Collaborative Computational Technologies for Biomedical Research

Edited by
Sean Ekins
Maggie A.Z. Hupcey
Antony J. Williams

Foreword by Alpheus Bingham
COLLABORATIVE COMPUTATIONAL TECHNOLOGIES FOR BIOMEDICAL RESEARCH
Wiley Series on Technologies for the Pharmaceutical Industry
Sean Ekins, Series Editor

Editorial Advisory Board

Dr. Renée J.G. Arnold (ACT LLC, USA)
Dr. David D. Christ (SNC Partners LLC, USA)
Dr. Michael J. Curtis (Rayne Institute, St Thomas’ Hospital, UK)
Dr. James H. Harwood (Delphi BioMedical Consultants, USA)
Dr. Maggie A.Z. Hupcey (PA Consulting, USA)
Dr. Dale Johnson (Emiliem, USA)
Prof. Tsuguchika Kaminuma, (Tokyo Medical and Dental University, Japan)
Dr. Mark Murcko, (Vertex, USA)
Dr. Peter W. Swaan (University of Maryland, USA)
Dr. Ana Szarfman (FDA, USA)
Dr. David Wild (Indiana University, USA)

Computational Toxicology: Risk Assessment for Pharmaceutical and Environmental Chemicals
Edited by Sean Ekins

Pharmaceutical Applications of Raman Spectroscopy
Edited by Slobodan Šašić

Pathway Analysis for Drug Discovery: Computational Infrastructure and Applications
Edited by Anton Yuryev

Drug Efficacy, Safety, and Biologics Discovery: Emerging Technologies and Tools
Edited by Sean Ekins and Jinghai J. Xu

The Engines of Hippocrates: From the Dawn of Medicine to Medical and Pharmaceutical Informatics
Barry Robson and O.K. Baek

Pharmaceutical Data Mining: Applications for Drug Discovery
Edited by Konstantin V. Balakin

The Agile Approach to Adaptive Research: Optimizing Efficiency in Clinical Development
Michael J. Rosenberg

Pharmaceutical and Biomedical Project Management in a Changing Global Environment
Scott D. Babler
COLLABORATIVE COMPUTATIONAL TECHNOLOGIES FOR BIOMEDICAL RESEARCH

Edited by

SEAN EKINS
MAGGIE A. Z. HUPCEY
ANTONY J. WILLIAMS
For Mum and Dad with thanks for letting me follow a route of my own.
Sean Ekins

For Motts, short but loud.
Maggie A. Z. Hupcey

For my twin sons, Taylor and Tyler—two of the best collaborators I know.
Antony J. Williams
In the long history of human kind (and animal kind, too) those who have learned to collaborate and improvise most effectively have prevailed.

Charles Darwin
CONTENTS

FOREWORD xi
Alpheus Bingham

PREFACE xv

CONTRIBUTORS xix

PART I GETTING PEOPLE TO COLLABORATE 1

1. The Need for Collaborative Technologies in Drug Discovery 3
   Chris L. Waller, Ramesh V. Durvasula, and Nick Lynch

2. Collaborative Innovation: The Essential Foundation of Scientific Discovery 19
   Robert Porter Lynch

3. Models for Collaborations and Computational Biology 39
   Shawnmarie Mayrand-Chung, Gabriela Cohen-Freue, and Zsuzsanna Hollander

