Structural and Functional Basis of Red Palm Weevil: Pathogen Interactions Leading Towards Enhanced Crop Production in Date Palms

Meshari Alazmi (✉ ms.alazmi@uoh.edu.sa)

University of Hail <https://orcid.org/0000-0001-9074-1029>

N. Alshammari

University of Hail College of Sciences

Naimah A. Alanazi

University of Hail College of Sciences

Abdel Moneim E. Sulieman

University of Hail College of Sciences

Research Article

Keywords: red palm weevil, date palm, host-pathogen, protein:protein interactions, molecular docking, molecular dynamics

Posted Date: August 23rd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-610617/v2>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Food safety remains a significant challenge despite the growth and development in agricultural research and the advent of modern biotechnological and agricultural tools. Though the agriculturist struggles to aid the growing population's needs, many pathogen-based plant diseases by their direct impact on cell division and tissue development have led to the loss of tons of food crops every year. Though there are many conventional and traditional methods to overcome this issue, the amount and time spend are huge. Scientists have developed systems biology tools to study the root cause of the problem and rectify it. Host-pathogen protein interactions (HPIs) have a promising role in identifying the pathogens' strategy to conquer the host organism. In this paper, the interactions between the host *Rhynchophorus ferrugineus* (an invasive wood-boring pest that destroys palm) and the pathogens *Proteus mirabilis*, *Serratia marcescens*, and *Klebsiella pneumoniae* are comprehensively studied using protein-protein interactions, molecular docking, and followed by 200 ns molecular dynamics simulations. This study elucidates the structural and functional basis of these proteins leading towards better plant health, production, and reliability.

Introduction

For nearly 5000 years, *Phoenix dactylifera* L. Arecaceae (the date palm) is an ancient plant, providing high nutritious food for the Indus valley and Middle East people (1,2). These date palms are considered as livelihood security of the rural farmers as they are extensively grown as economic or ornamental trees (3,4). *Rhynchophorus ferrugineus* (Olivier 1790) or red palm weevil (RPW), the most devastating pest of all palm trees worldwide, belongs to the order Coleoptera of the Curculionoidea family. It was once a native of Melanesia and Southeast Asia, but to an accidental anthropogenic introduction in the Caribbean, USA, Mediterranean Basin, and the Middle East, it has spread across the globe (5). These insects have turned out to be herbivorous crop pests during evolution (by specialized interactions with its host plants) (6,7). The distinct genetic makeup and life cycle of the RPW and its cryptic nature have made it hard to detect and, therefore, difficult to control the pest (8). Food and Agriculture Organization in the United Nations has identified RPW as a category-1 quarantine pest in the Middle East and North Africa (MENA) region as it is a threat to the date palm farmers (9). Currently, around 40 palm species (globally) are affected with *R. ferrugineus* (10). It uses date palm for feeding and mating regulated by a male-produced pheromone consisting of 4(RS)-methylnonan-5-one (ferrugineone) and (4RS, 5RS)-4-methylnonan-5-ol (ferrugineol) (11).

Host-pathogen protein interactions (HPIs) are highly complex processes, often involved in the pathogen's approach to invade and breach the host immune defenses and multiply and endure within the host organism (12,13). Traditional biological research such as isolating and studying small sets of components (14) and other experimental procedures for identifying interacting proteins like binary approaches and co-complex methods (15) are expensive as well as time-consuming (16) and may not provide a proper insight on HPI on a larger scale (14). Pathogens manipulate the cellular functionalities of host organisms via HPI in order to take advantage of the host cell capabilities, leading to infections (17). A major goal of host-pathogen studies is to elaborately study the specific interactions that determine success or failure of the pathogen. While omics studies (transcriptomics and proteomics) can register changes that occur in the infected cell, an unsatisfactory list of alterations alone does not provide meaningful mechanistic insights (18). To overcome this, systems biology approaches, which have been developed to enhance the accuracy, coverage, and efficiency in recognizing protein pairs, are an upcoming tool to support understanding the underlying mechanisms during HPIs (19). Unlike the traditional approach of considering the host or pathogen separately, a systems-level approach (HPI) as a whole, is requisite to reveal the mechanisms of infection. The technological advances in the post-genomic era resulted in the production of large-scale HPI data. Systems biology-based methods for the inference and analysis of HPI metabolism and networking (which shed light on infection mechanisms) are in high demand. The knowledge derived via this protocol largely contribute to the identification of sophisticated therapeutics to prevent or cure infections (17).

Indeed, several different systems biology approaches have demonstrated their effectiveness (20). These approaches depend on an unbiased and complete knowledge of the host/pathogen transcriptomics and use high-throughput data along-with mathematical modelling for anticipating protein-protein interaction (PPI) (21)(22). The rapid development of genomics and other "omics" approaches has dramatically impacted the way we investigated host-pathogen interactions since the turn of the millennium. Various techniques such as next-generation sequencing, stem cell biology, and high-throughput proteomics have transformed the scale and sensitivity with which we interrogate biological samples. These technological upfronts are impacting experimental design in the laboratory (23).

Further computational analyses of protein sequence, structural and genomic profiles have partially revealed the mechanisms of interaction and functional relationships (24,25) between host and pathogen, leading to a deeper understanding of the infection process (26). Some of the techniques such as gene neighbor and gene cluster methods (27,28), interologs (29,30), and phylogenetic profile (31,32) are used for recognizing both PPIs (33,34) and host-pathogen protein interactions (35,36). Many of the techniques and

various computational tools of systems biology are modular, and the resulting data can be integrated in myriad ways to inform the question of biological origin (18).

