

La biologia strutturale in movimento

di Matteo Allegretti

Editor: Federico Forneris

Revisori Esperti: Nicola Portolano, Emanuele Conte, Giuseppe Ciossani

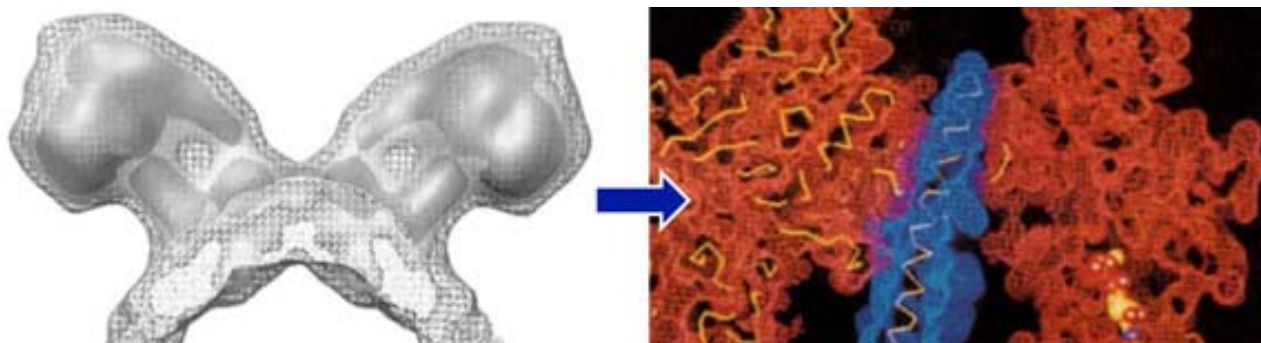
Revisori Naive: Annalisa Zuccotti, Andrea Mattia Marcelli



Parole Chiave: Acidi Nucleici, Biologia, Biologia Strutturale, Cristallografia, DNA, Drug Design, Macromolecole, Microscopia, Microscopia Elettronica, Modelli Matematici, NMR, Proteine, Ricerca di Base, Ricerca in Vitro

Permalink: <http://informa.airicerca.org/2014/10/04/biologia-strutturale-in-movimento/>

doi: 10.13140/RG.2.2.30623.87205



Pensate per un attimo di non conoscere come è fatto un motore, di non sapere come sono organizzati i componenti al suo interno, ma allo stesso tempo di avere la necessità di modificarne alcune sue parti per renderlo più efficiente, o per ripararlo. La biologia strutturale comprende quell'insieme di tecniche che ci consentono di "osservare le molecole biologiche", per comprenderne meglio il funzionamento ed eventualmente modificarle in caso cessino di funzionare, come ad esempio nel caso di molte malattie. Purtroppo non è così semplice indagare i dettagli delle molecole di cui siamo composti: non esistono strumenti ottici convenzionali che ci consentano di osservare direttamente oggetti piccoli come le molecole essenziali alla vita, e dobbiamo fare ricorso a strategie che richiedono applicazioni matematiche e tecnologie non convenzionali, come la microscopia elettronica o la cristallografia a raggi X. In questo momento stiamo attraversando una fase davvero entusiasmante per quanto riguarda lo sviluppo di nuove tecniche di biologia strutturale. Queste metodiche porteranno nel prossimo futuro ad una vera e propria "rivoluzione" nell'osservazione delle molecole. Questo articolo parla dell'evoluzione delle metodologie di biologia strutturale e delle moderne tecniche di investigazione delle macromolecole biologiche.

