Title
Efficient characterization of protein-protein interfaces involved in cellular signaling

Author
Reese, Michael Lloyd

Publication Date
2006

Peer reviewed|Thesis/dissertation

Efficient Characterization of Protein-Protein Interfaces Involved in cellular signaling

by

Michael Lloyd Reese

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biophysics

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Acknowledgements

The experimental methodology described in chapter I was published in *J Am Chem Soc.* 2003 125(47) pp. 14250-1. The characterization of the clathrin light chain/heavy chain interface appearing in chapter III was published in *EMBO J.* 2002 21(22) pp. 6072-82.

Collaborators and scientific contributions:

Chapter I: The models in Figure 4 were generated using a script developed jointly with Michael Kim of the Šali lab at UCSF.

Chapter II: The pull-downs in Figure 1b and 4a were completed by Srikanth Dakoji of the Bredt lab at UCSF.

Chapter III: Chih-Ying Chen generated the data found in Figures 1-2, 3a, and 4-5. Peter Hwang did the circular dichroism measurements found in Figure 3b-c. Nobuyuki Ota of the Agard lab at UCSF guided me in the generation of the structural model of the HC/LC complex. Sherri Newmyer did the pull-downs depicted in Figure 8b-c.

As my above list of collaborators attests, my experience at UCSF would have been far less successful without its incredibly cooperative environment. When I chose to study here, it was the friendliness and enthusiasm I saw in the other students that drew me most strongly. I would like to thank all of my classmates and colleagues who make it the welcoming, exciting institution it is.

A substantial part of my time as a UCSF student was spent in Frankfurt am Main "visiting" one of my advisors, Volker Dötsch, who moved to Germany in 2003. I was made to feel very welcome there, and value the friendships I made.

My jetsetting graduate lifestyle was made possible by an NSF Graduate Research Fellowship and a travel grant from Boehringer-Ingelheim Fonds.

The maintenance of my sanity was made possible with the help of Julie Ransom, who ensured that I was well supported, both while an expatriate and upon my return. Julie has an equally special knack for the smoothing of bureaucratic ruffles and for the soothing of personal troubles. Thank you.

Abstract

The transient interactions between the various domains that comprise a signaling network provide much of the complexity and adaptability of a biological system. It is, however, difficult to obtain structural information on macromolecular complexes. Furthermore, computational methods are sufficient to provide a reliable structural model only when restrained by experimental data. However, even limited biophysical data can be very time-consuming to collect.

To address this problem, I developed an experimental methodology to efficiently characterize protein-protein interfaces by NMR spectroscopy, described in the first chapter. By combining standard chemical shift perturbation with specific isotopic labeling of samples, the interfaces can be described in terms of residue content, skipping the time-consuming step of spectral assignment. Using these predicted interfaces to restrain computational docking generated models of the complex structure sufficiently accurate to guide further experiments. Additionally, I showed that comparing experimental data on the residue content of the interface with predicted data from a generated model informs on the model's quality.

The above method was initially designed for the study of the molecular scaffold PSD-95, which organizes the neuronal signaling apparatus to accelerate and regulate signal transduction. In chapter 2, work characterizing the interaction between MAP1a and SH3-GK domain of PSD-95 is described. Combining our method with additional NMR restraints, an approximate model of the SH3GK-MAP1a complex was generated and verified by biochemical analysis. This provides the first in-depth study into the

binding mode of the GK with its ligands and suggests manners in which the formation of the complexes may be regulated.

All signaling networks transfer information across membranes. The dynamic composition of signaling molecules in the plasma membrane is mediated by the molecule clathrin. Chapter 3 describes studies of the clathrin regulatory subunit, the light chain. From sparse genetic information an accurate model of the interface between the two subunits was generated. Further studies by NMR of the conformationally dynamic regions of the LC known to be involved in the recruitment of regulatory molecules are also presented.

In summary, I created a generalized methodology for efficient characterization of macromolecular complexes. Combining this with additional biophysical, biochemical, and genetic techniques in the study of various macromolecular complexes involved in neuronal plasticity and membrane trafficking has furthered our understanding of the molecular basis of signal transduction.

Table of Contents

Table of Contents

Table of Figures

Introduction

The transient interactions between the various domains that comprise a signaling network provide much of the complexity and adaptability of a biological system. It is, however, difficult to obtain structural information on macromolecular complexes; to date, only 15% of the structures deposited in the PDB are of complexes, and many of these are not transient interacting partners. Fewer still are structures of transient complexes where both (or all) partners are macromolecular (*i.e.* neither small molecules nor peptides).

Substantial inroads have been made into *de novo* structure prediction of individual domains through the use of various methods relying heavily on primary sequence homology to sequences with known structures(1, 2). Purely computational generation of structural models of complexes has proven more difficult(3), in part due to the dearth of experimental structural data for macromolecular complexes. In all cases, computational methods are sufficient to provide a reliable structural model only when restrained by experimental data. In fact, sparse NMR spectroscopic data has proven quite adaptable to computational docking of protein-protein complexes(3, 4). However, even limited biophysical data are very time-consuming to collect. Analysis of NMR data is especially slowed by requirement of spectral assignment. This is especially difficult as the size of the molecule of interest increases, producing more complicated spectra.

To address this problem, I developed an experimental methodology to efficiently characterize protein-protein interfaces by NMR spectroscopy, described in the first chapter. The method combines standard chemical shift perturbation with specific isotopic labeling of samples. This method enables the efficient identification of

interaction sites without prior backbone assignment of the protein of interest by comparing spectra of samples selectively labeled by amino acid type in free and complexed forms. By identifying the number of amino acids of each labeled type that display some spectral change upon binding, the amino acid content of the interface is determined. A patch with matching amino acid content on the surface of the known structure of the protein thus allows the interfaces to be determined.

In short, the interfaces can be described in terms of residue content, skipping the time-consuming step of spectral assignment. Furthermore, the specific labeling allows us to avoid problems with spectral dispersion common to larger proteins. Using these predicted interfaces to restrain computational docking generated models of the complex structure sufficiently accurate to guide further experiments. Additionally, I showed that comparing experimental data on the residue content of the interface with predicted data from a generated model informs on the model's quality. Using modern NMR experiments such as [specifically ILVM-labeled(5)] methyl-TROSY(6), it is possible that large molecular weight complexes may be accurately modeled by our method with data from a single experiment.

The above method was initially designed for the study of the molecular scaffold PSD-95, which organizes the neuronal signaling apparatus to accelerate and regulate signal transduction. A hallmark of eukaryotic signal transduction is the separation of recognition and catalytic functionality into modular domains. Thus recognition domains have evolved to do just that; such domains as WW, SH2, SH3, and PDZ are small, rigid, and catalytically inactive. In chapter 2, work characterizing the interaction between MAP1a and SH3-Guanylate kinase domain (GK) of PSD-95 is described. The GK is a

unique kind of protein recognition domain, in that its fold has evolved from an enzymatic kinase (now catalytically inactive) to recognize its ligands. Combining our method with additional NMR restraints, an approximate model of the complex was generated and verified by biochemical analysis. This provides the first in-depth study into the binding mode of the GK with its ligands and suggests manners in which the formation of the complexes may be regulated. Specifically, we demonstrate that the GK undergoes a conformational change upon recognition of MAP1a that may be reminiscent of a kinase "clamping down" on its substrate during catalysis.

All signaling networks transfer information across membranes. Endocytosis and trafficking of vesicles from the trans-golgi to the plasma membrane, and thus the dynamic composition of signaling molecules in those membranes, is mediated by the molecule clathrin. Chapter 3 describes studies of the clathrin regulatory subunit, the light chain. From sparse genetic information a model of the interface between the two subunits was generated, which has since been confirmed by electron miscroscopy(7). Further NMR studies of the conformationally dynamic regions of the LC known to be involved in the recruitment of regulatory molecules are also presented.

1. Sali, A. & Blundell, T. L. (1993) *J Mol Biol* **234,** 779-815.
2. Kim, D. E., Chivian, D. & Baker, D. (2004) *Nucleic Acids Res* **32,** W526-31.
3. Bonvin, A. M. (2006) *Curr Opin Struct Biol* **16,** 194-200.
4. Bonvin, A. M., Boelens, R. & Kaptein, R. (2005) *Curr Opin Chem Biol* **9,** 501-8.
5. Tugarinov, V. & Kay, L. E. (2004) *J Biomol NMR* **28,** 165-72.
6. Tugarinov, V., Hwang, P. M., Ollerenshaw, J. E. & Kay, L. E. (2003) *J Am Chem Soc* **125,** 10420-8.
7. Fotin, A., Cheng, Y., Sliz, P., Grigorieff, N., Harrison, S. C., Kirchhausen, T. & Walz, T. (2004) *Nature* **432,** 573-9.

Chapter I – Fast mapping: experimental and computational methodology

Preface

The formation of macromolecular complexes is fundamental to most every biological process. To date, an approximate 15% of the protein structures deposited in the PDB are of heteromultimers, and, of these, a large portion are proteins which are not transient in their associations (*e.g.* antibodies, multi-chain enzymes). It is the transient associations of macromolecules that create the complexity of a biological system(1-3), and it is the structures of these complexes that are often elusive. While understanding of the energetics involved in biological interaction is growing, the location of an interface on a given protein structure is difficult to predict(4, 5). Therefore, models of macromolecular complexes created without experimental restraints are often of tenuous value. It is, however, both time-consuming and technically challenging to gain information on transient complexes by most biophysical methods.

To address this problem, we developed an experimental methodology using NMR to identify an interface by its amino acid content. It is important to note that our method requires no spectral assignment, and can therefore be completed orders of magnitude more quickly than standard chemical shift perturbation mapping. Furthermore, we have shown that these unassigned "ambiguous" interfaces provide sufficient information to correctly model protein-protein complexes. In addition, the knowledge of the residue content of an interface can be used to effectively judge the value of a resulting model.

Experimental Methodology

One of the advantages of NMR spectroscopy is its ability to efficiently map interfaces(6, 7). This is based on the fact that the chemical shift of a particular nucleus is a very sensitive function of changes in its chemical environment. A prerequisite for using this chemical shift mapping method, however, is that the assignment of the nuclei showing chemical shift changes be known. While these assignments are a necessary byproduct of structural determination by NMR spectroscopy, they are unavailable for protein structures determined by X-ray crystallography. The development of a method that allows one to map the interface of protein complexes, without knowledge of the chemical shift assignments, would clearly be very useful and could open an avenue toward building structures of complexes in a very efficient way. Here, we report such a method and demonstrate its feasibility on the complex of the PDZ domain of neuronal nitric oxide synthase (nNOS) and the second PDZ domain of PSD-95(8-11).

To map the interface between two proteins, without the knowledge of the assignment, we have used a combinatorial approach. The method is based on preparing several protein samples, each one selectively ^{15}N labeled with one particular amino acid(12, 13). These samples are then used to form a complex with its unlabeled binding partner. By comparing [^{15}N,^{1}H]-HSQC spectra of the selectively labeled free protein and its complex, the number of shifted amino acids of a certain type can be identified. The combined results from several differently labeled samples allow one to define the minimum number of a certain type of amino acid located in the interface. Comparison of this list with the known structure of the protein is then used to identify the interface (**Figure 1**). This approach is further divided into two phases. First, we identify amino

acid types with low abundance in the protein's sequence. The combination of data from experiments with several of these rare residues allows us to identify the location of the binding site. Second, we use more common residues as probes to characterize the extent and shape of the interface.

In the case of the nNOS PDZ domain, we have used the five lysines, four phenylalanines, two histidines, and one tyrosine as initial probes to identify the site of the interface. **Figure 2** shows characteristic [^{15}N, ^{1}H]-HSQC spectra for a titration experiment. As the ^{15}N-lysine labeled nNOS PDZ was titrated with unlabeled PSD-95 PDZ2, two changes in the spectrum occurred. One peak disappeared while another peak appeared, as is characteristic of a complex in slow exchange. Two other peaks exhibited severe spectral broadening, as is characteristic of a complex in intermediate exchange. Finally, that the single shifted peak was not broadened shows that two distinct interfaces exist. To distinguish these two interfaces, we will refer to the binding site exhibiting slow exchange as "primary" and the site exhibiting intermediate exchange as "secondary".

In the nNOS, two pairs of lysines lie in the canonical PDZ domain, separated by one or four amino acids, and a single lysine lies in the -hairpin finger. This constellation makes the lysine in the β-finger the most likely candidate for the primary binding interface because it is likely that more than one significant shift would result should any of the other lysines be involved.

To further investigate the interface, we labeled the nNOS with ^{15}N-tyrosine, histidine, or phenylalanine. The single tyrosine, located in the canonical peptide-binding site, did not shift. Instead, it showed severe line broadening at higher PSD-95 concentrations. The experiment with the histidine-labeled sample revealed that one of the

two histidines shifted and neither histidine peak broadened. The phenylalanine-labeled sample showed two shifted peaks. We also observed significant line broadening of another phenylalanine peak at higher PSD-95 concentrations. These results predict that the primary binding site contains at least one lysine, one histidine, and two phenylalanines, but does not contain the single tyrosine. Thus, comparison with the structure of the nNOS (**Figure 2D**) suggests that the primary binding site includes the β-finger. The second binding site contains two lysines, one tyrosine, no histidines, and one phenylalanine. This constellation exists in the canonical peptide-binding site of the domain.

To further map the extent of the interface, we prepared samples of nNOS individually ^{15}N-labeled with leucine, valine, or isoleucine, which are more abundant in the nNOS sequence. In the titration experiments with these samples, we found four valine peaks that shifted, two isoleucine peaks, and one leucine peak. One leucine, one isoleucine, and one valine are located in the β-finger in the region between His 106 and Lys 118 and would, therefore, be expected to show differences in chemical shift. Isoleucines that are spatially close to the β-finger are found in strand 1 (Ile 16) or strand 4 (Ile 58). This arrangement provides two possibilities for the extended interface, including the β-finger and either strand 1 or strand 4. Leu 57, however, precedes Ile 58. As only one significant shift is observed in the leucine-labeled spectrum, and this leucine is likely the one located in the β-finger, these data suggest that strand 1 is part of the interface. This result is confirmed by the data from the valine-labeled nNOS. While strand 4 and its vicinity are devoid of valines, four are present in and around strand 1. These combined results predict that PSD-95 PDZ2 binds the upper part of the β-finger and strand 1.

Such interpretation is consistent with the crystal structure of the α-syntrophin/nNOS complex(8). nNOS residues in direct contact with syntrophin include 1 His, 1 Leu, 2 Ile, 1 Phe, and 3 Val. The additional 1 Lys and 1 Val are not in direct contact with syntrophin, but lie on the backside of the -finger, and are affected due to rearrangements in packing around the interface. It is therefore important to note, as with all chemical shift perturbation studies, that the region of perturbation will be slightly larger than the region of direct interaction.

In a similar way, we mapped the interface on the PSD-95 PDZ2 domain (**Figure 2E**). We identified the canonical peptide-binding site as the primary interface and the C-terminal tail of the construct as the secondary binding site. That nNOS can interact with PSD-95 through its β-finger while maintaining its competency to bind carboxy-termini with its canonical PDZ-domain has been previously demonstrated(9).

The mapping method described here is a very efficient tool for identifying interfaces if the NMR assignment is not known and is in principle also amenable to automated computer-based search algorithms. Initial tests with a simple surface-scanning script have indeed identified the correct interface. To provide final proof that the identified surface is indeed the interface, one can use a double-selective labeling method to unambiguously identify a specific amino acid of the interface(14). Alternatively, a mutagenesis approach can be used.

One prerequisite of the method is that the amino acid type selective labeling is specific and cross-labeling of other amino acid types can be suppressed. Amino acids that are at the end of biosynthetic pathways (Lys, Arg, His) can be added directly to the minimal medium without any further precautions. For other amino acids, we have, in the

past, used auxotrophic strains (DL39) that suppress any cross-labeling.8 These, however, also reduce the yield of expressed protein. Alternatively, cross-labeling can be suppressed in BL21(DE3) bacteria by adding unlabeled amino acids. For the current study, we have added unlabeled Val and Ile for labeling Leu, unlabeled Leu and Val for labeling Ile, und unlabeled Phe for labeling Tyr (and vice versa). Only for Val could cross-labeling not be suppressed, requiring expression in the DL39 strain.

An additional prerequisite is that the binding event should affect the immediate binding site without inducing major conformational changes. Such changes would lead to chemical shift differences outside of the interface. However, if conformational changes occur, the method proposed here can be combined with the saturation transfer method proposed by Takahashi *et al.*(6) This would require the use of ^{15}N-labeled and deuterated amino acids, which, however, are only recently becoming available.

Molecular Modeling

Assigned NMR chemical shift perturbation studies have previously been shown to provide sufficient information for the automated computational modeling of protein-protein complexes using so-called "ambiguous" distance restraints(15). That is, given an unambiguously determined interface on the surface of each binding partner, an accurate model of the complex can be predicted. The quality of the interfaces we predict from our data is much more ambiguous, as we have made very few, if any, assignments of the spectra. We thus sought to test whether our data could produce sufficient restraints to accurately model protein complexes.

As the HADDOCK protocol(15) drives docking with distance restraints between each residue in each of two interfaces, we found it prone to error in our case, in which the

identity of each residue in the interface was partially incomplete (see **Figure 3a**). We thus chose to identify our interfaces by determining the geometric centroid of the convex hull that enclosed each of the residues we had defined (**Figure 3b**). Using Modeller(16), we defined an upper-limit distance restraint of 5Å between the interface centroids of both subunits of a complex. In order to further orient the complex, we defined additional restraints between the collections of surface residues 5Å from each centroid. Akin to HADDOCK, ambiguous minimum-distance restraints were assigned between the two collections. The subunits' positions were randomized and initial models generated by rigid-body simulated annealing. Residues in either of the subunits within 5Å of the complex interface were refined by torsional dynamics. The resulting models were scored both by residue content and by using the Modeller scoring function using stereochemical, electrostatic, and van der Waals energy terms. To increase the conformational space that we sample, we created an initialization set in which we translated the centroid points between 1-4Å along the protein surface. Each member of the initialization set was subjected to the above protocol.

Results of the docking protocol are summarized in **Figure 4**. In the case of both PDB files 1GGR and 3EZA (the HADDOCK test set), our centroid method allowed us to generate models, starting from the unbound structures, with approximately 2Å RMSD from the deposited structure. By using interface centroids up to 6Å from their correct positions, we were able to demonstrate the robustness of our method; even in highly ambiguously defined cases, the correct solution was determined. Furthermore, the residue content as defined by the 9 amino acids we label experimentally is sufficient to identify the correct solution from decoy models. In short, this demonstrates that our

NMR method can quickly provide structural models of a protein complex sufficient to begin further experimental design.

References

1. Kuriyan, J. & Cowburn, D. (1997) *Annu Rev Biophys Biomol Struct* **26,** 259-88.
2. Cho, W. & Stahelin, R. V. (2005) *Annu Rev Biophys Biomol Struct* **34,** 119-51.
3. Cho, W. (2006) *Sci STKE* **2006,** pe7.
4. Bonvin, A. M. (2006) *Curr Opin Struct Biol* **16,** 194-200.
5. Bonvin, A. M., Boelens, R. & Kaptein, R. (2005) *Curr Opin Chem Biol* **9,** 501-8.
6. Takahashi, H., Nakanishi, T., Kami, K., Arata, Y. & Shimada, I. (2000) *Nat Struct Biol* **7,** 220-3.
7. Foster, M. P., Wuttke, D. S., Clemens, K. R., Jahnke, W., Radhakrishnan, I., Tennant, L., Reymond, M., Chung, J. & Wright, P. E. (1998) *J Biomol NMR* **12,** 51-71.
8. Hillier, B. J., Christopherson, K. S., Prehoda, K. E., Bredt, D. S. & Lim, W. A. (1999) *Science* **284,** 812-5.
9. Harris, B. Z. & Lim, W. A. (2001) *J Cell Sci* **114,** 3219-31.
10. Tochio, H., Zhang, Q., Mandal, P., Li, M. & Zhang, M. (1999) *Nat Struct Biol* **6,** 417-21.
11. Wang, P., Zhang, Q., Tochio, H., Fan, J. S. & Zhang, M. (2000) *Eur J Biochem* **267,** 3116-22.
12. McIntosh, L. P. & Dahlquist, F. W. (1990) *Q Rev Biophys* **23,** 1-38.
13. Ou, H. D., Lai, H. C., Serber, Z. & Dotsch, V. (2001) *J Biomol NMR* **21,** 269-73.
14. Weigelt, J., van Dongen, M., Uppenberg, J., Schultz, J. & Wikstrom, M. (2002) *J Am Chem Soc* **124,** 2446-7.
15. Dominguez, C., Boelens, R. & Bonvin, A. M. (2003) *J Am Chem Soc* **125,** 1731-7.
16. Sali, A. & Blundell, T. L. (1993) *J Mol Biol* **234,** 779-815.

Figure Legends

Figure 1: Schematic representation of the method to map the interface between two proteins described in the text.

Figure 2: (A) shows the [^{15}N,^{1}H]-HSQC spectrum of the free lysine-labeled PDZ domain from nNOS with five peaks, corresponding to the five lysines. (B) shows the spectrum of a sample in which 50% of the nNOS PDZ domain molecules are in complex with the PDZ2 domain from PSD-95. (C) shows a spectrum of a 1:1 complex of both PDZ domains. (D) shows the surface distribution of those amino acids on the nNOS PDZ domain that were used as initial probes in the determination of the interface with PSD-95 PDZ2. His are magenta, Lys are red, Phe are green, and Tyr are gold. (E) shows the surface distribution of those amino acids on the PSD-95 PDZ2 that were used as initial probes in the determination of the interface with nNOS PDZ. Arg are red, His are magenta, Met are cyan, Phe are green, and Tyr are gold. In summary, 2 His, 2 Tyr, 1 Phe, 1 Met, and no Arg were seen to be involved in the primary interface, and 1 His, 1 Tyr, 1 Arg, and neither Phe nor Met were involved in the secondary interface. All spectra were measured with a standard Watergate HSQC sequence on a 500 MHz Bruker Avance instrument equipped with a cryoprobe. For each spectrum, 512 complex points were measured in the acquisition and 50 complex points in the indirect dimension with 16 scans per increment at 25 C. The labeled protein was at a concentration of 0.3 mM in a 10 mM Hepes buffer, pH 7.0.

Figure 3: Demonstration of error due to weighting of clustered restraints in incompletely defined interfaces. A) depicts a complex modeled using ambiguous distance restraints between two interfaces in which the restraints have not been evenly distributed. The red

subunit is improperly drawn towards the cluster of four blue restraints. B) depicts the same interface modeled using the centroids of the interface as the source of the ambiguous distance restraints. Restraints are *de facto* evenly distributed around the centroids.

Figure 4: Summary of docking trials using centroid-based restraints, filtered by residue content. The residue content score is considered the sum of the absolute value of the difference between the expected number of each residue type within 5Å of the interface (*i.e.* for a given residue type, the number of peak shifts upon titration with ligand) and the actual number predicted by the model. Zero is thus the lowest value and highest score. A) The results of the docking of 1POH to 1F3G. The highest scored model (bold italics) is depicted in green, overlaid on the deposited structure, 1GGR, in blue. Average backbone RMSD is 1.5Å. B) The results of the docking of 1POH to 1ZYM. The highest scored model (bold italics) is depicted in green, overlaid on the deposited structure, 3EZA, in blue. Average backbone RMSD is 2.4Å.

