Proceedings of ICAD2009
The Fifth International Conference on Axiomatic Design
Campus de Caparica – March 25-27, 2009
ICAD-2009-05

AN AXIOMATIC DESIGN FOR MODELING BIOLOGICAL SYSTEMS

Harun Pirim and Burak Eksioglu
Department of Industrial and Systems Engineering
Mississippi State University
PO Box 9542
Mississippi State, MS 39762, USA

ABSTRACT

Precise mathematical modeling of biological systems may allow researchers to accurately estimate future states of these biological systems. Biological systems are regarded as complex systems to which mathematical modeling and Operations Research (OR) tools can be applied. Also, a general roadmap is essential for a biology scientist to decide which part of the system to focus on. In this study a recent review of computational biology and OR in biology literature is given. Modeling problems in biological sciences are classified and potential solution methodologies to these problems are addressed. An axiomatic design of a modeling procedure is provided.

Keywords: Modeling in Biology, Operations Research, Axiomatic Design

1 INTRODUCTION

Life science studies such as bioinformatics and molecular biology comprise a lot of cumbersome mathematical expressions. This is due to having produced a lot of data as a result of performing numerous experiments in the laboratory or using technological devices to understand what's going on at the molecular level. With all the data generated, biologists usually rely on software packages to understand and/or solve the problems encountered in biology. Unfortunately, a limited number of software packages are readily available for biologists, and those that are available have limited functionality [1]. Therefore, in many cases biologists may have to develop their own computer programs to solve the specific problem at hand. In most cases the custom-built computer programs will utilize mathematical programming and optimization tools. Thus, biologists will need to engage in fields like operations research and combinatorial optimization to be able to develop computer programs to tackle their problems. A typical problem for biologists who are eager to engage in optimization or mathematical modeling topics is picking the right mathematical method among numerous methods available that will help in solving their problem efficiently. Obviously, before they pick a method they first need to model their problem in mathematical terms. In other words, they need to translate their problem from biology to mathematics, and then look for the right solution method. Another issue between life sciences and mathematical disciplines is the difference the objectives they try to achieve. The following statement clearly shows that there are significant differences between what each disciple tries to achieve. "Difference between the usual combinatorial optimization problems and those which arise from scientific (as opposed to technological) contexts is that in science, the objective is to uncover the truth, rather than find a minimum cost solution to some objective function." [1]. However, this does not mean that the two disciplines should not interact.

In this study we provide a brief, up-to-date literature review on computational biology, bioinformatics, and molecular biology with a focus on mathematical modeling. The literature review provides an overview of mathematical techniques and models used in life sciences. Life science disciplines are classified based on the articles reviewed. In each discipline, problems and sub-problems encountered as well as solution procedures to these problems are given. Following the classification an axiomatic design approach is presented that is expected to help bio-scientists interested in learning more about mathematical modeling. The proposed axiomatic design also offers insights about introductory or advanced problem statements in biology to researchers whose major interests are in operations research or similar fields. The classification and the axiomatic design will also be helpful, to some extent, for scientists who are already involved in interdisciplinary studies comprising biology and mathematics by representing other related biology fields using the same mathematical techniques. The remainder of the paper is organized in four sections. In section 2 definitions of life science fields, biological problems, and solution methods are provided. The definitions are limited to the papers that are used in the literature review to classify the problems. In section 3, a brief overview of the operations research field and its relation to life sciences is given. Section 4 introduces the concept of axiomatic design, why it is important, and how it can be used. Section 4 also represents the axiomatic design developed for modeling biological systems. Section 5 concludes the study and gives directions for future studies.

2 PROBLEMS IN LIFE SCIENCES

There are many definitions about life sciences disciplines. Based on our review of the literature, we identified bioinformatics, computational biology, molecular biology, systems biology, and biochemistry as the main disciplines under life sciences. The definitions of these disciplines are as follows:

- Bioinformatics: The field of science in which biology, computer science, and information technology merge into a single discipline. There are three important sub-disciplines within bioinformatics: (1) the development of new algorithms and statistics with which to assess relationships among members of large data sets; (2) the analysis and interpretation of various types of data including nucleotide and amino acid sequences, protein domains, and protein structures; and (3) the development and implementation of tools that enable efficient access and management of different types of information [2].

