Berjot, M., Marx, J., & Alix, A. J. P. (1987) *J. Raman Spectrosc. 18*, 289–300.
Bussian, B. M., & Sander, C. (1989) *Biochemistry 28*, 4271–4277.
Byler, D. M., & Susi, H. (1986) *Biopolymers 25*, 469–487.
Chang, T. C., Wu, C. S. C., & Yang, J. T. (1978) *Anal. Biochem. 91*, 13–31.
Chen, M. C., & Lord, R. C. (1974) *J. Am. Chem. Soc. 96*, 4750–4752.
Chirgadze, Y. N., Fedorov, O. V., & Trushina, N. P. (1975) *Biopolymers 14*, 679–694.
Dong, A., Huang, P., & Caughey, W. S. (1990) *Biochemistry 29*, 3303–3308.
Dousseau, F., Therrien, M., & Pézolet, M. (1989) *Appl. Spectrosc. 43*, 538–542.
Eckert, K., Grosse, R., Malur, J., & Repke, K. R. H. (1977) *Biopolymers 16*, 2549–2563.
Elliot, A., & Ambrose, E. J. (1950) *Nature 165*, 921–922.
Haaland, D. M., & Thomas, E. V. (1988) *Anal. Chem. 60*, 1193–1202.
Kabsch, W., & Sander, C. (1983) *Biopolymers 22*, 2577–2637.
Krimm, S., & Bandekar, J. (1986) *Adv. Protein Chem. 38*. 181–364.
Lafleur, M., Pigeon, M., Pézolet, M., & Caillé, J.-P. (1989) *J. Chem. Phys. 93*, 1522–1526.
Lee, D. C., Haris, P. I., Chapman, D., & Mitchell, R. C. (1989) in *Spectroscopy of Biological Molecules: State of the Art* (Bertoluzza, A., Fagnano, C., & Monti, P., Eds.) pp 57–58, Società Editrice Escullapio, Bologna, Italy.
Levitt, M., & Greer, J. (1977) *J. Mol. Biol. 114*, 181–293.
Lindberg, W., Persson, J.-A., & Wold, S. (1983) *Anal. Chem. 554*, 643–648.
Lippert, J. L., Tyminsky, D., & Desmeules, P. J. (1976) *J. Am. Chem. Soc. 98*, 7075–7080.
Lord, R. C. (1971) *Pure Appl. Chem. Suppl. 7*, 179–191.
Miyazawa, T. (1958) *J. Chem. Phys. 29*, 246–248.
Miyazawa, T. (1960) *J. Chem. Phys. 32*, 1647–1652.
Miyazawa, T., & Blout, E. R. (1961) *J. Am. Chem. Soc. 83*, 712–719.
Pézolet, M., Pigeon-Gosselin, M., & Coulombe, L. (1976) *Biochim. Biophys. Acta 453*, 502–512.
Pézolet, M., Boulé, B., & Bourque, D. (1983) *Rev. Sci. Instrum. 54*, 1364–1367.
Powell, J. R., Wasacz, F. M., & Jakobsen, R. J. (1986) *Appl. Spectrosc. 40*, 339–344.
Provencher, S. W., & Glöckner, J. (1981) *Biochemisry 20*, 33–37.
Rey-Lafon, M., Forel, M. T., & Garrigou-Lagrange, C. (1973) *Spectrochim. Acta Part A 29*, 471–486.
Sen, A. C., & Keiderling, T. A. (1984) *Biopolymers 23*, 1519–1532.
Surewicz, W. K., & Mantch, H. H. (1988) *Biochim. Biophys. Acta 952*, 115–130.
Thomas, G. J., Jr., & Agard, D. A. (1984) *Biophys. J. 46*, 763–768.
Van Wart, H. E., & Scheraga, H. A. (1978) *Methods Enzymol. 49*, 67–149.
Williams, R. W. (1983) *J. Mol. Biol. 166*, 581–603.
Williams, R. W. (1986) *Methods Enzymol. 130*, 311–331.
Williams, R. W., & Dunker, A. K. (1981) *J. Mol. Biol. 152*, 783–813.
Yu, N. T., & Jo, B. H. (1973) *J. Am. Chem. Soc. 95*, 5033–5037.

Nucleotide Sequence and DNA Recognition Elements of *alc*, the Structural Gene Which Encodes Allantoicase, a Purine Catabolic Enzyme of *Neurospora crassa*[†,‡]

Hakjoo Lee, Ying-Hui Fu, and George A. Marzluf*
Department of Biochemistry, The Ohio State University, Columbus, Ohio 43210
Received April 25, 1990; Revised Manuscript Received June 20, 1990

ABSTRACT: The nitrogen regulatory circuit of *Neurospora crassa* contains structural genes that encode nitrogen catabolic enzymes which are subject to complex genetic and metabolic regulation. This set of genes is controlled by nitrogen limitation, by specific induction, and by the action of *nit*-2, a major positive-acting regulatory gene, and *nmr*, a negative-acting control gene. The complete nucleotide sequence of *alc*, the gene that encodes allantoicase, a purine catabolic enzyme, is presented. The *alc* gene contains a single intron, is transcribed from two initiation sites situated approximately 50 nb upstream of the translation start site, and encodes a protein comprised of 354 amino acids. Mobility shift and DNA footprint experiments identified a single binding site for the NIT2 regulatory protein in the *alc* promoter region. The binding site contains a 10 nucleotide base pair symmetrical sequence which is flanked by two possible core binding sequences, TATCT and TATCG. Mutant NIT2/β-gal fusion proteins with amino acid substitutions in a putative zinc-finger motif were shown to be completely deficient in the ability to bind to the *alc* promoter DNA fragment.

Neurospora crassa utilizes purines as secondary nitrogen sources, when primary sources such as glutamine and ammonia are limiting. Expression of the purine catabolic genes of *N. crassa* is controlled by uric acid induction, a pathway-specific regulation (Reinert & Marzluf, 1975). The purine catabolic genes and other structural genes of the nitrogen regulatory circuit of *N. crassa* are also controlled by nitrogen repression. Two regulatory genes, *nit*-2, a major positive control gene, and *nmr*, a negative control gene, together mediate nitrogen repression.

[†] This research was supported by Public Health Service Grant GM-23367 from the National Institutes of Health.
[‡] The nucleic acid sequence in this paper has been submitted to GenBank under Accession Number J02927.
* Corresponding author.

0006-2960/90/0429-8779$02.50/0

The *alc* gene, which encodes allantoicase, a purine catabolic enzyme, has been cloned (Lee et al., 1990). Expression of the *alc* gene was investigated by comparing mRNA content and enzyme activity of cells in nitrogen-repressed and in induced conditions. The *alc* gene is transcribed to give a single 1.2-kb poly(A+) RNA. Allantoicase activity of induced cells increased approximately 3-fold, compared to the enzyme activity in nitrogen-repressed cells. However, the mRNA content was approximately the same in both induced and N-repressed cells. These results suggested that expression of the *alc* gene is regulated at least in part at the posttranscriptional level. However, Northern blot analyses showed that mutants of the *nit*-2 and the *nmr* regulatory genes affect the cellular *alc* mRNA content, indicating that these two control genes act at the transcriptional level.

