

Decomposing Biological Complexity into a Conjunction of Theorems. The Case of the Melanoma Network

Giovanni Boniolo[†]
giovanni.boniolo@unife.it

Luisa Lanfrancone[‡]
luisa.lanfrancone@ico.eu

ABSTRACT

The complexity of intracellular molecular pathways can be simplified by the use of Network Biology that breaks down the intricacy of biological processes into components and interactions among them (interactome). In the paper we show that any complex interactome, that is, a biological network representing protein-protein, protein-DNA and DNA-RNA interactions, can be decomposed into a conjunction of logical theorems expressed in terms of *Zsyntax*, a formal language which allows representing (as long as you want) biological pathways. This result, illustrated with the case study of melanoma network, opens the possibility for a computable model of the cell expressed in a logical language and shows how a formal way of intending philosophy can be useful to cope with the complexity of the biological world.

keywords: network biology, philosophy of biomedicine, zsyntax.

1. Introduction

Over the last years Network Biology has increased its space and relevance in scientific journals and research centres. It mainly deals with biological networks which try to summarize biological complexity by means of graphs whose nodes represent biological molecules (proteins, DNA, RNA, etc.) and whose edges represent interactions among them. It is rather a shared opinion that such an approach offers advantages and disadvantages. On the one hand, it

[†] Dipartimento di Scienze Biomediche e Chirurgico Specialistiche, Università di Ferrara, Italy; Institute for Advanced Study, Technische Universität München, Germany.

[‡] Department of Experimental Oncology, Istituto Europeo di Oncologia, Milan, Italy.

gives the enormous benefit of joining, also with an immediate and intuitive visual rendering, information otherwise disconnected and of highlighting the most important ones (Lima-Mendez, van Helden, 2009; Qi, Ge, 2006). On the other hand, it involves severe simplifications of the real biological complexity. Nevertheless, notwithstanding these simplifications and sometime even thanks to these simplifications, some information on the systems represented can be obtained and it is not so rare that interesting discoveries can be achieved (Vidal, Cusick, Barabási, 2011; Hang et.al. 2014).

The mentioned simplifications concern especially semantic limitations. In particular, the network representations can be rather unspecific concerning the structure and variety of the real molecules in play, and many iconic and pictorial adjuncts (cartoons, arrows of different colours and shapes, added tables, etc.) have been introduced to provide some clearer and more complete information on the network (Kohn, 1999). The network community is obviously aware of these drawbacks and many efforts have been done to overcome them, even if there is no consensus on which could be the best way to schematically represent the network without losing details and information (Bruck, Ebenhoh, Heinrich, 2006; Faeder, Blinov, Hlavacek, 2005; Kitano et al. 2005; Cho, Kim, Przytycka, 2012).

We have previously proposed a logical and computer-implementable language, called *Zsyntax*, that can help the complete and rigorous representation of a biological network (Boniolo, Di Fiore, D'Agostino, 2010; Boniolo, D'Agostino, Piazza, Pulcini, 2013; Boniolo, D'Agostino, Piazza, G. Pulcini, 2015). It is an innovative tool since, due to its characteristics, allows a mathematically rigorous demonstration of molecular biology processes as theorems. *Zsyntax* allows representing (as long as you want) biological pathways belonging to a network as formal deductions starting from certain premises (a certain initial aggregate of molecules) and arriving at a conclusion (the final molecule of the pathway in question).

This means that given a biological pathway starting from molecule M and arriving at molecule M' , such a pathway can be represented as the theorem $IA \vdash M'$, where IA indicates the *initial aggregate* (i.e. the set containing all the molecules necessary to move from M to M') and \vdash indicates that there is a deduction from IA to M' . Thus any empirical biochemical step (reaction) from M to M' can be formally rewritten as an inferential step of the proof of the theorem $IA \vdash M'$ (Boniolo, Di Fiore, D'Agostino, 2010).

In the end, any *Zsyntax* theorem represents a given biological pathway, and any inferential step represents a movement from a formula representing an empirically grounded reaction to another.

Zsyntax language can be used to write algorithms and computer programs, thus rearranging biological information in a way that allows computer processing. Thus, the potential value of this language does not lie only in the formal representation of already known pathways but also in the field of text mining and in the domain of biological prediction, as it has been already discussed (Boniolo, Di Fiore, D'Agostino, 2010).

This is not the right place to argue for the relevance of a formal language for molecular biology. However, it is worth emphasizing that it ensures, how it is already shown by several examples coming from computational biology and mathematical biology, non-ambiguity and a degree of precision that cannot be achieved by ordinary language. Concerning *Zsyntax*, (i) it provides a mathematically rigorous representation of molecular biology processes, (ii) it is computer-implementable, which means that it may allow researchers to capitalize on the growing body of research carried out in the field of automated deduction, which aims to create computer programs to demonstrate theorems.

In this study, we show that any complex interactome, that is, any biological network representing protein-protein, protein-DNA, and/or DNA-RNA interactions, can be, first, decomposed into a series of *Zsyntax* theorems and, then, logically reconstructed as a conjunction of those theorems. This means that any biological network can be logically rewritten into computable theorems, with all the advantages of breaking down complex interactions saving all the intrinsic information. This result is particularly relevant for those who work in the field of Network Biology where interaction networks have to take into consideration dual relationships between individual nodes and the organization of cellular communication at the same time. Needless to say, this is an important step forward also for the molecular biologists who, especially now, have to do with big amounts of data which cannot be treated easily in a non-computable way.

