SAR and QSAR in Environmental Research
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t716100694

In silico method for identification of promising anticancer drug targets

O. N. Koborova [a]; D. A. Filimonov [a]; A. V. Zakharov [a]; A. A. Lagunin [a]; S. M. Ivanov [a]; A. Kel [b]; V. V. Poroikov [a]
[a] Institute of Biomedical Chemistry of Russian Academy of Medical Sciences, Moscow, Russia [b] BIOBASE GmbH, Wolfenbüttel, Germany

Online publication date: 17 December 2009

To cite this Article Koborova, O. N., Filimonov, D. A., Zakharov, A. V., Lagunin, A. A., Ivanov, S. M., Kel, A. and Poroikov, V. V.(2009) '*In silico* method for identification of promising anticancer drug targets', SAR and QSAR in Environmental Research, 20: 7, 755 — 766
To link to this Article: DOI: 10.1080/10629360903438628
URL: http://dx.doi.org/10.1080/10629360903438628

SAR and QSAR in Environmental Research
Vol. 20, Nos. 7–8, October–December 2009, 755–766

Taylor & Francis
Taylor & Francis Group

In silico method for identification of promising anticancer drug targets†

O.N. Koborova[a]*, D.A. Filimonov[a], A.V. Zakharov[a], A.A. Lagunin[a], S.M. Ivanov[a], A. Kel[b] and V.V. Poroikov[a]

[a]*Institute of Biomedical Chemistry of Russian Academy of Medical Sciences, Moscow, Russia;* [b]*BIOBASE GmbH, Wolfenbüttel, Germany*

(*Received 7 July 2009; in final form 1 October 2009*)

In recent years, the accumulation of the genomics, proteomics, transcriptomics data for topological and functional organization of regulatory networks in a cell has provided the possibility of identifying the potential targets involved in pathological processes and of selecting the most promising targets for future drug development. We propose an approach for anticancer drug target identification, which, using microarray data, allows discrete modelling of regulatory network behaviour. The effect of drugs inhibiting a particular protein or a combination of proteins in a regulatory network is analysed by simulation of a blockade of single nodes or their combinations. The method was applied to the four groups of breast cancer, HER2/neu-positive breast carcinomas, ductal carcinoma, invasive ductal carcinoma and/or a nodal metastasis, and to generalized breast cancer. As a result, some promising specific molecular targets and their combinations were identified. Inhibitors of some identified targets are known as potential drugs for therapy of malignant diseases; for some other targets we identified hits in the commercially available sample databases.

Keywords: regulatory networks; breast cancer; drug target identification

1. Introduction

The worldwide prevalence of cancerous diseases and the rather low effectiveness, toxicity and resistance of the current therapy require identification of new promising anti-tumour targets [1]. In recent years, the accumulation of genomic, transcriptomic, proteomic data for structural and functional organization of regulatory processes in a cell has provided the possibility of identifying the peculiarities related to a particular disease and of selecting the most promising targets for future drug development [2,3].

The regulation of cell processes is often considered as a regulatory network (RN) – a theoretical abstraction of biochemical reactions, intermolecular interactions and transport of molecules inside the cell through the membrane. RNs sufficiently well reflect the known cause-and-effect interactions between genes, proteins and small molecules. They consist of nodes connected by directed edges. Each node corresponds to a particular gene, mRNA, protein, protein–protein or protein–ligand complex, or small molecule.

*Corresponding author. Email: okoborova@gmail.com
†Presented at CMTPI 2009: Computational Methods in Toxicology and Pharmacology Integrating Internet Resources (Istanbul, Turkey, 4–8 July 2009).

ISSN 1062–936X print/ISSN 1029–046X online

DOI: 10.1080/10629360903438628
http://www.informaworld.com

Each edge corresponds to the interaction between cell components, so that the upstream nodes affect the downstream nodes. Different approaches exist for the analysis and modelling of the network's behaviour. Structural analysis methods give the possibility of working with large networks, and investigating properties of interesting structures such as hubs and bottlenecks, feedback loops, etc., but they do not allow the modelling of the processes in dynamics [4–7]. One of the most widely used dynamic methods is the method of differential equation modelling, which is applicable only to the small parts of a RN (pathways), as the experimental data for large RNs, e.g. kinetic parameters, reaction rates, etc., is limited [8,9]. It has been shown in several studies that some constructed RNs are very stable, and time is not a critical factor for the modelling of regulatory processes [10]; therefore, discrete modelling methods can be applied to large RNs reflecting such processes as the cell cycle, where the major information is concentrated in a sequence of cell states and the functioning of a particular gene(s)/protein(s) defines the selection of further cell behaviour [11]. Therefore, using a large mass of accumulated data on the interaction between proteins, the method of discrete modelling may sufficiently well describe the cell processes as opposed to the methods of differential equations, cellular automata, etc, that cannot provide the analysis of large RNs.

