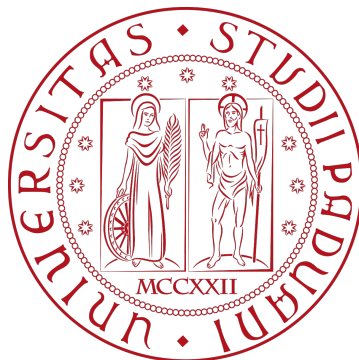


UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI BIOLOGIA

Corso di Laurea in Biotecnologie



ELABORATO DI LAUREA

Valutazione computazionale dell'espressione eterologa
della metalloproteasi IrrE di *Deinococcus radiodurans*
in *Escherichia coli* e *Anabaena* sp. PCC 7120

Tutor: *Prof. Livio Trainotti*

Dipartimento di Biologia

Laureanda: *Chiara Zanon*

Matricola: 2100837

ANNO ACCADEMICO 2025/2026

Abstract

L'utilizzo di microrganismi in ambienti estremi, con applicazioni in astrobiologia e biorisanamento, richiede lo sviluppo di ceppi ingegnerizzati caratterizzati da un'elevata tolleranza ai diversi stress a cui sono sottoposti. Il batterio poliestremofilo *Deinococcus radiodurans* si distingue per l'elevata tolleranza a radiazioni ionizzanti, mediata in parte dalla metalloproteasi pleiotropica IrrE. Nonostante sia stato dimostrato che la sua integrazione conferisca radio- e osmo-resistenza, i target proteolitici in *chassis* non nativi restano, ad oggi, ancora poco conosciuti.

Questo elaborato di tesi valuta, in *silico*, l'impatto sistemico dell'espressione eterologa di IrrE di *Deinococcus radiodurans* in *Escherichia coli* e *Anabaena* sp. PCC 7120. Dal momento che l'introduzione di una proteasi esogena può alterare l'omeostasi cellulare, è stata sviluppata una pipeline per valutare l'interazione tra la proteina ricombinante e il proteoma ospite e distinguere gli effetti citotossici da quelli vantaggiosi. I risultati delle predizioni strutturali e delle simulazioni di docking forniscono i presupposti molecolari necessari per l'ottimizzazione del design genetico prima della validazione sperimentale. Questa tesi delinea le basi teoriche per l'ingegnerizzazione di microrganismi destinati al biorisanamento e alle esplorazioni spaziali.

Indice

Abstract

1	Stato dell'arte	1
1.1	Radioresistenza batterica in ambienti estremi: implicazioni per l'astrobiologia	1
1.2	<i>Deinococcus radiodurans</i>	2
1.3	La proteasi IrrE e il repressore DdrO	2
1.4	Espressione eterologa del gene <i>IrrE</i>	4
2	Obiettivo	5
3	Materiali e Metodi	6
3.1	Chassis biologici	6
3.1.1	<i>Escherichia coli</i> K-12	6
3.1.2	<i>Anabaena</i> sp. PCC 7120	6
3.2	Ottimizzazione genica e regolazione termodinamica della traduzione	7
3.3	Screening proteomico per l'identificazione dei target	8
3.4	Analisi delle reti di interazione	8
3.4.1	STRING	8
3.4.2	Cytoscape	9
3.4.3	CyanoCyc e EcoCyc	9
3.5	Predizione strutturale di IrrE	9
3.6	Simulazione di docking molecolare	9
4	Risultati e discussione	10
4.1	Ottimizzazione del gene <i>irrE</i>	10
4.2	Specificità di taglio	12
4.3	Valutazione della tossicità off-target	14
4.3.1	Analisi degli off-target in <i>Escherichia coli</i> K12	15
4.3.2	Analisi degli off-target in <i>Anabaena</i> sp. PCC 7120	16
4.4	Serendipità molecolare	17
4.5	Il modello predittivo del destino cellulare	19
5	Conclusioni e prospettive future	21
	Bibliografia	22
A	Sequenze Ottimizzate	24

Capitolo 1

Stato dell'arte

1.1 Radioresistenza batterica in ambienti estremi: implicazioni per l'astrobiologia

Negli ultimi anni, gli studi astrobiologici si sono concentrati sulla comprensione dei meccanismi molecolari alla base della sopravvivenza degli organismi estremofili in ambienti caratterizzati da condizioni ostili di temperatura, pressione, radiazioni e stress osmotico. Tra le varie tipologie di stress a cui gli organismi sono soggetti, le radiazioni ionizzanti (raggi gamma, raggi X e particelle cariche ad alta energia) e gli effetti biologici da loro indotti assumono un ruolo di rilievo nelle ricerche in ambito spaziale. Le radiazioni ionizzanti interagiscono con l'ambiente acquoso intracellulare innescando il processo di radiolisi, un fenomeno chimico-fisico che comporta la formazione di specie reattive dell'ossigeno (ROS, radicali liberi altamente instabili in grado di ossidare le macromolecole biologiche). La formazione di ROS comporta citotossicità attraverso l'induzione di stress ossidativo a carico delle macromolecole (perossidazione lipidica delle membrane, inattivazione ossidativa delle proteine, ossidazione delle basi azotate e rotture a doppio filamento del DNA genomico). A dosi di radiazioni superiori a pochi Gray, la maggior parte degli organismi non specializzati subisce una perdita di vitalità irreversibile. La radioresistenza non dipende dalla schermatura fisica del DNA, ma dall'integrità funzionale del proteoma: la frammentazione del genoma avviene, infatti, in modo analogo sia nei ceppi sensibili che in quelli resistenti. La differenza risiede nella presenza, nei microrganismi radioresistenti, di complessi intracellulari di manganese (Mn^{2+}) e metaboliti organici che agiscono come scavenger di ROS, prevenendo l'inattivazione ossidativa degli enzimi deputati alla riparazione del genoma.[1] L'efficacia di queste strategie adattative è stata avvalorata empiricamente dai risultati del progetto BIOMEX (Biology and Mars Experiment) condotto dall'Agenzia Spaziale Europea (ESA) sulla Stazione Spaziale Internazionale (ISS) per testare la resistenza di microrganismi e biomolecole. Specifiche colture batteriche, esposte per 18 mesi a condizioni marziane simulate sulla piattaforma EXPOSE-R2 della ISS, hanno preservato l'integrità cellulare e mantenuto una vitalità misurabile. [2] Tra gli organismi analizzati emerge il batterio poliestremofilo *Deinococcus radiodurans*. [3]

1.2 *Deinococcus radiodurans*

La scoperta di *Deinococcus radiodurans* risale al 1956, quando il gruppo di ricerca di Arthur W. Anderson lo isolò da campioni di carne sottoposti a intensi trattamenti di sterilizzazione industriale con raggi gamma. Si tratta di un microrganismo poliestremofilo di grandi dimensioni (circa 1,5-3,5 μm di diametro) con morfologia coccoide che tende ad aggregarsi in caratteristiche tetradi. Il fenotipo resistente a condizioni estreme (quali dosi di radiazioni ionizzanti superiori a 12.000 Gy, radiazioni UV, stress ossidativo ed essiccamento), letali per la quasi totalità degli organismi viventi, non deriva da una protezione strutturale del materiale genetico ma è associato alla preservazione dell'apparato proteico.[1] A seguito dell'esposizione a stress genotossico, il suo genoma poliploide riporta frammentazioni e rotture a doppio filamento comparabili a quelle di organismi sensibili. Pertanto, la sopravvivenza cellulare è determinata dalla capacità degli enzimi riparatori di non riportare gravi alterazioni strutturali a seguito del danno iniziale in modo tale che il batterio possa riassemblare, in poche ore, l'intero genoma frammentato attraverso il meccanismo ESDSA (*Extended Synthesis-Dependent Strand Annealing*) e la ricombinazione omologa. L'integrità dell'apparato enzimatico è garantita dall'accumulo intracellulare di ioni manganese (Mn^{2+}) i quali, coordinandosi con peptidi a basso peso molecolare e ortofosfati, formano complessi che agiscono da *scavenger* passivi neutralizzando i ROS prima che possano ossidare le proteine. L'efficacia di questo meccanismo di difesa è garantita dal mantenimento di basse concentrazioni di ferro libero (Fe^{2+}) nel citoplasma, condizione che previene l'innescò della reazione di Fenton. A livello molecolare, l'attivazione dei meccanismi di riparo del genoma è determinata dallo stesso stress ossidativo che colpisce la cellula. I ROS generati dall'evento genotossico ossidano i residui tiolici di specifiche metalloproteine cellulari alterandone la conformazione strutturale e inducendo, di conseguenza, un immediato rilascio intracellulare di ioni zinco. Gli ioni Zn^{2+} rilasciati agiscono da secondi messaggeri attivando specifici bersagli proteici a valle. Tra questi si distingue la metalloproteasi IrrE che, innattivando il repressore DdrO, innesca il processo di riparazione del DNA. [4]

1.3 La proteasi IrrE e il repressore DdrO

Per contrastare gli effetti di stress genotossici, *Deinococcus radiodurans* utilizza il regulone RDR (*Radiation Desiccation Response*) in sostituzione del sistema di risposta SOS tipico dei batteri. L'attivazione e la modulazione di questo apparato protettivo è regolata dall'interazione tra la metalloproteasi IrrE

(indicata in letteratura anche come PprI) e il suo target specifico, il repressore trascrizionale DdrO.

