

Gli enzimi che troviamo in natura hanno una grande varietà di caratteristiche, di specificità di substrato e capacità di lavorare in condizione di reazioni blande. A volte però questi non sono adatti a processi di livello industriale, per quanto riguarda la stabilità, la specificità e l'attività.

In passato si tendeva a disegnare i processi di biocatalisi su misura per il catalizzatore, cercando un compromesso tra esigenze del processo e caratteristiche dell'enzima.

L'alternativa oggi è quella di progettare le condizioni ideali di reazione, per poi cercare l'enzima che meglio si adatta al nostro processo.

Ci sono caratteristiche della reazione sulle quali possiamo incidere in maniera diretta, ad esempio possiamo immobilizzare il biocatalizzatore. Questo può permettere spesso di stabilizzare l'enzima e migliorare alcune delle sue caratteristiche, aumentando ad esempio la specificità.

Un altro campo d'azione è il *medium engineering*, che si occupa di migliorare le condizioni del solvente in cui si effettua il processo.

Qualora gli enzimi a nostra disposizione non soddisfino i requisiti del nostro processo, possiamo sfruttare la biodiversità, cercando nuovi ceppi produttori di enzimi finora non utilizzati.

Frequentemente questi microorganismi si cercano in ambienti estremi, utilizzando anche tecniche di metagenomica.

Non è detto però che questo approccio ci dia il miglior risultato per i nostri scopi. A nostro supporto abbiamo in questo caso un'altra risorsa, il *protein engineering*. Questa disciplina fa riferimento ad una serie di tecniche che permettono di modificare le prestazioni di un enzima agendo sulla sua sequenza.

Questo approccio richiede i geni codificanti il catalizzatore di nostro interesse, in forma clonata.

Dovremo cercare un buon sistema di espressione per l'enzima, e padroneggiare le tecniche che servono per introdurre diversità nella sequenza, cioè i metodi di mutagenesi.

Abbiamo a disposizione diversi metodi per eseguire la mutagenesi.

Uno di questi è la mutagenesi sito specifica, che è una tecnica prettamente razionale e si basa sulla conoscenza dello specifico enzima, e di enzimi simili, che ci guida nella scelta di quali aminoacidi è opportuno cambiare.

Un esempio classico è la stabilizzazione dell'enzima.

Abbiamo ad esempio una esterasi che funziona benissimo nella reazione di nostro interesse, ma che è poco stabile nel solvente che utilizziamo. Quello che si fa nell'approccio razionale è confrontare sequenze e struttura del nostro enzima con quelle di enzimi omologhi ma un po' diversi.

Se abbiamo ad esempio informazioni su una esterasi più stabile nelle condizioni di nostro interesse, anche se con altre caratteristiche peggiori, ad esempio meno specifica, possiamo usare queste informazioni per progettare le mutazioni da inserire nel nostro enzima.

I mezzi di analisi bioinformatica possono essere utilizzati per ricavare questo tipo di informazioni.

Un questo corso parleremo poco di tecniche, perché già conosciute, ma si parlerà più di applicazioni. La mutagenesi sito diretta è data per scontata.

Nel dettaglio sarà approfondito il metodo dell'*evoluzione guidata*. Questa è una strategia di mutagenesi random. Oggi comunque raramente si lavora con tecniche completamente random.

Evoluzione Guidata

Nell'approccio razionale noi dobbiamo avere informazioni sulla proteina da ingegnerizzare, in modo da agire in maniera precisa sulla proteina, in base a delle ipotesi.

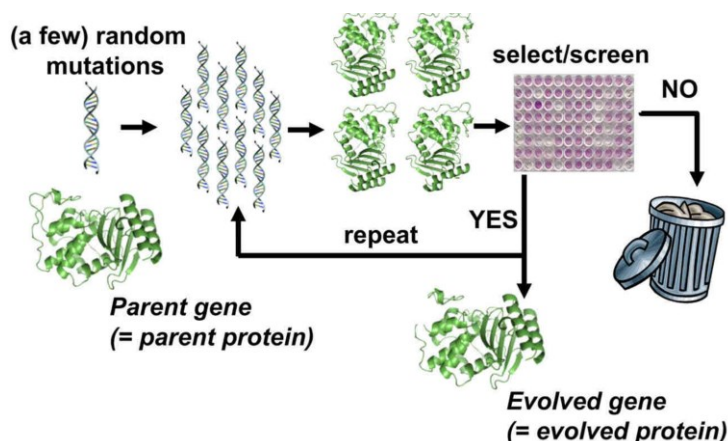
Nell'evoluzione guidata queste informazioni non sono necessarie. Bastano una sequenza genica ed un sistema di espressione adeguato.

Il metodo evolutivo ha un approccio ispirato a quello che avviene nell'evoluzione naturale.

Si parte in genere da una serie di sequenze della stessa proteina. Viene fatta una mutagenesi random, in modo da ottenere una serie di varianti della stessa proteina, di cui non sappiamo nessuna caratteristica.

Dato che cerchiamo una particolare proprietà della proteina, esprimiamo le sequenze mutagenizzate in un sistema di espressione e valutiamo le proprietà delle varianti. Selezioneremo le varianti che soddisfano le proprietà richieste. Ad esempio, se cerchiamo la stabilità, cerchiamo le proteine più stabili, mentre scartiamo le altre.

Le nuove sequenze possono essere nuovamente mutagenizzate e selezionate in più cicli. Alla fine avremo una sequenza evoluta.



La sequenza di partenza è detta *sequenza parentale*.

Ogni ciclo del processo sarà una generazione. Dopo il primo ciclo di mutagenesi avremo le sequenze di prima generazione.

