The First International Symposium on Optimization and Systems Biology (OSB'07)
Beijing, China, August 8–10, 2007
, pp. 100–109

New Insights into Network Motif Clusters from the Views of Cellular Localizations and Signal Pathways*

Guangxu Jin[1,2] Shihua Zhang[1,2] Xiang-Sun Zhang[1,†]
Luonan Chen[3,4,5*]

[1] Academy of Mathematics and Systems Science
Chinese Academy of Sciences, Beijing 100080, China
[2] Graduate University of Chinese Academy of Sciences, Beijing 100049, China
[3] Department of Electrical Engineering and Electronics,
Osaka Sangyo University, Osaka 574-8530, Japan
[4] ERATO Aihara Complexity Modelling Project, JST, Tokyo 153-8505, Japan
[5] Institute of Industrial Science, The University of Tokyo, Tokyo 153-8505, Japan

Abstract Biological complex networks, such as protein-protein interaction network (PPIN), represent the underlying living mechanisms in cell. Begin with the PPIN, distinct level modular organizations appear, like so-called network motifs, protein complexes and signal pathways. The network motifs defined as spandrels are highly clustered. Here, the motif clusters are divided into two categories, i.e. inner motif clusters (IMC) and outer motif clusters (OMC), based on protein complexes. IMC contains more network motifs inside some protein complex contrasting to OMC. In topological structure, the IMCs are connected more loosely than OMCs and in biology roles the IMCs mainly located in nucleus and serve for the signal pathways closely related to nucleus while OMCs mainly located in cytoplasm and reach for the signal pathways unrelated to nucleus, moreover, in contrast to OMCs, IMCs can reach less cellular localizations. Therefore, from the protein-protein interaction level, subsets of proteins that define functionally meaningful entities can reveal the underlying living mechanisms in cell.

Keywords Protein-protein interactome; hub; network motif; signal pathway; systems biology.

1 Introduction

Modular organization pervades biological complexity [1-3]. Complex biological networks show modularity at distinct scales, from small sets of genes or proteins controlling the signals to large groups of proteins involved in transcriptional of genes in nucleus or translation of mRNAs outside nucleus. At the relative small scale, small sets of interacting proteins or genes, so-called 'network motifs' [3-9], have been revealed as the functional building blocks of biology networks. At the relative large

*This work was partially supported by the National Natural Science Foundation of China under grant No.10631070, and the Ministry of Science and Technology, China, under grant No.2006CB503905.
†corresponding author: zxs@amt.ac.cn, chen@eic.osaka-sandai.ac.jp

scale, large groups of proteins, such as protein complexes or signal pathways [1, 10-17], can perform relatively independent functions.

Sole and Valverde have just suggested that network motifs are the spandrels of cellular complexity [3], where the spandrels are defined as the tapering triangular spaces formed, by the intersection of two rounded arches at right angles –are necessary architectural by-products of mounting a dome on rounded arches [18]. For network motifs, regardless of either building blocks or spandrels for biology networks, the architectural components construct the main structure of the networks. Without them, the biology networks will be destroyed into fragments and not be connected tightly any longer (see Figure 2).

Most network motifs are highly clustered, such as spandrels [3]. Here, we found that there are two types of network motif clusters, i.e. inner motif clusters (IMC) and outer motif clusters (OMC), controlling the network complexity in different manners due to whether most network motifs in motif cluster are located in the same protein complex. IMCs are mainly contained inside the protein complexes and OMCs are almost spread outside the protein complexes. Deletion of IMCs can't lead to the ruin of biology network while deletion of OMCs can destroy the network structure.

Many cell functions are carried out by subsets of units that define functionally meaningful entities[19]. In this context, the insights into the intrinsic relation between cell functions and subsets of biology networks provide us a new idea that the topological structure of a cell's network may represent the underlying principles of a cell's living mechanism. In this paper, identification of the relation between the motifs clusters and their roles in representing the underlying principles in a cell has been becoming the study focus for us.

