Systems Biology Study of Yeast Mitogen Activated Protein Kinase (MAPK) Cascade for Novel Drug Target Identification against Fungal Pathogens

Sheikh Md. Enayetul Babar, Md. Fahmid Islam, K. M. Taufiqur Rahman, and Sharmin Sultana Sumi

***Abstract*—Novel drug target identification and drug discovery are potential fields of systems biology which is recently being implemented against various pathogens based on differential biological processes of host and pathogen. Due to closer extent of biological similarity of eukaryotic fungal pathogens with their hosts, it has been urgent to find out novel drug targets against fungal pathogens. It is seen that mitogen activated protein kinase (MAPK) cascade transmits signals from outer cell surface to the nucleus and is involved in fungal survival mechanisms against environmental stress conditions. MAPK cascade molecules can be good targets of antifungal drugs to avoid fungal survival against conventional drugs. From these perspectives, systems biology approaches have been undertaken with an aim to assess the MAPK molecules as antifungal drug targets to find out the best one using Yeast (*Saccharomyces cerevisiae*) as model organism, Human as host and *Candida albicans* as pathogen. Comparative proteomic study, protein-protein interaction study, sequence and structural analysis study, molecular docking study and mathematical modeling study have been conducted in this regard. A combined prioritizing scoring system has been used to identify the most optimal MAPK target based on weighted decision matrix and combined matrix position score (CMPS), stability of the matrix position score (SD) and total score (TS) have been introduced. MAPKs having TS less than the average TS (72.55) were screened out as better MAPK targets. The approach predicts SLT2/MPK1 among the 11 MAPK molecules the best target of antifungal drugs having the lowest TS (54.49).**

***Index Terms*—Antifungal drug, MAPK cascade, novel drug target, signal transduction pathway, systems biology.**

I. Introduction

Deciphering the systemic properties of living entities has emerged as a central issue in recent years and systems biology is playing non-trivial role in the elucidation of form, function and behavior of life more effectively. Such systems biological approaches are necessary for us to make important breakthroughs in development of experimental, analytical and computational methods for achieving the deeper insight of life [1], [2]. Novel drug target identification and exploration of new drugs against emerging pathogens are important issues in recent medical science which are mainly based on differential biological processes of host and pathogens. Recent advancement in system biological approaches facilitated with bioinformatics, computational and mathematical biology has been a potential way in this regard [3]. Genomes of a large number of organisms have been sequenced over the last few years and the complete genome information allows analysis from different theoretical and practical perspectives.

Computational analysis of any particular gene provides a primary idea of its function and importance. Any gene necessary for the survival and viability of a pathogenic organism having little or no similarity to those of the host organism may be a possible potential drug target [4]. The differences in the proteins of the host and the pathogen can be effectively used for designing a drug specifically targeting the pathogen [5]. The computational approach based on comparative genomics and proteomics study has been used to investigate novel drug targets in some pathogenic organisms such as *Pseudomonas aeruginosa*, *Helicobacter pylori*, *Aeromonas hydrophila*, *Neisseria meningitidis etc.* [4]-[8]. Determination of the three dimensional structures of potential target proteins is a major task in drug discovery process which is a time consuming expensive process. Homology modeling is a useful tool to predict three dimensional structures of target proteins of unknown structures [9]. Molecular docking study is a tool to find out most effective drug ligands against target proteins based on three dimensional interactions between target protein and ligand [10]. Mathematical models can be a tool for global scale understanding of biological processes consequently unique drug target identification from dynamic point of view [11].

Resistance to antibiotics in pathogenic fungi is a problem of special importance in recent time. Though bacterial drug resistance has been studied over for the last few decades, such studies for pathogenic fungi have got recent interest [12]. Treatment with antifungal drugs often results in the appearance of resistant strains of fungi. Different signal transduction mechanisms are important for fungi in environmental sensing and survival that directly or indirectly lead to drug resistance or reduction of stress exerted by antifungal drugs [13]. MAPK signaling cascade that transmits signals from outer cell surface to the nucleus has known to be one of the major players in such processes [14]. The role of MAPK cascade in fungi is very central which predicts the MAPK molecules potential drug targets [15], [16].

