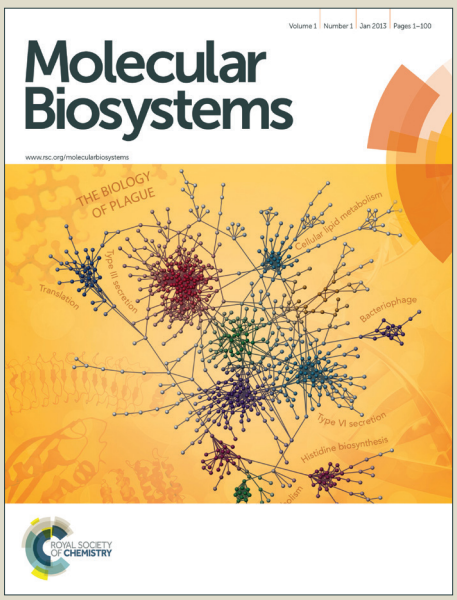
Molecular BioSystems

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/molecularbiosystems

Journal Name

Molecular BioSystems

Dynamic Article Links ►

Cite this: DOI: 10.1039/c0xx00000x

www.rsc.org/xxxxxx

ARTICLE TYPE

Identifying the potential extracellular electron transfer pathways from c-type cytochrome network

Dewu Ding,[a,b] Jun Xu,[a] Ling Li,[a] Jianming Xie[a] and Xiao Sun*[a]

Received (in XXX, XXX) Xth XXXXXXXXX 20XX, Accepted Xth XXXXXXXXX 20XX

DOI: 10.1039/b000000x

Extracellular electron transfer (EET) is the key feature of some bacteria, such as *Geobacter sulfurreducens* and *Shewanella oneidensis*. Via EET processes, these bacteria can grow on electrode surfaces and make current output of microbial fuel cells. C-type cytochromes can be used as carriers to transfer electrons, which play an important role in EET processes. Typically, from inner (cytoplasmic) membrane through periplasm to outer membrane, they could form EET pathways. Recent studies suggest that a group of c-type cytochromes could form a network which extended the well-known EET pathways. We obtained the protein interaction information for all of 41 c-type cytochromes in *Shewanella oneidensis* MR-1, constructed a large-scale protein interaction network, and studied its structural characteristics and functional significance. Centrality analysis has identified the top 10 key proteins of the network, and 7 of them are associated with electricity production in the bacteria, which suggests that the ability of *Shewanella oneidensis* MR-1 to produce electricity might be derived from the unique structure of the c-type cytochrome network. By modularity analysis, we obtained 5 modules from the network. The subcellular localization study has shown that the proteins in these modules all have diversiform cellular compartments, which reflects their potential to form EET pathways. In particularly, combination of protein subcellular localization and operon analysis, the well-known and new candidate EET pathways are obtained from the Mtr-like module, indicating potential EET pathways could be obtained from such a c-type cytochrome network.

1. Introduction

C-type cytochromes play an important role in the extracellular electron transfer (EET) processes, which are indispensable elements for extracellular respiration and current output of electricigens (microorganisms that are able to efficiently transfer electrons from themselves to electrodes, such as *Geobacter sulfurreducens* and *Shewanella oneidensis*).[1-2] In *Geobacter sulfurreducens*, c-type cytochromes OmcE and OmcS are responsible for the transmission of electrons to pili, then the pili further transfers electrons to the extracellular solid metal oxides.[3] In *Shewanella oneidensis*, the MtrCAB pathway is the critical EET pathway, which is posed by numerous of c-type cytochromes: via the menaquinone pool, inner membrane cymA obtains electrons which generated from substrate oxidation, and transfers them to periplasmic mtrA, then the outer membrane mtrB transfers electrons from mtrA to mtrC/omcA complexes, and ultimately to reduce the extracellular electron acceptors by direct contact.[4] In other words, c-type cytochromes could transfer the electrons from the inner (cytoplasmic) membrane through the periplasm to the outer membrane in these microorganisms.

Peng and colleagues' study on c-type cytochromes in *Shewanella* showed that the electrodes could influence the distribution and regulation of c-type cytochromes in the microbial cell surface, that is, the expression of c-type cytochromes is affected by the environment.[5] Research on *Shewanella* genomes suggested that there are a large number of c-type cytochrome genes,[6] which means c-type cytochromes have a rich diversity and could form different combinations in *Shewanella*, and thus constitute different EET pathways to achieve efficient electron transfer.

However, most of the current researches focused on the detail of a few specifical c-type cytochromes.[7-8] Until recently, it has been suggested that a group of c-type cytochromes could form a network which might extend the EET pathways.[9] For example, Zhang *et al*. reported the omcA–mtrC interaction network, but their network only contains 18 proteins.[10] Genome-wide study for the interactions among all of c-type cytochromes is not available. This study presented such a genome-wide c-type cytochrome network (section 3.1) that led us to investigate its structural characteristics and functional significance. Firstly, biological networks seem to have evolved towards a structure that is efficient to control cellular processes (such as promoting EET here), we wondered whether and how the purpose could be

[a] *State Key Laboratory of Bioelectronics, School of Biological Science and Medical Engineering, Southeast University, Nanjing, P.R. China. Fax: +86-025-83792349; Tel: +86-13951989906; E-mail: xsun@seu.edu.cn*

[b] *Department of Mathematics and Computer Science, Chizhou College, Chizhou, P.R. China*

Molecular BioSystems Accepted Manuscript

achieved by the c-type cytochrome network, the dominating nodes in the network were then determined by integrated centrality analysis (section 3.2). More importantly, we wondered whether potential EET pathways could be identified from the network (or functional modules in the network). The results indicate that the well-known and new candidate EET pathways could be identified based on the combination of protein subcellular localization and operon analysis (section 3.3). We thus proposed that potential EET pathways could be identified from such a c-type cytochrome network, which is our main finding in this work.

2. Materials and methods

2.1 Protein interaction network

A) c-type cytochromes C-type cytochromes are highly water soluble proteins, the basic function of such molecules is the transmission of electrons. They can covalently binding with heme by two cysteine (C), the sequence feature of heme-binding sites is CXXCH motif, and thus all c-type cytochromes contain a CXXCH sequence. Using genome annotation data [11] and pattern matching, Meyer *et al*. has identified 42 genes encoding c-type cytochromes in *Shewanella oneidensis* MR-1.[6]

B) protein-protein interaction Numbers of techniques have been developed to study protein-protein interaction, and researches in this field have accumulated a lot of data in the last two decades.[12-13] We used the famous protein interaction database STRING (http://string.embl.de) to analyze the interaction information of c-type cytochromes, and then built the protein interaction network for them.[14] There are two important parameters when STRING is used, the first one is minimum confidence score, where the confidence score is the approximate probability that a predicted interaction exists between two proteins in the same metabolic map in the KEGG database, the other is interactors which are other proteins predicted to be interacted with the seeds (i.e., the c-type cytochromes in our work). In order to obtain more comprehensive information, medium confidence (no less than 0.4) and 50 other interactors were used. As the complementarity, experimental identified and verified interactions were also considered by literature retrieval (see section 3.1).

