DOI: 10.1002/cbic.200800117

# An Indole-Binding Site is a Major Determinant of the Ligand Specificity of the GABA Type A Receptor-Associated Protein GABARAP

Yvonne Thielmann,[a, b] Jeannine Mohrlüder,[a, b] Bernd W. Koenig,*[a, b] Thomas Stangler,[a, b, c] Rudolf Hartmann,[a] Karin Becker,[a] Hans-Dieter Höltje,[d] and Dieter Willbold*[a, b]

*The role of tryptophan as a key residue for ligand binding to the ubiquitin-like modifier $GABA_A$ receptor associated protein (GABARAP) was investigated. Two tryptophan-binding hydrophobic patches were identified on the conserved face of the GABARAP structure by NMR spectroscopy and molecular docking. GABARAP binding of indole and indole derivatives, including the free amino acid tryptophan was quantified. The two tryptophan binding sites can be clearly distinguished by mapping the NMR spectroscopy-derived residue-specific apparent dissociation constant, $K_d$, onto the three-dimensional structure of GABARAP. The biological relevance of tryptophan-binding pockets of GABARAP was supported by a highly conserved tryptophan residue in the GABARAP binding region of calreticulin, clathrin heavy chain, and the gamma2 subunit of the $GABA_A$ receptor. Replacement of tryptophan by alanine abolished ligand binding to GABARAP.*

## Introduction

Recurrent protein-interaction domains recognize a specific amino acid sequence or a structural motif. Signature residues or common residue patterns are a hallmark of ligand sequences that bind to a specific domain type. For example, Src homology 3 (SH3) domains recognize a PxxP motif (where x denotes any natural amino acid) while binding of Src homology 2 (SH2) domains requires a phosphorylated tyrosine (pTyr) in the ligand. The presence of such key or anchor residues (e.g., Pro, pTyr) is crucial for binding affinity. Variable amino acid patterns in the immediate vicinity of the anchor ensure binding specificity. Nevertheless, the binding affinity of a protein interaction domain for an isolated anchor residue, that is, a key amino acid taken out of sequence context, is usually rather low. For example, millimolar concentrations of free pTyr are required to partially disrupt binding of the SH2 domain of abl protein tyrosine kinase to cellular proteins; this suggests that there is direct but weak binding of pTyr to abl-SH2.[1] The affinity of the SH2 domain of lymphocyte-specific protein tyrosine kinase (Lck) for pTyr is weaker than 1 mM according to isothermal titration calorimetry (ITC) data.[2] Binding of pTyr to Src-SH2 is temperature dependent with $K_d$ values between 200 and 333 μM based on ITC data.[3] In contrast, the affinity of SH2 domains to physiological pTyr-containing ligands is typically in the high nanomolar range.[2]

The γ-amino butyric acid receptor type A ($GABA_A$ receptor) is a ligand-gated chloride ion channel that mediates the effects of the major inhibitory neurotransmitter γ-amino butyric acid (GABA) in the central nervous system.[4] The γ2 subunit of $GABA_A$ receptors interacts with the $GABA_A$ receptor-associated protein (GABARAP). This interaction was first identified in a yeast two-hybrid screen, and confirmed by colocalization experiments in cultured cortical neurons and by coimmunoprecipitation of GABARAP with $GABA_A$ receptor subunits from brain extracts.[5]

GABARAP belongs to a family of proteins that are evolutionarily highly conserved among eukaryotic species from yeast over plants to mammals.[6] Sequence identities between human GABARAP and its orthologues are 100% (rat, mouse, bovine), 79% (*C. elegans*), 55% (*S. cerevisiae*), and 54% (*A. thaliana*), respectively; this suggests a critical function of the protein.[7,8] Several GABARAP-like homologues have been identified in humans including GATE-16 (Golgi-associated ATPase enhancer of 16 kDa), GEC1 (estrogen-induced 1.8 kb RNA-coded protein; this was first isolated in guinea pig endometrial cells) and LC3 (microtubule-associated protein light

[a] *Y. Thielmann,[+] Dr. J. Mohrlüder,[+] Dr. B. W. Koenig, Dr. T. Stangler, Dr. R. Hartmann, K. Becker, Prof. Dr. D. Willbold*
*INB-2/Molekulare Biophysik II, Forschungszentrum Jülich*
*52425 Jülich (Germany)*
*Fax: (+49) 2461-612023*
*E-mail: b.koenig@fz-juelich.de*
*d.willbold@fz-juelich.de*

[b] *Y. Thielmann,[+] Dr. J. Mohrlüder,[+] Dr. B. W. Koenig, Dr. T. Stangler, Prof. Dr. D. Willbold*
*Institut für Physikalische Biologie*
*BMFZ/Biologisch-Medizinisches Forschungszentrum*
*Heinrich-Heine-Universität Düsseldorf*
*40225 Düsseldorf (Germany)*

[c] *Dr. T. Stangler*
*Present address: Analytical and Pharmaceutical Development*
*Biopharmaceutical Operations, Sandoz GmbH*
*6250 Kundl (Austria)*

[d] *Prof. Dr. H.-D. Höltje*
*Institut für Pharmazeutische und Medizinische Chemie*
*Heinrich-Heine-Universität Düsseldorf*
*40225 Düsseldorf (Germany)*

[+] *These authors contributed equally to this work.*

  

chain 3).[9] The sequence similarities of these homologues to GABARAP range from ~87% (GEC1) to ~30% identity (LC3).[8, 10, 11]