Expression of the *nit*-3 gene, which encodes nitrate reductase, the first enzyme of the nitrate assimilatory pathway, has been studied in detail (Fu & Marzluf, 1988). The *nit*-3 gene is controlled by nitrogen repression and by pathway-specific induction and is expressed in wild-type cells only when they are subjected to nitrogen limitation and nitrate induction. *nit*-2 mutants fail to express nitrate reductase mRNA or enzyme, suggesting that the *nit*-3 gene is controlled by *nit*-2 at the transcriptional level.

The *nit*-2 gene encodes a regulatory protein of molecular weight of 110000 which contains a DNA binding domain comprised of a single zinc-finger element and a downstream basic region (Fu & Marzluf, 1990a). To investigate how the NIT2 regulatory protein activates transcription of structural genes of the nitrogen regulatory circuit, its control of the *alc* gene has been examined. Here we report the complete nucleotide sequence of the *alc* gene and show that it contains a small intron near the 5′-end of the gene. The *alc* gene encodes a protein comprised of 354 amino acids with a molecular weight of approximately 39000. We also present the results of gel band mobility shift experiments and DNA footprint studies which demonstrate that the NIT2 protein binds in a sequence-specific manner to the 5′-promoter region of the *Neurospora alc* gene. Both the zinc-finger domain and the positively charged region of NIT2 are essential for its DNA binding activity.

Materials and Methods

Strains. *N. crassa* wild-type strain 740R231A and the *nit*-2 mutant strain (allele nr37) were obtained from the Fungal Genetics Stock Center, Kansas City, KS. Cultures were grown in Vogel's liquid medium, with shaking at 30 °C as described previously (Lee et al., 1990).

S1 Nuclease and Primer Extension Mapping and DNA Sequencing. DNA sequencing was accomplished by the dideoxy chain termination method (Sanger et al., 1977) using [α-^{32}P]dATP and a modified T7 bacteriophage DNA polymerase, "Sequenase" (U.S. Biochemical Corp.). Plasmid DNA templates were prepared as minipreps (Birnboim & Doly, 1979). Deletion clones of pALC-1 and pALC-2 were constructed with exonuclease III and mung bean nuclease as described previously (Lee et al., 1990). Oligonucleotide primers for DNA sequencing and mobility shift experiments were synthesized on an Applied Biosystems Model 380B DNA synthesizer by the Ohio State University Biochemical Instrument Center. S1 nuclease mapping was conducted by the method of Birk and Sharp (1977). Primer extension experiments were accomplished by hybridizing a single-stranded, 5′-end-labeled *Cla*I–*Eco*RV restriction fragment of *alc* with poly(A+) RNA as a template for Molony murine leukemia virus reverse transcriptase (Bethesda Research Labs).

Site-Directed Mutagenesis. Site-directed mutagenesis of the *nit*-2 DNA fragment containing the zinc-finger motif was performed as described previously (Fu & Marzluf, 1990a; Kunkel, 1985).

Expression and Purification of a NIT2/β-Gal Fusion Protein. The *Bam*H1–*Eco*RI fragment of *nit*-2 deletion clone 22 (Fu & Marzluf, 1987) was cloned into the corresponding sites of pSKS106 (Casadaban et al., 1983) to obtain pN2SKS. pN2SKS was designed such that 217 amino acids of the NIT2 protein, which includes the entire zinc-finger domain and adjacent basic region, were fused to a short amino-terminal sequence (13 residues) coded for by the vector, and at its carboxy terminus the NIT2 finger domain was fused to the entire β-galactosidase protein. The *nit*-2 mutants obtained by site-directed mutagenesis were each introduced into the NIT2/β-gal fusion construct by replacing the wild-type *Kpn*I–*Eco*RI DNA fragment of pN2SKS with the corresponding mutated DNA fragment. These plasmids were each introduced into *E. coli* strain JM103 to express the fusion proteins (Studier & Moffatt, 1986). When the bacterial culture reached an absorbance of 0.5 at 600 nm, expression of the fusion protein was induced for 2 h with 0.5 mM IPTG. The cells were sedimented by centrifugation and resuspended in 20 mM Tris-HCl, 10 mM $MgCl_2$ (pH 7.4), and 1 mM PMSF and sonicated. Following centrifugation, the supernatant was subjected to aminobenzyl 1-thio-β-D-galactopyranoside affinity chromatography as described by Ullmann (1984). An aliquot of each purified protein was analyzed by SDS–PAGE. Except for a full-length β-gal (116 kDa) band and a shorter β-gal fragment (46 kDa), the fusion protein (140 kDa) was the only visible band on the gel (not shown).

Isolation of RNA and cDNA Clones. *N. crassa* total RNA was prepared by the method of Reinert et al. (1981); the poly(A+) fraction was isolated by oligo(dT)–cellulose chromatography (Aviv & Leder, 1972). A λ gt10 cDNA library of *Neurospora* was screened by plaque hybridization with pALC-1 as a probe. Several *alc* cDNA clones were isolated, and the DNA of the clone with the longest insert was purified and digested with *Eco*RI, and the insert cDNA was subcloned into the Bluescript plasmid vector and sequenced as described above.

Computer Methods. The handling of sequences, their analyses, translation, and hydropathy, and codon bias analyses were accomplished with Pustell software (International Biotechnology, Inc.).

Gel Band Mobility Shift Experiments. Radioactive DNA fragments used for gel band mobility shift experiments were labeled with ^{32}P by filling in with the Klenow fragment of DNA polymerase I. Mobility shift experiments were carried out as described by Hope and Struhl (1985). A NIT2/β-gal fusion protein (or no protein for free DNA) was added to a binding buffer [20 mM Tris-Hcl (pH 7.4), 50 mM KCl, 2 mM MgCl, 1 mM EDTA, and 100 μg/mL gelatin] supplemented with 1 μg of p(dI.dC) and an end-labeled DNA fragment (10000 cpm) with or without nonradioactive competitor DNA in a 15-μL total volume. The samples were incubated at 25 °C for 20 min, at which time 5 μL of loading buffer (binding buffer containing 20% glycerol, 1 mg/mL xylene cyanol FF, and 2 mg/mL bromophenol blue) was added. The samples were loaded onto 5% polyacrylamide [acrylamide:bis(acrylamide), 30:0.8] gels which were run with 90 mM Tris–borate buffer, pH 8.3. After electrophoresis, the gels were dried and exposed to Kodak XAR-5 film at −70 °C.

