETTERE



Rendiconti

CLASSE DI SCIENZE MATEMATICHE E NATURALI

Vol. 151
(2017)

Istituto Lombardo di Scienze e Lettere

MILANO
pISSN 1974-6989
eISSN 2384-986X

ISTITUTO LOMBARDO - ACCADEMIA di SCIENZE e LETTERE

RENDICONTI

CLASSE DI SCIENZE MATEMATICHE E NATURALI

Vol. 151
(2017)

Istituto Lombardo di Scienze e Lettere

MILANO

ISTITUTO LOMBARDO - ACCADEMIA DI SCIENZE E LETTERE

RENDICONTI

CLASSE DI SCIENZE MATEMATICHE E NATURALI

Responsabile scientifico
Carlo Domenico Pagani

Comitato di redazione
Silvio Beretta
Stefano Maiorana
Cinzia Bearzot
Carlo Domenico Pagani

Pubblicato con il contributo di
Politecnico di Milano
Università Cattolica del Sacro Cuore, Milano
Università Commerciale L. Bocconi, Milano
Università degli Studi dell'Insubria, Varese
Università degli Studi di Milano
Università degli Studi Milano-Bicocca
Università degli Studi di Pavia

pISSN 1974-6989
eISSN 2384-986X
Copyright 2020



PAGEPress Publications
via A. Cavagna Sangiuliani, 5
27100 Pavia, Italy
T. +39.0382.464340
F: +39.0382.34872

STORIA DELLA LUCE: DAL SOLE AL LASER

VITTORIO DEGIORGIO (*)

Nota presentata dal s.c. Luigi Lugiato
(Adunanza del 2 febbraio 2017)

SUNTO. – Le sorgenti di luce inventate e poi sviluppate negli ultimi decenni (laser e led) hanno dato origine a rivoluzionari sviluppi scientifici e tecnologici. L'articolo ripercorre in modo sintetico la storia della luce, iniziata studiando le proprietà della luce solare, e proseguita introducendo con grande successo il concetto di onda elettromagnetica. Alla fine del 1800 alcune evidenze sperimentali hanno reso necessario complicare il "panorama" introducendo l'ipotesi quantistica, che implica la discretizzazione dell'energia elettromagnetica. La nuova ipotesi ha portato alla invenzione ed alla realizzazione dei laser e dei led. Attraverso l'utilizzo di queste sorgenti di luce sono stati fatti progressi fondamentali nella fisica di base e nella tecnologia, approfondendo la descrizione del dualismo onda-corpuscolo e del concetto di coerenza, mettendo su basi completamente nuove le tecnologie dell'informazione e della comunicazione, e fornendo nuovi strumenti a tutta la sensoristica industriale e biomedicale.

***

ABSTRACT. – The new sources of light (laser and led) that have been invented and developed during the last fifty-sixty years have given origin to a true scientific and technological revolution. This article synthetically follows the main steps of the history of light, which started by studying the properties of solar emission, and introducing the concept of electromagnetic wave. At the end of the 19th century there was experimental evidence that it was necessary to complicate the approach by introducing the assumption that the electromagnetic energy is a discrete, and not continuous, quantity. The new concept led to design and realization of the laser and the led. Through the utilization of these sources fundamental progress was achieved in both science and technology, by deepening the description of wave-particle dualism and of the coherence concept, by putting on a completely new basis information and communication technology, and by giving new tools to industrial and biomedical sensing.

(*) Professore Emerito, Università degli Studi di Pavia, Italy.
E-mail: vittorio.degiorgio@unipv.it

1. Il mondo antico e medioevale

La luce è l'inizio di tutto! Si comincia con "FIAT LUX", quando? 5775 anni fa interpretando la Bibbia, ma, naturalmente, la data è controversa. Per gli antichi Egizi la natura della luce era chiara: la luce è lo sguardo di Dio. Per i filosofi greci (Empedocle, Platone, Aristotele, ...) la luce diventa una emanazione dell'occhio umano. Il matematico Euclide scrive attorno al 300 a.C. il primo libro di Ottica. Non si occupa della natura della luce, ma delle proprietà geometriche dei raggi luminosi.

Nel 1015 Ibn al-Haytham (noto come *Alhacen* or *Alhazen* in Europa) (965 Basra – c. 1040 Cairo) scrive un libro di Ottica, resuscitando il lavoro di Euclide e producendo un'analisi sistematica della teoria ottica dell'antica Grecia. Il suo libro di Ottica (*Kitab al Manazir* in arabo), tradotto in latino come *De aspectibus* o *Perspectiva*, ha avuto una grande influenza sullo studio dell'Ottica in Europa.

Lo sviluppo più importante dell'Ottica nel Medioevo è consistito nell'invenzione degli occhiali, che sembra sia avvenuta in Italia tra il 1268 e il 1289. L'inventore è sconosciuto. In quell'epoca gli occhiali erano usati prevalentemente da monaci, i quali fabbricavano artigianalmente le lenti usando cristalli di grande durezza, come il quarzo. La qualità delle lenti ebbe grande miglioramento quando la produzione fu intrapresa dale vetrerie di Murano.

2. Da Galileo a Newton

Galileo Galilei (1564 – 1642) cambia la storia della fisica (e della scienza in generale) introducendo un concetto fondamentale: il primo passo nella ricerca scientifica deve consistere in osservazioni sperimentali dei fenomeni da studiare. Seguendo il suo stimolo nasce in Europa una attività di ricerca sperimentale che ha un grande impatto sullo sviluppo dell'Ottica e, in particolare, sullo studio delle proprietà della luce. Nella seconda metà del 1600 ci furono alcuni importanti risultati sperimentali riguardanti le proprietà della luce. Ne cito due:

– La prima misura della velocità di propagazione della luce, da parte di Roemer
– La scoperta che la luce bianca è la somma di tanti colori, da parte di Newton.

2.1 La velocità della luce

All'inizio del XVII secolo, molti scienziati erano convinti che non esistesse una cosa come "la velocità della luce"; pensavano che la luce potesse percorrere qualsiasi distanza senza impiegare tempo. Galileo non era d'accordo, e ideò un esperimento per misurare la velocità della luce insieme ad un assistente: ognuno prese una lanterna schermata e andarono sulla cima di due colline che distavano poco più di un kilometro. Galileo scoprì la sua lanterna, e l'assistente che stava sull'altra collina, non appena vide la luce, scoprì a sua volta la sua lanterna. Galileo avrebbe quindi dovuto misurare il tempo necessario per vedere la luce dall'altra collina.

L'esperimento non diede risultato. Il problema è che la velocità della luce (300.000 km/s) è semplicemente troppo grande per poter essere misurata su brevi distanze; per percorrere un kilometro, la luce impiega un tempo così piccolo (circa 0.000003 secondi) che non poteva certo venire misurato dagli strumenti di cui disponeva Galileo. Quindi occorre una distanza molto più grande, milioni di kilometri, sulla quale far viaggiare la luce.

Una esigenza molto importante nella navigazione transoceanica è sempre stata quella di individuare la posizione in longitudine della nave. Nel secolo XVIII particolarmente interessata all'argomento era la massima potenza navale dell'epoca, l'Inghilterra. Si ricordi il "Longitude Act", emesso dal governo inglese nel 1714. Per fare questa misura era necessario ricorrere ad osservazioni astronomiche. Negli anni intorno al 1670, l'astronomo reale danese Ole Roemer (1644-1710), si dice seguendo un suggerimento di Galileo, intraprese uno studio del moto dei satelliti del pianeta Giove, che ruota intorno al Sole con un raggio dell'orbita di circa 800 milioni di chilometri. In particolare l'osservazione si concentrò sul satellite di Giove chiamato Io. Io compie un'orbita completa intorno a Giove in 1.76 giorni. Roemer si aspettava di poter predire il suo moto con buona precisione, ma scoprì che il satellite Io non era sempre dove ci si aspettava che fosse. In certi periodi dell'anno sembrava essere in ritardo, in altri, sembava anticipare. Molto strano. Roemer noto però che Io sembrava essere *avanti* rispetto all'orbita prevista quando la Terra si stava avvicinando a Giove, e *indietro* quando la terra si stava allontanando. Ora, avendo misurato di quanto sembrava variare la durata dell'orbita di Io e di quanto varia la distanza fra la Terra e il Sole,

Roemer fu in grado di calcolare la velocità della luce con una discreta approssimazione.

2.2 I colori

Negli anni 60 del 1600 Isaac Newton (1642- 1727) compì una serie di esperimenti di scomposizione di luce bianca usando prismi di vetro, come nell'esempio mostrato in *Fig. 1*. In quei tempi l'opinione comune era che il bianco fosse il colore naturale della luce, e che il passaggio attraverso il prisma degradasse irreversibilmente il fascio di luce, in modo tale che la presenza di colore indicasse una situazione intermedia tra luce e buio. Per controbattere questa ipotesi Newton utilizzò una sequenza di due prismi, mostrando che era possibile con il secondo prisma ricomporre il fascio di luce bianca che era stato decomposto dal primo prisma.

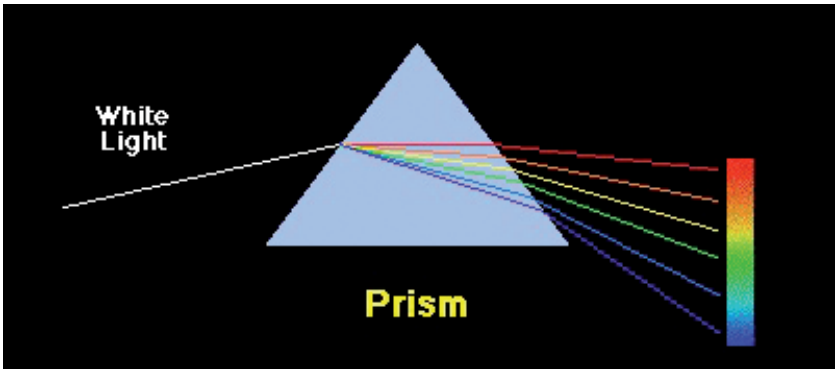


Fig. 1 – Scomposizione della luce bianca attraverso un prisma di vetro.

Nella *Fig. 2* è mostrata la distribuzione in lunghezza d'onda dell'intensità della luce solare, che presenta un picco in corrispondenza del giallo. Nella stessa Figura lo spettro solare è confrontato con quello di altre stelle che hanno una diversa temperatura superficiale. Si noti che esiste un perfetto adattamento biologico: il nostro occhio ha sensibilità centrata sullo spettro emesso dal sole. Se cambiamo stella diventiamo ciechi!

Newton si convinse che le proprietà mostrate dalla luce nei fenomeni di riflessione e rifrazione potevano essere spiegate solo

ricorrendo all'ipotesi che la luce fosse costituita di un flusso di particelle, e non da un'onda.

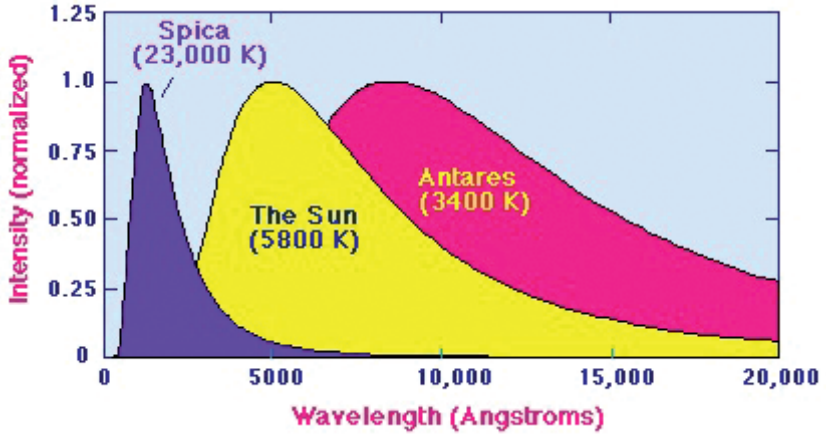


Fig. 2 – Spettro emesso da stelle a diversa temperatura.

3. Natura della luce

Che cosa è la luce? Un'onda (Christiaan Huygens, 1629-1695) o un flusso di particelle (Newton)? Diversi esperimenti svolti da Thomas Young, Augustin Fresnel, ed altri scienziati tra fine Settecento e inizio Ottocento mostrarono che la luce è un'onda, che trasporta energia e quantità di moto.

Una spettacolare unificazione di tutte le teorie concernenti l'elettricità, il magnetiamo e l'ottica fu compiuta da James Clerk Maxwell negli anni 60 del 1800, con la pubblicazione del lavoro "*A Dynamical Theory of the Electromagnetic Field*", e successivamente nel 1873 con il trattato "*Treatise on Electricity and Magnetism*" La teoria di Maxwell descrive il segnale luminoso come un insieme di onde elettriche e magnetiche che vibrano perpendicolarmente tra loro e perpendicolarmente alla direzione di propagazione.

La teoria generale dei fenomeni elettrici e magnetici proposta da Maxwell predice che possano esistere onde elettromagnetiche di qualsiasi frequenza, tutte viaggianti alla velocità $c = \lambda\nu$, dove λ è la lunghezza d'onda e ν la frequenza dell'onda elettromagnetica. A frequenza via

via crescente (lunghezza d'onda decrescente) le onde elettromagnetiche prendono il nome di onde radio, microonde, infrarosso, luce visibile, ultravioletto, raggi X.

La teoria di Maxwell ebbe subito ampie verifiche, ed aprì la strada ad importanti sviluppi tecnologici ed applicativi. Un passo molto significativo fu compiuto nel 1888 da Heinrich Hertz, che utilizzò la teoria di Maxwell per creare strumenti che potessero generare e ricevere impulsi a radiofrequenza. Di qui il successivo sviluppo delle radiocomunicazioni, della televisione, e degli apparecchi a microonde.

4. Nascita della fisica quantistica

La ricerca sperimentale svolta negli ultimi decenni del 1800 individuò però l'esistenza di alcuni fenomeni che non erano spiegabili con la teoria di Maxwell. Uno di questi è proprio la distribuzione in lunghezza d'onda (lo spettro) della radiazione solare riportata nella *Fig. 2*. Le misure effettuate su corpi riscaldati a diverse temperature T mostravano che la forma dello spettro e la lunghezza d'onda di picco non dipendono dalla natura del corpo, ma solo da T, ma non esisteva una teoria capace di spiegare i risultati sperimentali. Nella letteratura scientifica lo spettro della luce emessa è spesso chiamato spettro di corpo nero. Il corpo nero è un materiale ideale che assorbe completamente la radiazione incidente qualunque sia la sua frequenza. Max Planck escogitò una formula, che presentò il 19 ottobre 1900 in una riunione della Società di Fisica Tedesca, che descriveva con buona precisione i risultati sperimentali. L'aspetto sconvolgente della formula di Planck è che può essere giustificata solo ammettendo che l'energia posseduta dall'oscillatore a frequenza ν non sia espressa da un valore continuo, ma possa assumere solo valori discreti, che siano multipli di una quantità fondamentale $h\nu$, dove h è una costante universale, che è stata poi chiamata costante di Planck. La formula proposta da Planck per descrivere le curve della *Fig. 2* è la seguente:

$$\rho(\nu) = \frac{8\pi h\nu^3}{c^3} \frac{1}{\exp(h\nu / k_B T) - 1}$$

dove k_B è la costante di Boltzmann.

5. Laser e LED

5.1 Il laser

Qual'è l'origine dell'emissione di luce? Un insieme di atomi a temperatura T possiede una quantità di energia che si manifesta sotto forma di agitazione termica. Le collisioni fra atomi creano atomi eccitati. Adottando un semplice modello che vede l'atomo come un microscopico sistema planetario, in cui gli elettroni hanno il ruolo dei pianeti che ruotano attorno al nucleo, eccitare l'atomo significa portare uno, o più, elettroni su orbite più esterne rispetto a quelle percorse in stato di equilibrio. Con un tempo di rilassamento che dipende dal tipo di atomo e dal tipo di livelli energetici coinvolti l'atomo ritorna nella condizione di equilibrio emettendo l'energia in eccesso spesso sotto forma di onda elettromagnetica. Questo è il fenomeno cosiddetto di emissione spontanea di luce. L'onda emessa ha fase e direzione casuali. Un aspetto importante è che ogni tipo di atomo ha una serie discreta di livelli energetici, è quindi in grado di assorbire ed emettere solo un insieme discreto di frequenze.

In un articolo pubblicato nel 1917 Einstein presentò una nuova derivazione della formula di Planck utilizzando i concetti di assorbimento ed emissione, mostrando nel contempo che è però necessario ipotizzare anche l'esistenza di un nuovo processo, che chiamò emissione stimolata.

L'emisssione stimolata è un processo mediante il quale un quanto di luce (fotone) di frequenza appropriata interagisce con un atomo eccitato inducendolo a ritornare allo stato fondamentale emettendo l'energia in eccesso sotto forma di un fotone avente frequenza, direzione e polarizzazione identiche a quelle del fotone incidente (vedi *Fig. 3*). In termini ondulatori l'onda stimolata emessa ha anche la stessa fase dell'onda incidente, cioè è "coerente" con l'onda stimolante.

Il lavoro di Einstein non ebbe particolare eco nella letteratura scientifica per quasi 40 anni. Solo negli anni 50 del 1900 si è cominciato a riflettere sull'utilizzo dell'emissione stimolata per creare nuove sorgenti di onde elettromagnetiche. Nel 1951 Joseph Weber, all'epoca studente dell'Università del Maryland, propose di utilizzare l'emissione stimolata per realizzare un nuovo amplificatore di microonde. Nel 1953, Charles Hard Townes (Columbia University, USA) assieme ai suoi studenti James P. Gordon e Herbert J. Zeiger, riuscì a costruire

l'amplificatore di microonde, e gli diede il nome di "maser" (acronimo di "microwave amplification by stimulated emission of radiation").

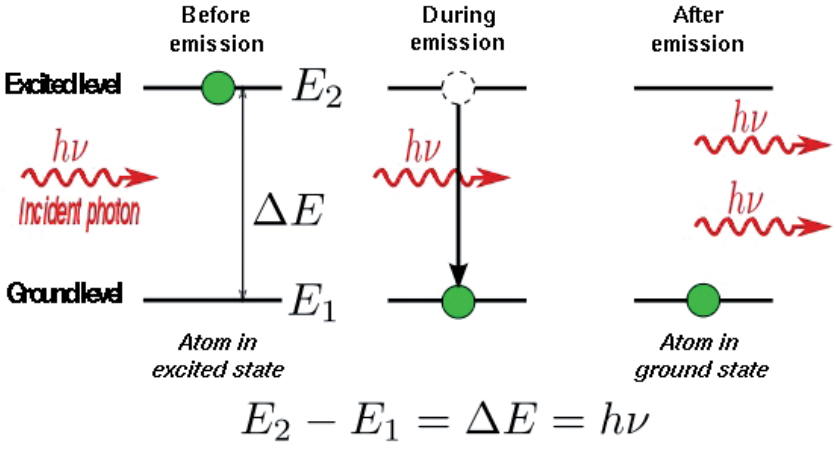


Fig. 3 – Emissione stimolata: il fotone incidente sull'atomo eccitato stimola l'emissione di un secondo fotone.

L'attività di ricerca successiva si indirizzò verso l'utilizzo dell'emissione stimolata per amplificare la luce invece delle microonde. Un ruolo importante fu giocato dai lavori dei ricercatori russi Nikolay Basov and Aleksandr Prokhorov, che mostrarono la necessità di coinvolgere più di due livelli energetici per ottenere un funzionamento in continua del dispositivo. In un lavoro pubblicato nel 1958 Townes e Schawlow proposero di costruire una sorgente di luce inserendo in una cavità costituita da due specchi paralleli un insieme di atomi eccitati allo stesso livello energetico. Grazie all'emissione stimolata, gli atomi eccitati presenti in cavità emettono luce in modo cooperativo e sincronizzato, generando, attraverso multiple riflessioni fra i due specchi, un fascio di luce direzionale (perpendicolare agli specchi) e monocromatico. Tecnicamente, si è costruito quello che si chiama un oscillatore, cioé un amplificatore ottico reazionato positivamente. Dalla frase "*light amplification by stimulated emission of radiation*" nasce l'acronimo "laser". Due anni dopo il lavoro teorico di Townes fu realizzato il primo laser da Theodore Maiman nei laboratory di ricerca della Hughes Aircraft Company (USA), utilizzando come mezzo attivo un cristallo di ossido di alluminio drogato con atomi di cromo (si tratta di una pietra preziosa, il rubino!) (*Fig. 4*).

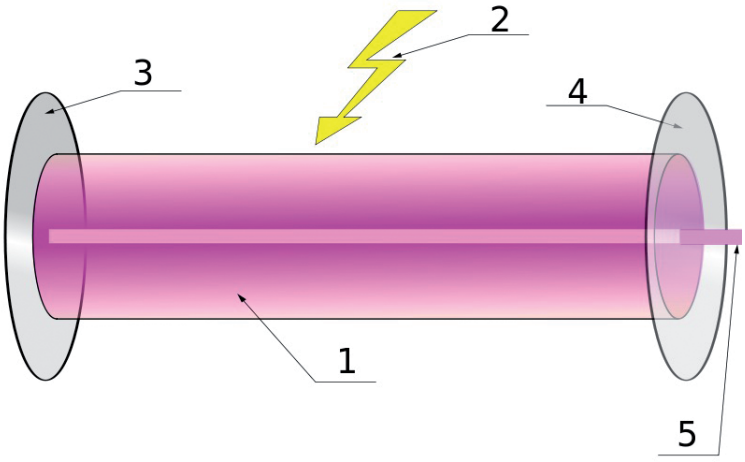


Fig. 4 – Laser a rubino: il cristallo contenente gli atomi eccitati è posto fra due specchi paralleli.

Nel giro di pochissimi anni furono poi realizzati molti tipi diversi di laser, utilizzando solidi cristallini ed amorfi, liquidi, e gas, coprendo un ampio spettro di lunghezze d'onda dall'infrarosso all'ultravioletto.

Una importanza particolare va ascritta al laser a semiconduttore: inventato nel 1962 presso i laboratori di ricerca della General Electric (USA) da Bob Hall e collaboratori. In alcuni tipi di semiconduttori i processi di ricombinazione elettrone-lacuna possono essere radiativi, cioè l'energia viene liberata sotto forma di onda elettromagnetica (emissione di un fotone). I processi di ricombinazione possono avvenire in un diodo a giunzione, dispositivo costituito da due zone drogate differentemente in modo da avere portatori di carica diversi. Il laser General Electric utilizzava una giunzione p-n di arseniuro di gallio (GaAs), ed emetteva nel vicino infrarosso alla lunghezza d'onda di 0.85 μm. Il laser è alimentato con una corrente elettrica continua o impulsata, l'asse della cavità laser ha direzione perpendicolare a quella del passaggio di corrente elettrica (vedi *Fig. 5*). Il laser ha dimensioni molto piccole, ed ha una efficienza enormemente superiore a quella di tutti gli altri tipi di laser: il rapporto tra la potenza luminosa emessa e quella elettrica utilizzata per il funzionamento può superare il 50%. Inoltre si tratta di un dispositivo comandato da una differenza di potenziale molto bassa, dell'ordine di 1 V.

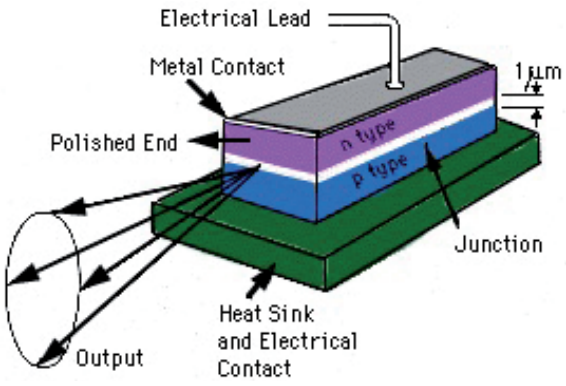


Fig. 5 – Laser a semiconduttore.

Il laser è una sorgente di luce che differisce profondamente dalle sorgenti tradizionali, perché emette un fascio di luce collimato e monocromatico. Tipicamente il fascio laser ha uno sparpagliamento angolare dell'ordine del milliradiante, cioè la macchia luminosa ha un diametro dell'ordine di un millimetro se osservata alla distanza di un metro. La lunghezza d'onda di emissione può essere definita con grande precisione, in un laser "singolo modo" la larghezza di banda rapportata alla lunghezza d'onda centrale, $\Delta\lambda/\lambda$, può valere 10^{-11}.

I laser possono funzionare in continua o anche impulsati. Impulsi laser ultrabrevi, dell'ordine dei femtosecondi, con intensità elevatissima, dell'ordine dei 10^{18} W/cm^2 , sono impiegati nelle più avanzate ricerche scientifiche.

Un aspetto di grande importanza scientifica è che la disponibilità di fasci laser di elevata intensità ha fatto nascere una nuova disciplina, l'ottica nonlineare. Inviando il fascio laser su opportuni cristalli, è ad esempio possibile, come illustrato nella *Fig. 6*, generare un nuovo fascio di luce a frequenza doppia.

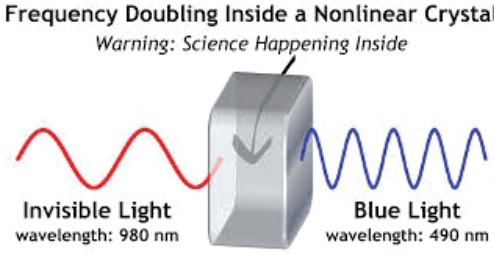


Fig. 6 – Ottica nonlineare: raddoppio di frequenza utilizzando un cristallo con proprietà ottiche nonlineari.

5.2 Diodo a emissione di luce, Light Emitting Diode (LED)

Il diodo a emissione di luce (noto con l'acronimo "led") è un sottoprodotto della ricerca sui laser a semiconduttore. Dal punto di vista della struttura il led è più semplice del laser perché non c'è una cavità ottica. L'emissione avviene su di una banda di frequenza più estesa e con un maggiore sparpagliamento angolare. Il primo led fu realizzato da Nick Holonyak, General Electric (USA), utilizzando una giunzione p-n di GaAsP, con emissione a $\lambda = 0.71$ μm. I pregi del led sono simili a quelli del laser a semiconduttore. Si tratta di una sorgente fredda, comandata a bassa tensione, di grande efficienza, e lunga durata. Sono disponibili led che emettono su tutti colori dello spettro visibile.

Un aspetto molto interessante dei dispositivi a semiconduttore è che esistono materiali organici che hanno proprietà simili a quelle dei semiconduttori "classici". Si possono realizzare oled (organic light emitting diode, diodo organico a emissione di luce) con materiali di tipo plastico, molto sottili e pieghevoli.

6. Applicazione della fotonica

E' impossibile qui descrivere, sia pure sommariamente, tutte le applicazioni che i dispositivi laser hanno e continueranno ad avere nei campi più disparati.

Mi limito ad alcuni cenni, scegliendo, con una certa arbitrarietà le applicazioni più diffuse e più semplici da spiegare. Le applicazioni del laser si possono dividere in tre categorie principali: a) tecnologie dell'informazione e delle comunicazioni ("Information and Communication Technology", ICT), che comprendono le comunicazioni ottiche (internet, telefonia a lunga distanza), le memorie ottiche (CD, DVD, blue-ray), le stampanti laser, e l'elaborazione ottica dei dati; b) strumentazione diagnostica e sensoristica, che comprende una grande varietà di tecniche ottiche per la telemetria, la velocimetria, la misura di grandezze meccaniche ed elettriche, la diagnostica ambientale e biomedicale; c) la lavorazione dei materiali con fasci laser intensi, che comprende operazioni di taglio, saldatura, foratura e marcatura di tutti i materiali di interesse industriale, ed anche la biomedicina: distacco della retina, eliminazione miopia, chirurgia laser.

Tra le applicazioni di tipo a) spiccano per la loro eccezionale importanza economica e sociale le comunicazioni ottiche. Il fascio di luce emesso dal laser a semiconduttore viene modulato in ampiezza e/o fase, in modo tale da inserire in esso l'informazione da trasmettere. Il fascio laser viaggia in fibra ottica attraverso la rete mondiale internet che copre tutto il globo terrestre utilizzando una fitta rete di cavi sottomarini con una capacità di trasmissione di alcuni Tbit/s. E' noto che un'onda elettromagnetica di dimensione trasversale finita si sparpaglia angolarmente propagandosi nel vuoto o in un mezzo omogeneo: questo è il fenomeno della diffrazione. E' possibile evitare lo sparpagliamento facendo propagare la luce in una guida d'onda. Lo schema più semplice di guida d'onda luminosa è costituito da un cilindro di vetro posto in aria, in cui i raggi luminosi che viaggiano formando angoli piccoli con l'asse del cilindro incontrano la parete vetro-aria con un angolo di incidenza superiore all'angolo limite, e vengono quindi totalmente riflessi. In realtà la struttura che viene utilizzata per le fibre ottiche comprende un nucleo ed un mantello fatti con due materiali vetrosi di diverso indice di rifrazione (vedi *Figg. 7* e *8*). Nella fibra ottica utilizzata per le telecomunicazioni viaggiano 10-50 miliardi di impulsi al secondo, ed il singolo impulso (un bit) dura 10-50 picosecondi. I cavi ottici sottomarini portano il 95% del traffico IP voce e dati tra continenti, ed il 100% del traffico internet internazionale. Un ruolo molto importante nella diffusione delle comunicazioni ottiche è stato giocato dall'affinamento delle tecniche di produzione industriale di fibre di altissima qualità a costi contenuti.

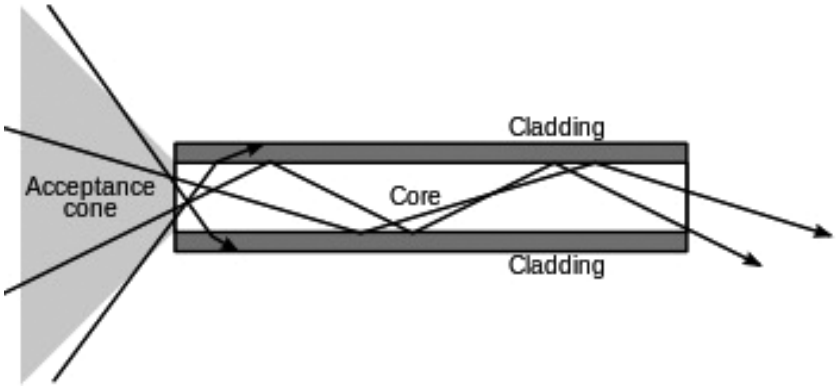


Fig. 7 – Fibra ottica: il fascio di luce si propaga per riflessioni totali interne.

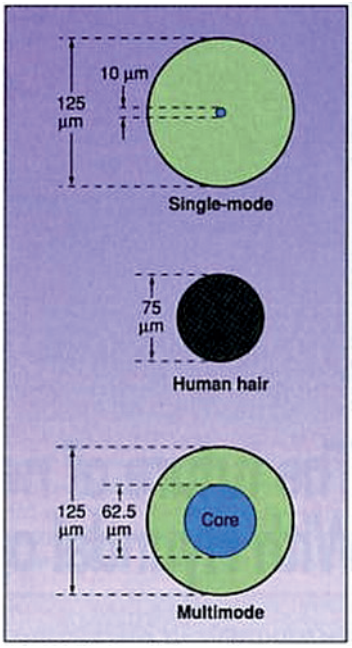


Fig. 8 – Sezione trasversale di una fibra ottica, le dimensioni sono paragonabili a quelle di un capello.

Le applicazioni di tipo b) coprono una vastissima gamma che si estende da aspetti fisici fondamentali ad aspetti industriali e biologici. Grazie alla monocromaticità e coerenza del laser si possono misurare con grande precisione distanza e velocità di oggetti in movimento, ed anche spostamenti estremamente piccoli. Riguardo agli aspetti di fisica fondamentale mi limito qui a citare l'interferometro laser chiamato LIGO che ha avuto un ruolo essenziale nel recentissimo esperimento di rivelazione di onde gravitazionali. In campo biologico segnalo la capacità del fascio laser opportunamente focalizzato di intrappolare oggetti microscopici, quali le cellule biologiche (*Fig. 9*). La spettroscopia laser è ampiamente utilizzata come metodo efficace e poco invasivo nella diagnostica medica.

Le applicazioni di tipo c) sono basate su processi termici che usano il raggio laser come fonte di calore, poiché esso può essere facilmente concentrato con delle lenti raggiungendo densità di potenza superiore a 1 MW/mm^2. L'elevatissima irradianza, data dal concentrare una grande potenza in un'area molto piccola, permette ai laser il taglio, l'incisione e

la saldatura di metalli. Nel momento in cui il laser interagisce con il materiale, l'energia dei fotoni viene assorbita dal materiale in lavorazione provocando localmente un rapido aumento di temperatura, che porta a fusione o ebollizione asportando il materiale senza contatto meccanico per espulsione del materiale fuso, vaporizzazione o meccanismi di ablazione. A differenza dei processi convenzionali non si ha usura e la rimozione del materiale non dipende dalla sua durezza, ma dalle proprietà ottiche del laser e le proprietà ottiche e termofisiche del materiale. Risulta per questo particolarmente indicato per lavorare leghe ad alta resistenza termica, carburi, compositi fibro-rinforzati, ceramici e stelliti.

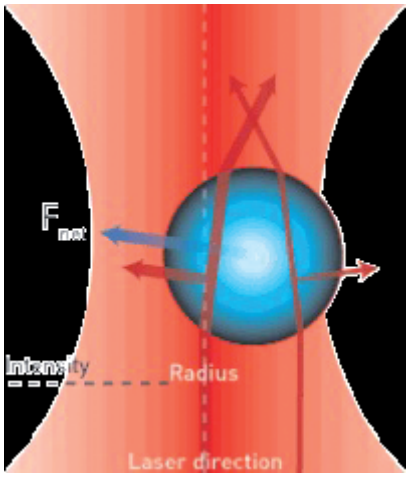


Fig. 9 – Particella di dimensioni micrometriche intrappolata focalizzando un fascio laser.

Nell'ingegneria meccanica spesso si impiegano pezzi e componenti in materiali resistenti. La marcatura laser è particolarmente indicata per garantire la tracciabilità, poiché è permanente e resistente anche a sollecitazioni meccaniche o agenti atmosferici. Può essere applicata a tutte le forme e praticamente su qualsiasi materiale, caratteristica che la rende adattabile ad un'ampia gamma di componenti. La marcatura di pezzi tramite laser può essere applicata a quasi tutti i materiali e in ogni forma. Il laser rappresenta quindi uno strumento universale per la marcatura dei pezzi più diversi.

Un settore importante delle lavorazioni con laser è rappresentato dalla chirurgia laser, soprattutto di tipo oculistico (eliminazione miopia, distacco della retina).

Uno sviluppo molto recente, di grande interesse applicativo, riguarda la fabbricazione tri-dimensionale (3D) di oggetti, denominata in inglese Additive Manufactoring. Oggetti di forma arbitrariamente complessa sono fabbricati strato per strato con programmazione al calcolatore.

Un enorme campo di applicazioni riguarda anche i led: schermi TV e megaschermi, illuminazione di ambienti, semafori, ... Mi limito qui ad accennare i vantaggi dei led dal punto di vista dell'illuminotecnica: la durata; l'assenza di costi di manutenzione; l'elevato rendimento; luce pulita, priva di componenti IR e UV; la flessibilità di installazione del punto luce; funzionamento a bassissima tensione, quindi in sicurezza; accensione a freddo (fino a -40°C); insensibilità a umidità e vibrazioni; durata non influenzata dal numero di accensioni/spegnimenti. Va aggiunto che lo spettro luminoso dei led varia molto in base alla tipologia e, se il led viene usato per l'illuminazione, lo spettro può essere sfruttato anche al 100%. L'utilizzo del led nell'illuminazione domestica punta a sostituire lampade a incandescenza, alogene o fluorescenti compatte (a risparmio energetico). Una normativa della Comunità Europea prevede che entro il 2016 non vengano più vendute lampade a incandescenza, in tutti i paesi della Comunità Europea, graduandone annualmente il divieto in base alla potenza in watt: le lampadine da 60 watt sono andate in pensione, in Europa, dal 1 settembre 2011; quelle da 100 sono state eliminate dal mercato nel settembre del 2009 e quelle da 75 nel 2010.

La straordinaria adattabilità degli oled apre le porte a campi di applicazioni diversissimi e finora immaginati solamente dalla mente degli autori di fantascienza. Gli oled possono infatti essere costruiti con materiali quasi del tutto trasparenti alla luce da essi prodotta, consentendo la realizzazione di schermi che, una volta spenti, appaiono come veri e propri vetri. Ancora più incredibile potrebbe sembrare la creazione di dispositivi flessibili, resa possibile dalla capacità degli oled di essere costruiti su substrati plastici (cioè flessibili). Uno schermo simile non ha limiti alle proprie applicazioni, in futuro potrebbe essere utilizzato facilmente per creare televisori pieghevoli, oppure essere applicato a vestiti o altro.

7. Commento finale

L'argomento trattato in questo articolo presenta nello stesso tempo un elevato contenuto scientifico ed un eccezionale potenziale applicativo. L'importanza scientifica è testimoniata efficacemente dalla

constatazione che premi Nobel in Fisica e Chimica legati ad argomenti di Fotonica sono stati assegnati, limitandoci agli ultimi 15 anni, negli anni 1999, 2000, 2001, 2005, 2009, 2012, e 2014.

In particolare i Premi Nobel 2014 per la Fisica e per la Chimica sono entrambi legati alla Fotonica;

- Fisica: Isamu Akasaki, Hiroshi Amano, Shuji Nakamura ***"for the invention of efficient blue light-emitting diodes which has enabled bright and energy-saving white light sources"***.
- Chimica: Eric Betzig, Stefan W. Hell and William E. Moerner ***"for the development of super-resolved fluorescence microscopy"***.

Vorrei esprimere un particolare ringraziamento al prof. Luigi Lugiato per avermi presentato ai membri dell'Istituto.

BIBLIOGRAFIA

Una referenza di carattere generale, con taglio divulgativo, è rappresentata dal testo di A. Frova, "*Luce colore visione*", Rizzoli (2000). Per un testo recente di carattere non-specialistico, si veda V. Degiorgio e I. Cristiani, "*Note di Fotonica*", Springer (2016).

Per quanto riguarda la parte storica (Sezioni 1-3) si può consultare ad esempio:

B.D. Guenther, "*Modern Optics*", Oxford University Press (2015).

Per gli articoli storici citati nelle Sezioni 4 e 5:

M. Planck "*On the Law of the Energy Distribution in the Normal Spectrum*", Annals of Physics, 4, 553 (1901).

A. Einstein, "*Zur Quantentheorie der Strahlung*", Physikalische Zeitschrift 18, 121-128 (1917). In inglese "*On the Quantum Theory of Radiation*", in D. ter Haar, "The old quantum theory", Pergamon (1967).

Per le sezioni 6 e 7:

R.J. Glauber, "*One hundred years of light quanta*", in K. Grandin, ed., The Nobel Prizes 2005, Nobel Foundation, 2006, pp. 75-98.

C.H. Townes, "*Production of Coherent Radiation by Atoms and Molecules*", in Nobel Lectures in Physics, 1963-1970, Elsevier 1972.

T.H. Maiman, "*Stimulation of Coherent Radiation in Ruby*", Nature 187, 493-404 (1960).

TRASPORTO DI CALORE ATTRAVERSO UN'INTERFACCIA: UN PROBLEMA CLASSICO RIVISITATO IN CHIAVE MODERNA

Nota del s.c. LUCIANO COLOMBO (*)

(Adunanza del 2 febbraio 2017)

SUNTO. – Descrivo alcuni esperimenti computazionali basati su simulazioni di dinamica molecolare, dimostrando che l'interfaccia tra due materiali solidi può essere descritta come un sistema termodinamico autonomo. Facendo uso della descrizione di Gibbs per tale interfaccia, discuto un modello teorico di termodinamica di non-equilibrio capace di fornire informazioni sulla sua resistenza termica e dimostro che tale resistenza può essere considerata come una proprietà specifica di interfaccia, unicamente determinata dalla temperatura locale.

***

ABSTRACT. – I describe a set of computational experiments using molecular dynamics simulations, showing that the interface between two solid materials can be described as an autonomous thermodynamical system. By making use of the Gibbs description for such an interface, I discuss a robust nonequilibrium thermodynamics theoretical framework providing information about its corresponding thermal boundary resistance. In particular, I show that the termal resistance of a junction between two pure solid materials can be regarded as an interface property, depending solely on the interface temperature.

1. DEFINIZIONE DEL PROBLEMA

La fisica del trasporto di calore è governata a livello macroscopico

(*) Istituto Lombardo, Accademia di Scienze e Lettere, Milano; Dipartimento di Fisica dell'Università di Cagliari, Cittadella Universitaria, Monserrato (CA), Italy. E-mail: luciano.colombo@dsf.unica.it

dalla legge fenomenologica di Fourier che in regime stazionario viene scritta nella forma [1]

$$\vec{J}_T = -\kappa \vec{\nabla} T \tag{1}$$

dove $\vec{J}_T$ rappresenta il flusso di calore (cioè la quantità di energia termica Q che attraversa una sezione A del campione nell'unità di tempo, dunque misurata in unità W m^{-2}), $\vec{\nabla} T$ è il gradiente di temperatura applicato ai capi di tale campione (misurato in K m^{-1}) e κ è la conducibilità termica del materiale (misurata in W m^{-1} K^{-1}). L'eq.(1) rappresenta la formulazione semplificata dell'equazione di Fourier valida se il materiale è omogeneo e isotropo e, dunque, solo quando la sua conducibilità può essere espressa da una grandezza scalare che, come dimostrato teoricamente, dipende unicamente da proprietà vibrazionali intrinseche del materiale. Per gli scopi di questa Nota, scritta in stile introduttivo e con intento principalmente didattico, è sufficiente ricordare che la teoria cinetica del trasporto dimostra che [2]

$$\kappa \sim \sum_\lambda C_V^{(\lambda)} \, l_\lambda \, v_\lambda \tag{2}$$

dove l_λ e υ_λ sono, rispettivamente, il libero cammino medio e la velocità di propagazione dello λ-esimo portatore microscopico di calore (fonone), mentre il termine $C_V^{(\lambda)}$ rappresenta il contributo di quest'ultimo al calore specifico (contributo reticolare).

L'esplosiva diffusione di tecnologie basate su nano-dispositivi (che sono sempre costituiti da molti strati di diversi materiali, di spessore usualmente compreso tra micron e nanometro) ha reso lo studio del trasporto di calore in sistemi eterogenei un problema di cruciale importanza applicativa. A livello nanometrico la situazione è resa molto complicata dall'esistenza di interfacce che, a seconda dei casi, possono formarsi tra materiali con differenti proprietà di conduzione termica (si pensi, ad esempio, ad un'interfaccia semiconduttore/metallo) oppure tra fasi differenti dello stesso materiale (si pensi ad un'interfaccia cristallo/amorfo). Conseguentemente, risulta di grande importanza capire il ruolo assunto da una interfaccia rispetto al passaggio di calore.

Queste necessità tecnologiche hanno aperto alcuni problemi di fisica fondamentale che, in estrema sintesi, possiamo riassumere in tre domande: (i) dove si colloca esattamente l'interfaccia tra due materiali

A e B? (ii) l'interfaccia A-B ha proprietà fisiche proprie, oppure è una entità puramente geometrica? (iii) come fluisce il calore attraverso un'interfaccia? Storicamente la risposta a queste domande è stata fornita in termini di uno schema puramente fenomenologico, basato sull'analogia tra la corrente termica e la corrente elettrica; nel seguito ne diamo un breve riassunto, come argomento propedeutico alla successiva discussione sul ruolo dell'interfaccia sviluppato alla luce delle più moderne teorie termodinamiche.

In conduttori metallici filiformi omogenei, il passaggio di corrente elettrica è descritto, nella sua forma più semplice, come

$$\Delta V = R_{el}\, I \tag{3}$$

dove ΔV è la differenza di potenziale elettrostatico applicata ai capi del filo e I è l'intensità di corrente (carica fluita nell'unità di tempo) che attraversa il mezzo di resistenza elettrica R_{el}. Anche nel caso elettrico possiamo definire un'equazione di trasporto in regime stazionario (legge di Ohm) simile all'eq.(1) nella forma

$$\vec{J}_V = -\,\sigma\, \vec{\nabla} V \tag{4}$$

dove $\vec{J}_V$ è il flusso di corrente elettrica (cioè la quantità di carica che attraversa una sezione A del campione nell'unità di tempo, dunque misurato in unità A m^{-2}) e σ è la conducibilità elettrica. Poiché in modulo vale la relazione $J_V = I/A$, possiamo sfruttare l'analogia formale tra le eq.(1) e (4) e porre $J_T = Q/A$ in modo tale che risulti definibile la resistenza termica R_T del sistema attraversato da un flusso termico $\vec{J}_T$ come

$$\Delta T = R_T\, Q/A \tag{5}$$

Si consideri ora la situazione rappresentata schematicamente in *Fig. 1*, corrispondente a un modello ideale di interfaccia (a questo livello definita come la semplice linea di separazione geometrica tra i due materiali A e B che la formano).

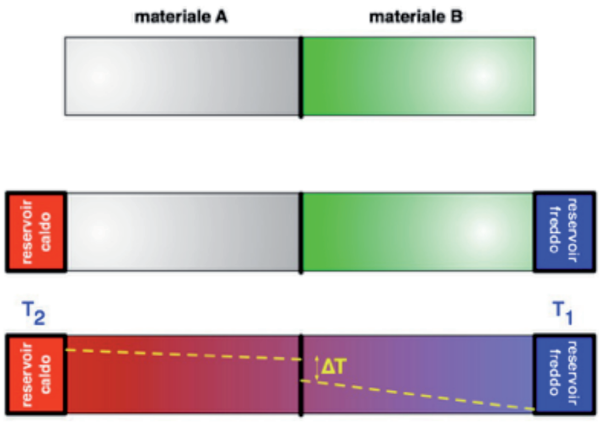


Fig.1 – Schema di interfaccia geometrica A-B e con relativa discontinuità ΔT di temperatura. Abbiamo supposto regime di conduzione termica stazionario.

Ai suoi capi si supponga applicata una differenza di temperatura $(T_2 - T_1)$, dove abbiamo supposto che $T_2 > T_1$ ovvero che il materiale a destra (A) sia accoppiato a un termostato a temperatura maggiore di quello a sinistra (B). In condizioni stazionarie, sperimentalmente si osserva che all'interfaccia il profilo di temperatura (linea tratteggiata gialla in *Fig. 1*) subisce una discontinuità ΔT. Ad essa viene associata una resistenza termica

$$R_T^{Kaptiza} = \frac{\Delta T}{Q/A} \tag{6}$$

detta resistenza di Kapitza [3]. Questa definizione fenomenologica non ha un vero fondamento microscopico e, inoltre, assume che l'interfaccia sia priva di dimensioni o proprietà. Nel seguito rivisiteremo criticamente entrambe queste assunzioni.

2. L'INTERFACCIA SECONDO LA COSTRUZIONE DI GIBBS

Supponiamo di poter identificare una certa proprietà fisica $P(z)$, per esempio legata a una caratteristica strutturale (fase cristallina del materiale o stato di aggregazione) o composizionale (specie chimiche presenti o concentrazione di una determinata popolazione atomica) dei

materiali costituenti l'etero-struttura di nostro interesse. Conveniamo di chiamare P_{left} e P_{right} i valori calcolati o misurati di tale proprietà nelle regioni di volume lontane dall'interfaccia stessa. Possiamo quindi definire l'interfaccia termodinamica A-B come quella regione in cui $P(z)$ si discosta significativamente dai valori P_{left} e P_{right}. Questa definizione porta alla costruzione di Gibbs [4] rappresentata in *Fig. 2*.

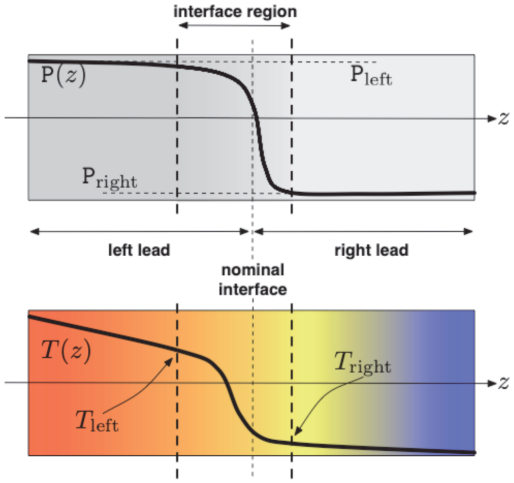


Fig. 2 – Alto: definizione della regione di interfaccia secondo la costruzione di Gibbs. I termini P_{left}, P_{right} sono definiti nel testo. Sotto: distribuzione di temperatura in condizioni stazionarie lungo la direzione normale alla interfaccia (direzione z). La linea tratteggiata verticale "nominal interface" rappresenta la posizione geometrica di confine tra materiale A e materiale B (si veda Fig. 1).

È importante notare due caratteristiche emergenti da tale costruzione: (i) l'interfaccia assume uno spessore finito che (ii) risulta distribuito in maniera non simmetrica a cavallo della interfaccia nominale (cioè assegnata con criterio geometrico). L'interfaccia, dunque, assume il ruolo di sistema fisico vero e proprio, per il quale è possibile una descrizione termodinamica autonoma.

3. La termodinamica dell'interfaccia

Seguendo gli argomenti sviluppati originalmente da Onsager nell'ambito della sua teoria termodinamica di non-equilibrio in regime

di risposta lineare [4], si dimostra che l'interfaccia di Gibbs è attraversato da una corrente di entropia che obbedisce a un'opportuna equazione di continuità. Questo concetto trova giustificazione fenomenologicamente ancora una volta invocando un argomento di analogia con il caso di una corrente di massa in un fluido (analogo idraulico).

Tralasciando gli aspetti matematici che possono essere trovati altrove [4], riassumiamo la situazione fisica in *Fig. 3*: nella regione di interfaccia il rateo totale di variazione temporale dell'entropia è dato dal bilancio tra il suo flusso in ingresso, il suo flusso in uscita e la sua produzione (con segno positivo o negativo) nella regione stessa. Quest'ultimo termine descrive l'azione di eventuali pozzi o sorgenti dell'analogo idraulico. Microscopicamente è associato a eventi specifici che influenzano localmente il flusso dei portatori microscopici di calore.

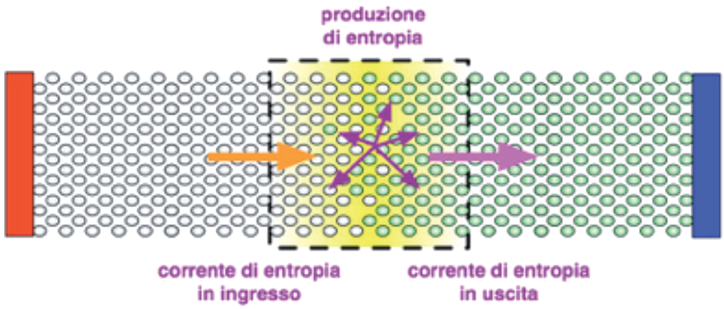


Fig. 3 – Rappresentazione grafica della equazione di continuità per il flusso di entropia. La regione di interfaccia (zona ombreggiata in giallo con contorni a linea tratteggiata) è interessata sia da una corrente in entrata (freccia in giallo-ocra) sia da una corrente in uscita (freccia in color magenta). Al suo interno si osserva produzione di entropia.

Sviluppando la matematica dell'equazione di continuità per la corrente di entropia, si dimostra che la resistenza termica di interfaccia è data da (le grandezze sono definite in *Fig. 2*)

$$R_T^{Onsager} = T_S^2 \frac{\left(\frac{1}{T_{right}} - \frac{1}{T_{left}}\right)}{Q/A} \tag{7}$$

dove le principali novità sono due, ovvero: (i) $R_T^{Onsager}$ dipende dalla differenza degli inversi delle temperature misurate in regioni remote ai due lati dell'interfaccia e (ii) compare in esplicito una proprietà termo-

dinamica dell'interfaccia, ovvero la sua temperatura T_S. Per quanto riguarda il primo punto, è evidente la differenza rispetto al caso di un sistema omogeneo trattato secondo la formulazione fenomenologica riportata in eq.(6) che lega la resistenza termica alla differenza di temperatura. Questo è uno dei risultati importanti della formulazione di Onsager che individua proprio nel gradiente di temperatura inversa la forza generalizzata responsabile del meccanismo di trasporto termico. Il secondo punto, invece, risalta la novità concettuale della costruzione di Gibbs: l'interfaccia è un sistema termodinamico autonomo, con sue proprie caratteristiche termodinamiche.

La nuova formulazione riassunta in eq.(7) ha il notevole vantaggio pratico di essere una teoria calcolabile, ovvero traducibile in un algoritmo computazionale risolvibile numericamente. In particolare, utilizzando la simulazione atomistica nella sua versione nota come "dinamica molecolare" è possibile il calcolo diretto delle tre temperature $\{T_{left}, T_S, T_{right}\}$ attraverso la stima dell'energia cinetica dei soli atomi compresi nelle rispettive regioni. La corrente termica Q/A che compare a denominatore è invece calcolata stimando il relativo flusso tramite la differenza di lavoro termico esercitato dai termostati caldo e freddo.

In conclusione, disponiamo ora di un dispositivo teorico che (i) consente di definire fisicamente cosa sia una interfaccia tra due materiali con diverse proprietà di trasporto termico, (ii) ha un fondamento teorico robusto inscritto nel dominio della termodinamica dei processi di non-equilibrio, (iii) ha carattere predittivo perché consente il calcolo numerico diretto di tutte le grandezze fisiche che regolano il fenomeno di nostro interesse.

4. Applicazione al caso dell'interfaccia Si/Ge

I sistemi a base di silicio e germanio sono oggetto di molte ricerche perché utili nel campo della conversione termoelettrica (trasformazione di gradienti di temperatura in corrente elettrica), considerata uno dei principi fisici più promettenti per sostenere il fabbisogno energetico riducendo l'uso di combustibili fossili, per la co-generazione di energia o anche solo per alimentare dispositivi dei più disparati generi da usare secondo la filosofia di impiego nota come "*embed and forget*".

I sistemi compositi Si/Ge sono costituiti da materiali abbondantemente disponibili, per i quali esiste una consolidata tecnologia di pro-

cesso e le cui caratteristiche di conduzione elettrica sono perfettamente sotto controllo grazie ai progressi ottenuti in campo microelettronico. Dei tre fattori che determinano la figura di merito termoelettrica [5]

$$ZT = \frac{\sigma S^2}{\kappa} \mathsf{T} \tag{8}$$

i due che compaiono a numeratore (rispettivamente, la conducibilità elettrica σ e il coefficiente Seebeck *S*) possono essere controllati alla bisogna con metodi ordinari di drogaggio chimico (la realtà è un più complessa, ma per la presente formulazione semplificata questo livello di dettaglio è sufficiente). Poiché maggiore è *ZT* maggiore risulta l'efficienza di conversione termoelettrica, l'imperativo tecnologico ancora da soddisfare è quello di minimizzare il trasporto di calore.

Come possibile strategia per raggiungere questo obiettivo è stata proposta la realizzazione di sistemi Si/Ge nano-strutturati quali, ad esempio, nano-compositi, nano-fili o sistemi multistrato [6,7]. Essi sono tutti caratterizzati da una stessa proprietà morfologica dominante, ovvero: contengono al loro interno un numero molto grande (e facilmente ingegnerizzabile) di interfacce Si-Ge. Poiché come discusso nella precedente Sezione ciascuna interfaccia è sorgente di resistenza termica, all'aumentare del loro numero si può immaginare di ottenere una significativa riduzione di κ. Da qui nasce la necessità di predire il valore della singola interfaccia, allo scopo di progettare la nano-struttura a minor valore di κ possibile.

Si consideri, dunque, una interfaccia tra due strati monocristallini di silicio e germanio, orientata secondo la direzione cristallografica [001], d'ora innanzi chiamata direzione *z*. Stante il carattere illustrativo di questo esempio modello, possiamo semplificare la geometria assumendo la situazione sperimentale di crescita pseudomorfica: imponendo che la costante reticolare nel piano di interfaccia sia quella del silicio, ricaveremo -secondo quanto previsto dalla teoria elementare della elasticità- che la parte germanio prossima all'interfaccia sia soggetta a un campo di deformazione che ne altera le tipiche distanze interatomiche lungo la direzione di crescita. Ci aspettiamo, inoltre, che questo campo di deformazione si espanda anche nello strato di silicio, ma anche in questo caso solo limitatamente a pochi piani atomici vicini all'interfaccia. Risulta abbastanza ovvio, a questo punto, individuare la proprietà $P(z)$ introdotta nella Sez.2 nella distanza inter-planare d_{in} misurata lungo *z*, ovvero perpendicolarmente all'interfaccia.

L'applicazione della costruzione di Gibbs al modello strutturale appena definito porta al risultato riassunto in *Fig. 4* [8].

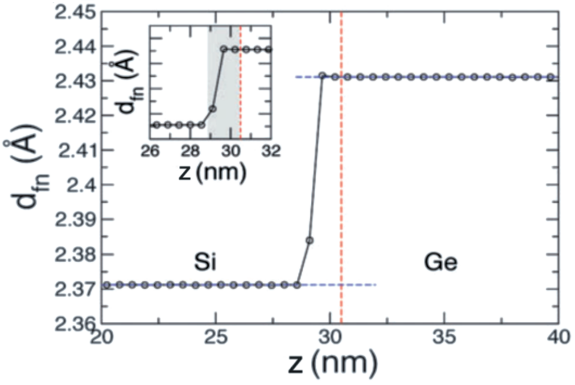


Fig. 4 – Definizione di interfaccia per una etero-struttura pseudomorfica Si/Ge cresciuta lungo la direzione z = [001] al passo reticolare del silicio, secondo la costruzione di Gibbs basata sulla distanza inter-planare d_{in} lungo z = [001]

Adesso imponiamo una differenza di temperatura ai capi della etero-struttura e, con i metodi tipici della dinamica molecolare, facciamo evolvere il sistema sotto questo vincolo fino al raggiungimento dello stato stazionario. Una volta raggiunta questa condizione, suddividendo il sistema in tanti strati di uguale spessore lungo la direzione normale all'interfaccia e calcolando la temperatura locale per ciascuno di essi dall'energia cinetica degli atomi ivi contenuti, possiamo finalmente ricavare il profilo di temperatura attraverso l'etero-struttura e, quindi, definire le temperature $\{T_{left}, T_{right}\}$ come riportato in *Fig. 5*.

Per quanto riguarda la temperatura dell'interfaccia è anch'essa calcolata sempre utilizzando le energie cinetiche atomiche, ma questa volta includendo nel calcolo tutte le particelle nella regione di *Fig. 5* [8] ombreggiata in grigio e corrispondente all'interfaccia di Onsager definita con simile artificio grafico in *Fig. 4*. Dobbiamo sottolineare un aspetto sottile, ma molto importante: la stessa differenza di temperatura ai capi della etero-struttura può essere ottenuta accoppiando i suoi estremi a due termostati aventi molte possibili coppie diverse di valori di temperatura. La simulazione atomistica dimostra che per ciascuna coppia, pur tutte corrispondenti ad un identico netto valore di differenza, risulta una diversa temperatura di interfaccia T_S. Dunque, ci dobbiamo aspettare che il valore di resistenza termica cambi al mutare della

condizione termodinamica in cui si trova la regione d'interfaccia di Gibbs.

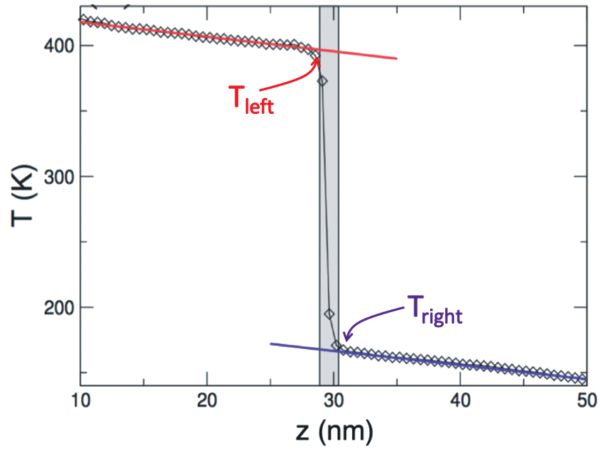


Fig. 5 – Profilo di temperatura per l'interfaccia di Fig. 4 e determinazione dei parametri necessari per l'applicazione della formula di Onsager data in eq.(7).

Disponiamo, finalmente, non solo di tutte le informazioni numeriche, ma anche dei criteri interpretativi necessari a per predire il valore di resistenza termica del sistema pseudomorfico Si/Ge cresciuto al passo reticolate del silicio lungo la direzione cristallografica [001] scelto come esempio applicativo. Il risultato è riportato in *Fig. 6* [8] e rappresenta il punto finale della nostra discussione.

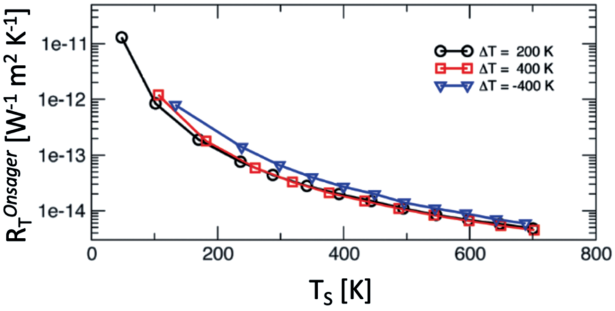


Fig. 6 – Resistenza termica calcolata per l'interfaccia di Fig. 4 corrispondenti a tre soli valori di differenza di temperatura ΔT applicata, corrispondenti a molte coppie diverse di temperatura dei termostati caldo e freddo: questo si riflette in altrettanti diversi valori di temperatura di interfaccia T_S.

Innanzitutto, osserviamo che la teoria Gibbs-Onsager permette una stima quantitativa della resistenza di interfaccia basata su informazioni calcolabili alla scala atomistica; inoltre, predice la sua diminuzione all'aumentare della temperatura di interfaccia. Questa in effetti è una caratteristica sorprendente e, per certi versi, addirittura contro-intuitiva se la si pone in relazione al risultato paradigmaticamente valido per tutti i materiali di volume che prevede la diminuzione della conducibilità termica all'aumentare della temperatura (nel regime di alte temperature).

In secondo luogo, si dimostra che la resistenza dipende unicamente dalla temperatura locale nella regione di Gibbs: è infatti evidente dalla *Fig. 6* che se la temperatura di interfaccia rimane inalterata, pur cambiando la polarizzazione termica ai capi dell'etero-giunzione, si osserverà un identico valore di resistenza. A parte il suo valore applicativo, è interessante notare che questo risultato corrobora *a fortiori* l'idea di considerare l'interfaccia come un sistema termodinamico autonomo: per esso è dunque definibile (e calcolabile!) l'energia interna e da essa tutte le proprietà termodinamiche di interesse.

5. Conclusioni

In questa breve Nota abbiamo dimostrato che un'etero-giunzione Si-Ge allo stato solido può essere considerata un sistema termodinamico autonomo, con proprietà esclusivamente dipendenti dalla temperatura dell'interfaccia.

Utilizzando informazioni strutturali alla scala atomica, è stato possibile identificare la regione interfacciale come uno strato di spessore pari a soli 16.6 Å, immerso nella regione di Si adiacente l'interfaccia nominale tra Si e Ge. La temperatura di interfaccia è ottenuta dall'energia cinetica media di questa regione.

I risultati derivano dall'uso di forze termodinamiche generalizzate, come definite nella termodinamica di non equilibrio che, dunque, va considerata come lo schema teorico più appropriato per lo studio dei fenomeni di trasporto del calore in sistemi eterogenei.

6. Ringraziamenti

Questo lavoro è basato su ricerche da me svolte durante l'anno

sabbatico passato tra l'anno 2015 e l'anno 2016 presso Institut de Ciència de Materials de Barcelona (ICMAB–CSIC) che ringrazio per l'ospitalità. I risultati di questa ricercar sono stati pubblicati in forma più estesa e completa in [8].

BIBLIOGRAFIA

[1] J.H. Lienhard IV e J.H. Lienhard V, "*A heat transfer texbook – 3rd edition*" (Phlogiston Press, Cambridge, Massachusetts, USA, 2006).

[2] N.W. Ashcroft e N.D. Mermin, "*Solid state physics*" (Holt, Rinehart & Winston, New York, USA, 1976).

[3] P.L. Kapitza, J. Phys., **4**, 181(1941).

[4] S. Kjelstrup e D. Bedeaux, "*Non-Equilibrium Thermodynamics of Heterogeneous Systems*"(World Scientific, Singapore, 2008).

[5] G.S. Nolas, J. Sharp e H. Goldsmid, "*Thermoelectrics: Basic Principles and New Materials Developments*" (Springer, New York, 2001).

[6] A.J. Minnich, M.S. Dresselhaus, Z.F. Ren, eG. Chen, Energy Environ. Sci. **2**, 466 (2009).

[7] M.S. Dresselhaus, G. Chen, M.Y. Tang, R. Yang, H. Lee, D. Wang, Z. Ren, e J.-P. Fleurial, Adv. Mater. **19**, 1043 (2007).

[8] R. Rurali, L. Colombo, X. Cartoixà, Ø. Wilhelmsen, T.T. Trinh, D. Bedeaux e S. Kjelstrup, Phys. Chem. Chem. Phys. **18**, 13741 (2016).

Scienze naturali
Botanica

Istituto Lombardo (Rend. Scienze) 151, 31-35 (2017)

NEW PLANT BREEDING TECHNOLOGIES FOR A SUSTAINABLE AGRICULTURE

MARTIN KATER (*)

Nota presentata dal m.e. C. Soave
(Adunanza del 9 marzo 2017)

SUNTO. – L'Agricoltura ha una storia lunga 10.000 anni. Durante questo periodo l'uomo ha selezionato mutanti naturali che risultavano essere utili per la coltivazione e il consumo umano. Tuttavia le caratteristiche selezionate non sono sempre favorevoli per le piante stesse, compromettendo spesso la loro capacità di sopravvivenza e adattamento nell'ambiente naturale. In questo modo le piante coltivate sono diventate dipendenti, per loro sopravvivenza, dall'uomo come l'uomo è dipendente dalle piante per sopravvivere. L'incremento della produzione agricola è stata notevole durante la rivoluzione verde e anche successivamente le nuove conoscenze genetiche hanno permesso di aumentare drasticamente la produzione di cibo. Tuttavia adesso con la rapida crescita della popolazione mondiale si è reso necessaria la selezione di nuove piante coltivate non solo in grado di produrre di più ma anche lo riescano a fare con meno risorse di acqua, fertilizzanti, insetticidi etc. Piante che siano più resistente a stress biotici e abiotici, sempre più frequenti in un mondo dove i cambiamenti climatici stanno diventando una grossa sfida per il futuro. Studiare i genomi delle piante e la funzioni dei loro geni è l'unico modo per poter accelerare il miglioramento genetico delle piante tutelando contemporaneamente l'ambiente. Una tecnologia molto importante per cercare di risolvere queste problematiche è il "genome editing", usando il sistema CRISPR-Cas. Purtroppo, in Europa l'uso di questa nuova tecnologia in agricoltura ha trovato non è di facile applicazione non tanto per motivazioni scientifiche ma per ideologie politiche contrarie a questo tipo di innovazione in agricoltura.

***

(*) Department of Biosciences, Università degli Studi di Milano, Italy.
E-mail: martin.kater@unimi.it

ABSTRACT. – Agriculture has a 10,000-year history. During this period, humans selected natural mutants that were found to be useful for cultivation and human consumption. However, the selected characteristics are not always favorable for the plants themselves, often compromising their ability to survive and adapt to the natural environment. In this way, cultivated plants have become dependent, for their survival, on man as man is dependent on plants to survive. The increase in agricultural production was notable during the green revolution and even afterwards new genetic knowledge made it possible to drastically increase food production. However, now with the rapid growth of the world population, the selection of new cultivars has become necessary, not only capable of producing more but also able to do so with less water, fertilizers, insecticides etc. Plants that are more resistant to biotic and abiotic stresses, which are increasingly frequent in a world where climate change is becoming a major challenge for the future. Studying the genomes of plants and the functions of their genes is the only way to accelerate the genetic improvement of plants while protecting the environment. A very important technology to try to solve these problems is "genome editing", using the CRISPR-Cas system. Unfortunately, in Europe the use of this new technology in agriculture has found it is not easy to apply not so much for scientific reasons but for political ideologies contrary to this type of innovation in agriculture.

Agriculture has a long history. About 10.000 years ago agriculture started in different parts of the world (*Fig. 1*). Important in this respect is the domestication of plants, which means that humans were selection natural mutants that showed favourable traits. The rare mutants or hybrids that they found in nature were not all used for consumption but they preserved them for the next season. In this way wild plant species were slowly on transformed into domesticated plants.



Fig. 1.

One of the most important characters that was independently selected in several grass species (cereals like wheat and rice) was the loss of seed shattering (*Fig. 2*). This allowed easy harvesting, instead of picking up seeds from the soil the seeds remained fixed to the plant. This

boosted food production. It is believed that this is one of the mutations that allowed human settlements instead of being nomads. The plants that can't spread their seeds anymore became dependent on humans and we became dependent on our domesticated plants for our food supply.



Fig. 2.

During the last 60 years, the green revolution (1950-1960) and subsequent integration of advanced research based innovation has drastically boosted agricultural output and has resulted in food abundance, particularly in the developing world. However, with the increasing world population, expected to reach 10 billion people in 2050, food production is becoming under pressure and yearly yield increases per hectare are likely to fall short the demand. Therefore, a new Green Revolution is needed to boost production per hectare but it has to be a sustainable "revolution" (*Fig. 3*). We will need crops with high yield potential, that need less water, less fertilisers, that are tolerant to biotic and abiotic stresses. Production per hectare needs to increase also because we can't keep on increasing agricultural land on the costs of natural habitats, like the tropical forests. We will have to produce "MORE" with "LESS". Is this possible?

Many technologies are being developed. Not only to improve our crops but also how we cultivate them. Satellite guided irrigation systems are a good example of high-tech agricultural practises.

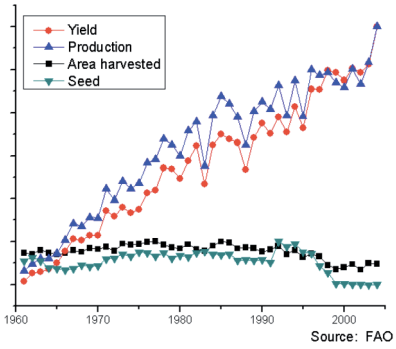


Fig. 3.

Plant Biotechnology for crop improvement is now approaching these challenges from a genomics perspective. Whole plant genomes are massively being sequenced through the application of high-tech DNA sequencing approaches, which are still rapidly increasing their capacity at a continuously reducing costs.

What can we do with the information that we get from these genomes. Many things, like comparing plant genomes, but important is that we can annotate the "complete" set of genes that a plant genome encodes. Considering that we have experimentally studied the function of about 5% of the genes of rice, you can imagine that obtaining functional knowledge about the remaining 95% of genes is an enormous resource for further crop improvement.

Summarising this part, in the past 10.000 years we have selected favourable mutations based on the plant phenotypes. Now in the new biotech era we select first the genes and create favourable mutations.

Genetically modified organisms (GMOs), by introducing beneficial genes from wild relatives or unrelated species is a technology that hasn't been accepted, especially in Europe, by a wide public and politicians. However, we already modify for years our plants by mutagenising them and subsequently selecting favourable mutations. Despite that this has a drastic impact on the genome, this technology has been broadly accepted and our daily food is largely coming from plants that have been mutagenised in this way.

Recently, a bacterial immune based system named CRISPR-Cas9 has substantially improved our tool kit to modify genomes (*Fig. 4*).

This system allows making mutations in specific locations in the

genome. We still have to introduce a transgene into the genome of a plant that contains the CRISPR-Cas based mutagenesis system. It makes mutations in a desired place of the genome and after that event we don't need the transgene any longer and it can simply being removed by crossing. This means that the transgene itself doesn't provide the favourable trait but it just creates the favourable mutations somewhere in the genome.

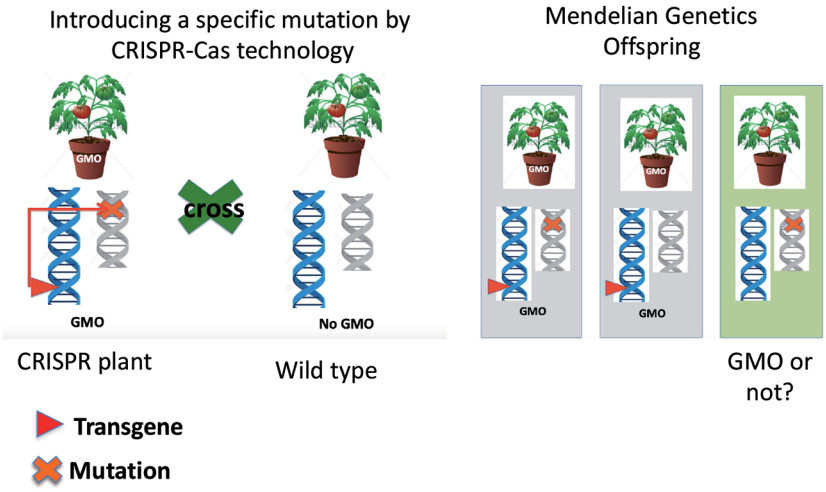


Fig. 4.

The question now is: such a mutagenised plant that doesn't contain a transgene is it a GMO? Scientifically speaking it isn't but the European parliament is not so sure…

TECNICHE PER MASCHERARE IL SAPORE AMARO DEI PRODOTTI FARMACEUTICI

ELISABETTA BRENNA (*)

Nota presentata dal m.e. C. Fuganti
(Adunanza del 23 marzo 2017)

SUNTO. – Nel testo vengono illustrate le principali tecniche usate in campo farmaceutico per mascherare il sapore amaro delle specialità medicinali contenute e rendere il prodotto più gradevole per l'assunzione orale da parte del paziente.

***

ABSTRACT. – The text describes the main techniques used in the pharmaceutical field to mask the bitter taste of active pharmaceutical ingredients and make the product more pleasant for oral intake by the patient.

INTRODUZIONE

Il sapore gioca un ruolo fattore fondamentale nel determinare la gradevolezza delle formulazioni farmaceutiche per uso orale, così da ottenere l'apprezzamento da parte del paziente. Il sapore conferisce al farmaco una caratteristica specifica ed unica, che lo può rendere preferibile ad altri prodotti. Questo aspetto è di particolare importanza nell'ambito dei medicinali da banco e di quelli per uso pediatrico.

La sensazione di amaro provocata in bocca da molti principi attivi farmaceutici è uno degli ostacoli maggiori che si incontrano nella messa

(*) Dipartimento di Chimica, Materiali ed Ingegneria Chimica "G. Natta", Politecnico di Milano, Italy. E-mail: mariaelisabetta.brenna@polimi.it

a punto di sistemi che consentano la somministrazione del farmaco per via orale (soluzioni, sospensioni, pastiglie/lastrine da sciogliere in bocca, gomme da masticare), modalità che consente maggior praticità di ingestione, oltre che rapidità di azione e di assorbimento. Le alternative possibili per i farmaci amari sono la somministrazione endovenosa, che è più laboriosa ed invasiva anche se più efficace, o mediante compresse rivestite, che però può presentare un problema per bambini e anziani, o per pazienti con particolari patologie. La legislazione europea corrente ha imposto un piano di sviluppo preciso per il controllo dei sapori amari e sgradevoli nei farmaci ad uso pediatrico. Il cattivo sapore di una medicina, infatti, può addirittura pregiudicarne l'assunzione e mettere a rischio i benefici ad essa correlati. D'altro canto, una medicina non deve essere "troppo buona", per evitare abusi ed assuefazioni.

Le tecniche per mascherare il sapore amaro dei medicinali si basano su tre approcci fondamentali: i) aggiunta di dolcificanti, aromi ed agenti effervescenti che coprano il sapore sgradevole; ii) sequestro della sostanza amara con metodi fisici, per evitare che essa venga in contatto con i calici gustativi; iii) utilizzo di composti che interagiscono con i recettori gustativi bloccandoli.

1. Aggiunta di dolcificanti, aromi ed altri eccipienti

L'aggiunta di dolcificanti ed aromi è il metodo più semplice per coprire il sapore amaro di un farmaco.

I dolcificanti sono di solito molto solubili in acqua, si sciolgono nella saliva e coprono i calici gustativi presenti sulla lingua, ritardando l'interazione con il farmaco amaro. Gli aromi conferiscono alla preparazione un sapore netto che tende a prevalere e sono usati in combinazione con altri agenti primari. Certi aromi raffrescanti, come il mentolo, "stordiscono" i calici gustativi e ritardano la percezione dell'amaro.

Anche lo zinco solfato si è rivelato efficace [1] nell'inibire il sapore amaro di alcune sostanze, come il chinico cloridrato, il tetralone e il benzoato di denatonio. Non ha invece effetto sul saccarosio ottaacetato, sulla pseudoefedrina e sul destrometorfano. La sua utilità è compromessa dal fatto che è anche un buon inibitore della dolcezza e potrebbe interferire in generale con il sapore di formulazioni complesse. L'unico dolcificante che non viene inibito dallo zinco solfato è il sodio ciclamato.

Il sapore sgradevole di certi colluttori e gocce per la tosse, contenenti medicinali e altre sostanze amare come l'olio di eucaliptolo, possono essere coperti usando fencone , borneolo e isoborneolo.[2] In alcuni dentifrici, il benzetonio cloruro viene mascherato con estratti di stevia con proprietà dolcificanti.[3] L'aggiunta di agenti anestetizzanti come il sodio fenolato a fili interdentali medicati con aspirina ha la funzione di stordire i calici gustativi per 4-5 secondi, rendendo impercettibile il sapore amaro dell'aspirina.[4] L'olio essenziale di chiodi di garofano, ad esempio, viene usato per mascherare il sapore amaro di analgesici, espettoranti, antitussivi, decongestionanti, grazie alla combinazione del suo sapore speziato e di un leggero effetto anestetico. Per sostenerne le capacità coprenti è possibile aggiungere aroma miele o vaniglia.

La combinazione di certi aromi con l'effetto di effervescenza, prodotto da acido citrico e sodio bicarbonato, è utile per coprire il sapore amaro di alcuni prodotti per i quali sia prevista la dissoluzione in acqua solo all'atto della somministrazione.

L'aspartame (*Fig. 1*) è il dolcificante più usato per ridurre il sapore amaro. Una concentrazione pari allo 0.8 % è sufficiente per ridurre del 25% la sensazione di amaro prodotta dal paracetamolo. Altri dolcificanti artificiali, come la neoesperidina diidrocalcone (*Fig. 1*) ed il corrispondente glucoside, sono in grado di mascherare amaro e salato grazie alla loro dolcezza duratura. La loro efficacia viene dal fatto che il profilo di gusto di una sostanza amara appare più lentamente nel tempo rispetto alla normale durata di un dolcificante.[5]

A prodotti farmaceutici contenenti saponine con proprietà anticolesterolemiche vengono aggiunti amminoacidi, quali la glicina e la alanina, per il controllo del sapore amaro.

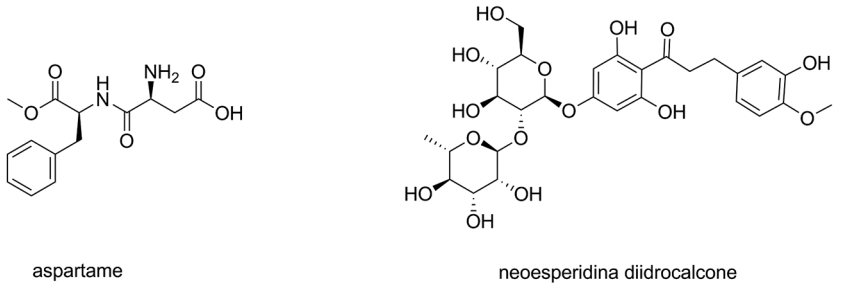


Fig. 1 – Strutture dell'aspartame e della neoesperidina diidrocalcone.

2. Formazione di complessi di inclusione

Tale approccio sfrutta la possibilità che la molecola di farmaco si adatti alla cavità di un agente complessante, a dare un complesso stabile. Il metodo ha efficacia nell'attenuare il sapore amaro, sia perché risulta diminuita la solubilità orale del farmaco, sia perché diminuisce la quantità di sostanza esposta ai calici gustativi. Gli agenti complessanti più usati sono le ciclodestrine a 6, 7 oppure 8 unità glucopiranosiche collegate da legami α-glicosidici (α, β, e γ-ciclodestrine, rispettivamente) (*Fig. 2*).

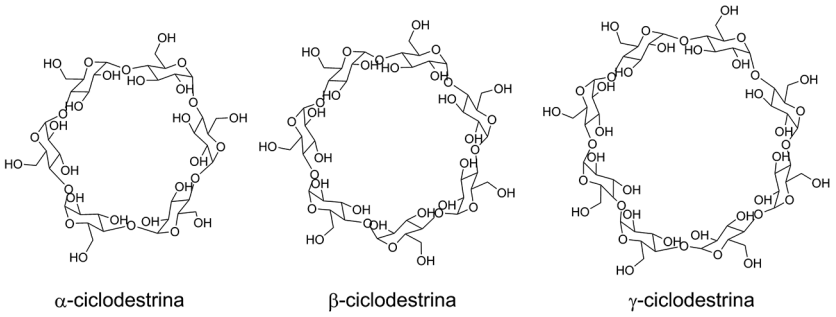


Fig. 2 – Struttura delle più comuni ciclodestrine.

Le ciclodestrine danno luogo ad una caratteristica struttura a tronco di cono con un diametro interno di 5-10 Å, una superficie interna idrofobica ed una esterna idrofilica. In soluzione il cono è idratato, ma piccole molecole idrofobiche o porzioni idrofobiche di molecole grandi possono legarsi mediante forze di Van der Waals alla cavità interna. L'affinità della ciclodestrina per una data sostanza dipende dal valore della costante di associazione (K_{ass}).

La velocità di formazione dei complessi si avvicina al limite di diffusione, cosicchè è possibile avere un buon eccesso di ciclodestrine presenti nel prodotto farmaceutico prima che sia assunto.

La cetrazina forma dei complessi 1:1 in soluzione con le ciclodestrine α β ε γ.[6] Le ciclodestrine più piccole α e β formano un complesso stabile con una parte della cetrazina, mentre la ciclodestrina più grande forma un complesso meno stabile con l'intera molecola. Le ciclodestrine α e β sopprimono il sapore amaro della cetrazina, mentre la ciclodestrina γ è inefficace, forse appunto per il basso valore di K_{ass}. E' stata messa a punto, ad esempio, una gomma da masticare contenente cetrazina e ciclodestrine.[7] Queste ultime favoriscono il rilascio del

principio attivo dalla gomma durante la masticazione. Tuttavia, a causa della velocità di interazione, avere semplicemente le ciclodestrine nella formulazione ha gli stessi effetti che usare i complessi pre-formati.

In alcuni casi, un terzo agente viene aggiunto per promuovere l'interazione ospite-ciclodestrina. Per esempio, la famotidina ha una K_{ass} relativamente bassa con la β-ciclodestrina, ma l'aggiunta di idrossipropil metil cellulosa aumenta l'interazione di legame e diminuisce il sapore amaro.[8]

3. Formazione di rivestimenti con materiali polimerici

Il rivestimento è uno dei metodi più efficaci usato comunemente per coprire il sapore amaro. Polimeri idrofobici, idrofilici, e lipidi possono essere impiegati come materiali per il rivestimento, da soli o in combinazione, in soluzioni mono o multistrato, in processi di preparazione condotti in acqua o in solvente organico. Con questa tecnica si impedisce che il farmaco amaro venga in contatto con i recettori gustativi quando viene introdotto nella cavità orale. La scelta del materiale di rivestimento deve essere fatta, però, tenendo conto che non deve alterare il profilo di rilascio del farmaco *in vivo*. Agenti inerti di rivestimento che vengono usati sono: il polivinilpirrolidone, la gelatina, la metil e la etil cellulosa.