Available crystal structures of GABARAP,[11–13] GATE-16,[14] and LC3[15] as well as the only NMR structure of GABARAP[16] manifest a high degree of structural similarity within the GABARAP family. All structures share a stable ubiquitin-like (UBL) fold in the C-terminal region (amino acids 30 to 117 or 120, respectively) and two additional helices ($\alpha$1 and $\alpha$2) at the N terminus (Figure 1A). The UBL core domain comprises two parallel $\beta$ strands ($\beta$2 and $\beta$3) that are flanked by one antiparallel $\beta$ strand on either side ($\beta$1 and $\beta$4), and $\alpha$ helices $\alpha$3 (between $\beta$2 and $\beta$3) and $\alpha$4 (between $\beta$3 and $\beta$4). Both $\alpha$3 and $\alpha$4 pack towards the concave surface of the central $\beta$ sheet, whereas $\alpha$1 and $\alpha$2 are located on the convex side. One face of the UBL core structure is highly conserved at the amino acid level, but the opposite side shows considerable divergence.[11, 14] The conserved surface features two prominent patches of partially exposed hydrophobic residues that constitute hydrophobic pockets and will be referred to as HP1 and HP2 in what follows. In the three-dimensional structure of GABARAP,[16] HP1 is lined by Ile21, Pro30, Leu50, Phe104 and the aliphatic portions of Glu17 and Lys48, while HP2 is formed by Tyr49, Val51, Pro52, Leu55, Phe60, Leu63 and Ile64 (Figure 1). Notably, pocket 1 is surrounded by several basic side chains (Arg28, Lys46, Lys48).

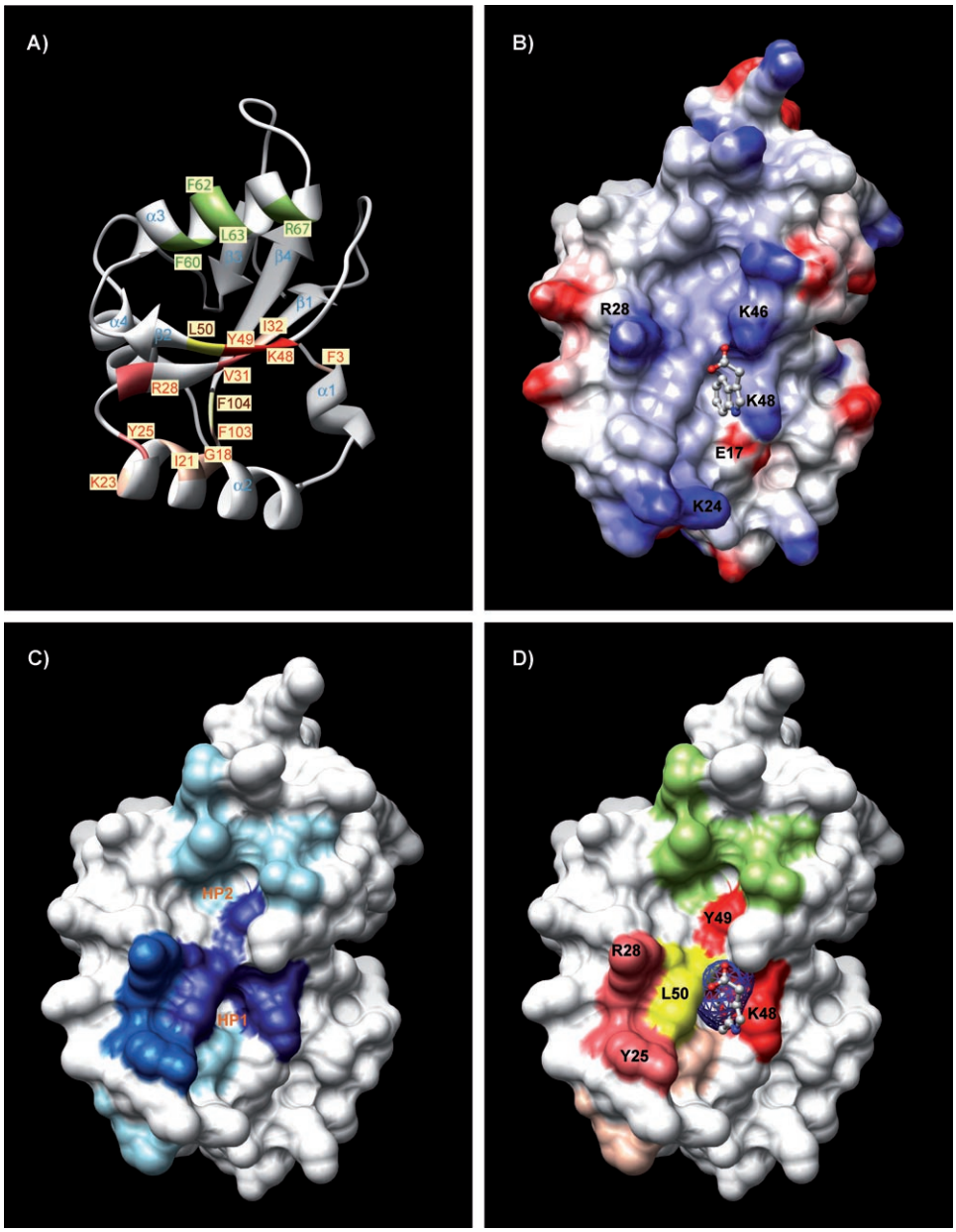


**Figure 1.** Localization of the IAA-binding site at the conserved face of GABARAP. Shown are surface and ribbon diagrams of the GABARAP crystal structure (PDB ID code 1gnu)[12] and the bound IAA molecule. A) Ribbon diagram of GABARAP; B) the surface of GABARAP with the colors based on the electrostatic potential, with basic and acidic side chains in blue and red, respectively; C) the protein surface is color coded according to the residue-specific CSP observed upon addition of IAA (28 mM). The darker the shade of blue, the stronger is the shift perturbation. Two hydrophobic patches (HP1 and HP2) on the surface of GABARAP are indicated. D) Color of shifting residues reflects the strength of the interaction between GABARAP and IAA: shades of red indicate $K_d$ ~6 mM; shades of green indicate a very weak and perhaps unspecific interaction with an apparent $K_d$ above 30 mM; yellow residues show an intermediate apparent $K_d$. Color saturation reflects the magnitude of the CSP. Residues that have a CSP above the arbitrary threshold of 0.18 are mapped onto the ribbon representation of GABARAP in panel A; the color coding is identical to panel D. Panels B and D also show a ball-and-stick representation of IAA that is bound to GABARAP in the energetically most favorable position and orientation, which was determined by molecular docking. In addition, panel D contains a J-surface that encloses the region that exhibited a dot density greater than 2.8$\sigma$ ($\sigma$ is the standard deviation of the dot density from the mean).

