

A general framework dedicated to computational morphogenesis Part II – Knowledge representation and architecture[☆]

Pridi Siregar^{a,*}, Nathalie Julen^a, Peter Hufnagl^b, George Mutter^c

^a Integrative BioComputing - IBC, Chantepie, France

^b Department of Digital Pathology and IT, Institute of Pathology, Charité – Universitätsmedizin Berlin, Berlin, Germany

^c Department of Pathology, Harvard Medical School and Brigham and Women's Hospital, Boston, MA, USA

ARTICLE INFO

Keywords:

Computational morphogenesis
Knowledge representation
Inference
Agents

ABSTRACT

In our previous paper we introduced morphogenesis and post-embryonic life as arising from cells interacting via coupled chemical, electrical and mechanical processes occurring across multiple organization levels. We reviewed these processes from the perspectives of developmental biology and how they relate to physics-based constitutive equations that are well suited to model intercellular interactions' fields. In this paper we will describe a knowledge representation and architectural design strategy that can organize and encode the biochemical, biological and biophysical data necessary to represent and model the highly specialized and diversified cells that constitute living tissues. Since there are about 200 different types of cells in mammalian tissues, a huge amount of molecular, cellular and tissue data must be accounted for. This data cannot be incorporated in an *ad hoc* manner but, on the contrary, must be organized according to some sound principles. We give an overview of these principles and describe how they can be incorporated as proper features of a Knowledge Base System (KBS) dedicated to computational morphogenesis (CM).

1. Introduction

Cells communicate via multimodal fields that can be modeled by physics-based constitutive equations (Siregar et al., 2003, 2018a). However, such equations are ill-suited to capture the reactivity and goal-driven behavior of living cells and organisms in general. In response to external cues, cells self-organize to form the intricate architectures of living tissues (Wolpert et al., 2002). There are thus two separate issues to consider in CM. One is modeling the multimodal interaction fields that, in our approach, are continuous. The other is modeling the response and self-organization of interacting cells as discrete agents that interpret and adapt their behavior in response to these fields. In addition, each individual cell is also a field-generating source and thus contributes to modify the environments in which it is embedded (Costantini and Kopan, 2010; Wolpert et al., 2002). Thus causality is simultaneously top-down, from the environment to the cell and its genome, and bottom-up, from the genome to the environment (Siregar, 2009; Siregar et al., 2018a). In short, there is no privileged level of causation (Noble, 2012). The pertinence of this view is not limited to living organisms but to any complex system where global properties have an impact on the system's constituents and vice versa.

The general view that nature as a whole is constituted by nested communicating entities draws its justification from a fundamental perspective –the Standard Model- of particle physics where matter particles (fermions) exchange information via Gauge bosons (Phillips, 2003) which are commonly viewed as the force carriers that are responsible for all dynamic changes of matter. The nature of the fermions and Gauge bosons determines which of the four known fundamental forces are impacting the fermions. These forces are (in order of strength) the strong nuclear force, electromagnetism, weak nuclear force and gravitation. They range across the very-short distances (1 fm, or 1.0×10^{-15} m for the strong nuclear force) to infinity for gravitational and electromagnetic force (Phillips, 2003). Hence a view according to which cells, tissues, organisms and societies of organisms are nested agents interacting with each other via multimodal fields that are efficient over different characteristic distances is a generalization, in the biological domain, of the “fermions-Boson” dynamics. Since the incredibly complex evolving and emerging forms observed at virtually all space-scales in the Universe are underpinned by simple fundamental interactions, it seems reasonable to posit that mesoscopic and macroscopic scale phenomena of complex systems could follow a similar pattern. For instance, a single ant is in itself, a complex system of

[☆] This is Part 2 of a 2-part series.

* Corresponding author at: Integrative BioComputing IBC, Centre d'Affaires Buro Club, Place du Granier, 35135 Chantepie, France.

E-mail address: pidi.siregar@ibiocomputing.com (P. Siregar).

interacting cells. Ants are social insects that form colonies and interact over short-distances, from physical contact and mutual antennation, and over medium and long-range distances via pheromone fields that evaporate at different rates (Jackson and Ratnieks, 2006). Emergent global features such as pheromone trails result from the actions of individual ants (bottom-up causation). Conversely, each ant's behavior is influenced by the organization of the trails that can provide positive and negative feedbacks to organize foraging at the colony level (Jackson and Ratnieks, 2006) (top-down causation). The same holds for any society of living organisms including humans where each individual impacts the society as a whole and vice versa. Morphogenesis follows the same causal pattern and the aim of Knowledge-base system (KBS) design in CM is to capture the unifying principles that underlie all multi-scale ecosystems while, at the same time, account for the specifics that distinguish organisms and their constituents. This poses a few challenges. First, biological sciences cover many knowledge-intensive domains such as organic chemistry, biochemistry, molecular biology, cell biology, anatomy and physiology. The amount of facts, ranging from molecular to ecosystems is simply daunting and cannot be incorporated within a KBS in an *ad hoc* manner. Second, any knowledge that will instruct generic computational models must be machine readable and interpretable, i.e., the facts must be well organized and machine-friendly. Third, inference mechanisms are necessary for AI agents to draw significant information from uncertain facts, and to develop plausible computational models of the target system. In addition, the "data gaps" must be circumvented by a fair amount of reasoning that must be robust to uncertainty. This situation has a direct impact on modeling. The generation of 3D *structurally* and *functionally* plausible models of, say, normal and abnormal tissues of endometrium, breast, colon, or brain tissues from virtual stem cells is not a simple task. In the following sections we will describe how we addressed these challenges by providing a general overview on what we believe are important aspects of knowledge representation, inference engines, and generating programs.

2. Knowledge representation

Knowledge representation is a complex and hard subject that is meant to capture in an articulate way our prior knowledge about the world. There are two complementary facets to knowledge representation: content and a language in which to express it. Ontologies deal with content while machine readability and intelligibility relate to language. We first will deal with content and will introduce language latter-on.

2.1. The need for a core ontology in biology

The life and medical sciences differ from that of physics and engineering. The physics world-view is *parsimonious*, *consistent* and at the same time aims at *completeness*. It is parsimonious insofar as classical and modern theories in physics are built around a few principles - such as mass, charge and momentum conservation- and fundamental equations that are derived from these principles are quite few (Einstein, 1920; Phillips, 2003). Physics is consistent because physical theories do not contradict their axioms that invariably include conservation laws. Physics aims at completeness because new theories explain more facts than older ones and thus gain in generality. The life and medical sciences have a different trajectory. The bulk of data and factual knowledge in biology and medicine is huge, increasing, fragmented, incomplete, uncertain and sometimes contradictory and has led to data and knowledge silos. Fortunately, data analytics have become very popular and widespread and play a critical role in making sense out of the data (LeCun et al., 2015; Sharma et al., 2015; Efron and Hastie, 2016). Unfortunately, biological theories from which computational models are built only play a small part in biology. In fact, since the advent of data analytics they have atrophied even further since today's mantra is to "let the data speak for itself". One can even get the

impression that the introduction of prior knowledge in any data-driven inference is like introducing a virus that carries the "curse of bias". This begs the question as to why research in biology should be so different from that of, say, physics to the point that prior knowledge becomes a liability rather than an asset. There is a strong unity in the universe and, until proven otherwise, life is part of it. Research for this unity in the life sciences could lead to the development of theories and computational frameworks that will efficiently support experimental research, data-mining and data analytics. That is, the synergy between theories and experimental research becomes a continuing hypothetico-deductive loop of scientific inquiry (Popper, 1963). Prior knowledge precedes theory-formation, and fortunately, large efforts have been devoted to the organization of the fragmented facts in the form of ontologies (Ashburner et al., 2000; Bodenreider and Stevens, 2006; Da Silveira et al., 2015). Current researches on biological and medical ontologies provide support to solving real problems such as clinical diagnosis. One of the difficulties encountered however is that there are an infinite number of possible reifications of world objects. Domain ontologies formalize the multiple world-views but do not reduce their numbers. Foundational work have been carried-out in the form of upper ontologies (Smith, 2003; Smith et al., 2003; Spear, 2016), but for our own purpose, we need to extend upper ontologies in order to design a KBS that will provide efficient guiding in modeling, experimental and data-mining. The question is where do we go from here? One possible solution is to look for core concepts that are both general and rich enough to be contextualized in many specialized domains. We define this corpus of knowledge as *core ontologies* designed to extend the scope of upper ontologies.

We expect core ontologies to provide support to the following goals.

Parsimony. Core concepts can play the role of an initial theory (axioms) from which new statements can be inferred. A theory is parsimonious if the axiom set is small and its extension of derived facts (theorems) is large.

Coherence. If the inference mechanisms are sound (e.g. logical deduction) then all new derived statements will be consistent with the initial theory and the coherency of its extension will be guaranteed.

Generalization and exploration. Core ontologies can help establish similarities in distinct subject matters by abstracting invariant relational patterns that can be re-contextualized in many domains and applications. This possibility to decontextualize and recontextualize rests at the root of analogical reasoning holds one of the keys to exploring new domains in a powerful way.

Unity and conceptual scaffolding. Core ontologies provide conceptual scaffolds that can help in the development of a unified theory in biology. From a KBS design perspective they can be translated into generic knowledge structures that constitute a prerequisite for designing general inference and problem-solving schemes.

2.2. A few elements of our core ontology¹

2.2.1. General scientific categories

Our framework was designed around general concepts borrowed from physics, chemistry, cybernetics, complex systems, and mathematics into our core ontology (Siregar, 2009). Physics provides the fundamental physical units (m, kg, sec, Amp) and formulates how they relate to each other in all quantified measures and processes (Phillips, 2003). Chemistry provides the conceptual framework to capture matter transformations and their relation to physics (McQuarie and Simon, 1997). Cybernetics provides the language to describe control and communication in artificial and natural systems (Wiener, 1948). Complex systems cover quite rich and distinct research topics including the self-organization within swarms (Jackson and Ratnieks, 2006),

¹ In this section, words that start with an uppercase letter designate categories (i.e. sets).

dynamic stability, dynamic coupling, and the study of emergence (Bocarra, 2009). For us, the starting point is physics (that includes but is not limited to Newtonian physics) since we posit that all natural processes are underpinned by physical laws even if from a phenomenological point of view, not all observables are reducible to physics. As for mathematics, they provide the common abstract language of most scientific disciplines, and constitute one of the focal points of our core ontology. Like some upper ontologies, the most abstract concept in our core ontology is the Object category that contains two subcategories: Physical Entity and Concept. All objects have properties and a Property is either an Intrinsic Property or a Relative Property. Relative properties are defined by relations between objects. A property can be a Qualitative Property or a Quantitative Property. Quantitative properties are defined in terms of numerical types such as Scalar, Vector, or Matrix all belonging to the Quantity type. Physical Entity includes Space, Time, Material Object, Physical System and Field. Material objects interact with each other via one or more Field and a Physical System is a portion of the universe composed of one or more material objects that have properties such as mass and charge. Physical systems have Energy that can be partitioned into different categories including Potential Energy and Kinetic Energy. Energy is related to Force and Motion by Physical Laws. A physical system's dynamics can be described as a succession of events that are localized in time and space. Subcategories of Event are Process and State. Processes are events that hold over a space-time interval and states are events that hold over a space interval and a time-instant. Time ordered State instances define a Trajectory in the Phase Space in a Dynamical System. States are partitioned into two sub-categories: Observable State and Non Observable State. By "observable" we mean sensed data through a Natural Sensor (e.g. visual) or via a Device Sensor. Any Process can formally be represented in terms of domain categories:

- Physics categories such that Force, Motion, or Energy
- Cybernetics categories that include Signal, Information, Feedback Loop, and Control
- Chemical categories such as Reaction, Reactant, Product, Catalyst, and Reaction Type
- Complex systems categories such as Network, Node, Link, Critical Point, Emergence
- Biological categories such as Development, Growth, and Homeostasis

The key to high-level interoperability lies in the overlapping of categorical domains and the explicit relations that can be expressed from them. For instance, a Physical System is a Network in complex systems. An instance of a Network is composed of instances of Node Object that interact via instances of Link. A Node Object can be a Material Object and a Link may be a Physical Process such as the binding of a ligand to a receptor. An Emergent Property is a property that holds for a Network but not for its Node instances. Since a node can designate any Material Object one can infer that emergence can be observed in physical, chemical, biological and social entities among many others. The formalization of complex systems must include processes that can be expressed in terms of the following categories:

- Mass Transport, Transformation and Storage
- Charge Transport, Transformation and Storage
- Energy Transport, Transformation and Storage

But also in terms of

- Information Transport, Transformation and Storage
- Control and communication

These process categories are of course related since mass and energy are fundamentally equivalent (Einstein, 1920), and control and

communication mechanisms are substantiated by physical ones. An open question is the ontological status of information as a fundamental category. Interestingly, some conceptual frameworks suggest that natural dynamical systems actually perform calculations which would entail equivalency and perhaps pre-eminence of information over matter/energy (Dodig-Crnkovic, 2011, 2012). Any of the above upper-level categories can of course be partitioned into subcategories, but the general idea is to develop a core ontology that remains as general as possible while rich enough to perform non-trivial inferences and computation. For instance, in our previous paper, we stepped down the ladder of generality to identify the cardinal concepts of our core biology ontology (Siregar et al., 2018a). What is noteworthy is the fact that there is always a possibility to map specific domain knowledge to core concepts and the type of mapping depends on the context. For instance, phosphorylation can be abstracted as a Molecular Transfer process (a phosphate group from a donor to an acceptor), an Energy Bond Transfer process (high energy bond from a donor to an acceptor), or both. Likewise, photosynthesis can be studied as Chemical Transformation process (conversion of carbon dioxide and water into glucose and oxygen), an Energy Transformation event (conversion of light energy to chemical energy), or both. The interconversion of potential energy and kinetic energy in a frictionless swinging pendulum can be abstracted as a cyclic potential to kinetic energy transformation process. Finally, the Concept category includes the Mathematics, Physical Law and Biology categories. Once categories are defined, it is the relations between them that constitute the real knowledge. For instance a physical law expressing a variational principle (e.g. light propagation in a gravitational field) can be mapped into the Self-Adjoint Operator category of Mathematics (Lanczos, 1986). In CM, the constitutive equations occupy the central-stage (Siregar et al., 2018a), and most of the above-mentioned core concepts have one or more mathematical objects related to them (ref. Sect 5.3.2).

