The next step in biology: A periodic table?

PAWAN K DHAR

Synthetic Genomics, Systems Biology Group, E209, RIKEN Genomic Sciences Centre, 1-7-22, Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

(Fax, 81-45-503-9552; Email, pkdhar@gsc.riken.jp)

Systems biology is an approach to explain the behaviour of a system in relation to its individual components. Synthetic biology uses key hierarchical and modular concepts of systems biology to engineer novel biological systems. In my opinion the next step in biology is to use molecule-to-phenotype data using these approaches and integrate them in the form a periodic table. A periodic table in biology would provide chassis to classify, systematize and compare diversity of component properties vis-a-vis system behaviour. Using periodic table it could be possible to compute higher-level interactions from component properties. This paper examines the concept of building a bio-periodic table using protein fold as the fundamental unit.

[Dhar P K 2007 The next step in biology: A periodic table?; *J.Biosci.* **32** 1005–1008]

Recently biology has seen a transition from a component-driven approach to a systems-driven approach (Dhar and Weiss 2007). This new way of doing biology has brought together sciences that mostly existed as "non-interacting" disciplines. These days it is common to use tools and concepts from computer science, mathematics and engineering to solve a biological problem. New terms like "systems biology" and "synthetic biology" have made their way into the common scientific jargon . At this unique confluence of trend and tradition, let us examine the future needs of biological community and make preparations to fulfill it.

Given the way systems and synthetic biology efforts have evolved in the recent past, it is easy to foresee an upcoming era of Computer Aided Design (CAD) of organisms. To enable CAD biology, one would need to decompose and modularize high-resolution qualitative and quantitative biological data. The non-decomposable layer would be comprised of unique elements–as much as possible. These unique elements could be tied together in the form of a relationship table or what I would call, a *periodic table*. In my opinion, a bio-periodic table could be defined as a tabular arrangement of unique biological elements that, in combinations, produce distinct higher-level properties of the system. This paper examines the possibility of using 'protein fold' as building block of a molecular periodic table. Before we delve into deeper aspects of a periodic table in biology, let us briefly examine the terms systems and synthetic biology and how these approaches can help in the construction of a bio-periodic table.

Systems biology describes a study of linked components in which the property of each part is mapped to the overall systems behaviour. At the fundamental level, biological components may be genes, as in regulatory networks, or proteins as in signalling networks, or indeed both. Going beyond cellular level, one may consider a group of interacting cells (working for a common purpose) as fundamental biological components. Subsequently, one may consider interactions at the level of organs, organisms and ecosystems as the basal layer, depending on the need and importantly, the availability of data.

Synthetic biology describes a practical application of systems biology. Synthetic biology is an approach to reverse engineer biology using the key concepts of modularity and hierarchy. The hope is to understand the rules of composition and employ these rules to create useful biological applications. Recently the term "constructive biology" has been used as a gentle and socially acceptable substitute to "synthetic biology". In engineering terms, systems biology

Keywords. Periodic table; systems biology; synthetic biology

resembles *process analysis* and synthetic biology, *systems engineering*.

Technological advancements in biology have reached a stage where living systems can be constructed from off-the-shelf chemicals (Smith *et al* 2003). In future, we will witness more such attempts to reverse engineer organisms. Biology will be increasingly viewed as a technology to fulfill human needs of energy, environment, health–to name a few. To meet such a grand vision we must ask: do we have enough knowledge of how a living machine operates? Do we know the best-case scenarios and boundary conditions of a biological system? Is it easier to create complexity than to reduce it? Is it possible to build a controllable system with both vector processes (transcriptional cascades) and scalar process (diffusion)? Probably some more questions that merit answer are: Given that biology is not absolutely dependent on precise numbers, are there any global thresholds that talk to local thresholds to drive cellular decisions? What is the role of noise in affecting physiological states in organisms? How can one distinguish random noise from a deliberate one? How much of loss-of-information allowance is adequate in abstracting and modelling biology with reasonable accuracy? Finally, what is the origin of emergence in biology?