DNA Footprinting Analysis. Tubes containing all reactants were incubated at 25 °C for 20 min. Conditions for binding

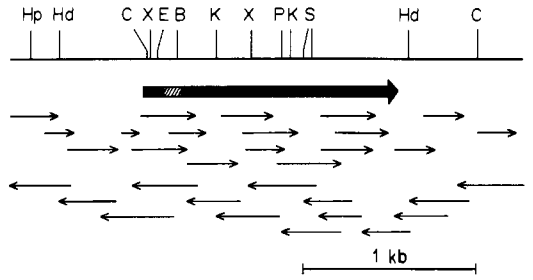

FIGURE 1: DNA sequencing strategy for the *alc* gene. The restriction map displays the 2.7-kp region of plasmids pALC-1 and pALC-2 which was sequenced. The thick arrow indicates the location of the cDNA. Key: thin arrows, extent and direction of each fragment sequences; striped area, intron; B, *Bgl*II; C, *Cla*I; E, *Eco*RV, Hd, *Hin*dIII; Hp, *Hpa*I; K, *Kpn*I; P, *Pst*I; S, *Sal*I; X, *Xho*I.

NIT2/β-gal fusion protein to end-labeled DNA were as described above, except that the total reaction volume was 20 μL, more probe (100 000 cpm) was used than in the mobility shift assay, and no carrier DNA was added (Hope & Struhl, 1985). A total of 2–7.5 μL of 20 μg/mL DNase I in the binding buffer was added to each sample. DNase I digestion was halted after 2 min at 25 °C by addition of an equal volume of stop solution (0.5% SDS, 10 mM EDTA, and 0.5 mg/mL yeast tRNA). The DNA samples were extracted with PCA, ethanol precipitated, resuspended in 2 μL of a loading solution (80% formamide, 0.1% xylene cyanol FF, and 0.1% bromophenol blue), and electrophoresed on a sequencing gel.

RESULTS

alc Nucleotide Sequence. A restriction map of the *alc* gene and the strategy used to sequence it are shown in Figure 1. The figure also shows the location of the cDNA whose nucleotide sequence was analyzed. The nucleotide sequence of both strands of the entire *alc* gene was determined with the dideoxy sequencing technique; overlapping clones were used to confirm the sequence of each segment. The *alc* cDNA was sequenced with the same technique, with four different oligonucleotide primers (17-mers). The nucleotide sequence of the *alc* gene and its flanking regions is presented in Figure 2.

Initiation of alc Transcription. The direction of transcription as shown in Figure 1 was determined by use of single-stranded probes in Northern blot experiments (results not shown). An S1 nuclease mapping experiment showed that the 5′-end of the *alc* transcript was located at positions −54 and −53. This result was confirmed by a primer extension assay, which revealed that transcription of *alc* is initiated mainly at −54 and, to a lesser extent, at −53 (Figure 3). The *alc* cDNA 3′-terminus occurs at position −12, indicating that the cDNA insert is almost full length and ends 42 bp downstream from the transcription initiation site determined by S1 nuclease and primer extension mapping.

Intervening Sequence in the alc Gene. Comparison of the genomic and cDNA sequences shows that the *N. crassa alc* gene is interrupted by one intron of 72 bp. The intron is located near the translation start site, contains two termination codons, UAG and UAA in the same reading frame as in the *alc* protein, and lacks the typical codon bias of *N. crassa*. The intron has good consensus 5′ and 3′ splice site sequences and an internal conserved sequence typical for *N. crassa* (Orbach et al., 1986).

Organization of the alc Gene. The *alc* promoter has a possible TATA element centered at position −66 and stretches of poly(T) close to the transcription start site. A protein-coding region locator program, which incorporates the coding bias typical of *Neurospora* proteins, e.g., a predominant use of codons ending in either C or T and an almost complete exclusion of codons ending in A (Orbach et al., 1986), demonstrated that the open reading frame which specifies the *alc* protein is highly favored as a coding region. No other reading frame in either orientation of the entire sequenced region shows any significant codon bias. Only one ATG codon (designated +1 in Figure 2) is found from −250 to +55, with a second ATG at +56. The ATG at +1 has a good consensus sequence for translation initiation codons of *N. crassa* (Roberts et al., 1988). The predicted allantoicase protein is composed of 354 amino acid residues, with a calculated molecular weight of 39 165 and a p*I* value of 5.71. The protein is composed of 42%, 31%, 13%, and 14% of nonpolar, polar, acidic, and basic amino acid residues, respectively. The amino acid sequence of the predicted allantoicase protein is shown along with the nucleotide sequence (Figure 2). Comparison of the sequences of genomic and cDNA clones revealed that polyadenylation of the *alc* transcript occurs 176 bp downstream of the UAA stop codon (Figure 2). The typical mammalian poly(A) additional signal, AATAAA, is absent from the *alc* gene. However, there do occur three AUUUA motifs in the 3′-untranslated region which might function in rapid turnover of the *alc* mRNA (Shaw & Kamen, 1986).

Binding of NIT2/β-Gal Fusion Protein to alc Promoter Fragment. To investigate whether the NIT2/β-gal fusion protein would bind to the *alc* promoter DNA, a labeled 410-bp *Acc*I–*Cla*I (−399 to +11) fragment of the *alc* gene was used as a probe. Mobility shift experiments demonstrated that the NIT2/β-gal fusion protein binds to the *alc* promoter DNA (data not shown). To locate the binding region more precisely, shorter restriction fragments were employed as probes. As shown in Figure 4, , the *Rsa*I–*Cla*I (−186 to +11) and the *Rsa*I–*Taq*I (−186 to −45) fragments were bound by the fusion protein, but the *Acc*I–*Alu*I (−399 to −152) and the *Alu*I–*Cla*I (−152 to +11) fragments were not. These results demonstrate that the binding region for NIT2/β-gal is located within the *Rsa*I–*Taq*I fragment and imply that the binding site includes the *Alu*I site. The results of these mobility shift experiments are summarized in Figure 4B.

DNase I Protection of alc DNA by NIT2/β-Gal Fusion Protein. DNA footprints were used to define the binding site for the NIT2 protein. The NIT2/β-gal fusion protein protects a 30-bp region from DNase I digestion (Figure 5). As predicted by the mobility shift assays, the protected region contains the *Alu*I site, which comprises part of a symmetrical dyad of 10 bp. The binding region, from −169 to −138 with respect to the *alc* translation start site, includes the sequence TATCT, implicated as a part of the NIT2 recognition site, and a second copy of the closely related sequence, TATCG. When bound by the NIT2 protein, the coding strand displays one hypersensitive site, which is located at the center of the symmetrical dyad. The DNA footprint, which defines a single NIT2 protein binding site, is depicted in Figure 5D.

