

## *L'RNA è uscito dal cono d'ombra e spiega l'altra faccia della vita*

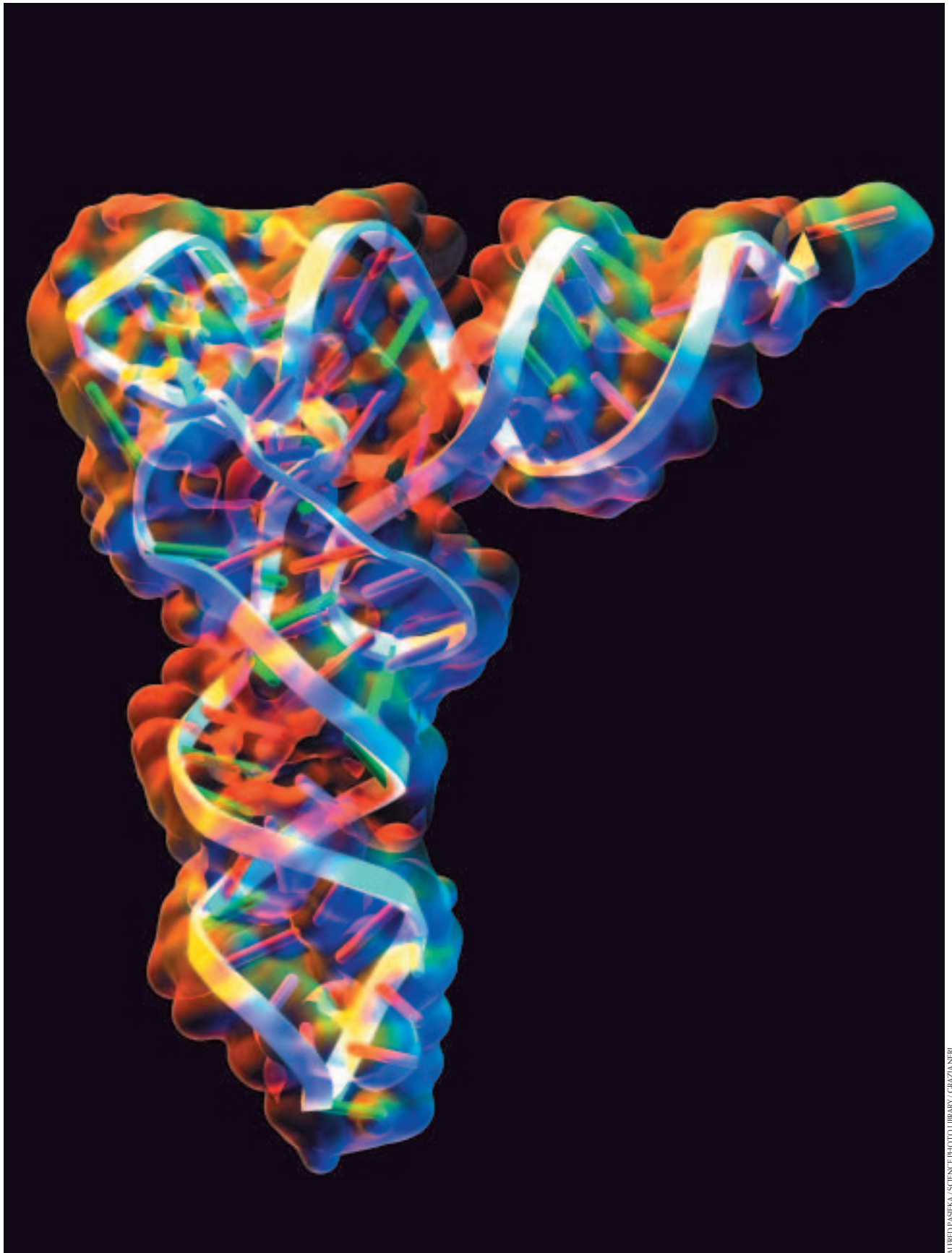
Mezzo secolo dopo la scoperta della doppia elica la ribalta torna agli acidi ribonucleici che come messaggeri adattatori ed enzimi avevano già contribuito alla nascita della biologia molecolare