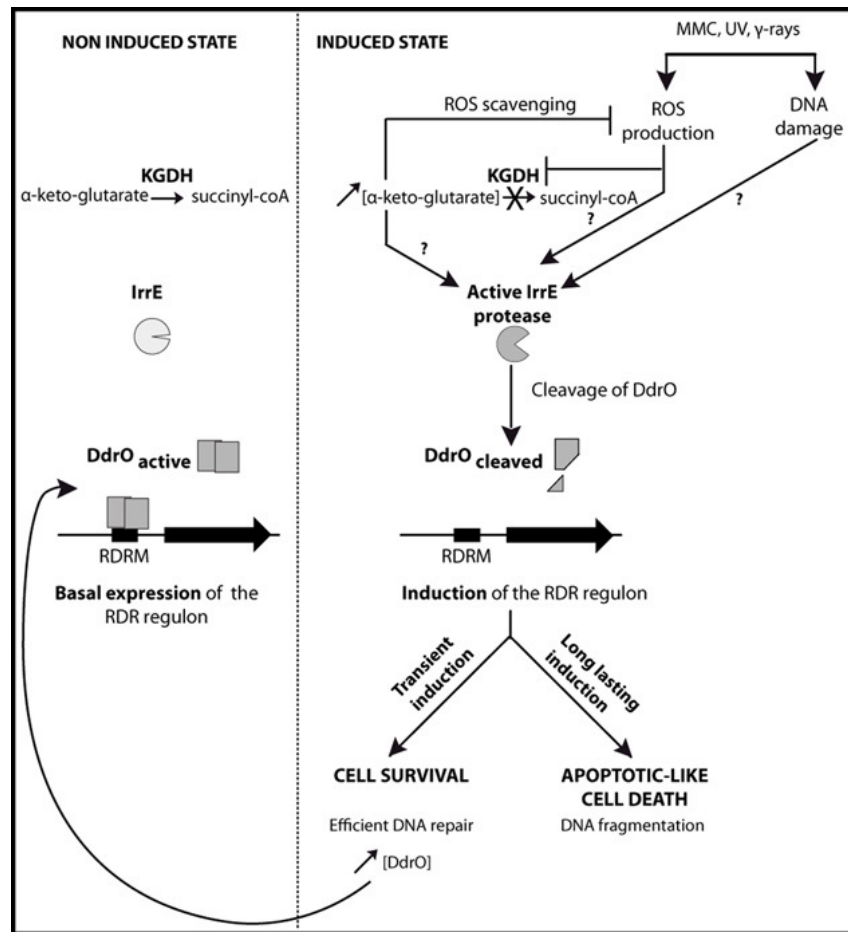


Figura 1.1: Modello del pathway di regolazione trascrizionale IrrE/DdrO. L'immagine è tratta da "*DdrO is an essential protein that regulates the radiation desiccation response and the apoptotic-like cell death in the radioresistant Deinococcus radiodurans bacterium*" di Devigne A. et al. in *Molecular Microbiology*(2015)[5]

La proteina IrrE (o PprI) è una metallo-endoropeptidasi pleiotropica caratterizzata da tre domini funzionali: un dominio catalitico N-terminale zinco-dipendente contenente il motivo conservato HEXXH, un dominio centrale Helix-Turn-Helix (HTH) con funzione strutturale e un dominio C-terminale GAF-like necessario per la trasduzione del segnale. IrrE viene attivata mediante una cascata redox indotta dallo stress genotossico: l'accumulo di ROS ossida i legami tiolici delle metalloproteine, innescando il rilascio di ioni Zn^{2+} che agiscono come effettori allosterici inducendo il riarrangiamento conformazionale necessario per l'attivazione enzimatica di IrrE.[4] Il substrato specifico di IrrE è DdrO, un repressore trascrizionale della famiglia XRE. La sua struttura comprende un dominio N-terminale con un motivo HTH, deputato al riconoscimento della sequenza palindromica RDRM

(*Radiation/Desiccation Response Motif*) di 17 pb presente nei promotori del regulone RDR e un dominio C-terminale, caratterizzato da un nucleo idrofobico, necessario per l'omodimerizzazione. In condizioni basali, DdrO agisce come omodimero, legandosi saldamente ai motivi RDRM presenti nei promotori di oltre venti geni di cui reprime (o riduce) la trascrizione (tra i quali *recA* e *pprA*). A seguito di stress genotossici, l'accumulo di Zn^{2+} attiva IrrE che effettua un taglio sito-specifico sul dominio C-terminale dei monomeri liberi di DdrO, scindendo il legame peptidico tra i residui Leu-108 e Arg-109. Questo comporta il distacco del repressore DdrO dal DNA e consente all'RNA polimerasi di avviare la trascrizione dei geni a valle dei promotori RDR. L'attività proteolitica di IrrE deve essere controllata e transitoria in quanto un'attività continuativa comporta la riduzione di DdrO con conseguente iper-espressione del regulone che può innescare esiti letali. [5]

1.4 Espressione eterologa del gene *IrrE*

L'espressione eterologa si configura come il processo biotecnologico attraverso il quale un gene esogeno viene inserito, trascritto e tradotto all'interno di un organismo ospite diverso da quello nativo, determinando la sintesi di una proteina ricombinante. Applicando questa tecnica è stato possibile esprimere il gene *IrrE*, isolato dal batterio estremofilo *Deinococcus radiodurans*, in organismi filogeneticamente distanti, conferendo loro fenotipi resistenti a diverse tipologie di stress abiotico. L'introduzione del gene *IrrE* in organismi ospiti determina un duplice effetto fisiologico. La sua espressione eterologa comporta evidenti vantaggi fenotipici, incrementando la radioresistenza verso raggi gamma e UV in *Escherichia coli*[6] e la tolleranza a stress osmotici e salini sia in *Escherichia coli* che in *Brassica napus*. Tuttavia, è emerso un effetto negativo associato all'elevato costo metabolico imposto all'ospite che, in condizioni ambientali ottimali e in assenza di stress, registra una riduzione del tasso di crescita basale.[7] Nel lievito *Saccharomyces cerevisiae* l'espressione di IrrE conferisce maggiore resistenza agli inibitori tossici caratteristici dei processi di fermentazione industriale (acido acetico, furfurale e fenolo). L'applicazione risulta, però, dipendente dal ceppo specifico e dalle condizioni operative.[8] Nonostante i fenotipi resistenti riscontrati in più *chassis* ospiti, permane una lacuna teorica relativa alle dinamiche molecolari coinvolte. La precisa identità dei bersagli di IrrE nell'ospite eterologo (dove la proteasi agisce in assenza dei suoi partner proteici nativi) rimane, ad oggi, ancora da chiarire.

Capitolo 2

Obiettivo

Il seguente elaborato di tesi si pone l'obiettivo di valutare, attraverso un rigoroso approccio *in silico*, l'impatto dell'espressione eterologa della metalloproteasi IrrE di *Deinococcus radiodurans* in *Escherichia coli* e in *Anabaena* sp. PCC 7120. L'inserimento di un enzima esogeno in un *chassis* non nativo altera l'omeostasi cellulare, comportando un costo metabolico. Pertanto, l'elaborato si propone di simulare a livello molecolare l'impatto dell'espressione eterologa. Viene definito un modello predittivo di destino cellulare, uno strumento utile per mappare le interazioni proteina-proteina tra l'enzima IrrE e il proteoma ospite, con l'intento di prevedere quali siano le possibili conseguenze funzionali prima di un'eventuale traslazione *in vitro*. L'espressione eterologa di IrrE in un organismo ospite può, in linea teorica, dar luogo a tre distinti esiti funzionali:

- **Tossicità off-target:** scenario sfavorevole in cui IrrE interagisce in modo aspecifico con proteine endogene essenziali determinando instabilità metabolica potenzialmente letale;
- **Inerzia funzionale:** condizione fenotipicamente neutra in cui l'enzima, pur essendo espresso correttamente, non identifica substrati compatibili all'interno del proteoma ospite e, pertanto, si accumula nel citoplasma comportando un costo metabolico privo di benefici adattativi;
- **Serendipità molecolare:** scenario in cui IrrE riconosce e inattiva selettivamente regolatori trascrizionali dell'ospite funzionalmente analoghi ai suoi target nativi conferendo al *chassis* un fenotipo resistente.

