Gene synthesis, high-level expression and assignment of backbone ^{15}N and ^{13}C resonances of soybean leghemoglobin

Stefan Prytulla, H. Jane Dyson, Peter E. Wright*

Department of Molecular Biology and Skaggs Institute of Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA

Received 7 October 1996

Abstract A synthetic gene for apoleghemoglobin-a from soybean, optimized for expression in *Escherichia coli* has been designed and synthesized by a recursive polymerase chain reaction technique. The protein has been expressed with high efficiency and a purification protocol has been developed. The holoprotein is readily reconstituted by the addition of heme. ^{15}N- and ^{15}N,^{13}C-labeled samples were produced and backbone ^{15}N and ^{13}C assignments were determined by 2D and 3D NMR spectroscopy. Comparison of the chemical shifts of $^{13}C^{\alpha}$ and ^{13}CO with random coil shifts revealed a pattern of secondary structure which correlates well with the one previously derived from homonuclear NMR data and low-resolution X-ray crystallography.

Key words: Leghemoglobin; Gene synthesis; NMR assignment; Secondary structure

1. Introduction

Leghemoglobins are genetically heterogeneous monomeric heme proteins found in legume root nodules. They are involved in the symbiosis between the plant and nitrogen-fixing *Rhizobium* bacteria by providing oxygen for bacterial respiration at a partial pressure low enough to prevent oxidative damage to the nitrogenase complex [1–4]. Soybean leghemoglobin-a (Lb-a) consists of 143 amino acids [5,6]. Only a low-resolution X-ray structure (3.3 Å) of its nicotinate complex is available [7]. However, X-ray structures of Lb-II from lupin (57% sequence homology with Lb-a) with several heme ligands have been solved to higher resolution [8,9]. The proton NMR spectrum of the carbon monoxide complex of soybean Lb-a (LbCO) has been almost completely assigned by homonuclear 2D and 3D NMR spectroscopy [10], and the secondary structure inferred from these data agrees well with the X-ray results. These structures show a folding topology typical of other globins such as myoglobin and hemoglobin. Despite the similarity of their tertiary structures, the sequence homology of leghemoglobins and myoglobins is low [11]. However, they do share a pattern of hydrophobic and polar residues considered to be determinants of the globin fold [12]. The O_2 affinity and association constant for leghemoglobin is much higher than that for myoglobin [1]. This has been attributed to a more open and flexible heme pocket [7–9]. The observed conservation of structure and function together with the very low sequence homology also prompts the question as to whether the kinetic folding pathways are the same in leghemoglobin and myoglobin. While the folding pathway of apomyoglobin was established by Jennings and Wright [13], nothing is known so far about that of apoleghemoglobin. An important step towards the investigation of the folding pathway is the detection of slowly exchanging amide protons in leghemoglobin well distributed throughout the protein [14], thus providing probes for hydrogen–deuterium exchange experiments. More detailed investigations of the solution structure and dynamics, as well as of the ligand-binding mechanism and the folding pathway of leghemoglobin require ^{15}N- and ^{13}C-labeled samples for heteronuclear NMR spectroscopy. As an important step in this direction, the present paper describes the development of a highly efficient bacterial expression system for soybean apoleghemoglobin-a together with a simple purification scheme. ^{15}N and ^{13}C chemical shifts of most backbone amide nitrogens as well as of most C^{α}, C^{β} and CO carbons were determined by heteronuclear 2D and 3D NMR spectroscopy utilizing the previously determined ^{1}H assignments [10].

2. Materials and methods

2.1. Materials

The oligonucleotides were purchased from Keystone Laboratory, Menlo Park, Ca. They were purified by PAGE (PagePURE, long oligos) or by reverse-phase chromatography (CartridgePURE, primers). The vector pCRII and competent *Escherichia coli* cells INVαF′ (F′ *end*A1 *rec*A1 *hsd*R17 (r_k^-, m_k^+) *sup*E44 *thi*-1 *gyr*A96 *rel*A1 ϕ80*lac*ZΔM15Δ(*lac*ZYA-*arg*F)U169) were purchased from Invitrogen. The vector pET24a and competent *E. coli* cells XL1-Blue (*sup*E44 *hsd*R17 *rec*A1 *gyr*A46 *thi rel*A1 *lac*$^-$ F′[*proAB*$^+$ *lacI*^{q} *lacZ*ΔM15 Tn10(*tet*^{r})]) and BL21(DE3) (*hsdS gal* [λcIts857 *ind*1 Sam7 *nin*5 *lac*UV5-T7 gene1]) were purchased from Novagen. *Vent* DNA polymerase and the restriction enzymes *Nde*I and *Bam*HI were obtained from New England Biolabs.

2.2. Gene design and synthesis

A gene coding for soybean apoleghemoglobin-a was designed using the Intelligenetics software. Eight overlapping oligonucleotides were used to assemble the gene by a recursive PCR technique [15–18]. Gene assembly and amplification were carried out in two separate PCR reactions in an Ericomp Twin Block thermocycler. The reaction mixture of the recursive assembly PCR contained 4 pmol of each of the eight oligonucleotides, 2.5 U *Vent* DNA polymerase and 250 μM of each of the 4 dNTPs in a final volume of 100 μl PCR buffer (10 mM KCl, 10 mM $(NH_4)_2SO_4$, 20 mM Tris-HCl (pH 8.8), 2 mM $MgSO_4$, 0.1% Triton X-100). The reaction began with a hot start of 4 min at 80°C. Five PCR cycles were carried out with 1 min denaturation at 95°C, 1 min annealing at 59°C and 1 min chain extension at 72°C. The reaction was completed with 5 min incubation at 72°C. Ten microliters of the reaction mixture were used for the second PCR. The reaction mixture of the amplification PCR contained 100 pmol

*Corresponding author. Fax: (619) 784-9822.

