MEDINFO 2004
M. Fieschi et al. (Eds)
Amsterdam: IOS Press


# Evolution of a Foundational Model of Physiology: Symbolic Representation for Functional Bioinformatics

**Daniel L. Cook [a], Jose L. V. Mejino [b], Cornelius Rosse [b,c]**

*Structural Informatics Group*
*Departments of [a]Physiology & Biophysics, [b]Biological Structure, and*
*[c]Medical Education and Bioinformatics, University of Washington, Seattle, WA, USA*

**Abstract**

*We describe the need for a Foundational Model of Physiology (FMP) as a reference ontology for "functional bioinformatics". The FMP is intended to support symbolic lookup, logical inference and mathematical analysis by integrating descriptive, qualitative and quantitative functional knowledge. The FMP will serve as a symbolic representation of biological functions initially pertaining to human physiology and ultimately extensible to other species. We describe the evolving architecture of the FMP, which is based on the ontological principles of the BioD biological description language and the Foundational Model of Anatomy (FMA).*

***Keywords***

Functional bioinformatics, knowledge representation, ontology, physiology, virtual human, computational modeling.

## Introduction

The cloning of the human genome and the success of bioinformatics to computationally represent genomic structure and function has amplified long-held beliefs that computational methods will be immensely beneficial, if not essential, for representing and analyzing the structure and function of entire humans. "Digital" or "virtual" humans have been envisaged as critical research and clinical tools that can ultimately provide comprehensive, searchable knowledge bases and an integrated platform for biosystem analysis and simulation. A computable representation of physiologic functions that is generalizable and scalable from molecules (e.g., enzymatic catalysis, intermolecular binding) to an entire organism (e.g., food- and mate-seeking behavior) as well as its macroscopic parts, must be a key component of such a biosystem but, to date, its specifics remain to be articulated. In this preliminary report we outline the elements of such an abstraction for representing physiological functions and propose an ontological architecture for a Foundational Model of Physiology (FMP) that we will develop as a companion to the existing Foundational Model of Anatomy (FMA) [1].

## Challenges of Functional Bioinformatics

The overwhelming scale and complexity of human biology presents daunting representational challenges because biomedical knowledge is largely qualitative, not quantitative, and because its roots in multiple disciplines share few organizing principles.

Where clear organizing principles exist, however, progress can be spectacular. Bioinformatics was itself born in the genome sciences by exploiting two key principles: 1) the "central dogma" of biology — genes encoded as DNA are transcribed to RNA which are translated to proteins that have biological functions, and 2) symbolic representations of DNA, RNA and protein sequences and functions form the computational basis for archiving and analyzing genome structure and function. Intermolecular binding is the organizing principle for a variety of ligand-receptor databases while linked biochemical reactions is the organizing principle for metabolic and cell-signaling pathway databases. Built on differing principles and for a variety of purposes, however, such knowledge resources have evolved into a segregated mix of knowledge dialects that resist unification for lack of common organizing principles.

Mathematical modeling (largely in terms of differential equations) has long been used to represent and analyze physiological functions but the lack of quantitative data severely restricts mathematical modeling to only a few very well-studied problems. Structures and functions critical to understanding human biology are often poorly defined, poorly understood and are known only in terms that are descriptive, tentative and often speculative. In such a knowledge environment, mathematics as a representational language is not well-suited for the kinds of qualitative queries that are the currency of most bioscience discourse: "Are there instances where hormone secretion is stimulated by adrenalin?", "What is the effect of myosin gene expression on cardiac contractility?", etc.

Thus, it is our belief that any bioinformatical approach to generating, reusing and accessing functional biomedical knowledge must necessarily rest on a logically structured symbolic representation that includes the descriptive, qualitative and quantitative aspects of physiology. Furthermore, such a symbolic representation must provide a framework that readily embraces all organizational levels from the molecular to the organismal. Toward this end, we describe progress in developing a Foundational Model of Physiology that is "foundational" in the same sense as is the existing Foundational Model of Anatomy — it is intended to be reused as a template for the development of more specific domain ontologies in biology and medicine, and to sup-

port the development on knowledge-based applications that call for physiological information. Moreover, the FMP, like the FMA, is based on broad organizing principles, which will make it possible to integrate quantitative mathematical models of physiological function with logical symbolic representations.

