[Frontiers in Bioscience 10, 961-965, January 1, 2005]

MULTI-SCALE METHODOLOGY: A KEY TO DECIPHERING SYSTEMS BIOLOGY

Xinhao Ye, Ju Chu, Yinping Zhuang, Siliang Zhang

State Key Laboratory of Bioreactor Engineering, National Engineer Research Center for Biotechnology(Shanghai), Department of Biochemical Engineering, East China University of Science and Technology, 130 Meilong Road, P.O.Box 329, Shanghai 200237, China P.R

TABLE OF CONTENTS

1. ABSTRACT

Presently, it is widely accepted complex systems couldn't be comprehended by studying parts in isolation without examining integrative and emergent properties, and system-level understanding thus has become the focus in biological science. However, it should also be noted that common systematic analysis was restricted to large-scale analysis at a certain level, while the facts that the nature of complex systems is their multi-scale structures was usually neglected or ignored. Therefore, this paper described a multi-scale methodology to investigate the nature of biological complexity and prospected this methodology could lead to a promising revolution in current system-level understanding and the integration of molecular biology databases.

2. INTRODUCTION

To date when we cheer up the imminent leap forward in various *omics* science, an era of 'big science' is stealthily but undoubtedly coming into being (1, 2). That is, most of research fields in life science become multidisciplinary (3, 4), drawing information and techniques not only from biochemistry, genetics, molecular biology and cell physiology, but from chemistry, chemical engineering, systems science and computer science as well. In the meanwhile, apart from this interdisciplinary endeavor, biologist, facing biological entangling networks, also has to shift his focus from individual properties of the components to collective properties of the systems (5, 6, 7). Thus, a systematic fashion emerged and bio-X science was highly advocated to map the multi-tiered molecular networks in biological systems (8).

In fact, however, system-level approaches have already had a long history to explore the complex systems in biology (9), but traditionally, biological science tended to capture the complexity of biological systems in a simpler formalism (5, 10). It was clearly a kind of reductionist, driven by the needs of simplicity as well as the limits of technology (11). Although these classical methods had played the important roles in progress of complexity science, it was safely hard to advance the systematic analysis of biological networks more with the reductionist vision now (5, 12). Luckily, statistics and other computational methods had come to our rescue and provided several important experimental tools to cell biologists (13). What is more, coupled with advances in high-throughput measurements, recent achievements in *omics* science have grandly raised hopes for system-level understanding and prediction to cellular spatial and temporal phenomena.

Consequently, on a hand, system-level analysis received the mainstream attentions and became the core approach in biology; On the other hand, numerous of novel grandiloquent terms ceaselessly came forth, such as large-scale organization, whole-cell simulation, global gene expression profiling. However, given the avalanche of information in those areas following this fashion, biologist inevitably had been pushed into such an awkward position that new technologies were providing information at a much faster rate than his ability to digest and understand it (7, 14). Hence being puzzled and even somewhat helpless, we had to question either whether it was essential to comprehend cellular mechanism by current system-level analysis, or what the thorough and rational systematic analysis methods is could completely determine the biological complexity in the intact scales?

3. THE NATURE OF BIOLOGICAL COMPLEXITY

From the philosophical viewpoint, the objective of science is to seek the facts or nature of the complex phenomena of interest, then to make it clear in theory and to apply it to the engineering practice. Clearly, it is the same case as systems biology. In order to perform a more concise and efficient investigation on life's complex systems, as well as to fulfill the tasks of systems biology, no one could overlook the nature of biological complexity.

As a rough definition, a complex system is *a system with a large number of elements, building blocks or*

agents, capable of exchanging stimuli with one another and with their environment (7). Thus, an effective way to find the nature out is to investigate their organizing principles, including how parts were connected and how systems were operated. Fortunately in the past decades several major mechanisms had been discovered inside most systems (14, 15, 16). As the common characteristic of complex systems, it is clear that complex systems have the collective properties, which are unequal to the sum of individual properties of their component sub-systems. That means complex systems must be studied in their entireties rather than by analyzing the parts in isolation in view of some emergent or integrated properties. Moreover, complex systems have the history and are in a dynamic state, of which the present behaviors are in part determined by their past behaviors. That is system has to be analyzed to follow *the arrow of time*. But to the special properties of their own, complex systems are that they *display organization without any external organizing principle being applied* (7). Each system thus has its particular organizing principles and it can't be understood out of itself. On this footing, it is evident that rational strategies should unite these features compatibly to explore complex systems and unravel their specific features from universal ones to address their complicated temporal and spatial properties.

