

SILVIO A BECCARA

LE PROTEINE ED IL RIPIEGAMENTO

ABSTRACT - A BECCARA S., 2014 - Proteins and their folding.

Atti Acc. Rov. Agiati, a. 264, 2014, ser. IX, vol. IV, B: 53-62.

In this paper basics about protein structure and protein folding are introduced. An account is given about the approach developed by the computational biophysics group at the LISC in order to sample protein folding and conformational transitions, along with a brief summary of most recent results.

KEY WORDS - Protein folding, Proteins, Computer simulations, Molecular biology.

RIASSUNTO - A BECCARA S., 2014 - Le proteine ed il ripiegamento.

In questo articolo vengono introdotti i concetti fondamentali riguardo alla struttura delle proteine e al loro ripiegamento. Si danno inoltre cenni riguardo al metodo sviluppato presso il LISC per campionare il ripiegamento delle proteine e le transizioni conformazionali, insieme ad un breve riassunto dei risultati più recenti.

PAROLE CHIAVE - Ripiegamento delle proteine, Proteine, Simulazioni al computer, Biologia molecolare.

Al laboratorio LISC (Laboratorio Interdisciplinare di Scienza Computazionale) della Fondazione Bruno Kessler e Università di Trento, ci occupiamo di proteine, ed in particolare del loro ripiegamento.

Che cosa sono le proteine? Probabilmente molti di noi sentendo questa parola penseranno all'alimentazione, e magari alle raccomandazioni che hanno ricevuto da piccoli, dai propri genitori: «Devi mangiare più proteine!». Questa associazione mentale non è sbagliata: le proteine sono tra le sostanze importanti che dobbiamo assumere con il cibo, ed il motivo è che svolgono una serie di funzioni fondamentali per il nostro organismo, come pure per quello di tutti gli esseri viventi.

Le proteine sono i “mattoni” delle nostre cellule, in quanto ne costituiscono la struttura portante. Le fibre dei nostri muscoli sono fatte

principalmente di proteine (come la miosina), come pure i tendini (perlopiù di collagene) e i capelli e le unghie (perlopiù di cheratina).

Le proteine però hanno anche altre funzioni, altrettanto importanti. Per esempio ci difendono dagli attacchi degli agenti patogeni: tutti gli anticorpi sono proteine, prodotte dai globuli bianchi del nostro sistema immunitario. La battaglia tra il nostro organismo e batteri e virus somiglia un po' a quella tra agenti segreti di nazioni contrapposte, con mascheramenti, inganni e agguati a tradimento. Gli strumenti di questa battaglia sono ancora proteine, generalmente agganciate alla superficie cellulare, che funzionano più o meno come dei codici segreti: possono essere riconosciuti ed interpretati, ma anche mascherati, e quindi ignorati.

Senza proteine le nostre cellule non potrebbero nemmeno respirare, in quanto l'ossigeno nel nostro sangue viene trasportato da una proteina, l'emoglobina, e lo stesso vale per l'anidride carbonica, o CO_2 , della quale si parla tanto ultimamente a proposito dell'effetto serra.

Le proteine funzionano poi come delle "nanofabbriche", all'interno delle quali si realizzano le reazioni chimiche che ci permettono di digerire il cibo, e quindi nutrirci, di distruggere eventuali sostanze nocive (come l'alcool), e di far funzionare le "centrali elettriche" delle nostre cellule, in modo da dare energia al nostro organismo.

Le proteine sono indispensabili anche per permettere alle piante di catturare la luce solare, e quindi di crescere e di dare così cibo a tutto il pianeta.

Una proteina, che si chiama rodopsina, vi permette di leggere questo articolo, e più in generale di vedere il mondo: si trova nei coni e bastoncelli, cellule della nostra retina sensibili alla luce, e produce un piccolo segnale elettrico ogni volta che viene colpita da un fotone, cioè un pacchetto di energia luminosa.

