The Crystal Structure of a Five-Stranded Coiled Coil in COMP: A Prototype Ion Channel?

Vladimir N. Malashkevich,* Richard A. Kammerer, Vladimir P. Efimov,† Therese Schulthess, Jürgen Engel

Oligomerization by the formation of α-helical bundles is common in many proteins. The crystal structure of a parallel pentameric coiled coil, constituting the oligomerization domain in the cartilage oligomeric matrix protein (COMP), was determined at 2.05 angstroms resolution. The same structure probably occurs in two other extracellular matrix proteins, thrombospondins 3 and 4. Complementary hydrophobic interactions and conserved disulfide bridges between the α helices result in a thermostable structure with unusual properties. The long hydrophobic axial pore is filled with water molecules but can also accommodate small apolar groups. An "ion trap" is formed inside the pore by a ring of conserved glutamines, which binds chloride and probably other monatomic anions. The oligomerization domain of COMP has marked similarities with proposed models of the pentameric transmembrane ion channels in phospholamban and the acetylcholine receptor.

Coiled-coil formation is a simple structural tool that facilitates oligomerization of biological macromolecules. Coiled-coil–forming amino acid sequences share a characteristic heptad repeat, $(a–g)_n$, with predominantly hydrophobic residues in the a and d positions and polar residues generally elsewhere. α-Helical bundles can be formed either from a single chain folding back on itself several times (in which case the constituent α helices may be parallel or antiparallel) or from a number of chains (typically two to four) in which the bundles normally contain only parallel α helices. To date, only two-, three-, and four-stranded coiled-coil structures have been studied with x-ray crystallography (*1*). Recent observations by electron microscopy of pentameric bouquet-like structures formed by the extracellular matrix proteins COMP (*2*) and thrombospondins 3 and 4 (TSP3 and TSP4) (*3*), as well as sequence analysis and circular dichroism studies of the oligomerization domain of COMP, suggest a pentameric coiled coil as the likely structure (*4*). The oligomerization domain in COMP is formed by the NH_2-termini provided by five identical subunits of 110 kD (*2*). The COOH-terminal parts form cell-binding domains, which interact with yet undefined receptors at chondrocytes. The multimeric state of COMP is assumed to be essential for recognition of clustered receptors, much as the first component of complement C1q specifically recognizes clusters of immunoglobulins (*5*).

We prepared peptides of different lengths from the oligomerization domain of rat COMP, and well-diffracting crystals were obtained from 46–amino acid residue fragments (*6*). Homologous sequences bearing six heptad repeats were also detected in TSP3 and TSP4 from different sources (Fig. 1). We report both the crystal structure of the oligomerization domain of COMP and its xenon derivative at 2.05 Å and 2.15 Å resolution, respectively (Table 1).

The pentameric bundle of COMP is built of 230 amino acid residues. Five chains, each with 46 residues, form a parallel coiled-coil structure with an average length of 73 Å and an average diameter of 30 Å (Fig. 2). The chains are predominantly folded into right-handed α helices, which are wrapped around each other in a left-handed superhelix with parameters as listed (Table 2). Four residues, including two cysteines involved in interchain disulfide bridges, are in a β-turn conformation, and five residues are in a random coil conformation. The observed α-helical content is close to that derived from circular dichroism spectra (*6*). The arrangement of the main chain and internal side chain atoms in the bundle obeys approximate fivefold symmetry. This is more precisely maintained within the six heptad repeats of the coiled coil than at the termini. Surface side chains also violate fivefold symmetry because of the pseudo-hexagonal arrangement of the molecules in the crystal in a plane lying perpendicular to the pentamer axis. The main chain conformation of the six heptad repeats is reasonably close to that predicted by Kajava (*7*), though the accuracy of prediction for the side chain positions and disulfide bridges was lower: the root-mean-square deviation between equivalent Cα atoms is 1.1 Å and that between all atoms is 3.1 Å.

