

Structural genomics and the organisation of open science

CRESO SÁ AND MERLI TAMTIK¹

Abstract

One of the most hotly debated issues in contemporary academic science is the tension between open scientific inquiry and the privatisation of knowledge through intellectual property. Nowhere is this tension more obvious than in the life sciences, where academic research and industrial R&D have become intermingled. This paper examines the open science approach employed in the research field of structural genomics, giving rise to specific arrangements to organise the production and dissemination of knowledge. Features of the structural genomics approach are identified and discussed as they relate to the characteristics of the field and more generally to scientific activity.

Introduction

One of the most hotly debated issues in science policy is the tension between open scientific inquiry and the privatisation of knowledge through intellectual property (IP). A varied literature has emerged examining this issue from multiple disciplinary perspectives.² In essence, this literature highlights how the conditions shaping academic science over the past generation have contributed to the increasing appropriation of research findings, methods, and tools by individuals and organisations, challenging the public ethos of science. Most obviously, firms sponsoring university research might create obstacles for the prompt dissemination of valuable knowledge. Universities too have become more aggressive in ascertaining their rights over the IP they generate, in many cases responding to deliberate government policy.³ There are growing concerns that this slows down the advancement of science and technological innovation.⁴

Reactions to these trends are apparent in movements to guard, restore, and expand the openness of science. Open science models have been proposed recently by various groups, including advocates of the “open access” and “open source” movements, scholars, and even biomedical and biotechnology companies.⁵ Maurer,⁶ acknowledging various related ideas and concepts, defines open science as comprehending “(a) full, frank, and timely publication of results, (b) absence of intellectual property restrictions, and (c) radically increased pre- and post-publication transparency of data, activities, and deliberations within research groups.” Such models have been advanced and advocated in a range of fields, including astrophysics, biology, geology, neuroscience, and physics.

This paper approaches and contributes to this debate from a perspective rooted in the sociology of science and institutional analysis.⁷ Critical to this perspective is how institutionally embedded actors deal with competing and often conflicting ideas about appropriate ways of organising scientific production. Scientists in fields at the forefront of entrepreneurial activity face hybrid norms, values, and incentives.

Commercial and academic orientations combine to create novel ideas about scientific activity and ways of organising research. Against this background, the contemporary debate around open science is seen as evidence of contestation and innovation arising from evolving ideas about scientific work and underlying practices.⁸

This paper examines how the emerging structural genomics research community elaborated and promoted specific arrangements for open science over the past decade. Structural genomics is a particular high-throughput approach to describe the three-dimensional (3D) structure of every protein encoded by a given genome, which is considered a critical step for research related to drug development. There is thus a strong nexus between science and innovation permeating research policy, industrial R&D strategies, and scientific activity in genomics. By taking a historical case study approach, this paper traces the development of practices that now inform structural genomics centres and consortia around the world. The context and arguments that contributed to the shaping of structural genomics as a scientific field are identified, as they are intertwined with the idea of promoting openness in scientific discovery. As such, the case of structural genomics suggests potentially valuable avenues for the organisation of the open production and dissemination of knowledge.

Theoretical approach

Contemporary sociological understanding of science is fundamentally informed by Merton's work.⁹ He famously proposed that science is (or should be) shaped by the norms of communalism, universalism, disinterestedness and organised scepticism. Scientists are driven towards discovery and recognition by peers, and only the open dissemination of new knowledge can lead to career advancement. Such orientation is fundamentally different from the economic logics guiding industrial research, where profit is pursued through the appropriation of research results. These ideal-typical characterisations have long provided useful markers for the sociological study of science, but they do not capture the diversity in research practices and cultures across space and time. The generic orientations of academics have long interacted with local concerns, interests, and traditions to produce diverse institutional cultures and settings.¹⁰

Conceptually, this paper considers the role of actors in scientific fields in the interpretation, adaptation, modification, and/or enactment of organisational models for scientific activity. Scientific actors in the life sciences are viewed as holding a heterogeneous set of normative views, values, and assumptions about the nature of science and the production and dissemination of knowledge. Multiple, sometimes competing institutional logics are at play – sets of “material practices and symbolic constructions” that constitute organising principles for scientific production and which are available to organisations and individuals to elaborate.¹¹ This is evident in the interplay between the open dissemination of knowledge and commercial objectives in the life sciences, where academic research and industrial R&D have become intermingled.¹² The life sciences have been described as a “new world” in which the traditional models of public and private science no longer accurately reflect research

practices in either academic or corporate settings.¹³ The distinction between the disinterested orientation to the advancement of public knowledge and entrepreneurial pursuits has become quaint, as university researchers and industrial scientists combine both sets of objectives.¹⁴

Institutional entrepreneurs are those actors that actively promote particular ideas and organising principles through individual and collective action. They mobilise symbolic and material resources to garner support for new ways of organising scientific work. At the individual level, the model of the “scientist-entrepreneur” has been assimilated into universities, where engagement with patenting, spin-off companies, and corporations is balanced with professors’ academic responsibilities.¹⁵ Institutional entrepreneurship may involve the creation of “new organisational models and policies that change the direction and flow of organisational activity”.¹⁶ That is evidenced in the life sciences by hybrid organisational and work arrangements combining elements of both public and private, shaped by shifting research funding patterns, new research techniques, and evolving organisational settings.¹⁷