Although biological network is characterized as symbiotic systems (18) relying on the cooperation between the network and the involved specific elements, which is dissimilar with the normal complex-system functioning by the whole networks, it also can't cast off above organizing principles. Their respective unique characteristics are dominated by their special genotype and corresponding environmental perturbations, while the general ones are for all life systems. Recently, Oltvai and Barabasi (19) compared it as life's complexity pyramid to illustrate the relation between the particular and the universal. At the top of the pyramid, large-scale organization with scale-free structure has been uncovered as inherent properties in broad organisms, complying with a power law (20). Various complex systems, like Internet and social networks, yet strikingly share the similar hierarchical architectures with biological systems (7, 21). Further at the lower level, even within cell, metabolic networks and protein interactions still have certain similar network topologies to construct corresponding functional modules in cellular organization (22). Whereas, there is little commonness at the bottom of pyramid, where cell's genome, transcriptome, proteome and metabolome vividly exhibits organism specificity, including itself distinct inherited information and the special principles of information processing.

However, it should be noted that this from-universal-to-particular pyramid is only a transverse section, if coalesced the temporal-spatial properties and extended this pyramid to a 3D structure by XYZ axis constructed by the temporal, spatial and organizing characteristics, respectively. Nevertheless by this 3D pyramid, it is evidently shown that multi-scale structure in complex systems is the uniform nature to incorporate above mechanisms. This conclusion not only provides series of platforms to draw diagrams of their interconnections and the assembly lines how system functions, but also implies the grand challenges for systems biology to cross and unite the different horizontal and vertical levels or scales in life's solid pyramid (16).

4. WHAT IS THE SYSTEM-LEVEL UNDERSTANDING?

At the risk of sounding too direct, the heart of systems biology is to the tight coupling with experimentation, data analysis, and hypothesis generation under the multi-scaling challenges described above (25). Initially, however, such three concepts were divided into distinct science approaches in systems biology. For instance, hypothesis-driven approach stems from the extending of theories; data-driven approach lies in experiments or databases, while model-driven approach depends on computer simulation. Taken the hypothesis-driven science for granted, much of last century biology used to reduce biological phenomena to the behavior of molecules and especially attempted to infer the existence of genes and their properties from the investigation of inheritance of variation (26). It shaped so-called hypothesis-driven science. By contrast, with advanced high-throughput technologies, biologist had to change the custom over from 'a gene = a paper' research to a single paper describing the properties of the whole genome or the comprehensive protein networks (2). That signaled a huge leap towards a systematic fashion from molecule-directed fashion, as well as the shift from hypothesis-driven science to non-hypothesis-driven or data-driven science (8, 25).

Therefore, if analogizing the development in hypothesis-driven science as the accumulation of discrete particles, recent achievements by date-driven *omics* science had established numerous longitudinal sections in the solid pyramid. Given the transverse sections built by organization principles, it is evident that current 'flat' systematic fashion should extend to the integrated tridimensional spaces from various severed planes of transverse and lengthways sections around corresponding spatio-temporal-organizing scales.

Furthermore, note that although large-scale measurement from data-driven science allows the rule parameters to be determined, it is invalid to realize the comprehensive examination to the structures and dynamics from the causal links between the inputs space and the outputs space, especially if based on blind trial-and-error methods and discrete measurement. Firstly, these causal links almost belong to a kind of 'N to N' networks rather than the '1 to 1' or 'N to 1' or '1 to N' links. Without rational methodologies, it cannot but fall into cluster analysis or statistics-based theories after a mass of experiments to assure the relation between various matrix of inputs and outputs space (11, 25); secondly, considered the networks properties *high cohesion and low coupling*, the causal links between the *hubs* of levels or scales in the system perhaps is not robust and adaptable as much as them

inside(20); thirdly, due to the particular hierarchical architecture in different inputs and outputs space, there are quite different organization principles in various scales(26). As an illustration, only in input space, if the relation among inputs is governed only by cascade control, it is appropriate to confer the higher-level module by the common large-scale analysis depended on mono-scale measurement, like simple gene or protein microarrays. Otherwise, if inputs or their subsets are independent on one another, it is safely inefficient to explore life's pyramid only with a flood of the system-level measurement without multi-scale apprehension.

Hence, in order to understand and especially predict the spatial and temporal phenomena among the Micro-(DNA, RNA), Meso- (Protein, Metabolite) and Macro-(Cell, Device) scale, the extended system-level understanding on the basis of multi-scale methodology should be highlighted to meet the challenges in the next 'space' era.

