GEORGIA INSTITUTE OF TECHNOLOGY OFFICE OF CONTRACT ADMINISTRATION

# PROJECT ADMINISTRATION DATA SHEET



[X] ORIGINAL [ ] REVISION NO. _____

Project No./(Center No.) G-33-622 (R6018-2A0) GTRC/GIT DATE 5 /25 /87

Project Director: Dr. Fred L. Suddath School/Lab Chemistry

Sponsor: DHHS/PHS/NIH/National Institute of General Medical Sciences

Agreement No.: Grant No. 5 R01-GM36610-03

Award Period: From 5/1/87 To 4/30/88 (Performance) 7/31/88 Reports

Sponsor Amount:

| | New With This Change | Total to Date |
|---|---|---|
| Contract Value: | $ | $ 114,344 |
| Funded: | $ | $ 114,344 |

Cost Sharing No./(Center No.) _____ Cost Sharing: $ _____

Title: High Resolution Structure of Pisum Sativum Lectin

## ADMINISTRATIVE DATA

OCA Contact E. Faith Gleason x4-4820

1) Sponsor Technical Contact:

Dr. John Norvell (301) 496-7260
NIGMS
National Institutes of Health
Westwood Bldg, Rm 936
5333 Westbard Avenue
Bethesda, MD 20892

2) Sponsor Issuing Office:

Jeff Carow (301) 496-7166
Grants Management Office
NIGMS, National Institutes of Health
Westwood Bldg., Rm 936
5333 Westbard Avenue
Bethesda, MD 20892

Military Security Classification: NA
(or) Company/Industrial Proprietary: NA

ONR Resident Rep. is ACO: _____ Yes X No
Defense Priority Rating: NA

## RESTRICTIONS

See Attached NIH Supplemental Information Sheet for Additional Requirements.

Travel: Foreign travel must have prior approval — Contact OCA in each case. Domestic travel requires sponsor approval where total will exceed greater of $500 or 125% of approved proposal budget category.

Equipment: Title vests with GIT. No equipment may be purchased in the last 6 months of the NIH Grant project period.

## COMMENTS:

Includes Subcontract to the University of Alabama.



## COPIES TO:

SPONSOR'S I.D. NO. 02.108.001.87.026

| | | |
|---|---|---|
| Project Director | Procurement/GTRI Supply Services | GTRC |
| Research Administrative Network | Research Security Services | Library |
| Research Property Management | Contract Support Div.(OCA)(2) Pat | Project File |
| Accounting | Research Communications | Other C. Stanford |

GEORGIA INSTITUTE OF TECHNOLOGY OFFICE OF CONTRACT ADMINISTRATION

SPONSORED PROJECT TERMINATION/CLOSEOUT SHEET

Date 8/5/88

Project No. G-33-622 R6018-2A0 School/~~Lab~~ Chemistry

Includes Subproject No.(s) N/A

Project Director(s) Fred L. Suddath GTRC/~~GIT~~

Sponsor DHHS/PHS/NIH/National Institute of General Medical Sciences

Title High Resolution Structure of Pisum Sativum Lectin

Effective Completion Date: 7/31/88 (Performance) 7/31/88 (Reports)

Grant/Contract Closeout Actions Remaining:

- [ ] None
- [x] Final Invoice or Copy of Last Invoice Serving as Final SF269
- [ ] Release and Assignment
- [ ] Final Report of Inventions and/or Subcontract: Patent and Subcontract Questionnaire sent to Project Director [ ]
- [ ] Govt. Property Inventory & Related Certificate
- [ ] Classified Material Certificate
- [ ] Other

Continues Project No. G-33-684 Continued by Project No. Not Assigned at this time-

COPIES TO:

Project Director
Research Administrative Network
Research Property Management
Accounting
Procurement/~~GTRI Supply Services~~
~~Research Security Services~~
Reports Coordinator (OCA)
~~Program Administration Division~~
Contract Support Division (2)

~~Facilities Management (EOB)~~
Library
GTRC
Project File
Other

ConA peptide sequence is the result of post-translational ligation of the initial amino and carboxyl termini as well as post-translational cleavage to create the new termini. [52] This ligation would occur at a position originally thought to be a site of proteolytic cleavage in ConA that may now be interpreted as a partial failure of ligation. Bowles *et.al.*[53] have recently shown by metabolic labeling of immature jackbean cotyledons, with $^{14}C$-amino acids, that ConA is indeed synthesized as a precursor peptide with the chain topology identical to PSA and favin. After folding it is then processed first by a proteolytic cleavage into β- and α-chains that are subsequently ligated to form the observed ConA sequence.

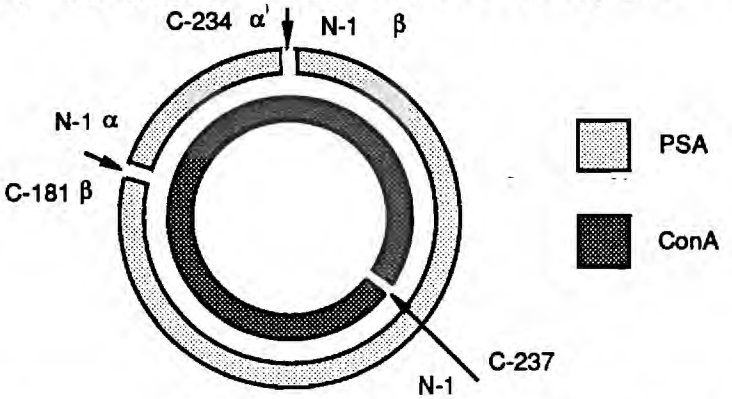



**Fig.1 A schematic representation of the circularly permuted amino acid sequences of PSA and ConA showing differences in chain termini.** The two chains of PSA (α and β) are represented by a circular band. A concentric *stippled* band represents the single chain of ConA positioned to maximize sequence homology with PSA.

The unusual biochemical complexities of these related lectins, differences in sequence permutation, differences in quaternary structure, and subtle differences in carbohydrate-binding specificity, can now be examined in terms of detailed molecular structures. The specific interaction of proteins and complex carbohydrates is equally important to the interactions of protein and nucleic acid. These interactions are of critical importance in cell-cell adhesion, cell-virus interaction, cellular recognition, and other cellular regulation processes. The information that is currently available concerning simple monosaccharide-protein interaction, favin-glucose, E.coli sugar binding proteins, and enzymes of carbohydrate metabolism, is essential, however, the lectin-carbohydrate and lectin-peptide-linked-carbohydrate structures will provide a more complete understanding of this important interaction.

The structure of pea lectin has been determined at 1.9 Å resolution, an extension of the results reported at 6.0 and 3.0 Å resolution.[54,55] A report of the structure of PSA at 5.0-Å has been published by a group at the USSR Academy of Sciences in Moscow.[56]

## Personnel

1. Dr. F.L. Suddath, Principal Investigator, 5/1/85 - present 33%
2. Dr. Beth Parks, postdoctoral fellow, 3 months, 5/1/85 -7/30/85, 75%
3. Dr. Susan Phillips, postdoctoral fellow, 6/1/86 - present 100%
4. Mr. Thomas Prasthofer, graduate research assistant, 5/1/85 - present 100% (UAB)

## C. Progress Report and Preliminary Studies - May 1, 1985 to July 1, 1988

The objectives of this project as stated 3 years ago have been partially completed. Additional objectives, not considered at that time, have also been accomplished. Items **1-5** below, summarize the progress in the areas specifically mentioned in the previous applications *Specific Aims.* Items **6-8** are other areas of significant progress that were not considered at the time of the previous review.

1. The structure of pea lectin (PSA) has been determined at 3.0 Å resolution and the refinement of the structure has progressed to the stage of 1.9 Å with good stereochemistry and a crystallographic

R=0.20. Much structural information remains to be realized from the current coordinates and we are in the process of analyzing these structural details.

2. Due to this laboratory being moved from UAB to Georgia Tech in the fall of 1985 and the move of John Helliwell from the University of Keele to the University of York during that same time period we have not collected the high resolution (1.2 Å) native data at Daresbury as we had planned. The native film data to 1.83 Å resolution has been rescanned-reprocessed and it is this data (to 1.9 Å) that is being used in the PSA refinement. **Diffraction data to at least 1.4 Å has been observed from crystals (0.3mm) of native PSA at room temperature on our area detector (June '88). Larger crystals are expected to allow us to collect higher resolution data, possibly to the limit observed using synchrotron radiation (1.2 Å).**

3. At 3.0 Å resolution the similarity between PSA and ConA is striking in spite of the unprecedented circular permutation of the primary sequence. A preliminary account of these similarities is described in the enclosed paper (appendix A). Much needs to be done in comparing the similarities and differences between the A and B monomers of the PSA dimer. Many interesting differences in the details of the two-fold (noncrystallographic) related monomers are obvious at 1.9 Å. We expect to add an addendum to this application, before review, that will summarize our refinement results at 1.9 Å.

4. In "collaboration" with J.P. Carver, Karl Hardman and Jim Rini of the University of Toronto the structure of pea lectin complexed with 3,6-di-O-(α-D-mannopyranosyl)-α-D-mannopyranoside has been solved at 2.5 Å resolution. The carbohydrate-lectin crystals were prepared in Toronto, data was collected in our laboratory (UAB) and the structure was solved in Toronto by molecular replacement methods using our native pea lectin coordinates as a search model.

5. No attempt has been made to refine the structure of the heavy atom derivatives of pea lectin as was proposed. These studies have been put aside due to more important tasks and a lack of man power.

6. Over the past 2.5 years we have been collaborating with Dr. Jeffery Engler at the University of Alabama at Birmingham (UAB) on the expression of pea lectin in *E.coli*. **The pea lectin has been expressed in E.coli as a functional molecule and crystals ($P2_12_12_1$, a=65.1, b=73.0, c=109.1) have been obtained that diffract to at least (2.8 Å).**

7. A number of interesting lectins related to PSA have been crystallized (see Table 1) and important collaborative contacts have been established with Dr. Richard Cummings of the University of Georgia and Dr. Alex McPherson University of California, Riverside.

8. Although of less scientific interest, but of considerable practical importance, substantial progress has been made in establishing the laboratory at Georgia Tech. During the first 9 months the laboratory physical facilities were completely renovated including benches, utilities and storage areas. The computer room and graphics room that houses the VAX 11/780 and the Evans and Sutherland PS330, respectively, were air conditioned and remodeled. Hardware and software are operational at this point and are being heavily used for the refinement and refitting of PSA native. The wet laboratory facilities are fully operational including centrifuges, electrophoresis, cold boxes, chromatography, HPLC, crystallization facilities, etc.
**During the winter and spring of this year a data collection facility was established at Georgia Tech. This facility is based on a Rigaku RU-200 X-ray source, Huber goniostat, μVAX II/GPX workstation, and a single San Diego Multiwire Systems area detector. The system is operated by the UCSD software and modeled after that configuration in most respects. The facility was operational as of June of 1988, approximately 9 months later than originally planned.**

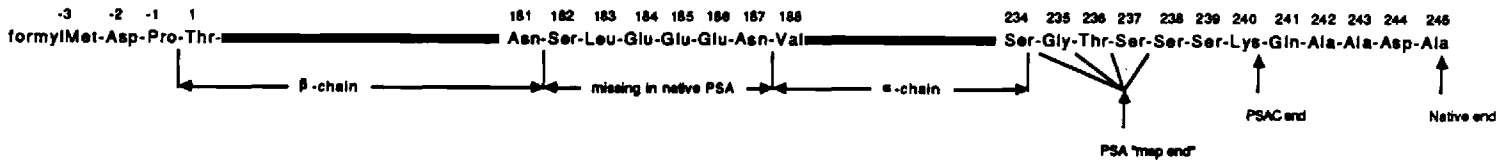
## C.I. The Structure of PSA

***The Pea Lectin Monomer:*** Pea lectin (PSA) crystallizes in the space group $P2_12_12_1$ with a=50.73 Å, b=61.16 Å, c=136.59 Å with a complete dimer of 49,740 daltons in the asymmetric unit. Each subunit of the dimer contains an α-chain of 7000 daltons and a β-chain of 17,000 daltons (see table 1). The structure of PSA was solved initially at 6.0 Å resolution using SIRAS phases[57] and later extended to 3.0 Å by MIR and map modification.[58] The model has been refined using a combination of restrained least squares methods developed by Konnert and Hendrickson[59] and a molecular graphics system consisting of an Evans and Sutherland PS330 connected to a VAX 11/780 running the program "FRODO" version 6.4.[60] We have recently begun to evaluate the results of this refinement on the graphics unit.