2.2 Centrality analysis

To identify the key c-type cytochromes, we used centrality analysis. Herein, centrality is a concept which is used to identify the relative importance of a node in a given network. Generally speaking, a centrality measure is a function which assigns a numerical value *C(v)* to each node of a network, and there are many different definitions and formulas used to measure centrality. For examples, the simplest measure is degree centrality which is used to indicate the number of links of each node in the network, betweenness centrality is corresponding to the number of shortest pathways going through the nodes, while closeness centrality is helpful to identify nodes in the core and periphery part of the network, and so on. It is suggested that the combination of several measures are more effective. We thus engaged ***Degree, Eccentricity, Closeness, Centroid Value, Stress, Shortest Path Betweenness, CF-Closeness*** and ***PageRank*** centrality measures in this study. All measures were performed with CentiBiN (http://centibin.ipk-gatersleben.de).[15]

2.3 Modularity analysis

In generally, functional modules can be clearly associated to certain biological processes, such as known biological pathways. Furthermore, the small reduced modules also should be helpful to understand the structure and function of complex biological networks. The basic principle for identifying such modules in biology networks is dense node–node links within modules but sparser links between them. An important measure related to detection of modules is modularity *M*. For a presumptive partition of a network into several modules, the modularity *M* of this partition is defined as following:

$$M \equiv \sum_{s=1}^{r}\left[\frac{l_s}{L}-\left(\frac{d_s}{2L}\right)^2\right] \qquad (1)$$

where *r* is the number of modules, *ls* is the number of links between nodes in modules, *ds* is the sum of the degrees of the nodes in module *s*, and *L* is the total number of links in the network. It is suggested that maximization of the modularity *M* would yield the most accurate results for real world complex networks, and thus is widely used for identification of modules in networks.[16] Guimera and Amaral introduced a simulated annealing method to obtain the best partition of a network by maximization of modularity *M*, which could be achieved by a command line tool "NetCarto" (kindly provided by prof. Roger Guimera).[17-18]

2.4 Subcellular localization

The subcellular localization of proteins could contribute to understand the EET processes and the role of different c-type cytochromes in EET. To perform the subcellular localization, we used PSORTb (http://www.psort.org/psortb), one of the best tools for current subcellular localization analysis.[19]

2.5 Operon prediction

Operon is made up of multiple adjacent genes with the related regulatory signals, and it could form the basic transcript unit to affect the gene expression. In fact, operon regulation mode is a major mode of regulation in prokaryotes, the relativity of the function and the regulation of the expression could be achieved via operon mechanism for many prokaryote genes. It has been indicated that the genes, which are constituted an operon, are often functional relativity or within the same metabolic pathway. Thus, operon is useful for deducing EET pathways. The operons in *Shewanella oneidensis* MR-1 were obtained by MicrobesOnline Operon Predictions Tool (http://www.microbesonline.org/operons/).[20]

2.6 Gene expression

Gene expression data analysis provides information on gene function. Genes with significant expression differences under different conditions could be filtered by expression profiling analysis, which might be helpful to understand different EET processes, as well as to verify potential EET pathways. To perform gene expression analysis, the $\log_2$ ratio comparison data

Molecular BioSystems Accepted Manuscript

This journal is © The Royal Society of Chemistry [year]

of gene expression level between DMSO + Lactate (Treatment) and LB + O_2 (Control) for the main c-type cytochromes were downloaded from MicrobesOnline (http://www.microbesonline.org/cgi-bin/microarray/viewExp.cgi?expId=2426). Furthermore, we also used the expression dataset GSE39462 for 15 other growth conditions, including aerobic growth in Luria-Bertani (LB) broth, aerobic growth in defined minimal media with 8 different carbon sources, aerobic growth in defined lactate medium with 4 different stresses and anaerobic growth in a defined medium with 2 different carbon sources.[21] The dataset was downloaded from GEO (http://www.ncbi.nlm.nih.gov/geo). All related analyses were conducted in R platform (http://r-project.org) using the Bioconductor package.

3. Results and discussion

3.1 Construction of protein interaction network

To construct the protein interaction network, we first obtained all of 42 c-type cytochromes from Meyer *et al*.' report,[6] and removed SO_4570 from the list, since it is a pseudo gene for dmsE (see http://www.ncbi.nlm.nih.gov/gene/?term=SO_4570 for details). All methods in STRING were used to obtain the protein interaction information. For a protein pair, a list of confidence scores (between 0 and 1, indicating the probability that a predicted interaction exists, see section 2.1 B) would be obtained by using these methods, based on which an integrated score could be obtained subsequently according to STRING's scoring rules.[14, 22] For example, for the protein pair omcA and mtrC, STRING gives a list of confidence scores based on different methods: neighborhood (0.421), gene fusion (0), co-occurrence (0.674), homology (0.498), co-expression (0), experiments (0), databases (0) and textmining (0.738). Accordingly, an integrated score of 0.751 is given by STRING. Since the integrated score is higher than selective minimum confidence score (0.4), omcA and mtrC were hence treated as an interaction pair (omcA pp mtrC). STRING includes both physical and functional interactions,[23] but all predicted interactions were represented as undirected links in this study, regardless of their types. Then, we obtained all interaction information for the 41 remaining c-type cytochromes. In order to obtain more comprehensive information, 50 other proteins (interactors in STRING's glossary, see also section 2.1 B) which were predicted to be interacted with these 41 c-type cytochromes were considered. As a result, 513 interactions among 91 proteins (41 c-type cytochromes and 50 interactors) were obtained from STRING (Dec 2013).

In theory, only experiments can verify these interactions, but large-scale experiments are beyond the scope of the present work. Nevertheless, experimental identified and verified interactions are also considered as the complementarity. Most of current researches focused on important c-type cytochromes in Mtr pathways (e.g., cymA, mtrC, omcA, *etc*), the interactions about these c-type cytochromes are investigated in detail. We found that most of these experimental verified interactions are included in our predicted interactions list. Nevertheless, 4 new interactions were discovered by literature retrieval,[7-10, 24-28] they are: mtrA pp mtrD, fccA pp cymA, fccA pp mtrD and fccA pp cctA, which make that there are 91 proteins and 517 interactions in the final protein interaction network.