Ad esempio il sapore amaro dell'ibuprofen è stato coperto usando delle microcapsule, preparate con la tecnica dell' air suspension coating. Le microcapsule hanno un cuore di ibuprofen microcristallino, rivestito di copolimero acrilico e sono usate per creare un prodotto masticabile senza sapore amaro.

Uno dei metodi più efficienti per ottenere le microcapsule rivestite è l'impiego dei processori a letto fluido. Gli ingredienti sono miscelati e preriscaldati da un flusso di aria calda diretto verso l'alto. La granulazione ha luogo nebulizzando un liquido nella polvere fluidificata. Successivamente i granuli sono essiccati con aria calda. Una seconda alternativa è rappresentata dalla deposizione di strati successivi di un composto attivo su un nucleo di materiale inerte, ad esempio sfere di zucchero. Il farmaco amaro viene sciolto o disperso in un solvente, che può essere anche acquoso, insieme con un legante che consenta l'aderenza delle molecole di farmaco al substrato inerte. Come leganti vengono usati l'idrossipropilcellulosa (HPC), polivinilpirrolidone, carbossimetilcellulosa. Le particelle rivestite di strati di farmaco sono successivamente ricoperte con lo strato di polimero. Come polimeri si usano:

etilcellulosa, Eudragit E-100, HPC, polivinilalcol e polivinilacetato.

Un terzo approccio consiste nell'effettuare la granulazione del farmaco e poi rivestirlo di polimero. La granulazione fa diminuire l'area superficiale delle particelle di farmaco, aumentandone le dimensioni. In tal modo diminuisce la quantità di polimero da usare. Per questi rivestimenti possono essere usate anche proteine, quali la zeina, la gliadina e l'ordeina o la gelatina.

4. Formazione di complessi con resine a scambio ionico

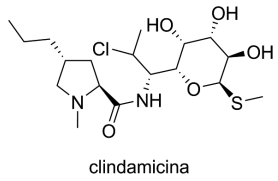
Le resine a scambio ionico sono polimeri ad alto peso molecolare con gruppi funzionali acidi o basici. Possono essere usate per legare i farmaci mediante esposizione ripetuta della resina alla soluzione contenente il farmaco in una colonna cromatografica, oppure mediante contatto prolungato della resina nella soluzione del farmaco. Le molecole di farmaco sono legate alla resina mediante interazioni ioniche, tali che la dissociazione non avvenga nelle condizioni di pH della saliva. Il rilascio del farmaco avviene più avanti nel tratto gastrointestinale. Le resine acide a scambio cationico sono usate con farmaci basici, quelle basiche a scambio anionico sono usate per farmaci acidi. Ad esempio, la formazione di un complesso chinolone-resina consente di eliminar il sapore amaro dei chinoloni e ottenere una formulazione liquida per uso orale.[9] Il chinolone viene sciolto in solvente acquoso, viene aggiunta la resina a dare il complesso. Questo viene poi sospeso direttamente in uno sciroppo (estratto di malto) con l'aggiunta di un agente anti-caking (silice colloidale) ed un conservante (acido sorbico).

5. Preparazione di pro-farmaci

Per profarmaco si intende un precursore inerte di un farmaco che mediante una reazione che avvenga *in vivo* possa liberare la sostanza farmacologicamente attiva. Il sapore amaro di una sostanza deriva dall'efficienza con cui il recettore dell'amaro interagisce con la sostanza stessa, che dipende a sua volta dalla struttura chimica della sostanza. Modificando la struttura della sostanza possiamo modificarne la percezione gustativa. L'altra modalità di azione è la riduzione della solubilità nella saliva.

Gruppi carbossilici ed alcolici si prestano ad essere derivatizzati

mediante esterificazione. Gli enzimi esterasi presenti poi nell'organismo possono rilasciare il principio attivo per idrolisi. Questo è l'approccio usato nel preparare l'estere palmitico della clindamicina (*Fig. 3*).



clindamicina

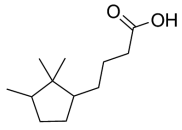
Fig. 3 – Struttura della clindamicina.

6. Uso di Bitter Receptor Blockers

Un approccio che può essere utilizzato per mascherare il sapore amaro è quello di utilizzare delle molecole che inibiscano l'attivazione dei recettori del gusto amaro presenti sulla nostra lingua.

Per 21 dei 25 recettori del gusto amaro sono stati identificati uno o più agonisti.[10] Sono invece assai rari i composti che sono stati trovati agire come antagonisti di tali recettori.[11]

Il composto più conosciuto e potente è l'acido 4-(2,2,3-trimetilciclopentil)butanoico (GIV3727, *Fig. 4*), preparato e studiato dalla Givaudan nel 2010: è in grado di attenuare il retrogusto amaro della saccarina e dell'acesulfame K, inibendo l'attivazione dei recettori of hTAS2R31 and hTAS2R43.



acido 4-(2,2,3-trimetilciclopentil)-butanoico (GIV3727)

Fig. 4 – Struttura del composto GIV3727.

Un altro composto in grado di diminuire l'attivazione dei recettori dell'amaro è l'acido *p*-(dipropilsulfamil)benzoico (noto come probenecid) (*Fig. 5*). Tale sostanza è in gardo di inibire l'attivazione dei recettori hTAS2R16, hTAS2R38, and hTAS2R43, e di sopprimere il sapore amaro della salicina nei test sensoriali. Esistono, invece, delle sostanze che si comportano come agonisti verso un certo set di recettori e come antago-

nisti verso un altro set. Si tratta dei due lattoni sesquiterpenici 3β-idrossidiidrocostunolide (3HDC) and 3β-idrossipelenolide (3HP) (*Fig. 5*).

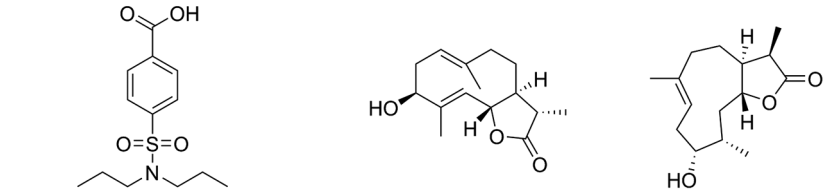


Fig. 5 – Struttura di alcuni composti usati per ridurre l'attivazione dei recettori del sapore amaro.

Nel 2013 ricercatori della Unilever hanno riportato che tre flavanoni, 4'-fluoro-6- metossiflavanone, 6,3'-dimetossiflavanone, e 6-metossiflavanone (*Fig. 6*), mostrano, in ordine di potenza decrescente, comportamento inibitorio verso l'attivazione del recettore hTAS2R39 ad opera della epigallocatechina gallato (ECG). Il 6-metossiflavanone inibisce anche l'attivazione del recettore hTAS2R14, sempre attivato dalla ECG. Lo stesso effetto è stato osservato con un agonista strutturalmente differente, il benzoato di denatonio.

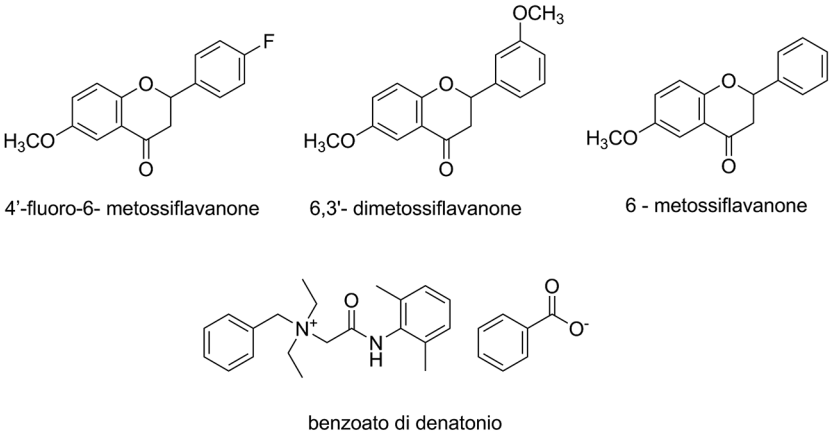
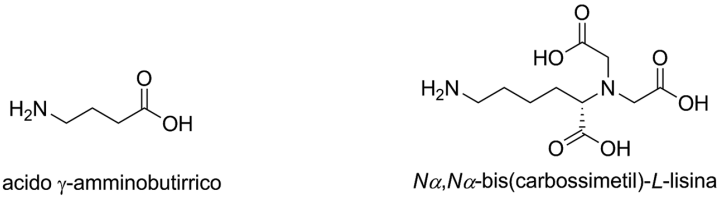


Fig. 6 – Struttura di alcuni inibitori di specifici recettori del sapore amaro.

Nel 2014, sono stati individuate altre due sostanze in grado di comportarsi come inibitori del recettore T2R4 attivato dalla chinina:

l'acido γ-amminobutirrico (GABA) e la *N*α,*N*α-bis(carbossimetil)-L-lisina (BCML) (*Fig. 7*). Lo studio farmacologico ha permesso di stabilire che il GABA agisce come antagonista, mentre BCML si comporta come agonista inverso.



acido γ-amminobutirrico — *N*α,*N*α-bis(carbossimetil)-*L*-lisina

Fig. 7 – Altri inibitori dei recettori del sapore amaro.

Per l'uso pratico di questi bloccanti dell'amarezza, devono essere soddisfatti i seguenti requisiti:

i) i bloccanti devono essere efficaci a basse dosi. Un antagonista che deve essere usato in quantità equimolare o maggiore rispetto all'agonista non è efficiente;
ii) il composto deve funzionare quando applicato simultaneamente al composto amaro e deve possibilimente inibire tutti i recettori del sapore amaro;
iii) deve essere possibilmente di origine naturale ed essere disponibile in quantità sufficiente e a basso costo;
v) devono essere riconosciuti come GRAS (Generally Recognized as Safe) dalla Federation of Flavors and Extracts Manufacturers (FEMA).

Conclusioni

Diverse sono le soluzioni a disposizione per coprire il sapore amaro dei farmaci, a seconda del tipo di specialità medicinale e della particolare applicazione.

I requisiti fondamentali richiesti perché l'uso in campo farmaceutico possono essere così riassunti:

a) richiedere il minor numero di passaggi di lavorazione;
b) poter essere condotto a temperatura ambiente, senza uso di apparecchiature specifiche;
c) fare uso del minor numero di eccipienti, che devono essere facil-

mente disponibili, a basso costo ed ammessi per l'uso in campo farmaceutico;

d) non devono avere effetti negativi sulla biodisponibilità del farmaco.

BIBLIOGRAFIA

[1] Keast RSJ, Breslin PAS. Bitterness Suppression with Zinc Sulfate and Na-Cyclamate: A Model of Combined Peripheral and Central Neural Approaches to Flavor Modification, *Pharmaceutical Research*, 2005: 22: 19070-1977.

[2] Hussein MM, Bareclon SA. Taste-Masking Agents for Bitterness of Volatile Oils, US Patent 4,983,394, January 8, 1991.

[3] Yokoo T, Hirohata H. Composition for Oral Cavity, JP 05,000,931, January 8, 1993.

[4] Fuisz RC. Taste-Masking of Pharmaceutical Floss with Phenol, US Patent 5,028,632, July 2, 1991.

[5] Kurtz RJ, Fuller WD. Ingestibles Containing Substantially Tasteless Sweetness Inhibitors as Bitter Taste Reducers or Substantially Tasteless Bitter Inhibitors as Sweet Taste Reducers, US Patent 5,232,735, August 3, 1993.

[6] Stojanov M, Wimmer R, Larsen KIML. Study of the inclusion complexes formed between cetirizine and α -, β -, and γ -cyclodextrin and evaluation on their taste-masking properties, *J Pharm Sci.,* 2011:100:3177-85.

[7] Stojanov M, Larsen KL. Cetrizine release from cyclodextrin formulated compressed chewing gum. *Drug Dev. Ind. Pharm*., 2012:38: 1061-7.

[8] Patel AR, Vavia PR. Preparation and evaluation of taste masked famotidine formulation using drug/beta-cyclodextrin/polymer ternary complexation approach, *AAPS Pharm Sci Technol*., 2008;9:544-50.

[9] Gao R, Shao ZJ, Fan, ACL, Witchey-Lakshmanan LC, Stewart DC. Taste Masking of Oral Quinolone Liquid Preparations Using Ion Exchange Resins. US Patent 6, 514,492, February 4, 2003.

[10] a) Meyerhof W, Batram C, Kuhn C, Brockhoff A, Chudoba E, Bufe B, Appendino G, Behrens M. The molecular receptive ranges of human TAS2R bitter taste receptors, *Chem. Sens*. 2010:35: 157-170; b) Thalmann S, Behrens M, Meyerhof W. Major haplotypes of the human bitter taste receptor TAS2R41 encode functional receptors for chloramphenicol. *Biochem. Biophys. Res. Commun*. 2013: 435: 267-273.

[11] a) Slack JP, Brockhoff A, Batram C, Menzel S, Sonnabend C, Born S, Galindo MM, Koh S, Thalmann S, Ostopovici-Halip L, Simons CT, Ungureanu I, Duineveld K, Bologa CG, Behrens M, Furrer S, Oprea TI, Meyerhof W. Modulation of bitter taste perception by a small molecule hTAS2R antagonist, *Curr. Biol.* 2010:20:1104–1109; b) Roland WS, Gouka RJ, Gruppen H, Driesse M, van Buren L, Smit G, Vincken JP. 6-Methoxyflavanones as bitter taste receptor blockers for hTAS2R39, *PLoS One*, 2014: 9: e94451; c) Brockhoff A, Behrens M, Roudnitzky N, Appendino G, Avonto C, Meyerhof W. (2011) Receptor ago-

nism and antagonism of dietary bitter compounds. J. Neurosci., 2011: 31 14775-14782; d) Greene TA, Alarcon S, Thomas A, Berdougo E, Doranz BJ, Breslin PA, Rucker JB. Probenecid inhibits the human bitter taste receptor TAS2R16 and suppresses bitter perception of salicin, PLoS One, 2011: 6: e20123.

Istituto Lombardo (Rend. Scienze) 151, 49-76 (2017)

UN METODO SPETTROSCOPICO PER LA QUANTIFICAZIONE DI FARMACI IN FLUIDI BIOLOGICI

PAOLO MARIA OSSI (*), *et al.*[#]

Nota presentata dal m.e. Carlo Bottani
(Adunanza del 6 aprile 2017)

SUNTO. – Viene discussa una tecnica spettroscopica innovativa per determinare la concentrazione di farmaci in fluidi biologici. Introduciamo il contesto dei farmaci caratterizzati da un ristretto indice terapeutico in relazione all'epilessia e al morbo di Parkinson. Ricapitoliamo quindi gli elementi essenziali della spettroscopia Raman e dell'intensificazione del segnale per effetto dell'interazione di un analita con una superficie metallica corrugata alla scala nanometrica. L'ottimizzazione dell'intensificazione dello spettro di un dato analita dipende dalla scelta del metallo e dai dettagli nanostrutturali della sua superficie. Consideriamo in dettaglio la sintesi di superfici di metalli nobili condotta mediante ablazione laser pulsata in gas inerte ad alta pressione. Mostriamo quindi il comportamento spettroscopico di queste superfici per determinare la concentrazione di farmaci in diversi fluidi, incluso il sangue umano. Consideriamo in particolare la carbamazepina (un farmaco antiepilettico ampiamente adottato nei Paesi in via di sviluppo) e l'apomorfina (un farmaco usato per il trattamento di pazienti affetti dal morbo di Parkinson).

***

ABSTRACT. – An innovative spectroscopic technique to determine the drug concentration in biological fluids is discussed. We introduce the context of drugs with narrow therapeutic index in relation to epilepsy and Parkinson's disease. We then recapitulate the essen-

(*) Politecnico di Milano, Dipartimento di Energia, Milano, Italy.
E-mail: paolo.ossi@polimi.it

(#) La nota è il risultato di un lavoro di ricerca condotto in collaborazione con: Chiara Zanchi, Matteo Pistaffa, Andrea Lucotti, Matteo Tommasini, Sebastiano Trusso, Ugo De Grazia, Emilio Ciusani, Marina Casazza, Silvana Franceschetti.

tials of Raman and enhanced Raman spectroscopy that makes use of a corrugated metallic surface. Optimizing the intensification of the spectroscopic signatures of a given analyte critically depends on the metal choice and on the fine details of the induced surface nanostructuring. We review the topic with emphasis on noble metal surfaces synthesized by pulsed laser ablation in inert gas at high pressure. The performance of optimized surfaces to determine the drug concentration in different fluids, including human blood, is discussed with reference to carbamazepine, an anti-epileptic drug widely adopted in Developing Countries and to apomorphine, a drug used to treat via subcutaneous injection patients with important manifestations of Parkinson's disease.

Introduzione

Il dosaggio di farmaci (Therapeutic Drug Monitoring, TDM) è una pratica clinica che consiste nella determinazione della concentrazione di un farmaco in un fluido biologico, solitamente il plasma sanguigno. Tale procedura è particolarmente rilevante per farmaci caratterizzati da un ristretto indice terapeutico (Narrow Therapeutic Index, NTI). In questi farmaci è minima la differenza tra la concentrazione alla quale si hanno effetti terapeutici e la concentrazione (di poco maggiore) alla quale cominciano ad essere osservati effetti collaterali negativi per il paziente (*Fig. 1*). È ovvio come, nelle terapie che utilizzano farmaci di questo tipo, sia rilevante per il medico poter disporre del dato di concentrazione plasmatica del farmaco. In questo modo il medico può associare le condizioni cliniche del paziente al dosaggio del farmaco in modo da garantire l'efficacia del trattamento e l'ottimizzazione della dose necessaria, così da evitare di trovarsi in condizioni per le quali si verificano effetti collaterali negativi per il paziente (*Fig. 1*).

L'indice terapeutico di un farmaco è rappresentato dalla distanza tra la dose necessaria per avere l'effetto farmacologico e la dose che provoca un effetto tossico. Tanto più sono distanti queste due dosi, tanto più sicuro è il farmaco, e viceversa.

TDM è utilizzato da tempo nella pratica clinica, per l'insieme ristretto dei farmaci NTI. Tra questi farmaci troviamo farmaci antitumorali[1] e farmaci antiepilettici (AED, Anti Epileptic Drug), per i quali

[1] S. Fornasaro, S. Dalla Marta, M. Rabusin, A. Bonifacio, V. Sergo, *Toward SERS-based point-of-care approaches for therapeutic drug monitoring: the case of methotrexate*, Faraday Discussions 187 (2016), 485.

una piccola variazione della dose può essere tossica, o inefficace nel controllo dell'insorgenza delle crisi epilettiche.[2]

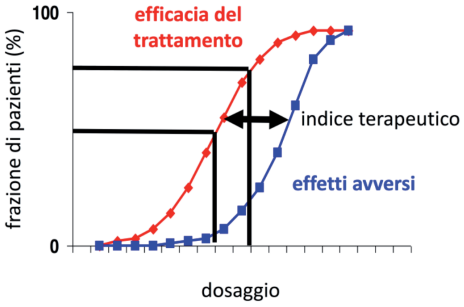


Fig. 1 – Definizione dell'indice terapeutico per un farmaco antiepilettico. Per una determinata frazione percentuale della popolazione di pazienti, l'indice terapeutico definisce l'intervallo di posologia del farmaco per il quale si osserva l'efficacia del trattamento. Tale intervallo è compreso tra la regione a bassa dose, in cui non si hanno effetti clinici, e la regione a dosi maggiori, in cui si osservano effetti clinici avversi e indesiderati. In linea di principio, più è ampia la regione compresa tra le due sigmoidi rossa e blu, più è agevole l'impiego del farmaco, dato che il dosaggio è meno critico nel determinare gli effetti clinici desiderati.

Epilessia

L'epilessia, tra le patologie neurologiche gravi, è la più diffusa. Il 7.6 per 1000 della popolazione del mondo presenta nel corso della sua vita una epilessia.[3,4] Nella maggior parte dei casi l'epilessia può essere

---

[2] B.W. Abou-Khalil, *Making Sense of Lamotrigine Serum Levels*, Epilepsy Currents 5 (2005), 115.

[3] K.M. Fiest, K.M. Sauro, S. Wiebe, S.B. Patten, C.S. Kwon, J. Dykeman, T. Pringsheim, D.L. Lorenzetti, N. Jetté, *Prevalence and incidence of epilepsy: A systematic review and meta-analysis of international studies*, Neurology 88 (2017), 296.

[4] D.J. Thurman, E. Beghi, C.E. Begley, A.T. Berg, J.R. Buchhalter, D. Ding, D.C. Hesdorffer, W.A. Hauser, L. Kazis, R. Kobau, B. Kroner, D. Labiner, K. Liow, G. Logroscino, M.T. Medina, C.R. Newton, K. Parko, A. Paschal, P.M. Preux, J.W. Sander, A. Selassie, W. Theodore, T. Tomson, S. Wiebe, and ILAE Commission on Epidemiology, *Standards for epidemiologic studies and surveillance of epilepsy*, Epilepsia 52 (2011), 2.

trattata con successo, ma il gap terapeutico, cioè la differenza tra i pazienti affetti da epilessia e quelli trattati positivamente, può essere molto ampio, specialmente nei Paesi in via di sviluppo[5] anche perché parecchi farmaci sono inaccessibili, o troppo costosi.[6]

Inoltre, molti pazienti sottoposti a trattamento farmacologico possono soffrire di effetti collaterali che influenzano negativamente la qualità della vita, quali iperattività, stato confusionale, insonnia e depressione.[7] Per ridurre l'insorgenza di effetti collaterali, ogni paziente dovrebbe assumere una dose personalizzata del farmaco, definita a partire dai risultati del dosaggio (TDM) e dall'esame critico dei dati da parte del neurologo.[8]

Nel caso degli AED la TDM è necessaria nelle fasi iniziali del trattamento, o a seguito di cambiamento del, o di uno dei farmaci, perché le crisi epilettiche si verificano a intervalli irregolari, per cui è difficile trovare la dose ottimale sulla sola base delle osservazioni cliniche. Inoltre, in assenza del dato di concentrazione plasmatica, i sintomi degli effetti tossici possono essere equivoci, specialmente in pazienti che soffrono di altre patologie neurologiche oltre all'epilessia.[9] Un altro dato rilevante che merita di essere evidenziato è che l'indice terapeutico, ovvero l'intervallo di dose adatto al trattamento di un paziente, dipende dal paziente stesso, come si osserva anche in *Fig. 1*. Dato un farmaco antiepilettico, non esiste una concentrazione plasmatica ottimale applicabile a ogni paziente. Alcuni pazienti possono rispondere positivamente a dosi che per altri pazienti sono insufficienti a con-

---

5 R. Quintas, A. Raggi, A.M. Giovannetti, M. Pagani, C. Sabariego, A. Cieza, M. Leonardi, *Psychosocial difficulties in people with epilepsy: A systematic review of literature from 2005 until 2010*, Epilepsy & Behavior 25 (2012), 60.

6 A.C. Meyer, T. Dua, J. Ma, S. Saxena, G. Birbeck, *Global disparities in the epilepsy treatment gap: A systematic review*, Bulletin of the World Health Organization 88 (2010), 260.

7 E. Beghi, *Efficacy and tolerability of the new antiepileptic drugs: comparison of two recent guidelines*, The Lancet Neurology 3 (2004), 618.

8 L. Mercolini, R. Mandrioli, M. Amore, M.A. Raggi, *Simultaneous HPLC-F analysis of three recent antiepileptic drugs in human plasma*, Journal of Pharmaceutical and Biomedical Analysis 53 (2010), 62.

9 S.I. Johannessen, D. Battino, D.J. Berry, M. Bialer, G. Krämer, T. Tomson, P.N. Patsalos, *Therapeutic drug monitoring of the newer antiepileptic drugs*, Therapeutic drug monitoring 25 (2003), 347.

trollare l'insorgenza di crisi epilettiche. Questa situazione generale dipende dalla severità delle crisi epilettiche del paziente e dalla sua risposta farmacocinetica.

Le tecniche analitiche standard attualmente impiegate nei laboratori di analisi clinica per il dosaggio dei farmaci sono il dosaggio immunometrico (immunoassay), o varie tecniche cromatografiche come l'HPLC (High Performance Liquid Chromatography, o la cromatografia associata alla spettrometria di massa (HPLC-MS). Si tratta di tecniche specifiche e sensibili, purtroppo associate a costi elevati, tempi di attesa lunghi e necessità di prelievi di volumi sanguigni dell'ordine di alcuni millilitri. In *Tab. 1* sono riportati per alcuni AED la tecnica usata per effettuare il dosaggio sanguigno, il tempo tecnico necessario per effettuarlo, il tempo richiesto per avere il risultato dal momento della consegna del campione al Laboratorio Analisi ed il costo del singolo esame. Osserviamo che la differenza tra tempo tecnico e tempo richiesto dipende dalla necessità di raccogliere (da diversi pazienti) un numero sufficiente di campioni da analizzare in gruppo per mantenere i costi entro i limiti indicati.

Tab. 1. Metodi di analisi, tempi richiesti e costi per diversi farmaci antiepilettici.

Principio attivo	**Metodo di analisi**	**Tempo tecnico (giorni)**	**Tempo richiesto (giorni)**	**Costo unitario (Euro)**
Oxcarbazepina	2; 3	2	3 – 4	29
Perampanel	2; 3	3 – 4	5	29
Lacosamide	1; 2; 3	3 – 4	7	29
Lamotrigina	1; 2; 3	3 – 4	7	29
Acido Valproico	1; 2	1	1	15
Levetiracetam	1; 2; 3	3 – 4	7	29
Fenitoina	1; 2; 3	1	1	15

Metodi di analisi: 1, Immunometria; 2, spettrometria di massa tandem da fase liquida; 3, HPLC-MS.
Fonte: Laboratorio di Patologia Clinica e Genetica Medica, Fondazione IRCCS Istituto Neurologico C. Besta, Milano.

Queste caratteristiche intrinseche limitano la diffusione del dosaggio farmacologico e la popolazione dei pazienti che può beneficiare degli esiti di questi esami clinici.

Malattia di Parkinson

La malattia di Parkinson (MP) è una patologia neurodegenerativa ad evoluzione lenta, ma progressiva. Essa coinvolge principalmente il controllo dei movimenti e dell'equilibrio. Tale malattia è la più frequente tra le patologie denominate "Disordini del Movimento". I sintomi di MP sono probabilmente noti da migliaia di anni (ne sono state trovate tracce in uno scritto indiano (5000 A.C.) e in uno cinese di 2500 anni fa). Il nome della malattia è legato a James Parkinson, un farmacista e chirurgo londinese che descrisse gran parte dei sintomi nel "Trattato sulla paralisi agitante" (XIX secolo). La malattia è diffusa in tutto il mondo ed in tutti i gruppi etnici. L'età media di esordio è intorno ai 58-60 anni, ma circa il 5 % dei pazienti può presentare un esordio giovanile tra i 21 ed i 40 anni. Prima dei 20 anni è estremamente rara. Sopra i 60 anni colpisce l'1-2% della popolazione, mentre la percentuale sale al 3-5% quando l'età è superiore agli 85.[10,11]

Uno dei problemi maggiori nel trattamento dei pazienti Parkinsoniani è lo sviluppo, a lungo termine, di gravi complicazioni motorie che non possono essere trattate adeguatamente con la somministrazione di farmaci per via orale a dosi tollerabili. In questi casi la somministrazione sottocutanea continua di apomorfina (APO) è un trattamento che permette di migliorare le capacità motorie dei pazienti, riducendo i periodi denominati *off*.[11] In questo scenario clinico è fondamentale l'aggiustamento del dosaggio di APO nel paziente, tenendo conto delle variazioni di concentrazione del farmaco nel sangue (farmacocinetica). A tale scopo, nella pratica si considerano solitamente una serie di osservazioni cliniche standardizzate, unitamente, quando sia possibile effettuarli, all'esito di test sulla condizione motoria.[12] In alternativa, analisi chimiche standard (HPLC-MS) su campioni sanguigni

---

10 Linea guida 24, Ministero della Sanità, Agosto 2013.

11 Sito italiano sulla malattia di Parkinson, Fondazione Grigioni per il Morbo di Parkinson, link consultato il 23 maggio 2017; http://www.parkinson.it/morbo-di-parkinson.html

12 A.E. Elia, C. Dollenz, P. Soliveri, A. Albanese, *Motor features and response to oral levodopa in patients with Parkinson's disease under continuous dopaminergic infusion or deep brain stimulation*, Eur. J. Neurol. 19 (2012), 76.

del paziente sono un metodo robusto e affidabile, ma non adottato di routine, per ottenere la concentrazione plasmatica del farmaco. Con tecniche che rendano tali misure più agevoli, ci si aspetta di poter migliorare la condizione di pazienti affetti da MP allo stadio avanzato, grazie alla possibilità di fornire informazioni utili alla determinazione della posologia ottimale e personalizzata del farmaco.

In tempi recenti, nella letteratura scientifica che tratta di dosaggio di farmaci, è emersa una tecnica promettente e complementare a quelle in uso clinico, basata sulla spettroscopia vibrazionale Raman/SERS (Surface Enhanced Raman Spectroscopy)[13-16] le cui caratteristiche salienti sono discusse nel paragrafo seguente. In un possibile scenario applicativo sarebbe possibile risalire alla concentrazione plasmatica di un farmaco in campioni di origine ematica sottoposti a trattamenti minimi e rapidi (es. centrifugazione, estrazione con metanolo). La misura SERS sfrutta l'interazione della molecola di farmaco con una superficie metallica nanostrutturata e l'intensità del segnale, in condizioni controllate, permette di risalire alla quantità di molecole adsorbite sul metallo, che è funzione della concentrazione del farmaco nella soluzione in esame. Ad esempio, sono stati riportati in passato risultati che mostrano una capacità analitica della tecnica SERS comparabile con la tecnica HPLC.[13] Allo stato attuale della ricerca, lo sviluppo di procedure per misure SERS quantitative richiede in ogni caso una messa a punto e validazione che si basa su determinazioni della concentrazione del farmaco condotte con metodi standard di riferimento (HPLC-MS o dosaggio immunologico). Gli sforzi attuali della ricerca nel campo SERS-TDM sono volti a superare le difficoltà

---

[13] C. McLaughlin, D. MacMillan, C. McCardle, W. Ewen Smith, *Quantitative analysis of mitoxantrone by surface-enhanced resonance Raman scattering*, Analytical Chemistry 74 (2002), 3160-3167.

[14] I. J. Hidi, A. Mühlig, M. Jahn, F. Liebold, D. Cialla, K. Weberac, J. Popp, *LOC-SERS: towards point-of-care diagnostic of methotrexate*, Analytical Methods 6 (2014), 3943-3947.

[15] C. Zanchi, A. Lucotti, M. Tommasini, S. Trusso, U. de Grazia, E. Ciusani, P.M. Ossi, *Laser tailored nanoparticle arrays to detect molecules at dilute concentration*, Applied Surface Science 396 (2017) 1866-1874.

[16] A. Jaworska, S. Fornasaro, V. Sergo, A. Bonifacio, *Potential of Surface Enhanced Raman Spectroscopy (SERS) in Therapeutic Drug Monitoring (TDM)*. A Critical Review. Biosensors (Basel) 6 (2016), 47.

legate alla debolezza dei segnali di alcuni farmaci e/o all'influenza negativa del segnale di fondo proveniente da quantità maggiori di biomolecole presenti insieme al farmaco nel fluido da analizzare. Un'altra importante linea di ricerca riguarda lo sviluppo di sensori SERS (ovvero superfici metalliche nanostrutturate) con caratteristiche di alta sensibilità e uniformità spaziale, controllo e riproducibilità del processo di fabbricazione e basso costo. Al momento, prima che si possa introdurre la tecnica SERS nei laboratori di analisi clinica, vi sono ancora alcune difficoltà da superare. Queste sono principalmente di natura tecnologica, per quanto riguarda le caratteristiche del sensore SERS e di natura fisico-chimica per quanto riguarda il trattamento dei campioni da sottoporre all'analisi spettroscopica. Nel seguito si studia come in due casi rappresentativi di un farmaco AED e di un farmaco AP, la tecnica SERS possa costituire un valido sistema per una determinazione veloce e minimamente invasiva della concentrazione plasmatica del farmaco. I risultati riportati in questo contributo sono volti a dimostrare le linee d'azione che sono state seguite da un team di ricercatori comprendenti fisici, chimici, ingegneri, biologi e medici, per avvicinare alla realtà clinica una tecnica spettroscopica molto promettente che finora è stata principalmente utilizzata solamente in laboratori universitari o in altri centri di ricerca scientifica.

Spettroscopia Raman

La spettroscopia Raman deve il suo nome a Sir Chandrasekhara Venkata Raman che pubblicò l'osservazione del fenomeno fisico alla base di questa tecnica sperimentale in uno storico articolo comparso su Nature nel 1928.[17] In questo lavoro Raman e Krishnan riportavano il fatto che, a seguito dell'interazione della luce con la materia, oltre alla grande maggioranza dei fotoni diffusi in modo elastico (con la stessa energia di quelli incidenti) si osserva anche una piccola frazione di fotoni la cui energia (e quindi lunghezza d'onda) è leggermente diversa rispetto a quella della radiazione eccitatrice. Già ai suoi albori, il fenomeno fu interpretato come un processo di scambio di energia tra i fotoni incidenti e gli stati quantici del campione illuminato. In questo modo, per conservare l'ener-

[17] C. V. Raman & K. S. Krishnan, *A New Type of Secondary Radiation*, Nature 121 (1928), 501.

gia totale prima e dopo l'interazione luce-materia, il fotone diffuso a seguito dell'interazione deve necessariamente cambiare la sua energia (e lunghezza d'onda) rispetto al fotone incidente. L'analisi teorica del fenomeno Raman fu essenzialmente dovuta, nel corso degli anni, ai contributi di Heisemberg, Kramers, Dirac e Placzek. Successivi contributi significativi alla teoria del fenomeno Raman in condizioni di risonanza della sorgente luminosa con transizioni elettroniche furono dovuti ad Albrecht, Peticolas, Nafie e Stein. Dal punto di vista teorico, nella descrizione dell'intensità Raman, il ruolo centrale è quello del tensore di polarizzabilità elettrica e della sua relazione con i modi normali vibrazionali della molecola, o del solido.[18]

La descrizione del processo Raman può essere schematizzata utilizzando il diagramma a livelli riportato in *Fig. 2*. Nel caso di una transizione Raman Stokes il sistema (molecola, o solido), descritto come un oscillatore armonico semplice quantistico, si trova inizialmente nello stato fondamentale, interagisce anelasticamente con il fotone incidente arrivando così in uno stato finale vibrazionale eccitato. Per la conservazione dell'energia, il fotone diffuso ha energia (hν') inferiore rispetto al fotone incidente (hν). La differenza di energia hν - hν' è pari al quanto di energia vibrazionale ($h\nu_{vib}$) dell'oscillatore che si è popolato ($h\nu - h\nu' = h\nu_{vib}$).

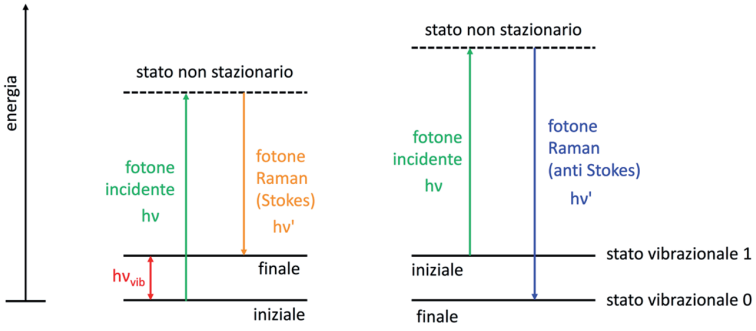


Fig. 2 – Schema dei due possibili processi di scattering Raman: Stokes (a sinistra) e anti-Stokes (a destra). In entrambi i casi la differenza in modulo tra l'energia del fotone incidente e quella del fotone diffuso anelasticamente è pari ad un quanto vibrazionale, $h\nu_{vib}$. Gli stati iniziale e finale dei processi Raman Stokes e anti Stokes risultano scambiati.

18 D.A. Long, *The Raman Effect*, Wiley, 2002.

L'altro possibile processo Raman, definito anti-Stokes, richiede che il sistema si trovi inizialmente in uno stato vibrazionale eccitato e che nello stato finale l'energia vibrazionale del sistema sia minore di quella iniziale. In tal caso l'energia del fotone diffuso dalla molecola, o dal solido è maggiore dell'energia del fotone incidente ed il bilancio energetico risulta $h\nu' - h\nu = h\nu_{vib}$. Dato che in condizioni ordinarie di temperatura la popolazione degli stati vibrazionali eccitati è trascurabile, nelle misure Raman si considerano solitamente le transizioni Stokes piuttosto che quelle anti-Stokes, definitamente meno intense.

Nel caso in cui l'energia del fotone incidente sia tale da poter essere assorbita dal sistema, oltre al normale processo di assorbimento di luce, si può continuare ad osservare il fenomeno Raman, che in queste condizioni prende il nome di Raman risonante e risulta intensificato rispetto alle condizioni in cui la radiazione laser non è assorbibile dal sistema. In certi casi si può arrivare ad una intensificazione del segnale di cinque ordini di grandezza rispetto alle condizioni fuori risonanza, il che corrisponde a significativi vantaggi dal punto di vista applicativo perché permette di misurare la specie risonante anche in presenza di altre specie che, di fatto, restano inosservate dato che il loro segnale è trascurabile rispetto a quello della specie risonante. È bene sottolineare che il processo Raman risonante non va confuso con un processo di fluorescenza, benché la fluorescenza sia un processo alternativo che può essere osservato insieme al Raman risonante e può quindi ostacolarne la misura. La differenza sostanziale tra Raman risonante e fluorescenza è che nel caso del Raman lo stato eccitato non è popolato e il processo di diffusione del fotone (scattering) è veloce (10^{-14} s); nel caso della fluorescenza lo stato eccitato è popolato e successivamente decade, con un tempo di vita media del processo molto più lungo che nel caso Raman ($10^{-8} – 10^{-5}$ s).

La tecnica Raman utilizza quindi la diffusione anelastica di luce monocromatica (laser) per la determinazione di transizioni vibrazionali di molecole, o solidi. In questo modo, dato il legame stretto che esiste tra gli stati vibrazionali di una molecola e la sua composizione e struttura, è possibile associare la presenza e l'intensità di specifici segnali Raman alla presenza e alla concentrazione di una data molecola nel campione analizzato. Nel caso di campioni solidi puri o quasi, la tecnica consente di riconoscere immediatamente la presenza della molecola di interesse (il principio attivo) nonostante la presenza di

altre sostanze come gli eccipienti (si veda ad esempio il caso della molecola di Perampanel, il principio attivo del farmaco Fycompa – *Fig. 3*).[19]

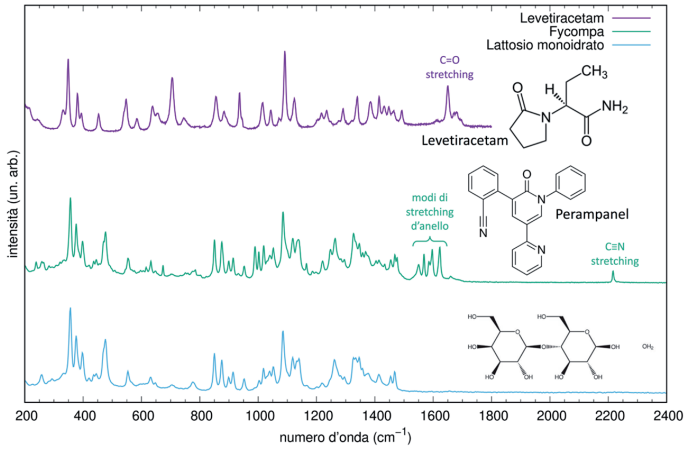


Fig. 3 – Esempi di spettri Raman di composti farmaceutici antiepilettici: (a) Levetiracetam (principio attivo del farmaco commercializzato come Keppra); (b) pastiglia del farmaco Fycompa, in cui oltre ai segnali del principio attivo Perampanel sono osservabili i segnali dell'eccipiente principale, il lattosio monoidrato (c). Sono evidenziati alcuni segnali Raman direttamente riconducibili a gruppi chimici presenti nella struttura dei due principi attivi.

Poiché le regole di selezione della spettroscopia Raman non sono le stesse che governano la spettroscopia vibrazionale di assorbimento infrarosso (IR), la tecnica Raman risulta complementare alla tecnica IR nel caratterizzare gli stati vibrazionali di molecole e solidi. Infatti, modi di vibrazione che sono deboli, o difficilmente osservabili all'IR possono risultare intensi al Raman, e viceversa.

Spettroscopia Raman intensificata da superficie – *Surface Enhanced Raman Spectroscopy* (SERS)

Il limite intrinseco della spettroscopia Raman è la debolezza del

[19] C. Pedrini, *Development of a spectroscopic approach to antiepileptic drug sensing: a case study on lacosamide and levetiracetam*, Tesi di Laurea Magistrale, Politecnico di Milano, 2017.

segnale, ovvero la scarsità dei fotoni diffusi anelasticamente rispetto alla maggioranza dei fotoni che subiscono diffusione elastica (detta diffusione Rayleigh). Benché in condizioni di risonanza il processo Raman diventi molto più efficiente, non è sempre possibile disporre di sorgenti laser adatte a stabilire condizioni di risonanza con il sistema di interesse. Per tale motivo, fin dalla scoperta dell'effetto SERS nei primi anni '70 del 900, molti sforzi sono stati rivolti all'utilizzo dell'intensificazione prodotta da superfici corrugate di alcuni metalli per estendere significativamente il campo di applicabilità di questa tecnica spettroscopica. In particolare, lo studio di analiti presenti in soluzioni a basse concentrazioni è uno scenario particolarmente adatto alle caratteristiche della tecnica SERS. Il meccanismo di intensificazione del segnale richiede la vicinanza della molecola da analizzare ad una superficie metallica che presenti un plasmone di superficie (SP) eccitabile dall'energia dei fotoni incidenti utilizzati nell'esperimento. Un plasmone è un'eccitazione collettiva quantizzata della densità del gas di elettroni di conduzione in un solido. Un plasmone può essere eccitato colpendo con una particella carica, o con un fotone il solido. Il campo Coulombiano della prima e quello elettromagnetico (EM) del secondo provocano una ridistribuzione della carica del gas elettronico che attiva un insieme di oscillazioni di plasma, la cui energia (E_p) dipende dalla densità di numero (n) del gas elettronico

$$E_p = \hbar\omega_p = \hbar\,(ne^2/\varepsilon_0 m)^{1/2} \tag{1}$$

In un metallo, per cui n per la banda di conduzione è tipicamente compresa nell'intervallo $10^{27} < n < 10^{29}$ m^{-3}, E_p è compresa fra circa 2 eV e circa 12 eV. Se un gas elettronico semi-infinito viene confinato da una superficie piana, risolvendo l'equazione di Laplace per le fluttuazioni possibili della densità di carica si prevedono oscillazioni di plasma periodiche nel piano della superficie, la cui ampiezza decade in modo esponenziale verso l'interno del gas elettronico, normalmente alla superficie del metallo. L'unità quantizzata di tali oscillazioni è definita plasmone di superficie (s*urface plasmon*, SP). Con le condizioni elettromagnetiche al contorno, per cui la componente tangenziale del campo EM e la componente normale dello spostamento devono essere entrambe costanti all'interfaccia fra la superficie ed il vuoto, applicando la relazione di dispersione per la costante dielettrica del

gas elettronico, si ottiene[20] la relazione fra la frequenza di un plasmone di superficie ω_s e quella di un plasmone di volume ω_p:

$$\omega_s^2 = ½\ \omega_p^2 \tag{2}$$

Le intensità relative delle eccitazioni plasmoniche di volume e di superficie dipendono da angolo di incidenza, energia, e polarizzazione della radiazione EM eccitatrice.

Quando le dimensioni tipiche delle strutture metalliche scendono alla scala nanometrica (5 – 50 nm), a causa di fenomeni quantistici che intervengono alle piccole dimensioni, si osserva uno spostamento delle frequenze plasmoniche a più basse energie. Nanoparticelle (NP) di argento e oro presentano i picchi di assorbimento relativi alla presenza di SP (associati alle oscillazioni di plasma degli elettroni di conduzione alla superficie della particella) nel visibile, rispettivamente attorno a 410 nm e 520 nm. Ciò rende possibile eccitare tali oscillazioni elettroniche utilizzando sorgenti laser operanti nello spettro del visibile. Quando un analita venga adsorbito alla superficie di nanostrutture metalliche, nel caso in cui la frequenza della radiazione EM incidente sia in risonanza con la frequenza del SP, si verifica un drammatico innalzamento dell'efficienza Raman indotto dalle oscillazioni degli elettroni di superficie. Infatti, queste oscillazioni generano una *ridistribuzione* del campo EM sulla superficie della NP amplificandolo di parecchi ordini di grandezza e amplificando allo stesso tempo l'efficienza Raman. Nel caso di singole NP l'innalzamento del segnale Raman è sufficiente alla rivelazione di molecole con sezione d'urto Raman molto bassa, ma l'effetto è ancora più eclatante nel caso in cui la molecola venga adsorbita nello spazio (*gap*) tra due NP fra loro adiacenti: qui la risonanza produce innalzamenti del segnale Raman dell'ordine di fattori $10^{8–10}$, tali da consentire la rivelazione, in casi favorevoli, di singole molecole. Ricordiamo che il fattore di intensificazione *(enhancement)* Raman Q è dato da:

$$Q_{SERS} = \left| \frac{E(\mathrm{r}, \omega_0)}{E(\omega_0)} \right|^4 \tag{3}$$

[20] H.M. Rosenberg, *The Solid State*, Clarendon Press, Oxford, UK, 1974.

dove $E(r, \omega_0)$ è il campo elettromagnetico totale alla posizione della molecola sulla nanostruttura e $E(\omega_0)$ è il campo elettromagnetico associato alla radiazione incidente. Oltre a questo meccanismo elettromagnetico di intensificazione del segnale Raman, si osserva in molti casi un ulteriore e spesso benefico innalzamento del segnale causato dall'interazione chimica dell'analita con la superficie metallica. Tale interazione è usualmente accompagnata da un trasferimento di carica tra metallo e analita e dalla possibilità che l'eccitatrice utilizzata nell'esperimento possa promuovere transizioni elettroniche metallo-analita, per le quali si possono instaurare le grandi polarizzabilità di transizione che sono all'origine dell'intensificazione dell'intensità Raman. I metalli nobili nanostrutturati (specialmente Ag ed Au) sono sistemi molto popolari nelle misure SERS perché presentano SP eccitabili con la radiazione nel visibile di alcuni dei laser più comunemente utilizzati. Essi, rispetto ad altri metalli, mostrano gli effetti di intensificazione maggiore del segnale Raman. In molti casi, per entrambi i metalli, è stato osservato anche l'effetto di intensificazione chimico.

Substrati SERS realizzati con deposizione a laser pulsato

La posizione (frequenza) e larghezza del SP sono le caratteristiche ottiche di specifico interesse di una superficie metallica corrugata quando se ne voglia esplorare il comportamento utilizzandola come substrato con attività SERS. Definita la scelta del metallo, caduta sui metalli nobili argento (Ag) e oro (Au) che fin dalla scoperta dell'effetto SERS sono noti per presentare i più cospicui incrementi di intensità dei segnali Raman, in linea di principio il problema consiste nell'ingegnerizzare la nanostruttura della superficie, controllandone lo sviluppo così da ottenere i massimi incrementi di segnale. La tecnica da noi utilizzata per realizzare una superficie corrugata artificialmente alla scala nanometrica è l'ablazione a laser pulsato (PLA, *pulsed laser ablation*) effettuata in un gas ambiente chimicamente inerte ad alta pressione. Schematicamente, un impulso laser di durata fra 1 e 10 ns viene focalizzato su un'area della superficie del bersaglio e ivi deposita una certa densità di energia. Questa provoca riscaldamento ed evaporazione di atomi e piccoli cluster dalla superficie irraggiata; il fenomeno è descritto dall'equazione di Hertz-Knudsen per l'evaporazione normale. Se la densità di energia depositata dal laser al bersaglio supera la soglia defi-

nita di ablazione, la quantità di materia espulsa dalla superficie cresce tanto da costituire un vapore denso, in cui prevalgono ancora atomi e piccoli *cluster* neutri, inizialmente posizionato sopra l'area irraggiata. Tale vapore è molto anisotropo, dato che esso ha circa la stessa superficie dell'area irraggiata, il cui raggio è molto superiore allo spessore. Con bersagli metallici la formazione del vapore ablato richiede un tempo molto minore (qualche 10^{-2}) della durata dell'impulso laser che lo ha prodotto. Tale nuvola di vapore assorbe fortemente la radiazione laser: il parametro significativo qui è la densità di numero delle particelle ablate n_a, il cui valore critico n_{cr} è

$$n_{cr} = (4\pi\varepsilon_0 m_e e^{-2} c^2) / \lambda^2 \qquad (4)$$

ove m_e ed e sono massa e carica dell'elettrone a riposo, ε_0 la costante dielettrica e c la velocità della luce nel vuoto, λ la lunghezza d'onda della radiazione laser. Se n_a è maggiore di n_{cr} il vapore è opaco rispetto alla radiazione laser e la assorbe, venendone energizzato attraverso i meccanismi di ionizzazione diretta a singolo fotone e di *bremsstrahlung* inversa degli elettroni liberi nel campo degli ioni. Contemporaneamente il bersaglio viene schermato rispetto all'azione dell'impulso laser. Il vapore energizzato viene convertito da neutro a ionizzato, con grado di ionizzazione 1, diventando un plasma da ablazione molto energetico (la temperatura elettronica equivalente T_e è compresa fra 2 eV e 10 eV) in espansione supersonica pilotata da forti gradienti di pressione normali alla superficie del bersaglio. Tale plasma, per analogia suggerita dalla sua forma, osservata in immagini di fotografia ultra-rapida, viene chiamato piuma di ablazione. Se la propagazione della piuma avviene in un gas inerte, al crescere della pressione di quest'ultimo, si sviluppa un'onda d'urto al fronte di propagazione della piuma, evidente da fotografie della piuma stessa, la piuma viene frenata, la sua forma diviene semi-sferica, le dimensioni sono ridotte rispetto a quelle osservate in espansioni in vuoto e si instaura un comportamento di propagazione diffusiva a tempi lunghi (microsecondi) contemporaneamente al moto retrogrado della coda della piuma verso il bersaglio. Questo comportamento, tipico di espansioni in gas massivi a pressioni fra 30 e 70 Pa, comporta anche una significativa inter-penetrazione fra specie del plasma ed atomi del gas ambiente che porta alla separazione della piuma in due gruppi di particelle: un gruppo veloce che propaga attraverso il gas con velocità simile a quella che avrebbe espandendosi

in vuoto praticamente senza collisioni ed un gruppo lento, ritardato a causa delle interazioni collisionali subite con gli atomi del gas.

A partire da una rassegna critica[21] dei modelli che sono stati proposti per descrivere la complessa fenomenologia appena sintetizzata, abbiamo elaborato un modello per l'espansione in un gas ad alta pressione di un plasma da ablazione, nel cui ambito si prevede il numero N di atomi costituenti una NP che si formi nella piuma durante la fase collisionale della sua espansione:

$$N = (\langle n_a \rangle \sigma_{a\text{-}a} \langle v \rangle t_f) \times (\langle n_g \rangle \sigma_{a\text{-}g} \langle v \rangle t_f) \quad (5)$$

n indica la densità di numero, σ la sezione d'urto di collisione elastica, <v> la velocità media delle particelle e t_f rappresenta il tempo di formazione delle NP nella piuma. I pedici a e g si riferiscono alle specie ablate e agli atomi del gas ambiente.

Una ricerca sistematica ci ha indicato i parametri di processo rilevanti per l'ablazione laser e per la successiva deposizione su un supporto rigido inerte di un numero crescente di piume di ablazione fra loro statisticamente uguali che dia luogo ad un substrato costituito da NP isolate, poi progressivamente interagenti tanto da coalescere e formare strutture sempre più estese, fino alla realizzazione di un film semi-continuo. Per definite condizioni sperimentali, quali lunghezza d'onda della radiazione laser, durata temporale ed energia degli impulsi, angolo di incidenza dell'impulso laser sul bersaglio, distanza bersaglio-supporto, natura del gas ambiente, temperatura di processo, i due parametri rilevanti per definire la nano-struttura di superficie del substrato in fase di deposizione sono il numero di impulsi laser N_{LP} e la pressione del gas ambiente p_g. Al crescere di p_g, mantenendo costante N_{LP}, la superficie appare ricoperta da nanoparticelle progressivamente più piccole, con forma sempre più regolare, sferica. Il valore di N_{LP} definisce il grado di ricopertura del substrato. Al crescere di N_{LP}, mantenendo costante p_g, si osserva principalmente un aumento della densità spaziale della popolazione di NP sul supporto; al progredire della coalescenza si ha una progressiva deviazione della loro forma da sferica verso geometrie più irregolari. Quanto alle proprietà ottiche delle nanostrutture ottenute, la

[21] F. Neri, P.M. Ossi, S. Trusso, *Cluster synthesis and assembling in laser generated plasmas*, Rivista del Nuovo Cimento 34 (2011), 103.

frequenza del SP aumenta, per fissato valore di N_{LP}, al crescere di p_g, così come, a fissata p_g, quando diminuisce N_{LP}.

In *Fig. 4* questo andamento è illustrato nel caso di arrangiamenti auto-assemblati di NP di Ag su supporti di (100)Si. Per NP di Au si osservano gli stessi andamenti parametrici appena discussi per NP di Ag. L'osservazione con microscopia elettronica in trasmissione (TEM) di assemblee di NP dei due metalli nobili, depositate in condizioni tali da evitare coalescenza, indica intervalli di dimensioni fra 1 e 4 nm per Ag e fra 2 e 6 nm per Au. I dati sono in accordo con le previsioni modellistiche (eq. 5). Siamo quindi in grado di controllare la formazione di tali NP e la loro deposizione in condizioni di integrità sul supporto ove formano un substrato con definite proprietà ottiche.

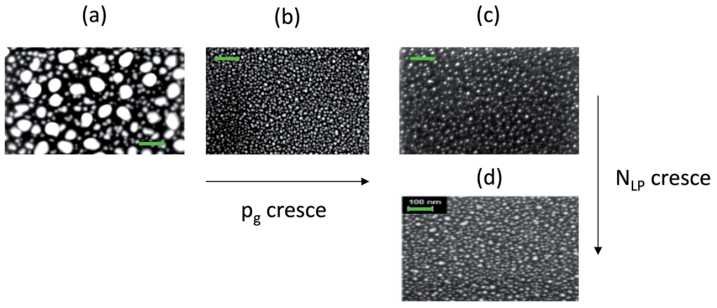


Fig. 4 – Micrografie SEM che mostrano la dipendenza della nanostruttura di superficie di un film di argento ablato in argon dalla pressione del gas ambiente (p_g) e dal numero di impulsi laser (N_{LP}). In (a), (b), (c) N_{LP} = 15000; p_g = 10 Pa (a), 40 Pa (b), 70 Pa (c). In (d) p_g = 70 Pa, N_{LP} = 30000. Le barre verdi corrispondono a 100 nm.

Rimane un parametro dell'ablazione parzialmente nascosto che influisce significativamente sulla nanostruttura di un film: la massa ablata da ogni impulso laser. Questa dipende, a parità di densità di energia depositata dall'impulso, dall'area su cui il fascio laser è focalizzato. Abbiamo studiato specificamente questo problema, considerando sia la propagazione della piuma,[22] sia la nanostruttura di superficie e le proprietà ottiche di un film.[23] Con impulsi che producono masse ablate

[22] M.C. Spadaro, E. Fazio, F. Neri, S. Trusso, P.M. Ossi, *On the role of the ablated mass on the propagation of a laser-generated plasma in an ambient gas*, Europhys. Lett. 109 (2015), 25002.

[23] M.C. Spadaro, E. Fazio, F. Neri, P.M. Ossi, S. Trusso, *On the influence of the*

in rapporto circa 1:2 (7.0 ng, M_p; 16.4 ng, M_g) si ottengono film di Ag morfologicamente simili, con nanostrutture a isole separate fra loro da canali di larghezza ed estensione variabili. Tuttavia, i film hanno proprietà ottiche molto differenti: nel caso della massa ablata M_p il SP presenta un evidente massimo a circa 590 nm, mentre si osserva un largo massimo a 820 nm nel caso della massa ablata M_g. Una verifica delle prestazioni SERS dei due substrati con un analita di riferimento, Rodamina 6G (R6G) in soluzioni acquose a concentrazione molare moderata (10^{-6} M), usando eccitatrice di lunghezza d'onda 785 nm, mostra che il film depositato con M_g non produce segnali apprezzabili dell'analita, a differenza di quello preparato con M_p, con il quale si ottiene uno spettro SERS ben definito.

In conclusione, siamo in grado di sintetizzare mediante ablazione laser in gas inerte ad alta pressione substrati SERS attivi di cui modificare la nanostruttura di superficie e, di conseguenza, le proprietà ottiche, in funzione delle caratteristiche dello spettro Raman dell'analita di interesse.

Applicazione della spettroscopia SERS alla misura della concentrazione di farmaci

Condizioni sperimentali

Gli esempi di seguito riportati riguardano l'applicazione della spettroscopia SERS, che fa uso di substrati nanostrutturati prodotti con la tecnica PLD sopra descritta, alla determinazione della concentrazione di farmaci in soluzione. Il dispositivo sperimentale, illustrato in *Fig. 5*, prevede poche e semplici operazioni che precedono la misura SERS. In particolare, il substrato nanostrutturato, utilizzato come sensore SERS, viene immerso nella soluzione di farmaco ad una data concentrazione (fase *dip*), lasciato in contatto con tale soluzione per un tempo fissato, asciugato (fase *dry*), ed infine posto sotto l'obiettivo dello spettrometro Raman, dove viene acquisito lo spettro SERS.

---

mass ablated by a laser pulse on thin film morphology and optical properties, Applied Physics A 117 (2014), 137-142.

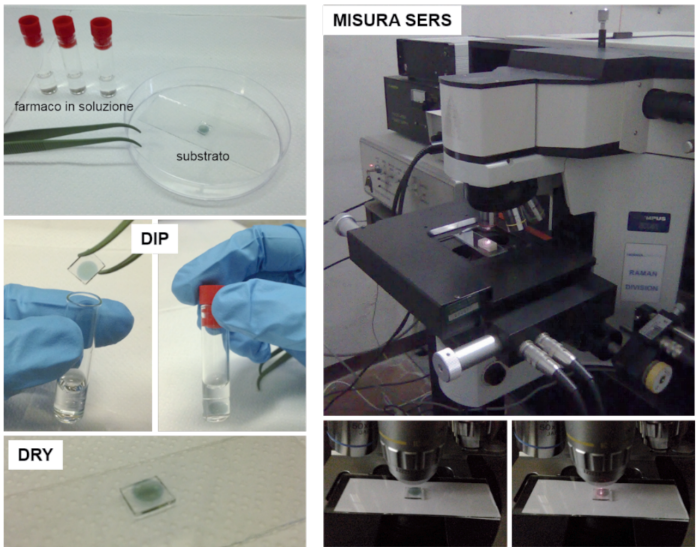


Fig. 5 – Dispositivo sperimentale. I pannelli di sinistra mostrano un substrato di oro nanostrutturato, deposto su vetro, tra quelli utilizzati per le misure SERS dei farmaci qui considerati. Il substrato viene immerso in un volume noto di soluzione di farmaco da analizzare, ad una data concentrazione, lasciato in contatto con la soluzione per un tempo fissato, quindi lasciato asciugare prima della misura SERS. I pannelli di destra mostrano la misura SERS, con uno strumento da laboratorio Jobin-Yvon HR800.

Messa a punto del metodo

Dato un insieme omogeneo di film nanostrutturati ottenuti mediante PLD, la messa a punto del metodo SERS richiede una serie di indagini chimico-fisiche, preliminari all'applicazione quantitativa, che consentono di utilizzare con sicurezza i segnali SERS di riferimento del farmaco in soluzione e di ottimizzare l'intensità di tali segnali. Tale serie di indagini richiede:

1) l'acquisizione di uno spettro Raman di riferimento per il farmaco in un dato solvente, ed il confronto con lo spettro SERS dello stesso farmaco in soluzioni diluite. Ciò permette l'assegnazione dei picchi principali nello spettro SERS, ovvero l'attribuzione a specifici modi normali di vibrazione dei principali segnali del farmaco oggetto di indagine.
2) La selezione, sulla base dello spettro SERS raccolto, di indicatori (*markers*) spettrali robusti, allo scopo di riconoscere e quantificare l'analita (farmaco), anche in matrici via via più complesse, dove la

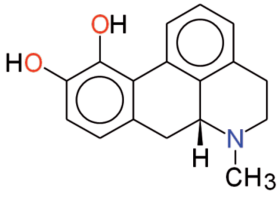
concomitante presenza di altre specie può dar luogo ad una risposta Raman/SERS (caso di biomolecole presenti insieme al farmaco nel fluido da analizzare) non trascurabile.

3) L'ottimizzazione del protocollo *dip & dry*, per la scelta migliore dei tempi di contatto del substrato (sensore) con la soluzione contenente il farmaco ad una data concentrazione.
4) L'ottimizzazione dei parametri di acquisizione dello spettro SERS, quali il tempo di integrazione, la potenza del laser e le dimensioni dell'area irraggiata, per avere spettri con un buon rapporto segnale/rumore, anche a basse concentrazioni del farmaco, pur senza rinunciare ad una misura rapida, prerequisito fondamentale per l'applicazione di interesse.
5) La verifica della riproducibilità e dell'uniformità spaziale del segnale SERS dell'analita. L'area di misura (tipicamente dell'ordine di 1 μm di diametro) è infatti molto più piccola della superficie dei sensori utilizzati (pochi mm di diametro), e la correlazione tra l'intensità del segnale SERS e la concentrazione del farmaco misurato è valutata sulla base di una media spaziale dell'intensità/area di picchi selezionati (a partire da un numero sufficiente di misure, esplorando la superficie del sensore).

Applicazione 1: Apomorfina

La prima applicazione riguarda lo sviluppo di un metodo per la quantificazione di apomorfina (APO) in soluzione acquosa mediante misure SERS, e la successiva prova di applicabilità del metodo alla rilevazione di APO in campioni di origine clinica. Il risultato finale è stato quello di ottenere una risposta SERS di APO ad una concentrazione limite inferiore di 3.3×10^{-7} M in soluzione acquosa, e di 3.3×10^{-6} M in siero.

APO, (6a*R*)-6-metil-5,6,6a,7-tetraidro-4H-dibenzo[de,g]chinolina-10,11-diolo, è un farmaco utilizzato per il trattamento del morbo di Parkinson. La sua struttura chimica è mostrata di seguito:



Nel caso di pazienti con gravi complicazioni (in una fase avanzata della malattia) l'infusione continua sottocutanea rappresenta una valida alternativa alla somministrazione orale. Il dosaggio di APO per via spettroscopica (SERS), da campione di sangue, o plasma, rappresenta una sostanziale innovazione nella pratica clinica, poiché attualmente la ricerca dei livelli del farmaco avviene a partire dall'osservazione clinica del paziente (test di condizione motoria).

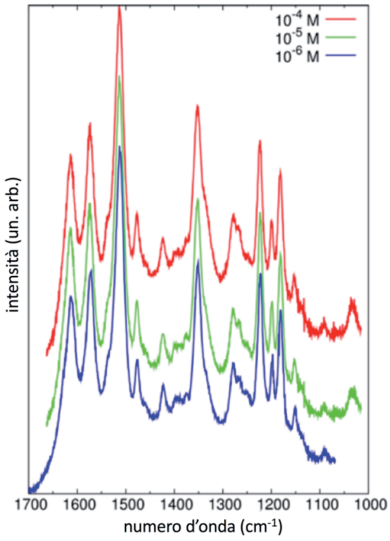


Fig. 6 – Spettri SERS di Apomorfina in soluzione acquosa, ottenuti da substrato di oro nanostrutturato a tre diverse concentrazioni molari (10^{-4}, 10^{-5} e 10^{-6} M). Gli spettri sono stati acquisiti con uno strumento da laboratorio Jobin-Yvon HR800, utilizzando un laser a stato solido a 785 nm, e un obiettivo 50x. Il tempo di integrazione è di 600 s.

La *Fig. 6* mostra una serie di spettri SERS di APO in soluzione acquosa, a tre diverse concentrazioni molari (10^{-4}, 10^{-5} e 10^{-6} M).[24] Il segnale SERS è ottenuto utilizzando substrati nanostrutturati di oro prodotti con $N_{LP} = 3\times10^{4}$ in atmosfera di Ar, alla pressione parziale di 100 Pa. L'esperimento, condotto in condizioni tali da massimizzare il

[24] N. Agarwal, E. Fazio, F. Neri, S. Trusso, C. Castiglioni, A. Lucotti, N. Santo, P.M. Ossi, *Ag and Au nanoparticles for SERS Substrates Produced by Pulsed Laser Ablation*, Crystal Research and Technology 46 (2011), 836-840.

segnale dell'analita (durata della fase *dip* pari a 60 minuti), dimostra l'elevata sensibilità dei substrati considerati nei confronti del farmaco, anche alle basse concentrazioni. Il segnale SERS di APO è riproducibile alle tre diverse concentrazioni, e in particolare è possibile raggiungere la concentrazione di interesse (10^{-6} M).

In una fase successiva è stato possibile verificare la capacità del metodo di discriminare, attraverso la valutazione di differenze nelle intensità degli spettri SERS, la concentrazione di farmaco in soluzione. Sono stati utilizzati, con risultati promettenti, substrati SERS attivi di oro prodotti con $N_{LP} = 1\times10^4$ in atmosfera di Ar, alla pressione parziale di 100 Pa, per la rilevazione di APO in soluzione acquosa in un intervallo di concentrazione compreso tra 3.3×10^{-4} e 3.3×10^{-7} M. I substrati sono stati immersi in un volume noto di soluzione per un tempo considerevolmente ridotto rispetto alla fase precedente. In particolare, per la fase *dip* è stata selezionata una durata ottimale di 5 minuti.

I picchi, selezionati tra i più intensi caratteristici dello spettro SERS di APO, utilizzati per correlare l'intensità alla concentrazione, sono il picco a 1353 cm^{-1}, e il picco a 1515 cm^{-1}.

Mentre alla più bassa delle concentrazioni considerate (3.3×10^{-7} M) si osserva solo un debole segnale dell'analita a 1515 cm^{-1}, e tempi di integrazione più alti non permettono di migliorare il rapporto segnale/rumore dello spettro, a partire dalla concentrazione di 3.3×10^{-6} M lo spettro di APO è di buona qualità, e l'intensità dei picchi aumenta all'aumentare della concentrazione del farmaco. L'andamento della risposta SERS del sensore di oro alle diverse concentrazioni di farmaco è illustrata, attraverso il confronto di spettri rappresentativi, in *Fig. 7*.[25] Dall'analisi di tali spettri è possibile ricavare un andamento, caratteristico, della risposta del sensore (intensità SERS) in funzione della concentrazione di APO.

L'ottimizzazione del numero di impulsi nella produzione mediante PLD dei substrati di oro, ha permesso infine di produrre sensori con caratteristiche superiori, in grado di mostrare una risposta SERS intensa per APO in soluzione acquosa alla concentrazione di 3.3×10^{-7} M. In *Fig. 8* è mostrata la morfologia rappresentativa di questa classe di substrati SERS attivi ottenuti mediante PLD.

---

[25] C. Zanchi, A. Lucotti, M. Tommasini, S. Trusso, U. de Grazia, E. Ciusani, P.M. Ossi, *Au nanoparticle-based sensor for apomorphine detection in plasma*, Beilstein Journal of Nanotechnology 6 (2015), 2224–2232.

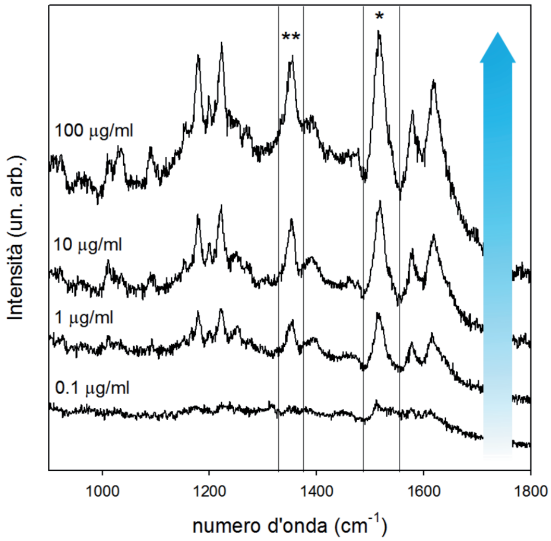


Fig. 7 – Spettri SERS di soluzioni acquose di Apomorfina a 100, 10, 1, 0.1 µg/mL (corrispondenti a concentrazioni da 3.3×10^{-4} M a 3.3×10^{-7} M) ottenuti con un tempo di immersione di 5 minuti. I picchi contrassegnati con gli asterischi a 1515 cm^{-1} e 1353 cm^{-1} sono i marker spettrali selezionati ai fini dello studio della dipendenza del segnale dalla concentrazione di analita.

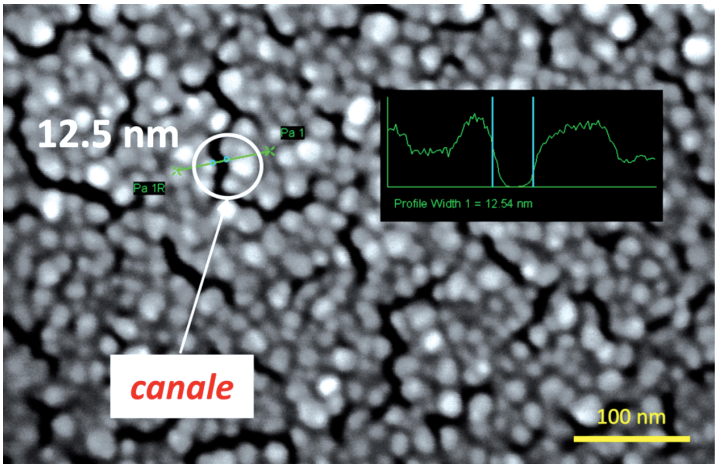


Fig. 8 – Micrografia SEM rappresentativa di un substrato di nanoparticelle di oro ottenuto con 10^4 impulsi laser, in Ar alla pressione parziale di 100 Pa.

La *Fig. 9* mostra il confronto fra gli spettri SERS di APO alla concentrazione di 6.6×10^{-5} M in plasma sanguigno e 3.3×10^{-6} M in siero. Gli spettri sono stati ottenuti dalle due nanostrutture di oro con N_{LP} rispettivamente pari a 1×10^{4} e 2×10^{4}.[26] Nel campione di siero sono riconoscibili tutte le caratteristiche spettrali principali dello spettro SERS di APO. In modo analogo a quanto già osservato in soluzione acquosa, il substrato deposto con un numero maggiore di impulsi laser mostra segnali meglio definiti e più intensi, benché quello preparato con N_{LP} minore sia stato utilizzato per analizzare una concentrazione di APO più alta di oltre un ordine di grandezza. La minore sensibilità del metodo nei confronti di APO in campione biologico, rispetto ai risultati in soluzione acquosa, è data dalla presenza, nel siero, di proteine (in particolare albumina), che legano una frazione importante di APO (fino al 70%), rendendone di fatto la concentrazione libera inferiore di quasi un ordine di grandezza rispetto a quella nominale.

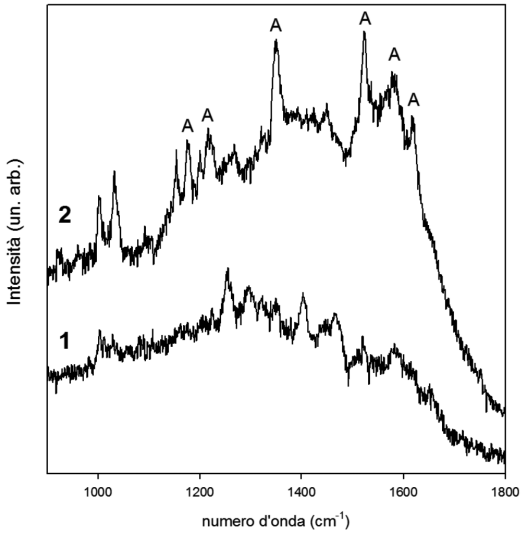
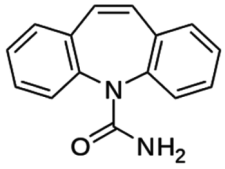


Fig. 9 – Confronto tra lo spettro SERS di apomorfina in plasma (6.6×10^{-5} M; curva 1) acquisito da substrato di oro deposto con numero di impulsi laser pari a 1×10^{4} (pressione parziale di Ar, 100 Pa), e lo spettro SERS di apomorfina in siero ($3.3\ 10^{-6}$ M; curva 2) acquisito da substrato di oro ottimizzato rispetto al numero di impulsi laser (2×10^{4}, pressione parziale di Ar, 100 Pa).

[26] S. Trusso, C. Zanchi, A. Bombelli, A. Lucotti, M. Tommasini, U. de Grazia,

Applicazione 2: Carbamazepina

Il secondo esempio riguarda lo studio della risposta SERS di Carbamazepina (CBZ) adsorbita su substrato di oro nanostrutturato, e la realizzazione di una curva di risposta del sensore (curva di calibrazione), in funzione della concentrazione di CBZ in metanolo. La CBZ, benzo[b][1]benzazepina-11-carbossamide, è un farmaco anti-epilettico largamente impiegato nei Paesi in via di sviluppo. La struttura chimica di CBZ è la seguente:



L'ottimizzazione delle condizioni di deposizione nella produzione mediante PLD di substrati nanostrutturati di oro su supporti di vetro (vedi Applicazione 1) ha permesso di ottenere un buon segnale SERS del farmaco anti-epilettico, e di attribuirne i segnali principali. In particolare, la *Fig. 10* riporta il confronto tra lo spettro SERS di CBZ ed il suo spettro Raman in metanolo (MeOH). I segnali SERS più intensi di CBZ si trovano nella finestra spettrale 650-1700 cm^{-1}. I modi SERS attivi principali, ovvero i picchi a 1619, 1599, 1564, 1309, 1222, 718, 697 cm^{-1} trovano una corrispondenza con lo spettro Raman di CBZ in MeOH, e possono essere tutti assegnati a vibrazioni caratteristiche della molecola di farmaco sulla base di calcoli DFT[15]. Differenze di pochi numeri d'onda dei segnali SERS rispetto a quelli Raman vengono attribuite al differente intorno chimico sperimentato da CBZ su oro rispetto a CBZ in soluzione. Tali piccole differenze confermano anche la debole interazione tra la CBZ e la superficie di oro SERS-attiva.

Proprio l'interazione debole tra il farmaco e la superficie di oro del sensore, ha reso possibile lo sviluppo di un metodo basato sull'utilizzo di un unico sensore, lavabile, per ottenere una curva di calibra-

---

E. Ciusani, L.M. Romito, P.M. Ossi, *Synthesis by pulsed laser ablation of 2D nanostructures for advanced biomedical sensing*, Journal of Instrumentation 11 (2016), C05006.

zione di CBZ in MeOH nella finestra di concentrazioni di interesse clinico (6 - 12 mg/l). Si è dimostrato infatti che, dopo opportuno lavaggio con MeOH, il sensore è riutilizzabile, ed è stata verificata la proporzionalità (linearità) tra la risposta SERS di CBZ e la sua concentrazione in MeOH, nell'intervallo di concentrazioni 6-50 mg/l.

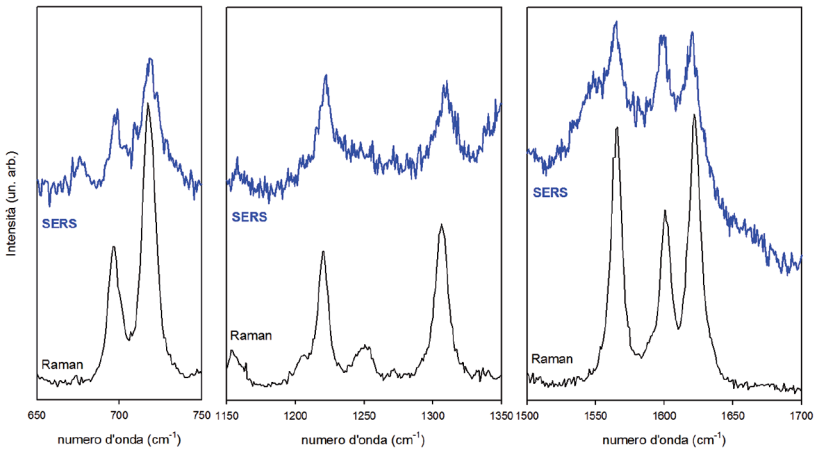


Fig. 10 – Confronto tra lo spettro Raman della carbamazepina in metanolo (20 mg/ml) e lo spettro SERS della carbamazepina (4.2×10^{-4} M) acquisito su substrato di oro. Tempo di contatto pari a 60 minuti, volume di soluzione pari a 1 ml.[15]

In *Fig. 11* sono illustrate in due diverse regioni spettrali, due dei picchi SERS utilizzati come marker di CBZ. L'intensità della risposta SERS in entrambe le regioni dipende dalla concentrazione di CBZ in MeOH. Acquisendo un numero sufficiente di spettri per ciascuna delle concentrazioni studiate è possibile stimare un valore medio di intensità per ciascuno dei segnali SERS selezionati (1621 cm^{-1} e 717 cm^{-1}). L'andamento di tali valori medi in funzione della concentrazione di CBZ è mostrato in *Fig. 12*.[27]

In *Fig. 13* è mostrato lo spettro SERS registrato da siero di paziente trattato con CBZ, dove sono evidenziati i segnali riconducibili alla presenza del farmaco.[27]

---

27 A. Bombelli, *Dosaggio SERS di farmaci con film nanostrutturati di oro*, Tesi di Laurea Magistrale, Politecnico di Milano, 2016.

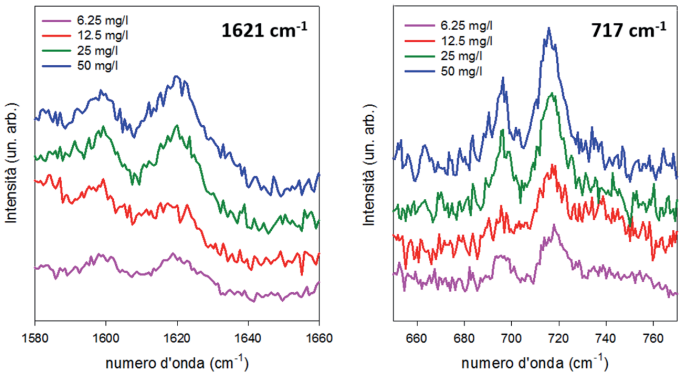


Fig. 11 – Segnali SERS di carbamazepina in due diverse regioni spettrali, in funzione della concentrazione di carbamazepina in MeOH (6.25 – 50 mg/l).

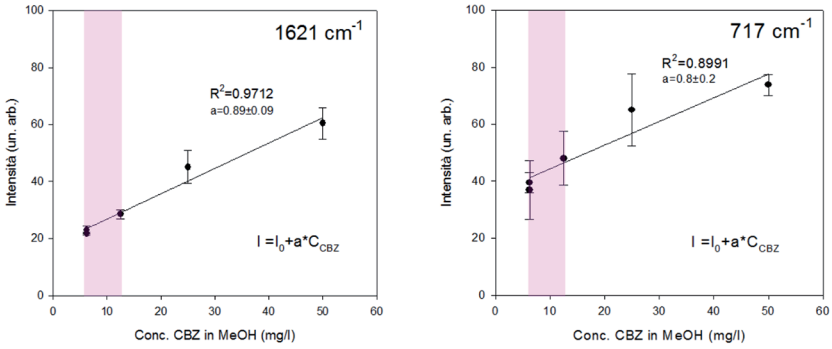


Fig. 12 – Valore medio dell'intensità dei segnali SERS di carbamazepina a 1621 cm^{-1} e 717 cm^{-1}, in funzione della concentrazione, ottenuti su singolo sensore, lavabile.

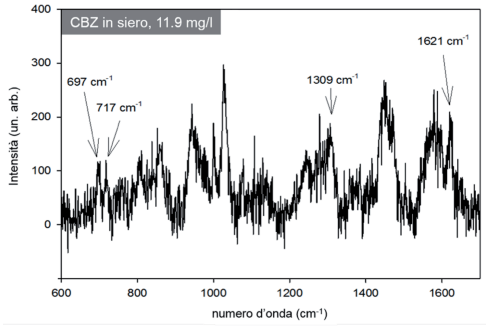


Fig. 13 – Spettro SERS di siero di paziente trattato con carbamazepina. Sono distinguibili segnali riconducibili alla presenza di farmaco (697, 717, 1309 e 1621 cm^{-1}).

Conclusioni

Abbiamo mostrato come sia possibile produrre in modo controllato, tramite tecnica PLD, film sottili di metalli nobili con proprietà plasmoniche *ad hoc* per applicazioni in spettroscopia SERS. Al variare di due parametri principali, la pressione di gas ambiente nella camera di deposizione PLD ed il numero di impulsi laser, si possono ottenere superfici nanostrutturate con caratteristiche variabili con continuità da nanoparticelle isolate, ad isole di nanoparticelle progressivamente coalescenti, fino a strutture percolate e film compatti. Nel caso di strutture consistenti in isole di forma irregolare con dimensioni entro poche decine di nanometri, separate fra loro da reti di canali di larghezza nanometrica, le caratteristiche SERS dei film ottenuti sono promettenti, e consentono la rivelazione quantitativa di farmaci di interesse medico clinico, quali Apomorfina e Carbamazepina. In particolare, la misura SERS di Carbamazepina in campioni di plasma sanguigno mostra la misurabilità di concentrazioni di interesse clinico. Il risultato è un buon punto di partenza per correlare l'intensità del segnale SERS del farmaco alla sua concentrazione nel sangue di pazienti, verso la realizzazione di una procedura SERS-TDM complementare alle tecniche esistenti per la quantificazione di farmaci nel sangue.

Ringraziamenti

Gli autori sono grati ad Alessandro Bombelli e Claudia Pedrini per i loro contributi alla ricerca ed alla Dott. Francesca Del Sorbo (FERB-Fondazione Europea Ricerca Biomedica, Centro di Riabilitazione Specialistica Neuromotoria, Cernusco s/N (MI)) per le discussioni scientifiche sulla malattia di Parkinson. Questo lavoro è stato parzialmente finanziato dal progetto "Controllare l'Epilessia nei Paesi in via di sviluppo" (Polisocial Award 2014 – Politecnico di Milano).

