Functional and Structural Genomics and Medicine

Annotation and Predictability of Cellular Pathways: LII. Computability and Potential Use of Parallel Ant Colony Optimization Algorithms

Wilfried Allaerts*

Biological Publishing A&O and Immunology Department, Erasmus MC Rotterdam, The Netherlands

Abstract

A central theme in this paper forms the question whether the cellular pathways in a multicellular organism are predictable and computable, especially during malignant transformation and tumor metastasis. This question is regarded as a complex optimization problem that, however, constitutes a multileveled so-called NP-hard problem. In this paper, the metaheuristics of parallel Ant Colony Optimization (ACO) algorithms are used as a template for modeling, similar to the use of ACO in the study of protein side-chain conformation. The estimation of a Minimum Energy Analogue (MEA) in a cell developmental framework can be described making use of a Probability of Next Step (PNS) function. Hereby, the analogy with the approximation method of Joint Spectral Radius (JSR) in matrix calculus is instructive. However, in order to evaluate the predictive value of a diagnosed skewness of signal propagation in cellular networks, clinical metabolic data from large parient groups are very important and access to these data is gravely missed. Finally, we present a case study to demonstrate the usefulness of ACO heuristics in predicting the role of glycosylation and epigenetic modification in cancer.

Keywords

Complexity Theory - Parallel Ant Colony Optimization algorithms – Cellular differentiation networks - Epigenetics of cancer.

Introduction

Previously, the problem of complex phenotypical behavior of cells during development as well as during malignant transformation or controlled reprogramming, was chosen as a starting point for multiscale modeling [1-3]. In order to make the architecture of cellular landscapes suitable for elucidating predictive studies or estimating the potential value of therapeutic intervention, two questions are considered of paramount importance. First, the question whether cellular phenotype decision steps are reducible to discrete optimization algorithms. Second, whether these optimization algorithms are computable or solvable within the limits of time and data storage capacity or space [4]. The problem of solvability of complex optimization problems is well known from protein side-chain modeling, e.g., for studying protein design, protein structure optimization and receptor docking studies [5-8]. In the group of Li-Jun Quan (7,8), the metaheuristics of (parallel) ant colony optimization is used (based on a single pheromone matrix) to combine different sources of energy functions and to generate protein side-chain conformation with the lowest energies jointly determined by the various energy functions.

The benefit of optimization algorithms results from the properties of mathematical logic and complexity theory, that allow for the translation of the results of a certain problem (like the 'satisfiability' or 'protein design' problems) to a new and distinct problem (4) (see ¶ Complex Optimization Problems). On the other hand, in order to validate the practical usefulness (at a metaheuristic level) of these abstract methods a thorough analysis of the theoretical framework and algorithm implementation procedures is considered pivotal. In this paper we will focus on the solvability and possible use of Ant Colony Optimization (ACO) algorithms for describing cellular transformation processes (see ¶ The metaheuristics of ACO). In order to implement this ACO approach into a cell biological framework, several adaptations are proposed (see ¶ Adaptations of ACO for Cellular Pathways).

When applied to the study of cancer, both experimental studies on the predictive potential of oncogenic markers as well as clinical metabolic profiling studies are important [9]. Moreover, several metabolites may have a profound effect on the progress of tumour growth and metastasis as shown in tumour hypoxia, due to a mechanism of DNA hypermethylation [10]. The usefulness of ACO is further exemplified in the analysis of sequences of events resulting in DNA-(de/hyper)methylation, O-glycosylation and O-phosphorylation [11, 12]. In contrast to the very low incidence of (potential) cancer cells escaping (programmed) cell death or immune surveillance, the chances of these escaped cells to develop into a full grown tumor (and especially after metastasis) are several orders

Article Information

DOI: 10.31021/fsgm.20181102
Article Type: Case Report
Journal Type: Open Access
Volume: 1 **Issue:** 1
Manuscript ID: FSGM-1-102
Publisher: Boffin Access Limited

Received Date: 01 March 2018
Accepted Date: 22 March 2018
Published Date: 05 May 2018

*Corresponding author:

Wilfried Allaerts
Biological Publishing A&O
Immunology Department
Erasmus MC Rotterdam
The Netherlands
E-mail: w.allaerts@planet.nl

Citation: Allaerts W. Annotation and Predictability of Cellular Pathways: III. Computability and Potential Use of Parallel Ant Colony Optimization Algorithms. Funct Struct Genomics Med. 2018 Apr;1(1):102

of magnitude higher [13]. As a result, the predictability of these teratogenic events in a multiscaled model system still challenges the future of medical discovery (see Case Study).

Complex Optimization Problems: Computable or not?

In complexity theory, the question of whether a problem can be solved by a certain algorithm or by hybridizing several searching intelligencies (at the so-called metaheuristic level) (7) is preceded by two primary questions:

a) Is it possible to describe the problem in an optimization form, e.g. is it possible to define a 'global minimum energy conformation' (GMEC) like in protein design studies (4)? The aim is then to find an expression for the energy or potential function that couples an interaction energy to all possible states of the system.

b) Is it possible to describe the problem in a decision form, i.e. can it be formulated in such a way that for each state of the system a 'yes/no' answer can be found for the question whether or not the energy or potential function exceeds a specified constant K (4).

In the decision form, the complex problem is defined "as a set of input-output pairs. Hereby, the algorithm solves a problem if it produces the correct output for every valid input (called 'instance')"(4). Then, algorithm efficiency is estimated according to a classification scheme: according to Pierce and Winfree (2002), solvable algorithms are basically categorized as polynomial-time and polynomial-space algorithms, depending on whether the algorithm requires a number of steps (polynomial-time) or an amount of data storage space (polynomial-space) that are "at most polynomials of the length of the data describing the instance"(4).

