BIOCHEMICAL CHARACTERIZATION OF THE BINDING REGION OF *SACCHAROMYCES CEREVISIAE* ALPHA-AGGLUTININ A MEMBER OF THE IMMUNOGLOBULIN SUPERFAMILY

by

MINHAO CHEN

A dissertation submitted to the Graduate Faculty in Biochemistry in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York.

1995

This manuscript has been read and accepted for the Graduate Faculty in Biochemistry in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

April 25, 1995
Date

Chair of Examining Committee

April 25, 1995
Date

Executive Officer

Supervisory Committee

The City University of New York

ABSTRACT

BIOCHEMICAL CHARACTERIZATION OF THE BINDING REGION OF *SACCHAROMYCES CEREVISIAE* ALPHA-AGGLUTININ A MEMBER OF THE IMMUNOGLOBULIN SUPERFAMILY

by

Minhao Chen

Advisor: Professor Peter N. Lipke

α-agglutinin of *S.cerevisiae* is a glycosylated cell adhesion molecule that mediates cell adhesion during mating. The N-terminal 331 residues of α-agglutinin (α-agglutinin$_{20\text{-}351}$) is sufficient for binding to its ligand **a**-agglutinin. A region of α-agglutinin$_{20\text{-}351}$ with amino acid residues 220 to 300 showed significant homology to the consensus sequence of immunoglobulin variable-type (IgV) superfamily.

In this study, this previously described Ig-like region is expanded to residues 200 to 330 to accommodate all the potential ß strands and the intradomain disulfide of an IgV domain. Two other regions, with amino acid residues 20-104 and 105-199, show significant sequence homology to each other and have sequence motifs common to IgV superfamily members. Therefore, α-agglutinin$_{20\text{-}351}$ consists of 3 IgV domains that together make up a binding environment for a-agglutinin. These regions are domain I: residues 20-104; domain II: residues 105-199; and domain III: residues 200-333.

α-agglutinin$_{20\text{-}351}$ was overexpressed and purified. The deN-glycosylated α-

agglutinin$_{20\text{-}351}$ is a 45 kDa protein containing O-linked carbohydrates. Circular Dichroism spectroscopy showed a ß-sheet content of 70 % in α-agglutinin$_{20\text{-}351}$ and 55 % in its proteolytic fragment α-agglutinin$_{155\text{-}351}$. The high β-sheet content of α-agglutinin$_{20\text{-}351}$ confirms that the binding region of α-agglutinin consists of anti-parallel β-sheet domains such as Ig-folds.

Proteolytic mixture containing fragments of amino acids 20-154 and 155-351 of α-agglutinin$_{20\text{-}351}$, retaining the overall secondary structure of the binding region, shows no activity. Purified α-agglutinin$_{155\text{-}351}$ containing the entire domain III gives no activity, either. Therefore, the activity of α-agglutinin is also dependent on structures of the first two domains.

The arrangement of the disulfide bonds in α-agglutinin$_{20\text{-}351}$ was established. Cys^{97} and Cys^{114} form a disulfide bond between domains I and II. Cys^{202} and Cys^{300} is an atypical intradomain disulfide bond between the A and F strands of domain III, whereas Cys^{227} and Cys^{256} have free sulfhydryls. Reduction of the disulfides is accompanied by a decrease in the ß-sheet content and loss of the binding activity. Much of the primary structure of α-agglutinin$_{20\text{-}351}$ has been confirmed by peptide mapping, which revealed two N-glycosylation sites in domain III. In additional, a Ser and Thr rich region that cluster in the C-terminus of α-agglutinin$_{20\text{-}351}$ are highly O-glycosylated.

PREFACE

This thesis is organized to illustrate the biochemical structure of the binding region of α-agglutinin: Chapter I is the general introduction of this work; Chapter II presents the internal sequence homologies and the homologies to variable type immunoglobulin superfamily domains in the N-terminal half of α-agglutinin; Chapter III illustrates the disulfide structure and glycosylation sites in the binding region; Chapter IV describes the secondary structure of this region and a proteolytic fragment; Chapter V is a discussion of results from Chapters II, III and IV; and Chapter V lists references cited in the text.

ACKNOWLEDGEMENTS

I would like to express my appreciations to those who have contributions to this work:

To my wife, Amy, for her years-long understanding and durable support throughout my Ph.D. candidacy career.

To my parents, especially to my mother, who passed away while I was away and pursuing this thesis work; otherwise she would be the happiest one to see that I finally made it.

To Dr. Peter Lipke, my thesis advisor, for his guidance of the thesis work and for all the scientific training he offered.

To Dr. Janet Kurjan and Dr. Peter Kahn for their invaluable collaborations in this work.

To Mandy Zhengming Shen and Steve Bobin for their outstanding technique support for this work.

To all the members of my thesis committee: Drs. Marie Filbin, Thomas Haines, Joseph Krakow, and Rivka Rudner.

This work is also a gift to my 2 year-old little, sweet daughter, Erica.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

Arg-C, endoprotease Arg-C

CD, circular dichroism

Con A, Concanavalin A

DMF, N,N-dimethylformamide

DTT, dithiothreitol

endo H, endoglycosylase H

HPLC, high performance liquid chromatography

Ig, immunoglobulin

IgV, variable-type immunoglobulin domain

P-2007, N-(1-pyrenemethyl) iodoacetamide

PGK, phosphoglycerate kinase

TCEP, tris(2-carboxyethyl)phosphine hydrochloride

Tris-HCl, tris(hydroxymethyl)aminomethane hydrochloride

Chapter I

Introduction

Sexual agglutination in the life cycle of *S. cerevisiae*

The mating process

Cell types of *S. cerevisiae*. In the life cycle of *S.cerevisiae*, cells may exist as either haploid (a or α) or diploid. Both haploid and diploid cells reproduce by budding and give rise to daughter cells in vegetative growth. Diploid cells can also enter meiosis and undergo sporulation (Malone 1990; Bresch et al., 1968). Haploid cells of opposite mating types (a or α) can mate and form diploid a/α zygotes.

The mating process. Haploid cells of the opposite mating type form aggregates upon contact (Wickerham, 1958; Brock, 1958). Pairs of opposite mating type cells in sexual aggregates fuse to form diploid zygotes (reviewed in Cross, et al., 1988; Sprague, et al., 1983; Yangishima and Yoshida, 1981).

During the mating process, cells secrete mating type-specific pheromones (Duntze et al., 1970). a Cells produce and secrete a-factor; α cells produce and secrete α-factor. Both a- and α- factor are peptide pheromones (Stotzler and Duntze, 1976; Betz et al., 1987). Haploid cells express specific receptors for pheromone made by cells of opposite mating type (Burkholder and Hartwell, 1985; Hagen et al., 1986). Both a and α factor receptors are transmembrane proteins. They mediate response to the pheromone induction by activating signal transduction through a G protein conjugated pathway (reviewed in Kurjan, 1992; Marsh et al., 1991), which is common to both mating types (Bender and Sprague, 1986). Pheromone induction is essential for mating (Jackson, et al., 1991; Caplan and Kurjan, 1991). A number of sterile mutants fail to response pheromones. These mutants are incompetent in processing signal transduction and they cannot mate (Kurjan, 1992). The consequences of pheromone induction include arrest of cells in the

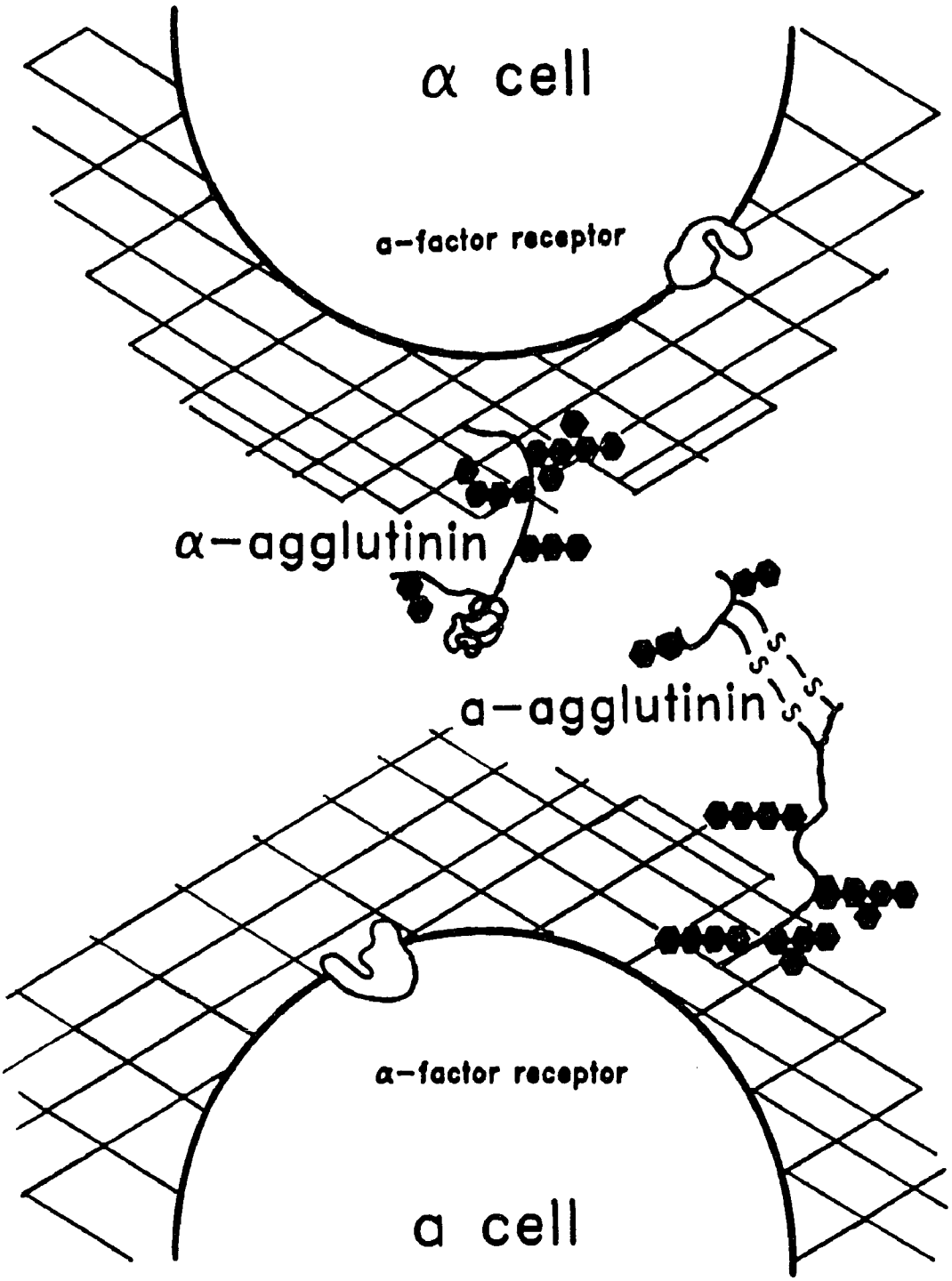
G1 phase of the cell cycle (Duntze et al., 1970; Chan and Otte, 1977, Moore, 1983), morphological changes of cell wall and cells (Lipke et al., 1976; Fujimura and Yanagishima, 1983), induction of agglutinability (Doi et al., 1979; Baffi et al., 1984; Terrance and Lipke, 1987), and expression of fusion-specific proteins and many other proteins as well (Trueheart et al., 1987). Fusion of a and α cell initiates at shmoo tips as a consequence of induction (Jackson and Hartwell, 1990a). After cells fuse to form the a/α zygote, the two nuclei fuse to form a a/α nucleus.

Sexual agglutination in mating

Sexual agglutination. Jackson and Hartwell showed that both mating types choose their sexual partner by responding to the strongest pheromone signal, and sexual courtship is an early step in mating (Jackson and Hartwell, 1990a and 1990b). The afterward sexual agglutination of opposite mating type haploids of *S. cerevisiae* and of other budding yeast species has been attributed to the specific cell surface glycoproteins expressed by each mating type (reviewed in Crandall et al., 1977; Yanagishima, 1984; Lipke and Kurjan, 1992). a Cells express a-agglutinin and α cells express α-agglutinin.

The complementary agglutinin-mediated sexual agglutination of a and α cells is shown (Fig.1). After pheromone induction, an increased level of a- and α-agglutinin is expressed on respective cells. Binding of α-agglutinin and a-agglutinin then promotes stable sexual aggregation. Interaction of complementary agglutinins allows direct cellular contact of opposite mating types (Moore, 1983; Terrance and Lipke, 1981), which is essential for fusion of a and α cells to form a/α zygotes (Lipke et al., 1989; Roy et al., 1991).

Figure 1. Diagram of sexual agglutinations in *S. cerevisiae*. Haploid mating type **a** and α cells form aggregates, mediated by the interaction of **a** and α-agglutinin on the respectively cells. In the diagram, cell walls are cross-hatched regions surrounding the semicircular plasma membrane, and consist of mainly glucans, mannoproteins, and chitins. Both **a** and α-agglutinins are extracellular glycoproteins, and hexagons represent carbohydrate moieties on the agglutinins. The binding subunit of **a**-agglutinin is attached to the matrix bonded core subunit through disulfide linkage. Diagram used with permission (Lipke and Kurjan, 1992).


α cell
a-factor receptor
α-agglutinin
a-agglutinin
S
S
S
S
a-factor receptor
a cell

Species-specific agglutination. The sexual agglutination during mating of other budding yeast species, including *Hansenula wingei*, *Pichia amethionina*, and *Saccharomyces kluyveri*, has been investigated. Complementary sexual agglutinins are also present in opposite mating type haploids of these species. Agglutination of opposite mating type cells is species specific (Yamaguchi et al., 1984). Agglutinins from one yeast species do not interact efficiently with those from other species.

Mating-type specificity and pheromone inducibility of agglutinins

Mating-type specific expression. The expression of **a**- and α-agglutinin is mating-type specific (Cross *et al.*, 1988; Yanagishima and Yoshida, 1981) and is under the control of mating-type locus (*MAT*). As other mating type specific products, regulation of mating-type specific expression of agglutinin is at the transcriptional level (Lipke et al., 1989). The *MATα1* gene product positively regulates expression of α-specific genes. Mutations in the *MATα1* locus result in defects in α-agglutinin production and agglutinability (Doi and Yoshimura, 1985). Both *MATα2* and *MATa1* gene products are repressors and together they negatively regulate expression of haploid-specific genes in **a**/α cells (Johnson and Herskowitz, 1985). *MATα2* alone negatively regulates **a**-agglutinin. The expression of the core subunit of **a**-agglutinin is haploid-specific but is not regulated by *MATα2* (Roy et al., 1991). Therefore, it is expressed in both **a** and α cells. The expression of the binding subunit of **a**-agglutinin, on the other hand, is **a** cell specific (de Nobel et al., 1995; Roy et al., 1991).

Pheromone inducible expression. The constitutive level of **a** and α-agglutinin is highly strain and species dependent. In *S. cerevisiae*, α-agglutinin is constitutively expressed in most α strains, whereas no constitutive level of **a**-agglutinin is detected in

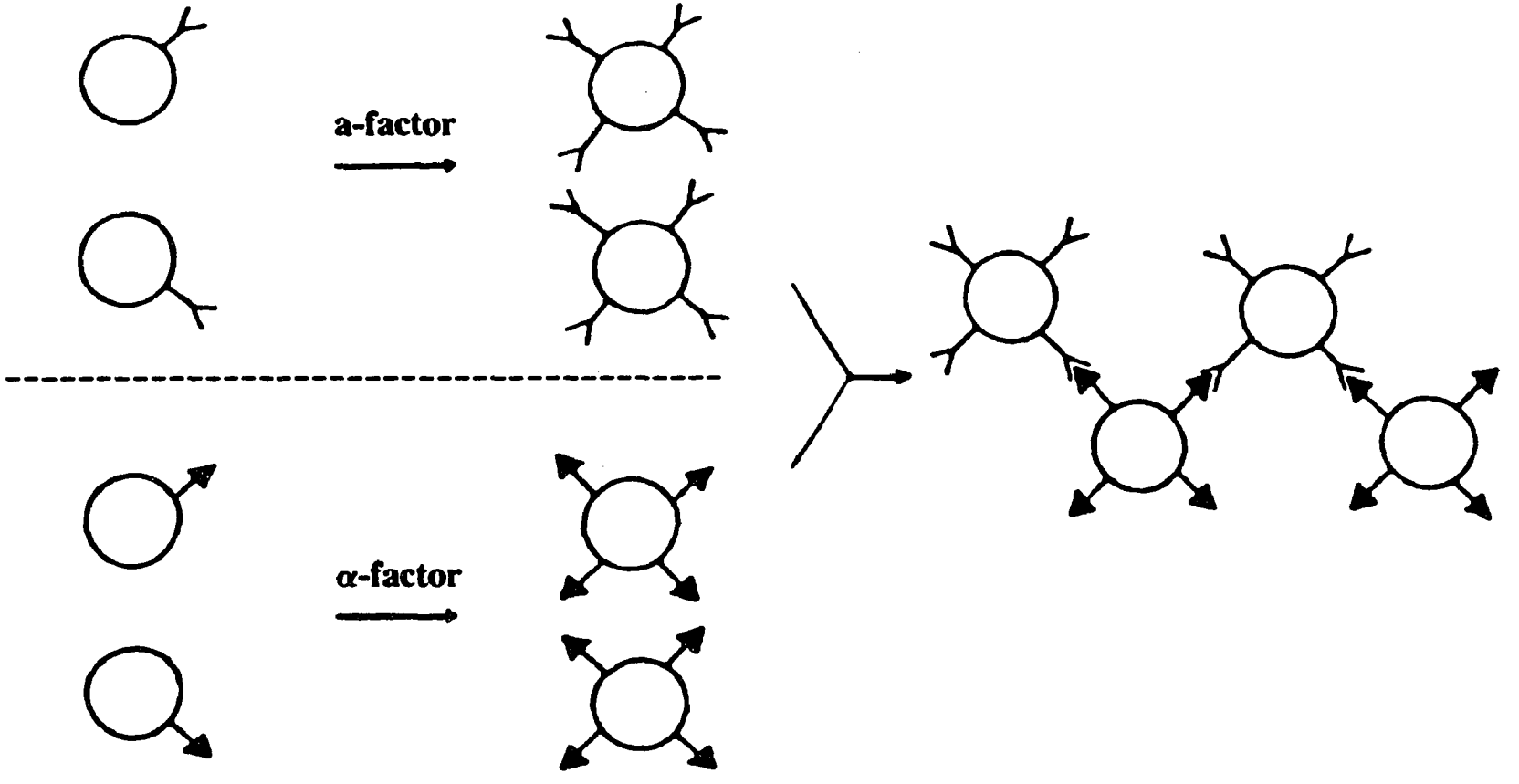
most **a** strains (Campbell, 1973; Crandall and Brock, 1968; Yanagishima and Yoshida, 1981; Manney and Meade, 1977). The expression of both agglutinins is pheromone-inducible in both inducible and constitutive strains of *S. cerevisiae* and also in other budding yeast species (Terrance and Lipke, 1981; reviewed in Yoshida et al., 1989). In *S. cerevisiae*, both constitutive and inducible forms of α-agglutinin are encoded by a single gene (Lipke et al., 1989). The induced expression of agglutinins is simultaneous with pheromone induced cell cycle arrest and cellular morphological change (Moore, 1983; Yanagishima and Fujimura, 1981). A 20-30 fold increased *AGα1* RNA levels and a 2-7 fold increased cell surface α-agglutinin level can be induced upon exposure of α cells to **a**-factor (Lipke et al., 1989; Wojciechowicz and Lipke, 1989). Increased expression level of **a**-agglutinin was also observed after exposure to α-factor (Cappellaro, et al., 1991). A diagram of pheromone induction of sexual agglutination process is shown (Fig.2). During mating, pheromones induce expression of cell surface **a**- and α-agglutinin. The interaction of complementary agglutinins mediates contact of **a** and α cells, and promotes sexual agglutination. Fusion between these sexual partners initiates among these sexual aggregates.

Role of agglutinins in mating

Sexual agglutinins are essential for mating under conditions that do not promote cell-cell contact (liquid medium). Mutations in the α-agglutinin structural gene, *AGα1*, or the structural genes of either **a**-agglutinin core subunit, *AGA1*, or **a**-agglutinin binding fragment structural gene, *AGA2*, resulted in deficient mating in liquid medium (Lipke et al., 1989, Roy et al., 1991, de Nobel et al., in press).

Figure 2. Pheromone induced sexual agglutination. Cellular agglutination is mediated by expression of agglutinins. *S. cerevisiae* expresses fewest agglutinin molecules per cell among four characterized yeast species. After exposure to pheromone (10^{-9} M) for 15 to 30 min, both mating types express $2X10^4$ to $5x10^4$ agglutinable sites per cell. Sexual aggregates are formed by the monovalent interaction of **a**- and α-agglutinin.

Basal Levels → Phermone Induction → Induced Levels → Sexual Agglutination



Pheromones induce an increased level of agglutinins, along with morphological change and cell cycle arrest of haploid cells. A high level of newly expressed agglutinins is observed on the shmoo tip of cells, which is the elongated cellular part induced by prolonged exposure to pheromone (Watzele et al., 1988; Wojciechowicz 1990). Increased levels of other mating and fusion related proteins, including pheromones, pheromone receptors, fusion protein and secreted membrane are also found at the shmoo tip (Trueheart, et al., 1987). Pheromone induction also occur at the tip. The interaction of complementary agglutinins may thereby enhance the efficiency of pheromone induction at least in the region of the tip. Lipke and Kurjan (1992) proposed that sexual agglutination mediates the mating process by 1. elevating the local concentrations of pheromone and pheromone receptors at the regions of cell-cell contact, and 2. enhancing close apposition of fusion proteins and secretory membrane at the shmoo tip, where fusion of a and α cells initiates.

Structure and binding features agglutinins

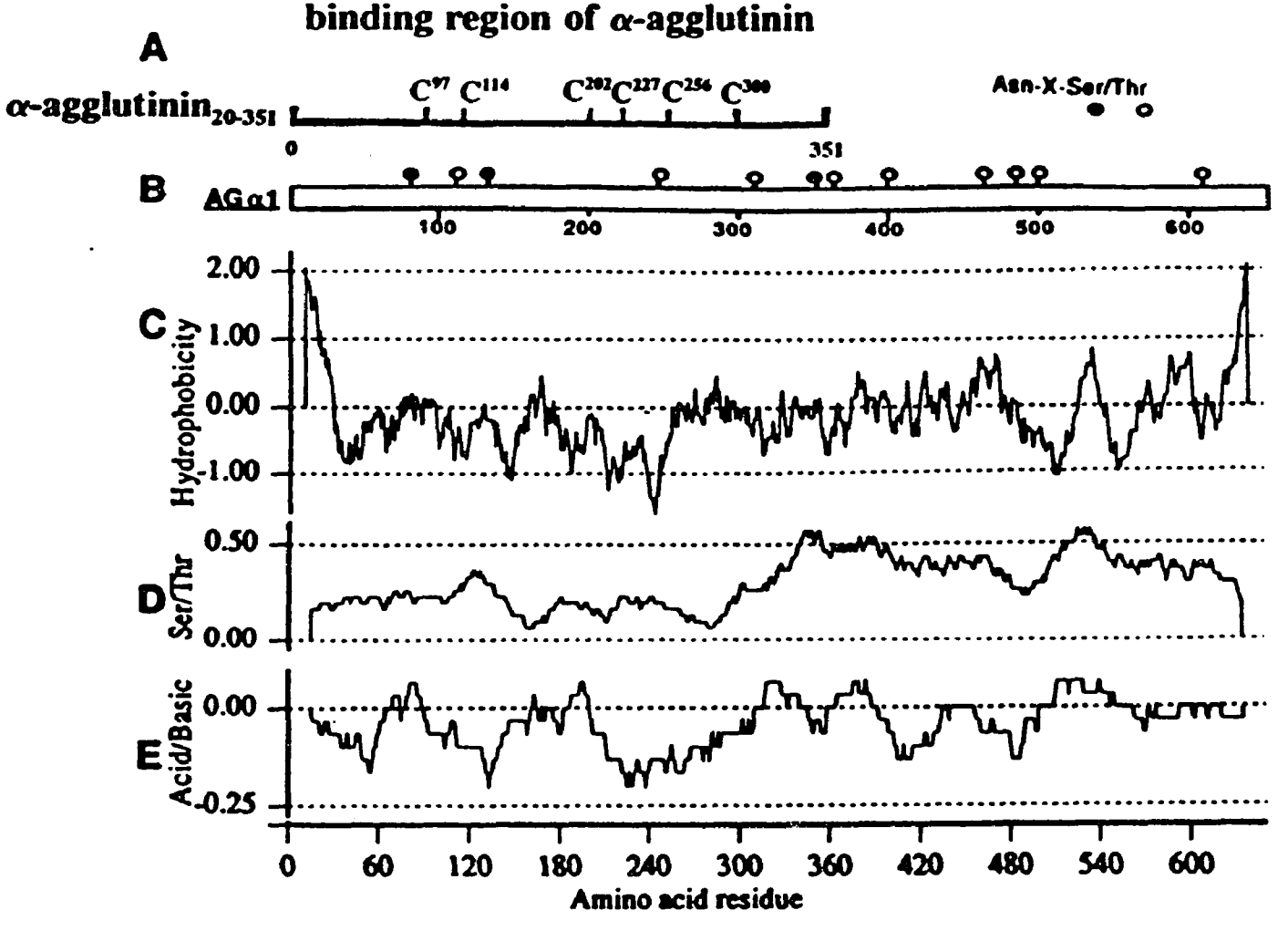
The structural genes of agglutinins

Genes for both α-agglutinin and a-agglutinin were cloned. α-agglutinin is a single polypeptide and is the product of *AGα1* gene. a-agglutinin, however, consists of core and binding subunits, each of which is the product of a separate gene, *AGA1* and *AGA2*, respectively.

α-agglutinin. *AGα1* encodes a polypeptide of 650 amino acids (Lipke et al., 1989 and Fig.3). It contains a hydrophobic N-terminal sequence of 19 amino acids, which is a signal for secretion (Hauser and Tanner, 1989). It also contains a C-terminal hydrophobic sequence of 20 amino acids, which is the signal for addition of a glycosylphosphatidyl-inositol (GPI) anchor involved in cell surface anchorage of the protein (Wojciechowicz, et al., 1993; Lu, et al., 1994 and 1995). The protein contains twelve cysteine residues that could form disulfide linkages. There are twelve potential N-linked glycosylation sites with sequence of Asn-Xaa-Ser/Thr. The protein has about one-third of its amino acid content as Ser or Thr, with the Ser and Thr rich region clustered in the C-terminal half of α-agglutinin. These residues are potential O-linked glycosylation sites.

a-agglutinin core fragment. The open reading frame of the a-agglutinin core fragment gene (*AGA1*) encodes a polypeptide with 725 amino acids, but does not contain the activity for binding to α-agglutinin (Roy et al., 1989). Like α-agglutinin, it contains hydrophobic sequences at both the N-terminal and C-terminal ends, which are indicative of secretion and GPI addition signal, respectively. The latter may also serve as signal for cell surface anchorage. The open reading frame consists of about 50% Ser and Thr

Figure 3. Structural feature of the α-agglutinin gene (*AGα1*) and its binding active region, α-agglutinin$_{1\text{-}351}$, (HAGα). (A) The binding active region, α-agglutinin$_{1\text{-}351}$, with position of cysteines indicated. (B) The *AGα1* open reading frame; potential N-glycosylation sites (Asn-X-Ser [o] and Asn-X-Thr [o]). (C) Hydrophobicity plot. (D) Frequency of Ser and Thr residues. (E) Acidicity plot. Plot (B) - (E) used with permission (Lipke et al., 1989).


binding region of α-agglutinin
A
α-agglutinin20-351
C97 C114 C202 C227 C256 C300
Asn-X-Ser/Thr
0
351
B
AGα1
100
200
300
400
500
600
C
Hydrophobicity
2.00
1.00
0.00
-1.00
D
Ser/Thr
0.50
0.00
E
Acid/Basic
0.00
-0.25
0
60
120
180
240
300
360
420
480
540
600
Amino acid residue

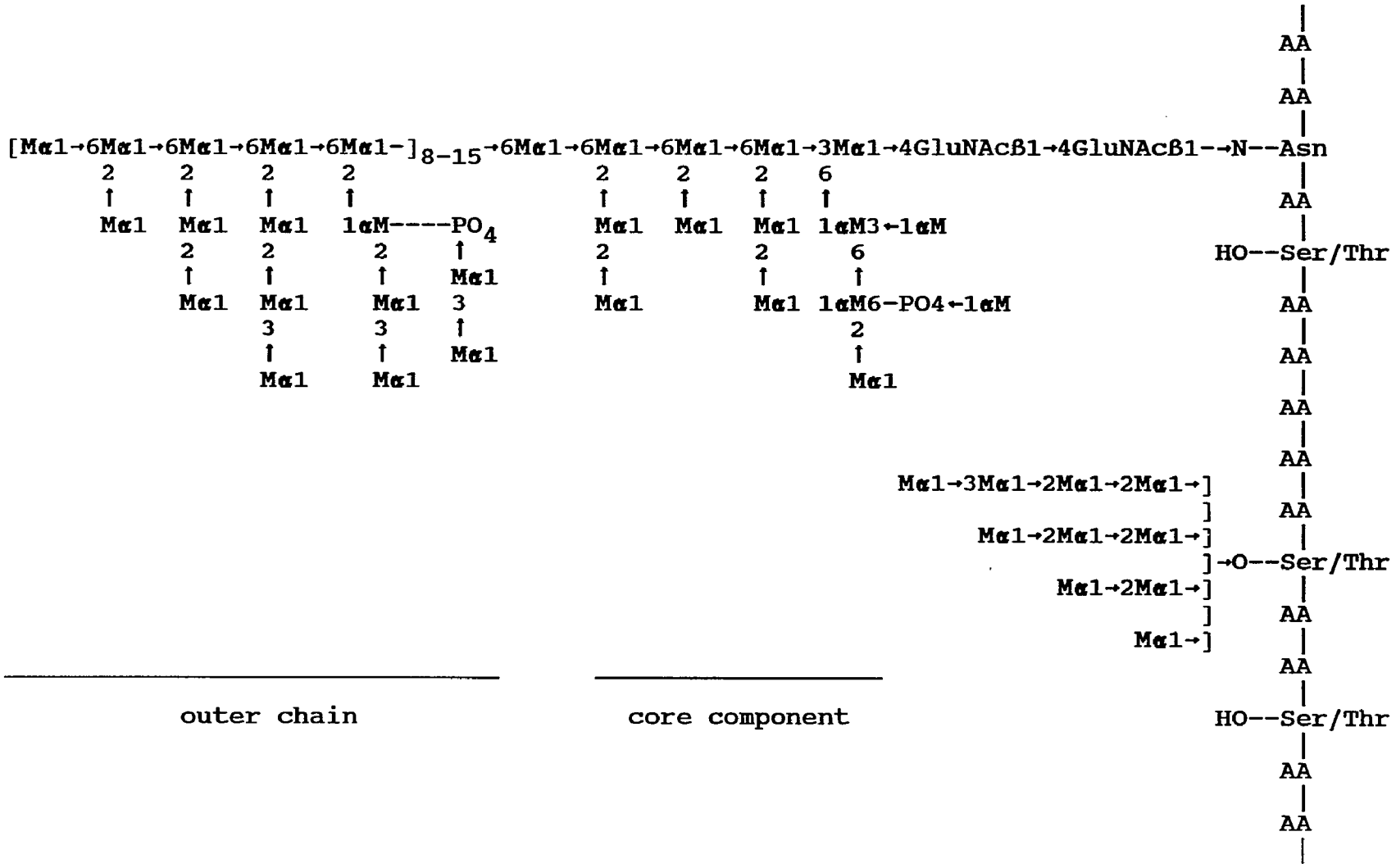
residues, but it does not contain potential N-glycosylation sites.

a-agglutinin binding fragment. Mutations in the *AGA1* gene result in secretion of active a-agglutinin binding subunit (Roy et al., 1989). The open reading frame of the a-agglutinin binding subunit gene (*AGA2*) encodes a small polypeptide with 87 amino acids including a N-terminal 18 amino acid secretion signal (Cappellaro, et al., 1991). It contains two cysteine residues, both of which are involved in disulfide binding to the core subunit of a-agglutinin (Cappellaro, 1994). It contains no potential N-glycosylation sites. The Ser and Thr residues account for about one-third of its total amino acid content. There are two Ser and Thr rich regions each at the N-terminal and C-terminal parts of the protein (Cappellaro et al., 1991 and 1994).