A discovery in late winter, that the Barry Finzel modification of the Hendrickson/Konnert refinement program, "PROFFT", apparently has some problems with refinement of the atomic B's as higher resolution data is added (2.1-1.9 Å) has been of some concern for us. The observation that we made was a number of main chain atoms refined to very low temperature factors ( 2.0) while other atoms seemed to have reasonable B's, the average being ~18 for all atoms. Also, when it appeared, by most criteria, we were near convergence the calculated shifts to the atomic B's were very large (>25). These large shifts persisted even with drastic damping of the B shifts. Extensive testing to assure that a proper linear scale factor was being used and a recheck of the film data processing/scaling all failed to correct the problem. The calculation of $F_{calc}$ to $F_{obs}$ anisotropic scale factors indicated an isotropic scale factor was sufficient[61]. Several helpful conversations with Barry Finzel indicated that others had experienced a similar problem and he suggested we run a few cycles of refinement using the non-FFT version of the "PROLSQ" program. Much to our delight this corrected the problem but also made us suspicious of our model since we had been using the FFT version of "PROLSQ" almost since refinement began. We restarted the refinement with the 2.48 Å coordinates and have continued to use the non-FFT version. A problem with this approach is the computer time per cycle increases from 24 min. to 31 hr. on a VAX 11/780.

At present we have a model with excellent stereochemistry (RMS delta in bond lengths of 0.025) and overall R factor of 0.201 (0.16 at lowest resolution and 0.24 at highest). Recently computed "OMIT", $2F_o$-$F_c$, and $F_o$-$F_c$ maps indicate a region of 3 residues, in both monomers, that is weak and possibly disordered. A number of new solvent molecules are obvious in these maps. At this resolution (1.9 Å) we are beginning to see alternate conformations of the side chains in several Thr residues. Our preliminary conclusion from this experience is that the FFT version of H/K is fine to ~2.5 Å but the non-FFT version should be used at higher resolution. This conclusion, however, is based on one case and a limited understanding, on our part, of the reasons for the FFT programs behavior. We now feel that refinement to the limit of our current resolution data (1.83 Å) will proceed more rapidly.

The cDNA complementary to pea lectin mRNA has been cloned and sequenced[62] and indicates that pea lectin is synthesized in pre-pro form. During processing to the mature form, a leader sequence is removed co-translationally and the resultant pro-lectin is cleaved post-translationally to yield the β- and α-chain. Apparently during the post-translational cleavage, several residues are removed from the linker region that connected the β- and α-chains. Protein sequencing results indicate that the amino-terminal residue of the α-chain is Val-188[63] while the carboxyl-terminal residue of the β-chain is Tyr-179.[64] The electron density in this region, however, suggests the presence of Pro-180 and possibly also Asn-181 in both monomers. These residues were not included in the model in early stages of refinement in an effort to avoid biasing our interpretation and the phases; however, density compatible with extra residues consistently returns in the maps. Prior to the latest cycles of refinement, Pro-180 was built into the model in both monomers as suggested by the density. Another possible post-translational modification is the removal of several amino acid residues from the carboxyl terminus of the α-chain. Protein sequencing results indicate that the α-chain terminates at Ser-237, Ser-238 or Ser 239 (C-terminal analysis did not distinguish the Ser position). The map indicates there is clear density for Ser-234. Past Ser-234 there is also weak but reasonably well connected density for about 3 more residues in both monomers. These residues are Gly-235, Thr-236 and Ser-237 based on the cDNA sequence and the possibility that this sequence is disordered will be investigated. The cDNA sequence indicates the precursor is terminated by Ala-245. The combination of C-terminal analysis and the electron density indicate extensive posttranslational modification has occurred at the C-terminus of the α-chain.




Throughout the refinement the A and B monomers in the asymmetric unit have been treated independently. At no stage has the noncrystallographic 2-fold symmetry been imposed on the model. In addition the two monomers were build independently into the initial 3.0 Å map by different individuals. In earlier stages of refinement, the electron density was weak in several areas of the model. These areas include residues 27-30, 54-58, and 75-77 in both monomers. These areas all consist of turns on the surface of the molecule. Residues 27-30 in both monomers were omitted from the refinement cycles and the subsequent electron density map calculation. The electron density around residues 27A-30A is still weak, where A's and B's are used to designate residues in the respective monomers. The connectivity of the main chain is good, but the side chain density is rather weak, especially for Glu-29A. In the B monomer, however, side chain density is much clearer. It was possible to fit all the side chains to the observed electron density. Electron density around residues 54A-58A is now very clear. Carbonyl positions are obvious, and all residues have good side chain density. In the B monomer, carbonyl positions are clear, but the density is still weak around Asp-54B. The region of electron density around residues 74A-77A is weak around the side chain of Asn-77A and Pro-74A. The corresponding region in the B monomer is very clear. Ramachandran plots for the two monomers (A and B) indicate there are 11 residues in each monomer that have torsion angles clearly outside normal values. Five of these residues are glycines. One residue is Asn-171, which is the fourth residue of a 4-residue turn. The geometry of this turn is similar to that of a class of 4-residue turns described by Sibanda and Thornton,[65] in which the conformation at the fourth residue, usually glycine or asparagine, commonly places it in the +/+ quadrant of the Ramachandran plot. Asn-39, Leu-101, Arg-133, and Asn-142 all appear to be clearly defined by the density. The equivalent residues to Leu-101 (Leu-230) and Asn-142 (Lys-30) in ConA also have positive $\phi$ values. There are no equivalent residues for Asn-39 and Arg-133 in ConA due to insertions or deletions.

Another unusual conformational feature in pea lectin is the presence of a *cis*-peptide bond between residues 80 and 81. The *cis* peptide bond is also present in ConA between residues 207 and 208 and is thought to be important for the $Ca^{++}$ binding site as described later. This *cis*-peptide region was omitted from early refinement, but was built into the model when the of 2Fo-Fc maps indicated it was present.

The structure of the pea lectin monomer, shown in Fig. 2, contains 226 amino acid residues. Beginning at the amino terminus of the β-chain, at the monomer-monomer interface, the polypeptide chain traces a convoluted pathway to form the two major and one minor antiparallel β-sheet structures that are the framework of the molecule. Along the way, extensive random coil region is overlayed onto the front face of the molecule, the chain breaks to end the β-chain and begin the α-chain at adjacent termini of the β-sheet in the back of the molecule, and the α-chain winds its way toward termination adjacent to the beginning of the β-chain. Altogether, there are six strands in the β-sheet in the back of the molecule; this sheet is referred to as sheet I and figures prominently in the architecture of the dimer (References to front and back of the pea lectin molecule are by analogy with the structure of ConA, where backs of dimers abut to form the tetramer.) The other major β-sheet II lies above the sheet I and is of seven strands, the outer two of which form the floor of the metal-binding region. The single crossover connection is right-handed. There are three strands in a smaller β-sheet seen in the lower left border of the molecule in Fig. 2. Two of the strands in this sheet are extensions of strands from the back sheet.

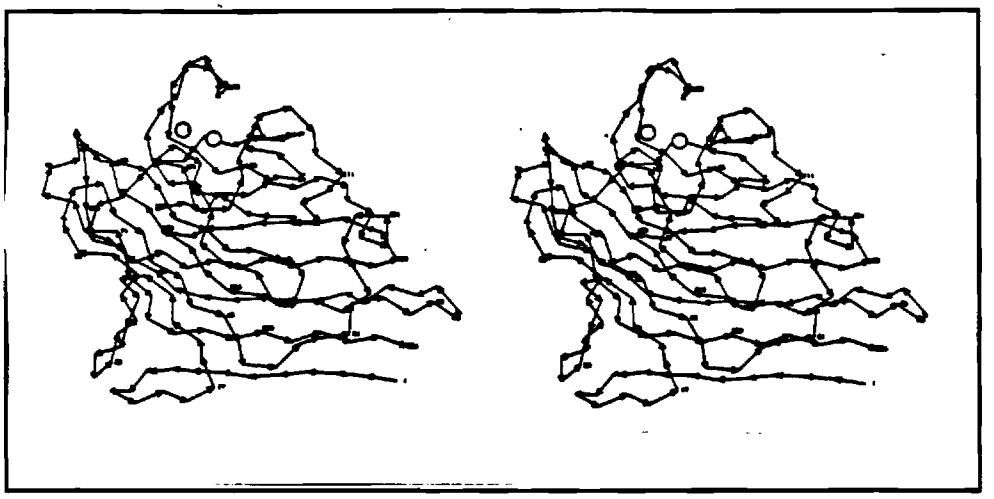


**Fig. 2. A stereo pair indicating the α-carbon positions of a PSA monomer.** Circles represent the position of the calcium and manganese ions in the metal binding loop.

***The Pea Lectin Dimer:*** The two monomers in the asymmetric unit are related by a molecular 2-fold axis to form the pea lectin dimer. These two monomers were modeled completely independently in an effort to identify structural differences permitted or imposed by crystal packing. The transformation that superimposes the B monomer onto the A monomer was calculated by least squares based on 226 equivalent α-carbon atoms plus the $Ca^{++}$ and $Mn^{++}$ ions and gave a root mean square difference in atomic positions of 0.45 Å. At only 10 of these positions are differences greater than 1.0 Å, with the largest difference being 3.8 Å for Val-188, the amino-terminal residue of the α-chain. Most of these large differences are located in terminal regions or loops at the surface of the structure.

The two pea lectin monomers of the dimer are joined such that the 6 stranded β-sheets of both monomers come together to form a continuous 12-stranded antiparallel β-sheet spanning the back of the dimer. Five main chain hydrogen bonds are formed between monomers at residues 1A and 9B, 3A and 7B, 5A and 5B, 3B and 7A, 1B and 9A, about the two-fold relating the monomers. Several other intermonomer contacts that may involve hydrogen bond formation, both side chain to side chain or side chain to main chain, are possible. These contacts include Thr-1A O to Thr-9B Oγ, Asp-54A Oδ1 to Lys-10B Nζ, Glu-2A Oε1 to Gln-15B Nε2, Trp-206A Nε2 to Asn 17B-Oδ1, Ser-48A Oγ to Tyr-46B OH, Tyr-46A OH to Ser-48B Oγ, and Asp-14A O to Trp-206B Nε. Apart from the interaction between residues 46A and 48B, these latter contacts do not reflect the molecular 2-fold symmetry. The details of these interactions are being studied at 1.9 Å resolution. The most striking overall feature of the protein-protein interface is the absence of extensive hydrophobic contacts. Most of the residues are neutral polar amino acids such as Asn, Gln, Ser, Thr and the hydrophilic "edges" of Tyr and Trp. Experiments are underway to determine influence of hydrogen bond breaking reagents like urea and guanidinium-HCl on the association of the monomers. (see draft paper in Appendix A)

***Comparison of Pea Lectin and ConA Tertiary and Quaternary Structure:*** As can be seen in Fig. 3, the structures of pea lectin and ConA are very similar, and both molecules have essentially the same overall β-sheet structure. However, the structures of the two molecules differ in several of the loops between β-strands and also in the regions of the carboxyl and amino termini. In the loops that differ between the two molecules, there are either insertions or deletions in pea lectin with respect to ConA In addition, the amino acid sequence of ConA is circularly permuted with respect to pea lectin and the ConA monomer consists of a single polypeptide chain, while the pea lectin monomer consists of α- and β-chains. Thus pea lectin and ConA have different carboxyl and amino termini and as expected these regions differ in the two molecules.