"This means that if $|x|$ is the length of the input data x, then there exist constants α and β such that there are polynomial bounds $\alpha|x|^{\beta}$, as opposed to exponential bounds $\alpha\beta^{|x|}$." (4)

The importance of the distinction between the two classes of algorithms is that polynomial-time (or P-) algorithms are efficient and easy to solve. On the other hand, problems for which polynomial-time algorithms are not necessarily known (or NP-algorithm) contain a subclass of polynomial-time reducible problems (NP-complete), satisfying the requirement "that for every 'yes' instance x of a problem A, there must exist a polynomial-space 'certificate' for x that can be checked for validity in polynomial time"(4).

However, following a 'longstanding conjecture' in complexity theory (that $P \neq NP$), it is believed that "No polynomial-time algorithms exist for problems that are NP-complete" (4). Moreover, NP-complete problems "do have the significant property that if there is a polynomial-time algorithm for any NP-complete problem, then there is a polynomial-time algorithm for all problems in NP"(4). Pierce and Winfree (2002) caution for the several decades of failure to find such a polynomial-time algorithm for any NP-complete problem, so leaving us "little cause for optimism".

On the other hand, they demonstrated that one well-known NP-complete problem, - namely the 'satisfiability' problem in mathematical logic, that can be solved in exponential time and for which no known polynomial-time algorithm exists (requiring up to 2^N tests for N Boolean variables) -, can be useful to demonstrate that the problem of protein design is NP-complete. This implies that also other (and all other) problems in the NP-complete class, according to Pierce and Winfree, can be polynomial-time-reduced to the problem of protein design (4). The corresponding optimization problem is therefore termed NP-hard (4). Moreover, exact methods for protein design optimization, such as the dead-end elimination method (5), are valuable approaches based on exponential-time algorithms, which however "increase the pessimism for finding an efficient polynomial-time algorithm." (4)

For the optimization of cellular (phenotype) decision cascades or cellular pathways, we suggest the previous primary questions could be clarified using the following formulation:

1) What is the 'minimum energy analogue' (MEA) for cellular (phenotype) decision cascades?

2) Can this 'minimum energy analogue' (MEA) be translated into a developmental framework or defined in a pathologically relevant time path?

Given that the MEA can be described in a decision form (see b) above), or,

$$f:\left\{V:\Re^2\rightarrow\Re:\forall i,j\rightarrow\exists K:MEA_{ij}\leq K\ \right\}\rightarrow\left\{T,F\right\} \qquad (1)$$

We may still need to find the biological analogue of the T/F (true/false) valued function f, or, the question that still needs to be answered is whether this MEA function has a prognostic value in the perspective of short or intermediate term survival of the patient.

Moreover, from recent advances in oncology research, it is well known that neoplastic and teratogenic transformations occur in a variable but often reproducible way [14]. Therefore, non-linear sequence dynamics of cellular development require a re-definition of the grid characteristics (in the formule above indicated by indices i,j), and hence, also of the MEAij function. From the principle of homeostasis, it follows that the probability of moving from one state (i) into another (j) is inversely related to the energy required for performing this step (within a certain time). In the present study, we therefore suggest to re-define the MEA into a 'probability of next step' function (PNS), defined within a (given) time frame (Δt). Formula (1) then becomes:

$$g:\left\{V:\Re^3\rightarrow\Re:\forall i,j,\Delta t\rightarrow\exists P_{\min}:PNS_{ij}(\Delta t)\geq P_{\min}\right\}\rightarrow\left\{T,F\right\} \qquad (2)$$

From the study of biomolecular 'roadblocks' that play a role in reprogramming cells into iPS, it appears that transient activation of developmental regulators at the genomic level and at the level of the transcriptome are pivotal cornerstones in cellular (reprogramming) pathways (see also 2, 3) [15]. Therefore, exact characterization of cellular decision cascades are considered as hard to solve as protein design algorithms (if any exponential-time algorithm allows for the exact characterization of the transcriptome in a transient environment). Moreover, also the epigenetic modification and 3D-architecture of the genome at kilobase resolution, should be taken into account in the MEA and/or PNS optimization procedures [16]. Similar to the problem of realistic protein design, the metaheuristic approach of combining multiple search intelligences (like in the Ant Colony Optimization algorithm) (7), might become a new asset for assessing the prognostic value of cellular decision cascades.

The probability of finding N specific molecular species in a certain volume (cell-free state) and the probability of transition of this system into a new (also cell-free) state within a certain period of time, in theory has been modeled using a probality density matrix equation system [17]. Alternatively, the dynamics of an N-species system (in ecological modeling) could be described by a Lotka-Volterra-type equation system, wherein the global stability of the system is based on proving the existence of a Lyapunov function [18].

Although the conditions for checking Lyapunov's theorem for polynomial systems are known to be NP-hard, it appears that some cases exist where they can be solved efficiently in polynomial time (using a so-called semidefinite program, SDP) This holds for iterations of probability density matrix equations as described (17), as well as for sum-of-squares [SOS] procedures for polynomial Lyapunov functions [19]. The problem is similar to the approximation method of the Joint Spectral Radius (JSR). The JSR represents the maximum growth rate obtained by taking arbitrary products of the matrices Ai from the iteration series $x_{k+1}=A_{\sigma(k)}x_k$ where the index σ(k) results from a mapping from the integers to a finite set of indices {1,... m}. The JSR is formally defined as:

$$\rho(A_1,\ldots,A_m):=\lim_{k\rightarrow\infty}\max_{\sigma\in\{1,\ldots,m\}^k}\left\|A_{\sigma k}\ldots A_{\sigma 2}A_{\sigma 1}\right\|^{1/k} \qquad (3)$$

Interestingly, the SOS polynomial Lyapunov functions have been used to prove upper bounds on this ρ (A1,A2,..) norm, whereas "the computation and even approximation of the JSR was found to be difficult", if not 'undecidable' (19) [20]. Also, the "computation of upper bounds is much more challenging task than that of lower bounds" (19). For the approximation of the PNS function, defined by a lower bound limit, this comes as good luck. In ¶ 4 we will discuss

how these theoretical concepts are implemented into a cell biological context.