The institutional context of the life sciences has generated debate and contestation. Consider the mounting concerns about expansive patenting practices in academic science. In the US in particular, where IP rights have expanded since the 1980s, an aggressive patenting regime has taken hold. Universities’ approach to technology transfer has in many respects been shaped by the growth of the patent-driven field of biotechnology, which represents a growing share of academic patents.¹⁸ Increasingly expansive patenting practices have led to IP claims on research tools and even basic scientific knowledge. The growing pursuit of academic patents has raised the spectre of a reduction in the free exchange of publicly available scientific knowledge. Many have criticised this trend, pointing to the costs and barriers it creates for both scientific research and corporate innovation.¹⁹ Heller and Eisenberg²⁰ describe an “anti-commons” scenario in the life sciences - a situation in which knowledge is underused because numerous IP owners can block each other. This is clearly unfavorable to traditional scientific practice in academic settings; there are indeed growing fears about secrecy in university research.²¹ Moreover, patent thickets (multiple overlapping IP rights) can be obstacles for firms seeking to commercialise new products.²² Initiatives to counter this trend have had multiple targets, including IP laws and regulations, university technology transfer policies and practices, and the behaviours of scientists.²³

The movement around open source is an example of institutional entrepreneurship. The open source approach is associated with non-restrictive types of licences and distributed modes of collaborations for product development. The idea of open source emerged in the 1980s in the software development community as a reaction to the growing commercialisation of software based on restrictive ownership and licensing practices.²⁴ The Open Source Initiative was established as an advocacy organisation to promote free software, explicitly working towards building the institution of open source licensing. From the start, the field of genomics was mentioned as an example

where open source principles would be especially beneficial.²⁵ Proponents of open source seek to expand traditional practices of research sharing and dissemination through special licensing arrangements and public databases that facilitate archiving and dissemination of information among researchers and potential users.²⁶ Many suggest that the idea of making fundamental information (such as software code) available to all potential users and innovators might be replicated in the life sciences. Bioinformatics has been a natural testing ground for the translation of open source models into biology, triggering several research initiatives.²⁷ Journals such as *PLoS Biology* have employed open access principles, and the US National Institutes of Health has implemented a public access policy governing the publication of research supported by the agency. Furthermore, open access to scientific information is increasingly seen as a promising opportunity for the scientific community to accelerate research discovery in life sciences.²⁸

Below, this paper examines how open science ideas and practices have been employed in the structural genomics community over the last decade. High expectations about the commercial value of research breakthroughs and close university-industry ties are *de rigueur* in genomics research.²⁹ Some assert that genomics is a “powerful new wave of health-related life sciences ... energized by the human genome project and the knowledge and tools it is spawning”.³⁰ The relevance of genomic information for drug discovery and development places the field at the intersection of scientific advance and technological development, highlighting the need for mechanisms to ensure the openness of research.

Proteins, drugs, and the making of a research field

For over four decades, structural biology research has examined the shape of proteins, which play critical roles in human health and diseases.³¹ A protein structure exhibits the arrangement of the protein's atoms, showing their design. This in turn indicates the normal functions of a protein, and how changes in form can cause illnesses. With the genome sequencing projects conducted in the 1980s and 1990s, a wealth of data on human genomes was made available, enabling more concerted research efforts to decipher the structure of human proteins. Knowledge gained from individual structures has impacted drug discovery, but only a fraction of new drugs has been developed through the systematic use of such data.³² Companies engaged in drug development, which is a costly endeavor with a high failure rate, are thus obvious beneficiaries from a better understanding of protein structures.³³ Some see the potential for comprehensive data on proteins structures to lead drug R&D towards a more directed and efficient process of discovery and design.³⁴

Nonetheless, deciphering the shape of proteins is technically demanding, requiring painstaking efforts from research teams to generate useful three-dimensional structures.³⁵ Technological advances in the 1990s have facilitated such work, making it faster and cheaper, although still laborious. Moreover, this research is not based on hypothesis testing, but on diligent problem solving that enables further scientific puzzles to be formulated. Added to the difficulty of the task, researchers may thus not

have the incentives to commit time and resources to this endeavor independently. By the late 1990s, less than 200 structures were deciphered annually, of which a tiny fraction was of human proteins.³⁶

With genome sequence data available, scientists, government research agencies, multinational companies, and private foundations involved in genomics research in many countries started to promote more systematic efforts in this area.³⁷ Advocates of structural genomics noted that without dedicated support, attention to deciphering protein structures was scattered and intermittent, and scientific rewards for this activity were questionable. For these reasons, proponents mobilised to call for collaborative and team science approaches to generating the scale and scope of expertise needed to advance research in the field.³⁸ In the USA, the National Institute of General Medical Sciences (NIGMS) of the National Institutes of Health (NIH) and the Department of Defense were early sponsors of pilot projects. In 1999, NIGMS launched the Protein Structure Initiative in the US. Japan, Canada, and Germany had initiated pilot projects by 1998, and were joined by other countries including Sweden, France, Italy, China, Brazil, Israel and Japan by the turn of the century.³⁹

Structural genomics research expanded as governmental funding agencies and private foundations moved to support long-range research programs in the 2000s (Table 1). The USA was home to the most expansive efforts; through the Protein Structure Initiative, the NIGM provided support for the creation of inter-institutional research centres in 2000. The first wave of investments totalled around \$341 million to support nine centres for five years.⁴⁰ In 2005, a second phase of the initiative renewed commitments to some of the initial units and supported five additional centres. A clear division of labour took shape among centres of different types, with some units focusing on research technologies and methodologies. Foreign institutions were also enlisted as collaborators. A total of \$325 million was allocated to the 14 centres.

