Photosynthesis *in silico*

Understanding Complexity from Molecules to Ecosystems

Advances in Photosynthesis and Respiration

VOLUME 29

The scope of our series, beginning with volume 11, reflects the concept that photosynthesis and respiration are intertwined with respect to both the protein complexes involved and to the entire bioenergetic machinery of all life. Advances in Photosynthesis and Respiration is a book series that provides a comprehensive and state-of-the-art account of research in photosynthesis and respiration. Photosynthesis is the process by which higher plants, algae, and certain species of bacteria transform and store solar energy in the form of energy-rich organic molecules. These compounds are in turn used as the energy source for all growth and reproduction in these and almost all other organisms. As such, virtually all life on the planet ultimately depends on photosynthetic energy conversion. Respiration, which occurs in mitochondrial and bacterial membranes, utilizes energy present in organic molecules to fuel a wide range of metabolic reactions critical for cell growth and development. In addition, many photosynthetic organisms engage in energetically wasteful photorespiration that begins in the chloroplast with an oxygenation reaction catalyzed by the same enzyme responsible for capturing carbon dioxide in photosynthesis. This series of books spans topics from physics to agronomy and medicine, from femtosecond processes to season long production, from the photophysics of reaction centers, through the electrochemistry of intermediate electron transfer, to the physiology of whole organisms, and from X-ray crystallography of proteins to the morphology or organelles and intact organisms. The goal of the series is to offer beginning researchers, advanced undergraduate students, graduate students, and even research specialists, a comprehensive, up-to-date picture of the remarkable advances across the full scope of research on photosynthesis, respiration and related processes.

For other titles published in this series, go to
www.springer.com/series/5599

Photosynthesis *in silico*

Understanding Complexity from Molecules to Ecosystems

Edited by

Agu Laisk
University of Tartu
Estonia

Ladislav Nedbal
Institute of Systems Biology and Ecology
Academy of Sciences of the Czech Republic
Nové Hrady
Czech Republic

and

Govindjee
University of Illinois at Urbana-Champaign
Urbana
Illinois
USA

Springer

Library of Congress Control Number: 2009921443

ISBN 978-1-4020-9236-7 (HB)
ISBN 978-1-4020-9237-4 (e-book)

Published by Springer,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

www.springer.com

Cover: Photosynthesis *in silico*. Photo by Eero Talts, University of Tartu

Printed on acid-free paper

From the Series Editor

I am highly obliged to each and everyone of the authors from fifteen countries(Australia, Austria, Brazil, Canada, China, Czech Republic, Estonia, France, Germany, Lithuania, Russia, Switzerland, The Netherlands, U.K., and U.S.A.) for their valuable contributions to the successful and timely production of this unique book, Volume 29 in *Advances in Photosynthesis and Respiration*: '**Photosynthesis *in silico*—**Understanding Complexity from Molecules to Ecosystems', edited by two of the pioneers in the field (Agu Laisk, of Estonia; and Ladislav Nedbal, of the Czech Republic).

The authors are: Niels P.R. Anten (The Netherlands; Chapter 16); Michael J. Behrenfeld (U.S.A.; Chapter 20); Carl J. Bernacchi (U.S.A.; Chapter 10); Joseph Berry (U.S.A.; Chapter 9); Jan Červený (Czech Republic; Chapter 2); Jan P. Dekker (The Netherlands; Chapter 3); Gerald Edwards (U.S.A.; Chapter 14); Hillar Eichelmann (Estonia: Chapter 13); Hadi Farazdaghi (Canada; Chapter 12); Graham Farquhar (Australia; Chapters 9 and 10); Arvi Freiberg (Estonia; Chapter 4); Andrew Friend (U.K.; Chapter 20); Richard J. Geider (U.K.; Chapter 20); Jeremy Harbinson (The Netherlands; Chapter 11); Hubert Hasenauer (Austria; Chapter 19) Michael Hucka (U.S.A.; Chapter 1); Manfred Küppers (Germany; Chapter 18); Agu Laisk (Estonia; Chapters 13 and 14); Jérôme Lavergne (France; Chapter 8); Dušan Lazár (Czech Republic; Chapter 5); Stephen P. Long (U.S.A.; Chapters 10 and 17); Ladislav Nedbal (Czech Republic; Chapter 2); Ülo Niinemets (Estonia; Chapter 16); Vladimir I. Novoderezhkin (Russia; Chapter 3); Vello Oja (Estonia; Chapter 13); Michael Pfiz (Germany; Chapter 18); Stephen A. Pietsch (Austria; Chapter 19); Carlos Pimentel (Brazil; Chapter 10); Ondřej Prášil (Czech Republic; Chapter 6); Galina Riznichenko (Russia; Chapter 7); David M. Rosenthal (U.S.A.; Chapter 10); Andrew Rubin (Russia; Chapter 7); James Schaff (U.S.A.; Chapter 1); Gert Schansker (Switzerland; Chapter 5); Henning Schmidt (Germany; Chapter 2); Christopher J. Still (U.S.A.; Chapter 20); Paul C. Struik (The Netherlands; Chapter 11); Gediminas Trinkunas (Lithuania; Chapter 4); Rienk van Grondelle (The Netherlands; Chapter 3); Susanne von Caemmerer (Australia; Chapter 9); Wim Vredenberg (The Netherlands; Chapter 6); Ian E. Woodrow (Australia; Chapter 15); Xinyou Yin (The Netherlands; Chapter 11), and Xin-Guang Zhu (China; Chapter 17).

I also thank Jacco Flipsen, Noeline Gibson and André Tournois of Springer (Dordrecht, The Netherlands) for their patience with me during the production of this book. I am particularly thankful to Albert André Joseph (of SPi Technologies India Private Limited) for his wonderful cooperation in efficiently taking care of the corrections in the proofs of this book. Finally, I am grateful to my wife, Rajni Govindjee, and the offices of the Department of Plant Biology (Feng Sheng Hu, Head) and of Information Technology, Life Sciences (Jeff Haas, Director), of the University of Illinois at Urbana-Champaign, for their constant support.

A list of books, published under the Series '*Advances in Photosynthesis and Respiration'* is available at the Springer web site: <http://www.springer.com/series/5599>. Members of the ISPR (International Society of Photosynthesis Research) receive 25 % discount. For the Table of Content of most of the earlier volumes, see my web site at: <http://www.life.uiuc.edu/govindjee/Reference-Index.htm>.

