Structure of Saccharomyces cerevisiae $\alpha$-Agglutinin

EVIDENCE FOR A YEAST CELL WALL PROTEIN WITH MULTIPLE IMMUNOGLOBULIN-LIKE DOMAINS WITH ATYPICAL DISULFIDES*

(Received for publication, June 1, 1995, and in revised form, August 8, 1995)

Min-Hao Chen‡§, Zheng-Ming Shen‡, Stephen Bobin‡, Peter C. Kahn‡, and Peter N. Lipke‡**

From the *Department of Biological Sciences and The Institute for Biomolecular Structure and Function, Hunter College of the City University of New York, New York 10021 and the **Department of Biochemistry and Microbiology, Cook College, Rutgers University, New Brunswick, New Jersey 08893

$\alpha$-Agglutinin of Saccharomyces cerevisiae is a cell wall-associated protein that mediates cell interaction in mating. Although the mature protein includes about 610 residues, the NH$_2$-terminal half of the protein is sufficient for binding to its ligand a-agglutinin. $\alpha$-Agglutinin$^{20-351}$, a fully active fragment of the protein, has been purified and analyzed. Circular dichroism spectroscopy, together with sequence alignments, suggests that $\alpha$-agglutinin$^{20-351}$ consists of three immunoglobulin variable-like domains: domain I, residues 20–104; domain II, residues 105–199; and domain III, residues 200–326. Peptide sequencing data established the arrangement of the disulfide bonds in $\alpha$-agglutinin$^{20-351}$. Cys$^{97}$ is disulfide-bonded to Cys$^{114}$, forming an interdomain bond between domains I and II. Cys$^{202}$ is bonded to Cys$^{300}$, in an atypical intradomain disulfide bond between the A and F strands of domain III. Cys$^{227}$ and Cys$^{256}$ have free sulfhydryls. Sequencing also showed that at least two of three potential N-glycosylation sites with sequence Asn-Xaa-Thr are glycosylated. At least one of three Asn-Xaa-Thr sequences is not glycosylated, whereas Ser$^{282}$, and all hydroxy amino acid residues COOH-terminal to this position were modified. Therefore O-glycosylated Ser and Thr residues cluster in the COOH-terminal region of domain III, and the O-glycosylation continues into a Ser/Thr-rich sequence that extends from domain III to the COOH-terminal of the full-length protein.

Sexual agglutinins are expressed on the surface of haploid budding yeasts, including Saccharomyces cerevisiae (Lipke and Kurjan, 1992; Pierce and Ballou, 1983; Hagiya et al., 1977; Crandall et al., 1974; Crandall and Brock, 1968). During mating, the interaction of complementary agglutinins of each species mediates direct cell-cell contact to promote fusion of pairs of mating partners to form diploid zygotes. Mutants defective in these sexual agglutinins are mating-deficient in liquid medium (Lipke et al., 1989).

S. cerevisiae $\alpha$-agglutinin is a highly glycosylated cell wall-anchored protein that is constitutively expressed on cells of the $\alpha$ mating type and is induced to greater expression levels in response to the mating pheromone, $\alpha$-factor (Terrance et al., 1987; Hauser and Tanner, 1989; Lipke et al., 1989). The open reading frame of the $\alpha$-agglutinin gene, AGU1, encodes a single polypeptide of 650 amino acids, including an NH$_2$-terminal secretion signal (residues 1–19) and a COOH-terminal glycosylphosphatidylinositol (GPI) addition signal that is involved in cell wall anchoring (residues 628–650) (Kodukula et al., 1993; Wojciechowicz et al., 1993; Kaptyn et al., 1994; Lu et al., 1994, 1995; Van Berkel et al., 1994). The NH$_2$-terminal part of the mature protein (residues 20–350) contains the binding region, which has been proposed to consist of three domains (Wojciechowicz et al., 1993). These features are summarized in Fig. 1.

Within the NH$_2$-terminal half, a segment (amino acid residues 200–326, designated domain III) shows significant similarity to variable domains of the immunoglobulin superfamily (IgV domains) based on the amino acid sequence and predicted $\beta$-sheet profile analysis (Wojciechowicz et al., 1993). A His residue essential for binding has been identified within this putative domain (Cappellaro et al., 1991), and other essential residues have been identified by site-specific mutagenesis. 2 We have proposed that domains I and II are also Ig-like, but evidence to support this contention has been lacking. In Ig domains, post-translational modifications help determine tertiary structure (Dwek et al., 1993; Williams and Bar- day, 1988). We have investigated the disulfide bonding pattern of the $\alpha$-agglutinin by analyzing the positions of the N- and O-glycosylations in the Ig-like region (Terrance et al., 1987; Hauser and Tanner, 1989). N-Linked glycans are not important for cell adhesion, because endo H treatment or synthesis in the presence of tunicamycin does not affect binding activity (Terrance et al., 1987). O-Linked glycans are also present and appear to account for a significant portion of the apparent molecular size of $\alpha$-agglutinin (Wojciechowicz et al., 1993; Lu et al., 1994).

We have now produced a 332-residue active fragment, $\alpha$-agglutinin$^{20-351}$, in quantities sufficient to allow investigation of the secondary structure and determine the positions of post-translational modifications. The results, along with those of a modified sequence alignment procedure, result in a model for $\alpha$-agglutinin.

*This work was supported by the National Institute for General Medical Science and the Research Centers in Minority Institutions Program of the National Institutes of Health. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

§ Current address: Dept. of Pathology, Harvard Medical School, 200 Longwood Ave., Boston MA 02115.

