行政院國家科學委員會專題研究計畫 成果報告

生物與複雜網路模組結構之解析與應用研究
研究成果報告(精簡版)

計 畫 類 別 ：個別型
計 畫 編 號 ：NSC 99-2112-M-216-001-
執 行 期 間 ：99 年 08 月 01 日至 100 年 09 月 30 日
執 行 單 位 ：中華大學生物資訊學系

計 畫 主 持 人：黃俊燕

計畫參與人員：大專生-兼任助理人員：許菁怡
大專生-兼任助理人員：戴竹儀
大專生-兼任助理人員：王韋&amp；#24312；
大專生-兼任助理人員：曾昱凱
大專生-兼任助理人員：盧柏宇

報 告 附 件 ：出席國際會議研究心得報告及發表論文

處 理 方 式 ：本計畫涉及專利或其他智慧財產權，2 年後可公開查詢

中 華 民 國 100 年 10 月 28 日

附件一

行政院國家科學委員會補助專題研究計畫 ■ 成 果 報 告 □期中進度報告

生物與複雜網路模組結構之解析與應用研究

計畫類別：■個別型計畫　□整合型計畫
計畫編號：NSC 99－2112－M－216－001－
執行期間：　99 年 08 月 01 日至 100 年 09 月 30 日

執行機構及系所：中華大學生物資訊學系

計畫主持人：黃俊燕
共同主持人：
計畫參與人員：戴竹儀、許菁怡、曾昱凱、盧柏宇、王韋廸

成果報告類型(依經費核定清單規定繳交)：■精簡報告　□完整報告

本計畫除繳交成果報告外，另須繳交以下出國心得報告：
□赴國外出差或研習心得報告
□赴大陸地區出差或研習心得報告
■出席國際學術會議心得報告
□國際合作研究計畫國外研究報告

處理方式：除列管計畫及下列情形者外，得立即公開查詢
□涉及專利或其他智慧財產權，□一年■二年後可公開查詢

中　華　民　國　100 年 10 月 28 日

附件二

Abstract

Biological systems are often organized by multi-scale functional subsystems(modules). Accurate system-level modularity organization can provide valuable information on isolated subsystem models of subcellular processes or physiological phenomena. Current methods for modularity detection are mostly optimization-based and it is difficult to trace the origin of the unsatisfactory results, which may be due to poor data, inappropriate objective function selection or simply result from natural evolution, and hence no system-level accurate modularity can be offered. Motivated by the evolution idea and using robustness and adaption as guiding principles, we propose a new approach that can identify significant multi-scale functional modules that are sufficiently accurate at the system level. The success of this evolution strategy is demonstrated by applying to the yeast protein-protein interaction network and the neuronal network of C. elegans. Several functional subsystems of important physiological phenomena can be revealed. For example, the cell cycle subcellular process in yeast can be successfully dissected into functional modules of cell cycle control, cell size check point, spindle assembly checkpoint, and DNA damage check point in G_2/M and S phases. The interconnections between these modules provide clues on the signal stimulus entries of check points into the cell cycle, which are consistent with experimental findings. For the C. elegans, biologically plausible subsystem models of sensorimotor, chemosensation and egg-laying, mechanosensation and locomotion were extracted from the whole neuronal network. Previous unknown pathways of how chemotaxis affects egg-laying rate, subtle insights into functions of neurons, and a simplified neural circuit model for thermotaxis, can be obtained from the detected modularity organization. This evolution strategy can also be applied adequately to multi-scale biological systems from mesoscopic scale, e.g cortical network in brain, to subcellular molecular networks.

Keywords:

Biological networks, modules, Robustness, evolution, community structure

摘要

本計劃提出功能關聯函數最佳化的想法建構一個偵測生物網路模組結構之計算理論，功能關聯函數最佳化方法的優點在於可偵測重疊模組結構、沒有模組大小解析極限以及100%的蛋白質遮蓋率，我們將此理論方法應用至酵母菌蛋白質交互作用網路與線蟲神經元網路上以證明其於實際生物網路之可應用性。對於酵母菌蛋白質交互作用網路，我們得到了與實驗數據吻合之蛋白質功能分佈、功能模組大小分佈之理論預測結果，進一步檢視酵母菌細胞週期生理機制，我們由所計算出的模組結構拼出一個包含細胞週期控制單元與檢查點的完整細胞週期子網路系統，並且從這個細胞週期子網路系統之模組間交互作用的關係，我們可以獲得許多細胞週期檢查點與控制單元之生物訊號傳遞模式。於線蟲神經元網路，我們提出的方法是首例可以解析線蟲神經元網路功能單元的工具，我們也因此從整個線蟲神經元網路抽取出與觸覺、化學感應與生蛋、移動、力學感應生理現象有關之神經元迴路，更進一步提出了線蟲熱感應等溫運動之簡化神經迴路模型與化學感應影響產蛋速率之神經元傳導路徑。

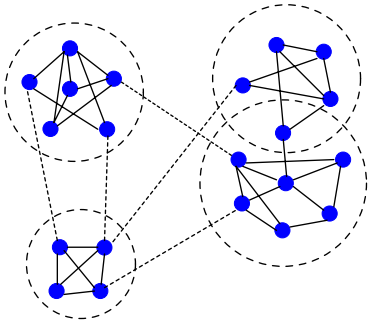
關鍵字:

生物網路、模組、穩健性、演化

研究報告內容:

I. 前言與研究目的

自從 1999 年 Barabasi 等人提出無尺度複雜網路(Scale free network)的概念後[1]，有相當多的學者開始從事複雜與生物網路之研究，並且這些網路也被證實為 scale free network[2,3]，E. Ravasz 與 A.L. Barabasi 進一步說明現實世界存在的複雜網路大多是階層式網路(hierarchical network)[4]，生物網路如蛋白質交互作用網路、新陳代謝網路也被驗證是階層式網路結構[5,6]。階層式網路有一個特點就是具有模組結構(modular structure)，如圖一所示，這些模組結構是由緊密連結之節點所組成，而模組間之交互作用相較於模組內稀疏許多，Dr. Lee Hartwell 指出細胞的功能例如訊息傳導、新陳代謝是由一些彼此有許多交互作用的分子所完成[7]，在生物網路當中這些模組結構更是具有真實生物功能的小單元，所以若能了解生物網路中具備獨立運作功能之功能模組必定對於生命運作現象的簡化模型建立與了解有很大的助益，於是有相當多的研究投入於複雜與生物網路之功能模組偵測之議題上。

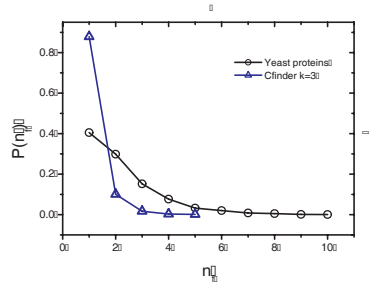


圖一. 網路模組結構示意圖。

目前要鑑定生物網路功能模組所需要克服的問題共計有，

(a) 重疊模組結構問題:

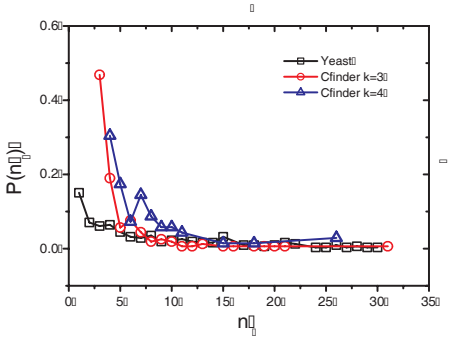
生物網路功能模組的特色是高度重疊模組結構，指一個節點可以被歸屬於多個模組結構，從圖二酵母菌蛋白質功個數分佈可以看出酵母菌蛋白質具有單一生物功能約只占 40%，具有兩個以上功能之蛋白質高達約 60%，而且一個蛋白質最多可同時擁有 13 種不同生物功能，所以生物網路是具有高度重疊模組結構之網路。



圖二. 酵母菌蛋白質功個數分佈圖。

(b) 模組大小解析度問題:

從圖三我們可以知道酵母菌功能模組大小分佈相當廣泛，小至兩三個蛋白質即可組成一個功能模組，大至約 300 個蛋白質方能組成一個功能模組，在酵母菌功能模組當中，組成蛋白質個數為 2、3、4、5 的模組佔了相當大的比例，所以一個能夠有效偵測生物網路功能模組之計算理論需要免除於模組大小解析度問題，並且其所獲得之模組大小分佈結果數據需與真實生物資料符合。



圖三. 酵母菌功能模組蛋白質個數分佈。

(c) 模組結構遮蓋率(coverage)與精確度(accuracy)問題:

好的偵測生物網路模組結構之方法必須要有 100%遮蓋率與高精確度，一個能夠有效偵測生物網路功能模組之計算理論需要兼顧遮蓋率與精確度，最好是達到遮蓋率 100%且具有相當高之精確度。有鑒於以上所提偵測生物網路模組結構將遇到的困難，本計劃提出一個基於演化思考的新方法於生物網路功能模組之計算理論，克服上述之困難，並將此方法應用於真實生物網路實例研究中。

II. 研究方法:

考慮與解決生物網路重疊模組結構，網路中的每一節點將賦予模組機率分佈函數 $P_\sigma^{(i)}$，即為節點 i 隸屬於模組 σ 之機率，如此一來我們可以把網路中的節點隸屬於多個模組事件轉換為機率事件，此外為了克服網路模組大小解析度問題，我們仿照 functional correlation 定義需考慮模組大小分布 f_σ，為了簡化計算量我們只考慮網路中的點與其有直接交互作用之點的功能關聯函數，而且採取疊代計算取代模擬退火來局部最佳化功能關聯函數，我們將此想法表示如下，

網路功能關聯函數 G 可定義為 $$G=\sum_i G^{(i)}=\sum_i\sum_\sigma G_{\sigma\sigma}^{(i)}=\sum_i\sum_\sigma P_\sigma^{(i)} g_{\sigma\sigma}^{(i)}$$

其中 $$G_{\sigma\sigma}^{(i)}=\frac{\sum_j P_\sigma^{(i)} A_{ij} P_\sigma^{(j)} \Big/ k_i}{f_\sigma} \qquad g_{\sigma\sigma}^{(i)}=\frac{\sum_j A_{ij} P_\sigma^{(j)} \Big/ k_i}{f_\sigma}$$ k_i 為節點 i 之 degree

網路模組結構計算內容與步驟條列如下，

(1) 隨機亂數產生初始模組機率分佈函數 $P_\sigma^{(i)}(0)$

計算初始模組大小分布 $f_\sigma(0)=\dfrac{\sum_i P_\sigma^{(i)}(0)}{N}$

計算初始功能關聯函數 $G_{\sigma\sigma}^{(i)}(0)=\dfrac{\sum_j P_\sigma^{(i)}(0)A_{ij}P_\sigma^{(j)}(0)\Big/k_i}{f_\sigma(0)}$

$$G(0)=\sum_i G^{(i)}(0)=\sum_i\sum_\sigma G_{\sigma\sigma}^{(i)}(0)=\sum_i\sum_\sigma P_\sigma^{(i)}(0)g_{\sigma\sigma}^{(i)}(0)$$

n=1； $\varepsilon=0.01$；

(2) 更新模組機率分佈函數 $P_\sigma^i(n)=\dfrac{f_\sigma(n-1)G_{\sigma\sigma}^{(i)}(n-1)}{\sum_\sigma f_\sigma(n-1)G_{\sigma\sigma}^{(i)}(n-1)}$

(3) 計算模組大小分布 $f_\sigma(n)=\dfrac{\sum_i P_\sigma^{(i)}(n)}{N}$

(4) 計算功能關聯函數 $G_{\sigma\sigma}^{(i)}(n)=\dfrac{\sum_j P_\sigma^{(i)}(n)A_{ij}P_\sigma^{(j)}(n)\Big/k_i}{f_\sigma(n)}$

$$G(n)=\sum_i G^{(i)}(n)=\sum_i\sum_\sigma G_{\sigma\sigma}^{(i)}(n)=\sum_i\sum_\sigma P_\sigma^{(i)}(n)g_{\sigma\sigma}^{(i)}(n)$$

