ACS OMEGA

http://pubs.acs.org/journal/acsodf

Article

Computational Proteome-Wide Study for the Prediction of *Escherichia coli* Protein Targeting in Host Cell Organelles and Their Implication in Development of Colon Cancer

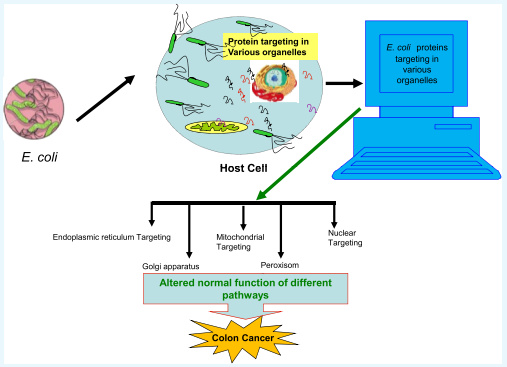
Shahanavaj Khan,* Sabika Zaidi, Abdulaziz Saleh Alouffi, Iftekhar Hassan, Ahmad Imran, and Rais Ahmad Khan

Cite This: *ACS Omega* 2020, 5, 7254−7261

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Enterohemorrhagic *Escherichia coli* infection is associated with gastrointestinal disorders, including diarrhea and colorectal cancer. Although evidences have established the involvement of *E. coli* in the growth of colon cancer, the molecular mechanisms of carcinogenesis of cancer growth and development are not well understood. We analyzed *E. coli* protein targeting in host cell organelles and the implication in colon cancer using in silico approaches. Our results indicated that many *E. coli* proteins targeted the endoplasmic reticulum (ER), ER membranes, Golgi apparatus, Golgi apparatus membranes, peroxisomes, nucleus, nuclear membrane, mitochondria, and mitochondrial membrane of host cells. These targeted proteins in ER, Golgi apparatus, peroxisomes, nucleus, and mitochondria may alter the normal functioning of various pathways including DNA repair, apoptosis, replication, transcription, and protein folding in *E. coli*-infected host cells. The results of the current in silico study provide insights into *E. coli* pathogenesis and may aid in designing new preventive and therapeutic strategies.

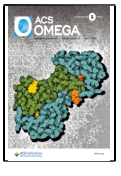


1. INTRODUCTION

Cancer is a big health challenge to the global populations. Lung cancer is diagnosed as the most common cancer in both sexes and the major cause of death followed by breast cancer, prostate cancer, colon cancer, stomach cancer, and liver cancer.[1] Recent evidence suggests that bacteria may contribute to the growth and development of various human cancers.[2,3] Colorectal cancer (CRC) is the third most common cancer diagnosis and the second deadliest malignancy in males and females.[4] Various factors are involved in colon cancer growth. The contribution of *Escherichia coli* in the etiology of colon cancer may be due to its promotion of chronic inflammation in the colon.[5] *E. coli* is connected to the development of colitis, which is a possible risk factor for colon cancer.[6,7] The diseases of the colon due to *E. coli*-mediated chronic inflammation include ulcerative colitis, inflammatory bowel disease, and Crohn's disease-induced colon cancer.[8−10] Host cells with these inflammatory conditions are colonized by enteroinvasive *E. coli* (EIEC), which has the potential of growth in the internal environment of the cell.[11,12] *E. coli* employs various strategies that enable the growth and development of CRC. The epidemiologic connection between EIEC and colon cancer is established. Recent reports have confirmed that colonization with *E. coli* is involved in the development of colitis and colon cancer.[7,10]

Polyketide synthase (PKS) pathogenicity islands increase the risk for *E. coli*-mediated development of colon cancer. The gene encoding PKS pathogenicity islands are responsible for the regulation of several important metabolites, including colibactin. Deletion of the PKS genes in *E. coli* decreased DNA damage and neoplastic lesions in IL10$^{-/-}$ mice colonized with *E. coli*.[13] The oncogenic potential of colibactin of *E. coli* B2 phylogenetic group strain has been confirmed.[14] Complete molecular studies of *E. coli* at the time of intracellular infection are needed. This may involve the induction of inflammatory diseases in colon cancer. *E. coli* proteins may work as components of the host cell proteome, and *E. coli* targeting of subcellular components alters the normal functioning of the host cells. Many novel insights have been made regarding the possible involvement of pathogenic strains of *E. coli* in colon cancer etiology. *E. coli* may alter several pathways responsible for the growth of colon cancer in susceptible persons. Proteins targeting intracellular pathogens in host cells play important

Received: November 27, 2019
Accepted: March 16, 2020
Published: March 30, 2020



https://dx.doi.org/10.1021/acsomega.9b04042
ACS Omega 2020, 5, 7254−7261

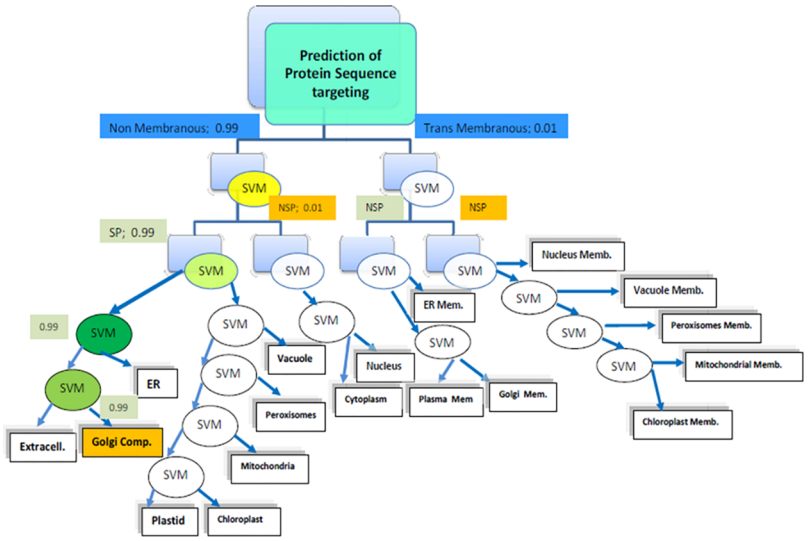


Figure 1. Scheme illustrated the *E. coli* protein targeting which may promote the growth of colon cancer.