Govindjee
Founding Editor of Advances in Photosynthesis and Respiration
Departments of Biochemistry and Plant Biology and Center of Biophysics & Computational Biology
University of Illinois at Urbana-Champaign, Illinois, USA
E-mail: gov@life.illinois.edu

Contents

Part II: Modeling of Light Harvesting and Primary Charge Separation

Part III: Modeling Electron Transport and Chlorophyll Fluorescence

Part IV: Integrated Modeling of Light and Dark Reactions of Photosynthesis

Part V: From Leaves to Canopies to the Globe

Preface

Scientific perception of nature relies on a process of transforming data to information, and then information into understanding. Data consist of observations and measurements and information is data organized according to some ontology, i.e. some set of assumptions about what entities exist and how they should be classified. Understanding is a model in the investigator's mind that describes how the entities relate to each other, a model created in the investigator's mind as a result of thinking. Thinking is thus a kind of self-programing of the brain, as a result of which understanding is achieved. When it "runs" in our brains, it allows us to predict the behavior of natural objects, e.g. in their temporal and spatial aspects. For communication within the scientific community, we first share new data, but then share the rigorous forms of the models, which may be verbal, graphic, or at their best, mathematical constructions, reflecting essential features of a natural system. The latter way of presentation of our understanding of photosynthesis is the subject of this book. In many chapters, the models are represented by differential equations that can reproduce the dynamics of the natural system, or in form of linear equations that define steady state fluxes or stoichiometries of such a system. A good model can not only reproduce already measured data about the behavior of the investigated system, but it can also predict results for future experiments.

By definition, models are approximations of nature that are by no means capable of capturing all aspects of the investigated system, no matter how powerful computers we may have used for it. In the early days of photosynthesis research, models were ingenious by their capacity to explain a prominent feature of the investigated process, such as, for example, the photochemical quenching of chlorophyll fluorescence. The early models were frequently relatively simple, not requiring a complex code or ontology. The closing of the reaction centers of Photosystem II during chlorophyll fluorescence induction was well described by Louis N. M. Duysens assuming a single component – the quencher Q. With increasing experimental accuracy and increasing complexity of the experimental protocols, this simple model was, in terms of Karl Popper's logic of scientific discovery, falsified or, in other words, its validity limits were found. The simple 'Quencher 'Q' model' of Duysens fell short, for example, in explaining the periodicity of four that occurs in chlorophyll fluorescence emission with multiple single turnover flashes, or in explaining the sigmoidal shape of the chlorophyll fluorescence induction curve. This and other models are perpetually expanding to explain new data obtained with new experimental protocols.

Such an expectation of the linear expansion of the models is by itself a simplified model. Sometimes an established, "generally accepted", feature of the model is replaced by another modification, a novel mechanism that explains already known data as well as the previously assumed mechanism, but widens and deepens the predictive power of the model. Thus, different models can explain similar or related phenomena, but only those are accepted for wider use that are able to accommodate new experimental data and more sophisticated protocols. The 'falsified' older and simpler models are not necessarily rejected and forgotten. Much more often they continue to be used with reservation about their range of applicability. For example, one does not need to consider the participation of pheophytin for the understanding of simple chlorophyll fluorescence induction curve on the time scale of seconds. In the area of whole photosynthetic process that specifically includes carbon fixation, Graham Faquhar, Susanne von Caemmerer and Joseph Berry have elegantly approximated it with two enzymatic reactions only. These ontogenetically older (in the sense of model development) models are typically easier to solve and can be obtained from the newer models by mathematically rigorous or empirical dimensionality reduction.

Photosynthesis is a complex process spanning from femtoseconds to days to seasons to centuries in time domain and from atoms to the global biosphere in spatial domain. No single model can describe photosynthesis in its full complexity and even approaching such an elusive goal would not be practical because such a mathematical model

would not be solvable, being as complex in its structure as nature itself. Rather, the process can be described in a mosaic of models such as the ones offered in this book. With increasing complexity of the models, we suggest that the readers consult the first two chapters of this book for a standardization of the model description, so that models become more than abstractions of individual modelers that are hard to share, merge or even compare with each other. We expect this book to be a beginning for creating a comprehensive modeling space of photosynthetic processes that would facilitate an ongoing 'falsification-upgrade' modeling spiral and would allow mergers between related model lines. The individual model areas represented here begin with the absorption of a photon and include electron transport, carbon assimilation, and product synthesis. With all these molecular models at hand one can upscale to cell, organ, plant, canopy, and eventually to global biosphere. Chapters presented in this book show how different levels of biological hierarchy overlay and interact in the amazing process of photosynthesis.

Photosynthesis *in silico* is a unique book in its integrated approach to the understanding of photosynthesis processes from light absorption and excitation energy transfer to global aspects of photosynthetic productivity – all interconnected by the use of mathematical modeling. The book is written by 44 international authorities from 15 countries. Chapters in this book are presented in a review style with emphasis on the latest breakthroughs. Instead of providing mathematical details, only the key equations, the basis for the novel conclusions, are provided, with references to the original work at the end of each chapter. Thus, *de facto* this is not a mathematical book of equations, but dominantly verbal discussion showing why the quantitative logic of mathematics has been so efficient for understanding the subject. Yet, in order to exploit the full potential of the book, we hope the models will eventually be translated to the universal format of the *Systems Biology Mark-Up Language* and made accessible also in their full mathematical form on the internet. As argued in Chapter 1, we stand here at the beginning of a qualitatively new scientific collaboration with its dynamics largely dependent on willingness of our research community to share resources to generate a free-access model database of photosynthesis. Such an endeavor is fully justified by an increasingly recognized role of photosynthesis in nature and lately also as an important alternative for technological solutions of currently surging energy needs of the humankind.

We thank our families and coworkers in our laboratories for their patience with us, and for their support during the preparation of this book. We also thank Noeline Gibson, Jacco Flipsen and André Tournois, of Springer, for their friendly and valuable guidance during the typesetting and printing of this book.

