Plus and Minus Sexual Agglutinins from Chlamydomonas reinhardtii

Patrick J. Ferris,a Sabine Waffenschmidt,b James G. Umen,a,1 Huawen Lin,a Jae-Hyeok Lee,a Koichi Ishida,a Takeaki Kubo,a,2 Jeffrey Lau,a and Ursula W. Goodenougha,3

a Department of Biology, Washington University, St. Louis, Missouri 63130
b Institut für Biochemie, Universität zu Köln, Köln, Germany 4750674

Gametes of the unicellular green alga Chlamydomonas reinhardtii undergo sexual adhesion via enormous chimeric Hyp-rich glycoproteins (HRGPs), the plus and minus sexual agglutinins, that are displayed on their flagellar membrane surfaces. We have previously purified the agglutinins and analyzed their structural organization using electron microscopy. We report here the cloning and sequencing of the Sag1 and Sad1 genes that encode the two agglutinins and relate their derived amino acid sequences and predicted secondary structure to the morphology of the purified proteins. Both agglutinin proteins are organized into three distinct domains: a head, a shaft in a polyproline II configuration, and an N-terminal domain. The plus and minus heads are related in overall organization but poorly conserved in sequence except for two regions of predicted hydrophobic α-helix. The shafts contain numerous repeats of the PPSPX motif previously identified in Gp1, a cell wall HRGP. We propose that the head domains engage in autolectin associations with the distal termini of their own shafts and suggest ways that adhesion may involve head–head interactions, exolectin interactions between the heads and shafts of opposite type, and antiparallel shaft–shaft interactions mediated by carbohydrates displayed in polyproline II configurations.

INTRODUCTION

Sexual adhesion between gametes of the green alga Chlamydomonas reinhardtii is one of the earliest cell–cell recognition systems to have been subjected to experimental analysis. In pioneering studies, Wiese documented that the mating-type-specific components that induce mating-type plus gametic flagella to adhere to mating-type minus gametic flagella could be recovered from the culture medium (Förster and Wiese, 1954; Wiese, 1965). Subsequent analyses of such gamone preparations revealed them to contain flagellar membrane vesicles (Bergman et al., 1975; Snell, 1976; Musgrave et al., 1981), and further studies eventually led to the purification and extensive characterization of their component agglutinin proteins in both C. reinhardtii and the distantly related C. eugametos (Adair et al., 1983; Cooper et al., 1983; Collin-Osdoby et al., 1984; Adair, 1985; Goodenough et al., 1985; Collin-Osdoby and Adair, 1985; Samson et al., 1987; Homan et al., 1988; Goodenough and Adair, 1989).

The agglutinins are displayed by nitrogen-starved gametes but not by mitotic vegetative cells. The plus and minus versions from C. reinhardtii are encoded by different genes and possess complementary adhesive properties, but they also share common features: (1) they are both huge monomeric glycoproteins (native molecular mass > 1000 kD; Adair et al., 1983); (2) their association with the flagellar surface is disrupted by EDTA (Adair et al., 1982), indicating that they are extrinsic membrane proteins; (3) they are both chimeric (Kieliszewski and Lamport, 1994), possessing a large globular head and a fibrous shaft (Goodenough et al., 1985); (4) they are members of the Hyp-rich glycoprotein (HRGP) family (Cooper et al., 1983) found as well in the cell walls of most green organisms (reviewed in Cassab, 1998; Serpe and Nothnagel, 1999) and implicated in sexual interactions in higher plants (reviewed in Wu et al., 2001).

The HRGPs of the Chlamydomonas cell wall self-assemble into dense fibrous meshworks, portions of which are subsequently stabilized by covalent cross-linking and portions of which are chaotrope-soluble (Hills et al., 1975; Homer and Roberts, 1979; Goodenough et al., 1986; Goodenough and Heuser, 1988; Waffenschmidt et al., 1993, 1999; Ferris et al., 2001). Agglutinin adhesion in Chlamydomonas also results in the formation of salt-sensitive plus/minus agglutinin meshworks (Goodenough, 1986; Goodenough and Heuser, 1999). These migrate in the plane of the membrane to the flagellar tips (Goodenough, 1983) in conjunction with an adhesion-associated rise in intracellular levels of cAMP (Pasquale and Goodenough, 1987; reviewed in Pan and Snell, 2000). cAMP elevation also rapidly elicits the downstream events of the mating reaction (cell wall disassembly and mating-structure activation) that culminate in fusion between pairs of plus and minus gametes to form diploid zygotes (reviewed in Snell, 1985; Goodenough, 1991; Beck and Haring, 1996).

Genetic screens of C. reinhardtii have yielded nonagglutinating mutants in both mating types (reviewed in Goodenough et al.,...
The three mutations that generate nonagglutinating minus strains (imp10, imp12, and agl) all map to the Sad1 (sexual adhesion) gene that resides in the mating-type minus (mt) locus in linkage group VI (Hwang et al., 1981; Matsuda et al., 1988; Ferris et al., 2002) and are hereafter called sad1-1 to sad1-3. (A functional allele of the Sad1 gene is also located in the mt locus [Ferris et al., 2002] but is expressed only when plus cells are induced to differentiate as minus [Galloway and Goodenough, 1985; Ferris and Goodenough, 1997] and is not considered further in this report.) Most mutations affecting plus agglutination map to the Sag1 (sexual agglutination) locus (Goodenough et al., 1978) that does not reside in the mt locus and has recently been mapped near the centromere of linkage group VIII (S.K. Dutcher, personal communication; see Bowers et al., 2003). These mutations, formerly known as imp2, imp5, imp6, imp7, and imp9, are herein designated sag1-1 to sag1-5. An insertional allele of the gene sag1-6 is described in this article.

We report here the cloning and predicted amino acid sequences of the Sad1 and the Sag1 genes from C. reinhardtii and suggest relationships between their predicted primary and secondary structures and the morphology of purified plus and minus agglutinin proteins.

RESULTS

Identification of the Sag1 Gene Encoding Plus Agglutinin

As detailed in Methods, an insertional mutation generating a non-adhesive phenotype was genetically mapped to the Sag1 locus, and DNA flanking the insertion was used to identify the wild-type Sag1 gene in genomic and cDNA libraries. By RNA gel blot analysis (Figure 1A), Sag1 probes hybridize to an ~11-kb transcript that is absent from vegetative cells and expressed exclusively in gametes that agglutinate as plus. Included in the blot are two mt− strains, mid-1 and mid-2, that carry mutations in the mt−-localized Mid gene necessary for minus gametic differentiation (Ferris and Goodenough, 1997); both strains agglutinate as plus and express the Sag1 gene as gametes. Included also is the iso1 (isoagglutination) mt− strain wherein some cells differentiate and agglutinate as plus and some as minus (Campbell et al., 1995); Sag1 expression is again observed. No transcript is detected in the sag1-6 insertional mutant (Figure 1C), whereas the five other UV-induced mutants (sag1-7 to sag1-12) that mark the Sag1 locus all express transcripts at varying levels (Figure 1C); DNA gel blot analysis indicates that none of these strains has incurred an indel > 0.5 kb (data not shown).

Because several of the sag1 mutants have been shown to lack agglutinin on their flagellar surfaces (Adair et al., 1983; Goodenough et al., 1985), they presumably carry either nonsense mutations or missense mutations that disallow proper folding or flagellar targeting.

Identification of the Sad1 Gene Encoding Minus Agglutinin

We have cloned both the mt− and the mt− loci and have recently subjected ~600 kb of each to RNA gel blot analysis (Ferris et al., 2002). Probes from the ~70 to ~90 region of the C domain of the mt− locus (Ferris and Goodenough, 1994; Ferris et al., 2002) hybridize to an ~12-kb message that is expressed exclusively in minus gametes (Figure 1B). As expected (see above), the iso1 mt− gametes also express the gene (Figure 1B). The mid-1 and mid-2 mutants fail to express the gene, documenting that Sad1, like other minus-specific genes, is dependent on the Mid protein for expression (Ferris and Goodenough, 1997). The sad1-1, sad1-2, and sad1-3 mutants express little if any transcript (Figure 1D); sad1-1 carries a 506-bp deletion in the proximal region of the coding sequence (Figure 2B); no indel >0.5 kb is detectable in sad1-2 and sad1-3 coding sequences (data not shown).

