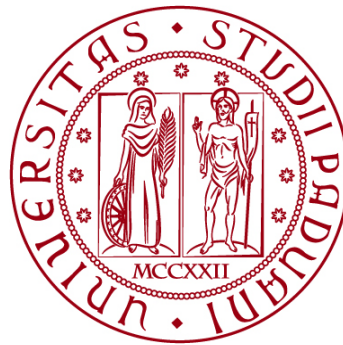


**UNIVERSITÀ DEGLI STUDI DI PADOVA**

**DIPARTIMENTO DI BIOLOGIA**

**Corso di Laurea in Biotecnologie**



**ELABORATO DI LAUREA**

**DESIGN COMPUTAZIONALE DI ENZIMI  
PER LA BIOCATALISI**

**Tutor: Prof. Francesco Filippini  
Dipartimento di Biologia**

**Laureando: Marco Filosi**

**ANNO ACCADEMICO 2021/2022**

# ***Indice***

## ***Introduzione***

### **Capitolo 1 - Economia circolare e biocatalisi**

- 1.1 Il passaggio da economia lineare a circolare
- 1.2 Biocatalisi

### **Capitolo 2 – Biometano**

- 2.1 Caratteristiche principali della metanogenesi
- 2.2 Reazioni enzimatiche della metanogenesi
- 2.3 Fisiologia degli archei metanogeni
- 2.4 Metanogenesi inversa per la produzione di Bio-GTL

### **Capitolo 3 - Design computazionale degli enzimi**

- 3.1 Reattività biochimica
- 3.2 Legame con il ligando
- 3.3 Tunnel, canali e trasporto del ligando
- 3.4 Dinamica proteica
- 3.5 Identificazione dell'hotspot e smart library design
- 3.6 Mutagenesi

### **Capitolo 4 - Applicazioni del design computazionale sugli enzimi**

- 4.1 Progettazione di biocatalizzatori efficienti
- 4.2 Progettazione di enzimi con attività nuove
- 4.3 Progettazione di enzimi specifici ed enantioselettivi
- 4.4 Design di enzimi stabili e solubili
- 4.5 Design di enzimi con flessibilità adattata

## ***Conclusioni***

## ***Bibliografia e sitografia***

## INTRODUZIONE

La tesi è incentrata sulla biocatalisi nell'ambito dell'economia circolare, indagando il miglioramento produttivo degli enzimi tramite un approccio computazionale. In particolare, ho approfondito la produzione di biometano a partire da sostanze di scarto come i rifiuti organici o gli scarti dell'industria zootecnica.

Questa opportunità si inserisce in un contesto mondiale in cui i cambiamenti climatici e la progressiva diminuzione delle risorse fossili presentano una grande minaccia per gli equilibri globali. Inoltre, più concretamente, nel primo trimestre 2022 abbiamo sperimentato a livello nazionale un notevole aumento del prezzo del gas (secondo ARERA l'aumento è del 94% rispetto all'anno precedente), che può essere potenzialmente impattante sul bilancio economico di famiglie ed imprese.

In questo quadro, è importante mirare ad una produzione di gas con due caratteristiche. La prima è che sia parte di un modello di economia circolare, con anche un'attenzione alla riduzione dell'emissione di gas ad effetto serra. Secondariamente, è importante la possibilità di produrre a livello locale per non essere soggetti alla variazione del prezzo dovuta a dinamiche internazionali di domanda e offerta oppure geopolitiche.

La tesi spazierà da una parte iniziale di introduzione all'economia circolare e alla biocatalisi, andando poi ad approfondire gli elementi della produzione del biometano: dai vantaggi della sua produzione fino alle vie metaboliche coinvolte. Infine, verranno elencate le principali tecniche utilizzate per il design computazionale degli enzimi e le loro applicazioni.

# 1. ECONOMIA CIRCOLARE E BIOCATALISI

## 1.1 Il passaggio da economia lineare a circolare

Il modello economico lineare, protagonista del grande aumento di produzione su scala globale nell'ultimo secolo, dev'essere ripensato a causa della sua insostenibilità. Infatti, l'utilizzo di questo modello produce:

- Consumo eccessivo di risorse non rinnovabili. Ad esempio, si stima che i giacimenti di petrolio potrebbero esaurirsi in qualche decina di anni;
- Generazione di rifiuti tossici. Questi rifiuti sono principalmente prodotti dall'industria chimica, ad esempio i componenti organici persistenti (POP), ossia dei composti organici che resistono molto alla decomposizione. Questi, rappresentano una minaccia per la salute, potendo causare cancro, malattie croniche, difetti nello sviluppo e, nei casi più estremi, la morte;
- Produzione di gas serra. Il dato più rilevante è rappresentato dalla concentrazione di CO<sub>2</sub> nell'atmosfera, la quale è passata da 280 ppm dell'era preindustriale fino all'attuale superamento di 410 ppm.

In particolare, gli ultimi due punti sopraelencati sono i protagonisti dell'inquinamento ambientale e del riscaldamento globale, il quale ha effetti sia sull'ambiente, sia sull'uomo. I principali effetti sull'ambiente sono: distruzione degli habitat con conseguente perdita della biodiversità; intensificazione degli eventi meteorologici estremi; e scioglimento dei ghiacciai. Dall'altra parte, le conseguenze sull'uomo sono: un impatto negativo sulla salute; migrazioni dovute alla siccità e alla sommersione di aree abitate a causa dell'innalzamento del livello del mare; e manifestazione di malattie infettive emergenti, ad esempio a causa della comparsa di insetti come vettori virali in regioni in cui non erano presenti precedentemente.

Per mantenere una notevole capacità produttiva che garantisca sia un mantenimento dell'alto tenore di vita, sia una riduzione dei danni causati dall'attività umana prima che diventino irreversibili, è quindi necessario superare il modello dell'economia lineare proponendone uno di tipo circolare. Nel modello economico circolare, lo scarto generato dalla produzione o il consumo viene riutilizzato come substrato di partenza per iniziare un nuovo ciclo produttivo. Questo elemento ha un duplice effetto: sostituisce il consumo di nuove risorse ed evita i costi di smaltimento od un potenziale inquinamento dato dalla sua dispersione



Figura 1 - Confronto tra economia circolare e lineare

nell'ambiente (come nel caso dell'inquinamento degli oceani dovuto alla dispersione della plastica) [8].

