GRADUATE SCHOOL
ISTANBUL MEDENIYET UNIVERSITY
DEPARTMENT OF BIOENGINEERING

Discovery of Small-Molecule Inhibitors and Therapeutic Peptide Design Targeting *Clostridium difficile* Infection

Master's Thesis

Damla Nur Çamlı

June 2025

GRADUATE SCHOOL
ISTANBUL MEDENIYET UNIVERSITY
DEPARTMENT OF BIOENGINEERING

Discovery of Small-Molecule Inhibitors and Therapeutic Peptide Design Targeting *Clostridium difficile* Infection

Master's Thesis

Damla Nur Çamlı

Supervisor

Asst. Prof. Saliha Ece Acuner Zorluuysal

June 2025

THESIS JURY APPROVAL

This Master thesis titled" Discovery of Small-Molecule Inhibitors and Therapeutic Peptide Design Targeting *Clostridium difficile* Infection" written by Damla Nur CAMLI at the Department of Bioengineering was accepted by our jury.

JURY MEMBERS **SIGNATURE**

Supervisor:

Asst. Prof. Saliha Ece Acuner Zorluuysal

İstanbul Medeniyet University

Other Jury Members:

Prof. Dr. Pınar Obakan Yerlikaya

İstanbul Medeniyet University

Assoc. Prof. Dr. Mehmet Özbil

Gebze Technical University

Date of Defence: 16/06/2025

STATEMENTS

Style and Reference Manual Statement

Having reviewed this thesis written under my supervision, I confirm that it has been written in accordance with Vancouver Manual of Style and used its foot-note reference format consistently throughout the entire text.

Signature
Asst. Prof. Saliha Ece Acuner Zorluuysal

Declaration of Originality

I hereby declare that all information in this dissertation has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conducts, I have fully cited and referenced all material and results that are not original to this work.

Signature
Damla Nur Çamlı

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to all the individuals and institutions who supported me throughout the process of preparing this thesis.

First and foremost, I am sincerely thankful to my advisor, Saliha Ece Acuner Zorluuysal, whose expertise and guidance have been invaluable, and whose continuous encouragement kept me motivated from start to finish. I would also like to extend my heartfelt thanks to all members of SEA Lab for their support and collaboration.

A very special thank you goes to Hacı Aslan Onur İşcil — more than a friend, he became a true companion and brother during this journey. His unwavering academic and emotional support made this experience not only more manageable but truly meaningful.

To my beloved mother who Meliha Çamlı and my entire family, your unconditional love has always been my greatest source of strength. And to the quietest yet most comforting presence throughout this process, my cat Mia your calm companionship and gentle purring softened even the toughest days.

With all my heart, thank you.

GENİŞ ÖZET

Clostridium Difficile Enfeksiyonuna Yönelik İnhibitör Potansiyeline Sahip Küçük-Molekül Keşfi ve Terapötik Peptid Tasarımı

Çamlı, Damla Nur

Yüksek Lisans Tezi, Biyomühendislik Anabilim Dalı

Danışman: Dr. Öğr. Üyesi Saliha Ece Acuner Zorluuysal

Haziran 2025

Clostridium difficile (*C. difficile*), insan gastrointestinal sisteminde ciddi enfeksiyonlara neden olan Gram pozitif, spor oluşturan ve anaerobik bir bakteridir (1). Özellikle antibiyotik kullanımı sonrası bağırsak mikrobiyotasının bozulması sonucu ortaya çıkan bu patojen, klinik olarak hafif ishalden, psödomembranöz kolit ve toksik megakolon gibi hayatı tehdit eden ağır durumlara yol açabilmektedir (2,3). Bu enfeksiyonlar, özellikle yaşlı bireyler ve hastane ortamında yatan hastalar arasında daha yaygındır ve yüksek morbidite ile mortaliteye neden olmaktadır (1). Dolayısıyla, *Clostridium difficile* enfeksiyonları (CDI) dünya çapında önemli bir halk sağlığı problemi olarak kabul edilmektedir (4).

Clostridium difficile enfeksiyonları'nın (CDI) patogenezinde başlıca virülens faktörleri, Toksin A (TcdA) ve Toksin B (TcdB) olarak bilinmektedir (5). Bu iki toksinden özellikle Toksin B, yüksek sitotoksisiteye sahip olması nedeniyle hastalığın oluşumunda daha baskın bir role sahiptir (6). TcdB, glukoziltransferaz aktivitesi yoluyla konak hücrelerdeki Rho ailesi GTPaz proteinlerini (Rac1, RhoA, Cdc42 gibi) modifiye eder (5). Bu modifikasyon, aktin iskeletinin bozulmasına yol açarak hücre iskeletinin dinamiklerini etkiler, hücrelerde apoptoz ve yapısal bütünlük kaybı gibi patolojik değişikliklerin ortaya çıkmasına neden olur (7).

TcdB'nin hastalık patogenezindeki merkezi rolü, bu toksini terapötik müdahaleler açısından öncelikli ve stratejik bir hedef haline getirmektedir. Mevcut antibiyotiklerin özellikle yüksek nüks oranları ve bağırsak mikrobiyotasına zarar verme potansiyeli nedeniyle, toksin hedefli tedavi yaklaşımları ön plana çıkmaktadır (1). Bu çalışmanın temel amacı, TcdB ve konak proteinleri arasındaki moleküler mekanizmayı anlamak ve bu doğrultuda TcdB'yi doğrudan hedef alarak inhibitör küçük moleküller ve terapötik peptidler tasarlamaktır. Nihai amaç olarak ise, *Clostridium difficile* enfeksiyonuna yönelik yenilikçi tedavi stratejileri geliştirmek hedeflenmektedir.

In Silico Yöntemlerin Araştırmadaki Rolü ve Önemi

Günümüzde özellikle hedefe yönelik ilaç keşfi süreçlerinde *in silico* yaklaşımlar, modern biyomedikal araştırmaların vazgeçilmez araçları haline gelmiştir. "*In silico*" terimi, bilgisayar destekli ortamda gerçekleştirilen moleküler modelleme, simülasyon ve hesaplamalı analizleri ifade eder (8). Bu yöntemler, geleneksel laboratuvar tekniklerinin zamansal ve finansal sınırlamalarını aşarak, geniş moleküler kütüphaneler içerisinden kısa sürede ve maliyet etkin biçimde potansiyel aday bileşiklerin taranmasına olanak tanımaktadır. Aynı zamanda, bu yaklaşımlar erken aşamada deneysel doğrulama gerektiren moleküllerin daha rasyonel bir şekilde seçilmesini sağlayarak, ilaç keşfi süreçlerinin etkinliğini artırmaktadır (8).

In silico yaklaşımların en önemli avantajlarından biri de, moleküler düzeyde atomik etkileşimlerin ve dinamiklerin yüksek çözünürlükte analiz edilebilmesidir (8). Özellikle moleküler dinamik simülasyonları, protein-ligand etkileşimlerinin zaman içindeki davranışlarını, bağlanma süreçlerini ve yapısal stabiliteyi ayrıntılı biçimde ortaya koyar. Bu sayede, deneysel olarak yakalanması zor olan geçici ara yapılar ve moleküler mekanizmalar hakkında kapsamlı bilgiler elde edilebilir (9).

Bu tez kapsamında, *Clostridium difficile* TcdB toksininin konak hücre GTPaz proteinleri (Rac1 ve Cdc42) ile olan etkileşimleri, moleküler dinamik simülasyonları ve serbest enerji hesaplamaları aracılığıyla incelenmiştir. Bu analizler doğrultusunda, toksin-inhibitör bağlanma bölgelerinin yapısal stabilitesi, seçiciliği ve etkileşim mekanizmaları moleküler düzeyde tanımlanmıştır.

Ayrıca *in silico* yöntemler, etik ve deneysel risklerin azaltılmasında da önemli bir rol üstlenmektedir. Yeni geliştirilen moleküllerin toksisite ve farmakokinetik profilleri, deneysel testler öncesinde sanal ortamda tahmin edilerek gereksiz hayvan deneyleri ve laboratuvar testlerinin önüne geçilebilir. Bu tezde de benzer şekilde, tasarlanan terapötik peptidlerin toksisite tahminleri ToxinPred gibi biyoinformatik araçlarla gerçekleştirilmiş ve yalnızca toksik olmayan adaylar değerlendirmeye alınmıştır (10,11).

Bunun yanı sıra, *in silico* analizler birden fazla parametrenin eş zamanlı değerlendirilmesine olanak tanır (12). Bağlanma enerjisi, yapı esnekliği, hidrojen bağları, elektrostatik etkileşimler gibi çok sayıda yapısal ve dinamik özellik birlikte analiz edilerek, yalnızca hedefe yüksek afiniteyle bağlanan değil, aynı zamanda yapısal kararlılığa sahip inhibitör adaylarının belirlenmesi mümkün olur (10,13). Böylece ilaç tasarım süreci daha bütüncül ve güvenilir hale getirilir.

TcdB'nin Yapısal ve Fonksiyonel Analizi

Clostridium difficile toksini B (TcdB), dört ana fonksiyonel bölgeden oluşur: N-terminal glukoziltransferaz domeni (GTD), otokatalitik proteaz domeni (CPD), ve C-terminal tekrarlayan oligopeptit dizileri (CROPs) (14). Bu modüler yapı, toksinin konak hücre proteinleriyle etkileşiminde ve toksik etkilerinin oluşmasında kritik rol oynar (15,16). GTD, Rho ailesi GTPaz proteinlerini (örneğin Rac1, RhoA, Cdc42) hedef alarak glukozilasyon yoluyla işlevlerini inhibe eder (16). CPD, inositol hekzafosfat (IP6) tarafından aktive edilen

otokatalitik aktivite ile GTD'yi serbestleştirir (5). TD, toksinin endozomal membranlardan sitozole geçişini sağlar (5). CROPs bölgesi ise hücre yüzeyindeki glikanlara bağlanarak hücreye özgü tanımayı gerçekleştirir (1). TcdB'nin konak GTPaz proteinleriyle etkileşimini anlamak amacıyla, Protein Data Bank (PDB) (17) veritabanındaki kristal yapılar kullanılarak yapısal analizler gerçekleştirilmiştir. Bu analizler, bağlanma bölgelerinin seçiciliğini tanımlamak ve yapı-temelli inhibitör tasarımı geliştirmek açısından önemlidir.

Bu tez kapsamında yapılan ve literatürde "MuGger Toxins: Exploring the Selective Binding Mechanism of Clostridial Glucosyltransferase Toxin B and Host GTPases" başlığıyla yayımladığımız çalışmada, TcdB'nin Rac1 ve Cdc42 ile etkileşimleri moleküler dinamik simülasyonlar ve serbest enerji hesaplamaları ile detaylı biçimde incelenmiştir (18). Elde edilen bulgular, toksin-seçicilik ilişkisini moleküler düzeyde ortaya koymuş ve hedefe yönelik inhibitör geliştirme sürecine katkı sağlamıştır (18).

Moleküler Dinamik Simülasyonları ve Modellenen Kompleks Yapıların Kararlılık Çalışmaları

TcdB ve konak proteinlerin oluşturduğu kompleksler, GTP ve GDP bağlanmış Rac1, Cdc42, Cd46 ve Nectin3 konak proteinleri kullanılarak modellenmiş ve moleküler dinamik simülasyonları yapılmıştır. GROMACS (19) programı ve CHARMM36m (20,21) kuvvet alanı kullanılarak üç farklı replikasyon halinde 500 ns'lik simülasyonlar yürütülmüştür. Bu simülasyonlar, toksin-konak etkileşimlerinin dinamik doğasını ve bağlanma kararlılığını (stabilitesini) ayrıntılı şekilde ortaya koymuştur.

Simülasyonların analizinde RMSD, RMSF, hidrojen bağı sayısı ve MM-PBSA bağlanma enerjisi gibi parametreler hesaplanarak, toksinin bağlanma bölgelerinin stabilitesi ve inhibitör tasarımındaki kritik önemi belirlenmiştir.

Gelişmiş Örnekleme Teknikleri: Metadinamik ve Umbrella Sampling

TcdB-Rac1 kompleksinin dinamik bağlanma süreçlerini daha detaylı incelemek için Well-Tempered Metadynamics (WT-MetaD) (22) simülasyonları gerçekleştirilmiştir. Konak GTPaza GTP ve GDP bağlı formdaki yapıları içeren kompleksler için yapılan simülasyonlar, potansiyel serbest enerji yüzeylerinin şekillenmesini ortaya koymuştur. Merkezler arası mesafe ve yarıçap gibi kolektif değişkenler kullanılarak, bağlanma yolları moleküler düzeyde tanımlanmıştır.

Buna ek olarak umbrella sampling yöntemiyle potansiyel serbest enerji (PMF) eğrileri hesaplanmış ve WHAM algoritması ile analiz edilmiştir (19,22). Bu çalışmalar, toksin-GTPaz seçiciliğinin enerji temelli anlaşılmasına katkı sağlamıştır.

Küçük Molekül Sanal Taraması ve İlaç Adayları

TcdB'nin aktif bölgesine bağlanabilecek inhibitör küçük moleküller, ZINC (21) veritabanından seçilen binlerce bileşik üzerinde sanal tarama yöntemi ile incelenmiştir. Moleküller, Lipinski kuralı, kan-beyin bariyeri geçişi ve ADMET özellikleri açısından değerlendirilmiştir.

Seçilen dört molekül, bağlanma enerjileri ve farmakofor özelliklerine göre belirlenmiş ve UDP-glukoz bağlanma bölgesine bağlanarak toksin aktivitesini engelleme potansiyeli göstermiştir. Bu bağlanma, toksinin hücre içine girmeden işlevini bloke etmesine olanak sağlamaktadır.

Terapötik Siklik Peptid Tasarımı ve Toksisite Analizleri

TcdB'nin konak proteinlerle etkileşimde bulunduğu bölgeler baz alınarak terapötik peptid dizileri tasarlanmıştır. Tasarlanan peptidlerin bağlanabilirlikleri moleküler kenetlenme (docking) analizleri ile değerlendirilmiş, toksisite ise ToxinPred (11) ve BLAST (23) gibi araçlarla analiz edilmiştir.

Toksik olmayan ve insan proteinleri ile düşük homoloji gösteren iki siklik peptid AMBER14SB (24) kuvvet alanı ve TIP3P su modeli altında moleküler dinamik simülasyonlarına tabi tutulmuştur. RMSD, RMSF, hidrojen bağları, zamanla bağlanma mesafeleri ve MM-GBSA bağlanma enerjileri hesaplanarak peptidlerin toksinin bağlanma bölgelerinde yüksek afinite ve stabilite gösterdiği tespit edilmiştir.

Entegre Enerji Analizleri ve Karşılaştırmalı Simülasyonlar

Küçük molekül ve peptid inhibitörlerin TcdB ile oluşturduğu komplekslerin bağlanma enerjileri MM-GBSA yöntemi ile karşılaştırılmıştır. Siklik peptidlerin hem fiziksel bağlanma hem de UDP bölgesi ile etkileşim yoluyla çift yönlü blokaj mekanizması geliştirdiği gözlemlenmiş, bu da inhibitör etkinliğini artırmıştır.

Sonuç ve Tartışma

Bu çalışmada, Clostridium difficile (C. difficile) bakterisinin toksinlerinden özellikle TcdB'nin yapısal ve fonksiyonel özellikleri detaylı bir şekilde moleküler düzeyde incelenmiştir. TcdB'nin, konak hücrelerdeki Rho ailesi GTPaz proteinlerine bağlanarak toksik etkisini gerçekleştirdiği bilinmektedir (14) . Bu çalışmada, bu bağlanma mekanizması bilgisayar ortamında simüle edilerek toksinin davranışları ve etkileşimleri daha derinlemesine analiz edilmiştir.

Gerçekleştirilen moleküler dinamik (MD) simülasyonları, TcdB'nin farklı GTPaz proteinleri ile oluşturduğu komplekslerin zaman içindeki davranışlarını ortaya koymuştur. Bu simülasyonlar sayesinde toksinin hedef proteinlere bağlanma stabilitesi ve seçiciliği hakkında önemli veriler elde edilmiştir. Özellikle, toksinin aktif bölgelerindeki bağlanma dinamiklerinin, inhibitör geliştirme sürecinde kritik rol oynadığı görülmüştür.

Bu noktada, bilgisayar destekli *in silico* yöntemlerin araştırmalardaki önemi oldukça büyüktür. Geleneksel laboratuvar deneylerine kıyasla bu yöntemler, çok daha hızlı, düşük maliyetli ve yüksek verimlidir (12). Örneğin, binlerce küçük molekülü ve peptidi sanal ortamda test etme olanağı, laboratuvar ortamında yıllar sürebilecek tarama işlemlerinin saatler içinde tamamlanmasını mümkün kılmaktadır. Böylece, en umut verici aday moleküller belirlenerek deneysel aşamalara odaklanılabilir; bu da ilaç keşfi süreçlerini hızlandırmakta ve maliyetleri önemli ölçüde azaltmaktadır (25).

Çalışmamızda, TcdB'nin hedef proteinlerle etkileşiminde kritik öneme sahip bağlanma bölgeleri moleküler modelleme ve simülasyon teknikleriyle detaylandırılmıştır. Bu sayede toksinin bağlanma şekli ve seçiciliği moleküler düzeyde daha net ortaya konmuştur. Ayrıca, toksin ile GTP ve GDP bağlı GTPaz proteinleri arasındaki etkileşimlerin termodinamik ve kinetik özellikleri gelişmiş simülasyon yöntemleriyle (Well-Tempered Metadynamics, Umbrella Sampling) analiz edilmiş ve bağlanma süreçlerinin ayrıntılı mekanizması ortaya konmuştur.

Bu kapsamlı moleküler analizler yalnızca toksinin doğal bağlanma dinamiklerini ortaya koymakla kalmamış, aynı zamanda toksini doğrudan hedef alan küçük moleküller ve peptidlerin tasarımına da zemin hazırlamıştır. Zengin moleküler kütüphanelerden seçilen potansiyel inhibitörler, sanal tarama ve farmakolojik filtreleme süreçlerinden geçirilmiştir. Bu moleküllerin toksinin işlevini engelleme potansiyeli, bağlanma enerjileri ve yapısal uygunlukları doğrultusunda değerlendirilmiştir.

Özellikle tasarlanan siklik peptidlerin, toksinin bağlanma bölgelerinde yüksek stabilite ve afinite gösterdiği belirlenmiştir. Bu peptidlerin toksisite analizleri, *in silico* biyoinformatik araçlar (örneğin ToxinPred) kullanılarak yapılmış ve insan proteinleri ile düşük sekans homolojisi gösteren, toksik olmayan adaylar saptanmıştır. Bu durum, terapötik uygulamalarda bu peptidlerin olası yan

etkilerinin düşük olacağını göstermektedir. Moleküler dinamik simülasyonları, bu peptidlerin toksinle oluşturduğu komplekslerin stabilite analizlerini doğrulamış ve bu yapıların hücresel ortamda etkin inhibisyon sağlayabileceğini ortaya koymuştur.

Bu çalışmanın önemli bulgularından biri de, siklik peptidlerin toksini hem fiziksel olarak bağlayarak hem de UDP-glukoz bağlanma bölgesine müdahale ederek çift yönlü bir inhibisyon mekanizması sergilemesidir. Bu durum, toksinin işlevinin baskılanmasında etkili bir strateji olarak değerlendirilebilir. Ayrıca, küçük moleküllerle peptidlerin karşılaştırmalı bağlanma enerjileri analiz edilmiş (MM-PBSA yöntemiyle) ve siklik peptidlerin bağlanma enerji profilinin oldukça olumlu olduğu gözlemlenmiştir. Bu da bu peptidlerin yüksek verimlilikte inhibitör adayları olduğunu ortaya koymaktadır.

Bu çalışmanın genel bulguları, toksin-hedefli tedavi yaklaşımlarının, özellikle yüksek nüks riski taşıyan ve antibiyotik direnci gelişmiş *C. difficile* enfeksiyonlarında alternatif tedavi stratejileri sunabileceğini göstermektedir. Mevcut antibiyotik tedavilerinin mikrobiyota üzerindeki olumsuz etkileri ve tekrar enfeksiyon riskleri göz önüne alındığında, toksin inhibisyonu enfeksiyonun etkilerini azaltmak için umut vadeden bir yaklaşımdır.

Bununla birlikte, bu *in silico* yöntemlerle elde edilen bulguların deneysel doğrulaması ve klinik öncesi testlerle desteklenmesi gereklidir. Tasarlanan moleküllerin hücre kültürü ve hayvan modellerinde etkinliği ve güvenliği ayrıntılı olarak değerlendirilmelidir. Bu süreç, ilaç geliştirme yolunda temel bir adımdır ve klinik uygulamalara geçiş için zorunludur. Ancak bu çalışma, moleküler düzeyde toksin etkileşimlerinin detaylı incelenmesi ve etkili inhibitörlerin rasyonel tasarımı açısından önemli bir başlangıç noktası oluşturmuştur.

Sonuç olarak, bu tezde kullanılan ileri düzey moleküler modelleme, simülasyon ve sanal tarama tekniklerinin, ilaç keşif süreçlerini hızlandırma ve maliyetleri düşürme açısından büyük potansiyel taşıdığı gösterilmiştir. Ayrıca, toksin-hedefli inhibitörlerin geliştirilmesi, C. difficile enfeksiyonları ile mücadelede yenilikçi ve etkili tedavi alternatifleri sunma potansiyeline sahiptir. Gelecekte, bu yaklaşımların laboratuvar ve klinik uygulamalarla bütünleştirilmesi, enfeksiyon hastalıkları alanında önemli ilerlemelere katkı sağlayacaktır.

Anahtar Kelimeler: Clostridium difficile enfeksiyonu, TcdB, moleküler dinamik simülasyon, küçük molekül keşfi, terapötik peptid tasarımı.

ABSTRACT

Discovery of Small-Molecule Inhibitors and Therapeutic Peptide Design Targeting *Clostridium difficile* Infection

Çamlı, Damla Nur

Master's Thesis, Department of Bioengineering

Supervisor: Asst. Prof. Saliha Ece Acuner Zorluuysal

June 2025

Clostridium difficile is a Gram-positive, anaerobic, spore-forming bacterium responsible for gastrointestinal infections that range from mild diarrhea to life-threatening conditions. Among the virulence factors, Toxin B plays a critical role due to its potent cytotoxic effects. Toxin B modifies host Rho and Ras family GTPases, selectively in their GDP-bound form, by glycosylation, disrupting the actin cytoskeleton, which leads to cellular damage and apoptosis. Current antibiotic treatments often result in high relapse rates and disrupt the gut microbiome, highlighting the need for alternative therapeutic strategies. Targeting TcdB directly represents a promising approach to mitigate disease progression. This study employs computational techniques to design small molecule inhibitors and therapeutic peptides that specifically inhibit TcdB activity. Molecular docking, MD simulations, and binding energy calculations were utilized to explore the interactions between TcdB and host GTPases, as well as potential inhibitors. Enhanced sampling methods provided insights into the binding processes and complex stability. Virtual screening of extensive compound libraries identified candidates that bind effectively to TcdB's active site. Furthermore, cyclic peptides engineered based on interaction hotspots showed strong binding affinity and stability, supported by toxicity prediction tools indicating low adverse effects. Comparative analyses suggest that cyclic peptides inhibit TcdB via a dual mechanism, not only by occupying the toxin's binding interface with the host GTPases but interfering also with the native UDP-glucose binding region. These findings support the development of TcdB-targeted therapies as viable alternatives to antibiotics in treating the infection. Experimental and *in vivo* studies are needed to translate computational results.

Keywords: *Clostridium difficile* infection, Toxin B, MD simulation, Small-Molecule Discovery, Therapeutic Peptide Design

List of Tables

List of Figures

ABBREVIATIONS

BBB: Blood–Brain Barrier.

