

## Targeting a key virulence factor in *Listeria monocytogenes*: An *in silico* discovery and pharmacokinetic profiling of novel internalin H inhibitors

Ayush Kumar Dewangan<sup>a</sup>, Usha Chouhan<sup>b,\*</sup>, Deedhiti Mistry<sup>b</sup>

<sup>a</sup> Amity Institute of Biotechnology, Amity University Chhattisgarh, Raipur, India

<sup>b</sup> Department of Mathematics, Bioinformatics and Computer Applications, Maulana Azad National Institute of Technology, Bhopal, Madhya Pradesh, India

### ARTICLE INFO

**Keywords:**

Drug design  
Computational chemistry  
Internalin H  
*Listeria monocytogenes*  
Molecular dynamics  
Anti-virulence therapy

### ABSTRACT

*Listeria monocytogenes* is a lethal food-borne bacterium with high mortality rates (20–30%) and growing antimicrobial resistance, highlighting the critical need for novel therapeutic interventions. This study explores an anti-virulence therapy neutralizing the bacterium by disrupting the Internalin H (InlH) surface protein, a major determinant of immune evasion, providing an alternative substitute to conventional antibiotics.

To validate and rank the potential candidates, a multi-layered *in silico* workflow was employed. This approach integrated high-throughput virtual screening (HTVS) and ADMET profiling with advanced validation techniques, including 125 ns Molecular Dynamics (MD) simulations, MM/GBSA free energy calculations, and Density Functional Theory (DFT) analysis. HTVS successfully identified triterpenoid analogues as top hits; however, the subsequent advanced validation corrected the initial docking rank, highlighting the limitations of static scoring.

The Hederagenin analogue (CID 137633443) was definitively identified as the most potent inhibitor, exhibiting a superior binding free energy ( $\Delta G_{bind}$ ) of  $-17.09\text{kcal/mol}$ . This potency was confirmed by its dynamic stability (low RMSD and a deep Free Energy Landscape basin) and its high chemical reactivity, as DFT analysis showed the Hederagenin analogue had the smallest HOMO-LUMO energy gap (4.92 eV).

This research successfully identifies a structurally and electronically validated small-molecule lead, offering a strong foundation for the development of novel anti-virulence therapies to combat listeriosis and address the increasing challenge of antibiotic resistance.

### Introduction

*Listeria monocytogenes* is a robust, Gram-positive bacterium ubiquitous in nature, known worldwide as a foodborne pathogen [1]. The bacterium stands out due to its ability to thrive, even grow in unfavorable conditions such as cold temperatures, low-pH environments, and moderate salt levels, thus posing a constant risk in food production [2]. The resulting infection, called listeriosis, is a dangerous and life-threatening disease. Though rarer than others, listeriosis has extremely high hospitalization and mortality rates, up to 20–30 % [3]. It is particularly perilous to immunocompromised individuals, as well as to pregnant women, newborn infants, older adults, and those suffering from weakened immunity [4,5]. In these individuals, it can cause life-threatening conditions such as meningitis, bacteraemia, and miscarriage, and can even lead to death in some cases, highlighting the critical need to formulate efficacious therapeutic interventions [6].

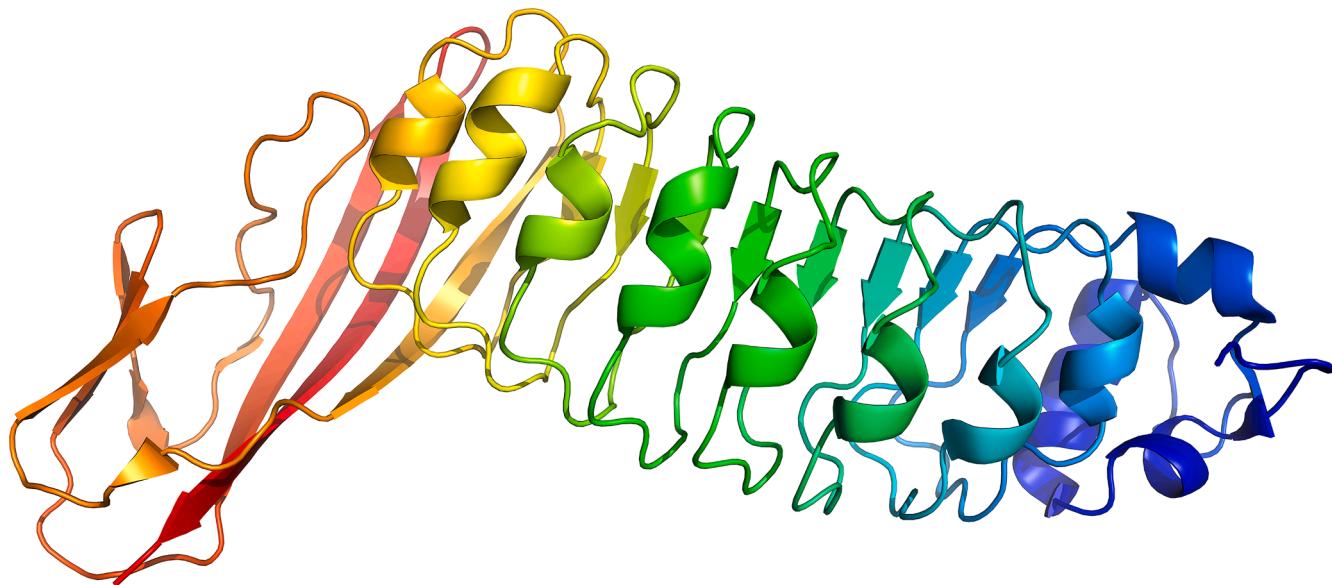
For decades, the main therapy for listeriosis has been antibiotics such

as high-dose intravenous ampicillin. However, these treatments are losing their effectiveness due to the emergence of antimicrobial resistance among strains of *L. monocytogenes* [7]. This increased resistance further narrows the already limited therapeutic options. Additionally, the bacterium's ability to survive and proliferate inside host cells, evade host immunity, and transmit quickly between cells without re-entering the extracellular environment creates a formidable obstacle to treatment [8]. Most antibiotics are ineffective in penetrating host cells to eliminate intracellular bacteria, rendering standard treatments less effective [9].

This presents a significant gap in existing medical research and highlights the current challenge of developing new treatment approaches. Rather than simply attacking the survival of the bacteria, one promising new strategy is to target its virulence factors, the very tools that it specifically employs to create disease. Through the neutralization of these factors, this approach disrupts the infection cycle and reduces the selective pressure that contributes to the evolution of antibiotic

\* Corresponding author.

E-mail address: [chouhanu@manit.ac.in](mailto:chouhanu@manit.ac.in) (U. Chouhan).



**Fig. 1.** 3D Structure of the Target Protein. The 3D ribbon structure of the *Listeria monocytogenes* Internalin H N-terminal domain (PDB ID: 1H6U). The putative binding pocket identified through blind docking is highlighted in the surface representation.

resistance [10].

The internalin protein family is an important group of surface proteins in *Listeria monocytogenes* that allows the bacterium to infect host cells, characterized by leucine-rich repeat domain proteins [11]. Primary members Internalin A (InlA) and Internalin B (InlB) are vital for the primary host cell invasion. InlA attaches to E-cadherin, the cell-cell adhesion protein molecule on human intestinal epithelial cells, an important requirement for passing through the intestinal barrier in a "zipper-like" mechanism [12]. InlB binds to the Met receptor, which is located on a broader variety of host cells, enabling the bacteria to infect more tissues and rapidly proliferate. Once internalized, *Listeria* escapes from the vacuole into the cytoplasm with the help of other virulence factors such as Listeriolysin O to manipulate the host's actin cytoskeleton through the ActA protein in order to proliferate, propel itself into neighboring cells to effectively spread the infection while evading the host immune system [8,13].

One such important virulence factor in *Listeria monocytogenes* is a surface protein known as Internalin H (InlH). Unlike other important proteins in the primary invasion, InlH is important in the later, systemic phases of infection. Its expression is induced by stress in the host environment, and through this, the bacterium evades the host's immune system by repressing inflammatory signal production such as interleukin-6 (IL-6). This facilitates the survival and dissemination of the bacteria throughout the body [14]. Due to the important role in pathogenesis, InlH is an excellent candidate for a new anti-virulence therapy. Blocking the action of InlH may boost the host's immune response to eliminate the infection.

To address this challenge, this study utilized *in silico* drug discovery methods. This is among the newer methods that utilized computer models and algorithms to predict how potential drug leads will respond to target proteins [15]. Methods like molecular docking and virtual screening, among other techniques, allow researchers to rapidly sort through extensive libraries of chemical compounds to identify promising leads in a very short time with high speed and efficiency [16]. The process reduces time spent carrying out experiments as well as their cost when compared to classical lab-based experiments [17].

To address this challenge, computational workflows have become indispensable in modern drug discovery, offering a rapid and cost-effective alternative to traditional high-throughput screening [15,17]. By filtering vast chemical libraries and predicting pharmacokinetic failures early in the process, *in silico* methods significantly accelerate the

identification of viable lead compounds [16]. Even though the structure of the crystal of InlH is available in PDB, there has been limited systematic screening of chemical compounds to find potential InlH inhibitors. One significant challenge in the process of discovering drugs is that most candidates fail in later stages due to unfavorable ADMET profiles [18]. This study seeks to bridge that gap through the application of a structure-based *in silico* method. Through the use of molecular docking software such as AutoDock Vina as well as ADMET (Absorption, Distribution, Metabolism, Excretion, and Toxicity) profiling, this study will systematically discover and characterize new small molecules that can inhibit InlH to pave the way for the formulation of new anti-listerial medicines.

However, while molecular docking is a powerful and efficient tool for high-throughput screening, it is fundamentally a static model [19, 20]. It provides a snapshot of a potential binding pose but often cannot fully account for the complex dynamic nature of protein-ligand interactions, such as protein flexibility or the crucial influence of a solvated, physiological environment [21]. The scoring functions used in docking, while excellent for rapid screening, also have limitations in their accuracy for the precise ranking of binding affinities [22,23].

To overcome these limitations and provide rigorous validation of our top candidates, we used a more robust, multi-layered computational pipeline. The first step was employing all-atom Molecular Dynamics (MD) simulations to assess the stability of the docked poses [21]. This involved simulating the physical motion of the protein-ligand complex for 125 ns, which allowed us to verify that the complex remained at a stable equilibrium and that its key intermolecular contacts were preserved. Following the MD simulations, the Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) method was employed, which is a well-accepted, "endpoint" calculation. This technique provides a more rigorous and physics-based estimation of the binding free energy ( $\Delta G_{bind}$ ) [22,23]. This step is crucial for re-scoring our hits and prioritizing candidates with the most favorable binding thermodynamics over those from docking alone.

Finally, the Density Functional Theory (DFT) was employed to understand the electronic-level properties that make our lead compounds effective. This quantum mechanical approach was used to analyze the "frontier molecular orbitals" (FMOs), specifically the Highest Occupied Molecular Orbital (HOMO) and the Lowest Occupied Molecular Orbital (LUMO) [24]. The resulting energy gap ( $\Delta E$ ) between these orbitals is a key descriptor of chemical reactivity and molecular stability. Generally,

a smaller HOMO-LUMO gap signifies a "softer," more reactive molecule that is a better electron donor, a favorable characteristic for a drug candidate [25]. This comprehensive workflow from rapid screening to rigorous biophysical and quantum mechanical validation provides a strong foundation for identifying high-affinity InlH inhibitors worthy of future experimental validation.

## Materials and methodology

### Computational software and environment

The computational work was divided between two environments. Initial tasks, such as molecular docking and system preparation, were performed locally on a machine with a 12th Gen Intel(R) Core(TM) i5-12500H 2.50 GHz processor and 16.0 GB RAM (running Ubuntu 24.04.2 via WSL). For the more demanding tasks, including all molecular dynamics simulations and DFT analyses, we utilized the cloud-based NVIDIA GPUs (T4, L4, and A100) available on Google Colab. The primary software packages and web servers employed for this study included: AutoDock 4 (v4.2.6) [26], utilized for the initial binding site identification and high-precision focused docking of the seven lead compounds, employing the Lamarckian Genetic Algorithm (LGA); AutoDock Vina (v1.2.3) [27], utilized for the high-throughput virtual screening (HTVS) of the analog library; AutoDockTools (ADT) (v1.5.7) [26], used for receptor and ligand preparation (PDBQT file generation); Open Babel (v3.1.1) [28], employed for file format conversion and ligand energy minimization; The PyMOL Molecular Graphics System (v2.5) [29], used for 3D structural visualization; the PLIP (Protein-Ligand Interaction Profiler) web server [30], used for 2D interaction diagram generation; and the SwissADME web server [31], used for in silico pharmacokinetic profiling.

### Receptor and ligand library preparation

The starting point for our receptor was the 3D crystallographic structure of the Internalin H (InlH) N-terminal domain (PDB ID: 1H6U) (Fig. 1), which was downloaded from the RCSB Protein Data Bank [32]. Preparation of the raw PDB structure was performed with AutoDockTools (ADT) [26]. This process included removing all non-receptor atoms (e.g., HOH), adding polar hydrogens, and assigning Kollman united-atom partial charges to the protein [33]. The final receptor was saved in the PDBQT format. An initial set of seven lead compounds was selected based on prior reports of biological activity. For the HTVS, a large analog library of 7000 structurally similar compounds was retrieved from PubChem [34]. To create a final, qualified library, this set was pre-filtered for compliance with Lipinski's Rule of Five [35], which yielded 209 candidate compounds. Finally, each of the 209 qualified ligands was prepared for docking. This involved energy minimization using Open Babel [28] (with the MMFF94 force field energy minimization [36]) and processing through ADT [26] to assign Gasteiger partial charge [37] and define rotatable bonds. These resulting structures were then saved in the PDBQT format.

### Molecular docking protocol

A multi-stage docking strategy was employed to accommodate the distinct binding preferences of the diverse lead compounds. First, to identify potential binding cavities across the entire protein surface, a blind docking protocol was performed on the seven initial lead compounds using AutoDock 4 [26]. A large grid box ( $126 \times 126 \times 126 \text{ \AA}$ ) was defined to encompass the entire protein structure; for example, in the screening of Oleanolic Acid, the grid was centered at x: 15.992, y: -7.163, z: 67.378. The Lamarckian Genetic Algorithm (LGA) was run for 50 iterations (ga\_num\_runs = 50).

Analysis of the blind docking results revealed distinct high-affinity binding pockets specific to each compound class. To refine these

**Table 1**  
Focused docking grid box center coordinates.

Sr. No	Compound Name	X center	Y center	Z center
1.	Chelerythrine	-9.518	-17.350	91.941
2.	Lupeol	5.379	-3.299	88.225
3.	Quercetin	-10.608	11.533	34.521
4.	Ursolic Acid	7.100	-7.531	95.115
5.	Maslinic Acid	-6.621	23.422	45.872
6.	Hederagenin	-16.284	-5.865	66.007
7.	Oleanolic Acid	-15.992	-7.163	67.378

**Table 2**  
Key Parameters used for GROMACS molecular dynamics simulations.

MD Parameters	Control	Hederagenin Analogue	Lupeol Analogue	Maslinic Acid Analogue
Simulation time (ns)	125	125	125	125
RMSD Protein (nm)	0.15 ± 0.03	0.15 ± 0.03	0.16 ± 0.03	0.15 ± 0.04
RMSF (nm)	0.08 ± 0.03	0.09 ± 0.03	0.09 ± 0.03	0.09 ± 0.03
Rg (nm)	2.55 ± 0.05	2.59 ± 0.01	2.55 ± 0.03	2.54 ± 0.03
Hydrogen bonds	0.30	0.14	0.07	0.31
H-bonds pairs (<0.35 nm)	1.23	1.34	1.04	1.26
SASA (nm <sup>2</sup> )	143.56 ± 3.12	145.40 ± 2.75	141.49 ± 2.86	140.04 ± 2.78

binding poses, a focused docking protocol was generated for each lead compound. Individual grid boxes were centered on the specific coordinates identified during the blind docking phase (using coordinates from Table 1) using high-precision LGA parameters (ga\_num\_runs = 100) to obtain the binding energies reported in Table 2. Subsequently, the 209-compound library was screened using the computationally efficient AutoDock Vina [27]. This step employed a targeted approach: analogues were docked specifically into the focused grid box of their structural parent compound (e.g., Maslinic Acid analogues were docked into the Maslinic Acid binding pocket). A hierarchical screening protocol was used: a primary screen (exhaustiveness = 8), followed by a high-precision refinement (exhaustiveness = 32) of the top-scoring candidates from each analog family. The final binding affinities from this refinement step are reported in Table 4.

