5. CBEN (Center for Biological and Environmental Nanotechnology, Rice University) (available at http://www.ruf.rice.edu/~cben/).

6. A. Dreger, When medicine goes too far in the pursuit of normality, *Bioethics Bulletin* **10** (1), (1999) (available at http://www.phen.ab.ca/materials/het/het10-01a.html).

7. G. Wolbring, Science and technology and the triple D (disease, disability, defect). In *Converging Technologies for Improving Human Performance*, ed. M. C. Roco, W. S. Bainbridge, Dordrecht, Netherlands: Kluwer (2002).

8. J. R. Lakowicz, B. Shen, Z. Gryczynski, S. D'Auria, I. Gryczynski, Intrinsic fluorescence from DNA can be enhanced by metallic particle, *Biochemical and Biophysical Research Communication* **286**, 875–879 (2001) (available at http://www.idealibrary.com).

9. J. Lee, P. Aich, Developers of a "molecular wire" win innovation award, *University of Saskatchewan News* (15 May 2002) (available at http://www.usask.ca/events/news/articles/20020515-2.html).

10. O.Harnack, W. Ford, F. Scholz, A. Yasuda, J. Wessels, Electrical conductivity of self-assembled metallic DNA nanowires, Presentation made at the Ninth Foresight Conference on Molecular Nanotechnology, 9–11 November, Santa Clara, CA (2002) (available at http://www.foresight.org/Conferences/MNT9/Abstracts/Harnack/).

11. W. F. Ogburn, *Social Change*, New York: B.W. Huebsch (1922).

THE USE OF ANALOGIES FOR INTERDISCIPLINARY RESEARCH IN THE CONVERGENCE OF NANO-, BIO-, AND INFORMATION TECHNOLOGY

Judith Klein-Seetharaman, Carnegie Mellon University

Modern research would be unthinkable without interdisciplinary approaches. For example, many of the biological disciplines provide proof by their names: biochemistry and biological chemistry, biophysics, biotechnology, biomaterials, biostatistics, bioinformatics, and computational biology all encompass aspects of biology that are influenced by at least one other scientific discipline. How do converging technologies fit into this existing, rich interdisciplinary framework of modern biological research? Technologies are based on science that has found a particular application area, where the driving force is a practical outcome. Different application areas can have the same scientific principles and approaches in common and only differ by the details of their implementation. Thus, in principle, the combination of two scientific disciplines that form the basis for two technologies in different application areas should yield the same benefit as the combination of the two technologies. While this is true in principle, in practice the convergence of technologies provides complementary benefits to existing interdisciplinary research. Technology in one application area is transferred to another application area by way of *analogy*. The use of analogies provides a link to the general public since analogies allow people to grasp concepts immediately when they can relate them to something they know. Thus, converging technologies have tremendous societal impact because they provide a direct means of communicating interdisciplinary research using analogies.

We will demonstrate this concept using our ongoing Biological Language Modeling Project, which exploits an analogy between language—familiar to everyone—and biology—familiar in detail only to experts. We expect that the same principles will hold true for other examples with more direct relevance to nanotechnology such as the manufacturing of miniature machines in analogy to biological machines.

The Language–Biology Analogy in Convergence of Information Technology and Biotechnology

The Biological Language Modeling project is a cross-disciplinary collaboration that converges Human Language Technologies with Biological Chemistry, in an effort to provide novel approaches to the mapping of biological sequences to the structure, function, and dynamics of proteins and, ultimately, of biological systems. Complex biological systems are built from cells that have differentiated to perform specialized functions. This differentiation is achieved through a complicated network of interacting biological molecules. The main action is carried out by proteins, which may be viewed as nano-sized biological machines (see below) that are composed of strings of characteristic sequences of the 20 amino acid building blocks. The sequences of the strings are encoded in their entirety in the genome. The linear strings of amino acids contain in principle all the information needed to fold a protein into a 3-D shape capable of executing its designated function. With the advent of whole-genome sequencing projects, we now have complete lists of all the protein sequences that define the complex function carried out by several organisms that have been sequenced hundreds to thousands in bacteria and tens of thousands in humans [1]. Individual proteins and functions have been studied for decades at various levels, atomic to macroscopic. Most recently, a new field has evolved, that of proteomics, which looks at all the proteins in a cell simultaneously. This multitude of data provides new opportunities for the applicability of statistical methods to yield practical answers in terms of likelihood for biological phenomena to occur.

