

The Intracellular Domain of the *Drosophila* Cholinesterase-Like Neural Adhesion Protein, Gliotactin, is Natively Unfolded

Tzviya Zeev-Ben-Mordehai,^{1,2} Edwin H. Rydberg,^{1,2} Ariel Solomon,¹ Lilly Toker,² Vanessa J. Auld,³ Israel Silman,^{2,†,*} Simone Botti,^{1,2} and Joel L. Sussman^{1,‡,*}

¹Department of Structural Biology, Weizmann Institute of Science, Rehovot, Israel

²Department of Neurobiology, Weizmann Institute of Science, Rehovot, Israel

³Department of Zoology, University of British Columbia, Vancouver, British Columbia, Canada

ABSTRACT *Drosophila* gliotactin (Gli) is a 109-kDa transmembrane, cholinesterase-like adhesion molecule (CLAM), expressed in peripheral glia, that is crucial for formation of the blood–nerve barrier. The intracellular portion (Gli-cyt) was cloned and expressed in the cytosolic fraction of *Escherichia coli* BLR(DE3) at 45 mg/L and purified by Ni-NTA (nitrilotriacetic acid) chromatography. Although migration on sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), under denaturing conditions, was unusually slow, molecular weight determination by matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry (MS) confirmed that the product was consistent with its theoretical size. Gel filtration chromatography yielded an anomalously large Stokes radius, suggesting a fully unfolded conformation. Circular dichroism (CD) spectroscopy demonstrated that Gli-cyt was >50% unfolded, further suggesting a nonglobular conformation. Finally, 1D-¹H NMR conclusively demonstrated that Gli-cyt possesses an extended unfolded structure. In addition, Gli-cyt was shown to possess charge and hydrophobic properties characteristic of natively unfolded proteins (i.e., proteins that, when purified, are intrinsically disordered under physiologic conditions *in vitro*). *Proteins* 2003;53:758–767.

© 2003 Wiley-Liss, Inc.

Key words: expression; cholinesterase-like; CLAM; intrinsically disordered; neural cell adhesion; acetylcholinesterase

INTRODUCTION

Cholinesterase-like adhesion molecules (CLAMs) are a family of neural cell adhesion proteins defined by the presence of an extracellular domain with high sequence similarity to acetylcholinesterase (AChE) (Fig. 1). This includes the presence of two to three of the conserved disulfide bonds and one or two members of the catalytic triad, although the catalytic serine is always absent (see recent review¹). Although no three-dimensional (3D) structures are available for members of this family, and little or no biophysical characterization has been performed, evidence indicates that the similarities between the cholinesterase-like (ChE-like) domains and AChE are more than

sequence-deep. Calculation of the electrostatic potentials of several CLAMs, with the use of homology models, suggested that their ChE-like domains possess an electrostatic surface distribution very similar to that of AChE.² Furthermore, a chimera in which the ChE-like domain of neurotactin (Nrt) had been replaced by AChE itself (from *Torpedo* or *Drosophila*) retained the cell adhesion properties of Nrt.^{3,4} This latter study also suggests a noncholinergic role, or roles, for AChE, as supported by other lines of evidence.^{1,5–8}

Although the CLAMs are defined by their ChE-like domain, several are transmembrane proteins with a single, membrane-spanning helix and a relatively small intracellular domain (typically 100–300 residues). Whereas some of these intracellular domains contain small motifs com-

Grant sponsor: European Commission Fifth Framework “Quality of Life and Management of Living Resources” program; Grant number: QLK3-2000-00650

Grant sponsor: “SPINE” Project; Grant number: QLG2-CT-2002-00988

Grant sponsor: Infrastructure Grant from The Israel Ministry of Science.

Grant sponsor: The Israel Ministry of Science for the Israel Structural Proteomics Center.

Grant sponsor: Divadol Foundation (Rehovot, Israel).

Grant sponsor: Kalman and Ida Wolens Foundation (Rehovot, Israel).

Grant sponsor: Jean and Julia Goldwurm Memorial Foundation (Rehovot, Israel).

Grant sponsor: The late Sally Schnitzer (Rehovot, Israel).

Grant sponsor: Helen & Milton A. Kimmelman Center for Biomolecular Structure and Assembly (Rehovot, Israel).

Grant sponsor: Benozio Center for Neurosciences.

Grant sponsor: Canadian Institutes for Health Research (Auld).

Grant sponsor: Golda and Dr. Yehiel Shwartzman and Sara and Haim Medvedi Families Postdoctoral Fellowship (Rydberg).

[†]Bernstein–Mason Professor of Neurochemistry

[‡]Morton and Gladys Pickman Professor of Structural Biology

*Correspondence to: Joel L. Sussman, Department of Structural Biology, Weizmann Institute of Science, Rehovot, 76100 Israel. E-mail: joel.sussman@weizmann.ac.il, or Israel Silman, Department of Neurobiology, Weizmann Institute of Science, Rehovot, 76100 Israel. E-mail: israel.silman@weizmann.ac.il

Received 14 February 2003; Accepted 25 March 2003

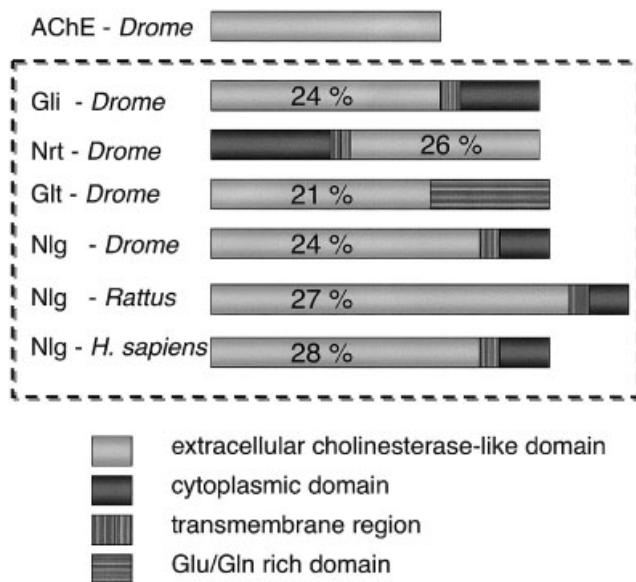


Fig. 1. Schematic representation of the general domain structure of the CLAMs. The percentage identity to *Drosophila* acetylcholinesterase (AChE) is shown in the corresponding domain box. Gli, gliotactin; Glt, glutactin; Nlg, neuroligin; Nrt, neurotactin.

mon to cytoplasmic proteins [e.g., ProGluSerThr (PEST) sequences found in Nrt,⁹ binding motifs for PDZ (a domain first found in the proteins PSD95, Dlg and ZO-1) in gliotactin (Gli) and neuroligin (Nlg),¹⁰ and putative protein-protein interaction LysGluLysGlu (KEKE) sequences that we have observed in Nrt], database searches have failed to reveal any significant sequence similarity between the CLAM intracellular domains and any other proteins (or even among the CLAMs themselves). However, whereas relatively little is known about the intracellular domains of the CLAMs, what is known suggests that they are crucial for neural cell adhesion and neural development in general. In a series of deletion experiments, the Nrt cytoplasmic domain was shown to be necessary for cellular adhesion.⁴ Likewise, interaction of Nlg with the PDZ domains of PSD-95 and S-SCAM is believed to be a critical event in the formation of glutamatergic synapses.¹⁰⁻¹²

Gli is a *Drosophila* CLAM expressed in glial cells during embryonic development.¹³ Embryos deficient in Gli appear morphologically normal but are paralyzed, showing uncoordinated peristaltic motion, and fail to hatch. The axons of such embryos have normal axonal transmission at low K⁺ concentrations but fail to transmit at physiologic K⁺ concentrations as a result of a breakdown in the blood-nerve barrier and incomplete wrapping of the peripheral axons by the glia.