All'interno del panorama dell'economia circolare, le biotecnologie giocano un ruolo di primo piano tramite la biocatalisi.

## **1.2 Biocatalisi**

La biocatalisi consiste nell'utilizzo di cellule (biocatalisi whole-cell) o enzimi per catalizzare reazioni chimiche. La biocatalisi whole-cell può fornirsi di microrganismi come batteri e funghi, mentre gli enzimi vengono purificati dopo essere stati prodotti da cellule. Negli ultimi anni, l'ingegneria metabolica e la biologia sintetica sono state usate per massimizzare la produttività di queste due tecniche ed in futuro sarà opportuno agire su altri fattori come la stabilità genetica, l'espressione genica e la tossicità del substrato o del prodotto.

Rispetto ai catalizzatori chimici, i principali vantaggi degli enzimi per la biocatalisi sono:

- L'essere biodegradabili. Oltre a rappresentare un vantaggio dal punto di vista ambientale, è anche vantaggioso economicamente, poiché permette di evitare un potenziale costo di smaltimento. Al contrario, quest'ultimo può essere necessario nel caso dei catalizzatori chimici, poiché le tracce di metalli nobili nei prodotti non possono superare determinati limiti imposti dalle normative;
- Sono prodotti a partire da risorse rinnovabili e facilmente reperibili. Al contrario, l'utilizzo di metalli catalizzatori è soggetto a forti variazioni del prezzo della materia prima, sia per il costo di estrazione, sia per la competizione con altri tipi di industria [10];
- Un alto tasso di chemo-, regio- ed enantio-selettività in condizioni moderate (pH fisiologico e normali valori ambientali di pressione e temperatura). Queste capacità rivestono una grande importanza nell'industria farmaceutica in quanto, ad esempio, differenti enantiomeri corrispondono ad attività biologiche diverse.

Nonostante questi vantaggi, la biocatalisi ha iniziato ad essere utilizzata più massicciamente solo negli ultimi vent'anni principalmente per due motivi: alcune classi di enzimi in precedenza non erano sufficientemente disponibili a livello commerciale e molti enzimi hanno una bassa efficienza nelle condizioni di temperatura e pH dei processi industriali [8].

## 2. BIOMETANO

Il biometano o gas naturale rinnovabile (GNR), è una fonte di energia rinnovabile con caratteristiche simili al gas di origine naturale, avendo una concentrazione di metano che supera il 90% (tabella 1). Può essere quindi utilizzato per il riscaldamento, come carburante per alimentare i veicoli e per la produzione di energia elettrica. Tra i metodi di produzione giocano un ruolo importante le biotecnologie, che sfruttano il metabolismo anaerobico di alcuni archeobatteri. In questo processo, chiamato metanogenesi o biometanazione, i microrganismi metanogeni producono biogas utilizzando come substrato biomasse derivanti da prodotti di scarto. In seguito, avviene una fase di “upgrading” del biogas durante la quale si rimuovono la CO<sub>2</sub> ed altri contaminanti, aumentando la concentrazione del metano oltre il 95% e permettendo così la formazione del biometano.

<b>Gas composition</b>	<b>Biogas</b>	<b>Biomethane</b>	<b>Natural Gas</b>
<b>Methane</b>	50-75%	94-99.9%	93-98%
<b>Carbon Dioxide</b>	25-45%	0.1-4%	1%
<b>Nitrogen</b>	<2%	<3%	1%
<b>Oxygen</b>	<2%	<1%	-
<b>Hydrogen</b>	<1%	Traces	-
<b>Hydrogen Sulphide</b>	20 – 20,000 ppm	<10 ppm	-
<b>Ammonia</b>	Traces	Traces	-
<b>Ethane</b>	-	-	<3%
<b>Propane</b>	-	-	<2%
<b>Siloxane</b>	Traces	-	-
<b>Water</b>	2-7%	-	-

*Tabella 1- Confronto della composizione di diverse tipologie di gas*

Per la produzione di biogas a livello industriale vengono utilizzati vari substrati:

- I residui del raccolto agricolo di determinate piantagioni, come quelle del riso e dello zucchero di canna;
- Letame animale proveniente dall'industria zootecnica;
- Materiale organico proveniente dai rifiuti urbani, come ad esempio il cibo o l'erba sfalciata;
- Fanghi da depurazione, ossia la parte solida presente nelle acque reflue.

Essendo simile al gas di origine naturale, il biometano ha diversi punti di forza:

- Può essere distribuito mediante le infrastrutture già presenti per il trasporto e lo stoccaggio del gas. Al contrario, il biogas può essere utilizzato direttamente per il suo consumo (ad esempio per produrre energia elettrica), ma non può essere immesso nella rete di trasporto data la sua differente composizione rispetto al gas di origine naturale;
- Si colloca nel modello dell'economia circolare, poiché vengono utilizzati come substrato per la sua produzione dei prodotti di scarto;
- Ha un basso costo di produzione. Infatti, il substrato non ha costi ed inoltre nell'intero processo produttivo non sono richiesti strumenti e tecnologie dispendiose;

- Riduce le emissioni di gas ad effetto serra. Nonostante la combustione del biometano liberi diossido di carbonio, questa rappresenta comunque un danno minore rispetto alla quantità di metano e CO<sub>2</sub> che verrebbe rilasciata nell'atmosfera in seguito alla semplice decomposizione dei substrati mediata dai microrganismi;
- Permette di arginare il fenomeno della saturazione delle discariche, sottraendo gli scarti che fungono da substrato per la produzione di biometano [3].

## 2.1 Caratteristiche principali della metanogenesi

La metanogenesi è un processo di respirazione anaerobica mediante il quale alcune specie di archeobatteri producono metano a partire da materia organica. Questi metanogeni sono anaerobi obbligati e perciò possono essere trovati in habitat privi di ossigeno come nel permafrost o nell'apparato digerente di alcuni animali.