As an exemplar case study of the potentialities of this outcome, we focus our attention on a specific melanoma network that can be found in the KEGG (Kyoto Encyclopaedia of Genes and Genomes) network, realized by Minoru Kanehisa (Kanehisa, 2013; Kanehisa et al., 2004; Kanehisa et al., 2006; Aoki-Kinoshita, Kanehisa, 2007). The biological database is a freely available web resource that integrates completely sequenced genomes with functional

information. Within the KEGG objects, such as genes and proteins, small molecules, reactions, pathways, diseases and drugs, we concentrate our attention on a specific pathway driving proliferation and survival in melanoma (www.genome.jp/kegg/pathway/hsa/hsa05218.html). This KEGG pathway has been chosen to illustrate how a simple and linear pathway can be broken down in theorems and then reconstructed holding an increased amount of information by the use of a novel mathematical language, *Zsyntax*. By this we mean that each theorem contains a small piece of information concerning a particular chain of reactions belonging to the pathway. Yet whenever we make the conjunction of all the theorems at issues, we regain, on the one hand, the complete information concerning the pathway but, on the other hand, we have also the information regarding how this complete information is composed and realised.

What we are going to present should be seen also as a way in which philosophy, especially formal philosophy, might positively cope with certain scientific questions by offering formal conceptual tools. In particular, we wish to show how complexity can be “deconstructed” in a series of computable theorems. This does not mean that we want to be reductionist. Actually we offer a tool that allows to bridge, as it will be illustrated, the complexity grasped by biological networks and the “simplicity” of the individual molecular pathways forming molecular mechanisms. Philosophy, indeed, is not just a way of abstractedly reflecting on, in our case, science, but also as a way of producing tools (even formal tools) to offer advancements in given scientific sectors. *Zsyntax*, the language here presented and illustrated with the case of melanoma network, should be seen exactly in this way: as a formal philosophical tool to increase, in the field of computational biology, our capacity of dealing with biological complexity by decomposing it in small pieces (each theorem) but without any reductionism intention.

2. Results

2.1. The *Zsyntax* Representation of a Melanoma Pathway

As said, in order to illustrate the use and the potentialities of *Zsyntax*, let us consider the melanoma KEGG network (see Figure 1) (Kanehisa, 2013). Let us begin by describing through the biologists’ usual ordinary language, the pathway concerning the proliferation and the survival of melanocytic cells

through MITF activation upon stem cell factor (SCF) stimulation (Lennartsson, Ronnstrand, 2012; Ronnstrand, 2004)¹.

SCF is a growth factor (GF) prototype in the KEGG scheme. The binding of SCF to c-KIT (that is, a tyrosine kinase receptor (RTK) prototype in the same scheme) leads to the dimerization and autophosphorylation of the receptor. In this way, c-KIT is allowed to bind downstream signalling effectors, such as SRC and SHCA adaptor proteins (Masson, Ronnstrand, 2009; Roskoski, 2005). SRC has the ability to bind c-KIT and by doing so, becomes activated, phosphorylates SHCA and suppresses differentiation to favour proliferation (Phung et al., 2011). SHCA provides association sites for the Grb2/SOS complex and activates RAS and RAF binding, thus sustaining MAPK activation. Grb2 is an adaptor protein involved in the signal transduction cascade downstream of several receptors, while SOS is a guanine nucleotide exchange factor that activates RAS. The stimulation of the RAS/MAPK pathway finally leads to MITF activation by its direct phosphorylation (Samayawardhena, Pallen, 2008). MITF is a transcription factor whose transactivation activity is increased when phosphorylated, resulting in increased proliferation of melanocytic cells (Phung et al., 2011; Vance, Goding, 2004).

In term of *Zsyntax*, what above means the theorem

$$(SCF, C) \vdash MAPK,$$

where (SCF, C) is the initial aggregate (IA), and MAPK is the final molecular outcome. In particular, C represents all the molecules, which should be taken into account if we want a complete representation of the pathway going from SCF to MAPK, that is,

¹ In this case, following the nomenclature indicated by the *HUGO Gene Nomenclature Committee*, (www.genenames.org) MITF is the Microphthalmia associated transcription factor; SRC is the sarcoma viral oncogene homolog; SHCA is the SRC homology 2 domain containing, transforming protein A; Grb2 is the growth factor receptor-bound protein 2; SOS is the Son of Sevenless; GTP is the guanosine triphosphate; ATP is the adenosine triphosphate; PI3K is the phosphatidylinositol-4,5-bisphosphate 3-kinase; AKT is the serine/threonine-specific protein kinase; PIP2 is the phosphatidylinositol bisphosphate; PIP3 is the phosphatidylinositol trisphosphate; PDK1 is the phosphoinositide-dependent kinase-1; MAPK is the mitogen-activated protein kinases, originally called ERK - Extracellular signal-regulated kinases -; MEK is the mitogen-activated protein kinase kinase; RAS is the Rat Sarcoma protein; RAF is a serine/threonine-protein kinase; GSK is the glycogen synthase kinase.

C (c-Kit, SRC, SHC1, Grb2, SOS, RAS, GTP, RAF, MEK, MAPK, ATP, ATP, ATP, ATP, ATP, ATP).