Figure 1

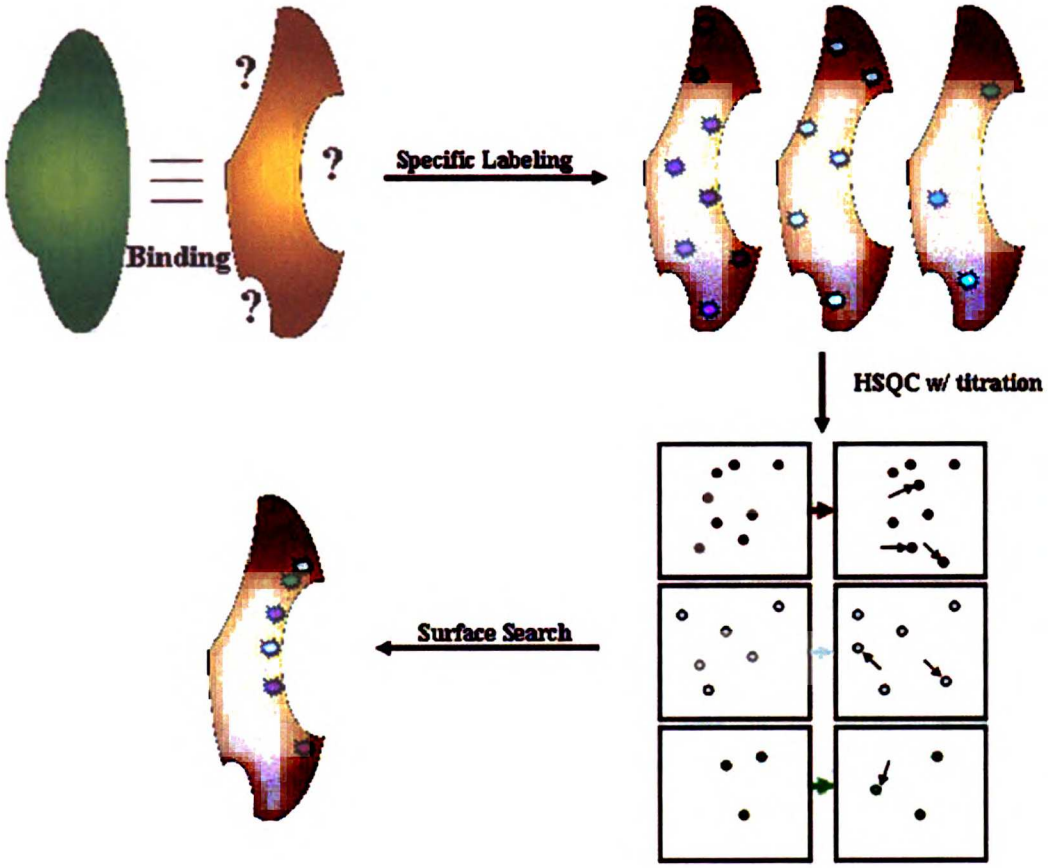

?
?
?
Binding
Specific Labeling
HSQC w/ titration
Surface Search

Figure 2

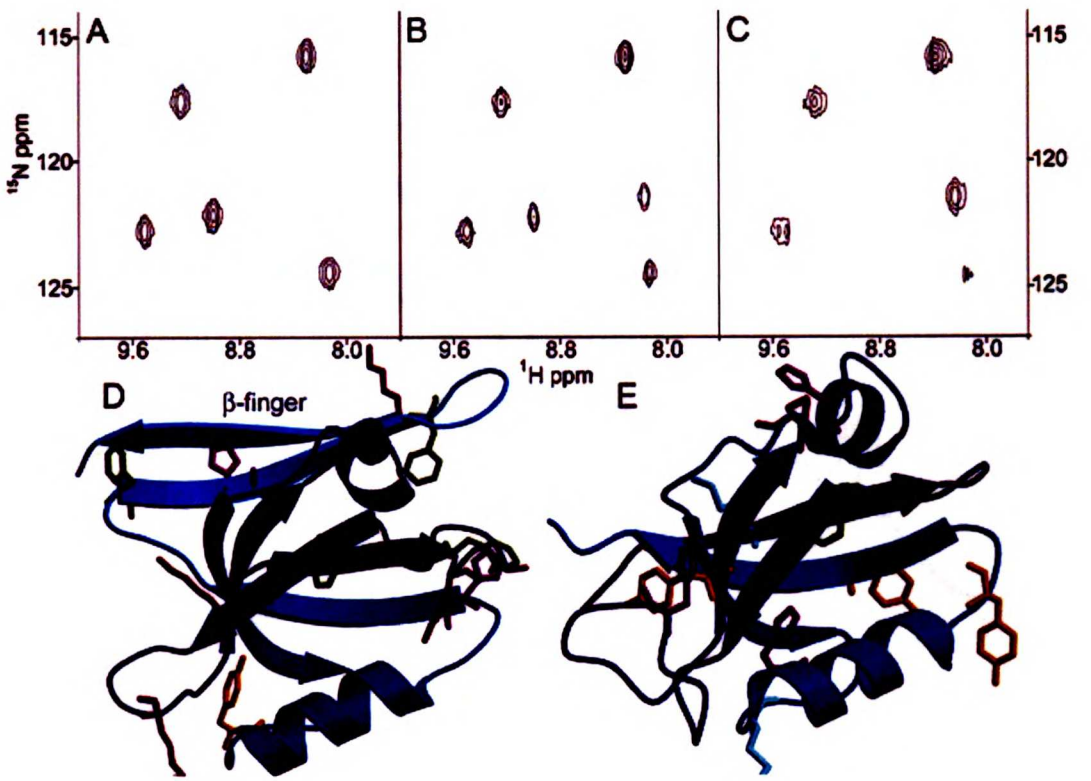


Figure 3

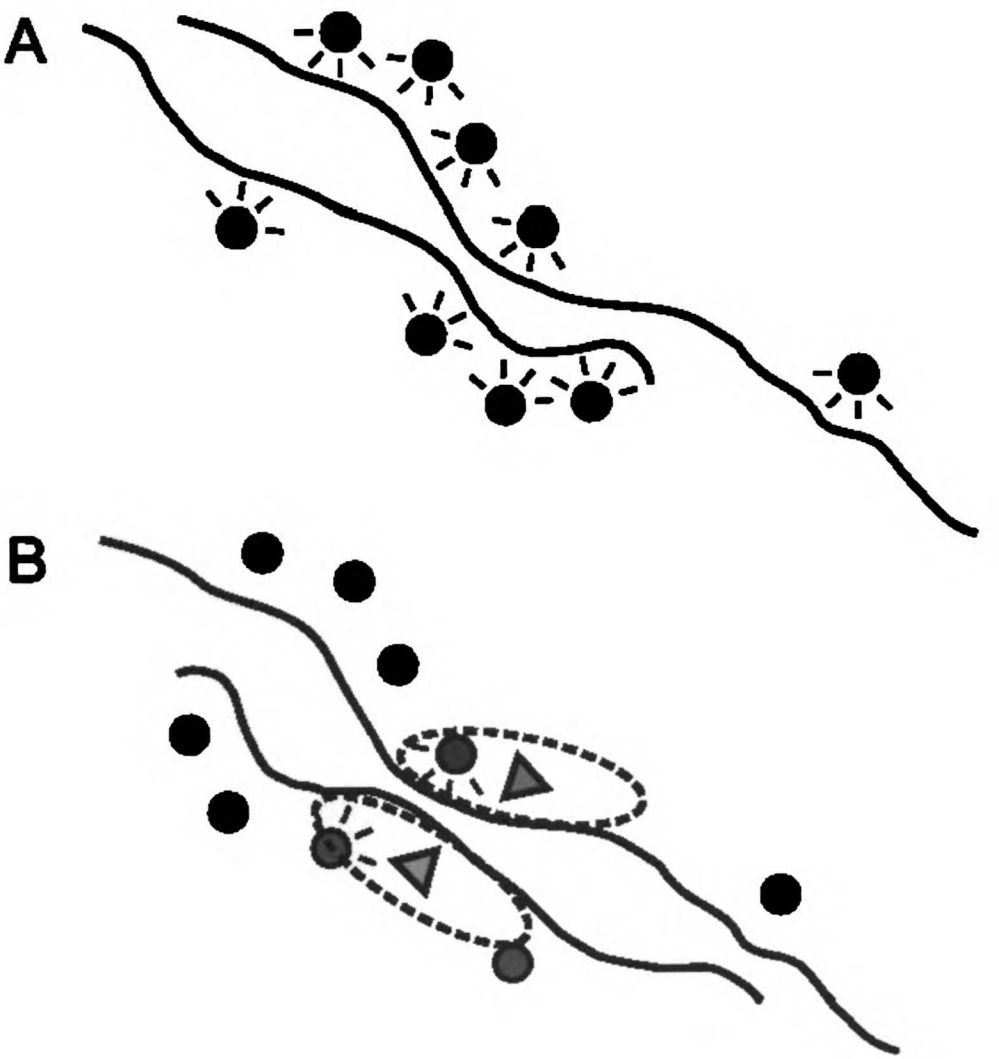
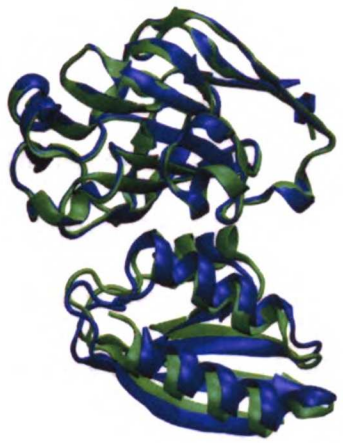


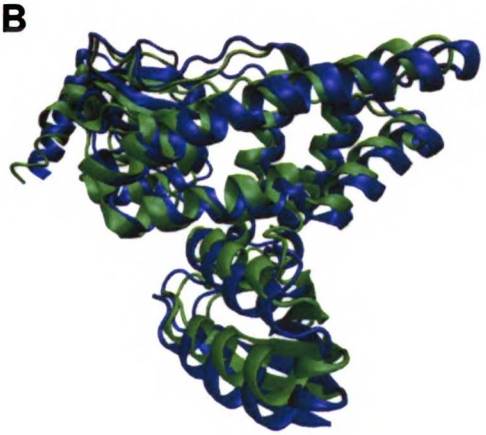
Figure 4

A



OVERLAP	DRMS	SCORE
1.000000	***0.827852***	***1***
1.000000	0.844412	1
0.706383	1.941129	2
0.693617	2.150901	2
0.638298	4.617433	4
0.646809	4.504601	5
0.642553	4.469667	6
0.651064	4.025195	6
0.651064	4.370425	6
0.638298	4.668548	6
0.638298	4.609100	6
0.655319	3.982547	6
0.642553	3.346966	6

B



OVERLAP	DRMS	SCORE
0.957831	***1.818859***	***1***
0.954819	1.856368	2
0.762048	2.649897	2
0.771084	2.115775	2
0.780120	2.109779	3
0.731928	4.541464	4
0.734940	2.951512	6
0.762048	2.132628	7

Chapter II – The guanylate kinase domain of PSD-95 binds dynamically to a conserved motif in MAP1a

Preface

The postsynaptic density protein PSD-95 and related membrane associated guanylate kinases (MAGUKs) are scaffolding proteins, whose modular interaction motifs organize protein complexes at cell junctions. The signature guanylate kinase domain (GK) contains elements of its GMP-binding site, but does not bind nucleotide. Instead, the GK domain has evolved from an enzyme to a protein-protein interaction motif. Here, we show that this canonical GMP-binding region interacts with microtubule associated protein 1a (MAP1a) and we present a structural model. We determine the consensus GK-binding sequence in MAP1a and demonstrate that PSD-95 can use a similar interaction mode to bind diverse protein partners. Furthermore, we show that PSD-95 GK has adopted the conformational flexibility of the ancestral enzyme to bind its varied ligands, which suggests a mechanism of regulation.

Introduction

Eukaryotic signals are often transduced through proteins with separate recognition and catalytic domains. Most protein recognition motifs, such as PTB, SH2, SH3, PDZ, and WW domains, are small and catalytically inactive(1). However, a small number of proteins with enzyme-like folds have evolved to recognize specific protein partners. Several signal transducers with enzyme-like recognition domains localize to the synapse of neurons, and, through these, interact with other signaling proteins, such as CASK, whose CaM kinase-like domain interacts with CaMKII(2), and the voltage-gated calcium channel β-subunit whose SH3-GK domain recognizes of the α-subunit(3). A subset of

these, including the membrane-associated guanylate kinases(4) (MAGUKs), employ multiple recognition motifs in addition to their enzyme-like fold. How these proteins integrate and propagate signals without a catalytic activity is not well understood.

MAGUKs mediate signal transduction by organizing multimolecular complexes at specific membrane sites(4). MAGUKs comprise a series of protein recognition motifs, including N-terminal PDZ domains and C-terminal SH3 and GK domains, the latter of which have been shown to be an integrated folded unit(3, 5, 6) (**Figure 1a**). Genetic screens in multiple species have shown that the SH3-GK domain is critical for MAGUK function as the majority of the deleterious point mutations occur in the SH3-GK(7, 8). Even so, whereas PDZ domains are structurally and functionally well characterized(4, 9), much less is known about the GK domain, in part due to the lack of a known consensus motif for interaction.

In mammalian neurons, a diverse set of MAGUKs organizes the complex signaling apparatus in the synaptic density. These neuronal MAGUKs (PSD-95/SAP-90, PSD-93/Chapsyn-110, hDlg/SAP-97, and SAP-102) are homologous to the *Drosophila* discs large protein (DLG) and localize to both pre- and post-synaptic sites. Two of the best characterized binding partners of the PSD-95 SH3-GK are MAP1a(10) (microtubule associated protein), which helps remodel the microtubule cytoskeleton in an activity-dependent manner(11) and the GKAP(12) (guanylate kinase associated protein) family of proteins, which links the MAGUK and Homer signaling scaffolds(13).

To better understand how signals are integrated at synapses we have investigated the interaction between the SH3-GK domain of PSD-95 (PSD-95 GK) and its ligands MAP1a and GKAP. An increasing body of evidence indicates that conformational

dynamics plays a role in signal transduction, and that important interactions are of relatively low affinity (Kd ~μM)(14-16). Both of these issues often confound standard structural techniques, thus we developed an NMR-based method to map the interface between PSD-95 GK and MAP1a. Our results indicate that the GMP substrate-binding site of PSD-95 GK evolved to recognize these protein ligands, which bind similarly to the nucleotide substrate in enzymatic GKs. Using biochemical methods, we have shown that an acidic residue of MAP1a binds to conserved GK residues that ancestral enzymatic GKs used to coordinate the nucleotide phosphate. Employing both NMR and molecular modeling we have demonstrated that PSD-95 GK uses multiple conformations in its binding. Modulation of the rate of exchange between these conformations may provide a mechanism of regulation adopted by the PSD-95 GK to accommodate distinct synaptic signals.

Results

GK binds a peptide fragment of MAP1a in the GMP-binding site

We created a series of GST-fusion constructs containing fragments of MAP1a and performed pull-downs of purified His_6-tagged PSD-95 SH3-GK (**Figure 1b**). We defined the minimal PSD-95 binding region as a 17 amino acid sequence from 1862-1878. A peptide corresponding to this sequence was used in crystallization trials with SH3-GK, but the complex proved unamenable to crystallization. Furthermore, the spectral dispersion and relaxation characteristics of SH3-GK complicated NMR assignment.

As an alternative strategy to define the mode of PSD-95:MAP1a interaction, we mapped the binding site using a recently developed strategy that uses specific isotopic

labeling combined with heteronuclear NMR experiments(17). This method enables the efficient identification of interaction sites without prior backbone assignment of the protein of interest by comparing spectra of samples selectively labeled by amino acid type in free and complexed forms. By identifying the number of amino acids of each labeled type that display some spectral change upon binding, we determined the amino acid content of the interface. We then located a patch with matching amino acid content on the surface of the known structure, in this case, the SH3-GK module from PSD-95(5, 6). Analysis of the spectra (**Figure 2** and **Supplementary Figure 1**) demonstrated that 4 Tyr, 4 Lys, 2 Phe, 2 Val, and 1 Ile were sensitive to MAP1a binding, suggesting they were near the interface. The only region of the structure containing such a cluster of residues is the GK substrate-binding site. To confirm our mapping results, we individually mutated three of the four Tyr in the substrate-binding site to Phe (Y580F, Y604F, and Y609F) and collected ^{15}N-HSQC spectra on ^{15}N-Tyr labeled samples of each of the SH3-GK mutants (**Supplementary Figure 2**). This mutational analysis allowed us to assign peaks for each Tyr residue in the substrate-binding site and confirmed that the MAP1a binding site of PSD-95 is the former nucleotide-binding site of the GK.

To further confirm the location of the binding site, we mutated two conserved arginines (R568 and R571) used by the ancestral enzyme to coordinate its substrate. While R571A SH3-GK bound MAP1a equivalently to wild-type protein, R568A interacted with an approximately 100-fold lower affinity (**Figure 3**). PSD-95 GK thus uses its conserved substrate recognition motif to bind its protein ligands.

The PSD-95 GK binding partner consensus is not strongly constrained

Examining sequences of other proteins that interact with the PSD-95 GK, we determined a consensus binding sequence. We demonstrated that the GKAP proteins bind to the GK competitively with the MAP1a peptide (**Figure 4a**) and contain 2-5 repeats of a SYxKA motif (**Figure 4b**), which is strikingly similar to the 5 C-terminal residues (DYRKA) of the MAP1a peptide. While the MAP1a peptide binds with a K_d of 25μM, peptides derived from single repeats of the GKAP sequence showed no detectable binding in either fluorescence anisotropy (**Figure 4c**) or in NMR titration experiments (data not shown). This corroborates previous work in which single GKAP repeats were unable to competitively inhibit binding of the full-length protein to PSD-95(12).

We then determined sequence determinants of binding affinity. When we mutated the MAP1a peptide Asp to Ser, it abolished binding to GK (**Figure 4c**). As the Asp thus provides the majority of the energy of binding (based on a 100-fold minimum reduction in affinity), we used it as a fixed position in an array of "SPOT"-synthesized peptides to determine which positions of the peptide were most permissive(18). Each of the 5 residues N- and C-terminal to MAP1a D1874 were systematically altered in the context of a membrane matrix. Interaction of this peptide array with purified GST-SH3GK was detected with an anti-GST antibody (**Figure 4d**). Combining this peptide array analysis with quantitative analysis of solution-state binding of specific mutant peptides (**Figure 4c**) yields an unexpected trend of permissiveness at many positions in the sequence. While basic residues were required at positions –1, +2, and +3 to the Asp in the SPOT analysis, mutation of K1873, R1876, or K1877 to an uncharged residue yielded peptides that still interacted relatively well, with, at most, a 3-fold drop in

affinity. In solution, mutation of Y1875F yielded no change in affinity, while Y1875R (which interacted strongly in the SPOT analysis) bound over an order of magnitude more weakly. This position is conserved between the MAP1a and GKAP repeats, indicating a preference for aromatic amino acids. Whereas the preferred sequence seems to be that of MAP1a, it should be noted that GKAP binds using repeated non-ideal sequences. In fact, GKAP-GK interaction has recently been shown to be quite stable *in vivo*(19). Thus GK exhibits different modes of binding in its various interactions.

MAP1a binds GK in an extended conformation

We next characterized the bound MAP1a peptide conformation. As the relaxation characteristics of the SH3-GK domain made it impossible to collect standard NOESY experiments, we performed an exchange transferred-NOESY experiment(20) to obtain information about the bound peptide structure (data not shown). In a transferred-NOESY experiment, an ^{1}H-^{1}H-NOESY is collected with the low molecular weight ligand in high excess of its larger binding partner. Such an experiment relies on the fact that the strength of an NOE is proportional to a molecule's rotational correlation time. While a freely tumbling peptide has undetectable NOEs, a larger protein or complex develops strong NOEs. Thus all crosspeaks detected in the transferred-NOESY correspond to the ligand in its bound, slowly tumbling, state.

In our experiment, no amide-amide NOEs were detected, indicating that the bound peptide exists in an extended conformation. Accordingly, we observed strong NOEs between the alpha proton and the amide proton of consecutive amino acids for the C-terminal 8 residues. In addition, strong NOEs between the aromatic protons of Y1875 and the side chains of D1874 and R1876 were detected. No NOEs were detected

between any of the more N-terminal residues in the peptide, indicating that the residues N-terminal to P1870 are in much faster conformational exchange than those C-terminal. These data suggest that the final 8 residues of the peptide tightly contact PSD-95, which is consistent with our SPOT analysis of the MAP1a sequence (**Figure 4b,d**).

Given the secondary structure of the bound MAP1a, but lacking NOE-derived distance restraints between the GK and its ligand, we required additional information to orient the MAP1a peptide within the GK substrate-binding region. Paramagnetic centers have long been known to increase the relaxation rates of spins in a distance-dependent manner. More recently, it has been shown that quantitative structural restraints can be obtained from paramagnetic relaxation enhancement (PRE) to yield long-range distance constraints in systems whose size or dynamics otherwise complicates their study(21, 22). To provide distance restraints between the MAP1a peptide and the GK substrate-binding domain, a MAP1a peptide C-terminally labeled with nitroxide was bound to ^{15}N-Tyr labeled SH3-GK and the line shapes of HSQC spectra in oxidizing and reducing conditions were compared to calculate the level of PRE (**Figure 5**). The peaks corresponding to Y604 and Y609 were completely bleached, while Y580 was broadened. From these data, the C-terminus of the peptide was calculated to lie within 15Å from Y604 and Y609 and 18±5Å from Y580.

The dynamics of PSD-95 GK:MAP1a interaction suggests a binding intermediate

Analysis of the peak line shape during an NMR titration experiment can yield information about the dynamics of the interaction. In this analysis, each Tyr in the GK substrate-binding site exhibits different kinetics upon peptide binding (**Figure 2d and Supplementary Figure 3**). While Y604 shows a 2-state slow exchange behavior, both

Y573 and Y580 exhibit a more complicated binding mechanism. The additional (third) peak corresponding to Y580 during the titration indicates that Y580 is sensitive to an intermediate binding state. The peaks corresponding to Y573 in both the bound and unbound states undergo extensive line broadening with titration, reducing the peak intensity. This indicates that Y573 is also sensitive to the binding intermediate, though the rate of exchange is faster than that of Y580. As Y604 and Y609 are insensitive to the intermediate, it is possible that this represents a conformational change in the GK with the greatest displacement around Y573, and the least around Y604 and Y609.

The complex line shapes observed in the ^{15}N-Tyr-labeled spectra are also seen in the spectra of other labeled amino acids. In the ^{15}N-Val-labeled spectra (**Figure 2b**), while one peak undergoes a slight, slow-exchange-type shift, another peak exhibits spectral changes similar to that of Y573 (slow exchange with an intermediate in intermediate exchange, resulting in broadening). Similarly, one Phe undergoes a complete shift, while another shifts and is simultaneously broadened due to an exchange intermediate (**Figure 2a**).

The number of peaks affected by the binding event exceeds the number of residues in the substrate-binding site, again suggesting the binding to MAP1a triggers a change in conformation in the GK. The majority of the additional chemical shift changes are minor, implicating a domain reorientation as a likely type of motion. This is consistent with the yeast GK binding of GMP, which triggers a "clamping-down" of the lid region onto the substrate-binding region(23). Interestingly, in each of the three crystal structures of the apo-PSD-95 SH3-GK (PDB accession 1JXM, 1JXO, and 1KJW), temperature factors in the GK are largest in the lid and substrate-binding regions,

demonstrating that these regions are most dynamic. PSD-95 GK thus uses the motion inherent in the GK fold in its recognition of protein ligands.

MAP1a binds a closed form of PSD-95 GK

To better understand how PSD-95 and MAP1a interact, a model of the complex was generated. The PSD-95 GK and MAP1a binding region (1868-1878) were docked first by rigid body refinement then by molecular dynamics. Distance restraints were derived from chemical shift perturbation and PRE data. A sample of 1000 initial models was generated, and the top-scored 200 structures were refined according to the HADDOCK protocol(24). The final models were clustered and a representative of the top-scored cluster is shown in **Figure 6a**. Combining this model of the bound peptide with a homology model of the PSD-95 GK in the closed conformation as a starting point, we generated a model of the MAP1a peptide bound to the closed GK through a molecular dynamics simulation (**Figure 6b**).

The resulting models show the peptide bound in an extended conformation, which is consistent with our NOESY and PRE data. The peptide sits in a negatively charged channel between the substrate-binding and lid regions of the GK, explaining the strong preference for basic (and against acidic) residues at several positions.

A closer examination reveals that MAP1a D1874 forms a salt-bridge to the single positively charged area in the center of the PSD-95 GK substrate-binding region, formed by two conserved Arg (R568 and R571) which coordinate the nucleotide phosphate in yeast GK. Without a negative charge to orient the peptide in the GK substrate-recognition site, binding is greatly weakened, as seen with the D1874S mutation. The reciprocal mutations in the PSD-95 GK confirm this salt bridge as critical to binding

(**Figure 3**). Mutation of GK R568A shows an approximate 100-fold decrease in affinity for MAP1a, whereas R571A binds equivalently to wild-type protein.

In our model, each of the basic residues in the peptide (K1873, R1876, and K1877) salt bridges with an Asp that is conserved in neuronal MAGUKs (**Figure 5c, d**). In our solution studies, mutation of either K1873A or K1877A exhibited a loss of binding greater than a simple missing charge, which can be explained by the additional loss of van der Waals interactions. Mutation of R1876L, however, shows a loss of binding somewhat less than expected from a loss of charge. Interestingly, this position is unconserved between the MAP1a and GKAP repeats. Mutation of either D545A or D549A in GK reduced its affinity for MAP1a 7-fold (**Figure 3**), whereas E600A showed no change (data not shown). In the open conformation of GK, the MAP1a peptide would be unable to pack efficiently, fulfill our distance restraints, and maintain interactions with each R568, D545, and D549. The closed conformation of GK, however, can accommodate each of these restraints, affirming the validity of our model.

Discussion

Combining the power of specific labeling and NMR chemical shift perturbation with spectral assignment by mutagenesis, we have characterized the interface between the GK and its MAP1a ligand. We defined binding-induced conformational changes in GK and found that GK binds promiscuously to a variety of peptide targets. This has important implications for regulation of the MAGUK family and its role in signal transduction.

