In silico Studies of Compounds Present in *Azadirachta indica (Neem)* and their Ability to Bind HIV Integrase Protein

Zhongrui Zhang, Yin Hei Lau
Sonia Arora (Faculty Advisor)

Abstract

Azadirachta indica (Neem) is an evergreen tree that belongs to the Meliaceae family. It is native to the Indian subcontinent and grows worldwide. It is also known as the "village pharmacy" in India for its wide range of therapeutic and pharmacological properties. An *in vitro* study indicated that *A. indica* showed anti-HIV properties. However, the exact mechanism for the supposed anti-HIV properties remains unknown. This study aimed to construct an *in-silico* database of the compounds present in *A. indica* and propose a computational analysis of these compounds against HIV integrase. We performed a thorough literature search to gather relevant information on the plant compounds, including chemical structure, location within the plant, extraction method, and percent yield of each compound found in the plant. We took a comprehensive approach to closely study the binding pockets of HIV integrase and performed molecular docking on *A. indica* compounds using Molecular Operating Environment. A deductive analysis of the docking energies of these compounds revealed thirty potential binders against HIV integrase proteins. We further validated these binders by comparing the ligand interactions to known inhibitors using Ligplot+, which identified the presence of numerous hydrogen bonds and hydrophobic interactions at the protein binding pocket. In conclusion, we propose an underlying binding potential for several *A. indica* compounds with HIV integrase, yielding a potential mechanism for the anti-HIV activity of *A. indica*.

Key terms: Neem, anti-HIV, medicinal plant, molecular docking

1 Introduction

Azadirachta indica (family: Meliaceae), commonly known as neem, is a large, fast-growing, tropical evergreen tree that has been widely used in traditional medicine since prehistoric times (Abdelhady et al., 2015; Paul, Prasad, & Sah, 2011; Sadeghian & Mortazaienezhad, 2007). *A. indica* is indigenous to the Indian subcontinent and is cultivated in at least 30 countries worldwide (Abdelhady et al., 2015). In India, *A. indica* is also known as "the village pharmacy," "the wonder tree," "nature's drug store," and "the life-giving tree" (Hossain et al., 2013; Patel et al., 2016; Paul et al., 2011). All parts of the tree can be used in disease treatments due to the presence of various phytochemicals.

Medicinal plants play an important role in the health of human society. *A. indica* is a medicinal plant with a broad spectrum of therapeutic applications. Bioactive phytochemicals, such as flavonoids, terpenoids, tannins, carbohydrates, and proteins, provide *A. indica* with its healing properties. Many unique compounds have been identified and isolated from all parts of *A. indica* (Sarah et al., 2019). Azadirachtin, nimbin, gedunin, and quercetin are some of the most studied compounds in *A. indica*. These compounds carry various biological and pharmacological properties such as antimicrobial, antiviral, antifungal, antimalarial, anti-inflammatory, antiulcer, and anticancer properties (Jerobin et al., 2015 and Paul et al., 2011). Clinical trials have affirmed various therapeutic properties of neem. For example, neem bark extract was found to control gastric hypersecretion as well as gastroesophageal and gastroduodenal ulcers (Bandyopadhyay et al., 2004). A recent randomized controlled trial also found the therapeutic potential of neem in preventing COVID-19 infection (Nesari et al., 2021). Therefore, *A. indica* has been used in the treatment of fe-

ver, malaria, intestinal infections, inflammation, arthritis, and skin diseases (Abdelhady et al., 2015; Anyaehie, 2009). Because of its medicinal properties in disease prevention and treatment, the United Nations proclaimed *A. indica* as the "Tree of the 21st Century" (Hossain et al., 2013).

The human immunodeficiency virus (HIV) is a retrovirus that attacks the human immune system. The genetic material of the retrovirus is inserted into the host genome by the retroviral integrase during the process of integration (Komal et al., 2020 and Smith & Daniel, 2006). As a result of a weakened immune system, various symptoms — such as fever, cough, swollen lymph nodes, mouth ulcers, and muscle aches — can develop. If HIV is not treated properly, it can lead to severe diseases such as tuberculosis, cryptococcal meningitis, cancers, and acquired immunodeficiency syndrome (AIDS) (World Health Organization [WHO], 2021; Centers for Disease Control and Prevention [CDC], 2021). Hence, urgent treatment is necessary for HIV patients. According to the WHO, there were approximately 37.7 million HIV cases in 2020. However, due to the development of resistance to current medication targeting integrase, the investigation of new integrase inhibitors is needed (Mesplède et al., 2012). In a previous *in vitro* study conducted by Udeinya et al. (2004), a fractionated acetone-water extract prepared from *A. indica* showed anti-cytoadhesion activity, which protects lymphocytes against invasion by HIV and suggests the anti-retroviral property of this plant.

In-silico screening, or computer-aided drug design (CADD), has become a crucial part of the modern drug discovery process. It uses a variety of bioinformatics applications and algorithms to efficiently screen for potential drug candidates and significantly reduce the time and resources needed in the traditional lab-bench-based drug delivery process (Rodrigues and Schneider, 2015). In addition, the aforementioned algorithms can be used to predict the pharmacological properties and interactions of molecules. Molecular docking is a type of CADD that predicts the protein-ligand interaction between the drug target and the drug candidate. It runs computer simulations of the potential drug candidates (ligands) with different 3D postures interacting with the drug targets (proteins) and measures the favorableness of such interactions in terms of binding energy. Such predictions could therefore be used as the first step of drug candidate screening and can eliminate unlikely candidates within a relatively short time frame using fewer resources.

Despite recent findings on *A. indica's* anti-HIV potential, the exact mechanism of action is still unknown. Previous studies from our lab have focused on several anti-HIV targets such as HIV protease and reverse transcriptase. However, none of these studies have led to conclusive data (unpublished observations). Therefore, this study aimed to construct an *in-silico* database of the compounds present in *A. indica* and propose a computational analysis of these compounds against HIV integrase — one of the important proteins in the HIV life cycle — to investigate the potential inhibitory activity in reducing viral load.