4. Precompetitive Collaborations in the Pharmaceutical Industry 55
   Jackie Hunter

5. Collaborations in Chemistry 85
   Sean Ekins, Antony J. Williams, and Christina K. Pikas

6. Consistent Patterns in Large-Scale Collaboration 99
   Robin W. Spencer
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Author(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>Collaborations Between Chemists and Biologists</td>
<td>Victor J. Hruby</td>
</tr>
<tr>
<td>9</td>
<td>Intellectual Property Aspects of Collaboration</td>
<td>John Wilbanks</td>
</tr>
<tr>
<td></td>
<td><strong>PART II METHODS AND PROCESSES FOR COLLABORATIONS</strong></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Scientific Networking and Collaborations</td>
<td>Edward D. Zanders</td>
</tr>
<tr>
<td>11</td>
<td>Cancer Commons: Biomedicine in the Internet Age</td>
<td>Jeff Shrager, Jay M. Tenenbaum, and Michael Travers</td>
</tr>
<tr>
<td>12</td>
<td>Collaborative Development of Large-Scale Biomedical Ontologies</td>
<td>Tania Tudorache and Mark A. Musen</td>
</tr>
<tr>
<td>13</td>
<td>Standards for Collaborative Computational Technologies for Biomedical Research</td>
<td>Sean Ekins, Antony J. Williams, and Maggie A. Z. Hupcey</td>
</tr>
<tr>
<td>14</td>
<td>Collaborative Systems Biology: Open Source, Open Data, and Cloud Computing</td>
<td>Brian Pratt</td>
</tr>
<tr>
<td>15</td>
<td>Eight Years Using Grids for Life Sciences</td>
<td>Vincent Breton, Lydia Maigne, David Sarramia, and David Hill</td>
</tr>
<tr>
<td>16</td>
<td>Enabling Precompetitive Translational Research: A Case Study</td>
<td>Sándor Szalma</td>
</tr>
<tr>
<td>17</td>
<td>Collaboration in Cancer Research Community: Cancer Biomedical Informatics Grid (caBIG)</td>
<td>George A. Komatsoulis</td>
</tr>
<tr>
<td>18</td>
<td>Leveraging Information Technology for Collaboration in Clinical Trials</td>
<td>O. K. Baek</td>
</tr>
<tr>
<td></td>
<td><strong>PART III TOOLS FOR COLLABORATIONS</strong></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>Evolution of Electronic Laboratory Notebooks</td>
<td>Keith T. Taylor</td>
</tr>
</tbody>
</table>
20. Collaborative Tools to Accelerate Neglected Disease Research: Open Source Drug Discovery Model
   Anshu Bhardwaj, Vinod Scaria, Zakir Thomas, Santhosh Adayikkoth, Open Source Drug Discovery (OSDD) Consortium, and Samir K. Brahmachari

21. Pioneering Use of the Cloud for Development of Collaborative Drug Discovery (CDD) Database
   Sean Ekins, Moses M. Hohman, and Barry A. Bunin

22. Chemspider: a Platform for Crowdsourced Collaboration to Curate Data Derived From Public Compound Databases
   Antony J. Williams

23. Collaborative-Based Bioinformatics Applications
   Brian D. Halligan

24. Collaborative Cheminformatics Applications
   Rajarshi Guha, Ola Spjuth, and Egon Willighagen

PART IV THE FUTURE OF COLLABORATIONS

25. Collaboration Using Open Notebook Science in Academia
   Jean-Claude Bradley, Andrew S. I. D. Lang, Steve Koch, and Cameron Neylon

26. Collaboration and the Semantic Web
   Christine Chichester and Barend Mons

27. Collaborative Visual Analytics Environment for Imaging Genetics
   Zhiyu He, Kevin Ponto, and Falko Kuester

28. Current and Future Challenges for Collaborative Computational Technologies for the Life Sciences
   Antony J. Williams, Renée J. G. Arnold, Cameron Neylon, Robin W. Spencer, Stephan Schürer, and Sean Ekins

INDEX
You have in your hands a book on collaboration, more specifically a book on scientific collaboration, and most specifically, a book on collaboration in the science of pharmaceutical development—the discovery of new therapies and medicines—products addressing the, as-yet, unmet medical needs of twenty-first century health. While only a few would take issue with the merits of collaboration, perhaps even most fail to appreciate the implications of collaborative technologies in the present day. The ability to fuse ideas—especially ideas that cross disciplines—is a crucial capability responsible for accelerating innovation and progress. Matt Ridley recently gave a TED talk entitled, “When Ideas Have Sex,” the salient point being that the fusion of ideas, each bringing its own set of memes, is a powerful way of creating new memetic material.

