TCABSE-J Research Article
ISSN: 2583- 2557, *Volume 1, Issue 6, pp9-14. 2023.*

TCABSE-J
Interdisciplinary Sciences Journal

Molecular dynamics simulations of cytotoxin-associated gene A coded protein from *Helicobacter pylori* to probe the flexibility of p53 binding pocket for inhibitor design.

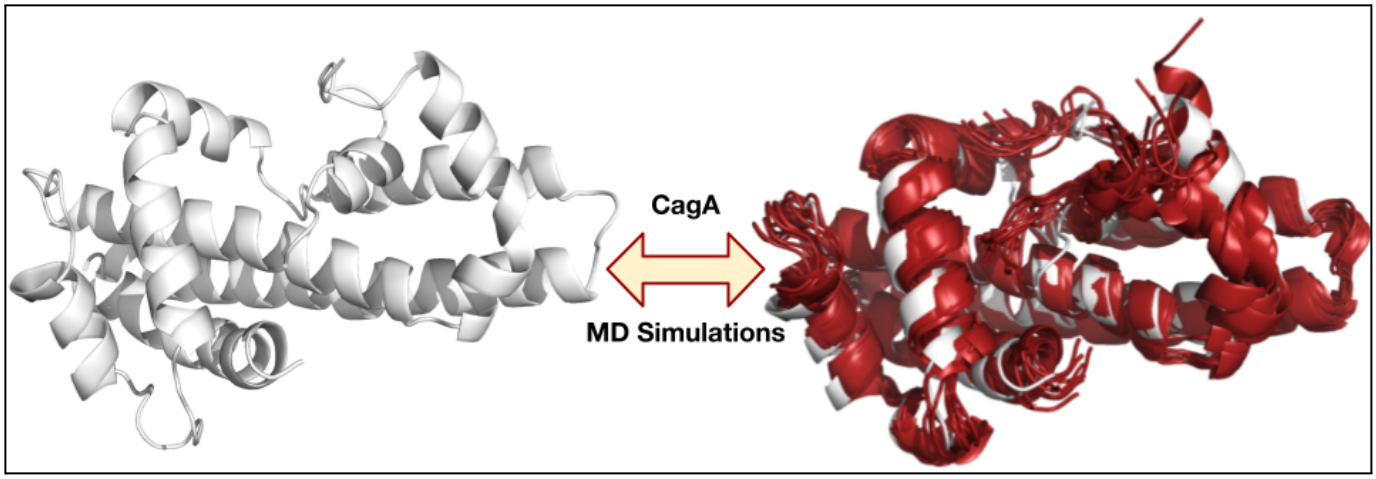
Madhumita Aggunna[1], Abhinav V. K. S. Grandhi[1, 2] and Ravikiran S. Yedidi[1, 3, *]

[1]Department of Intramural research core, The Center for Advanced-Applied Biological Sciences & Entrepreneurship (TCABS-E), Visakhapatnam 530016, A.P. India; [2]Koringa College of Pharmacy, Korangi 533461, A. P. India; [3]Department of Zoology, Andhra University, Visakhapatnam 530003, A.P. India.

(*Correspondence to R.S.Y.: tcabse.india@gmail.com).

Keywords: CagA, *H. pylori*, molecular dynamics simulations, inhibitor design.

Antibiotics and proton pump inhibitors are the only options for the treatment of *Helicobacter pylori* infections till date. However, antibiotic-resistance is already reported clinically against some of these regularly used antibiotics. In this context, novel antibiotics that specifically target the *H. pylori* are needed. Clinically, *H. pylori* infections are broadly classified into several types among which, the cytotoxin-associated gene A (CagA)-positive cases are identified as the ones where the bacterium contains CagA virulence protein that helps in hijacking the intracellular signaling cascades of the host gastric epithelial cells. In this study we explored the p53 binding pocket of CagA protein, taken from PDB ID: 4IRV, *in silico*. Our results indicate that the p53 binding pocket of CagA is flexible to potentially fit a diverse set of small molecules that can be used as inhibitors. Over a 10 ns molecular dynamics simulation we noticed multiple conformational changes leading to changes in the architecture of the p53 binding pocket but the pocket was stable without collapsing throughout the MD simulations indicating that this pocket can be targeted for virtual screening of small molecules. Based on our MD simulations, we conclude that the p53 binding pocket of CagA protein can be used for screening potential inhibitors of *H. pylori* infections.