Gli is a 109-kDa protein composed of an 86-kDa extracellular domain, itself composed of a 16-kDa domain of unknown function and a 70-kDa ChE-like domain, and a 23-kDa intracellular domain that is crucial for Gli function (D. Venema and V. Auld, unpublished data), connected by a 26-amino acid membrane-spanning helix.¹³ Apart from possessing a PDZ binding motif, the 207-amino acid

intracellular domain has no significant similarity to any protein in the sequence databases. In addition, no binding partners have been identified for either domain.

Here we report the subcloning and overexpression of Gli-cyt, the intracellular tail of *Drosophila* Gli. To characterize the physicochemical properties of the cytoplasmic domain of Gli, a series of biophysical studies were subsequently performed. As these studies progressed, it became increasingly obvious to us that Gli-cyt is a member of the growing family of natively unfolded proteins, namely, intrinsically disordered proteins that, when purified to homogeneity, have been shown to exist in an extended conformation, under physiologic conditions in vitro, with little or no ordered structure and high intramolecular flexibility.¹⁴⁻²¹

MATERIALS AND METHODS

Construction of the Gli-cyt Expression Vector

The portion of the Gli gene encoding the intracellular domain (as described¹³) was amplified by polymerase chain reaction (PCR) with the use of Gli cDNA cloned into pUAST as template, and the primers listed in Table I. The PCR product was purified, digested with *NdeI* and *HindIII*, and ligated between the *NdeI* and *HindIII* sites of the pET21b (Novagen) expression vector to yield pET21Glicyt. The sequence of the Gli coding region was confirmed by nucleotide sequencing.

Protein Expression and Purification

Escherichia coli BLR(DE3) transformed with pET21Glicyt was grown to saturation overnight at 37°C in 2 × YT (yeast extract trypton) medium containing ampicillin (100 µg/mL) and tetracycline (12.5 µg/mL). An aliquot of overnight culture was diluted 1/50 in 2 × YT medium containing ampicillin (100 µg/mL) and tetracycline (12.5 µg/mL), and grown at 37°C to an OD₆₀₀ of 0.7, at which time expression was induced by addition of isopropyl-β-D-galactoside (IPTG) to a final concentration of 0.5 mM. Cells were incubated for a further 3 h at 37°C before harvesting by centrifugation (7000 g, 15 min, 4°C), and stored at -80°C.

Cells were resuspended in buffer A (50 mM phosphate, 500 mM NaCl, pH 8.0) supplemented with 2 mM imidazole and 1% protease inhibitor cocktail (Sigma), and lysed by sonication (three 30-s pulses of 1700 J, with 1 min intervals). The supernatant was clarified by centrifugation (21,000 g, 30 min, 4°C).

Ni-NTA (nitrilotriacetic acid) beads (10 mL of 50% slurry, Qiagen), preequilibrated in buffer A + 2 mM imidazole, were incubated with the soluble cell extract (1 h, 4°C, with gentle mixing). The loaded beads were packed into a column (1.5 × 10 cm) and washed with 10 column volumes (CVs) of buffer A + 10 mM imidazole + 1% protease inhibitor cocktail. The recombinant protein eluted in 5 CV of buffer A + 100 mM imidazole, and the purified protein was subjected to buffer exchange against Dulbecco's phosphate-buffered saline (PBS) (10 mM phosphate, 137 mM NaCl, 3 mM KCl, pH 7.4) with the use of a Vivaspinn concentrator. We estimated the protein concentration from

TABLE I. Oligonucleotides Used for Cloning Gli-cyt Into pET21b^a

Name	Sequence
gliCt-5'	5'-TAATATCATATGATCATGTGGCGCAATGCGAAG
gliCt-3'	5'-TAATATAAGCTTTTAATGATGATGATGATGATGTACGGATGTCTG

^aThe restriction sites are underlined.

the absorbance at 280 nm, using a molar extinction coefficient of $9650 \text{ M}^{-1} \text{ cm}^{-1}$ (estimated with ExPASy proteomics tools²²).

Biophysical Characterization

Determination of mass and peptide mapping by matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF)

MS determinations were performed by the Weizmann Institute Biological Mass Spectrometry Service.

Determination of the Stokes radius by size exclusion chromatography (SEC)

We performed gel filtration on an analytic size exclusion column (Superdex75, Amersham Pharmacia Biotech) equilibrated with $2 \times \text{PBS}$ (Gibco), using an AKTA fast performance liquid chromatography (FPLC) system (Amersham Pharmacia Biotech). The elution volume (V_e) was monitored by absorbance at 280 nm. V_e for a particular molecular species was then converted to K_{av} by the following equation:

$$K_{av} = \frac{V_e - V_0}{V_t - V_0},$$

where V_0 is the exclusion volume, taken as the elution volume of dextran blue, and V_t is the total bed volume. We estimated the Stokes radius (R_{ST}) of Gli-cyt according to the method of Laurent and Killander (as modified by Siegel and Monty²³), using a linear calibration plot of R_{ST} versus $(-\log K_{av})^{1/2}$ obtained with standard globular proteins that included bovine pancreatic ribonuclease A ($R_{ST} = 16.4 \text{ \AA}$), bovine chymotrypsinogen A ($R_{ST} = 20.9 \text{ \AA}$), ovalbumin ($R_{ST} = 30.5 \text{ \AA}$), and bovine serum albumin ($R_{ST} = 35.5 \text{ \AA}$).

Circular dichroism (CD) spectroscopy

CD spectra were recorded on an AVIV Model 202 CD spectrometer. Far-UV spectra were recorded in a 0.1-cm pathlength cell and near-UV spectra in a 1-cm pathlength cell. All CD spectra were recorded in 10 mM Tris, 30 mM NaCl, pH 7.5, at 25°C, with a step size of 1 nm and an averaging time of 5 s. Each spectrum shown is derived from three spectra accumulated, averaged, and corrected for the contributions of the buffer.

¹H-NMR spectroscopy

NMR spectra were collected in 95% H₂O/5% D₂O. Data were acquired on a Bruker 500 MHz spectrometer and analyzed with the Bruker XWINNMR software. 1D-¹H spectra were collected at 30°C. For water suppression, we used the WATERGATE sequence (water suppression by

gradient-tailored excitation).²⁴ For resolution enhancement, we applied the sine-square window function to the free induction decay (FID).