The metaheuristics of Ant Colony Optimization (ACO)

In a 2D-random walk model, developed by Wang et al. [21], cell migration (e.g. during metastasis) is modeled following an attractiveness formula:

$$T_{ij} = \psi \cdot L_j + (1-\psi) \cdot \varepsilon_j \quad (4)$$

(see also 3)

where ψ is the so-called 'search precision parameter' (for $\psi = 0$, we have a purely random walk situation; for $\psi = 1$ the cells select the highest glucose concentration)(21).

There is a distinct mathematical similarity between the search precision parameter (ψ) and the attractiveness η_{xy} in the central algorithm of Ant Colony Optimization (see text box below) [22, 23].

Text Box 1:

Central Algorithm of Ant Colony Optimization (ACO) (Dorigo et al., 1996, 1997)

The probability for an ant k of moving from state x to state y, p^k_{xy} depends on two parameters, namely I) the attractiveness η_{xy} of the move (in fact the desirability of the state transition, which is inversely related to the distance d_{xy}), and II) the trail level τ_{xy} of the move. The trail level is defined a posteriori, and equals the amount of pheromone deposited for transition from state x to y. The parameter α , moreover, is introduced to control the influence of τ_{xy} , and ρ is the pheromone evaporation coefficient (β is a parameter to control the influence of η_{xy}). This results in the following expression for the probability of state transitions for ant k:

$$p^k_{xy} = \frac{\left(\tau^\alpha_{xy}\right) \cdot \left(\eta^\beta_{xy}\right)}{\sum_{allowed} \left(\tau^\alpha_{xy}\right) \cdot \left(\eta^\beta_{xy}\right)}$$

The procedure is run retrospectively, so the trails are updated after the ants have completed their 'solution', or:

$$\tau_{xy} \leftarrow (1-\beta)\tau_{xy} + \sum_k \Delta\tau^k_{xy}$$

Where $\Delta\tau^k_{xy} = \{Q/L_k$, if ant k uses curve xy in its tour and $\Delta\tau^k_{xy} = \{0$, if otherwise. L_k is the cost of the kth ant tour and is typically proportional to the tour length, and Q is a constant.

To incorporate the migratory behavior of a cell into the overall pattern of cellular behavior, it is suggested to make the migration decision a constitutive part of the ACO algorithm. The migrating 'ant' hereby represents either a germ of malignancy (e.g. a somatic mutation) or a (malignant) migratory cell. This generalization of the migratory model allows for incorporating preceding steps before a cell literally breaks away from its surrounding cells or tissue scaffold.

According to Murray, biological pattern formation (in organisms) can be either explained by reaction-diffusion mechanisms or by models based on cellular automata [24]. In both approaches, a mathematical simulation of the processes is based on the combination of stimulatory and inhibitory interactions between molecules or among cells, respectively. Moreover, recent studies have documented the importance of oxygen depletion or hypoxia in DNA hypermethylation and its resulting effect on several tumour promotors (10). In the case of cellular developmental processes, the 'a posteriori' defined trail level (see textbox 1 for central algorithm of ACO) should be replaced by a combination of:

A priori defined molecular transitions: e.g. the transitions proto-oncogene → oncogene → cell cycle-enhancing metabolites;

A posteriori defined attractiveness of cellular transitions: e.g. factors affecting neo-vascularization, tissue degradation, hypoxia, etc.

The combination of a priori and a posteriori causes and resulting effects requires the introduction of a proper spatial framework and distance notion, similar to the well known Voronoi diagrams (or generalized Dirichlet tessellations) [25]. But, whereas in Voronoi diagrams the framework is defined as a collection of cells x,y (associated with sites Pk, j), consisting of sets of all points in x, respectively in y, whose distance to Pk is not greater than their distance to other sites (like Pj), here it is the inverse of the distance between sites that defines the attractiveness of the move η_{xy} (see eq. (2) for definition of 'probability of next step' [PNS] in ¶ 2). Interestingly, when applying ACO to protein side chain packing, Quan et al. (8) have defined the attractiveness formula η_{xy} based on the energy difference (ΔE) induced by residue i picking up rotamer r_j and standardised according to the following formula:

$$\eta_{ij} = \frac{\pi}{2} - \arctan \Delta E \quad (5)$$

(For Δ E > 1 the value of η_{ij} asymptotically approaches 0.)

Moreover, the authors (8) use the energy function (ΔE) as provided by the Rosetta 3.4 platform [26]. The following steps of their procedure are based on the heuristics of energy minimization (7,8). In the following paragraph, we discuss some adaptations of the ACO approach for cellular pathways.

Adaptations of ACO for Cellular Pathways

In order to combine a priori and a posteriori causes (and resulting effects) (see ¶ 3), we suggest to make use of a multiscale system involving discrete network (node) formation – based on molecular connectivity patterns obtained from biochemical, molecular and, more specifically, tumor biology literature - and an attractiveness function defined in analogy with the preceding examples.

New data for discrete network (node) formation

Decades of molecular biological research undoubtedly yielded an enormous amount of molecular interaction data. Although meticulous attempts to generate integrated maps of the molecular interactions e.g. regulating mammalian cell cycle control and DNA repair systems, these maps unfortunately became obsolete due to numerous new discoveries [27]. To name a few, the discovery of the 'Ten Eleven Translocation' (TET) family of proteins or the 'Enhancer of Zeste Homolog2' (EZH2) related pathways (regulating e.g. the expression of Bone Morphogenetic Protein [BMP], leading to glioblastoma and brain tumours) were not known at the time of Kohn's comprehensive map (27) [28-30]. The TET family of proteins (TET1, TET2, TET3) were discovered on the basis of their fusion (of TET1) to the mixed leukemia gene in acute myeloid leukemia (AML) (28, 29), and were recently found to be involved in cytosine demethylation (TET1) [31]. Broadly speaking, studies of epi-genetic regulation mechanisms offered various new insights to the study of molecular interactions, and, more in particular, to the oncogenic effects of these proteins (30, 31) [32]. Also epigenetic mechanisms other than (de-)methylation, like e.g. O-GlcNAcylation, were found to be involved in the regulation of EZH2 protein stability, DNA- and histone methylation and tumour progression altogether (11,32) (see Case Study in ¶ 5) [33]. Moreover, general physiological conditions like hypoxia appeared to cause DNA hypermethylation for instance due to a reduction of TET activity (10).