Agu Laisk
Institute of Molecular and Cell Biology
University of Tartu, Riia 23, Tartu 51010
Estonia

Telephone: 372 736 6021
Fax: 372 742 0286
e-mail: alaisk@ut.ee

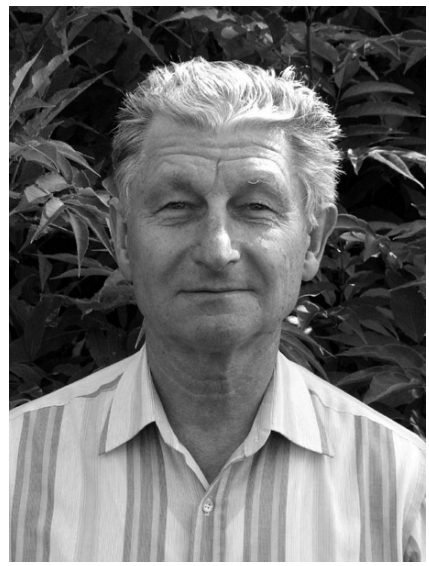
Ladislav Nedbal
Department of Biological Dynamics
Institute of Systems Biology and Ecology CAS
Zámek 136, 37333 Nové Hrady
Czech Republic

Telephone/Fax: 420 386 361 231
e-mail: nedbal@greentech.cz

Govindjee
Department of Plant Biology
University of Illinois at Urbana-Champaign
265 Morrill Hall, 505 South Goodwin Avenue
Urbana, IL 61801-3707, USA

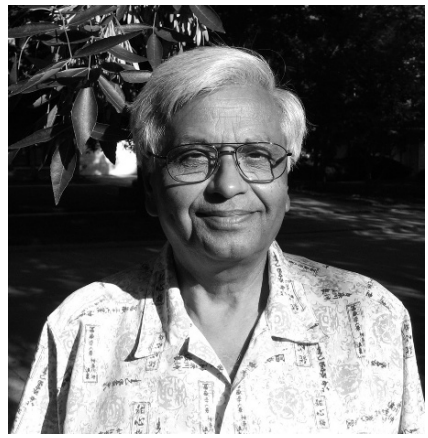
Telephone: 1 217 337 0627
Fax: 1 217 244 7246
e-mail: gov@illinois.edu

The Editors



Agu Laisk, born in 1938, obtained BS and MS degrees in Physics at the University of Tartu (Estonia) in 1961. He then joined the research group of Juhan Ross to study the penetration of sunlight into plant canopies for the purpose of modeling of plant productivity. His "candidate of science" work (equivalent to Ph.D., in the former Soviet Union), on the 'statistical distribution of light in the canopy', was completed in 1966. Since then he became interested in mechanisms that determine the rate of photosynthesis of a leaf. Together with his former student Vello Oja, he observed that in photosynthesis O_2 competes with CO_2 for one and the same acceptor and in 1970 published a mathematical model of photosynthesis and photorespiration, based on the competition of CO_2 and O_2 for ribulosebisphosphate (RuBP). Then he observed that at high CO_2 concentrations, O_2 enhances photosynthesis, showing the importance of the Mehler reaction. Soon thereafter, sophisticated experiments on "flashing" a leaf with short pulses of CO_2 showed that photosynthesis is limited by Rubisco at low CO_2, but by RuBP regeneration at high CO_2 levels. For these findings, the degree of Doctor of Science in Biology was awarded to him in 1976 by the Timiryazev Institute of Plant Physiology in Moscow (published as a monograph "*Kinetics of Photosynthesis and Photorespiration in C_3 Plants*" by "Nauka", Moscow, 1977). The specific approach of Laisk's group is in using only intact leaves as objects for measurements. This requires original equipment to be built in the laboratory – now appreciated in several other laboratories and, in principle, described in a book (together with Vello Oja) "*Dynamics of Leaf Photosynthesis. Rapid-Response Measurements and their Interpretations*", edited by Barry Osmond (CSIRO, Australia, 1998). A recent unexpected result from Laisk's laboratory is that cyclic electron transport around Photosystem I is much faster than necessary to cover the possible deficit in ATP synthesis – indicating that cyclic electron flow may be largely uncoupled from proton translocation or there must be a controllable proton leak. The interpretation of such kinetic experiments is unthinkable without the application of mathematical modeling. Agu Laisk is a Fellow Member of Estonian Academy of Science, life-time corresponding member of The American Society of Plant Biologists (ASPB), a member of the editorial board of *Photosynthesis Research* and of *Photosynthetica*. He has received National Science Awards from the Estonian Government. His international collaborators, who have deeply influenced his views, include: Ulrich Heber (Germany), David Walker (UK), Barry Osmond (Australia), Gerry Edwards (USA) and Richard Peterson (USA). At Tartu University, he teaches Bioenergetics.

Ladislav Nedbal, born in 1955, studied Biophysics at the Faculty of Mathematics and Physics, Charles University in Praha, Czech Republic. He graduated in 1981 with a thesis on the 'theory of the excitonic energy transfer in molecular crystals'. He learned about the fascinating process of photosynthesis from Ivan Šetlík, who is one of the founders of algal biotechnology. He moved over from doing modeling of energy transfer to research in experimental photosynthesis in the early years of his scientific career; this led to his present interest in modeling photosynthesis. Yet, preceding the present *déjà vu* with mathematical models were many more years of apprenticeship in experimental science that were marked with discreet advice from Govindjee. It was the present Series Editor of *Advances in Photosynthesis and Respiration*, who taught him the principles of technical writing in the late 1980s and introduced him, in 1990, to John Whitmarsh of the University of Illinois at Urbana-Champaign, USA. It was in John's lab where Nedbal discovered the photoprotective role of cytochrome b559. His other important tutors were Tjeerd Schaafsma in Wageningen, The Netherlands, and Anne-Lise Etienne in France. A significant inspiration came from David Kramer during the postdoc years in Urbana-Champaign and in Paris, where they collaborated in constructing a modulated light spectrophotometer and fluorometer. In that project he met Martin Trtilek, with whom he founded Photon Systems Instruments (PSI), a small company that has created a number of innovative instruments for photosynthesis research. The most important achievement was the development of the first commercial Pulse Amplitude Modulating (PAM) type imaging fluorometer – *FluorCam*. Recently, collaboration with Martin resulted in the construction of 'intelligent' photobioreactor for the cultivation of algae and cyanobacteria. The instrument collects, in real time, detailed information on the culture's photochemical yields and on its growth dynamics. The combination of mathematical modeling with experimental research in photosynthesis and engineering approaches logically led to another *déjà vu* in his career, this time with algal biotechnology. In this area, mathematical models bring to light yet unexplored pathways towards commercially viable use of algae and cyanobacteria. Further stimulating his interest in models are the mysterious dynamic features that he and his co-workers, including both his co-Editors of the present volume Agu Laisk and Govindjee, recently discovered in harmonically modulated light. Understanding plant behavior in dynamic light remains a major challenge that will be tackled by the current book *Photosynthesis in silico*.

