

A Facile Method for High-throughput Co-expression of Protein Pairs*[§]

Andrei Alexandrov‡§, Marissa Vignali¶||, Douglas J. LaCount¶||, Erin Quartley‡, Christina de Vries‡, Daniela De Rosa‡, Julie Babulski‡, Sarah F. Mitchell‡, Lori W. Schoenfeld¶||, Stanley Fields¶||, Wim G. Hol***, Mark E. Dumont‡§, Eric M. Phizicky‡§, and Elizabeth J. Grayhack‡§††

We developed a method to co-express protein pairs from collections of otherwise identical *Escherichia coli* plasmids expressing different ORFs by incorporating a 61-nucleotide sequence (LINK) into the plasmid to allow generation of tandem plasmids. Tandem plasmids are formed in a ligation-independent manner, propagate efficiently, and produce protein pairs in high quantities. This greatly facilitates co-expression for structural genomics projects that produce thousands of clones bearing identical origins and antibiotic markers. *Molecular & Cellular Proteomics* 3:934–938, 2004.

Co-expression of proteins is an important objective for biochemical and structural analysis of protein complexes because it often increases authenticity of biological activity and increases solubility of protein partners (1, 2). Although a variety of versatile systems are available for construction of large collections of expression plasmids, no convenient system exists for co-expression of protein pairs from such collections. Because the presence of an identical replication origin and antibiotic resistance marker precludes stable propagation of plasmid pairs in the same cell, one or both of the ORFs must be moved to a new plasmid to achieve co-expression. To this end, a number of methods have been developed, including the use of two plasmids with different selectable markers and compatible origins of replication (3, 4), a single plasmid containing two ORFs under control of separate promoters (5), or a single plasmid containing ORFs arranged in a polycistronic message (6). All of these methods are inconvenient for use in a high-throughput mode because they require *ad hoc* construction of the co-expression plasmid and often

sequencing of the ORFs in the new constructs.

We have developed a facile and general method for protein co-expression in *Escherichia coli* that can utilize sets of ORFs in identical expression plasmids with the simple requirement that the starting plasmid contains a 61-nucleotide sequence called LINK. This method takes advantage of our demonstration that two otherwise identical plasmids bearing different ORFs can be joined “head to tail” in a single tandem plasmid and propagated in *E. coli* (7). The LINK sequence expedites the joining of two plasmids using methods from ligation-independent cloning (LIC)¹ (8) and generalizes the method for any pair of ORFs. We demonstrate that the resulting tandem plasmid, with two identical replication origins and antibiotic resistance markers, efficiently propagates in *E. coli*, and that the two proteins are readily co-expressed in quantities that would satisfy the most demanding structural biology applications. The method is simple and rapid and does not require sequencing of the ORFs in the resulting tandem plasmid.

EXPERIMENTAL PROCEDURES

Materials—T4 DNA polymerase (LIC-qualified), pET14b vector, Nova Blue cells, and BL21(DE3) pLysS-competent cells were obtained from Novagen (Madison, WI); TALON resin was obtained from BD Biosciences (San Diego, CA); Criterion precast Gel 8–16% and broad range protein markers were obtained from Bio-Rad (Hercules, CA); 1Kb Plus DNA ladder was obtained from Invitrogen (San Diego, CA); BL21-Codon Plus (DE3)-RIL-competent cells were obtained from Stratagene (La Jolla, CA).

Plasmids and Plasmid Construction—BG1861 is a LIC vector that expresses proteins as N-terminal His₆-ORF fusion proteins. It was constructed from pET14b by replacing pET14b sequences between the NcoI site and the BamHI site with the sequence 5′-CCATGGCT-CACCACCACCACCACCATATGACGCGTTAACCACGTGAGTAAGATAGGATCC-3′ containing NcoI, NdeI, MluI, BbrPI, and BamHI sites (underlined). Replacement was accomplished in two steps by inserting complimentary oligonucleotides of sequence 5′-CATGGCTCAC-CACCACCACCACCATATGACGCGTC-3′ and 5′-TCGAGACGCGT-CATATGGTGGTGGTGGTGGTGAGC-3′ into the NcoI-XhoI sites in pET14b to create pET14b-LIC1, followed by insertion of oligonucleotides of sequence 5′-CGCGTTAACCACGTGAGTAAGATAG-3′ and

