ELSEVIER

Biophysical Chemistry 113 (2005) 275–288

Biophysical Chemistry

www.elsevier.com/locate/bpc

Automated oncogene detection in complex protein networks with applications to the MAPK signal transduction pathway

Dhruv Pant, Avijit Ghosh*

Drexel University, 3141 Chestnut st, Philadelphia, PA, 19104, United States

Received 3 August 2004; accepted 10 September 2004
Available online 18 October 2004

Abstract

Activation of the extracellular signal-regulated kinases (ERK1/2; p42/p44 mitogen-activated protein kinase (MAPK)) is one of the most extensively studied signaling pathways not least because it occurs downstream of oncogenic RAS. Here, we take advantage of the wealth of experimental data available on the canonical RAS/RAF/MEK/ERK pathway of Bhalla et al. to test the utility of a newly developed nonlinear analysis algorithm designed to predict likelihood of cellular transformation. By using ERK phosphorylation as an "output signal", the method analyzes experimentally determined kinetic data and predicts putative oncogenes and tumor suppressor gene products impacting the RAS/MAPK module using a purely theoretical approach. This analysis identified several modifiers of ERK/MAPK activation described previously. In addition, several novel enzymes are identified which are not previously described to affect ERK/MAPK phosphorylation. Importantly, the nonlinear analysis enables a ranking of modifiers of MAPK activation predicting their relative importance in RAS-dependent oncogenesis. The results are compared with a linearized analysis based on sensitivity analysis about the steady state or metabolic control analysis (MCA). The results are favorable, pointing to the utility of first-order sensitivity analysis and MCA in the analysis of complex signaling networks for oncogenes.

Keywords: Oncogene detection; MAPK signaling; Nonlinear analysis

1. Introduction

Intracellular signal transduction has been of interest to cancer researchers for many years based on the consideration that many oncogenes and suppressor genes affect signaling cascades. Only recently, however, it has been appreciated that signal transduction forms a complex network rather than being composed of linear and separate pathways. The net response of network activation governs normal cell growth and function by control of gene expression, a process that can lead to cell proliferation, arrest or death [1,2]. Mutant proteins situated at critical nodal points can be expected to affect multiple pathways and are likely to contribute to neoplastic transformation. Mutational events may affect protein function through modulation of enzymatic activity [3–5] or by affecting protein stability and, thus, steady-state expression levels [6,7].

The nonlinear nature of signaling cascades necessitates novel tools to dissect their complexity. Here, we describe a computational method that enables the analysis of complex networks with the intent of identifying enzymes of relevance to the transformed state. As a first step in testing the robustness of this approach, we made use of a large body of data on the canonical RAS/RAF/MEK/ERK pathway accrued experimentally over many years. We demonstrate that the model predicts several candidates for mutational events likely to contribute to transformation. In addition, several nonobvious candidates are identified that warrant future experimental exploration.

The theoretical model chosen upon which to base this study is one of several highly regarded computational

* Corresponding author. Tel.: +1 215 8952726; fax: +1 215 8955934.
E-mail address: avijit@physics.drexel.edu (A. Ghosh).

0301-4622/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.bpc.2004.09.004

studies of this central cascade. Recent studies of the mitogen-activated protein kinase (MAPK) signal transduction pathway have been developed by Schoeberl et al. [8], Kholodenko [9], Levchenko et al. [10] as well as others. The canonical MAPK pathway includes several major networks including PLC, PLC-γ, PLA2, PKC, Ras and calcium sequestering channels developed by Bhalla and Iyengar [11]. This network has several positive and negative feedback loops, multiple steady states and exhibits highly nonlinear behavior even within this subnetwork within the cell [11]. The interplay between the positive and negative feedback loops, attenuation and stimulatory factors all play a crucial role in the transformation of this central cascade.

