W562–W565 *Nucleic Acids Research, 2004, Vol. 32, Web Server issue*
DOI: 10.1093/nar/gkh422

LINKER: a web server to generate peptide sequences with extended conformation

Fan Xue[1], Zhong Gu[2] and Jin-an Feng[1,2,*]

[1]Department of Chemistry and [2]Center for Biotechnology, Temple University, 1901 N 13th Street, Philadelphia, PA 19122, USA

Received February 14, 2004; Revised and Accepted April 8, 2004

ABSTRACT

LINKER was developed as an online server to assist biomedical researchers to design linker sequences for constructing functional fusion proteins. The program automatically generates a set of peptide sequences that are known to adopt extended conformations as determined by X-ray crystallography and NMR. In addition to the desired linker sequence length, the web interface provides a number of optional input parameters so that the users may enhance sequence selection based on the requirements of specific applications. The output of LINKER includes a list of peptide sequences with specified length and sequence characteristics. A graphical subroutine was implemented to highlight the chemical features of every linker sequence by hydrophobicity plots. LINKER can be accessed at http://astro.temple.edu/~feng/Servers/BioinformaticServers.htm.

INTRODUCTION

The gene fusion technique has become an increasingly useful tool in biomedical research. Fusion proteins have been constructed to increase cellular stability and biological activity of functional proteins (1). In structural biology, the construction of recombinant fusion protein has often been used as a means to increase the expression of soluble proteins and to facilitate protein purification (2). In biotechnology, fusion proteins have been constructed to engineer bifunctional enzymes, and to select and produce antibodies (3–5). Rationally designed P450 enzymes were produced by gene fusion of the different domains of the cytochrome P450 BM3 from *Bacillus megaterium* (BMP) with flavodoxin and the domains of human P450 2E1 (6). These engineered fusion proteins showed improved solubility and electrochemical properties with catalytic activities of the P450 enzymes. Specialized functional fusion proteins have been constructed to target specific genes (7,8). It is expected that the gene fusion technology will continue to have significant impact on a variety of fields of biomedical research.

The construction of a fusion protein involves the linking of two macromolecules by a linker sequence. The selection of the linker sequence is of particular importance. Linker sequence composition could affect the folding stability of a fusion protein. It is often unfavorable to have a linker sequence with high propensity to adopt α-helix or β-strand structures, which could limit the flexibility of the protein and consequently its functional activity. Indeed, a more desirable linker is a sequence with preference to adopt extended conformation. In practice, most currently designed linker sequences have a high content of glycine residues that force the linker to adopt loop conformation. However, such monotonous linkers limit the functional activity of the fusion proteins. With the rapid advancing of biotechnology, it is envisioned that more sophisticated fusion proteins with diversified linker sequences will be created.

Polypeptide sequences specify their adopted secondary structures in proteins. Sequence analyses of proteins have revealed unique sequence patterns favorable to adopt α-helix and loop conformations (9,10). The loop sequences are actually quite diversified in proteins (9). It was found that a variety of loop sequences exist to accommodate different chemical environments of protein structures. For example, the surface loops in proteins are more hydrophilic while the interior loops of the proteins are relatively neutral. Loops of different sizes also exhibit different sequence patterns. Taking advantage of the knowledge learned from analysis of protein structures, we developed a program called LINKER that automatically generates a set of peptide sequences that are known to adopt extended conformations as determined by X-ray crystallography and NMR. A loop library was derived from the Protein Data Bank (PDB) that contained both globular and membrane proteins. Loop sequences of various lengths were extracted as previously described (11). Loop lengths of less than four residues, as well as the redundant loop sequences, were also removed from the library. LINKER is a web-based server that allows the user to select sequences of their choice.

*To whom correspondence should be addressed at Department of Chemistry, Temple University, 1901 N 13th Street, Philadelphia, PA 19122, USA. Tel: +1 215 204 7128; Fax: +1 215 204 1532; E-mail: feng@astro.temple.edu

LINKER SERVER

The goal to the design of LINKER was to provide a simple user interface for easy interaction. The workflow of the program can be described as a series of sequence filters (Figure 1) (11). The first filter, an input of the desired linker sequence length, directs the program to search the loop library and fetch linker sequences that match the input requirement. The second filter removes linker sequences that may contain sequence patterns sensitive to specified proteases. In constructing the fusion gene, it is often necessary to process the ends of the fusing genes and the linker gene sequence with restriction enzymes before ligating them together genetically. Linker sequences containing nucleotide sites sensitive to the same restriction enzyme as that of ends processing enzymes would be unfavorable choices. The third filter of LINKER allows the user to incorporate this consideration in their experimental design. LINKER also gives users the ability to select the chemical feature of the linker sequences by specifying desired amino acid compositions. The required

