7

Physics of Open Systems: A New Approach to Use Genomics Data in Risk Assessment

Viacheslav Ageev[1], Boric Fomin[1], Oleg Fomin[1], Tamara Kachanova[1], Chao Chen[2], Maria Spassova[2] and Leonid Kopylev[2]
[1]*Faculty of Innovations, Saint-Petersburg State Polytechnic University*
[2]*National Center for Environmental Assessment, Office of Research and Development Environmental Protection Agency*
[1]*Russia*
[2]*USA*

1. Introduction

Modern genomics technologies allow acquisition of huge volumes of gene expression data of the whole genome of different biological systems. It has created a new platform for understanding the exposure effects of toxic stressors on biological systems, for discovering the genetic markers of the negative impact of toxicants, for identifying the potential hazards, and for increasing the precision and efficiency of risk assessment.

Biological processes preceding the appearance of observable harmful effects due to exposures to a toxic chemical are very complex. Recognizing these biological processes as an early response is a key element in increasing the sensitivity of hazard and dose-response assessments for environmental agents. Such processes might be detectable at lower doses and may help to identify a hazard at a low-dose region that is relevant to environmental exposures. An important but often controversial issue in risk assessment is the shape of the dose-response curve at low doses: for example, is there a threshold effect? In the absence of sufficient scientific information, linear and non-threshold effects are assumed as the default position in the U.S. Environmental Protection Agency (EPA) cancer risk assessments (US EPA, 2005). Most attempts to establish a threshold effect were based on traditional toxicological data from animal bioassays, and *in vivo* and *in vitro* experiments for mode of action (MOA) studies that may involve animal or human cell cultures. Formaldehyde provides a good example to show how attempts have been made to evaluate effects at low doses (Conolly et al., 2004) and how uncertain they can be (Crump et al., 2008).

Emerging data from toxicogenomic studies has raised hope that the data may provide useful information about the shape of dose-response at low doses. Bioinformatics methods that can be used for dose-response analysis of high dimensional, multiple factors data are quite limited but are critically needed. Yu et al. (2006) introduced a method to apply Gene Ontology (GO) analysis to dose-response studies. Thomas et al. (2007) applied benchmark dose (BMD) tools (Crump, 1984) and BMD software (US EPA, 2006) to analyze such data. Using micro-array expression data from rats exposed to formaldehyde by inhalation for 6 hours at 0, 0.7, 2.0, 6.0, and 15 ppm, Thomas et al. (2007) identified a subset of genes that

shows a non-linear dose-response relationship with a non-significant gene activity at and below 6 ppm, implying a potential threshold effect. This dose-response behavior is similar to that of rat nasal tumors from inhaled formaldehyde which showed a very steep dose response (Kerns et al., 1983; Monticello et al., 1996). The subset of genes identified in Thomas et al. (2007) is obtained by test of statistically significant difference from the control using one-way analysis of variance. The identified subset of genes may be responsible for the observed tumors at high doses. However, it is possible that there is another subset of genes that may represent the MOA at low doses. Thus, the controversy about the shape of dose-response at low doses remains unless a unique MOA can be established for both low and high doses.

The Physics of Open Systems (PS) is an approach (Kachanova & Fomin, 2010) that is able to reveal the complexity of a bio-system. This paper introduces methods and technologies of PS and provides examples of its use in generation of system knowledge about gene expression on the basis of genomics microarray data. Methods of PS are applied to studying effects of bio-systems chemical exposure by analysis of genomic data. A detailed analysis of the Thomas et al. (2007) data using the PS approach is presented. The results are compared to the results from the original study and some advantages of the PS method are revealed.

2. Physics of open systems

Ideas and methods of PS are realized in the information technologies by providing complexity reduction and reconstruction of the whole empirical information in an open system (Kachanova & Fomin, 1999, 2002, 2003, 2009). Systemological concepts of PS were formed by solving a general problem of reconstructive analysis of complex systems by their empirical descriptions; this solution led to creation of a new method of obtaining scientific knowledge on open systems (*http://isd-consortium.ru/*). Theoretical basis of PS was formed as a result of creating a language of open system and qualimetry (measuring quality) of system knowledge. Development of the language of systems leads to understanding and rational explanation of obtained knowledge. PS is considered completed after the solution to the problem of synthesis of open systems is discovered. Based on this solution, the method is created for modeling states and for discovering emergent (means properties belonging only to the system; could not be derived of individual parts of a system in final number of steps) properties and patterns of an open system.

The PS-based solution for systems research applied problems is accomplished through performing the highly automated technological cycle including five subcomponents: (1) PS technology of context formation transforms experimental data into a format that is used in PS for initial empirical description of the systems; (2) PS technology of system reconstructions builds, based on empirical description of the system, its abstract representation by signed connection graph. The connection graph reflects characteristic for open systems heterogeneity of structures, states, and behavior of the system. This technology generates, based on connection graph, a family of formal system models that reveals the complexity of open systems in a family of its unique qualities. A family of such models contains system knowledge on structural invariants, behavioral invariants, roles and activities of parameters; (3) PS technology of system examination assesses sufficiency of empirical description for construction of complete set of system models, determines degree of completeness of each system model, constructs and verifies invariants of ideal system

states; (4) PS technology of system design generates reconstructions of actual system states. System reconstructions serve as base for obtaining knowledge on system mechanisms that explain magnitude and characteristic changes of parameters in each system state; and (5) PS technology of pattern formation transforms system knowledge into solutions of applied problems (Kachanova & Fomin, 1999).

2.1 Technology of context formation

PS technology of context formation represents the system in data obtained by observing and measuring interactions of the system with the environment. The system, in its natural scale and complexity, maps onto empirical description (empirical context). PS works best with complete empirical contexts. The key object of this technology is the empirical context of the system in the format of tabulated observations. In genomics, each row of the table describes states of one bio-object at given experimental conditions. Each column of the table contains particular gene levels of expression. The table size can be in tens, hundreds or thousands. The result of the technology is representation of a system in a space of its attributes. The coordinates of this space are represented by genes, and points of the space are states of bio-system.

2.2 Technology of system reconstructions

PS technology of system reconstruction automatically produces system knowledge based on empirical description of the system (Kachanova & Fomin, 2003). The empirical description is then transformed into abstract representation of the system in a form of signed connections graph (Fig. 1). The graph vertices are genes. The graph connections are statistically significant binary relationships ($p<0.05$). The edges have several attributes: strength, sign, and monotonicity (homotypic character of changes in gene expressions). These attributes are computed based on several different statistical approaches: Shannon index (Shannon, 1948), Kendall rank correlation coefficient (Kendall, 1970), Hoeffding independence test, and Blum-Kiefer-Rosenblatt independence test (Hollander & Wolfe, 1999). The structure of the binary relationships represents multiplicity of intra-system correlations. The signs of the binary relationships define different forms of system behavior.

The first axiom of PS (Kachanova & Fomin, 1999) states that changes in all system parameters are harmonic. The connection graph as representation of the system obtained from independently determined binary connections satisfies this axiom if all its cycles are even (signed balance). An out-of-balance condition of the connection graph shows heterogeneity of the system and its complexity. Connection graph with signs out-of-balance serves as a base for an automatic generation of complete set of system models. Each model determines system in its one quality. A complete set of models determines all qualities of the system.

Generation of system models starts with the finding in connection graphs of all unbalanced triangles (Fig. 2). Each unbalanced triangle is a minimal structure of binary relationship of parameters with out-of-balance connections signs. A system needs to be discovered in all its qualities and each quality of the system has to be homogeneous. Resolving lack of balance in the connection graph is realized in PS by finding symmetries of structures of relationships – singletons with the ability to harmonize connections between parameters (Kachanova & Fomin, 1999). A singleton is an unbalanced triangle with axial symmetry and system roles of

vertices. One vertex is special and identifies a quality of the system. Two other vertices serve as system-forming two-factor interaction. Vertices that belong to at least one singleton have leading system role. All singletons with the same special vertex form a kernel of system model with preservation of axial symmetry and two-factor relationships. The kernel determines a single quality of the system. System model with such a kernel represents the system as a whole in its one quality. The system as a whole in all its qualities is represented by the complete set of system models (Fig. 2). This complete set discovers complexity of the system.