Therefore, in order to define the discrete network nodes for implementing the ACO approach, not only the (proto)oncogenes and regulators of cell cycle and DNA repair genes like discovered in the eighties and nineties (14, 27), but also these recently discovered key proteins involved in epigenetic regulation should be taken into account (see ¶ 5).

Principles of network growth and architecture

A network architecture that allows for molecular transitions as well as for cellular growth and migration processes, requires a combination of the two paradigms of biological pattern modeling, as mentioned by Murray (24): reaction-diffusion mechanisms and

cellular automata.

The rationale for estimating the molecular interaction rates may be inferred from the equations of the differential form, as proposed by Wang et al. (21):

$$\frac{d(X_i)}{dt} = \sum v_{production} - \sum v_{consumption} \quad (5)$$

To describe the neural activity in a cell automaton, Murray (24) uses the following model equation:

$$A_{t+1}(x) = S\left[P_t(x)\right] - R_t(x) \quad (6)$$

where S[Pt] simply represents the net neural stimulation in the previous session, and Rt the previous session's inhibitory substance (Murray, 1989, p. 508).

The ACO-method is chosen for its potential usefulness for combining both approaches. The corresponding graph type of this combined cellular & molecular interaction network is obtained by merging open trail and loop motifs into a signaling network [34-36]. The use of directed graphs (or so-called Markov chains) allows for the quantitative estimation of probability of transition from one state into another (see also ¶ 4.3 below). In general, the total number of nodes (vortices) and degree of connectivity of these nodes (numbers of edges per vortex) is not a priori known. However, the combination of a posteriori node identification and experimental assessment of edge weight (34) in an iterative ACO procedure may become helpful. Whereas the transcription of active genes and the production of peptides and metabolites may become essentially predictable from detailed knowledge of the transcriptome and the corresponding biochemical pathways involved, the practical usefulness may be limited by the complexity and limited access to these data (9). However, regarding the quantitative characteristics of immune-surveillance and immune-escape of malignant 'germs', the resulting effects largely depend on the immune status of the patient and hence are, in principal, unpredictable. Epidemiological data on the effectiveness of immune defense mechanisms may become indicative for an a posteriori evaluation of the malignancy of a specific germ. Malignant 'germs' are either (proto-) oncogenes, conserved (somatic) mutations or of cellular nature, like metastatic cells. For each of these essentially teratogenic (and hence potentially lethal) germs an immune-suppression and/or immune-escape trade-off can be defined a posteriori.

In analogy with equations (5) and (6) above, the trade-off between immune suppression and immune escape of a malignant germ can be formally represented as:

$$M_{t+1}(x) = S\left[M_t(x)\right] - R_t(x) \quad (7)$$

where S [Mt (x)] represents the growth of a malignant germ during a certain time period, and Rt(x) the immune suppression during the preceding period. When S [Mt (x)] is smaller than Rt(x), the malignancy eventually dies out. In an a posteriori approach, thus the pattern of node connectivity could be estimated and modeled. In the absence of clinical evidence, however, the interaction network is a void and meaningless construct. Similarly, the insufficient public accessibility of metabolomics data, and also inadequate metabolic identification and reporting, may hamper the discovery potential of meta-analysis (of clinical metabolic profiling) in cancer patients (9).

However, an interesting, new hypothesis may shed some light on the complex interplay between metabolic features, the immune function of macrophages and the angiogenic versus metastatic behaviour of groups of cancer cells [37]. Briefly, the interaction of hypoxic tumor-associated macrophages (TAMs) and tumor cells results in re-establishment of abnormal angiogenesis and metastases following inhibition of the mTOR pathway (37). Inhibition of mTOR, an acronym for the mechanistic target of rapamycin and well-known sensor for cellular energy and key nutrients, has anti-tumor effects in cancer cells, but, unfortunately, these anti-tumor effects are countered by so-called pro-tumor effects of these inhibitory substances on TAMs (37) [38]. Inhibition of the mTOR pathway, e.g. as a result of hypoxia, occurs via transcription of the mTOR complex 1 (mTORC1) inhibitor REDD1 (acronym for regulated in development and DNA damage response 1) [39].

Therefore, theoretically, the combination of clinical metabolic data mining and elucidation of complex metabolic pathways and so-called tumor 'secretomes', including the parallel developments of cancer cells and tumor-associated macrophages (37), may generate the necessary input for multiscale tumor modeling [40].

Attractiveness of next move in parallel ACO

In the parallel implementation of ACO methods, the parallel developments of cancer cells and immunocompetent cells or tumor-associated macrophages are to be combined into an algorithm for defining the attractiveness of the 'next move' (see definition of PNS in ¶ 2) [41]. Herein, both the immune defense effects of immunocompetent cells and/or macrophages as well as the beneficial effects of the latter cells on tumor progression (37) should be taken into account. In the present paper we will however focus on the chemical dynamics of the system, since the mutual interdependence of the immune system on idiopathic and acquired immune mechanisms further augments its complexity.

It would be a too simplified view, however, to consider these developments as a purely chemical dynamic system, in which the probabilities of the 'moves' are considered to depend on a potential field around a primary source (one of the best documented sources is the ATPase-enzyme-complex, functioning as a proton pump) and on the distance to this source of the various biological compartments. The compartmentalization of biological systems is emergent and constitutes the complexity of the living organism [42].

At the genomic level, interactions between genes and their promoters and enhancers occur in so-called contact domains (of ~ 185 kb median length), which are also associated with distinct patterns of histone markers (16). To study the 3D-pattern of genome organisation, the authors used the so-called Hi-C technology, which combines DNA proximity ligation with high-throughput sequencing in a genome-wide fashion [43]. The genome appears to be subdivided in so-called "topologically associated domains" (TADs) of ~ 1 Mb, and it seems that Hi-C technology might become useful to study loops across the entire genome (16) [44]. Moreover, the domain borders are recognized following the interaction with loop anchors that bind CTCF (i.e. the CCCTC-Binding Factor, or Transcriptional repressor CTCF, a 11-zinc finger protein that regulates the 3D-structure of chromatin, by binding strands of DNA and thus forming chromatin loops). Herewith, the boundaries between active and heterochromatic or non-active DNA are topologically defined.