Manuscript received September 24, 2012; revised January 15, 2013. This work was supported by the Korea Institute of Science and technology (KIST), Korea.

Sheikh Md. Enayetul Babar, Md. Fahmid Islam, K. M. Taufiqur Rahman, and Sharmin Sultana Sumi are with Biotechnology and Genetic Engineering Discipline of Khulna University, Bangladesh (e-mail: babarku@yahoo.com, fahmid_islam9891@yahoo.com, taufiqbge@gmail.com, sharmin.sumi24@yahoo.com).

Candida albicans, adiploid fungus growing both as yeast and filamentous cells are causal agent of opportunistic oral and genital infections in humans [17]. Systemic fungal infections (fungemias) including those by *C. albicans* have emerged as important causes of morbidity and mortality in immunocompromised patients (e.g., AIDS, cancer chemotherapy, and organ or bone marrow transplantation) [18].

In this paper a drug target identification process has been described based on the previously described strategies. The major objective was to assess the MAPK cascade molecules of yeast as drug targets to predict similar type of target proteins of *Candida albicans* as best potential drug targets.

II. Materials and Methods

The whole study was conducted using *Saccharomyces cerevisiae* (baker's yeast) as model organism, targeting *Candida albicans* (pathogenic yeast/filamentous fungi) as pathogen in the context of *Homo sapiens* (human) as host. The methods were devised to find out best drug targets (MAPK molecules) in *S. cerevisiae MAPK* cascades. A series of studies were undertaken and results from different studies have been combined to identify the best targets among the MAPK molecules in a relative scale. As *S. cerevisiae* is a model system to study different eukaryotes (from fungus to human) and other fungus show basic genomic, morphogenetic, biochemical and physiological similarity with *S. cerevisiae* the predictions come out from the approach were theoretically consistent with the possibility to be implied to the pathogen *C. albicans* and other fungal pathogens as well.

Protein Lounge Pathway Database, KEGG, NCBI and UniProtKB were used to search in detail signal transduction pathways related to environmental information processing of yeast and MAPK cascade molecules [19]-[22]. For comparative proteomic study of yeast with host (human) and pathogen (*C. albicans*), BLAST to Proteome (protein blast) programs were used [23], [24]. Best 10 proteins were taken for further analysis from each BLAST run based on bit score (total score) and e-value (expected value). Expected threshold was 10 for each BLAST run. Lower homology with human proteome denotes to higher feasibility and higher homology with pathogen proteome show higher feasibility (here the bit scores of the least matched and best matched human or pathogen proteins were used to compare) Protein-Protein interaction study was done based on STRING program [25]. Best 20 proteins were taken for further analysis. To understand the extent of protein-protein interaction average interaction score (AVPIP) was measured by dividing the summation of individual interactions with the number of interacting proteins (here 20). Required confidence (score) was medium confidence (0.400). Higher interaction extent denotes to the involvement of the MAPKs in higher number of biological processes making the relative importance of the MAPKs. Homology modeling study was conducted based on Swiss-Model program (automated homology modeling) [26] as currently the 3-D Structures of MAPK molecules (except FUS3) are not available in the public databases. The models were further analyzed and visualized using Discovery Studio Visualizer software package [27]. The model quality was assessed based on QMEANscore4 and QMEAN Z-Score (average quality=AVQ score was measured). The quality of homology models definitely influences the predictions. Molecular docking study was conducted based on the DOCK Blaster program (clean–fragments database of 354309 molecules was searched [28]. Best ligands for each MAPK molecule was further visualized and analyzed by chimera software package based on binding energy (ligand binding=LB score) [29]. The availability of ligands of lower binding energy denotes to the higher feasibility.

Molecular pathway model were constructed for each MAPK cascade of the signal transduction pathways. The MAPK cascade for yeast assumed to be the same in various organisms (experimentally proved in different organisms and cell free systems as well) including one single phosphorylation stage and two dual phosphorylation stages. Literature survey was the key to construct such detailed structure of cascade [30], [31]. All the kinetic data were collected from literature [30]-[33].

