Proc. Natl. Acad. Sci. USA
Vol. 87, pp. 1937–1941, March 1990
Biochemistry

Overproduction and dissection of proteins by the expression-cassette polymerase chain reaction

(protein expression/CD4/domain)

KURTIS D. MACFERRIN, MICHAEL P. TERRANOVA, STUART L. SCHREIBER*, AND GREGORY L. VERDINE*

Department of Chemistry, Harvard University, Cambridge, MA 02138

Communicated by George M. Whitesides, December 15, 1989 (received for review August 21, 1989)

ABSTRACT We report an efficient, general approach for the construction of protein-overproducing strains of *Escherichia coli*. The method, expression-cassette polymerase chain reaction (ECPCR), allows the insertion of virtually any contiguous coding sequence between sequences that direct high-level protein biosynthesis in *E. coli*. The gene expression cassettes obtained by ECPCR are inserted into a regulated overexpression plasmid, and the resulting construct is used to transform *E. coli*. By effecting simultaneous 5′ and 3′ modification of a coding sequence, ECPCR permits the facile generation of mutant proteins having N- and/or C-terminal truncations. The method is a highly efficient way to dissect a multidomain protein into its component domains. The efficiency of the ECPCR approach is demonstrated in this study by construction of permuted overexpression vectors for the first two extracellular domains of the human CD4 protein.

The ability to rationally engineer protein-overproducing strains of *Escherichia coli* ("overproducers") has transformed the science of protein structure and function. Construction of an overproducer generally requires (*i*) replacement of the gene's (or cDNA's) native expression signals with strong, switchable signals; (*ii*) translocation of the refashioned gene to an overexpression plasmid that can be manipulated, characterized, and introduced into recipient cells; and (*iii*) stable transformation of *E. coli* by the recombinant plasmid (1). Here we describe a general procedure for protein overproduction, termed the expression-cassette polymerase chain reaction (ECPCR), which facilitates overproducer construction by effecting site-specific replacement of the 5′ and 3′ ends of the gene with expression sequences derived from synthetic oligonucleotides. ECPCR also takes advantage of the now-routine ability to introduce restriction sites via the polymerase chain reaction (PCR, ref. 2), yielding gene expression cassettes that are suitably equipped for cloning.

PCR permits targeted amplification of DNA sequences (reviewed in ref. 3). The specific segment of DNA amplified in PCR is dictated by the choice of base-pairing sites for oligonucleotide primers (Fig. 1). DNA sequences of the primers are incorporated entirely into the product DNA, thus providing the opportunity to replace sequence information of the native gene with new information borne on the synthetic primers. In ECPCR, sequences required for protein translation and restriction endonuclease digestion are incorporated into the primers, so that ECPCR of the target DNA results in synthesis of an expression cassette bearing all of the necessary information for cloning and translation in *E. coli* (Figs. 1 and 2). Transcriptional sequence elements—i.e., a strong, regulated promoter (4) and an efficient transcription terminator (5)—are supplied by the overexpression plasmid. Since the choice of protein-coding information inserted between

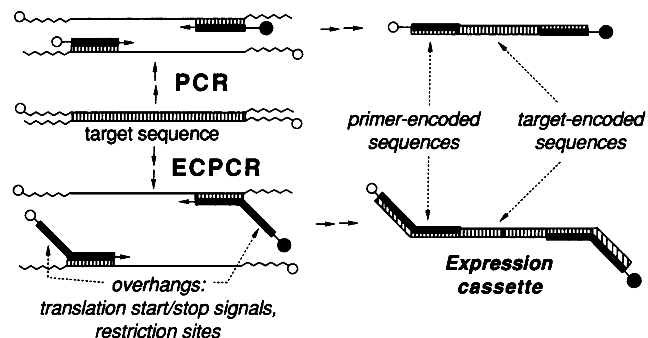

FIG. 1. Comparison of PCR and ECPCR. In PCR, primers that are complementary to the target sequence are used to amplify target DNA; in ECPCR, translational control sequences and restriction sites included in the primers are fused to the target sequence, thus providing an expression cassette ready for insertion into a bacterial overexpression vector.

the expression signals is dependent only upon the coding information incorporated into the ECPCR primers, this method can be used to dissect a multidomain protein into smaller proteins comprising its individual component domains. This domainal analysis strategy can facilitate the localization of ligand-binding and catalytic functions, in addition to determining the degree of interdomainal cooperativity. Functional modules identified by this process are, *inter alia*, targets for high-resolution structural analysis and can be incorporated into designed hybrid proteins (6–8).

