Aplysia Attractin: Biophysical Characterization and Modeling of a Water-Borne Pheromone

Catherine H. Schein,* Gregg T. Nagle,† Jason S. Page,‡ Jonathan V. Sweedler,‡ Yuan Xu,* Sherry D. Painter,† and Werner Braun*

*Sealy Center for Structural Biology, Department of Human Biological Chemistry and Genetics, and †Marine Biomedical Institute and Department of Anatomy and Neurosciences, University of Texas Medical Branch, Galveston, Texas 77555; and ‡Department of Chemistry and the Beckman Institute, University of Illinois, Urbana, Illinois 61801 USA

ABSTRACT Attractin, a 58-residue protein secreted by the mollusk *Aplysia californica*, stimulates sexually mature animals to approach egg cordons. Attractin from five different *Aplysia* species are ~40% identical in sequence. Recombinant attractin, expressed in insect cells and purified by reverse-phase high-performance liquid chromatography (RP-HPLC), is active in a bioassay using *A. brasiliana*; its circular dichroism (CD) spectrum indicates a predominantly α-helical structure. Matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) characterization of proteolytic fragments identified disulfide bonds between the six conserved cysteines (I–VI, II–V, III–IV, where the Roman numeral indicates the order of occurrence in the primary sequence). Attractin has no significant similarity to any other sequence in the database. The protozoan *Euplotes* pheromones were selected by fold recognition as possible templates. These diverse proteins have three α-helices, with six cysteine residues disulfide-bonded in a different pattern from attractin. Model structures with good stereochemical parameters were prepared using the EXDIS/DIAMOD/FANTOM program suite and constraints based on sequence alignments with the *Euplotes* templates and the attractin disulfide bonds. A potential receptor-binding site is suggested based on these data. Future structural characterization of attractin will be needed to confirm these models.

INTRODUCTION

Peptide pheromones coordinate reproductive activities in both vertebrate and invertebrate species. Little is known, however, about their structure or mode of action in higher organisms. We recently isolated a water-borne peptide pheromone (attractin) from the marine mollusk *Aplysia californica* (Painter et al., 1998). The peptide is synthesized by the albumen gland (Painter et al., 1998), a large exocrine organ in the female reproductive tract that packages the eggs into a cordon for deposition (Coggeshall, 1972). Once the cordon is laid, attractin elutes into the surrounding seawater and is slowly degraded.

The *A. californica* attractin cDNA encodes a signal peptide followed by a single copy of the 58-residue protein, with no transmembrane domain (Fan et al., 1997). Only NH_2-terminal fragments of up to 47 residues have been recovered from the surrounding seawater after deposition (Painter et al., 1998). *A. californica* albumen gland attractin is glycosylated at Asn-8, and fragments are de-glycosylated within 30 min after secretion (Painter et al., 1998). Northern blot analysis has demonstrated high levels of attractin mRNA in albumen glands, and ~1% of all clones in an albumen gland cDNA library encode attractin (Fan et al., 1997). As we show here, both attractin isolated from *Aplysia* and the recombinant peptide produced in a baculovirus system significantly attract conspecifics in T-maze bioassays at very low concentrations (1 pmol in 6 l seawater). Four attractin homologs have been isolated from other species of *Aplysia*, and partially or completely characterized (S. D. Painter and G. T. Nagle, manuscript in preparation). Two of the sequences are nearly identical to the *A. californica* sequence, while the other two are very similar to each other but ~40% identical to the others. All are attractive in one or more species, suggesting that they share a common 3D structure and receptor-binding site.

Attractin is a promising model system in which to determine the mechanism of protein pheromone action in marine organisms. A reliable 3D structure would provide insights about possible receptor binding sites on the surface, which could be further localized by site-directed mutagenesis. The 3D structure is likely to be novel, as the *Aplysia* sequences have no significant similarity to those of other known peptides or proteins. Several peptide pheromones from the protozoan *Euplotes raikovi* (E*r*) (Miceli et al., 1991) have been structurally characterized (Brown et al., 1993; Luginbuhl et al., 1996; Mronga et al., 1994; Weiss et al., 1995). This organism produces several proteins (E*r*-1, E*r*-2, E*r*-10, E*r*-11) that differ greatly in their mature primary sequence, but retain the same compact, "pyramid" 3D structure of three α-helices stabilized by three disulfide bonds (Miceli et al., 1991; Weiss et al., 1995).

We present here an initial biophysical characterization of attractin and possible 3D models. Recombinant attractin was purified from an insect cell expression system, assayed for bioactivity, and characterized by circular dichroism (CD). The disulfide-bonding pattern of the recombinant material was determined using limited proteolysis and ma-

Received for publication 1 December 2000 and in final form 3 April 2001.

Address reprint requests to Dr. Werner Braun, Sealy Center for Structural Biology, Department of Human Biological Chemistry and Genetics, University of Texas Medical Branch, Galveston, TX 77555-1157. Tel.: 409-747-6810; Fax: 409-747-6850; E-mail: werner@newton.utmb.edu.

0006-3495/01/07/463/10 $2.00

trix-assisted laser desorption/ionization mass spectrometry (MALDI-MS). BLAST searches with attractin sequences did not reveal a template structure with sufficient identity. Protein fold recognition and secondary structure prediction methods available on the WEB, including our MASIA program (Zhu et al., 2000), identified fragments and structures within the E*r* family as possible fold neighbors. CLUSTALW multiple sequence alignments revealed that attractins have some sequence similarity to the E*r* family, but that, consistent with the difference in the disulfide bonding pattern between the two families, only five of the six conserved cysteines align. We combined these sequence- and function-based considerations with the experimentally determined disulfide bonding pattern of *A. californica* attractin, and prepared preliminary models using our EXDIS/DIAMOD program suite (Soman et al., 2000). The models were further selected based on low energies following energy minimization with our FANTOM (Fraczkiewicz and Braun, 1998) program and the results of the PROCHECK program (Laskowski et al., 1993). CD data confirmed the structure was helical, as in our models. Visual analysis of the best models according to these criteria was consistent with a highly conserved region of the attractin structure forming a receptor-binding site at the surface of the molecule.

MATERIALS AND METHODS

Animals

Specimens of *Aplysia brasiliana,* ranging in weight from 100 to 500 g, were collected from South Padre Island, TX, during the summer reproductive season. They were maintained in individual cages in large aquaria containing recirculating artificial seawater (Instant Ocean Marine Salt, Longhorn Pet Supply, Houston, TX) at 20 ± 2°C, the summer water temperature for this species. A 14:10 light/dark cycle was maintained in the facility, with the light starting at 0600. The animals were fed dried laver at 1600–1800, after bioassays were finished.

Pheromonal attraction assay

The T-maze, removable stimulus cages, and experimental protocol were described previously (Painter et al., 1991). Individual animals for each assay were selected based on four criteria. The test animal must 1) be able to lay eggs but not have done so in the past 24 h; 2) not have been used in a bioassay during the preceding 24 h; 3) not have been exposed to attractin before; and 4) have been housed in the same aquarium as the stimulus animal (Painter et al., 1998).