Istituto Lombardo (Rend. Scienze) 151, 77-107 (2017)

QUANTUM FIELD THEORY FROM FIRST PRINCIPLES

PAOLO PERINOTTI (*)

Nota presentata dal m.e. Giacomo Mauro D'Ariano
(Adunanza del 4 maggio 2017)

Sunto. – La descrizione matematica dei sistemi quantistici ne identifica univocamente la natura. In altre parole, trattiamo un sistema come quantistico se, per descriverne il comportamento, adottiamo il formalismo degli spazi di Hilbert e delle relative strutture individuate dai postulati della teoria quantistica. La scelta di utilizzare sistemi quantistici quali sistemi elementari della fisica può essere giustificata in termini di principi informatici, grazie a risultati di ricerche svolte nell'ultimo decennio. Tali risultati concludono un programma di ricerca durato quasi un secolo, volto a trovare una formulazione della teoria quantistica in termini di principi operazionali. Questo risultato pone ora una nuova sfida, che viene qui descritta ed affrontata. Se i sistemi quantistici sono in prima battuta pensati come portatori elementari di informazione, piuttosto che come costituenti elementari della materia, e le loro connessioni sono connessioni logiche all'interno di un algoritmo, piuttosto che relazioni spazio-temporali, occorre trovare una opportuna giustificazione per i concetti meccanici— che caratterizzano la teoria quantistica come teoria dei sistemi fisici. A tal fine, illustreremo come si possa pensare una legge fisica quale algoritmo di aggiornamento del contenuto dei registri di memoria che compongono un sistema. Imponendo quindi le proprietà caratteristiche delle leggi fisiche a tale algoritmo, ovvero omogeneità, reversibilità ed isotropia, mostreremo che le leggi fisiche così selezionate sono particolari algoritmi noti sotto il nome di automi cellulari. Ulteriori assunzioni di massima semplicità dell'algoritmo portano a due soli automi cellulari, che in un opportuno regime possono essere descritti dalle equazioni differenziali di Weyl, alla base della dinamica di campi quantistici relativistici. Discuteremo infine come lo stesso automa cellulare possa dar luogo tanto alla dinamica di particelle Fermioniche, quanto alle leggi di Maxwell, che regolano la dinamica del campo elettromagnetico. Concluderemo con la discussione del principio di relatività, che, opportunamente riformulato, consiste nella definizione del concetto di cambio di sistema di riferimento inerziale, e permette quindi di

(*) Dipartimento di Fisica, Università degli Studi di Pavia, Italy.
E-mail: paolo.perinotti@unipv.it

ritrovare la simmetria dello spazio-tempo Minkowskiano, definita dal gruppo di Poincaré. Lo spazio-tempo emerge quindi in tale contesto, non come sfondo per le leggi fisiche, ma come loro manifestazione, con caratteristiche dettate dalla dinamica dei sistemi, inevitabilmente rivestito delle equazioni che descrivono tale dinamica. In breve, non esiste spazio-tempo senza un'equazione di evoluzione che lo richieda.

* * *

Abstract. – The mathematical description of quantum systems univocally identifies their nature. In other words we treat a system as quantum if we describe its behaviour adopting Hilbert spaces and structures thereof, as prescribed by the postulates of quantum theory. The choice of using quantum systems as the elementary systems of physics can be justified in terms of informational principles, thanks to results of the last decade. Such results come as the conclusion of a research program that lasted almost one century, with the aim of reformulating quantum theory in terms of operational principles. This achievement now poses a new challenge, that we face here. If the systems of quantum theory are thought of as elementary information carriers in the first place, rather than elementary constituents of matter, and their connections are logical connections within a given algorithm, rather than space-time relations, then we need to find the origin of mechanical concepts—that characterise quantum mechanics as a theory of physical systems. To this end, we will illustrate how physical laws can be viewed as algorithms for the update of memory registers that make a physical system. Imposing the characteristic properties of physical laws to such an algorithm, *i.e.* homogeneity, reversibility and isotropy, we will show that the physical laws thus selected are particular algorithms known as cellular automata. Further assumptions regarding maximal simplicity of the algorithm lead to two cellular automata only, that in a suitable regime can be described by Weyl's differential equations, lying at the basis of the dynamics of relativistic quantum fields. We will finally discuss how the same cellular automaton can give rise to both Fermionic field dynamics and to Maxwell's equations, that rule the dynamics of the electromagnetic field. We will conclude reviewing the discussion of the relativity principle, that must be suitably adapted to the scenario where space-time is not an elementary notion, through the definition of a change of inertial reference frame, and whose formulation leads to the recovery of the symmetry of Minkowski space-time, identified with Poincaré's group. Space-time thus emerges as one of the manifestations of physical laws, rather than the background where they occur, and its features are determined by the dynamics of systems, necessarily equipped with differential equations that express it. In brief, there is no space-time unless an evolution rule requires it.

The advent of Quantum Mechanics in the early twentieth century probably represents the most profound revolution in the history of natural sciences, surely comparable to the Copernican revolution or to the introduction of the Galilean method, and in most respects even more shaking than the development of general relativity. The consequences of the quantum world-view are still far from being grasped in their full

extent, even by the community of physicists themselves. This is true if one accepts the Copenhagen interpretation of quantum physics, as well as if one denies it and yearns for a consistent way of reconciling quantum theory with a classical imaginative world.

While there are attempts at formulations of quantum mechanics that can save a notion of predetermined elements of reality independent of observations, we will focus here on the research program that comes from acceptance of a world-view where physical events occur in an inherently probabilistic way, depending on the choices of observers that measure them, and the best physical theory we can wish for must provide rules for calculating correlations between possible events. In this scenario, concepts like mass, charge, or temperature of a system, are high-level structures as compared to registrations of detector clicks or pointer positions, or naked-eye observations. A further thought on the situation then leads to the conclusion that our bottom-level theory is rather a theory of information than a mechanical theory: information about operations performed on systems and about observed events. Mechanics is the high-level picture that we wish to recover at the end of our route towards a reformulation of fundamental physics.

It is then clear why most concepts, techniques and tools that we use in this endeavour are borrowed from that special experience in the history of quantum science that is quantum information theory. Since its very beginning, besides introducing new information-processing concepts and technologies, quantum information theory has represented a new way of looking at foundations of Quantum Theory (QT). The reconsideration of the structure of quantum theory and the exploration of areas that were neglected before, lead to a new axiomatization program, initiated in the early 2000 [1, 2, 3, 4] and lasted for more than one decade [5, 6, 7]. The goal of the endeavour was to reconstruct von Neumann's mathematical formulation of the theory in terms of Hilbert-spaces—neglecting the mechanical postulates such as Schrödinger's equation or its relativistic counterparts—starting from information-processing principles. A complete derivation of QT for finite dimensional systems has been finally achieved in Ref. [8] within the framework of Operational Probabilistic Theories (OPT), starting from six principles assessing the possibility or impossibility to carry out specific information-processing tasks. A book has later been published [9] that presents the framework of OPTs—theories that share the same structures for composition of systems and processes as QT and classical theory, and include suitable rules for calcu-

lating probabilities of composite processes—along with the derivation of QT.

Coming back to our metaphor, however, at this stage we are not even half-way through the journey from the low-level theory of physical systems to the high-level language of mechanics. The informational approach is then carried one step further, with the purpose of reconstructing quantum equations of motion, and space-time as their manifestation. In the simplest, non interacting case, we will recover Weyl, Dirac and Maxwell free field theories, along with the fundamental constants they involve, such as $\hbar$ and c.

The key idea from this stage on is to think of physical laws as information processing algorithms, which process the state of an array of quantum memory cells. If one follows the original proposal by Feynman [10], where physical laws are supposed to be amenable to an *exact* simulation by an algorithm that requires a bounded amount of resources per unit volume of space-time, then the distinction between simulation algorithm and actual physical law vanishes[1]. In this case, the information capacity of a finite physical system has to be bounded. From this perspective, then, any model of physical law will correspond to an algorithm where continuous quantum fields will be replaced by countably many finite-dimensional Fermionic[2] quantum systems. Such an algorithm, reversible and abiding by some weak form of homogeneity and locality, then corresponds to a Fermionic Cellular Automaton. Cellular automata, originally introduced in classical computation by von Neumann [11], were studied in the quantum domain since the late eighties of last century [12, 13, 14]. It was only in 2004 that the concept was rigorously formalised [15], and in the following we will use the term Quantum Cellular Automaton (QCA) to refer to this notion. Now the subject has been studied by various authors in further detail [16, 17, 18, 19, 20]. Very recently, the author

[1] Differently from the proposal of Feynman, we consider here physical laws exclusively as scientific tools to analyse physical phenomena, and distinguish them, in principle, from the actual rules governing the occurrence of phenomena. It is clear, however, that the goal of the formulation of physical laws is to predict the behaviour of physical systems.

[2] The motivation for the choice of Fermionic systems rather than qubits would require a very long discussion per se. Here we will only remark that while qubit algorithms are easily simulated by Fermionic algorithms, respecting their topological structure, for the converse simulation one needs to distort the topology of connections between elementary processes. Moreover, the requirement of linearity that we will discuss shortly is not satisfied by non-trivial qubit algorithms. For obvious reasons, Bosonic systems would violate the requirement of a bounded information density.

formalised cellular automata in the wider context of OPTs [21].

It is due to mention that studies similar in their technical development have been carried out before. The simulation of relativistic quantum dynamics via QCAs, in particular, was already discussed in Refs. [22, 23, 24]. However, we remark that the program that we are discussing here adopts a paradigm that is reversed with respect to those mentioned above, and aims at a derivation of quantum field theory (QFT) from an informational standpoint [25, 26]. Other authors, following these results, also addressed the foundations of QFT in the QCA framework [19, 20].

The present status of this research program is very successful, having derived from a few, very simple requirements, the equations of Weyl and Dirac in 1+1 [27, 28] and in 3+1 dimensions [25], along with Maxwell's equations [29] for suitable bilinear functions of the Weyl fields, also abiding by bosonic commutation relations to very good approximation. Weyl's, Dirac's and Maxwell's equations are the dynamical equations of the fundamental relativistic quantum fields. The above results then imply that the free, non-interacting fields describing the most elementary physical systems have been recovered.

In the same project we can list a stream of works that analyse the symmetries of the physical laws represented by the above mentioned QCAs [30], starting from a rigorous application of the relativity principle [31, 32, 33, 34]. The results of these analyses show that the relativistic Minkowski space-time can be recovered, as the natural manifold for the representation of physical laws represented by the above mentioned QCAs. Indeed, while the discrete nature of QCAs manifestly breaks covariance under the usual representations of Lorentz's group, upon identifying the notion of "reference frame" with that of "representation" of the dynamics, one can appeal to the relativity principle to define the "inertial representation" as the one for which the physical law retains the same mathematical form. In such a way the change of inertial reference frame leads to a set of modified Lorentz transformations that recover the usual ones when the observation scale is much larger than the discrete microscopic scale.

The future challenge consists in accounting also for the fundamental interactions between elementary fields. The program presents with technical difficulties, and preliminary studies have been carried out, analysing 1+1-dimensional toy-theories [35, 36], and introducing perturbative techniques that genralise their counterpart for continuous-time, Hamiltonian dynamics [37].

1. Cellular automata

The approach illustrated in the introduction is based on the assumption that physical systems and events are the manifestation of information processing occurring on an array of elementary memory cells. This paradigm is that of a general algorithm running on a computer, whose memory registers are systems of some yet unspecified kind. They might be classical bits—*bits* for short —or larger classical systems, or quantum bits—*qubits*—or Fermionic modes, or even systems of some more exotic kind. The choice of the type of system that we use for our model of physical law can be motivated in various ways, but will be justified, in the end of the day, only *posteriori*, based on its predictions. It is reasonable to suppose, however, that our theoretical model should not be unnecessarily complex, according to a loose actualisation of Occam's razor. Following Feynman [10] and Deutsch [38] we then suppose that the physical laws should be exactly replicated by a computer operating on finite means as long as a finite portion of space-time is to be described. Reasoning in this way, one can imagine that every elementary system should carry only a finite amount of information. While the last assumption is not strictly necessary, its violation seems to be at odds with the existence of fundamental physical constants such as c, $\hbar$ and G, that determine spatial scales separating regimes where physical systems behave in a very different way. The choice that we make is then for a kind of system with a finite number of levels.

Our choice for the elementary cells will be be quantum, because present day physics suggests that the behaviour of elementary physical systems is consistent with such a picture. However, we have to choose between the two simplest, finite-dimensional quantum systems: qubits or Fermionic modes. For the moment it is not strictly necessary to exclude one of the two, and we will come back on this question later.

The big question now is: what is the best candidate algorithm for representing physical laws? More precisely, what features must it cast? What are the constraints that we impose on our algorithm? While, also in this case, our assumptions are not strictly necessary, we analyse the features of a physical law, and use them as desiderata for our algorithm. In the first place, physical laws are homogeneous in space-time. The equations that express them hold independently of the space-time point that we are considering. How can we translate this requirement for our algorithm, as we do not allow space-time to enter our basic language? Homogeneity in our computational scenario is the requirement that, roughly speaking,

every memory system is treated in the same way by the algorithm, and the rule for updating the content of a cell in a computational step is then independent of both i) the cell address, and ii) the step counter of the algorithm. We formulate the homogeneity principle as follows

Homogeneous update rule: every two memory cells cannot be distinguished by the way in which the rule updates their state, unless one establishes a reference cell, which can be any.

In order to make the content of the principle clearer, let us think of a memory cell as a point in space (or space-time). The physical laws are such that one cannot set an absolute reference frame, where a special point p is chosen as the origin, based on a special behaviour of physical laws at p. The first part of the principle then extends this property to an abstract memory array evolved by an update rule. On the other hand, if any reference frame is fixed in space (or space-time), then it is possible to label points in space (or space-time) with respect to such a reference frame. Einstein's construction of an inertial frame via clock synchronisation is a constructive example of the desideratum that physical laws allow an observer to distinguish the physical role of points in space-time with respect to a reference point. The possibility of a similar construction is bound to the possibility of operationally discriminate two points, so that their "geometric label" in space-time, or their logical address in a memory array is not merely theoretical. In other words, it is meaningful to name two different cells by different labels, once we arbitrarily set a reference frame. This requirement is captured and generalised by the second part of the principle.

The second property of physical laws that we require for our update rule is reversibility. Despite its intuitive content, there might be different meanings for the reversibility requirement. The first meaning comes from the framework of OPTs where an agent is supposed to be able to perform on a system every transformation that is mathematically conceivable and not inconsistent. Then, a transformation is reversible if there exists another one that inverts its effects, and whose effect is inverted by it. This concept of reversibility, that we name *operational invertibility*, is clearly not applicable to a physical law, since there is no possibility of changing a physical law to its inverse *in practice*. On the opposite side, one has a notion of reversibility related to the possibility to use knowledge of the state of a cell or a group of cells to mathematically reconstruct the state of another group of cells at an earlier evolution step. We call this property *invertibility*. Invertibility might seem a reasonable requirement for a physi-

cal law, but we need to remind that in the theories we are considering it is in principle impossible to acquire the knowledge of the state of any group of cells [39]. The last notion that we discuss, that we call reversibility, lies in between the above two, and is the notion that is commonly adopted in quantum mechanics textbooks. According to this notion, an evolution is reversible if it is equivalent to its inverse modulo operationally reversible transformations acting on local memory cells.

Suppose now that we have an invertible evolution rule $\mathscr{U}$, whose mathematical inverse $\mathscr{U}^{-1}$ is another possible evolution rule. Suppose that the two rules are not reversible in the sense defined above. It is very easy to construct a reversible rule by taking two copies of the memory array, thus building a "two-layer" memory with addresses (a, i), where $a \in \mathbb{N}$ is the address of a cell in the single layer, while $i \in \{0, 1\}$ addresses the layer. The cell a in the new memory array is the composite systems made of the cells $(a, 0)$ and $(a, 1)$. Let then the evolution rule on this new array be $\mathscr{U} \otimes \mathscr{U}^{-1}$. This rule is reversible, since it is equivalent to $\mathscr{U}^{-1} \otimes \mathscr{U}$ modulo swapping the systems $(a, 0)$ and $(a, 1)$ for every a. The latter operation is a local, operationally reversible transformation on the cell a. The formulation of reversibility is very simple.

Reversible update rule: the update rule is reversible, *i.e.* it is equivalent to its inverse via one and the same operationally reversible transformation acting independently on every cell.

Let us then consider, in what follows, a denumerable memory array, with a reversible update rule that is homogeneous. We make a final requirement on our update rule, that is *locality*. If we started considering systems in a geometrically defined space-time, we could define locality as the property of an update rule to change the state of a system in a way that only depends on "close" systems, where closeness is defined by the underlying geometry. However, in our approach we are not introducing physical concepts at the fundamental level, and we are trying to show that the latter can be recovered as emergent features, including the geometry of space-time. Thus, we have to turn the usual approach to locality on its head, and define two memory cells a and b to be *neighbours* if the update rule is such that either a change in a at step t affects b at $t + 1$, or viceversa. Thus, the distance of two cells a, b is defined by the number d of steps that are required for a change in a (or b) at step t to affect b (or a) at $t + d$.

The geometry of our memory array can be made more explicit by constructing a graph (V, E), as follows. The vertex set is $V \equiv A$, the

set of all possible addresses in our array, while the (ordered) edge set E contains all pairs $(a, b) \in A \times A$ such that a change in a at step t can affect b at $t + 1$. For a homogenous update rule, the graph (V, E) is very regular: being all the cells identical, their neighbourhoods must have the same structure: one can then choose a reference cell e, and identify its neighbour cells by the symbols $\{h_i\}_{i\in S_+}$. The same set then identifies the neighbours of every other cell a. One can prove, moreover, that the graph of a homogeneous rule is the Cayley graph of a group [26, 21]. A Cayley graph of a group G, presented by a set of generators $S_+ = \{h_i\}_{i\in S_+}$ along with a sufficient set of relations[3] R, is a graph denoted as $\Gamma(G, S_+)$, whose vertex set is G, and the edge set $E \subseteq G \times G$ is defined as $E = \{(g, g') \mid g \in G,\ g' = gh_i,\ i \in S_+\}$. Notice that the edges of a Cayely graph are colored, the color of an edge (g, gh_i) corresponding to the generator $h_i \in S_+$ (see fig. 1). Moreover, the Cayley graph is identified by the neighbourhood structure and by any sufficient set of closed paths representing the elements of R.

We now want to use the graph representing causal relations of an update rule to define rigorously the notion of locality. In order to make this principle independent of homogeneity, we consider a graph that has no further structure, in particular it needs not be a Cayley graph. The notion of distance that we defined above coincides with the definition of distance on the graph. Let $p \coloneqq (a_1, a_2, \ldots, a_n)$ be a path in A *i.e.* a collection of vertices such that either $(a_i, a_{i+1}) \in E$ or $(a_{i+1}, a_i) \in E$, and $a_1 = a$, $a_n = b$. The set of all these paths is $\gamma(a, b)$. Then the distance of a and b on the graph will be defined as

$$d(a, b) \coloneqq \min_{\gamma(a,b)} l(p). \tag{1}$$

With the above definition of distance, it is tautological that the update rule is local, provided we adopt a naive notion of locality, *i.e.* the evolution is such that changes on a cell affect only neighbouring cells in a single step. However, we want our notion of locality to be such that exceedingly connected cells are forbidden: the neighbourhood of a cell a must be of bounded size. More than that, since the memory array is allowed

[3] A set of relators is a set of group elements of the form $h_{i_1}^{s_1} h_{i_2}^{s_2} \ldots h_{i_k}^{s_k}$, with $s_k \in \{+1, -1\}$, that are set equal to the group unit e. The set R is sufficient if every product of generators that is equal to e belongs to the conjugate closure of R.

to be infinite, we want the bound to be uniform. This avoids the possibility that every single cell has a finite number of neighbours, but overall the neighbourhood size is unbounded. We observe that the last remark is important only because we want to define locality independently of homogeneity. Before concluding the discussion of locality, an important remark is in order. Up to now we only referred to pointwise properties of the update rule, *i.e.* properties ruling the behaviour at a given cell or in a small neighbourhood of cells. This is in line with the usual notion of a physical law, that is typically formulated as some differential equation holding at each and every single point in space-time, and, as such, it can be tested with local experiments. In other words, local experiments are sufficient to gather enough information about physical laws. In our framework, in order for the latter statement to hold true, locality and homogeneity are not sufficient. Indeed, in order to specify an update rule completely, the large scale structure of the graph has to be specified. In other words, knowledge of the neighbourhood of any point and the rule that updates it must be supplemented by the specification of closed paths. The locality requirement then amounts to require that every closed path can be decomposed into "elementary" closed paths that have a uniformly bounded length. For a homogeneous rule, this amounts to say that a sufficient set R of relations is finite, and the length of its elements is bounded. The latter condition grants us the possibility of testing physical laws on regions that have a bounded size.

Local update rule: the size of the neighbourhood of every cell is uniformly bounded, and closed paths on the graph of the update rule are decomposable into elementary closed paths of uniformly bounded length.

Now, a memory array with a homogeneous, reversible and local update rule is a *Cellular Automaton* (CA). For a more formal definition, the reader is referred to Ref. [21].

We now want to stress that the metric space defined by the Cayley graph $\Gamma(G, S)$ with the distance d, besides carrying all the algebraic information about G, carries geometric information. In particular, it identifies an equivalence class of metric spaces that contains all the Cayley graphs of G, as well as other metric spaces including suitable smooth manifolds, where the Cayley graphs of G can be embedded with a uniformly bounded distortion of distances. The geometric information about G captured by such an equivalence class is generally bound to its algebraic properties. Rigorously speaking, the equivalence relation we are referring to is *quasi-isometry*. Let (M_1, d_1) and (M_2, d_2) be two metric spaces. We

say that a map $f : M_1 \to M_2$ is a quasi-isometry if one can find real constants $1 < a < \infty$ and $0 < b < \infty$ such that for every $x, y \in M_1$

$$\frac{1}{a} d_2(f(x), f(y)) - b \leq d_1(x, y) \leq a d_2(f(x), f(y)) + b, \qquad (2)$$

and a real constant $0 < k < \infty$ such that for every $z \in M_2$ there exists $x \in M_1$ so that

$$d_2(z, f(x)) \leq k. \qquad (3)$$

If such a map f exists, we say that M_2 is quasi-isometric to M_1. Quasi-isometry is easily shown to be an equivalence relation. Two quasi-isometric spaces can be very different, *e.g.* one can be discrete, like a graph, and the other can be smooth, like a differentiable manifold. However, they have some common intrinsic geometrical property. Points in two quasi-isometric spaces can be associated in a way that "almost" preserves distances. In this sense, every CA naturally identifies a class of metric spaces, typically including smooth manifolds, through their Cayley graph (see *e.g. Fig. 1*). This class bears the relevant geometric information about the

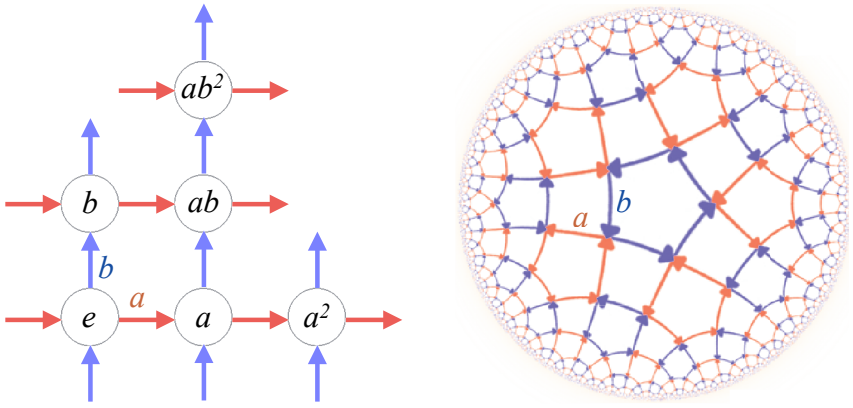


Fig. 1 — Two examples of a Cayley graphs. On the left: a Cayley graph $\mathbb{Z}^2$ with two generators a, b. The only non-trivial relation is the Abelianity condition $aba^{-1}b^{-1}$. On the right: the group is a Fuchsian group presented as $\langle a, b \mid a^5, b^5, abab \rangle$. The graph is quasi-isometrically embedded in the Poincaré disc, which makes it quasi-isometric to the hyperbolic plane $\mathbb{H}_2$.

structure of space (and also space-time) that emerges from the physical

law represented by the automaton. The theory that deals with quasi-isometry classes and the connection they allow one to draw between algebraic and geometric properties of groups is called *Geometric group theory*. Curiously, this theory seems to fill the gap lamented by A. Einstein in the following quote, which is amazing, yet not among the most famous of his:

But you have correctly grasped the drawback that the continuum brings. If the molecular view of matter is the correct (appropriate) one, i.e., *if a part of the universe is to be represented by a finite number of moving points, then the continuum of the present theory contains too great a manifold of possibilities. I also believe that this too great is responsible for the fact that our present means of description miscarry with the quantum theory. The problem seems to me how one can formulate statements about a discontinuum without calling upon a continuum (space-time) as an aid; the latter should be banned from the theory as a supplementary construction not justified by the essence of the problem, which corresponds to nothing "real". But we still lack the mathematical structure unfortunately. How much have I already plagued myself in this way!* [40].

2. Linear Fermionic cellular automata and quantum walks

A Fermionic CA has a memory array that is made of local Fermionic modes, whose transformations, states and effects can be expressed in terms of the Fermionic algebra generated by the local field operators $\psi_s(a)$ with their adjoints $\psi_s^\dagger(a)$, where $a \in A$ and $s \in H_a$ is an index that accounts for the internal structure of the cell a, which is composed by $|S_a|$ local Fermionic modes. The algebra is fully characterised by the Canonical Anticommutation Relations (CAR)

$$\{\psi_s(a), \psi_t^\dagger(b)\} = \delta_{ab}\delta_{st} I, \quad \{\psi_s(a), \psi_t(b)\} = 0, \tag{4}$$

where $\{X, Y\} \coloneqq XY + YX$ denotes the *anticommutator* of X and Y.

For homogeneous rules, the cell structure $H_a = H$ is independent of a. A reversible map $\mathscr{A}$ of a Fermionic memory array is defined as an automorphism of the algebra of operators, and it is then fully specified by the image of field operators $\mathscr{A}[\psi_s(a)]$. In the general case, the latter is a polynomial in the field operators of the neighbourhood of the cell a.

Normally, we would require an update rule that cannot create excitations, that is to say, a number-preserving automorphism:

$$\mathscr{A}(N) = N, \tag{5}$$

where $N \coloneqq \sum_{a,s} \psi_s^\dagger(a)\psi_s(a)$. In the following we will make the further restrictive assumption that the automorphism is *linear*, *i.e.*

$$\mathscr{A}[\psi_s(a)] = \sum_{h_i \in S} W_{ss'}(h_i)\psi_{s'}(ah_i^{-1}), \tag{6}$$

where $S \coloneqq S_+ \cup S_-$, and $S_- \coloneqq S^{-1}$. The matrix $W(h_i)$ with elements $W_{ss'}(h_i)$ is called the *transition matrix* corresponding to the generator $h_i \in S$. Since the automaton is linear, the transition matrix contains all the information about the dynamics of the system. In Fermionic theory, a basis of pure states, identified by the string $\boldsymbol{a_s} = (a_1 s_1, a_2 s_2, \ldots, a_k s_k)$ can be defined by

$$|\boldsymbol{a_s}\rangle \coloneqq \psi_{s_1}^\dagger(a_1)\psi_{s_2}^\dagger(a_2)\ldots\psi_{s_k}^\dagger(a_k)|\Omega\rangle, \tag{7}$$

where $|\Omega\rangle$ is the unique common eigenvector of the (commuting) number operators $\psi_s^\dagger(a)\psi_s(a)$ with null eigenvalues for every a and s. By some straightforward algebra, one can easily realise that the transition matrices $W^*(h_i)$ define the evolution of states of a single excitation

$$|\psi(t)\rangle = \sum_{a\in A, s\in H} \varphi_s(a,t)|as\rangle, \tag{8}$$

namely

$$|\psi(t+1)\rangle = \sum_{a\in A, s,s'\in H, h_i\in S} W^*_{ss'}(h_i)\varphi_s(a,t)|ah_i^{-1}s'\rangle, \tag{9}$$

in other words

$$\varphi_s(a,t+1) = \sum_{s'\in H, h_i\in S} W^*_{ss'}(h_i)\varphi_s(ah_i^{-1},t). \tag{10}$$

The dynamics of single excitations thus determines the full dynamics, and multiple excitations evolve independently. Treating $\varphi_s(a,t)$ as a "single particle" wavefunction in $\mathscr{H} = l^2(A)\otimes\mathbb{C}^{|H|}$, the space of square

summable sequences on $A \times H$, for groups having suitable finite quotients the transition matrices define a unitary operator[4] W on $\mathscr{H}$ (for details see Ref. [41]). One can easily verify [25, 42] that W can be expressed as

$$W = \sum_{h_i \in S} T_{h_i} \otimes W(h_i), \tag{11}$$

where T_g is the *right-regular* representation of the group A, *i.e.* $T_g|f\rangle \coloneqq |fg^{-1}\rangle$, with $\{|f\rangle \mid f \in A\}$ the canonical orthonormal basis in $l^2(A)$ and $g \in A$. A unitary operator of the form 11 on $l^2(A) \otimes \mathbb{C}^{|H|}$ is called *Quantum Walk* (QW), because it represents the coherent quantum version of a classical random walk.

We remark that not only a linear Fermionic CA can be represented by a QW, but also every QW corresponds to a linear Fermionic CA. Indeed, consider a general QW W as in Eq. (11), and define the linear endomorphism of the Fermionic algebra on the corresponding Cayley graph

$$\mathscr{A}[\psi_s(g)] \coloneqq \sum_{h_i \in S} W_{ss'}(h_i)\psi_{s'}(gh_i^{-1}).$$

Thanks to unitarity of W, the endomorphism $\mathscr{A}$ is actually an automorphism, and one can easily show that it corresponds to a Fermionic CA.

3. Isotropy

In this section we discuss a further principle that we impose on our CAs: isotropy. Once again, while the name of the property is borrowed from space-time geometry, in order to give it a meaning in a pre-geometric context — where we only have a memory array organised as a Cayley graph by virtue of causal connections introduced by a dynamical law — we need to rethink the notion of isotropy in depth, and abstract its essence avoiding to appeal to a background geometry. Notice that in the case of isotropy avoiding *every* geometrical notion is impossible, but we will strictly refer to those introduced by homogeneity, *i.e.* the Cayley graph representing causal connections of memory cells and its properties. The notion

[4] In the general case, preservation of the Fermionic algebra easily allows one to prove that $W^\dagger$ is an isometry, however the proof of unitarity can be carried out for CAs on Cayley graphs of groups satisfying the *wrapping lemma* (see Refs. [15, 21]).

of isotropy, in space-time, requires every direction to be equivalent. The translation fo this principle in our context leads to the requirement that, given the Cayley graph $\Gamma(A, S)$, every generator in S_+ is equivalent. This means that, if we permute the labels of elements in S_+ according to some transitive permutation group, we will observe a physical law that is equivalent to the original one. Equivalence of the two physical laws is intended, as in the case of reversibility, through one and the same operationally reversible transformation acting independently on every cell.

Let us now spell out isotropy in rigorous terms. Let L be a group of automorphisms of the Cayley graph $\Gamma(A, S)$ (*i.e.* permutations of the vertices that map edges to edges) that can be expressed as a permutation λ of S. This means that, if $l \in L$, for every $a = h_{i_1}^{s_1} h_{i_2}^{s_2} \dots h_{i_k}^{s_k} \in A$ one has $l(a) = \lambda(h_{i_1}^{s_1})\lambda(h_{i_2}^{s_2}) \dots \lambda(h_{i_k}^{s_k})$. It is easy to verify that a Cayley graph isomorphism that can be expressed as a permutation is also a group automorphism of A. We say that a quantum walk on $\Gamma(A, S)$ is L-isotropic if there exists a projective unitary faithful representation $U : L \to \mathcal{U}(\mathbb{C}^{|H|})$ such that

$$W_{l(h_i)} = U_l W_{h_i} U_l^{-1}, \tag{12}$$

for $l \in L$ and $W_{l(h_i)} \neq W_{h_i}$, with the group L acting transitively on S_+. This notion allows us to formulate a new principle, which we call *isotropy* (see ref. [42]).

Isotropic update rule. A Quantum Walk is isotropic if it is L-isotropic for some group L of automorphisms of the Cayley graph $\Gamma(A, S)$ that can be expressed as a permutation λ of S.

In the remainder, in addition to the homogeneity, reversibility and locality requirements, we will restrict to CAs that are linear and whose QW is isotropic.

4. Quantum Walks on Cayley graphs of $\mathbb{Z}^d$

We will now restrict attention to Cayley graphs of $\mathbb{Z}^d$. Up to now our approach was fully deductive, *i.e.* we analysed the simple consequences of our requirements on the update rule for our memory array. This, in fact, is the spirit of our approach. However, in this section we will reason in an opposite way. We start from simplest non-trivial cell structure ($|H| = 2$) and from the group (*i.e.* $\mathbb{Z}$), and ask ourselves what of its Cayley graphs can support an isotropic linear Fermionic CA. Notice that the

simplest cell structure corresponds to $|H| = 2$, because for Cayley graphs of Abelian groups one can prove [43] that for $|H| = 1$ one has only the trivial QW. Before restricting to the interesting physical case of $d = 3$, we discuss general properties of QWs on Cayley graphs of Abelian groups that allow us to simplify the analysis of the unitarity constraints and to make the problem of diagonalisation much simpler.

Indeed, the right-regular representation of Abelian groups is Abelian, and this implies that the operators T_g can be simultaneously diagonalised in the improper basis $|\boldsymbol{k}\rangle$. The distributions $|\boldsymbol{k}\rangle$ are easily computed after we introduce the basis $\tilde{h}_j$ defined as follows. First, we choose in $\mathbb{Z}^d$ a canonical basis, and express every $h_i \in S$ as a vector $\boldsymbol{h}_i$. We then define the sets $D_n \coloneqq \{h_{i_1}, \dots, h_{i_d}\}$, that are all the possible subsets of S complete and linearly independent. Now, for every D_n we construct the dual set $\tilde{D}_n \coloneqq \{\tilde{\boldsymbol{h}}_1^{(n)}, \dots, \tilde{\boldsymbol{h}}_d^{(n)}\}$, such that for $\boldsymbol{h}_i \in D_n$ one has $\tilde{\boldsymbol{h}}_j^{(n)} \cdot \boldsymbol{h}_{i_k} = \delta_{jk}$. Now, let $\tilde{D} \coloneqq \bigcup_n \tilde{D}_n$, and finally we can express the so-called (first) Brillouin zone as

$$B \coloneqq \bigcap_{\boldsymbol{h} \in \tilde{D}} \{\boldsymbol{k} \in \mathbb{Z}^d \mid -\pi \leq \tilde{\boldsymbol{h}} \cdot \boldsymbol{k} \leq \pi\} \tag{13}$$

Now, the common (improper) eigenvectors of the right regular representation are

$$|\boldsymbol{k}\rangle \coloneqq \frac{1}{|B|} \sum_{\boldsymbol{x} \in A} e^{-i\boldsymbol{k}\cdot\boldsymbol{x}} |\boldsymbol{x}\rangle, \tag{14}$$

where we adopted the vector notation $\boldsymbol{x}$ for $x \in A$, meaning that if $x = \prod_i h_i^{s_i}$, then $\boldsymbol{x} = \sum_i s_i \boldsymbol{h}_i$, and where $|B|$ is the Borel measure of B in $\mathbb{Z}^d$. It is easy to verify that

$$T_{\boldsymbol{x}} |\boldsymbol{k}\rangle = e^{-i\boldsymbol{k}\cdot\boldsymbol{x}} |\boldsymbol{k}\rangle. \tag{15}$$

A QW of the form (11) on a Cayley graph of $\mathbb{Z}^d$ can be decomposed in the representations $\tilde{W}(\boldsymbol{k})$ supported on $|\boldsymbol{k}\rangle \otimes \mathbb{C}^{|H|}$, as follows

$$\tilde{W}(\boldsymbol{k}) = \sum_{\boldsymbol{h}_i \in S} e^{-i\boldsymbol{k}\cdot\boldsymbol{h}_i} W(\boldsymbol{h}_i). \tag{16}$$

One can easily prove that $\tilde{W}(\boldsymbol{k})$ is unitary for every $\boldsymbol{k} \in B$, and this makes it easier diagonalising W, as well as studying its properties.

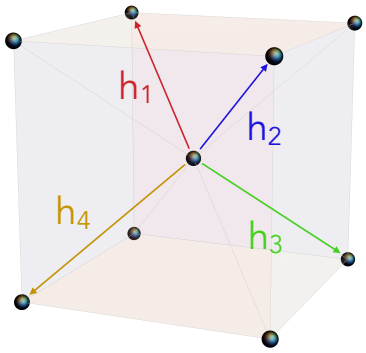


Fig. 2 — The elementary cell of the BCC lattice. The generators $\boldsymbol{h}_1$, $\boldsymbol{h}_2$, $\boldsymbol{h}_3$, $\boldsymbol{h}_4$ *are higlighted.*

5. Solutions of the unitarity conditions in the three dimensional case

The condition of reversibility in the linear Fermionic case requires unitarity of the walk operator W. In terms of the transition matrices one can express the unitarity conditions as follows

$$\sum_{\boldsymbol{h}_i \in S} W^\dagger(\boldsymbol{h}_i) W(\boldsymbol{h}_i) = \sum_{\boldsymbol{h}_i \in S} W(\boldsymbol{h}_i) W^\dagger(\boldsymbol{h}_i) = I_{|H|},$$
$$\sum_{\boldsymbol{h}_i - \boldsymbol{h}_j = \boldsymbol{h}'} W^\dagger(\boldsymbol{h}_i) W(\boldsymbol{h}_j) = \sum_{\boldsymbol{h}_i - \boldsymbol{h}_j = \boldsymbol{h}''} W(\boldsymbol{h}_j) W^\dagger(\boldsymbol{h}_i) = 0$$

This system of second degree equations has no easy solution. Exploiting isotropy, one can make the above system much easier to solve. In the following, we will consider the three-dimensional case $d = 3$. Isotropy, in the first place, selects a unique possible Cayley graph for a QW on $\mathbb{Z}^3$, *i.e.* the one represented in $\mathbb{R}^3$ by the Body-Centered Cubic (BCC) lattice, with four generators $\boldsymbol{h}_1 + \boldsymbol{h}_2 + \boldsymbol{h}_3 + \boldsymbol{h}_4 = 0$ (see *Fig. 2*). Secondarily, one can prove that the isotropic solutions of the above equations only four, modulo local unitary equivalence, *i.e.* modulo changes of basis in the Hilbert space $\mathbb{C}^{|H|}$ of the internal degrees of freedom. The solutions are the following, given in terms of the matrices $\tilde{W}(\boldsymbol{k}) = W^{\pm}_{\boldsymbol{k}}$ and

$\tilde{W}(\boldsymbol{k}) = Z^{\pm}_{\boldsymbol{k}}$:

$$\begin{aligned} &W^{\pm}_{\boldsymbol{k}} = d^{\pm}_{\boldsymbol{k}} I - i\boldsymbol{n}^{\pm}_{\boldsymbol{k}} \cdot \boldsymbol{\sigma}^{\pm}, \qquad Z^{\pm}_{\boldsymbol{k}} = [W^{\pm}_{\boldsymbol{k}}]^T, \\ &\boldsymbol{n}^{\pm}_{\boldsymbol{k}} = \begin{pmatrix} s_x c_y c_z \mp c_x s_y s_z \\ c_x s_y c_z \pm s_x c_y s_z \\ c_x c_y s_z \mp s_x s_y c_z \end{pmatrix}, \quad d^{\pm}_{\boldsymbol{k}} = c_x c_y c_z \pm s_x s_y s_z, \\ &s_i := \sin k_i, \quad c_i := \cos k_i, \quad i = x, y, z, \end{aligned} \tag{17}$$

where X^T denotes the transpose of X, $\boldsymbol{\sigma}^+ = (\sigma_x, \sigma_y, \sigma_z)$ are the Pauli operators, and $\boldsymbol{\sigma}^- = (\sigma_x, -\sigma_y, \sigma_z)$ are their transposes. The above walks are called Weyl quantum walks, because in a suitable regime their action can be approximated by the Weyl equations.

The spectrum of the Weyl walks is now very easily calculated to be

$$\begin{aligned} \mathrm{Spec}(W^{\pm}_{\boldsymbol{k}}) = \mathrm{Spec}(Z^{\pm}_{\boldsymbol{k}}) = \{e^{-i\omega^{\pm}_{\boldsymbol{k}}}, e^{i\omega^{\pm}_{\boldsymbol{k}}}\}, \\ \omega^{\pm}_{\boldsymbol{k}} = \arccos d^{\pm}_{\boldsymbol{k}}. \end{aligned} \tag{18}$$

The Brillouin zone is defined by the following equations

$$B = \bigcap_{\substack{i,j\in\{x,y,z\} \\ s=\pm 1}} \{-\pi \leq k_i + s k_j \leq \pi\} \tag{19}$$

and it is represented in *Fig. 3*. The function $\omega^+ : B \to [0, \pi)$ defined by $\omega^+ : \boldsymbol{k} \mapsto \omega^+_{\boldsymbol{k}}$ is the *dispersion relation* of the QW, and along with the *wave vector* $\boldsymbol{k}$ and the *helicity vector* $\boldsymbol{n}^{\pm}_{\boldsymbol{k}}$ contains all the information about its kinematics. In particular, a wave-packet with wave-vector distribution peaked around $\boldsymbol{k}$ propagates with a group velocity given by $\boldsymbol{v}^{\pm}_{\boldsymbol{k}} = \boldsymbol{\nabla}\omega^{\pm}_{\boldsymbol{k}}$.

The transition matrices for the Weyl walks are

$$A_{\boldsymbol{h}_1} = \begin{pmatrix} \zeta^{\pm} & 0 \\ \zeta^{\pm} & 0 \end{pmatrix}, \qquad A_{-\boldsymbol{h}_1} = \begin{pmatrix} 0 & -\zeta^{\mp} \\ 0 & \zeta^{\mp} \end{pmatrix}, \tag{20}$$

$$A_{\boldsymbol{h}_2} = \begin{pmatrix} 0 & \zeta^{\pm} \\ 0 & \zeta^{\pm} \end{pmatrix}, \qquad A_{-\boldsymbol{h}_2} = \begin{pmatrix} \zeta^{\mp} & 0 \\ -\zeta^{\mp} & 0 \end{pmatrix}, \tag{21}$$

$$A_{\boldsymbol{h}_3} = \begin{pmatrix} 0 & -\zeta^{\pm} \\ 0 & \zeta^{\pm} \end{pmatrix}, \qquad A_{-\boldsymbol{h}_3} = \begin{pmatrix} \zeta^{\mp} & 0 \\ \zeta^{\mp} & 0 \end{pmatrix}, \tag{22}$$

$$A_{\boldsymbol{h}_4} = \begin{pmatrix} \zeta^{\pm} & 0 \\ -\zeta^{\pm} & 0 \end{pmatrix}, \qquad A_{-\boldsymbol{h}_4} = \begin{pmatrix} 0 & \zeta^{\mp} \\ 0 & \zeta^{\mp} \end{pmatrix}, \tag{23}$$

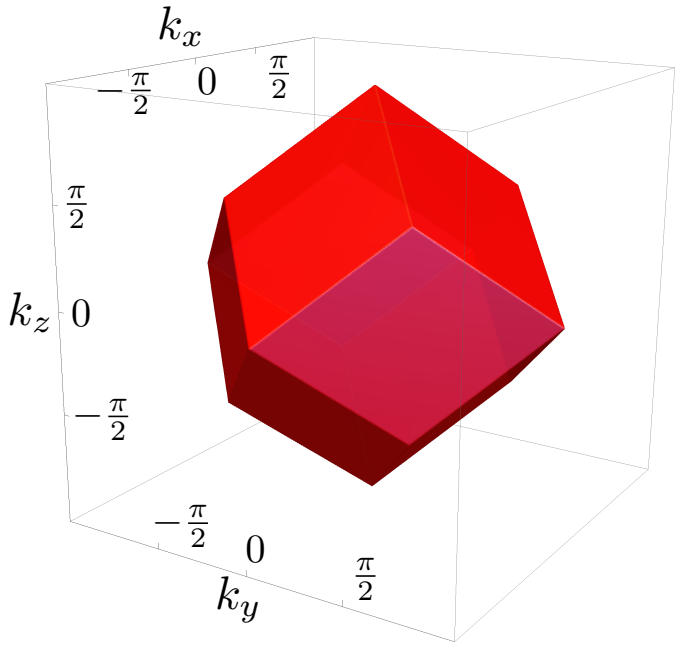


Fig. 3 — The Brillouin zone for the Weyl walks in Eq. (17).

where $\zeta^{\pm} = \frac{1\pm i}{4}$. It is now easy to check that the isotropy group is the Heisenberg group $\mathbb{Z}_2 \times \mathbb{Z}_2$ and its projective unitary representation on $\mathbb{C}^2$ is $\{I, i\sigma_x, i\sigma_y, i\sigma_z\}$.

The corresponding Fermionic CA are called Weyl automata, and we will denote them by the symbol $\mathscr{W}^{\pm}$.

6. Small wave-vector regime

As we mentioned earlier, the name "Weyl" quantum walk, or "Weyl" automaton is justified by the behaviour of the solutions of the walk dynamics in a suitable approximation, that we deem *small wave-vector regime*. This regime is defined by the condition that the state of the memory array is given by a vector $|\psi\rangle \in l^2(A) \otimes \mathbb{C}^2$ that is a superposition of plane

waves $|\boldsymbol{k}\rangle$ with amplitudes narrowly peaked around a value $\boldsymbol{k}_0$, *i.e.*

$$|\psi\rangle = \sum_{j\in\{\pm\}} \int_B d^3\boldsymbol{k}\psi_j(\boldsymbol{k})|\boldsymbol{k}\rangle|u^j_{\boldsymbol{k}}\rangle, \tag{24}$$

where $W_{\boldsymbol{k}}|u^j_{\boldsymbol{k}}\rangle = e^{ij\omega^{\pm}_{\boldsymbol{k}}}|u^j_{\boldsymbol{k}}\rangle$, and

$$|\psi_j(\boldsymbol{k})| < \varepsilon, \quad \forall j,\ \forall \boldsymbol{k}:\ |\boldsymbol{k}-\boldsymbol{k}_0| \geq \delta, \tag{25}$$

for some $|\boldsymbol{k}_0| \ll \pi$. Notice that we did not specify the sign $\pm$ in the expressions $W_{\boldsymbol{k}}$ and $|u^j_{\boldsymbol{k}}\rangle$, for the sake of a lighter notation. What we discuss in the following, indeed, holds independently of the special QW, unless otherwise specified. The symbol $W_{\boldsymbol{k}}$ can thus denote $W^{\pm}_{\boldsymbol{k}}$ and $Z^{\pm}_{\boldsymbol{k}}$. In order to analyse the evolution in such regime, let us introduce the *interpolating Hamiltonian* $H_I(\boldsymbol{k})$, defined by

$$W_{\boldsymbol{k}} = \exp\{-iH_I(\boldsymbol{k})\}. \tag{26}$$

If we embed the discrete lattice $(\boldsymbol{x}, t)$ quasi-isometrically in $\mathbb{R}^4$, and extend the domain of all functions of $\boldsymbol{k}$ from B to the whole $\mathbb{R}^3$, the operator $H_I(\boldsymbol{k})$ defines a Hamiltonian for a continuous-time evolution on $L^2(\mathbb{R}^3)\otimes\mathbb{C}^2$, given by

$$U(t) = \int_{\mathbb{R}^3} d^3\boldsymbol{k}|\boldsymbol{k}\rangle\langle\boldsymbol{k}| \otimes \exp\{-iH_I(\boldsymbol{k})t\}.$$

The term "interpolating Hamiltonian" refers to the fact that, if we apply $U(t)$ to the extension of a wave packet $\psi(\boldsymbol{k}) = (\psi_+(\boldsymbol{k}), \psi_-(\boldsymbol{k}))$, the solution $\psi(\boldsymbol{k}, t) = (\psi_+(\boldsymbol{k}, t), \psi_-(\boldsymbol{k}, t))$ satisfies Schrödinger's equation

$$i\partial_t\psi(\boldsymbol{k}, t) = H_I(\boldsymbol{k})\psi(\boldsymbol{k}, t). \tag{27}$$

Now, let us consider the Weyl automaton $W^{\pm}_{\boldsymbol{k}}$. For a small wave-vector packet, to first order in $|\boldsymbol{k}|$ one has $H_I(\boldsymbol{k}) = \boldsymbol{k}\cdot\boldsymbol{\sigma}$, and the differential equation becomes (modulo a unitary transformation in the case of $W^-_{\boldsymbol{k}}$)

$$i\partial_t\psi(\boldsymbol{k}, t) = \pm\boldsymbol{k}\cdot\boldsymbol{\sigma}\psi(\boldsymbol{k}, t). \tag{28}$$

Now, the inverse Fourier transform of the latter are Weyl's equations

$$i\partial_t\psi(\boldsymbol{x}, t) = \pm\boldsymbol{\sigma}\cdot\boldsymbol{\nabla}\psi(\boldsymbol{x}, t). \tag{29}$$

7. Reversibility and the Dirac automaton

Up to now we did only impose the unitarity condition, which is necessary for our strengthened notion of reversibility, but not sufficient. We observe that actually the Weyl quantum walks are not equivalent to their inverses through a local transformation acting identically and independently on every cell. Indeed, such a transformation can only rotate the vector $\boldsymbol{n}_{\boldsymbol{k}}^{\pm}$ by a fixed rotation, independent of $\boldsymbol{k}$, and this is not sufficient to turn $\boldsymbol{n}_{\boldsymbol{k}}^{\pm}$ into $-\boldsymbol{n}_{\boldsymbol{k}}^{\pm}$ for every $\boldsymbol{k}$. However, the construction that we presented in the discussion about reversibility will produce a reversible automaton that locally behaves as a Weyl automaton: it is sufficient to double every cell, thus having $|H| = 4$, and taking the automaton corresponding to the walk operator

$$A^{\pm} \coloneqq W^{\pm} \otimes (W^{\pm})^{\dagger}. \tag{30}$$

In the small wavevector regime, the above QW behaves as a massless Dirac field. Notice that one can take one further step, and introduce the family of Dirac automata, defined by the walk operators

$$D^{\pm} \coloneqq \begin{pmatrix} nW^{\pm} & imI \\ imI & n(W^{\pm})^{\dagger} \end{pmatrix}, \tag{31}$$

where $n^2 + m^2 = 1$. The parameter m parametrises the elements of this family of QWs, each of which has a small-wavelength behaviour governed by the following differential equation

$$i\partial_t \psi(\boldsymbol{x}, t) = (\pm in\boldsymbol{\alpha} \cdot \boldsymbol{\nabla} - m\beta)\psi(\boldsymbol{x}, t),$$

where one can express $\boldsymbol{\alpha} = \gamma^0\boldsymbol{\gamma}$, and $b = \gamma^0$ in terms of the Dirac matrices γ^μ representing the Clifford algebra for $SU(1,3)$, *i.e.* $\{\gamma^\mu, \gamma^\nu\} = 2\eta^{\mu\nu}$. For $m \ll n$, one can take a further approximation, truncating to first order in m, and rewrite $\boldsymbol{\alpha}$ and β, thus obtaining Dirac's equation

$$(i\gamma^\mu \partial_\mu - mI)\psi(\boldsymbol{x}, t) = 0. \tag{32}$$

We remark that the Dirac QWs correspond to the Dirac CAs which acts on the same Cayley graph as the Weyl's, with every cell doubled in two modes labelled u and d. The Dirac CA can be obtained as

$$\mathscr{D}^{\pm} = [\mathscr{W}^{\pm} \otimes \mathscr{I}]\tilde{\mathscr{S}}_m[(\mathscr{W}^{\pm})^{-1} \otimes \mathscr{I}]\mathscr{S}, \tag{33}$$

where

$$\tilde{\mathscr{S}}_m \coloneqq \prod_{g\in A} \tilde{\mathscr{S}}_m(g),$$

and $\tilde{\mathscr{S}}_m(g) \coloneqq \mathscr{Z}_{mg}\mathscr{S}_g$, $\mathscr{S}_g$ denoting the swap operators acting on the cell g, *i.e.* $\mathscr{S}_g[\psi_s^u(g)] = \psi_s^d(g)$, $\mathscr{S}_g[\psi_s^d(g)] = \psi_s^u(g)$, while

$$\begin{aligned}\mathscr{Z}_{mg}[\psi_s^u(g)] &= m\psi_s^u(g) + in\psi^d(g),\\ \mathscr{Z}_{mg}[\psi_s^d(g)] &= m\psi_s^d(g) + in\psi^u(g).\end{aligned}$$

Let us now define

$$\mathscr{S}_m^{\pm\prime}(g) \coloneqq [\mathscr{W}^{\pm}\otimes\mathscr{I}]\tilde{\mathscr{S}}_m[(\mathscr{W}^{\pm})^{-1}\otimes\mathscr{I}].$$

Since $\mathscr{W}^{\pm}\otimes\mathscr{I}$ is a local automorphism of the Fermionic algebra, one has that i) $\mathscr{S}_m^{\pm\prime}(g)$ acts on the neighbourhood of g of the u modes, and on g of the d modes; ii) since $[\tilde{\mathscr{S}}_m(g), \tilde{\mathscr{S}}_m(g')] = 0$, also $[\mathscr{S}_m^{\pm\prime}(g), \mathscr{S}_m^{\pm\prime}(g')] = 0$. This implies that $\mathscr{D}^{\pm}$ can be decomposed as

$$\mathscr{D}^{\pm} = \prod_{g\in A} \mathscr{S}_m^{\pm\prime}(g) \prod_{g'\in A} \mathscr{S}_{g'}.$$

Upon defining $N_g^{\pm}$ the neighbourhood of g in the Cayley graph of the CA, one can then find a subset $H \subseteq A$ such that $N_{h_1}^{\pm} \cap N_{h_2}^{\pm} = \emptyset$ for $h_1 \neq h_2 \in H$, and such that finitely many—say k—translations Hg_i^{-1} of H cover the whole A. In this way one can easily obtain that

$$\mathscr{D}^{\pm} = \prod_{i=1}^{k} \left(\prod_{h\in Hg_i^{-1}} \mathscr{S}_m^{\pm\prime}(h) \right) \prod_{g\in A} \mathscr{S}_g.$$

This means that the CAs $\mathscr{D}^{\pm}$ admit a decomposition into k layers of local, non overlapping unitary gates acting locally on the neighbourhoods of a cell, *i.e.* they admit a so-called *Margolus decomposition* [15].

8. Theory of light

We showed that our CAs reproduce in the small wavelength regime Weyl's and Dirac's equations. We now discuss how one can reconstruct also Maxwell's equations from the massless Dirac CA, following Ref. [29]. The idea is to study particular entangled states where the correlations are finely tuned. The second-quantisation analysis is easier than the QW one, thus, we start introducing some notation. First, we denote by $\varphi_s(\boldsymbol{x},t)$ the field operators for the "up" component of the $\boldsymbol{x}$ cell at step t, and $\psi_s(\boldsymbol{x},t)$ those for the "down" one. Correspondingly, we will denote by $\varphi_s(\boldsymbol{k},t)$ and $\psi_s(\boldsymbol{k},t)$ the field operators for the normal modes $\boldsymbol{k}$, *i.e.*

$$\varphi_s(\boldsymbol{k},t) = 1/|B|^{1/2} \sum_{\boldsymbol{x}\in A} e^{-i\boldsymbol{k}\cdot\boldsymbol{x}} \varphi_s(\boldsymbol{x},t),$$

and

$$\psi_s(\boldsymbol{k},t) = 1/|B|^{1/2} \sum_{\boldsymbol{x}\in A} e^{-i\boldsymbol{k}\cdot\boldsymbol{x}} \psi_s(\boldsymbol{x},t).$$

Now, we define the following bilinear operators

$$\begin{aligned}
\boldsymbol{F}_T(\boldsymbol{k},t) &\coloneqq \boldsymbol{F}(\boldsymbol{k},t) - \frac{\boldsymbol{n}_{\frac{\boldsymbol{k}}{2}}}{|\boldsymbol{n}_{\frac{\boldsymbol{k}}{2}}|}\left[\frac{\boldsymbol{n}_{\frac{\boldsymbol{k}}{2}}}{|\boldsymbol{n}_{\frac{\boldsymbol{k}}{2}}|}\cdot \boldsymbol{F}(\boldsymbol{k},t)\right]\\
F^i(\boldsymbol{k},t) &\coloneqq \sum_{ss'}\int_{B'} d^3\boldsymbol{q} f_{\boldsymbol{k}}(\boldsymbol{q})\varphi_s(\boldsymbol{k}/2-\boldsymbol{q},t)\sigma^i_{ss'}\psi_{s'}(\boldsymbol{k}/2+\boldsymbol{q},t),\\
\boldsymbol{E}(\boldsymbol{k},t) &\coloneqq |\boldsymbol{n}_{\frac{\boldsymbol{k}}{2}}|\{\boldsymbol{F}_T(\boldsymbol{k},t)+\boldsymbol{F}_T^\dagger(\boldsymbol{k},t)\},\\
\boldsymbol{B}(\boldsymbol{k},y) &\coloneqq i|\boldsymbol{n}_{\frac{\boldsymbol{k}}{2}}|\{\boldsymbol{F}_T(\boldsymbol{k},t)-\boldsymbol{F}_T^\dagger(\boldsymbol{k},t)\},
\end{aligned}$$

where B' is a suitable domain corresponding to the Brillouin zone rescaled by a factor $1/2$. One can then verify (see Ref. [29]) that the operators $\boldsymbol{E}$ and $\boldsymbol{B}$ approximately satisfy Maxwell's equations in the form

$$\partial_t \boldsymbol{E}(\boldsymbol{k},t) = i2\boldsymbol{n}_{\frac{\boldsymbol{k}}{2}} \times \boldsymbol{B}(\boldsymbol{k},t), \tag{34}$$

$$\partial_t \boldsymbol{B}(\boldsymbol{k},t) = -i2\boldsymbol{n}_{\frac{\boldsymbol{k}}{2}} \times \boldsymbol{E}(\boldsymbol{k},t), \tag{35}$$

$$2\boldsymbol{n}_{\frac{\boldsymbol{k}}{2}} \cdot \boldsymbol{E}(\boldsymbol{k},t) = 0, \tag{36}$$

$$2\boldsymbol{n}_{\frac{\boldsymbol{k}}{2}} \cdot \boldsymbol{B}(\boldsymbol{k},t) = 0. \tag{37}$$

Moreover, despite the Fermionic nature of the fields that define the operator $\boldsymbol{F}(\boldsymbol{k},t)$, one can prove that, upon suitable choice of the function $f_{\boldsymbol{k}}(\boldsymbol{q})$, its transverse components

$$\begin{aligned}&\gamma^i(\boldsymbol{k}) \coloneqq \boldsymbol{u}^i_{\boldsymbol{k}} \cdot \boldsymbol{F}(\boldsymbol{k},t_0), \quad i=1,2\\ &\boldsymbol{u}^i_{\boldsymbol{k}} \cdot \boldsymbol{n}_{\frac{\boldsymbol{k}}{2}} = \boldsymbol{u}^1_{\boldsymbol{k}} \cdot \boldsymbol{u}^2_{\boldsymbol{k}} = 0, \quad |\boldsymbol{u}^i_{\boldsymbol{k}}| = 1, \quad (\boldsymbol{u}^1_{\boldsymbol{k}} \times \boldsymbol{u}^2_{\boldsymbol{k}}) \cdot \boldsymbol{n}_{\frac{\boldsymbol{k}}{2}} > 0,\end{aligned}$$

approximately satisfy the canonical commutation relations

$$[\gamma^i(\boldsymbol{k}), \gamma^j(\boldsymbol{k}')] = \delta_{ij}\delta_B(\boldsymbol{k}-\boldsymbol{k}'). \tag{38}$$

9. Special relativity and space-time symmetries

The question that we discuss in the present section regards the emergence of space-time. We already stressed that the quasi-isometry class of the Cayley graph induced by a CA identifies in a family of manifolds, where the evolution rule can be described in terms of interpolating dynamics. The differential equations that we found, such as Weyl's, Dirac's or Maxwell's, have a common feature: covariance under the Poincaré group. The symmetry group of space-time in this case is not previously fixed, but is the result of various approximations. The nature of the emergent space-time is intrinsically connected with the symmetries of physical laws represented by CAs. However, for the time being, we do not have an operational interpretation for the symmetry group. This can be made clearer if we observe that we do not have an acceptable notion of a boost or a rotation connecting observers in two different reference frames. Indeed, the Cayley graph is a graphical representation of causal connections between events, and the CA represents a physical law. What does it mean to "move" with respect to a causal connection or with respect to a physical law? The detailed treatment of this question can be found in Refs. [31, 32, 34]

To make the role of symmetries consistent we start observing that a CA represents a physical law as described by an observer. Now, every "inertial" observer will describe the evolution of physical systems using the same rule. Symmetries are thus required only to map the state of systems as seen by one observer to the state of the same systems as seen by a different observer. Now, the conceivable transformations are once again dictated by our CA. The question is then: how does a CA enforce a group of transformations that can be interpreted as a change of reference frame?

Here comes into play the relativity principle. IN its original formulation, due to Galileo, the principle states that *every inertial observer describes physical laws in the same way*. Thus, the maps representing changes of reference frame are those that leave the CA invariant. How does this simple rule allow us to reconstruct the symmetry group? We proceed in the following way. First of all, our Fermionic CAs are completely described by QWs. Then, the relativity principle in our case will impose that changes of inertial frame must preserve the QW of interest.

Now, a QW is a unitary operator, and as such it is completely specified by its eigenvalues and corresponding eigenvectors. Moreover, an eigenvector of a QW represents a stationary physical condition, and every observer has to agree on stationarity. Thus, eigenspaces of the QW must be mapped to eigenspaces by a change of inertial frame. Similarly, if two observers change the labelling of the Cayley graph in the same way, shifting the "origin" labelled by the neutral element e to the node formerly labelled g, then the change of inertial frame between them must be independent of g. Let us analyse the consequences of this requirement in detail. If the origin is shifted from e to g, then a node formerly labelled f would now be labelled gf. Thus, the left-regular representation changes accordingly: $T_{\boldsymbol{x}} \mapsto T_{\boldsymbol{y}+\boldsymbol{x}} = T_{\boldsymbol{y}} T_{\boldsymbol{x}}$. This implies that a state that is stationary under $T_{\boldsymbol{y}}$ for one of the two observers must be stationary also for the other. This implies that the vectors $|\boldsymbol{k}\rangle$ must be mapped to vectors $|\boldsymbol{k}'\rangle$. In other words, a change of inertial frame must be given by a map $\boldsymbol{k}' : B \to B$ that maps $\boldsymbol{k} \mapsto \boldsymbol{k}'(\boldsymbol{k})$, so that the eigenvalue equation for the QW $W_{\boldsymbol{k}}$ of interest retains the same solutions

$$W_{\boldsymbol{k}} |\psi(\boldsymbol{k})\rangle = e^{-i\omega} |\psi(\boldsymbol{k})\rangle. \tag{39}$$

In order to study the maps $\boldsymbol{k}'$ allowed, we first observe that, using unitarity of $W_{\boldsymbol{k}}$, the eigenvalue equation can be rewritten as

$$(W_{\boldsymbol{k}} \pm W_{\boldsymbol{k}}^\dagger) |\psi(\boldsymbol{k})\rangle = (e^{-i\omega} \pm e^{i\omega}) |\psi(\boldsymbol{k})\rangle, \tag{40}$$

namely

$$\begin{cases} (d_{\boldsymbol{k}} - \cos\omega) |\psi(\boldsymbol{k})\rangle = 0, \\ (\boldsymbol{n}_{\boldsymbol{k}} \cdot \boldsymbol{\sigma} - \sin\omega I) |\psi(\boldsymbol{k})\rangle = 0. \end{cases} \tag{41}$$

Since unitarity imposes that $d_{\boldsymbol{k}} = 1 - |\boldsymbol{n}_{\boldsymbol{k}}|^2$, and the second equation requires that $\sin\omega = |\boldsymbol{n}_{\boldsymbol{k}}|$, the first equation in the system is redundant.

Let us then focus on the second one, and write it as

$$\begin{aligned} &n_\mu(k)\sigma^\mu|\psi(\boldsymbol{k})\rangle = 0, \\ &n_\mu(k) \coloneqq (\sin\omega, \boldsymbol{n_k}), \quad \sigma_\mu = (I, \boldsymbol{\sigma}). \end{aligned} \tag{42}$$

The form of Eq. (42) clearly recalls a Lorentz-invariant expression. However, suppose that a Lorentz transform is applied to the vector $n^\mu(k)$: in general the obtained four- vector is not in the range of the map $n : B \times [0, 2\pi) \to B_1(0) \times [-1, 1] \subseteq \mathbb{R}^4$. This problem can be easily overcome multiplying Eq. (42) by a non-vanishing function $f(k)$ such that the map $k \mapsto f(k)n(k)$ has the same invertibility domains as $k \mapsto n(k)$. Defining $p^\mu(k) \coloneqq f(k)n^\mu(k)$, we then have the following expression for the eigenvalue equation

$$p_\mu(k)\sigma^\mu|\psi(\boldsymbol{k})\rangle = 0 \tag{43}$$

Now, given a map $\boldsymbol{k}' : B \to B$ we can write

$$p_\mu(k')\sigma^\mu = e^{ia(k)}\tilde{\Gamma}_k p_\mu(k)\sigma^\mu\Gamma_k, \tag{44}$$

for suitable invertible matrices $\tilde{\Gamma}_k$ and Γ_k and a C^∞ map $a : B\times[0, 2\pi) \to \mathbb{R}$. Clearly the maps $\boldsymbol{k} \mapsto \boldsymbol{k'}$ and $|\psi(\boldsymbol{k})\rangle \mapsto \Gamma_k^{-1}|\psi(\boldsymbol{k})\rangle$ preserve Eq. (42). However, this scenario would leave room to an exceedingly wide range of possible changes of inertial frame, and thus we impose that the maps $\tilde{\Gamma}_k$ and Γ_k do not actually depend on k, *i.e.* we require

$$p_\mu(k')\sigma^\mu = \tilde{\Gamma} p_\mu(k)\sigma^\mu\Gamma. \tag{45}$$

One can then prove that the allowed maps k' form then a realisation of the semidirect product of the Poincaré group and a group of invertible radial diffeomorphisms of the Brillouin zone.

The existence of suitable functions f that allow one to define $p^f_\mu(k)$ is proved in Ref. [32]. Notice that we made the dependence of the function p^f_μ on f explicit, because in principle there might exist more than one function f (actually, there are continuously many). On the other hand, one can find the detailed derivation of the full group in Ref. [33]. In *Fig. 4* we show some orbits of vectors k under the subgroups of rotations, rotations around a fixed axis, and boosts.

In Ref. [34] the symmetries of the massive (Dirac) QWs in $1 + 1$-d are studied in the same approach, and it is shown that in that case the

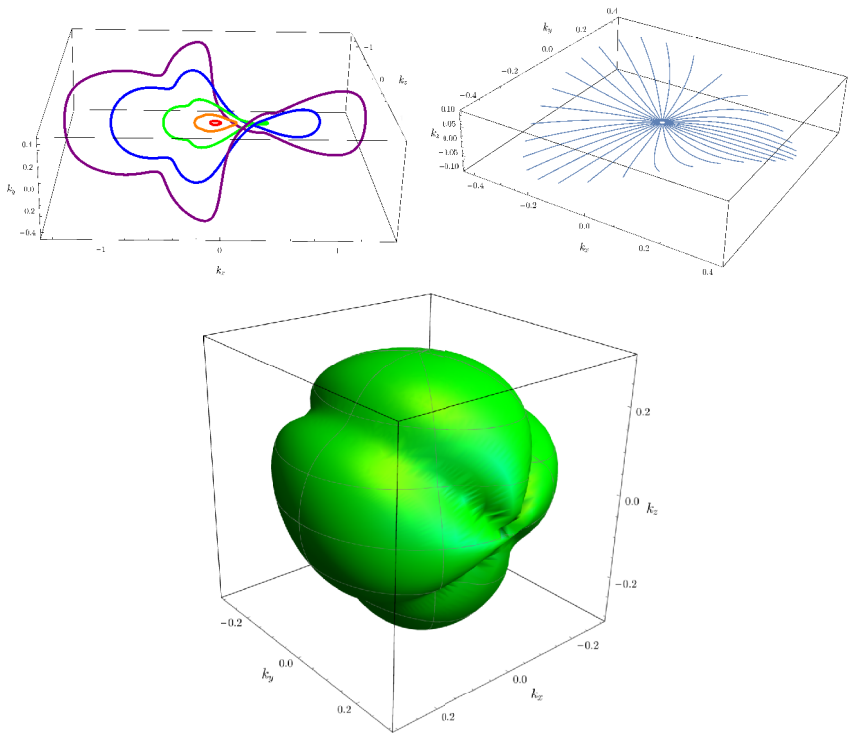


Fig. 4 — Three examples of orbits of a four-vector $k \in B$ under subgroups of the symmetry group of the Weyl CA. In the top left figure we show the orbits of various k's under the group of rotations under a fixed axis. Notice the vertical deformation with respect to the usual orbits under rotation. In the top right figure we show the orbits of small k's under boosts in various directions. In the bottom figure, we show the orbit of a vector k under the full group of rotations.

changes of inertial reference frame form a group that is a semidirect product as above, but involving the linear group $SO(1,2)$, *i.e.* recombining the mass parameter with the components of k.

All the realisations of the aforementioned symmetry groups provide non-linear diffeomorphisms of the Brillouin zone. As a consequence, there is no dual transformation on space-time coordintes, but rather a different transformation for every value of k. This phenomenon is known in the field of doubly special relativity, an approach to quantum gravity based on the relaxation of the usual Lorentz symmetry that allows one a

space-time symmetry group that preserves, besides the speed ofg light c, also a length (or equivalently an energy), that could be the Planck length l_P. Its most counterintuitive consequence of this kind of deformations leads to the phenomenon deemed *relative locality*: the space-time transformations describing the same change of inertial frame are different for systems in different regimes of the energy spectrum. As a result, events that coincide in one reference frame might not coincide in another one. Such a possibility was studied for the Weyl QW in 1+1-d in Ref. [31].

10. Conclusions

The success of the program that we discussed so far is very promising. We recovered the dynamics of free relativistic quantum fields and the Poincaré symmetry of space-time. To this end, a remark is in order: the dimension 3+1 is not derived, but assumed from the beginning. The future developments of the present approach require the construction of interacting theories from suitable principles, and an approximation toolkit in order to perform calculations for the sake of comparison with the standard model, along with prediction of possible deviations. In this respect, a quantum version of the gauge principle seems very appropriate in the context of our approach where time is treated as a discrete variable. In this case, indeed, there is no way to compare the "canonical basis" of the local Hilbert space at any site at step t with that at step $t+1$, and this implies that freedom must be allowed in the definition of such basis at every step, in a way that must not break homogeneity, nor create "particles" out of nothing. This requirement leads in 1+1-d to an essentially unique family of non-linear interactions that were studied in Ref. [35], where the model was analytically diagonalised in the two-fermion sector.

The study of symmetries in the presence of interactions is of crucial importance in view of the challenge of recovering a space-time geometry that depends on the state of the fields, in the perspective of reconciling a dynamic geometry with the quantum nature of elementary systems. The latter problem clearly exhibits very hard challenges, and tackling simplified versions of it, like those reported in Section 9, allows us to fill the gaps in our present understanding.

REFERENCES

[1] L. Hardy. Quantum theory from five reasonable axioms. *Arxiv preprint quant-ph/0101012*, 2001.

[2] Christopher A Fuchs. Quantum mechanics as quantum information (and only a little more). *eprint arXiv*, 2002.

[3] Giacomo Mauro D' Ariano. On the missing axiom of quantum mechanics. *AIP Conference Proceedings*, 810(1):114–130, 2006.

[4] Giacomo Mauro D'Ariano. Probabilistic theories: What is special about quantum mechanics? *Philosophy of Quantum Information and Entanglement*, 85, 2010.

[5] Borivoje Dakic and Caslav Brukner. Quantum theory and beyond: is entanglement special? In H. Halvorson, editor, *Deep Beauty: Understanding the Quantum World through Mathematical Innovation*, pages 365–392. Cambridge University Press, 2011.

[6] Lluís Masanes and Markus P Müller. A derivation of quantum theory from physical requirements. *New Journal of Physics*, 13(6):063001, 2011.

[7] Lucien Hardy. Reformulating and reconstructing quantum theory. *arXiv preprint arXiv:1104.2066*, 2011.

[8] G. Chiribella, G.M. D'Ariano, and P. Perinotti. Informational derivation of quantum theory. *Phys. Rev. A*, 84(012311):012311–012350, 2011.

[9] Giacomo Mauro D'Ariano, Giulio Chiribella, and Paolo Perinotti. *Quantum Theory from First Principles: An Informational Approach*. Cambridge University Press, 2017.

[10] R.P. Feynman. Simulating physics with computers. *International journal of theoretical physics*, 21(6):467–488, 1982.

[11] J. Von Neumann. *Theory of self-reproducing automata*. Urbana, University of Illinois, 1966.

[12] G. Grossing and A. Zeilinger. Quantum cellular automata. *Complex Systems*, 2(2):197–208, 1988.

[13] Yakir Aharonov, Luiz Davidovich, and Nicim Zagury. Quantum random walks. *Physical Review A*, 48:1687–1690, 1993.

[14] A. Ambainis, E. Bach, A. Nayak, A. Vishwanath, and J. Watrous. One-dimensional quantum walks. In *Proceedings of the thirty-third annual ACM symposium on Theory of computing*, pages 37–49. ACM, 2001.

[15] B. Schumacher and R.F. Werner. Reversible quantum cellular automata. *Arxiv preprint quant-ph/0405174*, 2004.

[16] D. Gross, V. Nesme, H. Vogts, and R. F. Werner. Index theory of one dimensional quantum walks and cellular automata. *Communications in Mathematical Physics*, 310(2):419–454, Mar 2012.

[17] P. Arrighi, V. Nesme, and R. Werner. Unitarity plus causality implies localizability. *Journal of Computer and System Sciences*, 77(2):372–378, 2011.

[18] M Freedman and M B Hastings. Classification of Quantum Cellular Automata. February 2019.

[19] Pablo Arrighi. An overview of Quantum Cellular Automata. April 2019.

[20] Terry Farrelly. A review of Quantum Cellular Automata. April 2019.

[21] Paolo Perinotti. Cellular automata in operational probabilistic theories, 2019.

[22] I. Bialynicki-Birula. Weyl, dirac, and maxwell equations on a lattice as unitary cellular automata. *Physical Review D*, 49(12):6920, 1994.

[23] D.A. Meyer. From quantum cellular automata to quantum lattice gases. *Journal of Statistical Physics*, 85(5):551–574, 1996.

[24] Jeffrey Yepez. Relativistic path integral as a lattice-based quantum algorithm. *Quantum Information Processing*, 4(6):471–509, 2006.

[25] Giacomo Mauro D'Ariano and Paolo Perinotti. Derivation of the dirac equation from principles of information processing. *Phys. Rev. A*, 90:062106, Dec 2014.

[26] Giacomo Mauro D'Ariano and Paolo Perinotti. Quantum cellular automata and free quantum field theory. *Frontiers of Physics*, 12(1):120301, 2016.

[27] Alessandro Bisio, Giacomo Mauro D'Ariano, and Alessandro Tosini. Dirac quantum cellular automaton in one dimension: *Zitterbewegung* and scattering from potential. *Phys. Rev. A*, 88:032301, Sep 2013.

[28] Alessandro Bisio, Giacomo Mauro D'Ariano, and Alessandro Tosini. Quantum field as a quantum cellular automaton: The dirac free evolution in one dimension. *Annals of Physics*, 354(0):244 – 264, 2015.

[29] Alessandro Bisio, Giacomo Mauro D'Ariano, and Paolo Perinotti. Quantum cellular automaton theory of light. *Annals of Physics*, 368:177 – 190, 2016.

[30] Giacomo [Mauro D'Ariano] and Alessandro Tosini. Emergence of space–time from topologically homogeneous causal networks. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 44(3):294 – 299, 2013.

[31] Bibeau-Delisle, A., Bisio, A., D'Ariano, G. M., Perinotti, P., and Tosini, A. Doubly special relativity from quantum cellular automata. *EPL*, 109(5):50003, 2015.

[32] Alessandro Bisio, Giacomo Mauro D'Ariano, and Paolo Perinotti. Special relativity in a discrete quantum universe. *Phys. Rev. A*, 94:042120, Oct 2016.

[33] Alessandro Bisio, Giacomo Mauro D'Ariano, and Paolo Perinotti. Quantum walks, weyl equation and the lorentz group. *Foundations of Physics*, 47(8):1065–1076, 2017.

[34] Luca Apadula, Alessandro Bisio, Giacomo Mauro D'Ariano, and Paolo Perinotti. Symmetries of the dirac quantum walk and emergence of the de sitter group, 2018.

[35] Alessandro Bisio, Giacomo Mauro D'Ariano, Paolo Perinotti, and Alessandro Tosini. Thirring quantum cellular automaton. *Phys. Rev. A*, 97:032132, Mar 2018.

[36] Alessandro Bisio, Giacomo D'Ariano, Nicola Mosco, Paolo Perinotti, and Alessandro Tosini. Solutions of a two-particle interacting quantum walk. *Entropy*, 20(6):435, Jun 2018.

[37] Alessandro Bisio, Nicola Mosco, and Paolo Perinotti. Scattering and perturbation theory for discrete-time dynamics, 2019.

[38] David Deutsch. Quantum theory, the church-turing principle and the universal quantum computer. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, 400(1818):97–117, 1985.

[39] G. M. D'Ariano and H. P. Yuen. On the impossibility of measuring the wave function of a single quantum system. 76:2832–2835, 1996.

[40] J. Stachel. *Einstein from 'B' to 'Z'*. Einstein Studies. Birkhäuser Boston, 2001.

[41] Paolo Perinotti and Leopoldo Poggiali. Scalar fermionic cellular automata on finite cayley graphs. *Phys. Rev. A*, 98:052337, Nov 2018.

[42] Giacomo Mauro D'Ariano, Marco Erba, and Paolo Perinotti. Chirality from quantum walks without a quantum coin. *Phys. Rev. A*, 100:012105, Jul 2019.

[43] Alessandro Bisio, Giacomo Mauro D'Ariano, Marco Erba, Paolo Perinotti, and Alessandro Tosini. Quantum walks with a one-dimensional coin. *Phys. Rev. A*, 93:062334, Jun 2016.

Scienze naturali
Zoologia
Istituto Lombardo (Rend. Scienze) 151, 109-120 (2017)

COMPETIZIONE E COOPERAZIONE NELLE STORIE VITALI DEGLI ORGANISMI MARINI

Nota del s.c. GIORGIO BAVESTRELLO (*)

(Adunanza del 25 maggio 2017)

SUNTO. – Le interazioni biotiche dirette e indirette che guidano la struttura della comunità possono essere positive, aumentando la fitness di entrambi i partner o negative, aumentando l'idoneità di uno solo dei partner. L'interazione tra i partner può produrre un'integrazione tra loro, dando così origine a un'associazione definita come simbiosi, con parassitismo e mutualismo ai due estremi di un continuum simbiotico. In passato, le interazioni competitive sono state considerate più diffuse all'interno delle comunità e responsabili della loro struttura. Al contrario, le interazioni cooperative sono state principalmente avvicinate con un approccio descrittivo spesso considerate al livello di curiosità biologiche. Sia il parassitismo che il mutualismo possono evolversi da un'interazione commensale a seconda del contesto ecologico e avere un ruolo diverso nel promuovere l'evoluzione. A sua volta, il commensalismo può evolversi dall'epibiosi, un fenomeno diffuso nell'ambiente bentonico marino in cui l'elevata competizione promuove la sovra-crescita degli organismi. Pertanto l'epibiosi è il punto di partenza per l'evoluzione di diversi tipi di interazioni che portano a un ulteriore aumento della biodiversità.

***

ABSTRACT. – Direct and indirect biotic interactions driving community structure, can be positive, increasing the fitness of both partners, or negative, increasing the fitness of only one of the participant to the interaction. Moreover, both partners may produce an integration among them, thus giving rise to an association defined as symbiosis, with parasitism and mutualism as the two extremes of a symbiotic continuum. In the past, negative interactions have been considered more widespread within communities and responsible for most of their structures. By contrast, cooperative interactions have

(*) Dipartimento di Scienze della Terra, dell'Ambiente e della Vita, Università degli Studi di Genova, Italy. E-mail: giorgio.bavestrello@gmail.com

mainly approached with a descriptive aspect. Both parasitism and mutualism can evolve from a commensal interaction depending from the ecological context and have a different role in promoting evolution. In turn commensalism can evolve from epibiosis, a widespread phenomenon in marine benthic environment where the high spatial competition promotes the overgrowth of organisms. Therefore, epibiosis is the starting point for further evolution of different kind of interactions leading to an increasing of biodiversity.

1) Introduzione: rapporti antagonistici

…un mondo dove la Vita uccide la Vita, mangia la Vita…
Oriana Fallaci, La forza della Ragione

Il 22 Novembre 1859 veniva stampata a Londra, per i tipi dell'editore John Murray, un' opera destinata a segnare una svolta nell'ambito del pensiero biologico. Il volume intitolato "*On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*" era l'opera matura di un ormai celebre naturalista inglese, Charles Robert Darwin, nella quale era descritta, corredata di un'ampia esemplificazione, quella che da lì a pochi anni sarebbe stata chiamata la teoria dell'evoluzione.

Darwin aveva elaborato la sua idea di origine delle specie basandola su due pilastri: la mutazione, apportatrice di variabilità casuale negli organismi, che tante conferme avrebbe avute un secolo dopo, grazie alla descrizione della struttura del DNA, e la selezione naturale. Questo ultimo elemento rappresentava probabilmente la novità più interessante del pensiero darwiniano che, in questo modo, aveva trovato l'agente in grado di indirizzare il corso del cammino evolutivo, favorendo gli individui più adatti ad affrontare le contingenze di un ambiente continuamente mutevole. Che Darwin stesso ponesse particolare attenzione su questo punto è evidenziato dal titolo stesso dell'opera, nel quale figurano esplicitamente la selezione naturale (*natural selection*) e la lotta per la sopravvivenza (*struggle for life*). In un ambiente reso competitivo dalla limitatezza delle risorse (è noto che Darwin fu molto colpito dal celebre "*An Essay on Principle of population"* che Thomas Malthus aveva pubblicato sessanta anni prima), ogni organismo si trova impegnato in una continua gara che premia i più adeguati ad ottenere cibo, a sfuggire ai predatori, a conquistarsi uno spazio, a riprodursi. Da allora, i rapporti antagonistici, sia a livel-

lo inter sia intra specifico, sono stati la chiave di lettura delle interazioni biotiche all'interno degli ecosistemi.

La lotta per la sopravvivenza è un concetto piuttosto generico che racchiude in sé diverse tipologie di fenomeni che sono state chiarite nel corso del secolo successivo tramite lo sviluppo di nuove discipline scientifiche. L'approccio ecologico moderno è nato dalla scoperta che le relazioni che legano gli organismi possono essere interpretate in chiave trofica (categorie trofiche, catene trofiche, reti trofiche…). In fondo, i celebri flussi di materia e di energia nascondono, sotto l'asettica forma del computo ragionieristico, la cruda realtà del "chi mangia chi". I rapporti tra prede e predatori hanno così colpito l'immaginazione del mondo scientifico che i primi approcci all'ecologia matematica sono stati tentati proprio in quest'ambito tramite il celebre modello, indipendentemente messo a punto da Alfred Lotka e Vito Volterra nel periodo tra le due guerre mondiali (Volterra, 1931).

D'altra parte, già Darwin aveva posto in luce che esistono numerose interazioni non trofiche, ad esempio quelle sessuali, che possono notevolmente influire sui processi evolutivi. Queste argomentazioni furono ampiamente descritte nella sua opera più discussa "*Descent of man*" pubblicato nel 1871 dal solito John Murray. In seguito, l'ecologia del comportamento e lo studio dello sviluppo delle comunità hanno messo in relazione un'ampia serie di rapporti competitivi sia di tipo intra che interspecifico legati alla possibilità di accesso a diversi tipi di risorse (cibo, spazio, partner sessuale).

Più recentemente, l'idea di competizione è stata applicata con successo addirittura a livello cellulare, come nel caso delle interazioni tra neuroni nella costruzione di reti neurali (Kohonen, 2001) o in quello della competizione spermatica nella fecondazione dell'uovo (Wedell *et al.*, 2002).