Se vi ho incuriosito un pochino riguardo alle proteine, adesso cerco di darvi un'idea di come sono fatte e di che cosa permette loro di svolgere una meravigliosa serie di compiti diversi. Le proteine prima di tutto sono molecole. Possiamo immaginare una molecola (dal latino: piccola quantità di materia) come un oggetto, molto piccolo, fatto di pezzi, ancora più piccoli, che si chiamano atomi, disposti secondo una struttura ben precisa. Lo stesso è vero del resto per gli oggetti che usiamo nella vita di tutti i giorni: per esempio un telefono è fatto di tasti, visore, batteria ed altre parti, e se per esempio metto le cuffie nel vano della batteria il telefono non funziona. Oppure una bicicletta deve avere due ruote ed un manubrio, con le ruote alle estremità del telaio. Gli atomi sono per la molecola ciò che il manubrio e le ruote sono per la bicicletta. Un

esempio di molecola (in apparenza) semplice è quella dell'acqua (Fig. 1): le palline che si vedono nella figura rappresentano, in maniera un po' rozza, l'atomo di ossigeno (in rosso) e quelli di idrogeno (in bianco), con i legami che li tengono assieme (le stecchette che collegano gli atomi tra di loro). Ci sono sempre due atomi di idrogeno per ogni atomo di ossigeno (come ci sono sempre

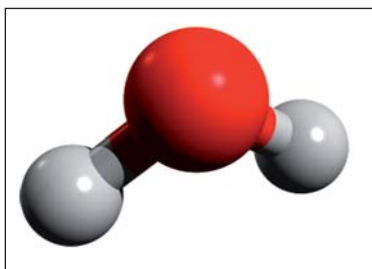


Fig. 1 - Una molecola di acqua.

due ruote per un manubrio nella bicicletta), e gli idrogeni si trovano ad una distanza ben definita dall'ossigeno (come le ruote si trovano ad una distanza ben definita tra loro e dal manubrio). Quanto è piccola una molecola di acqua? Per darvi un'idea, nel diametro (*non* nella lunghezza) di un vostro capello potrebbero entrare circa

300mila molecole di acqua messe in fila. Le proteine seguono lo stesso schema, cioè di atomi tenuti insieme da legami, con una struttura ben definita, ma sono un po' più grandi: le più semplici sono costituite da qualche centinaio di atomi, le più complicate possono arrivare a centinaia di migliaia.

Per fortuna è possibile suddividere le proteine, comunque complesse, in blocchi di base, che ci permettono di trattarle in maniera più semplice. Le proteine sono in sostanza molecole fatte di una singola catena

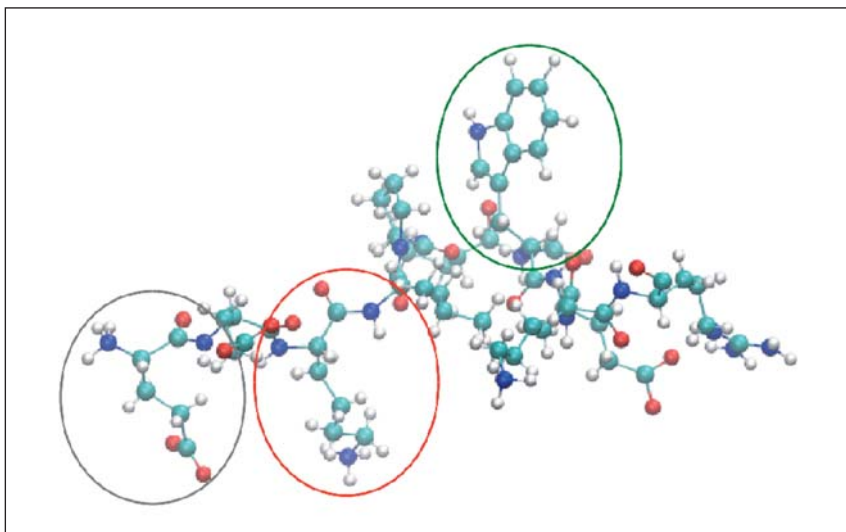


Fig. 2 - Struttura generale delle proteine.

di “anelli” (evidenziati da un cerchio nella Fig. 2), che si chiamano residui. In natura esistono 23 tipi diversi di residui, e tutte le proteine si possono costruire unendone insieme un numero qualsiasi, un po’ come in un Lego o un Meccano. La struttura di base di tutti i residui è sempre la stessa. Li possiamo considerare come degli “elementi componibili” che hanno tutti due attacchi identici: pensiamo magari ai fili con le lucine colorate dell’albero di Natale. Le “boccole di connessione” sono le palline blu in basso a sinistra e la rossa in basso a destra in tutte le illustrazioni qui sotto (atomi rispettivamente di azoto ed ossigeno); inoltre le boccole sono connesse tutte a due atomi di carbonio (le palline nere che stanno sopra). I residui si differenziano poi per la “ramificazione” laterale, cioè gli atomi che sono legati ai due carboni, indicati qui sotto da un’ellissi. Per esempio (Fig. 3), nel *triptofano* (primo da sinistra) la catena laterale è fatta di due anelli fusi insieme, nella serina (seconda da sinistra) di un atomo di ossigeno ed uno di idrogeno (gruppo *alcoolico*) e nella leucina (terza da sinistra) di una struttura ramificata di atomi di carbonio e di ossigeno.