The availability of enormous amounts of data has also transformed linguistics. The analogy between biology and language is shown in Figure 3.4. In language, instead of genome sequences, raw text stored in databases, Web sites, and libraries maps to the meaning of words, phrases, sentences, and paragraphs as compared to protein structure and function. After decoding, we can extract knowledge about a topic from the raw text. In language, success in this process has been demonstrated by the ability to retrieve, summarize, and translate text. Examples include powerful speech recognition systems, fast Web document search engines, and computer-generated sentences that are preferred by human evaluators in their grammatical accuracy and elegance over sentences that humans build naturally. The transformation of linguistics through data availability has allowed convergence of linguistics with computer science and information technology. Thus, even though a deep fundamental understanding of language is still lacking, e.g., a gene for speech has only been discovered recently, data availability has allowed us to obtain practical answers that fundamentally affect our lives.

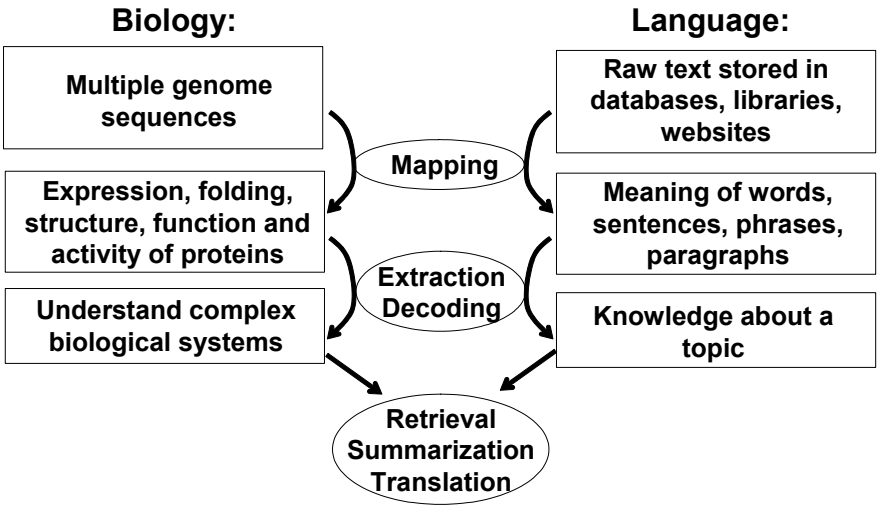

Figure 3.4: Concept of analogy between language and biology

In direct analogy, transformation of biological chemistry by data availability opens the door to convergence with computer science and information technology. This convergence between biological chemistry and computer science and information technology has already happened in the past—namely, when the disciplines of bioinformatics and computational biology emerged. In fact, human language is, just like biological chemistry—a domain of application for statistical approaches. What can the analogy to language provide that is not already provided by the previous, more general convergence of disciplines? The answer lies in the implementation details. For example, classification of protein sequences into families and subfamilies based on their functional and evolutionary relatedness is a typical problem in bioinformatics and has been studied extensively previously. Typical classification methods such as k-Nearest Neighbor approaches, Hidden Markov Models, Support Vector Machines, and Neural Networks have all been applied to this problem without reference to language technologies. At the same time, all of these methods are also used extensively in the language technologies domain. Previously, it had been concluded that it was necessary to apply classifiers of very high complexity, i.e. Support Vector Machines, to the protein classification problem to achieve high accuracy. However, application of the much simpler Naïve Bayes methods to the same problem, in combination with feature selection using chi-square that has been found to be a very successful combination in the language domain [2], has been shown to outperform the complex Support Vector Machines in this task also [3]. The reason for this success lies in the deeper analogy between language and biology that impacts methodical details, the feature selection process.

The principle for feature selection in the language domain as compared to the biology domain is illustrated in Figure 3.5. The left-hand side illustrates feature selection in language. The bags of words in three related (ball games) but different documents are shown. Some of the words identify the relatedness of the documents (“ball”), while some identify the differences (“hoop – basket”, “touchdown – tackle”, “glove – bat”), while some are irrelevant for distinguishing these texts from each other and from unrelated texts (“a – to – the”). The right-hand side of the figure shows feature selection in biology. Word equivalents used in protein sequence language are short stretches of amino acids, here shown mapped onto a secondary structure representation of a G protein coupled receptor. Only some amino acid positions (labeled) are useful in distinguishing different subtypes

of G protein coupled receptors, while the helices (center) are common to all G protein coupled receptors, and other areas cannot be distinguished from any other protein. This example demonstrates that experts in language technologies are able to augment current interdisciplinary research in bioinformatics, although the set of tools in principle is in common.