Per valutare in modo quantitativo queste ipotesi, la tesi sviluppa una pipeline bioinformatica strutturata in più fasi: ottimizzazione della sequenza codificante per garantire una sua espressione stabile in organismi ospiti, identificazione computazionale dei potenziali target proteolitici e simulazioni di docking molecolare tridimensionale. L'integrazione dei risultati consente di stimare le possibili interazioni fra IrrE e il proteoma ospite, fornendo le basi teoriche per l'ingegnerizzazione di microrganismi con potenziali applicazioni nel biorisanamento ambientale e in astrobiologia.

Capitolo 3

Materiali e Metodi

Il presente studio adotta il ciclo DBTL (Design–Build–Test–Learn), framework metodologico che prevede quattro fasi sequenziali: la progettazione del costruito genetico (Design), l’assemblaggio fisico dello stesso (Build), la caratterizzazione fenotipica in vivo e/o in vitro (Test) e la revisione del modello sulla base dei dati ottenuti (Learn). L’indagine si concentra sulle fasi di Design e Test, condotte integralmente *in silico* attraverso applicativi open-source.

3.1 Chassis biologici

La pipeline metodologica è stata testata parallelamente su due organismi ospiti: *Escherichia coli* K-12 e *Anabaena* sp. PCC 7120.

3.1.1 *Escherichia coli* K-12

Escherichia coli K-12 è un enterobatterio Gram-negativo consolidato come *chassis* eterotrofo standard nella biologia sintetica grazie alla rapidità di crescita, all’estesa caratterizzazione genetica e all’assenza di sistemi di restrizione endogeni nei ceppi di laboratorio. In questo studio funge da modello di validazione preliminare della pipeline bioinformatica. Il genoma di riferimento (NCBI Assembly: GCF_000005845.2) e il proteoma completo (Proteome ID: UP000000625) sono stati scaricati dai database pubblici (rispettivamente NCBI e UniProt) ed utilizzati come dataset per l’indagine computazionale.

3.1.2 *Anabaena* sp. PCC 7120

Anabaena sp. PCC 7120 (indicato in letteratura anche come *Nostoc* sp. PCC 7120) è un cianobatterio filamentoso fotoautotrofo capace di fissare autonomamente carbonio e azoto e di produrre ossigeno. Rispetto ad *Escherichia coli*, possiede tre endonucleasi di restrizione di Tipo II attive (AvaI, AvaII e AvaIII) che degradano il DNA esogeno non metilato. Costituisce il target applicativo finale del seguente studio per il suo potenziale astrobiologico e biotecnologico. Le sequenze genomiche e proteomiche di riferimento sono state estratte dai database pubblici NCBI e UniProt (NCBI Assembly: GCA_000009705.1, Proteome ID: UP000002483).

3.2 Ottimizzazione genica e regolazione termodinamica della traduzione

La sequenza wild-type del gene IrrE (NCBI Gene ID: 69516398, *Deinococcus radiodurans*) è stata sottoposta a ottimizzazione, in *silico*, attraverso un fork locale di TIsigner opportunamente modificato, insieme all'ingegnere Toni Kosumovic, per includere funzionalità non presenti nella versione web originale, previa autorizzazione degli sviluppatori. TIsigner è un tool bioinformatico progettato nel 2021 per ottimizzare l'espressione proteica migliorando l'accessibilità termodinamica dell'mRNA al sito di inizio della traduzione (Translation Initiation Site) e modificando i codoni sinonimi per prevenire la formazione di strutture secondarie stabili che impediscono il legame del ribosoma.[9] Poiché *Anabaena* sp. PCC 7120 non è un organismo modello standard come *Escherichia coli*, i suoi parametri genomici non risultano precaricati nella maggior parte dei tool open-source, limitandone le possibilità di ottimizzazione. Pertanto, il software è stato eseguito integrando due moduli supplementari al fine di garantire l'approccio "multi-criterio" supportato dalla letteratura che mostra come un'ottimizzazione genica fondata solo sulla massimizzazione del Codon Adaptation Index (CAI), tramite l'uso sistematico del codone più frequente, generi trascritti artificialmente ripetitivi.[10] Nello specifico, il codice originario è stato aggiornato aggiungendo un sistema di Codon Usage personalizzato che importa le tabelle in formato Kazusa specifiche per l'organismo analizzato (estrapolate dal Codon Usage Database del Kazusa DNA Research Institute). Questa modifica permette di allineare le sostituzioni sinonime dei codoni in funzione del relativo bias traduzionale. Una seconda modifica riguarda l'introduzione di un vincolo relativo al contenuto GC globale dell'intera sequenza, che porta l'algoritmo a scartare automaticamente qualsiasi mutazione alteri la percentuale nucleotidica al di fuori del range fisiologico tollerato dall'ospite. I parametri di ottimizzazione sono stati impostati per massimizzare l'accessibilità del trascritto a valle del promotore selezionato (rispettivamente P_{lac} per *Escherichia coli* e P_{petE} per *Anabaena* sp. PCC 7120), mantenendo il contenuto GC globale tra il 45% e il 60% per *Escherichia coli* K12 e fra il 36% e il 55% per *Anabaena* sp. 7120. L'algoritmo esclude inoltre una lista di siti di restrizione, comprendente i target di taglio per il clonaggio e, nel caso specifico di *Anabaena*, anche per le sue endonucleasi native (AvaI, AvaII e AvaIII). Il CAI associato a ciascuna sequenza è stato calcolato tramite il server E-CAI (<http://genomes.urv.es/CAIcal/E-CAI/>).

3.3 Screening proteomico per l'identificazione dei target

I proteomi di riferimento di *Escherichia coli* K-12 (UniProt ID: UP000000625; 4403 sequenze) e di *Anabaena* sp. PCC 7120 (UniProt ID: UP000002483; 6070 sequenze) sono stati scaricati in formato FASTA dal database UniProtKB. L'identificazione *in silico* dei potenziali substrati aspecifici di IrrE all'interno dei proteomi eterologhi è stata condotta inizialmente eseguendo, in locale, il tool bioinformatico PROSPEROusPlus (<https://github.com/lifuyi774/ProsperousPlus>). Si tratta di una piattaforma open-source progettata per la predizione, ad alta processività, dei target proteolitici che opera tramite algoritmi di ensemble machine learning e calcola uno score probabilistico di idrolisi per ogni singolo residuo. Dal momento che per IrrE non è presente un modello specifico, l'algoritmo è stato istruito fornendo in input una Position Weight Matrix (PWM) customizzata, generata allineando le sequenze dei siti di taglio riconosciuti nel substrato nativo DdrO, validati in letteratura per diverse specie del genere *Deinococcus*. La rappresentazione grafica del multiallineamento è stata generata utilizzando l'applicativo WebLogo 3.9.0 (<https://weblogo.threeplusone.com/>). A causa della ridotta numerosità del training set fornito (10 sequenze, riportate nella Tabella 4.3), è stato generato un modello degenere e non discriminante. Pertanto, è stata utilizzata una matrice di punteggio posizione-specifica (PSSM, Position-Specific Scoring Matrix), ottenuta stimando, per ogni posizione della sequenza di taglio, la frequenza dei residui amminoacidici nei siti positivi, normalizzandola rispetto alla frequenza degli stessi nel proteoma di riferimento. La selezione delle proteine target è stata ristretta alle sole sequenze i cui valori predittivi si collocano nel top 5% della distribuzione per il rispettivo proteoma di riferimento, al fine di ridurre il numero di falsi positivi.

3.4 Analisi delle reti di interazione

La classificazione funzionale e relazionale dei candidati emersi dallo screening proteomico ha consentito di valutare l'impatto sistemico di IrrE sull'omeostasi cellulare di *Escherichia coli* e *Anabaena* sp. PCC 7120.

3.4.1 STRING

STRING (<https://string-db.org>) è un database biologico strutturato per la previsione e l'integrazione delle interazioni proteina-proteina. Il tool è stato utilizzato per analizzare il profilo di interazione del set di proteine identificate come potenziali bersagli di IrrE. La soglia minima di *interaction score* è stata

fissata a 0.700 per filtrare i falsi positivi.

3.4.2 Cytoscape

Cytoscape (<https://cytoscape.org>) è una piattaforma software open-source progettata per la visualizzazione e l'analisi topologica delle reti di interazione molecolare. Il software è stato impiegato localmente per importare e generare i network tramite l'estensione integrata "STRING app". Gli strumenti di calcolo della piattaforma sono stati utilizzati per determinare il grado di connettività (node degree) dei diversi target.