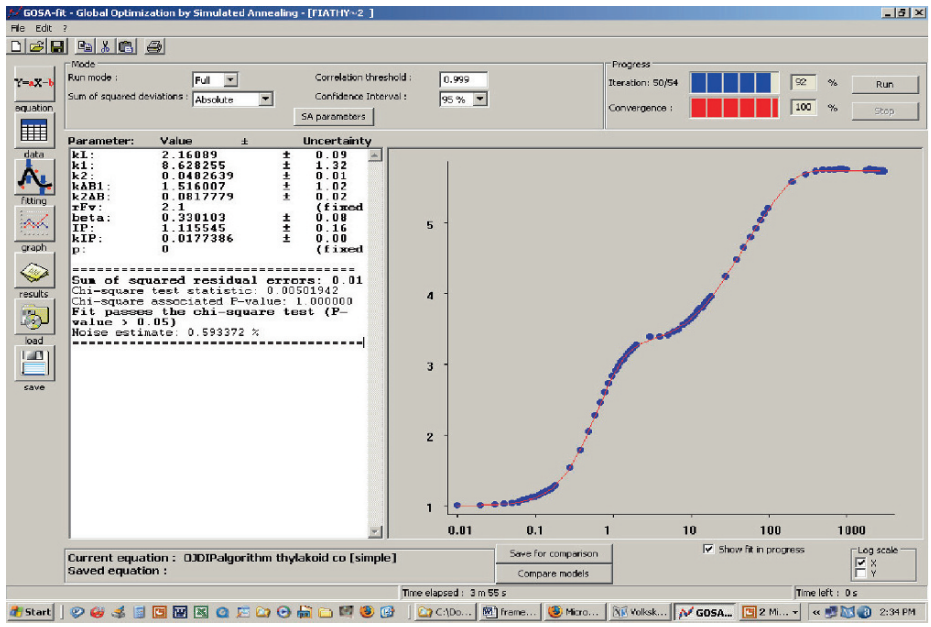


Govindjee, born in 1932, obtained his B.Sc. (Chemistry, Biology) and M.Sc. (Botany) in 1952 and 1954, from the University of Allahabad, India, and his Ph.D. (Biophysics) in 1960, from the University of Illinois at Urbana-Champaign (UIUC), IL, USA. His mentors were Robert Emerson and Eugene Rabinowitch. He is best known for his research on the excitation energy transfer, light emission, the primary photochemistry and the electron transfer in Photosystem II (PS II). His research, with many collaborators, has included the discovery of a short-wavelength form of chlorophyll (Chl) *a* functioning in the Chl *b*-containing system of PS II; of the two-light effects in Chl *a* fluorescence and in NADP reduction in chloroplasts (Emerson Enhancement); the basic relationships between Chl *a* fluorescence and photosynthetic reactions; the unique role of bicarbonate at the acceptor side of PS II. He provided the theory of thermoluminescence in plants, made the first picosecond measurement on the primary photochemistry of PS II and used Fluorescence Lifetime Imaging Microscopy (FLIM) of Chl *a* fluorescence in understanding photoprotection against excess light. His current focus is on the history of photosynthesis research, photosynthesis education, and possible existence of extraterrestrial life. He has served on the faculty of UIUC for about 40 years. Since 1999 he has been Professor Emeritus of Biochemistry, Biophysics and Plant Biology at the same institution. He is coauthor of '*Photosynthesis*' (with E. Rabinowitch; John Wiley, 1969), and editor of several books including *Bioenergetics of Photosynthesis* (Academic Press, 1975); *Photosynthesis*, Volumes I and II (Academic Press, 1982); *Light Emission of Plants and Bacteria* (with J. Amesz and D.C. Fork; Academic Press, 1986); *Chlorophyll a Fluorescence: A Signature of Photosynthesis* (with G.C. Papageorgiou, Springer, 2004); and *Discoveries in Photosynthesis* (with J.T. Beatty, H. Gest and J.F. Allen; Springer, 2005). His honors include: Fellow of the American Association of Advancement of Science; Distinguished Lecturer of the School of Life Sciences, UIUC; Fellow and Lifetime member of the National Academy of Sciences (India); President of the American Society for Photobiology (1980–1981); Fulbright Scholar and Fulbright Senior Lecturer; Honorary President of the 2004 International Photosynthesis Congress (Montréal, Canada); the 2006 Recipient of the Lifetime Achievement Award from the Rebeiz Foundation for Basic Biology; the 2007 Recipient of the Communication Award of the International Society of Photosynthesis Research (ISPR); and the 2008 Liberal Arts and Sciences Alumni Achievement Award of the University of Illinois. During 2007, *Photosynthesis Research* celebrated Govindjee's 50 years in Photosynthesis, and his 75th birthday through a two-Part special volume of the journal (Julian Eaton-Rye, editor). To celebrate his life-long achievement in Photosynthesis Research, Education, and its History, University of Indore, India, recently held a 3-day International Symposium (Nov. 27–29, 2008) on 'Photosynthesis in Global Perspective' (K.N. Guruprasad, Convener).

Govindjee has trained more than 20 Ph.D. students and about 10 postdoctoral associates.

Author Index

Color Plates



$$F(t)/F0 = 1+p+rFv^{*}(1-k1/(k1-kL)^{*}\exp(-kL^{*}t)+kL/(k1-kL)^{*}\exp(-k1^{*}t))^{*}((1-beta)^{*}kL/(kL+kAB1)+beta^{*}(1+(1-k1/(k1-kL)^{*}\exp(-kL^{*}t) +kL/(k1-kL)^{*}\exp(-k1^{*}t))^{*}\exp(-k2AB^{*}t)))+rFv^{*}(1-\exp(-k2^{*}t))+IP^{*}(1-\exp(-kIP^{*}t))^{*}(1+kIP^{*}t))\ [t\ in\ ms]$$

Fig. 1. Display of the curve fitting procedure, using global optimization simulation annealing (GOSA) with equation derived after summation of Eqs. (6.9–6.11) (*bottom line*). Symbols are experimental points, line is the simulated curve. Note that the fit was obtained after an about 4 min iteration time. See Chapter 6, p. 139

a

b

c		March 1996	October 1996	March 1997
	Tree height [m]	2.79	4.54	4.89
	Total leaf area [m^2] (gain/loss)	2.21	6.39 (9.46/5.28)	3.56 (6.70/9.54)
	Total leaf biomass [g] (gain/loss)	185.5	537 (794/443)	299 (564/802)
	Supporting tissue [g] (gain/loss)	374	1267 (1075/182)	2028 (1127/366)
	Aboveground biomass [g] (gain/loss)	560	1804 (1869/625)	2327 (1691/1168)

