An Approach to the Anatomical Correlation of Species through the Foundational Model of Anatomy

Ravensara S. Travillian MA[1,2]**; Cornelius Rosse MD, DSc**[1,2,3]**; Linda G. Shapiro PhD**[1,2,4]

[1]Structural Informatics Group, [2]Dept. of Medical Education & Biomedical Informatics,
[3]Dept. of Biological Structure, [4]Dept. of Computer Science & Engineering;
University of Washington, Seattle, WA 98195

The increasing need for extrapolating information from one species to another has been highlighted by contemporary research in bioinformatics, genomics, proteomics, and animal models of human disease, as well as other fields. We propose an approach to correlating the anatomy of Homo sapiens *with selected species, using the Foundational Model of Anatomy (FMA) as a framework, and graph matching as a method, for determining similarities and differences in the nodes and relationships (edges) defined by the attributed graph of the FMA. We illustrate our approach by comparing anatomical structures of mouse and human that present prototypical mapping problems.*

INTRODUCTION

Manifestations of normal physiologic function, as well as of disease processes, may be regarded as attributes of anatomical structures ranging in size and complexity from biological macromolecules to cells, tissues, organs, and organ systems. Therefore, we contend that the first correlations that must be established between species should be concerned with their structure or anatomy. Moreover, the rapidly emerging databases and knowledge bases that are evolving as reusable resources in bioinformatics mandate that both the species-specific structural information and the interspecies similarities and differences be navigable by computational methods. We hypothesize that the frame-based ontology of the Foundational Model of Anatomy (FMA)[1,2] furnishes a comprehensive set of concepts and relationships for correlating human anatomy, at all levels of structural organization, with the anatomy of any mammalian or vertebrate species. This contention is strengthened by the highly-conserved groups of structural genes that regulate the establishment of the body plan (*Bauplan*) of all vertebrates during their embryonic development. This genetically determined grand design accommodates the anatomical variations on the basis of which not only species, but also individual members of a species, may be distinguished, and these variations are also genetically determined. The challenge is to develop a correlated symbolic and computational model capable of representing species-specific embellishments of the basic vertebrate *Bauplan* without having to generate a separate abstraction for each of the species.

We begin by illustrating the levels of structural organization implemented in the FMA, discuss how to compare structures, and then propose an approach to mapping, as a prototype, instances of mouse anatomy to human anatomy. Before drawing our conclusions we illustrate our approach by comparing anatomical structures of mouse and human that present prototypical mapping problems.

STRUCTURAL LEVELS IN THE FMA

Similarities and differences between species exist and must be dealt with at all levels of structural organization. For example, the human hand, the mouse's paw and the horse's foot have many distinguishing features, yet at a high level of abstraction they are all similar in that they all retain the basic structural pattern of the terminal segment of the free limb of any vertebrate. In contrast, although epithelial cells lining the alveoli of the human and mouse lung may look entirely alike even by electron microscopy, molecular complexes of histocompatibility antigens inserted in their cell membrane distinguish them as mouse and human. Consequently it is important to specify the levels of both the abstraction and resolution at which the anatomical correlations are to be made.

Although the FMA has been developed and instantiated for human anatomy, its *Anatomy Taxonomy* (AT) component includes high-level, abstract classes that correspond to the generalized vertebrate *Bauplan*. For example, both hand and foot are represented as kinds of `Terminal segment of free limb`, a class that could equally subsume the mouse's paw or the horse's foot. In its second component (*Anatomical Structural Abstraction* or ASA) the FMA currently represents the anatomical characteristics of the human hand and human foot through relationships modeled as attributes and attribute values. Many of these attributes would correspond to those of the mouse's paw and the horse's foot, and only the values of certain attributes would change. Even the attribute values would remain the same for the mouse and human pulmonary alveolar cell, except for those of the histocompatibility antigen complexes.

