## Project design

**## Fase 1: Analisi del Target e Preparazione (1-2 settimane)**

L'obiettivo di questa fase è conoscere intimamente il tuo target. Non puoi colpire un bersaglio se non sai com'è fatto.

**I tuoi compiti:**

1. **Software Setup:** Assicurati di avere installato un visualizzatore molecolare. Ti consiglio **UCSF ChimeraX** (più moderno e potente) o **PyMOL**. Entrambi sono gratuiti per uso accademico.
2. **Esplorazione della Struttura:**
   * Vai sul sito del [Protein Data Bank (PDB)](https://www.rcsb.org/) e scarica la struttura con **PDB ID: 1YCR**.
   * Apri il file in ChimeraX/PyMOL. Vedrai due molecole: la proteina MDM2 (più grande) e un piccolo pezzo di p53 (un'alfa-elica).
   * **Identifica la tasca di legame:** Guarda dove si inserisce l'elica di p53. Quella è la tua "landing zone".
   * **Identifica i residui chiave:** L'interazione è dominata da tre amminoacidi di p53 che si infilano nella tasca di MDM2. Identificali usando il visualizzatore. Sono **Fenilalanina-19 (PHE), Triptofano-23 (TRP) e Leucina-26 (LEU)**. Questi sono i "punti caldi" (*hotspots*) che la tua mini-proteina dovrà mimare.
3. **Prepara il Target:** Rimuovi la catena del peptide p53 dalla struttura e salva un file PDB contenente **solo la proteina MDM2**. Questo sarà il tuo input per la fase di design.

**Risultato della fase:** Avrai una chiara comprensione tridimensionale del problema e un file PDB pulito del tuo target.

**## Fase 2: Design *de novo* con AI (2-3 settimane)**

Questa è la fase creativa. Useremo l'intelligenza artificiale per generare da zero una mini-proteina che si adatti alla tasca di MDM2.

**I tuoi compiti:**

1. **Generazione dello Scaffold con RFdiffusion:**
   * Userai **RFdiffusion**, probabilmente tramite un notebook Google Colab. Questo strumento è un modello generativo che disegna "forme" proteiche (scaffold).
   * Il tuo input sarà la struttura di MDM2 e una specifica sulla "zona target" (la tasca che hai identificato).
   * Il tuo "prompt" per l'AI sarà concettualmente: *"Disegna uno scaffold proteico di 40-60 amminoacidi che si leghi a questa superficie di MDM2"*.
   * Fai generare al modello circa 10-20 scaffold diversi.
2. **Design della Sequenza con ProteinMPNN:**
   * RFdiffusion crea la forma 3D (il backbone), ma non la sequenza di amminoacidi. Per questo userai **ProteinMPNN**.
   * Per ogni scaffold promettente generato da RFdiffusion, darai in pasto il suo backbone a ProteinMPNN.
   * Il suo compito è trovare la sequenza di amminoacidi ottimale che si ripiegherà spontaneamente in quella forma.

**Risultato della fase:** Avrai 10-20 sequenze di mini-proteine completamente nuove, progettate specificamente per legarsi a MDM2.

**## Fase 3: Validazione *in silico* e Filtraggio (1-2 settimane)**

Non tutti i design avranno successo. Ora dobbiamo filtrare i candidati migliori usando un "arbitro" computazionale.

**I tuoi compiti:**

1. **Controllo di "Piegabilità" con AlphaFold2:**
   * Per ogni sequenza ottenuta da ProteinMPNN, usa **AlphaFold2** per predire la sua struttura *da sola* (senza MDM2).
   * **Domanda chiave:** La struttura predetta da AlphaFold2 assomiglia allo scaffold disegnato da RFdiffusion? Questo test si chiama **"folding recyclability"**. Se la somiglianza è alta (basso RMSD), il design è buono. Se la proteina si piega in modo completamente diverso, scartala.
2. **Controllo del Legame con AlphaFold-Multimer:**
   * Prendi i 3-5 candidati migliori che hanno superato il test precedente.
   * Usa **AlphaFold-Multimer** per predire la struttura del complesso: la tua mini-proteina + la proteina MDM2.
   * **Domande chiave:** La mini-proteina si lega alla tasca giusta? L'orientamento è corretto?
   * **Metrica fondamentale:** Guarda il punteggio **ipTM** (interface predicted Template Modeling score). Un valore > 0.85 indica un'altissima confidenza che l'interfaccia di legame predetta sia corretta.

