Cell Research (2009) :1-3.
© 2009 IBCB, SIBS, CAS All rights reserved 1001-0602/08 $ 30.00
www.nature.com/cr

LETTER TO THE EDITOR

DOG 1.0: illustrator of protein domain structures

Cell Research advance online publication 30 January 2009; doi: 10.1038/cr.2009.

Dear Editor,

A good picture is worth a thousand words. Schematic diagram of protein domain structures with functional motifs/sites in a concise and illustrative drawing is greatly helpful for a broad readership to grasp the old and novel functions of proteins rapidly. To estimate how many papers contain protein domain graphs, we went through all original research papers (excluding reviews and other articles) in five leading journals in this field, namely *Cell*, *Nature Cell Biology*, *The Journal of Cell Biology*, *Molecular Biology of the Cell* and *Molecular and Cellular Biology*, published in 2007 and found that 27.6% (551 out of 1994 papers) contain at least one figure for protein domain structures (Supplementary information, Table S1). Thus, scientific articles, particularly those in the field of molecular and cellular biology, often present schematic diagram of protein domain structures.

Usually, experimentalists could draw a protein domain graph with Microsoft PowerPoint, PhotoShop or other graphic softwares. However, the scale of a functional domain is often approximately decided by eyes. Thus, diagram of protein domain structures in exact proportions is almost unavailable. Also, the position of a specific motif or functional site (*e.g.*, a phosphorylation site) could only be roughly pin-pointed. In this regard, preparation of an exquisite figure is quite time-consuming by hand-drawing. On the other hand, the protein domain graphs could be automatically generated by a dozen of computational programs. For example, the Pfam [1], SMART [2], PROSITE [3], InterPro [4] and other related databases or online tools could predict the functional domain organization of a given protein sequence, and generate the result into a figure. However, these tools could not help experimentalists to design their own graphs. Thus, development of computer software that can illustrate user-designated protein domain structures will be a great help for biological experimentalists to communicate their research results.

In this work, we present a novel software of DOG (Domain Graph, version 1.0) for experimentalists, to prepare publication-quality figures of protein domain structures. The scale of a protein domain and the position of a functional motif/site will be precisely defined. The DOG 1.0 software was written in JAVA 1.5 (J2SE 5.0) and packed with Install4j 4.0.8. Thus, DOG 1.0 could be easily installed on a computer. Then we developed several packages to support three major Operating Systems (OS), including Windows, Unix/Linux and Mac. The Windows XP, Fedora Core 6 OS (Linux), Apple Mac OS X 10.4 (Tiger) and 10.5 (Leopard) were chosen to test the stability of DOG1.0. For Windows and Linux systems, a Java Runtime Environment 6 (JRE) package of Sun Microsystems was also included.

DOG 1.0 is easy to use. Users simply input the length of a protein sequence and then add functional domains by specifying the start and end positions for each domain. Different functional domains can be marked in different colors, with domain names displayed above, under or inside of the functional domains. Domain names, start and end positions can also be hidden to make a figure more concise. Clicking the “Site” button allows a user to add short motifs (usually < 15 aa) or functional sites, while the “Note” button adds the name of the protein or some other comments. The completed protein domain graphs can be exported in PNG (Portable Network Graphics) or JPEG (Joint Photographic Experts Group) image formats for publication. Moreover, all information of a protein domain graph can be saved as a project file in XML (The Extensible Markup Language) format, enabling users to edit their project files later or share them with colleagues.