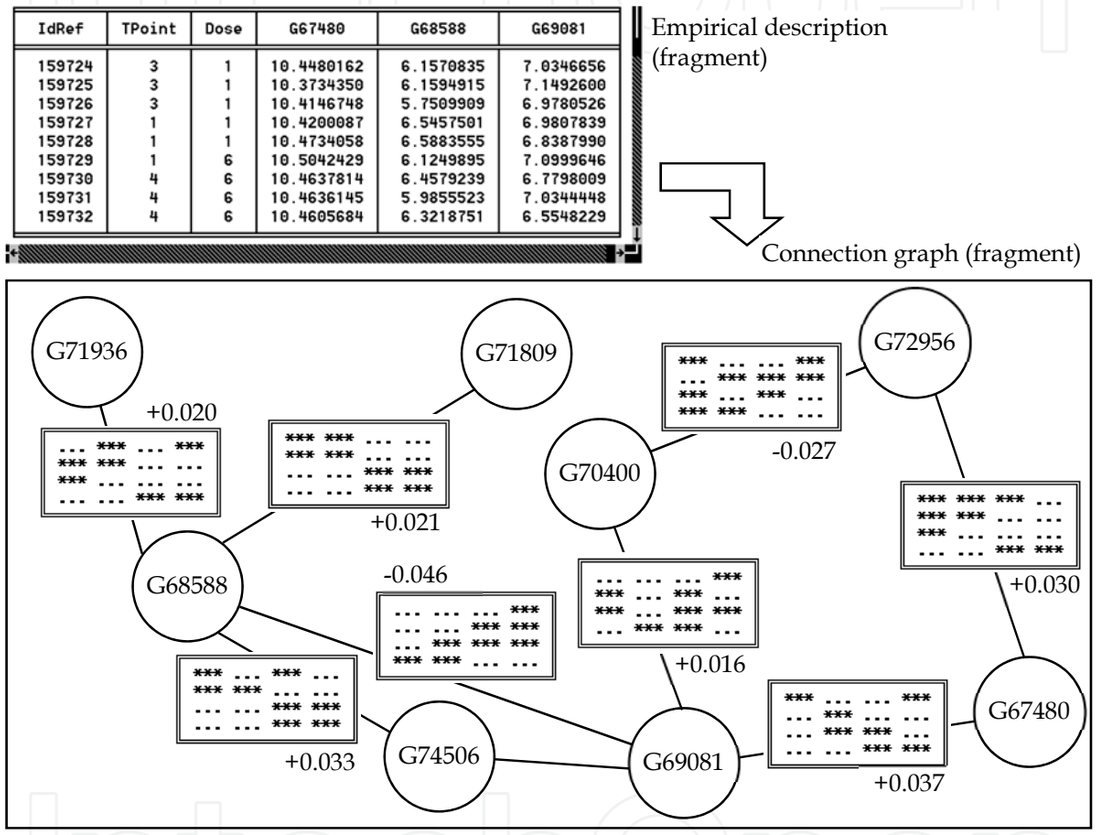


Fig. 1. Transformation of the empirical context into connection graph. Graph vertices are probes. Graph connections are statistically significant binary relationships with attributes: strength, sign, and monotonicity.

The result of PS technology is knowledge on space of qualities of the system which are images of family of abstract system models. Each system model matches a region in which particular quality of the system is represented. The structure of the region determines conceptual borders showing qualities with different intensity.

2.3 Technology of system examination

PS technology of system examination assesses generated system knowledge and constructs, based on system models, a complete set of ideal states of the system. It also maps each region of the space of qualities into the space of attributes and determines set of objects with

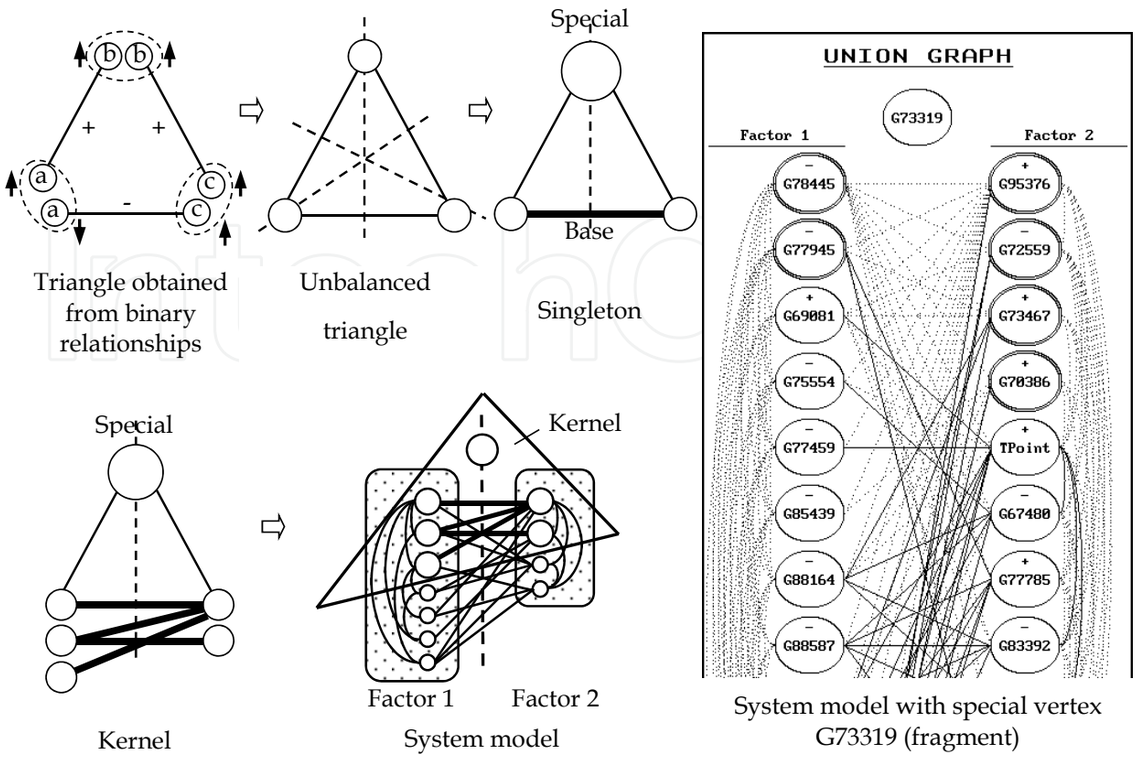


Fig. 2. Schematic description of a system model: finding in connection graphs all unbalanced triangles; determining singletons; obtaining kernel of a system model; building up factors of the system model by inclusion of vertices from the kernel neighborhood; and forming a system model.

quality characteristics for this region. The technology works with words, concepts and assessments of the language of systems (Kachanova & Fomin, 2009). It uses these objects to represent properties of each system using generated system knowledge.

The technology works with different forms of representation of the system: empirical description; complete set of system models; system model of each quality; and complete family of condensed triangles (condensed is in the same sense as condensed graph). For each system representation there is a special approach of assessing the properties of that representation. The technology obtains assessments of system properties of parameters, the structure of binary relationships, and system models. With this as a base, the technology constructs complex assessments of complex properties of the system. For comparative analysis, the technology introduces and uses preference relations and the rule of group choices.

The empirical description of the system is assessed based on its sufficiency for generation of complete system knowledge. In the complete system knowledge, heterogeneity of the system is completely revealed: unbalances are resolved, and changes in all parameters are explained by system mechanisms. The family of system models is assessed by its ability to express completely the space of system qualities. Each system model is characterized by absolute assessments of its ability to represent characteristic quality of the system. To

determine these assessments, the technology constructs a basic sample of the system model (the ideal representation of sense of qualitative determinacy in an abstract form). For this sample, assessments of shapeness and homogeneity are computed. The shapeness is determined by the morphological characteristics of the sample. The homogeneity characterizes conceptual equivalency of all parts of the basic sample. The condensed triangle is the ultimate concentrated image of a system quality expressed by a system model. The condensed triangle serves as an instrument that maps a region of the space of qualities into the space of attributes of the system. For assessment of the quality of the condensed triangle, the technology introduces complex quantitative characteristics of adequacy of representation of a system model in the condensed triangle.

The main purpose of technology of system examination is transformation of the family of system models into a set of models of the ideal states of the system. Main axial symmetry of a system model allows only two ways of concordance of its signs in agreement with the first axiom of PS. Each alternative gives rise to a model of stereotype of the system. The model of each stereotype is transformed into two models of ideal states of the system in accordance with the individualization axiom. This axiom establishes existence of a unique border between high and low values (Kachanova & Fomin, 1999). In a model of ideal state, each parameter obtains a level of value on a nominal scale High/Low. Two models of the ideal states of the same stereotype have the same signs of connections and opposite levels of the vertices values. Complete set of models of the ideal states determines the system as a whole with all its qualities and all ways of manifestation of these qualities in reality.

The direct mapping of regions of the space of system qualities into its space of attributes is achieved by mapping of the set of models of the ideal states on the empirical description of the system. This mapping is achieved by using condensed triangles and special scales of numerical levels of parameters values. The technology constructs scale for each parameter in each ideal model. Each parameter that possesses nominal level (High/Low) obtains a quantitative level. The set of all quantitative levels of parameters determines region of the ideal in the space of attributes. This region contains set of objects whose states correspond to that ideal with different intensity of manifestation of qualities of the ideal in reality. A set of such objects forms cluster of experimental objects (Fig. 3).

Joint set of singletons, system models, and models of the ideal states form complete layout of the system in which concepts of the system are given and revealed in abstract representations. The result of the technology of system examination is knowledge on the quality of the empirical description, the quality of all system and ideal models, and the quality of maps of regions of the qualities space into the space of attributes. The empirical description quality conceptualizes and assesses the system from the position of changing values and features of their correlations. The quality of the model assesses its shapeness, homogeneity, completeness, and ability to completely reveal and correctly transfer concepts of the system to observed data. The quality of mapping provides ability of forming sets of clusters to represent all system concepts revealed in each model of the ideal state.