The network is modelled as an undirected graph, in which nodes represent the proteins and the links represent interactions among them. The global topology structure of the protein interaction network of c-type cytochromes in *Shewanella oneidensis* MR-1 is shown in figure 1 with cytoscape (http://www.cytoscape.org/) .[29]

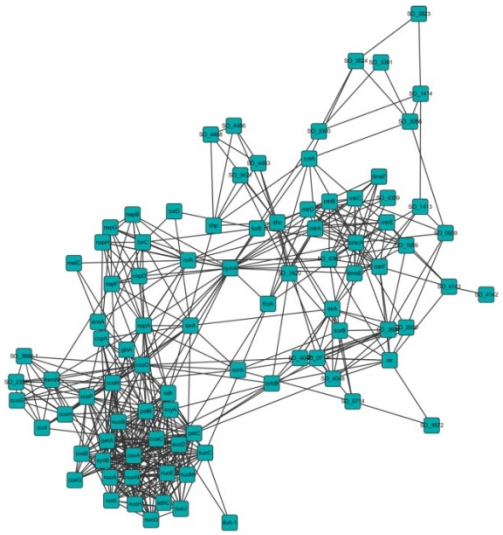


Fig. 1 The protein interaction network of c-type cytochromes in *Shewanella oneidensis* MR-1. The network consists of 91 proteins and 517 interactions. Several denser regions are clearly visible from the figure, indicating that the interaction modules of proteins participating in similar biological process might be existent.

Due to the protein interaction information for the 41 c-type cytochromes only might be deficient, and which will lead that the model contains many separate parts and can not form a whole network. The 50 interactors considered previously not only for more comprehensive interaction information, but also with the purpose to construct an integrated protein interaction network.

From topology and functional evolution view, there might be some key nodes (proteins) in the network. Furthermore, although this c-type cytochrome network is very sparse (density less than 0.13), several denser regions are clearly visible from figure 1. The protein interaction network would like to provide the information for the interaction modules of proteins participating in similar biological process. Thus, we need to identify key proteins and functional modules in the network next.

3.2 Centrality analysis

Topology of a network has been shown to strongly correlate with its function.[30] We thus studied the key proteins in network with multi-centrality measures. Firstly, we ranked the top 10 key proteins by every centrality measure, and due to different measures focus on different aspects of centrality, so the key proteins might be different according to different centrality measures (see table 1). Then, to select the top 10 key proteins from an integrated view, we ranked all proteins in table 1 by the number of they appeared there (for the same instance, the sum of rank number for the protein in table 1 are used). As a result, the top 10 key proteins among them are given in table 2.

Molecular BioSystems Accepted Manuscript

We found that such an integrated analysis seems effective, 7 of 10 key proteins are c-type cytochromes which are associated with electricity production. The first one is ***cymA*** which can get the electrons from substrate oxidation, this tetraheme cytochrome has been shown to act as a hub in EET processes, and it is required for almost all forms of anaerobic respiration in *Shewanella*. For example, cymA is the source for gaining electrons in the MtrCAB pathway.[7-9, 24] The other important element ***omcA*** which act as the sink for losing electrons in the MtrCAB pathway has also been identified as a key protein, this outer membrane c-type cytochrome could restore extracellular electron acceptors by direct contact with them,[24] or donate electrons to riboflavins or soluble extracellular redox compounds.[31-32] Furthermore, ***ccoO*** and ***petC*** are proteins involved in aerobic respiration in various organisms, and have been found playing critical roles in both aerobic and anaerobic respiration with highly toxic metals as electron acceptor;[33] the TMAO respiration system terminal reductase ***torA*** is located in the periplasm, it probably receives electrons from the pentahemic c-type cytochrome torC, and the periplasmic nitrate reductase ***napA*** has been used in respiratory nitrate ammonification;[34] and ***coxC*** plays multi-roles in aerobic electron transport chain, respiratory electron transport chain, as well as oxidation-reduction process. Despite of that there is no direct experimental evidence currently, the remaining 3 key proteins (petB, ndh and ccoN) might also be of great interest to electricity production, petB and ccoN might be associated with petC and ccoO respectively, while ndh is the respiratory NADH dehydrogenase.

Table 1 The top 10 key proteins corresponding to 8 different centrality measures (Degree, deg; Eccentricity, ecc; Closeness, clo; Centroid Value, cen; Stress, str; Shortest Path Betweenness, spb; CF-Closeness, cfc; PageRank, pr).

RANK	V_{deg}	V_{ecc}	V_{clo}	V_{cen}	V_{str}	V_{spb}	V_{cfc}	V_{pr}
1	petB	mtrA	cymA	cymA	cymA	cymA	cymA	cymA
2	petC	omcA	ccoO	ccoO	ccoO	ndh	petC	petC
3	petA	cymA	ndh	petC	ndh	ccoO	ccoO	omcA
4	ndh	petB	ccoN	ndh	petC	petC	petB	ccoO
5	ccoO	napF	petB	ccoN	ccoN	shp	ndh	petB
6	ccoN	torA	napA	petB	petB	ccoN	ccoN	ndh
7	coxA	SO_4048	torA	napA	napA	cctA	petA	ccoN
8	coxC	dmsE	petC	coxC	omcA	omcA	coxC	petA
9	cymA	dhc	omcA	torA	nrfA	petB	coxA	ccoP
10	ccoP	napA	coxC	cytcB	torA	torA	ccoP	coxC

Table 2 The top 10 key proteins corresponding to an integrated analysis. Notes: * indicates that the times of the protein identified as a key protein by 8 different centrality measures in table 1; ** indicates that, for the same instance, the sum of rank number for the protein in table 1 are used.