5. MULTI-SCALE METHODOLOGY

In retrospect, the multi-scale methodology was firstly proposed in astronomy(27) and chemistry(28) during the 90s last century. Then it has received more and more attention and been used broadly in various basic and applied scientific disciplines (17), such as mathematics, physics, chemistry, astronomy, biology, and in applied fields such as mechanics, chemical engineering, biomedical engineering and so forth. And a novel term, *multi-scale science*, has been proposed as an independent branch in system sciences, which was considered to be a grand challenge for 21st century (29, 30, 31).

Say the biotechnology. In fact, if traced back to the development of biotechnology, lots of progress was achieved also through some kinds of multi-scale analysis. A good case is the research evolution on Crabtree effect. Crabtree (or glucose repression of the respiratory chain) effects were firstly presented in 1929. In the initial research, there was no choice but to select other non-fermentative carbon source such as glycerol, instead of the fermentative carbon substrate glucose, in order to obviate these effects (32). However from the 60s to 70s last century, *golden era of antibiotics*, glucose became possible to be utilized as carbon resource with the development of fed-batch culture technique and SSF (simultaneous saccharification and fermentation) process (33). Meanwhile continuous culture system was investigated with the introduction of the oxygen electrode at device scale to control appropriate specific growth rate and oxygen density. On the other hand, PTS system and the active transport and permease system were discovered in the end of 70s (34, 35), which agitated an important revolution from engineering scale to molecule-scale with the development of molecule biology, including the progress in signal transudation. In the end of 90s last century, hexokinase P was reported to play a major role in the early part of the glucose-repression cascade (36). Then both positive and negative control aspects were found in glucose effects at the level of transcriptional control, which were mediated by the cyclic AMP (cAMP)-cAMP receptor protein complex (for catabolite repression) and by the specific repressor (for specific repression), respectively. And subsequent studies have revealed the specific repression is controlled by IIA^{Glc}-dependent PTS-mediated inducer exclusion (37). Recently entire sequence of genome, large-scale expression measurements and protein assays yet had exhibited a more comprehensive map enable us to dissect glucose signaling more systemically (38, 39, 40).

Similarly in other aspects of life science, multiscale analysis has grandly extended the spatio-temporal-organizing range of biological network. However, it is noteworthy that multi-scale analysis is just a kind of methodology, rather than some concrete quantitative methods, like system-level analysis which only supplies us series of principles for the investigation of the multi-scale structure of complex systems. The major ones are descriptive (*describing the appearance of structures without paying attention to the mechanism of the formation of the structures and the relationship between the different scales*), correlative (*formulating the phenomena at higher scales through analyzing the interaction at lower scales*), and variational (*revealing the relationship between scales by formulating the stability condition of structure*)(17). Fortunately, today with the development of modeling network dynamics or computational biology, it provided huge opportunities to apply those principles to direct the quantitative research, especially in database integration. Numbers of model systems has been implemented to accelerate the interconnectedness of biological research (41). And gene-enzyme-endproduct control unit model became a central method to describe the genetic regulation mechanism and to optimize biosynthesis (42, 43, 44).

6. PROSPECT

Science discovery is actually just a 'compilation strategy' as described by Selinger *et al.* (11), which is not a goal but a process of continual accumulation. That is, there are never the endpoint and so-called the eventual goals in science. However, today lots of biologists regard the whole-cell simulation (45, 46) as the highest ideal. That means in the future biological systems could be deciphered with comprehensive system-level understanding and cellular behavior simulation. Obviously until now under the limit of combinatorial capability and casual links among scales, even to move upward or downward harmoniously is not an easy task. Especially considering the recent controversies between ELMO (elementary flux modes) and EXPA (extreme pathways) (47, 48), although both methods are indeed identical under most situations of interest, the discrepancies, even in a tiny extent, challenge us to balance the reversible movement between the bottom-top and the top-bottom process with multi-scale analysis. Further, given the accumulation of noise, small discrepancies between model and reality could accumulate to cause completely inaccurate predictions (9), due to so-called 'butterfly effect'. So the central issue of systems biology should alter to enhance the further communication in scale or among scales,

obviating the conflicts and distorting, while multiscale analysis is compelled to complement a series of molecular, cellular, and physiological validation methodologies to adapt the new 'multiscale- synthetic-analysis' fashion. Undoubtedly these progresses will contribute to a new avenue to tackle the overall behavior of cell in the next era.