**Risultato della fase:** Avrai selezionato **1-2 candidati finali** di mini-proteine con un'alta probabilità computazionale di funzionare come previsto.

**## Fase 4: Simulazione di Dinamica Molecolare con GROMACS/OpenMM (3-4 settimane)**

Questa è la prova del nove. Verificheremo se il tuo miglior candidato è stabile e rimane legato al target in un ambiente dinamico e realistico.

**I tuoi compiti:**

1. **Setup del Sistema:**
   * Prendi il modello del complesso (il tuo miglior design + MDM2) dalla Fase 3.
   * Usa GROMACS/OpenMM per preparare il sistema: scegli un campo di forze (**force field**) come AMBER14SB, aggiungi le molecole d'acqua (solvatazione), aggiungi ioni per neutralizzare la carica.
2. **Simulazione:**
   * Esegui prima una minimizzazione dell'energia e poi una fase di equilibrazione (riscaldando lentamente il sistema).
   * Lancia la simulazione di "produzione" per almeno **100-200 nanosecondi**.
3. **Analisi della Traiettoria:**
   * **Stabilità:** Calcola l'**RMSD** per vedere se il complesso rimane stabile o se il tuo binder "vola via".
   * **Flessibilità:** Calcola l'**RMSF** (Root Mean Square Fluctuation) per vedere quali parti della tua mini-proteina sono rigide e quali flessibili.
   * **Interazioni:** Analizza i **legami idrogeno** e i contatti idrofobici tra il tuo binder e MDM2 per tutta la durata della simulazione. Rimangono stabili?

**Risultato finale del progetto:** Avrai un report completo che documenta il design, la validazione e l'analisi dinamica di un nuovo potenziale inibitore proteico di MDM2. Un lavoro di questo tipo sarebbe una base eccellente per una tesi o per una pubblicazione.

In bocca al lupo! È un progetto fantastico. Documenta ogni passaggio. 🚀

Chatgpt

**📅 Timeline Overview**

|  |  |  |  |
| --- | --- | --- | --- |
| **Phase** | **Goal** | **Expected Duration** | **Status** |
| 0 | Environment & toolchain up | 3–4 days | ☐ |
| 1 | Target analysis & cleanup | 1–2 weeks | ☐ |
| 2 | de novo scaffold design (RFdiffusion → ProteinMPNN) | 2–3 weeks | ☐ |
| 3 | in silico validation & filtering | 1–2 weeks | ☐ |
| 4 | MD simulation & free‑energy sanity check | 3–4 weeks | ☐ |

**Legend**   ☐ todo   🟡 in‑progress   ✅ done

Software

|  |  |  |
| --- | --- | --- |
| Category | Tool | Quick Install |
| Visualizer | **PyMOL** (open‑source) | conda install -c conda-forge -c schrodinger pymol-bundle |
| Gen Design | **RFdiffusion** (2024 commit) | devo capire che fare |
| Seq Design | **ProteinMPNN** | dovrebbe essere già isntallato |
| Structure Pred | **AlphaFold2** / **AlphaFold‑Multimer** (GPU) | uso su colab |
| MD Engine | **GROMACS 2025** | già installato con gpu |
| MD Engine | **OpenMM 8.1** | conda install -c conda-forge openmm |

# Theory

<https://aacrjournals.org/mcr/article/1/14/1001/232350/The-MDM2-p53-Interaction>

**Lunghezza e dominio**

* MDM2 umano: 491 aa; interazione mediata dal dominio N-terminale (aa 25–109).
* p53: transactivation domain N-terminale, α-elica di 15 aa (residui 18–26).

**Configurazione “key-lock”**

* Tasca idrofobica profonda in MDM2 (aa 26–108) formata da 14 residui idrofobici/aromatici.
* Elica di p53 si inserisce con il lato idrofobico (Phe19, Trp23, Leu26) nella tasca di MDM2.

**Hot-spots di interazione**

* **Residui p53 critici (18–26)**
  + Leu14, Phe19, Leu22, Trp23, Leu26 → **Phe19, Trp23 e Leu26** sono i più cruciali per binding e degradazione.
  + Thr18 stabilizza l’α-elica di p53; sua fosforilazione indebolisce l’interazione (~10-fold).
* **Residui MDM2 essenziali**
  + Mutazioni in Gly58, Glu68, Val75, Cys77 abolish binding a p53.

**Affinità e regolazione post-traduzionale**

* **Forza di legame**
  + K<sub>d</sub> 60–700 nM a seconda della lunghezza del peptide p53.
* **Fosforilazioni**
  + S15/S20 su p53: nessun effetto sul binding.
  + T18 su p53: riduce affinità 10×, necessario per dissociazione in risposta a danno al DNA.
  + In vivo: cascade S15→T18 necessaria per stabilizzazione di p53 dopo radiazioni.