La biologia strutturale è quel ramo della biologia molecolare che investiga la struttura di molecole come proteine o acidi nucleici (il DNA ad esempio), cercando di comprenderne l'architettura degli atomi nello spazio tridimensionale. Ciò è di fondamentale importanza per descrivere la materia vivente a livello microscopico e per comprenderne il meccanismo di funzionamento. La struttura di una molecola è lo specchio delle sue caratteristiche funzionali. Questo concetto è di grande importanza in biologia. Macromolecole come le proteine eseguono gran parte delle funzioni cellulari proprio grazie a "come sono fatte strutturalmente", ossia la loro specifica forma tridimensionale generata dall'avvolgimento nello spazio delle sequenze di aminoacidi che le compongono. Questa forma è definita struttura terziaria (*Box Struttura delle Proteine*), e dipende dalla specifica sequenza aminoacidica di cui la proteina è composta. I venti aminoacidi che costituiscono tutte le proteine possono organizzarsi in un numero letteralmente infinito di strutture tridimensionali (basti pensare che le possibili combinazioni di aminoacidi in una proteina di 70 aminoacidi sono 20^{70}); tuttavia, solo alcune di queste strutture sono state selezionate dalla natura (si pensa speculativamente che circa 10^{14} siano presenti in natura, un numero comunque considerevole). E' altresì interessante notare come nonostante vi siano centinaia di migliaia di proteine conosciute e caratterizzate finora, vi siano soltanto 1300 specifici motivi strutturali (detti domini) in cui catalogare tutte queste strutture. Il compito dei biologi strutturali è quello di determinare l'organizzazione tridimensionale delle proteine, cercando di descrivere quanti più possibili arrangiamenti dei diversi domini si siano evoluti per svolgere una particolare funzione biologica.

Ottenere informazioni strutturali di macromolecole biologiche quali le proteine o gli acidi nucleici con dettagli molecolari non è un compito facile: la fisica ci insegna che anche il più potente microscopio ottico non ci permetterà mai di visualizzare singole macromolecole, men che meno la distribuzione degli atomi che le compongono. Per questo, i biologi strutturali fanno ricorso a varie tecniche di indagine capaci di superare i limiti convenzionali di risoluzione della microscopia ottica. Tre tecniche classicamente si ascrivono alla biologia strutturale: la cristallografia a raggi X, la risonanza magnetica nucleare (RMN o NMR) e la microscopia elettronica (EM) (*Box Tecniche di Biologia Strutturale*). Queste tecniche presentano tutte delle grandissime potenzialità, ma anche delle limitazioni (legate in particolar modo alla complessità e alla dimensione delle molecole che vengono analizzate) che non permettono di selezionare una sola tecnica di indagine strutturale come "l'approccio definitivo". In realtà, queste metodologie possono essere viste attualmente come complementari nel fornire

informazioni sull'organizzazione tridimensionale di macromolecole biologiche di varia entità e dimensione. Per fare due esempi dell'influenza storica di queste tecniche, possiamo citare la determinazione della struttura elicoidale del DNA tramite cristallografia a raggi X ad opera di Rosalind Franklin (1951) e la caratterizzazione a risoluzione nanometrica di alcuni organelli cellulari tramite l'utilizzo di sezioni di cellule analizzate al microscopio elettronico (1945).

Struttura delle Proteine

La struttura primaria di una proteina è l'elenco, in forma sequenziale, degli aminoacidi che la compongono. Questa sequenza è codificata dalla sequenza di DNA attraverso un codice, il codice genetico, che associa in modo specifico ad ogni tripletta di nucleotidi del DNA uno dei 20 aminoacidi. Il numero degli aminoacidi in una proteina può variare da due a migliaia. La struttura secondaria è l'arrangiamento degli aminoacidi in tre tipi di strutture regolari: α -eliche (rappresentate come nastri avvolti a spirale o cilindri), filamenti- β (rappresentati come nastri terminanti con una freccia ad indicare la direzione) e loop di connessione. Le differenti strutture secondarie assumono nello spazio una certa conformazione che determina la forma tridimensionale (struttura terziaria o fold) e la funzione della proteina. Si parla di struttura quaternaria quando più proteine insieme si arrangiano a formare un complesso proteico.