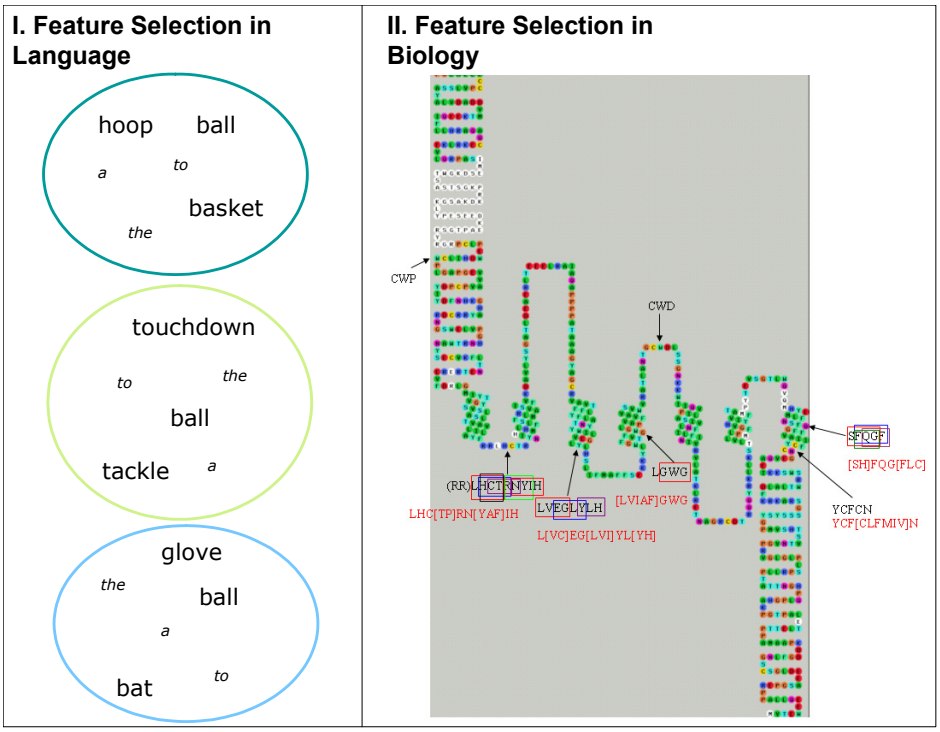

Figure 3.5. Example for the utility of details in language for biological questions.

Analogy Miniature Machines and Functional Biological Systems in Convergence of Nanotechnology and Biotechnology

Inasmuch as the application of language technologies to biological chemistry provides a novel view on old research problems, the application of other technologies may do the same. One particularly promising example is nanotechnology. The ability to control structures at a nanometer-scale level is a necessity for improving magnetic, electric, catalytic, or mechanical properties of materials for the next generation. Nanoscale cantilevers are being designed for applications in various future devices such as actuators for “smart gates” in drug-delivery systems or for data storage [4].

Methods traditionally employed in the nanotechnologies field such as Atomic Force Spectroscopy [5] and Laser Tweezers have found important applications in the protein science field, such as protein folding [6], pharmacology [7], and membrane protein structure studies [8]. However, because the field of nanotechnology is much newer than biological chemistry, there is great promise in the reverse process: the application of knowledge and principles from biological chemistry for nanotechnology. Proteins, cells, and organisms are machines of increasing complexity and size that carry out specific functions. They respond to an imposed energy field by an output that is often in the form of a different field. For example, certain organisms respond

electrochemically to chemical fields with an electrical signal; others respond with a structural and thus mechanical response to an optical perturbation. Biological systems range widely in sizes, ranging from nanoscale to microscale and can be potentially viewed as micro/nano-scaled actuators, sensors, or fuel cells for MEMS or nanoscale machines. Biological systems are usually classified by biologists according to evolutionary relatedness. We will attempt to carry out a classification of biological systems, i.e., individual proteins and protein complexes, organelles and single cells, from the eye of an engineer.

Cells and Organelles as Machines

An engineer's view of a system is based on the conversion of one field into another field, either as a result of a signal conversion or for energy requirements. For example, chemical, electrical, or optical energy and mechanical work can be used for this purpose. Organisms have been classified according to their energy source, and various nutrient requirements, patterns, or conversion abilities and sensitivities/toxicities (including carbon, phosphorus, sulfur, trace elements and nitrogen sources, essential amino acids, toxic metals). Some examples for classification based on energy source include auxotroph (chemoautotroph, photoautotroph) and heterotroph (chemoheterotroph, photoheterotroph). It is possible to extend on this classification and view a cell based on its capabilities. For example, we may view a cell as a battery or an actuator switch. This view may stimulate novel ideas to understand and interface with biological systems to make them amenable to nanotechnology.

