Vol. 199, No. 1, 1994 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
February 28, 1994 Pages 1-10

THE *STREPTOCOCCUS PYOGENES* HYALURONAN SYNTHASE: SEQUENCE COMPARISON AND CONSERVATION AMONG VARIOUS GROUP A STRAINS

Paul L. DeAngelis, Ning Yang and Paul H. Weigel

Department of Human Biological Chemistry and Genetics
The University of Texas Medical Branch, Galveston, Texas 77555-0647

Received January 3, 1994

We recently cloned the hyaluronan synthase gene (*hasA*) from *Streptococcus pyogenes* (DeAngelis *et al.*, J. Biol. Chem. **268**, 19181, 1993). Since this is the first glycosaminoglycan synthase gene to be cloned and these enzymes have also not been purified, nothing is yet known at the molecular level about the similarity or relatedness of hyaluronan synthase to other gene products. We found several proteins in the sequence database with substantial similarities to hyaluronan synthase (HasA), including NodC from *Rhizobium*, DG42 from *Xenopus*, and the three chitin synthases (Chs) from *Saccharomyces*. Like HasA, NodC and the Chs proteins all have at least an N-acetylglucosaminyl transferase activity in common and are membrane-associated. The polymerase chain reaction was also used to show that HasA, and probably the two gene operon for hyaluronan biosynthesis (*hasA/hasB*), is highly conserved among Group A streptococcal strains. In nine strains, isolated from 1917 through 1993, only five silent mutations were detected. The results show that hyaluronan synthase is highly conserved within Group A strains, although another virulence factor, the M protein, varies considerably in these same strains.

Group A streptococci (GAS) are human pathogens responsible for pharyngitis, impetigo, and scarlet fever as well as sequelae such as rheumatic fever and glomerulonephritis. The bacteria utilize several virulence factors, such as the M protein and the hyaluronan (hyaluronic acid; HA) capsule, to aid in the evasion of the host primary defenses (1,2). We have recently identified and cloned the HA synthase gene (*hasA*) responsible for polymerizing HA, which is an alternating polysaccharide copolymer of D-glucuronic acid and N-acetyl-D-glucosamine (3,4). HasA is the first member of the family of glycosyltransferases that synthesize glycosaminoglycans to be cloned and described at the molecular level. Upon

0006-291X/94 $5.00

comparison to the protein database, we found several polypeptides with similar sequences. Especially noteworthy are the *nodC* genes from the bacterial genus *Rhizobium*, which are strikingly similar to *hasA*. The NodC protein is postulated to synthesize N-acetylglucosamine oligomers that form the backbone of nodulation factors (5,6). Here we elaborate on the similarities of the HasA protein to other proteins as well as describe the conservation of HasA and the operon for HA biosynthesis among GAS strains belonging to a variety of M serotypes.

METHODS

Materials: Media components were from Difco. The Sequenase kit and urea were from U.S. Biochemical. Taq polymerase and DNA modifying enzymes were from Promega. Bacterial strains are described in Table I. All other reagents were from Sigma unless noted.

Sequencing: Chromosomal DNA from various streptococcal strains was obtained by the method of Caparon and Scott (7) and used as the template for the polymerase chain reaction (PCR; 8). The *hasA* region (4) was amplified with primers flanking the open reading frame using the manufacturer's buffer supplemented with 1.5 mM $MgCl_2$. The 5' sense primer ending in the Met start codon of *hasA* was ATCTTGATTTATCTAAATATG. The 3' antisense primer downstream of *hasA*, which originated in the adjacent *hasB*, the gene encoding the enzyme UDP-glucose dehydrogenase, was CCAACATATCCTGATCC-AGC (4,9). The temperature cycle was: 95°C, 40 sec; 40°C, 30 sec; 72°C, 3 min in a Perkin-Elmer DNA Thermal Cycler machine. After 30 cycles, the reaction mixture was removed from the oil overlay, the amplified product selectively precipitated in 2.5 M ammonium acetate and 50% ethanol, and the pellet was washed with 70% ethanol (10). The PCR products from all of the Group A strains tested migrated equivalently on agarose gels with a size of 1.3 kb as predicted from our reported sequence (4; data not shown). The standard Sequenase method for sequencing plasmids (alkaline treatment, ethanol precipitation, slow annealing), with the modification of the inclusion of 0.6 μl of extension mix in the A and T reactions, was used to determine the internal sequence of the amplified DNA using a sense primer within the *hasA* gene starting at Ala184 (CCTTTGGGGTGGAGCGTGC).