Abbreviations: dNTP, deoxyribonucleotide; DSS, sodium 2,2-dimethyl-2-silapentane-5-sulfonate; IPTG, isopropyl thiogalactoside; Lb, leghemoglobin; PAGE, polyacrylamide gel electrophoresis; PCR, polymerase chain reaction; TFA, trifluoro acetic acid; X-gal, 5 bromo-4-chloro-3-indolyl-β-D-galactoside

PII S0014-5793(96)01278-1

of each of the two primers, 2.5 U *Vent* DNA polymerase and 250 μM of each of the 4 dNTPs in 100 μl PCR buffer. The thermocycle was: 1 min at 95°C, 1 min at 60°C, 1 min at 72°C. Twenty-five cycles were carried out, preceded by a hot start of 4 min at 80°C and followed by 7 min incubation at 72°C.

2.3. Addition of A overhangs

The PCR mixture was diluted with 300 μl water and extracted with a mixture of 400 μl of Tris-saturated phenol and 400 μl of chloroform. The aqueous phase was concentrated with a Millipore UF ultrafiltration cell (molecular weight cutoff: 30 kDa) and diluted to 100 μl with PCR buffer. One unit of *Taq* DNA polymerase and 250 μM dNTPs were added. The mixture was incubated at 72°C for 20 min and placed on ice. The enzyme was removed by phenol/chloroform extraction as described above and the aqueous phase concentrated and diluted with water to 800 μl.

2.4. Cloning

The synthesized gene was inserted into the pCRII vector using the TA Cloning Kit from Invitrogen. INVαF′ cells were transformed with the reaction mixture and spread on X-gal LB-agar-plates supplemented with 30 μg/ml kanamycin. Taking advantage of the blue/white screening feature, white colonies were picked after overnight incubation at 37°C and amplified in 15 ml of kanamycin LB. Plasmids were isolated using the QIAGEN Midiprep Kit. The inserts were sequenced with an Applied Biosystems 373A sequencer. The *leghemoglobin* gene was excised from the plasmid with *Nde*I and *Bam*HI. It was purified by agarose gel chromatography and cloned into linearized pET24a. The hybrid plasmid was amplified in *E. coli* XL1-Blue. It was isolated with the QIAGEN Midiprep Kit and used to transform *E. coli* BL21(DE3) cells.

2.5. Expression

LB medium (3 ml, supplemented with 30 μg/ml kanamycin) was inoculated with a single colony of transformed BL21(DE3) cells and incubated at 37°C for 6 h. M9 minimal medium (50 ml, supplemented with 30 μg/ml kanamycin) was inoculated with the LB culture (3 ml) and incubated at 37°C overnight. Twenty milliliters of the preculture were used to inoculate M9 minimal medium (1 l, supplemented with 30 μg/ml kanamycin). The cells were grown until $OD_{550} = 1$ (4–5 h) and induced with 1 mM IPTG. After another 5 h of incubation the cells were harvested by centrifugation (5500 rpm, 5 min).

2.6. Protein purification

The cells were resuspended in 50 ml of lysis buffer (1 mM EDTA, 50 mM Tris-HCl, pH 8). Lysozyme solution (1.5 ml, 10 mg/ml in lysis buffer) was added and the suspension incubated on ice for 30 min. DNAse I solution (400 μl, 1 mg/ml) and $MgCl_2$ (400 μl, 1 M) were added. The suspension was incubated for another 30 min and centrifuged (10 min, 12000×*g*/8500 rpm). The pellet was resuspended in lysis buffer and the suspension centrifuged again. The pellet was thoroughly resuspended in water and the suspension mixed with the same volume of acetonitrile containing 0.2% TFA. A solution of 0.1% TFA in water was added to yield a final ratio of acetonitrile/water of 40:60 (v/v). Insoluble material was removed by filtration through a 0.2 μm membrane filter (Millipore) and the solution was applied to a preparative reverse-phase (C18) HPLC column equilibrated with acetonitrile/water (40:60) containing 0.1% TFA. After the sample was adsorbed to the column, it was washed with acetonitrile/water (40:60) containing 0.1% TFA and eluted with a linear gradient of acetonitrile in water containing 0.1% TFA. The apoLb-containing fractions eluted at approximately 51% acetonitrile and were lyophilized.

2.7. Formation of carbonmonoxy leghemoglobin

Recombinant apoLb (8 mg, 0.5 μmol) was dissolved in 75 ml of Tris-HCl buffer (10 mM, pH = 8) and mixed with heme dicyanide (0.4 mg, 0.625 μmol) in 25 ml of water and concentrated to 0.3 ml by ultrafiltration (Amicon YM10 membranes). Holo-leghemoglobin was transformed to the carbonmonoxide complex by a modification of a method described previously [19]. All buffers and solutions were equilibrated with carbon monoxide and experiments were carried out under carbon monoxide. Twenty microliters of sodium dithionite solution (20 mg/ml) were added to 0.3 ml of the leghemoglobin solution. The solution was applied to a Sephadex G-10 column, equilibrated with 10 mM potassium phosphate (pH* 7.0) in D_2O, and eluted directly into a carbon monoxide-flushed NMR tube.