I metanogeni sono classificati in tre gruppi in base al substrato che utilizzano:

- Idrogenotrofici. Ossidano H<sub>2</sub> e riducono CO<sub>2</sub> a CH<sub>4</sub>. Si possono trovare nei sedimenti oceanici e nell'apparato digerente di animali (in particolare ruminanti ma anche negli umani). Si stima che una singola mucca rilasci circa 200 litri di metano al giorno;
- Aceticlastici. Dividono l'acetato per formare CO<sub>2</sub> e CH<sub>4</sub>. Si possono trovare laddove i metanogeni idrogenotrofici riducono a sufficienza il livello di H<sub>2</sub> creando le condizioni adatte alla formazione di acetato in grandi quantità. Questi habitat, come paludi o campi di riso, contribuiscono a due terzi delle emissioni nell'atmosfera di metano generato biologicamente;
- Metilotrofici. Usano composti metilati come il metanolo e possono essere trovati nei sedimenti marini ricchi di solfato [5].

## 2.2 Reazioni enzimatiche della metanogenesi

Gli enzimi e i coenzimi coinvolti nella metanogenesi sono espressi da circa 200 geni. L'ultima reazione del processo è gestita dalla metil-coenzima M reduttasi (MCR), la quale riduce il metil-coenzima M in  $\text{CH}_4$  tramite il coenzima B formando un eterodisolfuro. L'MCR è un enzima molto sensibile all'ossigeno.

Inoltre, tutti i metanogeni hanno una eterodisolfuro reduttasi, la quale ricicla nuovamente

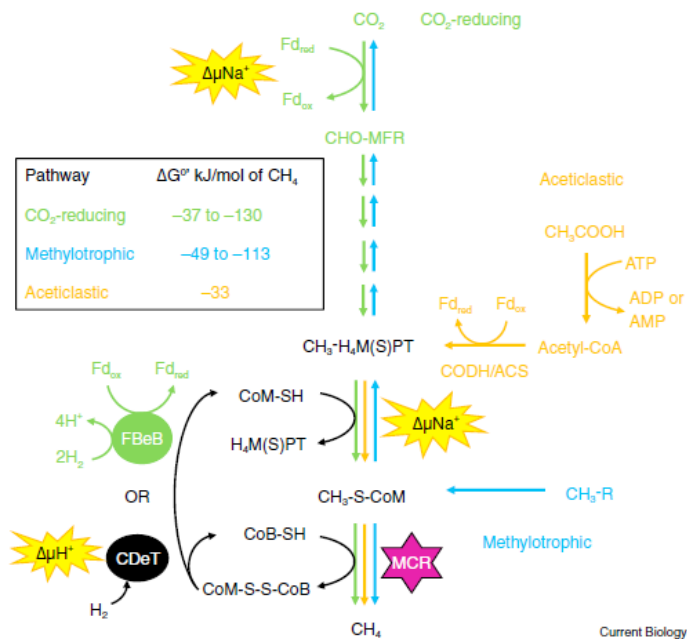


Figura 2 - Vie enzimatiche della metanogenesi

l'eterodisolfuro nei coenzimi B ed M. Ci sono due tipi di eterodisolfuro reduttasi: FBeB, la quale si trova nella maggior parte dei metanogeni idrogenotrofici; e CDeT, che si trova nei metanogeni metilotrofici, aceticlastici ed alcuni idrogenotrofici. La FBeB ha anche il ruolo di formare la feredossina ridotta necessaria alla riduzione della  $\text{CO}_2$ , congiungendosi così al primo passaggio della metanogenesi (ciclo di Wolfe). Dall'altra parte, un altro ruolo della CDeT è quello di formare un potenziale protonico elettrochimico ( $\Delta\mu\text{H}^+$ ), il quale può prendere parte alla riduzione della  $\text{CO}_2$  (tramite la generazione di una feredossina a basso potenziale) o alla sintesi di ATP.

Gli altri passaggi dipendono dal substrato. Nella via idrogenotrofica, inizialmente la  $\text{CO}_2$  si lega al metanofurano (MFR) venendo ridotta a CHO-MFR dalla formilmetanofurano deidrogenasi e una feredossina ridotta. In seguito, avvengono una serie di reazioni che portano alla riduzione a  $\text{CH}_3\text{-H}_4\text{MPT}$ , del quale gruppo metile sarà poi trasferito al CoM-SH. In questo modo, si forma il  $\text{CH}_3\text{-S-CoM}$  che costituisce il punto di incontro fra le tre vie enzimatiche. Il trasferimento del metile genera una forza sodium-motiva ( $\Delta\mu\text{Na}^+$ ), la quale può anche essere utilizzata da alcuni metanogeni metilotrofici per trasferire il metile nella direzione opposta.

Nella via aceticlastica, viene formato Acetil-CoA tramite l'attivazione di acetato con l'ATP. In seguito, l'Acetil-CoA viene diviso formando gruppi carbonile e gruppi metile. Questi due gruppi hanno compiti precisi nel legame con gli enzimi: il gruppo carbonile viene ossidato a  $\text{CO}_2$ , usando come accettore di elettroni la feredossina e congiungendosi alla reazione della metil-coenzima M reduttasi con la riduzione dell'eterodisolfuro; il gruppo metile viene trasferito al  $\text{H}_4\text{MPT}$  od al  $\text{H}_4\text{SPT}$ , dove si congiunge alla via della riduzione della  $\text{CO}_2$ .



Nella via metilotrofica, viene trasferito al CoM-SH il gruppo metile presente in  $\text{CH}_3\text{-R}$ . Quando non è presente  $\text{H}_2$ ,  $\text{CH}_3\text{-S-CoM}$  viene ossidato mentre con la sua presenza viene ridotto a metano.

L'energia libera standard varia in base alla tipologia di metanogenesi. La metanogenesi metilotrofica è la più favorevole, con un  $\Delta G^0$  che varia da -49 a -113 kJ/mole a seconda del substrato. Al contrario, quella della metanogenesi acetoclastica ha spesso un valore assoluto minore ( $\Delta G^0 = -33$  kJ/mole), il che potrebbe essere correlato al basso tasso di crescita di questi metanogeni [5].

