Protein Science (1994), *3*:1537–1546. Cambridge University Press. Printed in the USA.

Structure comparison of the pheromones Er-1, Er-10, and Er-2 from *Euplotes raikovi*

PETER LUGINBÜHL, MARCEL OTTIGER, SIGGI MRONGA, AND KURT WÜTHRICH

Institut für Molekularbiologie und Biophysik, Eidgenössische Technische Hochschule-Hönggerberg, CH-8093 Zürich, Switzerland

(RECEIVED March 23, 1994; ACCEPTED June 21, 1994)

Abstract

The NMR structures of the homologous pheromones Er-1, Er-10, and Er-2 from the ciliated protozoan *Euplotes raikovi* are compared. For all 3 proteins the molecular architecture is made up of an antiparallel 3-helix bundle. The preservation of the core part of the structure is directly manifested by similar patterns of slowed backbone amide proton exchange rates, hydrogen bond formation, and relative solvent accessibility. To align the 6 half-cystine residues in the individual sequences within the preserved 3-dimensional core structure, several deletions and insertions had to be introduced that differ from those previously proposed on the basis of the primary structures. Of special interest is a deletion in the second helix of Er-2, which is accommodated by a transition from an α-helix in Er-1 and Er-10 to a 3_{10}-helix in Er-2. The most significant structural differences are located in the C-terminal part of the proteins, which may have an important role in specific receptor recognition.

Keywords: α-helix to 3_{10}-helix transition; comparison of NMR structures; *Euplotes raikovi*; homology alignments; pheromones

Pheromones of the ciliated protozoan *Euplotes raikovi* function as cell signal molecules that control the processes of cellular conjugation by self/non-self recognition (Luporini & Miceli, 1986). Each pheromone determines a distinctive molecular phenotype ("mating type") of the cell of origin. Searches for the structural basis of pheromone functions have so far been mainly conducted on the primary structure level. The amino acid sequences for 5 of these proteins, which contain 37–40 residues (Fig. 1) (Miceli et al., 1991; Raffioni et al., 1992), suggested close relationships between the pheromones Er-1 and Er-10, and between Er-11 and Er-20, with 44% and 56% identity, respectively, whereas a larger divergence was observed for any other pair among the 5 proteins (Raffioni et al., 1992). This conclusion from the sequence comparisons was also supported by hydrophobicity analyses (Raffioni et al., 1992): Er-1 and Er-10 are distinguished from the other 3 pheromones by high hydrophilicity in the polypeptide segment 20–30, and Er-11 and Er-20 are unique among the 5 sequences by having high hydrophobicity in the region 10–20. Apart from 6 half-cystine residues (which will be referred to as Cys I to Cys VI) and the N-terminal Asp, which are common to all 5 sequences, there are 11 additional sites in which 3 or more of the pheromones share a common residue. Seven of these are clustered in the N-terminal region from residues 1 to 13, including all the residues between Cys I and Cys II (Fig. 1). The high conservation of this segment of the pheromone polypeptide chain may reflect a common scaffold-type role in the 3-dimensional structure, whereas the more variable C-terminal half of the sequence may be more directly related to unique properties of the individual pheromones, such as specific receptor interactions (Raffioni et al., 1992).

Structure–function correlations in the pheromones in Figure 1 can now also be investigated on the level of the 3-dimensional structures, since the solution structures of Er-1 (Mronga et al., 1994 [companion paper]), Er-10 (Brown et al., 1993), and Er-2 (Ottiger et al., 1994 [companion paper]) have been determined by NMR spectroscopy. At first glance, the 3 protein structures have identical architectures (Fig. 2; Kinemage 1); however, even the backbone fold alone indicates that there are subtle differences between the individual proteins. This paper presents a detailed comparison of the 3 NMR structures and identifies common traits as well as unique features of the individual proteins, which might provide a rationale for the different specificities of the 3 species.

Results and discussion

The comparative studies reported herein are based on earlier structure determinations (Mronga et al., 1994; Ottiger et al., 1994), except that the structure of Er-10 was recalculated from the previously published input data (Brown et al., 1993), using

Reprint requests to: Kurt Wüthrich, Institut für Molekularbiologie und Biophysik, Eidgenössische Technische Hochschule-Hönggerberg, CH-8093 Zürich, Switzerland.

```
Cys            I     II   III  IV          V           VI

                      10        20        30        40
Er-1     D--ACEQAAIQCVESACESLCTEGEDRTGCYMYIYSN--CPPYV

Er-10    D--LCEQSALQCNEQGCHNFCS-PEDKPGCLGMVWNPELCP

Er-2     DPMTCEQAMASCEHTMC-GYCQ-GPLYMTCIGITTDPE-CGLP

Er-11    D--ECANAAAQCSITLCNLYCG--PLIEICELTVMQN--CEPPFS

Er-20    D--ICDDAVAQCSMTLCQL-CY---NTEICELSVIGS--CQPPFS
```

Fig. 1. Amino acid sequences for 5 pheromones from *Euplotes raikovi*. The numeration at the top is for E*r*-1, and the Roman numerals indicate the positions of the Cys residues in all 5 proteins. The previously published sequence alignment (Raffioni et al., 1992) has been modified by shifting the E*r*-2 segment Gly 22–Thr 27 by 1 residue toward the C-terminus, and the E*r*-1 segment Met 30–Asn 35 by 1 residue toward the N-terminus. The sequence alignments of E*r*-11 and E*r*-20 were also modified accordingly (see text).

the same protocol as for E*r*-1 and E*r*-2 to enable an unbiased comparison of all 3 proteins. The effects of the recalculation of the E*r*-10 structure are briefly discussed in the following section. Table 1 gives a summary for all 3 proteins of the NMR constraints used as input for the structure calculations, and of the quantitative characterization of the structure determinations by residual constraint violations. These data and the RMS deviation (RMSD) values calculated for selected polypeptide segments (Table 2) show that high-quality structures were obtained for all 3 proteins. Therefore, even relatively small structure variations may be significant at the level of the precision of the structure determinations, depending in detail on the polypeptide segments considered.

Recalculation of the Er-10 structure

The structures of E*r*-1, E*r*-2, and E*r*-10 were all calculated with the program DIANA (Güntert et al., 1991) using redundant di-

