The Crystal Structure of Lysin, a Fertilization Protein

A. Shaw, D. E. McRee, V. D. Vacquier, C. D. Stout*

Lysin, a protein from abalone sperm, creates a hole in the envelope of the egg, permitting the sperm to pass through the envelope and fuse with the egg. The structure of lysin, refined at 1.9 angstroms resolution, reveals an α-helical, amphipathic molecule. The surface of the protein exhibits three features: two tracks of basic residues that span the length of the molecule, a solvent-exposed cluster of aromatic and aliphatic amino acids, and an extended amino-terminal hypervariable domain that is species-specific. The structure suggests possible mechanisms of action.

Abalone are marine mollusks (genus *Haliotis*) with eight species inhabiting the west coast of North America. The species have overlapping breeding seasons and habitats yet maintain themselves as distinct. Adults spawn gametes into seawater where fertilization occurs. Abalone eggs are surrounded by a rigid, elevated vitelline envelope (VE) 0.6 μm thick, composed of eight glycoprotein subunits of 32 to 44 kD. The glycoproteins form a tight fibrous network of filaments, each 13 nm in diameter (*1*). Spermatozoa must penetrate the VE in order to reach the egg plasma membrane. The sperm acrosome releases a 16-kD protein, lysin, onto the VE surface. Lysin disrupts the VE by a stoichiometric, nonenzymatic mechanism, creating a hole 3 μm in diameter through which the sperm pass (*1*). Lysin dissolution of the VE is species-selective in vitro and may be responsible for species-specific gamete recognition in nature (*2*).

A. Shaw, D. E. McRee, C. D. Stout, Department of Molecular Biology, The Scripps Research Institute, La Jolla, CA 92037–1093.
V. D. Vacquier, Scripps Institution of Oceanography, Center for Marine Biomedicine and Biotechnology, University of California at San Diego, La Jolla, CA 92093–0202.
*To whom correspondence should be addressed.

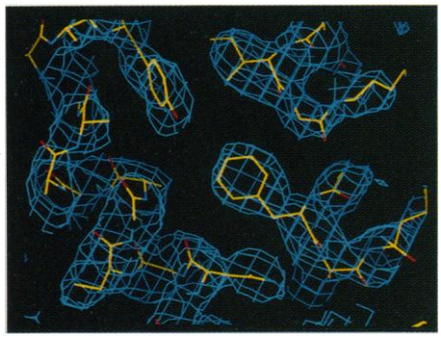

Fig. 1. A representative portion of the multiple isomorphous replacement (MIR) electron density map at 3.0 Å resolution, contoured at 1σ with the model for the refined structure superposed. Oxygen atoms are red and nitrogen atoms are blue.

As a first step in understanding the mechanism by which egg envelopes are dissolved, the crystal structure of lysin has been determined.

Abalone sperm lysin is an α-helical protein with three striking surface features (Table 1 and Fig. 1): (i) Two parallel tracks of positively charged residues line one face of the molecule, (ii) a cluster of aromatic and aliphatic residues is exposed to solvent, and (iii) the species-specific hypervariable domain of the NH_2-terminus extends away from the helical bundle but is ordered because of crystal packing contacts. The overall dimensions are 50 Å by 50 Å by 35 Å. The structure is 65% α helical and has no β sheet (Fig. 2). There are five α helices. Helices α1 to α4 form the core of the molecule and have extensive interactions along their internal faces. Helix α1 consists of seven turns and is essentially straight. Helix α2 has eight turns and bends 125° at His^{61}. Hydrogen bond analysis indicates that helix α3 and helix α4 are separate with a three-residue (96 to 98) link between them; the angle between their axes is 135°. The internal face of α5, with just two turns, is in contact with the NH_2-terminal segment of α2. The kink in α2 and the bend between α3 and α4 maximize the contacts among the first four helices to create a helical bundle.