## Hypothesis

Our hypothesis, very broadly stated, is that an integrated symbolic and mathematical representation of human functions can be based on three overarching organizing principles:

1. *Anatomical principle*: A canonical description of the structural organization of the entire human body can be represented as an ontology of anatomical entities; as in the FMA [1].
2. *Physiological principle:* Descriptions of human functions at all organizational levels can be represented in terms of an ontology of functional relationships between anatomical entities; as proposed here for the FMP.
3. *nstantiation principle:* The instances of the canonical FMA and FMP can be instantiated for analyzing and simulating individual cases by: a) associating qualitative and quantitative values with spatial and non-spatial physical properties of anatomical entities, and b) predicting functional outcomes using logical and mathematical expressions that represent cause-effect relationships.

We anticipate that such a foundational knowledge representation can support three fundamentally different kinds of queries: 1) *Descriptive:* What functions does the pancreas have? What cells secrete insulin? 2) *Qualitative logical inferences:* If insulin secretory rate is increased, what blood metabolite levels change? Find all negative feedback paths controlling insulin secretion. 3) *Quantitative simulation and analysis:* If 100 gm of glucose is infused in 20 min, what peak blood glucose level results? What are the control gains of the negative feedback paths?

## Background

Our hypothesis and organizing principles stem from a convergence of ongoing efforts at the University of Washington (UW) to develop structural and functional ontologies for biology. On one tack, BioD has been developed by one of us (DLC) as a schematic diagramming language for describing and analyzing the function of complex, biological systems that can only be studied in a multidisciplinary framework [2]. On the other tack, the Structural Informatics Group at the UW has created the Digital Anatomist FMA [1] as a canonical symbolic model of human of anatomy. While separately conceived, these approaches have shared a common vision of formalizing and generalizing symbolic representations of biological structure and function as open-source knowledge resources for the development of knowledge-based applications in structural and functional bioinformatics.

## Representation of Structure

Both BioD and the FMA represent anatomical structures (including microscopic and molecular entities) in terms of extensible class subsumption inheritance (*is a*) hierarchies or taxonomies, illustrated by a portion of the BioD taxonomy:

*chemical species*
*molecule*
*molecular functional site*
*catalytic site*
*kinase site*
*occupancy site*
*phosphorylation site*

Both systems build complex attributed graphs as *part of* hierarchies as well, where a functional site (e.g., a catalytic site) on a molecule (i.e., a *molecular functional site*) can, for instance, be a part of a *molecule*, but not *vice versa*. In BioD's diagrammatic approach, containment and connectivity relationships are represented graphically. In the FMA, an Anatomical Structural Abstraction (ASA) provides symbolic representations for spatial and topological relationships (containment, connectivity, etc.) and for shape and dimensional attributes. The similar representations of structure (including both *is a* and *part of* hierarchies) in BioD and the FMA, suggested to us that BioD's graphical representation of functional relationships could map to an ontological architecture patterned after the FMA.

## Representation of Function

BioD is based on the linguistic hypothesis that cause-effect relationships between biological structures can be expressed as syntactically simple noun-verb-noun sentences such as "Molecule A phosphorylates molecule B". In a sense, BioD simply formalizes the graphical grammar of diagrams in which icons represent biological objects and action arrows represent cause/effect relationships between biological objects. The BioD model in Figure 1 shows a *molecule* of type A which has, as its parts, *kinase functional sites* (K) that *phosphorylate* (arrow) the *phosphorylation functional sites* (P) that are parts of type B substrate *molecule.*

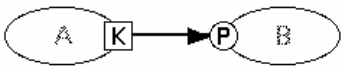


*Figure 1 - A simple BioD model of phosphorylation*

Fundamental to BioD is the idea that the *physical state* of an anatomical or physical entity can be defined as the set of the values of measurable *physical properties* of the entity. Accordingly, BioD nouns are assigned one or more *physical properties* that include both spatial (location, parts, etc.) and physical (mass, size, temperature, enzymatic activity, etc.) *physical properties.*

The *kinase functional site* represented in Figure 1 has two functional properties: an *amount* property (its value is inferred from the *amount* of its parent *molecule* A, and an *activity* property representing the degree of catalytic activation (as percent of maximal). In addition to *amount* and *activity* properties, degree of phosphorylation of a *phosphorylation site* is represented by an *occupancy* property (0–100%). Depending on prevailing knowledge or purposes of an analysis, functional properties can be as-

signed values that are either qualitative (e.g., present/absent) or quantitative (e.g., 1 mM).

