



UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI SCIENZE DEL FARMACO

CORSO DI LAUREA MAGISTRALE IN CHIMICA E TECNOLOGIA
FARMACEUTICHE

TESI DI LAUREA

Affrontare l'epidemia di vaiolo delle scimmie: un protocollo computazionale per l'identificazione di farmaci approvati da riposizionare come inibitori della proteasi virale.

RELATORE: PROF. STEFANO MORO

LAUREANDO: ANDREA DODARO

ANNO ACCADEMICO 2021/2022

SOMMARIO

INTRODUZIONE	9
2022 outbreak.....	10
Malattia	12
Trasmissione	14
Vaccinazione ed eradicazione del vaiolo.....	16
Trattamento.....	16
Orthopox virus	17
Ciclo replicativo e morfologia	18
Proteasi.....	20
Proteasi cisteiniche	21
Proteasi I7.....	23
Origine e scopo della tesi	24
MATERIALI E METODI	26
Hardware & software.....	26
Virtual screening	29
Homology modeling ³⁷	32
De novo prediction ³⁸	35
Meccanica molecolare.....	36
Contributi di legame (<i>Ebonded</i>).....	36
Contributi di non legame (<i>Enon – bonded</i>).....	38
Dinamica molecolare.....	39
Docking	41
- Docking rigido	41
- Docking semi-flessibile	42
- Docking flessibile	42
Funzione di scoring IFP _{CS}	45
RISULTATI E DISCUSSIONE	49
Modellazione della struttura della proteasi principale del Monkeypox.....	50
Mutagenesi	57
Analisi della possibile formazione di un ponte disolfuro	60
Analisi strutturale del sito catalitico	62
Caratterizzazione della binding mode del Ttp6171	65
Virtual screening (Drugbank).....	67
Virtual screening (composti in house).....	77
CONCLUSIONE	81

BIBLIOGRAFIA.....83

INDICE DELLE FIGURE

Figura 1. Casi totali registrati 1° maggio 2022 – 29 novembre 2022 ⁷	10
Figura 2. Casi totali registrati 1° maggio 2022 – 29 novembre 2022 ⁷	11
Figura 3. Sintomi dell'infezione da Monkeypox ⁸	13
Figura 4. Trasmissione del MPV ⁸	15
Figura 5. Formazione dei virioni immaturi ¹⁵	19
Figura 6. Maturazione dei virioni ¹⁵	19
Figura 7. Serine, Cysteine, Aspartic and Glutamic protease mechanism ¹⁸	20
Figura 8. Meccanismo di una proteasi cisteinica ²²	22
Figura 9. Schema dell'organizzazione informatica interna al laboratorio MMS ²⁶	27
Figura 10. Timeline del drug discovery ³⁶	29
Figura 11. Steps in homology modeling ³⁷	34
Figura 12. Porzione di una fingerprint d'interazione	46
Figura 13. Modello di homology modeling di phyre2	50
Figura 14. Modello monomerico di Alphafold	51
Figura 15. Modello dimerico di Alphafold	52
Figura 16. Analisi geometrica del modello monomerico di omologia strutturale.	53
Figura 17. Analisi geometrica del modello monomerico predetto con Alphafold.	54
Figura 18. Analisi geometrica del modello dimerico	55
Figura 19. Sovrapposizione del modello di phyre2 con il modello di Alphafold e la SUMO proteasi Ulp1	56
Figura 20. Capacità delle proteasi mutanti di tagliare le proteine del core virale ¹⁹	57
Figura 21. Mutazione D248A	58
Figura 22. Mutazione Q322A	58
Figura 23. Cys408	61
Figura 24. Cavità prossime alla triade catalitica	62
Figura 25. Conformazione a "ponte"	63
Figura 26. Conformazione aperta	64
Figura 27. Attività del Ttp6171 ²⁵	65
Figura 28. Struttura del Ttp6171 ²⁵	65
Figura 29. Docking pose Ttp6171	66
Figura 30. DB01748	72
Figura 31. DB02429	72
Figura 32. DB04888	72
Figura 33. DB07476	73
Figura 34. DB07531	73
Figura 35. DB07537	73
Figura 36. DB08399	74
Figura 37. DB08754	74
Figura 38. DB12027	74
Figura 39. DB12558	75
Figura 40. DB12882	75
Figura 41. DB13248	75
Figura 42. DB13948	76
Figura 43. DB16236	76

Figura 44. mg2646.....	77
Figura 45. mg2733.....	78
Figura 46. mg2734.....	78
Figura 47. mg2736.....	78
Figura 48. mg2792.....	79
Figura 49. mg2793.....	79
Figura 50. mg2816.....	79

INDICE DELLE EQUAZIONI

Equazione 1. Equazioni di potenziale	36
Equazione 2. Potenziale di stretching	37
Equazione 3. Potenziale di bending	37
Equazione 4. Potenziale dei diedri	37
Equazione 5. Potenziale elettrostatico.....	38
Equazione 6. Potenziale di van der Waals.....	38
Equazione 7. IFPcs scoring function.	47

INDICE DELLE TABELLE

Tabella 1. Software utilizzati durante il percorso di tesi	28
Tabella 2. Distanza rilevata in dinamica (Cys408).....	60
Tabella 3. Drug filter parameters from OpenEye.	70

INTRODUZIONE

Il vaiolo delle scimmie è una zoonosi virale provocata dal virus Monkeypox.

Questo virus è stato identificato per la prima volta nel 1958 a Copenaghen in Danimarca, a seguito del contagio di alcuni macachi provenienti da Singapore.¹ Nel corso degli anni è stato isolato numerose volte e nell'agosto del 1970 si è registrato il primo contagio umano, un bambino di 9 mesi nella Repubblica Democratica del Congo.² Sono proprio i paesi centro-occidentali dell'Africa i luoghi in cui è endemico questo virus. Nonostante questo, è riuscito a diffondersi più volte fuori da questi paesi^{3,4}, suscitando l'interesse nello sviluppare farmaci efficaci per contenere presenti e future epidemie.

2022 outbreak

L'outbreak del Monkeypox nel 2022 ha portato l'Organizzazione Mondiale Della Sanità a definirlo come un'emergenza per la sanità pubblica di interesse internazionale.⁵

Dall'inizio del 2022 al 30 novembre 2022, il Centers for Disease Control and Prevention, ha registrato 81.462 casi di Monkeypox distribuiti in 110 paesi, con 59 decessi. Il primo caso in un paese non endemico si è registrato il 6 maggio nel Regno Unito, quando un residente ritornato dalla Nigeria ha contratto la malattia.⁶ Altri focolai sono stati identificati e i casi sono rapidamente aumentati in tutto il mondo. Questa diffusione senza precedenti ha riportato questa zoonosi all'attenzione della comunità scientifica internazionale.

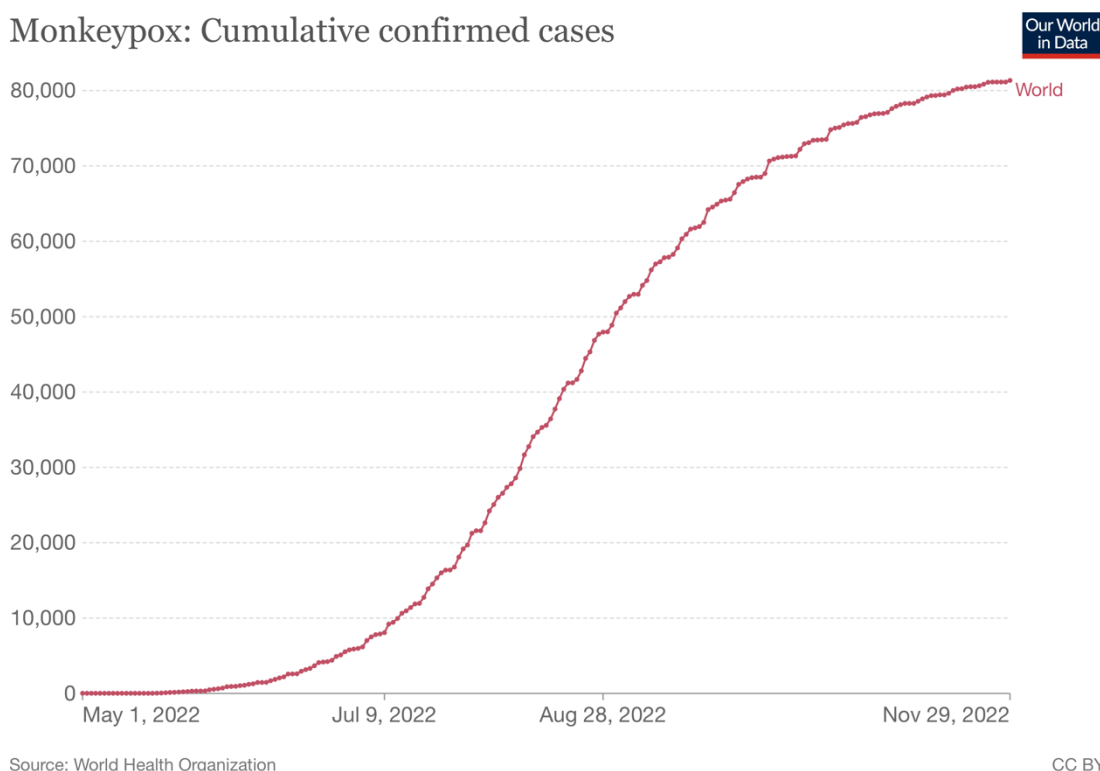
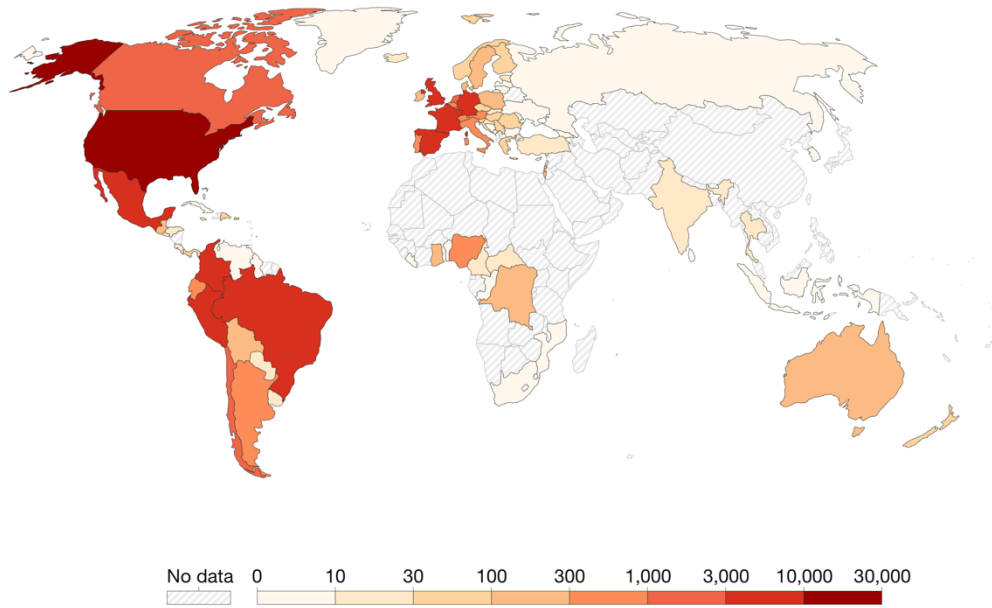


Figura 1. Casi totali registrati 1° maggio 2022 – 29 novembre 2022⁷

Monkeypox: Cumulative confirmed cases, Nov 29, 2022



Source: World Health Organization

CC BY

Figura 2. Casi totali registrati 1° maggio 2022 – 29 novembre 2022⁷

Malattia

Il vaiolo delle scimmie ha un periodo di incubazione compreso tra i 5 giorni e le 3 settimane, i sintomi permangono per 2-5 settimane.⁶

Sebbene abbia una letalità minore rispetto al vaiolo, questa rimane elevata nei paesi in cui è endemico e si stima possa arrivare al 10%.^{3,8,9} Il rischio è maggiore nei bambini, poiché questi tendono a sviluppare la malattia in una forma più grave. Inoltre, anche i soggetti immunodepressi sono esposti a importanti rischi.^{3,8}

Similarmente al vaiolo, la manifestazione più caratteristica della malattia sono i rash cutanei che colpiscono il volto, le mani e la zona genitale, portando a lesioni che possono arrivare ad avere una dimensione di 1 cm di diametro.¹⁰

Spesso chi contrae questa malattia presenta delle lesioni nella cavità orale, come ulcere all'interno della bocca o della gola. Queste lesioni rendono difficoltoso il raggiungimento del corretto apporto nutrizionale.^{4,10}

Inoltre, si registrano anche sintomi gastrointestinali come vomito o diarrea, che possono causare una severa disidratazione, oltre che a un peggioramento dell'apporto nutrizionale.

L'infezione dell'occhio può provocare la perdita della vista a causa della formazione di cicatrici a livello della cornea, mentre le lesioni cutanee, oltre a essere dolorose, possono portare a infezioni secondarie di tipo batterico.⁴

Altri sintomi comuni sono la febbre, linfadenopatia, emicrania, dolori muscolo-scheletrici e dolore ai linfonodi.³

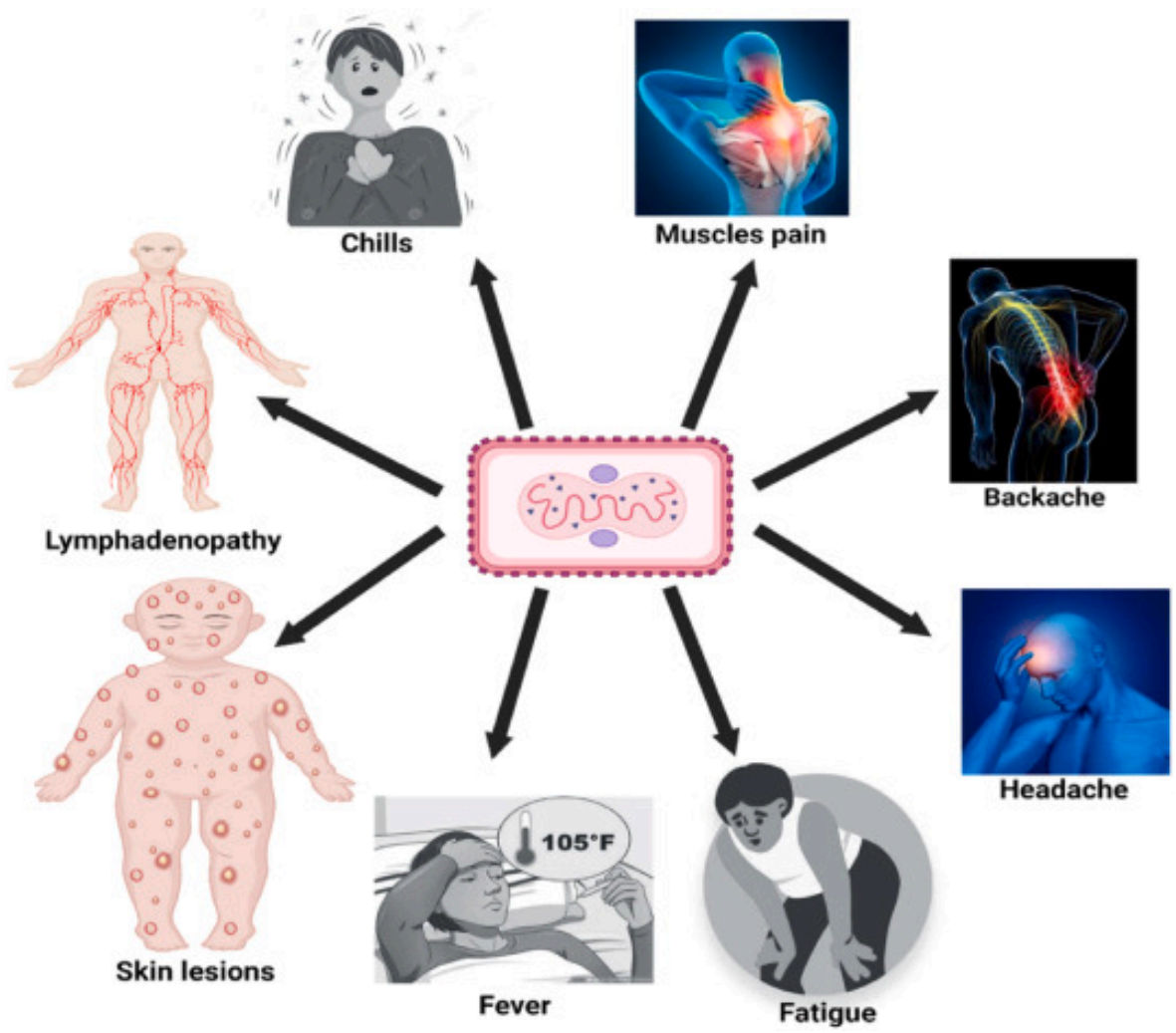


Figura 3. Sintomi dell'infezione da Monkeypox⁸

Trasmissione

La trasmissione può essere di tipo animale-animale, animale-uomo e uomo-uomo.

Il nome Monkeypox deriva dall'animale che ha portato alla sua prima identificazione.

Tuttavia, le scimmie sono degli ospiti secondari: si pensa i principali animali incubatori siano i roditori, anche se questo virus può contagiare un'ampia gamma di animali, tra i quali anche i primati.⁶

Le infezioni umane sono principalmente legate al contatto con questi animali, tramite l'esposizione a fluidi corporei come la saliva, le escrezioni respiratorie, gli essudati cutanei, le lesioni mucose e le feci. Alcuni di questi roditori presenti in Africa, infatti, vengono cacciati, dunque la manipolazione della carne e i morsi da parte di questi animali può condurre all'infezione.⁸

Per quanto riguarda il contagio uomo-uomo, questo può avvenire tramite un contatto stretto e prolungato o attraverso l'esposizione alle lesioni di individui infetti.

Il contatto prolungato faccia a faccia consente l'esposizione a goccioline respiratorie che possono veicolare il vaiolo delle scimmie.⁸

Similarmente alla modalità con cui si sviluppa l'infezione di vaiolo, il virus inizia a replicarsi a livello orofaringeo, arrivando a diffondersi fino ai linfonodi locali, raggiungendo successivamente i linfonodi più distanti e gli organi, attraverso la circolazione sanguigna.^{4,6,8}

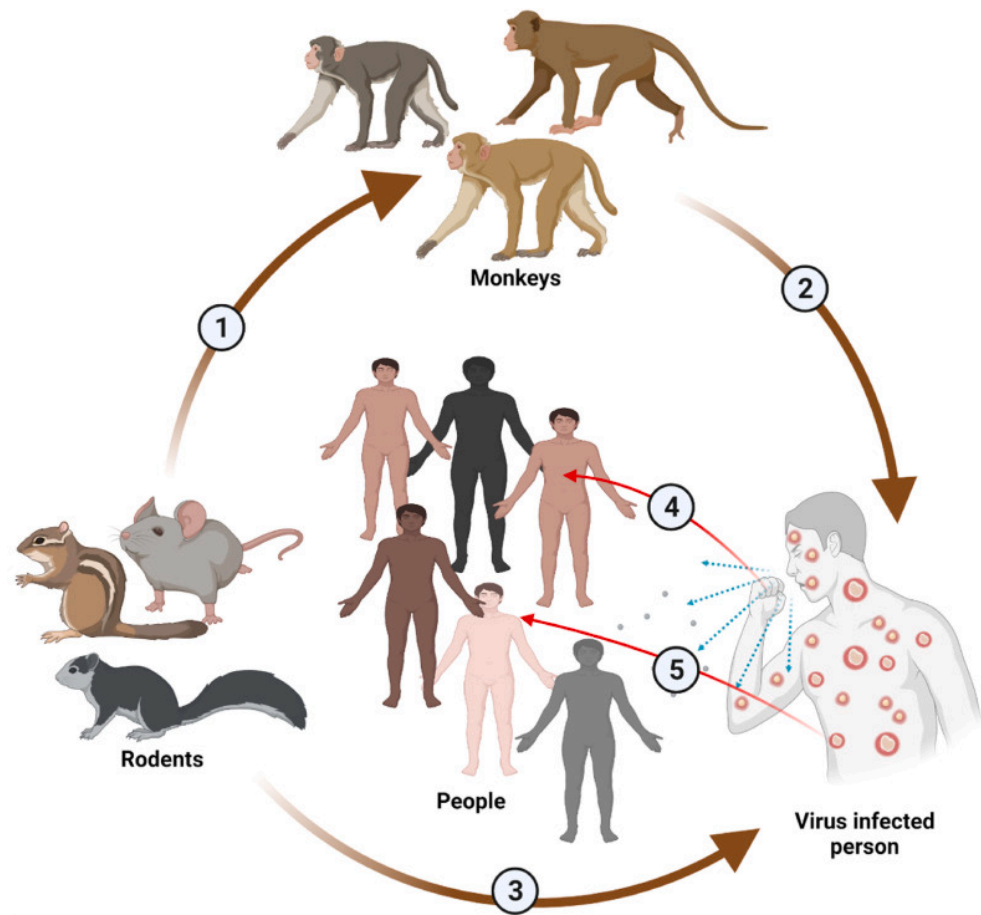


Figura 4. Trasmissione del MPV⁸

Vaccinazione ed eradicazione del vaiolo

Tramite la vaccinazione di massa si è attuato il programma di eradicazione del vaiolo dell'Organizzazione Mondiale della Sanità, che ha dichiarato il virus debellato nel 1979. La vaccinazione contro il vaiolo conduceva in realtà a un'immunità incrociata dato che il vaccino utilizzato si basava sul *vaccinia virus*, un virus appartenente allo stesso genere.⁹ Questa immunità incrociata si estende anche al Monkeypox e si pensa che questo sia uno dei motivi per cui i casi di vaiolo delle scimmie abbiano statisticamente un'incidenza maggiore nella popolazione più giovane che non presenta questa immunità, dato che a seguito del debellamento del vaiolo è venuta meno la necessità di un obbligo vaccinale.^{4,11,12,9}

Trattamento

Non esiste un trattamento sviluppato in maniera specifica per il Monkeypox e data la sua similarità con il *variola virus*, l'attenzione si è spostata agli antivirali usati contro il vaiolo.