The FMA is ideally suited for correlating the anatomy of different species, because the classes of the AT reflect the levels of structural organization. `Anatomical structure` is declared as the dominant class of the AT, and its subclasses `Organ`, `Cell` and `Biological macromolecule` are considered as the units of structural organization. Other anatomical structures either constitute cells (`Cell part`) or organs (`Organ part`), or are constituted of cells (`Tissue`) or organs (`Body part`, `Organ system`, and `Anatomical set`). As we illustrate, comparisons can and need to be made at each of these levels, as well as across some levels. The frame-based Protégé-2000 knowledge-acquisition system[3], in which the FMA is implemented, facilitates the comparison of concepts, their attribute values, and the relationships exhibited.

APPROACH

Because much of the current interest in animal models of human cancer focuses on gene expression patterns in transgenic mice[4], we set out to compare the anatomy of the mouse prostate and mammary gland with their human equivalents. Using the FMA as a template, we implemented in Protégé-2000 symbolic models of mouse anatomy, based on published as well as primary data. We relied on this exercise to define the problems we had to solve for mapping anatomical structures of a non-human species to human anatomy.

In Protégé, the concepts are frames, the names of their attributes are slots, and the values of those attributes are slot-values. Similarities and differences can occur at the representation levels of frame (concept), slot (attribute) and slot-value (attribute value). In this study we rely predominantly on concept comparisons and the inverse *-has part-* and *-part of-* ASA relationships represented by inverse slots.

In our comparisons we have found interesting similarities and differences for both the mammary gland and prostate, as well as for other organs, at all levels of structural organization. The human prostate meets the FMA definition of `Organ`; in the mouse, however, each of five anatomical structures can be identified as `Prostate`, and each is definable as an organ and distinguished by a different name (*e.g.*, `Right` and `Left dorsolateral prostate`). These five organs meet the FMA definition of `Anatomical set`, and therefore become a subclass `Set of prostates`. Consequently, in the FMA the human prostate is an `Organ`, whereas the mouse prostate is an `Anatomical set` and each member of the set is an `Organ`.

The largest organ parts into which the human prostate decomposes are lobes, whereas there are no lobes in mouse prostates. Both human and mouse prostates, however, are constituted of prostatic ducts,

stroma and capsule. It is on the basis of the prostatic ducts and their anatomical parts (layers of their wall, types of cells) that correspondences can be established between the lobes of the human prostate and the individual prostates of the mouse. Microscopically, the epithelial cells lining the ducts are structurally similar in both species, but there are differences in the termination of the ducts. Similarities and differences in the predilection of different lobes and prostates to benign and malignant neoplasia must be sought in the gene expression patterns exhibited by epithelial cells of the putatively corresponding anatomical structures.

The challenge is to glean from the frame-based symbolic models we established for the mouse and human prostate and mammary gland the elements of a computational model that can express similarities and differences in the anatomy of the two species in a way that can generalize to any pair of symbolic models, whether they represent corresponding anatomical parts of two species or different developmental stages of the same species.

COMPARING STRUCTURES

Our intent is to develop a computational model that can map the anatomical entities of one species to those of another and determine similarities and differences represented in the frames of these anatomical entities in the FMA. This mapping makes use of the AT and ASA components of the FMA. We hypothesize that applying set-theoretic and graph-theoretic approaches to comparing the frames, slots, and slot-values of the FMA will formally capture what is similar and what is different across a pair of species.

Set Comparisons. Although organ parts as such are not classified in the FMA as anatomical sets, in set theory they are regarded as a set. A mapping from a set A to a set B is a function $f : A \longmapsto B$ that assigns a single element of B to each element of A. A mapping f is *one-to-one* if it maps each element of A to a unique element of B $[f(a) = f(a') \implies a = a']$. Let A be a set of human organ parts {`Capsule`, `Stroma`, `Prostatic duct`}, and B be a set of animal organ parts {`Capsule`, `Stroma`, `Prostatic duct`}. The function $f : A \longmapsto B$ that maps each human organ to its animal counterpart is a one-to-one map, because all of the human organ parts in set A have a unique corresponding organ part in the animal set B. It is also *onto*, because all the elements of B are used. A set isomorphism is a one-to-one and onto mapping, as illustrated in Figure 1.