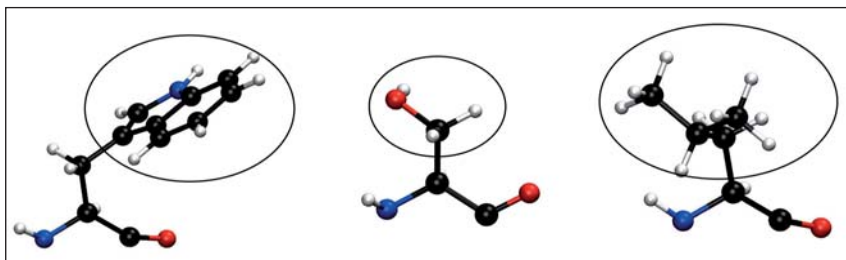


Fig. 3 - Tre amminoacidi diversi.

Fin qui tutto sembra semplice... ma come diceva qualcuno, il diavolo è nei dettagli.

E qui di dettagli ne mancano ancora parecchi. Prima di tutto una proteina ha una sua flessibilità intrinseca: in altre parole, una volta che ho costruito la mia catena, la posso piegare e ripiegare su se stessa, nelle tre dimensioni, in milioni di miliardi di modi diversi, tanto più numerosi quanto più è lunga (tre di questi “ripiegamenti” sono mostrati in Fig. 4, per una proteina abbastanza semplice). In secondo luogo, non tutti i ripiegamenti che posso realizzare vanno bene: in realtà uno solo sui milioni di miliardi possibili permette alla proteina di funzionare.

La cosa incredibile è che le proteine, una volta assemblate nel ribosoma (un organello cellulare), trovano l’unica struttura buona tra le milioni di miliardi possibili... da sole! È un po’ come se ordinaste una

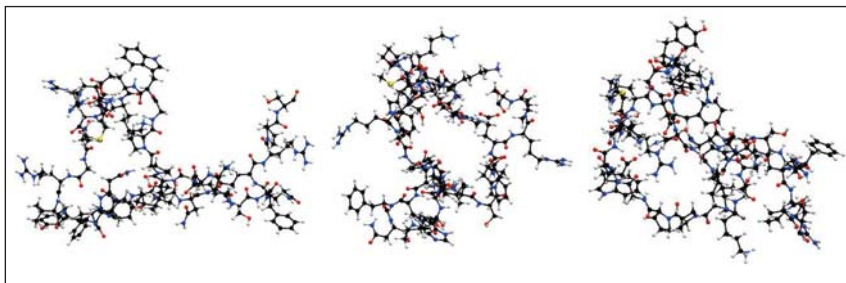


Fig. 4 - Tre ripiegamenti possibili per una proteina di piccole dimensioni, la fip35.

macchina qualsiasi in scatola di montaggio, per esempio un tagliaerba, e all'apertura della scatola i pezzi si avvitassero da soli uno con l'altro fino a costruirlo bello e pronto per funzionare.

Come fanno le proteine ad auto-ripiegarsi? Se dovessero provare tutte le possibilità una per una, anche se lo facessero molto in fretta, impiegherebbero un tempo più lungo dell'età dell'universo. Loro però ci riescono in frazioni di secondo (per fortuna per noi!): in media da qualche milionesimo di secondo a qualche millesimo di secondo.

A dire la verità, qualche volta – molto raramente – il ripiegamento non va per il verso giusto: la molecola assume una struttura che non è quella corretta, magari anche solo per qualche dettaglio. A volte questo non comporta danni per l'organismo, perché la proteina si spiega e si ripiega nuovamente. Se però ciò non accade, essa può diventare tossica, dando luogo a malattie come il “morbo della mucca pazza”, quello di Alzheimer, quello di Parkinson, o le sclerosi.

Insomma, capire come funziona il ripiegamento delle proteine è un bell'enigma per la scienza di base, ma ha anche implicazioni molto importanti per possibili applicazioni biomediche. Purtroppo studiare il ripiegamento con esperimenti fatti in laboratorio è molto difficile: si ottengono sempre informazioni parziali e indirette, che richiedono per di più un'elaborazione piuttosto complicata e non univoca. Ci sono sì nuove tecniche sperimentali in fase di sviluppo, ma sono ancora ben lontane dall'essere operative.