Il Tecovirimat è un antivirale approvato contro il vaiolo, agisce inibendo la proteina vp37, andando a bloccare la formazione dell'envelope dei virioni. Nel 2022 è stato approvato dall'EMA contro il Monkeypox.

Il Centers for Disease Control and Prevention ha suggerito l'uso del Cidofovir, un inibitore di DNA-polimerasi. In alternativa, si potrebbero impiegare le immunoglobuline contro il *vaccinia virus* (VIGIV) usate per trattare gli effetti avversi a seguito della vaccinazione con *vaccinia virus*. Sempre il CDC suggerisce come la vaccinazione entro i primi 4 giorni dall'esposizione al virus, grazie al lungo periodo di incubazione virale, potrebbe offrire una buona protezione contro il virus.¹³

Orthopox virus

Il Monkeypox appartiene al genere Orthopoxvirus della famiglia Poxviridae.

I virus appartenenti a questo genere sono di grandi dimensioni, tra i 220 e i 450 nm di lunghezza, e tra i 140 e i 260 nm di larghezza.¹⁴

Questi virus presentano un doppio filamento di DNA e a differenza dei comuni virus a DNA, che si replicano all'interno del nucleo della cellula, questi virus si replicano all'interno del citoplasma e dipendono fortemente dalle proteine virali che esprimono grazie ai ribosomi della cellula infettata.^{9,4}

Date le loro grandi dimensioni verrebbero presto riconosciuti dal sistema immunitario; tuttavia, questi virus possiedono dei geni che esprimono per delle molecole che hanno una funzione modulatrice nei confronti del sistema immunitario.¹⁵

A questo genere appartengono altri virus di grande interesse scientifico: il primo è il *variola virus*, agente eziologico del vaiolo, il secondo è il *vaccinia virus* che è diventato un modello per quanto riguarda lo studio del genere orthopox.⁹

Ciclo replicativo e morfologia

Il virus più caratterizzato in termini di ciclo replicativo e morfologia è il *vaccinia virus*.

Data la complessità strutturale del virus e del ciclo stesso, è necessaria ulteriore ricerca al fine di caratterizzare in maniera esaustiva i meccanismi connessi ad esso.

In termini generali possiamo individuare le seguenti fasi: entrata del virus nella cellula ospite, formazione della cosiddetta factory, formazione della membrana a mezzaluna, formazione del virione immaturo, transizione a virione maturo e formazione di virioni con l'envelope. Le forme virali infettive sono il virione maturo (MV) e il virione extracellulare con l'envelope (EV).

A seguito dell'entrata del virus nella cellula ospite, tramite la fusione della membrana virale con quella citoplasmatica, avviene la dissoluzione del core e la formazione delle "fabbriche virali". L'origine di queste strutture non è del tutto chiara, ma la loro funzione è essenziale per il ciclo virale; infatti, in queste strutture avviene la replicazione del DNA e la formazione del virione.

All'interno delle fabbriche avviene la formazione della membrana a mezzaluna, questa membrana semicircolare si chiuderà poi in un cerchio formando il virione immaturo, che contiene il DNA virale e le proteine del core.

La proteina D13 funge da scaffold per la formazione di questa membrana e interagisce con la proteina A17 per formare il precursore della membrana a mezzaluna.

La rimozione di D13 è necessaria alla formazione del virione maturo, assieme al processamento proteolitico di A17, ad opera della proteasi cisteinica I7.

Questa proteasi è anche responsabile del taglio che porta alla maturazione di alcune proteine del core virale, queste proteine costituiscono il 33% della massa del virione¹⁶. Un'altra proteina importante per la formazione del MV è la metallo proteasi G1, responsabile della condensazione del core.

Dopo l'uscita dalle fabbriche, alcuni virioni maturi si evolvono ulteriormente a EV, caratterizzati dalla presenza di un involucro esterno e coinvolti nell'infezione cellula-cellula.¹⁵

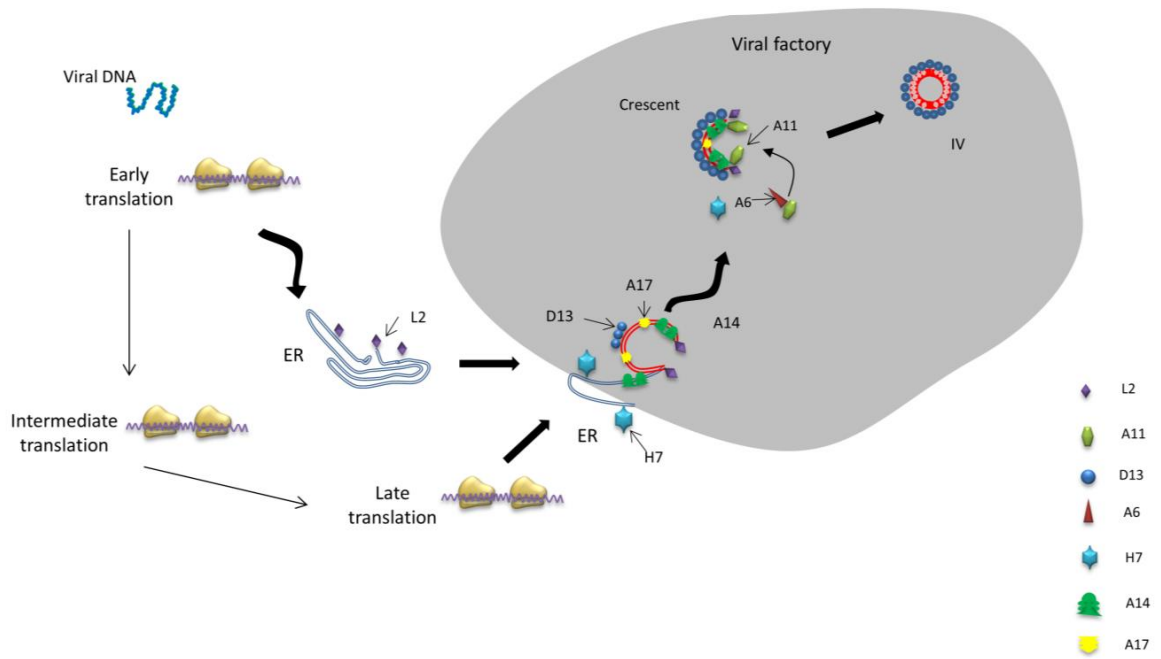


Figura 5. Formazione dei virioni immaturi¹⁵

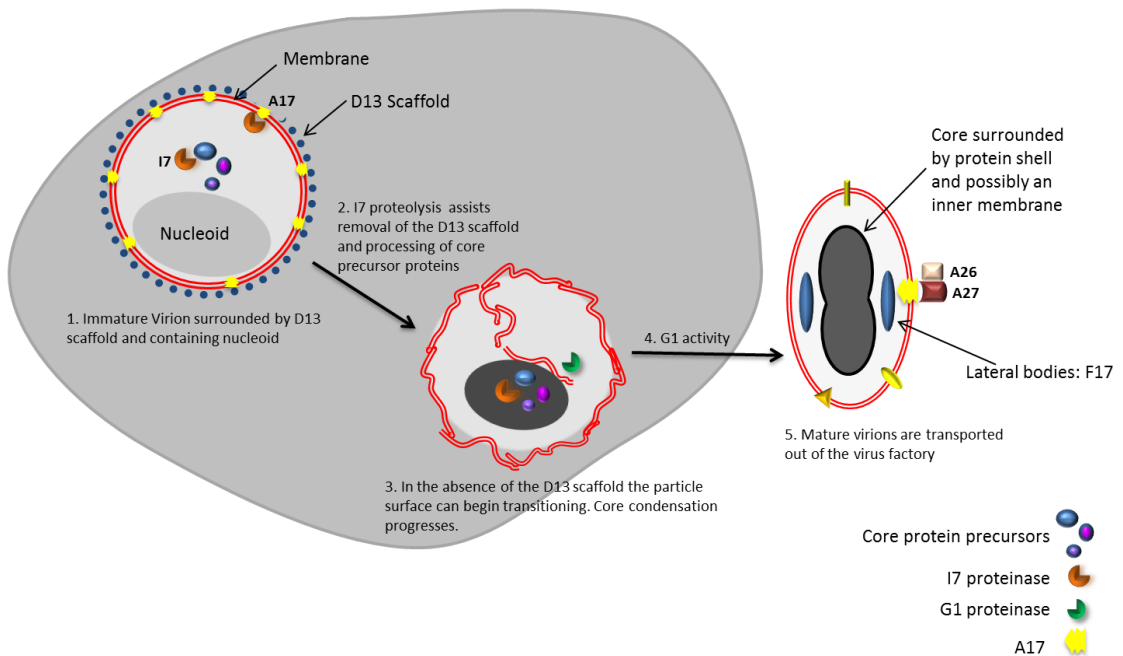


Figura 6. Maturazione dei virioni¹⁵

Proteasi

Le proteasi sono degli enzimi responsabili dell'idrolisi del legame peptidico.

Vengono classificate in sei classi differenti: seriniche, cisteiniche, treoniniche, aspartiche, glutammiche e metalloproteasi. Le prime tre prendono il nome dall'amminoacido con funzione nucleofila direttamente responsabile dell'attacco al legame peptidico, al quale fanno seguito la formazione di un intermedio acilico e la conseguente liberazione del carbossilato dall'estremità amminica, mentre le seconde sfruttano una molecola d'acqua attivata per portare un attacco indiretto al legame peptidico.¹⁷

Nell'ambito delle proteasi virali, possiamo distinguere la proteolisi formativa e quella morfogenica. Per quanto riguarda la proteolisi formativa, essa si riferisce al processamento proteolitico di varie proteine virali, mentre la proteolisi morfogenica consiste nel taglio di proteine virali strutturali, coinvolte nella maturazione dei virioni.¹⁶

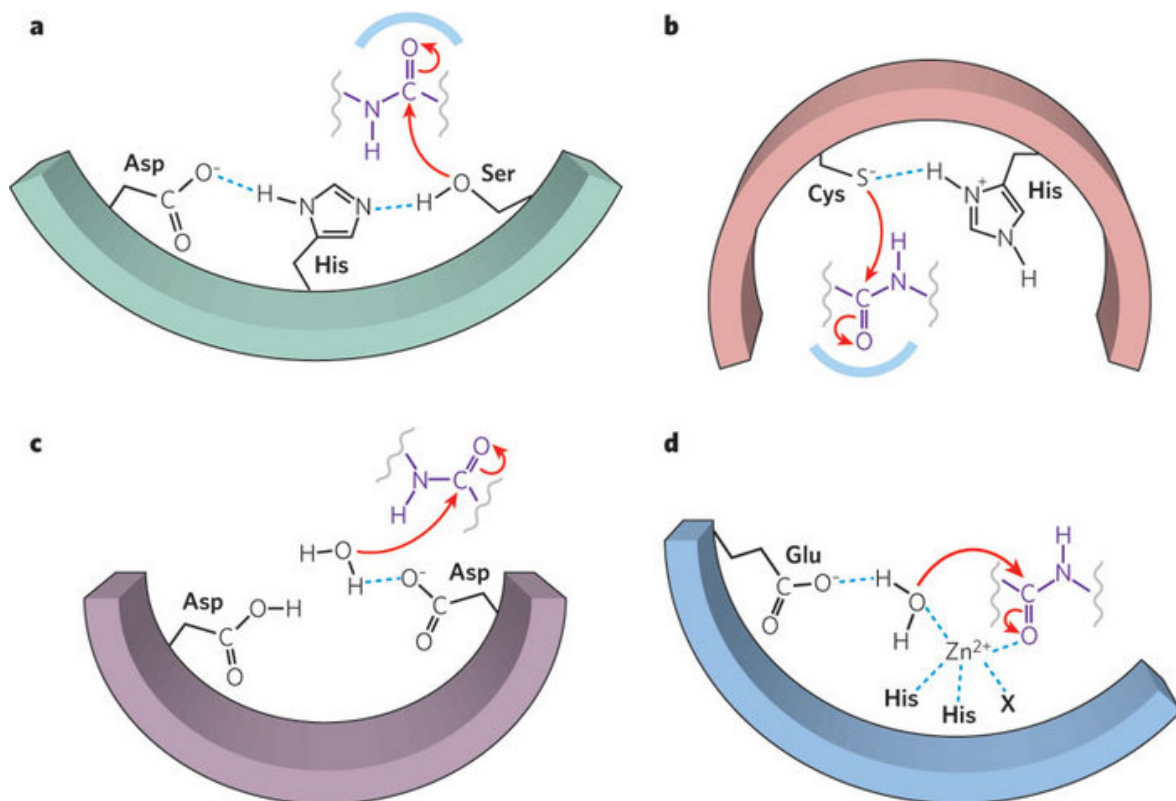


Figura 7. Serine, Cysteine, Aspartic and Glutamic protease mechanism¹⁸

Proteasi cisteiniche

La proteasi i7 del Monkeypox virus è una proteasi cisteinica conservata all'interno del genere orthopox.^{16,19,20}

Le proteasi cisteiniche presentano una triade catalitica composta da una cisteina, da una istidina e da un acido aspartico: quest'ultimo amminoacido non è necessariamente presente poiché la diade Cys-His ha di per sé attività catalitica.

Il taglio proteolitico inizia con la deprotonazione della cisteina catalitica ad opera dell'istidina prossimale, la quale rimuove il protone dal tiolo trasformandolo in tiolato, aumentandone la nucleofilicità. A seguire, avviene l'attacco nucleofilo al legame peptidico con la formazione dell'intermedio covalente tetraedrico e successivamente del tioestere, con la liberazione dell'estremità amminica del peptide.

Generalmente le proteasi seriniche che a livello meccanicistico si differenziano da quelle cisteiniche per l'amminoacido avente funzione nucleofila, necessitano di un buco ossianionico, composto da gruppi ammidici del backbone proteico e catene laterali di amminoacidi positivamente carichi, similmente alle proteasi cisteiniche. Inoltre, necessitano di amminoacidi carichi negativamente come l'aspartato per stabilizzare l'istidina. In termini generali queste proteasi compensano quindi l'energia di attivazione con la stabilizzazione dell'intermedio; differentemente, le proteasi cisteiniche sono più attivate nel loro stato iniziale, dunque non sempre presentano un aspartato. Il tioestere formato viene poi idrolizzato liberando l'estremità carbossilica del peptide e la cisteina in forma tiolica.²¹

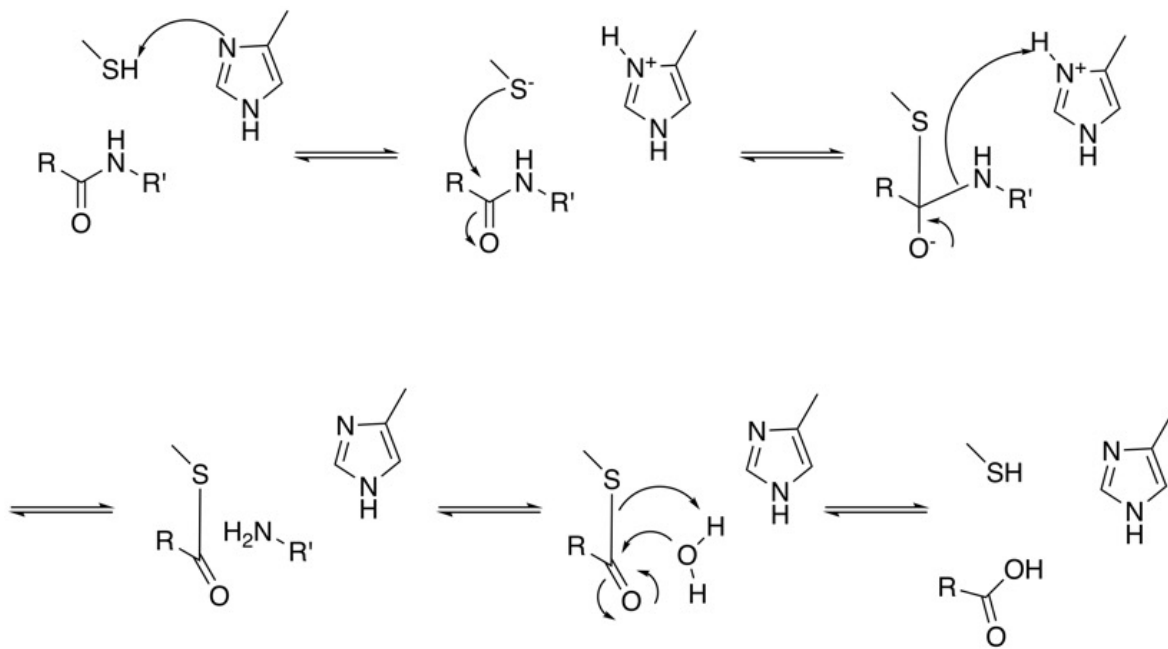


Figura 8. Meccanismo di una proteasi cisteinica²²

Proteasi I7

La proteasi I7 del Monkeypox virus è una proteasi di tipo cisteinico conservata all'interno del genere orthopox^{16,19,20}, l'identità di sequenza di questa proteasi tra il *vaccinia virus* e il *variola virus* è del 99%.¹⁹ Anche le proteine substrato sono altamente conservate all'interno del genere, tre proteine del core virale (P25K, P4a, P4b) che maturano grazie al processamento proteolitico ad opera di I7, differiscono tra il *variola virus* e il *vaccinia virus* in un range compreso tra 1% e 3%. Inoltre, le differenze a livello di sequenza amminoacidica non interessano il sito catalitico.²³

Questa proteasi è coinvolta in processi di proteolisi formativa e morfogenica²⁴: essa processa dunque sia alcune proteine del core virale che la proteina D13 coinvolta nel processo di maturazione dei virioni. Il taglio proteolitico avviene al livello di una sequenza AG/X, dove X indica l'amminoacido in posizione P1' nel peptide substrato: questo amminoacido caratterizzato preferenzialmente da catene laterali di ridotte dimensioni, come nel caso di residui di alanina, serina o treonina.²⁰

Lo studio del taglio del tetrapeptide substrato di I7 ha infatti indicato una chiara preferenza per l'alanina in posizione P2. Per quanto riguarda le posizioni P3 e P4 troviamo un'affinità maggiore per amminoacidi idrofobici, abbiamo poi una preferenza per amminoacidi carichi negativamente in posizione P7 e P8.²³

Dagli studi di mutagenesi effettuati su questa proteasi, è emerso come l'attività catalitica venga persa nel momento in cui viene mutato un amminoacido tra H241, W242, D248, Q322, C328 e G329.¹⁹ Inoltre è stato dimostrato come la dimerizzazione di questa proteasi porti ad un aumento di attività catalitica.²³

Origine e scopo della tesi

Il preoccupante aumento dei casi di vaiolo delle scimmie, registrati a partire dall'inizio del 2022 in tutto il mondo e la conseguente necessità di trovare un trattamento efficace, hanno guidato questo lavoro di tesi verso la strategia del repurposing, ovvero quella di individuare, tramite virtual screening, dei composti già approvati o candidati in fase clinica di sperimentazione da riposizionare contro la proteasi del Monkeypox. In un contesto emergenziale come questo è necessario trovare rapidamente un trattamento, e il riposizionamento permette di ridurre notevolmente i tempi rispetto ad un approccio basato sullo sviluppo razionale. Quest'ultimo caso oltre a richiedere più tempo da un punto di vista metodologico, una volta individuato il composto finale, richiede più tempo da un punto di vista regolatorio.

La scelta della proteasi come target nasce dal successo che questo approccio ha ottenuto con l'HIV e l'epatite C ^{20,25}, oltre che dall'esperienza maturata dal gruppo dell'MMS, guidato dal Professor Stefano Moro, nel campo delle proteasi virali, nello specifico quella del Sars-cov-2.

Il lavoro di tesi nasce dunque dall'intenzione di estendere le competenze acquisite nel campo delle proteasi e dall'urgenza di trovare potenziali candidati per contrastare la diffusione del Monkeypox, in collaborazione con il laboratorio del Professor Cristiano Salata del dipartimento di Medicina Molecolare dell'Università degli studi di Padova, che si occuperà di testare sperimentalmente i composti selezionati.

MATERIALI E METODI

Hardware & software

Il lavoro di tesi è stato svolto all'interno del laboratorio di modellistica molecolare MMS (Molecular Modeling Section) del Dipartimento di Scienze del Farmaco dell'Università degli Studi di Padova, sotto la supervisione del Prof. Stefano Moro. L'infrastruttura informatica del laboratorio comprende:

- 10 user workstation
- 3 nodi di calcolo CPU (64 core ciascuno, per un totale di 640 GB di RAM)
- 7 nodi di elaborazione GPU (21 GPU NVIDIA, serie Titan, GTX e RTX)
- 1 web server
- 3 nodi di storage con 250 TB di capacità

L'architettura software è basata sul sistema operativo Linux Ubuntu, aggiornato alla versione 16.04 LTS e alla versione 20.04 LTS.