As applications of DOG 1.0, we randomly re-drew domain graphs for 35 distinct proteins selected from the above five journals. These instances could be browsed in online or software demos. We also blindly picked out four examples shown as below (Figure 1). Smogorzewska A *et al*. identified the phosphoprotein (phosphory-

Correspondence :Xuebiao Yao[a], Yu Xue[b]
[a]Tel: +86-551-3606304; Fax: +86-551-3607141
E-mail: yaoxb@ustc.edu.cn
[b]Tel: +86-551-3607821; Fax: +86-551-3607141
E-mail: xueyu@ustc.edu.cn

lation sites of S556, S730, T952 and S1121) FANCI as a paralog of FANCD2 [5]. Its monoubiquitination (K523) and interaction with FANCD2 play important roles in response to DNA damage. A single mutation at R1285 will disrupt the Fanconi anemia (FA) pathway. Here, we re-drew the domain/motif/site structures of FANCI with DOG 1.0 (Figure 1A). FANCI contains a Lipocalin fold domain (612-650), a nuclear localization signal (NLS, 779-795, hidden in Figure 1A) and an ARM repeat (985-1207). The ubiquitination site, multiple phosphorylation sites and a disease-causing mutation are also shown (Figure 1A). TBC1D11/GAPCenA was identified as a GTPase activating protein (GAP) for Rab4 [6]. Here we re-diagrammed its functional domain structures in detail (Figure 1B). Martindill *et al.* discovered a novel Hand1 interacting protein of HICp40 [7]. HICp40 negatively regulates Hand1 activity as a molecular switch in the stem-cell differentiation pathway [7]. The domain organizations of HICp40 were re-drawn (Figure 1C). Ou *et al.* identified a novel ciliary protein of F35D11.11B in *Caenorhabditis elegans* [8]. We re-illustrated the functional domain structures of F35D11.11B (Figure 1D). Taken together, we propose that DOG 1.0 could be a great help for molecular and cellular experimentalists, allowing the presentation of protein domain structures in a more precise, convenient and concise manner.

In this article, we properly resolved a basic but important issue, allowing experimentalists to illustrate their own protein domain figures with ease. A four-step procedure makes DOG 1.0 an easy-to-use software. The proportion of a functional domain and the position of a motif/site are precisely decided from given inputs. And the output could be exported as a publication-quality figure in either PNG or JPEG format file. Although some information is not necessary to be shown in a figure, all information of a protein domain structure could be saved as a project file in XML format (*e.g.*, Supplementary information, Figure S1). We suggest that users could submit the project files together with their manuscripts as a supplementary material. It will be useful for a broader readership to understand the functional organizations of proteins and carry out further experiments. Although DOG 1.0 is mainly designed for experimentalists, it's also useful for database/program designers. For example, if the protein sequence data and other associated information are also included in a project file but not to be shown, it will be a novel file format for biological data visualization and storage. For future plan, the DOG soft-

