[Frontiers in Bioscience 10, 1070-1078, May 1, 2005]

INSIGHTS TO METABOLIC NETWORK EVOLUTION BY FUSION PROTEINS

Meena Kishore Sakharkar [1], Yu Yiting [1], Vincent T. K. Chow [2], Pandjassarame Kangueane [1]

[1] NCSV, MPE, Nanyang Technological University, Singapore 639798, [2] Department of Microbiology, NUS, Singapore 119260

TABLE OF CONTENTS

1. ABSTRACT

Human fusion proteins consisting of two or more fusion partners of prokaryotic origin exhibit accreted function. Recent studies have elucidated the importance of fusion proteins in complex regulatory networks. The significance of fusion proteins in cellular networks and their evolutionary mechanism is largely unknown. Here, we discuss the association of six fusion proteins with the citric acid cycle. We define possible gene fusion scenarios and show that they produce metabolites with high connectivity for complex networking. Complex networking of metabolites requires proteins with incremental structural architectures and functional capabilities. Such higher order functionality is frequently provided by fusion proteins. Therefore, evolution of fusion proteins capable of producing metabolites with greater connectivity for enhanced cross-talk between pathways is critical for the selection of multiple trajectories in maintaining a stoichiometric balance during regulation. The association of six fusion proteins with the citric acid cycle and their capability to produce metabolites with high connectivity index is intriguing. This suggests that fusion gene products and their evolution have had a key role in the selection of complex multifaceted networks. In addition, we propose that fusion proteins have gained additive biochemical function for a balanced regulation of metabolic networks.

2. INTRODUCTION

Fusion proteins in one species consist of two or more fusion partners from one or more other species and they exhibit accreted function compared to fusion partners (1). In recent years, several fusion proteins have been identified across distant phylogenetic distances and their accreted function is comprehensively discussed. They exhibit enhanced functional networks (2), substrate specificity (3), multi-functionality (4), simulate protein-protein interfaces (5) and acquire novel function (6). Therefore, the formation of fusion protein is evolutionarily selective and functionally critical. Hence, it is important to establish the mapping between events of fusion and fission across different species. However, this mapping is highly combinatorial, information demanding and computationally intensive.

In recent years, databases have been constructed to capture fusion events across distant phylogenies. These databases contain fusion proteins between human and yeast (7), human and prokaryotes (1) and within prokaryotes (8). It has been shown that many human proteins of prokaryotic origin mimic *operons* (a group of genes controlled by the same regulatory gene) and exhibit multiple functions (1). Subsequently, we identified 6 human proteins of prokaryotic origin that are associated with metabolic pathways. We further probed their accreted function by establishing connectivity to metabolites. Thus far, studies on metabolism have focused on their connectivity in networks and little is known about the origin and evolution of metabolic members that regulate network dynamics (9-10). Nonetheless, metabolic pathways are connected through an amazing diversity of compounds with different chemical structures and biological activities for a balanced stoichiometry. Therefore, the material balance is maintained through the networks by regulating a mosaic of metabolites (substrates and products) at levels of entry and