Computation: PC/Gene 5.03 programs were used to search for homology [FSTPSCAN] in the Swiss Protein Database release No. 26, to align sequences [CLUSTAL or ALIGN] and to determine the hydropathy profiles [SOAP] of the various gene products. Prediction of membrane-associated regions was performed by the [HELIXMEM] and [RAOARGOS] programs.

RESULTS AND DISCUSSION

HasA, the enzyme that polymerizes HA was found to be very similar to the NodC proteins (19.5% to 30.6% identical) by ALIGN depending on the particular isolate of *Rhizobium* used for comparison. NodC is postulated to synthesize N-acetylglucosamine

oligomers, which requires recognition of UDP-GlcNAc and N-acetylglucosaminyl transferase activity, two functions also possessed by HasA. Additionally, a developmentally regulated *Xenopus laevis* protein of unknown function, DG42 (11), was also significantly similar to HasA (19.5% identity by ALIGN). Numerous residues in several regions among the HasA, NodC, and DG42 proteins are absolutely conserved, which is strongly suggestive of a similar functional role for these regions of the proteins (Fig. 1). Furthermore, HasA and NodC are very similar with respect to their predicted hydrophobicity profiles (Fig. 2). Both proteins have at least one membrane-associating domain near the amino terminus as well as a cluster of three domains near the carboxyl terminus that are predicted to be membrane-spanning regions. The hydropathic profiles are in agreement with the cellular localization of both enzymes at the membrane (12,13).

A unique feature of HA synthesis, in comparison to other glycosaminoglycans, is that the growing polymer chain is extruded through the plasma membrane into the extracellular space. The mechanism of transport of their respective carbohydrate products to the exterior of the cell is not known, but NodC and HasA may perform that role as well as the saccharide polymerization reactions. Portions of the four hydrophobic regions are also conserved at the amino acid level between HasA and NodC (Fig. 1). This observation may be due to the requirements for transport of charged polysaccharides through a channel-like structure in the lipid bilayer. We predict, on the basis of hydropathy and sequence comparisons, that the intracellular catalytic domains reside in the central portion of HasA, while the domains in the vicinity of the N and C termini are involved in secretion/transport of HA through the membrane.

Since HasA is from a Gram-positive human pathogen and NodC is from a Gram-negative soil bacteria that colonizes plants, it is likely that the two enzymes were optimized independently for polysaccharide polymerization utilizing UDP-sugar nucleotides. The primary structures of NodC and HasA could be the result of convergent evolution to a common function. This hypothesis is supported by another observation that a class of fungal