CDI: Clostridium difficile Infection.

Cdc42: Cell Division Cycle 42, a member of the Rho GTPase family.

CHARMM-GUI: Graphical User Interface for CHARMM-based system preparation.

CVs: Collective Variables.

DEG: Differentially Expressed Gene.

DRBD: Delivery and Receptor Binding Domain.

ECFP: Extended-Connectivity Fingerprint.

FID / Fidaxomicin: A macrocyclic antibiotic used for treating CDI.

FMT: Fecal Microbiota Transplantation.

GIA: Gastrointestinal Absorption.

GDP: Guanosine Diphosphate.

GNP: Guanosine 5'- [β, γ-imido] triphosphate.

GROMACS: GROningen MAchine for Chemical Simulations.

GTP: Guanosine Triphosphate.

GTPase: Guanosine Triphosphatase (enzyme family).

GTD: Glucosyltransferase Domain.

HADDOCK: High Ambiguity Driven Biomolecular DOCKing.

HBA: Hydrogen Bond Acceptor.

HBD: Hydrogen Bond Donor.

InhA: Enoyl- [acyl carrier protein] reductase, target in tuberculosis.

InsP6: Inositol Hexakisphosphate.

ITP: Include Topology Parameters (GROMACS).

MD: Molecular Dynamics.

Mg^{2+}: Magnesium Ion.

Mn^{2+}: Manganese Ion.

MM-GBSA: Molecular Mechanics/Generalized Born Surface Area.

MM-PBSA: Molecular Mechanics/Poisson–Boltzmann Surface Area.

MMPBSA.py: Python implementation of MM-PBSA calculations.

MCP (CD46): Membrane Cofactor Protein.

NPT: Constant Number of Particles, Pressure, and Temperature.

NVT: Constant Number of Particles, Volume, and Temperature.

PAINS: Pan-Assay Interference Compounds.

PB: Poisson–Boltzmann.

PDB: Protein Data Bank.

PLUMED: Plugin for Molecular Dynamics simulations with enhanced sampling.

PMF: Potential of Mean Force.

RAC1: Ras-related C3 botulinum toxin substrate 1.

RBD: Receptor Binding Domain.

RDP: Relative Distance Parameter.

RhoA/B/C: Rho family small GTPases A/B/C.

RMSD: Root Mean Square Deviation.

RMSF: Root Mean Square Fluctuation.

Rg: Radius of Gyration.

SASA: Solvent Accessible Surface Area.

SDC: Sidechain Decomposition Contribution.

SMILES: Simplified Molecular Input Line Entry System.

TcdA: Toxin A.

TcdB: Toxin B.

TIP3P: Transferable Intermolecular Potential with 3 Points (water model).

TPSA: Topological Polar Surface Area.

UDP: Uridine Diphosphate.

UNIPROT: Universal Protein Resource.

WHAM: Weighted Histogram Analysis Method.

WT-MetaD: Well-Tempered Metadynamics.

ZINC: ZINC Is Not Commercial — ligand discovery database.

ΔG: Gibbs Free Energy Change

TABLE OF CONTENTS

INTRODUCTION

Clostridium difficile is a Gram-positive, anaerobic bacterium capable of forming spores, which primarily causes infections when the gut microbiota is disrupted, often because of antibiotic treatment (3). This microorganism is a leading agent of healthcare-associated infections and represents a significant risk especially for elderly and immunocompromised patients (26). The clinical condition it induces, known as *C. difficile* infection (CDI), is largely attributed to two major exotoxins: Toxin A (TcdA) and Toxin B (TcdB) (14,27). TcdB facilitates its entry into host cells by interacting with multiple surface receptors, subsequently inactivating small GTPase proteins inside the cell through its glucosyltransferase enzymatic activity (28,29). Among its targets, Rac1—a small GTPase involved in regulating the cytoskeleton—is notably affected (29). The interference of TcdB selectively with GDP-bound Rac1 leads to a range of cellular dysfunctions, impacting cell structure and function (30). Deciphering the atomic-level details of this toxin-protein interaction is essential for understanding the molecular mechanisms underpinning toxin action and disease progression. Thanks to advances in computational biology, these interactions can be studied comprehensively using *in silico* approaches such as molecular modelling, molecular dynamics simulations, docking and virtual screening, alongside traditional experimental methods. This thesis focuses on exploring the binding interaction between TcdB and Rac1 with the aim of identifying promising small-molecule and peptide inhibitors. By mapping the binding interfaces and characterizing structural interaction sites, therapeutically relevant target regions were delineated. The outcomes are intended to support the discovery of novel molecular strategies to combat *C. difficile* infections. This study investigates the structural and dynamic interactions between Clostridium difficile Toxin B (TcdB) and host GTPase proteins Rac1 and Cdc42. An integrative approach was employed, combining structural modelling, molecular dynamics (MD) simulations, advanced sampling techniques such as well-tempered meta-dynamics and umbrella sampling, binding free energy calculations

using MM-PBSA, and docking studies involving potential therapeutic small molecules and peptides. This multi-level strategy aims to elucidate the molecular mechanisms of TcdB-host interactions and to identify novel inhibitors targeting critical binding interfaces. This thesis focuses on exploring the binding interaction between TcdB and Rac1 with the aim of identifying promising small-molecule and peptide inhibitors. By mapping the binding interfaces and characterizing structural interaction sites, therapeutically relevant target regions were delineated. The outcomes are intended to support the discovery of novel molecular strategies to combat C. difficile infections.

1. LITERATURE REVIEW

1.1. Microbiota

The human microbiota is a broad-spectrum concept that can be defined as a small ecosystem within itself (31). In this community, which can be referred to as an ecosystem, bacteria are among the main organisms (31). Different species are concentrated in different regions of the human body (32). Among these regions, the most notable is the gut microbiota (33). In this area, a wide variety of microorganisms coexist with human intestinal cells, engaging in both informational and material exchange, thereby maintaining a certain balance (34). The disruption of this balance can lead to various diseases, a condition referred to as dysbiosis (35). One of the key bacteria involved in disrupting this balance is *Clostridium difficile* (36).

1.2. *Clostridium difficile* and *Clostridium difficile* Infection

Clostridia are motile, rod-shaped, obligate anaerobic, Gram-positive, spore-forming bacteria commonly found in nature, particularly in soil (31). *Clostridium difficile* exists in two distinct forms during its life cycle: the vegetative form, which replicates within the host but cannot survive in external environments, and the spore form, which is highly resistant to environmental stressors and can persist both in the environment and in food sources (37). Notably, the spore form is resistant to gastric acid, allowing it to pass through the digestive system intact and reach the intestines (30).

Under anaerobic conditions, such as those in the gut, the spores germinate into vegetative cells and begin to proliferate, producing their main virulence factors—Toxin A (TcdA) and Toxin B (TcdB) (14). The disease caused by this bacterium, *Clostridium difficile* infection (CDI), is classified among healthcare-associated infections and represents a significant public health concern (38).

The symptoms of CDI vary widely, ranging from mild diarrhea to life-threatening colitis, resulting from disruption of intestinal epithelial integrity, tissue damage, and inflammation (39). Individuals at higher risk include the elderly, those

with underlying medical conditions, long-term hospitalized patients, and individuals using broad-spectrum antibiotics or proton pump inhibitors (40).

Although CDI was previously considered primarily a healthcare-associated infection, recent years have seen a marked increase in community-acquired cases. In both Europe and the United States, CDI affects hundreds of thousands of people annually and is responsible for tens of thousands of deaths, making it a major public health threat (35). Moreover, the economic burden it places on healthcare systems is substantial.

Management of CDI is particularly challenging due to the pathogen's high transmissibility, frequent recurrence despite treatment, and increasing resistance to antibiotics (38). These challenges have led to a growing focus on therapeutic strategies targeting the bacterial toxins and the development of alternative treatment approaches (41).

1.3. Toxin B

TcdB is one of the two principal exotoxins produced by *Clostridioides difficile* and plays a key role in its pathogenicity. It is often significantly more cytotoxic than TcdA (28). This large, single-chain protein, with a molecular mass of approximately 270 kDa, is a member of the large clostridial toxin (LCT) family (28). TcdB comprises four major functional domains: the N-terminal glucosyltransferase domain (GTD), the autoprotease domain (APD), the delivery and receptor-binding domain (DRBD), and the C-terminal combined repetitive oligopeptides (CROPs) (14) (Figures 1 and 2).

The toxic mechanism of TcdB begins with its binding to specific receptors on the surface of host cells. The receptor-binding domain (RBD) enables interactions with particular membrane proteins and carbohydrates, facilitating the internalization of the toxin via endocytosis (42) (Figure 1). Once inside the endosome, the acidic environment triggers conformational changes that allow the translocation domain to insert into the endosomal membrane. This facilitates the transfer of the GTD into the host cell cytosol (37).

Within the cytoplasm, the cysteine protease domain (CPD) is activated by intracellular molecules such as inositol hexakisphosphate (InsP6), leading to cleavage of the holotoxin and release of the GTD (42). The GTD then glycosylates specific members of the Rho family of small GTPases—including Rho, Rac, and Cdc42—by transferring a glucose moiety from UDP-glucose to conserved threonine residues on these proteins (42, 43) (Figure 1). This modification disrupts the normal GDP–GTP exchange cycle, effectively locking the GTPases in their inactive forms (44).

Inactivation of these GTPases interferes with cytoskeletal dynamics (39). It leads to the collapse of actin filaments, disassembly of stress fibers, changes in cell morphology, and weakening of cell-cell junctions, all of which compromise epithelial barrier integrity (29, 39). In intestinal epithelial cells, these effects result in disruption of tight junctions and increased permeability (16). Additionally, TcdB promotes the release of pro-inflammatory cytokines and initiates cell death pathways such as apoptosis and necrosis, thereby exacerbating tissue damage (29, 39). Through these mechanisms, TcdB-mediated inactivation of GDP-bound Rho GTPases is a crucial contributor to the inflammation and clinical manifestations observed in *C. difficile* infection (5).

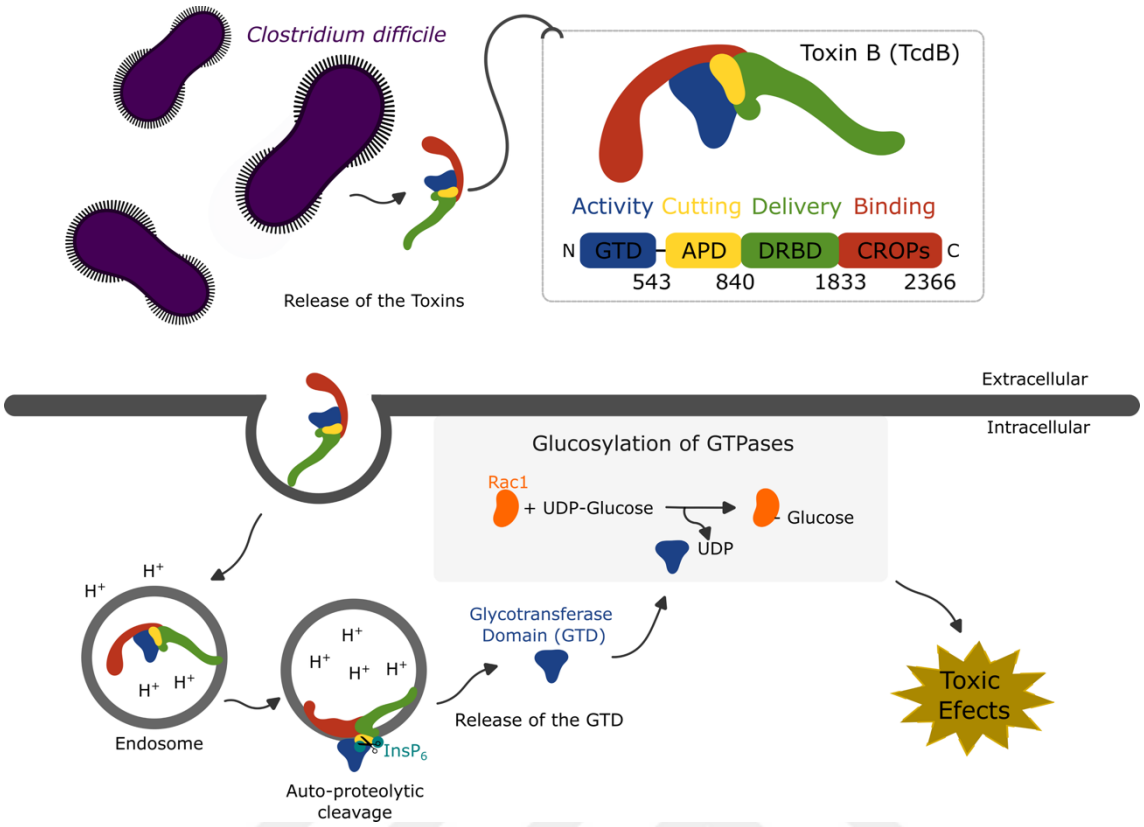


Figure 1. Mechanism of action of Clostridium difficile Toxin B (TcdB). TcdB is secreted by C. difficile and enters host cells through receptor-mediated endocytosis. Within the endosome, acidification triggers auto-proteolytic cleavage via the yellow Auto-protease Domain (APD), releasing the N-terminal Glucosyltransferase Domain (GTD, blue) into the cytosol. The GTD, which contains the conserved DXD motif, specifically glycosylates small Rho GTPases such as Rac1 using UDP-glucose. This modification inactivates Rac1, disrupts actin cytoskeleton dynamics, and induces cytopathic effects including cell rounding and death. The green Delivery and Receptor-Binding Domain (DRBD) mediates translocation, while the red C-terminal CROPs domain ensures host receptor recognition.

1.4. Glycosylation of Rho/Ras GTPases and the Cellular Effects of TcdB

TcdB exerts its intracellular action by glycosylating target proteins from the Rho and Ras GTPase families using UDP-glucose as a donor once it has entered the host cytosol (45). This modification results in the inactivation of these small GTPases, thereby disrupting key intracellular signaling pathways (29). Primary targets include several Rho GTPases such as RhoA, RhoB, RhoC, Rac1, Rac2, and

Cdc42, while secondary targets extend to Ras family members including Ral, Ras, Rap1, and Rap2 (42).

Rho GTPases are central regulators of actin cytoskeleton dynamics, whereas Ras GTPases participate in cellular processes such as proliferation, differentiation, angiogenesis, and adhesion. These proteins function as molecular switches that toggle between GDP-bound (inactive) and GTP-bound (active) states. TcdB glycosylates a conserved threonine residue in the switch I region of these GTPases—Thr35 in Rac1 and Cdc42, Thr37 in RhoA, and Thr61 in R-Ras (44). This threonine is essential for Mg^{2+} coordination and GTP binding, making it critical for their activation mechanism (15). Glycosylation at this site prevents the necessary conformational shifts for effector binding, effectively silencing downstream signaling (46).

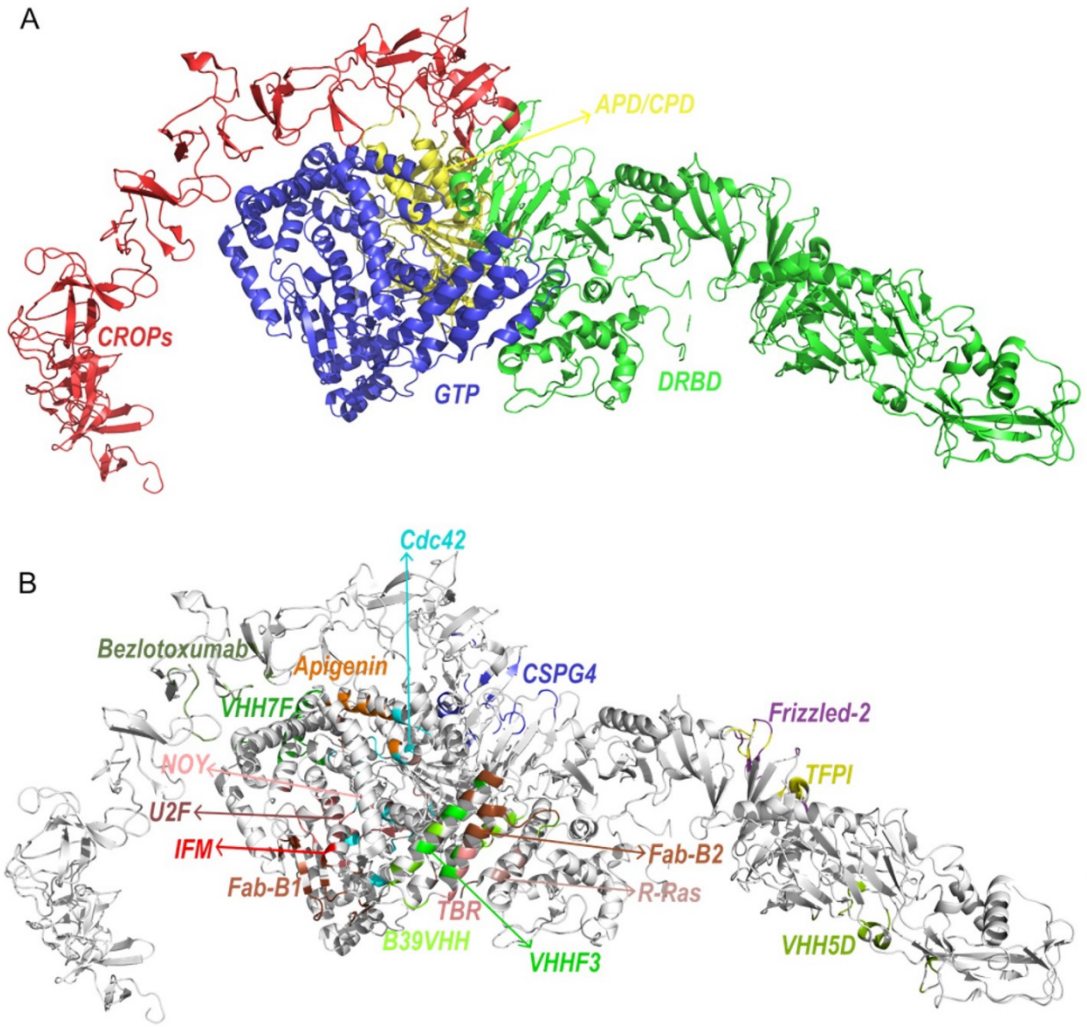


Figure 2: Structural domains and known partner representation of the full-length Toxin B protein structure. (A) TcdB comprises four main domains: the N-terminal glucosyltransferase domain (GTD; residues 1–543), the cysteine protease domain (CPD; residues 544–840), the central delivery and receptor-binding domain (DRBD; residues 841–1833), and the C-terminal combined repetitive oligopeptides (CROPs) region (residues 1834–2366), which facilitates receptor interaction. (B) The full-length crystal structure of TcdB (PDB ID: 6OQ5, shown in white) is used as a reference for structural studies involving interactions with host proteins.

Strain-specific variants of TcdB exhibit differing target specificities. For example, TcdB from strains such as UK1 (RT027) and VP10463 modifies RhoA, Rac1, and Cdc42 but not R-Ras (45). In contrast, variants from strains like 8864, NAP1V, VPI1470, and M68 (RT017) glycosylate R-Ras, Rac1, and Cdc42 while sparing

RhoA (42). These target preferences may influence disease progression and virulence in animal models (42).

Glycosylation of Rho and Ras GTPases leads to extensive reorganization of the actin cytoskeleton, resulting in dramatic changes in cell morphology (47). Cells exhibit loss of stress fibbers, rounding (cytopathic effect, CPE), and shrinkage with abnormal membrane structures (42). RhoA inactivation is a key driver of CPE; however, more recent studies highlight Rac1 glycosylation as equally critical (47). Cells expressing glycosylation-resistant Rac1 are notably protected from TcdB-induced cytotoxicity (47). Rho GTPase inactivation also interferes with the cell cycle. Loss of RhoA disrupts contractile ring formation during cytokinesis, resulting in binucleated cells (15). Inactivation of Rac1 delays mitotic entry by suppressing CDK1/cyclin B and Aurora A kinase activation (42). Both TcdA and TcdB reduce cyclin D1 expression, causing G1/S phase arrest (42). TcdA further activates p53 and p21, halting the cycle at the G2/M transition (5).

Disruption of Rho signalling impairs epithelial barrier function by affecting tight junctions (TJs), which are composed of proteins like occludin and claudin that associate with F-actin via ZO proteins (42). TcdA and TcdB destabilize actin filaments, compromising TJ integrity and increasing paracellular permeability (44). In human intestinal organoid (HIO) models, TcdA causes more pronounced epithelial barrier disruption than TcdB (37). Normally, epithelial polarity and tight junctions prevent bacterial adhesion, but TcdA weakens this defence, facilitating bacterial colonization (37). Restoration of the intestinal epithelium is vital for resolving CDI and preventing relapse (42). However, TcdA and TcdB impair this process by inhibiting the Wnt/β-catenin signalling pathway, which is essential for epithelial regeneration (42). TcdB binds to FZD-7 receptors, blocking Wnt3a signalling and stem cell function (42). Epidemic RT027 strains of TcdB impair stem cell activity even without FZD binding (48,49). TcdA disrupts β-catenin nuclear translocation via Rac1 inactivation, reducing expression of genes involved in proliferation (42). Both toxins also interfere with the Hippo signalling pathway

by promoting degradation of YAP and TAZ, key effectors in cell proliferation and stem cell maintenance (42,43).

Tissue damage during CDI can extend beyond the superficial epithelium to deeper layers, including enteric glial cells (EGCs), which are critical for gastrointestinal function (42). TcdB induces cytopathic effects and cellular senescence in EGCs (50). These senescent cells exhibit persistent cell cycle arrest, DNA damage, and altered gene expression, including upregulation of p27 and inactivation of cyclin B1/CDK1 complexes (42). Dysfunctional EGCs contribute to both impaired gut physiology and pathogenesis (42).

Additionally, TcdA and TcdB downregulate the expression of apical ion transporters such as SLC9A3 (NHE3) and SLC26A3 (DRA), which are essential for fluid and electrolyte absorption (51,52). This disruption leads to osmotic diarrhea. The reduced expression may result from impaired cytoskeletal transport of these proteins or enhanced degradation via the ubiquitin-proteasome pathway (51). TcdB induces two forms of cell death depending on its concentration. At lower levels, it triggers glycosylation-dependent apoptosis (53). At concentrations above 100 pM, glycosylation-independent necrosis occurs (54). Necrosis is characterized by mitochondrial swelling, loss of plasma membrane integrity, and leakage of cellular contents (51). During this process, TcdB stimulates calcium release, activating protein kinase C (PKC), which in turn activates NADPH oxidase complexes in endosomes, leading to reactive oxygen species (ROS) production. Elevated ROS levels cause ATP depletion, LDH release, reduced caspase-3/7 activity, and chromatin condensation, ultimately resulting in necrotic cell death. Unlike TcdB, TcdA does not typically induce lethal cell death and instead causes mucosal damage through distinct, less ROS-dependent mechanisms (51,52).

1.5. Therapeutic Perspectives on TcdB-Targeted Mechanisms in CDI

In summary, the glycosylation activity of TcdB on Rho and Ras GTPases leads to extensive cellular disruption, affecting structures and signalling pathways that govern cell morphology, cell cycle progression, epithelial barrier integrity, and cell death (4,14). Although TcdA contributes to these pathogenic processes, the molecular specificity and cytotoxic mechanisms of TcdB are particularly critical in determining disease severity and progression (29,53).

Despite antibiotics remaining the mainstay of *Clostridium difficile* infection (CDI) treatment, their efficacy is often limited by recurrent infections and toxin-mediated tissue damage (42). This has prompted the development of novel therapeutic strategies aimed at neutralizing the toxins directly (5). Current approaches include antitoxin antibodies, receptor blockers, and toxin-binding agents (42). Notably, compounds that inhibit the intracellular activity of TcdB or modulate the host inflammatory response are under clinical investigation (5). Additionally, interventions such as fecal microbiota transplantation (FMT) show promise in reducing recurrence by restoring gut microbial balance (55).