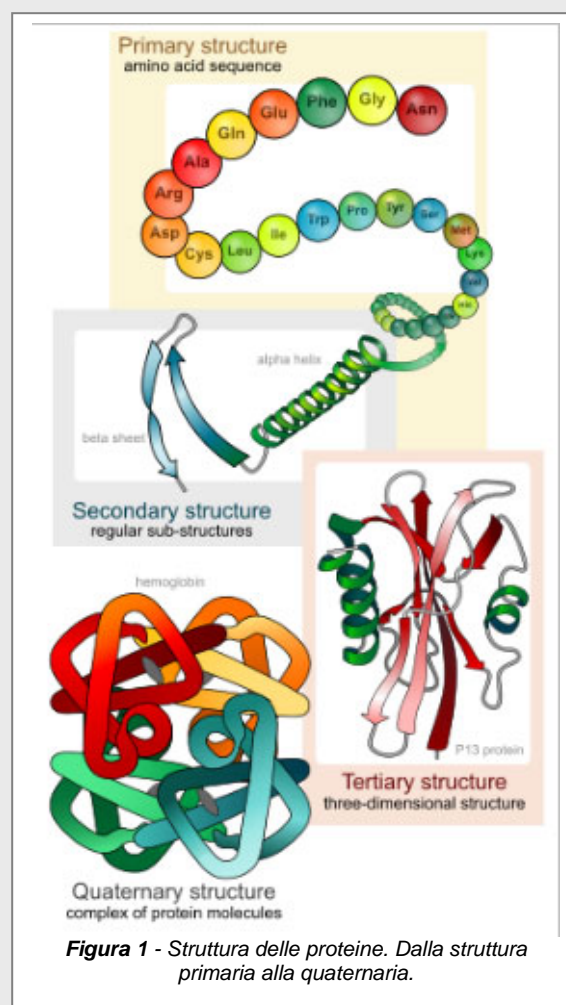


Figura 1 - Struttura delle proteine. Dalla struttura primaria alla quaternaria.

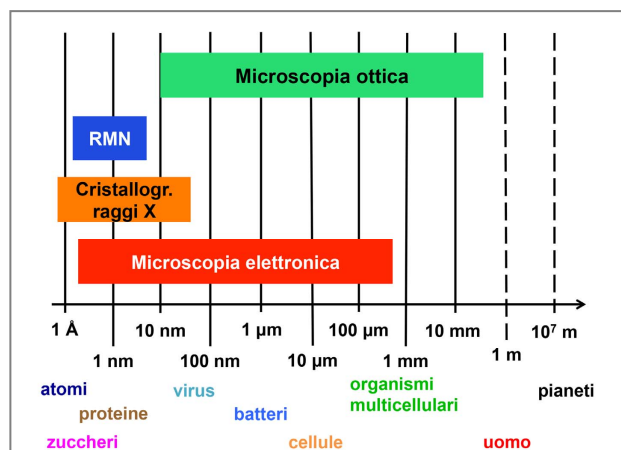


Figura 2 - La microscopia ottica permette di visualizzare oggetti da ~10 nm fino a decine di mm in tempo reale con un ampio campo visivo. La microscopia elettronica ha storicamente contribuito alla visualizzazione di campioni biologici fino a mm di dimensione, ma a bassa risoluzione (nanometrica). Recentemente sono state raggiunte risoluzioni atomiche nell'indagine di singole proteine (~3 angstrom, ossia tre decimi di nanometro), mostrando pertanto un grande potenziale di questa tecnica per indagini strutturali ad ampio spettro di risoluzione, da gruppi di cellule grandi centinaia di micron, fino a singole macromolecole (purché non troppo piccole). La cristallografia a raggi X raggiunge le più alte risoluzioni per i campioni biologici, essa può caratterizzare proteine, acidi nucleici, o loro complessi elucidandone i dettagli a livello di distanze tra gli atomi che li compongono. Tuttavia risente di importanti limitazioni dal punto di vista procedurale che non permettono una caratterizzazione immediata delle macromolecole oggetto di studio. La risonanza magnetica nucleare (NMR) può fornire dettagli strutturali ad alta risoluzione (simile alla cristallografia a raggi X), ma si tratta di una tecnica limitata esclusivamente a proteine o acidi nucleici piccoli (minori di 10 nm di grandezza).