roles in host cell regulation, including cell growth and apoptosis.[15] Bioinformatics tools are very important for the predictions of subcellular localization of proteins, structure of proteins, and function of different proteins. The isoelectric point (pI) of particular protein is the pH on which a protein carries zero net charge. The pI of specific enzymes or protein is important to understanding their biochemical utility and therefore analyzing the value of pI is a key aspect of proteomic studies. Numerous essential cellular processes are controlled in subcellular compartments of the host cell, including the nucleus, endoplasmic reticulum (ER), Golgi complex, and peroxisomes. *E. coli* proteins may adversely influence these compartments. Recent studies demonstrate that proteins targeting various bacterial proteomes in the nucleus, cytoplasm, ER, and mitochondria are implicated in cancer progression , including prostate, lungs, and colon cancers.[7,16−19] We analyzed the *E. coli* proteins targeting subcellular organelles, including the ER, Golgi complex, peroxisomes, nucleus, mitochondria, and cytoplasm, of the host cell that are implicated in the growth and development of colon cancer using bioinformatic approach.

2. RESULTS

2.1. Selection of Protein Database of *E. coli*. We have selected EcoProDB, which contains 561 *E. coli* proteins. These were used to predict *E. coli* protein targeting of subcellular organelles in *E. coli*-infected host cells.

2.2. Analysis of Subcellular Targeting of *E. coli* Proteins by LocTree3. LocTree3 predicted subcellular protein targeting in 18 classes of eukaryotes. The strategies used to predict protein targeting in subcellular organelles are illustrated in Figure 1. The diagram demonstrates the example of one of the proteins targeting the nonmembranous secretory pathway in the ER with an expected accuracy of 0.99 using an advanced profile kernel support vector machine (SVM). The results for the 561 *E. coli* proteins targeting organelles in the host cells are summarized in Table 1. LocTree3 analyses demonstrated that the highest number of proteins targeted by *E. coli* was located in the cytoplasm, followed by the mitochondrion. Lower numbers of proteins were targeted in other organelles, including the ER (3), ER membranes (4), Golgi apparatus membranes (3), mitochondrial membranes (11), nucleus (24), peroxisomes (9), plasma membranes (10), vacuole (3), and vacuole membranes (2). The results of the current protein targeting prediction study were demonstrated graphically in Figure 2. Descriptions and functions of the proteins targeting the ER, ER membranes, Golgi complex, Golgi complex membranes, peroxisomes, and peroxisome membranes are displayed in Table 1.

Table 1. Prediction of *E. coli* Protein Targeting in Subcellular Organelles of Host Cells by LocTree 3.0 Predictor

s. no.	sub-cellular organelles	no. of proteins targeted
1	chloroplast	92
2	chloroplast membrane	01
3	cytoplasm	210
4	ER	03
5	ER membrane	04
6	Golgi apparatus membrane	03
7	mitochondrion	103
8	mitochondrion membrane	11
9	nucleus	24
10	peroxisome	09
11	plasma membrane	10
12	plastid	04
13	secreted	69
14	vacuole	03
15	vacuole membrane	02

These subcellular organelles are dynamic and believed to be important for intracellular transportation and play a role in the growth of cancer. The ER is important for the regulation of housekeeping functions and maintaining homeostasis within the cell. When the normal functioning of the ER is disrupted, a phenomenon known as "ER stress" stimulates pathways to counteract this disruption, collectively termed the unfolded protein response (UPR). The UPR is associated with the growth of cancer due to chronic stress and alteration in the normal phenomenon of programmed cell death.[20,21] Similarly, various *E. coli* proteins target the Golgi complex, Golgi complex membranes, peroxisomes, and peroxisome membranes affecting the normal functioning of the host cell and inducing carcinogenesis. Complete descriptions of subcellular

https://dx.doi.org/10.1021/acsomega.9b04042
ACS Omega 2020, 5, 7254−7261

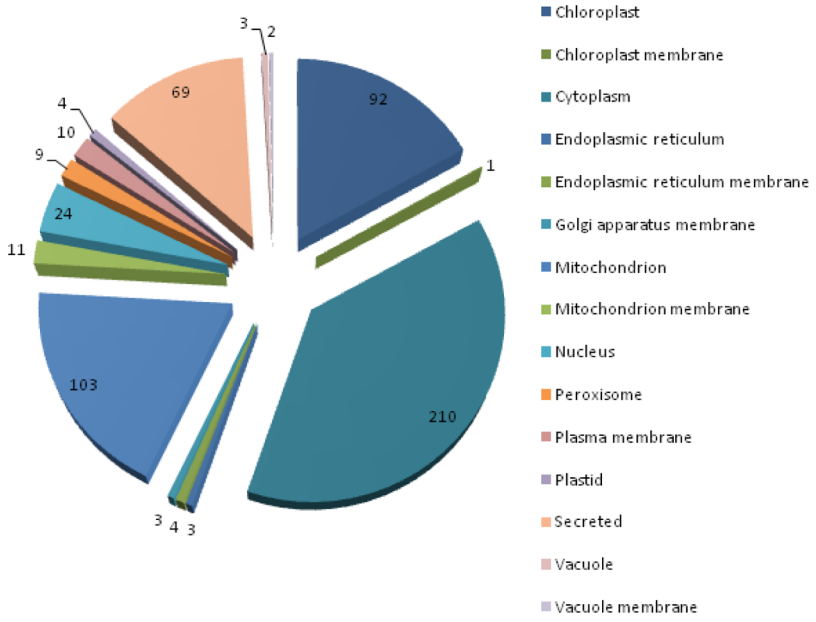


Figure 2. Image showing *E. coli* proteins targeting in different subcellular organelles.

protein targeting and their functions in bacterial cells are illustrated in the Supporting Information (Table S1).