Recombinant attractin expression

The *A. californica* albumen gland attractin cDNA (Fan et al., 1997) was subcloned into the baculovirus expression vector pFastBac 1, and recombinant virus was generated using the Bac-to-Bac Baculovirus Expression System (Gibco BRL, Rockville, MD). Attractin was expressed in Sf9 insect cells grown in Sf-900 II serum-free medium (Gibco BRL).

Purification of native and recombinant attractin

A. californica albumen gland attractin was purified by analytical C18 reverse-phase high-performance liquid chromatography (RP-HPLC) as previously described (Painter et al., 1998). To prepare recombinant attractin, expressing Sf9 cells were centrifuged and the pellet was resuspended in 20 ml ice-cold 0.1% heptafluorobutyric acid (HFBA) and sonicated. The resulting lysates were purified on C18 Sep-Pak Vac cartridges (5 g; Waters, Milford, MA) that were pretreated with 10 ml 100% acetonitrile (CH_3CN) containing 0.1% HFBA and rinsed with 20 ml 0.1% HFBA. The peptides were loaded, then eluted with 15 ml of 50% CH_3CN containing 0.1% HFBA and lyophilized. The lyophilysate was resuspended in 2.5 ml of 0.1% HFBA and applied to a Vydac analytical C18 RP-HPLC column (4.6 × 250 mm; Hesperia, CA). The column was eluted with a two-step linear gradient of 0.1% HFBA in water and 100% CH_3CN containing 0.1% HFBA. The first step was 0–10% CH_3CN in 5 min, followed by a shallower gradient from 10% to 34% CH_3CN in 85 min. The column eluate was monitored at 215 nm, and 1-min (1 ml) fractions were collected. The attractin-containing fractions were combined and lyophilized.

Before determination of the disulfide-bonding pattern, attractin was reduced with 2-mercaptoethanol and alkylated with 4-vinylpyridine (4-VP; Pierce Chemical, Rockford, IL) (Coligan, 1997; Hawke and Yuan, 1987), and purified by Vydac C18 RP-HPLC. The same gradient conditions were used as described above, except that 0.1% trifluoroacetic acid (TFA; Pierce) was the counterion. The peak of interest was characterized by amino acid compositional analysis, microsequence analysis, and MALDI-MS.

For CD and disulfide bonding pattern studies, fractions containing recombinant attractin were not reduced and alkylated before Vydac C18 RP-HPLC.

Circular dichroism

Lyophilized attractin was taken up in water and neutralized by adding NaOH. The remaining pellet was allowed to slowly redissolve into 10 mM sodium phosphate buffer, pH 6.5 (NaP), centrifuged to remove precipitated protein, then concentrated by centrifugation (7000 × *g*) in a Centricon 3 concentrator (Amicon) to a volume of 0.5 ml. Fresh NaP (1.5 ml) was added and the sample concentrated again to 0.5 ml. After one more addition of buffer and concentration, the protein was diluted 15-fold into NaP to a final concentration of 0.3 mg/ml. The CD spectrum was measured on an Aviv spectrophotometer model 62DS. Attractin ran as a single, ~6 kD band on an 18% SDS-polyacrylamide gel stained with Coomassie Blue (data not shown).

Amino acid compositional and microsequence analysis

Compositional analyses were performed using a PE Biosystems 420H Amino Acid Analyzer (Smith et al., 1991). For Edman degradation of recombinant attractin, samples were applied to Biobrene Plus-treated glass fiber filters and subjected to automated NH_2-terminal sequence analysis using a PE Biosystems Procise 494/HT protein sequencer. Pulsed liquid cycles were used for each analysis.

Determination of the disulfide bonding pattern

Lyophilized recombinant attractin (~3 nmol) was dissolved in 150 μl Milli-Q-purified water (Millipore, Boston, MA) to ~20 μM concentration. For the first step in enzymatic digestion, a small crystal of lyophilized *Staphylococcus aureus* V8 protease (V8; Sigma-Aldrich, St. Louis, MO) was dissolved in 50 μl of 10 mM ammonium phosphate, pH 7.6, made from ammonium phosphate monobasic (Fisher Scientific, Pittsburgh, PA) and ammonium phosphate dibasic (Aldrich, Milwaukee, WI). The attractin solution (10 μl) was combined with 10 μl of the V8 protease solution in a microcentrifuge tube. A second mixture of 10 μl of the enzyme solution and 10 μl purified water served as a blank. Both solutions were put on a

heating block set to 40°C and samples removed at time intervals for MALDI-MS. For the second step in the enzymatic digest, a small crystal of lyophilized trypsin (Sigma-Aldrich) was dissolved in 50 μl ammonium phosphate buffer, and 0.5 μl of this solution was added to the remaining attractin and control solution for the V8 digest. The solutions were incubated further and sampled as above.

MALDI-MS measurements

MALDI matrix solutions of 15 mg/ml 2,5-dihydroxybenzoic acid (DHB; ICN Biomedicals, Aurora, OH) in purified water and 15 mg/ml α-cyano-4-hydroxycinnamic acid (CHCA; Aldrich) in 6:3:1 CH_3CN/water/3% TFA were prepared. Samples (0.5 μl) were taken at various reaction times and spotted separately with 0.5 μl of the two different MALDI matrix solutions onto a standard gold-plated target (PE Biosystems, Framingham, MA). The spotted samples were mass-profiled using a Voyager DE STR mass spectrometer with delayed ion extraction (PE Biosystems). Mass spectra were obtained using a pulsed nitrogen laser (337 nm) as the ionization/desorption source. The instrument was used in linear mode with an acceleration voltage of 20 kV and a delay of 100 ns. The resulting mass spectra are unsmoothed and are an average of 80–128 acquisitions. Mass calibration was performed externally on the target by using low (α-bag cell peptide) and high (egg-laying hormone) mass peptides from *A. californica* (American Peptide Co., Sunnyvale, CA).

Determination of free sulfhydryl groups

Equal amounts (5 μl) of a 20 μM attractin solution and 5 mM iodoacetate (IAA; Sigma-Aldrich) in 10 mM Tris/HCl buffer (Sigma-Aldrich), pH 8.2 were combined in a microcentrifuge tube. After an initial incubation, 1 μl 30 mM dithiothreitol (DTT; Sigma-Aldrich) in water was added to the attractin/IAA solution and a protein-free control. Both solutions were kept at 50°C in a hot water bath. Samples of 0.5 μl were taken periodically from both solutions throughout the reactions and spotted onto a MALDI target with 0.5 μl CHCA matrix solution.