Downstream of DNA-transcription, several subcellular compartments are recognized that are all important for post-translational modification (see also ¶ 5). In order to make the 'nodes' of a developmental network consistent with the presumed discrete character of the network, chemical compounds are considered to be in dynamical equilibrium (constant concentrations) within each (subcellular) compartment, but may differ between the various membrane and cytoplasmic compartments of the cell.

For these (sub) cellular compartments, an average distance function d_{xy} (see Text Box 1, ¶ 3) is inferred that is inversely proportional to the effective (chemical) action range in the ACO procedure, and so is also the PNS.

For the τ_{xy} variables, the so-called trail level in the ACO procedure, the relative abundance of marker genes ($\zeta\eta_{nm}$) in a certain node (x,y), based on transcriptome expression data (from clinical data), may be introduced using the following formula's (8-10). Herein, (ζ, η) represent a selection (e.g. 20) of n genes with expression levels above an (a posteriori) fixed threshold. The β parameter here reflects how the expression of gene ζ is influenced by the (n-1) genes η(1..m) (see also Text Box 1, ¶ 3). Not only the influencing of gene ς^n by (n-1) other genes, formulated as :

$$\tau_{xy} \leftarrow \frac{\beta \arctan \varsigma_n}{n\left(\frac{\pi}{2} - \sum_{n-1} \oint \arctan \eta_m d\eta\right)} \quad (8)$$

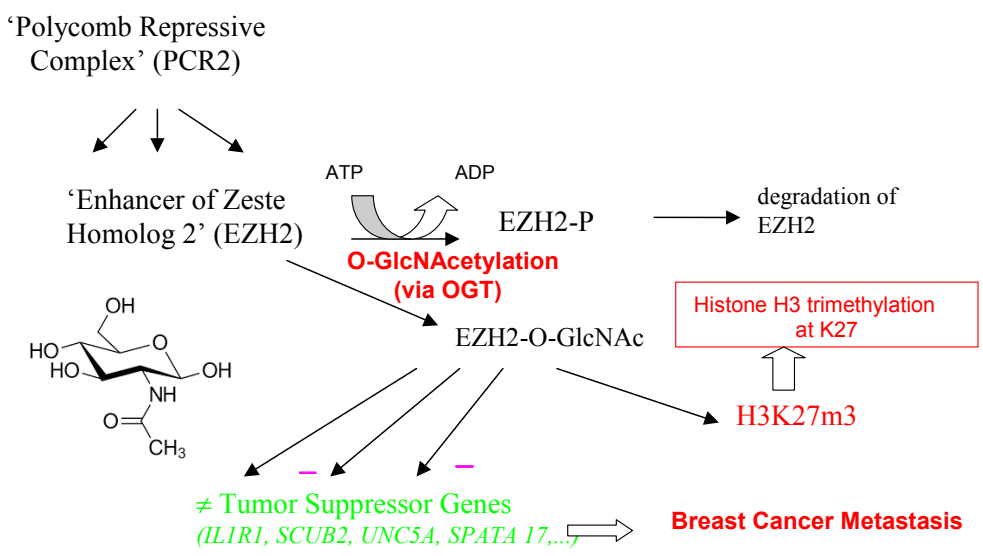

Figure 1: Schematic representation of sequence of events generating DNA-methylation and tumor suppressor gene inactivation

but also a factor ε determining the signal width (of propagation) is introduced:

$$\frac{\beta \arctan \varsigma_n}{n(\frac{\pi}{2} - \sum_{n-1} \oint \arctan \eta_m d\eta)} \to \varepsilon(>1) \to \tau_{xy}^{\varepsilon} \quad (9)$$

(for k>l steps)

or, alternatively, for ε < 1, a 'fading out' of the signal takes place:

$$\Delta \tau_{xy}^{k} = 0 \quad (10)$$

(for k>l steps)

From these d_{xy} and τ_{xy} parameters a so-called interactivity matrix for the x → y state transition has to be calculated in an iterative procedure (see also ¶ 2). At this level, a diagnosed skewness of signal propagation in the network could be introduced, enabling the coupling of clinical data to observed aberrations from the 'normal', healthy state like observed in allergic diseases [45].

However, at each distinct level of the network, the type and number of interacting gene transcripts (mRNA-molecules), but also protein factors and other chemical constituents differ which makes the probability matrices dependent on the network level. Hence, for each level an iteration series (see definition of Joint Spectral Radius in ¶ 2) has to be formulated. This implicates that predicting future developments in cellular pathways corresponds to a multiple (say ω times) of NP-hard problems. Therefore, most cell biological studies are necessarily restricted number of processes within a largely simplified system (see case study below).

Case Study: The role of Glycosylation and Epigenetic Modification in Cancer

For the study of the genetic polymorphisms, epigenetic modification, transcriptome and protein-protein interactions, the contemporary scientist in principle has access to a wide range of scientific databases, e.g. the GenBank database of the National Institutes of Health (NIH), the International Human Epigenome Consortium (IHEC) Data Portal, the various Rosetta Platforms, MethyCancer, etc. The conumdrum of scientific progress these days therefore has become more and more a problem of data 'mining': questions of relevance and completeness of the selected data are important and especially the question how various data sources are interconnected. Nevertheless, meta-analyses of clinical metabolic profiling studies have revealed a shortage of and limitations of the access to the relevant clinical data (9). Regarding the post-translational modifications of peptides, not only these processes are much more widespread than the well-known phosphorylation and glycosylation (11, 12), there appears to be also a complex interplay between the various chemical reactions and the biological effects that these reactions provoke. For a broader survey of the role of glycosylation in cancer-related biological processes in for instance the prostate, the work of J. Munkley is very instructive (12). These processes include cell adhesion, migration and interactions with the extracellular matrix, immune surveillance, cell signalling and cellular metabolism.