Ci sono delle fondamentali differenze tra l'evoluzione diretta e l'evoluzione naturale.

- Noi scegliamo in anticipo quali sono le caratteristiche che vogliamo evolvere.

Questo è un bene ed un male allo stesso tempo. Noi stiamo facendo mutagenesi casuale e poi selezioniamo per una caratteristica specifica, ad esempio la stabilità. Può essere che in alcune varianti la mutagenesi abbia introdotto caratteristiche interessanti per altri scopi, ad esempio la specificità. Ma se noi non selezioniamo per quella caratteristica potremmo scartare queste varianti.

- A differenza della natura, la scala temporale in cui ci serve il risultato è ridotta. Settimane o mesi, non milioni di anni.

Ci sono alcuni aspetti da attenzionare negli esperimenti di evoluzione guidata:

- Dobbiamo avere a disposizione una sequenza parentale da cui partire. Quale scegliamo?
- Quale metodo di mutagenesi scegliamo?
- Quanti mutanti vogliamo fare?
- Come valutiamo se la proteina è migliorata o no?

Tecniche di evoluzione guidata (studiare review Packer et al. Methods for directed evolution...)

Possiamo immaginare la proprietà di nostro interesse come un massimo nell'ipersuperficie di quella specifica attività. Ad ogni generazione di mutagenesi avremo diverse vie di evoluzione, per cui sceglieremo la via che tende ad un miglioramento dell'attività cercata.

Scelta della sequenza parentale

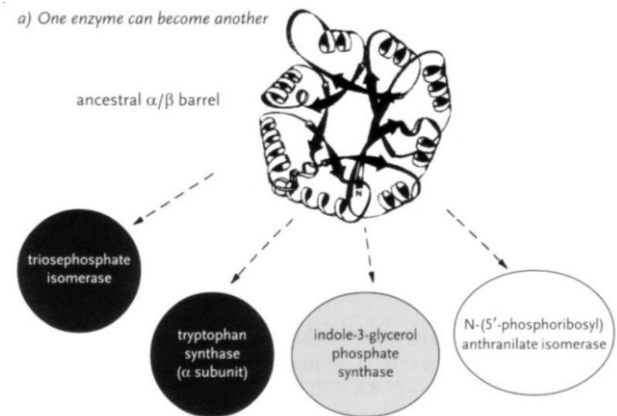
Abbiamo un'ampia quantità di sequenze proteiche su banche dati online, sia manualmente verificate (ad esempio su UniProt o SwissProt), sia annotate in maniera automatica (TrEMBL: ampia quantità di sequenze ottenute con i metodi di sequenziamento NGS. Molte di queste sequenze non sono mai state espresse).

Nel corso dell'evoluzione si sono affermati un numero limitato di tipi di organizzazione strutturale, su cui l'evoluzione ha poi basato tutta la ricchezza di funzioni proteiche ed enzimatiche che adesso vediamo.

Ipotizzando di dover produrre una proteina particolare e non esistente.

Ciò che sappiamo è che a partire da una certa struttura (un certo fold), organizzazione e funzione, nel corso dell'evoluzione si sono prodotte una serie di proteine *omologhe*, con caratteristiche differenti.

Un fold molto diffuso è il *TIM-barrel* (presente ad esempio nella *Trioso Fosfato Isomerasi*, un enzima del ciclo glicolitico), con un *foglietto-beta* interno circondato da *alfa-eliche*. Questa struttura si ritrova in vari enzimi con specificità molto diverse, come visibile in figura, essendo una organizzazione stabile e plastica, perché può accomodare siti attivi di enzimi anche molto diversi.



Di solito si parte da un enzima con almeno un po' di attività verso le reazioni di interesse. Spesso è la scelta più prudente.

Una alternativa è partire da strutture più solide, anche prive della funzione specifica cercata, con l'obiettivo di far emergere le nuove funzioni desiderate mediante l'evoluzione guidata.

Quanti mutanti facciamo?

Il problema dello *spazio di sequenza* consiste in questo:

se prendiamo una proteina, e vogliamo andare in modo del tutto casuale, e quindi valutare tutte le possibilità, devo testare tutte le combinazioni di ogni aminoacido con tutte le combinazioni di tutti gli altri aminoacidi, vedremo che dovremo testare un numero impressionante di mutanti, secondo la formula:

$$19^N \cdot x! / [(x - m)! m!]$$

dove N è il numero di mutanti, m è il numero degli aminoacidi sostituiti contemporaneamente ed x è il numero di residui che costituiscono l'enzima. (formula riportata come da slide, il numero di sequenze possibili. Utilizzando la notazione precedente è: x^{20})

Di questo si è molto discusso, ma nella realtà non è mai necessario esplorare tutto lo spazio di sequenza. Si può procedere per gradi, andandosi man mano a focalizzare sulle mutazioni di interesse.

Spesso adesso si inizia con una mutagenesi del tutto casuale, per poi andare a migliorare in maniera più specifica le varianti migliori.

Si possono usare approcci semi-random, in cui si sa che regioni specifiche della proteina abbiano maggior interesse per la nostra funzione. Questa regione se non conosciuta a priori può essere indicata dai primi cicli di mutagenesi.

Questo permette di focalizzare gli sforzi su una regione limitata della sequenza.

Selezione vs. screening

Applicare un processo di *selezione* è generalmente più facile.

Dopo la fase di mutagenesi si esprimono le sequenze mutate in organismi ospiti. Si possono a questo punto selezionare i cloni che esprimono la proteina di interesse. Questo però richiede che la cellula ospite abbia una proprietà vitale per l'organismo. Questo purtroppo capita raramente.