Recently, we screened a randomized library of phage-displayed dodecapeptide sequences for binding to GABARAP.[17] The resulting set of GABARAP-binding sequences shows at least one tryptophan residue in 80 out of the 85 dodecapeptides (94%) that were identified. This is even more remarkable considering that tryptophan is, with an observed frequency of 2.2%, significantly underrepresented in the phage-displayed peptide library (Ph.D.-12 manual, New England Biolabs). Two tryptophan-containing consensus motifs were derived

  

from subsets of the identified peptide sequences that led us to discover the previously unknown GABARAP-binding capacity of calreticulin[17] and clathrin heavy chain.[18] In the current manuscript we address the significance of the conserved tryptophan for the binding of calreticulin, clathrin heavy chain and the γ2 subunit of the $GABA_A$ receptor to GABARAP. We propose that tryptophan might play the role of a signature amino acid residue that must be present in protein domains that bind to the hydrophobic patch HP1 on the conserved face of GABARAP.

## Results and Discussion

The persistent appearance of a tryptophan residue in the phage display-selected GABARAP-binding peptides is very likely related to its peculiar side chain, which is comparably large and consists mainly of a flat, rigid, hydrophobic, and aromatic indole ring. Therefore, we characterized the interaction of GABARAP with indole and three indole derivatives. Binding between GABARAP and the uncharged indole, the zwitterionic free amino acid L-Trp, the anionic indole-3-acetic acid (IAA), and the uncharged *N*-acetyl-L-tryptophan amide (NATA) was experimentally confirmed by saturation transfer difference (STD) NMR spectroscopy (Figure 2).[19,20] A STD spectrum of a sample that contained glucose (10 mM) and GABARAP (0.1 mM) was recorded as a negative control. As expected, the STD spectrum did not show any NMR signals of glucose. The STD experiment is capable of detecting protein–ligand interactions with dissociation constants ($K_d$) between the upper nano- and the lower millimolar range. Selective saturation of a protein $^1$H resonance spreads rapidly to other nuclei of the protein by intramolecular spin diffusion. In the case of ligand binding, saturation is also transferred to the spins of the ligand. Observation of ligand signals in the difference spectrum that is calculated from two $^1$H NMR spectra recorded with and without selective saturation of protein signals, respectively, indicates ligand binding (Figure 2).

### Prediction of indole binding sites on GABARAP

A molecular docking algorithm was applied for localization of tryptophan-binding sites on the previously determined three-dimensional structure of GABARAP. Docking of indole to GABARAP resulted in two highly preferred sites that match the two hydrophobic patches, HP1 and HP2, on the conserved face of the GABARAP fold, which is described above. In 35 out of the 100 configurations that were calculated the indole was found in HP1. Moreover, the six configurations with the highest-scoring values ("fitness") belonged to this subset of 35 configurations. However, close inspection of the docked indole