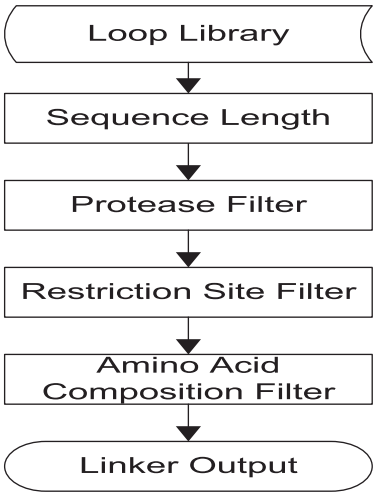
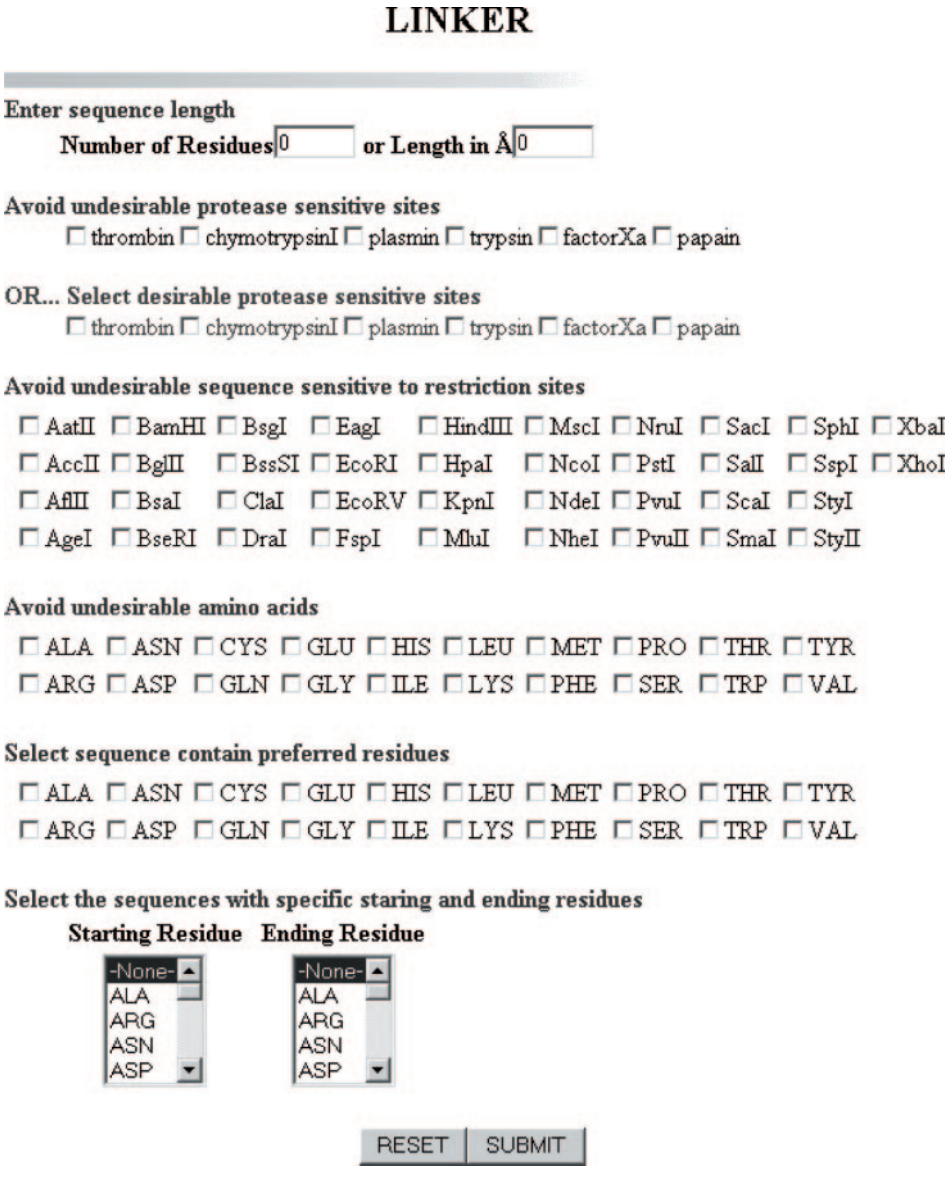


Figure 1. A flowchart of LINKER illustrating the sequence filtering process of the program.