Advancing our understanding of TcdB's molecular interactions selectively with GDP-bound inactive Rho/Ras GTPases is vital not only for elucidating the pathophysiology of CDI but also for designing targeted and more effective therapeutic strategies (56). Among these strategies, *in silico* approaches such as discovery of small-molecule inhibitors and therapeutic peptide design are increasingly significant in identifying novel anti-toxin agents.

1.6. *In Silico* Drug Discovery Approaches Targeting TcdB

Targeting TcdB—a major virulence factor of *C. difficile*—has become a key focus in in silico drug discovery strategies for combating *C. difficile* infection (CDI). Structural studies have identified druggable regions within TcdB, particularly the glucosyltransferase domain and receptor-binding interfaces that interact with host proteins such as Frizzled receptors and CSPG4 (42). These regions serve as

promising targets for both small molecule inhibitors and peptide-based therapeutics.

Recent structural investigations have provided detailed insights into the conformational dynamics and host interaction mechanisms of TcdB. Notably, Chen et al. (2019) resolved the full-length structure of TcdB, revealing critical architectural features of its functional domains (57). Similarly, Liu et al. (2021) elucidated the structural basis of TcdB's selective modification of host Rho and Ras GTPases, offering a molecular framework for inhibitor design (58).

In this thesis, both rationally designed therapeutic peptides and virtually screened small molecules targeting TcdB were evaluated using molecular docking, molecular dynamics (MD) simulations, and free energy calculations. These computational analyses provide valuable insights into the binding potential and dynamics of candidate therapeutics, contributing to the development of novel, mechanism-specific treatment options for CDI.

2. MATERIALS AND METHODS

2.1. Structural Preparation and Characterization with Docking, Molecular Dynamics, and Advanced Computational Techniques for Host GTPases-bacterial toxin TcdB

2.1.1. Modelling of Host Protein–TcdB Interactions

This section describes the structural analysis of TcdB and its interactions with host proteins, inhibitors, and antibodies (Table S2). All available structures of TcdB were obtained from the Protein Data Bank (PDB) (17) and AlphaFold (59)databases as of January 2023, using UniProt (60) accession numbers P18177 and Q9EXR0, corresponding to the TCDB_CLODI and TCDB2_CLODI genes, respectively. In total, 30 structures were analysed, comprising 28 experimentally determined and 2 predicted models (Table A1). Among these, 21 structures contain complexes with various ligands, including host proteins, small molecules, or antibodies. Binding interfaces within these complexes were identified via the PDBsum (61) server and visualized using PyMOL (62), with reference to the full-length crystal structure of TcdB (PDB ID: 6oq5) (Figure 2B).

To explore potential host protein interactions, the PHISTO database (63) was queried in September 2022, identifying three human proteins reported to interact with TcdB, namely Nectin-3, Rac1, and CD46. Due to the lack of experimentally resolved structural complexes for these interactions, docking-based modelling was employed. Emphasis was placed on the presence and role of metal ions—Mg^{2+} in GTPases and Mn^{2+} in TcdB—as these ions are critical for maintaining structural stability. Missing Mg^{2+} ions, such as in the GDP-bound Rac1 structure (PDB ID: 5o33), were modelled by structural alignment with homologous GTPases containing the ion (e.g., PDB ID: 7s0y). Docking simulations were conducted using the HADDOCK web server (64), with input files prepared through the PDB-tools suite. Interface residues were defined based on information from known protein complexes, and chain A of the TcdB–Cdc42 complex (PDB ID:

7s0y) was utilized as a modelling template. The resulting models were further validated by comparison of interface regions with predictions from AlphaFold-Multimer v2. Visualization and figure preparation were performed using PyMOL (62), ChimeraX (62,65), and Inkscape software's (66).

2.1.2. MD Simulations of Modelled Host GTPase–TcdB Complexes

To evaluate the structural stability and dynamic behaviour of the modelled host GTPase–TcdB complexes, molecular dynamics (MD) simulations were performed using the GROMACS software package (19). Initial structures were derived from docking models and subsequently prepared for simulation through the CHARMM-GUI server (67). The systems were solvated with the TIP3P water model, neutralized, and supplemented with 0.15 M NaCl to mimic physiological ionic conditions. The CHARMM36m force field (20) was employed throughout the simulations.

Following energy minimization, equilibration was carried out in two phases: an NVT ensemble (constant volume and temperature) followed by an NPT ensemble (constant pressure and temperature). Production simulations were then conducted at 310 K and 1 atm. Temperature and pressure were controlled using the Nose–Hoover thermostat and Parrinello–Rahman barostat, respectively. For each of the four complexes—Rac1-GDP/TcdB, Rac1-GTP/TcdB, Cdc42-GDP/TcdB, and Cdc42-GTP/TcdB—three independent molecular dynamics production runs of 500 nanoseconds were conducted, yielding a total simulation time of 6 microseconds. Among these, the Cdc42-GDP/TcdB and Cdc42-GTP/TcdB complexes, for which high-resolution experimental structures are available in the Protein Data Bank (PDB), were used as structural templates and references to model the corresponding Rac1 complexes.

Simulation analyses began with assessing global and local structural deviations via backbone root-mean-square deviation (RMSD) and per-residue root-mean

square fluctuation (RMSF) calculations. The following GROMACS (19) commands were utilized for processing trajectory and analysis:

gmx make_ndx -f system.tpr -o index.ndx

gmx trjconv -s system.tpr -f trajectory.xtc -o trajectory_nojump.xtc -pbc nojump

gmx trjconv -s system.tpr -f trajectory_nojump.xtc -o trajectory_noPBC.xtc -pbc mol -center -n index.ndx

gmx rms -s system.tpr -f trajectory_noPBC.xtc -o rmsd.xvg -tu ns -n index.ndx

gmx rmsf -s system.tpr -f trajectory_noPBC.xtc -o rmsf.xvg -n index.ndx

To monitor the compactness of the complexes, the radius of gyration (Rg) was calculated. The positional stability of cofactors, such as Mg^{2+} ions, was tracked by measuring distances between these ions and GTP/GDP ligands using the command:

gmx distance -s system.tpr -f trajectory_noPBC.xtc -select 'com of group "MG" plus com of group "GTP"' -oall mg_gtp_dist.xvg

Binding free energies were computed employing the gmx_MMPBSA tool (19) within a Conda environment. In these calculations, the TcdB protein together with the Mn^{2+} ion and any bound UDP/GDP ligands were defined as groups, while the host GTPase (RAC1 or CDC42) with Mg^{2+} and GTP/GDP ligands constituted the second group. An example command used for the energy calculations is as follows:

mpirun -np 5 gmx_MMPBSA -O -i mmpbsa.in -cs system.tpr -ci index.ndx -cg 23 22 -ct trajectory_noPBC.xtc -cp topol.top -o FINAL_RESULTS_MMPBSA.dat -eo FINAL_RESULTS_MMPBSA.csv

Additionally, snapshots extracted from the first and last 40 ns intervals of each simulation were analysed via Sidechain Decomposition Contribution (SDC) mode to determine residue-level binding energy components. These energetic

contributions were mapped onto the protein structures as B-factors to visualize key residues involved in binding.

All simulation data were exported in. xvg format and subsequently plotted using GraphPad Prism software(68) . Time-dependent analyses of RMSD, RMSF, and ion-ligand distances were presented as graphs, with appropriate colour coding and error bars representing standard error to facilitate comparative evaluation.

2.1.4. Enhanced Sampling Analysis of Rac1–TcdB Complexes

To gain detailed mechanistic insights into how TcdB influences GDP-bound Rac1, enhanced sampling approaches were employed, specifically well-tempered Metadynamics (WT-MetaD) and umbrella sampling. All simulations were performed using GROMACS 2021.4 in combination with the PLUMED 2.8 plugin (21,22).

In the Metadynamics simulations, two collective variables (CVs) were defined to capture essential structural transitions: (1) the centre-of-mass (COM) distance between Rac1 and TcdB, and (2) the radius of gyration (Rg) of the entire complex. Simulations were conducted under physiological conditions (310 K and 1 atm). The GDP-bound Rac1–TcdB complex was simulated for 100 ns, while the GTP-bound form was simulated for 62 ns. During these simulations, Gaussian hills with a height of 0.5 kcal/mol were added every 2 ps to accelerate exploration of the conformational landscape.

Post-simulation analysis included calculation of solvent-accessible surface area (SASA), measurement of COM distances, and quantification of intermolecular hydrogen bonds. These metrics provided molecular-level understanding of Rac1–TcdB binding dynamics under different nucleotide-bound states.

To estimate the binding free energy (ΔG), umbrella sampling simulations were performed for Rac1 complexes bound to either GDP or GTP. The reaction coordinate was defined as the linear distance between Rac1 and its respective ligand. This coordinate was divided into 20 windows; each spaced at 0.2 nm intervals. In each window, simulations were conducted with a harmonic restraint of 1000

kJ/mol·nm² applied along the x-axis, which ensured controlled sampling along the defined pathway. Each window underwent 1 ns equilibration followed by a 5 ns production run.

After all simulations, the results from each window were integrated using the Weighted Histogram Analysis Method (WHAM) to generate potential of mean force (PMF) curves. The binding free energy was derived as the energy difference between the highest and lowest points of each PMF profile, reflecting the strength of ligand binding.

2.2. Therapeutic Small Molecule and Peptide Inhibitor Design Approach

In this study, an in-silico strategy was employed to design potential inhibitors specifically targeting the interaction between *Clostridium difficile* toxin B (TcdB) and the host GTPase protein Rac1. Both small molecule and peptide-based inhibitors were designed and evaluated using computational approaches including virtual screening, molecular docking, and peptide modelling.

Rac1 was selected as the primary target based on molecular dynamics (MD) and Metadynamics simulation results, which indicated a stronger and more stable interaction between TcdB and the GDP-bound form of Rac1. This interaction was associated with conformational changes that may lead to cytoskeletal disruption and downstream apoptotic or necrotic effects. Although Cdc42 is also targeted by TcdB, it was used in this study as a structural reference due to its high sequence and structural similarity to Rac1, and the availability of experimentally resolved TcdB–Cdc42 complexes in the Protein Data Bank (PDB). These Cdc42-based complexes served as templates for building the Rac1–TcdB models used throughout the inhibitor design pipeline.

2.2.1. Discovery of Small-Molecule Inhibitors

2.2.1.1. Virtual Screening

a. Preparation of the Molecular Library

The initial molecular library for virtual screening was constructed using the ZINC database (21). Within the database's "Tranche Browser" interface, molecules were filtered based on the following parameters: 3D representation, standard reactivity, pH reference set to neutral (mid), and all available charge states were included. The molecular weight was restricted to the range of 200–500 Daltons. The LogP parameter was inclusive of all values ≥ –1, including those exceeding 5. This selection yielded a comprehensive dataset of approximately 14 million compounds distributed across 8,200 tranches, which were collectively downloaded in the SMILES (.smi) format.

b. Construction and Cleaning of the Molecular Library

To consolidate the molecular data, individual .smi files located in multiple subdirectories were merged into a single file via shell commands. Specifically, the concatenation was performed using the command:
cat /home/damla/sanaltarama/.smi > merged.smi*

Subsequently, a Python script was employed to remove any empty or malformed lines and extract only valid SMILES strings. The script iterated through each line of the merged file, writing cleaned SMILES entries to a new file (cleaned_merged.smi). For computational efficiency during virtual screening, this large file was partitioned into 40 smaller subsets. Each subset was processed using a Python script (sanaltarama.py) designed to automate the screening workflow.

c. Virtual Screening Procedure

A similarity-based virtual screening approach was implemented, with target compounds defined in SMILES notation. Reference ligands (existing TcdB inhibitors) were sourced from the PubChem database, including Apigenin, NOY (Noeuromycin), Vancomycin, and Fidaxomicin, with their canonical SMILES strings used as queries. The cleaned ZINC library was imported into a local

SQLite database (molecules.db) using Python's sqlite3 module to facilitate rapid querying and filtering. A table named library was created, storing unique identifiers alongside SMILES representations of each molecule. This approach enabled efficient management of the large dataset during screening.

d. Drug-Likeness Filtering via Lipinski's Rule

To ensure the identification of compounds with favourable drug-like properties, Lipinski's Rule of Five was employed using the RDKit cheminformatics library. Key molecular descriptors such as molecular weight (MW), hydrogen bond donors (HBD), hydrogen bond acceptors (HBA), the octanol–water partition coefficient (logP), and topological polar surface area (TPSA) were calculated. Compounds meeting the established criteria—MW ≤ 500, HBD ≤ 5, HBA ≤ 10, logP ≤ 5, and TPSA ≤ 140 $Å^2$—were considered to have acceptable drug-likeness and were retained in a separate ADME screening dataset.

e. ADME-Tox Profiling and Filter Application

Further pharmacokinetic profiling involved in silico prediction of blood-brain barrier permeability (BBB) and gastrointestinal absorption (GIA) using models based on SwissADME algorithms implemented in Python. Compounds meeting these criteria were catalogued within the PB_BBB_GIA table.
To ensure chemical specificity and reduce false positives, PAINS (Pan-Assay Interference Compounds) and Brenk toxicity filters were applied, removing molecules with potentially promiscuous or toxic substructures.

f. Molecular Fingerprinting and Similarity Scoring

Molecular fingerprints were generated using the Morgan fingerprint algorithm (Extended Connectivity Fingerprints, ECFP) with a radius of 2- and 1024-bits length to numerically encode chemical structures (69). Similarity calculations between the reference molecules and library compounds were performed using

the Tanimoto coefficient via RDKit's data structures. Compounds exhibiting a Tanimoto similarity score greater than 0.44 and favourable BBB or GIA properties were shortlisted for further evaluation.

```
from rdkit import Chem
from rdkit.Chem import AllChem, DataStructs
fp1 = AllChem.GetMorganFingerprintAsBitVect(mol1, 2, nBits=1024)
fp2 = AllChem.GetMorganFingerprintAsBitVect(mol2, 2, nBits=1024)
similarity = DataStructs.TanimotoSimilarity(fp1, fp2)
```

g. Automation of Virtual Screening Workflow

All processes were fully automated through Python scripting, enabling scalable, high-throughput identification of structurally similar and pharmacokinetically promising compounds. This workflow provides a robust and efficient strategy for large-scale virtual screening campaigns targeting the selected biological system.

2.2.1.2. Molecular Docking

a. Protein Preparation and Active Site Definition

The protein structure of TcdB used in this study was not directly retrieved from the Protein Data Bank (PDB) but was instead derived from a previously equilibrated molecular dynamics (MD) simulation model based on the crystal structure with PDB ID: 7S0Y. A representative and biologically relevant conformation of the protein was selected from the MD trajectory.

To obtain a clean receptor structure suitable for docking, all non-protein entities such as water molecules, metal ions, and co-crystallized ligands were removed. Missing residues were modelled and completed using the reference crystal structure via tLeap (AmberTools). Subsequently, hydrogen atoms were added, and

protonation states were assigned using the H++ web server (http://biophysics.cs.vt.edu/) at physiological pH 7.0, which corresponds to the intracellular environment of human intestinal epithelial cells. This ensured accurate representation of side-chain charge states and hydrogen bonding networks under biologically relevant conditions.

The target site for docking was defined based on previous literature and detailed structural analysis. Emphasis was placed on the Rac1 interaction interface and the flexible loop regions surrounding the manganese ion coordination site. This region was also chosen because it is located near the UDP-binding domain, which is considered functionally important for TcdB's pathogenic mechanism. Importantly, the same binding site was used for both small molecule inhibitors and peptide ligands, enabling a consistent comparison of binding affinities across compound classes.

A cubic docking grid box was defined to fully encompass this binding region, centred at coordinates (x = 34.000, y = 14.000, z = 40.000) with dimensions of 20 Å × 20 Å × 20 Å. This grid configuration was carefully selected to include all critical residues involved in ligand recognition and binding.

Molecular docking simulations were performed using AutoDock Vina, which enables flexible ligand docking by employing an efficient scoring function and stochastic global optimization algorithm. This allowed for the identification and ranking of potential binding poses within the predefined target site.

b. Flexible Docking Configuration

To better account for protein conformational flexibility during ligand binding, selected residues within the active site, particularly loop segments adjacent to the manganese-binding region critical for TcdB–Rac1 interaction, were treated as flexible. This allowed the docking protocol to simulate more realistic binding modes by enabling side-chain movements of these key residues.

c. Ligand Preparation

A compound library of small molecules in .sdf format was curated and processed for docking. Each ligand was standardized, with missing hydrogen atoms added using cheminformatics software. The ligands were then converted to the PDBQT format required for AutoDock Vina (70) simulations, during which partial charges and rotatable bonds were assigned. Molecules that failed conversion or contained structural errors were excluded from subsequent docking.

d. Docking Simulations

Docking was conducted using AutoDock Vina, targeting the predefined active site grid box on TcdB. The default exhaustiveness parameter was applied, and flexible docking protocol were explored. For each ligand, the binding affinity (expressed in kcal/mol) of the highest (the most negative) scoring pose was recorded. Complementary docking runs were also performed via the SwissDock web server (71), which employs a variant of the AutoDock Vina algorithm, and these results were integrated into the overall analysis.

e. Filtering and Prioritization

Post-docking, ligands were ranked by their predicted binding affinities, and the top 100 compounds with the lowest binding energies were shortlisted. These candidates were further filtered using the following criteria: binding affinity ≤ –8.0 kcal/mol, and target specificity confirmed by lack of predicted affinity towards RAC1. The 4 highest-ranked ligands passing these filters underwent additional evaluation using ADMETlab 2.0 and pkCSM platforms to predict their pharmacokinetic properties and toxicity profiles.

The post-docking filtering criteria included a binding affinity threshold of ≤ –8.0 kcal/mol, which was selected based on literature precedents describing it as an indicator of strong binding affinity in protein–ligand interactions. While the exact numerical value of docking scores can vary depending on the scoring function

and docking protocol used, binding energies below –8.0 kcal/mol are commonly associated with high-affinity interactions and are thus used as an empirical cutoff in virtual screening studies. This threshold served as a preliminary filter and was not the sole criterion for ligand prioritization. Candidates marginally above this threshold were not discarded outright; instead, additional parameters such as target selectivity (lack of affinity towards RAC1), ADME/Tox properties, and structural compatibility were also considered to support decision-making. Therefore, the filtering approach aimed to balance binding strength with pharmacological relevance and target specificity.

2.2.1. Peptide Design Process

In this study, cyclic peptides with high specificity for a defined region of the *Clostridium difficile* TcdB toxin were designed. The focus was placed on the Asp270 (D270) residue within the glucosyltransferase domain of TcdB, which has been demonstrated as functionally critical based on previous molecular dynamics simulations, ligand interaction studies, and structural analyses.

2.2.2.1. Target Binding Site Identification

The binding pocket surrounding the D270 residue in TcdB was analysed using PyMOL software. Amino acids within a 5 Å radius of D270 were identified using the following commands:

pymol
select target_residue, resi 270
select within_5A, (br. all within 5 of target_residue)
show sticks, within_5A
color cyan, within_5A

"L L A A A S I L R I N I T I S G P Y" This sequence represents the unique amino acid residues located within 5 Å of Asp270 in Toxin B protein, listed in one-letter

codes. This selection revealed **eight key residues** critical to interaction with candidate peptides, and this microenvironment was defined as the target site for subsequent peptide docking and optimization.

2.2.2.2. Initial Peptide Sequence Selection and Optimization

The peptide design process in this study was guided by structural insights obtained from the TcdB target site, which was previously used in small molecule docking (see Section 2.2.1.2). Specifically, the binding pocket located near the UDP-binding domain—encompassing the manganese coordination site and residues involved in Rac1 interaction—was selected as the docking target for peptide ligands as well. This region includes the functionally important Asp270residue, around which a local 5 Å radius sequence environment was extracted using PyMOL. The initial sequence used for peptide design was derived directly from this local environment of TcdB, thus ensuring structural relevance and site specificity.

Based on this region, an initial 10-amino acid linear peptide sequence (ADRINITSGY) was selected due to its proximity to the binding pocket and potential compatibility. However, considering the spatial constraints of the cavity and the requirement for conformational stability in a cyclic peptide design, this sequence was optimized to a shorter 8-mer, ADRITPSG. This size was chosen as a balance

between structural compactness for cavity accommodation and sufficient length for stable head-to-tail cyclization.

The rationale for each amino acid in the optimized sequence is as follows:

- D (Aspartic acid): Negatively charged, enables electrostatic interactions with basic residues within the binding pocket.
- R (Arginine): Positively charged, capable of forming hydrogen bonds and salt bridges with nearby polar residues.
- T (Threonine) and S (Serine): Contain polar side chains that enhance hydrogen bonding capacity.
- P (Proline): Introduces conformational rigidity, favourable for cyclic peptide stability.
- G (Glycine): Provides flexibility, aiding steric accommodation in confined regions.

a. Cyclization of the Peptide

The final 8-mer peptide was cyclized via head-to-tail linkage, a well-established approach for enhancing protease resistance, conformational stability, and binding affinity. Although side-chain cyclization strategies (e.g., disulfide bridges or stapling) were considered to further constrain the structure, these were not implemented due to synthetic complexity and focus on sequence-driven variation.

b. Generation of Peptide Variants

To investigate the influence of physicochemical properties—namely electrostatic charge, hydrophobicity, and hydrogen-bonding potential—a series of substitution rules was applied to selected positions in the peptide. The initial sequence did not include lysine (K), but it was introduced during this stage specifi-

cally within charge-altering substitutions to comprehensively explore basic–acidic residue interplay (e.g., D ↔ K, R), rather than being part of the original template. The substitution rules were:

- Charge-altering substitutions:
  D ↔ K, R;
  E ↔ K, R;
  K ↔ D, E;
  R ↔ D, E
- Hydrophobic substitutions:
  I, L, V ↔ F
- Hydrogen bonding optimization:
  S ↔ N;
  T ↔ N, Q

A custom Python script using the itertools module was developed to systematically generate all possible permutations and combinations of these substitutions. This resulted in a total of 120,960 alternative peptide sequences, which were stored in the following files:

- hdock_peptides.fasta (FASTA format)
- output_sequences.txt and output_sequences_3.txt (for docking input into HDOCK)

All designed peptides were subsequently subjected to docking analysis targeting the same predefined region of TcdB to enable consistent comparison with small molecule inhibitors.

c. Structural Modelling Using Rosetta

All cyclic peptide variants (n = 120,960) were structurally modelled using the Rosetta Cyclic Peptide Modelling tool, employing the SimpleCycpepPredict protocol. This protocol generates low-energy 3D conformations of cyclic peptides, focusing on backbone closure and realistic torsion angle sampling.

Following modelling, each structure underwent quality control to ensure suitability for downstream applications. This included:

- Backbone continuity and successful head-to-tail cyclization.
- Absence of steric clashes or unrealistic bond lengths/angles.
- Energetic favourability, based on Rosetta scoring.
  Only structures that passed all criteria were retained for further analysis.

d. Toxicity Screening and Local Docking

To prioritize non-toxic candidates, ToxinPred2, a machine learning-based prediction platform, was used to evaluate peptide toxicity. Sequences were scored according to a trained classification model, and all peptides with toxicity scores below the safety threshold ($p < 0.05$) were excluded. This initial filtering dramatically reduced the candidate pool from 120,960 to 15 non-toxic cyclic peptides.

These 15 peptides were then subjected to local docking analysis using the Linux version of the HDOCK platform, targeting the same TcdB binding site used for small molecule docking. Docking was executed via the following command-line instruction:

bash

```
./hdock_linux -rec tcdB_clean.pdb -lig peptide_model.pdb -out result_output/
```

Docking outputs were evaluated based on:

- Binding affinity scores (from HDOCK's hybrid scoring function),
- Interaction profiles with the D270-centered binding site, including hydrogen bonding and electrostatic contacts.