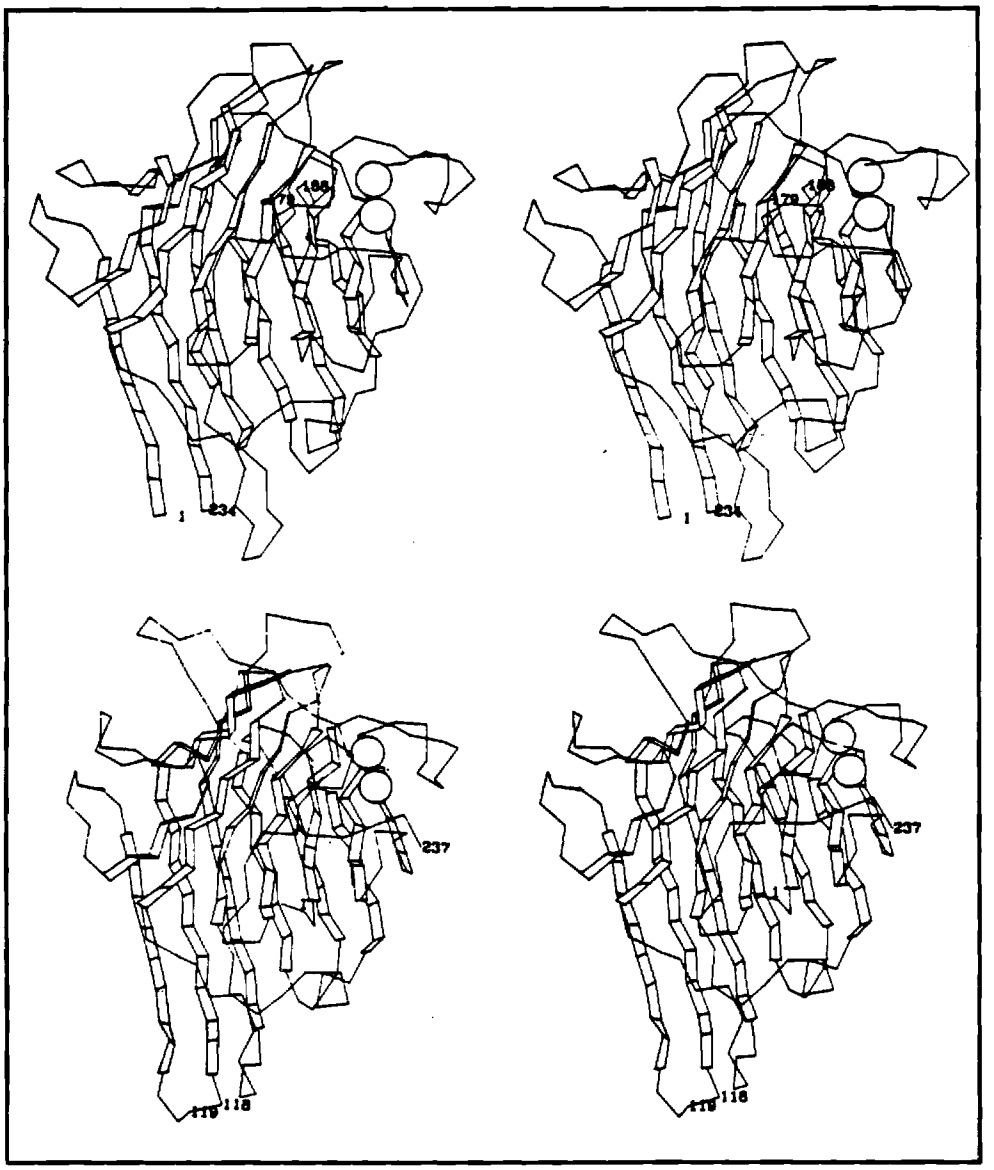



**Fig. 3. Stereo pairs of PSA monomer (top) and a ConA monomer (bottom).** Chain termini are numbered, as are the positions on ConA where ligation occurs.

It should be noted that all sites suggested to be subjects of post-translational modification lie on the surface of the pea lectin and ConA molecules easily accessible to enzymatic action. In particular, the proposed site of ligation in ConA, in the loop at about position 120, is firmly embedded in β-sheet I. In pea lectin, the corresponding site, at residues 1 and 234, is not connected by a peptide bond. The close correspondence between pea lectin and ConA structures in this region suggests that, were the ConA chains not ligated, the structure would nevertheless be very similar to that seen with termini lying in close proximity and in a not unfavorable orientation for ligation.

Pea lectin and ConA differ in that the ConA dimers associate to form tetramers at physiological pH, while pea lectin dimers do not. The 12-stranded β-sheets of the ConA dimers are twisted such that the two dimers are slightly wrapped around each other in the tetramer. This twist is different in the pea lectin dimer. It is possible that the change in twist is in part responsible for the lack of success using ConA as a search model in molecular replacement experiments on PSA, PSA-$(Man)_3$ complex and favin. It will be interesting to see if the twists of the 12-stranded β-sheets in lentil lectin, which also does not form tetramers, are closer to that in pea lectin than to that in ConA.

***Metal-binding Regions:*** The protein ligands to the $Ca^{++}$ and $Mn^{++}$ ions are conserved in the two proteins. Glu-119, Asp-121, and Asp-129 are unidentate ligands of the $Mn^{++}$ ion in pea lectin as are Glu-8, Asp-10, and Asp-19 in ConA. The fourth protein ligand to $Mn^{++}$ is provided by His-136 in pea lectin and His-24 in ConA. Asp-121 appears to be a bidentate ligand to $Ca^{++}$ in pea lectin as is Asp-10 in ConA. Phe-123 provides a carbonyl oxygen ligand to $Ca^{++}$ in pea lectin where Tyr-12 provides this same function in ConA. The remaining two protein ligands to $Ca^{++}$ in pea lectin are provided by Asn-125 and Asp-129, Asp-129 serving to bridge the metal ions. In ConA, these ligands are provided by Asn-14 and Asp-19, Asp-19 also bridging the metal ions.

In ConA, there are two water ligands for $Mn^{++}$ and two water ligands for $Ca^{++}$. In ConA, one of the water ligands to $Ca^{++}$ is also hydrogen-bonded to Oδ2 and O of Asp-208. The proximity of these two protein atoms to the binding site is a result of the *cis*-peptide bond between residues 207 and 208. In pea lectin, an equivalent *cis*-peptide is located between residues 80 and 80 and the carbonyl and carboxylate oxygen atoms of Asp-81 are at the appropriate distances from the $Ca^{++}$ ion for H-bonding to a bridging water molecule. Prior to the latest refinement cycles, two water ligands were added to the $Ca^{++}$ ion in the A monomer. One water molecule is hydrogen bonded to both the $Ca^{++}$ and a side chain oxygen of Asp-81. The other water molecule bridges between the $Ca^{++}$ and the carbonyl oxygen of Gly-99. In the B monomer, only the water molecule which bridges from $Ca^{++}$ to Asp-81 has been built into the model. No water molecules have yet been added to the $Mn^{++}$.

***Carbohydrate Binding Site:*** The structure of an α-methyl-D-mannoside ConA complex has been determined to 6-Å resolution.[66,67] The use of iodinated sugars permitted the location of the α-methyl-D-mannose binding site in the vicinity of the side chains of residues Tyr-12, Tyr-100, Asp-16, and Asp-208, some 10-14 Å from the $Mn^{++}$ position. This structure has been used along with the refined coordinates and binding specificity data to build a model with computer graphics to better define the orientation of the mannose in the binding site. The modeling studies suggest that residues Asn-14, Gly-98, Leu-99, Ser-168, Asp-208, Thr-226, and Arg-228 are involved in binding α-methyl-D-mannoside.[68,69] Of these residues, only 3, Asn-125 (Asn-14 of ConA), Gly-216 (Gly-98 of ConA) and Asp-81 (Asp 208 of ConA), are conserved in pea lectin.

Studies with a-methyl-D-mannoside-soaked pea lectin crystals have been successful, and x-ray diffraction patterns show them to tolerate soaking in concentrations of sugar as high as 3mM for 24 h. Data have been collected from these crystals, to 6.0 Å resolution, on a borrowed diffractometer under less than ideal conditions. The difference Fourier calculated was very noisy and although density could be found near logical surface sites on both monomers there also were noise peaks as high and higher than the putative monosaccharides. Data from these crystals will be recollected to higher resolution and under more suitable conditions this fall.

A pea lectin complex with a trimannoside has been crystallized in Toronto[70] and data were collected to 2.5-Å resolution in our laboratory in the summer of 1984. The structure would not yield to solution by molecular replacement methods, in Toronto, using ConA as a model; however, in the winter of 1985 we provided the Toronto group with preliminary coordinates of the PSA dimer (2.48 Å, R=28%) that immediately gave a solution. The PSA coordinates were used as a starting point for model building and refinement. Initial information indicated that the carbohydrate binds in a location similar to that in ConA. **All attempts to obtain more information about the structure from Dr. Jeremy Carver have failed.** We have **not** resolved the impasse with Dr. Carver over the past 18 months regarding the structure of pea lectin-trimannose complex determine from **our data** and **our coordinates.**

Since it appears that we will not have access to the coordinates of the lectin-complex, we have started a completely in-house solution of carbohydrate-pea lectin complexes. The first step in the process is to grow large crystals of pea lectin without trimannose in the same cell as the complex ($P2_12_12_1$, a=64.3 Å, b=73.4 Å and c=108.3 Å). This crystalline form has been prepared and characterized. This form of pea lectin has been reported earlier[71] and will be referred to as form II. Form II has been prepared in the presence of 10 fold excess of mannose and may prove to be the better form of pea lectin for carbohydrate-pea lectin complexes. Crystal soaking experiments with form II and oligosaccharides provided by Richard Cummings are underway. Our intentions are to collect native data from form II crystals and solve the structure by molecular replacement using the structure derived from form I. A preliminary analysis of the packing diagram of form I crystals suggest the binding site for carbohydrate is sterically hindered by symmetry related molecules (see appendix A). **The availability of large**

**crystals (0.5 x 0.5 x 4mm) of form II pea lectin should provide an ideal subject for investigation of various carbohydrate-pea lectin complexes. The trisaccharide, 3,6-di-O-(α-D-mannopyranosyl)-α-D-mannopyranoside, has been prepared by Richard Cummings and is pure by HPLC. Enough of this material is available to saturate several tens of milligrams of PSA.**