- Computational Biology: The field of science that "encompasses the use of algorithmic tools to facilitate biological analyses" [2].

- Molecular Biology: "The study of the biochemistry of cells, it is closely linked to cell biology, in particular the biochemistry of DNA and cogeners." "The branch of biology that studies the structure and activity of macro-molecules essential to life (and especially with their genetic role)." The study of biology at the molecular level, such as the chemical properties of DNA" [3].

- Systems Biology: "A field that seeks to study the relationships and interactions between various parts of a biological system (metabolic pathways, organelles, cells, and organisms) and to integrate this information to understand how biological systems function" [4].

- Biochemistry: "The study of organic chemistry of compounds and processes occurring in organisms; the effort to understand biology within the context of chemistry" [3].

Some problems can be classified in more than one sub-discipline. For example the protein structure alignment problem can appear in both bioinformatics and molecular biology. For these kinds of problems we chose only one sub-discipline randomly to keep our axiomatic design as simple and general as possible. In the other words, if a problem is assigned to a sub-discipline this problem is assumed to exist just in that discipline. In the following paragraphs we list some of the problems that we classified under each of the sub-disciplines other than systems biology given above.

Some bioinformatics problems that we have indentified are protein structure alignment, genetic regulatory interactions, biological activity, structural bioinformatics and protein-protein interaction. Protein structure alignment is one of the most important problems encountered not only in bioinformatics but also in molecular biology. This problem is known to be NP-hard [5]. In [5] a new method for solving this problem is introduced in which the protein structure alignment problem is formulated as a mixed integer linear program (MIP). Mathematical modeling as well as computational study of genetic regulatory interactions is a relatively new topic in bioinformatics. In [6] an approximation method is given to tackle the genetic regulatory interactions problem. The method computes steady-state probability distributions of probabilistic Boolean networks. A transition probability matrix is constructed and a probability distribution is determined. Efficiency of the proposed method is shown by numerical experiments based on a genetic network. In general only molecular structure is considered for biological activity whereas in [7] a Markov model is proposed considering both molecular structure and the specific biological system the drug affects.

Computational biology problems are generally related to infection biology, genetic structure of natural populations, genotype sequences, population variability, RNA secondary structure prediction, pairwise alignment, population genetics, structure prediction, and RNA folding prediction. In [8] a Bayesian network classifier is used to represent relations between random variables. A prediction about previously unseen data is made. Data is captured from infected mice. This study is related to 'infection biology'. In [9] a Bayesian clustering approach is extended to understand the mating structure of populations that is an important goal of population biology. In [10] probability-based parameter estimation methods are given for understanding the uncertainty and variability in biological models. Here these models are related to population variability. In [11] a web-based computational tool called "taveRNA" is provided to identify the structure and functionality of ncRNA molecules. In [12] the web server DIAL is described which is used for pairwise structural alignment of RNA.

Most of the molecular biology problems are related to genome research, mapping, gene regulatory networks, nucleic acids, ancestral genome reconstruction, genome sequences, genomics, proteomics, genetic epidemiology. In [13] a system for discovering and viewing syntenic regions of FPC maps, a topic in comparative genomics, is provided. In [14] an algorithm for pathway mapping across microbial genomes is presented. The algorithm deals with sequence similarity and genomic structure information where the problem is formulated as an integer program. In [15] a linear model of gene regulation is used to form an optimization model and a solution framework. In [16] the Gibbs Centroid Sampler software package is introduced. It is shown that centroid estimators yield efficient improvements to the prediction of RNA secondary structure and motif finding. In [17] a new measure of protein structural relationships (i.e. protein structural distance) is used. A structural alignment that uses

double dynamic programming is utilized as the calculation method. In [18] a pattern recognition algorithm for detecting patterns with different lengths from large data sets is given. In [19] a combinatorial optimization framework for motif finding is given.