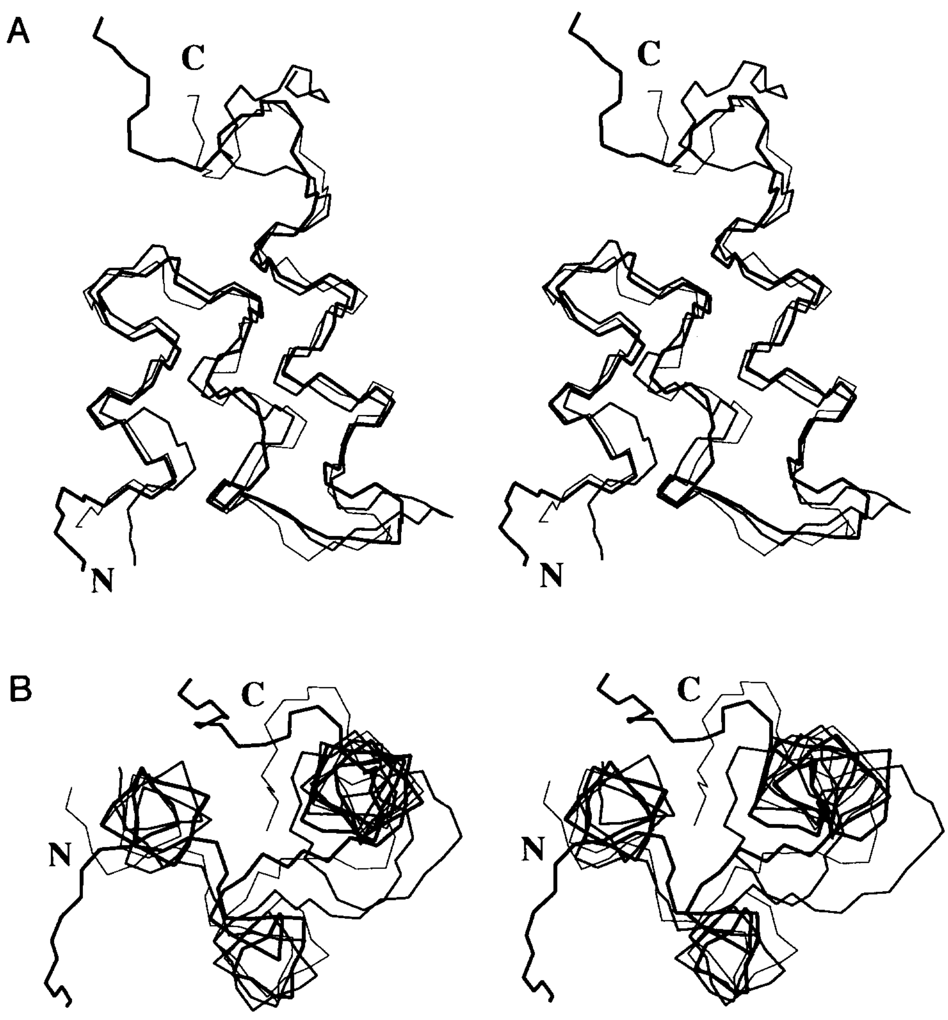

Fig. 2. Stereo views of the least-squares superposition of the mean structures of E*r*-1 (medium line), E*r*-10 (thin line), and E*r*-2 (thick line). The backbone atoms N, C^{α}, and C′ of residues 2–15 and 24–34 of E*r*-1, 2–15 and 23–33 of E*r*-10, and 4–17 and 24–34 of E*r*-2 have been superimposed for best fit. N and C indicate the approximate locations of the N- and C-termini of the pheromone structures. **A:** Front view. **B:** Top view.

Table 1. *Survey of the NMR constraints used as input for the structure calculation and quantitative characterization of the 20 best conformers of Er-1, Er-10, and Er-2 after energy refinement*[a]

	Er-1	Er-10	Er-2
NOE upper distance constraints	503	375	603
ϕ Dihedral angle constraints	29	28	27
χ^1 Dihedral angle constraints	14	19	19
DIANA target function (Å^2)[b]	1.01 ± 0.14	0.94 ± 0.22	0.68 ± 0.08
AMBER energy (kcal/mol)[b]	−1,106 ± 30	−1,146 ± 40	−896 ± 19
Residual NOE constraint violations (Å)[b]			
Maximum	0.10 ± 0.00	0.10 ± 0.01	0.10 ± 0.00
Sum	5.0 ± 0.3	3.3 ± 0.2	5.3 ± 0.2
Residual dihedral angle constraint violations (deg)[b]			
Maximum	2.0 ± 0.3	2.2 ± 0.2	2.0 ± 0.2
Sum	8.4 ± 2.2	15.8 ± 2.8	10.9 ± 2.5

[a] The structure calculations were performed with the program DIANA (Güntert et al., 1991), using standard weights for the different types of experimental constraints. For the energy refinement with the program OPAL, we used the AMBER all-atom force field (Weiner et al., 1986) with a constant dielectric permittivity of $\epsilon_r = 1$. Details are given in the original descriptions of the structure determinations for Er-1 and Er-2 (Mronga et al., 1994; Ottiger et al., 1994 [companion papers]).

[b] The values given are the mean ± standard deviation among the 20 conformers used to represent the solution structure.

hedral angle constraints (REDAC) (Güntert & Wüthrich, 1991) to improve convergence. A difference between the protocols of the original Er-10 calculation (Brown et al., 1993) and the calculations used by Mronga et al. (1994) and Ottiger et al. (1994) is that 2 REDAC cycles were originally applied and only 1 REDAC cycle was used in the new calculations. Of each protein, the 20 conformers with the lowest DIANA target function values were subjected to restrained energy minimization using the AMBER all-atom force field (Weiner et al., 1986). However, whereas Er-10 was originally minimized in vacuo using a distance-proportional dielectric permittivity, the presently used structures of all 3 proteins were minimized after immersing each conformer in a 6-Å-thick shell of water molecules and employing a constant dielectric permittivity for the electrostatic interactions.

Comparison of the original Er-10 structure (Brown et al., 1993) with the recalculated one shows that the latter has a slightly larger average backbone RMSD value. Superposition of the atoms N, C^α, and C′ of the best-defined polypeptide segments comprising residues 2–15 and 23–33 in the 20 conformers used to represent the NMR structure results in RMSD values relative to the mean of 0.33 ± 0.05 Å for the recalculated structure (Table 2) and 0.24 ± 0.04 Å for the original structure (Brown et al., 1993). The RMSD value for the corresponding superposition of the 2 mean structures is 0.32 Å. From these numbers it is clear that the deviation between the original and the new mean structure is not significant at the level of the precision of the 2 structure calculations. The most significant differences are found in the second helix, which is somewhat distorted in the original Er-10 structure but is more regular in the new structure. The backbone RMSD value calculated for the atoms N, C^α, and C′ of residues 14–18 (Table 3) of the new mean structure relative to a regular α-helix is 0.59 Å, as compared to a corresponding value of 0.75 Å for the original structure, and the previously observed non-α-helical hydrogen bond Gly 14 O′–Cys 19 NH is not present in the new structure.

Table 2. *RMSD values for the superposition of Er-1, Er-10, and Er-2 calculated for selected backbone atoms*[a]

	Er-1	Er-10	Er-2
Er-1	0.29 ± 0.06	1.00	1.11
Er-10		0.33 ± 0.05	0.77
Er-2			0.21 ± 0.04

[a] The RMSD values are given in Å. The backbone atoms N, C^α, and C′ of the following corresponding polypeptide segments (Fig. 3) have been considered in the RMSD calculations: 2–15 and 24–34 for Er-1, 2–15 and 23–33 for Er-10, and 4–17 and 24–34 for Er-2. On the diagonal, the average value and the standard deviation of the pairwise RMSD values for the 20 energy-refined conformers relative to their mean structure is given for each protein. Above the diagonal, the backbone RMSD values between the mean structures of the different proteins are given.

Table 3. *RMSD values calculated for the backbone atoms N, C^α, and C′ for superpositions of corresponding 5-residue segments of helix 2 from Er-1, Er-10, and Er-2 with a regular α-helix and a regular 3_{10}-helix*

	RMSD (Å) relative to	
	α-Helix[a]	3_{10}-Helix[b]
Er-1 (14–18)[c]	0.36	1.31
Er-10 (14–18)[c]	0.59	1.53
Er-2 (15–19)[c]	1.50	0.60
α-Helix	0.00	1.05
3_{10}-Helix	1.05	0.00

[a] α-Helix: $\phi = -57°$, $\psi = -47°$.