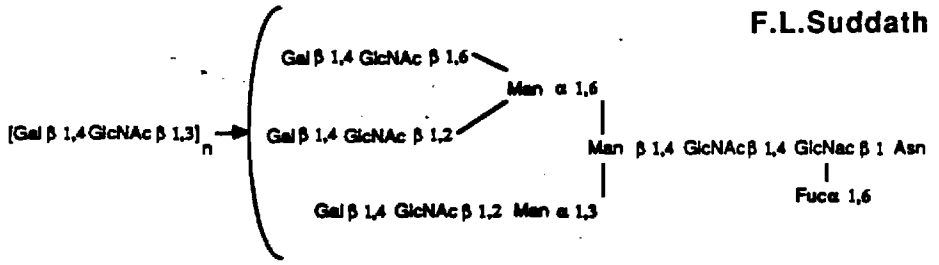
The crystals of recombinant PSA (PSAC) are of space group $P2_12_12_1$, a=65.1, b=73.0, c=109.1, and diffract to at least 2.8 Å. These crystals were obtained by hanging drop methods from 20mg/ml of PSAC at pH 5.1 using polyethylene glycol 3350 as the precipitating agent. The PSAC crystals were characterized on our area detector using very small crystals (0.08x0.1x0.15mm). Native data has not been collected from these crystals to date. We are waiting for crystals from a new batch of PSAC to grow larger. We currently have crystals (~5) that are 0.1x0.15x0.2mm in size. Careful measurement of these and other crystals indicate they are growing slowly and if they continue to grow at the present rate we should have crystals of reasonable dimension in early July (0.3mm smallest dimension). The cell parameters are almost identical with those found for form II of wild type PSA and PSA trimannoside crystals. The structure of PSA trimannose has been solved. The thesis of Jim Rini, University of Toronto, has sufficient information, we believe, to position the PSA structure in the PSAC or PSA form II cell. We anticipation that a few cycles of "CORELS" refinement treating the PSA form I structure as a rigid body against the PSAC native F's will result in a reasonable starting model for the interpretation of the PSAC structure. Model building to add the 6 additional amino acids between the β and α chains and the 3 "extra" residues at the N terminus of the β chain should be straightforward.

**C.II. Related Lectins**

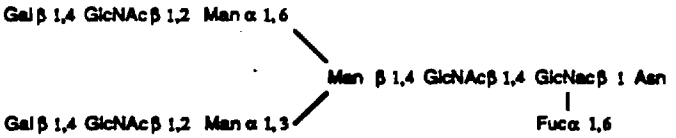
Several other lectin are under investigation because they are closely related to PSA or have some inherently interesting properties similar to PSA. The first protein in this category is the lectin from the common lentil. This lectin has a high degree of primary sequence homology with PSA (90% identity) and binds similar carbohydrates. Crystals of lentil lectin (LCA) suitable for high resolution diffraction studies have been prepared in this laboratory and others. This lectin is of particular interest because the C-terminal 20 amino acids (Glu-160 to Pro-180 in PSA) of the β-chain are reported to be missing from lentil lectin[72]. Since the sequence homology is so high it seems quite unlikely that the structure of LCA would be very different from PSA, yet the 20 amino acids that are supposedly missing constitute a β-strand-loop-β-strand in the 12 stranded β-sheet at the back of PSA dimer (Fig 3). These strands and the large β-sheet constitute the major structural framework of PSA as well as ConA and favin. We suspect the 20 amino acids could have been missed had there been an additional posttranslational cleavage at Gly-159 and the peptide was lost during isolation and purification for sequencing. On the other hand, if the peptide is missing the structural variation would be <u>very</u> interesting. Since the "missing" peptide (160-180) constitutes the $4^{th}$ and $5^{th}$ strand of the back anti-parallel β-sheet, it is a strand-loop-strand. If this element is removed from PSA β-sheet I, the $6^{th}$ strand (the last in the sheet) would have the right direction, in the N-terminal to C-terminal sense, to maintain a 4 strand anti-parallel sheet by moving over two strands and terminating the major sheets just as it does in PSA. The loop prior to the $6^{th}$ strand appears to have sufficient flexibility to accommodate this movement. The outcome of this putative rearrangement on the packing between the major β-sheet I and the inner β-sheet II is difficult to assess. An intriguing consequence of this speculation is that it would leave the $6^{th}$ strand, which constitutes the N-terminus of the α-chain (Val-188) juxtaposed to the C-terminus of the β-chain in a position primed for ligation.

Until recently, the inhibition of lectin-induced agglutination by various mono- and disaccharides was commonly used to define the carbohydrate-binding specificities of lectins. Whereas ConA, favin, PSA and LCA are all frequently classified as mannose-specific lectins, it is increasingly apparent that lectins can distinguish much more complex structural features of glycoconjugates than just single sugar residues. The carbohydrate specificity differences between PSA and LCA are subtle and interesting. PSA interacts with high affinity with certain complex-type N-linked oligosaccharides containing a fucose residue in the core.[73] Structures **I-III** are examples.

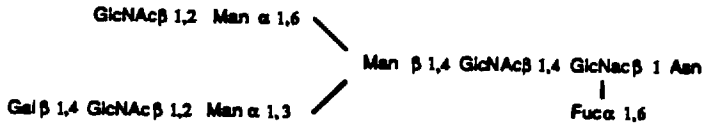
**Structure I**




**Structure II**




**Structure III**




Although a fucose residue attached 1,6 to the GlcNac closest to the peptide is required for high-affinity ($10^{-6}$ M) interactions with these N-linked oligosaccharides, fucose itself is **not** effective as a competing sugar in dissociating the complex. This suggests that the fucose residue is important in maintaining the optimal conformation, for binding, of other residue in the chain. The addition of a single 1,6 linked fucose to the GlcNac residue increases the binding constant by a factor of 50.[74] PSA also has a higher affinity for glycopeptides containing terminal mannose residues in addition to the core fucose residue than for those lacking terminal mannose residues. The nature of the linkage of oligosaccharide to peptide is important for interaction of oligosaccharides with PSA. The lectin will not bind, with high affinity, to free oligosaccharides with or without a reducing terminus and not attached to peptide even though the free oligosaccharide may contain a residue of core fucose.[75]

In contrast LCA has a higher affinity for glycopeptides with terminal N-acetylglucosamine residues rather than terminal galactose or mannose residues. LCA is also different from PSA in its ability to interact with free oligosaccharides **not** attached to peptide. LCA will interact with high affinity with free oligosaccharides if the oligosaccharides contain both the appropriate structures including a residue of fucose and a reducing terminus. After the reduction of the oligosaccharides reducing termini LCA will bind only weakly ($10^{-3}$).

A summary of these observations and many others suggest that lectins have extended binding surfaces with exquisite specificity. As a result the classification and general notion of lectins binding mono- and disaccharides is misleading in trying to understand the interactions at the molecular level. The specificity and affinity described by Quiocho[76] for the arabinose-binding protein is not manifest in the lectin at the monosaccharide level but possibly at the complex carbohydrate level.

Lectins from several other sources have been crystallized in this laboratory including:

1. The winged bean basic lectin, material supplied by Professor Kazuo Iwai of Kyoto University, Faculty of Agriculture. Not characterized but large, 0.8 mm crystals.

2. Two lectins isolated by Professor Debkumar Basu, Neurochemistry Division, Sree Chitra Tirunal Institute for Medical Sciences and Technology, from the jack fruit, have been crystallized. One of these lectins is described in Table 1, the other is a β-GalNAc specific lectin of 39,000 daltons that has not been characterized but yields large 0.6 mm crystals.

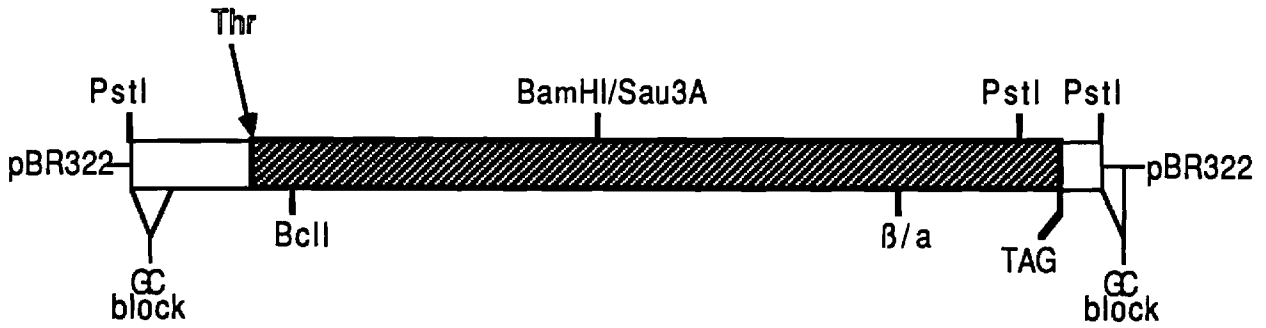
3. The lectin sainfoin has produced small irregular needles from material supplied by Dr. Ken Hapner of Montana State University.

4. Five lectins from *saphoria japonica* have been crystallized by Dr. Alex McPherson. Three of these crystalline lectins, described in Table 1, have been supplied to us.

Of the lectins mentioned above two are of particular interest. The *saphoria japonica* lectins are interesting because they are isolated from different plant tissue (leaf, bark, and seeds) and appear to be similar to PSA. The jack fruit lectin is interesting because it is composed of 4 small (10,500 dalton) identical subunits yet displays only 2 carbohydrate binding sites per tetramer. The carbohydrate specificity of this lectin is unique for the β-D-Gal(1-3)D-GalNAc of the T-antigen.

**C.III: Recombinant PSA**

Native pea lectin is formed by several posttranslational cleavages of a pre-pro form of the protein. Several modifications were made to the sequence of the gene so that the recombinant protein synthesized in *E.coli.* would resemble native pea lectin as closely as possible. In order to perform these modifications, the coding region was first cloned into the pUC 119 plasmid which allows site directed mutagenesis of the gene (Figs. 5 and 7). The first mutagenesis of pea lectin removed an internal *Bam HI* site to facilitate a later cloning step and did not alter the amino acid sequence. Next the codon for Lys 240 was converted into a TAA stop codon by insertional site-directed mutagenesis. This mutagenesis also generated a *Hind III* site immediately downstream of the stop codon. The final modification was the insertion of a synthetic linker into the *Bcl I* site seven codons into the gene. This linker provided a *Bam HI* site which allowed the in-frame ligation to the Lambda $P_L$ promoter as well as an initiator Met codon. As a result, nine base pairs were added to the 5' end of the gene which adds an additional Met-Asp-Pro to the N-terminus of the protein. The recombinant protein is now identical to native pea lectin with the exception above and amino acids 182-187 which are cleaved from the native protein to form two subunits but remain in the recombinant protein (Fig. 4). To generate the expression vector pPL 21, the Lambda $P_L$ promoter was cleaved from pAS 1 at the *Eco RI* and *Bam HI* sites and ligated to the *Bam HI* site at the 5' end of the pea lectin sequence. This expression vector, when transformed into the temperature sensitive Lambda repressor strain N4830, expresses recombinant pea lectin at the nonpermissive temperature (see appendix). The identity of the induced protein has been demonstrated by a number of methods. The molecular weight of the protein is within the expected range of about 25,000 Daltons. This is slightly larger than the 23,000 Dalton size of native pea lectin as would be expected because the presence of an additional nine amino acids in the recombinant. Western blot analysis shows that the protein is bound by rabbit antibody which specifically binds native pea lectin (see appendix). The recombinant protein can be purified from cell lysates based on its affinity for Sephadex just as is the case with native pea lectin (see appendix). Finally, partial amino acid sequencing reveals that the recombinant protein has the expected 15 N-terminal amino acids. The protein has been isolated in quantities sufficient for crystal growth experiments and hemagglutination assays (65 mg to date). The current construction and isolation procedure yields approximately 2.5 mg/liter of growth medium and is routinely being grown on a 28 liter scale. Facilities at the University of Alabama at Birmingham, Department of Biochemistry, permit routine growth of bacteria on a 500 liter scale.