*Action arrows* represent cause/effect relationships by which the *physical properties* of one entity change the *physical properties* of itself or of other entities. For instance, increasing the *amount* of the kinase molecule A would increase the *occupancy* of the *phosphorylation site* on B. Such *action* relationships are represented diagrammatically as arrows that are drawn from an inheritance hierarchy of *actions* in which, for instance, *phosphorylates* is a descendent of *reacts chemically* (i.e., *reacts chemically -> attaches covalently –> phosphorylates*).

A key concept in BioD is that *actions* require one or more *role players* that must satisfy a set of semantic linking rules which restrict the class and context for candidate *role players*. For instance, the *phosphorylates* verb requires 2 *role players*: a *kinase functional site* and a *phosphorylation functional site* which must coexist in a common *compartment*; i.e., they must have direct physical access to each other.

Verbs are specific for the functional properties they affect. For example, in Figure 2 the *molecule* labeled "gene" increases the *amount* of *molecule* A via a *produces action* with a corresponding increase in the *amount* of A's *kinase site*. The *activity* of the *kinase site* can, however, be independently *inhibited* when binding of a regulatory *molecule,* R, increases the *occupancy* of its *binding site* on *molecule* A.

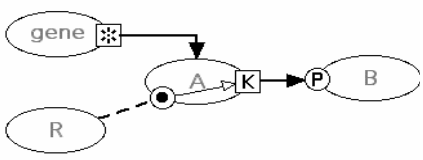


*Figure 2 - Dual regulation of a kinase.*

For simulation and analysis, each BioD verb defines both logical and mathematical computational expressions for calculating how the action affects the properties of its role players. Verbs have functional parameters that set the rate or extent of the changes that they impose on the functional properties of their role players. Thus, BioD is designed to support "hybrid" qualitative/quantitative analyses as a way to exploit qualitative knowledge while constraining predictions with quantitative knowledge as it becomes available.

We are developing a prototype Java application, Chalkboard, to test BioD principles and methods for creating and editing graphical models and for performing hybrid qualitative/quantitative analysis. Implemented as a "computer-aided design" application, a Chalkboard palette provides BioD noun icons that can be linked by verb arrows for which the nouns represent *role players* in the *actions* represented by arrows. Linking is constrained by semantic rules for each verb object while linking automatically connects the verb's computational expressions to the noun's functional properties.

In Chalkboard, representations of complex regulatory networks can be quickly built and analyzed. For example, a qualitative "Path Tracing" tool tracks the functional consequences of incrementing or decrementing functional property values (Figure 3).

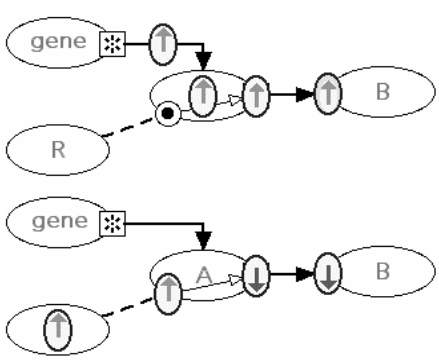


*Figure 3 - The Chalkboard "Path Tracing" tool applied to the model of Figure 2.* (Top) An increase (up-arrow) in the expression rate of *molecule* A increases the *amount* of A's *kinase site*. This increases the *occupancy* of *molecule* B's *phosphorylation site*. (Bottom) Increasing the *amount* of *molecule* R increases the *occupancy* of its *binding site* on A and reduces the *activity* (down-arrow) of A's *kinase site* and thus reduces the *occupancy* of B's *phosphorylation site.*

## Basic features of the FMP

The similar representations of structure in BioD and the FMA in terms of *is a* and *part of* hierarchies suggested that BioD's "vocabulary" could be elaborated and extended to the scale of the FMA. Figure 4 diagrams the key parts and relationships of the current vision of the FMP and the existing FMA (gray boxes), which provide the ontological model for developing the FMP (white boxes).