‡ Current address: Dept. of Pharmacology and Toxicology, Dartmouth School of Medicine, Hanover, NH 03755.

** To whom correspondence should be addressed: Dept. of Biological Sciences, Hunter College, 695 Park Ave., New York, NY 10021. Tel.: 212-772-5235; Fax: 212-772-4073; E-mail: lipke@genctr.hunter.cuny.edu.

1 The abbreviations used are: GPI, glycosylphosphatidylinositol; ConA, concanavalin A; DTT, dithiothreitol; endo H, endo-N-acetylglucosaminidase H; Ig, immunoglobulin; IgV, immunoglobulin variable domain; P-2007, N-(1-pyrenemethyl)iodoacetamide; PAGE, polyacrylamide gel electrophoresis; TCP, tris-(2-carboxymethyl)phosphine hydrochloride; PKG, phosphoglycerate kinase; HPLC, high performance liquid chromatography.

2 H. De Nadel, P. N. Lipke, and J. Kurjan, submitted for publication.
EXPERIMENTAL PROCEDURES

Chemicals and Reagents—All chemicals were from Sigma, unless otherwise stated, and of appropriate purity. Nitrocellulose membranes were from Schleicher & Schuell. Reagents for gel electrophoresis were from Kodak Scientific Imaging Systems. Protein standards and Bio-Gel P-60 were purchased from Bio-Rad. Reagents for polymerase chain reactions were obtained from Perkin-Elmer, and restriction enzymes were from New England Biolabs or U. S. Biochemical Corp. Enzyme Arg-C, sequencing grade Staphylococcus aureus V8, hydrophilic bead-bound trypsin, and endoprotease Asn-N were from Boehringer Mannheim. The cysteine-specific reagent P-2007 (N-(1-pyrrenemethyl)-iodobacetamide) and reducing reagent TCEP (tris-(2-carboxyethyl)phosphine hydrochloride) were from Molecular Probes. Immobilon-AV membranes were purchased from Micron Scientific. Molecules, Ura+ ade2-1 his3-11, 15 maa-2, 3,112 trp-1 ura-3 can1-100, was used to express the α-agglutinin (20–351)-encoding the secretion signal. The purified polymerase chain reaction product was cloned into the pPGK-H31 vector. Bioassays utilized test strain X2180-1A (MATa Suc2 mal63 gal2 CUP1) and X2180-1B (MATa Suc2 mal63 gal2 CUP1) (Terrance and Lipke, 1981). The expression vector, YEp-PGK, containing the pBR322-derived Ampr and OriE, the yeast URA3 gene, and 2μm replication origin, allowed the cloning of the AG1311 fragment between the constitutive phosphoglycerate kinase (PGK) promoter and terminator (Kang et al., 1990). Plasmids were propagated in Escherichia coli strain HB101.

Construction of pPGK-AG1311—Two single-stranded oligonucleotides were synthesized to use as primers for the construction of pPGK-AG1311. AG25-H3, TGC GCC AAG CTT TTA AAA ATG TCT ACT TTT CTC, and AGM-H3, AAA TGG AAG CTT TGG ATT ACG CAC TAG TGT TTA TAC TTG T, contain HindIII sites (underlined nucleotides) outside the open reading frame. The 3′ end primer included a stop codon (nucleotides with double underline) corresponding to Tyr230 in the deduced α-agglutinin protein sequence. The DNA fragment encoding α-agglutinin (20–351) was amplified using the AG21-containing plasmid pH27 (Lipke et al., 1989) as template in a polymerase chain reaction. The polymerase chain reaction product contained the open reading frame of AG1 from nucleotides 1 to 1053 and included the sequence encoding the secretion signal. The purified polymerase chain reaction product was cloned into the HindIII site of the expression vector YEPPGK. The orientation of the insert was checked by restriction mapping with EcoRI, HindIII, and BamHI, and the sequence of the insert in pPGK-AG1311 was verified by DNA sequencing.

Overexpression and Purification of α-Agglutinin (20–351) from Culture Supernatant—pPGK-AG1311, encoding α-agglutinin (20–351), was introduced into the aga1 mutant Lc21. Transfomants were grown to stationary phase in 1-liter cultures of synthetic uracil-less medium overnight at room temperature. The cells were centrifuged, and the culture supernatant was concentrated and dialyzed against 10 mM sodium acetate buffer, pH 7.5, at 4°C. The dialyzed material was partially purified by chromatography on a DEAE-Sephadex column (120-ml bed volume) which was previously equilibrated with 10 mM sodium acetate, pH 5.5, containing 0.15 M sodium chloride and 0.1% SDS (w/v). Peptide samples (3×1 μl) were spotted onto an Immobilon-AV membrane. The membrane was air-dried and incubated for 30 min in Tris-HCl buffer, pH 7.5, containing 0.15 M sodium chloride and 0.1% Tween 20 (TTBS) and then blocked in a 3-h incubation in with fresh 10% ethanolamine (w/v)
in 1 M sodium bicarbonate buffer, pH 9.5. After blocking, the membranes were incubated for 1 h with 0.5 μg/ml concanavalin A (ConA)-conjugated peroxidase in TTBS. After washing three times in 10 mM Tris-HCl buffer, pH 7.5, containing 0.15 M NaCl, the membranes were stained with 4-chloro-1-naphthol and hydrogen peroxide (Canas et al., 1993).