Biochemistry problems are typically related to bioterrorism [20] and artificial chemistry [21]. To tackle bioterrorism, differential equations are commonly used. For artificial chemistry problems, in general, cellular automata are used.

3 OPERATIONS RESEARCH IN LIFE SCIENCES

Operations Research (OR) is a field that is rooted to World War II. The British army was in need of using scarce resources in the most efficient way. As the name indicates, OR emerged as a scientific field as a result of searching for the best operations like assigning resources to people/places/machines, and finding optimum values that minimize cost or maximize profit. With the technological availability and computational ability of computers, OR philosophy became more precise in terms of solvability of mathematical models. OR tools are widely used in telecommunication networks, scheduling, vehicle routing, queuing, production planning, logistics applications, and financial problems. Recent OR applications are in healthcare and social topics. As described in Section 2, life sciences also benefit from OR tools such as integer programming, dynamic programming, and linear programming.

4 AXIOMATIC DESIGN

Axiomatic Design (AD) was introduced by Nam Pyo Suh in 1990 to address the need of a general framework for any kind of design work. Application areas of AD are systems, software design, materials, and material processing [22]. AD is based on two axioms: axiom of independence and information axiom. According to the independence axiom, independence of functional requirements should be maintained. Information axiom emphasizes minimizing the information content of the design. Functional requirements (FRs) are defined to satisfy the needs. FRs are minimum set of independent requirements to be satisfied by the design [22]. In case of existence of more than one FR, design solution must be such that each FR can be satisfied without affecting the other FRs. Design solution comes out with design parameters (DPs) to achieve a job that are the process variables (PVs). Consequently, a design can be in one of three states: Uncoupled, coupled or decoupled design. A design matrix is constructed to understand the state of a design.

An AD framework consists of four domains. These domains are customer, functional, physical, and process domains. The customer domain is mapped into a functional domain. The functional domain is mapped into a physical domain and the process goes on like this. FRs are in the functional domain, and DPs are in the physical domain. An AD study starts with identifying FRs. An FR is what we want to achieve where as a DP is how we want to achieve it. Each FR is satisfied by a corresponding DP. By the way, a DP may correspond to a further FR that makes the AD decoupled. If at least one DP corresponds to a previous FR then the AD is coupled. In case a FR is satisfied by just its corresponding DP, the AD is uncoupled. AD study is basically a mapping between each domain. These mappings are represented in a hierarchical decomposed zigzagging manner (see Figures 1 and 2). After defining the mappings, relationships between FRs, DPs and PVs are clarified via a matrix representation. The designer decides how detailed the design should be and stops the decomposition process accordingly. In our study we used the independence axiom. This allowed us to develop a roadmap for life science studies. Thus, our main functional requirement is to define a life science study. FRs and the corresponding DPs we defined are as follows:

FR0 = Define a life science study
FR1 = Involve in bioinformatics
FR2 = Involve in computational biology
FR3 = Involve in molecular biology
FR4 = Involve in systems biology
FR5 = Involve in biochemistry

We define a DP corresponding to each of the above FRs. For example; DP0 corresponds to FR0, DP1 corresponds to FR1, and so on. DP0 is defined as "Study by fields." The definitions of DP1, DP4, and DP5 are given in Figure 1, and those of DP2 and DP3 are given in Figure 2. Due to space limitations a complete picture of the AD is not provided. However, Figures 1 and 2 show the pieces of the AD we developed. Under DP0 we have FR1, FR2, FR3, FR4, and FR5. Also, as can be seen from Figure 1, DP1 has a number of sub FRs (FR11, FR12, …, FR15) associated with it. Each of these sub FRs also have a corresponding DP (e.g. DP11 for FR11, DP22 for FR22, etc.).