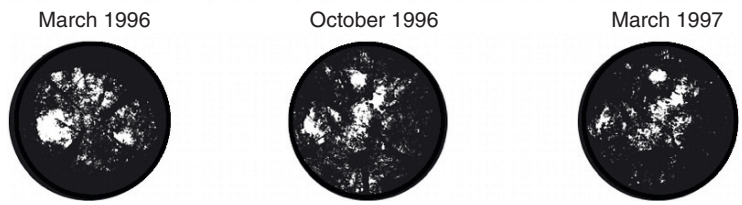
d	Time of investigation	380 d		
		via aboveground biomass monitoring	via 'steady-state' model	via 'dynamic' model
	Total carbon gain	3561 g	6393 g	4536 g
	Carbon allocation to supporting tissue	2202 g	35%	49%
	Carbon allocation to leaves	1359 g	21%	30%
	Resulting carbon allocation to roots		2832 g (44%)	975 g (21%)

Fig. 2. Development of an individual of the shade-intolerant pioneer *Ochroma lagopus* from an open site and deduction of its annual carbon allocation. Light green leaf area: sun exposed, dark green: (self-)shaded (from Timm et al., 2004). (**a**) Aboveground architectural development as reconstructed via the method described in Fig. 18.1; (**b**) change in the individual's light environment as indicated by hemispherical photography immediately above its uppermost leaves; (**c**) growth and biomass parameters of the respective individual; (**d**) deduction of annual assimilate flux balances (carbon allocation) as percentage of total annual crown carbon gain, either via a steady-state or a dynamic photosynthesis model. Carbon gain and allocation of biomass are given in equivalents of dry matter $(CH_2O)_n$. See Chapter 18, p. 423

a

Billia colombiana

b



c

	March 1996	October 1996	March 1997
Tree height [m]	2.1	2.26	2.45
Total leaf area [m^2] (gain/loss)	0.42	0.581 (0.204/0.045)	0.632 (0.089/0.038)
Total leaf biomass [g] (gain/loss)	27.6	38 (13.3/2.9)	41.3 (5.8/2.5)
Supporting tissue [g] (gain/loss)	58.7	90.0 (32.1/0.76)	103.0 (13.7/0.68)
Aboveground biomass [g] (gain/loss)	86.3	128.0 (45.5/3.7)	144.3 (19.5/3.2)

d

Time of investigation: 380 d

	via aboveground biomass monitoring	via 'steady-state' model	via 'dynamic' model
Total carbon gain	64.9 g	194.1 g	81.8 g
Carbon allocation to supporting tissue	45.8 g	24%	56%
Carbon allocation to leaves	19.1 g	10%	23%
Resulting carbon allocation to roots		129.2 g (66%)	16.9 g (21%)

Fig. 3. The same as Fig. 18.4, but for an individual of the mid- to late-successional shade-tolerant *Billia colombiana* below a closed canopy. Red leaf area: newly developed (from Timm et al., 2004). See Chapter 18, p. 424

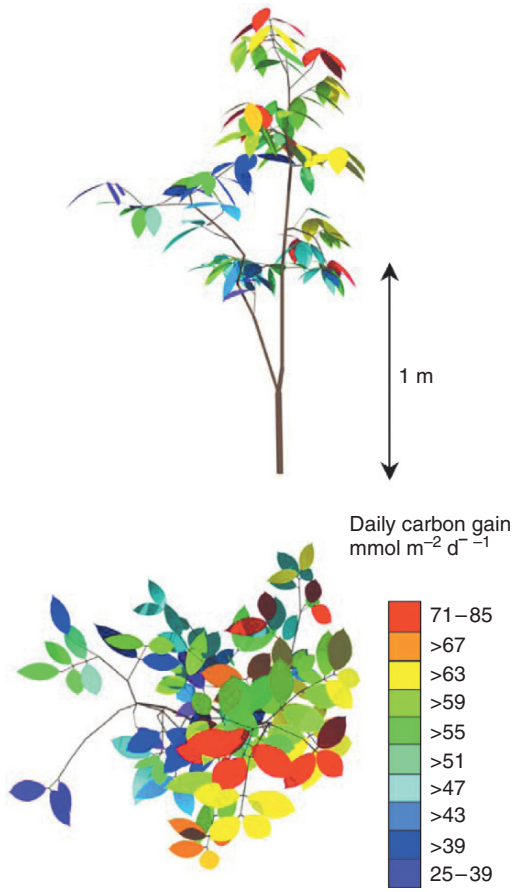


Fig. 4. Daily carbon balance of each individual leaf in the crown of a *Salacia petenensis* plant. Crown carbon gain was determined by summing up the individual balances. In the mean over 380 days carbon gain amounted to 426 mg day^{-1} (from Timm et al., 2004). See Chapter 18, p. 432