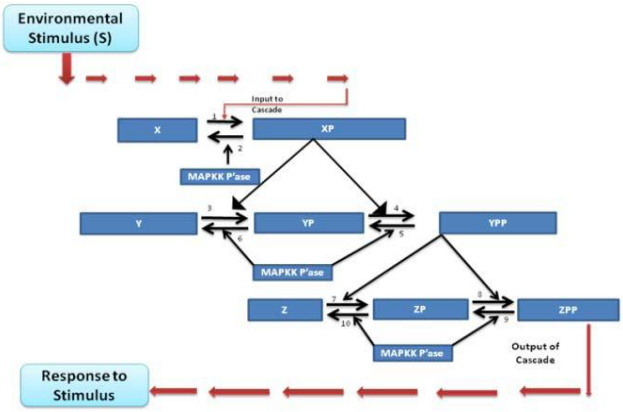

Fig. 1. A generalized MAPK cascade (Here, X = MAPKKK= STE11, BCK; Y = MAPK = STE7, PBS2, MKK1, MKK2; Z = MAPK = KSS1, FUS3, HOG1, SLT2, SMK1)

Finally a generalized MAPK cascade was built based on general notations (X= MAPKKK, Y= MAPKK, and Z= MAPK) (Fig. 1). The models by Huang and Ferrell [30] and Khodolenko [31] were downloaded from BioModels Database [32] in SBML and COPASI format. The models were simulated using copasi software package [34]. Time course simulation was done to analyze signal sensitivity. The models for mass action kinetics and enzymatic kinetics (Michaelis– Menton kinetics) for the yeast MAPK cascade were built and simulated using MATLAB environment [35]. Mathematical models for each reversible reaction were designed using ordinary differential equations "(1-6)". The simulations were done to satisfy steady state. Simulation was done for the entry points (3 points) of the cascade assuming concentration a function of time.

$$d/dt\,[\mathrm{X}] = v_2 - v_1 = -K_1\,[\mathrm{X}] + K_2\,[\mathrm{XP}] \qquad (1)$$

$$d/dt\,[\mathrm{Y}] = v_6 - v_3 = -K_3\,[\mathrm{Y}] + K_6\,[\mathrm{YP}] \qquad (2)$$

$$d/dt\,[\mathrm{Z}] = v_{10} - v_7 = -K_7\,[\mathrm{Z}] + K_{10}\,[\mathrm{ZP}] \qquad (3)$$

where: K_n = Reaction rate constant
$V_{n=}$ Reaction rate
(Equations for Mass action kinetics)

$$d/dt\,[Z] = k_{3c}YPPZ/(K_3+Z) - k_{3i}ZP/(ZP+K_{3i}) \quad (4)$$

$$d/dt\,[Y] = k_{2c}XPY/(K_2+Y) - k_{2i}YP/(YP+K_{2i}) \quad (5)$$

$$d/dt\,[X] = -k_{1c}SX/(K_1+X) + k_{1i}XP/(XP+K_{1i}) \quad (6)$$

where: k_{nc}= Catalytic activation rate
k_{ni} = Catalytic inactivation rate
K_n = Binding affinity
K_{ni} = Binding affinity of inactivation
(Equations for Enzymatic kinetics)

The MAPK molecules have been assigned relative ascending values (1-11 ranks) according to the feasibility of MAPK molecules as drug targets from each study (for signal sensitivity study the ranks were 1, 2 and 3 for first, second and third stage phosphorylation respectively as the signal sensitivity increases towards according to the progression of the cascade).To identify best MAPK molecule as drug target, an 11 by 6 weighted decision matrix [30] was formed to calculate the combined score for the six parameters above for each MAPK molecule. The relative values from previous studies were used to form the matrix. MAPK molecules were considered according to row and the parameters were considered according to column. The least scoring molecule was assumed to be the best target in comparison to others. A weighted matrix was formed assuming 3 for the comparative proteomic values, 2 for the protein-protein interaction values, 1 for homology modeling values and molecular docking values, and 3 for signal sensitivity values (this was done to represent the relative importance of the parameters). Combined matrix position score (CMPS) was calculated based on the summation of individual scores of 6 positions of each MAPK in the matrix. The standard deviations (SD) of matrix positions of different MAPK molecules were calculated to understand their positional stability based on different parameters [37]. Finally total score (TS) has been calculated as the summation of CMPS and SD.