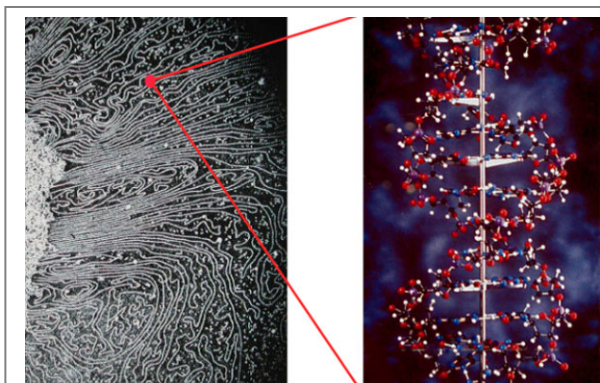


Figura 3 - DNA cromosomiale al microscopio elettronico: nell'ingrandimento a destra il modello a bastoncini e palline (ball-and-stick) della doppia elica risolto dalla cristallografia a raggi X.

Recentemente, la microscopia elettronica di macromolecole vetrificate in condizioni vicino alle fisiologiche (*Box Tecniche di Biologia Strutturale*) ha raggiunto risoluzioni molto vicine a quelle della cristallografia e della risonanza magnetica nucleare, con la sola limitazione di essere efficace esclusivamente per proteine di grandi dimensioni,

ossia con masse superiori ai 200 kDa. Tutto questo è stato ottenuto grazie a rivelatori di ultima generazione, cosicché il numero di strutture che sono investigate e risolte dalla microscopia elettronica sta aumentando esponenzialmente negli ultimi anni.

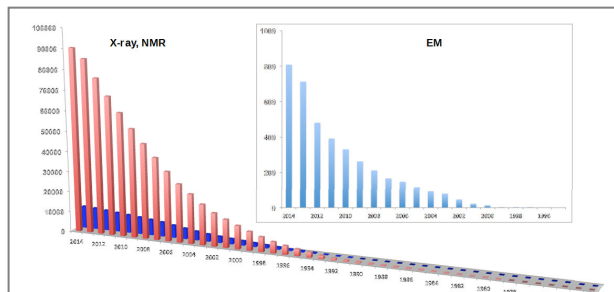


Figura 4 - Crescita nel numero di macromolecole (proteine o acidi nucleici) risolte dalle tre tecniche classiche negli ultimi 40 anni (dati provenienti dall'archivio mondiale delle strutture tridimensionali di proteine, protein data bank o PDB). Nel grafico 3D in rosso sono le strutture risolte dalla cristallografia (X-ray), chiaramente tecnica dominante con oltre 100000 strutture risolte ad oggi (asse y, i dati riportati si riferiscono ad inizio 2014). In blu la risonanza magnetica nucleare (NMR) con circa 10500 strutture in totale. In azzurro nel riquadro gli istogrammi della microscopia elettronica (EM), con sole 809 strutture depositate (aggiornato a luglio 2014), ma tendenzialmente in grande crescita.

Tecniche di Biologia Strutturale

La cristallografia a raggi X è una tecnica sperimentale basata sul fatto che i raggi X (lunghezza d'onda fino ad 1/100 Å) sono diffusi dai cristalli di proteine permettendo la determinazione della struttura tridimensionale (ossia, determinare la posizione nello spazio di ciascun atomo della macromolecola) a risoluzione atomica. La cristallografia è la tecnica più diffusa in biologia strutturale ed è anche quella che fornisce attualmente i dati a più alta risoluzione (addirittura sub-angstrom in alcuni casi), i principali svantaggi sono la spesso lunga attività di laboratorio necessaria per ottenere cristalli di proteine adatti agli esperimenti di cristallografia, ed il fatto che i cristalli possono non rispecchiare la struttura naturale della proteina in ambiente fisiologico.