2 Methods

Literature review and database building

A literature search on *A. indica* was first conducted to collect common compounds present in this plant with readily available structures. For each of these compounds, the percent yield and the location within the plant containing the highest abundance of compounds were also collected. The three-dimensional (3D) structures of the compounds were collected on PubChem (Kim et al., 2020) (SDF format) or ChemSpider (MOL format). For compounds without readily available 3D structures on these public sources, the 2D structures were collected and converted to 3D models using Discovery Studio (D.S.) Visualizer (BIOVIA & Dassault Systèmes, 2017). Hydrogen atoms were added to all 3D compound structures. The geometry of each compound was cleaned using the built-in Minimize Structure tool in UCSF Chimera (Pettersen et al., 2004) to reduce the internal energies. All optimized *A. indica* compounds were saved in a MOL2 format and ready for molecular docking.

Protein visualization and optimization

The 3D protein structures of HIV integrase were retrieved from the RCSB Protein Data Bank to serve as the target model of investigation (Berman et al., 2000). The human protein structures with higher resolution and known bounded ligands were prioritized during the collection process. Minor protein processing was performed to optimize the protein model and minimize errors. All selected proteins underwent energy optimization and geometry cleaning with the Structural Preparation and Protonate 3D tool in Molecular Operating Environment (MOE) ("Molecular Operating Environment (MOE), 2019.01," 2022), and the processed proteins were saved in MOE format. A pre-docking binding pocket analysis was also performed using Ligplot+(Laskowski & Swindells, 2011) to collect baseline protein-ligand interactions between the known inhibitor compounds and the HIV integrase proteins.

Docking parameter optimization and baseline building

The minimized HIV integrase protein files were re-docked using MOE Docking. Re-docking was performed as a suitability control experiment, which involved taking out the originally bounded ligands in each protein file and docking them back into the protein binding pocket with various binding parameters to identify the most optimal conditions for the experiment. The default placement method of Triangle Matcher with London dG scoring system and the refinement method of Rigid Receptor with GBVI/WSA dG scoring system were used as the docking methods. These methods and scoring systems are known to give reliable results, and the results were estimated in terms of free energy reported in kcal/mol (Corbeil et al., 2012; Galli et al., 2014). However, other parameters such as receptor region, docking site, and number of docked poses were tested for the most optimized parameter combinations for each protein target, which was measured by the RMSD values between the original ligand and the re-docked models. The docking energies for the most suitable docking parameters (often resulted in the lowest RMSD values) were also recorded for baseline purposes.

A. indica compound docking

A MOE database file in MDB format was created with the name and the structure in MOL2 format of each optimized *A. indica* compound. The optimized docking parameters were used albeit the MOE compound database file, which was used as the docking ligands. The docked poses with the best docking energies (most negative) for each compound were recorded. For each *A. indica* compound, the average docking energies against all protein targets were compared with the average docking energies in optimization. Since the docking energy was measured in free energy, any *A. indica* compounds with more negative average docking energies than optimization was more thermodynamically favorable to bind and therefore identified as potential HIV integrase binders.

Post-docking analysis

Ligplot+ was used to perform a post-docking analysis on all binders in complex with their protein binding pockets. Two of the most common and relatively strong protein-ligand interactions — hydrogen bonds and hydrophobic interactions — were examined to help explain and verify the favorable docking energies obtained by the binder compounds.

3 Results

Compound database building

A comprehensive database comprised of 50 compounds present in *A. indica* was created. Figure 1 shows the structures of a few compounds present in the database. In addition to the 3D structures of each compound, the database also collected the locations where these compounds are found on the plant, a brief categorization of each compound, the extraction methods, and the corresponding percent yield found in the plant (Table 1).

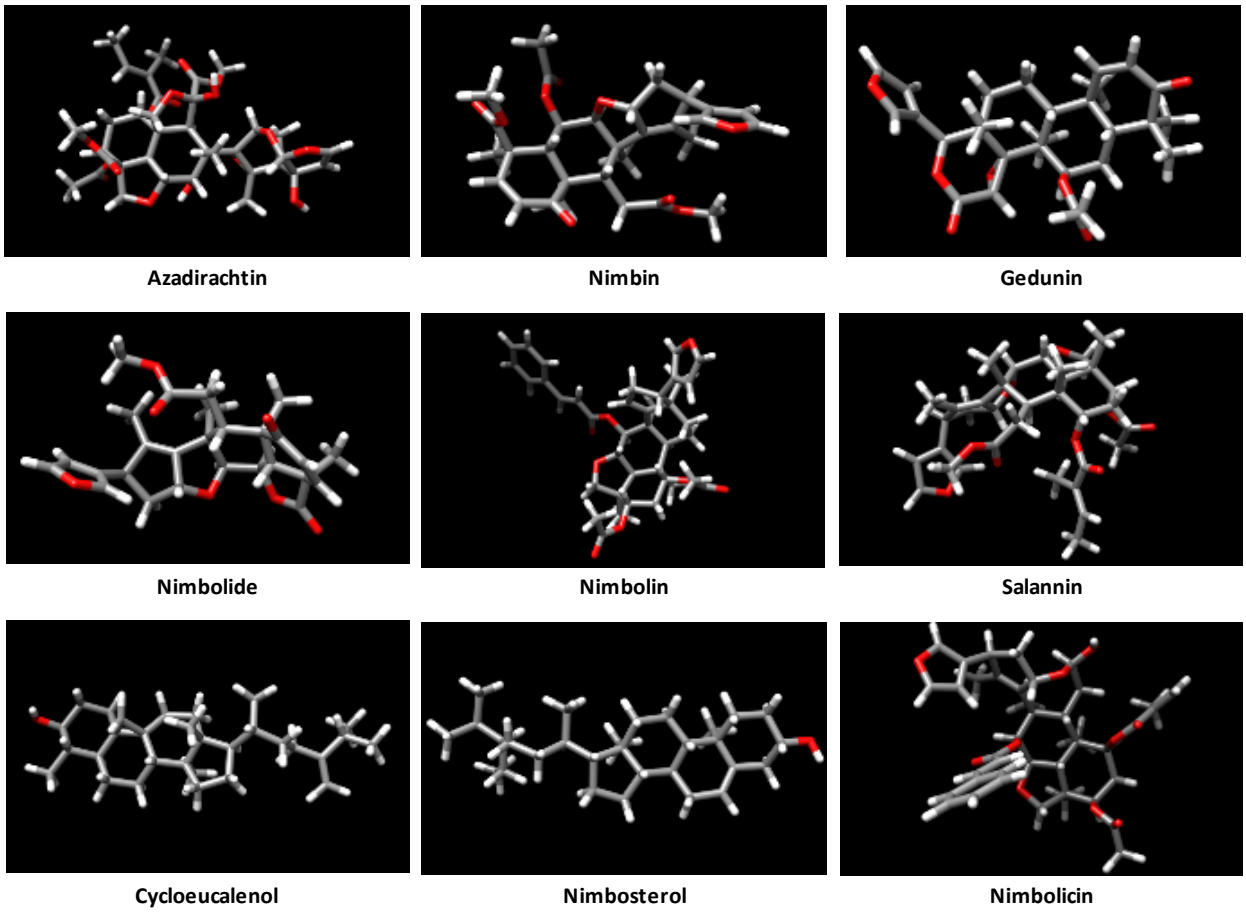

Figure 1. *The 3D structures of nine representative A. indica compounds collected in the database. Grey: carbon; white: hydrogen; red: oxygen.*