2.4 Technology of system design

PS technology of system design applies set of clusters for construction of models of observed states. Each ideal state of the system is realized in different experimental objects with different intensity. On the basis of each set of clusters, the technology generates models

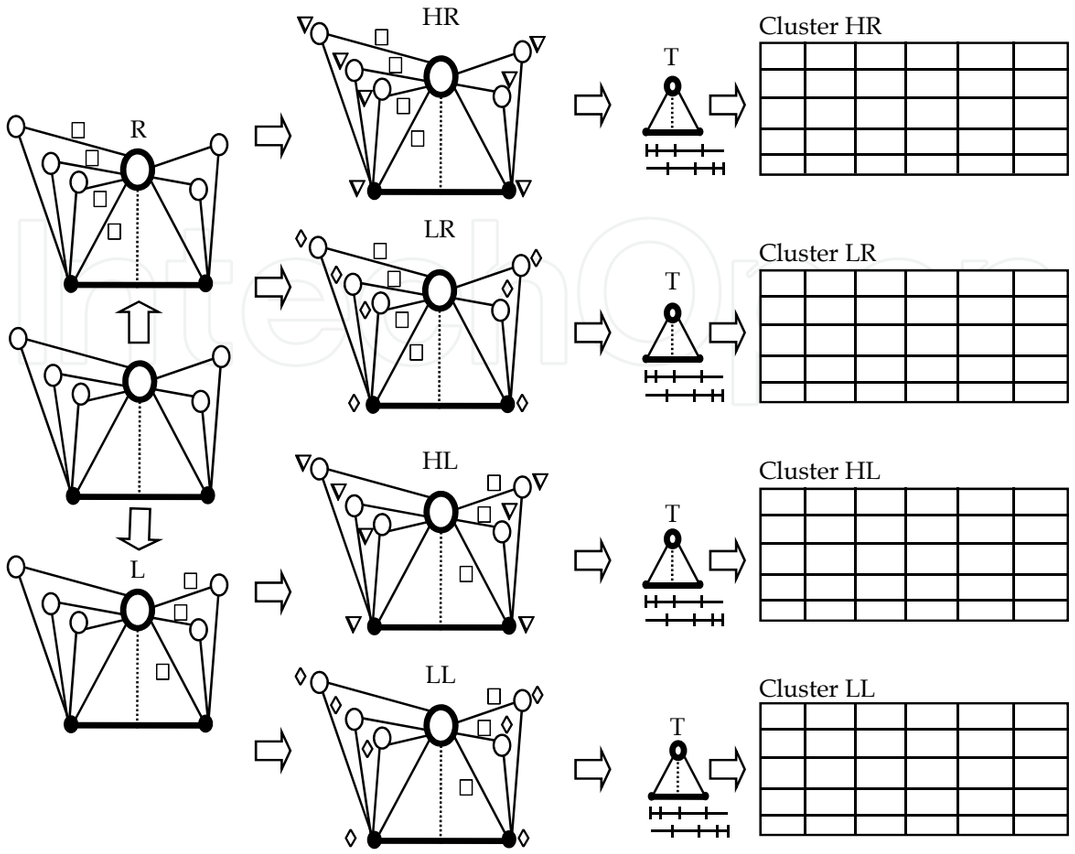


Fig. 3. Schematic of construction of clusters of experimental objects: R, L – models of stereotypes of system behavior; HR, LR, HL, LL – models of ideal states; «□» – balanced model; «∇» – balanced model with special vertex at High level; «◊» – balanced model with special vertex at Low level; and «T» – condensed triangle with numerical levels of values. Cluster is the set of experimental objects corresponding to the model of the ideal state of the system.

of implementation forms of the ideal. Such model includes cluster of experimental objects and assessments of degree of implementation of the ideal in these objects. Objects in the cluster occupy certain region in the space of system attributes generated by the ideal model. The ideal also belongs to this region. The degree of 'closeness' of each object in the cluster to the ideal is determined by a special 'closeness' scale. The ability of each parameter of the state associated with experimental objects to transfer systemic concept of the ideal is assessed by a set of values on quantitative scales (Fig. 4).

The main purpose of the technology of system design is automatic generation of reconstructions of actual system states that are represented in the system empirical description by states of experimental objects. The model of implementation forms of each ideal contains a cluster with experimental objects associated with that ideal. The result of the direct mapping of the ideal onto empirical description of the system is a set of clusters of objects which have intersections. All objects from the empirical description represent states of the system. All objects from each cluster represent different states from the position of one system quality.

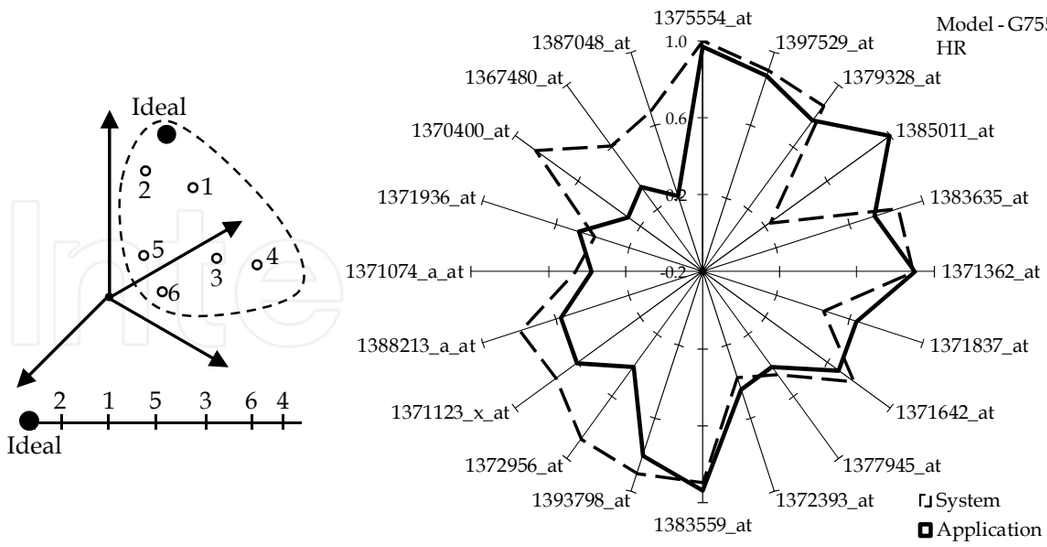


Fig. 4. Attributes of model of implementation forms of the ideal: measure of 'closeness' to ideal on scale of 'closeness' and system significance and application significance of parameters of the ideal model.

Each observed state of the system corresponds to a particular set of models of the implementation forms of different ideals. The reconstruction of implemented state appears as a result of assembly of all models of the given set. In a model of the ideal state, the system has one quality, generated by two-factor interaction that forms the kernel of the system model from singleton with common axial symmetry. In the reconstruction of the implemented state the system is (as a result of that assembly) multi-quality and generated by interactions that form the kernel of the model of the reconstruction from singletons of the ideal models. The kernel of the model reconstruction has not only axial symmetry of singletons but also all symmetries of the higher order, which harmonize qualities that form particular state of the system (Fig. 5).

The states of the system are revealed in reconstructions by parameters and mechanisms that characterize and determine these states. Each parameter has a set of attributes that are assessed from the position of the system as a whole by special quantitative scales. In an observed state of the system, each parameter of the state has a certain value on the measuring scale. In models of the ideal states all parameters are determined on levels High or Low. In models of implementation forms of the ideal states, the levels of parameters are assigned to levels High or Low with a degree of revealing intensity for this level. In the reconstruction of states of the observation, each parameter can be differently determined by different models of implementation forms (some model forms assess value as Low with some intensity and some as High with some intensity). Resolving such uncertainty is achieved by the scale of level prevalence. Each parameter of the state obtains a level on the ordinal 17-item scale. Each item of the scale matches certain intensity of manifestation of high, medium, low values of the parameter. In PS, special numeric scales are constructed based on Farey sums (Conway & Guy, 1996) and Saaty (1980) scale. The stability of parameter level in each observed state is assigned a value on the ordinal 4-item scale – the scale of predeterminacy.

System model	Objects (2 ppм, 6 hours)			
	GSM159757	GSM159758	GSM159759	GSM159760
G77945	HR	LL	LL	LL
G95990	HR	HR	HL	HL
G85439	HR		HR	LL
G74517	HL		LR	HR
G78445	HL		LR	LR
G79201	HL	HL		LL
G87048	HL	LR	LR	
G76337		LR	HR	LR
G75599	HR			HL
G97551	HR			LR
G97697	HR			HL
G73097	HR			LR

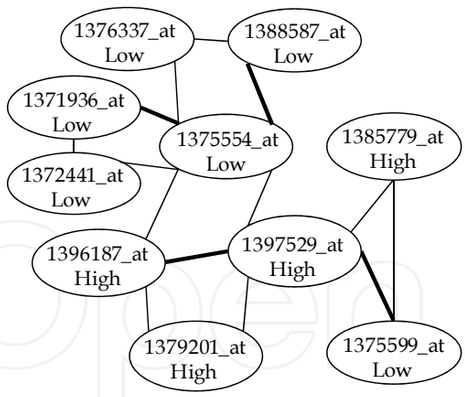


Fig. 5. Reconstructions of states (fragment): The tabular representation of models of reconstructions of four objects; cells of the table show codes of ideal states of system models; and drawing of kernel of model reconstruction of one object.

The mechanism that forms observed state of the system appears as a result of assembly of mechanisms represented by a set of models of implementation forms of the ideals. In the reconstruction of the observed state, each particular system mechanism assists in confirmation or changes of this state. The role of each mechanism in determination of this state is done by the reconstruction of the observed state. For mechanisms with role in changes of the state, it is determined along which parameters and directions these changes do happen. The complete set of reconstructions contains knowledge of all system as a whole and its emergent properties. Thus it represents knowledge on limitations and patterns of conjugacy on different qualities of the system in their observed states. The results of the technology are models of rational explanation of properties of each parameter in each state; properties of system as a whole; properties of observed states of the system as a whole; and mechanisms that form changes of each parameter and of global system properties.