Bioinformatic Characterization

Calculation of the mean net charge $\langle R \rangle$, mean hydrophobicity $\langle H \rangle$, and “folding index”

Based on the approach of Uversky et al.,¹⁷ the mean net charge of Gli was defined as the absolute value of the difference between the numbers of positively and negatively charged residues at pH 7.0 divided by the total number of residues. The mean hydrophobicity was calculated by the method of Kyte and Doolittle.²⁵

In addition, we found it useful to combine $\langle R \rangle$ and $\langle H \rangle$ into a single value that we have named the “folding index” (I_F). This index was derived by a simple rearrangement of the equation $\langle R \rangle = 2.785\langle H \rangle - 1.151$, determined by Uversky et al.¹⁷ to describe the boundary between the natively unfolded and natively folded regions in charge/hydrophobicity phase space:

$$I_F = 2.785\langle H \rangle - \langle R \rangle - 1.151$$

Thus, the boundary is now set as the x axis, with the folding index (I_F) being the distance (a measure combining the hydrophobicity and charge) and direction (positive is natively folded, negative is natively unfolded) normal to the axis (see the FoldIndex© server <http://bioportal.weizmann.ac.il/fldbin/findex>).

PONDR prediction of unstructured regions

We submitted the Gli sequence to the PONDR server (<http://www.pondr.com>) using the default integrated predictor VL-XT.^{26,27} Access to PONDR was provided by Molecular Kinetics (P.O. Box 2475 CS, Pullman, WA 99165-2475; E-mail: mhungerford@molecularkinetics.com) under license from the WSU Research Foundation. PONDR was copyrighted in 1999 by the WSU Research Foundation, all rights reserved.

RESULTS

Cloning, Expression, and Purification

The intracellular portion of Gli (Gli-cyt) was subcloned into pET21b, with removal of the T7-tag and addition of an in-frame C-terminal His6-tag. The transmembrane (TM) helix prediction algorithm TMHMM corroborated the previous TM helix prediction.¹³ Three residues of the TM helix were included in the construct on the chance that they might be necessary for folding. Overexpressed Gli-cyt was produced in the cytosolic fraction of *E. coli* BLR(DE3) at >40 mg/L. Purification to homogeneity was accomplished in a single pass through a Ni-NTA column, under non-denaturing conditions.

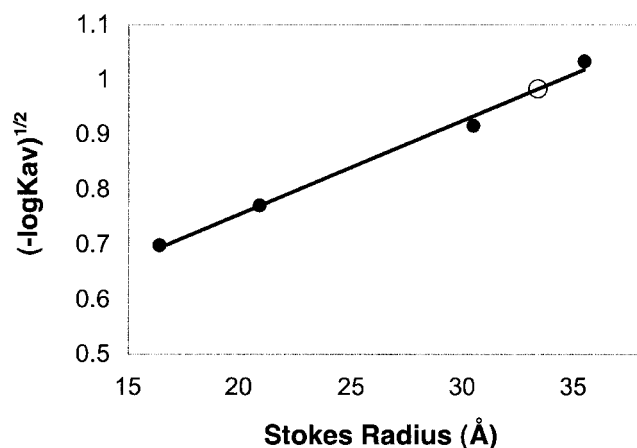


Fig. 2. Stokes radius of Gli-cyt determined by gel filtration chromatography. The open circle represents Gli-cyt. The protein standards are ribonuclease A ($R_{ST} = 16.4$ Å), chymotrypsinogen A ($R_{ST} = 20.9$ Å), ovalbumin ($R_{ST} = 30.5$ Å), and bovine serum albumin ($R_{ST} = 35.5$ Å).

The molecular weight (MW) of Gli-cyt was found to be $[M+H^+] = 23,564$ Da by MALDI mass spectrometry (MS) (data not shown), consistent with the theoretical MW of 23,579 Da. Peptide mapping, also using MALDI MS, matched 58% of the tryptic fragments to Gli-cyt (data not shown). However, on sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), under reducing conditions, the protein exhibited an apparent MW of ~ 30 kDa (data not shown). This anomaly may be attributed to abnormal binding of SDS by Gli-cyt, which is a rather basic protein (estimated $pI = 9.9$, based on amino acid composition with the use of ExPASy proteomics tools). Similarly abnormal mobility has been demonstrated for other highly charged proteins.^{14,28}

Biophysical Characterization

Gel filtration suggests a structure with low compactness

In the course of developing a purification procedure for Gli-cyt, we noted that it eluted with an anomalously low exclusion volume from a gel filtration column, if it was assumed to be a compact globular protein. The mobility of particles on SEC depends on both their size and their shape. Thus, larger or extended proteins elute at lower exclusion volumes than smaller and compact ones. To quantify the unusual mobility of Gli-cyt on SEC, we estimated its Stokes radius (R_{ST}) by comparing its mobility on a Superdex75 column to those of a series of standard proteins (Fig. 2). Gli-cyt eluted with an effective R_{ST} of 33.4 Å.

It was shown earlier that there is a linear correlation of $\log R_{ST}$ and $\log M$ (the logarithm of the molecular mass) for each of four principal folding states for globular proteins, namely, native (ordered), molten globule (MG), premolten globule (PMG), and denatured.¹⁸ Furthermore, it was shown that natively unfolded proteins can, likewise, be divided into two classes: PMG-like and random coil. To better characterize the hydrodynamic properties of Gli-cyt, we calculated the R_{ST} for a 23.5-kDa polypeptide in each

TABLE II. Calculated Stokes Radius (R_{ST}) for a 23.5-kDa Polypeptide in Various Conformational States Using the Equations of Uversky¹⁸

	Conformational state	Calculated R_{ST}
Globular	Native	22.75
	MG	25.55
	PMG	31.92
	Denaturant unfolded	44.80
Natively unfolded	PMG	33.35
	Random coil	40.24

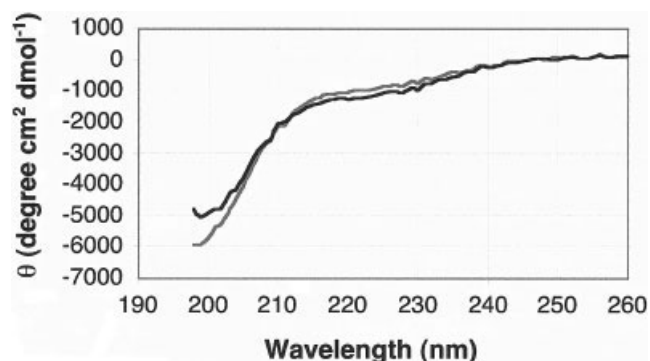


Fig. 3. CD spectrum of Gli-cyt in the far UV. The gray line was measured at 25°C, and the black is the spectrum measured at 50°C. The spectra are dominated by the contribution from random coil components.

conformational state using the equations of Uversky¹⁸ (Table II). The experimentally determined R_{ST} for Gli-cyt, 33.4 Å, is significantly larger than expected for either the native or MG states. However, it is significantly lower than expected for a random coil, suggesting that it may be in a PMG state.