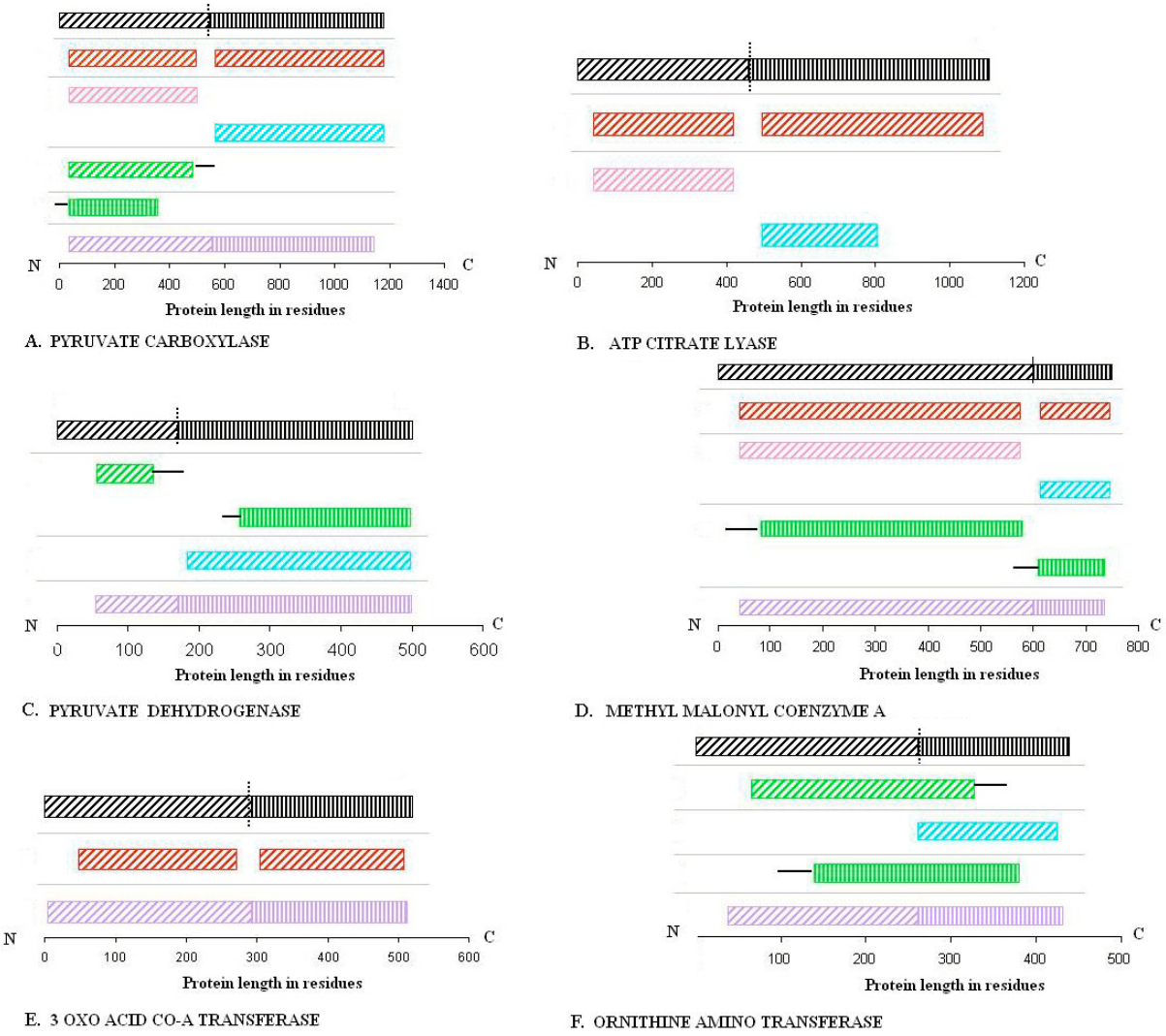

Figure 1. Fusion proteins and fusion partners. Human fusion protein [black], partners from same bacterial species [red], partners from different bacterial species [blue, pink], partners with non-homologous N or C terminal domains in a bacterial species [green], a similar fused structure in a bacterial species [purple], slanting bars indicate N terminal domains and vertical bars indicate C terminal domains. Each designated fusion pair is shown within two faint lines.

exit in a synchronized manner. Such an inter-connected synchronized design could lead to the simultaneous flow of metabolites in numerous directions with optimal kinetic rates (11-12). This is achieved through the tight regulation of proteins in networks through selective protein architectures (13). Evolution of protein architectures by gene fusion is proposed as a key mechanism for network development. Herein, we discuss the possible rationale for the evolution of complex metabolic networks with the origin and formation of fusion proteins.

3. MATERIALS AND METHODS

3.1 Human fusion proteins – definition and dataset source

By definition, a human fusion protein should show evidence of fusion partners in one or more prokaryotes (1). In this case, the N terminal and C terminal domains are made of independent fusion partners from the same or different bacterial species (Figure 1). It should be noted that the fusion partners have high sequence homology with their corresponding domains in the fused protein. The 141 human fusion proteins of prokaryotic origin presented in our previous report are used in this study (1).

3.2 Fusion proteins as metabolic enzymes

The 141 fusion proteins (1) were visually inspected and manually mapped to metabolic pathways using the KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway database interface (15). The KEGG database contains updated information on metabolic pathways, regulatory networks and molecular complexes. This information is used to identify the location of fusion proteins in the citric acid cycle of metabolic pathways (Figure 2). This exercise enabled us to select six fusion