2.5 Technology of pattern formation

PS technology of pattern formation works with the obtained system knowledge. It transforms system knowledge into informational, intellectual, cognitive, and technological resources of knowledge that serve as instruments of solving typical applied problems. Informational resource is knowledge that is the product of the system analysis and conceptualization of empirical facts (deficiencies and assessments of the quality of the empirical description, level of importance of the parameters, and relevancy of the parameters and the experimental objects to the problems that need solutions). The intellectual resource consists of the family of formal models that generate cognitive potential for research (system models as well as assessments of completeness of the system knowledge). The cognitive resource is the knowledge that serves for reasoning and action that has translational potential and provides generation of universally-notional ways of scientific communication such as models, objects, schematics, and language of systems. The technological resource is the objective knowledge on the system as a whole and its parts that allows rational explanation of system states and mechanisms of system changes (states and the space of states). The technology of pattern formation actualizes, organizes, and offers formed knowledge resources for the solution of applied systemic problems.

The technology of pattern formation transforms resources of knowledge into formats that could be applied for solution of applied problems. When working with genomic data, technology accounts for features of empirical descriptions of bio-systems: missing data, data on replicated experiments, and possibly insufficient representativeness. To establish patterns, formats are used in such a way that elements of the knowledge are represented in the special scales allowing for recovery of missing data, aggregation of levels, generalization of levels, and 3-item scale. In reconstructions of the states, levels of the parameters are determined by the system mechanisms. This allows creation of scale that allows recovering levels of values of missing data. Reconstruction of states models observed values of parameters by levels of their values on 17-items scale of level prevalence. Set of levels of the parameters obtained in replicated experiments is represented by an item on 13-item scale of level aggregation. This scale is effective for homogeneous data, because levels from the set lie in one zone of the scale of level prevalence. When data is not homogeneous, aggregation may be achieved by application of the scale of generalized levels. If heterogeneity is prominent, this scale results in level uncertainty. 3-item scale is a simplification of these scales. On 3-item scale, each region of high, medium, and low values obtains item of the scale (H, M, L).

The technology of pattern formation uses, in application to the type of parameters changes caused by external influences, technological resource and, accounting for features of the empirical material, develops methods of solving problems based on the obtained knowledge.

3. Profiles of gene activity for rats exposed to formaldehyde

3.1 Empirical description of GO-categories

The experimental data on gene expressions of rats exposed to formaldehyde (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE7002) are used in this article to illustrate how PS can be used to analyze this type of data. There were 90 rats total (including 25 in the control group). Parameters of the experiment are formaldehyde concentration exposure at various levels (0; 0.7; 2; 6; 15 ppm) and variable sacrifice times (6 hours; 24 hours; 5 days; 8 days; and 19 days) (See Table 1).

Formaldehyde concentration, ppm	Time					Total
	6 hours	24 hours	5 days	8 days	19 days	
0	8	4	4	5	4	25
0.7	4	4	4	4	4	20
2	4	4	5	4	4	21
6	4	4	4	4	4	20
15	4	-	-	-	-	4
Total	24	16	17	17	16	90

Table 1. Number of rats in each experimental group.

Thomas et al. (2007) conducted the experiment and analyzed the genomics data statistically. This article presents results of system analysis of the genomic data by PS methods.

Gene populations were structured according to NTNU GeneTools ontology (http://www.genetools.microarray.ntnu.no) that contains biological, molecular, and

cellular components. GO-categories, containing certain sets of genes, correspond to hierarchical structure from each part of ontology. In each GO-category, genes are endowed by systemic homogeneity. This allows for the consideration of each GO-category as a system and thus allows investigation of rat gene expressions by PS methods. These methods are used to resolve changes in gene expressions for genes from four GO-categories (Table 2). The two GO-categories from the biological process ontology are traditionally thought to be related to formaldehyde MOA (e.g. Thomas et al., 2007). The empirical description of each GO-category as a system is a table with rows showing experimental bio-objects (rats), and column showing values of gene expressions.

Ontology	ID of GO-category	Name of GO-category	Level in hierarchy	Number of genes
Molecular function	GO:0004386	Helicase activity	3	118
Biological process	GO:0006281	DNA repair	6	204
	GO:0006954	Inflammatory response	5	281
Cellular component	GO:0016604	Nuclear body	6	119

Table 2. Description of GO-categories.

3.2 System models

The empirical description of GO-category serves as an initial representation of the system in data; the technology of system reconstructions transforms it into attributed connection graph. The connection graph reveals the complexity (heterogeneity) characteristic of the bio-systems (Table 3). Each significant binary connection in a connection graph serves as projection of multiple intra-system correlations. Heterogeneity of the system is transferred by the population of all binary connections which mostly are weak and complex, resulting in a large number of unbalanced triangles.

Based on the connection graph of the GO-category, the technology of system reconstructions obtains complete family of the system models that reveal system complexity of that GO-category. Each model of this family is generated by the same scheme. All GO-categories are complex systems (Table 4).

The characteristic of heterogeneity of systems is number of the system models and is comprised of number of singletons describing a GO-category. The complexity of each GO-category is completely revealed in the system models. This is confirmed by the high fraction of unbalanced triangles with resolved unbalance (carriers of heterogeneity of the system). The family of the system models determines changes in all parameters with leading system role. In the opposite case, the system knowledge on pattern of changes in the models is not completely revealed. Such a deficiency of the system knowledge may be caused by incomplete representativeness or possible incompleteness of empirical contexts.

3.3 Quality of system knowledge

The system knowledge is represented by families of system models. The technology of system examination provides assessment of the quality of system knowledge (Table 5).

GO-category	Total number of connections	Connection measures		Connection Sign		Connection Strength		Triangles	
		Informational	Nonparametric	Positive	Negative	Complex	Strong	Total	Unbalanced
GO:0004386	3307	1617	1690	2079	1225	1912	5	29368	4472
GO:0006281	9065	4312	4753	5559	3506	5251	0	135690	17730
GO:0006954	14990	7046	7944	10806	4184	8719	18	2858755	26132
GO:0016604	3282	1615	1667	1951	1327	1869	2	28983	3841

Table 3. Quantitative characteristics of connection graphs.

GO-category	Number of models	Number of singletons	Fraction of triangles with unbalance resolved	Leading system role of parameter, %
GO:0004386	126	1379	0.94	100
GO:0006281	209	3667	0.93	100
GO:0006954	275	4476	0.85	100
GO:0016604	123	1455	0.97	100

Table 4. Characteristics of system models.

GO-category	The ideal of expression of system concept		Expression of context in system models	
	Quality of shapeness	Quality of homogeneity	With good quality of shapeness, %	With good quality of homogeneity, %
GO:0004386	0.70	0.82	29	100
GO:0006281	0.58	0.78	56	100
GO:0006954	0.68	0.80	51	100
GO:0016604	0.64	0.79	73	100

Table 5. Integral assessments of quality of the system knowledge.

GO-category	Quality of condensed triangle	Direct mappings %	Volume of implemented concepts	Average number of ideals per object
GO:0004386	0.33	91	3159	35
GO:0006281	0.36	85	4143	46
GO:0006954	0.31	90	6606	73
GO:0016604	0.34	90	2945	33

Table 6. Integral assessments of qualities of direct mappings.

A fraction of models with good assessments is used as general characteristics of completeness and finality of the system knowledge. A system model completely responds to the ideal if assessments of shapeness and homogeneity equal to the scale value of 1. A model has a good quality if its scale value exceeds 0.6. The main problem of the technology of system examination is construction of the set of clusters of bio-objects that are carriers of a particular quality of the system (Table 6). The condensed triangle serves as a tool of direct mapping of the ideals of the system onto objects belonging to clusters. A condensed triangle is constructed for each ideal. Average assessment of quality of all condensed triangles is satisfactory (the mapping tool is adequate when assessment value is 0). Quality of direct mapping is satisfactory if more than 90% of the ideal states obtained empirical confirmation as adequate. The purpose of the direct mapping of the family of models of the ideal system states is determining the volume of system concepts, implemented in observed system states (total number of objects in all clusters is a summed number of system qualities that are associated with experimental animals). Volume of implemented concepts and the average number of the ideals per object reveals noticeable heterogeneity of mechanisms that form observed states of experimental animals (bio-objects).

3.4 Reconstructions of observed states

The technology of system design completes automatic generation of the system knowledge. It constructs models of observed states of bio-objects and assesses their quality (Table 7).

Reconstructions are constructed for all ninety experimental animals. The changes in all parameters for all observed states are determined by the revealed system mechanisms (values of all parameters-obtained explanation). For almost all parameters, the level of value can be established (with unique definition of levels of values on the scale of level prevalence). Pre-determinicity of levels of values by system mechanisms in all observed states is sufficiently high.

GO-category	Number of reconstructions	Determinicity, %	Supportability of the level, %	Pre-determinicity of the level
GO:0004386	90	98	97	84
GO:0006281	90	99	99.8	86
GO:0006954	90	98	98	85
GO:0016604	90	98	97	83

Table 7. Characteristics of states reconstructions.