Evolution of protein-recognition in the context of an enzymatic fold

A series of crystal structures of enzymatic GK domains have documented the binding of both GMP and ATP substrates and the conformational change in the protein that their binding affects(23, 25). This conformational change can be described as a rigid body rotation of the core, lid, and substrate-binding regions around hinge regions (**Figure 7a**). While MAGUKs cannot catalyze phosphoryl transfer, it has been suggested that nucleotide binding may regulate MAGUK GK activity by inducing conformational changes(26) or by aiding in the coordination of protein ligand(5). We have shown that, in the case of the neuronal MAGUKs, the GK affinity for its protein ligand is independent of nucleotide concentration, and have previously demonstrated that MAGUKs lack a physiologically relevant affinity for guanylate nucleotide(27).

Bearing the above differences in mind, comparison of the MAGUK GK to its ancestral enzyme yields insight into the details of MAGUK GK function. While the core and substrate-binding regions of each of the four apo-MAGUK GKs (three structures of PSD-95 and one of CASK) overlay with each other and with the apo-yeast enzyme GK structure (**Figure 7b, Supplementary Figure 4**) the lid region shows the highest structural divergence. In the CASK and PSD-95 structures, the lid helices are rotated away from the substrate-binding region and closer to the P-loop. In PSD-95 but not in CASK, the second lid helix packs against the P-loop, completely occluding the region occupied by ADP in the enzyme structure. It is likely that this stabilizes the open conformational state of the MAGUKs, much as the binding of ATP stabilizes the fully closed state of the enzyme. The PSD-95 GK domains vary little between the three structures (average backbone RMSD = 0.54Å), compared to the yeast apo-GK (RMSD =

1.2Å between structures in the asymmetric unit; 2.8Å maximum displacement of the substrate binding domain, **Supplementary Figure 4**). Thus, contrary to the apo-enzyme GK, the MAGUK GK can be characterized as conformationally stable when unoccupied by ligand. Additionally, although the neuronal MAGUKs cannot bind nucleotide with their ATP-binding site, it is possible that CASK and related p55-family MAGUKs may have their affinity for protein ligand modulated by nucleotide occupancy in this site.

The region of the substrate-binding domain with the largest displacement in the apo-yeast GK corresponds to the loop in the PSD-95 GK containing Y573. This region undergoes the greatest motion in our model of the GK bound to MAP1a. Considering the substrate-binding domain motion as a rigid body rotation around helix 3 (PSD-95 residues 615-622, **Figure 7c**), as described by Sekulic, *et al.*, one would expect Y604 and Y609 to experience very little motion, lying near the axis of rotation. The amplitude of displacement would increase towards Y580 and further towards Y573. This is consistent with our NMR data, which demonstrate Y580 and Y573 both experience motion during ligand binding that Y604 and Y609 do not. This suggests that the three states captured by our NMR data may be described as: unbound GK, MAP1a bound to GK in an apo conformation, and MAP1a bound to GK in a closed conformation. Again, different from the yeast enzyme GK, the binding of ligand to MAGUKs activates an exchange between conformations.

Implications for Regulation

The residues that line the substrate recognition site are highly conserved in each neuronal MAGUK (**Figure 6c**). Thus it is likely that each of the neuronal MAGUK GKs recognizes the same consensus sequence. This is consistent with SAP-97 and SAP-102's

demonstrated functional equivalency to *Drosophila* DLG through their rescue of a *dlg*-line(28).

This sequence identity is reduced in the remainder of the GK and in the SH3, and the unstructured Hook domain shows divergence (**Figure 8a**). This area of dissimilarity contains sites for interaction of SAP-102 and SAP-97 with calmodulin(29) and interaction of p55 with 4.1R(30). Calmodulin binding is thought to mediate SH3-GK homo- and hetero-dimerization, by modulating the structure of the GK(5, 29, 31), which raises the intriguing possibility that this oligomerization also regulates GK binding of ligands through an alteration in the preferred motions of the substrate-binding region.

Several GK binding proteins have repeats of weak binding sites. As the GK binds sequences in regions of predicted low structural complexity, it is unlikely that these fold into a larger structure. Interestingly, these weak, repeated binding sites contain neither aspartates nor glutamates, but rather serines or threonines, which may be phosphorylated. In the case of GKAP, mutation of the serine to aspartate did not allow individual repeats to bind in either pull-downs or solution experiments (data not shown). In the context of solid-supported peptide binding studies, however, individual GKAP repeats were able to bind GST-SH3GK (data not shown). As diffusion is limited at synaptic sites, and one or more of the binding partners is often membrane associated, this solid-support experiment may better represent physiological circumstances. Regardless, these multi-site repeats likely bind through a mechanism of cooperatively enhanced local concentration, as in the multisite phosphorylation of Sic1(32, 33). The dynamic conformation of GK during binding to its ligand may aid in the recognition of non-ideal repeats by facilitating the local scanning of the peptide chain. However, the evolution of multiple non-ideal repeats

is complex and requires a benefit. For GK interactors, the concatenation of non-ideal repeats could serve as a point of regulation for binding. Alternatively, the repeats may provide a scaffold to cluster multiple MAGUKs through their GKs.

Whereas PSD-95 GK employs its substrate-binding site to bind protein ligands, the GK domain of $Ca_v\beta$ recognizes its ligand, the α-subunit of the channel, in its ATP-binding site(3). The $Ca_v\beta$ GK has lost the substrate-binding site, explaining why its ATP-binding site has become of primary importance. It is interesting to note that MAGUKs have maintained both regions, which together coordinate protein ligands. Additional regions in GK may recognize other ligands, such as GAKIN(34) and BEGAIN(35), neither of which contains the consensus sequence for GK binding.

Most intriguing is the conformational flexibility the GK employs to recognize ligands. While most peptide-recognition folds are small and rigid, the GK is relatively large and dynamic. Among biological macromolecules, motion underlies regulatory function. Phosphorylation of SAP-97 distal to the GK modulates its affinity for GKAP(36). Furthermore, splice variants of the Hook region of SAP-97 abrogate binding to GKAP(37). These larger Hook loops were postulated to occlude the GKAP-binding site on GK. However, we find the GKAP binding site lies on the side opposite the Hook (**Figure 1a**). In addition, in the previous study, the N-terminal L27 domain of SAP-97 could alleviate this inhibition in both *cis* and *trans* through an interaction with the Hook. Thus, the Hook regions of MAGUKs, which are sequence divergent, are the sites of differential splicing and provide sites for distal regulation of GK affinity, possibly through allosterically modulating the rate of GK conformational exchange (**Figure 8a-b**).

Finally, the GK domain interactions must be interpreted in light of the greater neuronal MAGUK scaffold. The repeated PDZs of MAGUKs cluster receptors, ion channels, and the down-stream machinery that serves to propagate the neuronal signal. We have shown that the GK domain itself can interact with a diverse set of ligands. In addition, the GK undergoes a conformational change upon binding, and is regulated by events distal to its substrate-binding region. Thus the MAGUK family may interact constitutively with partners through its PDZ domains and in a regulated fashion with its GK ligands. Such regulated interaction presents an important means of crosstalk among the various signaling networks of the synapse. In short, the complex interplay between the individual domains of MAGUKs demonstrate that these proteins must not be considered simple beads-on-a-string, as these intramolecular interactions have far-reaching implications for the networks which the scaffolds organize.

Methods

Protein Expression and Purification

To create ^{15}N-Lys-labeled SH3-GK, an *E. coli* strain BL21(DE3) transformed with the plasmid encoding His_6-SH3-GK under control of a T7 promoter (Novagen) was grown in 3L of LB to an OD_{600}=0.8, sedimented, and transferred to 1L of M9 media containing the appropriate amounts ^{15}N-labeled amino acids(38) and 0.3mM IPTG then grown overnight at 18°C. To label all other amino acids used in this study, expression was carried out in the DL39 auxotrophic strain as previously described(38). Bacteria were lysed by sonication in 50ml of 50mM HEPES pH 7.5, 500mM NaCl, and 15mM imidazole. Protein was purified on 1ml NiNTA resin (Qiagen) by elution with 150mM imidazole, after extensive washing with lysis buffer. Purity was assessed by SDS-PAGE. The protein was dialyzed against 20mM HEPES pH 7.5, 150mM NaCl, 2mM DTT, and concentrated to 350μM by centrifugal filtration (Millipore). GST-SH3-GK and GST-MAP1a were expressed in BL21(DE3) bacteria, as above. Bacteria were lysed by sonication in 50ml of 50mM HEPES pH 7.5, 200mM NaCl. Protein was purified on 1ml glutathione-sepharose resin (Pharmacia) by elution with 10mM glutathione, following extensive washing with lysis buffer. The purified protein was dialyzed, as above. GFP-GKAP was expressed in COS cells, as previously described(27).

NMR Methods

All spectra were measured on Bruker Avance NMR spectrometers operating at proton resonance frequencies of either 500 MHz or 600 MHz and equipped with cryogenic probes. Binding during the titration series was monitored with [^{15}N,^{1}H]-HSQC experiments, with a spectral width of 12.8 ppm in the proton and 32 ppm in the ^{15}N

dimension. 50 complex points were collected in the indirect dimension and 512 complex points in the direct dimension with either 128 or 256 scans per increment depending on protein concentration. Titrations were carried out in 5 steps, each one adding 0.3 mol equivalents of the peptide in a range from 1:0.3 to 1:1.5 protein:peptide molar ratio. The experiments were conducted at 30°C.

For PRE experiments, a MAP1a peptide corresponding to residues 1862-1878 was synthesized with a C-terminal Cys. This peptide was reacted with MTLS (Toronto Research) overnight and the resulting nitroxide-labeled peptide purified by reverse-phase HPLC. Purified His-tagged SH3-GK with all cysteines mutated to serines was titrated with the nitroxide labeled peptide in 20mM HEPES 7.5, 150mM NaCl to a 1:1.5 protein:peptide molar ratio. Spectra were measured as above on a Bruker Avance spectrometer operating at 800MHz equipped with a cryogenic probe. The nitroxide was reduced with an 8-fold molar excess of ascorbic acid to record a comparison spectrum without PRE. Distance restraints were calculated as described by Battiste, *et. al.*(21)

NOESY and TOCSY spectra of the free MAP1a peptide were measured on a 500 MHz spectrometer with 1024 complex points in the direct and 256 complex points in the indirect dimensions with a spectral width of 10.9 ppm in both dimensions. The mixing times were 400ms for the NOESY, measured with 64 scans per increment and 90ms for the TOCSY, measured with 16 scans per increment. Two transferred-NOESY experiments were measured with a sample containing 1mM peptide and 20μM SH3GK with either 200ms or 400ms mixing times and 64 scans per increment. All spectra were processed with TOPSPIN and analyzed with SPARKY.

Measurement of Binding Affinities

A peptide with a sequence corresponding to residues 1862-1878 of MAP1a was synthesized with an N-terminal Cys. This peptide was reacted with maleimide-rhodamine (Molecular Probes) and the resulting fluorescent peptide was purified by reverse-phase HPLC. Purified His-tagged wild-type or mutant SH3-GK was titrated against the fluorescent peptide in 100mM HEPES, pH 7.5, 150mM NaCl, 10mM β-mercaptoethanol, 0.01% Triton. Wild-type and mutant unlabeled peptides were titrated against a mixture of 1.25μM rhodamine-labeled peptide and 1.25μM wild-type His-SH3-GK in the above buffer. All data were fit with GraphPad Prism. Binding constants from competition binding assays represent K_I (taken to be equal to K_d) calculated from determined IC_{50} values.

Solid-state binding assays

A SPOT array scanning each position of the MAP1a peptide from 1869-1879, with D1874 used as an anchor, was obtained from Sigma-Genosys. The membrane was blocked with 100mM HEPES 7.5, 150mM NaCl + 5% BSA and probed with 10ml 10μM GST-SH3GK in buffer. The membrane was washed with buffer and probed with an HRP-conjugated anti-GST mAb (ICL) and detected with ECL (Amersham).

Molecular Modeling

A model of the MAP1a peptide (residues 1869-1878) was docked to the known PSD-95 GK structure (PDB: 1KJW) using a combination of rigid body and torsional dynamics according to the HADDOCK protocol(24) in CNS(39). Ambiguous constraints between the 4 Tyr in the GK binding site and peptide residues 1871-78 were used to drive the docking. 1000 initial models were generated, and the 200 lowest energy models refined

by simulated annealing in the context of explicit solvent. The final models were hierarchically clustered by backbone RMSD. The top-scoring model from the highest scored cluster was used as a starting point to generate a model of the peptide bound to the GK in its closed conformation. A homology model of the closed conformation of the PSD-95 GK structure was created from the structure of the GMP-bound yeast GK (PDB: 1EX7) using Modeller(40). The GK model's backbone was constrained in its initial position and the peptide left unconstrained. Distance restraints between the MAP1a peptide and the GK derived from the PRE and biochemical data were used to refine the model. 500 models were generated by simulated annealing in CNS and clustered as above.

References

1. Kuriyan, J. & Cowburn, D. (1997) *Annu Rev Biophys Biomol Struct* **26,** 259-88.
2. Lu, C. S., Hodge, J. J., Mehren, J., Sun, X. X. & Griffith, L. C. (2003) *Neuron* **40,** 1185-97.
3. Van Petegem, F., Clark, K. A., Chatelain, F. C. & Minor, D. L., Jr. (2004) *Nature* **429,** 671-5.
4. Funke, L., Dakoji, S. & Bredt, D. S. (2005) *Annu Rev Biochem* **74,** 219-45.
5. Tavares, G. A., Panepucci, E. H. & Brunger, A. T. (2001) *Mol Cell* **8,** 1313-25.
6. McGee, A. W., Dakoji, S. R., Olsen, O., Bredt, D. S., Lim, W. A. & Prehoda, K. E. (2001) *Mol Cell* **8,** 1291-301.
7. Woods, D. F., Hough, C., Peel, D., Callaini, G. & Bryant, P. J. (1996) *J Cell Biol* **134,** 1469-82.
8. Hoskins, R., Hajnal, A. F., Harp, S. A. & Kim, S. K. (1996) *Development* **122,** 97-111.
9. Harris, B. Z. & Lim, W. A. (2001) *J Cell Sci* **114,** 3219-31.
10. Brenman, J. E., Topinka, J. R., Cooper, E. C., McGee, A. W., Rosen, J., Milroy, T., Ralston, H. J. & Bredt, D. S. (1998) *J Neurosci* **18,** 8805-13.
11. Szebenyi, G., Bollati, F., Bisbal, M., Sheridan, S., Faas, L., Wray, R., Haferkamp, S., Nguyen, S., Caceres, A. & Brady, S. T. (2005) *Curr Biol* **15,** 1820-6.
12. Kim, E., Naisbitt, S., Hsueh, Y. P., Rao, A., Rothschild, A., Craig, A. M. & Sheng, M. (1997) *J Cell Biol* **136,** 669-78.
13. Tu, J. C., Xiao, B., Naisbitt, S., Yuan, J. P., Petralia, R. S., Brakeman, P., Doan, A., Aakalu, V. K., Lanahan, A. A., Sheng, M. & Worley, P. F. (1999) *Neuron* **23,** 583-92.
14. Cho, W. (2006) *Sci STKE* **2006,** pe7.

15. Cho, W. & Stahelin, R. V. (2005) *Annu Rev Biophys Biomol Struct* **34,** 119-51.
16. Nooren, I. M. & Thornton, J. M. (2003) *Embo J* **22,** 3486-92.
17. Reese, M. L. & Dötsch, V. (2003) *J Am Chem Soc* **125,** 14250-1.
18. Frank, R. (2002) *J Immunol Methods* **267,** 13-26.
19. Gerrow, K., Romorini, S., Nabi, S. M., Colicos, M. A., Sala, C. & El-Husseini, A. (2006) *Neuron* **49,** 547-62.
20. Post, C. B. (2003) *Curr Opin Struct Biol* **13,** 581-8.
21. Battiste, J. L. & Wagner, G. (2000) *Biochemistry* **39,** 5355-65.
22. Liang, B., Bushweller, J. H. & Tamm, L. K. (2006) *J Am Chem Soc* **128,** 4389-97.
23. Blaszczyk, J., Li, Y., Yan, H. & Ji, X. (2001) *J Mol Biol* **307,** 247-57.
24. Dominguez, C., Boelens, R. & Bonvin, A. M. (2003) *J Am Chem Soc* **125,** 1731-7.
25. Sekulic, N., Shuvalova, L., Spangenberg, O., Konrad, M. & Lavie, A. (2002) *J Biol Chem* **277,** 30236-43.
26. Li, Y., Spangenberg, O., Paarmann, I., Konrad, M. & Lavie, A. (2002) *J Biol Chem* **277,** 4159-65.
27. Olsen, O. & Bredt, D. S. (2003) *J Biol Chem* **278,** 6873-8.
28. Thomas, U., Phannavong, B., Muller, B., Garner, C. C. & Gundelfinger, E. D. (1997) *Mech Dev* **62,** 161-74.
29. Masuko, N., Makino, K., Kuwahara, H., Fukunaga, K., Sudo, T., Araki, N., Yamamoto, H., Yamada, Y., Miyamoto, E. & Saya, H. (1999) *J Biol Chem* **274,** 5782-90.
30. Marfatia, S. M., Leu, R. A., Branton, D. & Chishti, A. H. (1995) *J Biol Chem* **270,** 715-9.
31. Paarmann, I., Spangenberg, O., Lavie, A. & Konrad, M. (2002) *J Biol Chem* **277,** 40832-8.
32. Klein, P., Pawson, T. & Tyers, M. (2003) *Curr Biol* **13,** 1669-78.
33. Nash, P., Tang, X., Orlicky, S., Chen, Q., Gertler, F. B., Mendenhall, M. D., Sicheri, F., Pawson, T. & Tyers, M. (2001) *Nature* **414,** 514-21.
34. Hanada, T., Lin, L., Tibaldi, E. V., Reinherz, E. L. & Chishti, A. H. (2000) *J Biol Chem* **275,** 28774-84.
35. Deguchi, M., Hata, Y., Takeuchi, M., Ide, N., Hirao, K., Yao, I., Irie, M., Toyoda, A. & Takai, Y. (1998) *J Biol Chem* **273,** 26269-72.
36. Sabio, G., Arthur, J. S., Kuma, Y., Peggie, M., Carr, J., Murray-Tait, V., Centeno, F., Goedert, M., Morrice, N. A. & Cuenda, A. (2005) *Embo J* **24,** 1134-45.
37. Wu, H., Reissner, C., Kuhlendahl, S., Coblentz, B., Reuver, S., Kindler, S., Gundelfinger, E. D. & Garner, C. C. (2000) *Embo J* **19,** 5740-51.
38. Ou, H. D., Lai, H. C., Serber, Z. & Dotsch, V. (2001) *J Biomol NMR* **21,** 269-73.
39. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998) *Acta Crystallogr D Biol Crystallogr* **54 (Pt 5),** 905-21.
40. Sali, A. & Blundell, T. L. (1993) *J Mol Biol* **234,** 779-815.

Figure Legends

Figure 1: A) Domain organization of neuronal MAGUKs. While each of the neuronal MAGUKs has three PDZ domains followed by the SH3-GK domain, only certain splice variants of SAP-97 have an N-terminal L27 domain. The structure of the SH3-GK cassette is shown inset. Labeled are the disordered Hook of the SH3 and the lid helices and substrate-binding domain (SBD) of the GK. B) GST-tagged MAP1a constructs of various lengths were used as bait to pull-down His-tagged SH3-GK. Residues 1862-1878 of MAP1a define the minimal binding region to the GK. Bands are visualized with coomassie blue staining.

Figure 2: ^{15}N-HSQC titration experiments. A) ^{15}N-Phe labeled SH3-GK B) ^{15}N-Val labeled SH3-GK C) ^{15}N-Ile labeled SH3-GK. Spectra of free SH3-GK are shown in black, those bound to the MAP1a peptide are in red. Peaks undergoing significant chemical shift changes are starred. All spectra were collected at 500 MHz. D) ^{15}N-Tyr labeled SH3-GK, titrated with MAP1a peptide. These spectra were collected at 600 MHz.

Figure 3: Binding curves for rhodamine-labeled MAP1a peptide interaction with wild-type and SH3-GK carrying mutations in the substrate binding domain.

Figure 4: A) MAP1a binds competitively for GK with GKAP. A synthetic peptide with the sequence of MAP1a 1862-1878 competitively inhibited GKAP's binding to GK. B) The MAP1a binding sequence is significantly similar in sequence (boxed) to each of the five repeats in the GKAP GK recognition region. The region of MAP1a demonstrated to contact GK by NMR is boxed with a dotted line. C) Peptide half-maximal inhibitory concentration curves for competition of rhodamine-MAP1a peptide binding to SH3GK

by GKAP repeats and by wild-type and mutant MAP1a. D) Characterization of the GK consensus sequence with a membrane-bound array of peptides based on the MAP1a sequence. Wild-type residues are indicated in bold. Residues not permitted at a given position are bracketed in the consensus column. It is important to note that strength of signal is not necessarily correlated with strength of interaction.

Figure 5: ^{15}N-Tyr labeled SH3-GK with all cysteines mutated to serines was bound to a MAP1a peptide C-terminally labeled with nitroxide spin label and HSQC spectra were collected at 800MHz with the spin label in both an oxidized (black) and reduced state (red). The peaks corresponding to Y604 and Y609 were completely bleached by the nitroxide, putting an upper limit of 15Å on their distance from the peptide C-terminus, thus orienting the peptide in the binding site.

Figure 6: Molecular model of MAP1a bound to GK. A) A model of the MAP1a peptide bound in the GK GMP-binding site in an open conformation was constructed using molecular dynamics restrained by ambiguous NMR-derived constraints. The GK is represented as a molecular surface coloured with its electrostatic potential (negative in red, positive in blue). The MAP1a peptide (black) is shown in stick representation. MAP1a and GK residues of interest are labeled (GK in italics). B) A model of the MAP1a peptide bound to GK in a closed conformation was constructed by refining the peptide docked in the open GK in the context of a model for the closed GK. C) Peptide binding diagram. The GK side chains (grey) that contact the MAP1a peptide residues that provide specificity (black) in the model structure are illustrated. Charged and polar interactions are represented in red, hydrophobic interactions in blue. D) Alignment of PSD-95 family GK domains. The rat PSD-95, SAP-97, PSD-93, SAP-102, and CASK

GK domains were aligned with the *Drosophila* discs large (dlg) and yeast GK using CLUSTALW. Amino acid identities are shaded and marked in black, similarities in shaded blue. Amino acids that make contacts with the MAP1a peptide in the modeled structure are starred. The peptide-binding region is boxed with a dotted line, while the lid is boxed with a dashed line.

Figure 7: Motions inherent in the GK domain. A) Structural alignment of enzymatic GK in three states, apo (1EX6; blue), GMP-bound (1EX7; cyan), ADP and GMP-bound (1LVG; green). The motion of the protein can be described as a rigid body rotation of the lid and substrate binding (labeled SB) regions around hinges connected to the core of the fold. B) Structural alignment of two MAGUK GKs with apo-enzyme GK. The bound ADP of 1LVG is represented as licorice. While the PSD-95 (1KJW; grey; RMSD = 1.2Å) and CASK GK (1KGD; green; RMSD = 1.0Å) overlay with the yeast GK in its apo state (1EX6; cyan), their lid regions are rotated towards the ATP binding P-loop. The packing of the MAGUK lid with the P-loop may act to stabilize the apo-conformation. C) The complex line shapes of the tyrosines in the substrate binding site (**Figure 2d**) reveal dynamics consistent with the rotation of the substrate binding region around an axis (Helix 3) collinear with backbone of Y604 and Y609. Y580 and Y573 each demonstrate respectively faster dynamics during interaction with MAP1a, which is consistent with the greater displacement they experience during the swinging of the substrate-binding region. Here the closed conformation of GK is grey and the open is purple, with the positions of the tyrosine amides highlighted as spheres.

Figure 8: A model for regulation of inter-network crosstalk. A) Alignment of the SH3 and Hook domains of the neuronal MAGUK family. The rat PSD-95, PSD-93, SAP-97,

SAP-102, and *Drosophila* Dlg sequences were aligned using CLUSTALW. Identical residues are shaded, similar residues are displayed in black. The Hook region is boxed. B) The Hook region of SAP-97 has been shown to downregulate GK affinity for its ligands, possibly through the induction of a conformational change in the GK. The SAP-97 N-terminal L27 domain was demonstrated to interact with the Hook, relieving the effect. It is possible that other binding partners of the Hook region, including calmodulin and 4.1N, may have a similar regulating effect.

Supplementary Figure 1: Titration of ^{15}N-Lys labeled SH3-GK with MAP1a peptide. The spectrum of free SH3-GK is shown in black; that bound to the MAP1a peptide is in red. Peaks undergoing significant chemical shift changes are starred. Spectra were collected at 500 MHz.