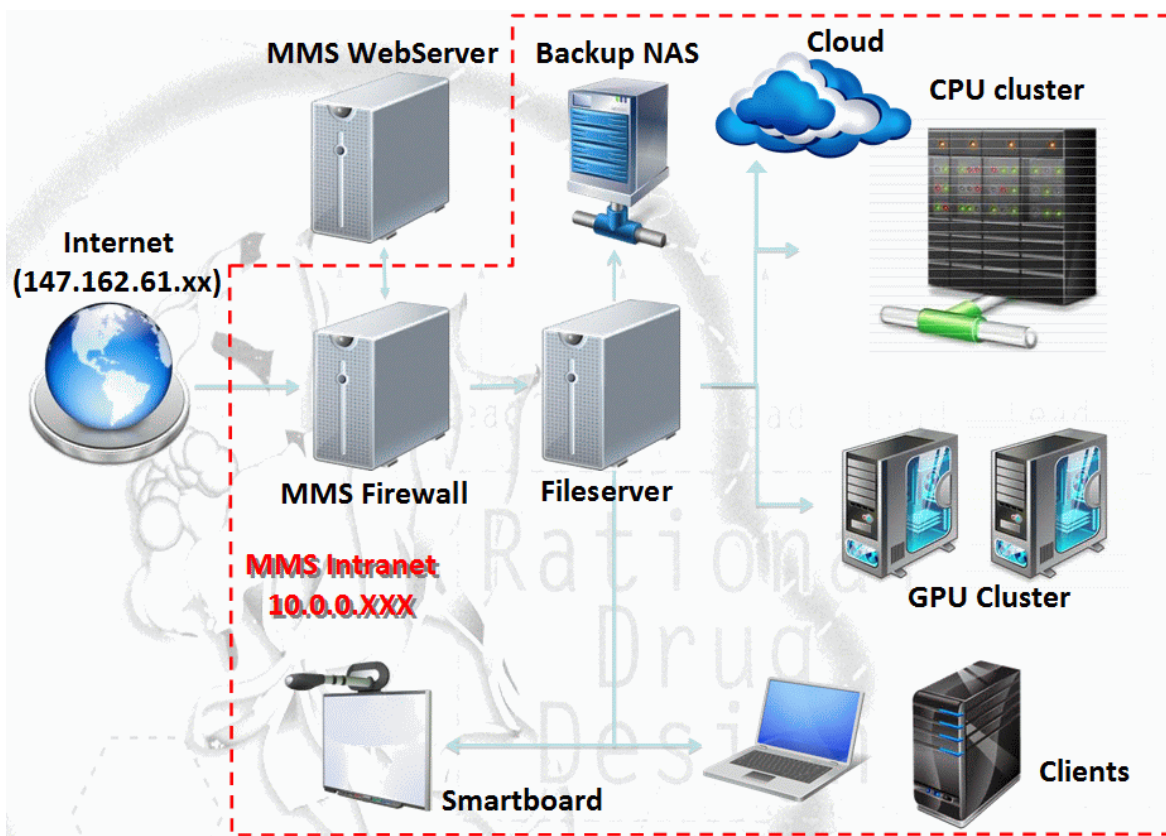


Figura 9. Schema dell'organizzazione informatica interna al laboratorio MMS²⁶

Per l'attività di tesi sono stati utilizzati una serie di script SVL (Scientific Vector Language), BASH (Bourne Again Shell) e Python, oltre ai seguenti software commerciali e non commerciali:

Software	Utilizzo
Alphafold ²⁷	Predizione struttura proteine
Phyre2 ²⁸	Predizione struttura proteine
MOE 2019: Molecular Operating Environment ²⁹	Preparazione strutture proteiche
QUACPAC Tautomers ³⁰	Preparazione ligandi: generazione tautomeri

QUACPAC Fixpka ³⁰	Preparazione ligandi: protonazione idrogeni titolabili
QUACPAC Molcharge ³⁰	Preparazione ligandi: assegnazione cariche parziali
PLANTS ^{31,32}	Docking molecolare
AmberTools ³³	Preparazione sistema per la dinamica molecolare
ACEMD ³⁴	Motore di dinamica molecolare
VMD ³⁵	Gestione delle traiettorie di dinamica molecolare

Tabella 1. Software utilizzati durante il percorso di tesi

Virtual screening

Nelle fasi precoci del drug discovery la fase di screening ricopre un ruolo fondamentale. L'obiettivo è quello di esplorare uno spazio chimico sufficientemente ampio da massimizzare le possibilità di identificare degli hit-compounds in grado di riconoscere il bersaglio molecolare (proteina o acido nucleico) di interesse.

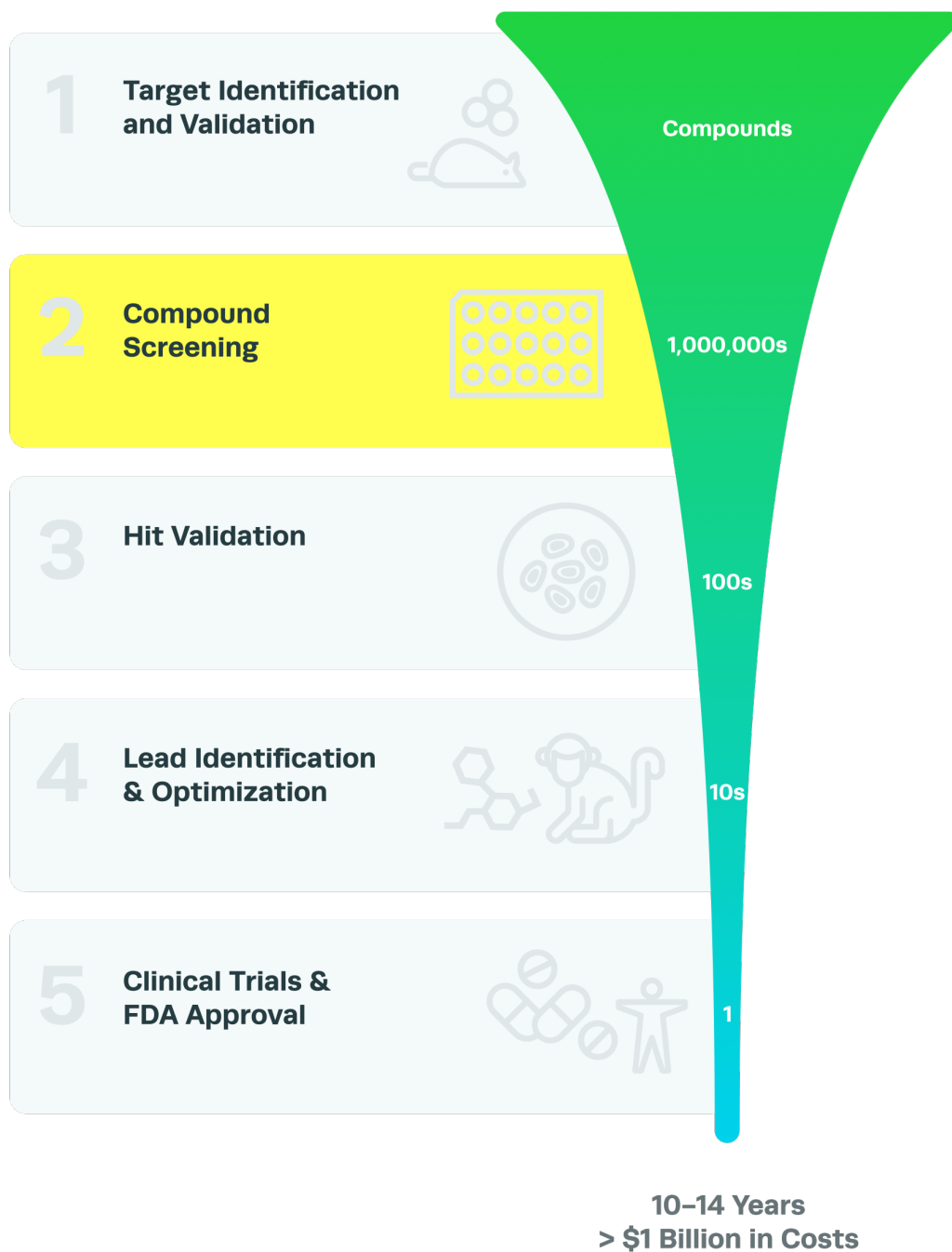


Figura 10. Timeline del drug discovery³⁶

Una delle tecniche più usate è l'HTS (high throughput screening), che permette di testare un numero elevato di composti. Nonostante ciò, questa metodologia presenta numerosi svantaggi, tra cui le difficoltà nel reperimento e nello stoccaggio dei composti da sottoporre a screening, i costi elevati per l'acquisto degli stessi e per l'esecuzione dei saggi e infine i tempi necessari all'effettuazione dello screening stesso. Per ovviare a queste limitazioni dell'approccio HTS, si può premettere alla fase di screening sperimentale una fase di screening in-silico, altresì detta "virtual screening".

Il principale vantaggio del virtual screening consiste nell'elevato numero di composti che possono essere saggiati, decisamente maggiore rispetto a qualsiasi screening sperimentale.

Lo scopo principale di un virtual screening, quindi, è quello di ispezionare database virtuali di milioni di composti con l'obiettivo di individuare tra essi i più promettenti ai fini dell'interazione col target di interesse, riservando a questo ridotto pool di molecole le economicamente e temporalmente dispendiose fasi di screening sperimentale.

Attualmente, data la grande disponibilità di strutture sperimentalmente risolte di biomolecole di interesse terapeutico, e grazie ai grandi progressi effettuati nello sviluppo di algoritmi atti alla predizione di struttura terziaria di proteine, la maggior parte dei virtual screening si basano sul docking, una metodica che permette di prevedere la modalità preferenziale di accomodamento di ligandi all'interno del sito attivo di un determinato target. Ciò avviene perché questa tecnica computazionale rappresenta, ad oggi, il miglior compromesso tra accuratezza nella predizione della bontà dell'interazione farmaco-recettore e la velocità di esecuzione del calcolo, quest'ultima decisamente superiore rispetto a tecniche più raffinate e computazionalmente dispendiose come la dinamica molecolare.

Dal momento che non esiste un protocollo univoco per l'esecuzione di un virtual screening basato sul docking, è importante svilupparne uno solido e affidabile per il sistema in questione, guardando ai dati in letteratura e valutando la capacità del protocollo di selezionare e premiare composti che si sono dimostrati sperimentalmente attivi. Un altro approccio utile alla calibrazione del protocollo è quello di valutare l'accuratezza del software di docking nel riprodurre la binding mode sperimentalmente determinata di composti ad attività nota sul target. Tale confronto può essere basato su un dato puramente geometrico, come lo scarto quadratico medio delle coordinate

atomiche (r.m.s.d.), oppure di natura interattiva, ad esempio tramite il confronto delle fingerprint di interazione.

Un ulteriore aspetto da tenere in considerazione è legato all'hardware che si ha a disposizione: maggiore è la potenza computazionale a disposizione, più è elevato il numero di composti che si possono sottoporre a virtual screening. Al contrario, minore è la disponibilità di risorse computazionali, minore è il numero di composti che potranno essere investigati tramite tale tecnica. Un possibile modo di ovviare a questo problema consiste nel clusterizzare il database di partenza, raggruppando tra di loro composti chimicamente simili tra di loro e scegliendo per ciascun gruppo un elemento rappresentativo dello stesso. In questo modo è possibile ridurre di molto la dimensione della libreria di partenza, pur mantenendo una buona efficienza di campionamento dello spazio chimico.

Homology modeling³⁷

In assenza di una struttura sperimentalmente risolta, per generare una struttura tridimensionale di una proteina si può crearne un modello per omologia, partendo dalla sequenza primaria. Questa tecnica si basa sull'assunzione che la struttura proteica sia determinata dalla sequenza primaria e che sequenze simili si organizzino tridimensionalmente nello spazio similmente tra di loro.

Gli step per la creazione di un modello di omologia sono:

- Identificazione e selezione di un template

Il primo step della creazione di un modello di omologia è la ricerca di un template. Partendo dalla sequenza amminoacidica vengono interrogati dei database come il Protein Data Bank, alla ricerca di proteine con una significativa similarità di sequenza, il cui valore non dovrebbe mai scendere al di sotto del 25%. Altri fattori che vengono considerati per la selezione del template sono l'albero filogenetico, la risoluzione e fattori ambientali quali pH, tipo di solvente e presenza di ligandi.

- Allineamento di sequenza

Dopo la selezione del template, la sequenza target viene allineata con quella del riferimento o con più riferimenti. Questa fase è quella più importante di tutto il processo, di conseguenza un corretto allineamento è essenziale per la bontà del modello.

- Creazione del modello

Le tecniche per la generazione del modello tridimensionale si suddividono in: rigid-body, segment matching, spatial restraint e artificial evolution.

Per quanto riguarda il metodo rigid body, la proteina è scomposta in regioni conservate che vengono assemblate formando la struttura finale. La tecnica del segment matching prevede l'utilizzo delle posizioni atomiche del template come impalcatura sulla quale

aggiungere dei segmenti selezionati per similarità di sequenza. Nelle tecniche di spatial restrain il punto di partenza è sempre il template, sul quale si applicano dei restraints sulla lunghezza di legame, angoli di legame, diedri e interazioni di van der Waals. I metodi di evoluzione artificiale utilizzano la tecnica rigid body, introducendo delle mutazioni fino a quando la sequenza del template diventa come quella target.

- Modellamento dei loop

Le strutture dei loop sono spesso divergenti tra proteine omologhe e la specificità della funzione di una proteina può essere determinata anche da questi loop. Per il modellamento dei loop vengono impiegate due tecniche: la ricerca nei database e la ricerca conformazionale.

- Side chain modeling

Il modellamento delle catene laterali viene operato mettendo le catene laterali sulle coordinate del backbone di una struttura "parente".

- Ottimizzazione del modello

Attraverso l'uso dei campi di forza appartenenti alla meccanica molecolare il sistema viene minimizzato a livello energetico.

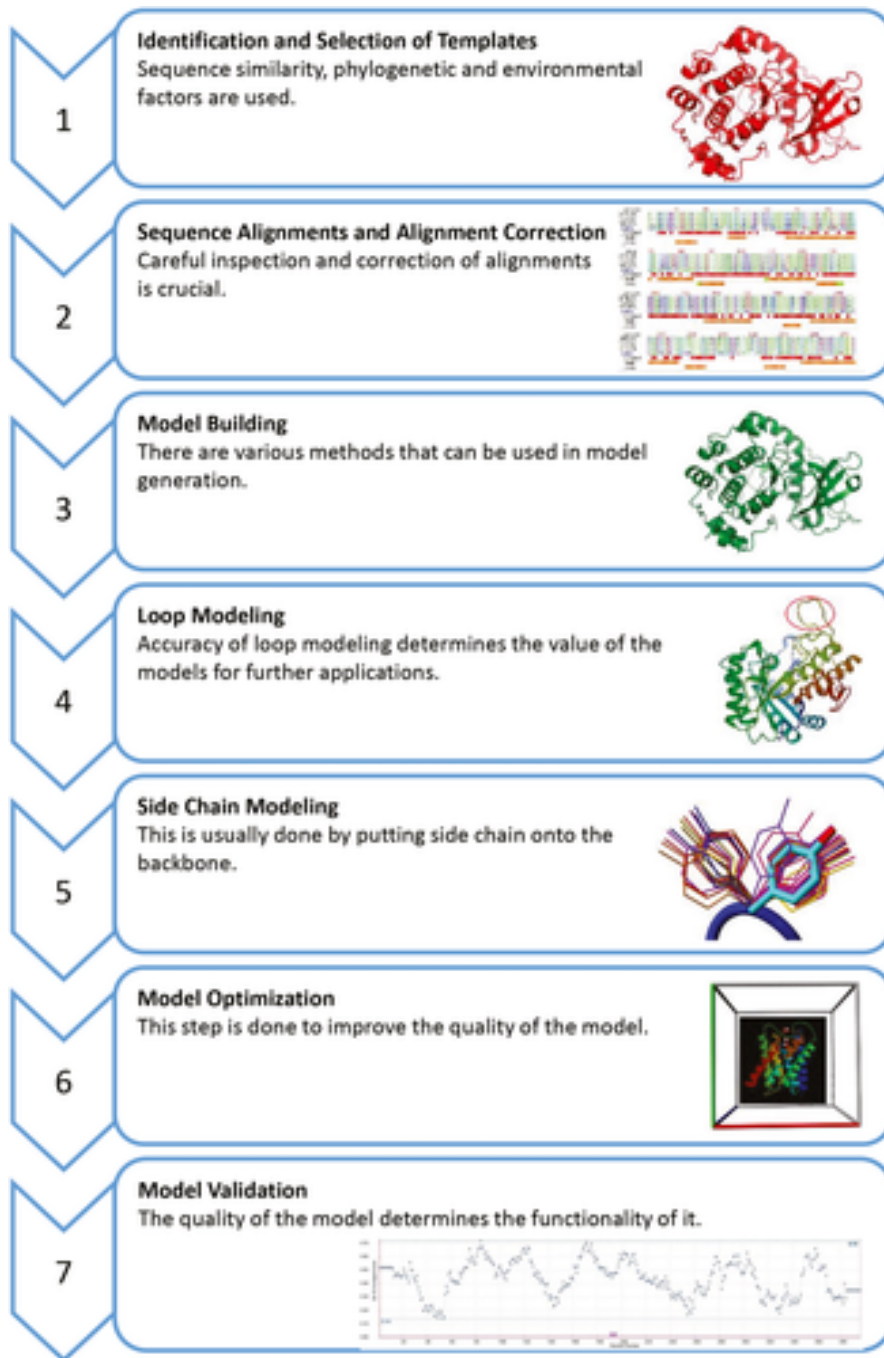


Figura 11. Steps in homology modeling³⁷

De novo prediction³⁸

La de novo prediction è diventata una valida alternativa ai modelli di omologia, permette infatti di predire la struttura tridimensionale di una proteina, utilizzando come input solamente la sequenza primaria e senza bisogno di utilizzare alcun template strutturale. Questo approccio può essere un'importante risorsa nel caso in cui non vi sia una struttura risolta della protein target e l'identità di sequenza con altre proteine omologhe a struttura nota sia troppo bassa.

Ad oggi, il sistema più avanzato ed efficace in questo campo è Alphafold2, un sistema di predizione de novo sviluppato dalla DeepMind. Si tratta della seconda versione di una rete neurale convoluzionale chiamata Alphafold. Nella sua evoluzione, l'architettura convoluzionale è stata sostituita con una attention-based, che permette di imparare il flusso di informazioni in maniera dinamica, tralasciando quelle non rilevanti.

Questa tecnologia si è dimostrata estremamente accurata nel predire strutture proteiche che sono state risolte in seguito alla predizione durante la competizione CASP14 (Critical Assesment of Protein Structure).

Alphafold2 prende come input la sequenza amminoacidica e restituisce come output la struttura predetta, indicando attraverso un codice colore l'accuratezza della predizione in funzione della posizione all'interno della struttura proteica.

Il limite principale nell'accuratezza della predizione fornita da Alphafold risiede nella composizione del training set. Essendo esso basato sulle strutture depositate nel PDB, alcune delle quali possono presentare equilibri conformazionali e/o essere state ottenute in condizioni non naturali che ne favoriscano la cristallizzazione, esse potrebbero non rispecchiare pienamente la reale struttura proteica riscontrabile in vivo.

Meccanica molecolare

Lo studio dell'energia di un sistema molecolare è un aspetto fondamentale della chimica computazionale. Per calcolare l'energia di un sistema si possono impiegare dei metodi basati sulla quantomeccanica oppure dei metodi basati sulla meccanica classica.

I metodi quantomeccanici considerano il sistema nella sua interezza, ma richiedono una grande potenza di calcolo. I metodi basati sulla meccanica classica descrivono gli atomi come entità puntiformi dotate di massa e carica; questa semplificazione rende questi metodi meno dispendiosi da un punto di vista computazionale.

L'energia potenziale di un sistema viene descritta da un insieme di equazioni parametriche chiamato force field. Esistono vari tipi di force field, ognuno dei quali è ottimizzato per descrivere un determinato sistema. Tra i vari campi di forze troviamo CHARMM utile a descrivere sistemi di membrana, troviamo poi AMBER che viene impiegato per la descrizione di proteine solubili e abbiamo infine MMFF che viene usato per la descrizione di piccole molecole organiche.

Ogni campo di forza tiene conto di due contributi per il calcolo dell'energia potenziale del sistema: il contributo di legame e quello di non legame.

$$E = E_{bonded} + E_{non-bonded}$$

$$E_{bonded} = E_{stretching} + E_{bending} + E_{dihedral}$$

$$E_{non-bonded} = E_{electrostatic} + E_{vdW}$$

Equazione 1. Equazioni di potenziale

Contributi di legame (E_{bonded})

- **Stretching**: lo stiramento del legame è dato dalla variazione tra i centri di massa di due atomi coinvolti in un legame. Questa relazione tra la posizione dei centri di massa dei due atomi e la lunghezza del legame, viene descritta attraverso la legge di Hooke, dove

K_r indica la costante di forza, mentre $(r - r_0)^2$ indica la deviazione della lunghezza del legame rispetto alla lunghezza all'equilibrio.

$$E_{stretching} = \sum K_b (r - r_0)^2$$

Equazione 2. Potenziale di stretching

- **Bending**: il bending viene descritto come la variazione dell'angolo tra due legami adiacenti. Come nel caso dello stretching la descrizione matematica è basata sulla legge di Hooke, dove K_θ indica la costante di deformazione angolare, mentre $(\theta - \theta_0)^2$ indica la variazione dell'angolo rispetto all'equilibrio.

$$E_{bending} = \sum K_\theta (\theta - \theta_0)^2$$

Equazione 3. Potenziale di bending

- **Dihedrals torsion**: la torsione dei diedri consiste nella rotazione del legame centrale in un diedro, ovvero un sistema a 4 atomi connessi da 3 legami. Il potenziale torsionale viene creato attraverso la seguente equazione trigonometrica, dove V_n indica la costante torsionale, η rappresenta la molteplicità, φ corrisponde all'angolo diedro e infine γ indica il fattore di fase.

$$E_{dihedrals} = \sum \frac{V_n}{2} [1 + \cos(\eta\varphi - \gamma)]$$

Equazione 4. Potenziale dei diedri

Contributi di non legame ($E_{non-bonded}$)

- **Potenziale elettrostatico:** Questo termine descrive le interazioni tra le cariche parziali degli atomi che compongono il sistema, interazioni descritte dalla legge di Coulomb. In questa relazione q_i e q_j rappresentano le cariche parziali dei due atomi, r_{ij} indica la loro distanza, mentre ϵ_0 indica la costante dielettrica del mezzo.

$$E_{electrostatic} = \sum_i \sum_j \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}$$

Equazione 5. Potenziale elettrostatico

- **Potenziale Lennard-Jones:** questo termine quantifica il contributo energetico dato dalle interazioni di Van der Waals, ovvero le interazioni di dipolo istantaneo. L'equazione di Lennard-Jones è costituita da due termini, uno di natura repulsiva, inversamente proporzionale a r_{ij}^{12} e uno di natura attrattiva, inversamente proporzionale a r_{ij}^6 , dove r indica la distanza interatomica.

$$E_{vdw} = \sum_i \sum_j \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6}$$

Equazione 6. Potenziale di van der Waals

Dinamica molecolare

La dinamica molecolare è una tecnica computazionale che permette di simulare l'evoluzione tempo dipendente di un sistema molecolare.