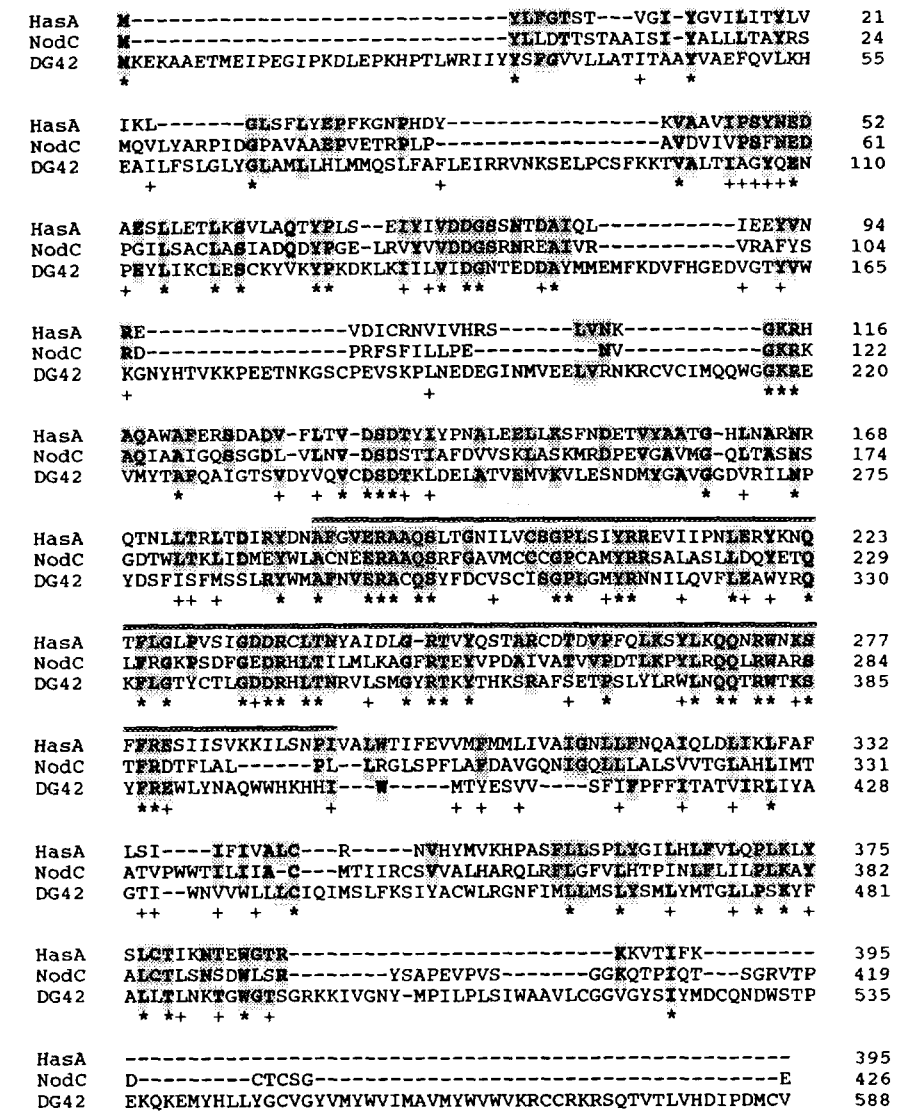

Figure 1. Sequence comparison of *hasA Streptococcus pyogenes*, *nodC* (*Rhizobium meliloti*), and DG42 (*Xenopus laevis*) gene products. The deduced amino acid sequences were aligned using CLUSTAL. Identical residues among all three sequences are marked with asterisks. Similar amino acids indicating conservative substitutions are denoted with plus signs. Residues in either DG42 or NodC that are identical to HasA are indicated by bold face and stiple. The region analyzed by PCR sequencing of various *Streptococcus pyogenes* strains as shown in Table I is indicated below the solid dark line. The following HasA sequences were scored by both HELIXMEM and RAOARGOS as membrane-associated: Val 9→Leu 29, Ile 294→Gly 314, Leu 326→Leu 341 and Ala 353→Gln 370. These membrane domains are indicated by the bars in figure 2A.

genes encoding chitin synthases, which have been noted to be very similar to NodC and DG42 in certain regions (14), also display similarity to HasA (Fig. 3). Three members of the yeast chitin synthase family, Chs1, Chs2, and Chs3, are 9.1%, 6.3%; and 9.6%, identical