In this paper we present an original algorithm of RN simulation based on a discrete dichotomy model, which may be used for identification of promising targets, and its application to the four groups of breast cancer: HER2/neu-positive breast carcinomas, ductal carcinoma, invasive ductal carcinoma and/or a nodal metastasis, and to generalized breast cancer.

2. Materials and methods

2.1 *Algorithm*

The aim of this work is to develop an approach for identification of promising drug targets. For this purpose RNs may be considered as artificial neural networks (ANNs). The usual ANN approach implies transformation of input data to the output results, so that the structure and inner behaviour of the ANN is not significant. To distinguish the difference between the usual ANN approach and our approach, we call the considered RN a dichotomic RN. Since the goal of our study is the identification of promising drug targets, we consider the states of particular nodes and the edges of the RN at different steps of the simulation; therefore, information about the behaviour of the neural network at the distinct initial states is of high priority.

Each node of the dichotomic RN may be in two states, active and inactive, and each edge may be of two types: activation and inactivation.

Let S_i be the state of the node i; then for the active node $S_i > 0$ and for the inactive $S_i = 0$ (Figure 1). Let the properties of the edges correspond to the matrix $\mathbf{B}$, where the array cell $b_{ik} = 1$ if the edge is an output of node i and if it activates node k, $b_{ik} = -1$ if the edge is an output of node i and if it inactivates node k, $b_{ik} = 0$ in all other cases and $b_{ii} = 0$ means that the node cannot activate itself. The direction of edges means that only one of two values, b_{ik} and b_{ki}, is not zero.

We consider the consecutive states $\mathbf{S}(0)$, $\mathbf{S}(1)$, $\mathbf{S}(2)$, ... as a *trajectory* and the subset of a network node's states $\{\mathbf{S}\} \in \mathbf{2}^n$ as an outcome (result).

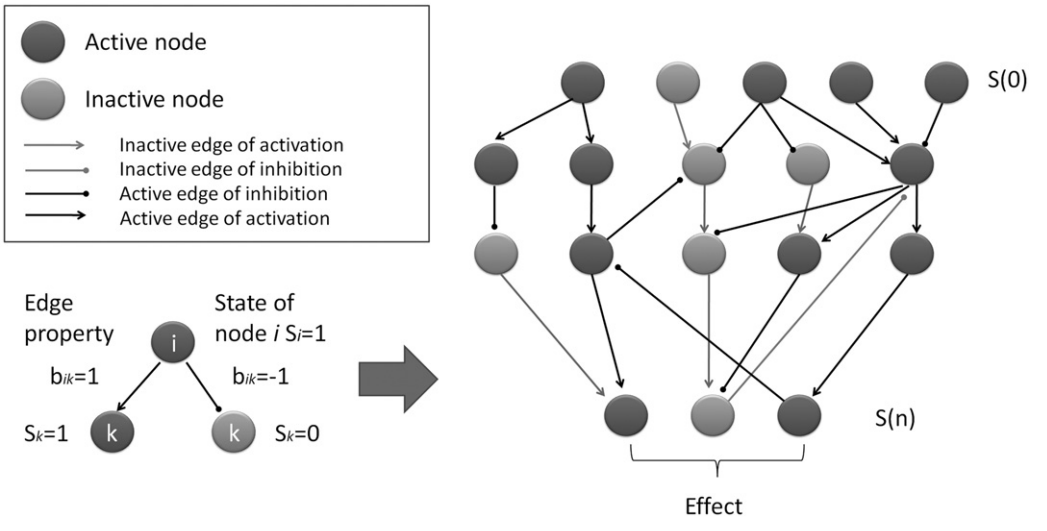


Figure 1. Scheme of regulatory network modelling.

Using such an approach, the modelling (simulation) of the regulation processes lies in the computation of the trajectories $\mathbf{S}(0), \mathbf{S}(1), \mathbf{S}(2), \ldots$ by the set of dichotomy functions:

$$S_i(t+1) = F_i(S_1(t), S_2(t), \ldots, S_n(t)), \quad i = 1, 2, \ldots, n, \tag{1}$$

where $S(0)$ is the given initial state of the RN, $S_{i(t)}$ is the state of node i in the network of n nodes during the step t; F_i is the function of transition determined by the network structure and by the rules of combination of node states, where up-stream nodes are connected with node i by their down-stream edges. If the number of working activating nodes connected with the node i exceeds that of the working inactivating nodes, then the node i is active, and *vice versa*.

