FEBS Letters 586 (2012) 2586–2596

journal homepage: www.FEBSLetters.org

Review

The application of modular protein domains in proteomics

Joshua A. Jadwin, Mari Ogiue-Ikeda, Kazuya Machida *

Department of Genetics and Developmental Biology, Raymond and Beverly Sackler Laboratory of Genetics and Molecular Medicine, University of Connecticut Health Center, 400 Farmington Avenue, Farmington, CT 06030, USA

ARTICLE INFO

Article history:
Received 24 March 2012
Revised 13 April 2012
Accepted 13 April 2012
Available online 21 April 2012

Edited by Marius Sudol, Gianni Cesareni, Giulio Superti-Furga and Wilhelm Just

Keywords:
Modular protein domain
Proteomics
Domainomics
Motif scanning
Domain scanning
Multiplex scanning

ABSTRACT

The ability of modular protein domains to independently fold and bind short peptide ligands both in vivo and in vitro has allowed a significant number of protein–protein interaction studies to take advantage of them as affinity and detection reagents. Here, we refer to modular domain based proteomics as "domainomics" to draw attention to the potential of using domains and their motifs as tools in proteomics. In this review we describe core concepts of domainomics, established and emerging technologies, and recent studies by functional category. Accumulation of domain–motif binding data should ultimately provide the foundation for domain-specific interactomes, which will likely reveal the underlying substructure of protein networks as well as the selectivity and plasticity of signal transduction.

1. Introduction

Eukaryotic proteins are modular in nature. Many proteins contain independently folding globular domains capable of binding short peptide motifs even when both domain and motif are removed from the context of their full-length protein [1,2]. Modular protein interacting domains facilitate protein–protein interactions required for a diverse set of cellular processes including signal transduction and subcellular localization. Domains are categorized based on structural and sequence homology, with each domain family recognizing motifs with similar characteristics, such as phosphorylated tyrosine (pTyr) or proline rich sequences (Table 1). The combination of modular domains within a protein contributes to its biological function by defining its protein interaction network. The post-translational modification (PTM) of amino acid side chains within a peptide motif can modulate domain–motif binding, providing the basis for the elegant and complicated protein signaling networks required for life [3]. Over the past decade, exploitation of a number of high-throughput proteomic technologies including increasingly sensitive mass spectrometry (MS) and protein microarrays has led to the dissection of vast protein interaction networks and the role of PTMs in altering network topology [4]. Accurate quantification of protein–protein interactions and PTM is now possible [5].

Because of the role of modular domains in assembling protein complexes, a significant portion of proteomics studies take advantage of modular domains as a means to assess protein–protein interactions (e.g., as bait in pull-down or probes in microarray). Here we refer to this modular domain-based proteomics as "domainomics." While this may be a somewhat artificial segmentation, it is meant to draw attention to the potential of domains and their motifs as tools in contemporary proteomics. The emergence of domainomics as a unique sub-genre of proteomics raises a number of important questions: (1) What are the characteristics of modular domains as a research tool? (2) How are assays tailored to address specific scientific questions? (3) What technologies are available for exploiting domains as tools? (4) What lessons can be

Abbreviations: pTyr, phosphotyrosine; PTM, post-translational modification; MS, mass spectrometry; PSSM, position specific scoring matrix; SH2, Src homology 2; SH3, Src homology 3; WW, tryptophan tryptophan; PDZ, post synaptic density protein 95-disks large protein 4-zonula occludens 1; FHA, forkhead-associated; PH, pleckstrin homology; PHD, plant homeo domain; PTB, phosphotyrosine binding; FACS, fluorescence-activated cell sorting; pBpa, *p*-benzoyl-L-phenylalanine; LC–MS/MS, liquid chromatography followed by tandem mass spectrometry; Y2H, yeast-two-hybrid; ELISA, enzyme linked immunosorbent assay; IRS-1, insulin receptor substrate 1; PKC, protein kinase C; SILAC, stable isotope labeling by amino acids in cell culture; EGF, epidermal growth factor; EGFR, epidermal growth factor receptor; EVH, Ena/Vasp homology; GYF, glycine-tyrosine-phenylalanine; GST, glutathione S-transferase; BRCT, BRCA1 C-terminus; BCR, breakpoint cluster; ATM, ataxia telangiectasia mutated; MDC1, mediator of DNA damage checkpoint protein 1; pS, phosphoserine; pT, phosphothreonine; PTIP, pax interacting protein 1; ROC, receiver operating characteristic; MBT, malignant brain tumor; PWWP, proline-tryptophan-tryptophan-proline; WD40, tryptophan-aspartic acid 40; SPOT, peptide synthesis on membrane; STAT, signal transducers and activators of transcription

* Corresponding author. Fax: +1 860 679 8345.
E-mail address: machida@neuron.uchc.edu (K. Machida).

0014-5793/$36.00
http://dx.doi.org/10.1016/j.febslet.2012.04.019

Table 1
Characteristics of modular binding domains. Representative modular protein domains grouped by functional categories are presented. Approximate amino acid sizes (a.a) and affinity ranges are based on our literature search. The number of human domains and domain-containing proteins were estimated using the SMART database.