Other Methods—SDS-polyacrylamide gel electrophoresis was carried out according to the method of Laemmli (1970), using 12 and 15% gels. Proteins were visualized by staining with Coomassie Blue or a Silver Staining Plus kit (Bio-Rad). Protein concentrations were determined by the bicinchoninic protein assay method (Pierce) using bovine serum albumin, fraction V as standard.

RESULTS

Expression and Secretion of α-Agglutinin—The plasmid pPGK-Ag1351 encodes a 351-residue form of α-agglutinin that lacks the COOH-terminal sequences which anchor α-agglutinin to the cell wall (Wojciechowicz et al., 1993), and therefore the product, α-agglutinin20–351, is secreted into the culture medium after cleavage of the 19-residue secretion signal. An aga1 mutant harboring this plasmid secreted 4.5 × 10^4 units (about 1 mg) of α-agglutinin/liter (Terrance et al., 1987; Wojciechowicz et al., 1993). α-Agglutinin20–351 in crude culture supernatants was identified by immunoblots before and after endoglycosidase H treatment (Fig. 2). The fully glycosylated protein had an apparent molecular size of 110 kDa. After removal of N-linked carbohydrates with endo H, the molecular size of α-agglutinin20–351 was reduced to 45 kDa. In some preparations, the protein was present as a doublet (Fig. 3), due to incomplete removal of N-linked glycan at one site (data not shown). The mobility of the deglycosylated α-agglutinin20–351 decreased after treatment with DTT (Fig. 2). This decrease implies an increase in the Stokes radius caused by reduction of disulfide bonds.

Elution of endo H-treated α-agglutinin20–351 from a Bio-Gel P-60 column gave purified α-agglutinin with an apparent molecular size of 45 kDa for the smaller species on SDS gels (Fig. 3). The deduced M_r of α-agglutinin20–351 from the predicted amino acid sequence is 37,108.

Therefore, N-linked carbohydrate accounts for two-thirds of the apparent 110-kDa molecular mass of α-agglutinin20–351 and the O-linked carbohydrate remaining after endo H digestion could account for an additional 8 kDa of apparent mass.

Endoprotease Arg-C Digestion of α-Agglutinin20–351—When purified α-agglutinin20–351 was digested with mouse endoprotease Arg-C from mouse submaxillary glands, proteolytic fragments of 31, 21, and 16 kDa were generated (Fig. 4).

The NH₂-terminal sequence of each fragment was determined by microsequence analysis after electroblotting onto polyvinylidene difluoride membranes. Both the 16- and 21-kDa fragments had the same NH₂-terminal sequence as mature α-agglutinin, beginning at Ile20, immediately following the secretion signal sequence (Table I). The 21-kDa form represented a species with some N-linked carbohydrate remaining and generated a 16-kDa fragment after additional treatment with endo H (data not shown). The NH₂-terminal α-agglutinin polypeptide from Ile20 to Lys154 would have a molecular mass of 15,119 daltons, close to the value for the 16-kDa peptide. The 31-kDa fragment, called α-agglutinin155–351, started with Ser150-Gly-Pro-Met-Leu-Val (Table I). The predicted molecular mass of this fragment was 37,108 daltons.
this peptide is 21,989 Da. The extra 7 kDa of apparent molecular mass in agglutinin155–351 may be attributed to the presence of multiple O-glycosylations (see below). No additional fragments were seen, including any of the predicted peptides following Arg residues (Fig. 4). Therefore, endoprotease Arg-C cleaved only at Lys154, instead of any of the six Arg residues in α-agglutinin20–351.

Endoprotease Arg-C from Clostridium histolyticum also cleaved at Lys154 only (data not shown). Peptide sequencing confirmed that the cleaved residue was Lys. No fragments were generated in α-agglutinin20–351 incubated without protease. Therefore, hydrolysis of α-agglutinin20–351 at Lys154 was endoprotease Arg-C specific and not due to proteolytic activity in the α-agglutinin preparations or in other reagents used for the digestion. Tosyl-lysyl chloroketone inhibits Arg-C (Mazzoni et al., 1991); therefore, Arg-C must have proteolytic activity toward Lys.

Agglutination Activity of Proteolytic Fragments of α-Agglutinin20–351—To examine whether any of the endoprotease Arg-C digested fragments retained agglutination activity, protease-treated α-agglutinin20–351 was reconstituted with sodium acetate buffer to pH 5.5 and assayed for activity. This material had no measurable agglutination activity at concentrations up to 6.7 μg/ml, whereas native α-agglutinin20–351 was active at 3.3 ng/ml. Therefore, the agglutination activity was less than 2 × 10−4 that of intact α-agglutinin20–351. Similarly, the 31-kDa α-agglutinin155–351 fragment purified on a Bio-Gel P-30 column had less than 10−4 of the binding activity of α-agglutinin20–351 (data not shown).

CD of Native α-Agglutinin20–351—α-Agglutinin20–351 has been proposed to consist of three Ig-like domains, which would consist of predominantly antiparallel β-sheets along with associated turns and loops, but little or no α-helix content (Williams and Barclay, 1988; Wojciechowicz et al., 1993). The CD spectrum of α-agglutinin20–351 (Fig. 5) showed a typical β-sheet structure profile, with a negative band at 217 nm (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-agglutinin20–351. Quantitative analysis of the CD spectrum of α-agglutinin20–351 indicated the presence of 6.8% α-helix, 69.4% β-sheet, 13.2% turns, and 10.5% random structure. This high β-sheet content suggests the presence of antiparallel β-sheet structures, consistent with Ig domains.