We assume the *time* to be a discrete variable and at each particular moment of time (step) the states of the nodes depend on the nodes of the network in the previous step of the simulation, so that $\mathbf{S}(t)$ are the states of the network nodes at step t.

The function of the transition of node i is as follows,

$$F_i(S_1, S_2, \ldots, S_n) = \Theta\Big(a_i + \sum\nolimits_k S_k b_{ki}\Big), \tag{2}$$

where the threshold function $\Theta(z) = 0$ if $z \leq 0$ and $\Theta(z) = 1$ if $z > 0$; a_i is used for more accurate adjustment of the model according to the known data on regulation processes (as described in the Results section). In addition, parameter a_i may reflect the influence of the unknown part of the RN.

For all network nodes of our model, the initial states $\mathbf{S}(0)$ are the input data and the output data are trajectories. The computation of the trajectories ends when the selected result is reached, which remains stable during all time steps. The selected result corresponds to activation/inhibition of a particular RN fragment (a set of corresponding nodes), e.g. on reaching cell cycle arrest and/or apoptosis stimulation of the cancer cells without influence on the normal cells.

A set of nodes that essentially change trajectories for the selected results can be picked out by changing the network node's states; e.g. when a particular protein is inhibited, its state corresponds to $S_i(0) = S_i(1) = S_i(2) = \cdots = S_i(t) = 0$.

We implemented the original algorithm of dichotomy network simulation as a computer program NetFlowEx for identification of promising antineoplastic targets.

2.2 *Data*

The data on protein–protein and protein–gene interactions were collected from the TRANSPATH® database (http://www.biobase-international.com). The database also contains a lot of information about signalling molecules, interactions and pathways [12]. The obtained network consists of 1405 nodes and 2336 edges. It includes different parts of the cell cycle regulation and apoptosis RN, such as the cascade of caspases, the TNFalpha pathway, the IL-6 pathway, the TGFbeta pathway, the HIF-1alpha pathway, the VEGF pathway, the Fas pathway, different cell cycle regulatory chains, the EGF pathway, MAP kinases cascades, the Wnt/beta-catenin pathway and other pathways involved in breast cancer progression and mammary gland development. The nodes reflecting the proteins have only direct reactions that have been experimentally confirmed; indirect reactions between genes and proteins (transrepression, transactivation and expression) were also included in the network. The average number of edges incident on a node in the created network equals 1.66.

The list of human housekeeping genes (467 genes) – the constitutive genes transcribed at a relatively constant level across many or all known conditions – was taken from the Explain database, which provides a number of predefined gene sets, such as Human, Mouse, and Rat promoters and housekeeping genes [13]. The ExPlain® Analysis System promotes biological interpretation of high-throughput experiments such as DNA microarrays, proteomics or proteome data, and ChIP-on-chip experiments (Chromatin-Immunoprecipitation on chip). Some of the genes found have no down-stream edges or encode polypeptides involved in the metabolic pathways, have no edges associated with cell cycle regulation or belong to the class of structure proteins. Some genes are out of our selected network. Finally, 48 housekeeping genes were selected for further simulation of normal cell processes (Table 1).

Table 1. Selection of up- and down-regulated genes for further simulation.

	Property of genes	*Results of meta-analysis*	*Available in Explain database*	*Selected for simulation*
Normal process	Housekeeping	467	467	48
HER2/neu-positive	Down-regulated	125	56	9
breast carcinomas	Up-regulated	16	10	2
Ductal carcinoma	Down-regulated	100	58	9
	Up-regulated	96	57	5
Invasive ductal carcinoma	Down-regulated	191	155	20
and/or a nodal metastasis	Up-regulated	124	83	15
Generalized breast cancer	Down-regulated	786	365	59
	Up-regulated	691	324	49

The microarray data on generalized breast cancer were obtained from the Cyclonet database (http://cyclonet.biouml.org), which also contains information about normal and pathological mammalian cell cycle regulation [14]. The data for breast cancer were analysed by meta-analysis, which is widely used for identification of differentially expressed genes by integrating the results of several independent microarray experiments [15]. The data for generalized breast cancer consisted of statistically significant 786 down-regulated genes and 691 up-regulated genes. We found that only 365 of the 786 down-regulated genes and 324 of the 691 up-regulated genes were presented in the Explain database (Table 1). Eventually, 49 up- and 59 down-regulated genes were presented in our network, due to the availability of the information about their interactions with other proteins of the selected network. The number of up- and down-regulated genes of generalized breast cancer and other groups of breast cancer, and the housekeeping genes used for further simulation, are presented in Table 1. The genes and the corresponding proteins with no interactions to other proteins were not considered. Complete lists of the analyzed genes and proteins, up- and down-regulated genes can be viewed in the supplementary material which is available via the supplementary material tab on the online article webpage.