Cloning and Sequencing the Sag1 and Sad1 Genes

As detailed in Methods, sequencing the enormous Sag1 and Sad1 genes entailed sequencing genomic DNA, identifying putative open reading frames (ORFs) and introns by codon usage, sequencing available cDNAs derived from 3’ ends of the genes, and the use of RT-PCR to verify intron positions and 5’ rapid amplification of cDNA ends (RACE) to obtain N-terminal sequences (Figure 2).

Figure 2A shows the structure of the Sag1 gene, 14,776 bp in length with 15 exons, where the position of the Cry1 plasmid insertion (sag1-6) that identified the gene is indicated. The predicted mRNA encodes a protein of 3409 amino acids; with the predicted signal peptide removed, the protein would be 3349 amino acids, with a molecular mass of 330 kD. An estimate of 400 kD was proposed for the hydrogen fluoride–deglycosylated plus agglutinin (Adair et al., 1983), where the discrepancy is likely attributable both to incomplete deglycosylation and to the reduced ability of (hydroxyl)proline-rich peptide domains to bind SDS (Godl et al., 1997).

Figure 2B shows the structure of the Sad1 gene, 17,843 bp in length with 30 exons; introns in similar locations to introns in the Sag1 gene are labeled A to D. The predicted mRNA encodes a protein of 3889 amino acids, or 3853 amino acids with the signal peptide removed, with a molecular mass of 385 kD. The position of the 506-bp deletion carried by the sad1-1 mutant is indicated.

General Relationships between the Plus and Minus Agglutinin Sequences and Agglutinin Protein Morphology

Figures 3A and 3B show the derived amino acid sequences of the plus and minus agglutinin proteins from C. reinhardtii. Three domains are evident: (1) a large C-terminal domain (2006 amino acids in plus, pl 9.71; 2404 amino acids in minus, pl 7.50) with scattered putative N-glycosylation sites (12 in plus, 14 in minus); (2) a central domain (934 amino acids in plus, pl 3.84; 873 amino acids in minus, pl 4.12) rich in P residues (60% P in both plus and minus); (3) a smaller N-terminal domain (409 amino acids in plus, pl 4.03; 576 amino acids in minus, pl 5.60, with signal peptides removed) with putative N-glycosylation sites (three in plus, 10 in minus). The locations of these domains are also indicated in Figure 2.

Purified agglutinin proteins from C. reinhardtii, when visualized by transmission electron microscopy (TEM) after adsorption to
mica, quick-freeze deep etching (QFDE), and rotary replication with platinum (Heuser, 1983; Goodenough et al., 1985), are shown in Figures 4A (plus) and 4B (minus) and are diagrammed in Figure 4F. Each displays distinct domains: a large globular head and a long fibrous shaft that terminates in a pronounced tail hook. Images of intact flagella show that the tail-hook ends associate with the membrane surface and the heads extend outward (Goodenough et al., 1985; Goodenough and Heuser, 1999).

Circular dichroism (CD) spectra of purified agglutinins display the characteristics of polyproline II (P II) helices (Figure 5), with diagnostic extrema at 187 and 217 nm and molar ellipticities of −10,800 and 2000 and −11,000 and 2100 deg cm² × dmol⁻¹. Polypeptides adopt the PII configuration when they are rich in Pro or its posttranslational derivative, Hyp (reviewed in Creamer, 1998). Therefore, the central P-rich domains are proposed to adopt PII configurations, undergo hydroxylation and glycosylation, and self-organize as fibrous shafts.

The globular head of the minus agglutinin measures 11.5 ± 1.1 nm in QFDE replicas (Goodenough et al., 1985); for comparison, the globular head of the γ-heavy chain of outer-arm dynein,
which possesses a hollow central core, measures 12.6 ± 0.9 nm using the same technique (Goodenough et al., 1987). The minus C-terminal domain contains 2404 amino acids; the γ-heavy chain contains 2405 amino acids. These correspondences indicate that the C-terminal domains correspond to the agglutinin heads. No significant matches to the head sequences are found in the GenBank database.

The smaller (~500 amino acids) N-terminal domains, not recognized in our previous study (Goodenough et al., 1985; see below), are adjacent to the tail hooks and positioned to mediate the binding of agglutinins to the flagellar membrane. These sequences also lack significant GenBank matches.

Comparison of the Heads of the Plus and Minus Agglutinins

By OFDE-TEM, the heads of the minus agglutinins are invariably globular or somewhat oblate after adsorption to mica (Figure 4B; Goodenough et al., 1985). By contrast, the plus heads, although often also globular (Figures 4A and 4D, head 1), frequently appear as bilobed structures (Figure 4D, heads 2 and 3), apparently as the consequence of denaturation at the time of mica adsorption. The lobe proximal to the shaft is usually globular, whereas the distal lobe tends to splatter (Figure 4D, heads 2 and 3).

Figure 6 presents a computer-generated alignment of the plus and minus head amino acid sequences (alignment length = 2423 amino acids), wherein 23% (557 positions) are scored as identical, 13% (312 positions) strongly similar, 17% (428 positions) weakly similar, and 46% (1126 positions, many generated via indels) without similarity. Except for the α1 and α2 subdomains described below, the limited sequence similarity between the two heads is patchy and of dubious significance. Of the putative N-glycosylation sites (12 in plus, 14 in minus; see Figure 3), six are in similar locations.

Figure 6 also displays computer-predicted secondary structures of the two head domains as determined by two algorithms.
Despite their low sequence similarity, the heads are organized into six comparable subdomains. (1) The α subdomain (aligned residues 1 to 599; see legend to Figure 6 for color coding of subdomains), adjacent to the C-terminal end of the shaft, carries scattered predicted hydrophilic α-helices and β-strands. Four C residues mark the first 50 amino acids in the shaft/α transition (two additional Cs are found in the minus sequence, residues 342 and 404), and half of the predicted N-glycosylation sites in the head (6/12 for plus, 7/14 for minus) are found in this subdomain. A predicted hydrophobic α-helix is found in a similar location around aligned residue 480 (asterisks). (2) The P-rich subdomain (residues 600 to 780), with strings of up to five contiguous P residues, carries two predicted N-glycosylation sites in minus and lacks C residues in both mating types. (3) The a1 subdomain (residues 781 to 905) includes long tracts of predicted hydrophobic α-helix, with two Cs in each sequence and no predicted N-glycosylation sites. (4) The long unstructured subdomain (residues 906 to 2097) displays little predicted secondary structure, carries two Cs in plus and three in minus, and contains most of the remaining predicted N-glycosylation sites (four in plus and four in minus). (5) The a2 subdomain (residues 2098 to 2265) carries long tracts of predicted hydrophobic α-helix. Most of the remaining C residues (three in plus and five in minus) also localize in a2, and there is one conserved predicted N-glycosylation site. (6) The short C-terminal subdomain (residues 2266 to 2423) is poor in predicted secondary structure and carries one predicted N-glycosylation site and one C residue in plus.

The a1 and a2 subdomains are of particular interest in that their predicted hydrophobic α-helices are the only head sequences that display strong plus/minus sequence homology (Figure 6). Of the 80 amino acids in aligned α-helixes in a1, 38 (46%) are identical and 38 (25%) are scored as strongly similar; of the 102 amino acids in aligned α-helixes in a2, 54 (53%) are identical and 15 (15%) strongly similar. Many of the identical amino acids are found in blocks of four to eight residues.

Figure 3. Predicted Amino Acid Sequences of the Sag1 and Sad1 Agglutinin Proteins.

(A) Sag1.
(B) Sad1.
Numbered arrowheads indicate the locations of introns (the first intron in both sequences is in the 5′ untranslated region); putative N-glycosylation sites are underlined, and predicted signal sequences are boxed. N-terminal, shaft, and head domains are described in text. a.a., amino acids.
Shafts of the Plus and Minus Agglutinins:

General Considerations

The 934–amino acid shaft of the plus agglutinin contains 135 iterations of the PPSPX motif; the 873–amino acid shaft of the minus agglutinin contains 122 iterations of the PPSPX motif (Figure 3). These features place the agglutinins in the PPSPX family first defined by the cell wall protein Gp1 (Figure 4C; Ferris et al., 2001) and a gamete-specific protein of unknown function called Mta2 (Ferris et al., 2001). ORFs encoding (PPSPK)20 have also been identified in the genome of a large Chlorella virus (Eriksson et al., 1999; Van Etten, 2003).