Taken at the appropriate level of abstraction, such a set isomorphism holds for certain parts of a prostate in both species, regardless of the number of prostates each actually possesses. This formalism generalizes to other entities, for example, the heart, its chambers, and
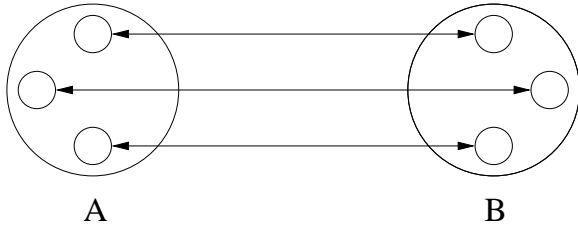
Figure 1: A set isomorphism for organ parts of the human (A) and mouse (B) prostate.

Figure 2: Mapping the human heart (H) to the mouse heart (M).

the walls of the chambers (Figure 2). There is a set isomorphism between the human and mouse heart at the organ level and also at the organ part level: each species has a heart, a corresponding set of cardiac chambers (right and left atrium, right and left ventricle) and the wall of each chamber has a corresponding set of layers (epicardium, myocardium, endocardium).

Graph Comparisons. We proceed from sets to graphs, in order to map between relationships as well as between structures. In our graph examples the AT concepts correspond to the nodes and the ASA relationships to the edges of the graphs. A graph isomorphism is more restrictive than a set isomorphism. Let $G_A = (A, E_A)$ be a graph with node set A and edge set E_A, and let $G_B = (B, E_B)$ be a second graph. A graph isomorphism is a one-to-one, onto mapping $f : A \longmapsto B$ such that $(a, a') \in E_A$ iff $(f(a), f(a')) \in E_B$. This means that if there is an edge between nodes a and a' in G_A, there must be an edge between the corresponding nodes $f(a)$ and $f(a')$ in G_B, and vice versa. This is a relational constraint.

Let Graph A be a representation of the human heart (H), and Graph B be a representation of the mouse heart (M), as depicted in Figure 2. The root of each graph is `Heart`, and it has four children, connected to `Heart` by the relationship *has-part*: `Left atrium`, `Left ventricle`, `Right atrium`, and `Right ventricle`. (For simplicity of illustration, we limit the graph to cardiac chambers). In mapping the nodes of Graph A to the nodes of Graph B, mouse `Heart` matches human `Heart`, `Right atrium` matches `Right atrium`, and so forth. Similarly, the four *has-part* edges match. The mapping is therefore one-to-one and onto, and the relational constraints are satisfied, which constitutes a graph isomorphism.

Similarly, although the entire prostate is not isomorphic across species, the prostatic ducts are isomorphic between mouse and human, as are the parts of the pro-
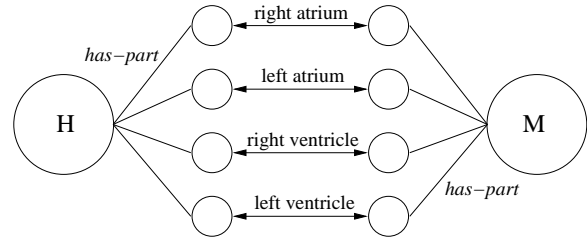
static duct. For example, the wall of the prostatic duct has the following isomorphic layers: the epithelium surrounds the lumen of the duct, the muscle layer surrounds the epithelium, and the adventitia surrounds the muscle layer.

Although these examples are of simple graphs, the frame-based representation of the FMA in Protégé is much more complex than a simple graph since 1) it has attributed nodes (*e.g.*, *has-mass*; *has-inherent-3D-shape*), and 2) it has multiple relationships (*e.g.*, *is-a*, *has-part*, *continuous-with*, *adjacent-to*). The edges of the complex graph structure of the FMA represent this rich mixture of structures and relationships. We have found that similarities and differences can occur at all levels between two graphs, as well as across levels, and that, as expected, there are more similarities than differences.

SYMBOLIC DIFFERENCES

We use set and graph isomorphism to illustrate anatomical similarity and any deviation from isomorphism to represent a difference in the anatomical entities compared. In this way, we can start with an organ, display the *part-of* hierarchy to the cellular level for each species under comparison, and determine the mappings at each level. If two structures are isomorphic at some level of abstraction and resolution, they are identical at that level. But if they are not isomorphic, how do we gauge the difference between two corresponding structures?