3.4.3 CyanoCyc e EcoCyc

L'annotazione funzionale delle proteine individuate come target è stata eseguita tramite i database di pathway e genomi CyanoCyc (<https://cyanocyc.org/>) ed EcoCyc (<https://ecocyc.org/>). I dati estratti sono stati integrati con STRING e Cytoscape per mappare le interazioni proteina-proteina e analizzare la topologia dei network, con lo scopo di valutare l'ontologia e l'essenzialità delle proteine indagate.

3.5 Predizione strutturale di IrrE

I target selezionati dalle analisi precedenti sono stati sottoposti a modellazione tridimensionale per valutarne l'accessibilità sterica. I modelli tridimensionali dei target oggetto di studio sono stati generati utilizzando AlphaFold 3 (<https://alphafoldserver.com>), un modello basato su reti neurali ad apprendimento profondo (Deep Learning) in grado di predire le strutture tridimensionali di proteine. L'analisi e la visualizzazione tridimensionale dei modelli generati da AlphaFold 3 sono state condotte attraverso il software UCSF ChimeraX (<https://www.cgl.ucsf.edu/chimerax/>).

3.6 Simulazione di docking molecolare

L'interazione tra IrrE e i target individuati è stata simulata tramite il server web HDock (<http://hdock.phys.hust.edu.cn/>), un tool bioinformatico che permette di generare modelli tridimensionali di docking molecolare proteina-proteina. I complessi ottenuti sono stati successivamente processati, in locale, con PRODIGY (<https://wenmr.science.uu.nl/prodigy/manual>), un applicativo ottimizzato per la stima dell'affinità di legame a livello atomico. L'analisi dei parametri termodinamici calcolati, tra i quali l'energia libera di legame (ΔG) e la costante di dissociazione (K_d), ha consentito di discriminare le interazioni aspecifiche da quelle stabili e biologicamente rilevanti.

Capitolo 4

Risultati e discussione

4.1 Ottimizzazione del gene *irrE*

La sequenza wild-type del gene *IrrE* di *Deinococcus radiodurans* (lunghezza: 987 bp; contenuto GC: 69.1%) è stata analizzata per valutare se fosse idonea per l'espressione eterologa in *chassis* non nativi. L'analisi preliminare, condotta tramite il server E-CAI, ha restituito un Codon Adaptation Index (CAI) di partenza pari a 0.668 per *Escherichia coli* e 0.557 per *Anabaena*. Tali valori denotano un marcato disallineamento rispetto al codon bias degli ospiti, indicando una potenziale inefficienza traduzionale. A causa dell'elevato contenuto GC del gene *IrrE* di *Deinococcus radiodurans* è necessaria una riprogettazione sinonima per l'espressione nei *chassis* *Escherichia coli* (contenuto GC fisiologico pari a 50.8%) e *Anabaena* (contenuto GC fisiologico pari a 41.4%). L'algoritmo TIsigner, integrando le rispettive Codon Usage Table in formato Kazusa, ha generato quattro diverse varianti nucleotidiche per ciascun organismo ospite. Nello specifico, l'espressione del gene in *Escherichia coli* è stata posta sotto il controllo del promotore P_{lac} mentre per *Anabaena* è stato impiegato il promotore P_{petE} . I dati comparativi relativi al CAI calcolato, al contenuto nucleotidico (%GC) e all'accessibilità termodinamica del sito di inizio (Opening Energy) sono riportati rispettivamente in Tabella 4.1 per *Escherichia coli* e in Tabella 4.2 per *Anabaena*.

Sequenza <i>IrrE</i>	CAI	Contenuto GC (%)	Opening Energy (kcal/mol)
Wild-Type	0.665	69.1	13.91
Variante 1	0.634	59.5	8.61
Variante 2	0.631	59.2	10.58
Variante 3	0.624	58.4	12.25
Variante 4	0.631	58.5	10.59

Tabella 4.1: Analisi comparativa dei parametri tra la sequenza wild-type del gene *IrrE* e le varianti ottimizzate per *Escherichia coli* K12. In grassetto è evidenziata la variante selezionata per i passaggi successivi.

Sequenza <i>IrrE</i>	CAI	Contenuto GC(%)	Opening Energy (kcal/mol)
Wild-Type (Originale)	0.557	69.1	5.69
Variante 1	0,766	51.3	5.63
Variante 2	0.559	69.1	5.59
Variante 3	0.773	51.3	5.62
Variante 4	0,784	51.0	5.62

Tabella 4.2: Analisi comparativa dei parametri tra la sequenza wild-type del gene *IrrE* e le varianti ottimizzate *in silico* per *Anabaena* sp. PCC 7120. In grassetto è evidenziata la variante selezionata per i passaggi successivi

Dall'analisi comparativa delle varianti prodotte, la Variante 1 per *Escherichia coli* (CAI: 0.634; GC: 59.5%) e la Variante 4 per *Anabaena* sp. PCC 7120 (CAI: 0.784; GC: 51.0%) sono state selezionate come sequenze ottimali per una successiva validazione sperimentale. Nello specifico, confrontando i valori dei parametri ottenuti nelle sequenze generate per *Escherichia coli*, è evidente una lieve riduzione generale dell'indice CAI rispetto alla sequenza wild-type in quanto sono stati introdotti codoni sinonimi ricchi di adenina e timina per abbassare %GC globale entro la soglia fisiologica (dal 69.1% della sequenza wild-type a valori inferiori alla soglia del 60.0%). La riduzione dell'indice CAI è compensata, nella sequenza 1, dal miglioramento delle proprietà termodinamiche del trascritto, riscontrabile nella marcata diminuzione dell'energia di apertura (da 13.91 a 8.61 kcal/mol). Per quanto concerne le sequenze ottenute per *Anabaena* sp. PCC 7120, la variante 4 è caratterizzata da un miglioramento sinergico dei parametri confrontati: è stato mantenuto un valore di Opening Energy ottimale (5.62 kcal/mol), è stato ridotto il contenuto GC (dal 69.1% al 51.0%) ed è evidente un incremento dell'indice CAI (da 0.557 fino a 0.784).

Le due sequenze nucleotidiche integrali sono consultabili nell'Appendice A. In vista di una successiva traslazione *in vitro*, la validazione del costrutto richiede la calibrazione del Ribosome Binding Site (RBS). *IrrE* agisce come un regolatore pleiotropico e, pertanto, l'uso di un RBS forte comporterebbe una sovraespressione incontrollata e potrebbe indurre citotossicità. Per limitare questo rischio, si consiglia di optare per l'integrazione di un RBS ad efficienza moderata (come ad esempio l'RBS BBa_B0032 depositato nel Registry iGem) per garantire un'espressione basale stabile.

4.2 Specificità di taglio

L'analisi bioinformatica della struttura primaria di IrrE e del suo substrato nativo DdrO ha consentito di mappare la Cleavage Site Region (CSR) e ricavare il modello predittivo da applicare alle analisi successive. Gli studi biochimici e strutturali descritti in letteratura indicano che il repressore trascrizionale DdrO viene inattivato da IrrE tramite un taglio endoproteolitico sito-specifico a livello del suo dominio C-terminale, tra i residui Leu e Arg.[5] Per definire il sito di taglio riconosciuto da IrrE nel substrato, sono state estratte, dal database UniProt, le sequenze amminoacidiche di DdrO appartenenti a dieci specie rappresentative del genere *Deinococcus*. Per ciascuna sequenza è stata isolata una finestra ottamerica, comprendente quattro residui a monte del sito di taglio (posizioni P4–P1, estremità N-terminale) e quattro residui a valle (posizioni P1'–P4', estremità C-terminale), indicata secondo la nomenclatura standard di Schechter e Berger. L'allineamento multiplo di queste sequenze è riportato nella Tabella 4.3.

Organismo	UniProt ID	Sito N-term					Sito C-term			
		P4	P3	P2	P1	↓	P1'	P2'	P3'	P4'
<i>D. radiodurans</i>	Q9RRB8	R	I	E	L		R	G	K	R
<i>D. gobiensis</i>	AFD24237.1	R	I	E	L		R	G	K	R
<i>D. deserti</i>	C1CYP4	R	I	E	L		R	G	K	R
<i>D. geothermalis</i>	H8GUQ4	R	I	E	L		R	G	K	R
<i>D. reticulitermitis</i>	A0A1H7BZ47	R	I	E	L		R	G	K	R
<i>D. soli</i>	A0A0F7JL75	R	I	E	L		R	G	K	R
<i>D. puniceus</i>	A0A172TC40	R	I	E	L		R	G	K	R
<i>D. altitudinis</i>	A0ACH3L5D7	R	I	E	L		R	G	K	R
<i>D. peraridilitoris</i>	K9ZYE9	R	I	E	L		R	G	K	R
<i>D. irradiatisoli</i>	A0A2Z3JRE3	R	I	E	L		R	G	K	R
Motivo Consenso		R	I	E	L	↓	R	G	K	R

Tabella 4.3: Allineamento multiplo della sequenza del sito riconosciuto da IrrE nel repressore DdrO di 10 diverse specie del genere *Deinococcus*. La freccia (↓) indica il legame peptidico scisso dalla metalloproteasi.