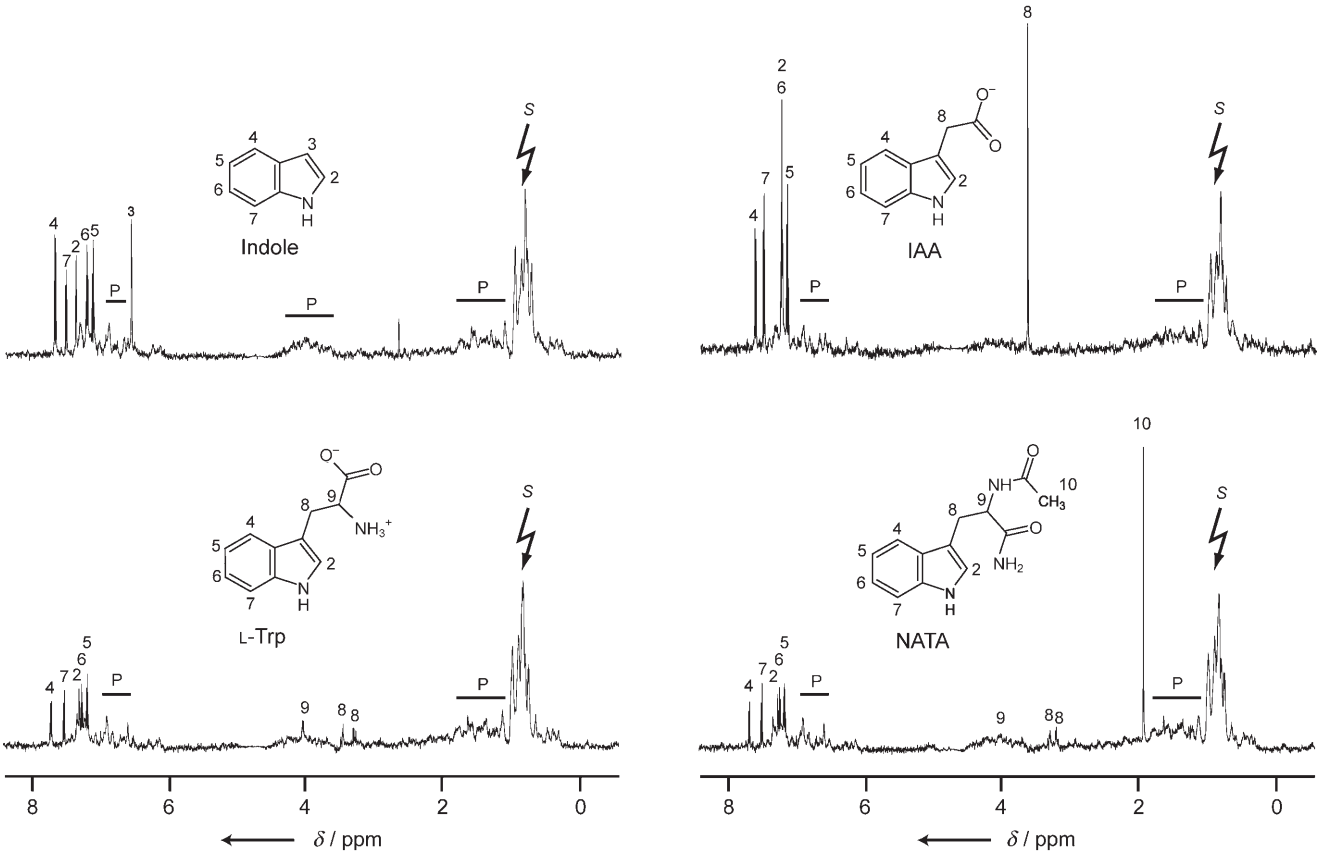


**Figure 2.** Saturation transfer difference (STD) NMR spectra prove the interaction of GABARAP with indole and the indole derivatives IAA, L-Trp, and NATA. The depicted ionization state reflects the predominant state at neutral pH according to published p$K_a$ values. The spectral region around the saturation frequency (*S*) is slightly disturbed. Very weak protein signals (P) are observed in addition to the marked proton resonances of all four ligands.

  

molecules revealed that the rigid molecules were deeply immerged in the hydrophobic pockets with their pyrrole moiety of the indole ring pointing towards the bottom of the tight cavities. Although this ligand orientation apparently reflects the energetically most favorable position of an isolated indole on the surface of GABARAP, it is incompatible with an indole side chain of a tryptophan in the context of a GABARAP-bound polypeptide for steric reasons.

In another set of docking experiments, we studied the interaction of GABARAP with IAA. This indole derivative has an acetic acid group covalently attached to position three of the indole ring (Figure 2). The carboxyl group is ionized at neutral pH, which results in a negative net charge of IAA. Docking revealed just one energetically highly favorable IAA-binding site on GABARAP, which matches the hydrophobic pocket HP1 (Figure 1C). Importantly, the orientation of the indole ring of IAA with respect to GABARAP is opposite to the one observed with the bare indole. In the case of IAA, the benzene ring points towards the bottom of the hydrophobic pocket. The indole ring of IAA is sandwiched between the side chains of Lys48 and Leu50 and the carboxyl group sticks out of the hydrophobic cavity. Cationic amino groups of the side chains of Lys46 and Lys48 directly flank HP1 and provide favorable interaction sites for the carboxyl moiety of IAA (Figure 1B). The predicted orientation of the IAA molecule indicates that an indolic tryptophan side chain of a GABARAP-bound peptide could be favorably accommodated in HP1.

### Experimental localization of interaction site

For experimental verification of the predicted binding site of the indolic side chain of tryptophan we conducted HSQC titration experiments on $^{15}$N-labeled GABARAP. The NMR frequencies of the protein backbone atoms are exquisitely sensitive probes of the chemical environment of the observed spins. Ligand binding usually disturbs the electron current density in the immediate vicinity of the binding site, and results in localized chemical shift changes of nearby protein spins, which are often referred to as chemical shift perturbation (CSP). Two-dimensional $^{1}$H,$^{15}$N HSQC spectra provide residue-specific amide $^{1}$H and $^{15}$N resonance frequencies. A series of protein HSQC spectra that were recorded with increasing amounts of ligand in the sample (often called HSQC titration) allows mapping of the ligand-binding site onto the protein surface. It provides insight into the kinetics and perhaps the mode of binding, and in favorable cases also yields the dissociation constant of the protein–ligand complex.[21–23] Aromatic ring currents can induce profound through-space NMR shielding of nearby spins that depend both on the distance and on the planar orientation of the aromatic ring relative to the observed spin. Isochemical-shielding surfaces of $\pm$0.1 ppm of an indole ring are roughly 9 Å away from the center of the fused ring.[24] Therefore, CSP mapping provides an excellent tool for the characterization of the interaction of indole derivatives with GABARAP.

Chemical shift perturbation at GABARAP backbone amide sites relative to ligand-free protein was studied at large ligand excess for indole, IAA, L-Trp, and NATA in aqueous buffer (Figure 3). Indole and NATA showed rather limited solubility and were added at their saturation concentration of 13 and 14 mM, respectively. Normalized $^{1}$H–N and $^{15}$N composite CSP values that were detected for each of these four ligands in GABARAP HSQC spectra are presented as a function of sequence position in Figure 3C. The overall pattern of ligand-induced chemical shift changes is very similar for all four ligands. The strongest perturbations occur in the $\beta$ strands $\beta$1 (K48, Y49, L50) and $\beta$2 (R28, V31) and in the loop that connects $\alpha$2 and $\beta$1 (Y25). Clusters with smaller but nevertheless significant changes are also observed in the C-terminal half of $\alpha$2 (G18, I21, K23), in $\alpha$3 (F60, F62, L63, R67), and in the loops between $\beta$3 and $\alpha$4 (N81, N82) and between $\alpha$4 and $\beta$4 (E101, F103, F104). Figure 1C maps the strength of the CSP observed upon addition of 28 mM IAA relative to ligand-free GABARAP on a surface contour of GABARAP. Surface-exposed residues are indicated in shades of blue if the normalized composite CSP of the corresponding backbone amide exceeded a threshold value of 0.18 (Figure 3C). The darkness of the blue reflects the strength of the CSP. The cluster of dark blue residues around HP1 is in agreement with the prediction of IAA binding at HP1. However, the surface-mapping approach is of limited use for exact localization of the bound ligand. Residues that show significant shift perturbations are spread over the entire face of GABARAP (Figure 1C). Significant CSPs are not unexpected even at a distance of ~10 Å from the center of the indole ring due to the exceptionally strong dipolar field caused by the $\pi$-electron currents of indole. The observed CSP pattern alone is apparently insufficient for precise localization of IAA.