From the Structural Genomics of Pathogenic Protozoa (SGPP) Consortium and ‡Center for Human Genetics and Molecular Pediatric Disease, University of Rochester Medical Center, Rochester, NY 14642; §Department of Biochemistry and Biophysics, University of Rochester Medical Center, Rochester, NY 14642; ¶Departments of Genome Sciences and Medicine, University of Washington, Seattle, WA 98195; ||Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195; and ***Department of Biochemistry, University of Washington, Seattle, WA 98195

Received, June 14, 2004, and in revised form, July 1, 2004

Published, MCP Papers in Press, July 6, 2004, DOI 10.1074/mcp.T400008-MCP200

¹ The abbreviations used are: LIC, ligation-independent cloning; IMAC, immobilized metal-affinity chromatography; MBP, maltose-binding protein; LB, Luria broth; IPTG, isopropyl-β-D-1-thiogalactopyranoside; BME, 2-mercaptoethanol; E2, ubiquitin-conjugating enzyme; UEV, ubiquitin-conjugating enzyme variant.

5'-GATCCTATCTTACTCACGTGGTTAA-3' into the MluI-BamHI sites of pET14b-LIC1.

The LINK sequence was originally cloned into a vector AVA0220, a derivative of the pET14b vector, by insertion of the LINK sequence into EagI-SalI sites using oligos FLIP_Top (5'-GGCCGTAACAACAC-CATTTAAATGGAGTGGTTACAAATGGAGTGGTTAATTAACAACAC-CATTTG-3') and FLIP_Bottom (5'-TCGACAAATGGTGTGTTAATTA-ACCACTCCATTTGTAACCACTCCATTTAAATGGTGTGTTAC-3') resulting in AVA0229 vector. The LINK sequence was moved to additional vectors (see below) by PCR amplification, digestion with appropriate restriction enzymes, and standard cloning procedures.

AVA0469 vector was constructed from BG1861 by insertion of a fragment that contained the LINK sequence into the NcoI-EagI sites. In this construction, the LINK sequence is upstream of the T7 promoter and does not interfere with expression or LIC of ORFs. The sequence of this vector is reported in the supplemental materials.

AVA0306 vector is a LIC vector containing a LINK sequence that expresses proteins as N-terminal His₆-MBP-3C^{protease site}-ORF fusion proteins. It was derived from H-MBP-3C vector (9) in three steps: insertion of sequences necessary for LIC of ORFs, deletion of an endogenous Swal site in the plasmid, and insertion of the LINK sequence into a M13 origin of replication, concomitantly destroying the M13 origin. First, to convert H-MBP-3C vector into a LIC vector, the multiple cloning site of the vector was replaced with the LIC-site-containing DNA fragment (5'-GAATTCCTGGAAGTTCTGTTCCAGG-GTCCTGGTTCGCGAATATTCTAGCTTTGTTTAAACAGCACGAACA-AGTTCTGCAG-3'), which was inserted into the EcoRI-PstI sites of the H-MBP-3C vector (9) to add NruI and Pml sites as well as sequences for LIC. Second, an existing Swal site present in the H-MBP-3C vector was deleted as follows: primers complimentary to the sequences flanking the Swal site (but not including it) were used to copy the plasmid by PCR, and the plasmid was rejoined via an added BamHI site encoded in the oligonucleotides. The primer pair was Swa_to_Bam_F (5'-GCGGGATCCGTAACGTTAATATTTGTTAA-ATTCGC-3') and Swa_to_Bam_R (5'-GCGGGATCCCAATCTTCTG-TTTTTGGGGC-3'). Third, the DNA fragment containing the LINK site was amplified from the vector AVA0229 using the primers BamHI_FLIP_F (5'-GCGGGATCCGCAACGCGGGCATCCC-3') and pMal_FLIP_R (5'-GAGGCCGTTGAGCACCGCACTACGTGATTCCTT-CTG-3') and inserted into the BamHI-DraIII sites of the vector resulting from step two. The sequence of vector AVA0306 can be found in the supplemental materials.

