EXPRESSION, STRUCTURAL AND FUNCTIONAL STUDIES OF FASCICLIN I

Thesis by

Wen-Ching Wang

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1992

(Defended May 22, 1992)

Dedicated to

my husband, Jong-Yih Lin

and to

my parents

Chen-Kuen Wang and Mei-Hei Chen

Acknowledgments

When I look back at these years as a graduate student at Caltech, many amazing things occurred in this world, including the Los Angeles riot. It also seems that I grew a lot while I was here, personally as well as scientifically. Many people at this academically energetic place have made my time very enjoyable and challenging. I would like to take this chance to express my gratitude to all of them.

My first thanks go to my advisers, Pamela J. Bjorkman and Kai Zinn, for their instruction, advice, support, encouragement, and ideas. Their constant curiosity and enthusiasm in science were inspiring. I am especially grateful for their patience in teaching me colorful scientific writing.

Several people have been especially valuable to me as sources of advice and as friends I could talk to about problems. I would like to thank Yueh-Ju Wang, Shin-Shey Tian, Louis Gastinel, Malini Raghavan and Susan Ou for their encouragement, many useful discussions and friendship.

I thank Jean-Paul Revel for his assistance in performing electron microscopic studies. I thank Ilana Tamir for her help in circular dichroism analysis of grasshopper fasciclin I and for sequencing the expression vector (pBJ5.GS.mCD4). I thank Pantelis Tsoulfas for his help in modifying the cDNA of grasshopper fasciclin I to generate a truncated form of this molecule using site directed mutagenesis. I also thank Barry Condrun for a pleasant collaboration in performing experiments using the primary culture of grasshopper embryos.

I thank John D. Balderschwieler, my advisor during my first year, for his interest, support and advice.

I thank David Penny for his technical assistance in tissue culture. I thank Roland Strong for providing editorial help in Chapter 4 and a schematic diagram of Cell Pharm II.

One of the most important factors in having a positive work environment at Caltech was being in the Bjorkman group. I thank all of my colleagues for providing a lively atmosphere, and for advice and discussions, scientific and non-scientific. Besides those already mentioned, these include Michael Blum, Peggy Fahnestock, Tippy Harlow, Andy Huber, and Dan Vaughn.

I would like to thank my thesis committee members for their taking time to sit and talk with me about scientific issues or counsel me on future career directions. Moreover, I would like to express my gratitude to the friends I have known while in graduate school. These people include many of those mentioned above and Ching-Hwa Kiang, Jen-Sue Chen, I-Ming Chen, Ming-Ji Fann, Te-Yi Kung, Amy Lin, Xiaotian Zhu, and Kyoung Joon Oh.

Finally, and most importantly, I want to acknowledge the love and patience of my husband, Jong-Yih Lin and my family. Being the husband of a professional student might not be the most fulfilling role in life, and his support to let me indulge my scientific curiosity and to finish a doctoral training in the States is greatly appreciated. I deeply thank my family, especially my parents, for their endless support and encouragement. It is only because of them that I am able to continue pursuing my dreams and doing something positive.

Abstract

Fasciclin I is a cell surface glycoprotein thought to be involved in growth cone guidance in the embryonic insect nervous system. It is expressed on the cell surfaces of all peripheral nervous system (PNS) axons, a subset of central nervous system (CNS) axons and on some nonneuronal cells. Fly embryos bearing mutations eliminating expression of both fasciclin I and the Abelson tyrosine kinase exhibit a severe phenotype in which many axon pathways fail to form. Fasciclin I mediates homophilic adhesion in transfected tissue culture cells, suggesting that it may affect growth cone guidance through homophilic interactions. To facilitate structure-function studies of fasciclin I, we have generated mammalian (CHO) cell lines expressing fasciclin I at a high level. The expressed fasciclin I protein was released from the cell surface in a soluble form by phospholipase C treatment. Milligram quantities of soluble expressed fasciclin I were purified on an immunoaffinity column. Large single crystals were obtained that diffracted to ~5 Å resolution which is insufficient for a structure determination to atomic resolution by x-ray crystallography. In an effort to produce a form of fasciclin I more amenable to crystallization, we also generated CHO and *Drosophila* cell (S2) lines that produce a truncated form of fasciclin I. The soluble fasciclin I expressed in S2 cells contains significantly less carbohydrate (~15 kDa) as compared to the molecules expressed in CHO cells. Therefore, S2-derived fasciclin I may be more suitable for crystallization. Biochemical characterization of the expressed fasciclin I indicates that fasciclin I exists as a monomer in solution, an observation consistent with homophilic interaction properties only if the interaction is of low affinity. Electron micrographs of fasciclin I suggest that it has a compact rectangular shape with no obvious flexible linker regions, in contrast to what has been seen in electron microscopic studies of other adhesion molecules. Circular dichroism analysis suggests that fasciclin I contains significant amounts of α-helical structure, which together with the electron microscopic results, suggests that its structure is substantially different from the β-

sheet structures predicted for adhesion molecules that are members of the immunoglobulin superfamily *and/or* contain fibronectin III repeats. Future structural and functional studies of fasciclin I will ultimately increase our understanding of neuronal cell surface recognition and axon guidance.

Table of Contents

List of Figures and Tables

CHAPTER 1

INTRODUCTION

Abstract

The nervous system regulates all aspects of bodily function. It is staggering in its complexity. In humans, there are about 10^{12} nerve cells in the brain, each with extensive wiring for an appropriate communication. How do neurons follow specific pathways and recognize their targets during development? This chapter first discusses a highly specialized and versatile navigational machine, the growth cone, that makes the pathway decisions, and the labeled pathway hypothesis proposed for growth cone steering. Section II focuses on discussing current knowledge concerning the properties and roles of cell adhesion molecules in vertebrates and in invertebrates underlying axonal migration and guidance. Finally, the primary structure and functions of fasciclin I, a cell surface molecule thought to be involved in neural development will be discussed.

I. Growth Cone Guidance and Labeled Pathway Hypothesis

How neuronal growth cones follow specific pathways and recognize their targets is a central issue in developmental neurobiology. In 1880, Ramón y Cajal already observed and named the growth cone as the structure at the end of embryonic axons. Although he had only fixed histological slides, Cajal proposed the growth cone in these static images as a "sort of club or battering ram, endowed with exquisite chemical sensitivity, with rapid ameboid movements, and with certain impulsive force, thanks to which it is able to proceed forward and overcome obstacles met in its way, forcing cellular interstices until it arrives at its final destination." He further stated "what mysterious forces precede the appearance of these prolongations, promote their growth and ramification, and finally establish the protoplasmic kisses, the intercellular articulations that appear to constitute the final ecstasy of an epic love story." Harrison confirmed Cajal's hypothesis that the outgrowth of axonal processes was not random by observing the extension of growth cones from living frog spinal cord explants in tissue culture (1910). He found that growth cones extended along stereotyped pathways to find and recognize their correct targets.

Advanced technologies developed in recent years, including the use of fluorescent, nontoxic lipophilic dyes to label growing axons in either cell cultures or living embryonic preparations, in combination with the use of confocal fluorescence and video-enhanced microscopy, make possible the observation of axonal navigation with remarkable clarity. From observations of both vertebrate and invertebrate systems, it appears that the growth cone is a highly motile structure composed of a broad, flat expanse of neurite, like a palm of the hand, with many long fingerlike filopodia and a ruffling membrane called a lamellipodium (Fig. 1). As axons grow, their growth cones continually extend and retract their filopodial and lamellipodial processes, in apparent exploration of the microenvironment, presumably for making navigational decisions among "environmental

cues" to steer it along a defined track (for review see Dodd and Jessell, 1988; Hynes and Lander, 1992).

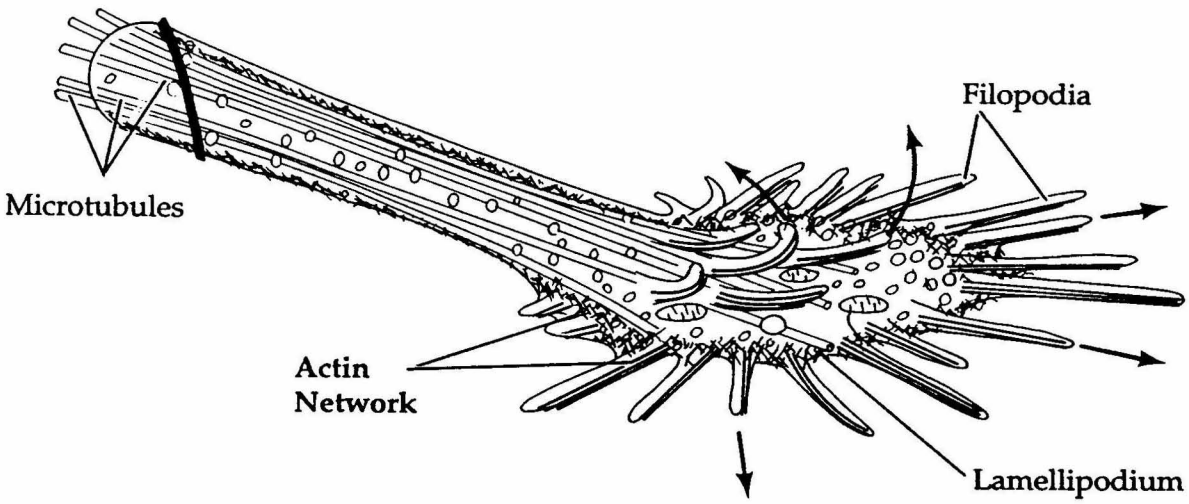

Fig. 1. A schematic view of the growth cone at the end of the axon. Microtubules in the axon are closely packed and axially aligned, but the many fewer and less organized in the growth cone. An actin network is prominent at the leading edge of the growth cone, and bundles of actin filaments fill the filopodia. The central region of the growth cone is rich in mitochondria and vesicles. The arrows indicate that while the leading edge of the growth cone is advancing in one direction, filopodia and lamellipodia are swept back over the surface of the growth cone in the opposite direction.

What are environmental cues and how do they work? To address these questions at the level of individual cellular interactions at a molecular level, the relatively simple nervous system of insect embryos has been used. This is because the mammalian nervous system is composed of hundreds of billions of nerve cells, each with extensive wiring. In contrast, at early stages of grasshopper neuronal development, there are less than 100 neurons in each hemisegment that extend single unbranched axons. Thus, insect embryos provide a useful system to identify and analyze pathway recognition by individual neurons, and to examine the molecular mechanisms involved in neuronal recognition (Goodman et al., 1984; Bastiani et al., 1985). Each developing segmental ganglion in the insect central nevous system (CNS) arises from the neuroepithelium and consists of a ventral layer of neuronal precursor cells, a middle layer of cell bodies of their neuronal progeny, and a dorsal layer of neuronal processes. As each neuron differentiates, it extends a growth cone dorsally into the region called neuropil in which all of the axonal and dendritic processes intertwine and interconnect. The neuropil initially develops as an orthogonal scaffold of axon bundles or fascicles organized into commissures that connect the two sides of a single segment, longitudinal tracts that connect neighboring segments, and nerves that connect the segmental ganglion to the periphery.

Initial insights into how growth cones work have come from a series of laser ablation experiments in grasshopper embryos (Raper et al., 1983a, 1984; Bastiani et al.,1984, 1986; du Lac et al., 1986; Doe et al., 1986), which showed that an identified growth cone, crossing over many bundles, could display a selective affinity for the surface of a specific fascicle of axons among neighboring fascicles (a process called selective fasciculation). In the grasshopper embryo, the long filopodial extensions from individual growth cones can sometimes contact as many as 25 different fascicles (containing in total about 100 different axons), yet these growth cones always choose to extend along one particular bundle of axons (Fig. 2). For example, the G growth cone extends across the embryo all the way on

a particular commissural bundle, and then invariably turns anteriorly on a longitudinal bundle containing the axons of the A and P neurons. If the P axons are eliminated by laser ablation, the G growth cone stalls at the end of the commissure. In contrast, if the A axons are ablated, the G growth cone behaves normally. Similar results were also obtained in a vertebrate CNS, the developing fish spinal cord (Kuwada, 1986). These observations strongly suggest that axon fascicles express different molecular labels that can be read by growing axons. This is called the labeled pathways hypothesis (Raper et al., 1983c, 1984), which predicts that each axon pathway is differentially labeled by surface recognition molecules that allow growth cones to identify it.

Neuronal growth cones must respond to cues directing them to follow the appropriate pathways and then stop at the correct destination. Cell surface molecules are likely to be of primary importance in mediating the recognition of specific extrinsic pathway and target cues. In recent years, an impressive amount has been learned about cell surface and extracellular matrix molecules that may serve as "adhesion molecules" for cell and substrate adhesion as well as axon fasciculation (Jessel et al., 1988). Many of these molecules are related, and can be grouped into families; this will be discussed in the next section. Although a simple model of growth cone navigation based on differential adhesion of growth cones to the substratum has been proposed for pathfinding, other possible mechanisms such as inhibitory factors cannot be excluded (Dodd and Jessell, 1988, Fig. 3). In fact, a large amount of data has emerged suggesting that specific inhibitors of axon growth also exist. The best characterized growth cone-collapsing activity was identified as a component of CNS myelin (Schwab and Caroni, 1988). As growth cones approach fragments of myelin or the surfaces of oligodendrocytes, the contact of one or a few filopodia with myelin results in a collapse of growth cone structure (loss of most filopodia and lamellipodia), as well as a profound and sometimes long-lasting loss of motility (Bandtlow et al., 1990). Collapse of a retinal growth cone upon contact with a particular

sympathetic neurite was also observed (Kapfhammer and Raper, 1987; Raper and Kapfhammer, 1990). Such inhibitory effects limit axon regeneration in adults as well as serve as a guidance function in normal development.

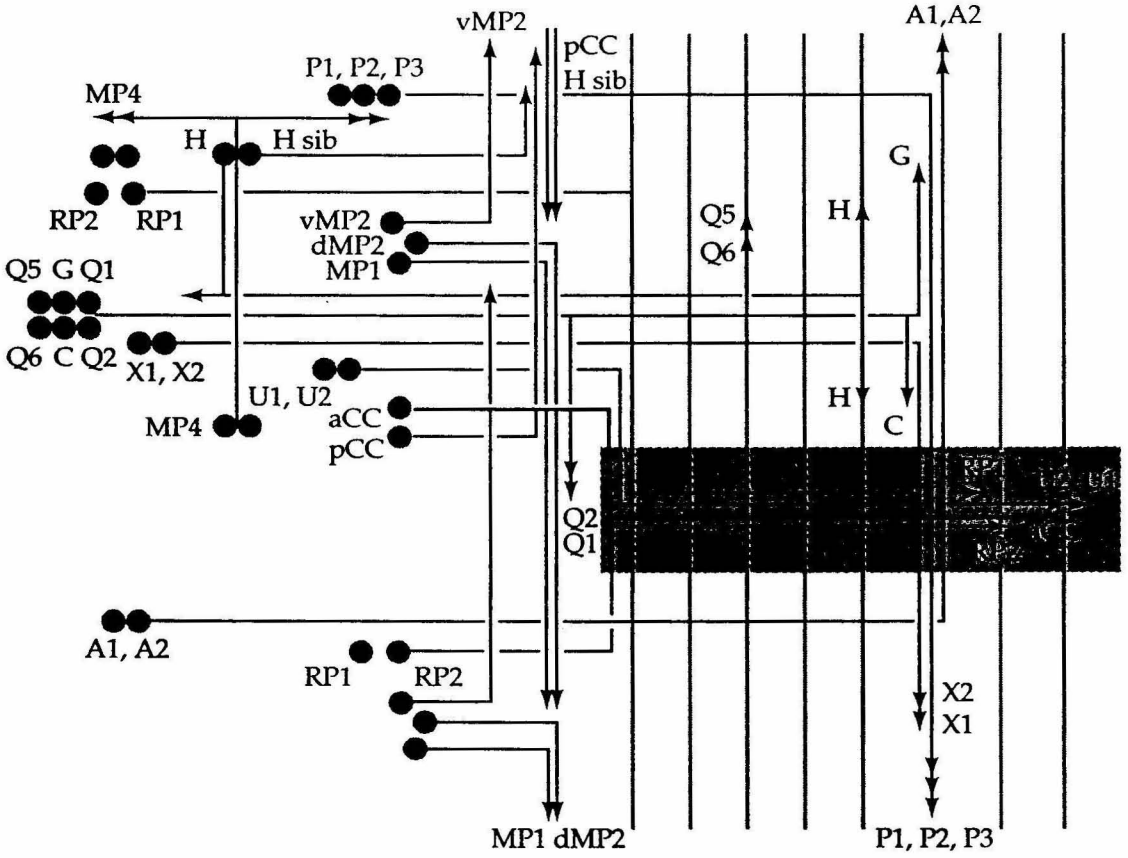

Fig. 2. Schematic diagram of selective fasciculation in the grasshopper embryo. A portion of the growth cones of the identified neurons in one hemisegment is labeled with their distinctive name. The axons form an orthogonal array of bundles by selective fasciculation, displaying the divergent and cell-specific choices. A given neuron, such as aCC, makes multiple choices as it grows, passing by some bundles (such as those formed by pCC, MP1 and dMP2) and joining up and growing along other bundles (such as those formed by U1, U2, RP1, and RP2) (from Goodman et al., 1984).

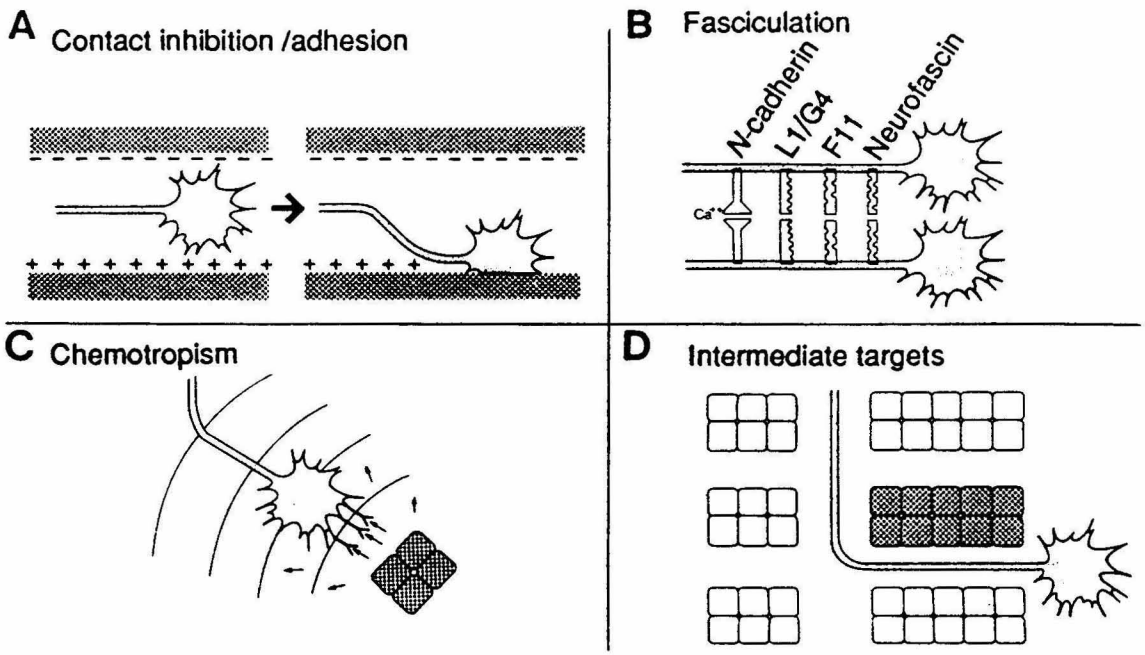

Fig. 3. Growth cones recognize specific guidance cues in their local environment. (A) Cell surface or ECM molecules that inhibit growth cones may contribute to their choice of substrate pathways and to selective fasciculations. (B) Adhesion molecules expressed on developing growth cones and axons may promote selective fasciculation and growth cone extension on preexisting axon tracts. (C) Growth cones may respond to gradients of diffusible molecules that are secreted by discrete cellular targets, thus orienting axonal projections by a chemotropic mechanism. (D) Intermediate cellular targets may mediate abrupt alterations in axonal pathways (from Dodd and Jessell, 1988).

In addition to protein-protein interactions, complex oligosaccharides can also mediate targeting interactions in the nervous system. Proteoglycans, for example, can interact with most of the extracellular matrix attachment proteins and at least two immunoglobulin superfamily adhesion molecules (N-CAM and myelin-associated glycoproteins). It is suggested that at least some proteoglycans may act as barriers to exert inhibitory influences on neural cell migration and axon growth (Hynes and Lander, 1992). It is also clear that N-CAM possesses a unique oligosaccharide, polysialic acid, that is predicted to extend a considerable distance beyond the membrane and may influence fasciculation and nerve-muscle interaction (Rutishauser et al., 1988).

The classes of molecules discussed above seem likely to provide many of the cues that are present on complex pathways along which neurons and growth cones travel. Some of the molecules are observed to be physically associated with cytoskeleton components in actively motile growth cones, suggesting that some receptors which have direct links to the cytoskeleton can alter growth cone motility (Sabry et al., 1991; Tanaka and Kirschner et al., 1991). There is also strong evidence that environmental signals may alter growth cone motility via second messengers, especially calcium whose concentration influences protein phosphorylation and which is required by many actin-binding proteins. Extensive studies in the large growth cones of cultured invertebrate neurites (*Helisoma*, a snail) have been done, using the fluorescent calcium sensitive intracellular dye fura-2 and whole-cell patch clamping to measure membrane potential (Kater and Mills, 1991). It appears that an optimum range of free intracellular calcium is necessary for growth cone extension in these experiments.

The growth cone is clearly a highly specialized and versatile navigational machine. In grasshopper embryo limb buds, O'Connor et al. (1990) observed that the growth cones of Ti1 neurons follow a highly stereotyped pathway (Fig. 4) with different kinds of migration

and steering events. Typically, the growth cone advances by extending veils between filopodia on a relatively homogeneous substrate, as is commonly observed on uniform substrate *in vitro*. At some choice points, for example, when growth cones confront an orthogonal border between substrates of dissimilar affinity, axon growth proceeds by extension of multiple lamellipodia, consolidation of one or a few lamellipodia into the preferred nascent axon, and subsequently retraction of inappropriate filopodia and lamellipodial processes. In contrast, a single filopodial contact with a very high-affinity substrate, such as a "guidepost cell," can rapidly expand in diameter, filling in with axoplasm and abruptly reorient the entire growth cone. Although many molecules that may mediate neural adhesion and recognition have been isolated, the molecular mechanisms that underlie different kinds of growth cone responses remain to be identified. For instance, how are a sufficient number and diversity of molecules expressed at the tip of a thin and elongated structure, what signaling pathways conveyed the external messages into the cells, and how do filopodial contacts withstand the powerful contractile forces? Only when this information is obtained will the mechanisms involved in growth cone behavior become clear.

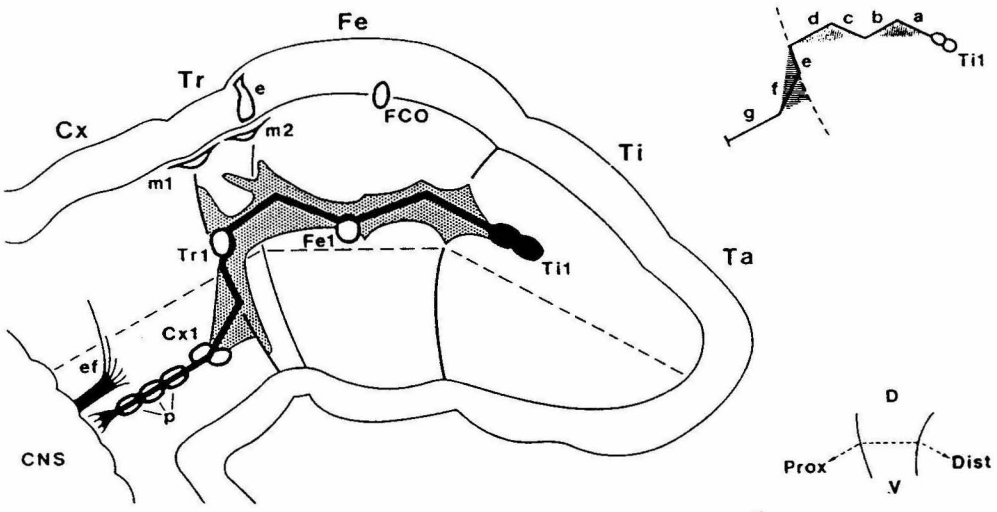

Fig. 4. Growth cones use many cues along their pathways. In grasshopper embryo limb buds, the sibling Ti1 pioneers are the first neurons to initiate axonogenesis. The pioneer gorwth cones migrate from the limb tip to the CNS along a highly stereotyped pathway that involves a series of well-defined alterations in direction (a-f) comprising discrete steering events (from Caudy and Bentley, 1986; O'Connor et al., 1990). The initial path of 50 μm, illustrated as segment a on the right, is reoriented in a proximal-ventral direction as contact is made with the first guidepost cell, neuron Fe1. This forms segment b of the path. At this point, the growth cones orient toward the m1 and m2 cells, forming the c segment. The d segment is formed by a contact with the second guidepost cell, Tr1. From this point, the growth cones turn sharply to follow the segment boundary in the epithelium, forming part e of the pathway. The f segment is formed by a distinct reorientation towards the next guidepost cells, the Cx1 cells. Leaving these cells the Ti1 growth cones follow a direct line into the CNS along the p cells.

II. Cell Adhesion Molecules

A. Families of Cell Adhesion Receptors in the Neuronal Processes

In recent years there have been major advances in identifying numerous cell surface molecules in vertebrates which mediate cell-cell or cell-substrate adhesion for the associations of neural cells (Dodd and Jessell, 1988; Lander, 1989). There are three major classes whose members appear in the different processes performing similar roles (Fig. 5); laminins and integrins (Hynes, 1992), cadherins (Takeichi, 1988, 1990, 1991), and the immunoglobulin superfamily (Jessell, 1988; Grumet, 1991). Another class of adhesion receptors, selectins (Bevilacqua et al., 1991; Lasky and Rosen, 1991), are known to be expressed on the surface of various blood cells and/or endothelial cells, and thus will not be included in this discussion.

Neurons recognize and adhere to molecules in the extracellular matrix via receptors on the surfaces of growth cones and axons, particularly in cell migration during gastrulation, and neural crest migration. One large group of receptors for these matrix proteins is the integrin family. The integrins are transmembrane αβ heterodimers. There are currently about 20 known integrin heterodimers, made up of various pairings of 8 β subunits and 13 α subunits, yielding heterogeneity in binding specificity (reviewed by Hynes, 1987, 1992; Hemler, 1990; Reichardt et al., 1991). Many integrins appear to recognize the Arg-Gly-Asp (RGD) peptide in the context of other sequences of the extracellular matrix (ECM) ligand to which they bind. Each integrin heterodimer has a distinct ligand specificity. A large number of adhesive extracellular matrix glycoproteins, such as fibronectins, laminins, tenascin, and thrombospondin, are known to be ligands of integrins (Lander, 1989). Two of these, laminin and fibronectin, which are major components of peripheral neural and non-neural extracellular matrices, promote neurite outgrowth by several types of neuron

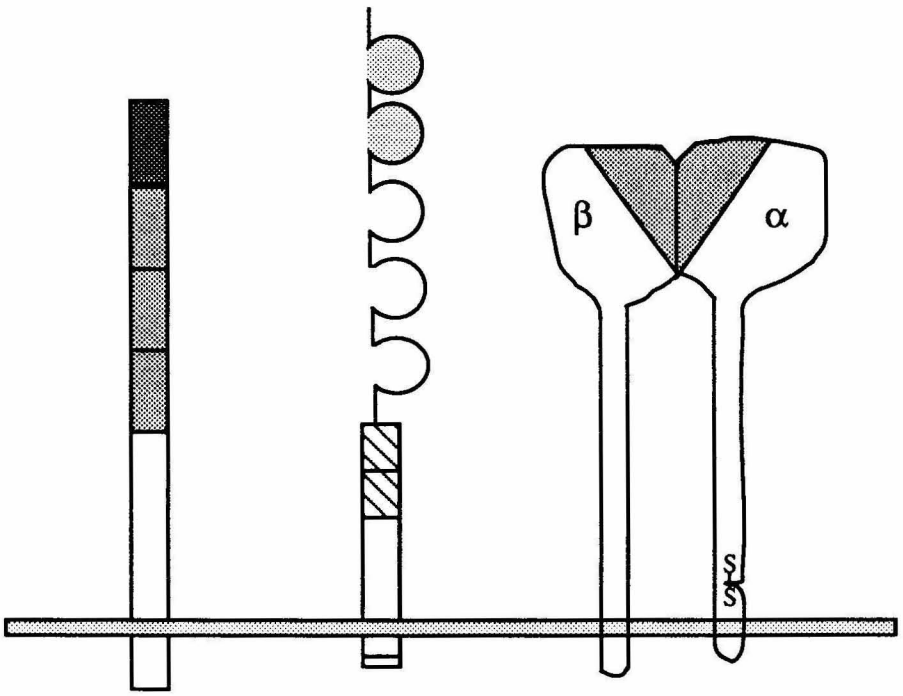

Fig. 5. Major families of cell adhesion receptors in neuronal processes. Cadherins are calcium-dependent homophilic cell-cell adhesion molecules. Immunoglobulin superfamily adhesion receptors contain immunoglobulin domains (half circles) and, frequently, fibronectin type III repeats (cross-hatched boxes). These cell-cell adhesion receptors are calcium-independent and participate in heterophilic and possibly also homophilic interactions. Integrins are heterodimeric receptors, some for extracellular matrix proteins and some for immunoglobulin superfamily counter receptors, and therefore can be involved in both cell-matrix and cell-cell adhesion. Most receptors are transmembrane proteins, and a few are PI-linked. Binding domains are indicated by the darkest colorings (from Hynes and Lander, 1992).

(Thiery et al., 1989). Antibodies to fibronectin block mesodermal migration in amphibians (Boucaut et al., 1984a) and birds (Harrisson, 1989). Similarly, gastrulation can be blocked either by RGD tripeptide (Boucaut et al., 1984b), or by antibodies against integrin subunits (Darribere et al., 1988, 1990). In addition, antibodies directed against a laminin/heparan sulfate proteoglycan complex also perturb cranial neural crest cell migration *in vivo* (Bronner-Fraser and Lallier, 1988). These observations provide evidence that ECM components promote neural crest migration *in vivo*, and that the appropriate expression of integrins and ECM molecules by crest cells is important for this migration.

When visualized by the electron microscope, integrins appear to be rod-shaped molecules extending about 20 nm from the membrane, with globular domains at their extracellular termini that include both α and β subunits (Nermut et al., 1988). This globular domain binds the ECM ligand, and the ligand-binding function depends on divalent cations (calcium, magnesium and manganesium) (Gailit and Ruoslahti, 1988). Each subunit also has a transmembrane and a cytoplasmic domain. Upon binding of ligand, integrins appear to transduce signals across the membrane, possibly by regulating second messenger levels as well as by direct interactions with cytoskeletal proteins. Among the cytoskeletal proteins thought to interact with the cytoplasmic domain of the integrins are talin, vinculin and α-actinin, each of which is found in neuronal growth cones (reviewed in Burridge et al., 1988).

Cadherins are encoded by a functionally related gene family and at least a dozen different cadherins are known that all share elements of sequence and structural homology (Takeichi, 1988, 1990, 1991). The various types of cadherin (E-type, P-type, and C-cadherin) cause aggregation in the presence of calcium when expressed on transfected L cells (Nagafuchi et al., 1987; Nose et al., 1988; Takeichi, 1988, 1990), and moreover, cells expressing different types of cadherins appear to be able to segregate from each other

to form separate aggregates. These results suggest that cadherins mediate calcium-dependent cell-cell adhesion in a homophilic manner and may function in the segregation of different cell types. Analyses of the predicted primary structures reveal that cadherins are integral membrane glycoproteins composed of four homologous domains of about 100 amino acids each in the extracellular domain and a highly conserved cytoplasmic domain (Takeichi, 1990). The determinants for binding specificity have been mapped to the most distal repeat using deletion mutants, domain swaps between E- and P-cadherins, and point mutants (Nose et al., 1990). Furthermore, truncation of E-cadherin's cytoplasmic domain resulted in a loss of cell adhesion, indicating that the conserved cytoplasmic region of cadherins may interact with cytoskeletal components (Nagafuchi and Takeichi, 1988). This agrees with the observation that E-cadherins colocalize with actin filaments (Boller et al., 1985; Hirano et al., 1987; Ozawa et al., 1990). Precisely how cadherins are controlled by cytoskeletal elements is still unknown. However, two to three molecules are found to associate with E-cadherins (Nagafuchi and Takeichi, 1989; Ozawa et al., 1989; McCrea and Gumbiner, 1991), which may be links to the cytoskeleton, and/or regulate E-cadherin function. These molecules, therefore, may form a structural and functional network essential for cadherin-mediated cell adhesion. It is possible that interactions between individual cadherin molecules are too weak to connect cells, since there is no observation of any dimer or oligomer formation in the extracellular domain of cadherin (Takeichi, 1990). Thus, aggregation of cadherins laterally at cell-cell contact site may be necessary to allow an adhesion event to occur. This might be achieved from support by cytoskeleton to increase these lateral interactions of cadherins. Alternatively, cytoskeleton might anchor these molecules to the cell-cell contact sites. It is interesting to note that T-cadherin, a member of the family, apparently has no cytoplasmic domain (Ranscht and Bronner-Fraser, 1991). Instead, it is linked to the membrane via a glycosyl-phosphatidylinositol (PI) anchor. Two forms of T-cadherin have been detected and the protein has been found to be expressed in a unique pattern during spinal cord development.

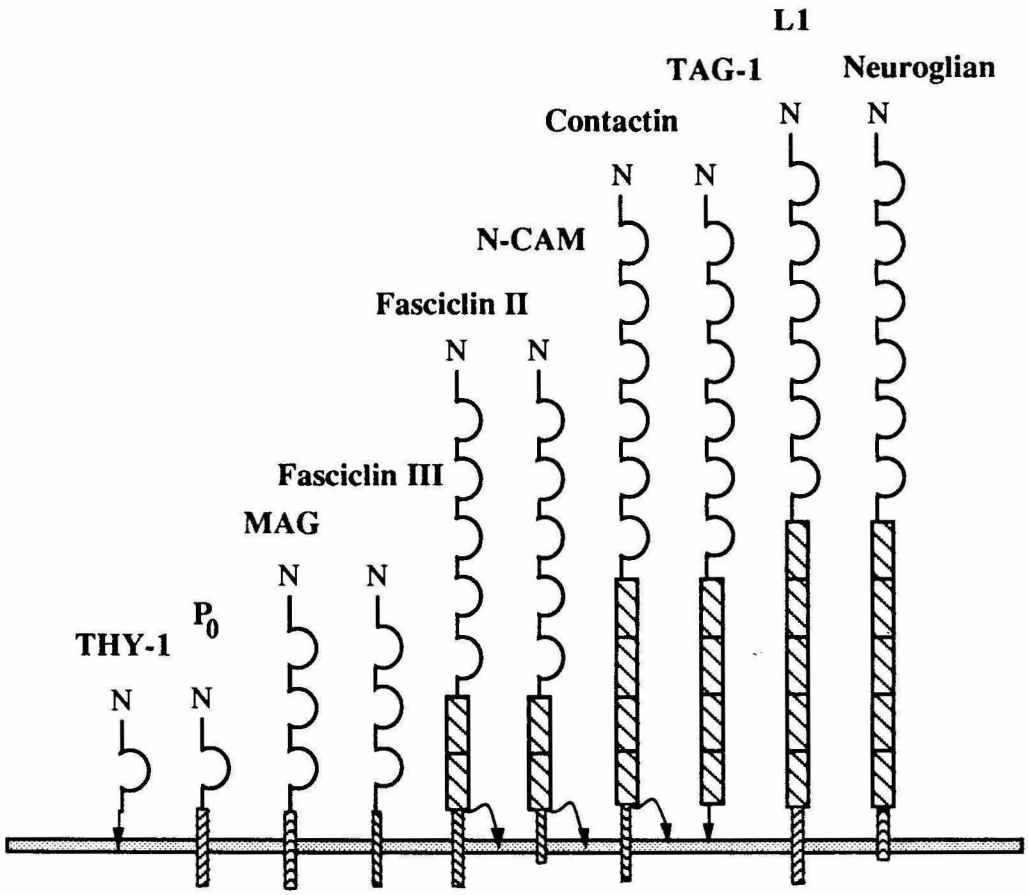

Fig. 6. Schematic diagram of neural molecules that are members of the immunoglobulin superfamily. Many of these have been indicated to function as cell surface adhesion molecules. The immunoglobulin fold domains are indicated by the half circles, and fibronectin type III repeats are indicated by cross-hatched boxes. Thy-1, TAG-1, variant forms of fasciclin II, contactin and N-CAM are attached to the plasma membrane by a phosphatidylinositol linkage as indicated by arrows.

The third group is comprised of members of the immunoglobulin superfamily (Williams, 1987). Each immunoglobulin-like domain consists of approximately 100 amino acids, with the two most highly conserved regions surrounding a pair of cysteines separated by 50-60 residues. Immunoglobulin superfamily proteins are by far the most populous and functionally diverse family of molecules present on the cell surface of lymphocytes as well as neural cells (Fig. 6). Many of these molecules, like N-CAM and L1, are thought to be homophilic and have dramatic effects in assays of cell aggregation and neurite outgrowth. The best evidence for homophilic binding is that transfection of N-CAM cDNA into fibroblast L-cells induces N-CAM dependent cell adhesion (Edelman et al., 1987); transfected L-cells expressing N-CAM aggregated with each other or bound membrane vesicles form chick brain. Both types of binding can be blocked by anti-N-CAM antibodies. Mapping of the binding specificity determinants of N-CAM show them to be in the two most distal immunoglobulin repeats (Frelinger and Rutishauser, 1986; Cole et al., 1986; Cunningham et al., 1987; Reyes et al., 1990). Unlike the cadherin family, members of immunoglobulin superfamily do not have a conserved cytoplasmic domain as depicted in Fig. 6. N-CAM, for instance, has three different isoforms different in the carboxyl terminal (Rutishauser, 1986) that are derived from alternative mRNA splicing; two are transmembrane proteins (180 kDa and 140 kDa), and another (120 kDa) is attached to the membrane via a phosphoinositol (PI) anchor. The polypeptides are expressed at different times and places during development; for instance, the largest N-CAM (180 kDa) first appears after neurulation and only on neural tissue, while the PI-linked form is expressed in significant amounts only later at the time of glial maturation (Cunningham et al., 1987). The functions of different isoforms are poorly understood. Probably the different cytoplasmic domains interact with different subsets of cytoskeletal components. The largest form of N-CAM has been shown to be unable to freely diffuse in the membrane and colocalize with myosin heavy chain in the skeletal muscle at early stages of myotube maturation (Figarella-Branger et al., 1992). Moreover, Ng-CAM, L1 and Nr-CAM have a

highly conserved cytoplasmic domain that may serve this function. Alternatively, they might affect cortical events involving second messengers; for example, changes in protein kinase C levels can affect responses of cells to CAMs (Bixby and Jhabvala, 1990). Moreover, recent studies on N-CAM and N-cadherin demonstrated that they might alter cell phenotype involving activation of intracellular second messenger pathways (Doherty et al., 1991).