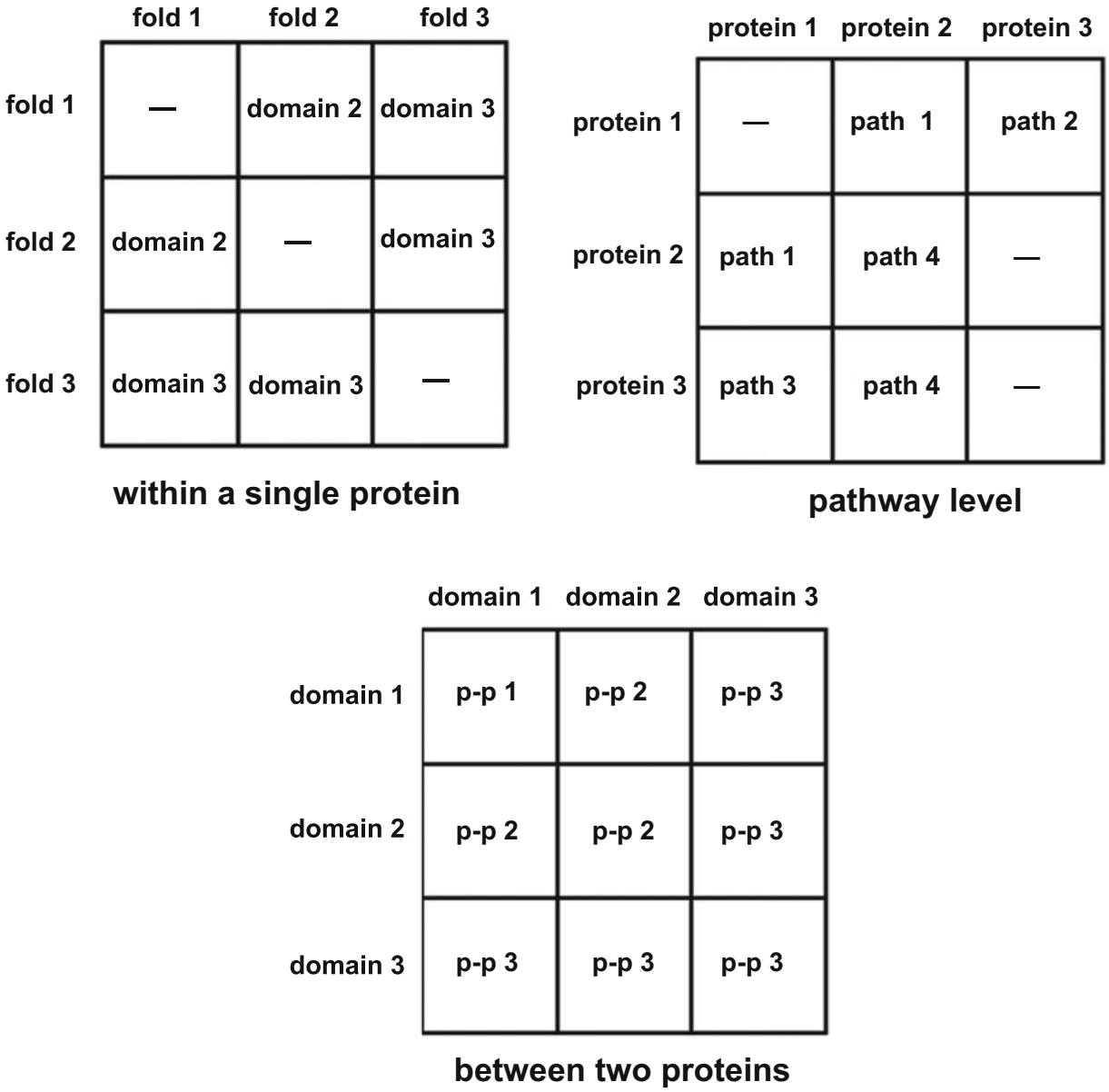

Figure 1. Function periodic table.