Saccharomyces cerevisiae ORFs TRM8 (YDL201w) and TRM82 (YDR165w) were PCR-amplified from yeast genomic DNA using primer pairs 201_F_SG (5'-GGGTCTCGTTCGATGAAAGCCAAGC-CACTAAGCC-3'), 201_R_SG (5'-CTTGTTCTGCTGTTTATTACAAT-ATGGCTGGCGTTGGTAATC-3'), and 165_F_SG (5'-GGGTCTCGTTCGATGAGCGTCATTCATCCTTTGCAG-3'), 165_R_SG (5'-CTTGTT-CGTGCTGTTTACGCCGCTTCAGCTAGAAACAGAG-3') and inserted into NruI- and PmlI-digested vector AVA0306 using standard LIC procedures (8).

Fragments of *Plasmodium falciparum* ORFs chr3.gen_223 and chr5.gen_178 were PCR-amplified from interacting clones obtained in a yeast two-hybrid screen, using primers Lic2F (5'-CTCACCACCAC-CACCACCATATGACATACTCATTGTTTCAGCCG-3') and Lic2R (5'-ATCCTATCTTACTCACTTATCCACCTCGAGAACGCGTTTGTCG-3'), and inserted into NdeI- and PmlI-digested vector AVA0469 using LIC (8). Likewise, *Leishmania major* ORFs, indicated in Fig. 2, were PCR-amplified from genomic DNA and cloned into AVA0469 using LIC. Sequences of these ORFs can be found at depts.washington.edu/sgpp/.

Tandem Plasmid Construction—Tandem plasmids were constructed in three steps. First, 1 μg of each plasmid was digested at the LINK site with 20 units of a restriction enzyme (Swal or PacI) in

appropriate buffers for 1 h in a 20-μl reaction, followed by heat-inactivation of the restriction enzyme (65 °C for 20 min). Second, 0.15 μg of digested plasmid (0.05 pmol, 3 μl of the heat-inactivated restriction reaction) was treated with 1 unit of T4 DNA polymerase in buffer containing 50 mM TrisCl, pH 8.0, 10 mM MgCl₂, 5 μg/ml BSA, and 5 mM DTT in the presence of 2.5 mM dGTP (for Swal-cleaved plasmid) or dCTP (for PacI-cleaved plasmid) at 22 °C for 20 min in a 20-μl reaction to form single-stranded 5' overhangs, followed by heat-inactivation of the enzymes. Third, 22.5 ng of T4 DNA polymerase-treated DNA from the heat-inactivated reactions (3 μl) was mixed at room temperature, annealed at 65 °C for 1 min, and supplemented with 2 μl of 25 mM EDTA. Then 1 μl of the resulting mixture was transformed into Nova Blue cells according to the manufacturer's protocol.

Cell Growth—Plasmid-transformed BL21 (DE3) pLysS cells were grown from a single colony overnight at 37 °C in 6.2 ml of Luria broth media supplemented with 100 μg/ml ampicillin, and then for 16 more generations using serial dilutions (equivalent to 406 liters of culture) and induced with 1 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG) at 18 °C for 16 h. Cells were harvested, lysed with SDS, and proteins were analyzed by SDS-PAGE.

Co-purification of Protein Complexes—The ubiquitin-conjugating enzyme (E2)/ubiquitin-conjugating enzyme variant (UEV) protein pair from *P. falciparum* was purified using IMAC and gel filtration. Frozen cells were lysed by sonication in buffer A (20 mM HEPES, pH 7.5, 1 M NaCl, 2 mM 2-mercaptoethanol (BME), and 5% glycerol) and centrifuged to generate crude extract. Extract was applied to TALON resin in buffer B (20 mM HEPES, pH 7.5, 0.5 M NaCl, 2 mM BME, and 5% glycerol), resin was washed with buffer, and proteins were eluted using a stepwise gradient of 5, 10, and 250 mM imidazole in buffer B. The protein pair was further purified using gel filtration chromatography on HiLoad Superdex 200 26/60 column (Amersham Pharmacia Biotech, Piscataway, NJ) and dialysed against buffer C (20 mM HEPES, pH 7.5, 0.5 M NaCl, 2 mM DTT, and 50% glycerol).