Central to the FMA is its Anatomical Taxonomy (AT), an *is a* hierarchy of anatomical entities whose subclasses include *body, organ, organ part, cell,* etc. To represent how these parts relate to each other, each AT entity participates in one or more structural relationships (SR), which collectively form an Anatomical Structural Abstraction (ASA). A key set of SRs form the basis for a total-body partonomy whereby *body parts* are *parts of bodies*, *organ parts* are *parts of organs*, and *cells* are *parts of organ parts*, and so on. The FMA currently represents an extensive body of canonical anatomical knowledge in the frame-based knowledge modeling environment Protégé [3] with over 70,000 AT concepts linked by more than 1.6 million ASA relationships.

To represent physiological actions we have designed the FMP to mirror and extend the FMA (Figure 4.) First, because the AT contains only anatomical entities that are biologically-derived, a Physical Entity Taxonomy (PET) is required to represent those non-anatomical entities (such as ions, essential nutrients, respiratory gasses, etc.) that participate in physiological actions, but are not themselves products of coordinated gene expression.

Second, a Physical Property Taxonomy (PPT) will provide symbolic representations of non-spatial physical state properties (mass, temperature, etc.) for entities of both the AT and PET. Note that while the ASA provides symbolic representations of

spatial dimensions, it does not represent values for these dimensions as are needed for quantitative and qualitative computations. Thus, as in BioD, each PPT property can be assigned slots for both qualitative (e.g., “absent" / ”present”) and quantitative (e.g., 1 mM) values.

The FMP will represent cause/effect relationships between entities (both anatomical and physical) as Physiological Actions (PA; that correspond to BioD verbs). PAs take on *has role player* relationships with both AT and PET entities (Figure 4). For example, a *secrete* action for a secretory cell has three role players: the *secretory product*, the *content*s of the *secretory cell* (i.e., the source of *secretory product*) and the *contents* of an *extracellular space* (i.e., the destination of *secretory product*). The *secrete* physiological action represents the movement of the *secretory product* (whether a *protein* of the AT or an *ion* of the PET, or both) from the cell *contents* to the extracellular *contents.*

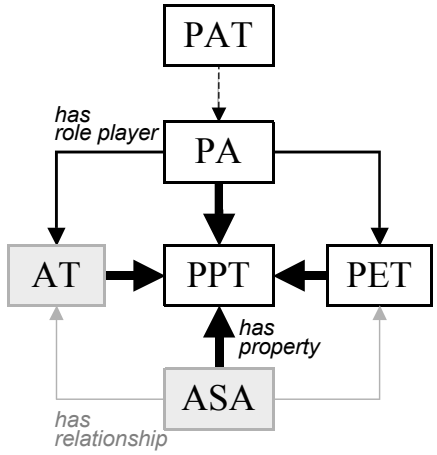


*Figure 4 - Key parts and parallel relationships within and between the FMP and FMA.* The proposed FMP (white boxes) consists of a Physiological Action Taxonomy (PAT) of Physiological Actions (PA), a Physical Entity Taxonomy (PET) and a Physical Property Taxonomy (PPT). The existing FMA (gray boxes) consists of an Anatomical Taxonomy (AT) and an Anatomical Structural Abstraction (ASA) that is a collection of Structural Relationships (SR).

PAs will be frame-based constructs that have a slot for each role player so that the PA has access to role player property values. In order to make quantitative and qualitative predictions, PAs will have: 1) a slot for a mathematical “action” function that returns a quantitative value for the action (e.g., a secretory flow rate) that depends on the physical property values of the role players (e.g., *concentration* of *secretory product*), 2) slots for quantitative functional parameters (e.g., a *diffusion rate constant* of the *parameter* class of the PPT) that are required by the action function and 3) a slot for a logical function (e.g., If the amount of A increases then the amount of B increases.) that returns a qualitative value (e.g., fast/slow) that predicts the effect of the action. In this manner, Physiological Actions will support both quantitative and qualitative analyses of physiological systems within a single framework.