One of the 15 peptides failed to produce a viable docking pose due to conformational incompatibility with the binding pocket, leaving 14 structurally and functionally compatible candidates.

e. Off-Target Risk Assessment

To minimize potential off-target interactions, off-target similarity was evaluated using BLASTp against the non-redundant (nr) human protein database. Due to their cyclic nature, each of the 14 peptide sequences was converted into eight linearized rotational isomers—each starting from a different amino acid position—to simulate all possible binding motifs. This approach generated approximately 120 linear sequences (14 peptides × 8 isomers + minor variation due to structural failure in one case).

BLASTp was configured for short peptide alignment, with the following parameters:

- E-value threshold: 1000 (high sensitivity),
- Substitution matrix: PAM30,
- Word size: 2,
- Gap penalties: Existence 9, Extension 1,
- Low complexity filtering and lookup table masking enabled.

BLASTp results were filtered by:

- ≥80% identity for general similarity,
- and query coverage thresholds to exclude low-confidence hits.

This process revealed that 12 out of 14 peptide candidates shared partial or high similarity with human protein sequences. Only two peptides demonstrated no significant homology (E-value < 0.01) with any known human protein, thus minimizing off-target interaction risk. These two final peptide candidates were selected for subsequent molecular dynamics simulations and functional validation.

f. Final Peptide Candidates

After comprehensive filtering based on binding affinity, toxicity, and off-target similarity, two peptide candidates were selected for further investigation. These candidates displayed:

- High binding affinity to the TcdB Asp270 site
- Non-toxic profiles
- No significant similarity to human proteins (E-value ≥ 0.01), minimizing off-target risks

These final candidates will undergo advanced molecular dynamics simulations and binding free energy calculations to fully characterize their interaction mechanisms and evaluate their therapeutic potential.

2.3. Molecular Model Preparation and Ligand Topology Generation

2.3.1 Addition of Hydrogen Atoms and Ligand Preparation

Accurate molecular dynamics (MD) simulations require complete and chemically realistic molecular structures. Ligand structures obtained from docking or experimental sources often lack explicit hydrogen atoms necessary for correct representation of hydrogen bonding and electrostatics. In this study, the selected ligand structure was first optimized and completed using Avogadro (v1.95) (72,74), where explicit hydrogen atoms were added automatically based on valency rules. This was followed by a brief geometry optimization using the MMFF94 force field to refine 3D structure.

a. 3D Structure Optimization of the Ligand

The geometry-optimized structure from Avogadro was further minimized using Open Babel with the following command to ensure conformational stability:

obabel input_ligand.pdb -O optimized_ligand.pdb --minimize --steps 1000

b. Conversion to Mol2 Format and Charge Assignment

To prepare the ligand for Amber simulations, the PDB file was converted to Mol2 format using Antechamber (75). Atomic partial charges were assigned using the AM1-BCC method:

antechamber -i optimized_ligand.pdb -fi pdb -o ligand_op_H.mol2 -fo mol2 -c bcc -s 2

The -c bcc flag specifies the AM1-BCC charge calculation, while -s 2 enables detailed error checking and optimization.

c. Identification of Missing Parameters and frcmod File Generation

Certain atom or bond types may lack predefined parameters in the Amber force field. To address this, the parmchk2 utility was used to scan the Mol2 file, identify missing parameters, and generate an additional parameter modification file (ligand.frcmod):

parmchk2 -i ligand_op_H.mol2 -f mol2 -o ligand.frcmod

2.3.2. Topology and Coordinate File Generation

Topology and coordinate files for Amber MD simulations were created using tleap:
tleap.in:
source leaprc.gaff
LIGAND = loadmol2 ligand_op_H.mol2
loadamberparams ligand.frcmod
check LIGAND
saveamberparm LIGAND ligand.prmtop ligand.inpcrd
quit
Execution:
tleap -f tleap.in

2.3.3. Simulation Issue Due to Docking Clashes

Although the ligand preparation steps were completed successfully, several docked poses resulted in steric clashes, particularly involving ligands positioned near the UDP-binding region of TcdB. Initially, the docking grid was centred around the manganese coordination site, which is known to participate in interactions with both Rac1 and the UDP ligand. Although this region was selected based on structural relevance and literature-supported functionality (rather than directly targeting the UDP binding pocket), the ligands—due to their structural similarity to UDP—frequently occupied overlapping spatial positions with the native UDP ligand observed in the crystal structure (PDB: 7S0Y).

These overlaps, which were not sufficiently penalized during rigid-body docking, led to non-physical atomic arrangements (e.g., violation of Van der Waals radii) in some of the docked complexes. During the subsequent equilibration phase of molecular dynamics simulations, these steric clashes caused the systems to become unstable, often leading to force field constraint violations and simulation failure. As a result, production MD simulations could not be successfully initiated, and essential downstream analyses—including RMSD, RMSF, and MM-GBSA—were not applicable for these specific ligand–TcdB complexes.

In future studies, a more refined binding site definition—excluding regions with high likelihood of natural ligand overlap—or the use of flexible docking protocols may help mitigate these issues.

2.4. Molecular Dynamics Simulations of Peptides and Ligands

2.4.1. System Preparation

a. Small-molecule Ligand-TcdB Complexes

Each system was solvated in an octahedral TIP3P water box with at least a 10 Å buffer around the solute. Charge neutralization was performed with counterions, and physiological ionic strength was maintained by adding 0.15 M NaCl using

the addIonsRand command in tleap. The number of Na^+ and Cl^- ions was calculated using the following formula:

source leaprc.protein.ff14SB
source leaprc.water.tip3p
mol = loadpdb complex.pdb
Solvateoct mol TIP3PBOX 10.0
addionsrand mol Na+ 27
addionsrand mol Na+ 27 and Cl- 40
saveamberparm mol complex.prmtop complex.inpcrd
quit

The ff14SB force field was applied to protein and peptide atoms.

b. Peptide-TcdB Complexes

Initial peptide–TcdB complex structures were built using the CHARMM-GUI web server. The output files were directly converted into Amber-compatible formats without structural alterations. Protonation states and hydrogen atoms were adjusted based on pH 7.4 using the H++ server (76) to ensure accurate protonation and electrostatics of ionizable residues.

Molecular dynamics simulations for these peptide–protein complexes were carried out using the Amber 22 software package. Although protein–protein complex simulations in this study were performed using GROMACS, Amber was chosen at this stage for its built-in support of binding free energy calculations (e.g., MM-GBSA, per-residue decomposition) and its ability to efficiently handle non-standard residues such as cyclic peptides.

Furthermore, the switch to Amber provided an opportunity to explore an alternative simulation engine and evaluate differences in force field performance. This approach also facilitated a more streamlined post-processing pipeline by

minimizing format conversions and enabling the direct use of Amber-native tools for trajectory analysis.

2.4.2. Minimization and Equilibration Protocol

Energy minimization and equilibration were conducted in three standard steps:

Minimization: Minimization of water and ions (1-550 residues). Full system minimization without restraints.

pmemd.cuda -O -i min.in -o min.out -p complex.prmtop -c complex.inpcrd -r min.rst -ref complex.inpcrd

Equilibration: NVT heating from 0 K to 310 K and NPT density equilibration at 1 atm

Heating input: *pmemd.cuda -O -i heat.in -o heat.out -p complex.prmtop -c min.rst -r heat.rst -ref min.rst -x heat.nc*

NPT input: *pmemd.cuda -O -i npt.in -o npt.out -p complex.prmtop -c heat.rst -r npt.rst -ref heat.rst -x npt.nc*

Production MD:
Production runs of 500 ns were performed for each replicate (3 replicates per peptide–TcdB system) using Amber24 (pmemd.cuda). This yielded a total of 3 µs cumulative simulation time per complex as (2 peptides × 3 replicates × 500 ns = 3 µs).

pmemd.cuda -O -i md.in -o md.out -p complex.prmtop -c npt.rst -r md.rst -x md.nc -inf mdinfo

2.5. Post-Simulation Analyses

All analyses were conducted using CPPTRAJ and MMPBSA.py tools in AmberTools. The following metrics were computed:

Root Mean Square Deviation (RMSD):

$$RMSF_i = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(r_i(t) - \langle r_i \rangle\right)^2}$$

Equation 1

Root Mean Square Fluctuation (RMSF):

$$RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(r_i(t) - r_i^{ref}\right)^2}$$

Equation 2

MM-GBSA Binding Free Energy (entropy excluded):

$$\Delta G_{\text{bind}} = G_{\text{complex}} - G_{\text{receptor}} - G_{\text{ligand}}$$

$$G = E_{\text{MM}} + G_{\text{polar}}$$

$$E_{\text{MM}} = E_{\text{vdW}} + E_{\text{ele}}$$

Note: Entropy contribution (–TΔS) was excluded due to high computational cost.

Scripts and Reproducibility

All trajectory and free energy analyses were automated using custom Python and shell scripts. Python scripts for RMSD, RMSF, H-bond, MM-GBSA, and system NaCl ion calculations (based on box volume) are available and can be shared upon request.

3. RESULTS AND DISCUSSION

This study investigates the structural and dynamic interactions between *Clostridium difficile* Toxin B (TcdB) and host GTPase proteins Rac1 and Cdc42. An integrative approach was employed, combining structural modelling, molecular dynamics (MD) simulations, advanced sampling techniques (well-tempered metadynamics and umbrella sampling), binding free energy calculations (MM-PBSA), and docking studies involving potential therapeutic small molecules and peptides.

3.1. Structural Analysis and Modelling of Toxin B Interactions

Structural analyses provide critical insights into the specific interactions between TcdB and host proteins, particularly those of the Rho family GTPases. Elucidating these interactions not only advances our understanding of the molecular effects of the toxin on host cell proteins but also facilitates the development of strategies aimed at mitigating TcdB's deleterious effects. The interaction interfaces identified in complexes formed between TcdB and host factors such as Apigenin, Cdc42, Frizzled-2, and CSPG4 serve as a structural reference framework for pathogen-host binding sites and functional domains (Figure 1 and Figure 2B).

Notably, Apigenin interacts with the N-terminal glucosyltransferase domain (GTD) of TcdB, suggesting its potential as an inhibitor of the toxin's enzymatic activity. Interactions with Rho GTPases like Cdc42 and Frizzled receptors reveal TcdB's multifaceted mechanism of action on diverse intracellular targets within human intestinal epithelial cells (48,49,64). Additionally, several neutralizing antibodies developed against TcdB were shown to bind distinct residues predominantly within the GTD, providing a molecular basis for the rational design of therapeutic agents targeting the toxin.

Expanding beyond experimentally validated complexes, this study modelled TcdB in complex with Rho GTPases, including both GTP- and GDP-bound forms

of Cdc42 and Rac1, utilizing HADDOCK and validated by AlphaFold MultimerV2 (64,77). Docking analyses revealed that these GTPases engage TcdB primarily through conserved 'Switch I' and 'Switch II' regions, with nucleotide and Mg^{2+} coordination sites retained within the complexes (Figure 3) (18). The results underscore conserved binding modes and highlight the significance of nucleotide-bound states in modulating TcdB affinity.

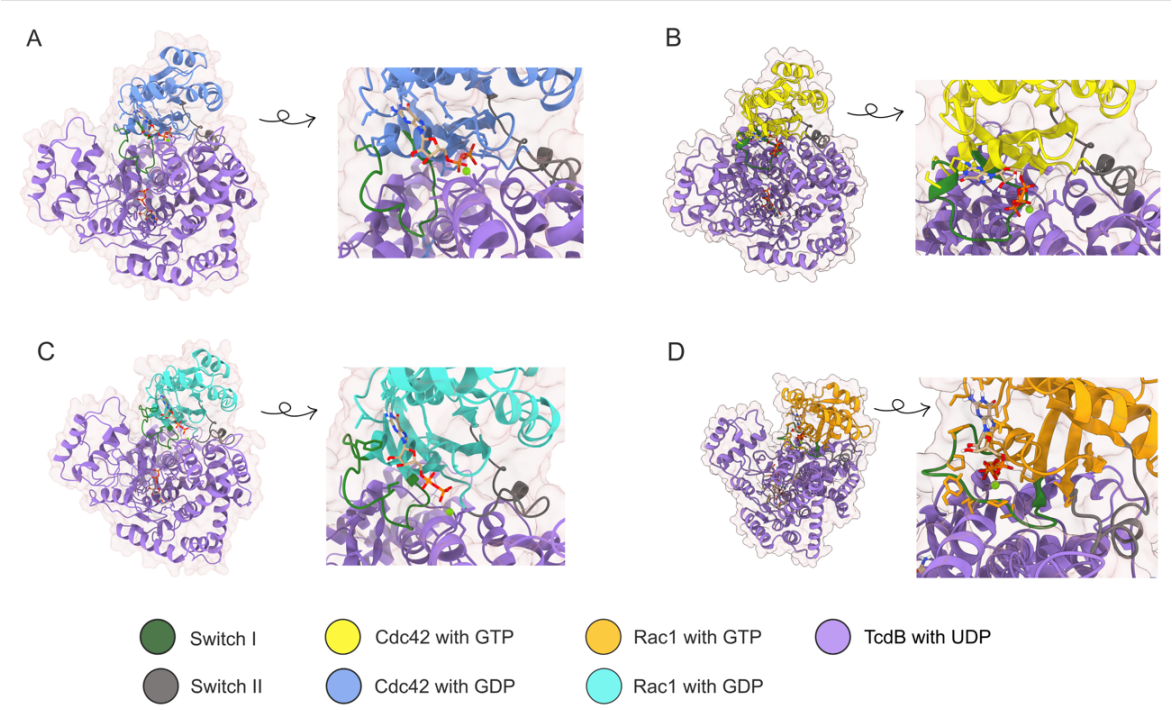


Figure 3 : Modelled binding conformations of TcdB with Rac1 and Cdc42. (A) TcdB complexed with GDP-bound Cdc42, (B) TcdB complexed with GTP-bound Cdc42, (C) TcdB complexed with GDP-bound Rac1, and (D) TcdB complexed with GTP-bound Rac1. In each inset image, magnesium ions are depicted as light green spheres, while GDP and GTP molecules are illustrated as stick representations.

3.2. Selective Binding Mechanism of TcdB to GDP-Bound Rac1

To investigate the dynamic aspects of TcdB interactions with host GTPases, a total of 6 microseconds of all-atom molecular dynamics (MD) simulations were performed. These simulations included TcdB in complex with GDP- and GTP-

bound forms of Rac1 and Cdc42, each system replicated in three independent 500-ns runs to ensure statistical robustness.

Root Mean Square Fluctuation (RMSF) and Root Mean Square Deviation (RMSD) analyses were used to evaluate the residue-level flexibility and overall conformational stability of these complexes. Results indicated that Rac1 exhibits markedly increased conformational fluctuations when bound to GTP, compared to its GDP-bound form (Figure 4C and 4F). This suggests that the nucleotide-bound state of Rac1 strongly influences its dynamic behaviour when interacting with TcdB. On the other hand, Cdc42 displayed relatively stable dynamics across both nucleotide states (Figure 4A, 4B, and 4E), indicating that the observed dynamic modulation is specific to Rac1.

All RMSD and RMSF calculations were carried out using Cα atoms across triplicate trajectories. The analysis revealed that the GDP-bound Rac1–TcdB complex consistently showed lower RMSD and RMSF values, suggesting a more stable and tightly bound conformation in comparison to the GTP-bound counterpart. These findings collectively point toward a ligand-dependent selectivity, wherein TcdB preferentially and more stably associates with the inactive GDP-bound form of Rac1, rather than the active GTP-bound form.

While this section specifically characterizes the structural and dynamic selectivity of TcdB toward Rac1, it also provides a functional rationale for the subsequent inhibitor design strategy. Although the TcdB–Rac1 interaction itself is crucial for substrate recognition, the enzymatic activity of TcdB requires UDP-glucose as a co-substrate. The glucosyltransferase domain (GTD) of TcdB becomes catalytically active in the presence of UDP-glucose, which is essential for transferring a glucose moiety to specific threonine residues on Rac1, thereby inactivating it.

Therefore, while Rac1 binding determines target selectivity, the presence of UDP-glucose governs TcdB's catalytic function. Inhibiting this enzymatic step—parti-

cularly by targeting the UDP-glucose binding pocket within TcdB—was the strategic focus of the therapeutic inhibitor design presented in later sections. By blocking the UDP-binding site, candidate molecules can potentially halt the glycosylation process, and thereby prevent Rac1 inactivation, even in the presence of TcdB.

To further investigate the structural basis of the observed binding energy differences, geometric distance analyses were performed to monitor the spatial relationships between the Mg^{2+} ion and key molecular components involved in nucleotide coordination. As presented in Figure 5, the distance between Mg^{2+} and the nucleotide (GDP or GTP) showed distinct patterns across complexes. In Rac1 systems, the Mg^{2+}–GDP complex exhibited increased fluctuations and larger average distances compared to its GTP-bound counterpart, suggesting weaker and more transient coordination. This observation aligns with the lower binding free energy of the GDP-bound Rac1–TcdB complex. Furthermore, Mg^{2+} displayed dynamic positional shifts relative to Asp461 of TcdB and Thr35 of Rac1, particularly in the GDP-bound state. In contrast, Cdc42 complexes showed relatively stable Mg^{2+} coordination regardless of the nucleotide bound, supporting the lack of significant energetic difference in those systems. These spatial metrics provide further structural support for the hypothesis that GDP-bound Rac1–TcdB interactions are uniquely sensitive to Mg^{2+} positioning, contributing to the differential stability observed in binding energy calculations.

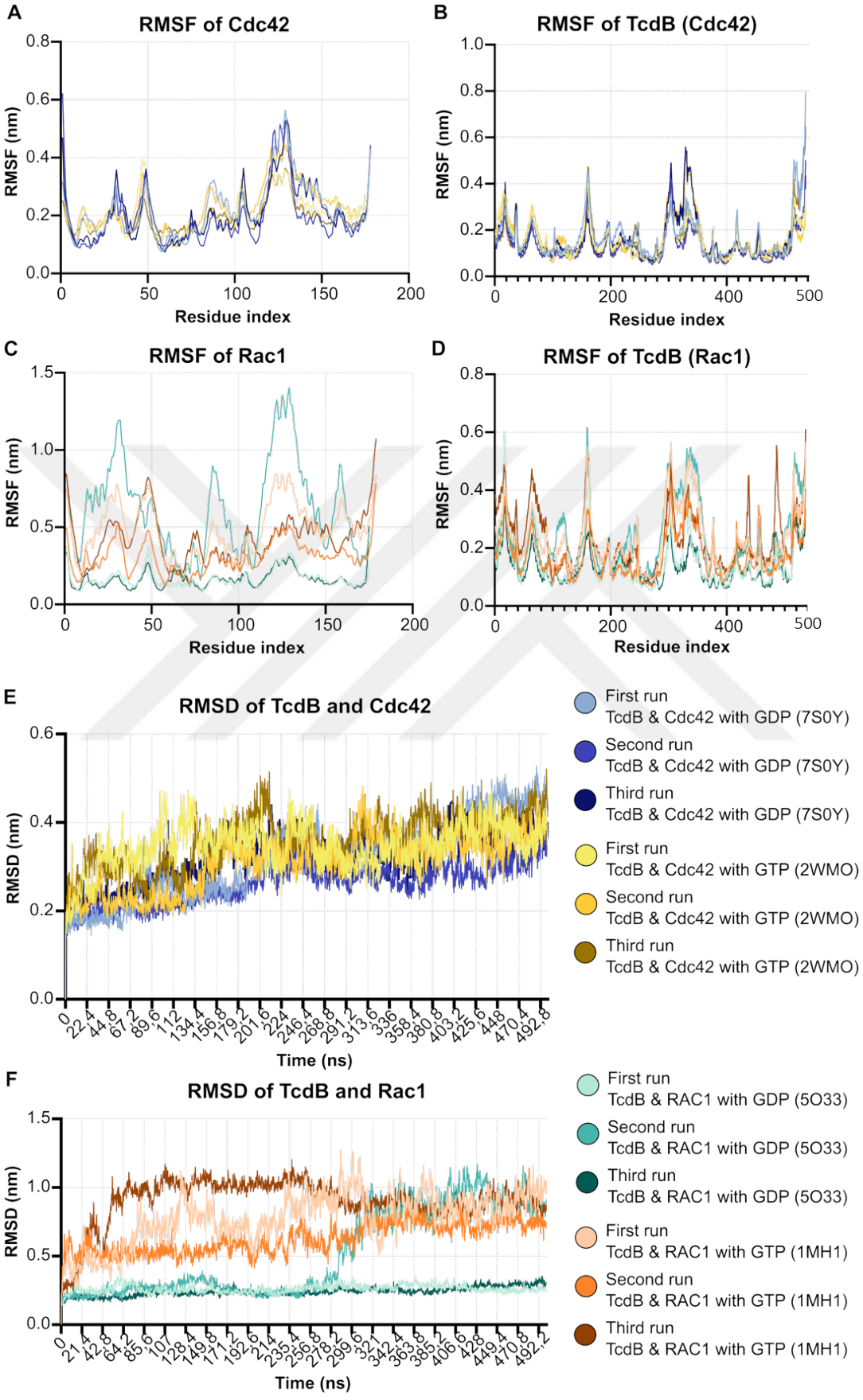

A
RMSF of Cdc42
RMSF (nm)
Residue index
B
RMSF of TcdB (Cdc42)
RMSF (nm)
Residue index
C
RMSF of Rac1
RMSF (nm)
Residue index
D
RMSF of TcdB (Rac1)
RMSF (nm)
Residue index
E
RMSD of TcdB and Cdc42
RMSD (nm)
Time (ns)
First run
TcdB & Cdc42 with GDP (7S0Y)
Second run
TcdB & Cdc42 with GDP (7S0Y)
Third run
TcdB & Cdc42 with GDP (7S0Y)
First run
TcdB & Cdc42 with GTP (2WMO)
Second run
TcdB & Cdc42 with GTP (2WMO)
Third run
TcdB & Cdc42 with GTP (2WMO)
F
RMSD of TcdB and Rac1
RMSD (nm)
Time (ns)
First run
TcdB & RAC1 with GDP (5O33)
Second run
TcdB & RAC1 with GDP (5O33)
Third run
TcdB & RAC1 with GDP (5O33)
First run
TcdB & RAC1 with GTP (1MH1)
Second run
TcdB & RAC1 with GTP (1MH1)
Third run
TcdB & RAC1 with GTP (1MH1)

Figure 4. RMSF and RMSD analyses of TcdB–Rac1 and TcdB–Cdc42 complexes across triplicate MD simulations. (A) RMSF plot of Cdc42 in GDP- and GTP-bound states, showing residue-wise flexibility. (B) RMSF plot of TcdB when in complex with Cdc42, comparing nucleotide-bound states. (C) RMSF plot of Rac1 in GDP- and GTP-bound states, highlighting increased fluctuations in GTP-bound form. (D) RMSF plot of TcdB when in complex with Rac1, illustrating differential stability across nucleotide-bound forms. (E) RMSD trajectories of the TcdB–Cdc42 complexes over time, indicating overall structural stability across replicates. (F) RMSD trajectories of the TcdB–Rac1 complexes over time, showing increased conformational variability in GTP-bound Rac1 systems.