In particolare, nell'ambito dell'ecologia del benthos marino, la competizione spaziale è considerata, almeno relativamente agli ambienti non stressati da fattori fisici, la chiave di volta della chiarificazione dei fenomeni che portano allo sviluppo di una comunità matura. Spugne, coralli e briozoi, organismi apparentemente inermi, lottano tenacemente per lo spazio con un'intensità che progressivamente aumenta con l'aumentare della biodiversità ambientale. Soprattutto per quanto riguarda i substrati rocciosi, già dagli anni 60 è apparso chiaro che, in acque superficiali, coperture del fondale superiori al 100% rappresentano la norma a qualunque latitudine.

2) Epibiosi

Noli me tangere!
Gv 20, 17

In numerosi ambienti non solo marini, una delle soluzioni più sorprendenti al problema della competizione spaziale è quella dell'epibiosi, nella quale alcuni organismi utilizzano il corpo di un ospite come substrato.

Questa strategia è particolarmente efficace perché numerosi organismi sono protetti da esoscheletri duri, più o meno mineralizzati, come le conchiglie dei molluschi, il carapace di numerosi artropodi, i tubi dei policheti, i coralliti delle madrepore, sui quali gli epibionti possono insediarsi riducendo al minimo le interazioni con l'ospite.

In molti casi, questa strategia apre alle specie epibionti imprevedibili opportunità come l'occupazione dei fondi mobili da parte di specie tipiche di fondi duri o quella di sfruttare le possibilità offerte da un substrato in movimento o comunque vivente. Sui fondali antartici incoerenti una ricca comunità, caratterizzata da poriferi, idrozoi e gorgonie, vive come epibionte delle conchiglie del comune pettine *Adamussium colbecky* o sugli aculei del riccio *Cidaris* sp. Per i filtratori passivi, come i poriferi, la vita sulle conchiglie dei bivalvi è particolarmente interessante per la possibilità di sfruttare le correnti indotte dall'attività dell'ospite (Cerrano *et al.*, 2006). Un altro esempio riguarda una comunità di idrozoi che riesce a vivere sui fanghi batiali del Mediterraneo a profondità superiori a 1000 m, insediandosi sulla faccia inferiore dell'esoscheletro del granchio di profondità *Geryon longipes.* In questa posizione i polipi degli idrozoi possono sfruttare le particelle organiche risospese dal movimento delle appendici del granchio (Di Camillo *et al.*, 2007).

Bisogna comunque notare che, se l'epibiosi apre nuove prospettive per la vita dell'epibionte, non è detto che avvantaggi anche l'ospite. Ad esempio, quando le conchiglie di *Adamussium* sono ricoperte da un popolamento troppo ricco, perdono le loro capacità di spostamento attuato tramite i movimenti delle valve (clapping) e divengono facili prede delle voraci stelle di mare antartiche (Cerrano *et al.*, 2006).

Un fenomeno ancora più complesso riguarda l'insediamento degli epibionti direttamente sui tessuti dell'ospite, provocando la reattività cellulare degli stessi. Sappiamo, infatti, che gli organismi del benthos lottano strenuamente per evitare l'insediamento di epibionti sui loro tessuti viventi. Il loro armamentario non è per nulla convenzionale e si compone

di vari tipi di armi meccaniche e, soprattutto, chimiche che possono essere specificamente iniettate nei tessuti del competitore (come nel caso delle nematocisti degli cnidari) (Mariscal, 1974) o più genericamente liberate nell'ambiente circostante (Gerhart *et al.*, 1988). Nell'ultimo decennio le scoperte dell'ecologia chimica hanno mostrato un campionario sterminato di queste molecole prodotte dal metabolismo secondario degli organismi marini, che vengono utilizzate come sostanze di difesa contro l'epibiosi, dimostrando ancora una volta come queste relazioni abbiano una base antagonistica.

D'altra parte, se un ipotetico epibionte evolve una qualche forma di resistenza nei confronti di alcune di queste sostanze antifouling, gli si apre improvvisamente la possibilità di occupare una nuova nicchia ecologica a bassissima competizione interspecifica. E' evidente che questa possibilità è stimolata da una lunga vicinanza tra ospite ed epibionte, suggerendo che in ambienti estremamente diversificati, dove gli organismi sono costretti a vivere strettamente addossati, si sviluppino fenomeni di resistenza chimica che possono essere alla base di più specifiche interazioni.

Bisogna notare in questo senso, come la possibilità della vita epibionte abbia, in alcuni gruppi, dato origine ad imponenti fenomeni di speciazione: è il caso degli idrozoi dove, intere famiglie o interi generi, composti da numerose specie, si sono evolute con uno stile di vita epibiontico in relazione a diversi gruppi di ospiti (la famiglia Hydractiniidae è specializzata per la vita sui gasteropodi, la famiglia Zancleidae vive quasi esclusivamente sui briozoi, il genere *Proboscidactyla* sul bordo dei tubi dei policheti, il genere *Sarsia* è epibionte di poriferi, il genere *Ralpharia* vive esclusivamente su ottocoralli, il genere *Eugymnanthea* sulle branchie dei bivalvi (*e.g.* Puce *et al.*, 2005).

3) Dalla competizione alla cooperazione

"parte di quella forza
che sempre vuole il male
e sempre il bene crea"
Goethe, Faust

Uno degli apici evolutivi dell' epibiosi antagonistica è alla base dello stile di vita di alcuni cnidari parassiti (generi *Gerardia, Alcyonium, Parazoanthus*) che ricoprono lo scheletro di vari coralli arborescenti dopo

averli progressivamente uccisi. La biologia di *Gerardia savaglia* del Mediterraneo è ancora poco conosciuta: quando la larva riesce ad insediarsi su una gorgonia, probabilmente in una piccola regione per qualche motivo privata del cenenchima vivente, comincia a crescere e contemporaneamente a produrre sostanze che uccidono i polipi della gorgonia nelle immediate vicinanze. Sul terreno così liberato, lo zoantideo cresce con una velocità che gli permette di eliminare definitivamente l'ospite nel giro di qualche anno. A questo punto *Gerardia* comincia a deporre sullo scheletro assile della gorgonia un suo proprio scheletro che irrobustisce quello preesistente e, contemporaneamente, inizia a riprodursi.

Altre specie di zoantidei tropicali del genere *Parazoanthus* hanno una strategia simile a quella di *Gerardia.* Questi polipi uccidono e ricoprono i coralli neri nei mari indonesiani ma, almeno nel rapporto con l'idrozoo tropicale *Plumularia habereri,* l'interazione dà origine ad un risultato inatteso. Anche in questo caso, lo zoantideo si insedia sul tronco dell'ospite e comincia a crescere ricoprendolo, ma non vengono mai attaccati i rami secondari su cui si trovano i polipi dell'idrozoo. Quest'ultimo, d'altra parte, reagisce crescendo in altezza e aumentando progressivamente il diametro del tronco e dei rami. Quindi, mentre *Plumularia* in condizioni normali è una piccola colonia di alcuni centimetri di altezza, quando è ricoperta dall'epibionte diventa un enorme ventaglio di un metro, nel quale lo zoantideo occupa i rami principali mentre i rami laterali secondari rimangono occupati dai polipi dell'idrozoo. L'associazione così composta aumenta enormemente, per entrambi i partners, la possibilità di ottenere cibo incrementando le dimensioni dell'area filtrante e contemporaneamente allontanandola dal substrato.

Questo caso è particolarmente significativo perché indica chiaramente come l'epibiosi possa evolvere in un rapporto cooperativo che risulta in una serie di proprietà emergenti, imprevedibili a partire da quelle dei due partners. Indica, inoltre, una possibile via filogenetica per lo stabilirsi di associazioni mutualistiche: la competizione per il substrato, risultante da un'elevata biodiversità, induce l'evoluzione di una resistenza alle difese chimiche di un organismo ospite da parte di un altro che risulta in un caso di epibiosi antagonistica. Da questa interazione può svilupparsi una simbiosi mutualistica con produzione di nuove proprietà da parte dei due organismi associati. E' evidente come i meccanismi regolativi di questi rapporti devono essere ricercati a livello di interazioni chimiche e cellulari tra i due organismi partecipanti all'associazione.

4) Cooperazione: proprietà emergenti

"Più nulla temo con te sotto al braccio,
perciò lancio al mio secolo la sfida"
Schiller

Dunque, i rapporti antagonistici legati alla competizione spaziale portano, talvolta, a risultati inattesi, quando i due partners non solo evitano vicendevolmente di uccidersi, ma dalla loro unione sorge un nuovo super-organismo con nuove proprietà impreviste rispetto a quelle dei due organismi isolatamente considerati.

Uno dei casi meglio studiati riguarda la relazione che si instaura tra numerose specie di briozoi e gli idrozoi del genere *Zanclea.* E' stato dimostrato sperimentalmente che nelle barriere coralline caraibiche il briozoo isolato è molto meno competitivo dell'associazione idroide-briozoo. I polipi, tramite le loro batterie di nematocisti, difendono il briozoo che può evitare la predazione e lottare più efficacemente per il substrato, diffondendosi più rapidamente e formando più colonie di dimensioni maggiori (Ostman & Haugsness, 1981).

Un altro caso è stato osservato nelle straordinarie barriere coralline dell'arcipelago indonesiano e riguarda la spugna *Mycale vansoesti* che vive in simbiosi con una macroalga corallinacea del genere *Amphiroa*. La spugna non è mai stata raccolta senza l'alga associata, mentre l'alga vive comunemente tra i coralli dove forma piccoli cuscinetti semisferici. Su questi ciuffi si insedia la larva della spugna che inizia a crescere forzando lo sviluppo stesso dell'alga che perde completamente la sua tipica struttura, ramificandosi all'interno dei tessuti dell'ospite fino a costituirne lo scheletro tridimensionale. La spugna utilizza, quindi, l'alga come struttura di sostegno e, d'altra parte, l'alga all'interno della spugna aumenta notevolmente le sue dimensioni; inoltre, i due partners uniti hanno una distribuzione batimetrica molto più ampia di quella dell'alga isolata. E' probabile che l'armamentario chimico della spugna permetta all'associazione di resistere efficacemente in ambienti ad elevata competitività dove l'alga isolata soccomberebbe (Calcinai *et al.*, 2006). E' evidente che questa maggiore competitività, ottenuta dall'insieme dei partners, tende a selezionare gli organismi più predisposti a cooperare stabilizzando, da un punto di vista evolutivo, il rapporto.

Sono stati osservati casi nei quali i fenomeni cooperativi determi-

nano la distribuzione geografica degli organismi coinvolti. Ad esempio, una piccola gorgonia caraibica, *Carijoa riisei,* tipica del fouling delle imbarcazioni, è diventata una specie invasiva in numerose scogliere coralline del Pacifico dove è stata trasportata insediata sulla carena delle navi. E' notevole il fatto che una spugna conosciuta esclusivamente per i Caraibi, sia diventata comune anche nel Pacifico dove si trova stabilmente in associazione con *Carijoa.* E' suggestivo pensare che questa spugna abbia viaggiato attorno al mondo in associazione con la gorgonia che ricopre difendendola dai predatori e dalla quale viene sostenuta (Calcinai *et al.*, 2004).

5) ENDOSIMBIOSI

"Ossia non è cambiato il dentro,
quello che prima era il fuori in cui nuotavo sotto il sole,
e in cui nuoto, nel buio, anche adesso che sta dentro"
Calvino, Ti con zero

Per alcuni aspetti l'endosimbiosi presenta tratti ancor più impressionanti. In ambiente marino questo fenomeno -microalghe o batteri che vivono intra- o intercellularmente all'interno del corpo dei metazoi - è incredibilmente diffuso e rappresenta la base nella strutturazione di alcuni dei più biodiversificati ecosistemi al mondo.

Probabilmente, il fenomeno endosimbiotico marino più noto riguarda la convivenza di un gruppo di microalghe, le zooxantelle, con numerosi organismi del benthos, in particolare i coralli costruttori di barriere. Il fenomeno non è ancora del tutto chiarito, ma è certo che le sostanze prodotte dalla fotosintesi delle alghe rappresentino il principale apporto trofico per i coralli che, in questo modo, possono prosperare costruendo imponenti strutture carbonatiche. Si tratta di una relazione ideale, che totalmente sfugge alla logica del "chi mangia chi", le zooxantelle, infatti, attuano la fotosintesi all'interno delle cellule dell'ospite utilizzando i cataboliti azotati quale fonte di nutrienti, mentre l'ospite utilizza direttamente gli essudati prodotti dalle alghe. La descrizione di questo meccanismo ha permesso di chiarire il paradosso delle barriere coralline che presentano un elevatissima produzione in alcune delle aree marine più oligotrofiche del mondo (Muscatine & Cernichiari, 1969).

L'altra faccia della medaglia risiede nel fatto che un meccanismo così sofisticato è estremamente fragile: i recenti bleaching verificatesi in alcune delle più importanti barriere coralline del mondo sono l'effetto più drammatico della rottura dell'equilibrio che regola il rapporto simbiotico. Durante questi fenomeni, a causa di cambiamenti ambientali, i polipi del corallo espellono i simbionti diventando, in questo modo, bianchi (to bleach = candeggiare). Normalmente, i polipi non sopravvivono all'espulsione dei propri simbionti causando gravi crisi alla stabilità delle barriere coralline (Brown, 1997).

Meno conosciuti sono altri casi di endosimbiosi che legano organismi procarioti, alghe azzurre e batteri, con la maggior parte delle specie di spugne. Tali relazioni sono talmente importanti che spesso la biomassa batterica presente in una spugna è superiore a quella dello stesso ospite (Taylor *et al.*, 2004). Ha fatto epoca, nell'ultima parte del secolo scorso, la scoperta che i lunghi pogonofori che rappresentano il più caratteristico elemento delle sorgenti calde abissali scoperte in molte regioni del mondo, hanno perso l'apparato digerente e si nutrono grazie a popolazioni di batteri chemiosintetici che vivono al loro interno.

Le relazioni tra simbionti ed ospiti sono estremamente complesse, con impressionanti fenomeni di integrazione genetica e metabolica che portano alla produzione di molecole sintetizzate in collaborazione da entrambi i partners (Stat *et al.*, 2006). D'altra parte, è ormai nota da lungo tempo e sempre più convincente, l'idea che la stessa origine della cellula eucariotica sia derivata da una interazione simbiotica tra diversi organismi procarioti (Margulis, 1993).

In particolare, bisogna notare che anche l'endosimbiosi nasce da meccanismi antagonistici legati alla predazione. Le larve di numerosi coralli costruttori nascono prive di zooxantelle che vengono ingerite dalla larva durante le prime fasi dello sviluppo. A questo punto, insorge però un meccanismo di riconoscimento tra il gastroderma della larva e l'alga unicellulare che, invece di essere digerita, viene incorporata nei tessuti dell'ospite dove, dopo una modificazione strutturale, inizia ad integrare il proprio metabolismo con quello della madrepora.

Anche in questo caso, quindi, un fenomeno "micro predatorio" evolve in un rapporto cooperativo, e questo tipo di eventi, accaduti indipendentemente più volte, potrebbero aver rappresentato alcune delle tappe fondamentali della storia della vita sulla terra.

6) Effetti

"Come tutto s'intesse nel gran Tutto
e ogni cosa nell'altra opera e vive!"
Goethe, Faust

A livello della comunità, il principale effetto dei fenomeni dell'epibiosi e dell'endobiosi é un aumento della biodiversità grazie ad un sostanziale aumento dell'eterogeneità spaziale e delle possibili interazioni. Queste relazioni non si limitano all'utilizzo di un particolare substrato a bassa competizione, ma hanno la caratteristica multidimensionalità delle nicchie ecologiche in cui l'interazione rappresenta lo spazio biologico in cui si svolge l'intera storia vitale di un organismo (insediamento, sviluppo, alimentazione, riproduzione).

Gli eventi speciativi sono forzati dalle relazioni producendo organismi adattati, da un punto di vista sia morfologico sia fisiologico e comportamentale, alla morfologia, fisiologia e comportamento dell'ospite. In acque tropicali, praticamente ogni specie di gorgonia, di attinia e di corallo nero presenta crostacei decapodi (granchi o gamberi) adattati alla vita tra i polipi. Alcuni gamberetti vivono sui coralli neri assumendo esattamente il colore dei polipi e i piccoli granchi della famiglia Maijdae vivono sugli stessi coralli, hanno una morfologia particolarmente mimetica, ma sanno anche staccare i polipi del corallo trapiantandoseli sul carapace e fissandoli tramite appositi peli uncinati. Gli anfipodi caprellidi che vivono in Mediterraneo sugli idroidi del genere *Eudendrium* non sanno catturare prede direttamente, mentre sono abilissimi ad aprire, senza danneggiarli, gli stomaci dei polipi per estrarre le prede da essi catturate grazie alle nematocisti dei tentacoli. E' stato recentemente dimostrato che questa infinita teoria di crostacei che popolano le differenti specie di cnidari migliora la sopravvivenza dei loro ospiti eliminando i sedimenti che si depositano sulle colonie (Stewart *et al.*, 2006)).

Una simile chiave di lettura può essere applicata al caso, particolarmente eclatante, della formazione delle scogliere da parte dei coralli costruttori. E' noto che i rapporti simbionte/ospite hanno una diretta influenza sulla precipitazione del carbonato che, poi, le singole specie di coralli modellano in accordo con la morfologia tipica di ogni specie. I coralli agiscono, quindi, come *ecosystem engineers,* organismi in grado di modificare l'ambiente attraverso le proprie strutture fisiche come

tessuti o scheletri vivi o morti in modo da creare habitat per altri organismi che sono in grado di vivere sia su di essi sia dentro ad essi (Jones *et al.* 1994). Il risultato dell'interazione tra coralli e alghe determina, in questo modo, uno dei più biodiversificati ecosistemi del mondo.

Quindi, l'effetto più impressionante delle relazioni simbiotiche in tutta la loro gamma, che va dalla competizione alla cooperazione, è quello dell'autocatalisi della biodiversità che, seguendo una geometria frattale, rimane autosimile al variare di scala: ogni organismo serve da substrato per epibionti più piccoli, a loro volta colonizzati da altri epibionti in una continua riduzione di taglia il cui limite raggiunge quello dell'organizzazione vitale. Lo studio delle comunità microscopiche con l'uso della microscopia elettronica ha dimostrato che sulle colonie di piccoli metazoi coloniali, come gli idrozoi, prosperano foreste di diatomee arborescenti tra le quali vivono diatomee sessili e negli spazi tra queste si osservano distese di batteri. Recentemente, è stata dimostrata l'importanza dei virus nell'ecologia degli ecosistemi marini (Danovaro & Serresi, 2000).

Quanto detto potrebbe rappresentare una nuova prospettiva per la comprensione dei sistemi viventi. Le relazioni tra gli organismi sono interpretabili solo in prima approssimazione sulla base di reti trofiche. Certo, i flussi di materia e di energia lungo le reti trofiche rappresentano lo scheletro dei sistemi ecologici, ma tali reti devono essere immaginate come intersecate da un'altra rete, una rete di relazioni "non trofiche", simbiotiche in senso generale, nelle quali la competizione e la cooperazione operano a diversi livelli (Montoya *et al.*, 2006).

Le reti simbiotiche rappresentano, rispetto alle reti trofiche, un potente sistema di regolazione in grado di modulare ogni rapporto tra organismi, aumentando la biodiversità, inducendo la speciazione, migliorando la convivenza, incrementando l'efficienza della ciclizzazione della materia e rappresentando una delle principali sorgenti di novità macroevolutive nella storia della vita.

BIBLIOGRAFIA

Brown BE (1997) Coral bleaching: causes and consequences. Coral Reef, 16: 129-138.

Calcinai B, Bavestrello G, Cerrano C (2004) Dispersal and association of two alien species in the Indonesian coral reefs: the octocoral *Carijoa riisei* and the demosponge *Desmapsamma anchorata*. J. Mar. Biol. Ass. U.K. 84: 937-941.

Calcinai B, Cerrano C, Totti C, Romagnoli T, Bavestrello G (2006) Symbiosis of Mycale

(Mycale) vansoesti sp. nov. (Porifera, Demospongiae) with a coralline alga from North Sulawesi (Indonesia). Invertebrate Biology 125(3): 195-204.
Cerrano C, Calcinai B, Bertolino M, Valisano L, Bavestrello G (2006) Epibionts of the scallop *Adamussium colbecki* in the Ross Sea, Antarctica. Chem Ecol. 22: 235-244.
Danovaro R, Serresi M , (2000) Viral density and virus to bacterium ratio in deep-sea sediments of the eastern Mediterranean. Applied and Environmental Microbiology. 66: 1857-1861.
Di Camillo C, Bo M, Puce S, Tazioli S, Froglia C, Bavestrello G (2007) The epibiontic assemblage of *Geryon longipes* (Crustacea: Decapoda: Geryonidae) from Southern Adriatic Sea. It. J. Zool. (in stampa).
Gerhart D.J., Rittschof (D.) & Mayo SW(1988) Chemical ecology and the search for marine antifoulants. J. Chem. Ecol. 14 (10): 1573-1561.
Jones CG, Lawton JH, Shachak M (1994) Organisms as ecosystem engineers. Oikos 69: 373-386.
Kohonen T (2001) Self-Organizing Maps Springer Series in Information Sciences, Vol. 30, Springer, Berlin, Heidelberg, New York.
Margulis L (1993) Symbiosis in Cell Evolution, 2nd Edition. Freeman, New York.
Mariscal RN (1974) Nematocysts. In: Coelenterate biology: reviews and new perspectives (Muscatine L, Lenhoff HM, eds.), Academic Press, New York, 129-178.
Muscatine L, Cernichiari E. (1969) Assimilation of Photosynthetic Products of Zooxanthellae by a Reef Coral. Biological Bulletin, 137: 506-523.
Montoya JM, Pimm SL, Sole RV (2006) Ecological networks and their fragility. Nature, 442: 259-264.
Ostman RW, Haugsness JA (1981) Mutualism among sessile invertebrates: a mediator of competition and predation. Science, New York 211: 846-848.
Puce S, Calcinai B, Bavestrello G, Cerrano C, Gracili C, Boero F (2005) Hydrozoa (Cnidaria) symbiotic with Porifera: a review. PSZN: Marine Ecology 26: 73-81.
Stewart HL, Sally Æ, Holbrook J, Schmitt RJ, J. Brooks ÆAJ (2006). Symbiotic crabs maintain coral health by clearing sediments. Coral Reefs 25: 609-615.
Stat M, Carter D, Hoegh-Guldberg O (2006) The evolutionary history of *Symbiodinium* and scleractinian hosts—Symbiosis, diversity, and the effect of climate change. Perspectives in Plant Ecology, Evolution and Systematics 8: 23-43.
Taylor MW, Schupp PJ, Dahllöf I, Kjelleberg S and Steinberg PD (2004) Host specificity in marine sponge-associated bacteria, and potential implications for marine microbial diversity. Environmental Microbiology 6: 121-130.
Volterra V (1931) Variations and fluctuations of the number of individuals in animal species living together. In: Animal Ecology. McGraw-Hill.

Istituto Lombardo (Rend. Scienze) 151, 121-131 (2017)

THE NATURE OF SUPERCONDUCTING FLUCTUATIONS AND NEW ASPECTS IN HIGH T_C SUPERCONDUCTORS

ALESSANDRO LASCIALFARI (*,**) e ATTILIO RIGAMONTI (**)

Nota presentata dal Prof. A. Lascialfari
(Adunanza del 29 giugno 2017)

SUNTO. – Dopo una breve introduzione sulle fluttuazioni superconduttive, si presentano tipici risultati sperimentali in diversi sistemi. Le fluttuazioni superconduttive sono note esistere sin dagli anni '60 e sono state evidenziate dapprima in nanoparticelle di superconduttori convenzionali. Nelle ultime decadi, attraverso misure di diamagnetismo precursore al di sopra della temperatura critica T_c, sono stati messi in evidenza particolari aspetti delle fluttazioni nei superconduttori ad alta T_c, particolarmente in quelli definiti "sotto-drogati". I risultati ottenuti, qui riassunti, sono stati interpretati in termini di fluttuazioni della fase del parametro d'ordine superconduttivo, una funzione complessa. Un modello teorico, già sviluppato per condizioni rigorose di dimensionalità nulla (quindi in sistemi di nanoparticelle) è stato da noi esteso a includere approssimativamente anche sistemi a dimensioni finite, dell'ordine della distanza di coerenza tra isole locali di superconduttività fluttuante. Sono cosi stati descritti e giustificati i risultati di diamagnetismo fluttuante in cuprati, pnictidi e anche in superconduttori a base Fe. Oltre a questi risultati, qui riassunti, vengono anche presentati risultati preliminari ottenuti in superconduttori a base Ca. Si mostra come un particolare parametro (il campo di upturn) ove la magnetizzazione diamagnetica inizia a diminuire al crescere del campo, consente di discriminare con sicurezza gli effetti di natura squisitamente superconduttiva da similari effetti di natura più convenzionale, legati a una distribuzione di T_c per disomogeneità di carattere chimico.

(*) Dipartimento di Fisica & INSTM, Università degli Studi di Pavia, Italy.
E-mail: alessandro.lascialfari@unipv.it
(**) Dipartimento di Fisica "A. Volta", Università degli Studi di Pavia, Italy.
E-mail: attilio.rigamonti@unipv.it

***

ABSTRACT. – After a brief introduction to thermodynamical superconducting fluctuations, some experimental results in different systems are presented. Superconducting fluctuations were predicted and revealed in nanoparticles of conventional superconductors since late 60s. However, just in the last two decades phenomena related to an anomalous diamagnetism have been discovered in high-T_c underdoped compounds, mainly by collecting the isothermal magnetization curves above the critical temperature T_c. These experimental data can be interpreted in terms of fluctuations of the phase of the order parameter and a related theoretical model can explain the results in cuprates, pnictides and Fe-based materials. Such results, here presented together with preliminary data on a Ca-based high-T_c system, allow one to extract a fundamental parameter, the upturn magnetic field above which the diamagnetic magnetization initiates to decrease on increasing the field. Its temperature dependence distinguishes among diamagnetic effects related to superconducting phase fluctuations and to analogous effects due to chemical inhomogenities.

1. FLUCTUATIONS, SUPERCONDUCTING FLUCTUATIONS AND RELATED DIAMAGNETISM

The concept of fluctuations of a specific quantity, is spreaded across different disciplines, ranging from *e.g.* economy, finance, health sciences, mathematics, chemistry and physics. In physics, one of well known experimental evidences of fluctuations (of thermodynamical type), is the density fluctuations in liquids, generating the critical opalescence effect, observed since centuries ago, that can be described by means of the Ginzburg-Landau (GL) theory. As a translation of the GL theory of fluctuations in correspondence of the transition critical temperature in different systems (magnetic, superconducting, superfluid, etc), superconducting fluctuations (SF) of the order parameter (the Cooper pair wave-function [1,2]) were evidenced by measurements mainly of the resistivity, the magnetization and the susceptibility just above T_c [3]. After the discovery of high-T_c systems [4], through STM microscope it was possible to observe directly superconducting fluctuations in the system $La_{2-x}Sr_xCuO_4$.

From the quantitative point of view, the order parameter of superconductors is represented by the wave-function of the Cooper pairs: $\Psi = |\Psi| e^{i\vartheta}$, where $|\Psi|$ is the modulus and ϑ is the phase [5]. The square modulus of this paramater can be locally different from zero for $T \geq T_c$, where Cooper pairs start to form in metastable states: the system is preparing for the transition to the superconducting phase (*Fig. 1*).

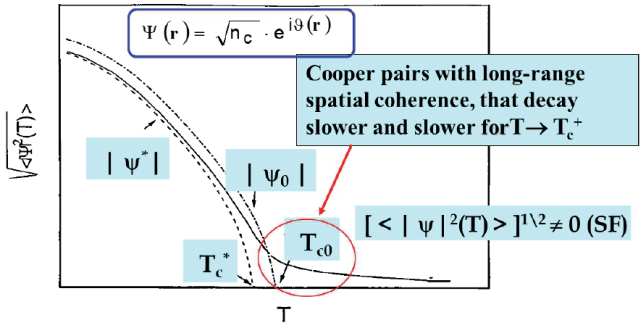


Fig. 1 – The order parameter is renormalized by the fluctuations. Dashed line: curve deduced from the Bardeen-Cooper-Schrieffer theory, with T_c renormalized by fluctuations. Continuous line: T-dependence of |Ψ| renormalized by fluctuations.*

The occurrence of superconducting fluctuations has relevant consequences: the "fluctuating" Cooper pairs above T_c imply the presence of (fluctuating) diamagnetism and of excess conductivity (*Fig. 2*). A qualitative understanding of the fluctuating diamagnetism (FD) can be achieved by considering the two electrons (of the Cooper pairs) at an average distance of the order of $\xi_{GL}(T)$ (the Ginzburg-Landau coherence length).

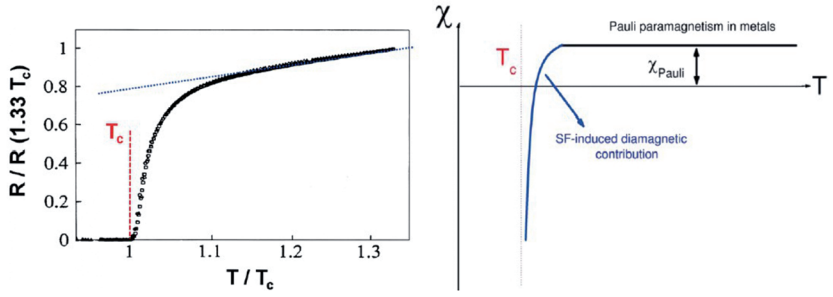


Fig. 2 – Left: renormalized eletrical resistance as a function of reduced temperature. Above T_c where the SF set in, the resistance is decreased with respect to metallic behaviour (blue line). Right: effect of magnetic susceptibility "smoothing" just above T_c, as consequence of the fluctuating diamagnetism.

The theory of FD in conventional superconductors (SC) in the limit of zero field has an exact solution, proposed by Prange [2,5,6], but no field quenching of fluctuating pairs is taken into account. In the

1960s, Gollub, Tinkham et al [7,8], demonstrated experimentally in metals the breakdown of a description which neglects the effect of the field in suppressing the fluctuating pairs. Further developments of the GL theory of SF in the gaussian approximation, allowed to extract the H-dependence of the magnetization (*Fig. 3*).

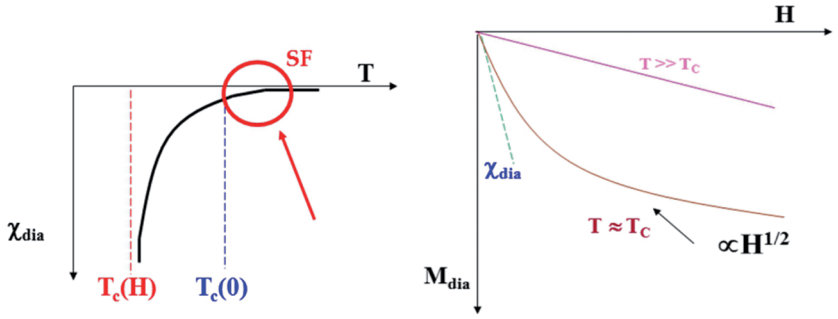


Fig. 3 – Sketchy T-behaviour of magnetic susceptibility (after subtraction of the Pauli term) and H-behaviour of the isothermal magnetization curves in the presence of SF. The curves on the right of the figure report the behaviour according to the GL theory in the gaussian approximation.

2. Superconducting fluctuations in Pb nanoparticles: the zero-dimensional case

A case where the theory of superconducting fluctuations has been directly verified, is the zero-D (0D) case for Pb nanoparticles [2,5]. In *Fig. 4* the diamagnetic magnetization M_{dia} is reported as a function of H for 3 distinct temperatures just above T_c ($T_c(0)$ = 7.09 K), in the sample with average diameter d=750 Å. The solid lines represent the best fit obtained by using the 0D equations [5], derived from complete GL theory in 0D (*i.e.* by including the $|\Psi_0|^4$ term in the free energy expression [2,5]). It is noted the presence of an experimental "upturn field" H_{up} that shifts toward lower fields as the temperature is decreased: H_{up} =110, 67, 60 Oe. The zero-dimensional condition is applied to the evanescent SC droplets of size which causes the maximum diamagnetic screening. The magnetic susceptibility data present clearly the diamagnetic effect, reported in *Fig. 5* for two different particles dimensions. In the 0D-case it could be noted that, by means of GL theory, a direct eval-

uation of the thermodynamical free energy can be made and the effect of H can be exactly taken into account [9].

Interestingly the data in MgB_2 are qualitatively explained for $\xi(T\rightarrow 0) \approx 100$Å and $10^{-3} < \varepsilon < 10^{-2}$ ($\varepsilon=(T-T_c)/T_c$); in this case one obtains $H_{up} \approx 100 \div 1000$ Oe by using the expression of diamagnetic magnetization in the 0D case [10]:

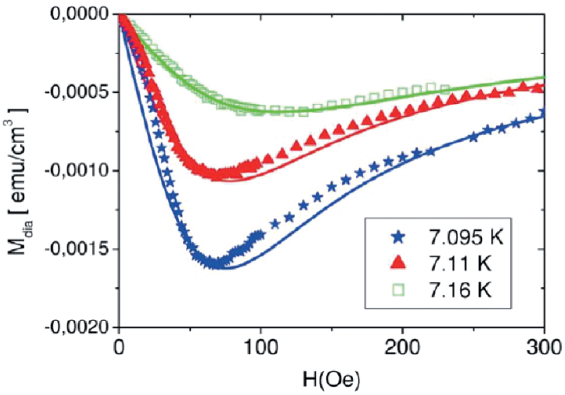


Fig. 4 – Pb nanoparticles with average diameter d=750 Å: magnetization curves for T just above $T_c(0)$=7.09K. The "upturn" field, clearly evidenced by experimental data, shifts toward higher values as the temperature is increased.

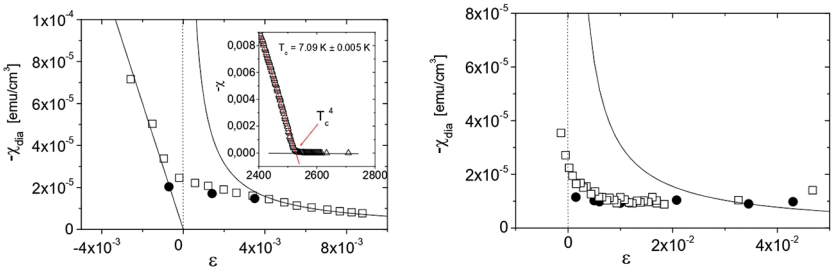


Fig. 5 – Temperature behavior of the magnetic susceptibility for Pb nanoparticles of average diameter 750 Å (left) and 160 Å (right) (figure from [9]). Solid circles: data from the isothermal magnetization curves; empty squares: data from the isofield measurements as a function of T. The solid lines report the behavior of M_{dia} obtained by neglecting nonlinear fluctuations. In the inset the curve from which the transition temperature has been determined is shown (ε=reduced temperature).

3. HIGH-T_c SUPERCONDUCTORS: ANOMALOUS FLUCTUATING DIAMAGNETISM

Since 1986, a new class of high-T_c SC (HTSC) compounds has been developed. A particular ensemble of such systems is constituted by the so-called high-T_c cuprates, the first example being the one of Bednorz and Muller [4], the $La_{2-x}Ba_xCuO_4$ oxide. These systems do not follow the Bardeen-Cooper-Schrieffer (BCS) theory [3], the main coupling mechanism still being substantially unknown. The main family of high-T_c compounds other than cuprates, are the Fe-based one [11].

In HTSC, the diamagnetic effects above T_c depend on the charge carriers concentration n_h [12-14]. One has to distinguish between optimally doped materials (highest T_c, optimal concentration n_h^{opt}), underdoped compounds ($n_h < n_h^{opt}$) and overdoped ones ($n_h > n_h^{opt}$). The behaviour of T_c *vs* n_h is plotted in *Fig. 6* in order to clarify the different dependence of T_c from the carrier concentration in the two branches.

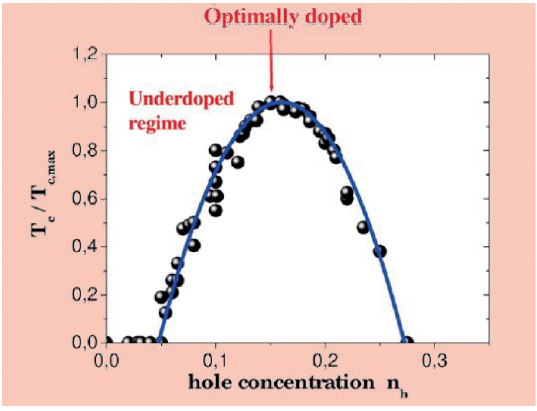


Fig. 6 – Renormalized superconducting transition temperature T_c vs charge carriers concentration n_h. The curve is universal, for a variety of HTSC.

In the underdoped HTSC, the SF has an anomalous diamagnetic behaviour above T_c. The main characteristics are (see *Fig. 7* for the system $SmBa_2Cu_{3-y}Al_yO_{6+d}$ [15]): (i) the absolute value of the diamagnetic magnetization is particularly high, presumably due to high T_c and to the short coherence length; (ii) there exists an upturn field H_{up} that assumes low values (with respect *e.g.* to the critical field ones), increases with T (for $T>T_c$) and depends on the magneto-thermal history. The origin of

this anomalous behaviour relies in the phase (ϑ) fluctuations of the order parameter, as evidenced by the model presented *e.g.* in Ref.s [14,16]. The model predicts the presence of vortexes and anti-vortexes above T_c, giving rise to mesoscopic superconducting droplets (and phase fluctuations). The irreversibility effect and the effect of the magnetic field on the thermally-activate vortexes has been predicted correctly as well. As a consequence of the existence of such phenomenon, the predicted M *vs* H behavior is sketched in *Fig. 8*, where the curves are not in scale. As seen, in addition to the observed "low" upturn field, there is a second "high" upturn field (> 2,5 $\times 10^4$ Oe) corresponding to the Cooper pairs breaking and related to the superconducting critical field.

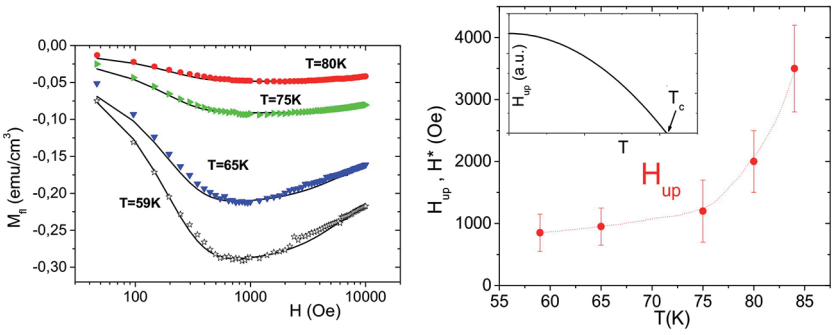


Fig. 7 – Fluctuating magnetization curves in $SmBa_2Cu_{3-y}Al_yO_{6+\delta}$ (left). The temperature behaviour of the upturn field H_{up} is reported on the right (for details on this aspect see Ref. [15]). It is noted the increase of H_{up} with T, typical of underdoped HTSC compounds.

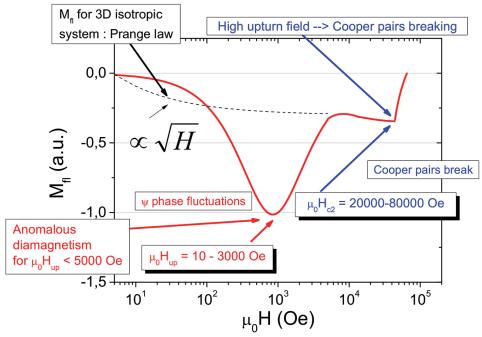


Fig. 8 – Sketchy view of the fluctuating magnetization behaviour as a function of the external field, above T_c in a system where anomalous diamagentism occurs. At low fields, the presence of an upturn field due to phase fluctuations of the order parameter is evidenced. At high field the conventional critical field upturn due to the Cooper pairs break is shown.

4. Superconducting fluctuation diamagnetism in pnictides, Fe-based systems and in CALABALACUO

The presence of an anomalous diamagnetism of sizeable strength and of an upturn field in the isothermal fluctuating magnetization curves, has been revealed also in pnictide SC systems [17] and in Fe-based superconductors [18]. The fluctuating magnetization curves are shown in *Fig. 9*, left and right side, respectively.

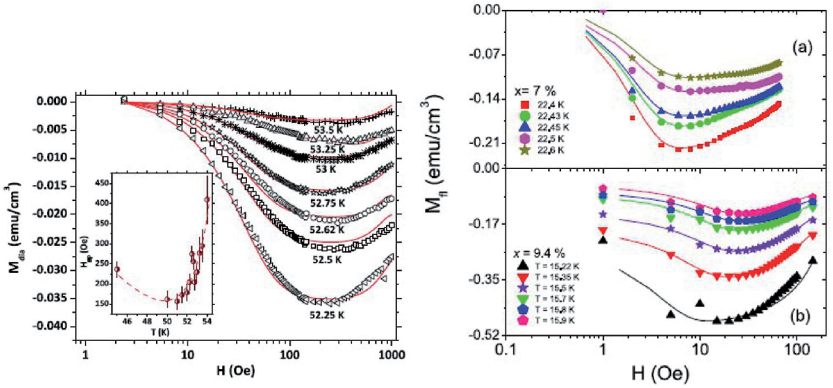


Fig. 9 – Left: isothermal fluctuating magnetization in $SmFeAsO_{0.8}F_{0.2}$. The upturn fields (ranging from 200 to 400 Oe) at different temperatures above T^c (T_c=52.3K) are displayed. The fitting red curves have been obtained by means of an early model reported in [14] and [15]. In the inset H_{up} behaviour, increasing by increasing T, is reported (dashed line is a guide to the eye). Right: isothermal fluctuating magnetization in $Ba(Fe_{1-x}Rh_x)_2As_2$ iron-based SC, for x=7% and x=9%. The solid lines are fits obtained using the model in Ref. [15].

A final example concerns a particular Ca-doped compound, $Ca_xLa_{1-x}Ba_{1.75-x}La_{1.25+x}Cu_3O_{6+y}$ (in brief CALABALACUO, first synthesized in 1993), whose peculiarity is to have the possibility of double doping (both Ba/Ca and/or oxygen content). The crystal structure of the system is reported in *Fig. 10a*, while *Fig. 10b* reports the M *vs* H fluctuating curves typical of an underdoped compound (x=0.4, y=1.15) for $T>T_c$.

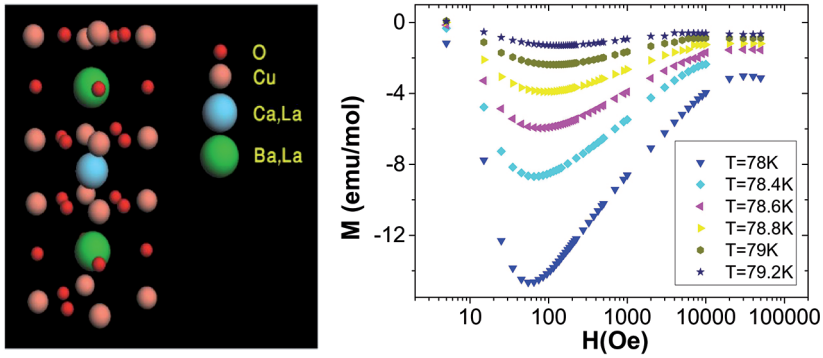


Fig. 10 – Left: Crystal structure of $Ca_xLa_{1-x}Ba_{1.75-x}La_{1.25+x}Cu_3O_{6+y}$*; Right: diamagnetic magnetization for the compound with x=0.4, y=1.15 (near to optimal doping but still underdoped). Even for this compounds, an upturn field is displayed for* $T>T_c$ *(*T_c *= 78.4K).*

5. Conclusions

It has been recalled how in underdoped HTSC compounds an anomalous diamagnetism was revealed by the experimental M *vs* H curves collected at constant temperature just above T_c (once corrected by Pauli paramagnetism and paramagnetic impurities contribution, thus giving M_{dia}, the diamagnetism fluctuating magnetization). The anomalous diamagentism is accompanied by irreversibility effects (field-cooled, FC, and zero-field-cooled, ZFC, curves different also above T_c) and can be explained in terms of mesoscopic charge inhomogeneities and of the simultaneous presence of an XY anisotropic liquid of vortexes, giving rise to phase fluctuations of the order parameter. An upturn field H_{up} that increases with T above T_c is the signature of such fluctuating diamagnetism, whereas for chemical inhogeneities the T-behaviour of the upturn is demonstrated to have the opposite T-behaviour. The effect of the field is qualitatively described by the 0-D model derived from the classical GL theory: above T_c evanescent droplets set in and an upturn field increasing with T is singled out. On the other hand, in optimal SC M_{dia} follows the conventional GL theory thus giving M_{dia} linear in H for $T>>T_c$, and $M_{dia} \propto H^{1/2}$ around T_c.

Finally one should note that the diamagnetic behaviour could be linked to the superconducting mechanism and eventually to the pseudo-gap formation, a strongly debated issue [see *e.g.* 19-21].

From the nature of SF, the origin of the SC mechanism could be ideally deduced and it is currently under further theoretical investigation.

ACKNOWLEDGMENTS

The authors acknowledge Samuele Sanna for having allowed the publication of CALABALACUO data, and P.Battistoni and G. Prando for CALABALACUO data analysis. A. Varlamov and L. Romanò are gratefully thanked for their contribution in the theoretical analysis of the experimental data.

REFERENCES

[1] "The Feynman lectures on Physics", R. Feynman, Chapter 21, vol III, ed. Addison Wesley (1965).
[2] "Structure of Matter", 3rd ed., A. Rigamonti and P. Carretta, ed. Springer Verlag (2015)].
[3] "Introduction to superconductivity" 2nd ed., M. Tinkham, ed. Dover publications (2004).
[4] J.G. Bednorz and K.A. Müller, Z. Phys. B 64, 189 (1986).
[5] A.I. Larkin and A.A. Varlamov, Theory of Fluctuations in Superconductors, Clarendon, Oxford, 2005.
[6] R.E. Prange, Phys. Rev. B1, 2349 (1970).
[7] J.P. Gollub, M.R. Beasley, and M. Tinkham, Phys. Rev. Letters 25, 1646 (1970).
[8] J.P. Gollub, M.R. Beasley, R.S. Newbower, and M. Tinkham, Phys. Rev. Letters 22, 1288 (1969).
[9] E. Bernardi, A. Lascialfari, A. Rigamonti, L. Romanò, V. Iannotti, G. Ausanio, C. Luponio, Phys. Rev. B 74, 134509 (2006).
[10] A. Lascialfari, T. Mishonov, A. Rigamonti, P. Tedesco, and A.A. Varlamov, Phys. Rev. B 65, 180501R (2002).
[11] Y. Kamihara, T. Watanabe, M. Hirano, and H. Hosono, J. Am. Chem. Soc. 130, 3296 (2008).
[12] C. Baraduc, A. Buzdin, J.Y. Henry, J.P. Brison, and L. Puech, Physica C 248, 138 (1995).
[13] A. Junod, J.-Y. Genoud, G. Triscone, and T. Schneider, Physica C 294, 115 (1998).
[14] A. Lascialfari, A. Rigamonti, L. Romanò, P. Tedesco, A. Varlamov and D. Embriaco, Phys. Rev. B 65, 144523 (2002).
[15] E. Bernardi, A. Lascialfari, A. Rigamonti, L. Romanò, M. Scavini, and C. Oliva, Phys. Rev. B 81, 064502 (2010).
[16] A. Lascialfari, A. Rigamonti, L. Romano', A.A. Varlamov, I. Zucca, Phys. Rev. B 68, 100505 (R) (2003).

[17] G. Prando et al., Phys. Rev. B 84, 064507 (2011).
[18] L. Bossoni et al., J.Phys.: Cond. Matter 26, 405703 (2014).
[19] Y. Wang, L. Li, and N.P. Ong, Phys. Rev. B 73, 024510 (2006).
[20] R.I. Rey, A. Ramos-Álvarez, J. Mosqueira, M.V. Ramallo, and F. Vidal, Phys. Rev. B 87, 056501 (2013).
[21] L. Li, Y. Wang, S. Komiya, Shimpei Ono, Y. Ando, G.D. Gu, and N.P. Ong, Phys. Rev. B 81, 054510 (2010).

Scienze naturali
Zoologia
Istituto Lombardo (Rend. Scienze) 151, 133-148 (2017)

STRATEGIE INNOVATIVE PER IL CONTROLLO DI ZANZARE, VETTORI DI ORGANISMI PATOGENI

Nota del s.c. GIULIANO GASPERI (*)

(Adunanza del 29 giugno 2017)

SUNTO. – Le zanzare sono tra le specie di insetti più pericolose essendo vettori di agenti patogeni, responsabili delle malattie più diffuse, quali malaria, febbre dengue, chikungunya, febbre del Nilo occidentale, febbre gialla e Zika. Alcune zanzare dei generi *Anopheles*, *Aedes* e *Culex* sono responsabili di milioni di casi di infezioni umane, con centinaia di migliaia di morti all'anno. Ne derivano conseguenze drammatiche di carattere socio-economico e sanitario, specialmente in Paesi in via di sviluppo. Il controllo degli insetti nocivi è da tempo basato sull'uso prolungato degli insetticidi che inducono resistenza negli insetti e determinano inquinamento dell'ambiente. Nella seconda metà del XX secolo, la consapevolezza dell'importanza della salute umana e dell'ambiente ha stimolato lo sviluppo di nuovi approcci di controllo che hanno come target il comportamento riproduttivo. In particolare, la cosidetta Tecnica dell'Insetto Sterile (SIT), applicata con successo contro alcune specie di ditteri nocivi in diverse aree del mondo, ha rappresentato il punto di partenza per ricerche legate alla conoscenza dettagliata della biologia delle specie da combattere. Approcci di biologia molecolare hanno stimolato miglioramenti nell'applicazione di SIT mediante strategie di trasformazione genetica. Tecniche innovative quali RIDL, incompatibilità citoplasmatica (CI) e metodologie di gene drive rivestono un evidente interesse applicativo per il controllo/eradicazione delle zanzare.

***

ABSTRACT. – Mosquitoes are the most dangerous insect species being the vectors of the pathogens causing the most widespread diseases as Malaria, Dengue fever, Chikungunya, West Nile fever, Yellow fever, Zika. Several species of the *Anopheles*, *Aedes* and *Culex* are responsable for millions of human infections causing hundreds of thousands of deaths per year. This causes dramatic socio-economic and health conse-

(*) Dipartimento di Biologia e Biotecnologie "Lazzaro Spallanzani", Università degli Studi di Pavia, Italy. E-mail: gasperi@unipv.it

quences, especially in developing countries. Harmful insects are controlled using chemical insecticides which cause insecticide resistance and environmental pollution, due to long term use. In the second half of the 20th century, new insect environmentally safe control stategies were been developed, based on the insect's reproductive behaviour. Notably, the Sterile Insect Technique (SIT) has been successfully applied against some dipteran pest species in several regions of the World. It drived several research efforts to increase our knowledge of the biology of the harmful species. SIT improvements have been also achieved with several molecular biology approaches and transgene technologies such as the innovative RIDL, cytoplasmic incompatibility (CI) and gene drive.

Introduzione

Gli insetti sono considerati gli animali di maggior successo evolutivo per la loro abbondanza e loro distribuzione in tutti gli ambienti, essendo in grado di adattarsi in modo efficace ad una grande varietà di habitat nei quali svolgono importanti funzioni ecologiche. Delle numerose specie di insetti descritte (1.5 milioni), molte sono vettori di agenti patogeni per animali e piante. Tra queste, le zanzare (famiglia Culicidae, Ordine Diptera), di cui sono state descritte circa 3.500 specie, rappresentano gli animali più pericolosi al mondo, essendo vettori di agenti patogeni (protozoi e virus). Essi sono tramessi dalle femmine infette all'ospite (uomo e/o vertebrato) mediante pasto di sangue, necessario per la maturazione delle uova. Tra le malattie più diffuse, trasmesse da alcune specie di zanzare dei generi *Anopheles*, *Aedes* e *Culex*, sono da menzionare malaria, febbre dengue, chikungunya, febbre del Nilo occidentale, febbre gialla e Zika. *Anopheles*, *Aedes* e *Culex* sono responsabili di milioni di infezioni umane, con centinaia di migliaia di morti all'anno [1; 2]. Queste patologie determinano drammatiche conseguenze di carattere socio-economico e sanitario, non solo in Paesi in via di sviluppo, ma anche in nuovi areali nel mondo, come conseguenza della globalizzazione dei traffici commerciali e dei cambiamenti climatici [3].

Tradizionalmente, le misure di controllo sono indirizzate a limitare l'esposizione dell'ospite alle zanzare mediante sostanze repellenti [4], insetticidi [5] o trattamenti medici [6]. Tuttavia, sebbene l'uso di insetticidi sia e sia stato per lungo tempo il più importante componente delle strategie di controllo delle popolazioni di zanzare, le malattie tramesse da zanzare stanno riacutizzandosi, generalmente a causa sia di resistenza agli insetticidi sviluppata nei vettori, che di resistenza dei patogeni. Numerosi studi hanno dimostrato che complessi meccanismi biochimici e genetici (in particolare, aumentata detossificazione metabolica degli

insetticidi e diminuita affinità per proteine bersaglio) sono responsabili della resistenza agli insetticidi [7]. Sebbene gli insetticidi chimici siano estensivamente usati in tutti i paesi del mondo, un'aumentata consapevolezza sociale tende a sostituirli gradualmente con bioinsetticidi specifici, più sicuri per la salute umana. Si tratta di composti biologici efficaci, basati su tossine di origine microbica, con spiccata specificità per alcuni ordini di insetti [8]. Tuttavia, fenomeni di resistenza degli insetti anche ai bioinsetticidi sono stati recentemente riportati [9].

Il controllo degli insetti nocivi su base biologica rappresenta attualmente una delle maggiori sfide che la scienza e la società devono sostenere per far fronte a possibili emergenze, senza tuttavia disattendere il rispetto dell'ambiente. Nella seconda metà del XX secolo, sono state introdotte tecniche ecologicamente valide a seguito dell'intuizione della Tecnica dell'Insetto Sterile (SIT)[10] e dell'introduzione dei nemici naturali di specie nocive [11]. L'avvento poi delle metodologie del DNA ricombinante e dell'ingegneria genetica, accompagnate da maggiori conoscenze delle specie considerate, ha stimolato miglioramenti nell'applicazione di SIT.

Negli ultimi anni sono in atto ricerche innovative, mediante approcci di trasformazione genetica e di "gene editing", che sfruttano caratteristiche e fenomeni biologici, e tendono a ridurre drasticamente la potenzialità riproduttiva e la capacità vettoriale delle popolazioni [12; 13; 14; 15; 16].

La tecnica dell'insetto sterile (SIT)

La Tecnica dell'Insetto Sterile è un metodo di controllo autocidico che sfrutta il comportamento riproduttivo degli insetti. Nella sua applicazione, SIT consiste nel rilascio sistematico di insetti sterilizzati, preferibilmente di sesso maschile, allevati in idonee biofabbriche. I maschi sterilizzati con mezzi fisici competono con quelli naturali per le femmine della popolazione bersaglio; di conseguenza, rilasciando continuamente maschi sterilizzati, il potenziale riproduttivo della popolazione bersaglio viene ridotto per la mancata produzione di progenie [17], teoricamente fino ad eradicazione della specie (*Fig. 1*).

L'efficienza di SIT dipende dalla possibilità di a) allevare in modo massivo la specie bersaglio; b) sterilizzare con radiazioni ionizzanti (raggi X, raggi g, neutroni) un notevole numero di insetti della

specie bersaglio in un adatto stadio di sviluppo; c) rilasciare un notevole numero di insetti sterilizzati, ragionevolmente competitivi rispetto ai maschi selvatici; d) distribuire gli insetti sterilizzati con l'uso di sistemi economici nell'area da trattare; e) valutare accuratamente la dimensione delle popolazioni native *prima* e *dopo* il rilascio degli insetti sterilizzati; f) escludere l'eventuale immigrazione di femmine feconde nell'area trattata (che deve essere molto ampia e/o abbastanza isolata) [18].

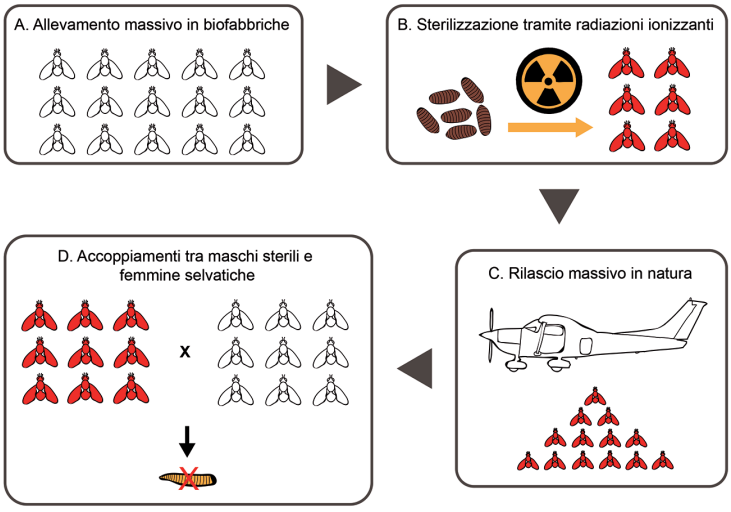


Fig. 1 – Applicazione della tecnica SIT (Tecnica dell'Insetto Sterile). A seguito di introduzione accidentale di una specie di insetto dannoso o di aumento della dimensione di una popolazione di una specie endemica, maschi sterilizzati della specie di interesse sono rilasciati in modo massivo nell'area bersaglio con opportuni mezzi di dispersione. I maschi sessualmente sterili, in numero superiore a quelli selvatici, si accoppiano con le femmine presenti nell'area, con il risultato che gli accoppiamenti non saranno fertili. Il risultato sarà un decremento della dimensione della popolazione che, se continuato nel tempo, porterà alla sua eradicazione.

Il concetto di SIT è stato sviluppato negli USA negli anni '30 del secolo scorso e la sua applicazione è stata realizzata con successo per la prima volta negli anni '50 contro *Cochliomyia hominivorax* (mosca assassina), dittero Calliphoridae, le cui larve si sviluppano come parassiti nei tessuti cutanei dei bovini, provocando pesanti miasi che causano perdite considerevoli all'allevamento di bestiame [10]. Il programma di eradicazione di questa specie mediante SIT è stato appli-

cato per la prima volta con successo nel 1954 nell'isola di Curacao. Con questa tecnica *C. hominivorax* è stata eradicata dagli USA a partire dal 1982 e dall'America Centrale dal 2001. Una barriera permanente di mosche sterili viene tutt'ora mantenuta nel Darien gap di Panama allo scopo di prevenire reinfestazioni della specie dai paesi dell'America meridionale.

Lo sviluppo di questa tecnologia è stato in gran parte guidato da USDA (United States Department of Agriculture) e dalla divisione congiunta FAO/IAEA (Food and Agriculture Organization of the United Nations - International Atomic Energy Agency), enti coinvolti a livello sovranazionale nel contenimento e/o eradicazione di numerose specie di insetti dannosi. SIT è stato applicato con successo su popolazioni di ditteri tefritidi (*Ceratitis capitata*, *Bactrocera dorsalis*, *Bactrocera cucurbitae*, *Bactrocera tryoni*, *Bactrocera oleae*, *Anastrepha suspensa*, *Anastrepha ludens*), di ditteri calliforidi (*Lucilia cuprina*), di ditteri glossinidi (mosche tse-tse), di ditteri culicidi (*Aedes aegypti*) e di lepidotteri delle fam. Lymantriidae (*Lymantrya dispar*) e Gelechiidae (*Pectinophora gossypiella*, tacna del cotone) [19].

Dal momento che solo i maschi sterili contribuiscono alla soppressione della popolazione, un limite di questa tecnica è rappresentato dal concomitante rilascio di femmine sterilizzate: sebbene le uova deposte da queste femmine non siano in grado di svilupparsi in embrioni, poiché danneggiate a livello di DNA dalle radiazioni, le femmine mantengono il proprio abituale comportamento. Nel caso di insetti fitofagi, le femmine depongono le uova nell'ospite e, nel caso di femmine ematofaghe, queste continuano a nutrirsi di sangue agendo così da vettori di patogeni. La presenza di femmine nell'allevamento di massa, inoltre, raddoppia il costo dell'allevamento e, poiché talvolta può realizzarsi una sterilizzazione incompleta, sussiste il rischio di immettere nell'ambiente femmine fertili che potrebbero ridurre in modo significativo l'efficenza della tecnica SIT [20].

Sebbene la tecnica SIT sia stata spesso applicata utilizzando mosche sterilizzate di ambedue i sessi, attualmente si tende ad impiegare solamente ceppi di soli maschi ("Genetic Sexing Strains", GSSs) ottenuti mediante un approccio di genetica classica: questi ceppi presentano traslocazioni reciproche tra il cromosoma Y (tipico del sesso maschile) e un autosoma che porta un marcatore recessivo selezionabile. In questo modo i due sessi vengono discriminati in stadi precoci dello sviluppo con conseguente realizzazione di GSSs da immettere in

natura [21]. E'noto che le radiazioni ionizzanti che portano a sterilizzazione tendono comunque ad abbassare la fitness dei maschi rilasciati; in base alla dimensione della popolazione da trattare con SIT, il rapporto dei maschi sterilizzati rispetto a quelli selvatici dovrebbe essere nel range di 1:2 – 1:15.

Le ricerche attualmente in atto sono intese a rendere la tecnica SIT più efficiente attraverso la generazione di ceppi di soli maschi mediante manipolazioni di genetica molecolare, e lo sfruttamento di particolari fenomeni biologici, anche quelli indotti negli insetti da microrganismi.

La tecnica RIDL (release of insects carrying dominant lethals)

Questa tecnica, descritta inizialmente in un lavoro condotto su *Drosophila melanogaster*, si basa sull'inserzione nel genoma dei maschi della specie di interesse di un gene letale dominante, mediante tecniche di biologia molecolare [22]. Nella formulazione originaria, il gene letale produce una proteina, chiamata tTAV (tetracycline transactivater variant), che si lega a un sito specifico nel genoma, chiamato TeTO (tetracycline operator), producendo una proteina tossica. L'accoppiamento della femmina con il maschio che porta il gene letale dominante (tTAV) determina così la morte della progenie nei primi stadi dello sviluppo. Se però gli insetti vengono allevati in presenza di tetraciclina (antidoto) assunta con il cibo, questo antibiotico si lega con la proteina tTAV inattivando il legame con TeTO, permettendo così la produzione di maschi e femmine vitali (condizione permissiva). Il gene letale dominante, e di conseguenza il sistema, è specifico per ciascuna specie di insetto di interesse [23]. Questa tecnica è stata sviluppata e brevettata da una ditta privata (Oxitec Ltd, Oxford Insect Technologies).

La condizione permissiva è assente in natura, e pertanto gli insetti omozigoti per il gene dominante letale, accoppiandosi con gli insetti selvatici, generano una progenie eterozigote che muore [24]. Il rilascio periodico di maschi con gene letale rappresenta un sistema efficiente di soppressione delle popolazioni di insetti basato su SIT. Rispetto alla tecnica classica SIT, RIDL ha il vantaggio che la sterilizzazione non richiede radiazioni ionizzanti e pertanto la competitività dei maschi è

mantenuta [25]. Tuttavia, il controllo in campo con il sistema RIDL contro *Aedes aegypti*, condotto con successo in diverse aree dove Dengue è endemica (Isole Cayman, Malesia e Brasile), ha richiesto per sicurezza la sterilizzazione dei maschi prima del rilascio in modo tale che il gene modificato non potesse diffondersi fra le generazioni successive [15; 26].

Controllo biologico mediante incompatibilità citoplasmatica (CI)

Si tratta di un fenomeno indotto da un gruppo di batteri gram-negativi del genere *Wolbachia*, trovati come endosimbionti nei tessuti riproduttivi degli Artropodi e di altri organismi [27]. Questi batteri sono in grado di manipolare, con vari meccanismi, la riproduzione dei loro ospiti, inducendo fenomeni quali incompatibilità riproduttiva (CI), partenogenesi e femminilizzazione dei maschi. L'incompatibilità citoplasmatica (CI) rappresenta il fenomeno più interessante per strategie di controllo dal momento che questa caratteristica viene ereditata per via materna con modalità uni- o bi-direzionale (*Fig. 2*). Nella incompatibilità uni-direzionale, l'incrocio tra un maschio infetto con *Wolbachia* e una femmina non-infetta dà origine ad uno zigote non-vitale. L'incrocio reciproco è invece fertile. *Wolbachia* induce incompatibilità nei maschi, mentre nelle femmine ristabilisce la compatibilità. Nella incompatibilità bi-direzionale, dove maschi e femmine portano differenti ceppi di *Wolbachia*, tutti gli incroci eterozigoti sono incompatibili. Essendo stato dimostrato che *Wolbachia* è in grado di diffondersi con estrema rapidità in popolazioni non-infette [28], questa caratteristica aggiunge un'ulteriore vantaggio nei programmi di controllo mediato da questo simbionte.

Oltre ad agire come un vero manipolatore della riproduzione dell'ospite, *Wolbachia* presenta alcuni ceppi in grado anche di inibire la replicazione di virus e protozoi nelle zanzare [29; 30; 31]. Queste condizioni rappresentano un potente mezzo per il controllo di popolazioni infette di specie di *Aedes* [32] e di *Anopheles* [33].

Con particolari tecniche biotecnologiche è possibile trasferire ceppi di *Wolbachia* da una specie all'altra, generando nuovi pattern di CI ed estendendo l'elenco di specie da controllare con strategie basate su *Wolbachia*. La tecnologia usata allo scopo si basa su microiniezione

di ooplasma infetto (transinfezione) contente l'appropriato ceppo di *Wolbachia* in embrioni recipienti [34]. Questa strategia di controllo è stata recentemente usata con successo contro popolazioni di *Aedes aegypti* in Australia e Vietnam [15], e prove di campo sono state condotte in Italia contro *Aedes albopictus* [35].

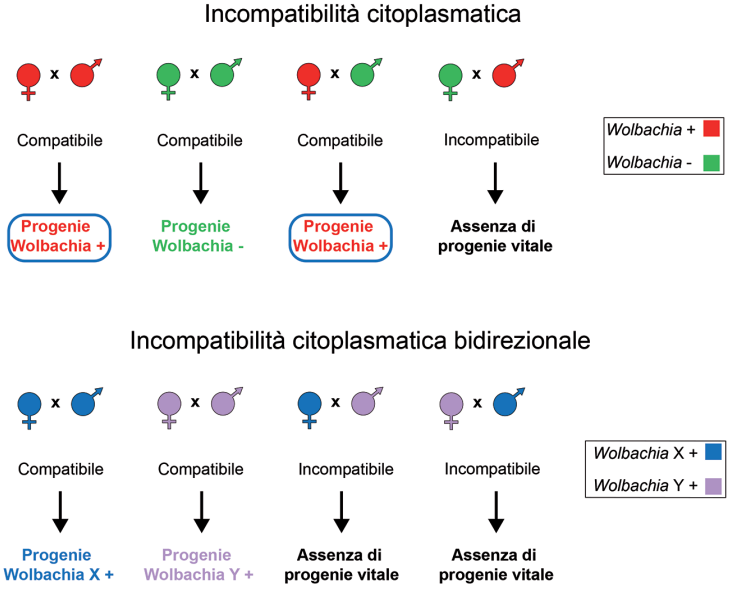


Fig. 2 – L'incompatibilità citoplasmatica (CI) è un fenomeno che impedisce agli spermatozoi e alle uova di formare una prole vitale. L'effetto deriva dai cambiamenti nelle cellule del gamete causati da Wolbachia. *a) CI uni-direzionale si realizza con femmine in grado di accoppiarsi con successo sia con maschi non-infetti che con maschi infetti con lo stesso ceppo di* Wolbachia*; b) CI bi-direzionale può realizzarsi in maschi e femmine infetti con due differenti ceppi di* Wolbachia*, incapaci di generare progenie fertile.*

Tecnologie di ingegneria genetica: gene drive

A differenza delle strategie soprariportate che necessitano di introduzioni continue di maschi sterili nelle popolazioni bersaglio, un promettente approccio di manipolazione genetica, il cosidetto "gene-drive", ha recentemente dimostrato la possibilità di selezionare alcune

caratteristiche ereditarie, nella fattispecie quelle che rendono sterili le femmine di zanzare o che ne riducono le possibilità di accoppiamento, o ancora che le rendono immuni al patogeno, e far sì che queste caratteristiche si diffondano rapidamente alle generazioni successive [36; 37]. In realtà "gene drive" è un fenomeno naturale, dovuto ad elementi genetici in grado di manipolare la gametogenesi e la riproduzione, e in modo tale da aumentare la propria trasmissione alle generazioni successive, fino a quando questi particolari elementi genetici non siano presenti in ogni discendente [38]. Questo fenomeno deriva da una varietà di meccanismi utilizzati nelle manipolazioni genetiche della linea germinale per far propagare una serie di geni in una popolazione, alterando la probabilità che un allele specifico venga trasmesso alla progenie dalla naturale probabilità del 50% (distorsione del sesso). Sistemi di gene drive altamente invasivi possono determinare una deviazione ereditaria vicino al 100%. Infatti, assumendo una popolazione con accoppiamenti random e senza costo di fitness, un elemento gene drive con 100% di trasmissione alla progenie aumenterebbe la sua frequenza fino a 0.99 dopo soltanto poche generazioni [39; 40].

Elementi di gene drive in natura sono rappresentati da elementi trasponibili (TEs), da microrganismi ereditabili e da endonucleasi di riferimento (homing endonucleases, HEGs) che tendono a diffondersi in una popolazione bersaglio per più generazioni, anche quando sono rilasciati a basse frequenze [14]. In particolare, HEGs si sono rivelate uno strumento ideale per distruggere geni essenziali per la fertilità della femmina, per la determinazione del sesso, per il riconoscimento dell'ospite o per la trasmissione del patogeno (knock-out scheme) e per diffondere modificazioni deleterie che innescano la soppressione della popolazione (knock-in scheme). HEGs sono localizzate entro la sequenza di riconoscimento del DNA, che lo rende resistente al taglio. Quando HEG viene in contatto con il suo cromosoma omologo intatto in cellule eterozigoti, esso taglia la sequenza bersaglio. Come risultato del processo di riparo del cromosoma (homology direct repair mechanism, HDR), indotto dalla rottura del doppio filamento, HEG può essere copiato convertendo le cellule eterozigoti in omozigoti. Questo meccanismo, noto come "homing", promuove un rapido aumento di frequenza di HEG in una popolazione [14].

Nell'ultimo decennio sono state proposti diversi sistemi artificiali di homing/gene drive basati su nucleasi modulari, di cui uno dei più promettenti è stato realizzato mediante editing genico CRISPR-Cas9.

Questo sistema di gene drive sintetico è stato designato in modo tale da essere attivo nella linea germinale di *Anopheles gambiae*, e in modo da essere capace di distruggere geni essenziali per la fertilità della femmina, come il gene di fertilità AGAP007280 [41]. In gene drive mediato da CRISPR, il costrutto comprende una singola nucleasi Cas9 e una guida di RNA che determina la specificità del taglio del sito bersaglio. Quando inserito nel genoma ed espresso nella linea germinale, la proteina Cas9 distrugge il gene AGAP007280. In risposta alla rottura del doppio filamento, la cellula impiega il sistema HDR (homology direct repair mechanism) per copiare il costrutto che include qualunque carico genetico nel cromosoma omologo intatto. Sebbene singole femmine di zanzara con una sola copia del gene difettivo sono ancora fertili, il sistema gene drive assicura l'ereditarietà della mutazione nel 99-95 % della progenie ed un rapido aumento di frequenza di AGAP007280 nella popolazione. Come risultato, la progenie ha un'alta probabilità di ricevere una copia del costrutto gene drive da ambedue i genitori: tutte le femmine omozigoti per la mutazione sono sterili, mentre i maschi continuano a trasmettere il gene drive. In tal modo, sono richieste poche generazioni per diffondere il tansgene/mutazione nelle popolazioni selvatiche a partire da bassa frequenza fino a raggiungere quasi la fissazione (*Fig. 3*).

Questa tecnologia ha le potenzialità per indurre un sostenibile decremento di densità di zanzare imponendo un carico riproduttivo sulla popolazione, anche se è stato osservata forte pressione selettiva per alleli resistenti come variazioni di sequenza al sito bersaglio di endonucleasi Cas9. Tuttavia, costrutti sintetici gene drive indirizzati a sequenze altamente conservate nel genoma decrescono la probabilità di resistenza [42].

L'eliminazione del gene *doublesex* (*Agdsx*) mediante il sistema CRISPR-Cas9 ha portato per la prima volta alla soppressione completa si popolazioni di *Anopheles gambiae* tenute in piccole gabbie dopo 7-11 generazioni (Andrea Crisanti, comunicazione personale). Il gene *Agdxs* è altamente conservato e codifica due trascritti alternativamente espressi, *Agdsx*F (dsx-femminile) e *Agdsx*M (dsx-maschile). Questo gene determina il dimorfismo sessuale nelle zanzare. L'eliminazione mediata da CRISPR-Cas9 della sequenza tra l'introne 4 e l'esone 5 in *Agdsx* inibisce la formazione del trascritto funzionale *Agdsx*F e induce sterilità nelle femmine, ma non influenza la fitness dei maschi. Femmine omozigoti nulle ($dsxF^{-/-}$) mostrano intersessi con caratteristiche maschio-

specifiche, quali antenne piumose, proboscide anormale e pterigopodi non ruotati. Queste femmine sono incapaci di pungere, di nutrirsi su sangue e di deporre uova. Dati i limiti funzionali del *Agdxs*, dopo 14 generazioni di zanzare tenute in piccole gabbie non è stata osservata selezione di alleli resistenti. Esperimenti in grandi gabbie che riproducano condizioni di ambienti tropicali sono in atto presso l'Imperial College London e il Polo d'Innovazione di Genomica Genetica e Biologia (sede di Terni) con lo scopo di valutare la dinamica di popolazione e il potenziale di soppressione di questo costrutto di gene drive sviluppato in zanzare del genere *Anopheles*.

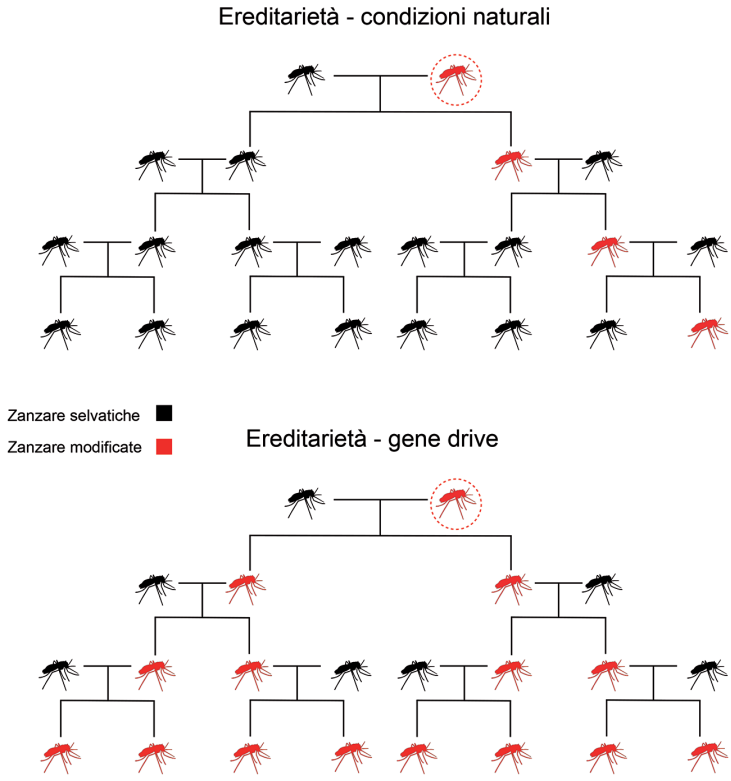


Fig. 3 – Le zanzare e altri organismi trasmettono un dato carattere alle generazioni successive con probabilità del 50% (eredità mendeliana), mentre un sistema gene drive copia e inserisce se stesso nei cromosomi di ambedue i genitori, assicurando una probabilità della sua trasmissione fino al 100%.

Conclusioni

Le strategie di controllo genetico sfruttano il comportamento riproduttivo degli insetti, attraverso l'accoppiamento, con l'uso di ceppi di soli maschi capaci di disperdersi attivamente nell'ambiente alla ricerca delle femmine selvatiche. Le strategie genetiche sono tipicamente specie-specifiche. Il controllo genetico non deve tuttavia essere visto come la panacea capace di eliminare da solo popolazioni di insetti nocivi ma, opportunamente integrato con tutti i metodi convenzionali, potrebbe portare a risultati efficienti.

Mentre alcune delle strategie, quali il classico SIT, sono applicate in campo in modo regolare a seguito di approvazione delle normative regolatorie vigenti [43], tutte quelle che fanno uso di insetti geneticamente modificati rispettano al momento il principio di precauzione. Alcuni Paesi (Isole Cayman, Malesia e Brasile, Australia e Vietnam) hanno dato autorizzazioni per usare la tecnica RIDL modificata o SIT integrata con *Wolbachia* [13; 24], mentre i ceppi di soli maschi generati mediante le promettenti manipolazioni di gene editing sono al momento trattati in ambienti confinati (Andrea Crisanti, comunicazione personale).

L'uso di insetti geneticamente manipolati da immettere in natura per il controllo/eradicazione di specie nocive, è argomento di opportune normative [44; 45; 46] a fronte di un forte dibattito a livello nazionale e internazionale relativamente alle tematiche della protezione dell'ambiente e della salute, nonostante ci sia ampio consenso in ambito scientifico nel ritenere che gli insetti geneticamente modificati non presentino rischi all'ambiente [47].

Ringraziamenti

Si ringrazia la Dr Francesca Scolari per la realizzazione delle figure incluse in questo scritto.

