

consistent with studies demonstrating Id expression in dividing neuroblasts (15, 31), presumptive interneurons, and motor neurons in the developing spinal cord (15). In *Drosophila melanogaster*, the Id homolog extramacrochaete regulates sensory organ patterning in the adult fly epidermis by negatively regulating the bHLH proteins achaete and scute (32, 33) in a dose-dependent manner (33, 34). In vertebrates, our results demonstrate that Id2 plays stage-dependent roles in ectodermal development, likely depending on the presence of other interacting bHLH molecules. Thus, Id2 heterodimers may maintain the balance between differentiation and proliferation in neural crest and ectodermal precursors.

References and Notes

1. N. Le Douarin, *The Neural Crest*, vol. 12 of *Developmental and Cell Biology Series* (Cambridge Univ. Press, Cambridge, 1982).
2. D. M. Noden, *Dev. Biol.* **42**, 106 (1975).
3. N. M. Le Douarin and M. A. Teillet, *ibid.* **41**, 162 (1974).
4. G. G. Leblanc, M. L. Epstein, M. E. Bronner-Fraser, *ibid.* **137**, 318 (1990).
5. H. Nakamura and C. S. Ayer-le Lievre, *J. Embryol. Exp. Morphol.* **70**, 1 (1982).
6. RNA was isolated from the presumptive midbrain cranial neural folds (four to six somites) and from dispase-isolated trunk (16 to 22 somites) neural tubes (spanning the last three somites up to Hensen's node) of White Leghorn chicken embryos, with RNazol B (Tel-Test). The resulting pools of RNA were compared by Differential Display (Display Systems). Subsequently, one differentially displayed band was used for whole-mount in situ hybridization and for a 12 to 15 HH stage chick embryo library screen.
7. X. H. Sun, N. G. Copeland, N. A. Jenkins, D. Baltimore, *Mol. Cell. Biol.* **11**, 5603 (1991).
8. S. Sawai and J. A. Campos-Ortega, *Mech. Dev.* **65**, 175 (1997).
9. R. Benezra, R. L. Davis, D. Lockshon, D. L. Turner, H. Weintraub, *Cell* **61**, 49 (1990).
10. H. Zhang, S. Reynaud, M. Kloc, L. D. Etkin, G. Spohr, *Mech. Dev.* **50**, 119 (1995).
11. Y. Jen, H. Weintraub, R. Benezra, *Genes Dev.* **6**, 1466 (1992).
12. T. J. Brennan, D. G. Edmondson, L. Li, E. N. Olson, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 3822 (1991).
13. L. A. Neuhoff and B. Wold, *Cell* **74**, 1033 (1993).
14. T. Neuman et al., *Dev. Biol.* **160**, 186 (1993).
15. Y. Jen, K. Manova, R. Benezra, *Dev. Dyn.* **208**, 92 (1997).
16. T. Ogata, J. M. Wozney, R. Benezra, M. Noda, *Proc. Natl. Acad. Sci. U.S.A.* **90**, 9219 (1993).
17. R. B. Wilson et al., *Mol. Cell. Biol.* **11**, 6185 (1991).
18. G. Condorelli et al., *Blood* **86**, 164 (1995).
19. A. Ishiguro et al., *ibid.* **87**, 5225 (1996).
20. B. A. Christy et al., *Proc. Natl. Acad. Sci. U.S.A.* **88**, 1815 (1991).
21. A. Iavarone, P. Garg, A. Lasorella, J. Hsu, M. A. Israel, *Genes Dev.* **8**, 1270 (1994).
22. W. Shoji, T. Inoue, T. Yamamoto, M. Obinata, *J. Biol. Chem.* **270**, 24818 (1995).
23. Vectors with or without a myc tag were made with VENT DNA polymerase. A cDNA fragment of the chick Id2 coding region was amplified with Eco R1 and Ava 1-Stu 1 primer linkers, and a cDNA fragment for the non-myc-tagged vector was amplified with Nco 1 and Ava 1-Stu 1 primer linkers. Between all cloning steps, the cDNA fragments were purified with a GeneClean Kit from BRL. The amplified chick Id2 coding region used to make the myc-tagged RCASBP (B) vector was cloned into the Eco R1 and Stu 1 sites of a CS2 + MT (myc tag) vector, cut with Nco 1 and Ava 1, cloned into a SLAX-13 shuttle vector at the Nco 1 and Ava 1 sites, cut with Cla 1, and cloned into

- the Cla 1 site of a RCASBP (B) vector. The non-myc-tagged vector was cloned directly into the SLAX-13 vector, bypassing the CS2 + MT (myc tag) vector. Infection of chicken embryonic line O fibroblast cells revealed expression of myc-tagged Id2 protein in the nucleus of nearly all cells by 2 days after infection, demonstrating proper expression.
24. B. C. Morgan and D. M. Fekete, in *Methods in Avian Embryology*, M. Bronner-Fraser, Ed. (Academic Press, San Diego, CA, 1996), vol. 51, pp. 186-217.
25. S. H. Hughes, J. J. Greenhouse, C. J. Petropoulos, P. Suttrave, *J. Virol.* **61**, 3004 (1987).
26. S. A. Homburger and D. M. Fekete, *Dev. Dyn.* **206**, 112 (1996).
27. V. Hamburger and H. L. Hamilton, *J. Morphol.* **88**, 49 (1951).
28. J. Sechrist, M. A. Nieto, R. T. Zamanian, M. Bronner-Fraser, *Development* **121**, 4103 (1995).
29. Y. Jen, K. Manova, R. Benezra, *Dev. Dyn.* **207**, 235 (1996).