III. Results and Discussions

The results from comparative proteomic study, protein-protein interaction study, structural analysis study, molecular docking study and mathematical modeling study are given below. Different proteins from different study show better effectiveness as drug targets. MKK2 shows least matching with human proteome, HOG1 shows best matching with pathogen proteome, SLT2 shows highest protein-protein interaction extent, FUS3 shows best model quality and STE11 shows best ligand binding (Fig. 2 (a-c). and Fig. 3(a-b)).

The above studies give us only static feature of the MAPK cascade system. But we should also consider the dynamic features of the cascade. The simple time course simulation shows the expected temporal sequence of kinase activation, from MAPKKK to the final effector MAPK. It shows that the activity of MAPK reaches its maximal level before MAPKKK and also hints at the increase in sensitivity along the levels of the cascade. The dose-response plot (Fig. 4) directly shows the strong increase in sensitivity along the levels of the cascade with the MAPK curve predicted to be the steepest.

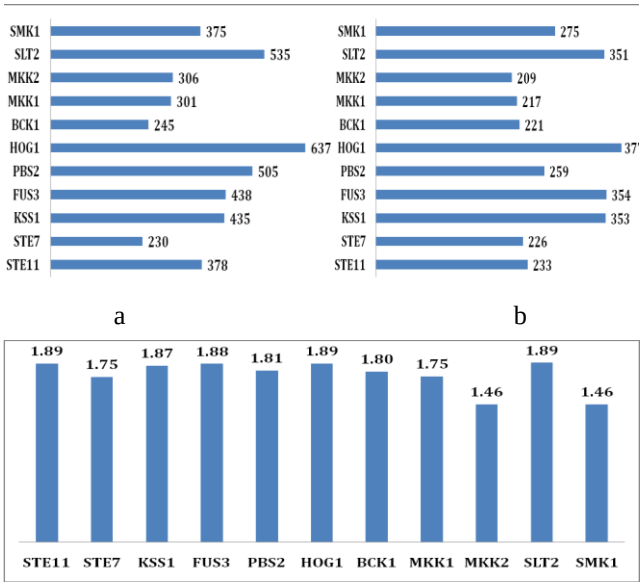

Fig. 2. Comparison of yeast MAPKs based on the human proteome dissimilarity (a), based on the pathogen proteome similarity (b), based on the average interaction score (AVPIP) (c).

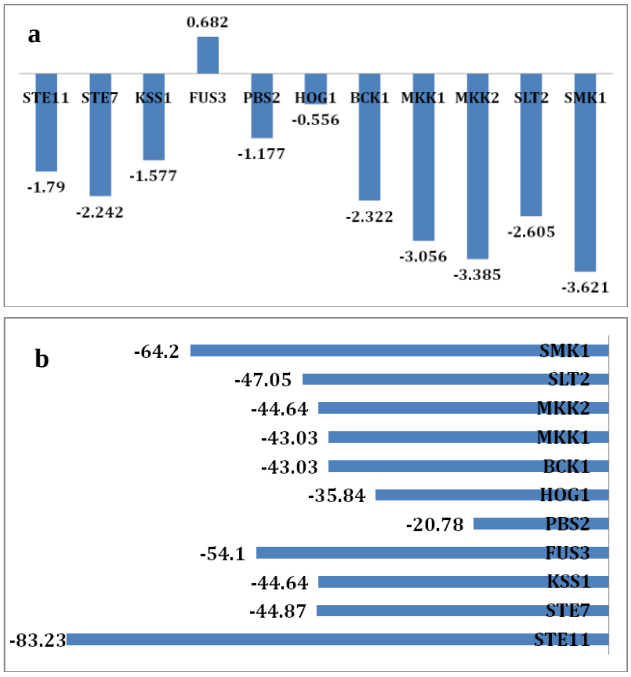

Fig. 3. Quality comparison yeast MAPKs based on the homology models (a), based on the ligand binding score (LB) (b).

