Use of T7 RNA Polymerase to Direct Expression of Outer Surface Protein A (OspA) from the Lyme Disease Spirochete, *Borrelia burgdorferi*.

John J. Dunn and Barbara N. Lade
Biology Department
Brookhaven National Laboratory
Upton, NY 11973

ABSTRACT

The *ospA* gene from a North American strain of the Lyme disease Spirochete, *Borrelia burgdorferi*, has been cloned under the control of transcription and translation signals from bacteriophage T7. Full-length OspA protein, a 273 amino acid (31kD) lipoprotein, is expressed poorly in *Escherichia coli* and is associated with the insoluble membrane fraction. In contrast, a truncated form of OspA lacking the amino-terminal signal sequence which normally would direct localization of the protein to the outer membrane is expressed at very high levels (≥ 100mg /liter) and is soluble. The truncated protein has been purified to homogeneity and is being tested to see if it will be useful as an immunogen in a vaccine against Lyme disease. Circular dichroism and fluorescence spectroscopy has been used to characterize the secondary structure and study conformational changes in the protein. Studies underway with other surface proteins from *B. burgdorferi* and a related spirochete, *B. hermsii*, which causes relapsing fever, leads us to conclude that a strategy similar to that used to express the truncated OspA can provide a facile method for producing variations of Borrelia lipoproteins which are highly expressed in *E. coli* and soluble without exposure to detergents.

INTRODUCTION

Lyme disease is caused by the spirochete *Borrelia burgdorferi* and this complex disorder is the most common vector-borne infection in the USA. Specific diagnostic tests and ultimately a vaccine for Lyme disease are of major importance. In this regard, the outer surface lipoproteins (OspA and OspB) are of interest because they are unique to *B. burgdorferi* and not shared with other Borrelia or spirochetes. These proteins are highly immunogenic in experimental animals, however, the antibody response to these antigens occurs late in the course of natural infection and only in a small minority of patients (<5%) [1-3]. While the exact functions of these proteins are unknown, Osp-enriched fractions have been utilized as immunodiagnostics for late disease and in the design of "selective Westerns". The genes for these two proteins (*ospA* & *ospB*) are encoded in tandem in a 2KB stretch of a 49 kB linear plasmid separated one from the other by only 12 base pairs and co-transcribed from the same promoter [4, 5]. The deduced amino acid sequences of OspA and OspB begin with N-terminal sequences believed to constitute signal sequences which normally direct localization of the protein products to the spirochetes outer membrane.

We have chosen OspA as a model system for studying expression of *B. burgdorferi* proteins, particularly surface proteins, in *E. coli*. These proteins most likely affect the antigenicity, immunological reactivity, host cell interactions of the spirochete, invasion of the host, and development of symptoms; information about these proteins is expected to be useful in the diagnosis, treatment and prevention of the disease. Our goal is to use recombinant DNA techniques to overexpress these proteins to standardize immunoassays and other diagnostic screening tests and to serve as potential antigens for a vaccine.

METHODS

T7/pET system

B. burgdorferi OspA protein was expressed using the T7/pET expression system developed at Brookhaven [6] in which the RNA polymerase of bacteriophage T7 is used to drive transcription of a cloned target sequence. Bacteriophage T7 RNA polymerase is a single-polypeptide enzyme that is highly selective for

specific promoters that are rarely encountered in DNA unrelated to T7 DNA. Efficient termination signals are also rare, so that T7 RNA polymerase is able to make complete transcripts of almost any DNA that is placed under control of a T7 promoter. A very active enzyme, T7 RNA polymerase elongates chains about five times faster than does *Escherichia coli* RNA polymerase. These properties, together with the availability of the cloned gene have been exploited as the basis of expression systems in *E. coli* and other cell types, including mammalian and yeast cells.

Target genes are cloned in pET plasmids downstream from a strong T7 promoter and adjacent to efficient translational initiation signals from the phage. A host strain containing a chromosomal copy of the gene for T7 RNA polymerase under control of the IPTG-inducible *lac*UV5 promoter is used for protein production. The T7 polymerase is so selective and active that, after IPTG induction, almost all of the cell's resources are converted to target gene expression; after a few hours, the desired protein product can constitute more than 50% of the total cell protein. The host strain we use, BL21(DE3)/pLysS, also contains a plasmid which specifies low levels of T7 lysozyme, a natural inhibitor of T7 RNA polymerase. In uninduced cells, lysozyme reduces the basal activity of the T7 RNA polymerase and increases the range of target genes that can be stably maintained in the expression host. We used a pET vector, pET9 that has a *kan* gene, conferring resistance to the antibiotic kanamycin, as the selectable marker instead of the more commonly used pET vectors with a gene conferring resistance to ampicillin, *bla*. Ampicillin is not used during growth, so ampicilloyl-target protein conjugates cannot be formed. Such conjugates would complicate projected immunological studies [7]. The host *E. coli* strain, BL21, lacks the *ompT* outer membrane protease that can degrade proteins during purification [8], and it is deficient in the *lon* protease that can degrade recombinant proteins as they are being synthesized [9].