### Validation of the docking protocol

Validation of the docking protocol was performed by assessing the convergence of binding modes. Since the target structure (PDB: 1H6U) is an apo-protein lacking a co-crystallized ligand, a standard redocking validation (RMSD calculation relative to a crystal pose) was not applicable. Instead, the robustness of the protocol was validated by comparing the results of blind docking (covering the entire protein surface) with focused docking (local search). For each lead compound, both protocols consistently converged on the same specific binding cavity. Furthermore, the validity of the generated poses was rigorously confirmed via 125 ns molecular dynamics simulations. The Minimum Distance analysis confirmed that the docked poses were stable and maintained secure positioning within their respective active sites throughout the simulation (Fig. 9), validating the predictive accuracy of the docking protocol.

### Binding mode visualization

To elucidate the molecular basis of binding, the protein-ligand interactions of the top-ranked compounds were analyzed. 3D structural visualizations were rendered using The PyMOL Molecular Graphics

System [29]. 2D diagrams detailing the specific hydrogen bonds, hydrophobic contacts, and  $\pi$ -interactions were generated using the Protein-Ligand Interaction Profiler (PLIP) [30] web server.

#### In silico admet profiling

To assess the "drug-likeness" of the top Vina hits, the *in silico* pharmacokinetic profiling was performed. The canonical SMILES strings for the lead compounds were submitted to the SwissADME web server [31]. This analysis predicted key ADMET parameters, including high gastrointestinal (GI) absorption, blood-brain barrier (BBB) penetration, potential as a P-glycoprotein (P-gp) substrate, and overall drug-likeness.

#### Molecular dynamics (MD) simulations

To validate the stability of the docked poses and to understand the dynamic behavior of the protein-ligand complexes, all-atom MD simulations were performed for the top three lead analogues (Maslinic Acid, Hederagenin, and Lupeol analogues) and positive control (Oleanolic Acid). All simulations were conducted using the GROMACS 2025.3 software package [38]. A dual-topology approach was used, employing the AMBER99SB-ILDN force field [39] for the protein and the General Amber Force Field (GAFF) [40] for the ligands. The protein topology was generated using gmx pdb2gmx, selecting the AMBER99SB-ILDN force field [39] and the TIP3P water model [41]. Ligand parameters were generated using the Antechamber and acpype workflow. antechamber [42] was used to calculate AM1-BCC charges [43], and acpype [44] was used to convert the AMBER-formatted topologies into GROMACS-compatible .itp files. For each of the four systems, the complex was centered in a triclinic simulation box using gmx editconf (1.0 nm margin) and solvated with TIP3P water molecules using gmx solvate [41]. The systems were then neutralized by adding  $\text{Na}^+$  and  $\text{Cl}^-$  counter-ions using gmx genion. Each system was then subjected to a four-stage protocol: (i) energy minimization using the steepest descent algorithm; (ii) a 1000 ps (1 ns) NVT (canonical) ensemble equilibration at 300 K using the V-rescale thermostat [45]; and (iii) a 1000 ps (1 ns) NPT (isothermal-isobaric) ensemble equilibration at 1 bar using the Berendsen barostat [46]. Position restraints were applied to the protein-ligand complex heavy atoms during both equilibration phases. Finally, all restraints were removed, and a 125 ns production MD simulation was run for each complex, using the Parrinello-Rahman barostat for pressure coupling [47]. Trajectories were saved every 10 ps for subsequent analysis.

#### Post-Simulation trajectory analysis

For the analysis of stability and dynamics, the production trajectories first required processing. They were concatenated into a single 125 ns trajectory (using gmx trjcat). Next, corrections for periodic boundary conditions (PBC) were applied using gmx trjconv, which centered the complex and removed rotational/translational "jumps" by fitting to a reference structure. The GROMACS [38] built-in modules, VMD [48], and the PLIP web server [30] were utilized for all the subsequent analyses. Several key metrics were calculated to assess structural stability. The Root Mean Square Deviation (RMSD) of the protein backbone and ligand was determined with gmx rms, while the protein's compactness was measured via the Radius of Gyration ( $R_g$ ) using gmx gyrate. To analyze per-residue flexibility, the Root Mean Square Fluctuation (RMSF) of  $\text{C}\alpha$  atoms was computed with gmx rmsf. Finally, the gmx sasa tool was used to calculate the Solvent Accessible Surface Area (SASA). Intermolecular hydrogen bond formation was analyzed using the VMD software package [48], and the total number of non-covalent contacts (within 0.6 nm) was calculated using the gmx mindist tool. Advanced dynamic analyses were also performed. Large-scale collective motions were identified through Principal Component Analysis (PCA) (using gmx covar and gmx ana eig). A Free Energy Landscape (FEL) was then

constructed from the resulting projections via gmx sham. Separately, correlated residue motions were investigated by calculating a Dynamic Cross-Correlation Matrix (DCCM); this involved converting the covariance matrix from gmx covar into a visual map using gmx xpm2ps.

Finally, a representative structure from the most stable cluster of the FEL was extracted and analyzed using the PLIP web server [30] to detail the persistent hydrogen bonds, hydrophobic contacts, and other non-covalent interactions.

#### Binding free energy (MM/GBSA) calculation

The Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) method was utilized to calculate the binding free energy ( $\Delta G_{bind}$ ). This calculation is defined as the difference between the free energy of the protein-ligand complex ( $G_{complex}$ ) and the free energy of the isolated receptor ( $G_{receptor}$ ) and ligand ( $G_{ligand}$ ):

$$\Delta G_{bind} = G_{complex} - (G_{receptor} + G_{ligand})$$

The free energy of each component is calculated as the sum of its gas-phase molecular mechanics (MM) energy ( $\Delta E_{MM}$ ), the solvation free energy ( $\Delta G_{solv}$ ), and an entropic term ( $-T\Delta S$ ). The final binding free energy is commonly expressed as:

$$\Delta G_{bind} = \Delta E_{MM} + \Delta G_{solv} - T\Delta S$$

An effective binding free energy was calculated. This was done because the entropic contribution ( $-T\Delta S$ ) is computationally demanding and often similar for related ligands, making it acceptable to exclude for relative ranking studies. The equation for effective binding free energy:

$$\Delta G_{bind} = \Delta E_{vdw} + \Delta E_{elec} + \Delta G_{GB} + \Delta G_{non-polar}$$

The gmx\_MMPBSA tool [49], which is an external package compatible with GROMACS [38] trajectories and AMBER topologies was used for this calculation. The binding free energy [50] ( $\Delta G_{bind}$ ) was determined by analyzing 501 frames, which were extracted at 20-frame intervals from the first 100 ns of the production trajectory. For the implicit solvent, the Generalized Born (GB) model was utilized igb=5, corresponding to GB-OBC2 [50], with a salt concentration of 0.150 M.

#### Density functional theory (DFT) analysis

Quantum mechanical calculations were performed with the ORCA (v5.0.4) computational chemistry package [51]. This was done to investigate the electronic properties and inherent chemical reactivity of the four lead compounds. Geometry optimization was first performed on the 3D structure of each ligand (as extracted from its complex). The optimization was conducted at the Density Functional Theory (DFT) level. The selected parameters included the B3LYP [52,53] functional, and the def2-SVP [54] basis set with the D3 dispersion correction [55] denoted as (! B3LYP-D3/def2-SVP).

After successful optimization, the energies of the Frontier Molecular Orbitals (FMOs) were calculated, specifically for the Highest Occupied Molecular Orbital (HOMO) and the Lowest Occupied Molecular Orbital (LUMO). These energies were used to derive the HOMO-LUMO energy gap ( $\Delta E = E_{LUMO} - E_{HOMO}$ ) which serves as a key descriptor of chemical reactivity and molecular stability. In addition, following recent computational protocols [56,57], a Molecular Electrostatic Potential (MEP) surface was generated for each optimized structure. This was done to visualize the charge distribution and to identify potential sites for electrophilic and nucleophilic interactions.

**Table 3**

Binding affinities of initial seven compounds from blind vs. focused docking.

Sr. No	CID	Compound Name	Source	Blind Dock Score (kcal/mol)	Focused Dock Score (kcal/mol)
1.	2703	Chelerythrine	Pubchem	-5.66	-6.72
2.	259,846	Lupeol	Pubchem	-5.74	-6.27
3.	5280,343	Quercetin	Pubchem	-5.17	-7.03
4.	64,945	Ursolic Acid	Pubchem	-5.71	-6.39
5.	73,659	Maslinic Acid	Pubchem	-5.90	-7.22
6.	73,299	Hederagenin	Pubchem	-5.26	-6.60
7.	10,494	Oleanolic Acid	Pubchem	-6.90	-7.63

## Result

### Identification of binding site via initial autodock 4 screening

The computational workflow began by screening an initial set of seven known biologically active compounds against the Internalin H protein (PDB: 1H6U) using AutoDock 4. A two-step protocol was employed: first, blind docking was performed to identify the most probable binding cavity on the entire protein surface. Following this, a more precise focused docking was centered on this putative active site.

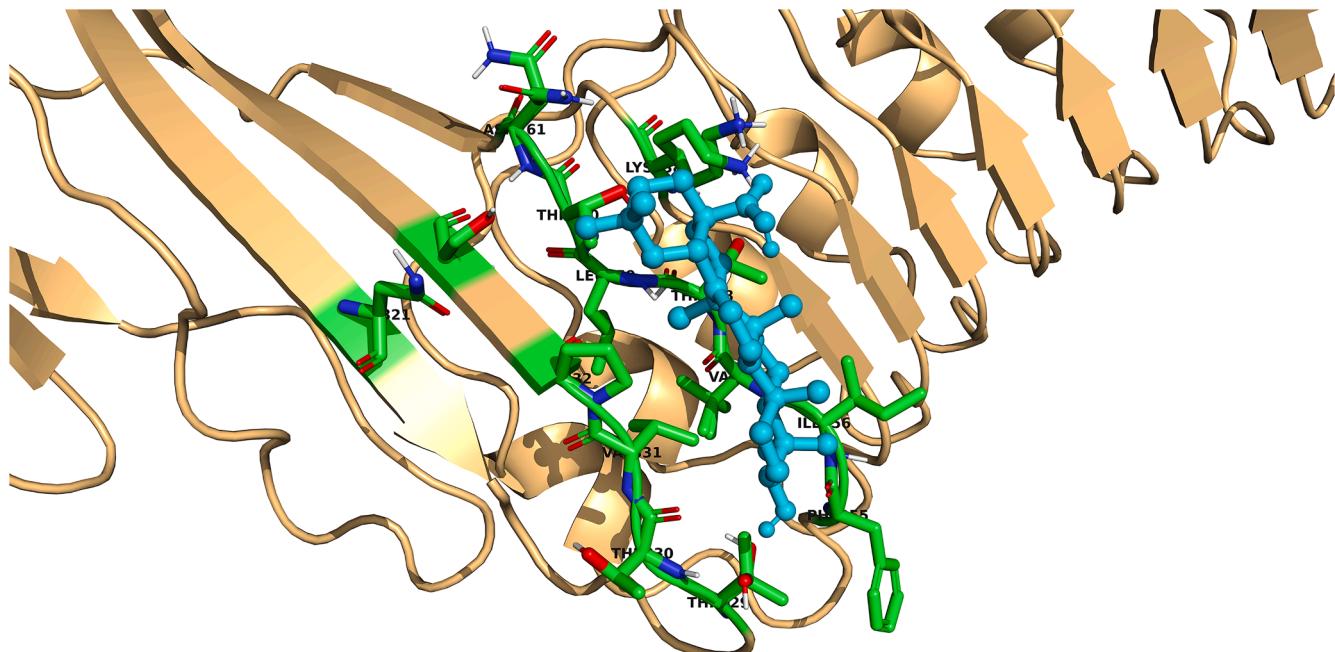
The binding affinities for both protocols are summarized in **Table 3**. The results show that the focused docking protocol yielded significantly

more favorable (more negative) binding energies for all seven compounds, confirming a single, preferential binding site. Of this initial set, Oleanolic Acid (CID 10,494) emerged as the top candidate, exhibiting the strongest binding affinity at -7.63 kcal/mol.

This optimal binding pose was selected from the largest and most stable conformation cluster identified in AutoDock 4 analysis. A visualization of this pose is presented in **Fig. 2**, showing Oleanolic Acid securely docked within the identified binding pocket of InlH. Analysis of this interaction, as detailed in the PLIP report, confirmed a stable binding mode maintained by several key hydrogen bonds and hydrophobic contacts with active site residues.

### Virtual screening for novel high-affinity inhibitors

Following the identification of the binding pocket and the establishment of Oleanolic Acid (-7.63 kcal/mol) as a control, a high-throughput virtual screening (HTVS) was performed to discover novel analogues with potentially superior binding affinities. A large library of 7000 structural analogues was retrieved from PubChem, which was then pre-filtered for compliance with Lipinski's Rule of Five. This crucial filtering step reduced the library to a final, drug-like set of 209 candidate compounds (as detailed in **Table 4**). This qualified library was then screened against the focused binding pocket using the computationally efficient AutoDock Vina.

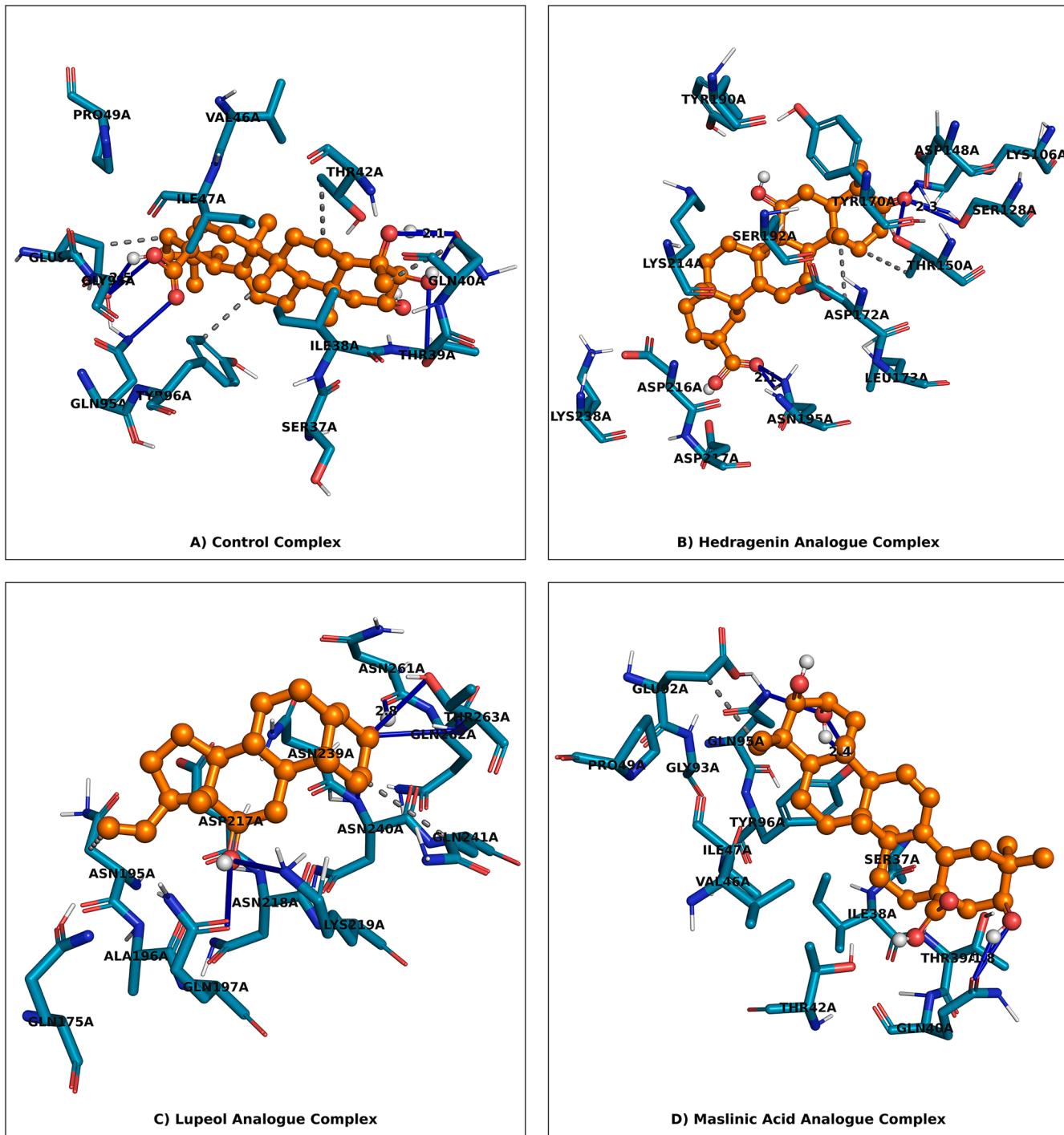


**Fig. 2.** Chemical Structures of the Seven Initial Lead Compounds. 3D structures of the seven biologically active compounds used for the initial AutoDock 4 screening. Oleanolic Acid, which had the top score, was selected as the control.