Spectroscopic measurements indicate that Gli-cyt is unstructured

Far-UV CD spectra of polypeptides with extensive α -helical structure have characteristic minima near 208 and 222 nm, β -sheet structures that yield a minimum at 215 nm, and random coil structures characterized by a negative peak near 200 nm and by low ellipticity at 222 nm.²⁹ The far-UV CD spectrum of Gli-cyt (Fig. 3) displays a maximum negative ellipticity at ~ 200 nm and a low ellipticity at 222 nm, which are indicative of a strong contribution from disordered structural elements. The CD spectra of Gli-cyt at 25 and 50°C were almost identical, demonstrating that mild heating does not affect the conformation. The CD spectrum was devoid of any significant ellipticity in the near UV (330–250 nm) at concentrations up to 130 μ M (data not shown), indicating the absence of oriented aromatic residues, which are typically present in the hydrophobic cores of globular proteins.

NMR spectroscopy provided unequivocal evidence that Gli-cyt is unstructured. The proton chemical shift (δ) in the 1D-¹H NMR spectrum is a measure of the resonance frequency of a given proton. It depends on the chemical

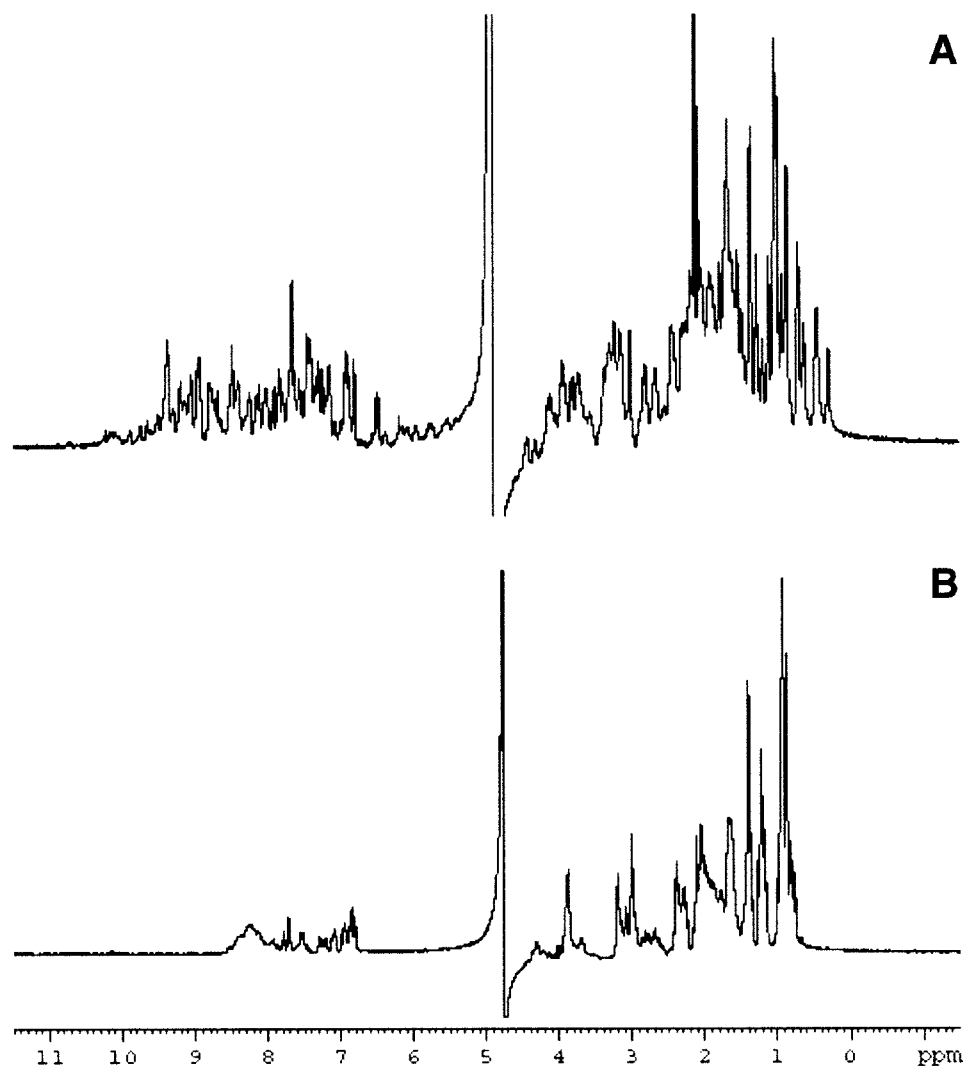


Fig. 4. 1D- ^1H NMR spectrum of (A) a 9-kDa complex of α -bungarotoxin with a 13-mer peptide; (B) Gli-cyt. The folded structure of the complex results in an expanded NMR spectrum, whereas the disordered nature of Gli-cyt produces a compressed spectrum characteristic of a random coil polypeptide.

environment of this proton and is primarily related to the chemical structure of the studied molecule. High electron density around a proton will cause a shielding effect, resulting in a shift of the resonance to a higher field (i.e., toward smaller values of δ), and low electron density will cause a deshielding effect. In folded proteins, protons are under the influence of such shielding and deshielding effects of not only the atoms bonded to them but also adjacent atoms. The outcome of these multiple influences is a spectrum displaying a wide dispersion of chemical shifts.³⁰ The chemical shifts for the common amino acid residues in "random coil" polypeptides serve as a reference for qualitative studies of the general features of a protein spectrum. Figure 4(a) shows a typical spectrum of a folded protein. The spectrum is of a complex of the snake venom three-finger neurotoxin, α -bungarotoxin, with a high-affinity 13-mer polypeptide.³¹ α -Bungarotoxin contains 74 amino acid residues and has a compact structure, with >40%

β -structure and five intrachain disulfide bridges. The change in the chemical environment as a result of folding causes dispersion of the chemical shift from "random coil" values. When a protein is in an extended conformation, any given proton is under the influence of only the atoms bonded to it, and will thus give rise to a spectrum similar to that of noninteracting individual amino acids, a summation of the spectrum of the individual residues, with relatively narrow dispersion. In the 1D ^1H -NMR spectrum of Gli-cyt in H_2O [Fig. 4(b)], the chemical shifts are in the narrow range between 0.8 and 8.5 ppm, demonstrating that Gli-cyt exists in solution as a mixture of rapidly equilibrating extended conformers. Altering the pH (5.0, 7.5, 10.0; data not shown) had no effect on the NMR spectrum, nor did addition of NaCl (100 mM), MgSO_4 (1 mM, 10 mM), or CaCl_2 (1 mM, 10 mM) (data not shown). Thus, the conformation cannot readily be stabilized by nonspecific ionic interactions.