Polymerase Chain Reaction

Polymerase chain reaction (PCR) technology [10] has made cloning of genes into the pET vectors efficient and straightforward. Not only does PCR permit targeted amplification of specific DNA sequences starting with very small quantities of template, but also it provides an efficient means to fuse signals, such as new, unique restriction sites, to the 5' and 3' ends of a target sequence. Synthetic oligonucleotide primers are designed to be complementary to a limited portion, typically about 15 nucleotides, at each end of the target sequence and to include the desired restriction enzyme recognition sequences near their 5'- ends. During amplification the DNA sequences of the primers are incorporated into the product, thereby, allowing new sequence information in the 5' ends of the primers to be fused to the target sequence. After cutting with the appropriate restriction endonuclease, the resulting amplified DNA is ready for insertion directly into the pET expression plasmid. The natural ATG initiation codon in the vector is part of an *Nde*I site (CA!TATG), so coding sequences are easily inserted directly at the initiation codon by including an appropriately positioned *Nde*I site in the forward PCR primer. Other vectors in the pET set have an *Nco*I site (C!CATGG) at their initiation codon in case the target gene has an *Nde*I site(s) in its sequence. PCR-mediated cloning also can be used to split a gene into regions that code for individual domains and antigenic epitopes and these segments can be expressed individually to eliminate cross-reacting antigenic domains. It is important to note that PCR amplification can be used to eliminate endogenous *B. burgdorferi* promoter signals in front of a protein coding sequence. In all likelihood these prokaryotic promoters would be recognized by the host polymerase and the plasmids containing them would be unstable due to significant levels of basal transcription of the foreign gene by *E. coli* RNA polymerase. Too high a level of unregulated basal expression probably explains why Isaacs and Radolf [11] were unable to clone the endoflagellar sheath protein gene from a related spirochete, *Treponema pallidum*, in *E. coli* if they included the native promoter in front of the coding sequence.

RESULTS

We have used two sequence specific sets of oligonucleotide primers to amplify and clone related forms of the *ospA* gene under the control of transcription and translation signals from T7 [12]. One set allowed the entire ospA sequence to be cloned, while the other primed amplification of a truncated form of *ospA* lacking the first 17 codons specified by the wild-type structural gene, *i.e.* the residues believed to act as a signal sequence to direct association of OspA with the Borrelia membrane. On the basis of its similarity with *E. coli* signal

sequences it seemed possible that the full-length OspA protein would be lipidated in *E. coli* and become membrane associated. Typically such proteins have poor solubility properties and consequently detergents are required to solubilize them.

The recombinant version of the full-length *ospA* gene was expressed at a low level, a few % of the total protein, even when the coding sequence was placed under control of the strong transcription and translation signals from bacteriophage T7 in the pET vector. The low level of expression presumably is due to the accumulated toxic effects of OspA protein localizing at the *E. coli* cell membrane during expression and, as expected, the full-length protein remains associated with the insoluble *E. coli* membrane fraction unless solubilized by detergents. In contrast, the truncated form of the OspA protein lacking the 17 amino acid long signal sequence is expressed at very high levels, greater than 50% of the total protein, and it is highly soluble. We have developed an efficient procedure for purifying large amounts of recombinant OspA to homogeneity; more than 100 mg of protein can be obtained from 1 liter of induced cells. The recombinant OspA is highly soluble in aqueous solution ($\geq$ 50 mg/ml) and it elutes from a gel filtration column as a monomer under nondenaturing conditions. The molar extinction coefficient at 280 nm of purified OspA has been determined, simplifying quantitation in immunological testing.

The recombinant protein begins with two amino acids from the vector (Met Ala) fused in frame to Lysine18 of the OspA sequence. Amino-terminal sequencing of purified protein demonstrates that the first methionine residue is removed, therefore, the recombinant protein is referred to as OspA-257. Western blotting demonstrated that, although the OspA-257 protein is missing a lipidated N-terminus, it still reacts with antibodies present in the synovial fluid of a patient with Lyme-induced arthritis and is recognized by a variety of polyclonal and monoclonal antibodies raised against whole *B. burgdorferi* cells. Thus, it seems highly likely that the protein will be useful in immunochemical analysis for detection of Lyme disease.