Sistemi con un numero ridotto di atomi possono essere descritti dalla quantomeccanica tramite la risoluzione dell'equazione di Schrödinger tempo dipendente; tuttavia, per sistemi più complessi non è possibile utilizzare questa descrizione quantistica, bisogna dunque approssimare la descrizione del fenomeno. Inoltre, nella descrizione quantomeccanica vengono considerate tutte le particelle del sistema come i protoni, i neutroni e gli elettroni, questo richiede una grande potenza computazionale e lunghi tempi di calcolo.

Il sistema molecolare viene quindi descritto come una serie di masse puntiformi legate da molle, e l'integrazione dell'equazione del moto newtoniana permette di valutare l'evoluzione temporale del sistema. Abbracciando questa descrizione appartenente alla meccanica classica, vengono accorciati i tempi necessari a completare simulazioni nell'ordine di grandezza dei nanosecondi.

In una simulazione di dinamica l'energia potenziale del sistema viene approssimata da una funzione che prende il nome di campo di forze. Questo campo è costituito dai termini energetici riguardanti le interazioni intramolecolari e intermolecolari. Tra le interazioni di legame troviamo lo stretching e i bending; per quanto riguarda le interazioni di non legame abbiamo l'interazione di van der Waals e le interazioni elettrostatiche coulombiane.

Per iniziare una simulazione di dinamica bisogna disporre delle coordinate degli atomi del sistema: queste vengono ottenute dalla struttura cristallografica, NMR o dal modello predetto. Successivamente, le strutture vengono minimizzate e ad ogni atomo viene assegnata una velocità randomica, ottenuta dalla distribuzione di Maxwell-Boltzmann, che porti il sistema ad una energia cinetica media che corrisponda ad una determinata temperatura, usualmente 300 K. In seguito, vengono eseguite delle simulazioni di pochi picosecondi e attraverso la legge di Newton e l'uso del campo di forze, viene calcolata la posizione atomica e riassegnata la velocità di ciascun atomo in maniera reiterata fino a quando dopo una serie di passaggi queste informazioni vengono salvate, andando a

costituire un timestep, ovvero un frame della traiettoria finale espressa come un video. La traiettoria finale sarà quindi data dall'unione di ogni frame in sequenza.

I progressi informatici legati all'hardware e in particolar modo alle GPU, hanno permesso di estendere la scala temporale delle simulazioni. Tuttavia, l'osservazione degli eventi di binding in una simulazione classica richiede tempi di simulazione nell'ordine dei μs o dei ms, che ad oggi, risultano ancora proibitivi. Per questo motivo sono state sviluppate delle tecniche di enhanced sampling che permettono di ridurre i tempi di calcolo rispetto alla dinamica molecolare classica.

Docking

Il docking è una tecnica computazionale che permette di indagare e predire la modalità di interazione ligando-proteina. Il punto di partenza di questa tecnica è la struttura della proteina target, che può essere ottenuta attraverso tecniche cristallografiche, NMR e Cryo-EM. In assenza di una struttura risolta sperimentalmente, si può ricorrere alla costruzione di un modello di omologia o, in alternativa, alla predizione de novo, avendo così la possibilità di lavorare su target proteici disponendo unicamente della sequenza primaria.

Una volta ottenuta una struttura, viene definito un volume da esplorare attraverso il raggio di una sfera o il lato di un cubo, includendo la porzione proteica di interesse per il binding. Vengono poi generate una serie di conformazioni e orientazioni nello spazio di ciascun ligando che prendono il nome di pose di docking, esplorando i gradi di libertà traslazionale, rotazionale e conformazionale dei composti. È fondamentale bilanciare la qualità di questo sampling con il tempo che serve a performare i calcoli necessari e tenere in considerazione la quantità di pose generate per ogni ligando e la loro diversità in termini di RMSD, al fine di ottimizzare la ricerca delle pose in relazione allo sforzo temporale e computazionale. Ad ogni pose viene associato un punteggio che indica la bontà riconoscimento tra ligando e sito attivo, in termini di qualità e quantità di interazioni stabilite col target. Ogni protocollo di docking è formato dunque da due componenti: un algoritmo di ricerca e una scoring function.

Gli algoritmi di ricerca si suddividono in:

- Docking rigido

I protocolli di docking rigido sono stati tra i primi a essere sviluppati, in questi protocolli sia la proteina che il ligando sono considerati come delle entità rigide. Per ridurre lo sforzo computazionale sono presi in considerazione solamente tre gradi di libertà rotazionali e traslazionali.

Questi protocolli descrivono efficacemente il cosiddetto meccanismo chiave-serratura, tuttavia il riconoscimento ligando-recettore è un fenomeno più dinamico e complesso che viene descritto accuratamente da nuovi protocolli non rigidi. Questo restringe il campo di applicazione di questa metodica al riconoscimento proteina-proteina, in cui i gradi di libertà conformazionale del sistema sarebbero troppo numerosi e computazionalmente dispendiosi se le due proteine non venissero considerate come entità rigide.

- Docking semi-flessibile

I protocolli di docking semi-flessibile sono stati sviluppati per superare le limitazioni di quelli rigidi e sono largamente impiegati nell'ambito dei virtual screening.

In questo caso viene considerata, oltre ai sei gradi di libertà roto-traslazionale, la libertà conformazionale del ligando data dall'esplorazione dei diedri rotabili del ligando.

Similarmente ai protocolli rigidi la proteina viene mantenuta rigida, di conseguenza il risultato del docking dipende fortemente dalla conformazione della proteina utilizzata.

- Docking flessibile

Questi protocolli tengono conto della flessibilità sia del ligando che della proteina, descrivendo il meccanismo induced-fit e selezione conformazionale. Tra le tecniche precedentemente descritte è la più versatile ma anche quella che richiede più potenza di calcolo e tempo. Per questo motivo il suo impiego è limitato e ostacolato da altre tecniche come la dinamica molecolare che descrivono efficacemente la dinamicità di un sistema ligando-proteina.

Gli algoritmi di scoring si suddividono in:

- Scoring function basate sul force field

Il campo di forze è un aspetto fondamentale della meccanica molecolare, in quanto esso approssima l'energia potenziale di un sistema ad una combinazione di interazioni intramolecolari e intermolecolari. Nel docking, differentemente dalla dinamica, viene considerata principalmente la componente intermolecolare, insieme alla componente torsionale del ligando. Queste scoring function includono principalmente due componenti: il potenziale elettrostatico descritto tramite la funzione di Coulomb e l'energia di van der Waals descritta dal potenziale di Lennard-Jones.

Un esempio di scoring function basata sul campo di forze è GoldScore, impiegata nel protocollo di docking GOLD, un'altra scoring function appartenente a questa categoria è AutoDock appartenente all'omonimo protocollo di docking.

- Scoring function empiriche

Queste scoring function costituiscono un buon compromesso tra la qualità del punteggio assegnato e la velocità di valutazione delle pose. Sono costituite dalla somma di più termini energetici ottenuti empiricamente, tra tutti ricordiamo il contributo elettrostatico, di van der Waals, di desolvatazione, entropico, idrofobico, legami a idrogeno ecc. Sulla base di un training set, vengono definiti dei coefficienti che pesano questi contributi.

Alcuni esempi di scoring function empiriche sono ChemScore, utilizzata nel protocollo di docking GOLD, e ChemPLP impiegata nel protocollo PLANTS.

- Scoring function knowledge-based

Queste scoring function si basano sull'assunzione che le interazioni ligando-proteina che vengono maggiormente osservate da un punto di vista statistico all'interno di grandi

database di strutture tridimensionali, come il Protein Data Bank, siano quelle energeticamente più stabili e favorevoli. Vengono quindi comparate le interazioni intermolecolari a corto raggio tra ligando e proteina con quelle presenti nei database, infine viene assegnato un punteggio delle pose ottenute. Tra le tipologie di scoring function precedentemente citate, queste sono le meno impiegate. Un esempio di scoring function knowledge based è la GOLD-ASP.

- Scoring function basate sul Machine-Learning

Le scoring function tradizionali possiedono una forma funzionale predeterminata. Nel caso del machine learning essa viene dedotta.

Queste scoring function sono state impiegate per determinare relazioni non lineari tra descrittori e affinità di binding. Rispetto alle scoring function tradizionali, quelle basate sul deep learning performano meglio per quanto riguarda la correlazione tra punteggio assegnato e affinità di binding sperimentale, ma nonostante questo le performance nei virtual screening è inferiore rispetto alle tradizionali.

Una branca del machine learning definita deep learning, ha ottenuto incredibili risultati nel riconoscimento di immagini e del linguaggio, grazie alle reti neurali convoluzionali e ricorrenti. Si pensa quindi che, con il passare del tempo, queste nuove metodiche di scoring possano essere impiegate anche nei virtual screening.

Funzione di scoring IFP_{cs}

Le comuni funzioni di scoring si basano sui contributi energetici delle interazioni di legame e di non legame. Ne consegue che molecole di grandi dimensioni e dunque ad alto peso molecolare, formeranno più interazioni aspecifiche con il target rispetto a molecole di dimensioni e peso inferiore. Queste molecole verranno quindi favorite per quanto riguarda lo scoring, senza tenere conto della reale specificità di riconoscimento dei composti nei confronti del target di interesse.

L'utilizzo di una scoring function basata sulle fingerprint di interazione permette di superare questo ostacolo, poiché le fingerprint sono meno influenzate dal peso molecolare e dalle dimensioni del ligando.

La fingerprint d'interazione consente di registrare l'informazione relativa al pattern di interazione tra proteina e ligando in una stringa monodimensionale. Questa stringa numerica da un punto di vista computazionale è estremamente vantaggiosa, poiché consente di immagazzinare e manipolare questa informazione in un linguaggio matematico.

Una fingerprint d'interazione è costituita da dei segmenti, uno per ogni residuo amminoacidico della proteina, e ogni segmento è composto da otto celle che codificano per otto tipi di interazione:

- contatto idrofobico
- stacking π - π *face-to-face*
- stacking π - π *edge-to-face*
- ponte idrogeno in cui la proteina funge da donatore
- ponte idrogeno in cui la proteina funge da accettore
- ponte salino in cui la proteina funge da specie carica positivamente
- ponte salino in cui la proteina funge da specie carica negativamente
- legame ionico con uno ione metallico.

Ogni cella indica il numero di atomi del residuo coinvolti nell'interazione, mentre il tipo di interazione è dettato dalla posizione nella sequenza del segmento. Una cella viene settata come "spenta" se il suo valore è pari a zero.

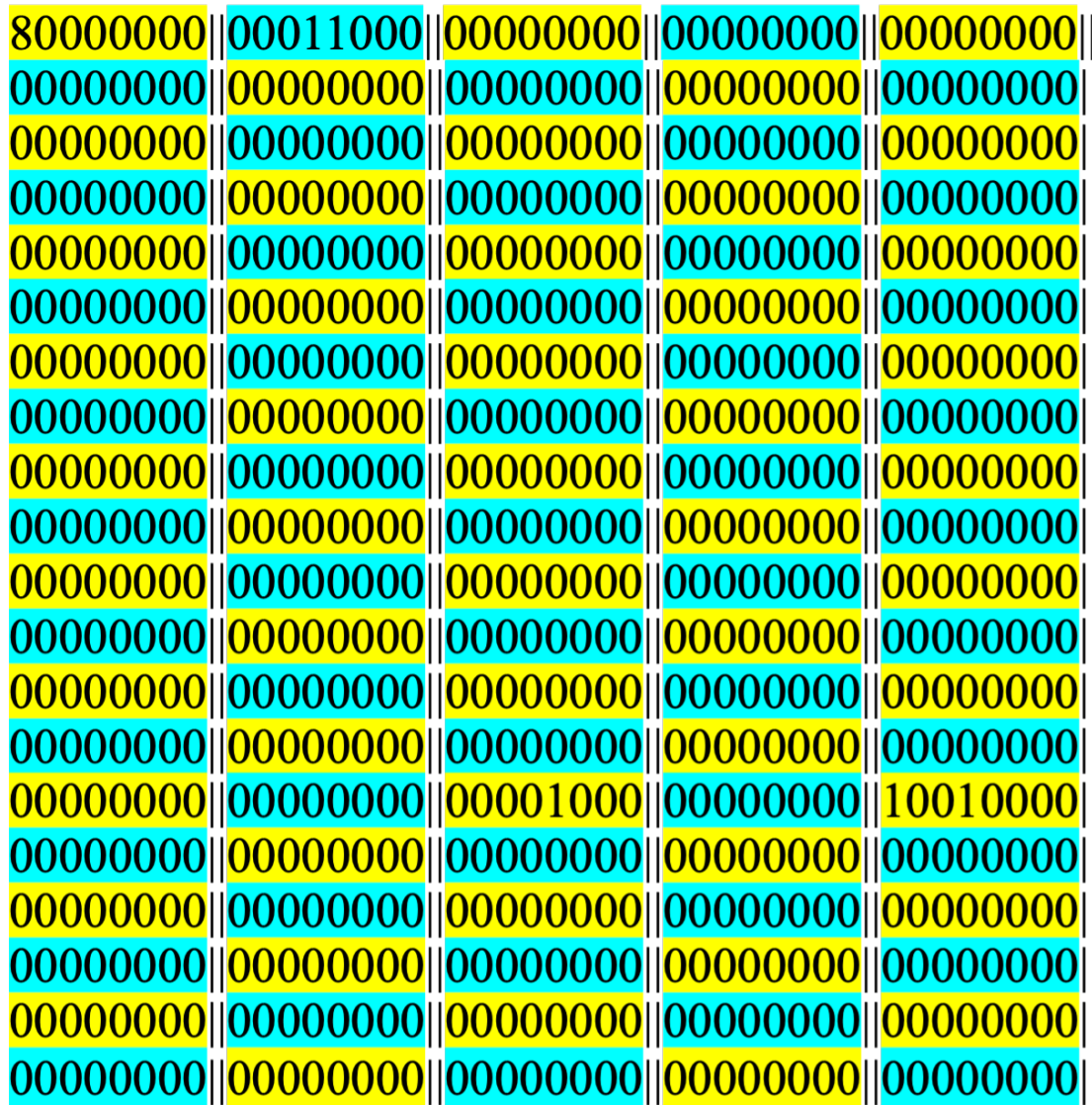


Figura 12. Porzione di una fingerprint d'interazione

L'informazione relativa al pattern di interazione tra la proteina e un ligando può essere comparata tra ligandi diversi. Per attuare questa comparazione le sequenze delle fingerprint vengono viste come dei vettori nello spazio e viene calcolata la similarità del

coseno, ovvero il coseno dell'angolo compreso tra i due vettori, quindi il prodotto scalare dei vettori diviso il prodotto delle loro lunghezze. Questa operazione porterebbe ad una bassa risoluzione se operata sulla fingerprint intera, infatti la quasi totalità della proteina, esclusa la porzione relativa al sito attivo, non interagisce con il ligando, di conseguenza i segmenti corrispondenti ad ogni residuo amminoacidico avrebbero ciascuno otto celle spente. Questo porterebbe ad una perdita di risoluzione, poiché il coseno tenderebbe al valore minimo. Per questo motivo le celle nulle analoghe, corrispondenti alla stessa interazione per il medesimo residuo di due fingerprint diverse, vengono eliminate creando così delle fingerprint "dense" che vengono impiegate nel calcolo della similarità del coseno.

$$IFP_{cs} = \frac{A \cdot B}{\| A \| \| B \|}$$

Equazione 7. IFPcs scoring function.

RISULTATI E DISCUSSIONE

Questo progetto di tesi ha portato all'identificazione di 14 potenziali inibitori della proteasi del Monkeypox, attraverso la determinazione di uno specifico protocollo di virtual screening basato sul docking, applicato ad un database di composti approvati o candidati farmaci in fase clinica di sperimentazione. La mancanza di un trattamento efficace e la necessità di contenere rapidamente la diffusione del virus, ci hanno portati a considerare la strategia del riposizionamento che si è già dimostrata efficace in scenari simili.³⁹ Grazie all'esperienza maturata con il primo virtual screening, ne è stato effettuato un secondo, questa volta su un database di composti interni sintetizzati nel laboratorio della professoressa Ferlin dell'Università di Padova, con l'obiettivo prospettico di trovare uno scaffold da cui partire per lo sviluppo razionale di farmaci specifici contro la proteasi, in una visione di lungo termine per contrastare eventuali future pandemie.

Modellazione della struttura della proteasi principale del Monkeypox

Dal momento che un virtual screening basato sul docking è fondato sulla conoscenza della proteina target e data la mancanza, ad oggi, di una struttura risolta sperimentalmente della proteasi del Monkeypox, ne sono stati creati tre differenti modelli, sia tramite homology modeling che tramite predizione *de novo*.

Il primo, consiste in un modello di omologia creato attraverso il webserver Phyre2, questo modello monomero è basato sulla SUMO proteasi Ulp1 del lievito *Saccharomyces*. Questa SUMO proteasi presenta una identità di sequenza del 17% con la proteasi del Monkeypox: una tale percentuale potrebbe indicare che il modello porti globalmente ad una approssimazione eccessiva a livello strutturale. Tuttavia, questo modello è già stato utilizzato in passato con successo per individuare dei candidati inibitori.^{20,25} Questo può essere spiegato da una elevata similarità localizzata a livello del sito catalitico, come suggerito dal fatto che il taglio proteolitico dei peptidi substrato avviene al livello di una sequenza GG/X, similmente alla proteasi del Monkeypox che presenta invece una specificità per la sequenza AG/X.

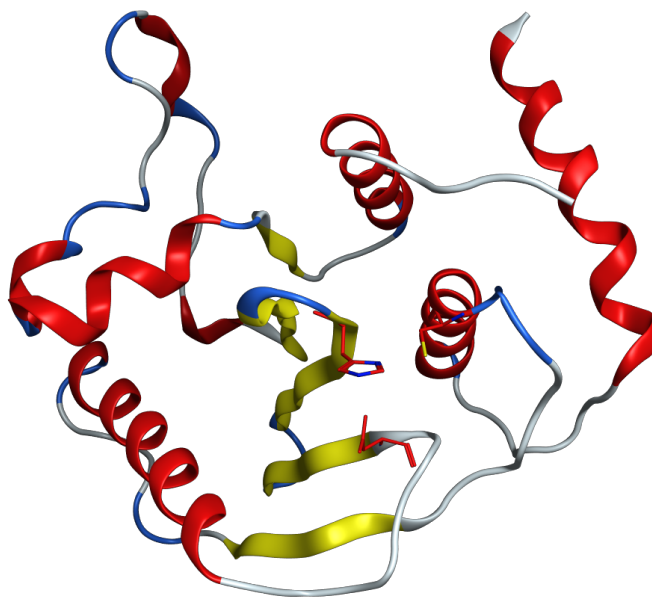


Figura 13. Modello di homology modeling di phyre2

Per ovviare all'eccessiva approssimazione strutturale a livello di porzioni della proteina distali rispetto al sito catalitico, ne sono stati creati altri due modelli attraverso la predizione de novo. Questa tecnica permette di creare una struttura terziaria o quaternaria solamente a partire dalla sequenza amminoacidica.

Il secondo modello è stato sviluppato attraverso Alphafold, un potente programma di predizione strutturale basato su intelligenza artificiale.

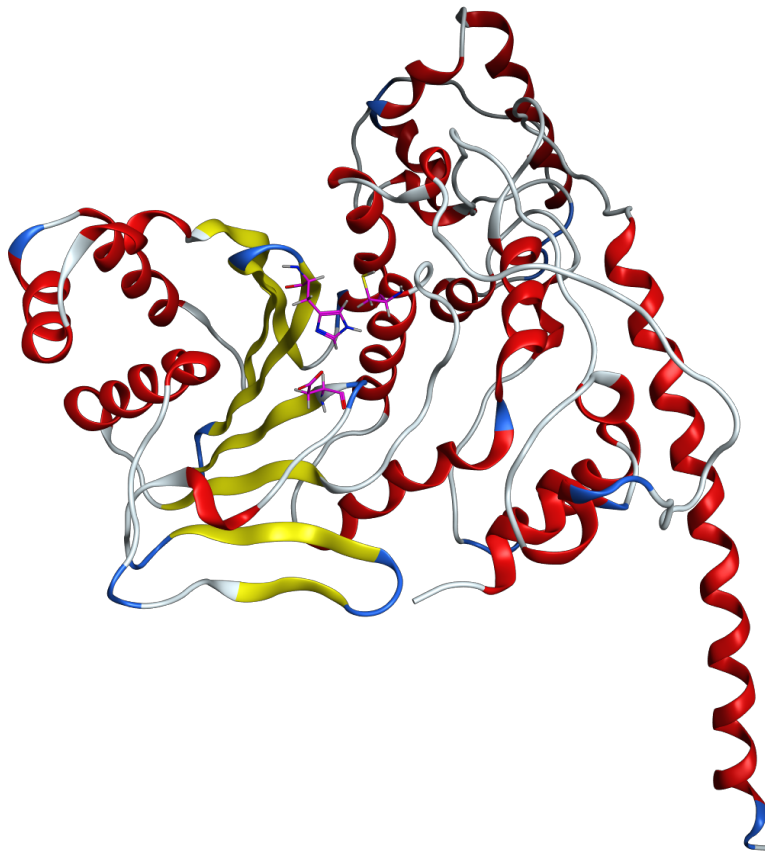


Figura 14. Modello monomero di Alphafold

Seguendo i dati di letteratura che indicano una dimerizzazione della proteasi con conseguente aumento di attività catalitica²³, è stato creato un secondo modello attraverso la predizione de novo, sfruttando nuovamente Alphafold, per descrivere la proteina nello stato dimerico.