Protein modeling

The modeling was done on SGI/R10000 workstations. The sequence of attractin was aligned to that of the template with the program CLUSTALW; www2.ebi.ac.uk/clustalw/ (Higgins et al., 1992; Thompson et al., 1994). The program EXDIS, developed in our group, was used to extract inter-atomic distance and dihedral angle constraints from the structure of the templates; http://www.scsb.utmb.edu/FANTOM/fm_home.html (Abe et al., 1984; Schaumann et al., 1990). During this process, short stretches corresponding to "gaps" or "loops" in the alignment are left out, and constraints are extracted from the remaining "fragments" of the protein. For a given atom, EXDIS selects a specified number of other atoms, chosen randomly, and calculates distances to them. For the two attractin models depicted ("I" and "III"), specifying 10 constraints per atom, a total of 2215 distances were extracted from the E*r*-11 structure (PDB structure 1ERY), and 2086 from E*r*-2 (1ERD). Each distance was used as an upper and a lower bound for that atom pair by adding a "tolerance" value of $\pm$0.1 Å. For dihedral angle constraints, EXDIS uses the following rule at each aligned position: if the amino acids in attractin and the template are identical, all dihedral angles are read from the template; if they differ, only the backbone dihedral angles are used. Values of unknown dihedral angles are assigned a starting value of 180°. These are converted to ranges by adding $\pm$10° to ω and $\pm$15° to the other torsion angles. For models I and III, 131 and 121 dihedral constraints were obtained from 1ERY and 1ERD, respectively. Models were then built with our self-correcting distance geometry program DIAMOD; http://www.scsb.utmb.edu/comp_biol.html/larisa/noah_diamod.html (Hänggi and Braun, 1994; Mumenthaler and Braun, 1995; Zhu et al., 2000).

The program FANTOM (Schaumann et al., 1990) was then used with the above constraints to minimize the conformational energy of the protein. Both template structures have three disulfide bridges, which differ in their assignments from those in attractin. These constraints were replaced with those for the disulfide bonds in attractin at residues 4–41, 13–33, and 20–26 in sequence.

In the next stage, the full energy function was applied and minimized. A fourth power energy function was used for distance constraints, which added kT/2 to the total energy for a violation by 0.2 Å in the regularization stage. This limit was raised in two steps to 1.0 Å by the end of the minimization. The distance constraints to the template were thus progressively relaxed. The dihedral angle constraint function added an energy of 10.0*kT/2 for every 5° violation. The minimization was done by the successive application of quasi-Newton and Newton-Raphson minimizers as implemented in FANTOM (Schaumann et al., 1990). Models were inspected and figures prepared using MOLMOL; www.mol.biol.ethz.ch/wuthrich/software/molmol (Koradi et al., 1996).

RESULTS

Expression and purification of bioactive recombinant attractin

The RP-HPLC peak fractions containing full-length recombinant attractin were characterized by amino acid compositional and Edman microsequence analyses; the 58-residue peptide sequence was identical to albumen gland attractin (data not shown). The only difference between the two was that, according to MALDI-MS, recombinant attractin lacks the glycosylation at Asn-8 found in the albumen gland peptide (Painter et al., 1998). As both proteins are equally active in T-maze bioassays (Fig. 1), glycosylation is not necessary for biological activity.

Determination of the number of disulfide bonds

Recombinant attractin, before and after reduction of disulfide bonds with DTT, was reacted with IAA to alkylate free sulfhydryl groups. When attractin was exposed to IAA there was no detectable mass shift from alkylation. However, when DTT was added to the attractin/IAA solution, all six cysteines were alkylated. Thus, all six of the cysteines in native attractin appear to be involved in disulfide bonds, with no accessible sulfhydryl groups.

Determination of the disulfide-bonding pattern of recombinant attractin

Attractin has several sites for enzymatic cleavage between each of the six cysteines. We chose to start with the V8 protease, which cleaves specifically after glutamic and aspartic acid, and then trypsin, which cleaves after lysine and arginine (Fig. 2 *A*). Small amounts of attractin (5–10 μl of a 20 μM solution) were digested with V8 protease and aliquots were periodically spotted onto a MALDI-MS target. Fig. 3 shows spectra from this reaction series (we show

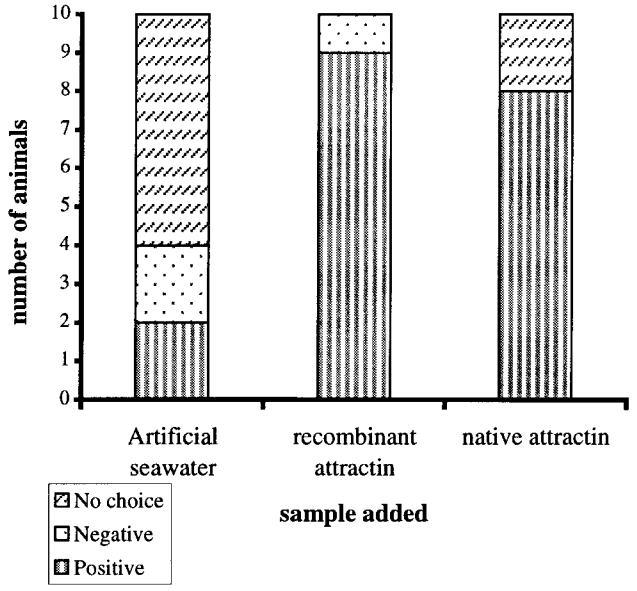


FIGURE 1 Native and recombinant attractin are biologically active. The number of *A. brasiliana* attracted to a non-laying conspecific was increased by placing 1 pmol of either native or recombinant attractin in the adjacent seawater. The response patterns for each differed significantly from that to a non-layer alone (recombinant: $\chi^2(2) = 13.75$; $p < 0.005$; native: $\chi^2(2) = 10.44$; $0.005 < p < 0.01$), but did not differ significantly from each other ($\chi^2(2) = 2.1$; $0.25 < p < 0.5$). This bar graph is based on 30 single-arm experiments, 10 per stimulus; animals chose between a stimulus in one arm and no stimulus in the other.

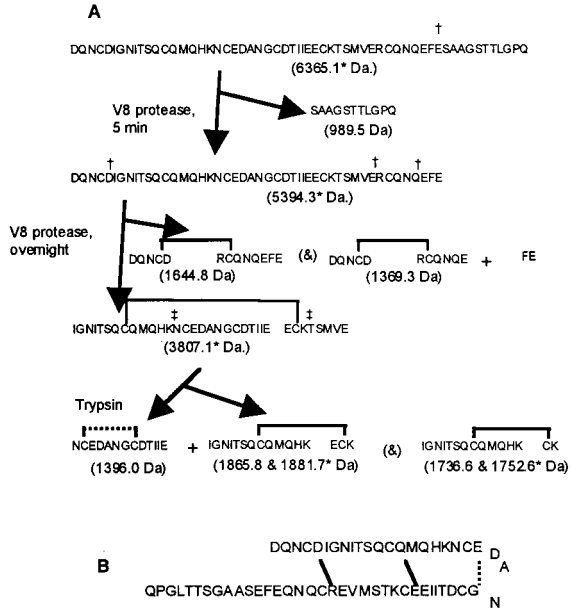


FIGURE 2 The molecular masses (from MALDI-MS analysis) of fragments of attractin formed after 5 min and 12 h of digestion with V8 protease and an additional 12 h of digestion with trypsin (*A*) indicate the disulfide bonding pattern shown in (*B*). The bonds between the first and sixth and second and fifth cysteine residues in the sequence are confirmed by detection of the indicated masses. The dotted line between the third and fourth cysteine residues indicates an inferred disulfide. Cleavage sites are indicated (†, V8 protease; ‡, trypsin; ∗, mass corresponding to the oxidation of methionine).

here only a few of more than 100 mass spectra obtained and analyzed). Within 5 min, the peptide SAAGSTTLGPQ was cleaved from the COOH terminus (Fig. 3 *A*). After 12 h (Fig. 3, *B* and *C*), peaks consistent with a disulfide-bond between the first and last cysteine (C4–C41) were found. The V8 digested sample was then further digested with trypsin. After 12 h (Fig. 3 *D*), six peaks corresponding to attractin peptides were found, two of which contain oxidized methionines. These peaks were consistent with a second disulfide bond occurring between the second and fifth cysteine residues in the sequence (C13–C33). The peptide containing the third and fourth cysteines (C20–C26; *m/z* 1396.0) was not cleaved further, so we could not determine whether these two cysteines were disulfide-bonded using the same approach. However, these appear, by the process of elimination, to be bonded to one another, as there are no free disulfides in attractin.