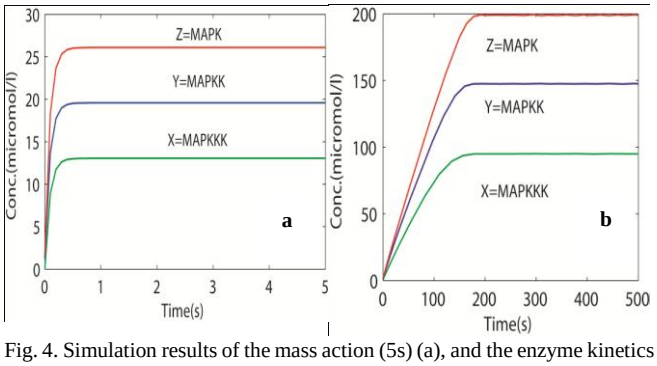

Fig. 4. Simulation results of the mass action (5s) (a), and the enzyme kinetics

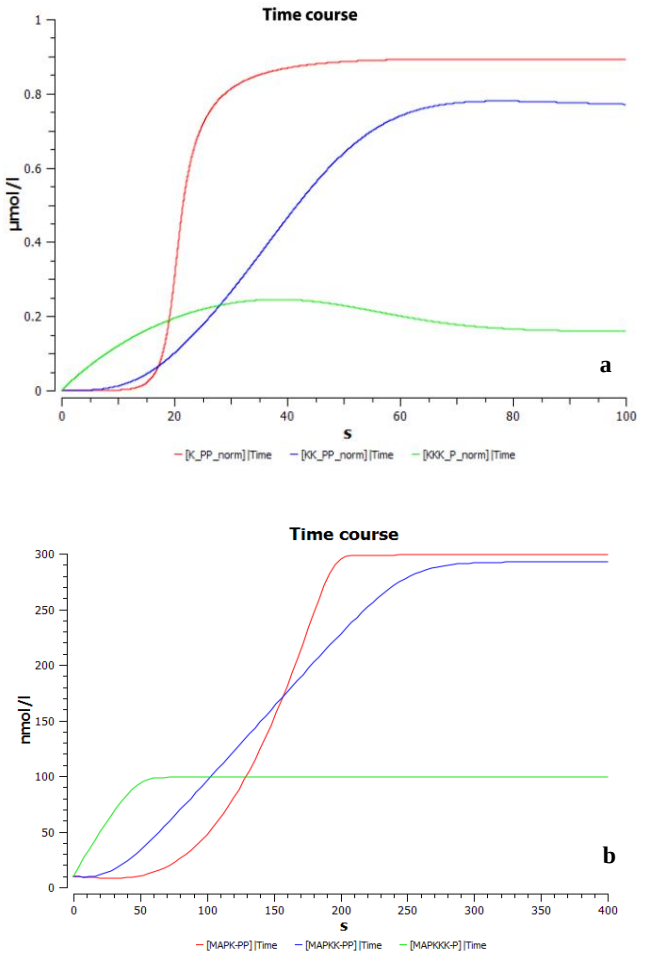

Fig. 5. Time course simulation curves for Huang and Ferrell Model (a) and Khodolenko model (b).

So, it can be inferred that MAPK stage is a comparatively important signaling point in the MAPK cascade.

TABLE I: A WEIGHTED DECISION MATRIX (11 BY 6 MATRIX) TABLE FOR ALL THE RELATIVE PARAMETERS (MAPK MOLECULES ARE ASSUMED TO BE ACCORDING TO THE COLUMN AND RELATIVE PARAMETER VALUES FOR EACH MAPK FROM THE STUDIES).