-30
NNNNNNNN RNAANNARNR ATG GCT TCT CTT CAA ACC CAAATGATC TCG
Met Ala Ser Leu Glu Thr Glu Met Ile Ser

[*Met Asp Pro*
-20 -10 *C CAT ATG GAT CCG*
TTC TAC GCG ATA TTT CTA TCC ATT CTC TTA ACA ACA ATC CTT TTC TTC AAG GTG AAC TCA
Phe Tyr Ala Ile Phe Leu Ser Ile Leu Leu Thr Thr Ile Leu Phe Phe Lys Val Asn Ser

1 *T* *C* **Bcl I** 10 20
ACT GAA ACC ACT TCC TTC TTG ATC ACC AAG TTC AGC CCC GAC CAA CAA AAC CTA ATC TTC
[Thr Glu Thr Thr Ser Phe Leu Ile Thr Lys Phe Ser Pro Asp Gln Gln Asn Leu Ile Phe

30 40
CAA GGA GAT GGC TAT ACC ACA AAA GAG AAG CTG ACA CTG ACC AAG GCA GTA AAG AAC ACT
Gln Gly Asp Gly Tyr Thr Thr Lys Glu Lys Leu Thr Leu Thr Lys Ala Val Lys Asn Thr

50 60
GTT GGC AGA GCC CTC TAT TCC TCA CCT ATC CAT ATC TGG GAT AGA GAA ACA GGC AAC GTT
Val Gly Arg Ala Leu Tyr Ser Ser Pro Ile His Ile Trp Asp Arg Glu Thr Gly Asn Val

70 80
GCT AAT TTT GTA ACT TCC TTC ACT TTT GTC ATA AAT GCA CCC AAC AGT TAC AAC GTT GCC
Ala Asn Phe Val Thr Ser Phe Thr Phe Val Ile Asn Ala Pro Asn Ser Tyr Asn Val Ala

90 100
GAC GGG TTT ACG TTC TTC ATC GCA CCT GTA GAT ACT AAG CCG CAG ACC GGC GGT GGA TAT
Asp Gly Phe Thr Phe Phe Ile Ala Pro Val Asp Thr Lys Pro Gln Thr Gly Gly Gly Tyr

110 120
CTC GGA GTT TTC AAT AGC GCA GAG TAT GAT AAA ACC ACT CAA ACT GTT GCT GTG GAG TTT
Leu Gly Val Phe Asn Ser Ala Glu Tyr Asp Lys Thr Thr Gln Thr Val Ala Val Glu Phe

*C* 130 140
GAC ACT TTC TAT AAT GCT GCA TGG GAT CCA AGC AAC AGA GAT AGA CAT ATT GGA ATC GAT
Asp Thr Phe Tyr Asn Ala Ala Trp Asp Pro Ser Asn Arg Asp Arg His Ile Gly Ile Asp

150 160
GTG AAC AGT ATC AAA TCC GTA AAC ACT AAG TCG TGG AAG TTG CAG AAT GGT GAA GAG GCT
Val Asn Ser Ile Lys Ser Val Asn Thr Lys Ser Trp Lys Leu Gln Asn Gly Glu Glu Ala

170 180
AAT GTT GTG ATA GCT TTT AAT GCT GCT ACT AAT GTG TTA ACT GTT AGT TTG ACC TAT CCT
Asn Val Val Ile Ala Phe Asn Ala Ala Thr Asn Val Leu Thr Val Ser Leu Thr Tyr Pro

190 200
AAT TCA CTT GAG GAA GAG AAT GTA ACT AGT TAT ACT CTT AGC GAC GTT GTG TCT TTG AAG
**Asn**]Ser Leu Glu Glu Glu Asn [ Val Thr Ser Tyr Thr Leu Ser Asp Val Val Ser Leu Lys

210 220
GAT GTT GTT CCT GAG TGG GTA AGG ATT GGT TTC TCA GCT ACC ACA GGA GCA GAA TAT GCA
Asp Val Val Pro Glu Trp Val Arg Ile Gly Phe Ser Ala Thr Thr Gly Ala Glu Tyr Ala

230 240
GCA CAT GAA GTT CTT TCA TGG TCT TTT CAT TCT GAG TTG AGT GGA ACT TCA AGT TCT T AAG

Ala His Glu Val Leu Ser Trp Ser Phe His Ser Glu Leu **Ser** Gly Thr Ser Ser **Ser**] Lys

*TT* **Pst I** 246
CAA GCT GCA GAT GCA TAG TTTTTTGCTT TTCATCAT
Gln Ala Ala Asp Ala Stop

**Fig. 4. Organization of pPS 15-50* and modifications made to express recombinant pea lectin in *E. Coli*.** The nucleotide and amino acid sequences of the pea lectin coding region of pPS15-50 are shown. Relevant restriction sites are underlined. Changes made to the sequence are shown in italics, those created by site-directed mutagenesis are in bold face. The brackets at Thr-1 and Val-188 indicate the positions of the native N-termini of the beta and alpha subunits respectively. The bracket preceding *Met* at the -3 position indicates the recombinant protein synthesis initiation site. A linker replacing sequences upstream of the *Bcl I* site provided the initiator ATG codon and a *Bam HI* site for attachment of the Lambda $P_L$ promoter. Two bases were changed to enhance the efficiency of translation in *E. Coli*. Site-directed mutagenesis was used to remove an internal *Bam HI* site and to create a stop codon followed by a *Hind III* site.
*pPS 15-50 was a generous gift from Dr. Thomas Higgins of the Commonwealth Scientific and Industrial Research Organization, Division of Plant Industry, Canberra, Australia.

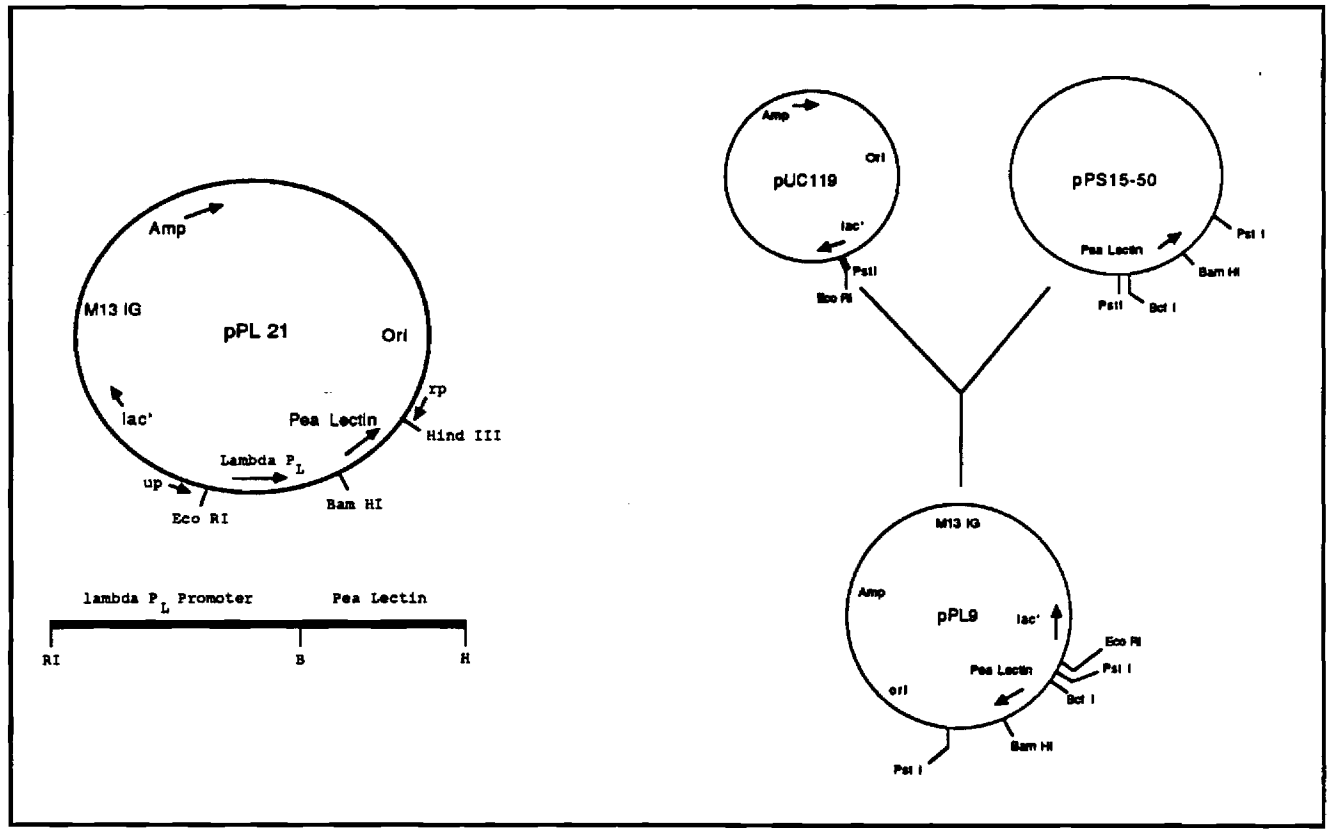



**Figure 5. (left)** Diagram of the expression vector pPL 21. Relevant restriction sites are shown. The positions of the universal and reverse primer annealing sites are shown (up and rp). pPL 21 is constructed so that the expression of the pea lectin gene is under the control of the Lambda repressor. When transformed into the *E. Coli.* strain N4830, a temperature sensitive mutant for the Lambda repressor, the expression of pea lectin can be induced at the nonpermisive temperature.
**Figure 6.(right)** pPL 9 was generated by ligating the gel-purified *Pst I* restriction fragment from pPS15-50 to pUC119 which had been digested with *Pst I* and treated with alkaline phosphatase to prevent reclosure of the vector. This plasmid was used to create an expression vector pPL21.

**Publications Resulting From Project:**

1. Einspahr, H., Suguna, K., Suddath, F.L., Ellis, G., Helliwell, J.R., and Papiz, M.Z., "The Location of Manganese and Calcium Ion Cofactors in Pea Lectin Crystals by Use of Anomalous Dispersion and Tunable Synchrotron X-Radiation", *Acta Cryst.* B**41**,336-341, (1985).

2. Einspahr, H., Parks, E.H., Suguna, K., Subramanian, E., and Suddath, F.L., "The Crystal Structure of Pea Lectin at 3.0 Å Resolution", *J. Biol. Chem.* **261**, 16518-16527, (1986).

3. Greenhough, T.J., and Suddath, F.L., "Oscillation Camera Data Processing. 4. Results and Recommendations for the Processing of Synchrotron Radiation Data in Macromolecular Crystallography", *J. Appl. Cryst.* **19**, 400-409, (1986).

4. Basu, Debkumar, Delucas, Lawrence, Parks, Elizabeth H., Suddath, F.L., (1988) Preliminary Crystallographic Study of the α-D-Galactose-specific Lectin from Jack Fruit (*Artocarpus Integra*) Seeds. *J. Mol. Biol.* **201**, 661-662.

5. Einspahr, H., Parks, E.H.,Phillips, S.R., and Suddath, F.L. (1988) Crystal Structure Studies of Legume Lectins. "Lectins in Biology, Biochemistry, Clinical Biochemistry", Vol 6, Ed. T.C.Bøg-Hansen and D. Freed, Sigma Library, St.Louis, MO. (in press).