In this paper, we not only found the distinction between IMCs and OMCs, but also found they represent different underlying principles in a cell. First, in the viewpoint of cellular localizations, contrasting to OMCs, more proteins in IMCs are located in nucleus and less proteins in IMCs are located in cytoplasm. Second, in the viewpoint of signal pathways, in contrast to OMCs, more IMCs are involved in the signal pathways in nucleus. Therefore, the different topological structures of IMCs and OMCs indicate their different roles in cell functions. The loosely connections of IMCs contrasting to OMCs determined they mainly located in nucleus and serve for the signal pathways in nucleus.

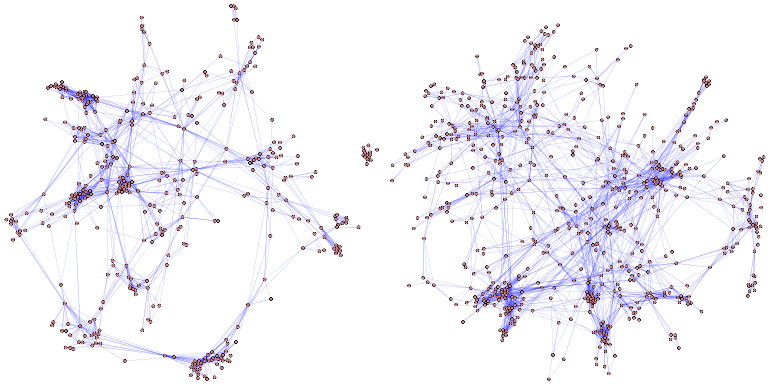
2 Results/Discussion

2.1 The topological structure of IMCs and OMCs

2.1.1 IMCs are connected relative loosely and OMCs are connected relative tightly

As the definitions of IMCs and OMCs, IMCs contain more network motifs inside the protein complexes while OMCs contain more network motifs outside the protein complexes. In Figure 1, it is easy to see that the IMCs in Figure 1a are connected loosely while the OMCs in Figure 1b are connected tightly. The reason is

that few network motifs in IMCs act as the connectors among the IMCs while many network motifs in OMCs play the roles in connecting the OMCs.



(a) (b)

Figure 1: The topological structures of IMCs and OMCs. (a) The connections between IMCs are sparse and IMCs are connected loosely. (b) The connections between OMCs are tight and OMCs are connected tightly.

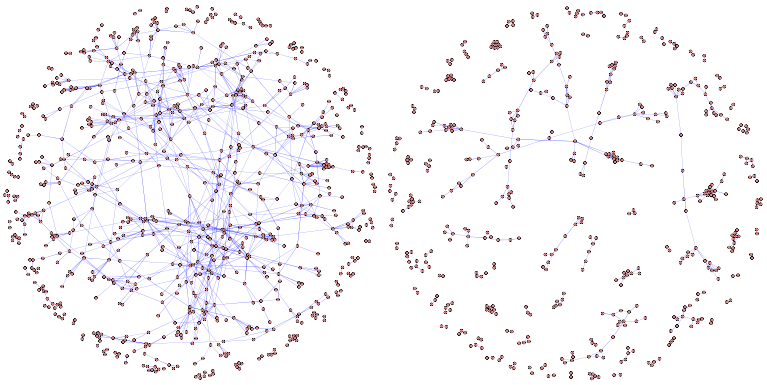
2.1.2 The partition of motif clusters is not random

To show the p-value of the partition of motif clusters, we performed a numerical experiment to test the distinct effects of deletion of IMCs and OMCs respectively on the original network. First, we deleted the IMCs and OMCs from the original network respectively and get the remained network of IMCs and the remained network of OMCs respectively (see Figure 2). Next, we divided the motif clusters into two equal groups randomly and repeated the same experiment as the first step 1000 times. We found the p-value of the partition of motif clusters, i.e. IMCs and OMCs, are less than 0.001 (see Figure 3). Therefore, the partition of motif clusters are significant due to it is based on the protein complexes.