RANK	ID	Protein	Description	Number* (Sum**)
1	SO_4591	cymA	membrane anchored tetraheme cytochrome c CymA	8 (18)
2	SO_0609	petB	ubiquinol-cytochrome c reductase cytochrome b subunit PetB	8 (40)
3	SO_2363	ccoO	Cbb3-type cytochrome c oxidase subunit II CcoO	7 (21)
4	SO_0610	petC	ubiquinol-cytochrome c reductase cytochrome c1 subunit PetC	7 (25)
5	SO_3517	ndh	respiratory NADH dehydrogenase II Ndh	7 (27)
6	SO_2364	ccoN	Cbb3-type cytochrome c oxidase subunit I CcoN	7 (39)
7	SO_1779	omcA	extracelllular iron oxide respiratory system surface decaheme cytochrome c component OmcA	5 (30)
8	SO_1232	torA	trimethylamine-N-oxide reductase TorA	5 (42)
9	SO_4609	coxC	aa3 type cytochrome c oxidase subunit III CoxC	5 (44)
10	SO_0848	napA	periplasmic nitrate reductase molybdopterin-binding subunit NapA	4

From network evolution perspective, the key positions of the network are taken by some c-type cytochromes which are associated with electricity production. This finding suggests that the ability of *Shewanella oneidensis* MR-1 to produce electricity might be derived from the unique structure of the protein interaction network of c-type cytochromes in the bacteria.

3.3 Modularity analysis

It is difficult to deduce potential EET pathways from the whole protein interaction network directly, since the combination explosion made that there are thousands of routes from a c-type cytochrome to another one. Small functional modules which might contain proteins participating in similar biological

Molecular BioSystems Accepted Manuscript

This journal is © The Royal Society of Chemistry [year]

processes would be helpful to deduce EET pathways, we thus need to obtain such small units by decomposing the network to several partitions.

Furthermore, the rules for deducing potential EET pathways are also needed. Unlike the classical MtrCAB pathway which moving electrons outside the cells, Bose *et al*. reported that *Rhodopseudomonas palustris* TIE-1 could take electrons in the cells via EET, the related protein system for electron transfer is encoded by the PioABC operon.[35] Despite the reversing electron transfer direction, there are 2 points shared by MtrCAB system and PioABC system.

First of all, genes that constituted an operon are often in the same functional pathway as mentioned in section 2.5, and the organization of genes in operons is believed to provide the advantage of coordinated regulation and production of functionally related genes. The proteins encoded by these genes might naturally form multi-protein complexes for promoting electron transfer, which composed the stem of the electron transfer chain, just like MtrCAB system and PioABC system here. Thus, to deduce potential EET pathways, we need identify the gene clusters controlled by the same operon firstly. The second, to form the pathways for EET, they should be located in diversiform cellular compartments. Typically, they should transfer electron from inner membrane, spanning periplasm, and getting out of outer membrane. For example, in MtrCAB pathway, inner membrane cymA and outer membrane omcA served as the source and the sink for electrons gaining and losing, they are linked by the stem which are composed by several periplasm and outer membrane c-type cytochromes encoded by the gene clusters mtrA, mtrB and mtrC. These 2 points form the basic for the identification of potential EET pathways in our work. Furthermore, verified experimental data could also be used to supply these EET pathways.

We obtained the decomposed results of the protein interaction network of c-type cytochromes in *Shewanella oneidensis* MR-1 based on simulated annealing algorithm.[17-18] The results showed that the network is decomposed to 5 modules (figure 2). Naturally, we expected that potential EET pathways could be found from these modules. By subcellular localization analysis, we found that the c-type cytochromes and their associated proteins in these modules located in diversiform cellular compartments, including cytoplasm, inner membrane, periplasm and outer membrane (table 3), which reflects their potential to form EET pathways that should be formed from inner membrane through periplasm to outer membrane. Since EET pathways may also collaborate with each other, the interactions between modules (termed cross-talk [36-38]) might indicate such cooperation among different EET pathways from different modules. Hence, theses cross-talk interactions are expected to enable further potential for the transmission of electrons. For example, there are 3 cross-talk interactions (sorA pp torA, sirA pp torA, and sorA pp napA) between module 1 and module 4 (yellow module and red module in figure 2, see also table 3 for module number). Considering the protein subcellular localization, these cross-talk interactions might facilitate electron transfer from cytoplasm (sorA) and inner membrane (sirA) to periplasm (torA and napA).

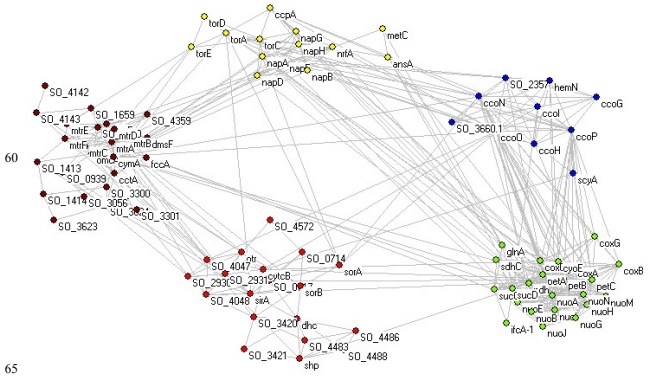


Fig. 2 Modules in the protein interaction network of c-type cytochromes in *Shewanella oneidensis* MR-1. Modules are shown in distinct colors.

Table 3 The protein subcellular localization information for the 5 modules. Proteins are located in diversiform cellular compartments (including cytoplasm, inner membrane, periplasm and outer membrane), which reflects their potential to form EET pathways that should be formed from inner membrane through periplasm to outer membrane.

Module	Cytoplasm	Inner Membrane	Periplasm	Outer Membrane
1	ansA,metC,napF,torD	napH,torC,torE	ccpA,napA,napB,napD,napG,nrfA,torA	——
2	hemN, ccoO	ccoN, ccoP, ccoI, SO_2357, SO_3660.1, ccoG	scyA	ccoH
3	glnA, nuoE, nuoI, sucC, sucD	coxA, coxB, coxC, cyoE, ndh, nuoA, nuoB, nuoH, nuoJ, nuoM, nuoN, petA, petB, petC, sdhC	ifcA	coxG, nuoG
4	sorA,SO_4486	sirA,SO_3421,SO_4483,SO_4488	cytcB,otr,SO_0714,SO_0717,SO_3420,SO_4047,SO_4048	dhc,shp,sorB,SO_2930,SO_2931,SO_4572
5	SO_3624	cymA	cctA, dmsE, fccA, mtrA, mtrD, SO_1413, SO_1414, SO_1659, SO_3056, SO_3300, SO_3301, SO_3623, SO_4142, SO_4360	dmsF, mtrB, mtrC, mtrE, mtrF, omcA, SO_0939, SO_4143, SO_4359

Collective observations were very interesting, we found that five primary MtrCAB-associated c-type cytochromes (omcA, mtrA, mtrB, mtrC and cymA suggested elsewhere [4, 39-40]) are all in module 5 (brown module in figure 2, see also table 3). So the emphasis was mainly conducted on this Mtr-like module, there are 25 proteins and 86 interactions among them.