In repetitive extensin-type HRGPs in higher plants, most of the shaft Pro are posttranslationally converted to Hyp (Kieliszewski and Lamport, 1994), and when Pro-rich peptides derived from two Volvox HRGPs were directly sequenced, the Pro residues were all found to be hydroxylated (Ertl et al., 1989, 1992). The following considerations indicate that Pro residues in the agglutinin shafts are largely if not completely hydroxylated as well. (1) Expressed as percentage of total amino acids, the purified plus agglutinin protein has been estimated to contain 12.3% Hyp and 4.4% Pro (Cooper et al., 1983). Of the 3349 amino acids in the deduced plus sequence (signal peptide removed), 560 of the 761 Pro residues reside in the shaft; hence, 16.7% of the total amino acids are predicted to be shaft-Pro and 6% non-shaft-Pro. The correspondence between these percentiles indicates that the Pro residues in the shafts are converted to Hyp. (2) By the same rubric, the purified minus agglutinin contains 12% Hyp and 6% Pro (Collin-Osdoby and Adair, 1985), and the minus agglutinin (3853 amino acids), with 526 of its 788 Pro residues in the shaft, carries 13.7% shaft-Pro and 6.8% non-shaft-Pro. The correspondences again indicate that the shaft Pro residues are converted to Hyp. (3) These correspondences are also seen in the cell wall protein Gp1, where amino acid analysis gives 14.7% Hyp/7.7% Pro (Goodenough et al., 1986), and sequence analysis gives 13.2% shaft-Pro and 7.4% non-shaft-Pro.

Although the two agglutinin shafts are different in primary sequence, five similar subdomains can be recognized (Figures 3, 4C).
7A and 7B); these are called 2A to 2E (where the shaft is the second of the three agglutinin protein domains). Each of the central 2C subdomains contains repeating motifs of a specific PPSPX motif (as detailed below). The flanking 2B and 2D subdomains are also largely made up of PPSPX motifs, but these are not reiterated in repetitive units; each 2D subdomain also carries two runs of non-PPSPX units (brackets). The 2A subdomains, contiguous to the N-terminal domains, lack PPSPX motifs, are dominated by PPX units, and contain a high density of basic amino acids. The 2E subdomains, contiguous to the head domains, are a mix of PPX, PPSPX, and nonreiterative sequences.

H. Tran and R. Pappu (personal communication) have analyzed the propensity of non-P residues to destabilize a PII helical configuration. By their algorithms, the block interruptions that are set off in brackets in each 2A subdomain (Figure 7) would fail to adopt the PII configuration. Their algorithms further predict that whereas the occasional presence of G, I, L, T, and V residues in the remaining shaft sequences would introduce local interruptions in helical integrity, the dominant amino acids in the sequences (A, E, K, P, Q, R, and S) are all fully compatible with PII helix formation.

The shaft domains of the agglutinin genes share two features with the shaft-encoding sequences of other HRGP-encoding genes in \textit{C. reinhardtii} listed in Supplemental Table 1 online. First, they are devoid of introns, in contrast with the globular domains of these genes, which invariably contain at least one and often numerous introns. Second, the usual strong bias in \textit{C. reinhardtii} toward the use of CCC or CCG as the P codon (Naya et al., 2001) is retained in the globular domains but relaxed in the shafts (see Supplemental Table 1 online). A possible interpretation of this observation is that the presence of the CCA and CCT codons renders the DNA less strongly hydrogen bonded and, hence, more readily opened up for replication and transcription; this feature may be under stronger selection in P-rich sequences than is selection for codon preference at the translational level.

The Shaft of the \textit{Plus} Agglutinin: Specific Features

In our original measurements (Goodenough et al., 1985), the purified \textit{plus} agglutinin was reported to be $228 \pm 7$ nm in length; current measurements of digitized images with computer-based measuring programs indicate that the visible \textit{plus} shaft is $244.5 \pm 6.2$ nm ($n = 17$) in length. If the full predicted shaft sequence, subtracting the block interruption, were to adopt a PII helix (3.34 amino acids/nm) it would be 275 nm in length. A possible explanation for the 30-nm discrepancy is offered in a later section.

The \textit{plus} shaft displays a consistent overall morphology after mica adsorption (Figures 4A and 4D): a tail hook at the nonhead

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure4}
\caption{Morphology of Agglutinins and Related Cell Wall Proteins.}
\textbf{(A)} \textit{Plus} agglutinin. Curved fibril protrudes beneath the large globular head.
\textbf{(B)} \textit{Minus} agglutinin.
\textbf{(C)} Gp1 (upper) and Gp2 (lower) cell wall HRGPs. The kink in Gp1 is curved (left) or acute (right).
\textbf{(D)} \textit{Plus} agglutinins. Arrows indicate globular domains at the proximal shaft termini. Head 1 is globular; heads 2 and 3 are bilobed; head 4 displays a protruding curved fibril, shown at higher magnification in the inset.
\textbf{(E)} \textit{Plus} shaft showing head loop free of its globular head domain.
\textbf{(F)} Diagram of the agglutinins. All images generated by QFDE-TEM. Bar = 50 nm.}
\end{figure}
terminus is followed by a more flexible region, followed by a straight region that extends to the head. In images where the head has dissociated from the shaft, a head loop can also be visualized. These regions are related to the shaft subdomains in the sections below.

The Tail Hook (Subdomain 2A)

By QFDE-TEM, the plus (and minus) shafts display a prominent kink at a conserved position, 12 ± 2 nm (n = 23) from the N terminus, whose angle ranges from acute to curved to nearly straight (Figures 4A, 4B, 4D, and 4F; Goodenough et al., 1985). As noted earlier (Figure 7), the predicted amino acid sequence of the plus shaft terminus carries a block interruption, TPVAR-CIOVGGICD, located at a position corresponding to 19 nm from the terminus. We therefore propose that the block interruption adopts a structural conformation that confers the shaft with the capacity to bend. In the shaft of the Gp1 cell wall protein, a sequence discontinuity, PRPPFPANTPM, also colocalizes with a morphological kink (Ferris et al., 2001), shown in Figure 4C, that has been posited to participate in cell wall assembly (Goodenough and Heuser, 1988).

The kink defines the midpoint of a terminal bend, 24.3 ± 3.7 nm (n = 23) in length, that we call the tail hook (Figure 4F), and we propose that subdomain 2A forms the tail hook, where the block interruption is flanked by a proximal (19 nm predicted) and a distal (14 nm predicted) segment. The plus 2A subdomain lacks PPSPX motifs, is dominated by PPX blocks (where X is often P), and is in general S-poor.

The Medial Region (Subdomains 2B to 2D)

The long medial region of the plus shaft is dominated by PPSPX repeats (Figure 7A). Of these, a central 2C subdomain carries iterated versions of particular repeats, whereas the flanking 2B and 2D subdomains are not reiterated.

The 2C reiterations can in some cases be recognized by repetitive use of particular X amino acids, but they are more reliably identified by the nucleotide sequences of codons. Using nucleotide-based analyses, the plus 2C sequence is found to have been generated by two sets of endoduplication events, indicated by the parentheses in Figure 7A, the first iterated three times and the second twice. Because these analyses have been conducted in the context of an ongoing comparative study of the shafts of C. reinhardtii and its sibling species C. incerta, the data generating these conclusions will be presented in a future publication.

The Head Loop (2E)

In our previous publication (Goodenough et al., 1985), we noted that the plus head occasionally denatures upon mica adsorption to reveal an underlying curved fibril (Figures 4A and 4D, head 4, magnified in the inset). When the heads are absent, a condition stimulated by protease digestion or disulfide reduction/alkylation, the fibril is seen to form a loop (Figure 4E). These head loops measure 26.3 ± 3.3 nm (n = 14) in length.