Based on the literature search, several microorganisms, such as *Bacillus thuringiensis* (37), *Metarhizium anisopliae*, and *Beauveria bassiana* (38,39), have been suggested to play an essential role in controlling the weevil. We hypothesize that *Proteus mirabilis*, *Serratia marcescens*, and *Klebsiella pneumoniae* directly interact with the host (RPW) based on protein-protein interactions. Hence, they might stop the growth of RPW and can eventually lead to its death. In this study, we want to identify proteins that are highly likely to interact, predict their function and 3D structures of the proteins, dock the interacting proteins, and finally do molecular dynamics simulations. Our goal is to find the top candidate enzymes used as a safe bio-pesticide to kill the RPW. We demonstrate structural analysis of molecular networks (utilizing disease modules as a novel strategy) to further discuss possible applications to host–pathogen interactions (HPI). With increasing number of good quality databases and repositories, the prediction of large-scale networks will also be feasible for multidimensional diagnostic studies that allows the researcher to determine what types of data are most informative when predicting a biological outcome of interest and design subsequent experiments accordingly (18)(22).

Materials And Methods

Host and Pathogen Sequence data retrieval

The host sequences, GenBank assembly accession: GCA_014462685.1, were retrieved from the NCBI website (40), using the following link: https://www.ncbi.nlm.nih.gov/assembly/GCA_014462685.1/#/st by clicking on "download assembly", "GenBank" and then choosing the file type "protein fasta (.faa)". The downloaded protein sequences contain more than 25,000 hypothetical proteins, which is their functions not yet annotated. Three pathogens, namely *Proteus mirabilis*, *Serratia marcescens*, and *Klebsiella pneumoniae*, were chosen for the study. The protein sequences for these three pathogens were downloaded in FASTA format from NCBI using <https://www.ncbi.nlm.nih.gov/genome/?term=Proteus+mirabilis+HI4320>, <https://www.ncbi.nlm.nih.gov/genome/?term=Serratia+marcescens>, <https://www.ncbi.nlm.nih.gov/genome/?term=Klebsiella+pneumoniae> links respectively.

Protein-Protein Interaction

PredHPI tool (41) was used for solving and studying the protein-protein interaction. The host and the pathogen's sequence identity and coverage were set to 90% and 80%, respectively, to obtain the best hits. PredHPI, a powerful tool consisting of the integrated framework for both detection and visualization of interaction networks within a single window. This facilitates the trepidation of model and non-model host-pathogen systems to further help the biologists in defining new hypotheses and designing experiments accordingly. PredHPI is built on high-end computational resources capable of handling proteome-scale sequence data from host and pathogen both. Data is displayed in an information-rich interactive visualization GUI, which can be further customized as per user-defined settings (41). **Flowcart 1** shows the working protocol of PredHPI.

Homology Modeling and Protein-protein Molecular Docking

Structures of the best hits for the host proteins and their respective pathogens were modeled using the PHYRE2 web-server (42). All the respective sequences were retrieved from NCBI and were subjected to homology modeling utilizing the intensive mode. Models obtained were subjected to Saves 6.0 webserver (consisting Procheck (43), Verify-3D (44,45), and Errat (46) based validation results) for structure validation (data not shown). After homology modeling, Prey-predator Protein-protein molecular docking was performed using the online web-server HawkDock (47). To interpret the structural feasibility and outcomes of protein-protein interactions (PPIs) is very important to gain a sound understanding of many crucial biological functions of living cells. Computational approaches like protein-protein docking are quite needed in this case to complement existing wetlab techniques. The reliability of a in-silico protein-protein docking experiment is reliant on the ability of the scoring function to accurately decipher the near-native binding structures from a pile of conformations (47,48). HawkDock works on the principle scoring function, called HawkRank (48) which is a novel scoring function designed for the sampling stage of protein-protein docking by adding various energy contributors including van der Waals potentials, electrostatic potentials and desolvation potentials (48).

Molecular Dynamics Simulations

System setup

The protein-protein complexes were prepared for molecular dynamics simulations. Six hundred ns accumulative simulations were performed with the amberff14sb force field (49). Solvated systems were prepared using scripts in the pdb2gmx module of Gromacs 2020.4 (50), and MD simulations using Gromacs 2020.4 (50) were performed. The system consisted of the protein complex, TIP3P water, counter ions Na⁺/Cl⁻. Then, several steps have been done on the system. These are energy minimization, which lasted for around 3200 steps following 1000 pico-second equilibration. The MD production run was set up for 200 ns each, totally to accumulative 600 ns MD production procedure. NPT ensemble was used (1 bar) with a time step of 2 femtoseconds. The temperature has been set up at 300K with a low damping coefficient, while pressure was controlled using the Nose-Hoover Langevin piston. Electrostatics was calculated using the particle mesh Ewald (PME) method. A total cut-off at 12 Å was given for short-range and van der Waals electrostatics. All simulations were replicated twice with initialized random seed to get average scorings.