## Hierarchical Organization of the FMP

An important feature of the proposed FMP is that both the PPT and PAT will be hierarchically organized as *is a* hierarchies that parallel the hierarchical organization of anatomical entities in the AT. Physiological Properties such as the *amount* of a *secretory product* will depend on the hierarchical level of the anatomical entity: the *amount* property of a *secretory vesicle* would be defined in terms of the *volume* of the *vesicle* times the *concentration* of *secretory product* in the *vesicle*; the *amount* in the *cell* defined as the sum of the *vesicle amounts* for all *vesicles* in the *cell*; and the *amount* in an *organ* defined as the sum of the *amounts* in the *organ's cells*. Properly defined, the accrual of extensive properties (e.g., mass, volume) according to hierarchical level can apply to all manner of quantitative analyses that take into account, for example, *cell-cell* variations of *vesicle number.*

The PPT will support two other kinds of property. *Relational properties* of an entity will be defined as relationships between the properties of the entity's parts. For instance, an arm's *orientation* property would depend on the independent *orientations* of the arm's bones. *Derived properties* will be defined to support more complex cases. The *posture* of a *body* could be defined in terms of a combination of specific *orientations* and *relative locations* of several *body parts.*

PAs are defined in terms of the hierarchical level of the role players in the action. The *cell*-level *secrete* action, as above, is defined in terms of the contents of *cell* and its *extracellular spaces.* The *organ*-level *secrete* action, however, is defined in terms of the secretory *organ*, and the contents of its arterial and venous supplies. At the *cell organelle* level, the *secrete* action is defined in terms of *secretory vesicles* and the *cell membrane.*

While the *secrete* action at one level is an aggregate of the rate at the next lower level, understanding how other actions accrue from the actions of an entity's parts will be challenging. For example, the *pump* action of the *heart* is the result of the concerted actions of many parts, none of which has a *pump* action itself. Cellular mitosis is another action, at the *cell* level, that must be carefully defined symbolically in terms of the actions of *cell* parts that participate in the action.

## Discussion

In this communication we introduce the Foundational Model of Physiology as a scaleable ontological framework capable of comprehensively representing canonical knowledge of human physiology. The FMP derives from a convergence of the BioD biological description language and the Foundation Model of Anatomy — heretofore independent but parallel approaches to biological knowledge representation. We anticipate that the combined FMA-FMP will be an open resource for a wide range of domain experts for expressing, organizing and analyzing biomedical phenomena in novel ways.

The FMP differs in fundamental ways from other approaches to representing functional knowledge of biological systems. The Gene Ontology [4] (GO), for example, annotates gene sequences with terms from ontologies for Molecular Function (e.g., “hormone binding” or “kinase activity”) or Biological Processes (e.g., “hormone secretion” or “adult walking behavior”), but the

GO does not attempt to relate high-level actions to constituent actions on an anatomical basis as we propose. A number of extensible markup language (XML) applications are currently available to support the uniform description and interchange of aspects of physiological function. While these XML-based languages such as the Systems Biology Markup Language [5] (SBML) and the Physiome Markup Languages [6] [Cell Markup Language (CellML), Field Markup Language (FieldML) and Anatomy Markup Language (AnatML)] offer considerable power for exchanging biological knowledge, none are designed, as are the FMA and FMP, primarily as ontologies. It will be important, however, that FMA/FMP development be coordinated with these XML applications to maximize representational power and to assure compatibility.

Our overall goal for the FMP is to establish a symbolic representation of human physiology as a knowledge resource for the development of research and clinical applications and tools that require access to a comprehensive source of human functional knowledge. By symbolic integrating descriptive representations with quantitative and qualitative representations, our design goal is to create an FMP that is an extensible knowledge base for human biology, which supports symbolic queries as well as mathematical and logical analyses of complex human biological system.

**Acknowledgements.**

This work was supported in part by grant GM064433 from the National Institute of General Medical Sciences of the NIH.

## References

[1] Rosse C, Mejino JVL. A reference ontology for biomedical informatics: the Foundational Model of Anatomy. *J Biomed Inform* 2003: 36(6): 478-500.

[2] Cook DL, Farley JF, Tapscott SJ. A basis for a visual language for describing, archiving and analyzing functional models of complex biological systems. *Genome Biol* 2001: 2(4):RESEARCH0012.

[3] Musen M, Crubezy M, Fergerson R, Noy N, Tu S, Vendetti J. Protégé-2000. 2003: http://protégé.stanford.edu/

[4] GO Consortium. Gene Ontology™ Consortium. 2003: http://www.geneontology.org/

[5] Systems Biology Workbench Development Group. The Systems Biology Markup Language (SBML), 2003: http://www.sbw-sbml.org/docs/index.html.

[6] Hedley W. The Physiome Markup Languages, 2003: http://www.bioengin.auckland.ac.nz/physiome.markup.php/

**Address for correspondence**

Daniel L. Cook, MD, PhD: raintown@halcyon.com