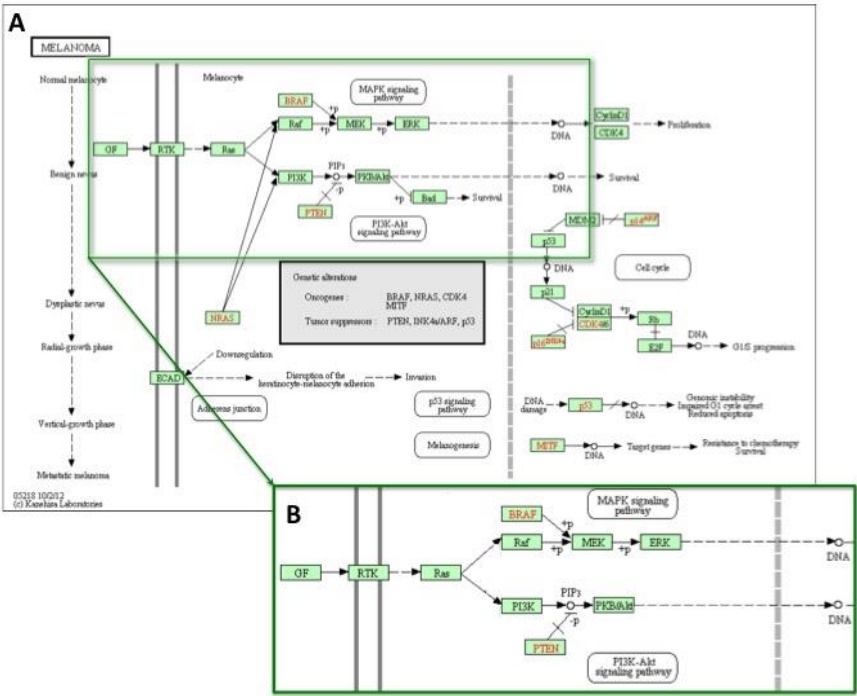


Figure 1: A: The melanoma KEGG pathways. B: inset representing the two specific pathways, the MAPK and PI3K signaling pathways, that we have considered to illustrate our proposal (from www.genome.jp/kegg/pathway/hsa/hsa05218.html).

Let us notice that ATP is indicated 6 times, that is, the exact number of times that this molecule occurs in the considered pathways. Thus, our theorem says:

$$(SCF, c\text{-Kit}, SRC, SHC1, Grb2, SOS, RAS, GTP, RAF, MEK, ATP, ATP, ATP, ATP, ATP, ATP) \vdash MAPK.$$

Up to now, we have identified the theorem representing the pathway we want to focus on. The real formal representation of the pathway from SCF to MAPK is given as soon as we provide the proof of such a theorem. In this proof all the necessary biochemical steps are represented by corresponding *Empirically*

Valid Formulas (EVF). Concerning this point, note that in *Zsyntax* there are two kinds of formulas: the *Empirically Valid Formulas* (EVF), representing real empirical biochemical processes, and the *Logically Valid Formulas* (LVF) indicating the valid inference rules allowing the move from one EVF to the successive one. Such a proof is shown in Table 1 of the Appendix.

Actually, it should be observed that Table 1 depicts only the abridged version of the demonstration of the theorem. It is a version in which there are only the steps involving the EVFs representing the reactions, but not those involving the LVFs. That is, in it there is no indication of the way by means of which we have moved from one EVF to another one. For space reason, we do not have provided here this complete (and longer) proof, but it could be obtained easily by a simple application of the rules governing the operators (Boniolo, Di Fiore, D'Agostino, 2010).

Another example of a conversion through the *Zsyntax* language may be given by considering the theorem representing a parallel biochemical pathway going from SCF to MITF, via PI3K. Again we begin describing the pathways in question by means of the biologists' usual ordinary language. In this case, we have a pathway in which PI3K and AKT plays a central role, especially concerning melanoma, since they regulate cell growth, survival, motility, and metabolism (Davies, 2012). PI3K is a heterodimer composed of a p110 catalytic subunit and a p85 regulatory subunit. Both subunits are responsible for the production of PIP2 and PIP3, whose function is to recruit signalling proteins to cell membranes and propagate the signal. Upon PI3K activation, both AKT and PDK1 bind PIP2 and PIP3 and translocate to the plasma membrane, where PDK1 phosphorylates AKT. The PI3K/AKT pathway mediates cell survival through MITF phosphorylation induced by MAPK. MITF phosphorylation stimulates both transactivation of the downstream effectors and degradation of the protein through the ubiquitin-proteasome pathway, depending on cell context (Wellbrock, 2008). Moreover the PI3K/AKT signalling regulates MITF protein levels in the cells through GSK activity (Terragni, 2011).

What is expressed above can be formally compacted into the theorem

$$(SCF, C) \vdash AKT,$$

where, as before, (SCF, C) is the initial aggregate (IA), and AKT is the final molecular outcome. In particular,

$$C (c\text{-Kit}, SRC, P85\text{-PI3K}, P110\text{-PI3K}, PIP2, PDK1, ATP, ATP, ATP, ATP, ATP).$$

Thus, the theorem is

$$(\text{SCF, c-Kit, SRC, P85-PI3K, P110-PI3K, PIP2, PDK1, ATP, ATP, ATP, ATP, ATP}) \vdash \text{AKT.}$$

Its proof, that is, the *Zsyntax* representation of the pathway from SCF to AKT, is given Table 2 of the Appendix. Again this is the abridged version. The complete demonstration, containing the LVFs, can be found in the Table 3.