**Strategie di inibizione e design di ligandi**

* **Peptidi stapled e mimetici**
  + Peptidi 12–15 aa derivati da phage display (consensus FxxL/YWxxL) con affinità superiore al p53 nativo.
  + Lead “12/1” (MPRFMDYWEGLN) → binding 50× più forte di p53 naturale.
  + Peptidi minimali (Ac-FMDYWEGLN) IC<sub>50</sub> ≈ 8.9 μM, successivamente ottimizzati 1 780-fold tramite non-natural aa e ciclizzazione.
* **Molecole small-molecule**
  + CP-31398: styrylquinazoline che ripristina conformatore wild-type di p53, blocca degradazione indipendentemente da MDM2.
  + Chlorofusin: metabolita di Fusarium sp. IC<sub>50</sub> ≈ 4.6 mM, lega dominio N-term di MDM2.
  + Derivati policiclici e chalcone in studio come inibitori non-peptidici.

## methods

### phase 1

**1.1 Grab & Inspect**

# fetch 1YCR directly from terminal (optional)

wget https://files.rcsb.org/download/1YCR.pdb

1. **Open** 1YCR.pdb in ChimeraX/PyMOL.
2. **Identify chains** – *A* = MDM2; *B* = p53 α‑helix (res 17‑28).
3. **Locate hot‑spots** – F19, W23, L26 (chain B). Use the “select” command:

select hotspots, chain B and resi 19+23+26

1. **Visual sanity check** – show MDM2 surface, cartoon the peptide, color hotspots.

**1.2 Target Cleanup**

# Remove peptide & junk, keep chain A only

delete chain B

remove solvent

save mdm2\_clean.pdb, chain A

Checklist:

* ☐ No peptide atoms
* ☐ No ligands/HOH unless needed
* ☐ Protonation states TBD (PropKa or pdb2gmx later)

**1.3 Pocket Mapping (optional bling)**

* Use **SiteMap** (Schrödinger) or **PockDrug** to map hydrophobic volumes.
* Screenshot pocket to include in thesis.

**Deliverables**  
• mdm2\_clean.pdb  
• 1–2 annotated screenshots highlighting FWL pocket

|  |  |  |  |
| --- | --- | --- | --- |
| Step | Code | Description | Note |
| In pymol | Select the protein sequence, delete che p53 sequence, delete water and save only as mdm2.pdb |  |  |
| * select p53\_interface, p53 within 4 of mdm2 * select mdm2\_interface, mdm2 within 4 of p53\_interface | Locate hot spots that my miniprotein have to mimick; define p53 and mdm2 interfaces |  |
| select p53 residue: select phe19, resn PHE and resi 19 and chain B  show sticks, phe19  color red, phe19  or  select hotspots, chain B and resi 19+23+26 |  |  |

Initialize repository in local:

* 1. git init in the project folder
  2. git remote add origin <https://github.com/fabianimarco98/mdm2-binder-design.git>
  3. git remote -v to verify the synchronization
     1. git add . add all the files
     2. git commit -m "Initial commit: Phase 1 files"
     3. git push (--set-upstream origin main)

### Phase 2: scaffold design

On google colab install RFdiffusion

|  |  |  |  |
| --- | --- | --- | --- |
| Step | Code | Description | Note |
| Terminal | touch RFdiffusion\_design.ipynb |  |  |

ColabDesign:

* **TrDesign** (TrRosetta-based design)
* **AfDesign** (AlphaFold-based design)
* **RfDesign** (RFdiffusion-based design)
* **ProteinMPNN** (sequence design da struttura)

A white background with black text

AI-generated content may be incorrect.

* il sito attivo è il sito dove si esplica l’attività della protein
* scelta dello scaffold è basata su un’analisi strutturale, cioè un confronto di strutture
* scegliere uno scaffold che sia strutturalmente simile, cioè che abbia similarità strutturali con il sito attivo della proteina di grandi dimensioni
* si sceglie lo scaffold bisogna trasferire il sito attivo dalla proteina di grandi dimensioni allo scaffold