S. cerevisiae Trm8/Trm82 protein complex was purified in buffer containing 20 mM HEPES, pH 7.5, 10% glycerol, and 2 mM BME, using IMAC, followed by protease 3C cleavage overnight at 4 °C to release the bound protein and gel filtration on a HiLoad Superdex 200 16/60 column, followed by dialysis.

RESULTS

To generate tandem plasmids from expression plasmids containing the LINK sequence, each ORF-expressing plasmid is cleaved at the LINK site with a restriction enzyme, and the plasmids are joined in a ligation-independent manner. The LINK sequence (Fig. 1a) features two octameric restriction sites (Swal and PacI), each flanked by sequences (labeled 1–4) that produce 14-bp 5' overhangs upon restriction digestion and treatment with T4 DNA polymerase (Fig. 1b). The LINK sequence is designed such that the single-stranded overhangs around the Swal site hybridize exactly with those around the PacI site, allowing for ligation-independent formation of the correct tandem plasmids, while preventing re-sealing of single plasmids and formation of incorrect products. Thus one plasmid must be cleaved with PacI and the other with Swal to generate the complimentary overhangs, but either plasmid can be cleaved with either enzyme. As with LIC (8), cloning with the LINK sequence is simple and results in an overwhelming majority of transformants that contain the desired tandem plasmids.

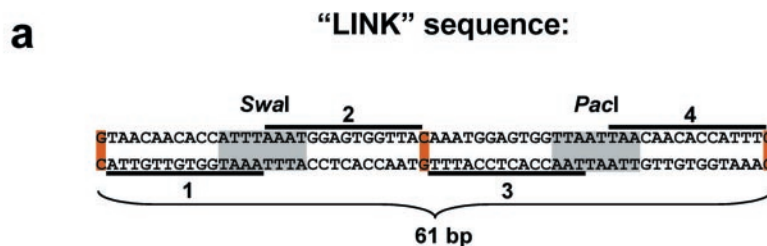


FIG. 1. Schematic of the LINK method of tandem plasmid construction. *a*, LINK sequence. Restriction sites are shaded in *gray*, and base pairs corresponding to T4 DNA polymerase exonucleolytic stop sites are shaded in *orange* (see below). *b*, construction of tandem plasmids. Plasmids are digested with *SwaI* or *PacI*, followed by heat-inactivation, treatment with T4 DNA polymerase in the presence of dGTP or dCTP, respectively, to form 5' overhangs, mixing of plasmids, and transformation into *E. coli*.