如果 $\Delta G(n)=\left|G(n)-G(n-1)\right|<\varepsilon$ 則跳至步驟 5

不成立則 n=n+1；跳至步驟 2 並重複步驟 2, 3, 4 ；

(5) 若 $P_\sigma^{(i)}>\lambda$，則點 i 隸屬模組 σ，λ 為選取之門檻值。

III. 結果與討論

我們將此計算理論應用至酵母菌蛋白質交互作用網路與線蟲神經元網路，獲得了許多有用且豐富的生物知識成果，由於詳細內容繁多請參考附件。

國科會補助專題研究計畫成果報告自評表

請就研究內容與原計畫相符程度、達成預期目標情況、研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）、是否適合在學術期刊發表或申請專利、主要發現或其他有關價值等，作一綜合評估。

1. 請就研究內容與原計畫相符程度、達成預期目標情況作一綜合評估

■達成目標

□ 未達成目標（請說明，以 100 字為限）

□ 實驗失敗

□ 因故實驗中斷

□ 其他原因

說明：

2. 研究成果在學術期刊發表或申請專利等情形：

論文：□已發表 ■未發表之文稿 □撰寫中 □無

專利：□已獲得 □申請中 □無

技轉：□已技轉 □洽談中 □無

其他：（以 100 字為限）

論文草稿已撰寫完成，目前投稿於期刊審稿中。

3. 請依學術成就、技術創新、社會影響等方面，評估研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）（以500 字為限）

本計劃研究成果為第一個理論工具解決目前生物網路模組結構計算與預測時所面臨到的，大小解析度極限、模組重疊、多尺度模組組織結構之困難，應用大尺度的生物網路可以將高通量的生物實驗數據與小規模的分子生物知識連結起來，提供日後基因體蛋白質體學家解希奇實驗數據之生物意義具有強大功用，最近 2010 八月起國際間亦開始重視過去以往理論工具在模組結構之重疊性與多尺度組織特性上之困難，開始有作者於 Nature 466, 761-765, 2010.等知名期刊討論此問題，本計劃成果能率先克服此些困難歸功於提出最佳化理論於模組預測上之不適用性，反觀目前各種計算模組的理論方法與工具數量非常多，其中幾乎所有理論工具都仰賴最佳化理論，所以本計劃成果提出最佳化理論於演化生物系統之不完全適用性，相信此一觀點也持續著會影響其他生物問題之研究。

附件三

附件:

Robustness and adaptation reveals significant functional complexity of evolving biological networks at multi-scales

Jiun-Yan Huang,[1] Chi-Wei Huang,[1] Kuo-Ching Kao,[1] and Pik-Yin Lai[†2]

[1] *Department of Bioinformatics, Chung Hua University, Hsin Chu 300, Taiwan, R.O.C*

[2] *Department of Physics, Graduate Institute of Biophysics and Center for Complex Systems, National Central University, Chung-Li, Taiwan 320, R.O.C*

(Dated: October 4, 2011)

[†]Corresponding author. Email: pylai@phy.ncu.edu.tw

Biological systems are often organized by multi-scale functional subsystems(modules). Accurate system-level modularity organization can provide valuable information on isolated subsystem models of subcellular processes or physiological phenomena. Current methods for modularity detection are mostly optimization-based and it is difficult to trace the origin of the unsatisfactory results, which may be due to poor data, inappropriate objective function selection or simply result from natural evolution, and hence no system-level accurate modularity can be offered. Motivated by the evolution idea and using robustness and adaption as guiding principles, we propose a new approach that can identify significant multi-scale functional modules that are sufficiently accurate at the system level. The success of this evolution strategy is demonstrated by applying to the yeast protein-protein interaction network and the neuronal network of C. elegans. Several functional subsystems of important physiological phenomena can be revealed. For example, the cell cycle subcellular process in yeast can be successfully dissected into functional modules of cell cycle control, cell size check point, spindle assembly checkpoint, and DNA damage check point in G_2/M and S phases. The interconnections between these modules provide clues on the signal stimulus entries of check points into the cell cycle, which are consistent with experimental findings. For the C. elegans, biologically plausible subsystem models of sensorimotor, chemosensation and egg-laying, mechanosensation and locomotion were extracted from the whole neuronal network. Previous unknown pathways of how chemotaxis affects egg-laying rate, subtle insights into functions of neurons, and a simplified neural circuit model for thermotaxis, can be obtained from the detected modularity organization. This evolution strategy can also be applied adequately to multi-scale biological systems from mesoscopic scale, e.g cortical network in brain, to subcellular molecular networks.

Understanding multi-scale functional complexity of biological systems can help to unravel the secret of life. Functional complexity of biological systems are usually performed by in-

teracting components(genes, proteins, metabolites and etc.) to form the so-called biological networks. These interacting components are organized spatially and temporally into interconnected subsystems(modules) in hierarchy, which in turn cooperate to perform multi-scale cellular functions. Using high throughput techniques, the interactomes of biological systems can be realized, computational identification of consistent modules in a biological system at the system level is one of the most significant and challenging problems in systems biology. The accurate information of interconnected modules of a biological system not only can reveal the functions of the members within modules, but also can shed light on how they are orchestrated together to form cellular machines in subcellular processes.

Modularity in biological networks has been discovered to possess overlapping and hierarchical structures [1, 2, 3, 4, 5, 6]. Modules in the same or different hierarchical levels can overlap heavily with each other, i.e. possess common members that carry out multiple functions. In a hierarchically organized biological network, a large module can be decomposed into several submodules, these submodules can be further subdivided into even smaller submodules and so on(See Fig. 1a). These modules are diverse in size, organized at various scales simultaneously or dynamically to perform multi-scale functions. Theoretically, there is no strict definition for modules, and the so called "assortative mixing" rule is widely accepted in detecting modules[7]. Modules are defined as clusters of densely intra-connected nodes with sparse links between them. Assortative mixing is a popular definition for modules, but there are alternative definition of modules using link clustering[8, 9].

In an evolving network, modular structures may facilitate evolutionary changes and are governed by robustness and adaptation[10]. For survival and adaptability, modularity organizations must be robust against environmental and genetic perturbations, but at the same time evolvable[11, 12]. To compromise between robustness and evolvability in biological systems, the generation of a variety of non-lethal phenotypes and genetic buffering produces many evolved modularity organizations. The core function embedded in a module is robust against changes, and adaptation would drive modularity to be even more robust. On the other hand the phenotypes are allowed to be changed by altering interactions between the

modules. However, due to limited resources that can be used to maintain robustness in a bio-system, modularity cannot evolve to be extremely robust against some perturbations or the system will be fragile against other unanticipated perturbations, which is harmful for survival[13, 14, 15]. Hence, modularity in an evolving biological network possesses conserved parts that were adapted optimally, and some non-adapted parts that deviated from optimality[16].

Most of the present modular detection theories are optimization-based methods, the overlapping and hierarchical characteristics of the modules make the detection of true system-level modular structures unsatisfactory[5, 17, 18, 19, 20, 21, 22] and suffered from the drawback of size resolution limit[23, 24]. The designed objective function could only take partial set of perturbations into account and the optimal modularity may even be undesirable for survival upon unanticipated perturbations, and thus fail to give the system-level plausible modularity. One can never know where the unsatisfactory modules originated from, it can be due to the optimization scheme, the selection of objective function, or evolution by nature (See Online supporting materials). Without an accurate system-level modularity, it is difficult to understand how the modules cooperated to build the cellular machines. Hence, accurate identification of modularity from interactomes at the system level is important and can provide powerful insight to unravel the pathways and signals stimuli between subcellular processes, however optimization is not a good scheme to achieve this goal.

Despite cells are open systems, the modularity organizations of adapted robustness trade-offs among robustness, fragility, limited resources and others, i.e robustness strength located within the adapted region that is not too high or not too low (as illustrated in Fig. 1b), possess an essential backbone. The functionally significant membership of these realistic modules can be revealed by the evolutionary conservative profiles of modularity in adapted robustness trade-offs with given interactomes. Evolutionary conservation rate of members in a module tells us their functional significance and reliability. Hence, investigating these adapted modularity organizations can offer us more biologically plausible system-level modularity than just by using optimazation. Robustness and adaptation can serve as fundamental

guiding principles to uncover biologically plausible modularity organizations at the system level. In this paper, this evolution strategy is applied to two real biological networks, the yeast protein-protein interaction(PPI) and C. elegans neuronal networks, to demonstrate its applicability on uncovering biological significant modules and revealing important biological processes from subcellular to cellular scales.

In our evolution strategy approach, modular classification for each node is described by several functional probability components P_σ if the node has some probability to be classified into the modules labelled by σ. A node becomes a member of the σ module if its P_σ component is larger than a chosen threshold. A node with several probability components larger than the threshold is assigned to multiple modules, and hence overlapped modular structures can be produced. To apply the evolution strategy to uncover the modular structure, a robustness function is used to evaluate and select modularity organizations of adapted robustness trade-offs. The correlation of a node with the σ-module is measured by the quantity $G_{\sigma\sigma}$ which is defined as the ratio of the percentage of directly interacting nodes that are in the same common module to the percentage of nodes belonged to the same module in the whole network(See Methods)[25]. The robustness function for modularity, $R_M(P_\sigma)$, is used to evaluate how robust the modularity(P_σ) is, and is defined as the sum of $G_{\sigma\sigma}$ over all nodes and modules. Presumably, a network would be highly efficient and more robust against perturbations if common module nodes tend to aggregate together in the network, i.e have higher value of robustness function $R_M(P_\sigma)$. A real network organized its modularity structures (defined by the functional probabilities P_σ for each node) toward the higher value of robustness function if such a network had been evolved and adapted for a sufficiently long time[25, 26, 16]. Therefore, the functional probability P_σ is hypothesized to be proportional to $G_{\sigma\sigma}$, i.e nodes within a module tends to evolve to interact directly.

In nature, real modularity is composed of conserved core of modules plus some modules that deviate strongly from optimality. In our theory, functionally significant modules are captured by conserved components within modules by examining modularity variants in adapted robustness trade-offs. Newly evolved and highly variational modules are difficult

to be detected accurately since they may not have adapted to be robust. Nodes associated with modules at various hierarchical levels possess different strengths of functional probabilities. The hierarchical organization is inferred by first identifying the so-called co-clustered groups which correspond to strongly cohesive modules at the lower level. These are nodes with high functional probabilities to be frequently classified into the same modules for adapted robust modularity realizations. Then these co-clustered groups associated to form more complex hierarchical modular structure. Fig. 1a illustrates how the overlapped modules can be detected at multi-scales. Module A has classification component σ_0, i.e. all members in modules A have their community probabilities in the σ_0 component higher than the threshold(See Methods for more details). In this example, module A contains three sub-modules A.1, A.2, A.3 with classification components $\sigma_1, \sigma_2, \sigma_3$ respectively. For instance, nodes belonging to submodule A.3 have two components of community probability, σ_0, σ_3 that are higher than the threshold. In general, nodes within each submodules are stronger functionally correlated than other nodes in module A, for example for nodes in A.1 the community probability in components σ_1 is larger than that of their σ_0 component. Hence, modules were detected from higher to lower hierarchies as the threshold is being varied from low to high values. Although our theory appears to be different from the "assortative mixing" rule, the assortative mixing rule can be shown to arise from the robustness and adaptation of modularity organization, but our method is free from the size resolution limit (See Online supporting materials).

First, we apply the evolution strategy to the yeast protein-protein interaction(PPI). In the **yeast PPI network**, embedding particular functions in detected modules is investigated by the *function accuracy* of modules, which is defined as the highest percentage of nodes within a module that have the same function annotation from experimental data. As shown in Fig. 2a, there are over 70% of modules with function accuracy higher than 0.5. It indicates that members of most detected modules have the same annotated functions. The k-clique percolation method has been proposed that could detect overlapped modules[5] and its results for the yeast PPI network are also shown in Fig. 2a $\sim$ 2d for comparison. Fig.