Figura 5 - Watson, Crick e la struttura della molecola di DNA determinata tramite cristallografia a raggi X, 1953 (da Nature.com)

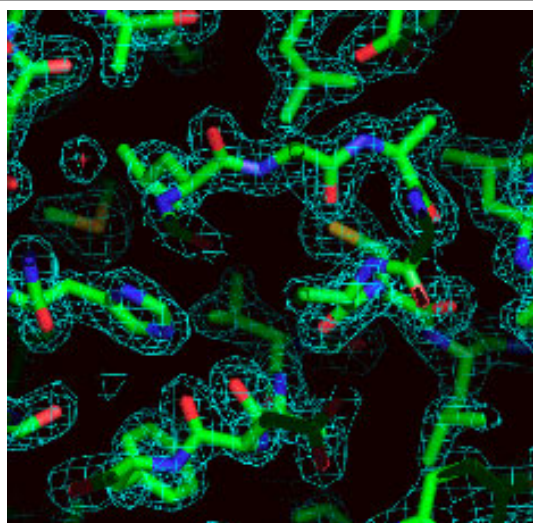


Figura 6 - Zoom nelle densità elettroniche a risoluzione atomica risolte dalla cristallografia a raggi X (reticolato azzurro) che mostra la possibilità di assegnare precisamente la posizione degli atomi di ciascun aminoacido (rappresentati come tubi verdi, rosso, blu, arancio) presente nella proteina.

La risonanza magnetica nucleare (NMR) è una tecnica di indagine della materia basata sulla misura della rotazione dello spin di protoni o di altri nuclei dotati di momento magnetico quando sono sottoposti ad un campo magnetico. Ogni nucleo mostra le sue caratteristiche perché ruota a velocità differente a seconda della sua posizione nella molecola, risuonando quindi a frequenze radio diverse. La NMR non utilizza cristalli di proteine come la cristallografia, ma proteine in soluzione più vicine al loro stato naturale; tale tecnica può dare informazioni anche sulla dinamica di varie parti della proteina, l'unico limite è che possono essere caratterizzate solo proteine di piccole dimensioni, ossia con meno di trecento aminoacidi.

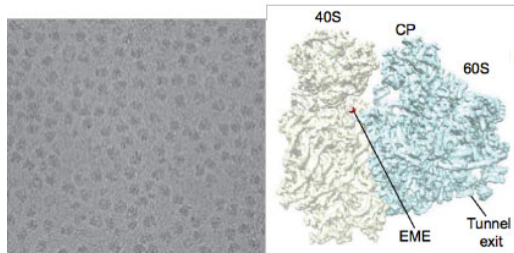


Figura 7 - Ricostruzione tridimensionale del ribosoma del *Plasmodium falciparum* tramite microscopia elettronica a particella singola in condizioni criogeniche (Wong et al., *elife* 2014). A destra la struttura del ribosoma con la posizione dell'emetina (antibiotico antimalarico) in rosso; a sinistra le proiezioni bidimensionali del ribosoma come viste al microscopio

La microscopia elettronica sfrutta come sorgente di radiazione un fascio di elettroni generato all'interno di un microscopio. Per studi di biologia strutturale ad alta risoluzione, si utilizza una sottobranca della microscopia elettronica, la microscopia elettronica criogenica (cryo-EM). Con questa tecnica, i campioni (siano essi soluzioni proteiche omogenee, organelli cellulari, o sezioni di cellule o tessuti), vengono posizionati su una griglia del diametro di circa 2 millimetri e poi rapidamente raffreddati in una soluzione di etano liquido in un processo cosiddetto di *vetrificazione*.