2.8. NMR spectroscopy

NMR spectra were recorded at 308 K on a Bruker AMX-500 NMR spectrometer. Proton chemical shifts were referenced to external DSS (0 ppm). ^{15}N and ^{13}C chemical shift were referenced indirectly to liquid NH_3 and DSS according to [20]. 2D $^1H,^{15}N$-HSQC [21,22] and 3D TOCSY-HSQC [21–23] spectra were recorded to assign the amide nitrogens. Spectral widths of the $^1H,^{15}N$-HSQC spectrum were 16.1 ppm in F_2 (^1H) and 40.1 ppm in F_1 (^{15}N). Two hundred and fifty-six complex data points were taken in t_1 and t_2. Zero-filling yielded a final matrix of 2048×1024 real data points. The TOCSY-HSQC spectrum was recorded with spectral widths of 8.3 ppm Hz in both ^1H dimensions (F_1 and F_3) and 41.1 ppm in F_2 (^{15}N). Ninety-six, 32 and 512 complex data points were taken in t_1, t_2 and t_3. Zero-filling yielded a final matrix of 128×128×512 real points in F_1, F_2 and F_3. A constant-time $^1H,^{13}C$-HSQC [24] spectrum and four different triple-resonance experiments (constant-time HNCA [25,26], CBCA(CO)NH [27], constant-time HNCO [25,26] and (HCA)CO(CA)NH [28]) were carried out for sequential assignment of the backbone atoms and to obtain ^{13}C chemical shifts. Spectral widths of the $^1H,^{13}C$-HSQC spectrum were 10.1 ppm and 79.5 ppm in F_2 (^1H) and F_1 (^{13}C). Two hundred and fifty-six complex data points were taken in both dimensions. Zero-filling yielded a final matrix of 1024×2048 real data points. All triple resonance spectra were recorded with spectral widths of 14.3 ppm in F_3 (^1H) and 39.5 ppm in F_2 (^{15}N). Spectral widths in F_1 (^{13}C) were 49.7 ppm for HNCA, 79.5 ppm for CBCA(CO)NH, 15.9 ppm for HNCO and (HCA)CO(CA)NH. In t_1, t_2 and t_3, 32, 32 and 1024 complex data points were recorded for all experiments except for HNCA with 44 complex points in t_1. All triple resonance spectra were zero-filled to yield final matrices of 128×128×1024 real points in F_1, F_2 and F_3. Data were processed with the Felix 95 software. Shifted sine-bells were used as window functions in all dimensions of all recorded spectra.

3. Results and discussion

3.1. Gene synthesis

The *Lb-a* gene from soybean has three intervening sequences and uses several codons rarely found in highly expressed *E. coli* proteins [6]. Since the main purpose of the expression is the preparation of relatively large amounts of isotopically labeled protein samples for NMR spectroscopy, good protein yields are required. The previous successful syntheses of genes in this laboratory coding for rusticyanin [18], stellacyanin (M. Lubienski, unpublished data) and the DNA-binding domain of human estrogen-related receptor (D. Casimiro, unpublished data) prompted us to design an optimized gene for the expression. In the current study a gene coding for Lb-a was designed following the same considerations as described in [18]. Only a subset of all possible 61 codons is actually used by *E. coli* for highly expressed genes [29]. Codon usage correlates with the abundance of the corresponding tRNAs [30]. Codons rarely found in highly expressed *E. coli* genes were avoided in this study because they are assumed to cause poor expression of the recombinant protein in *E. coli* due to low tRNA levels. Only the two most common codons for each amino acid were selected. Further requirements for a high expression level that were met by the designed gene are the absence of palindromic sequences which can form hairpin loops and a low overall G+C content (49.5%). Long sequence repeats were also not permitted to prevent mispriming events during the PCR reactions. The optimized sequence of the synthetic *Lb-a* gene is shown in Fig. 1. A total of eight oligonucleotides representing four fragments of the coding and four of the non-coding strand were chemically synthesized. Their lengths varied be-

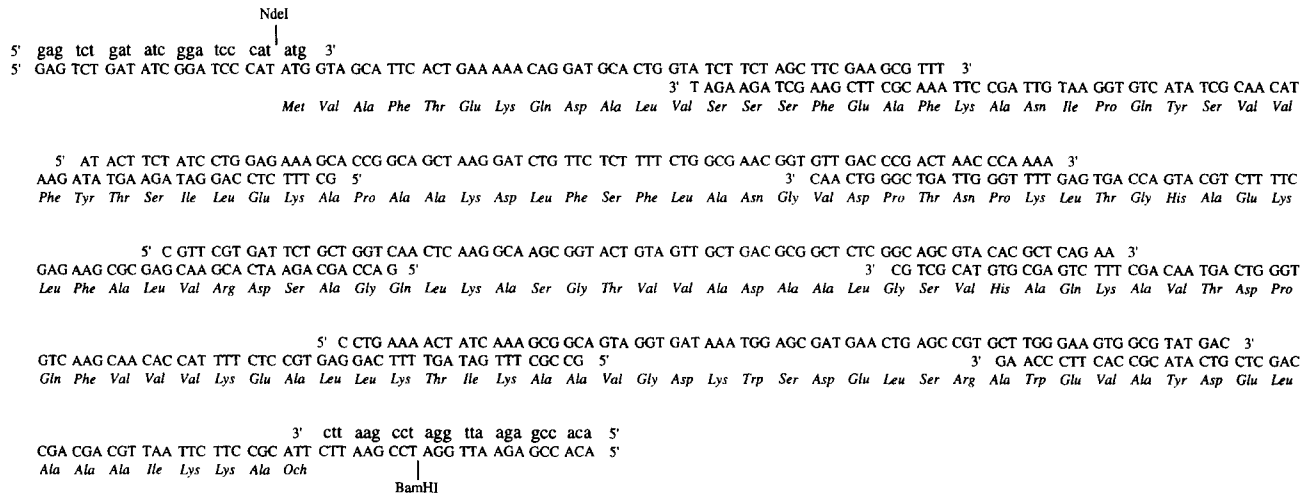

Fig. 1. Sequences of the eight overlapping oligonucleotides used to assemble the *Lb-a* gene. The four oligonucleotides in the upper row correspond to the coding strand, the ones in the lower row to the non-coding strand. The amino acid sequence is shown as well. The short sequences at the termini correspond to the amplification primers used in the second PCR. The *Nde*I and *Bam*HI restriction sites in the flanking cloning regions are also shown.

tween 73 and 83 nucleotides. Neighboring oligonucleotides overlapped by between 19 and 22 residues, resulting in calculated melting temperatures of 62°C for all priming sites. Both ends were flanked by restriction sites for *Nde*I and *Bam*HI (5′ and 3′ end of the coding strand, respectively) for the insertion in the desired linearized expression vector.