2a shows that the function accuracy for modules detected by k-clique percolation is higher than those by the evolution strategy. However, members of a module may participate in the activity of a biological process that is accomplished by various function proteins (as illustrated in Fig. 3 and discussed below). Those low function accuracy modules often participate in biological processes with hybrid function components. The results of k-clique percolation appears to have a higher function accuracy only because it detects the strongly connected parts of a network, but not the true modular structures. Fig. 2b and 2c show that our detected modular structures are consistent with real protein annotation data in module size distribution and number of function distribution for proteins. Such agreement reveals that both overlapping and hierarchical organizations of modularity are correctly captured. In Fig. 2b, the percentage of small size detected modules appears to be a little higher than the real data, but the agreement with the real data greatly improved when the unknown function modules, whose members are mostly unknown function proteins, are neglected in the statistics. Such unknown function proteins are clustered in modules, network-based bioinformatic approaches are difficult to infer accurate functions to these proteins[27].

Fig. 3a and 3b show two low function accuracy modules corresponding to the well-known cell cycle control and spindle assembly check point respectively. In Fig. 3a, it is a dynamically regulated module which controls the progress of cell cycle process. The cyclin dependent kinase(CDK) CDC28 sequentially binds and phosphorylates the cell cycle re-entry cyclin CLN3, G_1/S specific cyclins CLN1,2, S phase cyclin CLB5, G_2/M transition cyclins CLB1-4, to control the progress of cell division from $G_1 \rightarrow S \rightarrow G_2 \rightarrow M$. CDC28 binds to CLN3 to trigger the cell cycle process, then binds to CLN1, CLN2 promoting the cell to bud. After budding, CDC28 phosphorylates CLB5 to begin DNA replication, and mitosis follows the binding of CDC28 to CLB1-4. Finally, the two significant inhibitors SIC1 and CDH1 help the cell to return to G_1 state, to complete one cycle of cell division. Although, the cell cycle controlling module is biologically plausible, but only 29.6%(8/27) of proteins possessing the main function of budding in this module. Our result suggests that the unknown function protein YPR174C in the cell cycle control module which is localized

at the nuclear periphery, it probably participates in the cell cycle controlling mechanism. Fig. 3b shows the detected spindle check point module with function accuracy 0.455(5/11). The biological signal was propagated from mitotic arrest deficient proteins MAD1, MAD2, MAD3 to CDC20 and PDS1, then transmitted into cell cycle control module to accomplish the function of spindle assembly checkpoint[28, 29].

We select the cell cycle to demonstrate how important accurate system-wide consistent information of modules is for building the cellular machine of biological processes, especially the signal transmission between functional subsystems. The complete cell cycle machine can be dissected as a combination of several detected modules, including one controlling module, cell size check point, DNA damage check point in G_2/M and S phases, and spindle assembly checkpoint (see Fig. 3c). DNA damage response experiment revealed that DNA damage-induced DDC1 phosphorylation requires RAD24 protein, but RAD9 is not required for DDC1 phosphorylation, supporting the notion that RAD9 and RAD24 act in different pathways in DNA damage response[30]. The fact that RAD9 and RAD24 are located in two different detected modules related to DNA damage further supports such an experimental inference. It is worth-noting that our approach detected that the protein YDJ1 in the cell size check point module interacts with cyclin CLN3 to trigger cell cycle entry. Such a cell cycle triggering mechanism is consistent with the recent experimental discovery that a growth-associated chaperone YDJ1 releases CLN3 from endoplasmic recticulum to enter the nucleus and trigger the cell cycle event[31]. Although multi time scales are involved in the cell cycle process, our results demonstrate that only by considering protein interaction with typical subsecond time scale can still offer deep biological function knowledge about cell cycle if accurate and system-wide consistent information on functional subsystems(modules) was obtained.

The other application is on the **neuronal network in C. elegans**, it is the simplest brain connectome with only 281 neurons. In a neuronal network, the synapse network topology and bursting frequency of neurons are two important parameters to determine its physiological functions. Here the effect of bursting frequency is not considered, the

connections of chemical synapses and gap junctions are all treated as undirected edges, the excitatory or inhibitory nature of the synapses are also ignored, only the network structure can already reveal significant bioloigcal infomation. The C. elegans neuronal network is densely connected, its hierarchical and overlapping modularity organization is difficult to solve and there is still no any satisfactory result up to now. For example, the k-clique percolation method results in a single module for the whole neuronal network. Even the methods of optimization of modularity could only detect four large disjoint modules, which fails to understand their physiological functions[23]. Our theory detected 13 and 24 modules at thresholds of 0.1 and 0.3 respectively, and the modules are heavily overlapped, which means that each neuron is responsible for several functions. Fig. 4a, b show that common module neurons in C. elegans are mostly distributed widely in the worm body, unlike the mammalian network in which the same function neurons aggregate into cortical areas. The neurons within a detected module are often of multi neuron types and distributed to several ganglia (See Online Supporting materials Fig. S4.). It indicated that neurons in C. elegans such primitive animals are responsible for more functions than advanced mammals.

Fig. 4c $\sim$ f show several combinations of modules at a threshold value of 0.3, all of them correspond to specific physiological functions with some of them observed in experiments. Fig. 4c shows the detected modules can be identified as the sensorimotor modules. The module enclosed by red lines has sensory functions, e.g thermosensation, chemosensation, olfaction, etc. Most neurons in this module are amphids, and amphid interneurons. The other module enclosed by blue lines is the sensory/motor module composed of sensory neurons, ring interneurons and ring motor neurons. Its function is for transmitting the signals detected by sensory module to nerve ring motor neurons and then to innervate muscles. The thermotaxis, chemotaxis, olfactory behaviors of C. elegans are controlled by the sensorimotor modules.

Thermotaxis is an interesting phenomenon in C. elegans in which the worm can track along the isotherm of previously adapted cultivation temperature. The modules enclosed by red and blue lines in Fig. 4c are detected modules corresponding to thermosensation

and sensorimotor behaviors. The hybrid modules control the isothermal motion of C. elegans[32, 36, 33]. The detected hybrid modules can be further reduced to a simplified neural circuit model of thermotaxis (Fig. 4d) by selecting key neurons with severe effects on thermotactic phenotypes and motor neurons [32, 33]. Previous experiments revealed that AFD, AWC are the primary and secondary temperature sensory neurons, AIY is responsible for thermophilic motion and AIZ controls cryophilic motion. RIA integrates thermophilic and cryophilic signals from AIY and AIZ to motor output[32, 33, 34, 35]. The thermal sensory neurons in the simplified thermotaxis model are AFD(1,4), AWC(1,2), ASE(1,∞), ASG(1,1), ASI(1,1), ASH(∞,1), where the first and second number in the parentheses are the shortest path lengths from the sensory neuron to AIY and AIZ respectively, with ∞ indicates unreachable. The smaller distance to AIY/AIZ indicates the stronger correlation with thermophylic/cryophilic motion. Hence, our results indicate that AFD, AWC, ASE, ASG, and ASI are the key driving neurons to trigger thermophilic pathway (AFD and AWC that can trigger thermophilic pathway was previously observed in experiments[32, 33, 36]), whereas ASG, ASI, and ASH are the key neurons to trigger cryophilic pathway.

Fig. 4e shows the hybrid modules corresponding to the physiological function of how chemosensation correlates with egg-laying in C. elegans. Neurons enclosed by solid and dashed lines are members of modules at threshold value 0.3 and 0.4 respectively. The hybrid module is composed of three modules with the functions of chemosensation and egg-laying(green line), tail motion(orange), and ventral motion(purple). It has been reported that chemosensation could affect egg-laying rate in C. elegans[37], but little was known about how the signals are being conveyed from chemoreceptors to egg-laying motorneurons. The newly detected module enclosed by the green dashed lines in Fig. 4e suggests possible pathways on how the signals are transmitted from amphids, via the nerve ring, then to ventral cord and arrived at vulval motorneurons HSNL, HSNR, VC5, and VC4 in blue shadowed region 2. It was known that mechanical stimulation such as vibration of culture medium dish inhibits the egg-laying rate[37], this effect can be understood from mechanical sensory neurons PVM, AVM, PLMR, PLML, ALML in light blue shadow region 1 in our

detected chemosensation/egg-laying module. Our modular detection results can have more subtle explanation on the functions of neurons. For example, the six hermaphrodite-specific neurons VC1-6 in the ventral nerve cord can be subdivided into two groups[37]: vulval-proximal VC4, VC5 and vulval-distal VC1-3, VC6 within blue shadowed region 3. VC4 and VC5 have direct synaptic output to vulval muscles, but VC1-3 and VC6 make fewer neuromuscular junctions with vulval muscles, but have more junctions with ventral muscles. In our modular detection results, VC5 belongs to chemosensation and egg-laying module, VC1-3 and VC6 are in ventral motion module. The VC4 neuron is located in both modules. This result is consistent from the anatomy experimental findings.

The hybrid modules in Fig. 4f correspond to mechanosensation. C. elegans lives in dirt and eats bacteria, mechanical sensation is important for C. elegans to detect soil particles and help to find bacterial food sources. Hermaphrodite has 30 mechanoreceptor neurons(MRNs) that might be used to detect mechanical stimulation[39]. In the hybrid modules, the module enclosed by red lines is for the function of body mechanosensation. It is composed of MRNs and motor neurons in ventral and tail. The module enclosed by green lines corresponds to head mechanosensation. Most members in this module are MRNs and motor neurons in nerve ring. Neurons in the light blue shadow are the complete 30 MRNs in hermaphrodite. Fig. 4g illustrates the locomotion hybrid modules, these 3 modules correspond to head, ventral and tail motion respectively. Other neurons direct the synaptic outputs to these three modules to perform locomotion in C. elegans.

To summarize, our proposed evolution strategy can infer biologically plausible functional subsystems(modules) of a biological system at the system level. With the aid of accurate multi-scale modules, subsystems for specific physiological phenomena can be easily extract by combining relevant modules. In addition to network interactome data, integrating multi-sources high throughput data consistently at the system level to infer real functional sub-systems can accelerate the understanding in biology.

METHODS

The adjacency matrix A is used to describe the topology of a network with $A_{ij} = 1$ if node j interacts with i, and otherwise $A_{ij} = 0$. Each node i is assigned with a functional probability $P_\sigma^{(i)}$ which is the weight for node i to be classified into the σ module. The correlation of a node i with the σ-module is measured by the quantity $G_{\sigma\sigma}^{(i)}$ defined as [25]

$$G_{\sigma\sigma}^{(i)} = P_\sigma^{(i)} \frac{n_\sigma^{(i)}}{f_\sigma} = \frac{\sum_j P_\sigma^{(i)} A_{ij} P_\sigma^{(j)}/k_i}{f_\sigma} \tag{1}$$

where $n_\sigma^{(i)}$ is the percentage of nearest interacting neighbors of node i that belong to module σ, k_i is the degree of node i, $f_\sigma = \frac{\sum_i P_\sigma^{(i)}}{N}$ is the percentage of nodes belong to the σ module in the whole network of a total number of N nodes. The robustness function R_M, defined as the sum of $G_{\sigma\sigma}^{(i)}$ over all nodes and modules, is used to measure the robustness of the modularity organization,

$$R_M\left(P_\sigma^{(1)}, P_\sigma^{(2)}, \ldots, P_\sigma^{(N)}\right) = \sum_i \sum_\sigma G_{\sigma\sigma}^{(i)} = \sum_{i,j} \sum_\sigma \frac{P_\sigma^{(i)} A_{ij} P_\sigma^{(j)}/k_i}{f_\sigma}. \tag{2}$$

Thousands of modularity variants with adequate values of robustness function are sampled with different initial conditions. To sample the adapted modularity organizations, we start from a randomly generated initial modularity organization, i.e. assigned arbitrary functional probabilities P_σ to each node in the network. The robustness function R_M can be calculated using this initial modularity organization. The functional probability is hypothesized to be proportional to $G_{\sigma\sigma}^{(i)}$,

$$P_\sigma^{(i)} = \frac{f_\sigma G_{\sigma\sigma}^{(i)}}{\sum_\sigma f_\sigma G_{\sigma\sigma}^{(i)}}. \tag{3}$$

$P_\sigma^{(i)}$ would simply be proportional to f_σ if the modular structure is independent of the network topology. The influence of modular structure by the network topology is taken into account through the hypothesis of $P_\sigma^{(i)} \propto G_{\sigma\sigma}^{(i)}$. The functional probabilities are renewed by Eq. (3) after the robustness function was calculated from initial functional probabilities for each node. This process is repeated iteratively until the robustness function R_M is locally

optimized. If the robustness function falls within the adapted region as shwon schematically in Fig. 1b, then the corresponding functional probabilities are then used for figuring out the memberships of modules. A node with functional probability $P_\sigma > \lambda$ is classified to the σ module, where λ is the chosen threshold value.