**Table 4**

Top-scoring candidates from autodock vina virtual screening.

Sr. No	Compound Name	CID of similar compounds	Source	Blind Dock Score (kcal/mol)	Focused Dock Score (kcal/mol)	Compounds Filter according to Lipinski Rule of 5	
						Before filter	After filter
1.	Chelerythrine	25,093,273	Pubchem	-6.3	-6.2	1000	15
2.	Lupeol	70,626,379	Pubchem	-6.3	-7.3	1000	17
3.	Quercetin	168,296,925	Pubchem	-6.7	-6.6	1000	42
4.	Ursolic Acid	70,680,338	Pubchem	-7.3	-7.2	1000	28
5.	Maslinic Acid	163,053,220	Pubchem	-7.2	-7.4	1000	42
6.	Hederagenin	137,633,443	Pubchem	-7.3	-7.3	1000	37
7.	Oleanolic Acid	44,575,931	Pubchem	-7.2	-7.2	1000	28



**Fig. 3.** Static Protein-Ligand Interaction Diagrams (PLIP). 3D diagrams of the static (docked) interactions for the four lead candidates, generated by PLIP: (A) Hederagenin analogue, (B) Maslinic Acid analogue, (C) Oleanolic Acid (Control), and (D) Lupeol analogue, detailing the specific H-bonds and hydrophobic contacts.

The screening successfully identified several novel compounds with binding affinities that surpassed the original lead compounds. The top-scoring candidate from each of the seven analog families is presented in Table 4. Notably, a Maslinic Acid analogue (CID 163,053,220) was identified as the top-ranking hit with a binding energy of  $-7.4$  kcal/mol. This was followed by a Lupeol analogue (CID 70,626,379) and a Hederagenin analogue (CID 137,633,443), which both showed strong affinities of  $-7.3$  kcal/mol. Based on these promising results, these three lead analogues, along with the Oleanolic Acid control, were selected for full biophysical validation.

### *Analysis of predicted binding modes*

Before advancing the new lead compounds to MD simulation, their static binding poses from AutoDock Vina were analyzed to confirm their binding mode and elucidate their predicted interaction networks. The representative poses for all three analogues were visualized and found to bind securely within the same high-affinity pocket as the Oleanolic Acid control. The specific interactions were detailed using PLIP. These static interaction networks are visualized in Fig. 3 and summarized in detail in Table 5:

**Table 5**

Summary of key intermolecular interactions from static docking poses (PLIP analysis).

Compounds	Hydrophobic Interactions				Hydrogen Bonds		
	Residue	Distance (Å)	Ligand Atom	Protein Atom	Residue	Distance (Å)	
						H-A	D-A
Control	Gln40	3.85	C26	C79	Thr39	2.59	3.77
	Thr42	3.74	C8	C98	Gln40	2.08	2.93
	Glu92	3.30	C23	C521	Gln40	2.05	2.71
	Tyr96	3.66	C16	C560	Glu92	2.46	3.23
Hederagenin Analogue	Thr150	3.69	C14	C1062	Lys106	2.39	3.38
	Tyr170	3.65	C27	C1239	Ser128	2.30	3.12
	Asp172	3.79	C20	C1259	Thr150	2.85	3.74
Lupeol Analogue	Asn195	3.89	C21	C1456	Asn195	2.06	3.10
	Asp217	4.00	C10	C1662	Gln197	3.21	4.02
	Gln241	3.98	C18	C1908	Lys219	3.29	3.72
					Asn240	3.20	3.58
					Thr263	2.67	3.59
Maslinic Acid Analogue					Thr263	2.41	3.22
	Glu92	3.69	C23	C522	Gln40	1.82	2.73
					Gln95	2.09	3.02
					Tyr96	2.42	3.04

**Table 6**

Predicted physicochemical properties (from SwissADME).

Compound ID	C log <sub>p</sub>	Mol. Wt. (250 to 500 g/mol)	HBD (0–5)	HBA (0–10)	Rotatable Bonds (<10)	2TPSA (<140 Å)	Solubility (mg/ml)
73,193	4.33	488.7	4	5	1	97.99	5.76E-04
490,364	4.3	488.7	4	5	1	97.99	5.76E-04
11,156,036	4.3	488.7	4	5	1	97.99	5.76E-04
11,705,538	3.91	472.66	4	5	1	97.99	3.77E-03
12,067,090	4.08	472.66	4	5	2	97.99	1.57E-03
12,134,775	4.41	442.63	3	4	1	77.76	2.27E-03
12,442,765	4.08	488.66	4	6	2	115.06	7.11E-04
44,575,931	4.03	474.67	4	5	1	97.99	9.96E-04
53,253,851	4.29	488.7	4	5	1	97.99	7.26E-04
53,822,830	4.5	300.48	1	1	0	20.23	5.63E-03
68,335,034	3.73	304.47	2	2	0	40.46	1.38E-02
70,626,379	3.99	318.49	2	2	1	40.46	4.58E-03
70,680,338	4.35	486.68	4	5	2	97.99	7.61E-04
70,697,438	4.19	486.68	4	5	2	97.99	7.61E-04
89,593,574	3.74	304.47	2	2	0	40.46	1.38E-02
90,450,652	4.28	274.44	1	1	0	20.23	6.74E-03
101,844,542	4.01	472.66	4	5	1	97.99	9.50E-04
102,408,107	4.03	472.66	4	5	2	97.99	1.78E-03
118,580,276	4.26	274.44	1	1	0	20.23	6.74E-03
118,580,279	4.25	274.44	1	1	0	20.23	6.74E-03
137,633,443	4.53	484.67	2	5	1	91.67	8.09E-04
156,011,190	4.16	484.67	4	5	2	97.99	6.36E-04
162,962,832	4.15	474.67	4	5	1	97.99	7.56E-04
163,053,220	4.31	486.68	4	5	1	97.99	6.82E-04
168,296,925	2.5	328.32	3	6	4	100.13	2.38E-02

- Maslinic Acid Analogue: The top-scoring pose ( $-7.4$  kcal/mol) was stabilized by a robust network of interactions. Key interactions included three hydrogen bonds with Gln40, Gln95, and Tyr96, as well as a hydrophobic contact with Glu92.
- Hederagenin Analogue: This analogue ( $-7.3$  kcal/mol) also showed a strong interaction profile, forming a network of four key hydrogen bonds with Thr150, Asn195, Gln240, and Thr263.
- Lupeol Analogue: This pose ( $-7.3$  kcal/mol) was primarily stabilized by hydrophobic contacts with residues such as Asn195, Tyr241, and Thr263.

This analysis confirmed that all three lead analogues achieved their high affinity by forming strong, specific interactions within the target binding site, justifying their advancement to further validation.

#### Analysis of key binding residues

The interaction analysis highlights the critical role of specific

residues within the Internalin H binding pocket. As detailed in **Table 5**, residues such as Lys106, Thr150, and Gln40 act as primary anchors, forming persistent hydrogen bonds that lock the ligands into the active site. Structurally, these residues are located within the concave face of the Leucine-Rich Repeat (LRR) domain, which is the canonical protein-protein interaction interface responsible for host cell recognition. Other residues, such as Asp172 and Asn195, contribute to the hydrophobic enclosure of the triterpenoid scaffold. The engagement of these specific residues is biologically significant because it suggests the inhibitors physically obstruct the surface region required for InlH to interact with host receptors, thereby mechanically blocking the virulence mechanism.

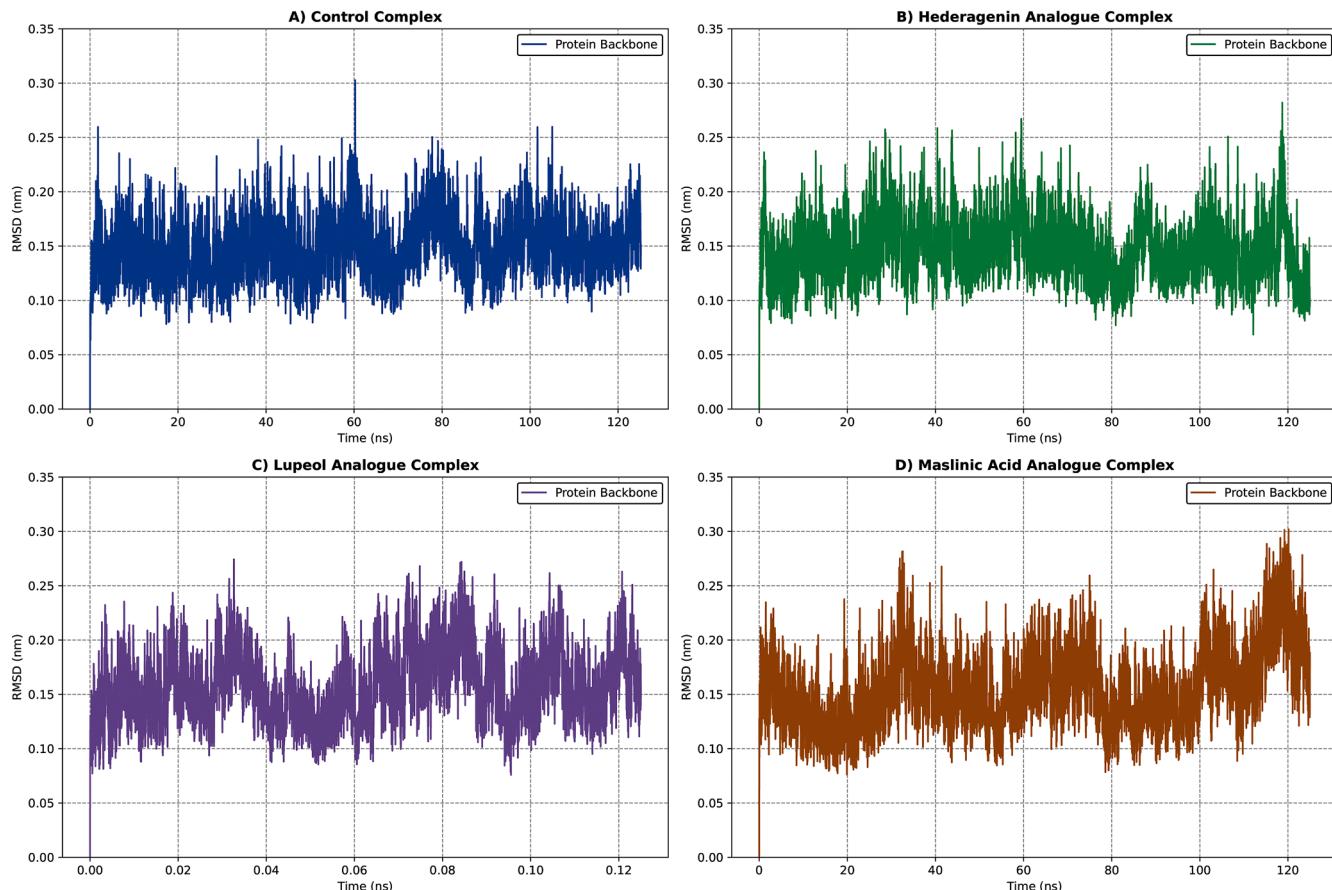
#### In silico *admet* and physicochemical profiling

After identifying the top-scoring candidates from the virtual screening, their "drug-likeness" and pharmacokinetic properties were evaluated *in silico* using the SwissADME web server. This step was crucial to ensure that the lead compounds possess favorable properties

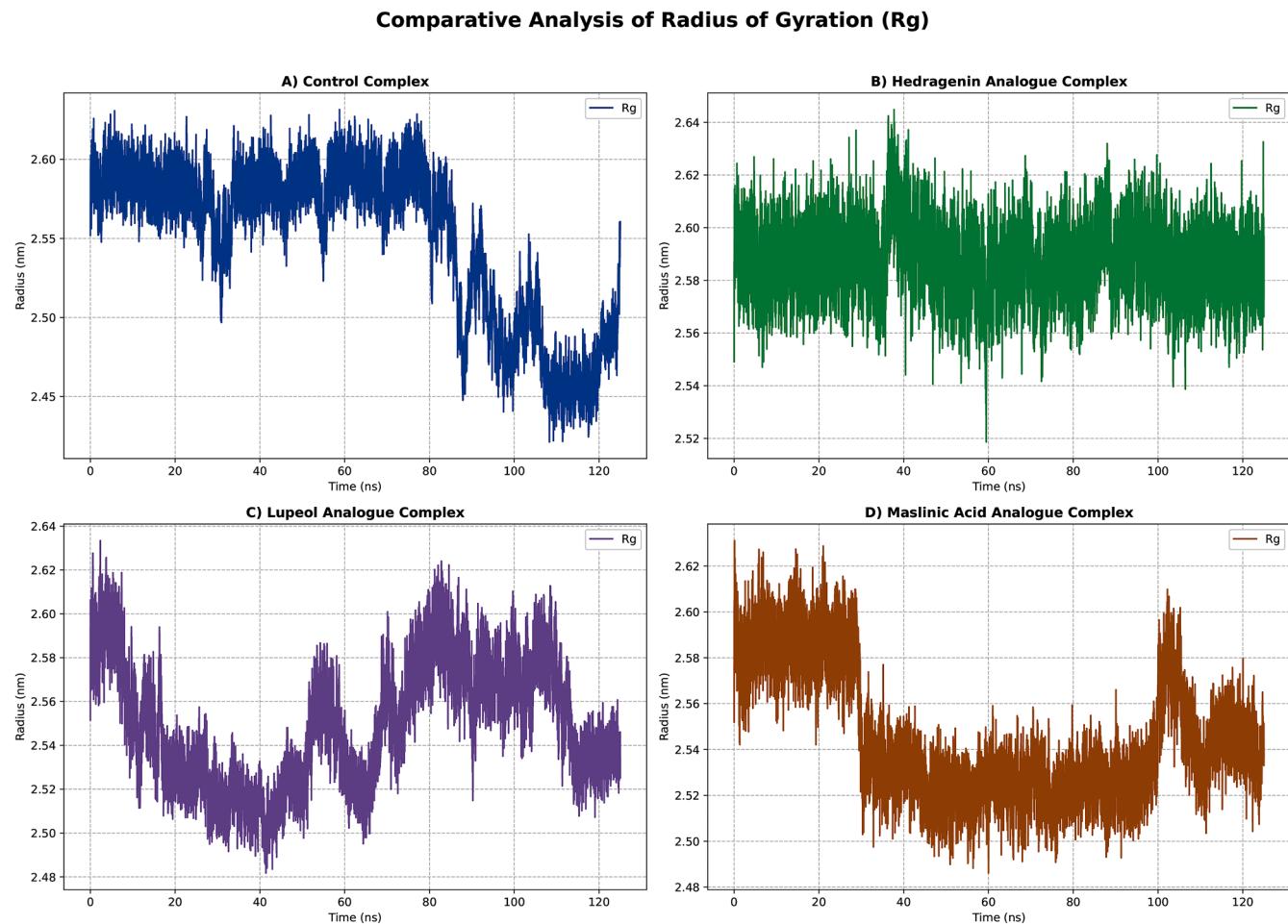
**Table 7**

Predicted pharmacokinetic (ADMET) properties (from SwissADME).