2.2.2. Teleology

The formal modeling of natural systems inevitably leads to the questions of change and causality. The Greek philosopher Aristotle (384-322 BCE) defined four types of causes (Gotthelf, 2012; Hennig, 2009): the material, efficient, formal and the final cause. The four causes, also known as modes of *explanation* models, answer the following types of questions: What is it made of? What instigated the creation/change of it? What is its form or structure? And final causation answers the question "what is it for" or "why"? Material and formal causes explain the static features of things, while efficient and final causes explain their dynamics. An intuitive illustration of what these causes may be in morphogenesis is as follows. The material cause includes the molecular and cellular elements of the undifferentiated mesenchyme (the initial conditions); the efficient cause includes biochemical and physical laws that drive the morphogenetic processes; the formal cause specifies how, when and where the biochemical and physical laws apply; and the final cause could be the grown differentiated organism that also serves as explanation of its own specifications. Aristotelian philosophy permeates functional analysis that plays an important role in contemporary engineering and the social sciences (Miller, 2017). Human artefacts and organizations have a purpose hence the question prior to any realization is why or what for? The "how" comes after and follows from the roles of each constituent and their normatively defined interactions. Biological systems are usually studied and explained using the same pattern of thought. For instance, the heart's function is to pump the blood and that of a kidney is to filter the blood. At smaller scales, myocytes are parts of the heart that contribute to the pumping activity of the heart, and at still smaller scale, intracellular metabolic functions insure that the myocytes' energy needs are satisfied. Physiological textbooks describe living organisms as integrated systems where organs have functions that must satisfy the organisms' end purposes (e.g. continued existence). Although the concept of purpose is still natural in every day discourse in biology, it has

largely fallen out of favor as a source of scientific explanation since Newton and Darwin. The Newtonian world is mechanistic and Darwin's theory of evolution based on chance and natural selection nullifies any notion of purpose. Specific reasons for rejecting teleology include backward causation wherein the effect would precede the cause, a return to vitalism, intentionality, Intelligent Design, and anthropomorphic interpretations of otherwise blind processes (Ribiero et al., 2015). Refutation of teleology is multiform and mirrors the diversity of its interpretations. Even today, opinions diverge as to whether teleology is a feature of things or a kind of discourse about things. At one end, teleology is an integral part of living organisms because they are closed to efficient causation (Louie, 2009; Rosen, 1985, 1991). That is, all constituents having the status of efficient causes are causally entailed by their own effects. This state of affairs is possible if causality forms closed paths (Louie, 2009; Rosen, 1985, 1991) and as such, Aristotle's final cause can be incorporated as an element of these paths (Igamberdiev, 2015). At the other end of the spectrum, some argue that even using the term *biological function* as a figure of speech (i.e., excluding any real scientific value) should be banned from the biological vocabulary (Bachelard, 2002; Ribiero et al., 2015). The authors cite evolutionary biologist Ernst Mayr that proposed to dismiss teleology and replace it by "teleonomy" (Mayr, 1976; Ribiero et al., 2015). In teleonomy, goals or ends are totally determined by the execution of a program (i.e. DNA). Since DNA is considered a product of chance and natural selection, teleonomy thus defined is consistent with the prevailing Newtonian mechanistic view. Mayr also proposes to use the term "biological role" instead of "function" (Ribiero et al., 2015). Setting aside the possibility that, in the latter suggestion, syntactic substitution rather than true semantic shift is at play, Mayr's teleonomy entails that what organisms are, and do, is solely driven by their genes. In other words, organisms are epiphenomenon of their genes with no autonomous causal work of their own to do (Nicholson, 2014; Shields, 2017). This raises ontological and/or epistemological issues insofar as the goal-directed quality of consciousness also becomes an epiphenomenon of clock-like Pavlovian mechanisms and thus constitutes an illusion. It is not our goal here to debate on hard scientific and philosophical issues but to explain why and how the GMSP was designed as a goal-directed teleological system (Siregar, 2009). First, we hold for true that living systems are anticipatory, goal-directed agents whose behavior goes beyond pure reactivity as can be argued for instance in robotics (Nasuto and Hayashi, 2016) and psychology (Bickhard, 2016). Second, sound arguments demonstrate that organisms are inherently normative because they are teleological systems in which end-states serve as "unifying principles for normatively characterized subordinate processes" (Shields, 2017). Cybernetics systems provide a good example of normative and purposeful systems. The main control mechanism consists of negative feedback loops that maintain such systems operating within set points in the presence of external disturbances (Wiener, 1948). In human thought, there is no such thing as a "malfunctioning planet" or a "dysfunctional rock". In biology and medicine it is virtually impossible to describe an organ without referring to its role(s), nor refer to disorders and diseases without the notion of *norms* against which proper function can be assessed (Spear, 2016; Shields, 2017). A *rapportement* can be made with the concept of self-determination in which the effects of an organism's activities contribute to determine the conditions of its own existence (Mossio et al., 2013). Finally, studies show that both genes and epigenetic factors play a role in adaptation, inheritance, and development (Perri et al., 2017; Osborne, 2017) thus re-introducing organisms -not just genes- as part of the explanation of liveliness. Causality (i.e. efficient cause) then becomes bottom-up and top-down where the parts drive the whole and the whole constrains the parts as can readily be observed during real (Gilmour et al., 2017) and *in virtuo* morphogenesis (Siregar, 2009; Siregar et al., 2018a). The closure of efficient causes with no preferred direction gives the quality of *wholeness* and *unity* to organisms (Noble, 2012; Shields, 2017) which abiotic matter seems to lack. From an ontological point of view one may

wonder how closure of causation and norm-satisfying mechanisms such as homeostasis, homeoresis (Waddington, 1968) or autopoiesis (Maturana and Varela, 1980) naturally came into being. Could it be an evolved and complex extension of Le Chatelier's principle? Simple reversible chemical reactions, when perturbed (e.g. by adding one of the reactants), "adjust" the direction of the reactions so that the system always evolves towards equilibrium (McQuarie and Simon, 1997). In Le Chatelier's principle, the "norm" is defined by the equilibrium constant and like most of chemistry, is explained by the law of mass action. Since organisms have evolved as a result of natural selection, so did the processes that allowed their development and maintain their livelihood. The natural selection of increasingly complex, stable and self-sustained interaction networks has led to teleological systems whose existence can be explained without having recourse to mysticism, Intelligent Design, or spooky backward causation.

3. The formal representation of knowledge

The formalization of knowledge is an important aspect of KBS and AI systems design, much like mathematics is to physics. Although explicit knowledge and its representation constitute only one aspect of intelligent systems (analogical non-verbal knowledge being the other) it arguably is an important one. Ontologies need to be represented in formal languages that can help unify and express concepts in a structured way and a good starting point is first order logic (FOL). FOL is the mathematics of objects and their relations. It is an expressive language with a rigorous syntax, a clear denotational semantics, and sound inference (Kleene, 1952; Genesereth and Nilsson, 1987). The expressive power of FOL can be assessed by the following example (Russell, 2015): "The key benefit of first-order logic is its expressive power, which leads to concise—and hence learnable—models. For example, the rules of chess occupy 10^6 pages in first-order logic, 10^5 pages in propositional logic, and 10^{38} pages in the language of finite automata". A full treatment of FOL, and *a fortiori* its extensions, is beyond the scope of a single paper. However we will outline some basic notions in order to make the paper self-contained.

3.1. Syntax of first order logic

3.1.1. Terms

Formally speaking, a conceptualization in FOL is a triple $\langle U, F, R \rangle$ where U is the *universe of discourse*, F a *functional basis set*, and R is a *relational basis set* (Kleene, 1952; Genesereth and Nilsson, 1987). The universe (or domain) of discourse contains all the objects of the conceptualization (i.e. the ontologies). As mentioned in section (2.2) the notion of "object" is broad and can cover physical and mathematical objects as well as algorithms. Objects are represented by *terms* of which there are three types: *object constants*, *functional terms* and *variables*. *Object constants* represent the elements of the universe of discourse that can be concrete (the sun) or abstract (the set of integers). Interrelationships among objects can be represented by *functional terms* that are defined by a *function constant* and n terms $\tau_1 \dots \tau_n$ and forming a *functional term* $f(\tau_1 \dots \tau_n)$ that map n functional terms into an object of the universe of discourse. For instance, the functional term $\text{Size}(RBC_1)$ maps a constant (denoting a particular red blood cell) into some real number. By definition, functional terms are thus recursively defined allowing them to be combined with one another. *Variable symbols* denote objects without explicitly naming them, e.g. $\text{Size}(x)$. A *ground term* is a term containing no variables, e.g. $\text{Size}(RBC_1)$ (see also Appendices A–C).

3.1.2. Predicates and formulas

Predicate symbols of arity n , represent relations between n objects or object attributes, i.e. $R(\tau_1 \dots \tau_n)$ where the τ_i are functional terms. A predicate of arity 0 is a *proposition* (e.g. "the earth is flat"). A proposition can only take two values: true or false. In predicate calculus, facts are stated in the form of formulas that are either *atomic* or *composite*. An

atomic formula is a single relation. A composite formula combines atomic formulas with the unary logical operator (\neg) that denotes negation and with binary logical operators ($\wedge, \vee, \Rightarrow, \Leftrightarrow, =$) that respectively denote conjunction, disjunction, material implication, material equivalence and equality. A *ground formula* is a formula made of only ground terms. Appendix A gives examples of atomic and composite formulas.

3.1.3. Quantifiers

Quantifiers allow speaking about objects and their properties without naming them explicitly (Kleene, 1952; Genesereth and Nilsson, 1987). The general form of *universally quantified* sentences is given by $\forall x P(x)$ meaning that $\forall x P(x)$ is true for all x . The general form of existentially quantified sentences is given by $\exists x P(x)$ meaning that there is at least one x such that $P(x)$ is true. The scope of a quantifier in a certain formula is the (sub)formula to which the quantifiers applies. In the formula $\forall x (P(x) \Rightarrow Q(y))$, the scope of the universal quantifier \forall is $P(x)$. In this formula the variable x is *bound* while the variable y is *free*. A *closed formula* is a formula which contains no free variables. (see also Appendices A–C).

3.2. Deduction

One of the main uses of FOL resides in the encoding of parts of the real world into formal systems that represents “irrefutable” facts as axioms and derived facts as theorems (Kleene, 1952). In FOL, the derivation of theorems from axioms uses an inference mechanism known as *modus ponens* (deduction) whose generic structure is given by: $A \Rightarrow B; A \therefore B$ meaning “ A implies B ; A is true therefore B is true. Here A and B can denote sentences of arbitrary complexity. For instance A can denote an entire knowledge base written in FOL sentences. The facts in a knowledge base are logically equivalent to a single sentence in conjunctive form. Hence one of the main purposes of core ontologies is to handcraft a knowledge base (KB) that is as compact as possible. That is, a KB containing implicit knowledge about a large piece of reality that can thereafter be explicitly derived (automatically) via inference procedures such as modus ponens.

3.3. Extensions of first order logic

We found that FOL is not expressive enough to capture our core ontology. Several extensions to FOL have been addressed and we very briefly describe them.

3.3.1. Higher-order logics

FOL quantifies over variables that range over individuals only. However, we found that the more expressive power of higher order logics is necessary in order to (i) quantify predicate and functional symbols, and (ii) define objects that designate *sets* or *categories*. Most elements of our core ontology do not represent individual objects but classes (sets). For instance the terms *EpithelialCell* and *EpithelialTissue* designates the class of epithelial cells and epithelial tissues respectively, not instances thereof. Likewise, most of inference and problem-solving engines have been defined from property and relations types, such as explicitly defining if a property is intrinsic or relative or if a relation is symmetric, transitive and/or reflexive. Reasoning about property and relation types is outside FOL but can contribute to powerful higher-order inferences.

3.3.2. Qualitative spatiotemporal reasoning (QSR)

Space-time is part of the fabric of the universe and virtually any piece of knowledge will refer to space, time or both in an explicit or implicit manner. QSR aims at capturing salient spatial and temporal aspects of things while abstracting away unimportant ones (Dylla et al., 2017). It provides methods to represent and reason about spatial and temporal knowledge. QSR addresses different aspects of space including

topology (Egenhofer and Franzosa, 1991; Dylla et al., 2017), geometry (Dylla et al., 2017) and kinematics (Glez-Cabrera et al., 2013). Qualitative spatial reasoning has many applications including planning (Bacchus and Kabanza, 2000), Geographic Information System (GIS), robotic navigation and high level vision (Cohn and Hazarika, 2001). Works in mereology (Casati and Varzi, 1999; Cohn and Hazarika, 2001) and mereotopology (Varzi, 1994, 1996) have provided the space primitives and axioms that should underpin all formal descriptions of space. In the GMSP, topological attributes of structures are based on these primitives (Appendix A). Time is also a fundamental concept through which change can be perceived and formalized. Logicians have extended FOL in different directions including adding temporal modalities such as “until”, “always”, “next” (Bacchus and Kabanza, 2000); by describing successions of events (history) such as in situation calculus (Levesque et al., 1998); or by explicitly representing time such as in reified temporal logics (Allen, 1983). In reified temporal logics, propositions are made into objects and the temporal primitives are either time points (McDermott, 1982) or time intervals (Allen, 1983). When the primitives are time intervals, Allen’s calculus can be applied to represent and reason about both space and time (Amaneddine and Condotta, 2013). We used this calculus in qualitative simulations (Siregar et al., 1995), inverse problems (Siregar and Sinteff, 1996b) and multi-scale modeling (Siregar et al., 1996, 1997). The GMSP incorporates point and interval calculi as well as histories in order to represent and control morphogenetic processes (see section 4.3). QSR can prove to be critical for medical imaging applications such as model-based digital pathology, especially when coupled with machine-learning techniques (Siregar and Julen, 2016a).

3.3.3. Approximate reasoning

Knowledge about the world is inherently *incomplete* hence always subject to revision. Theories and opinion can, in the light of new data, be refuted. Hence human reasoning is non-monotonic more often than not and cannot be formalized within the confines of FOL alone which was initially designed to formalize mathematical reasoning. In addition, natural sentences are often *imprecise*, *uncertain*, and even contradictory. Imprecision is generally concerned with the *content* of the propositions (e.g. antecedent and conclusion of a rule). For instance linguistic variables whose values are not numbers but labels such as *increased*, *decreased* and *normal* are imprecise and denote fuzzy subsets of the universe of discourse (Zadeh, 1983). Uncertainty is associated with the *truth values* of the propositions. In classical logic, a proposition is either true (1) or false (0). But in many concrete situations, binary logic may be too restrictive to capture rational but not necessarily sound inference. Alternatives to classical logic include multi-valued logic, non-monotonic logics that are extensions of FOL (McDermott and Doyle, 1980), fuzzy logic (Zadeh, 1983), and Bayesian inference (Pearl, 2009, 2014). Recent works are paving the way for combining FOL with probabilities that “holds enormous promise for AI” (Russell, 2015). In the GMSP approximate reasoning is currently focused on Bayesian inference where the truth-theoretic semantics of FOL is associated with a probability distribution on binary-valued random variables (Appendix D). We have incorporated Evidence Based Medicine inference (derived from Bayesian inference) in a medical education tool (Siregar and Julen, 2016b) currently being developed and that will serve as a testbed before its incorporation in our CM platform.