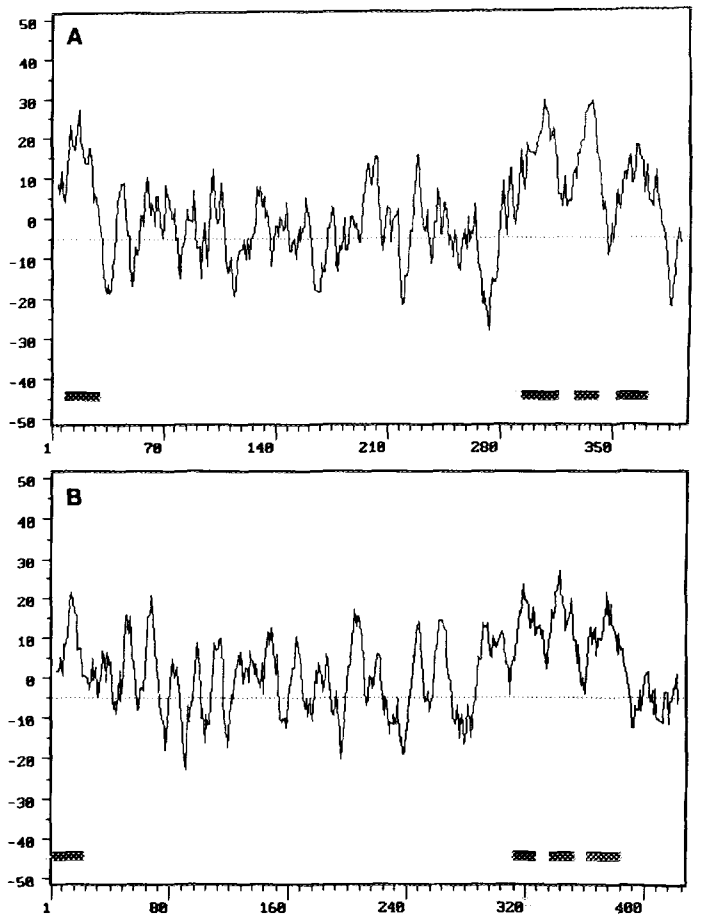

Figure 2. **Hydropathy plots of HasA from *Streptococcus pyogenes* and NodC from *Rhizobium meliloti*.** The hydropathic indices are on the Y-axis and amino acid numbers for HasA (1-395) and NodC (1-426) are on the X-axis. The two proteins display very similar profiles as well as locations for the predicted membrane-associated or transmembrane domains (marked with solid bars), which suggest similar membrane orientations. Panel A, HasA; Panel B, NodC.

respectively, to HasA as determined by ALIGN. At certain locations many residues of the chitin synthases are either identical or semiconserved with HasA.

An area of the *hasA* gene (4) that was very similar to the *nodC* gene sequence (coding for Thr195→Ile294) was compared among a variety of GAS isolates in our collection. When these nucleotide regions were examined by sequencing PCR products, no alteration of the HasA primary sequence was observed (Table I). Only several silent mutations were detected at the nucleotide level. This finding is quite remarkable due to the wide variety

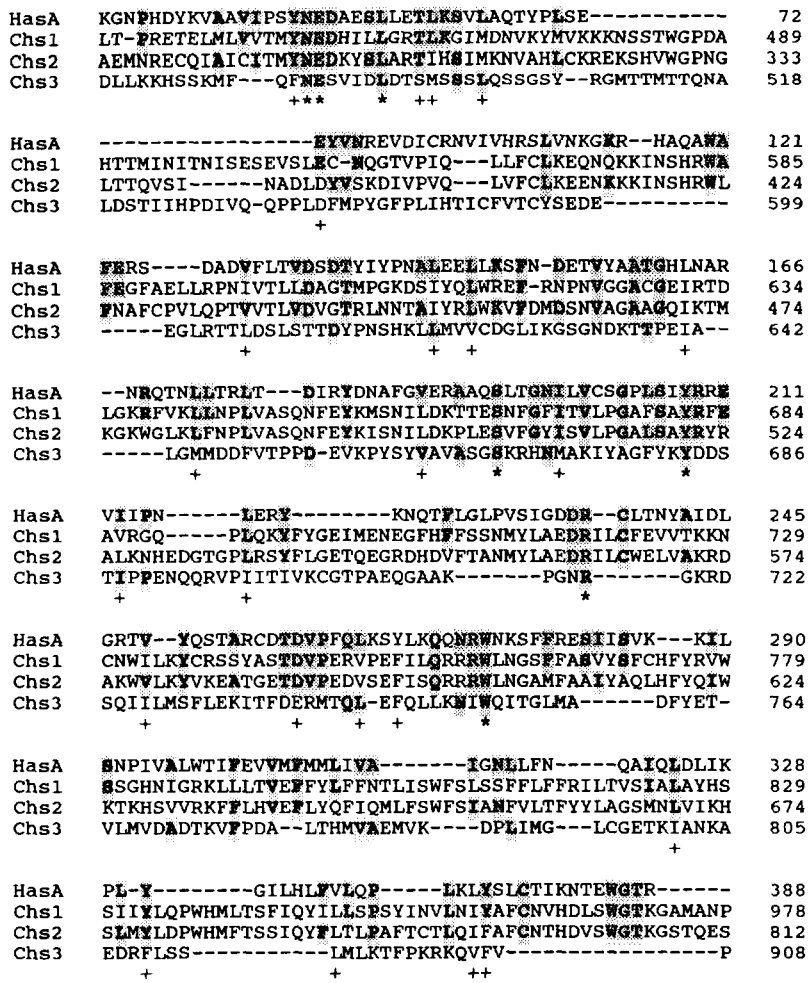

Figure 3. Sequence comparison of streptococcal HasA to *Saccharomyces cerevisiae* chitin synthases 1, 2, and 3. The predicted amino acid sequences were aligned using CLUSTAL. Asterisks indicate identical residues in all four proteins. Plus signs indicate sites of conservative substitutions. Residues in chitin synthase (Chs) 1, 2 or 3 that are identical to HasA are in bold face and stiple. Not all of the four protein sequences are shown.