For solving the problem of formaldehyde influence on changes in gene activity of bio-objects, reconstructions are used as the formal model that explains system patterns of joint harmonized changes of all parameters in each observed state. For each bio-object, each parameter obtains the level of value on the 17-item scale of level prevalence and value of the pre-determinicity attribute of that level (Table 8).

The observed values of each parameter in each observed state obtain formal definitions via sets of models of the system mechanisms that correspond to: determination of High and Low levels in a given state (class 1); formation of medium levels (class 2); determination of change potentials that explain type of changes in observed values (class 3). The quality of

modeling experimental values of parameters by levels is estimated by the concordance coefficient (Fig. 6).

Factor	Value	Level	Pre-determinicity	Model classes		
				1	2	3
1375599_at	3.776569	15	High	G85779/LR	G75599/HR, G89389/LR	G74011/HL, G73256/HL
1397405_at	6.805443	4	High	G73557/LL		G71936/LR, G74342/HL, G97405/LR, G85276/HR, G88213/HR, G73256/HL
1371123_x_at	9.592442	10	Sufficient	G73557/LL, G85349/HR, G75554/HR	G85439/HR, G74332/LL, G72377/LL	G71936/LR, G74342/HL, G74011/HL, G75901/HL, G71123/HL, G88213/HR, G68619/HL
1371642_at	11.87329	9	High	G73557/LL, G85349/HR, G96713/HR, G75554/HR	G89389/LR, G72377/LL	G74342/HL, G77945/LL, G77459/HL, G75633/HL, G94383/HR, G71123/HL, G74470/LR, G88213/HR
1397551_at	7.676976	15	Sufficient	G73557/LL, G98472/HL, G81683/LR	G88864/LL, G97551/HR, G89389/LR, G74332/LL	G71936/LR, G75633/HL, G75901/HL, G73256/HL, G68619/HL
1370979_at	7.183854	7	Sufficient	G85349/HR, G98472/HL, G81683/LR	G72377/LL	G71936/LR, G77945/LL, G75901/HL, G71123/HL, G88213/HR, G68619/HL

Table 8. Reconstruction of observed states of bio-object GSM159785 (fragment). The last three columns: codes of models of system mechanisms (PS notation).

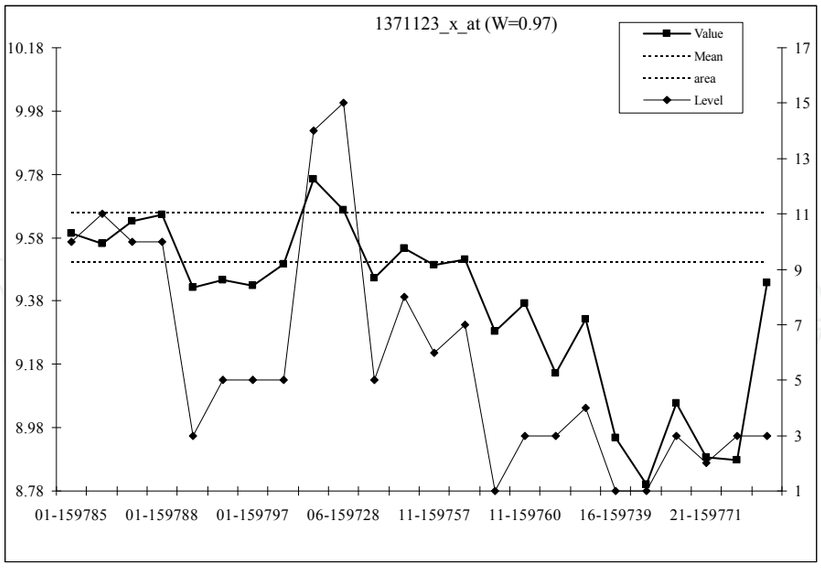


Fig. 6. Values of activity and levels of values for probe 1371123_x_at: x-axis – codes of objects; left y-axis – gene expression values; right y-axis – scale of level prevalence; and W – the concordance coefficient value.

Active probe set	Level of values	Formaldehyde concentration					Significance level	Transition type	Point of the first transition, ppm
		0 ppm 25 rats	0.7 ppm 20 rats	2 ppm 21 rats	6 ppm 20 rats	15 ppm 4 rats			
1367671_at	High	5	2	5	12	4	0.001	LH	6
	Low	9	8	8	1	0			
1368083_at	High	3	5	4	10	4	0.005	LH	6
	Low	10	7	8	2	0			
1368204_at	High	13	5	4	4	0	0.003	HL	2
	Low	1	4	10	8	2			
1368247_at	High	4	13	5	6	0	0.049	LHL	0.7
	Low	6	4	9	8	3			
1368311_at	High	11	9	4	6	0	0.021	HL	6
	Low	7	1	3	9	4			
1368410_at	High	8	8	6	4	0	0.016	HL	2
	Low	3	2	10	8	4			
1369965_at	High	13	8	4	2	0	0.000	HL	2
	Low	3	1	8	11	3			
1370910_at	High	10	9	5	3	0	0.042	HL	2
	Low	6	3	10	8	2			
1371217_at	High	2	9	6	3	4	0.036	C	0.7
	Low	6	4	7	8	0			
1371422_at	High	6	3	8	6	4	0.042	HLH	0.7
	Low	4	11	5	8	0			
1371911_at	High	4	4	8	8	4	0.021	LH	2
	Low	11	7	4	3	0			

1372181_at	High	8	8	5	4	0	0.028	HL	6
	Low	2	5	6	10	4			
1372393_at	High	7	7	6	3	4	0.040	HLH	6
	Low	6	3	9	10	0			
1372548_at	High	4	4	11	9	0	0.016	LHL	2
	Low	9	5	2	5	3			
1373094_at	High	3	5	6	10	4	0.048	LH	6
	Low	8	8	7	4	0			
1373280_at	High	8	5	4	6	4	0.021	HLH	2
	Low	2	3	11	8	0			
1373745_at	High	12	11	4	1	0	0.000	HL	2
	Low	4	1	8	10	3			
1374210_at	High	10	8	4	3	0	0.034	HL	2
	Low	7	2	8	6	4			
1374245_at	High	9	9	7	3	1	0.030	HL	6
	Low	3	4	8	12	2			
1374304_at	High	10	10	3	2	0	0.000	HL	2
	Low	2	2	7	13	2			
1375956_at	High	12	12	7	2	0	0.002	HL	2
	Low	3	4	9	10	2			
1375976_a_at	High	8	7	2	5	2	0.016	HL	2
	Low	2	1	9	4	1			
1376611_at	High	11	6	5	2	0	0.006	HL	2
	Low	2	3	5	10	2			
1377137_at	High	13	8	5	3	0	0.005	HL	2
	Low	4	3	7	11	3			
1377902_a_at	High	11	7	4	5	0	0.004	HL	2
	Low	2	3	8	11	4			
1379499_at	High	9	7	8	4	0	0.025	HL	6
	Low	3	4	7	11	4			
1379654_at	High	10	9	5	3	0	0.005	HL	2
	Low	6	1	7	9	4			
1382030_at	High	8	11	3	1	3	0.002	HLH	2
	Low	3	2	6	9	1			
1382783_at	High	7	11	3	2	0	0.006	HL	2
	Low	5	3	7	10	3			
1383251_at	High	11	6	5	3	1	0.015	HL	2
	Low	3	2	9	11	1			
1383953_at	High	14	7	5	2	0	0.000	HL	2
	Low	2	2	7	12	3			
1384029_at	High	10	7	8	0	0	0.001	HL	6
	Low	6	1	7	9	4			
1384257_at	High	11	9	6	2	1	0.028	HL	2
	Low	3	6	7	10	2			
1384378_at	High	13	8	6	4	0	0.016	HL	2
	Low	2	7	8	9	2			

1384523_at	High	11	9	5	2	1	0.006	HL	2
	Low	4	2	7	11	2			
1385006_at	High	3	5	10	9	0	0.013	LHL	2
	Low	7	9	1	6	2			
1385733_at	High	3	4	9	9	3	0.018	LH	2
	Low	13	5	6	4	0			
1385803_at	High	11	10	3	3	1	0.002	HL	2
	Low	5	0	8	9	1			
1386910_a_at	High	2	5	5	9	3	0.019	LH	0.7
	Low	8	3	10	3	0			
1388254_a_at	High	5	4	6	7	4	0.039	LH	6
	Low	9	9	8	2	0			
1388550_at	High	5	2	5	10	4	0.016	LH	2
	Low	7	11	5	5	0			
1389011_at	High	6	12	6	1	0	0.000	HL	2
	Low	4	1	8	12	4			
1389431_at	High	9	6	6	4	0	0.015	HL	6
	Low	1	4	5	10	3			
1389555_at	High	10	10	7	3	0	0.027	HL	6
	Low	4	4	9	7	4			
1390384_at	High	10	7	3	6	4	0.017	HLH	2
	Low	6	3	12	7	0			
1391078_at	High	12	7	4	4	0	0.005	HL	2
	Low	2	3	8	9	3			
1391491_a_at	High	12	5	5	2	4	0.000	HLH	2
	Low	2	3	8	13	0			
1393367_at	High	9	9	5	6	0	0.049	HL	2
	Low	4	3	8	7	4			
1393405_at	High	4	6	5	2	4	0.045	HLH	6
	Low	2	5	8	9	0			
1393798_at	High	1	2	12	11	0	0.000	LHL	2
	Low	9	6	4	3	4			
1393963_at	High	7	10	5	3	0	0.014	HL	2
	Low	5	3	9	10	4			
1394205_at	High	5	2	7	11	3	0.031	LH	2
	Low	11	8	5	6	0			
1395488_at	High	9	10	6	1	0	0.000	HL	2
	Low	6	1	6	11	4			
1395667_at	High	5	5	4	12	4	0.003	LH	6
	Low	6	7	10	1	0			

Table 9. Distribution of the number of experimental animals by levels of expressions of active genes and formaldehyde concentration for GO:0006281. Patterns of behavior of active genes: LH – monotonically increasing; HL – monotonically decreasing; LHL – convex; HLH – bent; C-complex. Genes determined to be active by both PS and Thomas et al. (2007) methods are shaded. For finding the pattern a statistical test is used (χ^2 test for proportions, $p<0.05$). The point of the first transition is established for each gene; it is defined as the smallest concentration at which changes in gene activity are determined (Fischer exact test, $p<0.05$).

