Bioinformatics, 31(22), 2015, 3700–3702
doi: 10.1093/bioinformatics/btv447
Advance Access Publication Date: 29 July 2015
Applications Note

OXFORD

Structural bioinformatics

SynLinker: an integrated system for designing linkers and synthetic fusion proteins

Chengcheng Liu[1], Ju Xin Chin [1] and Dong-Yup Lee[1,2,3,*]

[1]Bioprocessing Technology Institute, Agency for Science Technology and Research (A*STAR), Singapore 138668, Singapore, [2]Department of Chemical and Biomolecular Engineering, National University of Singapore, Singapore 117585, Singapore and [3]NUS Synthetic Biology for Clinical and Technological Innovation (SynCTI), Life Sciences Institute, National University of Singapore, 28 Medical Drive, Singapore 117456, Singapore

*To whom correspondence should be addressed.
Associate Editor: Anna Tramontano

Received on April 24, 2015; revised on July 20, 2015; accepted on July 24, 2015

Abstract

Summary: Synthetic fusion proteins have shown great potential in various biotechnological and (bio)pharmaceutical applications. They usually contain more than two protein domains joined by a linker peptide sequence which is often selected intuitively or in ad hoc manner. Thus, we developed an integrated web-based system, SynLinker, to provide appropriate linker candidates for constructing fusion proteins. We compiled a total of 2260 linker sequences comprising of natural linkers extracted from a set of non-redundant multi-domain proteins in Protein Data Bank and artificial/empirical linkers collected from literature and patents. Multiple query interface allows users to search for the desired linker candidates based on selection criteria and their preferences. In addition, a selected linker can be combined with two domain structures which are uploaded and appended at its N and C terminals, thereby predicting a *de novo* structure of the fusion protein. Hence, SynLinker can serve as a systematic tool for researchers who are interested in designing synthetic fusion proteins.
Availability and implementation: SynLinker is freely available at http://bioinfo.bti.a-star.edu.sg/synlinker.
Contact: cheld@nus.edu.sg
Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

A synthetic fusion protein is constructed to improve properties or new functionality by synergistically connecting multiple proteins into one single complex via a peptide linker sequence (Yu *et al.*, 2015). However, it is not straightforward to identify a suitable linker since its length or amino acid composition may affect the fusion proteins to be misfolded (Zhao *et al.*, 2008), expressed at low levels (Amet *et al.*, 2009) and reduce bioactivity (Bai *et al.*, 2005; Chen *et al.*, 2012; Zhang *et al.*, 2009). The linker selection is done mostly *ad hoc* and still based on random test in experiments. Thus, two interesting bioinformatics tools have been developed for helping rational design of fusion proteins and selection of their linkers. The first one is LINKER program, which was an online server for automatically generating linker sequences for fusion proteins (Crasto and Feng, 2000; Xue *et al.*, 2004). Unfortunately, this server website is not accessible any more. The second tool is LinkerDB, a database of linkers extracted from natural multi-domain proteins (George and Heringa, 2002). This database provides linker candidates via user-defined query interface with respect to their Protein Data Bank (PDB) code, sequence, length and secondary structure. However, no further update or improvement has been made ever since it was created although PDB database has expanded tremendously during the last decade. Hence, in order to provide most updated linker candidates, we built SynLinker for synthetic fusion protein design. For the comparison of linker design tools including SynLinker, refer to the supplementary information.

2 Synlinker features

SynLinker contains a total of 2150 natural linkers obtained from a non-redundant dataset which was collected by applying a blast *P*-value cutoff 10^{-7} on a clustering of chains from NCBI MMDB. The database also comprises of 110 artificial/empirical linkers collected from public domains including literature and patent information (See the Supplementary Information for dataset description). SynLinker encompasses a range of features for users to search, BLAST and select their desired linker candidates for constructing synthetic fusion proteins.

2.1 Linker selection criteria

On the 'Search' tab, SynLinker incorporates multiple query criteria that are critical for linker selection. (i) *Linker length*, in terms of number of residues or end-to-end distance in Angstroms, is important when the distance between the joined domains influences the fusion protein's bioactivity. (ii) *Solvent accessibility* of the linker is considered as another criterion since flexible linker regions in a protein are more accessible to solvent. (iii) *Terminal amino acids* of linker sequence are pivotal for the connection of joining genes. (iv) *Compositional bias* is a determining factor for linker selection since it can largely affect the folding stability of a fusion protein. (v) *Proteolytic site* also requires attention when selecting cleavable linkers for constructing fusion proteins. Once a query is submitted, both the natural and artificial/empirical linkers in SynLinker are searched simultaneously, giving rise to a list of linker candidates satisfying the desired selection criteria (Fig. 1A). The structure of a linker candidate can be visualized by clicking the 'show' button, which brings up a JSmol viewer displaying its conformation.

2.2 Fusion protein construction

The prediction of fusion protein conformation is a unique feature of SynLinker. At the outset, users select multiple linkers from the searched candidates by clicking the checkboxes based on their preferences. These selected linkers can be viewed for comparison and also used for constructing fusion proteins on the 'Protein Fusion' tab. Subsequently, two domain structures in PDB format can be uploaded and appended at N and C terminals of suggested linker sequences. In turn, the website server calls up a backend program '*fusedomains*', which *in silico* fuses the two domains together with the linker selected among candidates, followed by the energy minimization to further reduce the steric hindrance within the structure. Finally, a possible conformation of a fusion protein is generated along with calculated hydropathicity profile indicating the regions likely to be exposed on the protein's surface (Fig. 1B). The algorithm of '*fusedomains*' program is described in the supplementary information. In future, more options will be provided to enhance the quality of protein structure by incorporating conformational sampling and domain interactions, e.g. rigid-body docking on interacting domain pairs as implemented by py-DockTET (Cheng *et al.*, 2008).