Citation: Aggunna, M., Grandhi, A.V.K.S. and Yedidi, R.S. (2023). Molecular dynamics simulations of cytotoxin-associated gene A coded protein from *Helicobacter pylori* to probe the flexibility of p53 binding pocket for inhibitor design. *TCABSE-J*, Vol. 1, Issue 6:9-14. Epub: Aug 10th, 2023.

TCABSE-J Research Article
ISSN: 2583- 2557, *Volume 1, Issue 6, pp9-14. 2023.*

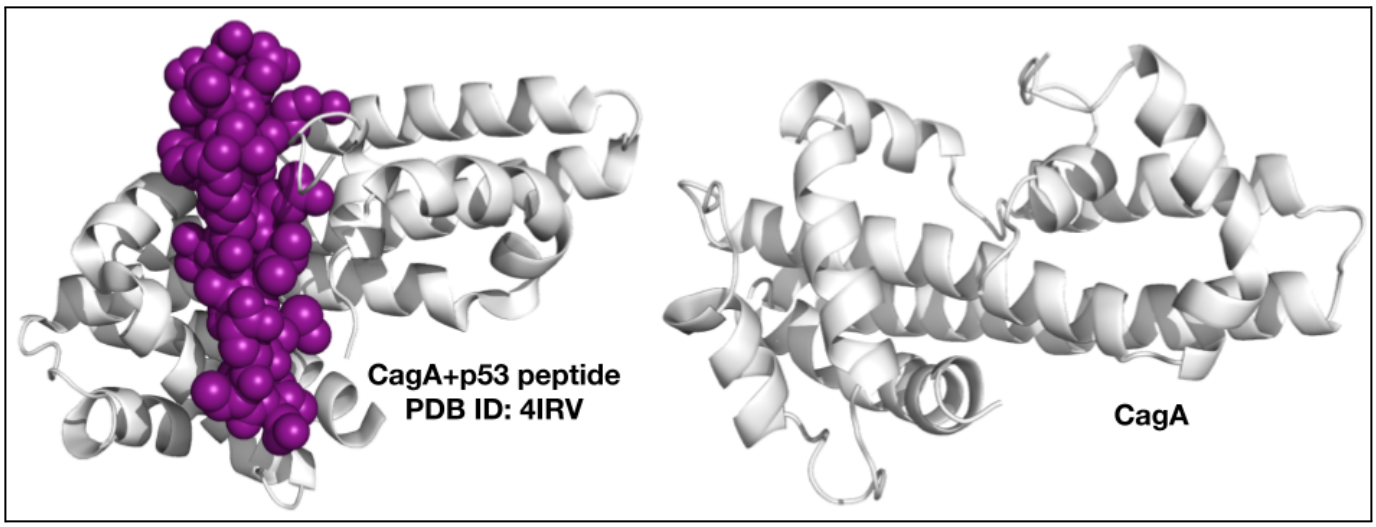


Figure 1. Structures of CagA +/- p53 peptide.