2.2. Interactomes as Conjunctions of *Zsyntax* Theorems

At this point we have all we need to argue for our thesis, that is that any interactome can be represented by a conjunctions of *Zsyntax* theorems. Let us begin with a simple observation on the structure of an abstract network, say G , with n nodes and such that any node is linked to any other node by an edge, that is, a pathway. Trivially enough, G can be thought of as the union of all its possible pathways. In this way, so to say, we can first decompose G into its component pathways and, then, we can recompose it as their union.

Now let us consider a real (that is, an empirically grounded) interactome B composed of n nodes, some of which are connected by ordered edges (each one representing a real biochemical pathway). Needless to say, B can be considered as a sub-network of the abstract network G having the same number n of nodes but a lower number of pathways. Of course, even the empirical network B can be decomposed into all the actual pathways composing it.

Yet we know that an empirical pathway can be represented as a *Zsyntax* theorem. Therefore, the original empirical molecular biology network B can be, first, decomposed into a set of *Zsyntax* theorems and, then recomposed as the conjunction of all these theorems.

2.3. The Case of the Melanoma Network

To illustrate this outcome, let us come back to the melanoma KEGG network (see Figure 1) (Kanehisa, 2013). We will focus only on the MAPK and PI3K signalling pathways. These pathways regulate melanoma survival and proliferation through the fine-tuning of MITF expression in melanocytic cells. As shown in Figure 1A, these pathways are activated upon binding of appropriate ligands, the GF, to RTK. The binding generates an intracellular cascade of events, involving the phosphorylation of RTKs, the stimulation of RAS and the subsequent binding to the RAF (Roskoski, 2005). RAF proteins

stimulate MEK and MAPK phosphorylation, thus transmitting the proliferative signal to the nucleus.

By taking into consideration what previously said, this part of the melanoma network can be decomposed into a set of *Zsyntax* theorems and then recomposed as the their conjunction. Let us indicate by $\&$ the conjunction of theorems (each one representing a component pathway), and by C_i the complement of the initial aggregate starting from the given molecule M (that is, the general initial aggregate is $IA = (M, C_i)$). Thus, we have:

$$GF \vdash GF \ \& \ RTK \vdash RTK \ \& \ RAS \vdash RAS \ \& \ RAF \vdash RAF \ \& \ MEK \vdash MEK \ \& \ MAPK \vdash MAPK \ \& \ (GF, C_1) \vdash RTK \ \& \ (RTK, C_2) \vdash RAS \ \& \ (RAS, C_3) \vdash RAF \ \& \ (RAF, C_4) \vdash MEK \ \& \ (MEK, C_5) \vdash MAPK \ \& \ (GF, C_6) \vdash RAS \ \& \ (RTK, C_7) \vdash RAF \ \& \ (RAS, C_8) \vdash MEK \ \& \ (RAF, C_9) \vdash MAPK \ \& \ (GF, C_{10}) \vdash RAF \ \& \ (RTK, C_{11}) \vdash MEK \ \& \ (RAS, C_{12}) \vdash MAPK \ \& \ (GF, C_{13}) \vdash MEK \ \& \ (RTK, C_{14}) \vdash MAPK \ \& \ (GF, C_{15}) \vdash MAPK \ \& \dots^2$$

That is, this formula is the logical reconstruction of the complex interacting pathways of melanoma network under investigation. It is evident, at this point, how this process of logical reconstruction could be applied to the entire melanoma network. Moreover, since the melanoma network is just an instantiation of a general biological network, nothing hinders to conclude that any biological network could be decomposed in *Zsyntax* theorems and then recomposed, along the same formal avenue, as conjunctions of all the theorems involved.

3. Discussion

Cellular processes, such as survival, proliferation, differentiation and apoptosis, can be exploited through the correct and timely regulated activation of a series of intracellular events. These events are organized in pathways, which are frequently interconnected. Cells communicate with their extracellular environment through receptors, whose signalling is mediated through the binding of ligands that can be soluble factors, but also extracellular

² It is worth noting that, for the sake of the logical completeness, in the conjunction of the theorems representing all the pathways of the melanoma network we have had to take into account also all the identity theorems (e.g., $GF \vdash GF$; $RTK \vdash RTK$; $RAS \vdash RAS$; etc.). The theorems represent the pathways starting and arriving at the same molecules without any intermediate step.

matrix proteins, or integral membrane proteins on other cells. Stimulation of cell surface receptors induces a variety of cellular responses, depending on the type of receptor that gets activated, its functional association with other receptors, or the type of modification that the receptor bears as a consequence of its activation. In any case, binding of ligands to cell surface receptors activates a variety of intracellular signalling pathways responsible for defining how the cell reacts to external stimuli. Complex interactions among pathways are then set in place in the cell to modulate and integrate each pathway. Increased understanding on the components of these pathways and on the interactions among members of different pathways may shed light on their physiological functions. Analogously, comprehensive knowledge of the intracellular signalling pathways in normal and cancer tissues can give a grasp on how to best interfere with variable and under-characterized signalling pathways to identify novel drug targets. It is therefore of utmost importance to deeply understand which are the cellular cascades that are activated by specific receptors, and also which are the interconnections among pathways. To this scope, network biology offers a number of tools, which are extremely useful, even if with some (both semantic and syntactic) limitations. To try to overcome some of these limitations we constructed a computational language, *Zsyntax*, which allows the conversion of biological pathways into theorems (Boniolo, Di Fiore, D'Agostino, 2010; Boniolo, D'Agostino, Piazza, Pulcini, 2013; Boniolo, D'Agostino, Piazza, G. Pulcini, 2015).