of M serotypes among the strains tested. Additionally, the various strains also originate from throughout North America, as well as one strain from New Zealand (NZ131). Furthermore, the bacteria have been continually growing in their human hosts, the only known mammalian reservoir of GAS, for decades without changing these regions of the HA synthase. The primary sequence of this central region of HasA in strain S43, isolated in 1917, is the same as in strains isolated in the 1930's and 1950's, as well as 1993. This conservation of sequence within the HA synthase is in contrast to the observations of great

Table I

Conservation of the hyaluronan synthase sequence (from Thr195→Ile294) between *Streptococcus pyogenes* strain S43[a] and other Group A strains

Strain	M-Type	Changes	Source
12344	1	Val1286 codon GTT→GTC	ATCC
12385	4	No change	ATCC
Manfredo	5	No change	J.C. Reitmeyer (UTMB)[b]
JRS4	6	No change	M.G. Caparon (Washington U., St. Louis)
Vaughn	24	Cys201 codon TGC→TGT	J.C. Reitmeyer (UTMB)[b]
NZ131	49	No change	J.J. Ferretti (U. of Oklahoma)
NSA156	49	No change	J.C. Reitmeyer (UTMB)[b]
DW1009	u[c]	Val1286 codon GTT→GTC	D.E. Wennerstrom (U. of Arkansas)
55 1320	u	Ile214 codon ATT→ATC Asp235 codon GAT→GAC	J.C. Reitmeyer (UTMB)

[a] S43 was obtained from the Rockefeller Collection and is M-type 6.
[b] Originally obtained from E. Whitnack (U. of Tennessee).
[c] u denotes an undetermined M-type.

heterogeneity in the gene for the M protein, *emm*, which is subject to both point substitution and insertion mutations (15,16). A possible rationale for the observation of *hasA* homology is that bacterial HA, the polysaccharide product of HasA, is not very immunogenic, since it is identical structurally to the human endogenous HA. Therefore, the synthase has not been put under selective pressure to change, but rather to conserve strictly the biosynthetic apparatus. On the other hand, the M protein is very antigenic, and after the host develops an antibody response against this molecule, the bacteria are more readily recognized and destroyed. There is thus selective pressure to alter the structure of the M protein, which is likely to have been a major factor in generating the ≥ 80 serotypes of the M protein and at least that many *emm* alleles.

In addition to the high degree of conservation in the HA synthase, it should also be noted that all of the strains examined (Table I) also appear to have conserved the basic operon needed for HA biosynthesis. This operon consists of two tandem genes, *hasA* and *hasB* (4). The HasB protein is a UDP-glucose dehydrogenase (9), which is required for the synthesis of one of the two sugar nucleotide precursors for HA synthesis, UDP-glucuronic acid. Since all eubacterial cell walls contain N-acetylglucosamine, the second sugar nucleotide precursor, UDP-N-acetylglucosamine, is present in both Gram-positive and Gram-negative bacteria. We have observed that the HA synthase and the dehydrogenase are the only two gene products needed to allow virtually any bacteria to synthesize HA (3,4). Our initial PCR experiments used primers spanning the region between HasA Met1 and HasB Val11. Products of identical size were obtained from all the strains, indicating that the start of the *hasB* open reading frame is present in each strain at the same distance from the end of the HasA open reading frame (not shown). The results also suggest that the nucleotides in the downstream primer corresponding to Ala6→Val11 of the UDP-glucose dehydrogenase from the assorted strains are identical to the sequence in strain S43, the strain from which we cloned these two genes (3,4). The entire primary sequence of the *hasB* gene of S43 (4) was identical to that reported for strain WF50 (9), except for two silent mutations.