Recognition	Name	Size (a.a)	Domain	Protein	Affinity
Phosphotyrosine	PTB	100–150	36	31	nM–μM
	SH2	~100	120	110	nM–μM
	PTP[a]	250–280	50	38	nM–μM
Phosphoserine & phosphothreonine	14-3-3	~250	8	8	nM–μM
	BRCT	90–100	46	24	nM–μM
	FF	50–60	25	6	nM–μM
	FHA	65–100	31	31	nM–μM
	MH2/DWB	~200	8	8	nM–μM
	POLO-Box	~200	5	5	nM–μM
Polyproline	EVH1/WH1	~115	8	8	μM
	GYF	~60	3	3	μM
	SH3	~60	291	217	nM–μM
	WW[b]	38–40	88	48	μM
Methyllysine	LRR	22–28	1967	228	μM
	PHD	~50	168	96	μM
	Chromo	30–70	43	31	μM
	MBT	100	29	9	μM
	Tudor	~50	55	27	μM
	PWWP	~135	17	14	μM
Acetyllysine	BROMO	~110	64	46	μM
C-Terminus	PDZ	~90	264	151	nM–μM
Miscellaneous: (β-propeller family)	WD40[c]	40–60	1682	272	nM–μM

[a] Catalytic domains of protein tyrosine phosphatases (PTP) are included because of their potential use in domainomics.
[b] Pin 1 WW domain can bind phosphoserine and threonine motifs.
[c] Typically seven WD40 repeats form a β-propeller module. These modules have been reported to recognize various ligand modifications including serine-phosphorylated, threonine-phosphorylated, lysine-methylated, and ubiquitinated residues.

learned from current domainomics studies? (5) Has in silico prediction become a reliable tool? (6) What insights can domainomics provide for the protein interactome? Providing comprehensive answers to all these questions would be a challenging task for any single review. However, we believe an overview of domainomics will provide some insight into these topics. We first describe the core concept of domainomics, then outline established and emerging technologies, and review recent studies by functional category. We finish with a perspective on the unique potential of domainomics and a discussion of how to enhance its role in proteomic studies.

2. Modular protein interacting domain as affinity reagents

Independent folding of domains, which preserves binding capabilities, allows for their use as affinity and detection reagents in a manner similar to antibodies. For example, in Western blotting, protein expression is visualized by probing a membrane-bound denatured lysate with a specific antibody. Similarly, a labeled domain probe is used to detect the presence of domain binding sites in far-Western blotting [6]. As an affinity reagent, an immobilized antibody can be incubated with a lysate to enrich for the target protein and its interacting partners (immunoprecipitation). Domains can also be used to pull-down binding partners within a lysate for identification by MS [7]. However, the functional characteristics of antibodies and domains differ in many ways. Antibodies are biochemically homogeneous, therefore procedures required in proteomics, such as purification, modification for labeling, and immobilization, can be shared. In contrast, each modular domain is a part of a different full-length protein, and as such, has distinct biochemical properties such as solubility and structural stability [8]. Thus, experimental procedures must be tailored to take advantage of each domain's physiological binding activity. Further, the specificity spectrum of antibodies and domains is qualitatively different. Antibodies are meant to recognize an epitope on target molecules; so off target cross-reactivity can badly affect quantitative results. Therefore, multiple validation and normalization steps are necessary to eliminate false positive signals in antibody-based proteomics [9,10]. On the other hand, modular domains naturally have wide-ranging specificity; promiscuity in ligand selection is considered a physiological propensity rather than experimental noise. Taken together, modular domains and antibodies both can serve as useful affinity reagents in biochemical research, though procedures and research applications are often quite different.

3. Application design

As affinity and detection reagents, modular domains and their short peptide ligands (motifs) can be used in three basic proteomic designs: motif scanning (motifs are scanned), domain scanning (domains are scanned), and multiplex scanning, each differing in their concept and execution (Fig. 1). "Motif scanning" surveys possible interaction partners containing a binding motif for a modular domain of interest. Typically, a modular domain is used as an affinity probe to either a library of synthesized peptides (e.g., SPOT arrays) or a whole proteome (e.g., far-Western and pull-down experiments). Motif scanning has been frequently used to define binding consensus motifs [11–13], to determine binding sites [14,15], and to identify interacting proteins within a particular cellular environment, such as during growth factor treatment [7,16]. In addition, a labeled domain can be used as a quantitative profiling tool for determining the presence or absence of modular domain binding sites in a group of cancer cell lysates or tissues [17].

"Domain scanning" uses a peptide binding motif as bait to screen a library of domains (e.g., domain microarray) or a proteome (e.g., pull-down experiments). This approach is often used to determine binding partners when a putative domain binding motif is known to play an important role. For example, downstream effector docking sites are often examined using domain scanning [18,19].

In "multiplex scanning," binary interactions between many domains and motifs are simultaneously analyzed in the same experimental system. By determining interactions between nearly all