Agglutinins as extracellular glycoproteins

Protein glycosylation in *S. cerevisiae*. Both the asparagine-linked (N-linked) and serine/threonine-linked (O-linked) carbohydrates are high mannose structures in yeast (reviewed in Tanner and Lehle, 1987; Kukuruzinska et al., 1987). The structure of N- and O-linked oligosaccharides are shown (Fig.4). The N-linked carbohydrate consists of core oligosaccharides and a highly branched outer chain. The core component contains $Man_nGlcNAc_2$ structure as illustrated, where the number of mannosyl groups is from eight to about fifteen. The two glcNAc are in ß1-4 linkage whereas the attachment of the high mannose structure to the second glcNAc is in α1-4 linkage. The outer chain consists of 50-150 highly branched mannose residues, which are attached to the α1-6 linked backbone mannose by α1-2 linkage. The terminal mannose often has an α1-3 linkage. Some of side chains contain phosphate. The O-linked oligosaccharides are linear mannose chains, which are short and consist of 1-4 mannoses in α1-2 and α1-3 linkage.

Figure 4. Carbohydrate structures of N- and O-linked oligosaccharide chains on *S. cerevisiae* glycoproteins. The N-linked oligosaccharides consist of core $Man_nGlcNAc_2$ structure, where the number of mannosyl groups ranges from 8 to 15, and highly branched outer chain with 50 to 150 mannose residues. The branched mannosyl group of both core and outer chains may contain phosphates. The O-linked oligosaccharides, on the contrary, are short and linear mannose chains consisting of 1-4 mannose residues.


[Mα1→6Mα1→6Mα1→6Mα1→6Mα1-]8-15→6Mα1→6Mα1→6Mα1→6Mα1→3Mα1→4GluNAcβ1→4GluNAcβ1→N--Asn
Mα1→3Mα1→2Mα1→2Mα1→]
Mα1→2Mα1→2Mα1→]
Mα1→2Mα1→]
Mα1→]
1αM----PO4
1αM3←1αM
1αM6-PO4←1αM
]→O--Ser/Thr
HO--Ser/Thr
HO--Ser/Thr
AA
outer chain
core component

α-agglutinin. The mature α-agglutinin is a cell surface glycoprotein and can be released from extracellular matrix by treatment with Zymolase, a lytic enzyme mixture containing ß1→3 glucanase and proteases, or laminarinase (ß1→3 glucanase). α-agglutinin has a molecular weight of more than 240 kDa and contains both N-linked and O-linked oligosaccharides (Terrance et al., 1981; Hauser and Tanner, 1989; Wojciechowicz et al., 1993; Lu et al., 1994). Treatment with endo-glycosidase F or H reduces more than half of its apparent molecular weight. The N-linked oligosacchrides are not required for its activity, since deN-glycosylated α-agglutinin is active in binding to a-agglutinin (Hauser and Tanner, 1989; Sijmons et al., 1987; Terrance et al., 1987). The mannosyl O-glycosylations account for ≥ 60 kDa, based on size difference between deN-glycosylated form and deduced molecular weight from amino acid sequence (Lu et al., 1994).

Some α-agglutinin is bound to extracellular membrane via the GPI anchor and is releasable by treatment with phosphatidylinositol-specific phospholipase C (PI-PLC) (Wojciechowicz et al., 1993; Lu et al., 1994). α-agglutinin is also found to be covalently bound to cell wall matrix through carbohydrate moieties, and treatment with ß-glucanase can release wall anchored high molecular α-agglutinin (Hauser and Tanner, 1989; Wojciechowicz et al., 1993; Lu et al., 1994). Cell wall association of α-agglutinin is accompanied by further glycosylation including addition of ß1,6-glucan moiety (Lu et al., 1995). This glucosylation is dependent on pre-existence of a GPI anchor to α-agglutinin.

a-agglutinin core fragment. As expected, the a-agglutinin core fragment is highly heterogenous (> 200 kDa) (Hagiya et al., 1977; Wagner and Lipke, 1983; Shimoda and Yanagishima, 1975; Yamaguchi et al., 1984). It contains a high proportion of O-linked carbohydrate but no detectable N-linked carbohydrates, as reports for a-

agglutinin analogs from *H.wingei*, *P.amethionia* and *S.kluyveri* (Mendonca-Previato et al., 1982; Pierce and Ballou, 1983; Yen and Ballous, 1974), which is consistent with the fact that there is no potential N-linked glycosylation sites in *AGA1* gene (Roy et al., 1991).

a-agglutinin binding fragment. Purified binding subunit is a 22 kDa glycoprotein, whereas the polypeptide backbone of 69 amino acids accounts for only 7 kDa (Cappellaro et al., 1991). It contains solely O-glycosylations, which accounts for more than 60% of the apparent molecular weight (Watzele et al., 1988). Ten or more Ser and Thr residues are O-glycosylated (Cappellaro et al., 1991). The activity of highly glycosylated a-agglutinin is heat stable (Wagner and Lipke, 1983). deO-glycosylated a-agglutinins from various treatments, including α-mannosidase, mild reaction with HF and ß-elimination, can form binding complex with α-agglutinin. The binding complex, however, is 5-times less stable than that formed by glycosylated a-agglutinin (Cappellaro et al., 1991). Synthetic peptide containing the binding active site of a-agglutinin is also 5-times less active than the glycosylated one from proteolytic digestion of the protein (Cappellaro et al., 1994). Therefore, the oligosaccharide moiety on a-agglutinin or its binding site is likely to have some contribution to the formation and stability of the a- and α-agglutinin binding complex (Cappellaro et al., 1994).

As described in other yeast species, *S. cerevisiae* a-agglutinin binding subunit is releasable by DTT treatment of α-factor induced a cells (Orlean et al., 1986). Site specific mutagenesis revealed that the binding subunit is attached to the cell surface anchored core subunit via two disulfide bridges (Cys^{7} and Cys^{50}, from the N-terminus of

matured protein) (Cappellaro et al. 1994).

The adhesive binding of agglutinins

The binding feature of α- and a-agglutinin. The complementary interaction of **a**- and α-agglutinin is monovalent (Crandall et al., 1974). **a**-Agglutinin and α-agglutinin forms 1:1 *in vitro* binding complex (Cappellaro, et al., 1991). The strength of cellular agglutination is dependent on both binding constant and number of agglutinin molecule per cell. *S. cerevisiae* agglutinins have relative large binding constant (10^8-10^9) and few binding site among the four characterized yeast species. The basal expression level varies from undetectable to 10^4 molecules per cell for **a**-agglutinin and 10^3 to 10^4 molecules per cell for α-agglutinin (Terrance and Lipke, 1987; Watzele et al., 1988). After exposure to pheromone, both mating types of *S. cerevisiae* express $2X10^4$ to $5x10^4$ agglutinable sites per cell (Terrance and Lipke, 1987; Watzele et al., 1988; Wojciechowicz and Lipke, 1989). An increased binding strength between *S. cerevisiae* **a** and α cells by pheromone induction is, thereby, strongly dependent on the number of binding site on each cell (Terrance and Lipke, 1987).

The interaction of **a**- and α-agglutinin may be ionic, and can be inhibited by ionic detergent such as SDS and high salt but not by nonionic detergents, like Tween-20 or Triton X-100 (Terrance and Lipke, 1981). The agglutinability of α-agglutinin is pH dependent (Terrance and Lipke, 1987). A pH range between 4 to 6 is optimal for binding. By histidine-specific labelling, Cappellaro et al. (1991) identified His^{292} on α-agglutinin as critical for binding to **a**-agglutinin. It is likely that charged residues are involved in agglutinins binding.

Kinetic studies on association of ^{125}I-labeled α-agglutinin to **a** cells illustrated that

binding of complementary agglutinins is in a complicated manner, and can be categorized into two steps: a quick but weak association ($< 10^8$ M^{-1}) followed by a slow but high affinity binding ($K_{A2} = 10^9$ M^{-1}) (Lipke et al., 1987). Lipke et al. (1987) proposed that the latter step of binding requires change of protein conformation, since it is cold sensitive.

The binding regions of α-agglutinin and a-agglutinin. A region of α-agglutinin from amino acids 128 to 356 was recognized to contain the binding site by antibody (Lipke et al., 1989). A truncated α-agglutinin containing the N-terminal 351 amino acids retains full binding activity as assayed by inhibition of agglutination (Wojciechowicz, et al., 1993), whereas truncation at amino acid 278 of α-agglutinin results in inactivation. Within the N-terminal binding region, His^{292} is essential for the binding activity. A segment around this His is proposed to interact with a-agglutinin (Cappellaro et al., 1991).

In a-agglutinin, the binding subunit consists of only 69 amino acids. A proteolytic peptide containing the C-terminal 10 amino acids and one or two O-glycosylated sites, with sequence of Gly-Ser-Pro-Ile-Asn-Thr-Gln-Tyr-Val-Phe, is capable of inhibiting agglutination. A 5-fold higher concentration (100 nM) is required to achieve a same inhibitory specific activity of the binding subunit (16 nM) (Cappellaro et al., 1994). The activity of this 10 amino acid peptide implies that much of the binding subunit is used for its attachment to the core subunit via disulfide linkages. In addition, this O-glycosylated proteolytic peptide at Ser^{60} and/or Thr^{64} is 4-5 time more active than the synthetic nonglycosylated peptide, which implies that glycosylation at either or both sites might somehow modulate the binding activity of a-agglutinin (Cappellaro et al., 1994).

The immunoglobulin-superfamily of cell adhesion molecules

The Ig-superfamily and its subsets

Immunoglobulin-like binding domains are a common feature of many cell adhesion molecules of animals. Molecules of this superfamily have a diversity of function (reviewed in Williams and Barclay, 1988). This superfamily includes not only groups key in immunity but also in mediation of cell surface recognition of various tissues. Based on a brief category of the functions, this superfamily consists of various immunoglobulins and their receptors, T lymphocyte adhesion molecules and T lymphocyte subset markers, MHC antigens and T cell receptors, brain/lymphoid antigens and neural-associated molecules (reviewed in Williams and Barclay, 1988). It also includes growth factor receptors, and tumor antigens, etc. In most cases, these molecules play a recognition role at the cell surface.

Sub-categories of the superfamily. Conserved sequence patterns are seen among the Ig-superfamily domains. These domains can be categorized into three subsets based on amino acid sequence: the V, C1, and C2 types (reviewed in Williams and Barclay, 1988; Williams, et al., 1989).

Basically, the length of amino acid sequence in an IgV domain is somewhat longer than in an IgC domain. A typical IgC domain consists of 70-80 amino acid residues whereas a typical IgV domain consists of 80 or more residues. Diagrams of all of Ig-superfamily molecules mentioned in the text, containing either or both IgV and IgC domains can be found in Williams and Barclay, 1988, and Dwek et al., 1993.

Ig-molecules containing the V type Ig-like domain include the variable domains of immunoglobulins (Amzel and Poljak, 1979; Edelman, 1970), domains I-IV of poly Ig

receptor (Mostov et al., 1984), domain I of CD2 and MAG (Jones et al., 1992; Lai et al., 1987), domains I and III of CD4 (Brady et al., 1993; Ryu et al, 1990; Wang et al., 1990), and the single domains of Thy-1 and CD8 (Williams and Gagnon, 1982; Lemke and Axel, 1985; Leahy et al., 1992).

Ig-molecules containing the C1 type Ig-like domain include the constant domains of immunoglobulins (Amzel and Poljak, 1979), α1 domain of the class I MHC molecules and domains α1 and α2 of the class II MHC (Lew et al., 1986; Kaufman et al., 1984; Bjorkman et al., 1987). The C2 subset molecules include domain II of CD2, domains II and IV of CD4, domains II-V of MAG, all the five domains of N-CAM, and others (Williams, 1987; Williams and Barclay, 1988).

Sequence and structure conservation

Sequence conservation. Both V- and C-type Ig domains share several conserved features. Conservations in the B and F strands of both V and C2 type Ig domains are the most distinct. A consensus sequence pattern of Val/Leu/Ile-Xaa-Val/Leu/Ile-Xaa-Cys-Ser/Thr-Xaa-Ser/Thr/Val is present in B strands, and a consensus sequence pattern of Asp-Xaa-Gly-Xaa-Tyr/Phe-Xaa-Cys-Xaa-Val/Ala in F strands, respectively (Williams and Barclay, 1988).

Other consensus residues and conserved substitutions in a typical IgV domain include a Trp lying about 11-16 amino acids downstream of the first consensus cysteine. An Arg/Lys residue lying about 23 amino acids upstream of the second consensus cysteine is also highly conserved among IgV domains (Williams and Barclay, 1988; Gardinier, et al., 1992).

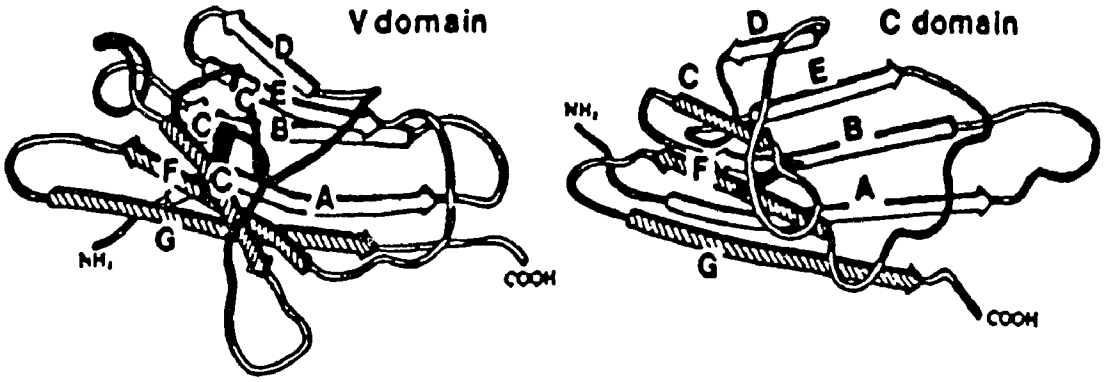
It has been noticed that although consensus sequence patterns are conserved in

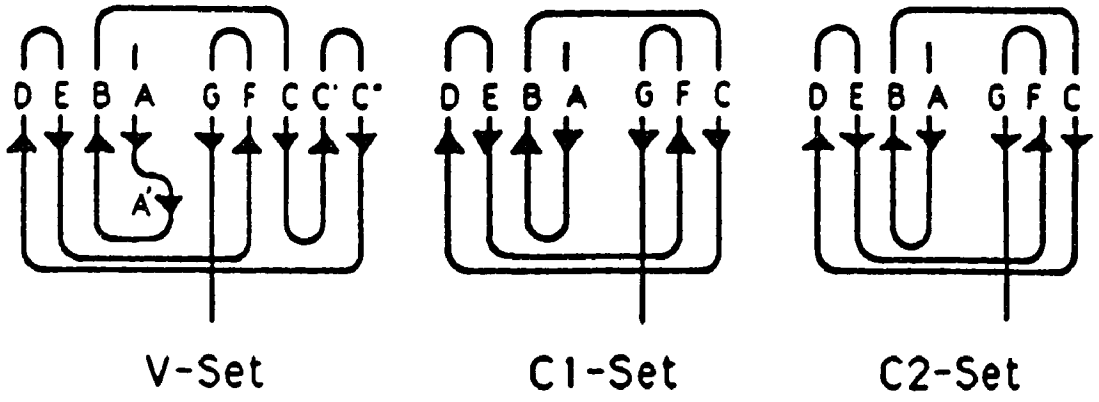
many of Ig-superfamily domains, no residue is invariant in all Ig-related domains. Even the conserved disulfide bonds are no longer the hallmark of Ig-like domains. IgV domains lacking the intradomain disulfide bond includes domain I of CD2 (Jones et al., 1992), LFA-3 (Seed, 1987), and CEA (Oikawa, et al., 1987), and domain IV of PDGFR (Yarden et al., 1986; Williams and Barclay, 1988).

Ig domain structure. The immunoglobulin superfamily molecules share a conserved structure, which is formed by two anti-parallel ß-sheets composed of several ß-strands. The variable (V) type Ig-folding structure consists of nine, while the constant (C1 and C2) type Ig-folding structure consist of seven ß strands. The schematic patterns of known V and C type Ig domain folds are shown according to Edmundson et al., 1975 (Fig.5, upper part). The schematic pattern of the three known Ig domain subtypes referenced from Amzel and Poljak (1979) also are shown (Fig.5, bottom part).

Strands of the two anti-parallel ß-sheets fold are named in a sequential order. The core of both V and C types Ig-fold consists of ß-strands A,B,E in one sheet and G,F,C in the other. These strands come from both N- and C-terminal parts of the domain sequence, which are common to both types but have systematic differences in length. Considerable sequence variation occurs in the middle of both C and V type domains, especially in sequence length. The C type Ig-like domains have 4 (A,B,E,D) and 3 (G,F,C) ß-strands in each of the anti-parallel ß-sheets. The V type domains have extra amino acids which form the C' and C" strands. Therefore, they consist of four ß-strands (A,B,E,D versus G,F,C,C') in each ß-sheet plus a short ß-strand segment of C" strand that crosses between sheets.

Figure 5. Diagrams of the folding patterns of IgV and IgC domains. The V, C1 and C2 domain folds as determined by Edmundson et al. (1975) are shown. In the lower part, the various domain types are shown in the schematic form as used by Amzel and Poljak (1979).


V domain
C domain
D
E
C
B
A
F
G
NH,
COOH


D E B A G F C C' C"
A'
D E B A G F C
D E B A G F C
V-Set
C1-Set
C2-Set

The overall structure and sequence pattern of C2 and C1 type is similar, and the division of the types is based on variation in conserved sequences. The C2 types have a slightly shorter strand D, and the turns linking the E and F strands are like those in variable domains (Jones et al., 1992; Wang et al., 1990; Ryu et al., 1990; Brady et al., 1993).

Each ß-strand of the two anti-parallel ß-sheets consists of 5-10 amino acids. Alternating hydrophobic amino acids in each strands form a hydrophobic core structure between the two anti-parallel sheets. The interaction between the two anti-parallel sheets can be further stabilized by a conserved intradomain disulfide bond. Across all the sequence of Ig-superfamily domains, identities or conservative amino acid substitutions are present in ß-strands B, C, E, and F; in particular, there is a characteristic pattern of alternating hydrophobic residues.

Criteria for inclusion of molecules in the Ig-superfamily. For the inclusion of molecules in the Ig-superfamily, the first criterion is the presence of a domain-sized sequence with significant similarity to Ig-superfamily consensus, including the presence of alternating hydrophobic amino acid residues in the ß-strands. In addition, there should be the probability that the sequence shares anti-parallel ß-sheet structure of an Ig-fold.

The IgV domain identities in α-agglutinin

Regions with significant identity to IgV domain. Within the binding region of α-agglutinin, the predicted amino acid sequence of *AGα1* between residue 220 and 300 shows homology to the immunoglobulin V-type fold (Wojciechowicz et al., 1993). The sequence and predicted ß-strand profile alignments of this domain against members of IgV domains is also shown (Fig.6). This region contains 12 identical amino acids and 2

conservative substitutions in comparision to the 19 conserved residues of the V-type consensus sequence. A segment with amino acid sequence of Ile^{198}-Asp^{199}-Leu^{200}-Asp^{201}-Cys^{202}-Ser^{203}-Ser^{204}-Val^{205} in α-agglutinin matches well the conserved sequence pattern of Val/Leu/Ile-Xaa-Val/Leu/Ile-Xaa-Cys-Ser/Thr-Xaa-Ser/Thr/Val in the predicted B strand of IgV domains. Tyr^{298} is present at position of a highly conserved Tyr/Phe residue in the F strand prior to the second conserved cysteine. Trp^{217} matches another highly conserved alignment lying 16 residues downstream from the first Cys. These homologies to the V-type Ig-consensus are statistically significant ($P < 10^{-7}$). The degree of the similarity to the consensus sequence is comparable to Ig fold domains of CD4, CD8, Thy-1, Po, and the CAM family, all of which are cell adhesion proteins.

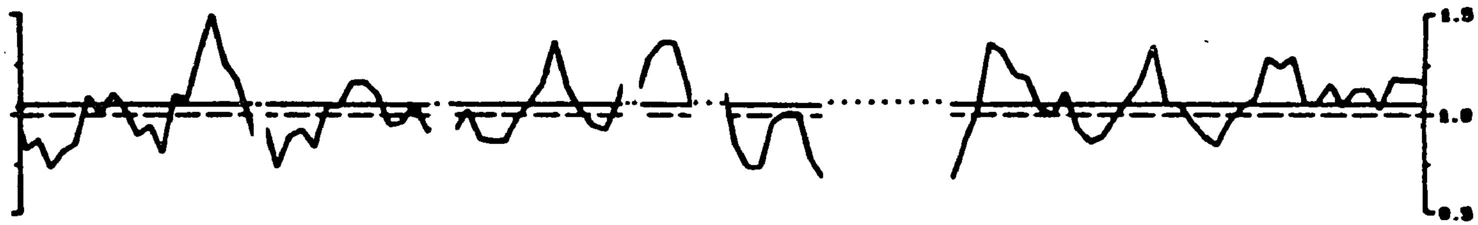
This region also shows strong ß-strand potential similar to the V-type Ig-fold by Chou-Fasman analysis, which can be categorized into 9 predicted ß-strands of potential anti-parallel ß-sheets structure as an Ig fold (Fig.6).

IgV domain identity in other regions of α-agglutinin. Members of Ig-superfamily usually have multiple Ig-fold domains within a single polypeptide. Amino acid sequence and ß-strand potential analyses of the *AGα1* sequence reveal two more potential Ig-domains in the N-terminal half of α-agglutinin binding region, at residues 20-104 and 105-199 (Wojciechowicz et al., 1993). Several conserved residues or conservative substitutions are seen in both regions. These observed homologies to the V-type consensus sequence, however, have not yielded a statistical significance.

Figure 6. Alignment of α-agglutinin (217-308) to members of the V-type immunoglobulin superfamily. The individual ß strands of the immunoglobulin fold are named and overlined above the alignment. The A and G strands of the alignment are omitted. This region of α-agglutinin sequence was aligned to the consensus sequence by the GCG Bestfit program. The Chou-Fasman ß-sheet potential of the region of α-agglutinin is shown underneath the sequence. Alignment used with permission (Wojciechowicz et al., 1993).

```
                   ___B___       ___C___     ___C'___    _C"_    ___D___     ___E___              ___F___
CD4 Domain      gkkGdtVeLtCtasq....kksiqfhWknsnqi..kilgnqgsFltk..gpsk...LndRadsrrslwdq..gnfpLiIkn..........lkieDsdtYiCevedqkee
CD8 Domain I    aeiGqeVkLtCevlrd....tsqgcsWlfrnssselllqptfiiYvss..srsk...LndildpnlfsarkennkyilLtlsk..........fstknqgyYfCsitsnsvm
Ig V-Heavy      vrpsqtlsLtCtvsgstf..sndyytWvrqp...pgrglewigYvfyhgtsddttpLrsRvtmlvdts...knqfsLrlss..........vtaaDtavYyCarnliagc
Ig V-Light      tspGetVtLtCrsstgav.ttsnyanWvqqkp..dhlftgliggtnnr.apgv...vpaRfsgslign.....kaaLtItg..........aqteDeaiYfCalwysnhw
MRC OX2         kllhttasLrCslkttq...eplivtWqkkka...vgpenmvtYskah.gvviqptYkdRinitelgl....lntsitfwn..........ttldDegcYmClfnmfgsg
Po              gavGsqVtLhCsfwssewvsddisftWryqpe.ggrdaisifhYakgqpyidevgtFkeRiqwvgdps...wkdgsivIh..........nldysDngtFtCdvknppdi
Poly Ig Rec     vleGdsVsitCyypttsvtr.hsrkfWcreeesgrcvtlastgYt....sqe....YsgRgkltdfpd...kgefvvtVdq..........ltqnDsgsYkCgvgvngrg
T Cell Rec α    vpeGartsLnCtfsds....asqyfwWyrqhs.gkapkalmsiFsn...geke....egRftihlnka...slhfsLhIrds..........qpsDsalYlCavtlyggs
T Cell Rec β    temGqeVtLrCkpisg.....hnslfWyrqt...mmrglellIYfnnnvpiddsgmpedRfsakmp....nasfstLkIqp..........seprDsavYfCassfstcs
Thy-1           clvnqnlrLdCrhennt.lpiqhefsltre....kkkhvlsgtLgv...peht...YrsRvnlfsdr.....fikvLtlanc.........fttkDegdYmCelrvsgqn

   Consensus    ---G--V-LtC---------------W--q----------Ø--Y------------Ø--RØ---------------L-I---------------D-g-Y-C-Ø------
                   |  | |               |  |          |  |            |  :|               | :                    | |
α-Agglutinin    dssnmnVdLdCssvqvys.sndfndwWfpqsy..ndtnadvtcFgsn..lwit...ldeklydg..........emLwVnalqslpanvntidhalefqYtCldtiantt
                217                                                                                                     308
```



Goal of this thesis work

The presence of conserved amino acid sequence motifs and predicted ß-strand patterns between the binding region of α-agglutinin and the V-type immunoglobulin superfamily suggested that it could be a new member of this superfamily. Members of the Ig-superfamily often consist of internally homologous multiple domains in tandem. Potential IgV domain identity in additional two regions suggests the binding region of α-agglutinin consist of 3 IgV domains. The goal of this thesis work is to elucidate if the binding region of α-agglutinin and its 3 domains show immunoglobulin-like conservations in secondary structure, disulfide bonding and glycosylation.

The possibility of IgV domain identity in the N-terminal 180 amino acids was analyzed by sequence alignments (Chapter II).

The disulfide structure and glycosylation sites in the binding region were mapped (Chapter III). Peptide mapping revealed a disulfide bridge between domain I and II, and an intradomain disulfide bond in domain III. The disulfide mapping data is consistent with the known variation of IgV domain superfamily. The identified disulfide structure and glycosylation sites should be useful in building up a three-dimensional model of this Ig-superfamily protein.

The secondary structure of the binding region and a proteolytic fragment are described in Chapter IV. Circular dichroism showed that the secondary structure of the binding region and its domains consist of mostly anti-parallel ß-sheets, consistent with the proposed Ig-fold structure in all the three domains. Adhesion studies showed that binding of α-agglutinin to a-agglutinin involves more than one domain.

Chapter II

Internal and IgV domain homologies in the N-terminal half of α-agglutinin

Introduction

IgV domains consist of 9 β strands, named A, B, C, C', C", D, E, F, and G, having strongly conserved residues in the B, C, D, and F strands (Williams and Barclay, 1988). Amino acid residues 220-300 of α-agglutinin show highly significant similarity to a consensus sequence for IgV domains, especially in the B, C, and F strands (Wojciechowicz et al., 1993). This entire IgV domain would include residues 200-327 to accommodate all 9 ß-strands. This domain will be called domain III in light of two more potential IgV domains in the N-terminal half of α-agglutinin, at residues 20 to 104 and 105 to 199 (Wojciechowicz et al., 1993; Chen et al., manuscript in preparation).

The N-terminal half of α-agglutinin contains the entire binding site for its ligand. Mutation, deletion and immunochemical evidence all suggested that the Ig-like region between amino acids 200 and 330 directly participates in binding (Lipke, et al., 1989; Cappellaro, et al., 1991). This domain contains His^{292}, which is essential for binding. However, a proteolytic fragment carrying the entire domain was inactive (Chapter IV). This implies that this single Ig-like domain is not sufficient for agglutination activity and there must be some essential residues that lie outside of this region in α-agglutinin$_{20\text{-}351}$.

The structure of the N-terminal region comprising residues 20 to 199 of matured protein, therefore, becomes important. Multiple sequential Ig domains that start from the N-terminal of the mature proteins are very common in the Ig superfamily (Barclay et al., 1990). A typical IgV domain consists of about 100 amino acids (Williams et al, 1989). The N-terminal 180 residues of α-agglutinin are enough to form two more IgV domains (Williams and Barclay, 1988; Williams et al., 1989; Ryu et al., 1990; Wang et al., 1990; Barclay et al., 1993). A preliminary analysis of amino acid sequence and Chou-Fasman

ß-strand potentials revealed that this region could contain two possible Ig domains (Wojciechowicz et al., 1993). No significant homology was found for these regions to other known families of cell adhesion molecules, e.g., Ig Constant domains, Fibronectin type III domains, Ca^{2+}-dependent cell adhesion domains, etc.

Members of Ig-superfamily often consist of multiple Ig-like domains that together mediate cell adhesion. PSG-CAM has three tandem IgV domains (Chen et al., 1992). IgA and IgM receptor (IgR) consists of four consecutive IgV domains from the N-terminus and one IgC2 domain at the C-terminus of the extracellular region (Mostov et al., 1984; Frutiger et al., 1986). In this chapter, I propose that the N-terminal half of α-agglutinin containing the binding region consists of 3 Ig-like domains. Internal sequence homologies of these regions and homology to the IgV domain consensus are examined.

OUTCOMES OF ALIGNMENTS

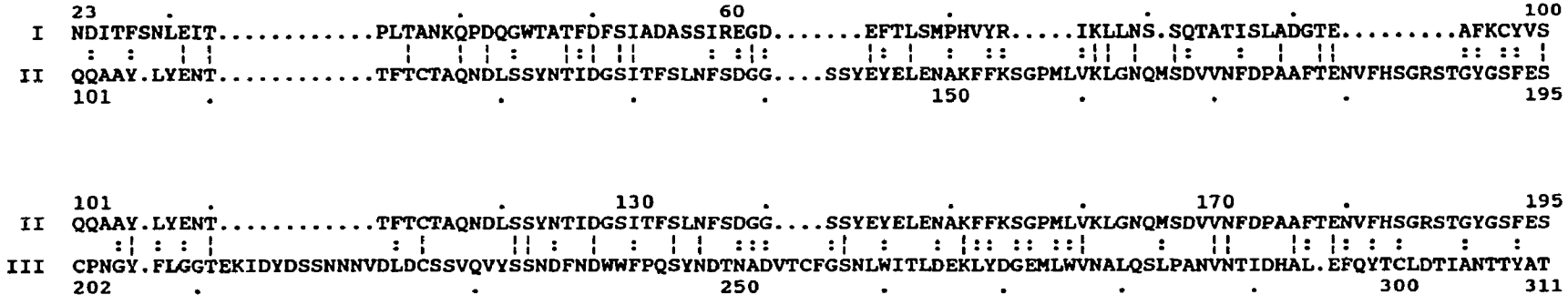
Internal sequence homologies

Analysis of amino acid sequence of the N-terminal 180 amino acids reveals that it contains two regions with apparent sequence similarity (amino acids 20-104 and 105-199). An alignment of these two regions has been made using GCG Bestfit program (Fig.7 upper part). There is 26.0% identity and 50.6% similar residues between these two regions. Analysis of residues 30-94 and 107-180 for the two regions yield a statistically significant alignment, which gives 4.7 SD units (z score) above the mean (GCG BESTFIT, gap weight 3.0, length weight 0.0; Gribskov and Devereux, 1991). The alignment result show that the internal sequence similarity between the two regions is highly significant ($P < 10^{-7}$).

Sequence analysis of region II (105-199) and III (200-330), using the Bestfit program under the same alignment criteria, reveals a 17.2% identity and 43.7% similar residues between the two domains (Fig.7 lower part). The alignment did not yield a Z score with statistical significance under conditions with a combination of different gap and length penalties. The pattern of sequence homology in this alignment, however, well matches the IgV domain consensus alignment of the two regions, based on a combination of sequence and predicted Chou-Fasman secondary structural analysis (Lipke, et al., in preparation; Fig.8). Regions II and III in the Bestfit alignment include most sequence of the predicted 9 ß strands in each proposed domain. The two independent methods yield identical alignment for regions II and III in strands A,B,C',C",E, and F. The only mismatch is in strand C and part of strand D. The coincidence of the alignment results strongly suggest that both regions share structural similarity.

Figure 7. Alignment of amino acid regions 23-100 to 101-195, and regions 101-195 to 202-311. Homologies were assigned and sequences were aligned using Bestfit program in GCG (Gribskov and Devereux, 1991).

Domain


23 60 100
I NDITFSNLEIT.....PLTANKQPDQGWTATFDFSIADASSIREGD.....EFTLSMPHVYR.....IKLLNS.SQTATISLADGTE.....AFKCYVS
II QQAAY.LYENT.....TFTCTAQNDLSSYNTIDGSITFSLNFSDGG....SSYEYELENAKFFKSGPMLVKLGNQMSDVVNFDPAAFTENVFHSGRSTGYGSFES
101 150 195
101 130 170 195
II QQAAY.LYENT.....TFTCTAQNDLSSYNTIDGSITFSLNFSDGG....SSYEYELENAKFFKSGPMLVKLGNQMSDVVNFDPAAFTENVFHSGRSTGYGSFES
III CPNGY.FLGGTEKIDYDSSNNNVDLDCSSVQVYSSNDFNDWWFPQSYNDTNADVTCFGSNLWITLDEKLYDGEMLWVNALQSLPANVNTIDHAL.EFQYTCLDTIANTTYAT
202 250 300 311

The homologous pattern between regions I and II, and regions II and III suggests that these regions share common domain structure.

IgV domain identity in the N-terminal 180 amino acids of α-agglutinin

In addition to internal sequence similarity, regions I and II also show sequence patterns similar to the IgV domain consensus. The alignment of these two regions against members of IgV domain is shown (Fig.8).