Although Prehm and coworkers have reported the primary sequence of a putative HA synthase from the Group C *Streptococcus equisimilis* strain D181 (17), they have not demonstrated that the protein cloned has HA synthase activity. In contrast, our work with the Group A HasA has shown HA synthase activity for this protein by several approaches (3,4). The two gene or protein sequences are not homologous. In ongoing studies using various oligonucleotide primers derived from the S43 *hasA* locus (4) to amplify D181 chromosomal DNA, we have obtained products that differ somewhat in size from those found with S43 DNA. However, when we used the same sequencing primer described in this report, we found that D181 possesses the identical nucleotide sequence as that seen for

the S43 central coding region of the hyaluronan synthase. Thus, Group C D181 DNA encodes a protein with at least 110 residues identical to the Group A HasA; this is highly likely to be the Group C *hasA* gene. Therefore, it appears that the HA synthases of Group A and Group C Streptococci are indeed similar, and that the Prehm group has pursued and characterized an unrelated molecule. Unfortunately, the same antibody used by Lansing *et al.* (17) to obtain their clone has been used erroneously as a probe for the eukaryotic HA synthase (18,19).

In conclusion, the amino acid residues of HasA that are identical or very similar in NodC and that are highly conserved among the various GAS isolates are likely to be important for the structure and/or function of the HA synthase. We also concur with previous speculation that DG42 may be a *Xenopus* hyaluronate synthase (20). Its sequence similarity to HasA, as well as the circumstantial linkage of DG42 expression patterns with increased levels of HA polysaccharide during morphogenesis make this hypothesis probable.

ACKNOWLEDGMENTS

We thank Sandy Capellen for help preparing the manuscript and the strain donors for their generous gifts. This research was supported by Grant 004952-11 from the Texas Higher Education Coordinating Board and National Institutes of Health Grant GM35978.

REFERENCES

1. Lancefield, R.C. (1962) *J. Immunol.*, **89**, 307-313.
2. Kass, E.H., and Seastone, C.V. (1944) *J. Exp. Med.*, **79**, 319-330.
3. DeAngelis, P.L., Papaconstantinou, J., and Weigel, P.H. (1993) *J. Biol. Chem.*, **268**, 14568-14571.
4. DeAngelis, P.L., Papaconstantinou, J., and Weigel, P.H. (1993) *J. Biol. Chem.*, **268**, 19181-19184.
5. Lerouge, P., Roche, P., Faucher, C., Maillet, F., Truchet, G., Promé, J.C., and Dénarié, J. (1990) *Nature*, **344**, 781-784.
6. Dénarié, J. and Cullimore, J. (1993) *Cell*, **74**, 951-954.
7. Caparon, M.G. and Scott, J.R. (1991) *Methods Enzymol.*, **204**, 556-586.
8. Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., and Erlich, H. (1986) *Cold Spring Harbor Symp. Quant. Biol.*, **51**, 263-273.
9. Dougherty, B.A. and van de Rijn, I. (1993) *J. Biol. Chem.* **268**, 7118-7124.
10. Landweber, L.F. and Kreitman, M. (1993) *Methods Enzymol.*, **218**, 17-26.

11. Rosa, F., Sargent, T.D., Rebbert, M.L., Michaels, G.S., Jamrich, M., Grunz, H., Jonas, E., Winkles, J.A., and Dawid, I.B., (1988) *Dev. Biol.*, **129**, 114-123.
12. Markovitz, A., Cifonelli, J.A., and Dorfman, A. (1959) *J. Biol. Chem.*, **234**, 2343-2350.
13. John, M., Schmidt, J., Wieneke, U., Kondorosi, E., Kondorosi, A., and Schell, J. (1985) *EMBO J.* **4**, 2425-2430.
14. Bulawa, C.E. (1992) *Molecular and Cell. Biol.*, **12**, 1764-1776.
15. Harbaugh, M.P., Podbielski, A., Hügl, S., and Cleary, P.P. (1993) *Molecular Microbiol.* **8**, 981-991.
16. Hollingshead, S.K., Readdy, T.L., Yung, D.L., and Bessen, D.E. (1993) *Molecular Microbiol.* **8**, 707-717.
17. Lansing, M., Lellig, S., Mausolf, A., Martini, I., Crescenzi, F., O'Regan, M., and Prehm, P. (1993) *Biochem. J.* **289**, 179-184.
18. Rittig, M., Lütjen-Drecoll, E., and Prehm, P. (1992) *Exp. Eye Res.*, **54**, 455-460.
19. Klewes, L., Turley, E.A., and Prehm, P. (1993) *Biochem. J.*, **290**, 791-795.
20. Atkinson, E.M., and Long, S.R. (1992) *Molecular Plant-Microbe Interactions* **5**, 439-442.