2.3. Isoelectric Point Value and Molecular Weight of *E. coli* Proteins. Proteins targeting the lumen of the ER had M_ws between 18.2 and 43.9 kD and pI values between 5.1 and 5.5 predicted with up to 84% expected accuracy. Proteins targeting the ER membrane had M_ws between 34.7 and 64.5 kD and pI values between 5.1 and 5.5 predicted with 84% expected accuracy (Table 2). Nine proteins targeted peroxisome interiors and had M_ws between 20 and 56 kD and pI values between 4.7 and 6.33 predicted with 93% expected accuracy (Table 3). Proteins with M_ws between 9.1 and 24.9 kD and pI values between 4.55 and 9.75 were found to target the cytoplasm of host cells. The expected accuracy for proteins targeting the cytoplasm of host cells typically decreased with increasing M_ws. The *E. coli* proteins targeting the cytoplasm of host cells may disturb the normal physiology of the cell. The results of our work showed the highest number of *E. coli* proteins targeting the cytoplasm of the host cells. Three proteins targeted the Golgi apparatus membrane. One hundred and three proteins were predicted to target the mitochondria and eleven proteins targeted the mitochondrial membrane. Twenty-four proteins were predicted to target the nuclear interior. Descriptions of these proteins are in the Supporting Information (Table S1).

3. DISCUSSION

Possible factors in cancer progression include radiation, chemical carcinogens, inflammation, tobacco, mutations, diet, and infections. Among these, infections are associated with approximately 16% of cancers.[22] Bacterial infections are associated with the growth of cancers through abnormal functioning of various pathways in host cells. Bacterial infections were originally not believed to be cancer promoters. However, the association between the growth of bacteria and cancer has been confirmed. Various bacterial proteins are related to cancer etiology. The gut microbiota may be involved in the progression of CRC through various strategies and mechanisms. *E. coli* is a crucial bacteria of the human intestinal microbiota, but some pathogenic *E. coli* strains may induce chronic inflammatory responses and produce various toxins, including cyclomodulins, which may be involved in carcinogenesis.[23] We employed the latest version of LocTree3 which used advanced analyses and SVMs, including a fast, novel SVM profile kernel (16,17); insertion of annotation transfer-derived homologs and experimentally annotated localization by BLAST (12); caching of results for faster processing in repeated searches (19,20); and gene ontology annotations for calculation of outcomes (18).

Subcellular protein targeting demonstrated reliability, which was quantified by percentage-expected accuracy. Cyclomodulin-positive pathogenic strains of *E. coli* are found more commonly on mucosal membranes of CRC patients than on mucosal membranes of healthy individuals. Previous studies confirmed that various nuclear targeting *E. coli* proteins are involved in the growth of colon cancer in host cells.[7]

The group proteins known as cyclomodulins have the ability to modulate the cell cycle of the host.[14,24,25] Cyclomodulins create breaks in double-stranded DNA in host cells and direct the progression of colon cancer.[26] We focused on nonnuclear targeting *E. coli* proteins during infection. Proteins with high M_ws targeted the interior of the ER and ER membranes with more than 80% expected accuracy (Table 2). Quinoprotein glucose dehydrogenase (P15877), ADP-L-glycero-D-manno-heptose-6-epimerase (P67910), and chain length determinant protein (P76372) targeted the interior of the Golgi apparatus membrane (Table 3). These proteins may alter normal protein folding in host cells and contribute to the progression of colon cancer. Accretion of unfolded *E. coli* proteins in the ER and ER membranes may trigger ER stress and encourage cleavage of ATF6 that subsequently directs the activation of UPR genes (Figure 2). Activation of UPR has been connected to the growth of colon cancer.[27] Golgi fragmentation was recognized in cancer more than fifty years ago.[28] Studies of peroxisome proteins and metabolites support their pro-tumorigenic functions through alterations in peroxisome receptors.[29,30] Many of these proteins are toxins involved in the growth of cancer in affected host cells. Detailed information, and the

https://dx.doi.org/10.1021/acsomega.9b04042
ACS Omega 2020, 5, 7254−7261

Table 2. *E. coli* Protein Targeting in Different Subcellular Organelles of Host Cell (ER, ER Membranes, and Golgi Apparatus Membranes) of Host Cells with (%) Expected Accuracy, pI Value, Molecular Weight (dalton), and Functions

accession no.	protein name	mol. weight (dalton)	pI value	protein targeting results using LocTree3	expected accuracy (%)	functions in bacteria
P0A9X4	rod shape-determining protein MreB	36 952.4	5.19	ER	80	involved in the formation of the rod shape of the cell. May act as a negative regulator of FtsI
P24232	Flavohemoprotein	43 867.66	5.48	ER	81	involved in nitric oxide (NO) detoxification in anaerobic processes, termed nitric oxide dioxygenase (NOD) reaction that utilizes O_2 and NAD(P) H to convert NO to nitrate, which protects the bacterium from various noxious nitrogen compounds. Induced by NO (under aerobic conditions), nitrite, nitrate(under anaerobic conditions), nitroso compounds, and paraquat. Induced by low pH during anaerobic growth
P23869	peptidyl−prolyl *cis*−*trans* isomerase B	18 153.47	5.51	ER	84	see PpiA
P0AB77	2-amino-3-ketobutyrate coenzyme A ligase	43 117.04	5.64	ER membrane	83	negatively regulated by Lrp
P3687	hypothetical ABC transporter ATP-binding protein YadG	34 647.07	8.44	ER membrane	82	not available
P6075	lipid A export ATP-binding/permease protein MsbA	64 460.71	8.62	ER membrane	82	involved in lipid A export and possibly also in glycerophospholipid export and for biogenesis of the outer membrane
PO2942	methyl-accepting chemotaxis protein I	59 442.99	4.88	ER membrane	89	not available
P67910	ADP-L-glycero-D-manno-heptose-6-epimerase	34 893.17	4.8	Golgi apparatus membrane	89	catalyzes the interconversion between ADP-D-glycero-β-D-manno-heptose and ADP-L-glycero-β-D-manno-heptose via an epimerization at carbon 6 of the heptose. Completely inhibited by ADP and ADP-glucose and partially inhibited by ATP and NADH. It is induced by heat shock
P76372	chain length determinant protein	36 454.74	5.43	Golgi apparatus membrane	88	confers a modal distribution of chain length on the O-antigen component of lipopolysaccharide (LPS). Gives rise to a reduced number of short-chain molecules and increased numbers of longer molecules
P1587	quinoprotein glucose dehydrogenase	86 747.35	5.4	Golgi apparatus membrane	82	may be involved in energy conservation rather than in sugar metabolism

