Molecular Biology of Lea Genes of Higher Plants

Progress Report

Dr. Leon S. Dure III

University of Georgia Research Foundation
Athens, Georgia 30602-4222

Covers Period From 1993 Through 1993

PREPARED FOR THE
U.S. DEPARTMENT OF ENERGY
UNDER CONTRACT DE-FG09-89ER14071

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

PROGRESS REPORT

Twenty months ago, we set out to begin the elucidation of the function of the D-7 Lea proteins by:

1. determining their 3D structure in various ionic environments;
2. determining the number of D-7 protein molecules per cotton seed;
3. determining the cell-type specificity (if any) in the expression of the D-7 genes.

Objectives 2 and 3 are not trivial considerations in the search for protein functions. The number of protein molecules per embryo and their cell type distribution could rule in/out possible functions. At present we speculate that the D-7, D-29 proteins function as ion-carriers in the dehydrated environment. Clearly, there should be many of these protein molecules in the embryo for such an idea to be tenable.

Objective 1 required large amounts of D-7 protein for the various biophysical assays of structure. Objectives 2 and 3 required the production of antisera monospecific for the D-7 proteins, since D-7 protein quantitation was to be accomplished via immunochemical methods and cell-type expression determined immunocytochemically. The production of this antisera also required purified D-7 protein. Aspects of objective 1 such as the computer modeling and energy minimizations utilizing the SYBYL and AMBER software programs and the biochemistry Iris mini computer could be carried out independently.

Against a background of these goals and requirements we have accomplished the following:

1. Completed sequencing a D-7 gene from cotton. This sequence was found to be identical to the coding sequence of the previously-obtained cDNA sequence. No introns in addition to the one already found in the partially sequenced gene was found.
2. Using pET-31 vectors, established transformants of *E. coli* that a) synthesize reasonable amounts of the D-7 protein, b) contain the D-7 sequence (cDNA) in the reverse orientation to serve as source for control material for protein purification and experiments involving antibodies.
3. Worked out a protocol for purifying to homogeneity the D-7 protein from *E. coli* crude extracts. D-7 has no known enzymatic activity and thus cannot be identified in extracts or fractions by assays of catalysis. Furthermore, it should be noted that of the 113 amino acids of D-7, there is but a single tyr and a single phe and no trp. Thus, D-7 cannot be followed during purification by absorbancy at wave lengths > 240 Å.

In addition, there are several abundant *E. coli* proteins with apparent molecular weights approximating D-7 which rule out following D-7 in purification via 1D SDS-PAGE. Its position on 2D gel electrophoresis (IEF/SDS-PAGE) has been our only means for working out the purification protocol.

The system used for D-7 synthesis in *E. coli* is based on the IPTG induction of the T3 RNA polymerase which, in turn, transcribes the D-7 (cDNA derived) coding sequence which is 3' to a T3 promoter. Only $\approx 0.5\%$ of the protein from induced *E. coli* crude extracts is D-7 (Figure 1). Furthermore, the induced cultures grow slowly. This we believe indicates that extensive synthesis of D-7 is inhibitory to the bacterium. (This is not unexpected, since D-7 presumably functions to allow for the survival of a severe loss of cellular water. The function itself may be deleterious to cell growth and division).

Figure 1. 2D gels of A: crude extract from *E. coli* induced to synthesize D-7 protein, B: crude extract from uninduced *E. coli*, C: purified D-7 protein. D-7 proteins in A are circled.

The purification is straightforward with $\approx 90-100\%$ yield up to the last step which is an electroelution from acrylamide gels. Yields from this step are pathetic, never reaching 30%. We continually modify conditions in hopes of improving the overall yield.

Note in Figure 1 that more than one D-7 "spot" is observed on 2D gels of induced *E. coli* crude extracts and in purified D-7. We think this signifies N-terminal modification of some of the molecules (N-blocked methionine, etc., that are routine phenomena for alien proteins produced in bacteria).

The identity of purified D-7 has been confirmed by amino acid sequencing of the purified fraction for the N terminal 20 residues.