BIBLIOGRAFIA

1. Hotez P.J., Fenwick A., Savioli L., Molyneux D.H. Rescuing the bottom billion through control of neglected tropical diseases. Lancet (2009) 373: 1570-1575.
2. Mayer S.V., Tesh R.B., Vasilakis N. The emergence of arthropod-borne viral diseases: a global prospective on dengue, chikungunya and zika fevers. Acta Trop. (2017) 166: 155-163.
3. Reiter P. Climate change and mosquito-borne disease. Environ. Health Persp. (2001) 109(suppl 1): 141-161. doi:10.1289/ehp.01109s1141
4. Paluch G., Bartholomay L., Coats J. Mosquito repellents: a review of chemical structure diversity and olfaction. Pest Manag. Sci. (2010) 66: 925-935.
5. Rivero A., Vézilier J., Weill M., Read A.F., Gandon S. Insecticide Control of Vector-Borne Diseases: When Is Insecticide Resistance a Problem? PLoS Path (2010) 6: e1001000. doi:10.1371/journal.ppat.1001000
6. Utzinger J., Becker S., Knopp S., Blum J. Neumayr A., Keiser J., Hatz C. Neglected tropical diseases: diagnosis, clinical management, treatment and control. Swiss Medical Weekly (2012) doi:10.4414/smw.2012.13727
7. Liu N. Insecticide Resistance in Mosquitoes: Impact, Mechanisms, and Research Directions. Annu. Rev. Entomol (2015) 60: 537–559. doi:10.1146/annurev-ento-010814-020828
8. Mazid S., Kalita J.C., Rajkhowa R.C. A review on the use of biopesticides in insect pest management. Int. J. Sci. Adv. Technol. (2011) 1: 169-178.
9. Tabashnik B.E., Brévault T., Carrière Y. Insect resistance to Bt crops: lessons from the first billion acres. Nature Biotech. (2013): 31: 510-521. doi:10.1038/nbt.2597
10. Knipling E. F. Possibilities of insect control or eradication through the use of sexually sterile males. J. Econ. Entomol. (1955) 48: 459-462.doi:10.1093/jee/48.4.459
11. Barratt B.I.P., Moran V.C., Bigler F., van Lenteren J.C. The status of biological control and recommendations for improving uptake for the future. BioControl (2017) 63: 155-167. doi:10.1007/s10526-017-9831-y
12. Scolari F., Siciliano P., Gabrieli P., Gomulski L. M., Bonomi A., Gasperi G., Malacrida A.R. Safe and fit genetically modified insects for pest control: from lab to field applications. Genetica (2011) 139: 41–52. doi: 10.1007/s10709-010-9483-7
13. Scolari F., Gomulski L.M., Gabrieli P., Manni M., Savini G., Gasperi G., Malacrida A.R.. How the 'omics' era will impact fruit fly pest control: the example of the Mediterranean fruit fly, *Ceratitis capitata*. BMC Genetics (2014) 15 (Suppl 2): S11.
14. Burt A. Site-specific selfish genes as tools for the control and genetic engineering of natural populations. Proc. Royal Soc. B: Biol. Sci. (2003) 270: 921–928. doi:10.1098/rspb.2002.2319
15. Alphey L. Genetic Control of Mosquitoes. Annu. Rev. Entomol. (2014) 59: 205-224. doi:10.1146/annurev-ento-011613-162002

16. Galizi R., Hammond A., Kyrou K., Taxiarchi C., Bernardini F., O'Loughlin S.M., Papathanos P.A., Nolan T., Windbichler N., Crisanti A. A CRISPR-Cas9 sex-ratio distortion system for genetic control. Sci. Rep. (2016) 6: 31139. doi:10.1038/srep31139
17. Hendrichs J., Robinson A. Sterile Insect Technique. pp. 953-957. In Encyclopedia of Insects (Resh V., Cardé R. Eds.) pp. 1168. Academic Press, Elsevier Publ. (2009). doi:10.1016/b978-0-12-374144-8.00252-6
18. Vreysen M.J.B., Robinson A.S., Hendrichs J. (Eds.) Area-Wide Control of Insect Pests: From Research to Field Implementation. pp 789. Springer Publ., 3300 AA Dordrecht, The Netherlands (2007).
19. Pereira R., Yuval B., Liedo P. Teal P.E.A., Shelly T.E., McInnis D.O., Hendrichs J. Improving sterile male performance in support of programmes integrating the sterile insect technique against fruit flies. J. Appl. Entomol (2011) 137: 178-190. doi:10.1111/j.14390418.2011.01664.x
20. McInnis D.O., Tam S., Grace C., Miyashita D. Population suppression and sterility rates induced by variable sex ratio, sterile insect releases of *Ceratitis capitata* (Diptera: Tephritidae) in Hawaii. Ann. Entomol. Soc. Am. (1994) 87: 231-240. doi:10.1093/aesa/87.2.231
21. Franz G. Genetic sexing strains in Mediterranean fruit fly, an example for other species amenable to large-scale rearing for the sterile insect technique. pp. 427-451. In: Sterile Insect Technique. Principles and Practice in Area-Wide Integrated Pest Management (Dyck V.A., Hendrichs J., Robinson A.S. Eds.) pp. 787. Springer Publ., The Netherlands (2005).
22. Thomas D.D., Donnelly C.A., Wood R.J., Alphey L.S. Insect Population Control Using a Dominant, Repressible, Lethal Genetic System. Science (2000) 287: 2474-2476. doi:10.1126/science.287.5462.2474
23. Harvey-Samuel T., Ant T., Alphey L. Towards the genetic control of invasive species. Biol. Invasions (2017) 19: 1683-1703. doi:10.1007/s10530-017-1384-6
24. Gong P., Epton M.J., Fu G., Scaife S., Hiscox A., Condon K. C., Condon G.C., Morrison N.I., Kelly D.V., Dafa'alla T., Coleman P.G., Alphey L. A dominant lethal genetic system for autocidal control of the Mediterranean fruitfly. Nature Biotech. (2005) 23: 453-456.
25. Black W.C., Alphey L., James A.A. Why RIDL is not SIT. Trends Parasitol. (2011) 27: 362–370. doi:10.1016/j.pt.2011.04.004
26. Carvalho D.O., McKemey A.R., Garziera L., Lacroix R., Donnelly C.A., Alphey L., Malavasi A., Capurro M.L. Suppression of a field population of Aedes aegypti in Brazil by sustained release of transgenic male mosquitoes. PLoS Negl. Trop. Dis. (2015) 9: e0003864. doi:10.1371/journal.pntd.0003864
27. Werren J.H., Baldo L., Clark M.E. *Wolbachia*: master manipulators of invertebrate biology. Nat. Rev Microbiol. (2008) 6: 741-751.
28. Turelli M. Evolution of incompatibility inducing microbes and their hosts. Evolution (1994) 48: 1500-1513.
29. Moreira L.A., Iturbe-Ormaetxe I., Jeffery J.A., Lu G., Pyke A.T., Hedges L.M., Rocha B.C., Hall-Mendelin S., Day A., Riegler M., Hugo L.E., Johnson K.N.,

Kay B.H., McGraw E.A., van den Hurk A.F., Ryan P.A., O'Neill S.L. A *Wolbachia* symbiont in *Aedes aegypti* limits infection with Dengue, Chikungunya, and *Plasmodium.* Cell (2009) 139: 1268-1278. doi:10.1016/j.cell.2009.11.042

30. Walker T., Johnson P.H., Moreira L.A., Iturbe-Ormaetxe I., Frentiu F.D., McMeniman C.J., Leong Y.S., Dong Y., Axford J., Kriesner P., Lloyd A.L., Richie S.A., O'Neill S.L., Hoffmann A.A. The *w*Mel *Wolbachia* strain blocks dengue and invades caged *Aedes aegypti* populations. Nature (2011) 476: 450-453. doi:10.1038/nature10355
31. Sinkins S.P. *Wolbachia* and arbovirus inhibition in mosquitoes. Future Microbiol. (2013) 8: 1249-1256.
32. Iturbe-Ormaetxe I., Walker T., O' Neill S.L. *Wolbachia* and the biological control of mosquito-borne disease. EMBO *rep* (2011) 12: 508-518. doi:10.1038/embor.2011.84.
33. Bian G., Joshi D., Dong Y., Lu P., Zhou G., Pan X., Xu Y., Dimopoulos G., Xi Z *Wolbachia* invades *Anopheles stephensi* populations and induces refractoriness to *Plasmodium* infection. Science (2013) 340:748-751. doi: 10.1126/science.1236192
34. Hughes G.L., Rasgon J.L. Transinfection: a method to investigate*Wolbachia*-host interactions and control arthropod-borne disease. Insect Mol. Biol. (2013) 23: 141-151. doi:10.1111/imb.12066
35. Puggioli A., Calvitti M., Moretti R., Bellini R. *w*Pip *Wolbachia* contribution to *Aedes albopictus* SIT performance: advantages under intensive rearing. Acta Trop. (2016) 164: 473-481. doi.org/10.1016/j.actatropica.2016.10.014
36. Windbichler N., Papathanos P.A., Crisanti A. Targeting the X chromosome during spermatogenesis induces Y chromosome transmission ratio distortion and early dominant embryo lethality in *Anopheles gambiae.* PLoS Genet. (2008) 4: e1000291. doi:10.1371/journal.pgen.1000291
37. Champer J., Buchman A., Akbari O.S. Cheating evolution: engineering gene drives to manipulate the fate of wild populations. Nature Rev. Genet. (2016) 17: 146-159. doi:10.1038/nrg.2015.34
38. Sandler L., Novitski E. Meiotic Drive as an Evolutionary Force. *Am. Nat.* (1957) 91: 105-110. doi:10.1086/281969.
39. Curtis C.F. Possible use of translocations to fix desirable genes in insect pest populations. Nature (1968) 218: 368-369.
40. Akbari O.S., Bellen H.J., Bier E., Bullock S.L., Burt A., Church G.M., Cook K.R., Duchek P., Edwards O.R., Esvelt K.M., Gantz V.M., Golic K.G., Gratz S.J., Harrison M.M., Hayes K.R., James A.A., Kaufman T.C., Knoblich J., Malik H.S., Matthews K.A., O'Connor-Giles K.M., Parks A.L., Perrimon N., Port F., Russell S., Ueda R., Wildonger J. Safeguarding gene drive experiments in the laboratory. Science (2015) 349: 927–929. doi:10.1126/science.aac7932
41. Hammond A., Galizi R., Kyrou K., Simoni A., Siniscalchi C., Katsanos D., Gribble M., Baker D., Marois E., Russell S., Burt A., Windbichler N., Crisanti A., Nolan T.A. CRISPR-Cas9 gene drive system targeting female reproduction

in the malaria mosquito vector *Anopheles gambiae*. Nature Biotech. (2016) 34: 78-83. doi:10.1038/nbt.3439

42. Hammond A.M., Kyrou K., Bruttini M., North A., Galizi R., Karlsson X., Kranjc N., Carpi F.M., D'Aurizio R., Crisanti A., Nolan T. The creation and selection of mutations resistant to a gene drive over multiple generations in the malaria mosquito. PLoS Genet. (2017) 13: e1007039. doi: 10.1371/journal.pgen.1007039.
43. IAEA (International Atomic Energy Agency) Sterile insect technique. https://www.iaea.org/topics/sterile-insect-technique, Vienna, Austria
44. Marshall J.M. The Cartagena Protocol and genetically modified mosquitoes. Nature Biotech. (2010) 28: 896-89.
45. Reeves R.G., Denton J.A., Santucci F., Bryk J., Reed F.A. Scientific standards and the regulation of genetically modified Iinsects. PLoS Negl Trop Dis (2012) 6: e1502. doi:10.1371/journal.pntd.0001502
46. WHO-TDR, FNIH. Guidance framework for testing of genetically modified mosquitoes. pp. 132. World Health Organization, Geneva (2014).
47. Reis-Castro L. Genetically modified insects as a public health tool: discussing the different bio-objectification within genetic strategies. Croatian Med. J. (2012) 53: 635-638. doi:10.3325/cmj.2012.53.635

Scienze naturali
Ecologia
Istituto Lombardo (Rend. Scienze) 151, 149-170 (2017)

BRONI E CASALE MONFERRATO: DUE COMUNITÀ DI FRONTE ALL'AMIANTO TRA MEMORIA, RIMOZIONE E RICONOSCIMENTO DEL DANNO

BRUNO ZIGLIOLI (*)

Nota presentata dal m.e. Marino Gatto
(Adunanza del 29 giugno 2017)

SUNTO. – L'articolo compara i *case study* degli stabilimenti di cemento-amianto "Eternit" di Casale Monferrato e "Fibronit" di Broni. Le due cittadine sono state investite dal dramma della contaminazione da amianto, che continuerà a colpire ancora per molto tempo i lavoratori e gli abitanti. A Casale, la memoria pubblica della fabbrica e delle sue conseguenze sanitarie si è cristallizzata in modo consapevole, tanto da rendere la vicenda della cittadina emblematica sul piano nazionale e internazionale, mentre a Broni si è verificata una profonda rimozione della vicenda. Dal quadro descritto, emerge con evidenza come le lacerazioni all'interno delle comunità si aprano in un ventaglio di "visioni cittadine" che vanno dalla "ipervisibilizzazione" del problema – la fabbrica come onnipresente causa di tutti i mali della città – alla sua completa rimozione dall'orizzonte urbano, rendendo molto difficoltosi gli interventi dei pubblici poteri.

***

ABSTRACT. – The article compares the case of the Eternit cement-asbestos factory in Casale Monferrato to that of the Fibronit factory in Broni. The two towns were assailed by the drama of asbestos contamination, which will continue to affect workers and local inhabitants for a long time to come. In Casale, the memory of the factory and its consequences on the health of the inhabitants has taken the form of a concrete awareness,

(*) Professore associato di Storia contemporanea, Dipartimento di Scienze Politiche e Sociali, Università degli Studi di Pavia, Italy.
E-mail: bruno.ziglioli@unipv.it

so much so that Casale has become an emblematic town on both national and international levels. However, the same cannot be said for Broni: here there has been a radical repression of the problem. This paper will demonstrate how tensions within the community open up a range of citizens'viewpoints, which span from a hyper-visualization of the problem – the factory as an omnipresent cause of all the evils of the city – to its complete removal from the urban horizon, making any kind of intervention on public authorities extremely difficult.

1. La fibra miracolosa

L'amianto è un silicato a struttura fibrosa, che si trova in natura frammisto ad altri minerali (la cosiddetta "rocciamadre"), dai quali deve essere separato mediante macinazione. Ne esistono molte varietà: la più diffusa, che costituisce circa il 95% del minerale estratto nel mondo, è del tipo crisotilo, o "amianto bianco"; ve ne sono poi altre con un potenziale cancerogeno molto superiore, come la crocidolite (o "amianto blu"), l'amosite, l'antofillite e la tremolite. Si tratta di un materiale a basso costo ma dalle grandi proprietà chimico-fisiche, conosciute fin dall'antichità: risulta resistente al calore e al fuoco, all'usura, all'azione di agenti chimici e biologici, ed è fonoassorbente.

Nel corso dell'età moderna, in una fase ancora preindustriale, l'amianto venne impiegato soprattutto nel settore tessile. Una data convenzionale di inizio dell'impiego industriale dell'asbesto in epoca contemporanea può essere fissata nel 1883, quando sulla rivista inglese *The Engineer* comparve un articolo intitolato *Asbestos and its Applications*, nel quale si descrive l'uso del materiale per isolare i motori a vapore del naviglio delle marine militari britannica e tedesca: tale utilizzo si diffuse tra le flotte militari e civili di tutto il mondo a cavallo della prima guerra mondiale. Nel frattempo, nel 1901, l'austriaco Ludwig Hatschek aveva brevettato una formula per un impasto di cemento con fibre di amianto che presentava caratteristiche eccezionali di leggerezza, resistenza alle temperature e al fuoco e flessibilità di impiego: proprio per sottolineare tali qualità, il prodotto era stato ribattezzato "Eternit", mentre il brevetto svizzero divenne di proprietà dell'industriale Alois Steinmann e poi, dal 1920, della famiglia Schmidheiny, che in breve tempo realizzò un colosso industriale e multinazionale proprio incentrato sulla produzione di cemento-amianto. Le si affiancarono altri tre grandi gruppi mondiali: la Johns-Manville negli Stati Uniti, la Turner & Newall in

Gran Bretagna, la Cape Asbestos, britannica ma sostanzialmente radicata in Sudafrica (v. [1], pagg. 46-47 e 304-310).

Negli anni del secondo dopoguerra l'impiego del minerale e dei suoi derivati e composti si fece massiccio in moltissimi settori: nell'edilizia, nella navalmeccanica e nella cantieristica, nell'industria elettrica, chimica, aeronautica, automobilistica, ferroviaria e tessile, quale materiale isolante e di coibentazione. Alla metà degli assi settanta, l'uso di amianto in Italia e nel mondo industrializzato era massiccio, anzi: tra il 1976 e il 1980 si sarebbe toccato il culmine in termini di estrazione, produzione e consumo, in una parabola ascendente iniziata con il cosiddetto miracolo economico e industriale italiano. Non vi era quasi settore produttivo che non utilizzasse amianto nei suoi cicli di lavorazione, e nell'edilizia – residenziale, commerciale o industriale che fosse – l'amianto veniva impiegato, in particolare sotto forma di fibrocemento, nei tetti e nelle coperture, negli isolamenti termici e acustici, nei prefabbricati, nelle controsoffittature, negli intonaci, nelle canne fumarie, nelle tubazioni di abitazioni e acquedotti, nelle canalette per impianti elettrici e nelle barriere antifiamma di case e capannoni ma anche di mezzi di trasporto (navi, treni, aerei), così come nei freni delle automobili e in molti altri ambiti. Circa il 40% degli edifici costruiti in Italia tra il 1967 e il 1975 risulta trattato con il «cemento armato dei poveri» (v. [2], pagg. 146-147).

L'impiego dell'asbesto si estese anche a oggetti di uso comune. Neppure i giochi per i bambini furono risparmiati dall'utilizzo del minerale: per esempio, una ricerca dell'Istituto per lo studio e la prevenzione oncologica dell'Azienda sanitaria e dell'Università di Firenze ha evidenziato come, tra il 1963 e il 1976, nella composizione della celebre pasta per modellare Das fosse presente amianto per il 30% circa del totale (v. [3], pagg. 80-85).

Tuttavia, parallelamente al vertiginoso aumento della produzione (v. [4], pag. 24), cominciarono a emergere le prime notizie sui pericoli connessi all'esposizione all'amianto. I risultati della prima ricerca medica rigorosa sulle malattie connesse all'asbesto, condotta dal medico inglese William E. Cooke, furono pubblicati nel luglio del 1924 dal *The British Medical Journal*, e individuarono una patologia a cui i lavoratori erano esposti: l'asbestosi [5]. Si tratta di una patologia cronica e progressiva da accumulo, dovuta all'inalazione delle fibre soprattutto da parte degli operai, che respiravano per molte ore al giorno, per molti anni della loro vita, la polvere di amianto. L'asbestosi comporta fenomeni di enfi-

sema, di ispessimento della pleura e di insufficienza respiratoria e cardiorespiratoria, anche a distanza di molti anni dall'esposizione, e può avere esito mortale dopo un lungo decorso. Il primo paese a includere tale patologia nella lista delle malattie professionali fu la Germania hitleriana nel 1936, seguita dall'Italia nel 1943 e dalla Francia nel 1945. Tale riconoscimento risultò più distanziato per la Svizzera (1953), per l'Austria (1955) e per il Belgio (nel 1969).

Sempre in Germania, nel 1938, il medico Martin Nordmann descrisse alcuni casi di cancro professionale degli operai dell'amianto: a seguito di questa ricerca, il tumore polmonare fu inscritto a sua volta nel registro delle malattie professionali. Tuttavia ci vollero molti anni affinché la relazione tra l'esposizione all'amianto e l'apparizione di un mesotelioma pleurico potesse essere scientificamente provata, grazie a inequivocabili studi compiuti in Sudafrica, negli Stati Uniti e altrove tra l'inizio degli anni Sessanta e la metà degli anni Settanta (v. [6], pp. 49-58). Il mesotelioma è un cancro della pleura la cui insorgenza non è legata necessariamente a un'esposizione continuata di tipo professionale, ma anche a esposizioni più sporadiche e limitate nel tempo. Per tale ragione, non risultano a rischio solo i lavoratori che operano a contatto con l'amianto, ma anche chi vive in prossimità degli impianti, chiunque venga a contatto con fibre e manufatti, i famigliari degli operai: in altre parole, l'intera comunità. Presenta una incubazione lunghissima, superiore ai vent'anni, in media dai quaranta ai cinquant'anni: a quasi trent'anni anni dalla messa al bando della lavorazione dell'asbesto, gli epidemiologi evidenziano che nel nostro Paese il picco della patologia verrà raggiunto intorno al 2025. Saranno gli ammalati del boom dell'amianto degli anni Settanta.

L'evidenza scientifica della correlazione asbesto-mesotelioma risultava acquisita già nella prima metà degli anni Settanta. Ciononostante, la lavorazione e l'utilizzo del minerale proseguì ancora per molti anni, prima che i lavoratori, i sindacati, le comunità che ospitavano gli impianti e più in generale l'opinione pubblica prendessero coscienza del grande pericolo connesso a quel genere di produzione, e prima che le legislazioni nazionali e sovranazionali intervenissero per limitarla o vietarla. I primi paesi a dotarsi di provvedimenti in tal senso sono stati la Svezia e la Norvegia, alla fine degli anni Settanta. Nel resto dell'Europa occidentale occorrerà attendere la prima metà degli anni Novanta: in Italia il bando alla lavorazione e all'utilizzo dell'amianto fu sancito con la legge 27 marzo 1992 n. 257, dopo una difficile battaglia

parlamentare, durata 5 anni. Molti ritardi e notevoli ambiguità si sono riscontrati sul piano comunitario: la Commissione europea ha deliberato la messa al bando dell'amianto nel 1999, e in seguito il provvedimento – entrato formalmente in vigore solo il 1° gennaio 2005 – è stato parzialmente derogato nel 2009. Il Canada, per molti decenni il principale estrattore di amianto al mondo, ha interrotto l'attività delle sue miniere nel 2011. Giappone, Corea del Sud e Hong Kong hanno optato per la messa al bando completa nel 2012. In Russia (principale esportatore), India (principale importatore), Indonesia, Cina, Brasile, Messico e nella maggior parte dei paesi dell'Africa e dell'Asia l'estrazione, la lavorazione e l'utilizzo delle fibre è tutt'ora legale.

A influire su questo drammatico ritardo – drammatico in termini di vite spezzate e di gravi sofferenze – hanno contribuito una serie di fattori fra loro interconnessi, che hanno agito su diversi livelli. Senza ombra di dubbio, la scarsa comunicazione tra accademia e sapere scientifico, da un lato, e società civile, dall'altro, ha giocato un ruolo estremamente negativo. In secondo luogo, come per altri tipi di rischi sanitari e ambientali connessi alle produzioni industriali, per molto tempo le patologie asbesto-correlate sono state considerate un inevitabile prezzo da pagare in nome del progresso. In terzo luogo, dati i lunghissimi tempi di incubazione della malattia, e considerando i cicli di crescita nella produzione mondiale di amianto, la "bomba mesoteliomi" scoppiò generalmente tra la fine degli anni Settanta e la metà degli Ottanta. È solo in quel momento che il problema dell'asbesto iniziò con difficoltà a "uscire alla fabbrica" e a diventare collettivo, comune, diffuso: non più solo una questione di sicurezza sul lavoro, di tutela della salute degli operai, al limite di "monetizzazione del rischio", ma un dramma che incide nella carne viva di una intera comunità per decenni a venire. Infine, ha pesantemente influito l'attività di *lobbyng* messa in atto dai produttori, dal 1928 riuniti nella Saiac (Società associata industria del cemento-amianto), che per molti anni ha preferito compiere una massiccia opera di minimizzazione, puntando a promuovere un presunto «uso controllato» dell'amianto e continuando a esaltarne le caratteristiche positive, attraverso capillari campagne di comunicazione, senza informare dei gravissimi rischi correlati (v. [7], pagg. 136-138).

Certamente alla metà degli anni Ottanta nessuno poteva più considerare l'asbesto un «materiale straordinario», quanto piuttosto un lascito pesantissimo in termini umani ed economici, con le complesse procedure di sostituzione di materiali, di produzione e di bonifica a tutti i

livelli, e soprattutto con le centinaia di *Spoon River* dell'amianto diffuse in tutto il mondo, in tutte le città e i paesi in cui erano insediati stabilimenti estrattivi o di lavorazione di quel minerale che fino a pochi decenni prima sembrava aver semplificato la vita di tutti.

2. Le Spoon River dell'amianto

Non era un problema legato solo all'asbesto e all'industria cementifera. Tra gli anni Settanta e gli anni Ottanta, l'Italia soffriva di un dissesto ambientale diffuso. I ritardi nello sviluppo industriale erano stati colmati nei decenni precedenti ricorrendo a produzioni dequalificate, a basso valore aggiunto, a basso costo della manodopera, fortemente inquinanti. La natura nella sua integralità (le acque, il suolo, l'atmosfera) e i luoghi di vita urbani erano stati completamente "mercificati" e subordinati alle priorità dell'industrializzazione: lo sfruttamento intensivo delle risorse naturali e l'insediamento di industrie fortemente impattanti a ridosso di centri densamente abitati veniva visto con favore dalla stessa opinione pubblica, tanto più nelle zone scarsamente sviluppate e con alti livelli di disoccupazione, come nel Mezzogiorno o in alcune zone depresse a economia prevalentemente rurale dello stesso Nord (v. [8], pagg. 135-140; [9], pagg. 37-41).

La grande carica cancerogena dell'amianto, la volatilità delle sue fibre, nonché la sua elevata ecletticità di utilizzo hanno reso particolarmente gravi i suoi effetti sulla salute umana, tanto da creare emergenze sanitarie di lunghissimo periodo. Intere categorie professionali sono state e sono tuttora interessate da incidenze elevate di asbestosi e mesotelioma: i militari, i marinai, gli addetti all'industria dei trasporti, della cantieristica navale e ferroviaria, gli operai siderurgici, i vigili del fuoco (che indossavano tute in fibra di amianto), nonché – in particolare – tutti i lavoratori dell'edilizia e della manutenzione degli immobili, dai muratori ai lattonieri, dagli idraulici agli antennisti (che dovevano lavorare sulle ricoperture in cemento-amianto e spesso perforarle, operazione per lungo tempo compiuta senza alcuna precauzione particolare). Si tratta di un elenco del tutto parziale, proprio perché – come si è accennato – la vastità degli impieghi dell'asbesto nel campo civile e industriale ha coinvolto una miriade di attività e di prodotti anche di uso comune e ludico.

Di estrema drammaticità si è rivelata la situazione delle città che ospitavano stabilimenti destinati all'estrazione e alla lavorazione del

minerale. Qui la contaminazione da amianto si è diffusa in tutto il tessuto urbano e ha colpito massicciamente, "nel mucchio", l'intera popolazione: non solo le maestranze che erano a diretto contatto con il minerale ma anche chi abitava a chilometri di distanza dalla fabbrica e non aveva direttamente mai avuto nulla a che fare con essa. In tali casi, il contatto con le fibre ha interferito profondamente con le strutture, le identità e le percezioni di sé delle comunità e dei centri urbani interessati. È possibile così delineare una mappa di decine di luoghi, sparsi su tutto il territorio nazionale, nei quali il contatto con l'amianto ha generato migliaia di storie di malattia, di sofferenza e di morte, che hanno finito per costituire una "narrazione urbana dell'asbesto" di forza tale da imporsi su tutte le altre narrazioni cittadine.

Casale Monferrato (Eternit), Broni (Fibronit), Grugliasco (Sia – Società italiana amianto), Cavagnolo (Saca), Balangero (Società amiantifera di Balangero, che gestiva in loco la più grande miniera di amianto a cielo aperto d'Europa), Monfalcone (Fincantieri), Rubiera dell'Emilia (Eternit), Casaralta di Bologna (Officine ferroviarie), Pistoia (Breda ferroviaria), Cisterna di Latina (Goodyear), Bagnoli (Eternit), Castellammare di Stabia (Fincantieri), Bari (Fibronit), Taranto (Ilva), Priolo Gargallo (Eternit), Avellino (Isochimica); Palermo (Fincantieri): queste sono solo alcune delle città italiane segnate dal contatto con l'asbesto, e molte altre potrebbero essere citate. Da queste località, si dipanano racconti che si ripetono nella loro tragicità: gli ambienti di lavoro impolverati dalle fibre, le scarse misure di sicurezza e prevenzione, l'affacciarsi dell'infermità tra i colleghi, la persistente sensazione di pericolo spesso rimossa per il timore di perdere il lavoro, le patologie contratte dai familiari che in via indiretta venivano a contatto con il minerale, magari attraverso le tute da lavoro lavate in casa. E poi, la scoperta che la malattia usciva dalla fabbrica, colpiva gli abitanti del centro storico così come quelli delle periferie più distanti dagli stabilimenti, chi vi aveva lavorato per anni come chi vi aveva lavorato qualche settimana e chi – in termini numericamente ancora più consistenti – non vi aveva lavorato mai. Ha scritto Enrico Bullian: «[...] il mesotelioma rappresenta un'evidente epidemia, dove […] a essere malata è la comunità ed il suo modello produttivo, non solo l'individuo» (v. [1], pag. 218).

Tale scoperta è avvenuta in modi e tempi diversi per ciascuno dei "luoghi dell'amianto", per varie ragioni che vanno dal livello consapevolezza circa le modalità del processo produttivo all'atteggiamento delle dirigenze aziendali, dal diverso ruolo giocato dai sindacati nelle

varie realtà al contesto economico e industriale "di contorno". Tuttavia, in ogni caso, è stata accompagnata da una sorta di "congelamento" del discorso pubblico urbano, schiacciato dalla presenza dell'impianto contaminante, dalla necessità della sua chiusura e, in seguito, dal continuo monitoraggio epidemiologico e oncologico sulla popolazione, dai lutti, dalla percezione di un rischio imminente. Le stesse scelte amministrative sono risultate condizionate dall'esigenza primaria di operare la riconversione e la complessa (e costosa) bonifica del territorio: in molti casi si sono innescati duri contrasti politici e fratture tra la popolazione circa le modalità e i metodi di tali interventi.

Tra i casi più gravi di contaminazione urbana spiccano quelli di Casale Monferrato e di Broni, connotati da una altissima incidenza di mesoteliomi e di altre malattie asbesto-correlate dovuta alla presenza degli stabilimenti Eternit e Fibronit a ridosso del centro abitato. Si tratta di due casi per molti aspetti comparabili: le due cittadine distano meno di cento chilometri l'una dall'altra e si trovano in territori morfologicamente ed economicamente simili – il Monferrato e l'Oltrepò Pavese – caratterizzati dalla presenza di un territorio collinare e agricolo alle spalle, con prevalente coltura vitivinicola, e dall'apertura verso la pianura e verso alcuni centri industriali di dimensioni medio-grandi; in entrambi i casi la fabbrica, per le sue dimensioni, la sua collocazione in pieno centro urbano, la sua capacità di attrazione di manodopera, era entrata a far parte integralmente dell'identità cittadina; inoltre, i dati epidemiologici relativi ai mesoteliomi sono del tutto confrontabili. Una ricerca del 2012 ha evidenziato un'incidenza pari a 82 casi ogni 100.000 abitanti per Broni (il più elevato tasso grezzo rilevato in Italia) e di 81 casi ogni 100.000 abitanti per Casale Monferrato (v. [10], pp. 1-26).

Tuttavia, i tempi e i modi del disvelamento del rischio divergono in modo evidente, tanto fra gli attori politico-sociali preposti al suo controllo e alla sua gestione quanto fra l'insieme degli abitanti: ciò ha generato una discorde modalità di percezione del pericolo, un diverso modo di confrontarsi alla fabbrica, una differente narrazione del rapporto tra cittadini, lavoratori e amianto, un difforme processo di costruzione sociale, psicologica e memoriale dell'identità cittadina, profondamente influenzato dal contatto con la fibra e con il pericolo tuttora incombente e attuale della malattia.

Non a caso, mentre la vicenda di Casale Monferrato è diventata un simbolo a livello nazionale e internazionale degli effetti dell'esposizione all'amianto e della lotta contro le aziende responsabili, quella di Broni

è rimasta lungamente confinata all'ambito strettamente locale e nella letteratura specialistica medica e tecnologica. Ciò si è riflesso nelle scelte politiche, nelle modalità di sviluppo e di bonifica, nonché nelle visioni future per il destino dei due centri urbani. Lo studio di tali difformità può consentire di mettere in luce alcuni aspetti che influiscono in modo determinante sulla strutturazione delle identità cittadine poste di fronte a un disastro ambientale e sanitario, attraverso l'analisi del comportamento di alcuni attori politici e sociali che svolgono ruoli chiave nella definizione di quelle medesime identità.

3. Casale Monferrato e la Eternit

La storia recente di Casale Monferrato è legata a doppio filo all'inizio dell'era dello sfruttamento industriale dell'amianto in Italia. Nei primissimi anni del Novecento, l'ingegner Adolfo Mazza acquistò da Ludwig Hatschek il brevetto Eternit per l'Italia e fondò nel 1906 la società Eternit Pietra Artificiale, con sede legale a Genova e stabilimento nella cittadina piemontese, sulla sponda destra del Po. La località monferrina presentava le caratteristiche ottimali per la riuscita dell'impresa: già costituiva il principale polo italiano del cemento, insieme a Bergamo, c'era acqua in abbondanza e la presenza di uno scalo ferroviario. Il terremoto di Messina e Reggio Calabria del 1908 offrì all'azienda una tragica quanto insperata occasione per decollare, con le grandi richieste di materiale per la ricostruzione che ne derivarono.

Nel 1911 l'ingegner Mazza brevettò un sistema di produzione per condutture in Eternit ad alta pressione per acquedotti che risultò molto redditizio, data la leggerezza del prodotto, la sua economicità e la sua facilità di trasporto. Gli altri manufatti dell'azienda erano costituiti da lastre e tegole per coperture, fioriere (dal 1915) e lastre ondulate (dal 1933). Nel 1917 la società venne quotata in borsa, e nel 1929 entrò a far parte del cartello Saiac. Nel 1932 lo stabilimento di Casale venne ampliato, e nel 1939 fu aperto quello di Bagnoli; dopo la Seconda guerra mondiale, furono inaugurati anche quelli di Rubiera nell'Emilia e di Priolo Gargallo, vicino a Siracusa. La produzione aumentò fino agli anni Settanta, seguendo la parabola ascendente dell'impiego dell'amianto nel mondo, e lo stabilimento raggiunse a metà degli anni Sessanta i duemila dipendenti per poi stabilizzarsi intorno ai mille fino agli anni Ottanta.

A partire dalla metà degli anni Settanta, il consiglio di fabbrica dello stabilimento, in accordo con la locale Camera del Lavoro, iniziò una attività rivendicativa sulle condizioni di salute all'interno della fabbrica e sulle ricadute sanitarie in città. Fu una operazione non facile, che mirava a superare il concetto della "monetizzazione del rischio" per puntare alla sua limitazione e poi alla sua eliminazione: per quanto oggi possa apparire terribile, il "rischio retribuito" era considerato dagli stessi lavoratori quasi come una sorta di *benefit*, che consentiva di arrotondare lo stipendio con l'incentivo inserito in busta paga per coloro che operavano a diretto contatto con la materia prima. A guidare le prime rivendicazioni, talvolta anche contro l'opinione di molti lavoratori e delle organizzazioni di categoria degli edili, timorose di ricadute occupazionali, furono il giovane delegato Nicola Pondrano e l'allora segretario della Camera del Lavoro Bruno Pesce. Per esempio, venne richiesto più volte all'azienda di gestire in modo più sicuro l'attività di frantumazione degli scarti di produzione provenienti da altri stabilimenti Eternit della Penisola: i pezzi difettati venivano demoliti a cielo aperto da una ruspa, e il materiale polverizzato veniva trasportato su camion scoperti che attraversavano il centro urbano, provocando un notevole spargimento di fibre (v. [11], pagg. 47-59).

Nel 1981 il patronato Inca di Casale, diretto dallo stesso Pondrano, si impegnò in una battaglia processuale contro il diniego espresso dalla dirigenza di versare la cosiddetta "rendita di passaggio" – ovvero l'indennizzo spettante a chi aveva lavorato per anni a contatto con l'amianto – a 120 dipendenti le cui dimissioni erano state incentivate attraverso accordi e premi *ad personam*. Nell'ambito di tale vicenda giudiziaria, che giunse a sentenza nel 1984 con il riconoscimento delle ragioni dei lavoratori e della situazione di rischio in tutti i reparti, venne ordinata una perizia sullo stabilimento, condotta dal professor Guido Salvini dell'Università di Pavia, i cui risultati evidenziarono senza ombra di dubbio l'estrema pericolosità della produzione. Il docente ritenne di dover trasmettere gli atti della sua indagine scientifica anche alla Procura della Repubblica, affinché procedesse a ulteriori accertamenti e assumesse i provvedimenti più opportuni. Ha scritto Nicola Pondrano: «La nostra forza fu di rompere quell'accerchiamento melmoso, quell'abbraccio mortale con il quale la proprietà Eternit ci aveva avvolto per decenni: la colonia marina, la Befana per i figli, l'olio d'oliva due volte l'anno, i superminimi, lo spaccio alimentare, i turni di lavoro tarati al fine di favorire anche doppie attività (soprattutto il lavoro nei

campi), stipendi con maggiorazioni per i lavoratori disagiati o polverosi. Tutto ciò aveva impedito, di fatto, di far nascere la consapevolezza dei pericoli che si correvano in fabbrica» (v. [12], pag. 16).

Proprio al fine di rafforzare e di rendere più efficace la battaglia giudiziaria, la Cgil di Casale ritenne di dover portare la questione dell'amianto "fuori" dalla fabbrica, coinvolgendo la comunità monferrina nel suo insieme e svolgendo opera di informazione e sensibilizzazione verso una opinione pubblica allarmata dai casi sempre più frequenti di mesotelioma (in seguito, tale patologia si rivelerà distribuita per i 2/3 proprio tra i non dipendenti), ma che nel contempo tendeva a rimuovere e a non accogliere una realtà troppo inquietante da accettare. Così, nel 1984, l'Inca-Cgil e la Camera del Lavoro promossero a Casale un convegno sulle malattie asbesto-correlate, al quale parteciparono i più importanti studiosi del tema. Inoltre, nel 1988, su iniziativa del medesimo sindacato, venne creata una associazione delle vittime dell'Eternit (in seguito denominata Afeva, Associazione familiari vittime amianto), la cui presidenza fu affidata a Romana Blasotti Pavesi, che aveva perduto il marito, la sorella, una nipote e una figlia per mesotelioma pleurico. La nascita dello strumento associativo risultò un modo per aprirsi ulteriormente alla città, e la stessa scelta di Romana Blasotti – esterna alla Cgil – costituì un mezzo per facilitare la comunicazione con l'intero tessuto cittadino, anche nelle sue parti più critiche e diffidenti verso il sindacato di sinistra. In questo modo si produsse un "effetto di trascinamento" dell'attenzione nei confronti della comunità tutta e delle sue rappresentanze politiche di ogni colore (v. [13], pagg. 57-59).

Nel frattempo, l'Eternit aveva chiuso i battenti. Il mercato sempre più in calo, la consapevolezza che i pericoli connessi all'asbesto avrebbero a breve indotto interventi normativi limitativi, la necessità di puntare su produzioni e materiali alternativi, la mobilitazione della fabbrica e della città: sono tutte ragioni che avevano condotto a un progressivo ridimensionamento della produzione e del personale e, infine, al fallimento su autoistanza della Eternit italiana. In quel momento, a Casale vi erano ancora 350 dipendenti della multinazionale. La Safe, ovvero la holding francese del gruppo Eternit, si candidò per rilevare la fabbrica e riprendere la produzione di fibrocemento: tuttavia il sindacato oppose il proprio diniego, vincendo il prolungato tabù della difesa a oltranza dei posti di lavoro, e la stessa opinione fu espressa dai rappresentanti della cittadinanza. Il 2 dicembre 1987 – oltre quattro anni prima della legge 257/1992 – il sindaco democristiano di Casale, Riccardo Coppo,

emanò una ordinanza che per la prima volta in Europa sanciva «il divieto, con decorrenza immediata, dell'impiego di lastre di cemento-amianto e di altri manufatti di amianto nelle costruzioni di qualsiasi genere nell'ambito del territorio comunale; il divieto dell'utilizzazione per qualsiasi uso di materiale anche residuo a precedenti processi di produzione, contenente fibre di amianto; che in caso di rimozione e smaltimento di materiali contenenti fibre di amianto le ditte esecutrici dei lavori dovranno attenersi alle prescrizioni tecnico-sanitarie che saranno disposte dalle autorità competenti» (v. [11], pagg. 78-79).

Il sindacato casalese si mostrò all'avanguardia rispetto alle confederazioni nazionali e di categoria, nella sua capacità di uscire dalla stretta logica di monetizzazione del rischio e della difesa dei posti di lavoro purchessia. Infatti in quegli anni restava diffusa la percezione del rischio ambientale quale problema esclusivamente "interno" alla fabbrica, legato alla tutela della salute dei lavoratori ma senza ricadute sul territorio circostante. Il disastro di Seveso del 1976 aveva costituito una prima occasione per un cambio di atteggiamento, contribuendo a una prima evoluzione nella condotta delle autorità politiche verso i temi ecologici. Tuttavia si trattò di un percorso lungo e accidentato: ancora per molti anni, gli stessi lavoratori dell'industria mostrarono un sentimento di ostilità verso l'ambientalismo e la salvaguardia della salute "esterna", spesso cercando di minimizzare o rimuovere rischi e danni per evitare la chiusura o il ridimensionamento delle fabbriche. Solo nella seconda metà Ottanta il movimento sindacale cominciò a porsi problemi diversi e ad allargare il campo della sua analisi a una riflessione generale sul legame tra industria, città e ambiente, andando oltre la semplice tutela dei posti di lavoro e della sicurezza interna agli stabilimenti. Si tratto comunque di un processo molto graduale: nel 1987 la stessa Cgil organizzò un convegno intitolato "Proposta ambiente", presentato come il primo incontro del genere promosso dal sindacato, dove si confrontarono esperti di ecologia di varia competenza (giuristi, economisti, medici) e rappresentanti degli interessi diffusi, come politici e sindacalisti (v. [9], pp. 39-40).

4. Broni e la Fibronit

Molto diversa risulta la parte finale della storia dello stabilimento Fibronit di Broni. La Cementifera italiana Fibronit SpA, con sede legale

nella stessa Casale Monferrato, inaugurò l'impianto di Broni, a ridosso dell'abitato, nel 1919; tredici anni dopo, nel 1932, la fabbrica intraprese la produzione di cemento-amianto, con la realizzazione di condutture "a bicchiere" monolitiche, lastre ondulate, canne quadre per camini, pezzi speciali e raccorderia di completamento. Anche in questo caso nel dopoguerra la produzione aumentò, seguendo le direttrici della "fame di amianto" in Italia, e conseguentemente gli impianti vennero potenziati: negli anni Sessanta lo stabilimento lavorava su sette linee di produzione e impiegava circa 1300 operai. Negli anni Settanta, in concomitanza con il picco mondiale di produzione del materiale, furono implementate due ulteriori linee produttive. Il periodo successivo, caratterizzato dal graduale calo della domanda di asbesto, vide la Fibronit reggere apparentemente bene il colpo: la progressiva riconversione della Eternit su materiali diversi, meno inquinanti ma a maggior costo, garantiva l'acceso a un mercato "di risulta", lasciato libero dalla concorrente.

Così, la lavorazione dell'asbesto a Broni proseguì ininterrottamente fino all'approvazione della legge del 1992, e – in deroga – anche oltre. La cessazione definitiva avverrà solo nel giugno del 1993. In questo caso non si segnalano tentativi da parte del sindacato di aprire un confronto con la comunità circa i potenziali pericoli e la possibile strada di uscita dall'amianto. Nel corso degli anni Settanta anche qui il consiglio di fabbrica aveva avanzato alcune rivendicazioni per migliorare la salubrità degli ambienti di lavoro, le quali avevano portato all'installazione di filtri e aspiratori e a protocolli di lavorazione a umido meno pericolosi. Allo stesso modo, tra il 1972 e il 1993, erano state effettuate 12 indagini sull'ambiente di lavoro che avevano evidenziato molti problemi di eccessiva esposizione e di assenza di alcuni dispositivi idonei a ridurne l'impatto: l'esito di tali analisi era stato trasmesso tanto alla dirigenza quanto al consiglio di fabbrica medesimo. Tuttavia, fino alla messa al bando dell'amianto in Italia e alla successiva chiusura dello stabilimento, ciò non condusse ad alcuna azione al di fuori della fabbrica, né venne ipotizzato un possibile percorso di dismissione della produzione e di ricollocazione dei lavoratori. Eppure, la stampa e i mezzi di comunicazione avevano già ampiamente riportato le notizie provenienti dalla non lontana Casale, e lo stesso sindacato – in prima fila nella cittadina monferrina nell'individuare una via di uscita dall'amianto – doveva essere in possesso di tutte le informazioni sulla pericolosità di quelle lavorazioni (v. [11], pagg 101-116).

Racconta Osvaldo Galli, segretario della Camera del Lavoro di Voghera dal 1990 al 1998: «Chiudere le fabbriche non apparteneva alla nostra cultura. C'erano state molte chiusure tra la fine degli anni Ottanta e l'inizio degli anni Novanta, nel settore delle valigerie, nel tessile, nella meccanica, e si ritenne necessario mantenere aperta la fabbrica di Broni a ogni costo […]. Come sindacato, non c'è stata la capacità di coinvolgere il tessuto locale, ma lo stesso può dirsi per la classe politica: non si è mai creato un sistema. Le forze politiche e istituzionali non si sono fatte carico della situazione. In Piemonte si è mossa la magistratura, da noi no, nonostante la presenza della pretura a Stradella, che non si è mai mossa, e neppure le forze politiche. Il sindacato ha tentato di tenere aperta la fabbrica, poi ha lasciato: ha sottovalutato la situazione: non siamo stati capaci di affrontare la deindustrializzazione di questa provincia».[1]

Romana Bianchi, deputato bronese del Pci dalla VII alla X legislatura, aveva seguito da vicino l'iter di approvazione della legge 257/1992, e racconta di un rapporto molto difficile con la città e la fabbrica: «Quando abbiamo gestito e poi approvato la legge, insieme a Bruno Pesce e al gruppo di Casale, nel sindacato a Broni c'era ancora la difficoltà a riconoscere che aver lavorato lì significava aver lavorato in un ambiente estremamente dannoso e che la chiusura della Fibronit sarebbe stata un bene per tutti. Anche tra i lavoratori della Cgil c'era, come dire, una reticenza a occuparsi di queste cose […] Quello era, ed è ancora oggi, ancora un tabù […], è come se ci fosse una rimozione […]. La questione vera è che a Broni non c'è mai stato, come a Casale, un movimento con una capacità di coinvolgimento di tutti i cittadini. […]. Quando abbiamo approvato la legge, gli unici contenti erano i lavoratori vicini alla pensione, perché la legge abbuonava otto anni di anzianità. Però era come se non si dovesse parlarne troppo, della legge in via di approvazione, di amianto e di mesotelioma. È mancata una coscienza sindacale e politica capace di unificare per un obiettivo. Sembrava una fissazione di alcuni del Pci o di ambientalisti un po' snob. Agli operai gli echi di quello che accadeva a Casale non arrivavano o, se arrivavano, li rimuovevano […]. La legge del 1992 è stata vissuta un po' come una disgrazia a Broni, e io ero il soggetto della disgrazia...».[2]

1 Testimonianza orale di Osvaldo Galli, Pavia, 3 giugno 2015.
2 Testimonianza orale di Romana Bianchi, Pavia, 27 marzo 2015.

In un contesto economico connotato da un fortissimo individualismo e da una sostanziale mancanza di identità unitaria in un territorio storicamente diviso e frastagliato – con una forte resistenza al cambiamento nelle zone dell'Oltrepò più strettamente lombarde (mentre quelle più vicine al Piemonte e all'Emilia risultavano maggiormente dinamiche) – e contornato da una deindustrializzazione più rapida e di impatto maggiore rispetto ai distretti circostanti (v. [14], in particolare pagg. 107 e sgg.), il danno ambientale venne sostanzialmente rimosso. Lo stesso sindacato, in fase di regressione a livello generale, non riuscì a fare quel salto di qualità che a Casale Monferrato aveva precocemente compiuto, e verso il quale stava evolvendo a livello nazionale: restò confinato nell'azione a difesa dei posti di lavoro, e in seguito non si è neppure costituito parte civile nei processi penali contro la dirigenza della Fibronit. In questo modo, l'amianto, la fabbrica chiusa, la malattia, restarono a lungo una presenza sottaciuta ma drammaticamente penetrante, scandita dai lutti elaborati a livello del tutto individuale: «Le famiglie di Broni che sono state toccate dal problema dell'amianto, con la morte di qualche congiunto, sono la quasi totalità, alcune in maniera terribile, ma ciò non ha portato mai a una elaborazione collettiva dei lutti: si pensa che ognuno debba curarsi la propria malattia o il proprio lutto in casa propria, quasi fosse una vergogna farlo conoscere».[3] L'amianto, rimosso a livello psicologico dal tessuto cittadino, ha finito per diventare un fattore totalizzante, in grado di condizionare la vita, lo sviluppo e l'identità della città.

5. Lo stigma dell'amianto

Sulla vicenda della Eternit di Casale Monferrato sono state scritte decine di libri, con ricostruzioni di taglio giornalistico, raccolte di memorie, studi psicologici, sociologici, giuridici e sulla criminalità di impresa, analisi psicologiche sulla popolazione, eccetera (v. per esempio [15], [16], [17], [18], solo per citare alcune opere). È stato persino realizzato un fumetto, una *graphic novel* intitolata *Eternit. Dissolvenza in bianco* [19]. Nel 2013 è stato rappresentato in tutta Italia il testo teatrale *Malapolvere*, tratto dall'omonimo libro di Silvana Mossano [20], e a

[3] Testimonianza orale di Romana Bianchi, Pavia, 27 marzo 2015.

fine 2015 è uscito nelle sale cinematografiche il film *Un posto sicuro*, diretto da Francesco Ghiaccio e interpretato, fra gli altri, da Marco D'Amore, Matilde Gioli e Giorgio Colangeli [21]. Sono attivi molti progetti nelle scuole, per mantenere viva la memoria ed educare allo sviluppo sostenibile, e la cittadina monferrina è diventata addirittura meta di un "turismo responsabile" di nicchia: dopo la visita alle colline vinicole piemontesi, i turisti si recano a conoscere la storia della città che, prima in Europa, ha sfidato l'amianto.

Anche in questo caso va rimarcata la differenza con Broni. Al di là dell'attenzione del quotidiano locale *La Provincia Pavese*, che peraltro ha iniziato a interessarsi costantemente del problema negli ultimi dieci anni, e di alcuni articoli e reportage apparsi sulla stampa nazionale, che raccontano le storie drammatiche (queste sì, del tutto comparabili con quelle di Casale) delle vittime dell'amianto, non è stata pubblicata alcuna opera specifica, né è stato svolto alcuno studio organico, se si escludono quelli di natura medico-epidemiologica. L'interesse per la contaminazione non va oltre i confini del territorio pavese e, mentre Casale Monferrato è nota in tutta Italia per la sua tragica vicenda, quale realtà-simbolo di una produzione mortale, la vicenda di Broni è sconosciuta ai più.

A Broni il timore dello stigma dell'amianto si è percepito con particolare forza, con tutto il carico di rimozioni che spesso si accompagnano ai casi di grave contaminazione ambientale: vi ha albergato a lungo la paura, da parte della comunità cittadina, di essere classificata e riconosciuta come "città dell'amianto", anziché come "città del vino", come recita la scritta affissa a tutti gli accessi del paese (e a loro volta le preoccupazioni per le ricadute sulla produzione vincola hanno contribuito ad accrescere i processi di occultamento). La rimozione è il mezzo che si utilizza per cercare di suscitare un rapido ritorno alla normalità e per la cancellazione dello stigma medesimo: si tratta di un meccanismo non molto diverso da quello evidenziato in molte zone teatro di stragi di civili nel corso della seconda guerra mondiale. Ma alla normalità non si poteva tornare in tempi rapidi: le morti per mesotelioma continuavano a succedersi (e proseguiranno ancora per molti anni), e lo stabilimento Fibronit, un gigante strapieno di amianto su un'area di 140.000 metri quadrati, restava lì (come resta lì tuttora, seppure in corso di bonifica), come una bomba innescata che gravava sulla comunità.

Fino al 1997 non esisteva neppure una commissione amianto in seno al consiglio comunale: verrà istituita solo in quell'anno, su richiesta dei consiglieri di opposizione. Per arrivare alla costituzione di una struttura

associativa dedicata alla lotta contro l'amianto occorre attendere il 2000, quando – su iniziativa di un gruppo di cittadini di orientamento politico molto diverso, in particolare Andrea Astranti, Domenico Novarini e Costanza Pace – venne creato un "Comitato Difesa Ambiente", che richiese alla Regione di assumere alcuni provvedimenti per la messa in sicurezza dell'area Fibronit (v. [11], pag. 128). Ma trascorsero ancora otto anni prima che si giungesse a una svolta. Gianluigi Vecchi, di Legambiente Pavia, racconta in una intervista filmata: «Quando la fabbrica è stata chiusa, l'atteggiamento della gente e degli operai, che poi erano mezzo paese, era di negare. Poi, quando hanno incominciato a esserci i primi morti, hanno rimosso. Finalmente, nel 2008 […] noi, il sindaco, la Cgil […], per la prima volta abbiamo parlato e abbiamo visto che la situazione era grave, e abbiamo indetto una assemblea, perché ormai le morti non era più possibile nasconderle. E lì, in quella assemblea stavolta affollatissima, finalmente si pose il problema e non fu più possibile nasconderlo».[4]

Solo da quell'anno iniziò a svilupparsi un tessuto associativo sul problema dell'amianto, con l'attività dell'Aiea (Associazione italiana esposti amianto), da cui poi si sono dipanati anche il Gruppo Vittime Amianto e l'Avap (Associazione vittime dell'amianto pavese), nata in quello stesso 2008 e trasformata l'anno successivo in Avani (Associazione vittime amianto nazionale). Una notevole parte dell'attività di queste associazioni è assorbita dall'organizzazione delle parti civili nei processi penali, ma hanno rivestito anche il ruolo di "imprenditori della memoria", che «selezionano, modellano e propongono definizioni del passato» (v. [22], pag. 82). In particolare, l'Avani ogni 28 aprile organizza manifestazioni in occasione della giornata mondiale delle vittime dell'amianto; promuove iniziative nelle scuole; chiede ai parenti dei deceduti di indicare sui necrologi la causa della morte; pungola gli amministratori e i politici sui temi della bonifica e dello smaltimento dell'asbesto (v. [11], pag. 129). Tale tessuto associativo, seppur di grande visibilità mediatica, è nato quando ormai la fabbrica era chiusa da quindici anni e si pone obiettivi precisi e ben demarcati, in una logica rivendicativa post novecentesca. Inoltre, riflette a livello locale le divisioni che contraddistinguono su scala nazionale i diversi sodalizi

[4] Intervista a Luigi Vecchi di Legambiente Pavia, caricata su Youtube il 27 luglio 2010 nel canale del gruppo Amiantobroni: https://www.youtube.com/watch?v=GJvHMn7uU80 (visionato il 27 febbraio 2020).

che si occupano di rappresentare le vittime dell'asbesto. Come scrive Enrico Bullian, si tratta di «un universo di piccoli gruppi non particolarmente ideologizzati […] Le rivalità tra gruppi sono frequenti e gli obiettivi delle diverse associazioni possono divergere e scontrarsi […]. I gruppi sono cellule autonome che costituiscono delle reti dai confini indefiniti […]» (v. [1], pag. 321).

A Casale, già alla metà degli anni Ottanta, il sindacato si era "preso in carico" la gestione della questione amianto tra i lavoratori, l'aveva traghettata nel tessuto civile cittadino attraverso l'associazione delle vittime, l'aveva posta all'attenzione generale prospettando e negoziando le possibili vie d'uscita. Cosicché in quel caso nessuna rimozione è avvenuta, e la paura dello stigma non ha condizionato a lungo la città. Anzi, lo stigma è stato a sua volta "preso in carico" dalla comunità intera, appuntato sul gonfalone comunale, per così dire: la città ha voluto diventare un emblema della contaminazione da asbesto. La presenza precoce di una leadership in grado di assumere le responsabilità e la guida nella pianificazione e nell'attuazione dei percorsi legali ha consentito di realizzare una sorta di "resilienza", intesa come quella capacità di riconoscere il prima e il dopo del vissuto traumatico, che permette di trasformare il danno sofferto in potenzialità (v. [16], pagg. 56 e 65). Insomma, l'azione del sindacato e dell'associazione ha consentito di costruire un «quadro» della memoria collettiva, che ha permesso «la conservazione, lo sviluppo e l'esplicitazione dei contenuti della memoria dei singoli» (v. [23], pag. 22). Tali «quadri» costituiscono «il prodotto della sopravvivenza di gruppi che, nella vita quotidiana, rappresentano i punti di riferimento più immediati e familiari per il soggetto […]. Essi propongono dei modelli di comportamento la cui struttura ed il cui significato trovano giustificazione nell'ambito di specifiche memorie collettive. Questi referenti hanno senso finché sopravvivono le identità collettive che, sull'idea di una memoria condivisa, hanno costruito i propri confini identitari» (v. [24], pag. 66).

I procedimenti giudiziari, quelli conclusi e quelli in corso, hanno costituito ulteriori momenti di identificazione della comunità nel dramma collettivo: la lotta comune contro l'amianto è diventata un elemento fondativo e unificante dell'identità cittadina, anche nei momenti più difficili e di fronte alle scelte più controverse. Invece, nel centro oltrepadano l'amianto risulta ancora oggi un elemento divisivo, che genera polemiche e contrasti nella lotta politica e tra i cittadini. Paradigmatica da questo punto di vista è la storia della bonifica dei due territori – il

Monferrato e l'Oltrepò – nei quali l'asbesto è stato utilizzato in modo ancor più massiccio che altrove, proprio per la presenza dei due poli produttivi: tra l'altro, gli abitanti potevano approvvigionarsi direttamente presso gli stabilimenti, dove venivano venduti a prezzi molto bassi o addirittura regalati manufatti non commercializzabili perché lievemente difettati; inoltre, potevano prelevare senza formalità i sacchi contenenti il cosiddetto "polverino", ovvero la polvere di tornitura delle condutture in cemento-amianto, con la quale venivano realizzate miscele "casalinghe" per la pavimentazione di strade private, per la copertura dei fondi dei cortili, per i campi da bocce, per i sottotetti. A Casale Monferrato i piani di bonifica sono stati molto precoci e hanno portato a risultati soddisfacenti già da anni: la stessa area dello stabilimento Eternit è stata completamente bonificata nel 2005, mentre nel 2001 era stata risanata la parte di sponda destra del Po, ove un canale di scolo della fabbrica aveva creato una vera e propria spiaggia contaminata con amianto misto a sabbia. Nell'Oltrepò Pavese la bonifica è in grave ritardo, e lo stabilimento Fibronit costituisce ancora una ferita urbanistica nel tessuto urbano di Broni, fonte di continua preoccupazione per gli abitanti. Tale ritardo è dovuto anche alle difficoltà riscontrate in provincia di Pavia nell'individuare un sito di smaltimento del minerale: i comitati bronesi hanno rifiutato decisamente l'ipotesi che esso possa essere localizzato nella loro città, perché essa «ha già dato a dismisura» in termini di presenza di amianto; anche le altre soluzioni ipotizzate si sono rivelate di non agevole realizzazione, per l'opposizione degli abitanti e di molti amministratori locali. Invece, a Casale Monferrato il sito di smaltimento è stato realizzato nel 2001 nel territorio comunale, con l'accordo dell'Afeva e il confronto con i cittadini (v. [11], pagg. 130-133). Si tratta di una ulteriore dimostrazione della volontà e della capacità di "farsi carico" del problema amianto fino in fondo, senza alcuna rimozione: «La presenza della discarica di amianto a Casale è fondamentale, sotto tutti i punti di vista, anche sotto quello del riconoscimento del danno subito. Noi lo abbiamo prodotto, noi lo abbiamo usato, noi ci siamo ammalati, noi lo eliminiamo».[5] E il cerchio, così, si chiude [25].

Insomma, mentre a Broni neppure i procedimenti penali sono riusciti a concorrere in maniera decisiva alla cristallizzazione di una

[5] Testimonianza orale di Nicola Pondrano, Casale Monferrato, 25 marzo 2015.

coscienza e di una memoria comune degli eventi, «a Casale il sostegno politico e sociale e la partecipazione ad azioni collettive […] hanno contribuito nelle persone alla creazione di un luogo mentale in cui confrontarsi con la propria capacità di non sentirsi completamente in balia degli eventi, anzi di poter sperimentare un'area di libertà di azione» (v. [16], pag. 69).

6. Conclusioni

Casale Monferrato è "uscita dall'amianto" proprio perché ha trovato la forza di riconoscersi nel suo dramma, senza occultamenti o rimozioni, integrandolo nel tessuto e nell'identità comunitaria. Broni ne è ancora "dentro", proprio perché per lungo tempo non è riuscita a riconoscerlo come parte integrante, per quanto subìta e non voluta, del vivere e del morire collettivo. La rinascita di Casale, pur con tutti i problemi economici comuni a tutti i centri urbani del paese, è stata evidente mentre, per gli abitanti dei dintorni, Broni resta un luogo pericoloso, e il suo caso andrebbe ulteriormente approfondito attraverso capillari studi sociologici sul territorio. Una differenza fondamentale nelle due vicende è ravvisabile nel diverso atteggiamento assunto tra gli anni Ottanta e i primi anni Novanta dalle forze di massa intermedie, in particolare in quello del sindacato: precocemente capace di farsi carico del *trade-off* lavoro-salute per i lavoratori e per i cittadini nel caso monferrino, lungamente fermo sulla difesa a oltranza dei posti di lavoro nel caso bronese.

La crisi delle identità collettive ha prodotto la progressiva perdita di mordente delle organizzazioni politiche e sindacali di massa, e la loro sostituzione nel processo rivendicativo con strutture associative sul modello dei comitati, per loro natura parziali e frammentate. Tale processo ha creato un problema ulteriore, anche in relazione alla possibilità della popolazione di essere coinvolta nei processi decisionali circa il destino delle città, del loro sviluppo industriale, della pianificazione urbanistica (v. [26], pagg. 82 e sgg; [27], pagg. 63 e sgg). Da questo punto di vista, oggi il caso italiano più eclatante è certamente quello dell'Ilva di Taranto, caratterizzato da una spaccatura verticale della città, tra chi "ipervisibilizza" la fabbrica, considerandola la fonte di tutti i mali del territorio (e finendo così per trascurare quelli dovuti ad altri fattori, anche di più lungo periodo) e chi invece la "invisibilizza",

rimuovendone gli aspetti più problematici, in nome della necessità di salvaguardare l'occupazione. Anche nella vicenda pugliese, il mondo sindacale e politico ha scontato un drammatico ritardo nella gestione complessiva dell'emergenza (v. [28], pagg. 286-290).

Ciò che emerge con chiarezza è che l'informazione e il coinvolgimento dei cittadini nelle scelte politiche e amministrative circa il destino delle città assume un'importanza ancora più forte, nel momento della crisi delle grandi ideologie e nella "società liquida" di oggi (v. [29], pagg. 427-429). Solo in questo modo le comunità possono riuscire a trasformare gli ostacoli in potenzialità. È una grande responsabilità collettiva posta in capo ai decisori, ai cittadini, ma anche ai possessori dei saperi in grado di orientare razionalmente le scelte: ovvero, a quel mondo scientifico e accademico che, conscio dei pericoli connessi all'utilizzo dell'amianto fin dagli anni Sessanta, non è riuscito a comunicarli e trasmetterli al di fuori dei suoi circoli.

BIBLIOGRAFIA

1. Enrico Bullian, *Il male che non scompare. Storia e conseguenze dell'uso dell'amianto nell'Italia contemporanea*, Il Ramo d'Oro, Trieste, 2008.
2. Mirco Volpedo, Davide Leporati (a cura di), *Morire d'amianto. L'Eternit di Casale Monferrato: dall'emergenza alla bonifica*, La Clessidra Editrice, Genova, 1997.
3. Stefano Silvestri, Francesco Di Benedetto, Corrado Raffaelli, Angela Veraldi, Asbestos in toys: an exemplary case, *Scandinavian Journal of Work, Environment and Healt*, anno: 2016 numero: 1 volume: 42.
4. Robert L. Virta, *Worldwide Asbestos Supply and Consumption Trends from 1900 through 2003*, U.S. Geological Survey, Reston-Virginia, 2006.
5. William Edmund Cooke, Fibrosis of the Lungs Due to the Inhalation of Asbestos Dust, *The British Medical Journal*, July 25, 1924.
6. Francesco Carnevale, Evoluzione e destino delle conoscenze sugli effetti dell'amianto per salute dei lavoratori, in: Ariella Verrocchio (a cura di), *Storia/Storia di amianto*, Ediesse, Roma, 2012.
7. Diego Alhaique, Fulvio Cavarian, Laurent Vogel, Amianto, tragedia senza confini, postfazione, in: Guido Iocca, *Casale Monferrato: la polvere che uccide. Voci dalla Chernobyl italiana*, Ediesse, Roma, 2011.
8. Saverio Luzzi, *Il virus del benessere. Ambiente, salute, sviluppo nell'Italia repubblicana*, Laterza, Roma-Bari, 2009.
9. Bruno Ziglioli, *La mina vagante. Il disastro di Seveso e la solidarietà nazionale*, FrancoAngeli, Milano, 2010.
10. Stato dell'arte e prospettive in materia di contrasto delle patologie asbesto-correlate, *Quaderni del Ministero della Salute*, anno: 2012 numero: 15 (maggio-giugno).

11. Bruno Ziglioli, *"Sembrava nevicasse". La Eternit di Casale Monferrato e la Fibronit di Broni: due comunità di fronte all'amianto*, FrancoAngeli, Milano, 2016.
12. Nicola Pondrano, Prefazione, in: Guido Iocca, *Casale Monferrato: la polvere che uccide. Voci dalla Chernobyl italiana*, Ediesse, Roma, 2011.
13. Bruno Ziglioli Romana e le altre. La narrazione femminile dell'amianto a Casale Monferrato, *La camera blu*, anno: 2018 numero: 18.
14. Antonella Zucchella, Stefano Denicolai, *Analisi strategico-organizzativa per lo sviluppo locale. Il Development Migration Path*, FrancoAngeli, Milano, 2006.
15. Rosalba Altopiedi, *Un caso di criminalità di impresa. L'Eternit di Casale Monferrato*, L'Harmattan, Torino, 2011.
16. Antonella Granieri, *Amianto, risorsa e dramma di Casale: risvolti psicologici nelle persone affette da mesotelioma e nei loro familiari*, Fratelli Frilli Editori, Genova, 2008.
17. Alberto Prunetti, *Amianto. Una storia operaia*, Alegre, Roma, 2014.
18. Alessandro Calvi, *Paracarri. Cronache da un'Italia che nessuno racconta*, Rubbettino, Soveria Mannelli, 2015.
19. Gea Ferraris, Assunta Prato, Eternit. *Dissolvenza in bianco*, Ediesse, Roma, 2011.
20. Silvana Mossano, *Malapolvere. Una città si ribella ai «signori» dell'amianto*, Sonda, Casale Monferrato, 2010.
21. *Un posto sicuro*, regia di Francesco Ghiaccio, Italia, 2015, 102 minuti.
22. Teresa Grande, Le origini sociali della memoria, in: Anna Lisa Tota (a cura di), *La memoria contesa. Studi sulla comunicazione sociale del passato*, FrancoAngeli, Milano, 2001.
23. Paolo Jedlowski, Introduzione alla prima edizione, in: Maurice Halbwachs, *La memoria collettiva*, Unicopli, Milano, 2001.
24. Marita Rampazi, *Storie di normale incertezza. Le sfide dell'identità nella società del rischio*, Led, Milano, 2009.
25. Barry Commoner, *Il cerchio da chiudere. La natura, l'uomo, la tecnologia*, Garzanti, Milano, 1972.
26. Gian Luigi Bulsei, *Ambiente e politiche pubbliche. Dai concetti ai percorsi di ricerca*, Carocci, Roma, 2005.
27. Gian Luigi Bulsei, *Le politiche ambientali. Intervento pubblico e regolazione sociale*, Rosenberg & Sellier, Torino, 1990.
28. Salvatore Romeo, *L'acciaio in fumo. L'Ilva di Taranto dal 1945 a oggi*, Donzelli, Roma, 2019.
29. Marta Dente, Giovanna Landi, Informazione e partecipazione. Il ruolo dell'opinione pubblica, in: Achille Cutrera, Giuseppe Pastorelli, Barbara Pozzo (a cura di), *Seveso trent'anni dopo: la gestione del rischio industriale*, Giuffrè, Milano, 2006.

Istituto Lombardo (Rend. Scienze) 151, 171-182 (2017)

CURRENT PERSPECTIVES ON THE MICROBIOTA-GUT-BRAIN AXIS IN CHRONIC NEURODEGENERATIVE DISORDERS

DIEGO ALBANI (*) e CARMEN GIORDANO (**)

Nota presentata dal m.e. Stefano Maiorana
(Adunanza del 29 giugno 2017)

SUNTO. – Le patologie neurodegenerative croniche sono prevalenti a causa dell'invecchiamento della popolazione. Queste condizioni invalidanti includono la demenza (come la malattia di Alzheimer - AD) e i disturbi del movimento (come il morbo di Parkinson - PD). Negli ultimi decenni, abbiamo accumulato prove sperimentali sulle possibili cause molecolari di AD/PD, ma sono disponibili pochi farmaci approvati, sfortunatamente senza alcuna reale azione di modifica della malattia. Mentre i casi di AD/ PD aumentano costantemente nel mondo, nuove ipotesi sulla loro patogenesi sono attivamente sotto esame. Una è basata sull'asse microbiota-intestino-cervello, che postula un'influenza diretta o indiretta della comunità microbica dell'intestino umano sul sistema nervoso e le sue funzioni. Questa interazione bidirezionale può avere un ruolo anche nei disturbi neurodegenerativi, in questo campo la ricerca richiede modelli e strumenti per crescere e giungere all'applicazione clinica. A tale riguardo, i modelli innovativi di malattia *in vitro* stanno guadagnando attenzione, in particolare perché presentano caratteristiche desiderabili come condizioni di crescita tridimensionale (3D) e movimento dei fluidi interstiziali grazie alla microfluidica. In questa nota riassumiamo i concetti fondamentali del ruolo del microbiota intestinale nelle malattie come AD/PD, con particolare attenzione ai progetti di ricerca in corso sull'argomento che stanno cercando di risolvere quesiti clinici o elaborare soluzioni tecniche per la ricerca di base e le applicazioni industriali in questo settore.

(*) Department of Neuroscience, IRCCS – Istituto di Ricerche Farmacologiche Mario Negri, Milano, Italy.

(**) Department of Chemistry, Materials and Chemical Engineering "Giulio Natta", Politecnico di Milano, Italy. E-mail: carmen.giordano@polimi.it

***

ABSTRACT. – Chronic neurodegenerative disorders are mostly due to population aging. These invalidating conditions include dementia (as Alzheimer's disease - AD) and movement disorders (as Parkinson's disease - PD). In the past decades, we have been accumulating experimental evidence on the molecular triggers of AD/PD and few approved drugs are available, unfortunately with no real disease-modifying action. As AD/PD cases constantly rise in the world, new hypothesis on their pathogenesis are actively under investigation. One of this is based on the microbiota-gut-brain axis, that postulates a direct or indirect influence of the microbial host community of human gut on the nervous system and its functions. This bidirectional interaction may play a role also in neurodegenerative disorders, even though research in this field requires models and tools for growing to a clinical translational perspective. To this respect, innovative *in vitro* engineered disease models are gaining attention, particularly because they feature desirable characteristics as tridimensional (3D) growing conditions and interstitial fluids movement recapitulation thanks to microfluidics. Here we summarize current data on the impact of gut microbiota on AD/PD models and patients, with a focus on ongoing research projects on the topic that are trying to unravel clinical questions or elaborate technical solutions for research and industry applications in the field.

1. THE MICROBIOTA-GUT-BRAIN AXIS AND NEURODEGENERATIVE DISORDERS

1.1. The gut microbiota

The human gastrointestinal tract hosts many different microorganisms, mostly bacteria, but also fungi, yeasts, viruses, and archaea [1]. The bacterial populations are collectively defined "microbiota", and their whole genome "microbiome" [2].

It seems that the gut is colonized by trillions of bacteria belonging to hundreds of different species that are dynamic and different between individuals. Moreover, the microbiota can change during the course of life and can be affected by different factors such as diet, life style, diseases and assumption of drugs [3,4]. The most abundant phyla in adults are *Bacteroidetes* and *Firmicutes* that sum-up 80-90% of the whole gut microbiota [1].

The microbiota in the gut has many functions important for human health. It is involved in modulation of inflammatory signaling and protection against infections, thanks to nutrient competition, leading to starvation of competing pathogens [5]. Another way is the pro-

duction of inhibitory substances such as bacteriocins, that act as antimicrobial peptides. Intestinal microbiota allows also the fermentation of complex non-digestible carbohydrates and the synthetizes of Short Chain Fatty Acids (SCFAs), as acetate, propionate and butyrate [6]. The SCFAs have anti-inflammatory properties; maybe anti-cancer effects, and also contributes to the control of mucin production [6,7].

The 'healthy gut microbiota' is very important also for human immune-system development and reactivity, thanks to the controlled exposition of the host to microbial strains [8]. Finally, the intestinal microbiota promotes the intestinal barrier homeostasis, keeping its thickness and stimulating the production of mucus layer and the expression of intestinal tight junctions.

1.2 Microbiota-gut-brain-axis

A connection between altered microbiota and changes in the central nervous system physiology are increasingly evident. Gut microbiota dysbiosis can increase the intestinal permeability, with translocation of endotoxins across the epithelial barrier and ensuing pro-inflammatory cytokines production [9]. The activation of enteric neurons and glial cells may result in neurological dysfunction spreading along the gut-brain-axis (for instance, by the vagus nerve) and might led to brain neurodegenerative disorders, such as Parkinson disease (PD) and Alzheimer disease (AD). As a matter of fact, recent studies have found a reduced abundance of the *Prevotellaceae* bacteria family in PD patients compared with healthy controls, and an increased abundance of *Enterobacteriaceae* among those patients with the postural instability and gait difficulty phenotype compared to those with tremor-dominant PD [10].

In AD, deposition of protein aggregates composed of amyloid-β peptide and tau protein in CNS tissues compromises cognitive function, mainly memory. We have evidence that alterations of microbiota population by antibiotic treatment may affect learning and memory abilities [11], even if this association is hard to study with available models, mostly germ-free (GF) animal models [12], as they suffer from main limitations as the different composition of the animal's microbiota and the lack of a strict control on microbiota composition once inocu-

lated. For this reason, future research in AD/PD centered on microbiota's pathologic role would greatly benefit from innovative solutions. In this scenario, bidimensional (2D) cell culture systems are the traditional tools, however, the availability of state-of-the art, innovative three dimensional (3D) and engineered cell-based models will strongly improve *in vitro* cell model reliability.

In the following paragraphs, an highlight on traditional and innovative 3D and engineered cell models related to AD/PD neurodegeneration will be provided, with a focus on an innovative multiorgan platform that, starting from state-of-the-art models, will represent the first *in vitro* model of the whole microbiota-gut-brain axis, useful to evaluate microbiota impact on brain cell pathophysiology.

2. Bidimensional and tridimensional cell culture systems

Cell culture is the process by which eukaryotic cells are grown outside their natural environment. After the cells of interest have been isolated from living tissue, typically a tumour, they can be maintained under controlled conditions, that vary from each cell type but generally consist of a plastic support with a liquid medium which supplies the essential nutrient and gasses, thus regulating the physio-chemical environment. Cell culture systems are indispensable tools for a wide range of basic and clinical *in vitro* research and can be classified between two-dimensional (2D) and three-dimensional (3D) cultures.

Bidimensional (2D) culture has been used since the early 1900s, when Ross Granville Harrison established the methodology of tissue culture adapted from bacteriology. He maintained tissue explants outside the animal and observed the outgrowth of nerve cells over time [13]. The classical 2D culture is still a static dish culture, which mainly generates cell monolayers. Most 2D cultures are adhesion dependent and cannot be grown in suspension cultures without mechanical support, so cells are often put onto coated surfaces of flat plastic dishes.

Two-dimensional cultures, due to their simplicity, are still extensively employed in the majority of cell cultures. They are inexpensive, very well established and a rich body of literature exists to compare

outcome measures. Moreover, 2D cultures are easy to control, observe, and analyze by imaging, and provide cells nutrients without any diffusional transport limitation [14].

On the other hand, they do not fully reflect the situation *in vivo*, where cells grow within a complex three-dimensional microenvironment, with vascular perfusion continuously supplying metabolites and removing catabolites. In addition, they exhibit increased drug sensitivity, due to their exposed surface, and generally show a reduced capability to rebalance when stimulated.

As a specific limitation for brain and its disease models, 2D neuronal networks show random connections with other types of neural cells and overlapping cellular responses, showing limitations in mimicking the 3D tissue specific characteristics.

Moreover, human neural cells cultured in flat, stiff, 2D environments typically display an irregular morphology and form unnatural cell-cell interactions. In this case, tissue specific architecture (3D neuronal network), mechanical and biochemical cues and cell-cell communication are lost to various degrees. This can lead to physiological inaccuracies that can be extremely problematic for disease modelling and pre-clinical drug screening.

In addition, 2D cultures, even if derived from sophisticated cell models (*e.g.* with induced-pluripotent stem cells - iPSC) show reduced viability, neurite outgrowth and synaptic density [15].

The 2D condition, in AD/PD disease models, may have the additional important limitation of lacking a time-dependent extracellular/intracellular accumulation of aggregated proteins, probably as a result of a culture condition not favourable conditions to protein accumulation, as the proteins might diffuse in the relatively large volume of cell culture media and are likely to be removed during regular media changes, preventing aggregation [16].

3. Engineered models of neurodegenerative disorders

Two key strategies are currently pursued to overcome 2D models also in neurodegeneration: 3D culturing and microfluidics. Below we describe the most relevant AD/PD disease models developed so far,

focusing on engineered systems and devices employed that take advantage of the just described technologies.

3.1 Alzheimer's disease models

Alzheimer's disease is characterized by two pathological features: Aβ plaques and neurofibrillary tangles, that are connected to AD's brain damage. In particular the amyloid hypothesis proposes that the accumulation of Aβ plaques within brain triggers AD's pathogenesis: it's necessary to develop *in vitro* AD's models to comprehend the complex mechanisms of AD's pathogenesis, and therefore to test therapeutic drugs. AD has been studied widely with patient-derived iPSCs: to overcome the observed 2D iPSC neuron cell culture models' inability to generate extracellular protein aggregations, the research has quickly passed to 3D and microfluidic systems.

The first 3D iPSC neuron culture model from familial AD patients (mutant APP and mutant presenilin) was developed by Choi and colleagues and it represented, for the first time, a disease model that displayed hallmarks of AD pathology: extracellular deposition of amyloid-β, including amyloid-β plaques, and aggregates of phosphorylated tau, as well as filamentous tau [17]. The model was based on a human neural stem cell-derived 3D culture system from neural progenitor cells (hNPCs). Progenitor cells were differentiated into neuronal and glial cells as they were planted onto a BD Matrigel™ layer, used to provide a 3D support matrix and to prevent diffusion of Aβ, allowing for high local concentrations sufficient to initiate aggregation [16].

More recent studies applied the 3D culture model to generate high throughput models for drug screening against tau aggregation [18], or to compare efficacy of drug candidates (β-secretase or γ-secretase inhibitors) in 2D vs. 3D culture systems [19], as 3D culture models of AD are also relatively cheaper and faster than AD transgenic mouse model. In particular, Medda and colleagues used commercially available cell lines to generate iPSC–derived cortical neurons to reproduce a robust and scalable tau aggregation model for drug discovery. Also in this case, they used Matrigel™ to have a thin-layer 3D culture [18].

The generation of brain organoids, initially developed to model human brain, has been used to study mechanisms of AD, too. In fact,

another way to reproduce 3D structures that can facilitate interstitial compartments of Aβ deposition is cerebral organoids or neuro-spheroid, self-organizing structures of stem cells cultivated *in vitro*.

Lee and colleagues developed the first iPSC-differentiated 3D neuro-spheroid model derived from AD patients' blood. In this model iPSC derived from ante-mortem blood samples obtained from five AD patients were cultured on Geltrex™ matrix-coated plates [19]. The use of 3D spheroid involves a reduced diffusion of drugs and chronic dosing and may constitute a disadvantage for large-scale drug screening, because of the increased time requirement for uniform bioavailability within the 3D spheroids. However, there are many offsetted advantages of 3D environment: besides a better anatomic resemblance to mature brain, it expands 3D cells that reliably represents the native target of the drug, it also allows for microscopic evaluation of spatial features related to drug effects and it eventually permits to quantify drug levels that is not available in our traditional 2D assay system [19].

In the same period, Raja and colleagues reported that brain organoids from FAD patients recapitulate AD disease phenotypes and pathologies including amyloid aggregation, hyperphosphorylated tau, and endosome abnormalities, all of which were reduced by treatment with secretase inhibitors [20]. One very important aspect of this model is the spontaneous and time-dependent appearance of both amyloid and tau pathology, in agreement with the observations from the scaffolded 3D cultures of Choi and colleagues [16].

The above described 3D models lack controlled flow dynamics, circulating blood cells, and a brain-specific microenvironment and ECM. Several *in vitro* AD models have been developed in microfluidic systems to deal with some of the just reported limitations [21].

For instance, Yoon Jung Choi and colleagues studied the neurotoxicity of Aβ oligomeric assemblies by developing a microfluidic system capable of generating a gradient of Aβ oligomeric assemblies within microchannels. Neuron cells were directly exposed to Aβ using a microfluidic culture system that could generate slow flow rates similar to those of interstitial flow in brain because of an osmotic pump into the platform [22]. The main result was the toxicity of oligomeric Aβ aggregates, a well-documented feature of *in vivo* AD models [22].

Park and colleagues from the same group extended this work to

develop a 3D AD model in a microfluidic system incorporating 3D neuro-spheres more closely mimicking the brain tissue [23], while tau propagation from neuron to neuron was examined within a microfluidic system composed of three distinct culture chambers connected by arrays of microgrooves [21]. This system examined uptake and propagation properties of different tau species derived from postmortem cortical extracts and brain interstitial fluid (ISF) of tau-transgenic mice, as well as human AD cortices [24].

3.2 Parkinson's disease models

Parkinson's disease symptoms result from the death of dopamine-generating neurons in a deep-brain region called *substantia nigra*, a pathologic process possibly initiated by the abnormal accumulation of -synuclein protein in the form of Lewy bodies.

PD cellular models rely on different cell lines having similarities to dopaminergic neurons and susceptible to toxicity deriving from protein aggregates containing α-synuclein.

Relevant lines are the neuroblastoma cell line SH-SY5Y and the pheochromocytoma cell line PC12, as well as Lund Human Mesencephalic cells (LUHMES) - which show a stable dopaminergic phenotype after differentiation protocols - and primary dopaminergic cultures from rat or mice [25]. In recent studies, similarly to AD, patient-derived iPSC have led to progresses taking a significant leap towards personalized medicine [26].

These cell models have been coupled to microfluidic systems to provide realistic *in vivo*-like models.

One of the first microfluidic platform virtually able to reproduce Parkinson's disease *in vitro* was developed by Lu *et al.* in order to study mitochondrial transport impairment as a contribute to PD neural degeneration. Lu *et al.* developed a microdevice with a large open chamber directly adjacent to an array of microchannels, which furthered axon isolation and orientation to better visualize the movement of mitochondria and detect their speed [27].

The microfluidic approach has been also important in models studying the proposed α-synuclein spreading among cells as possible propagation system of PD pathology.

For instance, Freundt *et al.* analysed the spreading of α-synuclein neuron to neuron with the aid of microfluidic device, even if medium change was manual and not continuous. Their model employed cortical mouse neurons from E17 embryonic stage growing them directly in the microfluidic device in order to separate soma from axonal projection in fluidically isolated microenvironments. In this study, it was demonstrated that α-synuclein fibrils are internalized, transported within axons, released, and subsequently taken up by additional neurons [28].

Finally, Fernandes *et al.* developed a microfluidic cell-culture platform for studying the communication between different cell populations, to allow in a dynamic environment the investigation of neurons/glia cell interactions. Of note, this dynamic platform avoids direct cell-cell contact, allowing cell to communicate either by diffusion or by perfusion of molecules [29].

3.3. MINERVA project as flexible platform to model microbiota-gut-brain axis

In the 2016, the European Research Council (ERC) Consolidator Grant Programme funded a technological project named "MIcrobiota-Gut-BraiN EngineeRed platform to eVAluate intestinal microflora impact on brain functionality" (acronym: MINERVA, ERC-CoG-2016 – ID 724734), having the main aim to evaluate microbiota metabolism impact on brain functionality linked to neurodegeneration. In fact, MINERVA proposes an *in vitro* multi-organ-on-chip microfluidic platform, based on miniaturized, optically accessible devices, serially connected to recapitulate the microbiota-gut-brain axis main players. These organ-on-chips are the "Microbiota-compartment", hosting one or more bacterial strains relevant for human gut microbiota, the "Gut-compartment", featuring gut epithelial and immune system cells, and the most innovative "Brain-compartment", reproducing a physiologic blood-brain barrier model and two 3D human brain cell models, where neurons, astrocytes and microglia are cultured alone or in combination. It is clear that MINERVA platform tries, thanks to a multidisciplinary approach, to offer a substantial contribution to speed-up the research on microbiota-gut-brain axis, in particular for neurodegenerative disorders as Alzheimer's disease or Parkinson's disease. To get these results,

it will develop the "Brain compartment" as a 3D *in vitro* disease model based on hydrogel scaffolds and human neural cells derived from iPSC, all features that, to a different extent, are currently already used in other available *in vitro* disease models that try to overcome limitations of standard 2D culturing techniques, as better detailed above.

4. Conclusions

State-of-the-art on the role of gut microbiota in neurodegeneration is now facing the need of reliable research tools helping in solving the basic question of the existence of a causal link between microbiota composition or metabolic products and the molecular triggers of Alzheimer's or Parkinson's disease. Of course, this is would be a revolutionary hypothesis, as research in the field has been focused up to now almost exclusively on the brain and the central nervous system. The availability of innovative solutions (as the platform MINERVA is), based on 3D *in vitro* culturing conditions and microfluidics are a current hot-spot, as demonstrated by the recent work in the field referenced above, and they will become real laboratory-scale routine in a few years, with the related perspective of clinical translation of results. In fact, a major outcome of research done in this field will be the identification of dietary or pharmacologic approaches able to keep the microbiota in a healthy condition, with a forced step forward also for precision medicine, as the individual microbioma may be unique and varies over time, so requesting a personalized treatment in presence of a neurodegenerative disorder or, even better, for its prevention. Finally, the microbiota-gut-brain axis challenge is also the need of a truly multidisciplinarity involving neuroscience, immunology, gastroenterology, material science and bioengineering, a philosophy fully present in MINERVA and other recent technological solutions.

Acknowledgements

"MINERVA" project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant Agreement No 724734)".

REFERENCES

[1] Sekirov I, Russell SL, Antunes LC, Finlay BB. Gut microbiota in health and disease. Physiol Rev. 2010, 90:859-904.

[2] Pagliari D, Saviano A, Newton EE, Serricchio ML, Dal Lago AA, Gasbarrini A, Cianci R. Gut Microbiota-Immune System Crosstalk and Pancreatic Disorders. Mediators Inflamm. 2018, 2018:7946431.

[3] Clarke SF, Murphy EF, Nilaweera K, Ross PR, Shanahan F, O'Toole PW, Cotter PD. The gut microbiota and its relationship to diet and obesity: new insights. Gut Microbes. 2012, 3:186-202.

[4] Maukonen J, Saarela M. Human gut microbiota: does diet matter? Proc Nutr Soc. 2015, 74:23-36.

[5] Rolhion N, Chassaing B. When pathogenic bacteria meet the intestinal microbiota. Philos Trans R Soc Lond B Biol Sci. 2016, 371(1707). pii: 20150504. doi: 10.1098/rstb.2015.0504.

[6] Louis P, Flint HJ. Diversity, metabolism and microbial ecology of butyrate-producing bacteria from the human large intestine. FEMS Microbiol Lett. 2009, 294:1-8.

[7] Miquel S, Martín R, Rossi O, Bermúdez-Humarán LG, Chatel JM, Sokol H, Thomas M, Wells JM, Langella P. Faecalibacterium prausnitzii and human intestinal health. Curr Opin Microbiol. 2013, 16:255-61.

[8] Mazmanian SK, Liu CH, Tzianabos AO, Kasper DL. An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. Cell. 2005,122:107-18.

[9] Wang B, Yao M, Lv L, Ling Z, Li L. The Human Microbiota in Health and Disease. Engineering, 2017, 3: 71–82.

[10] Mulak A, Bonaz B. Brain-gut-microbiota axis in Parkinson's disease. World J Gastroenterol. 2015, 21:10609-20.

[11] Fung TC, Olson CA, Hsiao EY. Interactions between the microbiota, immune and nervous systems in health and disease. Nat Neurosci. 2017, 20:145-155.

[12] Martin CR, Osadchiy V, Kalani A, Mayer EA The Brain-Gut-Microbiome Axis. Cellular and Molecular Gastroenterology and Hepatology. 2018, 6: 133-148.

[13] Millet LJ, Gillette MU. Over a century of neuron culture: from the hanging drop to microfluidic devices. Yale J Biol Med. 2012, 85:501-521.