Molecular BioSystems Accepted Manuscript

To deduce potential EET pathways from the module, we first identified the gene clusters controlled by the same operon, and got 8 gene clusters (table 4). Then, the information on subcellular localization was used to confirm that how many gene clusters could form the pathways for EET. We found that the proteins encoded by gene cluster SO_1413:SO_1414 (as well as the gene cluster SO_3300:SO_3301) are all located in the periplasm, that is, the proteins encoded by gene cluster SO_1413:SO_1414 and SO_3300:SO_3301 can not be used to transfer electrons to (and out of) the outer membrane, the proteins encoded by gene cluster SO_3623:SO_3624 spans cytoplasm and periplasm but they can not use any inner membrane cytochrome, gene cluster SO_4142:SO_4143 also can not use inner membrane cytochrome (see purple frames in figure 3), so these 4 gene clusters are excluded. Furthermore, experimental data has indicated that mtrA and cymA are bridged by fccA or cctA (see section 3.1), which has also been considered. That is, the inner membrane cymA get the electrons from substrate oxidation, and then bifurcate the electron transport chain towards periplasmic c-type cytochromes fccA or cctA, the bifurcated electron transport chain could be again congregated at mtrA, and which will be formed an electron transfer complex with the outer membrane localized MtrB and the terminal ferric reductases MtrC and OmcA in turn.

Table 4 Gene clusters with the same operon in the Mtr-like module.

Gene Clusters	Gene ID 1	Gene Name 1	Gene ID 2	Gene Name 2	Gene ID 3	Gene Name 3
1	SO_1413	SO_1413	SO_1414	SO_1414	—	—
2	SO_1427	dmsE	SO_1428	dmsF	—	—
3	SO_1776	mtrB	SO_1777	mtrA	SO_1778	mtrC
4	SO_1780	mtrF	SO_1781	mtrE	SO_1782	mtrD
5	SO_3300	SO_3300	SO_3301	SO_3301	—	—
6	SO_3623	SO_3623	SO_3624	SO_3624	—	—
7	SO_4142	SO_4142	SO_4143	SO_4143	—	—
8	SO_4359	SO_4359	SO_4360	SO_4360	—	—

In total, 5 possible EET pathways were obtained from this Mtr-like module (figure 3), which are similar to other studies for predicting EET pathways based on the homologues [41] or functional complexes [42] elsewhere. Among these 5 EET pathways, three integral metal reduction pathways (figure 3, red line) which were well-studied elsewhere are included: *(1)* ***MtrCAB pathway 1:*** *cymA – cctA – mtrA – mtrB – mtrC – omcA*, *(2)* ***MtrCAB pathway 2:*** *cymA – fccA – mtrA – mtrB – mtrC – omcA*; [4, 24, 41-44] and (3) ***MtrDEF pathway:*** *cymA – mtrD – mtrE – mtrF – omcA*. [42, 45-46]

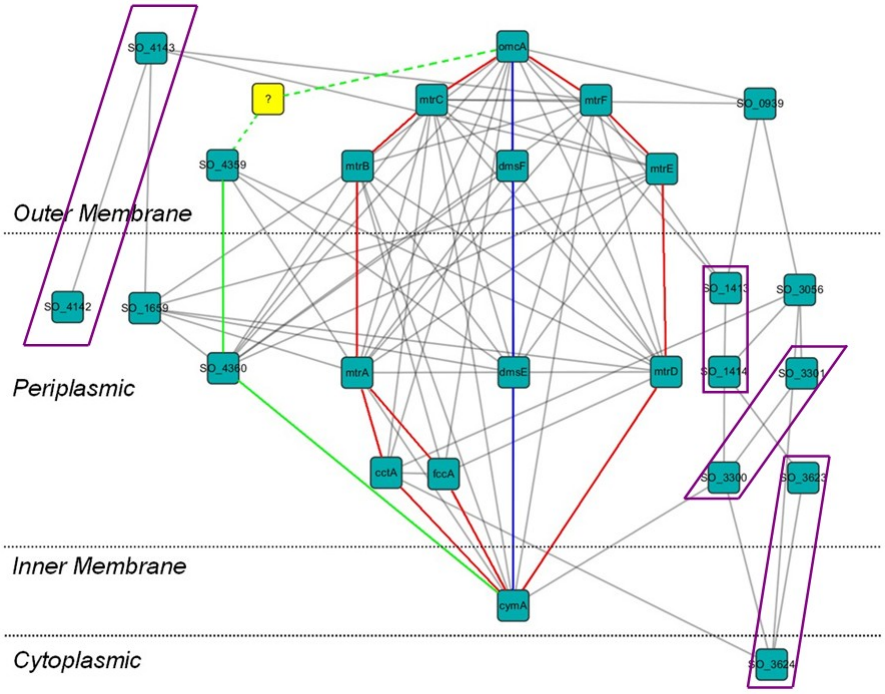


Fig. 3 5 EET pathways which were obtained from the Mtr-like module. Red lines indicate classical MtrCAB (1, 2) and MtrDEF pathways, blue lines indicate the DMSO pathway, and green lines indicate a new candidate pathway, "?" in green lines indicates that there might be a missing protein in this pathway (see the text). Purple frames indicate gene clusters coding for c-type cytochromes with the same operon but can not form EET pathways in this Mtr-like module.

Molecular BioSystems Accepted Manuscript

This journal is © The Royal Society of Chemistry [year]

(4) ***DMSO pathway:*** *cymA – dmsE – dmsF – omcA* (figure 3, blue line). It was suggested that dmsE do not exhibit significant activity and dmsF can not complement mtrB for iron(III) reduction, but dmsE and dmsF play an important role in DMSO reduction, and a model for DMSO pathway is then proposed: *cymA – dmsE – dmsF – dmsA/B.*[41-42] Through gene expression profiles analysis (figure 4), we found that when the electron acceptor is the DMSO, the expression of dmsE and dmsF are at a very high-level, which indicates that this EET pathway would be turned on at the situation. Here, the 4th pathway is similar but subtle difference to other works (e.g., Coursolle and Gralnick [41-42]), that is, cymA transfer electrons to periplasmic c-type cytochromes dmsE, and via outer membrane DMSO terminal reductase dmsF, ulteriorly transfer electrons to omcA rather than dmsA/B complexes. It indicates that DMSO pathway might have some new alternative processes with new situations, which should be actually existent. For example, Cheng *et al*. have demonstrated that DMSO could always stimulate electron acceptors (such as hydrous ferric oxide, HFO) reduction through metabolic and genetic regulation in *Shewanella oneidensis* MR-1, with or without dmsE.[47]