The 2E subdomain (Figure 7A) is predicted to be 25 nm in length and carries a high density of amino acids (G, L, V, and T) that are predicted to destabilize the P II helix (H. Tran and R. Pappu, personal communication) and hence might generate the propensity to curve. We therefore propose that the head loop corresponds to the 2E subdomain and suggest that the head normally assembles around the 2E sequences such that the head loop is masked from view. This could explain the 30-nm discrepancy, noted earlier, between the measured and predicted shaft length.

In platinum replicas, the agglutinin shaft proper is twice as thick (6 nm) as the head-loop shaft (3 nm) (Figure 4E). As detailed in the Discussion, these observations are consistent with the hypothesis that 2E is glycosylated differently from the bulk of the shaft.

The Shaft of the Minus Agglutinin: Specific Features

Measurements of native minus agglutinins yield a mean length of 224.9 ± 14.1 nm (n = 30) (predicted length 258 nm), with a wider range of values (202 to 248 nm) than the plus agglutinins (232 to 250 nm). Whereas, as noted above, the plus shafts display a consistent morphology, the minus shaft exits the head in...
Figure 6. Sequences of the Plus and Minus Agglutinins Heads Aligned by Computer Algorithms and Marked with Predicted Regions of Secondary Structure (See Icon Key at Top Left).

Aligned amino acids are color coded (key at top left) to indicate their level of similarity. The $\alpha\beta$ subdomains are shaded pink, followed by white P-rich subdomains, gray $\alpha_1$ subdomains, white long-unstructured subdomains, yellow $\alpha_2$ subdomains, and then white C-terminal subdomains. P, plus agglutinin heads; M, minus agglutinin heads.
Figure 7. Subdomain Organization of the Plus and Minus Agglutinin Shafts.
a variety of configurations, and the lengths and positions of straight versus flexible segments are varied. A possible contributor to this variability is the presence of more numerous helix-disrupting amino acids (G, I, L, T, and V) in the minus (60 residues) than the plus (48 residues) sequences and their high density in the 2D subdomain of the minus sequence where the shaft emerges from the head.

The central 2C subdomain initiates with 42 PPSPX units that reiterate the sequence PPSPPEPPSPA PPSPPE, with eight imperfect units (Figure 7B). The next 30 PPSPX units no longer display the serial E, A, and P reiterations. However, nucleotide-based alignments, to be presented in a future publication, indicate that this region is related to the reiterated repeat, with overlapping codon usage for PPSP but relaxed usage at the X position. Although the usage of A and P residues is no longer constrained, an E residue appears in every third motif, meaning that the minus 2C sequence, like the plus 2C sequence, is strongly negatively charged.

The head of the minus agglutinin has not been observed to denature on mica, and we have not subjected minus proteins to proteolysis or reduction/alkylation. We therefore have no direct evidence that the minus 2E subdomain (predicted length 25 nm) forms a head loop like plus 2E, but propose, by virtue of the presumed common ancestry of the two proteins, that this is the case and that the minus head associates with, and covers, the 2E end of the minus shaft.

Although different in primary sequence, the 2A subdomains of the plus and minus agglutinins are similar in predicted length (33 and 42 nm), lack PPSPX motifs, are S-poor and enriched in basic amino acids, and are dominated by blocks of PPX. They both also carry block interruptions in comparable positions, but whereas the plus version (TPVARIQVGGICD) carries two Cys and no aromatic amino acids, the minus version (TWPGKWE-GAWPFR) lacks Cys and has four aromatic amino acids. As argued for the plus protein, we propose that the minus block interruption functions to generate the kink in the minus tail hook (total length of minus tail hook = 22.9 ± 2.2 nm [n = 13]).

The N-Terminal Domains of the Plus and Minus Agglutinins

In our previous study (Goodenough et al., 1985), QFDE-TEM images showed the shafts to end abruptly at the termini of the tail hooks. We were therefore surprised to encounter, in each amino acid sequence, a non-P-rich domain extending from the end of the 2A sequence to the predicted signal sequence (Figures 3A, 3B, and 8). After cleaving the signal sequences, these domains would contain 409 and 576 amino acids, respectively. Both contain putative N-glycosylation sites (three in plus and 10 in minus, with one site conserved); each is enriched in C residues (10 in plus and 16 in minus, with six conserved); their intron profiles are strikingly different (Figure 2); neither displays significant sequence homologies to one another (Figure 8), to the head domains, nor to other proteins in the GenBank database; and neither carries predicted transmembrane α-helices nor putative N-myristylation sequences.

Reexamination of the QFDE images reveals globular domains at the termini of tail hooks in some agglutinin preparations (Figure 4D, arrows) but not others (Figure 4A). Possible explanations for this variability include a vulnerability of this domain to proteolysis during protein purification or the possibility that the domain may be capable of forming a sheath-like association with the tail hook so that it does not appear globular.

The N-terminal domains are positioned to mediate the association of the agglutinins with the flagellar membrane, an EDTA-sensitive interaction (Adair et al., 1983) that is presumably mediated by trans-flagellar membrane agglutinin-anchoring proteins.

DISCUSSION

The Plus and Minus Agglutinin Proteins: General Considerations

We have cloned and sequenced the genes encoding the plus and minus sexual agglutinins of C. reinhardtii, the former identified by insertional mutagenesis at a site allelic to mutations that generate nonagglutinating plus sag1 mutants, the latter by RNA gel blot analysis of the cloned mt- locus, where nonagglutinating minus sad1 mutants have been mapped. Gene expression is restricted to gametes of one mating type and is disrupted in some but not all of the nonagglutinating mutants: it is undetectable in the insertional mutant sag1-6 and in the sad1-3 mutant, faint in the sad1-1 and sad1-2 mutants, and detectable to various levels in the sag1-1 to sag1-5 strains, indicating that they carry nonsense or missense mutations.

Sag1 and Sad1 transcripts continue to be present in zygotes 30 min after plus and minus gametes are mixed, but they are undetectable 1.5 h later (data not shown). This same pattern of expression has also been observed for two other gamete-specific genes, Fus1 (Ferris et al., 1996) and Mid (Ferris and Goodenough, 1997). Therefore, transcription of gamete-specific genes appears to terminate rapidly in the zygote, and the transcripts are apparently short-lived.
The sequences of the plus and minus agglutinin genes put to rest earlier speculations that the agglutinin polypeptides might be assembled posttranslationally from smaller units (Goodenough et al., 1985): each ~12-kb transcript encodes a protein whose predicted secondary structure conforms to the morphology of the molecule visualized by electron microscopy.

Available evidence indicates that sexual adhesion in Chlamydomonas entails direct agglutinin-agglutinin interactions. If it were instead the case, for example, that the plus agglutinin recognized a second agglutinin receptor protein on the minus flagellum, one would expect that the sad1 mutants would continue to express this receptor and hence display at least
weak adhesive interactions with plus gametes, which is not the case. Antibody inhibition studies with C. eugametos also indicate that the agglutinins alone participate in adhesion (Homan et al., 1988). That sexual adhesion is effected by specific proteins distinguishes Chlamydomonas and yeasts (Shen et al., 2001) from organisms wherein the sexual (glyco)proteins displayed by the eggs are unrelated to those displayed by the sperm/pollen (reviewed in Vacquier, 1998; Swanson and Vacquier, 2002).

QFDE-TEM images of sexual adhesion in Chlamydomonas show complex meshworks of fibers that disallow identification of specific protein domains (Goodenough and Heuser, 1999). However, images of HRGP assemblages in Chlamydomonas cell walls (Goodenough and Heuser, 1988) and in hammock mastigonemes (Goodenough and Heuser, 1985) in this fashion suggest that shaft termini, shafts proper, shaft kinks, and globular heads all have the capacity to participate in intermolecular interactions between algal HRGPs.

Given that sexual adhesion in Chlamydomonas is a challenging interaction—flagella that are beating ~50 times per second must establish sufficient adhesiveness to arrest flagellar movement and bring the cell bodies sufficiently close to achieve cell fusion—it follows that agglutination may entail several modes of adhesive interactions that collectively achieve this outcome. Therefore, we consider below possible ways that protein–protein, protein–carbohydrate, and carbohydrate–carbohydrate interactions may participate in sex-specific and/or species-specific recognition/adhesion events, where it is now possible to test these ideas by expressing partial gene constructs (head-only and shaft-only) and analyzing the adhesive properties of their protein products.