People have collaborated as long as... well... as long as there have been people. Often nothing more than self-interest incites us to collaborate, to fill in portions of a solution important to us, portions we were not capable of creating on our own. Unfortunately, modern-day organizational structures very often serve as impediments to collaboration. Collaborating with those outside the walls of an institution may be more than culturally frowned upon, it may even be illegal under legislation written to hinder corporate espionage, or protect trade or national technological capabilities. (I guess if that were the only problem, it could be readily solved by a new set of policies or regulations.)

But institutional boundaries are not the only barriers that impede collaboration. Even within an institution—which should be legally, strategically, and financially incented for alignment, and for maximizing the opportunities for
internal collaboration—barriers still exist. The subunits of the institution: its departments, its divisions, its components produce collaboration “walls” of varying substantiality. Organizational lore and personal relationships add another layer of “not-invented here” (NIH) culture, and allegiances to local agendas, even to the point of disadvantaging the larger institutional unit. In fact, if we wish to pursue the elimination of collaboration barriers we have to realize that many barriers are not institutional at all. Choices to collaborate or not collaborate are sometimes based not just on current affiliations but on past affiliations, degrees obtained, reputations, and even a less than rational bias as to just who our collaboration partners should be.

A bright spot in recent history has been the open-source movement. It was loosely organized. It was NOT the project management assignment of any large corporate firm filled with project managers looking for substantial development programs like this one. While we acknowledge that there was a component of centralization, that is, Linus Torvald’s role in Linux, the majority of work was exercised in a distributed manner, each module remaining somewhat independent of the constraints often imposed by centralized planning functions. Most importantly, the basis upon which individuals contributed was informed solely by the contribution itself, not perceived qualifications or past reputations.

While the open-source movement has been associated primarily with the development of software, the demonstration that it can compete effectively with the traditional modes of corporate technology development raises the possibility that such collaborative forms will soon move well beyond software and into other arenas of complex development. This is more than mere speculation. In the chapters that follow you’ll see early endeavors to accomplish pharmaceutical development in a much more open manner. While these may still fall short of the phenomenon associated with Linux, they more than hint at a future to come. One barrier to this progression was highlighted in Harvard Business Review’s ten best business ideas for 2010; namely, the current lack of a well-accepted and digitized representation of this work. The vast majority of collaborative pharmaceutical development still remains primarily a local and classically social phenomenon.

While change is still impeded for the reasons described above, the corporate model of the fully integrated pharmaceutical company is under threat for very good reasons. In the past decade, it has shown its inability to create and sustain shareholder value. A closer examination of the business model itself reveals a variety of flaws (or features, if you’d prefer): long monetization cycles, large capital investments with high risks, and a complex union of both information and materials management. We might argue that a typical pharmaceutical company tries to operate, under one roof, three distinctive business entities. It is a high-tech manufacturer, producing exquisitely expensive fine chemicals or complex biotechnical products. It is a purveyor of information to the regulatory and medical communities, information with specifications and demands rarely matched in any other sector. And, finally, it is a high risk research
venture, which can only show returns when managed as a portfolio of complex assets demanding constant invention and breakthroughs.

Each of these three business entities would ideally be managed with a distinctive set of overarching strategies and yet such an approach is rarely accommodated. This book addresses, for the most part, only the unique challenge associated with managing large, complex, high-risk research endeavors. But of the three business-entity challenges cited here, a novel new approach to this one could transform the economics of the entire business.

Considering the present state the pharmaceutical industry finds itself in, the promise of innovative medicines for children and our children’s children may well depend on finding new collaborative paradigms with attendant business models. The material for this genesis, though nascent, may well be found in these pages.

Alpheus Bingham

April 2011