The full-length gene was assembled from the overlapping oligonucleotides by a recursive PCR method [15–17]. Gene assembly and amplification were carried out in two different PCR reactions as suggested in [18]. Only five and 25 PCR cycles were used for the assembly and amplification reactions. A low number of cycles is desirable, since PCR inherently incorporates mistakes. The final product shows a single band with the expected molecular weight on an agarose gel (Fig. 2a, lane 2). Gene assembly and amplification can also be carried out as a one-pot reaction, when the oligonucleotides are applied in 1/10 of their concentration in the separate assembly PCR, but the yield was significantly lower (lane 1). In both PCR reactions, *Vent* DNA polymerase was used instead of *Taq* DNA polymerase because of its higher fidelity of nucleotide incorporation due to exonuclease ('proof-reading') activity [31]. Due to the exonuclease activity, *Vent* DNA polymerase does not add an additional adenine at the 3′ end as does *Taq* DNA polymerase [32]. Since this was required in the following cloning procedure 'TA-cloning'), overhanging adenines were added in a separate incubation step in the presence of *Taq* DNA polymerase. The effort of an additional incubation step is redeemed by avoiding the otherwise necessary steps to assemble the gene from partially correct sequences [18] and by the fact that a far smaller number of clones needs

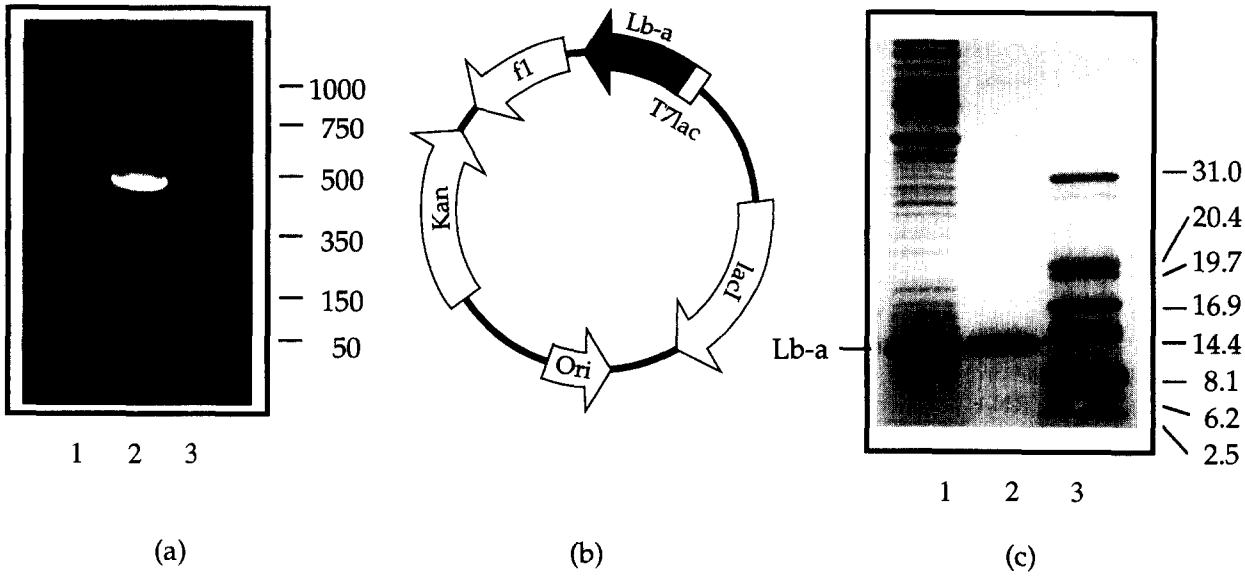

Fig. 2. a: Agarose-gel analysis of the PCR product (lane 2). Synthetic (PCR) DNA fragments (Promega) serve as molecular weight markers (lane 3). In lane 1, the product of the reaction carried out as a single PCR is shown. b: Map of the pET24a/Lb-a expression vector. The plasmid contains a kanamycin resistance gene for plasmid selection. The *Lb-a* gene is placed behind the T7lac promoter sequence to control expression. c: SDS-PAGE analysis of Lb-a expression and purification. Lane 1: Total lysate of *E. coli*. Lane 2: Lb-a-containing fraction of the reverse-phase HPLC. Lane 3: Molecular weight marker (Promega).