The structure of immunoglobulin domains in an antibody (Amzel and Poljak, 1979) is available at an atomic resolution. In addition, the three-dimensional structures of several other members of the immunoglobulin superfamily have been determined, including the human major histocompatibility complex protein, HLA-A2 (Bjorkman et al., 1987), CD4 (Wang et al., 1990; Ryu et al., 1990), CD2 (Driscoll et al., 1991) and CD8 (Leathy et al., 1992). The immunoglobulin fold is best described as two antiparallel β sheets packed tightly against each other. A common feature shared by immunoglobulin domains is that an immunoglobulin domain pairs with another immunoglobulin domain using the same face at almost right angles. Hence, domain-domain interactions analogous to the paired associations of immunoglobulin C domains may provide the structural basis for the homophilic binding of immunoglobulin superfamily molecules (Williams, 1987; Williams, 1988). In addition to the homophilic interaction, there is also clear evidence that N-CAM binds to heparan sulfate proteoglycans (Cole et al., 1986; Reyes et al., 1990). Thus, some immunoglobulin superfamily domains may mediate homotypic adhesion between like cells, whereas others could mediate heterotypic adhesion between two different cell types.

As more structure/function data on these adhesion molecules is generated, it has become clear that many adhesion molecules have a specific binding domain at the distal end of an extended extracellular region. Another interesting biochemical parallel between these three classes of integral membrane proteins, especially cadherins and integrins, is that their

cytoplasmic extensions appear to interact with the cytoskeleton. This suggests that the cytoskeleton may play a primary role in supporting and informing cells of their external contacts. The functions of the extracellular parts of these proteins membrane-proximal to their binding domains are unknown, but their conservation argues that they may play a role in modulation of cell-cell adhesion or in conformational changes and signal transduction events.

B. Adhesion Molecules in *Drosophila*

Although numerous adhesion molecules have been identified as described above, little was known about how these molecules were involved in growth cone guidance, due in part to the complexity of the neuronal network in vertebrates. Invertebrates have somewhat simpler nervous systems. Furthermore, genetic analysis in the fruitfly and nematode may facilitate an understanding of the role of adhesion molecules during the development. In the past four years, an increasing number of genes encoding adhesion molecules in invertebrates have been cloned and characterized. Several approaches have been used to identify adhesion molecules in *Drosophila*: screening monoclonal antibodies against dissected grasshopper and *Drosophila* embryos to search for putative recognition molecules; identification of a gene based on a mutant phenotype; and the use of polymerase chain reaction (PCR) or low stringency hybridization to clone genes by homology to their vertebrate counterparts. The molecules thus far identified are summarized in Table 1 (reviewed by Hortsch and Goodman, 1991). The common picture emerging from these recent studies is that many of the adhesion molecules found in *Drosophila* have considerable structural and sequence homology to the counterparts of vertebrates, indicating that these molecules probably evolved from the same ancestral proteins. Here, we will focus on the adhesion molecules of the fruitfly, *Drosophila melanogaster*, and compare with their vertebrate counterparts.

Table 1. *Drosophila* gene families containing cell and substrate adhesion molecules.

Name of molecule	Number of repeat units		Chromosomal position	Mutants exist?	References
Integrin family molecules					
PS2 α2 chain (*inflated*)[a]			15A1-5	yes	Bogaert et al. 1987
PS3 β1 chain [*lethal (l)myosperoid*]			7D1-5	yes	MacKrell et al. 1988
Laminin molecules	EGF repeats				
Laminin A subunit	(>16)		65A10-11	yes	Montell & Goodman 1988
Laminin B1 subunit	13		28D		Montell & Goodman 1988, 1989; Chi & Hui 1989
Laminin B2 subunit	12		67C		Montell & Goodman 1988
Immunoglobulin (Ig) domain molecules	Ig domains	Fn repeats			
Neuroglian	6	5	7F	yes	Bieber et al. 1989
Fasciclin II	5	2	4B1-2	yes	Grenningloh et al. 1990, 1991
Fasciclin III	3	0	36E1	yes	Snow et al. 1989
Amalgam	3	0	84B1-2	yes	Seeger et al. 1988
Cadherin-like molecules	Cadherin domains				
Fat	34		24D	yes	Mahoney et al. 1991
Dachsous	(>17)		21D	yes	H. Clark et al. (unpublished)
Other cell adhesion molecules					
Fasciclin I	4		89D	yes	Zinn et al. 1988; Elkins et al. 1990b

[a] Boldface names designate molecules that have either been directly shown to be adhesion molecules (e.g. in the S2 cell aggregation assay; Snow et al., 1989) or alternatively, are the *Drosophila* homologues of well known vertebrate adhesion molecules.

By screening a panel of monoclonal antibodies (mAbs) against dissected grasshopper and *Drosophila* embryos in an attempt to identify putative recognition molecules, four candidates were initially characterized: fasciclin I, fasciclin II, fasciclin III, and neuroglian (Bastiani et al., 1987; Patel et al., 1987; Harrelson and Goodman, 1988; Snow et al., 1989; Bieber et al., 1989; Grenningloh et al., 1990). These molecules are expressed on different overlapping subsets of fasciculating axons and growth cones during neuronal development in insects. cDNA clones for these proteins have been isolated and sequenced (Zinn et al., 1988; Grenningloh et al., 1991; Snow et al., 1989; Bieber et al., 1989). Fasciclin II, fasciclin III and neuroglian are shown to be members of the immunoglobulin gene superfamily. Fasciclin II displays a strong structural and sequence homology to N-CAM (Harrelson and Goodman, 1988; Grenningloh et al., 1991), whereas neuroglian is most closely related to L1 (Bieber et al., 1989). Fasciclin III has more divergent immunoglobulin domains and is not closely related to any other known member of this family (Snow et al., 1989). Fasciclin I, on the other hand, has a novel sequence that is unrelated to previously known molecules (Zinn et al., 1988).

To begin a genetic analysis of these molecules, mutations were identified in these four genes (Bieber et al., 1989; Elkins et al., 1990b; Grenningloh et al., 1990). At the gross level, the overall structure of the CNS in embryos develops in a relatively normal way. Recently, fasciclin II mutant embryos were characterized to have a defect in the MP1 pathway by using a specific monoclonal antibody (Grenningloh et al., 1991). In addition, double mutants of fasciclin I and the Abelson *(abl)* tyrosine kinase gene produce a lethal phenotype and show significant disruption in axonal organization (Elkins et al., 1990b). The CNS phenotypes of these null mutations suggest that fasciclins are involved in growth cone guidance.

Drosophila integrin-related proteins were first identified in the early 1980s when Wilcox et al. (1981) and Brower et al. (1984) found that three monoclonal antibodies immunoprecipitated a related group of cell surface antigens called PS1, PS2 and PS3, which are expressed in restricted regions of the *Drosophila* imaginal wing disc. However, it was not until 1987 that PS antigen complexes were demonstrated to be related to vertebrate integrins using N-terminal sequencing (Leptin et al., 1987). The genes encoding PS integrins were cloned and sequenced, which revealed a strong homology to vertebrate integrin sequences (Bogaert et al., 1987; MacKrell et al., 1988). PS1 and PS2 consist of two different *Drosophila* integrin α-subunits, called PS1α and PS2α, whereas PS3 is an integrin β-subunit, called PS3β, which dimerizes with the PS1α and PS2α subunits to form the PS1 and PS2 integrin complexes. The PS antigen ligands have not yet been identified. ECM molecules such as laminin (Fessler et al., 1987; Chi and Hui, 1989; Montell and Goodman, 1988; 1989) have been isolated and sequenced, and they are homologous to vertebrate counterparts. Genetic analysis of *Drosophila* integrin (PS3β) showed defects in muscle attachments and in germ band retraction (Newman and Wright, 1981; Leptin, 1989), indicating a role in tissue morphogenesis for integrins.

To identify cadherin-like molecules in *Drosophila*, degenerate oligonucleotide primers for PCR were used to search *Drosophila* cadherin domains (Mahoney et al., 1991). Two different cadherin genes were cloned, the *fat* gene (Mahoney et al., 1991) and the *dachsous* gene (Clark et al., unpublished). The *fat* gene encodes a protein that contains a putative signal sequence 34 tandem cadherin domains, four EGF-like repeats, a transmembrane domain, and a novel cytoplasmic domain that is unrelated to that found in the known vertebrate cadherins (Mahoney et al., 1991). Recessive lethal mutations in the fat locus of *Drosophila* result in hyperplastic, tumor-like overgrowth of larval imaginal discs and alterations in imaginal disc morphogenesis (Bryant et al., 1988), suggesting its function as a tumor-suppressor gene. Given all of these differences, the enormous transmembrane

protein encoded by the *fat* gene is probably not the *Drosophila* homologue of one of the known vertebrate cadherins. Instead, it may represent a new type of cadherin-like molecule of a much larger size. Whether a true vertebrate cadherin exists in *Drosophila* is still unknown. Other adhesion molecules with EGF-repeats, leucine-rich repeats and serine-esterase domain have been cloned and characterized (reviewed by Hortsch and Goodman, 1991). None of them is involved in neuronal development, thus not included in this discussion.

To test the adhesive properties of *Drosophila* surface molecules, a tissue culture approach was developed, using the *Drosophila* Schneider 2 (S2) cell line that grows as a single cell suspension. The idea is to transfect cloned cDNA encoding adhesion molecules into S2 cells under the control of an inducible promoter and then examine if large cell aggregates are formed after induction (Fig. 7). The ability to confer cell aggregation is presumably attributed to the potential adhesive properties of this molecule. This type of experiment was originally established by Takeichi and his colleagues to test the homophilic activities of cadherins (Nagafuchi et al., 1987) and was adapted to test the adhesive activities of *Drosophila* fasciclin III (Snow et al., 1989). Many molecules are defined as homophilic using this assay; fasciclin III (Snow et al., 1989), fasciclin II, neuroglian (Grenningloh et al., 1990), fasciclin I (Elkins et al., 1990a), and chaoptin (Krantz and Zipursky, 1990). Notch and Delta, molecules with EGF repeats, on the other hand, can interact with each other in a heterophilic fashion (Fehon et al., 1990).

The initial identification of mutations in genes encoding these molecules that have adhesive properties using tissue culture approaches revealed that members of gene families can serve different functions in neuronal development. For example, a null mutation in fasciclin II prevents the formation of the MP1 axon pathway, and double mutants of *fasciclin I* and *Abelson tyrosine kinase* gene result in a lethal phenotype with severe defects

in CNS, suggesting their roles in growth cone guidance. Moreover, null mutations in PS-integrins have a defect in muscle attachments as well as in germ band retraction. The diversity of extracellular motifs and domains found in adhesion molecules, their alternative modes of membrane attachment (transmembrane or PI-linked), and their multiple cytoplasmic domains, suggest that adhesion molecules may play a role in specific cell signaling as well as in simple adhesion. We do not understand how these molecules might function in growth cone guidance at a molecular level. One of the strengths of a genetic approach to this problem is the opportunity to use enhancer and suppressor mutant screens to identify downstream genes that might reveal the intracellular consequences of receptor occupancy, including effects on cytoskeletal organization and other signaling events.

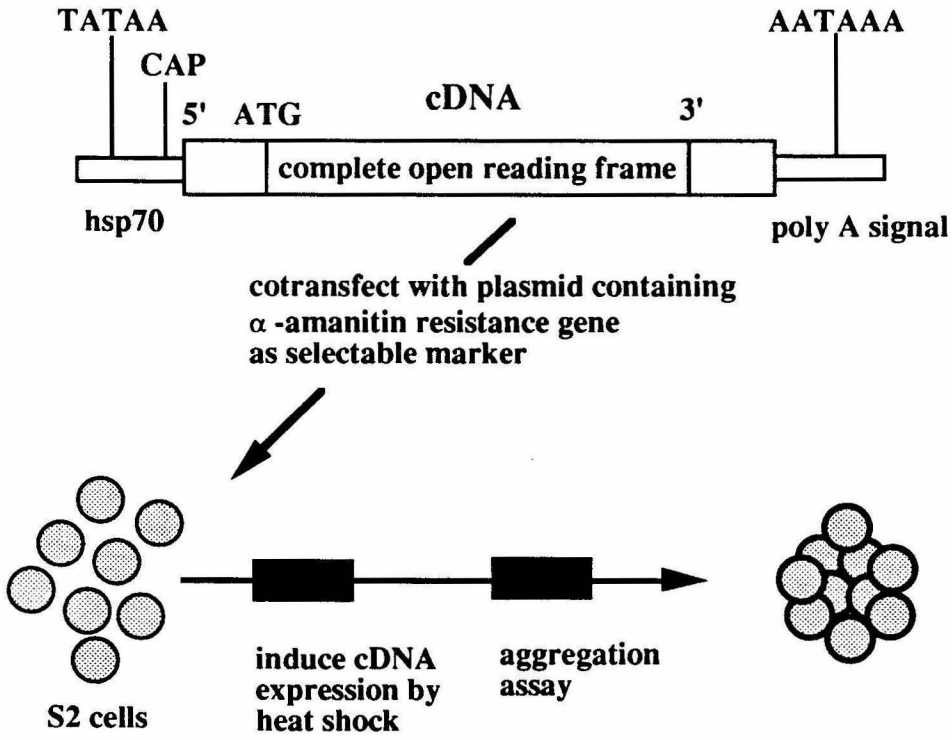

Fig. 7. Schematic diagram showing the strategy of cell aggregation assays. The cDNA clones under heat shock promoter of these molecules were transfected into Schneider 2 cells, induced by heat shock and then were tested positive for their ability to confer cell aggregation. Using this method, several *Drosophila* cell surface molecules have been shown to be cell adhesion molecules (Snow et al., 1989).

III. Fasciclin I

A. Primary Structure

The fasciclin I protein was first identified in the grasshopper embryo as a result in a monoclonal antibody screen for surface antigens expressed on subsets of axons. It is expressed on the surface of all PNS axons, a subset of fasciculating CNS axons, and on some nonneuronal cells during development (Bastiani et al., 1987). Subsequently, grasshopper fasciclin I was cloned and sequenced using oligonucleotide probes encoding N-terminal amino acid sequences of affinity purified proteins (Zinn et al., 1988; Snow et al., 1989). Using this clone in a low stringency screen, a cDNA encoding the *Drosophila* fasciclin I homologue was then isolated (Zinn et al., 1988).

Analysis of cDNAs of grasshopper and *Drosophila* fasciclin I proteins predicts that they have four homologous domains, each of approximately 150 amino acids (Fig. 8). Fasciclin I is a highly glycosylated protein: four potential N-linked glycosylation sites are in *Drosophila* fasciclin I, while six potential N-linked glycosylation sites are in grasshopper fasciclin I (Zinn et al., 1988). *Drosophila* fasciclin I is a 72 kDa protein and grasshopper fasciclin I is a 75 kDa molecule without disulfide bonds (Snow et al., 1989; Hortsch and Goodman, 1990). Fasciclin I is associated with the membrane via a phosphatidyl-inositol (PI) lipid anchor which can be cleaved by PI-specific phospholipase C (PI-PLC) (Snow et al., 1989; Hortsch and Goodman, 1990). The last C-terminal 30 amino acids of the sequence are mainly uncharged and provide the signal for this protein modification. In fact, a variety of cell-surface proteins are found to be attached to the cell membrane via a PI-anchor. These include neuronal proteins such as T-cadherin, Thy-1, TAG1, isoforms of fasciclin II, and NCAM, as well as proteins of the immune system such as CD59, CD55, CD48 and Ly 6 (reviews by Low, 1989; Walsh and Doherty, 1991). These

molecules have been implicated in neuronal development events and in lymphocyte activation after crosslinking with ligand or antibody. However, it is not known how these molecules, which lack intracellular domains, can transduce signals. An interaction of GPI-linked molecules with protein tyrosine kinase has been observed (Stefanova et al., 1991). Other mechanisms have also been proposed such as a relatively high lateral mobility in membranes, GPI anchor degradation and protein release by endogenous enzymes. However, it is still not known what is the biological significance of this anchor present in a large number of cell surface proteins.

B. Adhesion Properties of Fasciclin I

Does fasciclin I function as a neural cell adhesion molecule? This question is addressed by tissue culture experiments as described in the last section (Fig. 7). Transfection of *Drosophila* fasciclin I cDNA under the heat shock promoter into the S2 cell line followed by aggregation assays indicates that fasciclin I may be a calcium-independent homophilic adhesion molecule (Elkins et al., 1990a). When cells expressing fasciclin I are mixed with S2 cells expressing fasciclin III, the cells sort into separate aggregates for either fasciclin I- or fasciclin III-expressing cells, indicating the ability of fasciclin I to mediate selective cell sorting (Elkins et al., 1990a).

Recently, Jay and Keshishian (1990) demonstrated that grasshopper fasciclin I may be involved in axon fasciculation using a new technique, chromophore-assisted laser inactivation (CALI). Fasciclin I protein on the surface of the two pioneer neurons (Ti1) in the developing grasshopper limb was selectively inactivated using this method, which resulted in defasciculation without affecting their growth or guidance. This result agrees with the labeled pathways hypothesis that fasciclin I may function as a specific cue for neuronal recognition during the CNS development. However, a null mutation in fasciclin I

gene in *Drosophila* does not display gross defects in axon fasciculation in the developing PNS (Elkins et al., 1990b). Possibly, fasciclin I plays a role in one of the redundant pathways. Alternatively, CALI might inactivate complexes near fasciclin I in a different way than in a protein null mutation.

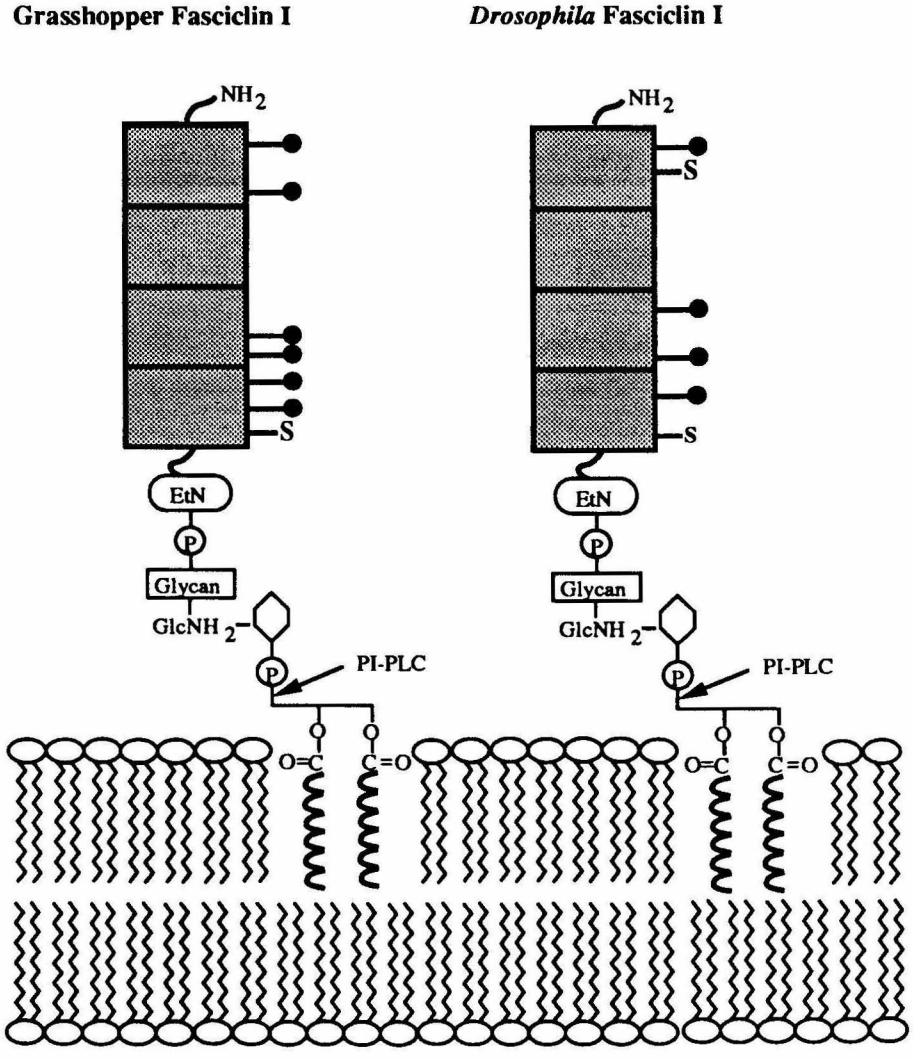

Fig. 8. Schematic diagram showing the primary structure of fasciclin I The protein comprises four homologous domains of approximately 150 amino acids each and is anchored in the membrane by glycophosphatidylinositol (GPI) that can be cleaved by GPI-specific phospholipase C (PI-PLC) indicated by an arrow. Cysteine residue is represented as S and N-linked potential glycosylation site is denoted as —●. EtN, ethanolamine; P, phosphate; glycan, formed by varying size of galscotose and probably attached to the core mannose; and $GlcNH_2$, glucosamine.

C. Fasciclin I May Be Involved in a Signal Transduction Pathway

A genetic analysis of *Drosophila* fasciclin I was done to determine its *in vivo* functions (Elkins et al., 1990b). It appears that a null mutation in the fasciclin I gene does not cause lethality, and that the overall structure of the central nervous system (CNS) in embryos develops in a relatively normal way, which is the same for mutations in fasciclin II gene and fasciclin III gene as discussed above (Grenningloh et al., 1990; Elkins et al., unpublished data). Recently, fasciclin II mutant embryos were characterized to have a defect in the MP1 pathway by using a specific monoclonal antibody (Grenningloh et al., 1991). It is thus possible that better probes may be needed to define a specific defect in fasciclin I mutants. Alternatively, fasciclin I may function in one of the redundant pathways and thus fasciclin I mutants alone do not have a visible embryonic phenotype. In an attempt to explore other molecules together with fasciclin I to be involved in the process of growth cone extension, double mutant combinations in a fasciclin I mutant background were analyzed. One of the double mutants eliminating expression of both fasciclin I and the Abelson *(abl)* tyrosine kinase gene produces a lethal phenotype having a significant disruption in axonal organization (Elkins et al., 1990b). It is known that Abelson tyrosine kinase protein is a membrane-bound cytoplasmic tyrosine kinase which may act as a signaling molecule (Hoffmann et al., 1983; Gertler et al., 1989). Therefore, this result suggests that fasciclin I and abelson tyrosine kinase may be involved in "parallel" signal transducing pathways associated with growth cone movement.

Although fasciclin I may function in selective fasciculation as well as signal transduction during axonal guidance, it is unclear how it works at a molecular level. The common picture emerging from all other adhesion molecules is that repeated structural motifs are often found in different gene families. Thus far, there is no structural information at atomic resolution of cell adhesion molecules involved in neuronal

recognition. We are therefore interested to study the structure-function relationships of fasciclin I. As a first step toward this goal, we have focused on obtaining biochemical quantities of protein, since fasciclin I is a rare protein in insect embryos. In Chapter 2, we report the establishment of Chinese Hamster Ovary (CHO) cell lines that express large quantities of fasciclin I using a glutamine-synthetase-based amplifiable expression system. Since fasciclin I is attached to the cell surface by a PI linkage, a soluble form can be derived by treatment with PI-PLC. Identification, purification and characterization of the soluble form of grasshopper fasciclin I will be discussed. The purified grasshopper fasciclin I can be crystallized in space group $C222_1$. However, the crystals diffract anisotropically, with diffraction to 3.5 Å resolution in one direction, and diffraction in the other two directions to only 5 Å resolution. To improve the quality of crystals, several strategies were used: (i) to remove extra carbohydrates by glycosidases. Grasshopper fasciclin I expressed in CHO cells has an extra 15 kDa of carbohydrate which may impede crystallization; (ii) to remove the remainder of the PI tail by genetic engineering; (iii) to express a secreted form of fasciclin I in insect expression system; (iv) to explore other crystallization conditions. In Chapter 3, we report the expression of secreted form in CHO cells and in insect cells. Biochemical and function studies of this protein will be discussed. Lastly, the preliminary structural studies of grasshopper fasciclin I by x-ray crystallography will be described in Chapter 4. Parts of the contents of Chapter 2 to Chapter 4 are described in a manuscript in preparation (W.-C. W., K. Z., and P. J. B.) to be submitted to J. Bio. Chem.

References

Amzel, L. M., and Poljak, R. J. (1979). Three-dimensional structure of immunoglobulins. Annu. Rev. Bioche. *48*, 961-997.

Bandtlow, C., Zachleder, T., and Schwab, M. E. (1990). Oligodendrocytes arrest neurite growth by contact inhibition. J. Neurosci. *10*, 3837-3848.

Bastiani, M. J., and Goodman, C. S. (1984). Neuronal growth cones: specific interactions mediated by filopodial insertion and induction of coated vesicles. Proc. Natl. Acad. Sci. USA *81*, 1849-1853.

Bastiani, M. J., Doe, C. Q., Helfand, S. L., and Goodman, C. S. (1985). Neuronal specificity and growth cone guidance in grasshopper and *Drosophila* embryos. Trends Neurosci. *8*, 257-266.

Bastiani, M. J., du Lac, S., and Goodman, C. S. (1986). Guidance of neuronal growth cones in the grasshopper embryo. I. Recognition of a specific axonal pathway by the pCC neuron. J. Neurosci. *6*, 3518-3531.

Bastiani, M. J., Harrelson, A. L., Snow, P. M., and Goodman, C. S. (1987). Expression of fasciclin I and II glycoproteins on subsets of axon pathways during neuronal development in the grasshopper. Cell *48*, 745-755.

Bevilacqua, M. P., Butcher, E., Furie, B., Gallatin, M., Gimbrone, M., Harlan, J., Kishimoto, K., Lasky, L., McEver, R., Paulson, J., Rosen, S., Seed, B., Siegelman, M., Springer, T., Stoolman, L., Tedder, T., Varki, A., Wagner, D., Weissman, I., and Zimmerman, G. (1991). Selectins: a family of adhesion receptors. Cell *67*, 233.

Bieber, A. J., Snow, P. M., Hortsch, M., Patel, N. H., Jacobs, J. R. , Traquina, Z. R., Schilling, J., and Goodman, C. S. (1989). *Drosophila* neuroglian: a member of the immunoglobulin superfamily with extensive homology to the vertebrate neural adhesion molecule L1. Cell *59*, 447-460.

Bixby, J. L., and Jhabvala, P. (1990). Extracellular matrix molecules and cell adhesion molecules induce neurites through different mechanisms. J. Cell Biol. *111*, 2725-2732.

Bjorkman, P. J., Saper, M. A., Samraoui, B., Bennet, W. S., Strominger, J. L., and Wiley, D. C. (1987). Structure of the human class I histocompatibility antigen, HLA-A2. Nature *329*, 506-512.

Bogaert, T., Brown, N., Wilcox, M. (1987). The *Drosophila* PS2 antigen is an invertebrate integrin that, like the fibronectin receptor, becomes localized to muscle attachments. Cell *5*, 924-940.

Boller, K., Vestweber, D., and Kemler, R. (1985). Cell-adhesion molecule uvomorulin is localized in the intermediate junctions of adult intestinal epithelial cells. J. Cell Biol. *100*, 327-332.

Boucaut, J. C., Darribere, T., Boulekbache, H., and Thiery, J. P. (1984a). Prevention of gastrulation but not neurulation by antibodies to fibronectin in amphibian embryos. Nature *307*, 364-367.

Boucaut, J. C., Darribere, T., Poole, T. J., Aoyama, H., Yamada, K. M., and Thiery, J. P. (1984b). Biologically active synthetic peptides and probes of embryonic development: a competitive peptide inhibitor of fibronectin function inhibits gastrulation in amphibian embryos and neural crest cell migration in avian embryos. J. Cell Biol. *99*, 1822-1830.

Bronner-Fraser, M., and Lallier, T. (1988). A monoclonal antibody against a laminin-heparan sulfate proteoglycan complex perturbs cranial neural crest migration *in vivo*. J. Cell Biol. *105*, 1321-1329.

Brower, D. L., Wilcox, M., Piovant, M., Smith, R. J., and Reger, L. A. (1984). Related cell-surface antigens expressed with positional specificity in *Drosophila* imaginal discs. Proc. Natl. Acad. Sci. USA *81*, 7485-7489.

Bryant, P. J., Huettner, B., Held, L. I., Ryerse, J., and Szidonyo, J. (1988). Mutations at the fat locus interfere with cell proliferation and epithelial morphogenesis in *Drosophila*. Dev. Biol. *129*, 541-554.

Burridge, K., Fath, K., Kelly, T., Nuckolls, G., and Turner, C. (1988). Focal adhesions: transmembrane junctions between the extracellular matrix and the cytoskeleton. Annu. Rev. Cell Biol. *4*, 487-525.

Cajal, S. R. (1960). In *Studies on Vertebrate Neurogenesis*, L. Guth, Translator (Thomas, Springfield, III).

Chi, H.-C., and Hui, C.-F. (1989). Primary structure of the *Drosophila* laminin B2 chain and comparison with human, mouse, and *Drosophila* laminin B1 and B2 chains. J. Biol. Chem. *264*, 1543-1550.

Cole, G. J., Loewy, A., Cross, N. V., Akeson, R., and Glaser, L. (1986). Topographic localization of the heparin-binding domain of the neural cell adhesion molecule N-CAM. J. Cell Biol. *103*, 1739-1744.

Cunningham, B. A., Hemperly, J. J., Murray, B. A., Prediger, E. A., Brackenbury, R., and Edelman, G. M. (1987). Neural cell adhesion molecule: structure, immunoglobulin-like domains, cell surface modulation, and alternative RNA splicing. Science *236*, 799-806.

Darribere, T., Yamada, K. M., Johnson, K. E., and Boucaut, J. C. (1988). The 140-kDa fibronectin receptor complex is required of mesodermal cell adhesion during gastrulation in the amphibian *Pleurodeles waltlii*. Dev. Biol. *126*, 182-194.

Darribere, T., Guida, K., Larjava, J., Johnson, K. E., Yamada, K. M., Thiery, J.-P., and Boucaut, J.-C. (1990). *In vivo* analyses of integrin β_1 subunit function in fibronectin matrix assembly. J. Cell Biol. *110*, 1813-1823.

Dodd, J., and Jessell, T. M. (1988). Axon guidance and the patterning of neuronal projections in vertebrates. Science *242*, 692-699.

Doe, C. Q., Bastiani, M. J., and Goodman, C. S. (1986). Guidance of neuronal growth cones in the grasshopper embryo. IV. Temporal delay experiments. J. Neurosci. *6*, 3552-3563.

Doherty, P., Ashton, S. V., Moore, S. E., and Walsh, F. S. (1991). Morphoregulatory activities of NCAM and N-cadherin can be accounted for by G protein-dependent activation of L- and N-type neuronal Ca^{2+} channels. Cell *67*, 21-33.

Driscoll, P. C., Cyster, J. G., Campbell, I. D., and Williams, A. F. (1991). Structure of a domain of rat T lymphocyte CD2 antigen. Nature *353*, 762-765.

du Lac, S., Bastiani, M. J., and Goodman, C. S. (1986). Guidance of neuronal growth cones in the grasshopper embryo. II. Recognition of a specific axonal pathway by the aCC neuron. J. Neurosci. *6*, 3532-3541.

Edelman, G. M. (1986). Cell adhesion molecules in the regulation of animal form and tissue pattern. Annu. Rev. Cell Biol. *2*, 81-116.

Edelman, G. M., Murray, B. A., Mege, R. M., Cunningham, B. A., and Gallin, W. A. (1987). Cellular expression of liver and neural cell adhesion molecules after transfection with their cDNAs results in specific cell-cell binding. Proc. Natl. Acad. Sci. USA *84*, 8502-8506.

Elkins, T., Hortsch, M., Bieber, A. J., Snow, P. M., and Goodman, C. S. (1990a). *Drosophila* fasciclin I is a novel homophilic adhesion molecule that along with fasciclin III can mediate cell sorting. J. Cell Biol. *110*, 1825-1832.

Elkins, T., Zinn, K., McAllister, L., Hoffmann, F. M., and Goodman, C. S. (1990b). Genetic analysis of a *Drosophila* neural cell adhesion molecule: interaction of fasciclin I and Abelson tyrosine kinase mutations. Cell *60*, 565-575.

Fehon, R. G., Johansen, K., Rebay, I., Artavanis-Tsakonas, S. (1991). Complex cellular and subcellular regulation of Notch expression during embryonic and amiginal development of *Drosophila*: implications for Notch function. J. Cell Biol. *113*, 657-669.

Fehon, R. G., Kooh, P. J., Rebay, I., Regan, C. L., Xu, T., Muskavitch, A. T., and Artavanis-Tsakonas, S. (1990). Molecular interactions between the protein products of the neurogenic loci Notch and Delta, two EGF-homologous genes in *Drosophila*. Cell *61*, 523-534.

Ferguson, M. A. J., and Williams, A. F. (1988). Cell-surface anchoring of proteins via glycosyl-phosphatidylinositol structures. Annu. Rev. Biochem. *57*, 285-320.

Fessler, L. I., Campbell, A. G., Cuncan, K. G., Fessler, J. H. (1987). *Drosophila* laminin: characterization and localization. J. Cell Biol. *105*, 2382-2391.

Figarella-Branger, D., Pellissier, J. F., Bianco, N., Pons, F., Leger, J. J., and Rougon, G. (1992). Expression of various N-CAM isoforms in human embryonic muscles: Correlation with myosin heavy chain phenotypes. J. Neuropath. and Exp. Neur. *51*, 12-23.

Frelinger, A. L., and Rutishauser, U. (1986). Topography of N-CAM structural and functional determinants. II. Placement of monoclonal antibody epitopes. J. Cell Biol. *103*, 1729-1737.

Gailit, J., and Ruoslahti, E. (1988). Regulation of the fibronection receptor affinity by divalent cations. J. Biol. Chem. *263*, 12927-12932.

Gay, D., Maddon, P., Sekaly, R., Talle, M. A., Godfrey, M., Long, E., Goldstein, G., Chess, L., Axel, R., Kappler, J., and Marrack, J. (1987). Functional interaction between human T-cell protein CD4 and the major histocompatibility complex HLA-DR antigen. Nature *328*, 626-629.

Gertler, F. B., Bennett, R. L., Clark, M. J., and Hoffmann, F. M. (1989). *Drosophila abl* tyrosine kinase in embryonic CNS axons: a role in axongenesis is revealed through dosage-sensitive interactions with *disabled*. Cell *58*, 103-113.

Ghysen, A., and Janson, R. (1980). Sensory pathways in *Drosophila*. *In* Development and Neurobiology of *Drosophila*, O. Siddiqi, P. Babu, L. M. Hall, and J. C. Hall, eds. (New York, Plenum Press), pp. 247-255.

Goodman, C. S., Raper, J. A., Ho, R., and Chang, S. (1982). Pathfinding by neuronal growth cones in grasshopper embryos. Symp. Soc. Dev. Biol. *40*, 275-316.

Goodman, G. S., Bastiani, M. J., Doe, C. Q., Lac, S. D., Helfand, S. L., Kuwada, J. Y., and Thomas, J. B. (1984). Cell recognition during neuronal development. Science *225*, 1271-1279.

Grenningloh, G., Bieber, A., Rehm., J., Snow, P. M., Traquina, Z., Hortsch, M., Patel, N. H., and Goodman, C. S. (1990). Molecular genetics of neuronal recognition in *Drosophila*: evolution and function of immunoglobulin superfamily cell adhesion molecules. Cold Spring Harbor Symp. Quant. Biol. *55*, 327-340.

Grenningloh, G., Rehm, E. J., and Goodman, C. S. (1991). Genetic analysis of growth cone guidance in *Drosophila*: fasciclin II functions as a neuronal recognition molecule. Cell *67*, 45-57.

Grumet, M. (1991). Cell adhesion molecules and their subgroups in the nervous system. Curr. Opin. Neurobiol. *1*, 370-376.

Harrison, R. G. (1910). The outgrowth of the nerve fiber as a mode of protoplasmic movement. J. Exp. Zool. *9*, 787-841.

Harrisson, F. (1989). The extracellular matrix and cell surface, mediators of cell interactions in chicken gastrulation. Int. J. Dev. Biol. *33*, 417-438.

Hemler, M. E. (1990). VLA proteins in the integrin family: structures, functions and their role on leukocytes. Annu. Rev. Immunol. *8*, 365-400.

Hirano, S., Nose, A., Hatta, K., Kawakami, A., and Takeichi, M. (1987). Calcium-dependent cell-cell adhesion molecules (cadherins): subclass specificities and possible involvement of actin bundles. J. Cell Biol. *105*, 2501-2510.

Hirokawa, N. (1991). Molecular architecture and dynamics of the neuronal cytoskeleton. In *The Neuronal Cytoskeleton*. R. D. Burgoyne (ed.). Wiley-Liss, New York, pp. 5-74.

Hoffman, F. M., Fresco, L. D., Hoffmann-Falk, H., and Shilo, B.-Z. (1983). Nucleotide sequences of the *Drosophila src* and *abl* homologs: conservation and variability in the *src* family oncogenes. Cell *35*, 393.

Hortsch, M. and Goodman, C. S. (1990). *Drosophila* fasciclin I, a neural cell adhesion molecule, has a phosphoinositol lipid membrane anchor. J. Biol. Chem. *265*, 15104-15109.

Hortsch, M., and Goodman, C. S. (1991). Cell and substrate adhesion molecules in *Drosophila*. Annu. Rev. Cell Biol. *7*, 505-557.

Hynes, R. O. (1987). Integrins: a family of cell surface receptors. Cell *48*, 549-554.

Hynes, R. O., and Lander, A. D. (1992). Contact and adhesive specificities in the associations, migrations and targeting of cells and axons. Cell *68*, 303-322.

Hynes, R. O. (1992). Integrins: versatility, modulation, and signaling in cell adhesion. Cell *69*, 11-25.

Jessell, T. M. (1988). Adhesion molecules and the hierarchy of neural development. Neuron *1*, 3-13.

Kapfhammer, J. P., and Raper, J. A. (1987). Collapse of growth cone structure on contact with specific neurites in culture. J. Neurosci. *7*, 201-212.

Kater, S. B., and Mills, L. R. (1991). Regulation of growth cone behavior by calcium. J. Neurosci. *11*, 891-899.

Kornfeld, R., and Kornfeld, S. (1985). Assembly of asparagine-linked oligosaccharides. Annu. Rev. Biochem. *54*, 631-664.

Krantz, D. E., and Zipursky, S. L. (1990). *Drosophila* chaoptin, a member of the leucine-rich repeat family, is a photoreceptor cell-specific adhesion molecule. EMBO J. *9*, 1969-1977.

Kuwada, J. Y. (1986). Cell recognition by neuronal growth cones in a simple vertebrate embryo. Science *233*, 740-746.

Lander, A. D. (1989). Understanding the molecules of neural cell contacts: emerging patterns of structure and function. TINS *12*, 189-195.

Lasky, L. A., and Rosen, S. D. (1991). The selectins: carbohydrate binding adhesion molecules of the immune system. In *Inflammation*: Basic Principles and Clinical Correlates, J. Gallin, I. Goldstein, and R. Snyderman, eds. (New York: Raven Press).

Leahy, D. J., Axel, R., and Hendrickson, W. A. (1992). Crystal structure of a soluble form of the human T cell coreceptor CD8 at 2.6 Å resolution. Cell *68*, 1145-1162.
and Wilcox, M. (1987). *Drosophila* position-specific antigens resemble the vertebrate fibronectin-receptor family. EMBO J. *6*, 1037-1043.

Leptin, M., Bogaert, T., Lehmann, R., and Wilcox, M. (1989). The function of PS integrins during *Drosophila* embryogenesis. Cell *56*, 401-408.