4. Produced antibodies to the D-7 protein purified from *E. coli*. The antisera obtained which was carried through the routine $(\text{NH}_4)_2\text{SO}_4$ precipitation step, is of intermediate titer, and is monospecific for the D-7 proteins of cotton as shown by the immunoblot of Figure 2.

Figure 2. 2D gels of buffer soluble protein extracted from mature cotton embryos that have been A: reacted with anti-D7 IgG, B: stained with Coomassie Brilliant Blue. D-7 proteins are enclosed in an ellipse.

We have used this antibody for two purposes:

a) determining if expression of D-7 in cotton embryos is equivalent throughout all embryo tissues, more prevalent in some cell types than others, or confined to specific cell types. The procedure we have used for the immunocytochemical localization of D-7 protein in embryos of various ages has been the goat anti-rabbit alkaline phosphatase system, which has given good results. We found early on that the large number of protein bodies in these tissues interfere with techniques employing silver enhancement. The D-7 antibody preparation produces precipitated material throughout all tissues and all cell types of late-stage embryos, including both the cotyledon and axis epidermis. Embryos treated with pre-immune IgG show no precipitated material. From this we concluded that the function of the D-7 protein is independent of cell type in the embryo.

b) determining the precise number of molecules of D-7 in the late-stage cotton embryo. Initially, we attempted to quantitate the amounts of D-7 protein in extracts and partially purified fractions via the ELISA immunochemical procedure. After many attempts this was abandoned in favor of the "Laurell rocket" immunoelectrophoretic procedures (5). The high and variable background signals that plagued the ELISA assays are not a problem with rocket electrophoresis, in which minute samples containing D-7 antigen are electrophoresed into stationary anti-D-7 IgG. Rocket-appearing precipitates form as antibody-antigen concentrations reach equivalency, and the distance from point of origin to the tip of the rocket precipitate is proportional to the log of the concentration of antigen in the sample.

The results of these experiments are presented in Figure 3. (The right-hand portion of this figure, b, d and f, shows the results obtained for another Lea protein D-113

Figure 3

ABUNDANCE OF D-7, D-113 PROTEINS IN MATURE COTTON EMBRYO

	D-7	D-113
molecular weight ¹	14,600	17,500
μgrams/embryo	99	297
percent cytosolic protein ²	1.29	3.86
molecules/embryo	4.1×10^{15}	1×10^{16}
molecules/cell ³	5.4×10^8	1.34×10^9
nmoles/embryo	6.8	16.97
cytosolic molarity ⁴	113 μMolar	283 μMolar

¹based on known sequence of gene family member (Baker et al., 1988).

²based on soluble protein after high-speed centrifugation being 7.7 mg/mature embryo.

³based on embryo cell number being 7.6×10^6 (Walbot and Dure, 1976).

⁴molarity is that of embryo immediately prior to desiccation. Cytosolic aqueous volume at that point is based on a loss of wet weight of 60 mg/embryo during desiccation.

which is a completely different type of protein with presumably an entirely different function in embryogenesis/desiccation tolerance from D-7. It does, however, serve as a foil for results obtained with D-7.).

Panel a of Figure 3 shows the standard curve (left) used to measure the amount of D-7 in extracts of mature cotton embryos. This curve was established using three different preparations of purified D-7 protein. Some of the values obtained with the embryo extracts are given in this panel as well (right). The slopes of the standard curve and of the experimental values are parallel, which verifies the validity of these measurements. Panel b shows the rockets formed during a typical electrophoretic run. A standard curve was run with each experimental measurement (wells 1-7) to verify that all the parameters of the electrophoretic run (temperature, humidity above the agarose slab, etc.) were properly controlled. A 150-fold range of pure D-7 protein was used to obtain the standard curve (30-4500 ngrams). The experimental wells (8-11) contained replicates of 20 and 60 μ grams of an embryo extract.

Panel e shows the rise and fall of D-7 antigen in extracts from embryos of increasing size and maturity (wells 1-5), from mature embryos (well 6) and from embryos germinated 12, 24, 36 and 48 hours (wells 7-10). Well 11 contained an extract of 60 hr germinated seedlings that were slowly desiccated for an additional 6 hr. No D-7 protein is evident in this post-embryogenetic water-stressed material.