Protein data collection and binding pocket analysis of HIV integrase

Four HIV integrase proteins (PDB IDs: 1QS4, 3NF6, 3NF7, 6WC8) were selected (Goldgur et al. 1999; Gorman et al. 2020; Peats et al. 2010). These PDB files had the highest resolution, were derived from human targets, and contained at least one known HIV integrase inhibitor. Clustal Omega multiple protein sequence alignment revealed that these selected integrase proteins were at least 95% identical between any two proteins (Sievers et al. 2011).

The difference in protein sequence was due to the presence of 2-4 unique mutation sites in each protein. The number of files used in this study was a balance between accuracy and resource, as previous studies from our lab had demonstrated that using four target files was sufficient to generate reliable results. Binding pocket analysis of each of the four proteins revealed the presence of numerous hydrogen bonds and hydrophobic interactions between the known inhibitor ligands and the integrase proteins, as shown in Table 2.

TABLE 1: *In-silico compound database for A. indica. All data was collected from publicly available journal sources as indicated in Reference column. The percent yields corresponded to the extraction method as listed. N/A: Data not available.*

	Compound	Source	Extraction method	Percent Yield	Reference
1	Azadirachtin	Flower, fruit, leaf, and seed	Aqueous	0.1-0.3% in seed	(Biswas et al, 2002; Kaushik, 2021; Morgan, 2009; Paul et al., 2011; Ponnusamy et al., 2015; Sadeghian & Mortazaienezhad, 2007; Singh et al., 2017)
2	Isomargolonone	Bark	N/A	N/A	(Biswas et al., 2002; Singh et al., 2017)
3	Azadiradione	Fruit, leaf, seed	Aqueous	0.3% in leaf	(Paul et al., 2011; Ponnusamy et al., 2015; Sadeghian & Mortazaienezhad, 2007)
4	Epicatechin	Bark	N/A	N/A	(Biswas et al., 2002; Singh et al., 2017)
5	Mahmoodin	Seed oil	N/A	N/A	(Biswas et al., 2002)
6	Azadirone	Fruit, leaf, seed oil	Aqueous	2.46% in leaf	(Paul et al., 2011; Ponnusamy et al., 2015; Sadeghian & Mortazaienezhad, 2007)
7	Flavanone	Flower	N/A	N/A	(Nakahara et al., 2003)
8	Margolone	Bark	N/A	N/A	(Biswas et al., 2002; Singh et al., 2017)
9	Catechin	Bark	N/A	N/A	(Biswas et al., 2002; Singh et al., 2017)
10	Nimbin	Leaf, seed oil, trunk and root bark	Aqueous	2.6% in leaf	(Biswas et al., 2002; Kaushik et al., 2021; Paul et al., 2011; Ponnusamy et al., 2015; Sadeghian & Mortazaienezhad, 2007; Singh et al., 2017)
11	Gedunin	Leaf, Seed oil	N/A	N/A	(Anand, 2017; Biswas et al., 2002; Paul et al., 2011; Ponnusamy et al., 2015; Sadeghian & Mortazaienezhad, 2007; Singh et al., 2017)
12	Nimbinin	Leaf, seed oil, trunk and root bark	N/A	N/A	(Koul, Isman, & Ketkar, 1990; Paul et al., 2011)
13	Nimbolide	Leaf, Seed oil	Aqueous	2.20% leaf	(Biswas et al., 2002; Kaushik et al., 2021; Sadeghian & Mortazaienezhad, 2007; Singh et al., 2017)
14	Nimbidin	Leaf, Seed	N/A	N/A	(Biswas et al., 2002; Koul et al., 1990; Singh et al., 2017)
15	Nimbolin A	Trunk wood	N/A	N/A	(Paul et al., 2011)
16	Nimbolin B	Trunk wood	N/A	N/A	(Paul et al., 2011)
17	Quercetin	Flower and leaf	N/A	N/A	(Kaushik et al., 2021; Paul et al., 2011)
18	Salannin	Leaf, seed oil	Aqueous	5.6% in leaf	(Paul et al., 2011; Ponnusamy et al., 2015; Sadeghian & Mortazaienezhad, 2007)
19	Nimbidol	Leaf	N/A	N/A	(Anand, 2017)
20	Cycloeucalenol	Wood oil	N/A	N/A	(Paul et al., 2011)