3. Results and discussion

3.1 *Simulation of normal processes*

We simulated the normal processes in the cell, using the housekeeping genes as input data. Some genes do not have input edges from other nodes, so they cannot maintain their stable states after the primary step of the simulation. For the housekeeping genes IKAPPAB-ALPHA, JUND, GSK3ALPHA, PIN1, 14-3-3BETA, 14-3-3ZETA, CKI-EPSILON, CYCLIN D3, MDIA1, RHOGDI-1, HRS, SIAHBP1, 14-3-3ETA, 14-3-3TAU, CKII-BETA, ADA3, E2F-4, RAD23A, NDKA, MM-1, APC5, RASSF1A, HMGB1, L11, RHOGAP1 and TRIM28, parameter a_i was set equal to 1 to sustain their active state. So as to ensure that all edges between the nodes are involved in the simulation, the maximum number of time steps was higher than the number of nodes (2500 time steps). The modelling results showed that from the 14th to the 19th step and from the 15th to 20th step CDK1 and Cyclin A : CDK1 are activated. However, after the 20th step the states of the cell cycle participants remain inactive and stable, and the cell does not enter into the M phase and does not divide (Figure 2). Caspases were not active and apoptosis was also not reached. Thus, the results of the simulation do not contradict the known processes in normal cells [16].

3.2 *Simulation of pathological processes*

The simulation of pathological processes was carried out individually for the four groups of breast cancer, HER2/neu positive breast carcinomas, ductal carcinoma, invasive ductal carcinoma and/or a nodal metastasis, and for generalized breast cancer, with the number of time steps equal to 2500.

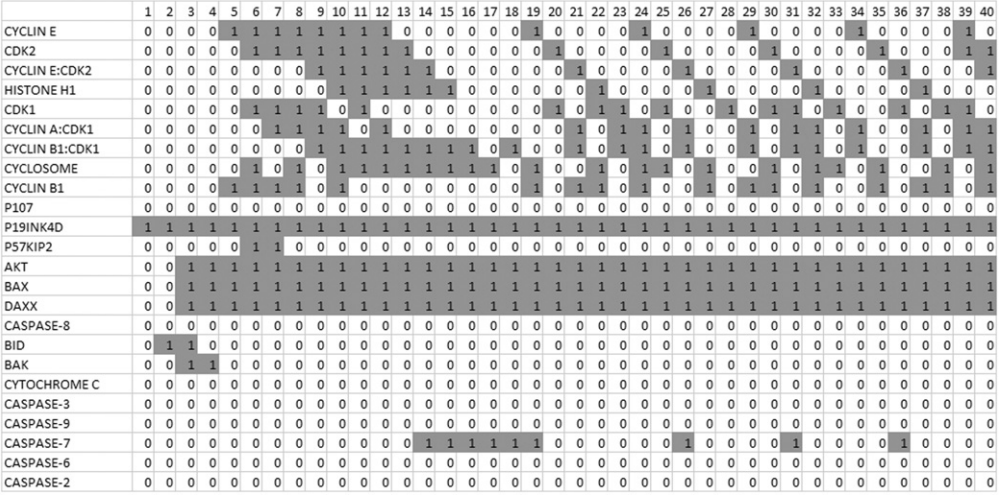
The results of the simulation for generalized breast cancer are presented in Figure 3. Input data for the initial states of the simulation contain five housekeeping genes that are up-regulated, EZRIN, NDKA, PIN1, PRDX1 and RHOGDI-1. The genes PRLR,

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
CYCLIN E	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
CDK2	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
CYCLIN E:CDK2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HISTONE H1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CYCLIN A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CDK1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0
CYCLIN A:CDK2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CYCLIN A:CDK1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
CYCLIN B1:CDK1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CYCLOSOME	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1
CYCLIN B1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
CYCLIN D1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
CDK4	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
CDK6	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
CYCLIN D:CDK4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CYCLIN D:CDK6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P107	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P19INK4D	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P57KIP2	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
AKT	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
BAX	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
DAXX	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
CASPASE-8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
BID	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
BAK	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CYTOCHROME C	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CASPASE-3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CASPASE-9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CASPASE-7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CASPASE-6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CASPASE-2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 2. Results of simulations using housekeeping genes as initial states for several cell cycle and apoptosis proteins. Rows correspond to proteins, columns correspond to time steps; a cell of the table reflects an active (1) or an inactive (0) state of the protein during the simulation time step. Active states are shaded in grey.