Japan, Canada and several European countries also developed large structural genomics centres over the decade. Japan established Protein 3000, a five-year initiative to decipher 3,000 protein structures. Most of the work, and half of the funds, was allocated to the RIKEN research institute, and specific target lists were addressed by other laboratories.⁴¹ In Germany, the Protein Structure Factory spun off of the German Human Genome Project to be a pioneering project in Europe. The Oxford Protein Production Facility in Oxford, UK and the Genopoles in France followed.⁴² The European Union funded a major collaboration involving up to 20 research institutions in 2002, the Structural Proteomics in Europe (SPINE) project, which was succeeded by SPINE2 in 2006.⁴³ In Canada, the Montreal-Kingston Bacterial Structural Genomics Initiative and the Ontario Centre for Structural Proteomics were created early in the decade. Another major project co-hosted in Canada, the Structural Genomics Consortium (SGC), was established by a transnational coalition of funding agencies and pharmaceutical companies.⁴⁴ Research sites were established in 2004 at the University of Toronto and the University of Oxford, and a year later the Karolinska Institutet joined the consortium with the backing of Swedish sponsors.

Table 1. Examples of Structural Genomics Centres

Centres	Lead Organisation	Collaborating Institutions
RIKEN Structural Genomics/Proteomics Initiative http://www.rsgi.riken.go.jp	Research Institute for Physical and Chemical Research (RIKEN), Genomic Science Center (JPN)	University of Tokyo (JPN), Hokkaido University (JPN), National Institute for High Energy Physics (JPN), Kyoto University (JPN), Osaka University (JPN), Ministry of Education (MEXT)(JPN)
Midwest Center for Structural Genomics http://www.mcsg.anl.gov/	Argonne National Laboratory (USA)	Northwestern University (USA), Washington University School of Medicine (USA), European Bioinformatics Institute (UK), University College London (UK), UT Southwestern Medical Center at Dallas (USA), University of Toronto (CAN), University of Virginia (USA)
Joint Center for Structural Genomics http://www.jcsg.org/	Scripps Research Institute (USA)	The Genomics Institute of The Novartis Research Foundation (USA), The University of California, San Diego (USA); the Sanford-Burnham Medical Research Institute (USA), The Stanford Synchrotron Radiation Lightsource (USA), SLAC National Accelerator Laboratory (USA).
New York Structural Genomics Research Consortium http://www.nysgxrc.org/	Structural GenomiX, Inc. (USA)	Albert Einstein College of Medicine (USA), SGX Pharmaceuticals, Inc. (USA), Brookhaven National Laboratory (USA), Case Western Reserve University (USA), University of California, San Diego (USA)
Northeast Structural Genomics Consortium http://www.nesg.org/	Rutgers University (USA)	Robert Wood Johnson Medical School (USA), Columbia University (USA), Miami University (USA), The State University of New York, Buffalo (USA), Hauptman Woodward Research Institute (USA), University of Toronto (Canada), University of Georgia (USA).
Structural Genomics Consortium http://www.thesgc.org/	University of Toronto (CAN)	University of Oxford (UK), Karolinska Institutet (SWE)

These efforts in North America, Europe, and Japan concentrated resources on large programmes, organised as centres or consortia involving investigators from multiple institutions. The large investments in structural genomics initiatives created a double challenge for research centres over the decade. The first was to meet the expectations created by proponents of the structural genomics of generating protein shapes efficiently. It involves focusing on organising protein sequences into families, selecting targets within those families, deciphering their three-dimensional structure employing X-ray crystallography or NMR spectroscopy, and finally building models from those structures to examine other similar proteins. Structural genomics uses high-throughput methods to decipher a large number of proteins structures in a systematic manner, as opposed to isolating individual proteins and focusing on them. By seeking economies of scale, this approach promised to lower the costs of identifying protein shapes, which may run at around \$100,000 per protein.⁴⁵

The rise and establishment of a scientific field is a dynamic social process, resulting from complex processes of negotiation, conflict and competition.⁴⁶ Rather than simply representing the “natural” unfolding of intellectual developments, research fields may emerge out of protracted struggles between competing groups over intellectual territory. As resources are scarce, already established fields contested rising scientific endeavours in structural genomics. Thus, the second challenge for the research community was to justify the scientific value of their achievements vis-à-vis traditional structural biology, particularly in light of the substantial investments made in structural genomics initiatives. Structural biologists question the fundamental premise of structural genomics as a distinctive field: the utility and desirability of engaging in large-scale discovery of protein shapes.⁴⁷ The structural genomics community acknowledges these concerns, and calls for greater exchange with the biosciences, but has defended its discovery-oriented approach as generating both significant new knowledge and technical advances.⁴⁸ These discussions reflect the perspectives of contending scientific approaches, in light of the voluminous resources invested on structural genomics centres over the decade. Alongside these debates, the structural genomics initiatives sparked a considerable amount of research, carried out according to open science principles and practices across research centres around the world.