Nella maggior parte dei casi si evolvono proteine non vitali per i cloni, che andranno quindi studiati uno ad uno per selezionare la proprietà desiderata. Questo è lo **screening**.

Dato il numero di sequenze da studiare servono metodi di screening semplici, oppure screening in più passaggi. Questo permette di ridurre il numero di campioni da analizzare in maniera più approfondita.

You get what you screen for.

La valutazione degli esiti della mutazione e la selezione delle varianti più promettenti è il passo che richiede più sforzi dal punto di vista sperimentale.

La scelta del metodo dipende appunto dall'attività cercata e dalla capacità di analisi del laboratorio in cui ci troviamo.

Screening

È il metodo che più spesso utilizzeremo per la selezione, perché frequentemente gli enzimi che andremo a valutare non sono vitali per la cellula che li esprime.

Attività

Ci baseremo spesso sull'attività delle varianti, comparandole con l'attività della sequenza parentale o con le caratteristiche della generazione precedente. Il caso migliore per lo sperimentatore è quando la proteina ricercata è fluorescente o luminescente, ha quindi caratteristiche che possiamo facilmente valutare. Non sempre è così, per cui possiamo affidarci a dei reporter per valutare più facilmente l'attività dell'enzima.

Quello che di solito cerchiamo è capire se la variante dell'enzima che analizziamo è attiva su un particolare substrato in determinate condizioni sperimentali. Cerchiamo quindi la trasformazione del substrato in prodotto. Il problema è che non tutti i prodotti sono facilmente rilevabili e misurabili.

Molto spesso invece che usare il vero substrato dell'enzima useremo un substrato simile, che ragionevolmente viene processato con la stessa reazione, ma il cui prodotto è misurabile facilmente, da ad esempio un segnale colorimetrico.

Si usano spesso substrati sintetici che rappresentano un modello su cui studiare l'attività. Dobbiamo comunque alla fine accertarci che l'enzima sia attivo su substrato su cui dovrà operare durante il processo reale.

Ci sono diverse tecniche per valutare le varianti enzimatiche più utili.

Un metodo è usare strumenti robotizzati, che permettono di lavorare con molti campioni in parallelo.

Ciò che conviene sempre fare è utilizzare metodi semplici.

Two step screening

Consiste nel fare una prima valutazione delle varianti col metodo più semplice a disposizione.

Va tenuto presente che quando si fa mutagenesi random, molte varianti non sono attive o sono meno attive del Wild Type.

Quindi il primo passo è fare un primo passaggio di screening che permette di rimuovere le varianti subottimali.

Il secondo passaggio è una valutazione specifica delle proprietà delle varianti, che a questo punto saranno diventate molte di meno.

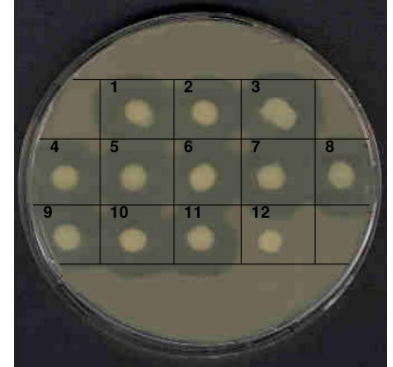
Esempio

Cerchiamo di modificare le proprietà un enzima appartenente alla classe delle *lipasi*, enzimi che idrolizzano i trigliceridi. In generale vogliamo alterare la specificità di substrato dell'enzima.

Il primo passaggio di screening è valutare quali sono le varianti attive. Possiamo usare la *tributirina* come substrato, aggiungendola ad esempio al terreno di coltura dell'organismo carrier.

La tributirina dà al terreno di coltura un aspetto opaco. Quando il clone produce una lipasi attiva, questa liserà il substrato, dando al terreno attorno alla colonia un alone di chiarificazione.

Dopo aver selezionato le proteine più attive rispetto al WT, ne potremo andare a studiare le caratteristiche in modo approfondito.



Quando si vogliono produrre ed analizzare migliaia o decine di migliaia di varianti, alcuni autori hanno studiato metodi alternativi.

Phage Display

Viene realizzata creando una proteina di fusione tra peptide del capsido virale e l'enzima di cui si vuole testare l'attività.

La variante enzimatica verrà esposta sulla superficie del fago. Questo permette di avere un'ampia quantità di cloni testabili.

Ciò che si fa dopo è l'isolamento delle varianti che mostrano le caratteristiche di interesse. Questa tecnica è stata utilizzata per la produzione di anticorpi, che vengono testati per l'affinità all'antigene, il quale può essere immobilizzato su una colonna cromatografica di filtrazione.

Lo stesso si può fare con gli enzimi. Usiamo una colonna con una matrice di immobilizzazione, nella quale abbiamo immobilizzato un substrato con cui vogliamo le nostre varianti enzimatiche reagiscano, tipicamente un *Transition State Analogue* (TSA), che è un substrato analogo a quello dell'enzima, ma che *non viene modificato dalla reazione enzimatica*.

Facciamo quindi scorrere la soluzione contenenti i fagi con le proteine di fusione nella colonna. Gli enzimi che hanno più affinità per il substrato resteranno legati alla colonna, mentre gli altri verranno eluiti.

Alla fine della filtrazione possiamo clivare il legame tra proteina del capsido ed enzima, in modo da poter estrarre solo l'enzima.