Low, M. G. (1989). Glycosyl-phosphatidylinositol: a versatile anchor for cell surface proteins. FASEB J. *3*, 600-608.

MacKrell, A. J., Blumberg, B., Haynes, S. R., and Fessler, J. H. (1988). The *lethal myospheroid* gene of *Drosophila* encodes a membrane protein homologous to vertebrate integrin beta subunits. Proc. Natl. Acad. Sci. USA *85*, 2533-2537.

McCrea, P. D., and Gumbiner, B. M. (1991). Purification of a 92-kDa cytoplasmic protein tightly associated with the cell-cell adhesion molecule E-cadherin (uvomorulin). J. Biol. Chem. *266*, 4514-4520.

Mahoney, P. A., Weber, U., Onofrechuk, P., Biessmann, H., Bryant, P. J., and Goodman, C. S. (1991). The *fat* tumor suppressor gene in *Drosophila* encodes a novel member of the cadherin gene superfamily. Cell *67*, 853-868.

Montell, D. J., and Goodman, C. S. (1989). *Drosophila* laminin: sequence of B2 subunit and expression of all three subunits during embryogenesis. J. Cell Biol. *109*, 2441-2453.

Montell, D. J., and Goodman, C. S. (1988). *Drosophila* substrate adhesion molecule: sequence of laminin B1 chain reveals domains of homology with mouse. Cell *53*, 463-473.

Nagafuchi, A., Shirayoshi, Y., Okazaki, K., Yasuka, K., and Takeichi, M. (1987). Transformation of cell adhesion properties by exgenously introduced E-cadherin cDNA. Nature *329*, 341-343.

Nagafuchi, A., and Takeichi, M. (1988). Cell binding function of E-cadherin is regulated by the cytoplasmic domain. EMBO J. *7*, 3679-3684.

Nagafuchi, A., and Takeichi, M. (1989). Transmembrane control of cadherin-mediated cell adhesion: a 94 kDa protein functionally associated with a specific region of the cytoplasmic domain of E-cadherin. Cell Reg. *1*, 37-44.

Nermut, M. V., Green, N. M., Eason, P., Yamada, S. S., and Yamada, K. M. (1988). Electron microscopy and structural model of human fibronectin receptor. EMBO J. *7*, 4093-4099.

Newman, S. M., and Wright, T. R. F. (1981). A histological and ultrastructural analysis of developmental defects produced by the mutation, *lethal (l)myospheroid*, in *Drosophila melanogaster*. Dev. Biol. *86*, 393-402.

Nose, A., Nagafuchi, A., and Takeichi, M. (1988). Expressed recombinant cadherins mediate cell sorting in model systems. Cell *54*, 993-1001.

Nose, A., Tsuji, K., and Takeichi, M. (1990). Localization of specificity determining sites in cadherin cell adhesion molecules. Cell *61*, 147-155.

O'Connor, T. P., Duerr, J. S., and Bentley, D. (1990). Pioneer growth cone steering decisions mediated by single filopodial contains *in situ*. J. Neurosci. *10*, 3935-3946.

Ozawa, M., Baribault, H., and Kemler, R. (1989). The cytoplasmic domain of the cell adhesion molecule uvomorulin associates with three independent proteins structurally related in different species. EMBO J. *8*, 1711-1717.

Ozawa, M., Ringwald, M., and Kemler, R. (1990). Uvomorulin-catenin complex formation is regulated by a specific domain in the cytoplasmic region of the cell adhesion molecule. Proc. Natl. Acad. Sci. USA. *87*, 4246-4250.

Patel, N. H., Snow, P. M. and Goodman, C. S. (1987). Characterization and cloning of fasciclin III: A glycoprotein expressed on a subset of neurons and axon pathways in *Drosophila*. Cell *48*, 975-988.

Ranscht, B., and Bronner-Fraser, M. (1991). T-cadherin expression alternates with migrating neural crest cells in the trunk of the avian embryo. Development *111*, 15-22.

Raper, J. A., Bastiani, M. J., and Goodman, C. S. (1983a). Guidance of neuronal growth cones: selective fasciculation in the grasshopper embryo. Cold Spring Harbor Symp. Quant. Biol. Mol. Neurobiol. *48*, 587-598.

Raper, J. A., Bastiani, M. J., and Goodman, C. S. (1983b). Pathfinding by neuronal growth cones in grasshopper embryos. I. Divergent choices made by the growth cones of sibling neurons. J. Neurosci. *3*, 20-30.

Raper, J. A., Bastiani, M. J., and Goodman, C. S. (1983c). Pathfinding by neuronal growth cones in grasshopper embryos. II. Selective fasciculation onto specific axonal pathways. J. Neurosci. *3*, 31-41.

Raper, J. A., Bastiani, M. J., and Goodman, C. S. (1984). Pathfinding by neuronal growth cones in grasshopper embryos. IV. The effects of ablating the A and P axons upon the behavior of the G growth cone. J. Neurosci. *4*, 2329-2345.

Raper, J. A., and Kapfhammer, J. P. (1990). The enrichment of a neuronal growth cone collapsing activity from embryonic chick brain. Neuron *4*, 21-29.

Reichardt, L. F., and Tomaselli, K. J. (1991). Extracellular matrix molecules and their receptors: functions in neural development. Annu. Rev. Neurosci. *14*, 531-570.

Reyes, A. A., Akeson, R., Brezina, L., and Cole, G. J. (1990). Structural requirements for neural cell adhesion molecule-heparin interaction. Cell Regulat. *1*, 567-576.

Robinson, P. J. (1991). Phosphatidylinositol membrane anchors and T-cell activation. Immunol. Today *12*, 35-41.

Rutishauser, U., Acheson, A., Hall, A. K., Mann, D. M. and Sunshine, J. (1988). The neural cell adhesion molecule (NCAM) as a regulator of cell-cell interactions. Science *240*, 53-57.

Ryu, S.-E., Kwong, P. D., Truneh, A., Porter, T. G., Arthos, J., Rosenberg, M., Dai, X., Xuong, N.-h., Axel, R., Sweet, R. W., and Hendrickson, W. A. (1990). Crystal structure of an HIV-binding recombinant fragment of human CD4. Nature *348*: 419-426.

Sabry, J. H., O'Connor, T. P., Evans, L., Toroian-Raymond, A., Kirschner, M., and Bently, D. (1991). Microtubule behavior during guidance of pioneer neuron growth cones in situ. J. Cell Biol. *115*, 381-395.

Schuch, U., Lohse, M. J., and Schachner, M. (1989). Neural cell adhesion molecules influence second messenger systems. Neuron *3*, 13-20.

Schwab, M. E., and Caroni, P. (1988). Oligodendrocytes and CNS myelin and nonpermissive substrates for neurite growth and fibroblast spreading *in vitro*. J. Neurosci. *8*, 2381-2393.

Snow, P. M., Zinn, K., Harrelson, A. L., McAllister, L., Schilling, J., Bastiani, M. J., Makk, G., and Goodman, C. S. (1988). Characterization and cloning of fasciclin I and fasciclin II glycoproteins in the grasshopper. Proc. Natl. Acad. Sci. USA 85, 5291-5295.

Snow, P. M., Bieber, A. J., and Goodman, C. S. (1989). Fasciclin III: a novel homophilic adhesion molecule in *Drosophila*. Cell *59*, 313-323.

Stefanova, I., Horejsi, V., Ansotegui, I. J., Knapp, W., and Stockinger, H. (1991). GPI-anchored cell-surface molecules complexed to protein tyrosine kinases. Science *254*, 1016-1019.

Tanaka, E. M., and Kirschner, M. W. (1991). Microtubule behavior in the growth cones of living neurons during axon elongation. J. Cell Biol. *115*, 345-363.

Takeichi, M. (1988). The cadherins: cell-cell adhesion molecules controlling animal morphogenesis. Development *102*, 639-655.

Takeichi, M. (1990). Cadherins: a molecular family important in selective cell-cell adhesion. Annu. Rev. Biochem. *59*, 237-252.

Takeichi, M. (1991). Cadherin cell adhesion receptors as a morphogenetic regulator. Science *251*, 1451-1455.

Thiery, J.-P., Duband, J. L., Dufour, S., Savagner, P., and Imhof, B. A. (1989). Roles of fibronectins in embryogenesis. In *Fibronectin*, D. F. Mosher, ed. (New York: Academic Press), pp. 181-212.

Walsh, P. S., and Doherty, P. (1991). Glycosylphosphatidylinositol anchored recognition molecules that function in axonal fasciculation, growth and guidance in the nervous system. Cell Bio. Interna. Rep. *15*, 1151-1166.

Wang, J., Yan, Y., Garrett, T. P. J., Liu, J., Rodgers, D. W., Garlick, R. L., Tarr, G. E., Husain, Y., Reinherz, E. L., and Harrison, S. C. (1990). Atomic structure of a fragment of human CD4 containing two immunoglobulin-like domains. Nature *348*, 411-419.

Wilcox, M., Brower, D. L., Smith, R. J. (1981). A position-specific cell surface antigen in the *Drosophila* wing imaginal disc. Cell *25*, 159-164.

Williams, A. F. (1987). A year in the life of the immunoglobulin superfamily. Immunol. Today *8*, 298-303.

Williams, A. F., and Barclay, A. N. (1988). The immunoglobulin superfamily-domains for cell surface recognition. Annu. Rev. Immunol. *6*, 381-405.

Zinn, K., McAllister, L., and Goodman, C. S. (1988). Sequence analysis and neuronal expression of fasciclin I in grasshopper and *Drosophila*. Cell *53*, 577-587.

CHAPTER 2

EXPRESSION, IDENTIFICATION, PURIFICATION AND CHARACTERIZATION OF FASCICLIN I

Abstract

To facilitate a structure-function analysis of fasciclin I, a glutamine synthetase-based expression system was developed. Chinese Hamster Ovary (CHO) cells were transfected with the full-length cDNA of fasciclin I along with a dominant selectable marker gene, the glutamine synthetase gene. With increasing concentrations of methionine sulfoximine, a potent inhibitor of glutamine synthetase gene function, stable transformant lines were selected and amplified. Using this system, fasciclin I, an insect molecule that is anchored to the membrane by a glycosyl phosphatidylinositol linkage, was expressed in a lipid form at high levels on the surface of mammalian cells. A soluble form of fasciclin I was obtained by treatment of intact CHO cells with phosphatidylinositol-specific phospholipase C. This CHO-derived fasciclin I molecule contains an extra 15 kDa of carbohydrate not found on fasciclin I from insect embryos. Milligram quantities of soluble expressed fasciclin I have been purified on an immunoaffinity column. Biochemical characterization of the purified fasciclin I indicates that it is very charge heterogeneous and proteolytically resistant.

I. Introduction

Crystallization of any protein requires a large amount of that protein (at least 5 mg to start with) in a biochemically pure and functionally active state. Fasciclin I is a very rare protein; only 100 μg can be obtainable from 1 kg of embryos. To overcome this problem, it is necessary to develop a system to express the protein at high levels. Although bacterial expression systems, such as pET-3 T7 expression vectors (Studier et al., 1990), are known to produce milligram quantities of expressed protein, the protein is usually aggregated in the form of inclusion bodies, from which it is difficult to obtain high yields of properly folded protein. Thus, we have chosen to express fasciclin I in eukaryotic cells, using a glutamine-synthetase-based amplifiable expression system developed in Celltech, Berkshire, U. K. (Bebbington, and Hentschel, 1987). This expression system allows amplification of the transfected gene in the presence of a potent inhibitor of glutamine synthetase, methionine sulfoximine (MSX), thereby increasing the levels of protein expression.

In this chapter, the establishment of stable cell lines that express either grasshopper fasciclin I or *Drosophila* fasciclin I at high levels will be described. We show that grasshopper and *Drosophila* fasciclin I can be expressed in a GPI-linked form on the surface of mammalian cells, and can be cleaved quantitatively by phosphatidylinositol-specific phospholipase C (PI-PLC). The molecule contains an extra 15 kDa of carbohydrate not found on fasciclin I from insect embryos. Up to 1 mg of soluble grasshopper fasciclin I protein per daily harvest from a cell line grown on a hollow-fiber bioreactor device has been purified by immunoaffinity chromatography. Finally, we will report some biochemical characterization of the purified grasshopper fasciclin I, including characterization of glycosylation heterogeneity and proteolytic resistance.

II. Materials and Methods

Reagents. Restriction enzymes, Klenow fragment and T4 DNA ligase were obtained from Stratagene, Boehringer Mannheim or New England Biolabs. Endoglycosidase F/N-glycosidase F, neuraminedase, and endo H glycosidase were from Boehringer Mannheim. 6D8, a mouse monoclonal antibody (MAb) against *Drosophila* fasciclin I and 3B11, a mouse monoclonal antibody against native grasshopper fasciclin I. The 3B11 hybridoma cell line that produces 3B11, a mouse monoclonal antibody against grasshopper fasciclin I, and rat anti-grasshopper fasciclin I antiserum were gifts of Michael J. Bastiani (University of Utah). The ascites of the 3B11 antibodies were from Monoclonal Antibody Facility at Caltech. Goat anti-mouse fluorescein-conjugated IgG and goat anti-mouse IgG-peroxidase conjugate were from Cappel Products. Mouse anti-rat IgG-alkaline phosphatase conjugate for Western blots was from Boehringer Mannheim. Anti-mouse *Elite* ABC kits for immunoblots were from Vector Lab. Affi-gel protein A agarose was from Bio-Rad. Protein A-Sepharose was from Pharmacia. Methionine sulfoximine (MSX), phospholipase C were from Sigma. Dulbecco's modified Eagle's medium (DMEM), α-minimum essential medium (αMEM), fetal bovine serum (FBS), dialyzed fetal bovine serum (DFBS), geneticin (G418) and other chemicals for tissue culture were from Sigma, Irvine Scientific and GIBCO/BRL.

Construction of lipid-linked forms of grasshopper fasciclin I and *Drosophila* fasciclin I. Full-length cDNA clones encoding the surface protein fasciclin I from grasshopper and *Drosophila* (Zinn et al., 1988) were subcloned into the EcoRI site of the polylinker of two expression vectors pBJ1 and pBJ5 (Lin et al., 1990), respectively, to generate four plasmids pBJ1.Df1, pBJ5.Df1, pBJ1.Gf1, and pBJ5.Gf1 (Fig. 1). Df1 represents *Drosophila* fasciclin I and Gf1 represents grasshopper fasciclin I. pBJ1 and pBJ5 both contain the SRα promoter (Takebe et al., 1988) composed of the SV 40 early

promoter and the R-U5 segment of the human T-cell leukemia virus type 1 (HTLV-1) long terminal repeats. This promoter is 1 or 2 orders of magnitude more active than the SV40 early promoter (Takebe et al., 1988). The major difference of these two vectors is that pBJ5 has a 16S splice junction while pBJ1 does not.

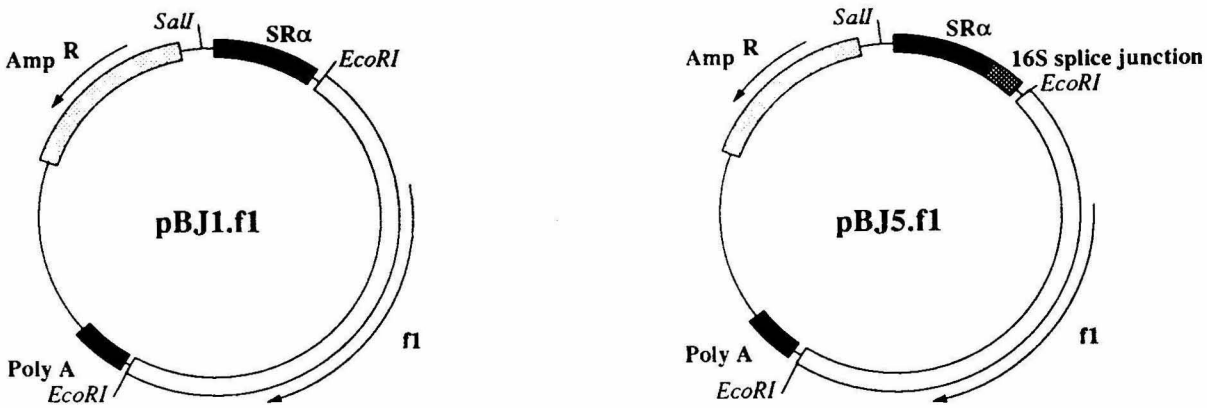

Fig. 1. Expression vectors pBJ1.f1 and pBJ5.f1. pBJ1.f1 and pBJ5.f1 are composed of the full-length cDNAs of either *Drosophila* fasciclin I (3 Kb) or grasshopper fasciclin I (3.16 Kb) under the control of SRα. The construction was done by EcoRI fragment containing fasciclin I cDNA into the EcoRI site of the polylinker of pBJ1 or pBJ5.

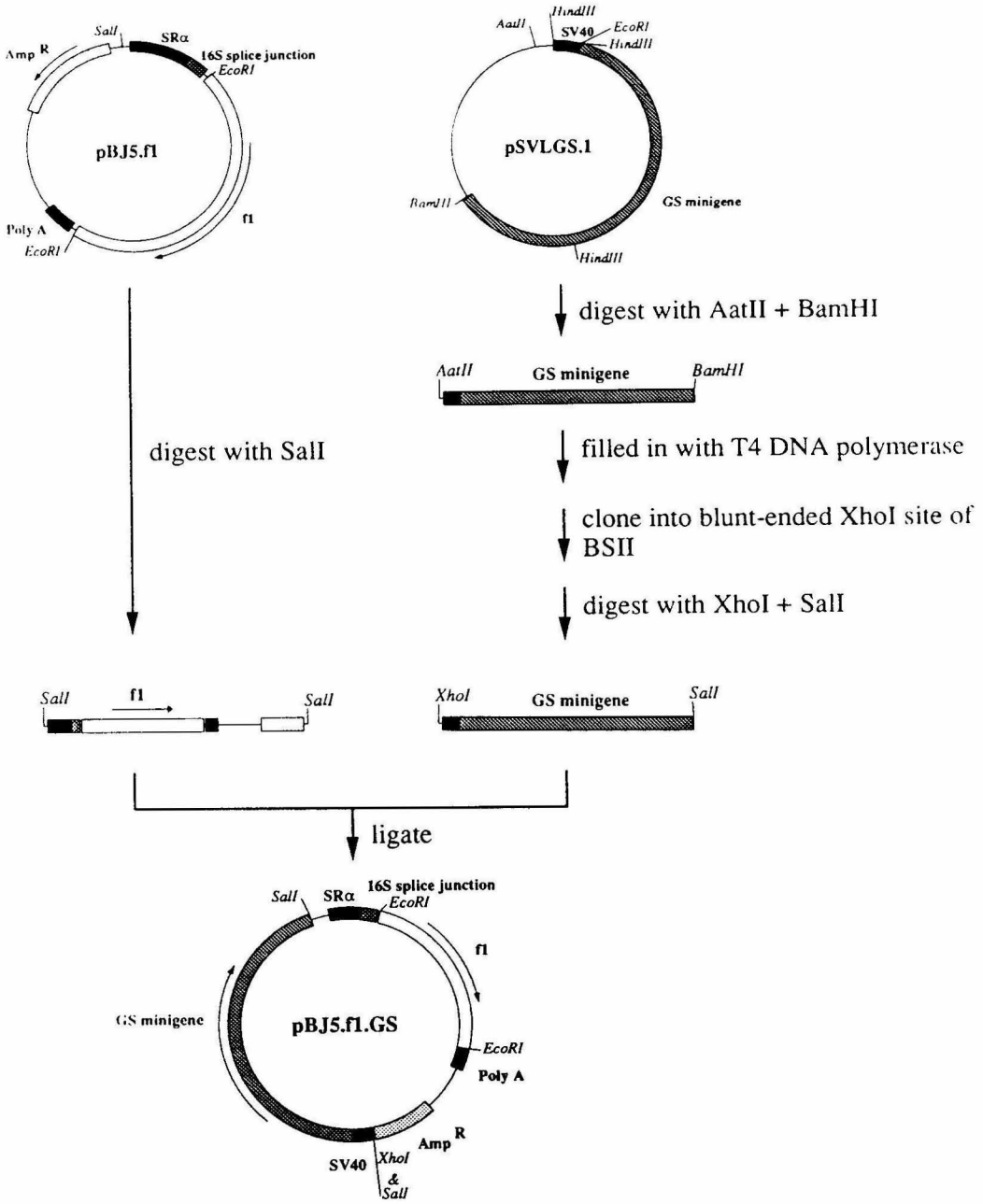

Fig. 2. Construction of pBJ5.GS.f1. pBJ5.GS.Df1 and pBJ5.GS.Gf1 contain the full-length cDNAs of fasciclin I (*Drosophila* or grasshopper) under the control of SRα and the glutamine synthetase minigene from pSVLGS.1.

The glutamine synthetase (GS) minigene from the Celltech vector pSVLGS.1 was excised using AatII and BamHI, the ends filled in using T4 polymerase and the resulting fragment was blunt end cloned into the filled in XhoI site of pBluescript KS(+) (Stratagene). A 5.5 Kb XhoI-SalI fragment containing the GS minigene was then cloned into the unique SalI site of pBJ5.Df1 to generate pBJ5.GS.Df1, or into one of the two SalI sites of pBJ5.Gf1 in a partial digest to obtain pBJ5.GS.Gf1 (Fig. 2). Molecular biological experiments were performed by standard methods (Sambrook et al., 1989).

Cell culture and transfection. COS7 cells were grown in monolayer culture in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum, 2 mM glutamine and 100 U/ml of penicillin/streptomycin (Gibco Labs, or Irvine Scientific). Chinese Hamster Ovary (CHO) K1 cells were grown in aMEM supplemented with 10% FBS, 2 mM glutamine and 100 U/ml penicillin/streptomycin.

For transient transfections, COS7 cells were transfected by the DEAE-dextran method (12 mg DNA/ 60 mm 70% confluent plate) (Gorman, 1985; Ausubel et al., 1989). Mock transfection was performed without plasmid DNA. A 10% DMSO shock for 2 minutes following transfection with DEAE-dextran was employed to increase expression levels. Three days after transfection, the cells were harvested, washed with PBS and suspended in PBS or homogenization buffer (20 mM Tris, pH 7.5, 1 mM Imidazole, 1 mM DTT, 1 mM magnesium chloride, 20 mM sodium diphosphate, protease inhibitor cocktail). The cell extracts then were either subjected to a CAT assay (Gorman et al., 1982) to determine the effectiveness of the transient transfection, or were prepared for Western blot analysis.

For stable transfection using the GS-based system, pBJ5.Df1.GS and pBJ5.Gf1.GS were introduced into CHO cells respectively by lipofectin-mediated method (Felgner et al., 1987). Selection and amplification of the glutamine synthetase gene were carried out

mainly according to the protocol established by Celltech. In brief, 200 μl of DNA-lipofectin complexes (30 μg DNA plus 80-100 μg of lipofectin) were introduced into a 10-cm plate with 50% confluent cells cultured in 1% DFBS, αMEM without glutamine. After 18-24 hours of incubation at 37°C, the medium was replaced with 10% DFCS, αMEM without glutamine and incubation continued for an additional 24 hours. The cells were then split into six 96-well plates and selected with 15 to 30 μM of MSX in 10% DFBS, αMEM without glutamine. MSX-resistant clones appeared and were isolated about 2 weeks later. The positive transfectants were recloned and amplified under 100-500 μM of MSX in 10% DFBS, αMEM without glutamine. The highest expressing cell populations were then selected by immunofluorescence with 3B11 or by a flow cytometeric analysis with cells stained with 3B11. The highest expressing lines were grown in a Cell Pharm I hollow-fiber bioreactor device (Unisyn Fibertec, San Diego; Fig. 3), allowing growth of up to 10^{11} cells in a 250-ml bioreactor.

Cell staining. Fixed cell staining was done by fixing cells in a solution of methanol:acetone (1:1) for 2 minutes. After being washed with PBS; the cells were incubated with the primary antibody solution (dilution of 1:200 to 1:500 with PBS containing 3% BSA) for 1 hour at 25°C and washed thoroughly with PBS. Either horseradish-peroxidase conjugated or fluorescently labeled goat anti-mouse antibody (Cappel Labs) was then added to the cells at a dilution of 1:100 (in PBS, 3% BSA). The cells were developed by DAB-peroxide hydrogen solution as described previously if horseradish peroxidase-conjugated secondary antibody was used.

Alternatively, live cell staining was performed in essentially the same manner as above, except that all procedures were carried out at 4°C and cells were not treated with methanol:acetone. Cells were observed under an inverted microscope equipped with fluorescence (Nikon, Diaphot).

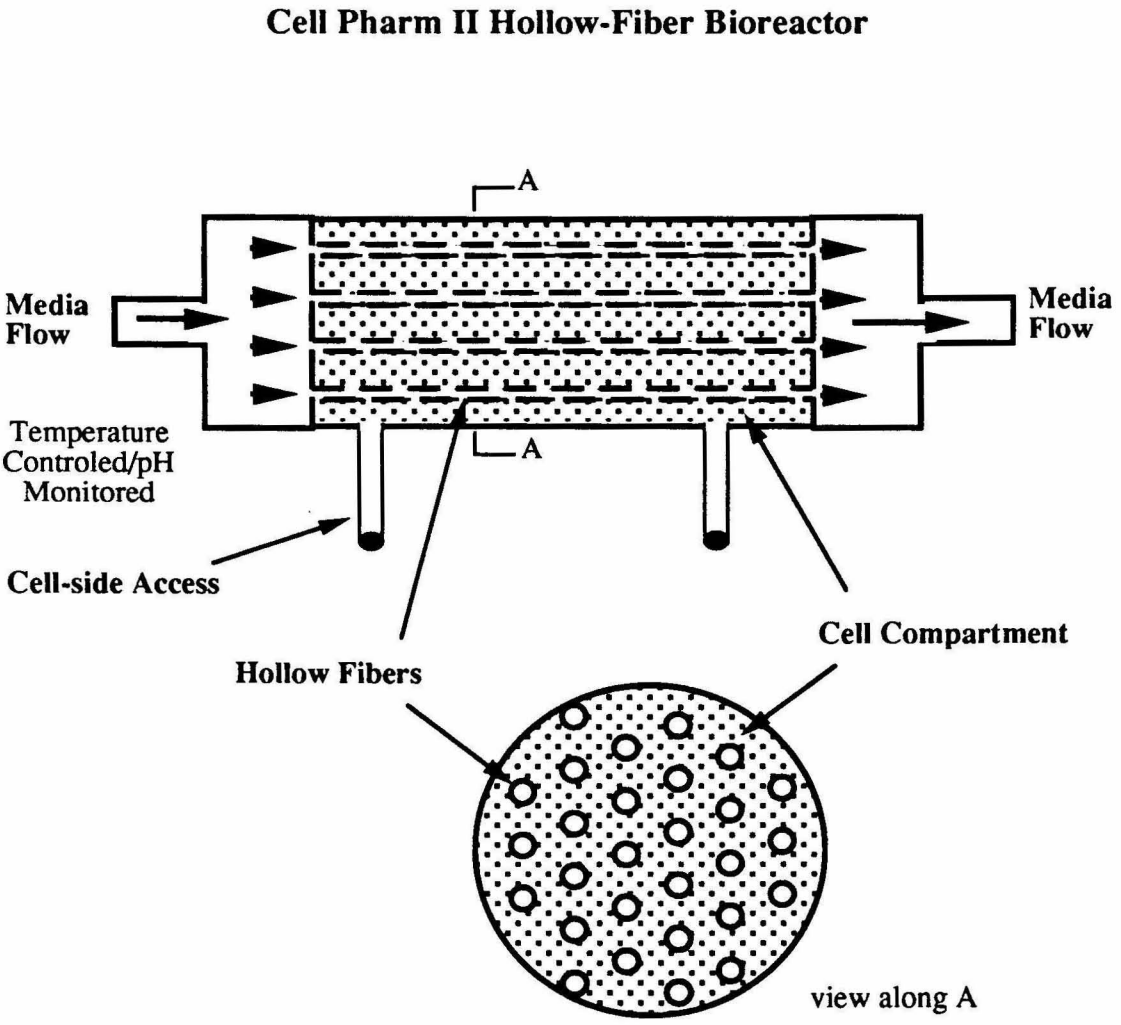

Fig. 3. Schematic diagram of a Cell Pharm II hollow-fiber bioreactor (Unisyn Fibertec, San Diego).

Gel electrophoresis, isoelectric focusing, and Western blot analysis. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was performed according to Laemmli (1972) and silver staining was performed essentially according to Morrissey (1981). Isoelectric focusing was performed as described in Giulian et al. (1984).

After SDS-PAGE, proteins were blotted to nitrocellulose paper according to Towbin et al. (1979). Protein transfer was done at 100 V for 3 hours in transfer buffer (25 mM Tris, 190 mM glycine, 10% methanol). The blots with transferred proteins were immersed in 5% non-fat dry milk in TBST solution (10 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.05% Tween 20) for 30 minutes and then incubated with mouse anti-*Drosophila* fasciclin I MAb (6D8), or rat (or mouse) anti-grasshopper fasciclin I antiserum at dilution of 1:1000 with TBST. After 1 hour, the blots were washed three times with TBST, two times with TBS (10 mM Tris-HCl, pH 8.0, 150 mM NaCl), each for 10-15 minutes and incubated with affinity-purified horseradish-peroxidase-conjugated goat anti-mouse or rat-anti-mouse IgG (1:200, diluted with TBS, Cappel) for 30 minutes. After three washes with TBS, twice with PBS and once with 50 mM Tris, pH 7.6, the blots were visualized using diaminobenzidine (DAB)-hydrogen peroxide staining solution [0.06% DAB (w/v), 0.03% H_2O_2 (v/v) in 50 mM Tris, pH 7.6] until the bands were suitably dark (about 5 minutes). Alternatively, *Elite* ABC kits for immunoblots from Vector Lab were used. All procedures were carried out at room temperature.

Flow cytometric analysis. Trypsinized cells ($2x10^6$ cells per sample) were incubated with MAbs in ascites fluid at a dilution of 1:100 at least 15 minutes on ice, washed three times with PBS and incubated with fluorescein conjugated goat anti-mouse IgG. After 15 minutes, cells were washed three times with PBS, resuspended in 1 ml of PBS containing 0.02% sodium azide and passed through a 50 μM filter. Flow cytometric

analysis was carried out on Ortho Cytofluorograf 50H with a Data General 2150 Computer System.

Deglycosylation of membrane proteins and phospholipase C treatment. Protein from transfected cells and from 10 to 14 hr old of *Drosophila* embryos were deglycosylated with endoglycosidase F carried out according to the manufacturer's specifications (Boehringer Mannheim Biochemica). Deglycosylated protein was analyzed on Western blots.

For analytical assays, PI-PLC from *Bacillus cereus* (Boehringer Mannheim Biochemica) was added to a final concentration of 1 U/ml, incubated for 1 hour at 37°C and the solution was centrifuged. For releasing fasciclin I from cells grown in the hollow fiber bioreactor device, phospholipase C containing PI-PLC (Sigma P6135; 1 mg/ml) was injected into the cell site at a 50-fold dilution for 3 hours at 37°C.

Cold immunoprecipitation, ^{35}S metabolic labeling, and N-terminal amino acid sequencing. Positive cell transfectants expressing PI-linked grasshopper fasciclin I were washed and digested with PI-PLC as described above. Supernatants containing fasciclin I were then reacted with 3B11-protein A beads for 2 hours at room temperature with a gentle agitation (in a nutator). The immunoprecipitates were washed six times with PBS and were electrophoresed on 10% SDS-PAGE.

For metabolic labeling, 10^5 transfected cells expressing PI-linked grasshopper fasciclin I were cultured overnight with 100 μCi ^{35}S methionine. Cells were then washed and digested with PI-PLC. The resultant supernatants were concentrated 10 fold and were electrophoresed on 10% SDS-PAGE. Gels were then fixed and prepared for

autoradiography using the EN^{3}HANCE^{TM} system according to the manufacturer's directions (Dupont).

For N-terminal sequencing, cells of two 175-cm^2 flasks were washed thoroughly and treated with PI-PLC and the supernatants were concentrated 30 fold on a centricon (30K cut off). The protein was isolated on a 10% SDS PAGE gel and electroblotted in a Bio-Rad transblot system onto a polyvinylidede difluoride membrane (PVDF, Millipore) and inserted in an Applied Biosystems Model 4778 sequencer reaction catridge.

Purification of soluble grasshopper fasciclin I by a 3B11 immunoaffinity chromatography and the anionic Mono Q column. The 3B11-protein A coupled immunoaffinity column was prepared essentially according to Harlow and Lane (1988). Briefly, 6 ml of 3B11 ascites (2 mg/ml of 3B11) were adjusted to pH 9.0, 3 M NaCl and then reacted with 5 ml of protein A beads (Pharmacia or Bio-Rad) for 1-2 hours in a nutator. After coupling, beads were washed thoroughly with 50 mM sodium borate, 3 M NaCl, pH 9.0 for four times and were resuspended in 10 volumes of 0.2 M sodium borate, 3 M NaCl, pH 9.0. Crosslinking of 3B11 and protein A beads was done using 20 mM of dimethyl pimelimidate (DMP, Pierce) in 0.2 M sodium borate, 3M NaCl, pH 9.0 for 30 minutes. The reaction was stopped by washing once in 0.2 M ethanolamine (pH 8.0) and incubation was continued for an additional 2 hours. The beads were washed three times with phosphate buffer saline (PBS) and resuspended in PBS with 0.04% sodium azide. All the procedures were carried out at room temperature.

To purify grasshopper fasciclin I, supernatants (about 500 ml) were passed over the column at a flow rate of 20 ml/hr at 4^{o}C, and the column was washed with 500 ml of PBS, 0.05% sodium azide. The fasciclin I protein was then eluted by 0.1 M sodium acetate, pH 3.4. Fractions of 3 ml were collected in test tubes containing 0.5 ml of 1 M phosphate, pH

8.2 to neutralize the elution buffer immediately. Generally, about 1 mg of grasshopper fasciclin I was obtained per daily harvest from the Cell Pharm containing a 250-ml hollow-fiber bioreactor.

For further purification, a solution of protein in 20 mM bisTris propane (pH 7.2) was applied on an FPLC Mono Q column preequilibrated with the same buffer, and eluted with a gradient of NaCl salt (0-150 mM) at a flow rate 1 ml/min. All procedures were performed at room temperature.

Proteolysis. Purified grasshopper fasciclin I expressed in CHO cells was treated with a battery of proteases, including trypsin, chymotrypsin, elastase, and papain. Protein was mixed with each protease at a 100:1 mass ratio in phosphate buffered saline (PBS) solution and incubated at 37°C. Aliquots of digested mixture were withdrawn at different time intervals, for example, 0, 30, 60 and 120 minutes, and the reaction was stopped by adding immediately phenylmethanesulfonyl fluoride (PMSF).

Deglycosylation by treatment with neuraminidase, endo H and endo F. Purified soluble grasshopper fasciclin I (400 μg) was added with different glycosidases in reaction conditions suggested by the Boehringer Mannheim manual: 0.1 U of neuraminidase (Sigma) in PBS, 0.2 U of endo H in 50 mM sodium phospate (pH 5.5), and 2 U of endo F in 0.25 M acetate (pH 6-7) and incubated at 4°C, room temperature or 37°C. Progress of the reaction was monitored as a size decrease on 10% SDS-PAGE by withdrawing aliquots of reaction mixture at different time intervals.

Size-exclusion chromatography and anionic chromatography. Following immunoaffinity chromatography, purified grasshopper fasciclin I was concentrated approximately 0.2-1 mg/ml using a vacuum dialysis apparatus (Schleicher & Schuell) and a

Centricon centrifugal concentrator (Amicon, 25 K cut). Fifty ml of the concentrated protein was subjected to a Superose 12 fast pressure liquid chromatography (FPLC) column equilibrated with PBS.

Purified grasshopper fasciclin I from CHO cells (0.2-1 mg/ml) was dialyzed against 20 mM bisTris propane, pH 7.2 and then loaded onto a 1 ml Mono-Q Sepharose fast pressure liquid chromatography anion-exchange chromatography preequibibrated with 20 mM bisTris propane (pH 7.2). Grasshopper fasciclin I was eluted with a salt gradient from 20 mM bisTris propane (pH 7.2) to 1 M NaCl, 20 mM bisTris propane (pH 7.2).

III. Results

A. Expression of PI-linked Form of Fasciclin I in Chinese Hamster Ovary (CHO) Cells

To express fasciclin I in mammalian cells, full-length cDNAs encoding the grasshopper and *Drosophila* fasciclin I proteins (Zinn et al., 1988) were subcloned into the expression vectors, pBJ1 and pBJ5 (Takebe et al., 1988; Lin et al., 1990) which contain a strong SV40/HTLV-1 hybrid promoter, as well as the SV40 origin of replication. To examine if fasciclin I was properly expressed under the SRα promoter, these four plasmids, pBJ1.Df1, pBJ5.Df1, pBJ1.Gf1, and pBJ5.Gf1 (Fig. 1), were transiently transfected into COS7 cells, in which plasmids bearing SV40 origins are efficiently replicated. Plasmid pSV2CAT was also transfected into COS7 cells to examine the transfection efficiency using a CAT assay (Fig. 4). After 3 days, the transfected cells were analyzed by CAT activity, or by immunofluorescent staining with the MAb 3B11, which recognizes grasshopper fasciclin I in live cell staining, or with the MAb 6D8, which recognizes *Drosophila* fasciclin I in fixed cell staining. As shown in Fig. 4, CAT activity

was observed, indicating that the transfection was successful. 1-5% of the cells displayed bright surface staining (data not shown). Membrane proteins of COS7 cells were analyzed by Western blotting with the MAb 6D8, which recognized *Drosophila* fasciclin I. It appeared that the expressed *Drosophila* fasciclin I in COS7 cells migrated at an apparent weight of 85 kDa compared to ~72 kDa to the *Drosophila* embryo fasciclin I (Fig. 5). After deglycosylation with endoglycosidase F (endo F), *Drosophila* fasciclin I protein from either COS7 cells or from embryos migrated at the same position, indicating the anomalous migration in the expressed fasciclin I is due to a higher extent of glycosylation in COS7 cells. To determine if fasciclin I can be released from the trasfected cells by PI-PLC, transfected COS7 cells were treated with PI-PLC, and samples of supernatant and membranes were analyzed for the presence of fasciclin I by Western blotting with the MAb 6D8. Almost all of the fasciclin I appeared in the supernatant fraction, indicating that most of the fasciclin I could be released from the transfected cells by PI-PLC (Fig. 6).