21	Nimbosterol (beta-sitosterol)	Leaf, wood oil	N/A	N/A	(Kaushik et al., 2021)
22	Nimbinone	Bark	N/A	N/A	(Ara, Siddiqui, Faizi, & Siddiqui, 1988)
23	Nimbolicin	Bark	N/A	N/A	(Read & French, 1993)
24	Margocin	Root bark	N/A	N/A	(Ara et al., 1990)
25	Gallic acid	Bark	N/A	N/A	(Biswas et al., 2002; Singh et al., 2017)
26	2-methyl-5-ethylfuran	Leaf	Butanol	4.8273%	(Hossain et al., 2013)
27	Arabinose	Bark	N/A	N/A	(Kumar et al., 2017)
28	m-toluylaldehyde	Leaf	Methanol	22.7669%	(Hossain et al., 2013)
29	2-methyl-benzaldehyde	Leaf	Butanol	11.8674%	(Hossain et al., 2013)
30	Levoglucosenone	Leaf	Butanol	7.1217%	(Hossain et al., 2013)
31	Methyl isoheptadecanoate	Leaf	Hexane Chloroform Methanol	2.1921% 11.6299% 12.2749%	(Hossain et al., 2013)
32	Methyl petroselinate	Leaf	Hexane	11.2380%	(Hossain et al., 2013)
33	Phytol	Leaf	Hexane Ethyl Acetate Chloroform	2.6170% 61.2401% 10.0515%	(Hossain et al., 2013)
34	Butyl palmitate	Leaf	Hexane	6.6981%	(Hossain et al., 2013)
35	Isobutyl stearate	Leaf	Hexane	4.2521 %	(Hossain et al., 2013)
36	Oxalic acid	Leaf	Hexane	13.7094%	(Hossain et al., 2013)
37	Methyl 14-methylpentadecanoate	Leaf	Methanol Ethyl Acetate Chloroform Butanol	38.1251% 6.4278% 31.8674% 13.4471%	(Hossain et al., 2013)
38	Hexahydrofarnesyl acetone	Leaf	Ethyl Acetate	2.5888%	(Hossain et al., 2013)
39	Lineoleoyl chloride	Leaf	Methanol Chloroform Butanol	26.8329% 11.3587% 13.6057%	(Hossain et al., 2013)
40	Nonacosane	Leaf	Chloroform Butanol	20.6575% 12.8752%	(Hossain et al., 2013)
41	Stearic acid	Kernel Oil	N/A	18%	(Do et al., 2022)
42	Palmitic acid	Kernel Oil	N/A	16.9%	Do et al., 2022)
43	Oleic acid	Kernel Oil	N/A	45.9%	(Do et al., 2022)
44	Linoleic acid	Kernel Oil	N/A	15.69%	(Do et al., 2022)
45	Pyroligneous acid	Heartwood	N/A	38.4%	(Kumar et al., 2017)
46	Hentriacontane	Leaf	Butanol	13.9887	(Hossain et al., 2013)
47	Heptacosane	Leaf	Hexane	8.1010%	(Hossain et al., 2013)
48	Octacosane	Leaf	Hexane	7.0926	(Hossain et al., 2013)
49	Eicosane	Leaf	Hexane	10.0136	(Hossain et al., 2013)
50	Nonadecane	Leaf	Hexane	3.7587%	(Hossain et al., 2013)

TABLE 2: *Binding pocket analysis of the four selected HIV integrase proteins. The hydrogen bonds and hydrophobic interactions were identified using Ligplot+. All ligands present in these proteins had shown existing binding activity toward HIV integrase.*

100: 1-(5-chloroindol-3-yl)-3-hydroxy-3-(2h-tetrazol-5-yl)-propenone
IMV: 5-[(2-oxo-2,3-dihydro-1H-indol-1-yl)methyl]-1,3-benzodioxole-4-carboxylic acid
CIW: 5-[(5-chloro-2-oxo-2,3-dihydro-1H-indol-1-yl)methyl]-1,3-benzodioxole-4-carboxylic acid
TQM: {5-(3-fluorophenyl)-2-[(thiophen-2-yl)ethynyl]-1-benzofuran-3-yl}acetic acid

PDB #	Ligand	Hydrogen Bond	Hydrophobic		
1QS4	100 (5ClTEP)	Thr66	Asp64	Asn155	
		Lys159	Gln148	Lys156	
			Ile151		
			Glu152		
3NF6	IMV	Glu170	Gln95	Ala128	Lys173
		His171	Tyr99	Ala129	Met178
		Thr174	Leu102	Trp132	
			Thr125	Ala169	
3NF7	CIW		Val77	Val150	Leu158
			Val79	Ser153	His183
			Gly82	Met154	
			Ile84	Glu157	
6WC8	TQM		Gln95	Leu102	
			Glu96	Ala128	
			Ala98	Ala129	
			Tyr99	Trp132	

DOCKING PARAMETER OPTIMIZATION AND BASELINE BUILDING

Docking optimization was performed on each of the four HIV integrase targets. The re-docked models were compared with the original ligands to identify the best docking parameters to use for A. indica compounds. After the optimization process, the protein atoms without the surrounding solvent were set as the docking receptor. The protein residues within the 5Å space of the original ligand were defined as the protein active sites for docking. Thirty placement poses and five refinement poses were deemed the best parameters for later studies. For each of the four protein targets, these parameters yielded RMSD values of 1.99, 0.35, 0.73, 1.63, respectively, which were low enough to generate accurate docked results (FIGURE 2). The average docking energy of redocked ligands was -5.6358 kcal/mol.

MOLECULAR DOCKING OF A. INDICA COMPOUNDS INTO HIV INTEGRASE BINDING POCKET

Molecular docking was performed on all 50 *A. indica* compounds against each of the four HIV integrase proteins using the optimized docking parameters. After a comparison between the average docking energy for the plant compounds and the redocked ligands, 30 *A. indica* compounds were predicted to have a more favorable binding energy and were identified as potential HIV integrase binders (TABLE 3).

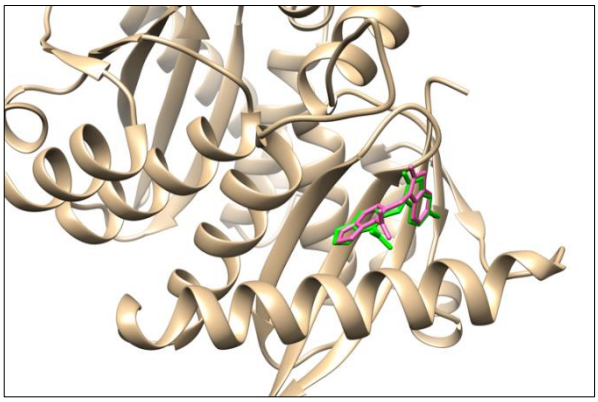

FIGURE 2A

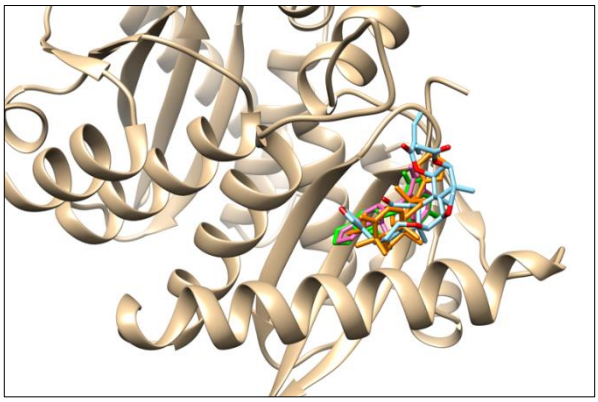

FIGURE 2B

FIGURE 2: *Ligand-bounded HIV integrase binding pocket (PDB# 3NF7).* (A) *Docking optimization with original ligand (green) and redocked ligand (pink).* (B) *A. indica compound salannin (blue) and gedunin (orange).*