Supplementary Figure 2: Peak assignment through mutation. Spectra were collected at 600 MHz. A) Unbound ^{15}N-Tyr labeled Y580F SH3-GK. B) Unbound ^{15}N-Tyr labeled Y609F SH3-GK. C) ^{15}N-Tyr labeled Y604F SH3-GK. Spectrum of free SH3-GK is in black, that bound to MAP1a peptide is in red. Note that in the free SH3-GK all four peaks corresponding to the Tyr in the GK substrate binding regions exhibit severe line broadening, and have disappeared. In the bound state, 3 of the 4 have reappeared, allowing for assignment of Y604. D) ^{15}N-Ile labeled unbound SH3-GK. Spectrum of wild-type SH3-GK is shown in black, that of Y604F is in red. Note that one peak exhibits severe line-broadening in the mutant protein, which corrresponds to I575, the single Ile in the substrate-binding site. Note that this peak exhibits slow-to-intermediate exchange in the wild-type bound spectrum (**Figure 2c**). Also note that the peak that exhibits line-broadening upon binding of MAP1a peptide in the wild-type protein is

unaffected by change in dynamics caused by Y604F. Thus the binding of MAP1a induces conformational exchange outside of the immediate substrate-binding region.

Supplementary Figure 3: Spectra of titration of ^{15}N-Tyr labeled wild-type SH3GK with MAP1a have been separated from overlay to allow easier examination of line shape changes. A) Unbound spectrum B) 5:2 protein:peptide C) 5:4 protein:peptide D) 1:1 protein:peptide

Supplementary Figure 4: Structural alignment of GK domains. A) Yeast GK is conformationally dynamic in the apo state. The two molecules of the assymetric unit of PDB 1EX6 were aligned with an RMSD of 1.2Å. B) The PSD-95 GK is rigid in its apo state. Each of the three structures of the PSD-95 GK (1JXM, 1JXO, 1KJW) were aligned with an RMSD of 0.54Å.

Figure 1

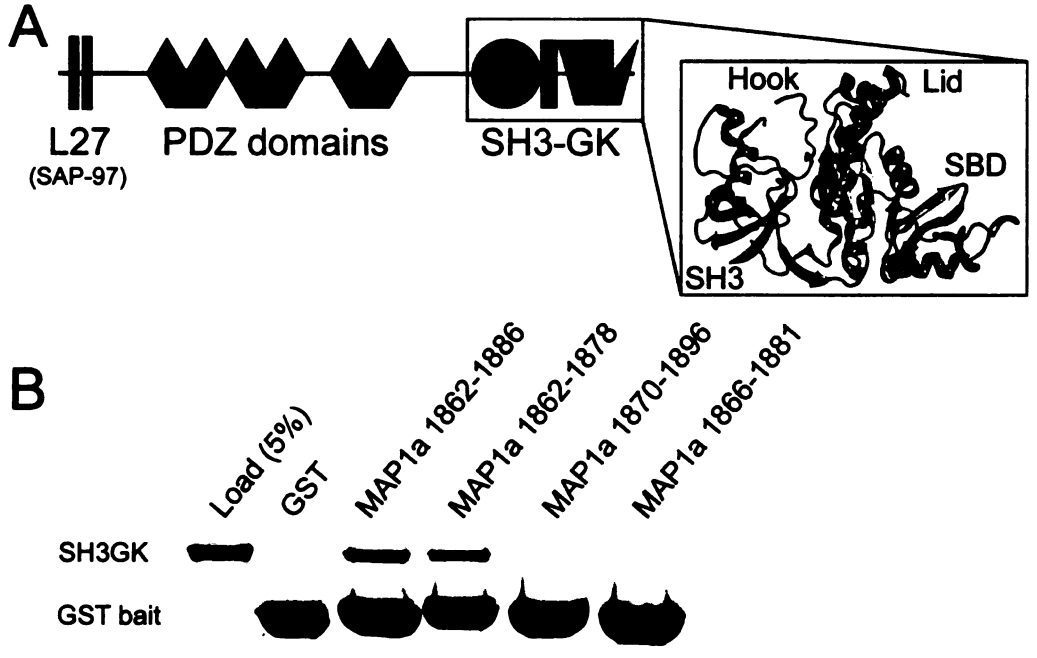

A
L27
(SAP-97)
PDZ domains
SH3-GK
Hook
Lid
SBD
SH3
B
Load (5%)
GST
MAP1a 1862-1886
MAP1a 1862-1878
MAP1a 1870-1896
MAP1a 1866-1881
SH3GK
GST bait

Figure 2

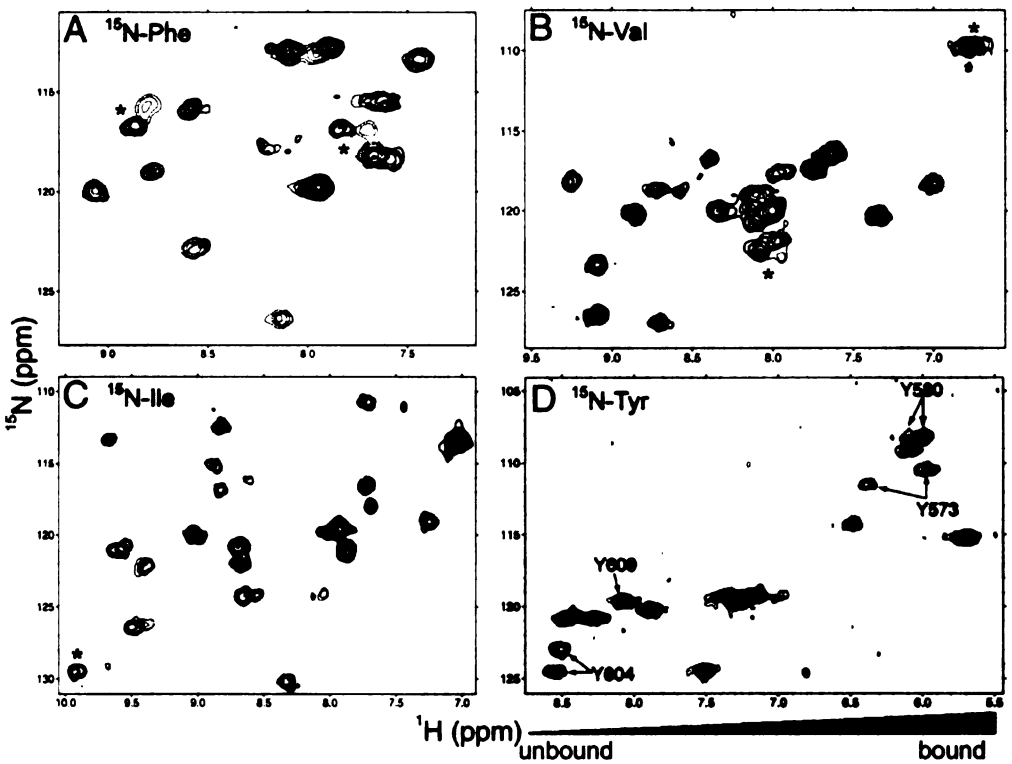

A ^{15}N-Phe
B ^{15}N-Val
C ^{15}N-Ile
D ^{15}N-Tyr
Y580
Y573
Y604
^{15}N (ppm)
^{1}H (ppm)
unbound
bound

Figure 3

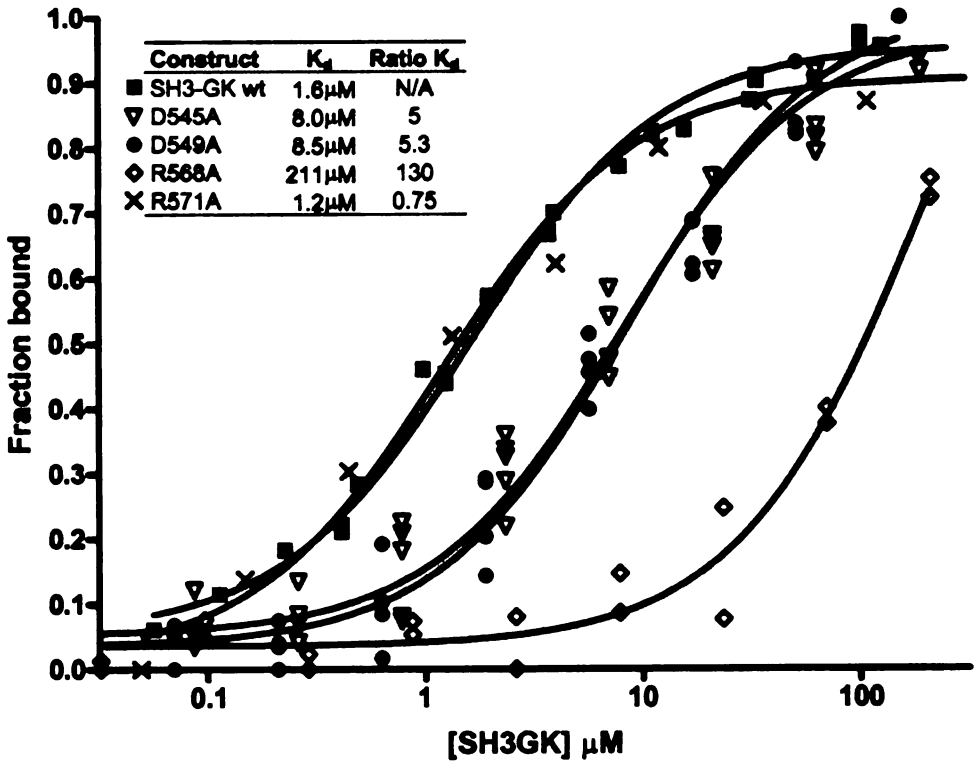

Construct
K_d
Ratio K_d
SH3-GK wt 1.6μM N/A
D545A 8.0μM 5
D549A 8.5μM 5.3
R568A 211μM 130
R571A 1.2μM 0.75
Fraction bound
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.1
1
10
100
[SH3GK] μM

Figure 4

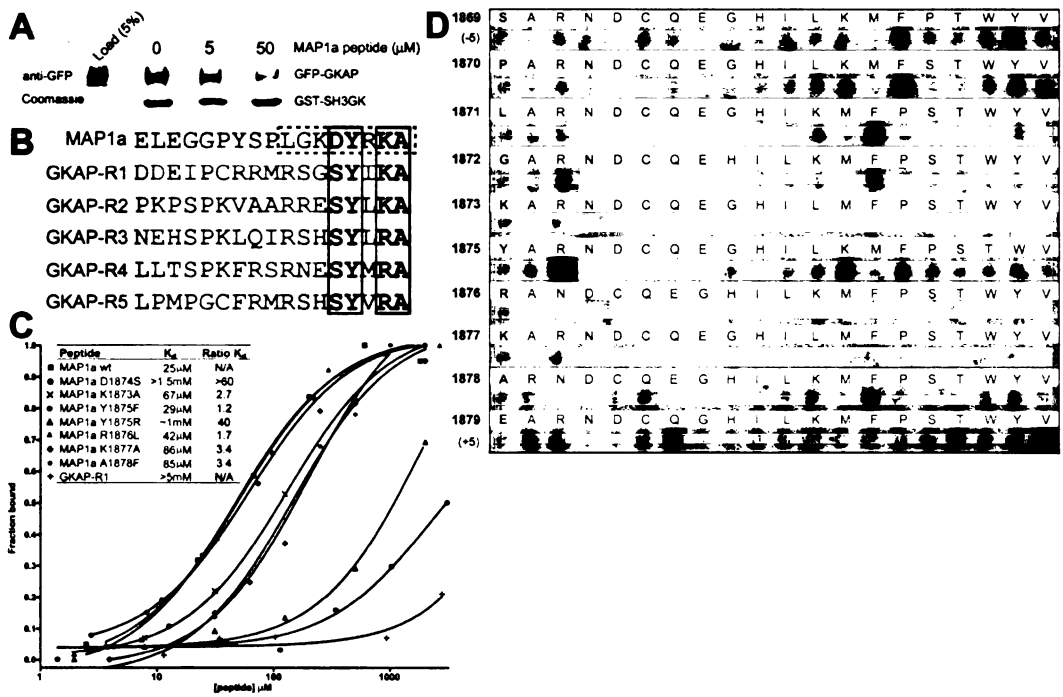

A
Load (5%)
0
5
50
MAP1a peptide (μM)
anti-GFP
GFP-GKAP
Coomassie
GST-SH3GK
B
MAP1a ELEGGPYSPLGKDYRKA
GKAP-R1 DDEIPCRRMRSGSYLKA
GKAP-R2 PKPSPKVAARRESYLKA
GKAP-R3 NEHSPKLQIRSHSYLRA
GKAP-R4 LLTSPKFRSRNESYMRA
GKAP-R5 LPMPGCFRMRSHSYMRA
C
Peptide
Kd
Ratio Kd
MAP1a wt 25μM N/A
MAP1a D1874S >1.5mM >60
MAP1a K1873A 67μM 2.7
MAP1a Y1875F 29μM 1.2
MAP1a Y1875R ~1mM 40
MAP1a R1876L 42μM 1.7
MAP1a K1877A 86μM 3.4
MAP1a A1878F 85μM 3.4
GKAP-R1 >5mM N/A
Fraction bound
[peptide] μM
D
1869
(-5)
1870
1871
1872
1873
1875
1876
1877
1878
1879
(+5)

Figure 5

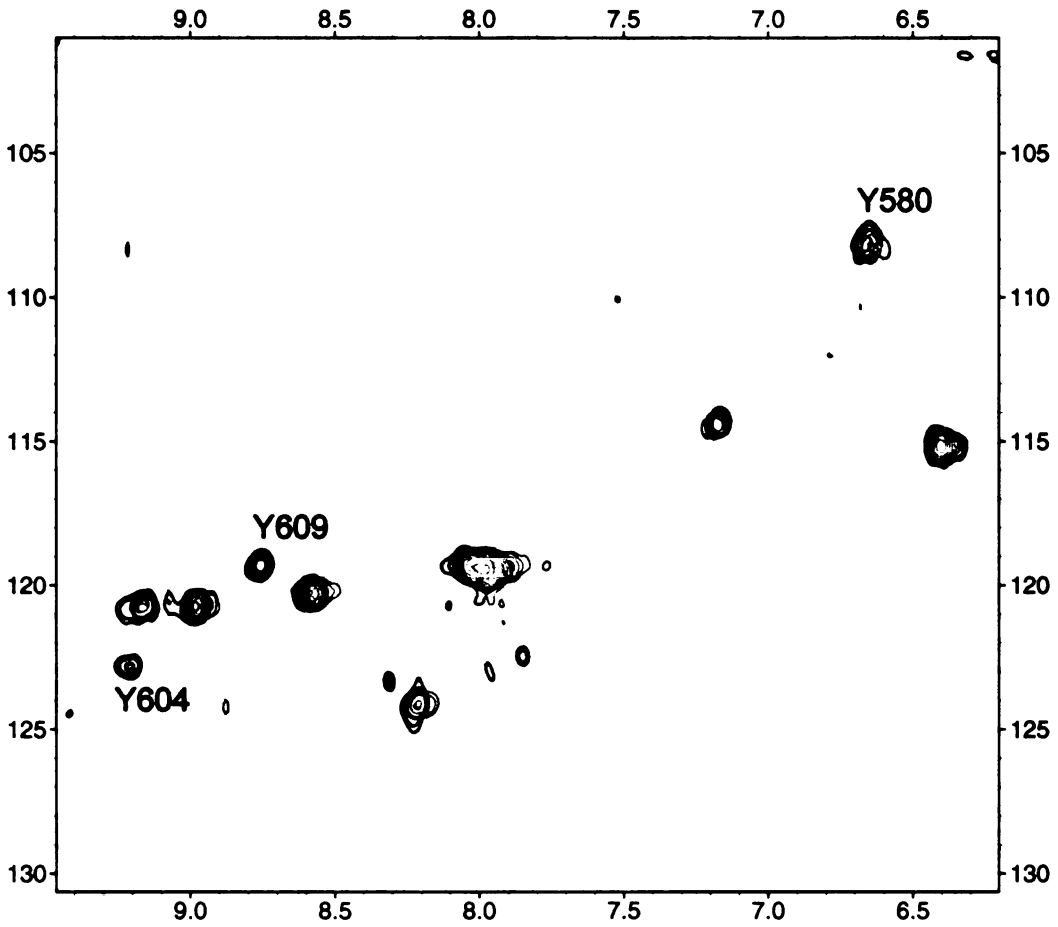


Figure 6

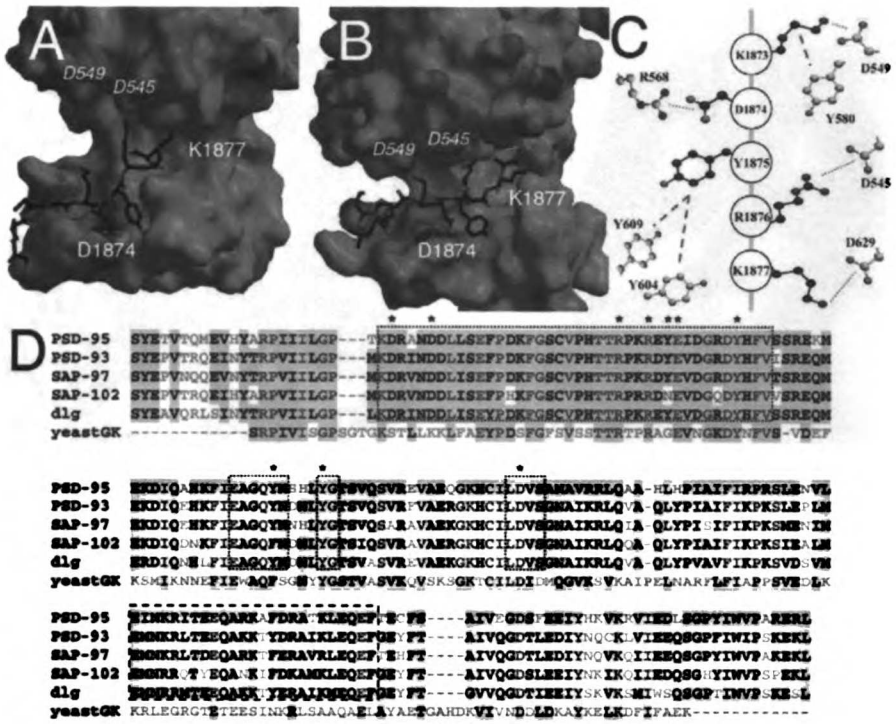

A
D549
D545
K1877
D1874
B
D549
D545
K1877
D1874
C
K1873
R568
D549
D1874
Y580
Y1875
D545
R1876
Y609
D629
K1877
Y604
D
PSD-95
PSD-93
SAP-97
SAP-102
dlg
yeastGK

Figure 7

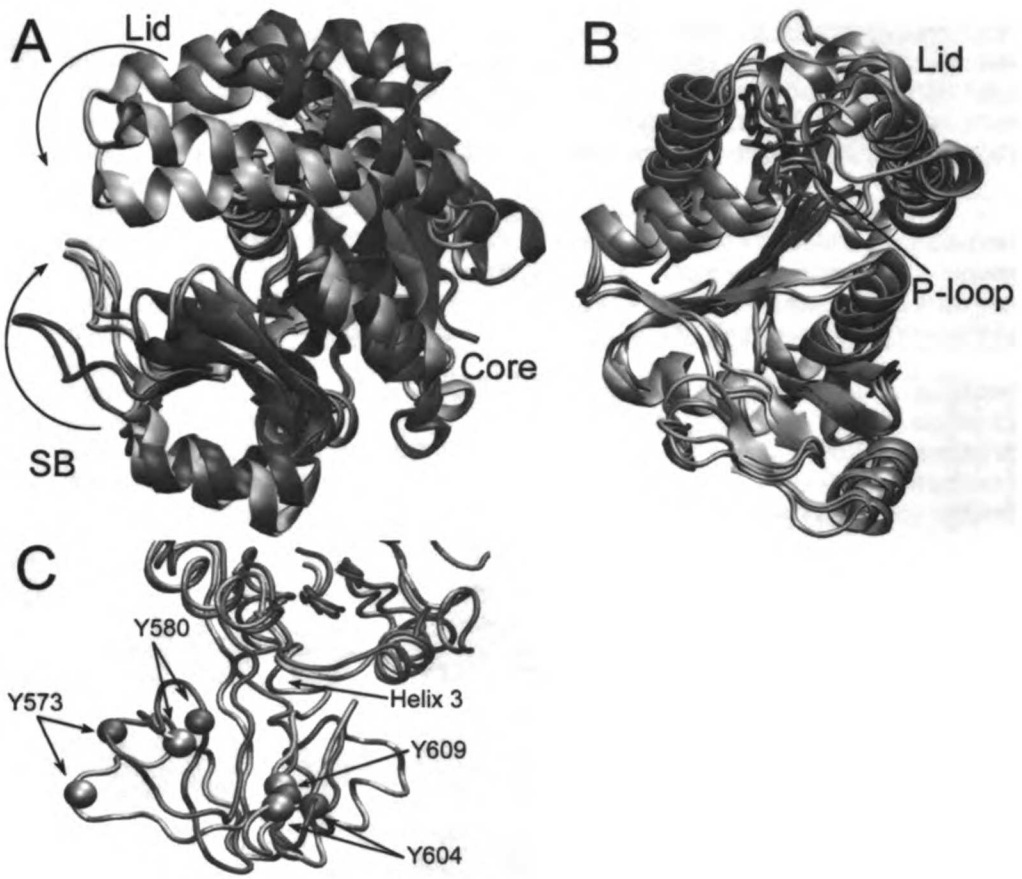
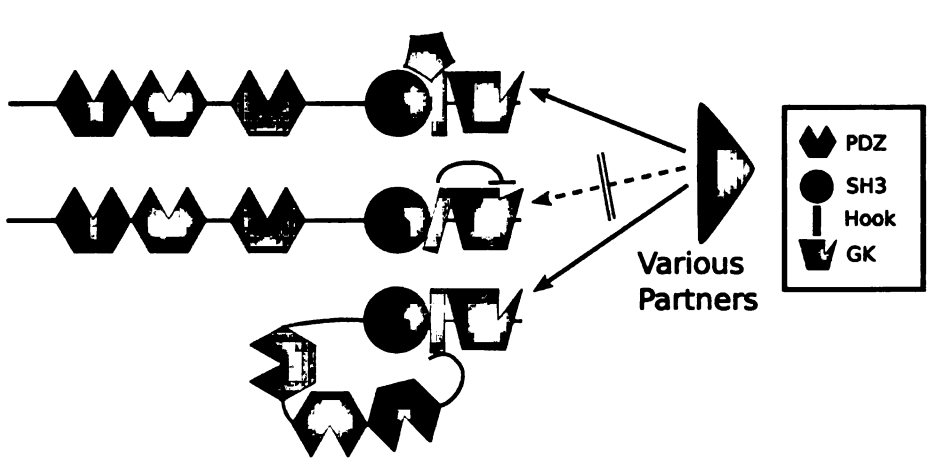