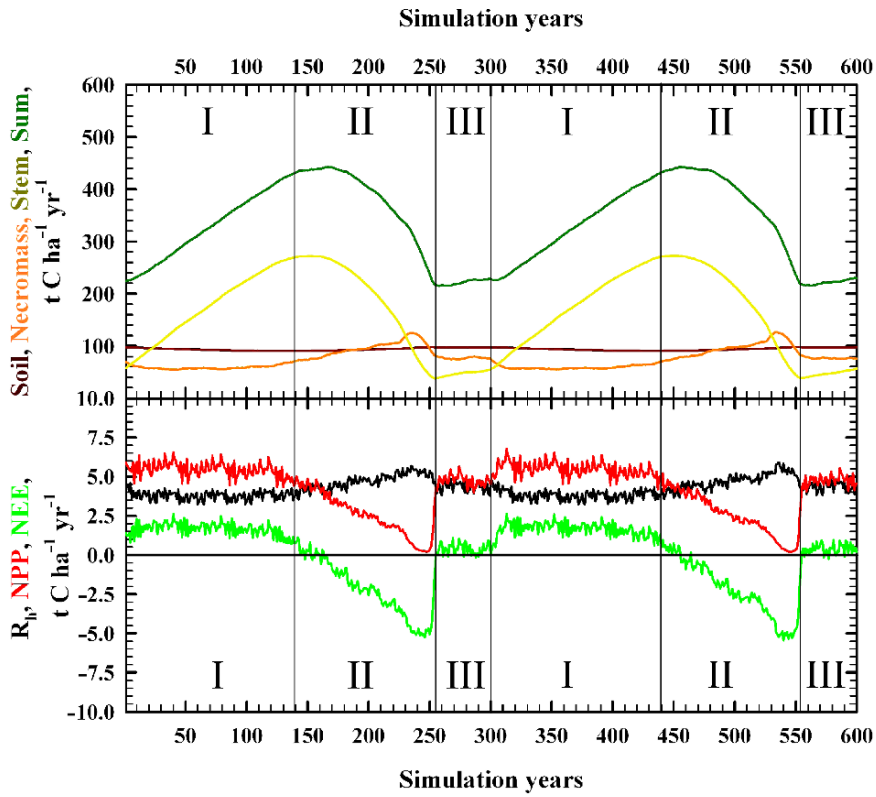


Fig. 5. Modeled pools and fluxes for the virgin forest reserve Rothwald using model parameters for Common beech forests. Upper graph: Comparison of the temporal development of modeled soil, necromass, stem and total ecosystem carbon content for 600 simulation years at landscape level steady state. Lower graph: Corresponding annual C fluxes from heterotrophic respiration (R_h), net primary production (NPP) and net ecosystem exchange (NEE). I – optimum phase; II – breakdown/regeneration phase; III – juvenescence (Pietsch and Hasenauer, 2006). See Chapter 19, p. 457

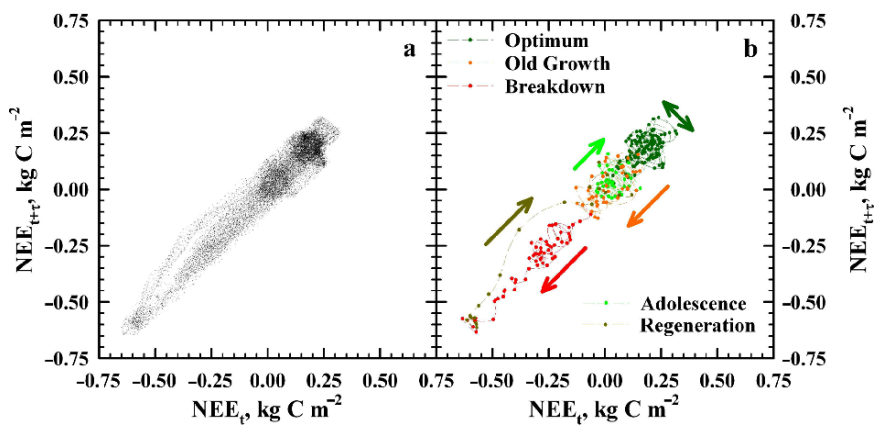


Fig. 6. Attractor of modeled NEE for the successional cycle evident within the virgin forest reserve Rothwald. **a**: NEE-Attractor for the virgin forest successional cycle reconstructed from model results using site and climate conditions of 18 research plots. **b**: Attractor reconstructed for one single plot and one successional cycle. Arrows indicate the trends of model behavior during different phases of the successional cycle. (S.A. Pietsch, unpublished) See Chapter 19, p. 460

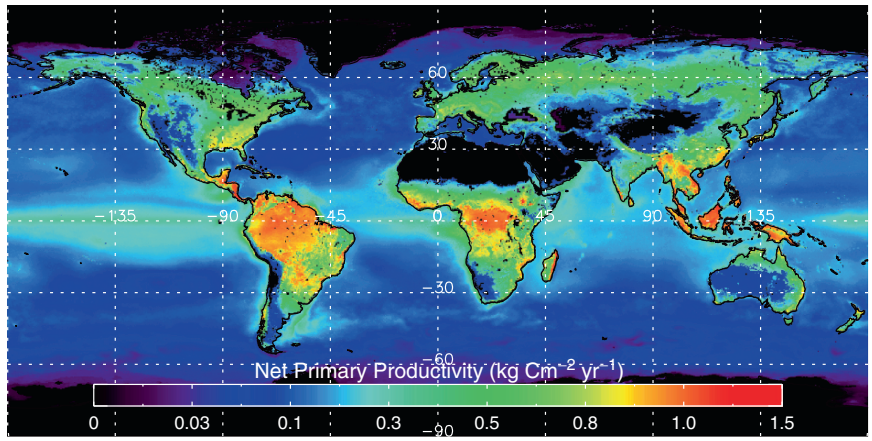


Fig. 7. Mean annual net primary productivity (NPP) simulated by Hybrid6.5 (land) and the CbPM (ocean) for the period 2000–2007. Total mean annual NPP is 107.3 Pg C year^{-1}, with 51.1% coming from land and 48.9% from the oceans. Land pixels simulated with $^1/_4°$ resolution and ocean pixels with $^1/_{12}°$ resolution. Land leaf area dynamics prescribed from MODIS satellite retrievals, ocean production calculated using data from the SeaWiFS instrument. Full simulation details are given in the text. See Chapter 20, p. 486

Part I

General Problems of Biological Modeling

Chapter 1

Trends and Tools for Modeling in Modern Biology

Michael Hucka*
Control and Dynamical Systems, California Institute of Technology, Pasadena, CA 91125, USA

James Schaff
Richard D. Berlin Center for Cell Analysis and Modeling, University of Connecticut Health Center, Farmington, CT 06030, USA

Summary

Computational modeling in biology requires sophisticated software tools. Precise communication and effective sharing of the models developed by researchers requires standard formats for storing, annotating, and exchanging models between software systems. Developing such standards is the driving vision behind the Systems Biology Markup Language (SBML) and several related efforts that we discuss in this chapter. At the same time, such standards are only enablers and ideally should be hidden "under the hood" of modeling environments that provide users with high-level, flexible facilities for working with computational models. As an example of the modern software systems available today, we discuss the Virtual Cell and illustrate its support for typical modeling activities in biology.

* Author for correspondence, e-mail: mhucka@caltech.edu

A. Laisk, L. Nedbal and Govindjee (eds.), Photosynthesis in silico: Understanding Complexity from Molecules to Ecosystems, pp. 3–15.