Binding of the NIT2/β-Gal Fusion Protein to a Synthetic Oligonucleotide. In order to confirm the identity of the NIT2 binding site, a synthetic 36-bp double-stranded oligonucleotide was synthesized. The 36-mer contains the *alc* 5′-sequence identified by the DNase I protection experiments with the NIT2/β-gal fusion protein. Figure 5D shows a comparison between the promoter segment and the oligonucleotide. Binding of the fusion protein to the 36-bp oligonucleotide was demonstrated by mobility shift experiments as shown in Figure 6. With increasing amounts of the nonradioactive 36-mer oligonucleotide as a competitor, the intensity of the shifted bands decreased significantly. In contrast, the unrelated se-

```
      -240       -230       -220       -210       -200
TTGCAGATCG AATACGACGG TTAGGTACGA CGAAGAAGGA CCACGATTGT CGTTG

 -190       -180       -170       -160       -150       -140
CTGTT ACGTACTTTG ACCTCCTCAA CGCACTATCT TGCTTAAGCT ATCGCTCTTG

       -130       -120       -110       -100       -90
TCTGTCGCTG TGGTGATATA AATTCTGTCC GCCTGCTCTT GGTTTATTCC GAGGA

   -80        -70        -60        -50        -40        -30
CGCTC GTTCCATCTC TGTTTTTTTT TTCTCTCTGT GACATCGAGG ACTGAAGTCT

       -20        -10          1          10          20
CACTTATTCA AATACACATT TCCCTCACC  ATG ACC GAC ATC GAT TAC AAG
                                 Met Thr Asp Ile Asp Tyr Lys

         30          40          50          60
CTC GAG GCT GTT CCG GCC ACT CGG ATT GCC GCC GAT GAT ATC GAC
Leu Glu Ala Val Pro Ala Thr Arg Ile Ala Ala Asp Asp Ile Asp

   70          80          90              100        110
AAG ACT TTC CGT TCC AGC ACC ATC G      GTCCGTAGC ATCCATCTCA
Lys Thr Phe Arg Ser Ser Thr Ile A

       120        130        140        150        160
CCAAACATGG CAACCCAAAC CTTTCAACTA ACGGAAGTCG AGCTGGGATA CAG

       170          180          190          200
 AT CTT ATC TCA GGG GCT CTC GGT GGC AAG GTT TCC GGT TTC TCG
 sp Leu Ile Ser Gly Ala Leu Gly Gly Lys Val Leu Gly Phe Ser

210          220          230          240          250
GAC GAA TGG TTC GCC GAA GCA GCC AAC CTC CTC ACT CCT ACA GCC
Asp Glu Trp Phe Ala Glu Ala Ala Asn Leu Leu Thr Pro Thr Ala

       260          270          280          290
CCA ATC CGC CAG CCG GGA AAG ATG GTT TAC ACC GGC GCC TGG TAT
Pro Ile Arg Gln Pro Gly Lys Met Val Tyr Thr Gly Ala Trp Tyr

300          310          320          330          340
GAC GGA TGG GAG ACA AGG AGA CAC AAC CCT GCC GAG TTC GAC TGG
Asp Gly Trp Glu Thr Arg Arg His Asn Pro Ala Glu Phe Asp Trp

       350          360          370          380
GTT GTG ATC CGT CTG GGC GTC GCC TCG GGT ACC GTC GAG GGT GTC
Val Val Ile Arg Leu Gly Val Ala Ser Gly Thr Val Glu Gly Val

390          400          410          420          430
GAG ATT GAC ACG GCT TTC TTC AAC GGC AAC CAT GCG CCC GCC ATC
Glu Ile Asp Thr Ala Phe Phe Asn Gly Asn His Ala Pro Ala Ile

       440          450          460          470
TCG GTC GAG GGT TGC TTC AGC CAA AAC GAC GAT GAG GTT CTG TCA
Ser Val Glu Gly Cys Phe Ser Gln Asn Asp Asp Glu Val Leu Ser
```

```
480           490           500           510           520
  *             •             •             *             •
TGG AAG GGC GAG CTG GGT GGA TGG GAG ACT ATT CTT GGC GTT CAA
Trp Lys Gly Glu Arg Gly Gly Trp Glu Thr Ile Leu Gly Val Gln

       530           540           550           560
         •             •             •             •
GAG TGC GGC CCT TCG CAG AGA TTC TGC TGG AAA CTC GAG AAC CCT
Glu Cys Gly Pro Ser Gln Arg Phe Gly Trp Lys Leu Glu Asn Pro

 570           580           590           600           610
  *             *             •             *             •
ACC AAG AAG CAG TAC ACC CAT GTG CGA CTA AAC ATG TAC CCC GAC
Thr Lys Lys Gln Tyr Thr His Val Arg Leu Asn Met Tyr Pro Asp

       620           630           640           650
         •             *             *             *
GGC GGC ATT GCC AGG TTC CGT CTG TTT GGA CAC GCC GTA CCG GTC
Gly Gly Ile Ala Arg Phe Arg Leu Phe Gly His Ala Val Pro Val

660           670           680           690           700
  *             •             •             •             •
TTC CCC GAC AAT ACG GAT GCC ATC TTT GAC TTG GCG GCT GCC CAG
Phe Pro Asp Asn Thr Asp Ala Ile Phe Asp Leu Ala Ala Ala Gln

       710           720           730           740
         *             •             •             *
AAC GGC GGA GTT GCG ATC TCC TGC AGT GAC CAG CAC TTT GGT ACC
Asn Gly Gly Val Ala Ile Ser Cys Ser Asp Gln His Phe Gly Thr

750           760           770           780           790
  *             •             *             •             *
AAG GAC AAC CTT ATC CTT CCG GGC CGC GGC AAG GAC ATG GGC GAC
Lys Asp Asn Leu Ile Leu Pro Gly Arg Gly Lys Asp Met Gly Asp

       800           810           820           830
         •             *             •             •
GGT TGG GAG ACA GCA CGC TCG CGC ACC AAG GGC CAC GTC GAC TGG
Gly Trp Glu Thr Ala Arg Ser Arg Thr Lys Gly His Val Asp Trp

840           850           860           870           880
  •             •             •             *             •
ACC ATC ATC AGA CTC GGC GCG CCC GGC TAC ATT CAG AAT TTC ATG
Thr Ile Ile Arg Leu Gly Ala Pro Gly Tyr Ile Gln Asn Phe Met

       890           900           910           920
         *             *             *             *
GTC GAC ACG GCT CAC TTC CGC GGT AAC TAC CCC CAG CAG GTC AAG
Val Asp Thr Ala His Phe Arg Gly Asn Tyr Pro Gln Gln Val Lys

930           940           950           960           970
  •             *             *             •             •
CTG CAA CGT ATC GAG TGG AAG AGC GAG GGC AGG CCG GGA GCG GAT
Leu Gln Arg Ile Glu Trp Lys Ser Glu Gly Arg Pro Gly Ala Asp

       980           990          1000          1010
         •             •             *             *
TCT GAG GGC TGG ACA GAG GTT GTT GAG CCC ATC AAG TGC GGT CCC
Ser Glu Gly Trp Thr Glu Val Val Glu Pro Ile Lys Cys Gly Pro

1020          1030          1040          1050          1060
  *             •             •             *             •
GAT CAG GAA CAC CCT GTC GAG AGC TTG GTG AAG GAC AAG CCG TTC
Asp Gln Glu His Pro Val Glu Ser Leu Val Lys Asp Lys Pro Phe

      1070          1080          1090          1100
         •             •             •             *
ACC CAC GTC AAG CTC ATC ATT GTG CCT GAC GGC GGA GTG AAA AGA
Thr His Val Lys Leu Ile Ile Val Pro Asp Gly Gly Val Lys Arg
```

```
1110          1120          1130                1140          1150
 *             *             *                   *             *
CTG CGG GTG TTT GCG AAG AGG GCT GTT TAA        GAA ATTACCAAGC
Leu Arg Val Phe Ala Lys Arg Ala Val End

      1160          1170          1180          1190          1200
        *             *             *             ●             *
TATATATCTG AAGGCAATTA TTCGGTGAGA GCAGCATTTA CGGGGAGCCA TCAAC

 1210          1220          1230          1240          1250          1260
   *             ●             *             *             ●             *
AGCGA GCGATCCACA TAAAAAGGGG GAGGACCTCA TTTAGTATGA TGGGCAACGA

      1270          1280          1290          1300          1310
        *             *             *             *             *    ■
GTGCAGTCAT TTAGCCGCGA AGAATCGAAA TCTCTCAGAT CTTTGATTGT CTGCG

 1320          1330          1340          1350          1360          1370
   ●             *             *             *             *             *
CTTAA GTAACAAAGT CTAATTCTCA ATCAGCTTTC GTCGTAGAGT AAAATTAGAA
```

