Modularized study of human calcium signalling pathway

LOSIANA NAYAK and RAJAT K DE*

Machine Intelligence Unit, Indian Statistical Institute, 203 B. T. Road, Kolkata 700 108, India

**Corresponding author (Fax, 91-33-2578 3357; E-mail, rajat@isical.ac.in)*

Signalling pathways are complex biochemical networks responsible for regulation of numerous cellular functions. These networks function by serial and successive interactions among a large number of vital biomolecules and chemical compounds. For deciphering and analysing the underlying mechanism of such networks, a modularized study is quite helpful. Here we propose an algorithm for modularization of calcium signalling pathway of *H. sapiens*. The idea that "a node whose function is dependant on maximum number of other nodes tends to be the center of a subnetwork" is used to divide a large signalling network into smaller subnetworks. Inclusion of node(s) into subnetworks(s) is dependant on the outdegree of the node(s). Here outdegree of a node refers to the number of relations of the considered node lying outside the constructed subnetwork. Node(s) having more than c relations lying outside the expanding subnetwork have to be excluded from it. Here c is a specified variable based on user preference, which is finally fixed during adjustments of created subnetworks, so that certain biological significance can be conferred on them.

1. Introduction

Modularization is a process by which we can break a network into small units for better analysis of the original network. The idea is used here to break human calcium signalling pathway into simple entities known as *modules*. Since there is no single definition of a module, we have followed certain criteria to create them. We have assumed that a module is a part of the original biochemical network, which tends to be self-sufficient and has minimal dependency on the rest part of the network. The justification for dividing a network into a number of modules lies in the fact that the complexity of each module is much less than that of entire pathway and is an easier means of studying the network by part. Thus analysing all the modules separately, generated from a pathway, we can have a better operational view of the whole network. When all modules of the pathway are generated by the algorithm, the network can be pictorially represented by modules as nodes and interactions among the modules as edges. This kind of representations roughly give idea of dependence among the created modules.

Calcium signalling pathways are very peculiar in nature. When there is an extracellular change, cells get the message either by introduction of calcium ions into cytoplasm or by evacuation to outside through ion channels. This mechanism is aided by organellar, and nuclear storage of calcium ions. Normal intracellular Ca^{2+} level is 10^{-7} M, much lower than the extracellular concentration of 10^{-3} M and lower than that of individual organellar concentrations. Cytoplasmic calcium ion concentration must be maintained at low levels as it precipitates phosphate, the established energy currency of cells. Prolonged high intracellular calcium levels even lead to cell death. Hence cells first evolved techniques for free calcium ion binding to reduce its effect towards cytosol which later is used as well for signal transduction across and inside the cell (Clapham 1995).

In this article, we have developed an algorithm that can be used for creating modules from biological networks. The algorithm views an entire biochemical pathway as a graph having gene products and chemical compounds as vertices and edges being different kind of interactions as shown in figure 1. An edge can be a protein-protein interaction or

Keywords. Biological networks; network modularization; nucleoplasmic reticulum; signal transduction

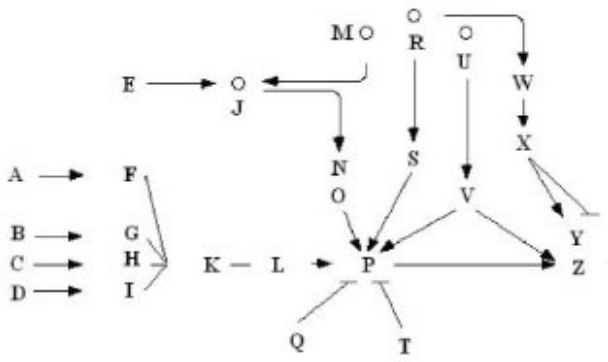


Figure 1. A Hypothetical Network (chemical compounds denoted with a circle, → activation, – binding, –| inhibition).

protein-compound interaction or a link to another map. For simplicity, here we have not taken the inter-network links into account. Thus the proposed algorithm can be applied to a signal transduction pathway to get a reduced network that enables better understanding. The methodology has been applied to a sample pathway (figure 1) and to calcium signalling pathway of *H. sapiens* (figure 6).

2. Role of calcium ions in signal transduction

Intracellular free calcium ion concentration gets maintained by calcium binding proteins and calcium pumps present in unit membranes. Calcium binding proteins (buffer and trigger proteins) bind with free calcium ions when present in abundance, and maintain its low level in cytoplasm. Buffer proteins bind with free calcium ion(s) as its concentration goes up in intracellular environment (Ex: Calsequestrin) and trigger proteins upon binding with free calcium ion(s) change their confirmation to modulate enzymes and ion channels, e.g., Calmodulin (Baimbridge *et al* 1992; Heizmann and Hunziker 1991). But these proteins serve for temporary solution. Calcium pumps are necessary to maintain low cytologic calcium level against the highly calcium rich extracellular environment (Zacharias *et al* 1995). Different kinds of calcium pumps are present in plasma membrane, mitochondrial membrane and smooth endoplasmic reticular membrane. Details are shown in figure 2 and 3 (Alberts *et al* 2002; Clapham 1995; Pozzan *et al* 1994).