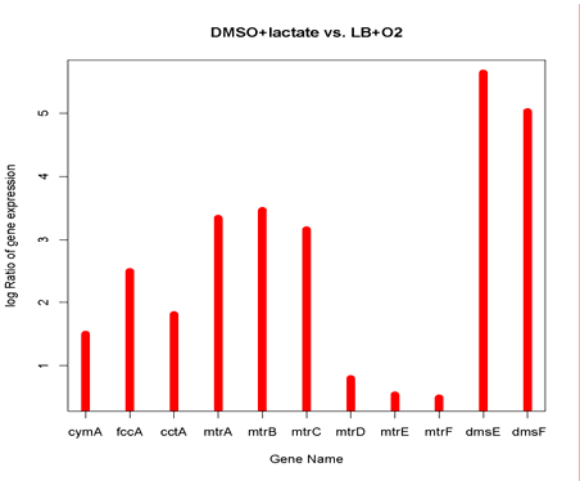


Fig. 4 The comparison of gene expression level between DMSO+Lactate (Treatment) and LB+O_2 (Control) for the main c-type cytochromes in Mtr-like module, the $\log_2$ ratio comparison data are downloaded from MicrobesOnline (http://www.microbesonline.org/cgi-bin/microarray/viewExp.cgi?expId=2426).

It also should be noted that the expression of mtrA, mtrB and mtrC are also at a high-level, but the mtrD, mtrE and mtrF seem to have a low-level expression at the situation, which indicates that the DMSO pathway would compete with MtrDEF pathway at some situations. In other words, the DMSO pathway might have some similar function to the MtrDEF pathway, and thus been an alternative electron route to the MtrDEF pathway at some particular situations (e.g., DMSO as the electron acceptor here). This is consistent with gene expression profiles from *Shewanella oneidensis* MR-1 by 15 growth conditions: aerobic growth in LB broth, aerobic growth in defined minimal media with 8 different carbon sources, aerobic growth in defined lactate medium with 4 different stresses and anaerobic growth in a defined medium with 2 different carbon sources. Figure 5 gives the heat-map summarizing of the gene expression profiles for 12 main c-type cytochromes in Mtr-like module under these 15 conditions, and it has shown that dmsE and dmsF have similar expression patterns with mtrD, mtrE and mtrF in most cases.

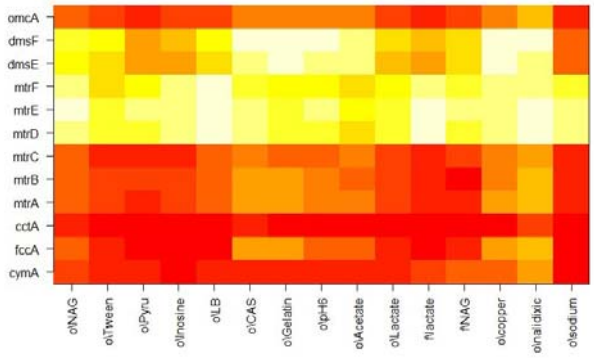


Fig. 5 Heat-map summarizing of the gene expression profiles for 12 main c-type cytochromes in Mtr-like module under 15 conditions with dataset GSE39462. The y axis shows the 12 c-type cytochromes coding genes, and the x axis shows the 15 growth conditions for *Shewanella oneidensis* MR-1: "o" means under aerobic condition using oxygen as eletron acceptor, "f" means under anaerobic condition using fumarate as the eletron acceptor. In both aerobic and anaerobic condition, *Shewanella oneidensis* MR-1 smaples were grown in different medium and various carbon sources. Expression levels of selected 12 genes were normalized to their color (yellow, low expression; red, high expression).

(5) ***New candidate pathway:*** *cymA – SO_4360 – SO_4359 – ? – omcA* (figure 3, green line). Even if the well-studied MtrCAB pathway also has new alternative processes with new situations (such as reversible as demonstrated by Ross *et al*. [48]), some new pathways which are unknown currently might also be existent, such as the 5th pathway here. Periplasmic decaheme cytochrome c SO_4360 is a homologous protein to mtrA (52% identity), dmsE (48%) and mtrD (46%), outer membrane SO_4359 is a homologous protein to dmsF (30%), mtrB (25%) and mtrE (24%), which might evolve from mtrA/mtrD/dmsE and mtrB/mtrE/dmsF with subtle different function, respectively. Furthermore, the subcellular localization of SO_4360 and SO_4359 are consistent to mtrA/mtrD/dmsE and mtrB/mtrE/dmsF, respectively. From homologous protein view, due to the similar identity between SO_4360 and SO_4359 to their paralogues, this new pathway might be branched from either MtrCAB, MtrDEF or DMSO pathway. This would be similar to Coursolle and Gralnick's work, in which they deduce that "SO_4359-62 may share overlapping functionality with DMSO reductases and MtrDEF with metal reductases".[42]

However, considering that there is no direct interaction between SO_4359 and omcA, we infered that there might be a missing protein whose subcellular localization and function is similar to mtrC or mtrF, and thus the pathway is more likely to be branched from MtrCAB or MtrDEF pathway. We assumed that the protein need be coordinated regulation with SO_4360 and SO_4359, and therefore we performed operon prediction for the genes which encoding these proteins, and found that there are 4 related genes in the same operon: SO_4357, SO_4358, SO_4361

Molecular BioSystems Accepted Manuscript

and SO_4362. Function analysis indicated that the proteins encoded by these genes are all extracellular oxidoreductase associated proteins. However, only SO_4358 has the similar size to mtrC and mtrF, which would make these proteins form the similar numbers of transmembrane domains (mtrC - 671aa, mtrF - 639aa; SO_4357 - 205aa, SO_4358 - 862aa, SO_4361 - 155aa, SO_4362 - 225aa). That is, the missing protein might be SO_4358. So the 5th pathway in Mtr-like module would more likely be "cymA – SO_4360 – SO_4359 – SO_4358 – omcA". This pathway would like to share some similar mechanism and function with mtrCAB or mtrDEF pathway, and thus facilitate electron transfer to outer membrane in some particular situations.