[b] 3_{10}-Helix: $\phi = -49°$, $\psi = -26°$.

[c] For each protein, the backbone atoms N, C^α, and C′ of the mean structure have been used for the calculation of the RMSD. The residue numbers of the segments considered are given in parentheses.

Comparison of the NMR structures of Er-1, Er-10, and Er-2

The structures of all 3 pheromones show the characteristic traits of highly stable, small, disulfide-rich proteins, in which the regular secondary structures are distorted by the high density of disulfide bonds (Richardson, 1981). The molecular architecture of the pheromones is characterized by a 3-helix bundle with an up-down-up topology. The axes of the 3 helices are nearly parallel, the first and second helices are connected by an extended dipeptide segment, and the second and third helices are connected by an extended loop of 2–4 residues. The residue pairing of the disulfide bridges was reported for Er-1 and Er-2 (Stewart et al., 1992), but for Er-10 only the disulfide bond Cys 10–Cys 37 can so far be identified with chemical methods (A. Stewart, S. Raffioni, & R.A. Bradshaw, unpubl. results). In the following we assume identity to Er-1 and Er-2 for the other 2 disulfide bonds in Er-10. Two of the disulfide bonds are strategically located to connect helices 1 and 2 (Cys 3–Cys 19 for Er-1 and Er-10, and Cys 5–Cys 20 for Er-2), and 2 and 3 (Cys 15–Cys 28 for Er-1, Cys 15–Cys 27 for Er-10, and Cys 17–Cys 28 for Er-2), respectively. The third disulfide bridge establishes a connection between the C-terminal region and the dipeptide link between helices 1 and 2 (Cys 10–Cys 36 for Er-1, Cys 10–Cys 37 for Er-10, and Cys 12–Cys 37 for Er-2). In all 3 proteins the 2 disulfide bridges connecting the 3 helices are in close spatial proximity, and the 2 alternative residue pairings would be nearly equally consistent with the NMR constraints. Therefore, it could not be excluded on the basis of the 3-dimensional NMR structure determination that the natural proteins might consist of mixtures of species containing 2 or even all 3 possible disulfide pairings involving Cys residues I, III, IV, and V (Fig. 2). This appeared to be a possible origin of minor conformations observed in the NMR spectra of Er-10 (Brown et al., 1993). Minor conformations were also observed in the NMR spectra of Er-1 (Mronga et al., 1994) and Er-2 (Ottiger et al., 1994), and since the chemical studies would not have detected minor components with different disulfide pairings (R.A. Bradshaw, pers. comm.), disulfide isomers could not a priori be excluded as the cause for the sample heterogeneity. However, in Er-2 the minor conformation could be shown to be a consequence of C-terminal -Xxx-Pro-OH *cis-trans* isomerism (Ottiger et al., 1994). Since Er-10 also contains a C-terminal proline (Fig. 1), it is very likely that the minor conformation is again due to *cis-trans* isomerism -Xxx-Pro-OH (Grathwohl & Wüthrich, 1976b). Based on studies with model peptides, the Pro residues near the C-terminus in Er-1 could also form conformational isomers with NMR-observable populations of species containing *cis*- and *trans*-peptide bonds (Grathwohl & Wüthrich, 1976a).

The superposition of the mean structures of Er-1, Er-10, and Er-2 in Figure 2 was calculated for minimal RMSD (Table 2) between the 2 best-defined polypeptide segments comprising the first helix and the 6 following residues, and the entire third helix, respectively, which corresponds to 64% of all residues. Whereas the best fit for superposition of the individual helices is obtained for the pair Er-1/Er-10 (data not shown), the more extensive superposition of Table 2 gives the lowest RMSD value for the pair Er-10/Er-2. This result can be rationalized by differences in the relative orientations of helix 1 and helix 3 in the 3 structures (see below).

Figure 3 provides a survey of the sequence distribution of hydrogen bonds and residues with low solvent accessibility, which presents a basis for detailed comparisons of individual regions of Er-1, Er-10, and Er-2. All 3 proteins start with a poorly defined N-terminal Asp. In Er-1 and Er-10 the first helix starts

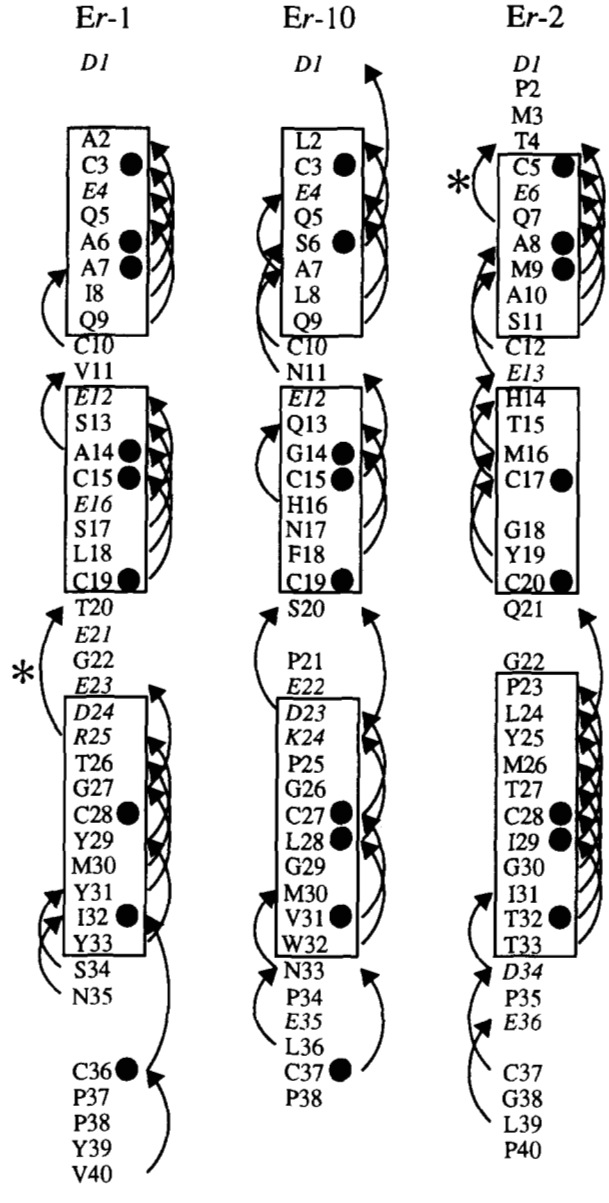

Fig. 3. Schematic comparison of structural features between Er-1, Er-10, and Er-2. The sequences are presented from top to bottom; residues with charged side chains are in italics, and the boxes indicate the locations of the helices. Backbone-backbone hydrogen bonds are shown by arrows pointing from the donor residue to the acceptor residue. Regular α-helical hydrogen bonds O_i'–HN_{i+4} are shown on the right-hand side, and other backbone-backbone hydrogen bonds on the left-hand side of the individual sequences. Two backbone–side chain hydrogen bonds that are discussed in the text are also shown; they are identified by asterisks. The filled circles indicate residues with small relative solvent accessibility. The previously published sequence alignment (Raffioni et al., 1992) has been modified by shifting the Er-2 segment Gly 22–Thr 27 by 1 residue toward the C-terminus, and the Er-1 segment Met 30–Asn 35 by 1 residue toward the N-terminus (see text).