3.5 Profiles of gene expression by formaldehyde concentration

The system effect of formaldehyde exposure is expressed in the pattern of changes in gene activity depending on the formaldehyde concentration. The reconstructions of all states of bio-objects are obtained for each GO-category. All parameters with High (Low) levels of values that have pre-determinicity degree no less than sufficient are obtained in the reconstructions. The complete set of reconstructions is structured by the concentration of formaldehyde: for each gene, a binary relationship "Level of values of gene – Concentration" is constructed (Table 9).

Each active gene is characterized by changes in levels of its values (level transitions) depending on formaldehyde concentration. Five types of transitions are introduced. These types help to understand system effects of exposure. Points of the first transition and types of transitions for all active genes in GO-categories are shown in Table 10. In these profiles the points of the first transition for large number of active genes is below 6 ppm.

GO-category	Point of the first transition				Transition type				
	0.7 ppm	2 ppm	6 ppm	15 ppm	LH	HL	LHL	HLH	C
GO:0004386	6	12	5	0	5	6	4	7	1
GO:0006281	3	36	14	0	9	32	4	7	1
GO:0006954	13	11	13	0	18	7	3	9	0
GO:0016604	2	14	3	0	8	8	1	2	0

Table 10. Number of active genes with characteristic behaviors for GO-categories.

This approach allows finding active genes and constructing profiles of their expression to establish the pattern of response to the formaldehyde concentration, when time is not considered.

3.6 Profiles of gene expressions by formaldehyde concentration accounting for the time

There is another parameter in the experiment, besides concentration. This parameter is the time point of observation. The concentration and the time are independent parameters of the experiment, but they become interconnected parameters in the system mechanisms that determine gene activity patterns. Actions of these parameters may be in different directions and patterns of their influence are complex. For finding the combined effect of two experimental parameters, it is required to establish significant differences of gene activity of exposed groups compared to the control group in each time point. In this problem, it is appropriate to use the whole scale, not just High and Low values. As a result, the solution would involve all bio-objects and all levels of parameters of their states.

Differences with the control group at each time point are found by using the Mann-Whitney criterion to the levels obtained by PS as the system knowledge. When difference is statistically significant for a gene at a time point, all bio-objects in the control group and that exposed group should show the same gene activity. This is verified by the scale of level

aggregation (see Section 2.5). A gene is considered active, if the difference between aggregated levels of activity of that gene for the control and exposed groups is sufficiently large (not less than half of the scale diapason).

Application of this approach allows establishing of the following facts. There is a pattern in how genes in each GO-category reveal their activity at different time points, and the effect of the concentration compared to the effect of the time is insignificant (Fig. 7). The differential gene expression is mostly expressed in the first and last time points. At the lowest concentration (0.7 ppm), differential expressions are present for the most genes.

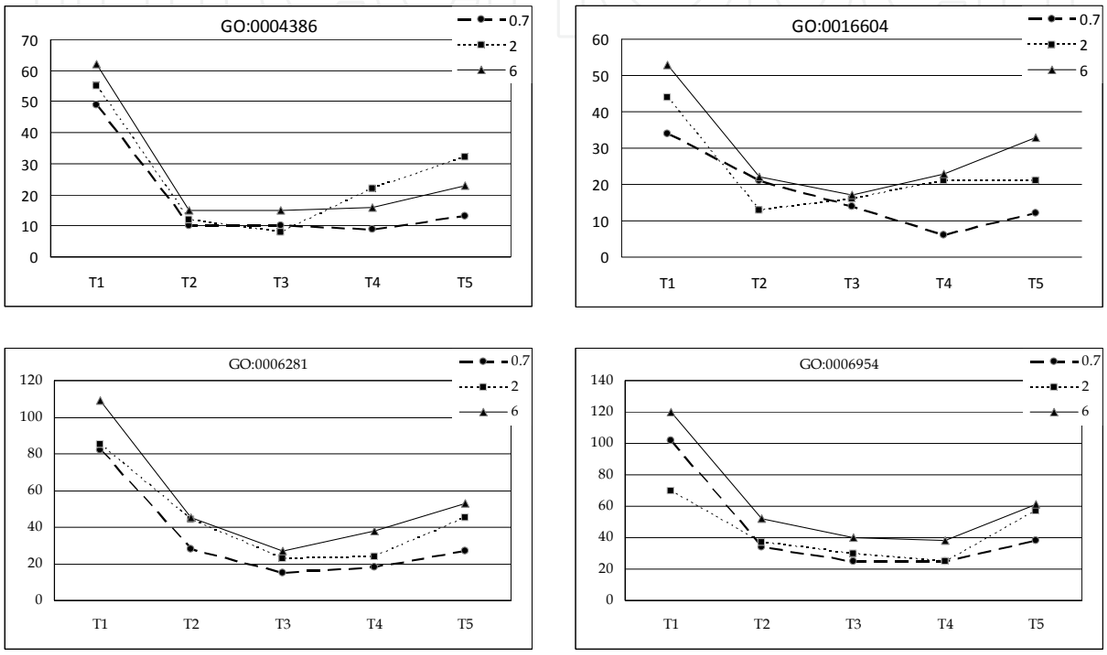


Fig. 7. Number of genes such that exposed groups (0.7 ppm, 2 ppm, 6 ppm) are different (according to the Mann-Whitney test) from the control group at each time point: x-axis – time points; y-axis – number of genes.

The procedure of transition to the aggregated levels for each gene is successful if all bio-objects at a time point are in one zone of the scale of level prevalence, i.e. express the same activity. For the simplified 3-item scale, this fact is reflected by Low, Medium, or High levels. When gene activity is different at the time point, the result of aggregation is not defined (0).

At the first time point (T1) the large number of genes in the control group has an undetermined level on the scale of level aggregation (Fig 8). The time parameter for this group introduces regularity in changes of gene activity. At the low formaldehyde concentration (0.07 ppm), at the first time point, almost every gene exhibited its characteristic level of activity, common for all bio-objects. In this time point of the experiment, the medium level of activity was the most common. At concentration 2 ppm, at the first time point, uncertainty of aggregated levels is lower than in the control group, but higher than in 0.7 ppm group. For 6 ppm concentration, the low level of activity is the most common. The distribution of the number of genes by aggregated levels, time points and concentrations varies considerably (Fig. 8).

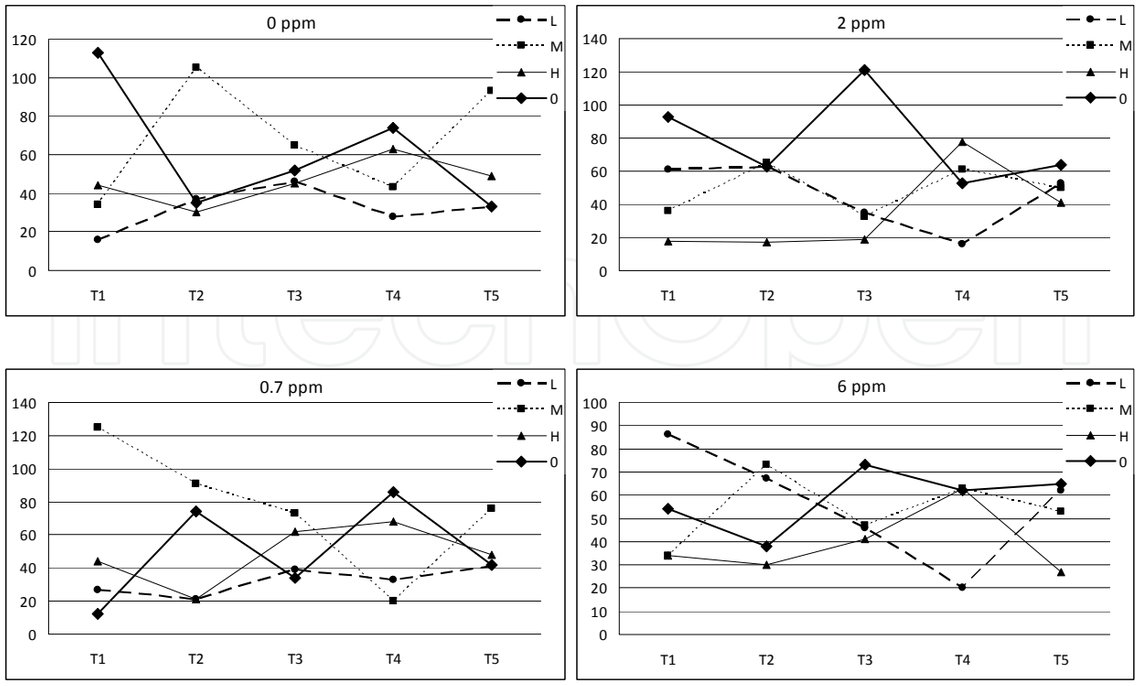


Fig. 8. Distribution of number of genes with aggregates values for concentration by time points (GO:0006281): x-axis – time points; y-axis – number of genes.