4. The GMSP framework

There are two separate but complementary aspects to consider in CM. One is modeling the multimodal intercellular interaction/communication fields and the other is modeling how cell respond to these fields. These two aspects combined model the dynamic couplings between hierarchical entities. In our previous paper we focused on the physics and mathematical underpinnings of the chemical, electrical and mechanical interaction fields (Siregar et al., 2018a). They permeate

intra and intercellular spaces, and their spatiotemporal patterns are constantly changing in shape, modality and intensity. What remains to be closely examined are the mechanisms that control those exquisitely precise spatiotemporal patterns in 3D space. The equations of physics described in (Siregar et al., 2018a) are mute with respect to this question since they describe how fields propagate but not when and where they should be generated. This is where biology steps in. Individual cells continuously “scan” the time-varying fields. Their expression such as modality or intensity correspond to the net response of a huge number of cells (neighbor and distant) that may include distant sensory cells processing information generated outside the organism. The information flow is top-down (environment to genes) and will elicit a cell response mediated by gene regulatory networks that are specific to the cell lineage and state. The effectors of this regulatory process are field-generating sources (e.g. morphogens) that realize a bottom-up (gene-to environment) process where each single cell modifies the environment in which it is embedded. Biology teaches us that genes constitute very dense information-packed structures that embody, in the abstract DNA code, all the potential responses of a cell. However, the actual cell responses depend on cross-talks between the cell and its environment. The GMSP was designed to represent, model, and simulate this multi-scale circular causation using as few building blocks and assumptions as possible. It was conceived as a teleological system whose architecture is intended to mirror living matter characterized by parts, functions and processes. The whole knowledge structure is based on a recursive modular composition of only three base components:

structural units, functional units, and problem-solving units (Siregar, 2009) (Fig. 1). A singular object is designated as a structural unit. Teleological knowledge is encoded in the functional units. The processes that support the functions are also symbolically represented in the functional units. Numerical computations that implement the processes are attached to the problem solving units. They are specialized goal-directed AI agents that are associated to structural and functional units and are designed to bridge domain-concepts with mathematical and computational ones. Their role is to assist model-generation as well as other complex tasks such as model-calibration, diagnosis and learning (Fig. 1).

4.1. Structural units

4.1.1. Structural attributes

In the GMSP, the structural units (SU) are intended to represent material objects defined by their spatial, physical and biological properties. Spatial features can be derived from topological and metric spaces. Topological properties include *continuity*, *genus* and *connectedness* (Isaeva et al., 2012; Presnov et al., 2014). Metric spaces extend topological spaces by the introduction of a distance measure. They thus allow defining geometric features as *shape* and *dimension*. Topological features can be defined by a limited set of axioms (rules) that are expressed using primitives such as “interior” and “border” of continuous domains (Casati and Varzi, 1999; Cohn and Hazarika, 2001; Varzi, 1996; Dylla et al., 2017). These idealized concepts are important in biology because major events such as mass/energy/information

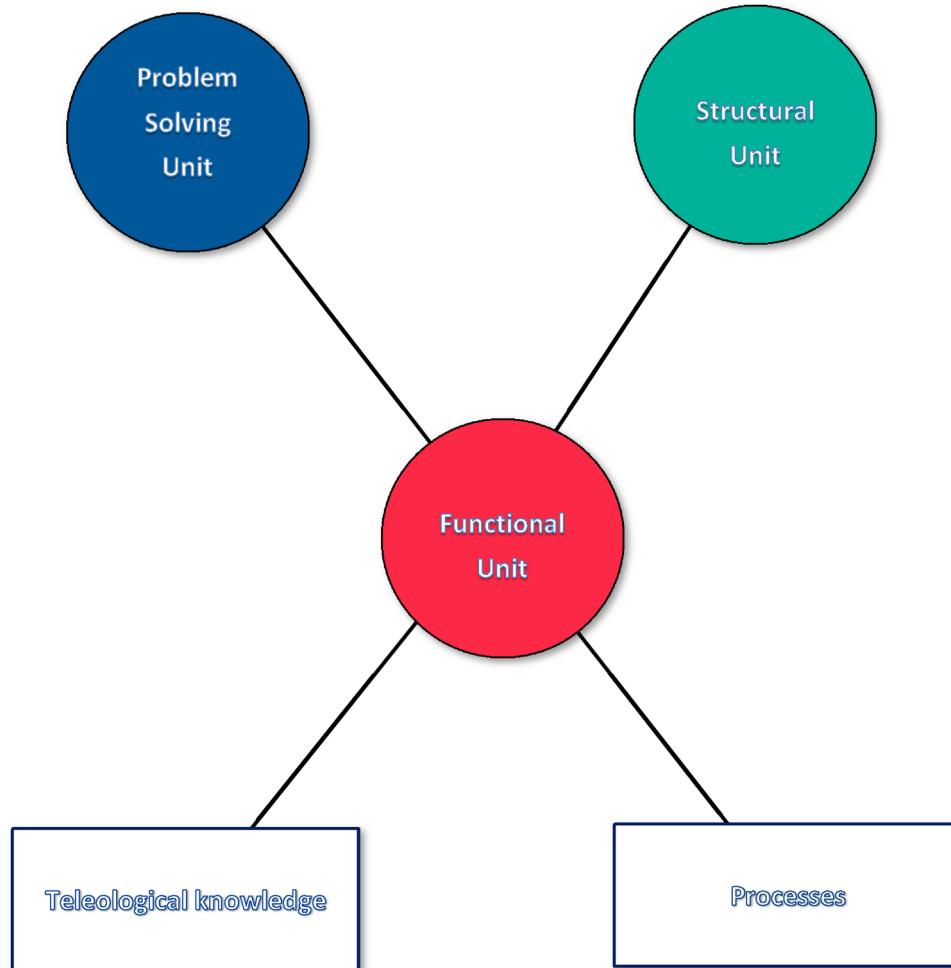


Fig. 1. The canonical structure of the GMSP. The whole architecture is based on only three types of components: structural units that describe static properties of objects, functional units define the dynamic functions of structural units, and the problem-solving units that solve highly specific tasks; e.g., model calibration, decision-support for clinical applications, etc. Hence, many problem-solving units can be assigned to a single functional unit.

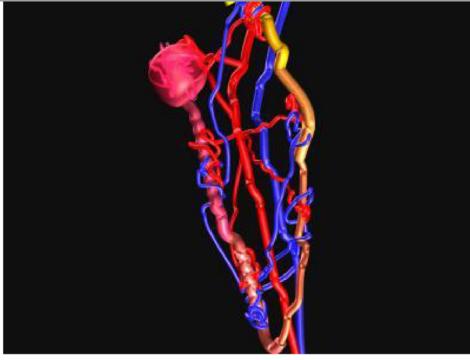
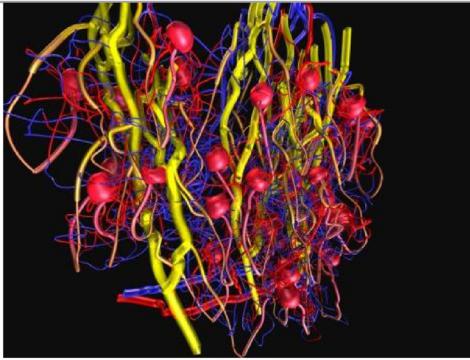
Generated 3D model	Topology in FOL
	<p><i>Afferent(AfferentArteriole, Nephron),</i> <i>Efferent(PeriTubularCapillary, Nephron),</i> <i>ContinuesInto(BowmanCapillary, ProximalTubule),</i> <i>ContinuesInto(ProximalTubule, DescendingHenleLoop),</i> <i>ContinuesInto(ProximalTubule, AscendingHenleLoop),</i> <i>ContinuesInto(AscendingHenleLoop, DistalConvTubule)</i></p> <p>...</p>
	<p>Nephrons have identical topologies but have unique geometries.</p>

Fig. 2. Class invariance and individual uniqueness. Knowledge about the “nephron” class includes topology that is shared between each class member. However, morphogenesis generates unique class members: each class instance (a particular nephron) has unique attributes such as position of its constituents, its actual dimensions, and geometry (e.g. different degrees of tubular convolutedness).

transfer happen at the borders of structural units. For instance, an organism’s “border” is the interface between self and non-self. The borders of domains can, from a geometrical perspective, be represented as surfaces in n-dimensional metric space. Surfaces can be defined explicitly, by sets of sampling points, or implicitly by mathematical expressions borrowed from analytical and differential geometry (Siregar, 2000). The parameters of the mathematical expressions can be used as features that may distinguish a normal structure from an abnormal one. Hence both topology and geometry provide important “biomarkers” especially because topology and geometry are closely linked to biological functions. Topology tends to feature the scale-free organization of living matter while geometry highlights the diversity of tissues (Figs. 2–4). Material objects can also be characterized by physical attributes. Mechanical features of interest in CM include Young’s modulus, elasticity and viscosity; electrical features include electrical conductivity; and biophysical properties include diffusion coefficients for given chemical species (Siregar et al., 2018a). Living matter is also characterized by cell and tissue biological attributes. These include markers of cell identity and lineage such as transcription factors or membrane receptors, and tissue features such as cell-types, the ratios between cell-type numbers, or the presence or absence of lumen, etc. Data on, and simulations about, morphogenesis clearly demonstrate how cell properties determine the emerging features of tissues. In the GMSP, there are two types of SUs: class SUs and individual SUs. The class SU embodies knowledge that pertains to all class instances (e.g. epithelial cell). In addition to the description that is common to its class, each SU instance is uniquely defined by a private logbook that summarizes its history in symbolic form, and makes SUs *self-descriptive*. During the simulation of, say, kidney tissue morphogenesis, many instances of virtual nephrons will be generated. Each virtual nephron will

have class features that will be common to all other nephrons, but also exhibit specific (mostly unique) geometrical attributes. Fig. 2 gives illustrative examples of common topological attributes versus unique geometrical ones that can characterize nephrons.

4.1.2. The nested organization of structural units

Most material objects are in fact compound objects made of constituents. In the GMSP, the recursive aspect of matter composition is captured by nested structural units (SU) formalized by *parthood* type relations (Appendix A). Explicit relationship between the properties of matter and the level of description in which the properties are expressed are also encoded in the SUs. For example, the substance “water” exhibits, at the macroscopic scale, bulk uniformity, electrical neutrality and nonpolar qualities while at a microscopic scale, the substance is as a discrete medium made of polarized objects: the water molecules. Likewise, an organ, viewed with the naked eye, may appear as a smooth continuous medium. A switch to higher resolution reveals spatially organized discrete tissue components (“modules”) that, at a still finer level of description, will reveal cellular organizations that exhibit their own specific attributes.

4.2. Functional units

4.2.1. Biological functions and processes

In the GMSP, the *functional units* (FU) incorporate teleological and symbolic knowledge about the processes that carry-out the function. Functions are strongly related to material objects insofar as functions are what they are because material objects are doing what they do (process) with the properties that they have. This seemingly trite statement summarizes in everyday parlance the relationship between functions,

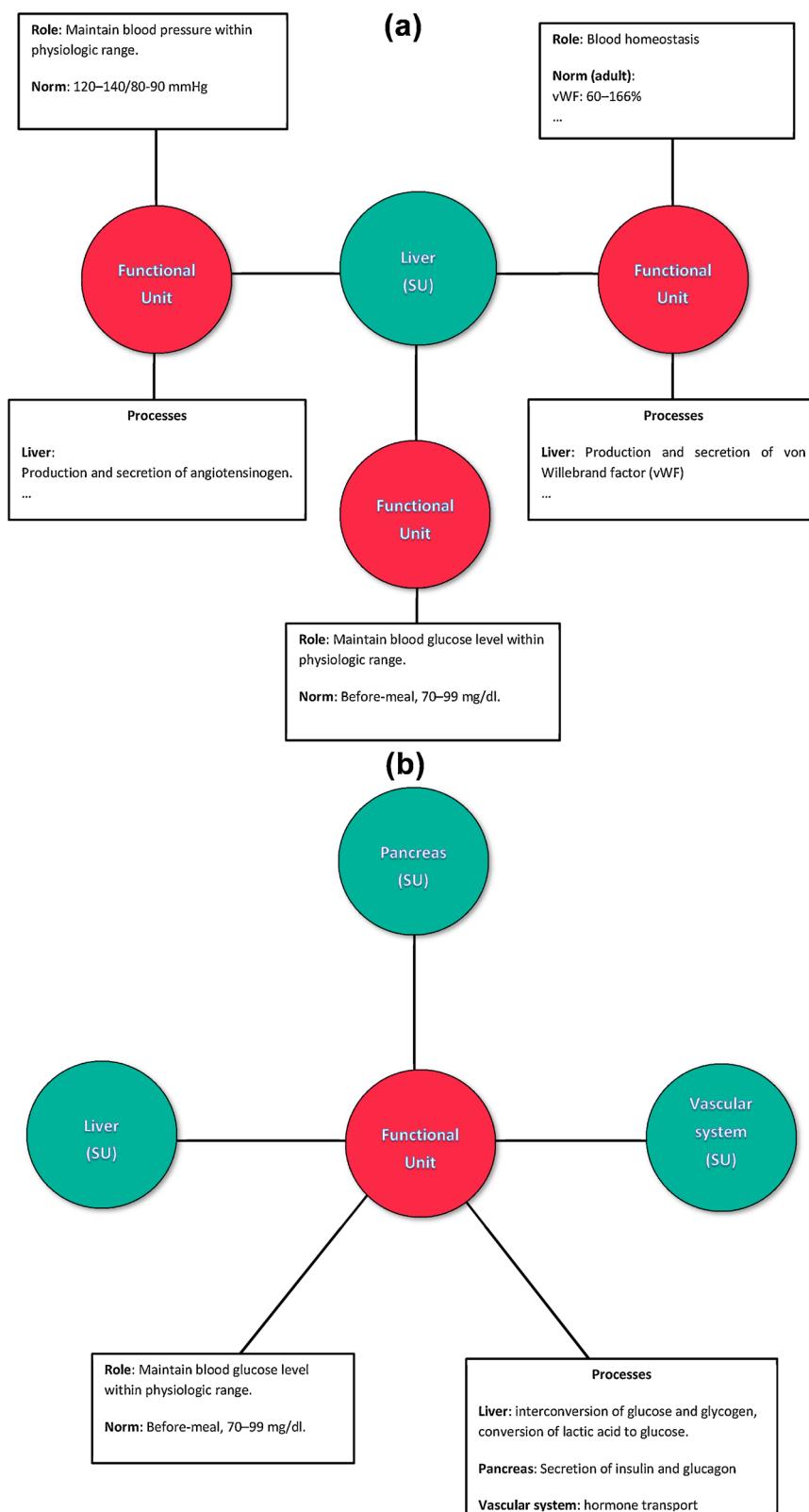


Fig. 3. (a) A structural unit can contribute to one or more functions. (b) A functional unit can be carried-out by one or more structural units.