### 2.3 Fisiologia degli archei metanogeni

I metanogeni assumono caratteristiche diverse a seconda del phylum di appartenenza.

Gli euryarchaeota metanogeni, ad esempio, possono essere mesofili, termofili o ipertermofili e crescono a pH neutro. Il tempo di raddoppiamento va dai 20 minuti per gli ipertermofili fino alle 2 ore per i mesofili.

Gli halobacterota metanogeni sono invece caratterizzati dall'utilizzo di molti tipi di substrato e possono essere idrogenotrofici, acetoclastici o metilotrofici. Inoltre, molte specie possono crescere in ambienti freddi, come *Methanogenium frigidum* che si riproduce ad una temperatura compresa tra gli 0 e i 17 °C. La velocità di crescita è invece più bassa rispetto agli euriarcheoti [5].

### 2.4 Metanogenesi inversa per la produzione di Bio-GTL

Il rilascio accidentale di metano, soprattutto durante la fase del trasporto, può comportare grossi problemi in termini di liberazione di un gas ad effetto serra nell'ambiente. Da ciò deriva l'importanza di convertirlo in biocarburanti più facilmente trasportabili.

L'ossidazione anaerobica del metano (AOM) può essere mediata dagli archei ANME (ANaerobic MEtanotroph), i quali utilizzano il metano come substrato per produrre acetato. Quest'ultimo può essere convertito in moltissimi composti, come il butanolo che funge da biocarburante. Il processo può essere considerato come una metanogenesi inversa ed è mediato da una MCR diversa da quella presente nei metanogeni, chiamata ANME MCR. Ne consegue che l'ingegnerizzazione di questo enzima può essere fondamentale per aumentare l'efficienza dell'intero processo.

Soo et al. (2016) hanno ingegnerizzato degli archei metanogeni per crescere anaerobicamente ed esprimere una ANME MCR (o aMCR) che gli permettesse di convertire il metano nell'acetato. Questo è stato necessario poiché non sono ancora state isolate colture microbiche pure che permettano la conversione del metano anaerobicamente. I processi anaerobici per la conversione di metano in carburante sono più rilevanti perché garantiscono una più alta energia e resa di carbonio rispetto ai processi aerobici [9].

### 3. DESIGN COMPUTAZIONALE DEGLI ENZIMI

Il miglioramento degli enzimi è fondamentale per aumentare le loro prestazioni in ambienti diversi da quelli nativi. Al fine di un'efficiente produzione su scala industriale, che permetta anche un abbattimento dei costi, è quindi necessario migliorare alcune caratteristiche degli enzimi per una loro maggiore efficienza nel contesto industriale. Queste caratteristiche, come la stabilità enzimatica, verranno trattate nel capitolo 4.

I metodi computazionali per il design di biocatalizzatori riguardano vari ambiti:

- Reattività biochimica
- Legame con il ligando
- Trasporto del ligando e vie di accesso
- Dinamica delle proteine
- Identificazione dell'hotspot e smart library design
- Mutagenesi

#### 3.1 Reattività biochimica

Lo studio del meccanismo di reazione di un enzima gioca un ruolo centrale per l'ingegnerizzazione di determinate caratteristiche come la stabilità, l'efficienza e la specificità. Lo studio avviene tramite la valutazione del profilo energetico nelle varie fasi della reazione enzimatica: 1) enzima e substrato non

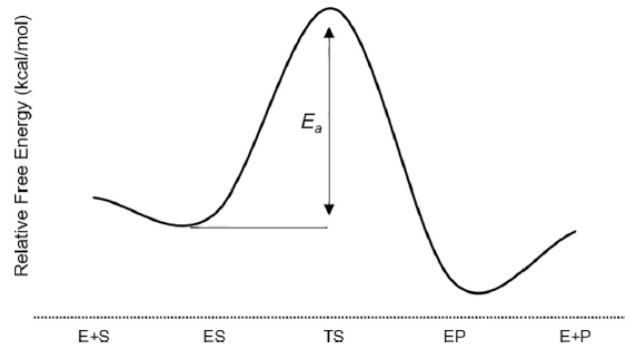


Figura 3 - Cinetica enzimatica ed energia libera

ancora uniti (E + S); 2) formazione del complesso enzima-substrato (ES); 3) stato di transizione (TS); 4) formazione del complesso enzima-prodotto (EP); e 5) liberazione del prodotto dall'enzima (E + P). Poiché il livello di energia libera è inversamente correlato alla stabilità, ne consegue che lo stato di transizione rappresenta il passaggio meno stabile, mentre il complesso EP è il più stabile.

In questo contesto, il design degli enzimi può avere come bersaglio l'energia di attivazione ( $E_a$ ), ossia la differenza tra l'energia libera di ES e di TS.  $E_a$  rappresenta la barriera alla formazione del prodotto a partire dal substrato ed un suo maggior valore si accompagna ad una maggior difficoltà nella formazione del prodotto e ad un maggior tempo richiesto per la sua formazione. L'energia di attivazione può essere diminuita mediante due strategie: abbassamento dell'energia libera del TS e progettazione di un complesso ES con un valore energetico più vicino a quello del TS.

La valutazione dell'energia può essere fatta tramite quattro approcci: l'approccio a cluster di meccanica quantistica; metodi ibridi; Empirical valence bond (EVB); e simulazioni di dinamica molecolare [7].

### 3.2 Legame con il ligando

La predizione delle modalità di interazione tra le proteine e i loro ligandi è molto importante per l'ingegneria proteica, poiché quest'interazione ha spesso un ruolo centrale nella funzione. Ad esempio, può permettere di valutare se un enzima progettato legherà meglio o peggio il substrato.

Il metodo computazionale più utilizzato in quest'ambito è il docking molecolare, il quale permette una rapida identificazione dell'affinità dei ligandi coi loro target e della modalità di legame. Per studiare il legame proteina-ligando è necessario conoscere le loro strutture tridimensionali, preferibilmente tramite via sperimentale (cristallografia, NMR) ma anche teorica (predizione ab initio).

Il processo del docking molecolare può essere definito in due passaggi: 1) la considerazione delle possibili conformazioni di ligando e recettore; 2) la valutazione della qualità del docking.