2.2 The biological roles of IMCs and OMCs

2.2.1 The distinct preferences for cellular localizations of IMCs and OMCs

In contrast to OMCs, more proteins in IMCs are located in nucleus and fewer proteins in IMCs are located in cytoplasm. Due to the distinction between the topological structures of IMCs and OMCs, i.e. IMCs are connected loosely and OMCs are connected tightly, they are located in different cellular localizations in cell. The proteins in motif clusters can be classified into three groups according to their cellular localizations, i.e. the proteins in nucleus, the proteins in cytoplasm and the proteins located in both nucleus and cytoplasm. By numerical experiments, we found that more proteins in IMCs are located in nucleus and fewer proteins are located in



(a) (b)

Figure 2: The remain network derived from deletion of IMCs and OMCs respectively. (a) The remain network is not ruined and the main component is still connected. (b) The network are break into fragments and the main component is disappear.

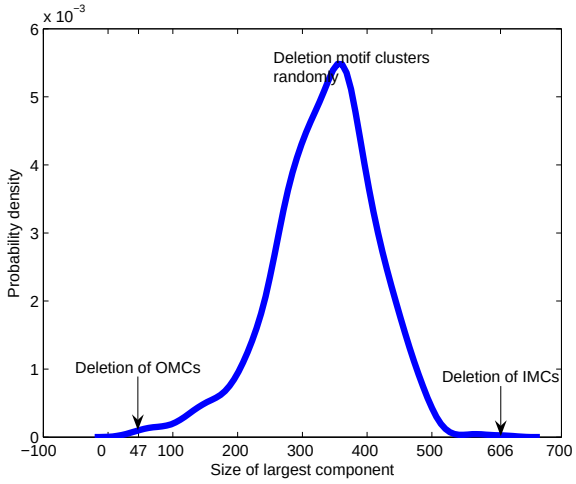
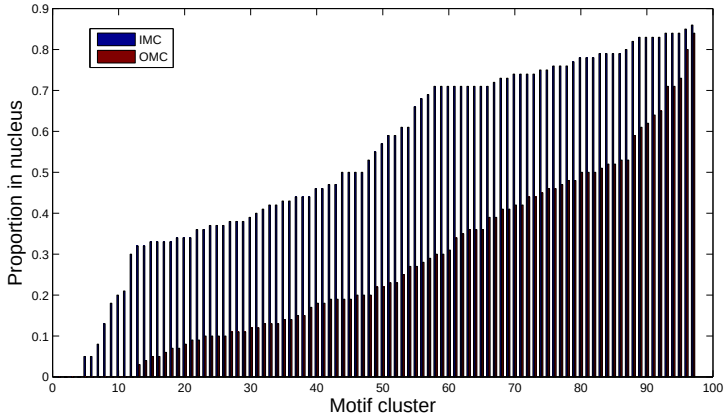
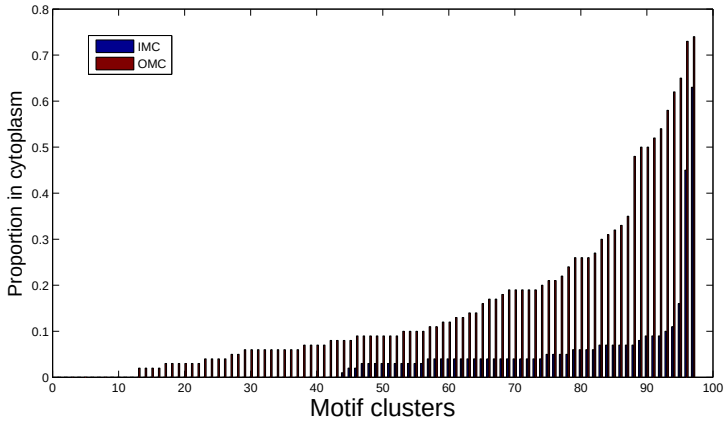


Figure 3: To test the difference between the topological structures of IMCs and OMCs, we randomly delete the equal number, i.e. 97 network motif clusters, from the original network 1000 times. The probability density curve is shown in the figure. It is easy to see that both the p-value of deletion of IMCs and the p-value of deletion of OMCs are less than 0.001. The result suggest there are significantly distinct topological structures between IMCs and OMCs.

cytoplasm contrasting to OMCs. In Figure 4a, the proportions of proteins located in nucleus in IMCs are significantly higher than ones in OMCs ($P < 10^{-13}$, Wilxocon rank sum test).