We have chosen the human CD4 protein and its corresponding cDNA as a model system for overexpression and domainal analysis by ECPCR. CD4, a membrane-bound glycoprotein expressed on the surface of helper T lymphocytes, augments the interactions of the T-cell receptor with antigen-bound class II major histocompatibility proteins (9–11). CD4 also functions as a specific cell-surface receptor for human immunodeficiency virus (12); consequently, considerable interest has arisen in development of non-membrane-bound ("soluble") CD4 derivatives as competitive inhibitors of viral adhesion (6–8, 13–19). On the basis of its cDNA-derived amino acid sequence (refs. 20–22; corrected in ref. 23), the extracellular segment of CD4 ("soluble CD4") has been proposed to possess four immunoglobulin-like domains, of which the first two (outermost from the membrane) have been strongly implicated in binding to human immunodeficiency virus and class II major histocompatibility antigens (19, 24–26).

To elucidate the recognition of macromolecular ligands by CD4, we wished to dissect the protein into its component domains, so that these individual proteins could be analyzed

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: (EC)PCR, (expression-cassette) polymerase chain reaction; RBS, ribosome binding site; TSE, translational spacer element; IPTG, isopropyl β-D-thiogalactopyranoside.
*To whom reprint requests should be addressed.

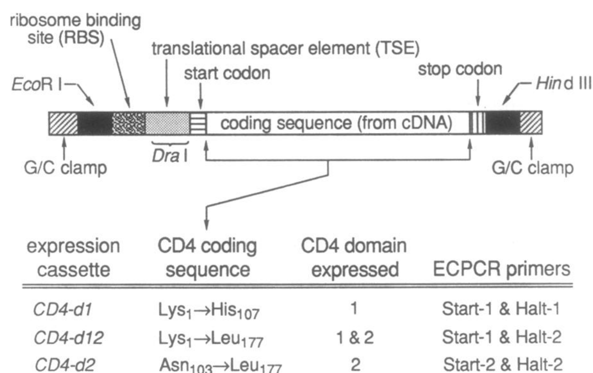

FIG. 2. CD4 expression cassettes generated in this study. (*Upper*) General structure, with relevant sequence features denoted. (*Lower*) Specific parameters of the three expression cassettes.

by high-resolution structural methods. Such highly targeted cleavages can, in practice, only be effected by manipulations of the protein's coding sequence, rather than cleavage of the protein itself. Instead of using an *ad hoc* approach to obtain an overproducer for each CD4 domain, we decided to explore the development of a general overexpression protocol that would access high-level biosynthesis of any contiguous coding sequence. These efforts led to the development of the ECPCR method, which is demonstrated here in the generation of three *E. coli* strains that respectively overproduce CD4 domain 1, domain 2, and domains 1 plus 2.

MATERIALS AND METHODS

Gene Sources and Preparation. A human CD4 cDNA insert in pZIPneoSVX (courtesy of S. Burakoff, Harvard Medical School) was excised by *Hin*dIII digestion and inserted into *Hin*dIII-digested pBS(+) (Stratagene); single-stranded DNA was isolated and purified according to the supplier's protocol. Phage DNA from the human KBM-7 library in λgt10 (courtesy of S. Reddy and B. Cochran, Massachusetts Institute of Technology) was used as obtained; the human peripheral blood acute lymphoblastoid leukemia (HPB-ALL) phagemid library (courtesy of B. Seed, Harvard Medical School) was digested with *Not* I and ethanol-precipitated prior to use. Human Jurkat T-cell RNA (courtesy of B. Bierer and S. Burakoff) and murine BALB/c RNA (courtesy of M. L. Hedley and T. Maniatis, Harvard University) were primed with poly(dT) and extended with reverse transcriptase (27); the RNA·DNA hybrid mixtures were used directly in PCR.

ECPCR Amplifications. Amplifications were carried out in 0.5-ml microcentrifuge tubes containing 10 mM Tris (pH 8.0 at 25°C), 50 mM KCl, 1.5 mM $MgCl_2$, 0.1% SeaPlaque agarose (FMC), 0.1% Triton X-100, 200 μM each dNPT, 1 μM primers, ≈25 ng of template DNA, and 4 units of *Thermus aquaticus* (*Taq*) DNA polymerase (Promega). Reaction mixtures (25 or 50 μl) were overlaid with an equal volume of paraffin oil and subjected to 10–30 cycles in a programmable thermal cycler (MJ Research, Cambridge, MA) using the following sequence: 94°C for 2 min, 42°C for 2 min, and 70°C for 5 min. The amplification products were analyzed in a 2% agarose (BRL) gel in TBE buffer (90 mM Tris/90 mM boric acid/2 mM EDTA). The identity of ECPCR cassettes from various sources was verified by restriction mapping (data not shown).