First of all, highly conserved sequence patterns in predicted stands B and F of a typical IgV domain are present in these two regions. In region I, Leu^{30}, Ile^{32} and Thr^{33} match the B strand consensus motif of Val/Leu/Ile-Xaa-Val/Leu/Ile-Thr-Cys-Ser/Thr (Fig.8 and Table I). Meanwhile, conservative substitutions for the typical cysteines in B strand (and also in F strand) by amino acid residues with similar structural or hydrophobic properties, like Ser, Val, and Pro, are often found for Ig-like domains that do not have intradomain disulfide bonds (Williams and Barclay, 1988). Pro^{34} is the substitute for the conserved cysteine in B strand. On the other hand, Phe^{95}, Cys^{97} and Val^{99} match the F strand consensus motif of Gly-Xaa-Gly-Xaa-Tyr/Phe-Xaa-Cys-Xaa-Val/Ala (Fig.8 and Table I).

Other consensus residues and conserved substitutions in region I include a Trp in predicted C strand, lying about 11-16 amino acids downstream of the first consensus cysteine, Trp^{45} matches the consensus pattern (Fig.8). An Arg/Lys residue in D strand, lying about 23 amino acids upstream of the second consensus cysteine is also highly conserved among members of IgV-superfamily (Williams and Barclay, 1988; and

Figure 8. Sequence and ß-strand potentials alignment of the α-agglutinin IgV domains with each other and with IgV reference molecules. Sequence alignment of the three regions of α-agglutinin to the reference sequences of IgV domains. Also shown are the Chou-Fasman ß sheet potentials of each sequence, in the same order as the alignment. Alignment adopted with permission from Lipke.

Domain

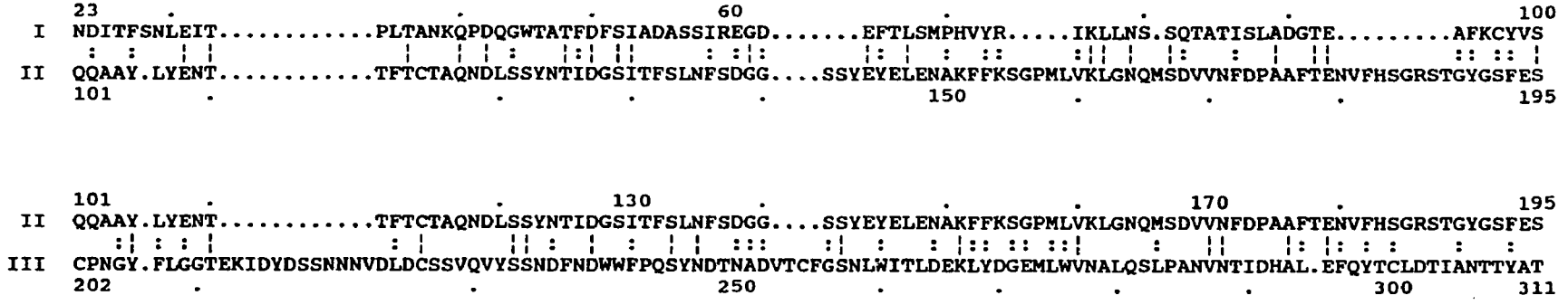

23 60 100
I NDITFSNLEIT.....PLTANKQPDQGWTATFDFSIADASSIREGD.....EFTLSMPHVYR.....IKLLNS.SQTATISLADGTE.....AFKCYVS
II QQAAY.LYENT.....TFTCTAQNDLSSYNTIDGSITFSLNFSDGG....SSYEYELENAKFFKSGPMLVKLGNQMSDVVNFDPAAFTENVFHSGRSTGYGSFES
101 150 195
101 130 170 195
II QQAAY.LYENT.....TFTCTAQNDLSSYNTIDGSITFSLNFSDGG....SSYEYELENAKFFKSGPMLVKLGNQMSDVVNFDPAAFTENVFHSGRSTGYGSFES
III CPNGY.FLGGTEKIDYDSSNNNVDLDCSSVQVYSSNDFNDWWFPQSYNDTNADVTCFGSNLWITLDEKLYDGEMLWVNALQSLPANVNTIDHAL.EFQYTCLDTIANTTYAT
202 250 300 311

TABLE I **Alignment of segments in each of the three N-terminal domains of α-agglutinin to the V-type Ig-superfamily consensus sequence around conserved cysteine residues.** Conserved amino acids in each domain and consensus sequence are underlined. Positions at which conserved residues were found in one or more domains are lined up. Conserved cysteines are also underlined.

Domain:	B strand Position:		F strand Position:	
1.	I---T---F---S---N---L---E---I---T---P	34	D---G---T---E---A---F---K---C---Y---V	97 99
2.	Y---L---Y---E---N---T---T---F---T---C	114	F---T---E---N---V---F---H---S---G---R	184 186
3.	S---S---N---N---N---V---D---L---D---C	227	A---L---E---F---Q---Y---T---C---L---D	300 302
Consensus Sequence:	G---X---X-V/L/I-X-I/L/V-T---C		D---X---G---X--Y/F--X---C---X-V/A	

Gardinier, et al., 1992). Met^{67} might be the substitute for this residue, as it appears in similar position in all 3 domains of α-agglutinin (Fig.8).

In region II, Cys^{114} and Thr^{115} are conserved for the B strand consensus motif (Fig.8 and Table I). Phe^{182} is conserved for the F strand consensus motif, whereas Ser^{184} is the conserved substitute for a consensus cysteine. Met^{158} might be the conserved substitute for the consensus Arg/Lys, which is about 23 amino acid upstream of the second consensus cysteine and in D strand.

In the IgV consensus alignment, the proposed domains I and II fuse at the adjacent ends. The amino acid sequence of Gln^{101} to Thr^{110} is shared as the G strand of domain I and the A strand of domain II, and a close association of these two domains is expected.

Comparison of outcomes of the Bestfit and IgV domain consensus alignments

The observed differences in the outcomes of the Bestfit alignment for internal homology of region I and II (Fig.7 and Fig.8) versus IgV domain alignment of these two regions reflects a difference in criteria used for each alignment. The BESTFIT program aligns regions with identical and similar amino acid residues in best sequential order. Residue properties, like size, charge, polarity, hydrophobicity and hydrophilicity, are scored in this alignment, regardless the geometric or three dimensional structure of regions aligned (Gribskov and Devereux, 1991). This alignment yielded a high Z score to regions with amino acid 30-94 and 107-180, revealing a significant internal sequence homology and likely origin from a common ancestral gene. However, this alignment method does not yield a significant alignment for diversified sequences that share structural conservation. BESTFIT analysis failed to yield a high Z score to the alignment

of regions II and III, which shows significant similarity under alignment algorithms for both sequence and predicted secondary structure similarity (Lipke, et al., in preparation).

Ig-superfamily proteins share conserved Ig folds regardless of sequence diversity. In an Ig-related domain alignment, both the Ig domain structure and sequence conservation become critical. The probability of ß-strand formation and the presence of alternating hydrophobic amino acid residues in the predicted ß-strands are key criteria in addition to a consensus sequence. Alignment of the N-terminal three regions against IgV domain consensus reveals IgV domain identity in region I and II (Fig.8; Lipke et al., in preparation).

Summary

In alignments of potential IgV domains to the Ig-superfamily consensus, both sequence and structure conservation have to be taken into account. A combination of sequence and predicted secondary structure alignment of the N-terminal half of α-agglutinin have been achieved (Lipke et al., in preparation). In such an alignment, both regions I and II can be delineated into nine ß-strands, as in domain III of α-agglutinin and in other known IgV domains.

Multiple sequential Ig domains are very common in the Ig superfamily. The binding region of α-agglutinin (20-200) contains 2 additional regions with size, residues and sequence motifs similar to the IgV domains. The apparent internal homology of these two regions is consistent with the proposed IgV-like domain identities.

We, therefore, propose that the binding region of α-agglutinin contains 3 Ig-like domains and assign residues 20 to 104 to domain I, residues 105-199 to domain II, and residues 200 to 327 to domain III of α-agglutinin. Structurally, residue Gln^{101} to Thr^{110} are predicted to be shared as the G strand of domain I and A strand of domain III. Structural evidence for the presence of Ig fold(s), mainly the anti-parallel ß-sheets core structure, in all three domains is provided in Chapter IV.

Chapter III

Disulfide Bonding and Glycosylation in the binding region of α-agglutinin

Introduction

S. cerevisiae α-agglutinin is a highly glycosylated extracellular matrix protein. The N-linked oligosaccharides account for more than half of the molecular weight of the cell surface extractable α-agglutinin (Terrance et al., 1987; Hauser and Tanner, 1988; Wojciechowicz et al., 1993; Lu et al., 1994;). The affinity of deN-glycosylated α-agglutinin to ConA and a reduction of molecular weight on HF treatment suggest the presence of O-linked carbohydrate on α-agglutinin (Terrance et al., 1987; Hauser and Tanner, 1989).

The N-terminal region of the protein, amino acid residues 20-350, contains the binding site for the ligand a-agglutinin. There are 6 Cys residues in the binding region of α-agglutinin. Four of these cysteines have sequence position and motif typical of disulfide bonded Cys residues in IgV domains (Wojciechowicz et al., 1993; Lipke et al., in preparation; Chapter II). Determination of disulfide patterns in the binding region of α-agglutinin should be useful to elucidate the role of disulfide bonds in the proposed IgV domains structure. It should also provide structural information for the regions of α-agglutinin which remain to be characterized.

N-linked carbohydrate is one of the dominant features in many Ig-superfamily cell adhesion molecules. The N-glycosylation sites in Thy-1 and in domains of CD2 and CD4 that are involved in binding have been characterized (Dwek et al., 1993). The glycosylation sites that are common in these Ig-related domains can, therefore, be established. The binding region of α-agglutinin contains six potential N-glycosylation sites (Asn-Xaa-Ser/Thr). The presence of a Ser and Thr rich region between amino acid residue 300 and 350 suggests the potential for O-glycosylation. Similarly, the

analysis of N- and O-linked glycosylation sites would be useful to provide structure information for the binding region of α-agglutinin.

α-agglutinin$_{20\text{-}351}$, a secreted fragment of α-agglutinin containing the entire binding region, has been constructed. Peptide mapping and sequencing method has been used to analyze disulfide bondings and to identify glycosylation sites.

Materials and Methods

Materials

Yeast strains and expression vector. The *agα1-3* mutant (Lα21), which is isogenic to wild-type strain W303-1B (*MATα ade2-1 his3-11, 15 leu2-3,112 trp1-1 ura3-2 can1-100*), was used to express the α-agglutinin$_{20\text{-}351}$ construct (Lipke et al., 1989). Bioassays utilized tester strain X2180-1A (*MATa SUC2 mal mel gal2 CUP1*) and X2180-1B (*MATα SUC2 mal mel gal2 CUP1*) (Terrance and Lipke, 1981). The expression vector, pPGK, containing the pBR322 $\underline{\text{Amp}^{R}}$, $\underline{\text{Ori}^{E}}$ (an *E. coli* replication origin), a yeast *URA3* gene, and the yeast 2-μm replication origin allows expression of genomic clone between the constitutive phosphoglycerate kinase (PGK) promoter and terminator (Kang et al., 1990).

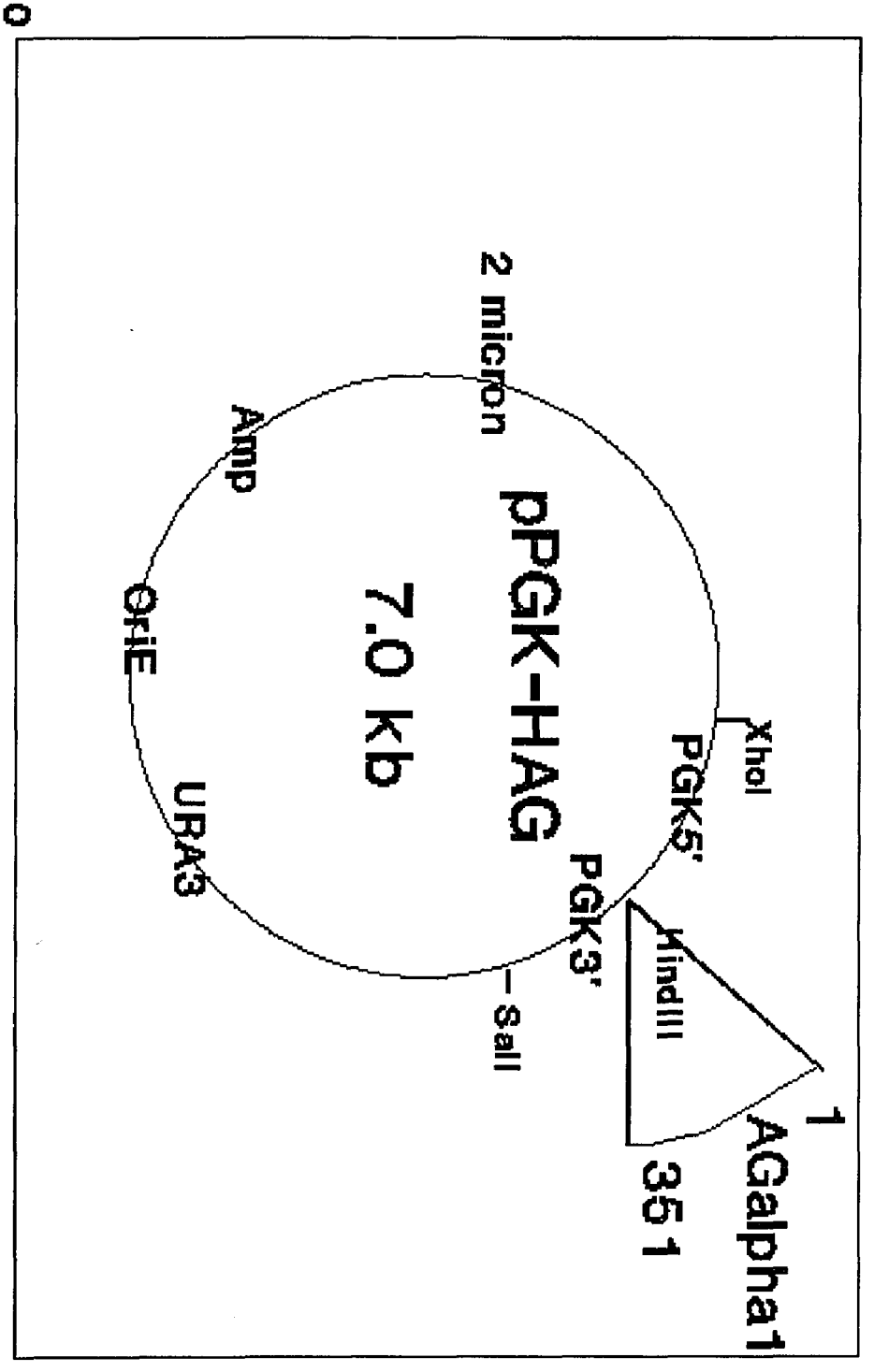
Chemicals and regents. All chemicals were from Sigma, unless otherwise stated. Regents for polymerase chain reactions were obtained from Perkin-Elmer. All the restriction enzymes and DNA molecular weight standard were from New England Biolab or US Biochemical. Reagents for gel electrophoresis were from Bio-Rad for DNA analysis and from IBI for protein analysis. Protein standards were purchased from Bio-Rad. Endoprotease Arg-C, sequencing grade Staph V8, hydrophilic bead-bond trypsin, and endoprotease Asn-N were purchased from Boehringer Mannheim Biochemical Corp. The cysteine-specific reagent P-2007 (N-(1-pyrenemethyl) iodoacetamide) and reducing reagent TCEP (tris(2-carboxyethyl)phosphine hydrochloride) were from Molecular Probe. Immobilon-AV membranes were purchased from Millipore.

Methods

Construction of pPGK-HAGα. Two single-stranded oligonucleotides were synthesized as primers for replication of DNA encoding α-agglutinin$_{20\text{-}351}$. AGα5'-H3', TTC GCC AAG CTT TTC AAA ATG TTC ACT TTT CTC, and AGαM-H3', AAA TGG AAG CTT TGG ATT ACG CAC TAG TGT TTA TAC TTG T, both contained a *HindIII* site (nucleotides with underline) outside the open reading frame. The 3' end primer included a stop codon (nucleotides with double underline) corresponding to Tyr-352 in the deduced α-agglutinin protein sequence. The cDNA for α-agglutinin$_{20\text{-}351}$ with added HindIII sites at both 5' and 3' was amplified from the *AGα1* containing plasmid, pH27 (Lipke et al., 1989). The PCR product flanked the open reading frame of *AGα1* from nucleotides 1 to 1053 and included the sequence encoding the secretion signal. The purified PCR product was cloned in-frame into the corresponding HindIII site of the expression vector, YEpDT-PGK. Diagram of this construct is shown (Fig.9). The orientation of this expression construct, pPGK-hAGα, was checked by mapping with EcoRI, HindIII, and BamHI. The construct was propagated in *Escherichia coli* strain of HB101 under ampicillin selection (Sambrook et al., 1989). The resulting cDNA expression constructs were partially verified by DNA sequence analysis and confirmed by peptide sequencing more than 70% of the polypeptide.

Overexpression and purification of α-agglutinin$_{20\text{-}351}$ from culture supernatant. pPGK-hAGα, encoding α-agglutinin$_{20\text{-}351}$, was introduced to α-agglutinin-deficient mutant Lα21 for the expression of secreted α-agglutinin$_{20\text{-}351}$. Lα21 transformants were grown in one liter of synthetic uracilless medium overnight at room temperature to stationary phase. The cells were centrifuged, and the supernatant was concentrated 10-fold through

Figure 9. Diagram of pPGK-hAGα overexpression construct. The expression shuttle vector, YEpDT-pPGK, contains the pBR322 *Amp*^{R}, *Ori*^{E} (an *E. coli* replication origin), a yeast *URA3* gene, the yeast 2-μm replication origin, and yeast constitutive phosphoglycerate kinase promoter and terminator. The cDNA for α-agglutinin$_{20\text{-}351}$ with added *HindIII* sites at both 5' and 3' was amplified from the *AGα1*-containing plasmid, pH27, using PCR. The PCR product, flanking the open reading frame of *AGα1* from nucleotides 1 to 1053 and including the sequence encoding the secretion signal, was cloned in-frame into the corresponding *HindIII* site of YEpDT-PGK.


0
pPGK-HAG
7.0 kb
2 micron
Amp
ORIE
URA3
PGK5'
PGK3'
XhoI
HindIII
SalI
1
AGalpha1
351

a Millipore filtration apparatus equipped with a membrane with a 100 kDa molecular weight cutoff. Aliquots of concentrated supernatant were dialyzed overnight against 4 liters of 10 mM sodium acetate (pH 5.5) at 4°C. The dialyzed material was first semipurified by chromatography on a DEAE-Sephadex column (120-ml bed volume) which was previous equilibrated with 10 mM sodium acetate (pH 5.5). The column was washed with the equilibrium buffer and eluted with 300 mM sodium chloride, 10 mM sodium acetate (pH 5.5) in 3 ml fractions. The α-agglutinin$_{20\text{-}351}$ content of each eluted fraction was monitored by an agglutination activity assay (Terrance and Lipke, 1981). Fractions containing activity were pooled for further purification.

The active material was dialyzed and lyophilized. The dry powder was reconstituted in 10 mM potassium chloride and incubated with 1:200 to 1:500 molar ratio of endoglycosylase H (endo H) for 4-6 hr at 25°C or overnight at 4°C. The deN-glycosylation α-agglutinin$_{20\text{-}351}$ was chromatographed on a Bio-gel P-60 size exclusion column (60-ml bed volume) which had been previously equilibrated with 30 mM sodium acetate (pH 5.5).

Assay of α-agglutinin$_{20\text{-}351}$ activity. α-Agglutinin$_{20\text{-}351}$ was assayed by its ability to inhibit the agglutinability of **a** cells, as described by Terrance et al. (1987). Samples to be assayed were placed in assay tubes with 0.2 ml (2 X 10^7) of induced **a** cells and assay buffer (0.1 M sodium acetate, pH 5.5, containing 10 μg of cycloheximide per ml) to bring the total volume to 2.8 ml. The tubes were incubated on a rotary shaker at 30°C for 90 min. After incubation, 0.2 ml of 2 x 10^7 cells was added to each tube, vortexed, compacted, suspended, and examined by taking O.D. reading at 660 nm. At doses below 3.5 U, the effect of α-agglutinin on a cells was linear. One unit of α-agglutinin activity

was defined as the amount needed to lower the agglutination index of induced a cells 0.1 AI unit below the control value.

Immunoblot assays. Polyclonal antisera against α-agglutinin were raised by injection of purified deN-glycosylated α-agglutinin$_{20\text{-}351}$. Immunoblots were performed using a method modified from Wojciechowicz et al. (1989). Briefly, after overnight transferring proteins to nitrocellulose membranes, the membranes were blocked with 3% gelatin in PBS and then incubated with 1:1000 dilution of a-cell pretreated anti-α-agglutinin$_{20\text{-}351}$ antibody with 1% gelatin, 0.1% Tween-20 in PBS followed by incubation with 1:1000 dilution of peroxidase-conjugated goat anti-rabbit-IgG antibody (Sigma Chemical Co.) in the same buffer. The blots were stained by peroxidase-mediated reaction of 4-chloro-1-naphthol with hydroperoxide.

Endoprotease digestions. Proteolytic digestions were initially conducted on heat denatured α-agglutinin$_{20\text{-}351}$ in presence of 10% acetonitrile and at 25°C for 18 hours, according to the manufacturers' suggested protocol. The ratio of protease to substrate was 1:25 for trypsin and staph V8, 1:100 for endoprotease Arg-C, and 1:5 for endoprotease Asn-N, respectively. Some trypsin and staph V8 digestions were performed on nondenatured α-agglutinin$_{20\text{-}351}$. Endoprotease Arg-C digestions were performed in 0.1 M NH_4HCO_3, pH 7.8. Staph V8 Digestions were performed in 0.1 M sodium phosphate buffer, pH 7.8. Under these conditions, V8 should cleave at the C-terminal site of both glutamic and aspartic acid, with some cleavage of Asn and Gln residues. Trypsin digestions were performed in 0.1 M tris buffer, pH 8.0. For the endoprotease Asp-N digestion of trypsin-digested α-agglutinin$_{20\text{-}351}$ fragments, HPLC purified tryptic peptides were lyophilized to remove acetonitrile and TFA, and the digestion was performed in 100

mM Tris buffer at pH 7.0.

HPLC. Peptide mixtures derived from digests and reduced digests were fractionated by reversed phase HPLC on an Applied Biosystems instrument used a fully end-capped Microbore Vydac C18 (3 cm x 3 mm i.d., 5 μm) with a Brownlee RP-300 guard column. Solvent A was 0.1%(v/v) aqueous TFA and solvent B was 90% acetonitrile in water (v/v) containing 0.1% TFA (v/v). The solvent elution rate was at 50 μl/min. The column effluent was monitored by absorbance at 220 nm, and peptide peaks were collected manually. For most tryptic digestions, products were fractionated with a linear gradient from 0 to 60% solvent B in 180 min. For identification of cysteine-specific labelled tryptic peptides, the gradient was programmed linearly from 0% B to 100% B in 100 min. For staph V8 digestion, all digests were fractionated with a linear solvent gradient from 0 to 45% solvent B in 180 min.

Cysteine specific labeling of tryptic α-agglutinin$_{20\text{-}351}$. The cysteine-specific reagent, P-2007 (N-(1-pyrenemethyl) iodoacetamide), was used. Fresh P-2007 saturation solution was prepared in N,N-dimethylformamide (DMF). The α-agglutinin$_{20\text{-}351}$ digest was boiled in 8 M urea and 0.1 M phosphate buffer, pH 7.2, for 5 min. An equal volume of DMF was added to the above solution to make 50% DMF. An 1.5 volume of saturated P-2007 DMF solution was added to the above digest containing 50% DMF. The final solution is saturated with P-2007 in 80% DMF. The reaction was performed under dark at room temperature with overnight agitation. In some cases, the digestion mixture was treated with the reducing reagent TCEP, 7 mM, before labeling. Additives and trypsin beads were removed by a Bio-Gel P-2 spin column after labeling. The labelled mixture was separated on HPLC and the eluting materials were detected at 341 nm,

characteristic of P-2007.

NH2-terminal sequencing of peptides. Peptides were sequenced by automated Edman degradation in a gas-phase sequenator (Model 470A, Applied Biosystems Inc.). The resultant phenylthiohydantion-(PTH)-derivatized amino acid residues were separated on a Vydac C18 column using a 120A PTH Analyzer (Applied Biosystems Inc.). Individual amino acid residues were identified and quantitated by comparison with standards.

Reduction of disulfide bonds. α-agglutinin$_{20\text{-}351}$ was treated with 0.1%, 0.2%, 1% and 5% DTT (7 mM to 350 mM) in 50 μl of 10 mM Tris buffer, pH 7.8 for 30 min at 37°C. DTT containing solutions were then removed by ultrafiltration in Millipore MC microtubes and α-agglutinin$_{20\text{-}351}$ was reconstituted in 100 mM sodium acetate pH 5.5 to 500 μl. 5 μg of α-agglutinin$_{20\text{-}351}$ was used for each treatment and a solution with 5% DTT but without α-agglutinin$_{20\text{-}351}$ was treated the same as all the samples as a control. The samples were serially diluted and assayed as previously described (Terrance and Lipke, 1981).

Dot blot analysis of O-linked proteolytic α-agglutinin$_{20\text{-}351}$ peptides from HPLC Each fraction from the reverse phase column was vacuum evaporated to dryness and reconstituted in 30 μl of 0.1% SDS in 0.5 M sodium phosphate pH 8.0. Each peptide sample (3x1 μl) was spotted onto an Immobilon-AV membrane. The membrane was air-dried and incubated with 10 mM Tris-HCl, pH 7.5, containing 0.15 M sodium chloride and 0.1% Tween 20 (TTBS) for 30 min and then blocked with fresh 10% ethanolamine in 1M sodium bicarbonate, pH 9.5, for 2 hour. After blocking, the membranes were incubated for 1 hour with 0.5 μg/ml Concanavalin A (Con A) conjugated peroxidase

(1:2000 from 1 mg/ml stock solution) in TTBS. After washing 3 times with TBS, the membrane were stained with 4-chloro-1-naphthol and hydrogen peroxide (Canas, et al., 1993).

Other procedures. SDS-polyacrylamide gel (12-15%) electrophoresis was performed according to the method of Laemmli and proteins were visualized by Comassie Blue.

RESULTS

Expression of α-agglutinin$_{20\text{-}351}$ as a glycoprotein

The level of α-agglutinin activity in culture supernatants of Lα21[pPGK-hAGα] was examined. Each liter of the supernatant contained an average of $4.5X10^4$ units of agglutination activity, a yield of α-agglutinin$_{20\text{-}351}$ of about 1 mg per liter (Terrance et al., 1987; Wojciewchowicz et al, 1993). α-agglutinin$_{20\text{-}351}$ in crude culture supernatants was identified by immunoblots before and after endo H treatment (Fig.10). The fully glycosylated α-agglutinin$_{20\text{-}351}$ had an apparent molecular size of 110 kDa (Lane a). After removal of N-linked carbohydrates with endo H, the molecular size of α-agglutinin$_{20\text{-}351}$ was reduced to 45 kDa (Lane c). The deduced Mr of α-agglutinin$_{20\text{-}351}$ from the predicted amino acid sequence is 37,108. Therefore, N-linked carbohydrates accounts for two thirds of the apparent 110 kDa molecular mass of α-agglutinin$_{20\text{-}351}$, and the O-linked carbohydrate remaining after Endo H digestion could account for an apparent 8 kDa.

In addition, the endo H-treated form of α-agglutinin$_{20\text{-}351}$ showed a clear molecular weight shift after dithiothreitol treatment (Lane d), suggesting that α-agglutinin$_{20\text{-}351}$ contains one or more intramolecular disulfide bonds.

Fig.11 shows the elution pattern of endo H-treated α-agglutinin$_{20\text{-}351}$ from a Bio-Gel P-60 column. The deN-glycosylated α-agglutinin$_{20\text{-}351}$ appeared as a 45 kDa doublet.

Identification of disulfides in endoprotease Arg-C digestions. Endoprotease Arg-C cleaves α-agglutinin$_{20\text{-}351}$ at Lys^{154} and releases a 20 kDa fragment containing residues 20-154 and a 30 kDa fragment containing residues 155-351 (α-agglutinin$_{155\text{-}351}$, detailed in Chapter VI). Fig.12 shows that the 20 and 30 kDa fragments were separable

Figure 10. Immunoblot of α-agglutinin$_{20\text{-}351}$ from culture supernatant. Supernatant from a culture of Lα21[pPGK-hAGα] (17 ml) was lyophilized to dryness, reconstituted in 200 μl distilled water, and passed through a Bio-gel P-10 column preequilibrated with 0.01 M sodium acetate pH 5.5 for desalting. 20 μl of desalted material was treated with or without 0.5 μl of 1 U/ml endo H at room temperature for 2 hours. Samples without (*Lane a* and *b*) and with (*Lane c* and *d*) Endo H treatment were analyzed by electrophoresis in presence (*Lane b* and *d*) or absence (*Lane a* and *c*) of the reducing reagent dithiothreitol.

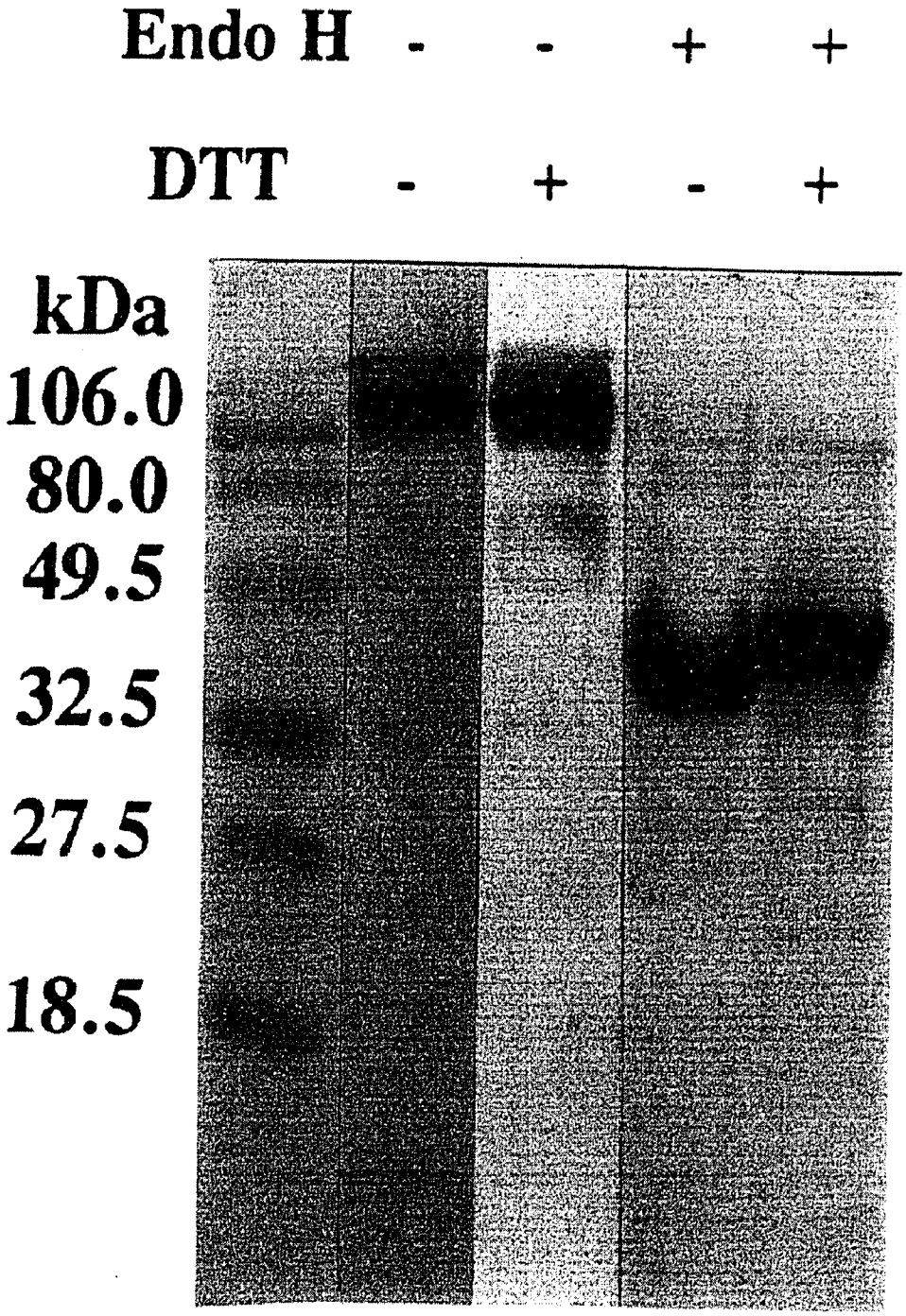
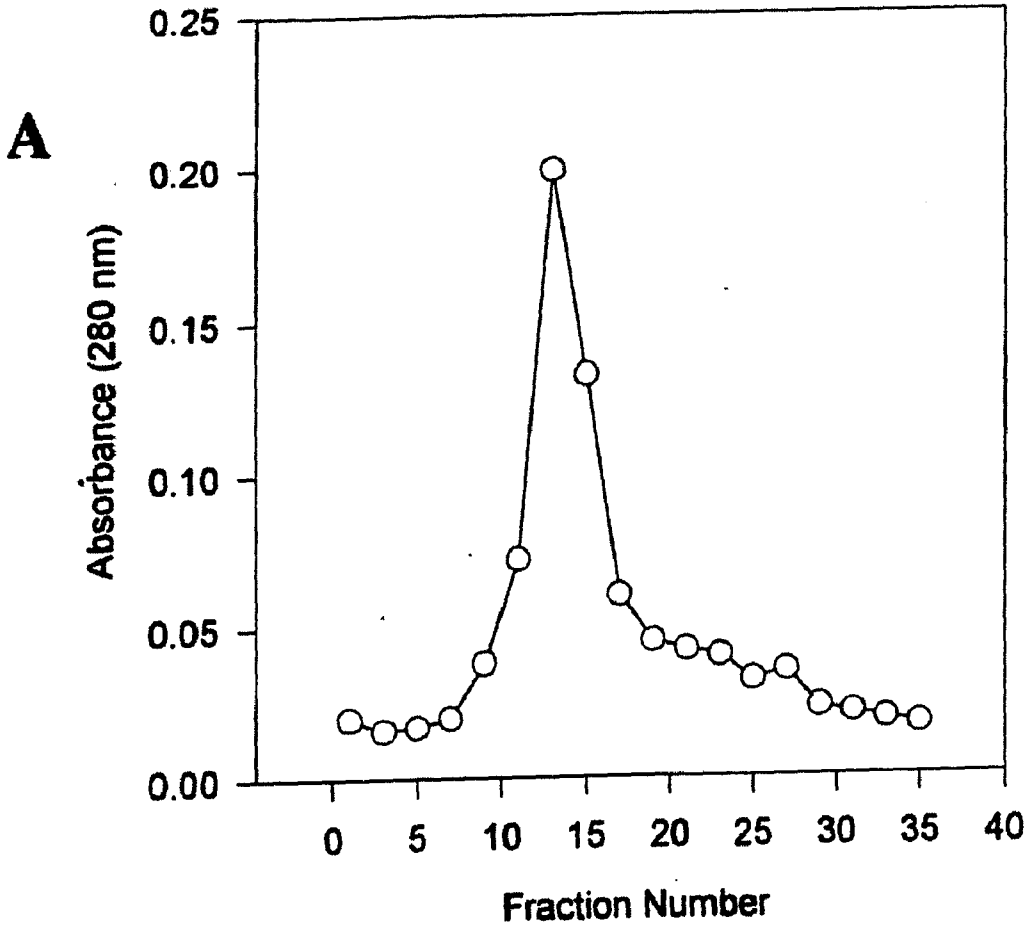

Endo H - - + +
DTT - + - +
kDa
106.0
80.0
49.5
32.5
27.5
18.5

Figure 11. Bio-gel P-60 chromatography of endo H treated α-agglutinin$_{20\text{-}351}$. The active material from DEAE-Sephadex A-25 was lyophilized to dryness. The material was reconstituted and dialyzed against 0.03 M sodium acetate, pH 5.5, treated with endo H (15 μl of 1 U/ml endo H to 2000 U of α-agglutinin activity), and loaded onto a Bio-Gel P-60 column preequilibrated with the same buffer. Fractions (3 ml) were collected and monitored at 280 nm (*Panel A*). Aliquots of fractions were electrophoresed on a 12% SDS-PAGE gel and visualized by staining with Coomassie blue (*Panel B*). Molecular size markers are shown on the left.