The pentamer structure is stabilized by complementary hydrophobic interactions between neighboring helices. Because of the favorable relative orientation of the α helices, four types of "knob-into-hole" interactions (*1*) are formed: knobs formed by the side chains at positions a, d, e, or g pack into holes formed between the side chains at positions a′–g′, e′–d′, c′–d′, and a′–b′, respectively (Fig. 3; primes indicate residues from an adjacent chain). Only residues at positions a and d are predominantly hydrophobic, whereas those at positions b, c, e, and g are amphipathic. Nevertheless, the hydrophobic interface between helices is preserved through the involvement of the aliphatic parts of side chains with their polar ends exposed to the solvent. A large buried surface of 11,091 $Å^2$, or 47.7% of the total solvent-accessible area of the five isolated helices together, confirms the major role played by hydrophobic interactions in maintaining the stability of the pentamer. The side chains in the a layers are packed in a so-called perpendicular manner, whereas those in the d layers are packed in a parallel manner (*1*). Relatively few polar interactions between helices were detected in the structure. However, a prominent interaction was the circular chain of hydrogen bonds formed between the amide groups of Gln^{54} residues inside the axial hydrophobic

```
                                 abcdefg abcdefg abcdefg abcdefg abcdefg abcdefg
                                                          *
COMP  (Rat, 27-72)     GDL APQMLRE LQETNAA LQDVREL LRQQVKE ITFLKNT VMECDAC G
COMP  (Human)          SDL GPQMLRE LQETNAA LQDVRDW LRQQVRE ITFLKNT VMECDAC G
TSP3  (Mouse)          GEQ TKALVTQ LTLFNQI LVELRDD IRDQVKE MSLIRNT IMECQVC G
TSP3  (Human)          GEQ TKALVTQ LTLFNQI LVELRDD IRDQVKE MSLIRNT IMECQVC G
TSP4  (Human)          GDF NRQFLGQ MTQLNQL LGEVKDL LRQQVKE TSFLRNT IAECQAC G
TSP4  (Xenopus)        GDV SRQLIGQ ITQMNQM LGELRDV MRQQVKE TMFLRNT IAECQAC G

PLB   Human, (26-52)       QKLQ NLFINFC LILICLL LICIIVM LL
```

Fig. 1. Sequence alignment of the coiled-coil domains of COMP, TSP3, TSP4, and PLB. Heptad positions, indicated by lowercase letters, are assigned according to Efimov *et al.* (*4*) and the results of the current work. Conserved residues are in bold. The position of Gln^{54} is marked by an asterisk. In the expressed fragment of rat COMP, Gly^{27} is replaced by Met. Assignment of heptad repeats in PLB was done according to Simmerman *et al.* (*14*). Underlined residues in the PLB sequence are critical for pentamer formation. Abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

V. N. Malashkevich, Department of Structural Biology, Biozentrum, University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland.
R. A. Kammerer, V. P. Efimov, T. Schulthess, J. Engel, Department of Biophysical Chemistry, Biozentrum, University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland.

*To whom correspondence should be addressed.
†Present address: Department of Pharmacology, Robert Wood Johnson Medical School, University of Medicine and Dentistry of New Jersey, 675 Hoes Lane, Piscataway, NJ 08854–5635, USA.

channel. Other interchain interactions were mainly between the heptad position g and position e′ of the successive heptad. The amide group of Asn^{41} is hydrogen-bonded to the carboxyl group of $Glu^{36'}$ of a neighboring helix and also to the hydroxyl of $Thr^{40'}$. Given that Gln^{54} and Asn^{41} are also conserved in the homologous structures of TSP3 and TSP4, they might be responsible for specific formation of the pentamer. Indeed, the tendency to form a pentamer is strongly decreased in the $Asn^{41} \rightarrow Leu^{41}$ mutant of COMP, and the formation of a tetramer is favored (*8*). Salt bridges are observed between Asp^{46} and $Arg^{48'}$ (two of five possible c–e′ interactions) and between Glu^{57} and $Lys^{62'}$ (three of five possible g–e′ interactions) (*9*).

In most extracellular matrix proteins, the subunits are connected by disulfide bonds that stabilize the oligomeric structure. In the oligomerization domain of COMP, the ring of disulfide bonds is formed between Cys^{71} of one chain and $Cys^{68'}$ of the neighboring chain (*4*). Although oligomerization is independent of disulfide bridge formation, the latter increases the midpoint of the thermal transition (T_m) to >100°C (*6*). The increase is apparently attributable to the decreased entropy of the molecule. In the current structure, the α-helical conformation is disrupted after Glu^{67}, and residues Cys^{68} to Cys^{71} form a type III β turn (Fig. 4). Chain direction reversal brings Cys^{71} into a position favorable for disulfide bridge formation with $Cys^{68'}$ from a neighboring helix.

The axial pore of the pentameric bundle is the most unusual feature of the molecule (Fig. 5A). It is lined almost exclusively with aliphatic side chains with the exception of an internal ring of five glutamines at position 54. The diameter of the pore, as defined by the van der Waals radii, varies between 2 and 6 Å because of regular constrictions formed by the rings of internal side chains at positions a or d. There are 10 such constrictions, starting at Met^{33} and ending at Val^{65}. The opening is maximal at Val^{47} and Val^{65} and is minimal at Met^{33} and Gln^{54}. Thirteen water molecules are found scattered along the pore (*10*). Because of the lack of buried polar groups and the resulting weak water-protein interactions, these water molecules are clustered into groups of two or more. The individual temperature factors of the water molecules were typically higher than those of the nearest protein groups, indicating high mobility and probably only partial occupancy. Four water molecules are bound near the threonine side chains at position 40, which are partially buried but have their hydroxyl groups pointing to the helix interface. Because of its nonpolar nature, the pore is expected to have high affinity for small hydrophobic, rather than water molecules. Consequently, the latter could be replaced by xenon atoms in the corresponding heavy atom derivative, where eight xenon atoms were bound along the axis at the pore widenings located between adjacent a and d layers (Fig. 5B).