The following criteria were used for obtaining set of active genes accounting for both experimental parameters: (1) significant difference with control group and (2) sufficient difference between aggregated levels for the control group and at least one exposed group. Results of this method of determining gene activity accounting for both experimental parameters are compared with results of the method accounting only for the formaldehyde concentration in Table 11 (cf. Table 9). Genes belonging to a GO-category are distributed over six classes. Class 1 includes genes whose activity is determined by both methods and both methods resulted in the same point of the first transition. Class 2 includes genes whose activity is also determined by both methods, but the points of the first transition are different. Class 3 includes genes that are active only according to the second method. Class 4 includes genes active according to the first method, but only weakly active according to the second method. Class 5 contains genes such that according to the second method there is a difference with control group, but activity is weak (the difference between aggregated levels of the exposed and the control groups is not sufficient). Class 6 includes genes activity of which is not found by any of the two methods. Almost all genes in the GO-categories demonstrated activity caused by exposure when accounting for time. Degree of this activity could be measured.

The combined effect of both experimental parameters is illustrated with profiles of expression, obtained for different classes of genes. Profiles are unfolded in the time series for different formaldehyde concentration (Fig. 9). Probe 1383953_at belongs to class 1. Profiles of expression of this gene by time points are analogous for almost all concentrations of formaldehyde. Clear effect of changes in gene activity is revealed at 2 ppm. Probe

1368204_at belongs to class 2. The first approach found point of the first transition at 2 ppm, but for the second approach that point is at 0.7 ppm. Probe 1370172_at belongs to class 3. The time dependence of the gene expression profiles of the exposed groups are in anti-phase with the time dependence of the gene expression profile of the control group. Activity determined by the second method. Activity of probe 1391078_at from class 4 is determined by the second approach as weak. In all time points there are no sufficient differences with control group, however the weak activity for the 2 ppm concentration which is the point of the first transition determined by the first method is revealed. Probe 1395419_at belongs to

Gene classes	Number of genes in a class	Formaldehyde concentration			Distribution of active genes according to Thomas et al. (2007) in a class
		0.7 ppm	2 ppm	6 ppm	
GO:0004386					
1	13	3	9	1	8
2	4	2	2	0	2
3	43	14	19	10	14
4	6				1
5	47				7
6	8				2
GO:0006281					
1	17	2	9	6	11
2	18	0	16	2	7
3	68	27	7	34	11
4	18				7
5	68				8
6	15				0
GO:0006954					
1	15	7	4	4	7
2	12	2	5	5	4
3	108	42	33	33	8
4	10				2
5	108				12
6	28				2
GO:0016604					
1	11	2	7	2	6
2	6	0	5	1	2
3	41	13	9	19	10
4	2				1
5	52				18
6	10				2

Table 11. The distribution of all genes by their classes of activity. In the last column, intersection of genes determined by PS in certain class and active genes determined by Thomas et al. (2007) are shown.

class 5. The second approach determined difference in gene expression from the control for all exposed concentrations, but these differences are small. Neither the first, nor the second method determined probe 1367455_at as active.

In the problem of determining points of the first transition and type of transition, the time parameter is important, as both parameters of experiment influence activity complexly. Accounting for both parameters allows: to divide genes into classes of activity that characterize the form of revealing of combined effect of both parameters; to reveal larger number of active genes; and to reveal a substantial number of genes with small changes in their activity.

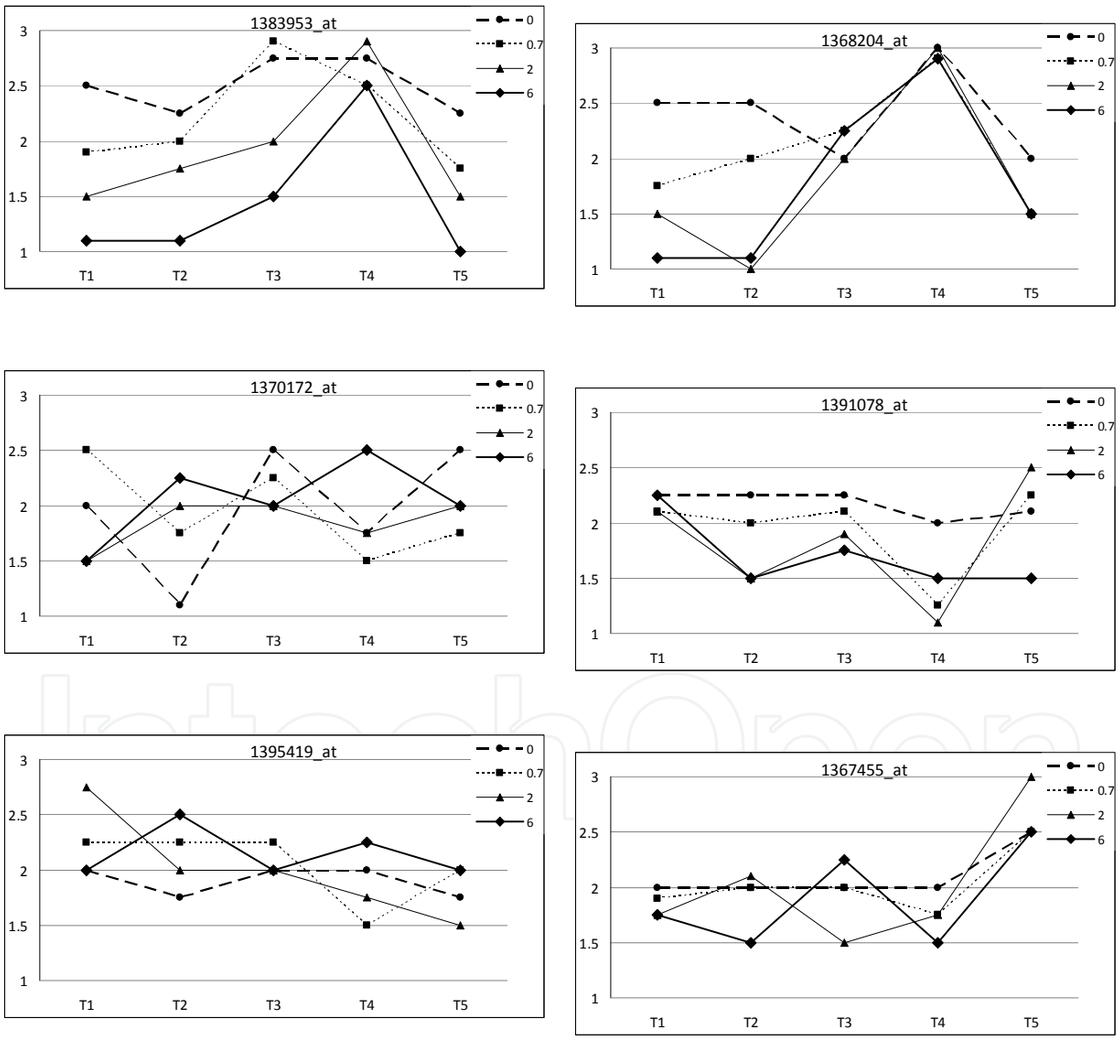


Fig. 9. Examples of profiles of gene expression for 6 classes of genes: x-axis – time points; y-axis – aggregated levels.

4. Conclusion

Physics of Open Systems overcomes complexity of bio-entities by understanding the nature of heterogeneity among them and using that understanding to create the reconstructions of their states. The reconstructions of states of bio-entities allow determination of levels of gene activity. This results in the profiles of gene activity. The reconstructions of states open the path to a deeper understanding of the mechanisms that determine levels of gene activity and determine the systematic responses of experimental animals to the formaldehyde exposure.

The data used in this article comes from the experiment described in Thomas et al. (2007). We considered an example of four GO-categories. For each active gene, we determined the point of the first transition, i.e. where the level of activity of the gene changes significantly (up or down). Our results indicate that the point of the first transition was most often found below 6 ppm, suggesting the absence of a threshold at 6 ppm. We conclude that there is a substantial gene activity below what had been proposed as a threshold of gene activity based on statistical BMD analysis (Thomas et al., 2007).

Some discussion about dose-response controversy is in order. A contentious issue in health risk assessment is whether or not effects observed at high doses in a bioassay can be extrapolated to low level of environmental exposures because MOAs may differ between high and low doses. Therefore, the issue of interest to us is to determine whether gene expression pattern is similar between high and low doses rather than finding only a subset of genes that may be responsible for the effects observed in bioassays. This issue is particularly important if the attempt has been made to use gene expression pattern to establish a threshold effect.