### J-surface localization of binding site

McCoy and Wyss introduced two approaches for the spatial localization of weakly interacting aromatic ligands from binding-induced CSP data. Protein chemical shift changes can be simulated for a range of ligand positions and orientations relative to the surface of the target with the aim of minimizing the difference between the measured and predicted CSP data.[25] Alternatively, the effect of an aromatic ring current that originates from the ligand on the resonance position of neighboring protein spins can be approximated by a point-dipole at the center of the ring.[26] The detected CSP of a single protein spin is compatible with a multitude of possible positions (dots) of the point-dipole that causes the perturbation. These positions define a sphere that is centered at the monitored protein spin and occupied by N dots. The local dot density is a measure of the probability of finding the source of the shift perturbation at a given voxel. The superposition of dot densities derived from the measured CSPs of all affected protein spins exhibit maximal dot density close to the center of the aromatic ring that is causing the shifts.[27] Surfaces that enclose the regions of substantially elevated dot density are referred to as "J-surfaces" because the electron current density J that is responsible for the observed CSPs is most probably localized within these surfaces.[27] Using the software Jsurf we derived a J-surface for IAA, which is centered within pocket HP1 (Figure 1D). As expected, this J-surface also engulfs the location of

  

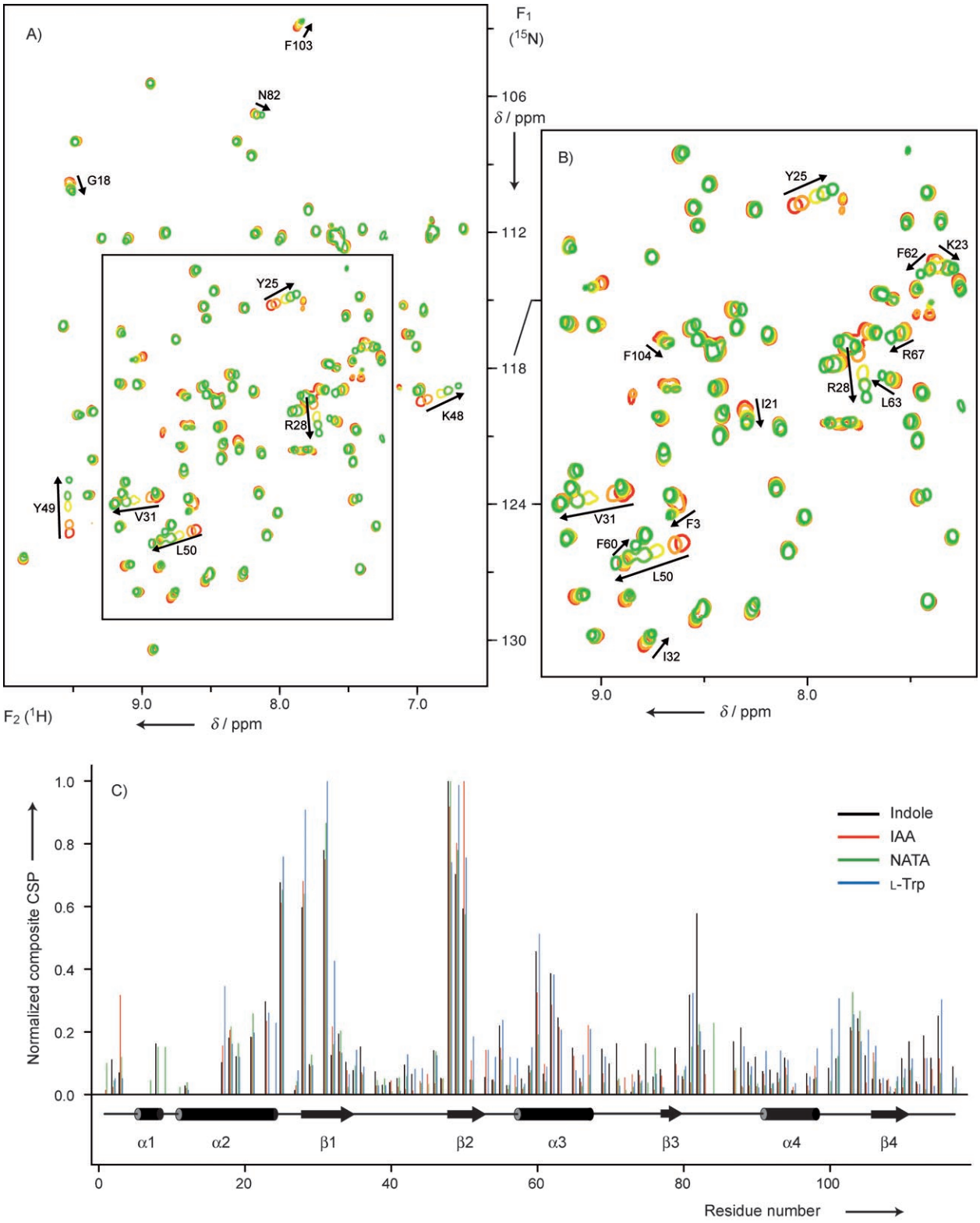


**Figure 3.** A) and B) $^{1}H,^{15}N$ HSQC titration of GABARAP with IAA and C) normalized chemical shift perturbation (CSP) of GABARAP resonances at the maximum ligand concentrations studied as a function of sequence position for indole (13 mM, black), IAA (28 mM, red), NATA (14 mM, green), and L-Trp (42 mM, blue) in DMSO-free buffer. Secondary structure elements of GABARAP are indicated at the bottom of panel C. GABARAP resonances characterized by a normalized CSP of more than 0.18 upon addition of IAA (28 mM) are labeled in the HSQC spectra. Five out of ten titration steps are shown in panels A and B that correspond to 0 mM (red contours), 1 mM (orange), 4 mM (yellow), 9 mM (light green), and 28 mM IAA (dark green) in the NMR spectroscopic sample. The following protocol was used to generate the normalized composite CSP values shown in panel C: for each ligand, all backbone $^{1}HN$ and $^{15}N$ CSPs of GABARAP resonances observed at the maximum ligand concentration were normalized to the largest $^{1}HN$ and $^{15}N$ CSP value, respectively, that was observed for this particular ligand. The normalized $^{1}HN$ and $^{15}N$ chemical shifts were then summed up for each residue, and the resulting composite shifts were normalized again by scaling the maximum shift observed for each ligand to a value of 1.0.[36]

  

the indole ring derived by molecular docking. Together with the similarity of the CSP pattern observed for all four indole derivatives (Figure 3C), these results indicate that HP1 is the dominating binding site of the tryptophan analogues that were studied.