(a)



(b)

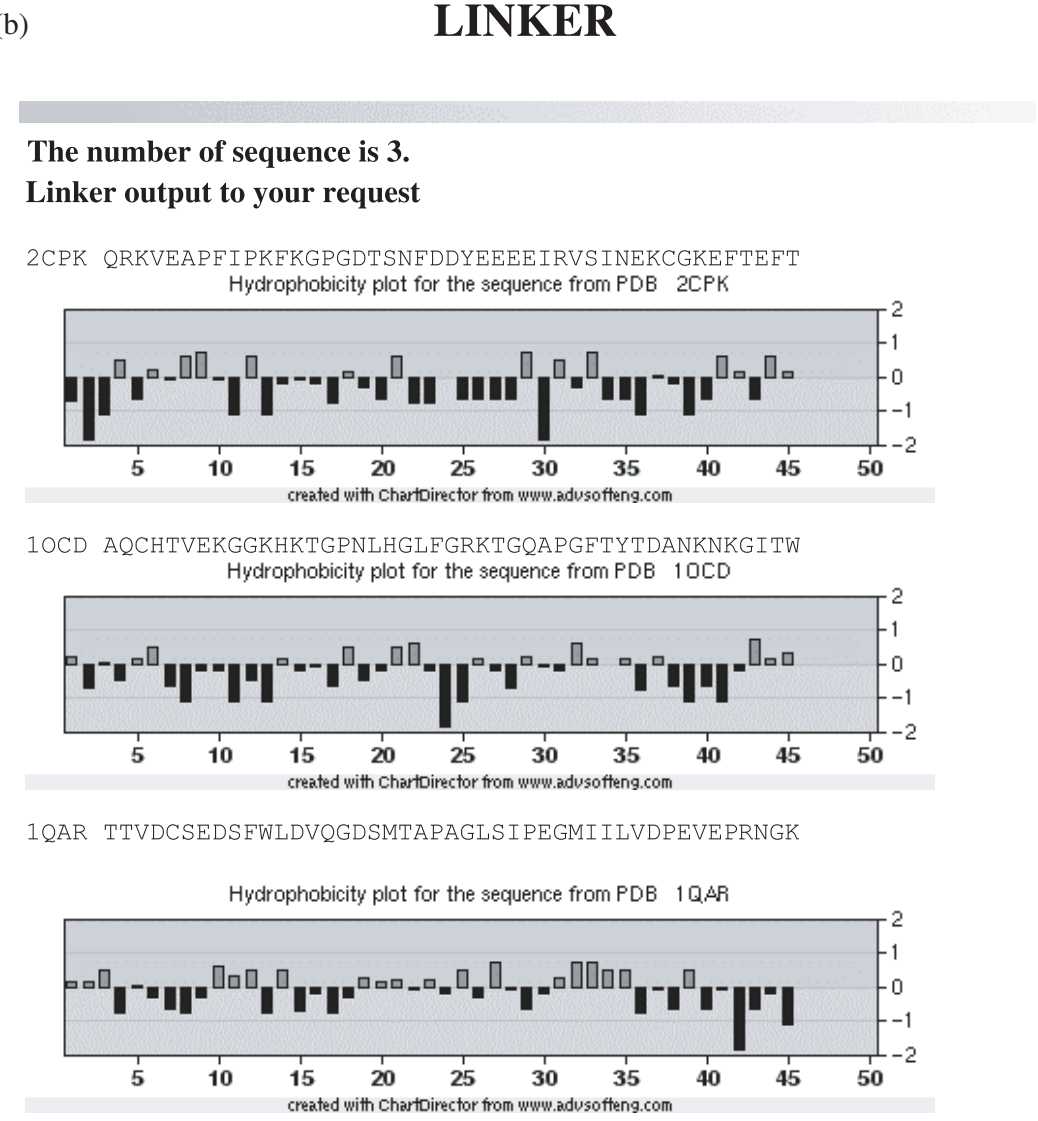


Figure 2. An intuitive user interface of LINKER. (**a**) The input page of the LINKER with a layout of various input options; (**b**) an example of LINKER output with linker sequences of 45 residues.

restriction enzyme ends processing also limits the selection of amino acids for both ends of linker sequences, since restriction sites only encode corresponding amino acids. Users may specify this limitation in the input page of LINKER.