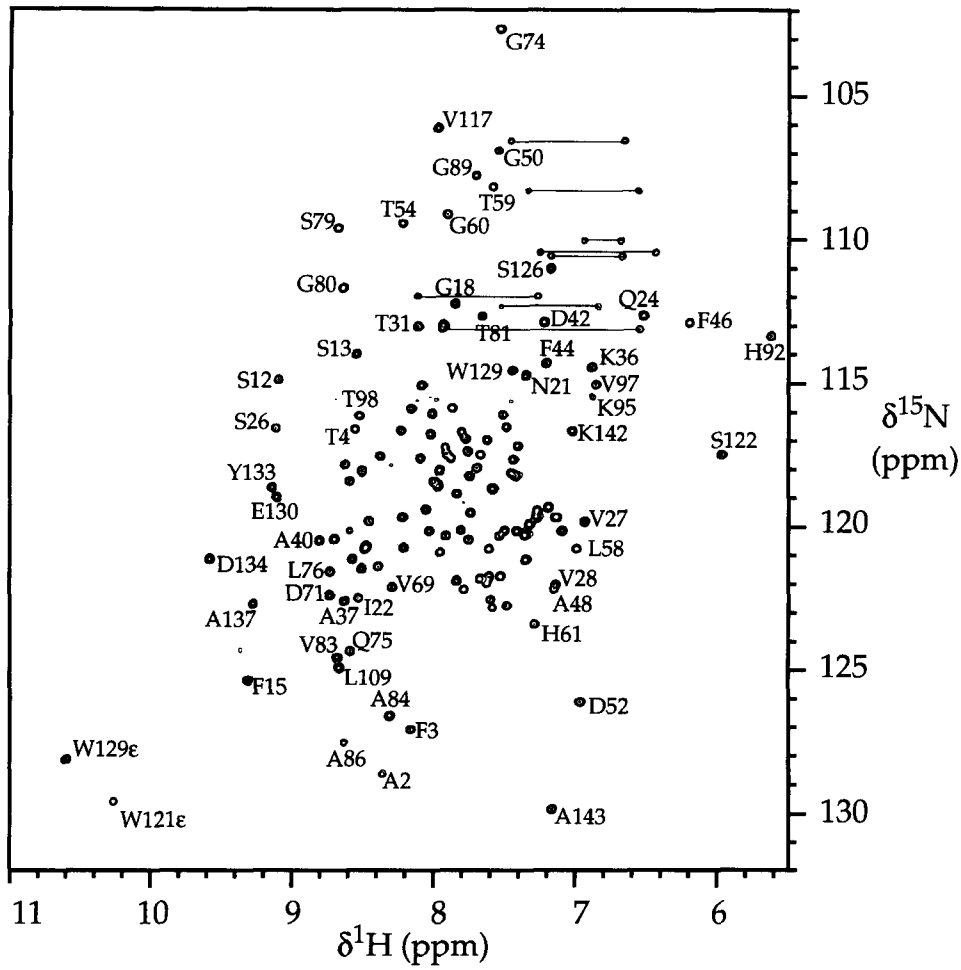

Fig. 3. $^{1}H,^{15}N$-HSQC spectrum of ^{15}N-labeled LbCO. The concentration of the sample was 1 mM in 10 mM sodium phosphate buffer at pH 7.0. The spectrum was recorded at 308 K. The pairs of peaks connected with horizontal lines arise from glutamine and asparagine side-chain amide groups.

to be sequenced. A further simplification of the described cloning procedure would be the replacement of the TA-cloning technique by an efficient blunt-end cloning method without the need for A overhangs.

The PCR product was inserted into the linearized pCRII vector and the hybrid plasmid was used to transform *E. coli* INVαF′. Plasmids were isolated and the inserts were sequenced. Three out of 10 clones had the correct nucleotide sequence over the entire length of the synthetic gene. The gene was excised from the plasmid with *Nde*I and *Bam*HI and cloned into the pET24a vector. The plasmid was amplified in *E. coli* XLI Blue and subsequently transformed into lysogenic *E. coli* BL21(DE3). In pET24a the *Lb-a* gene is under control of the strong T7lac promoter (Fig. 2b). Transcription of the gene requires T7 RNA polymerase which is not encoded on the plasmid. Instead, the expression host BL21(DE3) carries this gene in its chromosomal DNA where it is controlled by the lacUV5 promoter. Both lac promoters are blocked by lac repressors in the uninduced state ensuring a low level of background expression [33]. Both genes are induced by IPTG yielding a protein with the expected apparent molecular weight (Fig. 2c, lane 1). The cloning host XL1 Blue does not carry the T7 RNA polymerase gene and therefore cannot express the *Lb-a* gene. pET24a also has a kanamycin resistance gene for selection of correctly transformed bacteria.

3.2. Protein purification and reconstitution

Recombinant apoLb forms inclusion bodies in the cytoplasm. The inclusion bodies are solubilized by acetonitrile/water (40:60, v/v) containing 0.1% TFA and fractionated with reverse-phase HPLC. This procedure yields a pure protein (Fig. 2c, lane 2) which can be redissolved in water and a variety of buffers following lyophilization. The molecular mass of the recombinant apoprotein (15374 Da) was confirmed by electrospray ionization mass spectroscopy. The protein retains the initiator methionine residue. Approximately 40 mg were isolated from 1 l of bacteria culture. It readily binds heme or heme dicyanide to form the holoprotein. For characterization by spectroscopic methods, the bound heme was reduced to Fe(II) and transformed to the stable carbon monoxide complex. Electronic absorption spectra of native soybean LbCO [34] and recombinant LbCO are indistinguishable, indicating the correct coordination of the heme prosthetic group.

3.3. NMR spectroscopy

The $^{1}H,^{15}N$-HSQC spectrum of ^{15}N-labeled Lb-a is shown in Fig. 3. The triple resonance spectra of the doubly labeled protein provide two more independent data sets for sequential assignment besides the previously available sequential assignments based on inter proton NOEs [10]. The constant-time

Table 1
Backbone ^{15}N and ^{13}C resonance assignments[a] for carbonmonoxy Lb-a in 10 mM Na-phosphate (pH 7.0) at 308 K