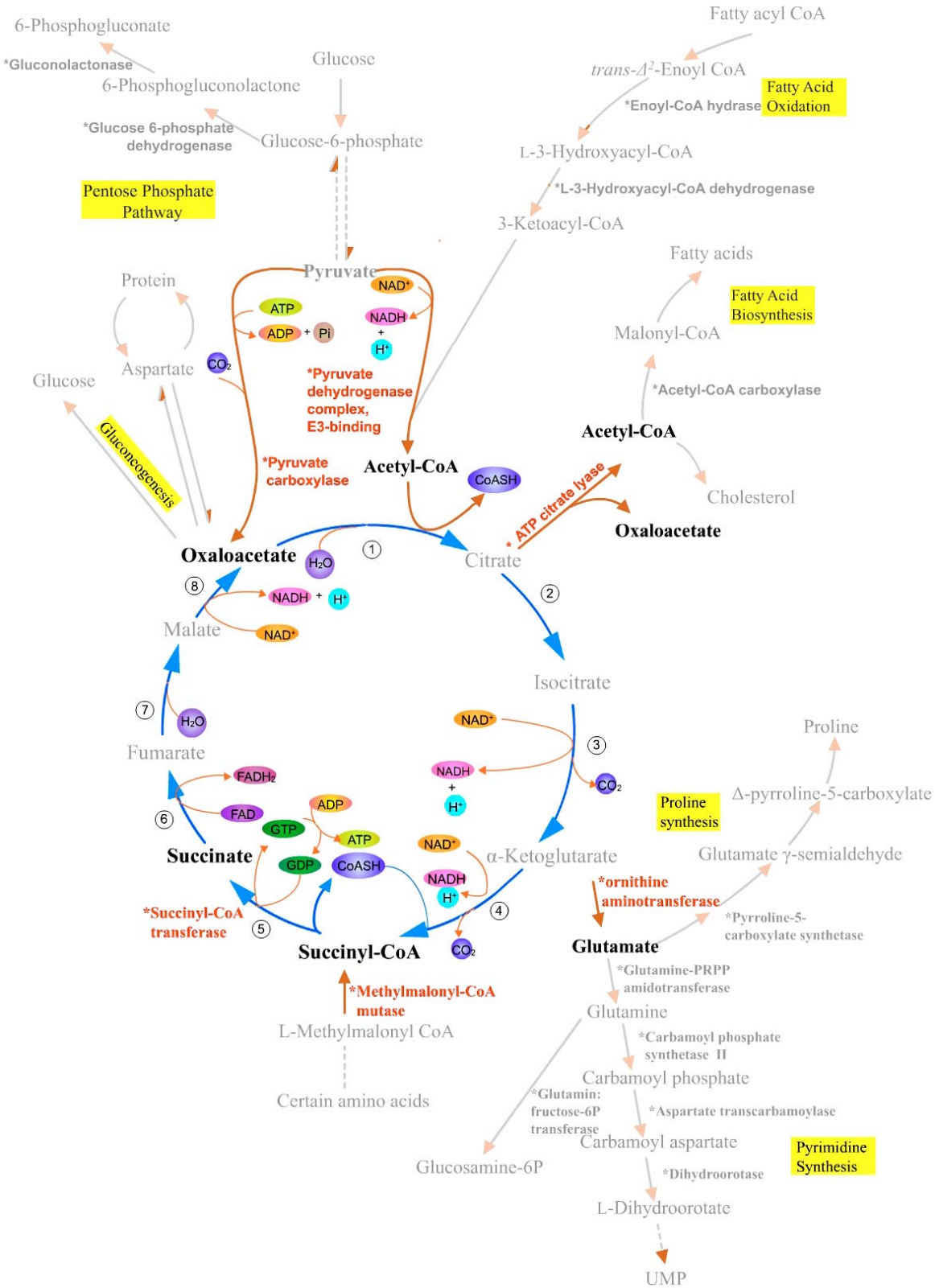

Figure 2. Fusion proteins and metabolites. A diagram showing citric acid cycle in carbohydrate metabolism . The six fusion proteins discussed in this study are indicated using an asterisk (∗) in red. The metabolites associated with the fusion proteins are indicated using BOLD font. The associated pathways with the metabolites are highlighted in yellow.

proteins (from the list of 141 fusion proteins) that function as metabolic enzymes in the citric acid cycle (Figure 2).

3.3. Fusion scenarios linked with metabolic enzymes having fusion structures

The fusion proteins consist of two or more fusion partners (Figure 1). In each of these fusion proteins [black bars in Figure 1], the N terminal domain [slanting bars in Figure 1] represents one fusion partner and the C terminal domain [vertical bars in Figure 1] represents another fusion partner. In Figure 1, each designated fusion pair is illustrated within two faint lines.

3.3.1. Scenario 1 and scenario 2

As shown in Figure 1, the fusion pairs may either come from the same bacterial species [red – scenario 1] or from different bacterial species [blue, pink – scenario 2]. The fusion partners shown in red, blue and pink represents a highly homologous (by measure of sequence similarity) full length ORF (open reading frame) in one or more bacterial species.

3.3.2. Scenario 3

The fusion partners shown in green (refer to Figure 1) have matching regions to either N or C terminal domains of the fused protein. In addition, the fusion partners represented in green also contain either preceding (N terminal region for bacterial protein) or following (C terminal region for the bacterial protein) domains or regions that are non-homologous (indicated by a protruding line from the center of a bar) to regions of a fusion protein. Thus, these categories of fusion partners have both homologous and non-homologous domains with reference to a fusion protein.

3.4 PRODOM domain assignments to fusion proteins

Fusion proteins consist of two or more domains for additive or novel role. Therefore, it is important to document the different domains that constitute a fusion protein (Table 1). For this purpose, we used the PRODOM database, which is a comprehensive set of protein domain families automatically generated from the SWISS-PROT and TrEMBL (14). This automatic assignment exercise indicates that each fusion protein consists of two or more PRODOM domains. This suggests fusion proteins consist of several domain-like units as building blocks.

3.5. Metabolites associated with fusion proteins

The metabolic enzymes having fusion structures use metabolites as substrates and produce metabolites as products (Table 2). It is our interest to establish the link between fusion proteins and the associated metabolites (Figure 2). For this purpose, we used the KEGG interface (15).

3.6. Connectivity index for metabolites of fusion proteins

Each metabolite associated with a fusion protein is connected with many other metabolic enzymes in pathways. Here, we hypothesize that the multifaceted connectivity of these metabolites with other members of the pathway is the driving force for domain accretion in fusion proteins. For this purpose, we define connectivity index for a metabolite associated with a fusion protein. By definition, the connectivity index of a metabolite is defined as its ability to connect (number of known links) with other enzymes/proteins in pathways. The connectivity index was calculated from the KEGG ligand database (Table 2).