Data analyses

Data analysis for the produced trajectories was performed using Gromacs scripts previously implemented in Gromacs (50), and data were plotted using GnuPlot (<http://gnuplot.info>). We have also calculated RMSF α alignments for carbons for all residues and structural changes by RMSD throughout the simulation run. Calculation between the hydrogen donor and acceptor was set with a cut-off at 3.2 Å, including the backbone and side-chain. Other analyses such as Radius of gyration (ROG), Solvent accessible surface area (SASA), and H-bond formations upon ligand binding were calculated using Gromacs scripts. RMSD, RMSF, total energy, SASA, Radius of gyration, and bonds were plotted using GnuPlot.

Results And Discussion

Protein-protein interactions for prey and predator systems

Based on the predictions from the PredHPI tool, we found several proteins of the host and pathogens interacting with each other. The host and the pathogens' most critical interacting proteins are tabulated in **Table 1** and **Table 2**, respectively, with their designated function (based on functional analysis with similar proteins of that kind).

Table 1 shows the function of homologous proteins for the host interacting proteins.

S. No.	Host homologous protein	Function
1	P61586	Transforming protein RhoA GTP-bound
2	P60953	Cell division control protein 42 homolog GTP-bound
3	P84243	Variant histone H3 replaces conventional H3 in a wide range of nucleosomes in active genes.
4	Q10567	Mediate both the recruitment of Catherin to membranes and the recognition of sorting signals within the cytosolic tails of transmembrane cargo molecules.
5	P61158	ATP-binding component
6	P0A6E4	Catalytic activity
7	P0ACD5	Iron-sulfur cluster assembly scaffold protein IscU

Table 2 shows the function of homologous proteins for the pathogen interacting proteins.

S. No.	Pathogen homologous protein	Function
1	P0A825	Serine hydroxymethyltransferase Catalyzes the reversible interconversion of serine and glycine
2	P0A698	The UvrABC repair system catalyzes the recognition and processing of DNA lesions.
3	P0A6B8	Master enzyme that delivers sulfur to several partners involved in Fe-S cluster assembly,
4	A0A384L4X0	Putative modulator of DNA gyrase
5	A0A384LGT3	Cofactor: Proline-specific aminopeptidase
6	Q8ZAG5	Magnesium transport protein CorA
7	Q8D1R0	Periplasmic serine endoprotease DegP-like
8	Q74R28	sn-glycerol-3-phosphate import ATP-binding protein UgpC
9	A0A384KKB9	Putative ferric siderophore receptor
10	Q8D184	Phospho-2-dehydro-3-deoxyheptonate aldolase

Table 3, Table 4, and Table 5 demonstrated the interaction profile and similarity score between the host and the three different pathogens (*Klebsiella pneumoniae*, *Proteus mirabilis*, and *Serratia marcescens*) proteins. We can observe the protein KAF7276687.1 as a common protein targeted by all three pathogens. It directly suggests the crucial and vital role of this protein in the host body. As per **Table 1**, the homolog of this protein is P0ACD5 which has a function of Iron-sulfur cluster assembly scaffold protein IscU. This protein shows that this protein works in the structural integrity of the host organism.

Table 3 shows the top protein interactions between *Proteus mirabilis* and red palm weevil (*Rhynchophorus ferrugineus*) based on the PredHPI tool. The host similarity is between the host protein sequence and the homologous protein of the host protein. The pathogen similarity is between the pathogen protein sequence and the homologous protein of the pathogen protein.

Host protein	host sequence length	Homolog of Host	Host similarity (%)		Pathogen Protein	Homolog of pathogen	pathogen sequence length	Pathogen similarity (%)	
			Local	global				local	global
KAF7276551.1	134	P0A6E4	90.70	26.40	WP_004243862.1	P0A825	417	95.00	95.00
KAF7276551.1	134	P0A6E4	90.70	26.40	WP_004249165.1	P0A698	944	94.60	94.20
KAF7276687.1	131	P0ACD5	93.60	89.30	WP_004243849.1	P0A6B8	404	94.80	94.80
KAF7269484.1	420	Q10567	95.70	42.30	WP_004245322.1	A0A384L4X0	481	90.90	90.90
KAF7281149.1	418	P61158	91.60	90.90	WP_004244046.1	A0A384LGT3	436	88.50	87.90

Table 4 shows the top protein interactions between *Klebsiella pneumoniae* and red palm weevil (*Rhynchophorus ferrugineus*) based on the PredHPI tool. The host similarity is between the host protein sequence and the homologous protein of the host protein. The pathogen similarity is between the pathogen protein sequence and the homologous protein of the pathogen protein.

Host protein	Host sequence length	Homolog of Host	Host similarity (%)		Pathogen protein	Pathogen sequence length	Homolog of pathogen	Pathogen similarity (%)	
			local	global				local	global
KAF7272596.1	202	P61586	94.40	81.70	YP_005224459.1	316	Q8ZAG5	94.90	94.90
KAF7269484.1	420	Q10567	95.70	42.30	YP_005229084.1	481	A0A384L4X0	94.00	94.00
KAF7281149.1	418	P61158	91.60	90.90	YP_005228709.1	438	A0A384LGT3	89.40	88.40
KAF7276687.1	131	P0ACD5	93.60	89.30	YP_005228233.1	404	P0A6B8	98.00	98.00

Table 5 shows the top protein interactions between *Serratia marcescens* and red palm weevil (*Rhynchophorus ferrugineus*) based on the PredHPI tool. The host similarity is between the host protein sequence and the homologous protein of the host protein. The pathogen similarity is between the pathogen protein sequence and the homologous protein of the pathogen protein.