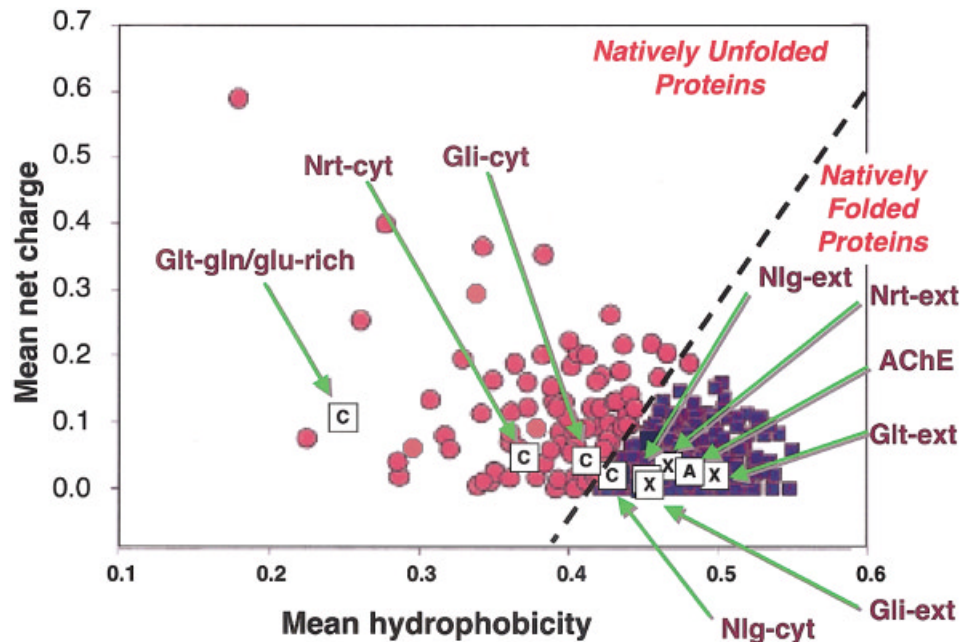


Fig. 5. Charge-hydrophobicity phase space separation between natively folded and natively unfolded proteins (adapted from Uversky et al.¹⁷). Superimposed are the data for the *Drosophila* CLAM intracellular and extracellular domains, and for *Drosophila* AChE. In addition, for the extracellular CLAM, Gli, division was made between the ChE-like domain and the Glu/Gln-rich domain. Gli-cyt and Nrt-cyt fall within the natively unfolded section, as does the Glu/Gln-rich domain of Gli. Nlg-cyt, however, lies just within the folded section. Interestingly, only the Nlg-cyt domain of rat (not shown) falls within natively unfolded space.

Gli-cyt Has Sequence Characteristics Common to Natively Unfolded Proteins

Natively unfolded proteins occupy a well-defined region of the charge-hydrophobicity phase space.^{17,18} To locate Gli-cyt within this phase space, its hydrophobicity²⁵ and net charge were calculated. In addition, for several of the known CLAMs, similar calculations were performed both for the extracellular ChE-like domain and for the intracellular domain (Fig. 5). All extracellular domains fall within the “folded” region of charge-hydrophobicity phase space. In addition to Gli-cyt, Nrt-cyt and one of the isoforms of rat Nlg-cyt fall within the “unstructured” phase space, whereas the cytoplasmic domains of human and *Drosophila* Nlg fall within the “folded” phase space. Although unlikely, an error in prediction of the TM-helix might result in the cloning and expression of a protein that is aberrantly folded compared to its native state. However, further analysis of the TM-cytoplasmic boundary with use of the Uversky et al.¹⁷ method suggested that an error in TM-helix prediction of 9 residues would have been necessary to potentially change the intrinsically unfolded state of the protein (Fig. 6).

The low hydrophobicity and high net charge of natively unfolded proteins result in a difference in amino acid composition between them and natively folded proteins. Compared to sequences of ordered proteins, disordered protein sequences are substantially depleted in I, L, V, W, F, Y, and C, which were therefore designated as “order-promoting” amino acids, and enriched in E, K, R, G, Q, S, P, and A, which have been designated as “disorder-

promoting.”²⁷ For example, Gli-cyt contains only three tyrosines (Y), one tryptophan (W), and no cysteines (C). Overall, 18.1% of the Gli-cyt amino acid sequence is composed of “order-promoting” amino acids, compared to an average of 29.3% for all proteins in the *Drosophila* proteome (Fig. 7, dark bars). Conversely, 58.6% of the Gli-cyt sequence is composed of residues from the group of “disorder-promoting” amino acids, compared to an average of 50.2% for all *Drosophila* proteins (Fig. 7, light bars). Thus, the amino acid composition of Gli-cyt also suggests that it has a natively unfolded structure.

Another tool for predicting disordered regions in proteins is a collection of neural network predictors of naturally disordered regions (PONDR).³² These predictors utilize several amino acid attributes, such as coordination number, hydrophobicity, or flexibility, and they are trained on experimentally reported disordered proteins or regions of proteins. The predictor VL-XT was applied to the entire Gli sequence. Figure 8 shows that whereas Gli-ext is predominantly folded, Gli-cyt is predicted to have three disordered stretches: residues 764–781, 810–840, and 853–950, with respective strengths of 0.74, 0.86, and 0.89. These three stretches account for 71% of the C-terminal sequence.

DISCUSSION

Although members of the CLAM family have been studied for just over a decade,^{9,13,33–37} and an understanding of their roles in neural development is beginning to emerge, virtually no structural studies have been per-