**Table 1** Properties of Lectins and status of x-ray analysis

| Lectin | *Pisum sativum* ("wild type") pea lectin (PSA) | *Pisum sativum* (cloned) PSAclone (PSAC) | *Lens Culnaris* lentil lectin LCA | *Artocarpus integra* Jack Fruit JSA |
|---|---|---|---|---|
| Source | Asgrow Seed Co. | Engler/Prasthoffer | Asgrow Seed Co. | Deb Basu |
| M.W. | 49,740 (dimer) α-chain=7000 β-chain=17,000 $(\alpha,\beta)_2$ | 51,720 (dimer) α=25,860 | 47,200 (dimer) α-chain=5928 β-chain=17,572 $(\alpha,\beta)_2$ | 39,500 (tetramer) α-chain=10,500 $\alpha_4$ |
| Specificity | α-methyl mannose α-methyl glucose | α-methyl mannose α-methyl glucose | α-methyl mannose α-methyl glucose | Galβ1⇒3GalNAc T-antigen disaccharide |
| Space Group | $P2_12_12_1$ | $P2_12_12_1$ | $P2_12_12_1$ | $P2_12_12$ |
| Unit Cell | | | | |
| a= | 50.73 | 65.1 | 50.0 | 77.1 |
| b= | 61.16 | 73.0 | 67.5 | 123.3 |
| c= | 136.59 | 109.1 | 130.3 | 78.7 |
| Status | 1.9 Å refined (R=0.20) | diffract to 2.8 Å | diffract to 2.5 Å | diffract to 2.8 Å |

| Lectin | *Sophora japonica* seed lectin SJA(seed I) | *Sophora japonica* seed lectin SJA(seed II) | *Sophora japonica* leaf lectin SJA(leaf I) | *Pisum sativum* Pea Lectin/Carbohydrate PSA-CHO |
|---|---|---|---|---|
| Source | Shannon/ McPhearson | Shannon/ McPhearson | Shannon/ McPhearson | Carver/Rini/Hardman |
| M.W. | 130,000(tetramer) α-chain=30,000 $\alpha_4$ | 130,000(tetramer) α-chain=30,000 $\alpha_4$ | 130,000(tetramer) α-chain=30,000 $\alpha_4$ | 47,000+CHO |
| Specificity | galactose | galactose | galactose | |
| Space Group | $P2_12_12_1$ | $P2_12_12_1$ | $P2_12_12_1$ | $P2_12_12_1$ |
| Unit Cell | | | | |
| a= | 60.8 | 64.2 | 105.0 | 64.3 |
| b= | 140.0 | 132.4 | 106.9 | 73.4 |
| c= | 149.0 | 139.2 | 133.3 | 108.5 |
| Status | large crystals diffract to 3.0 Å | large crystals diffract to 2.8 Å | large crystals diffract to 2.8 Å | model built at ~2.5 Å |

## D. Experimental Design and Methods

### D.I. X-Ray Structure Determination

The studies proposed will be carried out using well established and straightforward crystallographic techniques. Few preliminary experiments are necessary because excellent diffraction quality crystals are available for PSA, PSAC (PSA cloned), PSA monosaccharides, LCA, JSA (jack fruit lectin), SJA (seed I), SJA (seed II), and SJA (leaf I). We isolate and purify PSA, LCA, and PSA cloned proteins ourselves and we purify material supplied for the jack fruit. The SJA lectins are supplied as crystalline material suitable for data collection. An excellent set of starting model coordinates are available for PSA. These coordinates have been recently used to solve two other lectin structures, the favin-glucose structure by Reeke and Becker[77] and the PSA-$(Man)_3$ complex by Hardman. An excellent set of coordinates for ConA are also available in the laboratory based on Hardman's refinement at 1.7 Å.[78] If crystals of the various lectin, lectin-carbohydrate complexes are isomorphous with the parent or native protein crystals the structures can be solved straightforwardly by Fourier analysis using the calculated phases of the native-parent protein. Otherwise, the structures will be solved by molecular replacement technique or if necessary multiple isomorphous replacement. The amino acid sequence is available for PSA mutants, LCA, SJA (seed I), and sainfoin.

The data for these experiments will be collected on our area detector. We do not have extensive experience collecting data on an area detector but we do have experience with photographic film and diffractometer methods. This background we think will be sufficient to learn the characteristics of the new facility and produce high quality diffraction data.

We have been involved in a data comparison study with John Helliwell of York University whereby data sets were collected from PSA crystals by conventional single counter diffractometer (Picker-sealed tube), photographic film methods at the Daresbury Synchrotron Facility (wiggler beam line- flat film cassettes), and the FAST area detector system on the wiggler line (data not collected at this time). PSA crystals were chosen as a standard because in has a reasonably large cell, diffracts to 1.2 Å (detected by film on the synchrotron) and last for hours in the synchrotron beam. As a result of these ongoing studies we will have a variety of data sets to compare with our in house area detector data to assess the quality.

### D.II. PSA Refinement

As outlined in the progress report we have recently completed the computing aspects of PSA refinement at 1.9 Å resolution. There is much to be done in evaluating the results of the restrained least squares refinement. The data being used for the refinement was collected several years ago at the Daresbury Synchrotron Facility in collaboration with Drs. John Helliwell and Trevor Greenhough. Data was collected over a 100° rotation for PSA crystals mounted about a* and b* ($\lambda$=1.46 Å). To 1.83 Å the data is 92% complete and 89% of the data from 2.0 Å to 1.83 Å is above 2 sigma. A large number of redundant observations were made for all reflections and sigmas were based on the variations in measurements of the same and symmetry related reflections. There were some problems with the data because too large an oscillation angle was used for particular orientations of the reciprocal lattice. The data has been recently reproccessed using the profile fitting method described by Greenhough and Suddath.[79] We plan to recollect the native PSA data on the area detector and compare it to both the Daresbury film data ($R_{merg}$=0.072 to 1.83 Å) and the original Picker diffractometer data ($R_{merg}$=0.037 to 3.0 Å). We expect to eventually carry the refinement to the limit of the crystals. Whether the highest resolution data is collected at Daresbury or Georgia Tech will depend on the results of experiments this summer conducted on the in-house area detector. It is a major objective of this project to carry the refinement of PSA to maximum resolution in order to observe the differences between the two monomers, fully define the solvent structure as best we can at this resolution, and observe the variations in isotropic temperature factors as a function of crystal contacts and local environment. The diffracting quality of the PSA crystal is rather unusual for a molecule of 50K daltons to diffract to such high resolution. We hope to take full advantage of the opportunity and possibly understand in a more detailed way the structural reasons for this order.

**D.III. PSA Mutants**

The recombinant PSAC (for PSA cloned) protein has 3 additional amino acids at the N-terminus, formylMet, Asp, Pro and the 6 amino acids, Ser, Leu, Glu, Glu, Glu connecting what in native PSA are the C-terminus of the β-chain and the N-terminus of the α-chain. The position of the C-terminus of the α-chain of native PSA is difficult to determine. The cDNA before modification has a stop codon at Ala-245 but C-terminal analysis indicates Ser. There are 3 Ser residues between the stop codon and the last residue that can be unequivocally assigned in the current map, which is Ser-234 (both monomers). The stop codon in the pPL 21 construction is where Lys-240 was in the native cDNA leaving the C-terminus at Ser-239. Even though crystals of PSAC have already been obtained it may be desirable to prepare a different molecule by terminating the PSAC at Ser-234. That can easily be accomplished by introducing a stop codon by a single base mutation at Gly-235 (GGA). This would yield a molecule with a closely "trimmed" C-terminal end. N-terminal amino acid analysis of PSAC indicates heterogeneity at the N-terminal end, due to removal of the fMet and Asp, thus it might be possible to enhance this "processing" to yield a PSAC with only the linking 6 amino acids.

Our intentions are to continue efforts to prepare diffraction quality crystals of PSAC mutants. When crystals are obtained the structures will be solved by Fourier analysis if the crystals are isomorphous with PSAC or PSA, by molecular replacement or MIR methods otherwise.

It will be a challenge to introduce site specific mutations into PSA that illuminate our understanding of protein-protein and protein-carbohydrate interactions and not merely create addition structural work. We will be guided, but not restricted, in our design efforts by a three fundamental considerations:

1. Replacement of one amino acid by an isosteric residue of different function (*i.e.*, Asp by Asn, a replacement that maintains most of the hydrogen bonding characteristics, but removes the charge).

2. Replacement of one amino acid by another of identical function but different structure (*i.e.*, Glu by Asp, were the carboxyl group would be moved by approximately 1 Å).

3. Is it possible for the substituted residue to assume the $\phi\,\psi$ angles that the replaced residue has in the native structure (*i.e.*, can a Pro residue assume the same conformation of a specific Gly).

The proven and suggested carbohydrate bindings sites on ConA, favin, lentil, PSA, and other lectins will also guide us in selecting critical residues for alteration. We are interested in using the method of site specific mutations to investigate at least four aspects of the PSAC molecule.

1. By analogy with ConA, favin, incomplete information from the PSA-trimannoside complex, model building we know some of the residues that are involved in binding monosaccharides and are suggested to be involved in oligosaccharide binding. Dr. Engler has prepared the following mutants:
   (i) Asp81 - Glu81. Asp81 is probably involved in binding mannose by accepting H-bonds on the side chain. The effect of changing 81 to a Glu will restrict the bottom of the binding "pocket" yet still provide the H-bonding capability.
   (ii) Gly99 - Ala99. This will change the nature of backbone conformation and possibly prevent the donation of main chain H-bonds to the mannose.
   (iii) Tyr100 - Phe100. The Tyr100 appears to be in a position to donate a H-bond from the ring. The replacement with Phe will significantly alter the nature of this part of the extended "pocket".
   (iv) Trp128 - Phe128. The Try appears to provide a hydrophobic "patch" for the binding of fucose. The substitution with Phe will change the extent of this hydrophobic "patch" yet still maintain a similar nature.
   Dr. Engler has prepared ~25mg of each of these mutants and they have been isolated by affinity chromatography on Sephadex. This indicates the simple carbohydrate binding site is still functional enough to bind to Sephadex. Attempts to crystallize the mutants both complexed with 3,6-di-O-(α-D-mannopyranosyl)-α-D-mannopyranoside and Asn-linked-oligosaccharides is underway.

2. As noted in the progress report the interface between monomers appears to be polar and hydrophilic rather than hydrophobic (see Appendix A). We plan to investigate alterations at the interface and monitor the stability of dimer-monomer interaction as a function of solvent, ionic strength,

denaturants, temperature, etc. Many of these mutants can be investigated by biochemical methods but attempts will always be made to crystallize the PSACs. Should a mutant be found that prevents dimer formation and it crystallizes with a monomer in the asymmetric unit it might serve as a new and simpler system for further crystallographic studies.

3. A number of different ions can be substitute for $Ca^{++}$ and $Mn^{++}$ in PSA. There is an interesting alteration in the conformation of the peptide bond at Ala-80-Asp-81 on binding metals. The metal ions are required for monosaccharide binding and the trigger for producing the "cryptic" binding site is the isomerization of *trans* Ala-80⇒Asp-81 to *cis* Ala-80⇒Asp81. We plan to investigate the effects on mono- and oligosaccharides binding to PSACs that have been altered at the interface between the metal loop and the monosaccharide binding site. Dr. Engler has cloned the replacement of Asn125, that provides a ligand to the calcium ion, With an Arg125. The side chains positive charge could substitute for the calcium as has been seen in some serine proteases. This mutant has not been expressed.