4. RESULTS

4.1. Human fusion proteins

Human fusion proteins of prokaryotic origin have been identified and their role in biological system is implied (1). These proteins exhibit enhanced or novel functions through additive structural architectures. They are shown to mimic operons, simulate protein-protein interfaces, perform multiple functions and exhibit alternative splicing (1). We further examined these proteins and identified six of them that are involved in the citric acid cycle (Figure 2). These proteins consist of two or more PRODOM domains from two or more fusion partners of bacterial origin (Table 1). In these fused proteins, the N and C terminal domains represent physically separated fusion partners in one or more prokaryotes (Figure 1). The origin and mechanism of fusion protein is puzzling. In Figure 1, each designated fusion pair is illustrated within two faint lines. A detailed analysis of the fusion pairs suggests a number of fusion scenarios (Figure 1). These scenarios (see Materials and Methods) were illustrated as scenario 1 (fusion partners from same bacterial species), scenario 2 (fusion partners from different bacterial species) and scenario 3 (fusion partners have both homologous and non-homologous domains with reference to a fusion protein). This illustration indicates that fusion of partners to form a fused structure takes several possible routes as shown by different scenarios. It should be noted that a similar fusion structure could be found in one or more bacterial species [purple bars in Figure 1]. This implies that the fusion structures given in this report are not exclusive fused entities for human proteins. However, these human fusion proteins consist of fusion partners from one or more bacterial species. This suggests that fusion events are not only seen between bacterial and human but also within bacterial species. Hence, the origin and accreted role of human fusion proteins is evolutionarily interesting and functionally puzzling. Therefore, an understanding of their structural and functional evolution is critical.

4.2. Fusion proteins as metabolic enzymes

We selected six metabolic enzymes with fusion structures from a list of 141 fusion proteins reported elsewhere (1). The KEGG pathway database was used to identify the location of fusion proteins in the citric acid cycle (Figure 2). The six metabolic enzymes consisting of fusion structures are given in Table 1. This association between fusion proteins and members of metabolic networks is interesting. It is our interest to establish a detailed understanding of their role in metabolic networks in the light of their domain accretion and gene fusion. As given in Table 1, these human fusion proteins have N and C terminal domains matching independent subunits in bacterial genomes. This observation implies that metabolic enzymes with fusion structures have incremental role in

Table 1. Human fusion proteins in carbohydrate metabolism

	Human fusion (composite) proteins					Fusion partners (component) of prokaryotic origin									
					Prodom domain	N terminal					C terminal				
	RefSeq Accession	PL	Protein name	#	ID	PL	MR	Protein	#	Prodom domains	PL	MR	Protein	#	Prodom domains
1	NP_000911	1178	Pyruvate carboxylase precursor	14	PD000820 PD435299 PD000180 PD000755 PD002908 PD004644 PD002904 PD429392 PD327958 PD435081 PD328972 PD519428 PD007602 PD000268	477	38..483	Pyruvate carboxylase, subunit A	5	PD000820 PD000180 PD489184 PD000755 PD206296	567	563..1178	Pyruvate carboxylase, subunit B	8	PD002904 PD429392 PD327958 PD414942 PD328972 PD414942 PD440923 PD000268
2	NP_001087	1105	ATP citrate lyase	9	PD579887 PD593323 PD263361 PD003397 PD010086 PD140365 PD014373 PD002034 PD013870 PD027540 PD055009 PD033409 PD588241 PD510157 PD310830 PD001115 PD594236	398	41..418	Citrate lyase, subunit 1	2	PD003397 PD010086	610	496..1089	Citrate lyase, subunit 2	4	PD337294 PD014373 PD002034 PD013870
3	NP_003468	501	Pyruvate dehydrogenase	8	PD037584 PD583163 PD588193 PD477798 PD003187 PD004976 PD130469 PD130468	473	57..136	Biotin/lipoyl attachment	2	PD055009 PD000268	177	324..500	FAD-dependent pyridine nucleotide-disulphide oxidoreductase	3	PD510157 PD001115 PD564191
4	NP_000427	520	Succinyl CoA transferase	6	PD004447 PD004310 PD005702 PD592542 PD481683 PD002527	250	36..284	B-etoadipate: succinyl CoA transferase, subunit α	5	PD515695 PD477798 PD583163 PD588193 PD003187	219	302..516	B-ketoadipate :succinyl-CoA transferase subunit β	1	PD004976
5	NP_000246	750	Methylmalonyl Co-A mutase	9	PD595004 PD021187 PD000493 PD082173 PD000465 PD066084 PD308202 PD556298 PD001337	681	42..519	Methylmalonyl Co A mutase, subunit α, N-terminus	7	PD188365 PD191238 PD004447 PD004310 PD005702 PD537237 PD592542	144	611..749	Methylmalonyl Co A mutase, subunit α, C-terminus	2	PD002527 PD514726
6	NP_000265	439	Ornithine amino-transferase	8		426	58..330	Ornithine-oxo-acid aminotransferase	4	PD528736 PD082173 PD000465 PD455654	181	261..425	Probable ornithine aminotransferase (C-terminal part)	4	PD000465 PD066084 PD236513 PD001337