BMPR-IB, P19INK4D, MYT1, ErbB2, FIBRONECTIN, RON, RHOGDI-1, FGFR-3, TRAF2, KAP, JAB1, AR (NR3C4), TGFBETA3, IMPORTIN-ALPHA2, NDKA, 14-3-3ZETA, PIN1, CATHEPSIN B, MUC1, AURORA-A, GATA-3, SECURIN, BAF47, PRLR and BMPR-IB were permanently up-regulated as they cannot reproduce themselves in the model, as the information on their feedback regulation is limited.

Although the cell cycle inhibitor p19INK4D is permanently active, the cell cycle complexes are active iteratively. Cyclin E : CDK2 is active from the ninth to the 14th time steps, at the 21st time step and then iteratively active every fifth time step (Figure 3). Cyclin A : CDK1 and Cyclin B1 : CDK1 are active from the 7th to the 14th time step and from the 9th to the 16th step, respectively. Beginning with the 20th and 21st step, the complexes become active every second or first time step. Major apoptotic proteins are inactive during all of the 2500 time steps, in spite of the active BAX responsible for Cytochome C release considered as an indicator of apoptosis. The AKT protein of cell survival remains active from the third time step.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
CYCLIN E	0	0	0	0	1	1	1	1	1	1	1	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0
CDK2	0	0	0	0	0	1	1	1	1	1	1	1	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	1	1
CYCLIN E:CDK2	0	0	0	0	0	0	0	0	1	1	1	1	1	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	1
HISTONE H1	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0
CDK1	0	0	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	1	0	1	1	0	1	0	0	1	0	1	1	0	1	0	0	1	0	1	1	0
CYCLIN A:CDK1	0	0	0	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	1	0	1	1	0	1	0	0	1	0	1	1	0	1	0	0	1	0	1	1	1
CYCLIN B1:CDK1	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	0	1	0	0	1	0	1	1	0	1	0	0	1	0	1	1	0	1	0	0	1	0	1	1
CYCLOSOME	0	0	0	0	0	1	0	1	0	1	1	1	1	1	1	1	1	0	1	0	0	1	0	1	1	0	1	0	0	1	0	1	1	0	1	0	0	1	0	1
CYCLIN B1	0	0	0	0	1	1	1	1	0	1	0	0	0	0	0	0	0	0	1	0	1	1	0	1	0	0	1	0	1	1	0	1	0	0	1	0	1	1	0	1
P107	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P19INK4D	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
P57KIP2	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AKT	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
BAX	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
DAXX	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
CASPASE-8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
BID	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
BAK	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CYTOCHROME C	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CASPASE-3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CASPASE-9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CASPASE-7	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0
CASPASE-6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CASPASE-2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 3. Results of simulations using up- and down-regulated genes for generalized breast cancer as initial states for several cell cycles and apoptosis proteins. Rows correspond to proteins, columns correspond to time steps; a cell of the table reflects an active (1) or an inactive (0) state of the protein during the simulation time step. Active states are shaded in grey.

The results of the simulation for all the considered groups of breast cancer showed the inactivity of the apoptotic proteins and activation of the Cyclin/CDK complexes and major growth factors such as the ERBB family, EGFR and VEGFR. The behaviour of the Cyclin/CDK states of the breast cancer groups differs by intervals of periodic activation of the cell cycle complexes (the more aggressive the type of tumour, the faster the cell cycle generation).

3.3 *Drug target identification*

Drug target identification includes two different options: (1) inhibition of the cell cycle and (2) initiation of apoptosis.

3.3.1 *Inhibition of cell cycle progression*

We selected the blockade of the Cyclin/CDK complexes as the terminal nodes as their inhibition naturally leads to cell cycle arrest [17]. They can be inhibited either directly or indirectly through inhibition or activation of some other up-stream nodes.

At first, we considered each of the Cyclin/CDK complexes as terminal nodes, whose inhibition led to cell cycle arrest. We simulated inhibition of each other node one by one. As is seen from Table 2, Cyclin A : CDK2 and Cyclin E : CDK2 are typically up-regulated during the S phase. For all of the considered breast cancer groups, several common targets were found: CYCE, CYCLIN E, CDK2, PLK1 and AKT-1. The targets were not found for the complex of Cyclin B1 : CDK1, except for SYK which was identified for HER2/neu positive breast carcinomas.

Table 2. Results of target identification using Cyclin : CDK inhibition as terminal nodes for cell cycle arrest.