GILBERTO CORBELLINI

**L'**ENFASI CON CUI NEL 2003 È STATO CELEBRATO il cinquantenario della scoperta che il DNA possiede una struttura a doppia elica ha in qualche modo mascherato il fatto che, negli ultimi vent'anni, è entrata in crisi l'idea che il DNA rappresenti la chiave per capire le basi molecolari della vita. Il sequenziamento del genoma ha sollevato soprattutto nuove domande. A cominciare da come viene conservata, letta, ma soprattutto elaborata e amplificata l'informazione genetica. Di certo nel DNA c'è molta più informazione di quella che è stata e sarà trovata nelle sequenze che codificano per le proteine. Questo "di più" con ogni probabilità sarà trovato nel modo in cui il DNA è organizzato spazialmente, nel modo in cui gli apparati di lettura – cioè di trascrizione e traduzione – accedono all'informazione e la elaborano nelle fasi post-trascrizionali, e nel ruolo che la grande quantità di RNA non tradotto in proteine svolge nella regolazione epigenetica dell'espressione genica.

Se gli ultimi cinquant'anni della biologia sono stati all'insegna del DNA, i prossimi decenni potrebbero essere all'insegna dell'RNA. Come mostrano gli articoli di Giuseppe Macino e Carlo Cogoni nelle pagine precedenti, il mondo dei piccoli RNA si sta rivelando quanto mai ricco di sorprese, così che il ruolo biologico di questa molecola acquista sempre più importanza. Il che non dovrebbe meravigliare se fosse vero che, come ormai pensa la maggior parte dei biologi molecolari, l'informazione genetica e il codice del DNA sono stati inventati dall'evoluzione molecolare a partire da un mondo che aveva come protagonista l'RNA. Agli inizi degli anni Ottanta infatti la scoperta dei ribozimi, cioè di RNA dotati di funzioni enzimatiche, ha reso verosimile l'ipotesi avanzata nel 1968 da Francis Crick e Leslie Orgel: prima che il DNA diventasse il depositario dell'informazione genetica esisteva il mondo dell'RNA, un mondo in cui questa molecola racchiudeva in sé le funzioni che in seguito si sarebbero separate, ovvero la memoria genetica e l'attività enzimatica.

Anche se la maggior parte delle cognizioni storiche sulle origini della biologia molecolare e la scoperta del codice genetico ruotano attorno al DNA, di fatto sono state una serie di indagini chimiche sull'RNA, e alcune scoperte sul suo ruolo nel trasferimento dell'informazione genetica, a produrre concettualmente e operativamente gran parte delle conoscenze e delle tecniche alla base della biologia molecolare. Il DNA fu scoperto tra il febbraio e l'agosto del 1869 dal fisiologo te-



Raffigurazione al computer di un RNA di trasferimento (tRNA).

desco Friedrich Miescher, che estrasse dal nucleo di cellule di pus, e poi dallo sperma di salmone, una sostanza chimicamente diversa dalle proteine, che chiamò nucleina. Prima della fine del secolo furono caratterizzate chimicamente le quattro basi azotate di quello che era chiamato acido nucleico: adenina, timina, citosina e guanina. Nel 1900 il biochimico italiano Alberto Ascoli, mentre lavorava nel laboratorio di Albrecht Kossel a Berlino, isolò dalle cellule di lievito una nuova base, l'uracile, e qualche anno dopo Phoebus Levene, il più importante chimico degli acidi nucleici nei primi decenni del Novecento, dimostrò che nell'acido nucleico del lievito l'uracile è presente al posto della timina.