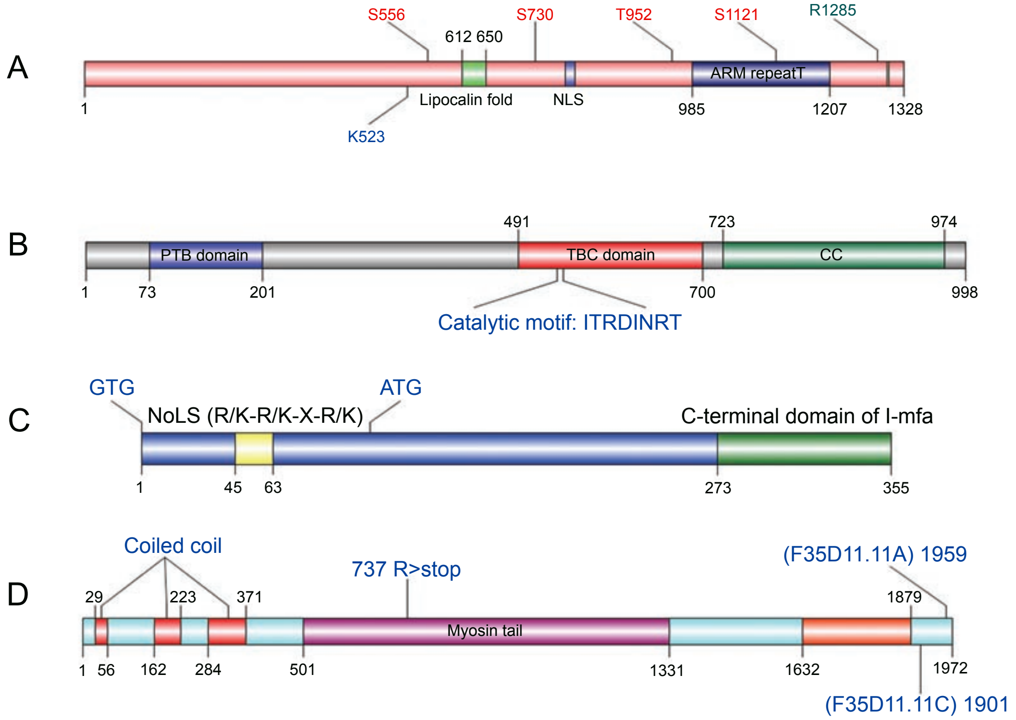

Figure 1 Domain structures for four proteins. **(A)** A monoubiquitinated and phosphorylated protein of FANCI [5]. **(B)** A GTPase activating protein (GAP) of TBC1D11/GAPCenA [6]. **(C)** A novel Hand1 interacting protein of HICp40 [7]. **(D)** A novel ciliary protein of F35D11.11B in *Caenorhabditis elegans* [8].

ware will be routinely updated based on suggestions and comments from both experimentalists and database/program designers. The DOG 1.0 software is freely available from: http://bioinformatics.lcd-ustc.org/dog.

Acknowledgment

The authors thank two anonymous reviewers, whose suggestions have greatly improved the presentation of this manuscript. We also thank Dr. Fernan Aguero (San Martin, Argentina) for his important feedback during DOG 1.0 development. This work was supported by grants from the National Basic Research Program (973 project) (2006CB933300, 2007CB914503, 2007CB947401), National Natural Science Foundation of China (30700138, 30830036, 30721002), Chinese Academy of Sciences (KSCX1-YW-R65, KSCX2-YW-21, KSCX2-YW-R-139, KJCX2-YW-M02), the Cultivation Fund of the Ministry of Education of China (NO706035), and National Science Foundation for Post-doctoral Scientists (20080430100).

Jian Ren[1], Longping Wen[1], Xinjiao Gao[1], Changjiang Jin[1], Yu Xue[1], Xuebiao Yao[1]

[1]Hefei National Laboratory for Physical Sciences at Microscale and School of Life Sciences, University of Science & Technology of China, Hefei, Anhui 230027, China

References

1 Finn RD, Tate J, Mistry J, *et al*. The Pfam protein families database. *Nucleic Acids Res* 2008; **36**:D281-D288.

2 Letunic I, Copley RR, Schmidt S, *et al.* SMART 4.0: towards genomic data integration. *Nucleic Acids Res* 2004; **32**:D142-D144.

3 Hulo N, Bairoch A, Bulliard V, *et al.* The PROSITE database. *Nucleic Acids Res* 2006; **34**:D227-230.

4 Mulder N, Apweiler R. InterPro and InterProScan: Tools for Protein Sequence Classification and Comparison. *Methods Mol Biol* 2007; **396**:59-70.

5 Smogorzewska A, Matsuoka S, Vinciguerra P, *et al.* Identification of the FANCI protein, a monoubiquitinated FANCD2 paralog required for DNA repair. *Cell* 2007; **129**:289-301.

6 Fuchs E, Haas AK, Spooner RA, *et al.* Specific Rab GTPase-activating proteins define the Shiga toxin and epidermal growth factor uptake pathways. *J Cell Biol* 2007; **177**:1133-1143.

7 Martindill DM, Risebro CA, Smart N, *et al.* Nucleolar release of Hand1 acts as a molecular switch to determine cell fate. *Nat Cell Biol* 2007; **9**:1131-1141.

8 Ou G, Koga M, Blacque OE, *et al.* Sensory ciliogenesis in Caenorhabditis elegans: assignment of IFT components into distinct modules based on transport and phenotypic profiles. *Mol Biol Cell* 2007; **18**:1554-1569.

(**Supplementary Information** is linked to the online version of the paper on the Cell Research website.)