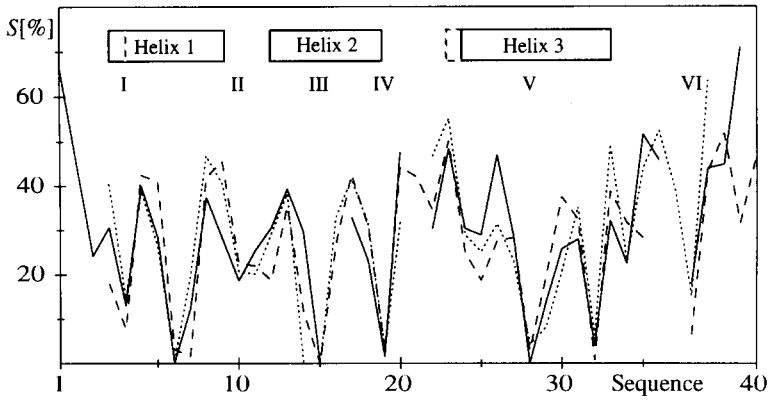

Fig. 4. Plot of the average solvent-accessible surface area (Richmond, 1984; Billeter et al., 1990) relative to the total surface area for individual residues versus the amino acid sequence. The numeration at the bottom is for E*r*-1. The boxes indicate the locations of the helices, and the Roman numerals identify the positions of the Cys residues. Broken line, E*r*-1; dotted line, E*r*-10; solid line, E*r*-2.

with residue 2, whereas in E*r*-2 the N-terminal residue is separated from helix 1 by a tripeptide loop. Helix 1 is an α-helix ending with 1 residue in a 3_{10}-helix conformation indicated by the O_i'-HN_{i+3} hydrogen bond to Cys II (Fig. 3). The only other non-α-helical hydrogen bond is found in E*r*-10, i.e., Glu 4 O′-Ala 7 NH, and an O_i'-HN_{i+5} hydrogen bond to the residue following Cys II is found in E*r*-10 and in E*r*-2 (Fig. 3). In E*r*-2, the backbone amide proton of Gln 7 forms a hydrogen bond with γO of Thr 4, indicating the formation of a "potential capping box" (Harper & Rose, 1993) where Thr 4 would act as N-cap (Richardson & Richardson, 1988). Although the hydrogen bond Thr 4 NH-Gln 7 ϵO has not been identified in the E*r*-2 structure, the slow exchange of the backbone amide proton of Thr 4 (Ottiger et al., 1994) supports the presence of such a hydrogen bond. Two residues in helix 1 are protected from the solvent in all 3 pheromone structures and may thus serve as anchor residues (Heinz et al., 1993): Cys I has a relative solvent accessibility of less than 15% in all 3 structures (Fig. 4); Ala 6 in E*r*-1, Ser 6 in E*r*-10, and Ala 8 in E*r*-2 are completely buried, with relative solvent accessibilities of 3.1%, 0.5%, and 0.1%, respectively (Fig. 4; Kinemage 2).

In all 3 proteins the start of helix 2 is separated by a single residue from Cys II, and Cys III and Cys IV are both located within helix 2 and are completely protected from solvent contact. When compared with the sequences of E*r*-1 and E*r*-10, E*r*-2 has a deletion between Cys III and Cys IV, which results in the formation of 3_{10}- rather than α-helical secondary structure (Fig. 3), as evidenced by the following observations. (1) The second helix of E*r*-2 contains all of the expected 3_{10}-helix hydrogen bonds, with the sole exception of Thr 15 O′-Gly 18 NH. (2) The Cys residues III and IV in the sequences of E*r*-1 and E*r*-10 are separated by 3 intervening residues, resulting in a translation along the α-helix axis of about 6.0 Å. In E*r*-2, the 2 Cys residues are separated by only 2 residues, for which the 3_{10}-helix provides again a translation along the helix axis of about 6.0 Å. In this way, the same positions of the Cys residues are maintained in E*r*-2 as in the other 2 pheromone structures (Fig. 5; Kinemage 2). (3) The RMSD value for the superposition of E*r*-2 residues 15–19 with a regular 3_{10}-helix is 2.5 times smaller than for the superposition with a regular α-helix (Table 3). Conversely, the backbone RMSD values for the corresponding segments of E*r*-1 and E*r*-10 are 3.6 and 2.6 times smaller for the

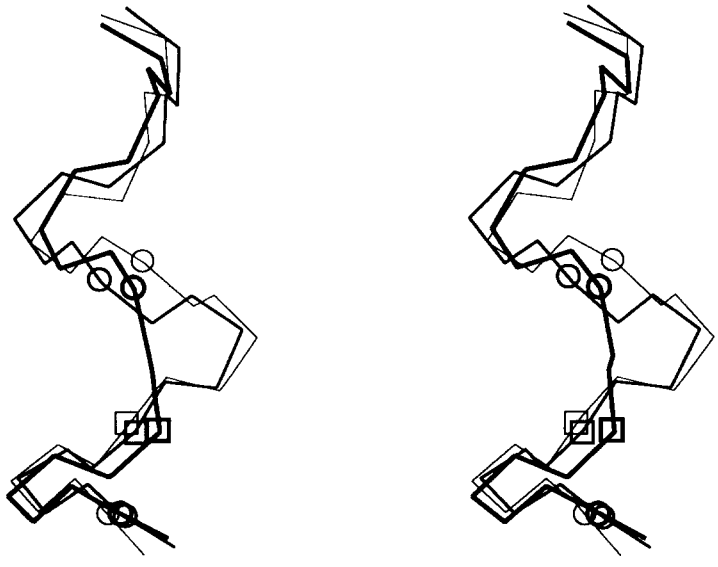

Fig. 5. Stereo view of helix 2 in the mean structures of E*r*-1 (medium line), E*r*-10 (thin line), and E*r*-2 (thick line). The backbone atoms N, C^{α}, and C′ of residues 12–15 and 18–19 for E*r*-1 and E*r*-10, and 14–17 and 19–20 for E*r*-2 have been superimposed for minimal RMSD. The circles identify the C^{α} positions of Cys III and Cys IV. The squares identify the C^{α} positions of Ser 17 (E*r*-1), Asn 17 (E*r*-10), and Gly 18 (E*r*-2).

superposition with the regular α-helix than for the superposition with the regular 3_{10}-helix (Table 3).

In the loop leading from helix 2 to helix 3, E*r*-1 contains a 1-residue insertion when compared to E*r*-10 and E*r*-2, and relative to E*r*-1 and E*r*-10 helix 3 in E*r*-2 is N-terminally elongated by the addition of Pro 23 (Fig. 3). The insertion leads to a bulging out of the loop in E*r*-1 (Fig. 2; Kinemage 2), and a Thr 20 O′–Arg 25 εNH backbone–side chain hydrogen bond is formed in the place of the backbone–backbone hydrogen bonds Ser 20 O′–Lys 24 NH in E*r*-10 and Gln 21 O′–Tyr 25 NH in E*r*-2 (Fig. 3). The insertion in the loop also causes a slight tilt of helix 3 in E*r*-1 when compared to E*r*-10 and E*r*-2 (Fig. 2). Overall, the relative orientation of helices 1 and 3 is therefore more similar between E*r*-10 and E*r*-2 than between E*r*-1 and either one of the 2 other proteins (Table 2; Fig. 2). As described in detail below, the structure comparisons in this loop region prompted us to propose a modification of the sequence alignments previously suggested on the basis of the amino acid sequences (Raffioni et al., 1992).