A
0.25
0.20
0.15
0.10
0.05
0.00
Absorbance (280 nm)
0
5
10
15
20
25
30
35
40
Fraction Number

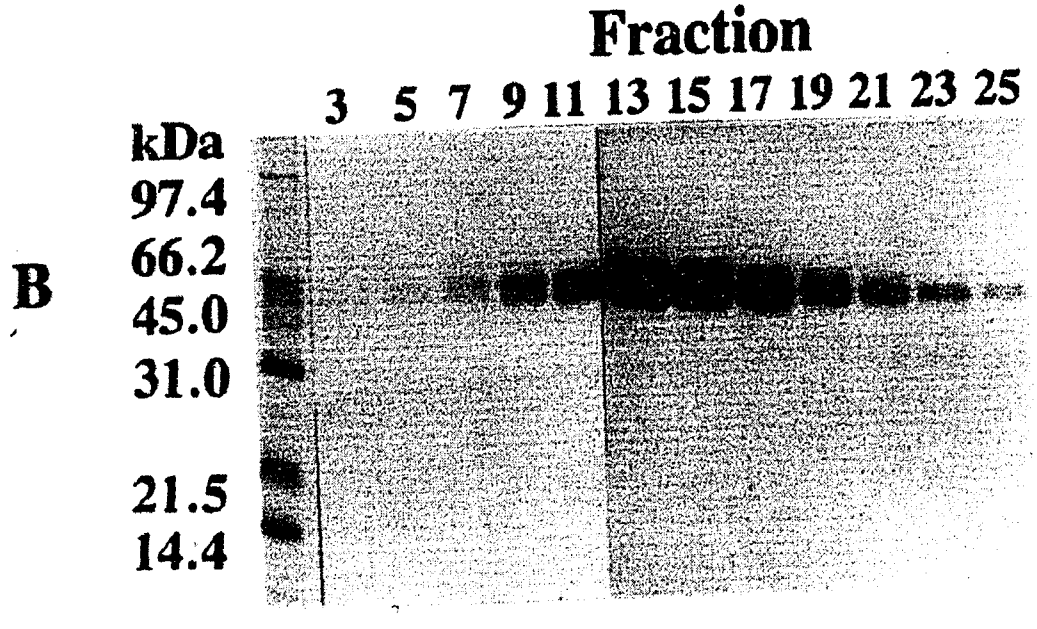
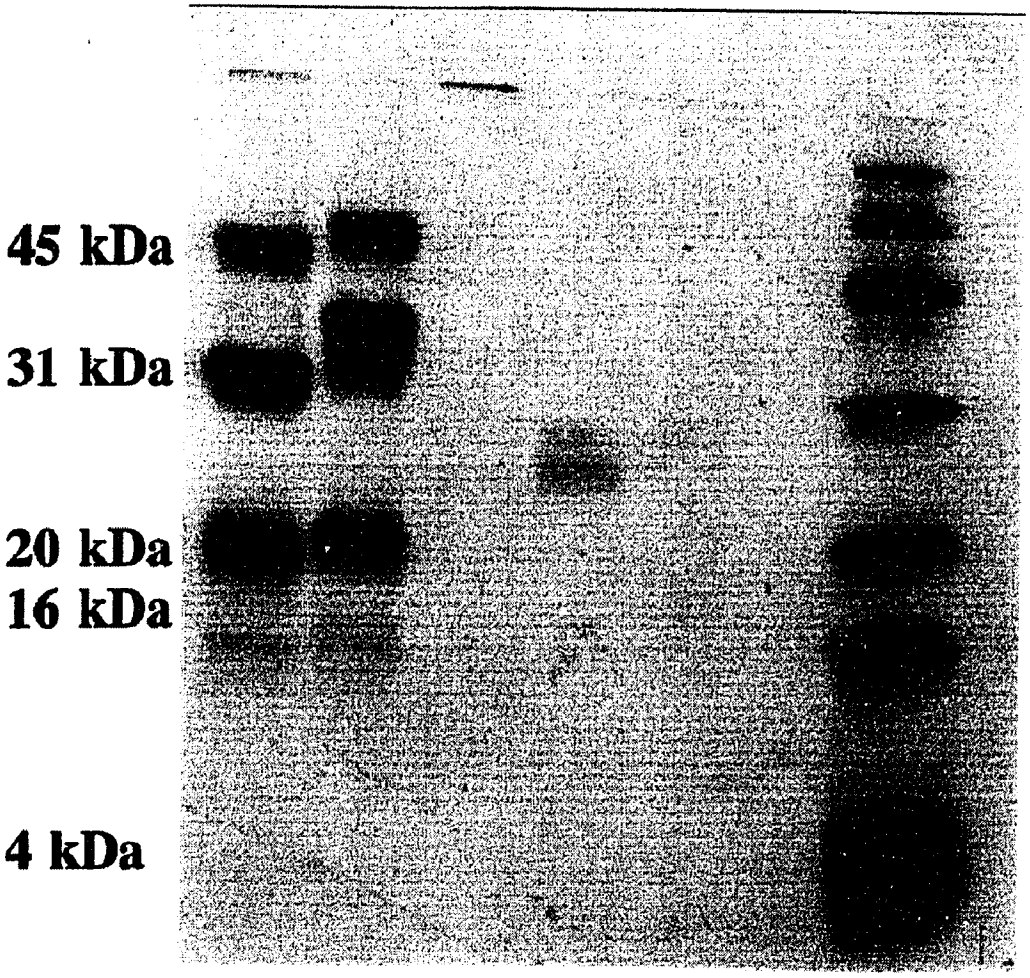

Fraction
3 5 7 9 11 13 15 17 19 21 23 25
B
kDa
97.4
66.2
45.0
31.0
21.5
14.4

Figure 12. SDS-PAGE analysis of endoprotease Arg-C digested-α-agglutinin$_{20\text{-}351}$. Samples of endoprotease Arg-C digested-α-agglutinin$_{20\text{-}351}$ (*Lane 1* and *2*) and endoprotease alone (*Lane 4* and *5*) were treated with (*Lane 2 and 5*) or without (*Lane 1 and 4*) DTT and electrophoresed on a 15% SDS-polyacrylamide gel, and the gel was stained with Coomassie Blue. Molecular weight standards on the right were from 97,400 to 4,000 Da.


45 kDa
31 kDa
20 kDa
16 kDa
4 kDa
Enzyme
DTT - + - +

in the absence of reducing agent, and that each gave a higher apparent molecular size on SDS-PAGE after DTT treatment, suggesting that each fragment contained disulfide bond(s) but that there is no disulfide linkage between them.

Based on deduced amino acid sequence, the 20 kDa fragment contained Cys^{97} and Cys^{114}, while the 30 kDa fragment, α-agglutinin$_{155\text{-}351}$, contained Cys^{202}, Cys^{227}, Cys^{256} and Cys^{300}. The result of Arg-C digestion along with molecular weight shift on DTT treatment suggested that Cys^{97} and Cys^{114} in the 20 kDa fragments form a disulfide bond. The disulfide conjugate in α-agglutinin$_{155\text{-}351}$ could not be determined from this data, since 4 cysteines could result in various disulfide linkages.

Mapping disulfide bonds in α-agglutinin$_{20\text{-}351}$

Outlines for identification of disulfide bonds. The procedures for peptide mapping of the disulfide structure in α-agglutinin$_{20\text{-}351}$ are summarized in Fig.13. α-agglutinin$_{20\text{-}351}$ was subject to proteolytic digestion with trypsin or staph V8 (Fig.13 panel A left and panel B). The disulfide bonded peptides were identified by change of retention times on the reverse phase HPLC chromatograms upon reduction. Peptides that were unique to the non-reduced condition were likely to contain disulfide bonds, while peptides unique to the reduced condition represent peptides separated by the reduction of the disulfide.

Fig.14 shows HPLC chromatograms of tryptic digests under reduced and nonreduced conditions. Three tryptic peptides (T1, T2, T2') were unique to the nonreduced chromatogram (Panel A), whereas tryptic peptides DT1, DT2, and DT3 were unique to the reduced chromatogram (Panel B). These peptides were subject to microsequencing analysis and the results were summarized in Tables II and III. The

Figure 13. Summary of procedures for mapping disulfide bonds in α-agglutinin$_{20\text{-}351}$. Disulfide bonds and free cysteines were mapped with trypsin (*Panel A*) and staph V8 (*Panel B*). Free sulfhydryls were confirmed by cysteine-specific labelling and endoprotease Asn-N (*Panel C*).

A: α-agglutinin$_{20\text{-}351}$ Digested with Trypsin (1:25)

↓

Aliquot of the Digest Treated with DTT

↓

Identification of Peaks Unique to Reduced/Nonreduced Digests on HPLC

↓

Amino Acid Sequence Analysis of All Candidate Peptide Fractions

↓

Reduced or Nonreduced Digests Labelled with P-2007

↓

Pass through Bio-gel P-6 Column to Remove Reagents for Labelling

↓

Identification of Cysteine Containing Peptides on HPLC

B: α-agglutinin$_{20\text{-}351}$ Digested with V8 (1:25)

↓

Aliquot of the Digest Treated with DTT

↓

Identification of Peptide Peaks Unique to the Reduced/Nonreduced Digests on HPLC

↓

Amino Acid Sequence Analysis of All Candidate Peptide Fractions

C: Isolation of Tryptic α-agglutinin$_{20\text{-}351}$ Fragment Containing Cys227 and Cys256 from RP-HPLC

↓

Labelled with P-2007 under Nonreduced Condition

↓

Re-chromatography

↓

Digestion of the Labelled Peptide with Endoprotease Asp-N

↓

Rechromatograph on HPLC

↓

Identification of the Two Labelled Cys-Containing Peptides

Figure 14. Chromatogram of reduced and nonreduced trypsin-digested α-agglutinin$_{20\text{-}351}$. Mixtures of trypsin digested peptides without (*Panel A*) or with (*Panel B*) DTT treatment were chromatographed as specified in "Materials and Methods". Peaks unique to the nonreduced (T1, T2, and T2') and reduced (DT1 - DT3) profiles were labeled. The peptide containing Cys^{227} and Cys^{254} is peak T3 in the nonreduced profile and peak DT4 in the reduced profile. The amino acid sequences of these peptides are listed in Table II, III, and IV.

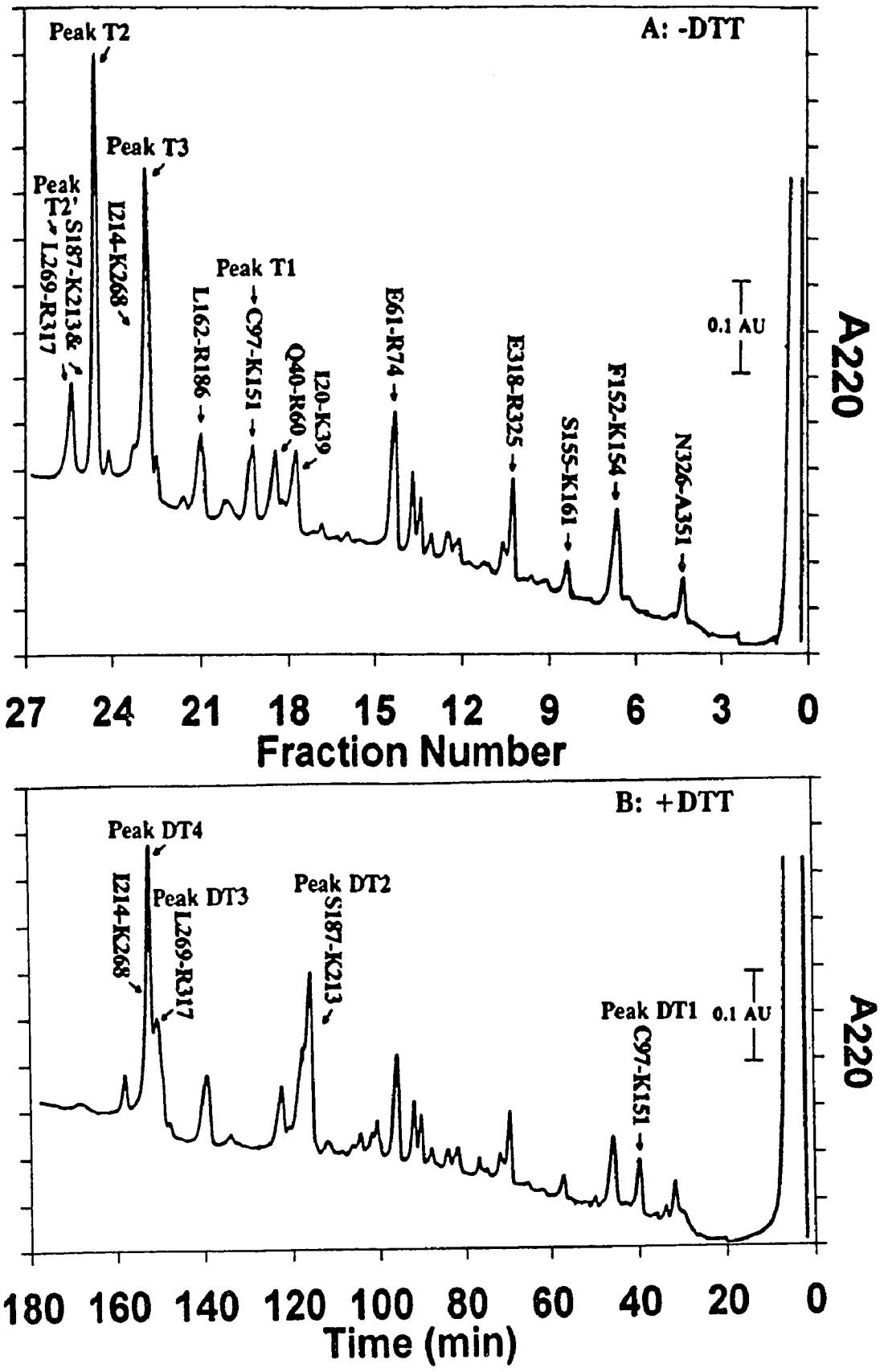

A: -DTT
Peak T2
Peak T3
Peak T2'
S187-K213&
L269-R317
I214-K268
Peak T1
L162-R186
C97-K151
Q40-R60
I20-K39
E61-R74
E318-R325
S155-K161
F152-K154
N326-A351
0.1 AU
A220
27 24 21 18 15 12 9 6 3 0
Fraction Number
B: +DTT
Peak DT4
I214-K268
Peak DT3
L269-R317
Peak DT2
S187-K213
Peak DT1
C97-K151
0.1 AU
A220
180 160 140 120 100 80 60 40 20 0
Time (min)

Figure 15. Chromatogram of reduced Staph V8 digested α-agglutinin$_{20\text{-}351}$. A peptide mixture after DTT treatment was chromatographed as specified in "Materials and Methods". The resulting peaks unique to the reduced profiles are labeled (DS1 - DS3). The peptide Cys^{227} and Cys^{254} is labeled as peak DS4.

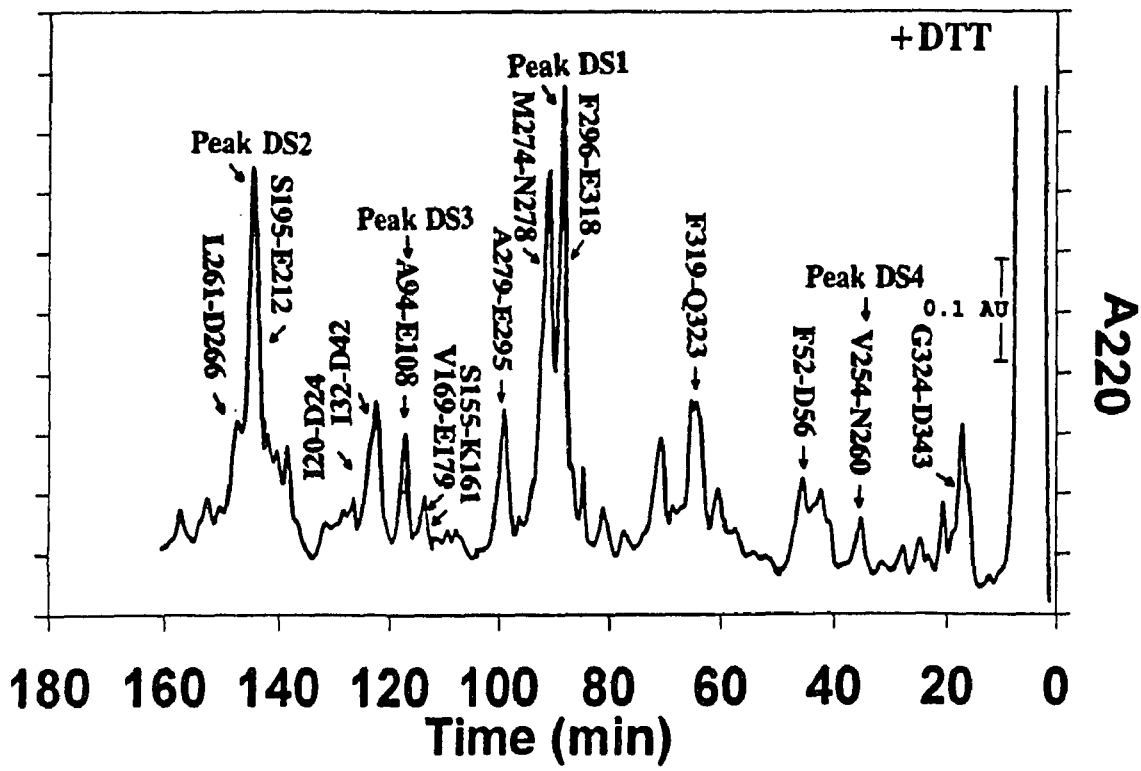

+DTT
Peak DS1
M274-N278
F296-E318
Peak DS2
S195-E212
L261-D266
Peak DS3
A279-E295
A94-E108
S155-K161
V169-E179
I20-D24
I32-D42
F319-Q323
Peak DS4
F52-D56
V254-N260
G324-D343
0.1 AU
A220
180 160 140 120 100 80 60 40 20 0
Time (min)

TABLE II **NH_2-terminal sequence analysis of tryptic peptides unique to non-reduced digest[a]**

	Peak T1	*Peak T2*		*Peak T2'*	
Residue	Cys^{97} & Cys^{114} Containing peptide	Cys^{202} / Cys^{300} Containing peptides		Cys^{202} / Cys^{300} Containing peptides	
1	Xc	S(63)	L(80)	S(32)	L(37)
2	Y(19)	T(52)	Y(72)	T(29)	Y(33)
3	V(27)	G(70)	D(51)	G(38)	D(27)
4	S(14)	Y(59)	G(55)	Y(31)	G(28)
5	Q(16)	G(64)	E(28)	G(27)	E(23)
6	Q(13)	S(27)	M(28)	S(19)	M(21)
7	A(16)	F(39)	L(35)	F(24)	L(25)
8	A(14)	E(32)	W(12)	E(17)	W(15)
9	Y(11)	S(18)	V(21)	S(8)	V(19)
10	L(12)	Y(25)	N(23)	Y(15)	N(14)
11	Y(8)	H(10)	A(24)	H(9)	A(12)
12	E(6)	L(18)	L(18)	L(15)	L(15)
13		G(20)	Q(17)	G(10)	Q(8)
14		M(13)	S(8)	M(7)	S(5)
15		Y(16)	L(20)	Y(6)	L(7)
16		Xc	P(17)	Xc	P(4)
17		P(20)	A(16)	P(5)	
18		N(10)	N(10)		
19		G(13)	V(12)		
20		Y(9)	N(17)		

[a] Tryptic peptides unique to the non-reduced digest were sequenced. X represents residues missing from the sequence. Lower case letters following X represent the residue deduced from the DNA sequence at the corresponding position. The numbers in parentheses refer to the pmole yield of each residue.

TABLE III **NH2-terminal sequence analysis of tryptic and staph V8 digested α-agglutinin$^{20\text{-}351}$ peptides unique to reduced digests.**[a]

	Staph V8 Digestion			*Tryptic Digestion*		
Residue	Peak DS3 Containing Cys^{97}	Peak DS2 Containing Cys^{202}	Peak DS1 Containing Cys^{300}	Peak DT1 Containing Cys^{97} & Cys^{114}	Peak DT2 Containing Cys^{202}	Peak DT3 Containing Cys^{300}
1	A(26)	S(14)	F(53)	Xc	S(7)	L(17)
2	F(18)	Y(16)	Q(54)	Y(19)	T(6)	Y(7)
3	K(11)	H(13)	Y(60)	V(23)	G(9)	D(13)
4	Xc	L(22)	Xt	S(14)	Y(9)	G(17)
5	Y(6)	G(31)	Xc	Q(17)	G(8)	E(11)
6	V(8)	M(17)	L(54)	Q(13)	S(2)	M(9)
7	S(4)	Y(18)	D(15)	A(12)	F(6)	L(10)
8	Q(7)	Xc	Xt	A(10)	E(5)	W(7)
9	Q(8)	P(9)	I(65)	Y(7)	S(2)	V(7)
10	A(12)	N(13)	A(52)	L(8)	Y(5)	N(13)
11	A(11)	G(14)	Xn	Y(5)		A(12)
12	Y(8)	Y(11)	Xt			L(12)
13	L(5)	F(8)	Xt			Q(8)
14	Y(7)	L(10)	Y(43)			S(7)
15	E(2)	G(9)	A(36)			L(11)
16		G(8)	Xt			P(6)
17		T(5)	Q(32)			A(11)
18		E(4)	F(39)			N(8)
19			Xt			V(9)
20			Xt			N(6)

[a] Endo H treated and reduced α-agglutinin$_{20\text{-}351}$ (100 μg) was cleaved with trypsin or Staph V8. The digests were separated and sequenced as specified under "Materials and Methods". Designations are as in Table II.

Table IV Predicted tryptic and Staph V8 α-agglutinin$_{20-351}$ peptides containing Cysteines.[a]

A: Tryptic group

Cys	Peptide Sequence	Sequence Position	Number of Residues
97 & 114	▼----------▼ CYVSQQAAYLYENTTFTCTAQNDLSS- -YNTIDGSITFSLNFSDGGSSYEYELENAK	97 - 151	55
300	LYDGEMLWVNALQSLPANVNTIDHAL- -EFQYCLDTIANTTYATQFSTTR ▼ ---------- ▲	269 - 317	48
202	STGYGSFESYHLGMYCPNGYFLGGTEK	187 - 213	37
227 & 256	IDYDSSNNNVDLDCSSVQVYSSNDFNDW- -WFPQSYNDTNADVTCFGSNLWITLDEK	214 - 268	55

B: Staph V8 group

Cys	Peptide Sequence	Sequence Position	Number of Residues
97	AFKCYVSQQAAYLYE	94 - 108	15
114	NTTFTCTAQND	109 - 119	11
202	SYHLGMYCPNGYFLGGTE	195 - 212	18
300	FQYTCLDTIANTTYATQFSTTR	296 - 318	23
227	CSSVQVYSSND	227 - 237	11
256	VTCFGSNLWITLD	254 - 266	13

[a] tryptic (A) and Staph V8 (B) fragments were predicted from the primary sequence of α-agglutinin$_{20-351}$.

amino terminal sequences of these tryptic peptides were then compared to the sequences of the predicted Cys-containing tryptic fragments from the primary structure of the protein (Table IV).

Fig.15 is the HPLC chromatograms of the staph V8 digest under reduced conditions. Three peptides showed unique to the reduced condition (DS1, DS2, DS3). These peptides are subject to sequencing analysis and the results were summarized in Table III. The NH2-terminal amino acid sequences of these peptides were also compared to the sequences of the predicted Cys-containing staph V8 fragments from the primary structure of the protein (Table IV).

The identified disulfide patterns (detailed below) in α-agglutinin$_{20\text{-}351}$ were confirmed by cysteine-specific labeling of the trypsin digestion with P-2007 in presence or absence of the reducing reagent TCEP, as indicated in the right part of panel A in Fig.13. After the free probe was removed, the reaction was chromatographed on HPLC and monitored at wavelength of 341 nm for labelled peptide. Only one P-2007-derivatized peptide was observed in the profile of nonreduced digest (Fig.16A), whereas 4 labelled species were observed in the profile of reduced digest (Fig.16B).

Cys^{97} and Cys^{114} are disulfide bonded. Tryptic peptide peak T1 (retention time, 127 min) was unique to the nonreduced condition and showed a shift in retention time after DTT treatment (Fig.14A, peak T1 and Fig.14B, peak DT1). This disulfide bonded peptide contains Cys^{97} (Table II). Based on the primary sequence (Table IV), there is no tryptic site between Cys^{97} and Cys^{114}, and consequently the single tryptic peptide containing Cys^{97} also contains Cys^{114}. These results suggest that Cys^{97} forms an intra-fragment disulfide bond with Cys^{114}.

The involvement of Cys^{114} in a disulfide bond was confirmed by sequencing analysis of peptide peak DS3. This peptide was present only after reduction of the V8 digest (Fig.15 and Table III).

Cys^{202} and Cys^{300} are disulfide bonded. Tryptic peptide peaks T2 and T2' (retention times 160 and 170 min, respectively) were unique to the nonreduced chromatogram (Fig.14A). Peptide sequencing analysis of these two disulfide bonded peptide peaks illustrated that both peaks yielded two sequences, consisting of Cys^{202}-containing and Cys^{300}-containing peptides (Table II). Both sequences were present in approximately equimolar quantities. This disulfide bonded peptide consistently chromatographed as a doublet. Identical peptides substituted with different oligosaccharide components can form more than one reverse-phase peak (Arterburn et al., 1990). It is therefore possible that this dual peptide peak may be the result of minor variation in the number of sugar residues remaining after deN-glycosylation, or of some other posttranslational modification for this portion of the protein.

Tryptic peptide peaks DT2 and DT3 (retention times 151 min and 115 min, respectively) were unique to the reduced profile (Fig.14B), which indicated that each of them was involved in a disulfide bond. Sequencing of these two half cystine containing peptides illustrated that peak DT2 and DT3 included Cys^{202} and Cys^{300}, respectively (Table III).

A staph V8-generated peptide that was unique to the nonreduced chromatogram also had two sequences, confirming that Cys^{202} and Cys^{300} formed a disulfide bond (data not shown). Staph V8-generated peptide peaks DS2 and DS1 (retention times 87 min and 145 min, respectively) were unique to the reduced chromatogram (Fig.15) and contained

Cys^{202} and Cys^{300} (Table III), respectively.

Identification of free sulfhydryls

A peptide including a single Cys residue (256) was obtained and sequenced from staph V8 digestion under both nonreduced (data not shown) and reduced conditions (peak DS4 in Fig.15 and Table V). Therefore, Cys^{256} has a free sulfhydryl.

Tryptic peptide peak T3 from nonreduced α-agglutinin$_{20\text{-}351}$ and peak DT4 from profile for reduced α-agglutinin$_{20\text{-}351}$ had the same retention times (155 min, Fig.14). Both peaks were the same Cys^{227} containing peptide (Table V). There is no tryptic site between Cys^{227} and Cys^{256} (Table IV). This tryptic peptide, therefore, contained both cysteines. Since this tryptic peptide peak did not show a shift in its retention time in the reduced HPLC chromatogram (Fig.14A and 14B), there appeared not to be an intra-fragment disulfide bond. These results indicate that Cys^{227} as well as Cys^{256} has a free sulfhydryl.

Confirmation of disulfide pattern and free cysteines by cysteine-specific labeling

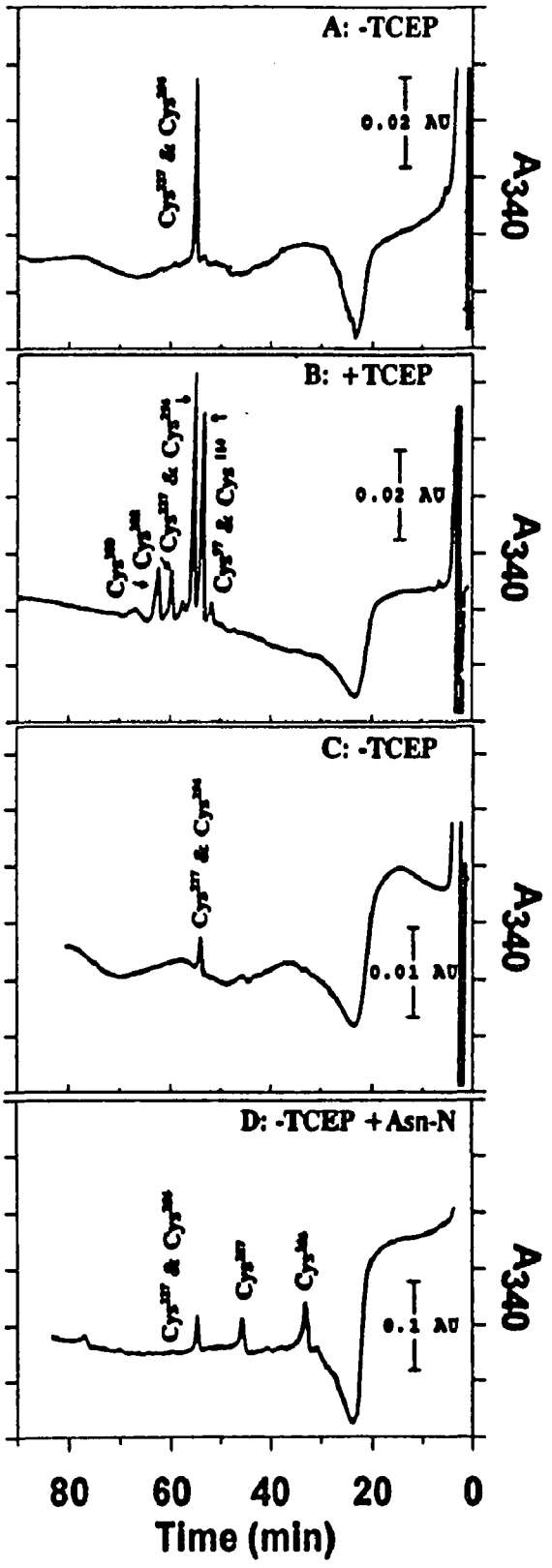
Cys^{227} and Cys^{256} can be labelled under nonreduced conditions. To verify that peptide peak T3 in the nonreduced profile contained Cys^{227} and Cys^{256} as free sulfhydryls, this peptide was labelled with P-2007 and analyzed using the procedure summarized in panel C of Fig.13. Fig.16C shows that this tryptic peptide can be labelled under nonreducing conditions. The alkylation of this peptide proved the presence of free cysteines in this peptide.