Figure 8

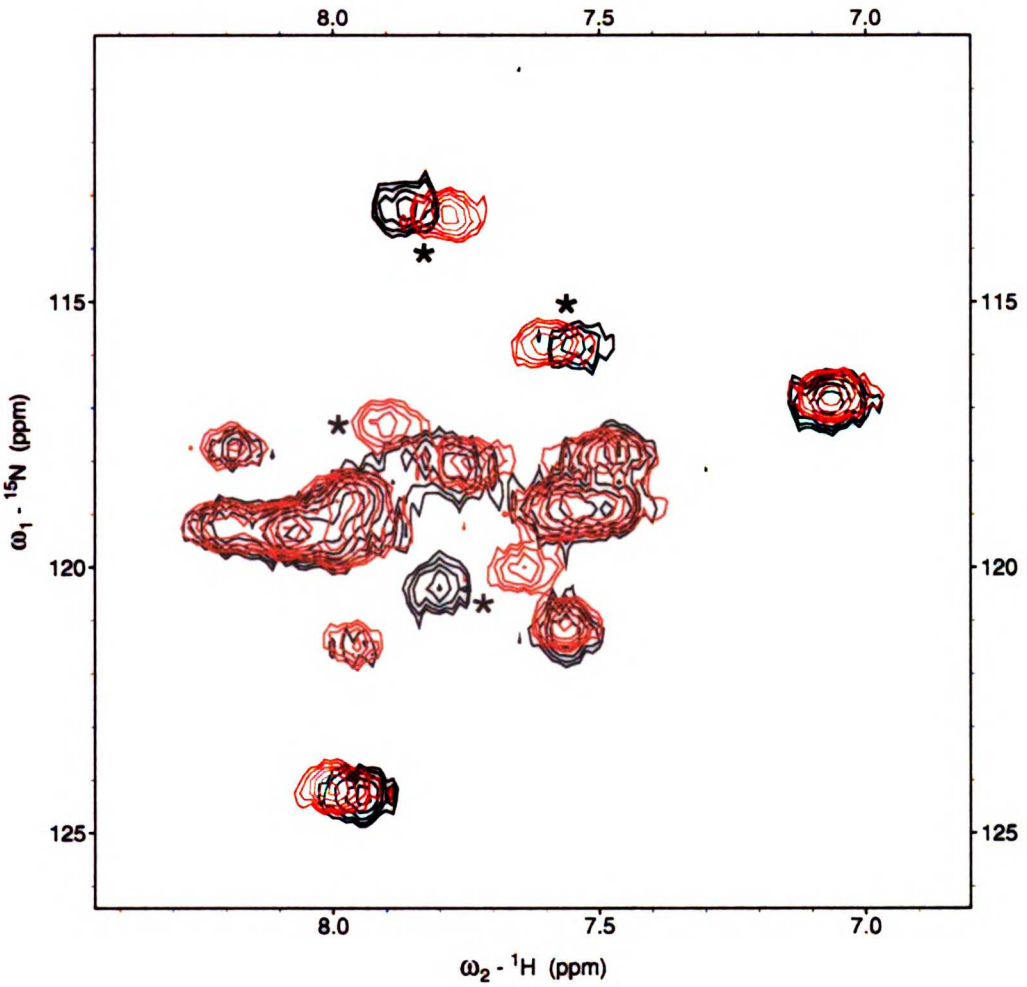
A

```
PSD-95   GFYIRALFDYDKTKDCGFLSQALSFRFGDVLHVIDAGDEEWWQARRVHSDSE
PSD-93   SLYVRAMFDYDKSKDSGLPSQGLSFKYGDILHVINASDDEWWQARRVILDGD
SAP-97   SLYVRALFDYDKTKDSGLPSQGLNFKFGDILHVINASDDEWWQARQVTPDGE
SAP-102  SLYVRALFDYDRTRDSCLPSQGLSFSYGDILHVINASDDEWWQARLVTPHGE
dlg      SLYVRALFDYDPNRDDGLPSRGLPFKHGDILHVTNASDDEWWQARRVLGDNE

PSD-95   TDDIGFIPSKRRVERREWSRLKAKDWG---SSS---GSQ------------
PSD-93   SEEMGVIPSKRRVERKERARLKTVKFN---AKPGVIDSKGDIPGLGDDGYGT
SAP-97   SDEVGVIPSKRRVEKKERARLKTVKFN---SKT--RGDKGEIP----DDMGS
SAP-102  SEQIGVIPSKKRVEKKERARLKTVKFH---ARTGMIESNRDFPGLSDDYYGA
dlg      DEQIGIVPSKRRWERKMRARDRSVKFQGHAAAANNNLDKQSTLDRKKKNFTFS

PSD-95   ------------------------------------------------GREDSVL
PSD-93   KTLR--------------------------------------------GQEDLIL
SAP-97   KGLKHVTSNASDSESSYHE--------------------YGCSKGGQEEYVL
SAP-102  KNLKGVTSNTSDSES-------------------------SSKGQEDAIL
dlg      RKFPFMKSRDEKNEDGSDQEPNGVVSSTSEIDINNVNNNQSNEPQPSEENVL
```

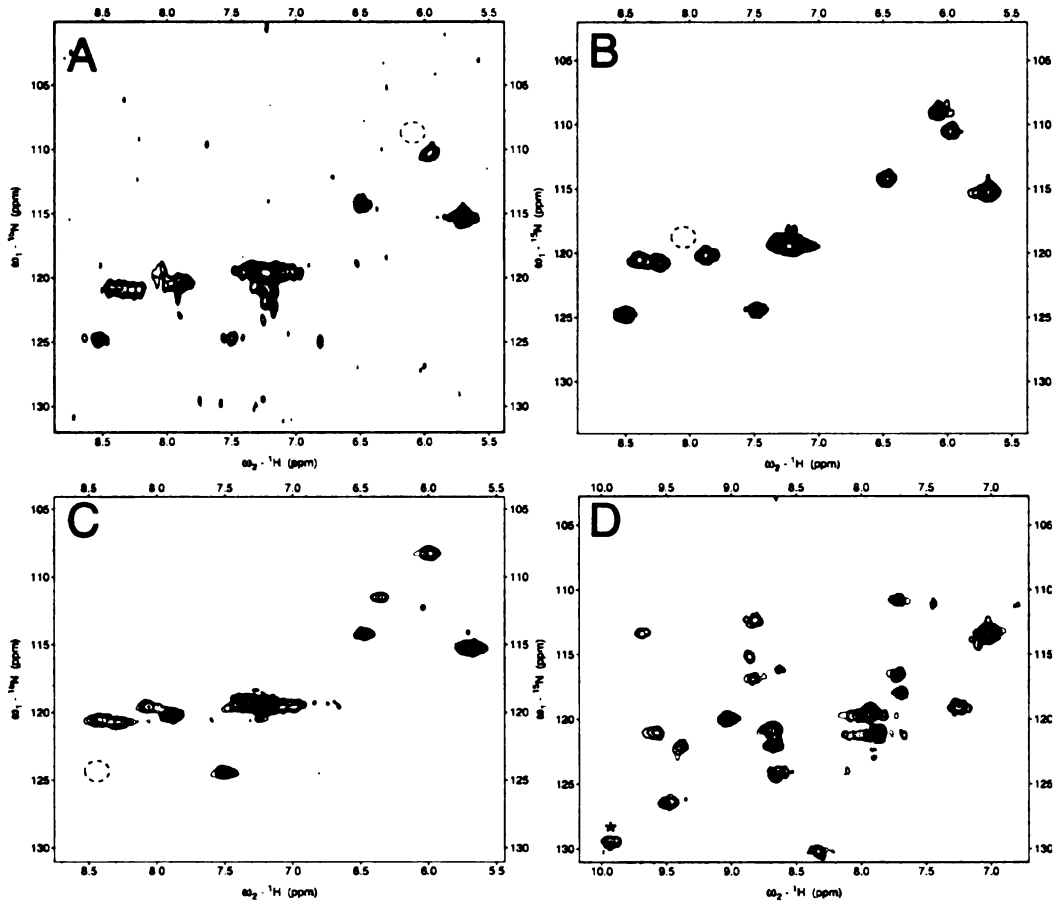
B



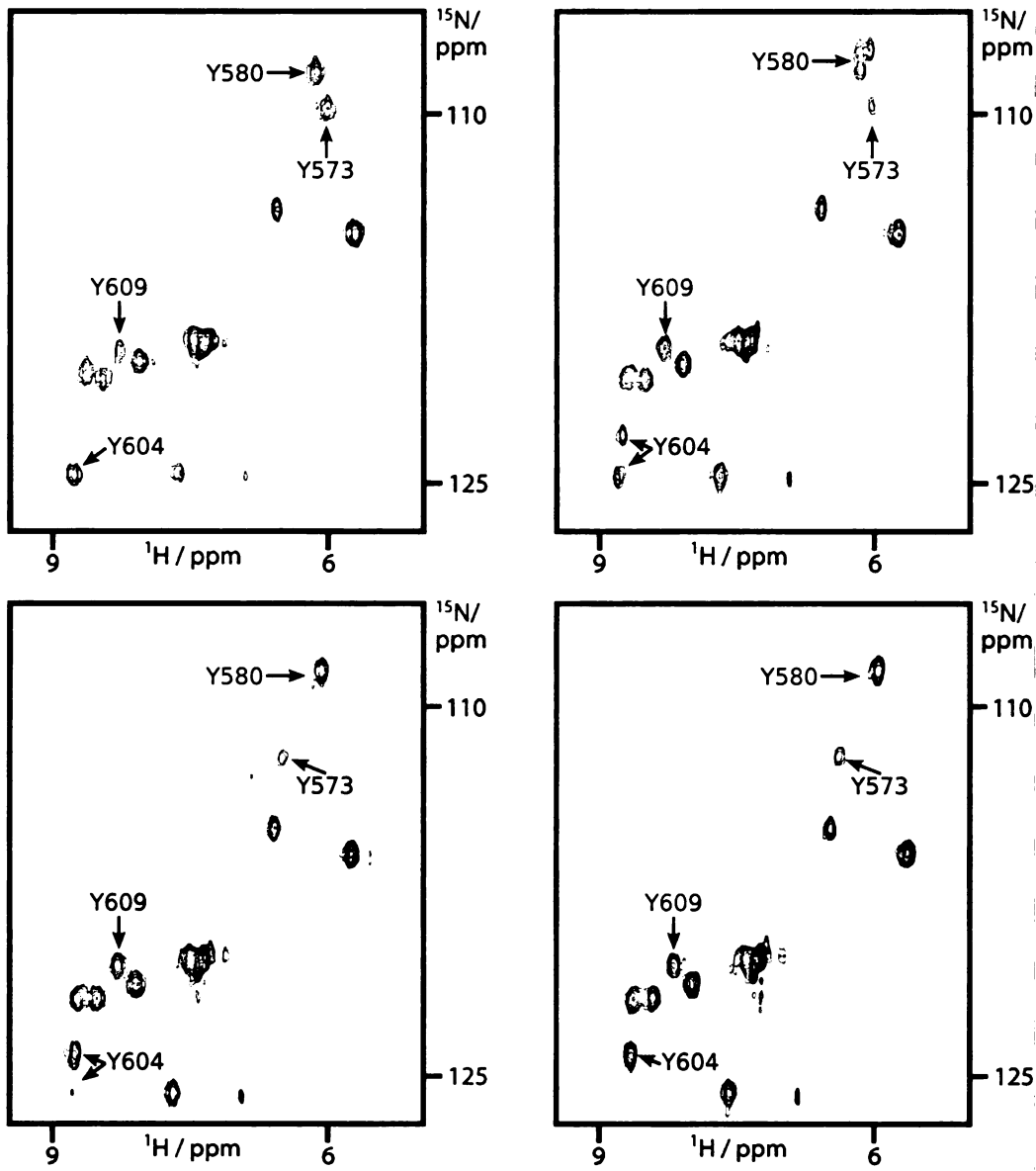
Supplementary Figure 1



Supplementary Figure 2



Supplementary Figure 3



Supplementary Figure 4

A

B

Chapter III: Towards a structural understanding of the regulation of clathrin assembly by its light chain

Preface

In eukaryotic cells, the sorting of proteins between organelles, the recycling of components in the plasma membrane, and the endocytosis of extracellular components are accomplished through the action of the ubiquitous protein clathrin, which forms coated vesicles. Clathrin is a hexamer composed of three trimerized heavy chains each associated with a single regulatory light chain. While the heavy chain is known to form the skeleton of the clathrin coat which drives vesicular budding, little is known about the full mechanism of clathrin action. The light chains are known to bind several auxiliary proteins(1-3), and have been implicated both as promoters(1, 2) and inhibitors(4) of clathrin assembly. However, the true physiological role of light chain remains uncertain. We determined the critical points of contact between the heavy and light chains of clathrin and created a structural model of the complex that recently was confirmed by EM(5). Additionally, we have demonstrated that the N- and C-termini of the light chain weakly associate using a region of the light chain previously shown to downregulate clathrin assembly. Furthermore, physiological data suggest that this interaction may provide the link between calcium signaling and clathrin assembly. Using NMR, we are working to create a structural model of the complex that will provide a better understanding of the mechanism of light chain regulation of clathrin.

Introduction: ***Clathrin light and heavy chain interface: alpha-helix binding superhelix via critical tryptophans***

Clathrin-coated vesicles (CCVs) mediate selective budding of receptors and ligands for endocytosis from the plasma membrane and for protein sorting from the trans-Golgi network during organelle biogenesis. On the cytosolic side of the membrane, the clathrin coat assembles into a polyhedral lattice that concentrates cargo proteins to specific sites within the membrane and contributes to membrane deformation for vesicle budding. Recent studies have identified regulatory proteins that influence cargo selection, membrane deformation, fission and uncoating(6). However, clathrin itself has intrinsic properties regulating self-assembly that are not yet fully defined. In particular, critical self-regulation is imposed by the interaction of the clathrin light chain (LC) subunit with the heavy chain (HC) subunit. In this study, we characterize the structural features of the LC–HC interface to gain further insight into the process of CCV formation.

The assembly unit of the polyhedral clathrin lattice has a triskelion (three-legged) shape and is formed by three HC subunits, each with an associated LC subunit. The center of the triskelion, also known as the hub, comprises the C-terminal third of the HC (mammalian residues 1074–1675) and includes the region where three HCs trimerize (1550–1615) and radially extend into proximal leg domains (1074–1522) where LCs bind(7). Beyond the hub, the legs kink and the HCs fold into linear distal leg domains)(6) adjacent to ß-propeller N-terminal domains (TDs; residues 1–546)(8). TDs interact with adaptors and accessory factors involved in CCV formation. Clathrin HCs constitute the lattice backbone, and in vitro studies have demonstrated that LCs exert negative regulation over the spontaneous assembly of HCs(4, 7, 9). In cells, this LC-dependent

inhibition is overcome by participation of adaptor molecules in the assembly reaction. The heterotetrameric adaptor molecules induce clathrin lattice formation, link the lattice to membranes and sequester receptor cargo(6, 10).

In vertebrates (mammals, fish and birds), there are two types of LCs with 60% sequence identity, LCa and LCb, encoded by different genes. In invertebrates (yeast, worms, Aplysia and Drosophila), there is only a single LC, but several LCs appear to exist in plants(6, 11). The two vertebrate LC forms are expressed in all tissues at varying relative levels(12) and are distributed heterogeneously in clathrin triskelia(13). In neuronal cells, alternative splicing generates LCa and LCb isoforms, inserting 30 or 18 additional amino acids, respectively(14, 15). A linear arrangement of structural and/or functional domains in LCs has been defined and suggests multiple regulatory roles for LCs in CCV formation(16). From the N- to C-termini, there is a conserved 22 residue sequence shared by all mammalian LCs with three negatively charged residues that regulate HC assembly(4), a calcium-binding site, the HC-binding region, neuronal inserts and a calmodulin-binding site. There are also sequences unique to each of the LCs, including phosphorylation sites for casein kinase II on LCb and an hsc70-binding site in LCa.

LCs were shown to associate with the HC proximal leg domain by antibody localization(13, 17) and by *in vitro* binding assays(7, 18). Genetic studies revealed that yeast LC binding depends on additional residues in the HC trimerization domain (Txd)(19). In contrast, mammalian LCs will bind to HC fragments missing the Txd. Mammalian LCs bind the HC via their central region(20, 21), which was previously predicted to participate in a coiled-coil interaction with the HC, based on sequence motifs

detected in both subunits(14, 15, 18). However, the crystal structure of the LC-binding portion of the proximal leg domain of HC revealed that it is an elongated superhelix coil of short α-helices(22). This finding, that the HC structure is not compatible with a classical coiled-coil interaction with the LC, prompted further analysis of the HC–LC interaction reported here.

Using yeast two-hybrid assays in combination with random PCR mutagenesis, we defined molecular contacts between the LC and the superhelical leg of the HC. This strategy selected first for mutants in LC that lost HC binding, and then for mutants in the HC partner that regained binding to the initial mutants. The first half of the scheme, also called a reverse two-hybrid assay, has been used often to analyze binding specificity(23). The other half takes advantage of the genetic concept of suppressors. This is, to our knowledge, the first example using these combined screens to dissect the molecular contacts of a protein–protein interface. Our results indicated that for LC to interact with HC, its central domain must maintain an a-helical structure. In addition, two contact sites between HC and LC were mapped at the molecular level, allowing us to generate a detailed structural model for subunit interaction. This model defines the extent of the HC–LC interface and suggests mechanisms for LC regulation of clathrin self-assembly.

Results

Defining minimal sites of HC and LC interaction

Boundaries of HC and LC interaction suggested by previous studies were defined by loss of activity upon proteolysis and deletion analysis(7, 18, 21). We therefore undertook a 'positive' mapping of the minimal interacting domains. Fragments of bovine clathrin HC and LCb (neuronal form) were cloned into two-hybrid bait (pGBT9) and

prey (pGAD424 or pACT2) vectors, respectively, to express the peptides fused to the GAL4 DNA-binding domain or the GAL4 activation domain. Consistent with previous findings(7), both the hub fragment (1073–1675) and the HC proximal leg (1073–1522) interacted effectively with full-length LCb when tested by two-hybrid assay (Figure 1A). Shorter HC fragments (1204–1522 and 1267–1522) apparently bound better to LCb than hub or the proximal domain, indicating possible inhibitory effects of HC regions beyond what is required for binding. Deletions at the N-terminus, after residue 1267, or at the C-terminus, before residue 1522, decreased HC fragment binding to LCb. When both ends were truncated simultaneously, the interaction of 1313–1481 with LCb was substantially weaker, suggesting that both ends of the 1267–1522 minimal domain are involved in LCb binding. Nevertheless, the remaining weak interaction indicated that contacts with LCb must extend through the central portion of the HC minimal domain.

A shorter fragment of LCb (77–165), comprising the predicted central HC-binding region, bound most HC fragments but produced a weaker signal than full-length LCb (Figure 1A). No binding was detected to the most C-terminal HC fragments (1480–1615 and 1523–1675), which interacted weakly with full-length LCb, suggesting that the C-terminus of HC does not interact with the central region of LCb. Additionally, LCb 77–165 interaction with HC fragments was more sensitive to the presence of HC residues 1267–1313 than full-length LCb, establishing an important contact site in that region. These data localize the minimal HC region required for efficient LC binding to between 1267 and 1522, overlapping most of the regions implicated in previous in vitro deletion studies(7) (1213– 1313, 1438–1481 and 1513–1522 (18)), and define the N-terminal boundary more precisely to between 1267 and 1313.

The central region of LCb (90–157), comprising 10 heptad repeats and previously predicted to participate in a coiled-coil interaction(14, 15), bound HC fragment 1204–1522 (Figure 1B). Binding of LCb fragments truncated at the N-terminus after residue 90, or at the C-terminus before residue 157, was dramatically reduced or not detectable. LCb fragments extending just beyond the central region bound almost equally well to HC 1204–1522, establishing 90–157 as the minimal HC-binding region in LCb. However, full extension to the LCb N- and/or C-termini always resulted in better binding, suggesting that these additional sequences facilitate LC–HC interaction. Neither the calcium-binding site (82–93 in LCb) nor the neuronal insert (155–172 in LCb) was critical for HC binding.

Parallel interaction at the C-termini of HC and LC

It has been shown by electron microscopy that monoclonal antibodies recognizing the C-terminus of LCa decorate the C-terminal vertices of clathrin triskelia(13). Moreover, mutagenesis studies indicated that the C-terminus of yeast LC interacts with the Txd of the yeast HC(19). To determine whether this is also the case for mammalian clathrin, a C-terminal HC fragment (1523–1675) comprising the vertex and the Txd was tested in combination with different LCb fragments. Only fragments containing the C-terminal portion of LCb (1–228, 77–228 and 154–228) were able to interact with the very C-terminus of HC (Figure 2). This correlates with the lack of binding by LCb 77–165 to C-terminal HC fragments shown in Figure 1A. Taken together, these results suggest that mammalian HC and LC contact in a parallel manner, at least at the C-termini, and that, similarly to yeast, the LCb-binding site in mammalian HC extends into the Txd.

Helical structure of the central region of LC is essential for HC interaction

To identify residues critical for interaction, in vitro random PCR mutagenesis(24) was used to screen for mutations, in the central region of LCb, that abolished the interaction with HC fragment 1204–1522. Among ~2 x 104 colonies screened, 28 clones that failed to give positive interactions in the yeast two-hybrid assay were analyzed further. Most of the inserts from these clones either incorporated an early stop codon (Figure 3A) or generated out-of-frame/frameshift mutations (not shown). Ten clones, shown in Figure 3A, contained in-frame inserts with mutations in protein sequence. Interestingly, most of these non-binding mutant fragments incorporated helix breakers, such as proline or glycine, into their sequences. This supports the prediction that the central region of LCb is a-helical, and suggests that maintenance of this structure is crucial for HC binding.

To probe its helix-forming propensity, we monitored the circular dichroism (CD) of LCb in the presence of increasing concentrations of trifluoroethanol (TFE). TFE is a co-solvent commonly used to increase the helicity of peptides. LCb in an aqueous buffer exhibited a CD profile indicative of mixed helix/random coil conformation (Figure 3B), consistent with previous CD analyses(17, 25). The titration of LCb with TFE promoted conformational rearrangement that reduced random coil content and increased a-helical structure, evident in the spectral shifts at peak wavelengths 208 and 222 nm. This demonstrated that LCb has a tendency to form helices. Above 30% (v/v) TFE, the structure of LCb reached a maximal helical composition and was no longer sensitive to higher TFE concentrations.

To investigate whether HC binding affects LC structure, CD spectra of LCb were compared in the presence and absence of the HC proximal leg fragment (residues 1073–1522). The CD spectrum data for the proximal leg alone (Figure 3C) was consistent with

the solved structure for residues 1210–1516, which fold into tightly packed α-helices(22). The spectrum of a stoichiometric (molar ratio 1:1) mixture of LCb and the proximal leg exhibited stronger negative peaks at 222 and 208 nm than the calculated sum of the spectra of each protein alone. This clear difference between the model for non-interacting proteins (sum of spectra) and the experimental mixture indicates that secondary structure is induced upon interaction of LCb with HC. We attribute this change to LCb because crystallographic analysis and structural predictions suggest that the HC folds independently into a superhelix throughout the proximal and distal leg segments.

K1415 of HC interacts with LCb W127 and LCa W130

One isolated LC mutant (number 16) that was unable to bind HC did not include any helix breakers, but instead harbored three substitutions, N80Y, K111E and W127R (Figure 3A). Single mutations were constructed by site-directed mutagenesis to determine which of these mutations was responsible for loss of binding to HC. We found that the mutation at residue 127, from tryptophan to arginine, completely abolished the ability of full-length LCb to interact with wild-type HC fragment 1204–1522 (Figure 4A), while the other two mutations did not affect LC binding (data not shown). Apparently, LCb W127 plays a crucial role in contacting the HC.

To find the complementary binding site of LCb W127 on the HC, suppressor mutagenesis was performed. HC fragment 1204–1522 was subjected to error-prone PCR, and the product was co-transformed into yeast with the prey vector encoding LC mutant 16. The transformants that recovered binding to mutant 16 were selected and most of these contained a mutation of lysine to glutamate at 1415 (Figure 4B). Since LC mutant 16

used in the screen also included mutations other than W127R, single mutations subsequently were generated to confirm residues responsible for W127R binding. The single mutation of K1415E in HC fragment 1204–1522 rescued binding to full-length LCb W127R in a two-hybrid plate assay (Figure 4A).

To determine whether LCa shares HC-binding properties with LCb, the tryptophan residue corresponding to LCb W127 was mutated to arginine in LCa (W130R). Similar to its effect on LCb, this mutation in full-length LCa eradicated its binding to HC fragment 1204–1522, and mutation of HC K1415E recovered the binding (Figure 4A). The mutant HC fragment, however, still bound wild-type LCa and LCb, indicating that the mutation did not affect overall HC conformation or the conformation of the binding pocket for W127 or W130. To confirm the interaction at the protein level, an in vitro binding assay was performed. Purified His6-tagged recombinant hub fragments (residues 1074–1675, wild-type or K1415E) first were immobilized on nickel affinity resin and incubated with lysates of bacteria expressing either wild-type LCb or the W127R mutant. Wild-type LCb bound efficiently to the hub fragment and the K1415E mutant hub (Figure 4C). In accord with the results described above, the LCb W127R fragment failed to bind hub, but did bind the K1415E mutant hub (Figure 4C). Furthermore, these protein assays established that LCb W127 is essential for LC binding even when additional C-terminal LC–HC contacts are present. The two-hybrid and binding data together demonstrated that HC K1415E can complement mutation of LCb W127R or LCa W130R and restore their loss of binding. However, mutation of this HC residue alone is not sufficient to abrogate LC binding by HC.

The K1326E mutation in HC rescues interaction with LCb W105R, W138R

Since the a-helical conformation of LC appears to be critical for HC interaction, we modeled the region of LCb 98–145 as an extended helix. When viewed from the C-terminus of this LC helix, two other tryptophan residues, W105 and W138, project from the same side as W127 (Figure 5A). The distance predicted between 105 and 127 is roughly twice that between 127 and 138. Given that LCb W127 is in contact with HC, we investigated whether these other two tryptophan residues also play a role in HC binding. LCb fragments with W105 and W138 residues mutagenized to arginine were generated for analysis by yeast two-hybrid assay. Single mutation of either W105R or W138R in LCb did not affect binding to HC fragment 1204–1522, yet LCb with both mutations W105R and W138R failed to interact with this HC fragment (Figure 5B, double mutant hereafter referred to as W105, 138R).