30. M. A. Selleck and M. Bronner-Fraser, *Development* **121**, 525 (1995).
31. Y. Nagata, W. Shoji, M. Obinata, K. Todokoro, *Biochem. Biophys. Res. Commun.* **207**, 916 (1995).
32. J. Garrell and J. Modolell, *Cell* **61**, 39 (1990).
33. H. M. Ellis, D. R. Spann, J. W. Posakony, *ibid.*, p. 27.
34. J. Botas, J. Moscoso del Prado, A. Garcia-Bellido, *EMBO J.* **1**, 307 (1982).
35. E. Hara et al., *Dev. Genet.* **18**, 161 (1996).
36. We thank S. Fraser, C. Baker, and M. Dickinson for critical reading of the manuscript; J. Sechrist for critical analysis of ectoderm conversion data; C. Marcelle, T. Moreno, M. Barembaum, and C. LaBonne for advice on molecular biology protocols; and K. Min and J. Neri for cryostat sectioning. We are grateful to S. Bryant and D. Gardiner for the stage 13 to 24 chick embryo library. Supported by U.S. Public Health Service grants NS34671 and NS36585.

16 March 1998; accepted 1 July 1998

Crystal Structure of Hemolin: A Horseshoe Shape with Implications for Homophilic Adhesion

Xiao-Dong Su,* Louis N. Gastinel,† Daniel E. Vaughn,‡ Ingrid Faye, Pak Poon, Pamela J. Bjorkman§

Hemolin, an insect immunoglobulin superfamily member, is a lipopolysaccharide-binding immune protein induced during bacterial infection. The 3.1 angstrom crystal structure reveals a bound phosphate and patches of positive charge, which may represent the lipopolysaccharide binding site, and a new and unexpected arrangement of four immunoglobulin-like domains forming a horseshoe. Sequence analysis and analytical ultracentrifugation suggest that the domain arrangement is a feature of the L1 family of neural cell adhesion molecules related to hemolin. These results are relevant to interpretation of human L1 mutations in neurological diseases and suggest a domain swapping model for how L1 family proteins mediate homophilic adhesion.

Insects have developed highly efficient innate forms of immunity against invading microorganisms such as bacteria and fungi (1). In the giant silkworm *Hyalophora cecropia* and the tobacco hornworm *Manduca sexta*, many

proteins are up-regulated in larvae or pupae upon bacterial infection. Hemolin is present in low amounts in the hemolymph of naïve insects, but is highly induced upon bacterial infection, and is assumed to be an integral component of the insect immune response (2).

Hemolin is a member of the immunoglobulin superfamily (IgSF), containing four Ig-like domains (3). It shares significant sequence similarity with the first four domains of the IgSF portion of transmembrane cell adhesion molecules (CAMs) of the L1 family, whose extracellular regions consist of six IgSF domains followed by five fibronectin III repeats (4) [~38% amino acid sequence identity between hemolin and the four NH₂-terminal IgSF domains of neuroglian (3), the insect ortholog of mammalian L1]. L1 family members mediate homophilic and heterophilic adhesion events that facilitate neurite outgrowth and fasciculation. Mutations in the human L1 gene are found in a variety of neurological disorders (4, 5).

X.-D. Su, L. N. Gastinel, P. J. Bjorkman, Division of Biology 156-29 and Howard Hughes Medical Institute, California Institute of Technology, Pasadena, CA 91125, USA. D. E. Vaughn, Division of Biology 156-29, California Institute of Technology, Pasadena, CA 91125, USA. I. Faye, Department of Genetics, Stockholm University, S-106 91 Stockholm, Sweden. P. Poon, Department of Chemistry and Biochemistry, University of California at Los Angeles, Los Angeles, CA 90095, USA.

*Present address: Department of Molecular Biophysics, Center for Chemistry and Chemical Engineering, Post Office Box 124, Lund University, S-221 00 Lund, Sweden.

†Present address: AFMB, CNRS UPR 9039, 31 Chemin Joseph Aiguier, 13402 Marseille Cedex 20, France.

‡Present address: Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724, USA.

§To whom correspondence should be addressed. E-mail: bjorkman@cco.caltech.edu

REPORTS

Protein phylogenetic analyses suggest that hemolin evolved from an L1-like ancestor, developing an immune system function independently of vertebrate members of the IgSF (6). Although its exact function in insect immunity remains elusive, it shares homophilic adhesion properties with related neural CAMs; for example, hemolin is found in a membrane form on hemocytes and can mediate homophilic adhesion (7). Secreted hemolin binds to hemocytes, inhibiting their aggregation (3, 8), perhaps by preventing homophilic interactions of the membrane form (7). Hemolin also binds to bacteria (3, 9) and to lipopolysaccharide (LPS) (10), a component of bacterial outer membranes.

We determined the 3.1 Å crystal structure of *H. cecropia* hemolin (Table 1) (11, 12). Hemolin is composed entirely of β -structure, consisting of four Ig-like domains (D1, D2, D3, and D4) that adopt the I-set folding topology (Fig. 1A) (13). The D2-D3 interface of hemolin includes a bound phosphate ion near two positively charged residues (D2 Arg¹⁵³ and D3 Arg²⁶⁶) (Fig. 1B) and is primarily associated with a loop between D3 strands D and E that contains residues typically found in phosphate-binding sites (Fig. 2A legend) (14). The two arginines and the sequence of the D-to-E loop are conserved in hemolin sequences, but not in the related L1 family members (Fig. 2A). Phosphate binding may be a feature related to hemolin's interactions with negatively charged LPS (10), such that this site, which is in a particularly basic portion of the protein (Fig. 1B), could represent the binding site for phosphate groups of LPS.