In this paper we have shown how an entire network (in particular an interactome) can be, first, decomposed into a series of *Zsyntax* theorems and, then, logically reconstructed as a conjunction of those theorems.

This outcome has a series of fruitful consequences, of course as soon as we move from the theoretical proposal here discussed to its practical implementation. As it is evident, this could occur whenever the appropriate software is realized. In particular, we need two software programs: 1) one allowing the conversion from the biological information concerning a given pathway into the corresponding *Zsyntax* theorem, in the form exemplified in Tables 1 and 2 (if we want the abridged version) or in Table 3 (if we want the complete version); 2) one able of decomposing the given interactome into a series of *Zsyntax* theorems and, then rebuilding it as a conjunction of those theorems. It is important to note that all theorems that describe the proliferative pathway of melanocytic cells have been written manually by mining literature for detailed information on individual cellular pathways.

4. Philosophical Relevance

As seen, we have a tool, which is able to transform the biological ordinary language, by means of which nowadays a molecular pathway is described, into a rigorous computable logical language, formally representing the pathways as theorems in a mathematical sense. With this language we can integrate the information that we have at systemic level, where complexity is sovereign, with the information that we have at molecular level, where the usual simplicity due to an inevitable practical reductionism is at the core.

Starting from a pathway of an interactome, we are immediately able to have all its component biochemical steps, detailed at molecular level, as representation of steps which form the demonstration of the corresponding theorem. Moreover, since any interactome can be represented by a conjunction of *Zsyntax* theorems, we could, at least in principle, think the interactome as a conjunction of theorems, synthesising all the interactions present in a cell. It means that, with the help of *Zsyntax*, we could have that computable representation of a cell that many authors are trying to pursue (Karr et al. 2012; Isalan, 2012; Gunawardena, 2012).

Finally, *Zsyntax* could be also an answer to the biologists who are asking for powerful mathematical and formal tools to put together the vast knowledge that we have on molecular interactions and on their products in a friendly way (Rzhetsky, Seringhaus, Gerstein, 2008). Here, of course, is not the right place to discuss the advantages of a mathematical and formal representation of reality. It suffices to recall what were the progresses when science moved from the “monde de l’*à-peu-près* à l’univers de la *precision*” (Koyré, 1948).

We would conclude by highlighting how this tool is coming from the philosophical reflection on what biological complexity is and on how it could be treated in order to manage it, without any reductionist intent. Surely we do not have solved all the problems concerning complexity, but we have shown that we can, first, deconstruct a very complex building into its compounding bricks, then reasort these bricks into small modules according to their logical relationships, and, finally, reconstruct the original complex building by logically connecting those modules.

5. Appendix

Here we give the abridged demonstrations (Table 1) of the first theorem mentioned in the text. While of the second we give both the abridged (Table 2) and the complete (Table 3) demonstration. In the case of the complete demonstrations, it is interesting to note how the inference rules governing the introduction and the elimination of the operators are used (see right column) and how it is possible to move from an EVF to another EVF (in bold) until the thesis of the theorem is proved.

Theorem:	
SCF@c-Kit, SRC, SHC1, Grb2, SOS, RAS, GTP, RAF, MEK, MAPK, ATP, ATP, ATP, ATP, ATP, ATP MAPK@P	
1.	SCF@c-Kit → SCF@c-Kit
2.	SCF@c-Kit@ATP@ATP → SCF@c-Kit@P₅₆₈@P₅₇₀@ADP@ADP
3.	SCF@c-Kit@P₅₆₈@P₅₇₀@SRC → SCF@c-Kit@P₅₆₈@P₅₇₀@SRC
4.	SCF@c-Kit@P₅₆₈@P₅₇₀@SRC@ATP → SCF@c-Kit@P₅₆₈@P₅₇₀@(SRC@P@ADP)
5.	SCF@c-Kit@P₅₆₈@P₅₇₀@(SRC@P@SHC1) → SCF@c-Kit@P₅₆₈@P₅₇₀@(SRC@P@SHC1)
6.	SCF@c-Kit@P₅₆₈@P₅₇₀@(SRC@P@SHC1@ATP) → SCF@c-Kit@P₅₆₈@P₅₇₀@(SRC@P@SHC1@P@ADP)
7.	SCF@c-Kit@P₅₆₈@P₅₇₀@(SRC@P@SHC1@P@Grb2@SOS) → SCF@c-Kit@P₅₆₈@P₅₇₀@(SRC@P@SHC1@P@Grb2@SOS)
8.	SCF@c-Kit@P₅₆₈@P₅₇₀@(SRC@P@SHC1@P@Grb2@SOS@RAS@GDP@GTP) → SCF@c-Kit@P₅₆₈@P₅₇₀@(SRC@P@SHC1@P@Grb2@SOS@RAS@GTP@GDP)
9.	SCF@c-Kit@P₅₆₈@P₅₇₀@(SRC@P@SHC1@P@Grb2@SOS@RAS@GTP@GDP) → SCF@c-Kit@P₅₆₈@P₅₇₀@(SRC@P@SHC1@P@Grb2@SOS@RAS@GTP@GDP)
10.	GTP@RAS@RAF@MEK → GTP@RAS@RAF@MEK
11.	GTP@RAS@RAF@MEK@ATP → GTP@RAS@RAF@MEK@P@ADP
12.	GTP@RAS@RAF@MEK@P → GTP@RAS@RAF@MEK@P
13.	RAF@MEK@P@MAPK → RAF@MEK@P@MAPK
14.	RAF@MEK@P@MAPK@ATP → RAF@MEK@P@MAPK@P@ADP
15.	MAPK@P