Red abalone sperm lysin with 12 arginine and 13 lysine residues has a net charge at neutral pH of +13 (discounting histidines). Figure 2 maps the distribution of these residues in red abalone lysin (Arg^1 and Lys^{136} are not observed in the crystal). These residues are arranged in two parallel tracks running the length of the molecule. From top to bottom (Fig. 2), one track comprises nine residues from the COOH-terminal ends of helices α1 and α2, all of helix α3, and the loop connecting α2 and α3. The other track

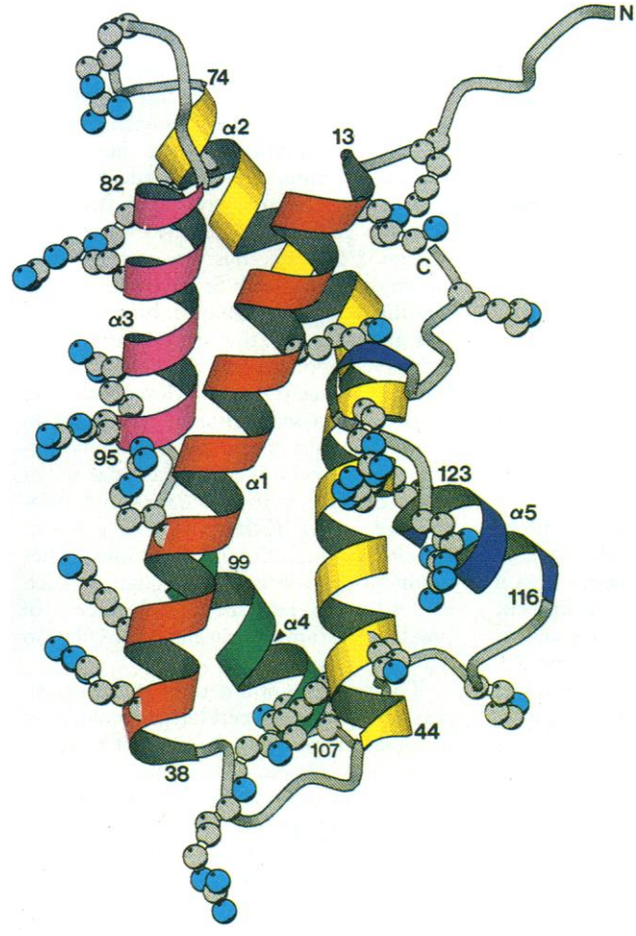

Fig. 2. The structure of red abalone sperm lysin and the side chains of 23 arginine and lysine residues in the two basic tracks. One track comprises residues 78, 71, 72, 87, 94, 95, 29, 33, and 36 (left side, top to bottom); the other comprises residues 9, 13, 132, 20, 125, 55, 56, 123, 48, 113, 106, 108, 47, and 40 (right side, top to bottom). The residue numbers defining the five α helices are indicated. Side chain carbon atoms are gray, nitrogen atoms cyan (figure created with Molscript).

comprises 14 residues from the NH_2-terminal ends of helices α1 and α2, the COOH-terminal ends of helices α4 and α5, and nonhelical segments adjacent to these helices. Measured between α carbons, each track is 45 Å in length. When viewed from the side, the face of the molecule containing the two basic tracks is essentially flat.

The hydrophobic patch is formed by the crossing over of the NH_2- and COOH-terminal segments of helix α2 with helices α4 and α3 (Fig. 3A). It is centered where these three helices bend around each other at residues 61 and 96 to 98 (Fig. 2). The combined solvent-exposed surface area of the 11 side chains is 1077 $Å^2$, which is 10% of the total surface area of the protein (*3*). Although tyrosine and tryptophan occur at protein surfaces, it is unusual for leucine, isoleucine, and phenylalanine to be solvent-exposed. The large number of charged residues in lysin, as well as the polar and hydrophobic interactions, must provide sufficient stability to compensate for exposure of aliphatic and aromatic residues.

The NH_2-terminus extends out from the protein core and is free to interact with other molecules. For red abalone lysin, the first 12 residues are Arg, Ser, Trp, His, Tyr, Val, Glu, Pro, Lys, Phe, Leu, and Asn (*2*). Although the first three residues are disordered, the remaining amino acids are ordered by contacts with a symmetry-related molecule in the crystal lattice (Fig. 3B). As a consequence, hydrophobic residues of the NH_2-terminus and the hydrophobic patch are occluded from solvent. In solution, the conformation of the NH_2-terminal residues may be different. For instance, residues 1 through 12 could fold into an α helix contiguous with helix α1; residues 3, 6, and 10 are hydrophobic or aromatic, and residues 1, 4, and 7 are hydrophilic and charged, typical of sequences that form amphipathic α helices (*4*). In such a conformation, the NH_2-terminus would still extend away from the core of the protein.