In order to investigate the evolutionary conservation rate of these sampled modularity in adapted robustness trade-offs, we first analyze the classification results for the obtained adapted modularity realizations at threshold 0.1 and identify the so-called co-clustered groups. The constituents of a co-clustered group are nodes with high functional probabilities to be frequently classified into the same modules, here we set the threshold for co-clustered groups as 0.7, these co-clustered groups constitute the functional modules in the lower level. The *classification component for a co-clustered group* is defined as the component of functional probability used for figuring out the members of the corresponding co-clustered group. Suppose that the classification components corresponding to each co-clustered groups are $\sigma_1,\ \sigma_2, \ldots,\ \sigma_m$, the community probability $\Pi^{(i)}(l)$ of the node i at the l^{th} trial was defined as the collection of functional probabilities for the node in classification components corresponding to each co-clustered group, $\Pi^{(i)}(l) = \left(P^{(i)}_{\sigma_1},\ P^{(i)}_{\sigma_2}, \ldots,\ P^{(i)}_{\sigma_m}\right)$. With a total of L trials, the community probability is then averaged over all sampled suboptimal modularity organizations, $\langle \Pi^{(i)} \rangle = \frac{1}{L} \sum_{l=1}^{L} \Pi^{(i)}(l)$. Finally, this resultant average community probability is used for figuring out members of each detected modules, i.e. nodes are assigned to modules σ if the σ-component of $\langle \Pi^{(i)} \rangle > \lambda$. The modular classification result is independent of the number of modules for the initial modularity organization as long as it is taken to be sufficiently large.

Acknowledgements. Supports by National Science Council of Taiwan under grant nos. 98-2112-M-008-023-MY3, 99-2112-M-216-001 and NCTS of Taiwan are acknowledged.

REFERENCES

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

[2] Ravasz, E., & Barabási, A.L. Hierarchical organization in complex networks. *Phys. Rev. E* **67**, 026112 (2003).

[3] Yook, S.H., Oltvai, Z.N., & Barabási, A.L. Functional and topological characterization of protein interaction networks. *Proteomics* **4**, 928-942 (2004).

[4] Sales-Pardo, M., Guimera, R., Moreira, A.A., & Amaral, L.A.N. Extracting the hierarchical organization of complex systems. *Proc. Nat. Acad. Sci. USA* **104**, 15224-15229 (2007).

[5] Palla, G., Derényi, I., Farkas, I., & Vicsek, T. Uncovering the overlapping community structure of complex networks in nature and society. *Nature* **435**, 814-818 (2005).

[6] Clauset, A., Moore, C., & Newman, M.E.J. Hierarchical structure and the prediction of missing links in networks. *Nature* **453**, 98-101 (2008).

[7] Newman, M.E.J. Assortative mixing in networks. *Phys. Rev. Lett.* **89**, 208701 (2002).

[8] Ahn, Y. Y., Bagrow, J. P., & Lehmann, S. Link communities reveal multiscale complexity in networks. *Nature* **466**, 761-765 (2010).

[9] Evans, T. S. & Lambiotte, R. Line graphs, link partitions and overlapping communities. *Phys. Rev. E* **80**, 016105 (2009).

[10] Hartwell, L.H., Hopfield, J.J., Leibler, S., & Murray A.W. From molecular to modular cell biology. *Science* **402**, c47-c52 (1999).

[11] Gerhart, J., & Kirschner, M. in *Cells, Embryos, and Evolution: Toward a Cellular and Developmental Understanding of Phenotypic Variation and Evolutionary Adaptability* (Blackwell Science, Malden, MA) (1997).

[12] Kirschner, M., & Gerhart, J. Evolvability. *Proc. Nat. Acad. Sci. USA* **95**, 8420-8427 (1998).

[13] Kitano, H. Towards a theory of biological robustness. *Mol. Syst. Biol.* **3**:137 (2007).

[14] Carlson, J.M., & Doyle, J. HOT Robustness and design in complex systems. *Phys. Rev. Lett.* **84**, 2529-2532 (2000).

[15] Carlson, J.M., & Doyle, J. Complexity and robustness. *Proc. Nat. Acad. Sci. USA* **99 suppl1**, 2538-2545 (2002).

[16] Pe′rez-Escuderoa, A., Marta Rivera-Albaa, A., de Polaviejaa, G. Structure of deviations from optimality in biological systems. *Proc. Nat. Acad. Sci. USA* **106**, 20544-20549 (2009).

[17] Newman, M.E.J., & Girvan, M. Finding and evaluating community structure in networks. *Phys. Rev. E* **69**, 026113 (2004).

[18] Duch, J., & Arenas, A. Community detection in complex networks using extremal optimization. *Phys. Rev. E* **72**, 027104 (2005).

[19] Li, Z., Zhang, S., Wang, R.S., Zhang, X.S., & Chen, L. Quantitative function for community detection. *Phys. Rev. E* **77** , 036109 (2008).

[20] Lancichinetti, A., Fortunato, S., & Kertesz, J. Detecting the overlapping and hierarchical community structure in complex networks. *New J. Phys.* **11**, 033015 (2009).

[21] Reichardt, J., & Bornholdt, S. Detecting fuzzy community structures in complex networks with a Potts model. *Phys. Rev. Lett.* **93**, 218701 (2004).

[22] Girvan, M., & Newman, M.E.J. Community structure in social and biological networks. *Proc. Nat. Acad. Sci. USA* **99**, 7821-7826 (2002).

[23] Fortunato, S., & Barthelemy, M. Resolution limit in community detection. *Proc. Nat. Acad. Sci. USA* **104**, 36-41 (2007).

[24] Kumpula, J.M., Saramaki, J., Kaski, K., & Kertesz, J. Resolution limit in complex network community detection with Potts model approach. *Eur. Phys. J. B* **56**, 41-45 (2007).

[25] Huang, J.Y. Tomography of functional organization in protein-protein interaction network. *Physica A* **388**, 2072-2080 (2009).

[26] Kao, K.C., & Huang, J.Y. Accurate and fast computational method for identifying protein function using protein-protein interaction data. *Mol. BioSyst.* **6**, 830-839 (2010).

[27] Sharan, R., Ulitsky, I., & Shamir, R. Network-based prediction of protein function. *Mol. Syst. Biol.* **3**:88 (2006).

[28] Hoyt, M.A. A new view of the spindle checkpoint. *J. Cell Biol.* **154**, 909-912 (2001).

[29] Ibrahim, B., Diekmann, S., Schmitt, E., & Dittrich, P. In-Silico modeling of the mitotic spindle assembly checkpoint. *PLoS ONE* **3**:e1555 (2008).

[30] Longhese, M.P., Foiani, M., Muzi-Falconi, M., Lucchini, G., & Plevani, P. DNA damage checkpoint in budding yeast. *The EMBO Journal* **17**, 5525-5528 (1998).

[31] Vergés, E., Colomina, N., Gari, E., Gallego, C., & Aldea, M. Cyclin Cln3 is retained at the ER and released by the J chaperone Ydj1 in late G1 to trigger cell cycle entry. *Mol. Cell* **26**, 649-662 (2007).

[32] Mori, I., & Ohshima, Y. Neural regulation of thermotaxis in Caenorhabditis elegans. *Nature* **376**, 344-348 (1996).

[33] Kuhara, A., et al. Temperature Sensing by an Olfactory Neuron in a Circuit Controlling Behavior of C. elegans. *Science* **320**, 803-807 (2008).

[34] Kimura, K.D., Miyawaki, A., Matsumoto, K., & Mori, I. The C. elegans Thermosensory Neuron AFD Responds to Warming. *Curr. Biol.* **14**, 1291-1295 (2004).

[35] Biron, D., Wasserman, S., Aravinthan, J.H.T., Samuel, D.T., & Sengupta, P. An olfactory neuron responds stochastically to temperature and modulates Caenorhabditis elegans thermotactic behavior. *Proc. Nat. Acad. Sci. USA* **105**, 11002-11007 (2008).

[36] Mori, I., Sasakura, H., & Kuhara, A. Worm thermotaxis a model system for analyzing thermosensation and neural plasticity. *Curr. Opin. Neurobiol.* **17**, 712-719 (2007).

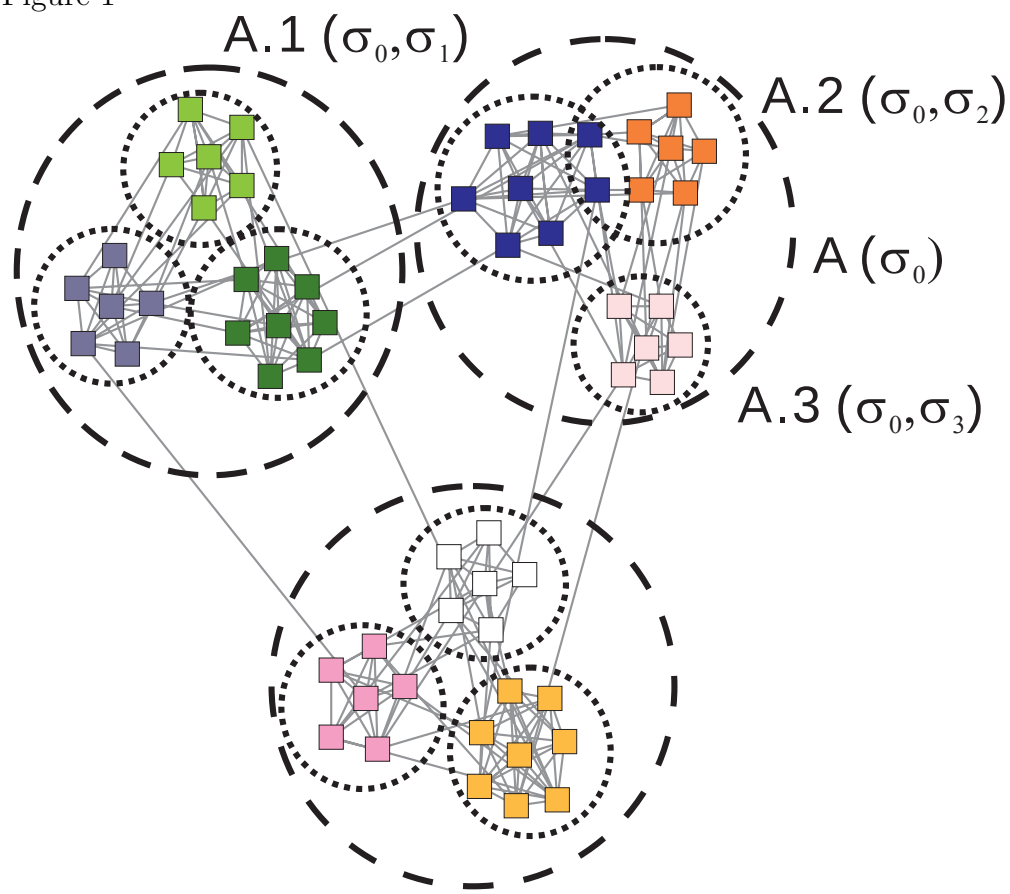
[37] Schafer, W.R. in *Egg-laying*, WormBook, ed. The C. elegans Research Community, WormBook, doi/10.1895/wormbook.1.38.1, http://www.wormbook.org/ (2005).

[38] White, J.G., Southgate, E., Thomson, J.N., & Brenner, S., FRS (eds) The structure of the nervous system of the Nematode Caenorhabditis Elegans. *Phil. Trans. R Soc. B* **314**, 1-340 (1986).