I risultati ottenuti dall'allineamento multiplo delle sequenze di DdrO indicano che il motivo consenso riconosciuto da IrrE è conservato nelle specie analizzate dell'ordine *Deinococcaceae*. La rappresentazione grafica mediante Sequence Logo (ottenuta utilizzando l'applicativo WebLogo 3.9.0 e visualizzabile in fig. 4.1) rivela la totale conservazione del motivo R-I-E-L ↓ R-G-K-R nel dataset analizzato. L'altezza massima in bit raggiunta da ciascuna colonna conferma l'assenza di variabilità nella sequenza amminoacidica e la conseguente rigidità del nucleo catalitico in corrispondenza del sito di taglio (posizioni P4-P4').

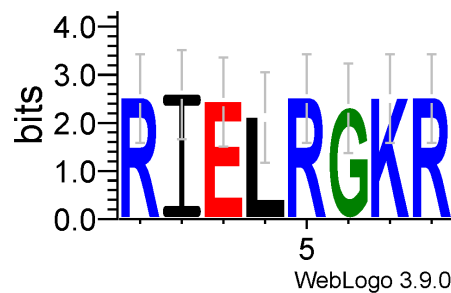


Figura 4.1: Sequence Logo del sito di taglio di IrrE nel substrato nativo DdrO, generato attraverso WebLogo 3.9.0 con schema di colorazione *Chemistry*

Nelle diverse specie analizzate di *Deinococcus*, la proteina DdrO non è interamente conservata. Pertanto, la conservazione riscontrata nel sito di taglio riconosciuto da IrrE suggerisce che quest'ultima agisca come un'endopeptidasi a specificità ristretta (riconosce un motivo conservato). Questa selettività è dettata dalle interazioni steriche ed elettrostatiche che avvengono all'interno della tasca catalitica di IrrE: il carbonile della Leucina in P1 si coordina con lo ione Zn^{2+} ancorato al motivo conservato HEXXH, mentre l'Arginina in P1' e gli amminoacidi con cariche positive in P3' e P4' stabilizzano il complesso. L'espressione di IrrE in organismi eterologhi non coevoluti, quali *Escherichia coli* e *Anabaena* sp. PCC 7120, può comportare potenziali interazioni *off-target*, derivanti dal riconoscimento di substrati endogeni da parte dell'enzima stesso. In particolare, la presenza nel proteoma ospite di sequenze strutturalmente analoghe al motivo consenso potrebbe comportare la proteolisi aberrante di proteine essenziali per il *chassis* studiato. È stata, perciò, eseguita una valutazione predittiva mediante PROSPEROus Plus per mappare il sito di taglio nei proteomi ospiti. Il modello iniziale ha adottato una matrice di peso posizionale (PWM) derivata dal multiallineamento delle sequenze native di DdrO (Tabella 4.3). Tuttavia, a causa dell'esiguo numero di sequenze allineate, PROSPEROus Plus ha generato un modello degenerare che assegna un punteggio uniforme e non discriminante a ciascuna proteina analizzata (score pari a 0,0971). Per ovviare a questo limite, dal medesimo

multiallineamento è stata ricavata una matrice di punteggio posizione-specifica (PSSM, Position-Specific Scoring Matrix) al fine di massimizzare la sensibilità dello screening proteomico.

4.3 Valutazione della tossicità off-target

La metalloproteasi IrrE esplica la propria attività catalitica attraverso il riconoscimento del motivo consenso R-I-E-L ↓ R-G-K-R. Tuttavia non tutte le proteine contenenti la sequenza specifica costituiscono bersagli desiderati. La caratterizzazione del profilo off-target è essenziale per escludere che l'espressione eterologa di IrrE in (*Escherichia coli* K12 e *Anabaena* sp. PCC 7120) provochi il taglio accidentale di proteine endogene essenziali per la sopravvivenza cellulare. Il potenziale di tossicità off-target cresce al convergere di tre fattori: (i) l'accuratezza del sito di taglio, valutata mediante il punteggio della matrice; (ii) il ruolo della proteina bersaglio nella fisiologia cellulare; (iii) la sua centralità nella rete di interazioni proteiche. L'idrolisi di un hub proteico, quantificato dal suo alto grado di connettività (node degree), amplifica l'effetto disfunzionale a livello sistemico a causa dell'interruzione delle vie molecolari ad esso collegate (effetto pleiotropico). L'intero proteoma dei due organismi indagati è stato processato mediante uno scoring basato su una matrice PSSM. L'applicazione del filtro del top 5% sulla distribuzione decrescente dei punteggi derivati dalle matrici PSSM ha consentito di isolare due librerie di potenziali target: rispettivamente 221 proteine per *Escherichia coli* K12 (con punteggi compresi tra 16,77 e 2,73 bit) e 304 per *Anabaena* sp. PCC 7120 (tra 16,88 e 2,73 bit). Lo score deriva esclusivamente dall'omologia della sequenza lineare e ignora la struttura tridimensionale della macromolecola. Un sito affine potrebbe, però, trovarsi nel core idrofobico della proteina nativa e non essere accessibile. I candidati che, a seguito di validazione strutturale condotta mediante AlphaFold 3 e ChimeraX, presentano siti inaccessibili vengono identificati come falsi positivi strutturali. Per valutare l'impatto sistemico di un eventuale taglio off-target, le librerie dei candidati sono state interrogate nel database STRING contro i rispettivi proteomi di riferimento, imponendo una soglia di confidenza alta ($\geq 0,700$). I risultati sono stati esportati in Cytoscape per il calcolo del node degree di ciascun bersaglio, parametro utile per stimare l'effetto associato alla potenziale degradazione dello stesso. L'analisi, riportata sinteticamente nella Tabella 4.4, evidenzia una marcata asimmetria. In *E. coli* il 100% dei candidati estratti trova corrispondenza nella rete, con 117 nodi che partecipano attivamente ad almeno un'interazione ad alta confidenza. Al contrario in *Anabaena* soltanto 53 dei 304 candidati (17,4%) mappa nel network. Questa discrepanza riflette la natura del genoma cianobatterico, ad

oggi, ancora costituito da molte proteine ipotetiche e accessorie che non sempre sono integrate nei modelli dei database. Dal punto di vista topologico, non emergono hub primari dal momento che nessun candidato raggiunge o supera la soglia critica di connettività, convenzionalmente fissata a 20 interazioni. Il node degree massimo riscontrato è 17 in *E. coli* (corrispondente alla subunità β della RNA polimerasi, *rpoB*) e 14 in *Anabaena* (proteina ribosomiale L4, *rplD*). L'assenza di hub primari suggerisce che i potenziali bersagli di IrrE siano localizzati in posizioni marginali o intermedie all'interno dell'interattoma cellulare. La valutazione dell'impatto sistemico (relativa ai 31 target essenziali in *E. coli* e 15 in *Anabaena*) rivela una netta anti-correlazione. I complessi macromolecolari primari e altamente connessi, come le subunità ribosomiali, mostrano punteggi PSSM deboli, denotando una probabilità di taglio minima, mentre i candidati con score più alto risultano topologicamente isolati (degree ~ 0) e non essenziali per la sopravvivenza dell'ospite.

<i>Chassis</i>	Candidati PSSM	Mappati STRING	Connessi (≥ 1 edge)	Node degree Max	Target Essenziali
<i>E. coli</i> K12	221	221 (100,0%)	117 / 221	17 (<i>rpoB</i>)	31 (Sperimentali)
<i>Anabaena</i> PCC 7120	304	53 (17,4%)	28 / 53	14 (<i>rplD</i>)	15 (Predetti)

Tabella 4.4: Riepilogo quantitativo e topologico delle librerie off-target (top 5%) estratte per i due *chassis* analizzati

4.3.1 Analisi degli off-target in *Escherichia coli* K12

In *Escherichia coli* K12, l'intersezione della libreria formata dai 221 target (caratterizzata da una copertura del 100% in STRING) con i dati sperimentali di EcoCyc ha isolato 31 proteine essenziali. Sono stati individuati, confrontando i punteggi PSSM e il node degree, alcuni potenziali bersagli off-target tra i quali, in ordine decrescente di tossicità, il fattore di inizio della traduzione IF-2 (*infB*, 10,35 bit), la DNA girasi B (*gyrB*, 9,47 bit), la subunità β' della RNA polimerasi (*rpoC*, 9,18 bit). Tra i potenziali bersagli identificati nello screening spicca anche il gene *rpoB*, codificante per la subunità β della RNA polimerasi. Presenta il massimo grado di connettività (node degree pari a 17) nell'interattoma di *Escherichia coli*, tuttavia lo score PSSM minimo pari a 3,05 bit riduce il rischio di taglio effettivo e, pertanto, la subunità viene classificata come falso positivo. Il fattore di inizio della traduzione IF-2 (*infB*) costituisce il target a massima tossicità (node degree 10): è stato rilevato, mediante l'ausilio del tool ChimeraX, che il sito di taglio si colloca su un loop esposto all'azione di IrrE. La DNA Girasi subunità B (*gyrB*) mostra, invece,

un motivo di taglio inserito all'interno di un foglietto β . Dal momento che il sito risulta parzialmente orientato verso il nucleo idrofobico si suppone che la probabilità di taglio da parte di IrrE sia ridotta. Gli impedimenti sterici nativi offrono un'efficace protezione alla subunità β' della RNA Polimerasi (rpoC): il sito di taglio è inserito in un'alfa-elica altamente strutturata, inglobata nel core idrofobico dell'enzima stesso e, pertanto, non accessibile ad IrrE. L'inattivazione proteolitica di questi specifici bersagli priverebbe la cellula delle sue funzioni fisiologiche, comportando l'arresto combinato di replicazione, trascrizione e traduzione con conseguente manifestazione di un possibile fenotipo letale.