The Head Domains

The agglutinin heads are the most outward-facing domains of the agglutinins on the flagellar surface (Goodenough et al., 1985; Goodenough and Heuser, 1999), and they are therefore positioned to initiate the adhesion reaction. Adhesive proteins often carry internal repeats of putative adhesive motifs (Bierman, 1998; Swanson and Vacquier, 1998; Gao and Garbers, 2001; Swanson et al., 2001), but the agglutinin heads carry no discernable repeats, nor do they carry sequences with suggestive homologies to known adhesive domains.

Confounding any analysis of the heads is their remarkable, and unexplained, size. Even if the head proves to carry several functional subdomains that participate in distinct facets of a multistep adhesion process, it is difficult to imagine a role for ~2400 amino acids, particularly because the head of its sister cell wall protein, Gp1, binds to several ligands using a mere 180 amino acids (Goodenough and Heuser, 1988; Ferris et al., 2001).

With the exception of the hydrophobic α-helix–rich regions considered below, the plus and minus heads display no significant sequence homology to one another, and whereas the minus head is invariably globular in QFDE-TEM images, the plus head is often bilobed (Figure 4D). Despite these differences, the agglutinin head domains share four intron positions (Figure 2), and their overall topological organization is homologous: six subdomains each display similar predicted secondary structures and similar endowments of C residues and N-glycosylation sites (Figure 6), indicating that both heads derive from a common, albeit distant, ancestral protein domain. Of the six head subdomains, the most N-terminal (α9) carries the majority of the predicted N-glycosylation sites and hydrophilic α-helical domains and is predicted to be surface localized, whereas the α1 and α2 subdomains, each carrying long tracts of hydrophobic α-helix, are predicted to localize within the head interior. More than half of the amino acids reside in the long unstructured subdomains in the central portions of the sequences.

The α1 and α2 subdomains include predicted hydrophobic α-helices that are ~50% identical in sequence between the plus and minus proteins of C. reinhardtii. These observations suggest that the α1 and α2 regions in general, and the conserved α-helical residues in particular, confer important head domain properties that are independent of, albeit perhaps necessary for, sex-specific adhesion. One possibility is that they participate in the postulated association of the heads with the head-loop subdomains of the shafts (see below), either directly or by generating a conformation that allows such associations to occur. A second possibility is that helix–helix interactions within each subdomain, and perhaps between α1 and α2, create folds necessary to carry the long unstructured subdomains.

Head–Shaft Interactions

While head–head interactions may initiate adhesion, images of fully adhered flagellar membranes show that they can be far closer than the >500 nm predicted if head–head interactions alone are involved (Goodenough and Heuser, 1999). A second possible agglutination modality, therefore, is that the heads make adhesive contacts with opposite-type shafts.

In Solanaceous plants, chimeric HRGPs (the solanaceous lectins) have been shown to recognize sugar residues in chitin (Kieliszewski et al., 1994; Van Damme et al., 2004), and the Cys-rich extensin-like proteins expressed in the flowers of Nicotiana tabacum have also been suggested to function as lectins (Wu et al., 2001). By analogy, the agglutinin heads may prove to include surface-localized lectin motifs that recognize distinctive carbohydrate moieties carried by the shafts of the opposite mating type, perhaps displayed in a recognized pattern. If these moieties are repetitive, then the heads may ratchet along the shafts during the adhesion process, abetting the apposition of interacting flagella.

Shaft Glycosylation and Shaft–Shaft Interactions

The plus and minus shafts, and the shaft of the cell wall protein Gp1, adopt the P11 helical configuration (Ferris et al., 2001; Figure 5) and are dominated by the PPSX motif over most of their lengths. The similarly wide caliber of the Gp1 and agglutinin shafts (Figures 4A to 4C) suggests that they carry sugar endowments that trap the platinum applied during rotary shadowing in a similar fashion, but given that the three shaft sequences are different from one another at the amino acid level, notably at their X positions, they may acquire distinctive patterns of sugar residues that participate in wall assembly and in sexual adhesion.
Motifs found in Hyp-rich sequences of plant proteins have been proposed to serve as glycomodules (Shpak et al., 1999; Kieliszewski, 2001), directing Hyp-glycosyltransferases to add specific sugar residues to specific Hyp residues in the polypeptide chain (Kieliszewski and Lamport, 1994). Studies identifying the sugars added to the products of synthetic glycogenes suggest two features of such a glycosylation code in higher plants: long branching sugars are found to be attached to alternating Hyp residues via O-galactosyl linkages, whereas short unbranched sugars are attached to contiguous Hyp residues via O-arabinosyl linkages (Shpak et al., 2001; Zhao et al., 2002; Tan et al., 2003, 2004).

Candidate glycomodules have not yet been identified experimentally in Chlamydomonas, but available data indicate that glycosylation patterns will also prove to be specified by the motifs found in HRGP shafts. For example, the PPSPX-rich Gp1 protein carries a complex mixture of sugar side chains, 37% of which are long and branching (Ferris et al., 2001), whereas Gp2, a second cell wall protein that lacks PPSPX motifs and carries numerous blocks of PPP and PPSP (P.J. Ferris, unpublished data), has a very different carbohydrate profile wherein only 11% of the sugars are long and branching (S. Waffenschmidt, unpublished data). Perhaps reflecting this difference, the shaft width of Gp2 (Figure 4C) is distinctly narrower than that of Gp1 (Figure 4C) or the agglutinins (Figures 4A and 4B). It follows that the unique plus and minus agglutinin shaft sequences (Figure 7) may also specify distinctive patterns of sugar addition.

Once sugars are added to such HRGP shafts, their surface topology is expected to be driven by the PII conformation. A PII helix carries three amino acids per helical gyre, meaning that a given amino acid is separated from its two neighbors by 120°; therefore, when viewed en face, each helix displays three longitudinal faces separated from one another by 120°. The amino acid sequence along each face will in most cases be different from the primary amino acid sequence of the polypeptide: for example, PPSPX repeats will generate PPPXS repeats along each face. Perhaps reflecting this difference, the shaft width of Gp2 (Figure 4C) is distinctly narrower than that of Gp1 (Figure 4C) or the agglutinins (Figures 4A and 4B). It follows that the unique plus and minus agglutinin shaft sequences (Figure 7) may also specify distinctive patterns of sugar addition.

Once sugars are added to such HRGP shafts, their surface topology is expected to be driven by the PII conformation. A PII helix carries three amino acids per helical gyre, meaning that a given amino acid is separated from its two neighbors by 120°; therefore, when viewed en face, each helix displays three longitudinal faces separated from one another by 120°. The amino acid sequence along each face will in most cases be different from the primary amino acid sequence of the polypeptide: for example, PPSPX repeats will generate PPPXS repeats along each face.

Glycosylation defines the interactive molecular surface of an HRGP shaft (Shpak et al., 1999), and glutaraldehyde-fixed gametic flagella continue to display mating-type-specific adhesiveness when mixed together (Goodenough, 1986). Hence, the agglutinin shafts may well associate with one another, and perhaps ratchet along one another, by making transient carbohydrate–carbohydrate interactions along their three faces. These interactions may be individually weak but additively serve to stabilize the whole, a proposal supported by images of adherent flagella that are dominated by meshworks of overlapping shafts (Goodenough and Heuser, 1999). The agglutinin shafts possess an inherent chirality because of the asymmetry of their PPPXS iterations, and they interact with one another in an antiparallel orientation during the mating reaction (Goodenough and Heuser, 1999), features that may also contribute to shaft–shaft interactions.

The shaft of the Gp1 cell wall protein carries discrete subdomains: its dominant PPSPX repeat is interrupted by a tract of poly-P and by a tract of PS repeats (Ferris et al., 2001). The agglutinin shafts are similarly heterogeneous: the terminal 2A and 2E subdomains carry distinctly different motifs from the rest of the shaft (Figure 7) and are predicted to form long faces of poly-P in the P₈ configuration. Hence, the 2A and 2E sequences may dictate unique glycosylation patterns and hence unique modes of protein–sugar or sugar–sugar interactions.