Compound ID	Formula	GI absorption	BBB permeant	Pgp substrate	Skin permeation	Bioavailability	Synthetic Accessibility
73,193	C <sub>30</sub> H <sub>48</sub> O <sub>5</sub>	High	No	Yes	-5.76	0.56	6.59
490,364	C <sub>30</sub> H <sub>48</sub> O <sub>5</sub>	High	No	Yes	-5.76	0.56	6.59
11,156,036	C <sub>30</sub> H <sub>48</sub> O <sub>5</sub>	High	No	Yes	-5.76	0.56	6.59
11,705,538	C <sub>29</sub> H <sub>44</sub> O <sub>5</sub>	High	No	Yes	-6.49	0.56	6.42
12,067,090	C <sub>29</sub> H <sub>44</sub> O <sub>5</sub>	High	No	Yes	-5.98	0.56	6.43
12,134,775	C <sub>28</sub> H <sub>42</sub> O <sub>4</sub>	High	No	Yes	-5.88	0.56	6.09
12,442,765	C <sub>29</sub> H <sub>44</sub> O <sub>6</sub>	Low	No	Yes	-5.79	0.56	6.4
44,575,931	C <sub>29</sub> H <sub>46</sub> O <sub>5</sub>	High	No	Yes	-5.86	0.56	6.32
53,253,851	C <sub>30</sub> H <sub>48</sub> O <sub>5</sub>	High	No	Yes	-5.87	0.56	6.53
53,822,830	C <sub>21</sub> H <sub>32</sub> O	High	Yes	No	-4.72	0.55	4.73
68,335,034	C <sub>20</sub> H <sub>30</sub> O <sub>2</sub>	High	Yes	Yes	-5.21	0.55	4.41
70,626,379	C <sub>12</sub> H <sub>34</sub> O <sub>2</sub>	High	Yes	Yes	-4.76	0.55	4.59
70,680,338	C <sub>30</sub> H <sub>46</sub> O <sub>5</sub>	High	No	Yes	-5.8	0.56	6.6
70,697,438	C <sub>30</sub> H <sub>46</sub> O <sub>5</sub>	High	No	Yes	-5.8	0.56	6.6
89,593,574	C <sub>20</sub> H <sub>32</sub> O <sub>2</sub>	High	Yes	Yes	-5.21	0.55	4.41
90,450,652	C <sub>19</sub> H <sub>36</sub> O	High	Yes	No	-4.52	0.55	4.07
101,844,542	C <sub>29</sub> H <sub>44</sub> O <sub>5</sub>	High	No	Yes	-5.81	0.56	6.33
102,408,107	C <sub>29</sub> H <sub>44</sub> O <sub>5</sub>	High	No	Yes	-6.05	0.56	6.43
118,580,276	C <sub>19</sub> H <sub>36</sub> O	High	Yes	No	-4.52	0.55	4.07
118,580,279	C <sub>19</sub> H <sub>36</sub> O	High	Yes	No	-4.52	0.55	4.07
137,633,443	C <sub>30</sub> H <sub>44</sub> O <sub>5</sub>	High	No	Yes	-5.88	0.56	6.15
156,011,190	C <sub>30</sub> H <sub>46</sub> O <sub>5</sub>	High	No	Yes	-5.69	0.56	6.96
162,962,832	C <sub>29</sub> H <sub>46</sub> O <sub>5</sub>	High	No	Yes	-5.72	0.56	6.34
163,053,220	C <sub>30</sub> H <sub>46</sub> O <sub>5</sub>	High	No	Yes	-5.82	0.56	6.77
168,296,925	C <sub>18</sub> H <sub>16</sub> O <sub>6</sub>	High	No	No	-6.01	0.55	3.21

**Protein Backbone Structural Stability Analysis (RMSD)**

**Fig. 4.** Root Mean Square Deviation (RMSD) of Protein Backbone. Root Mean Square Deviation (RMSD) of the protein backbone (C<sub>α</sub> atoms) over the 125 ns simulation time. The plots confirm that all four apo-enzyme systems reached a stable structural equilibrium after approximately 25 ns, maintaining a compact fold with an average RMSD of 0.15–0.16 nm throughout the trajectory.



**Fig. 5.** Radius of Gyration (Rg) of Protein. Radius of Gyration (Rg) of the protein  $\text{Ca}$  atoms for the four complexes. The stable and overlapping plots (fluctuating  $\sim 2.45 \text{ nm}$ ) confirm that all complexes maintained a compact globular fold.

before proceeding to more computationally intensive validations.

The physicochemical properties of the top hits are summarized in **Table 6**. The compounds were found to be in favorable ranges, with molecular weights between 274.44 and 488.7 g/mol and ClogP values generally between 2.5 and 4.53. All lead compounds showed full compliance with Lipinski's Rule of Five.

The predicted pharmacokinetic (ADMET) parameters are presented in **Table 7**. Key findings were that the top-ranked compounds were all predicted to have high gastrointestinal (GI) absorption. Furthermore, none of the lead triterpenoid analogues were predicted to be blood-brain barrier (BBB) permeant, a favorable safety feature that suggests a low risk of central nervous system side effects. While several compounds were identified as P-glycoprotein (P-gp) substrates, they all showed a high predicted oral bioavailability score (0.55–0.56). The selection of the lead analogues as high-priority candidates for MD simulation was justified by their strong binding affinities combined with their favorable ADMET profiles.

#### Molecular dynamics simulation

All-atom molecular dynamics (MD) simulations of 125 ns were performed on the four complexes. The purpose of these simulations was to rigorously validate the static binding poses and interaction networks (previously identified in *Analysis of Predicted Binding Modes*) and to explore the dynamic behavior of the lead compounds within a more realistic physiological environment. Static docking, while powerful for screening, provides a single, high-scoring snapshot. This allowed for the

stability of protein-ligand contacts, the system's convergence to an equilibrium state, and the overall structural integrity of the bound protein to be assessed, thus providing a much more rigorous test of binding stability.

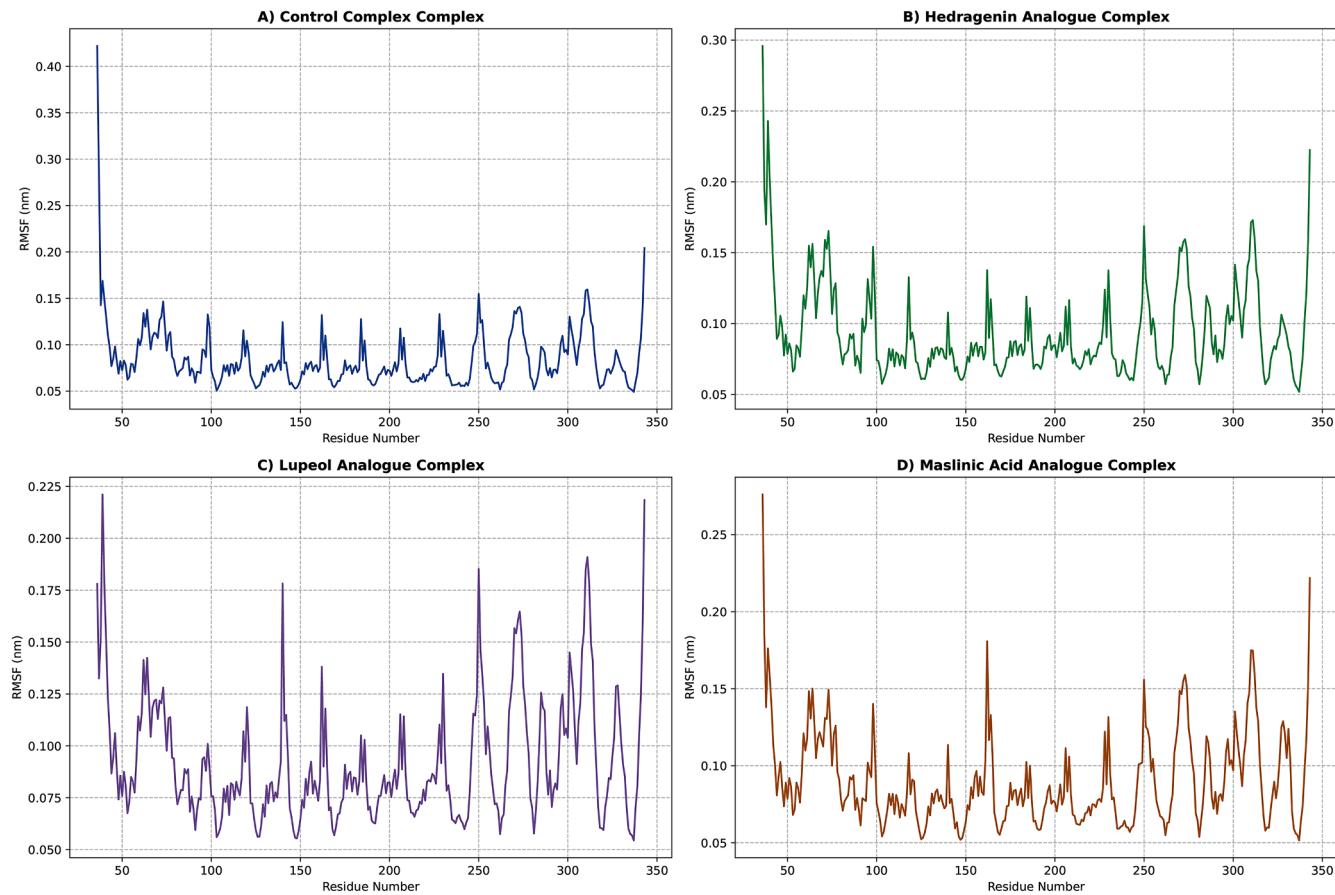
#### Trajectory pre-analysis and correction

Before any biophysical analysis could be performed, the raw simulation output required essential pre-processing. The five 25 ns production runs for each of the four complexes were first concatenated into a single 125 ns production trajectory using `gmx trjcat`. This raw trajectory, however, contains simulation artifacts from the Periodic Boundary Conditions (PBC), such as the complex appearing to "jump" across the box as it traverses the boundary. To create a continuous, analyzable trajectory, these artifacts were corrected using `gmx trjconv`. This step centered the protein-ligand complex in the simulation box and performed a rotational and translational fit to the reference starting structure. This "cleaning" process removes all non-physical jumps and ensures that all subsequent analyses, such as RMSD and Rg, are calculated on a meaningful, continuous trajectory.

#### Structural stability analysis

Following trajectory correction, the global structural stability of the protein-ligand complexes was assessed to confirm convergence to a stable equilibrium. The Root Mean Square Deviation (RMSD) of the protein backbone serves as a primary metric for structural deviation

### Comparative Analysis of RMS Fluctuation



**Fig. 6.** Root Mean Square Fluctuation (RMSF) of Protein Residues. *RMSF of the protein  $\text{C}\alpha$  atoms, showing per-residue flexibility. Major peaks correspond to exposed loop regions, while binding site residues show low fluctuation, indicating stabilization by the ligands..*

from the initial crystal conformation.

#### Root mean square deviation (RMSD)

The RMSD plots (Fig. 4) illustrate the system's convergence to equilibrium. An initial rise in RMSD was observed for all four complexes during the first 20–25 ns as the systems relaxed in the solvated environment. Following this equilibration phase, a stable plateau was achieved by all four trajectories, with tight fluctuations around an average of 0.15–0.16 nm for the remaining 100 ns. This absence of significant drift served as confirmation that a stable equilibrium was reached and that the protein's overall fold was not disrupted or destabilized by ligand binding.

#### Ligand binding stability

To assess the stability of the ligand within the binding pocket, a Minimum Distance analysis was performed (Fig. 9) rather than ligand RMSD, as the latter can be influenced by rotational symmetries within the pocket. The results definitively show that the Hederagenin and Maslinic Acid analogues maintained a consistent proximity of ~0.15–0.20 nm to the protein residues throughout the 125 ns simulation. This stable distance, coupled with the persistent intermolecular contacts (Fig. 10), confirms that the ligands remained securely anchored within the active site without dissociation.

#### Radius of gyration ( $R_g$ )

The Radius of Gyration ( $R_g$ ) analysis (Root **Mean Square Deviation (RMSD) of the protein backbone ( $\text{C}\alpha$  atoms) over the 125 ns simulation time**) The plots confirm that all four apo-enzyme systems reached a stable structural equilibrium after approximately 25 ns, maintaining a compact fold with an average RMSD of 0.15–0.16 nm throughout the trajectory.

Fig. 5, which is a measure of overall protein compactness, provided further support for this finding. Highly stable  $R_g$  plots were observed for all four complexes from the start of the simulation, with narrow fluctuations around ~2.45 nm. This stable  $R_g$  value is indicative that the protein's compact, globular fold was maintained and that no unfolding, swelling, or significant conformational collapse occurred while bound to any of the ligands.

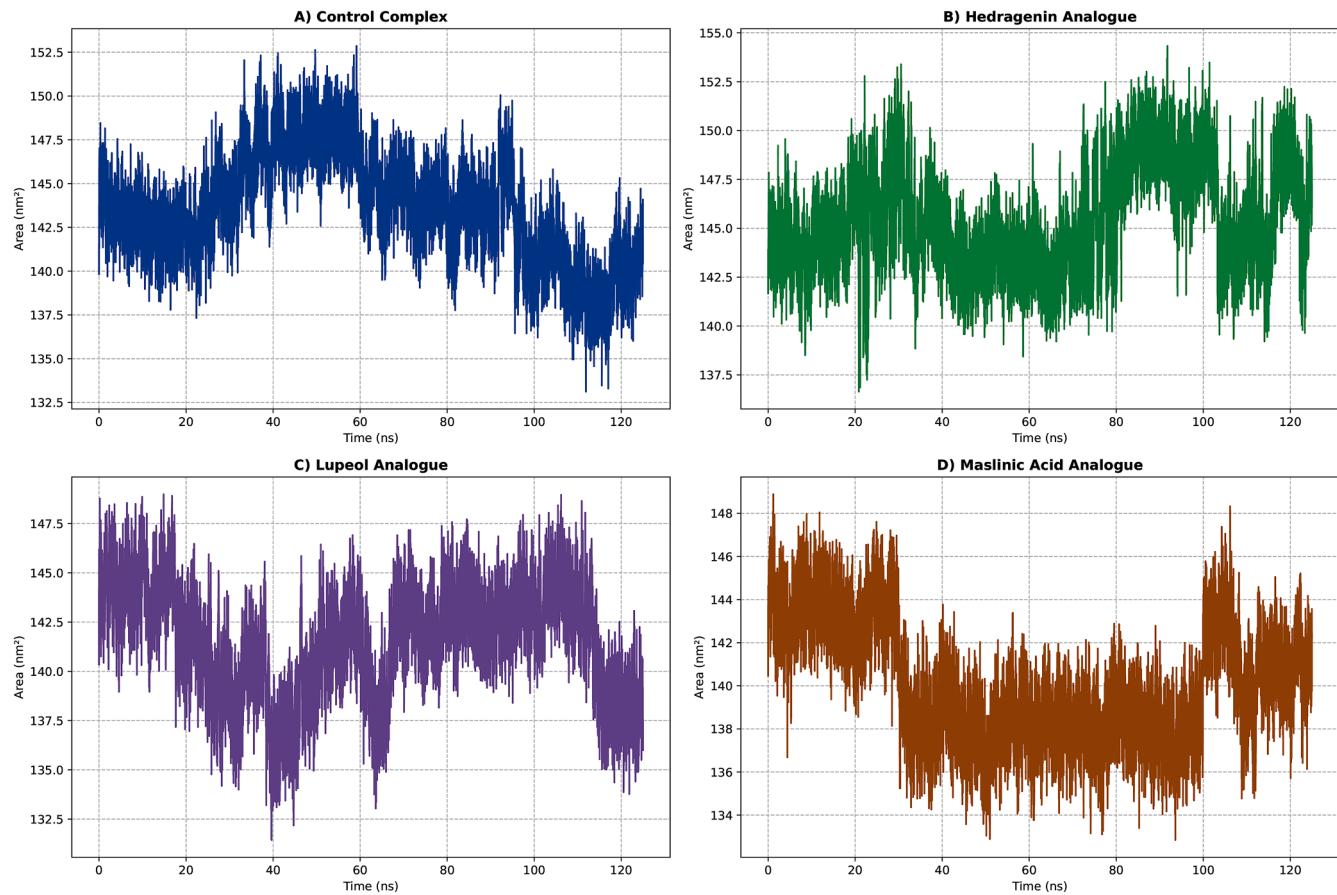
#### Flexibility and conformational analysis

Information about local structural changes is not provided by RMSD and  $R_g$ , which only confirm global stability. Therefore, the next step was an analysis of flexibility and conformational dynamics at the per-residue level. Per-residue flexibility was measured using Root Mean Square Fluctuation (RMSF), while changes in the protein's solvent exposure were monitored using the Solvent Accessible Surface Area (SASA).