15	18	4	5	1	9
12	33	18	6	5	6
27	15	10	4	6	3
30	12	6	1	3	3
18	9	12	3	9	6
33	3	8	2	8	3
9	30	14	7	7	9
6	27	16	9	7	6
3	24	22	10	6	6
24	6	2	8	4	3
21	21	20	11	2	3

As the lower total score (TS) which is the summation of Combined Matrix Position Score (CMPS) and Standard Deviation (SD) of matrix position scores, denotes to lower host similarity, higher pathogen similarity, higher protein-protein interaction, higher homology model quality, higher best ligand availability and higher signal sensitivity, MAPKs having lower TS can be predicted as the better drug targets.

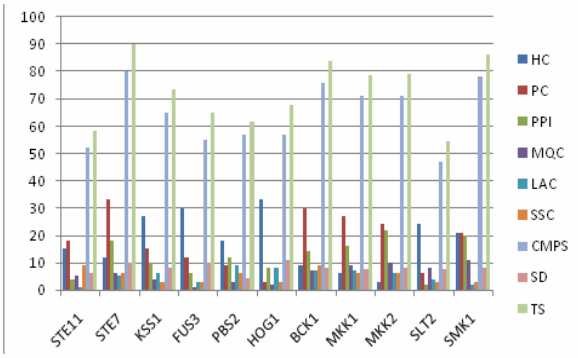

Fig. 6. Combined comparison of MAPKs based on their weighted scores from different studies. (HC= host comparison, PC= pathogen comparison, PPI= protein-protein interaction, MQC= model quality comparison, LAC= ligand availability comparison, SSC= signal sensitivity comparison, CMPS=combined matrix position score, SD=standard deviation, TS=total score).

TABLE II: CMPS, SD AND TS FOR DIFFERENT MAPKS (CMPS= COMBINED MATRIX POSITION SCORE, SD= STANDARD DEVIATION, TS= TOTAL SCORE).

MAPKs	CMPS	SD	TS
STE11	52	6.07	58.07
STE7	80	9.89	89.89
KSS1	65	8.27	73.27
FUS3	55	9.96	64.96
PBS2	57	4.5	61.5
HOG1	57	10.78	67.78
BCK1	76	8.1	84.1
MKK1	71	7.6	78.6
MKK2	71	8.17	79.17
SLT2	47	7.49	54.49
SMK1	78	8.18	86.18

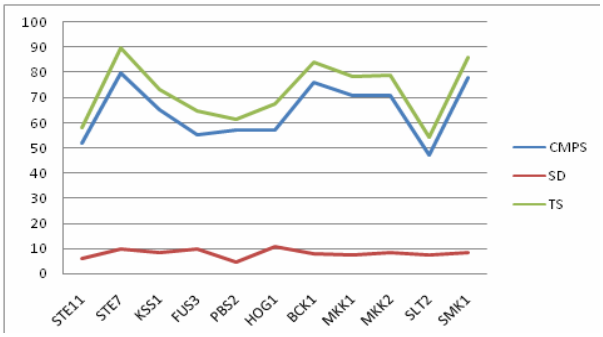

Fig. 7. Comparison of MAPKs based on CMPS, SD and TS.

MAPKs having TS less than the average TS (72.55) can be predicted as better MAPK targets. Thus SLT2 (54.49), STE11 (58.07), PBS2 (61.5), FUS3 (64.96) and HOG1

(67.78) have better potential as antifungal drug targets. Here SLT2/MPK1 shows least score (54.49) and can be predicted as the best target among the 11 MAPK molecules (Table I and Table II, Fig. 6 and 7]

IV. Conclusion and Recommendation

The study was conducted for the preliminary evaluation of the MAPK molecules as drug targets against fungal pathogens. The result comprises the understanding of a number of parameters in an integrated fashion. The predicted targets can be further analyzed and experimented for clear confirmation about their possibility as drug targets. Further analysis can be done based on more accurate structural analysis of the proteins to search out more accurate drugs. It can also be suggested that the MAPK molecules with lower TS values can also be used as targets for synergistic antifungal drugs in addition with conventional ones to avoid the evolution of resistance. Here *Candida albicans* was taken as target genome. Other fungal pathogens also can be assessed in this process to find out better drug target against them. Such *in silico* study will definitely help us to discover drugs with reduced cost, time and enhanced efficiency and accuracy.