PL = protein length; MR = matching region in the fusion protein; # = number of PRODOM domains; ID = PRODOM domain identifier

pathways. Although, an association between fusion proteins and metabolic networks is realized through this observation, it is important to establish the significance of this relation in specific quantitative terms. Additionally, the automatic PRODOM assignment procedure indicates that fusion protein consists of two or more PRODOM domains, suggesting that domain-like units serve as building blocks in their evolution. In conclusion, fusion proteins acquire complex structural architectures through the modular arrangements of building blocks at multiple layers of organization.

4.3. Metabolites of fusion proteins

The six metabolic enzymes having fusion structures use or produce metabolites that have multiple trajectories in pathways (Figure 2). These metabolites have high connectivity indicating greater involvement in cross talk between networks (Table 2). Hence, these metabolites are associated with a mosaic of reactions in networks. Therefore, a stoichiometric pressure is built on these metabolites and the need to establish a material balance is critical. Herein, we relate the high connectivity index of metabolites with their corresponding fusion proteins in the citric acid cycle. The five metabolites produced by these six enzymes are summarized in Table 2. Data in Table 2 suggests that these five metabolites have multifaceted role by participating in a mosaic of reactions and pathways.

The first metabolite, oxaloacetate is produced by pyruvate carboxylase and ATP citrate lyase (Figure 2) and

Table 2. Fusion proteins as metabolic enzymes and associated metabolites

Metabolites	Fusion proteins as metabolic enzymes	Number of associated metabolic reactions	Number of associated metabolic pathways	Number of enzymes associated with the metabolite		
				Total	*Enzyme uses metabolite as a substrate*	*Enzyme produces metabolite as a product*
Oxaloacetate	Pyruvate carboxylase & ATP citrate lyase	41	9	40	16	24
Acetyl co-A	Pyruvate dehydrogenase complex (E3-binding) & ATP citrate lyase	132	28	98	80	18
Succinate	Succinyl co-A transferase	79	11	40	5	35
Succinyl co-A	Methylmalonyl co-A mutase	26	6	18	13	5
Glutamate	Ornithine amino-transferase	121	11	82	25	57

it is associated with 41 reactions in 9 pathways (Table 2). It acts as substrate at 16 points and as product at 24 points. Clearly, oxaloacetate is under high stoichiometric pressure and data suggests that this metabolite is produced by two fusion proteins with accreted domains. However, the fusion partners are physically separated in one or more prokaryotes (Figures 1A and 1B). Hence, the mere observation demonstrating fusion/fission events is greatly intriguing. Pyruvate carboxylate consists of 14 PRODOM domains, where the N and C terminal domains are structurally similar to the physically separated subunits A and B of fusion partners in bacteria (Table 1). However, a similar fused structure is also seen in the bacterium *A. tumefaciens*. In *Methanosarcina barkeri*, pyruvate carboxylase exhibits an *operon* like structure (16). Similarly, ATP citrate lyase is made up of 9 PRODOM domains, where the N and C terminal domains are structurally similar to the physically separated subunits 1 and 2 of bacterial fusion partners. In *Klebsiella pneumoniae*, ATP citrate lyase also exhibits a unique *operon* like structure (17). Thus, the fused structure is analogous to an *operon* like arrangement. Data in Table 2 indicates that these two fusion proteins are associated with oxaloacetate which is severely constraint through several inter-connections in metabolism. In such an environment, oxaloacetate is formed from pyruvate by pyruvate carboxylase and from citrate by ATP citrate lyase. The second metabolite, acetyl co-A is produced by ATP citrate lyase and pyruvate dehydrogenase (Figure 1). This metabolite is associated with 132 reactions in 28 pathways (Table 2). It acts as substrate at 80 points and as product at 18 points. Thus, acetyl co-A is associated with many reactions in metabolic networks. However, acetyl co-A is also product of fusion proteins ATP citrate lyase and pyruvate dehydrogenase. Pyruvate dehydrogenase is made up of 8 PRODOM domains with an N terminal subunit A and a C terminal subunit B and ATP citrate lyase is made up of 9 PRODOM domains with an N terminal subunit 1 and a C terminal subunit 2 (Table 2).