An Open Science Approach

The NIH, the UK medical charity the Wellcome Trust, and the Japanese Ministry of Science were key agents in coalescing the international structural genomics community. They sponsored scientific meetings in the early 2000s, bringing together researchers and sponsors from North America, Japan and Europe.⁴⁹ At the first international meeting held at the Wellcome Trust's Genome Campus in 2000, a diverse range of participants deliberated on critical foundational issues for the nascent research community. Those included overarching research goals, the boundaries of the field and the promotion of international cooperation. Considering that the audience included scientists, funding agencies, and pharmaceutical companies, it

comes as no surprise that the discussion also centred upon ways “to ensure the openness” of research, and “the relationship to industrial activities”.⁵⁰

The need for shared principles for dealing with intellectual property came to the forefront as one of the main issues in the debate.⁵¹ It was agreed that the primary goal of structural genomics is to generate publicly available data and tools to support further scientific and technological advances. The role of biomedical innovations from this knowledge base was duly noted in the 2000 conference; “Raw fundamental data ... should be made freely available to researchers everywhere. However, intellectual property protection for inventions based on these can play an important role in stimulating the development of important new health care projects. Policies should be established to permit an appropriate balance between these goals”.⁵² From the very start, actors in the structural genomics community were well aware of the nexus between science and drug discovery. Negotiating the balance between these activities became one of the focal points of the emerging research community.

The second international meeting held in Airlie, Virginia in 2001 consolidated this debate. The meeting resulted in the “Airlie agreement”, a set of principles guiding the development of the structural genomics field, adhered to voluntarily by participating organisations. The agreement clearly recognises the importance of commercial outcomes of structural genomics research, but reacts to the growing trend around expansive patenting practices. It encourages “efforts to strengthen the utility requirement for patentability,” as the “community is concerned about the implications of the granting of patents based solely on the submission of three dimensional structural coordinates, without any identified non-trivial utility”.⁵³ With such cautions, the agreement sought to delineate what constitutes a scientific contribution to the knowledge commons (3D structural coordinates) for the emerging research community. That this needed to be discussed and elaborated is evidence of the fuzzy separation between basic knowledge and patentable inventions in the life sciences described above.

Besides, the Arlie agreement set detailed guidelines for the public dissemination of data, information, and methods by the structural genomics community. Those include the public release of structure data in the Protein Data Bank, a longstanding data repository;⁵⁴ progress updates on research projects; the exchange of information about proteins targeted, to avoid duplication of efforts across centres; and the exchange of technology, including research protocols and software.⁵⁵ Short papers and electronic publications were encouraged to accompany the release of protein structure data, maximising the dissemination of relevant information, in addition to fully fledged journal articles. The International Structural Genomics Organization (ISGO) was created to champion these principles. As they chart new directions, institutional entrepreneurs often need to develop new organisational models and templates to guide action. In this case, open science was operationalised into specific practices and protocols for the production and dissemination of knowledge.

The Airlie agreement also acknowledges the tensions between these open science principles and the pursuit of IP. Regarding the disclosure of structure data, it recommends the immediate release of all data, but recognises that at times there may be delays of up to six months after structures are deposited for all accompanying data to be made public. This time was considered adequate for researchers to “assess intellectual property prospects and to file a patent application if desired.”⁵⁶ Individual sponsors have, of course, made their own choices in light of the IP regulations in their countries.⁵⁷ These provisions reflect and recognise the hybrid practices that prevail in the life sciences.

Moreover, the Airlie agreement has shaped the *modus operandi* of structural genomics centres over the past decade. As stipulated in the meetings leading to the agreement, information on the research goals of each centre – their list of protein targets – is made available on institutional websites and deposited on a public database, TargetDB.⁵⁸ At a minimum, centres also make project information available on their websites including updates on their protein targets. Large initiatives such as SPINE2⁵⁹ and SGC⁶⁰ post detailed technical information on their projects as well. By sharing this information, the centres can help avoid duplication of efforts across research groups, as scientists can consult TargetDB for the lists of targets being worked on elsewhere. Moreover, making goals publicly available allow research centres to have their efforts assessed. Research centres update the status of research on protein targets, allowing for an evaluation of the structural genomics research effort.⁶¹

The systematic approach to data disclosure proposed in the Airlie agreement quickly yielded dividends in terms of the availability of data. The most visible of these are the protein shapes deposited in the Protein Data Bank, a public data repository (see Table 2). Protein structure information is retrievable from the PDB in a variety of formats by anyone with access to the Internet. Structural genomics centres have also made a range of research tools and methods available, as part of deliberate efforts of initiatives such as PSI and SPINE.⁶² For instance, PSI created a website called KnowledgeBase,⁶³ to facilitate access to raw data, research tools, methods, and data analyses produced by sponsored centres.