To determine if this peptide contains two labelled cysteines, the fraction containing the labelled peptide was further digested with endoprotease Asp-N and rechromatographed (Fig.16D). The original labelled peptide with a retention time of 53 min was still present, probably due to incomplete digestion. Two additional labelled

TABLE V **NH_2-terminal sequence analysis of peptides containing free sulfhydryl cysteines. Designations as in Table II.**

Residue	Tryptic Peak T3	Staph V8 Peak DS4
	Cys^{227} & Cys^{256} Peptide	Cys^{256} peptide
1	I (53)	V (18)
2	D (15)	T (9)
3	Y (15)	Xc
4	D (12)	F (13)
5	S (7)	G (12)
6	S (9)	S (5)
7	N (11)	N (5)
8	N (11)	
9	N (11)	
10	V (9)	
11	D (6)	
12	L (13)	
13	D (6)	
14	Xc	
15	X	
16	X	
17	X	
18	Q (3)	
19	V (3)	
20	Y (3)	

Figure 16. HPLC chromatograms of P-2007-labelled tryptic peptides of α-agglutinin$_{20\text{-}351}$. Tryptic peptides were labelled with P-2007 presence or absence of the reducing regent TCEP were fractionated by reversed phase HPLC using the standard program (*Panel A* and *Panel B*). Peptide T3 of figure 14, containing Cys^{227} and Cys^{254}, was labelled under nonreduced conditions (*Panel C*), isolated, digested with endoprotease Asn-N, and rechromatographed (*Panel D*).


A: -TCEP
0.02 AU
A340
B: +TCEP
0.02 AU
A340
C: -TCEP
0.01 AU
A340
D: -TCEP +Asn-N
0.1 AU
A340
80
60
40
20
0
Time (min)

peptides were detected at 35 min and 45 min, as a result of the digestion. Therefore, both Cys^{227} and Cys^{256} are free cysteines.

Cys^{97}, Cys^{114}, Cys^{202}, and Cys^{300} can be labelled only after reduction. To confirm the presence of the two disulfide bonds, tryptic digests were subject to P-2007 labeling in presence or absence of the reducing reagent TCEP, using procedures described in the right part of panel A Fig.13. There was only one P-2007-derivatized peptide in the profile of nonreduced digest, with retention time characteristic of the peptide containing the labeled Cys^{227} and Cys^{256} residues (Fig.16A and Fig.16C). There were 4 labelled species in the profile of reduced digest (Fig.16B). These peptides were corresponded to all the identified Cys-containing peptides.

Peak P1 should contain Cys^{97} and Cys^{114}, since it showed an equivalent absorbance molar ratio to that of peak P2, which contained two labelled cysteines. Peak P1 eluted at the predicted retention time for this peptide (Gribskov and Devereux 1991).

Peak P3 and P4 both were smaller peaks, showing equivalent absorbance and retention time of 56.5 and 59.5 min. These two labelled peptides should contain Cys^{202} and Cys^{300}, respectively, based on the composition. Therefore, alkylation analysis confirmed the assignment of disulfides: Cys^{97}-Cys^{114} and Cys^{202}-Cys^{300}, with free sulfhydryls in Cys^{226} and Cys^{257} in α-agglutinin$_{20\text{-}351}$.

Disulfide linkages and binding activity

To determine whether disulfide bonds in α-agglutinin$_{20\text{-}351}$ play a role in the binding activity, α-agglutinin$_{20\text{-}351}$ was treated with DTT. DTT was removed by ultrafiltration and α-agglutinin$_{20\text{-}351}$ was reconstituted in sodium acetate buffer at pH 5.5. Treatment with 0.1% (7 mM) DTT did not affect the activity (Table VI). Treatment with

TABLE VI The agglutination activity of α-agglutinin$_{20\text{-}351}$ after DTT treatment.[a]

α-agglutinin$_{20\text{-}351}$ treatment	Lowest active concentration (ng/ml)
10 mM tris/pH7.8	1
+ 0.1% DTT	1
+ 0.2% DTT	3
+ 1.0% DTT	$>10^2$
+ 10.0% DTT	$>10^3$

[a] α-agglutinin$_{20\text{-}351}$ were treated with various concentrations of DTT at pH 7.0 for 1 hour and reconstituted in 0.1 M sodium acetate buffer for assay.

0.2% (13mM) DTT, however, resulted in loss of 67% of the activity. The DTT treatment with a concentration of 1% or above inactivated α-agglutinin$_{20\text{-}351}$ completely.

Identification of O-linked glycosylations

Identification of O-linked glycosylation sites by peptide sequencing. Glycosylated Ser or Thr residues are more hydrophilic than the corresponding amino acids, resulting in lack of detection by the sequencer. Therefore, peptide sequencing provides an indirect method to identify O-linked glycosylation sites.

Tryptic or staph V8 peptides resolved by reverse-phase HPLC (Fig.14 and Fig.15) were sequenced. Absence of signals at positions predicted to be Thr and Ser, based on the gene sequence, were assumed to be due to glycosylation. Table VII summarizes the results from sequencing of staph V8 and tryptic α-agglutinin$_{20\text{-}351}$ peptides from two or more independent peptide sequences. A total of 4 staph V8 peptides and 2 tryptic peptides were found to contain modified Ser and Thr residues.

There are a total of eight Ser and fifteen Thr residues from Ser^{282} to the C-terminus of α-agglutinin$_{20\text{-}351}$ and all of them were modified. Modification of five Ser and five Thr residues (Ser^{282}, Ser^{331}, Ser^{334}, Ser^{335}, and Ser^{338}; Thr^{329}, Thr^{339}, Thr^{340}, Thr^{341}, and Thr^{342}) were detected in both staph V8 and tryptic peptides (Table VI), and modifications at additional three Ser and ten Thr residues (Ser^{314}, Ser^{346}, and Ser^{350}; Thr^{289}, Thr^{299}, Thr^{303}, Thr^{307}, Thr^{308}, Thr^{311}, Thr^{315}, Thr^{316}, Thr^{345}, and Thr^{349}) were identified by sequencing of tryptic peptides alone. These results indicate that all of the Ser and Thr residues between positions 282-351 of α-agglutinin were modified.

Confirmation of O-glycans with Con A. O-linked carbohydrates in yeast consist of one to four α-linked mannoses and therefore interact with Con A. To examine if O-

TABLE VII Identified N- and O-linked glycosylation sites in tryptic and Staph V8 peptides of α-agglutinin$_{20\text{-}351}$.

	Staph V8 Digests					Tryptic Digests	
Residue	*241-249*	*279-295*	*296-318*	*319-337*	*324-343*	*269-288*	*326-351*
1	W	A	F	F	G	L	N
2	W	L	Q	I	R	Y	L
3	F	Q	Y	V	N	D	G
4	P	Xs	Xt	Y	L	G	Xt
5	Q	L	Xc	Q	G	E	A
6	S	P	L	G	Xt	M	Xs
7	Y	A	D	R	A	L	A
8	Xn	N	Xt	N	Xs	W	K
9	D	V	I	L	A	V	Xs
10		Xt	A	G	K	N	Xs
11		I	Xn	Xt	Xs	A	F
12		D	Xt	A	Xs	L	I
13		H	Xt	Xs	F	Q	Xs
14		A	Y	A	I	Xs	Xt
15		L	A	K	Xs	L	Xt
16		E	Xt	Xs	Xt	P	Xt
17			Q	Xs	Xt	A	Xt
18			F	F	Xt	N	D
19			Xs	I	Xt	V	L
20			Xt		D	N	Xt
21			Xt				Xs
22							I
23							N
24							Xt
25							Xs
26							A

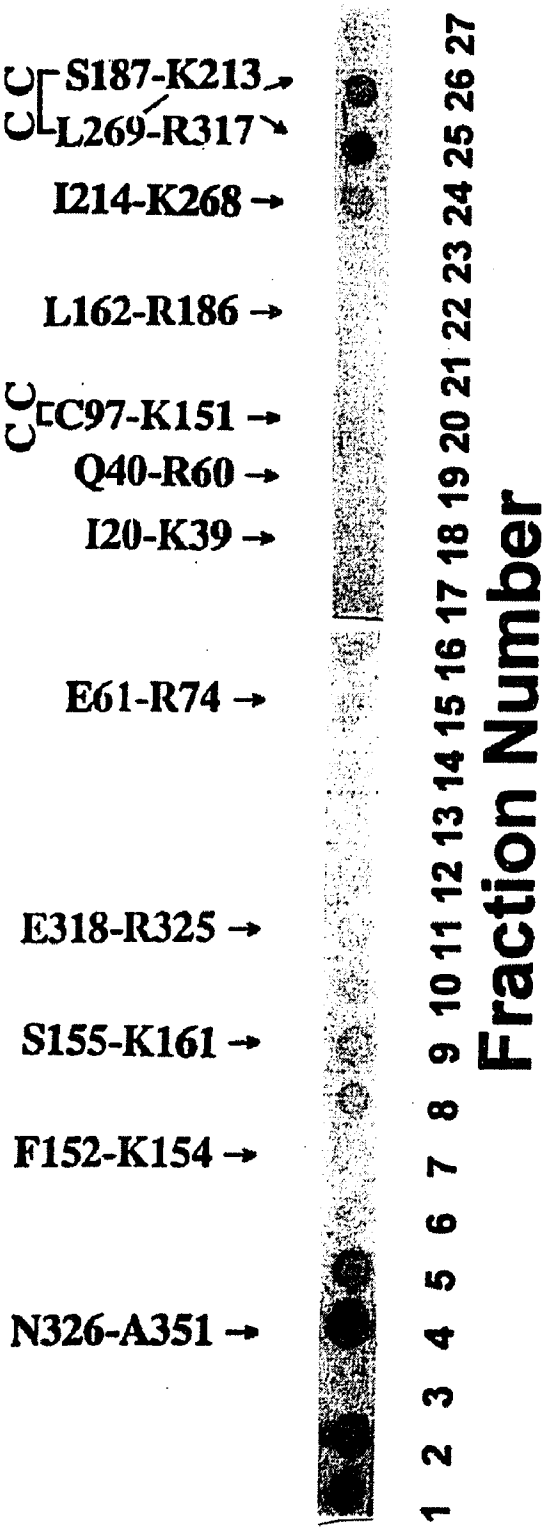
linked glycosylations were responsible for the failure to detect Ser and Thr residues, peroxidase conjugated ConA was used to probe peptides from the nonreduced tryptic digest. Fig.17 is the dot-blot analysis of tryptic fractions of HPLC fractions of nonreduced digest. Five peptides containing modified Ser or Thr residues reacted positively with ConA (fractions 4, 5, 24, 25, and 26 of Fig.14A). These peptides were those containing modified Ser and Thr residues listed in Table VII. Peptide fractions that chromatographed near the glycopeptides showed some reactivity, probably due to contamination. As the dot-blot experiment cannot specify particular O-glycosylated Ser or Thr residues in these peptides, we cannot conclude that how many Ser or Thr residues are modified in these peptides.

Identification of N-linked glycosylation sites

Endo H cleaves between the two glcNAc residues of N-linked oligosaccharides, leaving one glcNAc attached to Asn. The modified Asn residue is not detectable by the sequencer, and therefore provides an indirect assay for N-glycosylation.

There are six potential N-linked glycosylation sites with Asn-Xaa-Ser/Thr sequence in α-agglutinin$_{20\text{-}351}$. Among them, the glycosylation status of three potential sites were examined by peptide sequencing. Table VII illustrates that Asn^{248}-Asp^{249}-Thr^{250} and Asn^{306}-Thr^{307}-Thr^{308} were N-glycosylated in staph V8 peptides. Asn^{348}-Thr^{349}-Ser^{350}, however, was found not to be N-glycosylated in a tryptic peptide.

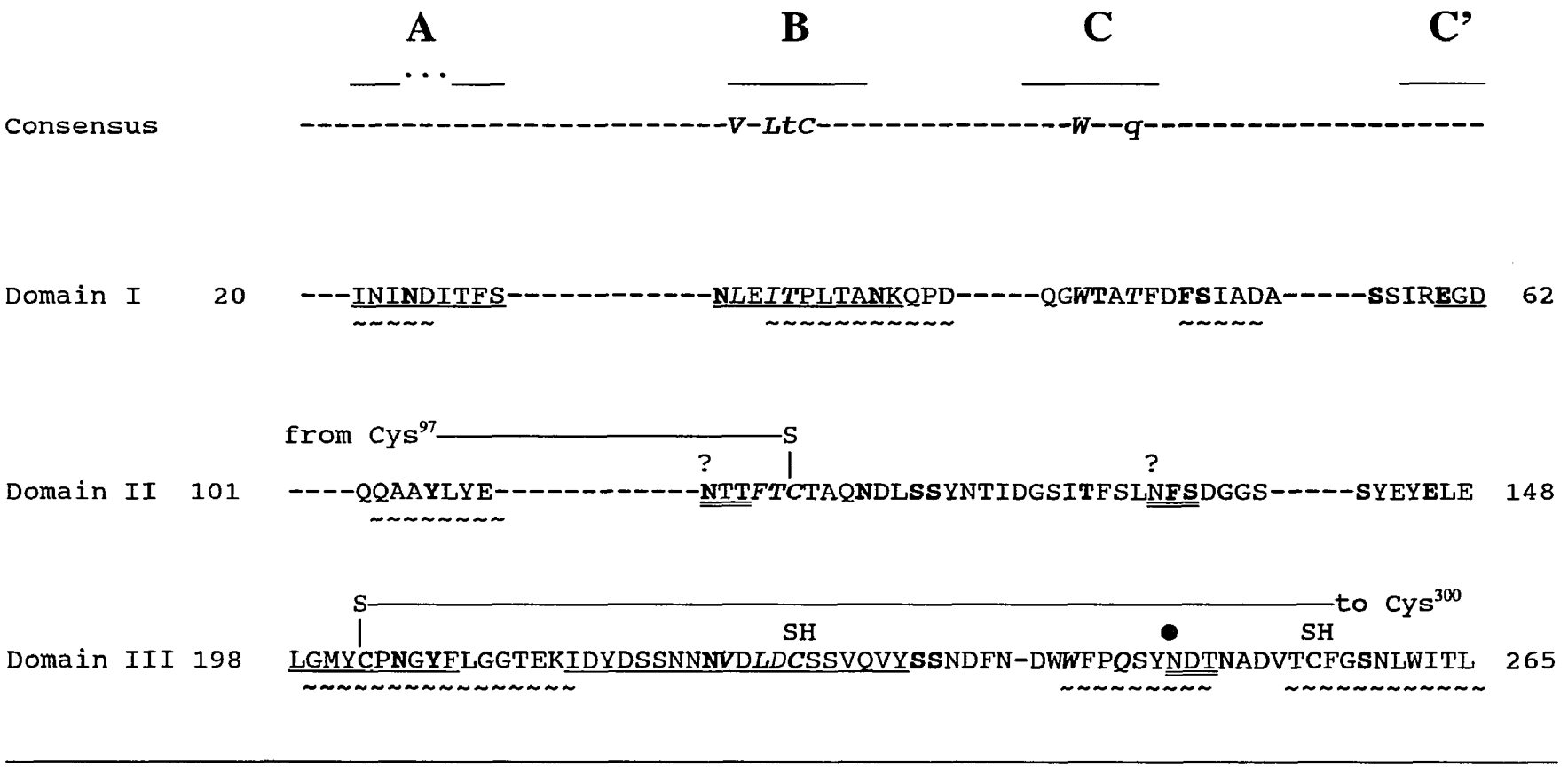
Figure 17. Con A binding of tryptic peptides of α-agglutinin$_{20\text{-}351}$. HPLC fractions from a tryptic digest under non-reduced conditions (Fig.14) were lyophilized to dryness, reconstituted in 0.1% SDS in 0.5 M sodium phosphate pH 8.0 and spotted onto an Immobilon-AV membrane. The membranes were incubated with Con A-conjugated peroxidase, and visualized with 4-chloronaphthol.

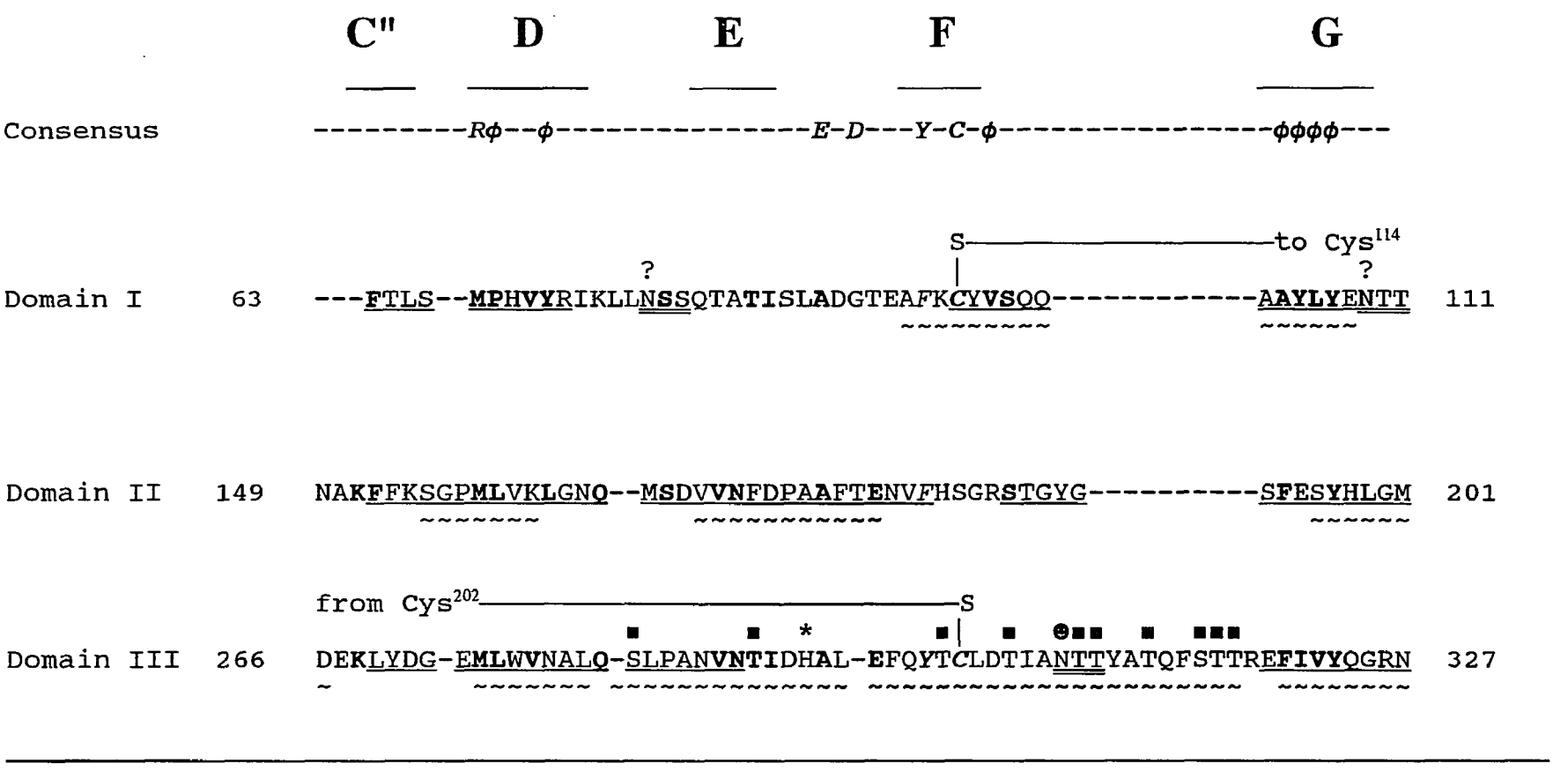

CC S187-K213
L269-R317
I214-K268
L162-R186
CC C97-K151
Q40-R60
I20-K39
E61-R74
E318-R325
S155-K161
F152-K154
N326-A351
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27
Fraction Number

Summary

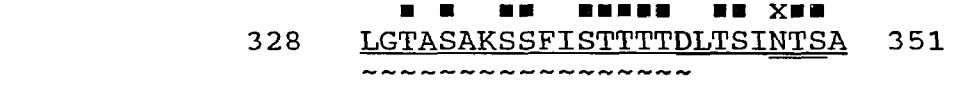
Peptide mapping of α-agglutinin$_{20\text{-}351}$ reveals important structural features of the binding region. Sequenced α-agglutinin$_{20\text{-}351}$ peptides in the proposed 3 IgV domains alignment are summarized (Fig.18). α-agglutinin$_{20\text{-}351}$ contains two disulfide bonds. Cys^{97} and Cys^{114} form an interdomain disulfide bond between the C-terminus of domain I and the N-terminus of domain II. Cys^{202} to Cys^{300} is an intradomain disulfide bond. Cys^{227} and Cys^{256} are two excess cysteines in domain III that contain free sulfhydryl groups. Peptide mapping also identified a number of O-glycosylation sites from Ser^{282} (D strand of domain III) to the C-terminus. All Ser and Thr residues in this segment are modified. Finally, peptide mapping also revealed two N-linked glycosylation sites at Asn^{248} (Asn^{248}-Asp^{249}-Thr^{250}) and Asn^{306} (Asn^{306}-Thr^{307}-Thr^{308}). The C-terminal potential site (Asn^{348}-Thr^{349}-Ser^{350}), however, is not glycosylated.

Figure 18. Summary of sequenced α-agglutinin$_{20\text{-}351}$ peptides in proposed three IgV domain alignments. Regions sequenced with trypsin and staph V8 are underlined with solid or wave-shape respectively. Sulfhydryl groups are labelled with SH and disulfide bonds are marked with S-S. Potential N-linked glycosylation sites are double underlined, among which ● stands for identified N-linked glycosylation sites, x stands for unused potential N-linked glycosylation site, and ? stands for site with unknown N-glycosylation status. Identified O-linked glycosylation sites are marked with ■. The known critical His^{292} is marked with *. These sequenced peptides are aligned in ß-strands typical of IgV domains and labelled by the standard convention. Alignments are based on ß-strand prediction and conservation of residue type, with all gaps in non-conserved regions of the consensus (data not shown). Identical residues are bold face, and residues identical to consensus are boldfaced and italicized. Residues similar to consensus are italicized only. ϕ stands for consensus hydrophobic residues.


A
B
C
C'
Consensus
Domain I 20
62
Domain II 101
148
from Cys97
to Cys300
SH
SH
Domain III 198
265



Post Domain III



Chapter IV

Structure and Binding Regions of the N-terminal Half of α-agglutinin

Introduction

The N-terminal half of α-agglutinin consists of three domains with internal homologies and IgV domain identity (Chapter II). These three contingent domains, therefore, have been proposed to contain the Ig-like fold formed by anti-parallel ß-sheet structure. This region contains two disulfide bonds, an intradomain disulfide bond (Cys^{202}-Cys^{300}) in domain III and an interdomain disulfide bond (Cys^{97}-Cys^{114}) between domain I and II (Chapter III). The disulfide bond in domain III could stabilize the interaction of the anti-parallel ß-sheets of the Ig-fold as in other members of Ig-superfamily (Williams and Barclay 1988). The domain I-II interdomain disulfide bond is likely to bring independently folded structures together to form higher level structures that may be functionally important.

Circular dichroism (CD) spectroscopy is sensitive to secondary structure and thus provides a tool to assess the structural features of α-agglutinin. The three domain model was tested on the purified α-$agglutinin_{20\text{-}351}$ as well as a proteolytic fragment that contains Ig-like domain III, for the anti-parallel ß-sheet structures. The role of disulfide bonds in maintaining the conformation of α-$agglutinin_{20\text{-}351}$ was investigated. The stability of the secondary structure to extreme pH treatments was examined.

Residues between 200 to 300 of α-agglutinin contains structure essential for binding activity (Wojciechowicz and Lipke 1989; Wojciechowicz et al., 1993). His^{292} is essential for binding activity (Cappellaro et al., 1991). Biological assays in conjunction with structural studies were carried out to determine role of the three domains in adhesion to the binding fragment of a-agglutinin.

Materials and Methods

Treatment of α-agglutinin$_{20\text{-}351}$ at different pH. Purified α-agglutinin$_{20\text{-}351}$ was incubated in 20 mM Tris pH 8.5, or 10 mM sodium acetate at pH 3.0 at room temperature for 30 minutes. Aliquots of each sample were then used for circular dichroism analysis and agglutination assay. To reconstitute α-agglutinin$_{20\text{-}351}$ after pH 8.5 treatment, the solution was washed away by centrifugation in Microcon tubes (Amicon) and α-agglutinin$_{20\text{-}351}$ was reconstituted in 10 mM sodium acetate pH 5.5.

Reduction of disulfide bonds. Purified α-agglutinin$_{20\text{-}351}$ and α-agglutinin$_{155\text{-}351}$ were treated with 70 mM dithiothreitol in 10 mM sodium phosphate buffer, pH 7.0 for 30 min at 37°C. The dithiothreitol containing solution was then washed by centrifugation in Microcon tubes. Treated α-agglutinin$_{20\text{-}351}$ and α-agglutinin$_{155\text{-}351}$ were reconstituted in 100 mM sodium acetate pH 5.5 for agglutination assay and circular dichroism.

Circular dichroism and structure analysis. Cells of 0.1 cm path length were used for far UV spectra, with typically 5 spectra being accumulated, averaged, and baseline-corrected on an AVIV circular dichroism spectrometer Model 60 DS (Lakewood, NJ) interfaced to an IBM personal computer. All spectra were acquired at 25°C. For conversion to mean residue ellipticity, a mean residue weight of 111.77 was used. The program PROSEC, which was derived from spectra studies by Yang et al., 1986, was used to analyze secondary structure distribution from recorded CD data.

Other methods. Proteins on SDS-PAGE gels were visualized by staining with Coomassie blue or Silver Staining Plus kit (Bio-Rad). Protein concentrations were determined by bicinchoninic protein assay method (Pierce) using bovine serum albumin, fraction V as standard. All other methods were described in Chapter III.

RESULTS

The secondary structure profile of α-agglutinin$_{20\text{-}351}$

The circular dichroism spectrum of α-agglutinin$_{20\text{-}351}$ (Fig.19) showed a typical ß-sheet structure profile. The characteristic negative band at 217 nm of the CD spectrum of the magnitude demonstrated the presence of a substantial amounts of ß-sheet (Brahms and Brahms, 1980). The absence of the intense negative peaks at either 208 or 222 nm, which are the characteristic of α-helix, indicated very little α-helix content in α-agglutinin$_{20\text{-}351}$.

Quantitative analysis of the CD spectrum of α-agglutinin$_{20\text{-}351}$ illustrated the presence of a 6.8% α-helix, 69.4% ß-sheet, 13.2% turns and 10.5% random structure (Table VIII). The mean residue ellipticity was $[\Theta_M]_{217}=-4.5\text{x}10^3$ degrees.cm^2.dmol^{-1}. This high ß-sheet content suggests the presence of anti-parallel ß-sheets core structure, consistent with Ig domains.

Effects of pH and heat on the secondary structure

The activity of α-agglutinin is optimal between pH 4.5 and pH 6.5. α-agglutinin is inactive below pH 3 or above 8.5 (Terrance and Lipke, 1981). To characterize the effects of extreme pH on conformation of the protein, CD spectra of α-agglutinin$_{20\text{-}351}$ were taken at either pH 3.0 in sodium acetate buffer or pH 8.5 in Tris buffer.

The effect of high pH treatment. In the CD profile of α-agglutinin$_{20\text{-}351}$ at pH 8.5 (Fig.19), the amplitude of the 217 nm negative band was reduced and the bandwidth of this peak became broadened towards low wavelength in comparison to that under native condition, which indicated an increase in turns and random structure content. Quantitative analysis revealed a decrease in the ß-sheet content (from 69.4% to 40.3%)

Figure 19. Far-UV circular dichroism spectra of α-agglutinin$_{20\text{-}351}$ at pH 5.5 (—), pH 8.5 (----), and reconstitution from pH 8.5 to pH 5.5 (···). Each spectrum represents the average of 5 individual spectra taken at 1.0-nm intervals as specified in "Materials and Methods". Equivalent molar concentration of each sample were examined.

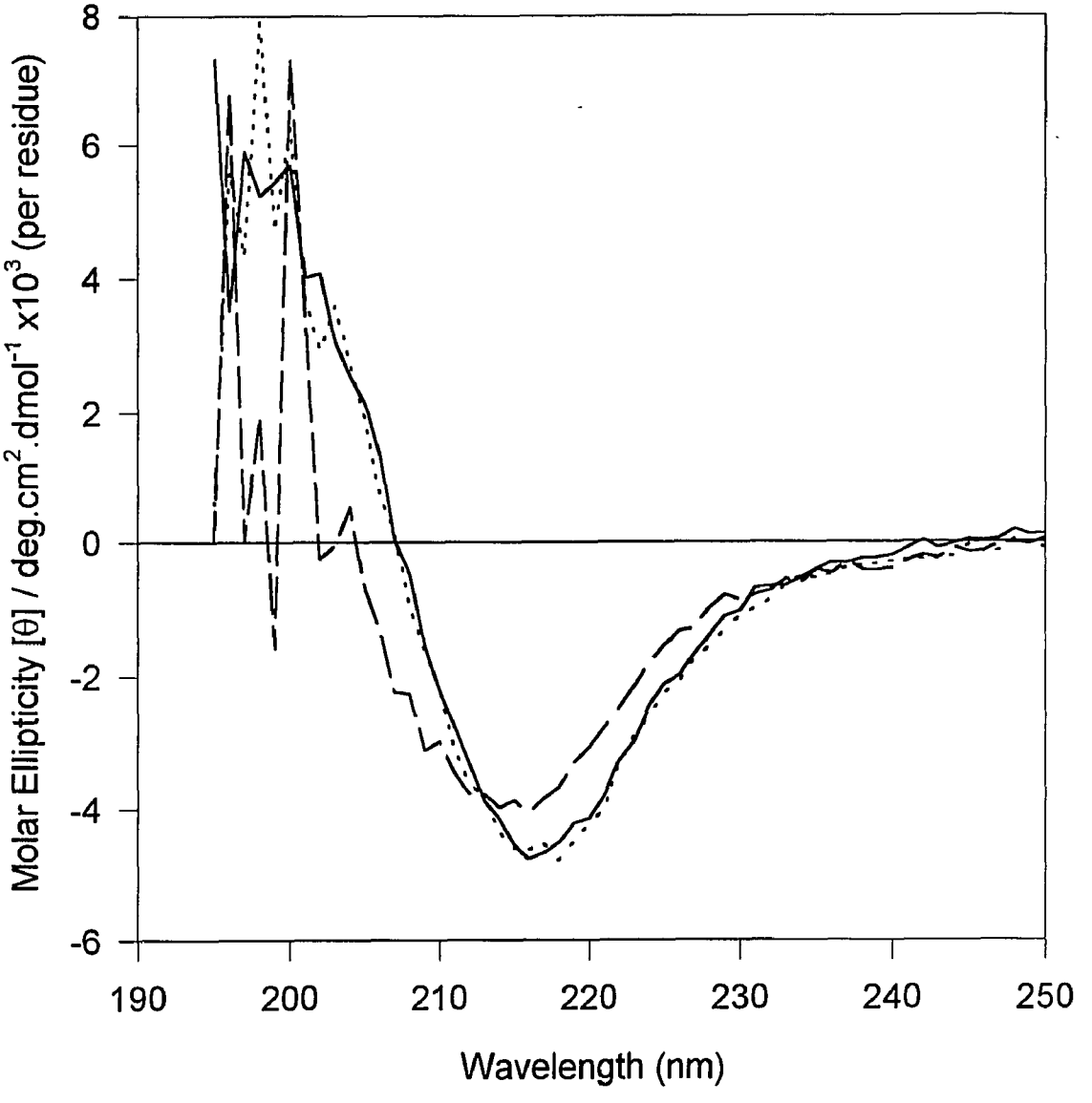

8
6
4
2
0
-2
-4
-6
190
200
210
220
230
240
250
Molar Ellipticity [θ] / deg.cm^2.dmol^{-1} x10^3 (per residue)
Wavelength (nm)

TABLE VIII Distribution of secondary structures from α-agglutinin$_{20\text{-}351}$ and α-agglutinin$_{155\text{-}351}$ CD spectra.