The third metabolite, succinyl co-A is produced by methyl malonyl co-A mutase (Figure 1). It is involved in 26 reactions in 6 pathways. It acts as substrate at 13 points and as product at 5 points. The enzyme methyl malonyl co-A mutase is made up of 9 PRODOM domains where the N and C terminal domains are analogous to subunits A and B (Table 2). These results indicate that metabolites of fusion proteins are severely constraint in the network. It has been shown that methylmalonyl co-A mutase simulates protein-protein interactions in *Propionibacterium shermanii* (19). The fusion partners are physically separated in *Propionibacterium shermanii* and they associate through weak interactions without covalent bonding. Thus, fusion imparts stability to protein-protein associations through the formation of a more stable covalently linked domain-domain interfaces. The fourth metabolite, succinate is produced by succinyl co-A transferase (Figure 1). It is associated with 79 reactions in 11 pathways (Table 2). It also shows that succinate acts as substrate at 5 points and as product at 35 points. Succinyl co-A transferase consist of two fusion partners (Figure 2). It is made up of 6 PRODOM domains with an N terminal subunit A and a C terminal subunit B (Table 2). This fusion protein is shown to simulate protein-protein interactions in *Pseudomonas putida* (20) and mimic operon like structure in *C. acetobutylicum* (21). The fifth metabolite, glutamate is produced by *ornithine aminotransferase* (Figure 1) and is involved in 121 reactions in 11 pathways (Table 2). It acts as substrate at 25 points and as product at 57 points. This protein is made up of 8 PRODOM domains with an N terminal subunit A and a C terminal subunit B (Table 2). These data show the possible link between fusion proteins and pathways with reference to oxaloacetate, acetyl co-A, succinyl co-A, succinate and glutamate. The hypothesis is that fusion proteins could meet the demand for metabolites by fusing two or more fusion partners that are physically separated in one or more bacteria. This is consistent with a previous report which suggested that biotin carboxylase family of enzymes have evolved into a complex multifunctional protein from smaller mono-functional precursors through successive gene fusions (18). Therefore, the fusion proteins with modular organization possibly acquire selective incremental functions for optimized role in complex networks.

5. DISCUSSION

Protein evolution is extremely efficient in generating systems that are optimally adapted in cellular environment (22-25). Optimality can be achieved by changing the topology of metabolic networks by tuning enzymatic or regulatory materials (11). Here, we show that metabolites like oxaloacetate, acetyl co-A, succinyl co-A, succinate and glutamate are products of fusion proteins. These metabolites have high connectivity index, suggesting their greater degree of involvement within networks. This observation implies the association of fusion proteins with complex metabolic networks. The association between human fusion proteins and metabolites with high connectivity is intriguing (Table 2). Detailed analysis of fusion proteins highlights the transition from a 'protein-protein interface' to either a 'domain-domain interface' or

an operon structure (a group of genes all controlled by the same regulatory gene) (1). This evolutionary transition is intriguing and it is important to systematically investigate the functional link between fusion partners and fused proteins using thermodynamics calculations. The transition may be thermodynamically favorable as fusion proteins acquire reduced entropy compared to their physically separated fusion partners. Therefore, it is envisaged that fusion proteins confer selective advantage in the evolution of regulating metabolic dynamics. This is specifically advantageous for multi-enzyme complexes as it selects kinetic advantages over unassociated enzyme components by increasing connectivity with metabolites. It is also reported that fusion of components into a single polypeptide ensures stability between physically connected domain structures and active sites for a balanced stoichiometric production of intermediates in complex networks (26-27). The physical proximity of multiple active centers in the same metabolic pathways alleviates molecular diffusion and reduces side reactions in cellular environment (28). Our data for the six metabolic enzymes having fusion structures aligns well with these observations. This enables fusion proteins to catalyze sequential steps in a biochemical pathway because association of two active sites enhances the efficiency of two consecutive reactions. Thus, fused protein architecture illustrates an evolutionary strategy accreted for maintaining complex stoichiometric balance. It is often thought that the function of fused genes is simply an addition of function to pre-existing component genes. However, chimerical genes generate proteins with novel function (29). A recent structural analysis of the Histidine biosynthesis components HisA and HisF indicates that the protein structure after gene fusion was also subject to structural and functional adaptation (30). In this case, gene fusion produced a new protein fold with novel function. Thus, most fusion proteins reveal that they have acquired separate functional domains from each component protein through domain accretion by gene fusion. Physical connection between fused domains increases structural propensity between active centers for the regulation of material balance (31). Since fusion proteins help in the evolution of complex networks, even a modest addition of domains could significantly increase numerous new interactions. This strategy helps to maintain equilibrium in a dynamic network with huge nodes. Thus, large networks of molecular interactions are regulated by relatively few genes in some organisms (32). That is, the 'gene number' (number of genes in a species genome) is negligible to 'reaction number' (number of reactions in a species cell) in higher eukaryotes. Our results illustrate an important evolutionary phenomenon that involves the formation of cellular network dynamics with the help of fusion proteins having multi-domain structural architectures with incremental functions.

6. CONCLUSION

The current analysis provides new insights into the relation of fusion proteins and network evolution. Thus, evolution of fusion genes has a key role in the selection and design of multifaceted network associated with complex genomes. This might have enabled fusion proteins to accrete incremental biochemical function for a balanced regulation of metabolic networks. Hence, protein fusion confers a selective advantage in evolution. We believe that an understanding of fusion scenarios and their association with members of a pathway should enable us to appreciate the role played by them to combine or to share metabolites across networks, creating novel pathways with functional diversity. In reality, there are many more fusion structures of varying phylogeny and a majority of them have not been captured as the networks are extremely large and complex. Although, the insights drawn from a detailed study of six fusion proteins is found interesting and valuable, additional evidence is required to establish a comprehensive relationship between additive biochemical function and network dynamics.