structures and processes. The well-being and livelihood of a living organism depends on the coupled dynamics of its constituents that must remain within certain operating norms. For instance, the role of the heart is to pump blood, but proper functioning entails that state variables such as blood pressure must be maintained within certain limits. These limits

impose constraints on the electromechanical processes that underlie the pumping function as can be assessed by the relationships that hold between blood pressure, cardiac output and heart rate (Fig. 3). This interdependency between function and normatively defined processes holds across description levels that differ by orders of magnitude. Any

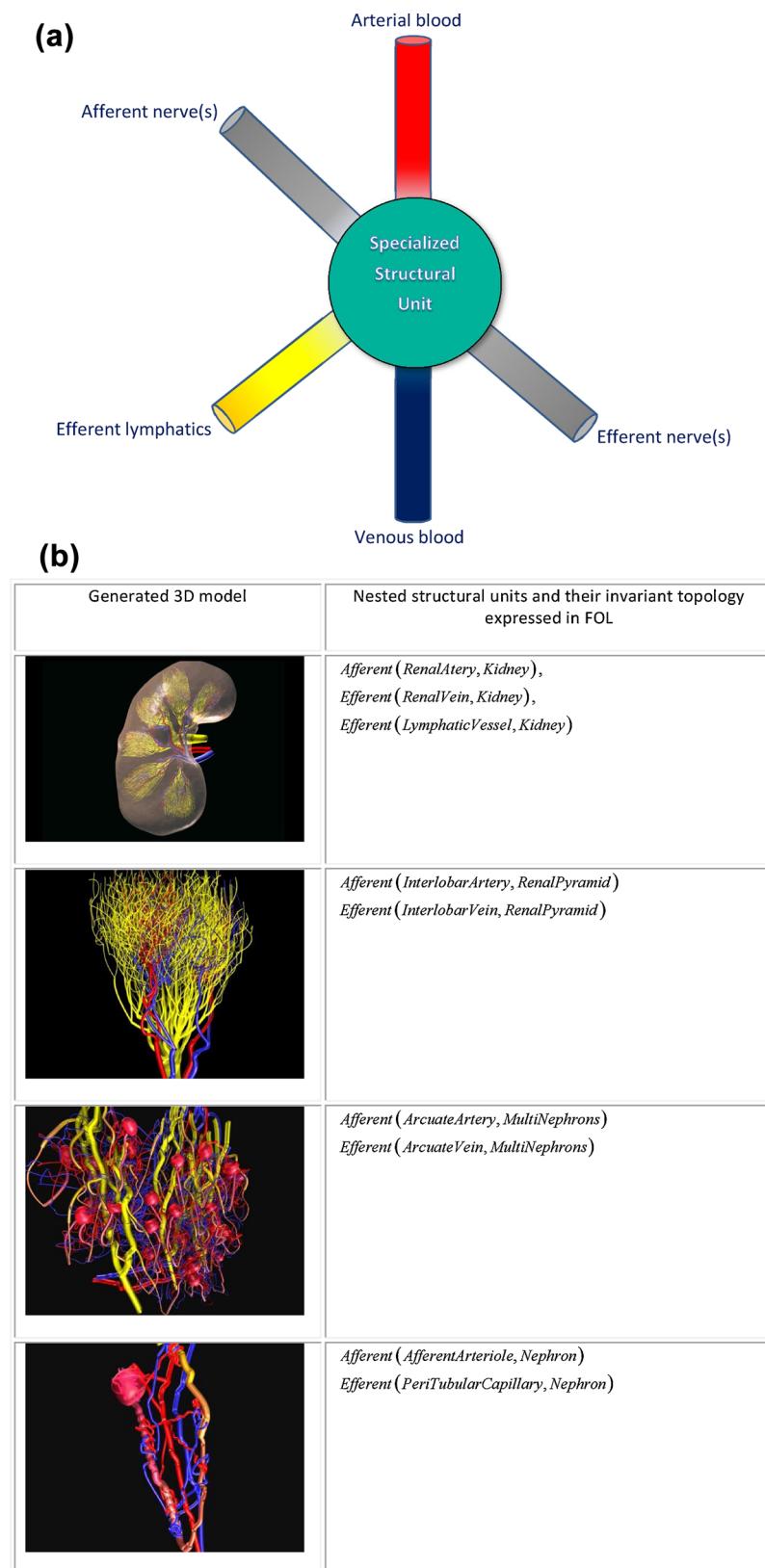


Fig. 4. (a) Scale-free topological and functional properties of living tissues. At all levels of description, organs, tissues down to small groups of cells are characterized by the same topological and functional pattern. They are open systems with matter inputs (nutrients) and output (waste elimination) requirements. Communication and control is mediated by blood vessels carrying hormones as well as nervous system components while host defense “logistics” include lymphatic vessels. These modular scale-invariant aspects are characteristic of living matter. **(b) The scale-free, fractal organization of the generated models.** The geometry and sizes of structural units vary within and across organizational levels while fundamental topological features remain unchanged. The nervous system network and lymphatics were not generated in our simulations since these were preliminary tests but their realization depends on morphogenetic principles that hold for any tissue.

parts of the heart down to single cardiomyocytes also contribute to heart's pumping activity, must operate within physiological ranges, and exhibits electro-mechanical activity. What holds for the heart holds for any system or organ. At the individual cell's organization level, biological function is intimately associated to the expression of cell membrane

receptors, the activation of transduction pathways, the on/off switching of gene regulatory networks, and the activation of effector processes that materialize the cells' responses to external cues. Other than fulfilling its role for the organism, the other main cell constraint is to satisfy its own metabolic needs.

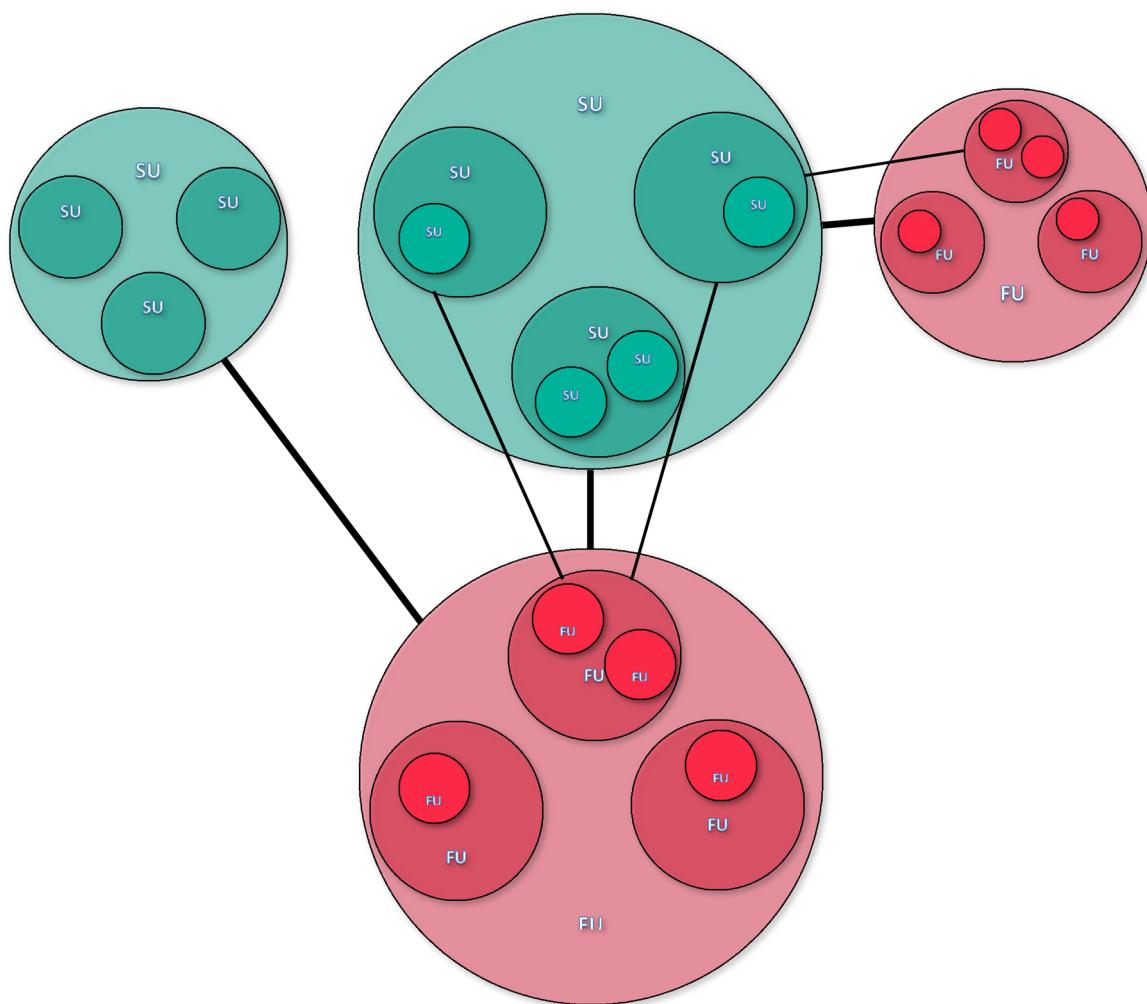


Fig. 5. Multi-scale structure/function/process relationships. The recursively nested structural and functional units can create complex and rich multi-scale networks of relations between structures, functions and processes. This organization allows addressing the question of pleiotropy in a clear and articulate manner.

4.2.2. The concepts of dysfunction and disorder: an example

In teleological systems, a dysfunction can be defined as one or more system norms not being satisfied and a disorder as one or more end-goals of the system, *as a whole*, not being realized. As an example consider the following concrete example and its relation to our core ontology. The heart is composed of myocytes that at each instant need to have balanced ATP consumption and production rates (cell energy norm) and survive within a certain pH range (cell environmental norm). In order to satisfy these constraints, the blood flow must be adjusted accordingly (flow control process). The presence of a coronary artery stenosis causes a reduction of the blood flow (flow process change) and thus a reduction of the influx of oxygen and nutrients in the heart tissue downstream of the artery (flow process change). This may result in an imbalance between the production and the consumption of ATP (cell energy norm being violated = dysfunction). The occurrence of this event induces a switch to anaerobic glycolysis (compensatory energy control process) and the production of lactic acid (molecular transformation process) (Kalogeris et al., 2012). Eventually, the decreased pH may move beyond the lower allowable threshold (cell environmental norm being violated = dysfunction). As a result, cell apoptosis, autophagy and tissue necrosis (cell transformation process) occurs leading to physical pain and impaired mental activity due to the brain metabolic needs not being satisfied (metabolic norms being violated) leading to end goals of well-being, normal mental activity, and even livelihood being challenged (disorder).

4.2.3. The cardinality relations between functions, structures and processes

- A structural unit can contribute to one or more functions. The liver anabolizes glucose into glycogen, catabolize glycogen into glucose and thus contributes to the control of the blood glucose level. The liver also produces and secretes angiotensinogen that contributes to the control of blood pressure (Fig. 3a).
- A function can be carried-out by one or more structures. The heart and kidneys contribute to the control of blood pressure. The maintenance of the blood glucose level within a physiologic range (function) is carried-out by the pancreas (insulin, glucagon) and the liver (Fig. 3b).

It thus follows that a function can be carried-out by one or more processes and vice versa.

4.2.4. The fractal organization of living tissue

Multi-scale functional self-organization can be substantiated by assessing the modular self-similar organization of living matter. Living organisms are dissipative and open systems subject to continuous matter input/output fluxes (Prigogine and Stengers, 1979). Any organ, tissue region or small groups of individual cells is always composed of the following: blood vessels involved in the transport of nutrients, hormones, and metabolic waste; lymphatic vessels involved in defense mechanisms and fluid transport; parenchymal cells that carry-out the

tissue-specific functions; supporting cast of protective cells that constitute the stroma; sensory and control neurons that, along with the hormones, regulate the processes that carry-out the functions. For example, the kidneys are vascularized by the renal artery and renal vein and innervated by sympathetic end sensory branches of the splanchnic nerve. The kidneys can recursively be decomposed into renal pyramids, sub-regions of pyramids, down to the smallest functional units that are the nephrons (Tortora and Derrickson, 2006; Vize et al., 2003) (Fig. 4). Like the whole organ, each pyramid is also vascularized by branches of the renal artery and innervated by branches of the splanchnic nerve. Likewise, each nephron is vascularized by the afferent arteriole, and venous blood flows through the peritubular capillaries, while sympathetic nerve fibers control, for instance, the diameter of the afferent arteriole. The same holds for many organs and tissues such as the liver, bones and muscles. Their smallest functional units are, respectively, the lobules, the osteons, and the muscle fibers (Tortora and Derrickson, 2006). Thus a complete functional unit is minimally composed of a central structural unit composed of parenchymal and supportive stroma cells, accompanied by transport, sensory and control units (Fig. 4a). In the GMSP, structures that carry-out the specialized (organ-specific) functions are represented as central SUs while transport and control structures are represented as auxiliary SUs associated to the central SU. There is a strong underlying unity behind the immensely diversified biosphere. This unity manifests itself through scale-free (or scale-invariant) structural and functional attributes of living matter (Fig. 4b). It is this unity and self-similar quality that we have attempted to capture in the GMSP's architecture and knowledge structure (Fig. 5).