Tale procedura, non solo immobilizza il campione nel suo stato nativo in soluzione acquosa, ma lo rende anche più resistente al getto di elettroni del microscopio. Si produce così un solido amorfo della soluzione proteica nativa che viene inserito nel microscopio elettronico mantenendo condizioni di bassa temperatura (circa $-170\text{ }^{\circ}\text{C}$) e di alto vuoto per permettere la raccolta di immagini del campione all'interno delle fessure di questa griglia di rame. Con questa tecnica, negli ultimi due anni, sono stati visualizzati interi organelli cellulari o sezioni di cellule a risoluzione intorno ai 40 Å, mentre ribosomi e virus ed alcuni complessi multiproteici sono stati caratterizzati a risoluzione semi-atomica ($\sim 3\text{ Å}$). Recentemente anche proteine di membrana (gli "ossi duri" della biologia strutturale) sono state determinate ad alta risoluzione mediante questa tecnica, aprendo enormi prospettive alla tecnica per gli anni a venire. Il principale svantaggio della microscopia elettronica è il limite di dimensione degli oggetti osservabili (mai minori di 100 kDa) ed il limite di risoluzione (teoricamente circa 2.5 Å per campioni biologici), che ne impedisce l'applicabilità indiscriminata a tutti i sistemi biologici.

Un altro approccio molto utilizzato in biologia strutturale è quello della biofisica computazionale, in particolare le simulazioni di dinamica molecolare. Proprio per questi studi, il premio Nobel per la chimica 2013 è stato assegnato a Martin Karplus, (università di Strasburgo, Harvard), Michael Levitt (università di Stanford) e Arieh Warshel (università della California del Sud). Queste metodologie possono predire la dinamica conformazionale delle macromolecole nel tempo favorendone la comprensione del meccanismo di funzionamento. Grazie a queste metodologie è possibile, in alcuni casi, predire come una sequenza di aminoacidi si ripiega nello spazio nella sua forma proteica funzionale, oppure prevedere i possibili effetti legati a mutazioni genetiche (e conseguenti variazioni di sequenza aminoacidica) nella funzione di determinate proteine, oppure ancora prevedere i passaggi intermedi necessari nell'assemblaggio dei virus, e molto altro. Questi approcci predittivi, fino a qualche anno fa estremamente limitati a causa dell'elevato costo computazionale necessario per effettuare le predizioni, stanno prendendo sempre più piede grazie alla crescente disponibilità di computers con processori capaci di lavorare in parallelo. La potenza di calcolo di questi super-computer è elevata ma ancora limitata per simulazioni che utilizzano tutti gli atomi presenti nel sistema: poche centinaia di nanosecondi per giorno in un sistema acqua-proteina di circa 25000 atomi (proteina < 200 kDa), quindi nuove generazioni ed architetture di super-computer e nuovi algoritmi stanno nascendo con lo scopo di aumentare al rapidità di processamento di dati decine di volte con la speranza di poter raggiungere presto millisecondi di simulazione in poche settimane per sistemi biologici di dimensioni maggiori.

Voglio menzionare infine la microscopia ottica, tecnica antichissima (il primo microscopio ottico fu creato nel 1611 da Keplero) che non è ascrivibile classicamente alla biologia strutturale perché

incapace di risolvere proteine ed acidi nucleici ad alta risoluzione spaziale. Nonostante le limitazioni di risoluzione, questa tecnica è divenuta un ottimo ponte tra biologia cellulare e strutturale per lo studio della morfologia e dell'organizzazione funzionale di organelli, tessuti, organi e per localizzazione di macromolecole come proteine caratterizzate dalle tecniche della biologia strutturale ad alta risoluzione. La complementarità di tali approcci può fornire dettagli precisi del funzionamento di un organello o di una cellula a vari livelli.