The presence of an exposed hydrophobic surface is a distinctive feature of the structure (Fig. 3A) that accounts for the physical properties of the protein. Lysin is absorbed to paraffin and organosilane-treated surfaces (*1*); it causes entrapped carboxyfluorescein to be released from neutral liposomes (*5*); it fuses negatively

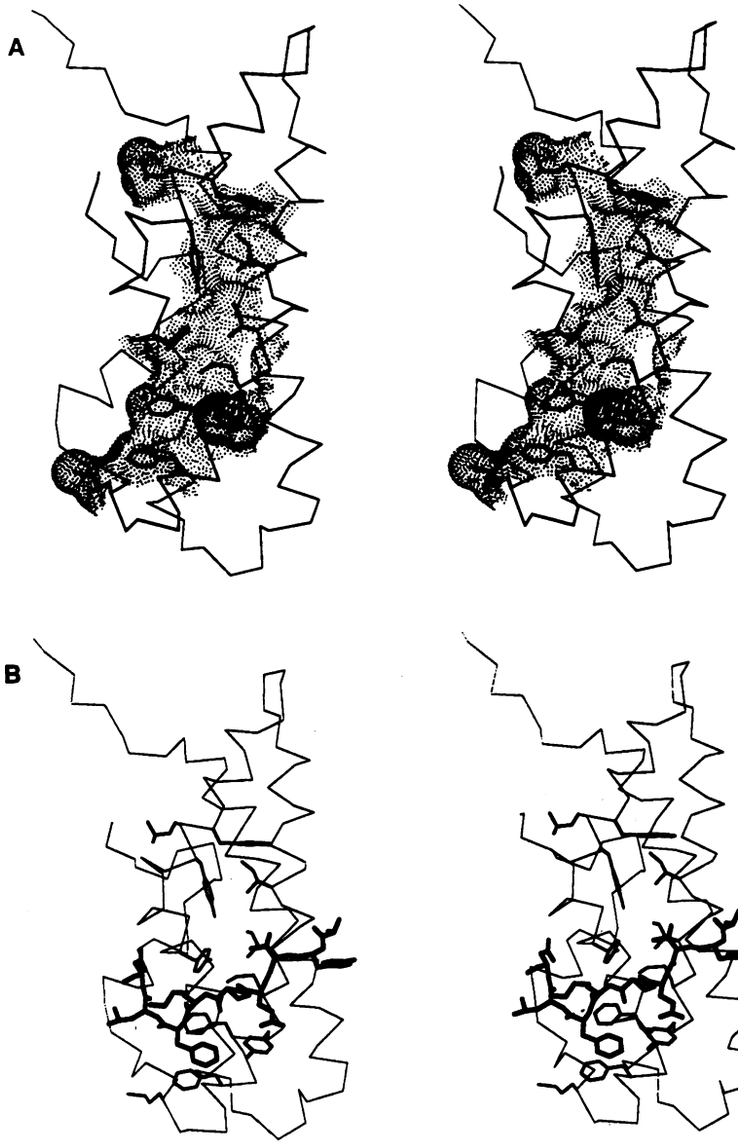

Fig. 3. (**A**) Stereo figure of the solvent-accessible surface (*3*) of 11 aliphatic and aromatic residues. From top to bottom, the side chains shown are for Leu^{67}, Trp^{68}, Tyr^{65}, Ile^{92}, Ile^{96}, Tyr^{57}, Met^{98}, Tyr^{100}, Phe^{101}, Phe^{104}, and Met^{110}. (**B**) Stereo figure of the interaction of the NH_2-terminus (heavy lines) of one molecule with the hydrophobic patch of a symmetry-related molecule (thin lines) as observed in the crystal. Residues 4 to 12 of the NH_2-terminus are shown. View of the hydrophobic patch is the same as in (A). Four residues of the NH_2-terminus interact with the hydrophobic surface of another molecule. The phenyl ring of Phe^{10} fits into a cavity formed by Tyr^{100}, Phe^{101}, and Phe^{104} and stacks on Phe^{104}; Pro^{8} fits between Tyr^{57}, Ile^{96}, and Tyr^{100}; the aliphatic side chain atoms of Glu^{7} stack on Tyr^{100}; and the side chains of Val^{6} and Ile^{96} are in van der Waals contact. Leu^{67} and Met^{110} from the same pair of molecules are also in contact. This extended interaction also involves hydrogen bonds between the amide of Val^{6} and the carbonyl of Arg^{95}, the carbonyl of Val^{6} and the amide of Asp^{97}, and the carbonyl of His^{4} and the side chain of Arg^{95}.