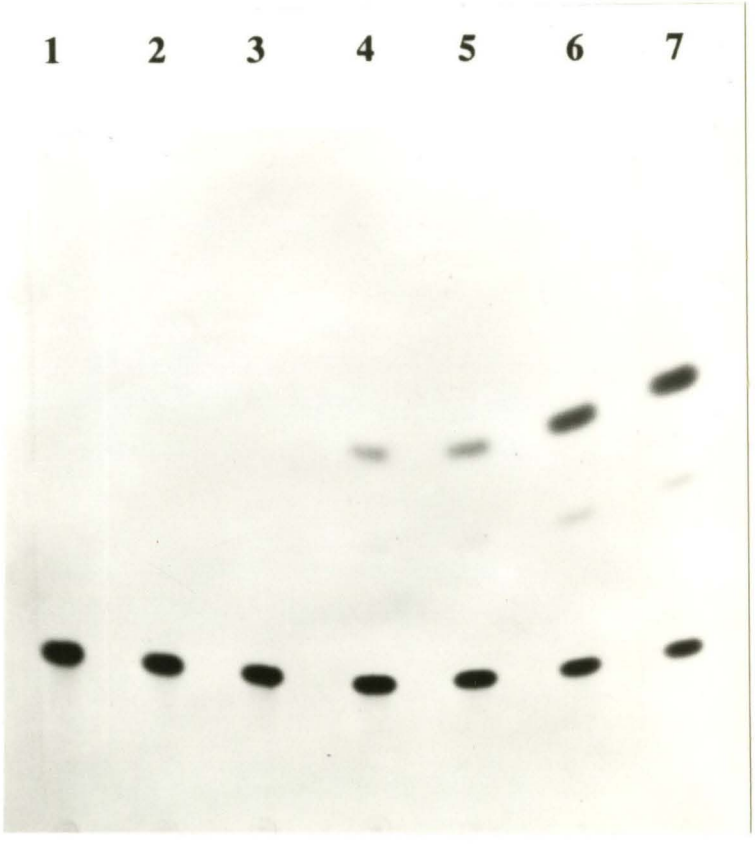

Fig. 4. Effectiveness of transfection into COS7 cells by CAT assays. COS7 cells were transfected with pSV2CAT (lanes 4 and 5), or cotransfected with pSV2CAT and pBJ5.Df1 (lanes 6 and 7) using DEAE-dextran method. Lanes 1 and 2 were non-transfected cell lysates; lane 3 was mock transfected cell lysates.

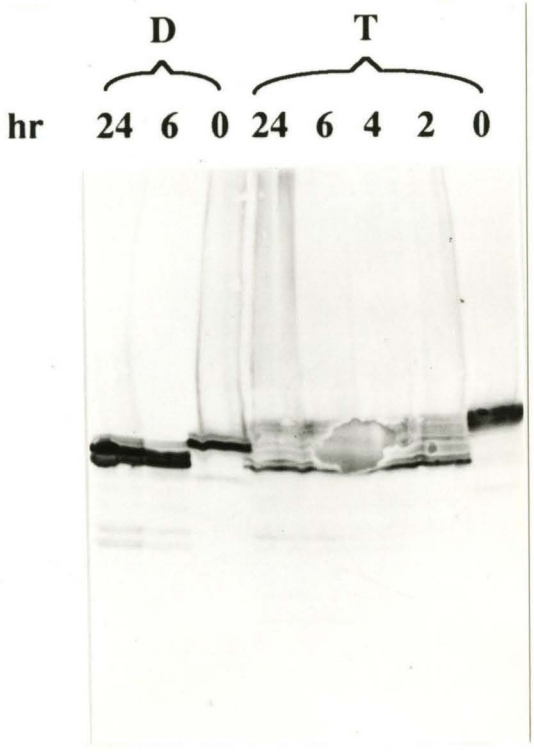

Fig. 5. Western blotting analysis of deglycosylated fasciclin I from transfected COS7 cells or from *Drosophila* embryos. Membrane fractions from transiently transfected COS7 cells (T) with pBJ5.Df1 or from 10 to 15 hr *Drosophila* embryos (D) were treated with endo F for 0 to 24 hr as indicated. Digested proteins were subjected to 10% SDS-PAGE and fasciclin I protein was detected on Western blots by the MAb 6D8.

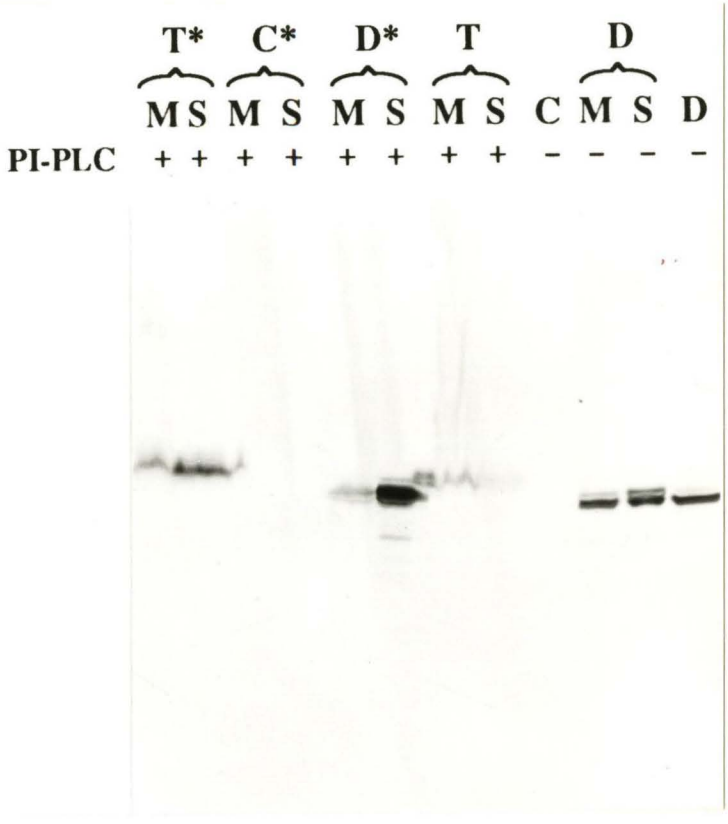

Fig. 6. Western blotting analysis of PI-PLC released *Drosophila* fasciclin I from transiently transfected COS7 cells with pBJ5.Df1 or from membrane fractions of *Drosophila* embryo. Membrane fractions prepared from transfected COS7 cells (T, T*) , from mock transfected COS7 cells (C, C*) or from 10 to 14 hr old *Drosophila* embryos (D, D*) were treated with PI-PLC. Membrane pellets (M) and supernatants (S) were isolated after digestion with PI-PLC, separated on a 10% SDS-PAGE, and analyzed by Western blotting. T*, C* and D* were incubated with PI-PLC, whereas T, C, and D were not.

After demonstrating that fasciclin I could be expressed on the surface of mammalian cells, stable transformant lines were created in Chinese Hamster Ovary (CHO) cells using a glutamine-synthetase-based amplifiable expression system developed by Celltech (Bebbington and Hentschel, 1987). This system allows selection and amplification of the transfected genes in the presence of the drug, methionine sulfoxime (MSX). Two plasmids pBJ5.Df1.GS and pBJ5.Gf1.GS (Fig. 2) were created by subcloning the GS minigene into pBJ5.Gf1 and pBJ5.Df1, and introduced into CHO cells. Stable transformant lines were selected and amplified with 15-30 μM of MSX. 146 MSX-resistant clones transfected with pBJ5.Df1.GS and 88 MSX-resistant clones transfected with pBJ5.Gf1.GS were selected for further amplification with 10-fold increase in the concentration of MSX. Levels of amplification varied, but some clones could be grown in the presence of up to 800 μM of MSX. The positive clones expressing grasshopper fasciclin I were initially screened by immunostaining using the MAB 3B11 (Fig. 7). The highest-expressing clone was then selected by fluorescence activated cell sorting (FACS) and FACS analysis showed that the fasciclin I could be released from these cells by PI-PLC (Fig. 8). Intact cells of the highest-expressing cell line were treated with PI-PLC and analyzed on a 10% SDS-PAGE gel to estimate the expression level; ~10 μg per 10^7 cells were produced, representing 6 x 10^6 molecules per cell. The level of expression of this cell line remains constant for at least 2 months as analyzed by FACS.

We were unable to analyze the *Drosophila* fasciclin I-expressing lines by FACS because none of the existing anti-fasciclin I MAbs, which had been raised against denatured protein expressed in E. coli (Hortsch and Goodman, 1990), were capable of staining the surfaces of live cells. Twenty positive clones expressing *Drosophila* fasciclin I were selected by immunostaining with the fixed cells using the MAb 6D8. However, we did not further pursue purification of *Drosophila* fasciclin I expressed in CHO cells since no MAbs could be used for purification by affinity chromatography. We have since used native

protein released from the CHO-*Drosophila* fasciclin I line to generate MAbs that stain live cells as well as bring down the *Drosophila* fasciclin I using immunoprecipitation (Te-Yi Kung and W-C. W., unpublished results). These MAbs should be useful for future studies.

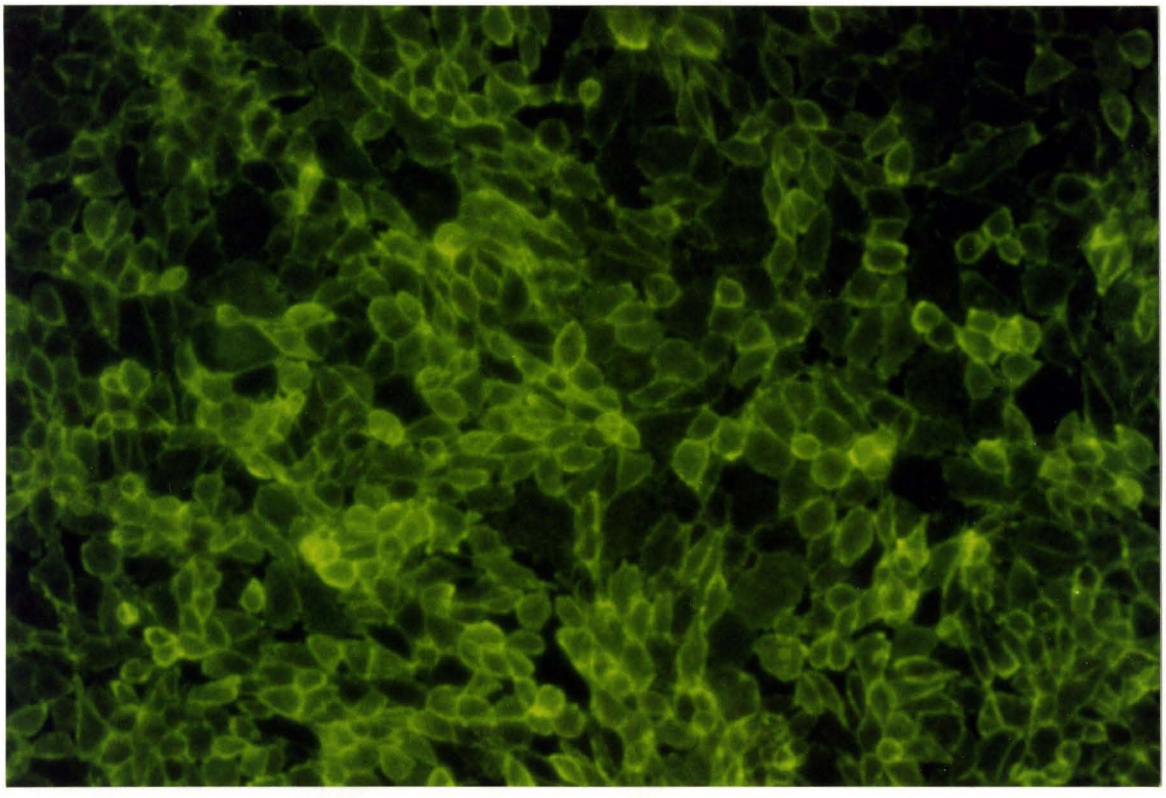

Fig. 7. Immunostaining of live grasshopper fasciclin I-expressing CHO cells by MAb 3B11.

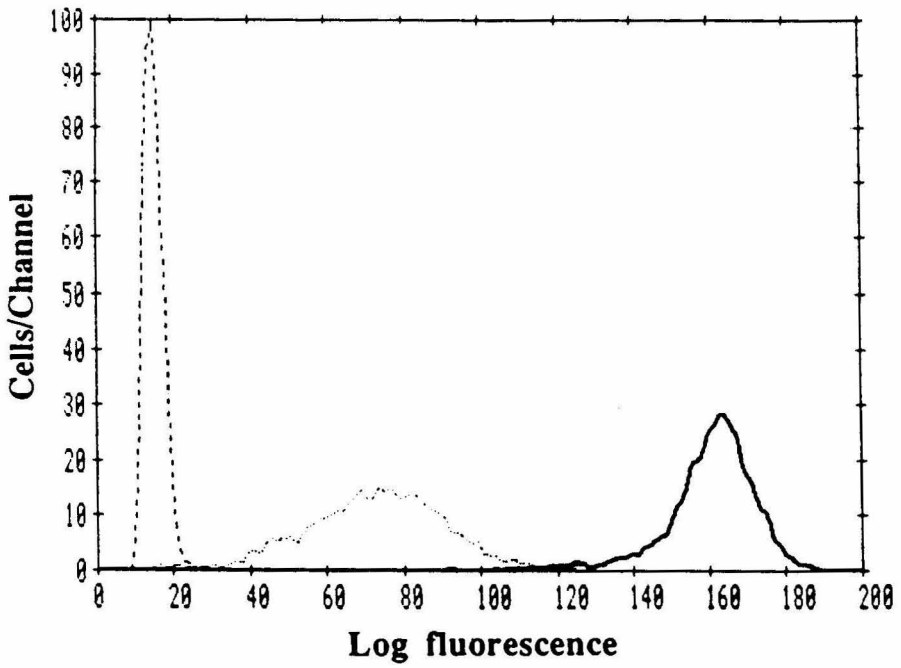

Fig. 8. Flow cytometric analysis of the PI-linked grasshopper fasciclin I proteins on the highest-expressing CHO cell line digested with PI-PLC. Cells were treated with PI-PLC (0.5 U/ml) for 1 hr at 37°C, washed and immunostained with the MAb 3B11. Dashed (- - -) line represents negative control; dotted line (····) represents sample after PI-PLC treatment; solid line (——) represents the highest-expressing CHO cells.

B. Identification of Fasciclin I Expressed in CHO Cells

To identify the CHO-derived grasshopper and *Drosophila* fasciclin I, cells were metabolically labeled with ^{35}S methionine, and supernatants from cells after washing and PI-PLC treatment were analyzed on a 10% SDS-PAGE gel (Fig. 9). The grasshopper and *Drosophila* fasciclin I migrated with apparent molecular masses of 90 and 85 kDa respectively, as compared to PI-PLC released fasciclin I from insect embryos which migrates with an apparent molecular mass of 75 kDa (grasshopper) or 72 kDa (*Drosophila*) (Snow et al., 1988; Hortsch and Goodman, 1990). Cold immunoprecipitation (using the MAb 3B11) or Western blot (using a rat anti-grasshopper fasciclin I antiserum) of PI-PLC released proteins from CHO-grasshopper fasciclin I-expressing cells also showed a 90 kDa band that corresponds to grasshopper fasciclin I (Fig. 10). The 90 kDa band was isolated (~30 μg) after transfer onto a PVDF membrane and subjected to N-terminal amino acid sequencing. The sequence obtained, KGEKSLEYKIRDDPDL, exactly corresponds to the first 16 residues of the mature grasshopper fasciclin I (Snow et al., 1988; Zinn et al., 1988). The difference in molecular weight between CHO-derived grasshopper and *Drosophila* fasciclin I is probably due to the presence of six potential N-linked glycosylation sites in the grasshopper fasciclin I sequence, as compared to four in *Drosophila* fasciclin I (Zinn et al., 1988).

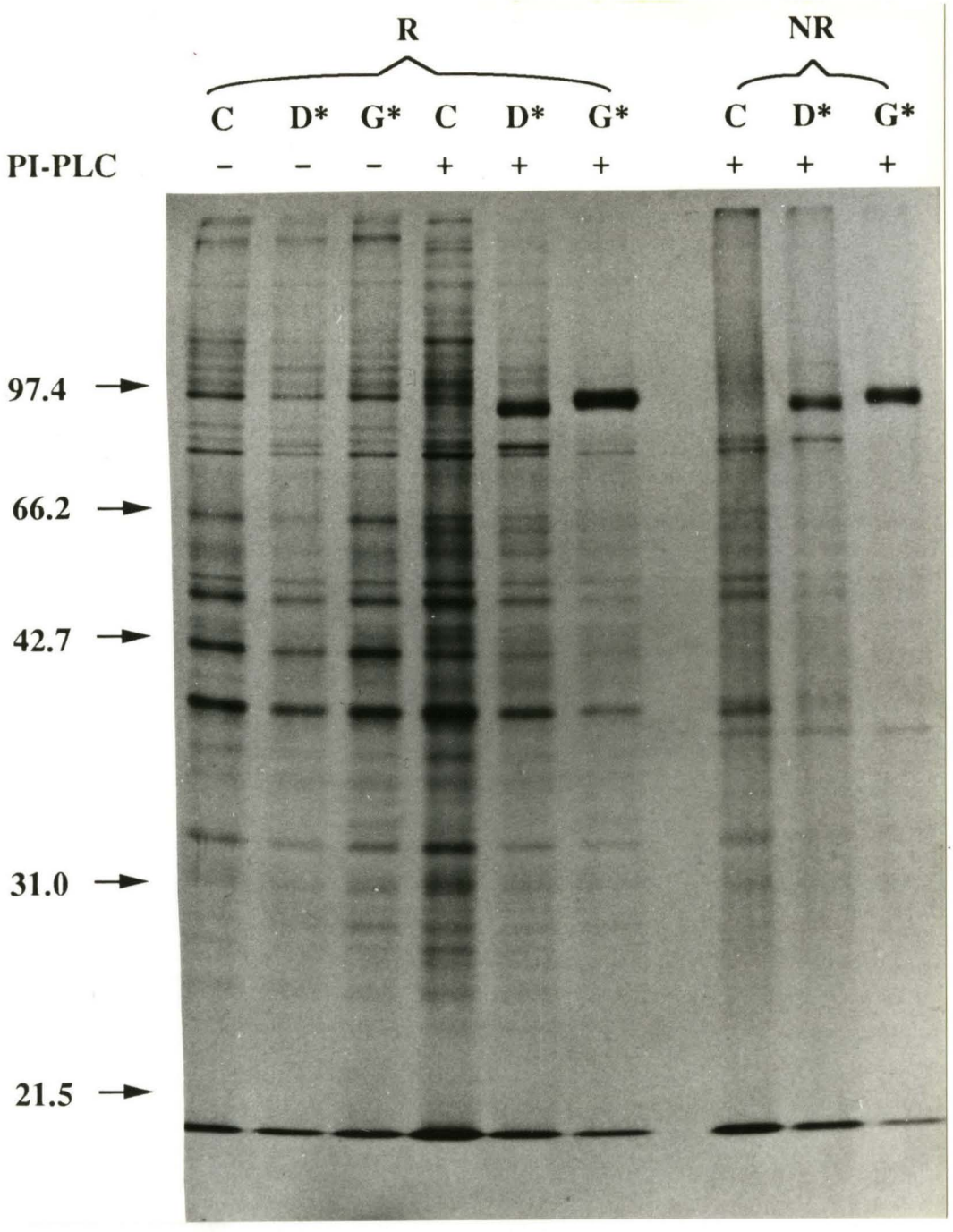

Fig. 9. ^{35}S metabolic labeling analysis of the expressed fasciclin I. PI-PLC released fasciclin I proteins (*Drosophila* or grasshopper) from 10^5 transformed cells that cultured overnight with ^{35}S methionine (100 μCi), were subjected to 10% SDS-PAGE in either reducing (R) or nonreducing (NR) condition for autoradiography. C, original CHO cells; D*, transformed cell line expressing *Drosophila* fasciclin I; G*, transformed cell line expressing grasshopper fasciclin I; -, without treatment of PI-PLC, and +, with treatment of PI-PLC.

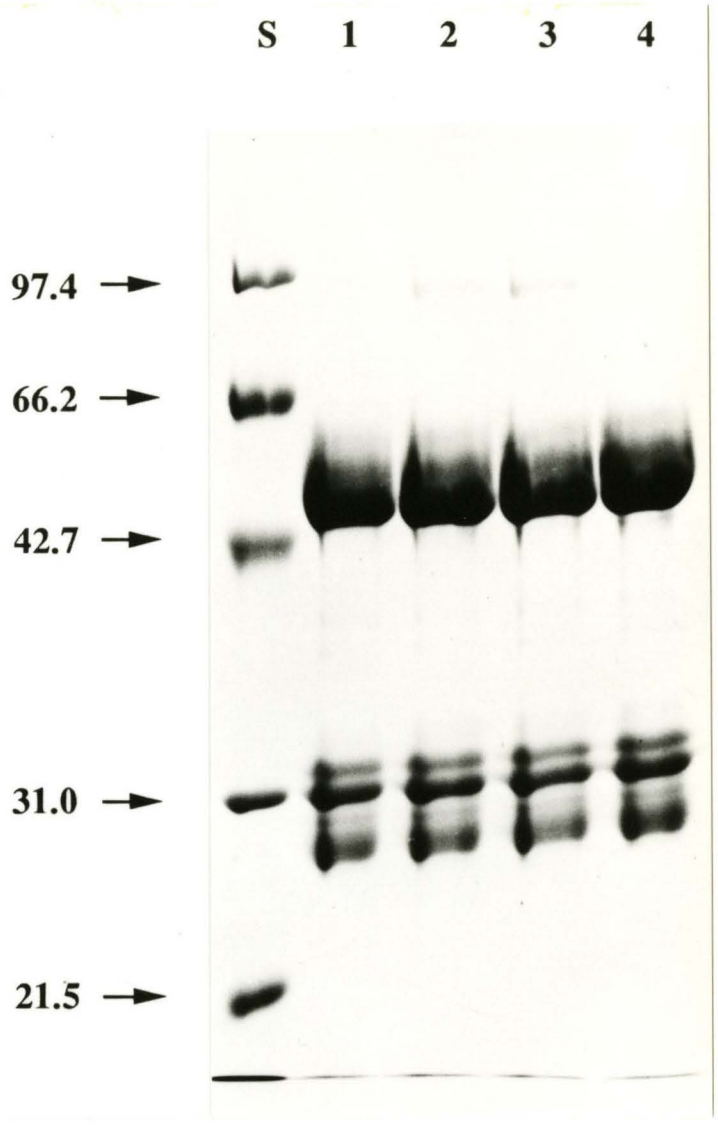

Fig. 10. Cold immunoprecipitation of CHO-derived grasshopper fasciclin I. PI-PLC released material from CHO cells expressing PI-linked grasshopper fasciclin I was reacted with protein G-3B11 beads. The washed immunoprecipitates were subjected to a 10% SDS-PAGE gel. Lane 1 was without PI-PLC treatment; lane 2 (17 mU/ml) and lane 3 (83 mU/ml) were with PI-PLC treatment and lane 4 was a control for interaction of 3B11 and protein G without cell supernatants, and S are protein standards.

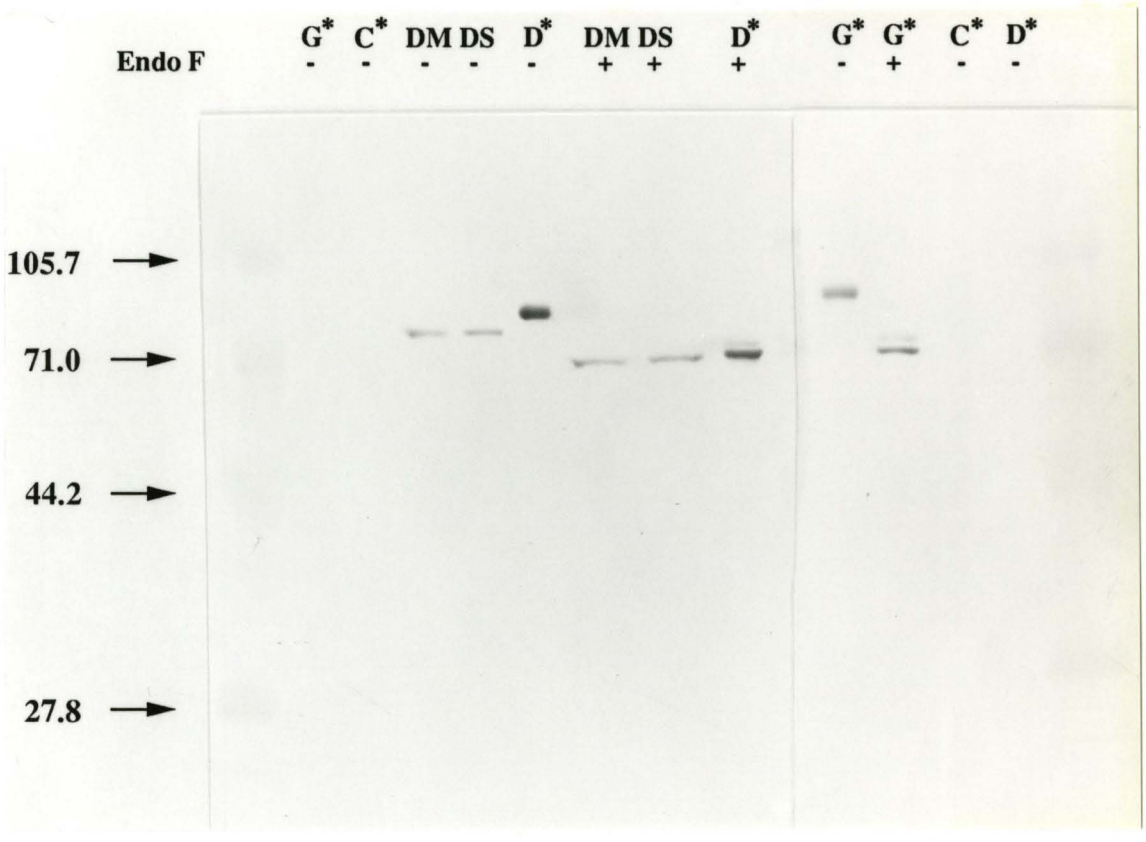

Fig. 11. Deglycosylation of PI-PLC released fasciclin I from CHO cells. PI-PLC released protein from 10^4 cells were deglycosylated with endo F, subjected to 8% SDS-PAGE and analyzed by Western blotting using the MAb 6D8 for *Drosophila* fasciclin I (left) or the rat anti-grasshopper fasciclin I for grasshopper fasciclin I (right). C*, mock transformed cell line; G*, transformed cell line that expresses grasshopper fasciclin I; D*, transformed cell line that expresses *Drosophila* fasciclin I; DM, membrane fractions of *Drosophila* embryo; DS, supernatant of *Drosophila* embryo membrane preparations after digestion of PI-PLC.

Complex oligosaccharides are added to proteins synthesized in mammalian cells, as compared to small truncated oligosaccharide cores added to proteins synthesized in insect cells (Carr et al., 1989; Kuroda et al., 1990). Thus, it seemed likely that the difference in molecular weight between the insect and CHO-derived fasciclin I proteins was due to additional glycosylation in CHO cells. To examine this possibility, PI-PLC released proteins from CHO-derived lines were treated with endoglycosidase F and compared on a Western blot to similarly treated fasciclin I from *Drosophila* embryos. After deglycosylation, *Drosophila* fasciclin I from CHO cells comigrates with fasciclin I derived from embryos. Deglycosylated CHO-derived grasshopper fasciclin I also migrates at a similar position, corresponding to an apparent molecular mass of 68 kDa (Fig. 11).

C. Purification and Characterization of Grasshopper Fasciclin I

The highest expressing CHO clones producing the PI-linked proteins were grown in a Cell Pharm II hollow fiber bioreactor device (Unisyn Fibertec, San Diego) in the presence of 300 mM MSX. The expressed protein was harvested from the medium after a 3 hour phospholipase C incubation (0.5 ml of 1 mg/ml) and isolated by passing over a 3B11 MAb immunoaffinity column (Fig. 12). ~0.7-1.0 mg of pure protein per daily harvest was obtained (Table 1). Purified PI-linked grasshopper fasciclin I migrates with an apparent molecular mass of ~90 kDa (Fig. 13).

Purification Scheme of Grasshopper Fasciclin I

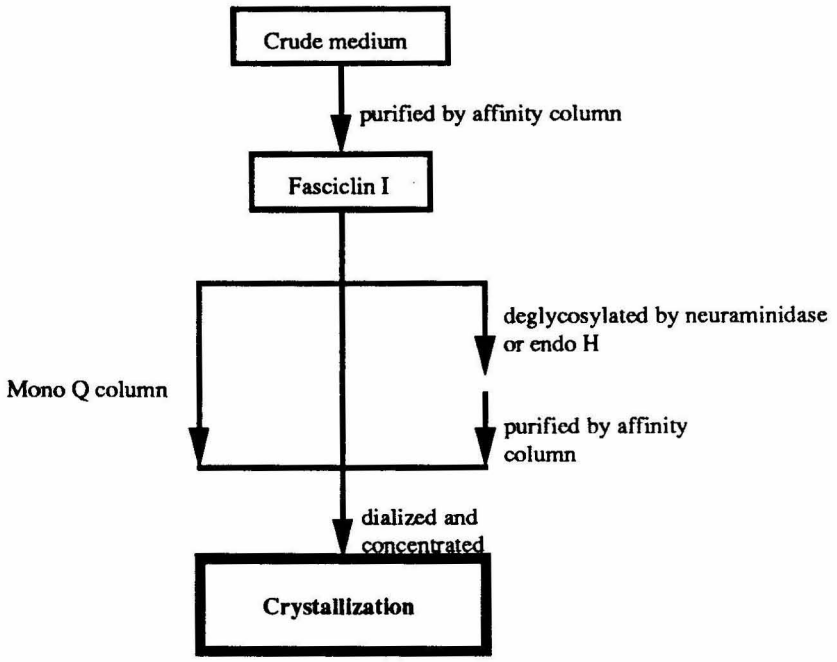

Fig. 12. Purified scheme of grasshopper fasciclin I.

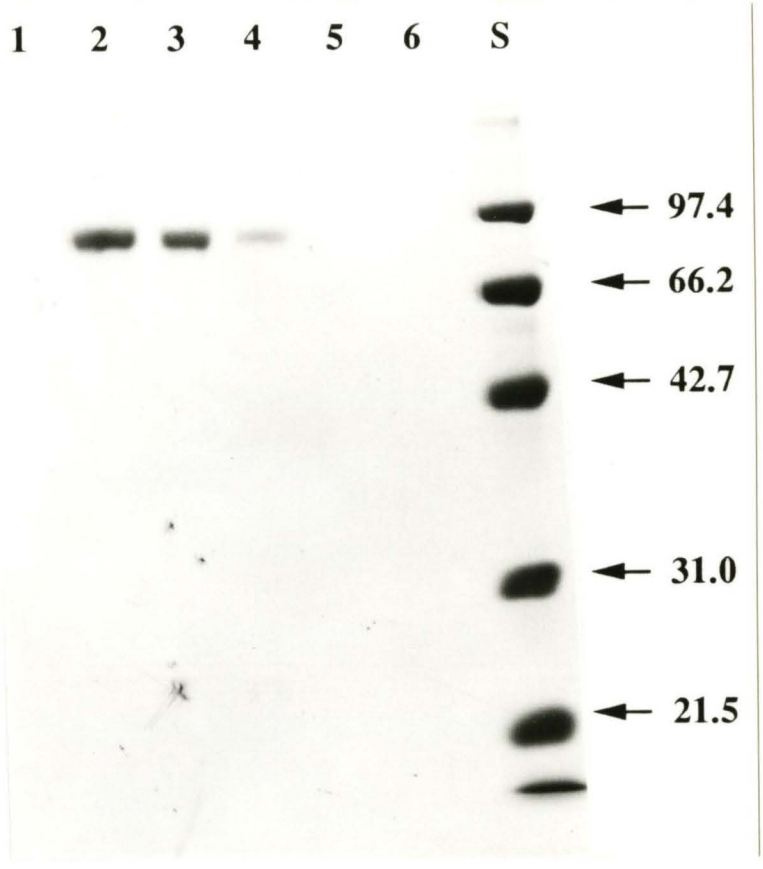

Fig. 13. SDS-PAGE analysis of purified grasshopper fasciclin I (PI-form) from CHO cells. PI-PLC released material from a Cell Pharm I hollow-fiber bioreactor device was isolated by passing over a 3B11 immunoaffinity column and eluted with acid conditions (pH 3.4). Fractions (lane 1 to lane 6) containing purified fasciclin I proteins were subjected to 10% SDS-PAGE and visualized by coomassie brilliant blue R-250 staining. S, protein standards.

Further purification was attempted with anionic exchange chromatography using an FPLC Mono Q column. As shown in Fig. 14A, purified grasshopper fasciclin I after a 3B11 immunoaffinity column was eluted as many overlapping peaks. Fractions of these peaks all migrated with an apparant molecular mass of 90 kDa on a 10% SDS-PAGE gel, indicating that CHO-derived grasshopper fasciclin I is quite charge heterogeneous. In an effort to obtain a homogeneous species for crystallization, we have used step gradients of NaCl salt (30 mM for 5 min, 100 mM for 5 min, 125 mM for 5 min and 1M for 5 min) to elute fasciclin I. A major peak appeared when eluted with 100 mM salt (Fig. 14B). Analysis of protein of this peak on a native isoelectric focusing (IEF) gel showed a pI from 5.5 to 6.5, as compared to pI from 5.5 to 7.5 for the CHO-derived grasshopper fasciclin I (Fig. 15), suggesting that CHO-derived grasshopper fasciclin I is very charge heterogeneous.

In an attempt to remove the extra 15 kDa of carbohydrate in the CHO-derived fasciclin I, we have analyzed different glycosidases, including neuraminidase, endo H and endo F, to deglycosylate the purifed grasshopper fasciclin I. Deglycosylation of the protein with neuraminidase and endo H reduced the apparent molecular mass to ~85 kDa and ~75 kDa, respectively, as shown by SDS-PAGE/10% (Fig. 16A). Deglycosylation with endo F, however, resulted in three species with apparent molecular masses of 66, 68, and 70, as well as apparent degradation of the protein. Analysis of the deglycosylated fasciclin I protein on a native IEF gel revealed an increasing pI from 5.5 to 7.5 with only four species, as compared to pI from 5.5 to 7.5 with 11 species in nondeglycosylated protein (Fig. 15). To optimize the reaction conditions, kinetic studies for neuraminidase and endo H have been done (Fig. 16). Both enzymes remove carbohydrate at 4°C after an overnight incubation.

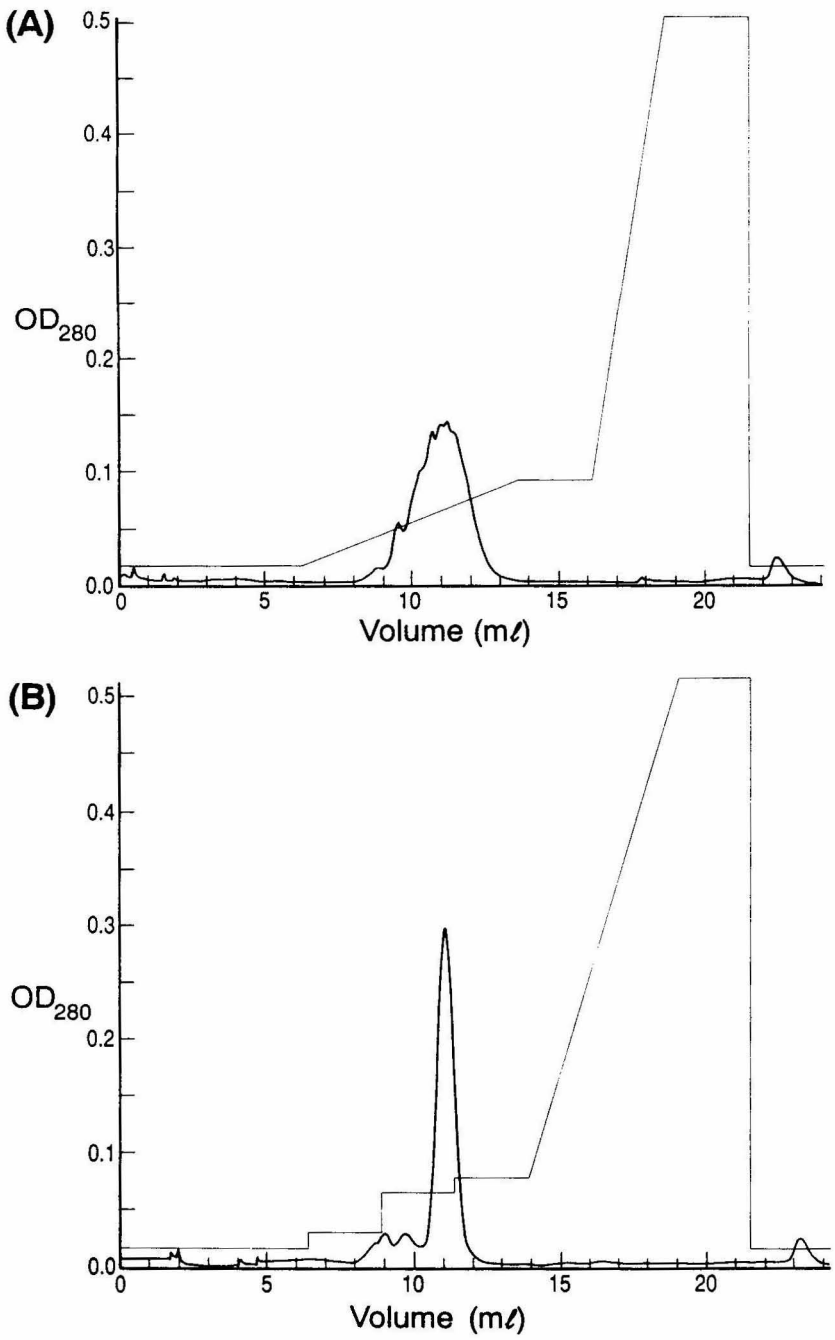

Fig. 14. Anionic exchange chromatography of the CHO-derived grasshopper fasciclin I. Immunopurified grasshopper fasciclin I material was applied on an FPLC Mono Q column pre-equilibrated with 20 mM bisTris propane, pH 7.2 and eluted with a salt gradient 90-150 mM NaCl (A) or step gradients, 50 mM, 100 mM, 125 mM of NaCl (B).