Although the individual hemolin domains resemble IgSF domains in other proteins, they are arranged into an unusual globular shape resembling a horseshoe. A sharp bend at the D2-D3 domain interface is responsible for the horseshoe shape, such that the almost linearly arranged D3-D4 segment folds back upon the D1-D2 segment, which is also almost linearly arranged. Hemolin's shape resembles the four domain structures of T cell receptors or the F_{ab} portions of antibodies rather than the "beads on a string" arrangement of domains in CD4, the only other single-chain four-domain IgSF protein of known structure (15). Unlike T cell receptors and F_{ab}'s, however, the interacting segments of hemolin are antiparallel to each other (Fig. 1A).

The angle relating the D2 and D3 domains of hemolin is the most acute interdomain angle in the available structures of IgSF or IgSF-related proteins. Specifically, hemolin D2 and D3 are related by an angle of 25°, whereas other bent IgSF structures show interdomain angles of 65° or more (Fig. 1A). As a result of the sharp bend between D2 and D3, the hemolin domains

interact strongly in pairwise combinations: D1 with D4 and D2 with D3. A total of 1217 and 1382 Å² of surface area is buried (12) at the D1-D4 and D2-D3 interfaces, respectively; this is slightly more than is buried at the killer inhibitory receptor (KIR) D1-D2 interface (1076 Å²) (15), but less than is buried at the interfaces in an F_{ab} [the V_H-V_L interface buries 1634 Å² and the C_{H1}-C_L interface buries 2051 Å² in the F_{ab} (15) shown in Fig. 1A]. Both the D1-D4 interface of hemolin and the KIR interdomain interface are formed largely by contacts between the β -sheets containing strands G, F, and C (GFC sheets), whereas the ABED sheets of hemolin mediate the D2-D3 contact (Fig. 1A). The use of opposite faces to mediate the two lateral interdomain contacts is reminiscent of the organization of an F_{ab}. By contrast to the extensive interactions between the paired hemolin domains, the adjacent hemolin interdomain interfaces (D1-D2 and D3-D4) bury little surface area: 576 and 280 Å², respectively, similar to surface areas buried between adjacent domains in elongated rod-like multidomain IgSF structures such as CD4, CD2, and VCAM-1 (~400 to 950 Å²) (15). Thus in isolation, the D1-D2 and D3-D4 segments of hemolin resemble more elongated IgSF molecules such as CD4. Hemolin's unusual shape is therefore a di-

rect result of the sharp D2-D3 bend and pairwise interactions between the D2-D3 and D1-D4 domains.

The significant sequence identity between hemolin and the first four domains of neuroglian ensures that the hemolin-related domains of L1 family members fold into tertiary structures resembling their counterpart hemolin domains (16). To determine if the four hemolin-related domains of L1 proteins share a common interdomain quaternary arrangement with hemolin (that is, an antiparallel interaction of D1 with D4 and D2 with D3), we compared residues at the hemolin D1-D4 and D2-D3 interfaces with their L1 and neuroglian counterparts. Many are identical or chemically similar in hemolin, neuroglian, and human L1 (Fig. 2, A and B). Sequence conservation is notable in the strand F and G regions of D1 and D4, the strand D to E region of D2, and the strand B to C region of D3, which are main areas of interdomain contacts. The length (but not the sequence) of the D2-D3 linking region is conserved between hemolin and the L1 family, allowing the third and fourth domains of L1 proteins to fold back and make antiparallel interactions with the first two domains. The conservation of critical residues at the hemolin and L1 interdomain interfaces justifies use of the hemolin structure as a model for the

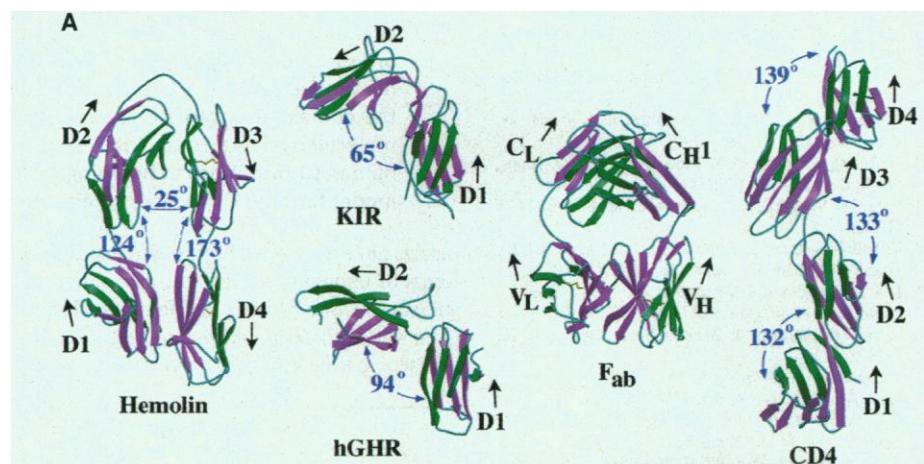


Fig. 1. (A) Structure of hemolin compared with other IgSF or IgSF-related structures (15) (ABE sheets are green; GFC sheets are purple). Bend angles (indicated with blue arrows between domains related by the angle) were calculated by determining the angle between the long axes of adjacent domains, approximated by ellipsoids calculated from the coordinates using the program Dom_angle (23). Arrows beside domain names indicate the NH₂- to COOH-terminal directions. KIR and human growth hormone receptor (hGHR) were oriented based upon the superposition of their D1 domains upon the hemolin D2 domain. **(B)** Bound phosphate ion is shown on the molecular surface of hemolin (left) [colors highlight the electrostatic potential calculated by GRASP (12); negative potential is in red and positive potential is in blue] and on a 3.1-Å 2F_{obs} - F_c annealed omit electron density map (12) (right; contoured at 0.9 σ).

REPORTS

organization of the first four domains of L1 family proteins.