4. It is not possible to do reversible protein folding-unfolding experiments with PSA since it is composed of two chains and will not renature after having been denatured. The PSAC molecule is a single chain and these types of experiments could be done. The reason one might want to do these experiments come from the following observation. ConA, a single chain molecule, will reversibly refold under proper conditions, but ConA's amino acid sequence has been permuted by the protolysis-ligation step show by Bowles. ConA can refold to the same three-dimensional structure from topologically different arrangements of the same amino acids. Since the native form of ConA is unavailable we propose to construct the topological equivalent of ConA for PSA (See Fig. 7.). In fact a family of PSACs could be formed where the N-terminal and C-terminal ends could be anywhere around the circumference of the circularized linear sequence of PSAC. The crystallization of these circularly permuted mutants would provide a unique opportunity to observe the effects of chain termini at positions within the three-dimensional structure. Many of these mutants would not be expected to fold into stable structures; however, the data relating the position of the ends and the ability to form stable structures may suggest critical steps in the folding pathways.

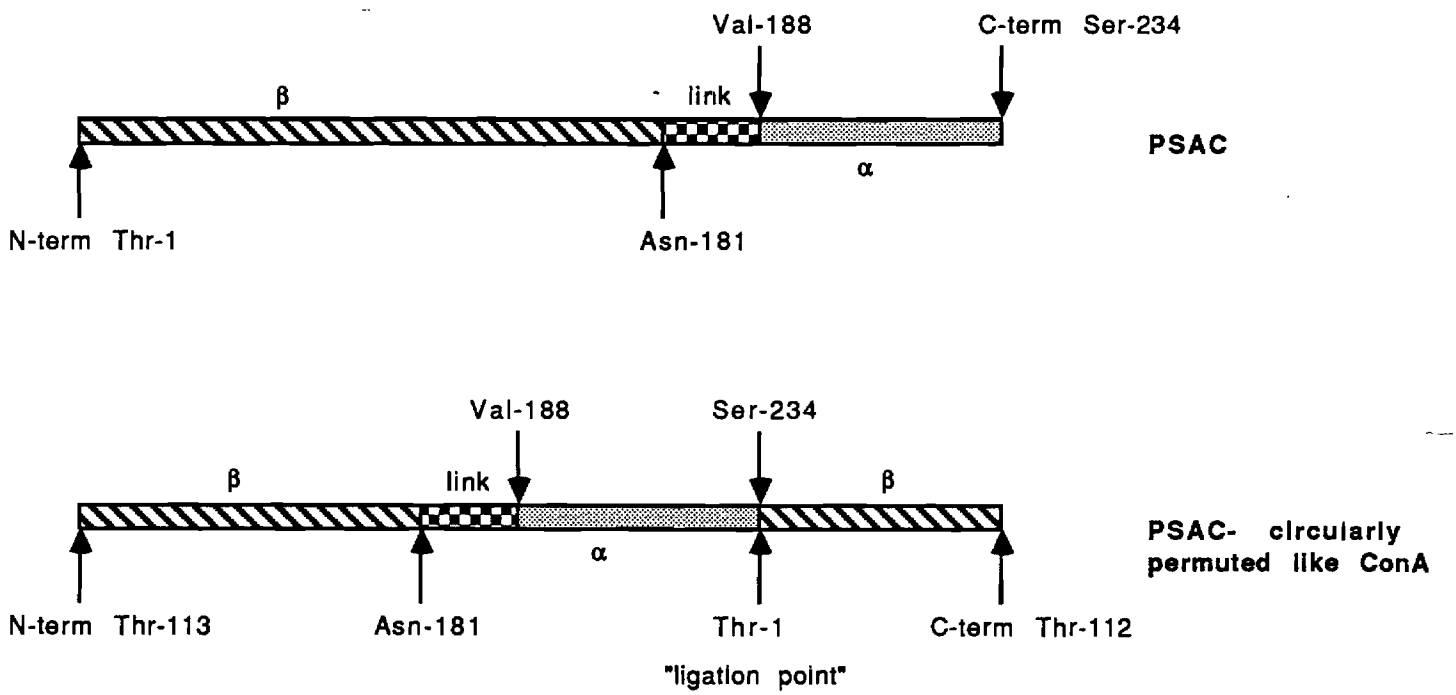



**Fig. 7.** Linear representation of the sequence of PSAC (top) and a circularly permuted sequence that is topologically equivalent to native ConA (bottom).

## D.IV. PSA and LCA-Carbohydrate Complexes

Dr. Richard Cummings has agreed to provide us with 1-5μmoles of several Asn-linked oligosaccharides shown in figure 9. These oligosaccharides will be used to soak crystals of PSA and PSAC with the expectation that they might diffuse into the crystals in spite of their rather large size. We will also form complexes of PSA or PSAC and selected oligosaccharides in solution and attempt to co-crystallize these complexes. We already know that PSA forms good crystals with 3,6-di-O-(α-D-mannopyranosyl)-α-D-mannopyranoside. Many of the more complex oligosaccharides will no doubt present a challenge. They do form quite stable complexes with PSA, especially when they contain the branched fucose. The opportunity to observe the structure of any of these carbohydrates at high resolution would be unique. Considerable interest has been expressed in using the lectins as molecular "scaffolding" in order that conformationally mobile oligosaccharides be observed in at least one stable conformation. No doubt the binding of these complex carbohydrates to the surface of the lectin will induce or select particular conformations that may only have transient existence in free solution.

An attempt to predict the success of the efforts to obtain crystals, as outlined above, would be foolish. However, we have large quantities of all lectins of interest and adequate supplies of N-linked oligosaccharides with a variety of different sugar substituents of varying structural complexity. We are optimistic that some combinations of these variables will yield crystals of good diffracting quality.

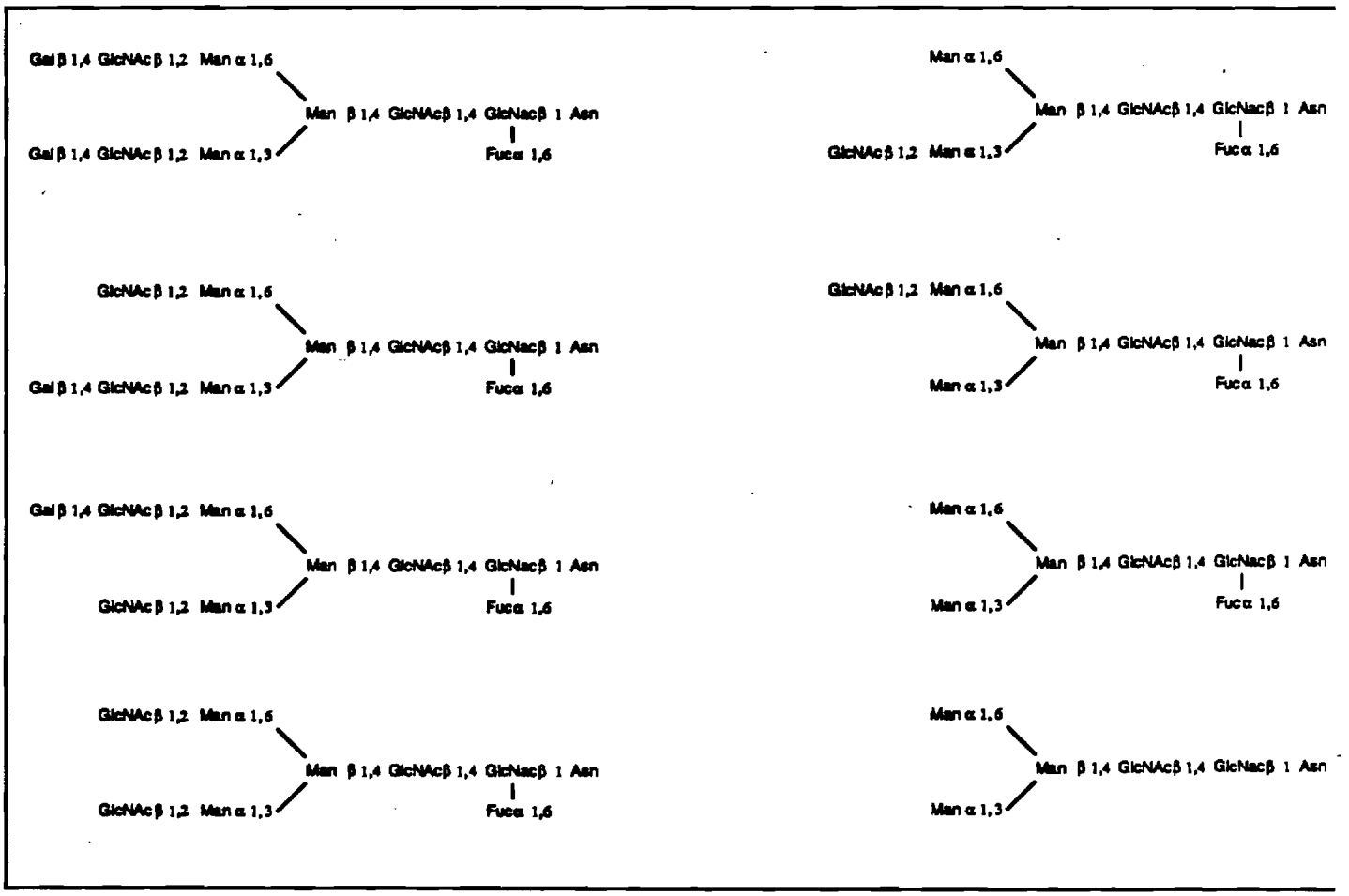



**Fig. 8. N-linked Oligosaccharides.** This family of oligosaccharides form complexes with PSA and LCA with varying binding constants. This family of oligosaccharides can be prepared with the fucose deleted but with a significantly diminished binding constant.

**D.V. Other Lectins**

There are several other lectins that we plan to pursue if time permits. These are the interesting tetrameric lectin from jack fruit (JSA). We have large crystals of this lectin that diffract to at least 2.8 Å resolution (see table 1). JSA has only two carbohydrate binding sites per tetramer unlike most other lectins that would have four, one per subunit. This suggest that at least in the quaternary structure the 4 subunits are nonequivalent or two subunits are required to form a carbohydrate binding site. Biochemical methods indicate the four 10,500 dalton subunit are identical.

Likewise the *sophora japonica* lectins are a very interesting series of almost identical proteins but isolated from different plant tissue. The differences in amount of covalently attached carbohydrate, amino acid sequence and subtle carbohydrate specificities would may an interesting complement to our studies of PSA and LCA. The investigation of these lectin will be pursued also if time permits.

**E. Human Subjects** - None

**F. Vertebrate Animals** - None

**G. Consultants:** Letters attached (see appendix).

1. Dr. Jeff Engler, Assistant Professor of Biochemistry, The University of Alabama at Birmingham. Dr. Engler has expressed PSA in *E.coli* and is training Mr. Thomas Prasthofer in recombinant DNA methods. Dr. Engler will continue to collaborate with us by preparing site specific mutants of PSA.

2. Dr. Alexander McPherson, Professor and Chairman, Department of Biochemistry, University of California, Riverside. Dr. McPherson has a wealth of proteins crystallized in his laboratory and has generously agreed to supply us with 3 **crystalline** *saphoria japonica* lectins purified in collaboration with Dr. Lee Shannon of UCR.

3. Dr. Richard Cummings, Associate Professor, Department of Biochemistry, University of Georgia. Dr. Cummings is an expert in the isolation and purification of complex carbohydrates. In addition he has an interest in the molecular details of lectin-carbohydrate complexes. He has generously agreed to supply us with several oligosaccharides that very specifically bind pea and lentil lectin.

**H. Consortium Arrangements** - None

**I. Literature Cited**

(Multiple entries for the same reference are a limitation of the word processor).

[1] Weis, W., Brown, J.H., Cusack, S., Paulson, J.C., Skehel, J.J., and Wiley, D.C., *Nature*, **333**, 426-431 (1988).