In Fig. 1 a schematical network is drawn, representing the different levels of events generating DNA-methylation and the various effects on cancer metastasis (after 34) (Figure 1). Schematically four levels are represented:

Interplay between phosphorylation and glycosylation in regulation of enzyme stability

A well-studied example of regulation of enzyme stability with respect to the metabolic control of signal transduction and transcription is found in the case of the 'Enhancer of Zeste Homolog 2' (EZH2) enzyme (11, 33). EZH2 is linked to the 'Polycomb Repressive Complex' (PRC2) and plays an important role in tumour progression (32). It was found that phosphorylation of EZH2 leads to degradation of EZH2 (11), whereas O-glycosylation through the activity of a so-called 'O-linked-N-acetyl-glucosamine (GlcNAc) transferase' (OGT) establishes an increased stability (33). The corresponding graph is that of a bifurcation pattern. The resulting EZH2-O-GlcAc is involved in histone H3 methylation at specific locations (33). O-linked GlcNAc has been shown to be elevated in numerous cancer types and has been described as a hallmark of cancer [46].

Enzyme stability and oncogenesis through inhibition of tumor suppressor genes

The effect of the Polycomb complex on cancer metastasis is effected through inhibition of several tumor suppressor genes, as is well known for breast cancer cells [47]. Suppressor genes that are involved in cancer metastasis are e.g. IL1R1, SCUBE2, UNC5A, SPATA17 (48). The graphical correlate is that of a multiple, negative

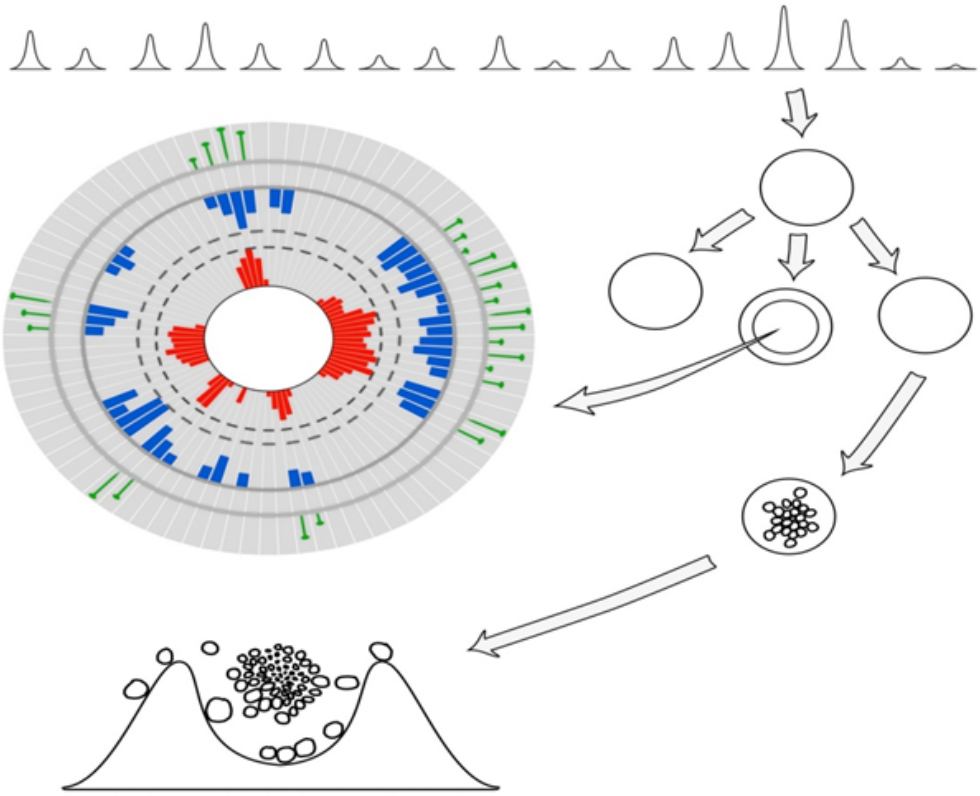

Figure 2: Schematic representation of signal propagation via gene transcription, protein-protein interactions and cellular events in a multicellular network system. The attractiveness of the next move for a migrating 'germ' through the network is represented by a rotating wheel selection system, that also defines the metabolic output at that level *(middle).* Metastasis of tumor cells is suggested to depend on a local sink and threshold mechanism *(bottom)*"

feedback loop system enabling the signal width and speed during propagation (35). In ACO terminology, it means that sideways detracting from cell multiplication or tumor growth are shut off and a fast lane for tumor progression results.

Transcriptional repression and DNA-methylation

The process of so-called 'chromatine silencing' may be effectuated through several mechanisms, as is well studied in the case of OGT [48]. Chromatine silencing is a broad term to designate the multiple forms of transcriptional repression, ensuring that a gene is turned off in an efficient and specific manner (48). For instance, histone deacetylases (HDACs) remove acetyl groups from histone proteins, which in mammalian cells may occur through the Sin3-HDAC omplex (48). Herein, Sin3 appears to act as a scaffold for several of these HDACs and related proteins [49].

One of the effector mechanisms of transcription repression results from the association with the DNA demethylase TET family proteins (TET2, TET3) [50]. TET1, an acronym for Ten Eleven Translocation 1, is a dioxygenase involved in cytosine demethylation [51]. So, besides histone methylation at several specific locations (16, 33), affecting the pattern of active loci in a domain of the chromatin, DNA demethylation through the TET protein family forms another branch of mechanisms affecting the transcriptional landscape [52].

Tissue Inhibitors of Metalloproteinases (TIMP) and Cell Invasion

Moreover, TET1 has a suppressive effect on cancer invasion by activating the Tissue Inhibitors of Metalloproteinase's (TIMP (31). TIMP's are known to have a profound effect on the preservation of extracellular matrix components (ECM), whereas ECM degradation is an important step in cancer cell invasion graphically, the latter effects may become represented as a sink for the propagation of tumor cells [53, 54].(See Fig.2)

Obviously, the present case study gives only a very restricted example of the biological and chemical mechanisms that may become involved in tumor biology. But narrowing down our focus is instrumental for the aim of modeling. Moreover, in view of the funneling effect of the transition of cell malignancy to full-blown (metastatic) tumors (13), it is most instrumental to focus on these transitions where an elevated likelihood greatly determines the outcome.