https://dx.doi.org/10.1021/acsomega.9b04042
ACS Omega 2020, 5, 7254−7261

Table 3. *E. coli* Protein Targeting in Peroxisomes of Host Cell, Including Percentage (%) Expected Accuracy, pI Value, Molecular Weight (dalton), and Functions

accession no.	protein name	mol. weight (dalton)	pI value	protein targeting results using LocTree3	expected accuracy (%)	functions in bacteria
P33232	L-lactate dehydrogenase	42 728.19	6.33	peroxisome	83	induced by L-lactate, aerobically
P0AET8	7α-hydroxysteroid dehydrogenase	26 778.59	5.22	peroxisome	83	7α-dehydroxylation of cholic acid, yielding deoxycholic acid and lithocholic acid, respectively. Highest affinity with taurochenodeoxycholic acid
P0A6F3	glycerol kinase	56 099.57	5.36	peroxisome	86	a key enzyme in the regulation of glycerol uptake and metabolism. The activity of this regulatory enzyme is affected by several metabolites. The inhibition by fructose 1,6-biphosphate causes alterations in the quaternary structure of the enzyme. Induced by L-α-glycerol-3-phosphate. Increases following exposure to the uncoupler of oxidative phosphorylation 2,4-dinitrophenol
P0ABU0	naphthoate synthase	31 633.08	5.99	peroxisome	93	converts O-succinylbenzoyl-CoA (OSB-CoA) to 1,4-dihydroxy-2-naphthoic acid (DHNA). Involved in menaquinone biosynthesis
P0ADG7	inosine-5′-monophosphate dehydrogenase	52 022.45	6.02	peroxisome	85	involved in the first step of GMP biosynthesis from IMP. Induced by low pH during anaerobic growth
P0ABH7	citrate synthase	48 014.99	6.21	peroxisome	87	involved in the tricarboxylic acid cycle and is allosterically inhibited by NADH
P0AD96	Leu/Ile/Val-binding protein	36 772.53	5.28	peroxisome	82	component of the leucine, isoleucine, valine, (threonine) transport system, which is one of the two periplasmic binding protein-dependent transport systems of the high-affinity transport of the branched-chain amino acids. Increases following exposure to the uncoupler of oxidative phosphorylation 2,4-dinitrophenol and in the physiological short-term adaptation to glucose-limitation
P04816	leucine-specific binding protein	36 982.71	5.07	peroxisome	82	component of the leucine-specific transport system, which is one of the two periplasmic binding protein-dependent transport systems of the high-affinity transport of the branched-chain amino acids. Increases following exposure to the uncoupler of oxidative phosphorylation 2,4-dinitrophenol. It decreases after benzoic acid treatment
P65556	putative Nudix hydrolase YfcD	20 375.86	4.7	peroxisome	80	not available

functions of proteins targeting the ER, ER membranes, Golgi complex, Golgi complex membranes, and peroxisomes, is shown in Tables 2 and 3. The Golgi apparatus is the main site of glycosylation and contains more than 250 glycosyltransferases. These enzymes work in a highly organized way due to the requirements of biosynthetic pathways.[31] Perturbation has occurred in the morphology of Golgi due to *E. coli* proteins altering the normal functioning of enzymes that help in the synthesis of glycosyl epitopes. Anomalous glycosylation in cancer arises through enhanced sialylation and has been observed in experimental models and clinical settings.[32,33]

Some proteins have the ability to alter the activity of mRNA, and protein degradation and translation, which alter the biochemical properties of host cells. Some of these proteins, discussed below, may be implicated in colon cancer etiology. The growth of cancer is strictly regulated by RNA binding proteins through alterations in expression at the posttranscriptional level.[34] The implication of RNA binding proteins in the etiology of ontogenesis is discussed in many studies.[35−37] Here, we observed various *E. coli* proteins targeting the host cytoplasm, which have the potential to bind to RNA and affect their functioning. For example, tRNA (uracil-5-)-methyltransferase (P23003) has the ability to form 5-methyl-uridine at position 54 (M-5-U54) in all tRNA which alters gene expression through growth rate-dependent regulation of transcription. Similarly, seryl-tRNA synthetase (P0A8L1) is also targeted. Alterations in the expression level of seryl-tRNA synthetase are implicated in cancer progression (Figure 2). Surprisingly, the etiology of various autoimmune disorders, neuronal pathologies, and colon cancer is connected with the altered expression of aminoacyl-tRNA synthetases.[38,39] DNA-directed RNA polymerase omega chain (P0A80) and transcription elongation factor GreA (P0A6W5) are also predicted to target the cytoplasm of host cells. Various proteins, including transcription antitermination protein NusG (P0AFG0), 16S rRNA processing protein RimM (P0A7X6), ribonuclease T (P30014), and ribonuclease III, are involved in the processing of ribosomal RNA precursors and some mRNAs (P0A7Y0). Ribosomal small subunit pseudouridine synthase A (P0AA43) of *E. coli* targeted the cytoplasm and possibly alters the normal functioning of host cells. Our prediction that ribonuclease T (P30014), which is involved in the biosynthesis of tRNA and removal of an AMP residue from uncharged tRNA, would target the cytoplasm of the host cells, was confirmed. Currently, research into the complex topic of ribonuclease-activated prodrugs is underway,[40] and the implication of *E. coli* ribonucleases in host cells should be reviewed because of their cytoplasmic targeting and crucial role in cancer development. Ribonucleases are enzymes involved in RNA cleavage through the action upon phosphodiester bonds, which directs various cellular functions. Mammalian ribonucleases are implicated in cancer prevention,[41] but little is known about the role of bacteria-derived ribonucleases in humans and hence requires rigorous wet-lab experimentation to reach a final conclusion. These targeted proteins may be factors in the growth and development of colon cancer. Discovering the role of RNA-binding proteins in cancer etiology and cytoplasmic targeting by *E. coli* originating from RNA-binding proteins requires evaluation of bacterial RNA-binding proteins. In addition, the role of translation-associated proteins in cancer is established.[42] Translation is an important step in regulating the expression of proteins. We predicted that ribosomal-protein-serine acetyltransferase (P13857) in the