Table 1. Data collection, phasing, and refinement statistics. Crystals of rat COMP were grown as described (*6*). Peptides containing 65, 52, and 46 amino acid residues were subjected to crystallization. Well-diffracting crystals were obtained only with the shortest construct (residues 28 through 72 of rat COMP). They belong to space group $P2_1$ with a = 38.47 Å, b = 49.47 Å, c = 54.98 Å, and β = 103.84°. We collected diffraction data at 4°C on an imaging plate (MarResearch, Hamburg, Germany), using copper K_α radiation from a rotating-anode generator, a crystal-to-detector distance of 100 mm, and 2° oscillation. Raw data were processed with MOSFLM (*19*), scaled with SCALA (*20*) and AGROVATA (*20*), and converted to amplitudes with TRUNCATE (*20*). Statistics of the merged data are given. The search for heavy atom derivatives was not straightforward because of persistent deviations (>5%) between unit cell dimensions of native and potential derivative crystals. Considering the tight crystal packing (only 36% solvent), this is not surprising (*21*). The derivative data were scaled to the M66A data with SCALEIT (*20*). Initial heavy atom positions were obtained from difference Patterson maps and were brought to a common origin with cross-phased difference Fourier maps. Heavy atom parameters were refined with MLPHARE (*20*). The MIRAS phases (10.0 to 2.8 Å; mean figure of merit = 0.651) were greatly improved and extended to 2.4 Å, by use of DM (*22*) with solvent-flattening, histogram matching, and fivefold averaging. Electron density interpretation and model building were done with FRODO (*23*) and O (*24*). The structure was initially refined against M66A data by simulated annealing with X-PLOR (*25*). The refinement was continued with TNT (*26*) against native data that had higher resolution and were of better quality. The structure of the xenon derivative was refined similarly. Rigid-body refinement indicated that the COMP molecule in the M66A crystal was rotated by more than 5.5° and translated by 5 Å relative to its position in the native crystal. All residues fall in allowed regions of the Ramachandran plot except Asp^{28} in chain D. The present native model includes 230 amino acid residues, one chloride ion, and 169 water molecules. The xenon derivative structure includes 230 amino acids, one chloride ion, eight xenon atoms, and 94 water molecules. All color figures were produced with the program GRASP (*27*). TLA, trimethyllead acetate; PA, praseodymium acetate.

	Native	M66A	Xenon 10 bar for 2 days	TLA 10 mM for 3 days	PA 0.1 mM for 1 day
			Data collection statistics		
Resolution (Å)	2.05	2.4	2.15	2.2	2.8
Observed reflections	46638	30068	39796	24348	10460
Unique reflections	12378	7908	10282	9637	4981
Completeness (%)	95.1	98.0	97.0	90.8	96.2
R_{sym}*	0.04	0.07	0.07	0.05	0.14
R_{deriv}†		0.441	0.217	0.153	0.215
			Phasing statistics (10.0–2.8 Å)		
Heavy atom sites			8	3	2
R_{Cullis}‡§			0.71/0.63	0.71/0.65	0.87/0.92
$R_{Cullis\text{-}anomalous}$			0.80	0.89	0.97
Phasing power‡‖			1.50/1.26	1.56/1.30	0.91/0.76
			Refinement statistics		
R factor (%)¶	17.6		21.2		
Mean B factor (Å^2)	31.7		46.2		
Bond distance (Å)#	0.012		0.015		
Bond angle (°)#	1.43		1.68		
Planar group (Å)#	0.008		0.010		

*$R_{sym} = \Sigma|I - \langle I\rangle|/\Sigma I$. †$R_{deriv} = \Sigma||F_{PH}| - |F_P||/\Sigma|F_P|$. ‡Values are for acentric/centric reflections. §$R_{Cullis} = \Sigma||F_{PH}| - |F_P + F_H||/\Sigma||F_{PH}| - |F_P||$. ‖Phasing power = $[\Sigma|F_H|^2/\Sigma(|F_{PH}| - |F_P|)^2]^{1/2}$. ¶$R$ factor = $\Sigma||F_{obs}| - |F_{calc}||/\Sigma|F_{obs}|$. #Root-mean-square error.

Table 2. Superhelical and helical parameters (*1*) were obtained by fitting the Cα coordinates of the ideal model calculated according to Crick (*28*) to the experimental structure. Number of residues per helix turn was derived with the formula $2\pi/[\omega_1 - \omega_0 \cdot \cos(\chi/2)]$, where χ is the superhelix crossing angle (*1*) and ω_0 and ω_1 are fitting parameters (*28*).