Table 1: Abridged proof the theorem: (SCF, c-Kit, SRC, SHC1, Grb2, SOS, RAS, GDP, GTP, RAF, MEK, MAPK, ATP, ATP, ATP, ATP, ATP, ATP) | MAPK@P.

By \odot we have indicated the biochemical interactions; by \otimes the union of the molecules in questions; by \rightarrow the paths between molecules. P is the phosphate group (whether we want to indicate the site to which its binds we indicate the site number as subscript; e.g. P₅₆₈ and P₅₇₀). By means of the brackets, we have indicated, if useful, what binds what. Note that during the deduction of the final compound, molecules not belonging the initial aggregate are considered (for example, ADP - adenosine diphosphate - and GDP - guanosine diphosphate). Of course, they are the outcomes of reactions occurring between the initial aggregate and the final result.

Theorem:	
$[SCF, c\text{-Kit}, P85\text{-PI3K}, P110\text{-PI3K}, PIP2, AKT, PDK1, ATP, ATP, ATP, ATP, ATP] \vdash AKT \odot P \odot P$	
1.	$SCF \otimes c\text{-Kit} \rightarrow SCF \odot c\text{-Kit}$
2.	$SCF \odot c\text{-Kit} \otimes ATP \rightarrow SCF \odot (c\text{-Kit} \odot P_{121}) \otimes ADP$
3.	$SCF \odot (c\text{-Kit} \odot P_{121}) \otimes P85\text{-PI3K} \rightarrow SCF \odot (c\text{-Kit} \odot P_{121}) \odot P85\text{-PI3K}$
4.	$SCF \odot (c\text{-Kit} \odot P_{121}) \odot P85\text{-PI3K} \otimes P110\text{-PI3K} \rightarrow SCF \odot (c\text{-Kit} \odot P_{121}) \odot (P85\text{-PI3K} \odot P110\text{-PI3K})$
5.	$SCF \odot (c\text{-Kit} \odot P_{121}) \odot (P85\text{-PI3K} \odot P110\text{-PI3K}) \otimes ATP \rightarrow$ $SCF \odot (c\text{-Kit} \odot P_{121}) \otimes (P85\text{-PI3K} \odot P110\text{-PI3K} \odot P) \otimes ADP$
6.	$P85\text{-PI3K} \odot P110\text{-PI3K} \odot P \otimes PIP2 \rightarrow (P85\text{-PI3K} \odot P110\text{-PI3K} \odot P) \odot PIP2$
7.	$(P85\text{-PI3K} \odot P110\text{-PI3K} \odot P) \odot PIP2 \otimes ATP \rightarrow (P85\text{-PI3K} \odot P110\text{-PI3K} \odot P) \otimes PIP3 \otimes ADP$
8.	$PIP3 \otimes AKT \rightarrow PIP3 \odot AKT$
9.	$PIP3 \odot AKT \otimes PDK1 \rightarrow PIP3 \odot AKT \odot PDK1$
10.	$PIP3 \odot PDK1 \odot AKT \otimes ATP \rightarrow PIP3 \otimes AKT \odot P \odot PDK1 \otimes ADP$
11.	$AKT \odot P \otimes ATP \rightarrow AKT \odot P \otimes P \otimes ADP$
12.	$AKT \odot P \odot P$

Table 2: Abridged proof the theorem: $(SCF, c\text{-Kit}, P85\text{-PI3K}, P110\text{-PI3K}, PIP2, AKT, PDK1, ATP, ATP, ATP, ATP, ATP) \vdash AKT \odot P \odot P$.³

³ By means of the brackets, we have indicated, if useful, what binds what.