A suppressor mutagenesis screen was carried out to identify HC mutations (in fragment 1204–1522) that rescued binding to LCb W105, 138R. The K1326E mutation appeared in the sequences of almost all the rescuing clones (Figure 5C). By generating the single mutation of K1326E in the HC fragment, we confirmed that this mutation was sufficient to rescue binding to LCb W105, 138R (Figure 5B). Correspondingly, in a protein-binding assay, full-length LCb with the mutations of W105, 138R bound less efficiently to wild-type hub, but hub carrying the K1326E mutation recovered robust binding to the double mutant LCb (Figure 5D). As seen for the single K1415E mutation, the K1326E mutation in HC did not affect wild-type LCb binding. Unlike the non-binding LCb W127R mutant, some residual binding between LCb W105, 138R and hub was still detectable in vitro. This indicates that W105 and W138 are not as critical in LC–

HC interaction as W127. Because mutation of neither W105 nor W138 alone was sufficient to abrogate HC binding, it is reasonable that a single mutation in HC restored the interaction with the double mutant in the two-hybrid assay. In fact, K1326E binding to LCb W105, 138R was sufficient to survive the challenge of 3-aminotriazole up to 60 mM (data not shown), a stringency test for this assay. Given the parallel orientation between the C-termini of HC and LC and the aligned interaction between HC K1415 and LCb W127, we predict that HC K1326E is most probably rescuing interaction with LC W105R, and not with W138R.

Structural model of HC and LC interaction

The minimal HC-binding regions of LCa (93–160) and LCb (90–157) were modeled as a-helices and manually docked to the crystallographically determined structure of the HC proximal leg, fixing the mapped interaction pairs of HC K1415 with LCb W127 or LCa W130, and HC K1326 with LCb W105 or LCa 108, as constraints (Figure 6A). To create molecular models of the interfaces, a computer-generated simulation was performed for each LC–HC combination. A total of 150 copies of each protein fragment pair were equilibrated at 2000 K with the LC side chain atoms reduced in atomic radius, bond lengths, and electrostatic and van der Waals energies. The equilibrated molecules were then cooled slowly to 300 K, while growing the LC side chains to their original sizes and energies. During this annealing phase, the distances of the LC backbone carbonyls and amides were weakly restrained at 2.86 ± 0.5 Å in order to promote the helicity of the LCs. Additional restraints were applied between carbons of the aromatic rings of the LC tryptophan residues and their partner lysine amines. For each LC, a probability density map was constructed from the 150 model structures generated

by this molecular dynamics approach, and a single model of the HC interface was best fit using crystallographic refinement techniques(26). In the resulting models, the HC-binding region of each LC was characterized as two contiguous a-helical segments that curve slightly along the length of the HC and bind via an essentially hydrophobic face. This is consistent with the prediction that the central region of the LCs forms an amphipathic helix. However, the models indicate that the primary LC-binding interface lies along the loop regions of the HC, rather than one of the helical faces, and the two subunits do not participate in a conventional coiled-coil interaction.

A closer look at the region predicted to surround LCb W127 (or LCa W130) showed the proximity of the LC tryptophan to HC K1415 (Figure 6B and D), consistent with the mutagenesis and binding data reported above. Notably, even when the constraint of LCb W127 or LCa W130 being near HC K1415 was dropped, the LC–HC docking models generated continued to place these residues in contact in the complexes. In these models, the LC tryptophan apparently stabilizes HC K1415 in an overall aromatic pocket of the HC core. The packing of the hydrocarbon arm of the lysine against the HC hydrophobic core was increased substantially from the unbound state, and the amine was desolvated, allowing it to form strong cation–π interactions(27) with HC F1410 and the LC tryptophan.

Interestingly, while LCb W105 or LCa W108 were near HC K1326 in the models, they were too distant to be interacting directly. The predicted distance was ~8 Å while that between HC K1415 and LCb W127 (or LCa W130) was ~3 Å. In addition, the amine of K1326 was fully solvated, mitigating any interaction the charged group could have with an aromatic residue. Rather, LCb W105/LCa W108 appeared to be interacting with

a pair of phenylalanine residues (F1296 and F1327), and LCb 101/LCa 104 were hydrogen-bonded with the backbone of the HC (Figure 6C and E). This is consistent with the mutagenesis data, which demonstrated a relatively weak interaction between W105 and the HC, but localized it to the area of K1326. It is important to note that the weak restraints used in the simulation were not sufficient to force an unfavorable interaction between the two residues in spite of their close proximity.

Discussion

Combining the screens of reverse two-hybrid and suppressor mutagenesis, we have established a novel approach to investigating the structural details of protein–protein interaction. This analysis has positively defined the essential regions for interaction of the clathrin HC and LC subunits, which were previously only inferred from negative deletion and proteolysis analysis. In addition, we have identified several molecular features critical for HC–LC binding. These data allowed the development of a three-dimensional model for the interface of HC and LC using molecular dynamics. The model suggests how the entire LC is oriented on the clathrin triskelion and has implications for LC control of clathrin assembly, discussed below.

Clathrin subunit interaction

Sequence analysis suggested that the central region of clathrin LCs forms amphipathic helices(14, 15). Earlier CD spectral analyses indicated a high a-helical content for purified clathrin (HC–LC complex) and a significantly lower content for isolated LCs(17, 25). The CD spectral analysis reported here demonstrates that free LC increases its helical structure in solvents favorable to helix formation or when mixed with its partner HC fragment. This is consistent with the mutagenesis data showing disruption

of LC–HC binding by helix breakers. The emerging picture is that, during cellular biosynthesis of a clathrin triskelion, unstructured regions of LC fold in concert with docking to clathrin HC. This is similar to other well-characterized cases of protein folding induced by binding(28).

The contact mapped between HC 1267–1522 and LCb 90–157 is fundamental for the core interaction of triskelion subunits. The structure of the proximal leg domain of HC is a superhelical rod of modular constituents [clathrin HC repeats (CHCRs)], each composed of 10 winding antiparallel α-helices(22). Based on the axial length of HC fragment 1210–1516 (115 Å), each residue spans an average 0.375 Å. Accordingly, the HC residues 1267–1516, involved in minimal LC binding, span ~100 Å, which accommodates LCb residues 90–157 (or LCa residues 93–160) modeled into an extended a-helix of 100 Å (Figure 6A). Thus the core LC–HC interaction covers ~60% of the proximal leg of the triskelion (Figure 7). Interestingly, the extended LC helix can be visualized as two contiguous helices of ~50 Å, each in contact with the loops of 10 HC helices, the length of a CHCR. Each LC helix segment is encoded by a separate exon(29), possibly reflecting complementary structural duplication in the two subunits.

Interactions independent of core binding were detected between the C-termini of HC and LCb, contributing positively to HC–LC binding affinity. This result is consistent with an earlier observation that the C-terminus of LC enhances its interaction with mammalian HC(21). In studies of yeast clathrin mutants, LC-binding regions in HC were demonstrated within the proximal leg and within the C-terminal Txd(19), and both sites were required for subunit interaction. We show here that mammalian LCb makes similar contacts, although it binds to the HC core in the absence of Txd sequences, suggesting

stronger affinity between the mammalian subunits. Correspondingly, by two-hybrid analysis, we have observed that the whole extent of yeast hub (a trimer) is required for detectable yeast LC binding (C.-Y.Chen and F.M.Brodsky, unpublished data), while mammalian LCs bind shorter monomeric HC fragments (Figure 1). Specific binding between the vertex segment of HC and the C-terminus of LC reveals a parallel orientation of their interaction domains.

Molecular features of the HC–LC interface

A primary feature of our molecular models of LC–HC interaction is that the LC helix interacts with a surface comprised almost entirely of loops on the HC. The helix–turn–helix–loop structure of the HC proximal domain is a 24 × 28 Å oval in cross-section(22) (Figure 6A). This creates two major helical faces spanning the length of the fragment and two thinner faces of loops between them. The LC contact points identified in this study, K1326 and K1415, are both located on one HC loop face (Figure 6A).

The region of the loop interface containing the K1415 pocket for W127 or W130 was well defined in both models of LC–HC interaction, as demonstrated by their low B-factors, values that indicate the degree to which their positions are consistent in each individual structure determined (data not shown). The docking of LC W127/130 to HC K1415 was characterized by stabilizing cation–π interactions, common in protein interfaces(30). In addition, tryptophan is the most enriched residue in hot spots of protein–protein contact(31), suggesting a key role in binding specificity. Mutation of the central tryptophan, in the context of the full-length LC, completely abolished binding to hub (Figure 4), demonstrating that the cation–p interactions surrounding LC W127/130

provide both the specificity and a large portion of the binding energy that drive subunit interaction.

The proximity constraint for LCb W105 and HC K1326 generated models in this interface region which aligned neighboring LC residues with the HC loop face (e.g. LCb S101 forms three hydrogen bonds with the HC backbone at residues G1294, F1296 and E1297). The predicted hydrophobic pocket for HC binding to LCb W105 or LCa W108 was adjacent to but did not include K1326. Further modeling of the subunits with the mutations showed that HC K1326E rescued the binding of LCb W105R because the side chain of the arginine in LCb W105R could form a salt bridge with HC K1326E (data not shown). LCb W105 or LCa W108 apparently contributes less affinity to the HC–LC interface than the cation–p interaction of W127/130. LCb W105R reduced HC binding only when coupled to W138R, and LCa W108R reduced but did not eliminate HC binding (data not shown). The single mutation in LCa of W141R also reduced HC binding (data not shown). Thus, the third tryptophan (LCb W138 or LCa W141), that aligns with the other two (Figure 5A), is also located near or in the LC–HC interface. The slight displacement of this tryptophan relative to the HC interface is evident in both LC–HC models and emphasizes the curvature of the LC axis relative to the HC axis. This curvature might explain how distant mutations such as W105R and W138R could be synergistic in disrupting binding by altering the LC twist.

Though the models of the region around HC K1326 are apparently similar for LCb and LCa, binding of the LCa W108R mutant was not restored by the HC K1326E mutation (data not shown). This could be attributed to the subtle differences observable in the predicted disposition of each LC relative to the HC superhelix. Although such

differences could be due to modeling limitations, they may actually reflect functional differences resulting from sequence divergence of LCa and LCb in their HC-binding regions.

Implications of the HC–LC interface for clathrin assembly

From analysis of the 21 Å resolution cryo-electron microscopy structure of a clathrin coat(32), it has been suggested that the proximal–proximal interactions between clathrin HCs involve the loops of the HC superhelix(33). However, our data clearly demonstrate that one of the loop 'edges' of the HC is occupied by the LC. It has been shown previously that LCs are readily dissociated from assembled clathrin baskets and are oriented on the lattice surface. Thus HC–HC interactions in the assembled lattice do not directly involve LCs, except for regulatory interactions. Consequently, we propose, based on our modeling, that the HCs most probably interact via their helical faces. This interaction would orient the LCs to the surface of the lattice, yet give them access to the interface to regulate assembly.

Our combined data and modeling establish the relative orientations of regulatory domains in mammalian LC with respect to the HC triskelion (Figure 7). This orientation suggests mechanisms of LC regulation of clathrin assembly in response to local intracellular calcium fluxes. Our model locates the predicted EF-hand of the LC Ca2+-binding domain(34) towards the N-terminus of the HC proximal leg. Structural EF-hands (e.g. those in calmodulin) experience large conformational changes upon Ca2+ binding. If this were also true for the LC EF-hand, it could act as a switch to control the N-terminus of LC folding back towards the C-terminus, along the segment of HC that is involved in the HC–HC interface. This bent LC conformation, predicted by antibody

blocking studies(18), potentially could inhibit proximal leg interaction with either the proximal or distal legs of neighboring triskelia by orienting residues in the N-terminal LC conserved region that have a negative regulatory effect on clathrin assembly(4). All LCs contain the EF-hand motif, from mammalian to yeast, supporting the suggestion that the sensitivity of LCs to Ca2+ concentration is involved in their function. Although the binding constant of LC for Ca2+ is relatively weak compared with that of calmodulin, it is conceivable that local intracellular Ca2+ concentrations could rise to a level that might influence clathrin assembly.

In order to transform from a planar lattice into a three-dimensional basket, the clathrin triskelia involved must undergo a change in pucker(35), i.e. the relative angles of the legs to the plane of the trimerization domain must be altered. These angles may be controlled by Ca2+ concentration via calmodulin binding. Our data corroborate previous studies suggesting that the calmodulin-binding site of the LC C-terminus is near the triskelion vertex(3). With the binding of calmodulin, the triskelion vertex would become substantially more rigid, constraining the angle that triskelion legs could assume. Calmodulin appears to bind only the neuronal forms of LC in the context of a CCV(3), suggesting that such a mechanism could contribute to the uniformity of size observed for neuronal CCVs compared with CCVs in other cell types.

Our models place the Src family kinase HC phosphorylation site at Y1477(36) near the C-terminus of the minimal HC–LC interface. This proximity allows the possibility of additional regulation of clathrin self-interaction through phosphorylation of this site and subsequent modification of the HC–LC interface. An orchestration of these

regulatory mechanisms for clathrin self-assembly probably contributes to controlling CCV formation.

Introduction: ***Towards an NMR structure of the LC "auto-inhibitory" complex***

Clathrin-mediated endocytosis (CME) facilitates the internalization of cell surface molecules, including hormones and cell surface receptors. Neuron-specific expression of endocytic proteins also promotes CME to recycle synaptic vesicles during neurotransmission(37, 38), but how these proteins couple exocytosis with endocytosis and accelerate endocytic rates remains obscure. In vertebrates, clathrin subunits are expressed in a tissue-dependent manner. Two clathrin light chain (LC) subunits, LCa and neuron-enriched LCb, regulate heavy chain (HC) assembly *in vitro* through an invariant LC N terminal EED motif. These residues may suppress HC assembly by either binding HC directly or interacting with another LC region to modulate HC indirectly. We have found, through coprecipitation assays with GST LC fusions, that the N and C termini of LCs, which flank the HC binding region, interact in a LC-specific manner. N-C termini LC-LC interactions are enhanced with LCb phosphorylation and are disrupted by Ca^{2+}-calmodulin. We hypothesize that distinct LC-specific N-C termini interactions differentially regulate tissue-specific clathrin assembly. In contrast to adaptor proteins(39), such as AP-2, that function in receptor-mediated endocytosis, LCs are uniquely poised to robustly regulate synaptic clathrin because they are present in stoichiometric amounts to HC. Thus LCb may be a major protein that accelerates HC assembly and endocytic rates at the presynaptic terminal.

Results and Discussion

Reviewing previous work demonstrating the interaction of calmodulin with the C-terminal region of both LCa and LCb(3), it seemed possible that the N-terminus of the LC was competing with calmodulin for its C-terminal binding site. To test for this

interaction, we designed constructs of human LCa (**Figure 8a**) containing sequences derived from either the N- or C-terminus. GST-tagged N-terminal fragments of LCa were then used to pull-down HA-tagged C-terminal fragments (**Figure 8b**). We defined the region of interaction to N-terminal residues 1-62 of LCa and C-terminal residues 148-248. Interestingly, the N-terminal residues include the most phylogenetically conserved LC sequence, the LC consensus that contains the EED motif involved in negatively regulating HC assembly. As expected, the C-terminal residues responsible for the LC N-/C-terminal interaction include the calmodulin-binding domain. By adding calmodulin to the pull-down conditions, we were able to inhibit the LCa N-/C-terminal complex in a calcium-dependent manner (**Figure 8c**). This raises the possibility that the interaction between the LCa termini acts physiologically to integrate Ca^{2+}-signals, triggering clathrin cage formation.

To confirm the binding results and define the minimal sequences required to constitute the N-/C-terminal interaction, we analyzed the binding in solution by NMR spectroscopy. N- and C-terminal LC sequences were fused to staphylococcal GB1 to protect the sequences from proteolysis. These proteins were expressed as both ^{15}N-labelled and NMR-inactive species. As the NMR chemical shifts are exquisitely sensitive to chemical environment, changes in a spectrum upon titration with a ligand can be used to monitor binding. In addition, the resolution of the binding event is on the atomic order; only peaks affected by the binding event will exhibit chemical shift changes.

^{15}N-GB1-LCa 1-62 exhibited small, but significant, spectral changes upon titration with GB1-LCa 148-248 (**Figure 9**), with a K_d on the order of 0.5-1 mM. As both

the N- and C-terminal regions of LCa are predicted to form α-helices (**Figure 8a**), we repeated the titrations in 25% trifluoroethanol (TFE), which we have previously shown to stabilize helicity in LC (40). As expected, stabilizing the secondary structure of the LC served to stabilize the interaction. The majority of the peaks in the spectra recorded in TFE (**Figure 10a,b**) exhibited changes upon titration with the C-terminal LCa, indicating that the majority of the N-terminal residues are sensitive to interaction with the C-terminus.

To avoid problems with the poor spectral dispersion characteristic of unfolded and helical proteins, GB1-LCa 148-248 was specifically labeled with ^{15}N-Lys. Titration in 25% TFE with unlabeled GB1-LCa 1-62 demonstrated that 3 Lys are strongly affected by the N/C-terminal interaction (**Figure 10c**). There are 4 Lys present in the calmodulin-binding domain of LCa. These data therefore suggest that LCa residues 200-248, which contain the calmodulin-binding domain, also comprise the LC consensus-region interaction domain.

As each of the N- and C-terminal regions identified above contain two predicted helices in approximately 40 amino acids of sequence, it is attractive to consider the possibility that the interaction occurs via a coiled-coil motif in which the acidic N-terminal consensus sequence masks the basic region of the C-terminus responsible for recruitment of calmodulin. Preliminary analysis of an ^{15}N-NOESY-HSQC of ^{15}N-LCa 220-248 bound to unlabeled GB1-LCa 20-62 in 20% TFE demonstrates that in these conditions, the LC does, in fact, form helices. Further work towards modeling the solution structure of the N-/C-terminal complex is outlined below.

Future Directions

As it is particularly difficult to gain structural information on weakly interacting, transient protein complexes by crystallography, we have begun structural studies of the LCa N-/C-terminal complex by NMR. It is difficult, however, to gain long-range distance restraints based on NOE measurements on helical proteins, as the backbone amides fulfill their hydrogen bonds in forming local secondary structure. Recently, long-range restraints derived from paramagnetic relaxation enhancement (PRE; see Ch 2. pp. 24, 34) have been demonstrated sufficient to provide a high resolution fold of a helical protein(41). PREs are measured as broadening of specific peaks that fall off with a sixth-order dependence on distance from the paramagnetic center. By labeling specific positions of the N- and C-termini of the LC constructs (labeled with arrowheads in **Figure 8a**) through introduction of cysteines by mutagenesis and subsequent labeling with a cysteine-specific nitroxide label, we expect to gain sufficient restraints to orient the complex. This is similar to the orientation we determined with sparse genetic restraints for the HC/LC complex, presented above. The model of the LC N-/C-terminal complex may be further refined through the use of orientational restraints derived from residual dipolar couplings(42).

In addition to modeling the N-/C-terminal complex *in vacuo*, we will examine the complex in the context of the clathrin cage. Using our previously generated model of the HC/LC interface(40) and the 8Å cryoelectron microscopy model of the clathrin cage(5) as a starting point, a model of the N-/C-terminal complex can be generated using either intra- (the termini of the same polypeptide chain) or intermolecular (the termini of LCa bound to different triskelia) restraints. These studies will provide insight as to the actual

function of the complex formation. For example, does the complex serve to inhibit and/or promote assembly? As the intermolecular interaction could only occur in the context of a clathrin cage (and may therefore more appropriately be called "quasi-intramolecular"), it would be more likely to stabilize the assembled cage rather than initiate cage formation.

Importantly, the model of the complex may be used to generate mutations that modify the strength of the N-/C-terminal interaction without disrupting other known interactions (*e.g.* the LC interaction with calmodulin or Hip1(1)). Such constructs could be used to probe *in vitro* assembly of clathrin cages and *in vivo* constitutive endocytosis in cell culture. Correlation of the results of such studies may suggest a molecular mechanism for the trigger of constitutive clathrin-mediated endocytosis.

Methods

Yeast two-hybrid assays

HC fragments were amplified from bovine brain cDNAs(7) by PCR and cloned into the BamHI–SalI sites of pGBT9 (Clontech). LCb fragments were amplified from bovine brain cDNAs(14) and cloned into either the EcoRI–BamHI sites of pGAD424 (Clontech) or the BamHI site of pACT2 (Clontech). Similarly, bovine LCa(14) was amplified and cloned into the EcoRI site of pGAD424. Single residue mutations were generated with the quick-change site-directed mutagenesis kit (Stratagene). Yeast transformation (SFY526 or AH109), quantitative liquid ß-galactosidase [using o-nitrophenyl ß-d-galactopyranoside (ONPG) as substrate], filter and plate growth assays were performed following the Clontech yeast protocols handbook. For each indicated two-hybrid interaction, three different colonies were picked and each was assayed in duplicate.

PCR mutagenesis screens

Error-prone random PCR mutagenesis was carried out as described previously(24). Briefly, target DNA was amplified using Taq polymerase (Promega) in a reaction mixture containing 3–6 mM MgCl2, 0.1–0.6 mM MnCl2, 1 mM dGTP, dCTP, dTTP and 0.2 mM dATP. The designed primers generated 85, 88 bp (or 90, 118 bp for suppressor screens) of homology flanking the gap of the vector for the recombination. In the reverse two-hybrid assay, the mutated PCR products of LCb 77–157 were transformed with gapped pGAD424 into AH109 cells carrying pGBT9 HC 1204–1522. Approximately 2 × 104 transformants appeared on SD –Leu –Trp plates after 3 days at 30°C and were then replica-plated onto SD –Leu –Trp –His –Ade plates. About 73 colonies that did not grow on the second plates were selected and tested again by streaking. The prey vectors from

28 confirmed colonies were isolated and the inserts were sequenced. Nine clones contained the exact length of the in-frame inserts with mutations, one with a longer insert (listed in Figure 3A), and the rest incorporated early stops or out-of-frame mutations. In the suppressor screens, the mutated PCR products of HC 1204–1522 and gapped pGBT9 were co-transformed into AH109 cells carrying either the prey mutant 16 or pGAD424 LCb W105, 138R. Transformants were selected on SD –Leu –Trp –His –Ade plates, and for each screen >90% of the rescued recombinant bait vectors shared the same mutation.

Circular dichroism spectroscopy

His6-tagged proximal leg (1073–1522 in pET15b) was expressed in BL21 (DE3) cells and purified using Ni2+ affinity resin, followed by size exclusion chromatography, as described previously(7). Neuronal LCb co-expressed with the hub in pET15b(7) was partially purified by boiling and centrifugation(18). CD spectra were obtained using a JASCO 715 spectropolarimeter equipped with a Peltier type cell holder for accurate temperature control. All measurements were conducted at 21°C in a 2 mm path length cell with protein concentrations of 0.5–1.5 μM in 20 mM sodium phosphate buffer, pH 7.1. Three to four averages were taken for each spectrum with the following parameters: step resolution, 0.2 nm; speed, 50 nm/min; response, 1 s; bandwidth, 1.0 nm. All spectra were corrected by subtraction of the corresponding spectra of blank samples.

Protein binding assay

Bovine hub fragment with an N-terminal His6 tag (in pET15b) were expressed in bacteria and affinity purified using nickel resin, as described previously(7). Bovine brain LCb cDNA was cloned into the NcoI–BamHI sites of pET15b (Novagen) to leave out the His6 tag. LCb constructs were prepared in phosphate-buffered saline [pH 7.4, including 1 μg/μl protease inhibitors and 0.2 mM phenylmethylsulfonyl fluoride (PMSF)] as crude bacterial lysates. Protein concentration was determined by micro-BCA assay (Pierce). The LC-binding assay was modified from Liu et al. (7). About 100 μg of purified recombinant hub fragment were bound to 100 μl of 50% Ni2+ affinity resin in 50 mM Tris–HCl pH 7.9 for 30 min. Lysates of bacteria expressing LCb (400 μg) were added to the bound resin and incubated for at least 30 min. After washing three times, bound proteins were eluted with 50 mM Tris–HCl pH 7.9, 0.5 M NaCl, 0.5 M imidazole. The bound fractions were subjected to SDS–PAGE and immunoblotted with a rabbit polyclonal antiserum against the conserved region of LCs (for LCb detection)(12) and a mouse monoclonal antibody to the His6 tag (Clontech) or a rabbit polyclonal antiserum against the purified recombinant proximal leg of HC (produced by a contract laboratory for hub detection). Following incubation with horseradish peroxidase-conjugated goat anti-mouse/rabbit antiserum (ZyMed), the binding was detected by enhanced chemiluminescence (Amersham).