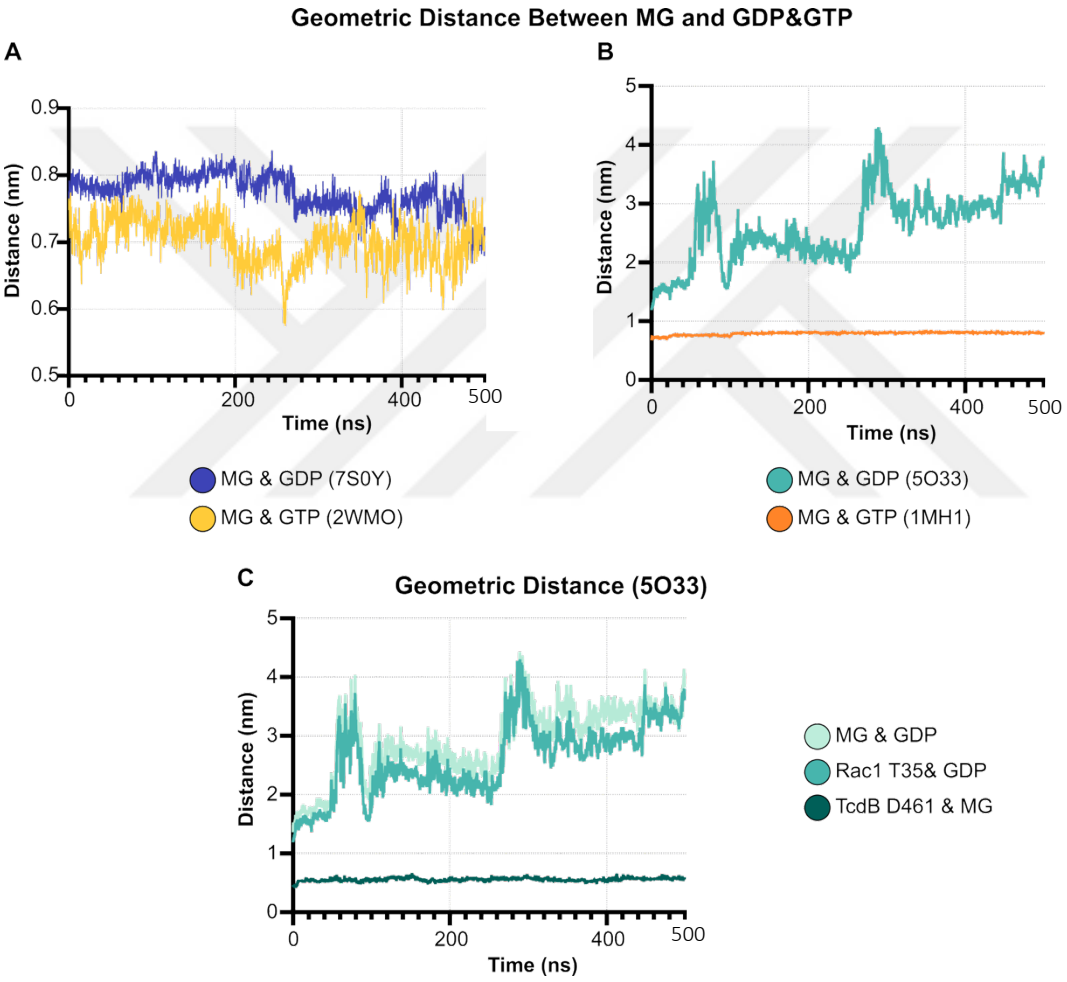


Figure 5: Mean geometric distance measurements between Mg^{2+} ion and key molecular components in GTPases interacting with TcdB. (A) Distance between Mg^{2+} and the nucleotide (GDP or GTP) in Cdc42 complexes, (B) Distance between Mg^{2+} and the nucleotide (GDP or GTP) in Rac1 complexes, (C) Distance between Mg^{2+} and the side chains of TcdB residue D461, Rac1 residue T35, and GDP.

Binding free energy calculations via MM-PBSA further substantiated the differential interactions of TcdB with distinct Rac1 nucleotide states. The GDP-bound Rac1–TcdB complex exhibited significantly more favourable binding energies, ranging from −100 to −200 kcal/mol, compared to the GTP-bound form, which showed binding energies close to 0 kcal/mol (Figure 6). In contrast, Cdc42 showed no significant binding energy difference between its GDP- and GTP-bound states Figures 6 and A1). Per-residue energy decomposition analysis identified Asp461 of TcdB as a critical contributor to the Rac1 interaction interface. Notably, this residue formed a coordination bond with Mg^{2+}, which appeared to play a dual role: stabilizing the cation within the complex, while at the same time reducing the binding affinity of Rac1 for GDP (Figures A1–A2). Comparative analysis of energy profiles over the first and last 40 ns of the simulation trajectory revealed significant shifts following GDP dissociation. These shifts support the hypothesis that the binding interface is GDP-sensitive and structurally responsive to the presence or absence of the nucleotide. All energy values were calculated in units of kcal/mol.

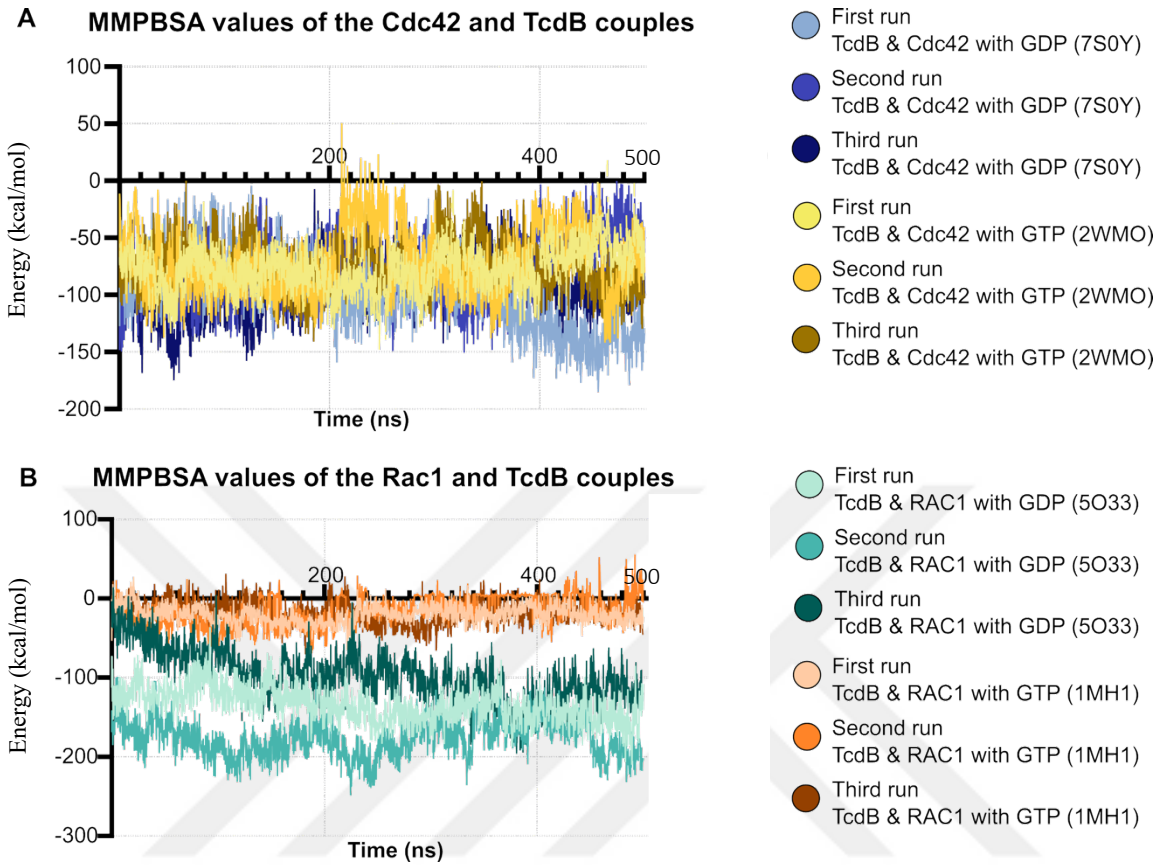


Figure *6* : MMPBSA analyses of TcdB/ Rac1 or Cdc42 and their complexes across triplicate MD simulations. (A) MMPBSA binding free energy scores for the TcdB–Cdc42 complex calculated from three independent molecular dynamics simulations. (B) MMPBSA binding free energy scores for the TcdB–Rac1 complex calculated from three independent molecular dynamics simulations.

3.3. Enhanced Sampling Analysis of Rac1–TcdB Complexes

The binding stability of Rac1 in its GDP- and GTP-bound states with TcdB was comprehensively evaluated through enhanced sampling methods including umbrella sampling and well-tempered meta-dynamics (WT-MetaD). In agreement with MM-PBSA calculations, umbrella sampling revealed markedly more favourable binding free energies for the GDP-bound Rac1–TcdB complex (−99.45 kJ/mol, equivalent to −23.77 kcal/mol), compared to the significantly weaker in-

teraction observed in the GTP-bound form (−27.11 kJ/mol, ≈ −6.48 kcal/mol) (Figure 7E–F). The PMF profiles indicated a deeper energy well and higher dissociation barrier for GDP-bound complexes (Figure 7C–D), whereas the GTP-bound forms showed shallow landscapes, suggesting rapid dissociation kinetics.

Note: ΔG values shown on Figure 7 are reported in kcal/mol, whereas MM-PBSA results are expressed in kJ/mol to maintain consistency with force field energy units. For direct comparison, 1 kcal/mol ≈ 4.184 kJ/mol.

WT-MetaD simulations further corroborated these findings, demonstrating that the GTP-bound complex gradually dissociates over time, as evidenced by increased solvent-accessible surface area (SASA; Figure 8A), a decline in hydrogen bonding (Figure 8B), and a growing distance between the centres of mass of Rac1 and TcdB (Figure 8C). In contrast, the GDP-bound form remained structurally stable with consistently lower SASA, more persistent H-bonding, and minimal centre-of-mass displacement—indicative of a robust, tightly associated complex.

Importantly, Figure 7 also illustrates the time-resolved harmonic pulling forces used in umbrella sampling, reflecting the physical resistance encountered during GDP or GTP dissociation (Panels A–B and E–F). Stronger forces and longer resistance times in GDP-bound systems further support the energetic preference observed.

It is important to note that the simulation of the GTP-bound complex terminates earlier than the GDP-bound trajectory due to spontaneous dissociation events occurring during the enhanced sampling. Once critical interactions between Rac1 and TcdB were lost, no meaningful data could be extracted beyond this point. Therefore, the shorter trajectory duration for the GTP-bound system is not a limitation, but rather a reflection of the actual physical instability of the complex, which is also visibly supported by the sharp divergence trends in the SASA and distance plots.

Detailed residue-level analyses (Figure 9) identified flexible hotspots near residues Asp461 and Asp524, which appear to facilitate GDP release by coordinating Mg^{2+}. These findings support a proposed "ion mugging" mechanism, where TcdB sequesters Mg^{2+} to destabilize the GDP within Rac1's nucleotide-binding pocket. Energy decomposition analyses confirmed significant energetic contributions from these regions, particularly electrostatic interactions critical for ion-mediated control of binding stability.

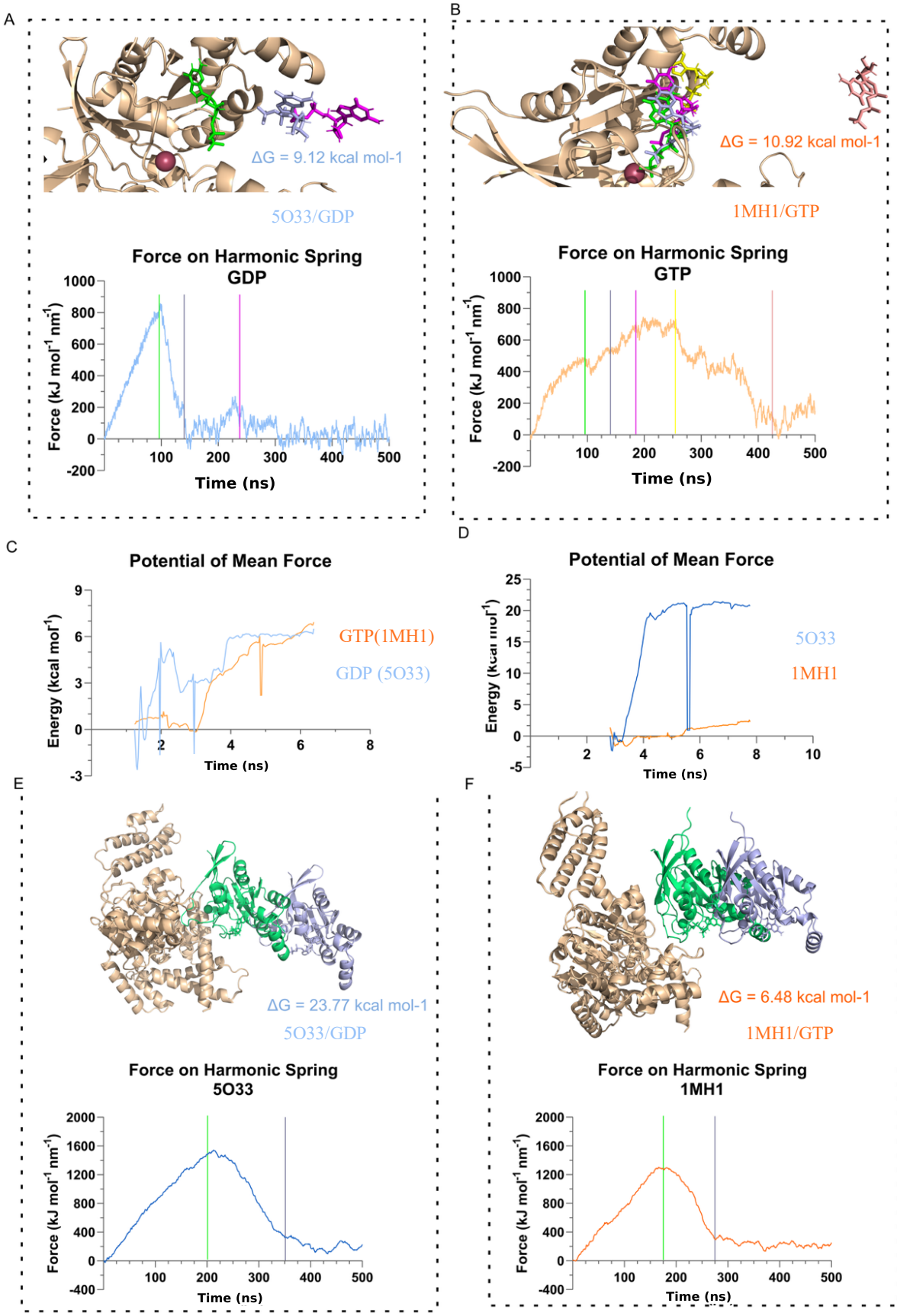

A
ΔG = 9.12 kcal mol-1
5O33/GDP
Force on Harmonic Spring
GDP
Force (kJ mol-1 nm-1)
Time (ns)
B
ΔG = 10.92 kcal mol-1
1MH1/GTP
Force on Harmonic Spring
GTP
Force (kJ mol-1 nm-1)
Time (ns)
C
Potential of Mean Force
Energy (kcal mol-1)
Time (ns)
GTP(1MH1)
GDP (5O33)
D
Potential of Mean Force
Energy (kcal mol-1)
Time (ns)
5O33
1MH1
E
ΔG = 23.77 kcal mol-1
5O33/GDP
Force on Harmonic Spring
5O33
Force (kJ mol-1 nm-1)
Time (ns)
F
ΔG = 6.48 kcal mol-1
1MH1/GTP
Force on Harmonic Spring
1MH1
Force (kJ mol-1 nm-1)
Time (ns)

Figure 7 : Umbrella sampling analysis for ligand and complex dissociation from Rac1 and TcdB complexes. (A–B) Harmonic biasing forces and associated conformations applied to extract GDP and GTP from Rac1 in the TcdB-bound state, respectively. (C–D) Potential of mean force (PMF) profiles illustrating the free energy landscapes for detaching GDP/GTP from Rac1, and from the Rac1–TcdB complexes. The ΔG values (calculated as $E_{max} - E_{min}$, in kcal/mol) are shown on each PMF plot: −2.61 to 6.51 for GDP–Rac1, −2.69 to 8.23 for GTP–Rac1, −2.35 to 21.42 for Rac1(GDP)–TcdB, and −6.48 to 0.00 for Rac1 (GTP)–TcdB. (E–F) Harmonic Spring forces and corresponding structural snapshots during the extraction of Rac1 bound to GDP and GTP from TcdB, respectively.

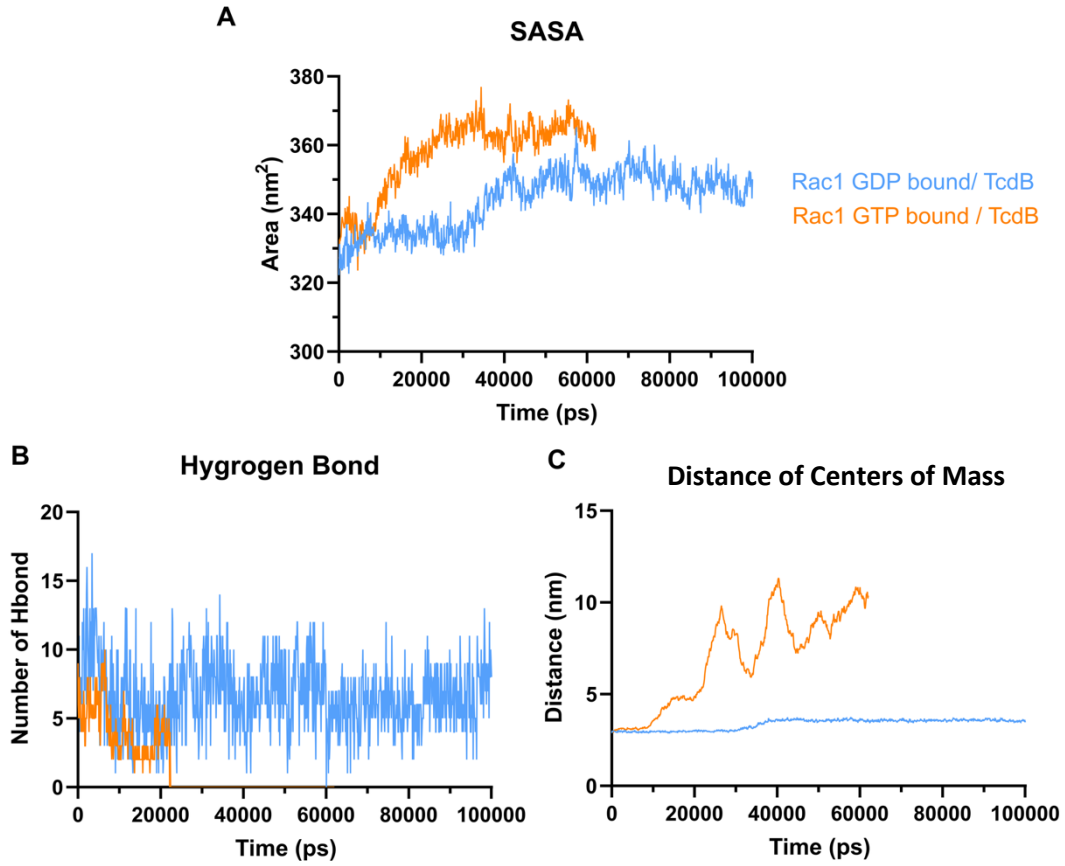


Figure 8: Results of well-tempered Metadynamics (WT-MetaD) simulations. (A) Solvent accessible surface area (SASA) variations, (B) Number of hydrogen bonds formed during the simulation, (C) Distance between the centres of mass of Rac1 and TcdB throughout the trajectory

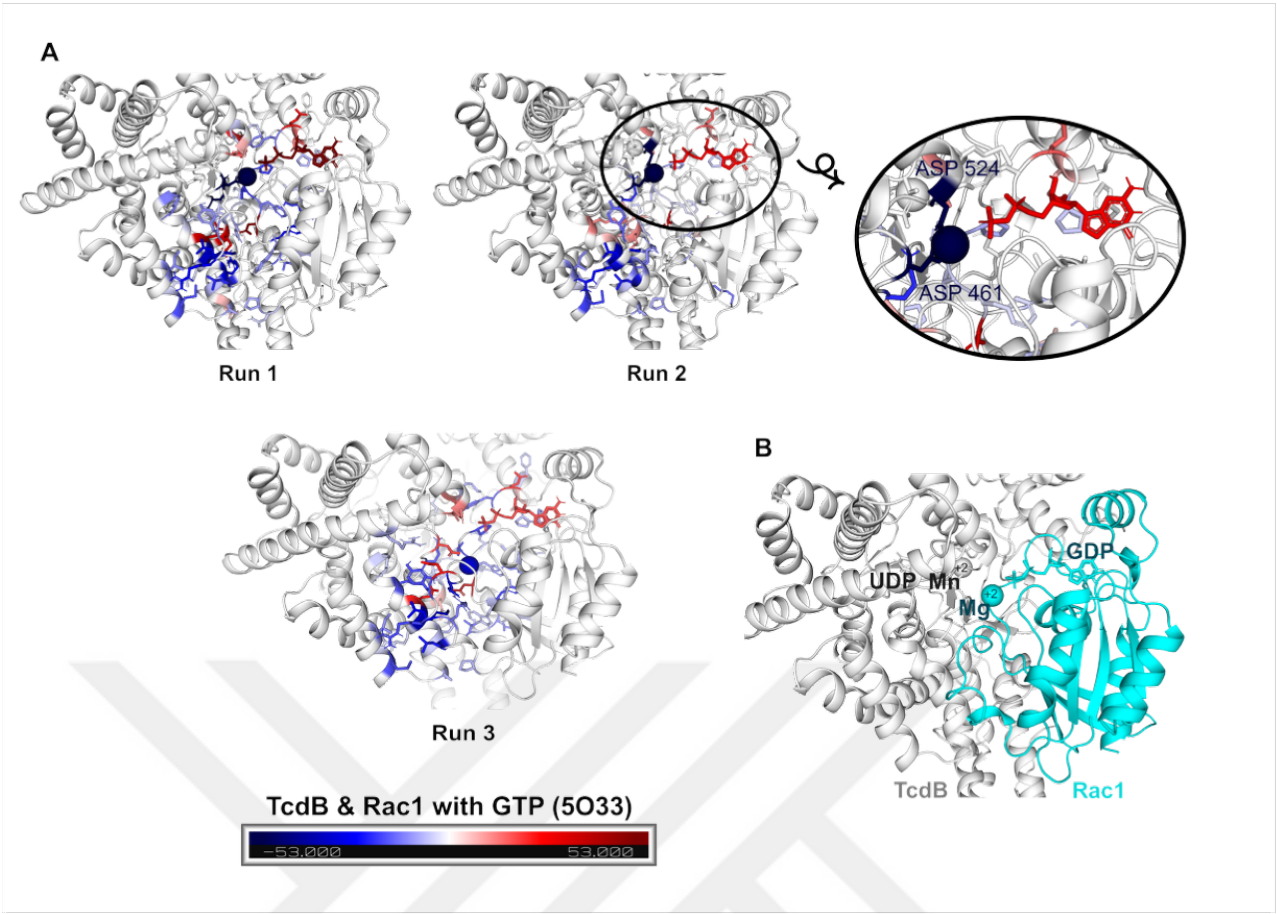


Figure 9: Visualization of residue-level energetic contributions within the TcdB–GDP-bound Rac1 complex. (A) Average per-residue interaction energies across three independent simulations, depicted using a blue-to-red colour gradient spanning −53 to +53 kcal/mol. (B) Representation of the bound nucleotide and ion components; the cyan-coloured region denotes the GDP-bound Rac1 structure including the coordinated Mg^{2+} ion.

3.4. Small Molecule Ligands

In this study, a structure-based virtual screening strategy was implemented to identify potential small-molecule inhibitors targeting the UDP-glucose binding pocket of the glucosyltransferase domain (GTD) of *C. difficile* toxin B (TcdB). TcdB, once internalized into host cells, inactivates members of the Rho family of small GTPases—specifically RhoA, Rac1, and Cdc42—via glycosylation, disrupting the actin cytoskeleton and impairing cellular morphology and intercellular junctions (5,53).

