TASK QUARTERLY 18 No 4, 321–325

PROTEIN-PROTEIN INTERACTION AND COARSE GRAINED SIMULATION STUDY OF GLIOBLASTOMA MULTIFORME REVEALS NOVEL PATHWAYS OF GPR17

MUTHARASU GNANAVEL[1], OLLI YLI-HARJA[1,2] AND MEENAKSHISUNDARAM KANDHAVELU[1]

[1] *Molecular Signaling Lab, Computational Systems Biology Signal Processing Department, Tampere University of Technology 33101 Tampere, Finland*

[2] *Institute for Systems Biology 1441N 34th St, Seattle, WA 98103–8904, USA*

(Paper presented at the CBSB14 Conference, May 25–27, 2014, Gdansk, Poland)

Abstract: Studies of receptor mediated signaling networks in neuronal cells provide a unique opportunity to uncover the basis of many diseases. Receptor signaling cascades proceed from the cell surface, where extra cellular factors interact with their specific receptors *e.g.* G-Protein Coupled Receptors (GPR). Recent studies have shown that the activation or suppression of GPR17 in diseased neuronal cells has potential impact in altering the tumor conditions. We identified many hundred times expressions of GPR17 in Glioblastoma Multiforme (GBM) from the RNA-Seq data. We also observed many other genes having similar expression patterns with GPR17, indicating possible connections of this receptor with diverse gene products. We performed a coarse-grained simulation of ~ 500 proteins inside a cytoplasm like a box with solvent water molecules. The summarized protein interaction networks resulted from a coarse-grained simulation and large scale protein-protein docking reveals novel molecular connections and pathways.

Keywords: signaling networks, RNA-Seq, coarse-grain, Glioblastoma Multiforme, pathways

1. Introduction

Receptor mediated signaling networks in normal and diseased neuronal cells provide a unique opportunity to uncover the basis of many neurological diseases. It helps in the development of novel therapeutic strategies to treat neurological disorders, neuronal tumors and cancers [1]. Receptor signaling cascades usually proceed from the cell surface, where extra cellular factors interact with their specific receptors e.g. G-Protein Coupled Receptors (GPR). GPR17 is one of the

important receptors reportedly having diverse functionalisms [2, 3]. GPR17 is a seven helical trans-membrane protein encoded by a gene located in chromosome 2. This protein is suggested to have a dual specificity receptor for uracil nucleotides and cysteinyl leukotrienes (CysLTs) and signals through the G(i) protein and the inhibition of adenylyl cyclase. Previous reports have shown the role of this protein in brain damage regulation by nucleotides and CysLTs following ischemia [4]. Diseases associated with GPR17 include suppression amblyopia, and amblyopia. In other GPCR signaling, Rhodopsin-like receptors (Class A/1) are its related super-pathways. Recent studies have shown that the activation or suppression of GPR17 in diseased neuronal cells has potential impact in altering the tumor conditions.

Many neuronal cancer [5] cells showed differing expression levels of GPR17. We observed many (up to many hundred) fold expressions of GPR17 in 169 samples of Glioblastoma Multiforme (GBM) from Next generation sequencing data. We also observed many other genes having similar expression patterns with GPR17, indicating possible connections of this receptor with diverse gene products. It is difficult to retrieve the structure-function relationship of GPR17 with other proteins and diverse small molecules due to the unavailability of an experimentally solved structure. Homology modeling is a potential alternative method to reach the 3D structure of proteins which are hard to crystallize with reasonable accuracy. We developed a 3D structure of GPR17 using comparative modeling which was further used in simulation and docking studies. Further in this study, we exploited the differential expression analysis for the RNA-Seq data to find genes playing a critical role in the disease phenotype, coarse-grained simulation to observe the protein aggregation and finally protein-protein docking methods to verify the molecular interaction. The obtained information helped in identifying novel molecular connections of GPR17 mediated signaling networks. Current study is a key development in understanding the diverse role of this receptor as well as the disease origin and progress of GBM at molecular resolution.

2. Materials and Methods

The GPR17 protein sequence containing 367 amino acids was retrieved from uniprot [6] (Uniprot ID **Q13304**). BLAST search against Protein Data Bank (PDB) [7] with GPR17 resulted in the Turkey beta1 adrenergic receptor (PDB ID: **4AMJ**) as a best template having 32% sequence identity. Query template sequence alignment in the PIR format was loaded into the MODELLER [8] program and 10,000 models of the GPR17 structure were constructed in different energy conformations. An ensemble of models generated from the MODELLER was sorted based on the DOPE score in an ascending order. From the ensemble, model with the lowest DOPE score (-40242.10) was chosen as the best model. Selected model was further fixed by energy minimization using NAMD [9]. The protein model was solvated in a water medium and the temperature was set to 298 K. The simulation parameters were used from par_all27.inp having CHARMM force field

constants for the bond stretches, angle bends and torsional coordinates. The input parameter file had charge and non-bonding information as well. Minimization was done at 2 fs/step using Langevin dynamics up to 500,000 steps (1 nano second). The minimized structure was further analyzed using the Ramachandran Plot, which showed that 282 (96.9%) residues were in the favored region ($\sim 98.0\%$ expected) and 7 (2.4%) residues in the allowed region ($\sim 2.0\%$ expected). The resulting structure was a seven alpha helical trans-membrane protein.