By using velocity sedimentation analytical ultracentrifugation to compare the shapes of hemolin and a soluble version of the related portion of *Drosophila melanogaster* neuroglian (Nrg-4D) (11), we also obtained experimental evidence that domains 1 through 4 of an L1 family member are arranged similarly to hemolin. Sedimentation coefficients for each protein were determined and used to calculate frictional coefficients (17). The resulting comparison of hemolin and Nrg-4D with elongated and globular proteins of similar molecular masses (17) supports the conclusion derived from the sequence data, corroborating that the solution structures of both hemolin and Nrg-4D are the horseshoe shape observed in the hemolin crystal structure. In addition, recent studies of homophilic adhesion mediated by *Drosophila* neuroglian are consistent with the hypothesis that the hemolin-related domains of neuroglian are folded into a shape requiring all four domains for structural stability and function. In these studies, the first four neuroglian domains were both necessary and sufficient to mediate homophilic adhesion when expressed at the surface of S2 cells, whereas single domains alone or molecules in which any single domain was deleted did not mediate significant adhesion (18). The final confirmation of the postulated structure of Nrg-4D awaits a crystallographic analysis, because the limited resolution of electron microscopic studies precludes detailed structural interpretations of interdomain arrangements (19).

Having obtained sequence-based and experimental evidence that hemolin and the related domains of L1 proteins are structurally similar, we can use the hemolin structure as a first-order model to predict the structural effects of pathological missense mutations in this region of human L1. L1 mutations were previously mapped onto models of individual domains (5). The structural consequences of substitutions can now be interpreted, assuming a horseshoe shape for the hemolin-related L1 domains. Six of 13 mutations within the D1 to D4 region of L1 (all six of which affect residues that are identical or chemically similar in hemolin and human L1) fall at the predicted D2-D3 or D1-D4 interfaces (Fig. 2, A and B), consistent with the hypothesis that pairwise D1-D4 and D2-D3 interactions are important for the functions of L1 protein in cell adhesion.

Structure-based models proposed for cell adhesion mediated by other proteins include head-to-head interactions, as postulated for CD2-related proteins (15), or formation of a zipperlike structure, as proposed for homophilic recognition by cadherins (20). These models were based on interactions observed in crystals, in which the millimolar protein concentration environment was as-

sumed to reproduce weak adhesive interactions that would normally occur only at the cell surface. The packing in the hemolin crystals does not suggest any obvious mechanism for homophilic adhesion that would be induced solely by high concentrations of pro-

tein (21). However, the significance of the antiparallel packing of tandem Ig-like domains observed in hemolin and predicted for L1 family members may lie in the potential ability of such structures to form oligomers that could function in homophilic adhesion