The GTD of TcdB possesses a specific pocket for binding UDP-glucose, its natural substrate. Molecules capable of competitively occupying this site may inhibit the enzymatic activity of TcdB by preventing substrate access. Such compounds are defined as competitive inhibitors and hold promise for neutralizing the toxic intracellular effects of the enzyme (30).

Following protein structure preparation, more than 14 million drug-like compounds from the ZINC database were virtually screened against the UDP-glucose binding site of TcdB using AutoDock Vina (70). The top four candidate molecules exhibited docking scores around −6.3 to −6.6 kcal/mol. While these values are comparable to, and in some cases slightly better than, known reference compounds such as apigenin (−6.28 kcal/mol) and vancomycin (−6.27 kcal/mol), the differences are modest and within the expected scoring range of docking algorithms (Table 1). Interestingly, two of these top candidates—Molecule_295 and Molecule_98—were chemically identical, despite originating from different compound scaffolds (apigenin and noeuromycin, respectively), suggesting convergence toward the same optimal binding structure.

Nevertheless, the comparable or slightly improved binding affinities suggest that these candidate molecules may hold promise as potential inhibitors of TcdB activity, particularly in the context of structure-based optimization and further validation through molecular dynamics simulations and binding free energy calculations.

Binding mode analyses revealed that these molecules established key hydrogen bonds and hydrophobic interactions with critical residues such as Asp2865, Glu449, and His448—residues known to interact with the diphosphate group of UDP-glucose (14). The candidates structurally mimicked the binding orientation of the native substrate, suggesting potential to directly block TcdB's enzymatic activity. This mechanism could preserve actin dynamics and signalling pathways by preventing GTPase glycosylation (29).

This inhibitor-based strategy offers distinct advantages over traditional antibiotic therapies. By directly targeting TcdB's toxic function rather than bacterial viability, such inhibitors may spare the gut microbiota and reduce the likelihood of resistance development, which are crucial therapeutic considerations in recurrent *C. difficile* infections (CDI) (3,26).

Pharmacokinetic and toxicological assessments further supported the drug-likeness of these candidates (Table 2). Furthermore, to address reviewer concerns regarding ranking methodology and reference selection, it is also important to clarify the rationale behind the use of binding affinity as the primary ranking metric and the –8.0 kcal/mol threshold. While other physicochemical and structural parameters were evaluated downstream, the initial cutoff was informed by established virtual screening protocols, where binding energies ≤ –8.0 kcal/mol are broadly indicative of strong and potentially biologically relevant interactions. Affinity scores slightly above this threshold (e.g., –7.9 kcal/mol) were not automatically discarded but re-evaluated within a multi-parametric decision framework that included ADMET properties, off-target binding (particularly RAC1), and structural alerts.

The reference compounds—apigenin and the soy-derived UDP-glucose analogue (with known PDB structures bound to TcdB)—were selected because they have documented experimental interaction with the toxin's GTD domain, serving as mechanistic benchmarks. Two additional compounds, a large antibiotic (vancomycin) and a CDI-used small molecule without direct evidence of TcdB binding, were included to test the specificity and selectivity of the docking pipeline. Notably, vancomycin yielded a high docking score; however, its large molecular size (>1 kDa) likely prevents cellular uptake, indicating a probable false-positive docking result that would not translate into intracellular efficacy.

Ultimately, our strategy prioritized not only docking rank but also a combination of biological plausibility, drug-likeness, and off-target selectivity. By doing so,

we mitigated the risk of overlooking active compounds due to marginal numerical differences and eliminated false positives arising from non-viable drug structures. All five molecules complied with Lipinski's Rule of Five (78,79), showed high predicted oral bioavailability, and did not present hepatotoxic or mutagenic alerts. Moreover, the candidates tested negative for PAINS and Brenk filters and exhibited logP and aqueous solubility values within acceptable drug design ranges, making them viable candidates for further pharmaceutical development (80,81).

Nevertheless, a major limitation of this study was the inability to perform molecular dynamics (MD) simulations due to system instability. As a result, the dynamic stability and time-resolved interaction profiles of the ligand-protein complexes could not be assessed. Future work should involve MD simulations and binding free energy calculations such as MM-GBSA to enhance the reliability and accuracy of the proposed inhibitor interactions (82).

Furthermore, the docking protocol used here assumes a static protein conformation and does not account for protein flexibility or induced fit effects. This introduces limitations in accurately predicting binding modes. Therefore, incorporating enhanced sampling techniques such as Metadynamics or umbrella sampling in future studies would be crucial to better evaluate binding free energy landscapes.

Table 1: Docking Scores of References and Candidate Molecules

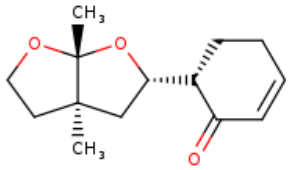
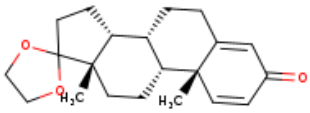
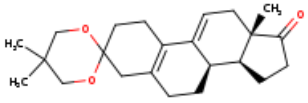
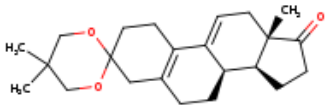
	AutoDock Vina (References)	SMILES	Auto Dock Vina (Candidates)	SwissDock
Molecule_47 Candidate From Vancomycin 	-29.268 (Vancomy-cin)	C[C@]12CCO[C@]1(C)O[C@H]([C@@H]1CCC=CC1=O)C2	-6.269	-5.811
Molecule_175 Candidate From Fidaxomicin 	-5.745 (Fidaxomicin)	C[C@]12CC[C@@H]3[C@@H](CCC4=CC(=O)C=C[C@]34C)[C@H]1CCC12OCCO1	-6.57	-7.309
Molecule_295 Candidate From Apigenin (same as Candidate 98) 	-10.195 (Apigenin)	C[C@]12CC=C3C4=C(CC[C@@H]3[C@@H]1CCC2=O)CC1(CC4)OCC(C)(C)CO1	-6.28	-8.142
Molecule_98 Candidate From Noeuromycin (same as Candidate 295) 	-4.924 (Noeuromy-cin)	C[C@]12CC=C3C4=C(CC[C@@H]3[C@@H]1CCC2=O)CC1(CC4)OCC(C)(C)CO1	-6.32	-8.134

Table 2 : Pharmacokinetic and Drug-Likeness Properties of Selected Small-Molecule Inhibitors.

Parameters	**Molecule175**	**Molecule 98**	**Molecule 47**
Molecular Formula	$C_{23}H_{32}O_3$	$C_{21}H_{28}O_3$	$C_{14}H_{22}O_3$
Molecular Weight	356.50	328.45	236.31
TPSA	35.53	35.53	35.53
Rotatable Bonds	0	0	1
H-bond Acceptors / Donors	3 / 0	3 / 0	3 / 0
Molar Refractivity	103.04	93.43	64.74
Fraction Csp^3	0.78	0.76	0.79
LogP	4.08	3.64	2.26
Water Solubility (LogS, ESOL)	-3.91	-4.15	-2.47
GI Absorption	High	High	High
BBB Permeability	Yes	Yes	Yes
P-gp Substrate	Yes !	No	No
CYP1A2 Inhibition	No	No	No
CYP2C19 Inhibition	No	No	No
CYP2C9 Inhibition	Yes !	Yes !	No
CYP2D6 Inhibition	No	No	No
CYP3A4 Inhibition	Yes !	No	No
Lipinski Rule Violations	0	0	0
Other Drug-Likeness Rules	All passed	All passed	All passed
Bioavailability Score	0.55	0.55	0.55
PAINS Warning	None	None	None
Brenk Alert	None	None	None
Lead-Likeness	1 violation !	1 violation !	1 violation !
Synthetic Accessibility	5.72	5.41	4.46

Overall, the findings of this section highlight the therapeutic promise of small molecules designed to inhibit the glucosyltransferase activity of TcdB. Given the limitations of currently approved antibiotics for CDI—such as vancomycin and fidaxomicin—in terms of microbiota disruption and high recurrence rates, toxin-directed inhibitors offer a more selective, resistance-resilient, fecal microbiota transplantation and microbiota-sparing treatment alternative (55).

3.5. Therapeutic Cyclic Peptides

Toxin B (TcdB), one of the two main exotoxins secreted by *C. difficile*, plays a central role in disrupting the host cytoskeleton and epithelial integrity through the

inactivation of small Rho-family GTPases, including Rac1, RhoA, and Cdc42. This inactivation occurs via a glucosyltransferase reaction that transfers a glucose moiety from UDP-glucose to key threonine residues (e.g., Thr35 in Rac1), thereby blocking GTPase signalling and leading to cytopathic effects and apoptosis (18,39).

The glucosyltransferase domain (GTD) of TcdB contains a catalytically essential pocket centred around Asp270, which coordinates substrate (UDP-glucose) binding and GTPase interaction (42). This site was identified as an ideal therapeutic target for peptide-based inhibition. The goal of this section of the study was to design cyclic peptides capable of competitively and sterically blocking this site, thereby neutralizing TcdB's function.

Initially, the ADRINITSGY sequence—previously reported to interact near Asp270—was used as a starting scaffold for peptide design. However, structural modelling revealed steric clashes within the binding cleft, indicating that the original 10-mer linear peptide was too long to fit optimally. To address this, the sequence was refined to a shorter 8-mer, ADRITPSG, balancing structural compactness with the ability to form a stable head-to-tail cyclic conformation. This design leveraged residue-specific rationales, such as incorporating proline for conformational rigidity and charged/polar residues (Asp, Arg, Thr, Ser) to enhance binding interactions.

To guide this rational design, residues surrounding Asp270 in TcdB were analyzed. The sequence 'L L A A A S I L R I N I T I S G P Y' represents the unique amino acid residues located within 5 Å of Asp270 in the TcdB protein, listed in one-letter codes. This selection revealed eight key residues critical to interaction with candidate peptides, and this microenvironment was defined as the target site for subsequent peptide docking and optimization.

Building upon this optimized core, a comprehensive peptide library was generated through systematic substitutions targeting physicochemical properties: elect-

rostatic charge (e.g., D ↔ K, R), hydrophobicity (I, L, V ↔ F), and hydrogen bonding capacity (S ↔ N; T ↔ N, Q). Using a custom Python script implementing all combinatorial permutations, over 120,000 cyclic peptide variants were produced and structurally modelled using Rosetta's SimpleCycpepPredict protocol to ensure proper cyclization and conformational feasibility.

Subsequently, an in-silico toxicity screen via ToxinPred2 filtered out peptides with predicted adverse profiles, and homology screening using BLASTp against the human non-redundant proteome excluded candidates with significant similarity (E-value < 0.01) to endogenous proteins to minimize off-target effects. These filters reduced the candidate pool drastically, leaving 15 non-toxic, highly specific cyclic peptides.

Docking simulations against the Asp270-centered binding site of TcdB were performed using the HDOCK platform to evaluate binding affinities and interaction patterns. Among these, two peptides—Peptide 2 (AGNADRVN) and Peptide 15 (GNVRADAN)—stood out with top docking scores of approximately –210 and –229 kcal/mol, respectively, coupled with minimal toxicity and negligible homology to human proteins.

These two candidates were subjected to triplicate 500 ns molecular dynamics simulations using the AMBER14SB force field with TIP3P explicit water under physiological ionic conditions (0.15 M NaCl, pH 7.4), with protonation states assigned via the H++ server. MD analyses confirmed stable binding conformations and favourable interaction profiles with TcdB at the target site, supporting their potential as effective cyclic peptide inhibitors.

Table 3: Potential Cyclic Peptides

	Peptide Number	Sequence	Docking Score (kcal/mol)	ML Score
1	2759	RAADGNVN	-225	0.045
2	**8006**	**AGNADRVN**	**-210**	**0.045**
3	14142	AGNVNADR	-218	0.045
4	23193	RVNADGNA	-231	0.045
5	28895	NRVEAAGN	-242	0.045
6	41054	NADRVNAG	-227	0.045
8	41438	ARADGNVN	-216	0.04
9	44899	NVRADAGN	-221	0.045
10	56158	GNADRVNA	-210	0.045
11	58697	ADRAGNVN	-228	0.045
12	65869	ADGARVNN	-220	0.045
13	66285	ADRVNAGN	-245	0.045
14	73783	GNRLNEAA	-230	0.045
15	**116167**	**GNVRADAN**	**-229**	**0.035**

3.6. Dynamic and Structural Stability of Peptide–TcdB Complexes

Distance analyses between key residues (Asp270, Glu449) and peptide heavy atoms showed a maintained interaction range of 3.2–4.5 Å throughout the simulation, highlighting sustained binding within the active pocket (Figure 10). The RMSD (Root Mean Square Deviation) analysis revealed that both peptides maintained structural stability over 500 ns, with average RMSD values remaining below 2.5 Å (Figure 11). Peptide 15 displayed less deviation, indicating a more stable binding pose. RMSF (Root Mean Square Fluctuation) results further confirmed this, showing low flexibility across residues in the binding pocket, especially surrounding Asp270, Glu449, and His448 (Figures 12&13). These regions remained rigid during the simulation, implying a “lock-in” effect due to peptide

binding. Most interactions involved residues in the UDP-glucose coordination site, supporting their role as competitive inhibitors.

3.7. MM-GBSA Binding Energy and Dual Inhibition Mechanism

MM-GBSA binding free energy calculations (Figure 14) yielded ΔG_bind values of –52.3 kcal/mol for Peptide 15 and –45.6 kcal/mol for Peptide 2, indicating highly favourable, spontaneous interactions. These energies surpass the performance of previously identified small-molecule candidates that lacked dynamic stability. Importantly, structural snapshots (Figure 15) confirmed a dual inhibition mechanism. The peptides not only occupy the UDP-glucose binding pocket but also extend toward the Rac1-binding surface, likely blocking protein–protein interaction required for GTD's glucosyltransferase function. This double-binding mode mimics a steric and functional blockade, inhibiting both substrate turnover and effector protein engagement. Given that the TcdB-induced glycosylation of GTPases is a central step in epithelial barrier disruption, inhibition of both binding surfaces by a single ligand presents a powerful neutralization strategy. Additionally, the cyclic nature of the peptides may offer improved stability against proteolysis in the gut environment, a critical consideration for CDI therapy.

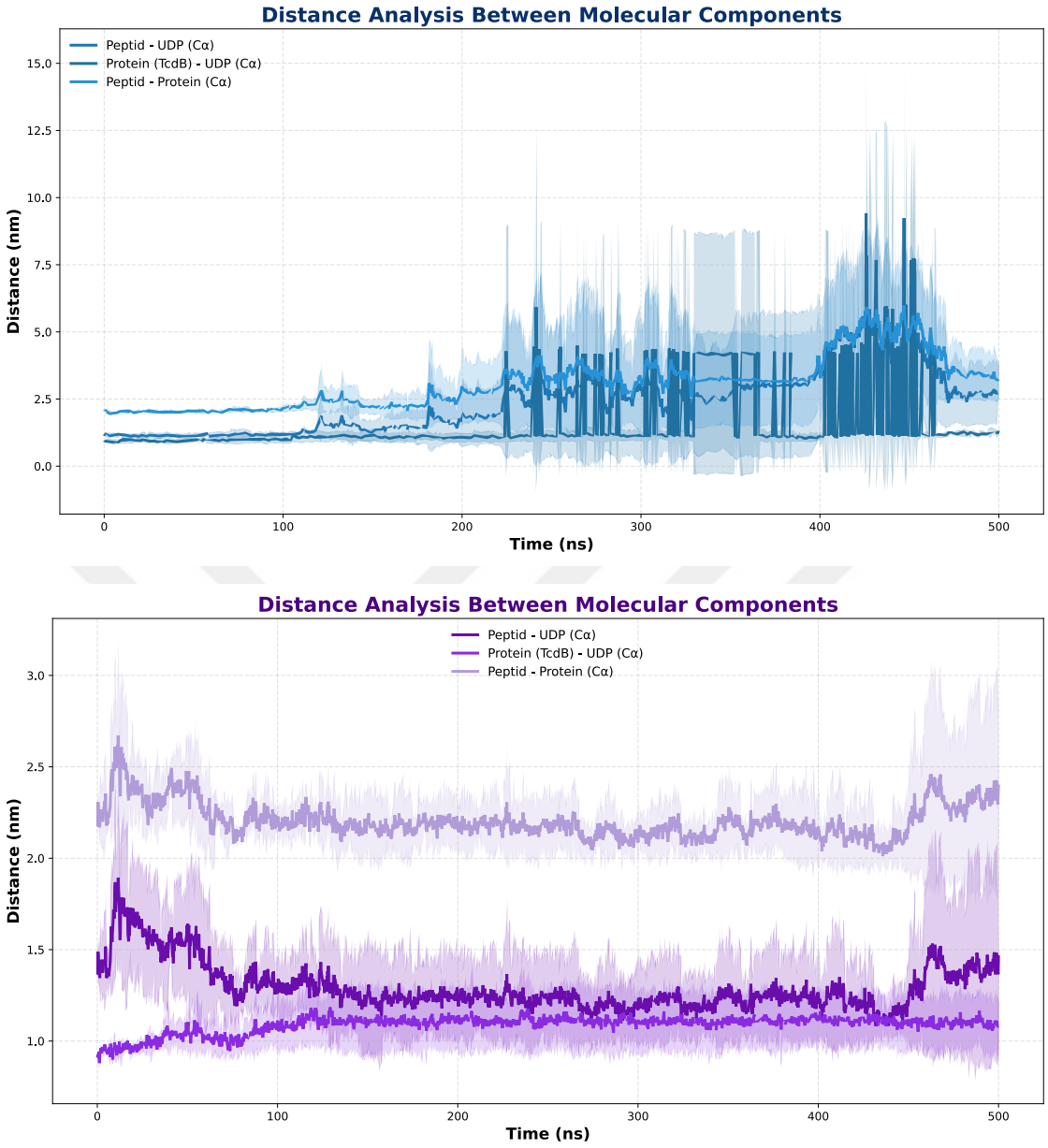


Figure 10 : Distance Analysis Between Molecular Components. Distances between peptide 2 –UDP (Cα), protein –UDP (Cα), and peptide 2 –protein (Cα) atoms throughout a 2500-frame molecular dynamics simulation. Increased fluctuations in the protein–UDP distance after frame 1000 indicate potential destabilization or competitive interference at the UDP binding site. Peptide 15 –UDP and peptide 15 –protein distances remain consistently low, suggesting a stable ternary interaction. The protein–UDP distance also remains relatively constant, implying that UDP may remain bound despite partial peptide occupancy of the binding site. Shaded areas represent standard deviation across trajectory frames.

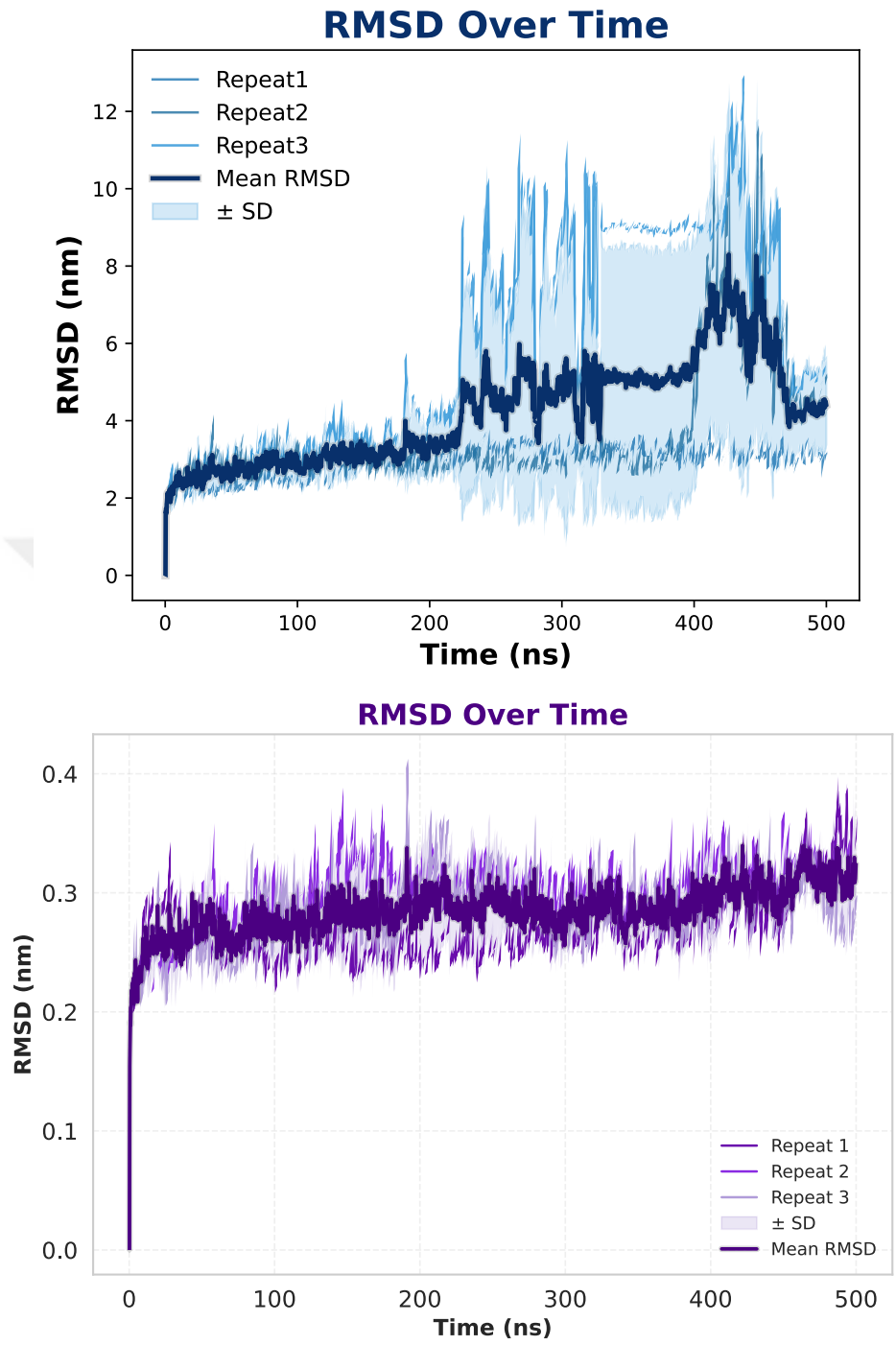


Figure 11 : Root Mean Square Deviation (RMSD) profiles for peptide–protein complexes during 500 ns molecular dynamics simulations. The upper panels depict RMSD trajectories for Peptide 2, and the lower panels correspond to Peptide 15. Each graph represents the mean RMSD calculated from three independent simulation replicates, with the associated standard deviation (SD) shown as shaded regions around the mean. These analyses provide insight into the conformational stability and dynamic behaviour of the peptide complexes over the simulation timeframe.