7. ACKNOWLEDGEMENTS

The authors gratefully acknowledge support from the China's State 9th and 10th Five Year Key Program (Serial No. 96-C03-04-02 and 2001BA708B07-05), China's National Hi-Tech Program (863 program, Serial No.2002AA217021), and China's National Foundation of Natural Science (Serial No.29976013). Meanwhile, the authors are indebted to Prof. Ying Hu and Prof. Honglai Liu (Department of Chemistry, ECUST) for inspiring comments during the preparation of this article. Also, YXH would like to extend thanks to Dr. Mingzhi Huang, Dr. Meijin Guo, Dr. Haifeng Hang, Dr. Syra Zhu and Mr. Mingfei Shao for their encouragement and constructive advices.

8. REFERENCES

1. Periwal V & Z. Szallasi: Trading 'web-work' for network. *Nature Biotech* 20, 345-346 (2002)

2. Lucchini S, A. Thompson & J.C.D. Hinton: Microarrays for microbilogists. *Microbiology* 147, 1403-1414 (2001)

3. Shimizu, K: Metabolic Pathway Engineering: systems analysis methods and their application. In: Advances in Applied Biotechnology. Ed: Zhong J. J. East China University of Sci & Tech Press, Shanghai (2000)

4. Lenhoff, A.M: A Natural Interaction: chemical engineering and molecular biophysics. *AIChE J* 49, 806-812 (2003)

5. Ricard, J: What do we mean by biological complexity? *C. R. Biologies* 326, 133-140 (2003)

6. Kitano, Hiroaki: Systems Biology: a brief overview. *Science* 295, 1662-1664 (2002)

7. Ottino, J.M: Complex systems. *AIChE J* 49, 292-299 (2003)

8. Ideker, T. and D. Lauffenburger: Building with a scaffold: emerging strategies for high- to low-level cellular modeling. *Trends Biotechnol* 21, 255-262 (2003)

9. Csete, M.E. and Doyle, J.C: Reverse Engineering of Biological Complexity. *Science* 295 1664-1665 (2003)

10. Krupa, B: On the Number of Experiments Required to Find the Causal Structure of complex systems. *J Theor Biol* 219, 257-267 (2002)

11. Selinger D.W, M.A. Wright & G.M.Church: On the complete determination of biological systems. *Trends Biotechnol*. 21, 251-254 (2003)

12. Auyang, Sunny Y: Foundations of Complex System Theories: in Economics. Evolutionary Biology and Statistical Physics. Cambridge University Press (1998)

13. Mellman, I. and T. Misteli: Computational cell biology. *J Cell Biol* 161, 463-464 (2003)

14. Hatzimanikatis, V: Bioinformatics and Functional Genomics Challenges and Opportunites. *AIChE J* 46, 2340 (2000)

15. Sethna J.P, A.K. Dahmen, C.R. Myers: Crackling noise. *Nature* 410, 241-250 (2001)

16. Auger, Pierre: Dynamics and Thermodynamics in Hierarchically Organized Systems: application in Physics, Biology and Economics, Pergamon Press (1989)

17. Li J.H. & M. Kwauk: Exploring complex systems in chemical engineering: the multi-scale methodology. *Chemical Engineering Science* 58, 521-535 (2003)

18. Kitnao, Hiroaki: Computational systems biology. *Nature* 420, 206-210 (2002)

19. Oltval, Z.N. and Barabasi, A-L: Life's Complexity Pyramid. *Science* 298, 763-764 (2002)

20. Jeong, H., B. Tombor, R. Albert, Z.N. Oltvai & A.L. Barabasi: The large-scale organization of metabolic networks. *Nature* 407, 651-654 (2000)

21. Barabasi, A-L: Emergence of scaling in random networks. *Science* 286, 509-512 (1999)

22. Ravasz E, A.L. Somera, D.A. Mongru, Z.N. Oltvai & A.L. Barabási: Hierarchical organization of modularity in metabolic networks. *Science* 297, 1551-1555 (2002)

23. Aitchison, J.D and T. Galitski: Inventories to insights. *J Cell Biol* 161, 465-469 (2003)

24. Hartwell, L.H: From molecular to modular cell biology. *Nature* 402, C47-C52 (1999)

25. Brent, R: Functional genomics: learning to think about gene expression data. *Curr Biol* 9, R338-R341 (1999)

26. Hartwell, L.H: From molecular to modular cell biology. *Nature* 402, C47-C52 (1999)

27. Esaclera, E. and MacGillivray, H.T: Topology in galaxy distributions: Method for a multi-scale analysis. A use of the wavelet transform. *Astronomy and Astrophysics* 298, 1-21 (1995)

28. Gobbert M.K, T.P. Merchant, L.J. Borucki & T.S. Cale: A multiscale simulator for low pressure chemical vapor deposition. *Journal of the Electrochemical Society* 144, 11, 3945-3951 (1997)