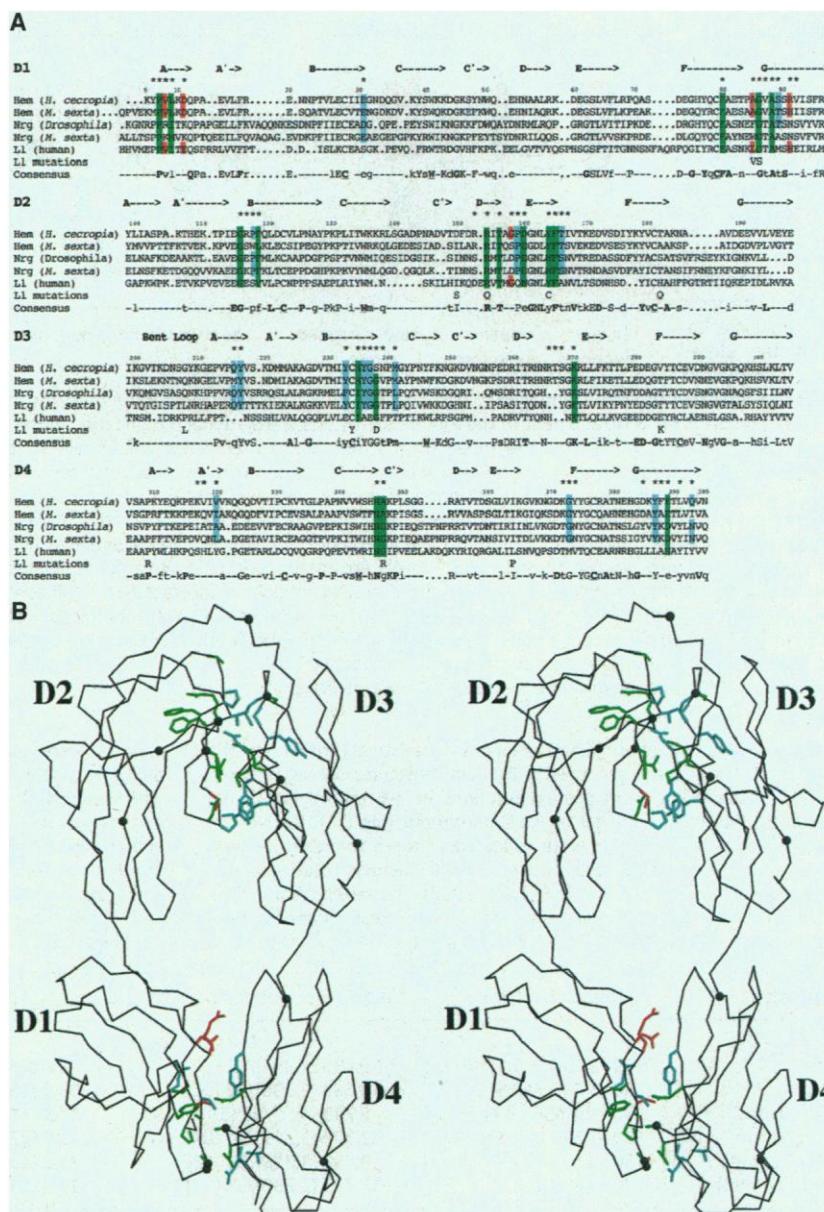


Fig. 2. (A) Sequence alignments (3, 24). Numbers refer to *H. cecropia* hemolin. Locations of β -strands in hemolin are indicated above the sequences with letter names. Residues at the D2-D3 and D1-D4 interdomain interfaces (those that contribute more than 10 \AA^2 of buried surface area to the interfaces) are indicated with an asterisk and are green if they are identical or chemically similar in L1 and a hemolin and neuroglian sequence, blue if they are identical or chemically similar in a hemolin and L1 sequence. Many of the highlighted residues at the D2-D3 and D1-D4 interfaces are also conserved in axonin 1, a vertebrate axon surface protein with which hemolin shares significant sequence identity (28%) (7). Substituted residues in L1 mutants (4, 5) are indicated below the L1 sequence. Hemolin's interactions with the bound phosphate ion include the side chains of His²⁶⁴, Arg¹⁵³, and Tyr²⁴³, and the main-chain nitrogens of Asn²⁶⁵, Arg²⁶⁶, Thr²⁶⁷, and Ser²⁶⁸. **(B)** Stereoview of the $\text{C}\alpha$ backbone of hemolin. Highlighted residues at the D2-D3 and D1-D4 interfaces [color-coded as in (A)] are identical or chemically similar in hemolin and L1 family proteins. $\text{C}\alpha$ atoms at positions corresponding to pathological mutations in human L1 (4, 5) are marked with a black sphere.

REPORTS

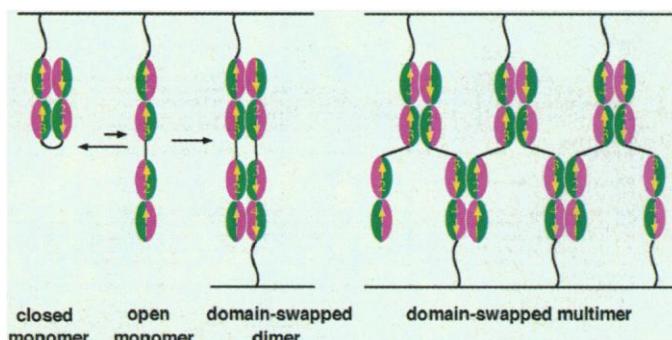
events. Hemolin dimers and oligomers with the same interdomain contacts as observed in the monomeric version that was crystallized could form by the mechanism of three-dimensional (3D) domain swapping, which has been documented for a variety of protein structures as occurring when a domain from a

monomeric protein is replaced by the same domain from an identical protein chain (22). A domain-swapped dimer of hemolin or an L1 family member would consist of intermolecular (rather than intramolecular) D2-D3 and D1-D4 pairs, which could be formed between open (straight) monomers with a

minor repositioning of the loop joining the D2 and D3 domains (Fig. 3). Repositioning of the D2-D3 loop would also allow formation of higher-order domain-swapped multimers of hemolin-related proteins.

Formation of domain-swapped multimers of hemolin and L1 family proteins is an attractive model for a structural mechanism of homophilic adhesion between two cell membranes, which can be stated as follows: (i) At the cell surface, the closed (bent) and open (straight) forms of the protein are in equilibrium favoring the bent form. (ii) Transient formation of the open (straight) form (perhaps in response to an extracellular signal) results in pairing with an open (straight) protein on another cell. (iii) Formation of additional domain-swapped dimers is facilitated by the close proximity of the adhering cell membranes. Alternatively, unpaired domains of the open (straight) proteins could nucleate formation of a ribbon of domain-swapped proteins, resulting in cell adhesion. Structural features of hemolin and L1 family members that are relevant for this model are that the D2-D3 linker is long enough to allow repositioning to form dimers or oligomers, or both, and that the D1-D4 and D2-D3 domain interfaces are fairly hydrophilic (Fig. 2A) and could therefore tolerate transient formation of open (straight) proteins before oligomerization.

Fig. 3. Schematic representation of a mechanism for homophilic adhesion mediated by 3D domain swapping (22) in hemolin and related proteins. On the left, the four NH₂-terminal domains of an L1 protein or the hemolin monomer (color coded interfaces as in Fig. 1A) are depicted in the closed (bent) conformation.



The black line indicates the remaining Ig-like, fibronectin type III, and transmembrane domains in the case of the L1 proteins (4) or attachment to the membrane by posttranslational modification in the case of hemolin (7). Transient formation of an open form would lead to formation of domain-swapped dimers (middle) or multimers (right) through homophilic interactions with open proteins on another cell. This model predicts that formation of a ribbon of domain-swapped proteins is more likely on a membrane than in solution: on a membrane where molecules are tethered to a surface, a ribbon of domain-swapped proteins could be nucleated through interactions of neighboring molecules with open proteins, rationalizing why soluble hemolin and Nrg-4D are monomeric (21) and soluble versions of homophilic IgSF-containing CAMs are generally monomeric. An antiparallel interaction of IgSF domains may be a general mechanism for homophilic adhesion mediated by neural CAMs in addition to proteins in the L1 family. For example, recent studies of homophilic adhesion mediated by N-CAM are consistent with an interaction of its five IgSF domains to create antiparallel D1-D5, D2-D4, and D3-D3 pairs (25), resulting in N-CAM dimers or multimers similar to those depicted in this figure.