FIGURE 2: DNA sequence of the *alc* gene. The nucleotide sequence is numbered from the ATG initiation codon. The amino acid sequence of the protein was deduced from the DNA sequence. A single intron is located at codon 31. A possible TATA element at −123 is enclosed in a box. The transcription start site is indicated with a filled circle. The polyadenylation site for the cDNA is indicated with a filled square. ATTTA (AUUUA) elements are underlined.

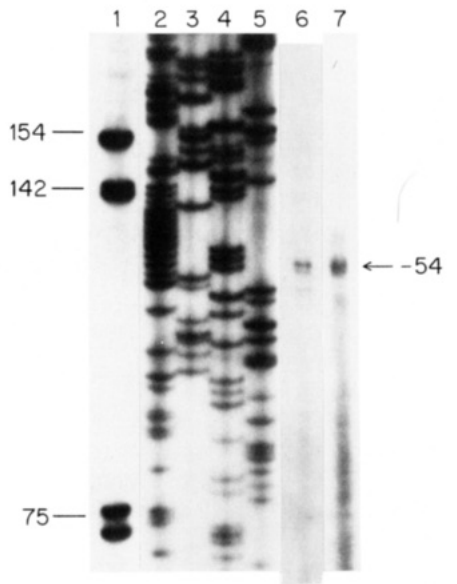

FIGURE 3: S1 nuclease mapping and primer extension mapping of the 5′-end of the *alc* transcript. Poly(A+) RNA from wild-type cells was the template for primer extension (cDNA synthesis) with the single-stranded *Cla*I–*Eco*RV fragment as a primer. Lane 6 shows the primer extension products. Lane 7 shows the DNA fragments protected by RNA from digestion by S1 nuclease; poly (A+) RNA from wild-type cells was hybridized with the end-labeled 1-kb *Kpn*I–*Eco*RV fragment and treated with S1 nuclease as described under Materials and Methods. Lane 1 contains end-labeled 1-kb DNA ladder as size markers. Lanes 2–5 show DNA sequencing T, G, C, and A, respectively, carried out with the oligonucleotide primer starting from the *Eco*RV site. Lane 6 was positioned in relation to the lanes for sequencing, as determined from a separate experiment.

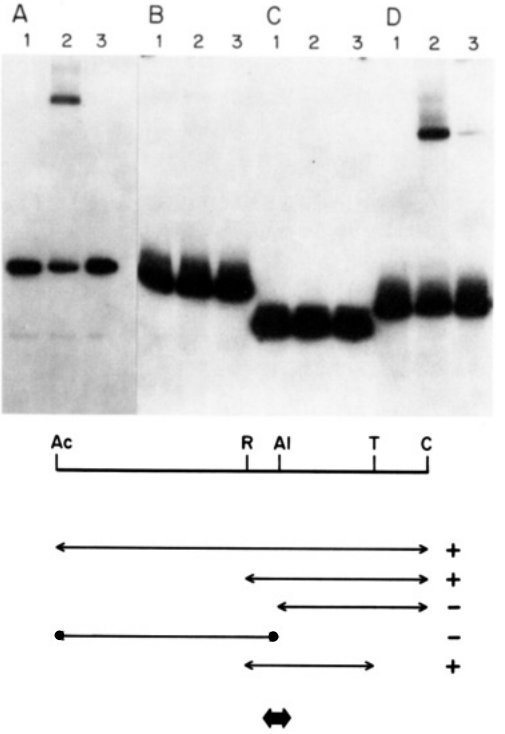

FIGURE 4: Binding of different *alc* promoter fragments by the NIT2/β-gal fusion protein. Upper: panels A–D show the results with the *Ral*–*Taq*I (−186 to −45), the *Acc*I–*Alu*I (−399 to −152), the *Alu*I–*Cla*I (−152 to +11), and the *Rsa*I–*Cla*I (−186 to +11) fragments, respectively. Lanes: 1, no protein; 2, 0.3 μg of fusion protein; 3, 0.3 μg of fusion protein and 0.25 μg of the unlabeled *Acc*I–*Cla*I fragment of *alc*, as a competitor. Lower: summary of mobility shift assays with the NIT2/β-gal fusion protein. The arrows indicate the restriction fragments employed as probes. The thick arrow indicates the region protected from digestion with DNase I by binding of the fusion protein. The restriction map of the promoter indicates only the sites used for preparation of these fragments. Key: (+) fusion protein specific binding; (−) no binding; Ac, *Acc*I; Al, *Alu*I; C, *Cla*I; R, *Rsa*I; T, *Taq*I.

quence p(dI.dC) competed very poorly. These results indicate that the 36-mer includes a specific binding site for the NIT2/β-gal fusion protein and suggest that DNA binding by NIT2 is sequence specific. The nonradioactive 36-mer also competed efficiently against a restriction fragment of the *alc* promoter containing the binding region, indicating an authentic binding site is present within the 36-mer (data not shown).

Mobility Shift Assays with Mutant NIT2/β-Gal Fusion Proteins. The importance of the possible zinc-finger motif and the adjacent basic region of NIT2 for DNA binding was examined by creating NIT2/β-gal mutants which have mutations in five different sites within the region of interest (Figure 7). The five mutant fusion genes were created with site-directed mutagenesis and expressed in *Escherichia coli*.

