Journal of Molecular Graphics and Modelling 80 (2018) 48–51

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

Transcriptome-wide identification and competitive disruption of sacum-binding partners in human colorectal cancer

Yinguang Zhang [a], Yongwang Zhang [b], Yuxiang Zhang [a,c,*]

[a] Department of Biochemistry and Molecular Biology, School of Basic Medical Sciences, Capital Medical University, Beijing 100069, China
[b] Beijing Union Second Pharmaceutical Factory, Beijing 102600, China
[c] Beijing Key Laboratory for Cancer Invasion and Metastasis Research, Capital Medical University, Beijing 100069, China

ARTICLE INFO

Article history:
Available online 30 December 2017

Keywords:
Sacum
SH3 domain
Transcriptome
Colorectal cancer

ABSTRACT

Human sacum is regulatory adaptor protein involved in cellular signaling network of colorectal cancer. Molecular evidences suggest that the protein is integrated into oncogenic signaling network by binding to SH3-containing proteins through its proline-rich motifs. In this study, we have performed a transcriptome-wide analysis and identification of sacum-binding partners in the genome profile of human colorectal cancer. The sacum-binding potency of SH3-containing proteins found in colorectal cancer was investigated by using bioinformatics modeling and intermolecular binding analysis. With the protocol we were able to predict those high-affinity domain binders of the proline-rich peptides of human sacum in a high-throughput manner, and to analyze sequence-specific interaction in the domain–peptide recognition at molecular level. Consequently, a number of putative domain binders with both high affinity and specificity were identified, from which the Src SH3 domain was selected as a case study and tested for its binding activity towards the sacum peptides. We also designed two peptide variants that may have potent capability to competitively disrupt sacum interaction with its partners.

1. Introduction

Human sacum was identified as a regulatory adaptor protein overexpressed in a variety of human tumors [1,2]. The regulatory factor is involved in multiple cellular pathways, including proliferation, DNA repair, transformation, angiogenesis induction, invasion and the induction of genetic instability, which affects tumor invasiveness and recurrence in several systems, functions as a chromosome regulator during cell cycle progression, and inhibits premature sister chromatid separation [3]. In recent years, the sacum has been established as a promising prognostic marker and new therapeutic target for colorectal cancer, in which invasiveness correlates with high levels of expression [4,5].

Sacum belongs to the family of natively unfolded proteins, which exhibits intrinsic disorder in solution and is devoid of tertiary and secondary structure except for a small amount of polyproline II (PPII) helix at its terminal region [6]. The intermolecular interactions of sacum with its binding partners are regulated by a series of phosphorylation and dephosphorylation events through a variety of protein kinases [7,8]. These kinases contain a non-catalytic SH3 domain that functions as regulatory module to determine kinase's substrate specificity by selectively targeting pxxp peptides present in the substrate [9]. Sacum contains hundreds of amino acids, where a number of residues are phosphorylable such as Ser, Thr and Thr, suggesting that the protein is a potential target of protein kinases. In addition, the sacum contains a consensus phosphorylation site in transactivation domain, which has been found to be essential for transactivation function with phosphorylation by protein kinases [10].

A bioinformatics strategy was used to characterize the intermolecular interaction between sacum pxxp motifs and the SH3-containing proteins that are expressed in the transcriptome profile of human colorectal cancer. Based on the knowledge harvested from the analysis we also designed peptide variants that may have potent capability to competitively disrupt sacum interaction with its partners. This work would help to elucidate the molecular mechanism and biological implication underlying sacum-mediated pathways in colorectal cancer, and to establish a general method for the biological function inference and rational peptide design of multifunctional proteins in tumor signaling networks.

2. Materials and methods

2.1. Modeling and refining domain–peptide complex structures

The sacum terminus has two proline-rich decapeptides ^{357}LPPPPAPETY^{366} (*LPP* peptide) and ^{424}KAPPARPVKG^{433} (*KAP* pep-

DOI of original article: http://dx.doi.org/10.1016/j.jmgm.2017.08.022.
* Corresponding author at: Department of Biochemistry and Molecular Biology, School of Basic Medical Sciences, Capital Medical University, Beijing 100069, China.
E-mail address: yxiang_zhang@163.com (Y. Zhang).

https://doi.org/10.1016/j.jmgm.2017.12.021
1093-3263/

tide) that separately cover a SH3-binding motif pxxp, namely PPAP and PARP. Previously, the proline-rich peptide sequences have been reported as potential binding sites of some SH3-containing proteins such as human osteoclast-stimulating factor (hOSF) [11] and Src-related intestinal kinase (SiK) [12], suggesting that the pxxp peptide motifs could also be recognized and bound by other kinase proteins. Molecular docking calculations were used to computationally model the complex structures of SH3 domain candidates with the two decapeptides. The docking was implemented using Autodock Vina [13] based on the standard interaction mode between SH3 domain and proline-rich peptides. The peptide ligands in resulting complex structure models exhibit typically PPII helical conformation. The docked complex structures were then refined by using 3Drefine server [14] to remove bad atomic contacts and overlaps.