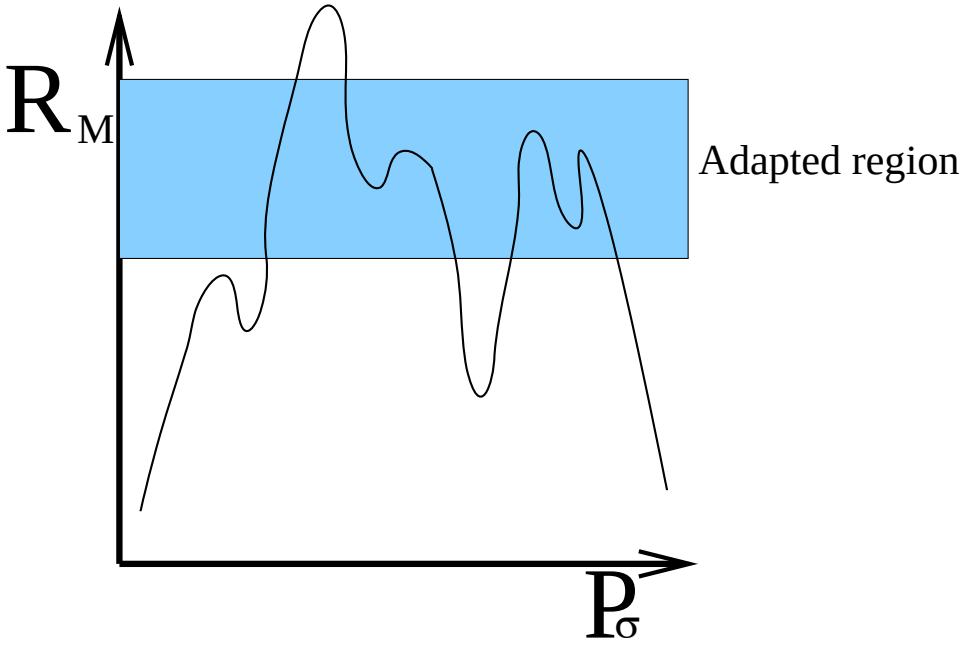
[39] Goodman, M. B. Mechanosensation. (January 06, 2006), WormBook, ed. The C. elegans Research Community, WormBook, doi/10.1895/wormbook.1.62.1,

http://www.wormbook.org.

Figure 1



(a)



(b)

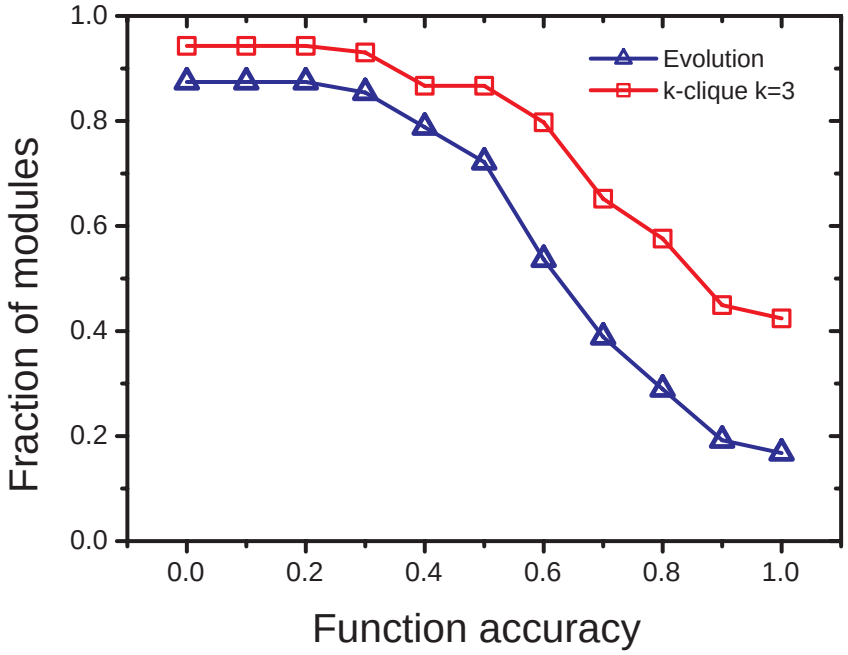
FIG. 1: Schematics illustrating how overlapping modules can be detected at multi-scales. (a) Module A containing a co-clustered group with classification component σ_0. There are three submodules A.1, A.2 and A.3 within A with classification components $\sigma_1, \sigma_2, \sigma_3$ corresponding to these three co-clustered groups. Usually, the community probabilities in components $\sigma_1, \sigma_2, \sigma_3$ are larger than that of component σ_0. The multi-scale modular structures can be obtained as the threshold is being varied from low to high values. (b) Modules were driven to have higher value of robustness function R_M if they were adapted for a sufficiently long time. Only modularity organizations with adequate value of robustness function are advantageous in surviving for cells. Too high or too low value of robustness is harmful for survival.(see also the supporting information).

FIG. 2: Consistency of detected modular structure and real protein annotation data in yeast PPI network. The results of k-clique percolation method are also shown for comparison[5]. (a) Most modules detected by evolution strategy have the main functions, which means that our detected modules are functional units. The results of k-clique percolation appears to have a higher function accuracy only because it detects the strongly connected parts of a network, but not the true modular structures. k-clique percolation, like other methods, fails to be able to recognize the pathways and signal triggering entries of biological processes: over half of the proteins detected in yeast are without modular classification (See supporting information Table SI). (b) The consistency of detected modules and real data in modular size distribution reveals that the true hierarchical organization was captured, while results of k-clique percolation method deviates a lot. (c) Evolution strategy classification membership, i.e. number of modules that a node participated, agrees well with the data on the distribution of number of functions for proteins, but k-clique percolation method does not. It indicates that the overlapping structures of modules was correctly uncovered. (d) The average number of functions for proteins with given node degree k agrees well with real annotation data for degree $k \leq 20$. The strong fluctuation for degree $k > 20$ is due to the low number statistics of these proteins.

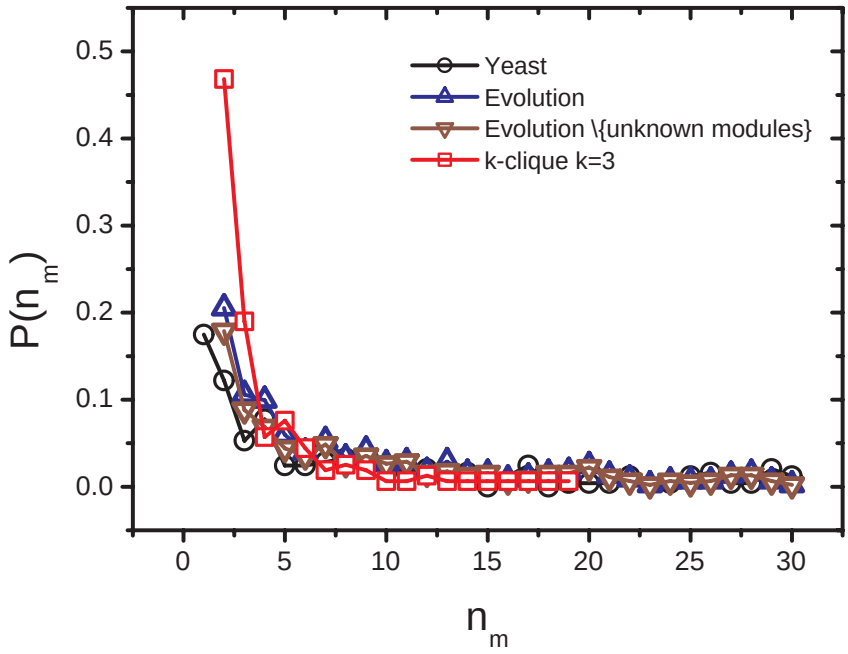
FIG. 3: Examples of detected modules in yeast PPI network. (a) The G_1, S, G_2, M phase cyclins CLN3, CLN1,2, CLB1-5 and two significant inhibitory proteins SIC1, CDH1 are included in the cell cycle control module except CLB6 which is absent from the DIP core data set. CDC28 binds to different phase cyclins to control the progress of cell cycle. The main function of this module is 43.01.03.05, the budding, cell polarity and filament formation, only 29.6%(8/27) of proteins within this module possess this function. Such a low function accuracy is due to the hybrid functions nature in the cell cycle progression control process. (b) Another example of low function accuracy module is a module that functions as the spindle assembly check point. The main function is cell cycle check points 10.03.01.03 with function accuracy 45.5%(5/11). (c) The complete cell cycle process in budding yeast is composed of the cell cycle control, cell size check point, DNA damage check point and spindle assembly check point modules. The signal of spindle assembly check point starts from MAD1, MAD2, MAD3 to CDC20 and PDS1, then entering into the cell cycle control module to trigger the check of spindle assembly[28, 29]. From the cell size check point module, cell growth triggers the cell cycle process by sending signal from YDJ1 to CLN3.

FIG. 4: Detected modules in C. elegans neuronal network reveal significant pathways of physiological phenomena. (a) Soma positions(projected onto the AP axis) for each neuron for the 24 modules detected at threshold 0.3. (b) Average soma position of neurons within each module for the 24 detected modules at threshold 0.3. (c) Sensorimotor modules. Six types of neurons are colored coded[38]: Red: amphids, Orange: other sensory receptors in head, Brown: motorneurons in the nerve ring, Purple: motorneurons in ventral cord, Yellow: neurons in tail ganglia, Green: egg-laying neurons. Neurons that belonged to two of the above six types are in Pink, neurons that belonged to three of the above six types are in Light blue, neurons that belong to none of these six types are in Grey. Neurons belonging to other modules were grouped into a circle at the right lower corner, in e∼g are similar. (d) Model neural circuit for thermotaxis in C. elegans. The chemical synapses and gap junctions are represented in directed and undirected edges respectively. The self undirected edge of a neuron class correspond to gap junctions between left and right neurons within this neuron class. Sensory neurons, interneurons and motor neurons are represented by triangles, squares and circles respectively. (e) Chemosensation/Egg-laying modules: neurons enclosed by green, orange and purple lines are members of Chemotaxis-Egg laying, tail motion, and ventral motion modules respectively. Neurons in blue shadowed region 1, 2, 3 are mechanical sensory, vulva motor and ventral motor neurons respectively. The neurons enclosed by green dashed lines is the submodule at threshold 0.4, which suggests previously unknown pathways of how signals are conveyed from amphid receptors to motorneurons HSNL, HSNR, VC5 and VC4 which innervate vulval muscles and modulate the egg-laying rate. (f) Head and body mechanosensory modules: neurons located within the light blue shadow are the 30 mechanoreceptor neurons in hermaphrodite[39]. (g) Locomotion modules: three modules corresponding to head, ventral and tail motion which cooperate to perform locomotion in C. elegans.

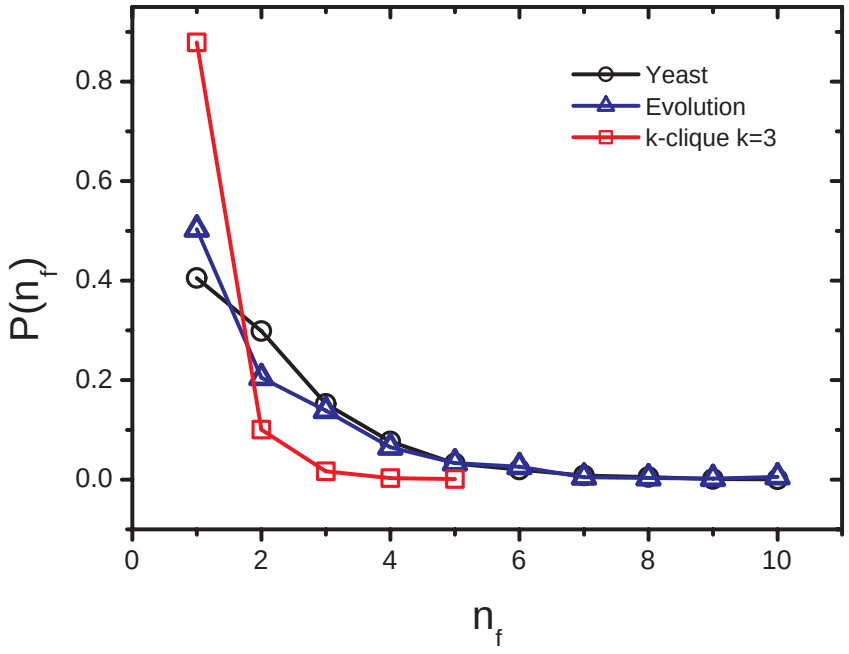
Figure 2


Fraction of modules
Function accuracy
Evolution
k-clique k=3
0.0
0.2
0.4
0.6
0.8
1.0