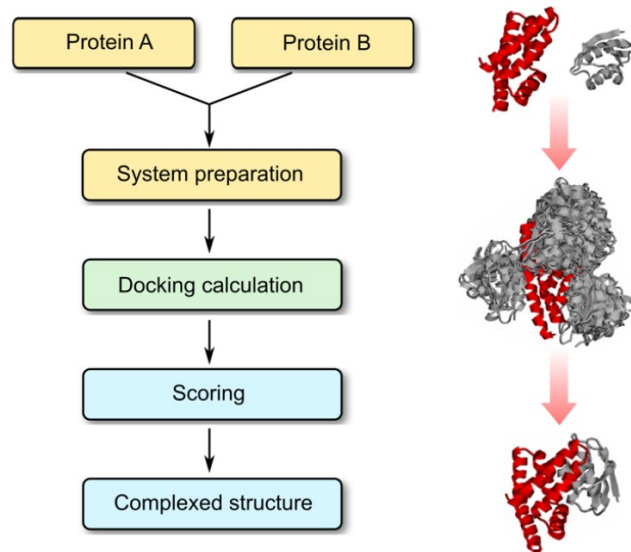


Figura 4 - Processo del docking molecolare proteina-proteina

1. Tradizionalmente, vengono prese in considerazione solamente le possibili conformazioni del ligando tramite vari algoritmi. Tuttavia, tramite il flexible docking è possibile anche applicare lo stesso principio ai recettori. Questo, permette il superamento di una più tradizionale concezione secondo la quale il recettore viene considerato come rigido e solamente il ligando è considerato flessibile, ma aumenta i tempi di calcolo computazionale.
2. Lo scoring serve a valutare le più probabili conformazioni del complesso proteina-ligando e quantificare l'affinità di legame. I metodi di assegnazione del punteggio sono vari e possono essere di natura fisica, statistica o di machine-learning.

Lo screening virtuale è correlato al docking molecolare e nell'ambito dell'ingegneria proteica consente di predire quali proteine, estratte da librerie che ne contengono milioni, si legano più propriamente ad un substrato di interesse. Di conseguenza, permette di definire un gruppo ristretto di candidati evitando il passaggio di screening sperimentale che può essere dispendioso sia in termini economici, sia di tempo [7].

### 3.3 Tunnel, canali e trasporto del ligando

L'analisi delle vie di accesso (tunnel e canali) e della traiettoria di trasporto del ligando permette di migliorare il trasporto di prodotti e substrati tra i siti catalitici e la superficie degli enzimi.

Il folding proteico lascia spazio alla formazione di aree vuote, le quali di solito non hanno una funzione biologica. Tuttavia, talvolta possono avere ruoli importanti come i tunnel ed i canali. I tunnel hanno diverse funzioni: permettono il trasporto verso il sito attivo di prodotti, substrati, cofattori e inibitori; e possono connettere due siti attivi presenti nell'enzima. Dall'altra parte, i canali sono caratteristici delle proteine transmembrana e formano lo spazio per il passaggio di ioni e piccole molecole. Queste importanti funzioni di tunnel e canali, li rendono determinanti per caratteristiche come la specificità di substrato, la termostabilità e l'enantioselettività, le quali possono essere modificate agendo sugli stessi tunnel e canali. Questi ultimi, possono essere identificati con vari software, come ChExVis.

Altri metodi si focalizzano sulla traiettoria del trasporto del ligando, utilizzando tecniche avanzate come l'iterative docking e algoritmi di intelligenza artificiale, oppure simulazioni di dinamica molecolare. Queste ultime verranno descritte nella sezione 3.4 e racchiudono tecniche come il random accelerated molecular dynamics, il quale permette di stimare la posizione di un ligando in seguito all'applicazione di una forza verso una direzione casuale. Tuttavia, le simulazioni di dinamica molecolare presentano svantaggi come l'essere dispendiosi in termini di tempo e la necessità di esperienza da parte degli operatori.

Gli algoritmi di intelligenza artificiale costruiscono in maniera ripetitiva un albero di esplorazione per trovare una via plausibile tra lo stato senza ligando e lo stato con il legame.

I due metodi descritti sopra, ossia l'identificazione di vie di accesso (tunnel e canali) e l'analisi del trasporto del ligando, tipicamente vengono svolti in modo separato. Tuttavia, ci sono strumenti come quello messo a disposizione da Caver Web, che riunisce i suoi software deputati alle due azioni in un'interfaccia unica [7].

### 3.4 Dinamica proteica

La dinamicità delle proteine si manifesta in varie modalità sia su scala temporale (alcune variazioni conformazionali richiedono nanosecondi, mentre altre millisecondi), sia per quanto riguarda le dimensioni (alcune variazioni coinvolgono la scala degli Å, altre i nm). Il suo studio è importante perché la dinamica è alla base di molti processi fondamentali per le interazioni e la funzionalità della proteina, tra cui la catalisi enzimatica.

La dinamica molecolare (MD) è il metodo computazionale più comune per studiare la dinamica delle proteine sia a livello atomico, sia a livello molecolare. Essa consiste nell'applicare agli atomi che costituiscono la proteina la seconda legge di Newton, ossia  $F = ma$  (dove  $m$  ed  $a$  sono rispettivamente la massa e l'accelerazione

della particella, mentre  $F$  si riferisce alle forze che agiscono sulla stessa). Nella MD, la forza viene associata al gradiente della funzione energia potenziale, la quale descrive una superficie chiamata PES (Potential Energy Surface) o superficie di energia potenziale.

Per determinare la traiettoria delle componenti è necessario conoscere alcune proprietà iniziali: la posizione iniziale degli atomi, ottenute tramite metodi sperimentali come la NMR o la cristallografia a raggi X; la velocità iniziale; e le accelerazioni (tramite PES).

La superficie di energia potenziale si traduce in un modello ad imbuto, dove l'asse  $z$  corrisponde all'energia libera. Ne consegue che i punti di minimo sono quelli in cui la molecola è più stabile mentre, più si sale lungo l'asse, più vi è instabilità.