α-agglutinin fragment	Treatment	Structure element (%) [a,b]			
		α-helix	ß-sheet	turn	random
α-agglutinin$_{20\text{-}351}$	Native (pH 5.5)	6.8	69.4	13.2	10.5
	Arg-C digest [c]	4.0	68.8	10.8	16.4
	Tris (pH 8.5)	10.4	40.3	30.2	19.0
	Tris (pH 5.5) [c]	9.2	57.4	20.4	13.0
	Sodium acetate (pH 3.0)	5.1	77.5	9.9	7.5
	DTT treatment [c]	9.0	61.1	17.8	12.1
	Heat denatured [d]	9.0	33.9	27.4	29.7
α-agglutinin$_{155\text{-}351}$	Native	5.7	46.9	18.5	29.0
	DTT treatment [c]	8.6	35.6	24.3	31.4

[a] Percentages are expressed as summations of each structure contribution divided by 332 residues.

[b] Data for structure calculation from Figs. 19-21, 23-24 from 240 nm to 195 nm.

[c] The final solutions were reconstituted to pH 5.5 after individual treatment at pH 8.5.

[d] Heat denaturation was performed in 2 mM DTT, pH 7.0.

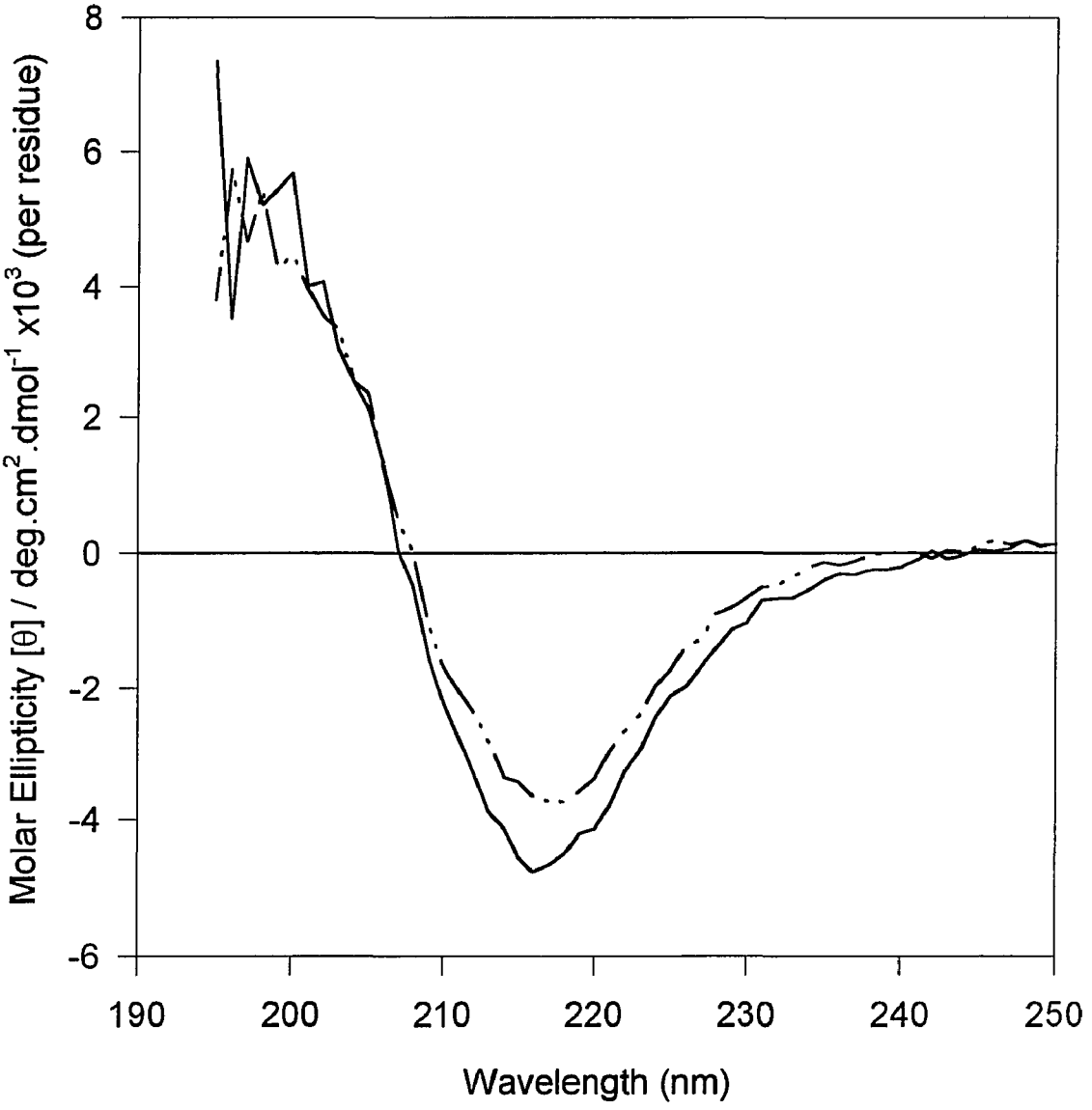
accompanied by a corresponding increase in turns (from 13.2% to 30.2%) and random structure (10.5% to 19.0%) content, respectively. The degree of such a reduction in ß-sheet content was similar to that after heat denaturation (below). There was no significant increase in α-helix content (Table VIII).

To determine if the alteration of the secondary structure and inactivation of α-agglutinin$_{20\text{-}351}$ was reversible, α-agglutinin$_{20\text{-}351}$ was reconstituted to pH 5.5 for 30 min after denaturation at pH 8.5. The CD spectrum of the pH reconstituted α-agglutinin$_{20\text{-}351}$ showed a recovery of the major CD feature of native α-agglutinin$_{20\text{-}351}$. The spectrum illustrated an increase in the ß-sheet content from 41.9% to 57.4%, with a decrease in turns and random structure content from 39.2% to 33.4% (Fig.19 and Table VIII). Therefore, this change of the secondary structure of α-agglutinin$_{20\text{-}351}$ by pH 8.5 was partially reversible. 82.7% of the original ß-sheet structure was recovered after reconstitution of the pH to 5.5. The recovery of the adhesive binding activity was also examined. 70% of the original activity was recovered. Therefore, this recovery of the binding activity correlated with restoration of the ß-sheet content.

The effect of acidic treatment. In contrast to the high pH treatment, the CD spectrum of α-agglutinin$_{20\text{-}351}$ at pH 3.0 (Fig.20) showed a typical ß-sheet profile with a slightly reduced magnitude and bandwidth of the negative peak at 217 nm in comparison to that of native α-agglutinin$_{20\text{-}351}$. There was a ß-sheet content of 77%, with almost no change in the proportion of other structural elements (Table VIII).

The effect of heat denaturation. The agglutination activity of α-agglutinin is heat labile (Terrance and Lipke, 1981). Since the heat denaturation of α-agglutinin was known to be irreversible, inactivation is expected to be the consequence of randomization of the

Figure 20. Far-UV circular dichroism spectra of α-agglutinin$_{20\text{-}351}$ at pH 5.5 (—) and pH 3.0 (-··-). Spectra taken condition as indicated in figure 19.


8
6
4
2
0
-2
-4
-6
190
200
210
220
230
240
250
Molar Ellipticity [θ] / deg.cm^2.dmol^{-1} x10^3 (per residue)
Wavelength (nm)

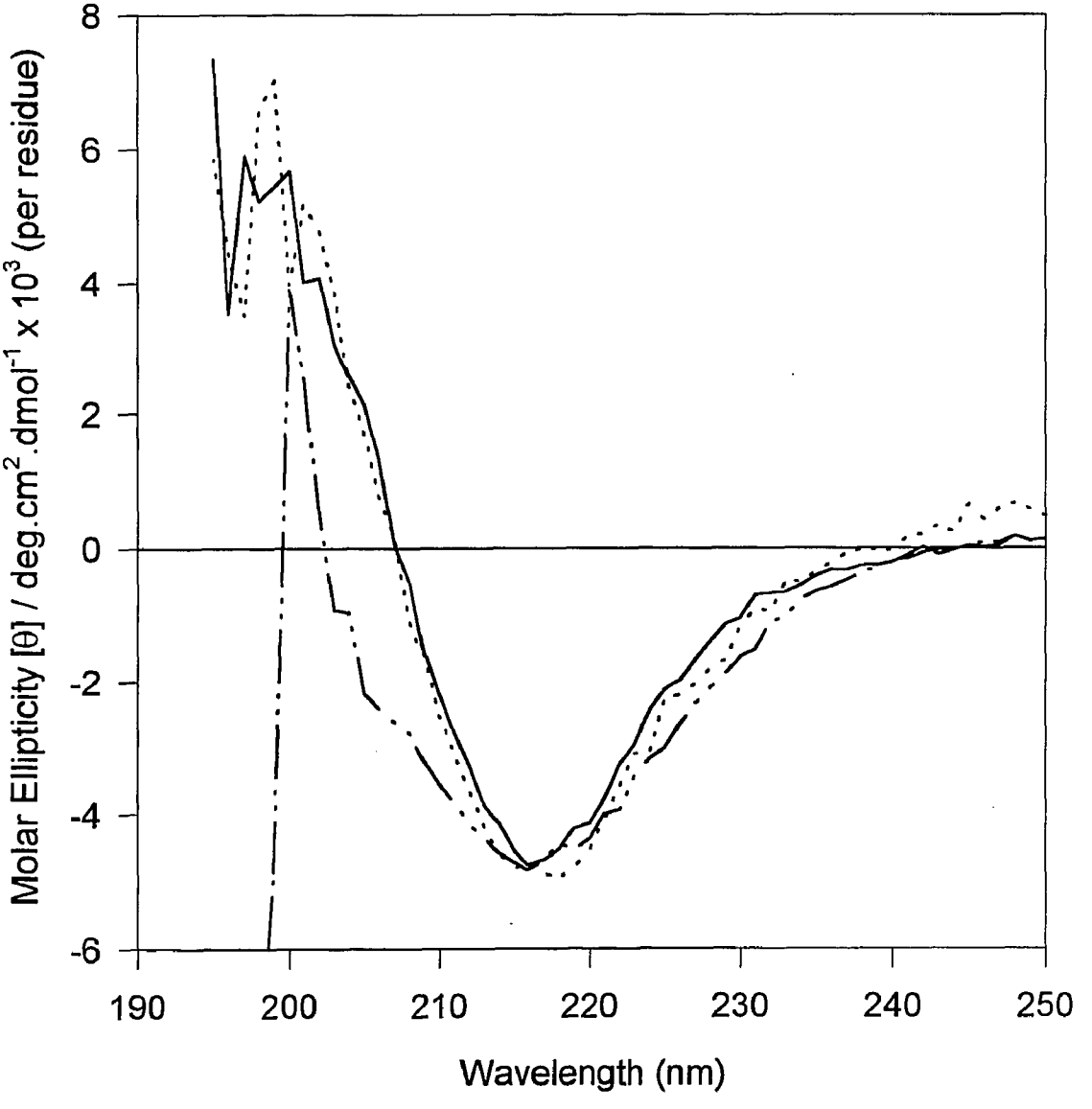
native conformation.

α-agglutinin$_{20\text{-}351}$ was heat denatured at 100°C for 5 minutes in dithiothreitol solution at pH 7.3 prior to taking a CD spectrum. The denaturation resulted in complete loss of activity and gave a different CD spectra from that of native α-agglutinin$_{20\text{-}351}$. It showed a significantly broadened negative band at 217 nm (Fig.21). A second negative peak with greater amplitude indicated the presence of a high proportion of random structure. The heat denatured α-agglutinin$_{20\text{-}351}$ contained only 33.9% ß-sheet, a decrease of more than 50% from that of native α-agglutinin$_{20\text{-}351}$. The decrease was accompanied by two-fold and three-fold increases in turns and random structure content, respectively.

Role of disulfide bonds in the conformation of α-agglutinin$_{20\text{-}351}$

α-agglutinin$_{20\text{-}351}$ contains disulfide bonds between Cys^{97}-Cys^{114} and Cys^{202}-Cys^{300} (Chapter III). Reduction of disulfide bonds leads to inactivation of α-agglutinin$_{20\text{-}351}$. α-agglutinin$_{20\text{-}351}$ was treated with 70 mM dithiothreitol at pH 7.3. Dithiothreitol was then removed, and the sample was reconstituted in pH 5.5 sodium acetate buffer. The CD profile of dithiothreitol treated α-agglutinin$_{20\text{-}351}$ was similar to that of native α-agglutinin$_{20\text{-}351}$, except that the width of the characteristic negative peak was slightly broadened (Fig.21). Dithiothreitol treated α-agglutinin$_{20\text{-}351}$ had a slightly lower ß-sheet content (61.1%) in comparison with that of native α-agglutinin$_{20\text{-}351}$ (69.4%), while the proportion of α-helix and ß-turns content were slightly increased (Table VIII).

Figure 21. Far-UV circular dichroism spectra of native (—), DTT treated (•••), and heat denatured (-•••-) α-agglutinin$_{20\text{-}351}$. Spectra taken condition as indicated in figure 19.


8
6
4
2
0
-2
-4
-6
190
200
210
220
230
240
250
Molar Ellipticity [θ] / deg.cm^2.dmol^{-1} x 10^3 (per residue)
Wavelength (nm)

The agglutinability of endoprotease Arg-C digest and α-agglutinin$_{155-351}$

Endoprotease Arg-C digestion. When purified α-agglutinin$_{20-351}$ was digested with Endoprotease Arg-C from mouse submaxillary glands, two major proteolytic fragments of 30 and 20 kDa, and one minor fragment of 16 kDa were generated (Fig.22, fraction 32). The same digestion pattern was obtained from digestions under either denaturing or nondenaturing conditions. No additional proteolytic fragments were observed. All the three fragments appeared to have a higher molecular weight on SDS-PAGE after dithiothreitol treatment and each contained one disulfide bond (Fig.12, Chapter III).

A single site digestion by Arg-C. The N-terminal sequence of each fragment was identified by electroblotting onto polyvinylidene difluoride membranes followed by microsequence analysis. Both the 16 and 20 kDa fragments were found to have the same unambiguous NH_2-terminal sequence corresponding to the NH2-terminal sequence of mature α-agglutinin. The 30 kDa fragment started after Lys-^{154} with Ser^{155}-Gly-Pro-Met-Leu-Val (Table IX). This fragment was called α-agglutinin$_{155-351}$. Therefore, instead of any of six Arginine residues in α-agglutinin$_{20-351}$, Arg-C hydrolyzed only at Lys-^{154}. Digestion of α-agglutinin$_{20-351}$ at Lys-154 was also found with endoprotease Arg-C from *Clostridium histolyticum* (data not shown) but not α-agglutinin$_{20-351}$ preparation incubated without protease. Therefore, hydrolysis of α-agglutinin$_{20-351}$ at Lys-154 was endoprotease Arg-C specific and not due to an autoproteolytic activity, or other reagents used for the digestion.

Based on the DNA sequence, the polypeptide from Ser^{155} to the C-terminus Ala^{351}

Figure 22. Gel electrophoresis of gel filtration fractions of endoprotease Arg-C-digested α-agglutinin$_{20\text{-}351}$. A digested sample was applied to a Bio-Gel P-30 column and eluted as described above. Even number fractions from 20 to 36 were electrophoresed on a 15% gel and visualized by silver staining. Molecular size markers are shown on the left.

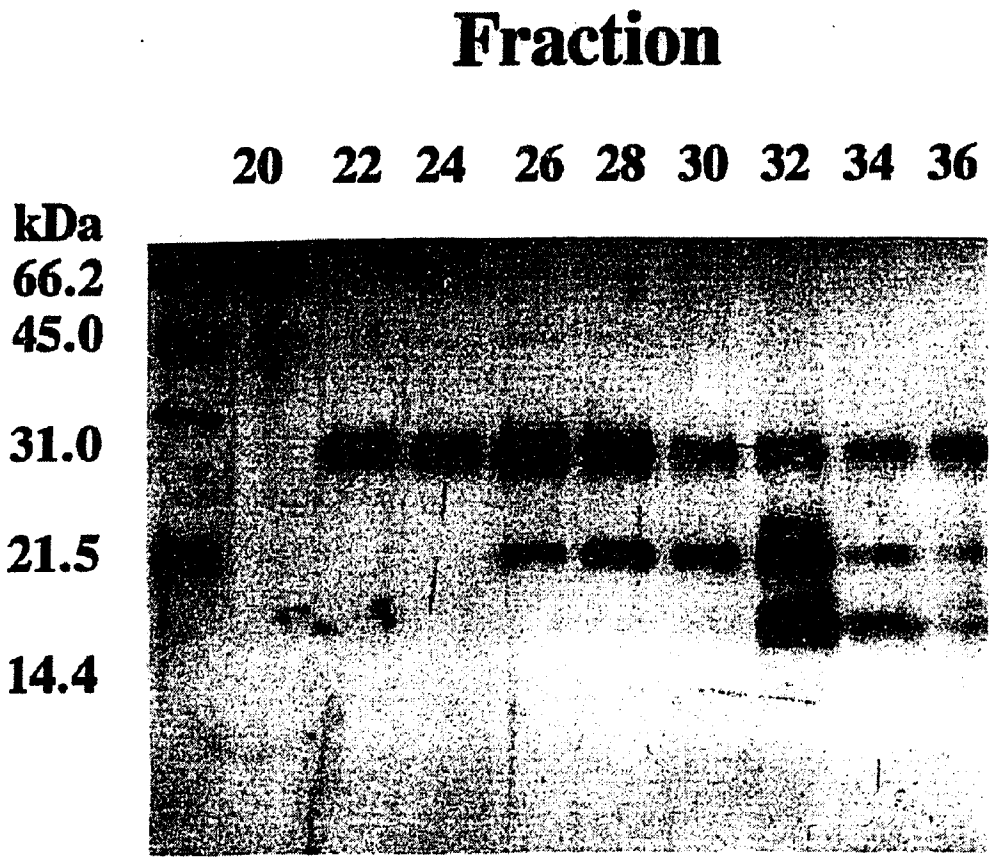

Fraction
20 22 24 26 28 30 32 34 36
kDa
66.2
45.0
31.0
21.5
14.4

TABLE IX N-terminal peptide sequence of α-agglutinin$_{20\text{-}351}$-Arg-C fragments

Arg-C fragment[a]	Sequence	Position in α-agglutinin	# Residues Sequenced
16 kDa	ININDITFSNLEITPLTANK	20 - 39	20
20 kDa	ININDITFSNLEITPLTANK	20 - 39	20
30 kDa	SGPMLVKLGNQMSDVVNFDV	155 - 174	20

[a] Endo H deglycosylated α-agglutinin$_{20\text{-}351}$ (100 μg) was cleaved with 1:25 endoprotease Arg-C in 0.1 M NH_4HCO_3, pH 7.8 for 18 hours at 25°C. The digest was subject to SDS-PAGE and electroblotted to polyvinylidene difluoride membranes followed by N-terminal microsequence analysis as specified under "Materials and Methods".

would have a molecular mass of 21,989 daltons. Therefore, the molecular mass of α-agglutinin$_{155\text{-}351}$ is large enough to cover the entire C-terminal region of α-agglutinin$_{20\text{-}351}$. There is no cleavage in domain III, otherwise a 4 kDa fragment from Arg^{317}-Ala^{351} would be observed after digestion. The extra 8 kDa molecular mass of agglutinin$_{155\text{-}351}$ may be attributed to the presence of multiple O-glycosylations in the C-terminal region of α-agglutinin$_{20\text{-}351}$ (Chapter III). As The N-terminal α-agglutinin$_{20\text{-}154}$ would have a predicted molecular mass of 15,119 daltons. The 16 and 20 kDa N-terminal fragments are large enough to contain this region.

Agglutinability of Arg-C digested α-agglutinin$_{20\text{-}351}$ and α-agglutinin$_{155\text{-}351}$. To examine if any of the endoprotease Arg-C digested fragments retained agglutination activity, the α-agglutinin$_{20\text{-}351}$-endoprotease Arg-C digests were reconstituted with sodium acetate to pH 5.5 and assayed for activity. Table X shows that α-agglutinin$_{20\text{-}351}$ digests with endoprotease Arg-C retained no measurable agglutination activity at concentration up to 10 μg. The agglutination activity of digested α-agglutinin$_{20\text{-}351}$ was less than 5 X 10^{-4} that of intact α-agglutinin$_{20\text{-}351}$.

The α-agglutinin$_{155\text{-}351}$ fragment of α-agglutinin$_{20\text{-}351}$-Arg-C digest contained the entire domain III as indicated above and was purified on a Bio-Gel P-30 column (Fig.22, fractions 22-24). Purified α-agglutinin$_{155\text{-}351}$ fragment was also assayed and no binding activity was found.

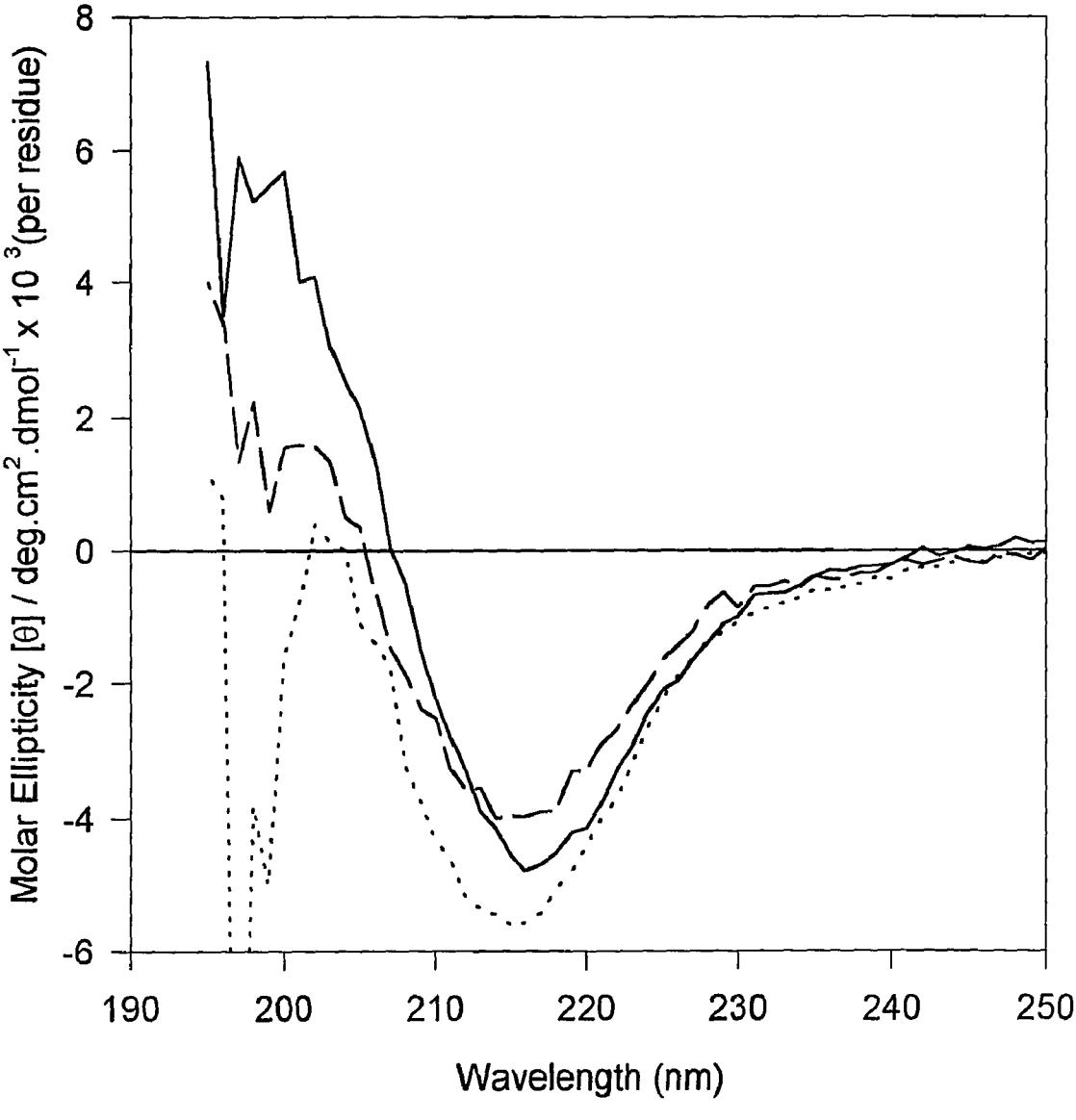
The secondary structure of Arg-C digested α-agglutinin$_{20\text{-}351}$

As α-agglutinin$_{20\text{-}351}$ can be inactivated by endoprotease Arg-C digestion at Lys^{154}, the effect of the digestion on the structure of α-agglutinin$_{20\text{-}351}$ was examined. The digestion product was reconstituted to pH 5.5 for 30 min before taking the CD spectrum.

TABLE X Agglutination activity of α-agglutinin$_{20\text{-}351}$ and proteolysis products

α-agglutinin$_{20\text{-}351}$ sample	Lowest active concentration (ng/ml)
Native α-agglutinin$_{20\text{-}351}$	10^{-1}
Endoprotease Arg-C digests	$> 10^{4}$
Purified α-agglutinin$_{155\text{-}351}$	$> 10^{4}$
Sodium acetate, 100 mM, pH 7.8	0

Figure 23. Far-UV circular dichroism spectra of native (—) and endoprotease Arg-C digested (----) α-agglutinin$_{20\text{-}351}$, and α-agglutinin$_{155\text{-}351}$ (···). Spectra taken under conditions indicated in figure 19.


8
6
4
2
0
-2
-4
-6
190
200
210
220
230
240
250
Molar Ellipticity [θ] / deg.cm^2.dmol^{-1} x 10^3(per residue)
Wavelength (nm)

The spectrum of the digest (Fig.23) was very similar to that of native α-agglutinin$_{20\text{-}351}$, in both the negative peak position at 217 nm and the corresponding bandwidth. Again, the spectrum lacked the local maxima at 208 and 222 nm, which is the characteristic of α-helix. The mean residue ellipticity was $[\Theta_M]_{217}$=-4.2x10^3 degrees.cm^2.dmol^{-1}. Quantitative analysis of the CD spectrum implied that the secondary structural profile in Arg-C digested α-agglutinin$_{20\text{-}351}$ was similar to native α-agglutinin$_{20\text{-}351}$, with 68.8% ß-sheet, and a slightly higher random structure content (Table VIII).

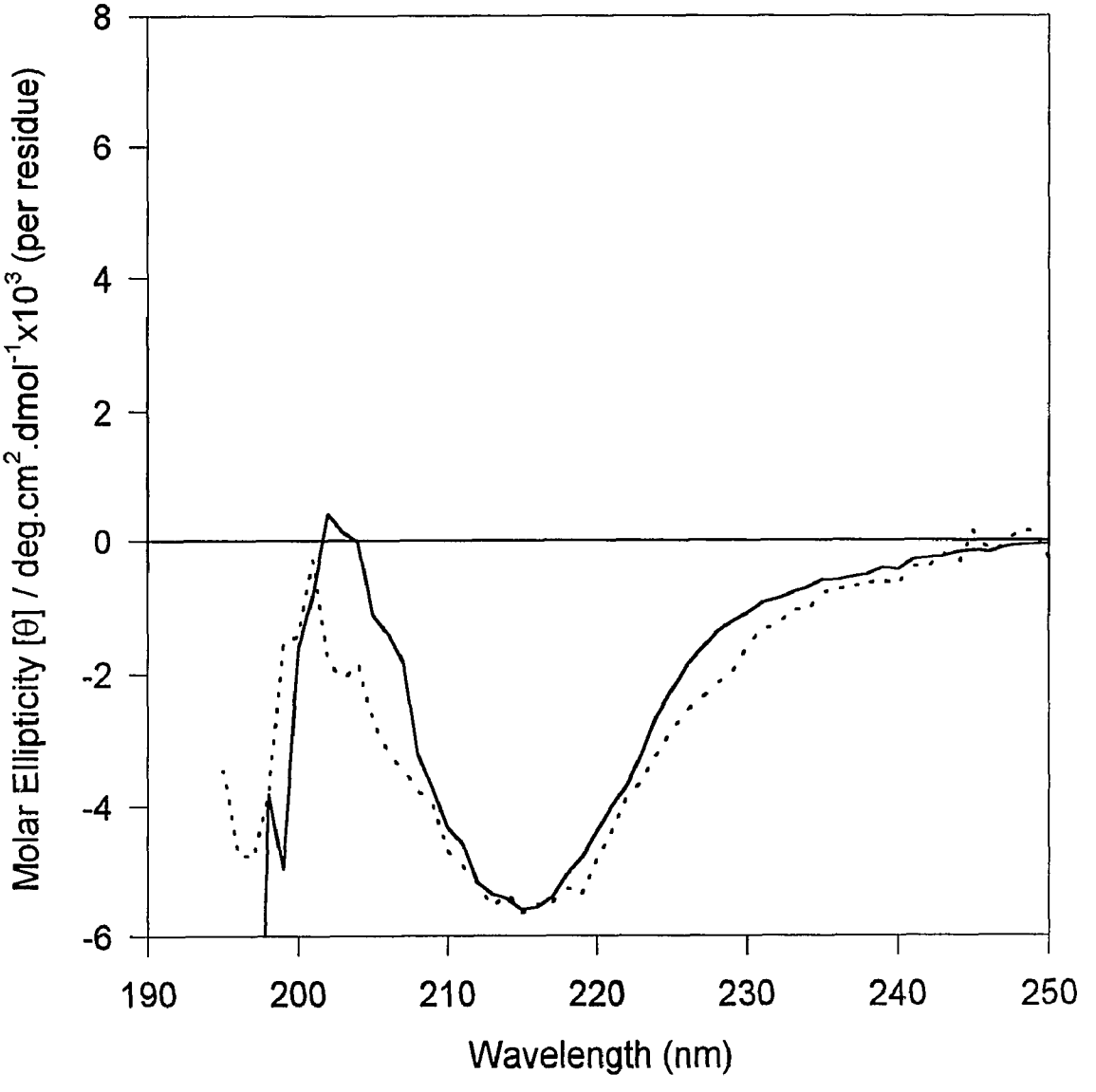
This CD profile indicated that the single site digestion at Lys-154 of α-agglutinin$_{20\text{-}351}$ did not greatly alter the native secondary structure of the protein. Therefore, the inactivation of the binding activity is not due to gross structural change during the Arg-C digestion.

The secondary structure of native and dithiothreitol reduced α-agglutinin$_{155\text{-}351}$

Circular dichroism of α-agglutinin$_{155\text{-}351}$. α-agglutinin$_{155\text{-}351}$ contains one-half of the proposed second domain and the entire third domain of the N-terminal half of α-agglutinin. The CD spectrum of purified α-agglutinin$_{155\text{-}351}$ showed a typical ß-sheet rich profile with an enhanced magnitude of the negative peak at 217 nm (Fig.24). The bandwidth of this negative peak became broadened towards low wavelength. A second negative peak below 210 nm indicated the presence of a relatively high proportion of random structure. The mean residue ellipticity was $[\Theta_M]_{217}$=-5.8x10^3 degrees.cm^2.dmol^{-1}. Analysis of the CD data illustrated that α-agglutinin$_{155\text{-}351}$ contains 46.9% ß-sheet, which is a 33% decrease from that of intact α-agglutinin$_{20\text{-}351}$, accompanied by a two fold increase in random structure content from α-agglutinin$_{20\text{-}351}$ (Fig.24). This change suggested loss of native structure during purification of the fragment.

Circular dichroism of dithiothreitol reduced α-agglutinin$_{155-351}$. After α-agglutinin$_{155-351}$ was treated with dithiothreitol, a significant bandwidth broadening of the negative peak at 217 nm was observed in comparison to that under nonreduced condition (Fig.24) and the CD profile was like that of heat denatured α-agglutinin$_{20-351}$ (Fig.21), which indicated an increase in the random structures. Quantitative analysis of the CD spectrum illustrated a reduction of ß-sheet content to 35.6% from 46.9% of native α-agglutinin$_{20-351}$, which is a 24% decrease (Table VIII). The distribution of each of the secondary structures was very similar to that of denatured α-agglutinin$_{20-351}$, with a predominance of turns and random structure (55.8%).

Figure 24. Far-UV circular dichroism spectra of native (—) and DTT treated (···) α-agglutinin$_{155\text{-}351}$. Spectra taken under conditions indicated in figure 19.


8
6
4
2
0
-2
-4
-6
190
200
210
220
230
240
250
Wavelength (nm)
Molar Ellipticity [θ] / deg.cm^2.dmol^{-1}x10^3 (per residue)

Summary

Biophysical evidence shows the presence of a high ß-sheet content in the region of α-agglutinin between amino acid residues 20 to 351. This data is consistent with the proposal that the N-terminal half of α-agglutinin contains 3 Ig-like domains. Such a ß-sheet-rich structure is also supported by evidence from extreme pH treatment and is compatible with the structure of a proteolytic fragment and the role of disulfide bonds in Ig-like domains. This high ß-sheet structure is important for the binding activity. Treatment with either basic pH or reducing reagent that decreases the ß-sheet content causes inactivation. We have also demonstrated that domain III of α-agglutinin is not sufficient for binding to a-agglutinin. Multidomain cooperation is, therefore, likely to contribute to agglutination activity.