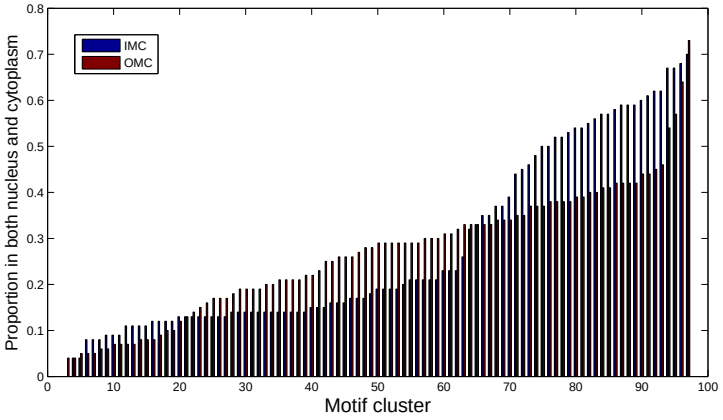
(a) More proteins in IMCs are located in nucleus in contrast to OMCs (P-value: $P < 10^{-13}$, Wilcoxon rank sum test).



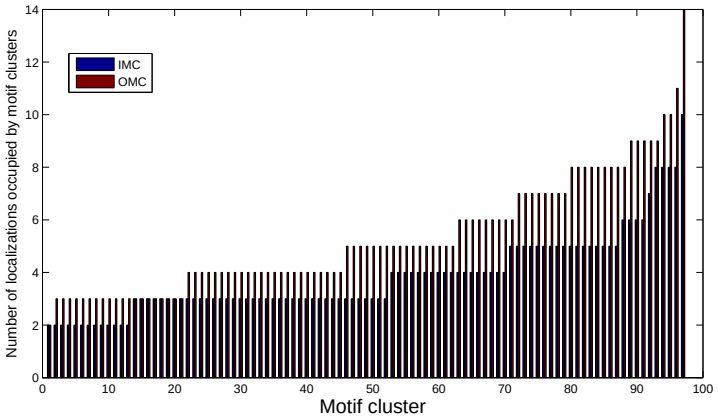
(b) Fewer proteins in IMCs are located in cytoplasm contrasting to OMCs (P-value: $P < 10^{-7}$, Wilcoxon rank sum test).

Figure 4: The distinct biological roles of IMCs and OMCs in cellular localizations.

In Figure 4b, the proportions of the proteins located in cytoplasm in IMCs are significantly lower than ones in OMCs ($P < 10^{-7}$, Wilxocon rank sum test). Additionally, in Figure 4c, for the proteins in both nucleus and cytoplasm, there are no significant difference between IMCs and OMCs ($P = 0.6821$, Wilxocon rank sum test). Thus, we can conclude that IMCs and OMCs prefer to different cellular localizations.



(c) There are no significant distinction between the IMCs and the OMCs, which can locate in both nucleus and cytoplasm (P-value: $P = 0.6821$, Wilxocon rank sum test).



(d) The number of cellular localizations that IMCs located in is significantly distinct from the one of OMCs (P-value: $P < 10^{-6}$, Wilcoxon rank sum test).

Figure 4: The distinct biological roles of IMCs and OMCs in cellular localizations.

In contrast to OMCs, the proteins in IMCs can reach less cellular localizations. Besides the cellular localizations, i.e. nucleus and cytoplasm, there are still other cellular localizations in cell. By numerical analysis, we found the proteins in IMCs can reach relatively less cellular localizations than ones in OMCs. In figure 4d, the numbers of cellular localizations for IMCs are significantly lower than the ones for OMCs ($P < 10^{-6}$, Wilxocon rank sum test).

The results suggest that one property in protein-protein interaction network, such as motif clusters based on protein complexes, reflect one underlying principles

in cell's living mechanism.

2.2.2 The distinct preferences for signal pathways of IMCs and OMCs

In consideration of the difference in preference of cellular localizations for IMCs and OMCs, we made further studies on the preferences for signal pathways for them. In a cell, nucleus is an essential part of it that regulates all its activity and be critical to its normal living. Replication of DNA, transcription of DNA and the process of cell cycle all happen there. Surely, the signal pathways involved in nucleus has special meanings for cell, such as basal transcription factors, RNA polymerase, purine metabolism, pyrimidine metabolism, proteasome and cell cycle. In turn, we consider a connection model in which the motif clusters and the pathways are connected by at least one protein. By such a connection model, we found the IMCs can reach more signal pathways involved in nucleus while the OMCs can connected with more signal pathways involved in other cellular localizations ($P < 10^{-2}$, Chi-square test). The result is in Table 1.