We of course would like to know the three-dimensional structure of OspA since such information would assist in elucidation of antigenic sites and eventually may lead to development of a peptide-based vaccine. X-ray diffraction is still the only method for determining the exact three-dimensional structure of proteins in the size range of OspA and attempts to crystallize OspA are in progress. In lieu of a crystal structure, we are using other methods to characterize the structure of OspA such as circular dichroism (CD) spectroscopy. Unlike X-ray diffraction, CD can be applied to proteins in solution, under physiological conditions. Each type of secondary structure is characterized by a distinctive CD spectrum, and recent work by Johnson [13] showed that when CD spectra of proteins are extended below 200 nm to 178 nm, five types of secondary structure can be predicted with accuracy comparable to X-ray diffraction: α-helix, anti-parallel ß-sheet, parallel ß-sheet, ß-turns (all types) and aperiodic structures (including random coil).

We used the CD spectrometer at the National Synchrotron Light Source at BNL to measure the CD spectra of OspA to wavelengths as short as 175 nm; conventional CD spectrometers measure spectra reliably in the wavelength region $\geq$ 200 nm. The spectra were analyzed with a computer program supplied by C.W. Johnson, Jr. This algorithm uses matrix techniques (singular value decomposition) and statistical procedures (variable selection) to fit the CD spectra of the protein to a linear combination of orthogonal CD spectra derived from a library of reference proteins, whose secondary structures are known from X-ray diffraction. The results [14] show that OspA contains mostly ß-sheet (27% anti-parallel, 9% parallel) configurations, ß-turn (21%), and random coil (34%), but little α-helix (11%). These values are quite different from those obtained using purely predictive methods based solely on the amino acid sequence of OspA, which predict a much higher percent of α-helix than is present (Table 1).

Description of Method	Type of structure %		
	α-helix	ß-all types	others
UV-CD (our work)	11±1	57±2	34±1
PCGene	24 (27)*	22 (21)	54 (52)
Chou-Fasman	45 (40)	28 (34)	27 (26)
Garnier-Robson	54 (53)	17 (18)	29 (29)

* (predicted values for full-length, unprocessed OspA)

TABLE 1.

We have also used steady-state and time-resolved fluorescence to investigate the static and dynamic aspects of OspA conformation. OspA contains a single tryptophan residue at position 200 (residue 216 in the full-length protein). Tryptophan can be selectively excited at wavelengths ≥ 295 nm and its fluorescence emission is particularly sensitive to the local environment. The fluorescence signal provides a probe for changes in protein conformation since changes in protein conformation are expected to cause changes in the environment of tryptophan. The emission peak (330 nm) of OspA and its response to various ionic and nonionic quenchers indicate that Trp200 is buried within the native protein in a relatively hydrophobic environment. Trp200 is completely exposed to the solvent, as expected for a random coil, when OspA is denatured by high temperature (≥ 80°C) or guanidine.

Both CD and fluorescence measurements reveal that the native conformation of OspA is highly stable; no significant changes are seen from pH 3-11, and the protein is stable at low and high salt concentrations. The protein returns to its native state, as reflected by CD and fluorescence measurements, after being held at 100°C for up to 10 min. Interestingly, the OspA-257 protein is exceptionally resistant to digestion by trypsin and human plasmin, even though it is rich in lysine residues (16 mol %) which are distributed rather uniformly throughout the protein, but it is extremely susceptible to proteinase K.

The same strategy used to clone *ospA* was used to clone recombinant versions of *ospB* directly into a pET vector with similar results, *i.e.* the full-length lipidated protein is poorly expressed while its truncated counterpart is overproduced[1]. The truncated OspB protein is currently being characterized by CD and fluorescence; it also has a single tryptophan.

In a related spirochete, *B. hermsii*, antigenic variation is the consequence of sequential expression of genes for a set of variable outer surface proteins known as Vmps [15]. Recently, it has been demonstrated that the Vmp genes have many features in common with the *B. burgdorferi* Osp gene family including N-terminal sequences that are typical of lipoproteins and evidence has been presented that the Vmp proteins are recognized and processed as lipoproteins in *E. coli* [16]. As might be expected, full-length Vmp's are expressed at relatively low levels in *E. coli*[2] presumably because the protein is toxic when it becomes anchored in the membrane. Vmp overproduction was possible if the region of the gene encoding the hydrophobic signal sequence was excised before the recombinant gene was placed under control of T7 signals in a pET plasmid[3]. Moveover, the truncated Vmp protein is soluble and it can be readily purified to homogeneity without the use of detergents.