Agli inizi del Novecento si sapeva dunque che gli acidi nucleici contengono, oltre al fosforo (nella forma di un derivato dell'acido fosforico), due purine (adenina e guanina) e tre pirimidine (timina, citosina e uracile), più alcuni zuccheri di cui però non erano chiare le caratteristiche. Che, comunque, apparivano diverse a seconda che gli acidi nucleici fossero ottenuti dalle cellule di lievito o dal timo di vitello (le due principali fonti da cui venivano estratti). Le differenze tra i due tipi di acidi nucleici furono messe in relazione con la loro origine: quelli che contenevano l'uracile, derivati dal lievito, erano considerati caratteristici delle cellule vegetali, mentre quelli contenenti la timina erano ritenuti caratteristici delle cellule animali. Occorsero diversi anni perché questa erronea nozione venisse superata. Nel 1922 il presunto acido nucleico delle piante fu isolato anche nelle cellule del pancreas e due anni dopo si riuscì a mettere in evidenza il supposto acido nucleico animale anche in cellule vegetali. Nel 1930 Levene riuscì inoltre a differenziare gli zuccheri dei due diversi acidi nucleici, che risulteranno essere il ribosio per l'acido contenente l'uracile (da cui acido ribonucleico) e il deossiribosio per quello contenente la timina (da cui acido deossiribonucleico). Le abbreviazioni RNA, per l'acido ribonucleico, e DNA, per il deossiribonucleico, si affermeranno a metà degli anni Cinquanta.

Fino alla fine degli anni Venti gli acidi nucleici erano considerati degli aggregati colloidali e, a parte le intuizioni di qualche citologo della fine dell'Ottocento, a nessuno era venuto in mente che avessero qualcosa a che fare con l'ereditarietà genetica. Si pensava infatti che la base chimica dell'ereditarietà si trovasse nelle proteine, assai più variabili nella loro composizione e dotate di un'alta specificità conformazionale e di una spiccata selettività chimica. Le caratteristiche macromolecolari degli acidi nucleici, in particolare l'alto peso molecolare, emersero solo dagli studi fisico-chimici e cristallografici iniziati negli anni Trenta. E il primo acido nucleico di cui fu stabilita la natura macromolecolare fu proprio l'RNA,

grazie all'analisi biochimica della componente nucleoproteica del virus del mosaico del tabacco, cristallizzato da Wendell Stanley nel 1935. Il virus è composto da una componente proteica e da una nucleica, cercando di chiarire quale delle due fosse responsabile del trasporto di cellula in cellula dell'attività virale – questione risolta definitivamente solo nel 1956 – fu appurato che la componente nucleica era costituita da RNA, il cui peso molecolare venne calcolato nel 1942, da Seymour Cohen e Wendell Stanley, nell'ordine di 300.000 dalton.

### La sintesi proteica

Dopo la dimostrazione, da parte di Oswald Avery nel 1945, che il DNA è il materiale chimico che trasmette le caratteristiche ereditarie, e ben prima che Watson e Crick proponessero il modello della doppia elica, diversi biochimici ipotizzavano che il DNA potesse sintetizzare in qualche modo l'RNA, che a sua volta guidava la sintesi degli enzimi. Tra le ipotesi avanzate una, proposta da Alexander Dounce nel 1952, anticipava alcuni concetti chiave della spiegazione genetico-molecolare della sintesi proteica. Dounce propose che l'ordine degli amminoacidi in ogni proteina derivasse dall'ordine dei nucleotidi nella corrispondente molecola di acido nucleico, e che il DNA fungesse da stampo per la sintesi dell'RNA, che a sua volta era il modello su cui venivano costruite le proteine. Fu così formulato per la prima volta il fondamentale concetto della colinearità tra la sequenza dei nucleotidi nel DNA e la sequenza degli amminoacidi nelle proteine.

La prima prova che l'RNA è coinvolto nella sintesi proteica risaliva ad alcune osservazioni effettuate già nel 1939 da Torbjörn Caspersson e nel 1941 da Jean Brachet, i quali rilevarono che le cellule impegnate attivamente nella sintesi proteica sono ricche di RNA. In particolare, Brachet stabilì che le quantità di DNA aumentano proporzionalmente in rapporto alla divisione cellulare, mentre quelle dell'RNA sono proporzionali all'attività biosintetica della cellula. Nel 1955 Brachet verificò, attraverso esperimenti di enucleazione, che nelle cellule private del nucleo la sintesi proteica si protraeva per diversi giorni. Tra il 1950 e il 1955 fu definitivamente stabilito, anche sulla base di studi di microscopia elettronica, che la sintesi proteica avviene all'interno di componenti microsomiali costituiti in gran parte da particelle sferiche formate soprattutto da RNA (i ribosomi). Nel 1952 Alfred Harshey rilevò che nella cellula batterica, dopo l'infezione del fago T2, nuove molecole di RNA venivano sintetizzate a tasso elevato. Era dunque evidente che l'RNA funzionava da stampo nell'assemblaggio di amminoacidi specifici in proteine.