Protein Expression. The expression vector was pHN1 (H. Nash and G.L.V., unpublished work), a phagemid in which cloned genes are under the control of the *tac* promoter (28) and $rrnBT_1T_2$ transcription terminator (29). Blunt-ended expression cassettes were digested with *Eco*RI and *Hin*dIII and inserted into *Eco*RI/*Hin*dIII-cut pHN1 to yield the overproducing phagemids pCD4-d1, -d2, and -d12. These were transformed into *E. coli* XA 90 F′$lacI^{Q1}$, insert-positive clones were isolated, and the sequence of each insert was verified by sequencing (30) of single-stranded phagemid DNA with the Sequenase kit (United States Biochemical). A 10-ml overnight culture of each recombinant was used to inoculate 1 liter of LB/ampicillin; growth was monitored until OD_{550} = 0.6–0.7, and the culture was induced by the addition of 10 ml of sterile-filtered isopropyl β-D-thiogalactopyranoside (IPTG) solution (48 mg/ml). After 16 hr, the cells were harvested by centrifugation. Whole-cell pellets from 1 ml of culture were lysed by boiling for 3 min in 100 μl of SDS/PAGE sample buffer (4% SDS/0.1% bromophenol blue/20% glycerol/10% 2-mercaptoethanol/0.12 mM Tris, pH 6.8). The wet cell paste from 1 liter of culture (5–8 g) was resuspended in 100 mM Tris, pH 8/1 mM EDTA/5% glycerol/10 mM 2-mercaptoethanol and lysed in a French pressure cell. The crude lysate was centrifuged, and the pellet was washed to remove entrapped soluble material. In all cases, the recombinant protein was present in the cell pellet and absent from the supernatant (data not shown), as judged by SDS/PAGE (31). Proteins CD4-d1, CD4-d2, and CD4-d12 were solubilized and purified by modifications of a literature procedure (19), which will be reported elsewhere.

RESULTS

Locus of Interdomain Incision of CD4. Dissection of CD4 to yield intact domainal fragments requires severance within the joining (J) regions, which for the two N-terminal domains of CD4 are found at residues 93–109 (J_1 region: connects domains 1 and 2) and 165–179 (J_2 region: connects domains 2 and 3) (Fig. 3; ref. 25). Modeling studies (K.D.M. and S.L.S., unpublished) suggest that the most independent peptide locus between CD4 domains 1 and 2 is the hydrophilic pentapeptide Asn^{103}-Ser-Asp-Thr-His^{107}. Nonetheless, since even this segment may be integral to one or both of the appended domains, it has been included in both the C terminus of domain 1 and the N terminus of domain 2 (*vide infra* and Fig. 3).

The precise locus of connection between domains 2 and 3 is less well predicted because that region is less homologous to crystallographically determined immunoglobulins. Nonetheless, since the location of introns in genes of the immunoglobulin superfamily generally correlates with the location of domainal boundaries in the proteins (32), we used this criterion to dictate the dissection of domains 2 and 3. An intron occurs in the codon for Ala^{178} (21), leading us to sever domains 2 and 3 between Leu^{177} and Ala^{178} (Fig. 3).

Structure of CD4 Expression Cassettes. The general structure for the CD4 expression cassettes is shown in Fig. 2. Three cassettes have been constructed, each bearing different CD4 coding sequence sandwiched between cloning and *E. coli* expression sequences. The *CD4-d1* expression cassette carries the CD4 cDNA sequence specifying Lys^1–His^{107}, corresponding to the native protein's first domain; *CD4-d12* encodes Lys^1–Leu^{177}, corresponding to domains 1 and 2; and *CD4-d2* encodes Asn^{103}–Leu^{177}, corresponding to domain 2.

The expression cassettes of Fig. 2 were generated by ECPCR of the native CD4 cDNA using two primers, one that installed 5′ sequence information (Start primer) and another that installed 3′ sequence information (Halt primer). The Start primers (Fig. 4) had a CGCGC sequence ("G/C clamp"), the GAATTC recognition site for *Eco*RI endonuclease, the AGGAGG consensus *E. coli* RBS, an 8-base A/T-rich TSE (for reviews, see refs. 33 and 34), and the universal ATG start codon, followed by coding information for the desired N terminus of the recombinant protein. In addition, a *Dra* I restriction site (TTTAAA) has been incorporated into the TSE, so that a periplasmic signal sequence may be added by insertion of a double-stranded oligonucleotide cassette: restriction by *Dra* I leaves a downstream AAA