Item	Pentamer	Tetramer (*1*)
	Superhelix parameters	
Supercoil radius, R_0 (Å)	8.6 ± 0.2	7.6
Residues per turn, $2\pi/\omega_0$	140 ± 1	139
Supercoil pitch, P (Å)	204 ± 1	205
Orientation of position a, φ (°)	19.5 ± 0.5	19.8
	α *Helix parameters*	
Residues per turn	3.58 ± 0.02	3.59
Rise per residue, d (Å)	1.52 ± 0.03	1.52
α Helix radius, R_l (Å)	2.20 ± 0.02	2.26
Helix-crossing angle, Ω (°)	18.5 ± 0.5	18.3
Interhelix distance, D (Å)	10.2 ± 0.4	10.6

One of nine widenings serves as an "ion trap" for a chloride ion in both the native- and the xenon-derivative structures. Although positions a and d are usually occupied by hydrophobic residues in canonical coiled-coil–forming sequences, some important exceptions were observed. An asparagine residue is conserved at position a in the sequence of the GCN4 leucine zipper (*1*). Glutamine at position 54 is also conserved in pentamer-forming sequences (Fig. 1) (*3*); however, possible reasons for its conservation were unclear. We found that in COMP, the side chains of Gln^{54} formed an internal ring interlocked by hydrogen bonds between $O^{\varepsilon 1}$ and $N^{\varepsilon 2}$ of neighboring residues (Fig. 6). The planes of the five amide groups deviated slightly from the plane of the ring in such a way that the partial positive charges on the nitrogens and the partial negative charges on the oxygens together create a dipole oriented parallel to the dipole moment of the α helices. This unusual structure acts as an ion trap for chloride, which is caught at the positive pole of the trap. The ion was identified on the basis of an average distance of 3.27 ± 0.16 Å between the ion and the interacting nitrogen atoms (in

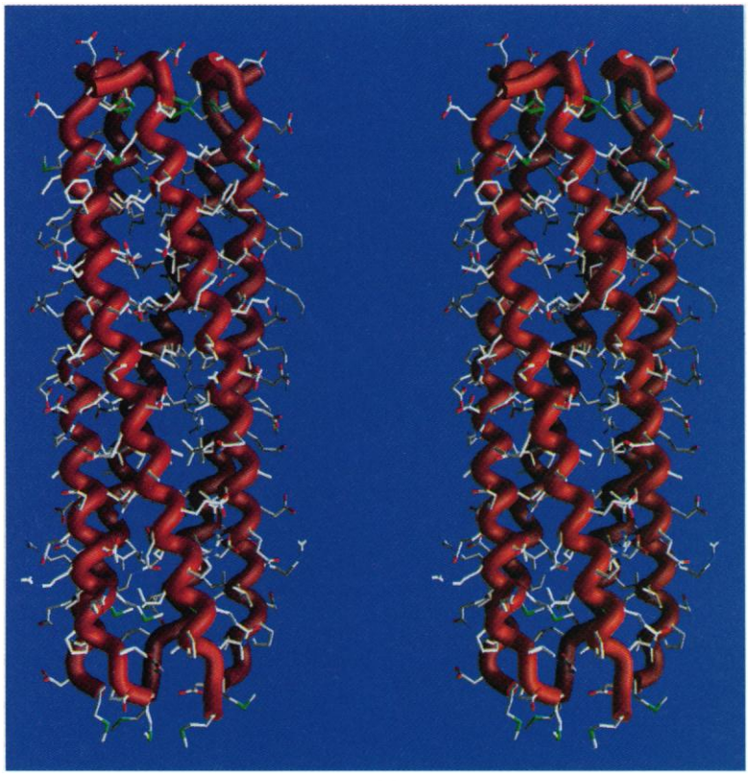


Fig. 2. Stereoview of the oligomerization domain of COMP seen perpendicular to the pentamer axis with its NH_2-terminus down. Ribbon representation is used for the main chain; side chains are colored according to atom type. Each helix makes about one-third of a superhelical turn over the length of the bundle. Deviations from fivefold symmetry of the positions of side chains can be seen.

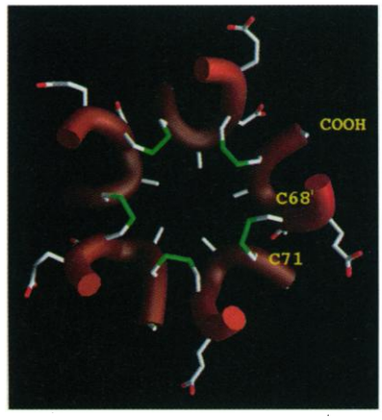
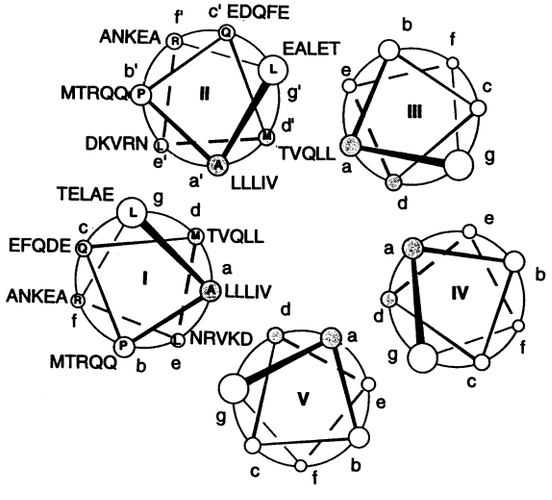