4.3. Qualitative models and qualitative process simulations

By “qualitative modeling” we refer to non-quantitative computation. With this definition in mind, many distinct approaches, methods and goals can be cited. In the context of dynamical systems theory, qualitative analysis includes studying trajectories and the typology of critical points in phase space (Bocarra, 2009). A special branch of mathematical biology, baptized relational biology by its inventors, proposes to study living systems within the context or category theory (Rosen, 1985, 1991). In relational biology, time, space, and in fact all of physics are abstracted-out, leaving information as the only primitive. Modeling consists of building and examining causal structures based on inferential entailment that is intended to mirror causal entailment in natural systems. Relational biology, has initiated research and initiatives that attempt to formalize complex systems and living organisms within a category theory framework (Ehresmann and Vanbremersch, 2009; Louie, 2009; Simeonov et al., 2012, 2013), and attempts to combine different formalisms have been undertaken (Ehresmann and Simeonov, 2012). One of the most illuminating results of relational biology, the closure of efficient causes, has sparked interest in theoretical biology, and notably in characterizing biological organization in terms of closures (Mossio et al., 2013; Montévil and Mossio, 2015). In CM however, space and time cannot be abstracted out. Positional and directional information play a most central role, notably as a determinant of cell fate (Gilmour et al., 2017). Thus CM -as it is defined here- requires a metric space, or one that combines both distances (i.e. scalars) and vectors such as a Hilbert space. Since time is also a critical factor, CM is inherently associated to the space-time of physics. In our previous paper (Siregar et al., 2018a) we examined constitutive equations that describe the multimodal fields that permeate the extracellular space of living tissues. The mechanisms underlying the cells' responses to these fields including transcription, splicing and translation are obviously of the same multimodal nature as the extracellular fields and thus could also be modeled by the same equations in a molecular dynamics setting. However, this would incur an impossible computational load. In our current framework, the cells' behavior and internal mechanisms are currently modeled by qualitative models based

on logic-circuit type formalisms that have been implemented in rules, Boolean networks or Petri nets with similar results. Qualitative models can emulate dynamic processes using symbolic representations of system behavior instead of numbers (Appendices B, C). They provide accounts of temporally ordered salient events and can make predictions of system behavior in contexts of uncertain or incomplete data. Qualitative simulations are very much akin to humans making mental predictions about system behavior (Forbus, 1984; Siregar et al., 1995) and can be grounded on spatiotemporal reasoning formalisms such as situation calculus (Levesque et al., 1998), and interval-based calculi (Siregar et al., 1995; Siregar and Sinteff, 1996b; Dylla et al., 2017). Qualitative models can also provide a primer to more detailed quantitative simulations in a multi-stage multi-resolution process (Siregar, 1996; Siregar et al., 1995, 1996, 1997), and could provide robustness to real-time applications such as patient monitoring (Siregar et al., 1993). Qualitative reasoning about physical systems includes handling qualitative values and their tendencies (qualitative derivatives), performing causal reasoning, or defining conditions for processes to occur. Qualitative terms such as “ x is increased” and “ x is decreasing” are the qualitative counterpart of the relations “ $x > \tau$ ” (τ being an upper “normality” bound) and “ $dx/dt < 0$ ”. Domains such as physiology and pathophysiology are replete with statements involving common-sense causal reasoning such as “high blood glucose level will trigger (cause) the secretion of insulin into the blood compartment by the beta cells of the Langerhans pancreatic islets”. This short sentence combines a qualitative variable, a causal relation, processes, material objects, implicit temporal ordering and space-related terms. Its formal representation is given in Appendix C. Very schematically, qualitative prediction corresponds to defining initial conditions and making inferences over possible future qualitative states using the transitivity of causal relations. Causal reasoning is, however, defeasible since unknown variables may prevent the anticipated effect of a cause to occur, but this issue will be left-out of the present discussion. In our CM platform, qualitative representations of events play an important role in the model validation task. All generated agents (SUs and FUs) write their own individual history in the form of time-stamped sequences of events expressed in symbolic terms (Appendices B & C). The goodness-of-fit of the generating system can be assessed by comparing the generated history of its components with a small set of events that constitute check-points of a reference history also in symbolic form (see also section 5.3).

5. The cell in the GMSP

5.1. The cell as a nested structural unit

A cell is a structural unit composed of subcellular components that include the cell membrane and its domains (e.g. apical), the cytosol, mitochondria, nucleus and genome. Each component is an SU that interacts with other units in order to implement cellular functions that include transduction and metabolism (Fig. 6). The “genome” is itself a SU made of “genes” wherein each “gene” actually represents (possibly large) sets of genes that control whole pathways and major cellular events like cell polarization, differentiation, etc. For instance a “gene” determines if a given cell should migrate. Another will determine the direction of migration, etc. All our virtual cells have the same “genome”. What will distinguish a cell's lineage and state are the transduction pathways and the (functional) topology of the gene regulatory networks. Pathways, genes, and molecules are represented by SUs that have a common template. Each SU embodies a boolean variable that will indicate if a unit is activated/inactivated. Each SU also embodies a set of variables that represent scalar quantities (e.g. concentrations) or vector quantities (e.g. cell polarization, net mechanical force, etc). This greatly reduces the number of “genes” and simplifies all cellular computations. Cell identity is represented by paths defined

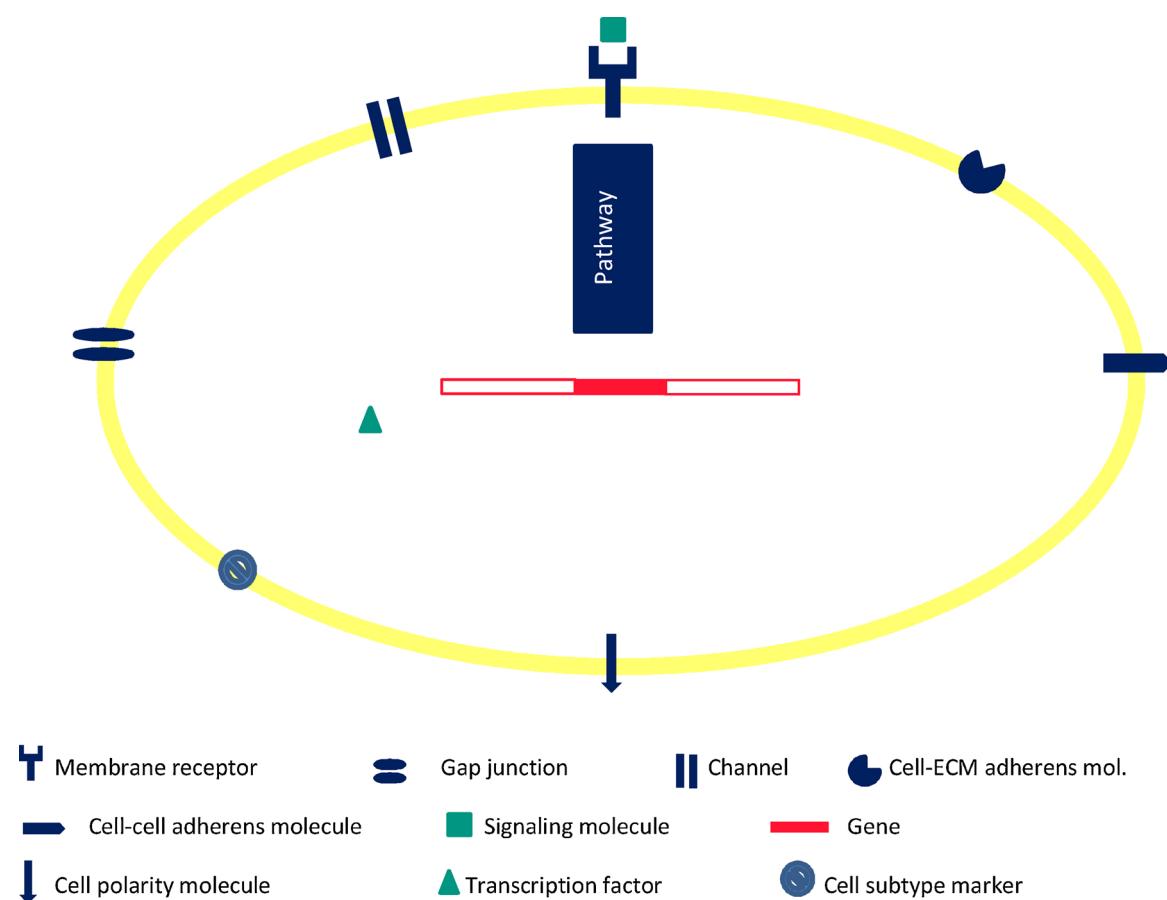
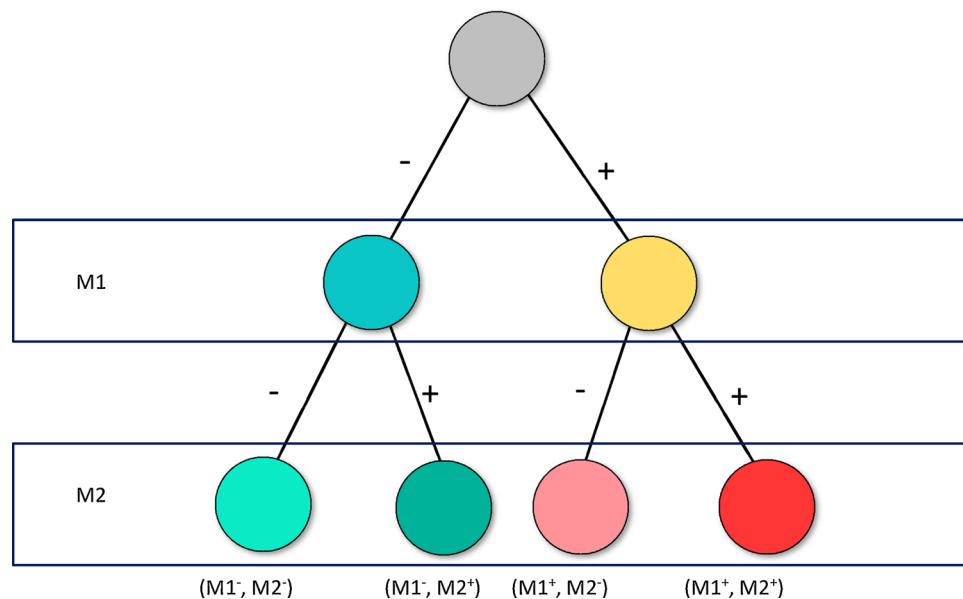


Fig. 6. Current classes of cell components. The number of classes has been limited to those that were deemed useful to our current simulations. However, each class represents a set that can have many members. For instance, kidney morphogenesis featured over 18 cell subtypes.



from the root node to the leafs of a tree structure where each level denotes a transcription factor and leafs denote cell lineage (Fig. 7). Likewise, a cell's state is assessed by the transcripts of membrane receptors, signaling molecules, cell-ECM adhesion receptors (integrins) and cell-cell adhesion molecules (cadherins). Cell identities and states are each summarized by an integer that ranges between 0 and 2^n where n is the number of transcripts.

5.2. The cell as a functional unit

Living cells are structural units that can play many roles and each one of them is represented by a functional unit (FU). Hence cells have many functional units attached to them and whenever knowledge about constraints is available it is also encoded in FUs. From a process point of view, our virtual cells will respond to a given external signal if two

Fig. 7. Cell identity. The expression of molecular markers such as transcription factors define cell identity that combines cell lineage, subtypes and state. Cell identity is formalized and computed as a path -from root to leafs- of a binary tree. The number of levels reflects the number of markers accounted for in our model. For lack of real data some of these markers can be fictitious.

conditions are met: the cell expresses the appropriate receptors and the signals intensity is above a given threshold. If these conditions are met, the cell's response is determined by the following internal events: signal transduction, activation of gene regulatory networks, gene product(s) that include signaling molecules and transcription factors that define the resulting cell identity and state. The events are modeled by simple logic-circuit type formalisms (e.g. Petri nets) in which the genes play the central role. Like real genes, they encode *all* the potential responses of a virtual cell and they embody an internal image of the system as a whole. The triggering of genes is determined by the activation of the transduction pathways as well as the expression of transcription factors. Gene expression reprogramming and emulating real experimental research (Teague et al., 2016) is modelled at this level.

Since Boolean variables characterize our virtual cell's internal components, which pathway/factor interacts with which genes and how genes interact with each other are defined by the logic of the gene regulation circuitry that is specific to cell lineage and state. Once, the cell state has been computed, global cell behavior is determined by a set of rules that match cell states to behavioral phenotypes such as cell division, differentiation, migration, secretion of molecules, apoptosis, and interconversion. Deterministic as well as stochastic rules, based on rates, have been implemented.

5.2.1. Cell division and apoptosis

Cell proliferation can occur with or without the presence of mitogens. Cell can divide symmetrically or asymmetrically. The general transformation rules for any distinct species A and B_j is as follows:

Symmetric cell division: $C_i; A \xrightarrow{\alpha_i} A + Ai = 0, \dots, n$

Asymmetric cell division: $C_i; A \xrightarrow{\beta_i} A + B_j; j = 1, \dots, p$

Apoptosis: $C_i; A \xrightarrow{\gamma_i} \phi$

where C_i is a context, with C_0 considered the "normal" default context, e.g., division in the absence of mitogenic cues or "natural" cell death. The parameter α_i and β_i are context dependent growth rates and γ_i are death rates. For asymmetric cell division we provide the possibility that a cell of type A can self-renew while producing distinct sibling cells B_j . The context of a given cell is formally represented as states, events and processes (Appendices A–D).

5.2.2. Cell transition and interconversions

In our model "cell differentiation" covers two types of reality: the differentiation from one cell type to another, e.g. mesenchyme to epithelial, and the transition from one subtype to another, e.g. epithelial (RET^-) to epithelial (RET^+). Similar to cell proliferation and apoptosis, we provide the possibility that any given cell may transition to one or more distinct cell lines at rates that can vary in a context-dependent fashion: $C_i; A \xrightarrow{\delta_i} B_j; i = 0, \dots, n; j = 1, \dots, p$

Cell interconversion is a special kind of cell transition: $C_i; A \xrightarrow{\delta_i} B; B \xrightarrow{\eta_i} Ai = 0, \dots, n$ that play a role in embryogenesis where, for instance, epithelial cells will transition to adopt the mesenchyme phenotype (ETM) during migration. In breast cancer tissues, equilibrium proportions in heterogeneous cancer cell populations have been explained in terms of transitions and interconversion between basal, stem-like and luminal cells (Gupta et al., 2009). In our framework, cell behavior is described by deterministic and probabilistic rules (Appendix C). Population dynamics can also be modeled by ODE following (De Pillis and Radunskaya, 2001), stochastic ODE following (Gillespie, 1976; Gillespie and Petzold, 2003) adapted to cell lines, and Markov chains (Gupta et al., 2009). A systematic study combining stochastic cell dynamics with the minute spatial specificity of positional information conveyed by multimodal morphogenetic fields could demonstrate that in some sense and at any given time, each cell is unique (including its epigenome and mutability), and as a consequence, the

"same" tissues within and between individuals can exhibit a great diversity of cell phenotypes (heterogeneity).

5.2.3. Cell-cell interactions

How the virtual respond to external cues is determined in the "genes" of each cell. Thus the logic of cell-cell interactions and their potential coordination are implicitly embodied in the cells' "genomes". There are no arbitrary upper-level rules that impact an individual cell's behavior although higher-level tissue-states that manifest themselves in the local interaction fields do have a profound influence as already discussed. The genomes implicitly encode autocrine, paracrine and long-range chemical interactions by coding for the ligand/receptor pairs between "sender" and "receptor" cells. Likewise, cell communication by direct contact involving transmembrane receptor/ligand pairs and/or gap junctions employs the same representational scheme. In our mesh-free particle-based cell model, cell neighbors are defined by "topological adjacency" that is established by cell-cell contact or by the formation of adherens junctions. Homotypic pairs of cells that are topologically adjacent can form adherens junctions and become mechanically linked by a spring system that connects their centroids (Siregar et al., 2018a). Gap junctions are formed concomitantly with adherens junctions. Collective behavior can then be orchestrated by message passing through gap junctions. For instance, in collective migration, if a lead cell says "go North", the message will be passed to follower cells if and only if gap junctions have been previously formed. Interactions such occurring in the Frizzled/Flamingo system that results in planar cell polarity, or the Delta/Notch system that participate in cell fate determination via lateral inhibition (Siregar et al., 2018a) can also be modeled by similar message-passing mechanisms.