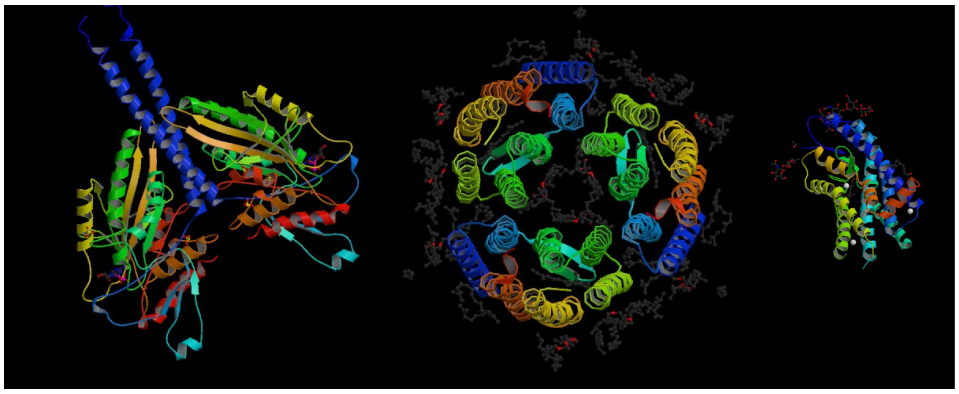

Figure 3.6: Examples of molecules representing engineering functions: motor, switch, pump (figures courtesy Protein Data Bank; also [middle figure] *Nature* [11], [right figure] *Journal of Molecular Biology* [12])

Proteins and Protein Complexes as Machines

Molecular studies of protein mechanisms have provided detailed understandings of the molecular mechanisms underlying functions in individual proteins. Three examples are shown in Figure 3.6. motor, switch and pump. On the left is an image of a kinesin microtubule motor, downloaded from the Protein Data Bank (PDB) [9] (PDB ID 2NCD [10]). In the middle is a proton pump bacteriorhodopsin (PDB ID 1C3W [11]). And on the right is a light switch rhodopsin (PDB ID 1L9H [12]). For example, the light-sensitive molecule rhodopsin converts light-energy into chemical energy by way of conformational changes in the protein [13, 14]. The nature of these conformational changes has recently been shown to be very much mechanical [15], and is therefore directly analogous to human machines. Testing such hypotheses may further increase our understanding of this molecular process, but in turn it may be possible to mimic this mechanism

invented by nature within the nanotechnology arena. While it is not possible to build proteins man-made, and even less so cells and organisms, due to the high complexity, it seems more feasible that one can use the natural systems within a nano-technology context. The details of this process are a major challenge for current and future convergence of nano- and biotechnology.

References

1. J. C. Venter, et al., The sequence of the human genome, *Science* **291**(5507), 1304-1351 (2001).
2. Y. Yang, J. T. Pedersen. A comparative study on feature selection in text categorization. In Proceedings of the Fourteenth International Conference on Machine Learning, ed. D. Fisher, 8-12 July, Nashville, TN (1997).
3. B. Cheng, J. Carbonell, J. Klein-Seetharaman, Chi-square feature selection leads to a simpler, more accurate G protein coupled receptor family classifier. *Proteins - Structure, Function and Bioinformatics*. In press.
4. Hellemans, Cantilever tales, *Science* **290**(1529) (2000).
5. G. Binnig, C. F. Quate, C. Gerber, Atomic force microscope, *Physical Review Letters* **56**(9), 930-933 (1986).
6. F. Oesterhelt, et al., Unfolding pathways of individual bacteriorhodopsins, *Science* **288**(5463), 143-146 (2000).
7. R. S. Pereira, Atomic force microscopy as a novel pharmacological tool, *Biochemical Pharmacology* **62**(8), 975-983 (2001).
8. D. Leckband, Force as a probe of membrane protein structure and function, *Current Opinion in Structural Biology* **11**(4), 433-439 (2001).
9. Research Collaboratory for Structural Bioinformatics, Protein Data Bank (available at http://www.rcsb.org/index.html).
10. E. P. Sablin, Direction determination in the minus-end-directed kinesin motor ncd, *Nature* **395**(6704), 813-816 (1998).
11. H. Luecke, et al., Structure of bacteriorhodopsin at 1.55 A resolution, *Journal of Molecular Biology* **291**(4), 899-911 (1999).
12. T. Okada, et al., Functional role of internal water molecules in rhodopsin revealed by X-ray crystallography, *Proceedings of the National Academy of Sciences USA* **99**(9), 5982-59877 (2002).
13. W. L. Hubbell, et al., Rhodopsin structure, dynamics, and activation: A perspective from crystallography, site-directed spin labeling, sulfhydryl reactivity, and disulfide cross-linking, *Advances in Protein Chemistry* **63**, 243-290 (2003).
14. J. Klein-Seetharaman, Dynamics in rhodopsin, *ChemBioChem* **3**(10), 981-986 (2002).
15. B. Isin, et al., Predisposition of the Three-Dimensional Dark State Structure of Rhodopsin for Functional Conformational Changes. In preparation.

CONVERGING TECHNOLOGIES: INNOVATION, LEGAL RISKS, AND SOCIETY

Sonia E. Miller, Converging Technologies Bar Association

> I am going to have the Society outlawed, every member removed from any responsible post. And all executive and technical positions, henceforward, can be filled only by applicants signing a non-Society oath. [1, p.268]
>
> Isaac Asimov in *I, Robot*