CD of α-Agglutinin20–351 Digested with Endoprotease Arg-C—Because α-agglutinin20–351 is inactivated by endoprotease Arg-C cleavage at Lys154, the effect of the digestion on the structure of α-agglutinin20–351 fragments was examined. The digestion product showed substantial reduction in β-sheet content when spectra were taken at pH 7.8 (Fig. 5). However, after reconstitution at pH 5.5 for 30 min, the CD spectrum of the digest was very similar to that of native α-agglutinin20–351, in both the negative peak position at 217 nm and the corresponding peak width (data not shown). Quantitative analysis of the CD spectrum revealed that the secondary structural profile was similar to native α-agglutinin20–351, with 68.8% β-sheet, and a slightly higher aperiodic structure content. This CD profile indicated that the single site digestion at Lys154 of α-agglutinin20–351 did not substantially alter the secondary structure of the protein fragments. Therefore, the inactivation of the binding activity is not due to gross structural change during the Arg-C digestion.

Disulfides in Endoprotease Arg-C-digested α-Agglutinin20–351—The products of endoprotease Arg-C cleavage of α-agglutinin20–351 were separable in the absence of reducing agents (Fig. 4), showing that there is no disulfide linkage between them. Both the 21- and the 31-kDa fragments showed lower mobility on SDS-PAGE after DTT treatment, suggesting that each fragment contained one or more internal disulfide bonds. Based on the deduced amino acid sequence, the 21-kDa fragment contained Cys97 and Cys114, implying that these residues form a disulfide bond. The 31-kDa fragment, α-agglutinin155–351, contained four Cys residues (Cys202, Cys227, Cys256, and Cys300). Therefore, the disulfide bonds in this fragment could not be determined from the endoprotease Arg-C data.

Identification of Disulfide Bonds—Identification of these disulfide bonds was accomplished by sequencing of tryptic and S. aureus V8 peptides that had different HPLC retention times in the presence and absence of DTT. Free sulfhydryls were identified in peptides that were not affected by DTT and confirmed by labeling with the iodoacetamide derivative P-2007.

α-Agglutinin20–351 was digested with trypsin in the presence or absence of DTT, and the products were separated by reversed phase chromatography on a C18 column. Three tryptic peptides (T1, T2, T2') were unique to the nonreduced chromatogram (Fig. 6A), and three peptides (DT1, DT2, and DT3) were unique to the reduced chromatogram (Fig. 6B). These peptides were sequenced and compared with the sequences of the Cys-containing tryptic fragments predicted from the gene sequence (Tables II–IV). Peaks T1 and DT1 had the sequence of the predicted peptide containing both Cys97 and Cys114. As with
peaks were peptides predicted to include Cys202 and Cys300, peaks appeared with retention times of 117 and 154 min (labelled DT2 and DT3) profiles are labeled. The peptide containing Cys227 and Cys256 is peak T4 in nonreduced and peak DT4 in the reduced profile. The amino acid sequences of these peptides are listed in Tables I, II, III, and IV. Both chromatograms were obtained under standard conditions, and the retention times shown in B apply to both chromatograms. Fraction numbers shown in A correspond to those mentioned in the text for concanavalin A blotting.

the change in gel mobility, the change in retention time in the presence of DTT implied that these two Cys residues formed an internal disulfide. Similar chromatography and sequencing analyses of peptides from S. aureus V8 digests confirmed this assignment (Tables III and IV); peptide DS2 was seen only after reduction and contained Cys97. As expected, tryptic peptide T1 containing Cys97 and Cys114 was labeled with P-2007 (Fig. 7, A versus B). To determine if the peptide contained two labeled cysteines, the isolated labeled peptide (Fig. 7C) was further digested with endoprotease Asp-N and rechromatographed (Fig. 7D). Two additional labeled peptides were detected at 35 and 45 min, as a result of the digestion. These peptides had the retention times expected for the labeled peptides containing Cys256 and Cys227, respectively. The original labeled peptide with a retention time of 53 min, however, was still present, probably due to incomplete digestion. Therefore, both Cys227 and Cys256 are free cysteines.

To verify that peptide peak T4 in the nonreduced profile contained Cys227 and Cys256 as free sulfhydryls, this peptide was labeled with P-2007. This peptide alone was labeled in reactions of tryptic digests with P-2007 under nonreducing conditions (Fig. 7, A versus B). To determine if the peptide contained two labeled cysteines, the isolated labeled peptide (Fig. 7C) was further digested with endoprotease Asp-N and rechromatographed (Fig. 7D). Two additional labeled peptides were detected at 35 and 45 min, as a result of the digestion. These peptides had the retention times expected for the labeled peptides containing Cys256 and Cys227, respectively. The original labeled peptide with a retention time of 53 min, however, was still present, probably due to incomplete digestion. Therefore, both Cys227 and Cys256 are free cysteines.

Identification of O-Linked Glycosylation Sites by Peptide Se- quencing—We have sequenced all recovered tryptic and S. aureus V8 peptides from α-agglutinin20–351, resulting in a peptide sequence that is about 76% complete, and including three of six potential N-glycosylation sequences and 52 of 74 Ser and Thr residues (Fig. 8). Glycosylated Ser or Thr residues are not detected by the sequencer; therefore, peptide sequencing provides an indirect method to identify O-linked glycosylation sites. Absence of a signal for Thr and Ser was interpreted to indicate glycosylation when the expected residues were observed at levels of 20 pmol or greater in the cycles immediately preceding missing Ser or Thr residues. Table VI summarizes the results from sequencing of S. aureus V8 and tryptic α-agglutinin20–351 peptides from two or more independent peptide sequences. A total of four S. aureus V8 peptides and two tryptic peptides contained modified Ser and Thr residues.