Many intracellular signal transduction pathways rely on elevated Ca^{2+} ion concentration as an important indicator (Berridge 1997). In addition, intercellular Ca^{2+} wave

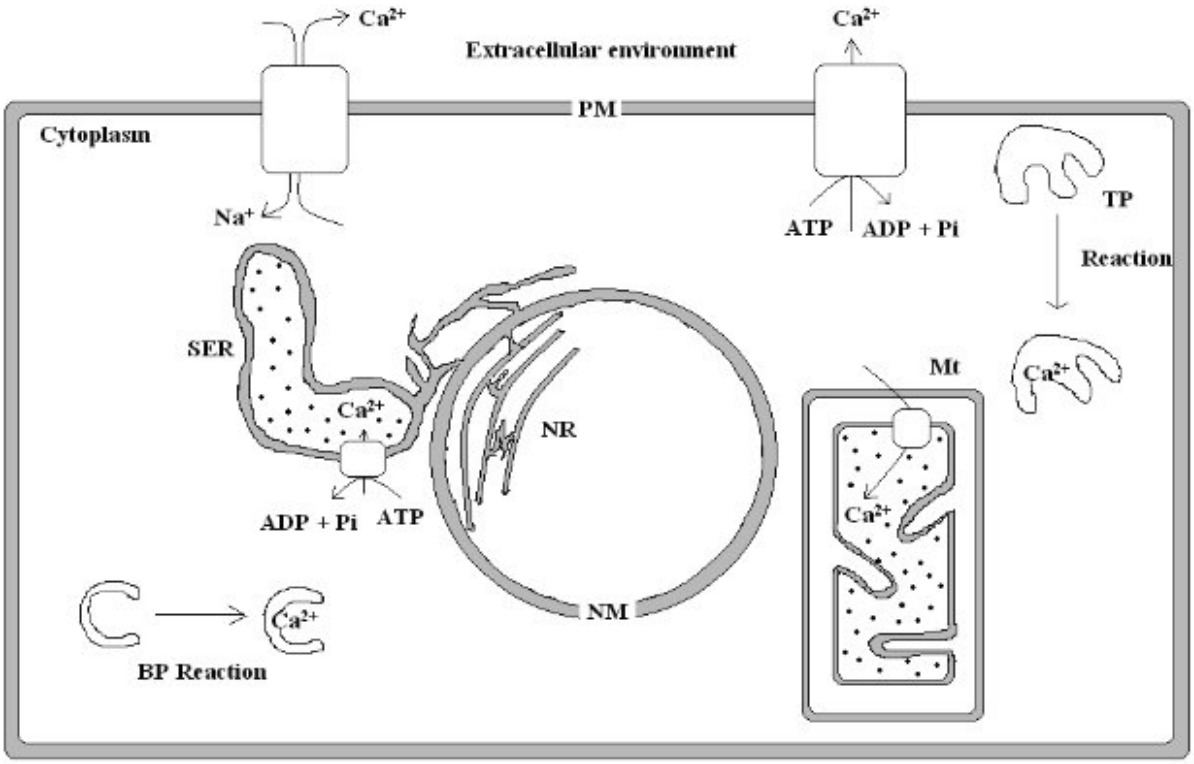


Figure 2. Diagrammatic view of various mechanisms for [Ca^{2+}] balance in a cell. (PM - Plasma Membrane, SER - Smooth Endoplasmic Reticulum, Mt - Mitochondria, TP - Trigger Protein, BP - Buffer Protein, NM - Nuclear membrane, NR - Nucleoplasmic reticulum).