Residue	N	CO	C^{α}	C^{β}
Met[0]				
Val[1]	174.4	61.6	33.7	
Ala[2]	128.6	175.9	52.1	18.5
Phe[3]	127.1	175.7	58.0	43.5
Thr[4]	116.6	175.1	61.1	72.3
Glu[5]	118.9	178.5	59.6	
Lys[6]		178.5	58.7	31.9
Gln[7]	118.2	177.6	60.0	27.8
Asp[8]	118.6	177.5	58.7	43.2
Ala[9]	120.7	181.3	54.9	17.5
Leu[10]	120.2	180.6	57.7	41.7
Val[11]	122.2	176.9	66.4	30.6
Ser[12]	114.9	178.0	62.9	63.1
Ser[13]	114.0	178.0	61.5	62.8
Ser[14]	117.4	176.7	61.1	64.0
Phe[15]	125.4	177.4	60.9	38.9
Glu[16]	117.6	179.0	59.8	28.6
Ala[17]	121.2	180.3	55.2	17.3
Phe[18]	120.5	175.5	59.7	38.0
Lys[19]	116.0	176.6	58.1	31.7
Ala[20]	119.3	178.1	53.4	18.6
Asn[21]	114.8	171.9	51.6	36.8
Ile[22]	122.5	175.9	66.7	
Pro[23]		177.3	67.3	30.9
Gln[24]	112.7	179.3	58.2	28.9
Thr[25]	117.4	178.3	59.6	36.7
Ser[26]	116.6	175.3	63.4	[62.2]
Val[27]	119.8	179.4	66.7	31.5
Val[28]	122.0	177.8	66.5	31.4
Phe[29]	121.8	176.5	61.1	38.6
Thr[30]	113.1	178.5	62.6	38.7
Thr[31]	113.0	176.5	67.4	68.5
Ser[32]	116.1	176.7	62.7	62.6
Ile[33]	120.1	176.6	65.8	37.9
Leu[34]	116.9	178.7	57.2	41.3
Glu[35]	116.7	178.1	58.9	29.9
Lys[36]	114.4	176.5	56.0	34.5
Ala[37]	122.6	175.7	49.7	
Pro[38]	179.2	65.4	31.2	
Ala[39]	117.9	179.4	53.8	18.0
Ala[40]	120.5	178.3	54.6	17.0
Lys[41]	117.7	175.3	59.5	32.1
Asp[42]	112.9	176.8	54.9	41.6
Leu[43]	118.7	176.9	56.5	44.2
Phe[44]	114.3	176.9	56.4	[40.3]
Ser[45]	[122.7]	176.4	62.0	[62.0]
Phe[46]	112.9	175.2	56.0	37.6
Leu[47]	119.7	177.2	53.8	41.4
Ala[48]	122.2	178.5	55.9	18.6
Asn[49]	112.8	175.1	51.9	39.1
Gly[50]	106.9	172.9	43.9	–
Val[51]	119.0	173.9	62.8	30.5
Asp[52]	126.1	174.6	49.8	
Pro[53]		176.0	63.9	31.6
Thr[54]	109.4	174.2	62.1	68.8
Asn[55]	122.8	174.5	50.2	
Pro[53]		179.9	63.9	31.6
Lys[57]	118.4	177.7	58.1	31.9
Leu[58]	120.8	177.3	57.9	42.3
Thr[59]	108.2	178.1	64.9	67.5
Gly[60]	109.1	176.0	46.8	
His[61]	123.4	176.1	61.6	30.2
Ala[62]	122.0	179.5	55.9	17.3
Glu[63]	115.1	177.6	61.0	29.6
Lys[64]	117.0	179.1	59.7	33.1
Leu[65]	120.7	178.8	61.1	41.2
Phe[66]	117.8	178.2	63.3	39.0
Ala[67]	120.4	181.2	55.3	18.0
Leu[68]	121.5	180.4	57.9	42.4

Table 1
(continued)

Residue	N	CO	C^{α}	C^{β}
Val[69]	122.1	177.6	67.0	31.2
Arg[70]	121.4	181.0	60.8	27.1
Asp[71]	122.4	177.6	57.9	41.7
Ser[72]	116.8	175.7	60.9	[62.8]
Ala[73]	119.6	177.8	54.9	18.1
Gly[74]	102.6	178.0	47.4	–
Gln[75]	124.3	179.5	60.2	29.3
Leu[76]	121.6	180.5	57.8	41.0
Lys[77]	120.8	177.6	59.2	31.8
Ala[78]	118.2	179.6	54.6	19.3
Ser[79]	109.6	175.3	59.1	67.3
Gly[80]	111.7	172.6	46.6	–
Thr[81]	112.7	170.3	60.0	71.3
Val[82]	118.9	174.1	60.6	34.7
Val[83]	124.6	174.9	60.6	34.1
Ala[84]	126.6	175.5	49.8	21.1
Asp[85]	121.1	177.0	54.3	42.4
Ala[86]	127.5	180.2	55.1	18.1
Ala[87]	120.1	180.0	54.8	17.8
Leu[88]	118.1	180.7	56.5	41.5
Gly[89]	107.8	175.9	47.4	–
Ser[90]	117.2	176.7	61.2	[62.8]
Val[91]	119.4	176.1	65.2	31.2
His[92]	113.4	175.5	61.4	25.8
Ala[93]	122.5	181.6	54.5	17.5
Gln[94]	117.6	177.8	58.3	28.0
Lys[95]	115.5	174.4	54.7	31.8
Ala[96]	120.8	176.1	52.3	16.6
Val[97]	115.0	177.4	63.2	31.4
Thr[98]	116.1	175.6	59.9	[72.2]
Asp[99]	[120.4]	[177.1]		
Pro[100]		180.4	66.3	31.2
Gln[101]	115.9	179.0	59.8	28.6
Phe[102]	118.4	180.9	63.6	39.1
Val[103]	120.9	177.0	66.9	31.4
Val[104]	120.3	178.6	66.7	31.7
Val[105]	119.4	176.3	67.3	30.4
Lys[106]	121.8	177.7	60.3	31.2
Glu[107]	116.6	179.5	59.5	29.0
Ala[108]	121.9	180.0	55.1	19.7
Leu[109]	124.9	178.5	59.8	41.4
Leu[110]	119.8	178.9	58.3	41.0
Lys[111]	118.0	178.9	59.7	32.6
Thr[112]	118.0	175.4	67.7	69.2
Ile[113]	120.3	177.6	61.6	34.3
Lys[114]	119.5	178.6	60.5	31.4
Ala[115]	120.3	179.0	54.5	17.3
Ala[116]	118.7	179.1	54.1	17.0
Val[117]	106.1	178.6	61.2	32.4
Gly[118]	112.3	176.3	47.7	–
Asp[119]		176.8	55.8	39.9
Lys[120]	117.5	176.0	56.0	31.6
Trp[121]	120.1	174.4	59.0	29.9
Ser[122]	117.5	173.0	56.3	
Asp[123]		[56.6]	[40.3]	
Glu[124]	[119.2]	178.8	59.4	29.3
Leu[125]	121.7	177.3	58.4	41.6
Ser[126]	111.0	177.1	61.4	62.5
Arg[127]	115.9	177.6	58.3	29.7
Ala[128]	120.1	178.7	55.7	18.1
Trp[129]	114.6	178.5	61.6	29.9
Glu[130]	119.0	177.7	61.1	30.0
Val[131]	117.5	177.7	66.5	32.5
Ala[132]	120.2	180.5	55.2	17.4
Tyr[133]	118.6	176.7	63.1	38.7
Asp[134]	121.1	181.0	57.8	39.7
Glu[135]	118.1	179.0	58.4	29.1
Leu[136]	122.7	177.7	57.4	40.7
Ala[137]	122.7	179.0	55.4	18.1