La stessa operazione si può fare in maniera indiretta. Si fissa alla colonna la *streptavidina*, si inserisce poi il substrato dell'enzima attaccato alla *biotina*. La biotina legherà la streptavidina, mentre il substrato sarà legato dalla proteina di fusione capsido-enzima, nel momento in cui la variante enzimatica abbia una buona affinità col substrato.

Questo metodo ha mostrato dei problemi riguardo alla dimensione degli enzimi che si possono esporre sulla superficie del fago.

Risultati positivi si sono ottenuti con la *penicillina acilasi*, con la *lipasi A* e con alcuni enzimi industriali.

Yeast Display

È stato sperimentato negli anni scorsi l'espressione di proteine di fusione sulla superficie del lievito. Questo facilita la produzione di proteine da eucarioti.

Il concetto è simile. Si fonde la sequenza dell'enzima di interesse con una proteina che si localizza sulla superficie esterna della cellula di lievito. Spesso sono proteine che mediano l'adesione tra cellule.

Per purificare ed isolare le sequenze, si associa alla tecnica FACS.

FACS è un metodo di citometria a flusso, che permette di separare le cellule in base a caratteristiche di fluorescenza. La fluorescenza è ottenuta mettendo dei *tag* nella proteina di fusione. Questi tag sono riconosciuti da anticorpi specifici legati a *label fluorescenti*. Questa tecnica è stata applicata a diverse proprietà delle proteine, ad esempio alle interazioni proteina-proteina.

Microfluidica

Ogni variante viene prodotta in maniera isolata attraverso la formazione di microdroplets in dispositivi microfluidici. Ogni droplet funge da singolo bioreattore, contenente una singola variante proteica, il che facilita lo screening delle varianti, specialmente con substrati fluorescenti.

Questi metodi hanno vantaggi e svantaggi, e ne vanno scelti in base alle caratteristiche dell'enzima ed alle capacità del nostro laboratorio. I più usati sono i metodi di screening con saggi di attività, come il metodo two-step.

Come introdurre varietà di sequenza

Ci sono diversi metodi utilizzati per generare la varietà nella sequenza.

Per eseguire la mutagenesi random inizialmente si usavano metodi chimici e fisici per generare mutazioni, oppure l'uso di ceppi mutatori, cioè ceppi di microorganismi che introducono parecchie variazioni nella sequenza durante la replicazione.

Queste tecniche non vengono utilizzate, perché soffrono di una bassa frequenza di mutazione e non sono facili da utilizzare.

Un metodo oggi molto usato è la epPCR (*PCR error prone*).

Altri metodi sfruttano la ricombinazione omologa, ad esempio lo *shuffling di DNA*.

Si può anche usare la *mutagenesi a saturazione*, che si applica quando si vuole sostituire un amminoacido con tutti gli altri aminoacidi. Questo non si fa sull'intera sequenza, ma su regioni di particolare importanza, oppure su uno o pochi aminoacidi che sono stati identificati come molto importanti per una particolare proprietà. Queste regioni sono chiamate *hot-spot*.

Si sono sviluppati anche metodi computazionali, che risultano molto importanti per identificare regioni di particolare importanza nella funzione della proteina, e risultano utili ad indirizzare il ricercatore nelle seguenti fasi di laboratorio.

epPCR

La PCR consiste nell'amplificazione di sequenze target di DNA. La sequenza target viene denaturata, e si effettua l'amplificazione per mezzo di primer, DNA polimerasi e nucleotidi in soluzione.

La PCR ha varie funzioni. La principale è sicuramente l'amplificazione del DNA.

Viene utilizzata anche per introdurre *mutazioni sito specifiche*. A questi scopi si usano DNA polimerasi molto fedeli, cioè che amplifichino il DNA in modo da inserire il minor numero di errori rispetto alla sequenza originale.

Allo scopo dell'evoluzione guidata serve introdurre mutazioni. Cercheremo quindi di indurre la DNA polimerasi a compiere errori durante il processo di amplificazione.

Abbiamo due alternative:

- Usiamo una DNA polimerasi fedele, ma la mettiamo in condizioni di replicazione non ottimali. Le polimerasi richiedono precise concentrazioni di ioni e nucleotidi nelle condizioni sperimentali. Quando cambiamo queste condizioni la polimerasi tende ad introdurre nucleotidi non complementari durante la replicazione.
- Usiamo DNA polimerasi poco fedeli. Un enzima di questo tipo è Mutazyme (Agilent Technologies)

In entrambi i metodi il numero di mutazioni inserite può essere controllato dallo sperimentatore. Ne primo caso la frequenza di mutazioni può essere più o meno sbilanciata. Mentre nel caso di Mutazyme la regolazione va fatta in un altro modo.

Sappiamo che la PCR si esegue in cicli. Più cicli vengono fatti con le polimerasi error prone, più mutazioni verranno introdotte.

Con enzimi simili a Mutazyme controlliamo la frequenza di mutazioni regolando la concentrazione del template. Se forniamo poco template all'inizio della reazione, faremo più cicli e quindi introdurremo più mutazioni.

La domanda è: quante mutazioni facciamo?

Molti laboratori preferiscono lavorare con basse frequenze di mutazione. Ad ogni ciclo si cerca di non alterare più di tre triplette, sostituendo quindi al massimo 3 aminoacidi per ciclo di mutagenesi. Questo approccio prudente è usato perché si ritiene che introdurre molte mutazioni tutte insieme possa alterare molto la funzionalità e la stabilità della proteina, per cui si preferisce procedere in maniera incrementale.