In all 3 pheromones, helix 3 is the longest helix. It is a regular α-helix, which ends with 1 (E*r*-10 and E*r*-2) or 2 (E*r*-1) residues in 3_{10}-helix conformation, as indicated by O'_i–HN_{i+3} hydrogen bonds (Fig. 3). In each pheromone structure, 2 residues in corresponding locations of helix 3 are well protected from the solvent (Fig. 4) and serve as anchor residues, i.e., Cys V and the residues corresponding to Ile 32 in E*r*-1 (Fig. 3; Kinemage 2).

The polypeptide segment from the end of helix 3 to the C-terminus is the most variable part of the pheromones, on the level of both the amino acid sequence (Fig. 1) and the 3-dimensional structure (Fig. 2). The corresponding residues Ser 34 in E*r*-1, Asn 33 in E*r*-10, and Asp 34 in E*r*-2 are separated from Cys VI by 1, 3, and 2 sequence positions, respectively. In E*r*-1, E*r*-10, and E*r*-2, Cys VI is separated from the C-terminus by 4, 1, and 3 residues, respectively, so that overall helix 3 is followed by a C-terminal hexapeptide in E*r*-10 and a C-terminal heptapeptide segment in both E*r*-1 and E*r*-2. The structure of the C-terminus seems to be determined primarily by the position of Cys VI within this 6- or 7-residue segment, which forms a disulfide bridge with Cys II in the extended dipeptide connecting helices 1 and 2. In E*r*-10 and E*r*-2, which contain, respectively, the segments Asn 33–Pro 34–Glu 35–Leu 36–Cys 37 and Asp 34–Pro 35–Glu 36–Cys 37 immediately following helix 3, the C-terminal part has to fold back on top of the dipeptide link between helices 1 and 2 to form the disulfide bridge (Fig. 6). In E*r*-1, however, with the sequence Ser 34–Asn 35–Cys 36, the disulfide bridge can be readily formed with the C-terminal part forming an extended loop on top of helix 3 (Fig. 6; Kinemage 2).

Revised sequence alignment for the E. raikovi pheromones

As indicated in the legends to Figures 1 and 3, the present analysis of the 3-dimensional structures of E*r*-1, E*r*-10, and E*r*-2 suggested some modifications of the sequence alignment relative to that proposed on the basis of the amino acid sequences (Raffioni et al., 1992). In E*r*-1, the hexapeptide segment corresponding to positions 30–35 in Figure 1 was shifted toward the N-terminus by 1 position, so that relative to E*r*-10 the 2-residue deletion immediately preceding Cys VI is now the only deletion between Cys V and Cys VI (Fig. 1). In E*r*-2, the hexapeptide corresponding to positions 22–27 in Figure 1 was shifted toward the C-terminus by 1 position, so that relative to E*r*-1 the deletion between Cys IV and Cys V is now in the nearest-next position to Cys IV rather than immediately preceding Cys V. In Figure 1, corresponding shifts were made for E*r*-11 and E*r*-20, i.e., the peptide segments preceding Cys V (PLIEI in E*r*-11, NTEI in E*r*-20) were shifted by 1 position toward the C-terminus, and the segments following Cys V (ELTVMQN in E*r*-11, ELSVIGS in E*r*-20) were shifted by 1 position toward the N-terminus.

In the group of E*r*-1, E*r*-10, and E*r*-2, the near-perfect coincidence of residues with small solvent accessibility (Fig. 4) is perhaps the most convincing single argument in favor of the revised alignment. The following are additional improved fits of structural details. (1) With the revised alignment it is readily rationalized that the same type of backbone–backbone hydrogen bond is observed between Ser 20 and Lys 24 in E*r*-10, and between Gln 21 and Tyr 25 in E*r*-2 (Fig. 3), and related to this one has closely similar relative orientations of helices 1 and 3 in these 2 proteins (Fig. 2; Table 2). (2) The revised alignment of the C-terminal part of helix 3 ensures that the anchor residue Ile 32 of E*r*-1 coincides with the corresponding anchor residues Val 31 in E*r*-10 and Thr 32 in E*r*-2.

The nature of the α- to 3_{10}-helix transition of helix 2 in Er-2

Recently, recombinant techniques have been employed to probe the 3-dimensional structure response to insertions and deletions in proteins, and some general rules on structure variations upon insertions or deletions in helices have been proposed (Sondek & Shortle, 1990, 1992; Marti et al., 1992; Heinz et al., 1993; Keefe et al., 1993). Here we shall consider the structure variations observed in helix 2 of the 3 natural proteins E*r*-1, E*r*-10, and E*r*-2 (Fig. 5; Kinemage 2) in light of the rules proposed from work with other proteins.

An important factor determining the structural response to an insertion seems to be the maintenance of the buried hydrophobic interface between the helix and the rest of the protein (Heinz et al., 1993). Insertions may lead to translocations of residues from the helix to the preceding or following loop, resulting in a register shift, or they may cause a "looping-out" in the form of an "α-aneurism" (Keefe et al., 1993), an "$\alpha \rightarrow \pi$ bulge" (Schiering et al., 1991; Harrison et al., 1994) or a "$3_{10} \rightarrow \alpha$ bulge" (Kavanaugh et al., 1993). While a register shift preserves the geometry of the helix, it results generally in new packing interactions between the helix and the remainder of the protein. For example, in amphipathic helices, a register shift may cause hydrophilic residues to exchange position with hydrophobic ones. Buried hydrophobic anchor residues in helices have been found to strongly resist translocation unless the substituting residue is also hydrophobic (Heinz et al., 1993). The implicated preservation of the hydrophobic interface agrees with general principles found from the analysis of the structural evolution of proteins (Lesk & Chothia, 1980). In E*r*-2, the deletion could clearly not be accommodated by a translocation, because the anchor residues of helix 2 are the 2 Cys residues that form disulfide bridges to the other 2 helices (Fig. 3). Each translocation by 1 position would rotate these Cys residues by about 100°, making the formation of the disulfide bridges impossible. The observed transition from an α-helix to a 3_{10}-helix or "$\alpha \rightarrow 3_{10}$