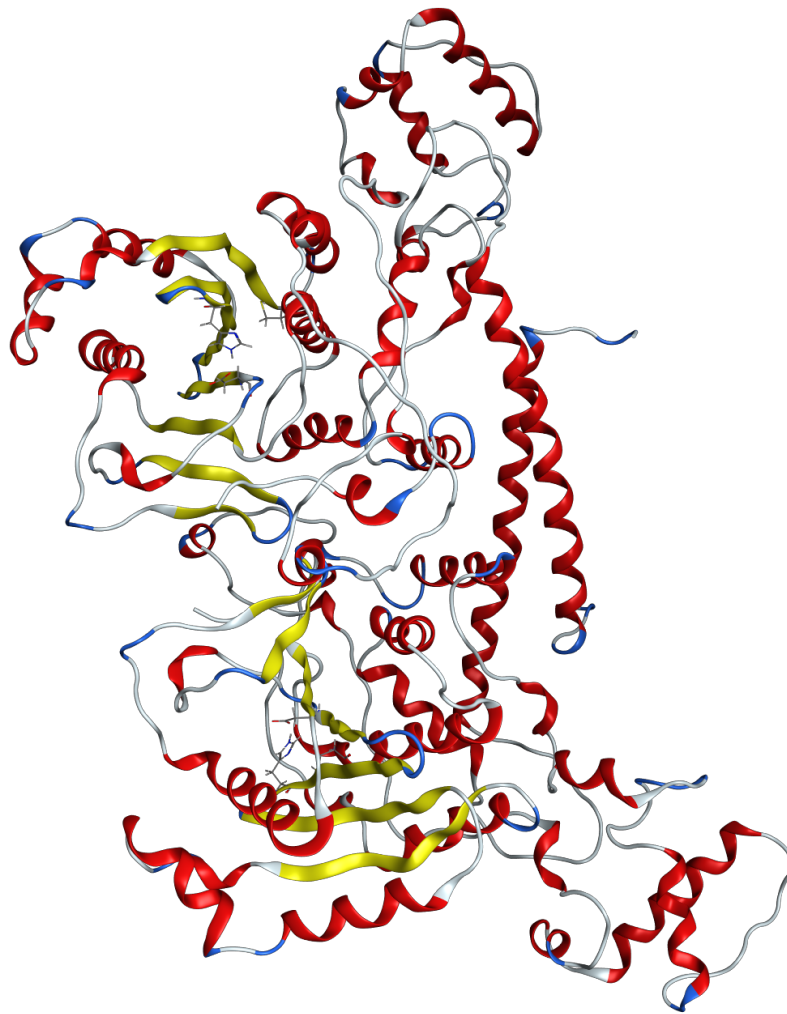


Figura 15. Modello dimerico di AlphaFold

I modelli sono stati preparati attraverso il tool Preparation di MOE. Nelle strutture di partenza non sono presenti i protoni, di conseguenza sono stati aggiunti attraverso il tool Protonate 3D. Dato che questi protoni vengono aggiunti in modo tale da soddisfare valenza, stato di protonazione e tautomeria previsto al pH target, ma non necessariamente in una posizione spaziale tale da minimizzare l'energia potenziale del sistema, è necessario minimizzare le loro coordinate da un punto di vista energetico attraverso lo strumento Energy Minimize con campo di forze AMBER 10:EHT. Lo stesso campo di forze è stato inoltre utilizzato per calcolare le cariche parziali del sistema.

Per valutare la stabilità e la qualità dei modelli, ognuno di questi è stato fatto poi rilassare in dinamica molecolare. Sono state effettuate quindi 3 simulazioni della durata di 100ns ciascuna per ogni modello. La stabilità dei modelli nel corso di queste traiettorie è stata poi analizzata tramite il calcolo di descrittori geometrici quali r.m.s.d. ed r.m.s.f. del backbone proteico.

Come è possibile osservare nei grafici riportati in Figura 16, Il modello di omologia sviluppato tramite phyre2 ha dimostrato di non avere una buona stabilità in dinamica, raggiungendo valori di RMSD intorno ai 7 Å; ciò può essere spiegato dall'intrinseca instabilità di un modello troncato come questo.

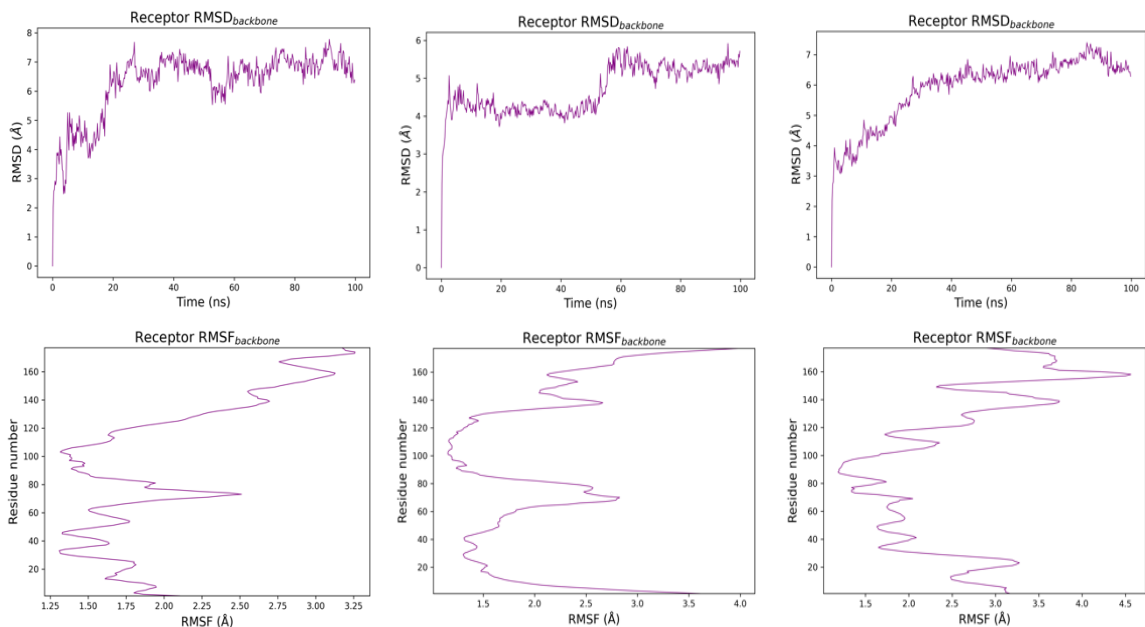


Figura 16. Analisi geometrica del modello monomero di omologia strutturale.

Come previsto, e come osservabile dall'andamento dei grafici in Figura 17, il modello monomero predetto da AlphaFold è più stabile rispetto al modello di phyre2, sia in termini di RMSD, sia in termini di RMSF.

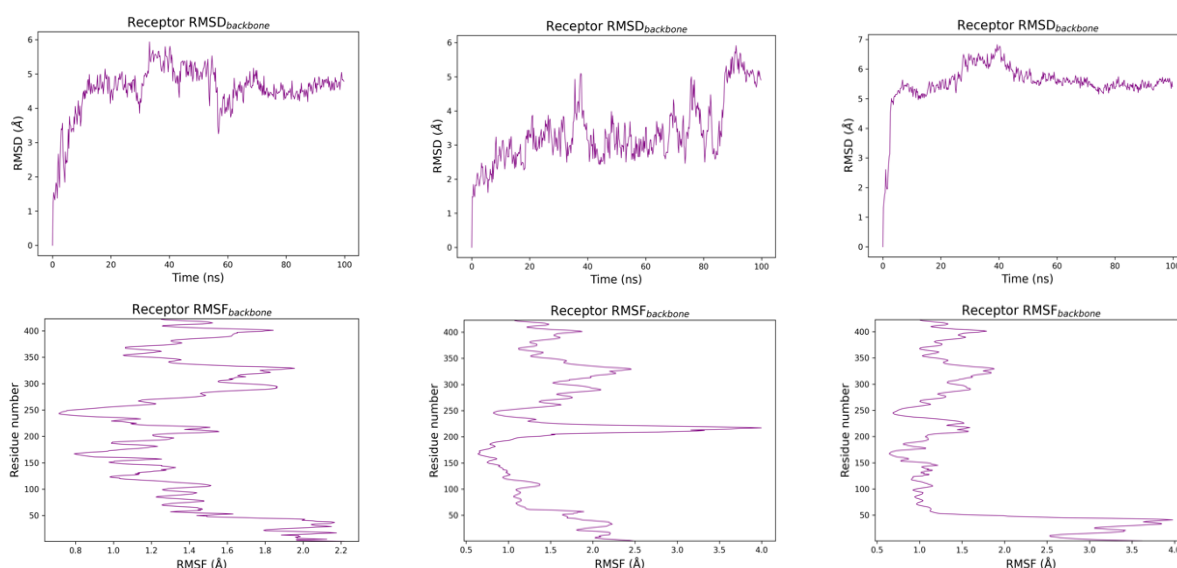


Figura 17. Analisi geometrica del modello monomero predetto con AlphaFold.

In accordo con i dati di letteratura che mostrano come la dimerizzazione aumenti l'attività catalitica, il modello dimerico si è dimostrato essere il più stabile nel corso delle simulazioni di dinamica (Figura 18). Dopo un iniziale rilassamento in cui il valore di RMSD aumenta, si raggiunge un livello costante che permane durante le simulazioni intorno a 1.7 Å: questo dato, oltre ad essere indicativo di una maggiore stabilità del modello in oggetto rispetto agli altri due, potrebbe essere una conferma indiretta di come la dimerizzazione porti a una stabilizzazione della proteasi, con conseguente aumento della sua attività enzimatica.

Per quanto riguarda il valore di RMSF degli amminoacidi, si può notare come il valore massimo raggiunto sia di 2.5 Å, inferiore rispetto al modello monomero di Phyre2.

Si può inoltre osservare come le porzioni relative al C-terminale e all'N-terminale, solitamente caratterizzate da ampie fluttuazioni strutturali dovute ai minori vincoli posti sulla loro posizione, vengano stabilizzate nella struttura dimerica rispetto alla forma monomera, poiché localizzate all'interfaccia tra i due monomeri.

Queste osservazioni hanno portato alla selezione del modello dimerico come quello di riferimento per il virtual screening.

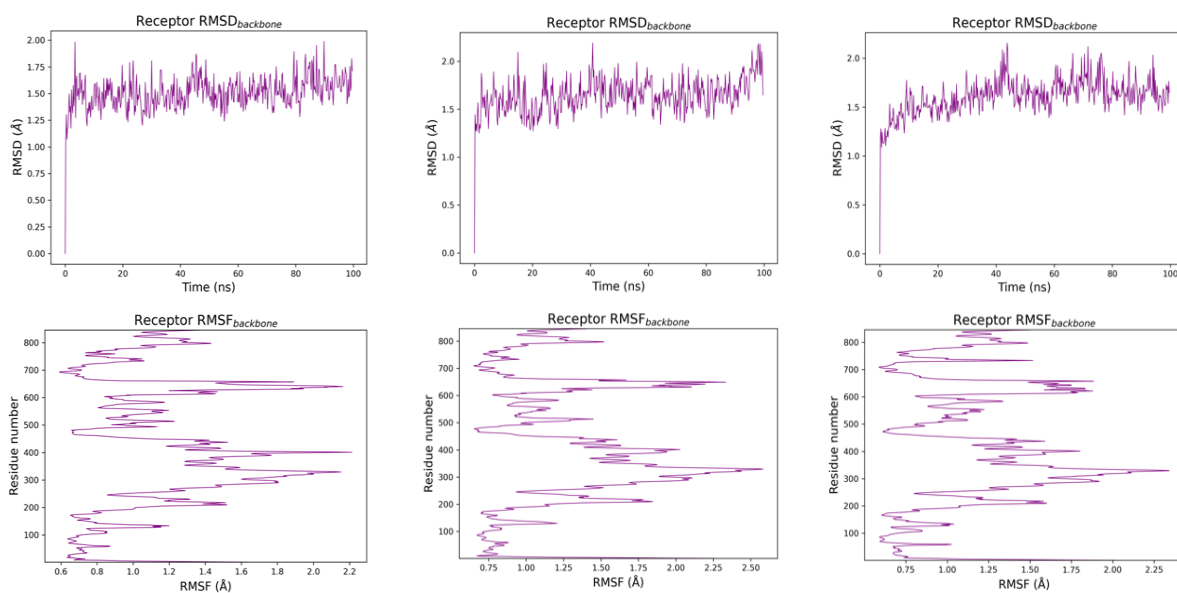


Figura 18. Analisi geometrica del modello dimerico

Nonostante il modello dimerico si sia stato selezionato quale più rappresentativo in funzione della sua maggiore stabilità in dinamica molecolare, è rilevante sottolineare come tutti i modelli mostrino comunque una buona sovrapposizione al livello del sito catalitico (Figura 19), validando indirettamente precedenti lavori basati su modelli di omologia.^{20,25}

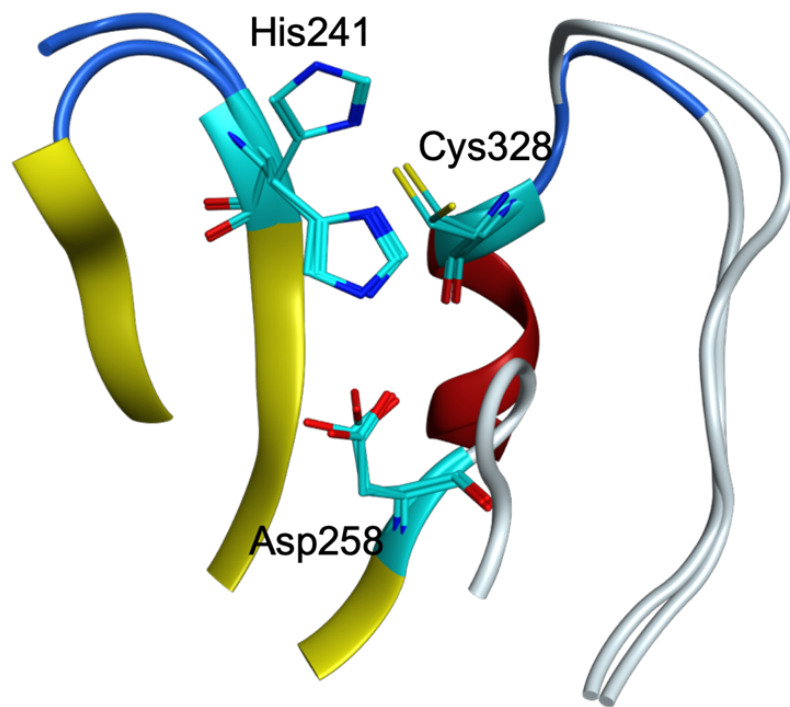


Figura 19. Sovrapposizione del modello di phyre2 con il modello di Alphafold e la SUMO proteasi Ulp1

Mutagenesi

Secondo i dati di letteratura relativi alla mutagenesi della proteasi¹⁹, amminoacidi sequenzialmente adiacenti alla triade catalitica (C328, H241, D258) o amminoacidi appartenenti ad essa, se mutati ad alanina, portano ad una perdita di attività proteolitica. Anche la mutazione dell'acido aspartico 248 e della glutammina 322 porta ad una perdita di attività.¹⁹

Questi ultimi due amminoacidi sono stati selezionati per uno studio di mutagenesi in silico, con lo scopo di osservare una perturbazione della triade catalitica che possa giustificare la perdita di attività emersa in letteratura.

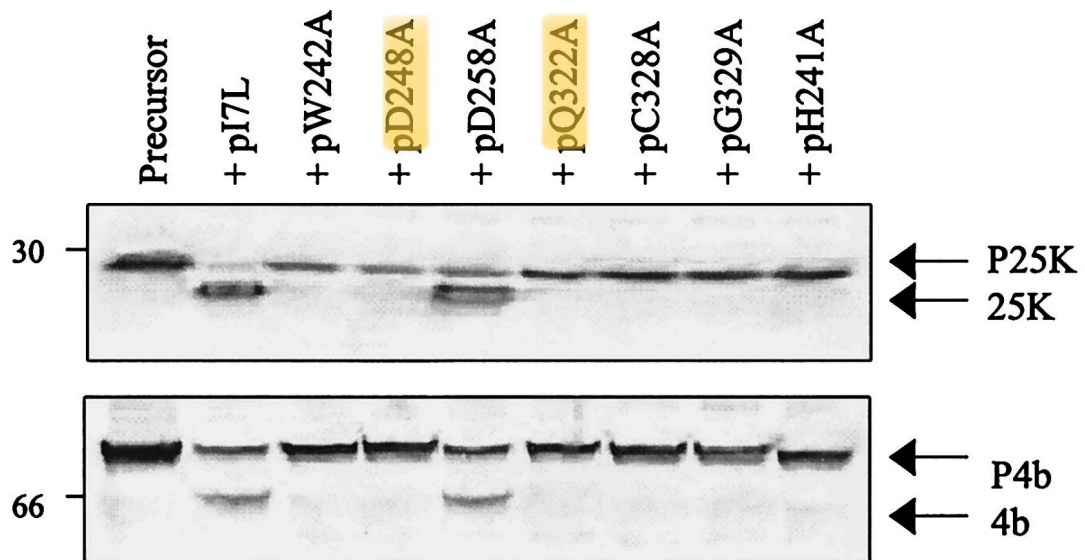


Figura 20. Capacità delle proteasi mutanti di tagliare le proteine del core virale¹⁹

L'acido aspartico 248 è coinvolto in un ponte salino con l'arginina 230: entrambi gli amminoacidi fanno parte dello stesso sistema secondario a beta foglietto e si trovano su due loop di due sequenze antiparallele (Figura 21). La glutammina 322 si trova invece nell'intorno chimico della cisteina catalitica (Figura 22).

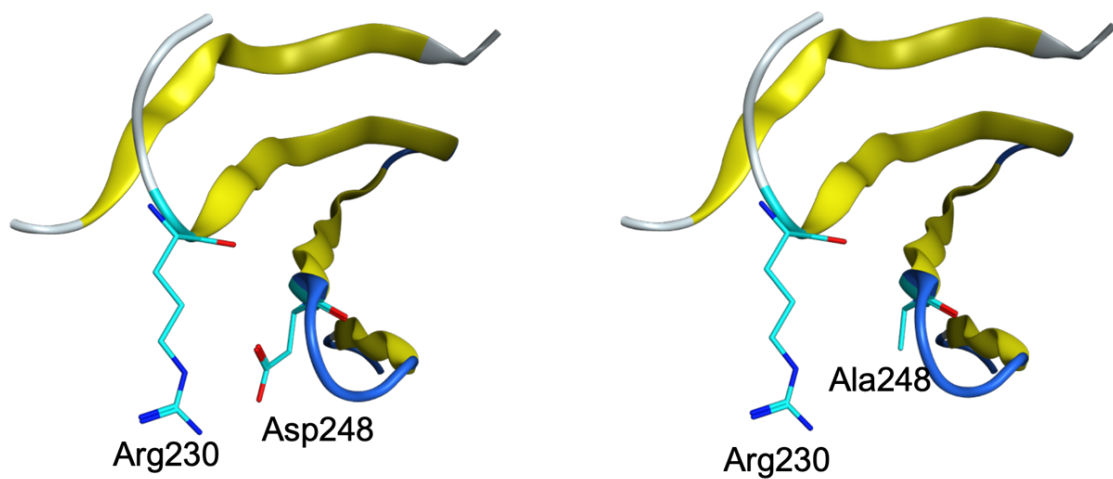


Figura 21. Mutazione D248A

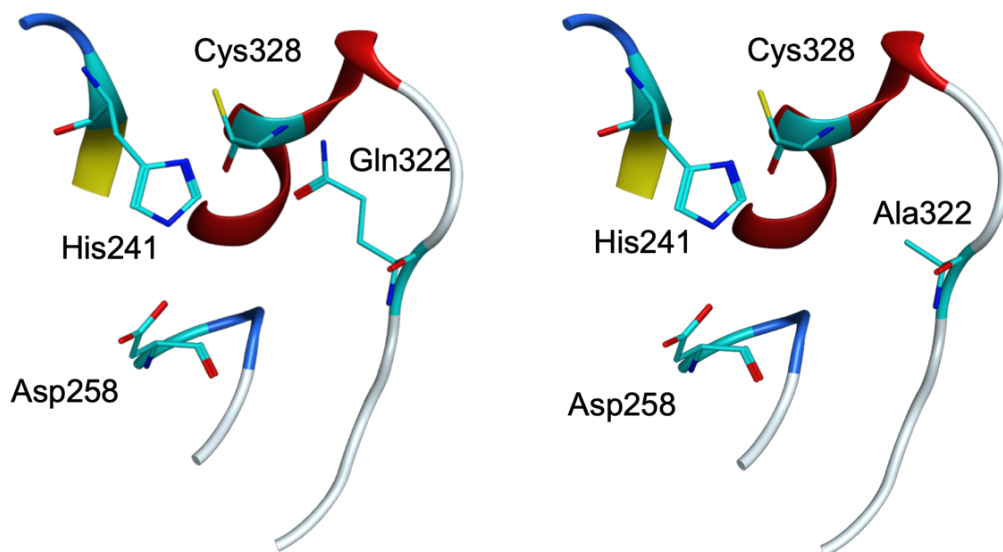


Figura 22. Mutazione Q322A

Tramite il modulo Protein Builder di MOE, sono state create quindi due proteasi mutate. La prima proteasi presenta quindi la mutazione D248A, mentre la seconda Q322A. Analogamente a quanto fatto per la validazione dei modelli di omologia della proteasi wild-type, la stabilità dei due mutanti è stata analizzata tramite simulazioni di dinamica da 100 ns ciascuna.

Nel caso della prima mutazione, nonostante la potenziale importanza del ponte salino per la stabilità del beta foglietto, non sono stati rilevati importanti cambiamenti conformazionali. Questo può essere in parte giustificato dalla presenza di ulteriori interazioni intermolecolari che contribuiscono a stabilizzare quella porzione della proteasi, che non permettono di osservare cambiamenti rilevanti nei tempi di simulazione impiegati.

Nel caso della mutazione in posizione 322, invece, si osserva un perturbamento dell'intorno chimico della cisteina catalitica. Lo spazio precedentemente occupato dalla glutammina viene esplorato parzialmente dalla cisteina catalitica che si ritrova in una posizione meno ottimale, rispetto a quella di partenza, per essere deprotonata dall'istidina.