Summary and Conclusions

Theoretically, the implementation of parallel Ant Colony Optimization algorithms may become a valuable metaheuristic approach that allows for the complexity and multileveled character of cellular pathways, especially in pathological aberrations of the multicellular organism such as in cancer. Important findings of complexity theory, moreover, have also predicted that complex optimization problems in biological modeling are analogous to so-called NP-complete problems. The corresponding optimization algorithms are regarded as NP-hard. The metaheuristics of combining several parallel ACO algorithms for complex multicellular modeling therefore may become comparable to a multiple of NP-hard problems, each of them similar to the Joint Spectre Radius (JSR) method.

In our application of the met heuristics of parallel ACO we ran into several methodical and theoretical difficulties. One major drawback, however, is the lack of accessibility of clinical metabolic data from sufficiently large patient groups. With the input from metabolic data from specific patient groups, the skewness of signal propagation (necessarily restricted to parts of the network) has to be modeled in accordance with known clinical manifestations. Therefore, the

method presented will become mainly instrumental to modeling the impact of interactions between different metabolic pathways and may become helpful for calculating the likelihood of neoplastic disorientation in a multiple balanced cellular network system.

References

1. Allaerts W. Towards the Development of Annotation Grids for Cell Lineage Studies: I. Study of Paradigm Shifts. Med Hypotheses Res. 2008;5:37-51.
2. Allaerts W. Annotation Grids for Cell Lineage and Tumor Metastasis: II. Defining the Architecture for Multiscale Probabilistic and Deterministic Models. Med Hypotheses Res. 2014;9:1-15.
3. Allaerts W. Studying Multiscale Probabilistic and Deterministic Models for the Annotation of Cellular Networks and Tumor Metastases. Journal of US-China Medical Science 2015;12:1-11.
4. Pierce NA, Winfree E. Protein Design is NP-hard. Protein Engineering. 2002;15(10):779-782.
5. Desmet J, De Maeyer M, Hazes B, Lasters I. The dead-end elimination theorem and its use in protein side-chain positioning. Nature. 1992;356:539-542.
6. Desmet J, Spriet J, Lasters I. Fast and Accurate Side-Chain Topology and Energy Refinement (FASTER) as a New Method for Protein Structure Optimization. Proteins: Structure, Function, and Genetics. 2002;48(1):31-43.
7. Lü Q, Xia XY, Chen R, Miao DJ, Chen SS, Quan LJ et al. When the Lowest Energy Does Not Induce Native Structures: Parallel Minimization of Multi-Energy Values by Hybridizing Searching Intelligences. PLOS One. 2012;7(9): e44967.
8. Quan LJ, Lü Q, Li H, Xia X, Wu H. Improved packing of protein side chains with parallel ant colonies. BMC Bioinformatics. 2014;15(Suppl 12):55.
9. Goveia J, Pircher A, Conradi LC, Kalucka J, Lagani V, Dewerchin M, et al. Meta-analysis of clinical metabolic profiling studies in cancer: challenges and opportunities. EMBO Mol Medicine 2016 Oct;8(10):1134–1142.
10. Thienpont B, Steinbacher J, Zhao H, D'Anna F, Kuchnio A, Ploumakis A, et al. Tumour hypoxia causes DNA hyper-methylation by reducing TET activity. Nature 2016 Sep;537(7618):63-68.
11. Kamemura K, Hart GW. Dynamic interplay between O-glycosylation and O-phosphorylation of nucleocytoplasmic proteins: a new paradigm for metabolic control of signal transduction and transcription. Prog Nucleic Acid Res Mol Biol. 2003;73:107-136.
12. Munkley J. Glycosylation is a global target for androgen control in prostate cancer cells. Endocrine-Related Cancer. 2017;24(3):R49-R64.
13. Nguyen DX. Tracing the origins of metastasis. Journal of Pathology. 2011;223(2):195-205.
14. Mukherjee S. The Emperor of All Maladies. A Biography of Cancer. New York, Scribner's Sons.
15. Polo JM, Anderssen E, Walsh RM, Schwarz BA, Nefzger CM, Lim SM, et al. A molecular roadmap of reprogramming somatic cells into iPS cells. Cell. 2012;151:1617-1632.
16. Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. Cell. 2014;159(7):1665-1680.
17. Munsky B, Khammash M. The finite state projection algorithm for the solution of the chemical master equation. Journal Chemical Physics. 2006;124(4):044104.
18. Law R. Theoretical aspects of community assembly. In: McGlade J (editor), Advanced Ecological Theory. Principles and Applications. London, Edinborough: Blackwell Science Ltd. 1999;143-171p.
19. Ahmadi AA, Parrilo PA. Non-monotonic Lyaponov Functions for Stability of Discrete Time Nonlinear and Switched Systems. Proc. 47th IEEE Conference on Decision and Control. 2008;614-621p.
20. Parrilo PA, Jadbabaie A. Approximation of the joint spectral radius using sum of squares. Linear Algebra and its Applications. 2008;428(10):2385-2402.
21. Wang Z, Zhang L, Sagotsky J, Deisboeck TS. Simulating Non-Small Cell Lung Cancer with a Multiscale Agent-Based Model. Theoretical Biology and Medical Modelling 2007;4(1):50-63.
22. Dorigo M, Maniezzo V, Colorni A. The ant system: optimization by a colony of cooperating agents. IEEE Transactions on Systems, Man and Cybernetics, Part B. 1996;26(1):29-41.
23. Dorigo M, Gambardella LM. Ant colony system: A cooperative learning approach to the travelling salesman problem. IEEE Transactions on Evolutionary Computation. 1997;1(1):53-66.
24. Murray JD. Mathematical Biology. Berlin, Heidelberg, New York: Springer-Verlag; 1989;767p.
25. Ashl PF, Bolker ED. Generalized Dirichlet tessellations. Geometriae Dedicata. 1986;20(2):209-243.
26. Rohl CA, Strauss CE, Misura KM, Baker D. Protein structure prediction using rosetta. Methods Enzymol. 2004;383:66-93.
27. Kohn KW. Molecular Interaction Map of the Mammalian Cell Cycle Control and DNA Repair Systems. Molecular Biology of the Cell. 1999;10(8):2703-2734.
28. Ono R, Taki T, Taketani T, Taniwaki M, Kobayashi H, Hayashi Y. LCX, leukemia-associated protein with a CXXC domain, is fused to MLL in acute myeloid leukemia with trilineage dysplasia having t(10;11)(q22;q23). Cancer Research. 2002 Jul;62(14):4075-4080.
29. Lorsbach RB, Moore J, Mathew S, Raimondi SC, Mukatira ST, Downing JR. TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23). Leukemia. 2003 Mar;17(3):637-641.
30. Lee J, Son MJ, Woolard K, Donin NM, Li A, Cheng CH, et al. Epigenetic-mediated dysfunction of the Bone Morphogenetic Protein pathway inhibits differentiation of Glioblastoma-initiating cells. Cancer Cell. 2008 Jan;13(1):69-80.
31. Hsu CH, Peng KL, Kang ML, Chen YR, Yang YC, Tsai CH, et al. TET1 suppresses cancer invasion by activating the Tissue Inhibitors of Metalloproteinases. Cell Reports. 2012;2(3):568-579.
32. Chang CJ, Hung MC. The role of EZH2 in tumour progression. British Journal of Cancer. 2012;106(2):243-247.
33. Chu CS, Lo PW, Yeh YH, Hsu PH, Peng SH, Teng YC, et al. O-GlcNAcylation regulates EZH2 protein stability and function. Proc Natl Acad Science (PNAS). 2013 Jan;111(4):1355-1360.
34. Wilson RJ. Introduction to Graph Theory. 4th ed. Harlow (UK): Pearson Education Ltd. 1996;171p.
35. Ma'ayan A, Jenkins SL, Neves S, Hasseldine A, Grace E, Dubin-Thaler B, et al. Formation of regulatory patterns during signal propagation in a mammalian cellular network. Science. 2005;309(5737):1078-1083.
36. Allaerts W. New cellular concepts in Neuroendocrinology: The anterior pituitary case. Current Trends in Endocrinology. 2011;5:19-28.
37. Wenes M, Shang M, Di Matteo M, Goveia J, Martin-Pérez R, Serneels J, et al. Macrophage Metabolism Controls Tumor Blood Vessel Morphogenesis and Metastasis. Cell Metabolism. 2016;24(5):1-15.
38. Laplante M, Sabatini DM. mTOR signaling in growth control and disease. Cell. 2012;149(2):274-293.
39. Brugarolas J, Lei K, Hurley RL, Manning BD, Reiling JH, Hafen E, et al. Regulation of mTOR function in response to hypoxia by REDD1 and the TSC1/TSC2 tumor suppressor complex. Genes Development. 2004 Dec;18(23):2893-2904.
40. Hernandez-Fernaud JR, Ruengeler E, Casazza A, Neilson LJ,