Theorem:		
SCF _c -Kit,P85-P13K,P110-P13K,PIP2,AKT,PKD1,ATP,ATP,ATP,ATP,ATP ⊢ AKT⊙P⊙P		
Proof's steps		Inferential rule
1. SCF _c -Kit,P85-P13K,P110-P13K,PIP2,AKT,PKD1,ATP,ATP,ATP,ATP,ATP		Initial Aggregate
2. SCF _c -Kit		From 1 by ⊙E
3. P85-P13K		From 1 by ⊙E
4. P110-P13K		From 1 by ⊙E
5. PIP2		From 1 by ⊙E
6. AKT		From 1 by ⊙E
7. PDK1		From 1 by ⊙E
8. ATP		From 1 by ⊙E
9. ATP		From 1 by ⊙E
10. ATP		From 1 by ⊙E
11. ATP		From 1 by ⊙E
12. ATP		From 1 by ⊙E
13. SCF _c -Kit → SCF _c -Kit		EVF
14. SCF _c -Kit		From 2,13 by →E
15. SCF _c -Kit⊙ATP		From 8,14 by ⊙I
16. SCF _c -Kit⊙ATP → SCF _c -Kit⊙(c-Kit⊙P ₁₁₀)⊙ADP		EVF
17. SCF _c -Kit⊙(c-Kit⊙P ₁₁₀)⊙ADP		From 15,16 by →E
18. SCF _c -Kit⊙P ₁₁₀		From 17 by ⊙E
19. SCF _c -Kit⊙(c-Kit⊙P ₁₁₀)⊙P85-P13K		From 3,18 by ⊙I
20. SCF _c -Kit⊙(c-Kit⊙P ₁₁₀)⊙P85-P13K → SCF _c -Kit⊙(c-Kit⊙P ₁₁₀)⊙P85-P13K		EVF
21. SCF _c -Kit⊙(c-Kit⊙P ₁₁₀)⊙P85-P13K		From 19,20 by →E
22. SCF _c -Kit⊙(c-Kit⊙P ₁₁₀)⊙P85-P13K⊙P110-P13K		From 4,21 by ⊙I
23. SCF _c -Kit⊙(c-Kit⊙P ₁₁₀)⊙P85-P13K⊙P110-P13K → SCF _c -Kit⊙(c-Kit⊙P ₁₁₀)⊙(P85-P13K⊙P110-P13K)		EVF
24. SCF _c -Kit⊙(c-Kit⊙P ₁₁₀)⊙(P85-P13K⊙P110-P13K)		From 22,23 by →E
25. SCF _c -Kit⊙(c-Kit⊙P ₁₁₀)⊙(P85-P13K⊙P110-P13K)⊙ATP		From 9,24 by ⊙I
26. SCF _c -Kit⊙(c-Kit⊙P ₁₁₀)⊙(P85-P13K⊙P110-P13K)⊙ATP → SCF _c -Kit⊙(c-Kit⊙P ₁₁₀)⊙(P85-P13K⊙P110-P13K)⊙ADP		EVF
27. SCF _c -Kit⊙(c-Kit⊙P ₁₁₀)⊙(P85-P13K⊙P110-P13K)⊙ADP		From 25,26 by →E
28. P85-P13K⊙P110-P13K⊙P		From 27 by ⊙E
29. P85-P13K⊙P110-P13K⊙P⊙PIP2		From 5,28 by ⊙I
30. P85-P13K⊙P110-P13K⊙P⊙PIP2 → (P85-P13K⊙P110-P13K⊙P)⊙PIP2		EVF
31. (P85-P13K⊙P110-P13K⊙P)⊙PIP2		From 29,30 by →E
32. (P85-P13K⊙P110-P13K⊙P)⊙PIP2⊙ATP		From 10,31 by ⊙I
33. (P85-P13K⊙P110-P13K⊙P)⊙PIP2⊙ATP → (P85-P13K⊙P110-P13K⊙P)⊙PIP3⊙ADP		EVF
34. (P85-P13K⊙P110-P13K⊙P)⊙PIP3⊙ADP		From 32,33 by →E
35. PIP3		From 34 by ⊙E
36. PIP3⊙AKT		From 6,35 by ⊙I
37. PIP3⊙AKT → PIP3⊙AKT		EVF
38. PIP3⊙AKT		From 36,37 by →E
39. PIP3⊙AKT⊙PDK1		From 7,38 by ⊙I
40. PIP3⊙AKT⊙PDK1 → PIP3⊙AKT⊙PDK1		EVF
41. PIP3⊙AKT⊙PDK1		From 39,40 by →E
42. PIP3⊙PDK1⊙AKT⊙ATP		From 11,41 by ⊙I
43. PIP3⊙PDK1⊙AKT⊙ATP → PIP3⊙AKT⊙P⊙PDK1⊙ADP		EVF
44. PIP3⊙AKT⊙P⊙PDK1⊙ADP		From 42,43 by →E
45. AKT⊙P		From 44 by ⊙E
46. AKT⊙P⊙ATP		From 12,45 by ⊙I
47. AKT⊙P⊙ATP → AKT⊙P⊙ADP		EVF
48. AKT⊙P⊙ADP		From 46,47 by →E
49. AKT⊙P⊙P		From 48 by ⊙E
SCF, c-Kit,SRC,P85-P13K,P110-P13K,PIP2,AKT,PKD1,ATP,ATP,ATP,ATP,ATP → Q.E.D	AKT⊙P⊙P	From 1,49 by →I

Table 3: Complete proof the theorem:(SCF,c-Kit,P85-P13K,P110-P13K,PIP2,AKT,PKD1,ATP,ATP,ATP,ATP,ATP) ⊢ AKT⊙P⊙P.

It highlights all the logical steps, and the relative *Logically Valid Formulas* (LVR), which permits to move from the initial aggregate (the hypothesis) to the final aggregate (the thesis to be proved). In bold the *Empirically Valid Formulas* (EVF) which are present in the abridged version. By means of the brackets, we have indicated, if useful, what binds what. In our case the LVR are the *Elimination of the conditional* (\rightarrow E): if $A \rightarrow B$ is true and we have A, then we have B; the *Introduction of the conditional* (\rightarrow I): if B can be derived from $C \otimes A$, then $A \rightarrow B$ can be derived from C alone; the *Elimination of conjunction* (\otimes E): if the conjunction of A and B ($A \otimes B$) can be derived from C, then both A and B individually can be derived from C; the *Introduction of conjunction* (\otimes I): if A can be derived from C, and B can be derived from D, then the conjunction of A and B can be derived from $C \otimes D$.