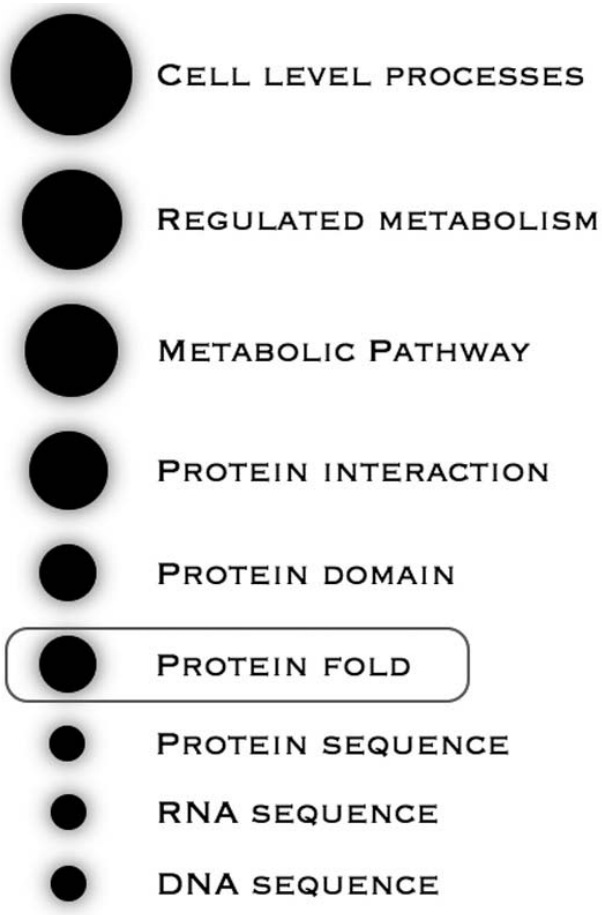

Figure 2. Building block of periodic table in biology

Some of the answers to these questions may actually be found in *old-fashioned* biology that is increasingly becoming more relevant than before. A good biology based application can be built only if the fundamental knowledge of molecular and cellular interactions is unambiguous and independent of interpretation, i.e., an observation is clearly described in terms of content and context. Thus, to bring uniformity in biology, we need to annotate each component and place them in some kind of a series that would explain their higher order interaction and the resulting behaviour of the system. In other words, it is time to build a periodic table in biology ! Though it looks like a distant goal, due to challenges in accurate (i.e. context based) data collection and interpretation, the bio-periodic table would be useful in providing well-defined modules based on biological rules of composition. By referring to the periodic table it would be possible to assemble new systems and/or edit the existing ones with reasonable ease.

In my opinion, the fundamental unit of such a periodic table should be a *protein fold*. One may ask: why not DNA sequence? There are several reasons why DNA sequence may not be the best candidate. Before we delve into that, we need to appreciate a fundamental distinction in cellular transactions: process vs regulation. A periodic table in chemistry deals with the 'process' description. Regulation of interaction among elements governed by atomic shells and sub-shells is the domain of physics. Likewise pathways and networks are processes. The role of gene regulation is to manage these processes. Due to this reason, an interacting protein unit represented by a certain fold would therefore be a suitable candidate . Another reason why DNA sequence is an unattractive candidate is its frequent non-uniqueness. We know that alternative splicing generates various forms of RNA molecules from a single gene sequence. Furthermore, there are situations where several DNA sequences contribute towards the formation of a single protein fold. Due to these reasons DNA sequence is an unattractive candidate for the position of a fundamental unit of the bio-periodic table.

Moving onwards from the fold level, a bio-periodic table would connect fold-description with the cell-level response through a series of hierarchical information transfers as shown in figure 1. Two key issues arise in this approach: the need to build '*interaction*' table and the need to build '*interaction management*' table. At the cellular level, the simplest form of interaction table (figure 2) would connect protein folds with pathways. On the other hand, 'interaction management table' would set boundary conditions to these processes by adding regulatory loops, quantitative thresholds and contextual descriptions (figure 3). Unlike the periodic table in chemistry, bio-periodic tables may have to be frequently edited in terms of processes and/or regulations. Also, unlike periodic table in chemistry, a change from one element to the next may not be uniformly incremental i.e. it may not be *periodic* in the sense of a periodic table in chemistry. The degree of change would depend completeness and accuracy of data.

Though we are still far off from that goal but once created, the benefits of building a bio-periodic table would be many: biological systems will be understood, manipulated and designed with greater ease. The bio-periodic table would connect context-based structure with topology and function. The table would allow knock-outs, knock-ins and use of conditional search statements. The bio-periodic table would be a good place to look for variables and constants in biology, build whole cell-scale models and generate hypotheses. Furthermore, such a table would help uncover molecular interaction patterns leading to identification of new rules of composition in biology.

While physicists and chemists enjoy a number of well-defined rules, principles and laws, biologists must contend with the rules of Mendelian Inheritance, which have essentially remained unchallenged for more than 100 years. One of the reasons why Mendel was successful in discovering the Laws of Inheritance, is because he dealt only with *constants* i.e. phenotypes. The phenotypes did