3 Implementation

SynLinker has a web interface written in a LAMP solution stack (Linux, Apache, MySQL, PHP), and designed for standard client/server architecture. The MySQL database stores linker information using the InnoDB engine. The website also uses JSmol for molecular structure visualization. SynLinker stores the 3D structures of linkers having coordinates in the PDB files. For those linkers that do not have resolved structures, SynLinker also provides a possible extended conformation generated by TraDES (http://trades.blueprint.org/)—an *ab initio* structure prediction method (Feldman and Hogue, 2000).

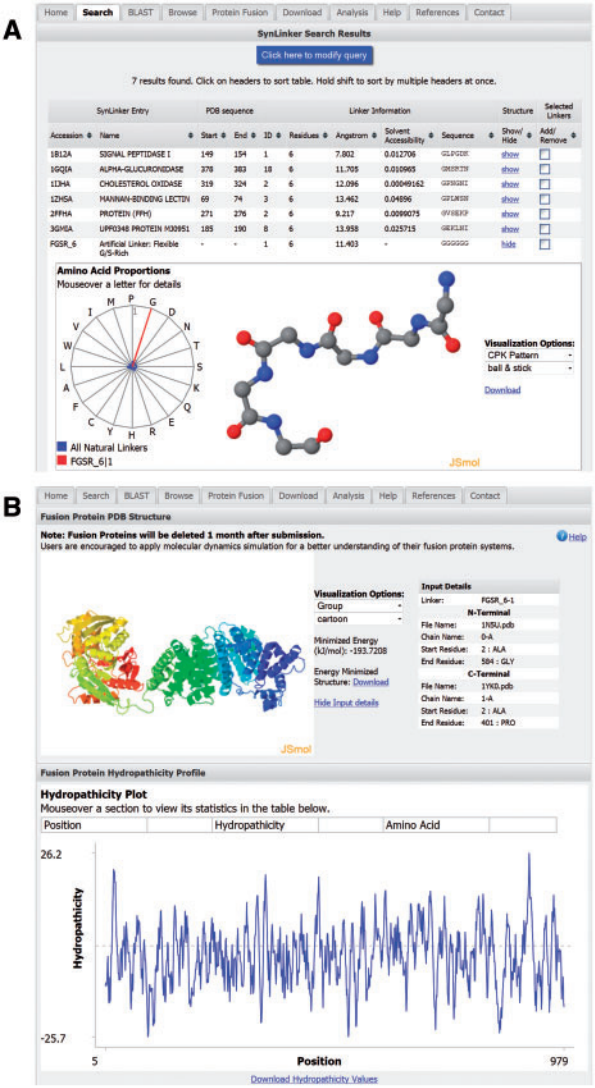

Fig. 1. SynLinker interface. (**A**) A list of linker candidates identified by setting the linker length to be six residues and starting residue to be Gly. The amino acid composition radar chart and the conformation of the selected linker FGSR_6-1 are displayed. (**B**) Fusion protein structure and hydropathicity plot. The conformation of fusion protein 'human serum albumin – atrial natriuretic factor (HAS-ANF)' (de Bold *et al.*, 2012) is predicted by fusing [PDB:1N5U] and [PDB:1YK0] together with the linker, FGSR_6-1

Funding

This work was supported by the National University of Singapore, the National Research Foundation of Singapore (NRF2013-THE001-035), Biomedical Research Council of A*STAR (Agency for Science, Technology and Research), Singapore, and a grant from the Next-Generation BioGreen 21 Program (SSAC, No. PJ01109405), Rural Development Administration, Republic of Korea.

Conflict of Interest: none declared.

References

Amet,N. *et al.* (2009) Insertion of the designed helical linker led to increased expression of tf-based fusion proteins. *Pharm. Res.*, **26**, 523–528.

Bai,Y. *et al.* (2005) Recombinant granulocyte colony-stimulating factor-transferrin fusion protein as an oral myelopoietic agent. *Proc. Natl. Acad. Sci. USA*, **102**, 7292–7296.

Chen,X. *et al*. (2012) Fusion protein linkers: property, design and functionality. *Adv. Drug Deliv. Rev.*, **65**, 1357–1369.

Cheng,T.M. *et al*. (2008) Structural assembly of two-domain proteins by rigid-body docking. *BMC Bioinformatics*, **9**, 441.

Crasto,C.J. and Feng,J.A. (2000) LINKER: a program to generate linker sequences for fusion proteins. *Protein Eng.*, **13**, 309–312.

de Bold,M.K. *et al*. (2012) Characterization of a long-acting recombinant human serum albumin-atrial natriuretic factor (ANF) expressed in *Pichia pastoris*. *Regul. Pept.*, **175**, 7–10.

Feldman,H.J. and Hogue,C.W. (2000) A fast method to sample real protein conformational space. *Proteins*, **39**, 112–131.

George,R.A. and Heringa,J. (2002) An analysis of protein domain linkers: their classification and role in protein folding. *Protein Eng.*, **15**, 871–879.

Xue,F. *et al*. (2004) LINKER: a web server to generate peptide sequences with extended conformation. *Nucleic Acids Res.*, **32**, W562–W565.

Yu,K. *et al*. (2015) Synthetic fusion protein design and applications. *Biotechnol. Adv.*, **33**, 155–164.

Zhang,J. *et al*. (2009) Design and optimization of a linker for fusion protein construction. *Prog. Nat. Sci.*, **19**, 1197–1200.

Zhao,H.L. *et al*. (2008) Increasing the homogeneity, stability and activity of human serum albumin and interferon-alpha2b fusion protein by linker engineering. *Protein Exp. Purif.*, **61**, 73–77.