La scoperta che un RNA funge da messaggero per trasportare l'informazione dal nucleo ai ribosomi



JAMES KING/HOLMES/SCIENCE PHOTO LIBRARY / GAZZANERI

Sidney Brenner, uno degli scopritori dell'RNA messaggero.

emerse nel corso dello stesso esperimento che portò a un'altra scoperta fondamentale: quella che la sintesi di una particolare proteina è regolata geneticamente, attraverso l'azione di geni detti regolatori, che rispondono alle diverse situazioni ambientali impedendo o consentendo la produzione della proteina da parte del gene strutturale. L'esperimento in questione, concepito da François Jacob e Jacques Monod e realizzato nei laboratori dell'Institut Pasteur di Parigi, divenne noto come "esperimento Pigiama" dalla pronuncia inglese della corretta denominazione, "Pajamo", derivata dai cognomi dei tre ricercatori che lo realizzarono: Arthur Pardee, Jacob e Monod. I dati sperimentali portavano a ipotizzare che tra i geni e le particelle microsomiali, dove avviene la sintesi proteica, esistesse un intermediario a vita corta. Nel 1961 Sydney Brenner e lo stesso Jacob, insieme a Meselson, dimostrarono che l'RNA sintetizzato subito dopo l'infezione di un batterio da parte di un batteriofago è diverso dall'RNA ribosomiale e che questo RNA (battezzato messaggero) si associa ai ribosomi, che svolgono un ruolo passivo nel processo di sintesi delle proteine.

Prima della scoperta dell'RNA messaggero, era stato appurato che la traduzione della sequenza nucleotidica in una sequenza amminoacidica coinvolge anche un altro RNA. Al protagonista di quella scoperta, Paul Zamecnik, si deve anche l'invenzione del sistema sperimentale che ha consentito di decifrare

il codice genetico: un protocollo che rendeva possibile la sintesi proteica *in vitro*. Cercando di stabilire quali fossero i componenti della cellula coinvolti direttamente nella sintesi proteica, Zamecnik ottenne un sistema contenente amminoacidi, ATP (molecole in grado di fornire l'energia per il legame peptidico), microsomi e un surnatante derivato dalla centrifugazione ad alta velocità di estratti cellulari. Questo surnatante era un RNA solubile e risultò immediatamente corrispondere agli "adattatori" ipotizzati da Francis Crick nel 1954 per spiegare in che modo gli amminoacidi potevano essere montati secondo l'ordine stabilito dalla sequenza degli acidi nucleici. In pratica Mahlon Hoagland e Paul Zamecnik avevano scoperto i tRNA, o RNA di trasporto, e nel 1957 Robert Holley dimostrò che questi erano in numero corrispondente a quello degli amminoacidi.

A quel punto era possibile tentare di decifrare il codice genetico, sempre lavorando con l'RNA. La prima cifra (tripletta) del codice venne letta alle 6 di mattina del 27 maggio 1961 grazie a un elegante esperimento realizzato da Johann Heinrich Matthaei e ripetuto da Marshall Nirenberg. Essi fabbricarono un RNA messaggero (mRNA) costituito da una lunga sequenza di un solo nucleotide, l'uracile (poli U). Quando questo venne aggiunto a estratti cellulari contenenti ribosomi, si osservò che veniva sintetizzata una catena amminoacidica formata solo da polifenilalanina. Quindi la sequenza UUU doveva codi-

ficare per la fenilalanina. Immediatamente si scoprì che la sequenza (poli A) codificava per una catena di lisina e (poli C) per un polipeptide contenente solo prolina. Così procedendo, in pochi anni fu decifrato l'intero codice.