https://dx.doi.org/10.1021/acsomega.9b04042
ACS Omega 2020, 5, 7254−7261

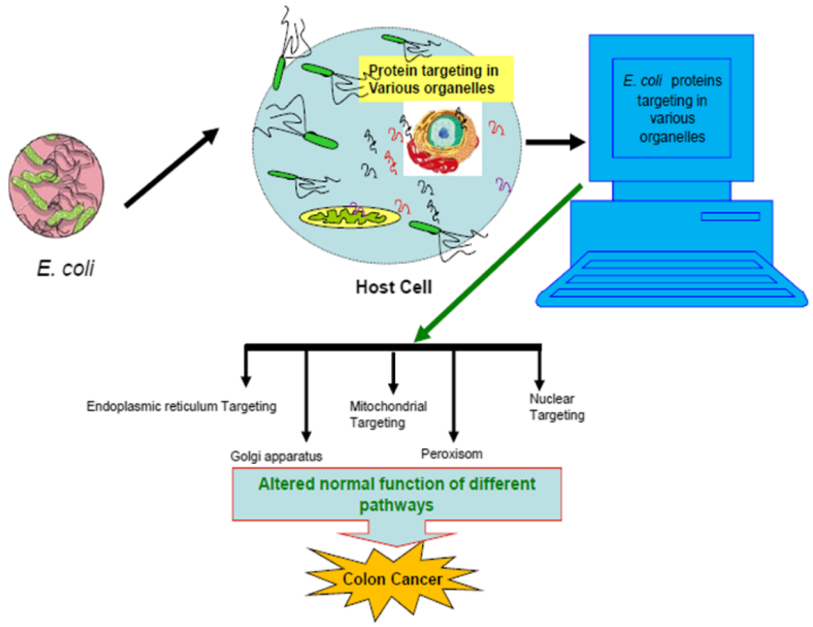


Figure 3. Strategies applied the in current in silico study to predict *E. coli* proteins targeting in subcellular organelles of host cells.

cytoplasm of host cells would provide a binding site for the regulation of protein synthesis.

Studies have demonstrated that increased expression of heat shock proteins (HSPs) are related to the growth and development of cancer. HSPs are involved in various aspects of carcinogenesis, including cell differentiation, cell proliferation, invasion, cell death, immune recognition, and metastasis.[43] We predicted various cytoplasm-targeting HSPs, including chaperone protein HchA (P31658), 33 kDa chaperonin (P0A6Y5), universal stress protein G (P39177), and trigger factor (P0A850), are necessary for proper protein folding in stress situations. Several bacterial HSPs have adverse effects in humans at the cellular level. Conversely, several bacterial HSPs protect epithelial cells from apoptosis and help to increase cell proliferation through the regulation of ERK and MAP kinase.[44,45] HSPs are highly conversed proteins, for example, Hsp70 has approximately 50% sequence similarity between humans and *E. coli*. The most highly conserved sequences may be 96% similar across species.[46] The early reports have revealed that the infection of a pathogenic strain of *E. coli* decreases or alters the expression of various tumor suppressors and oncogenes which may act as a potential factor for carcinogenesis through the induction of cellular proliferation. The BAX protein is considered as a strong factor to control the pro-apoptotic activities. BAX is usually dispensed in the cytoplasm of cells or loosely connected with the membrane of mitochondria. Upon stimulation of apoptosis, BAX is induced to undergo conformational change and regulate the normal functioning of apoptosis. Downregulation in the expression of BAX alter the normal phenomenon of apoptosis by the regulation of activation of caspase proteases. Therefore, different proteins of *E. coli* in various subcellular organelles play an important role in the growth of colon cancer.

4. CONCLUSION AND FUTURE WORK

The connection between *E. coli* and colon cancer is established. However, the *E. coli* target or *E. coli* linked factor, contributing to the progression of colon cancer, remains unknown. We concluded that various *E. coli* proteins targeting subcellular organelles of host cells play an important role in the growth of colon cancer. Although various prominent factors are associated with the growth of colon cancer but the infection with *E. coli* and their protein targets could not be ignored. Our results indicated that 24 *E. coli* proteins targeted into the nucleus of host cells. Therefore, future research should help to identify the potential functions of each protein. Some of these proteins may be pathogen-secreted host nuclear proteins, and then focus will be on studying the interactions between these proteins and host cells. In near future, the selected proteins may be used as a target vaccine eliciting an immune response against the *E. coli* and CRC. Available literature demonstrates that various toxins, such as cyclomodulin, have altered the normal functions of host cells. Additionally, we reported that *E. coli* proteins targeted the ER, ER membrane, Golgi apparatus, and peroxisomes, and this may be connected with the growth of colon cancer through abnormal pathway functioning. Alteration of ER and Golgi environments results in Golgi fragmentation that substantially alters the functions of residential Golgi glycosyltransferases and leads to the development of cancer-specific glycosyl epitopes. Alteration in the expression of oncogenes and tumor suppressor genes potentially associated with the growth of colon cancer. Emerging research in bacterial protein targeting in host cells confirms that *E. coli* protein targeting is a dynamic factor in the growth of colon cancer.