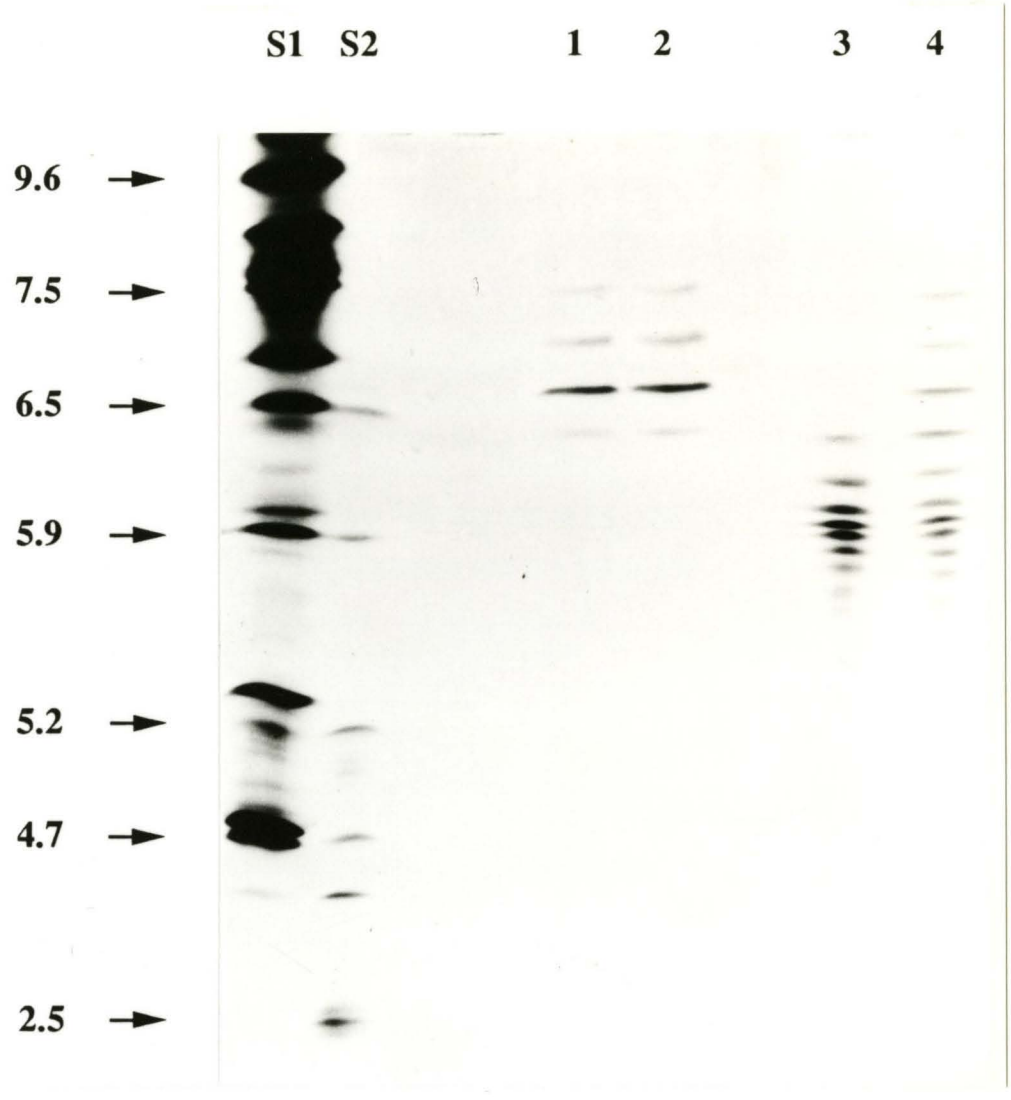

Fig. 15. Isoelectric focusing analysis of the CHO-derived grasshopper fasciclin I. PI-PLC released grasshopper fasciclin I from CHO transformed cells was subjected to a native isoelectric focusing gel. Lane 1, pure PI-form grasshopper fasciclin I treated with neuraminidase; lane 2, pure fasciclin I treated with endo H; lane 3, pool II after the Mono Q column; lane 4, the PI-form grasshopper fasciclin I purified from the immunoaffinity chromatography; S1, IEF standards, pI 4.6-9.6 (BioRad); S2, IEF standards, pI 2.5-6.5 (Pharmacia).

Table 1. *Harvest of PI-form grasshopper fasciclin I in the Cell Pharm II*[1]

##	Harvest Date	Harvest Volume (ml)	Purified Protein (mg)
1	10-15-90	230	0.5
2	10-18-90	300	0.5
3	10-22-90	290	1.0
4	10-23-90	270	0.5
5	10-24-90	240	0.8
6	10-26-90	350	1.0
7	10-27-90	360	0.5
8	10-28-90	310	0.2
9	10-29-90	185	0.2
10	10-30-90	340	0.8
11	10-31-90	430	0.5
12	11-02-90	314	0.3
13	11-04-90	316	0.9
14	11-05-90	272	0.2
15	11-06-90	352	0.7
16	11-07-90	328	0.5
17	11-08-90	420	0.6
18	11-09-90	380	0.5
19	11-11-90	340	0.3
20	11-12-90	330	0.3
21	11-13-90	240	0.4
22	11-14-90	320	0.2
23	11-15-90	360	0.3
24	11-16-90	360	0.5
25	11-17-90	250	0.2
26	11-18-90	440	0.7
27	11-20-90	220	0.4

[1] 10^8 cells of clone expressing the highest PI-form grasshopper fasciclin I were inoculated into Cell Pharm II on 10-05-90.

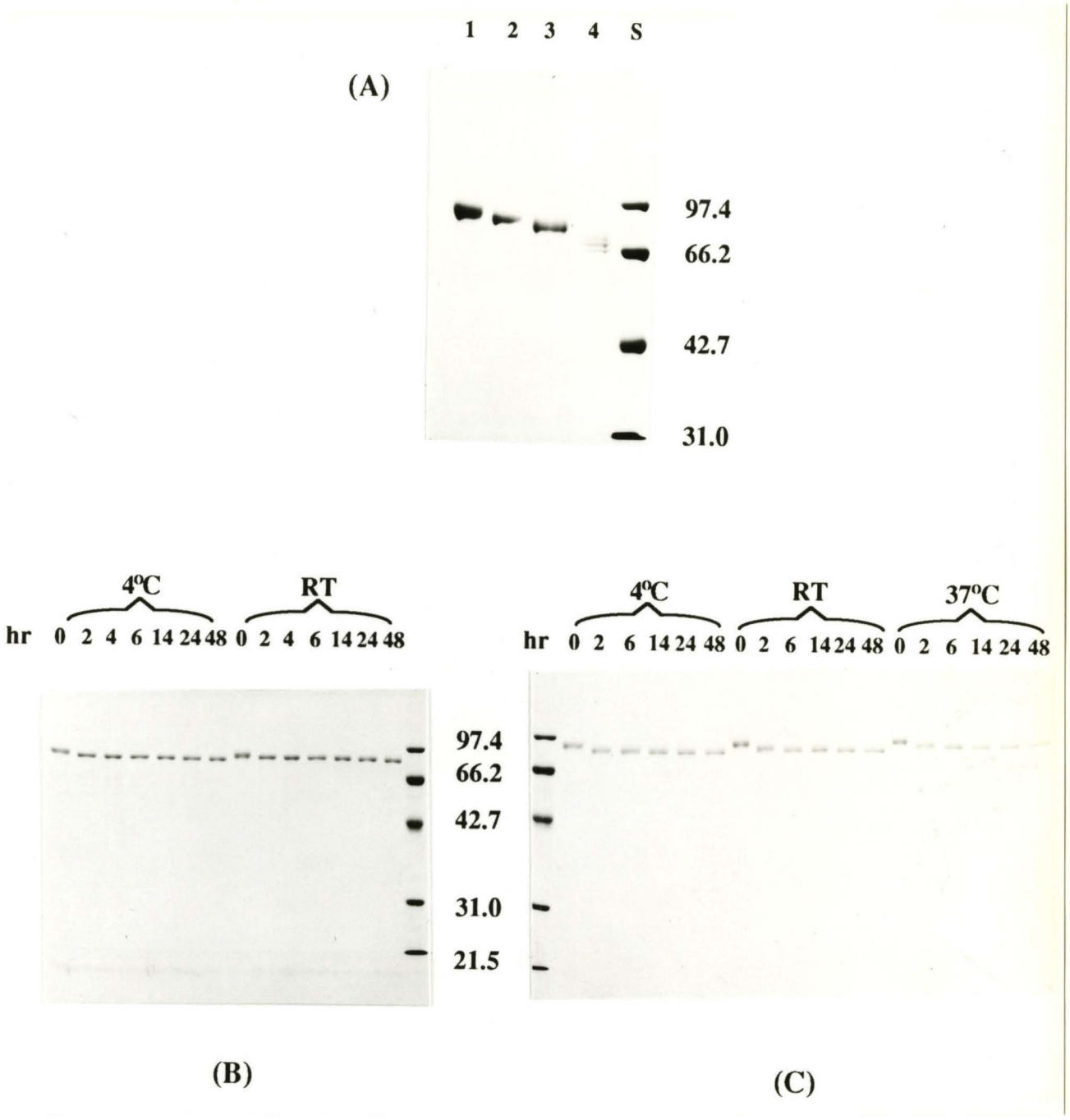

Fig. 16. Deglycosylation of the PI-form of grasshopper fasciclin by glycosidases. Pure proteins (lane 1) were glycosylated with neuraminidase (lane 2), endo H (lane 3) and endo F (lane 4) at a ratio of 1:100 at 37°C overnight (A). To optimize the reaction conditions, proteins were incubated with neuraminidase (B) or endo H (C) at different temperatures and aliquots of sample were withdrawn at different intervals as indicated for 10% SDS-PAGE.

To examine the stability and possible domain structure of purified fasciclin I, samples of purified protein were treated with trypsin, chymotrypsin, papain, and elastase (Fig. 17). Fasciclin I was resistant to digestion by these proteases, suggesting that the protein is folded into a stable structure without obvious flexible linker sequences between its domains, or that potential cleavage sites were protected by an extra 15 kDa of carbohydrate.

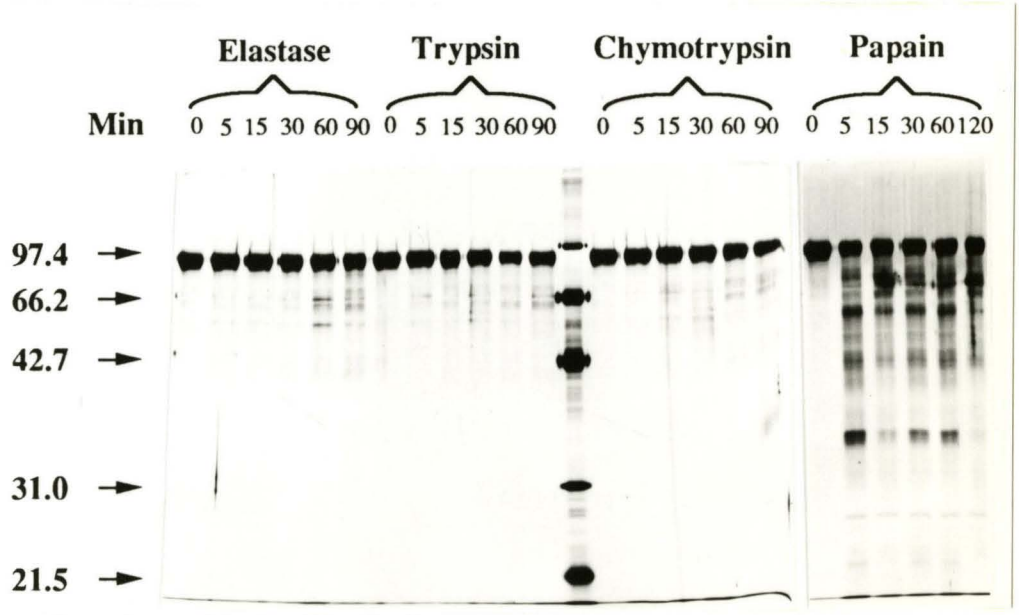

Fig. 17. Proteolytic digestion of the CHO-derived pure grasshopper fasciclin I. Pure proteins were treated with elastase, trypsin, chymotrypsin and papain as indicated at 37°C. Aliquots of treated sample were withdrawn at different time intervals and subjected to 10% SDS-PAGE.

IV. *Discussion*

As a first step towards a structural and functional analysis of grasshopper fasciclin I, we have established mammalian cell lines that express the PI-linked grasshopper fasciclin I using a glutamine-synthetase-based amplifiable expression system. With this expression system, a high level of expression is obtained in the initial selection and amplification in the presence of MSX, and a much higher expression is rapidly achieved after the second-round amplification. Expression of grasshopper fasciclin I was detected at the cell surface of CHO transfected populations by flow cytometry, using a conformation-sensitive monoclonal antibody, 3B11. The highest expression clone producing ~10 μg per 10^7 cells, representing 6 x 10^6 molecules per cell, was selected at 300 mM MSX. Staining was diminished by treatment of positive cell populations with PI-PLC, indicating that the CHO-derived grasshopper fasciclin I was associated with the cell membrane by its PI anchor.

The PI-forms of CHO-derived grasshopper fasciclin I and *Drosophila* fasciclin I were identified to have apparant molecular masses of ~90 kDa and ~85 kDa, respectively, as analyzed by cold immunoprecipitation, Western blot, ^{35}S metabolic labeling and N-terminal amino acid sequencing. As compared to ~75 kDa for grasshopper fasciclin I in grasshopper embryos (Snow et al., 1988) and ~72 kDa for *Drosophila* fasciclin I in *Drosophila* embryos (Hortsch and Goodman, 1990), there is an extra ~15 kDa molecular mass. The difference in molecular weight between the insect and CHO-derived fasciclin I proteins was demonstrated to be due to the additional glycosylation in CHO cells (Carr et al., 1989; Kuroda et al., 1990). This is consistent with six predicted N-linked carbohydrate structures in grasshopper fasciclin I and four in *Drosophila* fasciclin I (Zinn et al., 1988). Up to 1 mg of pure grasshopper fasciclin I per daily harvest in a Cell Pharm has been obtained for further biochemical characterization (Chapter 3) and a structure determination (Chapter 4).

With large amounts of complex oligosaccharides added on proteins, CHO-derived grasshopper fasciclin I is characterized to be very charge heterogeneous as shown in the profile of an anionic exchange chromatography and an isoelectric focusing analysis. The extra ~15 kDa carbohyrates of fasciclin I can be easily removed with different glycosidases, yet there are still four species with different pI after deglycosylation, probably due to *O*-linked glycosylation or other post-translational modifications. We have also shown that the expressed grasshopper fasciclin I is proteolytically resistant, indicating that the protein is not an extended structure with potential linker sequences between its domains, which is in contrast to a common feature suggested for adhesion molecules (Hynes and Lander, 1992). Alternatively, the large amount of carbohydrate attached to this protein may prevent the proteolytic digestion with proteases.

References

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K. (1989) Introduction of DNA into mammalian cells. *In* Current protocols in molecular biology. Volume I, Chapter 9.

Bebbington, C. R., and Hentschel, C. C. G. (1987). In *DNA Cloning*, ed. Glover, D. M. (IRL, Oxford), Vol. 3, pp. 163.

Carr, S. A., Hemling, M. E., Folena-Wasserman, G., Sweet, R. W., Anumula, K., Barr, J. R., Huddleston, M. J., and Taylor, P. (1989). Protein and carbohydrate structural analysis of a recombinant soluble CD4 receptor by mass spectrometry. J. Biol. Chem. *264*, 21286-21295.

Felgner, P. L., Gader, T. R., Holm, M., Roman, R., Chan, H. W., Wenz, M., Northrop, J. P., Ringold, G. M., and Danielsen, M. (1987). Lipofection: A highly efficient, lipid-mediated DNA-transfection procedure. Proc. Natl. Acad. Sci. USA *84*, 7413-7417.

Giulian, G. G., Moss, R. L., and Greaser, M. (1984). Analytical isoelectric focusing using a high-voltage vertical slab polyacrylamide gel system. Anal. Biochem. *142*, 421-436.

Gorman, C. (1985). High efficiency gene transfer into mammalian cells. *In* DNA cloning: A practical approach. Volume II. Ed. D. M. Glover. IRL press, Oxford. Chapter 6, pp. 143-190.

Gorman, C., Moffat, L. F., and Howard, B. H. (1982). Recombinant genomes which express chloramphenical acetyltransferase in mammalian cells. Mol. Cell Biol. *2*, 1044-1051.

Harlow, E., and Lane, D. (1988). In Antibodies: a laboratory manual, Cold Spring Harbor Laboratory. Cold Spring Harbor, NY, Chapter 13.

Hortsch, M., and Goodman, C. S. (1990). *Drosophila* fasciclin I, a neural cell adhesion molecule, has a phosphoinositol lipid membrane anchor. J. Biol. Chem. *265*, 15104-15109.

Hynes, R. O., and Lander, A. D. (1992). Contact and adhesive specificities in the associations, migrations and targeting of cells and axons. Cell *68*, 303-322.

Kunkel, T. A., Roberts, J. D., and Zakour, R. A. (1987). Rapid and efficient site directed mutagenesis without phenotypic selection. Meth. Enzymol. *154*, 367-382.

Kuroda, K., Geyer, H., Geyer, R., Doerfler, W., and Klenk, H.-D. (1990). The oligosaccharides of influenza virus hemagglutinin expressed in insect cells by a Baculovirus vector. Virol. *174*, 418-429.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacterophage T4. Nature *227*, 680-685.

Lin, A. Y., Devaux, B., Green, A., Sagerstrom, C., Elliott, J. F., and Davis, M. M. (1990). Science *249*, 677-679.

Morrissey, J. H. (1981). Silver stain for proteins in polyacrylamide gels: a modified procedure with enhanced uniform sensitivity. Anal. Biochem. *117*, 307-310.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual, 2nd ed. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory.

Snow, P. M., Zinn, K., Harrelson, A. L., McAllister, L., Schilling, J., Bastiani, M. J., Makk, G., and Goodman, C. S. (1988). Characterization and cloning of fasciclin I and fasciclin II glycoproteins in the grasshopper. Proc. Natl. Acad. Sci. USA *85*, 5291-5295.

Studier, F. W., Rosenberg, A. H., Dunn, J. J., and Dubendorff, J. W. (1990). Use of T7 RNA polymerase to direct the expression of cloned genes. Meth. Enzymol. *185*, 60-89.

Takebe, Y., Seike, M., Fujisawa, J.-I., Hoy, P., Yokota, K., Arai, K.-I., Yoshida, M., and Arai, N. (1988). SRα promoter: an efficient and versatile mammalian cDNA expression system composed of the simian virus 40 early promoter and the R-U5 segment of human T-cell leukemia virus type 1 long terminal repeat. Mol. Cell. Biol. *8*, 466-472.

Towbin, H., Staehelin, T., and Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc. Natl. Sci. USA *76*, 4350-4354.

Zinn, K., McAllister, L., and Goodman, C. S. (1988). Sequence analysis and neuronal expression of fasciclin I in grasshopper and *Drosophila*. Cell *53*, 577-587.

Chapter 3

Expression of Secreted Grasshopper Fasciclin I in Chinese Hamster Ovary Cells and in Schneider 2 Cells, and Functional Studies of the Expressed Grasshopper Fasciclin I

Abstract

This chapter describes a directly secreted form of grasshopper fasciclin I produced in two expression systems using Chinese Hamster Ovary (CHO) cells, or *Drosophila* Schneider 2 (S2) cells. The truncated form of grasshopper fasciclin I expressed in CHO cells and S2 cells migrates with apparent molecular masses of 87 kDa and 72 kDa respectively, as compared to PI-form of CHO-derived grasshopper fasciclin I that migrates with an apparent molecular mass of 90 kDa. Biochemical characterization of fasciclin I indicates that it exists as a monomer in solution, an observation consistent with homophilic interaction properties only if the interaction is of low affinity. Structural studies by electron microscopy and circular dichroism suggest that fasciclin I has a compact structure with a significant content of α-helix. Fasciclin I is therefore structurally distinct from adhesion molecules that are members of the immunoglobulin superfamily and/or contain fibronectin III repeats, which are predicted to be elongated structures with flexible hinge regions containing predominantly β-sheet secondary structure.

I. Introduction

In an effort to eliminate the heterogeneity that may result from the portion of the PI tail that remains after release from the membrane with PI-PLC, we generated a directly secreted form by modifying the cDNA sequence of grasshopper fasciclin I to introduce a stop codon before the PI attachment signal. We have shown that CHO cell lines transfected by an expression vector containing the truncated construct produces the truncated protein which is secreted into the medium. An insect expression system was also used to obtain a secreted form of protein containing a physiological amount of carbohydrate. The expression vector pRmHa3.sGf1 was created and introduced into *Drosophila* Schneider 2 (S2) cells. Preliminary studies show that the S2-derived grasshopper fasciclin I migrates at the expected molecular weight, and thus does not contain the extra ~15 kDa of carbohydrate moieties present in the CHO-derived protein.

With the milligram quantities of pure fasciclin I available, a biochemical characterization was initiated to understand the role of fasciclin I in neuronal recognition. *Drosophila* fasciclin I has been suggested to act as a homophilic adhesion molecule in tissue culture cells (Elkins et al., 1990a). It was therefore of interest to determine whether this adhesion activity would correlate with oligomerization in solution using the pure CHO- or S2-derived protein. However, soluble fasciclin I was found to exist as a monomer in solution, as analyzed by gel filtration, crosslinking experiments and electron microscopy that will be discussed in this chapter. An attempt to utilize S2- or CHO-expressing PI-form grasshopper cells to perform cell aggregation assays developed by Elkins et al. (1990a) did not succeed. Finally, circular dichroism analysis of the soluble protein suggests that the primary sequence of fasciclin I contains a significant amount of α-helical content and therefore its structure must differ from the β-sheet structures predicted for adhesion

molecules that are members of the immunoglobulin superfamily and/or contain fibronectin III repeats (de Vos et al., 1992; Williams, 1988; Hynes and Lander, 1992; Jessell, 1988).

II. Materials and Methods

Reagents. DSP and NHS-biotin were from Pierce. The S2 cell line and the phsneo vector were gifts of Michael Jackson (Scripps Clinic). The pRmHa3 vector was the gift of Allen Bieber (Purdue). An S2 cell line expressing the PI-linked grasshopper fasciclin I under metallothionein promoter was a gift of Mark Sieger (UC Berkeley). Affi-gel 10 was from BioRad. Antibodies and other reagents were described in Chapter 2. All other chemicals were reagent grade.

Construction of truncated forms of grasshopper fasciclin I. To obtain a secreted version of grasshopper fasciclin I, a stop codon followed by a BamHI site was inserted after residue 634 by site-directed mutagenesis (Pantelis Tsoulfas, W.-C. W., unpublished) (Kunkel et al., 1987). pBJ5.GS.sGf1 was generated by introducing an EcoRV-NotI fragment containing the modified cDNAs introduced into the filled XhoI and NotI sites of the polylinker of pBJ5.GS (Gastinel et al., 1992) (Fig. 1).

For expression in insect cells, the secreted form was constructed by introducing the modified cDNA encoding truncated grasshopper fasciclin I as an EcoRI-BamHI fragment into the EcoRI and BamHI sites of the polylinker of the pRmHa3 containing the metallothionein promoter to generate pRmHa3.sGf1. Molecular biological experiments were performed by standard methods (Sambrook et al., 1989).

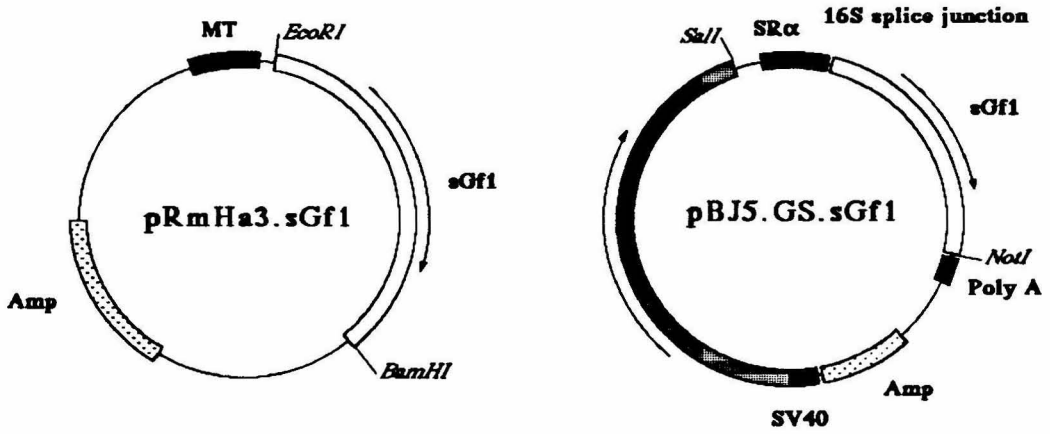

Fig. 1. Schematic diagram showing the structures of the secreted version of two expression vectors, pBJ5.GS.sGf1 and pRmHa3.sGf1.

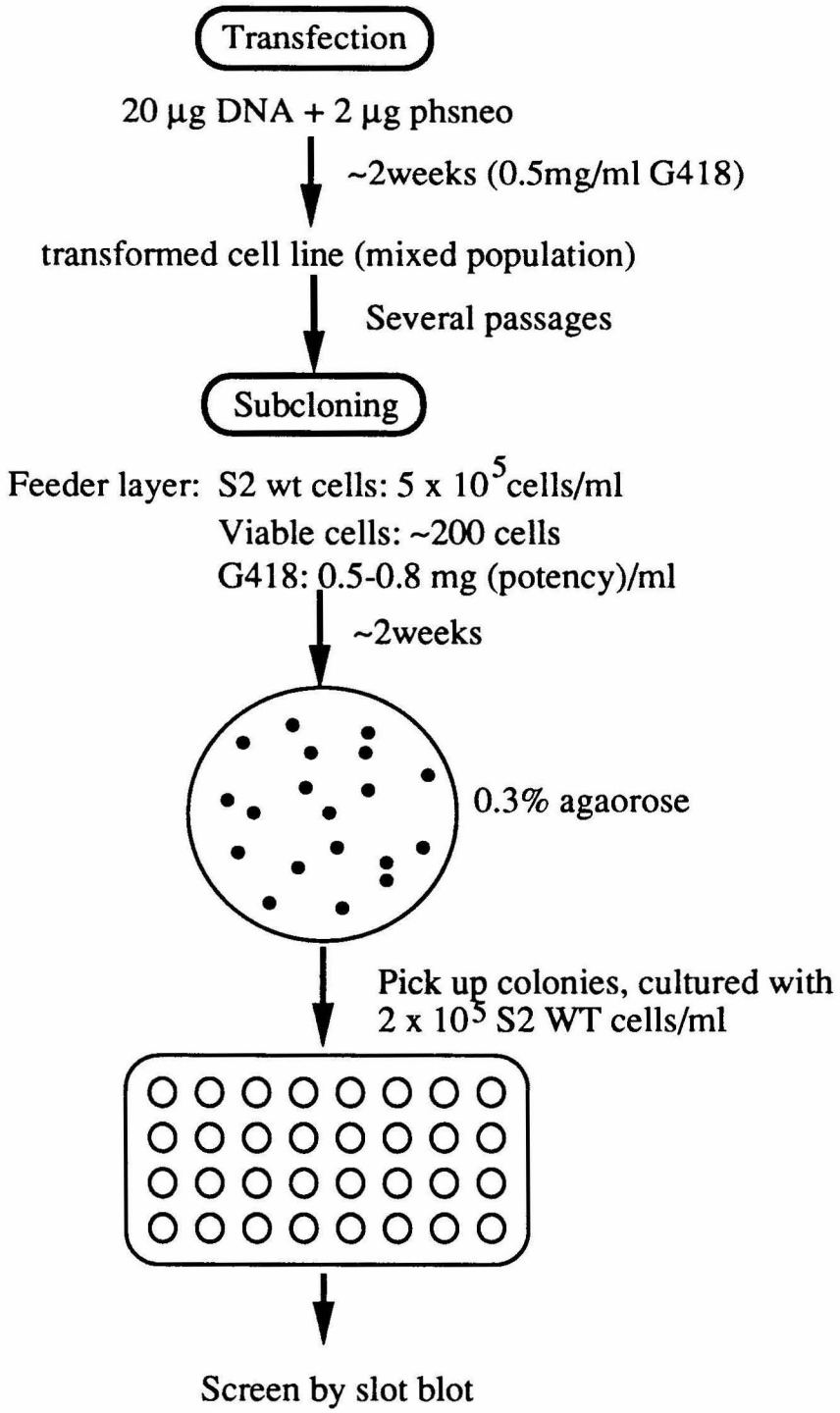

Fig. 2. Schematic diagram showing the strategy to clone Schneider 2 cell lines expressing fasciclin I.

Cell culture and transfection. Chinese Hamster Ovary (CHO) K1 cells were grown in αMEM supplemented with 10% FBS, 2 mM glutamine and 100U/ml penicillin/streptomycin. S2 cells were grown in Schneider's medium supplemented with 12.5% FBS, 2 mM glutamine and 100U/ml penicillin/streptomycin.

To generate CHO cells producing the secreted form of grasshopper fasciclin I, the pBJ5.GS.sGf1 expression vector was introduced into CHO cells, selected and amplified essentially the same as described in Chapter 2. Cell supernatants of transformant CHO cell lines were filtered through a nitrocellulose paper using a Minifold II slot blot system (Schleicher & Schuell #SRC072/0), and immunostained wtih the MAb 3B11 to examine the presence of grasshopper fasciclin I. Positive transformant lines were verified by Western blotting with a rat anti-grasshopper fasciclin I antiserum.

For stable transfection using an insect expression system (Fig. 2), 25 μg of pRmHa3.sGf1 plus 2.5 μg of phsneo, a selection vector, were cotransfected into S2 *Drosophila* cells (10^6 cells/ml) cultured in 10-cm plate (10 ml of medium) according to a calcium phosphate procedure (Stratagene). Two days after transfection, cells were split into two 175-cm^2 flasks and selected with 0.5-1 mg (potency)/ml of G418 (Gibco). After two weeks, transformed cells were passaged for a few generations and were ready to clone in a soft agar (A. Bieber, unpublished). In brief, 8 ml of wild type S2 cells (10^5 cells/ml) and transformed cells (100 to 800 cells) in Schneider's medium, 12.5% FCS, penicillin/streptomycin, with 0.8 to 1 mg/ml (potency) of G418 were pipetted onto one side of a 10-cm bacteriological grade petri dish, leaving most of the surface of the plate uncovered. 2 ml of a 1.5% molten agar solution (48°C; Difco #0142-02) were pipetted onto the opposite side of the petri plate. The contents of the plate were mixed thoroughly by swirling, allowed to solidify, and the plates were sealed by parafilm. Within 2 weeks after being cultured at 26°C, clones became visible and were isolated under an inverted

microscope and grown in individual wells of a 96-well plate. Expression of protein was induced in plates containing cells equilibrated to 2 x 10^6 cells/ml for 2 to 3 days in serum free medium with 0.7 mM $CuSO_4$ and examined using Western blotting by a mouse anti-fasciclin I antiserum or using the Minifold II slot blot system by the MAb 3B11.

Gel electrophoresis, and Western blot analysis. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was performed according to Laemmli (1972) and silver staining was performed according to Morrissey (1981).

After SDS-PAGE, proteins were blotted to nitrocellulose paper according to Towbin et al. (1979). Protein transfer was done at 100 V for 3 hours in transfer buffer (25 mM Tris, 190 mM glycine, 10% methanol). The blots with transferred proteins were immersed in 5% non-fat dry milk in TBST solution (10 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.05% Tween 20) for 30 minutes and then incubated with mouse anti-*Drosophila* fasciclin I MAb (6D8), or rat (or mouse) anti-grasshopper fasciclin I antiserum at dilution of 1:1000 with TBST. After 1 hour, the blots were washed three times with TBST, two times with TBS (10 mM Tris-HCl, pH 8.0, 150 mM NaCl), each for 10-15 minutes and incubated with affinity-purified horseradish-peroxidase-conjugated goat anti-mouse or rat-anti-mouse IgG (1:200, diluted with TBS, Cappel) for 30 minutes. After three washes with TBS, twice with PBS and once with 50 mM Tris, pH 7.6, the blots were visualized using diaminobenzidine (DAB)-hydrogen peroxide staining solution (0.06% DAB (w/v), 0.03% H_2O_2 (v/v) in 50 mM Tris, pH 7.6) until the bands were suitably dark (about 5 minutes). Alternatively, *Elite* ABC kits for immunoblots from Vector Lab were used. All procedures were carried out at room temperature.

Size-exclusion chromatography. Following immunoaffinity chromatography, fasciclin I was concentrated approximately 0.2-1 mg/ml using a vacuum dialysis apparatus

(Schleicher & Schuell) and a Centricon centrifugal concentrator (Amicon, 25,000 MW cutoff). 50 µl of the concentrated protein were run on an FPLC Superose 12 column equilibrated with PBS.

Crosslinking experiments. Purified grasshopper fasciclin I protein (~0.01-0.02 mg/ml) was treated for 30 min at 4°C with 3,3'-dithio-bis-propionic acid *N*-hydroxysuccinimide ester (DSP, 100 mM as a stock in dimethylformamide), a thio-cleavable homobifunctional crosslinking reagent, at concentrations ranging from 100 to 800 µg/ml. Crosslinking reactions were stopped by adding glycine to 50 mM, and samples were analyzed by Western blotting under reducing and nonreducing conditions.

Electron microscopy. To prepare sample for electron microscopy, grasshopper fasciclin I solutions at 0.1 to 0.3 mg/ml in 50 mM ammonium formate, pH 7.0, were mixed with 0.1 mg of ferritin in a 4:1 ratio. A 10-µl portion of the mixture was sprayed on a 400-mesh copper grid using a bulb nebulizer. The specimens were rotarily shadowed with platinum at an angle of 1:5 and examined under an electron microscope (Model 420; Philips Electronic Instruments, Inc., Mahwah, NJ).

Cell aggregation assays and embryo staining. S2 transformant lines expressing the PI-form or the secreted form of grasshopper fasciclin I were titered using a hemocytometer and cell densities were equilibrated for each experiment to 5 x 10^6 /ml in 0.5 ml HBSS and were induced by 0.7 mM $CuSO_4$ overnight. The cells were then agitated at 100 rpm on a shaker platform for 1 hour and examined under microscope by putting on a microscope slide with an 18-mm square coverslip. Alternatively, transformed S2 cells expressing PI-linked fasciclin I were mixed with a monolayer of confluent CHO cells expressing a high level of PI-linked fasciclin I, followed by agitation at 100 rpm on a

shaker for an hour. Cells were washed gently three times with PBS and the formation of rosettes was examined under microscope.

For embryo staining, the CHO-derived fasciclin I was biotinylated essentially according to the manufacturer's directions (Pierce). In brief, 0.5 ml of purified fasciclin I (0.76 mg/ml) was reacted with 5 μl of 6 mg/ml of NHS-biotin (in DMSO) for 1 hour at room temperature. Free NHS-biotin was removed by centrifuging with a Centricon-30 microconcentrator (Amicon) three times. The coupling efficiency was 7 biotins per molecule of grasshopper fasciclin I under these conditions. Biotinylated bovine serum albumin (3-4 biotins per molecule) was used as a negative control. Grasshopper embryo staining was performed essentially the same as described by Bastiani et al. (1987) in collaboration with Barry Condrun For analysis of embryonic perturbation, 0.2 mg/ml of purified CHO-derived fasciclin I was incubated with the primary culture of grasshopper fasciclin I at an approximately 1:10 dilution (Bastiani et al., 1987). Embryos were observed under microscope to analyze whether there was altered pathways.

Circular dichroism. The CD spectrum of the CHO-derived truncated grasshopper fasciclin I (0.21 mg/ml in 5 mM sodium phosphate, pH 7.5) was collected in collaboration with Ilana Tamir using a Jasco J-600 spetropolarimeter. A cuvette with an 0.1 cm pathlength was used, and the mean residue ellipticity (deg·cm^2/dmol) was determined assuming a mean residue weight of 141. All spectra were recorded from 260 to 185 nm and were determined as the average of five scans. The experiments were carried out at room temperature.

III. Results

A. Expression of Secreted Forms of Fasciclin I.

In order to remove the PI tail remaining in the PI-PLC released fasciclin I, a version of the cDNA encoding a secreted form of grasshopper fasciclin I was created using site directed mutagenesis (Kunkel et al., 1987; Pantelis Tsoulfas and W.-C. W., unpublished) by introducing a stop codon after amino acid 634 of grasshopper fasciclin I, which corresponds to the position presumably encoding the beginning of the PI linkage signal in the fasciclin I cDNA (Ferguson and Williams, 1988). The CHO expression vector, pBJ5.GS.sGf1, was reconstructed, introduced into CHO cells, selected and amplified with the presence of the drug MSX (see Chapter 2) to obtain stable CHO transformant cells. The highest-expressing lines were identified by the Minifold II slot blot analysis using the MAb 3B11 or by Western blotting of cell supernatants using the rat anti-fasciclin I antiserum. To obtain large amounts of protein, the highest-expressing CHO clones were grown in a Cell Pharm II hollow-fiber bioreactor device (Unisyn Fibertec, San Diego) in the presence of 100 μM MSX (Table 1). The truncated grasshopper fasciclin I was isolated from the medium collected from the cell site of the bioreactor (~500 ml) by passing over a 3B11 MAb immunoaffinity column. About 0.5-1.0 mg of pure protein per daily harvest has been obtained. The purified secreted form of grasshopper fasciclin I migrates with an apparent molecular mass of ~88 kDa as compared to ~90 kDa for the PI-form (Fig. 3A).

Since CHO-derived fasciclin I contains complex oligosaccharides, resulting in a large excess of carbohydrate (~15 kDa, see Chapter 2), an insect expression system that might give rise to less carbohydrate corresponding to the physiological amount of carbohydrate was developed. The modified cDNA encoding the secreted version of grasshopper fasciclin I was introduced into an expression vector, pRmHa3, containing a metallothionein

promoter that functions at high levels in copper-induced *Drosophila* Schneider cell line (Bunch et al., 1988), to generate pRmHa3.sGf1 (Fig. 1). This plasmid was stably introduced into S2 cells by cotransfection with a selection vector, phsneo (Rio and Rubin, 1985). The highest-expressing clones were identified by the Minifold II slot blot analysis or by Western blotting of cell supernatants, and secreted fasciclin I was purified as described above. About 0.5 mg of fasciclin I per liter can be obtained from supernatants of these cells (5 x 10^6/ml) after $CuSO_4$ induction, and this protein migrates with the expected molecular mass of 75 kDa on a 10% SDS-PAGE gel (Fig. 3B), indicating the small truncated oligosaccharide cores are added to the proteins synthesized in the *Drosophila* expression system.

Table 1. Harvest of the secreted grasshopper fasciclin I in the Cell Pharm II[1]

##	Harvest Date	Harvest Volume (ml)	Purified Protein (mg)
1	5-25-91	220	1.4
2	5-27-91	180	0.8
3	5-31-91	420	1.5
4	6-03-91	275	0.8
5	6-06-91	440	1.3
6	6-09-91	360	0.3
7	6-10-91	770	0.5
8	6-17-91	480	0.3
9	6-24-91[2]	400	0.4

[1] 10^8 cells of clone expressing the highest PI-form grasshopper fasciclin I were inoculated into Cell Pharm II on 5-19-91.
[2] Cells were contaminated and were eliminated.