Chapter V

Discussion and Conclusion

Part I: Functional diversity and evolutionary conservation in the Ig-superfamily

Diversified function in cell surface recognition

Most Ig-superfamily molecules are involved in cell-cell interactions. The immunity function, as the original identity of this class, is only one of the highly diversified functions of this superfamily. The functional diversity of this superfamily has been expanded to include cell surface recognition for the movement and differentiation of many types of animal tissues and cells (Williams and Barclay, 1988; Hunkapiller and Hood, 1989).

The interactions of Ig-superfamily molecules with other Ig-superfamily molecules or molecules of other superfamilies can be either homophilic or heterophilic. The homophilic interactions of both N-CAM and L1 regulate neurite outgrowth (Hoffman et al., 1984; Moos et al., 1988), whereas the homophilic interaction of Po mediates myelination in peripheral nervous system (Kuhn et al., 1991). The heterophilic interaction of ICAM-1 to LFA-1 (Staunton et al., 1988) and CD2 to LFA-3 (Recny et al., 1990) facilitate cellular immunity and T-lymphocyte division. Though N-CAM on different neurons homophilically interacts with each other, it also interacts with L1 on the surface of one cell during neurite outgrowth (Horstkorte et al., 1993). The sites for the two types of interaction on N-CAM are different. The N-terminal four IgC2 domains of N-CAM are involved in the homophilic interaction, while the fifth domain contains a C-type lectin and is responsible for interaction with oligomannosidic glycans on L1 (Horsktorte et al., 1993).

Other Ig-superfamily proteins have more than one type of ligand for adhesive binding. CD4 and CD8 are coreceptors on T-helper cells and coordinately engage MHC

class II molecules (Miceli et al., 1991). Human CD4 also binds to HIV envelope glycoprotein gp120 during viral infection (Pollard 1991). ICAM-1 recognizes LFA-1 on T-lymphocytes, and also serves as a major surface receptor for rhinoviruses (Staunton et al., 1989). In both cases, the viral binding sites on CD4 and ICAM-1 are distinct from their adhesive sites for cellular immunity.

The cell surface recognition mediated by this superfamily is not only restricted to protein-protein interactions, but includes protein-carbohydrate interactions. Members of the sialoadhesion family, which are also members of Ig-superfamily, including CD22, MAG, and sialoadhesin are sialic acid-binding proteins with different specificities for sialylated glycans (Kelm et al., 1994). Among them, CD22 mediates sialic acid-dependent adhesion of cells of haemopoietic origin including lymphocytes, monocytes, neutrophils and erythrocytes (Stamenkovic and Seed, 1990), sialoadhesin is a macrophage receptor (Crocker et al., 1991; Van den Berg et al., 1992), and MAG mediates myelination in central nervous system (Owens and Bunge, 1989; Trapp, 1990). As described, all these molecules play a recognition role at the cell surface.

Conservation of Ig fold in the superfamily

The evolutionary and structural basis of the Ig-superfamily rests on proposed gene duplication and divergence of domains of about 100 amino acids. These sequences show a characteristic anti-parallel ß-sheets folding pattern. The Ig fold is stabilized by a hydrophobic interior and often by a characteristic disulfide bond between the two sheets.

The Ig fold structure provides a stable platform for a diversity of sequences. The Ig-superfamily molecules vary the amino acids that are exposed on the external faces of the ß sheets and in the loops of sequence connecting the ß-strands. Conserved sequence

patterns are seen in the hydrophobic face of the ß strands (Williams and Barclay, 1988; Williams et al., 1989).

It has been argued that the functional recognition events observed from a diversity of recognition interactions does not obviously lead to conservation of the Ig-related sequence patterns. It could be that a requirement for resistance to proteolysis leads to conservation of the Ig fold and, thereafter, a diversity of recognition determinants has arisen from this stable folding structure (Williams, 1987; Williams, et al., 1989). The Ig fold is a structure type that may allow a greater diversity of sequence than any of the other structural motifs in protein superfamilies involved in cell-surface events, based on the growing number of new members of this superfamily.

Individual members of this superfamily consist of varying numbers of Ig-related domains. The number of Ig domains varies from one domain in Thy-1 and Po, to seventeen domains in sialadhesin (Williams and Barclay, 1988; Kelm et al., 1994). This variation in multiple Ig domains brings further diversity to this superfamily.

Evolutionary conservation

The identification of Ig-like sequence and folding structure in prokaryotic chaperones and cellulases (Holmgren *et al.*, 1992; Juy et al., 1992) implied that the Ig-like fold could originate earlier than the evolutionary split between the eukaryotes and the prokaryotes (Knoll 1992). Though these prokaryotic Ig-like molecules are not involved in cell-cell recognition, chaperone molecules bind to immature polypeptides to mediate the protein folding process.

α-agglutinin is a *S. cerevisiae* cell adhesion molecule. During the mating process, it mediates recognition of a cells, by interacting with a-agglutinin (reviewed in Lipke and

Kurjan, 1992). Different from the conserved Ig-like features in prokaryotic molecules, the presence of apparent IgV domain identity in three domains of the binding region of α-agglutinin, based on the sequence alignments, suggests that the Ig-like fold structure was used as a basis for cell-cell recognition prior to the divergence of fungi and animals. This thesis work provided structural evidence for the presence of characteristic anti-parallel ß-sheets structure of an Ig-fold.

Part II: Inclusion of α-agglutinin in the Ig-superfamily

Sequence alignment showing three IgV domains in binding region of α-agglutinin

The N-terminal half of α-agglutinin (up to amino acid 350) is sufficient for the binding activity (Wojciechowicz et al., 1993). This half of α-agglutinin contains three domains: residues 20-104 as domain I, 105-199 as domain II, with fusion of domains I and II at the adjacent ends, and 200-327 as domain III. The first two domains shows significant internal sequence homology, with a 26% sequence identity and 51% similarity. Domain II also shows a 17% sequence identity and 44% similarity to domain III. The pattern of such a sequence similarity also matches the IgV domain consensus, which is based on a combination of sequence and predicted secondary structure and shows that domains II and III share key structure similarities (Lipke et al., in preparation). These alignment results suggest that the three domains are evolutionarily related in both sequence and structure.

Domain III consists of 135 amino acids and shows strongest homology to the consensus sequence of variable-type Ig-superfamily among the three proposed IgV domain (Wojciechowicz et al., 1993; Lipke et al., in preparation). The sequence alignment in this work elucidated that domain I and II also show apparent IgV domain identity. Each domain can be delineated into the typical nine ß-strands of IgV domains, based on sequence and predicted secondary structure alignment (Fig.8; Lipke, et al, in preparation). Consensus sequence patterns are present in the predicted B, C, and F strands of both domains, especially in B and F strands (Fig.8 and Table I). An alternating hydrophobic residue pattern, which is typical for an Ig-folding structure, is also seen in strands A, B, E, and G, F, C of these two domains that presumably form

the core of the Ig fold. Fusion of two domains occurs at strand G of domain I and strand A of domain II, which result in a tight association of both domains. Therefore, the binding region of α-agglutinin contains 3 domains in tandem, having IgV domain similarity. As many members of Ig-superfamily often consist of multiple Ig-like domains for cell surface recognition (Williams and Barclay, 1988; Staunton et al., 1990), the IgV domains in α-agglutinin may together mediate cell adhesion.

Secondary structure support for inclusion of α-agglutinin in the Ig-superfamily

In the alignment of the N-terminal 3 domains of α-agglutinin to the IgV domain consensus, the predicted ß-sheet potentials have been used as in many other cases. The ß-sheet potential alignment is useful to test whether a domain assignment, based on the sequence alignment, is improbable or not from a structural point of view (Zvelebil et al., 1987). This method remains imprecise. The sequence alignment, therefore, needs to be buttressed by structural proof. Circular dichroism is sensitive to secondary structure. It has been used convincingly to examine the presence of pure ß-sheet structure for members of Ig-superfamily, including Thy-1 (Campbell et al., 1979). Thus it should provide a logical step in assessment of the structural features of the corresponding regions of α-agglutinin.

A predominant ß-sheets content in α-agglutinin$_{20\text{-}351}$ and the similarity in CD profile to Ig-superfamily molecules. The CD spectrum of α-agglutinin$_{20\text{-}351}$ demonstrated the presence of a characteristic anti-parallel ß-sheets structure with little of the intense optical activity of α-helical structures. The magnitude of the negative peak at 217 nm for ß-sheet was greater in α-agglutinin$_{20\text{-}351}$ than in the spectrum of human myeloma IgD-Ha and was in the range of that for many Ig molecules or their proteolytic fragments (Cathou

and Dorrington, 1975; Jefferis et al., 1978).

The CD profile of α-agglutinin$_{20\text{-}351}$ is similar to that of Thy-1 and CD2 (Recny et al., 1990), both of which are known to be Ig-superfamily proteins. The mean residue ellipticity at 217 nm for α-agglutinin$_{20\text{-}351}$, Thy-1, and CD2 are -4.7×10^3, -4.8×10^3, and -6.6×10^3 degrees.cm^2.dmol^{-1}, respectively. This high ß-sheet content of α-agglutinin$_{20\text{-}351}$ is also close to that of silk fibroin (Demura and Asakura, 1991) and human plasma fibronectin (Oesterlund, 1988), both of which are mostly anti-parallel ß-sheet structure (65% and 79%, respectively) and may be close the maximum possible ß-sheet content. Therefore, the ß-sheet content of α-agglutinin$_{20\text{-}351}$ (70%) is among the highest for known proteins and proteolytic fragments with pure anti-parallel ß-sheet structures. The high content of anti-parallel ß-sheets structure in α-agglutinin$_{20\text{-}351}$ suggests the presence of a similar amount of anti-parallel ß-sheet structure in each of the three domains. Such a high ß-sheet content is consistent with the sequence alignment of each of the three domains of α-agglutinin$_{20\text{-}351}$ to an Ig-superfamily consensus sequence. The presence of a predominantly ß-sheet structure, the similarity of CD spectra among Thy-1, CD2 and α-agglutinin$_{20\text{-}351}$, and the presence of Ig-consensus sequence, taken together present a strong case that α-agglutinin has structural similarities to the Ig-superfamily.

A predominant ß-sheet content in all the three IgV domains. The CD spectrum of α-agglutinin$_{155\text{-}351}$ indicated that it contains only 47% ß-sheet structure, accompanied by a corresponding increase in random structure. This structural alteration was probably due to partial unfolding of α-agglutinin$_{155\text{-}351}$ during purification, since no substantial structural alteration was detected the in Arg-C cleaved α-agglutinin$_{20\text{-}351}$.

In the 3-domain model, α-agglutinin$_{155\text{-}351}$ consists of one-half of domain II and

all of domain III. The increase in random structure in α-agglutinin$_{155-351}$ would arise from unfolding of domain II during purification, since the digestion product retains structure similar to native α-agglutinin$_{20-351}$, and half of domain II is missing from the digestion product. α-agglutinin$_{155-351}$ would retain only strands D, E, F, and G of domain II, which would not form a stable ß-sheet structure. Therefore, the 47% of ß-sheet structure in α-agglutinin$_{155-351}$ is likely to be contributed by domain III, which is stabilized by a disulfide bond. If the unfolding of domain II is indeed the major source of increased random structure in α-agglutinin$_{155-351}$, the ß-sheet content in domain III would be similar to that of α-agglutinin$_{20-351}$ as a whole, which is 70%. On the other hand, even if domain III is composed of pure anti-parallel ß-sheet structure (100 % β sheet), domain I and II would still have a ß-sheet content of at least 50%, based on the ß-sheet content in α-agglutinin$_{20-351}$. Therefore, ß-sheet is the predominant structure in all the domains.

Part III: The importance of secondary structure and IgV domains for binding activity

Disulfides and the conformation of α-agglutinin$_{20\text{-}351}$

α-agglutinin$_{20\text{-}351}$ contains six Cys residues. Among them, Cys^{97} in domain I and Cys^{114} in domain II form an interdomain disulfide bond, while Cys^{202} and Cys^{300} form an intradomain disulfide linkage in domain III. One or both of disulfide bonds are essential for agglutination activity: reduction leads to inactivation. Reduction of the disulfides in α-agglutinin$_{20\text{-}351}$ was accompanied by a small reduction of the ß- content from 69% to 61%. A similar reduction of ß-sheet content in α-agglutinin$_{155\text{-}351}$ (from 46.8% to 35.6%) was observed as a consequence of reduction of the intradomain disulfide bond between Cys^{202} and Cys^{227} in domain III. This reduction of ß-sheet content in α-agglutinin$_{155\text{-}351}$ would reflect a reduction of β sheet content in domain III roughly from 70% to 52%, if the ß-sheet content of α-agglutinin$_{155\text{-}351}$ is all contained in domain III. The loss of secondary structure on reduction implies that this disulfide bond is critical in stabilizing domain III.

The observed 18% drop in ß-sheet content in domain III after reduction of its intradomain disulfide bond would correspond to a 7% reduction in ß-sheet content of α-agglutinin$_{20\text{-}351}$. This value is very close to 8% drop detected in α-agglutinin$_{20\text{-}351}$ upon reduction of disulfide bonds. Thus, the change of ß-sheet content of α-agglutinin$_{20\text{-}351}$ appeared to be mostly from reduction of the domain III disulfide bond.

The effects of pH on the ß-sheet content

The agglutination activity of α-agglutinin is pH dependent (Terrance and Lipke, 1981). CD studies illustrated that the ß-sheet content of α-agglutinin$_{20\text{-}351}$ is also pH

dependent. Treatment at pH 8.5 caused a reduction of ß-sheet content from 69% to 40% along with inactivation of α-agglutinin$_{20\text{-}351}$. It is notable that the ß-sheet structures are usually labile to high pH (Yada and Nakai, 1986). It is also interesting to notice that change of the conformation and inactivation of α-agglutinin at this pH is reversible at pH 5.5 (Terrance and Lipke, 1981). On the contrary, a slight increase of ß-sheet content was observed at pH 3.0, which agreed with the phenomena that the ß-sheet structure is stable at acidic pH (Worobec et al., 1987) and can be induced by strengthened hydrogen bonds at low pH (Cabiaux et al., 1989). Inactivation of α-agglutinin at this pH is not likely to be attributable to a conformational change. Protonation of critical amino acid residues that directly participate in binding to a-agglutinin could be the cause of inactivation. Taken together, pH effects supported the hypothesis that α-agglutinin$_{20\text{-}351}$ has the characteristic properties of a ß-sheet structure and provided a rationale for the loss of binding activity at pH value above neutrality.

Multiple domains involved binding mechanism

Endoprotease Arg-C cleaves α-agglutinin$_{20\text{-}351}$ at Lys-154, not a normal site for this enzyme. The sequence at and around this cleavage site was confirmed by sequencing a tryptic peptide containing this Lys (Table IX). Arg-C hydrolysis at this site is, therefore, not a consequence of mutation of the DNA encoding α-agglutinin$_{20\text{-}351}$ changing from Lys to Arg. Tosyl-lysyl chloroketone inhibits Arg-C (Mazzoni et al., 1991), therefore, Arg-C must have some proteolytic activity towards Lys. The existence of this Arg-C digestion site led to successful isolation of the α-agglutinin$_{155\text{-}351}$ fragment containing the known binding region and the Ig-like domain III and, therefore, allowed us to study the agglutination properties and secondary structure of this proteolytic

fragment.

Activity and conformation of the Arg-C digest and α-agglutinin$_{155-351}$. Domain III (amino acid 200-330) contains the only element previously proposed to contribute to the binding site. Purified α-agglutinin$_{155-351}$ fragment containing the entire domain III, however, showed no binding activity. Nor did the α-agglutinin$_{20-351}$-Arg-C digest.

The CD spectrum of α-agglutinin$_{20-351}$-Arg-C digest was almost the same as that of intact α-agglutinin$_{20-351}$. The inactivity of the cleaved α-agglutinin$_{20-351}$ is not likely due to an alteration of the active binding conformation. The very similar CD spectra before and after Arg-C digest further suggests that strands in domain II are associated by the stable anti-parallel ß-sheets structure even after digestion, thereby retaining the native structure. Purification of α-agglutinin$_{155-351}$ from the digest would depend on denaturation of domain II and would result in the observed increase in random structure in α-agglutinin$_{155-351}$.

Ligand binding sites are usually at the interface between Ig domains, and the presence of multiple domains is consistent with the observation that binding of α-agglutinin to a-agglutinin is a multiple domain mechanism. Multiple domains cooperate in binding of cell adhesion molecules are common in Ig-superfamily proteins (William and Barclay), as seen in the homophilic adhesion of N-CAM, heterophilic adhesion of ICAM-1, and antibody-antigen interaction (Hoffman et al., 1984; Staunton et al., 1989; Amit et al., 1986).

Part IV: Disulfide bonds in the binding region of α-agglutinin

The results of mapping disulfide bonding and glycosylation sites are interpreted in light of the proposed domain structure of the protein.

Disulfide bonding between domain I and II

Domain I and II are among IgV domains lacking intradomain disulfide bonds. The amino acid sequence of domains I and II reveals that each domain contains only one cysteine. Cys^{97} is at the conserved position in strand F of domain I, and Cys^{114} is at the conserved position in strand B of domain II. Pro^{34} in B strand of domain I and Ser^{184} in F strand of domain II are substitutes for the respective conserved cysteine counterparts. Therefore, no typical intradomain disulfide linkage can occur in these two domains.

α-agglutinin is not the only Ig-superfamily protein that lacks an intradomain disulfide bond. Many members of the V-type Ig-superfamily domains, including CD2, LFA3, CD4, CEA, PDGFR, CSF1 and PSG/CEA, etc., have no cysteines in putative ß-strands B or F (Williams and Barclay, 1988; Chen et al., 1992). Instead, substitution of the cysteine by amino acid residues with similar structural or hydrophobic properties, like Ser, Val, and Pro, is often found for those Ig-like domains. In these cases, the hydrophobicity of these amino acid substitutions contribute to stabilization of the Ig-fold.

Interdomain disulfide bond between domain I and II. According to the domain alignment, the Cys^{97}-Cys^{114} disulfide would be an interdomain bridge between domains I and II. Such interdomain disulfides are known in other members of IgSF, including the lymphoid differentiation antigen CD33 (Simmons and Seed, 1988), the B cell adhesion molecule CD22 (Stamenkovic and Seed, 1990) and the myelin-associated glycoprotein

(MAG) (Pedraza et al., 1990). Each of these cell adhesion molecule has an interdomain disulfide bridge between the two Ig-like domains closest to the N-terminal. Both MAG and CD22 are members of sialic acid-dependent adhesion molecules (Kelm et al., 1994). The formation of this type of interdomain disulfide bridge brings the two domains close together, presumably, creating a site for adhesive binding. The interdomain disulfide bond between the first two domains of α-agglutinin should also stabilize the predicted fusion of both domains at adjacent ends. The common interdomain disulfide structure in members of sialoadhesion molecule family together with our finding implies that interdomain disulfide linkage may be fairly common among Ig-superfamily members.

Disulfide bonding and free sulfhydryls in domain III

Domain III contains an atypical intradomain disulfide bond. There are four cysteine residues in domain III in the A, B, C', and F strands. For many Ig-like domains, the typical intradomain disulfide linkage is between cysteines of B and F strands.

Though Cys^{227} and Cys^{300} of B and F strands are aligned in positions for the consensus intradomain disulfide bond (Fig.8 and Table I), Cys^{202} in strand A and Cys^{300} in strand F form the actual disulfide linkage. The position of the disulfide in IgV and IgC2 domains is not as restrictively conserved as in IgC1 domains. Atypical intradomain disulfide bonds are found in several molecules belong to these two subsets. MAG consists of one N-terminal IgV domain and four C2 type Ig domains. An intrasheet disulfide linkage between cysteines in strands B and E is present in the IgV domain (Pedraza et al., 1990). CD4 is a T cell antigen receptor that consists of four alternating IgV and IgC2 domains. The C2 type domain II of CD4 contains an intradomain disulfide bond

present between strands C and F (Ryu et al., 1990; Wang et al., 1990).

Domain III contains two free sulfhydryl groups. Though domain III of α-agglutinin shows the highest similarity to the IgV consensus, it has about 30 more amino acids than most other IgV domains. It also has additional unique features. Besides the atypical intradomain disulfide bond between Cys^{202} in strand A and Cys^{300} in strand F of domain III, Cys^{227} in strand B and Cys^{256} in strand C' of this domain are free sulfhydryls. Both cysteines can be derivatized under non-reducing conditions. However, they appear not to be exposed to solvent, since they were labelled only under denaturing conditions (data not shown). α-$agglutinin_{20\text{-}351}$ is a secreted protein. Cysteines on extracellular region of proteins tend to form disulfide bonds whenever possible, due to an oxidative extracellular environment. Cys^{227} and Cys^{256} are, therefore, likely to be in spatial positions that are unfavorable for a disulfide linkage.

This domain is not the only Ig-superfamily member with more than one pair of cysteines. Other members of the Ig-superfamily with an even number of excess cysteines in the extracellular region include MAG, CD22, CD33, c-Kit (Quarles, 1983/84), ICAM-1 (Staunton et al., 1988), rat and human MRC OX45, rat MRC OX2, and the poly Ig receptor (Mostov et al., 1984). However, cysteines in all of these molecules have been presumed to be disulfide bonded in pairs. No disulfide mapping data is available for most of these molecules, except for MAG and rat MRC OX45. The extra cysteine residues in the two N-terminal domains of MAG form an interdomain disulfide, similar to the interdomain disulfide bond between domain I and II of α-agglutinin. The two extra Cys residues of MRC OX45, however, form a second intradomain disulfide bond. On the other hand, CD8α, OX8, and Lyt-2 each contains three cysteines in one Ig domain

(Leahy et al., 1992). Cysteines in strands B and F of CD8α form a consensus intradomain disulfide bond, while cysteine in strand C is a free sulfhydryl, based on the crystal structure of human CD8.

Disulfide bonds required for binding activity

Intradomain disulfide bonds in Ig-like domains stabilize the hydrophobic core structure of the two anti-parallel ß-sheet (Williams and Barclay, 1988). In accord with this role, α-agglutinin$_{20\text{-}351}$ was inactivated by reduction of disulfide bonds, with accompanying denaturation. The critical disulfide linkage(s) for the binding activity are buried inside the core structure of α-agglutinin$_{20\text{-}351}$, since a high concentration of DTT was required to inhibit the activity. Inactivation of α-agglutinin$_{20\text{-}351}$ by reduction of disulfide bonds cannot, however, distinguish whether one or both disulfide bonds is essential for activity.

Both the agglutinability and CD spectra of reduced α-agglutinin$_{20\text{-}351}$ suggests the role of disulfide bonds for both binding activity and for stabilizing the anti-parallel ß-sheet structure, which is consistent with the role of disulfide bonds in IgSF domains.

Part V: Glycosylation in α-agglutinin$_{20\text{-}351}$

Carbohydrate moieties add molecular mass to α-agglutinin$_{20\text{-}351}$

α-agglutinin contains both N- and O-linked oligosaccharides (Hauser and Tanner, 1989; Terrance et al., 1987). Although the molecular mass of the amino acid sequence of α-agglutinin$_{20\text{-}351}$ is 36 kDa, the apparent molecular size on SDS-PAGE is 110 kDa, due to the presence of both N-linked and O-linked carbohydrates. There are six potential N-glycosylation sites in α-agglutinin$_{20\text{-}351}$ with tripeptide sequence of Asn-Xaa-Ser/Thr. The apparent molecular mass reduces to 45 kDa when the N-linked glycans are removed by endo H. These N-linked glycans on α-agglutinin are not important for cell adhesion, as removal of them by endo H does not affect binding activity (Terrance et al., 1987).

The apparent molecular size of de-N-glycosylated α-agglutinin$_{20\text{-}351}$ implies the presence of 9 kDa of O-linked carbohydrate, and its presence is confirmed by affinity of deN-glycosylated α-agglutinin for concanavalin A (Terrance et al., 1987) and by reduction in molecular weight on treatment with HF (Hauser and Tanner, 1989). The presence of a Ser and Thr rich region towards the C-terminus of α-agglutinin$_{20\text{-}351}$ suggests the potential for O-glycosylation in this region.

Function of O-glycosylations in α-agglutinin

O-glycosylation is common for cell surface proteins, with O-linked oligosaccharides often in Ser/Thr-rich regions. Many known cell surface O-glycosylated proteins, like LDL receptor (Goldstein et al., 1985), decay accelerating factor (DAF) (Reddy et al., 1989), and muscle specific domain of N-CAM (Walsh et al., 1989), contain clusters of Ser/Thr enrichment segments in the extracellular membrane-proximal regions. Expression of LDL receptor or DAF in O-linked glycosylation defective mutants

results in rapid degradation of the binding region (Kozarsky et al., 1988; Reddy et al., 1989). In α-agglutinin, the region rich in hydroxy amino acids extends from about residue 300 (the F-strand Cys of domain III) through the entire C-terminal half of α-agglutinin to residue 620, the C-terminus of the mature protein.

α-agglutinin expressed in the presence of tunicamycin, which inhibits N-glycosylation, reacts with Con A, showing that there are O-linked mannose residues (Terrance et al., 1987). The O-linked glycans of yeast cell wall mannoproteins consist of one to four mannose residues (reviewed by Tanner and Lehle, 1987; Kukuruzinska et al., 1987). In this study, O-glycosylation sites were detected as Ser and Thr residues missing from the sequences in multiple analyses. Residues were reported as glycosylated if they were absent from at least two independent sequences (Table VII). The result was further confirmed by Con A affinity of tryptic peptides from de-N-glycosylated α-agglutinin$_{20\text{-}351}$. Table XI interprets these O-glycosylation sites in light of the 3 IgV domains structure. There are multiple O-glycosylated sites starting from the second half of domain III after residue 282. Position 282 is in the loop between the proposed D and E strands of domain III, while the Ser/Thr rich region begins in the loop between strands F and G. Therefore, the boundary for O-glycosylations is within a domain. The O-glycosylation results are illustrated in a shape model of the binding region containing 3 IgV domains (Fig.25).

The Ser-Thr rich region of α-agglutinin continues after residue 351 to the C-terminal cell surface anchorage region. O-glycosylation could continue through this Ser-Thr rich region after α-agglutinin$_{20\text{-}351}$, which is implied by the high molecular mass of

TABLE XI Glycosylation of residues in α-agglutinin$_{20\text{-}351}$

Sequence	Domain I (20-100)			Domain II (105-199)			Domain III (200-265)			Domain III (265-351)		
	G	N	U	G	N	U	G	N	U	G	N	U
N(XT)	-	-	-	-	-	1	-	-	1	2	0	0
N(XS)	-	-	-	-	-	-	-	-	-	0	1	0
T	0	4	5	0	2	6	0	4	0	15	0	0
S	0	4	5	0	5	8	0	6	2	8	0	0

G, glycosylated; N, not glycosylated; U, glycosylation status unknown

Figure 25. Model of α-agglutinin showing proposed domain structure, Cys residues, and glycosylation sites. Disulfide bridges between Cys^{97} and Cys^{114}, Cys^{202} and Cys^{300} are shown. Identified N-linked oligosaccharide moieties are indicated by branched figures. O-linked oligosaccharide moieties are indicated as wavy lines.

α-Agglutinin

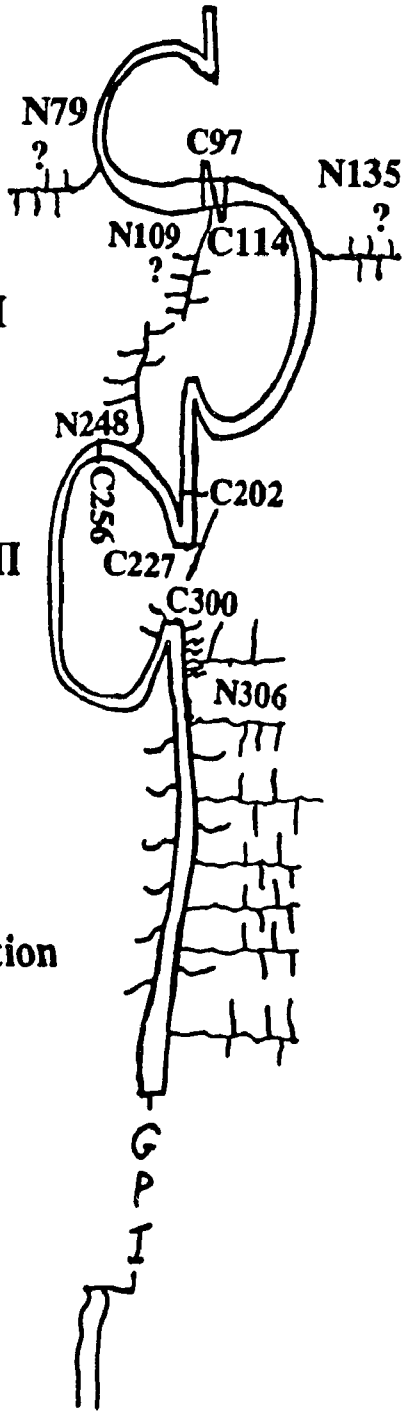
1-19 **Signal Sequence**

20-104 **Domain I**

105-199 **Domain II**

200-351 **Domain III**

352-650 **Glycosylation Region**



truncated versions of α-agglutinin, even after deN-glycosylation with endo H (Wojciechowicz et al., 1993). Clusters of O-glycosylations can linearize the polypeptide backbone structure of the C-terminal 300 amino acids, and would hold the binding region of α-agglutinin out from the wall surface, as has been seen by immunoelectron microscopy (Cappellaro et al., 1994).

N-linked glycosylation: a common feature in the Ig-superfamily

N-linked carbohydrate is one of the dominant features in many Ig-superfamily proteins. There are up to six possible N-linked sites in one IgV domain of human CEA (Oikawa et al., 1987). In mammals, the same protein can be differentially glycosylated in various tissues. The N-linked glycosylation structures of Thy-1 differ between brain and thymus (Parekh, et al., 1987). Glycosylation of N-CAM changes during differentiation (Cunningham et al., 1987). N-CAMs expressed at embryonic and fetal stage are extensively polysialylated, whereas the polysialic acid content is greatly reduced in adult. The homophilic binding affinity of N-CAM increases as the polysialic acid content decreases with age. Thus, a change of glycosylation in N-CAM modulates the adhesive properties of the molecule. The well characterized N-glycosylation sites in IgV domains include a single site in the F strand of human and rat CD4 domain 3; in the loop between E and F strands of Thy-1; and also include 3 sites in the F strand, the loops between E and F strands, and between F and G strands of rat CD2 domain 1 (Williams and Barclay, 1988; Dwek et al., 1993).

The glycosylation status of three of the six potential N-glycosylation sites was examined. The identified N-glycosylation sites are cited in table XI in accordance with the 3 IgV domains assignment. The results conform to the finding that Asn-Xaa-Thr sites

are used as N-glycosylation sites in yeast, but Asn-Xaa-Ser sequences are not glycosylated (Moehle et al. 1987).

The structural features of α-agglutinin as a whole are schematically illustrated in Fig.25, where the 3 IgV domains are shown, with the fused ß-strand between the first two domains. Both the disulfide bonding sites and N-glycosylation sites are marked, with an additional likely N-glycosylation site in the A to B loop of domain II. The two N-glycosylation Asn-Xaa-Thr sequences are present at residues 248-250 and 306-308. These positions in the loops between C and C' stands, and between F and G strands in domain III. Glycosylation sites in corresponding positions are common in IgV domains (Dwek et al., 1993). In summary Fig.25 represents a model of α-agglutinin containing three sequential IgV domains. The model is consistent with CD spectra, sequence alignments, and peptide mapping data.

Part VI: Conclusion

● Alignment of the N-terminal half of α-agglutinin reveals two additional IgV domains in the first 180 amino acids, in addition to the domain previously reported. The alignment also shows that these three Ig-like domain share apparent sequence similarity and could originate from a common ancestral domain.

● Biophysical evidence from α-agglutinin$_{20\text{-}351}$ and a proteolytic fragment support the inclusion of α-agglutinin in the Ig-superfamily. The presence of predominant anti-parallel ß-sheet structure in the entire binding region (70% of all residues) and also in domain III (≥ 55%) is consistent with the proposed Ig-fold structure for all three domains. Disruption of the ß-sheet structure caused loss of activity.

● Peptide mapping revealed the disulfide bonding pattern in the 3 IgV domains: an interdomain disulfide bond between domain I and II, and an intradomain disulfide bond between A and F strands of domain III. At least one of the disulfide bonds is essential for binding activity, probably stabilizing the anti-parallel ß-sheet structure as in other Ig domains.