5. EXPERIMENTAL SECTION OR COMPUTATIONAL METHODS

5.1. Selection of *E. coli* Proteins Database. We utilized the *E. coli* protein database (EcoProDB) supported by the Korean Advanced Institute of Science and Technology (KAIST).[47] This database contains 561 *E. coli* proteins with descriptions available from Swiss-Prot and the *E. coli* cell envelope protein data collection.[48,49]

5.2. Analysis of Subcellular *E. coli* Protein Targeting by LocTree3. We predicted subcellular *E. coli* protein targeting in host cells using LocTree3.[50] This prediction

https://dx.doi.org/10.1021/acsomega.9b04042
ACS Omega 2020, 5, 7254−7261

system uses kernel-based SVMs and homology-based inference to analyze native subcellular protein targeting in eighteen classes of eukaryotes, six classes of bacteria, and three classes of Archaea. LocTree3 predicted three localization classes: three in Archaea (plasma membrane, cytoplasm, and extracellular), six in bacteria (plasma membrane, cytoplasm, periplasmic space, outer membrane, fimbrium, and extracellular), and eighteen in Eukaryota (nucleus, nuclear membrane, cytoplasm, ER, ER membranes, mitochondria, mitochondrial membrane, Golgi apparatus, Golgi membranes, peroxisomes, peroxisome membranes, vacuole, vacuole membranes, chloroplast, chloroplast membrane plastids, plasma membranes, and extracellular). The scheme of prediction of protein targeting of *E. coli* in different subcellular organelles of the host cell and their implication in growth of cancer are shown in Figure 3.

5.3. Isoelectric Point Value and Molecular Weight of *E. coli* Proteins. Isoelectric point (pI) values and molecular weight (M_w) of *E. coli* proteins were predicted by ExPASy using the compute pI/M_w tool. This computational tool predicts the theoretical pI and M_w from a list of UniProt Knowledgebase (Swiss-Prot or TrEMBL) entries or for user-entered sequences.[48]

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.9b04042.

Analysis of *E. coli* proteins targeting in various subcellular organelles of the host cell and their potential functions in bacterial cells (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Shahanavaj Khan − *Bioinformatics and Biotechnology Unit, Department of Biosciences, SRGC, Muzaffarnagar 251001, UP, India; Department of Pharmaceutics, College of Pharmacy, King Saud University, Riyadh 11451, Saudi Arabia;* orcid.org/0000-0002-4049-2244; Phone: +91 9219993262; Email: khan.shahanavaj@gmail.com

Authors

Sabika Zaidi − *Bioinformatics and Biotechnology Unit, Department of Biosciences, SRGC, Muzaffarnagar 251001, UP, India*

Abdulaziz Saleh Alouffi − *King Abdulaziz City for Science and Technology, Riyadh 11442, Saudi Arabia*

Iftekhar Hassan − *Department of Zoology, College of Science, King Saud University, Riyadh 11451, Saudi Arabia*

Ahmad Imran − *King Abdullah Institute for Nanotechnology, King Saud University, Riyadh 11451, Saudi Arabia*

Rais Ahmad Khan − *Department of Chemistry, College of Science, King Saud University, Riyadh 11451, Saudi Arabia*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.9b04042

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors extend their appreciation to the Deanship of Scientific Research at King Saud University for funding this work through research group no (RG-1440-059).

■ REFERENCES

(1) Bray, F.; Ferlay, J.; Soerjomataram, I.; Siegel, R. L.; Torre, L. A.; Jemal, A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Ca-Cancer J. Clin.* **2018**, *68*, 394−424.

(2) Xia, J.; Chiu, L.-Y.; Nehring, R. B.; Bravo Núñez, M. A.; Mei, Q.; Perez, M.; Zhai, Y.; Fitzgerald, D. M.; Pribis, J. P.; Wang, Y.; Hu, C. W.; Powell, R. T.; LaBonte, S. A.; Jalali, A.; Matadamas Guzmán, M. L.; Lentzsch, A. M.; Szafran, A. T.; Joshi, M. C.; Richters, M.; Gibson, J. L.; Frisch, R. L.; Hastings, P. J.; Bates, D.; Queitsch, C.; Hilsenbeck, S. G.; Coarfa, C.; Hu, J. C.; Siegele, D. A.; Scott, K. L.; Liang, H.; Mancini, M. A.; Herman, C.; Miller, K. M.; Rosenberg, S. M. Bacteria-to-Human Protein Networks Reveal Origins of Endogenous DNA Damage. *Cell* **2019**, *176*, 127−143.e24.

(3) Mager, D. L. Bacteria and cancer: cause, coincidence or cure? A review. *J. Transl. Med.* **2006**, *4*, 14.

(4) Recio-Boiles, A.; Cagir, B. Cancer, Colon. *StatPearls [Internet]*; StatPearls Publishing: Treasure Island (FL), 2019.

(5) Zarei, O.; Arabestan, M. R.; Majlesi, A.; Mohammadi, Y.; Alikhani, M. Y. Determination of virulence determinants of Escherichia coli strains isolated from patients with colorectal cancer compared to the healthy subjects. *Gastroenterol. Hepatol. Bed Bench* **2019**, *12*, 52−59.

(6) Bonnet, M.; Buc, E.; Sauvanet, P.; Darcha, C.; Dubois, D.; Pereira, B.; Dechelotte, P.; Bonnet, R.; Pezet, D.; Darfeuille-Michaud, A. Colonization of the human gut by E. coli and colorectal cancer risk. *Clin. Cancer Res.* **2014**, *20*, 859−867.

(7) Khan, S. Potential role of Escherichia coli DNA mismatch repair proteins in colon cancer. *Crit. Rev. Oncol.-Hematol.* **2015**, *96*, 475−482.

(8) Chassaing, B.; Darfeuille−Michaud, A. The commensal microbiota and enteropathogens in the pathogenesis of inflammatory bowel diseases. *Gastroenterology* **2011**, *140*, 1720−1728.e3.