The Cytotoxin Associated gene-A (Cag-A) protein is a virulence factor of the *Helicobacter pylori (H. pylori)* which gets transported to the human gastric cells through the type-4 secretion system (T4SS) by developing a pilus that gets injected to host gastric epithelial cells [1,2,3]. The Cag-A protein coded from the Cag Pathogenic islands, a group of several types of Cag proteins [4,5,6], travels through the pilus. Once the Cag-A protein enters into the human body it participates in the interactions with other host proteins and infects the humans initially as gastritis, then gastric ulcers. The ulcers in the long term cause gastric cancer. The Cag-A protein is an oncogenic protein with a molecular weight between 120 kDa & 140 kDa [6,7,8]. This protein enters the host cell and participates in the downstream signaling by undergoing the kinase phosphorylation dependent and independent. The N-terminal region is kinase phosphorylation independent and inhibits tumor suppressors [6]. The N-terminal region of the Cag-A protein interacts with the Apoptosis stimulating protein of p53-2 protein (ASPP2) and inactivates and degrades the p53 protein [9,10]. The structure of the Cag-A protein is taken further for the molecular dynamic simulations to observe the conformational changes in the binding pocket region [6,11]. The binding pocket of the Cag-A at N-terminus is analyzed to understand the flexibility of the protein which makes the drug design easy. The Cag-A protein is observed to be a protein that is hard to be considered as a drug target and the binding pocket in the N-terminal region that is focused in this paper is also observed to be a long binding pocket which may lead to the complexity for the ligands to bind with low binding affinities. Thus the conformational flexibility of CagA should be explored further.

In this study, the Cag-A structure is taken from the PDB ID: 4IRV (Figure 1) and also the homology model of the same FASTA sequence of the Cag-A [11]. The Cag-A structure and the models were both used for the Molecular Dynamics (MD) simulations using the Desmond Molecular Dynamics System, version 2018-4, D. E. Shaw Research, New York, NY [Schrödinger, LLC, New York, NY].

Materials & Methods:

Preparation of CagA structure: The 3D structure of CagA protein was prepared from the X-ray crystal structure of CagA protein bound to p53 peptide, PDB ID: 4IRV [11]. There were four chains of CagA each bound to one p53 peptide in the structure. Chain A along with the p53 peptide that was bound in the pocket of chain A were kept and the coordinates for the remaining chains were deleted. The final prepared structure file contains one chain of CagA bound to the one chain of p53 peptide. This final file was used further for processing and building the system in order to perform the MD simulations.

CagA homology model building: The FASTA sequence of the Cag-A is taken from the PDB ID: 4IRV and is used to build the 3D homology model of the same using SWISS-MODEL tool (https://swissmodel.expasy.org/) choosing 4IRV as the template sequence. The homology model was built due to main chain breaks in the crystal structure.

Protein pre-processing: CagA protein pre-processing was performed using the Maestro molecular graphics window from Schrödinger LLC, NY. Both CagA from the crystal structure and the CagA homology model were imported into Maestro. Protein preparation wizard was used to pre-process the CagA which includes multiple steps of optimization.