Fig. 4. Structure of the interchain disulfide bridges seen from the NH_2-terminus of the molecule. The tetrapeptide Cys^{68} through Cys^{71} has a type III β-turn conformation, which allows disulfide bridge formation between Cys^{71} and $Cys^{68'}$ of the neighboring chain. The disulfide bridges have stereochemically optimal conformations (ϕCys^{68} = −83 ± 7°, ψCys^{68} = 107 ± 7°; ϕCys^{71} = −96 ± 20°, ψCys^{71} = −59 ± 57°; S-S dihedral angle is −85 ± 5°). The COOH-terminal Gly^{72} residue is not seen in the map (*29*).

Fig. 3. Schematic "helical wheel" representation of the oligomerization domain of COMP as seen from the NH_2-terminus. Heptad positions are labeled as in Fig. 1; a and d positions are highlighted in gray.



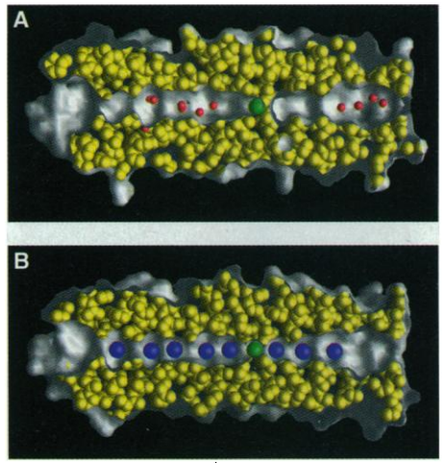


Fig. 5. Cross section of the axial pore of COMP. The protein part is represented by yellow van der Waals spheres. The corresponding molecular surface is shown in white. Constrictions formed by the a and d layers and widenings between them can be seen. (**A**) The native structure, with 13 water molecules bound in the pore (small red spheres; not all can be seen). The green sphere is a Cl^- ion caught at the positive pole of the "ion trap." (**B**) The xenon complex. Chloride remains in the trap, and eight xenon atoms are also bound between the a and d layers. The COOH-terminus is on the right.

the case of metal binding, the distance between metal and oxygen would be much shorter) and an individual *B* factor of 25.7 Å^2, which correlates well with the average value of 17.8 Å^2 for the interacting nitrogen atoms (*11*). Similar chloride binding by the ring of asparagines was detected in the structure of the trimeric parallel α-helical bundle from the transmembrane subunit (Mo-55) of the Moloney murine leukemia virus (*12*). Chloride and bromide stabilized the Mo-55 structure, but fluoride did not.

Despite differences in biological function and lack of obvious sequence identity, a strong similarity was found between the structure of the oligomerization domain of COMP and the structural models of pentameric transmembrane ion channels in phospholamban (PLB) and acetylcholine receptor (AChR). A coiled-coil structure with a completely hydrophobic internal pore was proposed for the Ca^{2+}-specific ion channel in PLB on the basis of systematic mutational analysis (*13*, *14*) and structural modeling (*15*). Sequence alignment methods did not reveal any significant homology between the corresponding sequences of COMP and PLB. However, the heptad assignment of Simmermann *et al.* (*14*) made possible the feasible alignment of the COMP and PLB sequences (Fig. 1). In contrast, the energy-minimized model of PLB (*15*) lacks heptad assignment. As shown by structural superposition, it has superhelix parameters and interchain distances similar to ours, but each helix in the PLB model is rotated clockwise by about 45° (ϕ = 65°, Table 2). This rotation is unusual, taking into account that the value of ϕ is well conserved in the coiled-coil structures from dimer to pentamer studied so far (Table 2) (*1*). We conclude that the crystal structure of the COMP channel represents the structure of the PLB channel in great detail and thus can be used as a structural prototype for pentameric channels formed from parallel α helices. A related cation-selective pore structure was also derived on the basis of mutational analysis (*16*) and electron microscopy of AChR at 9 Å resolution (*17*). This channel is made up of α helices contributed from five different but homologous subunits. The channel contains three polar residues in the a and d positions, in addition to a hydrophobic region with a leucine in position a, possibly occluding the pore in the closed, arresting state. The presence of water in the hydrophobic channel of COMP suggests that ions may penetrate the channels in PLB and AChR in hydrated form. The hydrodynamic radius of a completely hydrated Ca^{2+} ion is 2.7 Å compared with an ion radius of 1 Å for Ca^{2+} in a nonhydrated form (*18*).