Table 1. Data collection, phasing, and refinement statistics. Purified hemolin (11) was crystallized (space group P2₁2₁2₁, two molecules per asymmetric unit) using macroseeding from protein solutions at ~6 mg/ml in 50 mM phosphate, 0.15 M NaCl, and 1.8 M Na,K phosphate (pH 8.1). Native and heavy-atom derivative data sets were collected at room temperature on a Xenotronics multiwire area detector mounted on a Siemens rotating anode generator. Crystals soaked in 1.6 M (NH₄)₂SO₄ (Native I; *a* = 85.0 Å, *b* = 90.3 Å, *c* = 143.1 Å) diffract to higher resolution but are nonisomorphous with untreated crystals (Native II; *a* = 87.3 Å, *b* = 90.3 Å, *c* = 141.3 Å). Untreated

crystals were used for MIR phase determination, and the data from (NH₄)₂SO₄-soaked crystals were used for refinement. Crystals derivatized with xenon were collected as described (12). Data were processed with XDS and merged and scaled using CCP4 programs (12). Heavy-atom refinement and phasing were performed with the CCP4 version of MLPHARE to a figure of merit of 0.398 to 3.5 Å resolution. The MIR map was improved using NCS averaging and solvent-flipping using the program Solomon (12). The program O (12) was used for all model building, and the model was refined as described (12). Statistics in parentheses refer to the highest resolution bin.

Derivative	Resolution (Å)	Unique*/Percent complete†	<i>R</i> _{merge} (%)‡	<i>I</i> / <i>σ</i>	MFID (%)§	<i>f</i> _h / <i>E</i>
<i>Data collection</i>						
Native I	29–3.1 (3.3–3.1)	19,430 (1,241)/92 (56)	7.7 (24.6)	9.4 (3.1)	–	–
Native II	30–3.5 (3.7–3.5)	14,380 (1,998)/98 (96)	13.8 (35.0)	5.1 (2.0)	–	–
Xenon I (30 atm)	30–4.4 (4.6–4.4)	6,740 (427)/88 (39)	17.3 (18.3)	4.3 (4.0)	15.7	0.8
Xenon II (26 atm)	30–3.6 (3.8–3.6)	12,309 (1,215)/91 (64)	13.7 (27.0)	5.5 (2.7)	12.0	0.8
K ₂ PtCl ₄ I (0.2 mM)	30–4.0 (4.2–4.0)	9,789 (1,279)/98 (92)	25.0 (43.0)	3.0 (1.7)	23.1	1.6
K ₂ PtCl ₄ II (0.1 mM)	30–3.3 (3.5–3.3)	16,738 (2,230)/96 (91)	11.8 (34.7)	6.2 (2.2)	24.9	1.3
<i>Refinement statistics</i>						
Resolution	20.0–3.1 Å		rms Δ <i>B</i> bonded atoms			15.6 Å ²
Reflections in working set	17,966		rms Δ <i>φ</i> all NCS residues			3.2°
Reflections in test set	1,110		rms Δ <i>ψ</i> all NCS residues			9.3°
<i>R</i> _{cry} [¶]	21.8%		Number of nonhydrogen atoms			
<i>R</i> _{free} [#]	26.4%		Protein			6,098
rms deviations from ideal			Phosphate ion			10
Bond lengths	0.010 Å		Non-glycine residues in allowed			
Bond angles	1.4°		regions of Ramachandran plot as defined (12)			93%

*Unique, number of unique reflections. †Complete = (number of independent reflections)/(maximum number theoretical). ‡*R*_{merge} = $\sum |I - \langle I \rangle| / \sum I$, where *I* is the observed intensity and $\langle I \rangle$ is the average intensity from several measurements. §MFID (mean fractional isomorphous difference) = $\sum |F_{ph} - F_p| / \sum F_p$, where *F*_p = protein structure factor amplitude and *F*_{ph} = heavy-atom derivative structure factor amplitude. ¶*R*_{cry} = $\frac{\sum_h ||F_{obs}(h)| - |F_c(h)||}{\sum_h |F_{obs}(h)|}$, where *F*_{obs} and *F*_c are the observed and calculated structure factor amplitudes, respectively, for the *hkl* reflection. #*R*_{free} is calculated for a set of reflections that were not included in atomic refinement (12).

The hemolin structure reveals a new arrangement of IgSF domains that may be shared by related portions of neural CAMs. The antiparallel arrangement of tandem IgSF domains observed in hemolin and hypothesized to occur in L1 family members suggests a testable model for the mechanism of homophilic adhesion.