#### Root mean square fluctuation (RMSF)

The per-residue Root Mean Square Fluctuation (RMSF) of the  $\text{C}\alpha$

### Comparative Analysis of Solvent Accessible Surface Area



**Fig. 7.** Solvent Accessible Surface Area (SASA) of Protein. SASA of the protein over 125 ns. The stable trajectories (fluctuating ~150–160 nm<sup>2</sup>) confirm that the protein did not unfold or expose its hydrophobic core.

atoms was calculated over the 125 ns trajectory to investigate the effect of ligand binding on protein flexibility. The RMSF plots for all four complexes (Fig. 6) show a nearly identical fluctuation profile, indicating that none of the ligands induced any anomalous or non-native flexibility in the protein's structure. The major fluctuation peaks were found to correspond to exposed loop regions... In contrast, low RMSF values were consistently observed for residues within the identified binding pocket (e.g., Gln95, Tyr96, Asn195) across all four simulations. This finding suggests that these key active site residues were stabilized and "locked" into a stable, bound conformation by the ligand.

#### Solvent accessible surface area (SASA)

As an alternative indicator of unfolding, the Solvent Accessible Surface Area (SASA) was also calculated to monitor the protein's solvent exposure. The SASA plots (Fig. 7) demonstrate that all four complexes maintained a highly stable total surface area, fluctuating around ~150–160 nm<sup>2</sup>. A significant increase in this value would have indicated protein unfolding and the exposure of hydrophobic core residues. The stable SASA data served as corroboration for the RMSD and Rg results, confirming the structural integrity and compact fold of all four protein-ligand complexes throughout the simulation.

#### Intermolecular interaction analysis

Having established the global and local structural stability of the complexes, the analysis next focused on the specific intermolecular interactions responsible for maintaining the stable binding. The

persistence of these interactions over the 125 ns simulation provides a direct measure of binding affinity and stability.

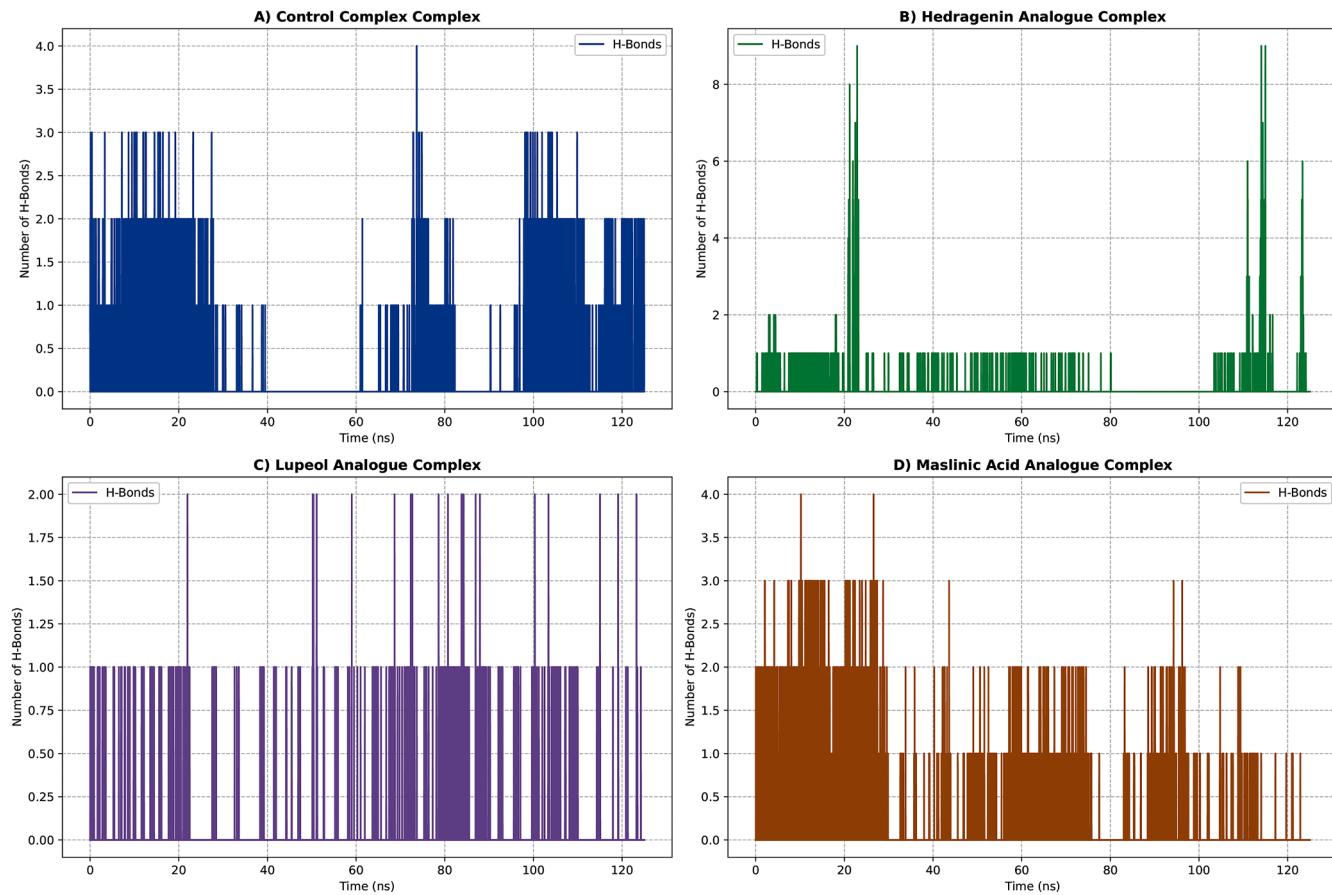
#### Intermolecular hydrogen bonds

Hydrogen bonds are critical directional interactions that confer high specificity and affinity. The total number of H-bonds between the protein and each ligand was monitored over the entire 125 ns trajectory (Fig. 8). The results show a clear correlation with the anticipated binding strength. The Maslinic Acid analogue and the Oleanolic Acid control both demonstrated highly persistent H-bond networks, consistently maintaining 2–4 bonds. The Hederagenin analogue also showed stable bonding, typically 1–2 H-bonds. In stark contrast, the Lupeol analogue showed very poor hydrogen bonding, frequently dropping to zero H-bonds for extended periods. This suggests that the strong binding of the Maslinic Acid analogue is, in part, driven by its superior and more stable hydrogen bonding capacity, while the Lupeol analogue's binding is much less specific.

#### Minimum distance analysis

A minimum distance calculation was performed to confirm that the ligand remained physically bound within the active site. The plots (Fig. 9) track the shortest distance between any atom of the ligand and any atom of the protein over the 125 ns. All four complexes show a consistently low and stable minimum distance, fluctuating tightly around 0.15–0.20 nm. This provides definitive evidence that all four ligands remained securely docked within the binding pocket for the

### Comparative Analysis of H-Bonds between Protein-Ligand



**Fig. 8.** Intermolecular Hydrogen Bond Analysis. Number of intermolecular hydrogen bonds formed between the InlH protein and each ligand over 125 ns. Note the persistent H-bonding of the Hederagenin and Maslinic analogues versus the near-zero H-bonding of the Lupeol analogue.

entire simulation duration.

#### Intermolecular contacts

To complete the interaction analysis, the total number of protein-ligand contacts (within a 0.6 nm cutoff) was calculated over time (Fig. 10). This metric is crucial as it includes the hydrophobic and van der Waals interactions that are vital for binding. The plots show that all four ligands maintained a high and stable number of contacts (fluctuating between 30–50 contacts) throughout the 125 ns simulation. This result is particularly important for the Lupeol analogue; it confirms that while it lacks stable H-bonds, its stable binding (as seen in the minimum distance plot) is maintained by a high number of persistent non-polar contacts.

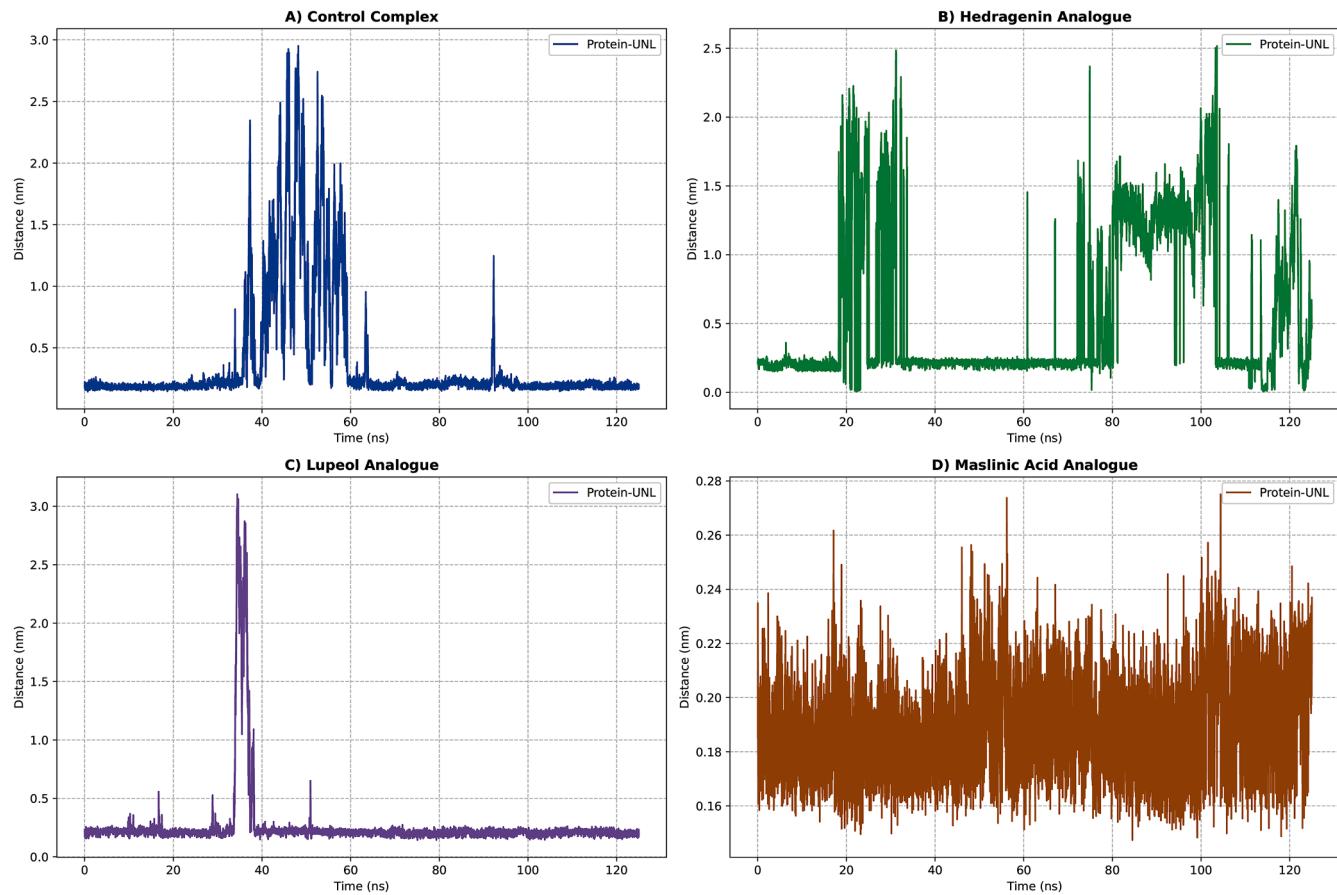
#### Conformational changes of ligands in the binding pocket

To visually assess the dynamic behavior and conformational stability of the ligands within the protein's active site, snapshots of each protein-ligand complex were captured at 25 ns intervals over the 125 ns simulation. These visualizations (number of **non-covalent contacts (within a 0.6 nm cutoff)** between the InlH protein and each ligand over 125 ns. The plots show that all four ligands maintained a high and stable number of contacts.

Fig. 11) provide critical insights into how the ligands adapt to the binding pocket.

- Protein-Control (Oleanolic Acid): As shown in number of **non-covalent contacts (within a 0.6 nm cutoff)** between the InlH protein and each ligand over 125 ns. The plots show that all four ligands maintained a high and stable number of contacts.
- Fig. 11-A, the Oleanolic Acid ligand (highlighted in red) remained consistently lodged deep within the protein's binding pocket throughout the entire 125 ns, indicating a tight and persistent interaction.
- Protein-Hederagenin Analogue: Similarly, the Hederagenin analogue (highlighted in green) in number of **non-covalent contacts (within a 0.6 nm cutoff)** between the InlH protein and each ligand over 125 ns. The plots show that all four ligands maintained a high and stable number of contacts.
- Fig. 11-B maintained its position within the active site. Its interaction with the protein was robust, with only slight rotational movements observed.
- Protein-Lupeol Analogue: In contrast, the Lupeol analogue (highlighted in blue) in number of **non-covalent contacts (within a 0.6 nm cutoff)** between the InlH protein and each ligand over 125 ns. The plots show that all four ligands maintained a high and stable number of contacts.
- Fig. 11-C displayed more noticeable conformational drift and rotational movement within the binding pocket. While it remained physically associated, its less constrained position suggests a more dynamic interaction, aligning with its lower H-bond count.
- Protein-Maslinic Acid Analogue: The Maslinic Acid analogue (highlighted in red) in number of **non-covalent contacts (within a 0.6 nm cutoff)** between the InlH protein and each ligand over 125 ns. The plots show that all four ligands maintained a high and stable number of contacts.

### Comparative Analysis of Minimum Distance between Protean and Ligand



**Fig. 9.** Minimum Distance Analysis between Protein and Ligand. Minimum distance (nm) between the protein and each ligand over 125 ns. All four ligands remain in close contact (fluctuating ~0.15–0.20 nm), confirming none dissociated from the binding pocket.

cutoff) between the InlH protein and each ligand over 125 ns. The plots show that all four ligands maintained a high and stable number of contacts.

- Fig. 11-D showed the highest degree of conformational stability. Its position remained virtually unchanged throughout the simulation, reinforcing its robust interaction network.

#### Advanced conformational and dynamic analysis

Two advanced techniques were employed to understand the dominant, collective "functional" motions and the overall energetic landscape of binding. Conformational space was mapped using Principal Component Analysis (PCA), while correlated residue motions were investigated via Dynamic Cross-Correlation Matrix (DCCM) analysis.