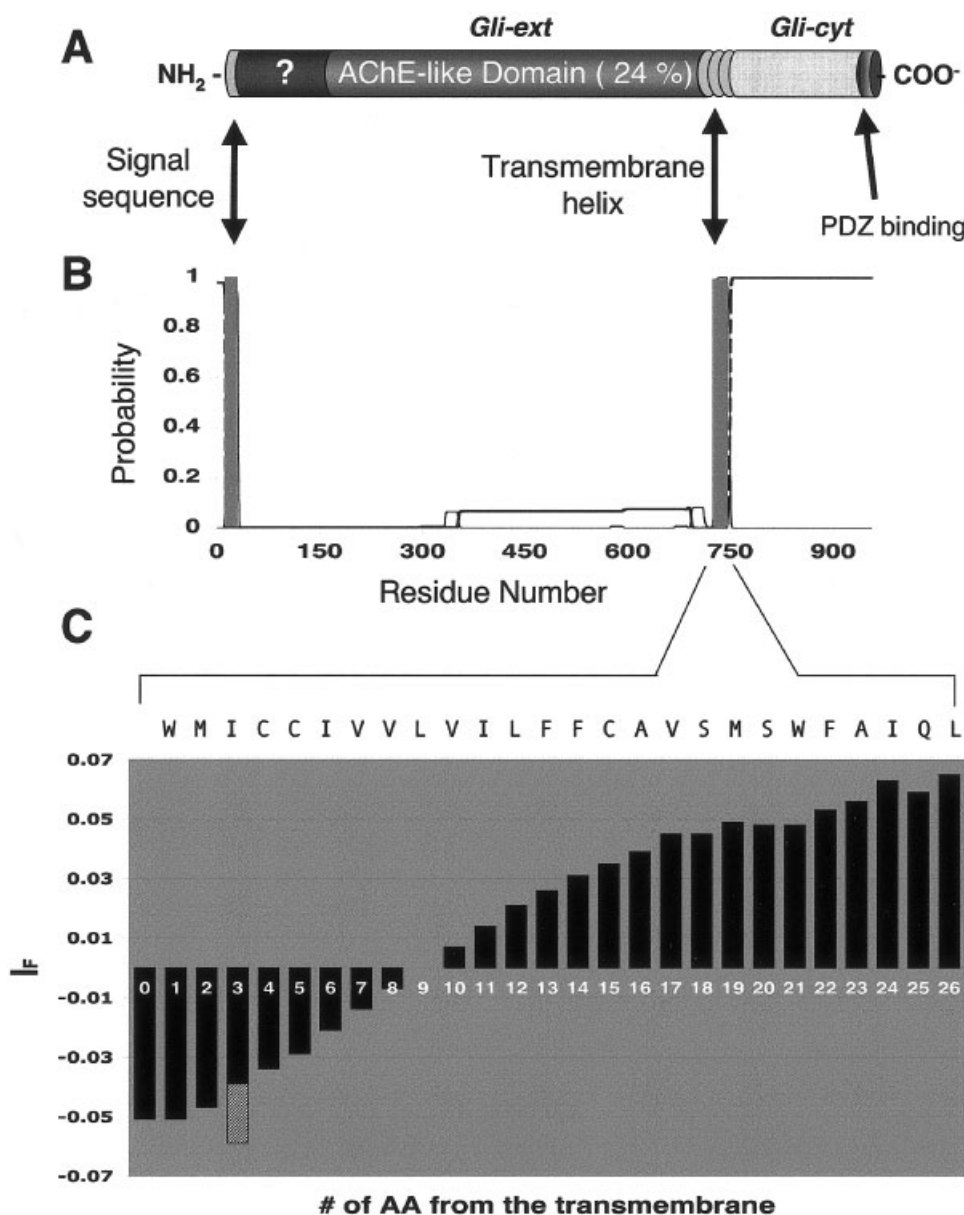


Fig. 6. (A) Schematic representation of *Drosophila* Gli domain organization. The extracellular portion of the Gli precursor contains a 39-residue secretion sequence, followed by a 140-residue uncharacterized domain and a 70-kDa ChE-like domain. The intracellular portion is 207 residues long, and the 4 amino acids at the COOH-terminus (QTSV) form a PDZ-binding motif. The extracellular and intracellular regions are separated by a 26-residue membrane-spanning helix; (B) Transmembrane hidden Markov model (TMHMM) analysis of Gli.^{46,47} The predicted signal sequence (residues 1–39) and the TM sequence (residues 724–749) are represented by shaded bars. The thin line represents the probability of a region being intracellular (i.e., residues after 749 have a probability of 1); (C) Effect of sequential addition of residues from the transmembrane sequence on the folding index (I_f) calculated for the cytoplasmic domain. The bar with the hatched segment represents the folding index of our His6-tagged construct.

formed on these proteins. Considering their vital role in early development, with respect to both the extracellular and intracellular portions,^{4,11} and their potentially important relationship with AChE, we felt it crucial that such studies should be undertaken. Our results, presented here, represent the first such study on any CLAM.

Biophysical measurements were initiated to characterize the structure of Gli-cyt in preparation for X-ray crystal-

lographic analysis. It was demonstrated by gel filtration that Gli-cyt possesses an extended conformation with a hydrodynamic radius substantially greater than would be expected for a 23.5-kDa globular protein. In addition, its CD spectrum in the far UV revealed that its secondary structure is dominated by a random coil contribution, whereas the CD spectrum in the near UV is devoid of ellipticity, supporting the absence of any tertiary struc-

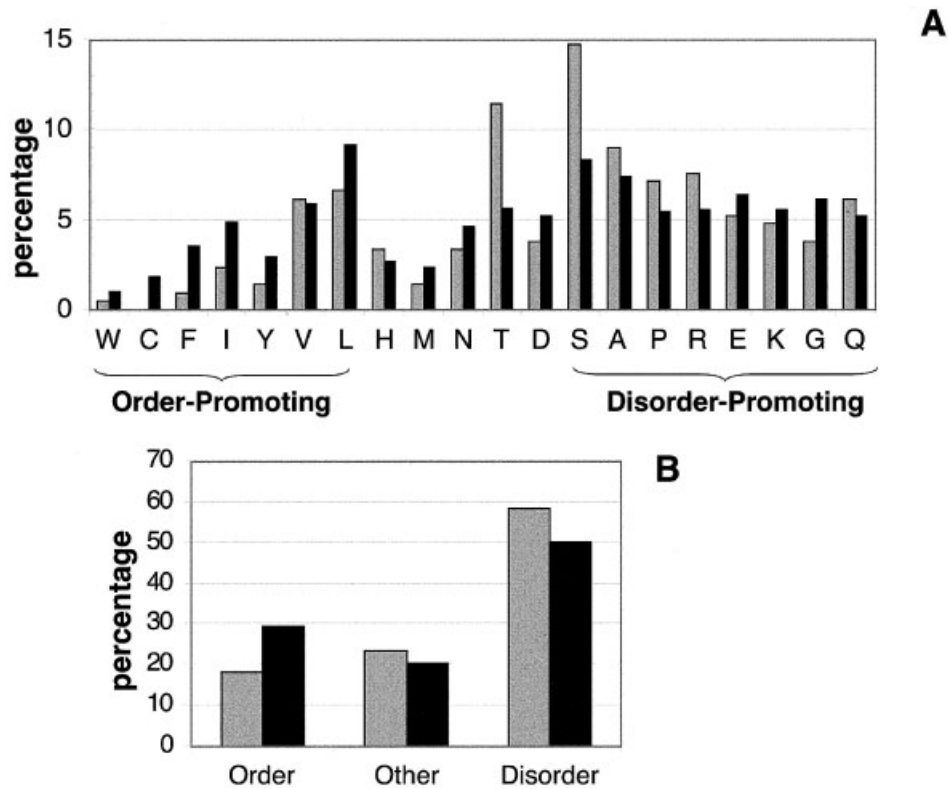


Fig. 7. Content of order-promoting and disorder-promoting amino acids in the *Drosophila* proteome and in Gli-cyt. (A) Overall amino acid percentage composition of the *Drosophila* proteome (black bars) and Gli-cyt (gray bars). (B) Relative percentage content in the *Drosophila* proteome and in Gli-cyt of order-promoting, disorder-promoting, and other amino acids.

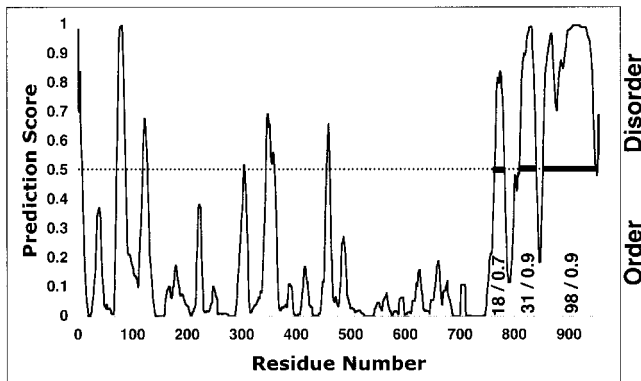


Fig. 8. PONDR prediction of unstructured regions in Gli with use of the VL-XT predictor. The prediction score is plotted against residue number. The threshold, 0.5, is marked by a broken line; residues with a higher score are considered disordered.³² The three strongest disordered regions are marked by thick bars, and the length and strength values for each are indicated.

ture. Finally, the narrow distribution of chemical shifts in the 1D-¹H NMR spectrum provided definitive evidence for the disordered nature of Gli-cyt. The results of these studies, in conjunction with high net charge and modest hydrophobic character, led us to conclude that Gli-cyt belongs to the family of natively unfolded proteins.