Il calcolo di energia avviene tramite due tipologie di metodo: metodi quantomeccanici, in cui si considera la velocità sia dei nuclei, sia degli elettroni; e metodi classici, in cui si tiene conto solo della velocità dei nuclei [7].

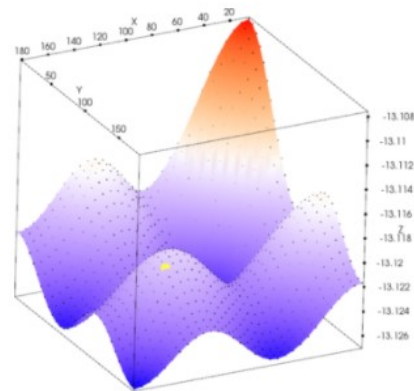


Figura 5 - PES e modello ad imbuto

### 3.5 Identificazione dell'hotspot e smart library

L'identificazione di siti che possono accogliere mutazioni è importante per una modifica mirata a livello della sequenza genomica, la quale possa migliorare l'efficienza dell'enzima. Unendo la metodologia dell'evoluzione diretta e del design razionale, è possibile ottenere delle smart library con un numero limitato di varianti, ma di alto interesse. La strategia scelta per l'identificazione degli hotspot dipende dalla proprietà alla quale si è interessati. Ad esempio, gli hotspot implicati nella stabilità possono essere identificati col software B-fitter, il quale si basa sull'analisi della struttura cristallografica degli enzimi.

Il metodo più comune per identificare gli hotspot è l'analisi filogenetica, la quale può essere fatta mediante software come ClustalX. Infatti, creando dei multiallineamenti tra sequenze di proteine appartenenti ad organismi differenti, è possibile individuare le regioni che sono maggiormente conservate. Il significato della loro conservazione è correlato ad un ruolo rilevante dal punto di vista funzionale o strutturale e di conseguenza non saranno regioni sicure da mutagenizzare. Al contrario, le regioni che possono accogliere mutazioni, sono quelle che dai multiallineamenti risultano molto variabili poiché non hanno un ruolo fondamentale. In seguito, l'insieme delle sostituzioni possibili può essere utilizzato per creare delle smart library di codoni che codificano per le varianti proteiche ricercate.

Il software HotSpot Wizard unisce varie tecniche e database per l'identificazione degli hotspot e la costruzione di smart library. Infatti, tramite l'input della struttura tridimensionale dell'enzima scelto, il software compie una serie di analisi identificando i residui conservati, mutabili e catalitici, individuando poi le sostituzioni possibili. Infine, darà come output la costruzione di smart library e una

valutazione della stabilità delle varie strutture con mutazioni negli hotspot identificati [7].

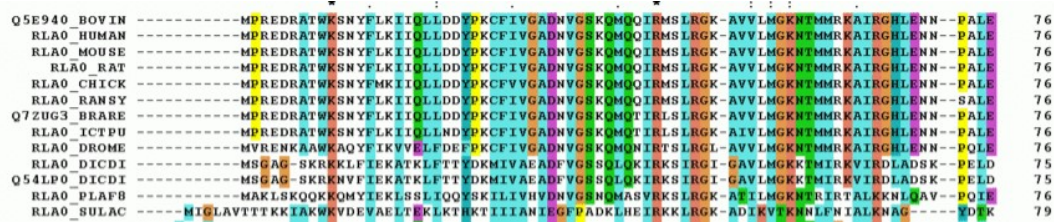


Figura 6 - Multiallineamento effettuato con ClustalX

### 3.6 Mutagenesi

Nel contesto della pianificazione delle mutazioni sugli enzimi, è importante valutare quali effetti le mutazioni possano produrre in termini di solubilità e stabilità proteica. Tra gli strumenti che permettono la predizione degli effetti di una mutagenesi, si trovano il machine learning, il calcolo del force-field e l'analisi filogenetica.

Le tecniche di machine learning non richiedono una conoscenza dei principi meccanicistici alla base della catalisi enzimatica, ma bisogna fare attenzione alla qualità dei dati sperimentali che vengono appresi dalla macchina. Infatti, nel caso in cui venga proposta una quantità di dati troppo piccola, c'è il pericolo di over-training del modello, il quale non considera la diversità che dovrebbe rappresentare il training set.

I metodi force-field si rifanno a modelli energetici derivanti dalle leggi fisiche e permettono di predire potenziali variazioni strutturali.

L'analisi filogenetica mira a prevedere gli effetti delle mutazioni identificando regioni conservate in multi-allineamenti di proteine omologhe dell'enzima di interesse [7].

## **4. APPLICAZIONI DEL DESIGN COMPUTAZIONALE SUGLI ENZIMI**

### **4.1 Progettazione di biocatalizzatori efficienti**

Prima di modificare l'efficienza catalitica degli enzimi, è necessario comprenderne i meccanismi tramite metodi computazionali, dove l'efficienza è definita tramite l'energia di attivazione  $E_a$ . Secondo il modello della cinetica di Michaelis-Menten, l'efficienza catalitica di un enzima è invece associata al rapporto  $k_{cat}/K_M$ . In questa formula,  $k_{cat}$  si riferisce alla velocità catalitica, mentre  $K_M$  rappresenta la quantità di substrato necessaria per raggiungere la metà della velocità massima di reazione.

Un esempio è costituito dal lavoro svolto da Zheng et al. (2014), i quali hanno modificato l'enzima butirrilcolinesterasi, un enzima prodotto dal fegato per idrolizzare l'acetilcolina ma che può anche degradare la cocaina. L'ingegnerizzazione ha permesso che l'enzima idrolizzasse la cocaina con un'efficienza catalitica aumentata di un ordine di grandezza, aprendo una nuova strada al trattamento clinico dei pazienti in sovradosaggio [7].

### **4.2 Progettazione di enzimi con attività nuove**

La progettazione di attività enzimatiche nuove che non sono presenti in natura, può essere raggiunta tramite modificazioni di vario tipo, che spaziano dal solo sito attivo fino alla progettazione di un enzima da zero.