I. Introduction

Understanding the dynamic processes that are the essence of a living cell stands as one of the most important and most difficult challenges of twenty-first century biology. Today, it is widely appreciated that we can only hope to meet that challenge through the development and application of computational methods (Hartwell et al., 1999; Fraser and Harland, 2000; Arkin, 2001; Tyson et al., 2001; Noble, 2002; Alm and Arkin, 2003; Zerhouni, 2003), particularly the creation of mechanistic, explanatory models illuminating the functional implications of the data upon which they are built.

Models are not substitutes for experiments and data; rather, they are faithful teammates in the process of scientific discovery. A realistic computational model represents a modeler's dynamic understanding of the structure and function of part of a biological system. As the number of researchers constructing realistic models continues to grow, and as the models become ever more sophisticated, they collectively represent a significant accumulation of knowledge about the structural and functional organization of the system. Moreover, using them, the assimilation of new hypotheses and data can be done in a more systematic way because the additions must be fitted into a common, consistent framework. Once properly constructed, the models become a dynamic representation of our current state of understanding of a system in a form that can facilitate communication between researchers and help to direct further experimental investigations (Bower and Bolouri, 2001).

Today's models are large (and growing ever larger) and complex (and getting ever more complex). We are now long past the point of being able to communicate and exchange real-world models effectively by simply summarizing them in written narratives featuring a few equations. The precise communication of computational models between humans and between software is critical to being able to realize modeling's promise. Achieving this requires standardizing the electronic format for representing computational models in a way independent of any particular software – after all, different research goals are often best served by different software tools, yet modelers still need to share their results with their colleagues. At the same time, today's researchers need powerful software environments that offer a range of capabilities to support the creation, analysis, storage and communication of models, all the while hiding the details of the model representation format and providing biological modelers with high-level user interfaces and capabilities matched to the tasks they need to do.

In this chapter, we discuss both standards and software for computational modeling in biology. We summarize the *de facto* standard format, the Systems Biology Markup Language (SBML), as well as ongoing related efforts to standardize the representation of model annotations through MIRIAM (the Minimum Information Requested In the Annotation of biochemical Models) and SBO (the Systems Biology Ontology). As critical as they are, however, such standards are in the end only *enablers*; they are (hopefully) not what users interact with directly. We therefore also discuss software systems, focusing on one in particular, the Virtual Cell, as a way to present typical modeling activities in the context of one of today's most full-featured, interactive modeling environments. The advanced capabilities of systems such as Virtual Cell also help drive further development of SBML and adjunct efforts, and so we close with a summary of present work to extend SBML as well as standardize other areas of modeling and simulation exchange, such as the description of simulations.

Abbreviations: DOI – digital object identifier; MIASE – minimum information about a simulation experiment; MIRIAM – minimum information requested in the annotation of biochemical models; SBGN – systems biology graphical notation; SBML – systems biology markup language; SBO – systems biology ontology; SSA – stochastic simulation algorithm; UML – unified modeling language; URN – uniform resource name; VCell – virtual cell; XML – eXtensible markup language

II. Representing Model Structure and Mathematics

Until the late 1980s, publication of a computational model almost universally involved publishing only the equations and parameter values, usually with some narrative descriptions of how

the model was coded in software and how it was simulated and analyzed. The systems of equations were, with few exceptions, directly implemented in software: in a very direct sense, the program *was* the model. Authors sometimes even wrote their own numerical integration code. This general approach was necessary because of the primitive state of computational platforms and electronic data exchange, and it was fraught with problems. The most significant problem is simply the opportunities for errors that arise when a model must be recapitulated by humans into and back out of natural language form. The degree to which this is a real problem is startling. Curators for databases of published models such as BioModels Database (Le Novère et al., 2006) and JWS Online (Snoep and Olivier, 2003; Olivier and Snoep, 2004), report by personal communication that when they first began operation in the 2000–2004 timeframe, over 95% of published models they encountered had something wrong with them, ranging from typographical errors to missing information (even today, the problem rate is greater than 60%). A second problem is that, when a model is inextricably intertwined with its software implementation, it is difficult to examine and understand the precise details of the actual model (rather than artifacts of its particular realization in software). A third problem is that having to reconstruct a model from a paper is an extremely tall hurdle to fast, efficient and error-free reuse of research results.

Some areas of biological modeling improved on this situation in the 1990s. The field of computational neuroscience was particularly advanced in this regard, having two freely-available simulation packages, GENESIS (Bower and Beeman, 1995; Bower et al., 2002) and NEURON (Hines and Carnevale, 1997), supported on a variety of operating systems. These simulation platforms made it possible for modelers to distribute abstract definitions of their models and simulation procedures in the form of scripts that could be interpreted automatically by the platform software. The approach vastly improved the reusability of models. However, there remained the limitation that the formats were specific to the simulation package in which they were developed. Whoever wanted to reuse the models had to run the same software in order to reuse the model (assuming they were able to get the necessary files from the model's authors – electronic publishing of models as supplements to journal articles was still rare).

With the surge of interest in computational systems biology at the beginning of this century, software tools evolved one step further with the creation of *application-independent* model description formats such as CellML (Hedley et al., 2001) and SBML (Hucka et al., 2003, 2004). This form of representation is not an algorithm or a simulation script; it is a declarative description of the model structure that is then interpreted and translated by each individual software system into whatever internal format it actually uses. No longer tied to a particular software system, such software-independent formats permit a wider variety of experimentation in algorithms, user interfaces, services, and many other aspects of software tool development, by virtue of allowing multiple software authors to explore different facilities that all use the same input/output representation. In addition, and even more significantly, it enables practical publication of models in public databases.

The Systems Biology Markup Language (SBML; http://sbml.org) has become the de facto standard for this purpose, supported by over 120 software systems at the time of this writing. SBML is a machine-readable lingua franca defined neutrally with respect to software tools and programming languages. It is a model definition language intended for use by software – humans are not intended to read and write SBML directly. By supporting SBML as an input and output format, different software tools can all operate on the identical representation of a model, removing opportunities for errors in translation and assuring a common starting point for analyses and simulations. SBML is defined using a subset of UML, the Unified Modeling Language (Booch et al., 2000), and in turn, this is used to define how SBML is expressed in XML, the eXtensible Markup Language (Bray et al., 1998). Software developers can make use of a number of resources for incorporating SBML support in their applications (Bornstein et al., 2008).