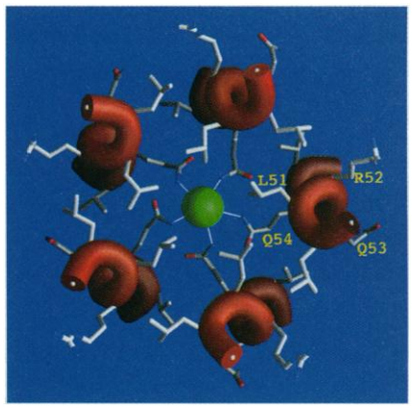


Fig. 6. View of the "ion trap" along the pentamer axis from the NH_2-terminus. The dipole is formed because of the systematic deviation of the amide groups of Gln^{54} residues from the plane of the ring, with its moment parallel to the dipole moment of the pentameric bundle. Chloride is bound, at the positive pole of the trap, between the layers formed by the side chains of Leu^{51} and Gln^{54}.

Previously, no channel function has been demonstrated for COMP. Direct integration of COMP into lipid membranes seems unlikely because of the highly hydrophilic outer surface that distinguishes it from PLB and other true channel domains. The question of whether the oligomerization domain of COMP has diverged from some pentameric ion channel at an early stage of evolution yet retaining some rudimentary properties of it, or whether similar structures have evolved independently for different functional purposes, remains open. Nevertheless, the unusual features observed in the three-dimensional structure of COMP exhibit a way of engineering a new class of stable molecules with unique structural and binding properties.

REFERENCES AND NOTES

1. P. B. Harbury, T. Zhang, P. S. Kim, T. Alber, *Science* **262**, 1401 (1993); P. B. Harbury, P. S. Kim, T. Alber, *Nature* **371**, 80 (1994).
2. M. Mörgelin, D. Heinegård, J. Engel, M. Paulsson, *J. Biol. Chem.* **267**, 6137 (1992); P. E. DiCesare, M. Mörgelin, K. Mann, M. Paulsson, *Eur. J. Biochem.* **223**, 927 (1994); G. Newton *et al.*, *Genomics* **24**, 435 (1994).
3. A. Qabar, L. Derick, J. Lawler, V. Dixit, *J. Biol. Chem.* **270**, 12725 (1995); N. Hauser, M. Paulsson, A. A. Kale, P. E. DiCesare, *FEBS Lett.* **368**, 307 (1995); J. Lawler, K. McHenry, M. Duquette, L. Derick, *J. Biol. Chem.* **270**, 2809 (1995).
4. V. P. Efimov, A. Lustig, J. Engel, *FEBS Lett.* **341**, 54 (1994).
5. J. Tschopp, T. Schultess, J. Engel, J.-C. Jaton, *ibid.* **112**, 152 (1980).
6. V. P. Efimov, J. Engel, V. N. Malashkevich, *Proteins* **24**, 259 (1996).
7. A. V. Kajava, *ibid.*, p. 218.
8. R. A. Kammerer, unpublished results.
9. Charged side chains in the crystal structure do not show a strong tendency to form interhelical ion pairs, even though most side chains are not involved in direct crystal contact. Weak pH-dependence of T_m under reducing conditions (at neutral pH, T_m = 55° ± 1°C, whereas at acidic and alkaline pH, T_m = 45° ± 1°C) indicates that although ionic interactions do play a moderate role in stabilizing the structure of the pentameric bundle, they may be especially important in guiding the correct assembly of the helices.
10. About 10% less surface area is buried in the COMP pentamer than in the tetrameric coiled coil (48.2 Å^2 compared with 51.9 Å^2 per residue). This results from the larger axial pore diameter in COMP, which makes residues at positions a and d more accessible to water molecules, whereas in the tetramer (*1*), water is excluded by the small channel radius of 1.0 to 1.3 Å. Water in the narrow pentameric channel cannot form clathrate cages around hydrophobic groups, as it does in bulk solvent (*18*). The resulting entropy gain may explain why a pore of the size found in the COMP pentamer does not destabilize the structure. Larger hydrophobic pores would be entropy-opposed, making hexameric or larger coiled-coil structures less probable.
11. Test refinement rules out fluoride or bromide as giving nonsensical *B* factor values of 5.7 and 59.2 Å^2, respectively. In addition, Cl^- is the most likely anion, because crystallization was performed in the presence of 0.5 M NaCl.
12. D. Fass, S. C. Harrison, P. S. Kim, *Nature Struct. Biol.* **3**, 465 (1996).
13. I. T. Arkin *et al.*, *J. Mol. Biol.* **248**, 824 (1995); I. T. Arkin *et al.*, *EMBO J.* **13**, 4757 (1994).
14. H. K. Simmerman, Y. M. Kobayashi, J. M. Autry, L. R. Jones, *J. Biol. Chem.* **271**, 5941 (1996).
15. P. D. Adams, I. T. Arkin, D. M. Engelman, A. T. Brünger, *Nature Struct. Biol.* **2**, 154 (1995).
16. H. A. Lester, *Annu. Rev. Biophys. Biomol. Struct.* **21**, 267 (1992); C. Labarca *et al.*, *Nature* **376**, 514 (1995).
17. N. Unwin, *Nature* **373**, 37 (1995).
18. P. W. Atkins, *Physical Chemistry* (Oxford Univ. Press, New York, ed. 5, 1994).
19. A. G. V. Leslie, *MOSFLM Users Guide* (Medical Research Council, Laboratory of Molecular Biology, Cambridge, 1994).
20. Collaborative Computing Project Number 4, *Acta Crystallogr. D* **50**, 760 (1994).
21. To overcome the problem of nonisomorphism, we expressed selenomethionine-substituted COMP. Because the peptide contains three methionines per chain (15 per asymmetric unit), we also expressed a Met^{66} → Ala^{66} (M66A) mutant to decrease the number of heavy atoms in the asymmetric unit and to obtain an additional derivative. Several diffraction data sets were collected in the laboratory, as described above, and others were collected on the BW7B beamline at the European Molecular Biology Laboratory outstation, Hamburg, Germany (λ = 0.862 Å) and on the Swiss-Norwegian beamline at the European Synchrotron Radiation Facility, Grenoble, France (λ = 0.875Å). Although an amino acid analysis indicated that selenomethionine was successfully incorporated in both wild-type and M66A COMP, we were not able to detect any signal from the Se atoms, using either isomorphous difference or anomalous Patterson techniques. We continued the search for heavy atom derivatives, using a traditional approach with selenomethionine M66A COMP as the native protein, because it demonstrated a higher stability of unit cell dimensions upon soaking (note the significant nonisomorphism between the native and M66A data). This search produced two useful derivatives. Alternatively, we attempted to make use of a putative axial hydrophobic cavity in the model pentameric coiled coil by preparing the xenon derivative. After subjecting native crystals to a xenon pressure between 2 and 5 bar, only weak binding was detected. However, with an improved system, we were able to apply pressure of 10 bar to the M66A mutant crystal in a standard 0.5-mm glass capillary, which resulted in a useful derivative.
22. K. Cowtan, *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography* **31**, 34 (1994).
23. T. A. Jones, *Methods Enzymol.* **115**, 157 (1985).
24. _____, J.-Y. Zou, S. W. Cowan, M. Kjeldgaard, *Acta Crystallogr. A* **47**, 110 (1991).
25. A. T. Brünger, *X-PLOR Version 3.1. A System for Crystallography and NMR* (Yale Univ. Press, New Haven, CT, 1992).
26. D. E. Tronrud, L. F. Ten Eyck, B. W. Matthews, *Acta Crystallogr. A* **43**, 489 (1987).
27. P. J. Kraulis, *J. Appl. Crystallogr.* **24**, 946 (1991).