Host protein	Host sequence length	Homolog of Host	Host similarity (%)		Pathogen protein	Pathogen sequence length	Homolog of pathogen	Pathogen similarity (%)	
			local	global				local	global
KAF7272595.1	217	P61586	94.40	80.20	WP_004934262.1	316	Q8ZAG5	97.20	97.20
KAF7272596.1	202	P61586	94.40	81.70	WP_004934262.1	316	Q8ZAG5	97.20	97.20
KAF7269484.1	420	Q10567	95.70	42.30	WP_025304368.1	456	Q8D1R0	90.60	89.40
KAF7269484.1	420	Q10567	95.70	42.30	WP_025304909.1	356	Q74R28	90.40	90.20
KAF7269484.1	420	Q10567	95.70	42.30	WP_025301655.1	746	A0A384KKB9	89.70	89.70
KAF7269484.1	420	Q10567	95.70	42.30	WP_025304416.1	481	A0A384L4X0	95.40	95.40
KAF7269484.1	420	Q10567	95.70	42.30	WP_025301643.1	358	Q8D184	93.50	92.20
KAF7281149.1	418	P61158	91.60	90.90	WP_025304153.1	438	A0A384LGT3	92.90	92.00
KAF7276687.1	131	P0ACD5	93.60	89.30	WP_016929764.1	404	P0A6B8	98.30	98.30

Homology modeling of all unknown proteins of host and pathogens:

All unknown host and pathogen proteins are modeled based on the homology modeling approach utilizing Phyre2 web-server (42). The homology model of the protein WP_004244046.1 was made using the template 1W7V chain D (51). The homology model of KAF7276687.1 was modeled using the templates: Solution NMR Structure of the Haemophilus Influenzae Iron-Sulfur Cluster Assembly Protein U with Zinc Bound at the Active Site. Northeast Structural Genomics Consortium Target IR24. (PDB ID: 1R9P) (52) and Solution structure of Iron-sulfur cluster protein U (PDB ID: 1WFZ). WP_004243849.1 protein has been modeled using PDB ID: 1P3W, X-ray crystal structure of *E. coli* IscS (resolution of 2.10 Å) and PDB ID: 3LVM, Crystal Structure of *E. coli* IscS (resolution of 2.05 Å). Transmembrane helices have been predicted in the sequence from 242 to 257 residues. KAF7281149.1 protein has been modeled utilizing PDB ID: 2P9L, Crystal Structure of bovine Arp2/3 complex with a resolution of 2.65 Å. KAF7279278.1 protein was modeled using PDB ID: 2NGR (i.e.) X-Ray Diffraction 1.9 Å. Transition state complex for GTP hydrolysis by CDC42: comparisons of the high-resolution structures for CDC42 bound to the active and catalytically compromised forms of the CDC42-GAP as a template. KAF7279371.1 and KAF7286850.1 were modeled using the same templates 1P3I-E, 1TZY-C, 1ID3-A, 1KX5-A, 1EQZ-G. Similarly, pathogen protein WP_000036873.1 was modeled using the templates 5D07-B, 6AN7-D, 5NJG-B, 5D07-A, 6JBH-C, 6AN7-C, and 5NJ3-B. The transmembrane helices predicted in this protein are: 40-55, 65-89, 107-137, 144-174, 185-203, 228-245, 255-278, 328-349, 368-389, 414-439, 443-467, 488-514, and 518-541.

Another pathogen protein WP_001062737.1 was homology modeled using the Crystal structure of a DUF3571 family protein (ABAYE3784) from *Acinetobacter baumannii* AYE at 1.95 Å resolution (PDB ID: 4L3U-A). The rest of the information can be obtained in **Table 6**.

Table 6. Homology models for all the host and the pathogens' selected proteins along with the template used for model preparation and the trans-membrane region predicted while modeling by the web-server.

Protein	Template	Transmembrane region predicted
WP_004244046.1	1W7V_D	
KAF7276687.1	1R9P, 1WFZ	-
WP_004243849.1	1P3W, 3LVM	242-257
KAF7281149.1	2P9L	-
KAF7279278.1	2NGR_A	-
KAF7279371.1	1P3I-E, 1TZY-C, 1ID3-A, 1KX5-A, 1EQZ-G	-
KAF7286850.1		-
WP_000036873.1	5D07-B, 6AN7-D, 5NJG-B, 5D07-A, 6JBH-C, 6AN7-C, 5NJ3-B	40-55, 65-89, 107-137, 144-174, 185-203, 228-245, 255-278, 328-349, 368-389, 414-439, 443-467, 488-514, 518-541
WP_001062737_1	4L3U-A	-
WP_001064089_1	3K1R-A, 1OBY-A, 2KOG-A, 1P1E-A, 2DLU-A, 2HKL-B, 1KWA-A, 2GZV-A, 1Z87-A	29-48
WP_001073054.1	2RCNA, 1U0LB, 4A2IV, 5UZ4Z, 2YV5A, 6H4DA	-
WP_025304153.1	1W7V-D	-
YP_005228233.1	1P3W-A, 3LVM-B	242-257
YP_005228709.1	1W7V-D	
WP_016929764.1	1P3W-A, 3LVM-B	241-256

Molecular docking

After obtaining all homology models of unknown proteins, molecular docking has been performed between the most common and most critical interacting proteins of the host and the predator (data as provided in **Table 7**) to understand the mechanism of interaction of two proteins. Docking score as the measure of docking interaction has been recorded. The best docking score has been observed in the KAF7276687.1 protein of the Host and WP_016929764.1 protein of the pathogen. The probable function of both the proteins also signifies the cause of the better binding. KAF7276687.1 has the probable function of Iron-sulfur cluster assembly scaffold protein IscU, and WP_016929764.1 has the Master enzyme that delivers sulfur to several partners involved in Fe-S cluster assembly as its potential role. A little similar score has been obtained for KAF7276687.1 and YP_005228233.1 interacting proteins which has same probable functions as just discussed. **Table 7** shows that Same functional probability resulted in the achievement of similar docking score.