Acknowledgment

The authors like to thank Md. Al Mamun, Md. Shaifur Rahman, Mahbub-E-Sobhani, Mehedi Rahman, Rajat Suvra Banik, Abdullah Mahmud-Al-Rafat, Tauhid Siddiki Tomal and Ahmad Ullah for their kind co-operation during the work.

References

[1] H. Kitano, "Systems biology: A brief overview," *Science*, vol. 295, no. 5560, pp. 1662-1664, 2002.
[2] H. Kitano, "Computational systems biology," *Nature*, vol. 420, no. 6912, pp. 206-10, 2002.
[3] E. C. Butcher, E. L. Berg, and E. J. Kunkel, "Systems biology in drug discovery," *Nature Biotechnology*, vol. 22, no. 10, pp. 1253-9, 2004.
[4] V. Sharma, P. Gupta, and A. Dixit, "In silico identification of putative drug targets from different metabolic pathways of Aeromonas hydrophila," *In Silico Biology*, vol. 8, pp. 0026, 2008.
[5] K. R. Sakharkar, M. K. Sakharkar, and V. T. Chow, "A novel genomics approach for the identification of drug targets in pathogens, with special reference to Pseudomonas aeruginosa," *In Silico Biology*, vol. 4, pp. 355-360, 2004.
[6] D. Perumal, C. S. Lim, K. R. Sakharkar, and M. K. Sakharkar, "Differential genome analyses of metabolic enzymes in Pseudomonas aeruginosa for drug target identification," *In Silico Biology*, vol. 7, pp. 453-465, 2007.
[7] A. Dutta, S. K. Singh, P. Ghosh, R. Mukherjee, S. Mitter, and D. Bandyopadhyay, "In silico identification of potential therapeutic targets in the human pathogen Helicobacter pylori," *In Silico Biology*, vol. 6, pp. 0005, 2006.
[8] K. Shobana and B. Thiagarajan, "In silico identification of putative drug targets in Neisseria meningitidis through metabolic pathway analysis," *Scholar's Research Journal*, vol. 1, no. 1, pp.13-24, 2011.
[9] Y. Zhang, "Progress and challenges in protein structure prediction," *Current Opinion in Structural Biology*, vol. 18, no. 3, pp. 342–348, 2008.
[10] D. B. Kitchen, H. Decornez, J. R. Furr, and J. Bajorath, "Docking and scoring in virtual screening for drug discovery: methods and applications," *Nature reviews. Drug Discovery*, vol. 3, no. 11, pp. 935–49, 2004.
[11] V. Mandlik, S. Shinde, A. Chaudhary, and S. Singh, "Biological network modeling identifies IPCS in Leishmania as a therapeutic target," *Integrative Biology*, vol. 4, pp. 1130-1142, 2012.
[12] K. Gulshan and W. S. Moye-Rowley, "Multidrug Resistance in Fungi," *Eukaryotic Cell*, vol. 6, pp. 1933-1942, 2007.
[13] S. P. Palecek, A. S. Parikh, and S.J. Kron, "Sensing, signalling and integrating physical processes during Saccharomyces cerevisiae invasive and filamentous growth," *Microbiology*, vol. 148, no. 4, pp. 893-907, 2002.
[14] C. Widmann, S. Gibson, M. B. Jarpe, and G. L. Johnson, "Mitogen-activated protein kinase: conservation of a three-kinase module from yeast to human," *Physiol Reviews*, vol. 79, no. 1, pp. 143-80, 1999.
[15] M. R. Junttila, S. P. Li, and J. Westermarck, "Phosphatase-mediated crosstalk between MAPK signaling pathways in the regulation of cell survival," *FASEB J.*, vol. 22, no. 4, pp. 954–65, 2008.
[16] K. J. Ryan and C. G. Ray (editors), *Sherris Medical Microbiology*, 4th ed.: McGraw Hill, 2004.
[17] C. dEnfert and B. Hube (editors), *Candida: Comparative and Functional Genomics*, 1st ed: Caister Academic Press, 2007.