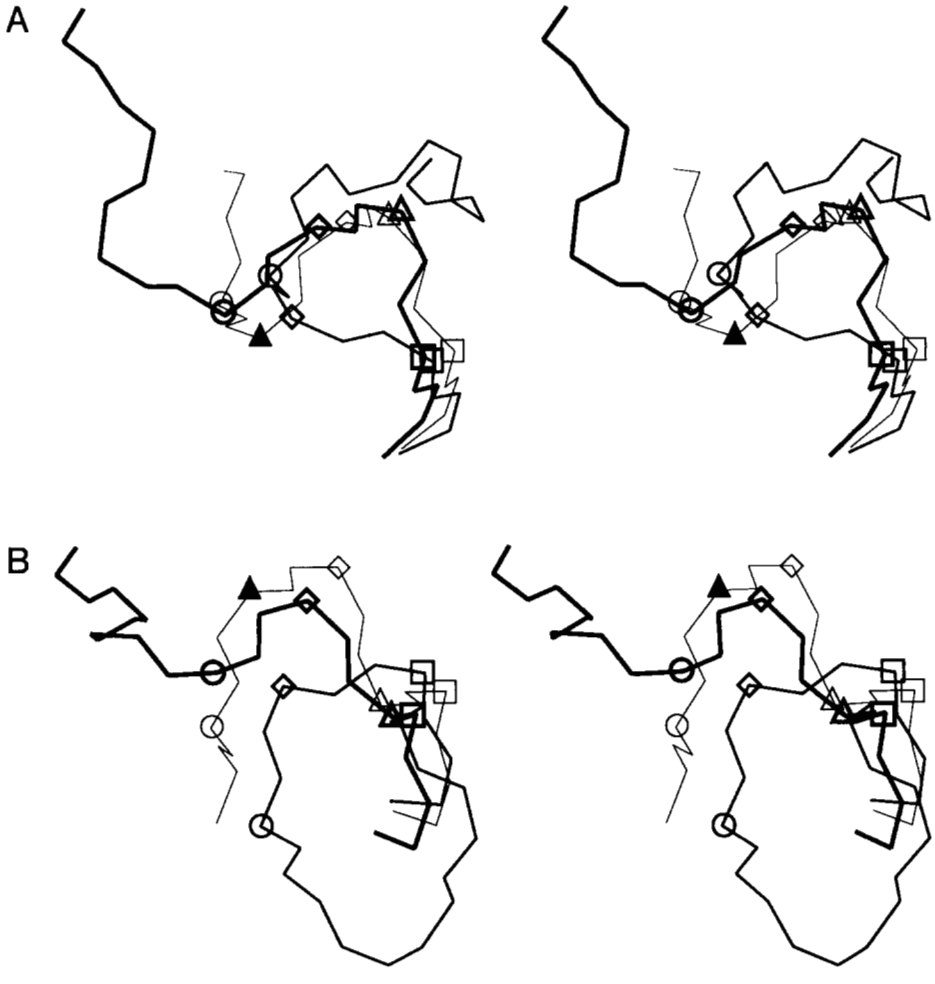

Fig. 6. Stereo views of the C-terminal region for the mean structures of E*r*-1 (medium line), E*r*-10 (thin line), and E*r*-2 (thick line). **A:** Front view. **B:** Top view. The same superposition as in Figure 2 has been used. The squares identify the C^{α} positions of Ser 34 (E*r*-1), Asn 33 (E*r*-10), and Asp 34 (E*r*-2); the open triangles the C^{α} positions of Pro 34 (E*r*-10) and Pro 35 (E*r*-2); the diamonds the C^{α} positions of Asn 35 (E*r*-1), Glu 35 (E*r*-10), and Glu 36 (E*r*-2); the filled triangle the C^{α} position of Leu 36 (E*r*-10); and the circles the C^{α} positions of Cys VI.

indentation" (Fig. 5) represents a localized interconversion between different helix types, which is a common response to insertions and deletions that leave anchor residues in place (Kavanaugh et al., 1993). The ease of such interconversions is quite readily rationalized by the fact that the backbone dihedral angles of 3_{10}-, α-, and π-helices are all located in the same low-energy region of the ϕ–ψ conformation space.

Detailed characterization of additional structural differences

So far in this paper the discussion has focused mainly on aspects that are of general interest with regard to protein structure and 3-dimensional protein architecture. In an earlier evaluation of the E*r*-10 structure alone (Brown et al., 1993), some additional structural features were identified that appeared to be of special relevance with regard to the implication from biochemical studies (Luporini & Miceli, 1986; Ortenzi et al., 1990) that the observed characteristics of the *E. raikovi* pheromone-receptor system might be explained in terms of homologous and heterologous association of domains with 3-dimensional structures similar to those determined for E*r*-10 and, now, for E*r*-1 and E*r*-2. In this section we compare these features in the 3 presently known pheromone structures.

An intriguing feature of the surface of E*r*-10 is observed when the molecule is rotated by 180° about a vertical axis relative to the orientation in Figure 2A, where there is a deep cleft at the bottom of the molecule (Fig. 7A). This cleft is capped at the top by the disulfide bridge 10–37, and the base of the cleft contains the side chains of Val 31 and of Cys 3, Cys 15, Cys 19, and Cys 27. The sides of the cleft are made up of helix 1 on the right side, and helix 3 plus the turn between helices 2 and 3 on the left side (see also Fig. 2B). At the bottom of the molecule, the walls of the cleft are charged and polar, with the α-amino group and the carboxyl groups of Asp 1 and Glu 4 on the right-hand wall, and the carboxyl groups of Glu 22 and Asp 23 as well as the side chain of Ser 20 on the left-hand wall. The upper portion of the cleft is less polar, with the side chains of Ala 7 and Leu 8 on the right-hand wall and the side chains of Met 30 and Leu 36 as well as the backbone atoms of the residues Gly 26 and Cys 27 on the left-hand wall.

A corresponding cleft with a similar distribution of amino acid types is observed in E*r*-1 (Fig. 7B). The lower part of the cleft is again made up of the side chains of Asp 1, Glu 4, Glu 23, and Asp 24. The tripeptide Glu 23–Asp 24–Arg 25 corresponds to Glu 22–Asp 23–Lys 24 in E*r*-10 (Fig. 3), and these tripeptide segments are mainly responsible for the high hydrophilicity in the polypeptide segment 20–30 of these 2 proteins (Raffioni et al., 1992). In the upper part of the right-hand wall of the cleft,