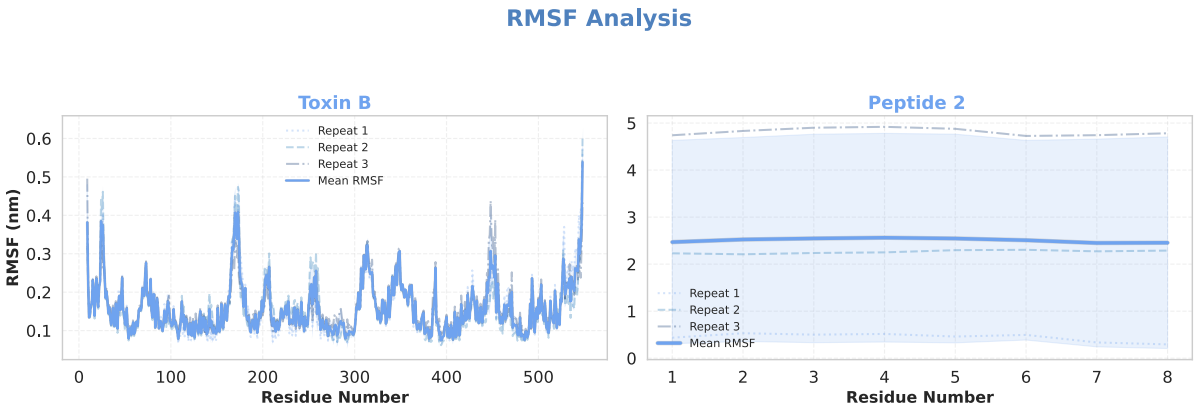


Figure 12 : RMSF analysis of the TcdB–Peptide 2 complex. TcdB shows low residue fluctuations, indicating structural stability. Peptide 2 remains stable, with slight mobility at the termini.

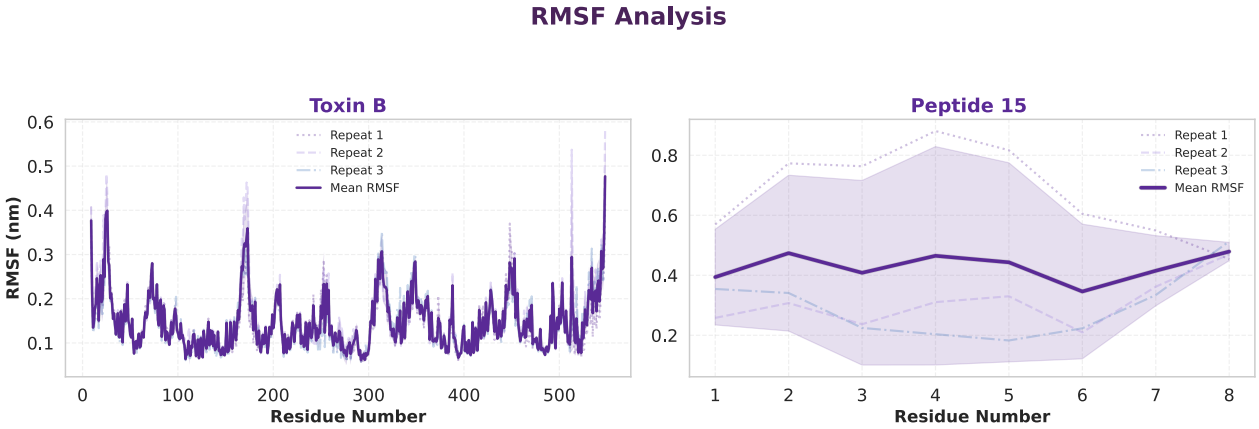


Figure 13 : RMSF analysis of the TcdB–Peptide 15 complex. TcdB shows low residue fluctuations, indicating structural stability. Peptide 15 remains stable, with slight mobility at the termini.

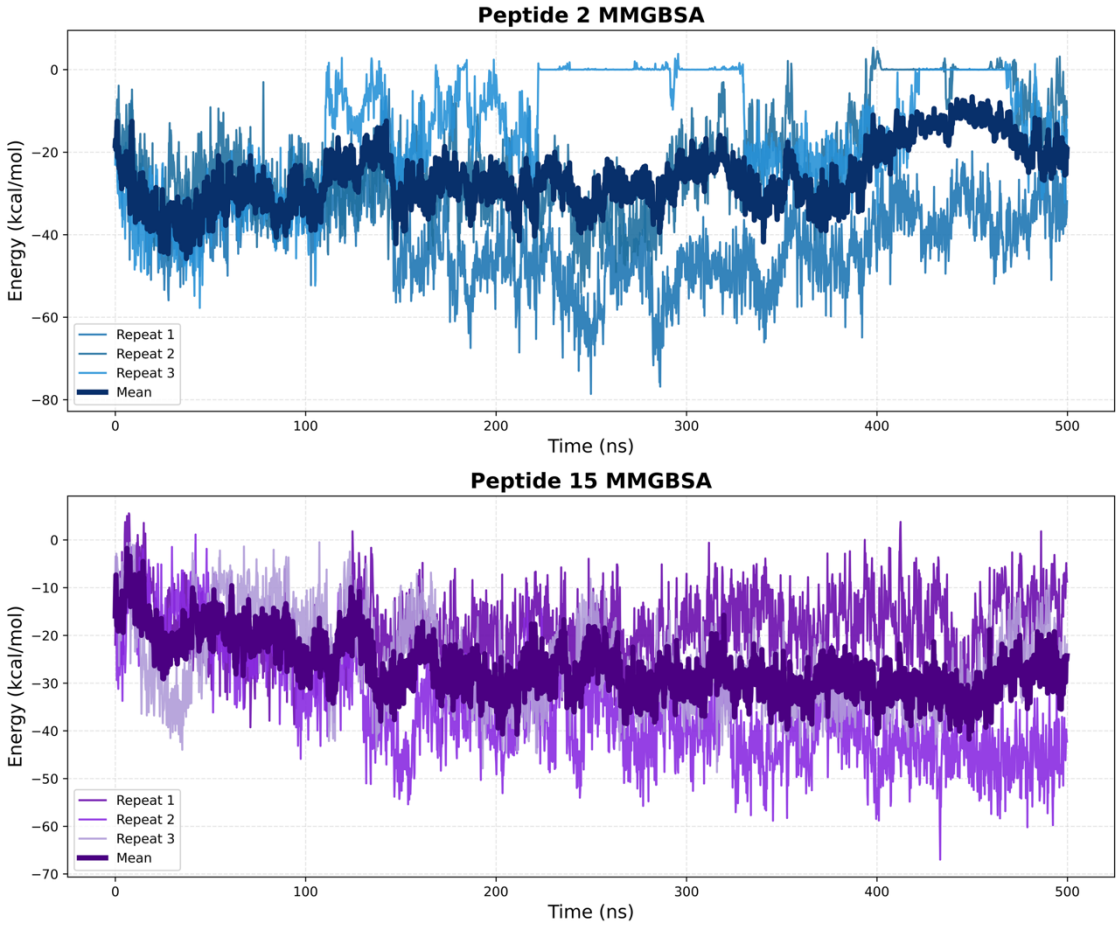


Figure 14 : MM-GBSA binding free energy analysis of TcdB–peptide complexes. Binding free energy profile of Peptide 2 interacting with TcdB over 2500 frames, shown across three independent replicates and their average (blue). Binding free energy profile of Peptide 15 under the same conditions. Energies are presented in kcal/mol. Each line represents an individual trajectory, with the bold line indicating the mean binding energy(purple).

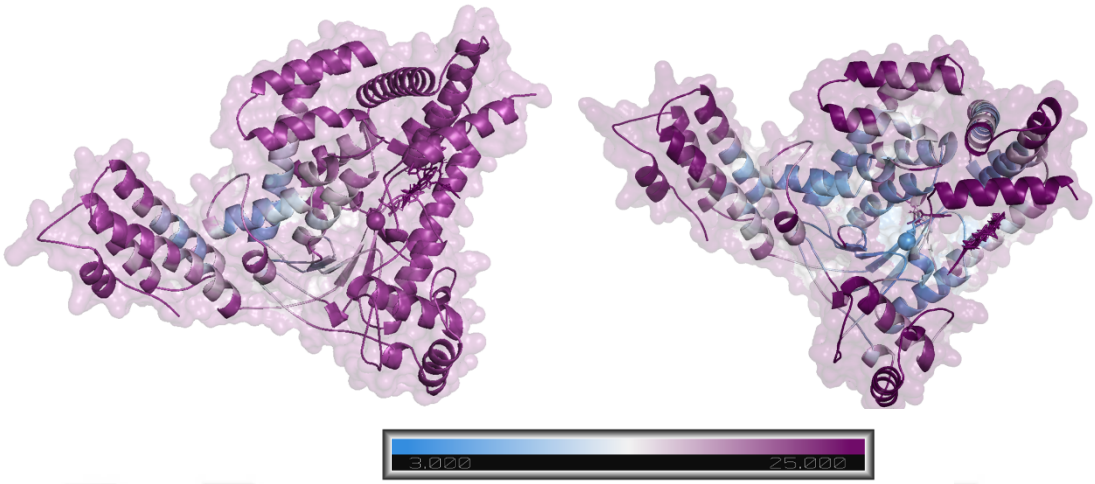


Figure 15. B-factor visualization of TcdB in complex with cyclic peptides. Left: TcdB bound to Peptide 2. Right: TcdB bound to Peptide 15. The color scale represents atomic displacement parameters (B-factors), where blue indicates low flexibility and white/pink indicates higher mobility. Protein structures are shown in cartoon and surface representation (from 3.02 to 25.00).

4. CONCLUSIONS AND FUTURE WORK

In this thesis, a multi-layered *in silico* approach was employed to understand selective interactions with host proteins and identify and characterize potential inhibitors targeting *C. difficile* Toxin B (TcdB), a central virulence factor responsible for epithelial damage and inflammatory signalling in *C. difficile* infection (CDI). The study focused on the rational design and evaluation of both small-molecule ligands and cyclic peptide inhibitors, aiming to disrupt TcdB's glucosyltransferase activity on host proteins.

Through structure-based virtual screening of over 14 million drug-like compounds from the ZINC database, multiple small-molecule candidates were identified with high predicted binding affinities for the UDP-glucose binding pocket of TcdB's glucosyltransferase domain (GTD). Detailed binding mode analyses revealed strong hydrogen bonding and hydrophobic interactions with key catalytic residues of TcdB—Asp2865, Glu449, and His448—suggesting that these molecules may act as potential competitive inhibitors by occupying the same site as the native substrate, UDP-glucose. Pharmacokinetic evaluation further demonstrated that all top candidates complied with Lipinski's Rule of Five and exhibited favourable ADME/T profiles, including high gastrointestinal absorption, absence of major toxicity alerts, and negative PAINS/Brenk filters.

Nevertheless, a major limitation of this study lies in the inability to perform molecular dynamics (MD) simulations for these small-ligand–protein complexes. Initial simulations failed to equilibrate, likely due to steric clashes caused by rigid docking poses within the constrained active site. As such, key dynamic properties—such as RMSD, RMSF, hydrogen bonding persistence, and MM-GBSA free energy estimates—could not be obtained. Therefore, while competitive inhibition appears structurally plausible, further validation is required to confirm binding stability and physiological relevance.

In a complementary strategy, therapeutic cyclic peptides were designed to block TcdB's interaction with host GTPases. Based on interface epitope mapping, homology filtering, and toxicity screening, Peptides 2 and 15 were selected for full 500 ns MD simulations. Both peptides demonstrated stable and specific binding to the toxin, supported by low RMSD fluctuations, consistent hydrogen bonding profiles, and favourable MM-PBSA binding free energies. Moreover, these peptides exhibited a dual inhibitory potential by simultaneously engaging the protein–protein interaction interface and partially obstructing the UDP-glucose binding region—thereby providing a more comprehensive inhibitory mechanism. Toxicity predictions confirmed the safety of these peptides, making them strong candidates for further therapeutic development.

Taken together, the data generated in this study support the viability of targeting TcdB's enzymatic and interface domains using rationally designed small molecules and cyclic peptides. The small molecules offer a faster route toward drug-like optimization, while the peptides provide highly specific and stable binding profiles. Both strategies present promising alternatives to traditional antibiotic therapies, especially considering their potential to preserve gut microbiota and minimize recurrence rates.

To further advance the therapeutic applications of this research, the following steps are recommended:

1. **Molecular Dynamics Simulations for Small Molecules**: Redocking using flexible docking protocols followed by full MD simulations should be undertaken to resolve steric clashes and assess ligand stability over time.
2. **Binding Free Energy Estimation**: Methods such as MM-GBSA or free energy perturbation (FEP) should be applied to the docked poses to provide accurate comparative binding energy profiles.
3. ***In Vitro* Validation**: Functional assays—such as glucosyltransferase inhibition, TcdB–GTPase binding interference, and cellular cytotoxicity tests—should be conducted to confirm the computational predictions.

4. **Peptide Optimization**: Further modifications (e.g., PEGylation, sequence stabilization, or incorporation of non-natural amino acids) can be explored to enhance bioavailability and resistance to proteolysis.
5. **Variant Coverage Assessment**: The binding efficiency of both small molecules and peptides should be tested computationally against different clinically relevant TcdB variants to ensure broad-spectrum potential.

In conclusion, this work establishes a strong computational foundation for the development of non-antibiotic therapies that directly neutralize TcdB function. By selectively disrupting toxin activity at the molecular level, these inhibitors may provide safer, microbiota-sparing, and resistance-resilient treatment options for CDI.

5. REFERENCES

1. Kelesidis T, Pothoulakis C. Clostridioides difficile infection. Gut Instincts: A Clinician's Handbook of Digestive and Liver Diseases [Internet]. 2024 Apr 10 [cited 2025 May 28];179–85. Available from: https://www.ncbi.nlm.nih.gov/books/NBK431054/
2. Klimko A, Tieranu CG, Curte AM, Preda CM, Tieranu I, Olteanu AO, et al. Clostridioides Difficile Enteritis: Case Report and Literature Review. Antibiotics [Internet]. 2022;11(2). Available from: https://doi.org/10.3390/antibiotics11020206
3. Zanella Terrier MC, Simonet ML, Bichard P, Frossard JL. Recurrent Clostridium difficile infections: The importance of the intestinal microbiota. World Journal of Gastroenterology : WJG [Internet]. 2014 Jun 6 [cited 2022 Oct 25];20(23):7416. Available from: /pmc/articles/PMC4064086/
4. Centers for Disease Control and Prevention. Vol. 027, Centers for Disease Control and Prevention. 2019 [cited 2022 Oct 25]. p. 2017–9 2017 Annual Report for the Emerging Infections Program for Clostridioides difficile Infection. Available from: https://www.cdc.gov/hai/eip/pdf/cdiff/2017-CDI-Report-H.pdf
5. Voth DE, Ballard JD. Clostridium difficile Toxins: Mechanism of Action and Role in Disease. Clin Microbiol Rev [Internet]. 2005 [cited 2025 May 28];18(2):247. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC1082799/
6. Carter GP, Rood JI, Lyras D. The role of toxin A and toxin B in the virulence of Clostridium difficile. Trends Microbiol [Internet]. 2012 Jan 1 [cited 2025 May 28];20(1):21–9. Available from: https://www.sciencedirect.com/science/article/abs/pii/S0966842X1100196X
7. Schwan C, Stecher B, Tzivelekidis T, Van Ham M, Rohde M, Hardt WD, et al. Clostridium difficile toxin CDT induces formation of microtubule-

based protrusions and increases adherence of bacteria. PLoS Pathog [Internet]. 2009 Oct [cited 2025 May 28];5(10). Available from: https://pubmed.ncbi.nlm.nih.gov/19834554/

8. Lionta E, Spyrou G, Vassilatis D, Cournia Z. Structure-Based Virtual Screening for Drug Discovery: Principles, Applications and Recent Advances. Curr Top Med Chem [Internet]. 2014 Oct 1 [cited 2025 May 28];14(16):1923–38. Available from: https://pubmed.ncbi.nlm.nih.gov/25262799/
9. Lounnas V, Ritschel T, Kelder J, McGuire R, Bywater RP, Foloppe N. Current progress in Structure-Based Rational Drug Design marks a new mindset in drug discovery. Comput Struct Biotechnol J [Internet]. 2013 [cited 2025 May 28];5(6):e201302011. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC3962124/
10. Gupta S, Kapoor P, Chaudhary K, Gautam A, Kumar R, Raghava GPS. In Silico Approach for Predicting Toxicity of Peptides and Proteins. PLoS One [Internet]. 2013 Sep 13 [cited 2025 May 28];8(9). Available from: https://pubmed.ncbi.nlm.nih.gov/24058508/
11. Sharma N, Naorem LD, Jain S, Raghava GPS. ToxinPred2: An improved method for predicting toxicity of proteins. Brief Bioinform [Internet]. 2022 Sep 1 [cited 2025 May 28];23(5). Available from: https://pubmed.ncbi.nlm.nih.gov/35595541/
12. Lionta E, Spyrou G, Vassilatis D, Cournia Z. Structure-Based Virtual Screening for Drug Discovery: Principles, Applications and Recent Advances. Curr Top Med Chem [Internet]. 2014 Oct 1 [cited 2025 May 28];14(16):1923–38. Available from: https://pubmed.ncbi.nlm.nih.gov/25262799/
13. Durrant JD, McCammon JA. Molecular dynamics simulations and drug discovery. BMC Biol [Internet]. 2011 Oct 28 [cited 2025 May 28];9(1):1–9. Available from: https://bmcbiol.biomedcentral.com/articles/10.1186/1741-7007-9-71

14. Pruitt RN, Chambers MG, Ng KKS, Ohi MD, Lacy DB. Structural organization of the functional domains of Clostridium difficile toxins A and B. Proc Natl Acad Sci U S A [Internet]. 2010;107(30):13467–72. Available from: www.pnas.org/cgi/doi/10.1073/pnas.1002199107
15. Reinert DJ, Jank T, Aktories K, Schulz GE. Structural basis for the function of Clostridium difficile toxin B. J Mol Biol. 2005;351(5):973–81.
16. Jank T, Aktories K. Structure and mode of action of clostridial glucosylating toxins: the ABCD model. Trends Microbiol [Internet]. 2008 May 1 [cited 2025 May 28];16(5):222–9. Available from: https://www.sciencedirect.com/science/article/abs/pii/S0966842X08000607
17. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. Nucleic Acids Res [Internet]. 2000 Jan 1 [cited 2022 Jun 1];28(1):235–42. Available from: https://academic.oup.com/nar/article/28/1/235/2384399
18. Camli DN, Iscil HAO, Acuner SE. MuGger Toxins: Exploring the Selective Binding Mechanism of Clostridial Glucosyltransferase Toxin B and Host GTPases. Proteins: Structure, Function and Bioinformatics [Internet]. 2024 Apr 1 [cited 2025 May 28];93(4). Available from: https://pubmed.ncbi.nlm.nih.gov/39670652/
19. Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. ScienceDirect GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. 2015 [cited 2024 May 4]; Available from: www.sciencedirect.comwww.elsevier.com/locate/softx
20. Huang J, Rauscher S, Nawrocki G, Ran T, Feig M, De Groot BL, et al. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. Nature Methods 2016 14:1 [Internet]. 2016 Nov 7 [cited 2025 May 28];14(1):71–3. Available from: https://www.nature.com/articles/nmeth.4067
21. Irwin JJ, Tang KG, Young J, Dandarchuluun C, Wong BR, Khurelbaatar M, et al. ZINC20 - A Free Ultralarge-Scale Chemical Database for Ligand

Discovery. J Chem Inf Model [Internet]. 2020 Dec 28 [cited 2024 Oct 25];60(12):6065–73. Available from: https://pubs.acs.org/doi/full/10.1021/acs.jcim.0c00675

22. Tribello GA, Bonomi M, Branduardi D, Camilloni C, Bussi G. PLUMED 2: New feathers for an old bird. Comput Phys Commun. 2014 Feb 1;185(2):604–13.
23. Sayers EW, Beck J, Bolton EE, Brister JR, Chan J, Connor R, et al. Database resources of the National Center for Biotechnology Information in 2025. Nucleic Acids Res [Internet]. 2024 [cited 2025 May 28];53(D1):D20. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC11701734/
24. Ponder JW, Case DA. FORCE FIELDS FOR PROTEIN SIMULATIONS. 2003;
25. Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and scoring in virtual screening for drug discovery: methods and applications. Nature Reviews Drug Discovery 2004 3:11 [Internet]. 2004 Nov [cited 2025 May 28];3(11):935–49. Available from: https://www.nature.com/articles/nrd1549
26. Viprey VF, Davis GL, Benson AD, Ewin D, Spittal W, Vernon JJ, et al. A point-prevalence study on community and inpatient Clostridioides difficile infections (CDI): Results from Combatting Bacterial Resistance in Europe CDI (COMBACTE-CDI), July to November 2018. Eurosurveillance [Internet]. 2022 [cited 2022 Oct 24];27(26). Available from: www.eurosurveillance.org
27. Coffing H, Priyamvada S, Anbazhagan AN, Salibay C, Engevik M, Versalovic J, et al. Clostridium difficile toxins A and B decrease intestinal SLC26A3 protein expression. Am J Physiol Gastrointest Liver Physiol [Internet]. 2018 [cited 2023 May 9];315:43–52. Available from: http://www.ajpgi.org
28. Lyras D, O'Connor JR, Howarth PM, Sambol SP, Carter GP, Phumoonna T, et al. Toxin B is essential for virulence of Clostridium difficile. Nature

2009 458:7242 [Internet]. 2009 Mar 1 [cited 2022 May 12];458(7242):1176–9. Available from: https://www.nature.com/articles/nature07822

29. Genth H, Huelsenbeck J, Hartmann B, Hofmann F, Just I, Gerhard R. Cellular stability of Rho-GTPases glucosylated by Clostridium difficile toxin B. FEBS Lett [Internet]. 2006 Jun 12 [cited 2023 Jan 19];580(14):3565–9. Available from: https://pubmed.ncbi.nlm.nih.gov/16730714/
30. He R, Peng J, Yuan P, Yang J, Wu X, Wang Y, et al. Glucosyltransferase Activity of Clostridium difficile Toxin B Triggers Autophagy-mediated Cell Growth Arrest. Sci Rep [Internet]. 2017;7(1):1–14. Available from: http://dx.doi.org/10.1038/s41598-017-11336-4
31. Willey, Sherwood, Woolverton. Prescott's Microbiology. 9th ed. Vol. 53, Journal of Chemical Information and Modeling. 2013. 1689–1699 p.
32. Madigan Michael T, Martinko John M., Bender Kelly S., Buckley Daniel h. SDA. Brock Biology of Microoranisms. Fourteenth. 2015. 151–153 p.
33. Bäckhed F, Fraser CM, Ringel Y, Sanders ME, Sartor RB, Sherman PM, et al. Defining a healthy human gut microbiome: Current concepts, future directions, and clinical applications. Cell Host Microbe. 2012;12(5):611–22.
34. Holzer P, Farzi A. Microbiome, HPA Axis and Production of Endocrine Hormones in the Gut | SpringerLink. 2018. 1–21 p.
35. Kho ZY, Lal SK. The human gut microbiome - A potential controller of wellness and disease. Front Microbiol. 2018 Aug 14;9(AUG):1835.
36. CDC - Centers for Disease Control and. What is C. diff? | CDC [Internet]. 2021 [cited 2022 May 3]. Available from: https://www.cdc.gov/cdiff/what-is.html
37. Chen P, Lam K ho, Liu Z, Mindlin FA, Chen B, Gutierrez CB, et al. Structure of the full-length Clostridium difficile toxin B. Nat Struct Mol Biol. 2019 Aug 1;26(8):712–9.
38. Fletcher JR, Pike CM, Parsons RJ, Rivera AJ, Foley MH, McLaren MR, et al. Clostridioides difficile exploits toxin-mediated inflammation to alter

the host nutritional landscape and exclude competitors from the gut microbiota. Nat Commun [Internet]. 2021;12(1):1–14. Available from: http://dx.doi.org/10.1038/s41467-020-20746-4