References and Notes

- H. G. Boman, *Cell* **65**, 205 (1991); J. A. Hoffmann, J.-M. Reichhart, C. Hetru, *Curr. Opin. Immunol.* **8**, 8 (1996); I. Faye and M. Kanost, in *Molecular Mechanisms of Immune Responses in Insects*, P. T. Brey and D. Hultmark, Eds. (Chapman & Hall, London, 1997).
- K. Andersson and H. Steiner, *Insect Biochem.* **17**, 133 (1987); N. E. Ladendorff and M. R. Kanost, *Arch. Insect Biochem. Physiol.* **15**, 33 (1990).
- References for protein sequences: *H. cecropia*: hemolin [S.-C. Sun, I. Lindström, H. G. Boman, I. Faye, O. Schmidt, *Science* **250**, 1729 (1990)]; revised sequence [I. Lindström-Dinnetz, S.-C. Sun, I. Faye, *Eur. J. Biochem.* **230**, 920 (1995)]; *M. sexta* hemolin sequence and inhibition of hemocyte aggregation experiments [N. E. Ladendorff and M. R. Kanost, *Arch. Insect Biochem. Physiol.* **18**, 285 (1991)]; *Drosophila* neuroglian [A. J. Bieber et al., *Cell* **59**, 447 (1989)]; *M. sexta* neuroglian [C.-L. Chen, D. J. Lampe, H. M. Robertson, J. B. Nardi, *Dev. Biol.* **181**, 1 (1997)]; and human L1 [M. Kobayashi, M. Miura, H. Asou, K. Uyemura, *Biochim. Biophys. Acta* **1090**, 238 (1991)].
- M. Hortsch, *Neuron* **17**, 587 (1996).
- A. Bateman et al., *EMBO J.* **15**, 6050 (1996).
- A. L. Hughes, *Immunogenetics* **47**, 283 (1998).
- R. Bettencourt, H. Lanz-Mendoza, K. Roxström Lindquist, I. Faye, *Eur. J. Biochem.* **250**, 630 (1997). Although adhesive activity mediated by L1 family members is calcium-independent, hemolin-mediated aggregation of microspheres requires calcium, suggesting that L1 proteins and hemolin mediate homophilic adhesion in response to different signals. This potentially important functional difference does not have profound structural implications, because calcium-binding sites comprise only a few residues (typically three to six) [C. A. McPhalen, N. C. J. Strynadka, M. N. G. James, *Adv. Protein Chem.* **42**, 77 (1991)].
- H. Lanz-Mendoza, R. Bettencourt, M. Fabbri, I. Faye, *Cell. Immunol.* **169**, 47 (1996).
- L. Zhao and M. R. Kanost, *J. Insect Physiol.* **42**, 73 (1996).
- S. Daffre and I. Faye, *FEBS Lett.* **408**, 127 (1997).
- Protein expression: Recombinant *H. cecropia* hemolin (full length, residues 1 to 395) was produced using a baculovirus expression system [M. D. Summers and G. E. Smith, in *A Manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures* (Texas Agricultural Experiment Station, College Station, TX, 1987), pp. 1-56] and purified from the supernatants of infected High 5 (Invitrogen) cells using a procedure similar to (7). A 6x-His tag and a stop codon were introduced into the *Drosophila* neuroglian cDNA (3) (gift of A. J. Bieber) following the codon for residue Gln⁴⁰⁵ (analogous to the COOH-terminal residue of hemolin, Asn³⁹⁵; Fig. 2A), and the protein (Nrg-4D) was expressed using a baculovirus expression system. Nrg-4D was purified from the supernatants of infected High 5 cells using nickel nitrilotriacetic acid (Ni-NTA) chromatography (Ni-NTA superflow; Qiagen) followed by gel filtration using a Superdex 200 fast protein liquid chromatography column (Pharmacia).
- Xenon derivatization: [M. H. B. Stowell et al., *J. Appl. Crystallogr.* **29**, 608 (1996)]; XDS: [W. Kabsch, *ibid.* **21**, 67 (1988)]. CCP4 programs: [CCP4: Collaborative Computational Project No. 4, Daresbury, UK (1994)]; *Acta Crystallogr. D* **50**, 760 (1994)]. Solomon: [J. P. Abrahams and A. G. W. Leslie, *ibid.* **52**, 30 (1996)]. O: [T. A. Jones and M. Kjeldgaard, *Methods Enzymol.* **277**, 173 (1997)]. *R_{free}*: [A. T. Brünger, *Nature* **355**, 472 (1992)]. The hemolin model was refined with X-PLOR [A. T. Brünger, *X-PLOR, Version 3.1: A System for X-ray and NMR* (Yale Univ. Press, New Haven, CT, 1992)] using strict noncrystallographic symmetry (NCS) constraints, and later tight NCS restraints (300 kcal/mol-Å²) between individual domains in the two molecules of the asymmetric unit. Residues that did not obey the NCS were not restrained (residues 5 to 7; 143 to 152, and part of the D2-D3 domain linking region, residues 203 to 210). Partial model phases were combined with experimental phases using SIGMA [R. J. Read, *Acta Crystallogr. A* **42**, 140 (1986)], and combined maps were used for further model building. Maximum-likelihood refinement, a bulk solvent correction, and grouped temperature (B) factors were used for the final refinement, as implemented in CNS (A. T. Brünger et al., *Acta Crystallogr. D*, in press). The current model contains 391 residues and a phosphate ion in each of the two NCS-related hemolin monomers (residues 5 to 395) (average B factor for protein atoms is 19.6 Å²). Excluding regions that deviate from the NCS, the domains in the two hemolin monomers are very similar (≤0.06 Å rms deviation for C α atoms). The close similarity suggests that hemolin's shape is unlikely to be an artifact of crystallization, because the two NCS-related molecules occupy different environments and are subjected to different packing forces. Electron density corresponding to carbohydrate is not seen at the one potential N-linked glycosylation site (Asn²⁶⁵) in the revised hemolin sequence (3). Ramachandran plot statistics (Table 1) are as defined by G. J. Kleywegt and T. A. Jones [*Structure* **4**, 1395 (1996)]. Molecular surface areas buried by interaction were calculated using X-PLOR with a 1.4 Å radius. Electrostatic calculations were done and Fig. 1B was made using GRASP [A. Nicholls, R. Bharadwaj, B. Honig, *Biophys. J.* **64**, A166 (1993)]. Figures 1A and 2B were made using MOLSCRIPT [P. J. Kraulis, *J. Appl. Crystallogr.* **24**, 946 (1991)] and RASTER-3D [E. A. Merritt and M. E. P. Murphy, *Acta Crystallogr. D.* **50**, 869 (1994)].