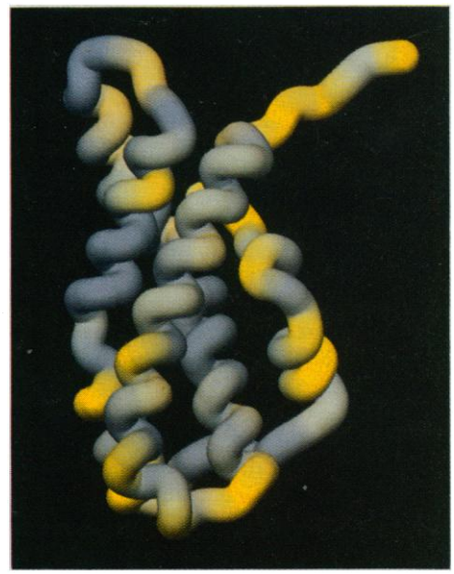

Fig. 4. Sequence variability and conservation in lysin from seven species of California abalone. On the basis of the sequence alignment (*2, 11*), the protein chain of red abalone lysin is shaded by residue in four colors: bright yellow, hypervariable; pale yellow, variable; light gray, conserved; and dark gray, strictly conserved. The NH_2-terminus extends to the upper right and is believed to be involved in species recognition.

Table 1. Structure determination. Crystals of red abalone (*Haliotis rufescens*) lysin were obtained as previously described (*12*). The crystals belong to space group $P2_12_12_1$, with a = 52.6 Å, b = 45.9 Å, and c = 81.0 Å, and contain one molecule in the asymmetric unit (V_m = 3.05 Å^3 per dalton). Complete diffraction data to 1.9 Å resolution were collected from four crystals of the native protein. Data to 2.7 Å resolution were collected from one crystal of each of five heavy-atom derivatives. Data were collected with CuKα radiation from an Elliot GX-21 x-ray generator equipped with focusing mirrors and a Rigaku Ru-200 x-ray generator equipped with a graphite monochromator. Data were recorded with Siemens area detectors and were indexed, integrated, merged, and scaled with the Xengen suite of programs (*13*). The phase problem was solved with multiple isomorphous replacement (MIR) on the basis of four independent derivatives and one double derivative. Calculations were carried out with the program suite Xtalview (*14*). Heavy-atom positions were located with isomorphous, difference Patterson syntheses and cross-phased difference Fourier maps. Refinement of heavy-atom parameters and MIR phase calculations were done with Xheavy (*14*). The figure of merit was 0.66 for all reflections to 3.0 Å resolution. The electron density map was readily interpretable (Fig. 1), and the polypeptide backbone was traced with a minimap. Electron density for the NH_2-terminus (residues 1 to 3) and the COOH-terminus (residues 135 and 136) was not seen, and that of residue 77 was very weak. Model building and density fitting was done with Xfit (*14*). The initial model was built with overlapping pentamers from a database of well-refined structures (*14, 15*) with the minimap-derived α-carbon positions serving as guide coordinates. The structure was refined with X-PLOR (*16*). The R factor for the starting model was 0.48 for all data in the range 8.0 to 3.0 Å resolution. After refinement with the 3.0 Å resolution data, the resolution was extended to 2.5 Å with alternate cycles of positional refinement and density fitting to $2|F_o| - |F_c|$ maps and was then extended to 1.9 Å with additional cycles of positional and B-factor refinement. The fit of the entire model to the density was confirmed by simulated-annealing omit maps in which 10-residue segments were removed from the refinement and map calculations. The final 1.9 Å resolution $2|F_o| - |F_c|$ map showed no electron density for residues 1 to 3 and 135 and 136; these residues are not included in the model. Residues His^4 and Met^{134} are modeled as alanine because no density was present for the side chains beyond β-carbon atoms. Ordered water molecules (74 with $B < 50$ Å^2) were identified in $|F_o| - |F_c|$ difference Fourier maps contoured at 3σ. The final model contains 1181 atoms and has an R factor of 0.187 for 14,806 reflections with $|F_o| > 0$ in the resolution range 6.0 to 1.9 Å. The root-mean-square deviations from ideality of bonds and angles are 0.016 Å and 2.85°, respectively. The average B factor for all atoms is 21.9 Å^2. There are no outliers in the Ramachandran plot.