(9) Kaser, A.; Zeissig, S.; Blumberg, R. S. Inflammatory bowel disease. *Annu. Rev. Immunol.* **2010**, *28*, 573−621.

(10) Khan, S.; Imran, A.; Malik, A.; Chaudhary, A. A.; Rub, A.; Jan, A. T.; Syed, J. B.; Rolfo, C. Bacterial imbalance and gut pathologies: Association and contribution of E. coli in inflammatory bowel disease. *Crit. Rev. Clin. Lab. Sci.* **2019**, *56*, 1−17.

(11) Prorok-Hamon, M.; Friswell, M. K.; Alswied, A.; Roberts, C. L.; Song, F.; Flanagan, P. K.; Knight, P.; Codling, C.; Marchesi, J. R.; Winstanley, C.; Hall, N.; Rhodes, J. M.; Campbell, B. J. Colonic mucosa-associated diffusely adherent afaC+ Escherichia coli expressing lpfA and pks are increased in inflammatory bowel disease and colon cancer. *Gut* **2014**, *63*, 761−770.

(12) Martin, H. M.; Campbell, B. J.; Hart, C. A.; Mpofu, C.; Nayar, M.; Singh, R.; Englyst, H.; Williams, H. F.; Rhodes, J. M. Enhanced Escherichia coli adherence and invasion in Crohn's disease and colon cancer. *Gastroenterology* **2004**, *127*, 80−93.

(13) Arthur, J. C.; Perez-Chanona, E.; Mühlbauer, M.; Tomkovich, S.; Uronis, J. M.; Fan, T.-J.; Campbell, B. J.; Abujamel, T.; Dogan, B.; Rogers, A. B.; Rhodes, J. M.; Stintzi, A.; Simpson, K. W.; Hansen, J. J.; Keku, T. O.; Fodor, A. A.; Jobin, C. Intestinal inflammation targets cancer-inducing activity of the microbiota. *Science* **2012**, *338*, 120−123.

(14) Secher, T.; Samba-Louaka, A.; Oswald, E.; Nougayrède, J.-P. Escherichia coli producing colibactin triggers premature and transmissible senescence in mammalian cells. *PLoS One* **2013**, *8*, No. e77157.

(15) Gao, L.-Y.; Kwaik, Y. A. The modulation of host cell apoptosis by intracellular bacterial pathogens. *Trends Microbiol.* **2000**, *8*, 306−313.

(16) Khan, A. A. In Silico Prediction of Escherichia coli Proteins Targeting the Host Cell Nucleus, with Special Reference to Their Role in Colon Cancer Etiology. *J. Comput. Biol.* **2014**, *21*, 466.

(17) Khan, S.; Zakariah, M.; Palaniappan, S. Computational prediction of Mycoplasma hominis proteins targeting in nucleus of host cell and their implication in prostate cancer etiology. *Tumor Biol.* **2016**, *37*, 10805−10813.