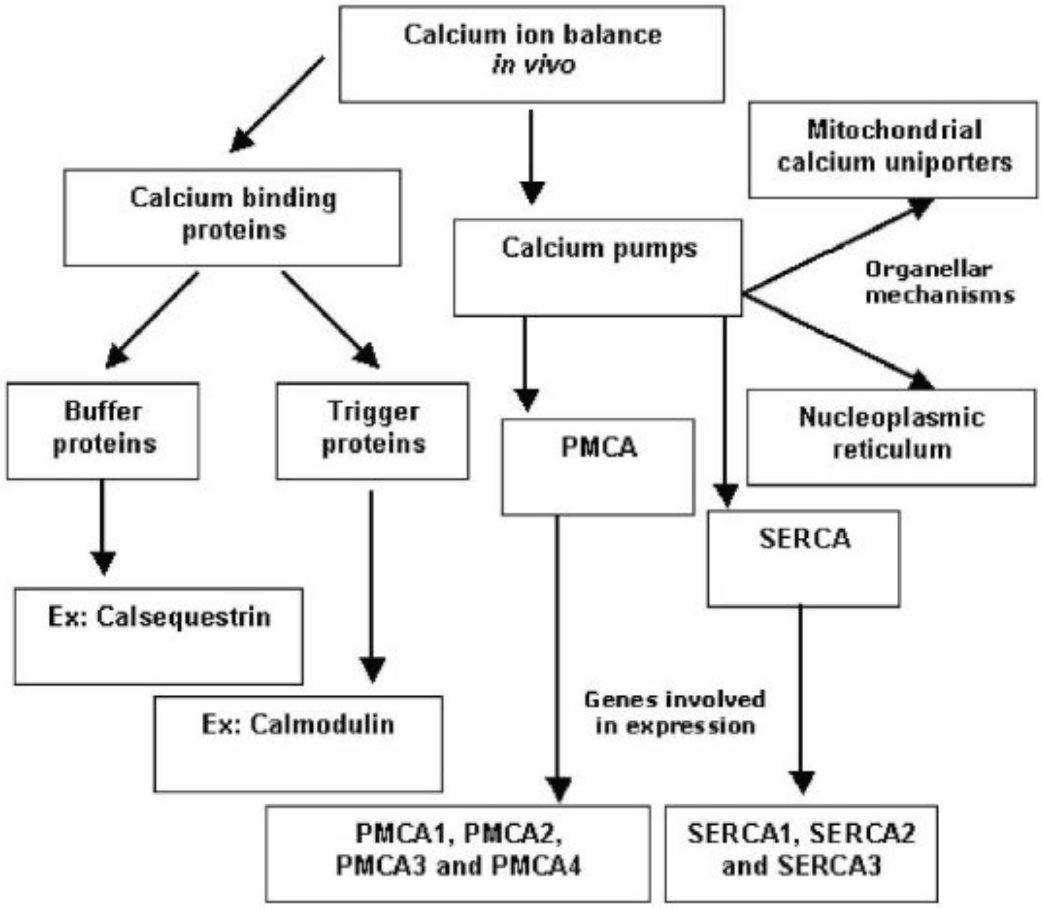


Figure 3. Mechanisms of intracellular calcium ion balance (Plasma Membrane Ca^{2+} ATPase (PMCA) pumps use ATP to pump calcium out of the cytosol. They are expressed by four genes (PMCA1, PMCA2, PMCA3 and PMCA4) in a tissue specific manner, with multiple alternatively spliced versions (Zacharias *et al* 1995). Ca^{2+} ATPase (SERCA) pumps enable the ER to take up a large amount of Ca^{2+} from the cytosol against a steep concentration gradient, even when Ca^{2+} levels in cytoplasm are low. SERCA pumps are the products of three different genes, known as SERCA1, SERCA2 and SERCA3. They are expressed in fasttwitch skeletal muscle, cardiac and slowtwitch skeletal muscle and nonmuscle tissues respectively. Mitochondrial Ca^{2+} uniporters have lower affinities for calcium ion than SERCA pumps. They have important role in returning the Ca^{2+} concentration to normal after a Ca^{2+} signal (Alberts *et al* 2002; Clapham 1995; Pozzan *et al* 1994). Also there is a Nuclear Ca^{2+} storing network that is continuous with the endoplasmic reticulum and nuclear envelope. This network is known as *nucleoplasmic reticulum*. It is an *InsP*3-gated calcium store that can give rise to local calcium signals in the nuclear interior).

propagation may act to coordinate the response of nearby cells in the tissue, leading to the concerted response of the tissue (Golumbfskie *et al* 2003). Most interestingly Ca^{2+} signalling pathway can have local as well as global effects depending on the concentration and mode of Ca^{2+} entry inside cell. When signal travels within submicromolar range, the concerned signalling pathway is essentially local. Information transferred in these pathways does not effect global changes like gene transcription and it terminates in the vicinity of signal origin (Levchenko 2002). Nuclear and cytologic Ca^{2+} signals can have effects that are independent of one another (Echevarria *et al* 2003).

Behaviour of calcium ions can be studied by calcium ion sensitive fluorescent indicators (aequorin/fura-2) inside a cell. Local openings of individual (or small groups of) calcium release channels in ER represent small and localized signal, seen in one or more discrete regions of cell. They are called as calcium blips, quarks, puffs, or sparks and represent elementary calcium signalling units. When extracellular change is strong and persistent, this localized signal can propagate as a regenerative calcium wave through cytosol. It is known as calcium spike. One such spike may be followed by a series of further spikes (each one lasting for some seconds). These oscillations can persist for as long

Table 1. List of modules in calcium signalling pathway of *H. sapiens* for different *c*-value

	c = 1		*c* = 2		*c* = 3,4		*c* = 5		*c* = 6		*c* = 7	
module name	node	rel	node	rel	node	rel	node	rel	node	rel	node	rel
(C00076)2	24	23	25	24	29	28	40	43	46	51	54	59
CALML6	08	07	08	07	08	07	08	07	08	07		
(C00076)1	06	05	07	06	07	06	06	05				
C01245	02	01	09	08	10	12						
C00165	01	Nil	01	Nil								
BST1	04	03	04	04								
PLCE1	02	01										
PLCG1	02	01										
PLCB1	03	02										
PLCD3	01	Nil										
RYR1	01	Nil										