Alignment of attractin sequences

Fig. 4 *A* shows an alignment of the complete attractin sequences that have been deposited in the International Protein Sequence Database (*A. californica* attractin, A59061; *A. brasiliana* attractin, B59060) and the partial sequences of three other *Aplysia* species. The purification, structure, and biological activity of attractin-related peptides from *A. brasiliana*, *A. fasciata*, *A. depilans*, and *A. vaccaria* will be described elsewhere (S. D. Painter and G. T. Nagle, manuscript in preparation). This alignment indicates that the major areas of conservation are the six cysteines and residues 29–36. Extensive search of all sequence databases revealed no closely related peptide sequences.

Protein fold recognition and secondary structure prediction

The attractin sequences were submitted to various protein fold recognition programs, including our MASIA program (Zhu et al., 2000) and the web-based programs PSA (White et al., 1994), JPRED (Cuff and Barton, 1999), PREDATOR (Frishman and Argos, 1997), PHD (Rost, 1996), GenTHREADER (Jones, 1999a), PSSM, PSIBLAST (Altschul et al., 1997), PSIPRED (Jones, 1999b) and hidden Markov models (Karplus et al., 1998). PSA, PREDATOR, and PHD all suggested helical regions at residues SQCQ and IEECKTS, with the limits of the helices differing slightly. The other methods suggested some β-sheet content but there was no consistency in the predictions.

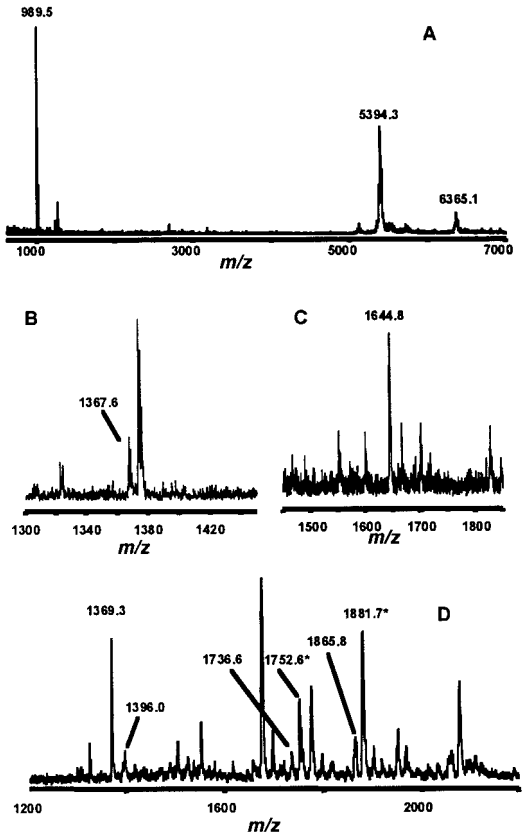


FIGURE 3 MALDI-MS spectra of samples from the enzymatic digest of attractin. (*A*) Spectrum of a sample from 5 min after enzyme addition, showing three peaks corresponding to a peptide removed from the COOH terminus of attractin, the shortened attractin, and full-length attractin. The intact and shortened attractin have masses labeled 16 Da greater than calculated as a result of different oxidation states of the two methionine residues in attractin. After further digestion by V8 protease (*B* and *C*), the two labeled peaks correspond to unique singly disulfide-bonded peptides and confirm one disulfide-bond arrangement in attractin. The other peaks seen in the spectra are from self-digestion of the enzyme, verified by MALDI-MS analysis of a blank solution containing only the enzyme. To further digest the peptides, trypsin was added to the solutions and analyzed by MALDI-MS (*D*). The peak at *m/z* 1369.3 is the same peptide as shown in Fig. 2 *B*. The two peaks at *m/z* 1881.7 and 1752.6 are similar to each other, both containing the same disulfide bond, with the mass difference from the loss of an E by the V8 protease, establishing the second disulfide bond (*, peak from the oxidation of methionine).

Although several PDB files were returned as possible fold templates by the various programs, two structure files, for *Euplotes* pheromones E*r*-2 and E*r*-11, were selected by PSIPRED as "high" probability. The PHD server selected *Er*-11 as a probable fold model, while hidden Markov methods selected *Er*-2 as one of three possible templates. However, none of these servers agreed on the alignment of the pheromones with the template. A multiple alignment (Fig. 4 *B*) of the *Euplotes* sequences for which there is a structure correctly aligned the former according to the location of the helices. A dendrogram of the CLUSTALW pair scores for an alignment of the *Euplotes* pheromones and the *Aplysia* sequences (Fig. 4 *C*) illustrates that the E*r* sequences individually align with attractins nearly as well as they align with each other. However, there is insufficient identity to designate the E*r*s as true structural homologs of attractin (Rost, 1999).

Other templates suggested by the different methods could be rejected based on our biophysical characterization of attractin. For example, zinc-finger motifs did not provide a compact and ordered structure to account for the high resistance of attractin to proteolysis. Also, mass spectroscopic characterization of attractin is not consistent with a tightly bound metal ion. (However, as the activity assay for attractin is performed in seawater, we cannot rule out involvement of a metal ion cofactor.) None of the suggested templates, using the alignments given by the programs, satisfied the disulfide-bonding pattern determined for attractin.

Three-dimensional modeling of the attractin structure

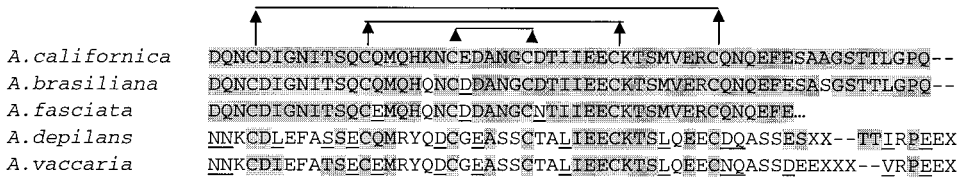
Although the disulfide bonds of attractin were different from that of the E*r*-pheromones, which have a I–IV, II–VI, III–V pattern (Brown et al., 1993; Stewart et al., 1992), a close inspection of the E*r* structures revealed that only one of the three helices needed to be moved in any of the structures to accommodate the disulfide bonding pattern of the *Aplysia* attractins. Indeed, for the structure calculation of E*r*-10 (Brown et al., 1993), using different disulfide-bond pairing constraints (or none at all) did not affect the overall topology of the helices in the final structure bundle. We thus tried various alignments with the E*r* sequences, substituting constraints for the attractin disulfide bonds, and built models using our EXDIS/FANTOM or EXDIS/DIAMOD/FANTOM program suites (Fig. 5 *A*). Of these models, on the basis of the energies after FANTOM minimization (Table 1) and the PROCHECK results (Table 2), we selected two models, based on the E*r*-2 or E*r*-11 structure with the disulfide-bond constraints determined as above for attractin (Fig. 5 *B*).