[14] Arber C, Lovejoy C, Wray S. Stem cell models of Alzheimer's disease: progress and challenges. Alzheimers Res Ther. 2017, 9:42.

[15] Siney EJ, Kurbatskaya K, Chatterjee S, Prasannan P, Mudher A, Willaime-Morawek S. Modelling Neurodegenerative Diseases *In Vitro*: Recent Advances in 3D IPSC Technologies AIMS Cell and Tissue Engineering. 2018, 2: 1-23.

[16] Choi SH, Kim YH, Hebisch M, Sliwinski C, Lee S, D'Avanzo C, Chen H, Hooli B, Asselin C, Muffat J, Klee JB, Zhang C, Wainger BJ, Peitz M, Kovacs DM, Woolf CJ, Wagner SL, Tanzi RE, Kim DY. A three-dimensional human neural cell culture model of Alzheimer's disease. Nature. 2014, 515:274-278.

[17] Ghaffari LT, Starr A, Nelson AT, Sattler R. Representing Diversity in the Dish:

Using Patient-Derived *in Vitro* Models to Recreate the Heterogeneity of Neurological Disease. Front Neurosci. 2018,12:56.

[18] Medda X, Mertens L, Versweyveld S, Diels A, Barnham L, Bretteville A, Buist A, Verheyen A, Royaux I, Ebneth A, Cabrera-Socorro A. Development of a Scalable, High-Throughput-Compatible Assay to Detect Tau Aggregates Using iPSC-Derived Cortical Neurons Maintained in a Three-Dimensional Culture Format. J Biomol Screen. 2016 , 21:804-15.

[19] Lee HK, Velazquez Sanchez C, Chen M, Morin PJ, Wells JM, Hanlon EB, Xia W. Three Dimensional Human Neuro-Spheroid Model of Alzheimer's Disease Based on Differentiated Induced Pluripotent Stem Cells. PLoS One. 2016,11:e0163072.

[20] Raja WK, Mungenast AE, Lin YT, Ko T, Abdurrob F, Seo J, Tsai LH. Self-Organizing 3D Human Neural Tissue Derived from Induced Pluripotent Stem Cells Recapitulate Alzheimer's Disease Phenotypes. PLoS One. 2016,11:e0161969.

[21] Osaki T, Shin Y, Sivathanu V, Campisi M, Kamm RD. *In Vitro* Microfluidic Models for Neurodegenerative Disorders. Adv Healthc Mater. 2018, 7

[22] Choi YJ, Chae S, Kim JH, Barald KF, Park JY, Lee SH. Neurotoxic amyloid beta oligomeric assemblies recreated in microfluidic platform with interstitial level of slow flow. Sci Rep. 2013, 3:1921.

[23] Park J, Lee BK, Jeong GS, Hyun JK, Lee CJ, Lee SH. Three-dimensional brain-on-a-chip with an interstitial level of flow and its application as an *in vitro* model of Alzheimer's disease. Lab Chip. 2015,15:141-50.

[24] Takeda S, Wegmann S, Cho H, DeVos SL, Commins C, Roe AD, Nicholls SB, Carlson GA, Pitstick R, Nobuhara CK, Costantino I, Frosch MP, Müller DJ, Irimia D, Hyman BT. Neuronal uptake and propagation of a rare phosphorylated high-molecular-weight tau derived from Alzheimer's disease brain. Nat Commun. 2015, 6:8490.

[25] Lázaro DF, Pavlou MAS, Outeiro TF. Cellular models as tools for the study of the role of alpha-synuclein in Parkinson's disease. Exp Neurol. 2017, 298:162-171.

[26] Torrent R, De Angelis Rigotti F, Dell'Era P, Memo M, Raya A, Consiglio A. Using iPS Cells toward the Understanding of Parkinson's Disease. J Clin Med. 2015, 4:548-566.

[27] Lu X, Kim-Han JS, O'Malley KL, Sakiyama-Elbert SE. A microdevice platform for visualizing mitochondrial transport in aligned dopaminergic axons. J Neurosci Methods. 2012, 209:35-9.

[28] Freundt EC, Maynard N, Clancy EK, Roy S, Bousset L, Sourigues Y, Covert M, Melki R, Kirkegaard K, Brahic M. Neuron-to-neuron transmission of -synuclein fibrils through axonal transport. Ann Neurol. 2012, 72:517-24.

[29] Fernandes JT, Chutna O, Chu V, Conde JP, Outeiro TF. A Novel Microfluidic Cell Co-culture Platform for the Study of the Molecular Mechanisms of Parkinson's Disease and Other Synucleinopathies. Front Neurosci. 2016,10:511.

Istituto Lombardo (Rend. Scienze) 151, 183-200 (2017)

LA CONVERSIONE DI ENERGIA SOLARE IN ENERGIA CHIMICA

ELENA SELLI (*)

Nota presentata dal m.e. Stefano Maiorana
(Adunanza del 5 ottobre 2017)

SUNTO. – L'energia costituisce la questione più importante del ventunesimo secolo. A causa di gravi problemi ambientali, legati soprattutto all'inquinamento dell'aria nelle zone urbane e all'effetto serra con conseguente riscaldamento globale, è necessario sostituire i combustibili fossili con fonti di energia rinnovabili. Accanto all'energia eolica, progressivamente sempre più utilizzata, l'energia solare rappresenta una forma di energia pulita, abbondante, diffusa ed economica, che può essere convertita in calore, elettricità o nei cosiddetti combustibili solari, utilizzando materiali in grado di assorbire la luce solare in modo efficiente e consentire una separazione delle cariche fotoprodotte. Questo è il prerequisito per ottenere dalla luce solare elettricità e i combustibili solari, come l'idrogeno dalla scissione fotocatalitica dell'acqua o sostanze organiche contenenti carbonio dalla riduzione del biossido di carbonio, quindi combustibili che possono essere immagazzinati e utilizzati al bisogno. La conversione dell'energia solare in energia chimica costituisce pertanto una grande sfida per la chimica.

***

ABSTRACT. – Energy is the most important issue of the 21st century. Due to severe environmental problems, mainly related to air pollution of urban areas and green house effects leading to global warming, fossil fuels need to be replaced by renewable energy sources. Beside the wind energy, which is progressively more and more exploited, solar energy represents a clean, abundant, diffuse and economical energy which can be converted into heat, electricity or in the form of so called solar fuels, provided special materials are developed able to efficiently absorb solar light and produce charge separation. This is the prerequisite for obtaining electricity from sunlight and solar fuels, e.g.

(*) Dipartimento di Chimica, Università degli Studi di Milano, Italy.
E-mail: elena.selli@unimi.it

hydrogen from water photosplitting or carbon containing species from carbon dioxide photoinduced reduction, which can be stored and used on need. Converting solar energy into chemical energy is thus a big challenge for chemistry.

I combustibili fossili hanno offerto enormi opportunità di sviluppo ai paesi del mondo occidentale durante il ventesimo secolo, ma già da tempo sono sorte grosse problematiche conseguenti al loro utilizzo. Non solo le riserve comprovate di combustibili fossili stanno progressivamente diminuendo e quindi occorre ricorrere in modo massiccio ad altre fonti di energia sostitutive, ma è ben evidente che il continuo utilizzo dei combustibili fossili comporta notevoli ricadute dannose, quali l'inquinamento ambientale che costituisce una minaccia per la salute e l'aumento dei gas serra con il conseguente riscaldamento globale del nostro pianeta e i disastrosi cambiamenti climatici che ne conseguono.

Senza dubbio l'energia costituisce il più importante problema del ventunesimo secolo [1]. Occorre trovare sorgenti di energia con cui sostituire i combustibili fossili e al tempo stesso far sì che tutti gli abitanti della terra possano raggiungere gli standard di vita delle nazioni sviluppate senza produrre devastazioni nel nostro pianeta, cercando quindi di lasciarlo in buono stato per le generazioni future. Tutti noi siamo quindi chiamati innanzitutto a evitare sprechi di energia e a contribuire allo sviluppo e al miglioramento delle tecnologie per l'energia perché il problema energetico possa essere risolto alla radice.

Probabilmente la conseguenza più pesante del consumo di combustibili fossili consiste nell'alterare il ciclo del carbonio che favorisce il riscaldamento globale. Anche se la quantità di carbonio immesso nell'aria a seguito dell'utilizzo dei combustibili fossili può sembrare piccolo in confronto con il suo continuo scambio che avviene naturalmente fra la biosfera e l'atmosfera, è tuttavia abbastanza, sul lungo termine, per far sì che la concentrazione di CO_2 aumenti costantemente. Dalla rivoluzione industriale, in effetti, la concentrazione di CO_2 nell'atmosfera è aumentata del 30%. Ciò comporta un'alterazione dell'equilibrio di irraggiamento della terra e l'induzione di variazioni climatiche che si sovrappongono alle oscillazioni naturali [2], con il riscaldamento globale e lo scatenarsi di eventi climatici estremi, quali uragani, siccità o ondate di caldo.

I combustibili fossili

Il petrolio è senza dubbio il combustibile ideale in quanto liquido e quindi facilmente estraibile e trasportabile. Dall'inizio del ventesimo secolo la domanda di petrolio è cresciuta costantemente negli anni ed il suo approvvigionamento è sempre stato in grado di far fronte alla domanda. Vi sarà un momento, tuttavia, in cui la fornitura di petrolio non sarà in grado di soddisfarne la domanda sempre crescente. In quel momento le riserve di petrolio non saranno esaurite, ma l'età d'oro del petrolio facile sarà finita.

Il picco nell'utilizzo del gas naturale, l'altra preziosa fonte di energia di origine fossile, è atteso più tardi rispetto al picco di utilizzo del petrolio. Tuttavia l'infrastruttura per la distribuzione del gas naturale è più rigida e fragile, rispetto a quella del petrolio. In linea di principio il problema della sua distribuzione può essere ridimensionato con l'utilizzo di navi a gas naturale liquefatto e di terminali costieri di rigassificazione, infrastrutture molto costose, che presentano gravi problemi di sicurezza.

Il carbone è il più abbondante, ma il più sporco dei combustibili fossili. Ma nei prossimi anni il problema del carbone non sarà tanto la sua disponibilità, ma la sua sostenibilità ambientale. L'uso dei combustibili che producono più gas serra è chiaramente in conflitto con gli attuali sforzi per una politica sul clima di scala mondiale.

La transizione già in atto dai combustibili fossili a qualche tipo di fonte di energia rinnovabile è quindi urgente e irrinunciabile. Né ci si può aspettare che la soluzione venga dall'utilizzo dell'energia nucleare. Infatti, oltre al problema, tuttora irrisolto, dello smaltimento sicuro delle scorie nucleari, che comporta la grossa preoccupazione di lasciare alle generazioni future depositi nucleari radioattivi per migliaia di anni, la possibilità concreta e comprovata del verificarsi di incidenti nelle centrali nucleari fa sì che ci si possa solo aspettare che il loro numero diminuirà, non aumenterà, negli anni futuri.

Fonti di energia rinnovabile

Esistono forme di energia cosiddette rinnovabili che sono già competitive con l'utilizzo dei combustibili fossili, quali l'energia eolica, l'energia termoelettrica e forme di energia che utilizzano fenomeni locali.

Produrre elettricità dal vento presenta una serie di notevoli vantaggi, quali, innanzitutto, il costo zero garantito del vento, l'assenza totale di emissioni nell'atmosfera e di calore da dissipare. Si tratta di una tecnologia relativamente semplice, con tempi di costruzione delle strutture relativamente brevi. Per questo il costo dell'elettricità prodotta per via eolica continuerà a diminuire sul lungo termine, rendendola sempre più competitiva rispetto ad altre tecnologie. Gli svantaggi sono correlati con la naturale variabilità dei venti, la distanza fra i campi eolici e i centri di consumo e le obiezioni di tipo estetico ed ecologico all'istallazione di pale eoliche. L'Europa ha un grosso potenziale di energia eolica, che in linea di principio dovrebbe poter soddisfare tutte le nostre necessità di elettricità.

L'energia idroelettrica è una delle tecnologie per la produzione di elettricità in uso da più tempo, per cui se ne possono facilmente valutare i pro e contro. Fra questi ultimi, per prima l'invasività. Nel ventesimo secolo la costruzione di grandi dighe ha comportato la necessita di dislocamento di popolazioni soprattutto nei paesi in via di sviluppo, con incalcolabili ricadute di tipo sociale ed economico. Sono stati inondati ecosistemi naturali unici, con conseguenti cambiamenti drammatici della vita nei fiumi a valle delle dighe. Nonostante questo, la produzione di energia idroelettrica presenta innegabili vantaggi, fra cui i bassi costi operativi e la durata più lunga degli impianti rispetto a qualsiasi altro modo di produzione di elettricità, il facile accumulo di energia potenziale che consente di far fronte alla domanda di picchi di elettricità, la possibilità di fornire acqua da bere e per irrigare, la prevenzione di inondazioni ricorrenti, talora distruttive. Enorme è in linea di principio il potenziale idroelettrico non sfruttato della terra; tuttavia soltanto una sua frazione molto limitata può essere utilizzata in modo sostenibile dal punto di vista economico e ambientale. Per quanto riguarda l'Europa, il potenziale per nuovi impianti idroelettrici è minimo, ma sistemi di piccole dimensioni largamente distribuiti sul territorio potranno in futuro dare un significativo contributo al fabbisogno di energia.

Fra le altre opzioni che madre natura ci regala per trarre energia dalla terra, la più importante è probabilmente l'energia geotermica derivante dal calore immagazzinato in profondità del nostro pianeta, che può essere estratta ed usata direttamente come calore o convertita in elettricità. Come sorgenti di energia possono essere utilizzati anche i

gradienti di temperatura di oceani e laghi, le correnti e le onde, mentre l'energia gravitazionale terra-luna può essere sfruttata in zone costiere dell'emisfero nord dove le maree muovono grossi volumi di acqua di mare in un tempo relativamente breve.

L'energia solare

La vita sulla terra non sarebbe possibile senza l'irraggiamento solare, che ci garantisce un flusso di energia inesauribile, in quantità che supera di gran lunga le necessità dell'umanità. Il sole è la stella madre del sistema solare (*Fig. 1*). È una nana gialla, la cui superficie è costituita da idrogeno (74%) ed elio (25%), con tracce di altri elementi più pesanti.

L'energia solare rende possibile la vita sul nostro pianeta. Senza la luce solare, infatti, non sarebbe possibile la fotosintesi, attraverso cui la CO_2 viene trasformata in composti organici e l'ossigeno ha raggiunto l'attuale concentrazione nell'atmosfera terrestre.



Fig. 1. Il Sole.

L'importanza del sole per la vita sulla terra è ben riconosciuto fin dall'antichità. In molte culture antiche e preistoriche, ad esempio quelle degli Inca e degli Aztechi, il sole era concepito come una divinità. Nella religione egizia il Sole era la divinità più importante e il faraone, considerato una divinità in terra, era ritenuto il figlio del Sole.

Antichi monumenti furono costruiti tenendo conto della posizione del sole nei vari periodi dell'anno: ad esempio i megaliti di Stonehenge (*Fig. 2*) segnano il solstizio d'estate. Nella mitologia greca la divinità solare principale fu Helios, in epoca storica confuso con Apollo, rappresentato alla guida del carro del sole, una quadriga tirata da cavalli che soffiano fuoco dalle narici (*Fig. 3*).



Fig. 2. I megaliti di Stonehenge.



Fig. 3. Monete antiche.

L'energia solare ha un enorme potenziale come fonte di energia pulita, abbondante ed economica, che, tuttavia, non può essere adoperata come tale, ma deve essere raccolta e convertita in altre forme di energia facilmente utilizzabili. Poiché l'energia solare è diffusa e intermittente, la sua conversione deve comportare la sua concentrazione ed il suo immagazzinamento. Ma attualmente nessuna delle vie utilizzate per convertire l'energia solare in calore, elettricità o combustibili è competitiva rispetto all'utilizzo dei combustibili fossili al loro prezzo attuale di mercato.

Il solare termico. Il modo con cui l'energia solare viene al momento maggiormente usata consiste nell'utilizzo dei collettori termici solari, in cui un liquido viene scaldato sotto irraggiamento solare per fornire acqua calda per uso diretto o per riscaldamento domestico (*Fig. 4*). Si tratta di sistemi molto semplici, che non richiedono la concentrazione della luce solare. Questo modo di utilizzo dell'energia solare potrà avere un ruolo molto rilevante, se si considera che nelle case moderne più del 50% dell'energia utilizzata viene consumata banalmente per scaldare l'acqua per riscaldarsi, lavarsi e cucinare. Quindi una larga diffusione del solare termico comporterà un sostanziale alleggerimento del conto dell'energia nel settore residenziale e delle relative conseguenze ambientali.



Fig. 4. Collettori solari.

L'energia solare può anche essere convertita in calore ad alta temperatura usando specchi che concentrano la radiazione solare (*Fig. 5*) e quindi in elettricità, o usata direttamente in processi termochimici.

La concentrazione di luce solare è stata usata fin dai tempi dell'antica Cina (*Fig. 6*). Si tramanda anche che Archimede usò scudi ben lucidati come specchi per concentrare la luce del sole sulla flotta romana che stava assaltando Siracusa nel 212 A.C.

L'elettricità solare. L'elettricità può essere prodotta dalla luce solare mediante celle fotovoltaiche e dispositivi correlati, oltre che da sistemi ad alta temperatura. Le celle fotovoltaiche costituiscono il sistema di gran lunga più promettente per la produzione di elettricità diffusa,

hanno molte applicazioni e sono particolarmente adatte a situazioni in cui non si ha a disposizione l'energia elettrica della rete. D'altra parte nei paesi sviluppati le celle solari, in forma di moduli o di pannelli solari sui tetti degli edifici, possono essere collegate alla rete di distribuzione dell'energia elettrica.



Fig. 5. Il sistema di concentrazione della luce solare alla Plataforma Solar de Almeria.

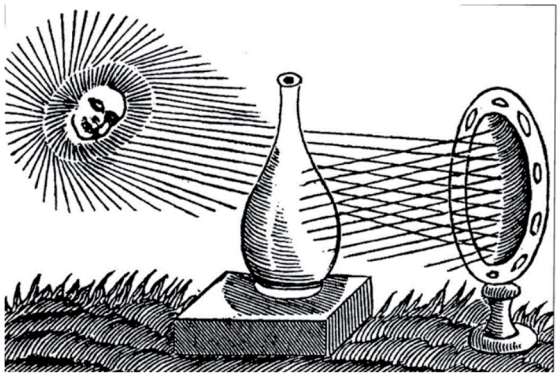


Fig. 6. Concentrazione di luce solare mediante specchio.

Il costo dell'elettricità prodotta dalla luce solare è già competitivo con quello dell'elettricità prodotta da sorgenti fossili e dal nucleare. In questo campo continua un'intensa e estesa ricerca per produrre nuovi materiali più performanti in grado di assorbire la luce solare e migliorare la tecnologia, per aumentare l'efficienza delle celle fotovoltaiche e ridurne i costi.

L'energia solare può essere convertita in energia elettrica anche in celle fotoelettrochimiche basate su elettrodi a semiconduttore in contatto con un elettrolita che consente scambio di elettroni (*Fig. 7*). Il semiconduttore più stabile e quindi di gran lunga più utilizzato fino a poco fa è il biossido di titanio, TiO_2, che tuttavia è in grado di assorbire solo una frazione relativamente piccola della luce solare perché il suo band gap è troppo largo (3.0 eV). Per superare questo problema, la luce può essere fatta assorbire da molecole di colorante adsorbite su film sottili di particelle nanocristalline di TiO_2 [3]. Le molecole di colorante, elettronicamente eccitate a seguito dell'assorbimento di luce solare, iniettano elettroni nel TiO_2, producendo processi di separazione di carica. Le celle solari sensibilizzate da coloranti, le cosiddette *dye sensitized solar cells*, sono interessanti per il basso costo del TiO_2 e per la potenziale semplicità del loro processo di costruzione. Tuttavia non hanno rese di conversione elevate e presentano problemi principalmente legati alla stabilità dei coloranti. Negli ultimi anni sono state superate da celle basate su materiali diversi, le perovskiti, con le quali posso essere ottenute rese di conversione molto maggiori, pur presentando problemi, soprattutto di stabilità, ancora non risolti.

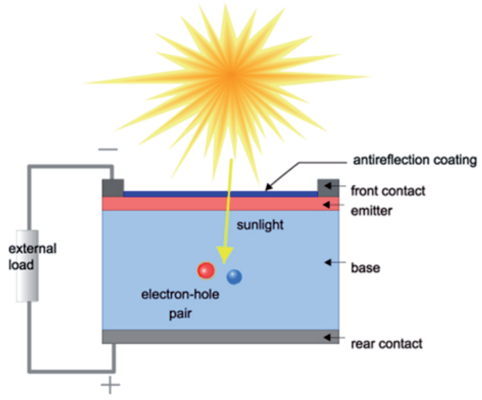


Fig. 7. Cella fotoelettrochimica.

Combustibili solari e fotosintesi artificiale. I combustibili liquidi rappresentano la miglior forma di energia in quanto possono essere facilmente conservati e trasportati. A causa della natura intermittente della luce solare, la sua conversione in combustibili chimici facilmente

immagazzinabili e trasportabili attraverso un processo analogo a quello che avviene nella fotosintesi naturale costituisce sicuramente la soluzione di gran lunga più vantaggiosa per l'utilizzo dell'energia solare.

La fotosintesi naturale, infatti, converte la luce solare in combustibili, producendo biomassa, che in tempi geologici si è poi trasformata in combustibili fossili. Tuttavia la velocità di produzione di combustibili fossili è 500.000 volte inferiore rispetto alla velocità con cui vengono attualmente consumati. Globalmente, solo lo 0.3% dell'energia solare che raggiunge la terra viene immagazzinata sotto forma di materia delle piante [1]. Aumentare l'efficienza della fotosintesi naturale sarebbe ovviamente un risultato importante, per raggiungere il quale è essenziale chiarire le basi molecolari del processo di fotosintesi naturale. Negli organismi fotosintetici la luce viene raccolta da sistemi antenna che consistono in complessi pigmento-proteina. L'energia di eccitazione così catturata viene poi trasferita ai centri di reazione dove viene convertita in energia potenziale elettrochimica in un processo di trasferimento elettronico allo stato eccitato. Gli equivalenti ossidanti e riducenti che ne derivano sono quindi trasportati ai siti catalitici, attraverso successivi *step* di trasferimento elettronico, e qui sono utilizzati per ossidare l'acqua e produrre combustibili, quali i carboidrati.

La necessità e la possibilità di ottenere la fotosintesi artificiale [4] è stata anticipata più di un secolo fa dal chimico italiano Giacomo Ciamician (*Fig. 8*), che in un famoso articolo [5] pubblicato su Science nel 1912 scrisse quanto segue:

> *When vegetation is rich, photochemistry may be left to the plants and, by rational cultivation, solar radiation may be used for industrial purposes. In the desert regions, unsuitable for any kind of cultivation, photochemistry will artificially put their solar energy to practical uses. On the arid lands there will spring up industrial colonies without smoke and without smokestacks; forests of glass tubes will extend over the plants and glass buildings will rise everywhere; inside of these will take place the photochemical processes that hitherto have been the guarded secret of the plants, but that will have been mastered by human industry which will know how to make them bear even more abundant fruit than nature, for nature is not in a hurry and mankind is. And if in a distant future the supply of coal becomes completely exhausted, civilization will not be checked by that, for life and civilization will continue as long as the sun shines!*



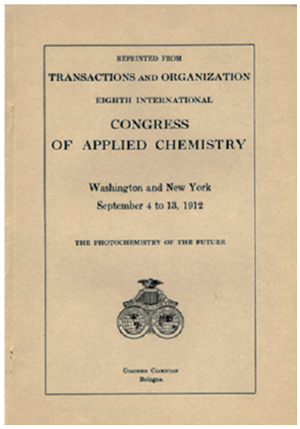


Fig. 8. Giacomo Ciamician.

La produzione efficiente di combustibili solari puliti rappresenta senza dubbio una delle più importanti sfide della scienza moderna. Perché la produzione di combustibili solari sia vantaggiosa dal punto di vista economico e ambientale, occorre ottenerli a partire da materiali di partenza abbondanti e non costosi, quali l'acqua e il biossido di carbonio. L'acqua può essere scissa in idrogeno e in ossigeno molecolari sulla superficie di un semiconduttore in grado di assorbire la luce solare e produrre una efficiente separazione di carica (*Fig. 9*). E il biossido di carbonio in soluzione acquosa può essere ridotto, ad esempio, a etanolo con la concomitante generazione di ossigeno molecolare.

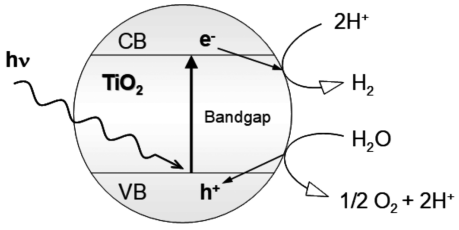


Fig. 9. Scissione fotocatalitica dell'acqua in idrogeno e ossigeno su biossido di titanio.

Su queste due semplici reazioni, e in particolare sullo sviluppo di materiali fotoattivi stabili in grado di assorbire in maniera efficiente la luce solare e consentire una buona separazione delle cariche fotogenera-

te, si è principalmente focalizzata negli ultimi anni la ricerca sulla conversione di energia solare in energia chimica. L'idrogeno molecolare H_2 fotogenerato per scissione fotocatalitica dell'acqua è il combustibile pulito per eccellenza, in quanto la sua combustione con ossigeno produce calore e acqua e la sua combinazione con ossigeno in una *fuel cell* genera elettricità, calore e acqua, senza alcuna immissione di CO_2 nell'atmosfera.

Il modo migliore per costruire sistemi fotosintetici artificiali per la produzione di combustibili solari consiste nel cercare di imitare l'organizzazione molecolare e supramolecolare del processo fotosintetico naturale: la raccolta di luce deve portare ad una separazione di carica, seguita da un trasporto di carica per trasferire gli equivalenti ossidanti e riducenti ai siti catalitici dove avviene separatamente lo sviluppo di ossigeno e lo sviluppo di idrogeno (o la riduzione di CO_2). Progressi sono indubbiamente stati fatti sui singoli aspetti della fotosintesi artificiale, con l'obiettivo di raggiungere l'integrazione dei vari componenti in un sistema attivo.

Il problema principale consiste nel fatto che la formazione di idrogeno e di ossigeno dall'acqua (così come la riduzione di CO_2) sono processi a molti elettroni, mentre l'assorbimento di luce è un processo a un fotone che comporta un processo di separazione di carica a un elettrone. Sono quindi necessari catalizzatori che accoppino eventi a singolo elettrone in processi in grado di accumulare gli equivalenti redox necessari per la produzione di carburanti. La ricerca è particolarmente focalizzata sullo sviluppo di catalizzatori efficienti per la generazione di ossigeno, che è un processo che implica il trasferimento di quattro elettroni.

Combustibili da biomassa. Occorre infine ricordare che la biomassa stessa, prodotta dalla fotosintesi naturale, è stata e continua ad essere un'importante risorsa per la produzione di energia, in particolare nei paesi in via di sviluppo, dove è spesso utilizzata per fornire energia in industrie su piccola scala e più in generale come legno combustibile da bruciare nelle stufe. Questo può comportare grossi problemi respiratori dovuti all'inquinamento dell'aria interna in conseguenza della cattiva ventilazione delle stufe alimentate con biomassa.

Nei paesi industrializzati la biomassa è utilizzata per produrre calore ed elettricità usando principalmente biomassa solida, o per ottenere combustibili liquidi, quali il bioetanolo e il biodiesel. L'etanolo viene estratto dalla canna da zucchero in Brasile e dal mais negli Stati Uniti. Una estensiva produzione di energia da biomassa richiede ovviamente aree di terra coltivabile molto grandi ed enormi quantità di acqua.

Sviluppo di materiali e dispositivi per la conversione di energia solare

Il gruppo di ricerca da me diretto nel Dipartimento di Chimica dell'Università degli Studi di Milano si dedica da anni alla conversione di energia solare in energia chimica, in particolare allo sviluppo di materiali fotocatalitici in grado di assorbire la luce solare e produrre idrogeno da soluzioni acquose, ovvero migliorare la separazione delle cariche prodotte nel materiale a seguito dell'assorbimento di luce, al fine di costruire fotoanodi in celle fotoeletrocatalitiche. Tali materiali sono stati utilizzati anche nella costruzione di celle e *devices* per la produzione di corrente e lo sviluppo separato di idrogeno e ossigeno mediante scissione fotocatalitica dell'acqua.

Come già accennato, il biossido di titanio è il materiale semiconduttore fin qui più largamente usato in fotocatalisi e fotoelettrocatalisi, sostanzialmente perché poco costoso e stabile anche sotto irraggiamento. Il principale limite nell'utilizzo del TiO_2 come materiale fotoattivo consiste nel fatto che è in grado di assorbire solo una piccola porzione della luce emessa dal sole. Pertanto sono stati studiati materiali a base di TiO_2 che assorbono una porzione più larga dello spettro solare e che consentono anche una migliore separazione delle cariche fotogenerate in tale materiale. Uno dei modi di spostare l'assorbimento del materiale a lunghezze d'onda più lunghe consiste nell'introduzione di stati all'interno del band gap del semiconduttore mediante drogaggio del materiale, ad esempio sostituendo parzialmente l'ossigeno nel TiO_2 con elementi quali carbonio o azoto [6]. Tuttavia in questo modo la mobilità delle cariche fotogenerate nel materiale per assorbimento di radiazione risulta molto limitata e i fotocatalizzatori ottenuti, pur assorbendo una frazione maggiore della luce solare, risultano meno efficienti.

Risultati molto interessanti sono stati invece ottenuti mediante fluorurazione del TiO_2 [6]. Mediante studi di fluorescenza risolta nel tempo, abbiamo dimostrato che in tali materiali si generano difetti superficiali che fungono da trappole per le lacune elettroniche fotogenerate, con effetti modesti sulle proprietà di assorbimento [7], ma con una migliore separazione di carica fotoindotta e conseguente aumento dell'attività fotocatalitica del materiale [8].

Per lo sviluppo fotocatalitico di idrogeno da soluzioni acquose particolarmente efficienti sono materiali a base di TiO_2 modificati per

deposizione di nanoparticelle di metalli [9], sia nobili (NM), quali oro [10-12] e soprattutto platino [13], che rame [14], preparati con un metodo innovativo di *flame spray pyrolysis* [11-13] sviluppato nei nostri laboratori (*Fig. 10*). L'efficienza di tali materiali risulta molto più elevata [15,16], principalmente perché le nanoparticelle di NM sono in grado di catturare gli elettroni fotopromossi nella banda di conduzione del semiconduttore e facilitare quindi la riduzione di acqua a idrogeno (*Fig. 11*). Tali materiali sono stati studiati principalmente nella reazione di produzione fotocatalitica di idrogeno da soluzioni acquose, ma ci siamo interessati anche della riduzione fotocatalitica di CO_2 [17,18].

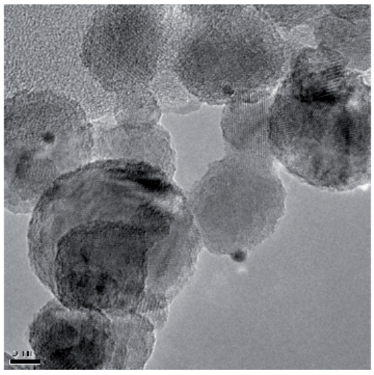


Fig. 10. Particelle di biossido di titanio contenenti nanoparticelle di metallo, ottenute per flame spray pyrolysis.

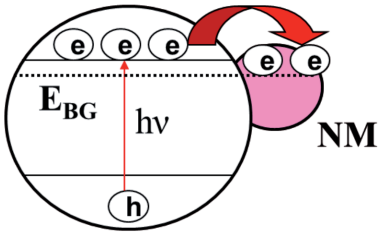


Fig. 11. Efficiente separazione di carica in semiconduttori modificati con nanoparticelle di metalli nobili (NM).

Recente è lo studio delle eterogiunzioni fra bismuto vanadato e triossido di tungsteno per la fabbricazione di fotoanodi per celle solari (*Fig. 12*), in cui l'efficienza di conversione in corrente della luce a diverse lunghezze d'onda è stata razionalizzata sulla base della dinamica

ultraveloce dei transienti fotogenerati nel materiale, studiata mediante spettroscopia di assorbimento al femtosecondo [19,20].

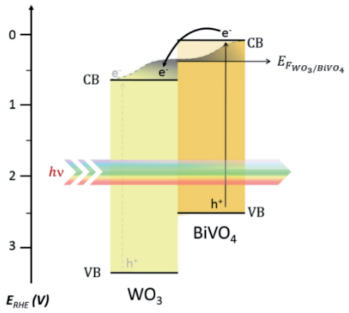


Fig. 12. Eterogiunzione bismuto vanadato-triossido di tungsteno per la fabbricazione di fotoanodi.

Sono stati infine sviluppati dispositivi per la scissione fotocatalica dell'acqua, a partire da una cella a due scomparti per l'evoluzione separata di idrogeno e ossigeno da soluzioni acquose (*Fig. 13*). L'elemento fondamentale di tale cella è costituito da un elettrodo di titanio ricoperto da un lato di TiO_2, che funge da fotoanodo, mentre gli elettroni fotopromossi nel sistema migrano attraverso l'elettrodo all'altro scomparto della cella dove avviene lo sviluppo di idrogeno per riduzione dei protoni dell'acqua [21]. I più recenti sviluppi di tale elettrodo consistono nell'utilizzo di TiO_2 in forma di nanotubi ottenuti per anodizzazione di una lastra di titanio [22], che hanno mostrato anche interessanti proprietà di cristalli fotonici [23] (*Fig. 14*).

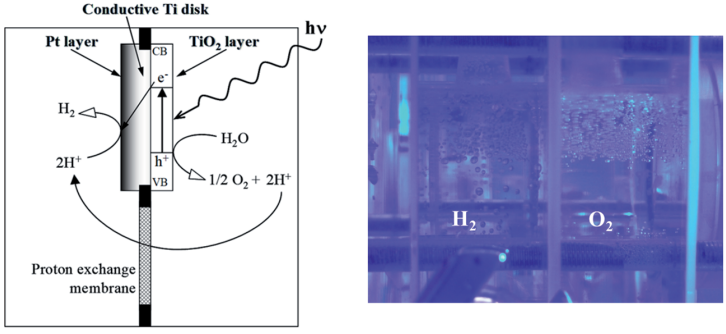


Fig. 13. Fotoelettrodo della cella a scomparti separati e immagine della cella stessa.

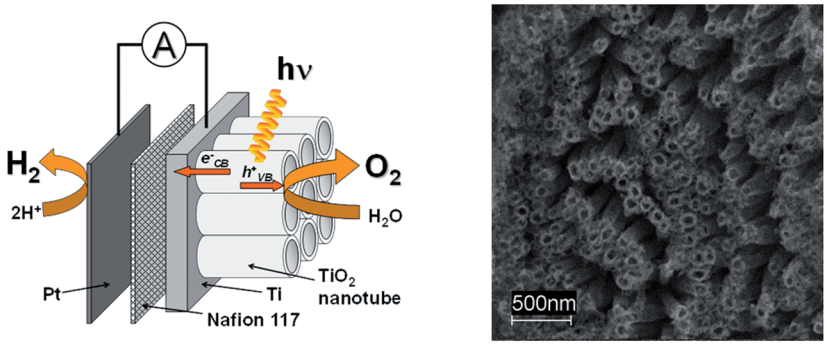


Fig. 14. Cella fotoelettrocatalitica per la scissione dell'acqua in idrogeno e ossigeno e simultanea misura della fotocorrente generata. Il materiale fotocatalitico è costituito da nanotubi di TiO_2 (immagine a destra) cresciuti per anodizzazione di una lastra sottile di titanio.

Conclusioni

L'energia solare è abbondante e inesauribile, è diffusa su tutta la terra e può essere trasformata in altre forme di energia utilizzabili dall'uomo attraverso una serie di tecnologie, anche relativamente semplici, che non implicano la possibilità di gravi incidenti, né un cattivo uso per scopi militari, ma che necessitano di ulteriori sviluppi per divenire competitive rispetto all'utilizzo dei combustibili fossili.

La chimica ha contribuito e sta contribuendo largamente a risolvere il problema energia principalmente giocando un ruolo chiave nello sviluppo delle tecnologie correlata con le tre strade di conversione dell'energia solare (calore solare, elettricità solare, combustibili solari), due delle quali implicano processi chimici, che richiedono lo sviluppo di materiali speciali. Le *performances* dei sistemi di assorbimento di luce, conversione e immagazzinamento di energia potranno essere largamente migliorate attraverso la progettazione razionale e l'applicazione sperimentale di nuovi materiali e processi. Ci stiamo lavorando in molti. Trovare una o più vie di uscita per risolvere il problema energetico e mitigare i problemi ambientali correlati con l'uso di combustibili fossili è una grande sfida per la chimica.

Ringraziamenti

I risultati delle ricerche sui materiali fotocatalitici qui brevemente presentati sono stati ottenuti in parte nell'ambito del progetto congiunto Regione Lombardia-Fondazione Cariplo *SmartMatLab*, di cui sono stata *Principal Investigator*, nonché grazie ad altri progetti di ricerca su materiali e dispositivi foto(elettro)catalitici finanziati da Fondazione Cariplo e dal Ministero dell'Università e della Ricerca (progetti PRIN).

Bibliografia

1. Armaroli N, Balzani V, The Future of Energy Supply: Challenges and Opportunities, *Angew. Chem. Int. Ed.*, 2007: 46: 52-66.
2. Kerr RA, A worrying trend of less ice, higher seas, *Science*, 2006: 311: 1698-1701.
3. Grätzel M, Solar energy conversion by dye-sensitized photovoltaic cells, *Inorg. Chem*., 2005: 44: 6841-6851.
4. Collings A, Critchley C, *Artificial Photosynthesis*, Wiley-VCH, Weinheim, 2005.
5. Ciamician G, The photochemistry of the future, *Science*, 1912: 36: 385-394.
6. Dozzi MV, Selli E, Doping TiO_2 with p-Block Elements: Effects on Photocatalytic Activity, *J. Photochem. Photobiol. C. Photochem. Rev.*, 2013: 14: 13-28.
7. Dozzi MV, Ohtani B, Selli E, Absorption and Action Spectra Analysis of Ammonium Fluoride-doped Titania Photocatalysts, *Phys. Chem. Chem. Phys.*, 2011: 13: 18217-18227.
8. Dozzi MV, D'Andrea C, Ohtani B, Valentini G, Selli E, Fluorine-doped TiO_2 materials: photocatalytic activity vs. time-resolved photoluminescence, *J. Phys. Chem. C*, 2013: 117: 25586-25595.
9. Chiarello GL, Dozzi MV, Selli E, TiO_2–based materials for photocatalytic hydrogen production, *J. Energy Chem.*, 2017: 26: 250-258.
10. Chiarello GL, Selli E, Forni L, Photocatalytic hydrogen production over flame spray pyrolysis-synthesised TiO_2 and Au/TiO_2, *Appl. Catal. B: Environ.,* 2008, 84: 332-339.
11. Chiarello GL, Forni L, Selli E, Photocatalytic hydrogen production by liquid- and gas-phase reforming of CH_3OH over flame-made TiO_2 and Au/TiO_2, *Catal. Today*, 2009: 144: 69-74.
12. Dozzi MV, Prati L, Canton P, Selli E, Effects of gold nanoparticles deposition on the photocatalytic activity of titanium dioxide under visible light, *Phys. Chem. Chem. Phys.*, 2009: 11: 7171-7180.
13. Chiarello GL, Dozzi MV, Scavini M, Grunwaldt JD, Selli E, One step flame-made fluorinated Pt/TiO_2 photocatalysts for hydrogen production, *Appl. Catal. B: Environ*., 2014: 160-161: 144-151.
14. Dozzi MV, Chiarello GL, Pedroni M, Livraghi S, Giamello E, Selli E, High photocatalytic hydrogen production on Cu(II) pre-grafted Pt/TiO_2, *Appl. Catal. B: Environ*., 2017: 209: 417-428.

15. Chiarello GL, Aguirre MH, Selli E, Hydrogen production by photocatalytic steam reforming of methanol on noble metal modified TiO_2, *J. Catal.*, 2010: 273: 182-190.
16. Chiarello GL, Ferri D, Selli E, Effect of the CH_3OH/H_2O ratio on the mechanism of the gas-phase photocatalytic reforming of methanol on noble metal-modified TiO_2, *J. Catal.*, 2011: 280: 168-177.
17. Riboni F, Selli E, Hoffmann MR, Colussi AJ, Homogeneous Reduction of CO_2 by Photogenerated Pyridinyl Radicals, *J. Phys. Chem. A*, 2015: 119: 4433-4438.
18. Grigioni I, Dozzi MV, Bernareggi M, Chiarello GL, Selli E, Photocatalytic CO_2 reduction vs. H_2 production: the effects of surface carbon-containing impurities on the performance of TiO_2-based photocatalysts, *Catal. Today*, 2017: 281: 214-220.
19. Grigioni I, Stamplecoskie KG, Selli E, Kamat PV, Dynamics of Photogenerated Charge Carriers in $WO_3/BiVO_4$ Heterojunction Photoanodes, *J. Phys. Chem. C*, 2015: 119: 20792-20800.
20. Grigioni I, Stamplecoskie KG, Jara DH, Dozzi MV, Oriana A, Cerullo G, Kamat PV, Selli E, Wavelength-Dependent Ultrafast Charge Carrier Separation in the $WO_3/BiVO_4$ Coupled System, *ACS Energy Lett.*, 2017: 2: 1362-1367.
21. Selli E, Chiarello GL, Quartarone E, Mustarelli P, Rossetti I, Forni L, A photocatalytic water splitting device for separate hydrogen and oxygen evolution, *Chem. Commun.*, 2007: 5022-5024.
22. Altomare M, Pozzi M, Allieta M, Bettini LG, Selli E, H_2 and O_2 photocatalytic production on TiO_2 nanotube arrays: effect of the anodization time on structural features and photoactivity, *Appl. Catal. B: Environ.*, 2013: 136-137: 81-88.
23. Yoo, JE, Lee K, Altomare M, Selli E, Schmuki P, Self-Organized Arrays of Single-Metal Catalyst Particles-in-TiO_2 Cavities: A Highly Efficient Photocatalytic System, *Angew. Chem., Int. Ed.*, 2013: 52: 7514-7517.

Istituto Lombardo (Rend. Scienze) 151, 201-218 (2017)

MARGHERITA D'AUSTRIA MUSICA E IDENTITÀ DI UNA PRINCIPESSA DEL RINASCIMENTO[1]

VINCENZO BORGHETTI (*)

Nota presentata dal m.e. Marino Gatto
(Adunanza del 26 ottobre 2017)

SUNTO. – Il presente saggio si concentra sullo *chansonnier* di Margherita d'Austria BrusBR 228 come oggetto cruciale per comprendere la politica culturale di Margherita, al di là delle sue funzioni più strettamente musicali. Sulla strada tracciata da alcuni recenti studi sull'argomento, nelle pagine che seguono propongo alcune riflessioni sul ruolo di questo manoscritto nella costruzione e performance dell'identità di Margherita come lettrice e come sovrana. A tale scopo mi soffermo prima sull'analisi dello *chansonnier* come libro di musica, cioè come un medium formato dall'interazione di diversi codici comunicativi, dedicando particolare attenzione alle pagine iniziali del manoscritto. Sulla base di questa analisi, formulo infine alcune ipotesi sia sull'uso e il significato di BrusBR 228, sia più in generale sulle abitudini di lettura di Margaret e sulle possibili finalità ideologiche e i significati culturali di questa sua attività.

***

ABSTRACT. – In this essay I concentrate on Margaret of Austria's chansonnier, BrusBR 228, reflecting on the role of this manuscript in the construction and performance of Margaret's identity as a princely reader and as a ruler. Following the path opened by recent studies on this manuscript, I propose an analysis of the chansonnier as a music book, that is, as an object constructed through the interaction of different communicative codes, pondering some elements particularly interesting for my perspective, such as the initial pages. On the basis of this analysis, I advance new hypotheses on the uses and meanings of BrusBR 228, as well as, more in general, on Margaret's reading habits and their potential cultural meanings and ideological ends.

(*) Università degli Studi di Verona, Italy. E-mail: vincenzo.borghetti@univr.it

1 Questo saggio è una versione concetrata di un una ricerca più vasta dedicata

L'argomento di questo saggio è l'analisi di un caso di mecenatismo musicale nel Rinascimento: la musica, o meglio, un libro di musica prodotto per Margherita d'Austria (1480-1530), duchessa di Savoia e reggente dei Paesi Bassi. La musica era una parte importante del lessico della preminenza condiviso dalle corti europee dell'epoca, pertanto appare oggi scontato che uno storico della musica del Rinascimento si applichi allo studio della musica destinata a una principessa dell'epoca. Questo saggio, però, non si occupa di mecenatismo musicale, non nel senso corrente della parola. Non è, infatti, «la protezione concessa generosamente alle arti e alle lettere» al centro delle pagine che seguono.[2] Anzi, lo stesso concetto di «mecenatismo» così come sopra definito è qui fuori luogo. La «protezione concessa» alla musica e alle arti in genere da parte esponenti della classe dominante non ha mai, e non solo nel Rinascimento, nulla di generoso, almeno non nell'accezione idealistica che connota il termine da alcuni secoli a questa parte. Questa «protezione» è, può essere, concessa sì con larghezza, ma non è solo frutto di magnanimità: essa è piuttosto il risultato di scelte interessate, che si servono del *bello* avendo però come fine l'*utile* – che invece l'uso comune del termine «mecenatismo» oggi esclude. Nelle pagine che seguono mi occuperò di questo tipo di uso della musica, analizzando il caso di Margherita d'Austria e illustrando, per mezzo di un suo *chansonnier*, le finalità del suo "mecenatismo" musicale.

Margherita d'Austria: l'ambizione femminile in una società patriarcale

Essere donne nel Rinascimento non era facile. Non lo era neppure se, come Margherita d'Austria (1480-1530), si era donne di altissimo

alla musica di Margherita D'Austria, e in particolare al suo *chansonnier*, i cui risultati si leggono in: V. Borghetti, *La sovrana lettrice: Margherita d'Austria e il suo chansonnier (Bruxelles, Bibliothèque royale de Belgique, Ms 228)*, in *Cara scientia mia, musica. Studi per Maria Caraci Vela*, a cura di A. Romagnoli, D. Sabaino, R. Tibaldi e P. Zappalà, Pisa, Edizioni ETS, 43-63 (Diverse voci, 14); e in V. Borghetti, *Trauer, Hoffnung der Frauen. Tod, Leiden und Macht in Margarete von Österreichs Chansonnier Brüssel, BR 228*, Die Tonkunst, (2017), 31-38.

[2] F. Palazzi, G. Folena, «mecenatismo», in *Dizionario della lingua italiana*, Firenze, Loescher, 1995, 1083.

rango. Nipote dell'Imperatore Federico III, figlia di Massimiliano d'Asburgo (poi a sua volta imperatore come Massimiliano I) e dell'erede del Ducato di Borgogna Maria di Borgogna, Margherita d'Austria (1480-1530) era uno dei migliori partiti d'Europa; il suo destino "naturale" sarebbero state nozze reali.[3] Come era all'epoca costume, fin dai primi anni di vita iniziarono a circolare suoi ritratti che avevano lo scopo di promuoverla tra le case regnanti d'Europa.[4] Questi dipinti, come quello oggi conservato alla National Gallery di Londra (illustrazione 1) mostravano la (futura) sposa ancora bambina, e ne mettevano in evidenza la bellezza come prova di sana e robusta costituzione, e quindi di predisposizione alla maternità, requisito fondamentale per assicurare la nascita di eredi.[5] Esibivano inoltre, come in questo caso, anche i vantaggi territoriali che il futuro consorte avrebbe conquistato sposandola. Rispetto però ad un analogo ritratto promozionale per un giovane principe dell'epoca (per esempio uno di quelli confezionati per suo fratello maggiore Filippo il Bello, nato nel 1478), i ritratti presentavano le principesse come persone a metà. Il ritratto di Margherita non potrebbe essere più esplicito, da questo punto di vista. In quanto donna, Margherita, non era considerata una persona completa, e il suo stemma lo rivela in modo inequivocabile: è vuoto nella sua parte sinistra, la prima. Solo un matrimonio avrebbe riempito questo vuoto, e avrebbe così completato lei, garantendole non solo un ruolo da sovrana, ma soprattutto uno *status* definito, *status* che, in quanto donna, Margherita da sola non avrebbe potuto conseguire.[6]

---

[3] Per una sintesi delle vicende storico-biografiche di Margherita d'Austria rimando alla cronologia di D. Eichberger, *Margaret of Austria 1480-1530*: "Fortune infortune fort une", e al suo *Margareta of Austria. A Princess with Ambition and Political Insight*, entrambi in *Women of Distinction: Margaret of York*, *Margaret of Austria*, ed. by D. Eichberger, Leuven, Brepols-Davidsfonds, 2002, risp. 26-27, e 49-55.

[4] Sull'iconografia dei ritratti di bambini e sul loro uso e significato nella Francia e nei Paesi Bassi all'epoca di Margherita si veda P. Lorentz, *Children's Portraits. Between Politics and Family Memories*, in *Women of distinction*, 115-123; sui ritratti giovanili di Margherita si veda D. Eichberger, *Leben mit Kunst, Wirken durch Kunst. Sammelwesen und Hofkunst unter Margarete von Österreich, Regentin der Niederlande*, Turnhout, Brepols, 2002 (Burgundica, V), 29-33.

[5] Bottega di Pieter van Conixloo, *Dittico con Filippo il Bello e Margherita d'Austria* (circa 1495), part., The National Gallery, London, inv. n. 2613.

[6] C'era in verità una via alternativa: la monacazione. Sebbene spirituale, anche

La costruzione dell'identità di genere per le donne dell'epoca era fondata sulla loro "naturale" imperfezione; essa traeva origine dalle sacre scritture e alimento da secoli di teologia e trattatistica misogine, e non faceva sconti nemmeno per le classi al potere. La costruzione di un'identità di genere basata sull'imperfezione ontologica della donna e sulla necessità femminile del completamento e del controllo maschile si perseguiva solo con la pedagogia esplicita dei testi dedicati a questi argomenti; essa era bensì realizzata - come oggi, come sempre - per mezzo di discorsi che, sotto forma di testi, enunciati, pratiche, immagini, oggetti e così via, pervadevano l'esistenza di donne e uomini educandone il pensiero e i comportamenti, e avevano tanta più efficacia quanto più si presentavano come ovvi e "naturali".



Fig. 1.

Il ritratto di Margherita della National Gallery sopra mostrato, è un esempio di questa pervasiva discorsività di genere: un'immagine che

questa era però una forma di matrimonio con una figura maschile; anche questa era dunque una forma di completamento dato da un uomo. Agli Asburgo, come è risaputo,

proclamava al mondo e ricordava a Margherita che lei era sì una principessa di ottimi natali, ma che era sempre una donna, una persona "incompleta" fintantoché il suo stemma fosse restato per metà vuoto, e che in quanto donna avrebbe potuto ambire solo ad essere una sposa, soggetta cioè al controllo di un marito, che le avrebbe conferito quella "completezza" di cui lei per "natura" doveva considerarsi priva (*Fig. 1*).

La biografia di Margherita per i primi decenni della sua vita coincide con la ricerca di questa "completezza" (e dei vantaggi politici che alla sua famiglia ne sarebbero derivati). A soli tre anni, Margherita venne promessa al futuro re di Francia, Carlo VIII, e partì alla volta della corte francese dove sarebbe stata educata come futura regina. Tuttavia le nozze con Carlo non avrebbero mai avuto luogo. Nel 1491 Carlo VIII decise di sposare Anna di Bretagna, rompendo così gli accordi precedenti, di conseguenza nel 1493 Margherita fece ritorno nei Paesi Bassi. Nel 1495 suo padre Massimiliano arrangiò per lei un nuovo matrimonio: Margherita avrebbe sposato Giovanni, primogenito di Ferdinando d'Aragona e Isabella di Castiglia. Celebrato nel 1497, il matrimonio ebbe tuttavia vita brevissima: Giovanni morì nel 1497 a pochi mesi dalle nozze. Nel 1501, Margherita si sposò ancora, stavolta con il duca di Savoia Filiberto II. Anche Filiberto, però, morì prematuramente (come Margherita era nato nel 1480) nel 1504, solo tre anni dopo la loro unione. All'età di 24 anni Margherita aveva già sulle sue spalle un ripudio e due vedovanze. A questo punto i piani del padre Massimiliano, e del fratello Filippo il Bello prevedevano per lei un ulteriore matrimonio, le cui trattative giunsero presto a buon punto: il re d'Inghilterra Enrico VII sarebbe stato il terzo marito di Margherita. Tuttavia, stavolta Margherita riuscì ad imporre il suo rifiuto. Quello con Filiberto sarebbe rimasto il suo ultimo matrimonio.[7] Non però l'ultimo

---

questa soluzione piaceva decisamente meno: vista la proverbiale propensione ad usare i matrimoni per ampliare i propri domini e le sfere di influenza, tendevano a far sposare il numero maggiore degli eredi, legittimi o illegittimi che fossero. Sulla condizione femminile all'epoca di Margherita d'Austria si veda G. Duby e M. Perrot, *Storia delle donne: il Medioevo*, a cura di C. Klapisch-Zuber, Roma-Bari, Laterza, 1990, in particolare C. Opitz, *La vita quotidiana delle donne nel tardo Medioevo (1250-1500)*, 330-402.

[7] Sul progetto di far sposare Margherita con Enrico VII si veda R. Wellens, *Un épisode des relations entre l'Angleterre et les Pays-Bas au début du XVIe siècle: le projet de mariage entre Marguerite d'Autriche et Henri VII*, Revue d'histoire moderne et contemporaine, 39/2 (1982), 267-290.

lutto importante della sua vita: nel 1506, infatti, morì, anch'egli prematuramente, suo fratello Filippo, il quale lasciava una moglie psicologicamente instabile (Giovanna "la Pazza"), un erede minore (Carlo, il futuro Carlo V), e, di conseguenza, un pericoloso vuoto dinastico.[8]

Gli eventi luttuosi furono per Margherita la fonte dei suoi dolori e al contempo delle sue fortune. A quell'epoca, infatti, la vedovanza, se senza figli maschi, poteva garantire ad una donna quell'accesso ai beni e alle proprietà derivanti dal matrimonio che non avrebbe potuto avere vivente il marito o un erede (maschio). Le vedove potevano cioè godere di quell'autonomia che era invece negata alle donne sposate o nubili, sempre soggette al controllo del marito, del figlio maschio adulto, o del padre o, in assenza, dell'autorità familiare. Fu quindi anche perché forte delle proprie disponibilità dotali che, nei lunghi negoziati con Enrico VII, Margherita, già vedova due volte, si trovò nelle condizioni di poter fare la scelta di non risposarsi.[9]

Tuttavia, non si trattava di un'autonomia di scelta fondata solo su basi economiche. Nella società dell'epoca la vedova godeva di uno *status* privilegiato tra le donne. Una tradizione che risaliva alla Bibbia sottolineava che, in precise circostanze, la condizione vedovile fosse associabile al possesso di valori morali positivi. Qualora la vedova fosse stata pia, fosse rimasta casta e devota alla memoria dello sposo defunto, si sarebbe votata ad una forma di monacazione non (del tutto) claustrale, assurgendo così a esempio di virtù non solo per le altre donne ma anche per gli uomini (per la sua dimostrata capacità di sopportare i colpi della sorte, degna di ammirazione perché dimostrata da una creatura "per natura" debole come una donna). A causa della sua statura morale, la figura della vedova era in quest'epoca spesso sovrapposta a quella della *virago*, l'eroina virile, come la Giuditta biblica, avrebbe saputo all'occorrenza lottare per difendere il proprio popolo e la propria famiglia.[10]

Dopo la morte di Filiberto, Margherita si trovò dunque nella

---

[8] Insieme a quella dell'erede Carlo, era necessario occuparsi dell'educazione delle sue sorelle rimaste nei Baesi Bassi (Eleonora, Isabella e Maria), vedi a riguardo Eichberger, *Margareta of Austria*, 51-52.

[9] Sulla condizione vedovile all'epoca di Margherita e sul suo caso specifico si veda B. Welzel, *Widowhood: Margaret of York and Margaret of Austria*, in *Women of Distinction*, 103-113.

[10] B. Welzel, *Widowhood*, 104-105.

"felice" posizione di essere vedova per la seconda volta. E proprio su questa reiterata vedovanza fece leva non solo per evitare un altro matrimonio (e quindi un marito-padrone), ma anche per riuscire ad ottenere quelle investiture politiche che la sua condizione avrebbe potuto adesso permetterle. Così, dopo la morte di Filiberto Margherita incentrò il suo discorso su questo suo *status*. Una delle sue prime azioni fu l'adozione di un nuovo motto che la identificase come vedova, cioè come donna virtuosa e forte colpita da sorte avversa (almeno secondo una delle sue possibili interpretazioni): «Fortune Infortune Fortune (o Fort Une)».[11]

Per Margherita conquistare un ruolo politico ufficiale non fu comunque facile, nonostante i vantaggi dati dall'essere vedova due volte. Nel 1508 il padre Massimiliano la nominò governatrice dei Paesi Bassi con la custodia dei suoi nipoti (i figli di Filippo il Bello e Giovanna la Pazza), poi reggente dal 1509. Nel 1515, dichiarato maggiorenne suo nipote Carlo, Margherita dovette cedergli il governo dei Paesi Bassi, ma nel 1519 lo stesso Carlo la reinstallò come reggente e governatrice, incarico confermato nel 1520 e che Margherita conservò fino alla sua morte nel 1530.[12]

Leggere per comandare

Come tutte le donne del suo rango, Margherita leggeva e possedeva libri; come vedova con ambizioni di potere, però, Margherita leggeva e collezionava libri in modi e con finalità diverse dalle principesse sue contemporanee. In Francia, sotto la tutela di Anna de Beaujeu, figlia di Luigi XI, e, in seguito, di Madame du Segré, le sue letture furono i testi *standard* per le principesse dell'epoca, testi dai quali Margherita avrebbe imparato a condursi come una regina cristiana, buona madre e moglie, saggia, umile e non incline ai piaceri mondani. Anne-Marie Legaré ricorda che, poco dopo il suo arrivo ad Amboise,

---

[11] Sul motto di Margherita si vedano D. Eichberger, *Leben mit Kunst*, 25-18; B. Welzel, *Widowhood*, 103-105.

[12] Sulle difficoltà che Margherita si trovò a fronteggiare per ottenere e conservare il suo ruolo di reggente si veda H. Meconi, *Margaret of Austria, Visual Representation, and Brussels, Royal Library*, *Ms. 228*, Journal of the Alamire Foundation, II/1 (2010), 23-24.

Margherita ricevette dalla tutrice uno dei testi-base dell'educazione delle principesse dell'epoca: il *Livre des Espitres et Euvangille de tout l'an*. Letture morali, devote, manuali di comportamento erano, come la studiosa sottolinea, quelle comuni e condivise all'epoca tra le principesse del nord Europa.[13]

La biblioteca che Margherita avrebbe costruito nel corso della sua vita, tuttavia, si distingueva da quelle che aveva conosciuto nei suoi anni di formazione, anzitutto per le dimensioni. Da adulta, infatti, Margherita era nota per possedere molti più libri rispetto alle principesse contemporanee, e per questo la sua biblioteca poteva competere con quelle dei famosi bibliofili della sua epoca (tutti rigorosamente uomini). È indicativo che Antonio de Beatis, quando nel 1516 al seguito del cardinale Luigi d'Aragona visitò la biblioteca di Margherita nel suo palazzo di Mechelen, non poté fare a meno di sottolineare l'eccezionalità della collezione proprio per essere opera di una donna. Nel suo diario di viaggio scrisse in proposito: «Llì [nella grande piazza della città] se vedde la casa di Madama Margarita, qual è molto bella et assai in ordine, benché non di molta vista, dove è una libraria per donne assai ornata e ricca».[14] La biblioteca di Margerita si distingueva poi per il suo contenuto. A parte le opere devozionali e morali tipiche delle biblioteche femminili, Margherita acquisì e commissionò soprattutto libri profani, di argomento storico, genealogico, cavalleresco, raccogliendo una collezione ragguardevole per la sua vastità, ma anche per la varietà delle materie che vi erano rappresentate, e differenziando così i propri interessi di lettrice e bibliofila da quelli comuni tra le principesse del suo tempo.[15]

Leggere e collezionare tanti libri - di così tanti generi e materie - erano attività all'epoca connotate in senso maschile, come indirettamente notava già de Beatis, e per questo furono necessarie a Margherita per costruire e consolidare ai suoi occhi e quelli del mondo la propria

---

13 A. M. Legaré, '*La librairye de Madame*', 214-215.

14 Il testo di Antonio de Beatis si legge nell'originale volgare in L. Pastor, *Die Reise des Kardinals Luigi d'Aragona durch Deutschland, die Niederlande, Frankreich und Oberitalien, 1517-1518, beschrieben von Antonio de Beatis*, Freiburg i. B., Herderische Verlagshandlung, 1905, 133. Secondo Legaré le uniche collezioni contemporanee paragonabili a quella di Margherita erano quelle del cardinale Alberto di Brandeburgo o del duca Federico il Saggio di Sassonia ('*La librairye de Madame*', 210).

15 Si veda a riguardo A. M. Legaré, '*La librairye de Madame*', 210-217.

identità di sovrana, di sovrana in quanto vedova, cioè di donna che può assumere ruoli tradizionalmente riservati agli uomini.[16] L'identità di vedova era quindi un presupposto per poter essere e sentirsi sovrana, e la biblioteca (sia come spazio sia come collezione), i libri e gli oggetti in essa contenuti furono alcuni tra i *media* con cui Margherita costruì e promosse il discorso di queste sue due identità. Chi si fosse recata o recato nella biblioteca di Margherita, inclusa la stessa Margherita, avrebbe potuto ripercorrere la storia dietro al suo ruolo di reggente e/o alla legittimità di questa sua aspirazione, vedendola prima sposa di Filiberto (nei busti di Conrad Meit che la ritraevano insieme al suo ultimo marito),[17] poi vedova devota alla sua memoria (l'armatura di Filiberto, e il busto, sempre di Meit, di Margherita in abito da vedova, entrambi esposti in stanze dove c'erano libri).[18] Poi avrebbe potuto vedere questa stessa storia narrata con grande varietà di modi e linguaggi nei suoi tanti e diversi libri, in particolare in quelli - da lei stessa commissionati e in cui a volte lei stessa compariva come autrice - che illustravano la sua condizione di donna virtuosa e colpita dalla fortuna, i libri cioè che ribadivano il suo diritto ad essere e promuoversi come persona degna del potere di cui era già - o avrebbe voluto essere – investita (tra cui per esempio Michele Riccio, *Changement de fortune en toute prosperité*, Jean Lemaire des Belges *La Couronne margaritique*, Jean Franco, *La genealogie de l'empereur Charles V*).[19]

---

16 *ibidem*.

17 I busti di Meit si trovano oggi al British Museum di Londra, Waddeston Bequest.

18 Come i busti di Meit sopra menzionati, l'armatura di Filiberto era esposta nello spazio ufficiale della biblioteca (vedi a riguardo D. Eichberger, *Leben mit Kunst*, 124-132; B. Welzel, *Widowhood*, 108-109). Il busto di Conrad Meit che ritrae Margherita da vedova (oggi al Bayerisches Nationalmuseum di Monaco, inv. n. R 420.) era invece esposto nel *petit cabinet*, lo studiolo, uno spazio più intimo rispetto alla biblioteca, e dove Margherita conservava alcuni libri, nello specifico otto manoscritti (su quest'ultimo busto di Meit e sugli spazi in cui era esposto si veda J. L. Burk, *Conrat Meit, Court Sculptor to Margaret of Austria*, in *Women of Distinction*, 284; D. Eichberger, *Leben mit Kunst*, 109-112.

19 Rispettivamente i manoscritti Wien, Österreichische Nationalbibliothek, Ms. 3441; Wien, Österreichische Nationalbibliothek, Cod. 3441; Paris, Bibliothèque Nationale de France, Ms. f. fr. 5616.

Leggere musica per comandare

Tra i libri che Margherita investì di un ruolo cruciale nella definizione e promozione del suo discorso di vedova/sovrana c'era anche un libro di musica: lo *chansonnier* Bruxelles, Bibliothèque Royale Ms 228 (in seguito BrusBR 228), confezionato per lei nello *scriptorium* di corte dal grande calligrafo musicale Petrus Alamire.[20] Non era l'unico libro di musica che Margherita possedesse. Gli inventari della sua biblioteca registrano infatti un altro *chansonnier* (BrusBR 11239), un grande libro corale con messe polifoniche (forse identificabile nel codice BrusBR 6428) e una raccolta di bassedanze (BrusBR 9085).[21] Mentre però questi ultimi erano doni, BrusBR 228 era l'unico che Margherita stessa avesse commissionato, e, come tale, esibisce caratteristiche che lo distinguono non solo dagli altri libri di musica della collezione, ma anche rispetto a libri analoghi dell'epoca.[22] Su queste caratteristiche vorrei adesso soffermarmi.

Nelle pagine d'apertura di BrusBR 228 (*Fig. 2*) compaiono il ritratto di Margherita e il suo stemma; alcuni dei motivi ricorrenti della decorazione sono le perle e soprattutto i fiori di margherita (in francese entrambi *marguerites*). Altri segni di appartenenza si trovano nel contenuto poetico e musicale: tra le composizioni alcune intonano suoi testi (si

---

[20] Della vasta bibliografia su BrusBR 228 ricordo qui l'articolo più recente: H. Meconi, *Margaret of Austria*, 11-36. Lo studio più completo su questo *chansonnier* è sempre *The Chansons Albums of Marguerite of Austria: Mss 228 and 11239 of the Bibliothèque Royale de Belgique, Brussels: A Critical Edition and Commnentary*, ed. by M. Picker, Berkeley-Los Angeles, University of California Press, 1965. Del manoscritto esiste un'edizione in facsimile: *Album de Marguerite d'Autriche: Brussel, Koninklijke Bibliotheek MS. 228*, introduction by M. Picker, Peer, Alamire, 1997.

[21] Sui tre manoscritti nella biblioteca di Margherita si veda M. Debae, *La bibliothèque de Marguerite d'Autriche. Essai de reconstitution d'après l'inventaire de 1523-1524*, Louvain-Paris, Peeters, 1995; i tre manoscritti sono ciatati rispettivamente 480-482, 61-62 e 457-460.

[22] BrusBR 11239 era un dono di Filiberto di Savoia (fatto verosimilmente prima delle nozze). BrusBR 9085 era un dono di Anna di Clèves, mentre il grande libro corale con messe polifoniche, forse BrusBR 6428 era un dono di suo padre Massimiliano. Sull'origine dei manoscritti e sui modi in cui giunsero nelle collezioni di Margherita si veda Debae, *La bibliothèque de Marguerite d'Autriche* alle pagine indicate nella nota precedente; in particolare sugli *chansonniers* polifonici si veda H. Meconi, *Margaret of Austria*, 21-23.

ipotizza anche alcune delle musiche possano essere sue), altre commemorano la scomparsa di persone di famiglia - *Proch Dolor/Pie Jesu* sulla morte di suo padre Massimiliano; *Se je souspire/Ecce iterum* per quella di suo fratello Filippo - o di membri della corte a lei vicini - *Cueurs desolez* sulla morte del consigliere Jean de Luxembourg - o del suo amatissimo pappagallo verde - *Soubz ce tumbel*. Come per altri libri commissionati da Margherita, i temi e i toni dominanti sono quelli del lutto, del dolore, del rimpianto; qualche composizione ha argomento e carattere diversi, ma come nel caso di *Plus nulz regretz*, che celebra l'alleanza degli Asburgo con l'Inghilterra nel Trattato di Calais del 1507, si riferisce ad avvenimenti politici che la coinvolgevano insieme alla Casa d'Austria.[23]
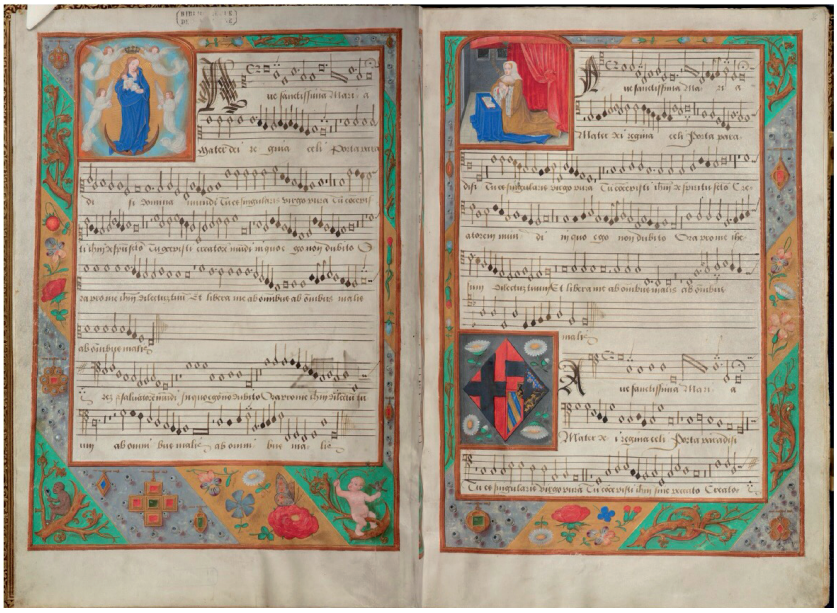


Fig. 2.

Che questo manoscritto fosse sotto molti aspetti diverso dagli altri libri di musica di Margherita, e, di conseguenza, avesse tra questi uno *sta-*

[23] Sulla paternità dei testi, sui loro dedicatari così come sulle occasioni che ne sono all'origine si veda a M. Picker, *The Texts of MS 228*, in *The Chansons Albums of Marguerite of Austria*, 52-57; H. Meconi, *Margaret of Austria*, 11-12.

tus speciale nella sua biblioteca, lo si ricava già dall'inventario del 1523, dove BrusBR 228 è definito «grant livre de chant».[24] Non è questa la definizione usata per l'altro *chansonnier* della collezione, BrusBR 11239, che l'inventario identifica invece come «petit livre de chant».[25] Nonostante siano simili per contenuto, i due manoscritti hanno dimensioni molto diverse: BrusBR 11239 misura 20x14,5 cm, mentre BrusBR 228 36,5x26 cm.[26] All'epoca gli *chansonniers* erano libri piccoli, un analogo musicale dei libri d'ore. Rispetto ai formati molto ridotti degli anni Settanta del Quattrocento, agli inizi del secolo successivo le dimensioni degli *chansonniers* crescono leggermente, tuttavia, anche tenendo conto di queste trasformazioni, BrusBR 228 resta un libro che eccede gli *standard* della tipologia libraria di appartenenza, come non si verifica invece con BrusBR 11239. È significativo che l'inventario usi una locuzione simile a quella di BrusBR 228 non per l'altro *chansonnier* della biblioteca, bensì per il grande libro corale con messe polifoniche, registrato come «ung grant covers de cuyr qui se dit Livre de chant[…]».[27] Con BrusBR 228 Alamire produsse per Margherita un libro di musica *sui generis*, in cui il repertorio di uno *chansonnier* è affidato non a un libro di piccolo formato, bensì ad un codice di grandi dimensioni come un libro corale.[28]

---

24 M. Debae, *La bibliothèque de Marguerite d'Autriche*, 11: «Item ung aultre grant livre de chant, couvert de velours tanner, a cloz dorez, et an la fin il y a que souvent change couleur commenceant Ave Sanctissima Maria».

25 M. Debae, *ivi*, 480: «Item, ung aultre petit, qui ce nomme Livre de champs. [sic = chants]».

26 Per le dimensioni di BrusBR 228 si veda *The Treasury of Petrus Alamire. Music and Art in Flemish Court Manuscripts 1500-1535*, ed. by H. Kellman, Ghent-Amsterdam, Ludion, 1999, 69; per quelle di BrusBR 11239 si veda la scheda del Digital Image Archive of Medieval Music (DIAMM) consultabile online <https://www.diamm.ac.uk/sources/1628/#/> (data accesso 26.11.2017).

27 M. Debae, *La bibliothèque de Marguerite d'Autriche*, 61: «Item ung grant couvert de cuyr qui se dit Livre de Champt que l'emepereur donna a madame, commenceant Kyrieleison. La premiere messe dudit Livre est de madame saincte Anne faicte par Pierchon de La Rue».

28 Un libro corale con musica sacra come il codice Chigi (Città del Vaticano, Biblioteca Apostolica Vaticana, Ms Chigi C VIII 234) misura 36,3x27,8 cm; dimensioni simili sono quelle del manoscritto Bruxelles, *Bibliothèque royale*, Ms 9126 (38,2x27 cm). Per i dettegli codicologici su questi manoscritti vedi *The Treasury of Petrus Alamire*, rispettivamente 73 e 125. La vicinanza del formato con i libri corali è notata anche da H. Meconi, *Margaret of Austria*, 19-20, e 22-23. Sul rapporti variabili tra

All'inizio di BrusBR 228 si trovano le due pagine più decorate dello *chansonnier*, dove il mottetto *Ave sanctissima Maria*, probabilmente opera di Pierre de la Rue, è accompagnato da una scena sacra, dal ritratto e dallo stemma di Margherita, ed è circondata da bordi miniati (vedi illustrazione 2). Si tratta di un uso frequente: spesso gli *chansonniers* dell'epoca riservavano l'apertura ad una composizione religiosa e, in genere, anche alla decorazione più ricca dell'intero libro. Si tratta per lo più di cosiddetti mottetti-cantilena, composizioni a tre voci, sia per formato che per stile simili alle composizioni profane di questi libri, ma con testi sacri o devozionali.[29] Aprendo BrusBR 228, Margherita, e chi con lei si fosse trovata o trovato a leggere dal manoscritto, avrebbe avuto di fronte una tipica situazione da *chansonnier*. Anzi, avrebbe avuto di fronte una situazione da *chansonnier* con un tratto esibito di modestia, poiché negli anni di BrusBR 228 anche per la musica profana lo *standard* erano ormai le quattro voci. Solo una lettura più attenta della musica avrebbe rivelato i *signa congruentiae*, svelando così che quello che appare come un mottetto-cantilena a tre voci è in realtà un grande mottetto canonico a sei voci, in cui in ciascuna delle tre voci ne contiene un'altra che legge la stessa parte, ma inizia dopo e ad un'altezza diversa da quella scritta, come indicano appunto i *signa congruentiae*. BrusBR 228 nelle sue pagine iniziali ripropone, quindi, a livello di *layout* il contrasto osservato tra le dimensioni del libro e il suo repertorio. L'apertura del codice ha l'aspetto visivo tradizionale di uno *chansonnier*, ma contiene un brano che per numero di voci e impegno compositivo eccede quello di un mottetto-cantilena, e che per queste sue caratteristiche avrebbe dovuto all'epoca essere consegnato idealmente ad un'altra tipologia di libro, cioè il libro corale di musica sacra, non ad uno *chansonnier*.

---

tipologia di libro, dimensioni, repertori e funzioni dei manoscritti musicali tra Quattro e Cinquecento si veda T. Schmidt-Beste, *Private of Institutional - Small or Big? Towards a Typology of Polypnonic Sources of Renaissance Music*, Journal of the Alamire Foundation, I/1 (2009), 13-26.

[29] Sull'*Ave regina celorum* di Frye e sul suo stile si veda D. J. Rothenberg, *The Flower of Paradise. Marian Devotion and Secular Song in Medieval and Renaissance Music*, Oxford-Newy York, Oxford University Press, 2011, 123-158; R. Strohm, *The Rise of European Music, 1380-1500*, Cambridge, Cambridge University Press, 1993, 394-397. In generale sulla cultura degli *chansonnniers* della seconda metà del Quattrocento si veda J. Alden, *Songs, Scribes, and Society. The History and Reception of the Loire Valley Chansonniers*, Oxford-New York, Oxford University Press, 2010.

Ave sanctissima Maria è accompagnata dal ritratto di Margherita e dal suo stemma. La committente compare all'inizio del *tenor* (fol. 2), inginocchiata in preghiera sotto un baldacchino, davanti ad un libro aperto. Le sue labbra pronunciano le parole «memento mei», rivolte alla Vergine col bambino *in sole* (l'iconografia dell'Immacolata, cui il testo di *Ave sanctissima Maria* si riferisce), che compare nella pagina a fronte all'inizio del *superius* (fol. 1v), e che Margherita può vedere attraverso la finestra davanti al suo inginocchiatoio (la miniatura ne mostra solo la cornice destra).[30] Il *bassus* è preceduto dallo stemma di Margherita (nella forma romboidale tipica per le donne) accompagnato da quattro margherite; nello stemma le sue armi d'Austria e di Borgogna seguono quelle del defunto marito Filiberto di Savoia (la croce d'argento su fondo rosso).

La presenza di Margherita come unica committente è di per sé una circostanza singolare nel contesto dei manoscritti musicali della corte dei Paesi Bassi. Petrus Alamire produsse, come è noto, diversi libri di musica di lusso nella cui decorazione compaiono i ritratti dei committenti. Quando compaiono committenti donne, queste sono presentate come mogli, sorelle, figlie, sempre cioè in relazione ad un'autorità maschile, ma non da sole.[31] Tuttavia la condizione e le ambizioni di Margherita non avevano corrispondenza in quelle di altre principesse della sua famiglia, che non sono raffigurate mai da sole nei codici di Alamire. In BrusBR 228 Margherita si mostra come una vedova osservante e devota alla memoria del marito: è ritratta in atto di preghiera alla Vergine, indossa il soggolo a pieghe tipico delle vedove. La presenza del marito è poi richiamata anche dallo stemma, che si riferisce al suo ultimo matrimonio: le armi del duca di Savoia sono sempre in prima posizione. Allo stesso tempo, però, Margherita si mostra anche come sovrana: il soggolo vedovile è infatti abbinato un abito cremisi (di cui si

[30] Sulle intonazioni di *Ave sanctissima Maria*, così come sull'iconografia dell'Immacolata in questo manoscritto si veda B. J. Blackburn, *The Virgin in the Sun: Music and Image for a Prayer Attributed to Sixtus IV*, Journal of the Royal Musical Association, 124/2 (1999), 157-195; Ead., *Messages in Minitures: Pictorial Programme and Theological Implications in the Alamire Choirbooks*, in *The Burgundian-Habsburg Court Complex of Music Manuscripts (1500-1535) and the Workshop of Petrus Alamire*, Colloquium Proceedings, Leuven 25-28 November 1999, ed. by B. Bouckaert, E. Schreurs, Leuven-Nerpelt, Alamire, 2003 (Yearbook of the Alamire Foundation, 5), 161-184.

[31] Vedi a riguardo H. Meconi, *Margaret of Austria*, 15-19.

spunta la parte finale delle maniche) sul quale Margherita indossa una ricca sopravveste dorata foderata di ermellino (si vedano per esempio i risvolti delle maniche).[32] Attraverso il ritratto e lo stemma, Margherita si mette in scena nelle prime pagine del suo *chansonnier* come principessa cristiana, sottolineando il suo *status* di vedova e, insieme e grazie a questo, la sua dignità ad un ruolo di potere.

La decorazione pittorica presenta altri elementi oltre al ritratto, lo stemma e la raffigurazione della Vergine col bambino *in sole*: nel bordo inferiore del fol. 1v compaiono nell'ordine una scimmia che si guarda allo specchio (nell'angolo sinistro) e un putto nudo a cavallo di un tronco, che in una mano stringe un ramo mentre sull'altra si posa un uccello con le ali spiegate (nell'angolo destro).[33]

Una tradizione iconografica che risale all'antichità, propagata poi dai bestiari medievali, attribuiva alla scimmia significati negativi e fino diabolici.[34] Raffigurate di frequente mentre scimmiottano azioni proprie degli esseri umani, nelle arti visive le scimmie hanno spesso una funzione di avvertimento morale diretto a chi le osserva: esse cioè mettono in guardia dal potenziale peccaminoso delle stesse attività in cui sono impegnate, perché condotte come un mero atteggiamento esteriore, ma senza avere cognizione del loro significato, come accade per l'appunto con le scimmie quando imitano i gesti degli esseri umani. La scimmia del fol. 1v di BrusBR 228 è dipinta nella pagina e nella posizione opposte rispetto al Margherita. Opposta a quella della committente è anche la sua occupazione. Mentre Margherita prega e così si rispecchia in uno specchio spirituale, lo *speculum sine macula* della Vergine Immacolata, la scimmia rimira solo se stessa in uno specchio terreno, che al contrario di quello di Margherita è «cum macula». In effetti, al centro dello spec-

---

[32] Sull'iconografia di Margherita come vedova e come sovrana si veda EICHBERGER, *Leben mit Kunst*, 33-42; B. Welzel, *Widowhood*; P. Matthews, *Apparel, Status, Fashion. Woman's Clothing and Jewellery*, in *Women of Distinction*, 147-153.

[33] Ne ho trattato brevemente in V. Borghetti, *The Listening Gaze: Alamire's Presentation Manuscripts and the Courtly Reader*, Journal of the Alamire Foundation, 7/1 (2015) (*Mise-en-Page in medieval and early modern music sources*, II, ed. by T. Schmidt-Beste and H. Vorholt), 47-66.

[34] Sull'iconografia e significati della scimmia si veda H. W. Janson, *Apes and Ape Lore in the Middle Ages and the Renaissance*, London, Warburg Institute, 1952 (Studies of the Warburg Institute, 20); si veda anche L. Amtower, *Engaging Words. The Culture of Reading in the Later Middle Ages*, New York, Palgrave, 2000, 57.

chio della scimmia è visibile una macchia scura, che è il riflesso realistico della scimmia e allo stesso tempo la sua "macchia" simbolica, frutto sia di un'attività non virtuosa sia, più in generale, della sua condizione terrena e bestialmente viziosa. Non a caso, per sottolineare la distanza tra le due scene di rispecchiamento, la scimmia è collocata agli antipodi di Margherita, nel bordo inferiore, in uno spazio basso, che è pertinente al basso livello morale della sua occupazione. Il putto è poi un altro esempio di umanità primitiva e imperfetta, e rafforza così il significato simbolico della scimmia.[35] Come questa, infatti, anch'esso si impegna in una attività piacevole ma vana: si direbbe giochi con quanto ha intorno nel bordo-giardino che lo ospita. Entrambe le creature non comprendono il valore spirituale di quanto li circonda, indifferenti alla scena sacra e alla preghiera in musica che li sovrasta, non si accorgono dell'apparizione della Vergine.

La scimmia e il putto di BrusBR 228 non sono figure neutre. Posti in apertura come esempi negativi di attività non volte ad un fine superiore, essi svolgono un'importante funzione, poiché costituiscono un *caveat* sia per Margherita, contro la tentazione di cedere ad un uso futile del suo bel *chansonnier*, sia per chiunque si fosse trovata o trovato a leggere il manoscritto, contro la tentazione di considerarlo solo un bel libro per ricche signore.[36]

L'analisi di BrusBR 228, anche concentrata sulle sue prime due pagine, spero abbia mostrato il ruolo cruciale di questo *chansonnier* nella biblioteca di Margherita. Non era un libro di musica qualsiasi, di quelli che si trovavano nelle collezioni di altre donne del suo rango. BrusBR 228 era piuttosto uno di quei libri commissionati da Margherita col fine di costruire e promuovere il suo discorso di donna colpita dalla fortuna, e affermare così la sua dignità per un ruolo di potere. Con BrusBR 228 la Margherita esprimeva non solo il suo interesse verso la musica, costruiva piuttosto la sua identità attraverso le possibilità che la musica e i suoi *media* le offrivano, sfruttando a questo scopo l'interazione di tutti gli elementi materiali, visivi, verbali e, ovviamente, musicali

[35] Sulla tradizione teorica medievale di considerare l'infanzia come una fase prerazionale e fino insensata nello sviluppo dell'essere umano si veda P. Riché, D. Alexandre-Bidon, *L'enfance au Moyen Âge*, Paris, Seuil-Bibliothèque Nationale de France, 1994, 12-23.