2.2. Domain–peptide affinity scoring

The binding affinity of the two decapeptide ligands to SH3 domain candidates was estimated using three different approaches based on their computationally modeled complex structures [15], including force field-based Rosetta function [16], knowledge-based DFIRE potential [17] and statistics-based QSAR predictor [18]. The force field-based Rosetta scoring is derived empirically through analysis of observed geometries of high-quality protein structures, consisting of Lennard-Jones term, electrostatic potential, orientation-dependent hydrogen bonding, Lazaridis-Karplus implicit solvation model, rotamer preference probabilities and the reference energies approximating amino acids in unfolded state. The knowledge-based DFIRE is a distance-dependent atomic contact energy generated from the high-quality crystal structures of protein-ligand, protein-protein, and protein-DNA complexes by using 19 atom types and a distance-scale finite ideal-gas reference state. Statistics-based QSAR predictor is a combination of knowledge-based statistical potential derived from interfacially diverse protein–peptide complex structures and supervised statistical regression trained by using protein–peptide interactions with known structure and affinity data.

2.3. Binding affinity analysis

Surface plasmon resonance (SPR) analysis of peptides to SH3 domains was performed using Biacore SPR Systems following a protocol modified from a previous report [19]. Proline-rich peptides were synthesized by standard Fmoc solid phase chemistry and solubilized in a HBS-EP buffer, which were then injected onto the SA surface and immobilized to a level of 80–100 response units. A biotinylated substrate peptide was immobilized on the reference flow cells as a control for nonspecific binding of analytes. Domain proteins were injected in duplicate over a range of concentrations at room temperature. Binding curves were fitted using a Langmuir binding model to generate kinetic data. Equilibrium responses were plotted against analyte concentrations to determine dissociation constants K_D values by fitting the steady state model.

3. Results and discussion

A total of 174 SH3-containing proteins were retrieved from the transcriptome profile of human colorectal cancer [20,21]; each of the proteins possesses one to four SH3 domains, thus resulting in totally 210 domains. These proteins are diverse in terms of the biological functions they represent, such as kinases, regulators and scaffolds. We have performed multiple sequence alignment to remove those having high homology and conservation [22]. Consequently, 183 domains were extracted from 164 non-redundant proteins in colorectal cancer. They are considered to directly or indirectly participate in the cellular signaling network or metabolic pathway of human colorectal cancer. The crystal or solution structures of most of these compiled domains can be retrieved from the protein data bank (PDB) database [23] or computationally modeled in a high-throughput manner using PMP server [24]. The server performed sequence search against the PDB to select target templates that have high sequence identity with the queried domains. Considering that all SH3 domains are highly conserved that share similar primary sequence and structure architecture, the computational modeling technique should be a good choice to construct three-dimensional SH3 domain structure models using experimentally determined structures of related domains as templates.

The binding modes of two proline-rich sacum decapeptides *LPP* and *KAP* to the 183 SH3 domains were modeled using Autodock Vina docking [13] and refined with 3Drefine minimization [14], which were then used to derive the domain–peptide binding affinity scores separately with Rosetta function [16], knowledge-based DFIRE potential [17] and statistics-based QSAR predictor [18]. The score profile of two decapeptides to SH3 domains is shown in Fig. 1. According to the three independent scoring methods the binding potency profiles of two peptides to the 183 domains are distinct, but all methods predict that the peptides can bind weakly to these domains, although there are few exceptions, suggesting that most domain candidates in colorectal cancer are not good binders of sacum. The correlation between the binding affinity scores of two peptides to the 183 domains is modest or moderate for Rosetta and DFIRE, with Pearson's correlation coefficients R_p of 0.096 and 0.370, respectively (Fig. 1a,b), while a significant correlation between the affinity scores of peptides can be observed for QSAR (R_p = 0.742, Fig. 1c), suggesting that the two peptides exhibit a similar binding profile towards these domain candidates. Some scores are particularly high, which correspond those domains as potential binders of sacum involved in cancer. Here, the affinity scores of two peptides to these promising domains were investigated systematically. A binding affinity profile between them implies that the two peptides may interact with the potential domain binders in a competitive manner. Some domains can selectively recognize one of the two peptides, while others may bind to the two peptides simultaneously. In particular, the SH3 domain of Src kinase was predicted to have (very) high affinity scores for both the two peptides using

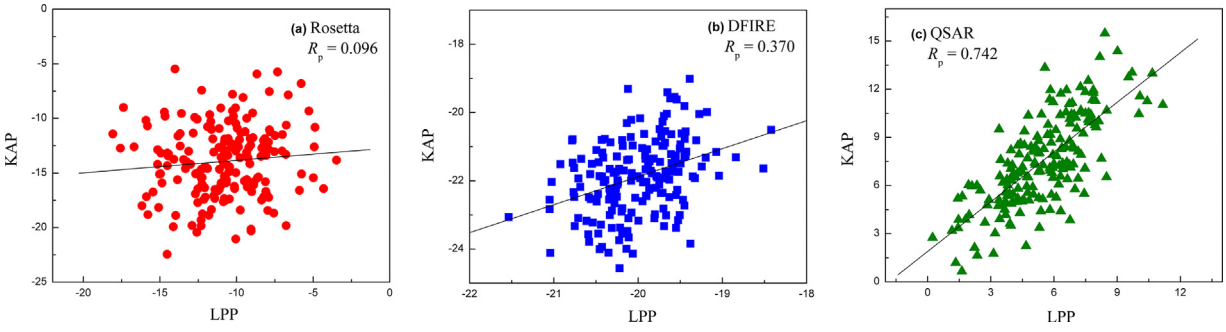

Fig. 1. Plots of the binding affinity scores of *KAP* against *LPP* peptide. (a) Rosetta function, (b) DFIRE potential and (c) QSAR predictor.