A number of genes from the 'DNA repair' GO category (GO:0006281), that is traditionally thought to be relevant to formaldehyde MOA (e.g. Thomas et al., 2007), were found active in our study in contrast to the Thomas et al. (2007) analysis (Table 9). Among these active genes in discrepancy with the analysis performed in the original study are: alpha 1 subunit of DNA directed polymerase (Gene symbol: Pola1 and probe ID:1376611_at), putative apurinic/apyrimidinic endonuclease 2 (RGD1565983, 1395667_at), APEX nuclease 1– a multifunctional DNA repair enzyme (Apex1, 1371217_at) and DNA non-homologous end-joining factor 1 (Nhej1, 1374245_at). Some of these enzymes are essential for various DNA repair processes. For example Nhej1 is an important component of non-homologous end-joining (NHEJ) repair of DNA double strand breaks (Buck et al., 2006, Ahnesorg et al., 2006, Tsai et al., 2007). Table 9 lists the active genes in 'DNA repair' GO category and characterizes them by the point of the first transition and type of change. Monotonic increase and monotonic decrease are labeled LH and HL respectively. Nhej1 and Pola1genes showed monotonic decrease with formaldehyde concentration, with Pola1 first transition at 2 ppm, and Nhej1 at 6 ppm (Table 9). Overall, 32 genes from 'DNA repair' category exhibit monotonic decrease and 25 of them have the first transition at a low, 2 ppm dose (Tables 9 and 10). Our results suggest that inhibition of key DNA repair proteins at transcriptional level might be an important component of formaldehyde genotoxicity at doses as low as 2 ppm. Formaldehyde inhibition of expression of proteins related to non-homologous end-joining (NHEJ) repair of DNA double strand breaks and base-excision repair (BER) of DNA

adducts has been observed previously (Zhang et al., 2011). In particular, reduction of Ogg1, a key participant in BER, and Ku 70, which regulates NHEJ processes, was detected in cell culture exposed to 150 μM formaldehyde (Zhang et al., 2011). However it is difficult to relate the concentrations of formaldehyde used in these *in vitro* experiments to the doses used in the *in vivo* study for the present analysis.

Important benefit of application of PS methods to the genomics data is its ability to account for both parameters (exposure concentrations and time) of the experiment, something that simpler methods of analysis (e.g. Thomas et al., 2007) are not able to accomplish. The time of observation is a very important parameter for determining gene activity. Almost all genes in the four GO-categories demonstrated activity caused by exposure when accounting for time (Table 11).

Simple methods such as the one proposed by Thomas et al. (2007) cannot be expected to adequately explain the complexity of genomics data. Such simple methods may lead to an erroneous conclusion of a threshold effect that does not exist. More appropriate methods, like the one proposed in this article, are needed for a proper analysis of genomics data.

5. Acknowledgement

Research of authors from Saint-Petersburg State Polytechnic University was supported by ISTC grant number 3476, 2006-2011. The views expressed in this article are those of the authors and do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency.

6. References

Ahnesorg, P., Smith, P. & Jackson, S.P. (2006). XLF interacts with the XRCC4-DNA ligase IV complex to promote DNA nonhomologous end-joining. *Cell*, Vol.124, No.2, pp. 301-313

Buck, D., Malivert, L., de Chasseval, R., Barraud, A., Fondanèche, M.C. & Sanal, O. (2006). Cernunnos, a novel nonhomologous end-joining factor, is mutated in human immunodeficiency with microcephaly. *Cell*, Vol.124, No.2, pp. 287-299

Conolly, R.B., Kimbell, J.S., Janszen, D., Schlosser, P.M., Kalisak, D., Preston, J., et al. (2004). Human respiratory tract cancer risks of inhaled formaldehyde: dose-response predictions derived from biologically motivated computational modeling of a combined rodent and human dataset. *Toxicological Sciences*, Vol.82, No.1, pp. 279–296

Conway, J.H. & Guy, R.K. (1996) *Farey Fractions and Ford Circles. The Book of Numbers*, Springer-Verlag, New York, NY

Crump KS. (1984). A new method for determining allowable daily intakes. *Fundamental Applications Toxicology*, Vol.4, No.5, pp. 854–871

Crump, K.S., Chen, C., Fox, J.F., van Landingham, C. & Subramaniam, R. (2008). Sensitivity analysis of biologicallymotivated model for formaldehyde-induced

respiratory cancer in humans. *The Annals of Occupational Hygiene*, Vol.52, No.6, pp. 481-495

Hollander, M. & Wolfe, D. (1999). *Nonparametric statistical methods*, John Wiley and Sons, New York, NY

Kachanova, T. & Fomin, B. (1999). *Basis of Systemology of Phenomenal,* SPbGETU, ISBN 5-7629-0293-5, St.Petersburg, Russia

Kachanova, T. & Fomin, B. (2002). *Meta-technology of System Reconstruction,* SPbGETU, ISBN 5-7629-0439-3, St.Petersburg, Russia

Kachanova, T. & Fomin, B. (2003). *Technology of System Reconstruction,* Polytechnica, ISBN 5-7325-0772-8, St.Petersburg, Russia

Kachanova, T. & Fomin, B. (2009). *Introduction to the Language of Systems,* Nauka, ISBN 978-5-02-025360-5, St.Petersburg, Russia

Kachanova, T. & Fomin, B. (2010). Physics of Systems is a postcybernetic paradigm of systemology, *Proceedings of WMSCI 2010 14th World Multi-Conference on Systemics, Cybernetics and Informatics*, pp. 244-249, ISBN-13: 978-1-936338-00-9 (Volume III), Orlando, Florida, USA, June 29th - Jule 12nd, 2010

Kendall, M. (1970). *Rank correlation methods*, Charles Griffin, London, UK

Kerns, W., Pavkov, K., Donofrio, D., Gralla, E. & Swenberg, J. (1983). Carcinogenicity of formaldehyde in rats and mice after long-term inhalation exposure. *Cancer Research*, Vol.43, No.9, (September 1983), pp. 4382-4392

Monticello, T., Swenberg, J., Gross, E., Leininger, J., Kimbell, J., Seilkop, S., et al. (1996). Correlation of regional and nonlinear formaldehyde-induced nasal cancer with proliferating population of cells. *Cancer Research*, Vol.56, No.5, (March 1996), pp. 1012-1022

Saaty, T.L. (1980). *The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation*, McGraw-Hill, New York, NY

Shannon, C.E. (1948). A mathematical theory of communication, *Bell System Technical Journal*, Vol.27, (July, October 1948), pp. 379-423 and 623-656

Thomas, R., Allen, B., Nong, A., Yang, L., Bermudez, E., Clewell, H., et al. (2007). A method to integrate benchmark dose estimates with genomic data to assess the functional effects of chemical exposure. *Toxicological Sciences*, Vol.98, No.1, (July 2007), pp. 240-248

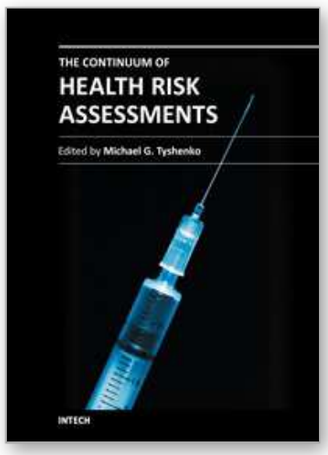
Tsai, C.J., Kim, S.A. & Chu, G. (2007). Cernunnos/XLF promotes the ligation of mismatched and noncohesive DNA ends. *Proceedings of National Academy of Sciences of the USA*, Vol.104, No. 19, pp.7851-7856

U.S. Environmental Protection Agency. (2005). *Guidelines for Carcinogen Risk Assessment*. EPA/630/P-03/001F

U. S. Environmental Protection Agency. (2006). *Help Manual for Benchmark Dose Software - Version 1.4*. EPA/600/R-00/014F, http://www.epa.gov/ncea/bmds/

Yu, X., Griffith, W.C., Hanspers. K., Dillman. J.F. III, Ong, H., Vredevoogd, M.A., et al. (2006). A system-based approach to interpret dose- and time-dependent microarray data: Quantitative integration of gene ontology analysis for risk assessment. *Toxicological Sciences*, Vol.92, No.2, (August 2006), pp. 560–577

Zhang, R., Kang, K.A., Piao, M.J., Kim, K.C., Lee, N.H., You, H.J., et al. (2011). Triphlorethol-a improves the non-homologous end joining and base-excision repair capacity impaired by formaldehyde. *Journal of Toxicological and Environmental Health Part A*, Vol.74, No.12, pp. 811-821



The Continuum of Health Risk Assessments
Edited by Dr. Michael G. Tyshenko

ISBN 980-953-307-582-7
Hard cover, 194 pages
Publisher InTech
Published online 16, May, 2012
Published in print edition May, 2012

This book presents a collection of health risk assessments for known and emerging hazards that span a continuum. Case studies for existing health risks include psychoactive drug usage in delivery truck drivers and using look-back risk assessment for accidental syringe re-use in healthcare settings. Case studies for emerging risks include precautionary actions to safeguard blood supplies; nanoparticle deposition in the lung; and the epistemic issues surrounding genetically modified organism risk assessments. The final section of the book deals with advancing health risk assessment analyses through a post-genomics lens and provides case studies on personalized genomics, new data analyses and improving in silico models for risk assessment. These case studies provide much insight into the ongoing evolution of health risk assessments.

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821

IntechOpen