Modeling HC–LC interaction

Using the interaction pairs of LCb W105-HC K1326 and LCb W127-HC K1415 (or the corresponding W108, 130 in LCa) as guides, a helical model created in Quanta (Accelrys, Inc.) of LCb 90–157 or LCa 93–160 was docked manually to the HC proximal leg crystal

structure(22) by Molmol(43). This docked model served as the starting point for the generation of an ensemble of 150 structures, which was refined to a single best-fit through multiconformational simulated annealing pseudo-crystallographic refinement(26). The final structure (Figure 6A) was refined against a pseudo-density map generated from the ensemble using standard crystallographic techniques.

Nuclear Overhauser effect (NOE) restraints were assigned between the Nζ and Cε of the HC lysine residues (K1326 and K1415) and each of the carbons of the aromatic ring of the partner tryptophan residues (W105/108 and W127/130, respectively). Structures were also calculated without the restraints in place to determine the degree to which the restraints forced the outcome (data not shown). These unrestrained structures were identical to the restrained structures calculated, within error. Additional NOE restraints were placed between the backbone carbonyls and amides of the LC in order to favor its helical structure. All calculations were carried out in X-PLOR, using the OPLS force field for polar hydrogens and the TIP3P model for water.

Plasmid construction

LCa and LCb were respectively PCR amplified from human lymphocyte or brain cDNA library. To obtain GST fusions of full length and fragment LCs, the corresponding sequences were PCR amplified and ligated into the EcoRI/XhoI sites of pGEX4T-2. GB1-His_6 constructs were created by ligating the PCR amplified N-terminal LCa constructs into the NdeI site of GB1-parallel and C-terminal LCa constructs into the EcoRI/XhoI sites of GB1-parallel.

Recombinant protein production and purification

All constructs were expressed in BL21 Rosetta2 pLysS (Strategene). Bacteria were grown in LB at 37°C to OD_{600}=0.8 and induced with 0.3mM IPTG for 3hrs at 30°C. ^{15}N-labeled protein was expressed by growing cells as above in 3L LB to OD_{600}=0.8, sedimenting, and transferring to 1L M9 media with 0.4% glucose and 1g $^{15}NH_4Cl$ (Cambridge Isotopes) and growing overnight at 18°C. ^{15}N-Lys labeled protein was expressed in 1L M9 with unlabeled NH_4Cl supplemented with 100mg ^{15}N-Lys (Cambridge Isotopes). His_6-fusion proteins were lysed by sonication in 50mM HEPES pH7.5, 0.5M NaCl, 15mM imidazole and purified on 1ml NiNTA resin (QIAGEN) by elution with 150mM imidazole, after extensive washing with lysis buffer. Purity was assessed by SDS-PAGE. Protein used in NMR was dialyzed extensively against water and concentrated by lyophilization.

Clathrin LC coprecipitation assays.

GST fusions of full length and truncated LC proteins (4 nmol) were immobilized on 20 µl of Glutathione Sepharose 4 fast flow (GE Healthcare) in 20 mM Hepes, pH 7.0, 100 mM NaCl, 0.02% Triton X-100, 5 mM DTT and 0.1 M PMSF at 4°C. Unbound GST-LCs were removed through two batch washes and the resin was mixed with 0.5 ml, 1 µM HA-tagged LC fragment in HEPES buffer for 2 h. The resin was subsequently washed three times, mixed with SDS PAGE sample buffer and boiled for analysis by SDS PAGE and Western blotting. In a subset of LC coprecipitation assays, calmodulin (Sigma) was first incubated with immobilized GST-LCa for 1h with mixing in the absence or presence of 0.5 mM $CaCl_2$. Subsequently $HisNLCa_{1-228}$-HA, which lacks the calmodulin binding domain, was added, incubated 1 h and processed as outlined above.

NMR Methods

All spectra were measured on a Bruker Avance NMR spectrometer operating at a proton resonance frequency 800 MHz and equipped with a cryogenic probe. Binding during the titration series was monitored with [^{15}N,^{1}H]-HSQC experiments, with a spectral width of 12.8 ppm in the proton and 32 ppm in the ^{15}N dimension. 128 complex points were collected in the indirect dimension and 512 complex points in the direct dimension with either 4 or 8 scans per increment depending on protein concentration. Titrations were carried out in 3 steps, each one adding 1 mol equivalents of the peptide in a range from 1:1 to 1:2 N-term:C-term molar ratio. The experiments were conducted at 28°C.

References

1. Chen, C. Y. & Brodsky, F. M. (2005) *J Biol Chem* **280,** 6109-17.
2. Xiao, J., Dai, R., Negyessy, L. & Bergson, C. (2006) *J Biol Chem* **281,** 15182-93.
3. Pley, U. M., Hill, B. L., Alibert, C., Brodsky, F. M. & Parham, P. (1995) *J Biol Chem* **270,** 2395-402.
4. Ybe, J. A., Greene, B., Liu, S. H., Pley, U., Parham, P. & Brodsky, F. M. (1998) *Embo J* **17,** 1297-303.
5. Fotin, A., Cheng, Y., Sliz, P., Grigorieff, N., Harrison, S. C., Kirchhausen, T. & Walz, T. (2004) *Nature* **432,** 573-9.
6. Brodsky, F. M., Chen, C. Y., Knuehl, C., Towler, M. C. & Wakeham, D. E. (2001) *Annu Rev Cell Dev Biol* **17,** 517-68.
7. Liu, S. H., Wong, M. L., Craik, C. S. & Brodsky, F. M. (1995) *Cell* **83,** 257-67.
8. ter Haar, E., Musacchio, A., Harrison, S. C. & Kirchhausen, T. (1998) *Cell* **95,** 563-73.
9. Ungewickell, E. & Ungewickell, H. (1991) *J Biol Chem* **266,** 12710-4.
10. Collins, B. M., McCoy, A. J., Kent, H. M., Evans, P. R. & Owen, D. J. (2002) *Cell* **109,** 523-35.
11. Scheele, U. & Holstein, S. E. (2002) *FEBS Lett* **514,** 355-60.
12. Acton, S. L. & Brodsky, F. M. (1990) *J Cell Biol* **111,** 1419-26.
13. Kirchhausen, T., Harrison, S. C., Parham, P. & Brodsky, F. M. (1983) *Proc Natl Acad Sci U S A* **80,** 2481-5.
14. Jackson, A. P., Seow, H. F., Holmes, N., Drickamer, K. & Parham, P. (1987) *Nature* **326,** 154-9.
15. Kirchhausen, T., Scarmato, P., Harrison, S. C., Monroe, J. J., Chow, E. P., Mattaliano, R. J., Ramachandran, K. L., Smart, J. E., Ahn, A. H. & Brosius, J. (1987) *Science* **236,** 320-4.
16. Brodsky, F. M., Hill, B. L., Acton, S. L., Nathke, I., Wong, D. H., Ponnambalam, S. & Parham, P. (1991) *Trends Biochem Sci* **16,** 208-13.
17. Ungewickell, E. (1983) *Embo J* **2,** 1401-8.
18. Nathke, I. S., Heuser, J., Lupas, A., Stock, J., Turck, C. W. & Brodsky, F. M. (1992) *Cell* **68,** 899-910.
19. Pishvaee, B., Munn, A. & Payne, G. S. (1997) *Embo J* **16,** 2227-39.
20. Brodsky, F. M., Galloway, C. J., Blank, G. S., Jackson, A. P., Seow, H. F., Drickamer, K. & Parham, P. (1987) *Nature* **326,** 203-5.
21. Scarmato, P. & Kirchhausen, T. (1990) *J Biol Chem* **265,** 3661-8.
22. Ybe, J. A., Brodsky, F. M., Hofmann, K., Lin, K., Liu, S. H., Chen, L., Earnest, T. N., Fletterick, R. J. & Hwang, P. K. (1999) *Nature* **399,** 371-5.
23. Vidal, M. & Endoh, H. (1999) *Trends Biotechnol* **17,** 374-81.
24. Muhlrad, D., Hunter, R. & Parker, R. (1992) *Yeast* **8,** 79-82.
25. Winkler, F. K. & Stanley, K. K. (1983) *Embo J* **2,** 1393-400.
26. Ota, N. & Agard, D. A. (2001) *J Mol Biol* **314,** 607-17.
27. Dougherty, D. A. (1996) *Science* **271,** 163-8.
28. Wright, P. E. & Dyson, H. J. (1999) *J Mol Biol* **293,** 321-31.
29. Stamm, S., Casper, D., Dinsmore, J., Kaufmann, C. A., Brosius, J. & Helfman, D. M. (1992) *Nucleic Acids Res* **20,** 5097-103.

30. Gallivan, J. P. & Dougherty, D. A. (1999) *Proc Natl Acad Sci U S A* **96,** 9459-64.
31. Bogan, A. A. & Thorn, K. S. (1998) *J Mol Biol* **280,** 1-9.
32. Smith, C. J., Grigorieff, N. & Pearse, B. M. (1998) *Embo J* **17,** 4943-53.
33. Kirchhausen, T. (2000) *Annu Rev Biochem* **69,** 699-727.
34. Nathke, I., Hill, B. L., Parham, P. & Brodsky, F. M. (1990) *J Biol Chem* **265,** 18621-7.
35. Nossal, R. (2001) *Traffic* **2,** 138-47.
36. Wilde, A., Beattie, E. C., Lem, L., Riethof, D. A., Liu, S. H., Mobley, W. C., Soriano, P. & Brodsky, F. M. (1999) *Cell* **96,** 677-87.
37. Sankaranarayanan, S. & Ryan, T. A. (2001) *Nat Neurosci* **4,** 129-36.
38. Ryan, T. A. (2006) *Curr Opin Cell Biol* **18,** 416-21.
39. Edeling, M. A., Smith, C. & Owen, D. (2006) *Nat Rev Mol Cell Biol* **7,** 32-44.
40. Chen, C. Y., Reese, M. L., Hwang, P. K., Ota, N., Agard, D. & Brodsky, F. M. (2002) *Embo J* **21,** 6072-82.
41. Liang, B., Bushweller, J. H. & Tamm, L. K. (2006) *J Am Chem Soc* **128,** 4389-97.
42. Lipsitz, R. S. & Tjandra, N. (2004) *Annu Rev Biophys Biomol Struct* **33,** 387-413.
43. Koradi, R., Billeter, M. & Wuthrich, K. (1996) *J Mol Graph* **14,** 51-5, 29-32.
44. Humphrey, W., Dalke, A. & Schulten, K. (1996) *J Mol Graph* **14,** 33-8, 27-8.

Figure Legends

Figure 1: Mapping minimal regions of clathrin HC and LC interaction. (A) Yeast SFY526 cells were co-transformed with the indicated bovine HC fragments (in pGBT9) plus full-length bovine brain LCb or the LCb fragment 77–165 (in pACT2). Quantitative ß-galactosidase (ß-gal) assays were performed, and results are shown in units as the mean ± SD of triplicate determinations. Above the fragments tested, a diagram of full-length HC is delineated, indicating the terminal, distal, proximal, trimerization and extreme C-terminal domains (Txd + C). Black bars below the HC diagram indicate sites previously implicated in LC binding (1213–1313, 1438–1481 and 1513–1522)(7, 18). Hatched lines within each fragment highlight the boundaries of the minimal region in HC (1267–1522) required for LC binding, mapped by these studies. (B) SFY526 cells were co-transformed with bovine HC fragment 1204–1522 (in pGBT9) and the indicated bovine brain LCb fragments (in pGAD424). Results of liquid ß-gal assays are shown in units as the mean ± SD of triplicate determinations. Above the fragments tested, a diagram of full-length bovine brain LCb is delineated, indicating the phosphorylation domain (P), conserved region (Cons), calcium-binding site (Ca), previously predicted HC-binding site (HC), neuron-specific insert (N) and calmodulin-binding site (Cam). Hatched lines within each fragment highlight the boundaries of the minimal region in LCb (90–157) required for HC binding, mapped by these studies. The ß-gal units shown in (A) are generally higher than those in (B) because LC fragments were constructed in different prey vectors in which pACT2 (for A) has a higher expression level than pGAD424 (for B).

Figure 2: Parallel interaction at the C-termini of clathrin heavy and light chains. Plate growth and ß-galactosidase (ß-gal) filter assays for the interactions between HC 1523–

1675 and various LCb fragments. The boundaries of the LCb fragments tested are diagrammed relative to the full-length LCb sequence, with the functional domains delineated as defined in Figure 1B. For the plate growth assay, yeast AH109 cells were co-transformed with HC 1523–1675 (in pGBT9) and the indicated LCb fragments (in pACT2). The transformants were spotted onto plates lacking Leu and Trp, with (+) or without (–) His and Ade. Trans formants expressing interacting constructs grew in the absence of His and Ade after 3–5 days. For the ß-gal filter assay, SFY526 cells were co-transformed with HC 1523–1675 (in pGBT9) and the indicated LCb fragments (in pACT2). Transformants were spotted onto filters directly in contact with media and incubated for 2 days. Blue color indicating positive interaction developed within 6 h following substrate application in the filter assay. The assays shown are typical of three independent experiments.

Figure 3: Requirement for an a-helical conformation of clathrin LC during interaction with HC. (A) Sequences of the mutant fragments of bovine brain LCb 77–157 that were unable to bind HC. Yeast AH109 cells harboring HC 1204–1522 (in pGBT9) were co-transformed with mutated PCR products of LCb 77–157 and a gapped prey vector, pGAD424. The inserts of the recombinant clones (numbered mutants) unable to interact with HC 1204–1522 were sequenced. Mutated residues are in color and, among those, helix breakers are shown in green. (B) Far-UV CD spectra of LCb in the presence of TFE. Samples of LCb at a final concentration of 1.5 μM were incubated in 20 mM sodium phosphate buffer pH 7.1 with varying concentrations of TFE (0–60% v/v). CD intensity is represented as mean residue molar ellipticity. (C) Gain of signal for the LCb–proximal leg complex recorded by far-UV CD. The concentration of each indicated

protein was 0.5 μM in 20 mM sodium phosphate buffer pH 7.1. The spectra of the isolated components, LCb, proximal leg and the complex are shown. The sum of the spectra of the two isolated proteins is also shown for comparison (dotted line). To facilitate the direct comparison of components with different numbers of residues, the CD intensity was calculated using the protein molar concentration rather than on a per residue basis.

Figure 4: Mutation of tryptophan to arginine at clathrin LC residue 127 in LCb or 130 in LCa abrogates binding to HC, which is rescued by HC mutation K1415E. (A) Plate growth assay of yeast AH109 cells co-transformed with HC fragment 1204–1522 (in pGBT9) and LCb/LCa (in pGAD424), with or without the indicated mutation. The transformants were spotted onto plates lacking Leu and Trp, with (+) or without (–) His. Transformants expressing interacting constructs grew in the absence of His after 2 days. (B) Sequences of bovine HC 1204–1522 mutants that rescued the binding to the LC mutant 16. AH109 cells harboring LCb mutant 16 (in pGAD424) were co-transformed with mutated PCR products of HC 1204–1522 and gapped bait vector, pGBT9. The transformants were selected for positive interactions. The inserts of the recombinant clones (1.2, 2.4 and 2.5) able to interact with the LC mutant 16 are shown. Mutated residues are in bold. The sequence of mutant 16 is in Figure 3A. (C) In vitro binding assays between recombinant hub fragments, with or without mutation K1415E, and wild-type or mutant LCb. Each purified hub fragment or no protein (–) was exposed to Ni2+ affinity resin and then incubated with lysates of bacteria expressing bovine brain LCb with or without the mutation W127R. Proteins bound to the affinity resin were eluted and analyzed for the presence of LC and hub fragments by immunoblotting with rabbit

polyclonal antisera raised against the conserved region of LCs or the proximal region of HC, respectively.

Figure 5: Double mutation of tryptophan to arginine at clathrin light chain LCb residues 105 and 138 abrogates the binding to HC, and the interaction is rescued by HC mutation K1326E. (A) The three-dimensional structure of the central portion of LCb is modeled as an a-helix, viewed from the C-terminus. The predicted orientation of the side chains of tryptophan residues at positions 105, 127 and 138 is shown. (B) Plate growth assay of yeast AH109 cells co-transformed with HC 1204–1522 (in pGBT9) and LCb (in pGAD424), with or without the indicated mutation(s). The transformants were spotted onto plates lacking Leu and Trp, with (+) or without (–) His and Ade. Transformants expressing interacting constructs grew in the absence of His and Ade after 2 days. These results are representative of three independent experiments. (C) Sequences of bovine HC 1204–1522 mutants that rescued binding to the LC mutant W105, 138R. AH109 cells harboring the LCb W105, 138R mutant fragment (in pGAD424) were co-transformed with mutated PCR products of HC 1204–1522 and gapped bait vector, pGBT9. The transformants were selected for positive interactions. The inserts of the recombinant clones (1, 2, 3 and 4) able to interact with LCb W105, 138R mutant fragment are shown. Mutated residues are in bold. (D) In vitro binding assay (as described for Figure 4C) between recombinant hub fragments, with or without mutation K1326E, and wild-type or mutant LCb. LC and hub fragments were detected by immunoblotting with rabbit polyclonal antiserum raised against the conserved region of LCs or mouse monoclonal antibody against the His6 tag epitope, respectively.

Figure 6: Model of the contact between clathrin HC and LC. (A) A model of the interface of the minimal HC-binding region of LC (red) with the HC proximal leg (blue) produced by molecular dynamics calculations after alignment according to the mapping data produced in this study. The positions of the LC tryptophan residues involved in HC binding are indicated for the modeled segments of LCb 90–157 and LCa 93–160. Note that the HC modeling extends N-terminally to the major contact region, which begins at around residue 1267, and the extra helices project back from the interface. (B) The hydrophobic pocket of HC K1415 (purple) in contact with LCb W127. HC K1415 is sandwiched between the LC tryptophan and F1410, forming a strong cation–p interaction. (C) Interactions with HC in the region of LCb W105. While LCb W105 does not interact directly with HC K1326 (purple), it packs against the aromatic side chains of F1296 and F1327, which do interact. LCb S101 and LCa S104 are oriented to hydrogen-bond with the HC backbone. (D) The hydrophobic pocket of HC K1415 (purple) in contact with LCa W130, a cation–p interaction similar to that predicted for LCb W127. (E) Interactions with HC in the region of LCa W108. The interactions shown are similar to those predicted for LCb though the model indicates they are not as favorable. As our simulations were carried out with explicit hydrogens, their positions are included in the (B–E) and predicted hydrogen bonds are indicated by dashed lines. Figures were created using the Visual Molecular Dynamics program(44) and rendered with Raster3D.

Figure 7: Model of clathrin LC regulatory domains and their position on the clathrin triskelion. The HC-binding region (red) of the LC orients the Ca2+-binding EF-hand towards the N-terminus of the proximal leg so that it may act as a structural switch to modulate the interaction of the N-terminal regulatory region (orange) of the LC with the

HC. The predicted EF-hand is placed at the N-terminal red/orange junction. The C-terminus of the LC, with its calmodulin-binding site (also orange), is localized to the vertex of the triskelion. For neuronal CCVs where the calmodulin-binding site is exposed(3), calmodulin (purple) may modulate the preferred angle of the triskelion legs as they extend from the vertex. The model of the hub and distal leg segments was generated to scale based on electron microscopy measurements(18), and the LC positioning is to scale based on the structural predictions discussed. TD and calmodulin are not drawn to scale.

Figure 8: A) The human LC sequence is shown aligned with its secondary structure prediction. The consensus EED motif and calmodulin binding region are highlighted. Positions that are good potential sites to be labeled with nitroxide to obtain PRE-based distance restraints are marked with arrowheads. B) The N-terminal half of LCa is shown to bind the C-terminal, and to require the calmodulin-binding region (residues 220-248) for the interaction. N-terminal GST-fragments were used to pull-down HA-tagged C-terminal fragments. Pull-downs were visualized by western blot anti-HA antibody. The GST-bait is shown coommassie stained. C) Calmodulin can compete with the N-/C-terminal LCa interaction in a Ca^{2+}-dependent manner. The above pull-downs were repeated in the presence of purified calmodulin and varying concentrations of Ca^{2+}.

Figure 9: [^{15}N,^{1}H]-HSQC spectra of ^{15}N-LCa 1-62-GB1 in 50mM HEPES 7.5. The titration with unlabeled GB1-LCa 148-248 is shown 0:1 (black), 1:1 (blue), 1:2 (red). Areas exhibiting significant spectral changes are starred.

Figure 10: A) [^{15}N,^{1}H]-HSQC spectra of ^{15}N-LCa 1-62-GB1 in 50mM HEPES 7.5 + 25% TFE. The unbound spectrum is shown in red, bound (1:2 with GB1-LCa 148-248)

is in blue. The central region of the spectrum is shown in here, while the Asn/Gln sidechain region is shown in B). C) [^{15}N,^{1}H]-HSQC spectra of ^{15}N-Lys-LCa 148-248-GB1. The titration with unlabeled GB1-LCa 1-62 is shown 0:1 (black), 1:1 (blue), 1:2 (red). Areas exhibiting significant spectral changes are starred.

Figure 1

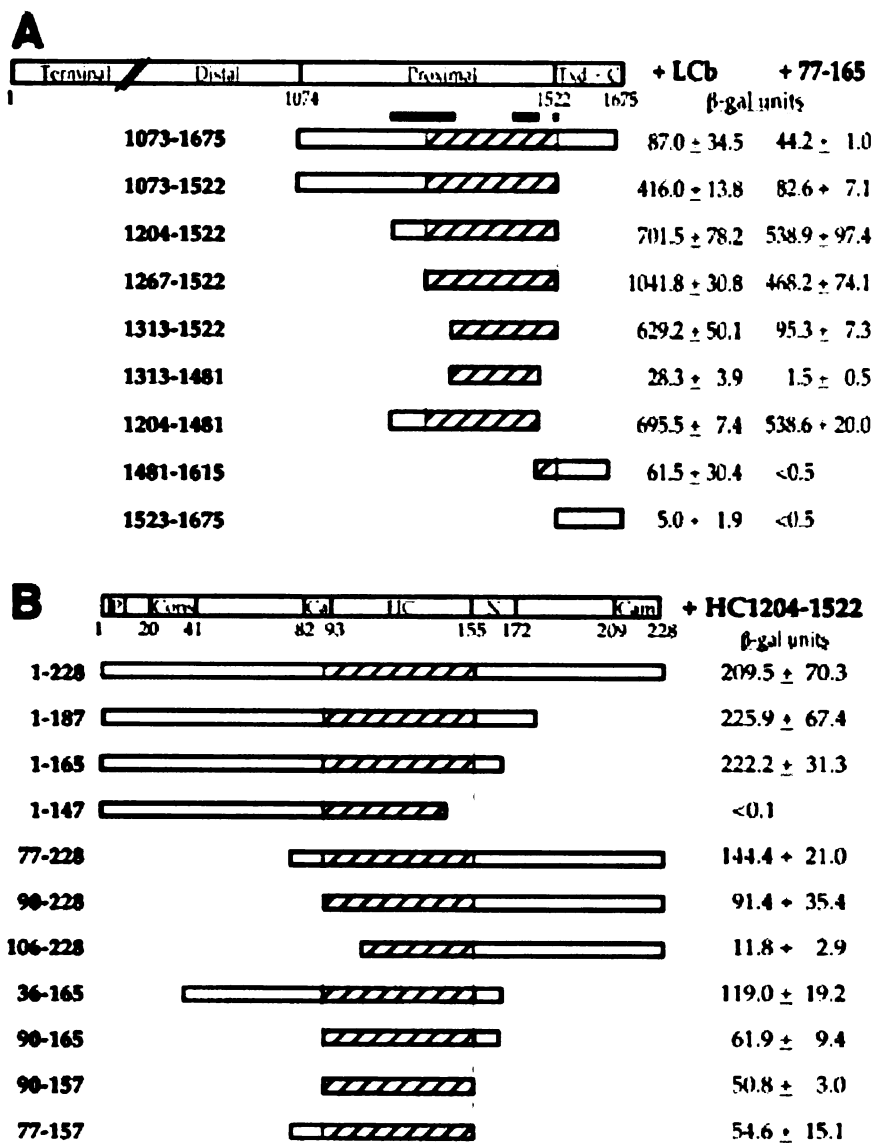

A
Terminal
Distal
Proximal
Ind - C
1
1074
1522
1675
+ LCb
+ 77-165
β-gal units
1073-1675
87.0 ± 34.5
44.2 ± 1.0
1073-1522
416.0 ± 13.8
82.6 + 7.1
1204-1522
701.5 ± 78.2
538.9 ± 97.4
1267-1522
1041.8 ± 30.8
468.2 ± 74.1
1313-1522
629.2 ± 50.1
95.3 ± 7.3
1313-1481
28.3 ± 3.9
1.5 ± 0.5
1204-1481
695.5 ± 7.4
538.6 + 20.0
1481-1615
61.5 ± 30.4
<0.5
1523-1675
5.0 + 1.9
<0.5
B
IP
Cam
Cal
UC
N
Cam
1
20
41
82
93
155
172
209
228
+ HC1204-1522
β-gal units
1-228
209.5 ± 70.3
1-187
225.9 ± 67.4
1-165
222.2 ± 31.3
1-147
<0.1
77-228
144.4 + 21.0
90-228
91.4 + 35.4
106-228
11.8 + 2.9
36-165
119.0 ± 19.2
90-165
61.9 ± 9.4
90-157
50.8 ± 3.0
77-157
54.6 ± 15.1