[18] B. R. Braun, "A Human-Curated Annotation of the Candida albicans Genome, "*PLoS genetics*, vol. 1, no. 1, pp. 36–57, 2005.
[19] M. Kanehisa and S. Goto, "KEGG: Kyoto Encyclopedia of Genes and Genomes," *Nucleic Acids Res.*, vol. 28, pp. 27-30, 2000.
[20] *Protein Lounge - Databases* for Biology, Pathway *Database*, and Protein Database, http://www.proteinlounge.com/Default.aspx
[21] *National Center for Biotechnology Information*, http://www.ncbi.nlm.nih.gov/
[22] The UniProt Consortium, "Reorganizing the protein space at the Universal Protein Resource (UniProt)," *Nucleic Acids Research*, vol. 40, no. D1, pp. D71-D7, 2011.
[23] S. F. Altschul, "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs," *Nucleic Acids Res.*, vol. 25, pp. 3389-3402, 1997.
[24] D. O. Inglis et al., "The Candida genome database incorporates multiple Candida species: multispecies search and analysis tools with curated gene and protein information for Candida albicans and Candida glabrata," *Nucleic Acids Res.*, vol. 40, no. 1, pp. D667-74, 2012.
[25] D. Szklarczyk et al., "The STRING database in 2011: functional interaction networks of proteins globally integrated and scored," *Nucleic Acids Res.*, vol. 39, pp. D561-8, 2011.
[26] K. Arnold et al., "The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling," *Bioinformatics*, vol. 22, pp. 195-201, 2006.
[27] Discovery studio visualizer 2.5. San Diego: Accelrys Software Inc., 2007.
[28] Irwin et al., "Automated docking screens: a feasibility study," *J. Med. Chem.* vol. 52, no. 18, pp. 5712-5720 , 2009.
[29] E. F. Pettersen et al., "UCSF Chimera--a visualization system for exploratory research and analysis." *J Comput Chem.*, vol. 25, no. 13, pp. 1605-12, 2004.
[30] C. Y. Huang and J. E. Ferrell Jr., "Ultrasensitivity in the mitogen-activated protein kinase cascade," P*roc Natl Acad Sci U S A*, vol. 93, no. 19, pp. 10078-83, 1996.
[31] B. N. Kholodenko, "Negative feedback and ultrasensitivity can bring about oscillations in the mitogen-activated protein kinase cascades," *Eur J Biochem*, vol. 267, no. 6, pp. 1583-8, 2000.
[32] C. Li et al., "BioModels Database: An enhanced, curated and annotated resource for published quantitative kinetic models. BMC Systems Biology," vol. 4, pp. 92, 2006.
[33] C. J. Thalhauser and N. L. Komarova, "Signal Response Sensitivity in the Yeast Mitogen-Activated Protein Kinase Cascade," *PLoS ONE*, vol. 5, no. 7, pp. e11568, 2010.
[34] S. Hoopset al., "COPASI- a COmplex PAthway SImulator," *Bioinformatics*, vol. 22, pp. 3067-74, 2006.
[35] MATLAB version 7.10.0. Natick, Massachusetts: The MathWorks Inc., 2010.
[36] Bronson and Richard, *Schaum's outline of theory and problems of matrix operations*, 2nd ed. New York: McGraw–Hill, 1989.
[37] Ghahramani, Saeed, *Fundamentals of Probability*, 2nd ed. Prentice Hall, pp. 438, New Jersey: 2000.

Sheikh Md. Enayetul Babar is a professor of Biotechnology and Genetic Engineering (BGE) Discipline of Khulna University, Bangladesh. He had obtained PhD from Korea Institute of Science and Technology (KIST). He has been working in this University since 1996. So far the author has published several articles in peer-reviewed journals and presented in many national and international conferences. Currently he is the head of the BGE Discipline of Khulna University, Bangladesh.