[2] Varghese, J.N., Laver, W.G., and Colman, P.M., *Nature*, **303**, 35-40 (1983).

[3] Vyas, N.K., Vyas, N.N., and Quiocho, F.A., *Nature*, **327**, 635-638 (1987) and references therein.

[4] Quiocho, F.A., *Annu. Rev. Biochem.*, Vol **55**, pp. 287-315 (1986).

[5] Stillmark, H.,"Uber Rizin, ein giftiges Ferment aus dem Samen von *Ricinus communis L.* and einigen anderen Euphorbiaceen," Inaug. Diss., Dorpat, (1888).

[6] Boyd, W.C. and Shapleigh, E. , *J. Immunol.*, **73**, 226-231 (1954).

[7] Boyd, W.C., and Shapleigh, E., *Science*, **119**, 419 (1954).

[8] Boyd, W.C., and Reguera, R.M., *J. Immonol.*, **62**, 333-339 (1949).

[9] Kauss, H., and Glaser, C., *FEBS Letts.*, **45**, 304-307 (1974).

[10] Hamblin, J., and Kent, S.P., *Nature (London) New Biol.*, **245**, 28-30 (1973).

[11] Albersheim, P., and Anderson, A.J., *Proc. Natl. Acad. Sci. USA.*, **68**, 1815-1819 (1971).

[12] Janzen, D.H., Juster, H.B., and Liener, I.E., *Science*, **192**, 795-796 (1976).

[13] Sharon, N. in "Mitogens in Immunobiology," J.J. Oppenheim and D.L. Rosenstreich, eds. Academic Press, New York, pp 31-41 (1976).

[14] Bird, G.W.G., *Br. Med. Bull.*, **15**, 165-168 (1959).
[15] Li, J.G., and Osgood, E.E., *Blood*, **4**, 670-675 (1949).
[16] Dorset, M., and Henley, R.R., *J. Agric. Res.* (Washington, D.C.), **6**, 333-339 (1916).
[17] Watkins, W.M., in "Glycoproteins," A. Gottschalk, ed., Elsevier, Amsterdam, 2nd Edition, 1972, Part B, pp. 830-891.
[18] Lis, H., and Sharon, N., *Annu. Rev. Biochem.*, 42, 541-574 (1973).
[19] Nicholson, G.,*Int. Rev. Cytol.*, 39, 89-190 (1974).
[20] Sharon, N., and Lis, H., *Methods Membrane Biol.*, 3, 147-200 (1975).
[21] Kristiansen, T., *Methods Enzymol.*, 34, (part B), 331-341 (1974).
[22] Lis, H. , and Sharon, N., *Annu. Rev. Biochem.*, Vol **55**, pp. 35-67 (1986).
[23] Reeke, G.N., and Becker, J.W. "Carbohydrate Binding Sites of Plant Lectins"
[24] Quiocho, F.A., *Annu. Rev. Biochem.*, Vol **55**, pp. 287-315 (1986).
[25] Sumner, J.B., *J. Biol. Chem.* **37**, 137-143 (1919).
[26] Wang, J.L., Cunningham, B.A., Waxdal, M.J., and Edelman, G.M., *J.Biol. Chem.* **250**, 1490-1502 (1975).
[27] Cunningham, B.A., Wang, J.L., Waxdal, M.L., and Edelman, G.M., *J. Biol. Chem.* **250**, 1503-1512 (1975).
[28] Edelman, G.M., Cunningham, B.A., Reeke, G.N., Becker, J.W., Waxdal, M.J., and Wang, J.L., *Proc. Natl. Acad. Sci USA*, **69**, 2580-2584 (1972).
[29] Hardman, K.D., and Ainsworth, C.F., *Biochemistry*, **11**, 4910-4919 (1972).
[30] Hardman, K.D., Agarwal, R.C., and Freiser, M.J., *J. Mol. Biol.* **157**, 69-86 (1982).
[31] Brewer, F., Bhattacharyya, L., Brown, R.D., and Koenig, S.H., *Biochem. Biophys Res. Comm.* **127**, 1066-1071 (1985).
[32] Edelman, G.M., Cunningham, B.A., Reeke, G.N., Becker, J.W., Waxdal, M.J., and Wang, J.L., *Proc. Natl. Acad. Sci USA*, **69**, 2580-2584 (1972).
[33] Hardman, K.D., and Ainsworth, C.F., *Biochemistry*, **11**, 4910-4919 (1972).
[34] Wright, C.S., *J. Mol. Biol.*, **111**, 439-457 (1977).
[35] Reeke, G.N., and Becker, J.W., *Science*, **234**, 1108-1111 (1986).
[36] Einspahr,H., Parks, E.H., Suguna, K., Subramanian, E., and Suddath, F.L., *J. Biol. Chem.*, **261**, 16518-16527 (1986).
[37] Riskulov,R.R., Dobrokhotova, Z.D., Kuzev, S.V., Lobsanov, Y.D., Lubnin, M.Y., Mokulskava, T.D., Myshko, G.E., Proskudina, L.T., Rogacheva, M.M., Saprykina, L.F., Khrenov, A.A., and Mokulskii, M.A., *FEBS Lett.*, **165**, 97-101 (1984).
[38] Boaz Shannan private communication. (1986).
[39] Salunke, D.M., Swamy, M.J., Khan, M. I., Mande, S.C., Surolia,A., and Vijayan, M., *J. Biol. Chem.* **260**, 13576-13579 (1985).
[40] Rini, J.M., Carver, J.P., and Hardman, K.D., *J. Mol. Biol.*, **189**, 259-260 (1986).
[41] Foriers, A., Lebrun, E., Van Rapenbusch, R., deNeve, R., and Strosberg, A. D. (1981) *J. Biol. Chem.* **256**, 5550-5560.
[42] Higgins, T. J. V., Chandler, P. M., Zurawski, G., Button, S. C., and Spencer, D. (1983) *J. Biol. Chem.* **258**, 9544-9549.
[43] Hopp, T. P., Hemperly, J. J., and Cunningham, B. A. (1982) *J. Biol. Chem.* **257**, 4473-4483.
[44] Hemperly, J. J., and Cunningham, B. A. (1983) *Trends Biochem. Sci.* **8**, 100-102.
[45] Kouchalakos, R. N., Bates, O. J., Bradshaw, R. A., and Hapner, K. D. (1984) *Biochemistry* **23**, 1824-1830.
[46] Hoffman, L. M., Ma, Y., and Barker, R. F. (1982) *Nucleic Acids Res.* **10**, 7819-7828.
[47] Hopp, T. P., Hemperly, J. J., and Cunningham, B. A. (1982) *J. Biol. Chem.* **257**, 4473-4483.
[48] Richardson, M., Campos, F. D. A. P., Moreira, R. A., Ainouz, I. L., Begbie, R., Watt, W. B., and Pusztai, A. (1984) *Eur. J. Biochem.* **144**, 101-111.
[49] Hemperly, J. J., Mostov, K. E., and Cunningham, B. A. (1982) *J. Biol. Chem.* **257**, 7903-7909.

[50] Higgins, T. J. V., Chandler, P. M., Zurawski, G., Button, S. C., and Spencer, D.(1983) *J. Biol. Chem.* **258**, 9544-9549.

[51] Carrington, D. M., Auffret, A., and Hanke, D. E. (1985) *Nature* **313**, 64-67.

[52] *ibid.*

[53] Bowles, D.J., Marcus, S.E., Pappin, D.J.C., Findlay, J.B.C., Eliopoulos, E., Maycox, P.R., and Burgess, J., *J. Cell. Biol.* **102**, 1284-1297 (1986).

[54] Meehan, E. J., Jr., McDuffie, J., Einspahr, H., Bugg, C. E., and Suddath, F. L., *J. Biol. Chem.* **257**, 13278-13282, (1982).

[55] Einspahr, H., Parks, E.H., Suguna, K., Subramanian, E., and Suddath, F.L., *J. Biol. Chem.* **261**, 16518-16527 (1986).

[56] Riskulov, R.R., Dobrokhotova, Z.D., Kuzev, S.V., Lobsanov, Y.D., Lubnin, M.Y., Mokulskava, T.D., Myshko, G.E., Proskudina, L.T., Rogacheva, M.M., Saprykina, L.F., Khrenov, A.A., and Mokulskii, M.A., *FEBS Lett.* **165**, 97-101 (1984).

[57] Meehan, E. J., Jr., McDuffie, J., Einspahr, H., Bugg, C. E., and Suddath, F. L., *J. Biol. Chem.* **257**, 13278-13282 (1982).

[58] Einspahr,H., Parks, E.H., Suguna, K., Subramanian, E., and Suddath, F.L., *J. Biol. Chem.* **261**, 16518-16527 (1986).

[59] Hendrickson, W. A., and Konnert, J. H. , in *Computing in Crystallography* (Diamond, R., Ramaseshan, S., and Venkatesan, K, eds.) pp 13.01-13.23, Indian Academy of Sciences, Banjalore (1980).

[60] Jones, T. A. (1978) *J. Appl. Cryst.* **11**, 268-272.

[61] Sheriff, S. and Hendrickson, W.A., *Acta Cryst.* **A43**, 118-121 (1987).

[62] Higgins, T. J. V., Chandler, P. M., Zurawski, G., Button, S. C., and Spencer, D.(1983) *J. Biol. Chem.* **258**, 9544-9549.

[63] Richardson, C., Behnke, W. D., Freisheim, J. H., and Blumenthal, K. M. (1978) *Biochim. Biophys. Acta* **537**, 310-319.

[64] Rini, J.M., Hofmann, T., and Carver, J.P., *Biochem. Cell Biol.*, Submitted, November 1986.

[65] Sibanda, B. L. and Thornton, J. M. (1985) *Nature* **316**, 170-174.

[66] Hardman, K. D. and Ainsworth, C. F. , *Biochemistry* **15**, 1120-1128 (1976).

[67] Becker, J. W., Reeke, G. N., Jr., Cunningham, B. A., and Edelman, G. M. (1976) *Nature* **259**, 406-409.

[68] Hardman, K.D., Agarwal, R.C., and Freiser, M.J., *J. Mol. Biol.* **157**, 69-86 (1982).

[69] Carver, J. P., MacKenzie, A. E., and Hardman, K. D. (1985) *Biopolymers* **24**, 49-63.

[70] Rini, J.M., Carver, J.P., and Hardman, K.D., *J. Mol. Biol.* **189**, 259-260 (1986).

[71] Wang, A.H., Quigley, G., and Rich, A., *BBRC* **76**,1261 (1977)

[72] Foriers, A., Lebrun, E., Van Rapenbusch, R., deNeve, R., and Strosberg, A. D. (1981) *J. Biol. Chem.* **256**, 5550-5560.

[73] Kornfeld, K., Reitman, M.L., and Kornfeld, R., *J. Biol. Chem.* **256**, 6633 (1981).

[74] Merkle, R.K. and Cummings, R.D., *Methods in Enzymology* **138** 232-259 (1987).

[75] Yamamoto, K., Tsuji, T., and Osawa, T., *Carbohydr. Res.* **110**, 283 (1982).

[76] Quiocho, F.A., and Vyas, N.K., *Nature*, **310** 381-386 (1984).

[77] Reeke, G.N., and Becker, J.W., *Science*, **234**, 1108-1111 (1986).

[78] Hardman, K.D., Agarwal, R.C., and Freiser, M.J., *J. Mol. Biol.* **157**, 69-86 (1982).

[79] Greenhough, T.J., and Suddath, F.L., *J. Appl. Cryst.* **19**, 400-409, (1986).