39. Gerhard R, Nottrott S, Schoentaube J, Tatge H, Oiling A, Just I. Glucosylation of Rho GTPases by Clostridium difficile toxin A triggers apoptosis in intestinal epithelial cells. J Med Microbiol. 2008;57(6):765–70.
40. Nakao S, Hasegawa S, Shimada K, Mukai R, Tanaka M, Matsumoto K, et al. Evaluation of anti-infective-related clostridium difficile-associated colitis using the Japanese adverse drug event report database. Int J Med Sci [Internet]. 2020 [cited 2022 May 4];17(7):921–30. Available from: /pmc/articles/PMC7163355/
41. Orth P, Xiao L, Hernandez LD, Reichert P, Sheth PR, Beaumont M, et al. Mechanism of action and epitopes of Clostridium difficile toxin B-neutralizing antibody bezlotoxumab revealed by X-ray crystallography. Journal of Biological Chemistry [Internet]. 2014;289(26):18008–21. Available from: http://dx.doi.org/10.1074/jbc.M114.560748
42. Pourliotopoulou E, Karampatakis T, Kachrimanidou M. Exploring the Toxin-Mediated Mechanisms in Clostridioides difficile Infection. Microorganisms 2024, Vol 12, Page 1004 [Internet]. 2024 May 16 [cited 2025 May 28];12(5):1004. Available from: https://www.mdpi.com/2076-2607/12/5/1004/htm
43. Smits WK, Lyras D, Lacy DB, Wilcox MH, Kuijper EJ. Clostridium difficile infection. Nat Rev Dis Primers. 2016 Apr 7;2:1–20.
44. Just I, Selzer J, Wilm M, Eichel-Streiber C Von, Mann M, Aktories K. Glucosylation of Rho proteins by Clostridium difficile toxin B. Nature [Internet]. 1995 Jun 8 [cited 2023 Jan 18];375(6531):500–3. Available from: https://pubmed.ncbi.nlm.nih.gov/7777059/
45. Liu Z, Zhang S, Chen P, Tian S, Zeng J, Perry K, et al. Structural basis for selective modification of Rho and Ras GTPases by Clostridioides difficile toxin B. Sci Adv [Internet]. 2021 Oct 1 [cited 2022 Jun 8];7(43):1–14. Available from: https://www.science.org/doi/full/10.1126/sciadv.abi4582

46. Qu L, Pan C, He SM, Lang B, Gao GD, Wang XL, et al. The ras superfamily of small gtpases in non-neoplastic cerebral diseases. Front Mol Neurosci. 2019 May 27;12:121.
47. Petersen L, Stroh S, Schöttelndreier D, Grassl GA, Rottner K, Brakebusch C, et al. The Essential Role of Rac1 Glucosylation in Clostridioides difficile Toxin B-Induced Arrest of G1-S Transition. Front Microbiol. 2022 Mar 7;13:662.
48. Tao L, Zhang J, Meraner P, Tovaglieri A, Wu X, Gerhard R, et al. Frizzled proteins are colonic epithelial receptors for C. difficile toxin B. Nature [Internet]. 2016 Sep 28 [cited 2022 May 12];538(7625):350–5. Available from: https://www.nature.com/articles/nature19799
49. Chen P, Tao L, Wang T, Zhang J, He A, Lam K ho, et al. Structural basis for recognition of frizzled proteins by clostridium difficile toxin B. Science (1979). 2018;360(6389):664–9.
50. Burke KE, Lamont JT. Clostridium difficile infection: A worldwide disease. Gut Liver. 2014 Jan;8(1):1–6.
51. Zhang P, Hong J, Yoon IN, Kang JK, Hwang JS, Kim H. Clostridium difficile Toxin A Induces Reactive Oxygen Species Production and p38 MAPK Activation to Exert Cellular Toxicity in Neuronal Cells. J Microbiol Biotechnol [Internet]. 2017 Jun 28 [cited 2025 May 28];27(6):1163–70. Available from: https://www.jmb.or.kr/journal/view.html?doi=10.4014/jmb.1702.02041
52. Alam MZ, Madan R. Clostridioides difficile Toxins: Host Cell Interactions and Their Role in Disease Pathogenesis. Toxins (Basel) [Internet]. 2024 Jun 1 [cited 2025 May 28];16(6):241. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC11209539/
53. Chumbler NM, Farrow MA, Lapierre LA, Franklin JL, Haslam D, Goldenring JR, et al. Clostridium difficile Toxin B Causes Epithelial Cell Necrosis through an Autoprocessing-Independent Mechanism. PLoS Pathog [Internet]. 2012 Dec [cited 2025 May 28];8(12). Available from: https://pubmed.ncbi.nlm.nih.gov/23236283/

54. Chumbler NM, Farrow MA, Lapierre LA, Franklin JL, Haslam D, Goldenring JR, et al. Clostridium difficile Toxin B Causes Epithelial Cell Necrosis through an Autoprocessing-Independent Mechanism. PLoS Pathog [Internet]. 2012 Dec [cited 2025 May 28];8(12). Available from: https://pubmed.ncbi.nlm.nih.gov/23236283/
55. Lan N, Ashburn J, Shen B. Fecal microbiota transplantation for Clostridium difficile infection in patients with ileal pouches. Gastroenterol Rep (Oxf) [Internet]. 2017 Aug 1 [cited 2024 Oct 25];5(3):200. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC5554389/
56. Chen P, Zeng J, Liu Z, Thaker H, Wang S, Tian S, et al. Structural basis for CSPG4 as a receptor for TcdB and a therapeutic target in Clostridioides difficile infection. Nat Commun [Internet]. 2021 Dec 1 [cited 2022 Jun 8];12(1). Available from: https://pubmed.ncbi.nlm.nih.gov/34145250/
57. Chen P, Lam K ho, Liu Z, Mindlin FA, Chen B, Gutierrez CB, et al. Structure of the full-length Clostridium difficile toxin B. Nat Struct Mol Biol [Internet]. 2019 Aug 1 [cited 2025 Jun 2];26(8):712–9. Available from: https://pubmed.ncbi.nlm.nih.gov/31308519/
58. Liu Z, Zhang S, Chen P, Tian S, Zeng J, Perry K, et al. Structural basis for selective modification of Rho and Ras GTPases by Clostridioides difficile toxin B. Sci Adv [Internet]. 2021 Oct 1 [cited 2025 Jun 2];7(43):eabi4582. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC8535798/
59. DeepMind, EMBL-EBI. AlphaFold Protein Structure Database [Internet]. 2021 [cited 2022 May 15]. Available from: https://alphafold.ebi.ac.uk/
60. Bateman A, Martin MJ, Orchard S, Magrane M, Agivetova R, Ahmad S, et al. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res [Internet]. 2021 Jan 8 [cited 2022 Jun 1];49(D1):D480–9. Available from: https://academic.oup.com/nar/article/49/D1/D480/6006196
61. Laskowski RA, Jabłońska J, Pravda L, Vařeková RS, Thornton JM. PDBsum: Structural summaries of PDB entries. Protein Science [Internet]. 2018 Jan 1 [cited 2022 Jun 1];27(1):129–34. Available from: https://pubmed.ncbi.nlm.nih.gov/28875543/

62. Schrödinger. LLC, The PyMOL Molecular Graphics System [Internet]. 2021 p. Version 1.2r3pre. Available from: https://pymol.org/2/
63. Durmuş Tekir S, Çakir T, Ardiç E, Sayilirbaş AS, Konuk G, Konuk M, et al. Vol. 29, Bioinformatics. Oxford Academic; 2013 [cited 2022 May 9]. p. 1357–8 PHISTO: Pathogen-host interaction search tool. Available from: https://phisto.org/search.xhtml
64. Van Zundert GCP, Rodrigues JPGLM, Trellet M, Schmitz C, Kastritis PL, Karaca E, et al. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. J Mol Biol. 2016 Feb 22;428(4):720–5.
65. Meng EC, Goddard TD, Pettersen EF, Couch GS, Pearson ZJ, Morris JH, et al. UCSF ChimeraX: Tools for structure building and analysis. Protein Science [Internet]. 2023 Nov 1 [cited 2024 May 2];32(11):e4792. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/pro.4792
66. Inkscape Project. Inkscape. 2020.
67. Jo S, Kim T, Iyer VG, Im W. CHARMM-GUI: A web-based graphical user interface for CHARMM. J Comput Chem. 2008 Aug;29(11):1859–65.
68. GraphPad Software. GraphPad Prism version 10.0.2 for Windows. Boston, Massachusetts, USA: GraphPad Software; 2023.
69. Rogers D, Hahn M. Extended-connectivity fingerprints. J Chem Inf Model [Internet]. 2010 May 24 [cited 2025 Jun 2];50(5):742–54. Available from: /doi/pdf/10.1021/ci100050t
70. Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. J Comput Chem [Internet]. 2010 Jan 30 [cited 2025 May 29];31(2):455. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC3041641/
71. Bugnon M, Röhrig UF, Goullieux M, Perez MAS, Daina A, Michielin O, et al. SwissDock 2024: major enhancements for small-molecule docking with Attracting Cavities and AutoDock Vina. Nucleic Acids Res [Internet].

2024 Jul 5 [cited 2025 May 29];52(W1):W324–32. Available from: https://dx.doi.org/10.1093/nar/gkae300

72. Bhardwaj G, Mulligan VK, Bahl CD, Gilmore JM, Harvey PJ, Cheneval O, et al. Accurate de novo design of hyperstable constrained peptides. Nature 2016 538:7625 [Internet]. 2016 Sep 14 [cited 2025 Jun 2];538(7625):329–35. Available from: https://www.nature.com/articles/nature19791
73. Yan Y, Tao H, He J, Huang SY. The HDOCK server for integrated protein–protein docking. Nat Protoc [Internet]. 2020 May 1 [cited 2025 May 29];15(5):1829–52. Available from: https://www.nature.com/articles/s41596-020-0312-x
74. Hanwell MD, Curtis DE, Lonie DC, Vandermeerschd T, Zurek E, Hutchison GR. Avogadro: An advanced semantic chemical editor, visualization, and analysis platform. J Cheminform [Internet]. 2012 Aug 13 [cited 2025 Jun 2];4(8):1–17. Available from: https://jcheminf.biomedcentral.com/articles/10.1186/1758-2946-4-17
75. Wang J, Wang W, Kollman PA, Case DA. Antechamber, An Accessory Software Package For Molecular Mechanical Calculations Correspondence to. J Chem Inf Comput Sci.
76. Gordon JC, Myers JB, Folta T, Shoja V, Heath LS, Onufriev A. H++: a server for estimating pKas and adding missing hydrogens to macromolecules. Nucleic Acids Res [Internet]. 2005 Jul [cited 2025 May 29];33(Web Server issue):W368. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC1160225/
77. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. Nature [Internet]. 2021 Jul 15 [cited 2022 Jun 1];596(7873):583–9. Available from: https://www.nature.com/articles/s41586-021-03819-2
78. Lipinski CA. Lead- and drug-like compounds: The rule-of-five revolution. Drug Discov Today Technol. 2004 Dec;1(4):337–41.
79. Agarwala R, Barrett T, Beck J, Benson DA, Bollin C, Bolton E, et al. Database resources of the National Center for Biotechnology Information.

Nucleic Acids Res [Internet]. 2016 Jan 1 [cited 2022 Jun 1];44(D1):D7–19. Available from: /pmc/articles/PMC4702911/

80. Baell JB, Holloway GA. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. J Med Chem [Internet]. 2010 Apr 8 [cited 2025 May 31];53(7):2719–40. Available from: /doi/pdf/10.1021/jm901137j

81. Bickerton GR, Paolini G V., Besnard J, Muresan S, Hopkins AL. Quantifying the chemical beauty of drugs. Nat Chem [Internet]. 2012 Feb 24 [cited 2025 May 30];4(2):90–8. Available from: https://www.nature.com/articles/nchem.1243

82. Genheden S, Ryde U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. Expert Opin Drug Discov [Internet]. 2015 May 1 [cited 2025 May 31];10(5):449–61. Available from: https://pubmed.ncbi.nlm.nih.gov/25835573/

APPENDIX

Table A1. Available Toxin B Structures

Identifier	Method	Resolution (Å)	Positions
PDB-2BVL	x-ray	2.20	2-541
PDB-2BVM	x-ray	2.55	2-542
PDB-4NC2	x-ray	2.50	2248-2366
PDB-4NP4	x-ray	2.89	1834-2099
PDB-5UQM	x-ray	2.03	1-543
PDB-5UQN	x-ray	2.06	1-543
PDB-5UQT	x-ray	2.75	1-543
PDB-6AR6	EM	9.00	4-2099
PDB-6C0B	x-ray	2.50	1285-1804
PDB-6OQ5	x-ray	3.87	1-2366
PDB-6OQ6	x-ray	2.97	1071-1432
PDB-6OQ7	x-ray	2.39	1-543
PDB-6OQ8	x-ray	2.20	1-542
PDB-7LOU	x-ray	1.82	2-543
PDB-7LOV	x-ray	2.50	2-545
PDB-7ML7	EM	3.17	1-1967
PDB-7N8X	EM	3.40	374-1876
PDB-7N95	EM	4.10	1-2366
PDB-7N97	EM	5.10	2-2366
PDB-7N9Q	EM	4.60	2-2366
PDB-7N9R	EM	5.90	1-2366
PDB-7N9S	EM	5.10	2-2366
PDB-7N9Y	EM	4.80	2-2366
PDB-7S0Y	x-ray	2.79	1-540
PDB-7SO5	x-ray	1.80	3-543

PDB-7SO7	x-ray	3.59	1-538
PDB-7S0Z	x-ray	2.34	1-541
PDB-7V1N	EM	3.20	1-2367
AlphaFold-AF-P18177-F1	Predicted	-	1-2366
AlphaFold-AF-Q9EXR0-F1	Predicted	-	1-2367

Table A2. TcdB interface residues in experimental complex structures

PDB ID	Molecule	TcdB interaction residues
2BVL	TBR	29,32,33,36
4NC2	B39 VHH	12,15,16,17,22,23,38,39,41,43,45,51,68,72,73,74,75,76
4NP4	Bezlo-toxumab	2009,2018,2021,2031,2032,2033,2034,2036,2070,2074,2075, 2076,2077,2092,2093,2033,2034,2070,2072,2074
5UQM 5UQN	U2F	101,102,103,139,265,266,269,270,273,284,286,287,288,383, 384,385,465,470,471,515,518,519,520
5UQT	Apigenin	173,425,426,428,429,430,432,433,436,448,455
6C0B	Frizzled2	1434,1437,1438,1440,1468,1488,1490,1491,1495,1501,1504, 1505,1506,1509,1511,1597,1598,1599
6OQ6	VHH 5D	1107,1110,1112,1305,1306,1307,1308,1310,1311,1313,1330, 1331,1332,1356,1387
6OQ7	VHH E3	22,23,25,26,29,30,33,48,51,54,55,58,61,63
6OQ8	VHH 7F	144,147,148,150,151,152,154,155,158,159,214,215,534,537, 538,540,541,542
7LOU	IFM	266,270,273,286,384,385,465,470,520
7LOV	NOY	266,270,273,286,383,384,385,465,470,471,520
7ML7	CSPG4	563,564,566,567,571,575,603,621,623,624,626,1754,1758,17 60,1808,1809,1810,1811,1812,1818,1819,1825,1829,1831,18 48,1849,1850

7N8X	CSPG4	563,564,567,568,571,603,621,623,624,626,1754,1755,1759,1760,1808,1810,1811,1812,1815,1818,1819,1825,1829,1831,1848,1850
7S0Y	Cdc42	173,310,314,378,379,380,381,382,383,429,432,433,436,439,441,444,447,448,451,452,455,459,462,463,466,471,472,475,493,494,495,498,515,516,518,520,521
7S0Z	R-Ras	1,2,3,4,7,8,11,12,73,74,77,78,84,85,173,380,382,383,437,446,450,456,473,492,493,495,496,497,498,501,517
7SO5	Fab-B2	17,18,19,20,21,22,23,25,26,29,58,62,63
7SO7	Fab-B1	93,95,96,97,123,124,125,126,128,130,238,243,280,364,365,367,391
7V1N	TFPI	1433,1434,1435,1439,1465,1489,1492,1494,1510,1512,1598,1599

*TBR (Hexatantalum Dodecabromide), U2F (Uridine-5'-Diphosphate-2-Deoxy-2-Fluoro-Alpha-D-Glucose), IFM (Isofagomine), NOY (Noeuromycin), CSPG4 (Chondroitin Sulphate Proteoglycan 4), TFPI (Isoform Beta of Tissue factor pathway inhibitor)

**Ligands are shown brown colour, monoclonal antibodies (mAbs) are shown green colour, and inhibitor are shown orange color.

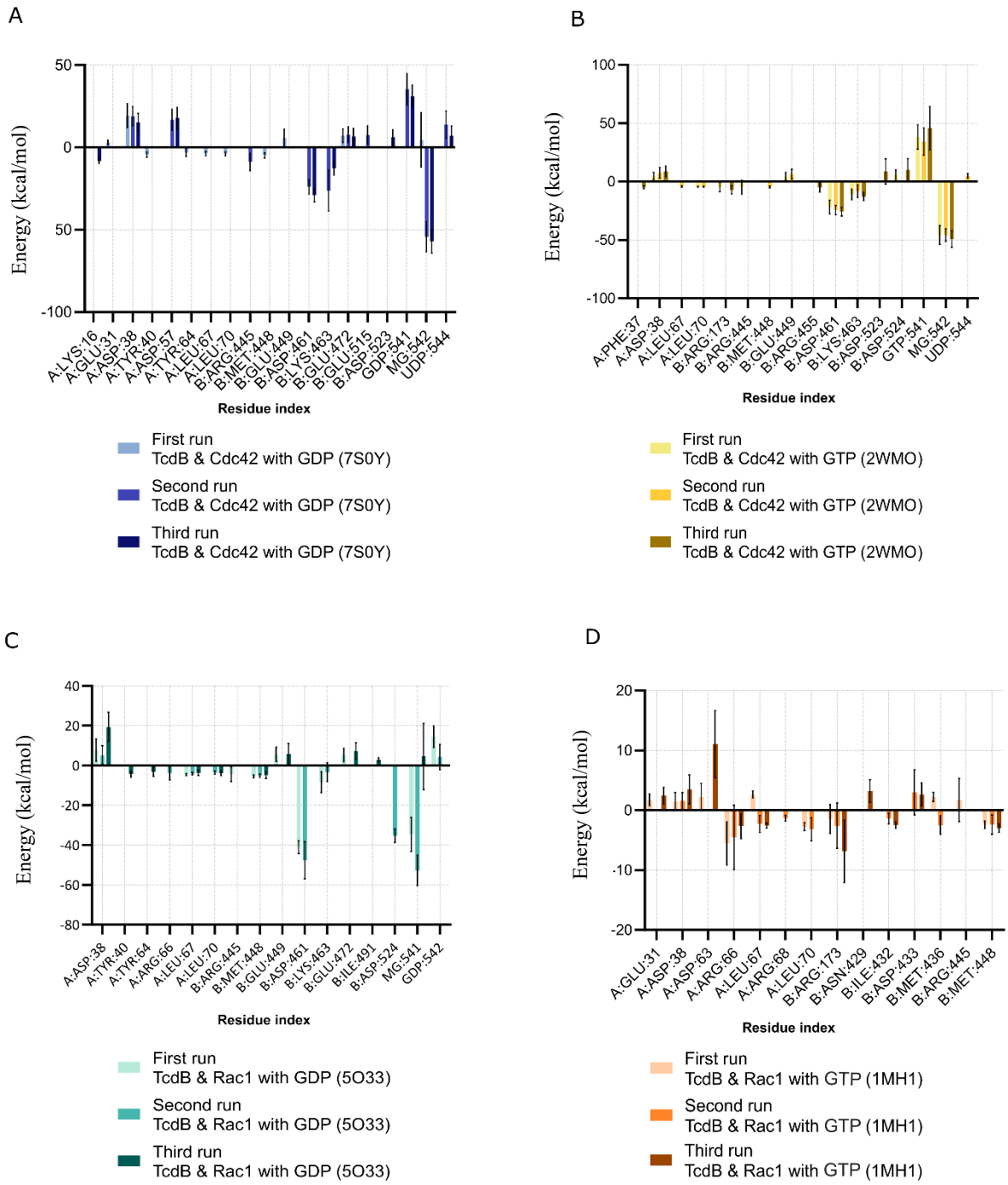


Figure A1. Decomposition analysis of top ten residues contributing to the total interaction energy.

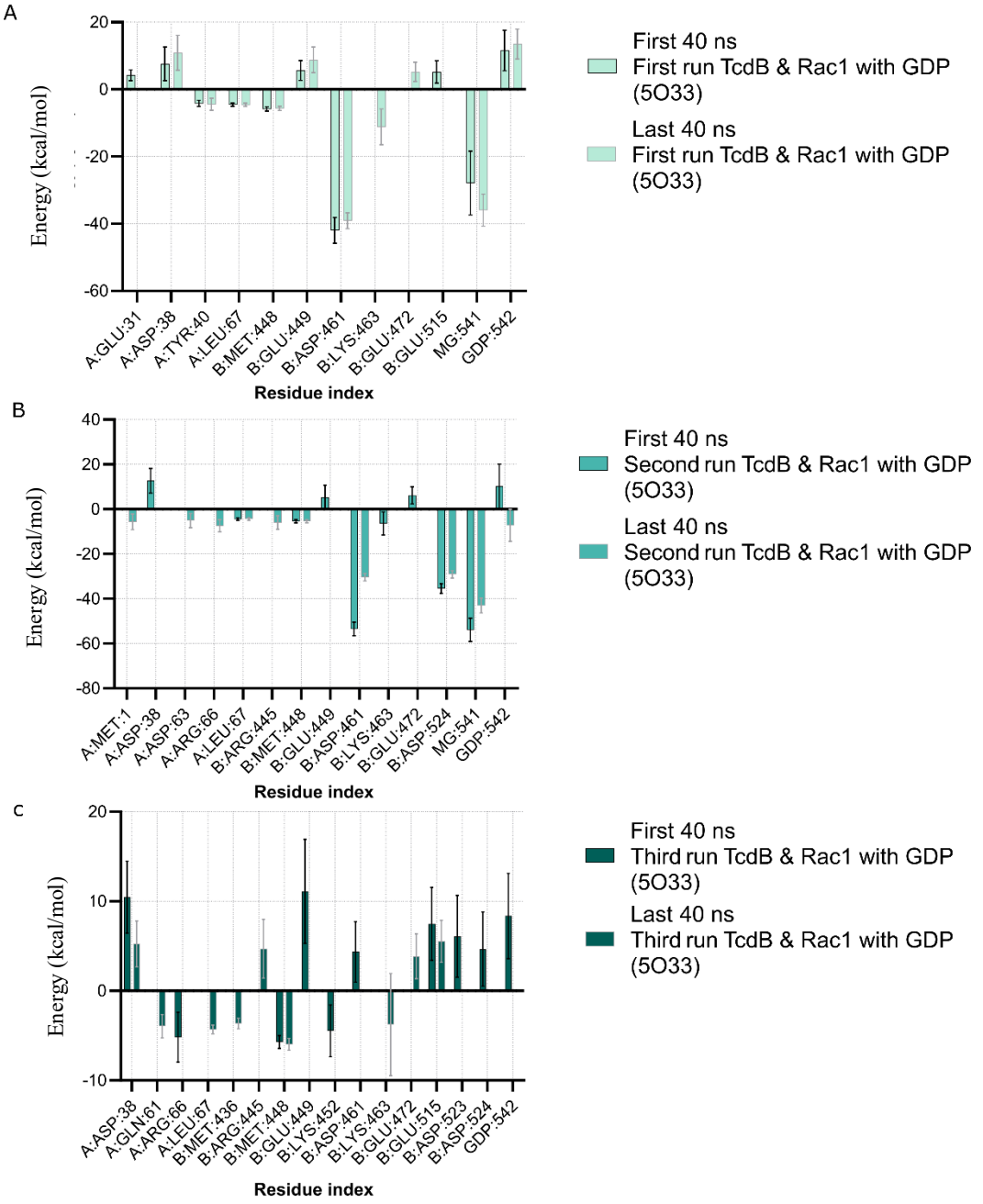


Figure A2. Per-residue decomposition analysis of top ten residues in GDP-bound Rac1 in the first and last 40ns of each parallel run.