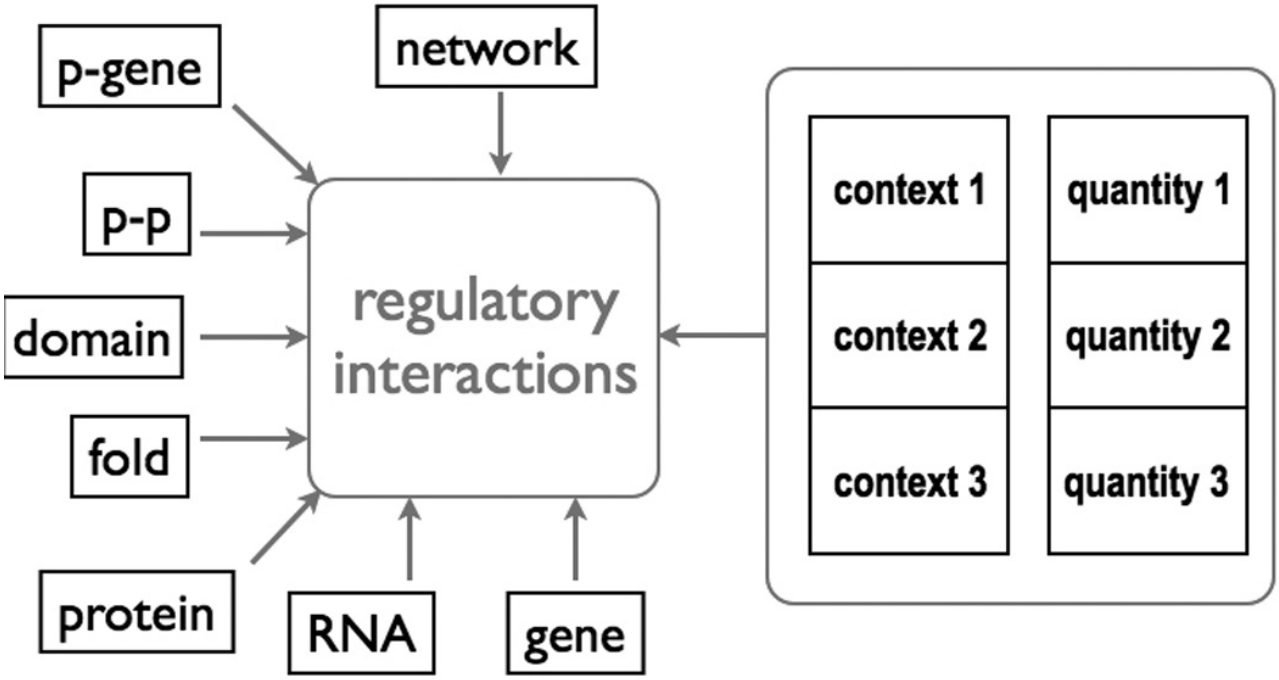

Figure 3. Regulation periodic table.

not change with ambient temperature, pressure, nutrients, pH etc. Due to strength in sample selection, Mendel only needed to use simple addition and division to discover Laws of Inheritance! Likewise, to understand biology at the molecular and network level, we need to collect constants at genomic, transcriptomic, proteomic, interactomic and metabolomic levels. By linking and statistically treating these constants it may be possible to get a better idea of how biological complexity operates.

At the molecular level, constants may be conserved sequence motifs and genetic code. At the gene expression level it may be difficult to directly visualize constants, as the expression is easily influenced by a number of intrinsic and extrinsic factors. Thus, there is a need to identify *expression constants*, *interaction-constants* and *network-constants*, to understand natural biology in its entirety (systems biology) and use it for designing new living systems (synthetic biology). Interestingly a constant concept in biology has always come rarely. Even the dogma that was considered central in biology has now morphed into a complex wiring of feedback loops. We do not yet know if the DNA is the *Director of Nuclear Affairs* or simply a *Database of Nuclear Artifacts* as the RNA camp would like us to believe. In my opinion, a bio-periodic table would help elucidate this key unsolved issue in biology.

Acknowledgement

I would like to express my warmest thanks to Dr Hao Zhu, University of Nottingham, UK and Dr.Chandra Verma, Bioinformatics Institute, Singapore for their critical comments and suggestions in improving the manuscript. My sincere thanks to Dr Yoshiyuki Sakaki, Genome Science Centre, RIKEN for supporting this work.

References

Dhar P K and Weiss R 2007. Enabling the new biology of the 21st century; *Synthetic Systems Biol.* 10.1007/s11693-006-9000-6

Smith H O, Hutchison C A 3rd, Pfannkoch C and Venter J C.2003 Generating a synthetic genome by whole genome assembly: phiX174 bacteriophage from synthetic oligonucleotides; *Proc. Natl. Acad. Sci. USA* **100** 15440–1544

*e*Publication: 6 August 2007