Effect	*Mechanism*	*HER2/neu positive breast carcinomas,*	*Ductal carcinoma*	*Invasive ductal carcinoma and/or a nodal metastasis*	*Generalized breast cancer*
Inhibition of cell cycle progression	Inactivation of Cyclin D1:CDK4, Cyclin D1:CDK6 (G1 phase)	CYCD1, CYCLIN D1			
	Inactivation of Cyclin E:CDK2 (G1/S phase), Cyclin A:CDK2 (S phase)	CYCE, CYCLIN E, CDK2, PLK1, AKT-1 SYK	N/A	SRC	N/A
	Inactivation of Cyclin B:CDK1 (G2/M phase)	SYK	N/A	N/A	N/A
Apoptosis activation	Activation of Cytochrome C	BCL-2			
		N/A	N/A	RAF-1, GRB-2, PKC, RACK1,	Fibronectin receptor ALPHA5 BETA1, Fibronectin
	Activation of Caspase-3	MKK4, PI3K, MKK6, P38ALPHA, CRKL, HPK1 N/A	N/A	VEGF-A, VEGFR-2, HIF-1ALPHA, VEGF	N/A

N/A – specific targets were not found.

It has been reported [18] that activating K-Ras mutations, frequently present in cancer, up-regulate the SYK kinase in several lung and pancreatic cell lines, and the treatment of K-Ras-dependent cells with either interfering RNA or small inhibitors induces cell apoptosis. Inhibitors of SYK activity and its pathways are expected to be useful in the treatment of inflammation diseases, and in the suppression of B-cell tumour growth [19]. The list of targets identified in this study mostly includes Cyclins, which are known to be crucial for cell cycle progression; thus this corresponds to well-known information [20,21]. PLK1 and AKT-1 might be more interesting targets as their combinations with Cyclins and CDKs are not so obvious, and they have a strong positive influence on cell cycle progression in normal cases and in pathology [22].

We have not found any non-trivial combination of targets whose blockade led to the simultaneous inhibition of particular complexes. As many different proteins are known to

take part in the regulation of Cyclin/CDK complexes, blockade of any one single target does not lead to cell cycle arrest.

3.3.2 *Apoptosis initiation*

Cytochrome C release was chosen as a terminal node for apoptosis initiation, as the activation of pro-apoptotic factors leads to formation of the multimeric complex known as apoptosome and initiates the caspase activation cascades [23]. Inactivation of either fibronection or fibronection receptor alpha5 beta1 leads to apoptosis in generalized breast cancer while inactivation of RAF-1, GRB-2, PKC or RACK1 leads to apoptosis in invasive ductal carcinoma and/or in a nodal metastasis group of cancer. BCL-2 is the common target for all considered breast cancer groups.

Caspase-3 was taken as the terminal node for possible apoptosis initiation. Table 2 shows the identified common targets for all of the considered breast cancer groups: MKK4, PI3K, MKK6, P38ALPHA, CRKL and HPK1. Inactivation of proteins responsible for blood vessel growth, VEGF-A, VGFR-2, HIF-1ALPHA and VEGF, leads to activation of caspase-3 in invasive ductal carcinoma and/or in a nodal metastasis, reflecting the experimentally confirmed importance of blood vessel growth in invasion processes [24].

3.4 *Algorithm verification*

We verified our results by modelling the inactivation of the identified targets in normal cases using active housekeeping genes as input data. At first, the identified promising targets were taken: fibronectin (hyperexpressed in generalized breast cancer), MKK4, PI3K, MKK6, P38ALPHA, CRKL, HPK1 and BCL-2 (common promising targets for all considered breast tumours). Fibronectin is involved in cell adhesion and migration processes including embryogenesis, wound healing, blood coagulation, host defence and metastasis [25]. It remains inactive in the simulation of normal cases. Thus, its inhibition does not affect the normal signalling processes. Inhibition of almost all of the identified targets does not lead to apoptosis initiation, except for three targets: MKK4, MKK6 and BCL-2.

In normal cases, the inhibition of MKK4 and MKK6 leads to activation of Caspases (Figure 4) via the FAS and FADD signalling chain, prompting the suggestion that inhibition of such kinases may be cytocidal.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
Cyclin E:CDK2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cyclin A:CDK2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cyclin B1:CDK1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DAXX	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
FAS	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
FADD	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
CASPASE-10	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
CASPASE-8	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	0	0	0	0	1	1	1	1	1	1	0	0	1	1	1	1	1	1	0	0	1	1	1	1	1
CASPASE-3	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	1	1	1	0	0	0	0	0	1	1	1	0
CASPASE-6	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	1	1	1	0	0	0	0	0	1	1	1

Figure 4. Results of simulations for inactivated MKK4 in normal cases. Rows correspond to proteins, columns correspond to time steps; a cell of the table reflects an active (1) or an inactive (0) state of the protein during the simulation time step. Active states are shaded in grey.

BCL-2 inactivation also leads to initiation of apoptosis via Cytochrome C, through the Cathepsin B chain (Figure 5). However, Cytochrome C release is finished after the 20th time step and remains inactive from the 21st to the 2500th time step.