28. F. H. C. Crick, *Acta Crystallogr.* **6**, 685 (1953).
29. As detected by SDS–polyacrylamide gel electrophoresis of freshly prepared COMP and of dissolved crystals, the concentration of reduced disulfides varied from 10 to 20%, depending on the preparation, and increased with time in the crystal. As a result, fragments of the polypeptide chain adjacent to disulfides were initially poorly defined, especially in the structure of the M66A mutant. In the wild-type structure, disulfide bridges were better defined because of the higher degree of oxidation and the stabilizing effect of Met^{66} side chains, which decreased the local protein mobility through higher packing density in the crystal. The initial electron density was significantly improved by a simple skeletonization-like procedure. Residues Glu^{67} to Gly^{72} were excluded from the phase calculations, and the density calculated with coefficients $(2F_{obs} - F_{calc})$, α_{calc} for this part of the structure was filled with dummy atoms. After three iterative cycles, the quality of the map was sufficient to model all of the remaining residues, except Gly^{72}.
30. Supported in part by Swiss National Science Foundation grant 31-32251.91 to J.E. We thank J. N. Jansonius for support and for providing access to x-ray equipment, and also for valuable comments on the manuscript. Coordinates have been deposited in the Brookhaven Protein Bank (entry code, 1VDF).

20 June 1996; accepted 19 September 1996

Nested Retrotransposons in the Intergenic Regions of the Maize Genome

Phillip SanMiguel, Alexander Tikhonov, Young-Kwan Jin, Natasha Motchoulskaia, Dmitrii Zakharov, Admasu Melake-Berhan, Patricia S. Springer,* Keith J. Edwards, Michael Lee,† Zoya Avramova, Jeffrey L. Bennetzen‡

The relative organization of genes and repetitive DNAs in complex eukaryotic genomes is not well understood. Diagnostic sequencing indicated that a 280-kilobase region containing the maize *Adh1*-F and u22 genes is composed primarily of retrotransposons inserted within each other. Ten retroelement families were discovered, with reiteration frequencies ranging from 10 to 30,000 copies per haploid genome. These retrotransposons accounted for more than 60 percent of the *Adh1*-F region and at least 50 percent of the nuclear DNA of maize. These elements were largely intact and are dispersed throughout the gene-containing regions of the maize genome.