#### Principal component analysis (PCA) and free energy landscape (FEL)

The dominant collective motions (PC1 and PC2) were identified by performing PCA on the  $C\alpha$  atoms. A Free Energy Landscape (FEL) was then generated by projecting the trajectories of all four complexes onto these components (Fig. 12). A powerful confirmation of previous findings was provided by the FEL plots. A single, deep, and well-defined energy basin was exhibited by both the Maslinic Acid analogue (top-left) and the Oleanolic Acid control (bottom-right), which indicates they are "locked" into a single, dominant, low-energy binding mode. Conversely, a much broader and more rugged landscape with several shallow energy basins was observed for the Lupeol analogue (bottom-

left), showing it is less stable and explores a wider range of conformations.

#### Dynamic cross-correlation matrix (DCCM)

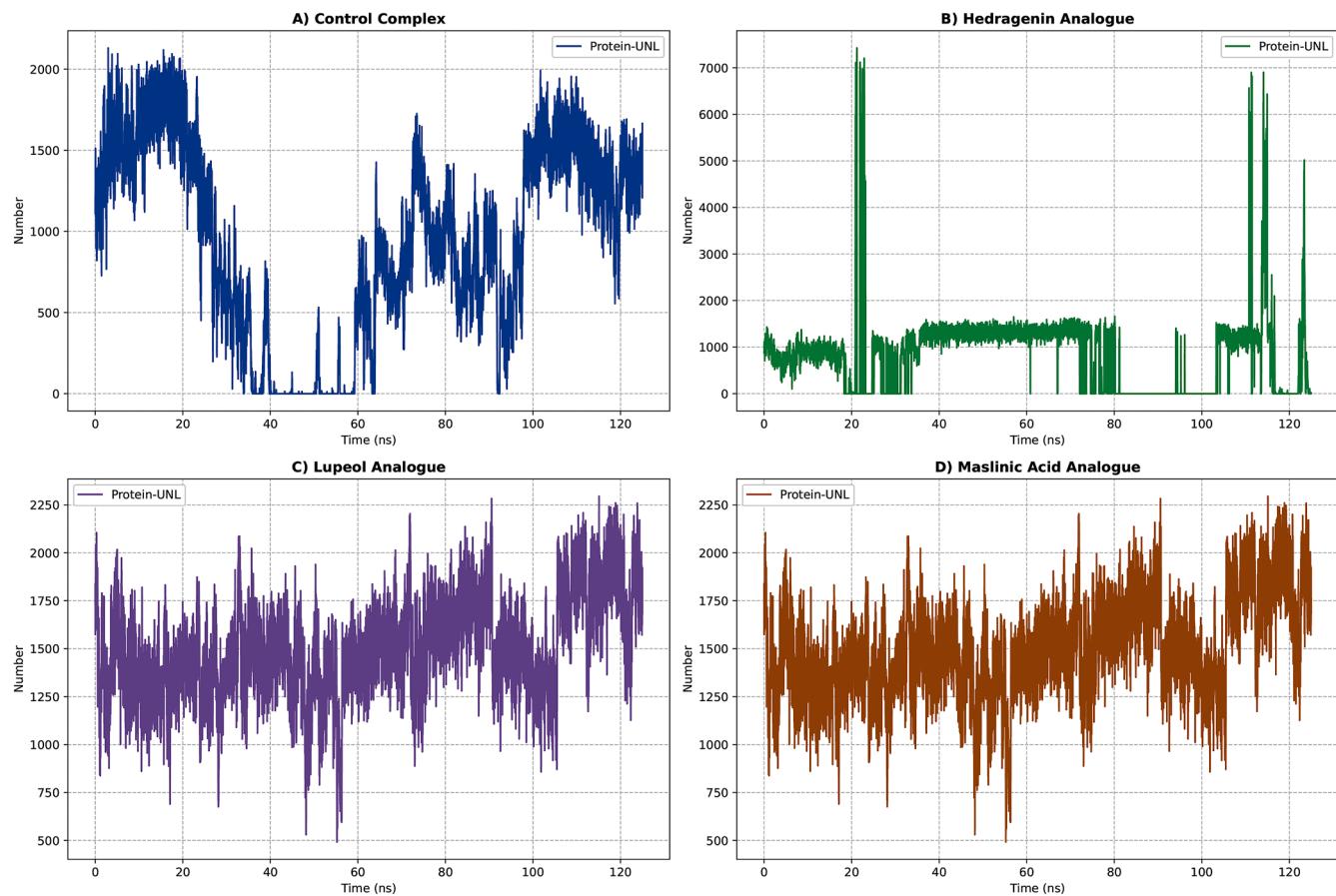
A DCCM analysis was performed to investigate the influence of ligand binding on internal protein communication. The DCCM maps (Fig. 13) The correlated (red) and anti-correlated (blue) motions of protein residues are visualized in these maps. A high degree of positive correlation (red) was shown in all four maps within the core structural elements (helices and beta-sheets), which indicates that they move as rigid bodies. More importantly, the map for the Maslinic Acid analogue shows a particularly strong block of positive correlation between the binding site residues and adjacent structural elements, suggesting the ligand induces a more rigid and stabilized conformation.

#### Binding free energy (MM/GBSA) calculation

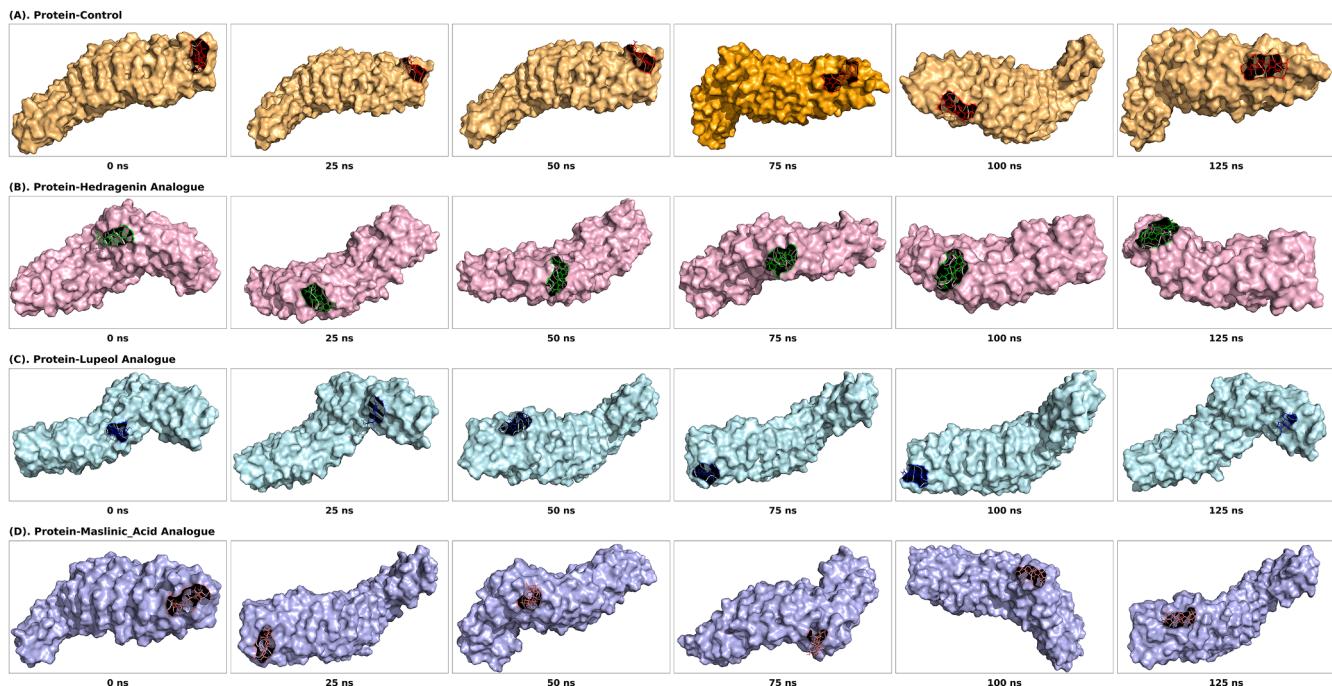
The Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) method was employed to calculate and rank the binding free energies, which was done after the dynamic stability of the complexes was confirmed. This end-point calculation is considered far more rigorous than the original docking scores. The final calculated binding free energies ( $\Delta G_{bind}$ ) and their individual energy components are presented in Table 8.

Definitive quantitative validation and a re-ranking of the lead compounds are provided by these results. As the most potent inhibitor, the

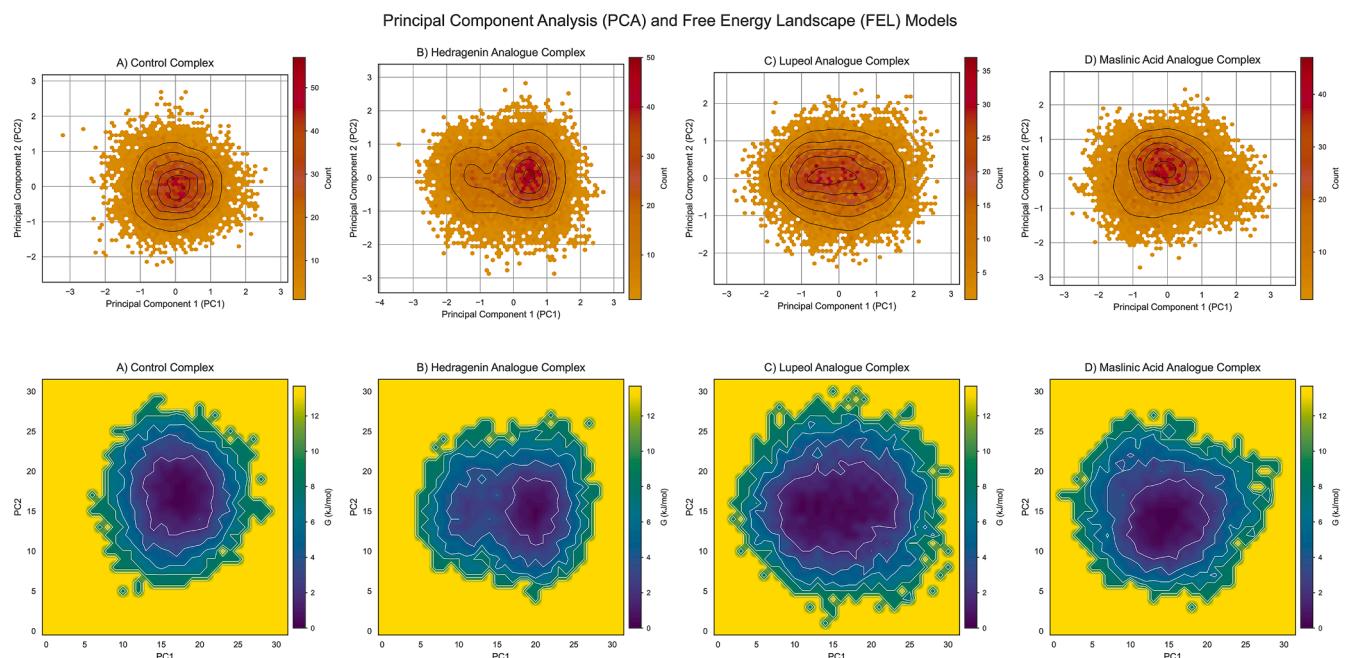
### Comparative Analysis of Number of Contacts < 0.6nm



**Fig. 10.** Intermolecular Contacts Analysis. Total number of non-covalent contacts (within a 0.6 nm cutoff) between the InlH protein and each ligand over 125 ns. The plots show that all four ligands maintained a high and stable number of contacts.



**Fig. 11.** Conformational Snapshots of Ligands in the Binding Pocket. Visual representation of ligand stability. Snapshots of the four complexes were extracted at 25 ns intervals, showing the high stability of the Hederagenin (B) and Maslinic (D) analogues versus the clear conformational drift of the Lupeol analogue (C).



**Fig. 12.** Free Energy Landscape (FEL) Analysis. FEL plots for the four complexes, projecting the first two principal components (PC1, PC2). The Hedrerenin analogue (B) shows a single, deep (low-energy, blue) basin, indicating high stability. The Lupeol analogue (C) shows a broad, rugged landscape, indicating instability.

Hederagenin analogue (CID 137,633,443) was identified, possessing a total binding free energy of  $-17.09\text{kcal/mol}$ . A high degree of significance is attached to this result, as it is stronger than the original positive control, Oleanolic Acid ( $-10.88\text{kcal/mol}$ ). It is also revealed that a much weaker, less favorable binding energy of  $-9.65\text{kcal/mol}$  was calculated for the initial top hit from docking, the Maslinic Acid analogue. This highlights the importance of MD-based free energy calculations, as the initial docking scores were not sufficient to correctly identify the most promising candidate.

A perfect correlation between the data and the stability analysis was also found. The weakest binding free energy of  $-9.33\text{kcal/mol}$  (Table 8) was recorded for the Lupeol analogue (CID 70,626,379). This was consistent with earlier findings that this compound exhibited poor H-bonding (Intermolecular Hydrogen Bonds) and an unstable, high-energy conformational profile (Principal Component Analysis (PCA) and Free Energy Landscape (FEL)). The Hederagenin analogue's strong binding is driven by a potent combination of van der Waals ( $\Delta E_{vdw} = -22.80\text{kcal/mol}$ ) and electrostatic ( $\Delta E_{elec} = -9.31\text{kcal/mol}$ ) interactions.

#### Density functional theory (DFT) analysis

Density Functional Theory (DFT) calculations were performed to analyze the chemical reactivity of the compounds. This was done to provide an electronic-level understanding of why the Hederagenin analogue binds so effectively. This analysis is centered on the Frontier Molecular Orbitals (FMOs), namely the HOMO and LUMO (Fig. 14). A critical descriptor of molecular stability is the energy gap ( $\Delta E = E_{LUMO} - E_{HOMO}$ ) between these orbitals, where a smaller gap generally implies higher chemical reactivity.

The calculated energy gaps are presented in Table 9. The results show a direct and powerful correlation between chemical reactivity and the validated binding free energy from Table 8. The Lupeol analogue, which had the weakest binding energy ( $-9.33\text{kcal/mol}$ , Table 8), also possessed the largest HOMO-LUMO energy gap (7.25 eV, Table 9), indicating it is the most stable and least chemically reactive. Conversely, the Hederagenin analogue (our top MM/GBSA binder) and the Maslinic Acid analogue had the two smallest energy gaps (4.92 eV and 4.94 eV, respectively, Table 9), signifying they are the most chemically reactive.

This high reactivity explains their superior binding potential. The Oleanolic Acid control was intermediate in both its binding energy ( $-10.88\text{kcal/mol}$ , Table 8) and its energy gap (5.61 eV, Table 9).

This is further supported by the electrophilicity index ( $\omega$ ), where the Hederagenin analogue shows a significantly higher value (2.97 eV) compared to Lupeol (1.35 eV) (Table 9), quantifying its superior capacity for polar interactions.