The summation of the results of these experiments is given in the table at the bottom of Figure 3. It shows that when the cell contents are averaged regardless of cell size or type, they contain $\approx$ 500 million molecules of D-7 protein each. This calculates to a molarity of 113 μ M in the embryo just prior to the loss of cellular water during the desiccation of the seed.

From the foregoing data, it would seem that the function of the D-7 protein requires a very large number of molecules per cell and that the function pertains to all cells regardless of cell type or function.

Most of the foregoing data is being assembled for publication (6).

5. Modeled D-7 and a 33 amino acid abstraction composed of three 11-mer consensus repeats (Figure 1, Appendix 2).

When D-7 was first sequenced, the existence of 5 tandemly arranged 11-mer repeats was noted in the molecule. As D-7 homologs from other species were sequenced, the 11-mer repeating unit was found to be their most shared feature. Comparison of D-7 homologs from 5 species allowed for a chemical consensus for the 11 positions to be constructed (Ref. 3 and Appendix 2). This consensus is remarkable in the fidelity of occurrence of certain amino acid residues in every position (save # 10) among all the species. Clearly the repeating unit constitutes the basis for the function of these proteins in that their molecular weights vary considerably. It is also almost certain that the repeating 11-mers exist as an α helix in solution.

When several repetitions of the 11-mer consensus are drawn in the helical wheel or helical net configuration, a stripe of methyl groups from ala and thr is observed moving diagonally across the length of the helix. Such a surface generally indicates that two such apolar surfaces interact through Van der Waal and hydrophobic forces to form an element of 3D structure called a "coiled coil". This structural motif is the basis for most of the structure of proteins comprising intermediate filaments, myosin, tropomyosin, laminins and many other structural proteins. In the case of these animal coiled coils, the superhelical coil is left-handed, resulting from a repeating 7-mer unit in which positions 1 and 4 are apolar.

In the case of the 11-mer unit, positions 1, 2, 5 and 9 are apolar, and a coiled coil dimer formed from the interaction of two of these units is a right-handed superhelix. Figure 4 illustrates these two different types of coils.

Figure 4. Diagrams of helical nets showing the interaction of apolar surfaces to form the coiled coil dimer. 7-mer shows interaction between the surfaces of helices containing the heptad repeat resulting in a left-handed coil. The 11-mer shows the helical surface interactions that result in a ~~left~~-handed coiled coil.

The notion that D-7 may exist as a coiled coil is attractive in that it puts to use the apolar surface of what is a very hydrophilic protein overall, and creates a structure that bristles with charged residues throughout the surfaces not involved in dimerization. However, with such a structure, a right-handed coiled coil formed from a pair of right-handed α helices, violates the handedness rule. This rule, based on supercoiled proteins characterized to date, predicts that the supercoil should be of the opposite handedness from the helix of the monomeric unit, e.g. collagen trimeric super helices as well as the coiled coil of the heptad repeat.

Dealing conceptually and mechanistically with this problem has monopolized our modeling efforts of the past months. Appendix 2 gives a full characterization of the 11-mer units and develops a possible new element of structure -- a right-handed coiled coil of right-handed α helices.

Our expenditure of time and effort on the 3D structure of repeating 11-mer units may be questioned. However, dimerization of the 33 residue abstraction and, thus, D-7 and its homologs, creates a molecule with remarkable binding sites for phosphate ions and their positive counter ions, an idea that will be explored in the Proposed Research section. But of paramount importance is the fact that we have preliminary evidence that D-7 does exist in solution as a dimer.