Table 1
(continued)

Residue	N	CO	C^{α}	C^{β}
Ala^{138}	116.5	179.7	54.7	17.5
Ala^{139}	119.9	180.4	54.5	18.1
Ile^{140}	119.7	177.3	65.3	37.4
Lys^{141}	117.2	177.0	58.5	32.5
Lys^{142}	116.6	175.0	56.1	32.9
Ala^{143}	129.9	181.9	54.0	

[a]Assignments in brackets are tentative.

HNCA [25,26] and CBCA(CO)NH [27] spectra connect pairs of backbone H^N and N via the intervening C^{α}. Additionally, the constant-time HNCO [25,26] and (HCA)CO(CA)NH [28] spectra provide connectivities of the amide groups via the intervening CO. Assignments were confirmed by comparing the intraresidue connectivities observed in the 3D TOCSY-HSQC spectrum with the corresponding homonuclear data [10]. ^1H chemical shifts of most detected protons of the recombinant protein are identical to those of the natural protein, providing strong evidence for the identity of their tertiary structures. The determined backbone ^{15}N and ^{13}C chemical shifts are summarized in Table 1.

Three corrections to the original ^1H assignments [10] had to be made. The resonances previously assigned to Asp^{99} have been reassigned to Gln^{101} because the triple-resonance experiments clearly establish connectivity to Phe^{102}. At the proton chemical shifts found for Gln^{101} by homonuclear methods (5.15 ppm), no matching peaks were found in the $^1H,^{15}N$-HSQC and triple resonance spectra. Instead, a peak at 5.15 ppm was found in the $^1H,^{13}C$-HSQC [23] spectrum. Because this resonance shows NOE connectivities to Phe^{102} it is most likely the H^{α} of Asp^{99}, but no connectivities to an H^N could be established leaving the assignment tentative. Finally, the backbone resonances of Phe^{46} were reassigned on the basis of clear sequential connectivities to Leu^{47} in the triple resonance experiments. Reinspection of the previously assigned NOESY spectrum [10] confirmed these corrections. Thr^{54} and Asn^{55} are flanked by two prolines (53 and 56) and could not be assigned previously due to the lack of sequential NOEs; both were assigned by connectivities in the triple resonance experiments. No $^1H,^{15}N$-HSQC peaks were found for Val^1, Lys^6, Asp^{119} and Asp^{123}, most likely because of rapid amide proton exchange due to their location in exposed loops.

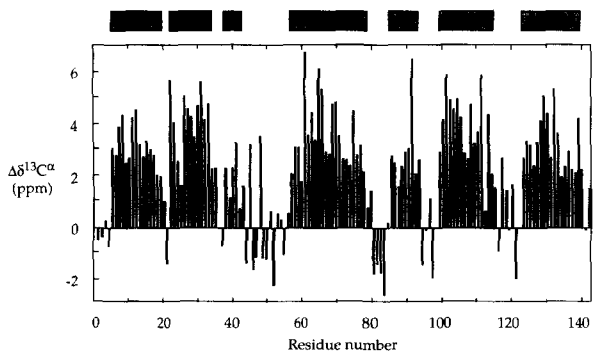

Fig. 4. Secondary $^{13}C^{\alpha}$ chemical shifts of LbCO. The seven helical regions are indicated.

3.4. Secondary structure

The chemical shifts of backbone $^{13}C^{\alpha}$ and ^{13}CO contain information about secondary structure elements. Downfield shifts of their resonance frequencies compared to random coil values [35,36] strongly indicate the presence of α-helix while β-sheets yield upfield shifts. The $^{13}C^{\alpha}$ secondary shifts of the carbonmonoxide complex of leghemoglobin reveal the presence of seven α-helical regions (Fig. 4). Residues 5–20 (helix A), 22–35 (helix B), 38–43 (helix C), 56–80 (helix E), 86–94 (helix F), 100–118 (helix G) and 124–141 (helix H). This is in excellent agreement with the crystal structure [7] and the data derived from sequential NOEs [10]. No $^{13}C^{\alpha}$ chemical shifts are available for Asp^{99} and Asp^{123}, leaving the starting points of helices G and H ambiguous. The ^{13}CO secondary chemical shifts (not shown) are also generally consistent with the helix locations determined by other methods but are a poorer diagnostic than the $^{13}C^{\alpha}$ chemical shifts.