Figure 2(a) shows the input page of LINKER. The user may specify the length of desired linker sequence with either the number of residues or the estimated length in angstroms. Proteolytic sites for six most commonly used proteases were incorporated in the LINKER. They include trypsin, chymotrypsin, thrombin, plasmin, papain and factor Xa. To activate this filter, the user may simply select the box next to the proteases. Alternatively, the user may select linker sequences that contain desirable protease sensitive sites. The input page also displays a complete list of all the restriction sites that are incorporated in the LINKER. The user may select the restriction enzymes used in the fusion gene construction. The program will automatically avoid selecting linker sequences that contain these restriction sites in their respective coding genes. In the amino acid composition filter, we implemented three selection categories in order to help users to specify their selections. For example, in constructing a membrane fusion protein, it is perhaps more favorable to include linker sequences with hydrophobic residues. On the other hand, in constructing a DNA-binding protein, it may be desirable to exclude highly charged residues (Lys and Arg) in the linker sequences since these residues could form salt bridges with the phosphate backbone of the DNA, thus influencing the DNA-binding property of the engineered fusion protein.

The output of LINKER is simple and easy to inspect [Figure 2(b)]. It contains a list of sequences that are known to adopt extended conformation. A PDB access code was placed at the beginning of each sequence identifying the source of the linker sequence. In this updated version of LINKER, we implemented a graphical representation of the chemical features of each sequence by hydrophobicity profiling. The hydrophobicity index for each amino acid was obtained from the Eisenberg consensus hydrophobicity scale (12).

AVAILABILITY

LINKER was written in Fortran with CGI interface. It was compiled on a LINUX-based workstation. The web server can be freely accessed on the World Wide Web at http://astro.temple.edu/~feng/Servers/BioinformaticServers.htm. A brief description of the program and a detailed user guide with examples are also available at the website.

ACKNOWLEDGEMENTS

We thank Junwen Wang and Liting Wen for advice on porting LINKER from SGI Irix 6.2 to Linux OS, and other members of the Feng group for helpful discussions. This research was supported by grants from the National Institutes of Health GM54630 and the American Cancer Society PRG9926301GMC, and an appropriation from the Commonwealth of Pennsylvania.

REFERENCES

1. Liu,N., Caderas,G., Deillon,C., Hoffmann,S., Klauser,S., Cui,T. and Gutte,B. (2001) Fusion proteins from artificial and natural structural modules. *Curr. Protein Pept. Sci.*, **2**, 107–121.
2. Caffrey,M. (2003) Membrane protein crystallization. *J. Struct. Biol.*, **142**, 108–132.
3. Murphy,J.R. (1996) Protein engineering and design for drug delivery. *Curr. Opin. Struct. Biol.*, **6**, 541–545.
4. Marshall,S.A., Lazar,G.A., Chirino,A.J. and Desjarlais,J.R. (2003) Rational design and engineering of therapeutic proteins. *Drug Discov. Today*, **8**, 212–221.
5. Graddis,T.J., Remmele,R.L., Jr and McGrew,J.T. (2002) Design proteins that work using recombinant technologies. *Curr. Pharm. Biotechnol.*, **3**, 285–297.
6. Gilardi,G., Meharenna,Y.T., Tsotsou,G.E., Sadeghi,S.J., Fairhead,M. and Giannini,S. (2002) Molecular Lego: design of molecular assemblies of P450 enzymes for nanobiotechnology. *Biosens. Bioelectron.*, **17**, 133–145.
7. Kim,J.S. and Pabo,C.O. (1997) Transcriptional repression by zinc finger peptides. Exploring the potential for applications in gene therapy. *J. Biol. Chem.*, **272**, 29795–29800.
8. Tang,L., Li,J., Katz,D.S. and Feng,J.A. (2000) Determining the DNA bending angle induced by non-specific high mobility group-1 (HMG-1) proteins: a novel method. *Biochemistry*, **39**, 3052–3060.
9. Crasto,C.J. and Feng,J.A. (2001) Sequence codes for extended conformation: a neighbor-dependent sequence analysis of proteins. *Proteins*, **42**, 399–413.
10. Wang,J. and Feng,J.A. (2003) Exploring the sequence patterns in the α-helices of proteins. *Protein Eng.*, **16**, 799–807.
11. Crasto,C.J. and Feng,J.A. (2000) LINKER: a program to generate linker sequences for fusion proteins. *Protein Eng.*, **13**, 309–312.
12. Eisenberg,D. (1984) Three-dimensional structure of membrane and surface proteins. *Annu. Rev. Biochem.*, **53**, 595–623.