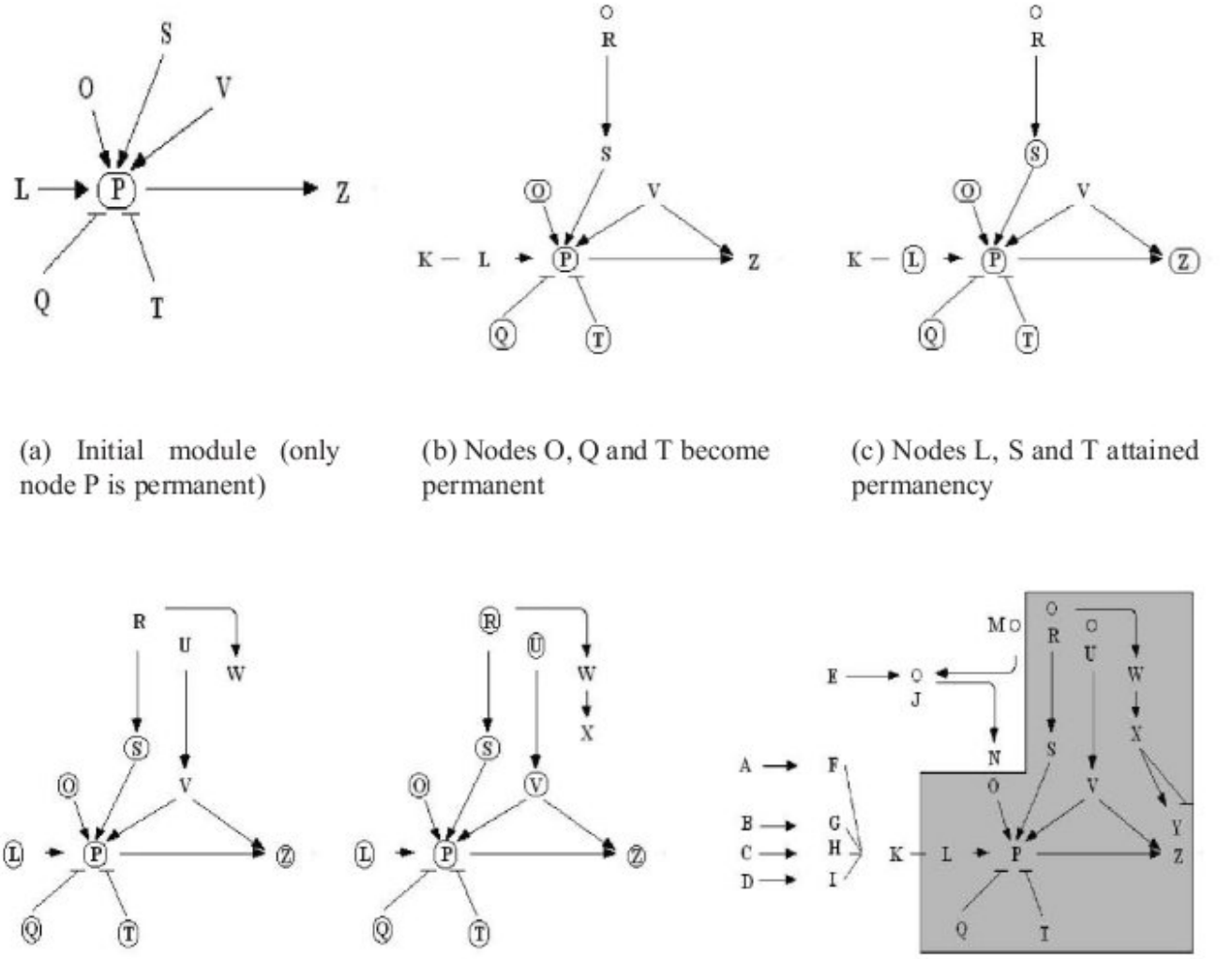
as receptors are activated at the cell surface. Ca^{2+} behaviour is peculiar in the sense that "released calcium ions initially stimulate more self-release, a process known as calcium-induced calcium release, but as its concentration gets high, inhibition of further Ca^{2+} release prevails" (Alberts *et al* 2002).

Ca^{2+} gradients within cells have been proposed to initiate cell migration, exocytosis, lymphocyte killer cell activity, acid secretion, transcellular ion transport, neurotransmitter release, gap junction regulation and numerous other functions (Tsien and Tsien 1990). Calcium is essential for cell growth and survival. Ca^{2+} affects the cell cycle in more than one way: depletion of the *InsP*3 receptor-gated Ca^{2+} pool results in cell cycle arrest at G0/G1 and S phases and Ca^{2+} is necessary and sufficient for resumption of meiosis in marine eggs; a spike of Ca^{2+} triggers completion of meiosis and initiation of mitosis (Means 1994). It was also found that the process of gene transcription depends on how Ca^{2+} enters the cell. Ca^{2+} entry through voltage-dependent L type Ca^{2+} channels and N-methyl-D-aspartic acid (NMDA) receptors initiates gene transcription through distinct DNA-regulatory elements (Bading *et al* 1993). Cellular Ca^{2+} levels quantitatively correlate with level of expression of transcription factors in single cells (Negulescu *et al* 1994). Calreticulin, a molecule previously thought to be a buffer protein, appears to regulate the glucocorticoid nuclear hormone receptor (Burns *et al* 1994). Increase in intranuclear Ca^{2+} initiates gene expression and cell cycle progression, but also can activate degradative processes in programmed cell death or apoptosis. Prolonged existence of high Ca^{2+} ion concentration activates nucleases that cleave DNA and degrade cell chromatin. Ca^{2+} dependent proteases, phosphatases and phospholipases break DNA, resulting in a loss of chromatin structural integrity (Nicotera *et al* 1994).

3. Algorithm

The proposed algorithm starts with detection of a node having maximum number of relations in the node pool of a given network. Considering the detected node as the starting point (the starting member is always a permanent member), an initial module is created where the central node is either a *predecessor* or *successor*. Thus the module is *extended* by including these nodes. Here an eventuality can arise where more than one node may have maximum number of relations. Then any one of the nodes (having maximum number of relations) that is encountered first by the algorithm is taken as the start point by default (followed by the others).

Once a module is initialized, the total number of relations of every individual member is considered. For a node in a module, if the number of relations lying inside the module is equal to total number of relations associated with the node, the member is considered to be *permanent*. If a node in a module has more than c relations that lie outside the module, it gets *excluded* from the module and hence the previous *nonpermanent nodes'* total relations decrease by 1. These *extension* and *exclusion* continue till there is no new node, or a node is present under consideration i.e. all the nodes of a module get declared as permanent. It is to be mentioned here that once a member is declared permanent, it is no more in the node pool. Hence a single member can not be included in more than one module. Also, if a member appears more than once in a network, its positional significance is taken into account. That is, if a member X is present four times in a network, it will be considered four times as X1, X2, X3 and X4. After successful completion of the creation of a module, the algorithm will search for another starting point and repeat the above mentioned steps to create another module. This process will continue till exhaustion of all the nodes



(a) Initial module (only node P is permanent)

(b) Nodes O, Q and T become permanent

(c) Nodes L, S and T attained permanency

(d) Node K is excluded from the module along with addition of two nodes

(e) Nodes R, U and V are declared permanent

(f) Module *P* (marked with gray color) in the whole network

Figure 4. Various stages in construction of module *P*.

present in the node pool. The process is described with an example network.