[1] Duray, P., Lade, B.N. and Dunn, J.J. (unpuplished).
[2] Barbour, A.G. (personal communication).
[3] Bundoc, V.G., Barbour, A.G., Lade, B.N. and Dunn, J.J. (unpublished)

SUMMARY

The T7 based pET expression system is a powerful and versatile system for cloning and expression of recombinant proteins in *E. coli*. PCR technology complements the pET system since it permits precise, sequence specific amplification of coding sequences which, when used in conjunction with appropriately designed primers allows amplification of DNA tailored for direct insertion into a pET vector. The net result is a more highly efficient method to rationally engineer and express open reading frames. In certain instances, the target protein may contain sequences which limit overproduction. If the sequence is a specific domain of the target protein then it may be possible to remove it without destroying the protein per say. A good example of such a domain is the class of hydrophobic leader signal sequences found at the beginning of Borrelia lipoproteins (OspA, OspB and the Vmp's). Elevated expression of the full-length proteins is toxic to the overproducing strains and they stop making the protein after a short period of time. In contrast, these same proteins without the hydrophobic leader sequences continue to be synthesized at high rates for hours following induction and they accumulate to very high levels. The ability to produce these recombinant proteins in large amounts and in a highly purified state without using detergents raises the possibility that they may be useful in immunoassays and as possible immunogens in vaccines. Such studies are currently in progress.

REFERENCES

1. Craft, J.E., Filcher, D.K., Shimamoto, G.T., and Steere, A.C. (1986). Antigens of Borrelia burgdorferi recognized during Lyme disease: appearance of a new immunoglobulin M response and expansion of the immunoglobulin G response late in the illness. J. Clin. Invest. **78**, 934-939.

2. Magnarelli, L.A., and Anderson, J.F. (1988). Enzyme-linked immunosorbent assays for the detection of class-specific immunoglobulins to *Borrelia burgdorferi*. Am. J. Epidemiol. **127**, 813-815.

3. Luft, B.J., and Dattwyler, R.J. (1989). Lyme Borreliosis. In Current Clinical Topics in Infectious Disease. Volume 10. J. S. Remington, M. N. Swartz (Editors), Raven Press, New York, NY.

4. Howe, T.R., LaQuier, F.W., and Barbour, A.G. (1986). Organization of genes encoding the outermembrane proteins of Lyme disease agent *Borrelia burgdorferi* within a single transcriptional unit. Infect. Immun. **54**, 207-211.

5. Bergström, S., Bundoc, V.G., and Barbour, A.G. (1989). Molecular Analysis of linear plasmid-encoded major surface proteins, OspA and OspB, of the Lyme disease spirochete *Borrelia burgdorferi*. Mol. Microbiol. **3**, 479-486.

6. Studier, F.W., Rosenberg, A.H., Dunn, J.J., and Dubendorff, J.W. (1990). Use of T7 RNA polymerase to direct expression of cloned genes. Meth. Enzymol. **185**, 60-89.

7. Yvon, M., Anglade, P., and Wal, J-M. (1990). Identification of the binding sites of benzyl penicilloyl, the allergenic metabolite of penicillin, on the serum albumin molecule. FEBS **263**, 237-240.

8. Grodberg, J. and Dunn, J.J. (1988). *ompT* encodes the *Escherichia coli* outer membrane protease that cleaves T7 RNA polymerase during purification. J. Bacteriol. **170**, 1245-1253.

9. Goff, S.A. and Goldberg, A.L. (1985). Production of abnormal proteins in E. coli stimulates transcription of *lon* and other heat shock genes. Cell **41**, 587-595.

10. Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B., and Erlich, H.A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science **239**, 487-491.

11. Isaacs, R.D., and Radolf, J.D. (1990). Expression in *Escherichia coli* of the 37-Kilodalton endoflagellar sheath protein of *Treponema pallidum* by use of the polymerase chain reaction and a T7 expression system. Infect. Immun. **58**, 2025-2034.

12. Dunn, J.J., Lade, B.A., and Barbour, A.G. (1990). Outer Surface protein A (OspA) from the Lyme Disease Spirochete, *Borrelia burgdorferi*: High Level Expression and Purification of a Soluble Recombinant Form of OspA. Protein Expression and Purification **1**, 159-168.

13. Johnson, W.C., Jr. (1990). Proteins: structure, function and genetics. **7**, 205-214.

14. France, L. L., Kieleczawa, J., Dunn, J. J., Hind, G., and Sutherland, J. C. Structural Analysis of an Outer Surface Protein from the Lyme Disease Spirochete, *Borrelia burgdorferi*, using Circular Dichroism and Fluorescence Spectroscopy. Biochem. Biophys. Acta. (in press).

15. Barbour, A.G., and Stonner, H.G. (1984). Antigenic variation of *Borrelia hermsii*. UCLA Symp Mol Biol New Ser **20**, 123-135.

16. Burman, N., Bergström, S., Restrepo, and Barbour, A.G. (1990). The variable antigens Vmp7 and Vmp21 of the relapsing fever bacterium *Borrelia hermsii* are structurally analogous to the VSG proteins of the African trypanosome. Mol. Microbiol. **4**, 1715-1726.