REFERENCES

- Aoki-Kinoshita KF, Kanehisa M (2007) Gene annotation and pathway mapping in KEGG. *Methods in molecular biology* 396: 71-91.
- Boniolo G, D'Agostino M, Piazza M, Pulcini G (2012) A Logic of Non-Monotonic Interactions, *Journal of Applied Logic* 11: 52-62.
- Boniolo G, D'Agostino M, Piazza M, Pulcini G (2015) Adding Logic to the Toolbox of Molecular Biology", *European Journal for the Philosophy of Science* 5: 399-417
- Boniolo G, Di Fiore PP, D'Agostino M (2010) Zsyntax: a formal language for molecular biology with projected applications in text mining and biological prediction, *PLOsONE*, 5(3): e9511;
- Bruck J, Ebenhoh O, Heinrich R (2006) Patterns of interactions of reaction pairs in metabolic networks. *Genome Inform* 17: 208-218.
- Cho DY, Kim YA, Przytycka TM (2012) Chapter 5: Network Biology Approach to Complex Diseases. *PLoS Comput Biol* 8(12): e1002820.
- Davies MA (2012) The role of the PI3K-AKT pathway in melanoma. *Cancer J*. 18: 142-7.

- Faeder JR, Blinov ML, Hlavacek WS (2005) Graphical rule-based representation of signal-transduction networks. In: Liebrock LM Editor. ACM Press, New York, NY. pp. 133-140.
- Gunawardena J. (2012) Silicon dreams of cells into symbols. *Nature Biotechnology* 30: 838-40.
- Isalan M (2012) Systems biology: A cell in a computer. *Nature* 488: 40-1.
- Kanehisa M (2013) Molecular network analysis of diseases and drugs in KEGG. *Methods in molecular biology* 939: 263-275.
- Kanehisa M, et al. (2004) The KEGG resource for deciphering the genome. *Nucleic acids research* 32: D277-280.
- Kanehisa M, et al. (2006) From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Research* 34: D354-357.
- Karr JR, et al. (2012) A Whole-Cell Computational Model Predicts Phenotype from Genotype. *Cell* 150: 389-401.
- Kitano H, Funahashi A, Matsuoka Y, Oda K (2005) Using process diagrams for the graphical representation of biological networks. *Nat Biotechnol* 23: 961-966.
- Kohn KW (1999) Molecular interaction map of the mammalian cell cycle control and DNA repair systems. *Mol Biol Cell* 10: 2703-2734.
- Koyré A (1948) Du monde de l' "à-peu-près" à l'univers de la précision. In Koyré A, *Études d'histoire de la pensée philosophique*. Paris: Gallimard 1971.
- Lennartsson J, Ronnstrand L (2012) Stem cell factor receptor/c-Kit: from basic science to clinical implications. *Physiological reviews* 92: 1619-1649. doi: 10.1152/physrev.00046.2011.
- Lima-Mendez G, van Helden J (2009) The powerful law of the power law and other myths in network biology. *Molecular bioSystems* 5: 1482-1493.
- Masson K, Ronnstrand L (2009) Oncogenic signaling from the hematopoietic growth factor receptors c-Kit and Flt3. *Cellular signalling* 21: 1717-1726. doi: 10.1016/j.cellsig.2009.06.002.

- Phung B, et al. (2011) C-KIT signaling depends on microphthalmia-associated transcription factor for effects on cell proliferation. *PLoS one* 6: e24064. doi: 10.1371/journal.pone.0024064.
- Qi Y, Ge H (2006) Modularity and dynamics of cellular networks. *PLoS computational biology* 2: e174.
- Ronnstrand L (2004) Signal transduction via the stem cell factor receptor/c-Kit. *Cellular and molecular life sciences : CMLS* 61: 2535-2548.
- Roskoski R Jr. (2005) Signaling by Kit protein-tyrosine kinase—the stem cell factor receptor. *Biochem Biophys Res Commun.* 337: 1-13.
- Rzhetsky A, Seringhaus M, Gerstein M (2008) Seeking a new biology through text mining. *Cell* 134: 9-13.
- Samayawardhena LA, Pallen CJ (2008) Protein-tyrosine phosphatase alpha regulates stem cell factor-dependent c-Kit activation and migration of mast cells. *The Journal of biological chemistry* 283: 29175-29185. doi: 10.1074/jbc.M804077200.
- Terragni J, et al. (2011) The E-box binding factors Max/Mnt, MITF, and USF1 act coordinately with FoxO to regulate expression of proapoptotic and cell cycle control genes by phosphatidylinositol 3-kinase/Akt/glycogen synthase kinase 3 signaling. *The Journal of Biological Chemistry* 286: 36215-36227.
- Vance KW, Goding CR (2004) The transcription network regulating melanocyte development and melanoma. *Pigment Cell Research* 17: 318-325.
- Vidal M, Cusick ME, Barabási AL (2011) Interactome Networks and Human Disease. *Cell* 144: 986-98.
- Wellbrock C, et al. (2008) Oncogenic BRAF regulates melanoma proliferation through the lineage specific factor MITF. *PLoS one* 3: e2734.
- Wuang JL, et al. (2014) Identification of Human Disease Genes from Interactome Network Using Graphlet Interaction. *PLoS ONE* 9: e86142.