(18) Zakariah, M.; Khan, S.; Chaudhary, A. A.; Rolfo, C.; Ben Ismail, M. M.; Alotaibi, Y. A. To Decipher the Mycoplasma hominis Proteins Targeting into the Endoplasmic Reticulum and Their Implications in Prostate Cancer Etiology Using Next-Generation Sequencing Data. *Molecules* **2018**, *23*, 994.
(19) Khan, S.; Imran, A.; Khan, A. A.; Abul Kalam, M.; Alshamsan, A. Systems Biology Approaches for the Prediction of Possible Role of Chlamydia pneumoniae Proteins in the Etiology of Lung Cancer. *PLoS One* **2016**, *11*, No. e0148530.
(20) Wang, W.-A.; Groenendyk, J.; Michalak, M. Endoplasmic reticulum stress associated responses in cancer. *Biochim. Biophys. Acta, Mol. Cell Res.* **2014**, *1843*, 2143−2149.
(21) Manalo, R. V. M. Anastasis and the ER stress response: Solving the paradox of the unfolded protein response in cancer. *Med. Hypotheses* **2017**, *109*, 25−27.
(22) de Martel, C.; Ferlay, J.; Franceschi, S.; Vignat, J.; Bray, F.; Forman, D.; Plummer, M. Global burden of cancers attributable to infections in 2008: a review and synthetic analysis. *Lancet Oncol.* **2012**, *13*, 607−615.
(23) Buc, E.; Dubois, D.; Sauvanet, P.; Raisch, J.; Delmas, J.; Darfeuille-Michaud, A.; Pezet, D.; Bonnet, R. High prevalence of mucosa-associated E. coli producing cyclomodulin and genotoxin in colon cancer. *PLoS One* **2013**, *8*, No. e56964.
(24) Nougayrede, J.-P.; Homburg, S.; Taieb, F.; Boury, M.; Brzuszkiewicz, E.; Gottschalk, G.; Buchrieser, C.; Hacker, J.; Dobrindt, U.; Oswald, E. Escherichia coli induces DNA double-strand breaks in eukaryotic cells. *Science* **2006**, *313*, 848−851.
(25) Cougnoux, A.; Dalmasso, G.; Martinez, R.; Buc, E.; Delmas, J.; Gibold, L.; Sauvanet, P.; Darcha, C.; Déchelotte, P.; Bonnet, M.; Pezet, D.; Wodrich, H.; Darfeuille-Michaud, A.; Bonnet, R. Bacterial genotoxin colibactin promotes colon tumour growth by inducing a senescence-associated secretory phenotype. *Gut* **2014**, *63*, 1932−1942.
(26) Cuevas-Ramos, G.; Petit, C. R.; Marcq, I.; Boury, M.; Oswald, E.; Nougayrede, J.-P. Escherichia coli induces DNA damage in vivo and triggers genomic instability in mammalian cells. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 11537−11542.
(27) Piton, N.; Wason, J.; Colasse, É.; Cornic, M.; Lemoine, F.; Le Pessot, F.; Marguet, F.; Sabourin, J.-C. Endoplasmic reticulum stress, unfolded protein response and development of colon adenocarcinoma. *Virchows Arch.* **2016**, *469*, 145−154.
(28) Petrosyan, A. Onco-Golgi: Is Fragmentation a Gate to Cancer Progression? *J. Biochem. Mol. Biol.* **2015**, *01*, 1−11.
(29) Gou, Q.; Gong, X.; Jin, J.; Shi, J.; Hou, Y. Peroxisome proliferator-activated receptors (PPARs) are potential drug targets for cancer therapy. *Oncotarget* **2017**, *8*, 60704−60709.
(30) Dahabieh, M. S.; Di Pietro, E.; Jangal, M.; Goncalves, C.; Witcher, M.; Braverman, N. E.; del Rincón, S. V. Peroxisomes and cancer: The role of a metabolic specialist in a disease of aberrant metabolism. *Biochim. Biophys. Acta, Rev. Cancer* **2018**, *1870*, 103−121.
(31) Stanley, P. Golgi glycosylation. *Cold Spring Harbor Perspect. Biol.* **2011**, *3*, a005199.
(32) Schultz, M. J.; Swindall, A. F.; Bellis, S. L. Regulation of the metastatic cell phenotype by sialylated glycans. *Cancer Metastasis Rev.* **2012**, *31*, 501−518.
(33) Yu, X.; Zhao, Y.; Wang, L.; Chen, X.; Su, Z.; Zhang, H.; Yuan, Q.; Wang, S. Sialylated β1, 6 branched N-glycans modulate the adhesion, invasion and metastasis of hepatocarcinoma cells. *Biomed. Pharmacother.* **2016**, *84*, 1654−1661.
(34) Kechavarzi, B.; Janga, S. Dissecting the expression landscape of RNA-binding proteins in human cancers. *Genome Biol.* **2014**, *15*, R14.
(35) Wurth, L. Versatility of RNA-Binding Proteins in Cancer. *Comp. Funct. Genomics* **2012**, *2012*, 178525.
(36) van Kouwenhove, M.; Kedde, M.; Agami, R. MicroRNA regulation by RNA-binding proteins and its implications for cancer. *Nat. Rev. Cancer* **2011**, *11*, 644−656.
(37) Lukong, K. E.; Chang, K.-w.; Khandjian, E. W.; Richard, S. RNA-binding proteins in human genetic disease. *Trends Genet.* **2008**, *24*, 416−425.
(38) Park, S. G.; Schimmel, P.; Kim, S. Aminoacyl tRNA synthetases and their connections to disease. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 11043−11049.
(39) Ikromov, O.; Alkamal, I.; Magheli, A.; Ratert, N.; Sendeski, M.; Miller, K.; Krause, H.; Kempkensteffen, C. Functional Epigenetic Analysis of Prostate Carcinoma: A Role for Seryl-tRNA Synthetase? *J. Biomarkers* **2014**, *2014*, 362164.
(40) Ellis, G. A.; McGrath, N. A.; Palte, M. J.; Raines, R. T. Ribonuclease-Activated Cancer Prodrug. *ACS Med. Chem. Lett.* **2012**, *3*, 268−272.
(41) Kim, W.-C.; Lee, C. H. The role of mammalian ribonucleases (RNases) in cancer. *Biochim. Biophys. Acta, Rev. Cancer* **2009**, *1796*, 99−113.
(42) Ruggero, D. Translational control in cancer etiology. *Cold Spring Harbor Perspect. Biol.* **2012**, *5*, a012336.
(43) Ciocca, D. R.; Calderwood, S. K. Heat shock proteins in cancer: diagnostic, prognostic, predictive, and treatment implications. *Cell Stress Chaperones* **2005**, *10*, 86−103.
(44) Zhang, L.; Pelech, S. L.; Mayrand, D.; Grenier, D.; Heino, J.; Uitto, V.-J. Bacterial heat shock protein-60 increases epithelial cell proliferation through the ERK1/2 MAP kinases. *Exp. Cell Res.* **2001**, *266*, 11−20.
(45) Zhang, L.; Pelech, S.; Uitto, V.-J. Bacterial GroEL-like heat shock protein 60 protects epithelial cells from stress-induced death through activation of ERK and inhibition of caspase 3. *Exp. Cell Res.* **2004**, *292*, 231−240.
(46) Schlesinger, M. J. Heat shock proteins. *J. Biol. Chem.* **1990**, *265*, 12111−12114.
(47) Yun, H.; Lee, J. W.; Jeong, J.; Chung, J.; Park, J. M.; Myoung, H. N.; Lee, S. Y. EcoProDB: the Escherichia coli protein database. *Bioinformatics* **2007**, *23*, 2501−2503.
(48) Boeckmann, B.; Bairoch, A.; Apweiler, R.; Blatter, M. C.; Estreicher, A.; Gasteiger, E.; Martin, M. J.; Michoud, K.; O'Donovan, C.; Phan, I.; Pilbout, S.; Schneider, M. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.* **2003**, *31*, 365−370.
(49) The UniProt Consortium. Update on activities at the Universal Protein Resource (UniProt) in 2013. *Nucleic Acids Res.* **2013**, *41*, D43−D47.
(50) Goldberg, T.; Hecht, M.; Hamp, T.; Karl, T.; Yachdav, G.; Ahmed, N.; Altermann, U.; Angerer, P.; Ansorge, S.; Balasz, K.; Bernhofer, M.; Betz, A.; Cizmadija, L.; Do, K. T.; Gerke, J.; Greil, R.; Joerdens, V.; Hastreiter, M.; Hembach, K.; Herzog, M.; Kalemanov, M.; Kluge, M.; Meier, A.; Nasir, H.; Neumaier, U.; Prade, V.; Reeb, J.; Sorokoumov, A.; Troshani, I.; Vorberg, S.; Waldraff, S.; Zierer, J.; Nielsen, H.; Rost, B. LocTree3 prediction of localization. *Nucleic Acids Res.* **2014**, *42*, W350−W355.

https://dx.doi.org/10.1021/acsomega.9b04042
ACS Omega 2020, 5, 7254−7261