5.3. Problem-solving units

5.3.1. General description

The role of a knowledge-based system is to make inferences and solve problems. It must make a clear distinction between proven facts, prior assumptions, system models, and refutable theories. In the GMSP, this role is assigned to the problem-solving units (PSU) (Siregar, 2009). A PSU is a cognitive agent that is assigned one and only one SU or FU. In other words a PSU dedicated to the SU "galaxy" will not be assigned to the SU "proton". Each PSU is associated to one particular object only. However, many PSUs can be assigned to a single FU or SU (Fig. 8). That is, PSUs are cognitive agents that have different world-views about a *same* object and thus address specific aspects of the said object. A PSU's role is defined by its type and can solve one class of problem *only*. For instance, the study of pleiotropy can be assigned to a particular class of PSUs that will analyze the functional role of a given SU across different levels of organization. Another PSU may inquire if phenotypic changes of the said SU are related to diseases. PSUs are teleological, goal-directed meta-level agents that monitor and act on the object-level FUs and SUs to which they are associated. Like functional units, PSUs incorporate structural (e.g. input and output types, algorithm structure) and teleological knowledge (function, constraints). PSUs of the same type have access to a shared KB in addition to a private log-book that records their own history (Siregar, 2009). There are two main types of PSU: ones that manipulate symbols and use logic and others that are biomimetic (e.g. swarms, neural networks) (Siregar, 2009). And like any set of objects, PSU can be organized into categories. In our conceptualization they are partitioned into the following main categories: Inference, Optimization, Learning and Discovery:

- Inference includes deduction, induction, hypothesis generation, abduction, temporal and spatial reasoning, causal reasoning, constraint-satisfaction, and Bayesian inference.
- Optimization includes classical optimization and meta-heuristics.
- Learning includes statistical inference and machine learning
- Discovery combines instances of the above categories



Fig. 8. Many PSUs can be associated to a single SU and/or FU. The role descriptions (teleology) are very succinct as this is an illustrative example only.

An example of a “Discovery” class PSU would attempt to answer the following question mentioned in (Siregar et al., 2018a). How can we relate the bulk mechanical properties of tissues with the mechanical attributes of their cellular and molecular constituents? This would involve accounting for the mechanical properties of collagen, elastin, and the viscoelastic properties of groups of cells forming adherens junctions via cadherins and other similar molecules. Here, a tissue-level PSU would collaborate with cell and molecular-level PSUs. As a group, they would explore the biological, physical and mathematical spaces jointly to arrive at some plausible explanatory models.

5.3.2. Model selection and instantiation

Specific classes of PSUs are assigned to model selection and instantiation. Their object-level combines knowledge about the FUs to which there are assigned (e.g. filtration by a nephron) as well as about equations, simulation methods and programs related to the FU. All simulators are coded in an imperative language (C++) and are linked to the PSUs by procedural attachment. Like all objects in a conceptualization, each simulator has a unique descriptor (referential unicity) and its selection is carried-out via the descriptor. More generally, PSUs dedicated to model-selection and instantiation do one of the following: select which formal model (e.g. equation) is adequate to solve a given problem, select the solver, check for initial conditions,

check for initial model parameter values etc. The PSUs draw their actions from the prevailing modeling context and the knowledge sources that link biological concepts with physical and mathematical ones (Fig. 9). The modularity and often self-similar properties of living matter have their counterparts in the physical and mathematical realm. The equations that describe molecular, electrical and mechanical processes are based on few physical principles and almost as few constitutive equations (Siregar et al., 2018a). What will differentiate the modeling of mass transports inside the cell, across cells and between distant organs are not the equations but their parameters. For instance the diffusion coefficient of given molecular specie within the cytosol may differ from that of interstitial space. Thus, part of the prior knowledge in the GMSP concerns the coupling between domain knowledge (e.g. the biophysics of a region) with knowledge about the computational models that simulate the processes.

5.3.3. Model calibration

Model calibration is assigned to a specific class of PSUs that can collaborate closely with those assigned to simulation (Fig. 10). The goodness-of-fit of generated models is assessed by comparing computed histories with *reference* histories defined at a few checkpoints. At each check-point, a “distance” measure between model features and real features of the partially developed tissue is computed. The goodness-of-

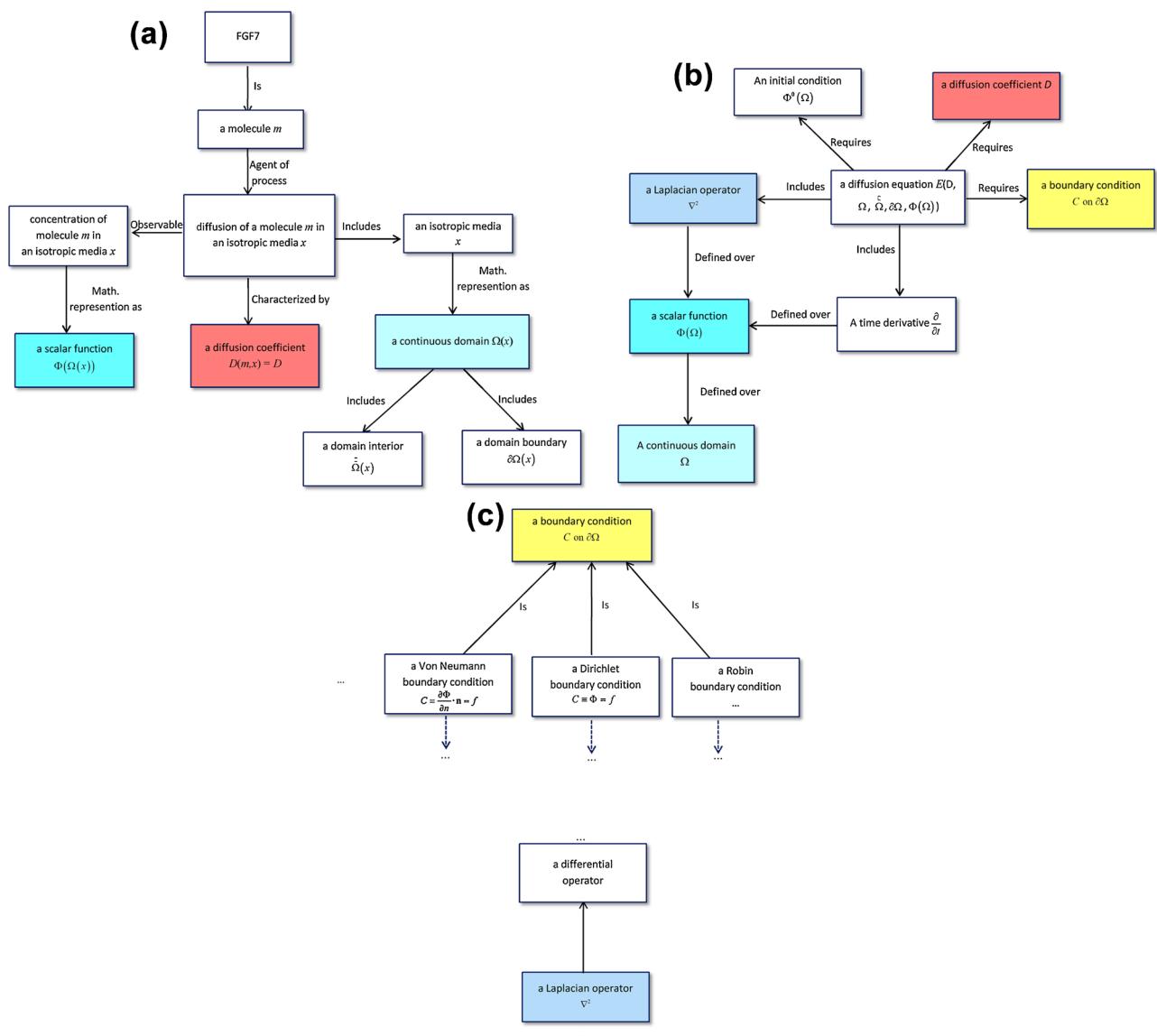


Fig. 9. (a) Establishing bridges between biology, chemistry, physics and mathematics. In Appendix B the functional term *Chemotaxis* (*EpithCell*, *Direction*(*Grad*(*Conc*(*FGF7*)))) translates into “chemotaxis of an epithelial cell in the direction of a gradient derived from a FGF7 concentration”. The functional term *Conc*(*FGF7*) designates “a FGF7 concentration” and the knowledge source associated to this term in the context of an isotropic media is schematically represented in this figure. Colored boxes correspond to matching boxes between this figure and Fig. 9b and c. (b) Part of the core ontology (knowledge source) corresponding to the diffusion equation for an isotropic media. (c) The diffusion equation knowledge source (cont'd).

fit is then defined by a distance measure between model features and real features which are mathematically incorporated in objective functions which either need to be minimized or maximized (e.g. Tett et al., 2017). If the discrepancy between simulated and target object is beyond an acceptable threshold, then the PSU backtracks to the previous check-point and proceeds to a tuning of the model parameter before proceeding to a new morphogenetic phase. The criteria for satisfaction embody quantitative and qualitative constraints. Quantitative constraints include the number of cells per unit volume for each cell-type, or the relative number of cells between cell-types. Qualitative constraints include tissue organization such as the relative position of different cell-types, topological attributes (e.g. genus, connectivity), geometrical features (e.g. “roundness”), the presence or absence of some compartments (e.g. lumen), and differential growth between cell types that can contribute to tissue buckling and shaping (Siregar et al., 2018a). In order to satisfy the constraints, PSUs recalibrate morphogenetic programs by accessing biophysical and biological parameters. Biophysical model parameters include the diffusion tensors (or

diffusion coefficients for isotropic media) of chemical species, and mechanical parameters such as Young's moduli. Cell-level biological parameters include proliferation and death rates for each cell lineage, interconversion rates between cell-types, and kinetic adaptation to morphogenetic gradients etc. Tissue-level parameters include bulk mechanical properties. At each level of organization, all parameters must reside within ranges that reflect known biophysics and biology. Stochastic components can also be incorporated by assigning probability densities or distributions to the said ranges. Many constraints in CM are “soft” ones since they pertain to classes and not individuals. For instance, objects such as nephrons will share common topological properties, but each nephron is geometrically unique (Fig. 2). Soft constraints hold the key to morphogenesis modeling for at least two reasons. First, exact object shape and sizes cannot be used as (hard) constraints since there are always geometrical variations between instances of a same class of objects. Second, model-calibration is currently riddled with missing data. The less data is available, the more we tend towards soft and even purely qualitative constraints (e.g. “epithelium

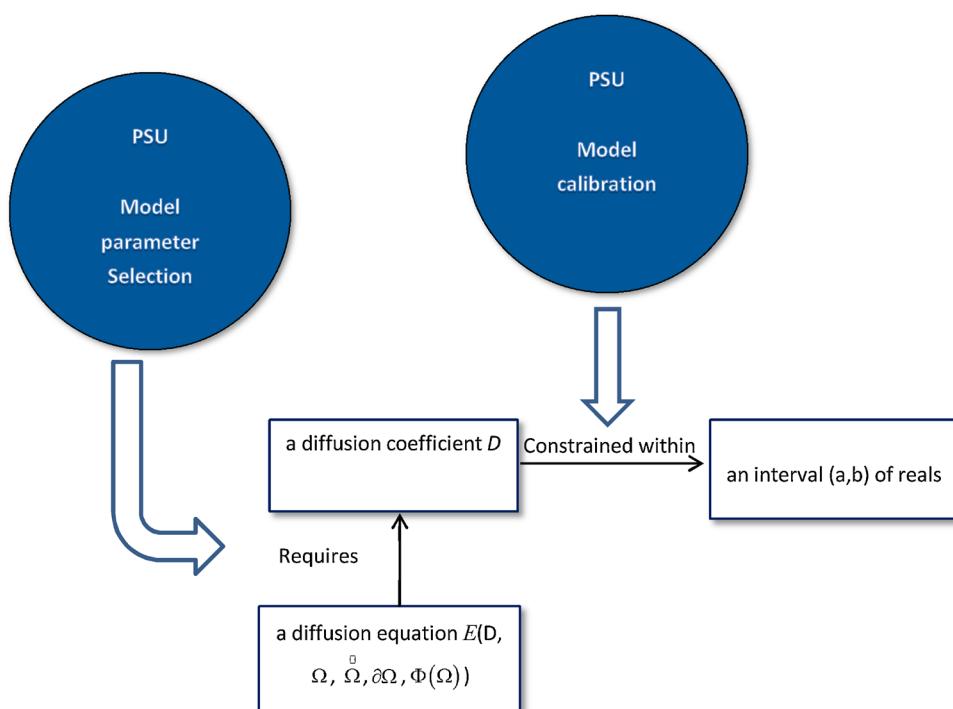


Fig. 10. Two PSUs associated to the diffusion of a molecular species. Many PSUs can be associated to a single FU. Each class of PSU reflects a particular world view or aspect. In this example, a *diffusion coefficient* is viewed as a required parameter of the diffusion equation by a “Model Parameter Selection” PSU while a “Model Calibration” PSU associates a constraint that must be satisfied during the model-calibration phase.

surrounds lumen”). In our framework, which is early-stage, model calibration has been limited to soft-constraints satisfaction (e.g. topological). The search through parameter space currently combines a Levenberg-Marquardt exact optimization method with inference procedures that exploit all available knowledge that may optimize the search (Siregar and Sinteff, 1996b; Siregar et al., 1997; Siregar, 2009). Model calibration can also apply Bayesian methods (Kano et al., 2003) as well as metaheuristics such as simulated annealing, genetic algorithms, and swarm-based optimization (Andrade-Cabrera et al., 2016). However, we found that the most critical factor in CM is the incorporation of prior knowledge, thus confirming the so-called No-Free-Lunch Theorem: search algorithms perform in accordance with the amount and quality of the prior knowledge they incorporate.

5.3.4. The nesting and self-organization of PSUs

As a consequence of the nesting of FUs, PSUs are also nested and are designed to interact like living cells, tissues and organs. Self-organization of PSUs corresponds to collaborating agents that adapt their individual behavior to satisfy constraints defined by higher-level PSUs. Since PSUs are nested so are the constraints and self-organization then becomes a multi-scale constraints-satisfaction process. PSUs can collaborate by exchanging messages (i.e. “bosons” fields) and by exhibiting the proper “receptors”. PSU actions can be schematized as follows: (i) interpret incoming message(s); (ii) define goals and actions (iii) implement actions; (iv) assess results; (v) if not satisfied return to (ii) or broadcast new queries if necessary and return to (i). In short, PSUs do in cognitive space, what cells or any goal-directed agent does in the material world. “Synaptic” links are created between co-active PSUs, and the strength of the link relates to how often two PSUs have been co-active in successfully solving a particular class of problem. Conversely, PSUs that are not co-active in successful problem-solving tasks will have their connection strength decreased. This process is reminiscent of the Hebbian rule in which co-active neurons see their synaptic gains increased (Caporale and Dan, 2008). The GMSP’s meta-level is organized as a fractal-like cognitive structure that is reminiscent of the mammalian brain. In the brain, most of the neurons form short-range synaptic contacts and feedback loops with nearby neurons. These small groups of neurons are organized into columns that are perpendicular to the

cortex’s surface (Mountcastle, 1997; Bureau et al., 2004). Columns then communicate with nearby columns via medium-length synaptic association, and neighboring columns form macro-columns that will exhibit long range excitatory associations with other distant macro-columns. In the GMSP, short-range associations correspond to local -same level- PSUs while long-range associations correspond to links between higher levels PSUs. As mentioned, these links are weighted, dynamic and will depend on both problem class and the goal satisfaction of the active PSUs. Hence our platform’s cognitive structure is designed to self-organize in function of individual and global goals satisfaction. This extends in a significant way current “deep learning” architectures such as convolution networks. To summarize, the GMSP has two isomorphic and parallel worlds: the world of (nested) computational models of reality, and the world of (nested) agents that perceive, reason and act on the computational models and on themselves. These two worlds attempt to capture salient features of a third one: the real world.