7. ACKNOWLEDGEMENTS

This work was supported by Nanyang Technological University, National University of Singapore and A*STAR-BMRC, Singapore (Grant # 01/1/21/19/191 and Grant # 03/1/22/19/242).

8. REFERENCES

1. Yiting Y., I. Chaturvedi, L.K. Meow, P. Kangueane & M.K. Sakharkar: Can ends justify the means? Digging deep for human fusion genes of prokaryotic origin. *Front Biosci e* 9, 2964-2971 (2004)
2. Yanai I., A. Derti.& C. DeLisi: Genes linked by fusion events are generally of the same functional category: A systematic analysis of 30 microbial genomes. *Proc. Natl. Acad. Sci. USA* 98, 7940-7945 (2001)
3. Katzen F., M. Deshmukh, F. Daldal & J. Beckwith: Evolutionary domain fusion expanded the substrate specificity of the transmembrane electron transporter DsbD. *EMBO J.* 21, 3960-3969 (2002)
4. Berthonneau E. & M. Mirande: A gene fusion event in the evolution of aminoacyl-tRNA synthetases. *FEBS Lett.* 470, 300-304 (2000)
5. Marcotte E.M., M. Pellegrini, H.L. Ng, D.W. Rice, T.O. Yeates & D. Eisenberg: Detecting protein function and protein-protein interactions from genome sequences. *Science* 285, 751-753 (1999)
6. Long M: A new function Evolved from Gene Fusion. *Genome Res.* 10, 1655-1657 (2000)
7. Truong K. & M. Ikura: Domain fusion analysis by applying relational algebra to protein sequence and domain databases. *BMC Bioinformatics.* 4, 16 (2003)
8. Suhre K. & J.M. Claverie: FusionDB: a database for in-depth analysis of prokaryotic gene fusion events. *Nucleic Acids Res.* 1, D273-D276 (2004)
9. Fernie A.R., F. Carrari & L.J. Sweetlove: Respiratory metabolism: glycolysis, the TCA cycle and mitochondrial electron transport. *Curr Opin Plant Biol.* 7, 254-261 (2004)
10. Blank L.M. & U. Sauer: TCA cycle activity in *Saccharomyces cerevisiae* is a function of the environmentally determined specific growth and glucose uptake rates. *Microbiology.* 150, 1085-1093 (2004)
11. Segre D., D. Vitkup & G.M. Church: Analysis of optimality in natural and perturbed metabolic networks. *Proc. Natl. Acad. Sci. USA* 99, 15112-15117 (2002)