6. Carried out preliminary measurements of D-7 molecular weight via sephadex chromatography.

The molecular weight of the D-7 protein synthesized by *E. coli* is 12,055 daltons. Its apparent size (denatured) on SDS-PAGE is 15,300 daltons. Its apparent size on chromatography on Sephacryl S100 HR is 35,-39,500 daltons in either 0.05 M phosphate buffer, pH 7.0 or phosphate buffer containing 0.5 M NaCl. The column eluates had to be monitored by SDS-PAGE because of the very low absorbancy of D-7. The sephadex-derived values would seem to indicate a trimeric molecule until it is recognized that sephadex chromatography measures Stokes radii. Should D-7 dimerize through the 11-mer repeat helical region, the long rod-like region formed would produce a highly asymmetric molecule with an unusually large Stokes radius. Thus, until further experiments are performed, we tentatively believe that D-7 exists as a dimer in phosphate buffer. The choice of phosphate buffer appears now to be unfortunate in view of more recent modeling. This is explained in the Proposed Research.

PROPOSED RESEARCH

As stated previously, we hope to determine the function of a subset of Lea proteins, typified by cotton Lea D-7, by determining their aqueous 3D structures, their absolute amounts per cell in embryo tissues and relative cell type distribution among the various cell types of the embryo. The last two tasks are almost complete. Thus, the 3D structure determinations contribute to the bulk of the research proposed here.

The experimental tools for accomplishing this are:

D-7 protein produced in *E. coli*

D-29 protein produced in *E. coli*

39 amino acid peptide produced in *E. coli*

33 amino acid peptide used in the computer molecular mechanics and dynamics programs (this is an abstract computer entity, not a tangible molecule).

The D-7 protein is being produced in *E. coli* on a routine basis and is being used as standards in the rocket immunoelectrophoresis measurements of D-7 extant in embryo tissue during embryogenesis germination. It is also being used in making these measurements in water stressed leaves. A computer-derived view of the *E. coli* product is shown in Figure 5. The 5x repeated 11-mer region is shown in the conventional α helix, the remainder is a random coil.

Bacterial Synthesis of Lea D-29

When it was recognized that the cotton Lea protein D-29 is composed principally of the tandemly repeated 11-mer unit, albeit without the stringent requirement for thr and ala in the apolar positions, it became obvious that the 3D structure of this protein should be included in this research. It became more obvious when the sequence of the protein pcC 3-06 from *Craterostigma plantagineum* (resurrection plant) was published. This latter protein is a member of a group of proteins that increase markedly in the leaves of this plant as they begin to desiccate. Analyzing its sequence has revealed that it, too, is composed principally of the 11-mer repeating unit, but, like D-29, it is not stringent for ala and thr in the apolar positions.

It should be noted that the size of these proteins are considerably different. D-7 has 113 amino acids, D-29,302, and the resurrection plant protein has 201. Furthermore, the D-7 homologs from other plants show a broad size range (Appendix 2). This suggests that their function is not strongly dependent on molecular weight, but perhaps on the total number of 11-mer repeats per cell.

We plan to produce D-29 in *E. coli*. Figure 6 shows this molecule in which the 11-mer containing regions are in the conventional α helix. Where hinge regions occur, the molecule is random coil as are the N and C terminal non-11-mer regions. In hopes of increasing synthesis over that of D-7 and of avoiding the losses in purification that are associated with electroelution from gel matrices, we plan to utilize the pMAL expression system (8). Here, the D-29 cDNA is inserted downstream from the product of the bacterial malE gene which

codes for the maltose-binding protein. This insertion interrupts a lacZa gene which allows for a blue-to-white screening of recombinants in the presence of IPTG.

In-frame insertion should produce a maltose-binding protein-D-29 fusion protein which can be purified to near homogeneity by means of affinity chromatography through an amylose column. D-29 can be cleaved from the fusion protein by the Factor Xa protease site occurring immediately upstream from the insertion site. Re-passage of this mixture through the amylose column binds the maltose-binding protein moiety and recovery of D-29 in the void volume. The apparent ease of purification makes this system particularly attractive since there are numerous proteins in *E. coli* crude extracts with molecular weights and isoelectric points similar to the cotton D-29.

Should these protocols prove recalcitrant, we plan to use the system developed by Hoffman-LaRoche (Invitrogen "Xpress System"). Here the Lea sequences would be fused 3' to a coding sequence that contains a translation start site, a poly-his stretch and an enterokinase cleavage site contained in a proprietorial vector. After expression induction, the bacterial cell lysate is passed over an affinity resin that sequesters the poly-his sector of the fusion protein by binding to Ni^{2+} . The Lea proteins are released from the resin by digestion with the protease.