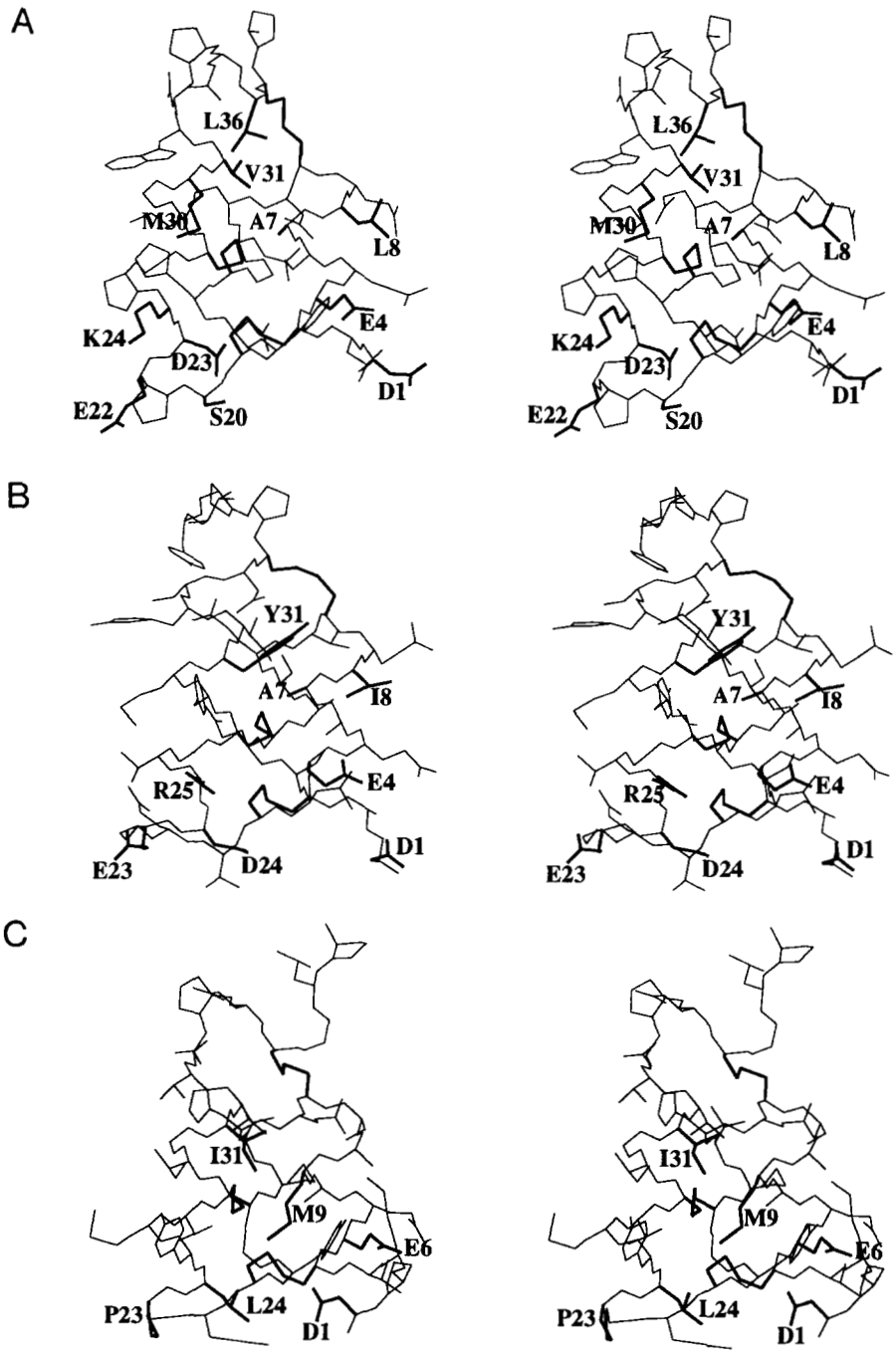

Fig. 7. Stereo views of the pheromones E*r*-10, E*r*-1, and E*r*-2 (**A**, **B**, and **C**, respectively) in an orientation obtained from that in Figure 2A by a 180° rotation about a vertical axis. This viewing angle enables detailed inspection of a cleft on the molecular surface that was previously identified in E*r*-10 (Brown et al., 1993). The 3 disulfide bonds and other side chains discussed in the text as essential components of this cleft are drawn with bold lines and, except for the Cys, these residues are identified with the 1-letter amino acid code and the sequence position. The polypeptide backbone and all other amino acid side chains are drawn with thin lines.

Leu 8 in E*r*-10 is replaced by Ile 8, and on the left-hand wall, Met 30 in E*r*-10 is replaced by Tyr 31. The side chain of Tyr 31 is located at the top of the cleft, where it protects Ala 7 completely from solvent contact (Fig. 4).

The surface cleft is also present in E*r*-2, but it shows distinctly different characteristics (Fig. 7C). Due to the insertion of the dipeptide segment Pro 2–Met 3 (Fig. 3), Asp 1 is shifted downward such that the lower part of the right-hand wall of the cleft contains Glu 6 as the only charged group. This supports the hypothesis that the conserved N-terminal Asp serves as site of recognition for the proteolytic processing of the propheromone (Miceli et al., 1991) and may not be of functional significance for receptor binding. The lower part of the left-hand wall of the cleft contains Pro 23 and Leu 24 in place of the acidic dipeptide Glu-Asp in E*r*-10 and E*r*-1, which contributes to the high hydrophobicity in the polypeptide segment 20–30 of E*r*-2 (Raffioni et al., 1992). In the upper part of the right-hand wall of the cleft, Met 9 occupies the position of Ala 7 in E*r*-10 and E*r*-1, and on the left-hand wall, Ile 31 corresponds to Met 30 in E*r*-10 and Tyr 31 in E*r*-1. The side chain of Met 9 lies entirely in the cleft and is partially protected from solvent contact (Fig. 4) by the side chain of Ile 31, which is located at the top of the cleft.

Another surface region identified in the E*r*-10 structure as being of potential functional significance is at the top of the molecule. This surface contains a number of hydrophobic side chains, i.e., Leu 8, Trp 32, Pro 34, Leu 36, and Pro 38, which

all have substantial surface accessibility (Fig. 4). The C-terminus is centrally located in this apolar surface, and functional importance is suggested by the observation that the allotypic substitution Pro 38 → Leu in Er-1 may result in changes of the ability to form mating pairs (Raffioni et al., 1992). The C-terminal backbone conformation in this top surface region is locally quite well defined in all 3 structures, which is probably due to the reduction of the accessible conformation space by the presence of 2 Pro residues and the disulfide bridge between Cys II and Cys VI. The structure of the C-termini contains further several hydrogen bonds (Fig. 3), which are also manifested by slowed backbone amide proton exchange. As was mentioned in the discussion of Figure 6, the C-terminal polypeptide segment following helix 3 is conformationally most variable among the 3 proteins, which can be traced to the different locations of Cys VI and the 2 Pro residues (Fig. 3). On the one hand, the C-terminus of Er-2, which encompasses 2 tight turns, is more isolated from the core of the protein than the C-termini of Er-10 and Er-1 (Fig. 2). Conversely, however, there are local conformational features that are most similar between Er-10 and Er-2, which both form type I tight turns immediately following helix 3, with the hydrogen bonds Asn 33 O′–Leu 36 NH and Asp 34 O′–Cys 37 NH, respectively (Fig. 3). Additionally, the side chain–backbone hydrogen bonds Asn 33 δO–Glu 35 NH in Er-10 and Asp 34 δO–Glu 36 NH in Er-2 are formed. In this part, where the C-termini fold back on top of the extended dipeptide segment between helices 1 and 2 to form the disulfide bridge, the structures of Er-10 and Er-2 follow each other particularly closely (Fig. 6). This first tight turn in the Er-2 structure is immediately followed by a type II tight turn with the hydrogen bond Glu 36 O′–Leu 39 NH, and the hydrogen bond Asn 33 O′–Cys 37 NH is formed in Er-10 (Fig. 3). It is evident already from the schematic presentation in Figure 3 that the combination of solvent-exposed hydrophobic side chains seen in Er-10 is only partially preserved in Er-1 and Er-2.

The global distribution of charged groups on the protein surface clearly distinguishes Er-1 and Er-10 from Er-2, which supports earlier corresponding implications from the primary structure (Raffioni et al., 1992). In addition to the chain-terminal charges, Er-1 and Er-10 contain 7 and 6 carboxylates, respectively. Five of these are in identical locations (Fig. 3), Glu 16 in Er-1 is replaced by His in Er-10, and only Glu 21 in Er-1 and Glu 35 in Er-10 do not have potentially charged counterparts in the other protein. In addition, in the 3-dimensional structure, Arg 25 in Er-1 corresponds to Lys 24 in Er-10. In the orientation of Figure 2A, nearly all charged side chains are clustered in the lower part of the molecules, thus causing a pronouncedly asymmetric charge distribution. In Er-2, there are 5 side chain carboxylates, of which only Asp 1 and Glu 6 coincide with the corresponding side chains in both other proteins, and His 14 corresponds to Glu 12 in Er-1 and Er-10. The charge distribution is much less asymmetric in Er-2. Most strikingly, the highly charged peptide segment at the start of helix 3 in Er-1 and Er-10 is replaced in Er-2 by a stretch of hydrophobic side chains.