12. Stelling J., S. Klamt, K. Bettenbrock, S. Schuster & E.D. Gilles: Metabolic network structure determines key aspects of functionality and regulation. *Nature* 420, 190-193 (2002)
13. Ihmels J., R. Levy & N. Barkai: Principles of transcriptional control in the metabolic network of *Saccharomyces cerevisiae*. *Nat. Biotechnol.* 22, 86-92 (2004)
14. Servant F., C. Bru, S. Carrere, E. Courcelle, J. Gouzy, D. Peyruc & D. Kahn: ProDom: automated clustering of homologous domains. *Brief Bioinform.* 3, 246-325 (2002)
15. Kanehisa M., S. Goto, S. Kawashima, Y. Okuno & M. Hattori: The KEGG resource for deciphering the genome. *Nucleic Acids Res*. 32, D277-D280 (2004)
16. Mukhopadhyay B., E. Purwantini, C.L. Kreder, & R.S. Wolfe: Oxaloacetate synthesis in the methanarchaeon *Methanosarcina barkeri*: pyruvate carboxylase genes and a putative Escherichia coli-type bifunctional biotin protein ligase gene (bpl/birA) exhibit a unique organization. *J. Bacteriol.* 183, 3804-3810 (2001)
17. Bott M. & P. Dimroth: *Klebsiella pneumoniae* genes for citrate lyase and citrate lyase ligase: localization, sequencing, and expression. Mol Microbiol. 14, 347-356 (1994)
18. Obermayer M. & F. Lynen: Structure of biotin enzymes. *Trends. Biochem. Sci.* 1, 169–171 (1976)
19. Marsh E.N., S.E. Harding, & P.F. Leadlay: Subunit interactions in *Propionibacterium shermanii* methylmalonyl-CoA mutase studied by analytical ultracentrifugation. *Biochem. J.* 260, 353-358 (1989)
20. Parales R.E. & C.S. Harwood: Characterization of the genes encoding beta-ketoadipate: succinyl-coenzyme A transferase in *Pseudomonas putida. J Bacteriol.* 174, 4657-4666 (1992)
21. Yeh W.K. & L.N. Ornston: Evolutionarily homologous alpha 2 beta 2 oligomeric structures in beta-ketoadipate succinyl-CoA transferases from Acinetobacter calcoaceticus and Pseudomonas putida. *J. Biol. Chem.* 256, 1565-1569 (1981)
22. Freeland S.J., R.D. Knight, L.F. Landweber & L.D. Hurst: Early Fixation of an Optimal Genetic Code. *Mol. Biol. Evol.* 17, 511-518 (2000)
23. Heinrich R. & S. Schuster: The modeling of metabolic systems. Structure, control and optimality. *Biosystems* 47, 61-77 (1998)
24. Stephani A., J.C. Nuno & R. Heinrich: Optimal Stoichiometric Designs of ATP-producing Systems as Determined by an Evolutionary Algorithm. *J. Theor. Biol.* 199, 45-61 (1999)
25. Akashi H. & T. Gojobori: Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis. Proc. Natl. Acad. Sci. USA* 99, 3695-3700 (2002)
26. McCarthy A.D. & D.G. Hardie: Fatty acid synthase: an example of protein evolution by gene fusion. *Trends Biochem. Sci.* 9, 60-63 (1984)
27. Rochet J.C., Oikawa, K., Hicks, L.D., Kay, C.M., Bridger, W.A. & Wolodko, W.T. Productive interactions between the two domains of pig heart CoA transferase during folding and assembly. *Biochemistry* 36, 8807-8820 (1997)
28. West I.C: Channelling in intermediary metabolism, ed. Auius, L.S. (H. S. A., London). 9, 13-69 (1997)
29. Brodsky G., T. Barnes, J. Bleskan, L. Becker, M. Cox & D. Patterson: The human GARS-AIRS-GART gene encodes two proteins which are differentially expressed during human brain development and temporally over-expressed in cerebellum of individuals with Down syndrome. *Hum. Mol. Genet*. 6, 2043-2050 (1997)
30. Lang D., R. Thoma, M. Henn-SAX, R. Sterner & M. Wilmanns: Structural evidence for evolution of the beta/alpha barrel scaffold by gene duplication and fusion. *Science* 289, 1546-1550 (2000)
31. Koonin E.V., Y.I. Wolf & G.P. Karev: The structure of the protein universe and genome evolution. *Nature* 420, 218-223 (2002)
32. Lander E.S., L.M. Linton, B. Birren, C. Nusbaum, M.C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J.P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J.C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R.H. Waterston, R.K. Wilson, L.W. Hillier, J.D. McPherson, M.A. Marra, E.R. Mardis, L.A. Fulton, A.T. Chinwalla, K.H. Pepin, W.R. Gish, S.L. Chissoe, M.C. Wendl, K.D. Delehaunty, T.L. Miner, A. Delehaunty, J.B. Kramer, L.L. Cook, R.S. Fulton, D.L. Johnson, P.J. Minx, S.W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J.F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, R.A. Gibbs, D.M. Muzny, S.E. Scherer, J.B. Bouck, E.J. Sodergren, K.C. Worley, C.M. Rives, J.H. Gorrell, M.L. Metzker, S.L. Naylor, R.S. Kucherlapati, D.L. Nelson, G.M. Weinstock, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, D.R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, H.M. Lee, J. Dubois, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, L. Hood, L. Rowen, A. Madan, S. Qin, R.W. Davis, N.A. Federspiel, A.P. Abola, M.J. Proctor, R.M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D.R. Cox, M.V. Olson, R. Kaul, C. Raymond, N. Shimizu, K. Kawasaki, S. Minoshima, G.A. Evans, M. Athanasiou, R. Schultz, B.A. Roe, F. Chen, H. Pan, J. Ramser, H. Lehrach, R. Reinhardt, W.R. McCombie, M. de la Bastide, N. Dedhia, H. Blocker, K. Hornischer, G. Nordsiek, R. Agarwala, L. Aravind, J.A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D.G. Brown, C.B. Burge, L. Cerutti, H.C. Chen, D. Church, M. Clamp, R.R. Copley, T. Doerks, S.R. Eddy, E.E. Eichler, T.S. Furey, J. Galagan, J.G. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L.S. Johnson, T.A. Jones, S. Kasif, A. Kaspryzk, S. Kennedy, W.J. Kent, P. Kitts, E.V. Koonin, I. Korf, D. Kulp, D. Lancet, T.M.

Lowe, A. McLysaght, T. Mikkelsen, J.V. Moran, N. Mulder, V.J. Pollara, C.P. Ponting, G. Schuler, J. Schultz, G. Slater, A.F. Smit, E. Stupka, J. Szustakowski, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y.I. Wolf, K.H. Wolfe, S.P. Yang, R.F. Yeh, F. Collins, M.S. Guyer, J. Peterson, A. Felsenfeld, K.A. Wetterstrand, A. Patrinos, M.J. Morgan, J. Szustakowki, P. de Jong, J.J. Catanese, K. Osoegawa, H. Shizuya, S. Choi & Y.J. Chen: Initial sequencing and analysis of the human genome. *Nature* 409, 860-921 (2001)

Key Words: Gene Fusion, Fusion Partners, Functional Accretion, Networks, Connectivity, Metabolites

Send correspondence to: Meena Kishore Sakharkar Ph.D, Nanyang Technological University, Singapore 639798; Tel: (+65) 6 790 5836; Fax: (+65) 6 774 4340; E-mail: mmeena@ntu.edu.sg

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