4. Conclusion

The successful design and synthesis of the *Lb-a* gene and the overexpression of the apoprotein in *E. coli* confirm the usefulness of this approach for producing labeled proteins of interest for NMR studies. The NMR and visible spectra conclusively show that the structure of recombinant reconstituted leghemoglobin-a is correct. The availability of a high-level bacterial expression system for leghemoglobin allows the production of large amounts of isotopically labeled protein for heteronuclear NMR spectroscopy and will facilitate mutagenesis for structure, function and folding studies.

Acknowledgements: We thank Dr. D. Casimiro for valuable discussions and assistance with the gene design and synthesis. This work was supported by Grants DK 34909 (P.E.W.) and GM 48374 (H.J.D.) from the National Institutes of Health. S.P. was supported by a fellowship from the Deutsche Forschungsgemeinschaft.

References

[1] Gibson, Q.H., Wittenberg, J.B., Wittenberg, B.A., Bogusz, D. and Appleby, C.A. (1989) J. Biol. Chem. 264, 100–107.
[2] Appleby, C.A. (1974) in The Biology of Nitrogen Fixation (Quispel, A., ed.) pp. 521–554, North Holland Publ., Amsterdam.
[3] Appleby, C.A. (1984) Annu. Rev. Plant Physiol. 35, 443–478.
[4] Appleby, C.A. (1992) Sci. Prog. 76, 365–398.
[5] Sievers, S.G., Huhtala, M.-L. and Ellfolk, N. (1978) Acta Chem. Scand. B32, 380–386.
[6] Hyldig-Nielsen, J.J., Jensen, E.Ø., Paludan, K., Wiborg, O., Garret, R., Jørgensen, P. and Marcker, K.A. (1982) Nucl. Acids Res. 2, 689–701.
[7] Ollis, D.A., Appleby, C.A., Colman, P.M., Cutten, A.E., Guss, J.M., Ventakappa, M.P. and Freeman, H.C. (1983) Aust. J. Chem. 36, 451–468.
[8] Arutyunyan, E.G., Kuranova, I.P., Vainshtein, B.K. and Steigemann, W. (1980) Sov. Phys. Crystallog. 25, 43–58.
[9] Harutyunyan, E.H., Safonova, T.N., Kuranova, I.P., Popov, A.N., Teplyakov, A.V., Obmolova, G.V., Rusakov, A.A., Vainshtein, B.K., Dodson, G.G., Wilson, J.C. and Perutz, M.F. (1995) J. Mol. Biol. 251, 104–115.
[10] Morikis, D., Lepre, C.A. and Wright, P.E. (1994) Eur. J. Biochem. 219, 611–626.
[11] Lesk, A.M. and Chothia, C. (1980) J. Mol. Biol. 136, 225–270.
[12] Bashford, D., Lesk, A.M. and Chothia, C. (1987) J. Mol. Biol. 196, 199–216.
[13] Jennings, P.A. and Wright, P.E. (1993) Science 262, 892–896.
[14] Morikis, D. and Wright, P.E. (1996) Eur. J. Biochem. 237, 212–220.
[15] Dillon, P.J. and Rosen, C.A. (1990) BioTechniques 9, 298–300.

[16] Ye, Qi-Zhuang, Johnson, L.L. and Baragi, V. (1992) Biochem. Biophys. Res. Commun. 186, 143–149.
[17] Prodromou, C. and Pearl, L.H. (1992) Protein Eng. 5, 827–829.
[18] Casimiro, D.R., Toy-Palmer, A., Blake, R.C. II and Dyson, H.J. (1995) Biochemistry 34, 6640–6648.
[19] Mabbutt, B.C. and Wright, P.E. (1983) Biochim. Biophys. Acta 744, 281–290.
[20] Wishart, D.S., Bigam, C.G., Holm, A., Hodges, R.S. and Sykes, B.D. (1995) J. Biomol. NMR 5, 67–81.
[21] Bodenhausen, G. and Ruben, D.J. (1980) J. Chem. Phys. Lett. 69, 546.
[22] Bax, A., Ikura, M., Kay, L.E., Torchia, D.A. and Tschudin, R. (1990) J. Magn. Reson. 86, 304–318.
[23] Marion, D., Driscoll, P.C., Kay, L.E., Wingfield, P.T., Bax, A., Gronenborn, A.M. and Clore, G.M. (1989) Biochemistry 28, 6150–6156.
[24] Vuister, G.W. and Bax, A. (1992) J. Magn. Reson. 98, 428–435.
[25] Kay, L.E., Ikura, M., Tschudin, R. and Bax, A. (1990) J. Magn. Reson. 89, 496–514.
[26] Grzesiek, S. and Bax, A. (1992) J. Magn. Reson. 96, 432–440.
[27] Grzesiek, S. and Bax, A. (1992) J. Am. Chem. Soc. 114, 6291–6293.
[28] Löhr, F. and Rüterjans, H. (1995) J. Biomol. NMR 6, 189–197.
[29] Gribskov, M., Devereux, J. and Burgess, R. (1984) Nucl. Acids Res. 12, 539–549.
[30] Ikemura, T. (1981) J. Mol. Biol. 146, 1–21.
[31] Kong, H., Kucera, R.B. and Jack, W.E. (1993) J. Biol. Chem. 268, 1965–1975.
[32] Clark, J.M. (1988) Nucl. Acids Res. 16, 9677–9686.
[33] Dubendorff, J.W. and Studier, F.W. (1991) J. Mol. Biol. 219, 45–59.
[34] Sternberg, H. and Virtanen, A.I. (1952) Acta Chem. Scand. 6, 1342–1352.
[35] Wishart, D.S., Sykes, B.D. and Richards, F.M. (1991) J. Mol. Biol. 222, 311–333.
[36] Wishart, D.S., Bigam, C.G., Yao, J., Abildgaard, F., Dyson, H.J., Oldfield, E. Markley, J.L. and Sykes, B.D. (1995) J. Biomol. NMR 6, 135–140.