6. Discussion

Computational morphogenesis (CM) is a very rich subject-matter where many disciplines like biology, physics and mathematics converge. There are two sides to CM: a practical one and a theoretical one. Our initial work in CM was motivated by practical goals: to design a computer program that generates models of complex systems from simple hand-crafted initial conditions. More specifically, the aim was to conceive a program that does the hard and lengthy job of modeling complex systems using morphogenetic principles. It turned-out that some deep issues needed to be addressed first, including symmetry breaking, teleology, the relation between forms and functions, multi-scale closure of causation, topological invariance versus geometrical specialization, self-organization and emergence, analog information (fields) and discrete/digital information (DNA), and dynamic fractal structures among others. The development of a multidisciplinary core ontology was one of the first actions that was undertaken. It served as the conceptual scaffold around which CM could be formalized. As in all ontology research, there is not one single “right” answer regarding the choice of knowledge that should be made explicit, how it should be represented, and how it should be organized. The important (and

obvious) observation is that core ontologies can serve as efficient bridges between domain ontologies, while the latter are essential for organizing knowledge out of primary data. Work on CM can explain the apparent tension between the incredible diversity of abiotic and biological forms and the parsimonious and elegant means that Nature employs to generate these forms. The core biological processes (e.g. planar cell polarization) that underpin all morphogenesis are relatively few; they constitute “subroutines” or “modules” that organism call upon during embryogenesis and in other biological settings. How cells exchange information to implement these basic modules can be modeled by a small number of constitutive equations that are themselves derived from even fewer first principles in physics. Constitutive equations however cannot cover all the ground necessary to explain living matter – at least not in any practical way. This is where cells - and more specifically - genomes enter into the scene as the main determinant of the diversity of life-forms insofar as they provide instructions on how, when and where information can be exchanged. Genomes however are not organisms; they are dense, evolved, information-packed and abstract representations of organisms. They encode, in a stable structure, the *potentialities* of the organism in which they are embedded. Morphogenesis establishes the bridge between the information stored in a single embryonic stem cell that represents all the potential aspects of an individual and the *material actualization* of this individual, i.e. the final phenotypes. In Aristotelian parlance, the material cause is the embryonic stem cell and the final cause is the mature individual. The efficient causes include the intercellular interaction fields that drive morphogenesis. Now, for an embryo to develop into a mature viable individual it must at each stage of its development, satisfy three main goals: maintain its own livelihood, be (structurally and functionally) in a state of transition, and be such that the end-state (or final cause) is reachable. This means that at each level of description, from the single cell to the whole developing embryo, all processes must be constrained to operate within certain norms that characterize normal development. Embryogenesis is inherently teleological where ends can explain the means, and one of the most obvious manifestations of this state of affairs is the fractal architecture of living matter: at each level of description, matter and energy needs, as well as waste elimination necessities are constraints that must be satisfied (the end), thence the architecture (the means). The study of short, medium and long-range feedback mechanisms that drive morphogenesis also provide insights about the relationship between multi-level feedback loops, closure, and the *unity* or *wholeness* of organisms. At each stage of embryonic development, a cell’s genome and its regulatory networks define the potential (future) fates of the cell, but the actual fate the cell will adopt is determined by the prevailing conditions within and at the boundaries of the embryo. Hence the parts, such as cells, are efficient causes of the whole (embryo), but the whole (including its form) constrains all the parts including the genome which thus establishes the unity that characterize organisms. Closely related to the concept of unity is the apparent contradiction between the observed simplicity and complexity of living matter. The fractal organization of living matter manifests itself by exhibiting different kinds of invariants including topological invariance as a testimony to simplicity and unity. However, organs can geometrically be very distinct even if they are topologically equivalent. Geometry highlights particular functions and thus complexity and diversification complements simplicity and unity. As in all dynamic

systems, time plays a critical role in biological systems. The very nature of the information substrate is a testimony of the way time is managed. The genetic information needs a stable “unchanging” substrate that DNA provides. Slow and transient propagating information will use molecules as its substrate and advection/diffusion as its transport means. Contemporaneously, fast and transient information will borrow the electrical modality materialized in fast-traveling action potentials and even much faster (approx. speed of light) volume conducted electrical and magnetic fields. It is interesting to note that stable genetic information is digital (in its primary structure), electrical signals are analog, and molecular signals can be viewed as somewhere in between. CM thus combines both digital and continuous/analog information storage, transformation and exchange. These are some of the fascinating topics that can and *must* be addressed for CM to be operational. The key is to explicitly identify the unifying aspects of morphogenesis in a theoretical setting that favors concepts that are universal and upon which general models can be built. Specific facts that characterize specialization and diversification can then be added and organized around a core ontology-based conceptual scaffold. Since data gaps still plague current data-intensive domains such biology, articulate and specialized AI-agents built around the conceptual scaffold can provide the support necessary to make CM a quickly evolving field from which a wealth of practical applications can be derived.

7. Conclusion

CM is a fascinating subject in itself as it can help explore what distinguishes abiotic matter from living organisms, combine theoretical and computational aspects in a single framework, as well as provide a fertile ground to study generic knowledge structures and reasoning mechanisms. From a more practical standpoint, the advantage of developing CM as a general modeling framework lies in the fact that the physical and biological principles that drive embryogenesis and growth also govern the developed organism. Thus computational models of development provide the tools for modeling normal adult homeostasis as well as disease initiation and progression (Siregar et al., 2018b). We make no assumptions about the intended roles of the generated models as these can range from pharmacokinetics and pharmacodynamics studies, synthetic biology, tissue engineering, cancer research, bio-inspired technologies or precision medicine. All these applications can benefit from powerful computational environment that, in addition to having generative competencies, can embody hypothesis-driven inference based on symbolic and quantitative knowledge coupled to bottom-up data-driven inferences such as machine learning and data mining. We believe that invariant knowledge structures on which generic hypothetico-deductive mechanisms can be performed constitute the necessary and perhaps sufficient elements to achieve this goal. During the conceptualization of our CM platform, a rather straightforward strategy came into being: to incorporate into the platform’s own architecture the distinguishing hierarchical and modular organization of living matter so that at some future point, the system’s cognitive structure could co-evolve with its object level knowledge. For now however, our CM platform is very much an early-stage prototype that needs to be fully developed in multi-disciplinary environments and deployed within high-performance computing infrastructures.

Appendix A

Examples of terms, predicates and rules concerning structural units

- Properties of structural units

Intrinsic properties of an object x are expressed by functional expressions of the form:

$f(x) = val; val \in Real/val \in Int/...$
 $f(x) = Qval; Qval = "low"/Qval = "normal"/Qval = "high"/...$
 $f(\mathbf{x}) = \mathbf{v}; \mathbf{v} \in R^n$
 $Unit(NormalRange(f(x)) \in [v_1, v_2], u)$
 ...

Examples: $Size(RBC) = 8\mu\text{m}$, $Unit(NormalRange(Size(RBC)) \in [6, 9]), \mu\text{m}$

- Relations on continuous domains

Unlike many ontologies that apply parthood in very distinct contexts, we use the parthood and proper parthood relations $Part(x,y)$ and $PPart(x,y)$ on continuous domains only. Tissues (e.g. regions of structural units) can, in certain circumstances be idealized as such, in which case the base relations defined in mereology and mereotopology apply.

$\forall x \forall y \neg PPart(x, x)$
 $\forall x \forall y \forall z PPart(x, y) \wedge PPart(y, z) \Rightarrow PPart(x, z)$
 $\forall x \forall y PPart(x, y) \Rightarrow \neg PPart(y, x)$

Example: $PPart(Interior(Lumen), Lumen)$

- Relations on discrete objects

When objects are viewed as discrete entities we use the predicate $UnitOf(x,y)$ to describe the nested hierarchies of the structural units. Like $PPart$, it is a non-reflexive, transitive and antisymmetric relation.

$\forall x \forall y \neg Subunit(x, x)$
 $\forall x \forall y \forall z Subunit(x, y) \wedge Subunit(y, z) \Rightarrow Subunit(x, z)$
 $\forall x \forall y Subunit(x, y) \Rightarrow \neg Subunit(y, x)$

Examples: $Subunit(Promoter, Gene)$, $Subunit(Gene, Genome)$, $Subunit(AcetylCoA, KrebsCycle)$

The topological connectivity between discrete objects is defined by the symmetrical relation $Connect(x,y)$ and the asymmetrical relations $Afferent(x,y)$ and $Efferent(x,y)$ with their obvious denotation. In our ontology, afferences and efferences provide an additional information with respect to connect: the notion of flow. The topology of structural units involved in the transport of mass, energy or information is expressed in these terms. The following rules hold:

Afference or efference implies symmetrical connectedness

$\forall x \forall y Afferent(x, y) \Rightarrow Connect(x, y)$
 $\forall x \forall y Connect(x, y) \Rightarrow Connect(y, x)$

Connectivity involving a nested structural unit

$\forall x \forall y \forall z Connect(x, y) \wedge \neg Subunit(x, z) \wedge Subunit(y, z) \Rightarrow Connect(x, z)$

Afference involving a nested structural unit (idem for efference)

$\forall x \forall y \forall z Afferent(x, y) \wedge \neg Subunit(x, z) \wedge Subunit(y, z) \Rightarrow Afferent(x, z)$

- Other structural relations

$BranchFrom(x, y)$, $OpenInto(x, y)$, $Invade(x, y)$, $Surround(x, y)$...

Appendix B

- Symbolic representations of processes

Processes are formally represented by predicate calculus functional expressions that can be nested to arbitrary levels. Some processes make explicit reference to time and space. They are formally functional terms, but can also be interpreted as (nested) predicates of higher-order logic.

- Spatial properties are defined by function symbols *Distance*, *Direction*, *Above*, *Front*, ...
- Temporal properties are defined by function symbols that designate time intervals, time points, and the relative position of time intervals (Before, After, Meets,...)

- Examples of formalized molecular processes*

“breaking of a phosphate group from an ATP

$BreakFrom(Phosph, ATP)$

“transfer a phosphate to a protein containing a tyrosine residue”

$TransferTo(Phosph, TyrProt)$

“breaking of a phosphate from an ATP and transfer of the phosphate to a protein containing a tyrosine residue”

Transfer(Phosph, ATP, TyrProt) ≡ BreakFrom(Phosph, ATP)&TransferTo(Phosph, TyrProt)

“catalysis of “a phosphate ...tyrosine residue” by a tyrosine kinase”

Catalysis(Transfer(Phosph, ATP, TyrProt), TyrKinase)

The occurrence of the “catalysis of “a phosphate ...” by a tyrosine kinase” at a given time interval

TimeInt(Catalysis(Transfer(Phosph, ATP, TyrProt), TyrKinase), Ti)

*Objects should be subscripted since they are instances of classes but we omitted them for readability concerns.

- Examples of formalized cellular processes

“secretion of FGF7 by mesenchyme cells surrounding a ureteric bud tip (UB)”

Secretion(Surround(MesanchCells, Tip(UB)), FGF7)

“chemotaxis of an epithelial cell in the direction of a FGF7 concentration gradient”

Chemotaxis(EpithCell, Direction(Grad(Conc(FGF7))))

“asymmetric cell division of a UB tip epithelial cell into its proper clone + a type2 cell”

AsymCellDiv(UnitOf(EpithCell, Tip(UB)), CellType2)

- An example of a formalized tissue process

“elongation of a ureteric bud in the direction of a GDNF concentration gradient”

Elongation(UreterBud, Direction(Grad(Conc(GDNF))))

Appendix C

- An example of formalized causal knowledge

“high blood glucose level will trigger (*cause*) the secretion of insulin into the blood compartment by the beta cells of the Langerhans pancreatic islets”.

Level(In(Gluc, Blood), High) → Secretion(UnitOf(BetaCells, UnitOf(LanghIslets, Pancreas)), Insulin)

In this example, parthood is defined within the functional term *UnitOf* which has the same meaning as its predicate counterpart *Subunit*.