1. **Preparazione del motivo p53**
   * Estrai dal PDB (1YCR) la α-elica p53 (resi 18–26).
   * Allinea in un sistema di riferimento locale (calcola asse, passo elicoidale).
2. **Costruzione del database di scaffold**
   * Usa strumenti di ricerca strutturale (es. **MASTER**, **GESAMT**)
     + Query: eliche 9–15 aa
     + Criterio: RMSD Cα < 0.5 Å con il motivo p53
   * Filtra per risoluzione (< 2.5 Å), B-factor, esposizione in superficie.
3. **Selezione degli scaffolds candidati**
   * Verifica che l’elica del candidato sia esposta (no core buried).
   * Controlla la presenza di loop adiacenti (per un grafting più pulito).
   * Mantieni 5–10 top hit per prova sperimentale.
4. **Grafting del motivo p53 sullo scaffold**
   * 1. **Rigid-body superposition**
        1. Allinea le Cα del motivo p53 sulle Cα dell’elica scaffold.
     2. **Sostituzione backbone**
        1. Rimpiazza i CA/N/C del scaffold con quelli del motivo p53 mantenendo i side-chain del motivo.
     3. **Loop closure**
        1. Chiudi i gap con Rosetta Remodel, KIC o Modeller per garantire continuità della catena.
5. **Ripianificazione dell’interfaccia**
   * Usa **ProteinMPNN** o **RosettaDesign**
     + Fissa i *hotspot* p53 (Phe19, Trp23, Leu26).
     + Riprogetta residui circostanti sullo scaffold per migliorare packing e idrofobicità complementare.
6. **Validazione strutturale**
   * Esegui **AlphaFold2-Multimer** sul complesso MDM2 + peptide graftato:
     + Controlla *pAE\_interaction* < 10
     + Verifica un buon pLDDT (> 70) sull’elica
   * Scarta i design con interfaccia mal posizionata.
7. **Refinement e stabilizzazione**
   * Applica **lactam-bridge** o **hydrocarbon staple** su due posizioni *i*–*i+4* (lato opposto hotspot).
   * Fai un breve **MD** (50–100 ns) con GROMACS/OpenMM
     + RMSD dell’elica
     + Numero di H-bond tra peptide e MDM2
     + RMSF dei residui di binding
8. **Selezione finale**
   * Ordina 2–3 design con:
     + Massima stabilità elica (basso RMSD/RMSF)
     + Interfaccia stabile (costanti H-bond, mindist < 4 Å)
   * Procedi con sintesi e test biologici.

python ~/programmi/proteinml/ProteinMPNN/protein\_mpnn\_run.py

|  |  |  |  |
| --- | --- | --- | --- |
| Step | Code | Description |  |
| Preparazione del motivo p53 | select p53\_helix, chain B and resi 18-26 |  |  |
| show cartoon, p53\_helix  zoom p53\_helix |  |  |
| save p53\_helix.pdb, p53\_helix |  |  |
| Costruzione del database di scaffold |  | Query contro pdb\_db della mia catena B frammento di p53 |  |
|  |  | Scarico risultati come csv |  |
|  | Script python per trovare i migliori match, scaricare i pdb, estrarre le catene e pulirle |  |  |

### Phase 3: transfer the motif on some scaffolds

 **Individua i residui chiave:** Analizzando la struttura, noterai che il frammento di p53 lega MDM2 tramite un'elica. I tre residui cruciali di p53 su questa elica sono:

* **Fenilalanina 19 (F19)**
* **Triptofano 23 (W23)**
* **Leucina 26 (L26)**

 **Obiettivo:** Il tuo scopo è posizionare questi tre residui su una alfa-elica del tuo scaffold in modo che mimino perfettamente la loro disposizione spaziale e le loro interazioni come si vedono nel complesso originale.

1. **Allineamento Strutturale:** Non basta un allineamento di sequenza. Devi eseguire un **allineamento strutturale** tra l'alfa-elica di p53 (presa dalla struttura 1YCR) e le eliche presenti sui tuoi scaffold. Strumenti come **PyMOL** o **UCSF Chimera** sono perfetti per questo. Devi trovare un'elica sullo scaffold che sia esposta al solvente e che si sovrapponga bene all'elica di p53.
2. **Identificazione delle Posizioni:** Una volta trovata la migliore sovrapposizione, identifica quali posizioni (numeri di residuo) sul tuo scaffold corrispondono a F19, W23 e L26 di p53.
3. **Mutagenesi In Silico:** Usando le funzioni di "mutagenesis" di PyMOL o Chimera, sostituisci i residui originali del tuo scaffold in quelle posizioni con Fenilalanina, Triptofano e Leucina. Presta attenzione a scegliere il rotamero (la conformazione della catena laterale) che meglio mima quello del p53 originale.