[36] Si veda a riguardo V. Borghetti, *The Listening Gaze*, 62.

di cui il suo libro di musica si compone. Spero poi che questa analisi contribuisca a mostrare quante e quali fossero le difficoltà per una donna, anche di altissimo rango come Margherita, nel crearsi uno spazio discorsivo autonomo in un mondo patriarcale. BrusBR 228 costituisce un esempio del suo successo, ma anche dei limiti che Margherita, come donna, non poté superare. Il suo libro restava per l'appunto uno *chansonnier*, un libro pertinente alla sfera privata della musica di corte, non a quella pubblica della cappella, che, come istituzione, faceva invece capo agli uomini della sua famiglia.[37] Le grandi dimensioni di questo *chansonnier*, le "grandi" composizioni in posizione esposta come *Ave sanctissima Maria* che esso contiene, rappresentano nel concreto il compromesso necessario per la posizione di preminenza conquistata da Margherita in quanto donna: BrusBR 228 eccede il formato di uno *chansonnier*, ma non eguaglia, anche se per poco, quello di un libro corale di musica sacra; *Ave sanctissima Maria* è un mottetto canonico, che però si nasconde dietro le apparenze di un mottetto-cantilena "da *chansonnier*", cioè un brano e un libro più "da donne" rispetto ad un mottetto e un libro corale da cappella.

In conclusione, parafrasando l'espressione di de Beatis, il libro di musica di Margherita era certo, come la sua biblioteca, "assai ornat[o] e ricch[o]", ma era pur sempre un libro "per donne". Ed era importante per Margherita poter dimostrare di saper stare nei ranghi che, come donna, la società le imponeva, seppure nella condizione privilegiata di sovrana. Con BrusBR 228 Margherita si metteva in scena come una buona vedova cristiana, che, se possedeva un prezioso *chansonnier*, non lo usava certo per il suo personale svago, ma per assolvere alla propria devozione (*Ave sacrissima Maria* in apertura e gli altri brani sacri distribuiti nel libro), e per leggere e (forse) fare musica meditando sui lutti e i dispiaceri della sua esistenza, quelli che le avevano permesso di conquistare una posizione di comando (le composizioni malinconiche e la presenza del marito defunto nello stemma). Tutto questo, però, non era

[37] Nel 1511 la cappella privata di Margherita contava un confessore e quattro cappellani e, come suggerisce Picker, era quindi troppo ridotta per avere funzioni principalmente musicali. Quando Margherita conquistò la reggenza, la sua cappella venne via via accresciuta, ma non raggiunse mai lo *status* di una istituzione musicale autonoma: per i cantori dipendeva da quella (molto più grande) di suo nipote Carlo. Vedi a riguardo M. Picker, *Marguerite of Austria and her Court*, in *The Chansons Albums*, 28-31.

sufficiente. Per essere credibile come vedova e sovrana, Margherita doveva dimostrarsi consapevole del fatto che, come donna, anche le sue attività virtuose erano sempre più esposte di quelle di un uomo ai pericoli del peccato. Le prime pagine del suo *chansonnier* servivano proprio a ribadire in modo chiaro a se stessa e al mondo che la sua preghiera e la sua meditazione articolate in musica nel suo bel libro nulla avevano a che vedere con quelle attività frivole, simboleggiate dalla scimmia e dal putto, tradizionalmente attribuite al suo sesso, e che lei mostrava di saper tener lontane, relegandole nella parte bassa del libro. Era questo il prezzo da pagare per una donna che voleva essere sovrana.

Istituto Lombardo (Rend. Scienze) 151, 219-240 (2017)

DALLE ESIGENZE ENERGETICHE SULLA TERRA ALL'ESPLORAZIONE DI MARTE PASSANDO PER I COLORI DELLA CAPPELLA SISTINA: PRESENTE E FUTURO DEL FOTOVOLTAICO E DEI COMBUSTIBILI SOLARI

ALESSANDRO ABBOTTO (*)

Nota presentata dal m.e. Stefano Maiorana
(Adunanza del 16 novembre 2017)

SUNTO. – La sfida delle energie rinnovabili è giunta oggi ad un passaggio strategico. Mentre è vero che l'80% dell'energia proviene ancora oggi dai combustibili fossili, il ruolo delle rinnovabili è in inarrestabile crescita. Tuttavia, non si tratta solo di una transizione tecnologica ma, da un lato, è la risposta necessaria alle esigenze del pianeta, quali la crescita demografica, le emergenze climatiche, la crescita dei paesi più poveri e, dall'altro, è un presupposto a prospettive di medio-lungo termine come la colonizzazione di Marte. I ricercatori sono quindi impegnati a sviluppare soluzioni innovative, pulite ed efficienti per conseguire tali obiettivi. In particolare i coloranti organici, sostanze organiche come i farmaci o i cristalli liquidi appositamente progettate, rappresentano una delle frontiere più affascinanti della chimica per produrre energia rinnovabile a basso costo e in modo pulito. Come i veri *colori* degli affreschi della Cappella Sistina, voluti da Michelangelo, sono ora emersi e visibili a tutti, anche in questo campo solo alcuni *coloranti* e non altri sono in grado di rispondere con efficacia alla sfida energetica. Lo scienziato, in particolare il chimico, con conoscenza dello stato dell'arte e fantasia, è impegnato nella progettazione di nuove molecole dotate dei *colori* opportuni.

***

ABSTRACT. – The challenge of renewable energies has now come to a strategic step. While

(*) Centro di Ricerca sull'Energia Solare, Dipartimento di Scienza dei Materiali, Università di Milano-Bicocca, Milano, Italy. E-mail: alessandro.abbotto@unimib.it

80% of the energy still comes from fossil fuels, the role of renewable energies is in constant growth. The challenge of renewables is not only a technological transition, but also the indispensable response to the needs of the planet, from the incessant demographic growth to climatic emergencies and to the development of the poorest countries, and eventually to medium to long term challenges such as the colonization of Mars. Researchers are therefore committed to developing innovative, clean and efficient solutions to meet such challenge. In particular, organic dyes, organic molecules such as those founds in drugs or liquid crystals, represent key actors of one of the most fascinating chemistry frontiers to produce low cost and clean renewable energy. As the true *colors* of the Sistine Chapel frescoes, wanted by Michelangelo, have now emerged and visible to everyone, also in this field only specific *colored* dyes are able to address effectively the energy challenge. The scientist, especially the chemist, starting from the state of the art and innovation, is engaged in the design of new molecules owning the correct *colors*.

1. Introduzione

Il dilemma energetico ha raggiunto oggi una fase critica. Importanti emergenze planetarie richiedono soluzioni energetiche efficaci ed immediate. L'esplosione della popolazione nell'ultimo secolo, l'emergenza climatica e l'inquinamento ambientale non consentono più di mantenere le strategie energetiche finora utilizzate. Nel 2014 circa 1.2 miliardi di persone (poco meno del 20% della popolazione mondiale) non hanno ancora accesso all'elettricità.[1] In particolare, in molte regioni centrali dell'Africa, nella zona sub-sahariana e nella regione oceanica, l'accesso alla corrente elettrica copre solo il 10-40% della popolazione, rendendo impossibili importanti operazioni quotidiane come la conservazione di farmaci sensibili al calore, vaccini e antibiotici o il semplice utilizzo di cibi deperibili. Le politiche energetiche del nuovo secolo sono quindi urgentemente chiamate a considerare e risolvere queste criticità. Una politica energetica sostenibile deve innanzitutto affrontare la povertà nel mondo, rendendo accessibile a tutta la popolazione mondiale energia elettrica, cibo e acqua. Dall'altra parte l'emergenza energetica si origina anche dall'inevitabile diminuzione, finanche esaurimento, delle fonti fossili di energia, intrinsecamente non rinnovabili se non nella scala dei milioni di anni. Secondo l'associazione scientifica ASPO (Association for the Study of Peak Oil) la produzione petrolifera mondiale ha già raggiunto o raggiungerà il suo picco nell'attuale decennio. Da questo momento quindi le disponibilità energetiche tradizionali continueranno a diminuire, rendendo impossibile il mantenimento delle attuali tecnologie. L'effetto della crescente scarsezza delle

fonti fossili non solo avrà inevitabili ripercussioni economico-finanziarie (già oggi le nuove tecnologie di reperimento delle fonti fossili, quali il *shale oil/gas*, le perforazioni artiche o oceaniche, le sabbie bituminose, consentono di accedere alle fonti fossili solo a costi molto superiori a quelli tradizionali) ma impedirà, tra l'altro, di accedere alla materia prima di importanti intermedi alla base dell'industria chimica.[2]

La sfida delle energie rinnovabili è giunta oggi ad un passaggio strategico. Mentre è vero che l'80% dell'energia proviene ancora oggi dai combustibili fossili, il ruolo delle rinnovabili è in inarrestabile crescita. Già in questi anni, sia in Italia che in altri paesi, nelle giornate soleggiate estive il 100% dell'energia elettrica è prodotta esclusivamente da fonti rinnovabili, un fatto noto a pochi.[3] Come detto, la sfida delle rinnovabili non è solo una transizione tecnologica, ma la risposta necessaria alle esigenze del pianeta, dalla crescita demografica alle emergenze climatiche e alla crescita dei paesi più poveri, fino ad affrontare sfide di medio-lungo termine come la colonizzazione di Marte. I ricercatori sono quindi impegnati per sviluppare soluzioni innovative, pulite ed efficienti per rispondere alla sfida.

La più importante fonte di energia oggi disponibile alla specie umana è il Sole. Il Sole invia oggi sulla terra una quantità di energia almeno 5000 volte superiore a quella che viene utilizzata. In termini di potenza energetica il consumo di energia mondiale è di circa 18 TW ed è previsto aumentare fino a oltre 30 TW entro il 2015. Per contro l'energia che ci proviene dal Sole risulta poco meno di 100,000 TW. L'energia solare, insieme all'energia geotermica e a quella gravitazionale (maree) costituiscono le tipologie di energia rinnovabili. In particolare possiamo classificare, in base alla tipologia di energia prodotta, l'energia solare in tre grandi categorie: 1) energia termica (riscaldamento, combustibili); 2) energia elettrica (fotovoltaico); 3) reagenti e combustibili per produzione e trasporto (fotosintesi naturale e artificiale) (*Fig. 1*). [4]

Nel campo dell'energia solare il ruolo delle molecole chimiche, in particolare delle molecole organiche, è in continua crescita. In particolare, la classe molecolare dei coloranti organici, sostanze organiche come i farmaci o i cristalli liquidi appositamente progettate, rappresenta una delle frontiere più affascinanti della chimica per produrre energia rinnovabile a basso costo e in modo pulito. Ad ogni molecola potenzialmente utile per catturare l'energia solare è associata una precisa struttura chimica, appositamente ingegnerizzata. Solo certe componenti molecolari e non altre, potranno conferire alla molecola le giuste proprietà per fungere

da componenti efficienti e stabili in dispositivi solari. In altri termini, poiché ad ogni molecola in grado di interagire efficacemente con la radiazione solare, è associata una precisa banda di assorbimento della luce solare, in particolare la porzione visibile, possiamo dire che un preciso colore, e non un altro, conferirà alla molecola le caratteristiche desiderate. Dopo secoli in cui i falsi colori della Cappella Sistina, a causa del fumo delle candele e dell'inquinamento ambientale, sono quelli che tutto il mondo ha visto, oggi un accurato procedimento di restauro e impianto di illuminazione (che ha riprodotto la naturale illuminazione ai tempi michelangioleschi), ha restituito i veri colori degli affreschi, quelli che Michelangelo ha progettato e dipinto. Come i veri colori degli affreschi della Cappella Sistina sono ora emersi e visibili a tutti, anche nel campo dell'energia solare solo alcuni "colori" e non altri sono in grado di rispondere con efficacia alla sfida energetica. Lo scienziato, in particolare il chimico, con la conoscenza dello stato dell'arte e molta fantasia, è impegnato nella progettazione di nuove molecole dotate dei "colori" opportuni.



Fig. 1.

L'utilizzo delle proprietà molecolari porta a ricadute non solo in ambito strettamente accademico ma consente di produrre dispositivi a costo minore e in modo più facile, di disporre di un'ampia varietà di caratteristiche estetiche ed architettoniche (si pensi ad esempio alla

varietà dei colori disponibili per un pannello fotovoltaico), di rendere possibile l'integrazione immediata con dispostivi, edifici, automobili e altri oggetti tecnologici, nonché materiali di qualsiasi genere, dalla plastica ai metalli e finanche tessuti e vestiti, ed infine di permettere di ispirarsi alla Natura (si pensi ad esempio, per quanto riguarda l'energia solare, alla fotosintesi clorofilliana).

Ingegnerizzare una molecola per dispositivi solari significa pertanto progettare accuratamente tutte le componenti chimiche della struttura, non solo per conferire le giuste caratteristiche intrinseche, ad esempio la capacità di assorbire efficacemente la luce solare e di convertirla, ma anche per dialogare e interfacciarsi opportunamente con materiali di altro tipo (ad esempio nanoparticelle di ossidi o contatti metallici) e altre componenti presenti nel dispositivo (*Fig. 2*).

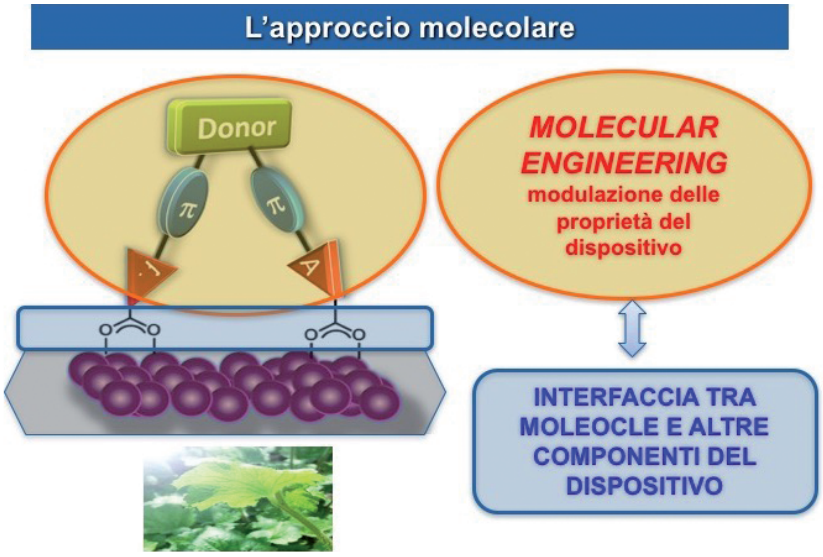


Fig. 2.

2. Molecole e fotovoltaico di ultima generazione

La *Fig. 3*, presa dal National Renewable Energy Laboratory statunitense,[5] riassume graficamente l'attuale stato della tecnologia fotovoltaica. Se ci si limita alle cosiddette celle singole allora è possibile classifi-

care la tecnologia fotovoltaica in tre generazioni: 1) prima generazione, a base di silicio; 2) seconda generazione, a base di film sottili; 3) terza generazione, a base di molecole organiche o ibride. Questa sezione tratterà della terza generazione, con particolare riferimento all'utilizzo dei coloranti organici. Per quanto riguarda questa tipologia troviamo essenzialmente tre tipi di dispositivi: 1) dispostivi a colorante molecolare (*dye-sensitized solar cells*, DSSC); 2) dispositivi completamente organici (*organic photovoltaics*, OPV); 3) celle solari a base di perovskiti ibride organiche-inorganiche (*perovskite solar cells*, PSC). Di queste le più promettenti sono la prima e la terza, con efficienze di conversione energetica, ovvero il rapporto tra l'elettricità prodotta (elettroni) e la luce assorbita (fotoni), ormai superiore al 20% e vicina al record delle celle a silicio del 26%. Si noti che questi valori si riferiscono alla scala di laboratorio, mentre per la scala industriale le efficienze sono leggermente inferiori.

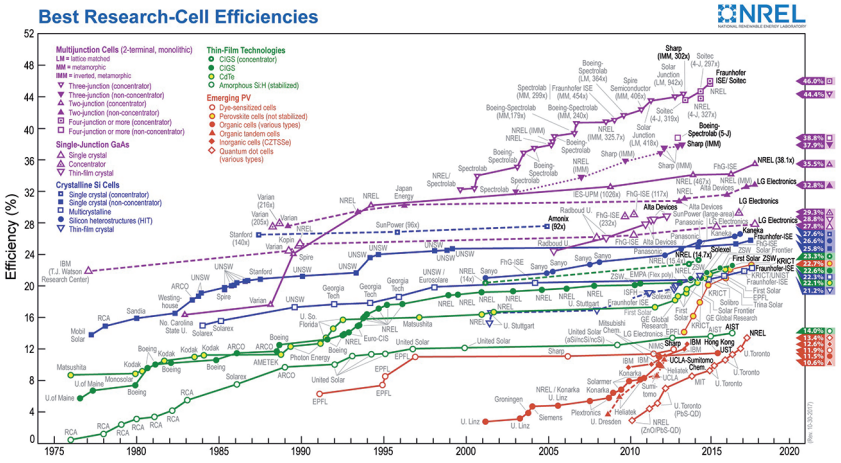


Fig. 3.

A fronte di efficienze leggermente inferiori, le celle di terza generazioni hanno tuttavia caratteristiche peculiari che le rendono uniche e alternative alle celle a silicio. Ad esempio possono dar luogo a dispositivi trasparenti e coloranti, per integrazioni architettonica (*building-integration*) al posto delle vetrate convenzionali, flessibili, leggeri e facilmente integrabili in altri dispostivi elettronici. Tra le caratteristiche più interessanti troviamo la possibilità di lavorare in luce diffusa (alba, crepuscolo, interni di edifici), dove il silicio non funziona, e di poter

essere prodotti a bassissimo costo. Si pensi infatti le nuove celle molecolari utilizzano circa 300 volte meno materiale di una cella a silicio, materiale inoltre che non deve avere gli eccezionali requisiti di purezza (99.99999%) del silicio solare. Di fatto l'industria chimica che produce cristalli liquidi per display e televisori e farmaci è già in grado di produrre i coloranti per fotovoltaico in modo economico ed efficiente.

Pur essendo stata introdotta solo da pochi anni questa tecnologia sta già arrivando ai primi esempi di installazioni esterne. Alcuni esempi sono riportati in *Fig. 4* tra questi evidenziamo le due installazioni a Milano, tra le prime al mondo, in occasione dell'EXPO 2015 (padiglione austriaco) e nella nuova piazza Gae Aulenti.



Fig. 4.

Negli ultimi anni il centro MIB-SOLAR si è occupato soprattutto di celle DSSC, in particolare dello sviluppo di nuovi coloranti organici che fungono da fotosensibilizzatori, ovvero le molecole deputate a raccogliere la luce solare e convertirla in un flusso di elettroni. I fotosensibilizzatori sono pertanto le componenti più importanti di una cella DSSC, la cui struttura è graficamente riportata in *Fig. 5*. Più che entrare nel dettaglio del funzionamento di una cella DSSC è qui importante descrivere la struttura molecolare del fotosensibilizzatore, schematizzata in *Fig. 6*. Il fotosensibilizzatore in *Fig. 6* è una molecola organica comprendente sistemi aromatici (derivati del benzene) e doppi e tripli legami, che conferiscono la peculiare proprietà di essere colorate e

assorbire la porzione visibile della luce solare in modo efficiente. Tuttavia questo non è sufficiente. Per poter trasformare efficacemente fotoni in elettroni la molecola deve essere opportunamente ingegnerizzata comprendendo al suo interno sotto-unità molecolari dotate di precise proprietà elettroniche (caratteristiche di ricchezza o deficienza elettronica, doppi e tripli legami adiacenti). Il chimico progetta e costruisce molecole di questo tipo proprio sviluppando nuovi frammenti molecolari elettron-ricchi o elettron-poveri e combinandoli in una struttura molecolare come quella riportata in *Fig. 6*. Ogni combinazione di frammenti molecolari porta a molecole diverse dotate, ciascuna, di un particolare "colore". E' questa caratteristica peculiare che farà si che alcune molecole funzioneranno meglio di altre.

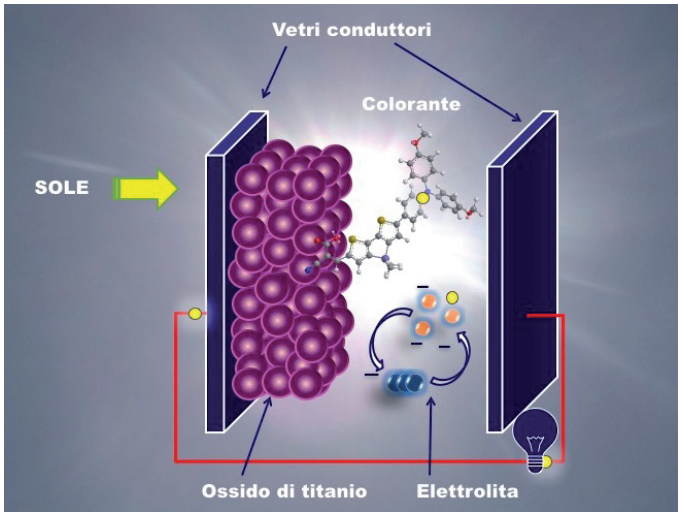


Fig. 5.

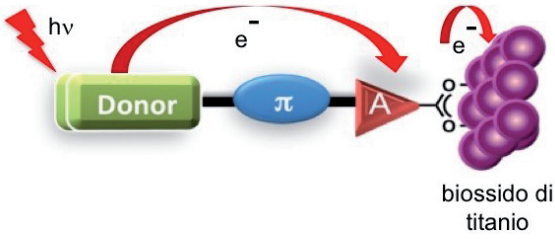


Fig. 6.

Recentemente il nostro gruppo di ricerca ha introdotto un nuovo tipo di architettura molecolare che oggi è ampiamente usato dalla comunità scientifica (*Fig. 7*).[6] Questo esempio ben rappresenta come una opportuna, ma allo stesso tempo semplice, progettazione molecolare possa portare benefici in termini di proprietà molecolari chiave e prestazioni finali del dispositivo. Nella parte in alto è rappresentata l'architettura convenzionale, riportata in letteratura. In questa architettura i frammenti molecolari sono disposti approssimativamente linearmente. Nella nuova architettura, da noi proposta e chiamata multi-ramificata, riportata nella parte inferiore della *Fig. 7*, il percorso degli elettroni lineare è sostituito da 2 o più percorsi molecolari, per favorire la trasformazione da fotoni ad elettroni e il trasferimento degli elettroni agli elettrodi e quindi al circuito elettrico esterno. Inoltre la struttura ramificata permette di migliorare le proprietà di assorbimento della luce e la stabilità dei legami (due al posto di uno) verso le nanoparticelle di biossido di titanio, rappresentate in viola, a loro volta presenti in forma di film sull'elettrodo. Nel complesso pertanto la nuova architettura ha permesso di ottenere sistemi molecolari e dispositivi con migliori proprietà di assorbimento della luce, trasporto di elettroni e stabilità. Alcuni composti di questa serie sono stati oggetto di alcuni brevetti internazionali depositati insieme a ENI SpA.[7]

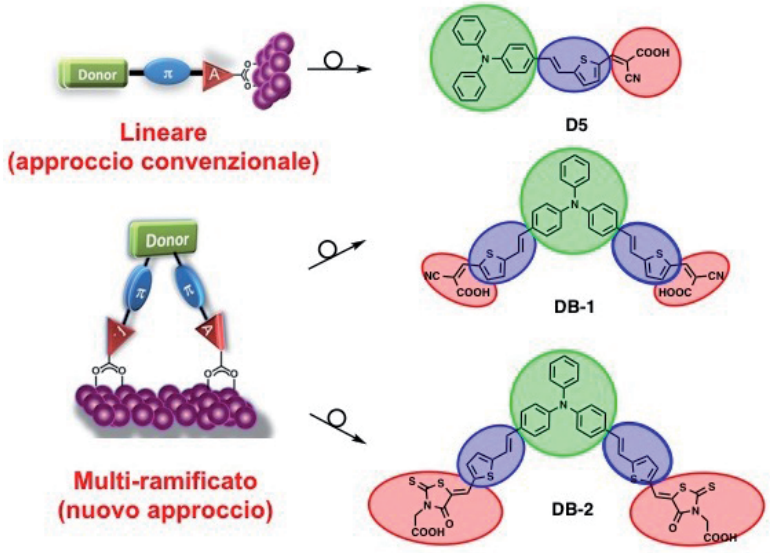


Fig. 7.

In un altro esempio più recente, del 2017, abbiamo invece legato chimicamente una molecola di colorante al grafene, al fine di sfruttare le importanti proprietà di quest'ultimo in ambito fotovoltaico (Fig. *8*).[8] Questo risultato, ottenuto insieme all'Università di Padova, pioniere della chimica del nanocarbonio (fullereni, nanotubi di carbonio e grafene) è il primo esempio al mondo di diade colorante-grafene per fotovoltaico di nuova generazione. Pur non raggiungendo efficienze elevate abbiamo verificato come la diade colorante-grafene sia più efficiente della molecola isolata. In entrambe i casi descritti, si è posto particolare attenzione alla stabilità, oltre che all'efficienza. Infatti solo un sistema molecolare che sia stabile per molti anni potrà consentire la commercializzazione dei dispositivi corrispondenti. La stabilità rappresenta pertanto un requisito essenziale che deve sempre accompagnare le altre proprietà molecolari anche se nella maggior parte dei report in letteratura questo aspetto viene, erroneamente, trascurato.

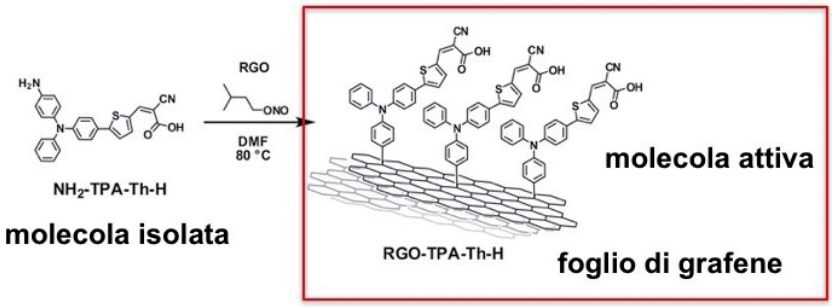


Fig. 8.

Nell'ultimo esempio che vi presentiamo, anch'esso del 2017, abbiamo affrontato un altro aspetto sempre più rilevante nell'industria chimica e tecnologica in generale, ovvero l'utilizzo di metodi di produzione completamente eco-sostenibili. In particolare, abbiamo affrontato il problema legato all'utilizzo di solventi organici volatili e tossici sia nella preparazione sia nella stessa costituzione del dispositivo (nel caso specifico l'acetonitrile). Sarebbe infatti un paradosso proporre nuovi dispositivi solari e sostenibili ma allo stesso tempo utilizzare composti tossici per la loro produzione. Sfruttando una chimica molto recente, abbiamo sostituito il solvente organico tossico con una miscela chiamata *Deep Eutectic Solvent* a base di colina (usato come mangime per i

polli) e la glicerina (o glicerolo), ampiamente usata in preparazione medicinali, cosmetiche e come additivo alimentare (*E422*) e facilmente ottenibile per saponificazione di grassi e olii (*Fig. 9*).[9] Entrambi i composti non solo non sono tossici ma sono ampiamente disponibili a costi molti bassi. Nella copertina che la rivista ci ha attribuito abbiamo intitolato "*drinkable solvents*", per sottolineare appunto l'aspetto di eco-sostenibilità. Le efficienze energetiche sono risultate minori che nel caso dell'utilizzo del solvente tradizionale ma, riteniamo, ampiamente compensate dall'elevata sostenibilità del prodotto, un requisito chiave per la sua ampia commercializzazione.

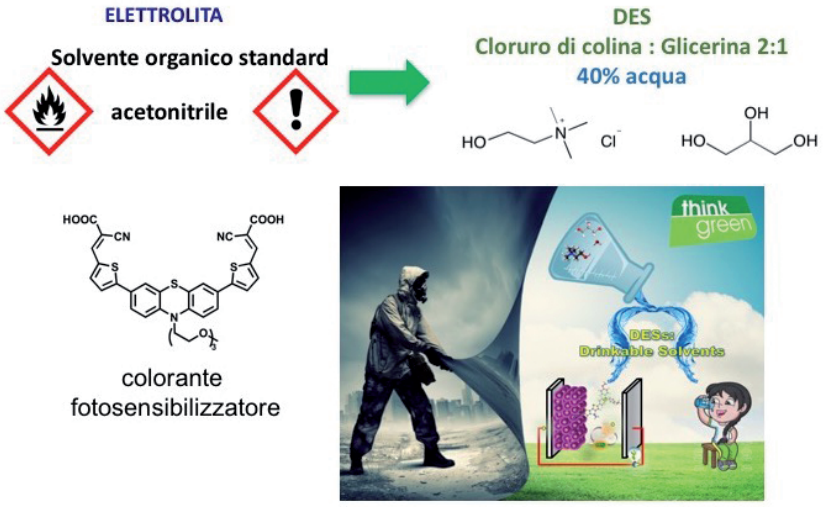


Fig. 9.

3. Molecole e combustibili solari di ultima generazione

Nel capitolo 2 abbiamo affrontato la produzione solare di elettricità. Tuttavia, seppur auspicabile,[4] la completa conversione verso l'elettricità di tutte le necessità energetiche è al momento non ottenibile allo stato delle conoscenze attuali. Di fatto oggi il 60% dell'energia prodotta non è legato all'elettricità ma ai combustibili, in particolare per trasporto (auto, navi ed aerei) e per la sintesi di intermedi chimici. Per quanto riguarda l'autotrasporto in questi mesi stiamo assistendo ad una rivolu-

zione. Dal 2016 in Norvegia, pur essendo il principale produttore europeo di petrolio, il 30% del parco automobilistico è elettrico o ibrido. Purtroppo l'Italia non rientra nella classifica dei primi dieci paesi col maggior tasso di trasporto elettrico. Importanti aziende automobilistiche (Volvo, Mercedes, Volkswagen) hanno annunciato l'abbandono completo di automobili a trazione fossile (benzina, diesel) oppure il completamento di tutta la gamma, compresi i brand affiliati, con modelli di autovetture elettriche. Ormai la transizione è segnata. Tuttavia la trazione elettrica non è la soluzione di tutti i problemi. A parte gli attuali alti costi delle vetture elettriche, l'autonomia è limitata a meno, in alcuni casi molto meno, di 500 km e i tempi di ricarica possono essere di alcune ore (alcune decine di minuti nei casi più tecnologicamente avanzati). Inoltre, pur esistendo prototipi di navi fotovoltaiche che hanno circumnavigato il globo (*MS Tûranor PlanetSolar*) e di aerei fotovoltaici che hanno coperto distanze ragguardevoli (*Solar Impulse 2*, percorso Giappone-Hawaii, circa 8000 km), non è pensabile né a breve né a medio-lungo termine, di disporre di navi e aerei commerciali a trazione elettrica (*Fig. 10*). I combustibili per trasporto rappresentano pertanto l'unica soluzione tecnologica per le prossime decine di anni.



Fig. 10.

Se non consideriamo le biomasse, l'unico combustile per trasporto sostenibile è l'idrogeno.[1] L'idrogeno è il combustibile ideale perché ha *carbon footoprint* nulla e per combustione produce solo vapore acqueo ed elettricità, quest'ultima nei motori a celle a combustibile. I primi esempi di utilizzo di serie dell'idrogeno per trasporto sono apparsi sul mercato proprio quest'anno (*Fig. 11*).



Fig. 11.

La Toyota ha commercializzato la prima autovettura di serie a idrogeno (Toyota Mirai) e nel 2018 in Germania prenderà servizio il primo treno al mondo a trazione a idrogeno. In entrambi i casi le prestazioni e l'autonomia sono comparabili, se non superiori, a quelle dei modelli a trazione convenzionale. Nel caso del modello automobilistico il completo rifornimento viene effettuato in circa 3 minuti, ovvero in tempi analoghi a quelli convenzionali. L'idrogeno è anche il combustibile di scelta, se non l'unico possibile, per la colonizzazione di Marte, dove il trasferimento dalla Terra di combustibili fossili è impensabile e lo sfruttamento dell'energia fotovoltaica è reso difficile dalla maggiore distanza dal Sole e dalle continue tempeste di sabbia. Esiste anche un ambizioso piano nazionale per la mobilità a idrogeno che prevede 27,000 veicoli a idrogeno per il 2025, 8 milioni (20% del totale) di veicoli per il 2050, 23,000 autobus e 20 stazioni di rifornimento per il 2020, che diventeranno 5,000 entro il 2050. Due anni fa è stata inaugurata nell'autostrada del Brennero

a Bolzano Sud la prima stazione di rifornimento di idrogeno (*Fig. 12*). Non mancano, infine, applicazioni per trasporti di dimensioni maggiori, come il trasporto su mare. Lo scorso luglio è partito da Parigi il giro del mondo di "*Energy Observer*", la prima nave ad idrogeno, che prende il combustibile direttamente dall'acqua su cui naviga dopo dissalazione e trasformazione in idrogeno con un processo di elettrolisi.



Fig. 12.

Tuttavia l'idrogeno non è una fonte di energia, non essendo presente in modo naturale sulla Terra, bensì un vettore energetico. Deve essere quindi ottenuto da altre fonti. Attualmente il metodo principale di produzione dell'idrogeno è lo *steam reforming* da metano, un combustibile fossile. Quindi l'uso dell'idrogeno è pulito ma non lo è la sua produzione. Il vantaggio netto in termini di eco-sostenibilità è pertanto nullo. Per accedere ad un'economia dell'idrogeno veramente sostenibile è necessario quindi produrre idrogeno in altro modo. Negli ultimi anni la ricerca sta concentrando gli sforzi nella produzione di idrogeno a partire da acqua e luce solare, in un processo, chiamato fotosintesi artificiale, che mima la fotosintesi naturale (*Fig. 13*). Nella fotosintesi artificiale l'acqua viene divisa nei suoi componenti, idrogeno e ossigeno. Purtroppo (o per fortuna, altrimenti avremmo nella Terra atmosfere di idrogeno al posto di oceani di acqua) questa reazione è energeticamente e cineticamente sfavorita. Per osservarla in rese soddisfacenti è necessario utilizzare opportuni catalizzatori e altre componenti moleco-

lari in quella che rappresenta senza dubbio una delle più affascinanti sfide della chimica moderna.

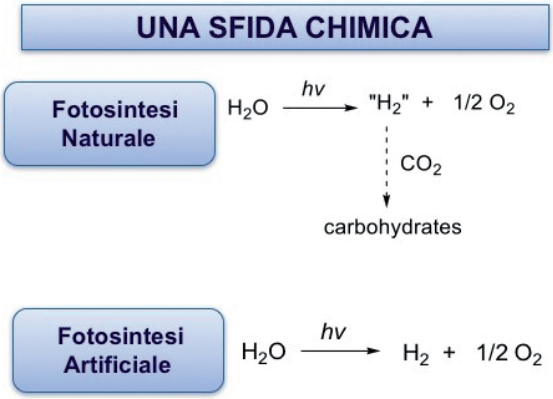


Fig. 13.

Tra i primi catalizzatori studiati, fin dai primi lavori pionieristici degli anni '70,[10] troviamo il biossido di titanio, un materiale economico e largamente utilizzato in molti ambiti commerciali, dalle vernici alle creme solari. Tuttavia il biossido di titanio assorbe solo la porzione ultravioletto dello spettro solare, rimanendo inerte nei confronti della parte dello spettro più importante, quella visibile o colorata (*Fig. 14*).

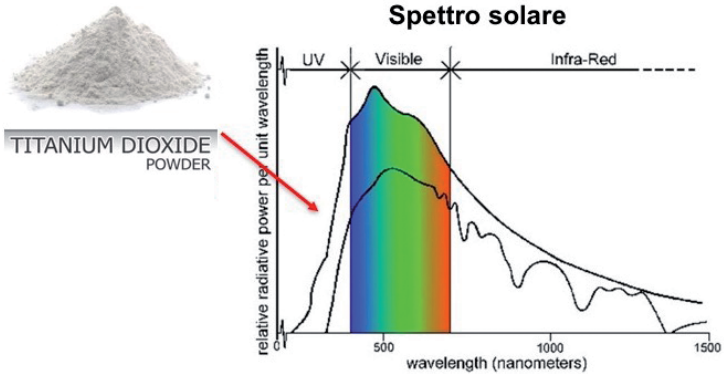


Fig. 14.

E' qui che entrano nuovamente in gioco i coloranti i quali, fungendo da antenne molecolari, catturano efficacemente l'energia solare

trasferendola ai centri reazionali dove avviene la scissione dell'acqua in idrogeno e ossigeno (*Fig. 15*).[11] Solo negli ultimi anni sono comparsi in letteratura i primi esempi di coloranti organici per la produzione di idrogeno via fotosintesi artificiale, ricerca che abbiamo recentemente raccolto in una *review* riepilogativa.[12]

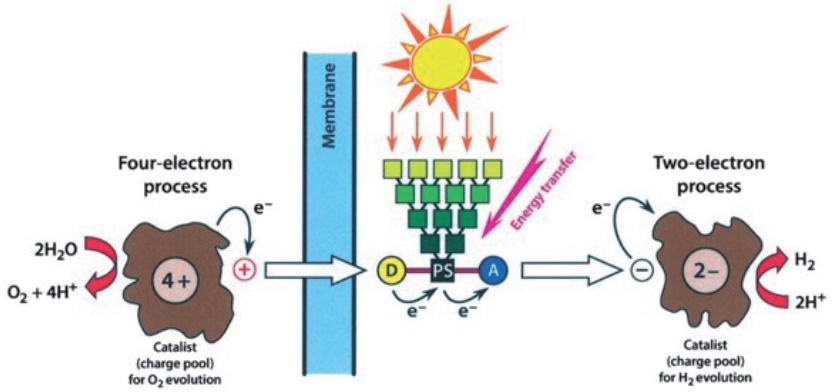


Fig. 15.

I requisiti progettuali per i fotosensibilizzatori organici per la produzione di idrogeno sono simili ma non identiche a quelle che abbiamo visto per i coloranti per le applicazioni fotovoltaiche. Con quest'ultimi condividono la necessità di assorbire efficacemente la radiazione solare ma, a differenza di questi, devono lavorare in acqua, un ambiente in genere poco compatibile con le molecole organiche aromatiche, a carattere idrofobo. Inoltre devono possedere legami chimici stabili in tale solvente (ovvero non *idrolizzabili*) e possibilmente mimare la Natura che ha perfezionato, in milioni di anni, il corrispondente processo naturale della fotosintesi a partire da acqua e sole. Per ottenere questi obiettivi il chimico ingegnerizza nuovamente la struttura molecolare, sia scegliendo opportuni *scaffold* o templati molecolari, sia sfruttando le funzionalizzazioni periferiche per indurre le proprietà desiderate (*Fig. 16*).

Le *Figg. 17* e *18* riportano alcuni esempi recenti di fotosensibilizzatori sviluppati nei nostri laboratori al MIB-SOLAR.[13] Nel primo caso viene studiato l'effetto della variazione del core elettron-ricco, mentre rimane invariata la rimanente parte della molecola. È interessante osservare che le tre molecole si comportano in maniera differente,

soprattutto in termini di efficienza energetica e produzione di idrogeno, come conseguenza della, seppur fine, differenza strutturale. Nel secondo esempio vengono invece studiati diversi spaziatori molecolari in corrispondenza dei due rami della molecola multi-ramificata. Anche in questo caso solo certi spaziatori, e non altri, funzionano bene. Nuovamente, solo le molecole col giusto *colore* presentano le proprietà desiderate. Il risultato è una produzione stabile e continua di idrogeno da acqua e sole, una caratteristica essenziale, come già sottolineato, per lo sviluppo industriale di questi materiali (*Fig. 19*).

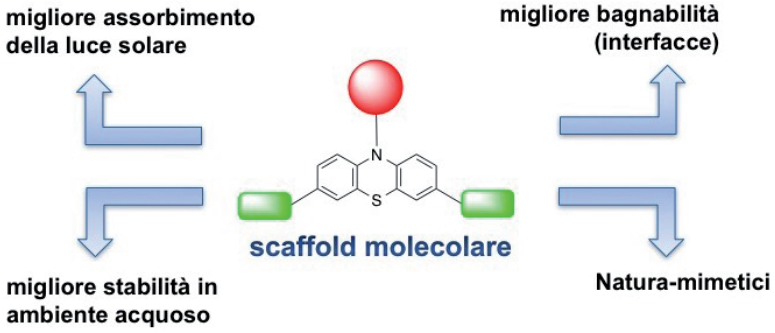


Fig. 16.

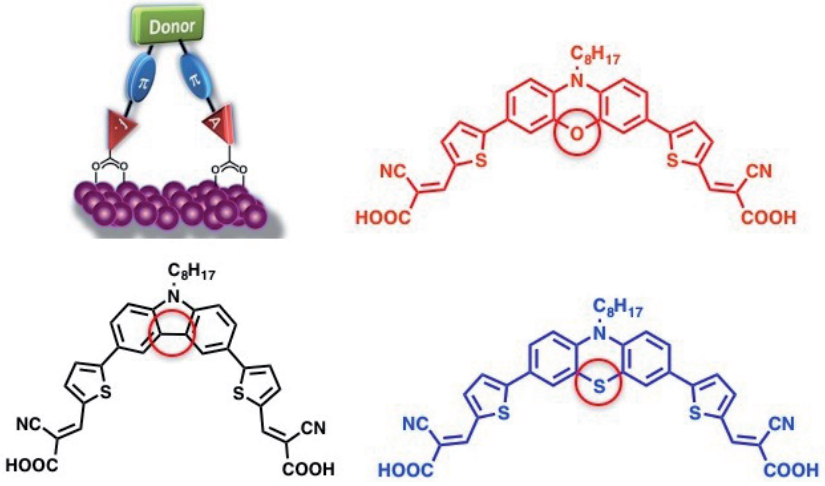


Fig. 17.

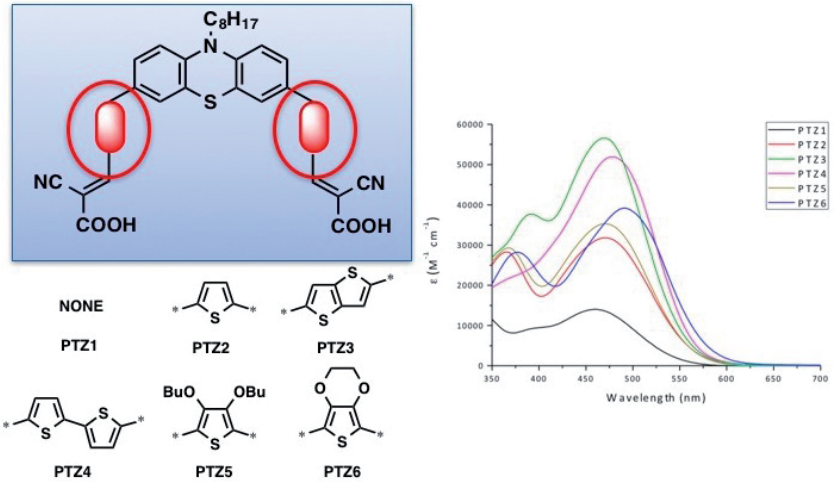


Fig. 18.

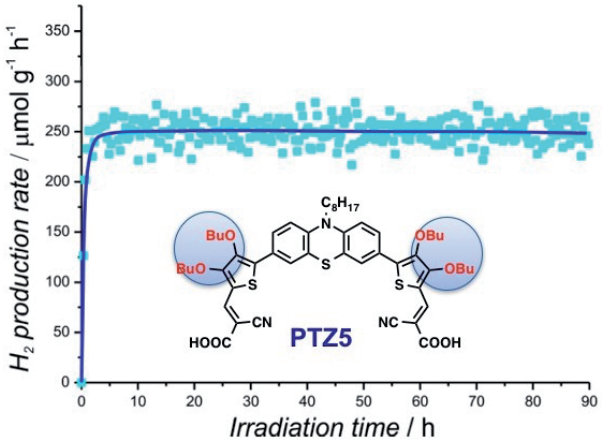


Fig. 19.

Per migliorare l'affinità della molecola all'ambiente acquoso è possibile introdurre in catena laterale delle funzionalità polari idrofile, tali da rendere tutta la molecola idrofila. Anche in questo caso è possibile sfruttare la varietà strutturale a disposizione del chimico, per migliore l'efficienza finale. In un nostro recente lavoro del 2016 (*Fig. 20*) abbiamo confrontato la capacità di produrre idrogeno utilizzando due coloranti molecolari, entrambi identici tranne che nella porzione periferica idrofi-

la.[14] Nel primo caso abbiamo utilizzato una catena polare convenzionale (in nero nella *Fig. 20*), mentre nel secondo caso abbiamo introdotto nella struttura molecolare un pendaglio zuccherino, ed esattamente il frammento del glucosio, che insieme al fruttosio è alla base del saccarosio, o zucchero da cucina. L'idea dietro allo studio è tanto semplice quanto efficace: sfruttare le ben note proprietà idrofile dello zucchero! I risultati confermano quanto sperato: il derivato zuccherino ha una affinità all'acqua ben superiore a quella della semplice catena polare, affinità che ha permesso una produzione di idrogeno praticamente raddoppiata. Gli interessanti risultati hanno valso all'articolo la copertina di una delle riviste chimiche più importanti, forse la più importante, della *Royal Society of Chemistry* (*Fig. 20*).

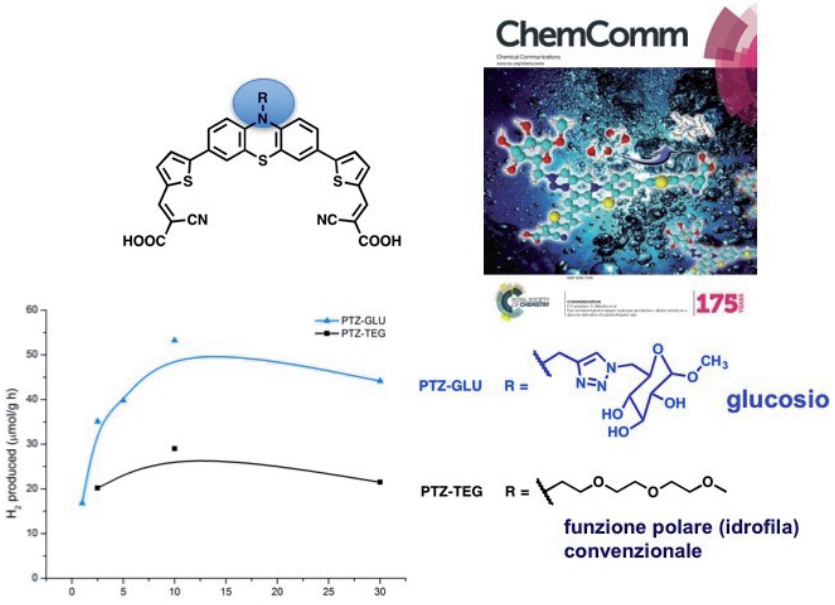


Fig. 20.

Nell'ultimo esempio, descritto in *Fig. 21*, il target è ancora più ambizioso. Sempre sfruttando il glucosio, ora lo scopo è quello di indurre le molecole a posizionarsi in modo ordinato sulla superficie dell'elettrodo. E' infatti noto che se le molecole sono vicine le une alle altre le prestazioni diminuiscono per un fenomeno di reciproco spegnimento dell'attività (*quenching intermolecolare*), se le molecole sono ordinate

parallelamente ma opportunamente spaziate allora le prestazioni sono massimizzate. Non è facile indurre le singole molecole a disporsi nello spazio come noi desideriamo. Nel nostro caso abbiamo sfruttato le deboli interazioni *intermolecolari* che si stabiliscono spontaneamente tra le molecole dotate del residuo periferico zuccherino e il glucosio stesso (qui aggiunto come acido glucuronico, un suo immediato derivato) (in basso a destra nella *Fig. 21*).[15] Dopo aver accertato l'ipotesi con sofisticate simulazioni computazionali (in basso a sinistra) abbiamo misurato la produzione di idrogeno senza (colonne rosse) e con (colonne blu) l'aggiunta di acido glucuronico. L'effetto è evidente. Nel secondo caso la produzione di idrogeno è raddoppiata, grazie alla costruzione dell'edificio *supermolecolare*.

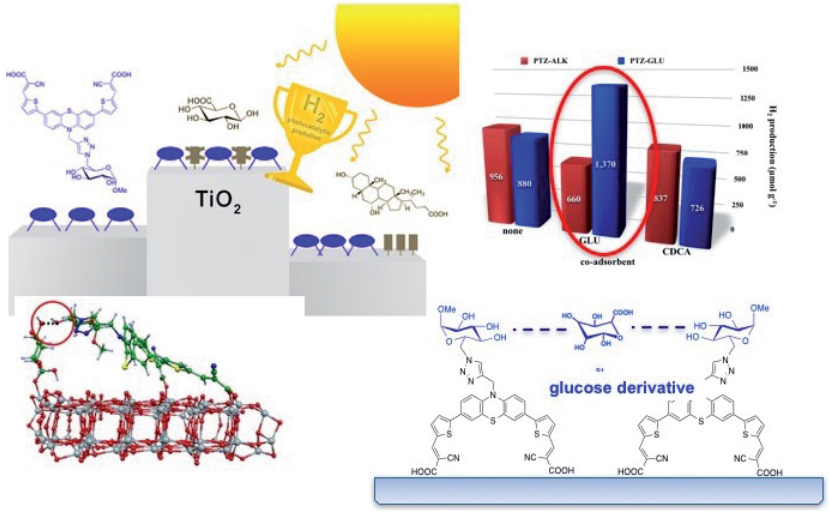


Fig. 21.

4. Conclusioni

La chimica molecolare, in particolare la chimica organica, è chiamata a partecipare alle nuove sfide scientifiche e tecnologiche nel campo dell'energia, contribuendo a risolvere le criticità attuali e a definire nuove soluzioni per l'approvvigionamento energetico dei prossimi decenni. Originali molecole innovative vengono oggi progettate per

rendere i dispositivi solari, sia per produzione di elettricità che per produzione di combustibili, più economici, efficienti, stabili e competitivi con le fonti fossili tradizionali. Pur esistendo in letteratura ormai diversi esempi, questa sfida è solo all'inizio. Da una parte, in termini di prove concettuali, le molecole organiche hanno già dimostrato di poter assumere il ruolo di coprotagonisti nella nuova sfida energetica del nostro secolo, dall'altro ancora molta strada deve essere percorsa, in termini soprattutto di efficienza e stabilità, ad oggi ancora non sufficiente per un impiego di massa ma, come descritto nei precedenti capitoli, in continua evoluzione passo dopo passo. La sfida è certamente complessa e richiederà anni per essere completamente risolta ma, al pari di altre tecnologie a base di molecole organiche, quali ad esempio i cristalli liquidi o gli OLED che hanno richiesto decine di anni prima di arrivare nel mercato, anche in questo campo, alla fine, il ruolo dei coloranti organici sarà senza dubbio rilevante.

5. Riconoscimenti

Gli autori desiderano ringraziare l'Università di Milano-Bicocca per il finanziamento che ha permesso la creazione del Centro di Ricerca dell'Energia Solare MIB-SOLAR[16] dove gran parte degli studi descritti sono stati sviluppati (*Fig. 22*), il Ministero dell'Istruzione, dell'Università e della Ricerca, la Regione Lombardia, l'Unione Europea e ENI SpA per il contributo finanziario alle ricerche qui descritte.



Fig. 22.

BIBLIOGRAFIA

1. REN21. 2017. Renewables 2017 Global Status Report (Paris: REN21 Secretariat) (il report è disponibile online sul sito di REN21 (http://www.ren21.net/).
2. Armaroli N., Balzani V., Energia per l'Astronave Terra, Zanichelli, Bologna, 2017.
3. Fonte: Terna SpA (http://www.terna.it/).
4. Balzani V., Armaroli N., Energy for a Sustainable World: From the Oil Age to a Sun-Powered Future, Wiley-VCH, Weinheim, 2010.
5. https://www.nrel.gov/
6. Manfredi N., Cecconi B., Abbotto A., Multi-Branched Multi-Anchoring Metal-Free Dyes for Dye-Sensitized Solar Cells, Eur. J. Org. Chem., 2014: 7069-7086.
7. (a) Abbotto A., Acciarri M., Biagini P., Binetti S., Organic dye for a dye sensitized solar cell. PCT Int. Appl., 2014: WO 2014053626; (b) Biagini P., Abbotto A., Manfredi N., Organic dyes and dye-sensitized solar cells using them, PCT Int. Appl., 2015: WO 2014125432.
8. Gatti T., Manfredi N., Boldrini C., Lamberti F., Abbotto A., Menna E., A D-π-A organic dye - reduced graphene oxide covalent dyad as a new concept photosensitizer for light harvesting applications, Carbon, 2017: 115: 746-753.
9. Boldrini C. L., Manfredi N., Perna F. M., Trifiletti V., Capriati V., Abbotto A., Dye-Sensitized Solar Cells that use an Aqueous Choline Chloride-Based Deep Eutectic Solvent as Effective Electrolyte Solution, Energy Technology, 2017: 5: 345-353.
10. Fujishima A., Honda K., Electrochemical Photolysis of Water at a Semiconductor Electrode, Nature, 1972: 238: 37-38.
11. (a) Armaroli N., Balzani V., Solar Electricity and Solar Fuels: Status and Perspectives in the Context of the Energy Transition, Chem. Eur. J., 2016: 22: 32-57; (b) Armaroli N., Balzani V., The Hydrogen Issue, Chemsuschem, 2011: 4: 21-36.
12. Cecconi B., Manfredi N., Montini T., Fornasiero P., Abbotto A., Dye-Sensitized Solar Hydrogen Production: The Emerging Role of Metal-Free Organic Sensitizers, Eur. J. Org. Chem., 2016: 5194-5215.
13. Cecconi B., Manfredi N., Ruffo R., Montini T., Romero-Ocaña I., Fornasiero P., Abbotto A., Tuning Thiophene-Based Phenothiazines for Stable Photocatalytic Hydrogen Production, ChemSusChem, 2015: 8: 4216-4228.
14. Manfredi N., Cecconi B., Calabrese V., Minotti A., Peri F., Ruffo R., Monai M., Romero-Ocana I., Montini T., Fornasiero P., Abbotto A., Dye-sensitized photocatalytic hydrogen production: distinct activity in a glucose derivative of a phenothiazine dye, Chem. Commun., 2016: 52: 6977-6980.
15. Manfredi N., Monai M., Montini T., Peri F., De Angelis F., Fornasiero P., Abbotto A., Dye-sensitized Photocatalytic Hydrogen Generation: Efficiency Enhancement by Organic Photosensitizer – Co-Adsorbent Intermolecular Interaction, ACS Energy Letters, 2017; DOI:10.1021/acsenergylett.7b00896.
16. www.mibsolar.mater.unimib.it

Istituto Lombardo (Rend. Scienze) 151, 241-253 (2017)

ESPLORANDO "FLATLANDIA": DAL GRAFENE ALLA SCOPERTA DI NUOVI MATERIALI BI-DIMENSIONALI

FRANCESCO TUMINO (*)

Nota presentata dal m.e. Carlo Bottani
(Adunanza del 16 novembre 2017)

SUNTO. – L'avvento delle nanotecnologie ha stimolato la ricerca verso nuovi nanomateriali per applicazioni innovative. Nel 2004, la produzione e lo studio del grafene — ovvero un singolo strato di atomi di carbonio — mise in luce, per la prima volta, le eccezionali proprietà di questo materiale e aprì la strada alla ricerca sui cosiddetti *materiali bi-dimensionali* (2D). Da allora, molti nuovi materiali 2D sono stati prodotti e studiati, evidenziando proprietà e comportamenti nuovi rispetto alle più note versioni "3D" degli stessi materiali. La ricerca sui materiali 2D è oggi uno dei settori più attivi e promettenti della scienza dei materiali e pone le basi per lo sviluppo di nuove tecnologie, ad esempio, nel campo dell'elettronica, dell'optoelettronica, dell'energia e della sensoristica molecolare. In questo articolo, verranno introdotti alcuni importanti aspetti di questa ricerca — come i metodi di sintesi e le tecniche di caratterizzazione dei materiali 2D — e verranno brevemente discusse alcune caratteristiche generali di questi materiali alla luce di recenti studi sperimentali effettuati tramite la tecnica di *microscopia a effetto tunnel* (STM).

***

ABSTRACT. – The development of nanotechnology has encouraged the research of new nanomaterials for innovative applications. In 2004, the production and study of graphene — that is a single layer of carbon atoms — showed, for the first time, the extraordinary properties of this material and opened the way to the exploration of the so-called *two-dimensional* (2D) *materials*. Since then, several 2D materials have been

(*) Dipartimento di Energia, Politecnico di Milano, Italy.
E-mail: francesco.tumino@polimi.it

produced and studied, revealing properties and behaviours, in general, very different from those of corresponding bulk materials. Research on 2D materials is nowadays one of the most active and promising fields of materials science, which is setting the basis for the development of novel technologies, such as in electronics, optoelectronics, energy and molecular sensing. In this paper, some important aspects of the study of 2D materials will be introduced — such as the synthesis methodologies and characterization techniques — and some of their properties will be shown, with the support of recent experimental results of *scanning tunneling microscopy* (STM) investigations.

1. Introduzione

La scoperta di nuovi materiali ha sempre avuto, dalla preistoria fino ai giorni nostri, un impatto enorme sul progresso tecnologico dell'uomo. In tempi recenti, la ricerca di nuovi materiali e il conseguente sviluppo di nuove tecnologie hanno ricevuto forte impulso, soprattutto grazie ai grandi passi avanti compiuti nella chimica e nella fisica della materia nel corso del XX secolo. Tale sviluppo ha permesso la nascita delle *nanotecnologie*, che si basano essenzialmente sulla manipolazione della materia su scala dimensionale dell'ordine del nanometro. Uno degli aspetti più innovativi di questo approccio risiede nella possibilità di accedere alle diverse proprietà che i materiali assumono quando le loro dimensioni sono ridotte fino alla scala nanometrica. La dimensione, infatti, gioca un ruolo fondamentale nel determinare la struttura elettronica di un materiale, dalla quale dipendono le sue proprietà ottiche, elettriche, meccaniche, termiche, magnetiche, chimiche. I *materiali bidimensionali* (2D) forniscono un esempio evidente di come ridurre la dimensione spaziale di un materiale — una sola dimensione in questo caso — possa far emergere proprietà e comportamenti nuovi rispetto a quelli dello stesso materiale "tri-dimensionale". Si consideri, ad esempio, la grafite, che, come molti sanno, è un materiale molto comune formato da atomi di carbonio. Se immaginiamo di ridurre lo spessore di un campione di grafite fino ad isolare un solo strato atomico, otteniamo un nuovo materiale dalle proprietà uniche: il *grafene*. Lo studio del grafene valse il premio Nobel per la fisica nel 2010 a Andre Geim e Konstantin Novoselov [1], e aprì la strada alla produzione e allo studio di nuovi materiali 2D, quali il nitruro di boro esagonale (hBN), il disolfuro di molibdeno (MoS_2), il silicene e altri ancora [2]. Oggi, la ricerca sui materiali 2D e sulle strutture che si possono ottenere combinandoli tra loro è uno dei campi più attivi della scienza dei materiali, potenzialmente in

grado di innovare radicalmente i futuri dispositivi elettronici e optoelettronici [3][4]. Proprio come in *Flatlandia* — l'universo bi-dimensionale immaginato da Edwin Abbott Abbott nel suo racconto del 1884 — l'esplorazione del mondo dei materiali 2D si profila come un'importante occasione di estendere la nostra conoscenza e di concepire nuovi scenari per il futuro sviluppo tecnologico e sociale. Una descrizione esaustiva e dettagliata della scienza dei materiali 2D va, naturalmente, oltre lo scopo di questo articolo, che, piuttosto, intende fornire una breve e generale introduzione a questa classe di materiali (sezione 2), mettendo in evidenza aspetti importanti della ricerca, quali i possibili metodi per la loro fabbricazione (sezione 3) e tecniche per la caratterizzazione delle loro proprietà (sezione 4). A scopo esemplificativo, verranno anche mostrati alcuni recenti risultati sperimentali ottenuti tramite osservazione al *microscopio a effetto tunnel* (STM) di alcuni materiali 2D (sezione 4.1).

2. Materiali ed eterostrutture 2D

Come accennato nell'introduzione, i materiali 2D sono materiali "super-sottili", aventi uno spessore dell'ordine del nanometro. Appartengono, quindi, a questa categoria quei materiali composti da uno — come il grafene — o, al più, pochi strati atomici. La bassa dimensionalità di questi materiali conferisce loro proprietà, in generale, molto diverse da quelle delle versioni "3D" (o *bulk*) dei corrispettivi materiali. A parte il grafene, un altro evidente esempio di come il numero di strati atomici influenza le proprietà di un materiale è fornito da un'importante famiglia di materiali 2D: i *dicalcogenuri di metalli di transizione* (TMD). Il principale, e più studiato, rappresentante di questa famiglia è il *disolfuro di molibdeno* (MoS_2), che si trova comunemente in natura sotto forma del minerale molibdenite. Si tratta di un materiale simile alla grafite, in quanto formato da strati atomici debolmente legati tra loro (da interazioni di tipo *van der Waals*), ciascuno dei quali composti da atomi legati tra loro da legami molto forti, di tipo covalente. A differenza della grafite, il singolo strato di MoS_2 ha, oltre alla composizione chimica, una struttura diversa, in quanto formato da tre piani atomici: un piano di atomi di molibdeno tra due piani di atomi di zolfo. Questo materiale, così come altri TMD (WS_2, $MoSe_2$, WSe_2, ecc.), si distingue per una peculiare modifica delle sue proprietà opto-elettroniche quando si passa da molti strati atomici a un solo strato atomico (cioè *al limite bi-dimen-*

sionale). Tale modifica interessa, in particolare, la sua struttura elettronica e si manifesta nella transizione da materiale semiconduttore a gap *indiretto* (nel caso 3D) a semiconduttore a gap *diretto* (nel caso 2D). Dal momento che il gap energetico è nell'intervallo di luce visibile (vale 1.9 eV), tale transizione implica che l'interazione con la luce — in termini, ad esempio, di assorbimento o fotoluminescenza — sia molto più intensa per il singolo strato di MoS_2 rispetto alla versione multi-strato, caratteristica che fa di questo materiale 2D un promettente candidato come principale componente di futuri dispositivi opto-elettronici [5]. Un altro aspetto che rende interessante lo studio dei materiali 2D è il fatto di avere proprietà molto diverse tra loro, oltre che rispetto ai materiali bulk. Ad esempio, con riferimento alla struttura elettronica, troviamo semiconduttori a zero gap (il grafene), semiconduttori con gap nel visibile (il MoS_2), metalli (il NbSe) e isolanti, come il *nitruro di boro esagonale* (hBN). Quest'ultimo è uno dei materiali 2D tradizionalmente più studiati oltre il grafene, del quale possiede la stessa struttura a nido d'ape, nella quale, però, si alternano atomi di boro e azoto, anziché di carbonio. Ciò è sufficiente a conferire allo hBN proprietà molto diverse da quelle del grafene, come un ampio gap energetico (6 eV). Esistono altri materiali 2D *graphene-like*, cioè con struttura simile a quella del grafene: di recente, ad esempio, sono stati studiati il *silicene*, il *germanene* e lo *stanene* (anche chiamati *X-eni*), ovvero le "versioni 2D" di silicio, germanio e stagno, rispettivamente. A questo punto è opportuno fare un'osservazione di carattere generale. Materiali 2D come grafene, MoS_2 o hBN non sono altro che il singolo strato di una struttura cristallina a molti strati, ed è quindi possibile produrli separando tra loro i vari strati fino ad isolarne uno solo (tramite tecniche di *esfoliazione*, come descritto nella sezione 3). Invece, dal momento che materiali come silicio, germanio o stagno non possiedono una struttura stratificata, i corrispondenti materiali 2D non possono essere ricavati per esfoliazione della struttura bulk, ma vanno *sintetizzati* tramite tecniche in grado di far "crescere" un singolo strato atomico su un substrato adatto a questo scopo. In questo caso, l'*interazione con il substrato* è un parametro che contribuisce significativamente, insieme al confinamento dimensionale, a determinare le proprietà strutturali ed elettroniche del materiale 2D supportato. Questo aspetto è centrale per la sintesi di molti materiali 2D, compresa un'importante classe di questi materiali: quella degli *ossidi 2D*. I materiali ossidi si trovano abbondantemente in natura, sono economici, non tossici, dotati di grande stabilità chimica e termica e, per questo, utiliz-

zati in moltissime applicazioni. Per la loro importanza, questi materiali occupano da decenni una posizione centrale nella scienza e tecnologia dei materiali. Di recente, lo studio di ossidi a bassa dimensionalità ha portato all'osservazione di nuove strutture 2D, che possono trovare applicazione in ambiti tecnologici avanzati [6]. La vasta gamma di proprietà e funzionalità offerte dagli ossidi, estende ulteriormente lo spettro di potenzialità dei materiali 2D. La grande varietà che si osserva all'interno di questa classe di materiali ha fatto emergere, negli ultimi anni, un nuovo paradigma nella ricerca e nel design delle future applicazioni di questi materiali: combinare insieme diversi materiali 2D per creare ad arte dei materiali più complessi — le cosiddette *eterostrutture 2D* — con nuove proprietà e funzionalità. Questi nuovi materiali possono essere realizzati impilando uno sopra all'altro diversi materiali 2D, a formare eterostrutture multi-strato, in cui, idealmente, ogni strato atomico corrisponde a uno specifico materiale 2D. Tali eterostrutture — solitamente chiamate *eterostrutture di van der Waals*, per via della natura del legame tra i vari strati — sono attualmente oggetto di un'intensa attività di ricerca sia di carattere teorico sia sperimentale, volta alla realizzazione di prototipi per lo studio sia delle loro proprietà fondamentali sia delle loro prestazioni in un'ottica più applicativa [7]. Naturalmente, l'obiettivo a lungo termine di realizzare dispositivi efficienti basati su eterostrutture 2D presuppone non solo uno sviluppo della conoscenza di base di questi sistemi, ma anche la ricerca di metodi per la loro fabbricazione che siano affidabili, scalabili e in grado di fornire un sufficiente grado di controllo sulle caratteristiche dei materiali prodotti.

3. Produrre i materiali 2D: metodi di fabbricazione

Storicamente, la prima tecnica messa a punto per ottenere il grafene fu la cosiddetta *esfoliazione micro-meccanica*. Questa tecnica prevede l'utilizzo di comune nastro adesivo per esfoliare ripetutamente un campione di grafite fino ad ottenere dei "pezzi" (*flakes*) sempre più sottili. Questa operazione produce anche dei flakes spessi uno strato atomico, cioè il grafene. Si tratta di una tecnica concettualmente molto semplice ed economica, che ha consentito di effettuare, nel 2004, i primi importanti esperimenti sul grafene e rivelarne le proprietà uniche. Successivamente, è stata applicata ad altri materiali esfoliabili, come MoS_2 e hBN, permettendo di produrre nuovi materiali 2D oltre il gra-

fene [8]. L'esfoliazione micro-meccanica è ancora oggi molto utilizzata nei laboratori di ricerca e, combinata a tecniche di trasferimento preciso dei flakes, può essere impiegata per la produzione di eterostrutture verticali di diversi materiali 2D. Tuttavia, questa tecnica presenta anche importanti limiti, tra cui quelli di non offrire nessun grado di controllo sullo spessore dei flakes ottenuti e di richiedere molto tempo per l'individuazione dei flakes mono-strato. Un passo avanti in questo senso è rappresentato dall'*esfoliazione in fase liquida*, ovvero l'esfoliazione del materiale bulk tramite specifici agenti chimici in soluzione. Tuttavia, è opportuno sottolineare che le tecniche di esfoliazione non consentono di produrre materiali 2D di elevata area superficiale e che non sono applicabili a materiali non esfoliabili, non permettendo, quindi, la produzione di molti materiali 2D la cui controparte bulk non sia formata da strati debolmente legati tra loro. Per produrre questi materiali — ad esempio il silicene — è necessario adottare un approccio differente, di tipo *bottom-up*, che consiste nel "crescere" il materiale a partire dai suoi costituenti elementari, atomi o molecole. Per fare ciò si possono utilizzare diverse *tecniche di deposizione*, che consentono di sintetizzare il materiale su substrati che ne favoriscano la crescita. Una delle tecniche di deposizione più investigate per la crescita di materiali 2D è la *deposizione chimica da vapore* (CVD), che permette di sintetizzare film di elevata qualità cristallina e ampia estensione (anche su scala del centimetro). In un tipico processo di CVD, uno o più precursori molecolari volatili vengono fatti reagire chimicamente in modo da produrre il film sul substrato. Quest'ultimo, al fine di promuovere le reazioni chimiche in gioco, viene tipicamente portato a temperature relativamente elevate (intorno a 1000 °C, indicativamente). Essendo la temperatura un parametro critico nell'ottica di una produzione su larga scala dei materiali 2D, lo sviluppo di metodi di sintesi che richiedano una minore temperatura di processo è altamente desiderabile. Per questo motivo, altre tecniche di deposizione, alternative alla CVD, vengono studiate e ottimizzate per la produzione di materiali 2D. In particolare, le tecniche di *deposizione fisica da vapore* (PVD) consentono generalmente di produrre film ultra-sottili di diversi materiali con temperature di sintesi più basse rispetto alla CVD. Per PVD si intendono processi di deposizione nei quali il materiale viene evaporato tipicamente da una sorgente solida in forma di atomi o molecole e trasportato in forma vapore o plasma fino al substrato su cui condensa. Tra le tecniche PVD, è interessante menzionare la *deposizione a impulsi laser* (PLD), che si basa sull'abla-

zione di un bersaglio solido da parte di impulsi laser ad alta energia. L'interazione laser-materia causa l'ablazione del bersaglio e la conseguente espulsione del materiale ablato sotto forma di una "piuma" di plasma che si espande verso il substrato, sul quale si forma il film. Si tratta di una tecnica di deposizione molto versatile che permette di produrre film e nanostrutture di vari materiali e morfologie al variare di diversi parametri di processo, quali l'energia degli impulsi, il numero totale di impulsi sul bersaglio, la loro frequenza di ripetizione, il gas all'interno della camera di deposizione e la sua pressione. Grazie a tale versatilità, si può applicare la PLD alla sintesi di materiali 2D su scala del centimetro, utilizzando temperature di crescita relativamente basse (400-600°C) e fornendo un ottimo grado di controllo sul numero di strati atomici e sulla stechiometria dei film prodotti [9].

4. Osservare i materiali 2D: tecniche di caratterizzazione

Una volta sintetizzati, è possibile "osservare" i materiali 2D tramite varie tecniche sperimentali, in grado di rivelarne le caratteristiche morfologiche, strutturali, elettroniche, ottiche, elettriche, magnetiche, chimiche, ecc. In generale, vista la natura bi-dimensionale dei materiali da studiare, risultano particolarmente efficaci le tecniche di caratterizzazione proprie della Scienza delle Superfici, che offrono la possibilità di "sondare" con sufficiente sensibilità pochi strati atomici. Alcune tra queste tecniche forniscono informazioni sulle proprietà strutturali: ad esempio, la *microscopia elettronica a trasmissione* (TEM) e le *microscopie a scansione di sonda* (SPM) permettono di osservare direttamente il reticolo cristallino con elevata risoluzione spaziale (nanometrica e subnanometrica), mentre tecniche di diffrazione quali la *diffrazione di elettroni a bassa energia* (LEED) e la *diffrazione superficiale di raggi X* (SXRD) consentono di analizzare la struttura del reticolo reciproco. La struttura geometrica del reticolo atomico che costituisce il materiale contribuisce in maniera fondamentale a determinarne la struttura elettronica. Quindi è importante poter disporre di tecniche in grado di rivelare sia le caratteristiche strutturali sia quelle elettroniche. La *microscopia a effetto tunnel* (STM), capostipite delle tecniche SPM, offre questa possibilità, in quanto permette di osservare la superficie del materiale con *risoluzione atomica* e di rivelarne le proprietà elettroniche locali. Questa tecnica fu inventata nel 1981 da Gerd Binnig and Heinrich

Röhrer, che vennero per questo insigniti del premio Nobel per la Fisica nel 1986 [10]. Il principio di funzionamento di questa tecnica si basa sull'*effetto tunnel*, noto effetto quantistico per cui gli elettroni hanno probabilità non nulla di attraversare una barriera di potenziale tra due elettrodi, se posti a una distanza sufficientemente piccola tra loro (dell'ordine di 1 nm). Nel microscopio a effetto tunnel i due elettrodi sono una sottile punta metallica e la superficie del campione: una differenza di potenziale tra i due (dell'ordine di 1 V) determina il passaggio di una corrente di tunnel (dell'ordine di 1 nA) che dipende sia dalla distanza punta-campione sia dalle proprietà elettroniche superficiali. La "scansione" della punta sulla superficie e la simultanea misura della corrente di tunnel permettono di ottenere, come risultato della misura STM, un'immagine della superficie caratterizzata da elevatissima risoluzione spaziale (fino a raggiungere la risoluzione atomica) e sensibile alla struttura elettronica locale. In particolare, la sensibilità agli stati elettronici superficiali può essere maggiormente sfruttata grazie alla tecnica di *spettroscopia a effetto tunnel* (STS), tecnica figlia della STM, che consente di ottenere uno spettro della *densità locale di stati elettronici* (LDOS) nell'intorno del livello di Fermi, fornendo così informazioni di carattere locale sulle proprietà elettroniche del materiale. Questa tecnica è complementare ad altre tecniche di spettroscopia elettronica comunemente utilizzate, come la *spettroscopia di fotoemissione.* Altre tecniche spettroscopiche, quali la *spettroscopia Raman* e la *spettroscopia di fotoluminescenza* (PL), consentono di studiare, rispettivamente, le proprietà vibrazionali e ottiche dei materiali 2D. La spettroscopia Raman rivela gli stati vibrazionali del materiale e risulta essere molto sensibile al numero di strati atomici di cui è composto un materiale 2D. La spettroscopia di fotoluminescenza misura a varie lunghezze d'onda la luce riemessa dal materiale in seguito ad un'eccitazione ottica, e si è dimostrata molto efficace nello studio delle proprietà optoelettroniche dei TMD, consentendo di osservare la transizione da semiconduttore a gap indiretto a diretto che caratterizza questi materiali al limite 2D.

4.1 Osservazioni STM di alcuni materiali 2D

In questa sezione verranno presentati alcuni risultati ottenuti da studi STM effettuati su tre materiali 2D: l'ossido di zinco (ZnO) 2D, l'ossido di titanio (TiO_x) 2D e il singolo strato di MoS_2. Tali studi sono stati realizzati presso il Laboratorio Materiali Micro e Nanostrutturati

del Politecnico di Milano. Il primo esempio riguarda lo ZnO, uno dei materiali ossidi più studiati e utilizzati in numerose applicazioni. La sua forma bulk più stabile ha una struttura di tipo *wurtzite*, in cui ogni atomo di Zn (O) è legato a quattro atomi di O (Zn) in configurazione tetraedrica. Recentemente, studi teorici [11], poi confermati sperimentalmente, hanno dimostrato che, al limite 2D, la struttura più stabile non è più quella wurtzitica, bensì una struttura di tipo grafite, in cui gli atomi di Zn e O si dispongono a nido d'ape in piani atomici perfettamente planari. La *Fig. 1(a)* mostra un'immagine STM in cui è possibile osservare dei *nanocristalli* di ZnO 2D, prodotti tramite PLD su una superficie di oro. Grazie all'elevata risoluzione spaziale, sia verticale che laterale, del microscopio STM è possibile determinare lo spessore dei nanocristalli e, quindi, il numero di strati atomici che li compone. In particolare, in *Fig. 1(b)*, si possono osservare due nanocristalli composti da uno e due strati atomici. Le immagini STM in *Fig. 1(c,d)* mostrano la superficie di questi nanocristalli a risoluzione atomica. Da immagini come queste è possibile osservare direttamente il reticolo atomico dello ZnO 2D, caratterizzato da simmetria esagonale e da una cella unitaria di circa 0.32 nm [12]. Per quanto detto, lo ZnO è un esempio evidente di come al limite 2D i materiali possano presentare strutture nuove rispetto a quelle note in forma bulk. Un ulteriore esempio di questa proprietà generale è fornito dall'ossido di titanio (TiO_x). Al limite 2D, questo materiale forma strutture molto diverse rispetto a quelle bulk (*rutilo* e *anatasio*) in seguito al confinamento dimensionale e all'interazione con il substrato sul quale sono sintetizzate. Ad esempio, le immagini STM in *Fig. 2(a,b)* mostrano una particolare struttura del TiO_x simile a quella del grafene, indotta dalla simmetria esagonale della superficie (111) dell'oro. Il modello strutturale di questa fase, ottenuto da un'analisi teorica, è raffigurato in *Fig. 2(c,d)* e mostra la disposizione a nido d'ape degli atomi di Ti e O, che formano un singolo strato atomico [13]. Il terzo esempio, mostrato in *Fig. 3*, si riferisce al singolo strato di MoS_2, uno dei materiali 2D più studiati. L'immagine STM in *Fig. 3(a)* mostra un tipico nanocristallo formato da un solo strato di MoS_2 sulla superficie (111) dell'oro. La caratteristica più evidente della superficie del MoS_2 è la presenza di una modulazione periodica esagonale, che si può meglio osservare in *Fig. 3(b)*, insieme al reticolo atomico. Tale modulazione può essere interpretata prendendo in prestito un concetto proprio dell'Ottica: la *figura (o pattern) di moiré*. Questa si forma quando due

reticoli aventi passo reticolare o orientazione diversi vengono sovrapposti. Nel nostro caso i due reticoli sono il reticolo atomico del MoS_2 sovrapposto al reticolo atomico dell'oro. Come mostrato nella simulazione di *Fig. 3(c)*, tale sovrapposizione dà luogo a un super-reticolo esagonale che descrive la modulazione periodica osservata nelle immagini STM [14]. La formazione di un pattern di moiré non è una caratteristica esclusiva del MoS_2 su Au(111), ma si osserva comunemente in molti sistemi composti da materiali 2D, dove può indurre modifiche nella struttura elettronica del materiale.

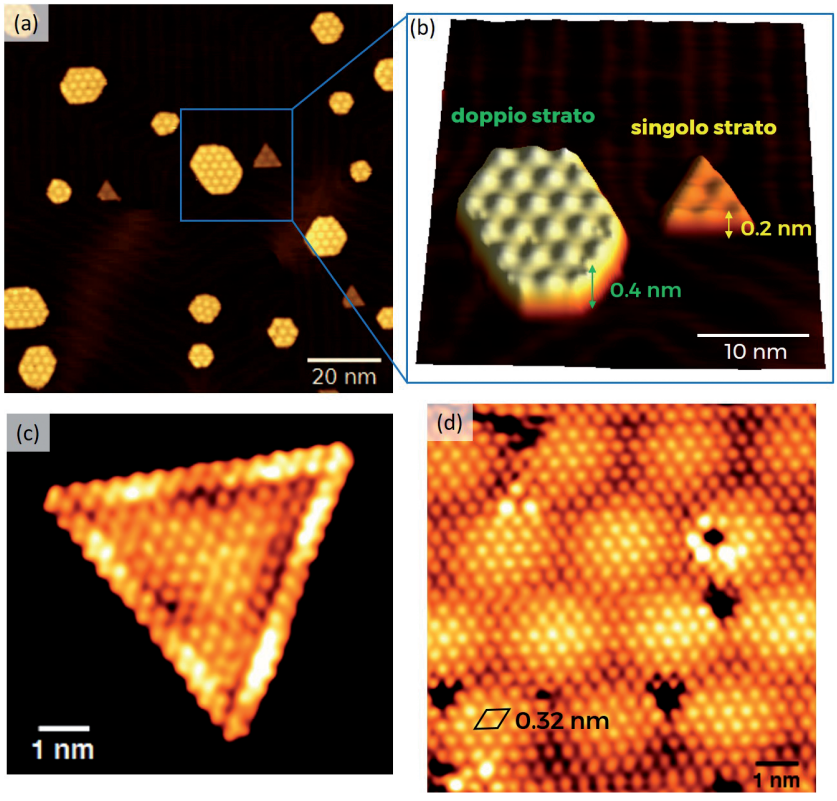


Fig. 1 – Immagini STM di nanocristalli di ZnO 2D sintetizzati tramite PLD sulla superficie (111) dell'oro. (a) Immagine STM di 200x200 nm². (b) Immagine STM tri-dimensionale di due nanocristalli di cui è indicato lo spessore misurato rispetto al substrato. Quello più sottile è composto da uno strato atomico, mentre l'altro da due. (c)-(d) Immagini STM a risoluzione atomica della superficie di due nanocristalli: la cella unitaria dello ZnO misura 0.32 nm.

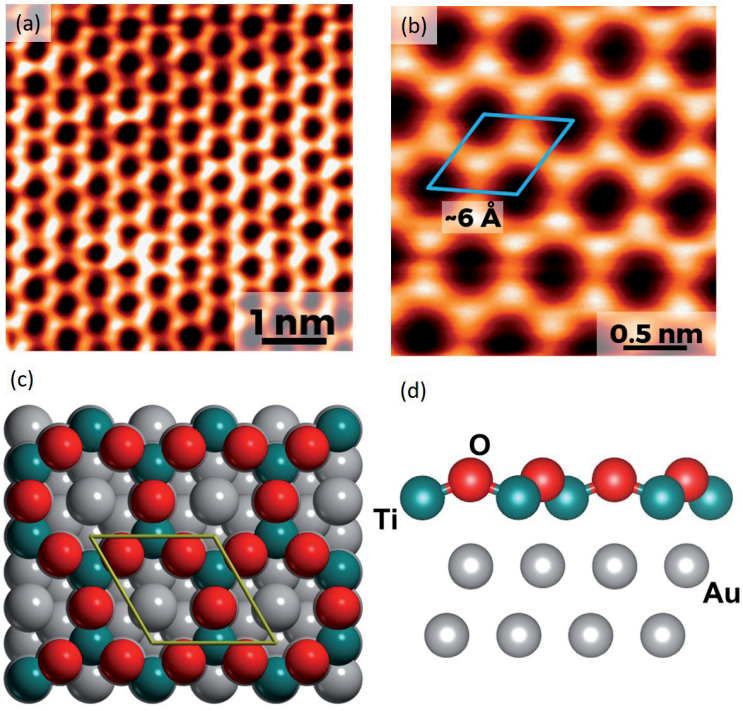


Fig. 2 – (a) Immagine STM della struttura a nido d'ape del TiOx 2D su Au(111). (b) Immagine STM a risoluzione atomica della stessa struttura. (c)-(d) Vista dall'alto e laterale del modello strutturale ottenuto da simulazioni teoriche. Legenda: rosso = ossigeno, verde = titanio, grigio = oro.

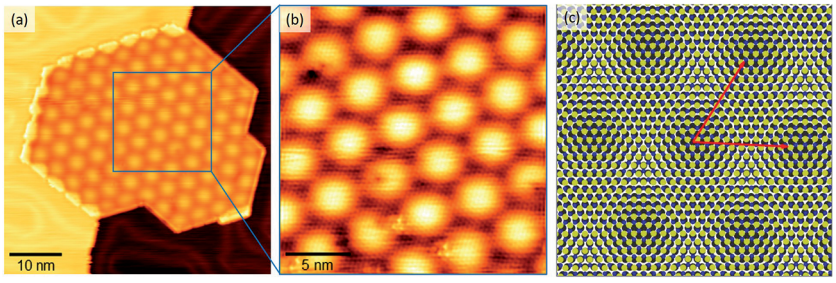


Fig. 3 – (a) Immagine STM di un nanocristallo formato da un singolo strato di MoS2 sintetizzato tramite PLD su Au(111). (b) Immagine STM a risoluzione atomica della superficie di MoS2, in cui si osserva il pattern di moiré esagonale. (c) Modello atomistico della sovrapposizione del reticolo di MoS2 su quello di oro: i due vettori (in rosso) generano il super-reticolo che descrive la geometria del pattern di moiré.