The verification of the algorithm was also made by reversing the direction of the edges in the network. Inactivation of the edges in the pathology did not give any promising targets. We also used the random microarray data with an unchanged network, so that the number of up- and down-regulated genes remained the same, but the genes were randomly selected.

As promising targets were not found, such data gave evidence that our algorithm identifies not random targets, but those responsible for the important signalling and cell cycle regulation processes.

As can be seen from the results of the modelling, inhibition of almost all of the identified targets leads to apoptosis in the pathology only. Furthermore, inhibitors for fibronectin are attributed to the group of oncolytic drugs, and these have already been used for therapy in clinics. Inhibitors of Cyclin-dependent kinases and Cyclins are also used for cancer therapy. For example, UCN-01, which is on the second stage of clinical trials, inhibits CDK1, CDK2, CDK6, CDK4, PI3K, CHK1, CHK2, PKC, PDK1 and some signalling pathways, and induces apoptosis [26].

4. Conclusion

We have developed an algorithm for identification of promising drug targets and their combinations based on a discrete approach to RN modelling. The approach provides modelling and analysis of large RNs using microarray data for many genes. The results of the simulation using housekeeping genes do not contradict with the processes known to occur in normal cells. Simulation of the pathology has shown iterative division, inactive states of apoptotic proteins and activation of major growth factors. Application of an *in silico* drug target identification algorithm for the four groups of breast cancer: HER2/neu-positive breast carcinomas, ductal carcinoma, invasive ductal carcinoma and/or a nodal metastasis, and for generalized breast cancer showed common and specific targets. Some of the identified targets for invasive ductal carcinoma and/or a nodal metastasis, VEGF-A, VGFR-2, HIF-1ALPHA and VEGF, are mostly important for the growth and maintenance of that type of cancer (e.g., vessel formation and growth). Most inhibitors of the identified targets belong to the group of oncolytic drugs, which are at different stages

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
CYCLIN E:CDK2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CYCLIN A:CDK2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CYCLIN B1:CDK1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
BAX	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
BID	0	0	0	0	0	0	0	1	1	0	1	1	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
BAK	0	0	0	0	0	0	0	0	1	0	0	1	1	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
CYTOCHROME C	0	0	0	0	0	0	0	0	0	1	1	0	1	1	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0
BCL-XL	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CATHEPSIN B	0	0	0	0	0	0	1	1	0	1	1	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 5. Results of simulations for inactivated BCL-2 in normal cases. Rows correspond to proteins, columns correspond to time steps; a cell of the table reflects an active (1) or an inactive (0) state of the protein during the simulation time step. Active states are shaded in grey.

of preclinical and clinical trials; some of these drugs have been launched (Gleevec, Nexavar, Torisel). Inhibitors for Cyclin D1, Cyclin E, AKT, SRC, BCL-2, RAF-1, PKC, VEGF-A. VEGFR-2, HIF-1ALPHA and VEGF have already been studied as drugs for breast cancer therapy (see the Integrity database, integrity.prous.com). Some of the identified targets involved in cell cycle regulation and gene expression do not possess the known inhibitors in the Integrity database, such as RACK1, CRKL and HPK1, and may be interesting for application in breast cancer therapy.

Acknowledgement

This work was supported by the European Commission project No. 037590 (FP6-2005-LIFESCIHEALTH-7).

References

[1] T.W. Hambley and W.N. Hait, *Is anticancer drug development heading in the right direction?*, Cancer Res. 69 (2009), pp. 1259–1262.

[2] A. Burgun and O. Bodenreider, *Accessing and integrating data and knowledge for biomedical research*, Yearb Med Inform. (2008), pp. 91–101.

[3] D. Ghosh and L.M. Poisson, *"Omics" data and levels of evidence for biomarker discovery*, Genomics 93 (2009), pp. 13–16.

[4] R.M. Gutierrez-Ríos, J.A. Freyre-Gonzalez, O. Resendis, J. Collado-Vides, M. Saier, and G. Gosset, *Identification of regulatory network topological units coordinating the genome-wide transcriptional response to glucose in* Escherichia coli, BMC Microbiol. 53 (2007), pp. 1–18.

[5] A.L. Sellerio, B. Bassetti, H. Isambert, and M. Cosentino Lagomarsino, *A comparative evolutionary study of transcription networks. The global role of feedback and hierachical structures*, Mol. Biosyst. 5 (2009), pp. 170–179.

[6] J.E. McDermott, R.C. Taylor, H. Yoon, and F. Heffron, *Bottlenecks and hubs in inferred networks are important for virulence in* Salmonella typhimurium, J. Comput. Biol. 16 (2009), pp. 169–180.

[7] A. Ma'ayan, *Insights into the organization of biochemical regulatory networks using graph theory analyses*, J. Biol. Chem. 284 (2009), pp. 5451–5455.