The hypothetical network in figure 1 is considered for generation of modules. The network contains 26 nodes and 26 relations existing among the nodes. The relations are of different kinds, i.e. activation (a), binding (b), indirect effect (d), inhibition (i), etc. Each member of the set given below shows the relation between a node and its following node. The term (A,F,a) denotes existence of activation relation from node A to node F. Thus we have a set N_R of members

N_R = {(A,F,a), (B,G,a), (C,H,a), (D,I,a), (F,K,b), (G,K,b), (H,K,b), (I,K,b), (K,L,b), (L,P,a), (Q,P,i), (T,P,i), (P,Z,a), (E,J,a), (J,N,a), (M,J,a), (O,P,a), (R,S,a), (S,P,a), (R,W,a), (W,X,a), (X,Y,a), (X,Y,i), (U,V,a), (V,P,a), (V,Z,a)}

We have to calculate total number of relations for all the nodes present in the network in order to choose the node with maximum number of relations as the starting point of

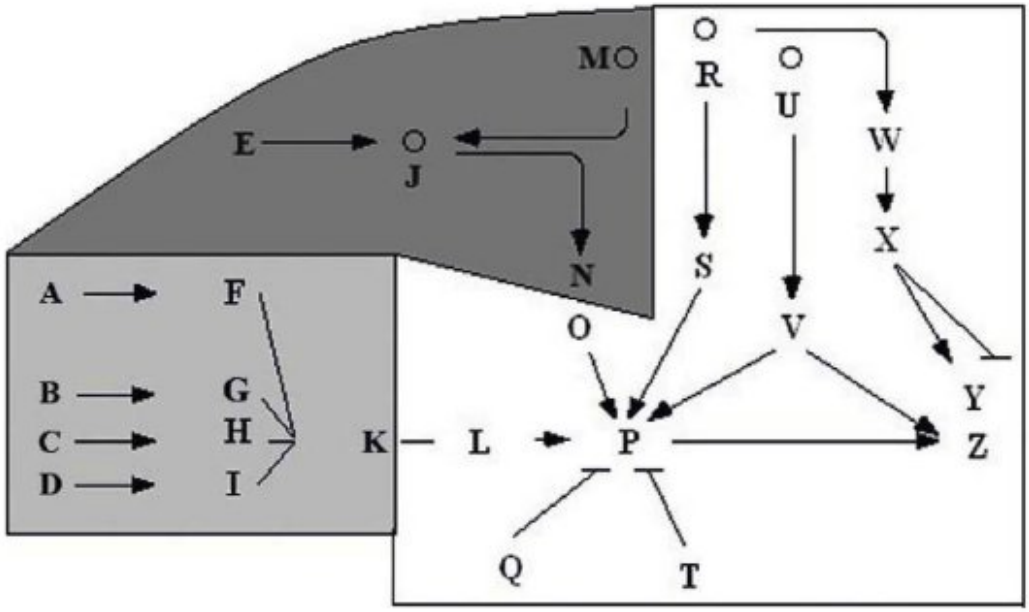


Figure 5. Modularized network for $c = 2$, dark gray part - module J, white part - module P and light gray part - module K.

an originating module. Total number T of relations of a node can be calculated by summing relations of a node first as predecessor and then successor.

$T_A = 1+0 = 1$, $T_B = 1+0 = 1$, $T_C = 1+0 = 1$, $T_D = 1+0 = 1$, $T_E = 1+0 = 1$, $T_F = 1+1 = 2$, $T_G = 1+1 = 2$, $T_H = 1+1 = 2$, $T_I = 1+1 = 2$, $T_J = 1+2 = 3$, $T_K = 1+4 = 5$, $T_L = 1+1 = 2$, $T_M = 1+0 = 1$, $T_N = 1+1 = 2$, $T_O = 1+0 = 1$, $T_P = 1+6 = 7$, $T_Q = 1+0 = 1$, $T_R = 2+0 = 2$, $T_S = 1+1 = 2$, $T_T = 1+0 = 1$, $T_U = 1+0 = 1$, $T_V = 2+1 = 3$, $T_W = 1+1 = 2$, $T_X = 2+1 = 3$, $T_Y = 2+0 = 2$, $T_Z = 0+2 = 2$.

Here node P is having the highest number of relations among the nodes of the given network. So P is the starting point of the module. After the first extension, with immediate neighbors, the module resembles figure 4a. Now we describe below the steps for determining the modules of the network. Here c is taken as 2 and we always named the modules after their starting node.

(i) All relations of Q, T and O lie in the created module. So they became permanent members as shown in figure 4b.

(ii) After second extension (figure 4c), respective relations of L, S and Z lie in the module. Hence they were also considered as permanent members of the module. Node K has more than 2 relations that lie outside the present module. So K cannot be a member of module P (figure 4d).

(iii) After third extension as shown in figure 4e, for V, R and U, their corresponding relations are inside the module. So except W, every member present in the module became permanent, and they were excluded from the node pool of the network.

(iv) Fourth extension made W permanent. Like wise after fifth and sixth extensions, all the members are permanent. Hence creation of the first module was complete. Module P (figure 4f) contains 13 permanent members.

(v) The whole process is again repeated taking K, i.e. the node with maximum relations from the left over nodes present in the node pool.