Implications for the biological functions of the pheromones

In mature cells, binding of a heterologous pheromone is an initial step in the process of cellular conjugation (Luporini & Miceli, 1986; Luporini et al., 1992). The fact that homologous and heterologous pheromones compete for the same receptors with essentially identical affinities (Ortenzi et al., 1990) implicates a requirement for conserved structural features in the different pheromones, so that the same receptor could recognize them. In addition, however, because the binding of the homologous pheromone in mature cells does not lead to consummation of conjugation, there must also be differences between the homologous and heterologous pheromone–receptor complexes. This aspect of intercellular communication in *E. raikovi* was discussed recently by Brown et al. (1993) on the basis of the NMR structure of Er-10. The hypothetical assumptions made in this earlier paper are clearly borne out by the present work: the different pheromones have extensive similarities in their 3-dimensional structures, which also contain, however, distinct differences that may provide the structural basis for specific receptor recognition by the individual proteins, in particular in the C-terminal region beyond helix 3 (Fig. 6), the loop between the helices 2 and 3 (Fig. 2), and the N-terminus (Fig. 2) (see also Kinemage 2).

Acknowledgments

We thank Dr. M. Billeter for help with technical aspects of the structure comparisons, R. Marani for careful processing of the text, and Drs. R.A. Bradshaw and P. Luporini for helpful comments on the manuscript. Financial support was obtained from the Schweizerischer Nationalfonds (project 31.32033.91) and the European Community (bursary from the Science program to S.M.).

References

Billeter M, Qian YQ, Otting G, Müller M, Gehring W, Wüthrich K. 1990. Determination of the three-dimensional structure of the *Antennapedia* homeodomain from *Drosophila* in solution by ^{1}H nuclear magnetic resonance spectroscopy. *J Mol Biol 214*:183–197.

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

Grathwohl C, Wüthrich K. 1976a. The X-Pro peptide bond as an NMR probe for conformational studies of flexible linear peptides. *Biopolymers 15*: 2025–2041.

Grathwohl C, Wüthrich K. 1976b. NMR studies of the molecular conformations in the linear oligopeptides H-(L-Ala)$_n$-L-Pro-OH. *Biopolymers 15*:2043–2057.

Güntert P, Braun W, Wüthrich K. 1991. Efficient computation of three-dimensional protein structures in solution from nuclear magnetic resonance data using the program DIANA and the supporting programs CALIBA, HABAS, and GLOMSA. *J Mol Biol 217*:517–530.

Güntert P, Wüthrich K. 1991. Improved efficiency of protein structure calculations from NMR data using the program DIANA with redundant dihedral angle constraints. *J Biomol NMR 1*:447–456.

Harper ET, Rose GD. 1993. Helix stop signals in proteins and peptides: The capping box. *Biochemistry 32*:7605–7609.

Harrison CJ, Bohm AA, Nelson HCM. 1994. Crystal structure of the DNA binding domain of the heat shock transcription factor. *Science 263*: 224–227.

Heinz DW, Baase WA, Dahlquist FW, Matthews BW. 1993. How amino acid insertions are allowed in an α-helix of T4 lysozyme. *Nature 361*:561–564.

Kavanaugh JS, Moo-Penn WF, Arnone A. 1993. Accommodation of insertions in helices: The mutation in hemoglobin Catonsville (Pro37α-Glu-Thr38a) generates a $3_{10} \rightarrow \alpha$ bulge. *Biochemistry 32*:2509–2513.

Keefe LJ, Sondek J, Shortle D, Lattman EE. 1993. The α aneurism: A structural motif revealed in an insertion mutant of staphylococcal nuclease. *Proc Natl Acad Sci USA 90*:3275–3279.

Lesk AM, Chothia C. 1980. How different amino acid sequences determine similar protein structures; the structure and evolutionary dynamics of the globins. *J Mol Biol 136*:225–270.

Luporini P, Miceli C. 1986. Mating pheromones. In: Gall JG, ed. *The molecular biology of ciliated protozoa*. London: Academic Press. pp 263–299.

Luporini P, Miceli C, Ortenzi C, Vallesi A. 1992. Developmental analysis of the cell recognition mechanism in the ciliate *Euplotes raikovi*. *Dev Genet 13*:9–15.

Marti T, Otto H, Rösselet SJ, Heyn MP, Khorana HG. 1992. Consequences of amino acid insertions and/or deletions in transmembrane helix C of bacteriorhodopsin. *Proc Natl Acad Sci USA 89*:1219–1223.

Miceli C, La Terza A, Bradshaw RA, Luporini P. 1991. Structural characterization of mating pheromone precursors of the ciliate protozoan *Euplotes raikovi*. *Eur J Biochem 202*:759–764.

Mronga S, Luginbühl P, Brown LR, Ortenzi C, Luporini P, Bradshaw RA, Wüthrich K. 1994. The NMR solution structure of pheromone E*r*-1 from the ciliated protozoan *Euplotes raikovi*. *Protein Sci 3*:0000–0000.

Ortenzi C, Miceli C, Bradshaw RA, Luporini P. 1990. Identification and initial characterization of an autocrine pheromone receptor in the protozoan ciliate *Euplotes raikovi*. *J Cell Biol 111*:607–614.

Ottiger M, Szyperski T, Luginbühl P, Ortenzi C, Luporini P, Bradshaw RA, Wüthrich K. 1994. The NMR solution structure of the pheromone E*r*-2 from the ciliated protozoan *Euplotes raikovi*. *Protein Sci 3*:0000–0000.

Raffioni S, Miceli C, Vallesi A, Chowdhury SK, Chait BT, Luporini P, Bradshaw RA. 1992. Primary structure of *Euplotes raikovi* pheromones: Comparison of five sequences of pheromones from cells with variable mating interactions. *Proc Natl Acad Sci USA 89*:2071–2075.

Richardson JS. 1981. The anatomy and taxonomy of protein structure. *Adv Protein Chem 34*:167–339.

Richardson JS, Richardson DC. 1988. Amino acid preferences for specific locations at the ends of α helices. *Science 240*:1648–1652.

Richmond TJ. 1984. Solvent-accessible surface area and excluded volume in proteins. Analytical equations for overlapping spheres and implications for the hydrophobic effect. *J Mol Biol 178*:63–89.

Schiering N, Kabsch W, Moore MJ, Distefano MD, Walsh CT, Pai EF. 1991. Structure of the detoxification catalyst mercuric ion reductase from *Bacillus* sp. strain RC607. *Nature 352*:168–172.

Sondek J, Shortle D. 1990. Accommodation of single amino acid insertions by the native state of staphylococcal nuclease. *Proteins Struct Funct Genet 7*:299–305.

Sondek J, Shortle D. 1992. Structural and energetic differences between insertions and substitutions in staphylococcal nuclease. *Proteins Struct Funct Genet 13*:132–140.

Stewart AE, Raffioni S, Chaudhary T, Chait BT, Luporini P, Bradshaw RA. 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.

Weiner SJ, Kollman PA, Nguyen DT, Case DA. 1986. An all atom force field for simulations of proteins and nucleic acids. *J Comp Chem 7*:230–252.