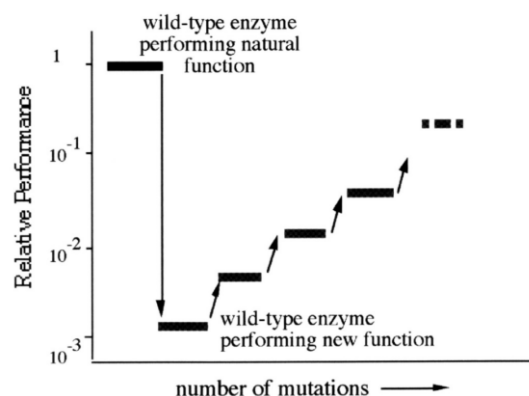
C'è anche chi ha provato ad introdurre molte mutazioni nelle sequenze. Il risultato è che la gran parte delle varianti così ottenute non sono funzionali.

Inserire mutazioni in modo graduale permette di controllare come l'enzima reagisce a poche mutazioni alla volta, e di individuare quelle con maggior potenziale.

Generata la variabilità si esprime l'enzima e si effettua lo screening. Scegliamo poche varianti promettenti (una o due) e possiamo eseguire un nuovo ciclo di mutagenesi, in cui nuove mutazioni si sommano alle mutazioni ottenute precedentemente.

Immaginiamo di voler evolvere un enzima che funziona bene per una funzione, e noi vogliamo introdurre una nuova funzione in questo enzima. All'inizio l'enzima parentale non performa bene per la nuova funzione. In seguito a cicli di mutagenesi e screening cerchiamo un aumento graduale delle prestazioni nella nuova funzione.

La epPCR è molto usata, perché relativamente semplice da usare. Inoltre permette di limitare il numero di varianti con cui lavoriamo. In genere al primo ciclo di mutagenesi considereremo 2000 o 3000 cloni da analizzare. Dopo lo screening selezioneremo il clone migliore su cui effettuare un altro ciclo di mutagenesi.



La epPCR ha anche degli svantaggi. La prima è che ad ogni ciclo miglioriamo di poco le prestazioni, perché sostituiamo appunto due o tre aminoacidi alla volta rispetto alla sequenza di partenza.

Un secondo svantaggio è che non riesce a coprire tutto lo spazio di sequenza. Questa tecnica non è del tutto random, ma ha dei bias nell'inserimento dei nucleotidi.

Dobbiamo considerare come sono composte le triplette. Il cambiamento di un solo nucleotide nella sequenza di DNA dà accesso ad un numero limitato di sostituzioni amminoacidiche, per via della degenerazione del codice genetico. Una singola sostituzione nucleotidica in una tripletta può dare un numero limitato di sostituzioni amminoacidiche, solo 7 ad esempio su una tripletta che codifica per l'*asparagina* (AAC).

In una tripletta che codifica per l'*arginina* (CGA) una sostituzione può dare al massimo 4 sostituzioni amminoacidiche.

Sappiamo che in ogni giro di mutagenesi è molto improbabile che la polimerasi error prone inserisca più di una mutazione. È statisticamente improbabile che in un secondo ciclo una seconda base della tripletta con già una mutazione venga sostituita.

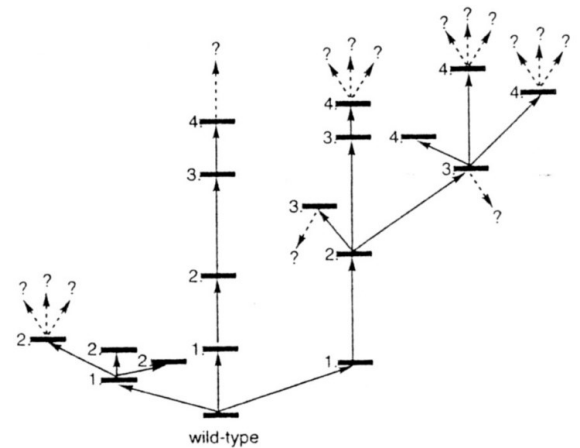
Quindi ci è precluso l'accesso a tutte le possibilità di mutazione per un singolo aminoacido.

Molte delle mutazioni inserite sono silenti, per via della degenerazione del codice genetico.

Abbiamo inoltre il rischio di una retromutazione.

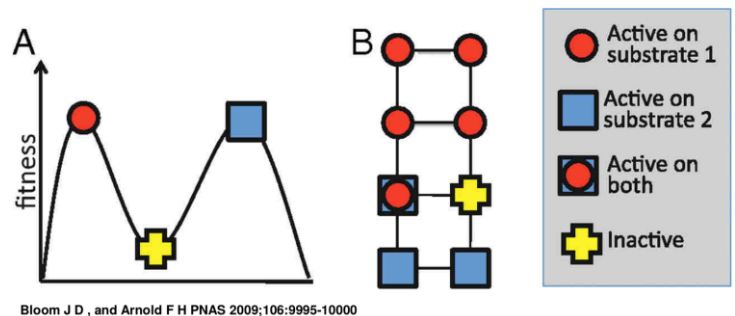
Un terzo svantaggio è che ad ogni ciclo si scelgono una o due varianti e si continua da lì. In questo modo si rischiano di perdere delle mutazioni che effettivamente avrebbero potuto dare vantaggi se sommate a mutazioni successive. Lo sperimentatore quindi si pone dei vincoli, perché sta indirizzando l'evoluzione in una direzione.

La epPCR resta comunque il metodo più utilizzato nei laboratori, perché semplice da utilizzare e perché comunque produce dei buoni risultati.



Ipotizzando di fare un esperimento di evoluzione guidata, partiamo da un enzima con caratteristiche ottimali per una certa attività, ad esempio la specificità per un substrato, e vogliamo ottenere una specificità per un altro substrato. Può succedere che il nostro enzima nel passaggio alla nuova funzione deve passare ad uno stato in cui l'attività è nulla o minima.