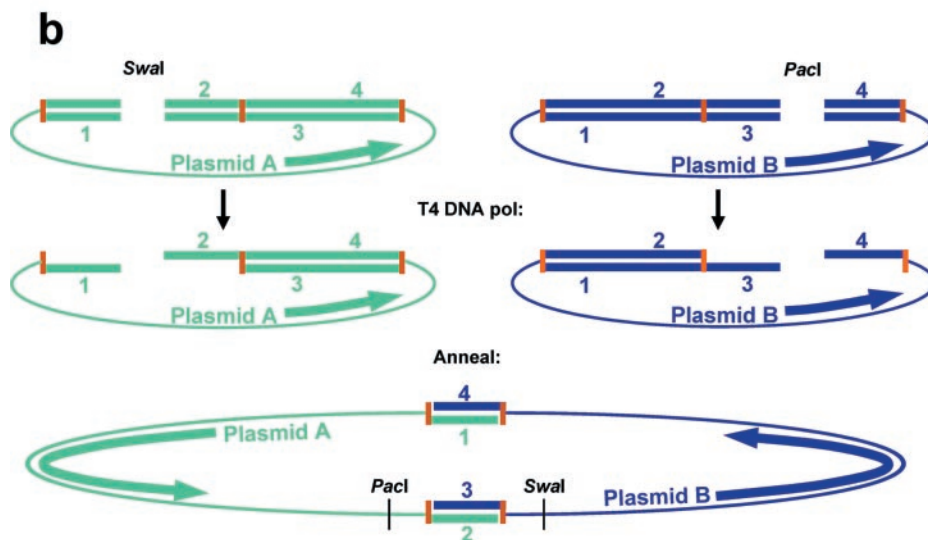


Fig. 2, *a* and *b* illustrate the analysis of six tandem plasmids constructed with the LINK sequence, containing all possible pairwise combinations of four plasmids, each expressing a distinct *L. major* ORF in a pET-derived vector. In each case, digestion of the tandem plasmids with restriction enzymes that excise the ORFs results in DNA fragments corresponding to the size of each ORF (Fig. 2*a*, compare lanes 1–4 with lanes 5–10), as well as a vector fragment of the expected size that is common to both parent and tandem plasmids. Furthermore, linearization of the plasmids with an enzyme that cleaves only once shows that the size of each tandem plasmid is distinctly larger than the size of the parent plasmid (Fig. 2*b*, lanes 5–10) and close to the size expected from the sum of the sizes of the corresponding plasmids of which it is made (Fig. 2*b*, lanes 1–4).

The tandem plasmids propagate efficiently to allow large scale co-expression of proteins. To demonstrate this, we monitored plasmid maintenance and protein expression after extended growth. Starting from a single colony of each transformant, we first grew a 6.2-ml overnight culture and then propagated each strain for 16 more generations by serial dilution of cultures, equivalent to the growth of 406 liters of culture. Then we assessed the plasmid content of the cells and induced the cells to express protein. Plasmids analyzed at this point were all the same size as before transformation into expression cells (compare Fig. 2, *b* and *c*), illustrating that the tandem plasmids propagate as a unit during this growth

period, with little observed recombination between the monomeric components. We speculate that the efficient propagation of tandem plasmids made from pBR322-derived vectors such as pET and pMal is due to the absence of a *cer* sequence, which is necessary for efficient recombinational resolution of dimers of ColE1 plasmids (10, 11). Following induction of expression after this prolonged growth, each strain containing a tandem plasmid expressed the corresponding pair of proteins at high levels, as judged by SDS-PAGE analysis of whole-cell lysates (Fig. 2*d*, compare lanes 1–4 with lanes 5–10). We note that tandem plasmids are more likely than other methods to yield comparable expression of ORF pairs, because the relative stoichiometry of the two ORFs is fixed and each ORF uses the same promoter.

Proteins expected to be members of a complex are readily co-purified after co-expression from a tandem plasmid made with LINK. Fig. 2*e* illustrates this for two different protein pairs, each in a different LINK-containing vector. Lanes 1–4 show the analysis of the predicted complex of E2 and UEV from *P. falciparum*, identified in a yeast two-hybrid screen. Each ORF was cloned into a LINK version of the pET-derived LIC vector BG1861 (AVA0469) to produce the corresponding His₆-ORF fusion protein when expressed alone (lanes 1 and 2). When combined as a tandem plasmid using the LINK sequence, both proteins are expressed (lane 3), and the E2/UEV protein pair is readily purified (lane 4). Similarly, lanes 6–9 illustrates this for the Trm8/Trm82 protein complex of *S. cerevisiae* (7),