Table 1: The different preference for pathways between IMCs and OMCs.

	Pathways involved in nucleus	Pathways involved outside nucleus
IMCs	90	62
OMCs	57	82

From the analysis of the paper, we can firmly conclude that the topological distinction between IMCs and OMCs reflects their different biological roles in cell.

3 Materials and Methods

Protein-protein interaction data

The HCfyi dataset of 3976 interactions among 1291 proteins was obtained from Batada NN et al [20]. It is based on an intersection method in which only the interactions observed at least twice are retained from various datasets. Dataset of the HCfyi was derived from all extent protein interaction datasets, which include all LC interaction data. Especially, the LC data were manually curated from over 31,793 abstracts and online publications [21], and there is no interaction derived from standard large-scale experiments in FYI.

Other data

The protein complex data were derived from MIPS [13]. Cellular localization were derived from Huh et al. [22], The signal pathway data were derived from KEGG [23] .

Finding network motifs

A three-motif (i.e. the motif is composed of three proteins and a four-motif (i.e. the motif is composed of four proteins) were found by mfinder2.1 [4, 5]. The

ID of the three-motif is 238 whose Z-score is 317.43. The ID of the four-motif is 13260 whose Z-score is 12.90. Indeed, there are other four motifs that were found by mfinder2.1. The reason why we choose motif 13260 is that others are either a combination of two three-motifs or a combination of one three-motif and a single edge.

Motif clusters (IMCs and OMCs)

In the subgraph composed of network motifs, we select the top 20% proteins that interact with more other proteins[24-28] (The cutoffs between 10% and 40% have no significant effect on the result). Then one protein with their network motifs, i.e. the network motifs all contain the protein are defined as a network motifs. All together there are 194 network motif clusters. For the network motif clusters, due to the proportion of network motifs whose proteins are all located in the same protein complex, we divided them into two equal groups, i.e. IMCs and OMCs.

Deletion of Motif clusters

Deletion of motif clusters from the network is deleting the proteins in motif clusters from the network [24, 29, 30].

Connection Model

To value the relation between a network motif cluster and a pathway, in the connection model, the network motif cluster and the pathway has at least one common protein.

Acknowledgments

We are grateful to Nizar N. Batada for providing us the data of HCfyi network.

References

[1] Hartwell LH, Hopfield JJ, Leibler S, Murray AW (1999) From molecular to modular cell biology. Nature 402: C47-C52.

[2] Alberts B (1998) The cell as a collection of protein machines: Preparing the next generation of molecular biologists. Cell 92: 291-294.

[3] Sole RV, Valverde S (2006) Are network motifs the spandrels of cellular complexity? Trends Ecol Evol. 21(8):419-22.

[4] Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, et al. (2002) Network Motifs: Simple Building Blocks of Complex Networks. Science 298: 824-827.

[5] Milo R, Itzkovitz S, Kashtan N, Levitt R, Shen-Orr S, et al. (2004) Superfamilies of designed and evolved networks. Science 303: 1538-42.

[6] Shen-Orr S, Milo R, Mangan S, Alon U (2002) Network motifs in the transcriptional regulation network of Escherichia coli. Nat Genet 31: 64-68 .

[7] Mangan S, Itzkovitz S, Zaslaver A, Alon U (2006) The Incoherent Feed-forward Loop Accelerates the Response-time of the gal System of Escherichia coli. JMB 356: 1073-81.

[8] Mangan S, Zaslaver A, Alon U (2003) The Coherent Feedforward Loop Serves as a Sign-sensitive Delay Element in Transcription Networks. JMB 334/2: 197-204.

[9] Li C, Chen L, Aihara K (2007) A systems biology perspective on signal processing in genetic motifs. IEEE Signal Proc Mag 24: 136-147.

[10] Papin JA, Price ND, Wiback SJ, Fell DA, Palsson BO (2003) Metabolic pathways in the postgenome era. Trends Biochem Sci 28: 250-258.