Eight Ser residues (positions 282, 316, 331, 334, 335, 338, 346, and 350) and 15 Thr residues (positions 289, 299, 303, 307, 308, 311, 314, 315, 329, 339, 340, 341, 342, 345, and 349) were found to be modified in tryptic peptides and/or S. aureus V8 peptides (Table VI). Therefore, all of the eight Ser and 15 Thr residues from Ser282 to the COOH terminus of α-agglutinin20–351 were modified. All other sequenced Ser and Thr residues were observed as expected (Fig. 8).

Confirmation of O-Glycans with ConA—O-Linked carbohydrates in yeast interact with ConA, because they consist of one to five α-linked mannose residues (Klis, 1994). To examine whether O-linked glycosylations were responsible for the masking of the undetected Ser and Thr residues, peroxidase-conjugated ConA was used to probe peptides from the nonreduced tryptic digest. Dot blot analysis of tryptic fractions of HPLC fractions of nonreduced digest showed that five peptides reacted positively with ConA (data not shown). These peptides (fractions 4, 5, 24, 25, and 26 of Fig. 6A) correlated with fragments containing modified Ser and Thr residues (Table VI). Because the dot blot experiment does not determine which Ser or Thr residues within a peptide were glycosylated, we cannot
definitively conclude that O-glycosylation accounts for all of the modification of Ser or Thr residues in these peptides, but it must account for some.

Identification of N-Linked Glycosylation Sites in α-Agglutinin20–351—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 (Asn-Xaa-Ser/Thr) in α-agglutinin20–351, Asn248, Asp249, Thr250, and Asn306, Thr307, Thr308 were N-glycosylated, whereas Asn348, Thr349, Ser350 was not N-glycosylated in (Table VI: tryptic peptide326–351). Other potential N-glycosylation sequences were in regions that were not successfully sequenced.

DISCUSSION

α-Agglutinin20–351 is fully active and must therefore form a correctly folded structure. A high proportion of β-sheet struc-
Structure of α-Agglutinin

The complete domain, including A and G strands, extends the alignment to residues 200–326 (Fig. 9). A threedimensional model of domain III based on homology to IgV domains has been constructed that accommodates the disulfide bond between Cys^{202} and Cys^{300}, positions of glycosylated residues and proteolytic sites, CD spectra, and site-specific mutagenesis results (Lipke et al., 1995). Thus, an Ig-like structure for domain III can accommodate all available data.

Assignment of domain III as an IgV-like domain suggests that there may be additional Ig-like domains in the NH_{2}-terminal region, because multiple sequential Ig domains are often present in members of the Ig superfamily. In members of the superfamily that are cell adhesion proteins, 2 to 5 sequential domains are common. These tandem domains are at the NH_{2} termini of the mature proteins in the vast majority of cases (Williams and Barclay, 1988). Furthermore, the Ig fold appears to be more widespread than the Ig superfamily itself and proteins with little or no sequence similarity to Ig domains form Ig-like folds. Most of these proteins are involved in cell adhesion or protein-protein interaction (Holmgren et al., 1992; Overduin et al., 1995; Shapiro et al., 1995).

The 180 NH_{2}-terminal residues of α-agglutinin^{20–351} are sufficient to form two more IgV domains, with the G strand of domain I being the A strand of domain II, as in CD4 (Fig. 9) (Williams and Barclay, 1988; Williams et al., 1989; Ryu et al., 1990; Wang et al., 1990; Barclay et al., 1993). A revised alignment procedure for α-agglutinin^{20–351} strongly supports a three-domain assignment (Fig. 9) (Lipke et al., 1995). When the sequences of the three proposed domains were aligned with each other and with an IgV consensus based on predicted strand profile (Fig. 9) and hydrophobic moment (Eisenberg et al., 1984) (data not shown), there was high conformity to the consensus in all three domains (Table VII). Although there is a low degree of identity in the alignment, the conserved residues include many of the IgV consensus residues. The alignments shown scored significantly better (Z > 3) than did random sequences of the same composition. Residues in α-agglutinin domains I and II corresponding to the consensus positions for the IgV domains include a Cys residue in each domain (the F strand Cys of domain II) and Trp^{55} corresponding to strand C of domain II. There are Met residues in all three proposed α-agglutinin domains in positions analogous to the conserved D-strand Arg in other IgV domains (resides 69, 158, and 274, Fig. 9). In IgV domains, an Asp residue at the beginning of the F strand forms a salt bridge with this Arg, which it could not do with the Met residue in the α-agglutinin. In the three proposed α-agglutinin domains, this Asp is also absent (resides 89, 176, and 293). Although the number of residues conserved among the three domain is low, the three sequences show about 40% similarity (Table VII). The conserved and identical residues are especially frequent at positions conserved in mammalian IgV domains (Fig. 9 and Table VII).

The similarity of domains I and II is also consistent with apparent sequence homology by a standard method. Residues 30–94 and 107–180 can be aligned with a Z score of 4.7 (GCGB EMBL, gap weight 3.0, length weight 0.0; Gribskov and Devereux, 1991). Such a score implies a common ancestral sequence and common structure for these regions, which correspond to strands B to F of domains I and II.