Table 7. Consensus table of host protein interaction to predator protein and the docking score obtained after molecular docking using HawkDock web-server (47).

Host organism	Host Protein	length	Probable function	Pathogen	Pathogen protein	Length	Probable function	Docking Score (kcal/mol)
Red Palm weevil	KAF7276687.1	131	Iron-sulfur cluster assembly scaffold protein IscU	Proteus mirabilis	WP_004243849.1	404	Master enzyme that delivers sulfur to many partners involved in Fe-S cluster assembly	-43.34
	KAF7281149.1	418	ATP-binding component	Klebsiella pneumoniae	WP_004244046.1	436	Cofactor: Proline-specific aminopeptidase	-36.43
	KAF7281149.1	418			YP_005228709.1	438		-36.54
	KAF7276687.1	131	Iron-sulfur cluster assembly scaffold protein IscU	Serratia marcescens	YP_005228233.1	404	Master enzyme that delivers sulfur to several partners involved in Fe-S cluster assembly	-53.16
	KAF7276687.1	131			WP_016929764.1	404		-65.03
	KAF7281149.1	418	ATP-binding component		WP_025304153.1	438	Cofactor: Proline-specific aminopeptidase	-32.89

Molecular dynamics simulations

The best docking pose with favorable interactions was selected for molecular dynamics simulations to study long-range/short-range interactions and stability of the complex over time. MD simulations each of 200 ns were performed for all protein-protein complexes using GROMACS software (50,53–57). The MD simulation trajectory of all three protein-protein complexes was analyzed, and interaction images are presented in Figure 1.

The RMSD, RMSF, Rg, SASA, energy, and hydrogen bonds plots were generated. The RMSD values of all complexes fall within a range of 3.2 to 3.8 Å against the function of time, thus, displayed high stability of protein-protein complexes during the simulation over 200 ns (Figure 2–left panel). The Radius of gyration (Rg) variations of the *Klebsiella pneumoniae* and *Serratia marcescens* complexes during the simulation period showed a similar fashion (Figure 2–right panel). A spike of >3 nm during ~40 ns time frame (in *Serratia marcescens*) and around ~90 ns time frame (in *Klebsiella pneumoniae*) point out a biological phenomenon that happened after the interaction between the two proteins of respective complexes. No such change has been observed in *Proteus mirabilis*. Such low variation in Rg values with respect to time revealed consistent compactness of the protein during simulation even when present in the complexed form with another protein, hence proved high stability of the complex. This phenomenon also explains why a lower time scale of around 50 ns is not enough for large protein complexes, which led us to run these simulations for 200 ns each.

The Root Mean Square fluctuations (RMSF) calculated for the protein backbone of the complexes shows similar results for *K. pneumoniae* and *P. mirabilis*. This provides evidence that the integrity of the protein backbone was maintained. The least violent displacements in the backbone were encountered by the protein when complexed with another protein. On the other hand, the RMSF for the *S. marcescens* complex shows a higher degree of fluctuations throughout the simulation run. Along with this, the high amplitude of fluctuations at the terminal residues was due to the high degree of freedom of these residues compared to that of the intermediate region of the protein chain. Further, the fluctuations in the residues of loop regions were higher compared to other forms of secondary structures (Figure 3-left panel).

To determine the complexes' structural stability, we also monitored the solvent-accessible surface area (SASA), as shown in Figure 3 (right panel). We observed that the structure of all the three complexes almost had a similar average SASA value of 240 nm².

These results suggested that the complex is very stable; thus, the conformational dynamics analyses RMSD, Rg, and SASA results were consistent, indicating the stable interaction of protein complex over time.

The number of hydrogen bonds formed between two proteins of the protein-protein complex during the simulation was plotted, proving effective interaction of the two proteins to each other during the 200 ns time of simulation (Figure 4-left panel). The higher numbers of hydrogen bonds (around 400) display significant interactions between the *K. pneumoniae* complexes during the simulation. All these results proved the highly stable binding of *K. pneumoniae* complex proteins, thus providing great scope for its potential to modulate the structural conformation of the complex to a greater extent. *P. mirabilis* and *S. marcescens* complexes also show equivalently strong and stable conformations.

The number of H-bonds between solute and solvent in the case of *K. pneumoniae* complex is higher (around 1160) than the other two complexes (around 1130 in *P. mirabilis* and *S. marcescens*). This evidence shows a greater chance for *K. pneumoniae* complex to interact freely with the environment and to adapt to any conformational change during the simulation (Figure 4-right panel).