2. Computational methods

The basis for discovering how certain mutations will affect the signaling behavior using a computational systems biology approach necessitates in first having an experimentally validated numerical model of the functioning signal transduction pathway. The parameters for the model of the functioning signal transduction network are taken from the pioneering work of Bhalla and Iyengar [11]. While the validation of certain features such as the proposed feedback loop remain to be tested, as the overall features of Bhalla and Iyengar's [11] kinetic model reproduce experimentally determinable behavior, it provides a reasonable test suite upon which to test the computational approach, with the acknowledgment that any refinements or corrections to the model will require a necessary correction in the analysis. Currently, the analysis does not include the negative feedback loop through MKP recently described by Bhalla et al. [12]. As this loop is not coupled to any members of the signaling pathway outside of ERK and MKP, this feedback loop is expected to have the same effect on mutations of upstream members. For this reason, the overall ranking (except for mutations directly in MKP and ERK) of enzymes provided by this analysis is expected to be the same, with an understanding that there would be a common attenuation effect on ERK of all mutations from this proposed negative feedback loop.

The goal here is to use energetic considerations of mutations to perform a nonlinear analysis of perturbations in the signaling process. The model derived by Bhalla et al. has been tested extensively against experimental data and is available publicly at the *cellml* website [13]. We emphasize that, while our own implementation is different from the implementation using the *Genesis* code base developed by Bhalla et al. [14], both the parameters and coupled set of elementary reactions are the same. Like *Genesis*, the numerical implementation and scripts necessary to reproduce the results of the mutational studies are available freely under the Gnu Public License with source code and binaries [15].

The complete cellular model is broken down into three localized compartments: the extra-cellular matrix, the cytosol and the calcium sequestering endosomal regions. Elementary chemical reactions describe the enzymatic and non-enzymatic reactions within each compartment. These reactions may be written as:

$$\sum_i n_i R_i \underset{k_b}{\overset{k_f}{\rightleftharpoons}} \sum_j n_j P_j \tag{1}$$

where a set of reactant species R_i with stoichiometric coefficients n_i interconverts into a set of product species P_j, with stoichiometric coefficients n_j with rate constants k_f and k_b. As collisions that are greater than bimolecular are rare, typically the order of an elementary chemical reaction is not greater than 2. Characteristics of signaling pathways are enzymatic reactions such as phosphorylation or dephosphorylation events. These reactions may be expressed as a combination of a reversible and a irreversible chemical reaction as follows:

$$E + S \underset{k_1}{\overset{k_2}{\rightleftharpoons}} E.S \xrightarrow{k_3} E + S^* \tag{2}$$

E represents an enzyme which catalyzes the substrate S. The intermediate species $E.S$ first forms reversibly with rate constants k_1 and k_2 followed by an irreversible catalytic step with rate constant k_3 which releases the activated substrate S^* and the enzyme for further catalysis.

One may express the time rate of change in concentration of *all* species in the complex, nonlinear cascade of reaction pathways as a coupled set of ordinary differential equations (ODEs), Eq. (3). The time rate of change of any species A is given as the sum of the time rate of change of that species involved in all of its elementary reactions of type Eq. (1). The change in concentration of any species C_i in a system of unimolecular and bimolecular reactions may then be written out as

$$\frac{d[C_i]}{dt} = \sum_j k_{ij}[C_j] + \sum_{l>m} k_{ilm}[C_l][C_m] + \sum_j T(C_i, C_j) \tag{3}$$

where k_{ij} is the rate constant for a unimolecular reaction involving species C_i and C_j at concentrations $[C_i]$ and $[C_j]$, respectively. If k_{ilm} is positive then k_{ilm} represents the rate constant of formation of species C_i from a bimolecular reaction between species C_l and C_m with concentrations $[C_l]$ and $[C_m]$. Conversely, if k_{ilm} is negative, then k_{ilm} represents the rate constant of disassociation of species C_i into two species C_l and C_m. $T(C_i,C_j)$ represents a function governing the passive transport of a species C_i into a