Led by the assumption that "since amino acid sequence determines 3-D structure, amino acid sequence should also

determine lack of 3-D structure,"²¹ specific sequence features shared by natively unfolded proteins have been evaluated and algorithms for their identification formulated.^{17,27} Natively unfolded proteins occupy the low hydrophobicity/high net-charge portion of charge-hydrophobicity phase space (Fig. 5), and have low sequence complexity relative to ordered proteins.^{17,27} Consistent with these characteristics, Gli-cyt has a very low frequency of hydrophobic amino acids (W, Y, I, F, L) in comparison to their frequency in the *Drosophila* proteome (Fig. 7). The underrepresentation of hydrophobic amino acids in a protein diminishes one of the basic thermodynamic forces known to be important for protein folding, namely, the hydrophobic interaction.³⁸ Because a hydrophobic core does not form, such proteins have large hydrodynamic dimensions, as indeed was shown to be the case for Gli-cyt. A high net charge (+7 for Gli-cyt), which is a complementary characteristic to the low hydrophobicity, also contributes to the unfolded nature through electrostatic repulsion, while ensuring that these proteins are not likely to self-aggregate, even though unfolded.

Several researchers have recently attempted to understand what environmental factors may lead natively unfolded proteins to fold. Induction of folding through alteration of the pH,³⁹ addition of alcohols such as trifluoroethanol (TFE), osmolytes such as trimethylamine *N*-oxide (TMAO),^{40,41} or crowding agents such as dextrans

and Ficoll 70⁴² have met with moderate success. It is also conceivable that membrane association or posttranslational modification(s) could induce folding. In our own preliminary trials, the disordered nature of Gli-cyt was found to be unaltered over a range of pH values and by addition of a number of salts. Undoubtedly, the folding of the unfolded domain into a biologically relevant conformation requires conditions similar to those the protein experiences *in vivo*. Thus, although the biologic relevance of folded conformations induced by alcohols or osmolytes might be questioned, there would be no argument with respect to a folded conformation induced through binding of a functional ligand or partner protein.

As already mentioned, no direct evidence is yet available concerning the putative extracellular and intracellular partners for Gli. However, Nrt, together with its extracellular partner amalgam, has been the object of a series of deletion and domain replacement experiments.^{3,4} These studies have demonstrated not only that approximately half of the ChE-like domain is required for the adhesive function of Nrt, but, surprisingly, that the entire cytoplasmic portion is also required. The intercellular interactions of Nlg, which with respect to both sequence and function are a much closer relative of Gli than Nrt,⁴³ have been studied in the most detail of any CLAM. Nlg, like Gli, possesses a C-terminal PDZ-binding motif. Recent studies suggest that the extracellular interactions of Nlg with neurexin (a TM protein that also possesses a PDZ binding motif), and its intercellular interaction with S-SCAM, may both be crucial for formation of glutamatergic synapses during development, in which a series of such interactions might "crystallize" the synaptic junction.¹¹ Although the extracellular partner of Gli is unknown, recent studies (V. Auld, unpublished) have suggested two possible intracellular partners for Gli. Both discs-large (DLG) and scribbled, each possessing multi-PDZ domains, are known to localize to the same septate junction regions as Gli.^{44,45} Thus, it is not unreasonable to suggest a similar developmental role for Gli at septate junctions as for Nlg at glutamatergic synapses. In addition, because the PDZ-binding motif is at the C-termini of both Nlg and Gli, it is not difficult to envisage that the high flexibility, large hydrodynamic radius, and binding plasticity of natively unfolded Gli-cyt, and by analogy, of the corresponding region of Nlg, provide them with an advantage for rapidly finding and interacting with their functional or structural partners.

In addition to providing the first biophysical characterization of a CLAM, the data presented here may also be relevant to the consideration of other CLAMs, as hinted in the preceding paragraph. The location in charge-hydrophobicity phase space of other CLAM intracellular domains suggests that intrinsic disorder may be a shared property. For example, the cytoplasmic portions of rat Nlg 1 and *Drosophila* Nrt occupy the natively unfolded section of hydrophobicity/net charge phase space (Fig. 5). We intend to address this question in recently initiated studies.

ACKNOWLEDGMENTS

Our thanks to Dr. Tali Scherf for her assistance in determining the 1D-¹H NMR spectra, to Dr. Clifford Felder for calculating the protein charge and hydrophobicity values, and to Dr. Kurt Giles for valuable discussions.

REFERENCES

- Soreq H, Seidman S. Acetylcholinesterase—new roles for an old actor. *Nat Rev Neurosci* 2001;2:294–302.
- Botti SA, Felder CE, Sussman JL, Silman I. Electrotactins: A class of adhesion proteins with conserved electrostatic and structural motifs. *Protein Eng* 1998;11:415–420.
- Darboux I, Barthalay Y, Piovant M, Hipeau-Jacquotte R. The structure–function relationships in *Drosophila* neurotactin show that cholinesterase domains may have adhesive properties. *EMBO J* 1996;15:4835–4843.
- Fremion F, Darboux I, Diano M, Hipeau-Jacquotte R, Seeger MA, Piovant M. Amalgam is a ligand for the transmembrane receptor neurotactin and is required for neurotactin-mediated cell adhesion and axon fasciculation in *Drosophila*. *EMBO J* 2000;19:4463–4472.
- Silver A. The biology of cholinesterases. In: Neuberger A, Tatum E, editors. *Frontiers of biology*. Amsterdam: North-Holland; 1974.
- Layer PG. Comparative localization of acetylcholinesterase and pseudocholinesterase during morphogenesis of the chicken brain. *Proc Natl Acad Sci U S A* 1983;80:6413–6417.
- Greenfield S. Acetylcholinesterase may have novel functions in the brain. *Trends Neurosci* 1984;7:364–368.
- Genever PG, Birch MA, Brown E, Skerry TM. Osteoblast-derived acetylcholinesterase: A novel mediator of cell-matrix interactions in bone? *Bone* 1999;24:297–303.
- de la Escalera S, Bockamp EO, Moya F, Piovant M, Jimenez F. Characterization and gene cloning of neurotactin, a *Drosophila* transmembrane protein related to cholinesterases. *EMBO J* 1990;9:3593–3601.
- Irie M, Hata Y, Takeuchi M, Ichtchenko K, Toyoda A, Hirao K, Takai Y, Rosahl TW, Sudhof TC. Binding of neuroligins to PSD-95. *Science* 1997;277:1511–1515.
- Scheiffele P, Fan J, Choij J, Fetter R, Serafini T. Neuroligin expressed in nonneuronal cells triggers presynaptic development in contacting axons. *Cell* 2000;101:657–669.
- Rao A, Harms KJ, Craig AM. Neuroligation: Building synapses around the neurexin–neuroligin link. *Nat Neurosci* 2000;3:747–749.
- Auld VJ, Fetter RD, Broadie K, Goodman CS. Gliotactin, a novel transmembrane protein on peripheral glia, is required to form the blood–nerve barrier in *Drosophila*. *Cell* 1995;81:757–767.
- Weinreb PH, Zhen W, Poon AW, Conway KA, Lansbury PT Jr. NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded. *Biochem J* 1996;35:13709–13715.
- Romero P, Obradovic Z, Kissinger CR, Villafranca JE, Garner E, Guillot S, Dunker AK. Thousands of proteins likely to have long disordered regions. *Pac Sympos Biocomput* 1998;437–448.
- Wright PE, Dyson HJ. Intrinsically unstructured proteins: Re-assessing the protein structure–function paradigm. *J Mol Biol* 1999;293:321–331.
- Uversky VN, Gillespie JR, Fink AL. Why are "natively unfolded" proteins unstructured under physiologic conditions? *Proteins* 2000;41:415–427.
- Uversky VN. What does it mean to be natively unfolded? *Eur J Biochem* 2002;269:2–12.
- Uversky VN. Natively unfolded proteins: A point where biology waits for physics. *Protein Sci* 2002;11:739–756.
- Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, Oldfield CJ, Campen AM, Ratliff CM, Hipps KW, Ausio J, Nissen MS, Reeves R, Kang C, Kissinger CR, Bailey RW, Griswold MD, Chiu W, Garner EC, Obradovic Z. Intrinsically disordered protein. *J Mol Graph Model* 2001;19:26–59.
- Dunker AK, Obradovic Z. The protein trinity—linking function and disorder. *Nat Biotechnol* 2001;19:805–806.
- Gill SC, von Hippel PH. Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem* 1989;182:319–326.
- Siegel LM, Monty KJ. Determination of molecular weights and