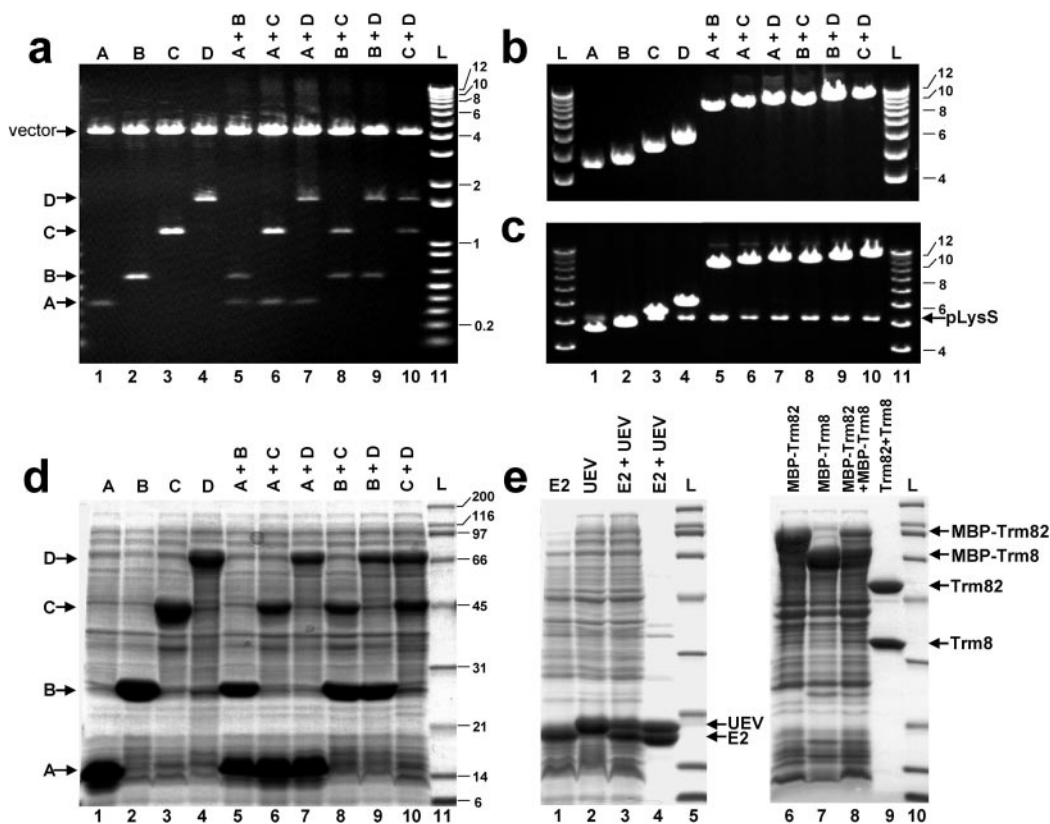


FIG. 2. Analysis of tandem plasmids and protein co-expression. *a*, LINK-generated tandem plasmids contain two inserts. Plasmids in vector AVA0469 containing *L. major* ORFs Lmaj007187AAA (*A*), Lmaj007352AAA (*B*), Lmaj007353AAA (*C*), and Lmaj007233AAA (*D*) were digested with BamHI and NdeI to release inserts (lanes 1–4), and products were compared with those of tandem plasmids (lanes 5–10) after gel electrophoresis; lane 11, DNA ladder. *b*, tandem plasmids are of the expected size. Plasmids in *a* were linearized with PacI, which cleaves each plasmid once (lanes 1–10), and resolved by gel electrophoresis; lane 11, DNA ladder. *c*, tandem plasmids propagate efficiently in expression cells. Plasmids in *a* were transformed into BL21 (DE3) pLysS cells, colonies were grown at 37 °C using serial dilutions equivalent to 406 liters of culture, and DNA was analyzed as in *b*. *d*, proteins are co-expressed at high levels from tandem plasmids. Cells were grown for multiple generations as in *c*, induced with IPTG at 18 °C for 16 h, harvested, and then lysed with SDS; proteins were analyzed by SDS-PAGE (lanes 1–10); lane 11, broad range standards (Bio-Rad). *e*, co-expression and purification of two protein complexes. Lanes 1–4, analysis of an E2/UEV protein pair using vector AVA0469 in BL21 (DE3) pLysS cells. Lanes 1 and 2, expression of E2 and UEV protein fragments from individual plasmids; lane 3, co-expression of E2 and UEV from the tandem plasmid; lane 4, purified complex after IMAC and gel filtration. Lanes 6–9, similar analysis of the yeast Trm8/Trm82 protein complex, cloned in vector AVA0306 and expressed in BL21-Codon Plus (DE3)-RIL cells. Lane 9, purified complex after IMAC, 3C protease cleavage, and gel filtration. Lanes 5 and 10, standards.

using a different LINK-containing vector (AVA0306), derived from a pMal vector, that expresses proteins as His⁶-MBP-3C^{protease site}-ORF fusion proteins.

DISCUSSION

The LINK method is ideal for structural genomics applications for two major reasons.

First, it has enormous potential as a high-throughput co-expression method for large sets of ORF-expressing plasmids containing LINK, because the method is not PCR-based and does not require any DNA purification steps and thus can be easily automated. Moreover, because the manipulations of each plasmid occur at a site remote from the ORF, the method does not require sequencing of ORFs in the tandem plasmid. We note that the LINK sequence can be introduced into any nonessential region of a plasmid for subsequent use.