TABLE 3: *The docking energy for each of the A. indica compounds and the redocked ligands. Each of these energy measures was the average docking energy of each compound against all four HIV integrase proteins (PDB# 1QS4, 3NF6, 3NF7, 6WC8). Highlighted compounds were identified as binders.*

** Redocked ligands from docking optimization.*

	Compound	Average Docking Energy (kcal/mol)			Compound	Average Docking Energy (kcal/mol)
0	Redocked*	-5.6358		26	2-methyl-5-ethylfuran	-4.1258
1	Azadirachtin	-6.3591		27	Arabinose	-4.1509
2	Isomargolonone	-5.3900		28	m-toluylaldehyde	-4.1068
3	Azadiradione	-5.4921		29	2-methyl-benzaldehyde	-3.9820
4	Epicatechin	-5.2782		30	Levoglucosenone	-3.5953
5	Mahmoodin	-5.9337		31	Methyl isoheptadecanoate	-6.2995
6	Azadirone	-5.6070		32	Methyl petroselinate	-6.4152
7	Flavanone	-5.0024		33	Phytol	-6.4347
8	Margolone	-5.3343		34	Butyl palmitate	-6.5009
9	Catechin	-5.2375		35	Isobutyl stearate	-6.7462
10	Nimbin	-6.0113		36	Oxalic acid	-3.0358
11	Gedunin	-5.8379		37	Methyl 14-methylpentadecanoate	-6.3077
12	Nimbinin	-5.6339		38	Hexahydrofarnesyl acetone	-6.0264
13	Nimbolide	-5.6445		39	Lineoleoyl Chloride	-6.2090
14	Nimbidin	-5.4625		40	Nonacosane	-7.4238
15	Nimbolin A	-6.6793		41	Stearic acid	-6.3773
16	Nimbolin B	-6.3832		42	Palmitic acid	-6.1378
17	Quercetin	-4.9746		43	Oleic acid	-6.2499
18	Salannin	-5.9750		44	Linoleic acid	-6.3096
19	Nimbidol	-5.1154		45	Pyroligneous acid	-3.3098
20	Cycloeucalenol	-6.0074		46	Hentriacontane	-7.3965
21	Nimbosterol	-6.3417		47	Heptacosane	-7.0865
22	Nimbinone	-5.1792		48	Octacosane	-7.3098
23	Nimbolicin	-6.4237		49	Eicosane	-6.4206
24	Margocin	-5.4672		50	Nonadecane	-6.1945
25	Gallic acid	-4.2497				

Post-docking analysis

A post-docking analysis was performed on all thirty *A. indica* binders against each of the four protein targets. The Ligplot+ images revealed the intermolecular interactions of these binders in complex with the HIV integrase binding pockets (Figure 3). Most of the binders were shown to be surrounded by large, hydrophobic clusters. Some hydrogen bonds were also observed with some *A. indica* compounds. The common interacting residues of a few representative compounds are shown in Table 4.

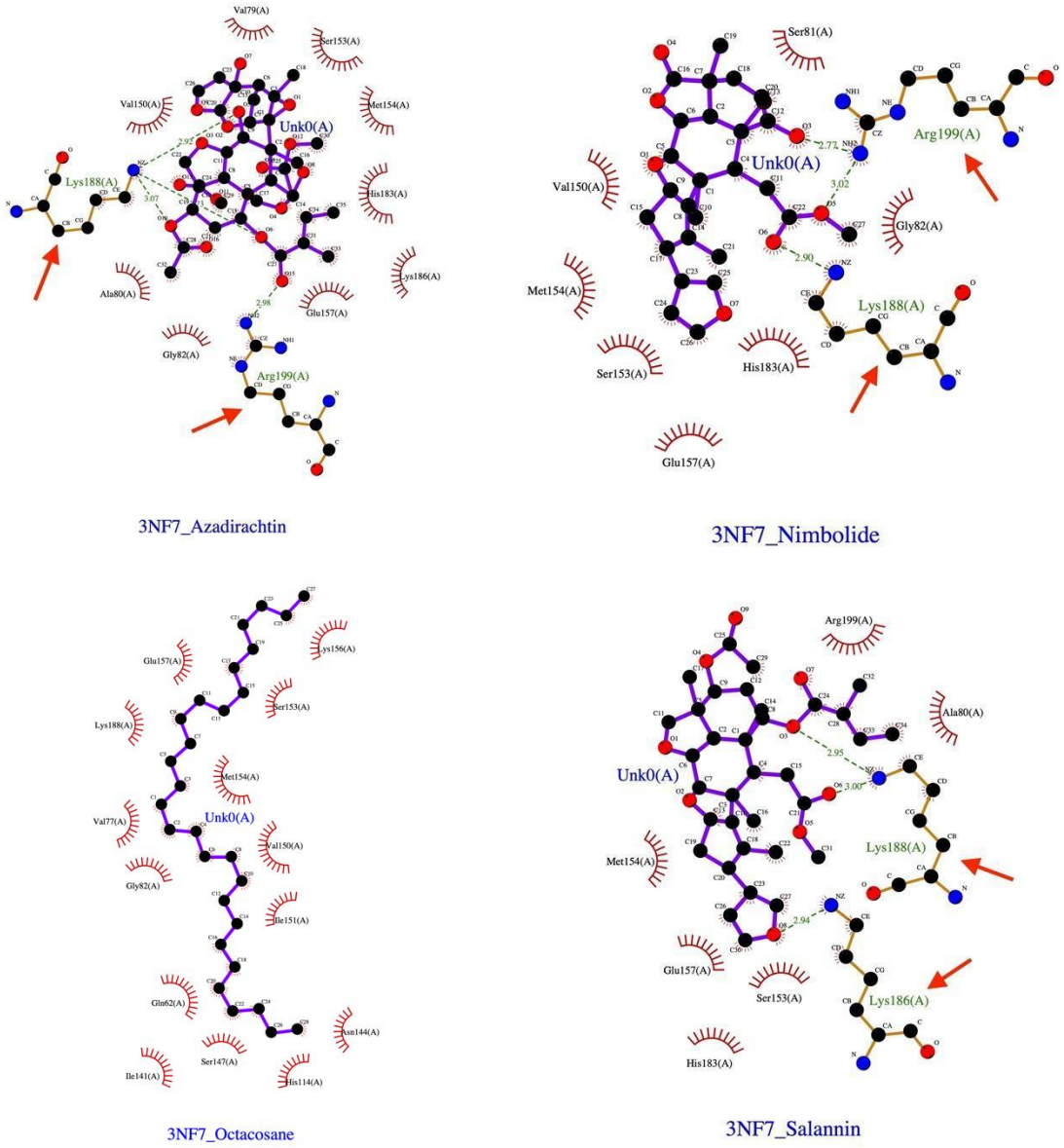

Figure 3: *Representative figure of the post-docking analysis on A. indica compound bounded HIV integrase proteins. Semi-circles indicate protein residue involved in hydrophobic interactions. Arrow-pointed orange compounds indicate protein residue involved in hydrogen bonds. The purple compounds are A. indica compounds of interest.*