Il design de novo richiede diversi passaggi (Figura 7): il backbone sampling, ossia la ricerca di un backbone proteico che si addica alla funzione richiesta; l'ottimizzazione della sequenza del backbone, tramite l'aggiunta di residui che stabilizzano la struttura; il design del sito funzionale; e l'assegnazione di un punteggio basato sull'energia delle possibili strutture, in cui il punteggio più basso sarà assegnato alla proteina più stabile [6].

Il software Rosetta è uno degli strumenti che possono essere usati nel design de novo, poiché è in grado di predire la struttura della proteina target tramite l'input della struttura primaria. Vengono applicati due approcci differenti in base alla tipologia di predizione, la quale può essere ab initio, in cui la predizione avviene da zero, oppure template-based, in cui si cambia una parte di una proteina nota.

La predizione ab initio consiste in tre passaggi:

1. La sequenza non nota viene divisa in tratti di 3-9 residui, i quali vengono confrontati con una banca di proteine a struttura nota. Per ognuno, viene quindi costruito un possibile elenco di strutture tridimensionali (Figura 8);
2. Il simulatore Montecarlo genera 1000-10000 possibilità di assemblaggio dei vari tratti di catene laterali, applicando delle sostituzioni casuali dei frammenti di 3-9 residui. La scelta casuale riguarda delle sostituzioni che secondo il programma conferiscono più stabilità;
3. Le varie conformazioni vengono valutate in base al loro valore di energia, il quale è correlato alla stabilità della struttura. La valutazione avviene favorendo: interazioni fra cariche opposte; conformazioni laterali che sono frequentemente osservate in natura; i legami ad idrogeno con una buona

geometria; e le interazioni di van der Waals. Inoltre, viene sfavorita la vicinanza tra gruppi polari [1].

Nonostante la predizione ab initio non sia ancora sufficientemente accurata, negli ultimi anni sono stati fatti molti passi avanti soprattutto grazie all'entrata in campo degli algoritmi basati sull'intelligenza artificiale. Infatti, a CASP14 (2020), ossia l'ultima edizione di una competizione in cui varie squadre sviluppano degli algoritmi per la predizione ab initio, ha vinto l'algoritmo AlphaFold. Questo algoritmo basato sul machine-learning è stato sviluppato da una squadra di DeepMind ed ha ottenuto un punteggio medio di 92.4 GDT. Il Global Distance Test (GDT) rappresenta la percentuale di amminoacidi con una distanza dalla posizione corretta al di sotto di una certa soglia. Le predizioni di AlphaFold sono state quindi relativamente vicine a strutture ottenibili per via sperimentale e in CASP15, che si terrà tra i mesi di maggio e giugno 2022, si potrebbero osservare ulteriori miglioramenti.

Per quanto riguarda la predizione template-based, per ottimizzare la struttura a livelli di energia più bassi viene usato il simulatore Montecarlo. Applicando una sostituzione a degli amminoacidi, si profilano due casi: abbassamento del livello di energia, dove la sostituzione viene automaticamente accettata; e innalzamento del livello di energia, dove la sostituzione viene accettata con un certo livello di probabilità definito dal criterio di Metropolis. Quest'ultimo abbassa la probabilità di accettazione del nuovo modello quando le variazioni di energia sono eccessivamente elevate ed il sistema è ad una più bassa temperatura [2].

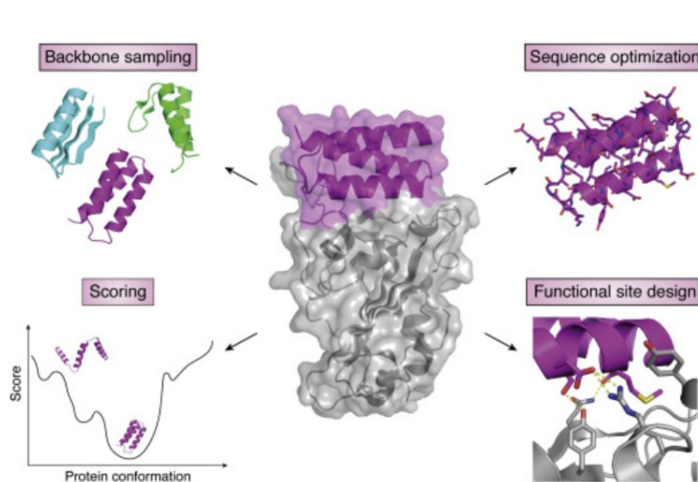


Figura 7 - Tappe del design de novo

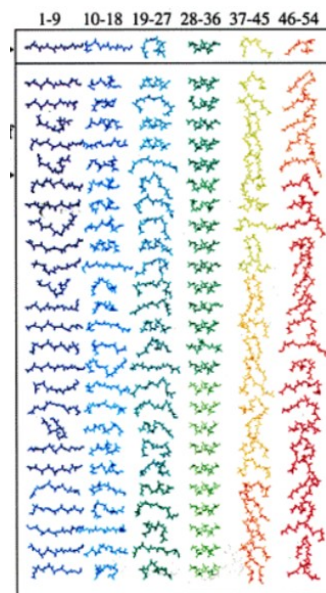


Figura 8 - Divisione in tratti di 9 residui da parte di Rosetta



### 4.3 Progettazione di enzimi specifici ed enantioselettivi

La specificità di substrato degli enzimi è rappresentata dalla capacità di legare uno o più ligandi. Il grado di specificità può essere: alto, in cui l'enzima si lega solo ad un substrato; basso, in cui l'enzima agisce su più substrati; e può esserci la presenza di promiscuità enzimatica, in cui l'enzima catalizza un substrato principale ma è anche in grado di legarne altri come ruolo secondario. Può essere quindi molto utile a livello industriale il design di enzimi che abbiano un'alta specificità per determinati substrati utilizzati.

Il software più utilizzato per l'ingegnerizzazione della specificità è Rosetta. Un esempio del suo utilizzo è costituito dal lavoro svolto da Murphy et al. (2009), il quale si articola nel design di una guanina deaminasi affinché compia una deaminazione della ammelide. In una prima fase, un'ipotetica struttura dello stato di transizione della ammelide è stata inserita a livello del sito attivo e sono stati aggiunti dei residui di ancoraggio per il nuovo ligando. In secondo luogo, si è mirato a stabilizzare la nuova struttura proteica e si sono predette le condizioni necessarie affinché l'ammelide potesse essere legata. Questo processo ha portato all'aumento di 6 ordini di grandezza della selettività ed un aumento di 100 volte dell'attività.