Analisi della possibile formazione di un ponte disolfuro

La cisteina 408 all'interno del modello dimerico è collocata al livello di una alfa elica nel C-terminale. La predizione della dimerizzazione porta la cisteina 408 dell'altro monomero a specchiarsi con essa in maniera simmetrica (Figura 23).

Si è dunque ipotizzato che qualora la predizione di dimerizzazione sia corretta, si possa verificare la formazione di un ponte disolfuro tra queste cisteine.

Al fine di verificare quindi la consistenza di questa condizione strutturale favorevole alla formazione del ponte disolfuro, si è quindi proceduto ad analizzare la distanza tra gli zolfi delle cisteine nel corso delle tre simulazioni di dinamica classica effettuate sul modello dimerico.

	Prima dinamica	Seconda dinamica	Terza dinamica
Distanza minima	3.51 Å	3.40 Å	3.40 Å
Distanza media	3.40 Å	4.14 Å	4.12 Å
Distanza massima	5.00 Å	5.28 Å	5.13 Å

Tabella 2. Distanza rilevata in dinamica (Cys408)

Come si può osservare in Tabella 2, la distanza tra gli zolfi delle due cisteine sopracitate rimane mediamente compresa tra 3 Å e 4 Å, mentre la distanza minima si aggira attorno ai 3.5 Å. Considerando che in dinamica molecolare non è contemplata la formazione di un legame tra due atomi e considerando la presenza dei protoni dei tioli che contribuiscono ad allontanare i due zolfi per ragioni steriche, si può ipotizzare che nelle giuste condizioni redox possa verificarsi la formazione di un ponte disolfuro, in cui la distanza tra gli zolfi misura 2.1 Å.

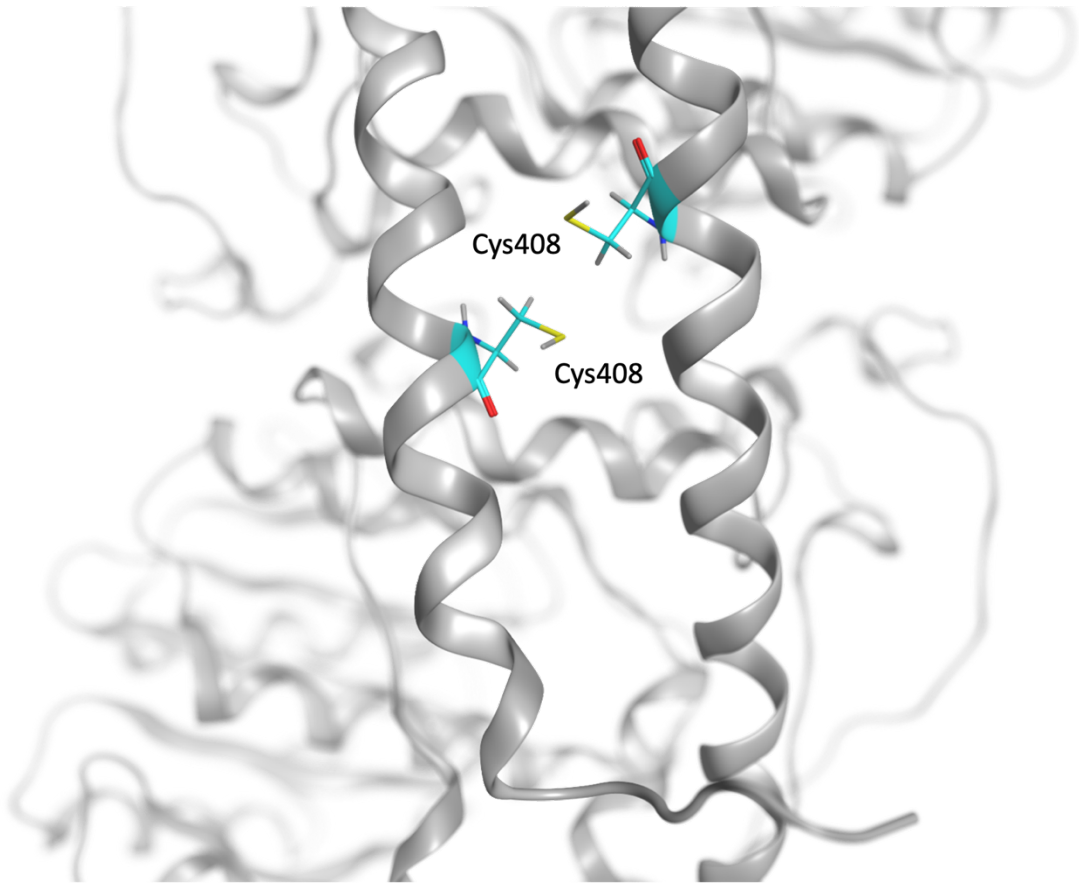


Figura 23. Cys408

Analisi strutturale del sito catalitico

Il sito catalitico è situato in una stretta porzione di raccordo tra due cavità, una inferiore di dimensioni maggiori e una superiore. Globalmente il sito risulta estremamente esposto al solvente (Figura 24).

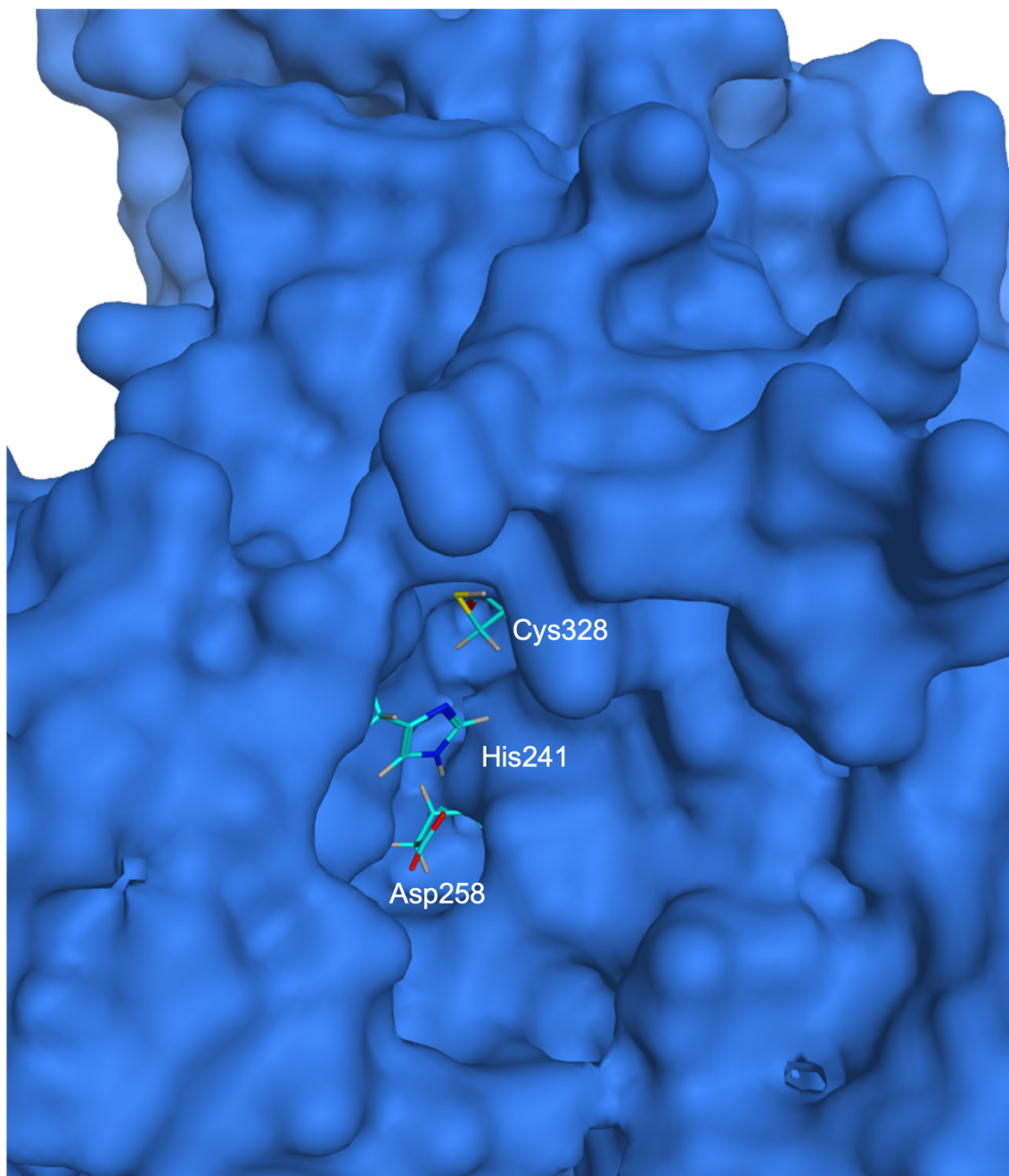


Figura 24. Cavità prossime alla triade catalitica

Da un'analisi visiva in dinamica è emerso come la presenza del triptofano 168 regoli l'accesso alla cisteina catalitica. Questo triptofano è collocato superiormente rispetto alla cisteina 328 in una conformazione a "ponte" (Figura 25), in equilibrio con la conformazione aperta dove il triptofano 168 interagisce stabilmente con il triptofano 242 (Figura 26). Questa interazione è determinata dal contributo idrofobico e di π -stacking tra i due residui amminoacidici.

Questa osservazione ha posto le basi per una ipotesi interattiva che un eventuale candidato inibitore dovrebbe avere nei confronti della proteasi: esso potrebbe quindi avere dei gruppi aromatici capaci di mimare l'interazione del Trp168 con il Trp242, andandosi ad "intercalare" tra i due amminoacidi.

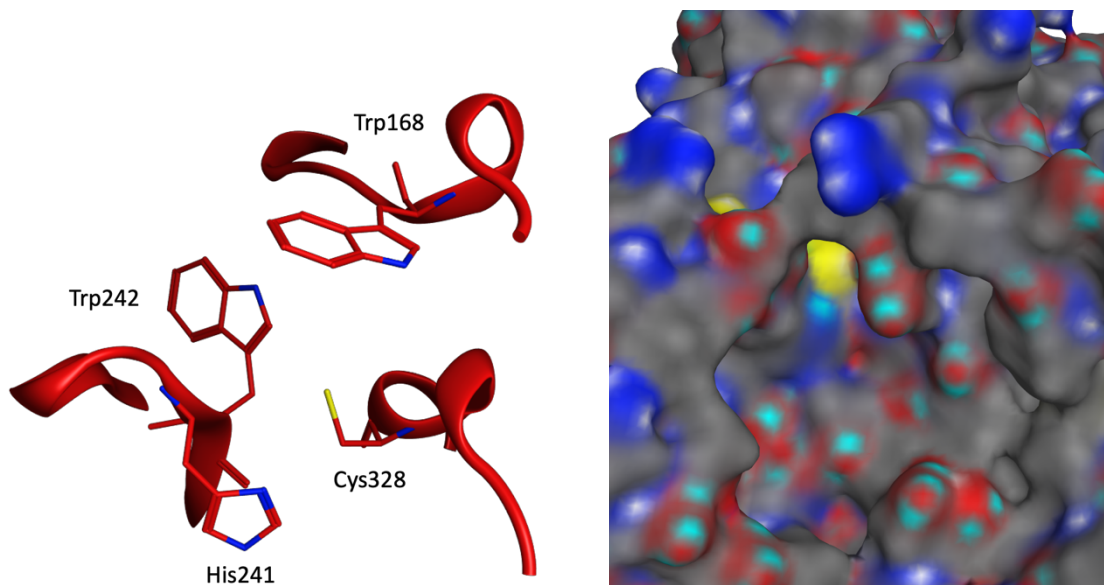


Figura 25. Conformazione a "ponte"

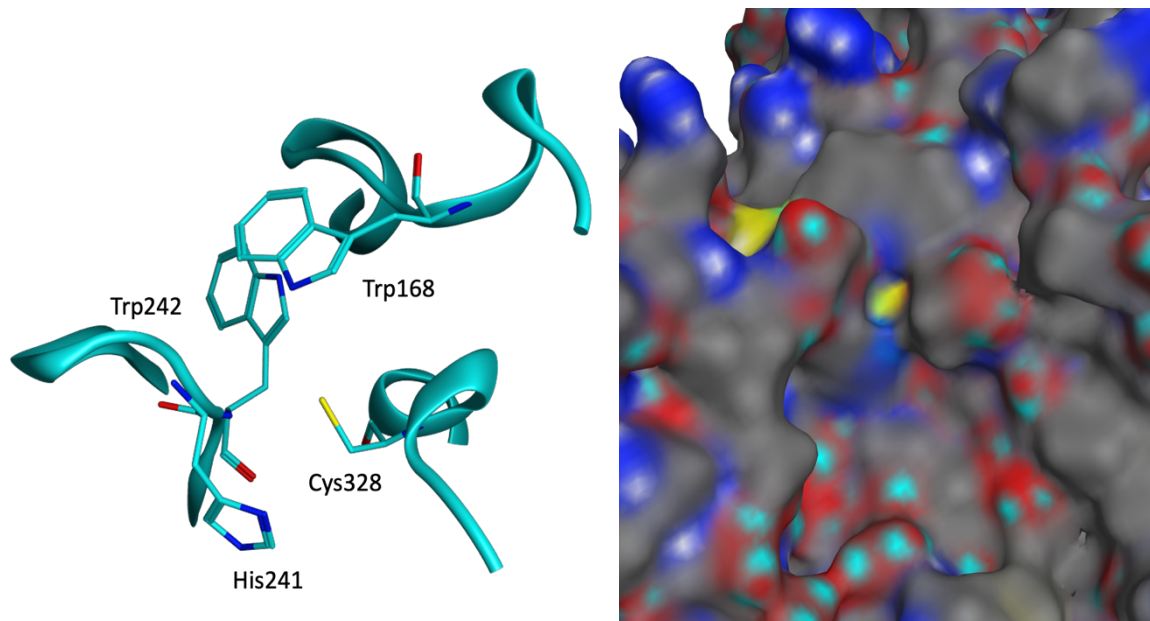


Figura 26. Conformazione aperta

Caratterizzazione della binding mode del Ttp6171

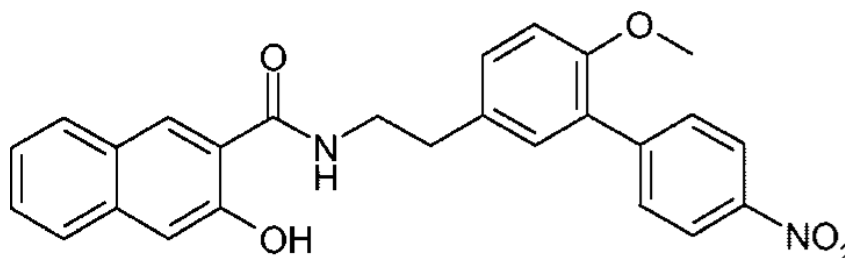
Il Ttp6171 è un promettente candidato inibitore, uscente da un virtual screening basato su un modello di omologia. Questo composto presenta un IC₅₀ di 12 µM e un buon indice terapeutico.²⁵

Compound	TC ₅₀ (µM)	IC ₅₀ (µM)	TI
6171	900	12	75
201018	600	200	3
200480	55	14	4
174878	400	50	8
130961	150	10	15
176510	400	56	7
123045	200	50	4

Figura 27. Attività del Ttp6171²⁵

È stato effettuato un docking di questa molecola, con l'obiettivo di razionalizzare la sua binding mode con la proteasi. Tra tutte le pose di docking ne è stata selezionata una in particolare, questa pose oltre ad avere un buon punteggio di scoring, elettrostatica e van der Waals, rispetta l'ipotesi interattiva precedentemente avanzata.

Il Ttp6171 presenta una stretta porzione centrale, caratterizzata da un legame ammidico, mentre le porzioni laterali sono più ingombrate e sono caratterizzate dalla presenza di cicli aromatici, in particolare un biciclo che potrebbe mimare il Triptofano 168 nella sua interazione con il Trp242.



TTP - 6171

Figura 28. Struttura del Ttp6171²⁵

Nella figura 29 viene mostrata la pose di docking del Ttp6171 selezionata. Il biciclo aromatico del composto si “intercala” tra il Trp168 e il Trp242 portando interazioni idrofobiche e di stacking π - π . Nella porzione centrale della molecola, troviamo un ponte a idrogeno tra il backbone della Ser240 e l’idrogeno legato all’azoto ammidico. Infine, la His241 media interazioni idrofobiche con i cicli aromatici esterni alla molecola e il ciclo terminale interagisce idrofobicamente con la Gln322.

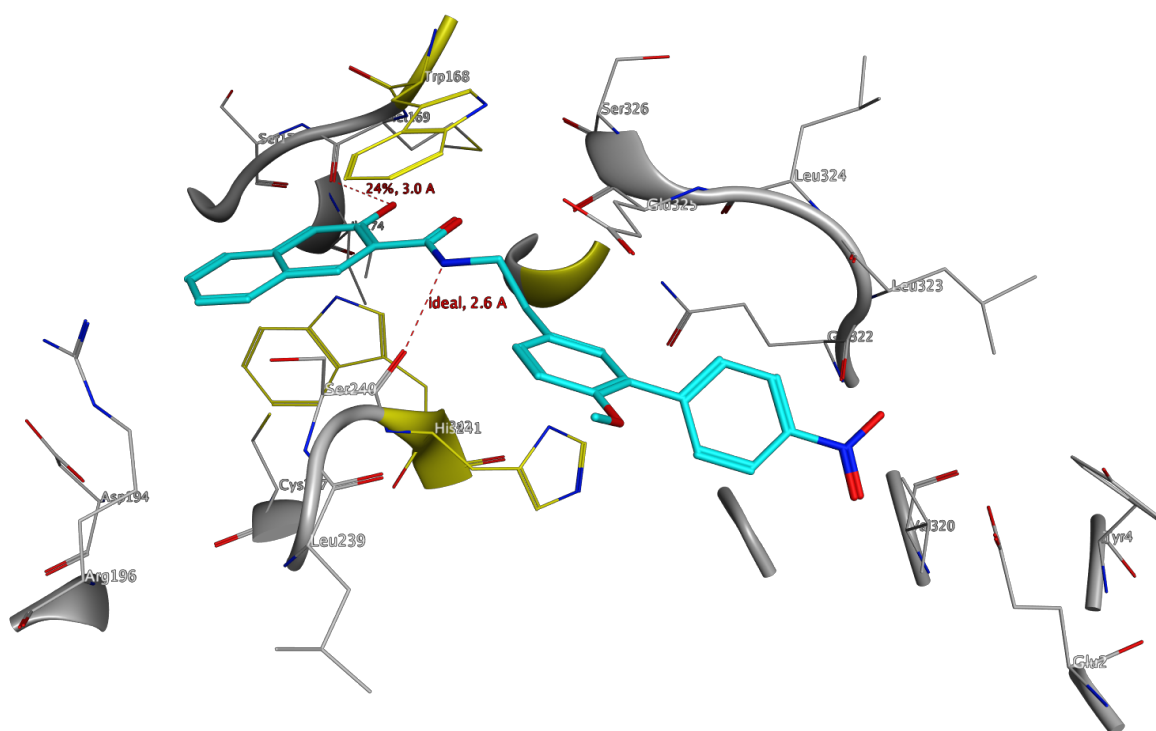


Figura 29. Docking pose Ttp6171

Questa pose di docking rappresenta la binding mode auspicata per il Ttp6171, di conseguenza è stata impiegata come riferimento per calcolarne la fingerprint di interazione con la proteina, con lo scopo di avere un descrittore che tenga conto delle interazioni precedentemente descritte, da utilizzare poi come filtro nel successivo virtual screening.

Virtual screening (Drugbank)

È stato effettuato un virtual screening di 11.300 composti provenienti da drugbank. Questi composti sono dei farmaci approvati o candidati nelle fasi cliniche di sperimentazione. La scelta di effettuare il virtual screening su questa libreria è dettata dalla necessità di trovare rapidamente un trattamento efficace contro il Monkeypox, ricorrendo alla tecnica del riposizionamento.