Table 2. Structural Genomics Centres Ranked by Number of Protein Structures Deposited on the Protein Data Bank, 2010⁶⁴.

Project	N
RIKEN Structural Genomics/Proteomics Initiative	2675
Midwest Center for Structural Genomics	1213
Joint Center for Structural Genomics	1014
New York SGX Research Center for Structural Genomics	928
Northeast Structural Genomics Consortium	854
Structural Genomics Consortium	843
Center for Eukaryotic Structural Genomics	198
TB Structural Genomics Consortium	196
Seattle Structural Genomics Center for Infectious Disease	192
Center for Structural Genomics of Infectious Diseases	176
Southeast Collaboratory for Structural Genomics	117
Structural Proteomics in Europe	115
Berkeley Structural Genomics Center	95
Montreal-Kingston Bacterial Structural Genomics Initiative	84
Structural Genomics of Pathogenic Protozoa Consortium	70
Structure 2 Function Project	52
Ontario Centre for Structural Proteomics	30
Medical Structural Genomics of Pathogenic Protozoa	24
New York Consortium on Membrane Protein Structure	23
Mycobacterium Tuberculosis Structural Proteomics Project	22
Oxford Protein Production Facility	22
Accelerated Technologies Center for Gene to 3D Structure	21
Israel Structural Proteomics Center	20
Center for Structures of Membrane Proteins	18
Integrated Center for Structure and Function Innovation	16
Marseilles Structural Genomics Program @ AFMB	11

The open science approach championed by ISGO and adopted through the initiatives described above has not been inimical to the involvement of the private sector. SGC, the sixth most prolific centre (Table 2), focuses on human protein structures with medical relevance, which are potentially valuable for the pharmaceutical industry. Spearheaded by the Wellcome Trust, SGC is funded by GlaxoSmithKline, Merck, and Novartis, in addition to funding agencies in Canada, the UK, and Sweden. SGC was set up as a non-profit organisation, and all participating sponsors and universities agreed to join the consortium following principles that resonate with the Arlie agreement. Crucially, the shapes of proteins targeted by the consortium are made publicly available without any restrictions on use as soon as they are discovered.⁶⁵ Withholding of such information, whether for the pursuit of publications or patents, is

strictly prohibited. The involvement of the pharmaceutical companies in the funding of the consortium is thus not predicated on privileged access to research discoveries, but on participating in the process of research agenda setting.⁶⁶ This is consistent with recent observations that an excessively patent-driven regime in the life sciences has at times been inimical to the interests of biotechnology and pharmaceutical companies.⁶⁷

Conclusion

The case of structural genomics reflects these ongoing efforts to create open science spaces in the life sciences, an aim shared by a range of actors in the public and private sectors.⁶⁸ In recent years there have been calls for open source biotechnology, open source genomics, and open source biology.⁶⁹ Critics point out that the opportunity to benefit from openly available data in the life sciences is often limited primarily to companies in the biotechnology and pharmaceutical industries. Others argue that promoting decentralisation of research efforts and open access to research results provides a cost-effective way to overcome the innovation crisis in new drug discoveries.⁷⁰ Moreover, proponents of open source science argue that its norms of widely sharing basic data (e.g. software code or 3D protein structures) can lead to much stronger scientific communities than disseminating research results through traditional channels (e.g. publications).⁷¹

Through the collective efforts of funding agencies, universities, and sponsors from the private sector, structural genomics centres have been created over the last decade employing similar organisational models. Operating in an arena where hybrid norms and practices have taken shape, such actors have actively confronted the challenges inherent in conciliating disparate orientations and often competing priorities of discovery and appropriation into a model of scientific production based on open science principles. The structural genomics centres follow a Big Science model whereby large teams of scientists have been funded in a coordinated fashion to achieve pre-determined goals. Open science has been pursued through a number of norms and practices advocated by public and private funding agencies, and codified in the Arlie agreement – a voluntary accord among participating organisations.

In some ways, the structural genomics community has built upon previous initiatives in the life science that provided relevant models for organising research activities. These include the genome sequencing projects that generated fundamental scientific knowledge for immediate release in the public domain through large-scale consortia, and the use of electronic databases for archiving data shared by other fields.⁷² The importance of these earlier projects cannot be underestimated, as they shaped the pool of ideas in circulation in the scientific community on how to create and sustain large-scale research programs with similar goals as those of structural genomics projects.

In other ways, the open science approach employed in the structural genomics community relates to the nature of the field. Sociological studies of scientific fields document the diversity in intellectual orientations, research processes and technologies, social structures, and forms of communication.⁷³ This variety impacts

how scientific work is organised and transmitted. The open science approach in structural genomics consortia is the outcome of the complex social process of negotiation among different stakeholders resulting in establishing a research community with distinct operational principles and research goals. Structural genomics can be described by Whitley's⁷⁴ concept of a “technologically integrated bureaucracy.” In such fields, the research technology integrates theory, methods, and relevant problems, leading to specialisation across research groups with a high division of labour. Scientists are expected to pursue narrowly defined problems, and knowledge is highly specific and empirical, focusing on a relatively large number of properties of particular phenomena rather than on highly general and abstract features of fundamental knowledge. Coordination and consensus in the research community on technical and strategic aspects is gained through a decentralised approach characterised by specific local objectives. These features are apparent in the focus on high throughput technologies, the shared research goals, and the features of structural genomics centres described above.