4.3.2 Analisi degli off-target in *Anabaena* sp. PCC 7120

Nel cianobatterio *Anabaena* sp. PCC 7120, lo screening ha estrapolato 304 candidati, di cui solo 53 (17,4%) risultano mappati in STRING, riflettendo la limitata annotazione funzionale del genoma e l'abbondanza di proteine ipotetiche o accessorie. La tRNA pseudouridina sintasi B, enzima deputato alla modificazione post-trascrizionale dei tRNA, presenta uno score di 9,85 bit e un node degree pari a 6. Costituisce il rischio off-target più critico rilevato in quanto il sito teorico di taglio è localizzato in un loop esterno esposto al solvente. L'analisi strutturale del fattore di allungamento EF-Ts (10,22 bit, node degree pari 11) ha individuato il motivo su un'alfa-elica esposta verso l'esterno nel modello surface, confermando la probabile accessibilità fisica del substrato a IrrE. Al contrario, la subunità γ della RNA polimerasi (9,45 bit, node degree 7) costituisce un falso positivo strutturale in quanto il sito di taglio è integrato in un'alfa-elica interna, in una porzione che ne preclude l'esposizione all'enzima. Infine, nella proteina ribosomiale bS16 (9,39 bit, grado 12), il motivo coincide con l'intero foglietto β (β -strand) N-terminale iniziale. Pur occupando una posizione geometricamente esterna, esso permane confinato e delimitato in una tasca della struttura e, quindi, non accessibile. L'alterazione dell'apparato trascrizionale e dei fattori di allungamento ribosomiale causata da causerebbe l'incapacità di preservare il proteoma basale, inducendo la morte dell'ospite. Il quadro emerso dal confronto dei parametri bioinformatici e strutturali è formalizzato nella Tabella 4.5.

Ospite	Gene (Proteina)	Score	Rank	Node Degree	Funzione	Accessibilità (ChimeraX/AF3)
<i>E. coli</i>	<i>infB</i> (IF-2)	10,35	3	10	Inizio traduz.	Esposto superf. (loop flessibile)
	<i>gyrB</i> (Girasi B)	9,47	16	8	Topol. DNA	Parz. interrato (foglietto β)
	<i>rpoB</i> (RNA pol β)	3,05	100	17	Trascriz. basale	Probabilità di taglio minima
	<i>rpoC</i> (RNA pol β')	9,18	26	10	Trascriz. basale	Schermato strutt. (α -elica)
<i>Anabaena</i>	<i>tsf</i> (EF-Ts)	10,22	10	11	Allung. traduz.	Superf. ed esposto (α -elica)
	<i>truB</i>	9,85	14	6	Modif. tRNA	Altam. esposto (loop esterno)
	<i>rpoC1</i> (RNA pol γ)	9,45	27	7	Trascriz. basale	Interrato e protetto (α -elica int.)
	<i>rpsP</i> (bS16)	9,39	36	12	Strutt. ribos.30S	Esterno, delimitato (β -strand iniz.)

Tabella 4.5: Profilo di alcuni candidati target dei due chassis.

4.4 Serendipità molecolare

La pipeline sviluppata per identificare i bersagli off-target di IrrE ha generato un esito inatteso ascrivibile a "serendipità molecolare". L'ipotesi di "serendipità" postula che IrrE potrebbe tagliare un regolatore endogeno del *chassis* ospite in modo analogo a quanto avviene in *Deinococcus radiodurans*. I bersagli ideali devono coniugare un sito di taglio accessibile, una funzione regolatoria coerente con lo scopo e un impatto minimo sulla rete cellulare globale. I complessi IrrE–target, predetti con AlphaFold 3 e HDOCK, sono stati valutati tramite PRODIGY (ΔG e K_d a 25 °C) e validati in ChimeraX. Affinché la reazione avvenga, il sito di taglio nella proteina target deve essere correttamente orientato verso il motivo di coordinazione dello zinco HEXXH di IrrE. Di conseguenza, i modelli caratterizzati da un'alta affinità ma da una conformazione non idonea sono stati scartati in quanto considerati cataliticamente inattivi (Tabella 4.6). I complessi caratterizzati sono presentati come casi esemplificativi, non come un elenco esaustivo; costituiscono i candidati più promettenti di un profilo serendipitico più ampio.

Target	Ospite	ΔG (kcal/mol)	K_d (25 °C)	Esito strutturale
<i>alr4567</i>	<i>Anabaena</i>	−6,5	16 μ M	Produttivo
<i>alr2176</i>	<i>Anabaena</i>	−11,1	7,1 nM	Produttivo
<i>torS</i>	<i>E. coli</i>	−12,1	1,4 nM	Non produttivo

Tabella 4.6: Esiti del docking IrrE–target

Lo screening del proteoma di *Anabaena* sp. PCC 7120 ha evidenziato due candidati principali: il regolatore trascrizionale *alr4567* e la proteina legante sideroforo-ferro *alr2176*.

Alr4567, repressore appartenente alla famiglia TetR/AcrR, costituisce il bersaglio a punteggio massimo dell'intero screening (16,88 bit) ed è

caratterizzato da un node degree pari a 0. L'analisi termodinamica rivela un'affinità elettrostatica moderata ($\Delta G = -6,5$ kcal/mol, $K_d \approx 16$ μ M), in grado di promuovere un rapido turnover proteolitico e prevenire il rischio del sequestro improduttivo dell'enzima (Figura 4.2). Dal punto di vista funzionale, il taglio di *alr4567* mima il meccanismo di DdrO in *Deinococcus radiodurans*: IrrE agisce da interruttore molecolare e promuove il distacco del repressore dal promotore, garantendo l'espressione del regulone a valle.

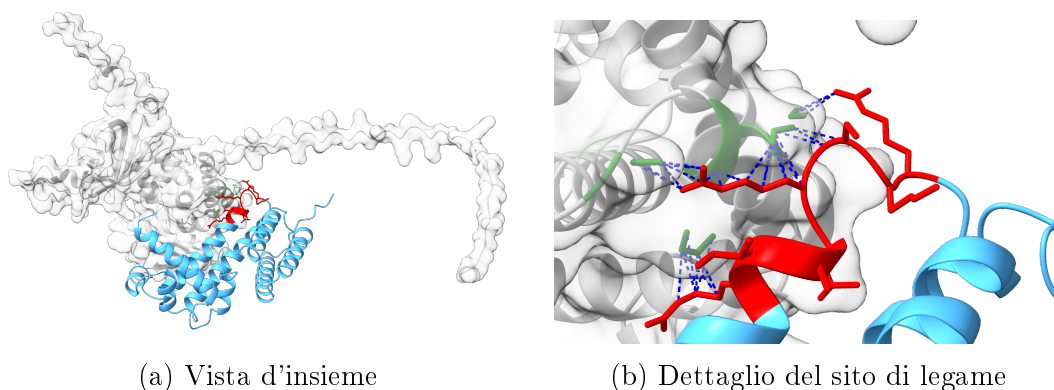


Figura 4.2: Complesso IrrE-*alr4567* (ChimeraX)(a) Vista d'insieme: la metalloproteasi IrrE è mostrata in formato surface in grigio chiaro, mentre il target *alr4567* è rappresentato a nastro (*cartoon*) in azzurro. Il frammento corrispondente al sito di taglio (residui 70-77) è evidenziato in rosso. (b) Il sito di taglio (rosso) è correttamente esposto ad IrrE

La proteina legante sideroforo-ferro *alr2176* (Trasportatore ABC) costituisce un secondo bersaglio serendipitico in *Anabaena*(10,22 bit). Questa posizione è associata alla funzione biologica assolta: il taglio da parte di IrrE comporterebbe l'inattivazione del trasportatore con conseguente riduzione dell'importazione di ferro all'interno della cellula. Ciò comporta la diminuzione della probabilità di innesco della reazione di Fenton mitigando, di conseguenza, il danno ossidativo e radiativo. La validazione tridimensionale ha richiesto una rigorosa selezione topologica: il modello 0 generato da AlphaFold 3 è stato scartato in quanto collocava il sito di taglio sul lato opposto rispetto alla tasca catalitica di IrrE. Il modello 1 (Figura 4.3), al contrario, ha confermato la formazione di un complesso termodinamicamente stabile ($\Delta G = -11,1$ kcal/mol, $K_d \approx 7,1$ nM).