To further explore the dynamic properties of the investigated structures in our simulations, the essential dynamics (ED) analysis on the backbone atoms was performed (58). ED reflects the overall conformational space of the protein-protein complex during simulations (Figure 5-left panel). The projection of trajectories onto the first two principal components (PC1, PC2) shows the motion of the investigated system in phase space and is illustrated in Figure 5 (left panel). In the ED analysis, all the three complexes show similar phase space behavior and explored the same conformational space. Minor irregular stretches or clusters of line in the ED plot of the complexes may be due to increased flexibility in the initial part of the simulation suggesting the more conformational dynamics of both partner proteins upon complex formation, and stabilizing over time, which remains stable until the end as seen in the RMSD plots.

We also calculated the H-bonds between Host protein and bacterial protein over 200 ns. As shown in figure 5 (right panel), the results of these analyses indicate the increase in h-bonds after 90 ns. The results implied that hydrogen bonds stabilize the interaction between both proteins over time and increase the complex's stability.

Conclusion

Critical challenges at present related to food safety has been a matter on concern on growth and development in agricultural research and the advent of modern biotechnological and agricultural tools. In this paper, host-pathogen interaction between *Rhynchophorus ferrugineus* (an invasive wood-boring pest that destroys palm) also known as red palm weevil and pathogen's *Proteus mirabilis*, *Serratia marcescens*, and *Klebsiella pneumoniae* have been comprehensively studied using structural and functional based research. The systems biology approach has also been applied, and it was found that the Iron sulfur cluster assembly protein (KAF7276687.1) in the host is targeted commonly by all three pathogens (*Klebsiella pneumoniae*, *Proteus mirabilis*, and *Serratia marcescens*) through their sulfur delivery master enzyme's. Following molecular dynamics simulations results show stable protein complexes over the course of 200 nanoseconds. This study provides a suitable starting point to carry forward the experimental work to make bio-pesticides or the control of red palm weevil, the most dangerous pest for the date's cultivation.

Declarations

Funding: This research project has been funded by the Deanship of Scientific Research at the University of Ha'il, Ha'il, Kingdom of Saudi Arabia (fund number: RG-191352).

Conflicts of interest/Competing interests: The authors declare no conflict of interest.

Ethics approval: N/A

Consent to participate: Yes

Consent for publication: Yes

Availability of data and material: N/A

Code availability: N/A

Authors' contributions: MA initiated and designed the project, MA, AN implemented the experiments, MA, AN, AS and NA wrote the manuscript.

References

1. Jones DL (1995) Palms - Google Scholar.
2. Purseglove JW (1972) Tropical crops: monocotyledons. - Google Scholar.
3. Peters HA, Pauw A, Silman MR, Terborgh JW. Falling palm fronds structure amazonian rainforest sapling communities. *Proc Biol Sci*. 2004 Aug 7;271 Suppl 5:S367–9.
4. Baker WJ, Norup MV, Clarkson JJ, Couvreur TLP, Dowe JL, Lewis CE, et al. Phylogenetic relationships among arecoid palms (Arecaceae: Arecoideae). *Ann Bot*. 2011 Dec;108(8):1417–1432.
5. Rugman-Jones PF, Hoddle CD, Hoddle MS, Stouthamer R. The lesser of two weevils: molecular-genetics of pest palm weevil populations confirm *Rhynchophorus vulneratus* (Panzer 1798) as a valid species distinct from *R. ferrugineus* (Olivier 1790), and reveal the global extent of both. *PLoS One*. 2013 Oct 15;8(10):e78379.
6. McKenna DD, Wild AL, Kanda K, Bellamy CL, Beutel RG, Caterino MS, et al. The beetle tree of life reveals that Coleoptera survived end-Permian mass extinction to diversify during the Cretaceous terrestrial revolution. *Syst Entomol*. 2015 Oct;40(4):835–880.
7. McKenna DD, Sequeira AS, Marvaldi AE, Farrell BD. Temporal lags and overlap in the diversification of weevils and flowering plants. *Proc Natl Acad Sci USA*. 2009 Apr 28;106(17):7083–7088.
8. Abdel Farag El-Shafie H, Romeno Faleiro J. Red Palm Weevil *Rhynchophorus ferrugineus* (Coleoptera: Curculionidae): Global Invasion, Current Management Options, Challenges and Future Prospects. In: El-Shafie H, editor. *Invasive Species - Introduction Pathways, Economic Impact, and Possible Management Options*. IntechOpen; 2020.
9. Scientific Consultation and High-Level Meeting on Red Palm Weevil Management.
10. Host Palms [Internet]. [cited 2021 Jun 1]. Available from: <https://www.savealgarvepalms.com/weevil-facts/host-palm-trees?lang=en>
11. Rochat D, Malosse C, Lettère M. Identification of new pheromone-related compounds from volatiles produced by males of four Rhynchophorinae weevils (Coleoptera, Curculionidae). *Comptes rendus de* 1993;
12. Dyer MD, Neff C, Dufford M, Rivera CG, Shattuck D, Bassaganya-Riera J, et al. The human-bacterial pathogen protein interaction networks of *Bacillus anthracis*, *Francisella tularensis*, and *Yersinia pestis*. *PLoS One*. 2010 Aug 9;5(8):e12089.
13. König R, Zhou Y, Elleder D, Diamond TL, Bonamy GMC, Irelan JT, et al. Global analysis of host-pathogen interactions that regulate early-stage HIV-1 replication. *Cell*. 2008 Oct 3;135(1):49–60.
14. Rizzetto L, Cavalieri D. Friend or foe: using systems biology to elucidate interactions between fungi and their hosts. *Trends Microbiol*. 2011 Oct;19(10):509–515.
15. Bonetta L. Protein-protein interactions: Interactome under construction. *Nature*. 2010 Dec 9;468(7325):851–854.
16. Liu X, Liu B, Huang Z, Shi T, Chen Y, Zhang J. SPPS: a sequence-based method for predicting probability of protein-protein interaction partners. *PLoS One*. 2012 Jan 26;7(1):e30938.
17. Durmuş S, Çakır T, Özgür A, Guthke R. A review on computational systems biology of pathogen-host interactions. *Front Microbiol*. 2015 Apr 9;6:235.
18. Zuck M, Austin LS, Danziger SA, Aitchison JD, Kaushansky A. The promise of systems biology approaches for revealing host pathogen interactions in malaria. *Front Microbiol*. 2017 Nov 16;8:2183.
19. Segata N, Boernigen D, Tickle TL, Morgan XC, Garrett WS, Huttenhower C. Computational meta'omics for microbial community studies. *Mol Syst Biol*. 2013 May 14;9:666.
20. Sturdevant DE, Virtaneva K, Martens C, Bozinov D, Ogundare O, Castro N, et al. Host-microbe interaction systems biology: lifecycle transcriptomics and comparative genomics. *Future Microbiol*. 2010 Feb;5(2):205–219.