CD Spectra Are Consistent with Induction of α-Agglutinin in the Ig Superfamily—The CD spectrum of α-agglutinin^{20–351} was similar to those of other members of the Ig superfamily, showing little or no α-helix and a predominance of β-sheet. The magnitude of the negative peak at 217 nm characteristic of β-sheet was greater in α-agglutinin^{20–351} than in the spectrum
Fig. 8. Summary of sequenced α-agglutinin20–351 peptides. Regions sequenced from with tryptic and S. auroreus V8 peptides are underlined with solid or wavy lines, respectively. Sulfhydryl groups are marked (SH) and disulfide bonds are marked. Identified O-linked glycosylation sites are marked (solid diamonds). Potential N-glycosylation sites are italicized and striped out; the two identified N-glycosylation sites are marked (stacked solid diamonds).

TABLE VI
Identification of glycosylated residues in α-agglutinin20–351

| Peptide (sequence position) | Sequencer cycle |
|----------------------------|-----------------|
|                            | 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 |
| Tryptic peptides            |                |
| 269–288                    | L Y D G E M L W V N A L Q X s L P A N V N |
| 326–351                    | N L G X t A X s A K X s X s F I X s X t X t X t D L X t X s I N X t X s A |
| S. aureus V8 peptides       |                |
| 241–249                    | W W F P Q S Y X n D |
| 279–295                    | A L Q X s L P A N V X t I D H A L E |
| 296–318                    | F Q Y X t X c L D X t I A X n X t X t Y A X t Q F X s X t X t |
| 319–337                    | F I V Y Q G R N L G X t A X s A K X s X s F I |
| 324–337                    | G R N L G X t A X s A K X s X s F I X s X t X t X t X t D |

of Igs themselves, but was in the range of that for many other members of the Ig superfamily (Cathou and Dorrington, 1975; J. Efferis et al., 1978; Killeen et al., 1988). The CD profile of α-agglutinin20–351 is similar to those of MRC OX-45, CD4, Thy-1, and CD2 (Campbell et al., 1979; Killeen et al., 1988; Chamow et al., 1990; Recny et al., 1990). The mean residue ellipticity at 217 nm for α-agglutinin20–351, Thy-1, and CD2 are −4.68 × 10−1, −4.8 × 10−1, and −6.6 × 10−1 degrees cm2 dmol−1, respectively. The high β-sheet content of α-agglutinin20–351 is also close to that of silk fibroin (Demura and Asakura, 1991) and human plasma fibronectin (Oesterlund, 1988), both of which are mostly antiparallel β-sheet structures and may be close to the maximum possible β-sheet content. Such a high β-sheet content can only be accommodated in globular proteins by antiparallel structures. Therefore, the β-sheet content of α-agglutinin20–351 (70%) is among the highest for known proteins with essentially pure antiparallel β-sheet structures. The unusually high content of antiparallel β-sheet also implies the presence of antiparallel β-sheet structure throughout the molecule and is therefore consistent with the three-domain alignment. It is worth noting that, even if domain III were composed of pure antiparallel β-sheet structure (100% sheet), domain I and II would still have a β-sheet content of at least 50% to yield an overall β-sheet content of 70% in α-agglutinin20–351. Therefore, β-sheet is the predominant structure in all of the domains.

Domain III (residues 200–326) was previously proposed to contribute to the binding site (Cappellaro et al., 1991; Lipke and Kurjan, 1992). Neither the purified α-agglutinin155–351 fragment nor the unpurified Arg-C digest of α-agglutinin20–351 retained activity, despite the retention of most of the secondary structure in the cleaved product. The inactivity of the cleaved product implies that regions of domains I and/or II are also essential for binding. Such contributions of multiple domains to the binding site is the rule in the Ig superfamily, with few exceptions (Williams and Barclay, 1988).

Disulfide Bonds and Free Sulfhydryls in α-Agglutinin20–351

Cys67 and Cys114 form an interdomain disulfide bond between the proposed COOH terminus of domain I and the NH2 terminus of domain II (Figs. 8–10). Interdomain disulfides are known in other members of the Ig superfamily, 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 (Pedraza et al., 1990), but α-agglutinin is unique in the position of the bond between the F and B strands on sequential domains. There are four cysteine residues in domain III, in the A, B, C′, and F strands. Intradomain disulfide linkages in Ig-like domains often form between cysteines of the B and F strands (Williams and Barclay, 1988). Although Cys227 and Cys300 are aligned in positions for the consensus intradomain disulfide bond, Cys202 in strand A and Cys309 in strand F form the actual disulfide linkage. The position of the disulfide Cys residues is not as highly conserved in the Ig superfamily as it is in the antibodies themselves. In domain I of myelin-associated glycoprotein, residues in strands B and E of the IgV domain form an intrasheet disulfide linkage (Pedraza et al., 1990). In domain II of CD4, there is a disulfide between strands C and F (Ryu et al., 1990; Wang et al., 1990). Thus, the bond between the A and F strands in domain III of α-agglutinin is a new position for intradomain disulfides in the Ig superfamily. These strands are close enough to allow formation of the bond (Lipke et al., 1995).