La libreria è stata inizialmente filtrata da 11.300 a 2.208 composti, attraverso la voce “drug” del tool FILTER di openeye:

Nome del filtro	Valore	Funzione
MIN_MOLWT	200.000.000	Peso molecolare minimo
MAX_MOLWT	600.000.000	Peso molecolare massimo
MIN_Num_Heavy	15	Numero minimo di atomi diversi dall'idrogeno
MAX_Num_Heavy	35	Numero massimo di atomi diversi dall'idrogeno
MIN_RING_SYS	0	Numero minimo di sistemi ad anello
MAX_RING_SYS	5	Numero massimo di sistemi ad anello
MIN_RING_SIZE	0	Numero minimo di atomi presenti in ogni sistema ad anello
MAX_RING_SIZE	20	Numero massimo di atomi presenti in ogni sistema ad anello
MIN_CARBONS	7	Numero minimo di carboni
MAX_CARBONS	35	Numero massimo di carboni
MIN_HETEROATOMS	2	Numero minimo di atomi diversi da carbonio e idrogeno
MAX_HETEROATOMS	20	Numero massimo di atomi diversi da carbonio e idrogeno

MIN_Het_C_Ratio	0.100000	MIN_HETEROATOMS/Numero di carboni
MAX_Het_C_Ratio	1.000.000	MAX_HETEROATOMS/Numero di carboni
MIN_ROT_BONDS	0	Numero minimo di legami rotabili
MAX_ROT_BONDS	20	Numero massimo di legami rotabili
MIN_RIGID_BONDS	0	Numero minimo di legami non rotabili
MAX_RIGID_BONDS	35	Numero massimo di legami non rotabili
MIN_CON_NON_RING	0	Numero minimo di atomi contigui legati (esclusi anelli)
MAX_CON_NON_RING	15	Numero massimo di atomi contigui legati (esclusi anelli)
MIN_UNBRANCHED	0	Numero minimo di atomi connessi e non ramificati legati (esclusi anelli)
MAX_UNBRANCHED	6	Numero massimo di atomi connessi e non ramificati legati (esclusi anelli)
MIN_FCN_GRP	0	Numero minimo di gruppi funzionali
MAX_FCN_GRP	18	Numero massimo di gruppi funzionali
MIN_COUNT_FORMAL_CRG	0	Numero minimo di atomi con una carica formale
MAX_COUNT_FORMAL_CRG	3	Numero massimo di atomi con una carica formale

MIN_SUM_FORMAL_CRG	-2	Numero minimo di carica formale totale
MAX_SUM_FORMAL_CRG	2	Numero massimo di carica formale totale
MIN_HBOND_DONORS	0	Numero minimo di idrogeni legati ad atomi di azoto, ossigeno e zolfo
MAX_HBOND_DONORS	6	Numero massimo di idrogeni legati ad atomi di azoto, ossigeno e zolfo
MIN_HBOND_ACCEPTORS	0	Numero minimo di atomi che rispettano i seguenti criteri: - degree 2, aromatic, non-positive nitrogens - electron rich or negative, valence less than 4, non-aromatic nitrogens - negatively charged or not electron withdrawn and neutral oxygens - degree 1, double bonded, electron rich sulfur
MAX_HBOND_ACCEPTORS	8	Numero massimo di atomi che rispettano i seguenti criteri: - degree 2, aromatic, non-positive nitrogens - electron rich or negative, valence less than 4, non-aromatic nitrogens - negatively charged or not electron withdrawn and neutral

		oxygens - degree 1, double bonded, electron rich sulfur
MIN_LIPINSKI_DONORS	0	Numero minimo di azoti e ossigeni che legano almeno un idrogeno
MAX_LIPINSKI_DONORS	5	Numero massimo di azoti e ossigeni che legano almeno un idrogeno
MIN_LIPINSKI_ACCEPTORS	0	Numero minimo di atomi di azoto e ossigeno
MAX_LIPINSKI_ACCEPTORS	10	Numero massimo di atomi di azoto e ossigeno
MIN_CHIRAL_CENTERS	0	Numero minimo di centri chirali
MAX_CHIRAL_CENTERS	4	Numero massimo di centri chirali
MIN_XLOGP	-5.000.000	LogP minimo
MAX_XLOGP	6.000.000	LogP massimo
MIN_SOLUBILITY	moderately	
MIN_2D_PSA	0.000000	Area superficiale polare minima
MAX_2D_PSA	150.000.000	Area superficiale polare massima
MAX_LIPINSKI	1	Numero di violazioni consentite alle regole di lipinski
MIN_ABS	0.500000	Score di biodisponibilità Yvonne Martine Abbott

Tabella 3. Drug filter parameters from OpenEye.

La libreria filtrata è stata successivamente preparata per il docking seguendo il seguente protocollo:

- Generazione delle coordinate 2D attraverso OpenBabel

- Generazione delle coordinate 3D attraverso Corina
- Determinazione dello stato di protonazione attraverso il tool Fixpka della suite Quacpac di OpenEye.
- Determinazione delle forme tautomeriche attraverso il tool Tautomers della suite Quacpac di OpenEye.
- Assegnazione delle cariche parziali attraverso il tool Molcharge della suite Quacpac di OpenEye.
- Assegnazione degli atom types attraverso SPORES (Structure Protonation and Recognition System)

Una volta preparato il database di strutture è stato eseguito un docking con PLANTS. Lo spazio di ricerca accessibile all'algoritmo di ricerca conformazionale è stato definitivo tramite una sfera di raggio 12 Å, centrata sul centro di massa della cisteina catalitica (Cys 328).

Le pose ottenute sono state minimizzate e successivamente è stato effettuato il calcolo delle fingerprint per ciascuna di esse. Le fingerprint ottenute sono state confrontate con quella ottenuta dalla pose selezionata del Ttp6171 (Figura 29). Ad ogni pose è stato associato un punteggio, ottenuto tramite l'utilizzo di una funzione di scoring basata sulla similarità del coseno dei vettori associati ad ogni fingerprint (Equazione 7). Il punteggio ottenuto tramite la scoring function IFP_{cs} è stato utilizzato come filtro, al fine di selezionare composti con un pattern di interazione con la proteina simile a quello del riferimento.

I composti sono stati filtrati secondo il seguente protocollo:

- Energia di interazione elettrostatica <10 kcal/mol
- Energia di interazione di van der Waals <10 kcal/mol
- Percentuale di superficie esposta al solvente <15%
- Scoring $IFP_{cs} > 0.97$
- Ispezione visiva

Sono stati ottenuti in questo modo 14 candidati inibitori.

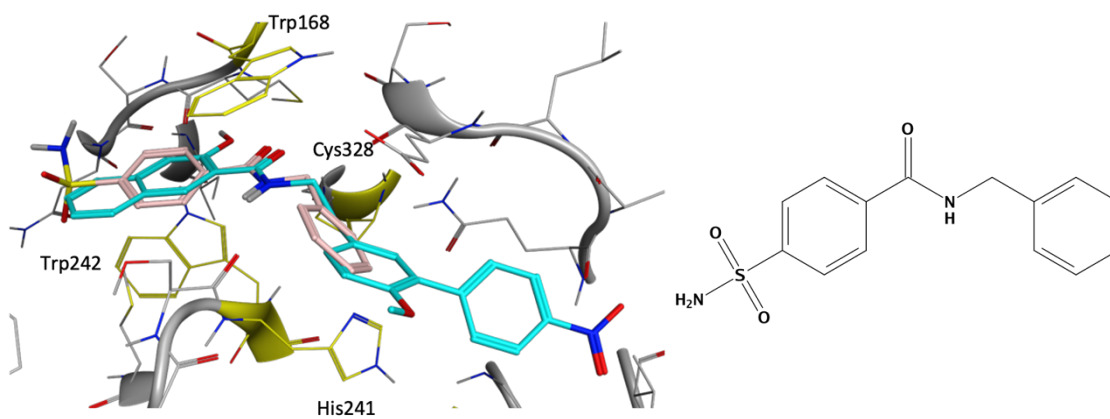


Figura 30. DB01748

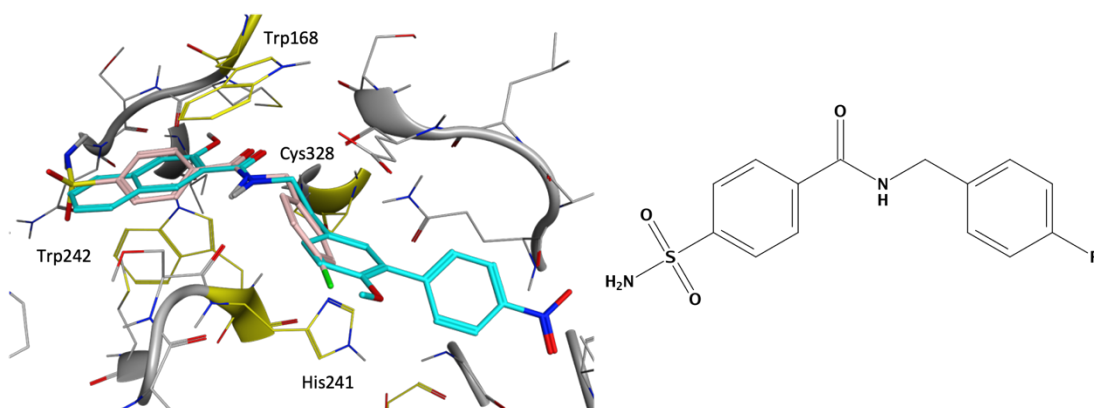


Figura 31. DB02429

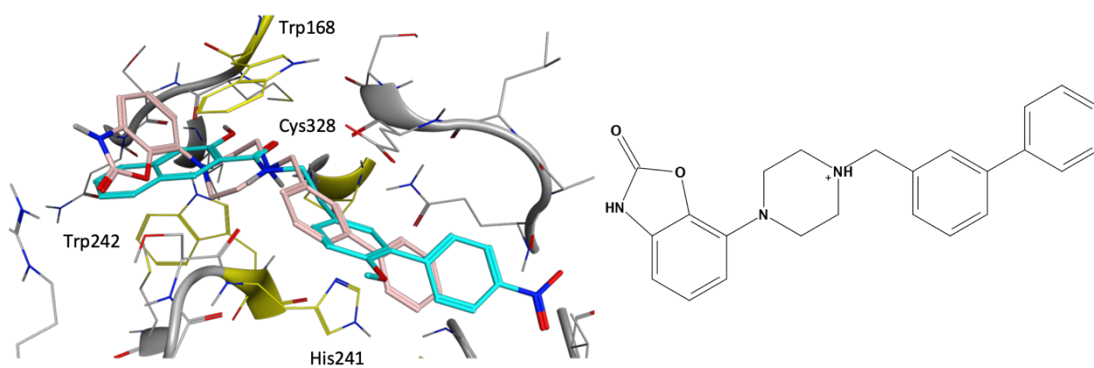


Figura 32. DB04888

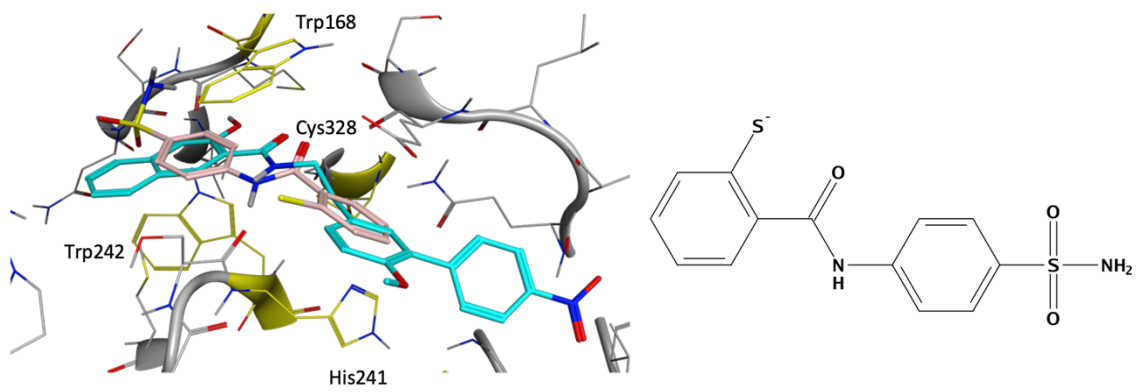


Figura 33. DB07476

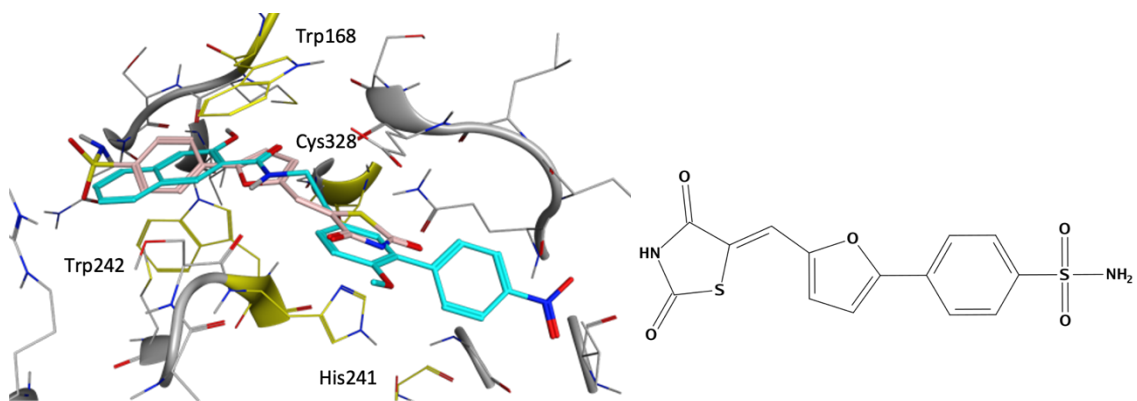


Figura 34. DB07531

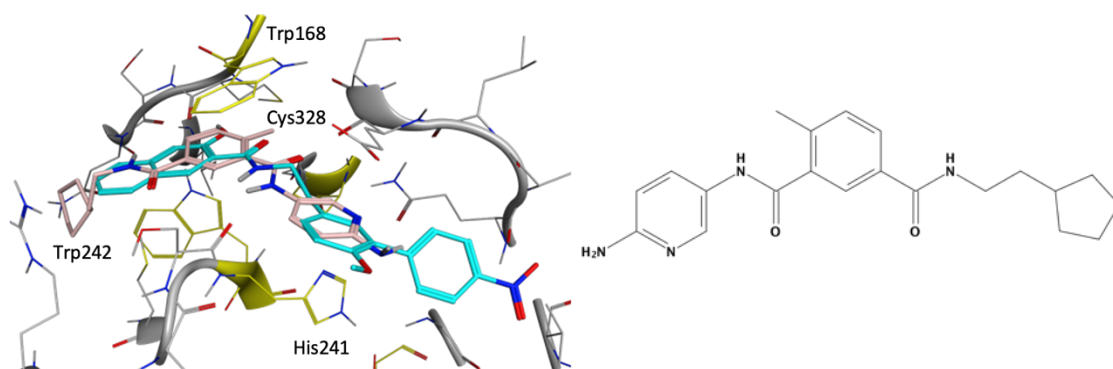


Figura 35. DB07537

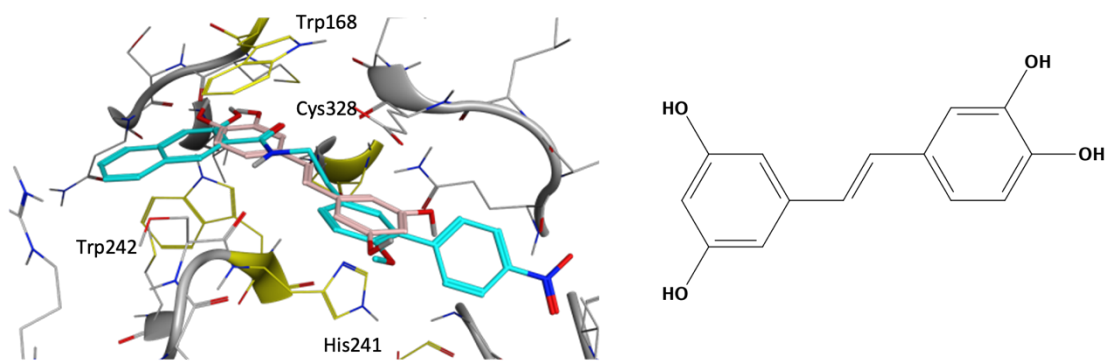


Figura 36. DB08399

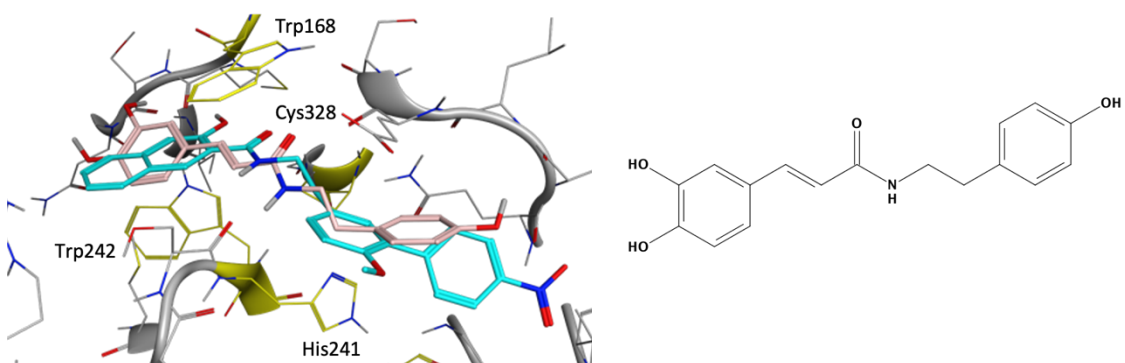


Figura 37. DB08754

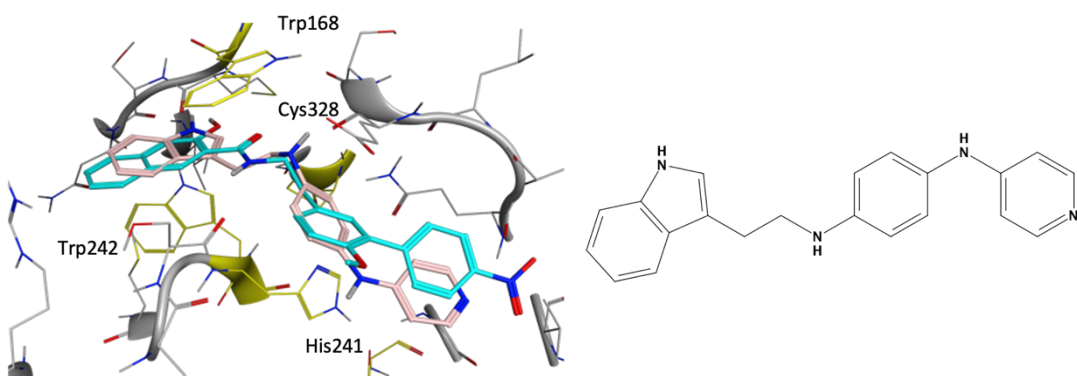


Figura 38. DB12027

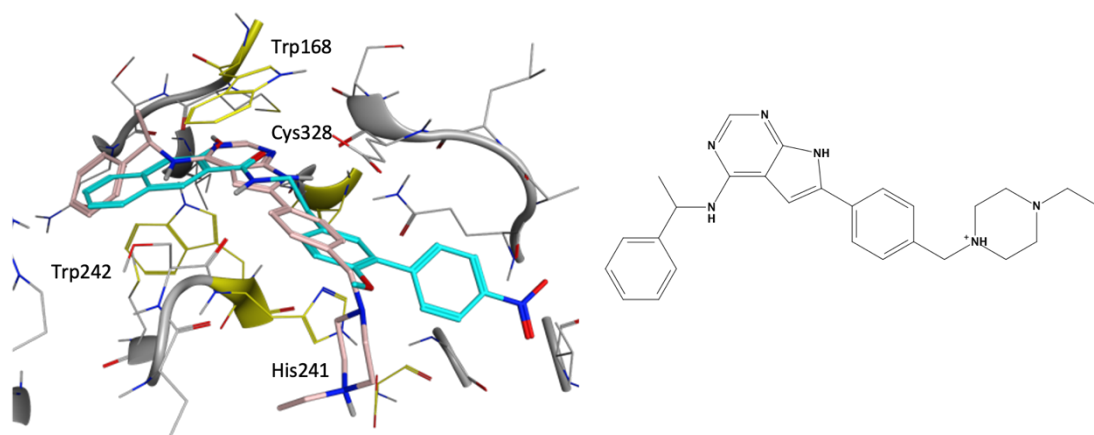


Figura 39. DB12558

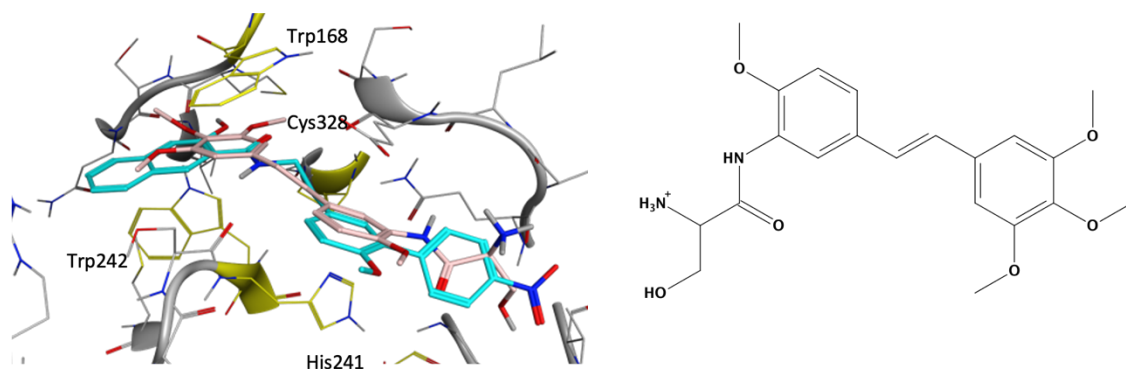


Figura 40. DB12882

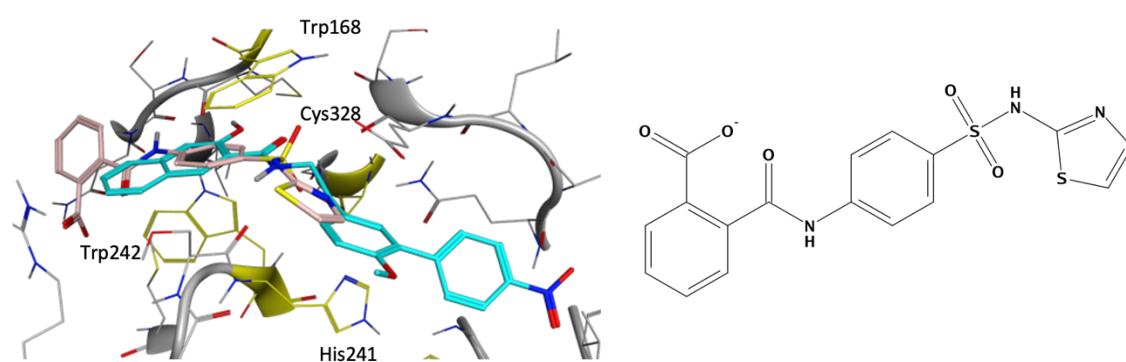


Figura 41. DB13248

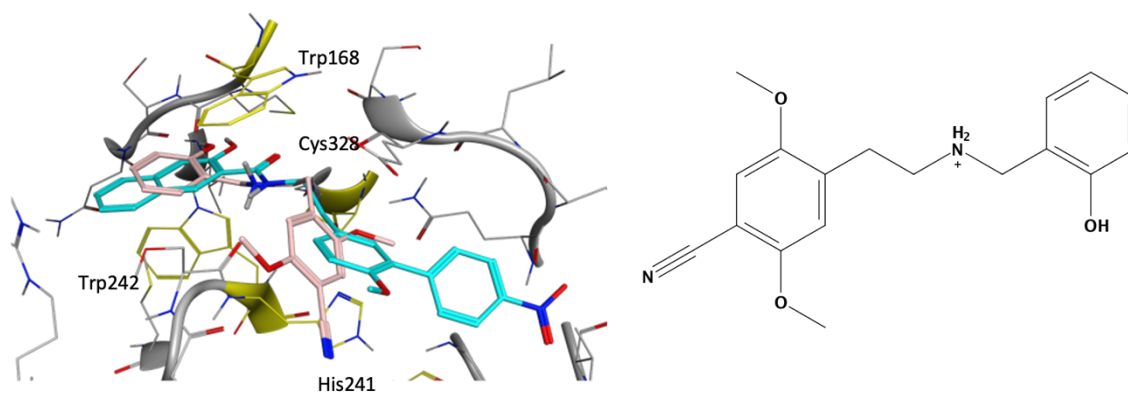


Figura 42. DB13948

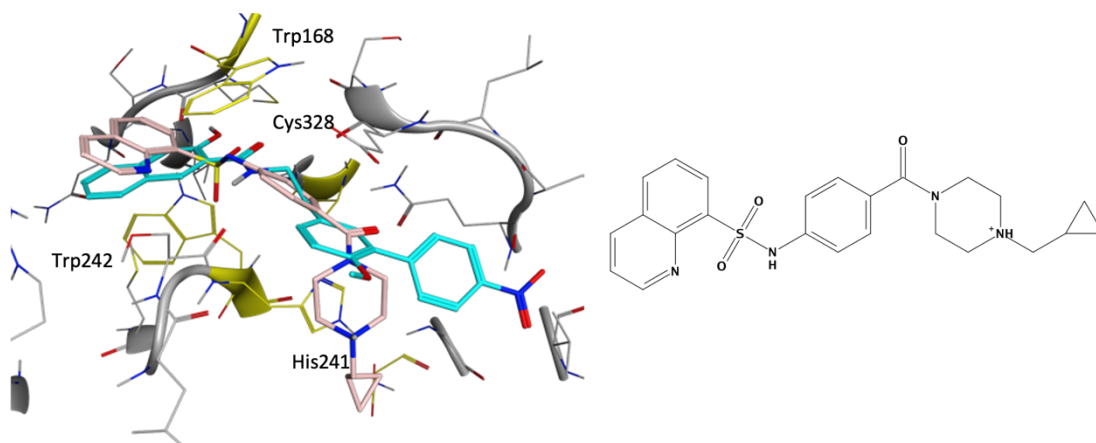


Figura 43. DB16236

Virtual screening (composti in house)

In maniera complementare al primo virtual screening, sempre nell'ottica di riposizionare composti già esistenti e pensati per bersagliare altri target di interesse terapeutico, è stato effettuato un secondo virtual screening di composti presenti "in house", derivati da 40 anni di attività scientifica del laboratorio di sintesi della professoressa M.G. Ferlin, presso il Dipartimento di Scienze del Farmaco dell'Università di Padova.