Data set	Resolution (Å)	Completeness (%)	Unique reflections (no.)	Redundancy	R_{symm}*	R_{merge}†	Sites (no.)	Phasing power‡
Native	1.9	100	15,624	6.4	0.070	—	—	—
$KAu(CN)_2$	2.7	85	5,028	2.8	0.055	0.167	1	1.41
$K_2Pt(SCN)_4$	2.7	88	5,155	4.0	0.051	0.236	2	1.90
$KAu(CN)_2$ and $K_2Pt(SCN)_4$	2.7	85	4,927	4.4	0.066	0.252	5	1.76
$K_2Pt(CN)_4$	2.7	81	4,687	3.6	0.055	0.387	1	1.31
$(CH)_3PbCl$	2.7	88	5,168	2.3	0.049	0.325	3	1.24

*R_{symm} is the unweighted R factor on intensities for multiple observations of symmetry-related reflections. †R_{merge} is the unweighted R factor on structure amplitudes for all reflections in common with the native data. ‡Phasing power is the ratio of the rms calculated heavy-atom structure amplitude to the rms lack of closure.

charged liposomes (*5*); and residues within the hydrophobic patch associate with aliphatic and aromatic amino acids (Fig. 3B). Although lysin interacts with lipid vesicles (*5*), it is important to note that in vivo it binds to a filamentous glycoprotein matrix, not a lipid bilayer. Lipid is not detected in isolated VEs (*1*). The function of the hydrophobic patch in VE dissolution remains unknown.

There is precedent for proteins having surface features similar to lysin. The apolipoprotein E receptor binding domain is a helical protein with a surface cluster of nine basic amino acids; this region is predicted to be the binding site for the low-density lipoprotein receptor (*6*). Colicin A exhibits a ring of basic residues expected to orient the protein on the lipid bilayer (*7*). Defensin is a cationic, amphipathic protein able to insert into lipid bilayers that also has solvent-exposed hydrophobic residues (*8*). The structure of *Limulus* antilipopolysaccharide factor possesses both an exposed hydrophobic surface and a marked clustering of basic residues (*9*). Tyrosine and tryptophan are found in abundance in antibody combining sites (*10*); in lysin, the hydrophobic patch contains three tyrosine and one tryptophan residues, consistent with the possibility that it is a biologically active surface.

Lysin is required for sperm to penetrate the egg envelope. It also defines the species specificity of fertilization. For example, black abalone lysin is poor at dissolving red abalone VEs, and the reciprocal cross mixture yields similar results (*2*). Red abalone lysin is poor at dissolving pink abalone VEs (*2*). The NH_2-terminal domain of residues 2 through 12 is strictly species-specific (*2, 11*). The COOH-terminal domain of residues 132 through 136, though not species-specific, is variable. These two most variable domains of lysin are adjacent in the structure (Fig. 2). Figure 4 depicts the hypervariable, variable, conserved, and strictly conserved amino acid positions of lysins from seven species (*2, 11*) in relation to the three-dimensional structure. The distribution of conserved residues within the molecule suggests that lysins of other abalone species have the same structure. The hypervariable positions may be those mediating species recognition between lysin and the VE.

Lysin disrupts the structural integrity of the egg VE by a binding interaction (*1*). Although detailed data do not exist on the structure of the VE glycoprotein matrix, the crystal structure of lysin suggests possible mechanisms of action. After species recognition, positive charges in the basic tracks may promote binding to the VE, allowing the hydrophobic patch to unfold VE glycoproteins. Alternatively, lysin binding may disrupt hydrogen bonding or induce allosteric changes in VE molecules, causing their dissociation. Although the mechanism by which sperm penetrate egg envelopes remains unknown in higher species, the dissolution of egg envelopes by surface-active sperm proteins may be a more general phenomenon of fertilization.