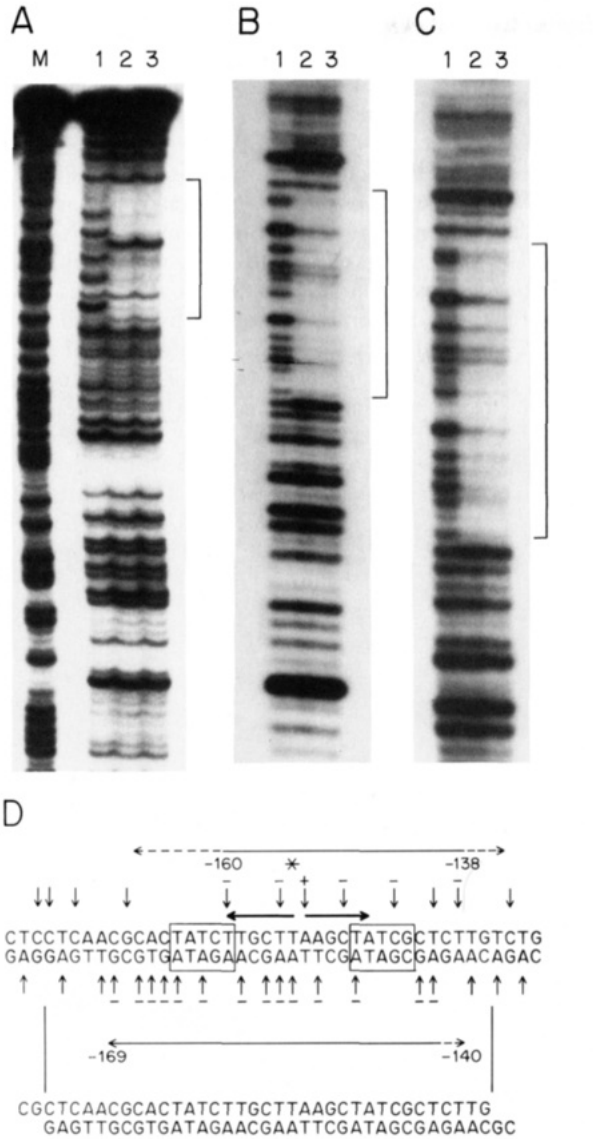

FIGURE 5: Protection of *alc* DNA from DNase I cleavage by the NIT2/β-gal fusion protein. Footprints of the coding (panel A) and noncoding (panels B and C) strands on the *alc* promoter region are shown. For the coding strand, the *Rsa*I–*Taq*I (−186 to −45) fragment was 3′ end labeled at the *Taq*I site. For the noncoding strand, the *Ava*II–*Cla*I (−212 to +11) fragment was 3′ end labeled at the *Ava*II site. Panels B and C show the same footprint except that in panel C the gel was run longer than in panel B. The probes were incubated with NIT2/β-gal fusion protein (0, 0.3, and 0.5 μg for lanes 1–3, respectively). After controlled digestion with DNase I, the products were examined on a sequencing gel. In each case, a sample of the same ^{32}P-labeled DNA fragment was subjected to the A+G reaction of the Maxam–Gilbert (1980) sequencing procedure and run in parallel in order to identify the protected region (lane M). Brackets indicate the protected region. Panel D: results of the footprinting analysis. Sites of DNase I cleavage on both strands of DNA between nucleotides −177 and −132 are indicated with vertical arrows. DNase I cleavages enhances or suppressed by NIT2/β-gal fusion protein are indicated as (+) or (−), respectively. the TATCT and the TATCG sequences are boxed, and the dyad is indicated with arrows, with its center marked with an asterisk. The protected regions are identified with solid lines. Dotted lines indicate other possible protected regions that are less clear because of poor digestion by DNase I. The 36-mer oligonucleotide is shown aligned with the binding site sequence.

As shown in Figure 7, all of the five mutant proteins failed to bind to the restriction fragment containing the *alc* promoter and to the 36-mer containing the NIT2 protein binding site. This indicates that DNA binding by the NIT2 protein indeed requires the zinc-finger motif and the adjacent basic region on its carboxyl side; the five mutations each identify amino

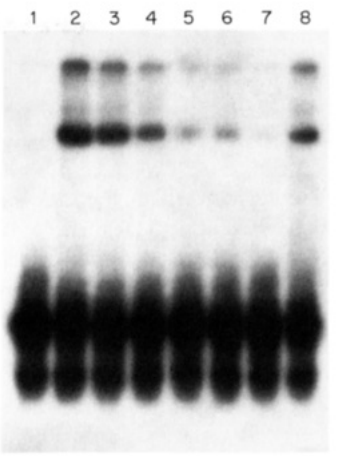

FIGURE 6: Binding of the 36-mer by the NIT2/β-gal fusion protein. The radioactively labeled 36-mer was present in samples for all lanes. Lane 1, 36-mer only. All other lanes contain the 36-mer plus 0.5 μg of the NIT2/β-gal fusion protein. Lanes 2–7 contain 0, 5, 10, 25, 50, and 125 ng of cold 36-mer as a competitor, respectively. Lane 8, 125 ng of p(dI·dC) as a nonspecific competitor.

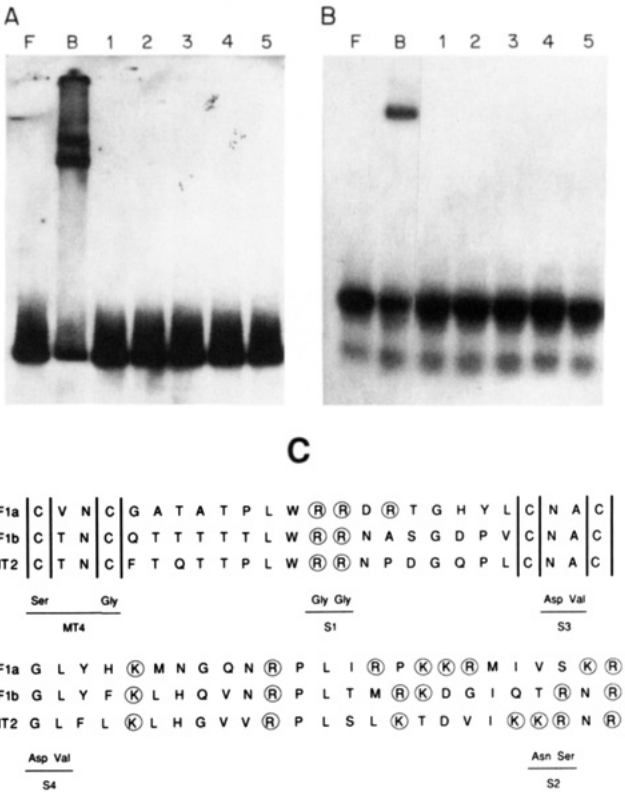

FIGURE 7: Mobility shift assay with mutant NIT2/β-gal fusion proteins. Panel A shows results with the Klenow-labeled *Rsa*I–*Taq*I fragment and 0.3 μg of each of the NIT2/β-gal fusion proteins. Panel B shows results with the 36-mer and 0.5 μg of the fusion proteins. Lanes: 1, free DNA; 2–7, NIT2/β-gal fusion proteins from wild-type and mutants MT4, S1, S2, S3, and S4, respectively (see panel C). Panel C: the amino acid sequence of the NIT2 zinc-finger motif and its immediate downstream basic region is compared with that of the two zinc fingers of the homologous mammalian GF1 protein; regions with strong sequence identity are underlined and in boldface. The identity and location of the substituted amino acids in each of the NIT2 mutants are shown under the wild-type sequence.

acid residues in NIT2 which are important for DNA binding.