The 2C subdomain is of particular interest in that it is composed of repeating sets of PPSPX motifs, with the plus sets totally different from the minus. In studies to be published elsewhere, we have established that these repeats have been generated by endoduplication events. We have also shown that the plus 2C sequences from C. reinhardtii are totally different from the plus 2C sequences of its sibling species C. incerta and that the minus 2C sequences are totally different in the two species. These findings point to the 2C subdomain as a candidate carrier of species-specific adhesive information.

The N-Terminal Domains

Unmated gametes display their agglutinins with the heads facing outwards and the shaft ends associated with the flagellar membranes (Goodenough et al., 1985; Goodenough and Heuser, 1999). Membrane association is therefore presumably mediated by the globular N-terminal domains that are found at the shaft termini (Figures 4D and 8). These domains carry no predicted transmembrane nor N-myristilation sequences (nor sites for glycosylphosphatidylinositol anchors, which are added to C termini). Therefore, they presumably associate with the flagellar surface by binding to transmembrane agglutinin-anchoring proteins, a conclusion reached earlier by biochemical analysis (Adair et al., 1982).

In contrast with the heads, the plus and minus N-terminal domains are not obviously homologous to one another in intron structure, sequence, or predicted secondary structure (Figures 2, 3, and 8), indicating that if they share a common ancestral domain at all, it has undergone extensive evolutionary divergence, and suggesting that they bind to different (i.e., plus-specific and minus-specific) agglutinin-anchoring proteins.

If agglutinin heads and/or shafts indeed ratchet along opposing shafts during the adhesion reaction, the heads will eventually reach the opposite ends of their partner agglutinins. Therefore, adhesive interactions between heads and N-terminal domains may also figure in the agglutination process.

Near each N-terminal domain is a block interruption sequence in subdomain 2A of the shaft (brackets in Figure 7) that is proposed to generate the kink in the shaft tail hook (Figure 4F); the plus and minus interruption sequences are totally different from one another. Kink domains in Gp1 (Figure 4C) are positioned to participate in cell wall assembly (Goodenough and Heuser, 1988), and the block interruption sequences may play some homologous role in sexual associations.

Organization of Chimeric HRGPs

The algal HRGPs that have been characterized to date all prove to be head-shaft chimeras (Kieliszewski and Lamport, 1994). Several higher plant HRGPs display this same organization,
notably proteins that participate in cell–cell interactions (lectins, AGPs, and proteins found in reproductive tissues). Because the P-rich and non-P-rich tracts in the algal and plant sequences are sharply demarcated, it has been assumed that the two domains form independent modules, and this is demonstrably the case for most regions of the agglutinins and Gp1 (Figure 4).

In addition, however, our analysis of the plus agglutinins indicates that the globular domains also have the capacity to assemble around adjacent shaft domains, analogous to a lollipop surrounding the end of its stick. When the plus head is disrupted before or upon adsorption to mica, a fibrisstructural is visible (Figure 4D, inset), which, in our previous study (Goodenough et al., 1985), we supposed to represent remnants of the denatured head. Reanalysis of such images with the protein sequence in mind suggests that the fibrisstructural is in fact the distal terminus of the shaft, corresponding to the 2E subdomain, around which the head ordinarily assembles. This structure curves back on itself to form a loop (Figure 4E), perhaps as a consequence of its endowment of amino acids that break the Pπ conformation (H. Tran and R. Pappu, personal communication).

The caliber of the head loop is distinctly narrower (3 nm) than the shaft proper (6 nm) (Figure 4E), but it is not as narrow as an unglycosylated Pπ helix (<1 nm), which would be barely visible against the mica surface (Heckman et al., 1988; Staehelin, 1986). We therefore assume that the 2E sequences of the head loops are hydroxylated and glycosylated, albeit differently from the shaft proper, and that the heads then bind to these glycosylated residues, presumably in the Golgi cisternae. It follows that the plus agglutinin can in this sense be considered an autolectin, its head binding to sugars displayed by its own glycopolyopeptide backbone. As posited earlier, the head may also function as an exolectin, binding to sugars displayed on opposite-type shafts.

If the globular domains of chimeric HRGPs prove to have both autolectin and exolectin capabilities, then these activities may alternate in different biological contexts. Thus, the head of an agglutinin may bind to the sugars of its own head loop and then, during the mating reaction, be induced to change conformation so as to bind to the sugars displayed on opposite-type shafts.

A more general concept emerges from these considerations. Whereas many chimeric HRGPs in algae and higher plants resemble the agglutinins and Gp1 in having a single long P-rich sequence interconnecting, and perhaps penetrating, globular N-terminal and/or C-terminal domains (Baldwin et al., 1992; Ertl et al., 1992; Rubinstein et al., 1995; Woessner and Goodenough, 1989; Waffenschmidt et al., 1993; Wu et al., 1993; Woessner et al., 1994; Godl et al., 1997; Schultz et al., 1997; Endler et al., 1999, 2002; Rodriguez et al., 1999; Bosch et al., 2001; Hallmann et al., 2001), other chimeric HRGPs have short P-rich segments that would at best form very short shafts (L. Song and W.L. Dentler, unpublished data, GenBank AF508983; Sumper and Hallmann, 1998; R.A. Bloodgood, unpublished data, GenBank AAO25117; Rodriguez et al., 1999), and others have several short P-rich sequences interspersed with several globular sequences (Cheung et al., 1993; Amon et al., 1998; Suzuki et al., 2000; Kubo et al., 2001; P. Ferris, GP2 sequence, unpublished data). Perhaps in many of these proteins, portions of the P-rich segments function not as protruding shafts but rather as organizational modules around which fold the globular domains. Such arrangements might also undergo biologically relevant conformational changes. In this regard, it is of interest that very short (5 to 12 residues) Pπ helices are present in ligands that interact with SH3 domains, WW domains, profilin, and Class II MHC proteins in animals (reviewed in Stanley and Creamer, 1999).

Sexual adhesion between flagellar membranes, although widespread in the algae (Pickett-Heaps, 1975), is by definition a strategy absent from the flagellaeless higher plants. It is, however, intriguing that most of the chimeric HRGPs that have been identified in higher plants are either AGPs, implicated in numerous kinds of cell–cell interactions (reviewed in Zhao et al., 2002), or lectins (Kieliszewski et al., 1994) or are preferentially expressed in reproductive tissues (Baumberger et al., 2003). Possibly the modes of interaction displayed by present-day algal chimeras originated early in the green lineage and continue to participate in conserved facets of plant biology.

METHODS

Identification of the Sag1 Gene

Strain 21gr mt+ was mutagenized by transformation (Kindle, 1990) with the EcoRI-linearized plasmid pJN4 carrying a dominant gene (Cry1) for emetine resistance (emER) (Nelson et al., 1994). Emetine-resistant colonies were initially screened for their ability to form viable zygotes after mating with an mt− tester strain. Those that failed were examined microscopically to identify mutants that are flagellated, fail to show flagellar agglutination as gametes, and are capable of forming zygotes when treated with cAMP (indicating that they are not defective in differentiation into gametocytes; Pasquale and Goodenough, 1987).

The 99E6 mutant, which met these criteria, was crossed with an mt− strain using cAMP, and random progeny were analyzed. The mutation was shown to be linked to emER, not linked to mt, and expressed in a mt− background only (mt− progeny agglutinated normally), the profile expected for a sex-limited mutation (Goodenough et al., 1978). Then, 99E6 was crossed with mt− strains carrying mutations in Sag1 (sag1−1) and Sag2 (imp8), a gene that influences O-glycosylation of the plus agglutinin (Vallon and Wollman, 1995). Normally agglutinating mt− progeny (14 out of 96 random progeny tested, where 1/8 or 12/96 are predicted by independent assortment) were recovered in the imp8 cross, whereas none (0 out of 96 progeny tested) were recovered in the cross to sag1−1, documenting Sag1 allelism.

A phage library from strain 99E6 (hereafter sag1−6) was screened with pJN4 sequences to identify the insertion-tagged locus. DNA flanking the insertion was used to identify the wild-type Sag1 gene from a second phage library. Analysis of mutant DNA identified the insertional site (Figure 2B).