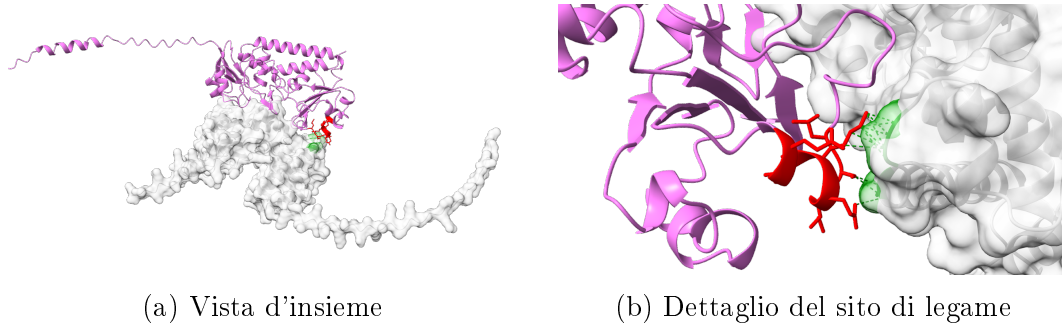


Figura 4.3: Complesso IrrE-alr2176 (ChimeraX). **(a)** Vista d'insieme: l'enzima IrrE è visualizzato in surface (grigio chiaro), il target *alr2176* a nastro in viola e il sito di taglio (residui 189-196) in rosso. **(b)** Dettaglio del sito di legame fra IrrE e alr2176

Nel proteoma di *Escherichia coli*, i bersagli serendipitici più probabili risultano proteine isolate, come AraA(16,77 bit). L'istidina chinasi sensoriale TorS rappresenta, invece, un controesempio paradigmatico. Nonostante mostri massima affinità termodinamica per IrrE ($\Delta G = -12.1$ kcal/mol; $K_d = 1.4$ nM), TorS non risulta un bersaglio valido poiché l'analisi con ChimeraX ha rilevato che il sito di taglio è orientato in direzione opposta rispetto alla tasca catalitica di IrrE comportando, perciò, un'interazione improduttiva.

4.5 Il modello predittivo del destino cellulare

Dal punto di vista metodologico, l'analisi basata solo sull'omologia della sequenza di taglio (punteggio PSSM), si dimostra intrinsecamente insufficiente in quanto non considera la reale accessibilità tridimensionale del sito. Per superare questo limite operativo, è stato sviluppato un modello predittivo del destino cellulare che integra lo score di taglio associato alla sequenza S , l'affinità termodinamica del complesso IrrE-bersaglio ΔG e l'accessibilità strutturale del sito a . La probabilità che il taglio avvenga è misurata dall'Indice di Suscettibilità Proteolitica (ISP), calcolato a partire dai suddetti parametri normalizzati ($[0, 1]$) :

$$ISP = S_n \cdot G_n \cdot a, \quad ISP \in [0, 1] \quad (4.1)$$

dove $S_n = S/S_{\text{rif}}$ è lo score di taglio calcolato tramite PSSM, normalizzato su un valore limite di riferimento ($S_{\text{rif}} = 20$); $G_n = |\Delta G|/|\Delta G|_{\text{rif}}$ è l'affinità termodinamica espressa in valore assoluto, normalizzata su un riferimento ad alta affinità ($|\Delta G|_{\text{rif}} = 15$ kcal/mol; valori più negativi indicano maggiore stabilità); $a \in \{0, 1\}$ indica l'accessibilità del sito di taglio (1 per sito esposto, 0 se non esposto). La normalizzazione basata su riferimenti fissi garantisce indipendenza dal campione dei dati analizzati. Come si evince dall'equazione 4.1, i termini concorrono in forma moltiplicativa poiché devono

essere soddisfatte simultaneamente tre condizioni affinché il taglio avvenga: (i) riconoscimento della sequenza espresso dallo score S ; (ii) accessibilità del sito; (iii) stabilità del complesso. Si definisce una soglia di significatività $\varepsilon = 0,25$, limite inferiore teorico che descrive un evento catalitico possibile. Questo valore corrisponde allo scenario in cui un target accessibile ($a = 1$) presenti parametri pari ad almeno la metà dei valori di riferimento massimi ideali ($S_n = 0,5$ e $G_n = 0,5$). Di conseguenza, le proteine target con $ISP < \varepsilon$ comportano inerzia funzionale mentre quelli con $ISP \geq \varepsilon$ sono identificati come candidati plausibili, da sottoporre a successiva caratterizzazione biologica. L'analisi sul proteoma di *Anabaena* posiziona alr4567 e alr2176 tra i target che possono essere teoricamente tagliati da IrrE ($ISP \approx 0,37$). Si nota come i due target raggiungano tale valore attraverso vie differenti: un elevato score di sequenza per il primo e una forte affinità di legame per il secondo. Il target *torS* di *Escherichia coli*, invece, determina inerzia funzionale a causa dell'impedimento sterico ($a = 0$). Il valore calcolato tramite l'equazione 4.1 stabilisce se il taglio può avvenire, ma non il suo impatto fenotipico. La valutazione dell'effetto del taglio di un target richiede un'analisi qualitativa della sua funzione specifica e della sua centralità nell'interattoma dell'ospite, quantificata dal node degree k . Un bersaglio con k elevato ($k > 20$) e una funzione biologica essenziale rappresenta un potenziale *off-target tossico* in quanto la sua proteolisi potrebbe compromettere pathway vitali, come ad esempio EF-Ts in *Anabaena*. Al contrario, un target caratterizzato da un k contenuto e una funzione non essenziale configura un possibile evento di *serendipità adattativa*. L' ISP è interpretabile come uno strumento da applicare *in silico* per escludere i falsi positivi strutturali (come *torS*) e isolare i candidati su cui concentrare l'annotazione funzionale. L'analisi suggerisce che i target alr4567 e alr2176 in *Anabaena* potrebbero costituire esempi di serendipità adattativa. Tuttavia, le conseguenze fenotipiche ipotizzate su base funzionale e topologica si configurano quali possibilità teoriche, non confermabili in assenza di validazione sperimentale *in vitro*.

Chassis	Target	S	a	ΔG (kcal/mol)	ISP	Taglio plausibile
<i>Anabaena</i>	alr4567	16,88	1	-6,5	0,37	Sì
<i>Anabaena</i>	alr2176	10,22	1	-11,1	0,38	Sì
<i>E. coli</i>	<i>torS</i>	9,62	0	-12,1	0,00	No

Tabella 4.7: Applicazione del modello ai target ipotizzati ($S_{\text{rif}} = 20$, $|\Delta G|_{\text{rif}} = 15$ kcal/mol). Il giudizio "Sì/No" è legato alla soglia di significatività $\varepsilon = 0,25$.

Capitolo 5

Conclusioni e prospettive future

Lo studio computazionale eseguito propone un quadro predittivo favorevole all'espressione eterologa della metalloproteasi IrrE di *Deinococcus radiodurans* in *Escherichia coli* e *Anabaena* sp. PCC 7120, fornendo una base teorica per le successive validazioni sperimentali *in vitro*. Sulla base dei risultati ottenuti è possibile affermare che analisi basate esclusivamente sull'omologia della sequenza amminoacidica sovrastimano il rischio sistemico. L'elaborato propone, pertanto, un modello predittivo del destino cellulare, sintetizzato nell'Indice di Suscettibilità Proteolitica (*ISP*). Si tratta di un'equazione che consente di stimare se il taglio può verificarsi, integrando affinità di legame, compatibilità strutturale e accessibilità del sito catalitico. L'applicazione del modello a *Escherichia coli* e *Anabaena* sp. PCC 7120 ha consentito di identificare i potenziali bersagli proteolitici di IrrE, ad ora, non ancora descritti in letteratura. Tra i risultati di maggiore interesse emergono, in *Anabaena* sp. PCC 7120, i target alr4567 e alr2176: la potenziale proteolisi del alr4567 *potrebbe* favorire la de-repressione di geni coinvolti nella risposta agli stress ambientali mentre l'inattivazione di alr2176 *potrebbe* limitare il danno ossidativo associato alla reazione di Fenton. È, tuttavia, opportuno ribadire il limite intrinseco dell'approccio adottato. Trattandosi di uno studio condotto esclusivamente *in silico*, il modello assume un valore di supporto decisionale, ma non probatorio. Gli effetti fenotipici ipotizzati restano possibilità puramente teoriche, dal momento che il modello non include tutte le variabili fisiologiche. Per confermare o smentire l'effettiva sussistenza di questi risultati è necessaria una successiva validazione sperimentale *in vitro* dell'attività di IrrE nei *chassis* indagati. Il passaggio verso la fase applicativa richiede la conferma dei tagli predetti attraverso spettrometria di massa e sistemi reporter *in vitro* e la caratterizzazione dell'essenzialità dei target tramite Tn-seq. Le evidenze di compatibilità e i potenziali vantaggi adattativi aprono prospettive per lo sviluppo di microrganismi ingegnerizzati destinati al biorisanamento ed applicabili in astrobiologia.