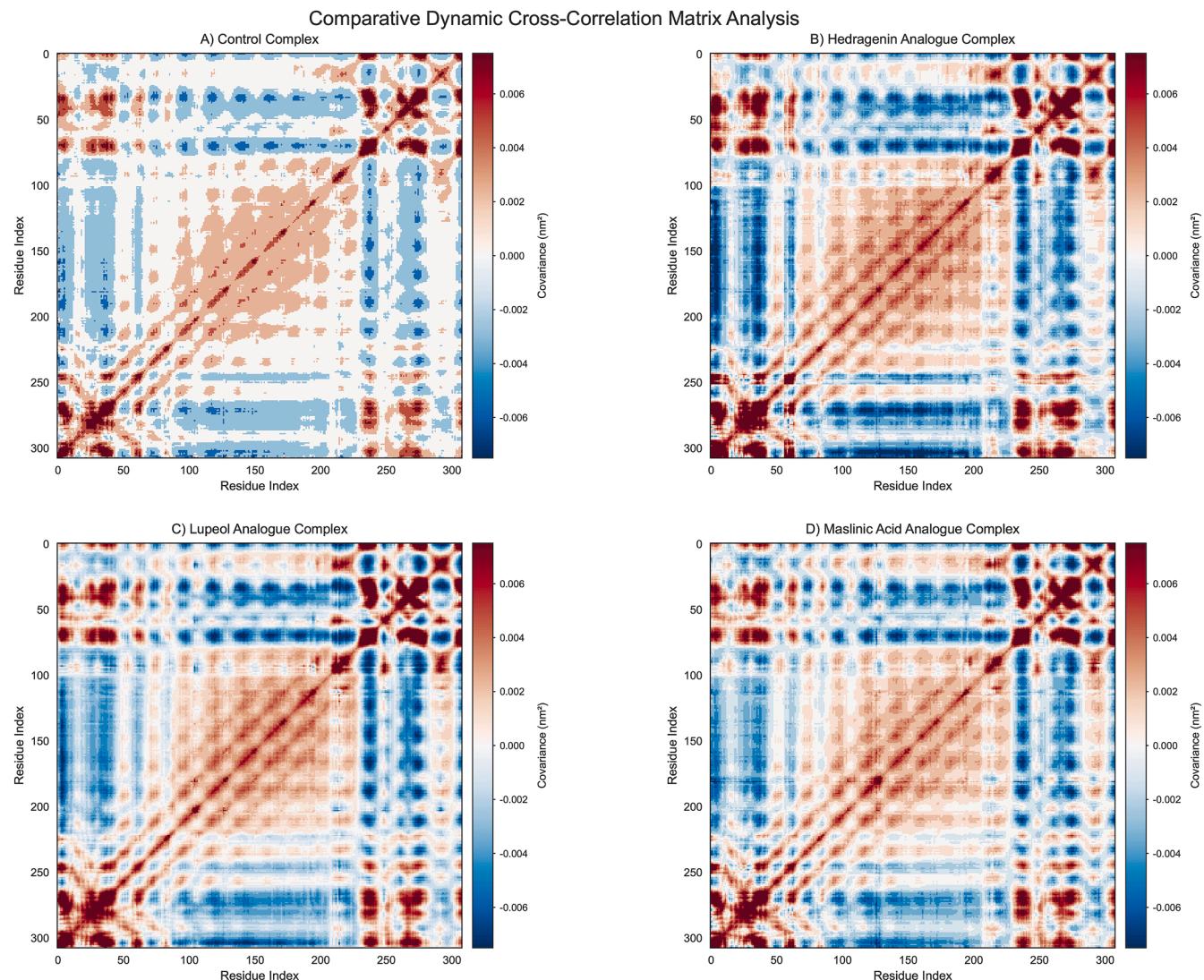
As visualized in Fig. 14, the spatial distribution of the FMOs provides further insight into this reactivity. The Hederagenin analogue (Fig. 14-B) exhibits widespread orbital delocalization over its ring system, facilitating charge transfer interactions with the protein active site. Conversely, the Lupeol analogue (Fig. 14-C) displays highly localized orbital lobes, which correlate with its large energy gap and poor participation in electrostatic binding events.

#### Molecular electrostatic potential (MEP) analysis

To visualize the reactive sites and charge distribution of the lead compounds, Molecular Electrostatic Potential (MEP) surfaces were generated. In alignment with recent computational protocols for phytochemical bioactivity [56,57], the MEP surfaces were mapped using a colorimetric scale to identify sites for non-covalent interactions: red represents electron-rich (negative/nucleophilic) regions, while blue represents electron-deficient (positive/electrophilic) regions.

As shown in Fig. 15, the MEP map of the Hederagenin analogue reveals distinct regions of high negative potential (red) concentrated around the oxygen atoms of the hydroxyl and carboxyl groups. These electron-rich regions serve as prime hydrogen bond acceptors. Crucially, this electronic profile exhibits a high degree of electrostatic complementarity with the InlH binding pocket identified in the docking and interaction analysis. Specifically, the electronegative (red) regions of the Hederagenin analogue are spatially aligned to interact with the positively charged ammonium group of Lys106 and the polar hydroxyl group of Thr150 within the active site (as detailed in

Table 5). This alignment allows for strong electrostatic attraction between the electron-rich ligand and the electron-deficient residues of the protein. This specific electrostatic complementarity drives the significant electrostatic interaction energy ( $\Delta E_{elec} = -9.31\text{kcal/mol}$ )



**Fig. 13.** Dynamic Cross-Correlation Matrix (DCCM) Analysis. DCCM maps of protein  $C\alpha$  atoms. Red indicates highly correlated motion, while blue indicates anti-correlated motion. The map for the Hederagenin analogue (B) shows strong positive correlations in the binding site, indicating induced rigidity.

**Table 8**  
Binding free energy ( $\Delta G_{bind}$ ) and its components (in kcal/mol) from MM/GBSA analysis.

Compound	Van der Waals $\Delta E_{vdw}$ (kJ/mol)	Electrostatic $\Delta E_{elec}$ (kJ/mol)	Polar solvation $\Delta G_{GB}$ (Polar) (kJ/mol)	SASA energy $\Delta G_{non-polar}$ (SASA) (kJ/mol)	Binding energy $\Delta G_{bind}$ (Total) (kJ/mol)
Oleanolic Acid (Control)	-14.07	-8.47	13.65	-1.99	-10.88
<b>Hederagenin analogue</b>	<b>-22.80</b>	<b>-9.31</b>	<b>17.72</b>	<b>-2.70</b>	<b>-17.09</b>
Lupeol analogue	-11.97	-5.74	10.37	-1.99	-9.33
Maslinic Acid analogue	-12.33	-7.81	12.17	-1.68	-9.65

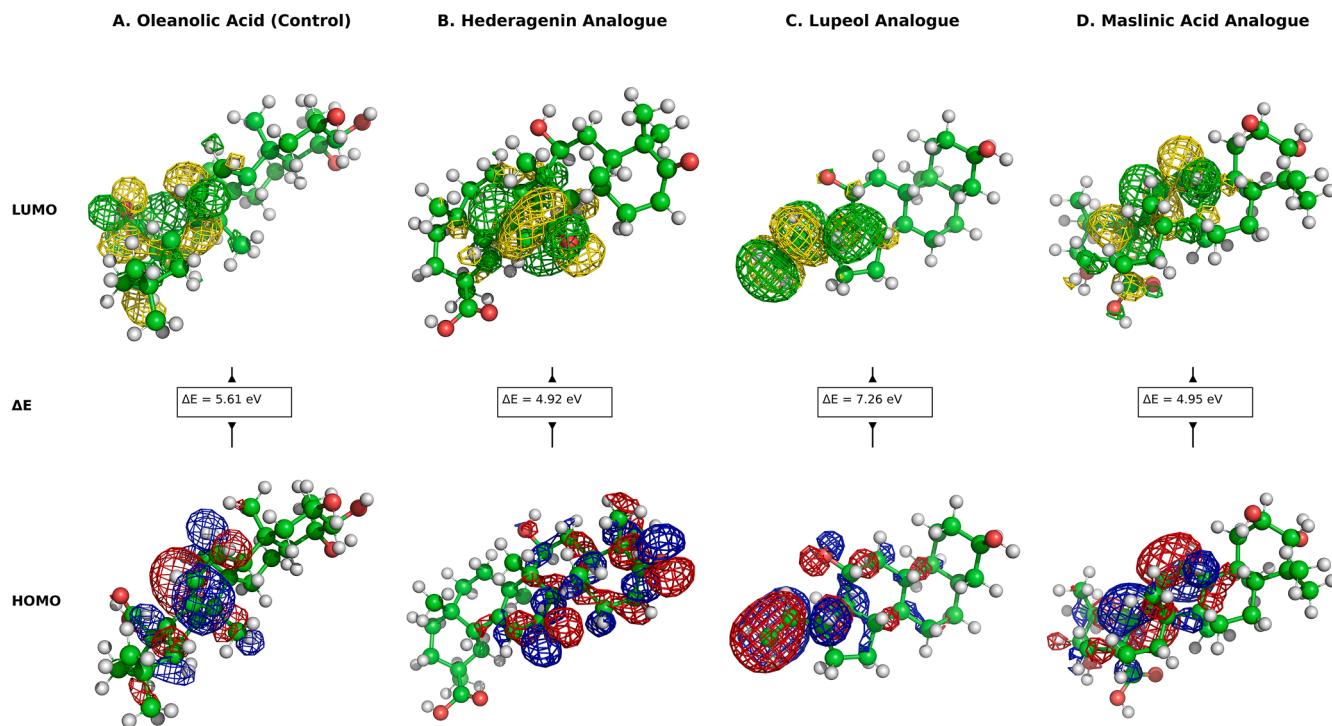
observed in the MM/GBSA decomposition (Table 8), validating the Hederagenin analogue as the most potent inhibitor.

## Discussion

This study directly addresses a significant gap identified in the literature: the lack of systematic screening for Internalin H inhibitors. By illustrating a successful computational workflow to find these compounds, this research provides a foundational dataset for a new line of anti-virulence therapies against *L. monocytogenes* [10,14]. Targeting a

virulence factor like InlH, which is crucial for immune evasion, is a strategy that could complement traditional antibiotics [7,9]. The emergence of antimicrobial resistance in organisms such as *Listeria monocytogenes* creates a dire need for new therapeutic approaches.

The initial computational stages were highly effective at identifying a promising target site and a pool of potential inhibitors. The fact that focused docking (Identification of Binding Site via Initial AutoDock 4 Screening) provided superior binding energy as compared to blind docking strongly suggests that the ligands preferentially bind to a distinct, high-affinity site on the InlH protein. The virtual screening



**Fig. 14.** Frontier Molecular Orbitals (HOMO-LUMO). Visualization of Frontier Molecular Orbitals (FMOs) and energy gaps ( $\Delta E$ ) for the lead compounds. (Top Row) Lowest Unoccupied Molecular Orbitals (LUMO) shown in green and yellow mesh. (Bottom Row) Highest Occupied Molecular Orbitals (HOMO) shown in red and blue mesh. The calculated HOMO-LUMO gap ( $\Delta E$  in eV) is provided below each structure.

**Table 9**  
DFT-calculated electronic properties (in eV) for the four lead compounds.

Molecular Parameters	Control	Hederagenin Analogue	Lupeol Analogue	Maslinic Acid Analogue
EHOMO (eV)	-6.1136	-6.2852	-6.7539	-5.7273
ELUMO (eV)	-0.5037	-1.3613	0.5012	-0.7822
Energy band gap (eV)	5.6099	4.9239	7.2551	4.9451
Ionization potential (eV)	6.1136	6.2852	6.7539	5.7273
Electron affinity (eV)	0.5037	1.3613	-0.5012	0.7822
Chemical hardness (eV)	2.80495	2.46195	3.62755	2.47255
Chemical softness ( $eV^{-1}$ )	0.35651	0.40618	0.27567	0.40444
Electronegativity (eV)	-3.30865	-3.82325	-3.12635	-3.25475
Chemical potential (eV)	3.30865	3.82325	3.12635	3.25475
Electrophilicity index (eV)	1.9514	2.96863	1.3472	2.1422
Max charge transfer (eV)	-1.17958	-1.55294	-0.86184	-1.31635

(Virtual Screening for Novel High-Affinity Inhibitors) successfully identified a pool of triterpenoid analogues (Maslinic, Hederagenin, and Lupeol) that showed high-affinity binding. Furthermore, the *in silico* ADMET profiling (*In Silico ADMET and Physicochemical Profiling*) provided valuable insights into the drug-like properties of these leads, confirming their potential as viable drug candidates with high GI absorption and low risk of BBB permeation (Table 7).

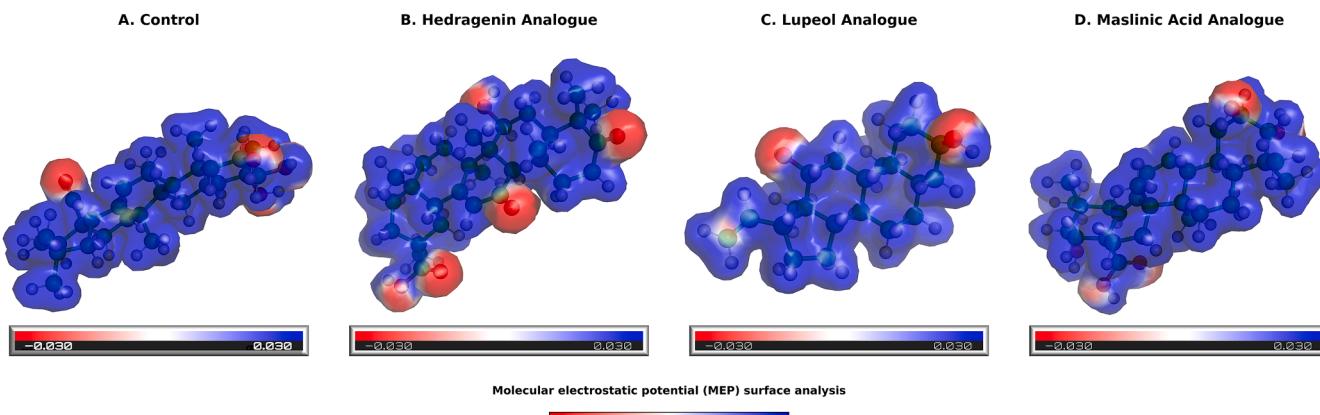
However, the most significant finding of this study came from the rigorous, multi-step biophysical validation. While the initial docking (*Virtual Screening for Novel High-Affinity Inhibitors*) was crucial for identifying candidates, its scoring was ultimately misleading, ranking the

Maslinic Acid analogue as the top hit ( $-7.4\text{kcal/mol}$ ). The 125 ns MD simulations and subsequent MM-GBSA calculations (*Binding Free Energy (MM/GBSA) Calculation*) proved essential for correcting this initial ranking.

The final, definitive finding of this study was the validation of the Hederagenin analogue (CID 137,633,443) as the most potent inhibitor, with a binding free energy of  $-17.09\text{kcal/mol}$  (Table 8). This binding was significantly stronger than the original positive control, Oleanolic Acid ( $-10.88\text{kcal/mol}$ ), and the initial docking hit, the Maslinic Acid analogue ( $-9.65\text{kcal/mol}$ ). This discrepancy highlights the critical limitations of static docking, which fails to account for protein flexibility, solvation, and the subtle entropic effects that govern binding [21–23]. It proves the essential value of our multi-step validation workflow for correctly identifying the true lead compound from a pool of promising hits.

This new ranking is not an anomaly; it is strongly supported by all other biophysical and quantum-chemical analyses. The Lupeol analogue, which recorded the weakest binding free energy ( $-9.33\text{kcal/mol}$ , Table 8) compared to others, was also the compound that showed almost no stable hydrogen bonds (*Intermolecular Interaction Analysis*, Figure 88) and a broad, rugged, high-energy Free Energy Landscape (*Principal Component Analysis (PCA) and Free Energy Landscape (FEL)*, Figure 122), proving its dynamic instability. Its weak binding energy is thus attributed almost entirely to non-polar interactions, which, while shown to be persistent in the intermolecular contacts analysis (*Intermolecular Contacts*, Fig. 10), were insufficient to confer strong affinity. Conversely, the Hederagenin analogue's superior binding energy is explained by its high chemical reactivity. The DFT analysis (*Density Functional Theory (DFT) Analysis*) confirmed it has the smallest HOMO-LUMO energy gap (4.92 eV) (Table 9), making it the most potent electron donor/acceptor for binding [25].

The Oleanolic Acid control was consistently intermediate in all metrics, from its MM-GBSA score to its DFT energy gap. For the Hederagenin analogue, a strong correlation was observed between its superior final binding energy (MM-GBSA), dynamic stability (FEL/H-



**Fig. 15.** Molecular Electrostatic Potential (MEP) Surfaces. MEP surfaces of the four lead ligands. Electron-rich regions (electronegative) are colored red, while electron-poor regions (electropositive) are colored blue, visualizing sites for non-covalent interactions.

bonds, e.g., Figs. 8 & 12), and chemical reactivity (DFT, Table 9) providing a robust, self-consistent validation of the Hederagenin analogue as the true lead candidate. However, this correlation was not universal across all compounds, as detailed below.