The structural genomics initiatives illustrate possible pathways for the deliberate adoption of open science norms and practices in other fields. They include: (a) codified formats and procedures for the release of research findings, (b) shared expectations among research sites about the public dissemination of project information in similar online platforms, (c) common use of electronic data repositories, and (d) the priority given to the sharing of research data and tools over the immediate pursuit of IP. Through these measures, the structural genomics community has not only reaffirmed the traditional public ethos of science, but it has advanced collective standards for managing the openness of various stages of the research process.

¹ University of Toronto, Ontario Institute for Studies in Education. Correspondence to: c.sa@utoronto.ca

² D. Blumenthal et al. Withholding Research Results in Academic Life Science. Evidence From a National Survey of Faculty. *The Journal of the American Medical Association* 1997; 277: 1224–1228; M. Heller and R. Eisenberg. Can Patents Deter Innovation? The Anticommons in Biomedical Research. *Science* 1998; 280: 698–701; E. Campbell, et al. Data Withholding in Academic Genetics—Evidence From a National Survey. *The Journal of the American Medical Association* 2002; 28: 473–480; R. Eisenberg and R. Nelson. R.S. Public vs. proprietary science: a fruitful tension? *Daedalus* 2002; 131: 89–101; P. David. Can “Open Science” be Protected from the Evolving Regime of IPR Protections? *Journal of Institutional and Theoretical Economics* 2003;160 (1): 9-34; F. Murray and S. Stern. When Ideas Are Not Free: The Impact of Patents on Scientific Research. *Innovation Policy and the Economy* 2006; 7: 33-69; D.

Rhoten and W. Powell. The Frontiers of Intellectual Property: Expanded Protection versus New Models of Open Science. *Annual Review of Law and Social Science* 2007; 3: 345-373.

³ D. Mowery et al. 2004. *Ivory Tower and Industrial Innovation: University-Industry Technology Before and After the Bayh-Dole Act*. Stanford, CA. Stanford University Press.

⁴ R. Nelson. The market economy, and the scientific commons. *Research Policy* 2004. doi:10.1016/j.respol.2003.09.008.

- ⁵ Rhoten & Powell, op. cit. note 2; A. Rai and R. Eisenberg. Bayh-Dole Reform and the Progress of Biomedicine. *Law and Contemporary Problems* 2003; 66 (1/2): 289-314.
- ⁶ S. Maurer. 2003. New Institutions for Doing Science: From Databases to Open Source Biology. A paper presented to the European Policy for Intellectual Property Conference on Copyright and database protection, patents and research tools, and other challenges to the intellectual property system. University of Maastricht, The Netherlands, November 24-25, 2003.
- ⁷ S. Shapin. 2008. The Scientific Life. In *The new institutionalism in organizational analysis* W. Powell and P. DiMaggio, eds. Chicago, IL: The University of Chicago Press: 143-163.
- ⁸ See for example: J. Colyvas. 2007. From divergent meanings to common practices: The early institutionalization of technology transfer in the life sciences at Stanford University. *Research Policy* 2007; 36 (4): 456-476; W. Powell and J. Owen-Smith. 2002. The New World of Knowledge Production in the Life Sciences. In *The Future of The City of Intellect*. S. Brint ed. Stanford, CA. Stanford University Press: 107-132.⁹ R. K. Merton. 1942. The Normative Structure of Science. In Merton, Robert King (1973) *The Sociology of Science: Theoretical and Empirical Investigations*. Chicago, IL: University of Chicago Press.
- ¹⁰ Shapin, op. cit. note 7.
- ¹¹ R. Friedland and R. Alford. Bringing Society Back In: Symbols, Practices, and Institutional Contradictions. *The New Institutionalism in Organizational Analysis*. 1991. Chicago, IL: University of Chicago Press: 232-263.
- ¹² Powell and Owen-Smith, op. cit. note 8; D. Kleinman. 2003. *Impure Cultures: University Biology and the World of Commerce*. Madison WI: The University of Wisconsin Press; L. Smith-Doerr. Institutionalizing the Network Form: How Life Scientists Legitimate Work in the Biotechnology Industry. *Sociological Forum* 2005; 20 (2): 271-299; Colyvas, op. cit. note 8.
- ¹³ Powell & Owen-Smith, op. cit. note 8.
- ¹⁴ Shapin, op. cit. note 7; Kleinman, op. cit. note 12.
- ¹⁵ Powell & Owen-Smith, op. cit. note 8; Kleinman, op. cit. note 12; Smith-Doerr, op. cit. note 12; Colyvas, op. cit. note 8; Shapin, op. cit. note 7.
- ¹⁶ H. Hwang & W. Powell. 2005. Institutions and entrepreneurship. In *Handbook of Entrepreneurship Research*. Z. Acs and D. Audretsch eds. Kluwer Publishers: 179-210.
- ¹⁷ Rhoten & Powell, op. cit. note 2.
- ¹⁸ Mowery et al, op. cit. note 3.
- ¹⁹ Heller & Eisenberg, op. cit. note 2, Rai & Eisenberg, op. cit. note 5; J. Reichman and P. Uhler. A Contractually Reconstructed Research Commons for Scientific Data in a Highly Protectionist Intellectual Property Environment. *Law and Contemporary Problems* 2003; 66: 315-462.
- ²⁰ Heller & Eisenberg, op. cit. note 2.
- ²¹ Blumenthal et al, op. cit. note 2; J. Walsh, C. Cho and W. Cohen. The View from the Bench: Patents, Material Transfers, and Biomedical Research. *Science* 2005; 309: 2002-3; W. Hong and J. Walsh. For Money of Glory? Commercialization, Competition, and Secrecy in the Entrepreneurial University. *The Sociological Quarterly* 2009; 50(1): 145-171.
- ²² G. Clarkson and D. DeKorte. The Problem of Patent Thickets in Convergent Technologies. *Annals of the New York Academy of Sciences* 2006; 1093:180-200.
- ²³ Nelson 2004, op. cit. note 4; Rhoten & Powell, op. cit. note 2.
- ²⁴ S. Weber. 2004. *The Success of Open Source*. Cambridge: Harvard University Press.
- ²⁵ Open Source Initiative (OSI). 2007. The Open Source Definition, version 1.9. Available at <http://www.opensource.org/docs/definition.php>