(vi) The node pool became a null set, after creation of three modules namely P, K and J. The modularized entire network is given in figure 5.

4. Modularization of human calcium signalling pathway

In this section our algorithm is applied to calcium signalling pathway of *H. sapiens*. The data is taken from KEGG/ Pathway database (*http://www.genome.jp/kegg/pathway.html#environmental*) (Ogata *et al* 1999; Kanehisa and Goto 2000). We have considered XML files representing the KGML (KEGG Markup Language) layouts for calcium signalling pathways. The pathway contains 55 nodes. One node *(C13050)* is isolated. So node pool for this network contains 54 nodes. These 54 nodes are having 59 relations among them as shown in figure 6. Modules are created from

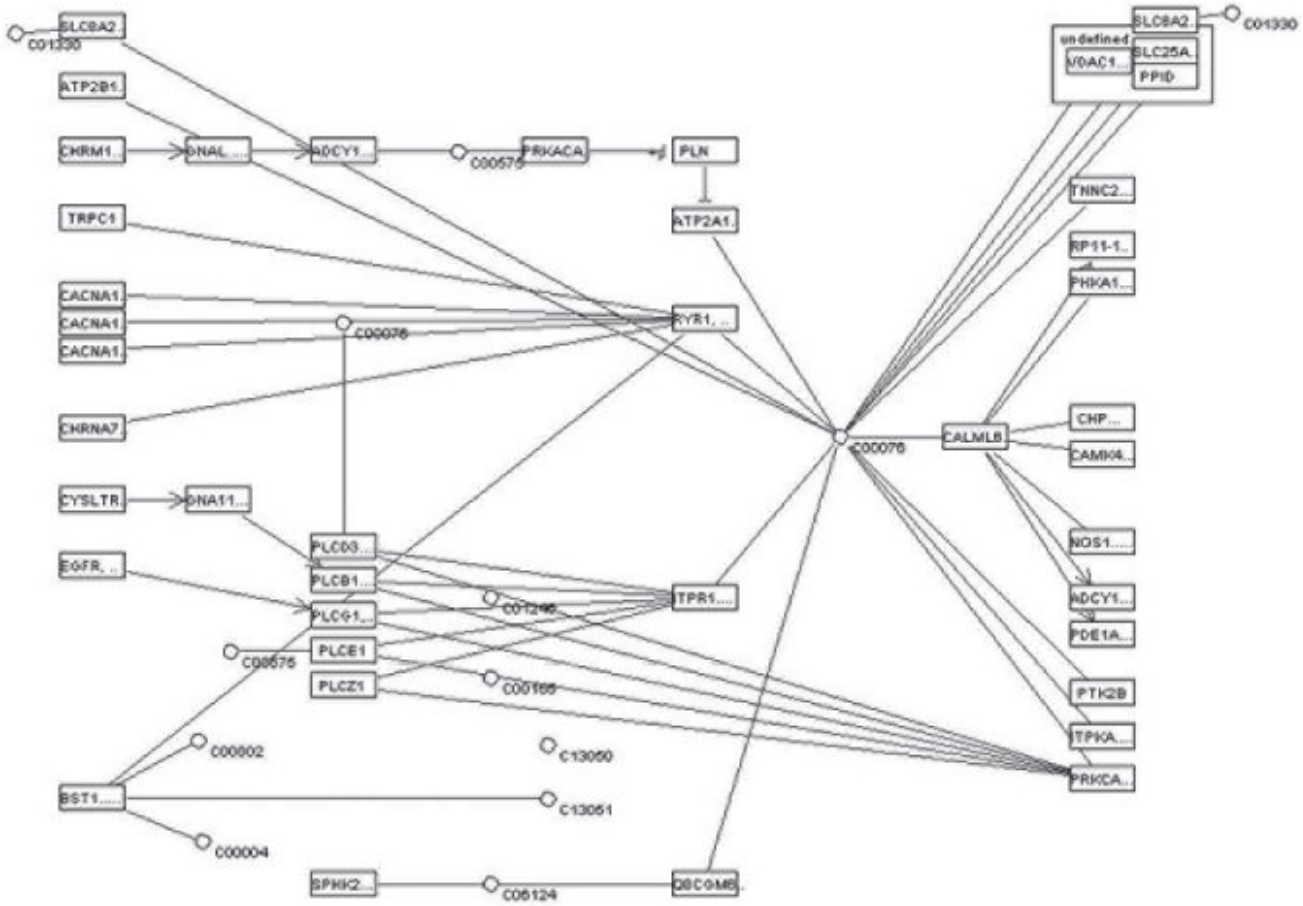


Figure 6. KGML layout for calcium signalling pathway of *H. sapiens*.

the same pathway at complexity level of 1, 2, 3, 4, 5, 6 and 7. For $c = 7$, the whole network emerges into a single module, so there is no need to consider the value of c beyond 7. The results are given in table 1.

For $c = 1$, we get 11 modules. Module *(C00076)2* emphasizes role of plasma membrane, endoplasmic reticulum, mitochondria and nucleoplasmic reticulum in calcium ion balance of cellular environment. But some receptors like RYR (Ryanodine receptors) present in ER membrane are not included in this module. *CALML6* represents role of calmodulin like proteins (calcium binding proteins) that upon binding with free calcium ions change conformation and trigger other enzymes and ion channels. *(C00076)1* module contains calcium channels present in plasma membrane for import purpose. Module *BST1* deals with calcium ion flow from outside to inside of bone marrow cells but how its intracellular balance is maintained is not clear in the module. For $c = 1$, the network is splitting profusely. Excessive splitting is giving rise to a lot of small modules. This is the reason why we are unable to assign any biological significance to the rest seven modules. So modularization of the same network is done for $c = 2$.