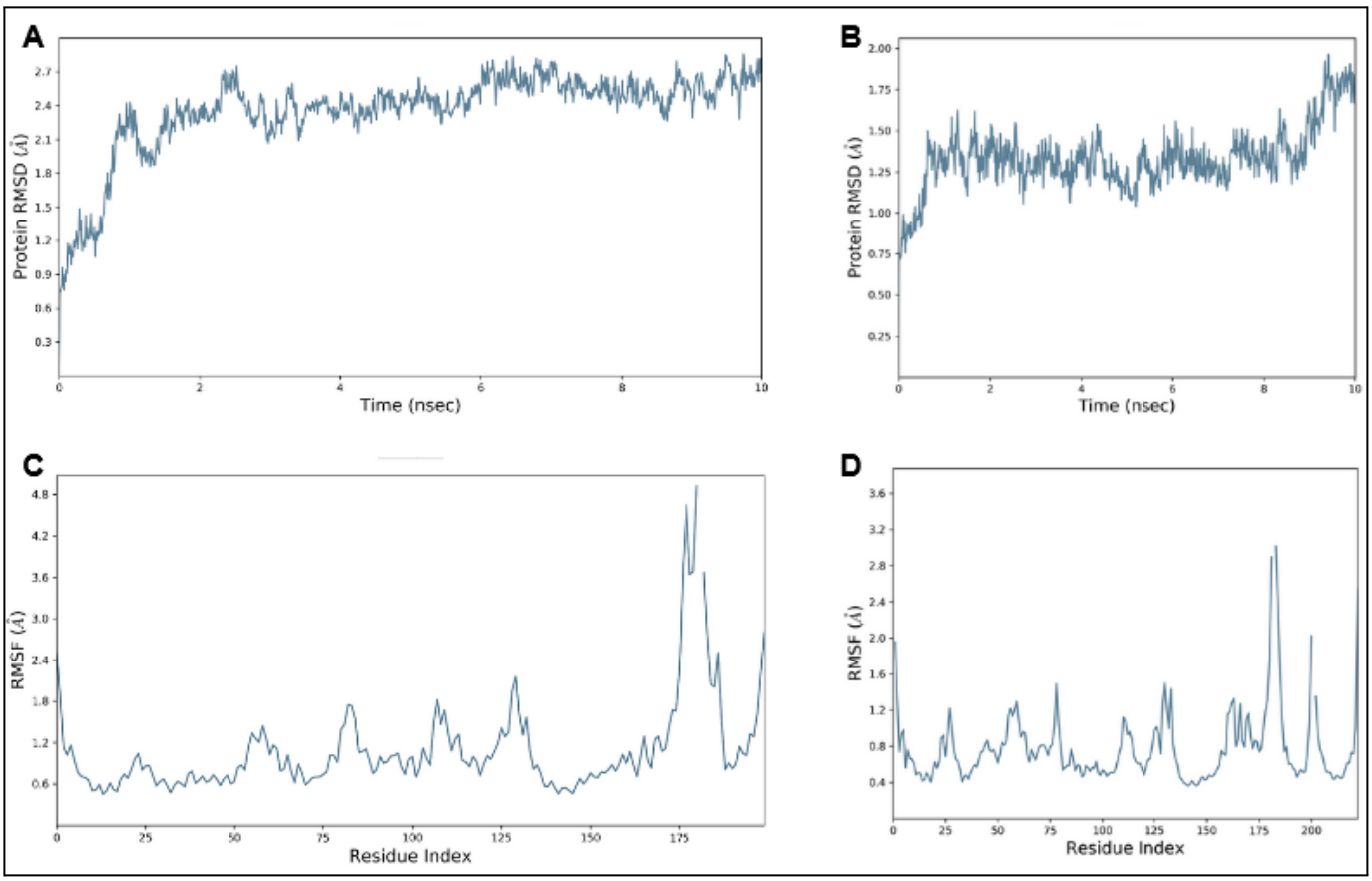


Figure 2. RMSD & RMSF analysis of CagA+/-p53 peptide.

Assignment of bond orders, capping the protein termini, addition of hydrogens, conversion of selenomethionine to methionine and generation of hydrogen bonds. The hydrogen atoms were then optimized by considering the water molecules in the structure. Any close contacts were rectified by relaxing the structure. The processed protein was then used to build the system for MD simulations.

System builder: The processed protein was used to set up the system with periodic boundaries containing the implicit solvent continuum using the SPC water model. The overall volume of the system was minimized by choosing the orthorhombic model. Ions were added and neutralized so that there is no net charge on the system. To mimic the physiological states, 0.15 M salt was added into the system. System builder option of Desmond (D.E. Shaw Research, NY) was used to build the systems for both CagA structure and its homology model.

MD simulations: The MD simulations were performed using Desmond on a 6th generation i5-quad core processor. The system built as described above is loaded into the Maestro molecular graphics window to display all the atoms. Periodic recording of the trajectory at fixed intervals along with the energy was set depending on the total length of the MD simulation. For example, a 10 picoseconds (ps) interval was set for recording the trajectory in order to perform a 10 nanoseconds (ns) MD simulation with an approximate number of 1000 frames per interval. OPLS2005 force field was used to perform the current molecular mechanics-based simulations.