### $K_d$ map of binding sites

A series of HSQC experiments of isotope-labeled protein recorded with varying amounts of ligand might allow the determination of $K_d$, provided that the dissociation of the protein–ligand complex occurs either in the fast or slow exchange range compared to the NMR chemical shift timescale.[23] The gradual shift of GABARAP resonances that are observed during HSQC titration with all four indole derivatives studied is typical for fast exchange and weak binding (Figure 3). The largest $^1$H chemical shift change that was observed for IAA corresponds to 192 Hz. Assuming that fast exchange is occurring, this observed maximum shift indicates a dissociation rate constant of the GABARAP–IAA complex of more than ~1200 Hz. Close inspection of the HSQC trajectories of individual peaks reveals qualitative differences (Figure 3). Some resonances (e.g., K48, Y49) shift strongly even at low IAA concentrations (red–orange–yellow), whereas others (e.g., F62, L63, R67) move predominantly at higher IAA content (yellow–light green–dark green). Different rates of peak movement indicate the presence of more than one binding site with unequal ligand affinities. This hypothesis was verified by determining $K_d$ values for all GABARAP backbone amides with strongly shifting peaks in the HSQC titration of IAA that were not compromised by strong overlap with neighboring peaks (peaks labeled in Figure 3). A graphical representation of the $K_d$ fits for the titration of GABARAP with IAA is shown in Figure 4. The derived $K_d$ values clearly indicate the presence of two ligand-binding sites on GABARAP. Titration with IAA results in a sizable cluster of GABARAP residues that experience ligand binding with a $K_d$ of $6\pm2$ mM (K48, Y49, V31, R28, Y25, F3, K23, I32, G18, I21, F103). A second cluster (F60, F62, L63, R67) shows very weak and arguably unspecific binding with fitted $K_d$ values between 38 and 208 mM. The latter $K_d$ estimate has a large uncertainty, partially because the range of ligand concentrations that was studied (up to 28 mM) is too low for a better characterization of such a weak interaction. However, realizing the existence of this second interaction site is important for the interpretation of the observed CSP pattern in Figure 1C irrespective of the question of the biological relevance of this second site. Finally, there are two residues that do not fit into either one of these two clusters. The most prominent is L50, which shows the strongest CSP of all GABARAP HSQC cross-peaks upon titration with IAA. Despite severe peak overlap with F60 in one of the ten HSQC spectra recorded during the titration, the remaining nine data points fit nicely to our simple binding isotherm and yield an apparent $K_d$ value of $13\pm1$ mM. The shift changes of F104 showed some scatter, which resulted in a poorly defined $K_d$ of $10\pm3$ mM. Inspection of the known structure of GABARAP revealed that L50 is located proximal to both identified ligand-binding sites. Also F104 might sense the influence of

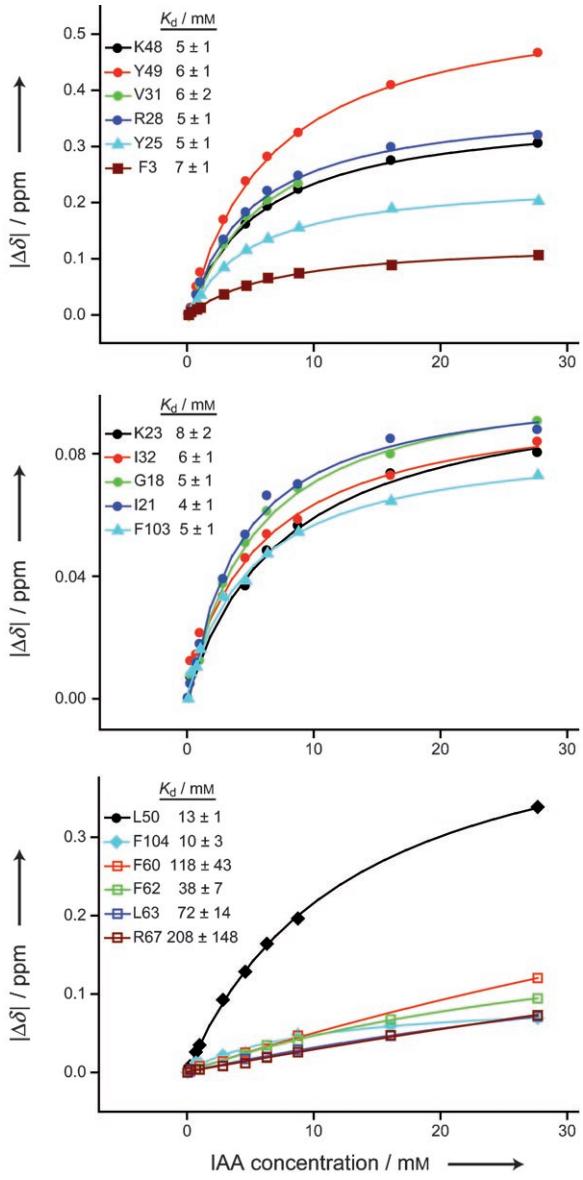


**Figure 4.** Representation of binding curves that were derived from the HSQC titration of GABARAP with IAA. GABARAP residues proximal to the primary IAA-binding site are characterized by $K_d$ in the range between 4 and 8 mM and are displayed in the upper two panels. The residues presented in the lower panel reflect weaker IAA affinity than the major binding site. F60, F62, L63, and R67 on helix α3 are in the vicinity of the hydrophobic patch 2. The estimated $K_d$ values might indicate unspecific binding of IAA at HP2. The observed CSP of L50 and F104 is most likely a superposition of shifts caused by IAA binding to both sites; this results in apparent $K_d$ values of 13 and 10 mM, respectively.