Con la epPCR noi potremmo perdere lo stato ad attività ridotta, ma con ottime potenzialità di evoluzione, perché selezioniamo solo le varianti con buona attività.



DNA Shuffling

Anche lo shuffling prende ispirazione di ciò che succede in natura. Nel corso dell'evoluzione, spesso per trasposizione, abbiamo la miscelazione di sequenze, spesso esoni, provenienti da diverse proteine, che hanno dato luogo a sequenze chimeriche con attività nuove o migliori.

In questa tecnica si usano sequenze omologhe, spesso da organismi diversi. È necessario che le sequenze abbiano almeno il 70% di omologia di sequenza.

Ottenute le sequenze queste vengono frammentate in segmenti. In questa tecnica **non** usiamo enzimi di restrizione per rompere il DNA, perché questi tagliano in posizioni specifiche, il che ci impedirebbe di ricostruire la sequenza nella sua interezza.

Possiamo usare metodi casuali, ad esempio una DNAsi non specifica, oppure mezzi fisici, come ultrasuoni.

Dopo la frammentazione avremo una miscela di frammenti di geni, da cui dovremo ricostruire una sequenza. Per riassemblare il gene si usa la PCR.

Tutte le sequenze di partenza hanno un certo grado di identità. Nel primo passaggio dell'amplificazione andremo a denaturare i nostri frammenti di sequenza, per poi fare l'annealing. In questo caso ci saranno eventi di self annealing tra sequenze abbastanza simili da legarsi. La PCR permette di completare zone che sono rimaste a singola catena, sulla base del template a disposizione.

Il self annealing permette quindi di ricostruire la sequenza completa.

Alla fine del processo si farà un passaggio di PCR con primer esterni per amplificare il gene completo, che sarà poi estratto.

Non sempre ci troviamo ad avere molte sequenze omologhe per un enzima. In questi casi possiamo noi generare la variabilità, facendo precedere lo shuffling da un passaggio di mutagenesi random con epPCR.

Il grafico dell'evoluzione è decisamente diverso rispetto all'epPCR, perché nelle generazioni successive si avranno molte combinazioni diverse di sequenza, che andranno poi valutate.

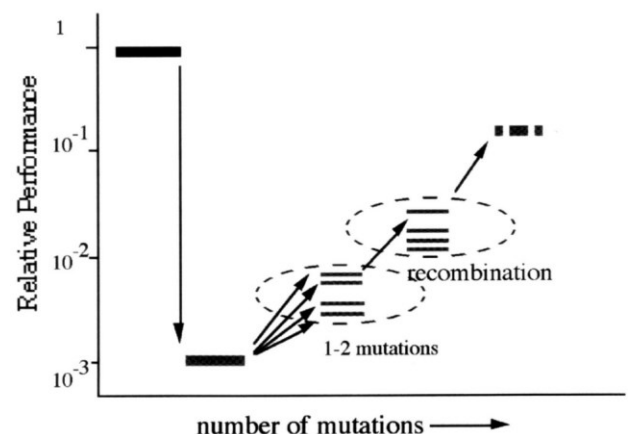
Il miglioramento delle proprietà con questo metodo è a salti. Si ha molta più variabilità nella sequenza e si può esplorare più a fondo lo spazio di sequenza.

Tra gli svantaggi dello shuffling troviamo la richiesta di avere sequenze codificanti ad alta identità di sequenza.

Bisogna analizzare molti cloni. Lo shuffling produce

tantissime varianti non attive, che vengono eliminate dai processi di screening.

Si ritrovano molte sequenze parentali, perché sequenze parentali tendono a fare annealing tra loro.



Anche lo shuffling è un metodo che funziona bene ed usato in molti laboratori.

Per ridurre la frequenza di parentali trovati nello shuffling si sono adottati vari metodi.

Uno di questi è il metodo StEP (Staggered Extension Process).

È un metodo ricombinativo, come lo shuffling, ma non richiede che le sequenze siano frammentate.

La chiave del suo funzionamento è lo *switch di template*.

Contrariamente alla epPCR, qui usiamo polimerasi ad alta fedeltà e primers esterni. Il processo inizia con la reazione di PCR, in cui il primer inserito dall'esterno si lega al template e permette alla polimerasi di iniziare la replicazione.

La reazione di allungamento viene interrotta prima di essere completata (qualche secondo). Si denatura quindi il DNA e si fanno riappaiare le sequenze, per poi farle riappaiare e ricominciare il ciclo. Nella fase di annealing la sequenza può anche appaiarsi ad un template diverso da quello di origine, generando variabilità di sequenza nel momento in cui questo si appaia a nuove sequenze. Alla fine avrò delle sequenze chimeriche.

Anche questo metodo è molto usato. Generalmente si fa partendo da poche sequenze parentali alternative.

Una limitazione è sull'omologia delle sequenze che possono apparirsi. Servono un minimo di circa 30 paia di basi per avere un annealing abbastanza robusto.

Synthetic Shuffling

Permette di evitare di usare sequenze parentali.

Assembly of designed oligonucleotides

Queste metodiche permettono di assemblare sequenze di oligonucleotidi disegnate in laboratorio, senza avere geni di partenza.

Si possono quindi usare per l'amplificazione oggetti che non sono sequenze geniche, ma sono oligonucleotidi sintetizzati in laboratorio, che in parte si appaiano tra di loro.

Questa tecnica è diventata popolare con i processi di sintesi di oligonucleotidi, e con la loro riduzione nei costi.