Thus, a plasmid containing the LINK sequence can be used for routine expression of individual proteins, because explicit testing showed that the sequence has no obvious negative effect on expression (data not shown).

Second, the method can be used for co-expression of virtually any pair of proteins because either of the two octameric restriction sites can be used for cleavage of a given plasmid, as long as the other site is used for the second plasmid. This feature allows the generation of tandem plasmids containing more than 99% of typical protein pairs (Supplemental Table I), the exact percentage depending on GC-content, average protein length, and distribution of sites within ORFs. In yeast, for example, 99.4% of all possible protein pairs can be obtained by this method. Indeed, the only pairs that cannot be joined are those exceedingly rare combinations of ORFs with the same octameric restriction site

within their sequences or those with both octameric recognition sequences in one ORF.

Finally, we note that the LINK method could be used to generate a random library of plasmids containing nearly every possible combination of protein pairs in a given genome for functional screening or selection. Applied to yeast, this would result in a library of 1.8×10^7 possible protein pairs, well within the transformation capability of *E. coli*.

* This work was supported by National Institutes of Health Grant P50 GM64655 (to W. G. H.). W. G. H. and S. F. are investigators of the Howard Hughes Medical Institute. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[S] The on-line version of this manuscript (available at <http://www.mcponline.org>) contains supplemental material.

‡‡ To whom correspondence should be addressed: Center for Human Genetics and Molecular Pediatric Disease, University of Rochester Medical Center, Rochester, NY 14642. Tel.: 585-275-2765; E-mail: elizabeth_grayhack@urmc.rochester.edu.

REFERENCES

1. Li, C., Schwabe, J. W., Banayo, E., and Evans, R. M. (1997) Coexpression of nuclear receptor partners increases their solubility and biological activities. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 2278–2283
2. Wang, H., and Chong, S. (2003) Visualization of coupled protein folding and binding in bacteria and purification of the heterodimeric complex. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 478–483
3. Gangloff, Y. G., Werten S., Romier C., Carre L., Poch O., Moras D., and Davidson I. (2000) The human TFIID components TAF(II)135 and TAF(II)20 and the yeast SAGA components ADA1 and TAF(II)68 heterodimerize to form histone-like pairs. *Mol. Cell. Biol.* **20**, 340–351
4. Selzer, G., Som, T., Itoh, T., and Tomizawa, J. (1983) The origin of replication of plasmid p15A and comparative studies on the nucleotide sequences around the origin of related plasmids. *Cell* **32**, 119–129
5. Lunin, V. V., Munger, C., Wagner, J., Ye, Z., Cygler, M., and Sacher, M. (2004) The structure of the MAP kinase scaffold MP1 bound to its partner p14: A complex with a critical role in endosomal MAP kinase signaling. *J. Biol. Chem.* **279**, 23422–23430
6. Stebbins, C. E., Kaelin, W. G., Jr., and Pavletich, N. P. (1999) Structure of the VHL-ElonginC-ElonginB complex: Implications for VHL tumor suppressor function. *Science* **284**, 455–461
7. Alexandrov, A., Martzen, M. R., and Phizicky, E. M. (2002) Two proteins that form a complex are required for 7-methylguanosine modification of yeast tRNA. *RNA* **8**, 1253–1266
8. Aslanidis, C., and de Jong, P. J. (1990) Ligation-independent cloning of PCR products (LIC-PCR). *Nucleic Acids Res.* **18**, 6069–6074
9. Alexandrov, A., Dutta, K., and Pascal, S. M. (2001) MBP fusion protein with a viral protease cleavage site: One-step cleavage/purification of insoluble proteins. *BioTechniques* **30**, 1194–1198
10. Summers, D. K., and Sherratt, D. J. (1984) Multimerization of high copy number plasmids causes instability: ColE1 encodes a determinant essential for plasmid monomerization and stability. *Cell* **36**, 1097–1103
11. Chatwin, H. M., and Summers, D. K. (2001) Monomer-dimer control of the ColE1 P(cer) promoter. *Microbiology* **147**, 3071–3081