A questo punto entriamo in gioco noi, biofisici computazionali (squillo di trombe...). Questo nome altisonante significa che il nostro lavoro consiste nel cercare di capire qualcosa sul funzionamento delle molecole importanti per la vita, come le proteine o il DNA, realizzando dei modelli fisici che mettiamo all'opera usando calcolatori, o *computers*. Al giorno d'oggi le conoscenze di meccanica quantistica e classica sono abbastanza avanzate da permetterci di costruire dei modelli abba-

stanza realistici delle forze che agiscono sugli atomi di una proteina, e abbiamo anche calcolatori abbastanza potenti da tener conto di tutti gli atomi, anche per proteine abbastanza grandi. In pratica, una volta che abbiamo costruito il modello possiamo fare una “simulazione” del movimento della proteina: pensatela come un video molto dettagliato (magari in alta definizione). Questa video si chiama “dinamica molecolare”. Se poi abbiamo bisogno di cambiare qualche dettaglio della nostra molecola possiamo farlo abbastanza facilmente, al contrario di quello che accade in un laboratorio sperimentale, dove magari bisogna costruire ex-novo una nuova macchina, o comprare reagenti rari e costosi, o usare sostanze pericolose, o spostare tutto l’esperimento in un laboratorio molto distante.

Tutto risolto allora? Basta schiacciare un tasto ed abbiamo tutte le risposte? Purtroppo no. Fra noi ed i risultati che ci interessano si para il problema delle “scale temporali”. No, non si tratta di gradini colpiti da tuoni e lampi... detto in soldoni, nel video della nostra proteina il tempo tra un fotogramma e l’altro, per una questione tecnica un po’ difficile da spiegare, è giocoforza brevissimo: dell’ordine del milionesimo di miliardesimo di secondo. Per vedere il ripiegamento della proteina più semplice avremmo quindi bisogno di almeno 10 miliardi di fotogrammi! Per molecole più complicate questo numero può aumentare anche di 1000 volte, diventando insostenibile. Nemmeno i calcolatori più potenti al mondo sono in grado di compiere questa impresa, neanche se li usassimo per anni di seguito. Al massimo potremmo avere i primi istanti del video: sarebbe un po’ come poter vedere solo i titoli iniziali del nostro film preferito, o solo il calcio d’inizio della partita dei mondiali!

Dato che un approccio “a forza bruta” non funziona, abbiamo cercato di inventarci un metodo più efficiente. Per spiegarlo, possiamo partire da una domanda che ci siamo posti (insieme a molti altri ricercatori): perché è necessario avere un numero così grande di “fotogrammi” nel nostro video? Cerchiamo di rappresentarci quello che succede. Con un’immagine un po’ colorita, ma realistica, è un po’ come se la proteina, per potersi ripiegare, dovesse scalare una collina e poi discendere dall’altro versante. La proteina però non è in grado di muoversi da sola: deve fare affidamento su una serie di piccole spinte che le vengono date dalle molecole d’acqua che la circondano. Gli urti però non la indirizzano necessariamente nella direzione della cima della collina, perché avvengono a casaccio. Inoltre, alle temperature tipiche degli esseri viventi, sono molto deboli. Per questo motivo la molecola segue un percorso più simile a quello di un ubriaco che cerca di tornare a casa che a quello di un campione della corsa in montagna: magari fa un passo ver-

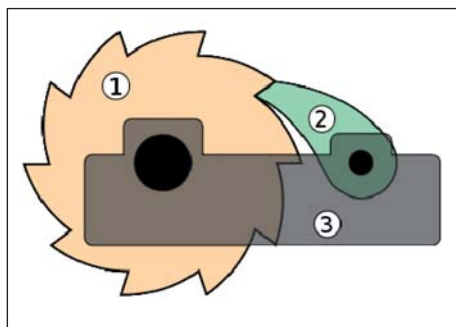


Fig. 5 - Ruota dentata con nottolino.

so l'alto, ma poi ne fa due all'indietro, poi va di nuovo avanti, poi magari sta ferma per un po', e così via. L'idea che verrebbe immediatamente è quella di "dare una mano" alla proteina: magari si potrebbe trainarla verso la cima della collina. In questo modo ridurremmo sì il tempo necessario per arrivare in cima e poi ridiscendere, però disturberemmo

anche in maniera pesante il suo moto naturale: i nostri risultati quindi non sarebbero corretti.