Figure 2

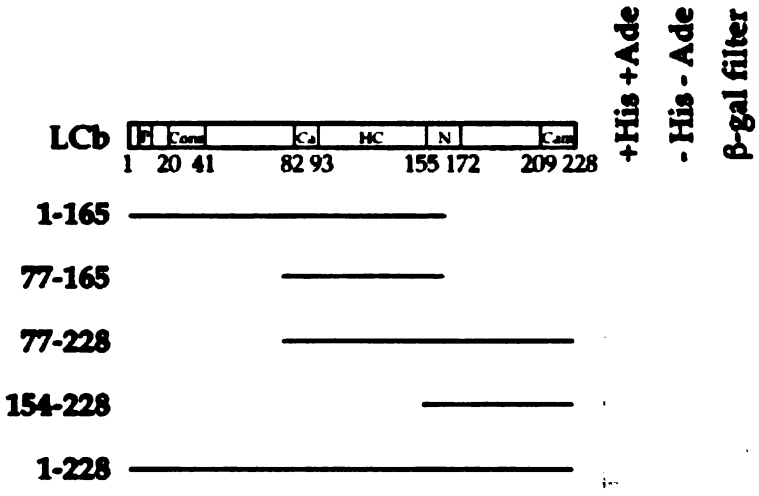

LCb
F
Ca
HC
N
1 20 41 82 93 155 172 209 228
+His +Ade
- His - Ade
β-gal filter
1-165
77-165
77-228
154-228
1-228

Figure 3

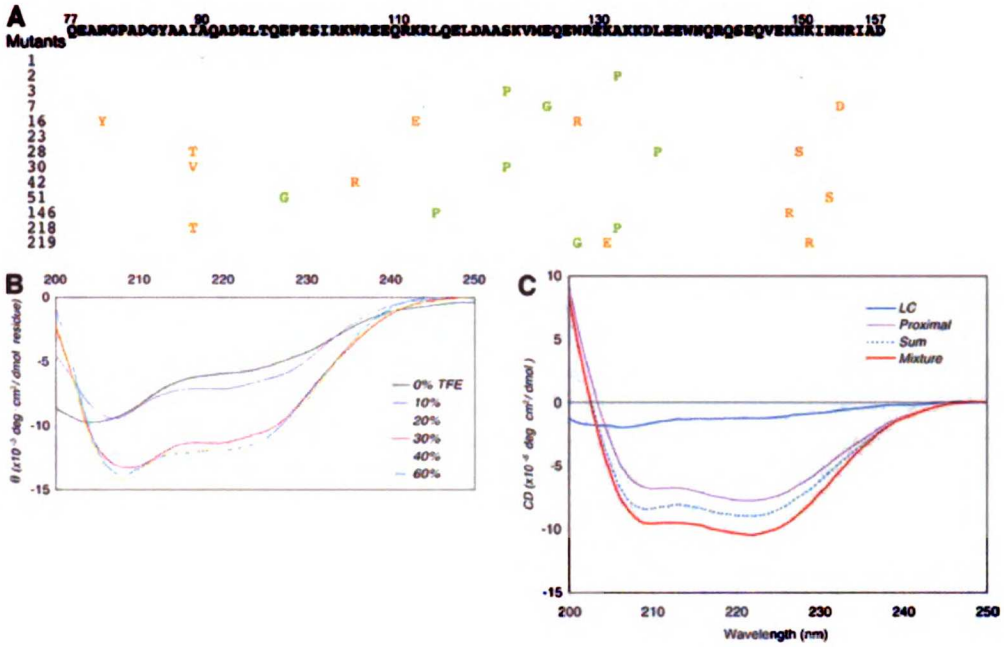

A
77
89
110
130
150
157
Mutants
1
2
3
7
16
23
28
30
42
51
146
218
219
B
200
210
220
230
240
250
0
-5
-10
-15
0% TFE
10%
20%
30%
40%
60%
C
10
5
0
-5
-10
-15
200
210
220
230
240
250
LC
Proximal
Sum
Mixture
Wavelength (nm)

Figure 4

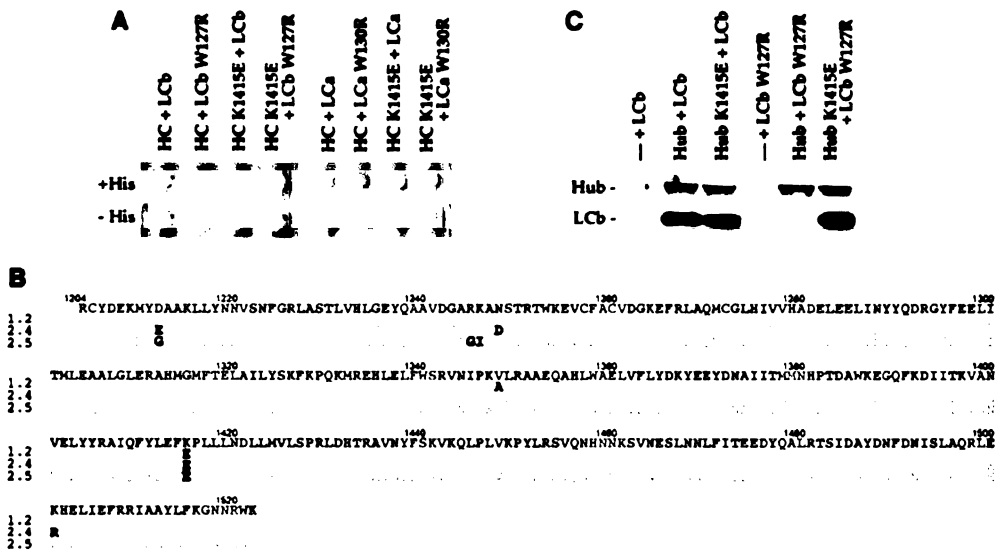

A
HC + LCb
HC + LCb W127R
HC K1415E + LCb
HC K1415E + LCb W127R
HC + LCa
HC + LCa W130R
HC K1415E + LCa
HC K1415E + LCa W130R
+His
- His
C
— + LCb
Hub + LCb
Hub K1415E + LCb
— + LCb W127R
Hub + LCb W127R
Hub K1415E + LCb W127R
Hub -
LCb -
B
1204 RCYDEKMYDAAKLLYNNVSNFGRLASTLVHLGEYQAAVDGARRANSTRTWKEVCFACVDGKEFRLAQMCGLHIVVHADELEELINYYQDRGYFEELI
1.2
2.4
2.5
1.2 TMLEAALGLERAHMGMFTELAILYSKFKPQKHREHLELFWSRVNIPKVLRAAEQAHLWAELVFLYDKYEEYDNAIITMGNHPTDAWKEGQFKDIITKVAN
2.4
2.5
1.2 VELYYRAIQFYLEFKPLLLNDLLMVLSPRLDHTRAVNYFSKVKQLPLVKPYLRSVQNHNNKSVWESLNNLFITEEDYQALRTSIDAYDNFDMISLAQRLE
2.4
2.5
KHELIEFRRIAAYLFKGNNRWK
1.2
2.4 R
2.5

Figure 5

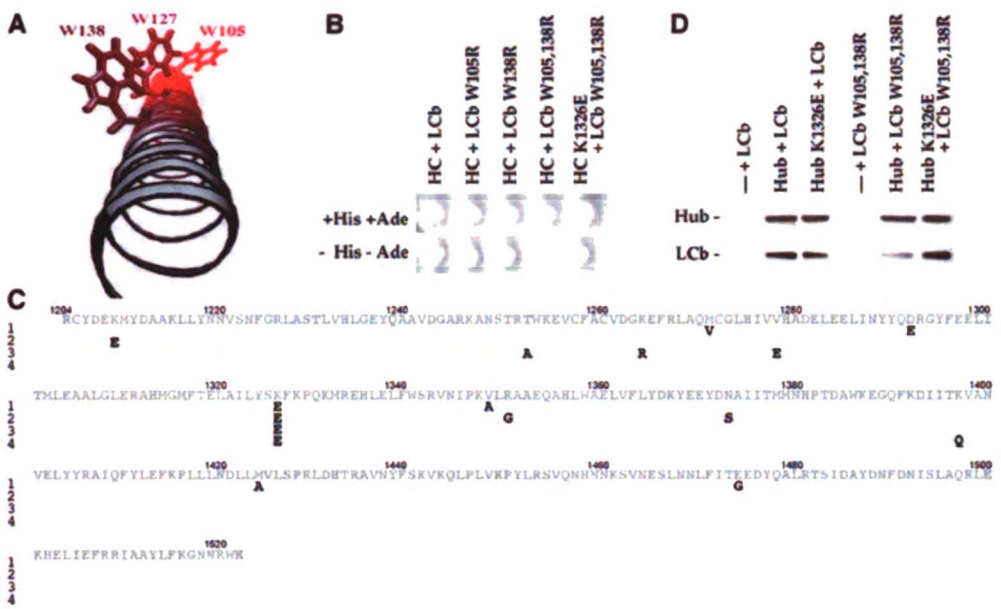

A
W138
W127
W105
B
HC + LCb
HC + LCb W105R
HC + LCb W138R
HC + LCb W105,138R
HC K1326E + LCb W105,138R
+His +Ade
- His - Ade
D
— + LCb
Hub + LCb
Hub K1326E + LCb
— + LCb W105,138R
Hub + LCb W105,138R
Hub K1326E + LCb W105,138R
Hub -
LCb -
C

Figure 6

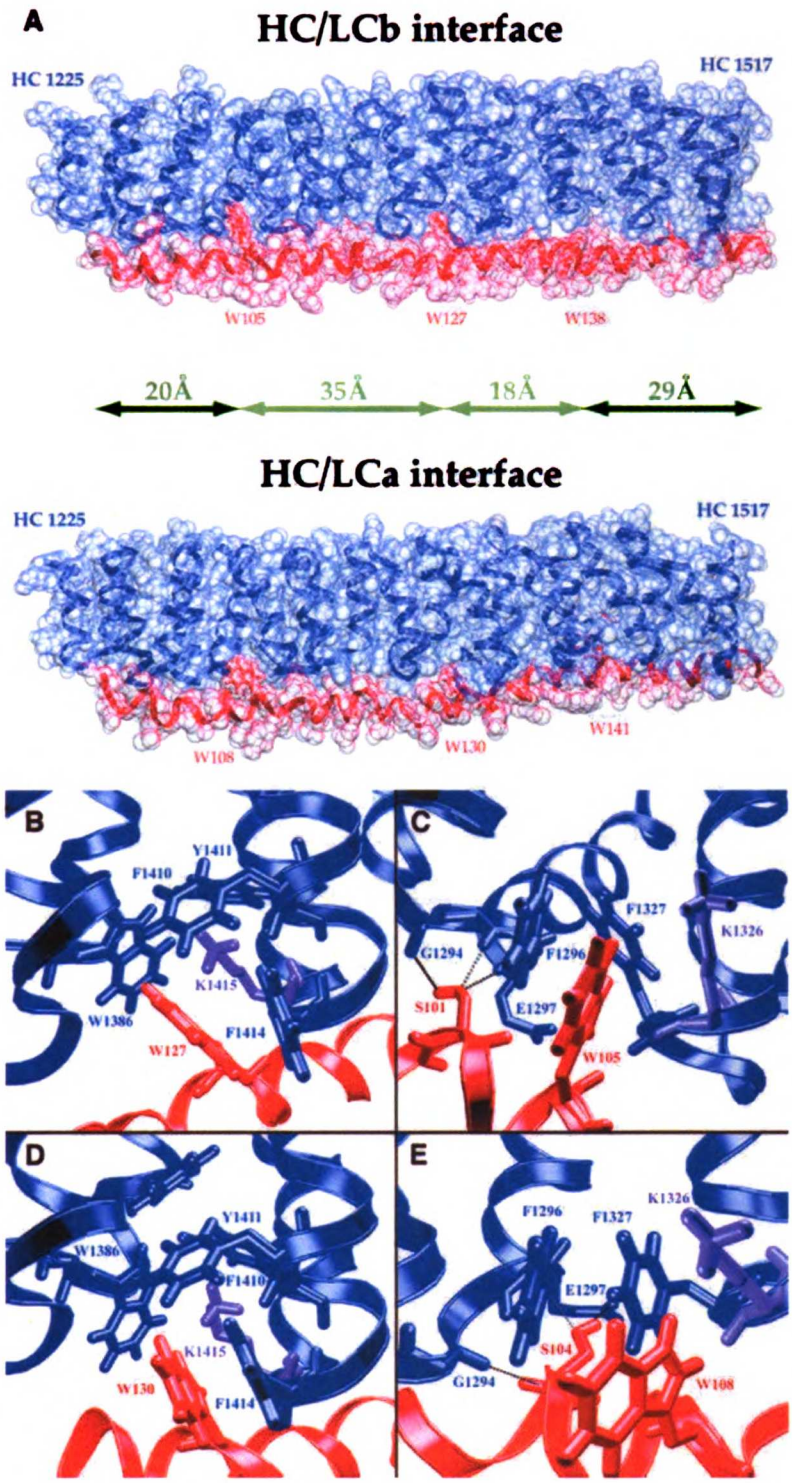

A
HC/LCb interface
HC 1225
HC 1517
W105
W127
W138
20Å
35Å
18Å
29Å
HC/LCa interface
HC 1225
HC 1517
W108
W130
W141
B
Y1411
F1410
K1415
W1386
F1414
W127
C
F1327
K1326
G1294
F1296
S101
E1297
W105
D
Y1411
W1386
F1410
K1415
W130
F1414
E
K1326
F1296
F1327
E1297
S104
G1294
W108

Figure 7

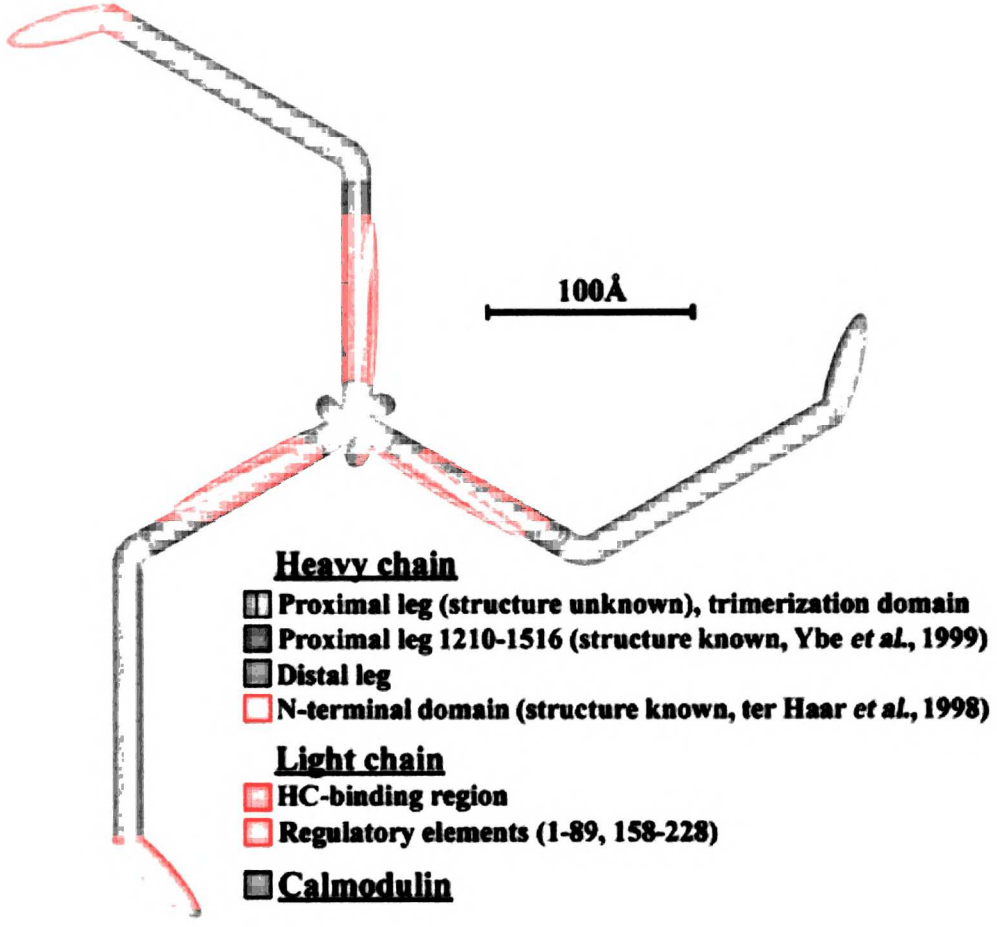
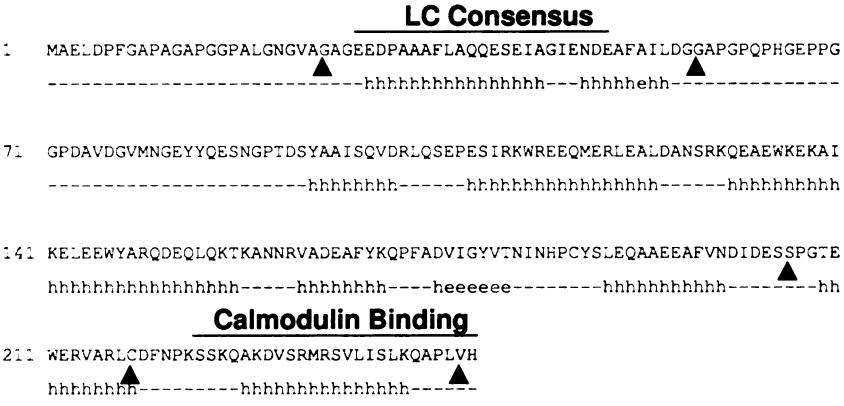

100Å
Heavy chain
Proximal leg (structure unknown), trimerization domain
Proximal leg 1210-1516 (structure known, Ybe et al., 1999)
Distal leg
N-terminal domain (structure known, ter Haar et al., 1998)
Light chain
HC-binding region
Regulatory elements (1-89, 158-228)
Calmodulin

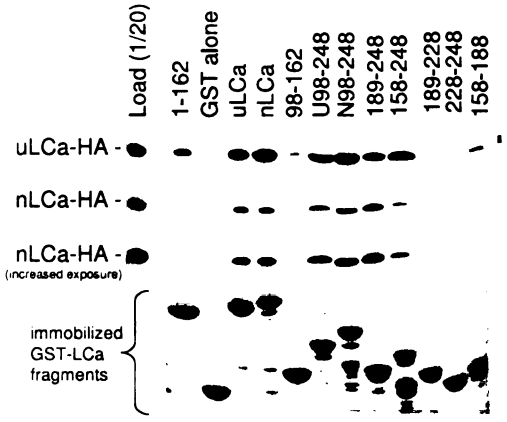
Figure 8

A

Human neuronal LCa



B



C

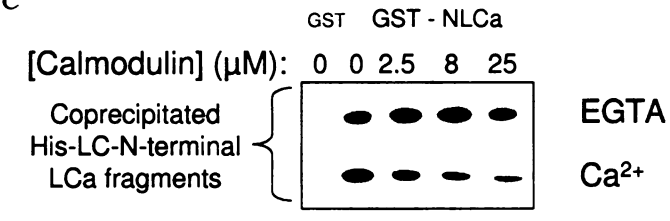


Figure 9

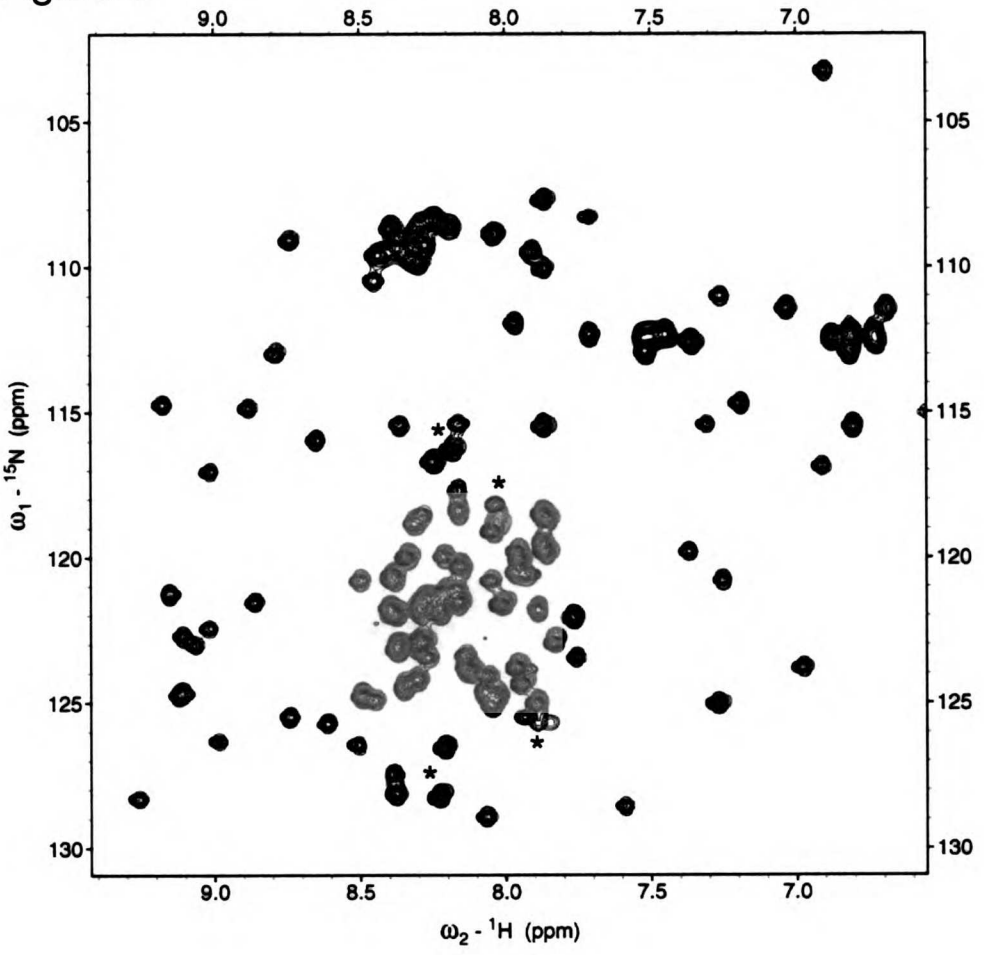
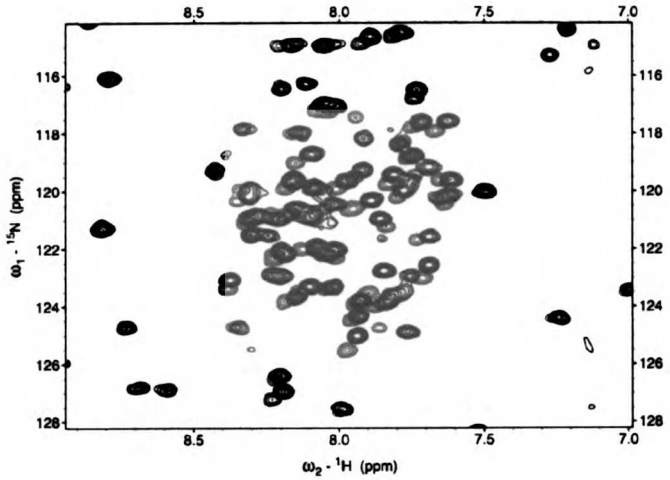
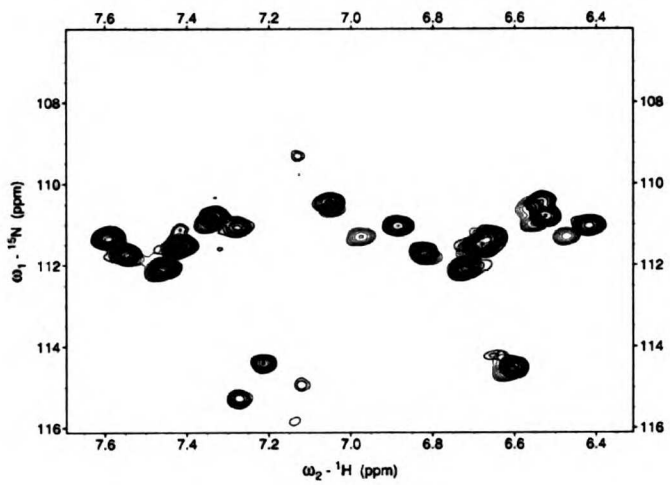

9.0
8.5
8.0
7.5
7.0
105
110
115
120
125
130
ω_1 - ^{15}N (ppm)
ω_2 - ^{1}H (ppm)

Figure 10

A



B



C