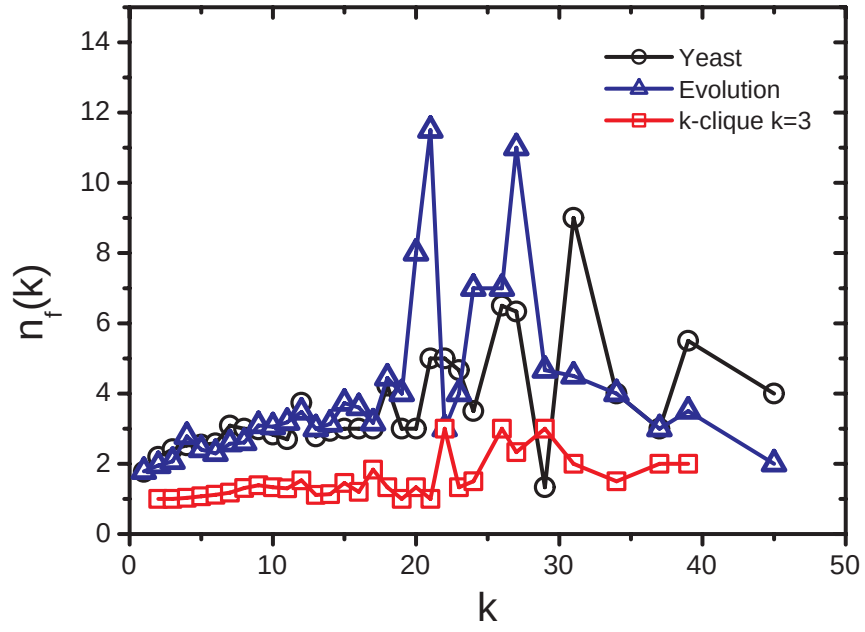
(a)


P(n_m)
n_m
Yeast
Evolution
Evolution \{unknown modules}
k-clique k=3
0.0
0.1
0.2
0.3
0.4
0.5
0
5
10
15
20
25
30

(b)

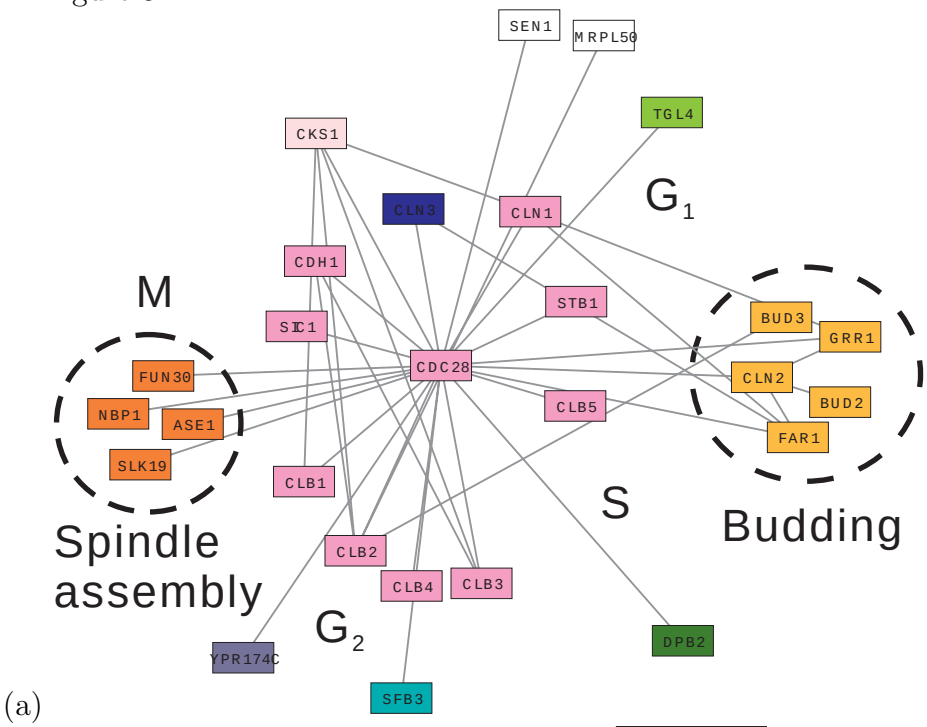

Yeast
Evolution
k-clique k=3
P(n_f)
n_f
0.0
0.2
0.4
0.6
0.8
0
2
4
6
8
10

(c)

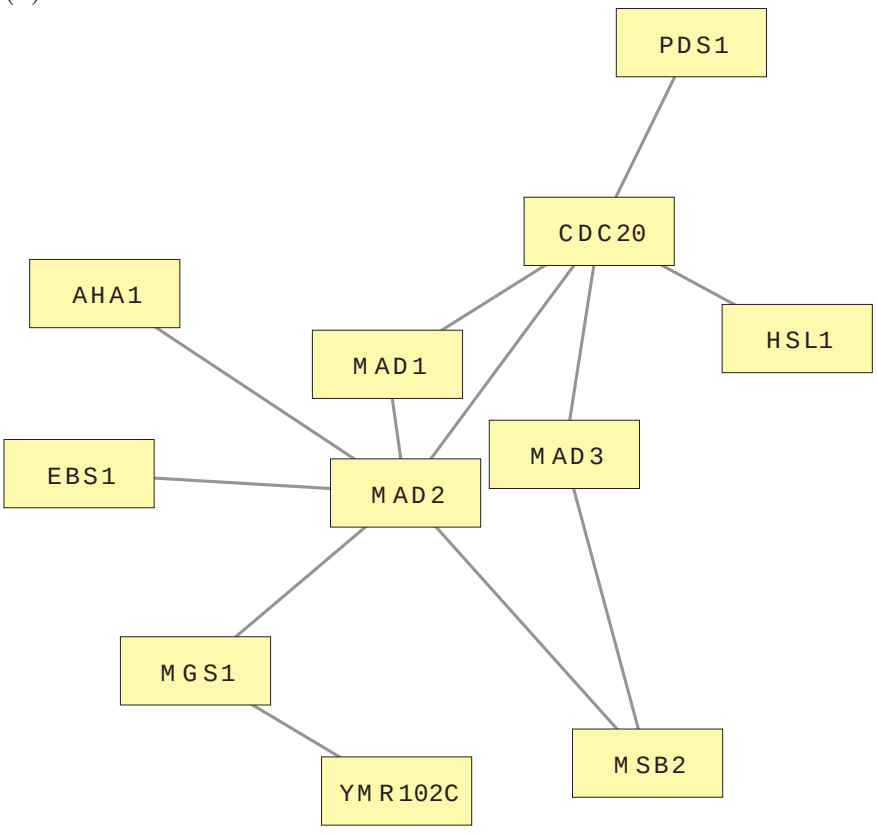

Yeast
Evolution
k-clique k=3
n_f(k)
k
0
2
4
6
8
10
12
14
0
10
20
30
40
50

(d)

Figure 3


SEN1
MRPL50
TGL4
CKS1
CLN3
CLN1
G1
CDH1
M
STB1
SIC1
BUD3
GRR1
CDC28
CLN2
FUN30
BUD2
NBP1
CLB5
ASE1
FAR1
SLK19
CLB1
S
Budding
Spindle
assembly
CLB2
CLB4
CLB3
G2
DPB2
YPR174C
SFB3

(a)


PDS1
CDC20
AHA1
HSL1
MAD1
MAD3
EBS1
MAD2
MGS1
MSB2
YMR102C

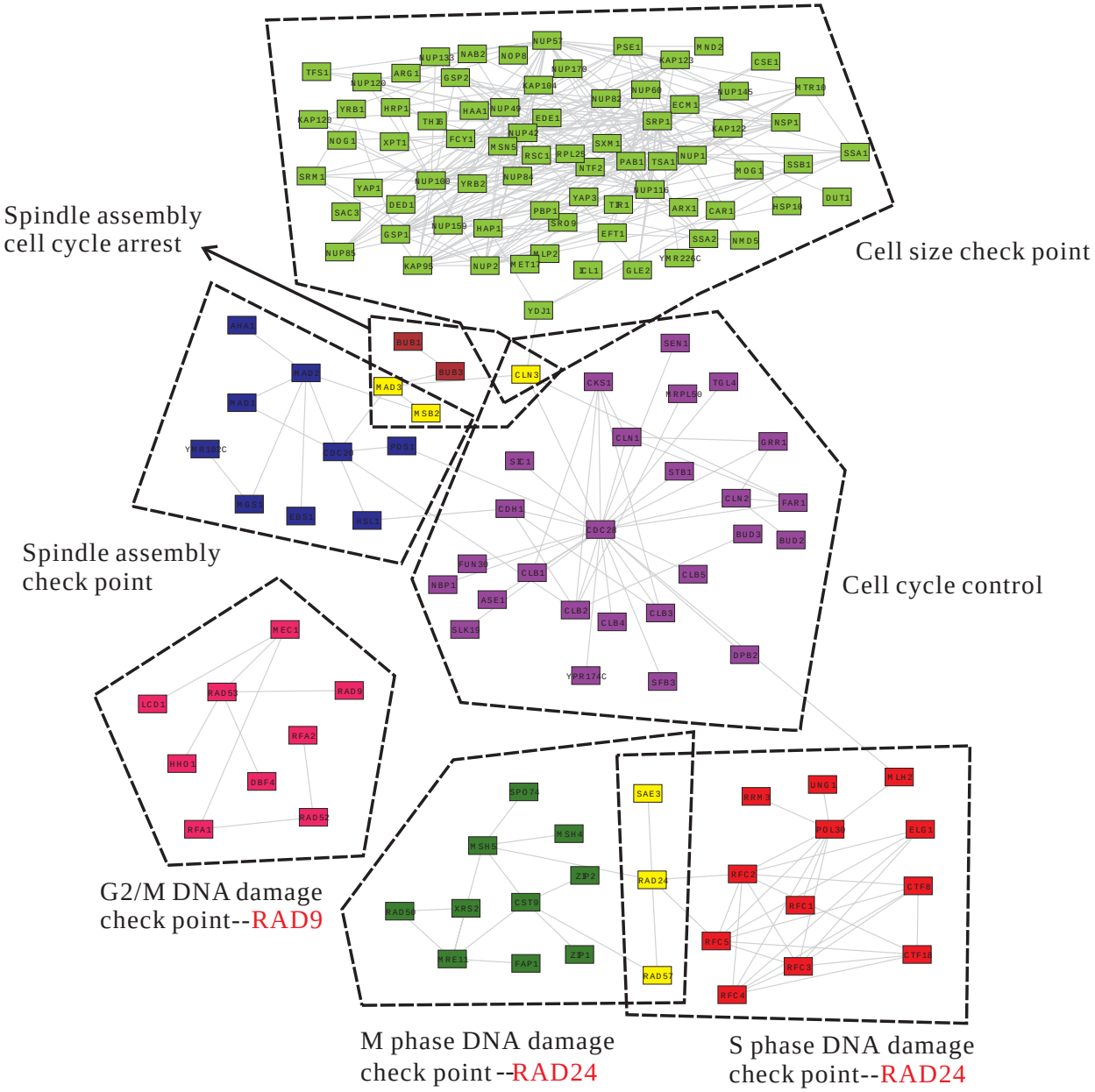
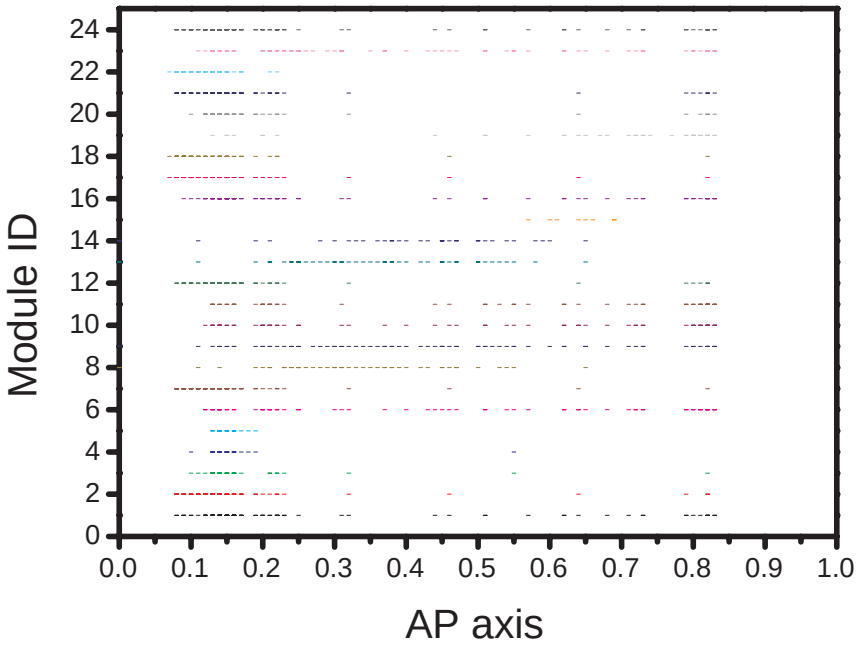
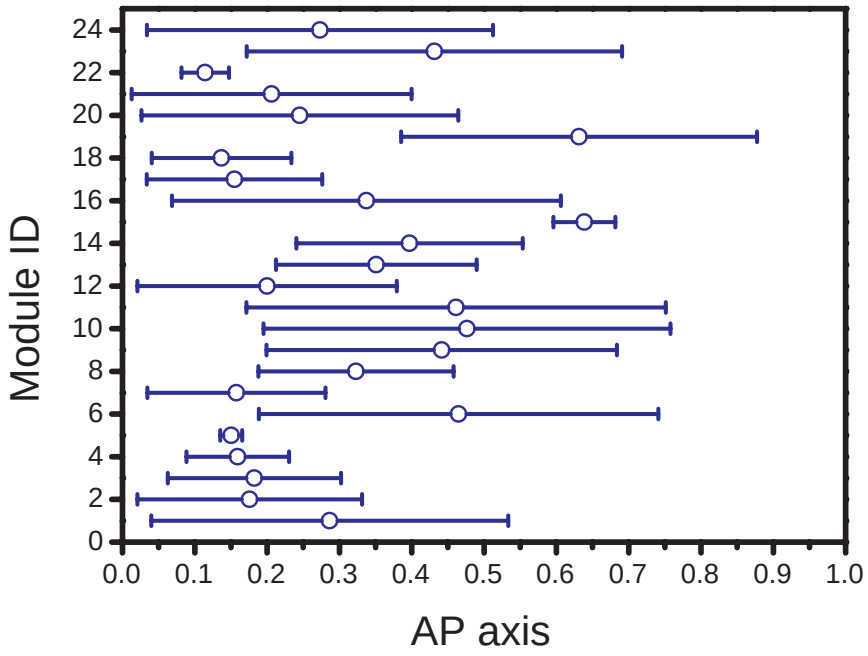