Based on our preliminary studies and the work of Shapiro and Haralick[5], we propose the following types of differences for our approach: node set differences, node attribute differences, node attribute value differences, and relationship differences. We illustrate each type of symbolic difference with examples.

Node set differences are differences between the number of entities in the source species and in the target species. Such mapping differences include null mappings, which may be one-to-zero (one mouse limiting ridge to none in human, discussed below) or many-to-zero (two areolae of breast in human to none in mouse mammary glands). Additionally, there are mappings that may be one-to-n (one human prostate to five mouse organs), or n-to-m (three lobes of human right lung to five lobes of mouse right lung; two mammary glands in human to twelve in mice).

Node attribute differences are differences in the existence of an attribute between two corresponding structures in the source and target species. For example, *has-member* (which is a specialization of the partonomic relationship constrained in the FMA to sets[6]) is an attribute of the node `Set of mouse prostates`. In this partonomic scheme, `Anatomical set` is made up of member organs. The class `Organ`, however, lacks the attribute *has-member*, and therefore a node attribute difference exists between the prostates of the two species.

Node attribute value differences are differences in values of corresponding attributes shared between corresponding nodes of two species. An isomorphism exists between the mouse and human stomachs at the levels of whole organ and organ part: the mapping is one-to-one and onto for {`Fundus of stomach`, `Body of stomach`, `Pyloric antrum`}. The isomorphism propagates to the next level, namely, the stomach wall, the layers of which are: mucosa (GM), submucosa (SM), muscularis (M) and serosa (S). The difference between mouse and human emerges in the attribute values for the node `Mucosa`. Unlike the body of the human stomach (HS), which is lined throughout by glandular mucosa (GM), the mucosa of the body of the mouse stomach (MS) is divided into two structurally-different regions: *glandular mucosa* (GM) and *non-glandular mucosa* (NM). GM and NM are separated by a `Limiting ridge` (LR), which has no corresponding human node[7].

Figure 3 depicts both node attribute value differences and node set differences. The mapping involving the `Serosa`, `Submucosa`, and `Muscularis` is an isomorphism, indicated by the two-headed arrows. The mucosa, however, is not isomorphic across species: in the human its attribute value is *glandular*, whereas in the mouse the values are *glandular* and *non-glandular*. The dashed line represents a mapping between nodes with different values for the same attributes. Additionally, there is no corresponding structure for the LR in the human: the difference in node mapping is represented by the dotted line. This is an example of a *null mapping*, and the non-existent structure is represented by the empty set notation {}.
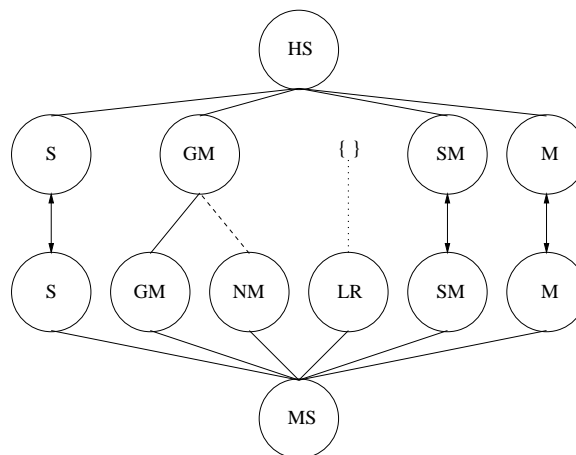

Figure 3: Layers of wall in mouse (MS) and human (HS) stomachs.

Relationship differences are differences in relationships (edges) between structures across species. For example, the dorsolateral prostates of the mouse are adjacent to the coagulating glands, which do not exist as organs in the human. Another example is the inguinal mammary glands of the mouse, which are adjacent to the inguinal ligament, whereas the human mammary glands are adjacent only to the pectoralis major muscle. Because they are located in different places in the body in different species, the spatial relationships among the anatomical entities are changed, and this change is reflected in the relationship differences across species.

RELATED COMPUTATIONAL MODELS

Although no two species are identical, there are varying degrees of similarity in the examples we cite at different levels, and more similarities than differences, as we would expect from the vertebrate *Bauplan*. The question of capturing these differences in a computational model poses a challenge.