IAA binding to both hydrophobic patches. Figure 1A and D shows residues of cluster 1 ($K_d$ ~6 mM) in shades of red, of cluster 2 ($K_d > 30$ mM) in green, and L50 and F104 in yellow. We speculate that the backbone amide resonances of L50 and F104 are influenced by ligand binding to both interaction sites and perhaps by localized conformational changes. In summary, mapping $K_d$ values to individual backbone amides of the spatial protein structure allows a more stringent interpretation of the binding topology than simple CSP mapping.

  

### Comparison of binding affinities of indole derivatives

Based on HSQC titration experiments we were able to estimate the dissociation constants for binding of IAA, indole and L-Trp to the hydrophobic pocket HP1 on the conserved face of GABARAP. The limited solubility of NATA (saturating concentration of ~14 mM) interfered with a quantitative characterization of its interaction with GABARAP. Of the three remaining compounds, the anionic IAA binds most strongly ($K_d$ ~6 mM), followed by the noncharged indole ($K_d$ ~12 mM) and the weakly interacting zwitterionic L-Trp ($K_d$ values between 30 and 60 mM were estimated).

The solubility of indole in buffer (13 mM as estimated by quantitative $^1$H NMR spectroscopy) was not sufficient for an accurate determination of $K_d$. Addition of 10% (*v*/*v*) [$D_6$]DMSO to the buffer increased the solubility of indole to ~27 mM and permitted the $K_d$ value for indole and GABARAP to be determined (~12 mM) after correction for DMSO-induced chemical shift changes. All other measurements were carried out in the absence of DMSO.

The negatively charged IAA shows a higher affinity than the bare indole ring and the zwitterionic free amino acid L-Trp. Apparently, there is an electrostatic contribution to the binding energy, which most likely results from attraction between the ionized carboxyl group of IAA and the basic side chain of either K46 or K48 of GABARAP (Figure 1B). Interestingly, the prominent tryptophan in phage display-selected GABARAP-binding peptides is frequently flanked by negatively charged amino acids in the polypeptide chain.[17] Thus, in addition to the absolutely required tryptophan, a negatively charged side chain might support specific binding to the hydrophobic patch HP1 of GABARAP.

### Tryptophan plays a key role for GABARAP binding

The undecapeptide P1 corresponds to the amino acid sequence of human calreticulin CRT(178–188), which was identified as a GABARAP-binding protein based on phage display experiments.[17] The only difference between P1 and peptide P1(W6A) is a tryptophan-to-alanine substitution of the single tryptophan in the sequence. The interaction between GABARAP and the two peptides was probed by surface plasmon resonance. GABARAP binds P1 with a $K_d$ of 11.4 μM. However, in case of P1(W6A) no interaction with GABARAP was detectable. This indicates that the affinity between P1(W6A) and GABARAP, if any, must be millimolar or weaker.

The loss of GABARAP binding of the P1 peptide upon tryptophan replacement strongly suggests a key role of the tryptophan within the binding motif despite the rather low GABARAP affinity of the free amino acid L-Trp and the indole derivatives studied. The importance of tryptophan for GABARAP binding is reminiscent of the key role of phosphotyrosine within the SH2 domain-binding motifs.

### Biological relevance of the indole-binding pocket

Compared to other amino acids, tryptophan occurs least frequently in proteins; further it is often found at protein surfaces or in the vicinity of the membrane–water interface in the case of membrane proteins. This interfacial exposure of tryptophan might suggest a preferential involvement of Trp in specific protein–protein interactions.[28] Indeed, statistical analysis of protein–protein interaction sites in a representative sample of protein complexes revealed that tryptophan is the amino acid with the highest propensity for residing in an interface patch, immediately followed by phenylalanine, another aromatic amino acid.[29] The peculiar interfacial preference of tryptophan might result from its profound dipolar character, the π electronic structure of the indole moiety and perhaps cation–π interactions,[30] or from its flat rigid shape.[31]

If the identified tryptophan-binding site on GABARAP is relevant for functional interactions with other proteins, then one would expect to find highly conserved tryptophan residues within the GABARAP-binding regions of at least some of GABARAP's interaction partners. Indeed, strongly conserved tryptophans are found in calreticulin (CRT), in the γ2 subunit of the $GABA_A$ receptor, and in the clathrin heavy chain. Multiple sequence alignment of human CRT (precursor protein; Swiss-Prot accession number P27797) with 24 orthologous from other eukaryotes ranging from plants to mammals revealed 100% conservation of W200 of *h*CRT (precursor protein). An arbitrary selection of six out of these 25 sequences is shown in Figure 5A. Interestingly, the CRT alignment shows also 92% conservation of aspartic acid residues both in the *i*−1 and *i*+1 positions relative to W200. Multiple sequence alignment of human $GABA_A$ receptor γ2 subunit (Swiss-Prot accession number P18507) with all five known homologues also indicates strong conservation of the tryptophan at sequence position 424 (Figure 5B). This tryptophan is part of a short stretch of amino acids (427–437) that was shown to be sufficient for binding to GABARAP.[13] Finally, W514 in the GABARAP-binding site of human clathrin heavy chain 1 (Swiss-Prot accession number Q00610) and 2 (Swiss-Prot accession number P53675) is conserved in the homologues of mouse and rat, but is conservatively substituted by phenylalanine or tyrosine in the homologues of three organisms, which interestingly do not code for GABARAP (Figure 5C). Even more intriguing, the reported GABARAP binding affinities for calreticulin, clathrin heavy chain and $GABA_A$ receptor γ2 subunit of 64 nM,[17] ~1 μM,[18] and >200 μM,[12] respectively, correlate with their local electrostatic net charge around the central tryptophan residue. The more negative the net charge, the lower is the reported $K_d$ value.