Bibliografia

- [1] Daly, M. J. "A new perspective on radiation resistance based on *Deinococcus radiodurans*", *Nature Reviews Microbiology*, 7(3), pp. 237–245, 2009, DOI: <https://doi.org/10.1038/nrmicro2073>
- [2] de Vera, J-P. et al. "Limits of Life and the Habitability of Mars: The ESA Space Experiment BIOMEX on the ISS", *Astrobiology*, 19(2), pp. 145-157, 2019, DOI: <https://doi.org/10.1089/ast.2018.1897>
- [3] Leuko, S. et al. "On the Stability of Deinoxanthin Exposed to Mars Conditions during a Long-Term Space Mission and Implications for Biomarker Detection on Other Planets", *Front. Microbiol.*, 8:1680, 2017, DOI: <https://doi.org/10.3389/fmicb.2017.01680>
- [4] Magerand, R. et al. "Redox signaling through zinc activates the radiation response in *Deinococcus* bacteria", *Scientific Reports*, 11, 4528, 2021, DOI: <https://doi.org/10.1038/s41598-021-84026-x>
- [5] Devigne, A. et al. "DdrO is an essential protein that regulates the radiation desiccation response and the apoptotic-like cell death in the radioresistant *Deinococcus radiodurans* bacterium", *Molecular Microbiology*, 96, pp. 1069-1084, 2015, DOI: <https://doi.org/10.1111/mmi.12991>
- [6] Gao, G., et al. "Expression of *Deinococcus radiodurans* PprI enhances the radioresistance of *Escherichia coli*", *DNA Repair*, 2(12), pp. 1419-1427, 2003, DOI: <https://doi.org/10.1016/j.dnarep.2003.08.012>
- [7] Pan, J. et al. "IrrE, a Global Regulator of Extreme Radiation Resistance in *Deinococcus radiodurans*, Enhances Salt Tolerance in *Escherichia coli* and *Brassica napus*", *PLOS ONE*, 4(2): e4422, 2009, DOI: <https://doi.org/10.1371/journal.pone.0004422>
- [8] Wang, L., et al. "Engineering prokaryotic regulator IrrE to enhance stress tolerance in budding yeast", *Biotechnology for Biofuels*, 13(1), 193, 2020, DOI: <https://doi.org/10.1186/s13068-020-01833-6>
- [9] Bhandari, B. K. et al. "TISIGNER.com: web services for improving recombinant protein production", *Nucleic Acids Research*, Volume 49, Issue W1, pp. W654–W661, 2 July 2021, DOI: <https://doi.org/10.1093/nar/gkab175>

-
- [10] Demissie et al. "Comparative Analysis of Codon Optimization Tools: Advancing toward a Multi-Criteria Framework for Synthetic Gene Design", *J. Microbiol. Biotechnol.*, 35, pp. 1-11, 2025, DOI: <https://doi.org/10.4014/jmb.2411.11066>

Appendice A

Sequenze Ottimizzate

Sequenza irrE ottimizzata per *Escherichia coli* K12:

ATGCCAAGCGCCAACGTGTACCTCCCTGCCCCGTCCGGTGTTCGT
 GGAGGTGGTATGGGCCCCAAAGGCCAAGGCTGAAGCGTCGAAACC
 ACATCCACAAATTCCGGTTAACTTCCGTTCGTGACGGCACCGGA
 CGCGTTAGCCGCCGCTAAGGCCAGAATGCGGGACCTAGCGGCGGC
 GTACGTAGCCGCGTTGCCTGGCCGTGACACGCACAGCCTTATGGC
 TGGCGTTCCGGGCGTTGATCTCAAATTTATGCCGCTTGGCTGGCG
 TGATGGCGCGTTTCGATCCAGAACATAATGTGATTCTGATCAACTC
 AGCGGCCCCGCCCTGAACGCCAGCGCTTTACGCTGGCGCACGAGAT
 CGGCCATGCAATCCTGCTGGGCGATGATGACCTCCTGTCAGATAT
 TCACGATGCGTACGAAGGCCGAACGCCTCGAGCAGGTAATCGAAAC
 CCTGTGTAATGTAGCGGCGGCAGCGATTTTGATGCCTGAACCGGT
 GATCGCGGAAATGTTAGAACGTTTTGGCCCGACCGGACGCGCTCT
 GCGGGAATTAGCAAAACGGGCGGAAGTATCGGCGTCCAGCGCGCT
 CTACGCGCTTACCGAACAGACTCCGGTCCCGGTGATTTATGCGGT
 GTGCGCGCCGGGCAAACCGCCGCGAGAGCAGGCAGCGTCAGATGA
 GGACGCAGGCCCATCTACTGAGAAAGTGCTCACTGTACGTGCGAG
 CTCAAGCACACGCGGCGTTAAATATACTCTGGCGTCAGGCACTCC
 CGTACCGGCAGATCACCCTGCCGCCTTAGCCCTGGCCACCGGCAT
 GGAAGTTCGTGAAGAGAGTTACGTTCCGTTCAGGAGCGGCCGTAA
 GATGAAAGCCGAGGTGGATGCTTATCCTTCGCGTGGAATTGTTGC
 AGTCAGCTTCGAATTTGACCCGGCGCGTCTGGGCCGAAAGGACAG
 TGAGCAAGCCGACAGGGACGAACCCCAAGACGCAGCGCAGTGA

Sequenza irrE ottimizzata per *Anabaena* sp. PCC 7120:

ATGCCCTCAGCCAACGTTTCTCCACCCTGCCCATCTGGTGTCCGC
 GGGGGTGGTATGGGGCCCCAAAGCCAAAGCTGAAGCCTCAAAACCT
 CACCCCCAGATTCCGGTTAAATTACCCTTTGTAACGGCGCCCGAT
 GCTTTAGCCGCTGCTAAAGCTAGAATGCGTGATTTAGCGGCTGCA
 TATGTGGCCGCATTACCAGGTCGTGACACACACAGTCTGATGGCG
 GGGGTGCCTGGAGTTGACCTGAAATTTATGCCACTGGGATGGCGC
 GACGGAGCTTTTGATCCTGAACATAATGTGATTCTAATTAATAGT
 GCCGCACGCCCTGAACGCCAAAGATTTACGTTAGCCCATGAAATT
 GGTACGCGATTCTACTGGGTGACGATGATTTATTGTCTGATATT
 CACGATGCTTATGAAGGTGAACGCTTAGAGCAGGTTATCGAGACA

TTATGTAATGTTGCAGCAGCAGCCATTCTAATGCCAGAACCAGTG
ATTGCTGAGATGCTGGAACGTTTCGGCCCCACCGGGCGTGCACTT
GCCGAACTCGCAAAACGTGCCGAGGTCAGTGCTAGTAGCGCTCTG
TACGCTCTTACAGAGCAAACCTCCCGTTCCTGTGATCTATGCCGTT
TGCGCACCAGGAAAACCTCCACGTGAGCAAGCAGCTAGTGATGAA
GATGCTGGGCCTAGTACGGAAAAAGTATTAACAGTTCGAGCAAGT
AGCAGCACACGCGGTGTTAAATATACACTAGCAAGTGGTACTCCC
GTTCCCGCTGATCATCCCGCAGCGTTAGCCTTAGCTACTGGCATG
GAAGTTAGAGAAGAGTCCTACGTTCCGTTTAGAAGTGGCCGCAAG
ATGAAAGCTGAAGTTGATGCCTATCCGAGCCGTGGCATAGTAGCC
GTAAGTTTTGAATTTGATCCAGCGCGTTTAGGACGTAAGGATAGT
GAACAGGCTGATCGGGACGAACCACAAGATGCTGCTCAATGA