Modules detected by modularity measure function *M* have been proved of great help to deduce potential EET pathways. However, it should be noted that *M* may be limited by resolution limit problem, i.e., the optimization of *M* might fail to detect modules smaller than certain scales (e.g., $\sqrt{2L}$ or smaller) [49], which may be responsable for the situation that no clear GO processes or KEGG pathways could be related to these 5 modules in our experiment. New improved measures such as modularity density function *D* [50-51] which did not show the resolution limit problem might be helpful for mining these smaller modules.

4. Conclusions

Considering the topology of a network has been shown to strongly correlate with its function, and previous studies have also shown the high efficiency about the prediction of gene clusters and biological pathways with topological property,[52-55] from which we constructed a large-scale protein interaction network for all of 41 c-type cytochromes in *Shewanella oneidensis* MR-1 and studied its structural characteristics and functional significance in this work. Centralization analysis suggests that the ability of *Shewanella oneidensis* MR-1 to produce electricity might be derived from the unique structure of the protein interaction network of c-type cytochromes in the bacteria. Modularity analysis suggests that the proteins in these modules locate in diversiform cellular compartments, reflecting their potential to form EET pathways. Furthermore, the well-known and new candidate EET pathways are obtained based on protein subcellular localization and operon analysis, indicating that potential EET pathways could be identified from such a c-type cytochrome network.

Acknowledgements

We would like to thank the two anonymous reviewers for their valuable comments on this study. The work was supported by the National Natural Science Foundation of China (No. 61472078, No. 51172043), the Open Research Fund of State Key Laboratory of Bioelectronics from Southeast University, and the Natural Science Foundation of Anhui Education Department (KJ2013B167).

References

1 T. Mehta, M. V. Coppi, S. E. Childers and D. R. Lovley, *Appl. Environ. Microbiol.*, 2005, **71**, 8634-8641.
2 J. A. Gralnick and D. K. Newman, *Mol. Microbiol.*, 2007, **65**, 1-11.
3 L. Shi, T. C. Squier, J. M. Zachara and J. K. Fredrickson, *Mol. Microbiol.*, 2007, **65**, 12-20.
4 D. Coursolle, D. B. Baron, D. R. Bond and J. A. Gralnick, *J. Bacteriol.*, 2010, **192**, 467-474.
5 L. Peng, S. J. You and J. Y. Wang, *Biosens. Bioelectron.*, 2010, **25**, 2530-2533.
6 T. E. Meyer, A. I. Tsapin, I. Vandenberghe, L. de Smet, D. Frishman, K. H. Nealson, M. A. Cusanovich and J. J. van Beeumen, *OMICS*, 2004, **8**, 57-77.
7 L. Shi, D. J. Richardson, Z. Wang, S. N. Kerisit, K. M. Rosso, J. M. Zachara and J. K. Fredrickson, *Environ. Microbiol. Rep.*, 2009, **1**, 220-227.
8 S. J. Marritt, T. G. Lowe, J. Bye, D. G. McMillan, L. Shi, J. Fredrickson, J. Zachara, D. J. Richardson, M. R. Cheesman, L. J. Jeuken and J. N. Butt, *Biochem. J.*, 2012, **444**, 465-474.
9 B. Schuetz, M. Schicklberger, J. Kuermann, A. M. Spormann and J. Gescher, *Appl. Environ. Microb.*, 2009, **75**, 7789-7796.
10 H. Zhang, X. Tang, G. R. Munske, N. Zakharova, L. Yang, C. Zheng, M. A. Wolff, N. Tolic, G. A. Anderson, L. Shi, M. J. Marshall, J. K. Fredrickson and J. E. Bruce, *J. Proteome. Res.*, 2008, **7**, 1712-1720.
11 J. F. Heidelberg, I. T. Paulsen, K. E. Nelson, E. J. Gaidos, W. C. Nelson, T. D. Read, J. A. Eisen, R. Seshadri, N. Ward, B. Methe, R. A. Clayton, T. Meyer, A. Tsapin, J. Scott, M. Beanan, L. Brinkac, S. Daugherty, R. T. DeBoy, R. J. Dodson, A. S. Durkin, D. H. Haft, J. F. Kolonay, R. Madupu, J. D. Peterson, L. A. Umayam, O. White, A. M. Wolf, J. Vamathevan, J. Weidman, M. Impraim, K. Lee, K. Berry, C. Lee, J. Mueller, H. Khouri, J. Gill, T. R. Utterback, L. A. McDonald, T. V. Feldblyum, H. O. Smith, J. C. Venter, K. H. Nealson and C. M. Fraser, *Nat. Biotechnol.*, 2002, **20**, 1118-1123.
12 S. Pierre and K. Scholich, *Mol. BioSyst.*, 2010, **6**, 641-647.
13 A. R. Kristensen and L. J. Foster, *Mol. BioSyst.*, 2013, **9**, 2201-2212.
14 A. Franceschini, D. Szklarczyk, S. Frankild, M. Kuhn, M. Simonovic, A. Roth, J. Lin, P. Minguez, P. Bork, C. von Mering and L. J. Jensen, *Nucleic Acids Res.*, 2013, **41**, D808-D815.
15 B. H. Junker, D. Koschutzki and F. Schreiber, *BMC Bioinformatics*, 2006, **7**, 219.
16 M. E. Newman and M. Girvan, *Phys. Rev. E*, 2004, **69**, 026113.
17 R. Guimera and L. A. N. Amaral, *Nature*, 2005, **433**, 895-900.
18 R. Guimera and L. A. N. Amaral, *J. Stat. Mech.-Theory. Exp.*, 2005, P02001.
19 N. Y. Yu, J. R. Wagner, M. R. Laird, G. Melli, S. Rey, R. Lo, P. Dao, S. C. Sahinalp, M. Ester, L. J. Foster and F. S. Brinkman, *Bioinformatics*, 2010, **26**, 1608-1615.
20 M. N. Price, K. H. Huang, E. J. Alm and A. P. Arkin, *Nucleic Acids Res.*, 2005, **33**, 880-892.
21 M. N. Price, A. M. Deutschbauer, J. M. Skerker, K. M. Wetmore, T. Ruths, J. S. Mar, J. V. Kuehl, W. Shao and A. P. Arkin, *Mol. Syst. Biol.*, 2013, **9**, 660.
22 D. Szklarczyk, A. Franceschini, M. Kuhn, M. Simonovic, A. Roth, P. Minguez, T. Doerks, M. Stark, J. Muller, P. Bork, L. J. Jensen and C. von Mering, *Nucleic Acids Res.*, 2011, **39**, D561-D568.
23 L. J. Jensen, M. Kuhn, M. Stark, S. Chaffron, C. Creevey, J. Muller, T. Doerks, P. Julien, A. Roth, M. Simonovic, P. Bork and C. von Mering, *Nucleic Acids Res.*, 2009, **37**, D412-D416.
24 B. M. Fonseca, C. M. Paquete, S. E. Neto, I. Pacheco, C. M. Soares and R. O. Louro, *Biochem. J.*, 2013, **449**, 101-108.
25 L. Shi, B. Chen, Z. Wang, D. A. Elias, M. U. Mayer, Y. A. Gorby, S. Ni, B. H. Lower, D. W. Kennedy, D. S. Wunschel, H. M. Mottaz, M. J. Marshall, E. A. Hill, A. S. Beliaev, J. M. Zachara, J. K. Fredrickson and T. C. Squier, *J. Bacteriol.*, 2006, **188**, 4705-4714.
26 X. Tang, W. Yi, G. R. Munske, D. P. Adhikari, N. L. Zakharova and J. E. Bruce, *J. Proteome Res.*, 2007, **6**, 724-734.
27 D. E. Ross, S. S. Ruebush, S. L. Brantley, R. S. Hartshorne, T. A. Clarke, D. J. Richardson and M. Tien, *Appl. Environ. Microbiol.*, 2007, **73**, 5797-5808.
28 J. Borloo, L. Desmet, J. V. Beeumen and B. Devreese, *Omics*, 2011, **1**, 260-267.
29 M. E. Smoot, K. Ono, J. Ruscheinski, P. L. Wang and T. Ideker, *Bioinformatics*, 2011, **27**, 431-432.
30 S. Cheng, B. Lian, J. Liang, T. Shi, L. Xie and Y. L. Zhao, *Mol. BioSyst.*, 2013, **9**, 2860-2868.
31 E. Marsili, D. B. Baron, I. D. Shikhare, D. Coursolle, J. A. Gralnick and D. R. Bond, *PNAS*, 2008, **105**, 3968-3973.