Note that expressing FRs in 'verb' form and DPs in 'noun' form is the standard convention. Following DPs and corresponding FRs are shown in figure 1 and figure 2 (See Appendix)

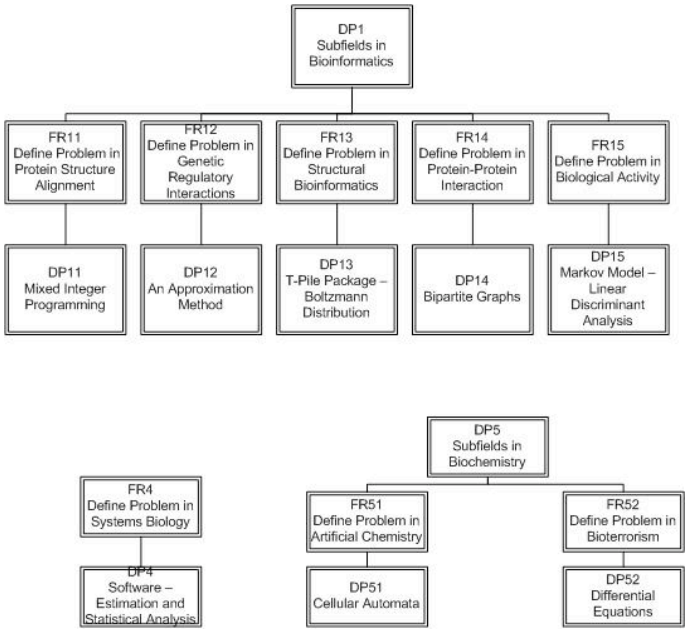

Figure 1. Decomposition of DPs and FRs

Having constructed the above decomposition figures, we can easily represent the relationship between FRs and DPs via design matrixes. Design matrixes for FRs are as follows:

$$\begin{bmatrix} FR1 \\ FR2 \\ FR3 \\ FR4 \\ FR5 \end{bmatrix} = \begin{bmatrix} X & & & & \\ & X & & & \\ & & X & & \\ & & & X & \\ & & & & X \end{bmatrix} \begin{bmatrix} DP1 \\ DP2 \\ DP3 \\ DP4 \\ DP5 \end{bmatrix}$$

$$\begin{bmatrix} FR11 \\ FR12 \\ FR13 \\ FR14 \\ FR15 \end{bmatrix} = \begin{bmatrix} X & & & & \\ & X & & & \\ & & X & & \\ & & & X & \\ & & & & X \end{bmatrix} \begin{bmatrix} DP11 \\ DP12 \\ DP13 \\ DP14 \\ DP15 \end{bmatrix}$$

Diagonal 'X's imply that each DP corresponds to only one FR. Such a design is known to be an uncoupled one. However we need to form all design matrixes to be able to make a comment about the category the overall design belongs to (i.e. the design may in fact be coupled or decoupled).

$$\begin{bmatrix} FR21 \\ FR22 \\ FR23 \\ FR24 \\ FR25 \\ FR26 \\ FR27 \\ FR28 \\ FR29 \\ FR210 \end{bmatrix} = \begin{bmatrix} X & & & & & & & & & \\ & X & & & & & & & & \\ & & X & & & & & & & \\ & & & X & & & & & & \\ & & & X & X & & & & & \\ & & & & & X & & & & \\ & & & X & X & & X & & & \\ & & & & & & & X & & \\ & & & & & & & & X & \\ & & & X & X & & X & & X & X \end{bmatrix} \begin{bmatrix} DP21 \\ DP22 \\ DP23 \\ DP24 \\ DP25 \\ DP26 \\ DP27 \\ DP28 \\ DP29 \\ DP210 \end{bmatrix}$$

$$\begin{bmatrix} FR31 \\ FR32 \\ FR33 \\ FR34 \\ FR35 \\ FR36 \\ FR37 \\ FR38 \\ FR39 \end{bmatrix} = \begin{bmatrix} X & & & & & & & & \\ & X & & & & & & & \\ & X & X & & & & & & \\ & & & X & & & & & \\ & X & X & & X & & & & \\ & & & & & X & & & \\ & X & X & & X & & X & & \\ & & & & & & & X & \\ & & & & & & & & X \end{bmatrix} \begin{bmatrix} DP31 \\ DP32 \\ DP33 \\ DP34 \\ DP35 \\ DP36 \\ DP37 \\ DP38 \\ DP39 \end{bmatrix}$$

We can notice from the above design matrixes that our design is actually decoupled. We assume that further DPs are not used in previous FRs. For example, DP29 'Dynamic Programming' is assumed to be different than DP24 'Dynamic Programming' because our design shows the relation between problems and solution methods rather than sequence of operations.