For $c = 2$, 6 modules are obtained. Module *CALML6* and *BST1* remain unchanged. In module (C00076)*2* Ryanodine receptors are included, making a clear picture of overall calcium ion flow and balance in a cell. *(C00076)1* is increased by one node PLCD3. The changed module includes plasma membrane based calcium import channels and interaction of the imported calcium ions with one of the phospholipase C (PLC) group members. *C00125* module shows activation of proteins belonging to PLC family and their relation with C01245. From prior knowledge we know PLC group members break into C01245 as a result of activation. C01245 molecule is a ligand for ITPR1 (inositol 1,4,5triphosphate receptor, type 1) present in ER membrane. This module is formed due to convergence of modules *C01245*, *PLCB1*, *PLCG1* and *PLCE1* found for $c = 1$. So the problem of over splitting noticed for $c = 1$ is reduced here. But still we are left with the problem of module *BST1* as described above and a small module *C00165* comprising of a single node to

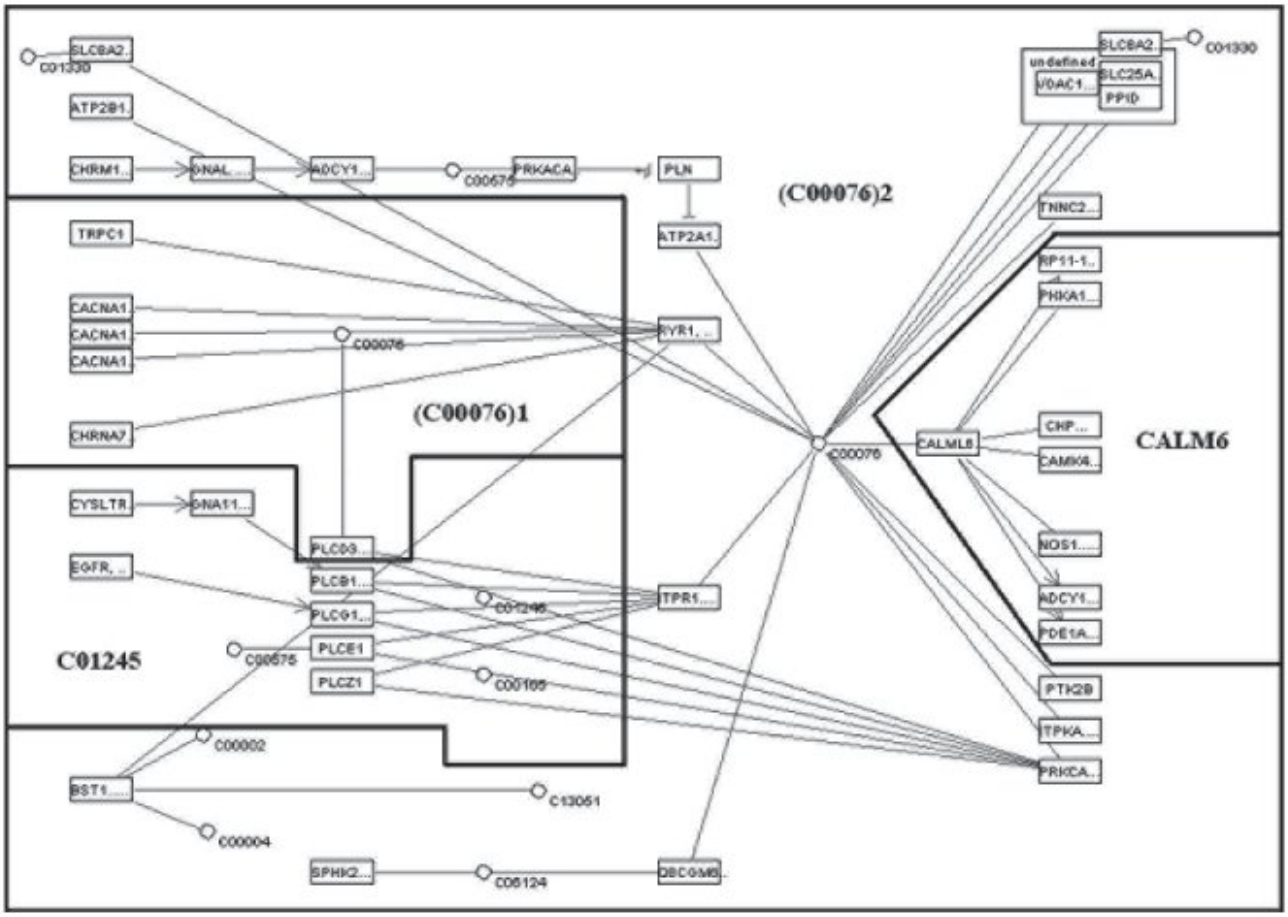


Figure 7. KGML layout for calcium signalling pathway of *H. sapiens* after modularization for $c = 3$ and 4 (black lines separate the modules).

describe. Its difficult to analyze such small modules. This lead to modularization of the network for $c = 3$.

Four modules are created for $c = 3$ as shown in figure 7. Module *BST1* for $c = 2$ is merged with *(C00076)2* that gives a complete explanation of calcium ion balance in bone marrow cells through ryanodine receptors present in ER membrane. Module *C00165* for $c = 2$ is merged with module *C01245*. C00165 is a byproduct when PKC group members break into C01245. Like C01245, it is not a ligand for ITPR1. It binds with PKC (protein kinase C) that takes part in controlling PM based calcium ion channels. But we were able to decipher its role clearly only after its emergence with module *C01245*. For $c = 2$, where C00165 is included in another module, it is confusing to decipher and understand this information. We are getting exactly similar modules for $c = 4$.