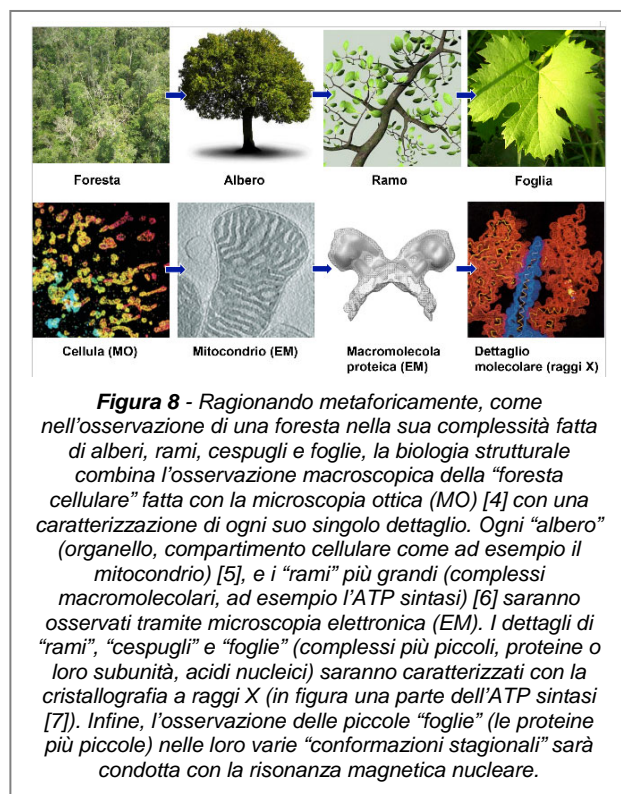


Figura 8 - Ragonando metaforicamente, come nell'osservazione di una foresta nella sua complessità fatta di alberi, rami, cespugli e foglie, la biologia strutturale combina l'osservazione macroscopica della "foresta cellulare" fatta con la microscopia ottica (MO) [4] con una caratterizzazione di ogni suo singolo dettaglio. Ogni "albero" (organello, compartimento cellulare come ad esempio il mitocondrio) [5], e i "rami" più grandi (complessi macromolecolari, ad esempio l'ATP sintasi) [6] saranno osservati tramite microscopia elettronica (EM). I dettagli di "rami", "cespugli" e "foglie" (complessi più piccoli, proteine o loro subunità, acidi nucleici) saranno caratterizzati con la cristallografia a raggi X (in figura una parte dell'ATP sintasi [7]). Infine, l'osservazione delle piccole "foglie" (le proteine più piccole) nelle loro varie "conformazioni stagionali" sarà condotta con la risonanza magnetica nucleare.

La collaborazione fruttuosa tra sviluppatori di strumentazioni e tecniche (principalmente fisici ed ingegneri) ed utilizzatori orientati allo studio di importanti problemi legati alle "molecole della vita" (biologi strutturali, biochimici e biologi molecolari), rende al giorno d'oggi il campo della biologia strutturale uno dei più multidisciplinari di tutte le scienze biologiche. Questo carattere multidisciplinare, unito alla ricerca continua di strumentazioni d'avanguardia, ha portato alla nascita di centri internazionali dove è possibile portare i propri campioni biologici ad analizzare. Questi centri sono già presenti e ben stabiliti per quanto riguarda la cristallografia a raggi X, che richiede enormi acceleratori di particelle (sincrotroni) per ottenere dati strutturali sulle proteine. Anche nel campo della microscopia elettronica stanno iniziando a comparire centri di ricerca internazionali dove esperti possono acquisire e inventariare dati con microscopi all'avanguardia una volta ricevuti i campioni da gruppi interessati a caratterizzazioni strutturali con

queste tecniche. Inoltre, metodi emergenti come la cristallografia con laser a elettroni liberi dopo tanti anni di sviluppo stanno prendendo piede in impianti giganti che a macchia d'olio stanno nascendo grazie ad investimenti di stati congiunti con enormi possibilità per il futuro.