L'ingegnerizzazione dell'enantioselettività avviene in due direzioni: promozione del legame con l'enantiomero target ed introduzione di ostacoli sterici al legame del secondo enantiomero. Questo, può avvenire generando delle librerie costituite da enzimi che presentano le due condizioni precedentemente esposte. In seguito, si valuteranno i vari enzimi della libreria tramite simulazioni di dinamica molecolare. In questo modo, si arriva alla selezione di un piccolo numero di mutanti da testare sperimentalmente [7].

### 4.4 Design di enzimi stabili e solubili

L'ingegnerizzazione degli enzimi che mira ad aumentare le proprietà catalitiche può portare come effetto collaterale ad una diminuzione della solubilità e stabilità. Tuttavia, l'ingegneria proteica può essere a sua volta usata per indurre mutazioni che stabilizzano e solubilizzano l'enzima.

La maggior parte dei metodi per ingegnerizzare la solubilità sono di machine learning. Come dati sperimentali possono essere forniti profili di sequenze e raccolte di mutazioni associate ad un punteggio di solubilità che contribuisce alla solubilità totale dell'enzima.

I metodi per la predizione degli effetti sulla stabilità a seguito di mutazioni sono principalmente di calcolo del force-field, machine learning ed analisi filogenetiche. Quest'ultimo metodo ha il vantaggio di non richiedere la conoscenza della struttura tridimensionale, al contrario del calcolo del force-field [7].

## **4.5 Design di enzimi con flessibilità adattata**

L'ingegnerizzazione della flessibilità è importante poiché i movimenti interni degli enzimi possono agire su diversi processi come l'azione catalitica, il binding del ligando ed il rilascio del prodotto.

La modificazione della flessibilità, che avviene sia a livello delle catene laterali sia del backbone, non è però facilmente attuabile per due motivi: le mutazioni che si introducono agiscono ad un livello molto fine, andando a variare ampiezze e frequenze specifiche dei movimenti target; e le interazioni a lungo raggio che controllano i movimenti target non sono ancora comprese a sufficienza.

I siti target possono essere sia a livello delle catene laterali sia del backbone, ma le regioni principalmente colpite dall'ingegnerizzazione sono i tunnel ed i loops [7].

## CONCLUSIONI

Le minacce del cambiamento climatico, dell'esaurimento delle risorse fossile e dell'inquinamento rendono fondamentale una revisione del modello economico, accogliendone uno di tipo circolare. In questo contesto, le biotecnologie giocano un ruolo di primo piano tramite la biocatalisi.

La produzione di biometano a partire da materiale di scarto, come rifiuti organici o scarti dell'industria zootecnica che vengono poi degradati dagli archei metanogeni, è una delle applicazioni biotecnologiche più promettenti. Tuttavia, è necessario migliorare la produzione per renderla più praticabile a livello industriale. Questo può essere possibile grazie al grande sviluppo che sta avendo la bioinformatica, la quale permette il design computazionale di enzimi migliorandone determinate caratteristiche.

Riguardo alle modifiche ad enzimi chiave del processo metanogenico, la letteratura scientifica è molto limitata. Secondo Lawton T. J.; Rosenzweig A. C. (2016) sarebbe una strada percorribile l'ingegnerizzazione di mMCR (MCR metanogeno) al fine di favorire la reazione inversa. Sarà però necessario comprendere bene quali fattori agiscano sulla direzione della reazione e se la aMCR proceda semplicemente nella via opposta rispetto alla mMCR [4].

Resta tuttavia largo spazio alla ricerca, con l'ingegnerizzazione di enzimi coinvolti nel processo metanogenico in cui si migliorino le caratteristiche per renderli più efficienti a livello industriale. Questo sarà possibile modificando la stabilità, l'efficienza, la specificità, la flessibilità e il design de novo di enzimi funzionalmente utili.

## BIBLIOGRAFIA

[1] Helmer Citterich M. [et al.], *Fondamenti di bioinformatica*, Zanichelli, 2018

## SITOGRAFIA

- [2] Kuhlman B., *Designing protein structures and complexes with the molecular modeling program Rosetta*, in “Journal of Biochemical Chemistry”, 7 novembre 2019, 10.1074/jbc.AW119.008144
- [3] Iea.org, *An introduction to biogas and biomethane*, visitato il 7/2/2022, <https://www.iea.org/reports/outlook-for-biogas-and-biomethane-prospects-for-organic-growth/an-introduction-to-biogas-and-biomethane>
- [4] Lawton T. J.; Rosenzweig A. C., *Biocatalyst form methane conversion: big progress on breaking a small substrate*, in “Chemical biology”, dicembre 2016, 10.1016/j.cbpa.2016.10.001
- [5] Lyu Z. et al., *Methanogenesis*, in “Current Biology”, 9 luglio 2018, 10.1016/j.cub.2018.05.021
- [6] Pan X.; Kortemme T., *Recent advances in de novo protein design: Principles, methods, and applications*, in “Journal of Biochemical Chemistry”, gennaio 2021, 10.1016/j.jbc.2021.100558
- [7] Planas-Iglesias J. et al., *Computational design of enzymes for biotechnological applications*, in “Biotechnology advances”, Marzo-Aprile 2021, 10.1016/j.biotechadv.2021.107696
- [8] Sheldon R. A., *Biocatalysis and biomass conversion: enabling a circular economy*, in “The Royal Society”, 6 luglio 2020, 10.1098/rsta.2019.0274
- [9] Soo V. W. C. et al., *Reversing methanogenesis to capture methane for liquid biofuel precursor*, in “BMC”, 14 gennaio 2016, 10.1186/s12934-015-0397-z
- [10] Truppo M. D., *Biocatalysis in the Pharmaceutical Industry: The Need for Speed*, in “ACS publications”, 18 aprile 2017, 10.1021/acsmmedchemlett.7b00114