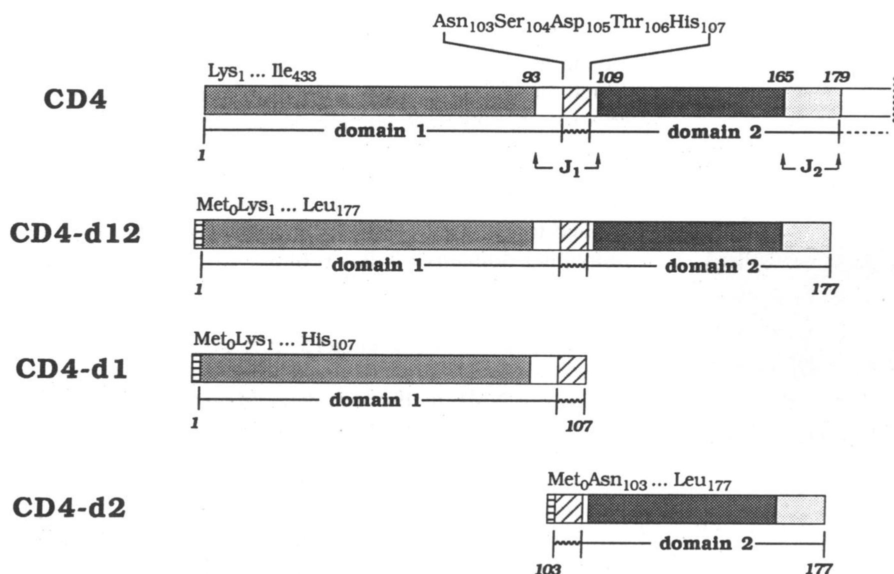

FIG. 3. Schematic representation of CD4 primary structure and the structure of recombinant CD4 proteins generated in this study. Segments corresponding to CD4 domains are indicated below the bar diagrams; locations of joining (J) regions are indicated. The N- and C-terminal residues of each protein (CD4 numbering as in ref. 23) are displayed at the top left of the respective structures. The sequence of the hydrophilic pentapeptide in the J_1 region is shown in the expansion above CD4. Met^0 refers to the primer-encoded (non-CD4) residue that initiates translation in the recombinant proteins.

blunt end that may be used to encode the lysine residue at the maltose-binding protein signal-peptide cleavage site (35). The Halt primers (Fig. 4) had a "G/C clamp" and the AAGCTT recognition sequence for *Hin*dIII, followed by a stop anticodon and anticoding sequence for the desired C terminus of the recombinant protein. The recombinant proteins specified by these expression cassettes are predicted to possess a Start primer-encoded N-terminal methionine (Met^0) followed by the cDNA-encoded CD4 protein sequence; translation terminates at the Halt primer-encoded stop codon. ECPCRs using different combinations of Start and Halt primers thus give rise to expression cassettes bearing different, predetermined coding sequences (Fig. 2).

Synthesis of CD4 Expression Cassettes by ECPCR. ECPCR of a CD4 cDNA using the Start-1 and Halt-2 primers was predicted to yield a 572-base-pair (bp) expression cassette (*CD4-d12*) that would direct the expression of CD4 domains 1 and 2 (Fig. 2). Similarly, use of the Start-1 and Halt-1 primers should give a 362-bp expression cassette (*CD4-d1*) for domain 1; Start-2 and Halt-2 should give a 251-bp expression cassette (*CD4-d2*) for domain 2. Agarose gel electrophoresis of parallel ECPCRs using these primer combinations showed DNA fragments of the predicted sizes (Fig. 5*A*). Expression-cassette synthesis by direct amplification from a human T-cell or myeloblastoid cell library was also efficient (Fig. 5*B*). The generality of this approach was further extended by using a crude Jurkat T-cell RNA preparation (converted to first-strand cDNA) as the source of CD4 coding sequence (Fig. 5*B*; only domain 2 cassette shown). Analogous sets of expression cassettes were generated using mouse spleen, liver, and thymus RNA samples (Fig. 5*C*; only domain 1 cassette shown), despite differences in the human and mouse CD4 sequences (21). The proteins encoded by the cassettes in Fig. 5*C* are mouse–human chimeras, with human CD4 sequence derived from the ECPCR primers, and mouse CD4 sequence from the amplified murine cDNA.

Expression of the Redesigned CD4 Genes. The latent cohesive ends of the expression cassettes were revealed by digestion with *Eco*RI and *Hin*dIII, allowing unidirectional insertion of the constructs into the phagemid expression vector pHN1. The three expression vectors (pCD4-d12, pCD4-d1, and pCD4-d2, with respective inserts *CD4-d12*, *CD4-d1*, and *CD4-d2*) were then transformed into *E. coli* XA 90 F'*lacI*^{Q1} and tested for induction of their respective recombinant proteins. Synthesis of proteins having approximately the predicted relative molecular mass (CD4-d12, M_r 19,708; CD4-d1, M_r 12,209; and CD4-d2, M_r 8254) was induced upon addition of IPTG (Fig. 6).

Preliminary Characterization of the Recombinant Proteins. Upon lysis of cells that express CD4-d12, -d1, or -d2, the recombinant proteins cosediment with the cell debris, thus indicating that they are present in inclusion bodies (Fig. 6). The inability of *E. coli* to biosynthesize members of the immunoglobulin superfamily in a soluble cytoplasmic form appears to be a general phenomenon (6, 18, 19, 36–39). Two approaches are commonly used to obtain native, correctly folded protein (40): (*i*) export of the protein to the periplasm by fusion of a signal sequence to the gene (22, 41–43); (*ii*) *in vitro* denaturation/renaturation of the improperly folded protein (6, 18, 19, 36–39). Both of these options have been successfully exercised in reported truncations of CD4 (19, 25). We chose the second option because the selective partitioning of the recombinant CD4 proteins into the solid cell pellet greatly simplified their purification (compare lanes I and Pt in Fig. 6). These polypeptides are readily renatured by conventional methods (19). Further details of the renaturation, purification, and full characterization of CD4-d12, -d1, and -d2, will be reported elsewhere.