A metà degli anni Sessanta veniva intanto determinata la prima sequenza nucleotidica di un tRNA utilizzando una tecnica simile a quella impiegata per il sequenziamento delle proteine, cioè spezzando la molecola in piccoli frammenti con enzimi non specifici (le nucleasi). Fred Sanger e i suoi collaboratori svilupparono quindi un metodo di sequenziamento basato su tecniche di frazionamento mediante ionoforesi bidimensionale su acetato di cellulosa, che consentiva di separare miscele complesse di piccoli nucleotidi ottenuti dalla degradazione di RNA più grandi.

### Il dogma infranto

Nel maggio del 1970, in occasione del X Congresso Internazionale di Oncologia tenutosi a Houston, Howard Temin e David Baltimore annunciarono la scoperta di un nuovo enzima virale, che denominarono "DNA polimerasi diretta da RNA" e che presto sarebbe diventato noto con il nome di trascrittasi inversa. L'enzima era in grado di costruire molecole di DNA a duplice filamento a partire da stampi di RNA a filamento singolo. Nel 1974, in seguito alla scoperta della trascrittasi inversa, veniva creata la famiglia *Retroviridae*, comprendente tutti i virus che contengono l'antigene specifico per la DNA polimerasi diretta da RNA (detti retrovirus). La trascrittasi inversa sarebbe presto diventata uno strumento essenziale dell'ingegneria genetica per costruire i cDNA (DNA complementari) a partire dall'mRNA.

La scoperta della trascrittasi inversa cominciò a far luce sul comportamento dei virus animali a RNA, molti dei quali provocano tumori e sono ampiamente distribuiti fra diverse specie. La spiegazione del meccanismo attraverso cui questi virus riescono a trasmettere l'informazione dall'RNA al DNA nella cellula vivente ha reso possibile l'unificazione delle due ipotesi sull'origine del cancro che si confrontavano fino agli inizi degli anni Sessanta: quella genetica e quella virale. Tuttavia, l'interesse principale suscitato dalla scoperta riguardò l'idea che si trattasse di una confutazione del cosiddetto dogma centrale, secondo cui l'informazione fluisce in un'unica direzione dal DNA all'RNA alle proteine. Si ipotizzava inoltre che l'esistenza della trascrittasi inversa rendesse in qualche modo possibile l'eredità lamarckiana, immaginando che l'enzima consentisse il passaggio di informazione dalla linea somatica a quella germinale. Lo stesso Temin, peraltro, accreditò questa interpretazione con una improbabile teoria del "protovirus" in cui spiegava la morfogenesi assumendo che fosse la quantità di RNA in una cellula, e quindi lo stato fun-

zionale della cellula stessa, a controllare il contenuto di DNA di quella cellula. Temin poi abbandonò l'idea che la trascrittasi inversa rappresentasse la chiave per spiegare il controllo dell'espressione genica. Ma alcuni anni dopo un'altra scoperta riguardante l'RNA avrebbe portato alla luce un potente meccanismo epigenetico di amplificazione dell'informazione genetica.

Nel 1977 si scoprì che in alcuni virus animali esistono geni frammentati (split genes), cioè geni contenenti sequenze che non vengono trascritte come mRNA (dette introni) interposte tra le sequenze effettivamente codificanti (gli esoni). Questa organizzazione del materiale genetico venne immediatamente riscontrata anche negli eucarioti e in particolare modo negli eucarioti multicellulari. I geni frammentati danno luogo a un singolo trascritto chiamato RNA nucleare eterogeneo, da cui gli introni vengono rimossi nel corso della maturazione ed elaborazione dell'mRNA, attraverso un processo denominato splicing. Lo studio dello splicing, a metà degli anni Ottanta, ha portato alla scoperta di un complesso meccanismo deputato a tagliare e risaldare i segmenti di mRNA in maturazione (pre-mRNA), chiamato spliceosoma. Lo spliceosoma, studiato in diversi sistemi biologici, si è via via rivelato una macchina macromolecolare estremamente complessa, costituita da cinque piccoli RNA e alcune centinaia di proteine.