RNA-SeqV2 generated using IlluminaHiSeq for 169 samples Glioblastoma Multiforme (GBM) was retrieved from Cancer Genome Atlas [10]. One control datum of the cerebral cortex gene expression was also included in this data for differential gene expression analysis. DESeq [11] The Bioconductor package was utilized to calculate the differentially expressed genes. Proteins with known structures among these candidate genes were chosen to study the protein-protein interactions causing or resulting from GBM. We used a homology based structure of GPR17, which is used with atomic resolution structures of other proteins to perform a high throughput protein-protein docking experiment.

Significance of interaction and stability of each complex from the docking was analyzed using molecular dynamics simulations. Further we used coarse grained models of all the over expressing proteins with GPR17 to sample the diffusion rate and binding affinities of molecules in the cytoplasm [12]. A microsecond simulation of crowded proteins in water medium was performed using the MARTINI [13] force field with the GROMACS [14] simulation software. We considered the relative copy number and sub-cellular localization of each molecule. Simulating GBM with such complex bio-molecular system captures the molecular details of wide protein-protein interactions. All the protein-protein complexes observed from the simulation and docking were listed and fed into the Cytoscape [15] network analysis platform. The summarized protein interaction networks resulting from this study showed novel molecular connections and pathways.

3. Results

We have identified 13 novel pathways involved/mediated through GPR17 from this comprehensive study (Table 1). At least 6 of those pathways are further confirmed with Gene Ontology annotation evidence to have interaction with GPR17. Genes which are showing functional relevance towards cancer phenotype are explained further here.

1. **MAP3K10**: The mitogen-activated protein kinase kinase kinase 10. This kinase functions preferentially on the JNK signaling pathway, and is reported to be involved in nerve growth factor (NGF) induced neuronal apoptosis. 2. **CASKIN1**: Scaffolding protein CASK to downstream intracellular effectors. 3. **CRKL**: Plays a role in fibroblast transformation by BCR-ABL. In addition, CRKL has oncogenic potential. 4. **EPHB1**: Ephrin receptors and their ligands, the ephrins, mediate numerous developmental processes, particularly in the nervous system. 5. **RIMS1**: Regulating synaptic membrane exocytosis protein 1, a synaptic vesicle protein

Table 1. List of genes identified as interacting partners with GPR17

No.	Gene	Function
1	EPHB1	Neuronally-Expressed EPH-Related Tyrosine Kinase
2	CRKL	V-Crk Avian Sarcoma Virus CT10 Oncogene Homolog-Like
3	CASKIN1	CASK Interacting Protein
4	MAST3	Microtubule-Associated Serine/Threonine-Protein Kinase 3
5	KCNJ4	Potassium Inwardly-Rectifying Channel
6	LMO4	Breast Tumor Autoantigen
7	ABI2	Abelson Interactor 2
8	SH3GL2	SH3-Domain GRB2-Like 2
9	ARHGAP21	Rho GTPase Activating Protein 21
10	PPP1R9A	Protein Phosphatase 1, Regulatory Subunit 9A
11	MYO10	Unconventional Myosin-10
12	RIMS1	Regulating Synaptic Membrane Exocytosis 1
13	MAP3K10	Mitogen-Activated Protein Kinase Kinase Kinase 10

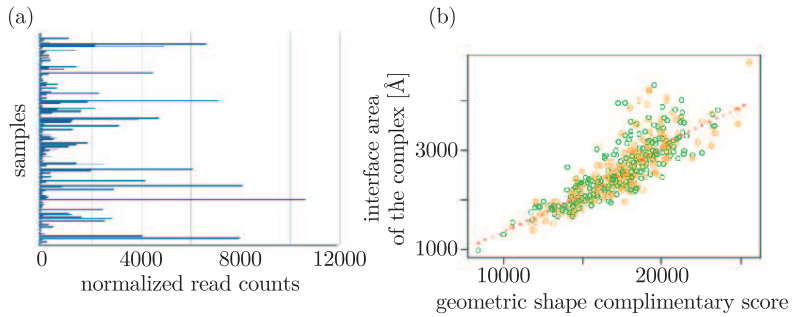

Figure 1. (a) Expression levels of GPR17 across 169 samples of GBM; (b) Correlation of interacting area against the binding energy score calculated for the positive entries selected from protein-protein docking; proteins used here are retrieved from differential expression analysis