Pulleine E, Santi A, et al. Secreted CLIC3 drives cancer progression through its glutathione-dependent oxidoreductase activity. Nature Communications. 2017 Feb;8:14206.

41. Randall M, Lewis A. A Parallel Implementation of Ant Colony Optimization. Journal of Parallel and Distributed Computing. 2002 Sep;62(9):1421-1432.
42. Finbow ME, Harrison MA. The vacuolar H+-ATPase: a universal proton pump of eukaryotes. Biochemical Journal. 1997;324(3):697-712.
43. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science. 2009 Oct;326(5950):289-293
44. Jin F, Li Y, Dixon JR, Selvaraj S, Ye Z, Lee AY, et al. A high-resolution map of the three-dimensional chromatin interactome in human cells. Nature. 2013 Nov;503(7475):290-294.
45. Allaerts W, Chang TW. Skewed exposure to environmental antigens complements Hygiene Hypothesis in explaining the rise of allergy. Acta Biotheoretica. 2017 June;65(2):117-134.
46. Zhang Z, Sun P, Liu J, Fu L, Yan J, et al. Suppression of FUT1/FUT4 expression by siRNA inhibits tumor growth. Biochimica et Biophysica Acta. 2008 Feb;1783(2):287-296.
47. Du J, Li L, Ou Z, Kong C, Zhang Y, Dong Z, et al. FOXC1, a target of polycomb, inhibits metastasis of breast cancer cells. Breast Cancer Res Threat. 2012 Jan;131(1): 65-73.
48. Yang X, Zhang F, Kudlow JE. Recruitment of O-GlcNAc transferase to promoters by corepressor mSin3A: Coupling protein O-GlcNAcylation to transcriptional repression. Cell. 2002 Jul;110(1):69-80.
49. Hassig CA, Fleischer TC, Billin AN, Schreiber SL, Ayer DE. Histone deacetylase activity is required for full transcriptional repression by mSin3A. Cell. 1997 May;89(3):341-347.
50. Vella P, Scelfo A, Jammula S, Chiacchiera F, Williams K, Cuomo A, et al. Tet proteins connect the O-linked N-acetylglucosamine transferase Ogt to chromatin in embryonic stem cells. Mol Cell. 2013 Feb;49(4):645-656.
51. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science. 2009;324 (5929):930-935.
52. Wu H, Zhang Y. Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation. Genes Dev. 2011;25(32):2436-2452.
53. Nemoto O, Yamada H, Mukaida M, Shimmei M. Stimulation of TIMP-1 production by oncostatin M in human articular cartilage. Arthritis and Rheumatology. 1996 Apr;39(4):560-566.
54. Lu P, Takai K, Weaver VM, Werb Z. Extracellular Matrix Degradation and Remodeling in Development and Disease. In: Hynes RO, Yamada KM (editors), Additional Perspectives on Extracellular Matrix Biology. Cold Spring Harb. Perspect. Biol. 2011;3(12).