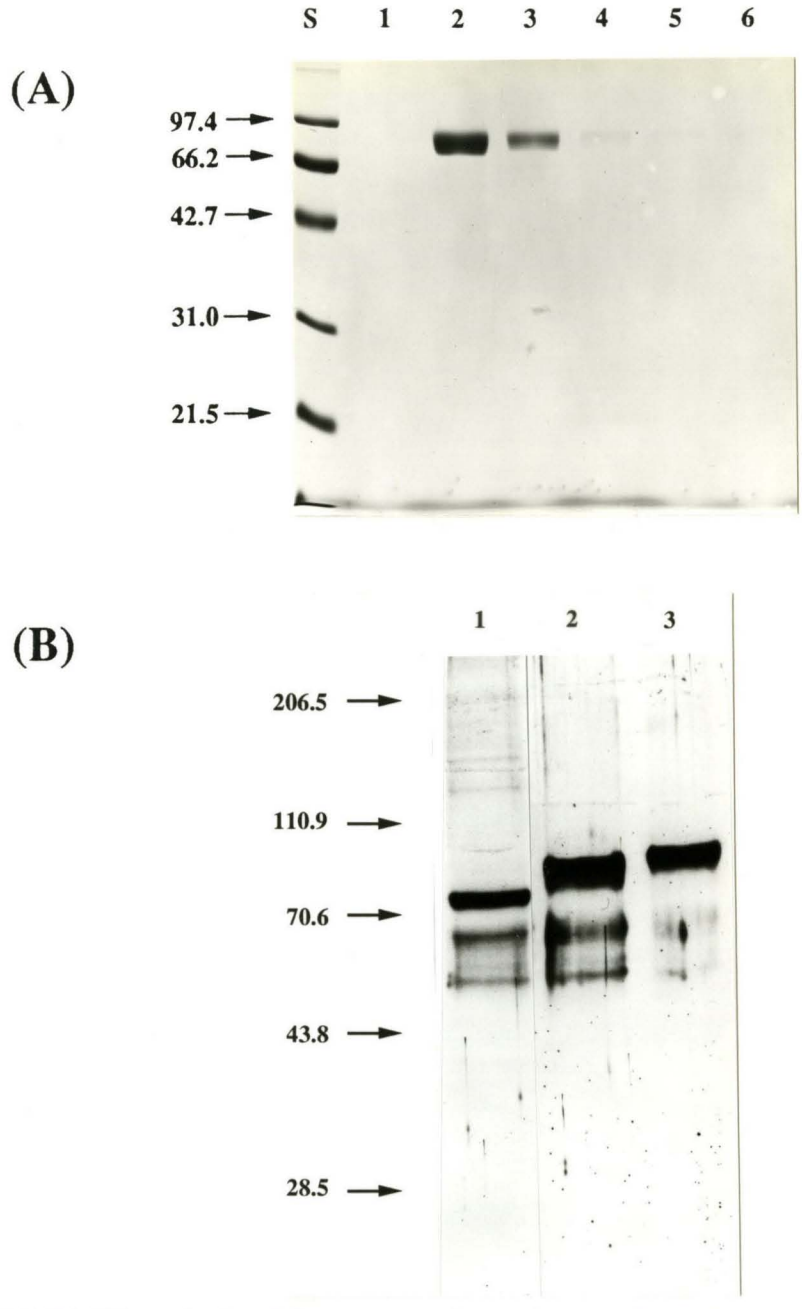

Fig. 3. SDS-PAGE analysis of the expressed grasshopper fasciclin I.
(A). Cell supernatants collected from the cell site of a Cell Pharm II hollow-fiber bioreactor device were isolated by passing over a 3B11 immunoaffinity column and eluted with acid conditions (pH 3.4). Fractions (lane 1 to lane 7) containing the secreted form of fasciclin I were subjected to 10% SDS-PAGE.
(B). Western blot of three forms of grasshopper fasciclin I, using a mouse anti-fasciclin I antiserum. Lane 1, the secreted form of fasciclin I from S2 cells; lane 2, CHO-derived truncated fasciclin I; lane 3, CHO-derived PI-linked fasciclin I.

B. Soluble Grasshopper Fasciclin I Acts as a Monomer in Solution

Previous observation has shown that *Drosophila* fasciclin I functions as a homophilic adhesion molecule in tissue culture cells (Elkins et al., 1990a). To determine whether the purified CHO- or S2-derived truncated grasshopper fasciclin I acts as an oligomer in solution correlated with this adhesion activity, we have characterized the pure material by gel filtration and crosslinking studies. As shown by the gel filtration profile (Fig. 4), concentrated fasciclin I protein from CHO or S2 transformant lines migrates in a position corresponding to a monomeric structure; ~88 kDa for CHO-derived protein and ~75 kDa for S2-derived protein, respectively, rather than a dimeric or higher oligomeric structure. In addition, crosslinking of the purified protein from CHO or S2 cells with the homobifunctional reagent DSP did not give rise to the appearance of a dimer or other oligomeric species on Western blots (Fig. 5). These data indicate that purified fasciclin I exists as a monomer in solution. Similar results have been observed for other homophilic adhesion molecules, N-CAM (Becker et al., 1989) and cadherins (Takeichi, 1990), suggesting that these homophilic interaction molecules have only weak self-affinity.

C. Visualization of Grasshopper Fasciclin I

In collaboration with J. P. Revel, rotary-shadowed fasciclin I preparations were examined by electron microscopy. Ferritin, a 24-subunit iron storage protein (Theil, 1989), was added in the specimens as a positive control and was shown to be a dominant spherical structure about ~200 Å (Fig. 6). Fasciclin I, in contrast, appeared to be a rectangle-like structure with dimensions of approximately 80 Å x 60 Å x 40 Å (after subtracting 10 Å at each end for the shell of metal). Since the image of contrast for fasciclin I is weak as well as measurements of lengths are small, the accuracy of dimensions is about ±10 Å. Nevertheless, the major fraction of these smaller rectangle-like

components is a monomer, which is consistent with our previous results that fasciclin I exists as a monomer in solution.

Adhesion molecules such as N-CAM, ICAM-1 and integrins have been suggested to have extended structures with potential flexible hinge regions, presumably formed by their tandem repeated domains (Hynes and Lander, 1992). Rotary shadowing electron micrographs of both N-CAM and ICAM-1 showed a bent rod with an apparent hinge (between domains 5 and 6 in N-CAM; between domains 2 and 3 in ICAM-1) (Becker et al., 1989; Staunton et al., 1990), indicating that they are flexible molecules. Although the primary sequence of fasciclin I suggests it contain four tandem repeated domains, our electron microscopic studies suggest that they are not arranged to form a long flexible structure as has been hypothesized for other adhesion molecules. Another indication that the fasciclin I structure differs significantly from adhesion molecules that are members of the immunoglobulin superfamily is the finding of a significant component of α-helical secondary structure from an analysis of the CD spectrum of purified fasciclin I (discussed later).

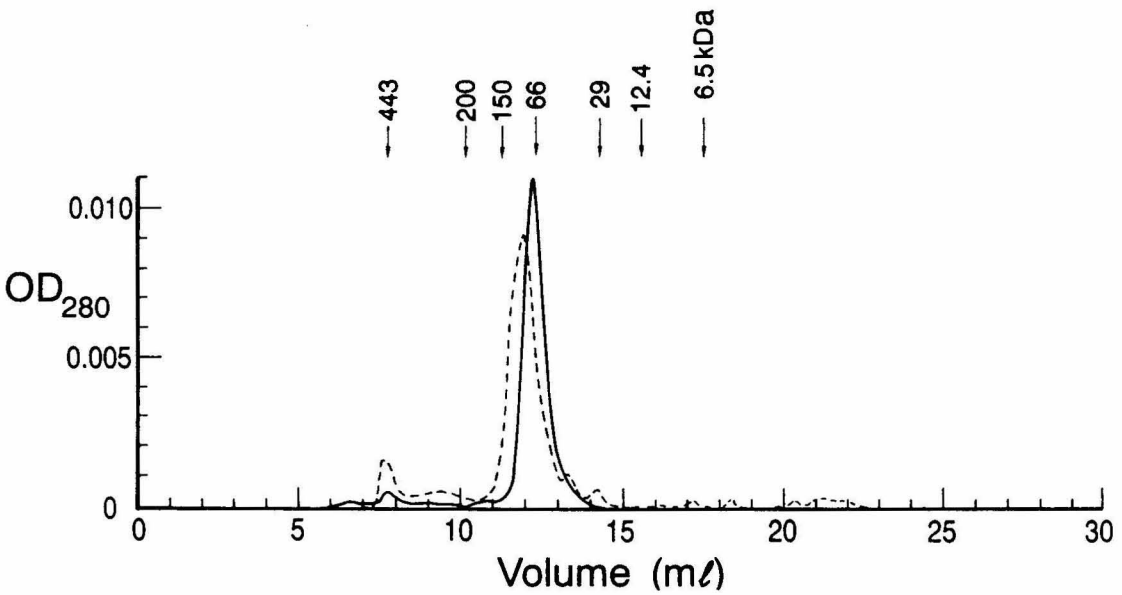

Fig. 4. Gel filtration of purified grasshopper fasciclin I expressed from CHO cells or S2 cells. Purified material of the secreted grasshopper fasciclin I expressed in CHO cells (- - -) or S2 cells (——) was run on the FPLC Superose 12 column in phosphate buffer saline solution at a flow rate of 0.4 ml/min.

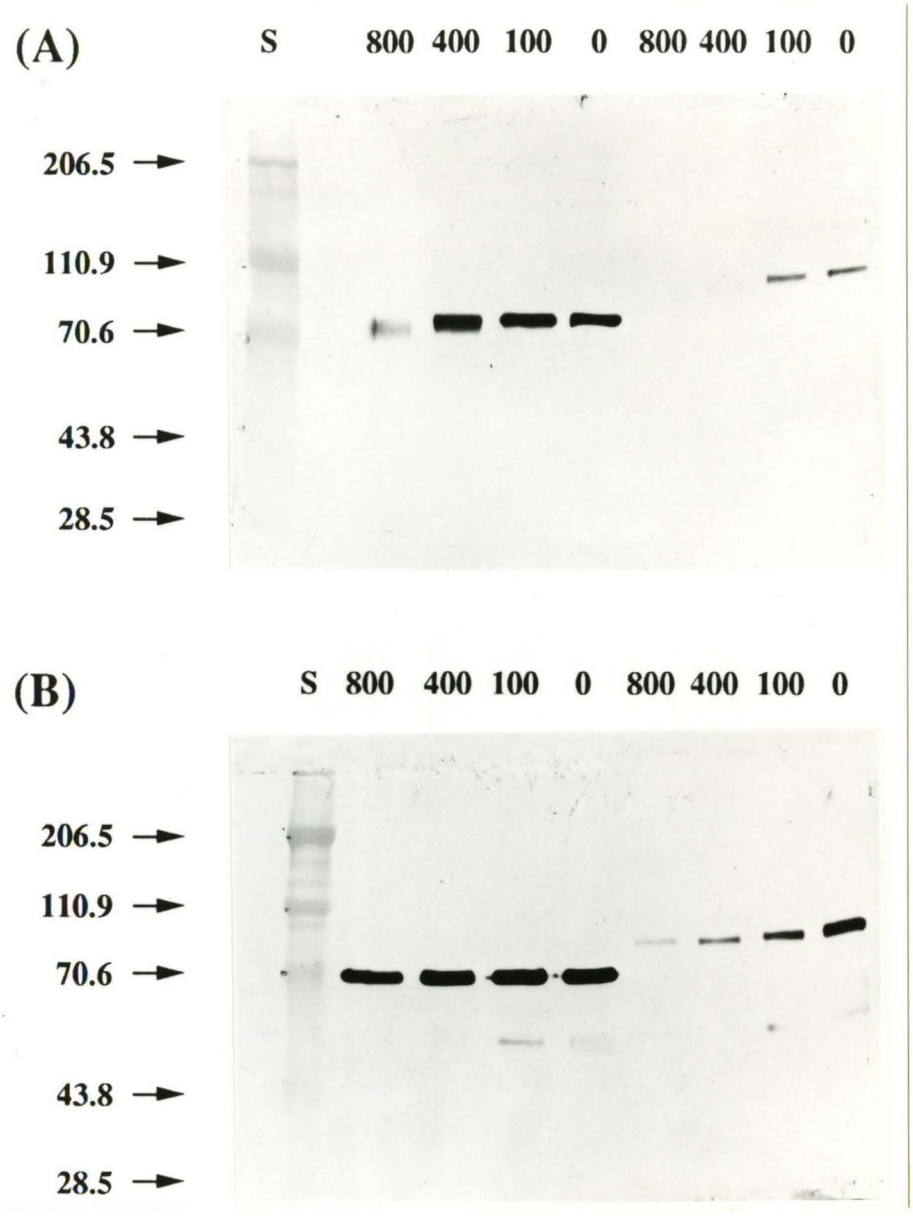

Fig. 5. Crosslinking of purified grasshopper fasciclin I expressed from CHO cells or S2 cells. Purified material was crosslinked using a reducible homobifunctional crosslinking reagent, DSP, at different concentrations (100 to 800 μg/ml) as indicated and was subjected to 8% SDS-PAGE in reducing or nonreducing condition. Lane 1 to lane 4, S2-derived truncated grasshopper fasciclin I; lane 5 to lane 8, CHO-derived truncated grasshopper fasciclin I.

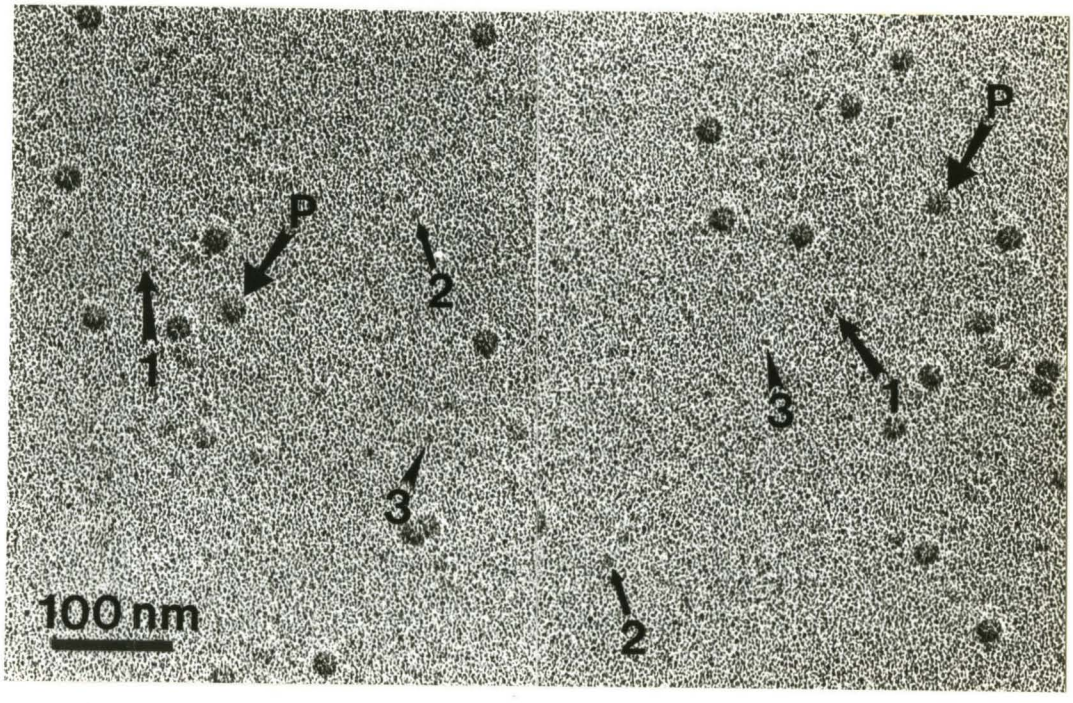

Fig. 6. Electron micrographs of rotary-shadowed soluble grasshopper fasciclin I. P indicates ferritin; 1, 2, and 3 indicate three different forms of fasciclin I.

D. Does Grasshopper Fasciclin I Mediate a Homophilic Interaction?

To address this question, cell aggregation assays developed by Elkins et al. (1990a) were attempted using S2- or CHO-expressing PI-linked grasshopper fasciclin I as described. S2 transformant lines expressing the PI-form (derived from Mark Seeger) or the secreted form of grasshopper fasciclin I were titered and induced by 0.7 mM $CuSO_4$ overnight. The cells were then agitated at 100 rpm on a shaker platform for 1 hour and examined if there were large aggregates. Alternatively, transformed S2 cells expressing PI-linked fasciclin I were mixed with a monolayer of confluent CHO cells expressing a high level of PI-linked fasciclin I, followed by agitation at 100 rpm on a shaker for an hour, washed and examined if rosettes were formed. It was expected that S2 cells expressing PI-form fasciclin I would form large aggregates, and the culture of S2-expressing PI-form fasciclin I with a monolayer of CHO cells expressing a high level of fasciclin I (6×10^6 molecules per cell) would give rise to many rosettes, which then could be inhibited by adding purified fasciclin I or by culturing with S2 cells that produce the secreted fasciclin I. However, no obvious aggregation was observed in S2-expressing PI form of grasshopper fasciclin I (data not shown) and no rosettes were formed in mixing with S2- and CHO-expressing PI-form fasciclin I. Attempts using biotinylated fasciclin I protein to stain grasshopper embryos did not succeed either (B. Condron, and W.-C. W., unpublished). These data again suggest that homophilic interaction properties exist only in the case of a low affinity interaction. Therefore, a modification of oligosaccharides on CHO-derived fasciclin I or a different induction condition [heat shock induction in Elkins' experiments (1990a) whereas $CuSO_4$ induction in our experiments] might result in a loss of homophilic activity. Alternatively, other molecules necessary for cell adhesion might only associate with *Drosophila* fasciclin I but not with grasshopper fasciclin I in S2 cells. Finally, no obvious altered pathways were observed in grasshopper embryonic cultures that were

incubated with purified CHO-derived grasshopper fasciclin I (B. Condron, and W.-C. W., unpublished), indicating that fasciclin I plays a redundant role in growth cone guidance.

E. CD Spectra

To obtain information regarding the secondary structure of fasciclin I, a far UV CD spectrum of fasciclin I was analyzed in collaboration with Ilana Tamir. As shown in Fig. 7, the CD spectrum of fasciclin I is characterized by a positive peak at 195 nm followed by deep negative doublets at 208 nm and 220 nm, indicating a strong α-helical character. The ellipticity of a protein at any given wavelength in the far UV region results from the sum of the contributions from the secondary structural elements with the molecule, thus allowing the estimation of the amounts of α-helix, β-sheet, and β-turn from the CD spectrum in this region (Greenfield and Fasman, 1969). Based on an equation developed by Greenfield and Fasman (1969), the α-helical content of a molecule can be calculated as

$$\% \ \alpha\text{-helix} = \frac{[\theta]_{208nm} - 4{,}000}{33{,}000 - 4{,}000}$$

where $[\theta]_{208nm}$ is mean residue ellipticity at 208 nm. $[\theta]_{208nm}$ can be derived using the equation given by

$$[\theta]_{208nm} = \frac{[\theta]_{\text{measured at 208 nm}} \times 100}{\text{C/mean residue weight} \times 1}$$

where C is the concentration of protein (0.21 mg/ml; determined by BCA assay), mean residue weight is [M.W./# amino acid residues (85,000/610 = 141)], and l is the pathlength (0.1 cm). Since [θ] is measured as 24.5 mdeg·cm^2/decimole at 208 nm, the mean residue ellipticity at 208 nm is derived as 16,450 (Fig. 7). The amount of α-helix is thus calculated as 43% using the equation of Greenfield and Fasman. In addition, using Chang's algorithm (1978) for determining the secondary structure of a protein based on its CD spectrum, it is estimated that grasshopper fasciclin I contains 37% α-helix, 0% β-sheet,

19% β-turn, and 34% irregular structure. CD is a very reliable method for detecting α-helix (Greenfield and Fasman, 1969), and the presence of the typical α-helical spectrum in fasciclin I as well as the high percentage of calculated α-helix suggests that α-helical structure dominates this molecule.

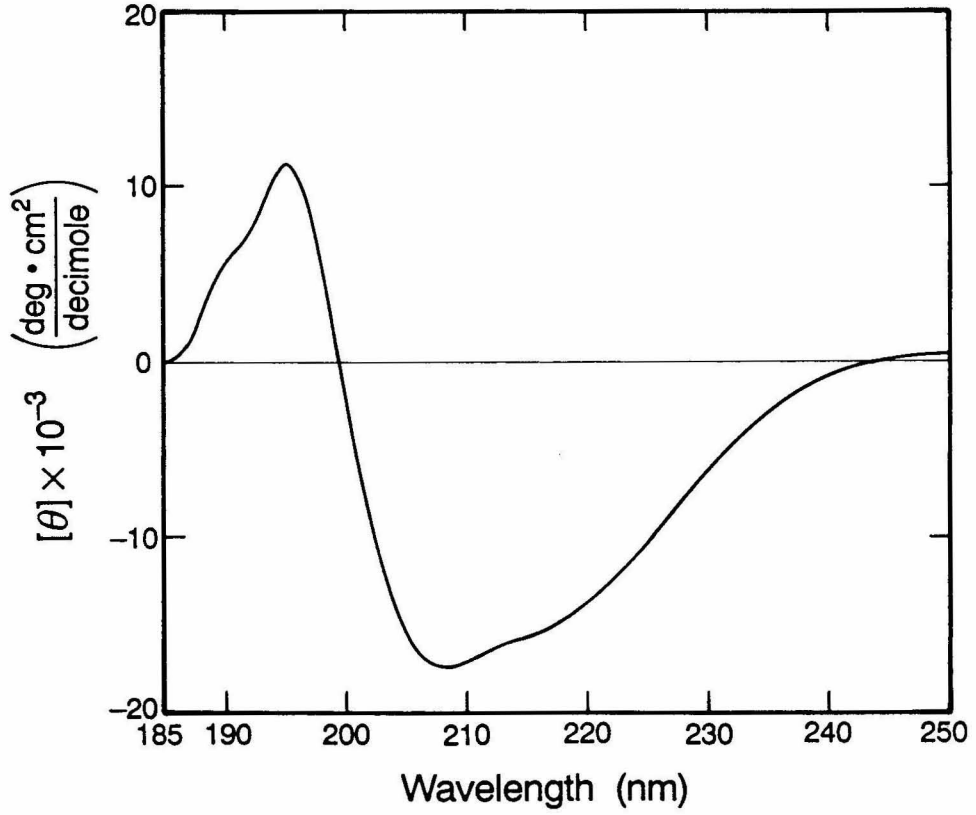

Fig. 7. Far UV CD spectra of the secreted form of grasshopper fasciclin I (0.21 mg/ml, in 5 mM sodium phosphate, pH 7.0). Data is expressed as the mean residue ellipticity [θ], calculated using a mean residue weight of 141.

IV. Discussion

CHO transformant cells expressing PI form grasshopper fasciclin I synthesize the molecule with complex oligosaccharides that result in an extra ~15 kDa carbohydrate being added to residual protein. In addition, the PI-PLC released protein has a PI-tail that may be a potential barrier to crystallization. Therefore, we have generated cell lines to produce a directly secreted form of fasciclin I. This was done by introducing a stop codon before the position presumably encoding the PI signal sequence (Ferguson and Williams, 1988) in the fasciclin I cDNA using site directed mutagenesis (Kunkel et al., 1987). Two secreted forms of expression vectors were constructed; pBJ5.GS.sGf1 for generating CHO-derived truncated fasciclin I, and pRmHa3.sGf1 for a direct production of fasciclin I from transformant S2 cells. We have shown that the truncated insect proteins can be efficiently expressed both in mammalian cell lines as well as in insect cell lines, with a apparent molecular weights of ~88 kDa and ~75 kDa, respectively. These results are consistent with the observations that insect cells synthesize proteins with smaller oligosaccharide cores, whereas mammalian cells produce proteins with complex oligosaccharides (Kuroda et al., 1990; Carr et al., 1989).

Cell adhesion assays (Elkins et al., 1990a) as well as genetic analysis (Elkins et al., 1990b) have suggested that fasciclin I may act as a homophilic adhesion molecule and function in one of the signal transduction pathways for growth cone guidance. To define the activities of fasciclin I that may be involved in axon pathfinding, several biochemical experiments have been done using the CHO- or S2-derived pure grasshopper fasciclin I. An initial characterization of this homophilic adhesion interaction in solution by gel filtration and crosslinking studies indicates that grasshopper fasciclin I exists as a monomer in solution. In addition, no homophilic interaction was observed in CHO cells or S2 cells expressing PI form grasshopper fasciclin I by cell aggregation assays. Attempts to stain

grasshopper embryos with the labeled fasciclin I did not show any staining (B. Condrun, W.-C. Wang, unpublished). Finally, perturbation studies in grasshopper embryo cultures by incubation with purified grasshopper fasciclin I failed to show any obvious altered pathways. These results suggest that if fasciclin I possesses homophilic adhesion activity, it is a low affinity interaction. Similar observations have been obtained for several adhesion molecules including N-CAM (Becker et al., 1989), and cadherins (Takeichi, 1990). Clustering of adhesion molecules laterally to form an organized patch thus may be necessary to achieve sufficient cell-cell binding forces for triggering signaling events (Hynes and Lander, 1992; Takeichi, 1991).

Electron microscopic images of fasciclin I show a compact rectangular structure with dimensions of approximately 80 Å x 60 Å x 40 Å instead of an extended structure, which is a common topology proposed for cell adhesion molecules (Hynes and Lander, 1992; Becker et al., 1989). Moreover, circular dichroism analysis suggests that the novel primary sequence of fasciclin I folds into a predominantly α-helical structure, whereas fasciclin III, a member of the immunoglobulin superfamily, possesses significant β-sheet characteristics (I. Tamir, P. Snow, W.-C. W., unpublished). Therefore, fasciclin I may represent a new structural motif with significant α-helical structure distinct from the β-sheet structures predicted for members of other adhesion receptors, including the immunoglobulin superfamily (Hynes and Lander, 1992).

References

Becker, J. W., Erickson, H. P., Hoffman, S., Cunningham, B. A., and Edelman, G. M. (1989). Topology of cell adhesion molecules. Proc. Natl. Acad. Sci. USA *86*, 1088-1092.

Bunch, T. A., Grinblat, Y., and Goldstein, L. S. B. (1988). Characterization and use of the *Drosophila* metallothionein promoter in cultured *Drosophila melanogaster* cells. Nucleic Acid Res. *16*, 1043-1061.

Carr, S. A., Hemling, M. E., Folena-Wasserman, G., Sweet, R. W., Anumula, K., Barr, J. R., Huddleston, M. J., and Taylor, P. (1989). Protein and carbohydrate structural analysis of a recombinant soluble CD4 receptor by mass spectrometry. J. Biol. Chem. *264*, 21286-21295.

Chang, T. C., Wu, C.-S. C., and Yang, J. T. (1978). Circular dichroic analysis of protein conformation: inclusion of the β-turns. Anal. Biochem. *91*, 13-31.

de Vos, A. M., Ultsch, M., and Kosslakoff, A. A. (1992). Human growth hormone and extacellular domain of its receptor crystal structure of the complex. Science *255*, 306-312.

Elkins, T., Hortsch, M., Bieber, A. J., Snow, P. M., and Goodman, C. S. (1990a). *Drosophila* fasciclin I is a novel homophilic adhesion molecule that along with fasciclin III can mediate cell sorting. J. Cell Biol. *110*, 1825-1832.

Elkins, T., Zinn, K., McAllister, L., Hoffmann, F. M., and Goodman, C. S. (1990b). Genetic analysis of a *Drosophila* neural cell adhesion molecule: interaction of fasciclin I and Abelson tyrosine kinase mutations. Cell *60*, 565-575.

Ferguson, M. A. J., and Williams, A. F. (1988). Cell-surface anchoring of proteins via glycosyl-phosphatidylinositol structures.Annu. Rev. Biochem. *57*, 285-320.

Gastinel. L. N., Simister, N. E., and Bjorkman, P. J. (1992). Expression and crystallization of a soluble and functional form of an Fc receptor related to class I histocompatibility molecules. Proc. Natl. Acad. Sci. USA *89*, 638-642.

Greenfield, N., and Fasman, G. D. (1969). Computed circular dichroism spectra for the evaluation of protein conformation. Biochem. *8*, 4108-4116.

Hynes, R. O., and Lander, A. D. (1992). Contact and adhesive specificities in the associations, migrations and targeting of cells and axons. Cell *68*, 303-322.

Jessell, T. M. (1988). Adhesion molecules and the hierarchy of neural development. Neuron *1*, 3-13.

Kunkel, T. A., Roberts, J. D., and Zakour, R. A. (1987). Rapid and efficient site directed mutagenesis without phenotypic selection. Meth. Enzymol. *154*, 367-382.

Kuroda, K., Geyer, H., Geyer, R., Doerfler, W., and Klenk, H.-D. (1990). The oligosaccharides of influenza virus hemagglutinin expressed in insect cells by a Baculovirus vector. Virol. *174*, 418-429.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacterophage T4. Nature *227*, 680-685.

Morrissey, J. H. (1981). Silver stain for proteins in polyacrylamide gels: a modified procedure with enhanced uniform sensitivity. Anal. Biochem. *117*, 307-310.

Rio, D. C., and Rubin, G. M. (1985). Transformation of cultured *Drosophila melanogaster* cells with a dominant selectable marker. Mol. Cell. Biol. *5*, 1833-1838.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual, 2nd ed. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory.

Staunton, D. E., Dustin, M. L., Drickson, H. P., and Springer, T. A. (1990). The arrangement of the immunoglobulin-like domains of ICAM-1 and the binding sites for LFA-1 and rhinovirus. Cell *61*, 243-254.

Takeichi, M. (1990). Cadherins: a molecular family important in selective cell-cell adhesion. Annu. Rev. Biochem. *59*, 237-252.

Takeichi, M. (1991). Cadherin cell adhesion receptors as a morphogenetic regulator. Science *251*, 1451-1455.

Theil, E. C. (1989). The ferritin family of iron storage proteins. Adv. Enzym. *63*, 421-449.

Towbin, H., Staehelin, T., and Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc. Natl. Sci. USA *76*, 4350-4354.

Williams, A. F., and Barclay, A. N. (1988). The immunoglobulin superfamily-domains for cell surface recognition. Annu. Rev. Immunol. *6*, 381-405.

Zinn, K., McAllister, L., and Goodman, C. S. (1988). Sequence analysis and neuronal expression of fasciclin I in grasshopper and *Drosophila*. Cell *53*, 577-587.

CHAPTER 4

STRUCTURAL STUDIES OF GRASSHOPPER FASCICLIN I BY X-RAY CRYSTALLOGRAPHY

Abstract

This chapter discusses the preliminary structural studies of the CHO-expressed grasshopper fasciclin I by x-ray crystallography. Large single crystals were obtained that diffracted to ~5 Å resolution anisotropically, with diffraction in one direction to 3.5 Å, which is insufficient for a structure determination to atomic resolution. Several strategies have been used to improve the quality of the crystals: (i) to produce a smaller fragment that may be more amenable to crystallization by proteases; (ii) to remove extra carbohydrate by glycosidases; (iii) to express a secreted form of fasciclin I in CHO cells to remove the remainder of the PI-tail; (iv) to express a secreted form in insect expression system; (v) to explore other crystallization conditions. The first three strategies have been tested, but crystals that diffract to higher resolution were not obtained. A new crystal form of CHO-derived secreted fasciclin I using a factorial screen was derived, but it has not been characterized due to the lack of protein. In addition, the truncated form of insect-derived fasciclin I has the expected molecular weight (~74 kDa), and thus contains significantly less carbohydrate, and may be more suitable for crystallization. Future studies will focus on a secreted form of grasshopper fasciclin I using a Baculovirus expression system that allow to produce 10-20 mg per liter of medium (using a multiplicity of infection to insect cells of 1 to 5). Moreover, a crystallographic investigation of *Drosophila* fasciclin I will be initiated because we are now able to purify this molecule (see Chapter 2) using a monoclonal antibody generated against *Drosophila* fasciclin I produced in CHO cells. The ultimate goal is to understand the structural basis of this molecule to help to elucidate its role in neuronal recognition at a molecular level.

I. Crystallization

In principle, the crystallization of a macromolecule is determined by solubility, an equilibrium thermodynamic factor, and by kinetic factors that control nucleation and growth. Crystals grow from nucleation centers that appear at an appropriate degree of supersaturation. This degree of supersaturation can be achieved by varying parameters to decrease the solubility of the molecule of interest, such as ionic strength, pH, temperature, and the presence of counter ions or organic solvents. Supersaturation is a metastable state; its conversion to a more stable state is kinetically controlled. The individual molecules in a crystal are rotationally and translationally constrained. The resulting loss of entropy is overcome by the formation of intermolecular bonds.

The most common approach to crystallizing macromolecules is to decrease the solubility of a concentrated purified protein solution gradually by adding salts or organic solvents as precipitating agents and/or changing the pH or ionic strength of the solution (McPherson, 1985). At high ionic strengths, an adequate quantity of water molecules surrounding the protein molecules is being utilized to form bonds with the small ions, thus proteins "salt-out" of solution. Organic solvent molecules bind water to themselves as do the salt ions, and, in addition, they lower the dielectric constant of the solvent to reduce the effective electrostatic shielding between the macromolecules, and the protein is less soluble. Changes in pH affect protein solubility because proteins are generally least soluble when they have a net charge of zero (the isoelectric point). Nonetheless, the molecule may exhibit a number of solubility minima as a function of pH resulting from its amino acid composition and tertiary structure. The solubility of a protein varies as a function of temperature because the dielectric constant of the solvent decreases with increasing temperature. Since the behavior of macromolecules in solution is complex and unpredictable owing to their various shapes, polyvalent surface character, and dynamic

properties, numerous conditions must be tried to discover a particular one that yields crystals.

Because of the limited amounts of protein available, the "hanging drop" method of Woldawer and Hodgson (1975) was used to screen for suitable crystallization conditions for the PI-form grasshopper fasciclin I. This method is ideally suited for screening a large number of conditions using small amounts of protein, and can also be used to grow large, single crystals for diffraction analysis. With this approach, a drop (usually 2 to 10 microliters) of grasshopper fasciclin I (4.2 to 4.6 mg/ml) in 10 mM Hepes, 0.02% NaN_3 (pH 7.5) was prepared by mixing equal volumes of the protein and the trial precipitant solution on a coverslip, which was then inverted and equilibrated over a well (24-well plate, Linbro) containing 1 ml of the undiluted precipitant solution. In this closed system, the microdroplet comes into equilibrium with the reservoir by slow evaporation. Examination of the crystals obtained was then carried out under a stereomicroscope (Nikon, SMZ-2T).

Initial attempts to crystallize PI-form grasshopper fasciclin I were implemented using precipitating agents including ammonium sulfate, ammonium citrate, ammonium phosphate, sodium phosphate, methylpentanediol and polyethyleneglycol (PEG) of molecular weights 400, 3350 and 8000. The pH was varied between 4.5 and 8.5 using appropriate buffers. Initially, single microcrystals were obtained from 2.0 M ammonium sulfate, pH 7.5, at room temperature. Since the average intensity of the x-ray diffraction pattern of a crystal is roughly proportional to the volume of the crystal and inversely proportional to the unit cell volume, it is necessary to grow crystals of a substantial size. Large single crystals (Fig. 1) with dimensions of 0.5 mm x 0.3 mm x 0.3 mm were obtained using 10-microliter hanging drops equilibrated against wells containing 1.84 M ammonium sulfate, pH 7.5-pH 8.5 or 1.6 M Na, K phosphate, pH 7.5-8.5 (Table 1).

These crystals diffract to 5 Å resolution, with anisotropic diffraction to 3.5 Å, using nickel-filtered CuKα radiation from a rotating-anode x-ray generator.

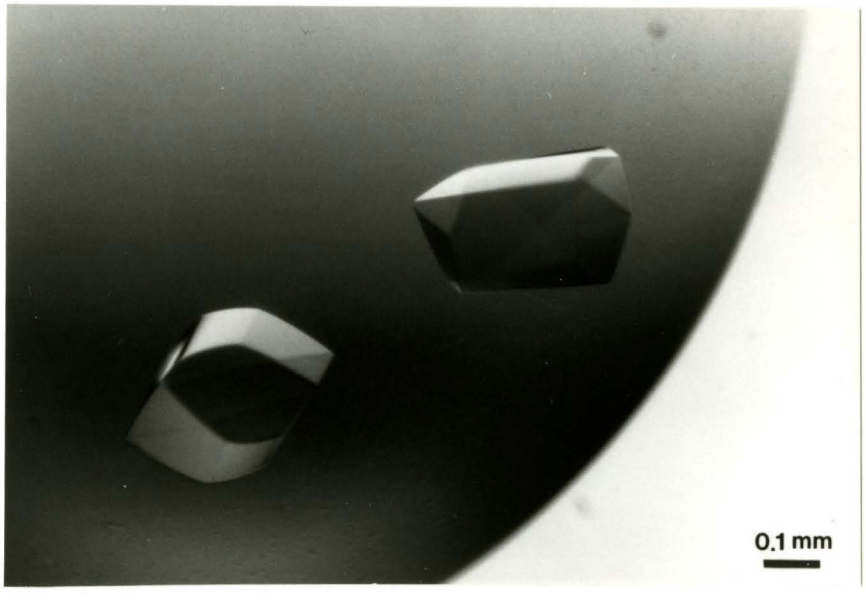

Fig. 1. Crystals of the grasshopper fasciclin I. Crystals were grown by vapor diffusion method and photographed with the crystallizing well.

Table 1. Crystallization of grasshopper fasciclin I

Protein	Precipitant	Characterization of Crystal
PI-linked Gf1/CHO[1]		
4.2-4.6 mg/ml in 10 mM Hepes, pH 7.0, 0.02% NaN_3	1.84 M $(NH_4)_2SO_4$, pH 7.0, 0.02% NaN_3	single crystals, diffract to 3.5 Å anisotropically, with diffraction in two directions to only 5 Å resolution; space group, $C222_1$ 156Å x 352Å x 168Å
	1.6 M Na, K phosphate, pH 7.5-8.5	the same as above
Secreted Gf1/CHO		
5.0 mg/ml in 10 mM Hepes, pH 7.0, 0.02% NaN_3	1.84 M $(NH_4)_2SO_4$, pH 7.0, 0.02% NaN_3	the same as above
	1.6 M Na, K phosphate, pH 7.5-8.5	the same as above
	12.5% PEG 8K, 0.1 M $CaCl_2$, 0.1M Hepes, pH 7.5	not available
	12.5% PEG 8K, 45 mM $MgCl_2$, 0.1M Tris, pH 8.5	not available
	3 M $(NH_4)_2SO_4$, 0.2 M Na,KTartatrate, 0.1 M Hepes, pH 7.5	not available

[1]Gf1 represents grasshopper fasciclin I.

II. Space Group Determination

The first single crystals of the PI-form grasshopper fasciclin I grown in two microliter hanging drops had dimensions of approximately 50 microns. X-ray photographs of this crystal showed diffraction to only 8 Å resolution. When large single crystals (0.5 mm x 0.3 mm x 0.3 mm) became available, screenless procession photographs taken down the c axis showed two mirrors in the zero and upper levels and systematic absences such that only reflections for which $h + k = 2n$ were present (Fig. 2a). Screenless precession photographs taken down the b axis also displayed *mm* symmetry (Fig. 2b), indicating that the space group is either $C222$ or $C222_1$. A 6° screened precession photograph (H0L) showed reflections only for which $00l = 2n$. Therefore the space group is $C222_1$. The unit cell dimensions are a = 156 Å, b = 352 Å, c = 168 Å. Fig. 3. shows the relative orientations of the axes with respect to the external morphology of the crystal.

To determine the number of molecules per asymmetric unit, an average crystal volume per unit molecular weight or V_m (Matthews, 1968) was determined for a range of values. The asymmetric unit of the crystal is estimated to contain between two molecules, corresponding to a V_m equal to 6.4 Å^3/dalton (which corresponds to a solvent content of 81%), and eight molecules, corresponding to a V_m equal to 1.6 (a solvent content of 23%). Based on a survey of 116 different protein crystals, Matthews (1968) found that the fraction of crystal volume occupied by solvent is usually near 43% (V_m equal to 2.4) but may vary between 27% (1.68) and 65% (3.53).

The low resolution diffraction limit of the grasshopper fasciclin I crystals might be the result of the potential structural flexibility between domains of the four-repeat structure (Zinn et al., 1988), and/or chemical heterogeneity from the high degree of glycosylation of this molecule. Alternatively, the remaining PI tail on the C-terminus which is also

heterogeneous and flexible might represent another potential barrier to growth of highly ordered crystals. The next section describes strategies to improve the quality of the crystals.