Discussion

The nitrogen regulatory circuit of *N. crassa* contains structural genes which encode catabolic enzymes for several

different pathways, each involved in the metabolism of secondary nitrogen sources. Expression of the genes of the circuit occurs during conditions of nitrogen limitation and is dependent upon the positive-acting NIT2 regulatory protein. The NIT2 protein consists of 1036 amino acids and contains a single Cys_2/Cys_2-type zinc-finger motif and an adjacent positively charged region, which together may comprise a DNA binding domain (Fu & Marzluf, 1990a). The NIT2 protein may function as a global transcriptional activator, suggesting that one or more NIT2 binding sites may be present in the promoters of many different nitrogen-regulated structural genes. The entire NIT2 protein is not expressed in bacterial cells, possibly because of its large size. However, we have expressed a segment of the NIT2 protein which contains its zinc finger and basic region in *E. coli* as a NIT2/β-gal fusion protein. The NIT2/β-gal fusion protein binds to three different sites upstream of the *Neurospora* nitrate reductase gene, *nit*-3, and also binds to the promoter regions of the related *Aspergillus* genes, *niaD* and *niiA*, which encode nitrate and nitrite reductase, respectively (Fu & Marzluf, 1990b).

Characterization of the *alc* gene, which encodes allantoicase, provided an opportunity to determine whether promoters for structural genes of the purine catabolic pathway also contain binding sites for NIT2. Mobility shift assays and DNA footprint experiments demonstrated that NIT2 indeed binds to a single site located at approximately position -150 upstream of the *alc* translation start site. A DNase I protected region of 32 bp was identified and possesses a symmetrical dyad of 10 bp; the finding that a DNase I hypersensitive site lies in the center of the dyad supports the concept that this region is recognized by NIT2 (Figure 5). Homologous pentanucleotide sequences, TATCT and TATCG, are located on opposite sides of the dyad and within the NIT2 binding region of the *alc* promoter. TATCT appears to represent a core binding sequence for NIT2 and the related mammalian protein GF1 and occurs at least twice in each of three binding sites upstream of the *Neurospora* nitrate reductase structural gene and in the four binding sites upstream of the homologous gene of *Aspergillus* (Fu & Marzluf, 1990b). The binding sites upstream of the nitrate reductase genes, however, lack a symmetrical dyad element as found in the case of the *alc* gene. The basis for sequence-specific recognition by NIT2 appears to be the presence of the core pentanucleotide sequences, since the various binding regions otherwise lack obvious homology. The mammalian GF1 protein also recognizes a core consensus DNA sequence, TATCT, present in promoter and enhancer elements of all members of the globin gene family (Tsai et al., 1989). Moreover, beyond the core sequence, the various binding sites for GF1 appear to lack any sequence homology (Tsai et al., 1989). It is well established that some regulatory proteins can recognize two or more distinct DNA sequences, e.g., the yeast HAP1 protein (Pfeifer et al., 1987), the octamer binding protein (OBP 100), the glucocorticoid receptor, and C/EBP (Johnson & McKnight, 1989).

The importance of the highly conserved amino acid residues in the zinc-finger motif and the downstream positively charged region of NIT2 for binding to the *alc* promoter was examined by use of site-directed mutagenesis. NIT2 mutant proteins with amino acid substitutions in these regions were all defective in binding DNA fragments carrying the *alc* promoter. These same mutant proteins did not function in vivo and were also incapable of binding to the upstream sites of the *nit*-3 gene, implying that the same domain of NIT2 is involved in binding to different genes of the nitrogen circuit. Since NIT2 can turn on various structural genes to widely different extents, we anticipate that individual sites show significant differences in binding affinity for NIT2, perhaps providing an explanation for the lack of obvious homology between the *alc* binding site and the binding sites for *nit*-3.

Expression of the allantoicase structural gene, as well as the genes encoding other purine catabolic enzymes such as xanthine dehydrogenase, uricase, and allantoinase, is also controlled by *nmr*, a negative-acting control gene, although its mode of action is unknown. In addition, the purine catabolic genes require induction by uric acid, an intermediate of the pathway. Attempts to identify a binding site in the *alc* promoter region for a possible pathway-specific regulatory protein which mediates induction were conducted with nuclear extracts for DNA footprint experiments. However, the presence of many DNA binding proteins prevented identification of any specific binding sites (results not shown). It appears that the purine catabolic genes of *N. crassa* may be controlled in a complex manner that involves the action of multiple control elements. Thus, further studies of the *alc* gene should reveal new insights concerning the operation of the nitrogen control circuit.

Acknowledgments

We thank Dr. F. William Studier for providing the pET3b expression vector and *E. coli* host strains. Song Han provided help in the mutagenesis of NIT2. We acknowledge Richard Swenson and Jane Tolley, Ohio State Biochemical Instrument Center, who prepared the oligonucleotides used in this work.

References

Avi, H., & Leder, P. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69*, 1408–1412.

Birk, A. J., & Sharp, P. A. (1977) *Cell 12*, 721–732.

Birnboim, H. C., & Doly, J. (1979) *Nucleic Acids Res. 7*, 1513–1523.

Casadaban, M. J., Martinez-Arias, A., Shapira, S. K., & Chou, J. (1983) *Methods Enzymol. 100*, 27–31.

Fu, Y. H., & Marzluf, G. A. (1987) *Mol. Cell. Biol. 7*, 1691–1696.

Fu. Y. H., & Marzluf, G. A. (1988) *J. Bacteriol. 170*, 657–661.

Fu, Y. H., & Marzluf, G. A. (1990a) *Mol. Cell. Biol. 10*, 1056–1065.

Fu, Y. H., & Marzluf, G. A. (1990b) *Proc. Natl. Acad. Sci. U.S.A. 87*, 5331–5335.

Hope, I. A., & Struhl, K. (1985) *Cell 43*, 177–188.

Johnson, P. F., & McKnight, S. L. (1989) *Annu. Rev. Biochem. 58*, 799–839.