- The sharing of strand A between the GFC and BED faces of the hemolin domains is a feature of V- and I-set, rather than C2, domains, as predicted by recent structure-based sequence alignments that suggest that many domains previously classified as C2, such as the hemolin domains and the IgSF domains of many CAMs, are members of the newly recognized I-set family [Y. Harpaz and C. Chothia, *J. Mol. Biol.* **238**, 528 (1994)]; D. E. Vaughn and P. J. Bjorkman, *Neuron* **16**, 261 (1996); C. Chothia and E. Y. Jones, *Annu. Rev. Biochem.* **66**, 823 (1997)]. Each hemolin domain includes a disulfide bond linking strands B and F; all domains except D3 contain a tryptophan at the position of the "invariant" tryptophan in Ig-like domains (D3 has a tyrosine at this position).
- R. R. Copley and R. R. Barton, *J. Mol. Biol.* **242**, 321 (1994). Although the structural details of hemolin's interactions with LPS remain to be elucidated, hemolin's phosphate-binding site and its structure in general bear no resemblance to the structure of bactericidal/permeability-increasing protein (BPI), a mammalian LPS-binding protein [L. J. Beamer, S. F. Carroll, D. Eisenberg, *Science* **276**, 1861 (1997)].
- Protein structures. CD4 [Protein Data Bank (PDB) code 1WIO], H. Wu, P. D. Kwong, W. A. Hendrickson, *Nature* **387**, 527 (1997); KIR (coordinates obtained from Q. R. Fan), Q. R. Fan et al., *Nature* **389**, 96 (1997); hGHR (PDB code 3HHR), A. M. De Vos, M. Ullsch, A. Kossiakoff, *Science* **255**, 306 (1992); F_{ab} (PDB code 1DBB), J. H. Arevalo, E. A. Stura, M. J. Taussigh, I. A. Wilson, *J. Mol. Biol.* **231**, 103 (1993); VCAM-1 (PDB code 1VCA), E. Y. Jones et al., *Nature* **373**, 539 (1995); CD2 (PDB code 1HNG), E. Y. Jones, S. J. Davis, A. F. Williams, K. Harlos, D. I. Stuart, *ibid.* **360**, 232 (1992), and ovalbumin (PDB code 1OVA), P. E. Stein, A. G. W. Leslie, J. T. Finch, R. W. Carrell, *J. Mol. Biol.* **221**, 941 (1991).
- The threshold of sequence identity that guarantees 3D similarity was established as 30% [C. Chothia and A. M. Lesk, *EMBO J.* **5**, 823 (1986)]. Length-dependent sequence identity thresholds are discussed by R. A. Abagyan and S. Batalov [*J. Mol. Biol.* **273**, 355 (1997) and references therein]. By these criteria, the hemolin domains and their counterparts in neuroglian are guaranteed to fold into similar tertiary structures.
- Ultracentrifugation experiments were performed at 20°C using a Beckman Optima XL-A analytical ultracentrifuge. Sedimentation coefficients were determined from sedimentation velocity experiments performed at 56,000 rpm using proteins [50 μg/ml in 5 mM phosphate and 0.1 M NaCl (pH 7.0)] monitored by absorbance at 220 or 230 nm. Hemolin (44,000 daltons) and Nrg-4D (51,000 daltons) (molecular masses determined by mass spectrometry) were compared to proteins of similar molecular masses: a globular protein [ovalbumin, 45,000 daltons; ~75 Å by 40 Å by 40 Å (15)] and an elongated protein containing four Ig-like domains arranged in tandem (invasin, 52,000 daltons; ~180 Å by 30 Å by 30 Å; Z. Hamburger, R. R. Isberg, P. J. Bjorkman, unpublished results). The shape-dependent translational friction coefficient (*f*) of each protein was calculated from the Svedberg equation [V. Bloomfield, W. O. Dalton, K. E. Van Holde, *Biopolymers* **5**, 135 (1967)], from which it can be derived that $f_1/f_2 \approx (s_2/s_1) (M_1/M_2)$, where *M* is molecular size and *s* is the experimentally determined sedimentation coefficient (*S*_{20,w}) in Svedberg units (1S = 1 × 10⁻¹³ s) (hemolin, 3.17S; Nrg-4D, 3.52S; ovalbumin, 3.32S; invasin, 2.75S). If two proteins have the same shape, $f_1/f_2 \approx 1$, while if protein₁ is more elongated than protein₂, $f_1/f_2 > 1$. We calculate that $f_{Nrg-4D}/f_{hemolin} = 1.04$; $f_{ovalbumin}/f_{hemolin} = 0.98$; $f_{invasin}/f_{hemolin} = 1.36$.
- G. Rutherford, Y.-m. E. Wang, A. J. Bieber, in preparation.
- Photomicrographs of rotary shadowed extracellular regions of neuroglian [H. P. Erickson, A. H. Huber, A. J. Bieber, P. J. Bjorkman, unpublished data; A. H. Huber, thesis, California Institute of Technology, Pasadena (1994)] and L1 [B. Drescher, E. Spiess, M. Schachner, R. Probstmeier, *Eur. J. Neurosci.* **12**, 2467 (1996)] are similar and indicate that both proteins include numerous points of flexibility. The L1 studies were interpreted as indicating that the NH₂-terminal Ig-like domains adopt an elongated structure. However, from our analysis of the published images, half are compatible with a folded structure at both ends of the protein.
- L. Shapiro et al., *Nature* **374**, 327 (1995).
- The two NCS-related hemolin monomers that show the strongest interaction (burial of a total of 2558 Å² of surface area) are related by an approximate two-fold axis (177° rotation, 0.4 Å translation). The resulting hemolin dimer packs in a parallel "side by side" arrangement rather than a "head to head" arrangement. Hemolin is monomeric at the concentrations used for equilibrium analytical ultracentrifugation and mass spectrometry (unpublished results), but could dimerize at higher concentrations.
- M. J. Bennett, M. P. Schlunegger, D. Eisenberg, *Protein Sci.* **4**, 2455 (1995).
- X.-D. Su and T. O. Yeates, unpublished data.
- Single-letter abbreviations for amino acid residues: 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.
- T. S. Ranheim, G. M. Edelman, B. A. Cunningham, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 4071 (1996).
- We thank T. O. Yeates for help with the Dom_angle program, H. P. Erickson for valuable discussions about ultracentrifugation data, A. J. Bieber for communicating results before publication, the Caltech PPMAL for mass spectrometry analyses, P. M. Snow and I. Nangiana for help with protein expression, Q. R. Fan for the KIR coordinates, and M. J. Bennett, L. M. Sánchez, and A. J. Chirino for discussions and help with crystallographic software. Hemolin coordinates have been deposited in the PDB (1BIH).

20 April 1998; accepted 1 July 1998