SBML can encode models consisting of biochemical entities (species) linked by reactions to form biochemical networks. An important principle in SBML is that models are decomposed into explicitly-labeled constituent elements, the

set of which resembles a verbose rendition of chemical reaction equations; the representation deliberately does not cast the model directly into a set of differential equations or other specific interpretation of the model. This explicit, modeling-framework-agnostic decomposition makes it easier for a software tool to interpret the model and translate the SBML form into whatever internal form the tool actually uses. The main constructs provided in SBML include the following:

Compartment and *compartment type*: a compartment is a container for well-stirred substances where reactions take place, while a compartment type is an SBML construct allowing compartments with similar characteristics to be classified together.

Species and *species type*: a species in SBML is a pool of a chemical substance located in a specific compartment, while species types allow pools of identical kinds of species located in separate compartments to be classified together.

Reaction: a statement describing some transformation, transport or binding process that can change one or more species (each reaction is characterized by the stoichiometry of its products and reactants and optionally by a rate equation).

Parameter: a quantity that has a symbolic name.

Unit definition: a name for a unit used in the expression of quantities in a model.

Rule: a mathematical expression that is added to the model equations constructed from the set of reactions (rules can be used to set parameter values, establish constraints between quantities, etc.).

Function: a named mathematical function that can be used in place of repeated expressions in rate equations and other formulae.

Event: a set of mathematical formulae evaluated at a specified moment in the time evolution of the system.

The simple formalisms in SBML allow a wide range of biological phenomena to be modeled, including cell signaling, metabolism, gene regulation, and more. Significant flexibility and power comes from the ability to define arbitrary formulae for the rates of change of variables as well as the ability to express other constraints mathematically.

SBML is being developed in "levels". Each higher level adds richness to the model definitions that can be represented by the language. By delimiting sets of features at incremental stages, the SBML development process provides software authors with stable standards and the community can gain experience with the language definitions before new features are introduced. Two levels have been defined so far, named (appropriately enough) Level 1 and Level 2. The former is simpler (but less powerful) than Level 2. The separate levels are intended to coexist; SBML Level 2 does not render Level 1 obsolete. Software tools that do not need or cannot support higher levels can go on using lower levels; tools that can read higher levels are assured of also being able to interpret models defined in the lower levels. Open-source libraries such as libSBML (Bornstein et al., 2008) allow developers to support both Levels 1 and 2 in their software with a minimum amount of effort.

III. Augmenting Models with Semantic Annotations

The ability to have meaningful exchange of complex mathematical models of biological phenomena turns out to require a deeper level of semantic encoding and knowledge management than is embodied by a format such as SBML, which encompasses only syntax and a limited level of semantics. This realization came early in the context of CellML, whose developers added a standard scheme for metadata annotations soon after CellML was developed (Lloyd et al., 2004). CellML's metadata scheme was adopted by SBML at the beginning of the development of SBML Level 2, but limitations with the scheme later led the SBML community to seek alternatives. These were found in the form of the Systems Biology Ontology (SBO; http://www.ebi.ac.uk/SBO; Le Novère et al., 2006), and the Minimum Information Requested in the Annotation of Biochemical Models (MIRIAM; Le Novère et al., 2005).

A. Systems Biology Ontology (SBO)

The rationale for SBO is to provide controlled vocabularies for terms that can be used to annotate components of a model in SBML (or indeed, any other formal model representation format).

It requires no change to the form of the basic model in SBML; rather, it provides the option to augment the basic model with machine-readable labels that can be used by software systems to recognize more of the semantics of the model. SBO provides terms for identifying common reaction rate expressions, common participant types and roles in reactions, common parameter types and their roles in rate expressions, common modeling frameworks (e.g., "continuous", "discrete", etc.), and common types of species and reactions. Recent versions of SBML Level 2 provide an optional attribute on every element where an SBO term may be attached. Table 1.1 lists the correspondences between major components of SBML and SBO vocabularies.

The relationship implied by the attribute value on an SBML model component is "is a": the thing defined by that SBML component "is an" instance of the thing defined in SBO by indicated SBO term. By adding SBO term references on the components of a model, a software tool can provide additional details using independent, shared vocabularies that can enable other software tools to recognize precisely what the component is meant to be. Those tools can then act on that information. For example, if the SBO identifier SBO:0000049 is assigned to the concept of "first-order irreversible mass-action kinetics, continuous framework", and a given reaction in a model has an SBO attribute with this value, then regardless of the identifier and name given to the reaction itself, a software tool could use this to inform users that the reaction is a first-order irreversible mass-action reaction.

Table 1.1. Correspondence between major SBML components and controlled vocabulary branches in the Systems Biology Ontology (SBO)

SBML component	SBO vocabulary
Model	Interaction
Function definition	Mathematical expression
Compartment type	Material entity
Species type	Material entity
Compartment	Material entity
Species	Material entity
Reaction	Interaction
Reaction's kinetic law	Mathematical expression → Rate law
Parameter	Quantitative parameter
Initial assignment	Mathematical expression
Rule	Mathematical expression
Event	Interaction

As a consequence of the structure of SBO, not only children are versions of the parents, but the mathematical expression associated with a child is a version of the mathematical expressions of the parents. This enables a software application to walk up and down the hierarchy and infer relationships that can be used to better interpret a model annotated with SBO terms. Simulation tools can check the consistency of a rate law in an SBML model, convert reactions from one modeling framework to another (e.g., continuous to discrete), or distinguish between identical mathematical expressions based on different assumptions (e.g., Henri-Michaelis-Menten vs. Briggs-Haldane). Other tools like SBMLmerge (Schulz et al., 2006) can use SBO annotations to integrate individual models into a larger one.

SBO adds a semantic layer to the formal representation of models, resulting in a more complete definition of the structure and meaning of a model. The presence of an SBO label on a compartment, species, or reaction, can also help map SBML elements to equivalents in other standards, such as (but not limited to) BioPAX (http://www.biopax.org) or the Systems Biology Graphical Notation (SBGN, http://www.sbgn.org). Such mappings can be used in conversion procedures, or to build interfaces, with SBO becoming a kind of "glue" between standards of representation.

B. Minimum Information Requested in the Annotation of Biochemical Models (MIRIAM)

While SBO annotations help add semantics, there remains a different kind of impediment to effective sharing and interpretation of computational models. Figure 1.1 illustrates the issue.

When a researcher develops a model, they often use simple identifiers for chemical substances, or at best, only one of a multitude of possible synonyms for the substance. The situation is even worse when it comes to the chemical reaction and other processes: these are often given names such as "R1", "R2", etc., or at best, generic