- frictional ratios of proteins in impure systems by use of gel filtration and density gradient centrifugation: Application to crude preparations of sulfite and hydroxylamine reductases. *Biochim Biophys Acta* 1966;112:346–362.
24. Piotto M, Saudek V, Sklenar V. Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J Biomol NMR* 1992;2:661–665.
 25. Kyte J, Doolittle RF. A simple method for displaying the hydrophobic character of a protein. *J Mol Biol* 1982;157:105–132.
 26. Li X, Romero P, Rani M, Dunker AK, Obradovic Z. Predicting protein disorder for N-, C-, and internal regions. *Genome Inform Ser Workshop* 1999;10:30–40.
 27. Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK. Sequence complexity of disordered protein. *Proteins* 2001;42:38–48.
 28. Hames BD. One-dimensional polyacrylamide gel electrophoresis. In: Hames BD, Rickwood D, editors. *Gel electrophoresis of proteins*. New York: Oxford University Press; 1990. p 1–147.
 29. Adler AJ, Greenfield NJ, Fasman GD. Circular dichroism and optical rotatory dispersion of proteins and polypeptides. *Methods Enzymol* 1973;27:675–735.
 30. Wider G. Structure determination of biological macromolecules in solution using nuclear magnetic resonance spectroscopy. *Biotechniques* 2000;29:1278–1282, 1284–1290, 1292.
 31. Scherf T, Kasher R, Balass M, Fridkin M, Fuchs S, Katchalski-Katzir E. A β -hairpin structure in a 13-mer peptide that binds α -bungarotoxin with high affinity and neutralizes its toxicity. *Proc Natl Acad Sci U S A* 2001;98:6629–6634.
 32. Li X, Obradovic Z, Brown CJ, Garner EC, Dunker AK. Comparing predictors of disordered protein. *Genome Inform Ser Workshop* 2000;11:172–184.
 33. Barthalay Y, Hipeau-Jacquotte R, de la Escalera S, Jimenez F, Piovant M. *Drosophila* neurotactin mediates heterophilic cell adhesion. *EMBO J* 1990;9:3603–3609.
 34. Hortsch M, Patel NH, Bieber AJ, Traquina ZR, Goodman CS. *Drosophila* neurotactin, a surface glycoprotein with homology to serine esterases, is dynamically expressed during embryogenesis. *Development* 1990;110:1327–1340.
 35. Olson PF, Fessler LI, Nelson RE, Sterne RE, Campbell AG, Fessler JH. Glutactin, a novel *Drosophila* basement membrane-related glycoprotein with sequence similarity to serine esterases. *EMBO J* 1990;9:1219–1227.
 36. Krejci E, Duval N, Chatonnet A, Vincens P, Massoulié J. Cholinesterase-like domains in enzymes and structural proteins: Functional and evolutionary relationships and identification of a catalytically essential aspartic acid. *Proc Natl Acad Sci U S A* 1991;88:6647–6651.
 37. Ichtchenko K, Hata Y, Nguyen T, Ullrich B, Missler M, Moomaw C, Sudhof TC. Neuroligin 1: A splice site-specific ligand for beta-neurexins. *Cell* 1995;81:435–443.
 38. Kauzmann W. Some factors in the interpretation of protein denaturation. *Adv Protein Chem* 1959;14:1–63.
 39. Dutta K, Alexandrov A, Huang H, Pascal SM. pH-induced folding of an apoptotic coiled coil. *Protein Sci* 2001;10:2531–2540.
 40. Baskakov I, Bolen DW. Forcing thermodynamically unfolded proteins to fold. *J Biol Chem* 1998;273:4831–4834.
 41. Baskakov IV, Kumar R, Srinivasan G, Ji YS, Bolen DW, Thompson EB. Trimethylamine *N*-oxide-induced cooperative folding of an intrinsically unfolded transcription-activating fragment of human glucocorticoid receptor. *J Biol Chem* 1999;274:10693–10696.
 42. Flaugh SL, Lumb KJ. Effects of macromolecular crowding on the intrinsically disordered proteins c-Fos and p27(Kip1). *Biomacromolecules* 2001;2:538–540.
 43. Gilbert M, Smith J, Roskams AJ, Auld VJ. Neuroligin 3 is a vertebrate gliotactin expressed in the olfactory ensheathing glia, a growth-promoting class of macroglia. *Glia* 2001;34:151–164.
 44. Bilder D, Perrimon N. Localization of apical epithelial determinants by the basolateral PDZ protein Scribble. *Nature* 2000;403:676–680.
 45. Mathew D, Gramates LS, Packard M, Thomas U, Bilder D, Perrimon N, Gorczyca M, Budnik V. Recruitment of scribble to the synaptic scaffolding complex requires GUK-holder, a novel DLG binding protein. *Curr Biol* 2002;12:531–539.
 46. Sonnhammer ELL, von Heijne G, Krogh A. A hidden Markov model for predicting transmembrane helices in protein sequences. In: Glasgow J, Littlejohn T, Major F, Lathrop R, Sankoff D, Sensen C, editors. *Proceedings of the Sixth International Conference on Intelligent Systems for Molecular Biology*. Menlo Park, CA: AAAI Press; 1998. p 175–182.
 47. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J Mol Biol* 2001;305:567–580.