Bernstein *et al.* have proposed a model-matching-and-merging approach to deal with the problem of merging two or more different schemas in a database environment[8]. Their schemas are represented as graph structures, as are ours. They allow a node in one graph to map to a node in the other graph if they are identical or "similar" concepts. Using a simple definition of similarity, they have developed a matching algorithm to find a mapping from one graph to another. The re-

sulting match is represented as a graph structure itself, an idea that we will pursue in our work.

Shapiro and Haralick[5] studied the problem of describing differences in the context of computer vision. They defined the *error* of a one-to-one, onto mapping $f : A \longmapsto B$ from graph G_A to graph G_B as the number of edges (a, a') of E_A whose corresponding edge $(f(a), f(a'))$ is not in E_B plus the number of edges (b, b') of E_B whose corresponding edge $(f^{-1}(b), f^{-1}(b'))$ is not in E_A. This error measure was used to compare two complex graph structures. The best mapping from G_A to G_B is the one with the least error. The error of the best mapping is called the *relational distance* between G_A and G_B.

With some modifications, this is the approach we will rely on for assessing the differences between structures across species. However, rather than returning a numeric value for the error as a measure of relational distance, we will experiment with *describing* the differences between the species in terms of the elements embedded in the frame-based symbolic representation.

CONCLUSIONS AND DISCUSSION

We have developed an approach to interspecies anatomical mapping based on the concept of graph matching, Shapiro and Haralick's relational distance between two structures, and the Bernstein *et al.* paradigm for model matching in database systems. Our approach is to 1) represent the structure of each species in the Protégé-2000 frame-based knowledge-representation system, as has been done for the FMA; 2) determine the best mapping from the graph representing the source species to the graph representing the target species in terms of mapping nodes to biologically-similar nodes and preserving structural relationships and 3) describe the difference between the two species in terms of the symbolic differences between the two graph structures. The examples we cite illustrate how our approach represents the first and necessary step in comparing human and mouse organs, as represented in the FMA. One problem with our approach is the need for symbolic models of other species, represented in Protégé. Several such models are currently being developed for the mouse, and we hope they will be forthcoming for other species. Once such models exist, representing them in Protégé is not difficult for us.

Our ultimate goal is to deliver to the user a structured presentation of what is similar and what is different between species being compared. This paper is a preliminary report, which forms the basis for developing this approach more comprehensively and then evaluating it. We recognize that this preliminary work is based on a few selected, even anecdotal, examples,

which are sufficient for illustrating the main ideas supporting the method, but are not adequate for its evaluation. Systematic and extensive testing is needed based on more than one pair of species. Our contribution is a new methodology for the anatomical correlation of species. We hope that the methodology we have developed will be useful in many applications.

REFERENCES

1 Rosse C, Mejino LVJ. A reference ontology for bioinformatics: the Foundational Model of Anatomy. *Journal of Biomedical Informatics* 2003. In press.

2 Rosse C, Shapiro LG, and Brinkley JF. The Digital Anatomist Foundational Model: Principles for defining and structuring its concept domain. *Proceedings 1998 American Medical Informatics Association Annual Symposium*, November 1998.

3 Mejino JLV, Noy, NF, Musen M, Rosse C. Representation of structural relationships in the Foundational Model of Anatomy. *Proceedings of the AMIA Annual Symposium* 2001:973.

4 NCI Mouse Models of Human Cancers Consortium (MMHCC) <http://emice.nci.nih.gov/>.

5 Shapiro LG, Haralick RM. A Metric for Comparing Relational Descriptions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-7, No. 1, 1985:90-94.

6 Mejino JLV Jr, Agoncillo AV, Rickard K, Rosse C. Representing Complexity in Part-Whole Relationships within the Foundational Model of Anatomy. *Proceedings of the American Medical Informatics Association Symposium* 2003. In press.

7 Robert A. Proposed terminology for the anatomy of the rat stomach. *Gastroenterology*. 1971 Feb; 60(2):344-5.

8 Bernstein, P.A., A.Y. Levy, R.A. Pottinger, "A Vision for Management of Complex Models," Microsoft Research Technical Report *MSR-TR-2000-53*, June 2000.