Now question arises once we get biologically significant modules at some value of c, whether we should proceed further and continue modularization at higher values of c to get more meaningful modules or stop the process. To get a logical answer we obtained modules for $c = 5$ and 6. For $c = 5$, we get three modules. Module *(C00076)2* is increased by several nodes and relations, that make it large and complex, hence our primary objective of dividing a complex network to simpler units is failed here. *(C00076)1* module is decreased by one node and one relation that again gives rise to the already discussed problem of calcium ion balance inside the cell. Module *CALML6* remains unchanged. For $c = 6$, we are getting only 2 modules. Module *CALML6* still remains unchanged and the rest part of the network gets included in the module *(C00076)2*. In quest of a better solution, it only aggravates our problem. For $c = 7$, the whole network rounds up to a single module.

So in general we can assume after a certain level, modularization with increasing c-value will yield similar results with that of previous complexity level or the modules will be enough large making their study and analysis difficult. As our objective is simplified study of a network and we are getting approximately biological significant modules at $c = 3$, we fixed c-value to 3 for calcium signalling

pathway. This value of c is used later for analysing calcium signalling pathways for further study.

5. Discussion

In this paper an algorithm for modularization of signal transduction pathways is described and the same is applied to calcium signalling pathway of *H. sapiens* for better study and analysis of the pathway. The method finds various modules from the pathway for different complexity values. We have successfully conferred biological significance to modules obtained from human calcium signalling pathway for c-value of 3. This validates effectiveness of the algorithm. The study is being continued with calcium signalling pathways of other species.

References

Alberts B, Johnson A, Lewis J, Raff M, Roberts K and Walter P 2002 *Molecular biology of the cell* Fourth edition (Garland Science)

Bading H, Ginty D D and Greenberg M E 1993 Regulation of gene expression in hippocampal neurons by distinct calcium signalling pathways; *Science* **260** 181–186

Baimbridge K G, Celio M R and Rogers J H 1992 Calcium binding proteins in the nervous system; *Trends Neurosci.* **15** 303–308

Berridge M J 1997 Elementary and global aspects of calcium signalling; *J. Physiol.* **499** 291–306

Burns K, Duggan B, Atkinson E A, Famulski K S, Nemer M, Bleackley R C and Michalak M 1994 Modulation of gene expression by calreticulin binding to the glucocorticoid receptor; *Nature (London)* **367** 476–480

Clapham D E 1995 Calcium signalling; *Cell* **80** 259–268

Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita K F, Itoh M, Kawashima S, Katayama T, Araki M and Hirakawa M 2006 From genomics to chemical genomics: new developments in KEGG; *Nucleic Acids Res.* **34** D354–D357

Echevarria W, Leite M F, Guerra M T, Zipfel W R and Nathanson M H 2003 Regulation of calcium signals in the nucleus by a nucleoplasmic reticulum; *Nat. Cell Biol.* **5** 440–446

Golumbfskie A J, Mundy C J, Kubota A K, Nichols A and Quong A A 2003 A three-dimensional model of intercellular calcium signalling in epithelial cells; *Proceedings of the 25th Annual International Conference of the IEEE EMBS,* 17–21 September, Cancun, Mexico

Heizmann C W and Hunziker W 1991 Intracellular calcium binding proteins: more sites than insights; *Trends Biochem. Sci.* **16** 98–103

Kanehisa M and Goto S 2000 KEGG: Kyoto Encyclopedia of Genes and Genomes; *Nucleic Acids Res.* **28** 27–30

Levchenko A 2002 Stochastic modeling of spatially localized events in protein kinase A and Ca2+ signalling; *Proceedings of the Second Joint EMBS/BMES Conference,* 23–26 October, Houston, TX, USA

Means A R 1994 Calcium, calmodulin and cell cycle regulation; *FEBS Lett.* **347** 1–4

Negulescu P A, Shastri N and Cahalan M D 1994 Intracellular calcium dependence of gene expression in single T lymphocytes; *Proc. Natl. Acad. Sci.USA* **91** 2873–2877

Nicotera P, Zhivotovsky B and Orrenius S 1994 Nuclear calcium transport and the role of calciumin apoptosis; *Cell Calcium* **16** 279–288

Ogata H, Goto S, Sato K, Fujibuchi W, Bono H and Kanehisa M 1999 KEGG: Kyoto Encyclopedia of Genes and Genomes; *Nucleic Acids Res.* **27** 29–34

Pozzan T, Rizzuto R, Volpe P and Meldolesi J 1994 Molecular and cellular physiology of intracellular calcium stores; *Physiol. Rev.* **74** 595–636

Bertram R, Greenstein J L, Hinch R, Pate E, Reisert J, Sanderson M J, Shannon T R, Sneyd J and Wilson R L 2005 *Tutorials in mathematical biosciences II* (ed.) James Sneyd (Heidelberg: SpringerVerlag)

Tsien R W and Tsien R Y 1990 Calcium channels, stores and oscillations; *Ann. Rev. Cell Biol.* **6** 715–760

Zacharias D A, Dalrymple S J and Strehler E E 1995 Transcript distribution of plasma membrane Ca2+ pump isoforms and splice variants in the human brain; *Mol. Brain Res.* **28** 263–272

*e*Publication: 6 August 2007