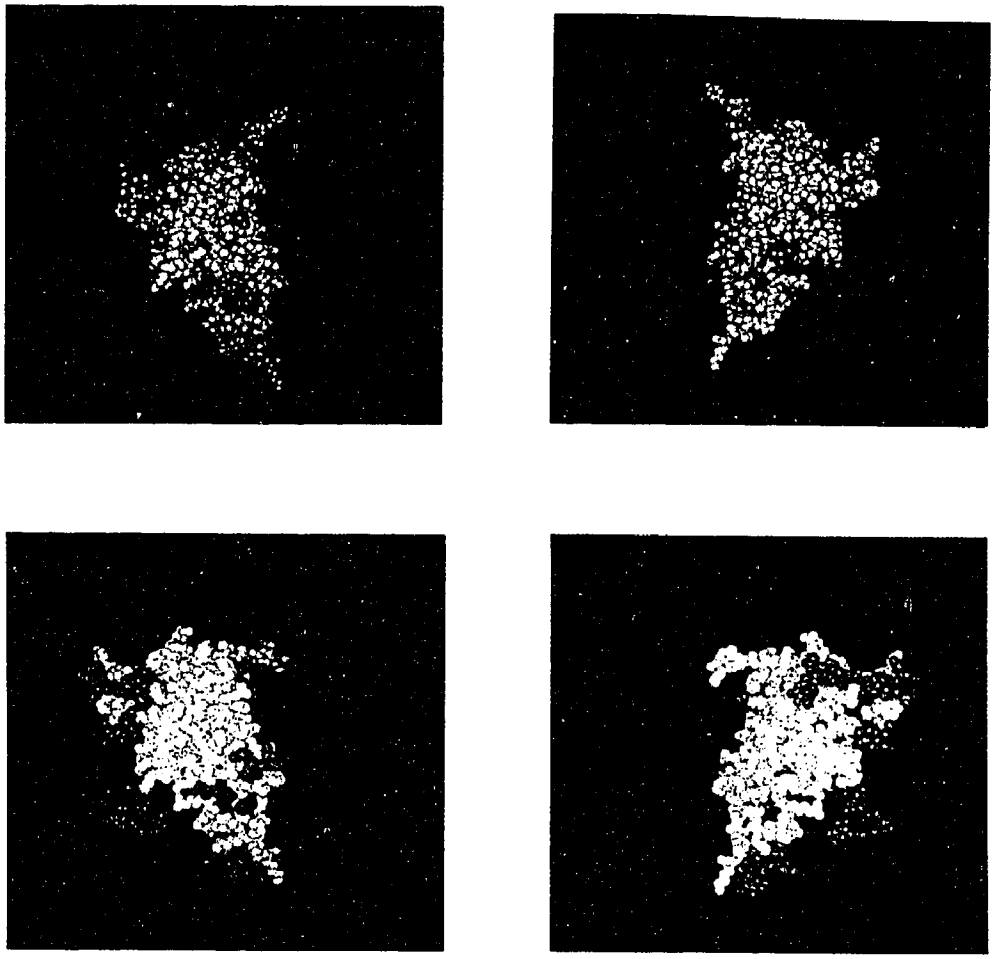
● A Ser/Thr rich region begins at the second half of domain III in the binding region is highly O-glycosylated. The Ser/Thr rich region continues in the C-terminal 300 residue sequence of α-agglutinin and O-glycosylation probably extends into this region, as a common feature for extracellular receptors and adhesion molecules including N-CAM.

● Asn^{248} and Asn^{306} are N-glycosylated. These two glycosylation sites are in the loop of C to C' and F to G strands in domain III, respectively. These positions are common in IgV domains.

● The sequence alignment, circular dichroism analysis, disulfide structure and glycosylation analysis of the N-terminal half of α-agglutinin together support a model of the binding region of α-agglutinin consisting of three IgV domains.

● These results have been used in building a 3-dimensional homolog model of domain III of α-agglutinin. The model was adjusted to accommodate the A and F strand disulfide bonding (Fig.26). The identified N- and O-linked glycosylation sites and identified proteolytic sites were used to test this model, for these residues should be located on the surface of a molecule. Results for the N-terminal first two domains will be very useful for building models of these two domains, and also for assembly of a model of the complete binding region.

Figure 26. Three dimensional models of domain III of α-agglutinin. The model is built based on homology of this domain to members of the variable-type Ig-superfamily. For all the four pictures, the N-terminal residue (Gly^{190}) of the domain is at top and the C-terminal residue (Ala^{333}) of the domain is at bottom. Each pair of pictures is the same version of model in 180° vertical rotation. His^{292}, Tyr^{298}, and Tyr^{309}, which are proposed to form part of the binding site for a-agglutinin, are shown in red. Staph V8 and tryptic sites are in purple and aqua, respectively. These identified proteolytic sites are on the surface of the model. In the lower model, glycosylations are shown. Oligosaccharides are in blue. O-glycosylated Ser and Thr residues are in aqua. N-glycosylated Asn are in green. These glycosylated residues are also on the surface of the model.



Chapter VI

REFERENCES

Amit, A.G., Mariuzza, R.A., Phillips, S.E.V., Poljak, R.J. 1986. Three-dimensional structure of an antigen-antibody complex at 2.8 A resolution. Science 233:747-53.

Amzel, L.M., and Poljak, R. J. 1979. Three-dimensional structure of Igs. Ann. Rev. Biochem. 48:961-97.

Arterburn, L.M., Earles, B.J., and August, T.J. 1990. The disulfide structure of mouse lysosome-associated membrane protein 1. J. Biol. Chem. 265:7419-7423.

Baffi, R.A., Shenbagamurthi, P., Terrance, K., Becker, J.M., Naider, F., and Lipke, P.N. 1984. Different structure-function relationships for α-factor induced morphogenesis and agglutination in *Saccharomyces cerevisiae*. J. Bacteriol. 158: 1152-1156.

Barclay, A., Brady, R., Davis, S., and Lange, G. 1993. CD4 and the immunoglobulin superfamily. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 342:7-12.

Bender, A. and Sprague, G.F.Jr. 1986. Yeast peptide pheromones, a-Factor and α-factor, activate a common response mechanism in their target cells. Cell 47:929-937.

Betz, R., J.W. Crabb, H.E. Meyer, R. Wittig, & W. Duntze. 1987. Amino acid sequences of a-Factor mating peptides from *Saccharomyces cerevisiae*. J. Biol. Chem. 262:546-548.

Bjorkman, P.J., Saper, M.A., Samraoui, B., Bennett, W.S., Strominger, J.L., Wiley, D.C. 1987. Structure of the human class I histocompatibility antigen HLA-A2. Nature 329:512-518.

Bork, P. 1992. The modular architecture of vertebrate collagens. FEBS Lett. 307:49-54. Brady, R.L., Dodson, E.J., Dodson, G.G., Lange, G., Davis, S.J., Williams, A.F. and Barclay, A.N. 1993. Crystal structure of domains 3 and 4 of rat CD4: relation to the amino-terminal domains. Science 260:979-983.

Bresch, C., Muller, G., and Egel, R. 1968. Genes involved in meiosis and sporulation of a yeast. Mol. Gen. Genetics. 102:301-306.

Brahms, S. and Brahms, J. 1980. Determination of protein secondary structure in solution by vacuum ultraviolet circular dichroism. J. Mol. Biol. 138:149-17.

Brock, T. 1958. Mating reaction in the yeast *Hansenula wingei*: preliminary observations and quantitation. J. Bacteriol. 75:697-701.

Burkholder, A. and Hartwell, L. 1985. The yeast α-factor receptor: structural properties deduced from the sequence of the STE2 gene. Nucleic Acids Res. 13:8463-8475.

Cabiaux, B., Brasseur, R., Wattiez, R., Falmagne, P., Ruysschaert, J.M., and Goormaghtigh, E. 1989. Secondary structure of diphtheria toxin and its fragments interacting with acidic liposomes studied by polarized infrared spectroscopy. J. Biol. Chem. 264: 4928-4938.

Campbell, D.A. 1973. Kinetics of the mating-specific aggregation in *Saccharomyces cerevisiae*. J. Bacteriol. 116:323-330.

Campbell, D.G., Williams, A.F., Bayley, P.M., and Reid, K.B.M. 1979. Structural similarities between Thy-1 antigen from rat brain and immunoglobulin. Nature 282:341-42.

Canas, B., Dai, Z., Lackland, H., Poretz, R., and Stein, S. 1993. Covalent attachment of peptides to membranes for dot-blot analysis of glycosylation sites and epitopes. Anal. Biochem. 211:179-182.

Caplan, S., and Kurjan, J. 1991. Role of a α-factor and the MFα1 α-factor precursor in mating in yeast. Genetics 127:299-307.

Cappellaro, C., Hauser, K., Mrsa, V., Watzele, M., Watzele, G., Gruber C., and Tanner W. 1991. *Saccharomyces cerevisiae* **a-** and α-agglutinin: characterization of their molecular interaction. EMBO J. 10:4081-4088.

Cathou, R.E. and Dorrington, K.J. 1975. in Subunits in Biological Systems Part C (eds Timasheff, S.N. and Fassman, G.D. 135, Dekker, New York.

Chan, R.K. and Otte, C.A. 1977. Isolation and genetic analysis of *Saccharomyces cerevisiae* mutants supersensitive to GI arrest by a-factor and α-factor pheromones. J. Bacteriol. 130:766-774.

Chen, H., Plouzek, C.A., Liu, J.-L., and Chen, C.-L. 1992. Characterization of a major member of the rat pregnancy-specific glycoprotein family. DNA Cell Biol. 11:139-148.

Crandall, M. and Caulton, J.H. 1975. Induction of haploid glycoprotein mating factors in diploid yeasts. Methods Cell Biol. 12:185-207.

Crandall, M. and Brock, T.D. 1968. Molecular basis of mating in Hansenula wingei. Bacteriol. Rev. 32:139-163.

Crandall, M., Lawrence, L.M., and Saunders, R.M. 1974. Molecular complementarity of yeast glycoprotein mating factors. Proc. Natl. Acad. Sci. USA. 71:26-29.

Crandall, M., MacKay, V.L., and Egel, R. 1977. Physiology of mating in three yeasts. Adv. Microbiol. 15:307-398.

Crocker, R.P., Kelm, S., Dubois, C., Martin, B., McWilliam, A.S., Shotton, D.M. 1991. Purification and properties of sialoadhesin, a sialic acid-binding receptor of murine tissue macrophages. EMBO J. 10:1661-1669.

Cross, F., Hartwell, L.H., Jackson, C., and Konopka, J.B. 1988. Conjugation in *S. cerevisiae*. Annu. Rev. Cell Biol. 4:429-57.

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

de Nobel, H., Pike, J., Lipke, P.N., and Kurjan, J. in press. Genetic analysis of **a**-agglutinin function in *S. cerevisiae* and identification of a plasmid expression mutant. Mol. Gen. Genet.

Demura, M. and Asakura, T. Porous membrane of *Bombyx mori* silk firoin: Structure characterization, physical properties and application to glucose oxidase immobilization. 1991. J. Membr. Sci. 59:39-52.

Doi, S., Suzuki, Y. and Yoshimura, M. 1979. Induction of sexual cell agglutinability of a mating type cells by α-factor in *Saccharomyces cerevisiae*. Biochem. Biophys. Res. Commun. 91:849-853.

Doi, S. and Yoshimura, M. 1985. A mutation affecting sexual agglutinability in *MAT*α locus of *Saccharomyces cerevisiae*. Curr. Genet. 9:191-196.

Duntze, W., MacKay, V. and Manney, T.R. 1970. *Saccharomyces cerevisiae*: a diffusible sex factor. Science 168:1472-1473.

Dwek, R., Ashford, D., Edge, C., Parekh, R., Rademacher, T., Wing, D., Barclay, A., Davis, S., and Williams, A. 1993. Glycosylation of CD4 and Thy-1. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 342:43-50.

Edelman, G.M. 1988. Morphoregulatory molecules. Biochemistry 27:3533-3543.

Edelman, G.M. 1970. The covalent structure of a human IgG.XI. Functional implications. Biochemsitry 9:3197-3205.

Edmundson, A.B., Ely, K.R., Abola, E.E., Schiffer, M., Panagiotopoulos, N. 1975. Rotational allomerism and divergent evolution of domains in Ig light chains. Biochemistry 14:3953-61.

Fields, S. 1990. Pheromone response in yeast. Trends Biochem. Sci. 15:270-73.

Goldstein, J.L., Brown, M.S., Anderson, R.G.V., Russell, D.W. and Schneider, W.J. 1985. Receptor-mediated endocytosis:concepts emerging from the LDL receptor system. Annu. Rev. Cell Biol. 1:1-39.

Frutiger, S., Hughes, G.J., Hanly, W.C., Kingzette, M., Jaton, J.-C. 1986. The amino-terminal domain of rabbit secretory component is responsible for noncovalent binding to immunoglobulin A dimers. 1986. J. Biol. Chem. 261:16673-81.

Fujimura, H., Yanagishima, N. 1983. Mating-type specific cell cycle arrest and shmoo formation by α pheromone of *Saccharomyces cerevisiae* in *Hansenula wingei*. Arch. Microbiol. 136:79-80.

Gardinier, M.V., Amiguet, P., Linington, C., and Matthieu, J.-M. 1992. Myelin/oligodendrocyte glycoprotein is a unique member of the immunoglobulin superfamily. J. Neurosci. Res. 33:177-187.

Gribskov. M. and Devereux, J. 1991 Sequence Analysis Primer USA Stockton Press.

Hagen, D., G. McCaffrey, and G. Sprague, Jr. 1986. Evidence the yeast STE3 gene encodes a receptor for the peptide pheromone **a** factor: gene sequence and implications for the structure of the presumed receptor. Proc. Natl. Acad. Sci. 83:1418-1422.

Hagiya, M., Yoshida, K., and Yanagishima, N. 1977. The release of sex-specific substances for sexual agglutination from haploid cells of *Saccharomyces cerevisiae*. Exper. Cell Res. 104:263-272.

Harlow, E., and D. Lane. 1988. Antibodies, a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Hauser, K. and Tanner, W. 1989. Purification of the inducible α-agglutinin and molecular cloning of the gene. FEBS Lett. 255:290-294.

Hoffman, S., Chuong, C.M., and Edelman, G.M. 1984. Evolutionary conservation of key structures and binding functions of neural cell adhesion molecules. Proc. Natl. Acad. Sci. USA. 81:6881-6885.

Holmgren, A., Kuehn, M.J., Branden, C.-I., and Hultgren, S.J. 1992. Conserved immunoglobulin-like features in a family of periplasmic pilus chaperones in bacteria. EMBO J. 11:1617-1722.

Horstkorte, R., Schachner, M., Magyar, J.P., Vorherr, T., and Schmitz, B. 1993. The fourth immunoglobulin-like domain of NCAM contains a carbohydrate recognition domain for oligomannosidic glycans implicated in association with L1 and neurite outgrowth. J. Cell Biol. 121:1409-1421.

Hunkapiller, T. and Hood, L. 1989. Diversity of the immunoglobulin gene superfamily. Adv. Immunol. 44:1

Jackson, C.L. and Hartwell, L. 1990a. Courtship in *Saccharomyces cerevisiae*: an Early Cell-Cell Interaction during Mating. Mol. Cell. Biol. 10:2202-2213.

Jackson, C.L. and Hartwell, L. 1990b. Courtship in *Saccharomyces cerevisiae*: both cell types choose mating partners by responding to the strongest pheromone signal. Cell 63:1039-1051.

Jackson, C.L., Konopka, J.B. and Hartwell, L. 1991. *S. cerevisiae* α pheromone receptors activate a novel signal transduction pathway for mating partner discrimination. Cell 67: 389-402.

Jefferis, R., Matthews, J.B., and Bayley, P.M. 1978. Immunochemistry 14:393-396.. Johnson, A.D. and Herskowitz, I. 1985. A repressor: *MAT*α2 product and its operator control expression of a set of cell type specific genes in yeast. Cell 42:237-247.

Jones, E.Y., Davis, S.J., Williams, A.F., Harlos, K., Stuart, D.I. 1992. Crystal structure at 2.8 Angstrom resolution of a soluble form of the cell adhesion molecule CD2. Nature 360:232-239.

Juy, M., Amit, A.G., Alzari, P.M., Poljak, R.J., Claeyssens, M., Beguin, P., and Aubert, J.-P. 1992. Three-dimensional structure of a thermostable bacterial cellulase. Nature 357: 89-91.

Kang, Y., Kane, J., Kurjan, J., Stadel, J.M., and Tipper, D.J. 1990. Effects of expression of mammalian G alpha and hybrid mammalian-yeast G alpha proteins on the yeast pheromone response signal transduction pathway. Mol. Cell. Biol. 10:2582-2590.

Kaufman, J.F., Auffray, C., Korman, A.J., Shackelford, D.A., Strominger, J. 1984. The class II molecules of the human and murine major histocompatibility complex. Cell 36:1-13.

Kawanabe, K., Yoshida, K. and Yanagishima, N. 1979. Sexual cell agglutination in relation to the formation of zygotes in *Saccharomyces cerevisiae*. Plant and Cell Physiol. 20(2):423-433.

Kelm, S., Pelz, A., Schauer, R., Filbin, M.T., Tang, S., de Bellard, M.-E., Schnaar, R.L., Mahoney, J.A., Hartnell, A., Bradfield, P., and Crocker, P.R. 1994. Sialoadhesion, myelin-associated glycoprotein and CD22 define a new family of sialic acid-dependent adhesion molecules of the immunoglobulin superfamily. Curr. Biol. 4:965-972.

Knoll, A.H. 1992. The early evolution of the eukaryotes: a geological perspective. Science 256:622-627.

Kozarsky, K., Kingsley, D., and Krieger, M. 1988. Use of a mutant cell line to study the kinetics and function of O-linked glycosylation of low density lipoprotein receptors. Proc. Natl. Acad. Sci. USA 85:4335-4339.

Kuhn, T.B., Stoeckli, E.T., Condrau, M.A., Rathjen, F.G., Sonderegger, P. 1991. Neurite outgrowth on immobilized axonin-1 is mediated by a heterophilic interaction with L1(G4). J. Cell. Biol. 115:1113-1126.

Kukuruzinska, M.A., Bergh, M.L.E., and Jackson, B.J. 1987. Protein Glycosylation in yeast. Ann. Rev. Biochem. 56:915-944.

Kurjan, J. 1992. Pheromone response in yeast. Annu. Rev. Biochem. 61:1097-1129.

Laemmli, U.K. 1970. Cleavage of structural proteins during the head assembly of the head of the bacteriophage T4. Nature 227:680-685.

Lai, C., Watson, Sutcliffe, J.F., and Milner, R.J. 1987. Neural protein 1B236/MAG defines a subgroup of the Ig superfamily. Immunol. Rev. 100:129-148.

Lemke, G. and Axel, R. 1985. Isolation and sequence of a cDNA encoding the major structural protein of peripheral myelin. Cell, 40:501-508.

Leahy, D.J., Axel, R., and Hendrickson, W.A. 1992. Crystal structure of a soluble form of the human T cell receptor CD8 at 2.6 A. Cell, 68:1145-1162.

Lew, A.M., Lillehoj, E.P., Cowan, E.P., Maloy, W.L., Van Schravendijk, M.R., Coligan, J.E. 1986. Class I genes and molecules: an update. Immunology 57:3-18.

Lipke, P.N. and Kurjan, J. 1992. Sexual agglutinins in budding yeasts: Structure, function and regulation of yeast cell adhesion proteins. Microbiol. Rev. 56:180-194.

Lipke, P.N., Terrance, K. and Wagner, N. 1983. Interactions of sexual agglutinins from yeast. Fed. Proc. 42:2128.

Lipke, P.N., Wojciechowicz, D., and Kurjan, J. 1989. AGα1 is the structural gene for the *Saccharomyces cerevisiae* α-agglutinin, a cell surface glycoprotein involved in cell-cell interactions during mating. Mol. Cell. Biol. 9:3155-3165.

Lipke, P.N., Taylor, A and Ballou, C. 1976. Morphogenic effects of α-factor on *Saccharomyces cerevisiae* **a** cells. J. Bacteriol. 127:610-618.

Lu, C.-F., Kurjan, J., Lipke, P.N. 1994. A pathway for cell wall anchorage of *Saccharomyces cerevisiae* α-agglutinin. Mol. Cell. Biol. 14:4825-4833

Lu, C., Montijn, R.C., Brown, J.L., Klis, F., Kurjan, J., Bussey, H., and Lipke, P.N. 1995. Glycosyl phosphatidylinositol-dependent cross-linking of α-agglutinin and ß-1,6-glucan in the *S. cerevisiae* cell wall. J. Cell Biol. 128:333-340.

Malone, R.E. 1990. Dual regulation of meiosis in yeast. Cell 61:375-378.

Manney, T.R., and Meade, J.H., 1977. Cell-cell interactions during mating in *S. cerevisiae.* In Reissig J.L (ed), Receptors and recognition, Chapman & Hall, London 3:281-321.

Marsh, L., Neiman, A.M., and Herskowitz, I. 1991. Signal transduction during pheromone response in yeast. Annu. Rev. Cell Biol. 7:699-728.

Mazzoni, M.R., Malinski, J.A., and Hamm, H.E. 1991. Structural analysis of rod GTP-binding protein, G sub(t). Limited proteolytic digestion pattern of G sub(t) with four proteases defines monoclonal antibody epitope. J. Biol. Chem. 266:14072-14081.

Mendonca-Previato, L., Burke, D. and Ballou, C.E. 1982. Sexual agglutination factors from the yeast *Pichia amethionina.* J. Cell. Biochem. 19:171-178.

Miceli, M.C., von Hoegen, P., and Parnes, J. 1991. Adhesion vs. coreceptor function of CD4 and CD8: role of the cytoplasmic tail in coreceptor activity. Proc. Natl. Acad. Sci. 88:2623-2627.

Moehle, C.M., Tizard, R., Lemon, S.K., Smart, J., and Jones, E.W. 1987. Protease B of the lysosomelike vacuole of the yeast *Saccharomyces cerevisiae* is homologous to the subtilisin family of serine proteases. Mol. Cell. Biol. 7:4390-4399.

Moore, S.A. 1983. Comparison of dose-response curves for α-factor-induced division arrest, agglutination, and projection formation of yeast cells. J. Biol. Chem. 258:13849-13856.

Moos, M., Tacke, R., Scherer, H., Teplow, D., Frueh, K., Schachner, M. 1988. Neural adhesion molecule L1 as a member of the immunoglobulin superfamily with binding domains similar to fibronectin. Nature 334:701-703.

Mostov, K.B., Friedlander, M., and Blobel G. 1984. The receptor for transepithelial transport of IgA and IgM contains multiple immunoglobulin-like domains. Nature 308:37-43.

Oesterlund, E. 1988. The secondary structure of human plasma fibronectin: conformational changes induced by acidic pH and elevated temperatures; a circular dichroic study. Biochim. Biophys. Acta. 955,330-336.

Oikawas, S., Imajo, S., Noguchi, T., Kosaki, G., Nakazato, H. 1987. The carcinoembryonic antigen (CEA) contains multiple immunoglobulin-like domains. Biochem. Biophys. Res. Commun. 144:634-42.

Orlean, P., H. Ammer, M. Watzele, and Tanner, W. 1986. Synthesis of an O-glycosylated cell surface protein induced in yeast by α-factor. Proc. Natl. Acad. Sci. USA. 83:6263-6266.

Owens, G.C. and Bunge, R.P. Evidence for an early role for myelin-associated glycoprotein in the process of myelination. 1989. GLIA. 2:119-128.

Parekh, R.B., Tse, A.G.D., Dwek, R.A., Williams, A.F., and Rademacher, T.W. 1987. Tissue-specific N-glycosylation, site-specific oligosaccharide patterns and lentil lectin recognition of rat Thy-1. EMBO J. 6:1233-1244.

Pedraza, L., Owens, G.C., Green, D.L.A., and Salzer, J. 1990. The myelin-associated glycoproteins: Membrane disposition, evidence of a novel disulfide linkage between immunoglobulin-like domains, and posttranslational palmitylation. J. Cell Biol. 111:2651-2661.

Pierce, M. and Ballou, C.E. 1983. Cell-cell recognition in yeast: characterization of the sexual agglutination factors from *Saccharomyces kluyveri*. J. Biol. Chem. 258:3576-3582.

Pollard, S.R., Meier, W., Chow, P., Rosa, J.J., and Wiley, D.C. 1991. CD4-binding regions of human immunodeficiency virus envelope glycoprotein gp120 defined by proteolytic digestion. MicroBio. 88:11320-11324.

Quarles, R.H. 1983/84. Myelin-associated glycoprotein in development and disease. Dev. Neurosci. 6:285-303.

Recny, M.A., Neidhardt, E.A., Sayre, P.H., Ciardelli, T.L., and Reinherz, E.L. 1990. Structural and functional characterization of the CD2 in an alph-beta protein folding class. J. Biol. Chem. 265:8542-8459.

Reddy, P., Caras, I., and Krieger, M. 1989. Effects of O-linked glycosylation on the cell surface expression and stability of decay-accelerating factor, a glycophospholipid-anchored membrane protein. J. Biol. Chem. 264:17329-17336.

Roy, A., Lu, C.F., Marykwas, D., Lipke, P.N. and Kurjan, J. 1991. The *AGA1* gene is involved in cell surface attachment of *Saccharomyces cerevisiae* cell adhesion glycoprotein a-agglutinin. Mol. Cell. Biol. 11:4196-4206.

Ryu, S.-E., Kwong, P.D., Truneh, A., Porter, T.G., Arthos, J., Rosenberg, M., Dai, X., Xuong, N., Axel, R., Sweet, R.W., and Hendrickson, W.A. 1990. Crystal structure of an HIV-binding recombinant fragment of human CD4. Nature 348:419-425.

Sambrook, J., Fritsch, E. F., and Maniatis, T. 1989. Molecular cloning: a laboratory manual 1.21-1.52, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Seed, B. 1987. An LFA-3 cDNA encodes a phospholipid-linked membrane protein homologous to its receptor, CD2. Nature 329:840-42.

Shimoda, C. and Yanagishima, N. 1975. Mating reaction in *Saccharomyces cerevisiae*. VIII. Mating type-specific substances responsible for sexual cell agglutination. Antonie van Leeuwenhoek. 41:521-532.

Sijmons, P.C., A.J.A. Nederbragt, F.M. Klis, and H. Van Den Ende. 1987. Isolation and composition of the constitutive agglutinins from haploid *Saccharomyces cerevisiae* cells. Arch. Microbiol. 148:208-212.

Simmons, D., and Seed, B. 1988. Isolation of a cDNA encoding CD33, a differentiation antigen of myeloid progenitor cells. J. Immunol. 141:2797-2800.

Sprague, G.F., Jr., Blair, L.C., and Thorner, J. 1983. Cell interactions and the regulation of cell type in the yeast *Saccharomyces cerevisiae*. Ann. Rev. Microbiol. 37:623-660.

Stamenkovic, I., and Seed, B. 1990. The B-cell antigen CD22 mediates monocyte and erythrocyte adhesion. Nature 345:74-77.

Staunton, D.E., Dustin, M.L., Erickson, 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.

Staunton, D.E., Martin, S.D., Stratowa, C., Dustin, M.L., and Springer, T.A. 1988. Primary structure of ICAM-1 demonstrates interaction between members of the immunoglobulin and integrin supergene families. Cell 52,925-933.

Staunton, D.E., Merluzzi, V.J., Rothlein, R., Barton, R., Marlin, S.D., and Springer, T.A. 1989. A cell adhesion molecule, ICAM-1, is the major surface receptor for rhinoviruses. Cell 56:849-853.

Stotzler, D. and Duntze W. 1976. Isolation and characterization of four related peptides exhibiting α factor activity from *Saccharomyces cerevisiae*. Enzymes 65:257-262.

Tanner, W. and Lehle L. 1987. Protein glycosylation in yeast. Biochimica et Biophysica Acta 906:81-99.

Terrance, K., Heller P., Wu Y.-S., and Lipke, P.N. 1987. Identification of glycopeptide components of α-agglutinin, a cell adhesion protein from *Saccharomyces cerevisiae*. J. Bacteriol. 169:475-482.

Terrance, K. and Lipke, P.N. 1987. Pheromone induction of sexual agglutinability in *Saccharomyces cerevisiae* **a** cells. J. Bacteriol. 169:4811-15.

Terrance, K. and Lipke, P.N. 1981. Sexual agglutination in *Saccharomyces cerevisiae*. J. Bacteriol. 148:889-896.

Trapp, B.D. 1990. Myelin associated glycoprotein, location and potential function. 1990. In Myelination and dysmyelination. Edited by Duncan ID, Skoff, R.P., Colman, D. The N.Y. Acad. Sci. 605:29-43.

Trueheart, J., Boeke, J.D., and Fink, G.R. 1987. Two genes required for cell fusion during yeast conjugation: evidence for a pheromone-induced surface protein. Mol. Cell. Biol. 7:2316-2328.

Van den Berg, T.K., Breve, J.J.P., Damoiseaux, J.G.M.C, Dopp E.A., Kelm, S., Crocker, P.R. 1992. Sialoadhesin on macrophages-its identification as a lymphocyte adhesion molecule. J. Exp. Med. 176:647-655.

Wagner, N., and Lipke, P.N. 1983. Properties of **a**-agglutinin from *S. cerevisiae*. Abstr. 83rd Annu. Meet. Am. Soc. Microbiol. 175 American Society for Microbiology, Washington, D.C.

Walsh, F.S., Parekh, R.B., Moore, S.E., Dickson, G., and Barton, C.H. 1989. Tissue specific O-linked glycosylation of the neural cell adhesion molecule (N-CAM). Development 105:803-811.

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 411:411-418.

Watzele, M., Klis, F. and Tanner, W. 1988. Purification and characterization of the inducible **a**-agglutinin of *Saccharomyces cerevisiae*. EMBO J. 7:1483-1488.
Wickerham, L.J. 1958. Sexual agglutination of heterothallic yeasts in diverse taxonomic areas. Science 128:1504-1505.

Williams, A.F. and Barclay, A.N. 1988. The immunoglobulin superfamily:domains for cell surface recognition. Ann. Rev. Immunol. 6:381-405.

Williams, A.F. Davis, S.J., and Barclay, A.N. 1989. Structural diversity in domains of the immunoglobulin superfamily. Cold Spring Harb. Symp. Quant. Biol. 54:637-647.

Williams, A.F. and Gagnon, J. 1982. Neuronal cell Thy-1 glycoprotein:homology with Ig. Science 216:696-703.

Wojciechowicz, D. 1990. Ph. D. Thesis, City University of New York.

Wojciechowicz, D., Lu, C., Kurjan, J., and Lipke, P. 1993. Cell surface anchorage and ligand-binding domains of the *S. cerevisiae* cell adhesion protein α-agglutinin, a member of the immunoglobulin superfamily. Mol. Cell. Biol. 13:2554-2563.

Wojciechowicz, D., and Lipke, P.N. 1989. α-agglutinin expression in *Saccharomyces cerevisiae*. Biochem. Biophys. Res. Commun. 161:45-51.

Worobec, E.A., Martin, N.L., McCubbin, W.D., Day, C.M., Brayer, G.D., and Hancock, R.E.W. 1987. Large-scale purification and biochemical characterization of crystallization-grade porin protein P from *Pseudomonas aeruginosa*. Biochim. Biophys. Acta. 939:366-374.

Yada, R.Y. and Nakai, S. 1986. Secondary structure of some aspartyl proteinases. J. Food Biochem. 10:155-183.

Yamaguchi, M., Yoshida, K., Banno., I., and Yanagishima, N. 1984. Mating-type differentiation in ascosporogenous yeasts on the basis of mating-type-specific substances responsible for sexual cell-cell recognition. Mol. Gen. Genet. 194:24-30.

Yanagishima, N, and Fujimura, H. 1981, Sex pheromones of the yeast *Hansenula wingei* and their relationship to sex pheromones in *Saccharomyces cerevisiae* and *Saccharomyces kluyveri*. Arch. Microbiol. 129:281-284.

Yanagishima, N., and Yoshida., K. 1981. Sexual interactions in *S. cerevisiae* with special reference to the regulation of sexual agglutinability. In O'Day D.H. and Horgen, P.A. (ed.), Sexual interactions in eukaryotic microbes, Academic Press, Inc., New York 261-295.

Yanagishima, N. 1984. In Linskens, H.F. and Haslop-Harrison, J.H. (ed.), Cellular interactions. Encyclopedia of plant physiology, series N, Springer-Verlag KG, Berlin 17:403-423.

Yang, J.T., Wu, C.C., and Martinez, H.M. 1986. Calculation of protein conformation from circular dichroism. Methods in Enzym. 130:228.

Yarden, Y., Escobedo, J.A., Kuang, W.-J., Yang-Feng, T.L., Daniel, T.O., Tremble, P.M., Chen, E.Y., Ando, M.E., Harkins, R.N., Francke, U., Fried, V.A., Ullrich, A., Williams, L.T. 1986. Structure of the receptor for platelet-derived growth factor helps define a family of closely related growth factor receptors. Nature 323:226-32.

Yen, P.H. and C.E. Ballou. 1974. Partial characterization of the sexual agglutination factor from *Hansenula wingei* Y-2340 type 5 cells. Biochemistry 13:2428-2437.

Yoshida, K., Hisatomi, T., and Yanagishima, N. 1989. Sexual behavior and its pheromonal regulation in ascosporogenous yeasts. J. Basic Microbiol. 29:99-118.

Zvelebil, M.J., Barton, G.J., Taylor, W.R., and Sternber, M.J.E. 1987. Prediction of protein secondary structure and active sites using the alignment of homologous sequences. J. Mol. Biol. 195:957-61.