Dal punto di vista applicativo, la possibilità di conoscere la struttura atomica delle proteine apre la strada a molteplici studi finalizzati a creare farmaci ad hoc per molecole target specifiche. Il "drug-design" basato su strutture atomiche di proteine rappresenta ormai la quotidianità nella ricerca farmaceutica contemporanea. Questo tipo di approccio è stato fondamentale per la scoperta di vari agenti terapeutici tra i quali figurano farmaci anti-cancro, antivirali o antinfiammatori. Tale metodologia si basa sulla possibilità di trovare una o più molecole (ligandi, di solito piccoli composti chimici) capaci di interagire stabilmente con le proteine target. L'individuazione di ligandi fa uso delle strutture tridimensionali delle proteine cercando molecole compatibili con la proteina target all'interno di grandi database di molecole esistenti (con potenziale d'azione quali ligandi già individuati per altri contesti biologici, oppure semplici archivi di composti chimici), oppure "disegnando" nuove molecole all'interno delle strutture tridimensionali delle proteine.

Disporre di informazioni strutturali su un numero sempre crescente (idealmente: tutte) di macromolecole biologiche e la caratterizzazione delle loro complesse interazioni all'interno di cellule e tessuti è davvero utile per comprendere le dinamiche dei vari processi che stanno alla base dei processi vitali degli organismi viventi, e per interpretare i meccanismi che determinano malfunzionamenti (malattie) e aggressioni da parte dell'ambiente esterno (come ad esempio nel caso di infezioni o altre invasioni da parte di agenti patogeni). Il vasto repertorio di tecniche descritte in questo articolo, insieme a molte altre che non sono state menzionate (come saggi biochimici e le tecniche di DNA ricombinante solo per citare due esempi) ci possono permettere di esplorare, comprendere e caratterizzare i principi emergenti che regolano l'organizzazione e la funzione di una cellula e di più cellule interagenti, con importanti possibilità applicative nel campo della medicina e della biotecnologia.

Bibliografia

- [1] Luisi PL. *The emergence of life. From chemical origins to synthetic biology* Cambridge Univ Press, (2006).
- [2] Chiarabelli C. *et al.*, Investigation of de novo totally random biosequences, Part II: On the folding frequency in a totally random library of de novo proteins obtained by phage display *Chemistry and Biodiversity*, 3, 840-859 (2006).
- [3] Porter, K.R., *et al.*, A study of tissue culture cells by electron microscopy: methods and preliminary observations *J. Exp. Med.*, 81, 233-241 (1945).

[4] Huang, B. *et al.*, Whole-cell 3D STORM reveals interactions between cellular structures with nanometer-scale resolution *Nature Methods*, 5, 1047-1052 (2008).

[5] Mourier, A. *et al.*, Loss of LRPPRC causes ATP synthase deficiency *Hum. Mol. Genet.*, 23, 2580-2592 (2014).

[6] Davies, KM. *et al.* Structure of the yeast F1Fo-ATP synthase dimer and its role in shaping the mitochondrial cristae *Proc Natl Acad Sci U S A*, 109, 13602-13607 (2012).

[7] Abrahams, JP. *et al.* Structure at 2.8 Å resolution of F1-ATPase from bovine heart mitochondria. *Nature*, 370, 621-628 (1994).

Autore: Matteo Allegretti

Matteo Allegretti è nato a Roma nel 1984. Laureato in biologia (triennale) e biochimica (magistrale) all'università degli studi di Roma Tre con una tesi sul tema dell'origine della vita e modelli cellulari, sta concludendo un dottorato al Max Planck Institute di biofisica di Francoforte dove utilizza la microscopia elettronica per studiare macromolecole, in particolare proteine di membrana coinvolte nella genesi dell'energia chimica cellulare. La sua curiosità è nella ricerca dei principi emergenti che regolano l'organizzazione molecolare e cellulare con una spinta verso un approccio inter-disciplinare alla ricerca scientifica che a suo parere ha molto bisogno di respirare. I suoi interessi spaziano dal biologia alla filosofia toccando saltuariamente tematiche sociali e escursioni a cielo aperto.