Bacterial Synthesis of a 39 Residue Artificial Protein

A great deal has been learned in the past few years about protein structure through the use of small "designer" peptides (9, 10, 11). Peptides synthesized chemically are still extremely expensive, considering the amounts needed to carry out the proposed biophysical experiments (see below). We plan to attempt the synthesis of a 39 residue peptide via the pMAL system outlined above. This peptide will contain 3 tandem repeats of the D-7 11-mer consensus sequence flanked by 3 N and C terminal residues. Using blunt-end ligation for insertion into the expression vector, this will require synthesis of a 120 bp DNA (including the stop codon). This can be done by coupling smaller fragments synthesized by the Molecular Genetics Instrumentation facility at the University of Georgia.

Very precise CD spectra should be obtainable with this molecule and it is especially attractive for NMR analysis.

As in the case of recombinant D-29 synthesis, should the pMAL expression system fail, we shall fall back on the "Xpress" system.

Biophysical Experiments

With these tools we plan the following procedures that in aggregate should establish the predominate 3D structure of D-7, D-29, and of the repeating 11-mer unit itself. Since a conceivable function for the D-7/D-29 proteins is that of an ion carrier in periods of very high ionic concentrations, all the 2 and 3D structural studies must be performed in a range of ionic strengths and with a range of likely ions (K_2HPO_4 , K_2NO_3 , etc.).

1. Determination of intrinsic molecular weight of D-7, D-29 by analytical ultracentrifugation. Fortunately, the Beckman "Model E" centrifuge of the Biochemistry Department is still maintained in operating status. Schlieren optics will be used for D-7 in view of its weak absorbancy. Values obtained here will yield a value for Stokes radius when compared with sephadex-derived data. This value should confirm the existence of a highly asymmetric molecule and confirm/deny D-7 holo-protein as a dimer -- at least in certain ionic environments. Speculation about the aggregation status of D-29 is premature.

A second method can be used to reinforce the data obtained from sedimentation analysis. It is based on "pore exclusion" or "migration limit" electrophoresis of undenatured proteins. In this case, acrylamide gradient gels of very high acrylamide concentration are used. In the initial report of this technique (14) gels of 4-24% acrylamide were used; we plan to use 12-26% gradients because of the relatively small size of these Lea proteins. In this procedure, proteins are allowed to migrate in the gel until they reach an acrylamide pore size where further migration becomes severely reduced. The initial speed with which proteins migrate is, of course, a function of their net charge at the pH of electrophoresis. Since they are not coated with SDS molecules, this speed can be slow. This usually requires electrophoresis for up to 24 hours for small proteins.

Since the position of D-7 and D-29 on 2D gels is in very good agreement with their calculated isoelectric points based on amino acid composition, the pH of electrophoresis is an easy decision. Determining the apparent size of these undenatured, pure Lea proteins by comparison with iso-electrophoresed standards should indicate homoaggregation if it occurs.

2. Obtain CD spectra for D-7, D-29 and the 39-mer. These spectra should indicate the extent to which these molecules are in the α helical configuration. Both the structure predictive algorithms of Chou and Fasman (12) and Argos (13) predict that the 11-mer repeating units are helical and with very high scores. Other portions of D-7 and D-29 may be helical also. These measures cannot distinguish between 3.6 and 3.667 pitched helices. These analyses will be carried out on the CD spectrometer of Dr. Michael Johnson of the Chemistry Department.
3. Analyze the structure of D-7 and the 39-mer by NMR spectroscopy. This is a technology beyond our capabilities. However, we are fortunate to have the collaboration of Dr. Herman van Halbeek of the Complex Carbohydrate Research Center and Biochemistry Department of The University of Georgia to carry out and interpret NMR measurements (see attached letter of collaboration). D-7 is small enough to be amenable to this analysis and the 39-mer is an ideal model compound. NOSEY spectra require 2 mM protein in 0.5 ml minimally. This translates to 6 mg of D-7 protein. Thus, it is a labor-intensive set of experiments, but they have the capability of distinguishing between the 3.6 and 3.667 pitched helices and also provide information on ions that may be bound in various ionic environments.