Trajectory analysis: The post MD simulation trajectory was loaded into the Maestro molecular graphics window and was analyzed frame by frame to evaluate any visible structural changes overall and with reference to the p53 binding pocket. The overall protein C_α root mean square deviations (RMSD) and per residue RMS fluctuations (RMSF) were visualized by using the simulation interactions diagram option in Desmond. Both the RMSD and RMSF plots were saved for each MD simulation. Coordinate files at intervals of 1 ns were saved as ".pdb" files to visualize the 3D conformational changes in the protein using Maestro. Figures and movies of the MD simulations were prepared while the trajectory is still loaded into the Maestro molecular graphics window.

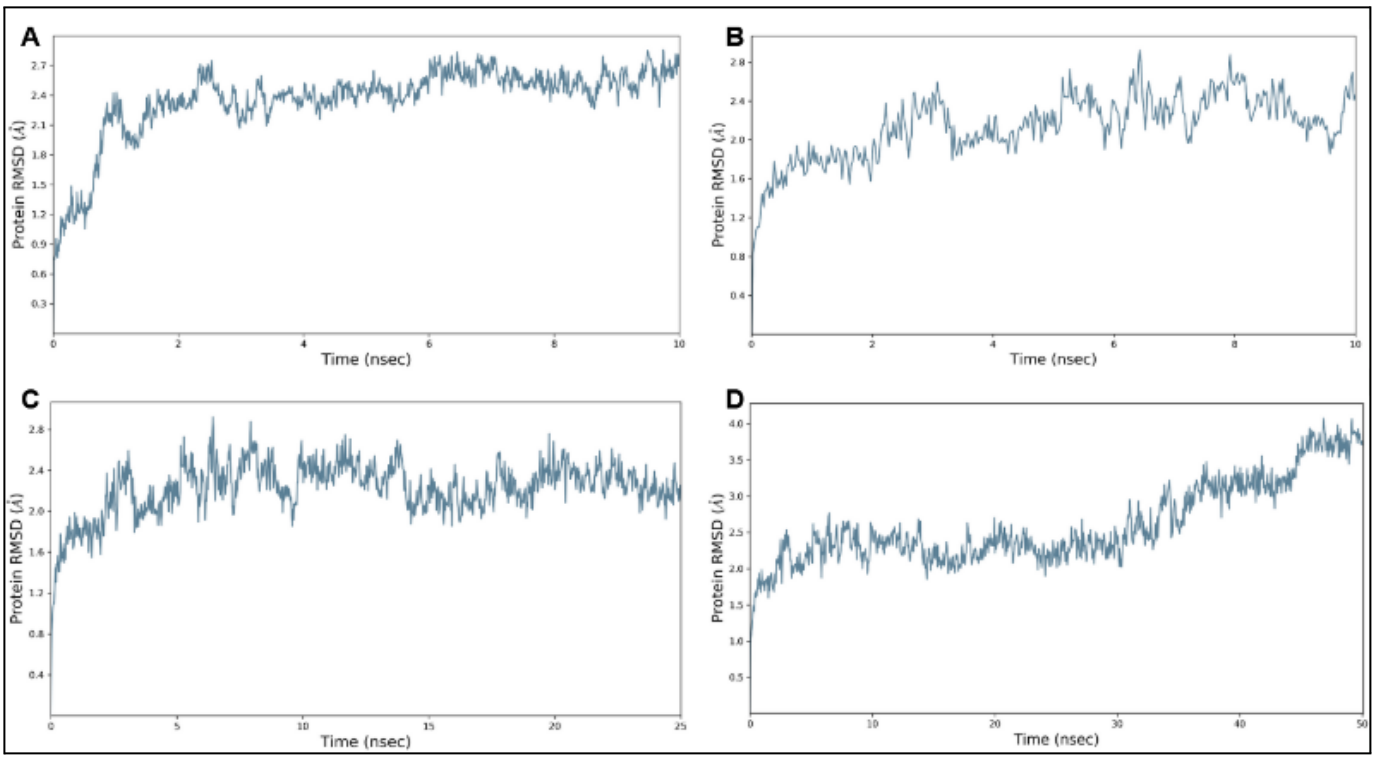


Figure 3. RMSD of CagA structure vs. homology model.