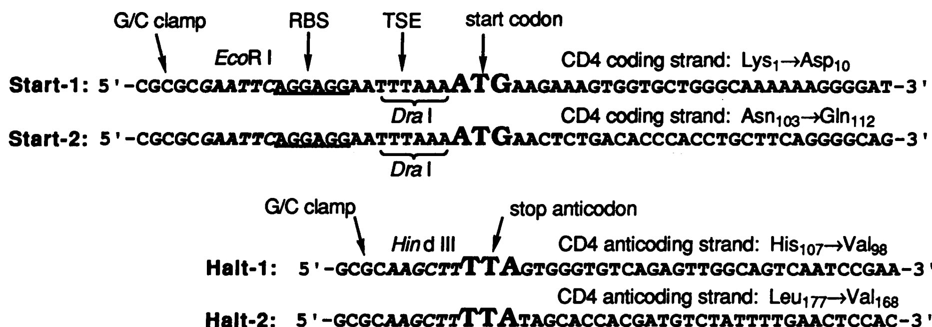

FIG. 4. Sequences of the PCR primers used to redesign the CD4 gene. Sequence features present in the primers are indicated. Use of these primers in various combinations gives permuted expression cassettes for the first two domains of CD4 (see Fig. 2).

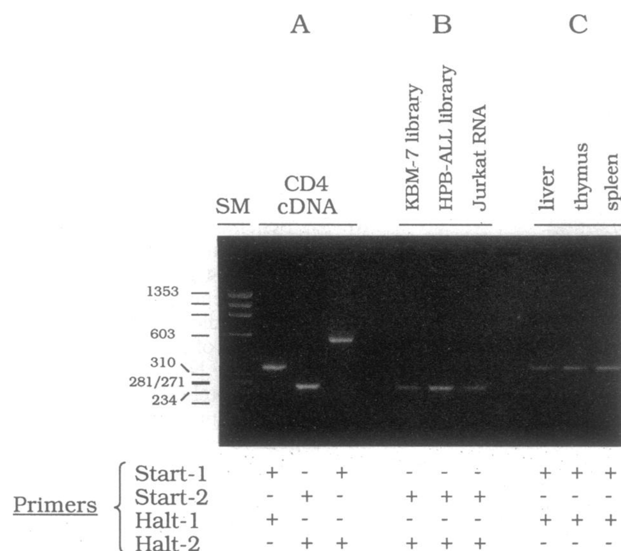

FIG. 5. PCR synthesis of CD4 gene cassettes using various sources of the native CD4 gene. (*A*) CD4 cDNA. (*B*) Human KBM-7 myeloblastoid cell cDNA library, human HPB-ALL T-cell cDNA library, and total Jurkat T-cell RNA. (*C*) BALB/c mouse RNA from various tissues. "RNA" samples are actually RNA·DNA hybrids arising from first-strand cDNA synthesis. Lane SM, DNA size markers (*Hae* III-digested ϕX174 phage DNA) with relevant fragment sizes indicated (bp). In *B* only the products from amplification using the Start-2 and Halt-2 primers (domain 2 cassette) are shown; the other two primer combinations give rise to their respective cassettes (data not shown). In *C*, only the products from amplification using the Start-1 and Halt-1 primers (domain 1 cassette) are shown; the other primer combinations were not tested.

N-terminal sequencing of CD4-d1 and -d12 yielded the predicted sequence, Met-Lys-Lys-Val-Val, and CD4-d2 gave the predicted Met-Asn-Ser-Asp-Thr, indicating that the recombinant species had not been posttranslationally cleaved. CD4-d1 thus is composed of CD4 residues Lys^1–His^{107} preceded by Met^0; CD4-d12 of Lys^1–Leu^{177} preceded by Met^0; and CD4-d2 of Asn^{103}–Leu^{177} preceded by Met^0 (Fig. 3).

DISCUSSION

We have developed a facile method that effects the formal insertion of coding DNA sequence information between efficient translational signals. Upon incorporation into an overproduction vector bearing a strong, regulated *E. coli* promoter and a transcription terminator, the resulting construct directs efficient biosynthesis of the cassette-encoded protein in *E. coli*. This method is applicable to the overproduction of entire proteins and fragments of proteins, and thus it is suitable for the rational generation of mutant proteins having N- and/or C-terminal truncations. In those cases in which biochemical or sequence information suggests the location of domainal boundaries in a protein—e.g., in the steroid hormone receptor (ref. 44 and references therein), complement control protein (ref. 45 and references therein), and adhesion (32, 46, 47) superfamilies—this method may be applied to the construction of *E. coli* expression vectors for individual domains of that protein, thereby facilitating the elucidation of its structure–function relationships.