Spindle assembly
cell cycle arrest
Cell size check point
Spindle assembly
check point
Cell cycle control
G2/M DNA damage
check point--RAD9
M phase DNA damage
check point--RAD24
S phase DNA damage
check point--RAD24

Figure 4

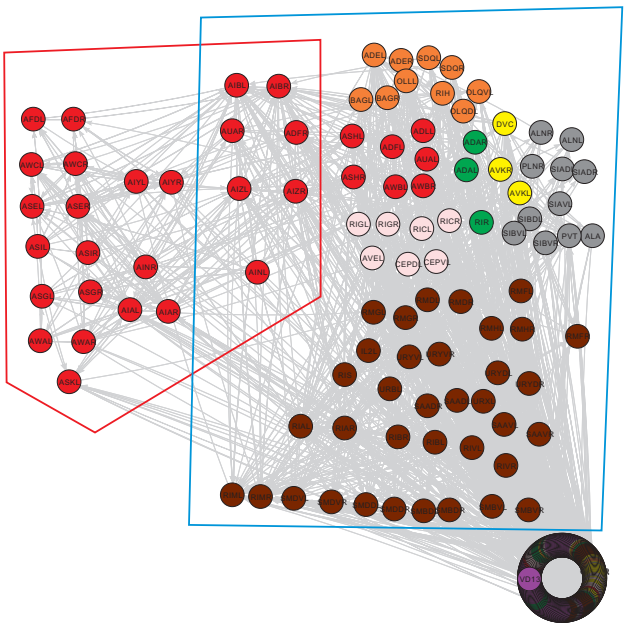

Module ID
AP axis
0
2
4
6
8
10
12
14
16
18
20
22
24
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0

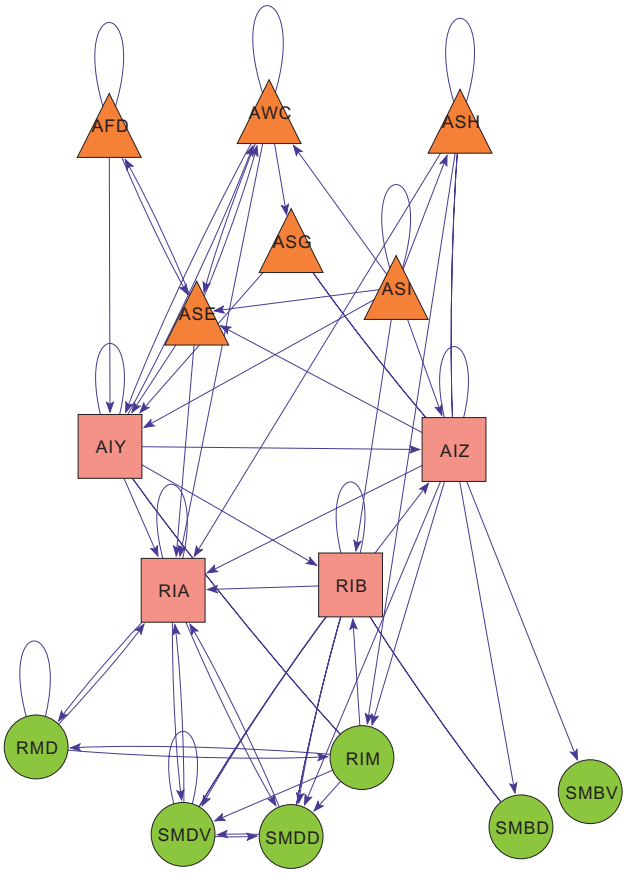
(a)


Module ID
AP axis
0
2
4
6
8
10
12
14
16
18
20
22
24
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0

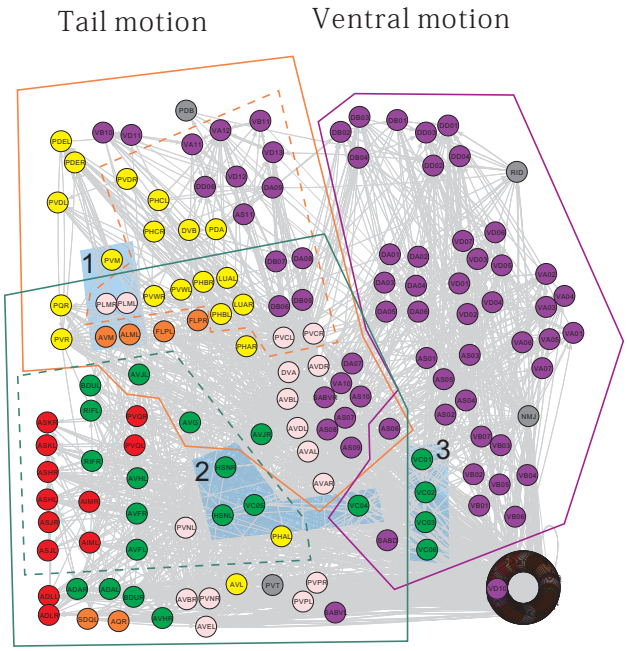
(b)

(c)

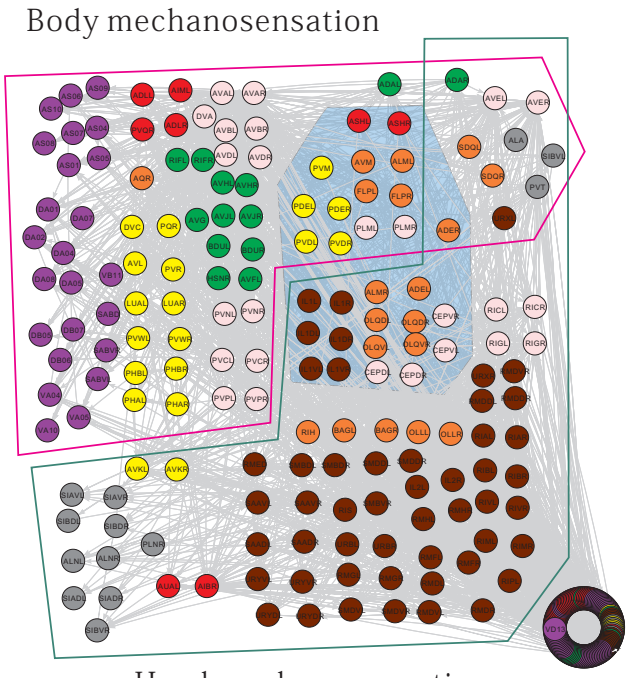




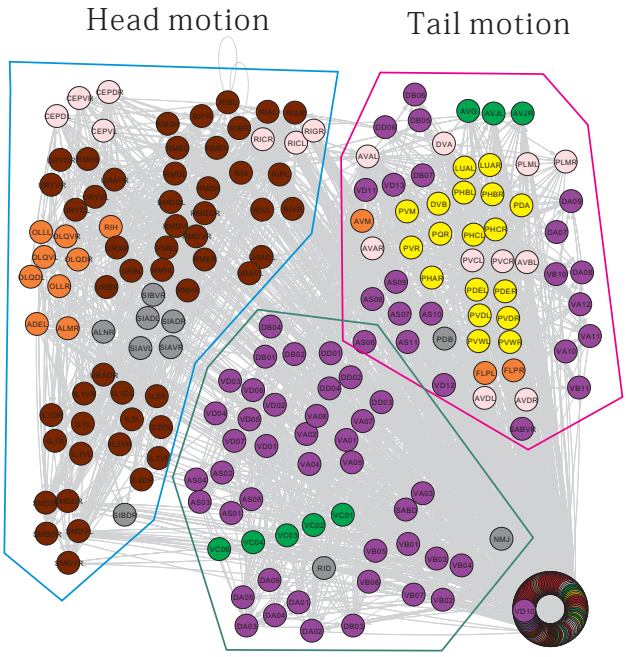
(d)



(e) Chemosensation/Egg-laying


Body mechanosensation
Head mechanosensation

(f)


Head motion
Tail motion
Ventral motion

(g)

行政院國家科學委員會補助國內專家學者出席國際學術會議報告

100 年 9 月 7 日

附件三

報告人姓名	黃俊燕	服務機構及職稱	中華大學生物資訊學系 副教授
時間 會議 地點	2011/8/28-2011/9/1 德國 Mannheim	本會核定 補助文號	NSC 99-2112-M-216-001
會議名稱	(中文)第十二屆國際系統生物研討會 (英文) The 12th International Conference on systems Biology		
發表論文題目	(中文) 以模塊法解析線蟲神經系統之功能組織 (英文) Modularity approach to decipher functional organization of neural system in C. elegans		

報告內容應包括下列各項：

一、參加會議經過

此次為本人首次參與國際系統生物研討會，該項國際系統生物研討會行之有年是系統生物學界一年一度的盛會，此回國際系統生物研討會由德國海德堡大學承辦於 Mannheim 的 Rosengarten 會議中心舉行，從 8/28 到 9/1 為期五天，由於此次德國行機會難得本人於 8/22-8/27 額外安排了訪問 Koln 與 Stuggart 的訪問行程，此行於 8/21 搭機前往德國法蘭克福，8/22-8/27 進行私人訪問行程，於 8/28 早上抵達 Mannheim 參加國際系統生物研討會，大會安排於 8/28 下午三點舉行開幕儀式，開幕後隨即接著四場精采的大會演講並於晚間八點舉行歡迎接待交誼餐會。

此回系統生物研討會，大會將口頭發表與張貼壁報論文依其性質區分為

1. Design principle of living systems
2. Model identification and discrimination
3. Systems Genomics and evolution
4. Systems Immunology
5. Host-Pathogen Interactions
6. Personalized medicine and drug development
7. Synthetic biology
8. Cell mechanics
9. Systems biology of aging
10. Computational tools and algorithms
11. Systems neuroscience