Results and Discussion:

Dimensions of the p53 binding pocket in CagA: The p53 binding pocket in CagA (Figure 1) was roughly measured to be 10 Å in width, 23 Å in length and 10 Å in depth thus adding up a total volume of 2,300 $Å^3$. The p53 peptide in this structure is 20 amino acids long which fits into the above mentioned volume (Figure 1). Such a large binding pocket might be stable with a bigger ligand such as the p53 peptide bound but without a ligand, the stability of such a binding pocket is questionable. In order to delineate this dilemma, we performed 10 ns MD simulations of CagA with and without p53 bound in the pocket.

The protein C_α RMSD of CagA with p53 is slightly lower than without p53: The protein C_α RMSD of CagA, with and without p53 bound, started at 0.75 Å and reached up to 1.6 Å and 2.4 Å at 1 ns, respectively (Figure 2). While the RMSD for CagA with p53 was stable continuously until 8 ns and then increased to 1.75 Å at 10 ns, the RMSD for CagA without p53 increased up to 2.7 Å at 2.5 ns and continued stably until 10 ns. This suggests that the apo protein (CagA without p53) had to go through more RMSD initially compared to the complexed protein. Thus the apo protein exhibited almost 1 Å higher RMSD compared to the p53 bound CagA (Figure 2).

The RMSF profile of apo CagA is different compared to the CagA bound to p53: The overall RMSF for the apo CagA and p53 bound CagA are <2.4 Å and <1.6 Å, respectively, suggesting that the p53 bound CagA has relatively lower RMSF compared to the apo CagA. RMSF values of residue beyond 175 were ignored because of a break in the main chain of the protein in the crystal structure. In order to circumvent this issue, we built a homology model of CagA and used that model for MD simulations.

MD simulations of CagA homology model: The C_α RMSD of CagA homology model is slightly different from the apo CagA that was taken from the crystal structure. As shown in Figure 3, the initial spike at 1 ns that was seen for the apo CagA from the crystal structure was no longer seen in the homology model at 1 ns. Taking this into consideration, we further investigated whether the homology model of CagA will show any major changes in the protein C_α RMSD. So, these MD simulations were further extended to 25 ns and 50 ns (Figure 3). No major spikes were seen within the 25 ns MD simulation. However, the C_α RMSD progressively increased to almost 4 Å from 30 ns to 50 ns in the 50 ns MD simulation.

Higher RMSF is seen in the 50 ns MD simulation compared to the 25 ns MD simulation: The homology model of CagA showed significant changes in the RMSF of its residues when simulated up to 50 ns instead of 25 ns. As shown in Figure 4,