REFERENCES AND NOTES

1. C. A. Lewis, C. F. Talbot, V. D. Vacquier, *Dev. Biol.* **92**, 227 (1982).
2. V. D. Vacquier, K. R. Carner, C. D. Stout, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 5792 (1990); V. D. Vacquier and Y.-H. Lee, *Zygote*, in press.
3. W. Kabsch and C. Sander, *Biopolymers* **22**, 2577 (1983). The solvent-accessible surface was calculated with the program MS by M. L. Connolly [*Science* **221**, 709 (1983)] with a 1.4 Å probe sphere. Surface areas are computed on the basis of an expected value of five points per square angstrom.
4. J. P. Segrest, H. D. Loof, J. G. Dohlman, C. G. Brouillette, G. M. Anantharamaiah, *Proteins* **8**, 103 (1990).
5. K. Hong and V. D. Vacquier, *Biochemistry* **25**, 543 (1986).
6. C. Wilson, M. R. Wardell, K. H. Weisgraber, R. W. Mahley, D. A. Agard, *Science* **252**, 1817 (1991).
7. M. W. Parker, J. P. M. Postma, F. Pattus, A. D. Tucker, D. Tsernoglou, *J. Mol. Biol.* **224**, 639 (1992).
8. C. P. Hill, J. Yee, M. E. Selsted, D. Eisenberg, *Science* **251**, 1481 (1991).
9. A. Hoess, S. Watson, G. R. Siber, R. Liddington, *EMBO J.* **12**, 3351 (1993).

10. I. S. Mian, A. R. Bradwell, A. J. Olson, *J. Mol. Biol.* **217**, 133 (1991).
11. Y.-H. Lee and V. D. Vacquier, *Biol. Bull. Woods Hole Mass.* **182**, 97 (1992).
12. M. L. Baginsky, C. D. Stout, V. D. Vacquier, *J. Biol. Chem.* **265**, 4958 (1990).
13. A. J. Howard, C. Nielsen, Ng. H. Xuong, in *Methods in Enzymology*, H. W. Wyckoff, C. H. W. Hirs, S. N. Timasheff, Eds. (Academic Press, Orlando, FL, 1985), vol. 114A, chap. 28.
14. D. E. McRee, *J. Mol. Graph.* **10**, 44 (1992).
15. T. H. Jones and S. Thirup, *EMBO J.* **5**, 819 (1986).
16. A. T. Brünger, M. Karplus, G. A. Petsko, *Acta Crystallogr. A* **45**, 50 (1989).
17. We thank M. Pique for computer graphics and C. Fisher for assistance with computer programs. We thank J. D. Bleil, J. A. Tainer, L. Kuhn, and N. B. Gilula for stimulating discussions. Coordinates have been deposited to the Protein Data Bank, accession number 1LIS. Supported by National Science Foundation grant MCB 9205020 (C.D.S.) and National Institutes of Health grant HD12986 (V.D.V.).

1 July 1993; accepted 19 October 1993

A Functional Recombinant Myosin II Lacking a Regulatory Light Chain–Binding Site

Taro Q. P. Uyeda* and James A. Spudich†

Myosin II, which converts the energy of adenosine triphosphate hydrolysis into the movement of actin filaments, is a hexamer of two heavy chains, two essential light chains, and two regulatory light chains (RLCs). *Dictyostelium* myosin II is known to be regulated in vitro by phosphorylation of the RLC. Cells in which the wild-type myosin II heavy chain was replaced with a recombinant form that lacks the binding site for RLC carried out cytokinesis and almost normal development, processes known to be dependent on functional myosin II. Characterization of the purified recombinant protein suggests that a complex of RLC and the RLC binding site of the heavy chain plays an inhibitory role for adenosine triphosphatase activity and a structural role for the movement of myosin along actin.

The enzymatic and motile activities of nonmuscle myosin II from many cell types, including *Dictyostelium* and various animal cells, are regulated by phosphorylation of RLC in vitro in a manner similar to smooth muscle myosin (*1*). Fluctuations in vivo of either the phosphorylation levels of RLC or that of the RLC-kinase have been demonstrated (*2*), and these observations have led to the hypothesis that RLC phosphorylation plays a crucial role in myosin-dependent motile activities in nonmuscle cells (*1*, *2*). We tested this hypothesis here by expressing a mutant *Dictyostelium discoideum* myosin II heavy chain with an internal deletion that removes the binding site for RLC.