Kunkel, T. A. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 488–492.

Lee, H., Fu, Y. H., & Marzluf, G. A. (1990) *Mol. Gen. Genet. 222*, 140–144.

Maxam, A. M., & Gilbert, W. (1980) *Methods Enzymol. 65*, 499–560.

Orbach, M. J., Porro, E. B., & Yanofsky, C. (1986) *Mol. Cell. Biol. 6*, 2452–2461.

Pfeifer, K., Arcangioli, B., & Guarente, L. (1987) *Cell 49*, 9–18.

Proudfoot, N. J., & Brownlee, G. G. (1976) *Nature 263*, 211–214.

Reinert, W. R., & Marzluf, G. A. (1975) *Arch. Biochem. Biophys. 166*, 565–574.

Reinert, W. R., Patel, V. B., & Giles, N. H. (1981) *Mol. Cell. Biol. 1*, 829–835.

Roberts, A. N., Berlin, V., Hager, K. M., & Yanofsky, C. (1988) *Mol. Cell. Biol. 8*, 2411–2418.
Sanger, F., Nicklen, S., & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.
Shaw, G., & Kamen, R. (1986) *Cell 46*, 659–667.
Studier, F. W., & Moffatt, B. A. (1986) *J. Mol. Biol. 189*, 113–130.
Tsai, S. F., Martin, D. I., Zon, L. I., E'Andrea, A. D., Wong, G. G., & Orkin, W. H. (1989) *Nature 339*, 446–451.
Ullmann, A. (1984) *Gene 29*, 27–31.

Sequential ^{1}H NMR Assignments and Secondary Structure of the B Domain of Staphylococcal Protein A: Structural Changes between the Free B Domain in Solution and the Fc-Bound B Domain in Crystal†

Hidetaka Torigoe,‡,§ Ichio Shimada,‡ Akiko Saito,∥ Moriyuki Sato,∥ and Yoji Arata*,‡

Faculty of Pharmaceutical Sciences, University of Tokyo, Bunkyo-ku, Hongo, Tokyo 113, Japan, and Tokyo Research Laboratories, Kyowa Hakko Kogyo Co., Ltd., Asahimachi 3-6-6, Machida, Tokyo 194, Japan

Received April 16, 1990; Revised Manuscript Received June 6, 1990

ABSTRACT: The recombinant B domain (FB) of staphylococcal protein A, which specifically binds to the Fc portion of immunoglobulin G (IgG), has been investigated with the use of two-dimensional proton nuclear magnetic resonance spectroscopy. All backbone and side-chain proton resonances of FB (60 amino acid residues), except the amide proton resonance of Ala2, were assigned by the sequential assignment procedures by using double-quantum-filtered correlated spectroscopy (DQF-COSY), homonuclear Hartmann–Hahn spectroscopy (HOHAHA), and nuclear Overhauser enhancement spectroscopy (NOESY). On the basis of the NOESY data, three helical regions, Glu9–His19, Glu25–Asp37, and Ser42–Ala55, were identified in the free FB in solution. Existence of two of the three helical regions, Glu9–His19 and Glu25–Asp37, is consistent with the X-ray crystallographic structure of the Fc-bound FB [Deisenhofer, J. (1981) *Biochemistry 20*, 2361–2370]. By contrast, in the Fc-bound FB as revealed by the X-ray analysis, the Ser42–Glu48 segment is extended and no structural information has been available in the Ala49–Ala55 segment. We suggest that a significant conformation change is induced in the C-terminal region of FB when it is bound to the Fc portion of IgG.

Protein A is a cell wall component of *Staphylococcus aureus*. It binds specifically to the Fc portion of immunoglobulin G (IgG)[1] from various mammalian species (Langone, 1982). The primary structure was determined by the analyses of the amino acid sequence of the trypsin-digested fragment of protein A (Hjelm et al., 1975; Sjodahl, 1976, 1977a,b) and of the nucleotide sequence of the gene coding for protein A (Uhlen et al., 1984). The N-terminal part of the mature protein consists of a tandem of five highly homologous domains designated as E, D, A, B, and C, starting from the N-terminal (Sjodahl, 1977a; Uhlen et al., 1984). Each segment, which contains about 60 amino acid residues, can bind to the Fc region (Sjodahl, 1977a; Moks et al., 1986). The C-terminal part is a cell wall binding domain designated as X (Sjodahl, 1977b). This segment does not bind to the Fc region and contains approximately 180 amino acid residues (Sjodahl, 1977b).

The three-dimensional structure of the B domain (FB) bound to Fc has been solved by an X-ray crystallographic analysis of the complex between FB and the Fc fragment of human polyclonal IgG at a resolution of 2.8 Å (Deisenhofer, 1981). The electron density for the Fc-bound FB was observed for the segment from Phe6 to Glu48; no information was available for the Ala2–Lys5 and Ala49–Lys59 segments. Two antiparallel helical regions, Gln10–Leu18 and Glu26–Asp37, are the predominant elements of the secondary structure of the Fc-bound FB. The structure of FB in the free state has not been determined at the atomic level.

In the present study, we use two-dimensional proton nuclear magnetic resonance (NMR) to discuss the solution conformation of FB. A gene coding for FB was chemically synthesized, and a high-level expression system of the synthetic gene in *Escherichia coli* was established (Saito et al., 1989). The recombinant FB is a single-chain polypeptide with 60 amino acid residues (M_r 6770) containing two additional amino acids, Thr at the N-terminal and Ala at the C-terminal. Sequence-specific resonance assignments were achieved by the standard procedure (Wüthrich, 1986) and used to determine

† This research was supported in part by grants to Y.A. from the Ministry of Education, Science, and Culture of Japan (62870089 and 63430022) and from the Mitsubishi Foundation.

* To whom correspondence should be addressed.
‡ University of Tokyo.
§ Present address: Department of Physiology, Jichi Medical School, Minamikawachi-machi, Tochigi-ken 329-04, Japan.
∥ Kyowa Hakko Kogyo Co., Ltd.

[1] Abbreviations: $d_{\alpha N}(i,j)$, $d_{\beta N}(i,j)$, $d_{NN}(i,j)$, and $d_{\alpha\beta}(i,j)$, intramolecular distances between protons αCH and NH, βCH and NH, NH and NH, and αCH and βCH on residues i and j, respectively; DQF-COSY, double-quantum-filtered correlated spectroscopy; DSS, sodium 2,2-dimethyl-2-silapentane-5-sulfonate; FB, B domain of staphylococcal protein A; Fc, C-terminal half of the heavy chain of immunoglobulin; HOHAHA, homonuclear Hartmann–Hahn spectroscopy; HPLC, high-performance liquid chromatography; IgG, immunoglobulin G; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser enhancement spectroscopy; Tris-HCl, tris(hydroxymethyl)aminomethane hydrochloride.

0006-2960/90/0429-8787$02.50/0