21. Liu Z-P, Chen L. Proteome-wide prediction of protein-protein interactions from high-throughput data. *Protein Cell*. 2012 Jul;3(7):508–520.
22. Dix A, Vlais S, Guthke R, Linde J. Use of systems biology to decipher host-pathogen interaction networks and predict biomarkers. *Clin Microbiol Infect*. 2016 Jul;22(7):600–606.
23. Yeung A, Hale C, Clare S, Palmer S, Bartholdson Scott J, Baker S, et al. Using a Systems Biology Approach To Study Host-Pathogen Interactions. *Microbiol Spectr*. 2019 Mar;7(2).
24. Shoemaker BA, Panchenko AR. Deciphering protein-protein interactions. Part II. Computational methods to predict protein and domain interaction partners. *PLoS Comput Biol*. 2007 Apr 27;3(4):e43.
25. Lewis ACF, Saeed R, Deane CM. Predicting protein-protein interactions in the context of protein evolution. *Mol Biosyst*. 2010 Jan;6(1):55–64.
26. Westermann AJ, Gorski SA, Vogel J. Dual RNA-seq of pathogen and host. *Nat Rev Microbiol*. 2012 Sep;10(9):618–630.
27. Galperin MY, Koonin EV. Who's your neighbor? New computational approaches for functional genomics. *Nat Biotechnol*. 2000 Jun;18(6):609–613.
28. Overbeek R, Fonstein M, D'Souza M, Pusch GD, Maltsev N. The use of gene clusters to infer functional coupling. *Proc Natl Acad Sci USA*. 1999 Mar 16;96(6):2896–2901.
29. Liu Z-P, Wang J, Qiu Y-Q, Leung RKK, Zhang X-S, Tsui SKW, et al. Inferring a protein interaction map of *Mycobacterium tuberculosis* based on sequences and interologs. *BMC Bioinformatics*. 2012 May 8;13 Suppl 7:S6.
30. Aloy P, Ceulemans H, Stark A, Russell RB. The relationship between sequence and interaction divergence in proteins. *J Mol Biol*. 2003 Oct 3;332(5):989–998.
31. Pellegrini M, Marcotte EM, Thompson MJ, Eisenberg D, Yeates TO. Assigning protein functions by comparative genome analysis: protein phylogenetic profiles. *Proc Natl Acad Sci USA*. 1999 Apr 13;96(8):4285–4288.
32. Barker D, Pagel M. Predicting functional gene links from phylogenetic-statistical analyses of whole genomes. *PLoS Comput Biol*. 2005 Jun 24;1(1):e3.
33. Wang F, Liu M, Song B, Li D, Pei H, Guo Y, et al. Prediction and characterization of protein-protein interaction networks in swine. *Proteome Sci*. 2012 Jan 10;10(1):2.
34. Shin CJ, Davis MJ, Ragan MA. Towards the mammalian interactome: Inference of a core mammalian interaction set in mouse. *Proteomics*. 2009 Dec 1;9(23):5256–5266.
35. Schleker S, Garcia-Garcia J, Klein-Seetharaman J, Oliva B. Prediction and comparison of *Salmonella*-human and *Salmonella*-*Arabidopsis* interactomes. *Chem Biodivers*. 2012 May;9(5):991–1018.
36. Krishnadev O, Srinivasan N. Prediction of protein-protein interactions between human host and a pathogen and its application to three pathogenic bacteria. *Int J Biol Macromol*. 2011 May 1;48(4):613–619.
37. Pu Y-C, Ma T-L, Hou Y-M, Sun M. An entomopathogenic bacterium strain, *Bacillus thuringiensis*, as a biological control agent against the red palm weevil, *Rhynchophorus ferrugineus* (Coleoptera: Curculionidae). *Pest Manag Sci*. 2017 Jul;73(7):1494–1502.
38. Hussain A, Rizwan-ul-Haq M, Al-Ayedh H, Ahmed S, Al-Jabr AM. Effect of *Beauveria bassiana* infection on the feeding performance and antioxidant defence of red palm weevil, *Rhynchophorus ferrugineus*. *Biocontrol*. 2015 Dec;60(6):849–859.
39. Jalinias J, Güerri-Agulló B, Mankin RW, López-Follana R, Lopez-Llorca LV. Acoustic Assessment of *Beauveria bassiana* (Hypocreales: Clavicipitaceae) Effects on *Rhynchophorus ferrugineus* (Coleoptera: Dryophthoridae) Larval Activity and Mortality. *J Econ Entomol*. 2015 Apr;108(2):444–453.