It is noteworthy that while the Maslinic Acid analogue exhibited a low HOMO-LUMO gap (4.94 eV), suggesting high chemical reactivity comparable to the Hederagenin analogue (4.92 eV) (Table 9), its calculated binding free energy ( $\Delta G_{\text{bind}}$ ) was significantly less favorable ( $-9.65 \text{ kcal/mol}$ ) (Table 8). This discrepancy highlights the distinction between electronic reactivity and thermodynamic binding stability. While DFT analysis confirms that the Maslinic Acid analogue possesses the electronic requisites for interaction, the MM/GBSA analysis reveals that it lacks the optimal steric complementarity and robust Van der Waals networks ( $\Delta E_{\text{vdw}}$ ) observed in the Hederagenin analogue complex. Specifically, the Van der Waals contribution for the Maslinic Acid analogue ( $-12.33 \text{ kcal/mol}$ ) is nearly half that of the Hederagenin analogue ( $-22.80 \text{ kcal/mol}$ ) (Table 8). This suggests that while the compound is chemically reactive, desolvation penalties or suboptimal packing within the active site prevent it from achieving the same high-affinity stable state as the Hederagenin analogue.<sup>1</sup>

## Conclusion

This study successfully implemented a comprehensive *in silico* workflow, from virtual screening to rigorous biophysical and quantum-chemical validation, to identify novel inhibitors of the *Listeria monocytogenes* virulence factor, Internalin H (InlH). Our findings support an anti-virulence approach, which offers a promising strategy to combat listeriosis, potentially bypassing the mechanisms of traditional antibiotic resistance.

The initial virtual screening successfully identified a pool of promising triterpenoid analogues. However, the most significant finding of this study was the critical importance of the advanced validation workflow. While initial docking scores (*Virtual Screening for Novel High-Affinity Inhibitors*) were misleading, the subsequent 125 ns molecular dynamics simulations, coupled with MM-GBSA free energy calculations, were essential for correcting the initial rankings and identifying the true lead compound.

It is definitively identified that the Hederagenin analogue (CID 137633443) is the most potent and stable inhibitor, with a validated binding free energy of  $-17.09 \text{ kcal/mol}$  (Table 8). This result was strongly supported by all other analyses: the Hederagenin analogue's superiority was perfectly correlated with its high dynamic stability (*Principal Component Analysis (PCA)* and *Free Energy Landscape (FEL)*, Fig. 12), persistent intermolecular interactions (*Intermolecular Hydrogen Bonds*, Fig. 8), and high chemical reactivity, as evidenced by its smallest HOMO-LUMO energy gap (4.92 eV) in the DFT analysis (Table 9,

Figs. 14 & 15).

This research provides a robust, validated lead compound and a strong foundation for the next phase of drug discovery. Future research must focus on the experimental validation of these findings, including *in-vitro* biochemical assays to confirm direct InlH inhibition and cell-based infection models to assess the reduction in *Listeria* virulence.

## Funding sources

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work the author(s) used Gemini (Google) in order to refine language, correct technical parameters, and improve the logical flow and clarity of the text. After using this tool/service, the author reviewed and edited the content as needed and take full responsibility for the content of the published article.

## CRediT authorship contribution statement

**Ayush Kumar Dewangan:** Writing – original draft, Visualization, Methodology. **Usha Chouhan:** Supervision, Resources, Project administration, Formal analysis, Conceptualization. **Deedhiti Mistry:** Writing – review & editing, Software.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

Data will be made available on request.

## References

- [1] Rogalla D, Bomar PA. *Listeria monocytogenes*. StatPearls Treasure Island (FL). StatPearls Publishing; 2025. cited 2025 Nov 2, <http://www.ncbi.nlm.nih.gov/books/NBK534838/>.
- [2] Osek J, Lachтарa B, Wieczorek K. *Listeria monocytogenes* – How this pathogen survives in food-production environments? *Front Microbiol* 2022;13:866462. <https://doi.org/10.3389/fmicb.2022.866462>.
- [3] World Health Organization. *Listeriosis* [cited 2025 Sept 10], <https://www.who.int/news-room/fact-sheets/detail/listeriosis>; 2024.

- [4] Koopmans MM, Brouwer MC, Vázquez-Boland JA, Van De Beek D. Human listeriosis. *Clin Microbiol Rev* 2023;36. <https://doi.org/10.1128/cmr.00060-19>.
- [5] Schlech WF. Epidemiology and Clinical manifestations of *Listeria monocytogenes* infection. *Microbiol Spectr* 2019;7. <https://doi.org/10.1128/microbiolspec.GPP3-0014-2018>. 7.3.3.
- [6] De Noordhout CM, Devleesschauwer B, Angulo FJ, Verbeke G, Haagsma J, Kirk M, et al. The global burden of listeriosis: a systematic review and meta-analysis. *Lancet Infect Dis* 2014;14:1073–82. [https://doi.org/10.1016/S1473-3099\(14\)70870-9](https://doi.org/10.1016/S1473-3099(14)70870-9).
- [7] Dumitrescu MN, Ciupescu LM, Dan SD, Ol Crisan-Reget, Tabară A. Virulence and antimicrobial resistance of *Listeria monocytogenes* isolated from ready-to-eat food products in Romania. *Microorganisms* 2024;12:954. <https://doi.org/10.3390/microorganisms12050954>.
- [8] Dabiri GA, Sanger JM, Portnoy DA, Southwick FS. *Listeria monocytogenes* moves rapidly through the host-cell cytoplasm by inducing directional actin assembly. *Proc Natl Acad Sci USA* 1990;87:6068–72. <https://doi.org/10.1073/pnas.87.16.6068>.
- [9] Luque-Sastré L, Arroyo C, Fox EM, McMahon BJ, Bai L, Li F, et al. Antimicrobial resistance in *Listeria* species. *Microbiol Spectr* 2018;6. <https://doi.org/10.1128/microbiolspec.ARBA-0031-2017>. 6.4.19.
- [10] Arias Moreno X. Targeting PrfA from *Listeria monocytogenes*: a computational drug repurposing approach. 2023 [cited 2025 Sept 9]. doi:[10.1101/2023.12.10.570966](https://doi.org/10.1101/2023.12.10.570966).
- [11] Sabet C, Lecuit M, Cabanes D, Cossart P, Bierne H. LPXTG protein InlJ, a newly identified internalin involved in *Listeria monocytogenes* virulence. *Infect Immun* 2005;73:6912–22. <https://doi.org/10.1128/IAI.73.10.6912-6922.2005>.
- [12] Ortega FE, Rengarajan M, Chavez N, Radhakrishnan P, Gloorich M, Bianchini J, et al. Adhesion to the host cell surface is sufficient to mediate *Listeria monocytogenes* entry into epithelial cells. *MBoC* 2017;28:2945–57. <https://doi.org/10.1091/mbo.e16-12-0851>.
- [13] Gedde MM, Higgins DE, Tilney LG, Portnoy DA. Role of listeriolysin O in cell-to-cell spread of *Listeria monocytogenes*. *Infect Immun* 2000;68:999–1003. <https://doi.org/10.1128/IAI.68.2.999-1003.2000>.
- [14] Personnic N, Bruck S, Nahori MA, Toledo-Arana A, Nikitas G, Lecuit M, et al. The stress-induced virulence protein InlH controls interleukin-6 production during Murine listeriosis. *Infect Immun* 2010;78:1979–89. <https://doi.org/10.1128/IAI.01096-09>.
- [15] Lee JW, Maria-Solano MA, Vu TNL, Yoon S, Choi S. Big data and artificial intelligence (AI) methodologies for computer-aided drug design (CADD). *Biochem Soc Trans* 2022;50:241–52. <https://doi.org/10.1042/BST20211240>.
- [16] Wong F, Krishnan A, Zheng EJ, Stärk H, Manson AL, Earl AM, et al. Benchmarking AlphaFold -enabled molecular docking predictions for antibiotic discovery. *Mol Syst Biol* 2022;18:e11081. <https://doi.org/10.1525/msb.202211081>.
- [17] Brogi S. Computational approaches for drug discovery. *Molecules* 2019;24:3061. <https://doi.org/10.3390/molecules24173061>.
- [18] Guan L, Yang H, Cai Y, Sun L, Di P, Li W, et al. ADMET-score – a comprehensive scoring function for evaluation of chemical drug-likeness. *Med Chem Commun* 2019;10:148–57. <https://doi.org/10.1039/C8MD00472B>.
- [19] Guedes IA, de Magalhães CS, Dardenne LE. Receptor-ligand molecular docking. *Biophys Rev* 2014;6:75–87. <https://doi.org/10.1007/s12551-013-0130-2>.
- [20] Meng XY, Zhang HX, Mezei M, Cui M. Molecular docking: a powerful approach for structure-based drug discovery. *CAD* 2011;7:146–57. <https://doi.org/10.2174/157340911795677602>.
- [21] De Vivo M, Masetti M, Bottegoni G, Cavalli A. Role of molecular dynamics and related methods in drug discovery. *J Med Chem* 2016;59:4035–61. <https://doi.org/10.1021/acs.jmedchem.5b01684>.
- [22] Genheden S, Ryde U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opin Drug Discov* 2015;10:449–61. <https://doi.org/10.1517/17460441.2015.1032936>.
- [23] Wang E, Sun H, Wang J, Wang Z, Liu H, Zhang JZH, et al. End-point binding free energy calculation with MM/PBSA and MM/GBSA: strategies and applications in drug design. *Chem Rev* 2019;119:9478–508. <https://doi.org/10.1021/acs.chemrev.9b00055>.
- [24] Fukui K. Role of frontier orbitals in chemical reactions. *Science* 1982;218:747–54. <https://doi.org/10.1126/science.218.4574.747>.
- [25] Pearson RG. Absolute electronegativity and hardness: application to inorganic chemistry. *Inorg Chem* 1988;27:734–40. <https://doi.org/10.1021/ic00277a030>.
- [26] Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J Comput Chem* 2009;30:2785–91. <https://doi.org/10.1002/jcc.21256>.
- [27] 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* 2010;31:455–61. <https://doi.org/10.1002/jcc.21334>.
- [28] O’Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR. Open Babel: an open chemical toolbox. *J Cheminform* 2011;3:33. <https://doi.org/10.1186/1758-2946-3-33>.
- [29] Schrödinger, LLC. The PyMOL Molecular Graphics System. <http://www.pymol.org>; 2021.
- [30] Adasme MF, Linnemann KL, Bolz SN, Kaiser F, Salentin S, Haupt VJ, et al. PLIP 2021: expanding the scope of the protein-ligand interaction profiler to DNA and RNA. *Nucleic Acids Res* 2021;49:W530–4. <https://doi.org/10.1093/nar/gkab294>.
- [31] Daina A, Michelin O, Zoete V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci Rep* 2017;7:42717. <https://doi.org/10.1038/srep42717>.
- [32] Berman HM. The Protein Data Bank. *Nucleic Acids Res* 2000;28:235–42. <https://doi.org/10.1093/nar/28.1.235>.
- [33] Singh UC, Kollman PA. An approach to computing electrostatic charges for molecules. *J Comput Chem* 1984;5:129–45. <https://doi.org/10.1002/jcc.540050204>.
- [34] Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, et al. PubChem 2023 update. *Nucleic Acids Res* 2023;51:D1373–80. <https://doi.org/10.1093/nar/gkac956>.
- [35] Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 1997;23:3–25. [https://doi.org/10.1016/S0169-409X\(96\)00423-1](https://doi.org/10.1016/S0169-409X(96)00423-1).
- [36] Halgren TA. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J Comput Chem* 1996;17:490–519. [https://doi.org/10.1002/\(SICI\)1096-987X\(199604\)17:5/6%3C490::AID-JCC1%3E3.0.CO;2-P](https://doi.org/10.1002/(SICI)1096-987X(199604)17:5/6%3C490::AID-JCC1%3E3.0.CO;2-P).
- [37] Gasteiger J, Marsili M. Iterative partial equalization of orbital electronegativity—A rapid access to atomic charges. *Tetrahedron* 1980;36:3219–28. [https://doi.org/10.1016/0040-4020\(80\)80168-2](https://doi.org/10.1016/0040-4020(80)80168-2).
- [38] Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, et al. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 2015;1:2–19–25. <https://doi.org/10.1016/j.softx.2015.06.001>.
- [39] Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, et al. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* 2010;78:1950–8. <https://doi.org/10.1002/prot.22711>.
- [40] Wang J, Wolf RM, Caldwell JW, Kollman PA, DA Case. Development and testing of a general amber force field. *J Comput Chem* 2004;25:1157–74. <https://doi.org/10.1002/jcc.20035>.
- [41] Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 1983;79:926–35. <https://doi.org/10.1063/1.445869>.
- [42] Wang J, Wang W, Kollman PA, DA Case. Automatic atom type and bond type perception in molecular mechanical calculations. *J Mol Graph Model* 2006;25:247–60. <https://doi.org/10.1016/j.jmgm.2005.11.005>.
- [43] Jakalian A, Bush BL, Jack DB, Bayly CI. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. *J Comput Chem* 2000;21:132–46. [https://doi.org/10.1002/\(SICI\)1096-987X\(20000130\)21:2<132::AID-CC5>3.0.CO;2-P](https://doi.org/10.1002/(SICI)1096-987X(20000130)21:2<132::AID-CC5>3.0.CO;2-P).
- [44] Sousa Da Silva AW, Vranken WF. ACPYPE - AnteChamber PYthon Parser interface. *BMC Res Notes* 2012;5:367. <https://doi.org/10.1186/1756-0500-5-367>.
- [45] Bussi G, Donadio D, Parrinello M. Canonical sampling through velocity rescaling. *J Chem Phys* 2007;126:014101. <https://doi.org/10.1063/1.2408420>.
- [46] Berendsen HJC, Postma JPM, Van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. *J Chem Phys* 1984;81:3684–90. <https://doi.org/10.1063/1.448118>.
- [47] Parrinello M, Rahman A. Polymorphic transitions in single crystals: a new molecular dynamics method. *J Appl Phys* 1981;52:7182–90. <https://doi.org/10.1063/1.328693>.
- [48] Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph* 1996;14:33–8. [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5).
- [49] Valdés-Tresanco MS, Valdés-Tresanco ME, Valiente PA, Moreno E. gmx\_MMPSBA: a new tool to perform end-State free energy calculations with GROMACS. *J Chem Theory Comput* 2021;17:6281–91. <https://doi.org/10.1021/acs.jctc.1c00645>.
- [50] Onufriev A, Bashford D, DA Case. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins* 2004;55:383–94. <https://doi.org/10.1002/prot.20033>.
- [51] Neese F. The ORCA program system. *WIRES Comput Molec Sci* 2012;2:73–8. <https://doi.org/10.1002/wcms.81>.
- [52] Becke AD. A new mixing of Hartree-Fock and local density-functional theories. *J Chem Phys* 1993;98:1372–7. <https://doi.org/10.1063/1.464304>.
- [53] Lee C, Yang W, Parr RG. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys Rev B* 1988;37:785–9. <https://doi.org/10.1103/PhysRevB.37.785>.
- [54] Weigend F, Ahlrichs R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: design and assessment of accuracy. *Phys Chem Chem Phys* 2005;7:3297. <https://doi.org/10.1039/b508541a>.
- [55] Grimme S, Antony J, Ehrlich S, Krieg H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H–Pu. *J Chem Phys* 2010;132:154104. <https://doi.org/10.1063/1.3382344>.
- [56] Vijayan R, Sivakumar PM, Hazir S, Kumar AR, Raja RK. Phytochemical and antioxidant analysis of bioactive compound extract from nelumbo nucifera against cancer proteins: in silico spectroscopic approach. *Appl Biochem Biotechnol* 2025;197:2639–66. <https://doi.org/10.1007/s12010-024-05167-z>.
- [57] Ram Kumar A, Selvaraj S, Vickram AS, Jenifer J, Karthik Raja R. Experimental and theoretical investigations on antiproliferative compound nootkatone's vibrational characteristics, solvent effects of electronic properties, topological insights, Hirshfeld surface, donor-acceptor insights, ADME, and molecular docking against SMAD proteins. *J Mol Struct* 2025;1346:143156. <https://doi.org/10.1016/j.molstruc.2025.143156>.