29. Karsch, M: Multi-scale phenomena and their simulation. Proceeding of international conference. World Scientific, Singapore (1997)

30. Krumhausl, J.A: Multiscale science: Materials in the 21st century. *Materials Science Forum* 327-328, 1-8 (2000)

31. Zhang, S. and Ju Chu: Multiscale methodology for modern biochemical engineering. Chemical Engineering Press, Beijing (2003)

32. Chu, Ju and Yourong Lee: Optimal strategies for modern fermentation engineering. Chemical Engineering Press, China (2002)

33. Ingram, L.O: Metabolic Engineering of Bacteria for Ethanol Production. *Bioeng Biotech* 58, 204-214 (1998)

34. Romano A.H, J.D. Trifone & M. Brustolon: Distribution of the phosphoenolpyruvate: glucose phosphotransferase system in fermentative bacteria. *J Bacteriol* 139:93-97 (1979)

35. Simoni R.D. & P.W. Postma: The energetics of bacterial active transport. *Ann Rev Biochem* 44, 523-554 (1975)

36. Gombert A.K., M.M.D. Santos, B. Christensen & J. Nielsen: Network identification and flux quantification in the central metabolism of Saccharomyces cerevisiae under different conditions of glucose repression. *J Bacteriol* 183, 1441-1451 (2001)

37. Holtman, C. K., A.C. Pawlyk, ND. Meadow & D.W. Pettigrew: In: Reverse Genetics of Escherichia coli Glycerol Kinase Allosteric Regulation and Glucose Control of Glycerol Utilization. Ed: Vivo. *J Bacteriol* 183, 3336-3344 (2001)

38. DeRisi, J.L., V.R. Iyer & P.O. Brown: Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* 278, 680-686 (1997)

39. Kuruvilla FG, A.F. Shamji, S.M. Sternson, P.J. Hergenrother, & S.L. Schreiber: Dissecting glucose signaling with diversity-oriented synthesis and small molecule microarrays, *Nature* 416, 653-657 (2002)

40. Mitchell, Peter: A perspective on protein microarrays. *Nature Biotech* 20, 225-229 (2002)

41. Macauley J, H. Wang & N. Goodman: A model system for studying the integration of molecular biology databases. *Bioinformatics* 14, 575-582 (1998)

42. Xiu ZL, Z.Y. Chang & A.P. Zeng: Nonlinear dynamics of regulation of bacterial trp operon: model analysis of intergrated effects of repression, feedback inhibition, and attenuation. *Biotechnol Prog* 18, 686-693 (2002)

43. Ye X.H., M.J. Guo, H.F. Hang, J. Chu, S.L. Zhang: Mutli-scale methodology for biochemical engineering. CNLS 23rd Annual Conference on Networks: structure, dynamics and function, New Mexico, USA (2003)

44. Chanda S.K, S. White, A.P. Orth, R. Reisdorph, L. Miraglia, R.S. Thomas, P. Dejesus & D.E. Mason: Genome-scale functional profiling of the mammalian AP4 signaling pathway. *Proc Natl Acad Sci* USA 100, 12153-12158 (2003)

45. Tomita, Masaru: Whole-cell simulation: a grand challenge of the 21st century. *Trends Biotechnol* 19, 205-210 (2001)

46. Simpson, M.L. Whole-cell biocomputing. *Trends Biotechnol* 19, 317-323 (2001)

47. Klamt, S. and J. Stelling: Two approaches for metabolic pathway analysis?. *Trends Biotechnol* 21, 64-69 (2003)

48. Palsson BO, ND. Price & J.A. Papin: Development of network-based pathway definitions: the need to analyze real metabolic networks. *Trends Biotechnol* 21, 195-198 (2003)

Key Words: Multiscale Methodology, Systems Biology, Complex Systems, Biological Networks

Send correspondence to: Xinhao Ye., Present affiliation: Department of Microbiology, Fudan University, Room 501, the 71th building, Nong 78, WuChuan Road, Shanghai, P.R. China. 200433, Tel: +86-21-55073009, E-mail: sanger_y@hotmail.com. or Prof. Siliang Zhang, State Key Laboratory of Bioreactor Engineering, National Engineer Research Center for Biotechnology (Shanghai), Department of Biochemical Engineering, East China University of Science and Technology, 130 Meilong Road, P.O.Box 329, Shanghai 200237, China P.R, Tel:+86-21-64253021, E-mail: siliangz@163.net

http://www.bioscience.org/current/vol10.htm