5 CONCLUSION

In this study we used an axiomatic design to provide a road map for a scientist to be able to involve in a life science study that requires mathematical modeling. Our design is decoupled which in a way implies that a scientist who is able to use a specific mathematical or OR tool can involve in another life science study using the same tool. This is a practical output of the AD. In future studies, the information axiom of AD can be considered for a life science study. A more detailed design for subfields can also be developed.

6 REFERENCES

[1] Warnow T., 1993, "Combinatorial Optimization in Biology" IEEE Engineering in Medicine and Biology

[2] http://www.epa.gov/NCCT/glossary.html , National Center for Computational Toxicology (NCCT) glossary, accessed on 12-13-2007 at 7:56 pm

[3] http://www.biology-online.org/dictionary/Branches_of_biology , accessed on 12-13-2007 at 8:00 pm

[4] http://publications.nigms.nih.gov/thenewgenetics/glossary.html , National Institute of General Medical Sciences Glossary, accessed in 12-13-2007 8:04 pm

[5] Chen L., Zhou T., Tang Y., *2005,* "Protein structure alignment by deterministic annealing" Bioinformatics, *Vol. 21 no. 1, pages 51–62*

[6] Ching W., Zhang S., K. M., Akutsu T., 2007, "An approximation method for solving the steady-state probability distribution of probabilistic Boolean networks" Bioinformatics Vol. 23 no. 12 pages 1511–1518

[7] Monteagudoa M. C., Gonz´alez-D´ıaza H., Uriartea E., 2006, "Simple Stochastic Fingerprints Towards Mathematical Modeling in Biology and Medicine 2. Unifying Markov Model for Drugs Side Effects" Bulletin of Mathematical Biology 68: 1527–1554

[8] Gilfeather F., Hamine V., Helman P., et. al., 2007, "Learning and modeling biosignatures from tissue

images" Computers in Biology and Medicine 37, 1539 – 1552

[9] Gao H., Williamson S. Bustamante C. D., 2007, "A Markov Chain Monte Carlo Approach for Joint Inference of Population Structure and Inbreeding Rates From Multilocus Genotype Data" Genetics 176: 1635–1651

[10] Banks H.T., Potte L. K., 2002, "Probabilistic methods for addressing uncertainty and variability in biological models: Application to a toxicokinetic model" Mathematical Biosciences

[11] Aksay C., Salari R., Karakoc E., et. al., 2007, "taveRNA: a web suite for RNA algorithms and applications" Nucleic Acids Research, Vol. 35

[12] Ferre F., Ponty Y., Lorenz W. A., Clote P., 2007, "DIAL: a web server for the pairwise alignment of two RNA three-dimensional structures using nucleotide, dihedral angle and base-pairing similarities" Nucleic Acids Research, Vol. 35

[13] Soderlund C., Nelson W., Shoemaker A., Paterson A., 2006 "SyMAP: A system for discovering and viewing syntenic regions of FPC maps" Genome Research, 16:1159–1168

[14] Mao F., Su Z., Olman V., et. al., 2006, "Mapping of orthologous genes in the context of biological pathways: An application of integer programming" PNAS, vol. 103, no. 1, 129–134

[15] Dasika M. S., Gupta A., Maranas C. D., VARNER J. D., 2004, "A Mixed Integer Linear Programming (MILP) Framework for Inferring Time Delay in Gene Regulatory Networks" PSB proceedings

[16] Thompson W. A., Newberg L. A., Conlan S., 2007, "The Gibbs Centroid Sampler" Nucleic Acids Research, Vol. 35

[17] Yang A., Honig B., 2000, "An Integrated Approach to the Analysis and Modeling of Protein Sequences and Structures. I. Protein Structural Alignment and a Quantitative Measure for Protein Structural Distance" J. Mol. Biol. 301, 665-678