Cys27 in strand B and Cys256 in strand C′ of domain III of α-agglutinin20–351 are free sulfhydryls and can be derivatized
Fig. 9. Alignment of three domains of α-agglutinin with each other and with a consensus sequence for IgV domains (Williams and Barclay, 1988). The positions of the β-strands in the consensus sequence are shown. The alignment is based on secondary structure prediction and alignment within prospective β-strands, with gaps allowed only between strands (Chou and Fasman; Lipke et al., 1995). The sequence between residues 101 and 110 is repeated as the G strand of domain I and the A strand of domain II, as discussed in the text. Identities are boxed and shaded, similarities are boxed without shading. Similarity sets are: A, F, I, L, M, V, Y; A, G; C, S, P; D, E; D, N; E, Q; H, K, R; H, W, Y; N, Q; S, T; φ represents a hydrophobic residue in the consensus and includes A, F, I, L, M, P, V, Y, and W.

| α-Agglutinin domain I | α-Agglutinin domain II |
|------------------------|------------------------|
| 50                     | 61                     |
| 61                     | 50                     |
| 33                     | 33                     |
| 50                     | 50                     |
| 67                     | 67                     |
| 14                     | 14                     |
| 20                     | 20                     |

TABLE VII
Comparison of domains of α-agglutinin

The table shows percent identity (boldface) and percent similarity according to the alignment shown in Fig. 6. Matches to the hydrophobic symbol φ in the consensus sequence are counted as identities.

α-agglutinin20–351. Endo H treatment converts the 21-kDa Arg-C digestion fragment to the 16-kDa fragment, so Asn79, Asn109, or Asn135 must be glycosylated. The 5-kDa size difference would accommodate less than 30 carbohydrate residues, the equivalent of a single N-linked chain in yeast (Hames, 1990; Klis, 1994). The glycosylated residue is probably Asn109 because it is the only Asn-Xaa-Thr sequence in this part of the molecule, and we have repeatedly failed to obtain the sequence from this residue (peptides T1, DT1, and DS2).

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 low density lipoprotein receptor (Goldstein et al., 1985), decay-accelerating factor (Reddy et al., 1989), the muscle-specific isoform of N-CAM (Walsh et al., 1989), and yeast Gas1p/Gpp1p (Gatti et al., 1994) contain clusters of Ser/Thr enrichment segments in the regions proximal to the membrane. Expression of low density lipoprotein receptor and decay-accelerating factor in mutant cells defective for O-glycosylation result in a rapid cleavage of the binding region from the extracellular surface (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) to the COOH-terminal signal for GPI anchor addition at approximately residue 627 (Lipke et al., 1989; Kodukula et al., 1993; Wójcichowicz et al., 1993).

α-Agglutinin expressed in the presence of tunicamycin, which inhibits N-glycosylation, reacts with ConA, indicating the presence of O-linked mannose residues (Terrance et al., 1987). This binding is not due to reaction with modified GPI anchors, because truncated fragments of α-agglutinin lacking the GPI anchor signal also bind ConA (Terrance et al., 1987; Hauser and Tanner, 1989; Wójcichowicz et al., 1993). The pattern of O-glycosylation in α-agglutinin20–351 indicates that there are multiple sites glycosylated after residue 282, which is
Possible O-glycosylation sites COOH-terminal to Asn\(^{348}\) have the sequence Asn-Xaa-Thr which extends to about residue 620. Six additional Asn-Xaa-Thr sequences in this Ser/Thr-rich region are probably glycosylated based on molecular size of truncated proteins in multicellular eukaryotes.

Asn-Xaa-Thr sequences in this Ser/Thr-rich region are probably glycosylated based on molecular size of truncated proteins in multicellular eukaryotes. Węczelewicz et al. (1993). Another possible N-glycosylation site at Asn\(^{338}\) is not shown. Only representative O-glycosylation sites are shown.

at the NH\(_2\)-terminal end of the E strand of domain III. O-Glycosylation is predicted to continue through the Ser/Thr-rich sequence which extends to about residue 620. Six additional Asn-Xaa-Thr sequences in this Ser/Thr-rich region are probably glycosylated based on molecular size of truncated proteins in multicellular eukaryotes. Węczelewicz et al. (1993). This highly glycosylated region (residues 300–627) would form a "stalk" holding the active site out from the wall surface, consistent with electron micrographs (J. entoft, 1990; Cappellaro et al., 1994). Finally, the stalk is predicted to continue to the COOH-terminal GPI anchor, which is processed in vivo to allow linkage to cell wall polysaccharides (L. et al., 1994, 1995).

A drawing of \(\alpha\)-agglutinin shows three sequential Ig domains, with N-glycosylation in sites common for such domains (Fig. 10). The binding site includes residues in domain III and at least one other region. The disulfide bonds between domains I and II and between the A and F strands in domain III are unique among Ig domains, and there are two free sulfhydryls in domain III. Following the Ig domains, there is a heavily N- and O-glycosylated stalk sequence, and the COOH-terminal of the protein is initially GPI anchored. Therefore \(\alpha\)-agglutinin has a structure that recapitulates many of the features of cell adhesion proteins in multicellular eukaryotes.

Acknowledgments—We thank J anet Kurjan and Joseph Krakow for helpful comments on the manuscript. We are grateful to Hans de Nobel for helpful comments and for sequencing of the pAG120–351 insert.