TABLE 4: *The common interacting HIV integrase residues of a few representative A. indica binders and the original redocked ligand via either hydrophobic interactions or hydrogen bonds.*

**Residues involved in hydrogen bonds.*

Compound	Average Docking Energy (kcal/mol)	Common Interacting Residues
Re-docked	-5.6358	Val77, Val79, Gly82, Ile84, Val150, Ser153, Met154, Glu157, Leu158
Azadirachitin	-6.3591	Val79, Gly82, Val150, Ser153, Met154, Glu157, His183, Lys188*, Arg199*
Isobutyl Stearate	-6.7462	Val77, Val79, Gly82, Val150, Met154, Glu157, His183, Lys186, Lys188
Nimbolicin	-6.4237	Val77, Val79, Ala80*, Gly82, Val150, Met154, Glu157, His183, Lys186, Lys188, Arg199
Nimbolide	-5.6445	Ser81, Gly82, Val150, Ser153, Met154, Glu157, His183, Lys188*, Arg199*
Nimbolin A	-6.6793	Val77, Val79, Ala80, Gly82, Val150, Ser153, Met154, Glu157, His183, Lys188, Arg199
Nimbosterol	-6.3417	Val77, Val79, Gly82, Val150, Met154, Glu157, His183, Lys186, Lys188
Octacosane	-7.3098	Val77, Gly82, Val150, Ser153, Met154, Glu157, Lys188
Salannin	-5.975	Ser153, Met154, Glu157, His183, Lys186*, Lys188*, Arg199

4 DISCUSSION & CONCLUSION

A comprehensive *in-silico* database of *A. indica* compounds was created, providing detailed information on many compounds regarding their sources, extraction methods, and percent yields extracted from the plant. Although the exact percent yield of these compounds varies greatly depending on the extraction method, this information could provide valuable insights for later drug discovery stages. The 3D structures collected for each of these compounds were also extensively used in the molecular docking studies against HIV integrase.

In the docking optimization process, the original ligands in each protein file were docked back into the protein binding pockets; the resulting model was referred to as re-docked ligands. The relative position and identity of the re-docked ligands were visually compared with the original ligands to determine the reliability of the docking methods. The re-docked ligands were shown to occupy a highly similar 3D space with the original ligand. This was also quantified via the RMSD values, which measured the average distance between the atoms of the original ligand and the re-docked ligand. Therefore, the low RMSD values also reflected highly similar postures between the predicted model and the original ligand. Both verification methods indicated that the optimized docking parameters and algorithm were highly accurate in predicting the binding affinity of the *A. indica* compounds.

Meanwhile, the average docking energy in optimization was also an important baseline for identifying the potential *A. indica* binders. A closer look at the 30 identified *A. indica* binders revealed that they were bound at the same binding pocket as the original ligands, suggesting a potentially similar allosteric effect. Interestingly, the average docking energies of six compounds (nonacosane, hentriacontane, octacosane, heptacosane, isobutyl stearate, and nimbolin A) were one standard deviation more favorable than the redocked ligands, indicating more efficient binding activity than the original ligands.

Post-docking analysis was performed to explain and validate the favorable docking energies predicted by the docking algorithm. The large hydrophobic clusters surrounding most *A. indica* binders and the presence of hydrogen bonds with some binders were both excellent indicators of strong intermolecular interactions. Some interacting residues in the original ligands, such as Val77, Val79, Gly82, Val150, Ser153, Met154, Glu157, and His183, were commonly retained across many *A. indica* binders. In addition to the retained interactions, most binders also gained new hydrophobic interactions; some of these binders, such as azadirachitin and nimbolide, also gained new hydrogen bonds. The *A. indica* binders that gained new interactions suggest a more potent binding ability to the target proteins compared to the corresponding binding ability of the original ligands. Therefore, the intermolecular interactions present in these new protein-ligand complexes validated the favorable binding energies predicted by the docking algorithms, which supported the identified *A. indica* binders against HIV integrase.

Overall, we have identified 30 out of 50 *A. indica* compounds as binders of HIV integrase proteins. The large proportion of the binder compounds present in this plant provides a feasible explanation of this plant's HIV viral reducing potential. Therefore, we propose a potential mechanism for the anti-HIV activity for *A. indica* which could offer insights into a novel HIV treatment candidate. However, the fact that the docking energies of the existing known binders were used as the cut-off point in *A. indica* binder identification may be a potential limitation this study. The possibility of the *A. indica* compounds with less-than-ideal docking energies binding to the targets still exists. This study only provides a computational screening of the potential HIV drug candidates; further bench testing on promising candidates is still necessary to validate the results. Therefore, future goals include further testing these compounds in a wet lab setting to validate the potential inhibitory potential against HIV integrase. The database created in this study may play an important role in future studies of this plant compared to other biological targets, which in turn enables exploration of other therapeutic targets■

5 Acknowledgements

We would like to express our sincere gratitude to our mentor and research advisor Dr. Sonia Arora for her continuous support throughout the duration of the project. This research work would not be possible without her enthusiasm and knowledge to the topic. Her guidance was always inspiring and this project was a great learning opportunity on *in-silico* approaches, which could have endless applications in drug discovery works.