The recombinant CD4 fragments overexpressed in the present study were present as 5–10% of the total cell protein, a level at which sufficient material for structural elucidation by x-ray or NMR methods may readily be obtained. While no particular effort was made to optimize the RBS or TSE sequences borne on the Start primers, the chosen sequences of these translational elements reflect currently accepted principles: (*i*) the consensus hexanucleotide RBS (AGGAGG) is stronger than its shorter variants (34); (*ii*) TSEs of 8–9 nucleotides appear to be optimal, although TSEs having from 6 to 12 nucleotides function well (34); and (*iii*) an A/T-rich TSE is preferred (48, 49). Increased expression of recombinant proteins is likely to be achievable through variation of the TSE and RBS sequences; however, ample evidence suggests that there is no universally optimal sequence element for the initiation of translation in *E. coli* (31).

The use of PCR to add new sequence information concomitant with amplification has found wide applicability in recombinant DNA technology, and the ECPCR protocol is representative of such methods (3). Particularly intriguing is the demonstration that the T7 promoter sequence can be added to target DNA in order to facilitate sequence analysis (50, 51); indeed, those results indicate that it should be possible to engineer T7-based overproduction systems using an ECPCR-like strategy in combination with previously described vectors (52). This study demonstrates that a sequence motif which directs efficient translation of several different coding sequences can be introduced successfully by PCR and that these sequences fused to the vector-borne *tac* promoter afford large quantities of recombinant proteins. The length of noncomplementary sequences added during ECPCR is primarily limited not by PCR amplification, but by the size limitation of automated DNA synthesis (routinely >100 nucleotides); hence, the potential exists to incorporate additional sequence motifs (e.g., synthetic promoters or periplasmic signal sequences) during the ECPCR procedure. Efforts to transplant the mammalian immune repertoire into *E. coli* (53, 54) can be greatly augmented by direct periplasmic expression of functional antibody fragments, and ECPCR could accomplish this without the need for construction of specialized vectors. Preliminary results from our

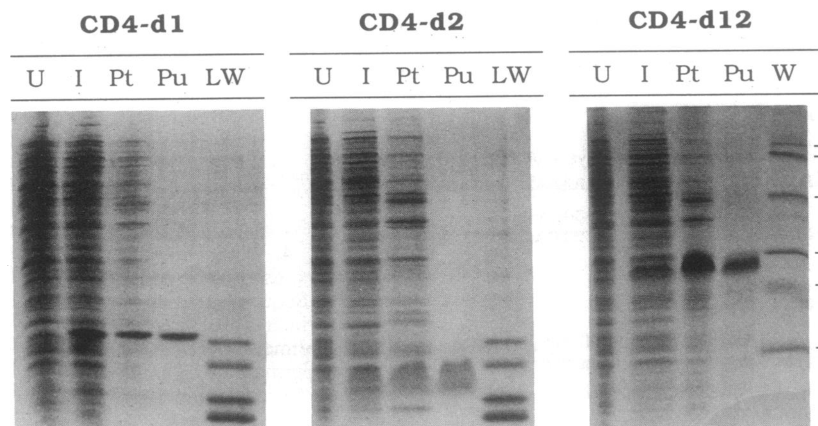

FIG. 6. Expression in *E. coli* of recombinant CD4 proteins CD4-d1, -d2, and -d12. Lanes: U, uninduced cells, total cell lysate; I, induced cells, total cell lysate; Pt, pellet from lysed induced cells; Pu, purified protein; LW, low molecular weight markers [Sigma; $M_r \times 10^{-3}$ (from top) 16.9, 14.4, 10.7, 8.2, and 6.2]; W, intermediate molecular weight markers [Bio-Rad, $M_r \times 10^{-3}$ (from top) 97.4, 66.2, 42.7, 31.0, 21.5, and 14.4]. All of the CD4 recombinant proteins have retarded mobilities relative to their known molecular weights; however, none of the molecular weight standards are of the immunoglobulin superfamily. In addition, CD4-d2 shows a characteristic tendency to blur upward when heavily loaded. Proteins were detected by staining with Coomassie blue.

laboratories indicate that construction of exportation cassettes by ECPCR is viable.

A distinct advantage of the ECPCR procedure is that it does not require the selection/screening of individual clones from a gene library, provided that the sequence of that gene is known. As described above for the case of CD4, the target gene can be amplified directly from a suitable cDNA library or RNA preparation to yield the desired expression cassette. This procedure thus provides ready access to overproducers for native and redesigned proteins, while eliminating the need for physical exchange of cDNA clones. Owing to the success of ECPCR observed in the CD4 system, we and others have used this method to construct two additional overproducers, which required less than 2 weeks from primer design to verification of overproduction: human cyclosporin A-binding protein, cyclophilin (ref. 55; J. Liu, M. Albers, C. M. Chen, S.L.S., and C. T. Walsh, unpublished results), and human FK506 binding protein, FKBP (refs. 56 and 57; R. F. Standaert, G.L.V., and S.L.S., unpublished results). These recent successes lead us to conclude that ECPCR is the most expedient method presently available for overproducing proteins and protein fragments in *E. coli*.