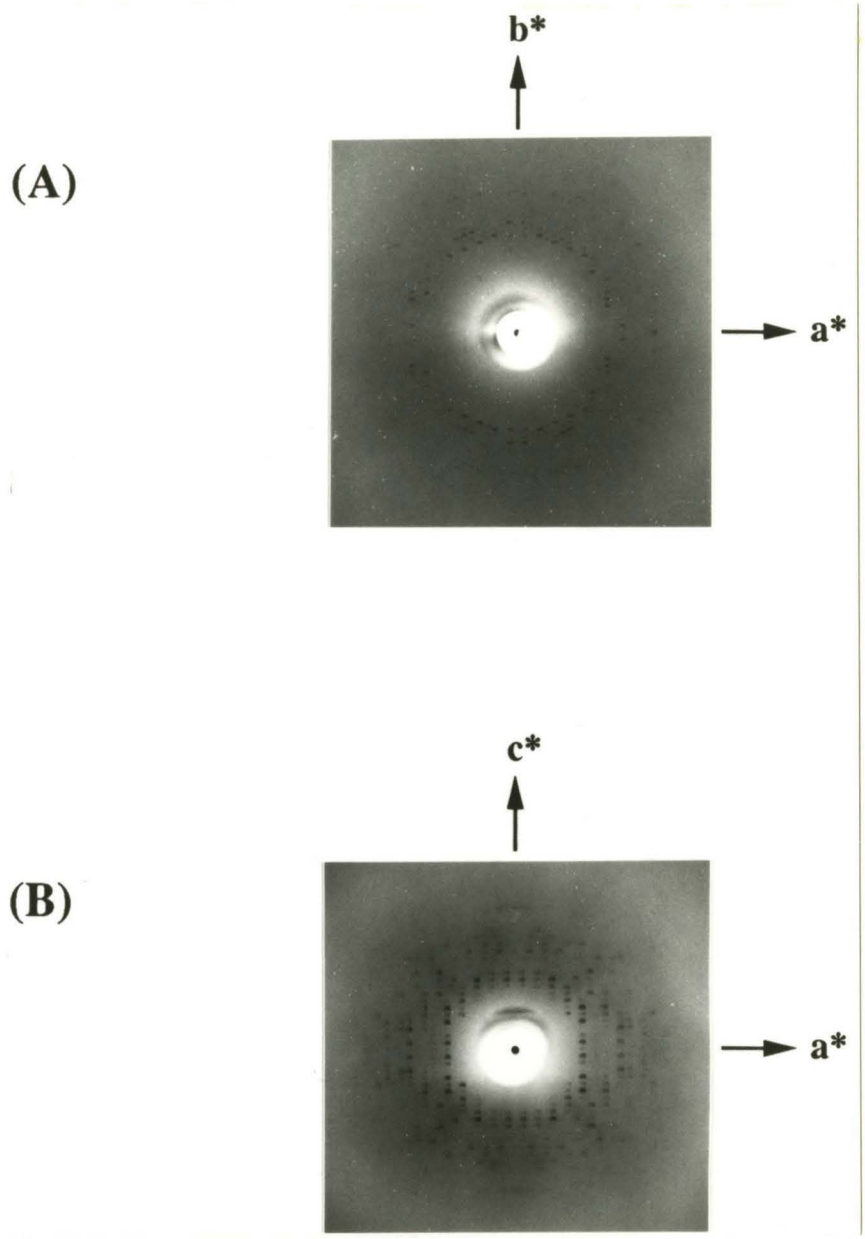

Fig. 2. The reciprocal lattice of the grasshopper fasciclin I crystals. Two 1° screenless precession photographs were taken down the c axis (A) or b axis (B) at a film-to-crystal distance of 10 cm. The exposure time was 3 hr. Reciprocal lattice of fasciclin I crystals has *mm* symmetry, and has symmetry *C2* with systematic absences such that only reflections for which $h + k = 2n$ are present. A 6° screened precession photograph along the b axis showed a $00l = 2n$ selection rule. This establishes the space group as $C222_1$. The unit cell dimensions are $a = 156$ Å, $b = 352$ Å, $c = 168$ Å.

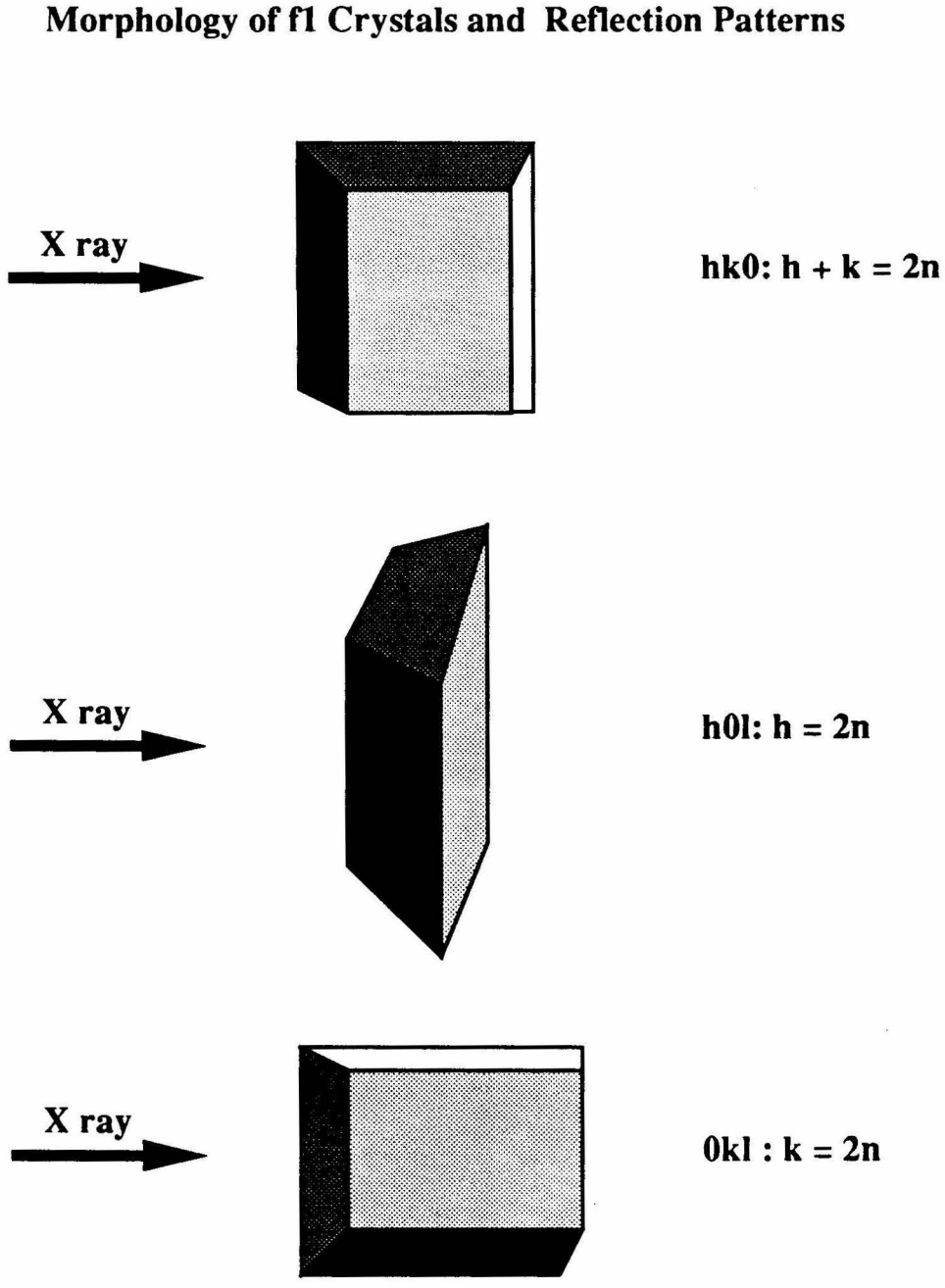

Fig. 3. The relationship of grasshopper fasciclin I crystal morphology and cell edges.

III. Attempts to Improve the Quality of the Grasshopper Fasciclin I Crystals

There are a number of examples in which proteolytic generation of a stable fragment has been used to obtain high quality crystals of a protein that crystallizes poorly in its native state. For example, the Fab fragments of several immunoglobulins (reviewed by Amzel and Poljak, 1979), the ATPase fragment of a 70 K heat-shock cognate protein (Flaherty et al., 1990), and the N-terminal two domains of CD4 (Ryu et al., 1990; Wang et al., 1990). As discussed in Chapter 2, however, the expressed grasshopper fasciclin I is quite proteolytically resistant and no proteases including trypsin, chymotrypsin, elastase and papain produced a stable fragment, suggesting that the protein is folded into a stable structure without obvious flexible linker sequences. This result is in accordance with our observation that the electron micrographs of grasshopper fasciclin I show a compact rectangular structure instead of an extended structure.

In a second approach, we have used glycosidases, including neuraminidase, endo H and endo F to remove extra carbohydrate moieties since CHO-derived grasshopper fasciclin I has approximately 22 kDa of carbohydrates and is very charge heterogeneous (at least 11 species on a native IEF gel; see Chapter 2). This type of strategy has been useful for a soluble form of the human CD8 which yielded diffraction quality crystals only after extensive treatment with neuraminidase, *O*-glycosidase, and V8-protease (Leahy et al., 1992). The glycosylated proteins (see Chapter 2) were repurified by immunoaffinity chromatography and put into crystallizational trials. Single crystals were obtained at the same conditions that yielded crystals of the untreated molecule. However, the resolution limit on a rotating-anode x-ray generator was not improved. These crystals were also examined at the Stanford Synchrotron Radiation Laboratory. Synchrotron radiation is a factor of several hundredfold more intense than x-rays produced by a rotating-anode

generator, and the resolution limits of some crystals have been extended using synchrotron radiation (for example see Leahy et al., 1992). Unfortunately, the fasciclin I crystals did not diffract to higher resolution using synchrotron radiation.

We next tried to express a secreted version of grasshopper fasciclin I in order to remove the PI tail. Two expression systems (CHO cells and *Drosophila* Schneider 2 cells) were used to produce the truncated grasshopper fasciclin I as discussed in Chapter 3. The apparent molecular mass of the truncated grasshopper fasciclin I from CHO cells is ~2 kDa less than the PI-form grasshopper fasciclin I. Crystallization trials with the truncated protein also yield crystals at the same condition. However, these crystals still diffracted anisotropically using either rotating-anode or a synchrotron radiation source. We have also obtained a new crystal form of CHO-derived fasciclin I using a factorial screen (Carter and Carter, 1979) (Table 2). However, we do not have enough protein at this time to optimize the crystallization conditions to produce crystals large enough to characterize.

The optimal protein for crystallization trials is the secreted form of fasciclin I expressed in *Drosophila* S2 cells. This form does not include the remainder of the PI tail since it is directly secreted; nor does it contain extra carbohydrate moieties (its molecular weight is ~75 kDa, in accordance with the expected molecular weight of grasshopper fasciclin I expressed *in vivo*; see Chapter 3). However, the low yield of secreted fasciclin I from the highest-expressing S2 clone (0.5 mg protein/liter after a 4 day induction) makes it impractical to isolate the required amount of protein for crystallization trials.

IV. Future Plans

In collaboration with Peter Snow (Suny Albany), a secreted version of grasshopper fasciclin I was expressed in a Baculovirus, producing 10-20 mg of protein per liter of medium using a multiplicity of infection of 1 to 5. With a large amount of grasshopper fasciclin I available, we may be able to continue pursuing other crystallization conditions.

We will also try to crystallize *Drosophila* fasciclin I because we are now able to produce this molecule (see Chapter 2). Since the *Drosophila* protein is only 48% identical to grasshopper fasciclin I, it may have very different crystallization properties. The ultimate goal is to understand the structure and function of fasciclin I to help in defining its role in neuronal recognition at a molecular level. Thus far, there is no structural information regarding molecules involved in cell-cell interactions in the nervous system. It is also interesting to note that there is a human gene homologous to fasciclin I that has recently been identified by a Japanese group (unpublished). Potentially, I may be able to obtain this cDNA (or clone it myself after the sequence is published) and then express the human analog of fasciclin I by using our well-established expression systems. It would then be of interest to perform structural and functional studies on the human protein.

Table 2. Crystallization conditions of a factorial screening

MPD 150λ (15%) $CaCl_2$ 30λ (45 mM) Acetate 100λ (100 mM) H_2O 720λ	PEG8k 250λ (12.5%) LS 50λ (100 mM) Acetate 100λ (100 mM) H_2O 600λ	PEG8k 750λ (37.5%) $MgCl_2$ 133λ (200 mM) Acetate 100λ (100 mM) H_2O 17λ	LS 500λ (1M) - Acetate 100λ (100 mM) H_2O 400λ	PEG8k 250λ (12.5%) AS 125λ (500 mM) Acetate 100λ (100 mM)) H_2O 525λ	AS 500λ (2M) $MgCl_2$ 30λ (45 mM) Acetate 100λ (100 mM) H_2O 370λ
AS 750λ (3M) Na/KPO_4 50λ (200 mM) Acetate 100λ (100 mM) H_2O 100λ	NaTartrate 500λ (1M) - Acetate 100λ (100 mM) H_2O 400λ	PEG8k 500λ (25%) NaTartrate 50λ (100 mM) Acetate 100λ (100 mM) H_2O 350λ	NaCl 500λ (2M) - Acetate 100λ (100 mM) H_2O 400λ	MPD 100λ (10%) NaTartrate 50λ (100 mM) Acetate 100λ (100 mM) H_2O 750λ	PEG8k 500λ (25%) LS 100λ (200 mM) Acetate 100λ (100 mM) H_2O 300λ
MPD 300λ (30%) NaCl 50λ (200 mM) Acetate 100λ (100 mM) H_2O 550λ	Na/KPO_4 750λ (3M) - Acetate 100λ (100 mM) H_2O 150λ	PEG8k 500λ (25%) LS 100λ (200 mM) Citrate 100λ (100 mM) H_2O 300λ	PEG8k 750λ (37.5%) - Citrate 100λ (100 mM) H_2O 150λ	AS 500λ (2M) NaTartrate 100λ (200 mM) Citrate 100λ (100 mM) H_2O 300λ	PEG8k 250λ (12.5%) - Citrate 100λ (100 mM) H_2O 650λ
PEG8k 250λ (12.5%) AS 125λ (500 mM) Citrate 100λ (100 mM) H_2O 525λ	Na/KPO_4 750λ (3M) NaCl 50λ (200 mM) Citrate 100λ (100 mM) H_2O 100λ	PEG8k 500λ (25%) $MgCl_2$ 133λ (200 mM) Citrate 100λ (100 mM) H_2O 267λ	NaTartrate 500λ (1M) $MgCl_2$ 133λ (200 mM) Citrate 100λ (100 mM) H_2O 267λ	AS 500λ (2M) Na/KPO_4 50λ (200 mM) Citrate 100λ (100 mM) H_2O 350λ	NaCl 500λ (2M) - Citrate 100λ (100 mM) H_2O 400λ
LS 500λ (1M) AS 125λ (500mM) Citrate 100λ (100 mM) H_2O 275λ	Na/KPO_4 750λ (3M) $MgCl_2$ 70λ (105 mM) Citrate 100λ (100 mM) H_2O 80λ	AS 750λ (2M) LS 100λ (200 mM) Citrate 100λ (100 mM) H_2O 50λ	MPD 300λ (30%) Na/KPO_4 50λ (200 mM Citrate 100λ (100 mM) H_2O 550λ	$CaCl_2$ 133λ (200 mM) - Citrate 100λ (100 mM) H_2O 767λ	PEG8k 500λ (25%) - PIPES 100λ (100 mM) H_2O 400λ
NaTartrate 500λ (1M) - PIPES 100λ (100 mM) H_2O 400λ	NaCl 500λ (2M) NaTartrate 100λ (200 mM PIPES 100λ (100 mM) H_2O 300λ	Na/KPO_4 750λ (3M) LS 100λ (200 mM) PIPES 100λ (100 mM) H_2O 50λ	PEG8k 750λ (35%) NaCl 50λ (200 mM) PIPES 100λ (100 mM) H_2O 100λ	AS 500λ (2M) - PIPES 100λ (100 mM) H_2O 400λ	AS 750λ (3M) - PIPES 100λ (100 mM) H_2O 150λ

PEG8k 750λ (35%) Na/KPO_4 50λ (200mM) PIPES 100λ (100 mM) H_2O 100λ	PEG8k 250λ (12.5%) NaTartrate 100λ (200mM) PIPES 100λ (100 mM) H_2O 550λ	PEG8k 500λ (25%) AS 125λ (500 mM) PIPES 100λ (100 mM) H_2O 275λ	MPD 300λ (30%) - PIPES 100λ (100 mM) H_2O 600λ	$CaCl_2$ 133λ (200 mM) - PIPES 100λ (100 mM) H_2O 767λ	NaCl 500λ (2M) Na/KPO_4 50λ (200 mM) PIPES 100λ (100 mM) H_2O 350λ
MPD 300λ (30%) $MgCl_2$ 133λ (200mM) PIPES 100λ (100 mM) H_2O 467λ	LS 500λ (1M) - PIPES 100λ (100 mM) H_2O 400λ	LS 500λ (1M) - HEPES 100λ (100 mM) H_2O 400λ	PEG8k 250λ (12.5%) $CaCl_2$ 67λ (100 mM) HEPES 100λ (100 mM) H_2O 583λ	AS 750λ (3M) NaCl 50λ (200 mM) HEPES 100λ (100 mM) H_2O 100λ	PEG8k 500λ (25%) - HEPES 100λ (100 mM) H_2O 400λ
AS 750λ (3M) NaTartrate 100λ (200mM) HEPES 100λ (100 mM) H_2O 50λ	AS 500λ (2M) LS 100λ (200 mM) HEPES 100λ (100 mM) H_2O 300λ	MPD300λ (30%) AS125λ (500 mM) HEPES 100λ (100 mM) H_2O 475λ	Na/KPO_4 750λ (3M) NaTartrate100λ (200mM) HEPES 100λ (100 mM) H_2O 50λ	NaTartrate 500λ (2M) Na/KPO_4 50λ (200mM) HEPES 100λ (100 mM) H_2O 350λ	PEG8k 750λ (37.5%) - HEPES 100λ (100 mM) H_2O 150λ
NaCl 500λ (2M) - HEPES 100λ (100 mM) H_2O 400λ	PEG8k 250λ (12.5%) $MgCl_2$ 133λ (200 mM) HEPES 100λ (100 mM) H_2O 517λ	PEG8k 750λ (37.5%) LS 100λ (200 mM) HEPES 100λ (100 mM) H_2O 50λ	LS 500λ (1M) NaTartrate 100λ (200 mM) TRIS 100λ (100 mM) H_2O 300λ	NaCl 500λ (2M) LS 100λ (200 mM) TRIS 100λ (100 mM) H_2O 300λ	PEG8k 500λ (25%) Na/KPO_4 50λ (200 mM) TRIS 100λ (100 mM) H_2O 350λ
PEG8k 500λ (25%) NaCl 50λ (200 mM) TRIS 100λ (100 mM) H_2O 350λ	AS 500λ (2M) NaCl 50λ (200 mM) TRIS 100λ (100 mM) H_2O 350λ	PEG8k 750λ (37.5%) NaTartrate 100λ (200 mM) TRIS 100λ (100 mM) H_2O 50λ	NaTartrate 500λ (1M) - TRIS 100λ (100 mM) H_2O 400λ	MPD 300λ (30%) LS 100λ (200 mM) TRIS 100λ (100 mM) H_2O 500λ	AS 750λ (3M) - TRIS 100λ (100 mM) H_2O 150λ
PEG8k 150λ (7.5%) $CaCl_2$ 30λ (45mM) TRIS 100λ (100 mM) H_2O 720λ	Na/KPO_4 750λ (3M) - TRIS 100λ (100 mM) H_2O 150λ	PEG8k 250λ (12.5%) $MgCl_2$ 30λ (45 mM) TRIS 100λ (100 mM) H_2O 620λ	MPD 300λ (30%) AS 125λ (500 mM) TRIS 100λ (100 mM) H_2O 475λ	$CaCl_2$ 133λ (200 mM) - TRIS 100λ (100 mM) H_2O 767λ	

Abbreviation: AS, ammonium sulfate; LS, lithium sulfate; MPD, methylpentanediol; PEG, polyethyleneglycol; λ, microliter.
Solutions:1.5M $CaCl_2$, 1.5M NaTartrate, 1.5M NaCl, 1.5M $MgCl_2$, 1.5M LS, sat. AS, MPD, 50% w/w PEG8k, 4.0M Na/KPO_4, {4M NaH_2PO_4 + 4M K_2PO_4}
Buffers:1.0M Acetate (pH 4.6), 1.0M Citrate(pH 5.6), 1.0M PIPES (pH 6.5), 1.0M HEPES (pH 7.5), 1.0M TRIS (pH 8.5).

References

Amzel, L. M., and Poljak, R. J. (48). Three-dimensional structure of immunoglobulins. Annu. Rev. Biochem.*48*, 961-997.

Carter Jr., C. W., and Carter, C. W. (1979). Protein crystallization using incomplete factorial experiments. J. Biol. Chem. *254*, 12219-12223.

Flaherty, K. M., Delucaflaherty, C., and Mckay, D. B. (1990). Three-dimensional structure of the ATPase fragment of a 70 K heat-shock cognate protein. Nature *346*, 623-628.

Kwong, P. D., Ryu, S. E., Hendrickson, W. A., and Axel, R. (1990). Molecular characteristics of recombinant human CD4 as deduced from phlymorphic crystals. Proc. Natl. Acad. Sci. USA *87*, 6423-6427.

Leahy, D. J., Axel, R., and Hendrickson, W. A. (1992). Crystal structure of a soluble form of the human T cell coreceptor CD8 at 2.6 Å resolution. Cell *68*, 1145-1162.

Matthews, B. W. (1968). Solvent content of protein crystals. J. Mol. Biol. *33*, 491-497.

McPherson, A. (1985). Crystallization of macromolecules: General principles. In *Meth. Enzymol.* ed. by Wyckoff, H. W., Hirs, C. H. W., and Timasheff, S. N., Academic Press., pp. 112-120.

Ryu, S.-E., Kwong, P. D., Truneh, A., Porter, T. G., Arthos, J., Rosenberg, M., Dai, X., Xuong, N.-h., Axel, R., Sweet, R. W., and Hendrickson, W. A. (1990). Crystal structure of an HIV-binding recombinant fragment of human CD4. Nature *348*, 419-426.

Wang, J., Yan, Y., Garrett, T. P. J., Liu, J., Rodgers, D. W., Garlick, R. L., Tarr, G. E., Husain, Y., Reinherz, E. L., and Harrison, S. C. (1990). Atomic structure of a fragment of human CD4 containing two immunoglobulin-like domains. Nature *348*, 411-419.

Woldawer, A., and Hodgson, K. O. (1975). Crystallization and crystal data of monellin. Proc. Natl. Acad. Sci. USA *72*, 398-399.

APPENDIX A

STRUCTURAL STUDIES OF A CD4 FRAGMENT CONTAINING TWO IMMUNOGLOBULIN-LIKE DOMAINS

Abstract

CD4, the cell-surface glycoprotein, found mainly on a subset of T lymphocytes and on macrophages. The primary function of CD4 is to associate with class II major histocompatibility (MHC) molecules on the surface of antigen-presenting cells, facilitating antigen recognition by the T cell receptor. The CD4/class II MHC interaction is also essential for thymocyte differentiation. Furthermore, human CD4 is the receptor for HIV infection. CD4 contains four immunoglobulin-like domains in its extracellular segment. The HIV-binding site on CD4 lies on the first N-terminal domain, whereas the class II binding sites are located on the first two domains. To elucidate a structure-function relationship at atomic resolution, a crystallographic study of the first two immunoglobulin-like domains of the human CD4 derived from a CD4 immunoadhesin chimeric molecule was initiated in collaboration with Steve Chamow at Genentech. Large single crystals were obtained with diffraction beyond 3 Å and a complete native data set to 2.7 Å was collected using a single crystal. However, the structure of the same CD4 fragment was solved by two other laboratories (Wayne Hendrickson and Steve Harrison), while we were screening heavy atom derivatives. Therefore we did not further pursue this project. Instead, a structural analysis of murine CD4 was initiated for an eventual mapping of the interaction between murine CD4 and a murine MHC class II protein, since our laboratory has been working on the structure of a murine MHC class II molecule. Section III describes the expression and purification of the first two domains of the murine CD4 (mV1V2) using the glutamine synthetase-based expression system (see Chapter 2). The CHO-derived mV1V2, however, acts as a covalent multimers, indicating improper folding of the mV1V2 fragment mediated by disulfide bonds, and thus is not suitable for crystallization. Attempts to examine the possible mechanism resulting in misfolding are discussed.

I. Introduction

T lymphocytes recognize peptide fragments of antigens bound to class I or class II major histocompatibility complex (MHC) molecules (reviewed in Townsend and Bodmer, 1989). Recognition of the peptide-MHC molecules complex is mediated by the T-cell receptor (TcR) and either the CD4 or CD8 glycoprotein on the surface of the T cells. Evidence is accumulating that the hypervariable regions of the TcR contact the top surface of the peptide/MHC complex, whereas CD4 and CD8 bind to monomorphic determinants on the MHC class II and MHC class I molecules, respectively, functioning as co-receptor along with the αβ TcR heterodimer (reviewed in Townsend and Bodmer, 1989; Fig. 1). Peptides bound by MHC class I molecules are recognized by $CD8^+$ cytotoxic T lymphocytes (CTLs), whereas peptides bound to MHC class II molecules are recognized by $CD4^+$ (primarily helper) T cells. Recognition of a specific peptide-class I complex usually results in the death of the cell displaying it. Recognition of a peptide-class II complex induces proliferation of the $CD4^+$ T cells and amplification of the immune response against the particular immunogen from which the peptide was derived. The engagement of CD4 and CD8 to MHC molecules may serve as a possible route for ligand-dependent signal transduction by association with a specific tyrosine kinase ($p56^{lck}$) inside the cells with CD4 and CD8, thus triggering an intracellular response (Veillette et al., 1988).

CD4 and CD8 are also critical in thymocyte differentiation and maturation, a process known as positive and negative selection. Biochemical and genetic studies suggest that interaction of non-differentiated $CD4^+CD8^+TcR^+$ T lymphocytes class I MHC eliminates the expression of CD4, whereas interaction with class II MHC extinguishes CD8 expression, in support of biochemical as well as genetic studies (Ramsdell and Fowlkes, 1989; Zuniga-Pflucker et al., 1990; Robey et al., 1991; Fung-Leung et al., 1991). This

results in the selection of class II-reactive cells (or class I-reactive cells) that are committed to a CD4 lineage (or a CD8 lineage).

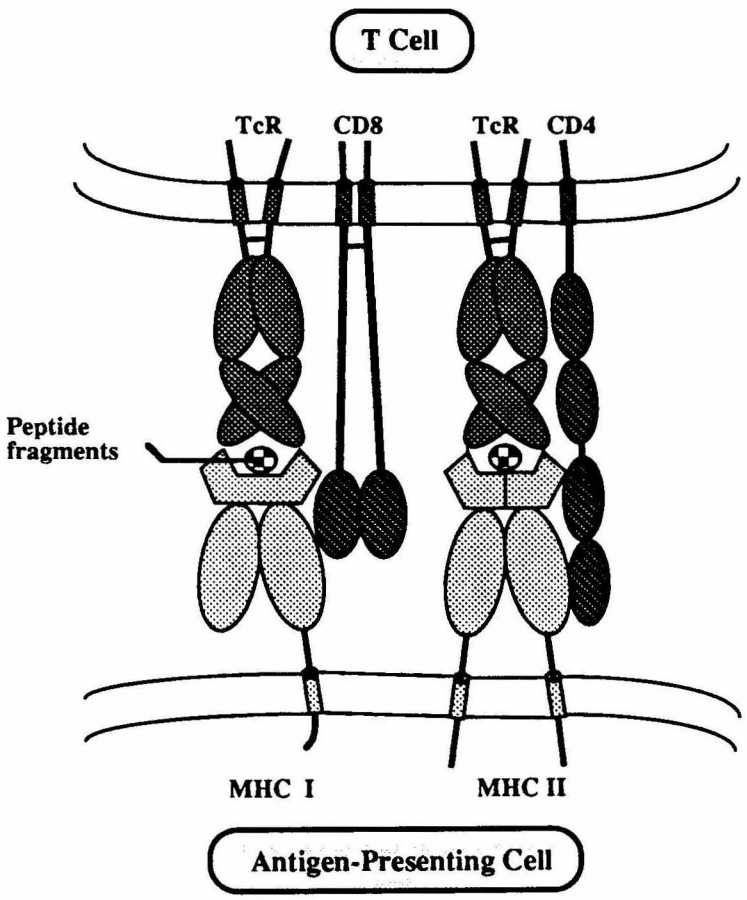

Fig. 1. A schematic diagram showing the interaction of CD4 or CD8 with T-cell receptor and MHC molecules. Petide fragments presented by MHC molecules are recognized by T cells through a T-cell receptor and its co-receptor, CD4 or CD8.

In humans, CD4 serves as the primary cellular receptor for infection by the human immunodeficiency virus (HIV), which is the causative agent of acquired immunodeficiency syndrome, AIDS (Klatzmann et al., 1984; Dalgleish et al., 1984; reviewed by Sattentau and Weiss, 1988). Infection by HIV is initiated by binding of the viral envelope protein, gp120, to CD4 on the T lymphocytes, followed by direct fusion of viral envelope with the cell membrane to introduce the viral RNA genome into the cytoplasm (reviewed by Robey and Axel, 1990).

CD4 is a 55 kDa molecule, consisting of a 372-amino acid extracellular portion, a 23-amino acid transmembrane domain, and a 38-amino acid cytoplasmic segment (Maddon et al., 1985, 1987; Clark et al., 1987). The extracellular portion of CD4 appears to have four tandem repeats with sequence similarity to immunoglobulin variable (V) regions (designated V1-V4) (Maddon et al., 1985). The gp120-binding site on CD4 lies on the first N-terminal domain (V1) (Peterson and Seed, 1988; Landau et al., 1988; Arthos et al., 1988; Clayton et al., 1989; Brodsky et al., 1990; Ashkenazi et al., 1990), whereas the class II binding site may be located on V1 and V2 domains (Clayton et al., 1989; Lamarre et al., 1989a, 1989b; Fleury et al., 1991).

To obtain a clear picture of how the V-like domains of CD4 are folded, and define their mode of interaction with gp120 and MHC class II molecule, many laboratories have attempted to crystallize a soluble fragment of CD4 composed of the V1-V4 domains. Although crystals have been obtained, they diffract to ~6 Å only (Davis et al., 1990; Kwong et al., 1990), perhaps because of a flexible linker between V2 and V3 domains. Proteolytic digests suggested that the first two domains (V1-V2) form a stable structure, and therefore might be more amenable to crystallization (Richardson et al., 1988; Ibegbu et al., 1989; Healey et al., 1990). In collaboration with Steve Chamow at Genentech, a human V1V2 fragment was generated by an enzymatic cleavage of a CD4 immunoadhesin

(Capon et al., 1984; Chamow et al., 1990), for a crystallographic structure analysis. Crystals of this fragment diffract beyond 3 Å, which we will report in Section II. While crystallographic studies were underway in our laboratory, structure of the same fragment obtained by direct expression (Arthos et al., 1989; Garlick et al., 1990) was solved by two other groups to 2.3 (Ryu et al., 1990) or 2.4 Å (Wang et al., 1990) resolution. We therefore did not pursue the structure determination of the human CD4 V1V2.

As reported by Ryu et al. (1990) and Wang et al. (1990), the V1V2 fragment forms a rigid rod of about 60-65 Å in length. As predicted, both domains resemble the immunoglobulin fold, two antiparallel β sheets packed against each other. An unusually long β strand forms the last strand of the V1 domain and continues as the first strand of the V2 domain, so that the two domains are arranged in tandem, one on top of the other. The V1 domain shares significant structural similarities with immunoglobulin Vκ domains; however, the counterpart of the immunoglobulin complementarity-determining region (CDR) 2 loop is elongated while the counterpart of the CDR3 loop is shortened. The folding of V2 resembles that of an immunoglobulin constant domain, but there is an unusual intra-sheet disulfide bond. The CDR2-equivalent loop in the V1 domain, protruding out from the core as a prominent ridge, is thought to interact with a complementary groove in the gp120. Most of the residues critical for gp120 binding lie in this ridge. The "real" function of the CD4 molecule, involving the binding of the MHC class II antigens, was recently mapped in a surface opposite the gp120-binding site, using a mutational analysis (Fleury et al., 1991).

Because our laboratory is working on a structural analysis of a murine class II MHC molecule, we decided to attempt an analysis of the structure of murine CD4 to allow an eventual mapping of their interaction. The sequence of murine CD4 is 55% homologous to human CD4 (Littman and Gettner, 1987), but does not serve as a receptor for HIV. It

would be therefore interesting to compare two structures. Using the glutamine synthetase-based expression system described in Chapter 2, a Chinese Hamster Ovary (CHO) cell line that expresses a fragment containing V1 and V2 domains (mV1V2) was generated. However, for unknown reasons, the expressed protein forms covalent multimers and therefore cannot be used for crystallization, as discussed in Section III.

II. Structural Studies of a Human CD4 Fragment, V1V2

A. Crystallization of V1V2

Based on the inability of other labs to crystallize V1-V4 in a form that diffracts to high resolution, a project was initiated in collaboration with Steve Chamow and Dan Capon at Genentech towards a structural analysis of the V1V2 fragment of human CD4. This fragment is derived from papain cleavage of a CD4-immunoglobulin hybrid molecule (CD4 immunoadhesin), which is a homodimer resembling an immunoglobulin without a light chain (Fig. 2) (Byrn et al., 1990). Each monomer of the CD4 immunoadhesin consists of the V1 and V2 domains of human CD4 joined at its carboxyl terminus to the hinge region and Fc domain of IgG. Cleavage with papain releases V1V2 from the Fc fragments. Two forms of V1V2 that differ by approximately 1 kDa in apparent molecular mass were consistently observed, presumably due to two cleavage sites in the hinge region (for details, see Chamow et al., 1990; Appendix B).

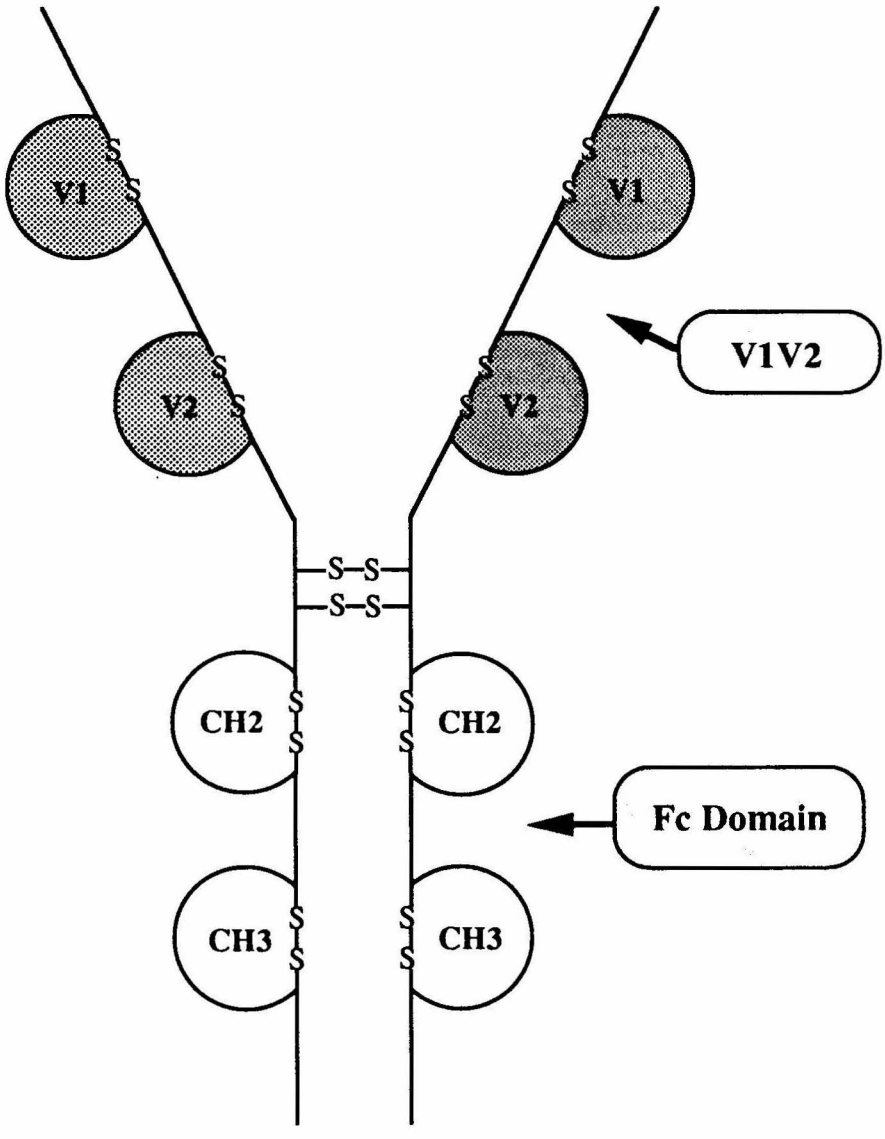

Fig. 2. Construction of the human CD4 immunoadhesin. The human CD4 immunoadhesin is constructed by the first two immunoglobulin-like domains (V1 and V2) of CD4 (shaded), and the Fc domain of IgG1.

Initial crystallization trials to find suitable conditions yielding crystals of the V1V2 fragment were done by the hanging drop method as described in Chapter 4 (Wlodwer and Hodgson, 1975). A droplet of the V1V2 fragment (7.9 mg/ml) in 50 mM MES, 0.1% NaN_3 (pH 6.0), was prepared by mixing equal volumes of the protein and the trial precipitant on a coverslip, which was then inverted and equilibrated over a well (24-well plate, Linbro) containing 1 ml of the undiluted precipitants. Crystallizations were attempted using precipitants including sodium citrate, sodium acetate, magnesium sulfate, polyethylene glycol (PEG) 600, 3350, 8000, methylpetanediol, ethanol, isopropanol, methanol, and acetone at room temperature. The growth of crystals was examined under stereomicroscope (Nikon, SMZ-2T).

Microcrystals (Fig. 3A) were obtained during the first attempted crystallization of V1V2 with 16%-20% PEG 3350 (4K) which grew as many long and thin plates extending from the same origin. PEG is a polymer of various lengths that can perturb the natural structure of water to form a network composed partly of water. Because of this restructuring of water, proteins are excluded into a phase where they may crystallize. To verify that the thin plate-like crystals were composed of protein, crystals in three drops were collected in a capillary tube, washed three times, and run on a 12% SDS-PAGE gel. Two bands corresponding to molecular mass 19 kDa and 20 kDa were shown on the gel, indicating that both forms of the V1V2 fragment were crystallized.

To improve the quality of crystals, crystallization attempts were tried in solutions containing PEG 4K at various pH (pH 6.5-8.5), as well as different salts (0.1 M ammonium sulfate, 0.2 M ammonium acetate, 0.2 M ammonium chloride, 0.2 M magnesium sulfate, 0.2 M lithium sulfate, 0.2 M magnesium chloride, 0.2 M calcium chloride, and 0.2 M sodium citrate). The best condition to produce bigger polycrystalline crystals was 16% -20% PEG 4K, 0.1 M ammonium sulfate, 0.1 M Tris, pH 8.5. The

biggest polycrystalline aggregate (~100 μm x 50 μm x 50 μm) diffracted beyond 3 Å using nickel-filtered CuKα radiation from a rotating-anode x-ray generator.