[8] H. Gea, H. Qian, and M. Qian, *Synchronized dynamics and non-equilibrium steady states in a stochastic yeast cell-cycle network*, Math. Biosci. 211 (2008), pp. 132–152.

[9] W. Materi and D.S. Wishart, *Computational systems biology in drug discovery and development: Methods and applications*, Drug Discov. Today 12 (2007), pp. 295–303.

[10] J. Tyson, K. Chen, and B. Novak, *Network dynamics and cell physiology*, Nature Rev. Mol. Cell. Biol. 2 (2001), pp. 908–916.

[11] S. Braunewell and S. Bornholdt, *Superstability of yeast cell cycle dynamics: Ensuring casuality in the presence of biochemical stochasticity*, J. Theor. Biol. 245 (2006), pp. 638–643.

[12] M. Krull, S. Pistor, N. Voss, A. Kel, I. Reuter, D. Kronenberg, H. Michael, K. Schwarzer, A. Potapov, C. Choi, O. Kel-Margoulis, and E. Wingender, *TRANSPATH: An information resource for storing and visualizing signaling pathways and their pathological aberrations*, Nucleic Acids Res. 34 (2006), pp. 546–551.

[13] A. Kel, N. Voss, T. Valeev, P. Stegmaier, O. Kel-Margoulis, and E. Wingender, *ExPlain: Finding upstream drug targets in disease gene regulatory networks*, SAR QSAR Environ. Res. 19 (2008), pp. 481–494.

[14] F. Kolpakov, V. Poroikov, R. Sharipov, Y. Kondrakhin, A. Zakharov, A. Lagunin, L. Milanesi, and A. Kel, *CYCLONET – an integrated database on cell cycle regulation and carcinogenesis*, Nucleic Acids Res. 35 (2007), pp. 550–556.

[15] Y.V. Kondrakhin, R.N. Sharipov, A.E. Kel, and F.A. Kolpakov, *Identification of differentially expressed genes by meta-analysis of microarray data on breast cancer*, In Silico Biol. 8 (2008), pp. 383–411.
[16] A.H. Wyllie, *Apoptosis and the regulation of cell numbers in normal and neoplastic tissues: An overview*, Cancer Metastasis. 11 (1992), pp. 95–103.
[17] P.S. Sharma, R. Sharma, and R. Tyagi, *Inhibitors of Cyclin dependent kinases: Useful targets for cancer treatment*, Curr. Cancer Drug Targets 8 (2008), pp. 53–75.
[18] A. Singh, P. Greninger, D. Rhodes, L. Koopman, S. Violette, N. Bardeesy, and J. Settleman, *A gene expression signature associated with "K-Ras addiction" reveals regulators of EMT and tumor cell survival*, Cancer Cell. 15 (2009), pp. 489–500.
[19] P. Ruzza, B. Biondi, and A. Calderan, *Therapeutic prospect of Syk inhibitors*, Expert Opin. Ther. Pat. 19(10) (2009), pp. 1–10.
[20] J. Massague, *G1 cell-cycle control and cancer*, Nature 432 (2004), pp. 298–306.
[21] M.B. Kastan and J. Bartek, *Cell-cycle checkpoints and cancer*, Nature 432 (2004), pp. 316–323.
[22] A. Lindqvist, V. Rodríguez-Bravo, and R.H. Medema, *The decision to enter mitosis: Feedback and redundancy in the mitotic entry network*, J. Cell Biol. 185 (2009), pp. 193–202.
[23] S. Gupta, G.E. Kass, E. Szegezdi, and B. Joseph, *The mitochondrial death pathway: A promising therapeutic target in Diseases*, J. Cell Mol Med. 13 (2009), pp. 1004–1033.
[24] G. Neufeld, T. Cohen, S. Gengrinovitch, and Z. Poltorak, *Vascular endothelial growth factor (VEGF) and its receptors*, FASEB J. 13 (1999), pp. 9–22.
[25] Y. Zhang, H. Lu, P. Dazin, and Y. Kapila, *Squamous cell carcinoma cell aggregates escape suspension-induced, p53-mediated anoikis: Fibronectin and integrin alphav mediate survival signals through focal adhesion kinase*, J. Biol. Chem. 279 (2004), pp. 48342–48349.
[26] S. Welch, H.W. Hirte, M.S. Carey, S.J. Hotte, M.S. Tsao, S. Brown, G.R. Pond, J.E. Dancey, and A.M. Oza, *UCN-01 in combination with topotecan in patients with advanced recurrent ovarian cancer: A study of the Princess Margaret Hospital Phase II consortium*, Gynecol. Oncol. 106 (2007), pp. 305–310.