The four models, regardless of which *Euplotes* pheromone was chosen, showed the highly conserved (I/L)IEECKTS (residues 29–36) area on the surface of one face of the 3-helical bundle. In model I, the area is on the surface exposed helix 3, while in model III it lies equally exposed on helix 1. We suggest that the conserved sequence may represent a consensus attractin "active site motif" or area for interaction with the receptor. The COOH-terminal area of the ciliate pheromones has been implicated in specific binding to their receptors (see Discussion).

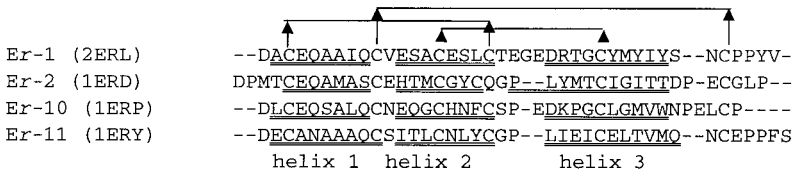
Circular dichroism

A spectrum of recombinant attractin (Fig. 6) revealed a largely helical structure for the pheromone, consistent with our models.

A.



B.



C.

Sequence:	1	2	3	4	5	6	7	8
1: *A. californica* 58 aa								
2: *A. brasiliana* 58 aa	93							
3: *A. fasciata* 47 aa	91	95						
4: *A. depilans* 55 aa	36	38	40					
5: *A. vaccaria* 54 aa	35	37	44	83				
6: Er-1 (2ERL) 40 aa	20	17	20	12	15			
7: Er-2 (1ERD) 40 aa	22	22	15	7	17	20		
8: Er-10 (1ERP) 38 aa	18	15	21	7	18	34	23	
9: Er-11 (1ERY) 39 aa	20	20	20	23	23	30	25	28

FIGURE 4 (*A*) CLUSTALW alignment of attractin sequences from five *Aplysia* species (S. D. Painter and G. T. Nagle, manuscript in preparation). Numbers above the sequences refer to the *A. californica* attractin sequence and the arrows indicate the position of the three disulfides determined for this protein. Residues identical with this sequence are shaded gray; similar residues are underlined, CLUSTALW inserted gaps are indicated by a hyphen, and the disulfide pattern is indicated by arrows. Sequence positions where the amino acid could not be absolutely identified in the sequencing cycle are designated by X; the sequence of *A. fasciata* was not completely determined. Note conservation of the six cysteine residues and an acidic/hydrophobic region (I/L)IEECKTS in all the attractins. (*B*) CLUSTALW accurately aligns the sequences of the four pheromones from *Euplotes raikovi (*E*r)* for which a structure is known according to their secondary structure (residues in each helix are underlined). The arrows show the pattern of disulfide bonds. (*C*) Matrix representation of the pairwise CLUSTALW scores for the sequences aligned in *A* and *B*. Shading has been used to emphasize that the *Aplysia* sequences form two distinct groups, with high (>90 and 83; *black on dark gray*) scores within, but only 38–44 scores between, the groups (*black on light gray*), and that the scores between the E*r*s and the attractins (*white on gray*) are in some cases as high as those between the E*r*s themselves (*white on black*). The E*r*-11 has approximately the same relationship to all of the *Aplysia* sequences, while E*r*-2 is most similar to the *A. californica* and *A. brasiliana* attractins.

DISCUSSION

Attractin is the first water-borne protein pheromone to be characterized in invertebrates. It thus provides an opportunity to better understand chemical communication among marine animals and pheromonal regulation of their reproductive behavior. We know from various types of evidence that 1) glycosylation is not necessary for function; 2) only the first 41–47 amino acids of attractin are required for activity; 3) all six cysteine residues in the first 41 residues are conserved among species, which suggests they may be bound in a pattern similar to *A. californica* attractin; 4) the structure of the non-reduced peptide is very compact and resistant to proteolysis; 5) although there is very low sequence identity between the *Aplysia* family of pheromones and any others known, several threading programs identified the *Euplotes* pheromones as having a similar fold; and 6) the CD spectrum of attractin indicates that a substantial part of the molecule is helical. We used all of this information to discriminate among models for the 3D structure of attractin.

Disulfide bonding patterns

The most crucial data we used in preparing these models were those of the disulfide bonding pattern determined by MALDI-MS characterization of proteolytic fragments of the non-reduced peptide. There are several ways to determine the disulfide bonding patterns of peptides and proteins (Chang, 1996; Jones et al., 1998; Takahashi and Hirose, 1990; Wu and Watson, 1997; Zhou and Smith, 1990). We chose to use MALDI-MS to analyze the peptides produced by the digestion of attractin because the method requires only a few nanomoles of protein to give reliable data. Furthermore, this mass spectrometric technique allows desorption/ionization of biological samples with minimal fragmentation (Hoffmann et al., 1996; Thiede et al., 1995).

A

```
I
attractin DQNCDIGNITSQCQMQHKNCEDANGCDTIIEECKTSMVERCQNQEFESAAGSTTLGPQ
1ERY      -------DECANAAAQCSITLCNLYCGPLIEICELTVMQNCEPPFS------------
                  ....  *   .     *  .** *  .... *.
II
attractin DQNCDIGNITSQCQMQHKNCEDANG---CDTIIEECKTSMVERCQNQEFESAAGSTTLGPQ
1ERY      ----------DECANAAAQCSITLCNLYCGPLIEICELTVMQNCEPPFS------------

III
attractin DQNcDIGNITSQcQMQHKNcEDANGcDTIIEEcKTSMVERcQNQEFESAAGSTTLGPQ--
1ERD      --------------------DPMTcEQAMAScEHTMcGYcQGPLYMTcIGITTDPEcGL
                              *   *.  .  *  .*   **.  . .. * **
IV
attractin --DQNcDIGNITSQcQMQHKNcEDANGcDTIIEEcKTSMVERcQNQEFESAAGSTTLGPQ
2ERL      DAcEQAAIQcVESAcESLcTEGEDRTGcYMYIYSNcPPYV--------------------
             ... *  . * *.      ** .**   *       *
```