Plant genomes, like those of other higher eukaryotes, consist of repetitive DNA sequences intermixed with genes (*1–4*), but neither the nature of these repetitive sequences nor the basis of their interspersion has been well defined. In general, larger eukaryotic genomes have higher percentages of repetitive DNAs. The maize nuclear genome, for instance, contains about 60 to 80% repetitive DNA (*1*).

In a contiguous 280-kb region flanking the maize *Adh1*-F gene, isolated on a yeast artificial chromosome (YAC) (*4*), 37 classes of repeats accounted for >60% of the DNA. Blocks containing different mixtures of these repeats make up most of the maize genome, range in size from 20 to 200 kb, and are hypermethylated in mature plant tissues (*3*). Fragment order, size, and composition in this region was further resolved (*4, 5*) (Fig. 1). Groups of repeats were observed more than once within the region. For instance, the L repeat was often found on nearby fragments (12 and 15, 41 and 45, 57 and 61, and 113 and 119; Fig. 1). Repeats M, N, O, and P were only found between two L repeats. Similarly, the Q, R, and S repeats were between pairs of I/J repeats, whereas repeat C was found between H repeats (Fig. 1). These patterns suggested that repeat pairs may be long terminal repeats (LTRs) and that the sequences between them may encode retroelement products (*6*).

We undertook diagnostic DNA sequencing of the presumptive LTRs of each of these proposed retrotransposons, and completely sequenced one element. In every case, the model was supported by the data. The complete element sequenced, which we named *Opie*-2, is 8987 base pairs (bp) with a 1271-bp 5′ LTR and a 1292-bp 3′ LTR. The LTRs correspond to the L repeat (Fig. 2).

Further sequencing uncovered three other families of LTR-retrotransposons: *Huck* (containing LTRs with the H repeat), *Kake* (containing the internal KK repeat), and *Fourf* (containing LTRs with the FF repeat) (Fig. 3). Three other novel retrotransposons were found by chance: *Milt*, *Reina*, and *Victim* (Fig. 3). We also identified three highly repetitive retroelements similar to those found in previous studies. *Ji* (with J/I LTRs) is related to retrotransposon PREM-2 (*7*); *Grande*-zm1 is homologous to *Grande* from *Zea diploperennis* (GenBank accession number X82087); and *Cinful* is homologous both to the defective retrotransposon *Zeon*-1 (*8*) and the solo

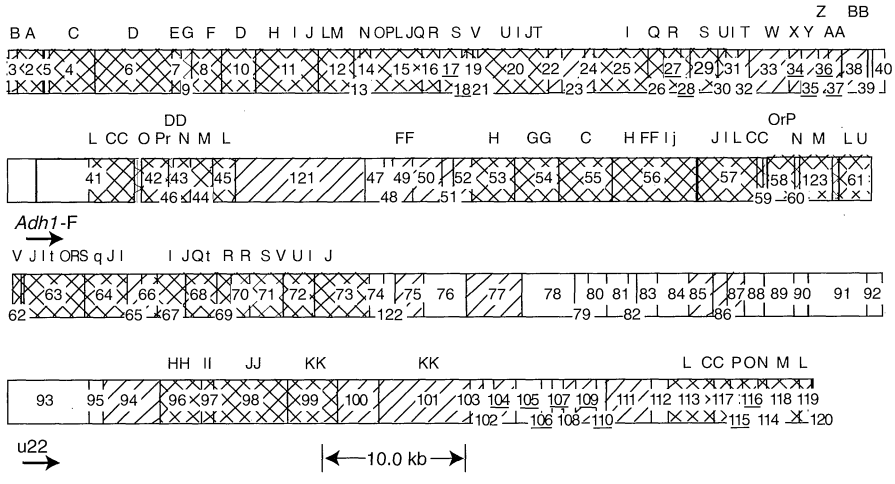


Fig. 1. Composition and arrangement of repetitive DNA flanking the maize *Adh1*-F locus. Bar patterns correspond to approximate copy numbers: cross-hatched bars, highly repetitive (thousands of copies); hatched bars, repetitive (hundreds of copies); and open bars, low copy number (<100 copies) (*4*). Letters above bars indicate the class of repetitive DNA, as defined by hybridization of each highly repetitive fragment to all fragments in the region (*4*). The arrows indicate two genes in this region, *Adh1*-F and u22 (*12*), and their direction of transcription. The order of adjacent underlined fragments is not known.

P. SanMiguel, A. Tikhonov, Y.-K. Jin, N. Motchoulskaia, D. Zakharov, A. Melake-Berhan, P. S. Springer, M. Lee, Z. Avramova, J. L. Bennetzen, Department of Biological Sciences, Purdue University, West Lafayette, IN 47907, USA.
K. J. Edwards, Long Ashton Research Station, Long Ashton, Bristol BS18 9AF, UK.

*Present address: Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA.
†Present address: Department of Agronomy, Iowa State University, Ames, IA 50011, USA.
‡To whom correspondence should be addressed.