Identification of the Sad1 Gene

The Sag1 gene was included in the mt− chromosome walk (Ferris and Goodenough, 1994) and identified by RNA gel blot analysis of the mt− locus (Ferris et al., 2002).

Isolation of cDNA Clones

cDNA clones for the Sag1 and Sag1 genes were identified by screening plaque lifts of a cDNA library in Uni-ZAPXR (Stratagene, La Jolla, CA) prepared from 1 h zygotic poly(A)+ RNA (Armbrust et al., 1993) and by hybridization with appropriate radiolabeled genomic probes. A library

Chlamydomonas Sexual Agglutinins 611
prepared from mt⁺ gamete poly(A)⁺ RNA (Kurvari et al., 1998), kindly provided by W.J. Snell, provided additional Sag1 cDNAs. Inserts from positive clones were excised as pBluescript II SK– plasmids using Stratagene’s rapid excision kit.

**DNA Sequencing and Analysis**

DNA sequencing entailed subcloning, construction of gene-specific primers, and use of the GPS-1 genome priming system (New England Biolabs, Beverly, MA). Sequence reactions were performed with the ABI PRISM Dye terminator cycle sequencing ready reaction kit (Applied Biosystems, Foster City, CA) using double-stranded plasmid DNA and subsequent analysis on an ABI DNA sequencer. Addition of 5% DMSO or 1 M betaine to the sequencing reactions was often necessary. dGTP was omitted from the sequencing reactions to prevent precipitation of magnesium during cycle sequencing. For sequencing reactions performed on an ABI DNA sequencer, 25 mM dGTP was added to the sequencing reactions. dGTP was omitted from sequencing reactions performed on a Bio-Rad sequencer. An ABI DNA sequencer was used for cycle sequencing reactions. The 5’ RACE (Rapid Amplification of cDNA Ends) of poly(A)⁺ RNA and reverse transcription were performed as for RT-PCR. After the RNase H treatment, cDNA was purified with a Prep-A-Gene DNA purification kit (Bio-Rad, Hercules, CA). A poly(C) linker was added to the 3’ end of cDNA by the terminal transferase reaction. The cDNA was resuspended in 50 μL of the buffer containing 50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate, 1 mM DTT, 0.25 mM CaCl₂, 2 mM dCTP, and 30 units of terminal transferase and incubated at 37°C for 30 min. The reaction was terminated by incubating at 70°C for 10 min. One microliter of resultant reaction solution was used for a 50-μL PCR reaction. The first round of PCR was performed with the primer 5’-GGCCACGCCTCGACTGACGGGIGGGIIGGGIIIG-3’, which binds the polyC tail and a gene-specific primer predicted to be near the 5’ end of the Sag1 or Sag1 mRNA. A second nested PCR was then performed with the primer 5’-GGCCACGCCTCGACTGACGGGIGGGIIGGGIIIG-3’ and then another gene-specific primer 5’ to the first primer. The resulting PCR fragments were cloned in pGEM-T vector and sequenced.

**Sequence and Secondary Structure Analysis and Image Processing**

Signal sequences were predicted in accordance with Nielsen et al. (1997). Sequences were aligned using the ClustalW program (Thompson et al., 1994) located at http://npsa-pbil.ibcp.fr/ with scoring of strongly similar, weakly similar, and identical amino acids according to the Gonnet matrix (Gonnet et al., 1992).

Secondary structure was predicted using two algorithms: the HNN secondary prediction method at http://npsa-pbil.ibcp.fr (Guermeur et al., 1999) and the PSIPRED method (Jones, 1999) at http://bioinf.cs.ucl.ac.uk/psipred/ (McCuffin et al., 2000). Secondary structures predicted by both methods (with a probability of five and higher by the PSIPRED method) were considered to be significant and included in the drawings and calculations.

Using the Adobe Photoshop 5.0 image processing tool kit (Mountain View, CA), the lengths of digitized images of agglutinin molecules were traced with a white line, and the length of the line was measured by the program.

**CD Spectra**

CD spectra were recorded on a Jasco Model J-175 (Jasco, Gross-Umstadt, Germany), calibrated using ammonium d10-camphorsulfonic acid. All measurements were performed in a heat-controlled 0.1-cm path length cylindrical cuvette at 180 to 260 nm at 25°C. Typically, 10 spectra were recorded at a scan speed of 50 nm/min with a step resolution of 0.1 nm and a nitrogen stream of 8 mL/min. All spectra were corrected for a protein-free spectrum obtained under identical conditions; noise reduction was applied according to the Jasco software.

Sequence data from this article have been deposited with EMBL/GenBank data libraries under accession numbers AY450930 and AY450929.

**ACKNOWLEDGMENTS**

We thank Linda Small for expert DNA cloning/sequencing, John Heuser for the QFDE-TEM images, and Rohit Pappu for stimulating discussions of P₃ helices. This work was supported by grants GM-26150 from the National Institutes of Health and MCB-9904887 from the National Science Foundation to U.W.G. and P.J.F. and by Wa 659/8-1 from Deutsche Forschungsgemeinschaft to S.W. Received September 29, 2004; accepted November 24, 2004.
REFERENCES


Analysis of proline codon usage in HRGPs of *Chlamydomonas reinhardtii*.