4. Carry out equilibrium dialysis measurements with D-7, D-29 and the 39-mer. We have alluded to the possibility that the 11-mer based proteins function to bind ions as seeds desiccate. This idea will be developed further in the next section. The idea can be tested by measuring the capability of the proteins to sequester ions from an adjoining compartment separated by a membrane permeable to ions but not to proteins. Hopefully, the 39-mer will be confined by standard membranes having a cut-off value of 3500 daltons.

The strength of ion binding, initially potassium phosphate and potassium sulfate, can be ascertained using $^{32}\text{PO}_4$ and $^{35}\text{SO}_4$ and Scatchard plots of the experimental data. Dissociations constants for phosphate, sulfate and other ions should be possible.

Modeling the 33-mer as a carrier of phosphate

Among the problems that cells encounter as they lose substantial amounts of water would seem to be that of handling the increasing ionic strength. The possibility for precipitation and crystallization seems likely. Furthermore, such precipitates and crystals may be deleterious to cell structures and functions. Phosphate salts in particular could be a problem, yet it would seem necessary for seeds to carry enough phosphate to provide for early seedling growth. Much of this is carried in the form of phytic acid (inositol hexaphosphate); crystals of which are found ubiquitously in seeds. The phosphate stored in these molecules is only slowly mobilized during seed germination. In cotton, the phytate fraction (easily seen as a layer in centrifugation pellets) is not depleted until the fourth day of germination at 25°C. Both solubilization of the phytate crystals and enzymatic hydrolysis of the phosphate is required. Thus, there would seem to be a need for readily available phosphate ions for the early events of germination and, further, a requirement that this phosphate not exist as a slowly soluble precipitate.

This argument is after-the-fact on our part. What caught our attention initially was the remarkable binding sites for phosphate (and additionally perhaps sulfate) that are created when two molecules of the 33-mer model dimerize in a parallel, in-phase fashion. The sites are formed by the hydroxyl of position 1 thr and the amine of position 8 lys of the 11-mer of one strand and the position 6 lys amine of the other strand. These thr and lys residues can be seen clearly in Fig. 4C of Appendix 2. For every two 11-mer repeats joined as a dimer (one 11-mer from each strand), 2 such binding sites exist. Hence, a dimer of the 33-mer would have 6 sites; a dimer of D-7, which contains five 11-mer repeats per peptide chain, would contain 10 sites.

Another feature of the 11-mer unit is that binding sites for the phosphate counter ions (K^+ or Na^+) exist along the surface of the helix opposite the apolar surface. These sites are the carboxylates of glu in positions 7 and 11 of the consensus sequence and can be seen in Fig. 4C (Appendix 2) on the right and left edges of the structure. The overall stoichiometry for totally binding K_2HPO_4 is remarkable.

Figures 7 and 8 show phosphate ions placed in the binding sites of the 33-mer. In Figure 7, which shows a hypothetical dimer from the N-terminus, the phosphorous atom is green, oxygens are red, nitrogens dark blue and protons light blue. The amines of the 2 lys and

the hydroxyl of thr can be seen to form a pocket that accommodates the 3 unprotonated oxygens of the phosphate. Figure 8 is another view of the phosphate-dimer interaction showing the multiple binding sites. In neither of the figures have the molecules been energy minimized, i.e., the phosphate has been positioned by the computer operator.

A recent example of phosphates binding across monomeric peptides by electrostatic bridging to form oligomeric structures is that of the honey bee protein melittin (15). In this case phosphate bridges have a profound stabilizing effect on the oligomer.

The foregoing provides the rationale for much of the modeling we propose to carry out. We now wish to characterize by energy minimizations the following:

1. The 33-mer in the conventional right-handed α helix.
2. The 33-mer parallel, in-phase dimer in right-handed helices induced to form a right-handed coiled coil.
3. The 33-mer dimer with phosphates in the 6 binding pockets and K^+ bound to the glu of positions 7 and 11.