- ²⁶ Maurer, op. cit. note 6; J. Hope. 2008. *Biobazaar: The Open Source Revolution and Biotechnology*. Cambridge: Harvard University Press.
- ²⁷ S. Maurer, A. Rai, and A. Sali. Finding Cures for Tropical Diseases: Is Open Source an Answer? *PLoS Medicine* 2004; 1(3): 56; J. Hope. 2008. *Biobazaar: The Open Source Revolution and Biotechnology*. Cambridge: Harvard University Press.
- ²⁸ S. K. Burley. An overview of Structural Genomics. *Nature* 2000; November: 932-934.
- ²⁹ W. Powell and J. Owen-Smith. Universities and the Market for Intellectual Property in the Life Sciences. *Journal of Policy Analysis and Management* 1998; 17 (2): 253- 277.
- ³⁰ A. Daar and P. Singer. 2005. Harnessing Genomics for Global Health: The Role of Higher Education. In *Creating Knowledge, Strengthening Nations: The Changing Role of Higher Education*. G. Jones, P. McCarney and M. Skolnik, eds. Toronto: University of Toronto Press: 246-264.
- ³¹ J.-M. Chandonia and S. Brenner. The Impact of Structural Genomics: Expectations and Outcomes. *Science* 2006; 20 (311): 347–351.
- ³² von Itzstein et al. Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature* 1993; 363: 418-423.
- ³³ R. Russell and D. Eggleston. New roles for structure in biology and drug discovery. *Nature Structural Biology* 2000; 7: 928-930.
- ³⁴ Rai and Eisenberg, op. cit. note 5; J. Weigelt et al. Structural genomics and drug discovery: all in the family. *Current Opinion in Chemical Biology* 2008; 12 (1): 32-39; M. Grabowski et al. Benefits of Structural Genomics for Drug Discovery Research. *Infectious Disorders - Drug Targets* 2009; 9 (5): 459-474.
- ³⁵ R. Stevens et al. Global efforts in Structural Genomics. *Science* 2001; 294 (5540): 89-92.
- ³⁶ A. Williamson. Creating a structural genomics consortium. *Nature Structural & Molecular Biology* 2000; 7: 953.
- ³⁷ Stevens et al., op. cit. note 35.
- ³⁸ Ibid; Chandonia and Brenner, op. cit. note 31.
- ³⁹ Ibid.
- ⁴⁰ NIGMS – National Institute for General Medical Sciences. 2007. Report of the Protein Structure Initiative Assessment Panel. Arlington: The National Institutes of Health.
- ⁴¹ D. Cyranoski. “Big science” protein project under fire. *Nature* 2006; 443. doi:10.1038/443382a.
- ⁴² D. Stuart et al. SPINE: Structural Proteomics in Europe - the best of both worlds. *Acta Crystallographica Section A*. 2006; doi:10.1107/S0907444906035347.
- ⁴³ S. Daenke, Y. Jones and D. Stuart. 2008. European Structural Proteomics – A perspective. In *Structural Proteomics and its Impacts on the Life Sciences*. J. L. Sussman and I. Silman, eds. Toh Tuck Link, Singapore. World Scientific Publishing: 463-504.
- ⁴⁴ K. Cottingham. The Structural Genomics Consortium makes its presence known. *Journal of Proteome Research* 2008; 7 (12): 5073.
- ⁴⁵ Chandonia and Brenner 2006, op. cit. note 31.
- ⁴⁶ R. Whitley. 2000. *The Intellectual and Social Organization of the Sciences*. New York, NY. Oxford University Press Inc.
- ⁴⁷ e.g. G. Petsko and A. Gregory. An idea whose time has gone. *Genome Biology* 2007; doi: 10.1186/gb-2007-8-6-107.
- ⁴⁸ L. Banci et al. An idea whose time has come. *Genome Biology* 2007; 8: 408; PSI – Protein Structure Initiative. 2008. Opportunities for Structural Genomics Beyond 2010: Creating Partnerships for the Future. Retrieved on December 2, 2009, from http://kb.psi-structuralgenomics.org/pdf/PSI_WhitePaper_100808.pdf; J. Weigelt. Structural genomics—Impact on biomedicine and drug discovery. *Experimental Cell Research* 2010; 316 (8): 1332-1338.
- ⁴⁹ T. Terwilliger, D. Stuart and S. Yokoyama. Lessons from Structural Genomics. *Annual Review of Biophysics* 2009; 38: 371–383.
- ⁵⁰ ISGO - International Structural Genomics Organization. 2000. Minutes, First International Structural Genomics Meeting sponsored by NIGMS and the Wellcome Trust, The Wellcome Trust Genome Campus, Hinxton Hall Conference Centre.
- ⁵¹ T. Terwilliger. Structural Genomics in North America. *Nature Structural Biology* 2000; 7: 935-939.
- ⁵² ISGO 2000, op. cit. note 50.