Molecular BioSystems Accepted Manuscript

This journal is © The Royal Society of Chemistry [year]

32 J. W. Voordeckers, B. C. Kim, M. Izallalen and D. R. Lovley, *Appl. Environ. Microbiol.*, 2010, **76**, 2371-2375.
33 H. Gao, S. Barua, Y. Liang, L. Wu, Y. Dong, S. Reed, J. Chen, D. Culley, D. Kennedy, Y. Yang, Z. He, K. H. Nealson, J. K. Fredrickson, J. M. Tiedje, M. Romine and J. Zhou, *Microb. Biotechnol.*, 2010, **3**, 455-466.
34 C. Baraquet, L. Theraulaz, C. Iobbi-Nivol, V. Mejean and C. Jourlin-Castelli, *Mol. Microbiol.*, 2009, **73**, 278-290.
35 A. Bose, E. J. Gardel, C. Vidoudez, E. A. Parra and P. R. Girguis, *Nat. Commun.*, 2014, **5**, 3391.
36 R. Donaldson and M. Calder, *Theoretical Computer Science*, 2012, **456**, 30-50.
37 S. Proietti, L. Bertini, A. M. Timperio, L. Zolla, C. Caporale and C. Caruso, *Mol. BioSyst.*, 2013, **9**, 1169-1187.
38 W. Li, L. N. Chen, X. Li, X. Jia, C. C. Feng, L. C. Zhang, Wei. M. He, J. J. Lv, Y. H. He, W. G. Li, X. L. Qu, Y. Y. Zhou and Y. C. Shi, *Mol. BioSyst.*, 2013, **9**, 3026-3035.
39 L. Shi, K. M. Rosso, J. M. Zachara and J. K. Fredrickson, *Biochem. Soc. Trans.*, 2012, **40**, 1261-1267.
40 L. Shi, K. M. Rosso, T. A. Clarke, D. J. Richardson, J. M. Zachara and J. K. Fredrickson, *Front Microbiol.*, 2012, **3**, 50.
41 D. Coursolle and J. A. Gralnick, *Mol. Microbiol.*, 2010, **77**, 995-1008.
42 D. Coursolle and J. A. Gralnick, *Front Microbiol.*, 2012, **3**, 56.
43 R. S. Hartshorne, C. L. Reardon, D. Ross, J. Nuester, T. A. Clarke, A. J. Gates, P. C. Mills, J. K. Fredrickson, J. M. Zachara, L. Shi, A. S. Beliaev, M. J. Marshall, M. Tien, S. Brantley, J. N. Butt and D. J. Richardson, *PNAS*, 2009, **106**, 22169-22174.
44 M. A. Firer-Sherwood, K. D. Bewley, J. Y. Mock and S. J. Elliott, *Metallomics*, 2011, **3**, 344-348.
45 M. Breuer, P. Zarzycki, J. Blumberger and K. M. Rosso, *JACS*, 2012, **134**, 9868-9871.
46 M. Breuer, K. M. Rosso and J. Blumberger, *PNAS*, 2014, **111**, 611-616.
47 Y. Y. Cheng, B. B. Li, D. B. Li, J. J. Chen, W. W. Li, Z. H. Tong, C. Wu and H. Q. Yu, *PLoS ONE*, 2013, **8**, e78466.
48 D. E. Ross, J. M. Flynn, D. B. Baron, J. A. Gralnick and D. R. Bond, *PLoS ONE*, 2011, **6**, e16649.
49 S. Fortunato and M. Barthelemy, *PNAS*, 2007, **104**, 36-41.
50 Z. P. Li, S. H. Zhang, R. S. Wang, X. S. Zhang and L. N. Chen, *Physical Review E*, 2008, **77**, 036109.
51 X. S. Zhang, R. S. Wang, Y. Wang, J. G. Wang, Y. Q. Qiu, L. Wang and L. N. Chen, *Europhysics Letters*, 2009, **87**, 38002.
52 J. Planas-Iglesias, E. Guney, J Garcia-Garcia, K. A. Robertson, S. Raza, T. C. Freeman, P. Ghazal and B. Oliva, *OMICS*, 2012, **16**, 245-256.
53 M. Jiang, Y. Chen, Y. Zhang, L. Chen, N. Zhang, T. Huang, Y. D. Cai and X. Kong, *Mol. BioSyst.*, 2013, **9**, 2720-2728.
54 C. Huang, Q. Ba, Q. Yue, J. Li, J. Li, R. Chu and H. Wang, *Mol. BioSyst.*, 2013, **9**, 3091-3100.
55 A. Mukhopadhyay and U. Maulik, *PLoS ONE*, 2014, **9**, e94029.

Molecular BioSystems Accepted Manuscript