Lo spliceosoma è verosimilmente uno dei sistemi che amplificano e controllano l'espressione dell'informazione genetica e gioca un ruolo fondamentale nel fenomeno dello splicing alternativo, che consente di ottenere più di un prodotto genico da ogni unità trascrizionale. Di volta in volta, infatti, nel processo di formazione di un dato pre-mRNA alcuni esoni possono essere inclusi e altri esclusi, per cui a partire da un singolo gene si possono produrre più mRNA con strutture primarie differenti. Questo meccanismo è particolarmente importante per amplificare l'informazione genetica utilizzabile, come dimostra il caso delle immunoglobuline, e si è rivelato specifico dei tessuti e degli stadi di sviluppo, consentendo quindi la sintesi da uno stesso gene di prodotti diversi a seconda del tipo cellulare e della fase di sviluppo. Lo splicing alternativo, controllato dallo spliceosoma, rappresenta un meccanismo efficace di regolazione della funzione genica, soprattutto per le cellule giunte allo stadio terminale di differenziamento e che hanno perduto la capacità di replicare il DNA, nonché per le cellule che devono rispondere rapidamente agli stimoli ambientali. Dato il ruolo essenziale dello splicing per la funzionalità cellulare, è del tutto comprensibile che diverse atrofie, distrofie e neoplasie siano risultate associate a mutazioni degli RNA e delle proteine che formano lo spliceosoma.



JAMES KING-HOLMES / SCIENCE PHOTO LIBRARY / GAZZANIERI

Frederick Sanger (a destra), pioniere del sequenziamento delle proteine e degli acidi nucleici.

### Aspettative e delusioni

Nel corso degli anni Ottanta e Novanta l'RNA è stato al centro di grandi aspettative per le possibili applicazioni terapeutiche. Nel 1978 Paul Zamecnik ha dimostrato infatti la possibilità di bloccare l'espressione di un gene responsabile di una patologia, sintetizzando oligonucleotidi di DNA antisenso – cioè di sequenza complementare all'mRNA che porta ai ribosomi il "senso" dell'informazione genetica – che agisce proprio legandosi all'mRNA e bloccandolo. Qualche anno dopo si scopriva che anche l'mRNA può produrre sequenze antisenso e ciò ampliava ulteriormente le possibilità di sfruttare il principio per scopi terapeutici. Di fatto, sia le malattie con una base ereditaria sia quelle virali usano come intermediario l'mRNA per produrre le loro manifestazioni fenotipiche, per cui il principio alla base della terapia antisenso ha subito creato grandi aspettative. Ma ha anche dato più di una delusione. Di fatto il principio ispiratore è molto semplice e del tutto logico, ma far entrare nell'ambiente cellulare gli oligonucleotidi e fare in modo che arrivino nel nucleo e li riconoscano la minisequenza specifica dell'mRNA bersaglio non è per niente facile. I successi, soprattutto per quanto riguarda la terapia antisenso contro diversi tumori, sono derivati più da eventi fortuiti (per tentativo ed errore) che dalla scoperta di un sistema efficace.

Le delusioni non hanno risparmiato nemmeno altre strategie terapeutiche basate sull'RNA, come quella concepita dal premio Nobel Thomas Cech che, dopo aver scoperto i ribozimi (RNA dotati di attività enzimatica), ne ha sviluppati alcuni in grado di distruggere l'mRNA di agenti patogeni – come il virus Hiv – o dei geni per proteine coinvolte nella progressione dei tumori. Cech ha addirittura fondato un'impresa, la Ribozyme Pharmaceutical. Ma pur dimostrandosi abbastanza efficaci *in vitro*, i ribozimi non hanno ancora corrisposto alle attese a livello clinico. Non sono stati più fortunati, finora, i tentativi di sviluppare terapie che riparino le mutazioni a livello degli mRNA utilizzando la macchina molecolare responsabile dello splicing, né i tentativi di sfruttare alcuni piccoli RNA, capaci di assumere conformazioni tridimensionali con cui si legano alle proteine, per inibirne l'azione.

Il recente annuncio che in modelli animali si è riusciti a somministrare alcuni piccoli RNA – che bloccano la traduzione dell'mRNA per una apolipoproteina implicata nel metabolismo del colesterolo – ottenendo una riduzione significativa dei livelli della proteina bersaglio, sposterà ora le aspettative di un'applicazione terapeutica su questi piccoli RNA e sui loro meccanismi di interferenza.

*Gilberto Corbellini, storico della medicina, Università di Roma La Sapienza*