B

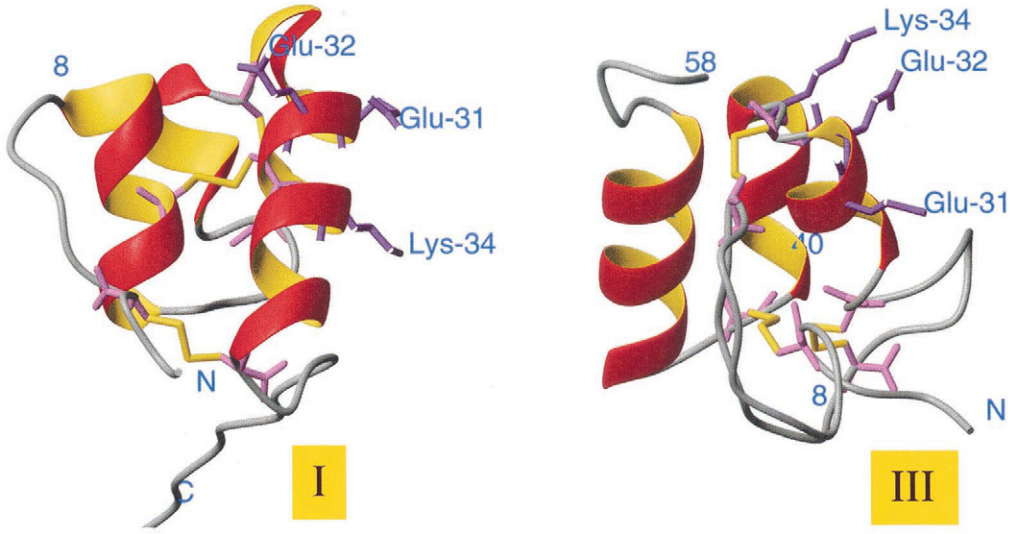


FIGURE 5 (*A*) Alignments for template-based models. CLUSTALW alignments of *A. californica* attractin with E*r*-1, E*r*-2, and E*r*-11 pheromones (PDB files 2ERL, 1ERD, and 1ERY, respectively) were prepared using a high gap penalty. Model II is from a multiple sequence alignment of the four E*r* sequences for which a structure is available with four attractin sequences. Template fragments used to extract backbone dihedral angle and distance constraints are indicated by shading. (*B*) Ribbon diagram of structural models I and III for attractin, using distance constraints extracted from the templates (1ERY and 1ERD, respectively) NMR structure files, and the disulfide constraints for attractin. The charged side chains in the conserved region are labeled.

Indeed, the method is sensitive enough for analysis of peptides from single cells and cellular subsections (Jimenez et al., 1994; Li et al., 1998; Worster et al., 1998). Cleaving attractin with trypsin and V8 protease produced several unique, singly disulfide-bonded peptides that both eliminated the E*r*-disulfide arrangement and allowed the determination of the attractin disulfide pattern.

The disulfide-bonding pattern is consistent with the resistance of most of the interior sequence of the peptide to proteolysis, as the first and last cysteines are bonded to one another. Multiple cleavage products were identified in the mass spectra that are consistent with the proposed attractin disulfide pattern.

Although *Euplotes* pheromones were used to model attractin, their disulfide-bonding pattern is different. In the MALDI-MS analysis of the enzymatic digestion of attractin, no cleavage peptides were observed that could be assigned to the *Euplotes* disulfide pattern. For example, the II–VI bond found for the E*r* pheromones is not possible in attractin based on the results of the V8 protease digestion shown in Figs. 2 and 3. Two peaks, *m/z* 1644.8 and *m/z* 1367.6, are assigned to peptides that include VI, where the latter arises from the first by V8 cleavage before the terminal FE. As there are no V8 sites between II and III, a VI disulfide-bonded to II would yield a large, easily observed cleavage product with a significantly different mass from any found.

Analysis of the disintegrins, another family of small, disulfide-rich proteins, has shown that disulfide bonding patterns are not necessarily conserved, even among proteins with high functional and sequence identity (Adler et al., 1993; Jia et al., 1997). Constraints from disulfide-bonding

TABLE 1 Energies (in kcal/mol) for the model structures from the alignments in Fig. 5*A* after mimimization with the FANTOM program. Models I–III all have acceptably low energy

Model	I	II	III	IV
Total energy	−181	−163	−147	1348
Electrostatic	−36	−13	46	65
Hydrogen bond	−87	−73	−82	−66
Lennard-Jones	−199	−253	−278	26
Torsional	110	124	155	226
Disulfide	32	27	12	1096
Dihedral angle	73	50	29	271
Upper distance limit	58	33	18	341
Lower distance limit	19	65	8	1263

patterns can facilitate structure determination, but on their own will not provide a structure.

Models of attractin structure

The attractin family of proteins is not closely related to any other sequences in the protein databases. The *A. depilans* and *A. vaccaria* sequences form one subfamily with ~40% identity to the *A. californica*, *A. brasiliana*, and *A. fasciata* sequences (Fig. 4). Our decision to use the *Euplotes* pheromones as template structures for the models was based on their similarity in biological function, areas of sequence similarity, and the results of fold-recognition methods. Other well-studied peptide pheromones, including yeast rhodotorucine A (YPEISWTRNGC(S-farnesyl)-OH; Akada et al., 1989) and sodefrin from the newt (SIPSKDALLK; Iwata et al., 1999; Kikuyama and Toyoda, 1999; Kikuyama et al., 1995) are much shorter and are cleaved from longer precursors.

The E*r*-pheromones are quite diverse in sequence (Fig. 4, *B* and *C*), with only the six cysteines and an NH_2-terminal aspartic acid absolutely conserved, yet their NMR structures are very similar, with differences primarily in the length of the helices (Luginbuhl et al., 1996; Luporini et al., 1996; Mronga et al., 1994). The third helix and COOH-terminal region of the E*r*-pheromones contain their most distinctive hallmarks and are believed to account for the specificity of receptor binding between the peptides. The E*r*-receptors, which probably arise by alternative splicing from the same gene for the E*r*-pheromones (Meyer et al., 1992; Miceli et al., 1991), each contain a copy of a pheromone sequence secreted by the cell (Luporini et al., 1996). The alignment of E*r*-1 molecules in a crystal lattice was used to suggest that the pheromone receptor interaction occurs by an antiparallel alignment of the third helix of the complementary sequences (Weiss et al., 1995). This helix, and part of the first helix, may also account for the E*r*-1 pheromone's ability to bind to the mammalian cell receptor for IL-2 (Vallesi et al., 1998).

TABLE 2 PROCHECK results for the four models

Model	I	II	III	IV
Ramachandran core/allowed/disallowed	100/00/00	96.1/2.0/1.9	92.1/3.9/4.0	94.1/3.9/2.0
Bad contacts	4	1	4	3
Planar groups	100%	100%	100%	100%
Bond angles	94.5%	94.2%	96.4%	96.3%
Bond lengths	100%	100%	100%	100%

The distribution of backbone dihedral angles (phi, psi) among core/allowed/disallowed regions in percentage are indicated for the Ramachandran plot. For all four models there were only a few bad contacts between two atoms, and the percentage of the bond angles, planar groups, and bond lengths within the allowed limits was high.

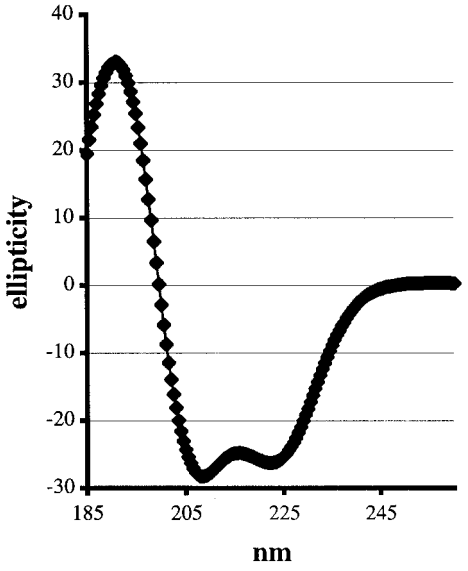


FIGURE 6 The CD spectrum of attractin (0.3 mg/ml in NaP) is consistent with the predominantly α-helical structure suggested by the models.