REFERENCES

Cappellaro, C., Hauser, K., Mrsa, V., Watzele, M., Watzele, G., Gruber, C., and Tanner, W. (1991) Mol. Gen. Genet. 229, 414–421

Cappellaro, C., Hauser, K., Mrsa, V., Watzele, M., Watzele, G., Gruber, C., and Tanner, W. (1991) EMBO J. 10, 4081–4088

Cattanach, B., Dorrington, K. (1975) in Subunits in Biological Systems, Part C (Timasheff, S. N., and Fasman, G. D., eds) pp. 91–224, Marcel Dekker, New York

Chow, P. Y., and Fasman, G. D. (1974) Biochemistry 13, 211–222

Crandon, M., and Brock, T. D. (1968) Bacteriol. Rev. 30, 139–163

At the NH\(_2\)-terminal end of the E strand of domain III. O-Glycosylation is predicted to continue through the Ser/Thr-rich sequence which extends to about residue 620. Six additional Asn-Xaa-Thr sequences in this Ser/Thr-rich region are probably glycosylated based on molecular size of truncated proteins in multicellular eukaryotes. Węczelewicz et al. (1993). This highly glycosylated region (residues 300–627) would form a "stalk" holding the active site out from the wall surface, consistent with electron micrographs (J. entoft, 1990; Cappellaro et al., 1994). Finally, the stalk is predicted to continue to the COOH-terminal GPI anchor, which is processed in vivo to allow linkage to cell wall polysaccharides (L. et al., 1994, 1995).

A drawing of \(\alpha\)-agglutinin shows three sequential Ig domains, with N-glycosylation in sites common for such domains (Fig. 10). The binding site includes residues in domain III and at least one other region. The disulfide bonds between domains I and II and between the A and F strands in domain III are unique among Ig domains, and there are two free sulfhydryls in domain III. Following the Ig domains, there is a heavily N- and O-glycosylated stalk sequence, and the COOH-terminal of the protein is initially GPI anchored. Therefore \(\alpha\)-agglutinin has a structure that recapitulates many of the features of cell adhesion proteins in multicellular eukaryotes.

Acknowledgments—We thank J anet Kurjan and Joseph Krakow for helpful comments on the manuscript. We are grateful to Hans de Nobel for helpful comments and for sequencing of the pAG120–351 insert.

REFERENCES

Cattanach, B., Dorrington, K. (1975) in Subunits in Biological Systems, Part C (Timasheff, S. N., and Fasman, G. D., eds) pp. 91–224, Marcel Dekker, New York

C reservoirs, and there are at least one other region. The disulfide bonds between domains I and II and between the A and F strands in domain III are unique among Ig domains, and there are two free sulfhydryls in domain III. Following the Ig domains, there is a heavily N- and O-glycosylated stalk sequence, and the COOH-terminal of the protein is initially GPI anchored. Therefore \(\alpha\)-agglutinin has a structure that recapitulates many of the features of cell adhesion proteins in multicellular eukaryotes.

Acknowledgments—We thank J anet Kurjan and Joseph Krakow for helpful comments on the manuscript. We are grateful to Hans de Nobel for helpful comments and for sequencing of the pAG120–351 insert.

REFERENCES

Cattanach, B., Dorrington, K. (1975) in Subunits in Biological Systems, Part C (Timasheff, S. N., and Fasman, G. D., eds) pp. 91–224, Marcel Dekker, New York

C reservoirs, and there are at least one other region. The disulfide bonds between domains I and II and between the A and F strands in domain III are unique among Ig domains, and there are two free sulfhydryls in domain III. Following the Ig domains, there is a heavily N- and O-glycosylated stalk sequence, and the COOH-terminal of the protein is initially GPI anchored. Therefore \(\alpha\)-agglutinin has a structure that recapitulates many of the features of cell adhesion proteins in multicellular eukaryotes.

Acknowledgments—We thank J anet Kurjan and Joseph Krakow for helpful comments on the manuscript. We are grateful to Hans de Nobel for helpful comments and for sequencing of the pAG120–351 insert.

REFERENCES

Cattanach, B., Dorrington, K. (1975) in Subunits in Biological Systems, Part C (Timasheff, S. N., and Fasman, G. D., eds) pp. 91–224, Marcel Dekker, New York

C reservoirs, and there are at least one other region. The disulfide bonds between domains I and II and between the A and F strands in domain III are unique among Ig domains, and there are two free sulfhydryls in domain III. Following the Ig domains, there is a heavily N- and O-glycosylated stalk sequence, and the COOH-terminal of the protein is initially GPI anchored. Therefore \(\alpha\)-agglutinin has a structure that recapitulates many of the features of cell adhesion proteins in multicellular eukaryotes.

Acknowledgments—We thank J anet Kurjan and Joseph Krakow for helpful comments on the manuscript. We are grateful to Hans de Nobel for helpful comments and for sequencing of the pAG120–351 insert.

REFERENCES

Cattanach, B., Dorrington, K. (1975) in Subunits in Biological Systems, Part C (Timasheff, S. N., and Fasman, G. D., eds) pp. 91–224, Marcel Dekker, New York

C reservoirs, and there are at least one other region. The disulfide bonds between domains I and II and between the A and F strands in domain III are unique among Ig domains, and there are two free sulfhydryls in domain III. Following the Ig domains, there is a heavily N- and O-glycosylated stalk sequence, and the COOH-terminal of the protein is initially GPI anchored. Therefore \(\alpha\)-agglutinin has a structure that recapitulates many of the features of cell adhesion proteins in multicellular eukaryotes.

Acknowledgments—We thank J anet Kurjan and Joseph Krakow for helpful comments on the manuscript. We are grateful to Hans de Nobel for helpful comments and for sequencing of the pAG120–351 insert.

REFERENCES