We thank B. E. Bierer and S. J. Burakoff for the CD4 cDNA (pZIPneoSVX) clone and for the human T-cell RNA preparation, S. Reddy and B. Cochran for the λgt10 cDNA library, B. Seed for the pH3M cDNA library, M. L. Hedley and T. Maniatis for the BALB/c RNA samples, and M. Ptashne for *E. coli* XA 90 F′*lacI*^{Q1}. We thank W. S. Lane of the Harvard Microchemistry Facility for protein analysis. The experimental assistance of K. Hayashibara, H. Nash, K. Ezaz-Nikpay, and R. Standaert is gratefully acknowledged. K.D.M. is a recipient of a National Science Foundation graduate fellowship. This work was supported by a National Institutes of Health grant (GM30738-09) to S.L.S. and by a Camille and Henry Dreyfus Foundation Distinguished New Faculty Award and a Du-Pont New Faculty Fellowship to G.L.V.

1. Perbal, B. (1988) *A Practical Guide to Molecular Cloning* (Wiley, New York), pp. 731–794.
2. Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A. & Arnheim, N. (1985) *Science* **230**, 1350–1354.
3. Erlich, H. A. (1989) *PCR Technology: Principles and Applications for DNA Amplification* (Stockton, New York).
4. McClure, W. R. (1985) *Annu. Rev. Biochem.* **54**, 171–204.
5. Brosius, J., Dull, T. J., Sleeter, D. D. & Noller, H. F. (1981) *J. Mol. Biol.* **148**, 1107–1127.
6. Choudhary, V. K., Queen, C., Junghans, R. P., Waldman, T. A., FitzGerald, D. J. & Pastan, I. (1989) *Nature (London)* **339**, 394–397.
7. Capon, D. J., Chamow, S. M., Mordenti, J., Marsters, S. A., Gregory, T., Mitsuya, H., Byrn, R. A., Lucas, C., Wurm, F. M., Groopman, J. E., Broder, S. & Smith, D. H. (1989) *Nature (London)* **337**, 525–531.
8. Traunecker, A., Schneider, J., Kiefer, H. & Karjalainen, K. (1989) *Nature (London)* **339**, 68–70.
9. Swain, S. L. (1983) *Immunol. Rev.* **74**, 129–142.
10. Gay, D., Maddon, P., Sekaly, R., Talle, M. A., Godfrey, M., Long, E., Goldstein, G., Chess, L., Axel, R., Kappler, J. & Marrack, P. (1987) *Nature (London)* **328**, 626–629.
11. Doyle, C. & Strominger, J. L. (1987) *Nature (London)* **330**, 256–259.
12. McDougal, J. S., Kennedy, M. S., Sligh, J. M., Cort, S. P., Malwe, A. & Nicholson, J. K. A. (1986) *Science* **231**, 382–385.
13. Smith, D. H., Byrn, R. A., Marsters, S. A., Gregory, T., Groopman, J. E. & Capon, D. J. (1987) *Science* **238**, 1704–1707.
14. Fischer, R. A., Bertonis, J. M., Meier, W., Johnson, V. A., Costopoulos, D. S., Liu, T., Tizard, R., Walker, B. D., Hirsch, M. S., Schooley, R. T. & Flavell, R. A. (1988) *Nature (London)* **331**, 76–78.
15. Hussey, R. E., Richardson, N. E., Kowalski, M., Brown, N. R., Chang, H.-C., Siliciano, R. F., Dorfman, T., Walker, B., Sodroski, J. & Reinherz, E. (1988) *Nature (London)* **331**, 78–81.
16. Deen, K. C., McDougal, J. S., Inacker, R., Folena-Wasserman, G., Arthos, J., Rosenburg, J., Maddon, P. J., Axel, R. & Sweet, R. W. (1988) *Nature (London)* **331**, 82–84.
17. Traunecker, A., Lüke, W. & Karjalainen, K. (1988) *Nature (London)* **331**, 84–86.
18. Choudhary, V. K., Mizakumi, T., Fuerst, T. R., FitzGerald, D. J., Moss, B., Pastan, I. & Berger, E. A. (1988) *Nature (London)* **335**, 369–372.
19. Chao, B. H., Costopoulos, D. S., Curiel, T., Bertonis, J. M., Chisholm, P., Williams, C., Schooley, R. T., Rosa, J. J., Fisher, R. A. & Maraganore, J. M. (1989) *J. Biol. Chem.* **264**, 5812–5817.
20. Maddon, P. J., Littman, D. R., Godfrey, M., Maddon, D. E., Chess, L. & Axel, R. (1985) *Cell* **42**, 93–104.
21. Maddon, P. J., Molineaux, S. M., Maddon, D. E., Zimmerman, K. A., Godfrey, M., Alt, F. W., Chess, L. & Axel, R. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 9155–9159.
22. Clark, S. J., Jefferies, W. A., Barclay, A. N., Gagnon, J. & Williams, A. F. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 1649–1653.
