

Analysis of the p42.3 protein structure and regulatory network

ZHANG JianHua^{1†}, MA Wang^{2†}, SHANG ZhiGang¹, XING Rui³, SHI Li^{1*} & LU YouYong^{3*}

¹ Department of Biomedical Engineering, School of Electrical Engineering, Zhengzhou University, Zhengzhou 450001, China;

² Oncology Department, The First Affiliated Hospital of Zhengzhou University, Zhengzhou 450052, China;

³ Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Laboratory of Molecular Oncology, Peking University Cancer Hospital & Institute, Beijing 100142, China

Received October 29, 2012; accepted December 19, 2012

p42.3 is a recently discovered gene that may participate in the regulation of gastric cancer cell generation and development. In this research, we analyzed the predicted *p42.3* protein structure using bioinformatics tools, established the regulatory network of the protein molecule and found the optimal pathway using a Bayesian network model.

p42.3, structure analysis, regulatory network, Bayesian network model

Citation: Zhang J H, Ma W, Shang Z G, et al. Analysis of the *p42.3* protein structure and regulatory network. *Chin Sci Bull*, 2013, 58: 869–872, doi: 10.1007/s11434-013-5691-8

p42.3 is a new gene that was cloned by synchronization, and identified by mRNA differential display and bioinformatics [1]. The 3877-bp long *p42.3* cDNA encodes a 389-amino acid protein with a molecular weight of 42.3 kD. *p42.3* is related to cell cycle regulation and has the characteristics of a tumor biomarker. The overexpression of the *p42.3* gene may be one of the early molecular events indicating gastric mucosal lesions that may advance to gastric cancer.

Network models, such as the nonlinear network model, graph theory model, Bayesian network model and Boolean network model, have been applied successfully to the modeling and simulation of gene regulatory networks [2].

Based on probabilistic reasoning, the Bayesian network, also called the belief network, is a mathematic model that has been proposed to solve the problems of uncertainty and incompleteness. Using the relationship between the spatial conformation of the *p42.3* and its function as the breakthrough point, we built a regulatory network by database searching and spatial modeling, and predicted the optimal pathway using a Bayesian network model. Our results provide a theoretical and experimental basis for the develop-

ment of gastric cancer molecular typing and individualized treatment with *p42.3* as the target molecule, and will help advance the research and treatment of gastric cancer and other tumors.

We used Phyre2 (<http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index>), a threading method forecasting tool, to predict the tertiary structure and to analyze the secondary structure of the *p42.3* protein (Figure 1). The spatial conformation of the *p45.3* protein was displayed in the Swiss-PdbViewer version 3.7 (<http://spdbv.vital-it.ch/>) as shown in Figure 2.

The predicted *p42.3* protein structure was found to be similar to the protein structure data sets containing the EF-hand or CC-domain functional domains when searched and sorted out using the similarity algorithm based on spherical coordinates space delamination in Phyre2. For the EF-hand structure data set, we found that parameter equal to 4 was most suitable, while parameter equal to 3 was most suitable for the CC-domain structure data set. The S100 family of calcium binding proteins (including S100A1, S100A11, S100A2 and S100A4), the small G protein, CIB (calmodulin-binding protein), ROCK1 (serine kinase), CENP-E (kinesin) and GCN4 protein were selected as the reference proteins for analyzing *p42.3* protein function.

†These authors contributed equally to this work.

*Corresponding authors (email: youyonglu@bjum.edu.cn; shili@zzu.edu.cn)



Figure 1 Secondary structure of p42.3.

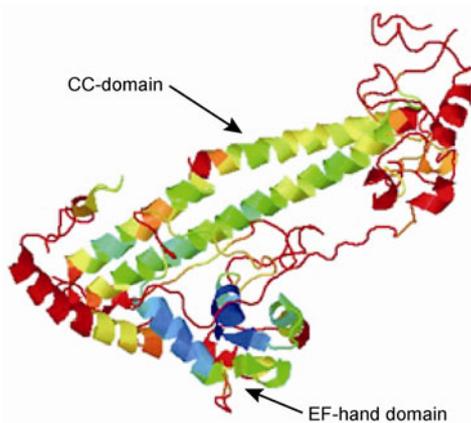


Figure 2 Spatial conformation of p42.3 protein.

The possible regulatory network of p42.3 protein was first determined through literature review and various materials. According to Bayes' theorem, when the possibility of a subevent is known the possibility of the parent event can be calculated; that is, the possibility that caused the occurrence of the subevent. In the regulatory network of p42.3, this method was adopted to look for the most possible event leading to the final subevent, namely, malignant cell proliferation. This prediction will be the optimal regulatory pathway of the mechanism of action of p42.3. After analysis, the optimal pathway (marked red in Figure 3) was selected. This pathway, which leads to cancer, involves the Ras protein, Raf-1, MEK, MAPK kinase, MAPK, tubulin, spindle protein, and centromere protein in that order (Figure 3), and

is the most possible action pathway of p42.3.

The involvement of the *p42.3* gene in gastric cancer development has been verified experimentally; however, its specific mechanism of action at the molecular level is unclear. The presence of an EF-hand structural domain that is similar to the domain in the tumor-associated calcium binding protein S100 family was predicted at the N-terminal end of the p42.3 protein. Furthermore, at the C-terminal end of p42.3, a CC domain was predicted. This domain may participate in protein-protein interactions and may be regulated by phosphorylation, which might influence the active site of the tumor suppressor APC protein and further influence the related cell signal pathway and its biological function [3]. Gene regulatory networks can also be established by predicting the transcription factor binding sites [4]. Tamada et al. [5] established the regulatory networks of yeast genes using a Bayesian network algorithm, and the structure analysis indicated that yeast cell function was closely related to the highly connected regulatory genes in the networks. In this study, we used a similarity algorithm based on spherical coordinates space delamination and performed programming realization using MATLAB software [6]. From every data set, we successfully screened several proteins with high similarity to the p42.3 protein that had functions that were associated with tumor development. These proteins were

regarded as the reference proteins [3,7–10]. An action regulatory network was built and the optimal regulatory pathway was found using a Bayesian network model. Our findings provide a novel insight into the structure and function of the tumor-associated *p42.3* gene.

This work was supported by the National Natural Science Foundation of China (60971110) and the Cooperation Project for Academicians from Henan Province and Other Provinces (122106000042).

- 1 Xu X, Li W, Fan X, et al. *Oncogene*, 2007, 26: 7371–7379
- 2 Chen T. Study on the construction algorithm of gene regulatory network. Dissertation for Master's Degree. Shanghai: Fudan University, 2009. 5
- 3 Lu Y X, Xue H P, Lu Y H, et al. *J Harbin Med Univ*, 2007, 41: 160–162
- 4 Huang J F, Yang J J, Wang G, et al. *Chin Sci Bull*, 2008, 53: 2054–2059
- 5 Tamada Y, Kim S, Bannai H, et al. *Bioinformatics*, 2003, 19: 227–236
- 6 Hu X D, Dong C H. *Master MATLAB from Layman to Professional Staff*. Beijing: Posts and Telecommunications Press, 2010. 1–5
- 7 Zhang Q. *Beijing Jiaotong Univ*, 2009, 6: 23–25
- 8 Uchida S, Yoshioka K, Kizu R, et al. *Cancer Res*, 2009, 69: 6438–6444
- 9 Tian Z D. *Inform Sci*, 2006, 7: 1049–1052
- 10 Pecevski D, Buesing L, Maass W. *PLoS Comput Biol*, 2011, 12, 7: e1002294

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.