The third helical region of E*r*-11 aligned with the conserved (I/L)IEECKTS (residues 29–36) area of the attractin sequences in the CLUSTALW alignment. This strictly conserved region was surface-exposed in all of our models based on the E*r*-2 and E*r*-11 structures. As attractins cross-react in multiple species of *Aplysia*, we suggest that this region is important for biological activity. The size and composition of multiple cysteine loops may also be an important determinant of receptor binding, as is the case for epidermal growth factor (van de Poll et al., 1998). In previous studies (Painter et al., 1998), *A. californica* attractin peptides purified from egg cordon eluates, down to ~41 residues, were active in T-maze bioassays. These studies also showed that biological cleavage of attractin occurs primarily from the COOH terminus. Post-translational modifications in the COOH-terminal region in *A. depilans* and *A. vaccaria* attractins (unpublished observations) may serve to prolong the half-life and biological activity of the peptides in seawater.

Predictions testable with site-directed mutagenesis

Although the models shown here for attractin differ, they can serve as starting points for design of peptides and structural mimics to illuminate the functional role of areas of the protein. The models suggest a very compact structure for attractin, consistent with the results of proteolysis. Based on our models, mutation of certain residues predicted to be at the surface, particularly the glutamic acid residues 31 and 32, should affect activity. Other areas of the peptide that may be important for activity, as suggested by sequence conservation, are the hydrophilic residues at the NH_2-terminus and around the last cysteine. Site-directed mutagenesis studies, as well as characterization of the attractin receptor, promise to reveal the function of this system in the mollusk and related organisms. We anticipate that this method of combining biophysical data with protein fold recognition will be particularly useful for analyzing the probable structure of other protein targets identified by the genomic initiatives.

Note added in proof: Preliminary characterization of recombinant attractin with NMR has confirmed one of the disulfide bonds and the helicity of the structure (Garimella, R., Y. Xu, C. H. Schein, G. Nagle, and W. Braun, unpublished data).

We thank B. Clough for expert technical assistance, Lucy Lee for circular dichroism spectra, C. J. Orlea for editing help, and acknowledge the University of Texas Medical Branch (UTMB) Protein Chemistry Laboratory, which is supported by the UTMB Educational Cancer Center, for compositional and microsequence analyses.

This work was funded by grants from the UTMB Sealy Center for Structural Biology (to G.T.N., W.B.); Texas Higher Education Coordinating Board Advanced Technology Program (004952-0084-1999 to W.B. and 004952-0002-1999 to G.T.N.); National Science Foundation (IBN-9985778 to S.D.P., DBI-9714937 to W.B., and CHE-9877071 to J.V.S.); National Institutes of Health (NS 31609 to J.V.S.); Department of Energy (DE-FG03-96ER62267 to W.B.); and the Sealy and Smith Foundation.

REFERENCES

Abe, H., W. Braun, T. Noguti, and N. Go. 1984. Rapid calculation of first and second derivatives of conformational energy with respect to dihedral angles for proteins. General recurrent equations. *Comput. Chem.* 8:239–247.

Adler, M., P. Carter, R. A. Lazarus, and G. Wagner. 1993. Cysteine pairing in the glycoprotein IIbIIIa antagonist kistrin using NMR, chemical analysis, and structure calculations. *Biochemistry.* 32:282–289.

Akada, R., K. Minomi, J. Kai, I. Yamashita, T. Miyakawa, and S. Fukui. 1989. Multiple genes coding for precursors of rhodotorucine A, a farnesyl peptide mating pheromone of the basidiomycetous yeast *Rhodosporidium toruloides*. *Mol. Cell. Biol.* 9:3491–3498.

Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25:3389–3402.

Brown, L. R., S. Mronga, R. A. Bradshaw, C. Ortenzi, P. Luporini, and K. Wüthrich. 1993. Nuclear magnetic resonance solution structure of the pheromone E*r*-10 from the ciliated protozoan *Euplotes raikovi*. *J. Mol. Biol.* 231:800–816.

Chang, J. 1996. The disulfide folding pathway of tick anticoagulant peptide (TAP), a Kunitz-type inhibitor structurally homologous to BPTI. *Biochemistry.* 35:11701–11709.

Coggeshall, R. E. 1972. The structure of the accessory genital mass in *Aplysia californica*. *Tissue Cell.* 4:105–127.

Coligan, J. E. 1997. Current Protocols in Protein Science. J. E. Coligan, B. M. Dunn, H. L. Ploegh, D. W. Speicher, and P. T. Wingfield, editors. John Wiley and Sons, Inc., New York.

Cuff, J. A., and G. J. Barton. 1999. Evaluation and improvement of multiple sequence methods for protein secondary structure prediction. *Proteins.* 34:508–519.

Fan, X., B. Wu, G. T. Nagle, and S. D. Painter. 1997. Molecular cloning of a cDNA encoding a potential water-borne pheromonal attractant released during *Aplysia* egg laying. *Mol. Brain Res.* 48:167–170.

Fraczkiewicz, R., and W. Braun. 1998. Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *J. Comp. Chem.* 19:319.

Frishman, D., and P. Argos. 1997. 75% accuracy in protein secondary structure prediction. *Proteins.* 27:329–335.

Hänggi, G., and W. Braun. 1994. Pattern recognition and self-correcting distance geometry calculations applied to myohemerythrin. *FEBS Lett.* 344:147–153.

Hawke, D. H., and P. M. Yuan. 1987. S-Pyridylethylation of Cystine Residues. User Bulletin 28, Report 28. Applied Biosystems, Foster City, CA.

Higgins, D. G., A. J. Bleasy, and R. Fuchs. 1992. CLUSTALW: improved software for multiple sequence alignment. *Comput. Appl. Biosci.* 8:189–191.

Hoffmann, E. D., J. Charette, and V. Stroobant. 1996. Mass Spectrometry. John Wiley and Sons, Chichester.

Iwata, T., K. Umezawa, F. Toyoda, N. Takahashi, H. Matsukawa, K. Yamamoto, S. Miura, H. Hayashi, and S. Kikuyama. 1999. Molecular cloning of newt sex pheromone precursor cDNAs: evidence for the existence of species-specific forms of pheromones. *FEBS Lett.* 457: 400–404.

Jia, L., X. Wang, J. D. Shannon, J. B. Bjarnason, and J. W. Fox. 1997. Function of disintegrin-like/cystein-rich domains of Atrolysin A. *J. Biol. Chem.* 272:13094–13102.

Jimenez, C. R., P. A. van Veelen, K. W. Li, W. C. Wildering, W. P. Geraerts, U. R. Tjaden, and J. van der Greef. 1994. Neuropeptide expression and processing as revealed by direct matrix-assisted laser desorption ionization mass spectrometry of single neurons. *J. Neurochem.* 62:404–407.

Jones, D. T. 1999a. GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. *J. Mol. Biol.* 287:797–815.

Jones, D. T. 1999b. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* 292:195–202.