Appendix D

- An example of probabilistic knowledge

“The probability that “UB tip epithelial cell divides asymmetrically into its proper clone + a type2 cell given a context *C*” = 0.3”

Pr(AsymCellDiv(UnitOf(EpithCell, Tip(UB)), CellType2)/C) = 0.3

Example of a context *C*

C ≡ Secretion(Surround(MesanchCells, Tip(UB)), FGF7)

References

- Allen, J.F., 1983. Maintaining knowledge about temporal intervals. *Commun. ACM* 26, 832–843.
- Amaneddine, N., Condotta, J., 2013. On the minimal labeling problem of temporal and spatial qualitative constraints. *FLAIRS Conference* 2013 16–21.
- Andrade-Cabrera, C., Turner, W.J.N., Burke, D.J., Neu, O., Finn, D., 2016. Lumped parameter building model calibration using particle swarm optimization. *3rd Asia Conference of International Building Performance Simulation Association (ASim2016)*.
- Ashburner, M., Ball, C.A., Blake, J.A., et al., 2000. Gene Ontology: tool for the unification of biology. *The Gene Ontology Consortium*. *Nat. Genet.* 25 (1), 25–29. <https://doi.org/10.1038/75556>.
- Bacchus, F., Kabanza, F., 2000. Using temporal logics to express search control knowledge for planning. *Artif. Intell.* 116, 123–191.
- Bachelard, G., 2002. The formation of the scientific mind: a contribution to a psycho-analysis of objective knowledge. *Clinamen*, Manchester.
- Bickhard, M.H., 2016. The anticipatory brain: two approaches. In: In: Müller, V.C. (Ed.), *Fundamental Issues of Artificial Intelligence* 376. Springer (Synthese Library), Berlin, pp. 259–281. https://doi.org/10.1007/978-3-319-26485-1_16.
- Bocarra, N., 2009. *Modeling Complex Systems*. Springer.
- Bodenreider, O., Stevens, R., 2006. Bio-ontologies: current trends and future directions. *Brief. Bioinform.* 7, 256–274.
- Bureau, I., Shepherd, G.M.G., Svoboda, K., 2004. Precise development of functional and anatomical columns in the neocortex. *Neuron* 42, 789–801.
- Caporale, N., Dan, Y., 2008. Spike timing-dependent plasticity: a Hebbian learning rule. *Ann. Rev. Neurosci.* 31, 25–46.
- Casati, R., Varzi, A., 1999. *Parts and Places: The Structures of Spatial Representation*. MIT Press, Cambridge, MA and London.
- Cohn, G., Hazarika, S.M., 2001. Qualitative spatial representation and reasoning: an overview. *Fundam. Inform.* 46, 1–29.
- Costantini, F., Kopan, R., 2010. Patterning a complex organ: branching morphogenesis and nephron segmentation in kidney development. *Dev. Cell* 18 (5), 698–712. <https://doi.org/10.1016/j.devcel.2010.04.008>.
- Da Silveira, M., Dos Reis, J.C., Pruski, C., 2015. Management of Dynamic Biomedical Terminologies: Current Status and Future Challenges. *IMIA Yearbook of Medical Informatics*.
- De Pillis, L.G., Radunskaya, A., 2001. A mathematical tumor model with immune

- resistance and drug therapy: an optimal control approach. *J. Theor. Med.* 3 (2), 79–100. <https://doi.org/10.1080/10273660108833067>.
- Dodig-Crnkovic, G., 2011. Significance of models of computation from turing model to natural computation. *Minds Mach.* 21, 301–322.
- Dodig-Crnkovic, G., 2012. Info-computationalism and morphological computing of information structure. In: Simeonov, P.L., Smith, L., Ehreshmann, A. (Eds.), *Integral Biomathics*. Springer Series on Computational Intelligence and Complexity.
- Dylla, F., Lee, J.H., Mossakowski, T., Schneider, T., van Delden, A., van de Ven, J., Wolter, D., 2017. A survey of qualitative spatial and temporal calculi - algebraic and computational properties. *ACM Comput. Surv. (CSUR)* 50, 1–39.
- Efron, B., Hastie, T., 2016. *Computer Age Statistical Inference*. Cambridge University Press, New-York.
- Egenhofer, M., Franzosa, R.D., 1991. Point-set topological spatial relations. *Int. J. Geogr. Inf. Sci.* 5, 161–174.
- Ehresmann, A.C., Simeonov, P.L., 2012. WLIMES, the wandering LIMES: towards a theoretical framework for wandering logic intelligence memory evolutive system. In: Simeonov, P.L., Smith, L.S., Ehresmann, A.C. (Eds.), *Tracing the Road to Reality*. Springer, Berlin, Heidelberg, pp. 105–122.
- Ehresmann, A.C., Vanbremersch, J.P., 2009. MENS a mathematical model for cognitive systems. *J. Mind Theory* 2, 129–180.
- Einstein, A., 1920. *Relativity, the Special and General Theory*. Henry Holt and Company, New York.
- Forbus, K., 1984. Qualitative process theory. *Artif. Intell.* 24, 85–168.
- Genesereth, M., Nilsson, N.J., 1987. *Logical Foundations of Artificial Intelligence*. Morgan Kaufmann, San Mateo, CA.
- Gillespie, D.T., 1976. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.* 22, 403–434.
- Gillespie, D.T., Petzold, L.R., 2003. Improved leap-size selection for accelerated stochastic simulation. *J. Chem. Phys.* 19, 8229–8234.
- Gilmour, D., Rembold, M., Leptin, M., 2017. From morphogen to morphogenesis and back. *Nature* 541, 311–320.
- Glez-Cabrera, F.J., Álvarez-Bravo, J.V., Díaz, F., 2013. QRPC: a new qualitative model for representing motion patterns. *Expert Syst. Appl.* 40, 4547–4561.
- Goethelf, A., 2012. *Teleology, First Principles, and Scientific Method in Aristotle's Biology*. Oxford University Press, New York.
- Gupta, P.B., Fillmore, C.M., Jiang, G., Shapira, S.D., Tao, K., Kuperwasser, C., Lander, E.S., 2009. Stochastic State Transitions Give Rise to Phenotypic Equilibrium in Populations of Cancer Cells. *Cell* 146, 633–644.
- Hennig, B., 2009. The four causes. *J. Philos.* 106, 137–160.
- Igamberdiev, A.U., 2015. Anticipatory dynamics of biological systems: from molecular quantum states to evolution. *Int. J. Gen. Syst.* 44 (6), 631–641.
- Isaeva, V.V., Kasyanov, N.V., Presnov, E.V., 2012. Topological singularities and symmetry breaking in development. *Biosystems* 109, 280–298. <https://doi.org/10.1016/j.biosystems.2012.05.004>.
- Jackson, D.E., Ratnieks, F.L.W., 2006. Communication in ants. *Curr. Biol.* 16, R570–R574.
- Kalogeris, T., Baines, C.P., Krenz, M., Korthuis, R.J., 2012. Cell biology of ischemia/re-perfusion injury. *Int. Rev. Cell Mol. Biol.* 298, 229–317.
- Kanso, A., Gromaire, M.C., Gaume, E., Tassin, B., Chebbo, G., 2003. Bayesian approach for the calibration of models application to an urban stormwater pollution model. *Water Sci. Technol.* 47, 77–84.
- Kleene, S.C., 1952. *Introduction to Metamathematics*. North Holland, New-York.
- Lanczos, C., 1986. *The Variational Principles of Mechanics*, 4th ed. Dover Publications.
- LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *Nature* 521, 436–444.
- Levesque, H., Pirri, F., Reiter, R., 1998. Foundations for the situation calculus. *Electron. Trans. Artif. Intell.* 2, 159–178.
- Louie, A.H., 2009. More Than Life Itself: A Synthetic Continuation in Relational Biology. Ontos Verlag, Frankfurt. <https://doi.org/10.1515/9783110321944>.
- Maturana, H.R., Varela, F.G., 1980. *Autopoiesis and Cognition: The Realization of the Living*. Boston Studies in the Philosophy of Science. Reidel, Dordrecht.
- Mayer, E., 1976. *Evolution and the Diversity of Life*. Belknap Press, Cambridge, MA.
- McDermott, D.V., 1982. A temporal logic for reasoning about processes and plans. *Cogn. Sci.* 6, 101–155.
- McDermott, D.V., Doyle, J., 1980. Non-monotonic logic I. *Artif. Intell.* 13, 41–72.
- McQuarie, D.A., Simon, J.D., 1997. *Physical Chemistry. A Molecular Approach*. University Science Books.
- Miller, T., 2017. Explanation in Artificial Intelligence: Insights from the Social Sciences. CoRR. <http://arxiv.org/abs/1706.07269>.
- Montévil, M., Mossio, M., 2015. Biological organisation as closure of constraints. *J. Theor. Biol.* 372, 179–191.
- Mossio, M., Bich, L., Moreno, A., 2013. Emergence, closure and inter-level causation in biological systems. *Erkenntnis* 78, 153–178.
- Mountcastle, V.B., 1997. The columnar organization of the neocortex. *Brain* 120, 701–722.
- Nasuto, S.J., Hayashi, Y., 2016. Anticipation: beyond synthetic biology and cognitive robotics. *Biosystems* 148, 22–31. <https://doi.org/10.1016/j.biosystems.2016.07.011>.
- Nicholson, D., 2014. The return of the organism as a fundamental explanatory concept in biology. *Philos. Compass* 9, 347–359.
- Noble, D., 2012. A theory of biological relativity: no privileged level of causation. *Interface Focus* 2 (1), 55–64. <https://doi.org/10.1098/rsfs.2011.0067>.
- Osborne, A., 2017. The role of epigenetics in human evolution. *Biosci. Horiz.* 10. <https://doi.org/10.1093/biohorizons/hzx007>.
- Pearl, J., 2009. Causal inference in statistics: an overview. *Stat. Surv.* 3, 96–146.
- Pearl, J., 2014. The deductive approach to causal inference. *J. Causal Inference* 2, 115–129.
- Perri, F., Longo, F., Giuliano, M., et al., 2017. Epigenetic control of gene expression: potential implications for Cancer treatment. *Crit. Rev. Oncol. Hematol.* 111, 166–172. <https://doi.org/10.1016/j.critrevonc.2017.01.020>.
- Phillips, A.C., 2003. In: Sandiford, D.J., Mandl, F., Phillips, A.C. (Eds.), *Introduction to Quantum Mechanics*. John Wiley & Sons.
- Popper, K., 1963. *Conjectures and Refutations*. Basic Books publishers, New York-London.
- Presnov, E., Isaeva, V., Kasyanov, N., 2014. Topological invariance of biological development. *Axiomathes* 24, 117–135. <https://doi.org/10.1007/s10516-013-9216-5>.
- Prigogine, I., Stengers, I., 1979. *La Nouvelle Alliance*. Gallimard, Paris.
- Ribeiro, M.G.L., Larentis, A.L., Caldas, L.A., Garcia, T.C., Terra, L.L., Herbst, M.H., Almeida, R.V., 2015. On the debate about teleology in biology: the notion of “teleological obstacle”. *História Ciências Saúd –Manguinhos* 22, 1321–1333.
- Rosen, R., 1985. *Anticipatory Systems: Philosophical, Mathematical and Methodological Foundations*. Pergamon Press, Oxford.
- Rosen, R., 1991. *Life Itself: a Comprehensive Inquiry into the Nature, Origin, and Fabrication of Life*. Columbia University Press, New York.
- Russell, S., 2015. Unifying logic and probability. *Commun. ACM* 58, 88–97.
- Sharma, H., Herze, N., Lohmann, S., Kayser, K., Hellwich, O., Hufnagl, P., 2015. A review of graph-based methods for image analysis in digital histopathology. *Diagn. Pathol.* 1, 61.
- Shields, C., 2017. What organisms once were and might yet be. *Philos. Theor. Pract. Biol.* 9 (7).
- Simeonov, P.L., Brezina, E.H., Cottam, R., et al., 2012. Stepping beyond the newtonian paradigm in biology towards an integrable model of life: accelerating discovery in the biological foundations of science. In: Simeonov, P.L., Smith, L.S., Ehresmann, A.C. (Eds.), *Tracing the Road to Reality*. Springer, Berlin, Heidelberg, pp. 319–417.
- Simeonov, P.L., Gomez-Ramirez, J., Siregar, P., 2013. On some recent insights in Integral Biomathics. *Prog. Biophys. Mol. Biol.* 113, 216–228.
- Siregar, P., 1996. Theoretical cardiology: from mathematical to qualitative models. *J. Biol. Sys.* 4, 131–150.
- Siregar, P., 2000. A Method for the three-dimensional construction of a virtual organ representing a real organ. European patent n° WO2000072272 A1.
- Siregar, P., (2009). Simulation of complex Systems. European Patent n° 08290535.7.
- Siregar, P., Chahine, M., Lemoulec, F., Le Beux, P., 1995. An interactive qualitative model in cardiology. *Comput. Biomed. Res.* 28, 443–478.
- Siregar, P., Julen, N., 2016a. Cross-fertilization between computational morphogenesis and digital pathology. *Diagn. Pathol.* 1 (8).
- Siregar, P., Julen, N., 2016b. Virtual patients and serious games in medicine. *Diagn. Pathol.* 1 (8).
- Siregar, P., Julen, N., Hufnagl, P., Mutter, G., 2018a. A general framework dedicated to computational morphogenesis Part I – Constitutive equations. *Biosystems*. <https://doi.org/10.1016/j.biosystems.2018.07.003>.
- Siregar, P., Julen, N., Hufnagl, P., Mutter, G.L., 2018b. Computational morphogenesis – Embryogenesis, cancer research and digital pathology. *Biosystems* 169–170, 40–54. <https://doi.org/10.1016/j.biosystems.2018.05.006>.
- Siregar, P., Julen, N., Sinteff, J.P., 2003. Computational integrative physiology: at the convergence of the life, physical and computational sciences. *Methods Inf. Med.* Suppl. 42, 177–184.
- Siregar, P., Mabo, P., Coatrieux, J.L., 1993. How can deep knowledge be used in CCU monitoring? Special issue on intelligent monitoring systems. *IEEE Eng. Med. Biol.* 4, 92–99.
- Siregar, P., Sinteff, J.P., 1996b. Introducing Spatio-temporal reasoning into the inverse problem in electroencephalography. *Artif. Intell. Med.* 8, 97–122.
- Siregar, P., Sinteff, J.P., Chahine, M., Le Beux, P., 1996. A Cellular Automata of the Heart and its coupling with a Qualitative Model. *Comput. Biomed. Res.* 29, 222–246.
- Siregar, P., Sinteff, J.P., Julen, N., Le Beux, P., 1997. Spatio-temporal reasoning for multi-scale modeling in cardiology. Special Issue on Cardiovascular Systems. *Artif. Intell. Med.* 10, 41–57.
- Smith, B., 2003. The logic of biological classification and the foundations of biomedical ontology. *Proc. 10th International Conference in Logic Methodology and Philosophy of Science*.
- Smith, B., Williams, J., Schulze-Kremer, S., 2003. The ontology of the Gene ontology. *Proc. Annual Symposium of the American Medical Informatics Association*. pp. 609–613.
- Spear, A.D., 2016. Functions in basic formal ontology. *Appl. Ontol.* 11, 103–128.
- Teague, B.P., Guye, P., Weiss, R., 2016. Synthetic morphogenesis. *Cold Spring Harb. Perspect. Biol.* 8, a023929.
- Tett, S.F.B., Yamazaki, K., Mineter, M.J., Cartis, C., Eizenberg, N., 2017. Calibrating climate models using inverse methods: case studies with HadAM3, HadAM3P and HadCM3. *Geosci. Model Dev.* 10, 3567–3589. <https://doi.org/10.5194/gmd-10-3567-2017>.
- Tortora, G., Derrickson, B., 2006. *Principles of Anatomy and Physiology*, 11th ed. John Wiley & Sons.
- Varzi, A.C., 1994. On the boundary between mereology and topology. In: Casati, R., Smith, B., White, G. (Eds.), *Philosophy and the Cognitive Sciences*. Hölder-Pichler-Tempky, Vienna, pp. 423–428.
- Varzi, A.C., 1996. A Parts, wholes and part-whole relations: the prospects of mereotopology. *Data Knowl. Eng.* 20, 259–286.

- Vize, P., Woolf, A., Bard, J., 2003. The Kidney: from Normal Development to Congenital Diseases. Academic Press, Amsterdam.
- Waddington, C.H., 1968. Towards a theoretical biology. *Nature* 218, 525–527.
- Wiener, N., 1948. Cybernetics, or Control and Communication in the Animal and the Machine. The MIT Press.
- Wolpert, L., Beddington, R., Jessell, T., Lawrence, P., Meyerowitz, E., Smith, J., 2002. Principles of Development, second edition. Oxford University Press ISBN 0-19-924939-3.
- Zadeh, L., 1983. A computational approach to fuzzy quantifiers in natural languages. *Comput. Math. Appl.* 9, 149–184.