5. Conclusioni

I materiali 2D sono materiali solidi ultrasottili, dello spessore al più di qualche strato atomico. Capostipite di questa classe di materiali è il grafene, il cui studio, a partire dal 2004, ha dato il via alla ricerca sui materiali 2D e lo loro uniche proprietà. Tali proprietà differiscono da quelle dei comuni materiali formati da molti strati atomici, per via di effetti fisici legati principalmente al confinamento bi-dimensionale. La ricerca sulle proprietà fondamentali e sui metodi di fabbricazione di questi materiali è in continua crescita e promette di aprire nuovi scenari di sviluppo tecnologico basati sulla manipolazione su scala nanometrica di materiali economici e facilmente reperibili. Tale sviluppo potrà interessare diversi ambiti applicativi, ma specialmente quello elettronico e optoelettronico, dove i limiti fisici imposti alla miniaturizzazione dei dispositivi rendono necessario concepire tecnologie alternative per realizzare sistemi sempre più avanzati. Oltre l'interesse applicativo, è importante sottolineare come lo studio di questi materiali ci metta a confronto con l'osservazione, l'analisi e la comprensione di proprietà e comportamenti fisico-chimici ancora sostanzialmente inesplorati. La ricerca sui materiali 2D segna, pertanto, un passaggio importante nella capacità dell'uomo di manipolare ed "ingegnerizzare" la materia su scala nanometrica, sulla quale si baserà lo sviluppo delle future nanotecnologie.

6. Ringraziamenti

Si ringrazia il Laboratorio Materiali Micro e Nanostrutturati del Dipartimento di Energia del Politecnico di Milano, presso il quale sono stati ottenuti i risultati sperimentali presentati.

BIBLIOGRAFIA

1. Geim AK, Novoselov KS. The rise of graphene. *Nature Materials*, 2007: 6(3): 183-191.
2. Butler SZ., et al. Progress, challenges, and opportunities in two-dimensional materials beyond graphene. *ACS Nano*, 2013: 7(4): 2898-2926.
3. Fiori G, Bonaccorso F, Iannaccone G, Palacios T, Neumaier D, Seabaugh A, Banerjee SK, Colombo L. Electronics based on two-dimensional materials. *Nature Nanotechnology*, 2014: 9(10): 768-779.

4. Xia F, Wang H, Xiao D, Dubey M, Ramasubramaniam A. Two-dimensional material nanophotonics. *Nature Photonics*, 2014: 8(12): 899-907.
5. Wang QH, Kalantar-Zadeh K, Kis A, Coleman JN, Strano MS. Electronics and optoelectronics of two-dimensional transition metal dichalcogenides. *Nature Nanotechnology*, 2012: 7(11): 699-712.
6. Pacchioni G. Two dimensional oxides: Multifunctional materials for advanced technologies. *Chemistry-A European Journal*, 2012: 18: 10144-10158.
7. Geim AK, Grigorieva IV. Van der Waals heterostructures. *Nature*, 2013: 499: 419-425.
8. Novoselov KS, Jiang D, Schedin F, Booth TJ, Khotkevich VV, Morozov SV, Geim AK. Two-dimensional atomic crystals. *Proceedings of the National Academy of Sciences of the United States of America*, 2005: 102(30): 10451-10453.
9. Yang Z, Hao J. Progress in pulsed laser deposited two-dimensional layered materials for device applications. *Journal of Materials Chemistry C*, 2016: 4: 8859-8878.
10. Binnig G, Rohrer H, Gerber C, Weibel E. Surface studies by scanning tunneling microscopy. *Physical Review Letters*, 1982: 49(1): 57-61.
11. Freeman CL, Claeyssens F, Allan NL, Harding JH. Graphitic nanofilms as precursors to wurtzite films: theory. *Physical Review Letters*, 2006: 96(6): 066102.
12. Tumino F, Casari CS, Passoni M, Bottani CE, Li Bassi A. Pulsed laser deposition of two-dimensional ZnO nanocrystals on Au (111): growth, surface structure and electronic properties. *Nanotechnology*, 2016: 27(47): 475703.
13. Tumino F, Carrozzo P, Mascaretti L, Casari CS, Passoni M, Tosoni S, Bottani CE, Li Bassi A. Two-dimensional TiOx nanostructures on Au (111): a scanning tunneling microscopy and spectroscopy investigation. *2D Materials*, 2015: 2(4): 045011.
14. Tumino F, Casari CS, Passoni M, Russo V, Li Bassi A. Pulsed laser deposition of single-layer MoS2 on Au(111): from nanosized crystals to large-area films. Nanoscale Advances, 2019: 1: 643-655.

Scienze naturali
Biologia
Istituto Lombardo (Rend. Scienze) 151, 255-258 (2017)

LA RIVOLUZIONE DELLA CRIO-MICROSCOPIA ELETTRONICA NELLO STUDIO STRUTTURALE DI PROTEINE E ACIDI NUCLEICI

Nota del s.c. MARTINO BOLOGNESI (*)

(Adunanza del 16 novembre 2017)

SUNTO. – La Biologia Strutturale si è occupata negli ultimi 50 anni di studiare i dettagli fini delle strutture molecolari subcellulari (es. proteine o acidi nucleici). In tempi recenti, lo sviluppo della crio-microscopia elettronica in singola particella ha portato ad una nuova rivoluzione in questo campo di studio, in precedenza dominio della cristallografia a raggi X e della risonanza magnetica nucleare. Oggi è possibile studiare direttamente le strutture dei grandi aggregati macro-molecolari (es. il ribosoma) con un potere risolutivo che, nei casi migliori, giunge a 2 Å, permettendo di distinguere i dettagli fini di aminoacidi e nucleotidi.

***

ABSTRACT. – Observing the fine details of molecular structures (*e.g.* in proteins and in nucleic acids) has been a central part of Structural Biology over the past 50 years. The recent advent of single particle cryo-electron microscopy brought a revolution in this field, that previously relied on X-ray crystallography and nuclear magnetic resonance. It is now possible to explore the structures of large subcellular assemblies, such as the ribosome, resolving details on the scale of amino acids and nucleotides, in favorable cases reaching the 2 Å resolution level.

La comprensione delle basi molecolari dei processi biologici si fonda sulla conoscenza delle modalità di interazione tra atomi/amino-acidi/nucleotidi e altre molecole diverse in uno specifico contesto cellulare o macromolecolare. Interessa anche come si formino/rompano

(*) Dipartimento di Bioscienze, Università degli Studi di Milano, Italy.
E-mail: martino.bolognesi@unimi.it

legami covalenti, come siano organizzate le strutture tridimensionali (3D) delle proteine, degli acidi nucleici e di altre (macro)molecole biologiche. Sono queste le sfide scientifiche della *Biologia Strutturale*, che nel corso degli ultimi decenni ha visto riconosciuti diversi premi Nobel nei campi della cristallografia a raggi-X e della risonanza magnetica nucleare (NMR) (per una panoramica dei risultati prodotti da questi due metodi di studio si consulti il data-base Protein Data Bank: www.rcsb.org/pdb). Cristallografia e NMR hanno fornito informazioni preziose, ad esempio per lo sviluppo di nuovi farmaci, producendo più di 140.000 strutture 3D di proteine, o loro complessi, a risoluzione atomica.

L'analisi di corpuscoli sub-cellulari, quindi "oggetti" molto piccoli, può essere svolta con un microscopio ottico. Tuttavia, la risoluzione spaziale di questo strumento cade in difetto quando il campione/dettaglio da analizzare sia di dimensioni inferiori a quelle della lunghezza d'onda della luce visibile (orientativamente minore di 300 nm). Grazie agli studi di de Broglie, nei primi decenni del XX secolo, sappiamo che particelle cariche in movimento godono delle proprietà ondulatorie caratteristiche di un fascio di luce. Essendo gli elettroni particelle elementari dotate di una carica negativa, essi possono essere diretti sul campione, realizzando uno strumento che ha le caratteristiche di un vero e proprio microscopio, ma con capacità risolutive molto maggiori del microscopio ottico. In accordo con questa osservazione, negli anni '30 del '900, Ernst Ruska, riuscì a focalizzare un fascio di elettroni attraverso il campo magnetico generato da un solenoide, realizzando di fatto la prima "lente per elettroni", e costruì il primo prototipo di microscopio elettronico. Tale scoperta fu riconosciuta con il Nobel per la Fisica nel 1986 e con l'inizio della diffusione di questi strumenti.

L'evoluzione tecnologica dei microscopi elettronici ha reso disponibile un potentissimo e nuovo metodo di indagine in Biologia, dapprima, e in tempi recenti anche nell'ambito della Biologia Strutturale. Gli elettroni infatti interagiscono con gli atomi in maniera più efficiente dei raggi X, possono essere focalizzati con lenti elettromagnetiche, presentano lunghezze d'onda molto brevi (dell'ordine di qualche centesimo di Å) dal punto di vista ondulatorio, e permettono di esaminare i campioni in soluzione o addirittura in cellula. In linea di principio, la microscopia elettronica può quindi permettere di risolvere atomi di carbonio legati covalentemente in una proteina o in una molecola di DNA, le cui distanze cadono nell'intervallo 1.2 – 1.6 Å.

Tuttavia, l'applicazione della microscopia elettronica a campioni biologici non è così immediata come poterbbe apparire. Infatti, L'energia degli elettroni accelerati nel microscopio viene trasferita al materiale organico in esame e lo danneggia sensibilmente. Inoltre, per consentire il percorso ottico degli elettroni attarverso lo strumento, la camera interna del microscopio deve essere mantenuta ad elevati livelli di vuoto ($< 10^{-7}$ Pa), notoriamente poco compatibili col mantenimento di condizioni native per un campione biologico.

Per ovviare a questi inconvenienti si è sviluppata la *cryo-electron microscopy* (Cryo-EM), metodo secondo cui il campione biologico in soluzione viene rapidamente congelato per immersione in un agente criogenico (ad esempio, etano o azoto liquidi), tramite un processo definito come "vitrificazione". A temperature dell'azoto liquido il danno da radiazione è ridotto, in quanto i radicai liberi creati dall'interazione del campione con gli elettroni hanno ridotte possibilità di diffusione, quindi di danno alle molecole presenti nel campione.

L'esperienza maturata nelle prime applicazioni Cryo-EM, tuttavia, ha indicato che il crio-congelamento da solo risolve in maniera sufficiente il problema del danno indotto dalla radiazione elettronica. Una vera "rivoluzione" in questa direzione si è riscontrata a partire dal 2012, con lo sviluppo di una nuova generazione di dispositivi, definiti *Direct Electron Detectors* (DED), caratterizzati da una elevata efficienza nella registrazione del segnale. A differenza di altri dispositivi (lastre fotografiche o CCD), in un DED l'elettrone viene registrato direttamente dal sensore costituito da ossidi metallici semiconduttori (tecnologia CMOS, che non richiede conversioni del segnale elettronico in radiazione del visibile); inoltre, il sensore DED presenta una velocità di acquisizione elevata, che permette di registrare centinaia di immagini/secondo. Per acquisire un segnale trattabile, un DED richiede che il campione venga irradiato con dosi bassissime di elettroni; pertanto si affianca in mnaiera sinergistica al processo di vitrificazione come espediente attivo nella riduzione del danno da radiazione al campione.

In un esperimento di crio-microscopia elettronica a singola particella (spCryo-EM) per studiare la struttura 3D di un complesso proteico (vitrificato in soluzione), migliaia d'immagini, dette *micrographs*, vengono registrate nello strumento (1-5 Tbytes di dati per esperimento). Ogni *micrograph* contiene centinaia di macromolecole, dette *single particles*, distribuite nella soluzione vitrificata. L'insieme complessivo dei dati acquisiti raccoglie centinaia di migliaia di campionature di sin-

gole molecole che vengono riconsosciute singolarmente e combinate secondo orientazione e conformazione assunta, al fine di sommare i loro segnali e minimizzare il rumore di fondo. Da questa elaborazione si passa poi ad una procedura computazionale più impegnativa, per teoria, carico di calcolo e tempi di esecuzione. Questa fase permette di ricostruire la struttura 3D del complesso proteico (la *single particle*) in esame; con gli strumenti più avanzati disponibili, la risoluzione raggiungibile si attesta tra 3 e 2 Å, a seconda delle qualità del campione, e in casi favorevoli l'analisi può richiedere 2-3 settimane.

L'impatto della spCryo-EM sulla Biologia Strutturale, e sul mondo della Biologia in generale, è stato visibile e sostanziale a partire dal 2015, con sviluppi e installazioni molto importanti nei principali laboratori di ricerca di tutto il mondo. In particolare, un intero campo di studio (quello delle proteine di membrana, che sono i bersagli di almeno il 30% dei farmaci maggiormente attivi) ha visto una progressione insperata e promettente, sia per le potenzialità applicative, sia per la rapidità delle risposte sperimentali. Non è un caso che il Premio Nobel per la Chimica 2017 sia stato conferito a Jacques Dubochet (Università di Losanna), Joachim Frank (Columbia University, New York) e Richard Henderson (MRC Laboratory of Molecular Biology, Cambridge, UK) "*for developing cryo-electron microscopy for the high-resolution structure determination of biomolecules in solution*".

Istituto Lombardo (Rend. Scienze) 151, 259-277 (2017)

A HOLISTIC APPROACH TO THE SPACE DEBRIS MITIGATION

ALESSANDRO ROSSI (*)

Nota presentata dal s.c. Antonio Giorgilli
(Adunanza del 30 novembre 2017)

Sunto. – Le attività spaziali in quasi tutti i regimi orbitali sono ora a rischio a causa della continua crescita della popolazione dei detriti spaziali. Per prevenire la proliferazione dei detriti spaziali in orbita terrestre è necessario affrontare il problema da prospettive diverse, sfruttando le più avanzate conoscenze teoriche e sperimentali in campi diversi come l'astrodinamica, l'ingegneria aerospaziale e la giurisprudenza spaziale. Questo approccio permette di affrontare quattro pilastri principali: prevenzione, protezione, mitigazione e regolamentazione. A tale riguardo, la Comunità europea ha finanziato un progetto H2020 chiamato ReDSHIFT il cui obiettivo è trovare metodi passivi (ovvero che non richiedano uso di propellente aggiuntivo e l'intervento di veicoli spaziali ad hoc) per mitigare la proliferazione dei detriti spaziali.

* * *

Abstract. – The space activities in almost every orbital regime are now jeopardized by the continuous growth of the space debris populations. To prevent the proliferation of the space debris in Earth orbit it is necessary to tackle the problem from different perspectives, exploiting the latest theoretical and experimental knowledge in different fields, such as astrodynamics, spacecraft engineering and legal studies, to address four main pillars: prevention, protection, mitigation and regulation. In this respect the European Community financed a large H2020 project named ReDSHIFT whose goal is to find passive means to mitigate the proliferation of space debris. A short summary of the project and of its main findings is given in the paper, with particular emphasis on the more theoretical part, related to the simulation of long term evolutionary scenarios of the whole space debris environment and to the mapping of the Low Earth Orbit phase space, looking for passive dynamical de-orbiting solutions.

(*) IFAC-CNR, Sesto Fiorentino (FI), Italy. E-mail: a.rossi@ifac.cnr.it

1. Introduction

Our everyday life is increasingly relying on space. Telecommunications, global positioning, meteorological forecasts, disasters prevention and mitigation, safety and security are just few examples of the fields in which space is now a crucial resource. Nonetheless, the increasing number of space debris in Earth orbit is seriously jeopardizing the space assets in all the orbital regimes, from the very low orbits (at around 400-500 km of altitudes) up to the Geostationary region (at around 36000 km of altitude) where the TV, telecommunication and meteorological satellites are located.

The impact of debris on the space activities has to be reduced by adopting a global strategy able to address the problem from different points of view, from the very beginning of the planning of a space mission. As an example, the choice of the orbit, of the spacecraft bus, of the spacecraft power system and propulsion, are all aspects that influence, and have to be optimized, having in mind not only the goal of the mission but also the minimization of the "environmental" impact of the spacecraft. The space debris related aspects can be summarized in terms of: prevention, protection, mitigation and regulation. These aspects will be analyzed in the following by describing a large Horizon 2020 project, named ReDSHIFT, which is currently undergoing and is addressing all the above issues.

We note that, if none of the above measures proves to be effective enough, the remediation measures should be considered. This can imply the need to realize a dedicated space mission to actively remove hardware from space. The Active Debris Removal (ADR) solution is not part of the ReDSHIFT project and will not be discussed here. The interested reader can refer to, e.g., [3].

2. The ReDSHIFT H2020 project

In order to tackle the issues mentioned in the previous section, a large research project, called ReDSHIFT (Revolutionary Design of Spacecraft through Holistic Integration of Future Technologies) has been funded by the European Union in the framework of the PROTEC Call of Horizon 2020. The ReDSHIFT project is coordinated by the author, at IFAC-CNR (see http://redshift-h2020.eu/). In ReDSHIFT the debris mitigation goals will be achieved through a holistic approach that considers,

from the outset, opposing and challenging constraints for the space environment preservation, the spacecraft survivability in the harsh space environment and the safety of humans on ground.

The main goal of ReDSHIFT is to tackle the space debris issue from a global perspective using expertises from several different fields: long term simulations of the space debris environment, astrodynamics, 3D printing, design for demise, protection and hypervelocity impact testing, legal and normative issues. The overall objective is to push for new, improved mitigation measures with the use of new theoretical results about the orbital dynamics of spacecraft coupled with new technologies, such as solar and drag sails (i.e., extremely light structures increasing the area of the spacecraft and exploiting either the gentle push of the solar radiation pressure -solar sail- or the drag produced by the thin residual atmosphere in case of low orbits, below about 800 km of altitude). The new paradigm of 3D printing, should be applied to enhance different aspects of the spacecraft design and manufacturing concurring to the mitigation efforts, such as, e.g., sail attachment and storage, shielding and design for demise.

The plot of the project is the following: first a through analysis of the currently adopted mitigation measures is performed to highlight their benefits and, possibly, their deficiencies in some aspects. This analysis is assisted by a number of simulations of the long term evolution of the space debris environment showing the overall effects of these measures in a measurable way. These simulations will serve also a the reference to assess the results achieved at the end of the project. This part will be described in Sec. 3.

While it is well known that the fundamental step to be performed to preserve the space environment is the disposal of the spacecraft at end-of-life, sometime the maneuvers needed to achieve this goal are not practicable also for energetic or technical reasons. Therefore, a thorough understanding of the orbital dynamics in all the circumterrestrial space will allow us to identify stability and instability regions, and to exploit them to find preferential routes (we called them “de-orbiting highways”) minimizing the energetic requirements for the operators, thus improving the applicability of the disposal maneuvers. Once identified, the best maneuvers needed to reach the “entrances” to the de-orbiting highways or the stable graveyard regions (in case the reentry option is not viable with the current technologies, e.g., for Geostationary satellites), coming from the used operational orbits, will be computed and the technical means to be

used to accomplish those maneuvers will be identified. This part will be described in Sec. 4.

As a matter of fact, the project is focusing mainly on a few specific passive technologies (that is technologies not requiring propellant on board) such as solar and drag sails. The studies concerning the sails will be twofold: the theoretical aspects of the sail dynamics and the technological aspects of sail manufacturing will be both tackled. Moreover, since our main mitigation solution will be the de-orbiting towards the atmospheric reentry, the focus will also be on spacecraft designed for demise, that is a design that will minimize the chances that chunks of the spacecraft might reach the ground.

In an effort to make the above technological solutions easier and more attractive to produce and to implement into future spacecraft design, ReDSHIFT will explore the possibility to use the now blossoming additive manufacturing (3D printing) technology, to actually realize a model spacecraft and, in particular, some specific parts related to the debris mitigation issues, such as, e.g., the shielding, a sail canister, the sail hatches and joints, etc.

Once designed and realized, the prototypes will undergo specific test such as hypervelocity impacts, radiation, vibration, thermal testing, etc. This part will be briefly described in Sec. 5.

In all the above process, a common thread linking highly theoretical analysis and more applied physics and engineering studies will be kept, in a constant attempt to stick to the proposed holistic vision.

At the end of the project the effectiveness of the proposed dynamical and technological solutions will be validated by repeating the long term simulations, and comparing the results with the ones obtained at the beginning of the project.

Beside the 3D parts, the final outputs of the projects will include also a software summarizing the theoretical and engineering findings, allowing the design of a space debris compliant mission (e.g., by suggesting the disposal trajectories and the technologies needed to achieve them, the best shielding opportunities for a given spacecraft and the possibility to produce it with additive manufacturing, etc.). It will output safe, scalable and cost-effective satellite and mission designs in response to operational constraints.

Last, but not least, all the experience gathered will be applied to the analysis of the normative related to space debris, in an effort to propose, to the proper forums new, improved mitigation practices and rules (Sec. 6).

3. Long term simulations of the space debris environment

The capability to model the long term evolution of the space debris environment is of paramount importance both to design the future missions and to test and assess the effectiveness of the proposed mitigation measures. In the last decades increasingly complex mathematical models were developed by the author's group, from the statistical one described in [10] to the semi-deterministic one, called *Space Debris Mitigation (SDM)* described in [11]. In these models, an initial population of objects present in space is evolved for time spans of the order of a few hundreds of years, integrating the orbits of every object and including all the source (e.g., launches, explosions, collisions, etc.) and sinks (e.g., natural perturbations such as air drag, active debris removal, mitigation measures, etc.) terms that are responsible for the evolution of the environment. Of particular importance is to be able to model the proposed mitigation measures, that is mainly operational practices aimed at reducing the orbital lifetime of spacecraft, once their mission is accomplished. In Low Earth Orbit (LEO, defined as the region of space from 0 to 2000 km of altitude) the main guideline is to dispose the spacecraft into an orbit whose residual lifetime, under the effects of the natural perturbations (atmospheric drag mainly), is not exceeding 25 year. This is the so-called *25-year rule* widely adopted by the space agencies worldwide as a pillar of the mitigation strategies aimed at protecting the LEO zone.

To assess and quantify the effectiveness of the most common measures, specific long term debris environment simulations were performed with SDM (and with a similar software developed by the Technical University of Braunschweig, called LUCA [9]). In particular, the simulations included an analysis of the effects of the level of compliance to the mitigation measures (i.e., how many missions actually respect the proposed guidelines - see Sec.6), of the residual lifetime of the spacecraft at their end-of-life (25 year with respect to 10 years), of the collision avoidance procedures, of the active debris removal and of the proposed launch in space of very large constellations of satellites in Low Earth orbit [8].

As an example, the left panel of *Fig. 1* shows the comparison of the effective number of objects in LEO for three simulated scenarios, over a time span of 200 years. The blue line is the "Reference" case and it depicts a scenario where the space operations are supposed to be performed in a way similar to the one currently adopted. In particular it is assumed that the spacecraft are de-orbited at the end-of-life respecting the 25-year rule,

but only with a 60 % compliance level. The magenta curve is based on the same Reference scenario, but it assumes the use of Active Debris Removal of two objects per year. In particular, every year, the two most dangerous objects (i.e., the objects having the highest product between their mass and their collision probability against the other objects in space) are removed from the simulation. Finally, the red line refers to a scenario like the Reference one, but the spacecraft are moved at the end-of-life to an orbit with only 10 years of residual lifetime and, moreover, a 90 % compliance to this 10-year rule is simulated. The right panel of *Fig. 1* shows the cumulative number of catastrophic collisions happening in the simulated time span for the three scenarios (with the same color code). A few comments to *Fig. 1* are in order:

- the improved compliance to the currently proposed mitigation measures gives the best results in terms of the growth of the LEO population and of the expected catastrophic collisions;

- notwithstanding the above mentioned improvement, the environment is still "unstable" and we observe a constant growth of the number of objects in LEO;

- the active removal of just two objects per year is not significantly improving the situation, but still produces a decrease of about 12 % of the final population. It is worth mentioning here that the technology for ADR is now almost mature and about to be tested. On the other hand the political and economic issues related to future ADR missions are still to be overcome.

The results of the simulations give the expected quantitative evaluation of the effectiveness of the different measures and represent the reference against which to compare the final results of the ReDSHIFT project.

Whereas the full details of the simulation results can be found in [13], here it is important to raise a warning that is evident looking at *Fig. 2*.

The Figure shows the spatial density of objects in LEO for the Reference and the improved mitigation scenarios for the epoch 2213 (end of the simulated time span) as a function of altitude, compared with the initial epoch. A general improvement of the situation can be noted, especially in the crowded critical regions around 900 and 1400 km, due to the increased number of satellites de-orbited at the end-of-life. Nonetheless, looking carefully at the higher end of the plot, around 2000 km of

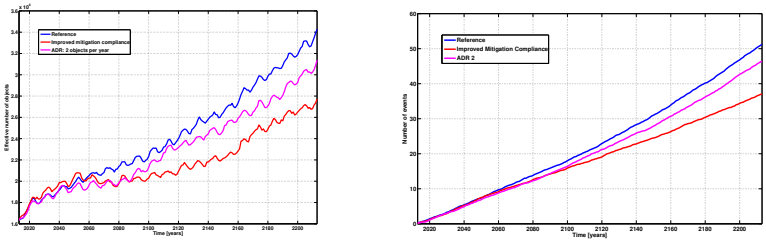


Fig. 1 – Effective number of objects (left panel) and cumulative number of fragmentations (right panel), in LEO, for three different evolutive scenarios. See text for details

altitude, an opposite behavior, with a growth of the density in the mitigation scenario, can be noticed. This is due to the following reason: for high LEO orbits (above about 1400 km of altitude) whenever an object has to be disposed at the end-of-life, the best solution, in terms of the ΔV required for de-orbiting, might be not the usual de-orbiting to a lower elliptical orbit with a defined residual lifetime. Instead, a re-orbiting into a circular "graveyard" orbit above the LEO protected zone (i.e., above 2000 km of altitude) can be energetically more convenient. Hence, a number of objects is usually moved, to this kind of graveyard. Therefore the mathematical model is able to choose the optimal solution between the two options (in terms of the required ΔV). In the plot of *Fig. 2* we are starting to see not only the simple effect of the accumulation of the disposed satellites in the super-LEO graveyard, but also the growth of fragments generated by the mutual collisions between the uncontrolled disposed spacecraft (unable to perform avoidance maneuvers).

This result reinforces the conclusion that novel disposal means should be devised also to lower the energetic requests of de-orbiting maneuvers, thus minimizing the recourse to the LEO graveyard zones. The mapping of the phase space described in the next Section, with the search for the "de-orbiting highways" is devoted exactly to this purpose.

4. Dynamical mapping of the circumterrestrial space

As shown in the previous section, the disposal of objects outside the LEO protected region is one of the most important mitigation measure needed to reduce the growth of space debris in the forthcoming years. In order to

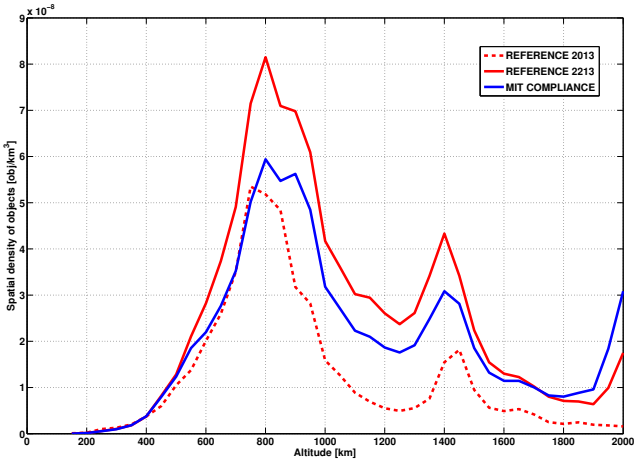


Fig. 2 – Spatial density of objects (objects/km^3) larger than 10 cm in LEO for the Reference (red line) and improved mitigation compliance cases (blue line) in the year 2213. The dashed red line represents the initial density at epoch 2013.

facilitate this action it is important to identify stable and unstable regions in the phase space where the objects could be moved to exploit either long term "graveyards" or, possibly and preferentially, faster escape routes. To this purpose, a dynamical mapping of the circumterrestrial space was performed within ReDSHIFT, including the whole space from LEO up to the geostationary orbit.

The circumterrestrial space was divided in a fine grid in the orbital elements space and about 20 millions of initial conditions were propagated, with three different orbital propagators, having similar dynamical models expressly tailored to the explored orbital regime, but different underling computational engines. All the orbital propagations were repeated twice: once assuming for the test object an Area-to mass ratio equal to the standard satellites (0.012 m^2/kg) and then assuming an augmented area (1 m^2/kg, thus simulating the presence of some kind of sail). The output of these massive orbital propagations were stored in terms of maps displaying, e.g., the lifetime or the maximum eccentricity growth for all the orbits during the 120 years time span of the propagation.

Here a short summary of the results for the LEO region will be pre-

sented. For more information on the mapping of the higher orbits the reader can refer to [14] and [5].

In LEO the main purpose of the dynamical disposal studies is to identify natural perturbations able to increase the orbital eccentricity in order to bring the spacecraft to interact with the atmosphere at perigee, hence shortening its lifetime. As a first step, it is worth identifying the orbits from which an abandoned object can "naturally" reenter, within a given time span, without any additional maneuver. In *Fig. 3*, for typical intact objects having a ratio between area and mass $A/m = 0.012$ m^2/kg, the minimum value of eccentricity (e), as a function of the semi-major axis, which guarantees a lifetime of at most 25 years is shown (blue line in the Figure). The red line shows instead the minimum value of eccentricity, for any value of inclination in the range explored and for each given semi-major axis, which ensures to reenter within the total time span of our exploration (120 years). Finally, the cyan stars represent cases where perturbations, different from drag, facilitate the reentry at specific values of inclination (see later).

Building on the results shown in *Fig. 3*, the LEO space is explored highlighting a strong network of dynamical resonances responsible for the growth of the orbital eccentricity able to improve the picture shown in the blue and red lines of *Fig. 3*, further shortening the residual lifetime. At difference from the dynamics of the higher orbits in the medium Earth Orbit region (MEO, at altitudes around 15000 - 20000 km), in LEO the perturbing effects can be obtained mostly not from lunisolar resonances, but also from high-degree zonal harmonics and solar radiation pressure.

The complete mapping of the LEO space, over the selected grid of initial conditions, produced a large number of phase diagrams such as the ones shown in *Fig. 4*.

The two panels show, as a function of initial inclination and eccentricity, the maximum eccentricity reached (see color bar) over the 120-year time span by an object with initial semimajor axis $a = R_\oplus + 1560$ km, where $R_\oplus$ is the radius of the Earth (initial right ascension of the ascending node $\Omega = 0$ and argument of perigee $\omega = 0°$). The left panel refers to an object having $A/m = 0.024\,\mathrm{m}^2/\mathrm{kg}$ and the right panel to an object having $A/m = 1\,\mathrm{m}^2/\mathrm{kg}$. A few things can be noted from this example. First the enhanced effects due to the increase in the A/m of the object is apparent. Then, the bright "corridors" that are visible mainly in the right panel, clearly show how, at specific inclinations (e.g., at $i \sim 40°$ or $i \sim 80°$) the perturbing effects are particularly enhanced. These corri-

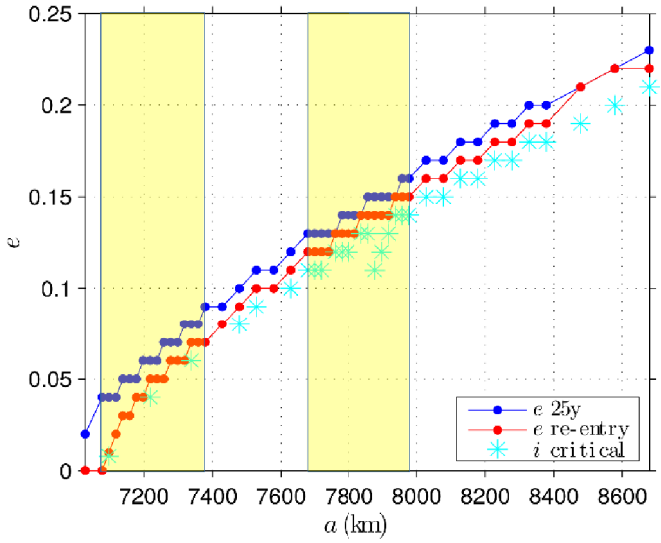


Fig. 3 – As a function of the semi-major axis, the minimum value of eccentricity required to reenter in 25 years (blue curve), and in less than 120 years (red curve), assuming $A/m = 0.012\ m^2/kg$, *and* $\Omega = 180°$ *and* $\omega = 0°$ *and the initial epoch 2020. The cyan stars represent cases where perturbations different from drag facilitate the reentry at specific values of inclination (see text for details). In yellow, the most populated altitudes are highlighted.*

dors correspond to resonant inclinations, associated with the cyan stars shown in *Fig. 3*. At these specific inclinations, the perturbation exerted by the geopotential or by the Moon or by the Sun (both gravitational or as solar radiation pressure) are able to alter the general behavior.

A specific analysis of these resonances was performed. All the details can be found in [2] and [1]. Here the main findings are briefly summarized. In Figure 5 and Figure 6 (top panel), we show, the maximum value, numerically computed, for the eccentricity in 120 years as a function of the initial inclination and semi-major axis, for prograde orbits starting from $e = 0.02$, $\Omega = 0°$, $\omega = 0°$, for spacecraft having $A/m = 0.012$ m^2/kg and $A/m = 1\ \mathrm{m}^2/\mathrm{kg}$, respectively.

In *Fig. 5* the maximum value, numerically computed, for the eccentricity in 120 years as a function of the initial inclination and semi-major

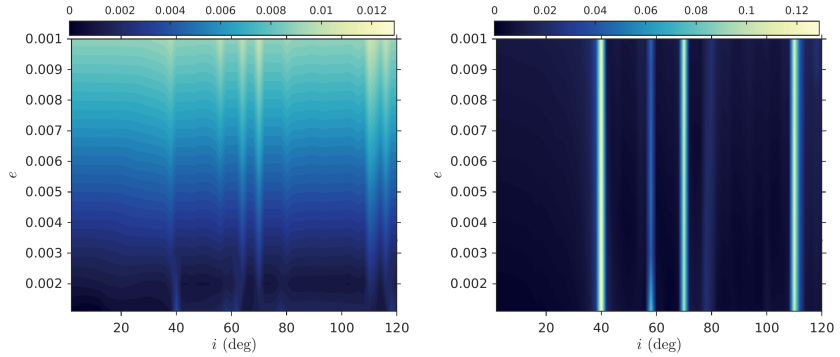


Fig. 4 – Maximum eccentricity computed over 120 years (color bar) for quasi-circular orbits as a function of initial inclination and eccentricity for $a = R_\oplus + 1560$ *km,* $\Omega = 90°$ *,* $\omega = 0°$ *at the initial epoch 2020. Left:* $C_R(A/m) = 0.012$ m^2*/kg; right:* $C_R(A/m) = 1$ m^2 */kg (*C_R *is the reflectivity coefficient varying between 1 and 2). Note the change of scale in the color bars between the left and the right panels.*

axis, for prograde orbits starting from $e = 0.02, \Omega = 0°, \omega = 0°$ for a spacecraft with A/m = 0.012 m^2/kg. Figure 6 shows the same quantities for a spacecraft with A/m = 1 m^2/kg. The yellow lines identify the locations of the following resonances:

- singly-averaged solar gravitational resonances (e.g., [7])
  - $\dot{\psi} = 2\dot{\omega} + \dot{\Omega} - 2n_S \approx 0$,
  - $\dot{\psi} = 2\dot{\omega} - n_S \approx 0$,
  - $\dot{\psi} = 2\dot{\omega} + 2\dot{\Omega} + n_S \approx 0$,
  - $\dot{\psi} = 2\dot{\omega} + 2\dot{\Omega} \approx 0$

  where ψ is the critical angle of the resonance, n_S is the apparent mean motion of the Sun and $\dot{\Omega}$ and $\dot{\omega}$ are the time derivatives of the right ascension of the ascending node and of the argument of perigee, respectively;
- $\dot{\psi} = 2\dot{\omega} + \dot{\Omega} \approx 0$, associated with doubly-averaged lunisolar gravitational perturbations (e.g., [7]);

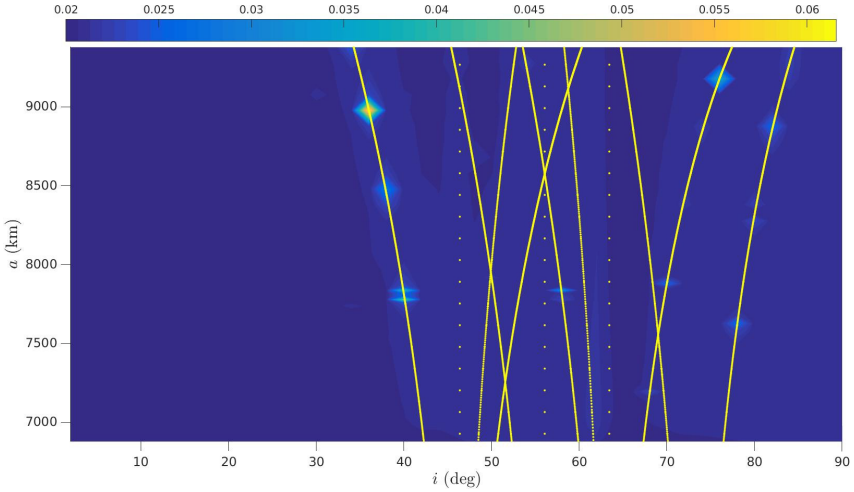


Fig. 5 – As a function of the initial inclination and semi-major axis, the maximum eccentricity value (color bar) computed in 120 years starting from $e = 0.02$, $\Omega = 0°$, $\omega = 0°$, *for the initial epoch 2018, considering* $C_R A/m = 0.12$ m^2/kg *and prograde orbits. The yellow lines denote the dominant resonances determining the long-term behavior.*

- $\dot{\psi} = \dot{\omega} \approx 0$, associated with lunisolar perturbations (e.g., [7]), but also with high-degree terms in the geopotential (e.g., [12]);
- $\dot{\psi} = \alpha\dot{\Omega} \pm \dot{\omega} \pm n_S \approx 0$, associated with SRP ($\alpha = 0, 1$), [4]-[6].

For $A/m = 1$ m^2/kg, the only resonances which matter are the six resonances associated with SRP, namely those in the last bullet, which can be rewritten, for the sake of clarity as,

$$\begin{aligned}
\dot{\psi}_1 &= \dot{\Omega} + \dot{\omega} - n_S \approx 0, \\
\dot{\psi}_2 &= \dot{\Omega} - \dot{\omega} - n_S \approx 0, \\
\dot{\psi}_3 &= \dot{\omega} - n_S \approx 0, \\
\dot{\psi}_4 &= \dot{\omega} + n_S \approx 0, \\
\dot{\psi}_5 &= \dot{\Omega} + \dot{\omega} + n_S \approx 0, \\
\dot{\psi}_6 &= \dot{\Omega} - \dot{\omega} + n_S \approx 0.
\end{aligned}$$

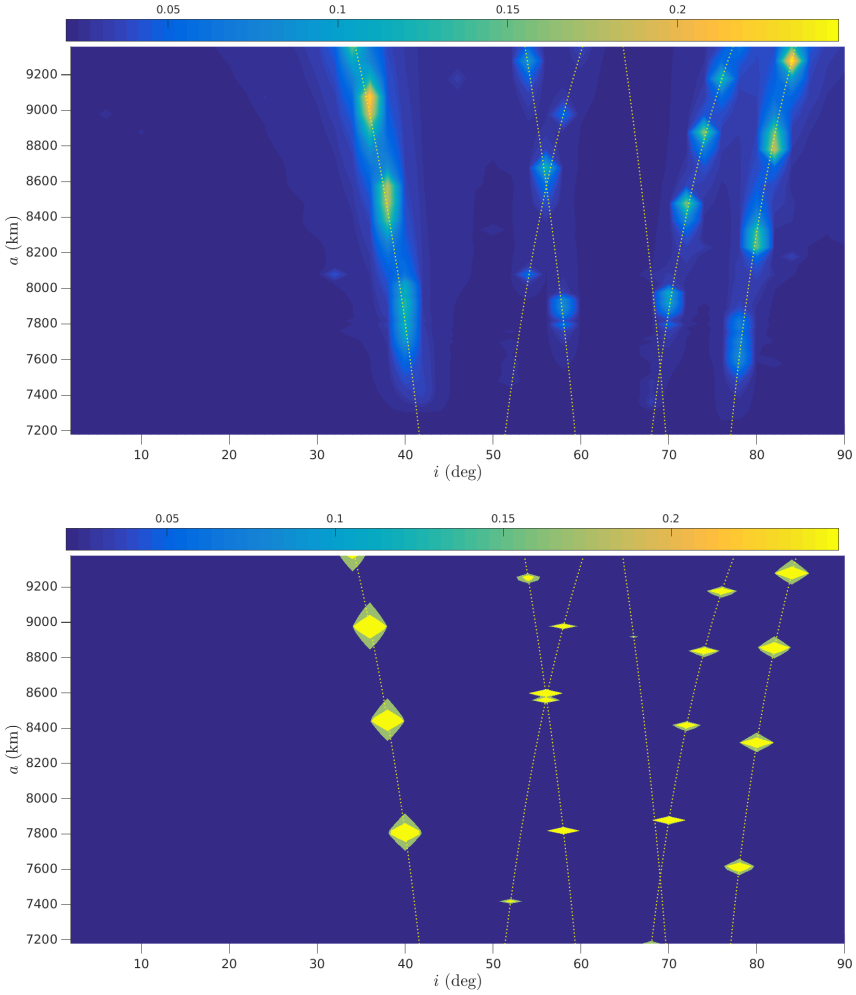


Fig. 6 – Top panel: as a function of the initial inclination and semi-major axis, the maximum eccentricity value (color bar) computed in 120 years starting from $e = 0.02$, $\Omega = 0°$, $\omega = 0°$. *for the initial epoch 2020, considering* $C_R A/m = 1\ m^2/kg$ *and prograde orbits. Bottom panel: analytical estimate of the maximum variation in eccentricity is shown. The yellow lines denote the dominant resonances determining the long-term behavior.*

From Figure 6 (top), it can be noted that, if an area augmentation device able to increase the A/M ratio is attached to the spacecraft, the variation in eccentricity that can be obtained at a resonance can definitely allow for a reentry. As a matter of fact, it turns out that the amount of eccentricity increase can be estimated analytically. The derivation is not presented here, but the reader can refer to [2] and [1]. The bottom panel of *Fig. 6* shows the location of the resonances computed analytically; the good correspondence with the numerical results of the top panel can be noted.

All the maps, similar to those of *Fig. 4*, for the whole LEO to GEO space, are now stored and will be assembled in an "atlas" which will be publicly available on the ReDSHIFT web site and will be used to identify, for every orbital regime the most convenient de-orbiting or re-orbiting options. Moreover the computation of the optimal maneuvers needed to steer, whenever possible, the spacecraft towards the resonances at the end-of-life are being computed.

5. Design, 3D printing and testing of novel spacecraft

As mentioned in Sec. 1, there is the need to tackle the space debris problem from different points of view. The design and building of the spacecraft has to be defined from the very beginning taking into account the harsh environment where the satellite will operate.

It is important to design satellites that can survive impacts against small particles (below $\sim$ 1 cm in size) in order to minimize the in orbit failure that, not only compromise the mission objectives, but also leave a stranded spacecraft in the operational orbits, preventing any disposal maneuvers. This can be accomplished by protecting the satellite, or specific vital section of it, with *ad-hoc* light-weight shields, with different materials. Whenever the shielding is not possible (i.e., for larger particles), it would be desirable to have a spacecraft which could possibly minimize the production of small debris after a fragmentation. This could be accomplished by designing the satellite with, e.g., pre-defined breakup lines bounded to fail in case of a large impact and less prone to transport the shock to the whole structure.

These complex engineering aspects are dealt with in ReDSHIFT. By using 3D printing, it is possible to test different designs and materials in a fast and relatively economical way. Multiple spacecraft parts can be printed, by changing some details, and then tested in specific facilities such as impact laboratories with light gas guns (to perform hypervelocity

impacts tests), vibration and radiation chambers, etc. Addictive manufacturing at component level will include also multifunctional structures testing novel solutions like embedded wiring (structronics), power generation, electronic shielding, heat dissipation,..... Moreover, as mentioned in the previous section, the possibility to add to a spacecraft, from the design phase, as de-orbiting device (such as a sail) can represent a fundamental tool to achieve a fast de-orbiting. Once again, 3D printing will allow the testing of novel interfaces between de-orbiting devices and the satellite body thus enhancing the applicability of these kind of solutions.

A parallel research activity is devoted to the so-called "design-for-demise" (D4D). While de-orbiting into the atmosphere is an ideal solution to prevent the proliferation of space debris, it is important to avoid the accidental fall of material on inhabited areas. Whenever a controlled re-entry of the spacecraft, targeting oceans or deserts, is not possible, it has to be checked that no spacecraft parts can survive the atmospheric heating and reach the ground. That is, the spacecraft has to be designed to demise into the atmosphere. This implies a number of preliminary theoretical studies to identify the re-entry trajectories that guaranty the highest heat load on the spacecraft structures that, in turns, have to be designed in such a way to facilitate the disruption during the re-entry. This can be a requirement conflicting with the above mentioned need for impact resistance and it is therefore a crucial design task to be defined at the very beginning of the spacecraft production. Every material used and every larger test part printed during the project will be therefore also tested for demise in an arc-heated hypersonic wind tunnel.

It is worth stressing again that, trying to keep the holistic approach as in the title of the project, all the design activities have to be performed in close collaboration with the theoretical findings described in the previous sections. E.g., the design and production have to take into account the debris environment expected during the mission lifetime, as derived from the long term simulations described in Sec. 3, to define the needed level of shielding and protection. Moreover, the de-orbiting options and trajectories, found for the specific orbital region, have to be considered in order to design the best de-orbiting devices able to steer the spacecraft towards the most favorable "de-orbiting highway" at the end-of-life. This interaction between the different aspects of the project will be engineered with the production of a software suite (made publicly available through the project web site) providing a complete debris mitigation analysis of a mission.

This phase of the project is now ongoing. The first detailed designs of the test spacecraft, shields and sail interface were produced. The first printed samples and tests shall start in the coming months.

6. Legal and political aspects of the space debris mitigation

As mentioned in Sec. 3, the potential benefits of any proposed mitigation measure is often weakened by its limited practical application. A clear understanding of the complex legal framework of the space activities is therefore essential to try and enforce any innovative mitigation procedure. The ReDSHIFT project will address also this aspect in an effort to maximize the effectiveness of its technical findings.

Currently, the enforcement of the existing mitigation rules, such as, e.g., the IADC (Inter Agency Space Debris Coordination Committee) and the UNOOSA ((Unites Nation Office for Outer Space) guidelines, is complicated by the fact that the guidelines are not legally binding (i.e., there are no legal consequences if they are not followed) but configure themselves as voluntary recommendations. That is, from a strictly legal perspective, it is the legal character of the existing mitigation rules that prevents their enforcement. i.e., the guidelines are explicitly dependent on voluntary compliance, hence no enforcement mechanism is foreseen. From an economic point of view, there are mostly cost-related and economic competition considerations which prevent from applying the guidelines as recommended. As the mitigation guidelines are, as mentioned above, not legally binding some other motive might be needed to stimulate their observance. "Steering" instruments comparable to the emissions trading in environmental policies could be the so-called "steering taxes", even if the application of this option is extremely delicate and of very difficult predictability with respect to their possible effects, as the space is a global market and a rather exclusive one due to the high costs involved. Other means could be state subsidies aiming at stimulating private actors to comply with mitigation standards. In the course of ReDSHIFT investigations in these direction will be performed.

Otherwise the guidelines have to be transformed to legally binding rules of international law. Mandatory legal character could be attributed to the guidelines also by virtue of national laws which foresee adherence to the standards contained in the guidelines. Individually some European countries have undertaken legal actions to deal with the space debris. Ex-

amples include the United Kingdom's Outer Space Act, German efforts at regulation, the French Space Act, the draft European Code of Conduct 2004, and the ESA Space Debris Mitigation Guidelines of 2008. In some countries, licenses can be issued only if operations are conducted in such a way to prevent the contamination of outer space with debris. All these national legislations are being analysed, together with the international guidelines, to identify benefits and flaws and to find common grounds between them trying to push Europe to act as a single entity to promote regulations on the global political forums

Since an object abandoned in space represents a risk for any operational satellite, another aspect of the problem is the liability in case a damaging collision would take place. According to the Liability Convention, a launching State remains the "owner" of a space object even after it has come to the end of its lifespan and has become an orbital debris. However, there is no concrete provision for a duty to remove debris in international law and the "legal" identification of the origin of small space debris (e.g., fragments) is practically unfeasible and the attribution of fault for liability is therefore actually impossible. The development of an insurance market for space projects would be a very important prerequisite for the practicability and applicability of space debris remediation and mitigation rules. Though technically and economically problematic innovative solutions could be hypothesized. E.g., the establishment of an insurance fund where all space actors pay a certain amount which is used for space environment remediation or for the consequences of space debris accidents.

In the analysis of an optimised legal and political framework able to better promote the application of improved guidelines it is important to explore how the new design and technologies and rules foster specific branches of the economy. In the environmental economy there are clear evidences that, contrary to common feelings and initial expectations, new rules are often fostering innovation urging manufacturers to move from the *status quo* towards new, often economically profitable technical solutions. The aim of the legal/political and economical analysis within ReDSHIFT is also to push the space manufacturers towards these innovative and possibly profitable directions.

7. Conclusions

To prevent the proliferation of the space debris in Earth orbit it is necessary to tackle the problem from different perspectives, exploiting the latest theoretical and experimental knowledge in different fields such as astrodynamics, spacecraft engineering, legal studies. The H2020 ReDSHIFT project was devised to answer to these needs. In particular in this note the main results reached up to now, especially in the dynamical mapping of the LEO region were described. A number of "corridors" related to specific resonances between gravitational and non-gravitational perturbation effects were identified. These resonances can be exploited to facilitate the de-orbiting of the spacecraft at the end-of-life by increasing the orbital eccentricity. The experimental part of the ReDSHIFT project has just started and the final results are expected by the end of the year 2018.

Acknowledgements. The research leading to these results has received funding from the Horizon 2020 European Union's Framework Programme for Research and Innovation (H2020-PROTEC-2015) under REA grant agreement n. [687500]- ReDSHIFT. The work presented here was performed by the ReDSHIFT team whose contribution is gratefully acknowledged. In particular E.M. Alessi and G. Schettino were responsible for most of the work described in Sec. 4

REFERENCES

[1] Alessi E.M., G. Schettino, A. Rossi, G.B. Valsecchi Solar Radiation Pressure Resonances in Low Earth Orbits, *MNRAS*, in press, 2017.

[2] Alessi E.M., G. Schettino, A. Rossi, G.B. Valsecchi Natural Highways for End-of-Life Solutions in the LEO Region, *CeMDA*, submitted, 2017.

[3] Bombardelli C., E.M. Alessi, A.Rossi, G.B. Valsecchi, Environmental effect of space debris repositioning, *Advances in Space Research*, 60, 28-37, 2017.

[4] Cook, G. E., Luni-solar perturbations of the orbit of an Earth satellite, *The Geophysical Journal of the Royal Astronomical Society*, 6, 271-291, 1962.

[5] Colombo C. and Gkolias I., Analysis of orbit stability in the Geosynchronous region for end-of-life disposal, *Proceedings of the 7th European Conference on Space Debris*, Darmstadt, Germany, 2017.

[6] Hughes, S., Satellite orbits perturbed by direct solar radiation pressure: general expan- sion of the disturbing function, *Planetary and Space Science*, 25, 809–815, 1977.

[7] Hughes, S., Earth Satellite Orbits with Resonant Lunisolar Perturbations. I. Resonances Dependent Only on Inclination, *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 372, 243–264, 1980.

[8] Lewis H., J. Radtke, A. Rossi *et al.*, Sensitivity of the Space Debris Environment to Large Constellations and Small Satellites,*JBIS*, 70, 105-107, 2017.

[9] Radtke J., S. Muller, V. Schaus, and E. Stoll, LUCA2 - an enhanced long-term utility for collision analy sis, in *Proceedings of the 7th European Conference on Space Debris*, Darmstadt, Germany, 2017.

[10] Rossi A., Cordelli A., Farinella P., Anselmo L., Collisional evolution of the Earth's orbital debris cloud, *Journal of Geophysical Research*, 99, 23195-23210, 1994.

[11] Rossi A., L. Anselmo, C. Pardini, R. Jehn, and G. B. Valsecchi, The new Space Debris Mitigation (SDM 4.0) long term evolution code, *in Proceedings of the Fifth European Conference on Space Debris*, Darmastadt, Germany, 2009.

[12] Roy, A. E., *Orbital Motion*, Institute of Physics Publishing, Bristol, 2005.

[13] Schaus V., J. Radtke, E. Stoll, A. Rossi, C. Colombo , S. Tonetti, I. Holbrough, Results of reference long-term simulations focussing on passive means to reduce the impact of space debris, in *Proceedings of the 7th European Conference on Space Debris*, Darmstadt, Germany, 2017.

[14] Skoulidou, A. J. Rosengren, K. Tsiganis, and G. Voyatzis, Cartographic study of the MEO phase space for passive debris removal, in *Proceedings of the 7th European Conference on Space Debris*, Darmstadt, Germany, 2017.

Scienze naturali
Biologia

Istituto Lombardo (Rend. Scienze) 151, 279-290 (2017)

PANDEMIE E VIRUS EMERGENTI: NUOVE SFIDE E ANTICHE PAURE

GIOVANNI MAGA (*)

Nota presentata dal s.c. Giuliano Gasperi
(Adunanza del 21 dicembre 2017)

SUNTO. – Le epidemie di AIDS, SARS, influenza aviaria, MERs-CoV and SARS-CoV2 sono esempi di infezioni virali che, a partire da piccoli focolai in regioni circoscritte, hanno fatto il giro del mondo. La notizia della scoperta di virus capaci di infettare l'uomo, provenienti da remote regioni del globo e dai nomi esotici come Nipah, Chikungunya, Hendra, non è più relegata ai bollettini medici specializzati, ma viene ripresa dalla stampa mondiale. Questo perché è ormai diffusa la consapevolezza che ogni ogni focolaio, per quanto piccolo e distante, può dare inizio a un incendio inarrestabile. Affrontare queste situazioni richiede la capacità di fare previsioni, di effettuare proiezioni attendibili di dati, di ipotizzare scenari possibili. In altre parole occorrono esperienza e conoscenze approfondite. Per combattere i virus, dunque, è di fondamentale importanza capire come si trasmettono, moltiplicano, evolvono, da soli o per effetto dell'impatto sull'ambiente delle attività umane. Sulla base di queste conoscenze, è possibile non solo ideare farmaci e vaccini per alleviare l'impatto delle infezioni virali sulla salute dell'uomo, ma anche manipolare geneticamente questi parassiti intracellulari per applicazioni di terapia genica.

***

ABSTRACT. – The epidemics of AIDS, SARS, avian flu, MERS-CoV and SARS-CoV2 are examples of viral infections which, starting from local outbreaks, spread in many different countries or even became pandemics. Reports of novel viruses emerging from remote areas of the world and able to infect the human population, are not anymore confined to medical bulletins, but are being found everyday in newspapers and public media. This, because it is now well understood that every outbreak, small and remote it may be, can rapidly evolve into a worldwide health emergency. Dealing with such situations requires

(*) Istituto di Genetica Molecolare IGM-CNR, Pavia, Italy.
E-mail: giovanni.maga@igm.cnr.it

high forecasting ability starting from the available data and deep knowledge of the many factors influencing the spread of a viral outbreak. In turn, such ability stems from the study of the life cycle and pathogenetic mechanisms of the viruses. Understanding the biology of viruses and the relationships between emerging viral pathogens and the anthropic influences on the environment is the key to develop new drugs and vaccines, but also to genetically manipulate viruses for gene therapy applications.

Introduzione

I virus sono antichi abitatori della Terra, forse fin dall'origine della vita, ma ancora non sappiamo rispondere del tutto a domande fondamentali quali: i virus sono vivi? Da dove vengono? Come ne nascono di nuovi?

In un mondo sempre più affollato, mobile e veloce, le opportunità che l'uomo incontri nuovi virus si moltiplicano. Epidemie come l'AIDS, l'influenza aviaria, Ebola o la SARS preoccupano gli esperti di salute pubblica: ma quali forze le hanno scatenate?

Nell'immaginario collettivo, i virus sono inestricabilmente legati a suggestioni negative: malattie, epidemie, bioterrorismo. Una visione giustificata ampiamente dal fatto che alcune delle malattie più diffuse e difficili da combattere a livello mondiale sono causate da virus e ulteriormente amplificata dal tono spesso terrorizzante che i mezzi di comunicazione riservano alla cronaca di episodi epidemici anche in paesi lontani.

A questa prospettiva antropocentrica, tuttavia, deve oggi sostituirsi una visione più globale, incentrata sul concetto di base che i virus, al pari di ogni altra forma biologica, sono parte integrante dell'ecosistema globale. I loro cicli vitali, quindi, sono strettamente connessi a quelli degli altri esseri viventi, in una rete di interdipendenza indissolubile. I virus sono parte fondamentale della vita sulla Terra: solo partendo da questo assunto fondamentale è possibile studiare e comprendere le complesse dinamiche che sottendono alla loro evoluzione e diffusione [1].

Caratteristiche biologiche dei virus e loro possibile origine

I virus sono parassiti endocellulari obbligati, devono cioè penetrare all'interno di una cellula per potersi riprodurre. I virus sono elementi acellulari: la loro struttura fondamentale è costituita da un acido nucleico (RNA o DNA) che contiene il patrimonio genetico, circondato da un

involucro proteico detto capside. Alcuni virus presentano poi un ulteriore rivestimento proteinaceo, detto matrice, sul quale prende contatto una membrana lipidica, derivata dalla membrana della cellula infetta, che prende il nome di tunica o "envelope". La particella virale così composta è detta virione. I capsidi virali hanno strutture altamente regolari, riconducibili a due simmetrie fondamentali: icosaedrica ed elicoidale. Questa caratteristica risponde a due esigenze: la necessità di compattare l'informazione contenuta nello spazio genetico ristretto caratteristico della maggior parte dei virus, che possiedono genomi di ridotte dimensioni, e la ottimizzazione del processo di assemblaggio dei capsidi, che avviene all'interno della cellula ospite. Le strutture di tipo icosaedrico od elicoidale, possono essere facilmente ottenute dall'assemblaggio di una o poche (due-tre) proteine in unità strutturali di ordine superiore che, avendo un'energia minima, possono anche autoassemblarsi senza bisogno di reazioni enzimatiche a supporto del processo [2].

Le dimensioni delle particelle virali vanno da 20 nm a circa 0.5 μm. Queste particelle, dette virioni, sono totalmente inerti nell'ambiente esterno, ma quando entrano in una cellula ne prendono il controllo, trasformandola in una fabbrica di virus. Durante questo processo, il patrimonio genetico dei virus (ovvero il loro RNA o DNA) può mescolarsi a quello della cellula ospite, oppure può inglobare geni cellulari, acquisendo così nuove proprietà e consentendo il trasferimento di geni da una cellula all'altra.

L'ingresso dei virus in una cellula è mediato dalle interazioni molecolari tra le proteine virali presenti alla superficie del virione e le proteine cellulari presenti sul lato esterno della membrana plasmatica [2]. Si distinguono due principali meccanismi di penetrazione delle particelle virali all'interno della cellula:

– Endocitosi, in cui il virione viene racchiuso in una vescicola formatasi dall'invaginazione della membrana plasmatica cellulare nel punto in cui il virione la contatta e che successivamente viene internalizzata nel citoplasma.
– Fusione, processo esclusivamente a carico dei virioni dotati di tunica, in cui la membrana lipidica esterna della particella virale si fonde alla membrana plasmatica cellulare, liberando il capside nel citoplasma.

Una volta all'interno della cellula, il virione va incontro a spoliazione, ovvero disgregazione, del capside e l'acido nucleico virale, così liberato, viene trascritto, tradotto in proteine e duplicato, al fine di

costituire nuovi virioni che poi usciranno dalla cellula. Il processo di fuoriuscita può essere mediato da gemmazione, in cui la particella virale acquisisce il rivestimento esterno lipidico da una porzione della membrana cellulare; da esocitosi, in cui il virione viene esportato all'esterno da una vescicola di trasporto; o per lisi cellulare che porta alla disgregazione della cellula stessa con liberazione di tutti i suoi componenti interni, incluse le nuove particelle virali.

Un elemento che caratterizza i virus è la loro capacità di infettare qualsiasi cellula vivente. Non esiste organismo vivente che non sia infetto da virus. Questa osservazione, da sola, suggerisce un'antichità estrema dei virus. Infatti, solo ipotizzando che questi parassiti siano comparsi all'epoca dei primi elementi cellulari batterici, oltre 3 miliardi e mezzo di anni fa, si può spiegare come abbiano potuto adattarsi a interagire con le cellule di organismi di ogni livello di complessità: dai semplici batteri, alle piante, agli animali. Attualmente esistono due ipotesi principali sull'origine dei virus [3]:

- *Ipotesi riduzionista*: i virus sono cellule degenerate che hanno perso la capacità di vita autonoma e sono divenute parassiti cellulari. Un meccanismo simile ha portato durante l'evoluzione alla generazione dei mitocondri a partire da antichi batteri endosimbionti.
- *Modello fuga dalla cellula:* i virus derivano da elementi genetici cellulari che hanno acquisito una certa autonomia. Tali elementi genetici mobili sono ancora presenti nella moderna biosfera, come ad esempio i plasmidi batterici e i trasposoni.

Indipendentemente dal preciso meccanismo evolutivo, i virus appaiono comunque elementi molto antichi della biosfera, apparsi prima della separazione delle due principali linee evolutive (procarioti e archea/eucarioti).

Abbondanza e distribuzione dei virus nella biosfera e loro importanza nell'evoluzione della vita

Un primo aspetto che riflette la pervasività dei virus nell'ambiente è la loro numerosità. Si stima che solo nelle acque degli oceani siano presenti 10^{30} virus e numeri simili caratterizzano tutti gli altri ambienti. Oggi sono note alla scienza solo lo 0.1% di tutte le specie virali esistenti. L'organismo umano ospita stabilmente migliaia di miliardi di virus. Un grammo di feci ne può contenere fino a 10^{11}. Si tratta di virus residenti

che non causano patologie, la maggior parte dei quali infetta i batteri dell'intestino umano. Il viroma umano, come viene definita la popolazione virale residente, è essenziale per il corretto sviluppo del sistema immunitario e per l'equilibrio della flora batterica intestinale [4].

Lo stretto connubio tra esseri viventi cellulari e virus è riflesso anche dal ruolo fondamentale che i virus hanno avuto nel plasmare la struttura genetica dei viventi, contribuendo così alla loro evoluzione. Tracce di questi trasferimenti genici si trovano nel genoma di tutti gli organismi, uomo compreso. Il genoma umano è infatti costituito per il 30% da sequenze derivanti da antichi virus, che durante l'evoluzione hanno mescolato i loro geni con quelle degli organismi della nostra linea evolutiva. Un esempio di quanto questo scambio di geni sia stato fondamentale per l'evoluzione è dato dalla proteina sincitina, essenziale per lo sviluppo della placenta. In origine era la proteina del rivestimento esterno di un virus, acquisito dai progenitori dei placentati circa 50 milioni di anni fa [5]. In generale esistono numerose testimonianze molecolari di un continuo flusso genico tra virus e organismi viventi, uomo incluso:

1) Basandosi sulle similitudini esistenti tra le proteine dell'apparato replicativo dei virus, dei batteri e degli eucarioti, è stato proposto che l'apparato per la conversione dell'RNA in DNA sia emerso nei virus (trascrittasi inversa) e poi "passato" alle cellule a RNA, avviando così il processo che ha stabilito il DNA come materiale genetico ereditabile comune a tutti gli esseri viventi [6].
2) E' stata avanzata l'ipotesi che il nucleo della cellula eucariote derivi da un virus dotato di tunica (simile al poxvirus), basandosi sull'osservazione che questi virus durante la trascrizione intracellulare formano degli pseudonuclei utilizzando le membrane interne della cellula infetta [7].
3) L'apparato replicativo e la RNA polimerasi dei mitocondri derivano da un proto-batteriofago (virus batterico) il cui DNA era integrato nel proto-procariote (simbionte intracellulare) che ha originato i mitocondri [8].

Potenziale epidemico e pandemico dei virus patogeni per l'uomo

Il fatto che i virus infettino tutte le specie viventi e quindi siano strettamente legati all'ecologia degli organismi che li ospitano, è in

grado di spiegare il fenomeno dell' "emergenza" di virus all'interno di popolazioni di nuovi ospiti. Ad esempio, tutti i virus che infettano l'uomo derivano da virus animali, adattatisi all'ospite umano. La definizione di "emergente" per un virus, quindi, è sempre relativa ad un preciso contesto geografico e temporale. In epoche diverse, cioè, ogni virus è stato "emergente". Il virus dell'epatite B è entrato nella popolazione umana dalle scimmie oltre 100.000 anni fa. Il virus del morbillo, invece, deriva da quello della peste bovina, passato all'uomo all'epoca dei primi allevamenti di bovini domestici nel neolitico, circa 10.000 anni fa. Per tutta la storia dell'uomo, quindi, si è avuta l'introduzione di nuovi virus. Tra i più recenti, entrati negli ultimi 100 anni, ricordiamo HIV, Ebola, il coronavirus della SARS, il virus dell'influenza aviaria H5N1 [1].

Il meccanismo è sempre lo stesso: l'uomo colonizza un ambiente modificandone l'ecologia e quindi entrando in contatto con nuove specie animali, portatrici di virus, che, grazie alla loro grande capacità di adattamento, possono diventare capaci di infettare l'uomo.

Un tipico esempio è il virus Ebola. Il suo serbatoio naturale sono pipistrelli equatoriali che si nutrono di frutta, a cui il virus è ben adattato e non causa nessuna patologia. Tramite il guano o la macellazione a scopo alimentare, il virus può occasionalmente passare all'uomo, causando un'infezione che può essere mortale in oltre l'80% dei casi. Tuttavia, proprio a causa della stretta relazione tra pipistrelli e virus, Ebola è confinato ai paesi Africani della fascia equatoriale. L'infezione dell'uomo è sempre accidentale e conseguente alle pratiche di caccia di animali selvatici, sia pipistrelli che primati, anche loro infettabili dai pipistrelli, e la conseguente macellazione con contatto con sangue infetto. Anche la diffusione del virus durante le epidemie è spesso conseguenza di comportamenti umani: ad esempio i riti funebri locali con esposizione pubblica dei morti per Ebola, altamente infettivi, che vengono lavati, toccati e baciati da decine di persone [9].

Molti virus possono essere trasmessi da vettori artropodi, come le zanzare. In questo caso, la diffusione di questi virus in nuovi areali è una diretta conseguenza dei cambiamenti nell'ecologia dei vettori. Un caso tipico è quello della zanzara tigre, *Aedes albopictus*. La sua diffusione dall'Asia alle zone temperate dell'area mediterranea ha aperto la possibilità a molti virus esotici, quali Zika, Chikungunya, Dengue, di causare episodi epidemici anche nei nostri paesi [10].

L'introduzione di nuovi virus in una popolazione suscettibile, ovvero che non ne ha mai fatto esperienza immunologica, può avere

effetti devastanti. Nel 1875, la prima epidemia di morbillo nelle isole Fiji, dove il virus era sconosciuto, ha causato oltre 50.000 vittime, un terzo della popolazione. Il virus era stato importato dal re e dai suoi dignitari che si erano recati in Australia in visita diplomatica e avevano contratto la malattia.

Anche le pandemie influenzali possono avere effetti importanti. Ricordiamo l'influenza Spagnola del 1918-19 che causò oltre 70 milioni di morti. I virus influenzali hanno come serbatoio naturale gli uccelli migratori anseriformi (anatre, cigni e congeneri). Da questi animali il virus è passato a tutti i mammiferi. Quando un virus aviario e uno umano coinfettano uno stesso ospite possono rimescolare i loro geni, generando varianti virali totalmente nuove. L'ospite classico è il suino, uno dei pochi organismi in grado di venire infettato contemporaneamente da virus aviari, suini e umani. I virus responsabili delle cinque pandemie influenzali degli ultimi 100 anni si sono tutti generati per riassortimento genico nei suini di allevamento, da cui sono passati all'uomo [11].

In generale, quando un nuovo virus entra in una popolazione, si possono stabilire diversi equilibri, in relazione al grado di suscettibilità immunologica degli individui. Si distinguono quattro tipi principali di interazione virus-ospite:

1) *Interazione stabile*: il virus si mantiene nella popolazione ospite per periodi indefiniti, in presenza di condizioni dell'ecosistema stabili.
2) *Interazione dinamica*: i cambiamenti nell'ecosistema o mutazioni geniche possono favorire l'introduzione di un virus in una nuova popolazione ospite (salto di specie).
3) *Interazione a vicolo cieco:* il virus occasionalmente infetta un ospite diverso da quello naturale, ma non avvia un nuovo ciclo stabile.
4) *Interazione non permissiva*: il virus non riesce ad infettare una data popolazione (pertanto definita non suscettibile)

L'attuale pandemia causata dal virus HIV-1: l'AIDS

Sebbene i mezzi di comunicazione spesso enfatizzino eventi epidemici di virus "esotici" come Ebola o Zika, in realtà il principale killer mondiale è il virus HIV-1, responsabile della pandemia di AIDS, la sindrome da immunodeficienza acquisita [12]. Oltre 35 milioni di persone nel mondo hanno contratto il virus, che causa oltre un milione di morti ogni anno e altrettante nuove infezioni.

Il virus HIV-1 appartiene al gruppo dei retrovirus, così indicati in quanto sono in grado di convertire il loro genoma a RNA in una molecola di DNA a doppio filamento, che viene poi fisicamente integrata nei cromosomi della cellula infetta. In questo modo i geni virali diventano parte del patrimonio genetico della cellula. In questa condizione, la cellula infetta è sempre potenzialmente in grado di produrre nuovi virus, ma può anche permanere in questa condizione, definita come infezione latente, per lungo tempo senza provocare alcuna reazione da parte del sistema immunitario. La maggior parte delle cellule infette iniziano subito a trascrivere e tradurre i geni virali, rilasciando nuovi virioni che alimentano così l'infezione. La conseguenza è la distruzione della cellula da parte del sistema immunitario o la sua morte a causa dell'effetto citopatico della replicazione virale. Una quota di cellule infette, tuttavia, permane in uno stato di infezione latente, costituendo un serbatoio permanente da cui il virus può emergere in qualsiasi momento. Questo è il motivo per cui non è possibile, ad oggi, eradicare il virus dall'organismo infetto.

Il bersaglio preferenziale del virus HIV-1 sono i linfociti T helper che presentano alla loro superficie il recettore CD4. Queste cellule hanno la funzione fondamentale di attivare le difese immunitarie che contrastano le infezioni. L'infezione da virus HIV-1 decorre in maniera del tutto asintomatica anche per 10-15 anni, ma la deplezione ad opera del virus di queste cellule T CD4 causa nel tempo una grave forma di deficienza immunitaria, l'AIDS, che causa l'instaurarsi di infezioni da parte di patogeni opportunisti. L'AIDS è una malattia mortale.

Oggi grazie alla terapia farmacologica possiamo evitare che un paziente sviluppi la fase di AIDS conclamato per cui, pur rimanendo sempre infetto, entra in una fase di cronicizzazione della malattia che ne garantisce una sopravvivenza ormai paragonabile a quella della popolazione non infetta. Tuttavia la terapia va assunta per tutta la vita e va iniziata idealmente prima di sviluppare l'immunodeficienza. Quest'ultimo punto è cruciale, soprattutto considerando che in molti casi l'infezione viene diagnosticata alla comparsa dei sintomi, quindi anni dopo l'infezione. In Italia ogni anno si infettano 4000 persone, ma questi casi per il 70% si riferiscono a infezioni diagnosticate in fase avanzata. Si stima che oltre 15.000 persone vivano nel nostro paese senza sapere di avere il virus HIV-1, contribuendo a diffonderlo. Il virus nei nostri paesi si trasmette in oltre l'80% dei casi tramite rapporti sessuali non protetti. Quindi solo l'utilizzo del preservativo è in grado di limitarne la diffu-

sione e solo la consapevolezza del rischio è in grado di guidare il paziente ad effettuare il test sierologico per verificare se è infetto. Attualmente è disponibile un test per autodiagnosi acquistabile in farmacia e in grado di fornire una risposta in 20 minuti. Se una persona ha contratto il virus, deve rivolgersi al medico in modo che, al momento opportuno, possa iniziare la terapia, bloccando così il progredire della malattia.

Prevenzione e terapia delle infezioni virali: i vaccini

L'arma più efficace per contrastare il diffondersi delle epidemie virali sono i vaccini [13].

Quando un agente patogeno, virus, batterio o protozoo, invade le cellule del nostro organismo, si attiva una complessa serie di eventi molecolari che portano all'attivazione del sistema immunitario. Cellule specializzate, come ad esempio le cellule dendritiche, sono in grado di fagocitare l'agente patogeno e processare le sue proteine generando corti peptidi che vengono poi esposti alla superficie complessati a speciali recettori e agiscono da antigeni. Queste cellule presentanti l'antigene o APC, sono in grado di spostarsi dal sito dell'infezione per andare ad incontrare altri elementi del sistema immunitario, in particolare i linfociti T helper. Questi sono dotati di speciali recettori alla loro superficie, in grado di riconoscere uno degli antigeni presenti sulle APC, formando così un complesso cellula-cellula. Come conseguenza, vengono rilasciate molecole di segnale come chemochine e citochine, che attivano altre cellule specializzate tra cui i linfociti T citotossici e i linfociti B. I primi sono in grado di aggredire le cellule infette che espongono alla loro superficie l'antigene corrispondente e distruggerle, mentre i linfociti B producono anticorpi in grado di legarsi specificamente all'agente patogeno che possiede l'antigene che le ha stimolate, neutralizzandolo. Una parte dei linfociti B che producono questi anticorpi specifici permane nell'organismo sotto forma di cellule di memoria. In occasione di un successivo incontro con lo stesso agente infettivo che le ha stimolate, queste cellule sono rapidamente in grado di produrre anticorpi contro di esso, bloccandone la diffusione. Al primo incontro con il patogeno la risposta immunitaria diviene pienamente efficace dopo due o tre settimane, mentre la seconda volta le cellule di memoria si attivano nel giro di pochi giorni. Il principio su cui si basa la vaccinazione è proprio quello di stimolare la memoria immunologica nei confronti di un agente

patogeno, senza però provocare la malattia. Esponendo l'organismo ad un virus o batterio attenuato o inattivato, così come agli antigeni purificati, si innesca la stessa cascata di segnali che verrebbe attivata in caso di infezione, portando così alla formazione di una popolazione di cellule di memoria. In questo modo, quando la persona vaccinata viene a contatto con il vero agente infettivo, il suo sistema immunitario è già "preavvertito" e quindi in grado di neutralizzare rapidamente l'aggressore, prima che questo abbia modo di diffondersi nell'organismo provocando la malattia.

Una cosa importante da ricordare è che la vaccinazione non protegge solo l'individuo che la fa, ma rappresenta uno strumento fondamentale per limitare la diffusione di malattie infettive all'interno della popolazione. E' il concetto della cosiddetta immunità di gruppo. Ogni agente infettivo, ad esempio un virus, è caratterizzato da uno specifico parametro detto numero di riproduzione o R_0. Questo numero riflette la capacità infettiva di un agente trasmissibile e, nella sua essenza, indica il numero di individui che possono contrarre la malattia a partire da una singola persona infetta. L'influenza ha un R_0 pari a 3-4, mentre il morbillo pari a 18. Questo vuol dire che, mediamente, una persona affetta da influenza può infettare 3-4 persone, mentre chi ha il morbillo può passare la malattia a circa 18 persone. O, in altre parole, che il morbillo è molto più infettivo dell'influenza. Questi valori si riferiscono ad una popolazione "ideale" in cui tutti gli individui siano ugualmente suscettibili al virus. Ma se adesso immaginiamo che una percentuale significativa delle persone sia immune all'infezione, allora il singolo paziente avrà una probabilità inferiore di entrare in contatto con un individuo che possa contrarre la malattia. Lo scopo delle vaccinazioni di massa è proprio questo: aumentare il grado di immunità generale all'interno della popolazione, così che, anche nell'evento che qualcuno si ammali, l'agente patogeno non possa diffondersi causando un'epidemia, trovandosi di fronte un "muro" di persone immuni. In generale, una quota di persone immuni pari o superiore al 95% è in grado di evitare il diffondersi di qualunque malattia infettiva. L'immunità di gruppo è molto importante per garantire la protezione delle persone che per motivi di età o salute non possono vaccinarsi.

Grazie alla vaccinazione mondiale, il vaiolo, che nei secoli si stima abbia causato oltre mezzo miliardo di vittime, è stato completamente eradicato in tutto il mondo. Ma ovunque si siano attuate vaccinazioni di massa, le malattie causate dai virus contro cui viene indotta l'immunità

si sono ridotte del 90-99%, o addirittura in alcuni casi sono totalmente scomparse, come la poliomielite che persiste ancora in solo poche nazioni del mondo.

Conclusioni

Studiare l'evoluzione e le dinamiche delle popolazioni virali riveste un interesse non solo per la salute pubblica, ma anche per la comprensione delle complesse relazioni tra i viventi, così come è in grado di offrire una chiave per lo studio di molti fenomeni biologici. Oggi sono note alla scienza meno di un millesimo delle specie virali che si stima esistano nella biosfera. Le nuove tecnologie biomolecolari, come la metagenomica, la proteomica e la biologia dei sistemi, offrono sempre migliori strumenti per caratterizzare nuove specie virali, predirne il potenziale patogenetico e identificare all'interno del ciclo vitale dei virus patogeni bersagli per approcci chemio- e immuno-terapici.

BIBLIOGRAFIA

1. Maga G., Occhio ai Virus (2012). Chiavi di Lettura, Zanichelli Ed. (Bologna).
2. Dimmock N.J., Easton A.J., Leppard K.N., Introduzione alla Virologia Moderna (2017). Casa Editrice Ambrosiana (Milano).
3. Harper D.R., Virus (2013). Zanichelli Ed. (Bologna).
4. Zarate S., Taboada B., Yocupicio-Monroy, M., Arias C.F., Human Virome (2017). Arch. Med. Res. 48:701-706.
5. Mi S., Lee X., Li X., Veldman G.M., Finnerty H., Racie L., La Vallie E., Tang X.Y., Edouard P., Howes S., Keith J.C., McCoy J.M., Syncytin is a captive retroviral envelope protein involved in human placental morphogenesis (2000). Nature 403: 785-789.
6. Cojocaru R., Unrau P.J., Transitioning to DNA genomes in an RNA world (2017). Elife 6: e32330.
7. Martin W.F., Garg S., Zimorski V., Endosymbiotic theories for eukaryote origin (2015). Philos. Trans. R. Soc. Lond. B Biol. Sci. 370: 20140330.
8. Shutt T.E., Gray M.W., Bacteriophage origins of mitochondrial replication and transcription proteins (2006). Trends Genet. 22: 90-95.
9. Coltart C.E., Lindsey B., Ghinai I., Johnson A.M., Heymann D.L., The Ebola outbreak 2013-2016: old lessons for new epidemics (2017). Philos. Trans. R. Soc. Lond. B Biol. Sci. 372: 20160297.
10. Huntigton M.K.m Allison J., Nair D., Emerging vector-borne diseases (2016). Am. Fam. Physician 94: 551-557.

11. Bavagnoli L., Maga G., The 2009 influenza pandemic: promising lessons for antiviral therapy for future outbreaks (2011). Curr. Med. Chem. 18: 5466-5475.
12. Maga G, AIDS: la Verità Negata (2015). Il Pensiero Scientifico Ed. (Roma).
13. Rappuoli R., Vozza L., I vaccini dell'era globale (2009). Chiavi di Lettura, Zanichelli Ed. (Bologna).

INDICE DEL VOLUME

DEGIORGIO V., Storia della luce: dal sole al laser 3

COLOMBO L., Trasporto di calore attraverso un'interfaccia: un problema classico rivisitato in chiave moderna 19

KATER M., New Plant Breeding Technologies for a sustainable agriculture 31

BRENNA E., Tecniche per mascherare il sapore amaro dei prodotti farmaceutici 37

OSSI P.M., *et al.*, Un metodo spettroscopico per la quantificazione di farmaci in fluidi biologici 49

PERINOTTI P., Quantum field theory from first principles 77

BAVESTRELLO G., Competizione e cooperazione nelle storie vitali degli organismi marini 109

LASCIALFARI A., RIGAMONTI A., The nature of superconducting fluctuations and new aspects in high T_c superconductors 121

GASPERI G., Strategie innovative per il controllo di zanzare, vettori di organismi patogeni 133

ZIGLIOLI B., Broni e Casale Monferrato: due comunità di fronte all'amianto tra memoria, rimozione e riconoscimento del danno 149

ALBANI D., GIORDANO C., Current perspectives on the microbiota-gut-brain axis in chronic neurodegenerative disorders 171

SELLI E., La conversione di energia solare in energia chimica 183

BORGHETTI V., Margherita d'Austria musica e identità di una principessa del Rinascimento 201

ABBOTTO A., Dalle esigenze energetiche sulla Terra all'esplorazione di Marte passando per i colori della Cappella Sistina: presente e futuro del fotovoltaico e dei combustibili solari 219

TUMINO F., Esplorando "Flatlandia": dal grafene alla scoperta di nuovi materiali bi-dimensionali 241

BOLOGNESI M., La rivoluzione della crio-microscopia elettronica nello studio strutturale di proteine e acidi nucleici 255

ROSSI A., A holistic approach to the space debris mitigation 259

MAGA G., Pandemie e virus emergenti: nuove sfide e antiche paure 279

INDICE PER MATERIA

Chimica

Abbotto A., Dalle esigenze energetiche sulla Terra all'esplorazione di Marte passando per i colori della Cappella Sistina: presente e futuro del fotovoltaico e dei combustibili solari 219

Albani D., Giordano C., Current perspectives on the microbiota-gut-brain axis in chronic neurodegenerative disorders 171

Brenna E., Tecniche per mascherare il sapore amaro dei prodotti farmaceutici 37

Selli E., La conversione di energia solare in energia chimica 183

Fisica

Colombo L., Trasporto di calore attraverso un'interfaccia: un problema classico rivisitato in chiave moderna 19

Degiorgio V., Storia della luce: dal sole al laser 3

Lascialfari A., Rigamonti A., The nature of superconducting fluctuations and new aspects in high T_c superconductors 121

Perinotti P., Quantum field theory from first principles 77

Rossi A., A holistic approach to the space debris mitigation 259

Ingegneria

Ossi P.M., *et al.*, Un metodo spettroscopico per la quantificazione di farmaci in fluidi biologici 49

Tumino F., Esplorando "Flatlandia": dal grafene alla scoperta di nuovi materiali bi-dimensionali 241

Scienze naturali

Biologia

Bolognesi M., La rivoluzione della crio-microscopia elettronica nello studio strutturale di proteine e acidi nucleici 255

Maga G., Pandemie e virus emergenti: nuove sfide e antiche paure 279

Botanica

Kater M., New Plant Breeding Technologies for a sustainable agriculture 31

Ecologia

Ziglioli B., Broni e Casale Monferrato: due comunità di fronte all'amianto tra memoria, rimozione e riconoscimento del danno 149

Zoologia

Bavestrello G., Competizione e cooperazione nelle storie vitali degli organismi marini 109

Gasperi G., Strategie innovative per il controllo di zanzare, vettori di organismi patogeni 133

Altro

Borghetti V., Margherita d'Austria musica e identità di una principessa del Rinascimento 201

Finito di stampare nel mese di maggio 2020

Decreto Pres. Tribunale 10 gennaio 1961, n. 5504 registro cancelleria
Direttore responsabile: Adele Robbiati Bianchi