- ⁵³ ISGO - International Structural Genomics Organization. 2001. Second International Structural Genomics Meeting. Airlie Center, Virginia, USA, April 4-6, 2001, p. 5.
- ⁵⁴ see H. Berman. The Protein Data Bank: a historical perspective. *Acta Crystallographica Section A* 2007; doi: 10.1107/S0108767307035623.
- ⁵⁵ ISGO 2001, op. cit. note 53, p. 3-5.
- ⁵⁶ Ibid, p. 3.
- ⁵⁷ For example, the NIH opted for a much shorter period of four to six weeks from completion of a protein structure to public release for centres sponsored by the PSI, consistently with other initiatives regarding the disclosure of research. C. Driscoll. 2008. *NIH Data and Resource Sharing, Data Release and Intellectual Property Policies for Genomics Community Resource Projects*. In *Legal Framework for e-Research: Realising the Potential*. B. Fitzgerald, ed. Sydney, Australia: Sydney University Press: 223-239.
- ⁵⁸ See <http://targetdb.sbkb.org/>
- ⁵⁹ See <http://www.spine2.eu>
- ⁶⁰ See <http://www.thesgc.org>
- ⁶¹ Terwilliger et al, op. cit. note 49.
- ⁶² Daenke et al, op. cit. note 43; Weigelt, op. cit. note 48.
- ⁶³ See <http://kb.psi-structuralgenomics.org/KB/>
- ⁶⁴ Protein Data Bank; centre with > 10 structures
- ⁶⁵ Cottingham, op. cit. note 44.
- ⁶⁶ A scientific committee involving representatives from the sponsors and the academic community recommends the protein structures that should be targeted to the consortium's governing body. Target structures are subsequently assigned among the sites according to their areas of specialization, with multiple teams working under the leadership of a principal investigator, and the overall coordination of the Chief Scientist in Toronto. This arrangement aims at balancing the particular interests of the sponsors involved, with the input of academic experts in the field; See M. Perkmann. 2009. Trading off revealing and appropriating in drug discovery: the role of trusted intermediaries. Proceedings of the 2009 Academy of Management Meeting, Chicago 7-11 August, 2009.
- ⁶⁷ Of course, protein structures are a building block for much subsequent research and development. The ability to influence the direction and priorities of research into such structures is a meaningful incentive for large corporations to participate in structural genomics centres such as SGC, for what are often modest sums relative to their R&D budgets. Rhoten and Powell, op. cit. note 2.
- ⁶⁸ Rhoten and Powell, op. cit. note 2.
- ⁶⁹ Maurer 2003, op. cit. note 6; A. Rai. 2005. Open and Collaborative Research: A New Model for Biomedicine. Intellectual Property Rights. In *Frontier Industries: Software and Biotech*. R. Hahn, ed. AEI-Brookings Press;
- ⁷⁰ S. Ekins and A. Williams. When pharmaceutical companies publish large datasets: an abundance of riches or fool's gold? *Drug Discovery Today* 2010; 15 (19-20): 812-815; B. Munos. Can Open-Source Drug R&D Repower Pharmaceutical Innovation? *Clinical Pharmacology & Therapeutics* 2010; 87 (5): 534-536.
- ⁷¹ A. Marturano. When Speed Truly Matters, Openness is the Answer. *Bioethics* 2009; 23 (7): 385-393.
- ⁷² Maurer, op. cit. note 6.
- ⁷³ A. Abbott. 2001. The chaos of disciplines. Chicago, IL: The University of Chicago Press; T. Becher and P. Trowler. 2001. *Academic Tribes and Territories. Intellectual enquiry and the culture of disciplines*. Open University Press; Whitley, op. cit. note 46.
- ⁷⁴ Whitley, op. cit. note 46.