Lo scopo di questo virtual screening è quello di trovare dei composti potenzialmente attivi che abbiano una via di sintesi ben caratterizzata, che possano dunque fungere da scaffold per lo sviluppo razionale di un inibitore della proteasi del Monkeypox.

La libreria di 427 composti è stata preparata e sottoposta a virtual screening, mantenendo le medesime condizioni del precedente.

Le pose di docking sono state poi filtrate secondo il seguente protocollo:

- Energia di interazione elettrostatica < 10 kcal/mol
- Energia di interazione di van der Waals < 10 kcal/mol
- Percentuale di superficie esposta al solvente $< 15\%$
- Scoring $I\text{FP}_{\text{CS}} > 0.95$
- Ispezione visiva

Sono stati selezionati 7 candidati inibitori.

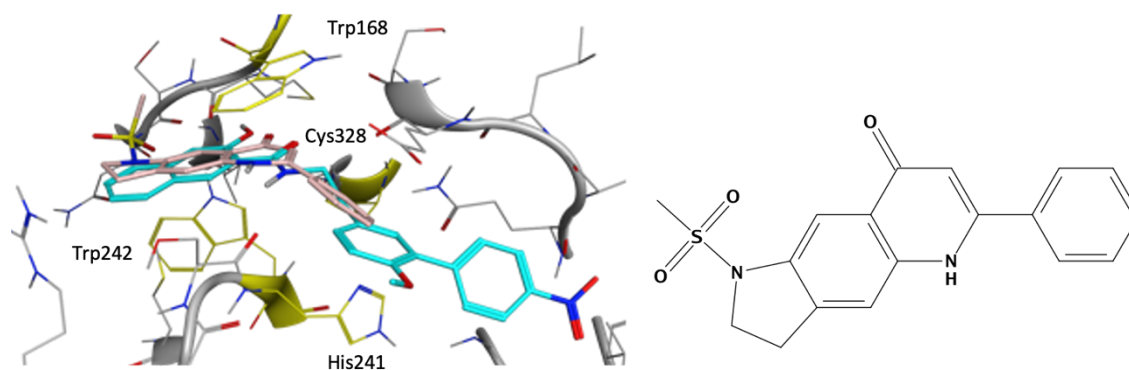


Figura 44. mg2646

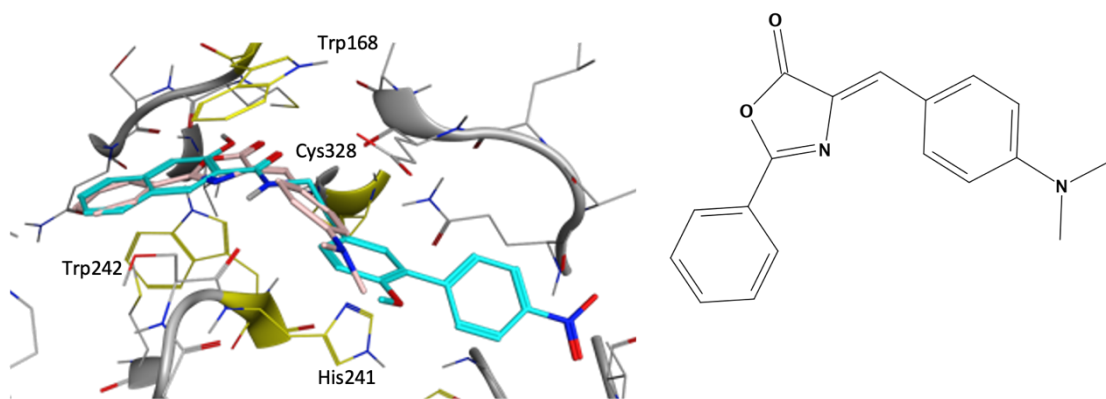


Figura 45. mg2733

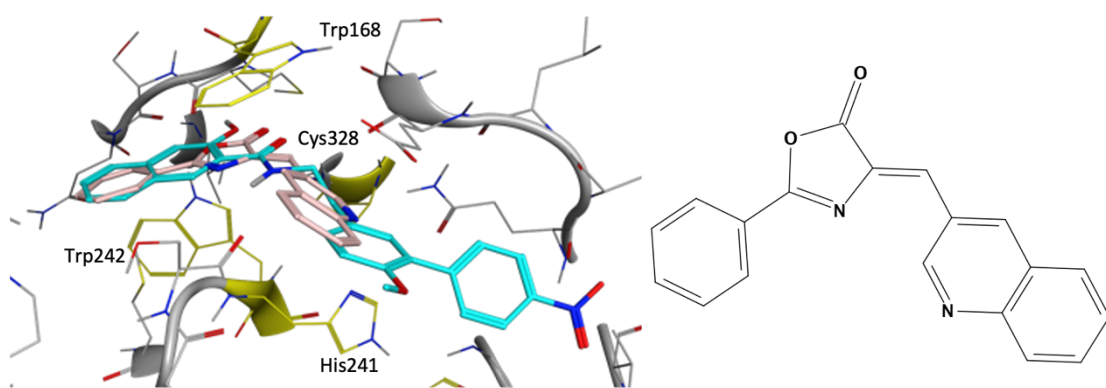


Figura 46. mg2734

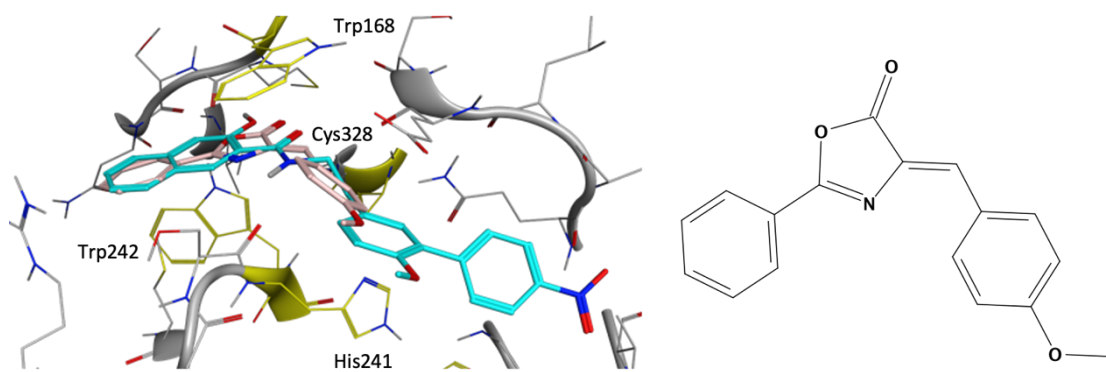


Figura 47. mg2736

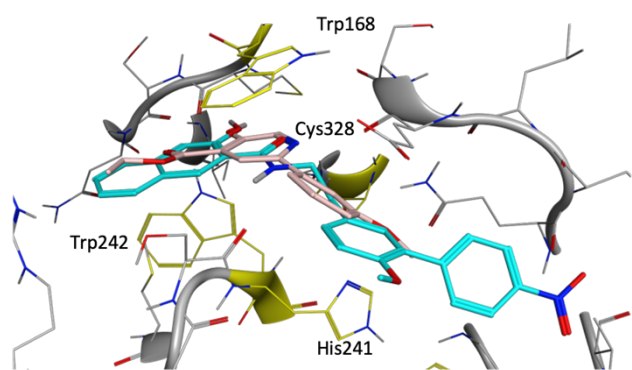


Figura 48. mg2792

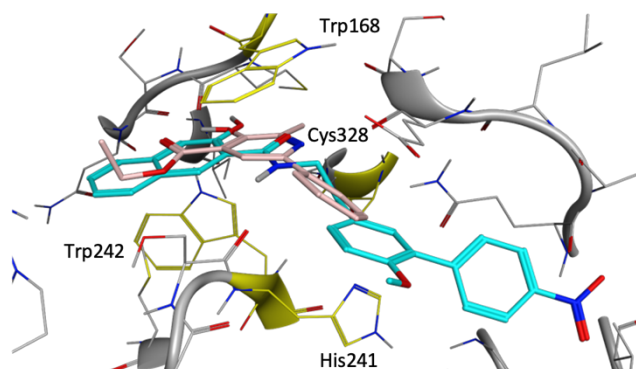
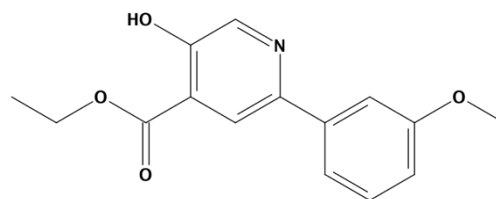


Figura 49. mg2793

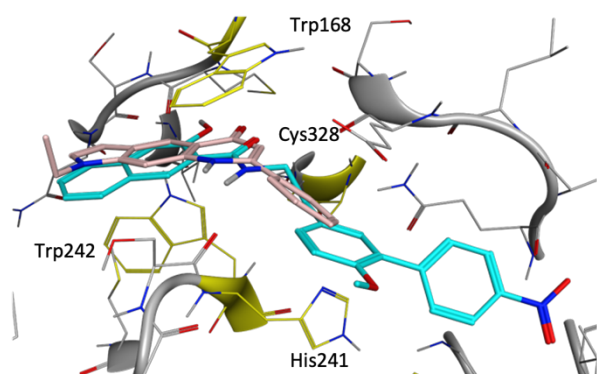
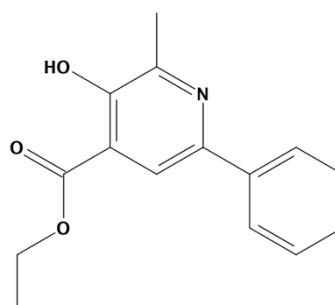
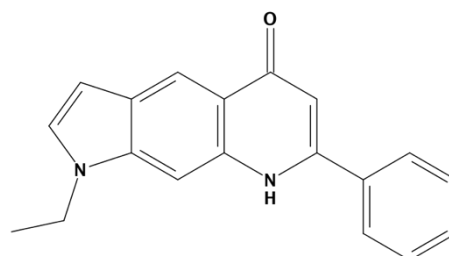


Figura 50. mg2816



CONCLUSIONE

Il presente lavoro di tesi si poneva l'obiettivo di sviluppare un protocollo utile al riposizionamento di farmaci approvati o candidati farmaci in fase clinica di sperimentazione, al fine di identificare dei composti utili in una situazione emergenziale come l'epidemia di vaiolo delle scimmie, per la quale non esiste un trattamento specifico ad oggi. I composti selezionati verranno in seguito testati sperimentalmente nel laboratorio del Professore Cristiano Salata del dipartimento di Medicina Molecolare dell'Università degli studi di Padova.

Il punto di partenza, ovvero la creazione di un modello della proteasi, è stato raggiunto tramite una tecnica innovativa di predizione de novo che, sfruttando una rete neurale e partendo da una sequenza primaria, fornisce una struttura terziaria o quaternaria.

Le informazioni raccolte durante l'analisi strutturale della proteasi e la caratterizzazione della binding mode con il Ttp6171, pongono le basi per un possibile sviluppo razionale di un inibitore specifico ad oggi inesistente. Inoltre, il secondo virtual screening ha portato alla selezione di composti che se dovessero dimostrare una attività nei confronti della proteasi, potrebbero essere un buon punto di partenza come scaffold per questo sviluppo.

BIBLIOGRAFIA

1. Parker, S. & Buller, R. M. A review of experimental and natural infections of animals with monkeypox virus between 1958 and 2012. *Future Virol* **8**, 129 (2013).
2. Marennikova, S. S., Seluhina, E. M., Mal'ceva, N. N., Cimiskjan, K. L. & Macevic, G. R. Isolation and properties of the causal agent of a new variola-like disease (monkeypox) in man. *Bull World Health Organ* **46**, 599 (1972).
3. Petersen, E. *et al.* Human Monkeypox: Epidemiologic and Clinical Characteristics, Diagnosis, and Prevention. *Infect Dis Clin North Am* **33**, 1027 (2019).
4. Kaler, J., Hussain, A., Flores, G., Kheiri, S. & Desrosiers, D. Monkeypox: A Comprehensive Review of Transmission, Pathogenesis, and Manifestation. *Cureus* **14**, (2022).
5. Singhal, T., Kabra, S. K. & Lodha, R. Monkeypox: A Review. *Indian J Pediatr* **89**, 955–960 (2022).
6. Hraib, M., Jouni, S., Albitar, M. M., Alaidi, S. & Alshehabi, Z. The outbreak of monkeypox 2022: An overview. *Annals of Medicine and Surgery* **79**, 104069 (2022).
7. Mathieu, E., Spooner, F., Dattani, S., Ritchie, H. & Roser, M. Monkeypox. *Our World in Data* (2022).
8. Kumar, N., Acharya, A., Gendelman, H. E. & Byrareddy, S. N. The 2022 outbreak and the pathobiology of the monkeypox virus. *J Autoimmun* **131**, 102855 (2022).
9. Haller, S. L., Peng, C., McFadden, G. & Rothenburg, S. Poxviruses and the evolution of host range and virulence. *Infect Genet Evol* **21**, 15–40 (2014).
10. Ortiz-Martínez, Y. *et al.* Monkeypox - a description of the clinical progression of skin lesions: a case report from Colorado, USA. *Ther Adv Infect Dis* **9**, (2022).
11. Rimoin, A. W. *et al.* Major increase in human monkeypox incidence 30 years after smallpox vaccination campaigns cease in the Democratic Republic of Congo. *Proc Natl Acad Sci U S A* **107**, 16262–16267 (2010).
12. Durski, K. N. *et al.* Emergence of Monkeypox — West and Central Africa, 1970–2017. *Morbidity and Mortality Weekly Report* **67**, 306 (2018).
13. Rizk, J. G., Lippi, G., Henry, B. M., Forthal, D. N. & Rizk, Y. Prevention and Treatment of Monkeypox. *Drugs* **82**, 957 (2022).
14. Craighead, J. E. Poxviruses. *Pathology and Pathogenesis of Human Viral Disease* 365–380 (2000) doi:10.1016/B978-012195160-3/50027-3.
15. Liu, L., Cooper, T., Howley, P. M. & Hayball, J. D. From Crescent to Mature Virion: Vaccinia Virus Assembly and Maturation. *Viruses* **6**, 3787 (2014).
16. Byrd, C. M. & Hruby, D. E. Vaccinia virus proteolysis—a review. *Rev Med Virol* **16**, 187 (2006).
17. Agbowuro, A. A., Huston, W. M., Gamble, A. B. & Tyndall, J. D. A. Proteases and protease inhibitors in infectious diseases. *Med Res Rev* **38**, 1295–1331 (2018).
18. Protease mechanisms | Learn Science at Scitable. <https://www.nature.com/scitable/content/protease-mechanisms-14462487/>.
19. Byrd, C. M., Bolken, T. C. & Hruby, D. E. Molecular Dissection of the Vaccinia Virus I7L Core Protein Proteinase. *J Virol* **77**, 11279 (2003).

20. Katritch, V. *et al.* Discovery of small molecule inhibitors of ubiquitin-like poxvirus proteinase I7L using homology modeling and covalent docking approaches. *J Comput Aided Mol Des* **21**, 549 (2007).
21. Otto, H. H. & Schirmeister, T. Cysteine Proteases and Their Inhibitors. *Chem Rev* **97**, 133–171 (1997).
22. Cysteine Protease Inhibitors | Cambridge MedChem Consulting.
https://www.cambridgemedchemconsulting.com/resources/hit_identification/focus/cysteine_protease_inhibitors.html.
23. Aleshin, A. E. *et al.* Activity, Specificity, and Probe Design for the Smallpox Virus Protease K7L. *J Biol Chem* **287**, 39470 (2012).
24. Byrd, C. M., Bolken, T. C. & Hruby, D. E. The Vaccinia Virus I7L Gene Product Is The Core Protein Proteinase. *J Virol* **76**, 8973 (2002).
25. Byrd, C. M. *et al.* New Class of Orthopoxvirus Antiviral Drugs That Block Viral Maturation. *J Virol* **78**, 12147 (2004).
26. MMS Infrastructures - MMS. <http://mms.dsfarm.unipd.it/mms-infrastructures.html>.
27. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **2021 596:7873** **596**, 583–589 (2021).
28. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modelling, prediction and analysis. *Nat Protoc* **10**, 845 (2015).
29. Molecular Operating Environment (MOE) | MOEsaic | PSILO.
<https://www.chemcomp.com/Products.htm>.
30. pKa and Tautomer Enumeration Software | Quacpac.
<https://www.eyesopen.com/quacpac>.
31. Korb, O., Stützle, T. & Exner, T. E. Empirical scoring functions for advanced protein-ligand docking with PLANTS. *J Chem Inf Model* **49**, 84–96 (2009).
32. Korb, O., Stützle, T. & Exner, T. E. An ant colony optimization approach to flexible protein–ligand docking. *Swarm Intelligence* **2007 1:2** **1**, 115–134 (2007).
33. Case, D. A. *et al.* The Amber biomolecular simulation programs. *J Comput Chem* **26**, 1668–1688 (2005).
34. Harvey, M. J., Giupponi, G. & de Fabritiis, G. ACEMD: Accelerating Biomolecular Dynamics in the Microsecond Time Scale. *J Chem Theory Comput* **5**, 1632–1639 (2009).
35. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J Mol Graph* **14**, 33–38 (1996).
36. CRISPR Screens: Approaches, Strategies, and Workflow.
<https://www.synthego.com/guide/crispr-methods/crispr-screen>.
37. Muhammed, M. T. & Aki-Yalcin, E. Homology modeling in drug discovery: Overview, current applications, and future perspectives. *Chem Biol Drug Des* **93**, 12–20 (2019).
38. Marcu, Ș. B., Tăbîrcă, S. & Tangney, M. An Overview of AlphaFold's Breakthrough. *Front Artif Intell* **5**, 112 (2022).
39. Mani, D., Wadhvani, A. & Krishnamurthy, P. T. Drug Repurposing in Antiviral Research: A Current Scenario. *Journal of Young Pharmacists* **11**, 117–121 (2019).

RINGRAZIAMENTI

Vorrei ringraziare i miei genitori che da sempre hanno creduto in me e mi hanno supportato non facendomi mai mancare niente, questo prezioso traguardo lo dedico a loro.

In generale vorrei ringraziare tutti i miei familiari e i miei amici, specialmente Giuseppe, Denis e Sally, per la pazienza avuta in questi anni, in particolar modo quest'ultimo, un anno di grandi sacrifici ma anche di grandi soddisfazioni.

Un ulteriore ringraziamento va ai compagni di corso che hanno deciso di condividere il loro cammino con il mio in questi cinque anni, tra tutti ci terrei a ringraziare Andrea che mi accompagna nel mio percorso di studi da ormai dieci anni.

Assieme ad Andrea, è d'obbligo un ringraziamento a Richard e ad Alessia che hanno vissuto con me sotto lo stesso tetto, condividendo momenti di ansia, tristezza e gioia.

Un ringraziamento speciale va a Sarah che da più di un anno mi sta accanto, sostenendomi nello studio e incoraggiandomi a seguire i miei sogni.

Per concludere vorrei dedicare un caloroso e profondo ringraziamento a tutto il gruppo dell'MMS: al professor Stefano Moro per la grande ispirazione e per la fiducia dimostrata incoraggiandomi a proseguire gli studi, a Davide, Anna, Veronica, Mattia e soprattutto Silvia e Matteo che mi hanno supportato e aiutato durante tutto il percorso di tesi, non facendo mai mancare attimi di sana goliardia che hanno reso leggere le giornate trascorse in laboratorio.