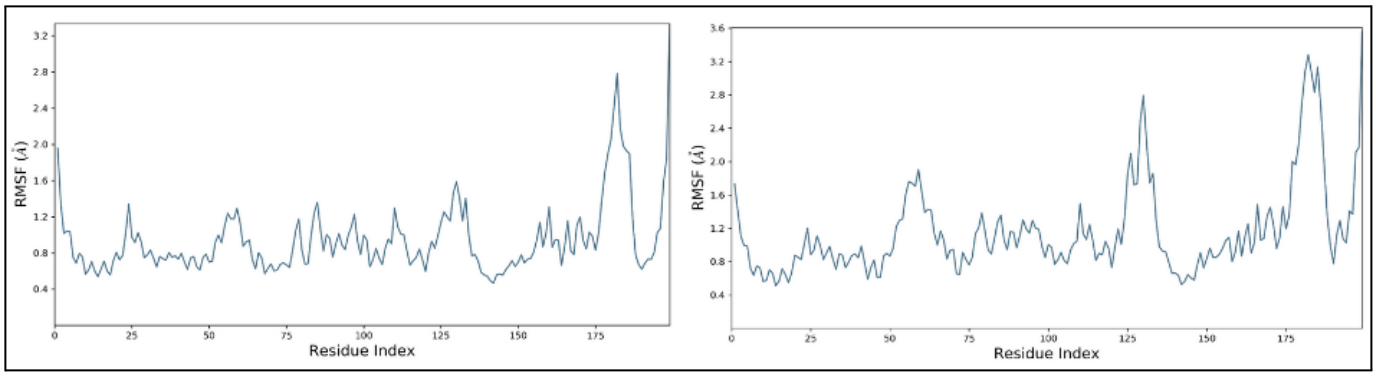


Figure 4. RMSF analysis of CagA homology model.

majorly, three regions were identified where there were significant changes in the RMSF values. These regions are 50-75 (1.2 Å vs. 2.0 Å), 125-150 (1.6 Å vs. 2.8 Å) and 175-185 (2.8 Å vs. 3.2 Å). Comparing Figures 3 and 4, it is safe to say that the residues 125-150 are probably contributing to the unusual behavior of the protein between 30 ns and 50 ns.

Conclusion and Future directions:

The current study clearly demonstrates that the apo CagA possesses more conformational flexibility compared to the p53 bound CagA. Further, the p53 binding pocket showed flexibility in the absence of the ligand suggesting that the volume of this pocket is large and thus, it poses a challenge to screen small molecule leads. We are currently in the process of using the p53 binding pocket of CagA for virtual screening of small molecules with a goal to identify any leads that can be used as potential drugs in the future for the treatment of *H. pylori* infections.

Acknowledgements: We thank The Yedidi Institute of Discovery and Education, Toronto for scientific collaborations.

Conflict of interest: The authors declare no conflict of interest in this study. However, this research article is an ongoing project currently at TCABS-E, Visakhapatnam, India.

Author contributions: M.A. and A.G. performed all the molecular dynamics simulations. R.S.Y. is the principal investigator who designed the project, trained M.A. and A.G., secured required material for the project, provided the laboratory space and facilities needed. M.A., A.G. assisted and R.S.Y. wrote and edited the final version of the manuscript.