40. Search: GCA_014462685.1 - NLM [Internet]. [cited 2021 May 30]. Available from: https://www.ncbi.nlm.nih.gov/search/all/?term=GCA_014462685.1
41. Loaiza CD, Kaundal R. PredHPI: an integrated web server platform for the detection and visualization of host-pathogen interactions using sequence-based methods. *Bioinformatics*. 2021 May 5;37(5):622–624.
42. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc*. 2015 Jun;10(6):845–858.
43. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr*. 1993 Apr 1;26(2):283–291.
44. Lüthy R, Bowie JU, Eisenberg D. Assessment of protein models with three-dimensional profiles. *Nature*. 1992 Mar 5;356(6364):83–85.
45. Bowie JU, Lüthy R, Eisenberg D. A method to identify protein sequences that fold into a known three-dimensional structure. *Science*. 1991 Jul 12;253(5016):164–170.
46. Colovos C, Yeates TO. Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci*. 1993 Sep;2(9):1511–1519.
47. Weng G, Wang E, Wang Z, Liu H, Zhu F, Li D, et al. HawkDock: a web server to predict and analyze the protein-protein complex based on computational docking and MM/GBSA. *Nucleic Acids Res*. 2019 Jul 2;47(W1):W322–W330.
48. Feng T, Chen F, Kang Y, Sun H, Liu H, Li D, et al. HawkRank: a new scoring function for protein-protein docking based on weighted energy terms. *J Cheminform*. 2017 Dec 28;9(1):66.
49. Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J Chem Theory Comput*. 2015 Aug 11;11(8):3696–3713.
50. Kutzner C, Páll S, Fechner M, Esztermann A, de Groot BL, Grubmüller H. More bang for your buck: Improved use of GPU nodes for GROMACS 2018. *J Comput Chem*. 2019 Oct 15;40(27):2418–2431.
51. Graham SC, Bond CS, Freeman HC, Guss JM. Structural and functional implications of metal ion selection in aminopeptidase P, a metalloprotease with a dinuclear metal center. *Biochemistry*. 2005 Oct 25;44(42):13820–13836.
52. Ramelot TA, Cort JR, Goldsmith-Fischman S, Kornhaber GJ, Xiao R, Shastry R, et al. Solution NMR structure of the iron-sulfur cluster assembly protein U (IscU) with zinc bound at the active site. *J Mol Biol*. 2004 Nov 19;344(2):567–583.
53. Pronk S, Páll S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, et al. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*. 2013 Apr 1;29(7):845–854.
54. Páll S, Abraham MJ, Kutzner C, Hess B, Lindahl E. Tackling Exascale Software Challenges in Molecular Dynamics Simulations with GROMACS. In: Markidis S, Laure E, editors. *Solving Software Challenges for Exascale*. Cham: Springer International Publishing; 2015. p. 3–27.
55. Huang Y, Chen W, Wallace JA, Shen J. All-Atom Continuous Constant pH Molecular Dynamics With Particle Mesh Ewald and Titratable Water. *J Chem Theory Comput*. 2016 Nov 8;12(11):5411–5421.
56. Liu X, Shi D, Zhou S, Liu H, Liu H, Yao X. Molecular dynamics simulations and novel drug discovery. *Expert Opin Drug Discov*. 2018;13(1):23–37.
57. Kutzner C, Páll S, Fechner M, Esztermann A, de Groot BL, Grubmüller H. Best bang for your buck: GPU nodes for GROMACS biomolecular simulations. *J Comput Chem*. 2015 Oct 5;36(26):1990–2008.
58. Prakash A, Dixit G, Meena NK, Singh R, Vishwakarma P, Mishra S, et al. Elucidation of stable intermediates in urea-induced unfolding pathway of human carbonic anhydrase IX. *J Biomol Struct Dyn*. 2018 Jul;36(9):2391–2406.

Figures

(A) Interaction of *K. pneumoniae* with predator protein

(B) Interaction of *P. mirabilis* with predator protein

(C) Interaction of *S. marcescens* with predator protein

Figure 1

shows the protein-protein interactions of (A) *Klebsiella pneumoniae*, (B) *Proteus mirabilis*, and (C) *Serratia marcescens* with the predator protein, respectively.

Figure 2

Root mean square deviation (RMSD) of C-alpha atoms of protein (left panel) and Radius of Gyration (R_g) (right panel) with respect to time throughout 200ns of MD simulation run.

Figure 3

Root Mean Square fluctuations (RMS fluctuation) (left panel) and solvent accessible surface area (right panel) of the complexes

Figure 4

An average number of Hydrogen bonds formed within the solute (protein-protein complex) and between solute and solvent during the entire 200 ns of simulations.

Figure 5

Essential dynamics provided the PCA graph (left) for all three host-pathogen systems. Projections for *S. marcescens* (in green) and *P. mirabilis* (in red) are exactly overlapping, and therefore red is not visible. The right panel shows the H-bonds between Host protein and bacterial protein.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [flowchart.jpg](#)