In una prima fase si disegnano gli oligonucleotidi.

Dal punto di vista bioinformatico posso scegliere delle sequenze geniche che desidero mescolare. Queste dovranno essere sequenze omologhe. Ci saranno comunque delle differenze tra queste sequenze, ma ci saranno anche zone altamente conservate.

Disegniamo quindi primer che coprono parte della sequenza. Nelle zone variabili della sequenza inseriremo mutazioni, mentre preserveremo le zone conservate tra i geni da mescolare.

Le sequenze oligonucleotidiche disegnate devono parzialmente sovrapporsi. Si usa la PCR, prima senza primer per l'assemblaggio degli oligonucleotidi, ed in fine con primer esterni per amplificare la sequenza.

In questo metodo serve una conoscenza un po' più approfondita delle sequenze rispetto ad altri metodi.

Ci sono varie implementazioni dello Shuffling. Una di queste è RACHITT (Random Chimeragenesis on Transient Templates).

Uno dei problemi dello shuffling è che rimangono molte sequenze parentali.

In RACHITT si parte da sequenze ad omologia inferiore rispetto allo shuffling classico. Queste vengono frammentate, ma non vengono riassemblate tramite PCR.

I frammenti single strand si legano ad un template transitorio single strand contenente uracile.

Alcune sequenze si legheranno bene, altre non si assoceranno. Queste regioni non associate sono dette *flap*, che verranno degradate da delle nucleasi.

I gap sono riempiti per mezzo della PCR. Il template artificiale viene degradato da opportuni enzimi.

Questa tecnica aumenta la variabilità di sequenza e riduce l'influenza delle sequenze parentali.

Non è semplice da mettere in pratica.

Si stanno sviluppando altre tecniche che permettono di mescolare sequenze con identità molto al di sotto del 70%.

Sono stati sviluppati anche metodi che permettono di cambiare l'ordine dei frammenti nella sequenza finale.

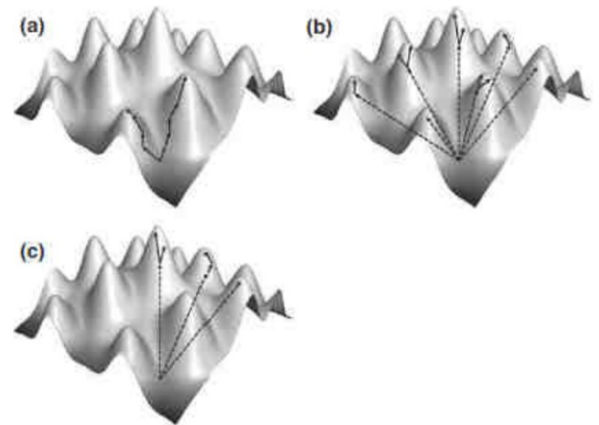
Migliorare l'efficienza

Il *fitness landscape* è una ipersuperficie caratterizzata da picchi e valli. I picchi sono le sequenze più adatte alla caratteristica che stiamo studiando, mentre le valli sono popolate da sequenze con proprietà meno adatte rispetto a quelle che stiamo cercando.

Nella epPCR (a) partiamo da una sequenza parentale, nella quale vengono introdotte poche mutazioni dopo ogni passaggio di mutagenesi. Selezioniamo il mutante migliore e procediamo ad un altro ciclo di mutagenesi. Risaliamo quindi un singolo picco dell'ipersuperficie; per questo si dice che la epPCR è indirizzata.

Nello Shuffling (b) partiamo da un gruppo di sequenze che vogliamo mescolare tra loro; potremo da queste ottenere proteine evolute di tanti tipi diversi, che esplorano il landscape in misura maggiore, e potrebbero avere proprietà di nostro interesse su picchi diversi.

Nella mutagenesi a saturazione (c) abbiamo già migliorato la proteina con altri metodi, e ci siamo accorti che ci sono regioni che hanno un rilievo maggiore per la funzione che cerchiamo. Con la mutagenesi a saturazione in queste regioni potremmo riuscire a migliorare ulteriormente le prestazioni dell'enzima.



Spesso negli esperimenti di evoluzione guidata si usano metodi diversi in combinazione. In genere la prima fase è random ed aiuta a trovare gli hot-spot, sui quali si concentreranno poi strumenti mirati per evolvere la sequenza.

Spesso la prima parte del lavoro può essere fatta con metodi computazionali, in modo da ottenere poi librerie di varianti più piccole e quindi agevoli da analizzare. In alcuni casi comunque la mutagenesi random può tornare molto utile.

Fatti da attenzionare

L'accumulo di mutazioni produce proteine non funzionali.

Può succedere con tutti i metodi di mutagenesi. Può succedere che una o più sostituzioni amminoacidiche portano a proteine che hanno difficoltà a ripiegarsi.

L'evoluzione guidata molto difficilmente porta alla selezione finale di proteine con modifiche al sito attivo, perché il sito attivo è la parte più conservata e più delicata dell'enzima, che tollera meno le alterazioni rispetto al resto della struttura. Mutazioni sul sito attivo spesso rendono l'enzima non funzionale, e per questo non superano i passaggi di screening e selezione.

Mutazioni positive possono andare perse nei successivi passaggi di mutagenesi.

Questo avviene più spesso con lo shuffling, quando varianti dell'enzima si mischiano e possono andar perse alcune proprietà interessanti.

C'è un trade-off tra stabilità ed attività?

Può essere che mutazioni positive per l'attività cercata portino alla destabilizzazione della proteina.

La mutagenesi può essere mirata a regioni specifiche, in cui c'è un'alta probabilità di influenzare la proprietà di interesse.