6 References

[1] Abdelhady, M. I. S., Bader, A., Shaheen, U., El-Malah, Y., Abourehab, M. A. S., & Barghash, M. F. (2015). Azadirachta indica as a Source for Antioxidant and Cytotoxic Polyphenolic Compounds. *Biosciences Biotechnology Research Asia, 12*(2), 1209-1222. doi:10.13005/bbra/1774

[2] Alzohairy, M. A. (2016). Therapeutics Role of Azadirachta indica (Neem) and Their Active Constituents in Diseases Prevention and Treatment. *Evid Based Complement Alternat Med., 2016*. doi:10.1155/2016/7382506

[3] Anand, N. (2017). Antifungal, Contraceptive, Anti-Cancer, Mosquito Repellent Properties of

Azadirachta Indica: A review. *Texila International Journal of Basic Medical Sciences, 2*(1). doi:10.21522/TIJBMS.2016.02.01.Art006

[4] Anyaehie, U. B. (2009). Medicinal properties of fractionated acetone/water neem [Azadirachta indica] leaf extract from Nigeria: a review. *Niger J Physiol Sci., 24*(2). doi:10.4314/njps.v24i2.52926

[5] Ara, I., Siddiqui, B. S., Faizi, S., & Siddiqui, S. (1988). Terpenoids from the stem bark of Azadirachta indica. *Phytochemistry, 27*(6), 1801-1804. doi:10.1016/0031-9422(88)80447-3

[6] Ara, I., Siddiqui, B. S., Faizi, S., & Siddiqui, S. (1990). Tricyclic diterpenoids from root bark of Azadirachta indica. *Phytochemistry, 29*(3), 911-914. doi:10.1016/0031-9422(90)80044-H

[7] Bandyopadhyay, U., Biswas K., Sengupta, A., Moitra, P., Dutta, P., Sarkar, D., Debnath, P., Ganguly, C. K. & Banerjee, R. K. (2004). Clinical studies on the effect of Neem (Azadirachta indica) bark extract on gastric secretion and gastroduodenal ulcer. *Life Sci, 75*(24):2867-78.

[8] Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., . . . Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Research, 28*, 235-242.

[9] BIOVIA, & Dassault Systèmes. (2017). Discovery Studio Visualizer. *San Diego: Dassault Systèmes*.

[10] Biswas, K., Chattopadhyay, I., Banerjee, R. K., & Bandyopadhyay, U. (2002). Biological activities and medicinal properties of Neem (Azadirachta indica). *Current Science, 82*(11), 1336-1345.

[11] Centers for Disease Control and Prevention [CDC]. (2021). About HIV. Retrieved from *HTTPS://WWW.CDC.GOV/HIV/BASICS/WHATISHIV.HTML*

[12] Corbeil, C. R., Williams, C. I., & Labute, P. (2012). Variability in docking success rates due to dataset preparation. Journal of computer-aided molecular design, 26(6), 775–786. *HTTPS://DOI.ORG/10.1007/S10822-012-9570-1*

[13] Council, N. R. (1992). *Neem: A Tree For Solving Global Problems*. Washington, DC: The National Academies Press.

[14] Do, D. N., Pham, H. D., Le, X. T., Le, M. T., Nguyen, N. P., Bui, Q. M. & Luu, X. C. (2022).Mechanical Extraction of Neem Seed Kernel (Azadirachta indica A. Juss.) Harvesting in Ninh Thuan, Viet Nam by Using Hydraulic Pressing: Effect of Processing Parameters. *Materials Science Forum, 1048*, 437-444.

[15] Galli CL, Sensi C, Fumagalli A, Parravicini C, Marinovich M, Eberini I. A computational approach to evaluate the androgenic affinity of iprodione, procymidone, vinclozolin and their metabolites. PLoS One. 2014 Aug 11;9(8):e104822. doi:10.1371/journal.pone.0104822.

[16] Goldgur, Y., Craigie, R., Fujiwara, T., Yoshinaga, T., Davies, D.R. (1999). Core domain of HIV-1 integrase complexed with Mg++ and 1-(5-chloroindol-3-yl)-3-hydroxy-3-(2H-tetrazol-5- yl)-propenone. doi: 10.2210/pdb1QS4/pdb

[17] Gorman, M.A. & Parker, M.W. (2020). HIV Integrase core domain in complex with inhibitor 2-(5-(3-fluorophenyl)-2-(2-(thiophen-2-yl)ethynyl)-1- benzofuran-3-yl)ethanoic acid. doi:10.2210/pdb6WC8/pdb

[18] Hossain, M. A., Al-Toubi, W. A. S., Weli, A. M., Al-Riyami, Q. A., & Al-Sabahi, J. N. (2013). Identification and characterization of chemical compounds indifferent crude extracts from leaves of Omani neem. *Journal of Taibah University for Science, 7*(4), 181-188. doi:10.1016/j.jtusci.2013.05.003

[19] Jerobin, J., Makwana, P., Kumar, R. S., Sundaramoorthy, R., Mukherjee, A., & Chandrasekaran, N. (2015). Antibacterial activity of neem nanoemulsion and its toxicity assessment on human lymphocytes in vitro. *Int J Nanomedicine, 10*(Suppl 1), 77-86. doi:10.2147/IJN.S79983

[20] Kaushik, P., Ahlawat, P., Singh, K., & Singh, R. (2021). Chemical constituents, pharmacological activities, and uses of common ayurvedic medicinal plants: a future source of new drugs. *Advances in Traditional Medicine*. doi:10.1007/s13596-021-00621-3

[21] Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., . . . Bolton, E. E. (2020). PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Research, 49*(D1), D1388-D1395. doi:10.1093/nar/gkaa971

[22] Komal, D., Khushboo, J., Aaftaab, S., Lakshmi, S., & Mallika, A. (2020). Targeting Integrase Enzyme: A Therapeutic Approach to Combat HIV Resistance. *Mini-Reviews in Medicinal Chemistry, 20*(3). doi:10.2174/1389557519666191015124932

[23] Koul, O., Isman, M. B., & Ketkar, C. M. (1990). Properties and uses of neem, Azadirachta indica. *Can. J. Bot., 68*(1), 1-11. doi:10.1139/b90-001

[24] Kumar, B., Bhaskar, D., Rajadurai, M., & Sathyamurthy, B. (2017). In vitro studies on the effect of Azadirachta indica Linn. in lung cancer A549

cell lines. *World J Pharm Pharm Sci, 6*(9), 1627-1640.

[25] Laskowski, R. A., & Swindells, M. B. (2011). LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *Journal of Chemical Information and Modeling, 51*, 2778-2786.

[26] Mesplède, T., Quashie, P. K. & Wainberg, M. A. (2012). Resistance to HIV integrase inhibitors. *Curr Opin HIV AIDS, 7*(5), 401-408. doi:10.1097/COH.0b013e328356db89

[27] Molecular Operating Environment (MOE), 2019.01. (2022). *Chemical Computing Group ULC, 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7*.