本人此回於 systems neuroscience 發表一篇壁報論文，難得的是有機會與從事蛋白質交互作用研究素富盛名的德國學者 P. Bork 做點短暫的意見交流。此次大會比較特殊的是於八月三十日下午安排了 Industrial session，並且於八月三十一日下午有安排了 Industrial workshop，並且於國際系統生物研討會前後於海德堡大學 BioQuant 與 DKFZ 安排了多達三十場左右的各種系統生物議題的 workshops，而且光是發表的壁報論文就多達六百多篇，足見此次系統生物研討會內容豐富程度。

二、與會心得

此行讓個人更能了解到目前系統生物學國際學者所關心的研究議題，而且也見證到了歐洲國家系統生物學雄厚的研究能量與研究題材的多樣性，更難得的是歐洲系統生物學界與業界緊密結合的程度值得國內學習效法，例如德國 Virtual liver network。由於國內從事系統生物研究的學者為數不多，而系統生物學又是當前生物學界公認的未來趨勢與主流，所以國內學者應該有多一點類似參與系統生物國際研討會的機會，以推廣系統生物於國內的研究量與質。

三、考察參觀活動(無是項活動者省略)

1. 海德堡大學
2. 海德堡大學 BioQuant 與 DKFZ

四、攜回資料名稱及內容

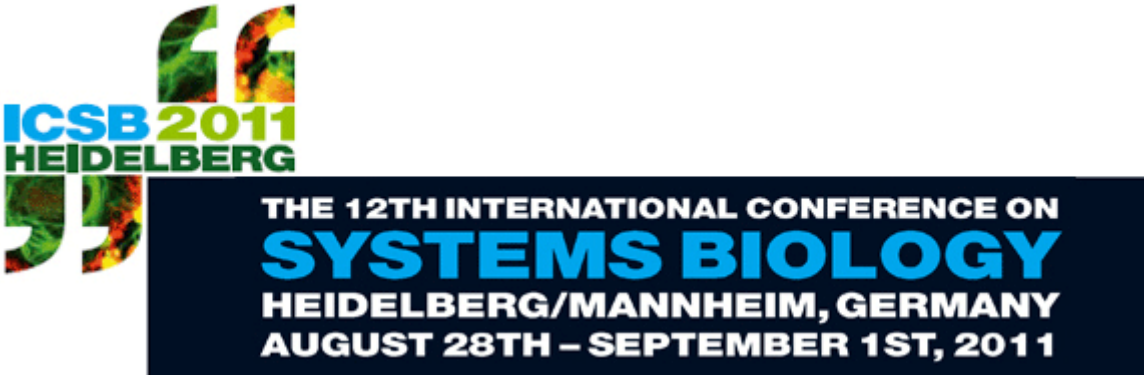
1. 大會秩序手冊(含時間手冊與發表亂文摘要)
2. 德國系統生物研究現況簡介資料，systembiologie.de 雜誌 2011 年第三期
3. IBM,Virtual liver network 等業界公司簡介

五、其他

表 Y04

Jiun-Yan Huang

寄件者: "ICSB 2011" <info@mcon.clvrmail.com>
收件者: <jyhuang@chu.edu.tw>
傳送日期: 2011年7月21日 上午 01:22
主旨: ICSB 2011 - Your Abstract for the 12th International Conference on Systems Biology



ICSB 2011 - Your Abstract for the 12th International Conference on Systems Biology **20.07.2011**

Dear Professor Huang,

Thank you for submitting an abstract for the **12th International Conference on Systems Biology (ICSB) in Mannheim, Germany 28th August - 01st September 2011.**

We are happy to inform you that your abstract was accepted as **"poster".**

The poster session during which you should be present is planned as follows:

Abstract Title: Modularity approach to decipher functional organization of neural system in C. elegans

Poster Number: PS 576

Scientific Topic: Systems Neuroscience

Date: 30.08.2011

Time: 13:00 - 16:00 h

The exhibition area for posters will be in the foyer of level 1 of the conference center.

Your poster should present for the duration of the conference. It should be hung on Sunday, August, 28th by 05:00 p.m.; pins for mounting of the poster will be provided.

Please remove your poster on Thursday, September, 1st by 1:30 p.m. otherwise your poster will be removed and disposed off by our staff.

The size of the poster board is: height: 150 cm width: 95 cm. Please do to exceed these dimensions with your poster.

For further information on the conference please visit: www.icsb-2011.net

We are looking forward to welcome you and your co-workers at the ICSB conference 2011.

Yours sincerely

Thilo Hübner
m:con project management

Tanja Berger
m:con abstract management

Online Version - unsubscribe

Impressum
Imprint
m:con - mannheim:congress GmbH Rosengartenplatz 2 D-68161 Mannheim Tel. +49 (0)621 / 4106-0 Fax +49 (0)621 / 4106-80121 E-mail: info@mcon-mannheim.de Internet: www.mcon-mannheim.de Chairman of the Supervisory Board: Lord Mayor Dr. Peter Kurz Manager: Michel Maugé, michel.mauge@mcon-mannheim.de Company headquarters: Mannheim Commercial register: Mannheim District Court, HRB 5582 VAT ID no.: DE 811 968 225 Tax number: 3 810 700 256 Bank details: Sparkasse Rhein Neckar Nord, Sort code 670 505 05, Account no. 30 179 498 IBAN: DE71 6705 0505 0030 1794 98 SWIFT-BIC: MANS DE66 XXX

Modularity approach to decipher functional organization of neural system in C. elegans

Jiun-Yan Huang[1](黃俊燕) , Pik-Yin Lai[2](黎璧賢)

[1] Department of Bioinformatics, Chung Hua University, Hsin Chu 300, Taiwan, R.O.C

[2] Department of Physics, Graduate Institute of Biophysics and Center for Complex Systems, National Central University, Chung-Li 320, Taiwan, R.O.C

C. elegans is a primitive model organism for neural system study. There are 302 neurons distributed from head to tail in C. elegans. The connectivity data had been assembled by J.G White et al. in 1986[1]. In the past ten years, the rapidly developed of complex network theory helps one to analyze subcellular molecular networks and neuronal networks. This connectivity data of C. elegans was analyzed via complex network approach by Lav R. Varshney et al.[2]. From connectivity to functionality becomes one of the most challenging and important issues in C. elegans neuroscience. However, a satisfactory method to unravel hierarchically modular structure of neuronal network is still lacking.

Here, we developed an evolution-based method to detect modular structure for C. elegans neuronal network. Our theory could detect overlapping modules at multi-scale in a network, which is appropriate to apply to hierarchically organized neuronal network system. From our results, functions of several modules can be recognized, e.g sensory, nerve ring motor, locomotion, chemosensation /egg-laying, mechanosensation, etc.. Furthermore, higher level organization of physiological behaviors, such as how chemosensation affects egg-laying rate, thermotaxis physiological behavior, were successfully resolved by interactions of these modules. Hence, our theory offers a systems way to decipher functional organization of neural system from connectivity data.

References:

[1] J. G. white et al., Phil. Trans. Royal Soc. London. Series B, Biol Scien. 314, 1-340, (1986).

[2] Lav R. Varshney et al., PLoS Compt. Biol. 7(2) :e10010066, (2011).

e-mail: jyhuang@chu.edu.tw, pylai@phy.ncu.edu.tw

國科會補助計畫衍生研發成果推廣資料表

日期:2011/10/24

國科會補助計畫	計畫名稱: 生物與複雜網路模組結構之解析與應用研究
	計畫主持人: 黃俊燕
	計畫編號: 99-2112-M-216-001- 學門領域: 軟物質及生物物理－理論

無研發成果推廣資料

99 年度專題研究計畫研究成果彙整表

計畫主持人：黃俊燕	計畫編號：99-2112-M-216-001-
計畫名稱：生物與複雜網路模組結構之解析與應用研究	

成果項目			量化			單位	備註（質化說明：如數個計畫共同成果、成果列為該期刊之封面故事...等）
			實際已達成數（被接受或已發表）	預期總達成數(含實際已達成數)	本計畫實際貢獻百分比		
國內	論文著作	期刊論文	0	0	100%	篇	
		研究報告/技術報告	0	0	100%		
		研討會論文	0	0	100%		
		專書	0	0	100%		
	專利	申請中件數	0	0	100%	件	
		已獲得件數	0	0	100%		
	技術移轉	件數	0	0	100%	件	
		權利金	0	0	100%	千元	
	參與計畫人力（本國籍）	碩士生	1	0	100%	人次	第八屆 台灣生物資訊與系統生物學研討會, Chu-Yi Tai, Yu-Kai Tseng, Wei-Ti Wang, Po-Yu Lu and Jiun-Yan Huang, The functional analysis of modular structures in C. elegans
		博士生	0	0	100%		
		博士後研究員	0	0	100%		
		專任助理	0	0	100%		
國外	論文著作	期刊論文	0	1	90%	篇	
		研究報告/技術報告	0	0	100%		
		研討會論文	1	0	100%		2011 International Conference on Systems Biology, Heidelberg
		專書	0	0	100%	章/本	
	專利	申請中件數	0	0	100%	件	
		已獲得件數	0	0	100%		
	技術移轉	件數	0	0	100%	件	

		權利金	0	0	100%	千元	
	參與計畫人力（外國籍）	碩士生	0	0	100%	人次	
		博士生	0	0	100%		
		博士後研究員	0	0	100%		
		專任助理	0	0	100%		
其他成果 (無法以量化表達之成果如辦理學術活動、獲得獎項、重要國際合作、研究成果國際影響力及其他協助產業技術發展之具體效益事項等，請以文字敘述填列。)		本計劃研究成果已撰寫完成標題為’’’’ Robustness and adaptation reveals significant functional complexity of evolving biological networks at multi-scales’’’’ 的論文草稿，此論文至目前為止為解決生物網路模組具有重疊與多尺度結構問題的第一個理論方法，針對不同尺度的生物網路，例如細胞內之蛋白質交互作用網路，線蟲神經元網路，以及哺乳類動物大腦皮質區網路都有非常好的結果且衍生出許多重要的生物知識，目前此草稿於 Nature Biotechnology 期刊審稿中。					

	成果項目	量化	名稱或內容性質簡述
科教處計畫加填項目	測驗工具(含質性與量性)	0	
	課程/模組	0	
	電腦及網路系統或工具	0	
	教材	0	
	舉辦之活動/競賽	0	
	研討會/工作坊	0	
	電子報、網站	0	
	計畫成果推廣之參與（閱聽）人數	0	

國科會補助專題研究計畫成果報告自評表

請就研究內容與原計畫相符程度、達成預期目標情況、研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）、是否適合在學術期刊發表或申請專利、主要發現或其他有關價值等，作一綜合評估。

1. 請就研究內容與原計畫相符程度、達成預期目標情況作一綜合評估

■達成目標

□未達成目標（請說明，以 100 字為限）

□實驗失敗

□因故實驗中斷

□其他原因

說明：

2. 研究成果在學術期刊發表或申請專利等情形：

論文：□已發表 ■未發表之文稿 □撰寫中 □無

專利：□已獲得 □申請中 ■無

技轉：□已技轉 □洽談中 ■無

其他：(以 100 字為限)

3. 請依學術成就、技術創新、社會影響等方面，評估研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）(以 500 字為限)

本計劃研究成果為第一個理論工具解決目前生物網路模組結構計算與預測時所面臨到的，大小解析度極限、模組重疊、多尺度模組組織結構之困難，應用大尺度的生物網路可以將高通量的生物實驗數據與小規模的分子生物知識連結起來，提供日後基因體蛋白質體學家解希奇實驗數據之生物意義具有強大功用，最近 2010 八月起國際間亦開始重視過去以往理論工具在模組結構之重疊性與多尺度組織特性上之困難，開始有作者於 Nature 466, 761-765, 2010.等知名期刊討論此問題，本計劃成果能率先克服此些困難歸功於提出最佳化理論於模組預測上之不適用性，反觀目前各種計算模組的理論方法與工具數量非常多，其中幾乎所有理論工具都仰賴最佳化理論，所以本計劃成果提出最佳化理論於演化生物系統之不完全適用性，相信此一觀點也持續著會影響其他生物問題之研究。