<table>
<thead>
<tr>
<th>Target</th>
<th>number of codons-a</th>
<th>%GC-b</th>
<th>%GC if normal bias-c</th>
<th>%GC 3rd position-d</th>
<th>number of proline codons</th>
<th>% prolines</th>
<th>%CCG/C</th>
<th>%CCG</th>
<th>%CCC</th>
<th>%CCA</th>
<th>%CCT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chlmy average</td>
<td>127,008</td>
<td>65</td>
<td>65</td>
<td>86</td>
<td>7392</td>
<td>6</td>
<td>80</td>
<td>31</td>
<td>49</td>
<td>8</td>
<td>12</td>
</tr>
<tr>
<td>SAD1 shaft</td>
<td>873</td>
<td>74</td>
<td>83</td>
<td>56</td>
<td>526</td>
<td>60</td>
<td>51</td>
<td>26</td>
<td>25</td>
<td>20</td>
<td>29</td>
</tr>
<tr>
<td>SAD1 N-term</td>
<td>612</td>
<td>62</td>
<td>64</td>
<td>82</td>
<td>42</td>
<td>7</td>
<td>74</td>
<td>36</td>
<td>38</td>
<td>19</td>
<td>7</td>
</tr>
<tr>
<td>SAD1 head</td>
<td>2404</td>
<td>72</td>
<td>74</td>
<td>78</td>
<td>221</td>
<td>9</td>
<td>70</td>
<td>37</td>
<td>34</td>
<td>13</td>
<td>17</td>
</tr>
<tr>
<td>SAG1 shaft</td>
<td>934</td>
<td>80</td>
<td>84</td>
<td>70</td>
<td>560</td>
<td>60</td>
<td>66</td>
<td>33</td>
<td>33</td>
<td>10</td>
<td>24</td>
</tr>
<tr>
<td>SAG1 N-term</td>
<td>469</td>
<td>63</td>
<td>65</td>
<td>79</td>
<td>37</td>
<td>8</td>
<td>67</td>
<td>32</td>
<td>35</td>
<td>27</td>
<td>5</td>
</tr>
<tr>
<td>SAG1 head</td>
<td>2006</td>
<td>73</td>
<td>74</td>
<td>80</td>
<td>167</td>
<td>8</td>
<td>81</td>
<td>54</td>
<td>27</td>
<td>11</td>
<td>8</td>
</tr>
<tr>
<td>GP1 shaft</td>
<td>342</td>
<td>79</td>
<td>84</td>
<td>66</td>
<td>208</td>
<td>61</td>
<td>64</td>
<td>16</td>
<td>48</td>
<td>3</td>
<td>34</td>
</tr>
<tr>
<td>GP1 rest</td>
<td>213</td>
<td>65</td>
<td>65</td>
<td>88</td>
<td>14</td>
<td>7</td>
<td>79</td>
<td>0</td>
<td>79</td>
<td>0</td>
<td>21</td>
</tr>
<tr>
<td>MTA2 shaft</td>
<td>166</td>
<td>78</td>
<td>83</td>
<td>66</td>
<td>100</td>
<td>60</td>
<td>62</td>
<td>27</td>
<td>35</td>
<td>17</td>
<td>21</td>
</tr>
<tr>
<td>MTA2 rest</td>
<td>220</td>
<td>61</td>
<td>65</td>
<td>75</td>
<td>12</td>
<td>6</td>
<td>59</td>
<td>17</td>
<td>42</td>
<td>7</td>
<td>25</td>
</tr>
<tr>
<td>GAS28 shaft</td>
<td>61</td>
<td>79</td>
<td>85</td>
<td>64</td>
<td>39</td>
<td>64</td>
<td>61</td>
<td>10</td>
<td>51</td>
<td>0</td>
<td>38</td>
</tr>
<tr>
<td>GAS28 rest</td>
<td>385</td>
<td>63</td>
<td>64</td>
<td>83</td>
<td>22</td>
<td>6</td>
<td>78</td>
<td>14</td>
<td>64</td>
<td>9</td>
<td>14</td>
</tr>
<tr>
<td>VSP1 shaft</td>
<td>190</td>
<td>70</td>
<td>73</td>
<td>77</td>
<td>61</td>
<td>32</td>
<td>70</td>
<td>26</td>
<td>44</td>
<td>8</td>
<td>21</td>
</tr>
<tr>
<td>VSP1 rest</td>
<td>110</td>
<td>66</td>
<td>70</td>
<td>74</td>
<td>4</td>
<td>4</td>
<td>50e</td>
<td>25</td>
<td>25</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>Gene</td>
<td>number of codons-a</td>
<td>%GC-b</td>
<td>%GC if normal bias-c</td>
<td>%GC 3rd position-d</td>
<td>number of proline codons</td>
<td>% prolines</td>
<td>%CCG/C</td>
<td>%CCG</td>
<td>%CCC</td>
<td>%CCA</td>
<td>%CCT</td>
</tr>
<tr>
<td>----------</td>
<td>--------------------</td>
<td>-------</td>
<td>---------------------</td>
<td>--------------------</td>
<td>--------------------------</td>
<td>-------------</td>
<td>--------</td>
<td>------</td>
<td>------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>VSP3 shaft</td>
<td>203</td>
<td>69</td>
<td>76</td>
<td>64</td>
<td>90</td>
<td>44</td>
<td>65</td>
<td>17</td>
<td>48</td>
<td>1</td>
<td>34</td>
</tr>
<tr>
<td>VSP3 rest</td>
<td>270</td>
<td>63</td>
<td>64</td>
<td>83</td>
<td>9</td>
<td>3</td>
<td>78</td>
<td>11</td>
<td>67</td>
<td>0</td>
<td>22</td>
</tr>
<tr>
<td>VSP4 shaft</td>
<td>66</td>
<td>80</td>
<td>78</td>
<td>89</td>
<td>34</td>
<td>52</td>
<td>88</td>
<td>32</td>
<td>56</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>VSP4 rest</td>
<td>925</td>
<td>65</td>
<td>64</td>
<td>89</td>
<td>58</td>
<td>6</td>
<td>94</td>
<td>34</td>
<td>60</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>ZSP1 shaft</td>
<td>123</td>
<td>73</td>
<td>74</td>
<td>81</td>
<td>42</td>
<td>34</td>
<td>81</td>
<td>29</td>
<td>52</td>
<td>5</td>
<td>14</td>
</tr>
<tr>
<td>ZSP1 rest</td>
<td>79</td>
<td>69</td>
<td>72</td>
<td>79</td>
<td>1</td>
<td>1</td>
<td>100e</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>GP2 shaft</td>
<td>241</td>
<td>79</td>
<td>85</td>
<td>65</td>
<td>167</td>
<td>69</td>
<td>60</td>
<td>19</td>
<td>41</td>
<td>0</td>
<td>40</td>
</tr>
<tr>
<td>GP2 rest</td>
<td>1017</td>
<td>65</td>
<td>65</td>
<td>86</td>
<td>85</td>
<td>8</td>
<td>79</td>
<td>14</td>
<td>65</td>
<td>1</td>
<td>20</td>
</tr>
</tbody>
</table>

Abbreviations:
a) number of codons in the indicated gene region  
b) GC content of the mRNA segment(s) coding for the indicated gene region  
c) the calculated GC content for the mRNA if the codons for all amino acids had been used in the same proportions as in an average *Chlamydomonas* gene  
d) % of codons that have a G or a C in the third position  
e) too few prolines for the result to be meaningful

Source of data:  
Chlamy average: A compilation of *Chlamydomonas reinhardtii* genes obtained from a codon usage database maintained by the Kazusa DNA Research Institute, http://www.kazusa.or.jp  
VSP1 L16461  
VSP3 L29029  
VSP4 AY036106  
ZSP1 S44199  
GP2 AY596305  
GAS28 AF015883  
GP1 AF309494  
MTA2 AF309495  
SAG1 AY450930  
SAD1 AY450929
Further notes:
The codon usage has been determined separately for the proline-rich portion (termed “shaft”) of each protein (which combines several segments in some proteins) and for the remainder of the protein (termed “rest”). For the two agglutinin proteins (SAG1 and SAD1) the head domain and the N-terminal domain were calculated separately.

Codon usage in *Chlamydomonas reinhardtii* is biased, and favors codons ending in G or C over those ending in A or T (86% of the time), consistent with the overall GC richness of coding regions (on average 65.4% GC). In regions rich in proline codons (CCN), as are found in HRGPs, this could lead to extremely high GC contents. For example, a region of 100 prolines would, on average, contain only 8 CCA and 12 CCT codons, resulting in a GC content of 280/300 or 93%. The data document that codon bias in the proline rich regions of cell wall and agglutinin genes is attenuated, with far more CCA and CCT codons used than expected. For example, the minus agglutinin shaft, which is 60% proline codons, has a GC content of 74%. This is, however, less than the GC content would be if the codon usage were typical of an average *C. reinhardtii* gene, in which case the GC content would have been 83% (these calculations are done for the codons of all the amino acids in the given region, not just the prolines). Thus the frequency of GC in the third position of all codons is only 56% rather than the usual 86%, and only 51% of proline codons end in G/C, rather than the usual 80%. This effect is not necessarily restricted to proline codons: for example, alanine codons (GCN), which are also GC rich, end in G or C 77% of the time on average, but 24% of the time in the SAD1 shaft (not shown).

The reduced 3rd position GC content is seen in all the proteins with long proline-rich shafts (i.e. SAD1, SAG1, GP1, MTA2, and GP2; all > 60% prolines). The shaft region of GAS28 is shorter but quite proline-rich, and also shows the effect.

The effect is less consistent in shafts of lower proline content. The 3rd position GC content for proline in VSP4, which has a short shaft of alternating serine and proline, is actually above average. In ZSP1 the effect is not seen perhaps because the shaft is fairly short. There is some effect in VSP1 and VSP3, whose shafts, while only 30-40% proline, are of greater length than those of ZSP1 and VSP4.
Plus and Minus Sexual Agglutinins from Chlamydomonas reinhardtii
Patrick J. Ferris, Sabine Waffenschmidt, James G. Umen, Huawen Lin, Jae-Hyeok Lee, Koichi Ishida, Takeaki Kubo, Jeffrey Lau and Ursula W. Goodenough
Plant Cell 2005;17;597-615; originally published online January 19, 2005;
DOI 10.1105/tpc.104.028035

This information is current as of October 20, 2017

Supplemental Data  /content/suppl/2005/01/31/tpc.104.028035.DC1.html
References  This article cites 100 articles, 46 of which can be accessed free at:
/content/17/2/597.full.html#ref-list-1
eTOCs  Sign up for eTOCs at:
http://www.plantcell.org/cgi/alerts/ctmain
CiteTrack Alerts  Sign up for CiteTrack Alerts at:
http://www.plantcell.org/cgi/alerts/ctmain
Subscription Information  Subscription Information for The Plant Cell and Plant Physiology is available at:
http://www.aspb.org/publications/subscriptions.cfm

© American Society of Plant Biologists
ADVANCING THE SCIENCE OF PLANT BIOLOGY