[28] Morgan, E. D. (2009). Azadirachtin, a scientific gold mine. *Bioorganic & Medicinal Chemistry, 17*(12), 4096-4105. doi:10.1016/j.bmc.2008.11.081

[29] Nakahara, K., Roy, M. K., Ono, H., Maeda, I., Ohnishi-Kameyama, M., Yoshida, M., & Trakoontivakorn, G. (2003). Prenylated flavanones isolated from flowers of Azadirachta indica (the neem tree) as antimutagenic constituents against heterocyclic amines. *J Agric Food Chem., 51*(22), 6456-6460. doi:10.1021/jf034666z

[30] Nesari, T. M., Bhardwaj, A., ShriKrishna, R., Ruknuddin, G., Ghildiyal, S., Das, A., Pandey, A. K., Chaudhary, N., Soman, G. & Barde, M. (2021). Neem (Azadirachta Indica A. Juss) Capsules for Prophylaxis of COVID-19 Infection: A Pilot, Double-Blind, Randomized Controlled Trial. *Altern Ther Health Med, 27*(S1):196-203.

[31] Patel, S. M., Venkata, K. C. N., Bhattacharyya, P., Sethi, G., & Bishayee, A. (2016). Potential of neem (Azadirachta indica L.) for prevention and treatment of oncologic diseases. *Semin in Cancer Biol., 40-41*, 100-115. doi:10.1016/j.semcancer.2016.03.002

[32] Paul, R., Prasad, M., & Sah, N. K. (2011). Anticancer biology of Azadirachta indica L (neem): a mini review. *Cancer Biol Ther., 12*(6), 467-476. doi:10.4161/cbt.12.6.16850

[33] Peat, T.S., Newman, J., Deadman, J.J., Rhodes, D. (2010). Structural basis for a new mechanism of inhibition of HIV integrase identified by fragment screening and structure based design. doi:10.2210/pdb3NF7/pdb

[34] Pettersen, E., Goddard, T., Huang, C., Couch, G., Greenblatt, D., Meng, E., & Ferrin, T. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. *Journal of Computational Chemistry, 25*(13), 1605-1612.

[35] Ponnusamy, S., Haldar, S., Mulani, F., Zinjarde, S., Thulasiram, H., & RaviKumar, A. (2015). Gedunin and Azadiradione: Human Pancreatic Alpha-Amylase Inhibiting Limonoids from Neem (Azadirachta indica) as Anti-Diabetic Agents. *PLoS One., 10*(10). doi:10.1371/journal.pone.0140113

[36] Read, Michael D. & French, H. James H., eds. (1993). Genetic Improvement of Neem: Strategies for the Future. *Proc. of the International Consultation on Neem Improvement held at Kasetsart University, Bangkok, Thailand, 18 - 22 January 1993*. Bangkok, Thailand: Winrock International. 194 + x pp.

[37] Rodrigues, T., Schneider, G. (2015). Chapter 6 - In Silico Screening: Hit Finding from Database Mining. *The Practice of Medicinal Chemistry, 4*, 141-160. *HTTPS://DOI.ORG/10.1016/B978-0-12-417205-0.00006-7*

[38] Sadeghian, M. M., & Mortazaienezhad, F. (2007). Investigation of Compounds from Azadirachta indica (Neem). *Asian Journal of Plant Sciences, 6*, 444-445. doi:10.3923/ajps.2007.444.445

[39] Sarah, R., Tabassum, B., Idrees, N., & Hussain, M. K. (2019). Bioactive Compounds Isolated from Neem Tree and Their Applications. In *Natural Bio-active Compounds*: Springer.

[40] Sievers, F., Wilm, A., Dineen, D., Gibson, T. J, Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D, Higgins, D. G, (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Molecular Systems Biology, 7. 539. doi: accession:10.1038/msb.2011.75

[41] Singh, H., Kaur, M., Dhillon, J. S., Batra, M., & Khurana, J. (2017). Neem: a magical herb in endodontics. *Stomatological Dis Sci, 1*, 50-54. doi:10.20517/2573-0002.2016.10

[42] Smith, J. A., & Daniel, R. (2006). Following the Path of the Virus: The Exploitation of Host DNA Repair Mechanisms by Retroviruses. *ACS Chem. Biol., 1*(4), 217-226. doi:10.1021/cb600131q

[43] Udeinya, I. J., Mbah, A. U., Chijioke, C. P., & Shu, E. N. (2004). An antimalarial extract from neem leaves is antiretroviral. *Transactions of The Royal Society of Tropical Medicine and Hygiene, 98*(7), 435-437. doi:10.1016/j.trstmh.2003.10.01

[44] World Health Organization [WHO]. (2021). HIV/AIDS. Retrieved from *HTTPS://WWW.WHO.INT/NEWS-ROOM/FACT-SHEETS/DETAIL/HIV-AIDS*

Zhongrui Zhang is a recent graduate from Rutgers University-New Brunswick. He has a B.S. in Biotechnology, bioinformatics from the School of Environmental and Biological Science. He had been conducting research in Dr. Sonia Arora's lab on two projects over the course of two years. One was to utilize *in-silico* techniques to study the compounds present in *Ocimum sanctum* against inflammatory pathways. The other was to investigate the potential anti-HIV activity of *Azadirachta indica* in reducing viral loads. He also assisted PhD students in Dr. James Simon's lab in conducting synthetic and analytical organic chemistry work, where he gained hands-on experience working with plant compounds. These research experiences sparked his interest in drug discovery, and he is currently working at Bristol Myers Squibb within the Biologics Department. In the future, Zhongrui would like to pursue graduate studies in the drug development-related field.

Zhongrui can be reached at zhongrui.zhang@rutgers.edu.

Yin Hei Lau is a graduate of Rutgers University. She has a B.S. degree in Biotechnology – Bioinformatics from the School of Environmental and Biological Science. She has a broad interest in health and medicine and would like to conduct further research in the medical field. Her research, under the guidance of Dr. Sonia Arora, investigated the anti-HIV properties of compounds in *A. indica* through *in-silico* approach. She also worked in Dr. Judith Storch's research lab for more than two years. She assisted a PhD student in functional analysis of enterocyte fatty acid binding proteins (FABP). For independent projects, she studied the hepatic lipid metabolism in the intestine-specific liver FABP (LFABP) knockout mice and the intestinal lipid metabolism in the liver-specific LFABP knockout mice.

Yin Hei can be contacted at: yinhei.lau@rutgers.edu.