[18] Feng J., Q. D., Cooper N. and B., "Probability-based pattern recognition and statistical framework for randomization: modeling tandem mass spectrum/peptide sequence false match frequencies" Bioinformatics, Vol. 23 no. 17 2007, pages 2210–2217

[19] Zaslavsky E., Singh M., 2006, "A combinatorial optimization approach for diverse motif finding Applications" Algorithms for Molecular Biology, 1:13

[20] Banks H.T., Bortz D., Pinter G., Potter L., 2003 "Modeling and Imaging Techniques with Potential for Application in Bioterrorism" Bioterrorism: Mathematical Modeling Applications in Homeland Security

[21] Hutton T. J., 2007, "Evolvable Self-Reproducing Cells in a Two-Dimensional Artificial Chemistry" Artificial Life 13: 11–30

[22] Suh, N. P. (2001). Axiomatic design: Advances and applications. New York: Oxford University Press

APPENDIX

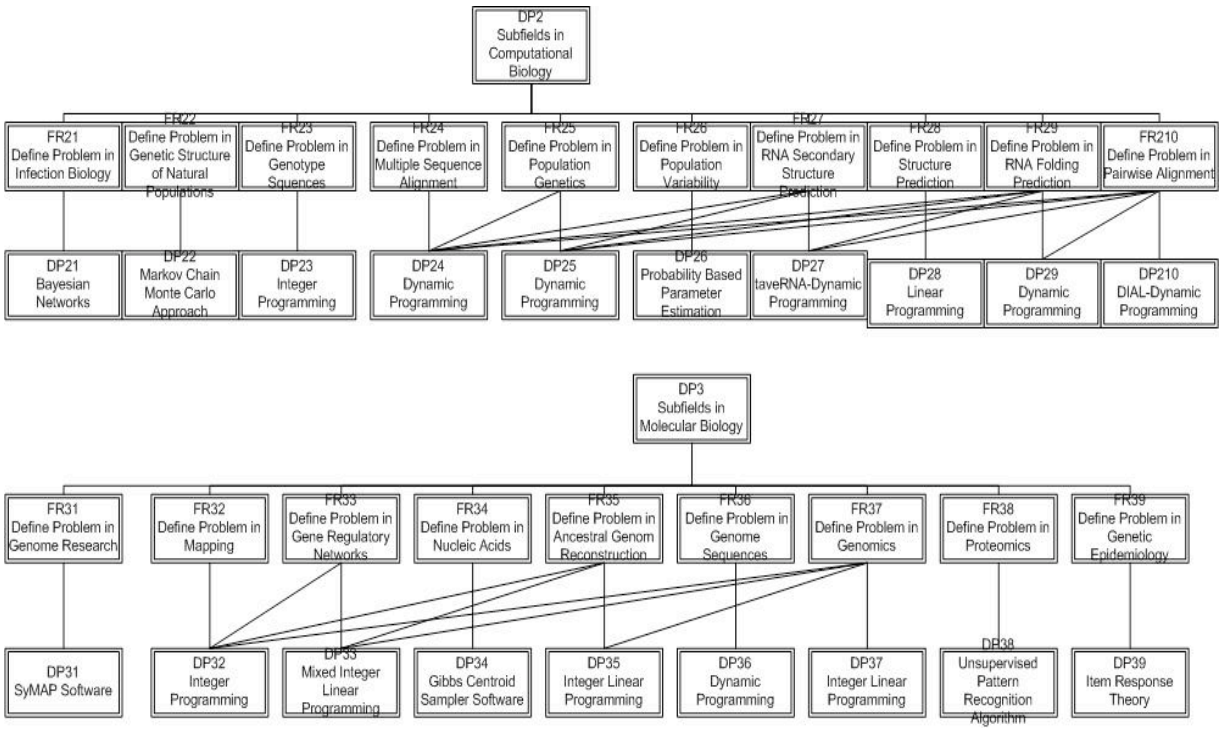

Figure 2. Decomposition of DP2 and DP3