References

1. Warren, J. R., & Marshall, B. (1983). Unidentified curved bacilli on gastric epithelium in active chronic gastritis. *Lancet (London, England)*, *1*(8336), 1273–1275.
2. Parsonnet, J., Friedman, G. D., Vandersteen, D. P., Chang, Y., Vogelman, J. H., Orentreich, N., & Sibley, R. K. (1991). Helicobacter pylori infection and the risk of gastric carcinoma. *The New England journal of medicine*, *325*(16), 1127–1131. https://doi.org/10.1056/NEJM199110173251603
3. Forman, D., Newell, D. G., Fullerton, F., Yarnell, J. W., Stacey, A. R., Wald, N., & Sitas, F. (1991). Association between infection with Helicobacter pylori and risk of gastric cancer: evidence from a prospective investigation. *BMJ (Clinical research ed.)*, *302*(6788), 1302–1305. https://doi.org/10.1136/bmj.302.6788.1302
4. Covacci, A., Censini, S., Bugnoli, M., Petracca, R., Burroni, D., Macchia, G., Massone, A., Papini, E., Xiang, Z., & Figura, N. (1993). Molecular characterization of the 128-kDa immunodominant antigen of Helicobacter pylori associated with cytotoxicity and duodenal ulcer. *Proceedings of the National Academy of Sciences of the United States of America*, *90*(12), 5791–5795. https://doi.org/10.1073/pnas.90.12.5791
5. Hatakeyama M. (2017). Structure and function of Helicobacter pylori CagA, the first-identified bacterial protein involved in human cancer. *Proceedings of the Japan Academy. Series B, Physical and biological sciences*, *93*(4), 196–219. https://doi.org/10.2183/pjab.93.013
6. Noto, J. M., & Peek, R. M., Jr (2012). The Helicobacter pylori cag Pathogenicity Island. *Methods in molecular biology (Clifton, N.J.)*, *921*, 41–50. https://doi.org/10.1007/978-1-62703-005-2_7
7. Hatakeyama, M. Oncogenic mechanisms of the *Helicobacter pylori* CagA protein. *Nat Rev Cancer* 4, 688–694 (2004). https://doi.org/10.1038/nrc1433
8. Yong, X., Tang, B., Li, BS. *et al. Helicobacter pylori* virulence factor CagA promotes tumorigenesis of gastric cancer *via* multiple signaling pathways. *Cell Commun Signal* 13, 30 (2015). https://doi.org/10.1186/s12964-015-0111-0
9. Wei, J., Nagy, T. A., Vilgelm, A., Zaika, E., Ogden, S. R., Romero-Gallo, J., Piazuelo, M. B., Correa, P., Washington, M. K., El-Rifai, W., Peek, R. M., & Zaika, A. (2010). Regulation of p53 tumor suppressor by Helicobacter pylori in gastric

epithelial cells. *Gastroenterology*, *139*(4), 1333–1343. https://doi.org/10.1053/j.gastro.2010.06.018

10. Wei, J., Noto, J. M., Zaika, E., Romero-Gallo, J., Piazuelo, M. B., Schneider, B., El-Rifai, W., Correa, P., Peek, R. M., & Zaika, A. I. (2015). Bacterial CagA protein induces degradation of p53 protein in a p14ARF-dependent manner. *Gut*, *64*(7), 1040–1048. https://doi.org/10.1136/gutjnl-2014-307295
11. Nešić, D., Buti, L., Lu, X., & Stebbins, C. E. (2014). Structure of the Helicobacter pylori CagA oncoprotein bound to the human tumor suppressor ASPP2. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(4), 1562–1567. https://doi.org/10.1073/pnas.1320631111

Full Figure Legends:

Figure 1. Structures of CagA +/- p53 peptide. CagA bound to the p53 peptide (purple color spheres) is shown on the left and the CagA after removing the coordinates for p53 peptide (apo CagA) is shown on the right.

Figure 2. RMSD & RMSF analysis of CagA+/-p53 peptide. Panels A and B show the protein C_α RMSD of apo CagA and CagA-p53 complex taken from PDB ID: 4IRV, respectively. Apo CagA shows an initial spike at 1 ns that was not seen in the complex but the complex showed a spike after 8 ns that was absent in the apo CagA protein. Panels C and D show the protein RMSF of apo and complex CagA, respectively. The overall RMSF for the apo CagA and p53 bound CagA are <2.4 Å and <1.6 Å, respectively

Figure 3. RMSD of CagA structure vs. homology model. RMSD of apo CagA from the crystal structure is shown in panel A as a reference. Panel B shows the RMSD of apo CagA homology model in which the initial spike at 1 ns is absent. Panels C and D show 25 ns and 50 ns MD simulations of the homology model. No major spikes were seen within the 25 ns MD simulation. However, there is a spike that progressively increased to almost 4 Å from 30 ns to 50 ns in the 50 ns MD simulation.

Figure 4. RMSF analysis of CagA homology model. The 25 ns and 50 ns MD simulations of the CagA homology model are shown on the left and right, respectively. Three regions: 50-75 (1.2 Å vs. 2.0 Å), 125-150 (1.6 Å vs. 2.8 Å) and 175-185 (2.8 Å vs. 3.2 Å) showed significant changes in the RMSF values.