Noi del laboratorio LISC abbiamo adottato una strategia un po' diversa, e meno "perturbante". In pratica lasciamo la proteina libera di muoversi in modo del tutto naturale finché procede verso la cima della collina, ma la rallentiamo notevolmente quando tende a ricadere verso il basso. Per ottenere questo risultato agganciamo una specie di "elastico" alla proteina, un po' come nei cartoni animati. L'elastico però funziona solo all'indietro; in altre parole non c'è un effetto fionda in avanti. Il nostro sistema si può paragonare anche alla ruota dentata con nottolino, rappresentata nella figura (Fig. 5, presa da Wikipedia): il nottolino, cioè il dente curvo che si ingrana sulla ruota dentata, non fa niente se la ruota gira in senso antiorario, ma le impedisce di tornare indietro, ruotando in senso orario.

Per tornare alla similitudine con il video con troppi fotogrammi per essere memorizzati, è un po' come se nel nostro filmato mettessimo solo i fotogrammi nei quali la proteina sale verso la cima della collina, tralasciando quelli in cui torna indietro. In questo modo realizziamo un enorme risparmio di tempo di calcolo, e riusciamo a simulare in maniera molto efficiente il ripiegamento di proteine anche piuttosto grandi: il nostro record attuale è 25.370 atomi, immersi in circa 60.000 molecole di acqua.

In ogni caso, anche il nostro metodo non è del tutto "naturale": anche noi disturbiamo, pur se in misura decisamente minore, il moto della proteina. Per fortuna ci vengono in aiuto due strumenti fondamentali della fisica quantistica: gli integrali di cammino di Richard Feynman (premio Nobel per la Fisica nel 1965) ed il principio variazionale. Gli integrali di cammino sostanzialmente ci dicono che per descrivere il movimento di una particella quantistica tra due punti dati è necessario

tener conto di tutti i possibili cammini che li collegano, e che ciascuno di questi cammini ha una sua probabilità: alcuni sono molto probabili, altri meno. Per fare un esempio, se una persona deve andare in macchina da Trento a Rovereto, è più probabile che segua la strada provinciale o l'autostrada, meno probabile che vada prima a Carbonare dalla strada della Fricca, poi a Folgaria e poi a Rovereto scendendo per Besenello, e meno ancora che vada a Schio passando per la Val d'Astico e poi a Rovereto passando per la Vallarsa, anche se tutti questi percorsi sono possibili. Noi sfruttiamo gli integrali di cammino in quanto ci permettono di scegliere il percorso più probabile, che prevale su tutti gli altri. Però non conosciamo a priori questa probabilità, e nemmeno i cammini (se no avremmo già in tasca la soluzione). E qui entra in gioco il principio variazionale: ci dice che in alcuni casi è possibile trovare una soluzione approssimata (che non significa *approssimativa*, ma la più vicina possibile a quella esatta) ad un problema, e che la migliore è quella che presenta il valore più basso possibile di un certo parametro, cioè quella alla quale possiamo assegnare, con un calcolo matematico, il numero più basso. È un po' come cercare di trovare la strada più breve tra Trento e Rovereto provando tutti i percorsi a caso – supponendo però di non conoscere la geografia del Trentino – e misurando volta per volta la distanza percorsa: il percorso migliore sarà quello più breve. Noi facciamo lo stesso con i percorsi di ripiegamento delle proteine, ma al posto della distanza abbiamo un parametro che si chiama “azione”. Per ogni punto di partenza e di arrivo simuliamo un certo numero di percorsi (fino ad un centinaio: vedere come esempio la Fig. 6). Per ogni percorso calcoliamo poi l'azione, e il percorso con l'azione più bassa è quello che “vince”, cioè è il più probabile. Tra parentesi, è anche quello lungo il quale il moto “naturale” della proteina è stato perturbato in misura minore.