The NH_2-terminal half of myosin II is called subfragment 1 (S1), and S1 alone is sufficient to hydrolyze adenosine triphosphate (ATP) and move actin filaments in vitro (*3*). The S1 molecule has a globular catalytic portion, which contains ATP- and actin-binding sites, and a neck region, which connects the catalytic domain to the α-helical coiled-coil tail of the myosin molecule (*1*, *4*, *5*). Other studies have localized the light chains to the neck region of S1 (Fig. 1B) (*6*). Crystallographic analysis of S1 from chicken skeletal muscle has identified amino acid residues of the heavy chain that interact with RLC (*5*). This region contains the "IQ" motif, which is found in sites that bind calmodulin and related polypeptides such as myosin light chains (*7*). A similarity between the chicken and *Dictyostelium* sequences allowed us to identify a region of the *Dictyostelium* heavy chain that is likely to bind RLC (Fig. 1A). The nucleotide sequence that corresponds to this region was deleted from the wild-type gene for myosin II heavy chain (*mhc*A) with polymerase chain reaction (PCR) technology. This mutant gene was cloned into a self-replicating vector (*8*) and used for transformation of an *mhc*A$^-$ *Dictyostelium* cell line (*9*). We used the transformants to study the mutant phenotype and to purify the mutant myosin MyΔRLCBS (Fig. 1C).

Wild-type *Dictyostelium* cells are mostly mononucleate and grow by binary fission in suspension culture with a doubling time of 5.1 ± 0.1 hours (n = 3; mean ± SEM) (*10*). When starved, they differentiate to form fruiting bodies (Fig. 2A). Both of these processes require functional myosin II (*11*). The *mhc*A$^-$ cells, which cannot perform cytokinesis, become multinucleated in suspension culture and eventually lyse. During the development of fruiting bodies, these cells are unable to proceed beyond the "mound" stage (Fig. 2B). The transformed cells that expressed the MyΔRLCBS *mhc*A gene in place of the native *mhc*A gene regained the ability to divide and grow in suspension culture with a doubling time of 5.5 ± 0.2 hours (n = 3; mean ± SEM) and stayed mostly mononucleate. Thus, the MyΔRLCBS cells undergo cytokinesis in suspension, unlike *mhc*A$^-$ cells. Expression of MyΔRLCBS also restored the ability of the cells to form fruiting bodies, but the

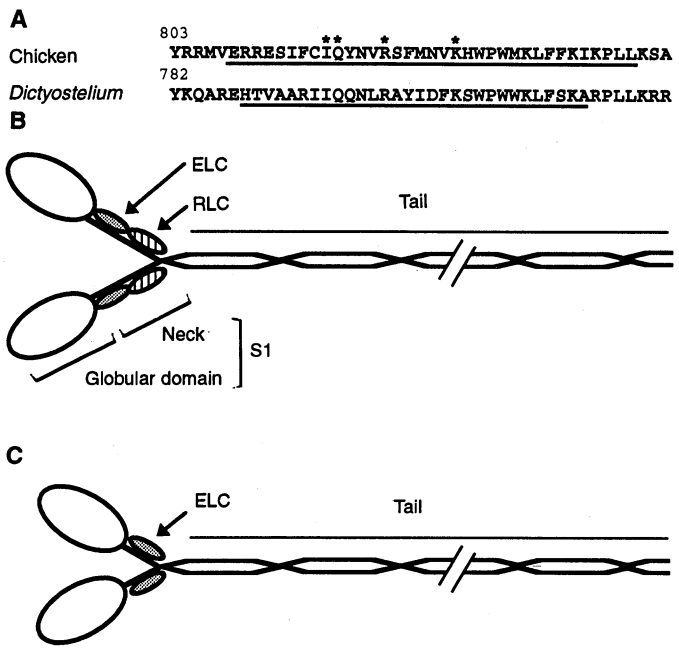

Fig. 1. (**A**) The amino acid sequences of the carboxyl portion of the neck region of S1 for chicken skeletal muscle (*30*) and *Dictyostelium* myosin II (*31*). The underlined residues of the chicken sequence interact with RLC in the crystal structure (*5*). The corresponding region in the *Dictyostelium* sequence (underlined) was deleted from the wild-type *mhc*A gene, resulting in the MyΔRLCBS *mhc*A gene. Asterisks indicate conserved residues of the "IQ" motif (*7*). (**B** and **C**) Schematic representations of the wild-type myosin II and MyΔRLCBS molecules, respectively. Numbers indicate the residue number counting from NH_2-terminus. 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.

Departments of Biochemistry and of Developmental Biology, Stanford University School of Medicine, Stanford, CA 94305.

*Present address: National Institute for Advanced Interdisciplinary Research, AIST, Tsukuba, Ibaraki 305, Japan.
†To whom correspondence should be addressed.