Jones, M. D., S. D. Patterson, and H. S. Lu. 1998. Determination of disulfide bonds in highly bridged disulfide-linked peptides by matrix-assisted laser desorption/ionization mass spectrometry with postsource decay. *Anal. Chem.* 70:136–143.

Karplus, K., C. Barrett, and R. Hughey. 1998. Hidden Markov models for detecting remote protein homologies. *Bioinformatics.* 14:846–856.

Kikuyama, S., and F. Toyoda. 1999. Sodefrin: a novel sex pheromone in a newt. *Rev. Reprod.* 4:1–4.

Kikuyama, S., F. Toyoda, Y. Ohmiya, K. Matsuda, S. Tanaka, and H. Hayashi. 1995. Sodefrin: a female-attracting peptide pheromone in newt cloacal glands. *Science.* 267:1643–1645.

Koradi, R., M. Billeter, and K. Wüthrich. 1996. MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graphics.* 14:51–55.

Laskowski, R. A., M. W. MacArthur, D. S. Moss, and J. M. Thornton. 1993. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 26:283–291.

Li, L., T. P. Moroz, R. W. Garden, P. D. Floyd, K. R. Weiss, and J. V. Sweedler. 1998. Mass spectrometric survey of interganglionically transported peptides in *Aplysia*. *Peptides.* 19:1425–1433.

Luginbuhl, P., J. Wu, O. Zerbe, C. Ortenzi, P. Luporini, and K. Wüthrich. 1996. The NMR solution structure of the pheromone E*r*-11 from the ciliated protozoan *Euplotes raikovi*. *Protein Sci.* 5:1512–1522.

Luporini, P., C. Miceli, C. Ortenzi, and A. Vallesi. 1996. Ciliate pheromones. *Prog. Mol. Subcell. Biol.* 17:80–104.

Meyer, F., H. J. Schmidt, and K. Heckmann. 1992. Pheromone 4 gene of *Euplotes actocarinatus*. *Dev. Genet.* 13:16–25.

Miceli, C., A. La Terza, R. A. Bradshaw, and P. Luporini. 1991. Structural characterization of mating pheromone precursors of the ciliate protozoan *Euplotes raikovi*. High conservation of pre and pro regions versus high variability of secreted regions. *Eur. J. Biochem.* 202:759–764.

Mronga, S., P. Luginbuhl, L. R. Brown, C. Ortenzi, P. Luporini, R. A. Bradshaw, and K. Wüthrich. 1994. The NMR solution structure of the pheromone E*r*-1 from the ciliated protozoan *Euplotes raikovi*. *Protein Sci.* 3:1527–1536.

Mumenthaler, C., and W. Braun. 1995. Predicting the helix packing of globular proteins by self-correcting distance geometry. *Protein Sci.* 4:863–871.

Painter, S. D., M. G. Chong, M. A. Wong, A. Gray, J. G. Cormier, and G. T. Nagle. 1991. Relative contributions of the egg layer and egg cordon to pheromonal attraction and the induction of mating and egg-laying behavior in *Aplysia*. *Biol. Bull.* 181:81–94.

Painter, S. D., B. Clough, R. W. Garden, J. V. Sweedler, and G. T. Nagle. 1998. Characterization of *Aplysia* attractin, the first water-borne peptide pheromone in invertebrates. *Biol. Bull.* 194:120–131.

Rost, B. 1996. PHD: predicting one-dimensional protein structure by profile based neural networks. *Methods Enzymol.* 266:525–539.

Rost, B. 1999. Twilight zone of protein sequence alignments. *Protein Eng.* 12:85–94.

Schaumann, T., W. Braun, and K. Wüthrich. 1990. A program, FANTOM, for energy refinement of polypeptides and proteins using a Newton-Raphson minimizer in the torsion angle space. *Biopolymers.* 29: 679–694.

Smith, J. S., B. T. Miller, S. L. Knock, and A. Kurosky. 1991. Biotinylated peptides/proteins. 1. Determination of stoichiometry of derivatization. *Anal. Biochem.* 197:247–253.

Soman, K. V., T. Midoro-Horiuti, J. C. Ferreon, R. M. Goldblum, E. G. Brooks, A. Kurosky, W. Braun, and C. H. Schein. 2000. Homology modeling and characterization of IgE binding epitopes of mountain cedar allergen jun a 3. *Biophys. J.* 79:1601–1609.

Stewart, A. E., S. Raffioni, T. Chaudhary, B. T. Chait, P. Luporini, and R. A. Bradshaw. 1992. The disulfide bond pairing of the pheromones E*r*-1 and E*r*-2 of the ciliated protozoan *Euplotes raikovi*. *Protein Sci.* 1:777–785.

Takahashi, N., and M. Hirose. 1990. Determination of sulfhydryl groups and disulfide bonds in a protein by polyacrylamide gel electrophoresis. *Anal. Biochem.* 188:359–365.

Thiede, B., B. Wittmann-Liebold, M. Bienert, and E. Krause. 1995. MALDI-MS for C-terminal sequence determination of peptides and proteins degraded by carboxypeptidase Y and P. *FEBS Lett.* 357:65–69.

Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22:4673–4680.

Vallesi, A., G. Giuli, P. Ghiara, G. Scapigliati, and P. Luporini. 1998. Structure-function relationships of pheromones of the ciliate *Euplotes raikovi* with mammalian growth factors: cross-reactivity between E*r*-1 and interleukin-2 systems. *Exp. Cell Res.* 241:253–259.

van de Poll, M. L., M. J. van Vugt, A. E. Lenferink, and E. J. van Zoelen. 1998. Identification of the minimal requirements for binding to the human epidermal growth factor (EGF) receptor using chimeras of human EGF and an EGF repeat of Drosophila Notch. *J. Biol. Chem.* 273:16075–16081.

Weiss, M. S., D. H. Anderson, S. Raffioni, R. A. Bradshaw, C. Ortenzi, P. Luporini, and D. Eisenberg. 1995. A cooperative model for receptor recognition and cell adhesion: evidence from the molecular packing in the 1.6-A crystal structure of the pheromone E*r*-1 from the ciliated protozoan *Euplotes raikovi*. *Proc. Natl. Acad. Sci. U.S.A.* 92: 10172–10176.

White, J. V., C. M. Stultz, and T. F. Smith. 1994. Protein classification by stochastic modeling and optimal filtering of amino-acid sequences. *Math. Biosci.* 119:35–75.

Worster, B. M., M. S. Yeoman, and P. R. Benjamin. 1998. Matrix-assisted laser desorption/ionization time of flight mass spectrometric analysis of the pattern of peptide expression in single neurons resulting from alternative mRNA splicing of the FMRFamide gene. *Eur. J. Neurosci.* 10:3498–3507.

Wu, J., and J. T. Watson. 1997. A novel methodology for assignment of disulfide bond pairings in proteins. *Protein Sci.* 6:391–398.

Zhou, Z. R., and D. L. Smith. 1990. Assignment of disulfide bonds in proteins by partial acid hydrolysis and mass spectrometry. *J. Protein Chem.* 9:523–532.

Zhu, H., C. H. Schein, and W. A. Braun. 2000. MASIA: recognition of common patterns and properties in multiple aligned protein sequences. *Bioinformatics.* 16:1–2.