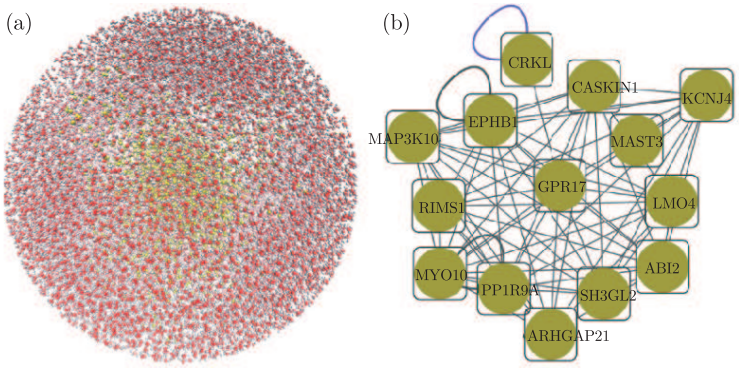

Figure 2. (a) Simulation box containing ~ 500 proteins in coarse-grained representation with water molecules surrounded at least for 5 angstrom distance. (b) Final network showing protein-protein docking constructed by integrating and simulation showing molecular connections of GPR17

that regulates synaptic vesicle exocytosis. Scaffolds formed by this protein help to regulate vesicle exocytosis during short-term plasticity. 6. **ARHGAP21**: Rho GTPase activating protein 21, diseases associated with this gene includes influenza, and glioblastoma. GO annotations related to this gene include phospholipid binding and GTPase activator activity.

Kinases are the major contributing class of proteins showing molecular interaction possibilities with this receptor. This indicates the presence of wide and complex signaling networks altered by the disease phenotype. Ion channels and phophatases are coming next to kinases in the network of disease triggered signals. The functional relevance of gene products interacting with GRP17 identified from this study demonstrates the efficiency of our novel approach in reaching the systems biological map of potential receptors using computational biophysical methods.

References

[1] Hennen S, Wang H, Peters L, Merten N, Simon K, Spinrath A, Blättermann S, Akkari R, Schrage R, Schröder R, Schulz D, Vermeiren C, Zimmermann K, Kehraus S, Drewke C, Pfeifer A, König G M, Mohr K, Gillard M, Müller C E, Lu Q R, Gomeza J and Kostenis E 2013 *Sci. Signal.* **6**, ra93

[2] Lecca D, Trincavelli M L, Gelosa P, Sironi L, Ciana P, Fumagalli M, Villa G, Verderio C, Grumelli C, Guerrini U, Tremoli E, Rosa P, Cuboni S, Martini C, Buffo A, Cimino M and Abbracchio M P 2008 *PloS One* **3** (10), e3579

[3] Benned-Jensen T and Rosenkilde M M 2010 *Br. J. Pharmacol.* **159** (5) 1092

[4] Marucci G, Lammi C, Buccioni M, Dal Ben D, Lambertucci C, Amantini C, Santoni G, Kandhavelu M, Abbracchio M P, Lecca D, Volpini R and Cristalli G 2011 *Anal. Biochem.* **414** 300

[5] Boda E, Viganò F, Rosa P, Fumagalli M, Labat-Gest V, Tempia F, Abbracchio M P, Dimou L and Buffo A 2011 *Glia* **59** (12) 1958

[6] Apweiler R, Bairoch A, Wu C H, Barker W C, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin M J, Natale D A, O'Donovan C, Redaschi N and Yeh L-S L 2004 *Nucleic Acids Res.* **32**, D115

[7] Berman H M, Westbrook J, Feng Z, Gilliland G, Bhat T N, Weissig H, Shindyalov I N and Bourne P E 2000 *The Protein Data Bank*, *Nucleic Acids Res.* **28** 235

[8] Sali A and Blundell T L 1993 *Comparative protein modelling by satisfaction of spatial restraints*, *J. Mol. Biol.* **234** 779

[9] Phillips J C, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel R D, Kalé L and Schulten K 2005 *Scalable molecular dynamics with NAMD* **26**

[10] Collins F S 2007 *The Cancer Genome Atlas (TCGA)*

[11] Anders S and Huber W 2010 *Genome Biol* **11**, R106

[12] McGuffee S R and Elcock A H 2010 *PLoS Comput. Biol.* **6**

[13] Marrink S J, Risselada H J, Yefimov S, Tieleman D P and De Vries A H 2007 *J. Phys. Chem. B* **111** 7812

[14] Pronk S, Páll S, Schulz R, Larsson P, Bjelkmar P, Apostolov R, Shirts M R, Smith J C, Kasson P M, Van Der Spoel D, Hess B and Lindahl E 2013 *Bioinformatics* **29** 845

[15] Shannon P, Markiel A, Ozier O, Baliga N S, Wang J T, Ramage D, Amin N, Schwikowski B and Ideker T 2003 *Genome Res.* **13** 2498