[11] Krogan NJ, Cagney G, Yu H, Zhong G, Guo X, et al. (2006) Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. Nature 440: 637-644.

[12] Karp PD (2001) Pathway databases: a case study in computational symbolic theories. Science 293: 2040-2044.

[13] Mewes HW, Frishman D, Guldener U, Mannhaupt G, Mayer K, et al. (2002) MIPS: a database for genomes and protein sequences. Nucleic Acids Res 30: 31-34.

[14] Ho Y, Gruhler A, Heilbut A, Bader GD, Moore L, et al. (2002) Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. Nature 415: 180-183.

[15] Gavin AC, Bosche M, Krause R, Grandi P, Marzioch M, et al. (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature 415: 141-147.

[16] Gavin AC, Aloy P, Grandi P, Krause R, Boesche M, et al. (2006) Proteome survey reveals modularity of the yeast cell machinery. Nature 440: 631-637.

[17] Krogan NJ, Cagney G, Yu H, Zhong G, Guo X, et al. (2006) Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. Nature 440: 637-644.

[18] Gould SJ, Lewontin RC. (1979) The Spandrels of San Marco and the Panglossian Paradigm: a critique of the Adaptationist Programme. Proc. R. Soc. B 205, 581-598.

[19] Sole RV, Ferrer R, Montoya JM, Valverde S (2002) Selection, tinkering and emergence in complex networks. Complexity 8, 20-33

[20] Batada NN, Reguly T, Breitkreutz A, Boucher L, Breitkreutz BJ, et al (2006) Stratus not altocumulus: A new view of the yeast protein interaction network. PLoS Biol 4(10): e317.

[21] Reguly T, Breitkreutz A, Boucher L, Breitkreutz BJ, Hon GC, et al. (2006) Comprehensive curation and analysis of global interaction networks in Saccharomyces cerevisiae. J Biol 5: 11.

[22] Huh WK, Falvo JV, Gerke LC, Carroll AS, Howson RW, et al. (2003) Global analysis of protein localization in budding yeast. Nature 425: 686-691.

[23] Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita KF, Itoh M,et al. (2006) From genomics to chemical genomics: new developments in KEGG. Nucleic Acids Res 34: D354-7.

[24] Han JD, Bertin N, Hao T, Goldberg DS, Berriz GF, et al. (2004) Evidence for dynamically organized modularity in the yeast protein-protein interaction network. Nature 430: 88-93.

[25] Ekman D, Light S, Bjorklund AK, Elofsson A (2006) What properties characterize the hub proteins of the protein-protein interaction network of Saccharomyces cerevisiae? Genome Biol 7: R45.

[26] Maslov S, Sneppen K (2002) Specificity and stability in topology of protein networks. Science 296: 910-913.

[27] Giaever G, Chu AM, Ni L, Connelly C, Riles L, et al. (2002) Functional profiling of the Saccharomyces cerevisiae genome. Nature 418: 387-391.

[28] Barabasi AL, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. Nat Rev Genet 5(2):101-13.

[29] Jeong H, Mason SP, Barabasi AL, Oltvai ZN (2001) Lethality and centrality in protein networks. Nature 411: 41-42.

[30] Albert R, Jeong H, Barabasi AL (2000) Error and attack tolerance of complex networks. Nature 406: 378-382.

[31] Ge H, Liu Z, Church GM, Vidal M (2001) Correlation between transcriptome and interactome mapping data from Saccharomyces cerevisiae. Nature Genet 29: 482-486.

[32] Enright AJ, Ouzounis CA (2001) BioLayout — an automatic graph layout algorithm for similarity visualization. Bioinformatics 17(9): 853-854.

[33] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13: 2498-2504.

[34] Bernard R (1986) Fundamentals of Biostatistic. Boston Mass: Duxbury Press.

[35] Li Z, Zhang S, Wang Y, Zhang XS, Chen L (2007) Alignment of molecular networks by integer quadratic programming. Bioinformatics, in press.

[36] Zhang S, Jin G, Zhang XS, Chen L (2007) Discovering functions and revealing mechanisms at molecular level from biological networks. Proteomics, in press.