Since none of the crystallization conditions gave rise to large single crystals, two forms of seeding were attempted: microseeding and macroseeding (McPherson, 1976). In the microseeding technique, polycrystalline crystals were crushed into small nucleation particles (1 to 10 μm) in the mother liquor. A dilution of the crushed seeds to the mother liquor is made from 1:10 to $1:10^9$ and then used as a precipitating agent to set up for crystallization. It is only by this technique that small yet single crystals (~50 μm) grew. To obtain bigger crystals, the macroscopic seeding technique was used (Thaller et al., 1981). In this technique, a small single crystal is carefully washed to remove crystalline debris and denatured molecules on the crystal surface. The crystal is then transferred to a fresh protein solution, allowing it to grow further. This process can be repeated until the crystals have reached the desired size. Crystals were usually observed in 1-2 days after macroseeding. These crystals grew to their full size (~0.5 mm) in 1-2 weeks (Fig. 3B).

B. Space Group Determination of V1V2

Precession photographs of single crystals with the x-ray beam along the c axis (HK0) showed reciprocal lattice symmetry *C2* with systematic absences such that only reflections for which $h + k = 2n$ are present (Fig. 4A). A precession photograph down *b* (H0L), coinciding with the long edge of the V1V2 crystal, showed reflections for which h = 2n are present (Fig. 4B). This establishes the space group as *C2* with unit cell dimensions a = 134.5 Å, b = 33.1 Å, c = 45.6 Å and $\beta = 96^{o}$. The asymmetric unit of the crystal is estimated to contain one molecule based on average volume to mass ratios ($V_m = 2.5$) of protein crystals with a solvent content of 52% (Matthews, 1968).

(A)

(B)

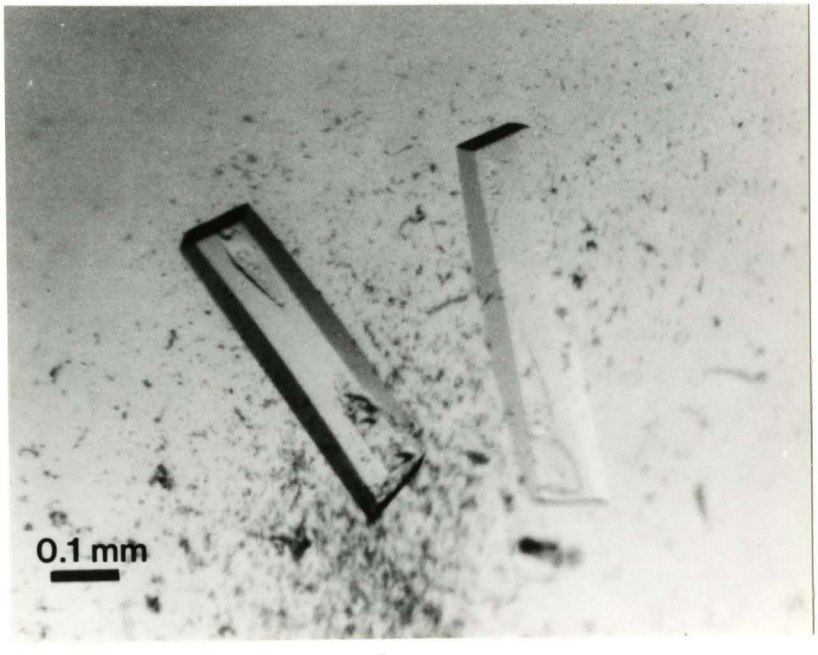

Fig. 3. Crystals of the human V1V2 fragment. Crystals were grown directly (A) or with seeding techniques (B) by vapor diffusion method and photographed with the crystallizing well.

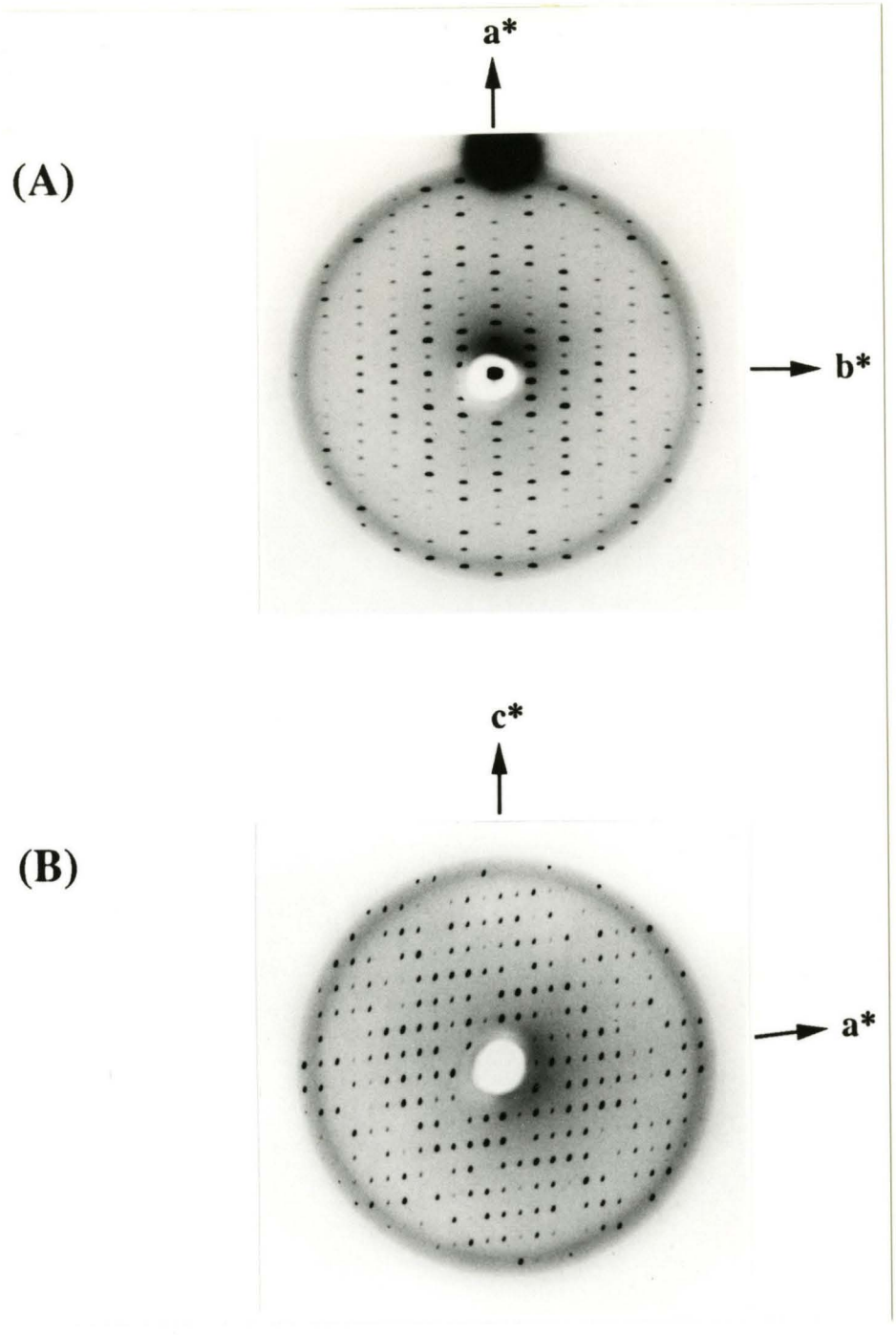

Fig. 4. The reciprocal lattice of the human V1V2 crystals. Two 8° precession photographs of the *hk0* (A) and *h0l* (B) projections of V1-V2 crystals taken at a film-to-crystal distance of 10 cm, operated at 50 kV, 150 mA. The exposure time was 15 hr. Reciprocal lattice of V1-V2 crystals has symmetry C_2 with systematic absences such that only reflections for which $h + k = 2n$ are present. This establishes the space group as $C2$. The unit cell dimensions are a = 134.5 Å, b = 33.1 Å, c = 45.6 Å and β = 96°.

C. Native Data Collection and Screening of Heavy Atom Derivatives

Crystals of V1V2 are relatively radiation resistant. A complete 2.7 Å native data set was collected from one crystal on a Siemens electronic detector mounted on a rotating anode generator producing nickel-filtered CuKα radiation in 16 hours and 40 minutes (1 minute/frame; 0.1 deg/frame). The crystal was mounted with the unique axis along the spindle so that each photograph contains a reflection and its Friedel reflection. In this orientation, 180 degree of scanning (ω) results in fourfold redundancy of the complete data set. The collected data set was processed with the XENGEN software package including autoindexing, integrating, scaling and reducing. Table 1 shows the data collection statistics including completeness of data and merging R values as a function of scattering angle.

Screening for heavy atom derivatives was done by comparison of screened precession photographs of major zones of crystals soaked in heavy-atom compounds to precession photographs of native crystals. Intensity changes of some reflections with no alteration in the unit cell dimensions as compared with the native precession photograph indicate a potential isomorphous heavy atom derivative. Several heavy-atom compounds, including K_2PtCl_4, TBA_4Pt_2, and $UO_2(Ac)_2$, were tried by soaking with the crystals in mother liquor. A potential derivative, obtained by soaking with 3 mM K_2PtCl_4 for 1-2 days, was identified by an *h0l* precession photograph (8°). A detector data set was collected. Significant changes in intensity as compared to the native data were not observed, perhaps because of the presence of 0.1 M $(NH_4)_2SO_4$ in the mother liquor, which could produce NH_3 and react with K_2PtCl_4 to produce $K_2Pt(NH_3)_4$. At this point in time, we learned that this structure had been solved by two other groups (Ryu et al., 1990; Wang et al, 1990) and further data collection was not pursued. Appendix B describes the purification and crystallization of the V1V2 fragment.

Table 1. Data collection statistics of the native V1V2

Bragg Spacing	# Reflections[1]	#Measured Total[2]	#Unique[3]	%Data[4]
∞ - 4.93	1014	3870	1009	99.5
4.93-3.91	966	3905	965	99.9
3.91-3.42	963	3959	963	100.0
3.42-3.11	946	3844	944	99.8
3.11-2.88	948	2829	743	78.4
2.88-2.71	934	1397	406	43.5
∞ - 2.71	5771	19804	5030	87.1

Data collection results are shown for various successive ranges of interplanar shells. Values for number of predicted unique reflections, total number of measured reflections, number of unique reflections that exist, and fraction of theoretically possible data measured.

[1]Number of predicted unique reflections.

[2]Total number of measured reflections.

[3]Number of unique reflections that exist

[4]Fraction of theoretically possible data observed.

III. Expression of a Murine CD4 Fragment in Chinese Hamster Ovary (CHO) Cells

A. Expression and Purification

A cDNA fragment encoding the first 184 amino acids of murine CD4 was created by polymerase chain reaction (PCR) (Saiki et al., 1988) using the following oligonucleotides as primers:
GTCGAATTCACCACCATGTGCCGAGCC and
GCTGGATCCCTATGTGCTCTGAAAAACCCCAAG. A PCR fragment was generated so that an EcoRI site was at the 5' end, followed by 639 bp encoding the V1 and V2 domains, and a stop codon followed by a BamHI site was introduced after amino acid 184. This EcoRI-BamHI fragment was then cloned into the EcoRI and BamHI sites of the polylinker of the expression vector pBJ5.GS (Gastinel et al., 1992) to obtain pBJ5.GS.mCD4. This cloning task was done by Pamela J. Bjorkman.

To obtain CHO cells expressing the V1V2 fragment of murine CD4, we used the glutamine-synthetase-based amplifiable expression system as described in Chapter 2. For the selection of clones expressing high levels of the truncated mV1V2, medium was filtered through a nitrocellulose paper using a Minifold II slot blot system (Schleicher & Schuell #SRC072/0), and the presence of mV1V2 was verified by immunostaining, using the monoclonal antibody, GK1.5. Cells of the clone producing the highest level of mV1V2 were grown in a Cell Pharm II hollow-fiber (Unisyn Fibertec, San Diego) bioreactor device in the presence of 100 μM of methionine sulfoxime.

To purify mV1V2, the monoclonal antibody GK1.5 was used for immunoaffinity chromatography. Ascites fluid containing GK1.5 (~10 mg/ml) was partially purified to

~70% pure by 50% $(NH_4)_2SO_4$ precipitation and dialyzed against 0.2 M $NaHCO_3$, 0.5 M NaCl, pH 8.6 (coupling buffer). About 50 mg of GK1.5 in 5 ml of coupling buffer were covalently linked to 7 ml of CNBr-activated Sepharose (~8 mg/ml) according to the manufacturer's directions (Pharmacia). Alternatively, Affi-Gel-10-GK1.5 coupled beads (~15 mg/ml capacity) were made in 0.1 M MES, pH 6.1, in which GK1.5 but not albumin (a major contaminant in the mouse ascites fluid) linked to the Affi-Gel 10 beads (BioRad). About 500 ml of supernatants from Cell Pharm (daily harvest) were put on the column at a flow rate of 20 ml/hr at 4°C, and the column was washed with 500 ml of PBS, 0.05% sodium azide with protease inhibitors (1 mM EDTA, 100 μM PMSF, 0.5 μg/ml leupeptin, 1 μg/ml aprotinin). The protein was then eluted by 0.1 M sodium acetate, pH 3.2. Fractions of 3 ml were collected in test tubes containing 0.5 ml of 1 M phosphate (pH 8.2) for immediate neutralization (Fig. 5). In general, about 0.3 mg of mV1V2 was obtained per daily harvest from the Cell Pharm containing a 250-ml hollow-fiber bioreactor (Table 2).

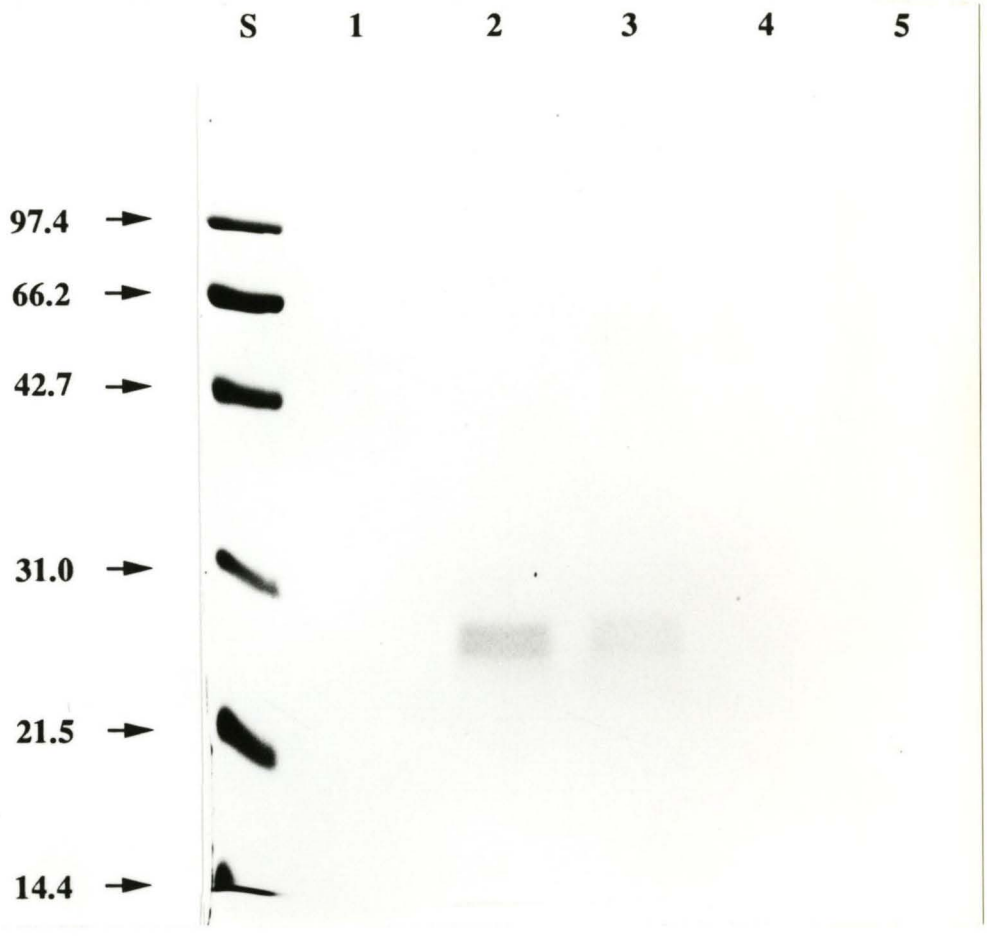

Fig. 5. SDS-PAGE analysis of eluted fractions of purified mV1V2 expressed in CHO cells. Fractions (lane 1 to lane 5) containing purified mV1V2 proteins eluted from the GK1.5 monoclonal antibody immunoaffinity column (pH 3.2) were subjected to 12% SDS-PAGE and visualized by coomassie brilliant blue R-250 staining.

Table 2. Harvest of mV1V2 in Cell Pharm II[1]

##	Harvest Date	Harvest Volume (ml)	Purified Protein (mg)
1	8-05-91	350	0.4
2	8-09-91	300	0.6
3	8-13-91	400	0.4
4	8-16-91	220	0.4
5	8-19-91	200	0.5
6	8-22-91	510	0.5
7	8-26-91	490	0.5
8	8-28-91	570	0.5
9	8-31-91	640	1.1
10	9-03-91	820	0.5
11	9-05-91	810	0.5
12	9-06-91	330	0.3
13	9-10-91	760	0.3
14	9-12-91	700	0.4
15	9-13-91	430	0.3
16	9-16-91	620	0.3
17	9-18-91	400	0.3
18	9-19-91	580	0.3
19	9-20-91	520	0.3
20	9-23-91	620	0.3
21	9-24-91	770	
22	9-26-91	550	
23	9-30-91	610	
24	10-02-91	500	
25	10-04-91	530	
26	10-07-91	600	
27	10-09-91	710	
28	10-11-91	420	
29	10-14-91	430	
30	10-15-91	450	

[1]10^8 cells of clone CD4.27 were inoculated into Cell Pharm II on 07-30-91.

B. Oligomeric Multimers of the Expressed Soluble mV1V2

Purified mV1V2 from CHO cells migrates in two forms with an apparent molecular mass 28 kDa and 29 kDa (Fig. 5) on a reducing 12% SDS polyacrylamide gel (SDS-PAGE). This might be due to glycosylation heterogeneity since there is one potential N-linked glycosylation site at amino acid 161. Similar results have been observed for the expressed rat V1V2 fragment from CHO cells, which is expected to have one-carbohydrate structure (Williams et al., 1989). When the purified mV1V2 was run on an FPLC Superose 12 column, however, peaks were observed at the position of the void volume, at a position corresponding to a monomer, and intermediate positions (Fig. 6). Fractions from the gel filtration chromatography were subjected to 12% SDS-PAGE in either non-reducing or reducing condition. As shown in Fig. 6, fractions of higher molecular weights appeared to correspond to covalently linked oligomers, since they could be reduced to a monomer in the presence of β-mercaptoethanol. To examine if these oligomers were produced during the low pH elution from the affinity column, fractions containing multimers were collected as pool I, and the monomer fractions were collected as pool II. Pool I was repurified by the immunoaffinity chromatography, eluted, concentrated and run on the FPLC Superose 12 column as described above. Fig. 7 shows a gel filtration profile of the repurified multimers, indicating that, first, the eluted oligomeric forms can interact again with the monoclonal antibody, GK1.5, and, second, that they remain as oligomers after repurification and re-elution. Pool II, composed of the monomer of mV1V2, similarly, was re-chromatographed using FPLC Superose 12 column. As expected, the major peak migrates in the position of a monomer. Nonetheless, proteolytic fragments, ~12 kDa and ~6 kDa, appeared, indicating that either mV1V2 monomer and/or its multimers are unstable. These results strongly suggest that the expressed mV1V2 forms covalent oligomers mediated by disulfide bonds, and may indicate improper folding of the V1V2 fragment.

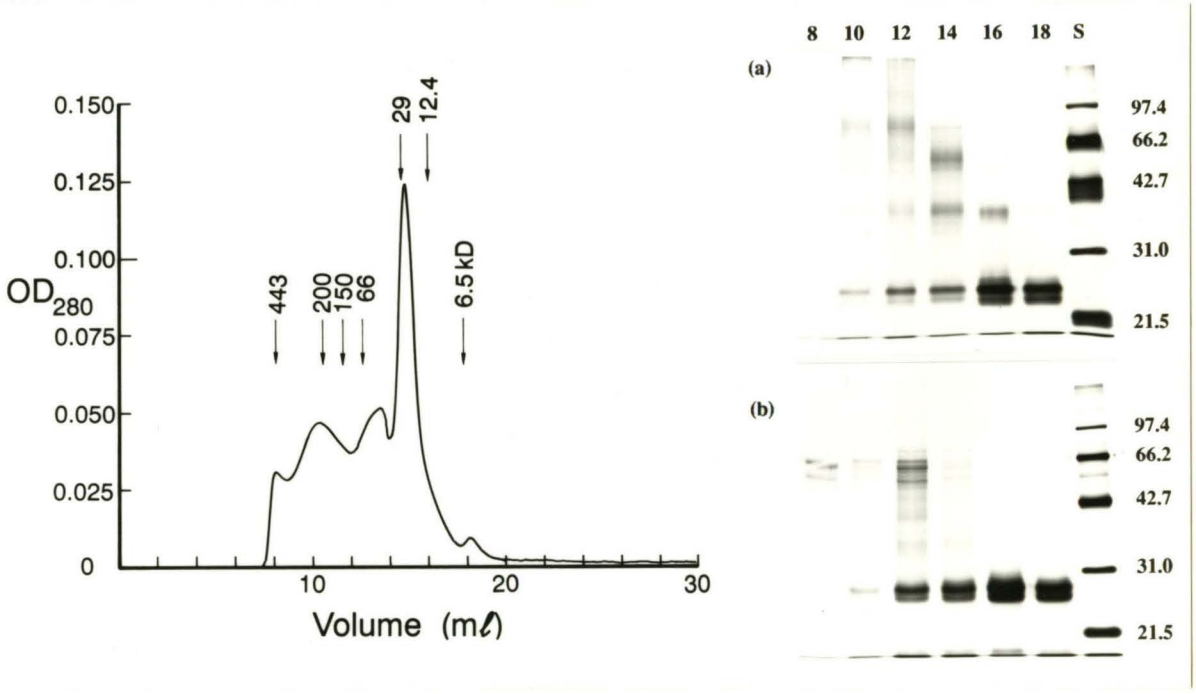

Fig. 6. Characterization of purified mV1V2 expressed in CHO cells. 100 µl of purified mV1V2 proteins (3 mg/ml) were applied on a FPLC Superose 12 column in 10 mM Hepes, 0.2 M NaCl, 1 mM EDTA, 0.04% NaN_3, pH 7.5, at a flow rate of 0.4 ml/min. Fractions (#8-#18) were subjected to 12% SDS-PAGE in either nonreducing (a) or reducing (b) conditions and were visualized by silver staining.

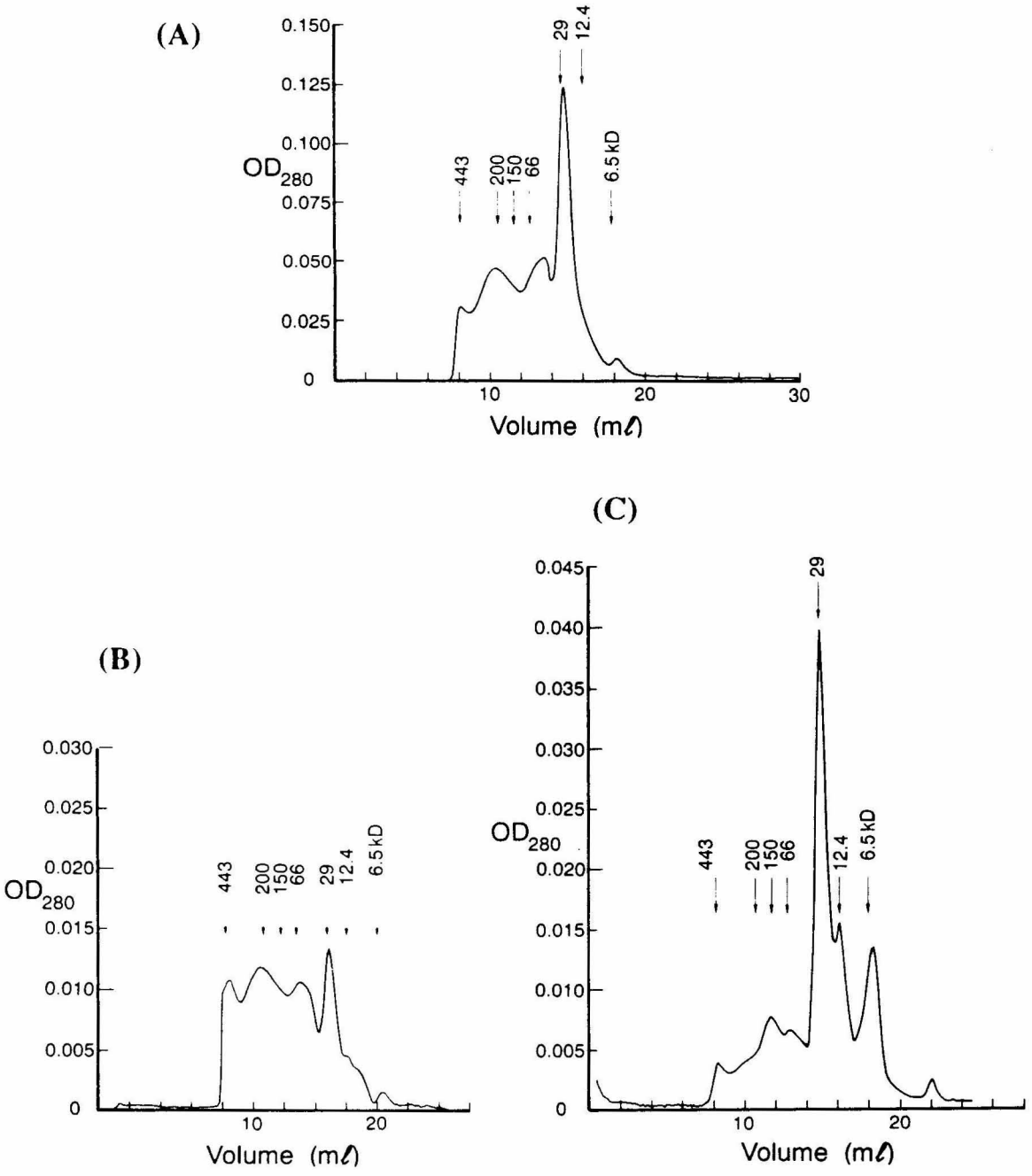

Fig. 7. Gel-filtration profile of repurified mV1V2. Fractions of the size exclusion chromatography (A) were collected into two pools, pool I (higher molecular weights) and pool II (most are monomers of mV1V2). Pool I was repurified, eluted, concentrated and was applied on anFPLC Superose 12 column (B). Pool II was concentrated and applied directly to the Superose 12 column (C).

Since human V1V2 has been successfully expressed in CHO cells (Arthos et al., 1988) without problems of oligomerization, it is surprising that the analogous murine CD4 fragment forms multimers. To check for possible point mutation(s) introduced during the PCR generation of the V1V2 coding sequence, mV1V2 in pBJ5.GS.mCD4 was sequenced completely by Ilana Tamis. The sequence corresponded exactly to the published sequence of murine CD4 (Littman and Gettner, 1987). Since a cysteine residue is found in the leader sequence, a protein with a free sulfhydryl would be generated if the signal sequence was not properly cleaved to generate the mature protein. To check for this possibility, purified mV1V2 was subjected to N-terminal sequencing (Caltech facility). The sequence obtained, KTLVLGKEGETAXL, exactly corresponded to the first 12 residues of murine CD4 (Littman and Gettner, 1987), thus ruling out the possibility due to an extra cysteine in the uncleaved signal peptide.

When comparing mV1V2 to the human V1V2, it is interesting to find that the human V1V2 is unglycosylated while mV1V2 contains a single N-linked glycosylation site. Surprisingly, we notice that the fourth cysteine (residue 162) is one residue C-terminal to the potential N-linked glycosylation site (residue 161 is Asn). Addition of N-linked glycosylation moieties may differ in CHO cells and murine T lymphocytes. It is possible that glycosylation might interfere with the formation of the disulfide bond. Both glycosylation of asparagine residues and the formation of disulfide bonds are cotranslational processes, occurring immediately after transport of the nascent polypeptide into the lumen of the rough endoplasmic reticulum (Rothman and Lodish, 1977; Glabe et al., 1980; Bergman and Kuehl, 1979). It therefore raises an interesting possibility as to whether such a mechanism results in misfolding for mV1V2. Future experiments will be attempted to examine this possibility using tunacamycin, an inhibitor for N-linked glycosylation. This will be important to design an efficient way to express mV1V2 for a structure determination, as well as important for understanding the cell biology.

References

Arthos, J., Deen, K. C., Chaikin, M. A., Fornwald, J. A., Sathe, G., Sattentau, Q. J., Clapham, P. R., Weiss, R. A., McDougal., J. S., Pietropaolo, C., Axel, R., Truneh, A., Maddon, R. J., and Sweet, R. W. (1989). Identification of the residues in human CD4 critical for the binding of HIV. Cell *57*, 469-481.

Ashkenazi, A., Presta, L. G., Marsters, S. A., Camerato, T. R., Rosenthal, K. A., Fendly, B. M., and Capon, D. J. (1990). Mapping the CD4 binding site for human immunodeficiency virus by alanine-scanning mutagenesis. Proc. Natl. Acad. Sci. USA *87*, 7150-7154.

Bergman, L. W., and Kuehl, W. M. (1979). Formation of an intrachain disulfide bond on nascent immunoglobulin light chains. J. Biol. Chem. *254*, 8868.

Brodsky, M. J., Warton, M., Myers, R. M., and Littman, D. R. (1990). Analysis of the site in CD4 that binds to the HIV envelope glycoprotein. J. Immunol. *144*, 3078-3086.

Byrn, R. A., Mordenti, J., Lucas, C., Smith, D., Marsters, S. A., Johnson, J. S., Cossum, P., Chamow, S. M., Wurm, F. M., Gregory, T., Groopman, J. E., and Capon, D. J. (1990). Biological properties of a CD4 immunoadhesin. Nature *344*, 667-670.

Capon, D. J., Chamow, S. M., Mordenti, J., Marsters, S. A., Gregory, T., Mitsuya, H., Byrn, R. A., Lucas, C., Wurm, F. M., Groopman, J. E., Broder, S., and Smith, D. H. (1989). Designing CD4 immunoadhesins for AIDS therapy. Nature *337*, 525-531.

Chamow, S. M., Peers, D. H., Byrn, R. A., Mulkerrin, M. G., Harris, R. J., Wang, W.-C., Bjorkman, P. J., Capon, D. J., and Ashkenazi, A. (1990). Enzymatic cleavage of a CD4 immunoadhesin generates crystallizable, biologically active Fd-like fragments.

Clark, S. J., Jefferies, W. A., Barclay, A. N., Gagnon, J., and Williams, A. F. (1987). Peptide and nucleotide sequences of rat CD4 (W3/25) antigen: evidence for derivation from a structure with four immunoglobulin-related domains. Proc. Natl. Acad. Sci. USA *84*, 1649-1653.

Clayton, L. K., Hussey, R. E., Steinbrich, R., Ramachandran, H., Husain, Y., and Reinherz, E. L. (1988). Substitution of murine for human CD4 residues identifies amino acids critical for HIV-gp120 binding. Nature *335*, 363-366.

Dalgleish, A. G., Beverley, P. C. L., Clapham, P. R., Crawford, D. H., Greaves, M. F., and Weiss, R. A. (1984). The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. Nature *312*, 763-767.

Davis, S. J., Brady, R. L., Barclay, A. N., Harlos, K., Dodson, G. G, and Williams, A. F. (1990). Crystallization of a soluble form of the rat T-cell surface glycoprotein CD4 complexed with Fab from the W3/25 monoclonal antibody. J. Mol. Biol. *213*, 7-10.

Fleury, S., Lamarre, D., Meloche, S., Ryu, S.-E., Cantin, C., Hendrickson, W. A., and Sekaly, R.-P. (1991). Mutational analysis of the interaction between CD4 and class III MHC: Class II antigens contact CD4 on a surface opposite the gp120-binding site. Cell *66*, 1037-1049.

Fung-Leung, W., Schilham, M. W., Rahemtulla, A., Kundig, T. M., Vollenweider, M., Potter, J., van Ewijk, W., and Mak, T. (1991). CD8 is needed for development of cytotoxic T cells but not helper T cells. Cell *65*, 443-449.

Garlick, R. L., Kirschner, R. J., Eckenrode, F. M., Tarpley, W. G., and Tomich, C.-S. C. (1990). *Escherichia coli* expression, purification, and biological activity of a truncated soluble CD4. AIDS Res. Hum. Retroviruses *6*, 465-479.

Gastinel, L. N., Simister, N. E., and Bjorkman, P. J. (1992). Expression and crystallization of a soluble and functional form of an Fc receptor related to class I histocompatibility molecules. Proc. Natl. Acad. Sci. USA. *89*, 638-642.

Glabe, C. G., Hanover, J. A., and Lennarz, W. J. (1980). Glycosylation of ovalbumin nascent chains. J. Biol. Chem. *255*, 9236-9242.

Healey, D., Dianda, L., Moore, J. P., McDougal, J. S., Moore, M. J., Estess, P., Buck, D., Kwong, P. D., Beverley, P. C. L., and Sattentau, Q. J. (1990). Novel anti-CD4 monoclonal antibodies separate human immunodeficiency virus infection and fusion of $CD4^+$ cells from virus binding. J. Exp. Med. *172*, 1233-1242.

Ibegbu, C. C., Kennedy, M. S., Maddon, P. J., Deen, K. C., Hicks, D., Sweet, R. W., and McDougal, J. S. (1989). Structural features of CD4 required for binding to HIV. J. Immunol. *142*, 2250-2256.

Klatzmann, D., Champagne, E., Chamaret, S., Gruest, J., Guetard, D., Hercend, T., Gluckman, J.-C., and Montagnier, L. (1984). T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV. Nature *312*, 767-768.

Kwong, P. D., Ryu, S. E., Hendrickson, W. A., and Axel, R. (1990). Molecular characteristics of recombinant human CD4 as deduced from phlymorphic crystals. Proc. Natl. Acad. Sci. USA *87*, 6423-6427.

Lamarre, D., Ashkenazi, A., Fleury, S., Smith, D. H., Sekaly, R. P., and Capon, D. J. (1989a). The MHC-binding and gp120-binding functions of CD4 are separable. Science *245*, 743-746.

Lamarre, D., Capon, D. J., Karp, D. R., Gregory, T., Long, E., and Sekaly, R. P. (1989b). Class II MHC molecules and the HIV gp120 envelope protein interact with functionally distinct regions of the CD4 molecule. EMBO J. *8*, 3271-3277.

Landau, N. R., Warton, M., and Littman, D. R. (1988). The envelope glycoprotein of the human immunodeficiency virus binds to the immunoglobulin-like domain of CD4. Nature *334*, 159-162.

Littman, D. R., and Gettner, S. N. (1987). Unusual intron in the immunoglobulin domain of the newly isolated murine CD4 (L3T4) gene. Nature *325*, 453-455.

Maddon, P. J., Littman, D. R., Godfrey, M., Maddon, D. E., Chess, L., and Axel, R. (1985). The isolation and nucleotide sequence of a cDNA encoding the T cell surface protein T4: a new member of the immunoglobulin gene family. Cell *42*, 93-104.

Maddon, P. J., Moleneaux, S. M., Maddon, D. E., Zimmerman, K. A., Godfrey, N., Alt, F. W., Chess, L., and Axel, R. (1987). Structure and expression of the human and mouse T4 genes. Proc. Natl. Acad. Sci. USA *84*, 9155-9159.

Maddon, P. J., McDougal, J. S., Clapham, P. R., Dalgleish, A. G., Jamal, S., Weiss, R. A., and Axel, R. (1988). HIV infection does not require endocytosis of its receptor, CD4. Cell *54*, 865-874.

Matthews, B. W. (1968). Solvent content of protein crystals. J. Mol. Biol. *33*, 491-497.

Peterson, A., and Seed, B. (1988). Genetic analysis of monoclonal antibody and HIV binding sites on the human lymphocyte antigen CD4. Cell *54*, 65-72.

Ramsdell, F., and Fowlkes, B. J. (1989). Engagement of CD4 and CD8 accessory molecules is required for T cell maturation. J. Immunol. *143*, 1467-1471.

Richardson, N. E., Brown, N. R., Hussey, R. E., Vaid, A., Matthews, T. J., Bolognesi, D. P., and Reinherz, E. L. (1988). Binding site for human immunodeficiency virus coat protein gp120 is located in the NH_2-terminal region of T4 (CD4) and requires the intact variable-region-like domain. Proc. Natl. Acad. Sci. USA *85*, 6102-6106.

Robey, E. A., and Axel, R. (1990). CD4: collaborator in immune recognition and HIV infection. Cell *60*, 697-700.

Robey, E. A., Fowlkes, B. J., Gordon, J. W., Kioussis, D., Von Boehmer, H., Ramsdell, R., and Axel, R. (1991). Thymic selection in CD8 transgenic mice supports an instructive model for commitment to a CD4 or CD8 lineage. Cell *64*, 99-107.

Rothman, J. E., and Lodish, H. F. (1977). Synchronized transmembrane insertion and glycosylation of a nascent membrane protein. Nature *269*, 775-780.

Ryu, S.-E., Kwong, P. D., Truneh, A., Porter, T. G., Arthos, J., Rosenberg, M., Dai, X., Xuong, N.-h., Axel, R., Sweet, R. W., and Hendrickson, W. A. (1990). Crystal structure of an HIV-binding recombinant fragment of human CD4. Nature *348*, 419-426.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., and Erlich, H. A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science *239*, 487.

Sattentau, Q. J., and Weiss, R. A. (1988). The CD4 antigen: physiological ligand and HIV receptor. Cell *52*, 631-633.

Thaller, C., Weaver, L. H., Eichele, G., Wilson, E., Karlsson, R., and Jansonius, J. N. (1981). Repeated seeding technique for growing large single crystals of proteins. J. Mol. Biol. *147*, 465-469.

Townsend, A. R. M., and Bodmer, H. (1989). Antigen recognition by class I-restricted T lymphocytes. Annu. Rev. Immunol. *7*, 601-624.

Veillette, A., Bookman, M. A., Horak, E. M., and Bolen, J. B. (1988). The CD4 and CD8 T cell surface antigens are associated with the internal membrane tyrosin-protein kinase $p56^{lck}$. Cell *55*, 301-308.

Wang, J., Yan, Y., Garrett, T. P. J., Liu, J., Rodgers, D. W., Garlick, R. L., Tarr, G. E., Husain, Y., Reinherz, E. L., and Harrison, S. C. (1990). Atomic structure of a fragment of human CD4 containing two immunoglobulin-like domains. Nature *348*, 411-419.

Williams, A. F., Davis, S. J., He, Q., and Barclay, A. N. (1989). Structural diversity in domains of the immunoglobulin superfamily. Cold Spring Harbor Symp. Quanti. Biol. Volume LIV, pp. 637-647.

Woldawer, A., and Hodgson, K. O. (1975). Crystallization and crystal data of monellin. Proc. Natl. Acad. Sci. USA *72*, 398-399.

Zuniga-Pflucker, J. C., Jones, L. A., Longo, D. L., and Kruisbeek, A. M. (1990). CD8 is required during positive selection of $CD4^-/CD8^-$ T cells. J. Exp. Med. *171*, 427-437.

APPENDIX B

ENZYMATIC CLEAVAGE OF A CD4 IMMUNOADHESIN GENERATES CRYSTALLIZABLE, BIOLOGICALLY ACTIVE Fd-LIKE FRAGMENTS