23. Littman, D. R., Maddon, P. J. & Axel, R. (1988) *Cell* **55**, 541.
24. Berger, E. A., Fuerst, T. R. & Moss, B. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2357–2361.
25. Arthos, J., Deen, K. C., Chaikin, M. A., Fornwald, J. A., Sathe, G., Sattentau, Q. J., Clapham, P. R., Weiss, R. A., McDougal, J. S., Pietropaolo, C., Axel, R., Truneh, A., Maddon, P. J. & Sweet, R. W. (1989) *Cell* **57**, 469–481.
26. Lamarre, D., Capon, D. J., Karp, D. R., Gregory, T., Long, E. O. & Sekaly, R.-P. (1989) *EMBO J.* **8**, 3271–3277.
27. Klickstein, L. B. & Neve, R. L. (1989) in *Current Protocols in Molecular Biology*, eds. Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Smith, J. A., Seidman, J. G. & Struhl, K. (Greene, and Wiley, New York), pp. 5.5.1–5.5.9.
28. Amann, E., Brosius, J. & Ptashne, M. (1983) *Gene* **25**, 167–178.
29. Brosius, J., Dull, T. J., Sleeter, D. D. & Noller, H. F. (1981) *J. Mol. Biol.* **148**, 107–127.
30. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
31. Laemmli, U. K. (1970) *Nature (London)* **227**, 680–685.
32. Williams, A. F. & Barclay, A. N. (1988) *Annu. Rev. Immunol.* **6**, 381–405.
33. Dahlberg, A. E. (1989) *Cell* **57**, 525–529.
34. Gold, L. (1988) *Annu. Rev. Biochem.* **57**, 199–233.
35. Bedouelle, H. (1980) *Nature (London)* **285**, 78–81.
36. Huston, J. S., Levinson, D., Mudgett-Hunter, M., Tai, M.-S., Novotny, J., Margolies, M. N., Ridge, R. J., Bruccoleri, R. E., Haber, E., Crea, R. & Oppermann, H. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 5879–5883.
37. Bird, R. E., Hardman, K. D., Jacobson, J. W., Johnson, S., Kaufman, B. M., Lee, S.-M., Lee, T., Pope, S. H., Riordan, G. S. & Whitlow, M. (1988) *Science* **242**, 423–426.
38. Boss, M. A., Kenten, J. H., Wood, C. R. & Emtage, J. S. (1984) *Nucleic Acids Res.* **12**, 3791–3806.
39. Cabilly, S., Riggs, A. D., Pande, H., Shively, J. E., Holmes, W. E., Rey, M., Perry, L. J., Wetzel, R. & Heyneker, H. L. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 3273–3277.
40. Schein, C. H. (1989) *Bio/Technology* **7**, 1147–1149.
41. Skerra, A. & Pluckthun, A. (1988) *Science* **240**, 1038–1040.
42. Better, M., Chang, C. P., Robinson, R. R. & Horowitz, A. H. (1988) *Science* **240**, 1041–1043.
43. Ward, E. S., Gussow, D., Griffiths, A. D., Jones, P. T. & Winter, G. (1989) *Nature (London)* **341**, 544–546.
44. Evans, R. M. (1988) *Science* **240**, 889–895.
45. Reid, K. B. M. & Day, A. J. (1989) *Immunol. Today* **10**, 177–180.
46. Edelman, G. (1989) *Sci. Am.* **260**, 76–88.
47. White, J. M. & Littman, D. R. (1989) *Cell* **56**, 725–728.
48. de Boer, H. A., Comstock, L. J., Hui, A., Wong, E. & Vasser, M. (1983) *Biochem. Soc. Symp.* **48**, 233–244.
49. Matteuci, M. & Heyneker, H. L. (1983) *Nucleic Acids Res.* **11**, 3113–3121.
50. Stoflet, E. S., Koeberi, D. D., Sarkar, G. & Sommer, S. S. (1988) *Science* **239**, 491–494.
51. Sarkar, G. & Sommer, S. S. (1988) *Nucleic Acids Res.* **16**, 5197.
52. Tabor, S. & Richardson, C. C. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 1074–1078.
53. Ward, E. S., Gussow, D., Griffiths, A. D., Jones, P. T. & Winter, G. (1989) *Nature (London)* **341**, 544–546.
54. Huse, W. D., Sastry, L., Iverson, S. A., Kang, A. S., Alting-Mees, M., Burton, D. R., Benkovic, S. J. & Lerner, R. A. (1989) *Nature (London)* **246**, 1275–1281.
55. Haendler, B., Hofer-Warbinek, R. & Hofer, E. (1987) *EMBO J.* **6**, 947–950.
56. Siekierka, J. J., Hung, S. H. Y., Poe, M., Lin, C. S. & Sigal, N. (1989) *Nature (London)* **341**, 755–757.
57. Harding, M. W., Galat, A., Uehling, D. E. & Schreiber, S. L. (1989) *Nature (London)* **341**, 758–760.