Energy minimization via the molecular mechanics program AMBER 5.5 examines the placement of each atom and maneuvers it so as to find the lowest free energy position, considering bond angles, bond lengths, and the field forces of adjacent atoms that impinge on this position. The influence of hydration water can be included. What is not considered is the influence of entropy. The values obtained are in terms of energy loss or gain as structures are formed, but do not represent an intrinsic ΔG . However, through comparisons of alternative states (i.e., α helix vs. random coil or monomer vs. dimer), the relative thermodynamic likelihood of a given structure existing can be arrived at. Hopefully, the thermodynamic driving force for ion binding and the effect of ion binding on dimerization can be measured.

At this stage small model molecules, e.g., the 33-mer, can be handled by our software/hardware. Because of the enormous number of values that must be handled, larger molecules are not readily amenable to this treatment.

Summation

Out of all the foregoing, we hope to rule in/out the following possibilities:

1. That D-7 and perhaps D-29 exist principally as α helices which in high ionic strength exist as dimers.
2. That D-7 upon dimerization forms a right-handed super helix -- a heretofore unobserved structure.
3. That D-7 and D-29 bind phosphate and/or other ions with dissociation constants low enough for this binding to be meaningful in lowering the ionic environment.

4. That the molar concentrations of D-7 and D-29 in the cytosol of embryo cells is sufficient for the ion sequestration to be meaningful in preventing ion precipitation or crystallization.

As always, we anticipate surprises, but whatever the outcome of this proposed research, we should know more about how the properties of these proteins relate to desiccation tolerance/water stress survival in plants.

REFERENCES

1. Dure, L. III, Greenway, S., and Galau, G. A. (1981). *Biochemistry* 20:4162.
2. Dure, L. III, Crouch, M., Harada, J., Ho, T.-H.D., Mundy, J., Quatrano, R., Thomas, T., and Sung, R. (1989). *Plant Mol. Biol.* 12:475.
3. Baker, J., Steele, C., and Dure, L. III (1988). *Plant Mol. Biol.* 11:277.
4. Galau, G. A., Hughes, D. W. and Dure, L. III (1986). *Plant Mol. Biol.* 7:155.
5. Laurell, C. B. (1966). *Anal. Chem.* 15:45.
6. Roberts, J. K. and Dure, L. III. In preparation.
7. Wampler, J. E. and Dure, L. III. In preparation.
8. Riggs, P. (1990). In Ausubel, F. M. *et al.* (eds.), *Current Protocols in Molecular Biology*. Greene Association/Wiley Interscience, New York.
9. Kaiser, E. T. (1987). In *Protein Engineering*, pp. 193-199. Alan R. Liss, Inc., New York.
10. DeGrado, W. F., Wasserman, Z. R. and Lear, J. D. (1988). *Science* 243:622
11. Hecht, M. H., Richardson, J. S., Richardson, D. C. and Ogden, R. C. (1990). *Science* 249:884.
12. Chou, P. Y. and Fasman, G. D. (1978). *Adv. Eng.* 47:45.
13. Argos, P. (1989). In *Editbase* (computer program). Purdue Research Foundation and USDA-ARS.
14. Anderson, L.O., Borg, H. and Mikaelsson, M. (1972). *FEBS Lett.* 20:199.
15. Wilcox, W. and Eisenberg, D. (1992). *Protein Science* 1:641.

FIGURE LEGENDS

- Figure 5.** Computer generated view of the D-7 protein synthesized in *E. coli*. N-terminal is at right. 11-mer containing domain is shown in an α helix.
- Figure 6.** Computer generated view of the D-29 protein. N-terminal is at right. 11-mer containing domains are in α helices.
- Figure 7.** Computer generated view of two 33-mer abstract peptides, dimerized in phase through their apolar surfaces (yellow). The N-terminal of both peptides is at top. A phosphate ion is shown bound to two lys and one thr at the top. See text for details.
- Figure 8.** Another view of Figure 7. The N-terminals are at bottom.