Approccio semirazionale

Spesso noi abbiamo a disposizione informazioni strutturali, biochimiche e di sequenza sull'enzima di interesse.

Ad esempio possiamo vedere che nella famiglia di proteine a cui appartiene l'enzima che stiamo studiando ci sono membri che, rispetto alla caratteristica che ci interessa, hanno mutazioni concentrate in una determinata zona.

Nei casi in cui abbiamo la struttura cristallografica della proteina possiamo andare a valutare il B-factor, che è un parametro cristallografico che ci dà informazioni sulla mobilità delle porzioni della proteina. Zone ad elevato B-factor sono spesso il target di mutagenesi mirata quando l'obiettivo è stabilizzare la proteina.

Dall'articolo di Luts (Beyond directed evolution - semi-rational protein engineering and design) vediamo che usando un sistema misto di razionalità e casualità possiamo arrivare a ridisegnare completamente le proprietà degli enzimi. Si riesce a cambiare la specificità di substrato della proteina, ridisegnando ad esempio la zona di legame al sito attivo. Questo intervento drastico ha portato alla perdita di funzionalità dell'enzima. L'enzima era un dimero, che non era stabile col nuovo sito attivo. Su questa nuova sequenza sono stati fatti esperimenti di evoluzione guidata che hanno permesso di introdurre stabilità e formare il dimero, rendendo l'enzima funzionante.

Un altro metodo consiste mischiare domini tra enzimi omologhi. Mischiare domini dà a volte problemi, perché i vari domini possono non essere compatibili dal punto di vista topologico. Questo problema si può risolvere con metodi computazionali, che permettono di predire quali proteine chimeriche possono essere funzionali.

Sostituzioni che migliorano l'attività ma peggiorano la stabilità

Il cambiamento funzionale che si può ottenere in una proteina dipende dalla sua stabilità iniziale. C'è una differenza di energia che separa la proteina nativa dalla proteina denaturata. Questa differenza non è tanta.

Quando lavoriamo su una proteina con un fold stabile questa avrà un'ampia differenza di energia tra lo stato nativo e lo stato denaturato. Se inseriamo una mutazione, con effetto positivo sulla funzione ma destabilizzante, e la barriera energetica per l'unfolding è alta, la proteina continua a funzionare.

Se la nostra proteina di partenza era al limite della stabilità, l'inserimento della stessa mutazione sposterà la proteina verso l'unfolding. Quindi una mutazione che in generale avrebbe potuto migliorare la funzione non viene selezionata, perché destabilizza la proteina e la rende non funzionale. Se la proteina fosse stata stabile avremmo avuto un miglioramento di funzione, ma nella proteina instabile perdiamo l'informazione sulla sostituzione utile.

Evolvability

Dobbiamo capire se una proteina può essere evoluta oppure no. Cioè se questa può accettare le mutazioni positive nella sua struttura, mantenendo la struttura e l'attività catalitica. Spesso la robustezza ad accettare mutazioni è correlata alla stabilità della proteina. Fold più stabili sopportano meglio interventi di evoluzione.

Come si migliora l'evolubilità della proteina? (Leggere articoli Tawfik)

Neutral drift

Le mutazioni vengono introdotte in modo random senza introdurre passaggi di selezione. In questo modo creiamo un background che mostra una diversità superiore.

Stiamo quindi facendo accumulare mutazioni neutrali.

In passaggi successivi queste mutazioni possono essere selezionate, per vari motivi.

Ad esempio alcune mutazioni rendono la proteina più stabile in alcune condizioni, oppure possono rivelarsi utili per stabilizzare la proteina a seguito di effetti destabilizzanti di mutazioni mirate. Questo passaggio di evoluzione neutra permette di avere più flessibilità nell'accettazione di cambiamenti nella sequenza.

Co-espressione di chaperoni

Questa è una tecnica indirizzata ad agevolare il folding della proteina.

So che interventi di mutagenesi destabilizzano la proteina. Se si esprimono le varianti in sistemi ospite che esprimono chaperoni, i quali permettono di evitare l'aggregazione di proteine poco stabili.

Catalisi di nuovi substrati

Sappiamo che nell'evoluzione di un enzima ci possono essere fasi intermedie in cui la proteina perde la sua attività. Per esempio se voglio passare da una specificità ad un'altra potrei partire da un enzima molto specifico per un substrato. In seguito all'evoluzione questo enzima può poi perdere parte della sua specificità e ad esempio ampliare il tipo di substrati che accoglie. In seguito con l'accumulo di altre mutazioni l'enzima potrebbe evolvere una nuova specificità di substrato. Questo fenomeno si è molto probabilmente presentato nel corso dell'evoluzione naturale delle sequenze odierne.

Questo fenomeno è alla base del processo di promiscuità enzimatica, per cui si è visto che molti enzimi oltre all'attività principale hanno anche altre possibili attività, sia per quanto riguarda i meccanismi catalitici che la specificità del substrato. La promiscuità può essere utilissima per evolvere nuovi enzimi, perché funge da base su cui impiantare nuove attività.

La promiscuità potrebbe essere utile per la catalisi per la nuova chimica, che usano enzimi per realizzare reazioni che gli enzimi naturali non fanno.

Negli ultimi anni alcuni hanno cercato di fare un'analisi evolutiva di una famiglia di proteine, cercando di risalire ai primi enzimi di questa famiglia. Questo perché si pensa che questi primi enzimi fossero più robusti (capaci di accettare nuove mutazioni) e più generalisti (ampia varietà di substrati, con possibilità di specializzazione), quindi migliori candidati ad essere evoluti.