Il nostro metodo è molto efficiente, ma abbiamo comunque bisogno di usare calcolatori piuttosto potenti, in particolare per molecole grandi (diciamo da qualche centinaio di atomi in su). In altre parole, i nostri “video” non si possono ottenere sul computer di casa o sul cellulare. I calcolatori che usiamo per il nostro lavoro si dicono “paralleli”. Hanno a disposizione un numero piuttosto alto di processori (fino a qualche migliaio), che lavorano insieme su un dato problema, purché sia stato prima scomposto in parti più piccole, e comunicano tra loro per mezzo di una rete molto veloce (almeno 100 volte più veloce della connessione Internet che normalmente abbiamo nelle nostre case). Una tipica simulazione può occupare 400 processori per una giornata o due: sarebbe come usare il vostro PC di casa per uno o due anni. Questa mole di calcolo potrebbe sembrare enorme, ma il ripiegamento delle

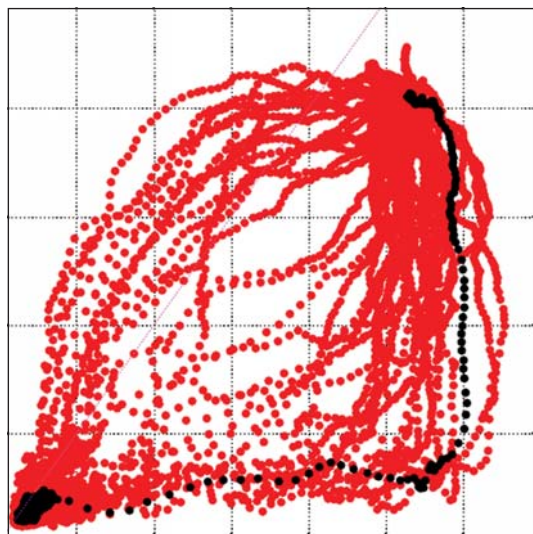


Fig. 6 - Diagramma dei diversi percorsi di ripiegamento per la proteina fip35.

proteine non può essere simulato altrimenti su nessun computer, nemmeno quelli più potenti al mondo. L'unica eccezione è data dal supercalcolatore Anton, costruito appositamente negli USA dal miliardario D. E. Shaw per fare questo tipo di operazioni; ma anche Anton è in grado di trattare al massimo molecole di un migliaio di atomi.

Per essere sicuri della correttezza del nostro metodo abbiamo confrontato i nostri risultati con dati sperimentali e simulazioni Anton, relativi a due proteine abbastanza piccole ma comunque complesse da simulare, e ben studiate dalla comunità scientifica: la fip35 (fatta di 562 atomi) e la villina (fatta di 583 atomi). In entrambi i casi l'accordo con i nostri risultati è molto buono, e conferma quindi la correttezza del nostro approccio.

In seguito abbiamo studiato il ripiegamento di una proteina annotata. Questa molecola si trova per esempio in batteri che vivono presso le sorgenti idrotermali, sul fondo dell'oceano. Lì la pressione e la temperatura raggiungono valori molto elevati, e pare che la molecola sia più resistente a queste condizioni estreme grazie al nodo che presenta. Ultimamente ci siamo occupati di una classe di proteine, le *serpine*, che sono presenti praticamente ovunque nel nostro corpo, in quanto servono a regolare funzioni come la coagulazione del sangue e la risposta del sistema immunitario. Queste molecole sono fatte di circa seimila atomi, quindi sono totalmente al di fuori della portata di Anton. Nella Fig. 7 si vede una serpina, il cui nome è PAI-1. Non sono mostrati tutti gli atomi, perché l'illustrazione sarebbe troppo confusa, ma solo le strutture



Fig. 7 - La serpina PAI-1.

principali della proteina. Siamo riusciti a visualizzare il meccanismo di transizione da una forma della serpina, cosiddetta *attiva*, a quella *latente*. La forma attiva è quella in grado di svolgere la funzione biologica, ma se presente in concentrazione troppo alta può risultare tossica, dando origine per esempio ad alcuni tipi di tumore e di malattie cardiovascolari. Per questo motivo è necessario tenere sotto controllo la concentrazione della forma attiva, diminuendola se necessario. A questo scopo si usano dei farmaci, costituiti da molecole di piccole dimensioni, dette *peptidi*: in pratica piccole proteine. Con le nostre simulazioni siamo riusciti a capire il modo in cui questi farmaci fun-

zionano. In particolare siamo riusciti a scoprire in quale zona della proteina si “agganciano” e la sua forma quando questo avviene. I nostri risultati possono quindi essere interessanti non solo dal punto di vista della scienza di base, ma anche per eventuali applicazioni biomediche. Attualmente, in collaborazione con colleghi dell’Università del Maryland, stiamo studiando il ripiegamento delle serpine, e in particolare stiamo cercando di capire che cosa succede quando la proteina si ripiega nel “modo sbagliato”, e che cosa provoca questo problema. I primi risultati sono molto buoni, e ci incoraggiano a proseguire in questa avventura entusiasmante.