## Conclusions

GABARAP exhibits a specific binding site for indole and indole derivatives. This binding site was traced to a hydrophobic pocket on the conserved face of the GABARAP structure by using NMR spectroscopy-based ligand affinity mapping. The presence of a tryptophan residue appears to be a major determinant of GABARAP ligands that bind to this site. Support for

  

A)

| | | |
|---|---|---|
| CALR_HUMAN | LEDD**W**DFLPPKKIKDPDA | 213 |
| CALR_MOUSE | LEDD**W**DFLPPKKIKDPDA | 213 |
| CALR_RAT | LEDD**W**DFLPPKKIKDPDA | 213 |
| CALR_DROME | LEDD**W**DFLAPKKIKDPTA | 213 |
| CALR_CAEEL | LEED**W**DLLPAKKIKDPDA | 208 |
| CALR_DICDI | LADD**W**ELLPSKQIKDPKQ | 213 |

B)

| | | |
|---|---|---|
| GBRG2_HUMAN | RTGA**W**RHGRIHIRIAKMD | 442 |
| GBRG2_PONY | RTGA**W**RHGRIHIRIAKMD | 442 |
| GBRG2_MOUSE | RTGA**W**RHGRIHIRIAKMD | 449 |
| GBRG2_RAT | RTGA**W**RHGRIHIRIAKMD | 441 |
| GBRG2_BOVIN | RTGA**W**RHGRIHIRIAKMD | 450 |
| GBRG2_CHICK | RTGA**W**RHGRIHIRIAKMD | 449 |

C)

| | | |
|---|---|---|
| CLH1_HUMAN | YTPD**W**IFLLRNVMRISPD | 527 |
| CLH2_HUMAN | YTPD**W**IFLLRGVMKISPE | 527 |
| CLH_MOUSE | YTPD**W**IFLLRNVMRISPD | 527 |
| CLH_RAT | YTPD**W**IFLLRNVMRISPD | 527 |
| CLH_YEAST | YQPN**F**LVLISSLIRSSPD | 531 |
| CLH_DROME | YTPD**Y**VFLLRSVMRSNPE | 528 |
| CLH_DICDI | YKPD**F**MFLLQRMANANPM | 526 |

**Figure 5.** Sequence alignments of the GABARAP-binding proteins calreticulin, the γ2 subunit of the $GABA_A$ receptor, and the clathrin heavy chain reveal a highly conserved tryptophan over a broad variety of species. A) Alignment of amino acids 196–213 of human calreticulin (precursor protein) with homologues from five other species indicates conservation of W200 of human calreticulin. The arbitrarily chosen examples represent the 25 calreticulin sequences that are currently deposited in the Swiss-Prot database. Notably, all 25 exhibit the conserved tryptophan. B) Alignment of amino acids 425–442 of human $GABA_A$ receptor γ2 subunit (precursor protein) with all five orthologues that are found in the Swiss-Prot database shows 100% conservation of the tryptophan at position 429 in the human protein. The stretch of underlined amino acids 427–437 of the human $GABA_A$ receptor γ2 subunit has been shown to be sufficient for GABARAP binding.[13] Interestingly, this stretch contains the conserved tryptophan. C) Alignment of amino acids 510–527 of the human proteins clathrin heavy chain 1 and 2 (precursor proteins) with all five orthologues that are currently present in the Swiss-Prot database. W514 of human clathrin is conserved in the homologue proteins of mouse and rat, but conservatively replaced by other aromatic residues in the homologues of yeast, drosophila melanogaster, and dictyostelium discoideum. Remarkably, the GABARAP protein has been identified in human, mouse, and rat but not in the latter three species.

this hypothesis comes from: 1) the high abundance of tryptophan in GABARAP binding peptides that were selected by phage display screens, 2) the loss of GABARAP affinity upon a tryptophan-to-alanine replacement in a phage display-selected peptide, and 3) the strong conservation of a tryptophan residue in the GABARAP binding regions of calreticulin, clathrin heavy chain and the $GABA_A$ receptor γ2 subunit. Additional studies are required to decide whether the indole-binding site on GABARAP is also relevant for the binding of low-molecular-weight indolic compounds that are ubiquitous in and outside of cells.

## Experimental Section

**Materials**: GABARAP was recombinantly expressed with an N-terminal GST affinity tag in *E. coli* and purified as described previously.[32] Uniformly $^{15}N$-labeled protein was obtained by growing the bacteria in minimal medium supplemented with $^{15}N$-labeled ammonium chloride (Euriso-Top, Saarbrücken, Germany) as the sole nitrogen source. Indole and indole-3-acetic acid (IAA) were purchased from Aldrich. L-Tryptophan (L-Trp) and N-acetyl-L-tryptophan amide (NATA) were from Sigma. Perdeuterated [$D_6$]DMSO and $D_2O$ were obtained from Euriso-Top. The two peptides $COCH_3$-SLEDDWDFLPP-$NH_2$ (P1) and $COCH_3$-SLEDDADFLPP-$NH_2$ (P1(W6A)) were custom synthesized and purified to >95% by the core facility "Analytisches Zentrallabor" of BMFZ at the Heinrich-Heine University in Düsseldorf, Germany, and Jerini BioTools, (Berlin, Germany), respectively.

**NMR spectroscopy experiments**: Data were recorded at 25 °C and 14.1 T with a Varian Unity INOVA 600 instrument by using a 5 mm $^1H\{^{13}C,^{15}N\}$ triple resonance probe that was equipped with an actively shielded *z*-axis pulsed field gradient coil. Saturation transfer difference (STD) spectra were calculated from two $^1H$ NMR spectra that were recorded in an interleaved manner with 64 scans each with and without, respectively, selective protein irradiation at ~1 ppm for 3 s. The water signal was suppressed with the WATERGATE sequence and convolution of the time-domain data. Samples that contained GABARAP (0.1 mM) and ligand (~10 mM) in PBS buffer (pH 7.4) were prepared with a mixture of $^2H_2O$ and $H_2O$ (10 and 90%, respectively). The buffer of the indole sample contained [$D_6$]DMSO (10 vol%) to facilitate solubility of the hydrophobic ligand. Two-dimensional $^1H,^{15}N$ HSQC spectra of $^{15}N$-labeled GABARAP were recorded on samples that contained protein (between 87 and 192 μM) and varying amounts of ligand (from 0 to 42 mM) in buffer (25 mM sodium phosphate, pH 7.0, 100 mM NaCl, 100 mM KCl). Indole samples were supplemented with [$D_6$]DMSO (10 vol%).

Ligand concentration was verified in all samples by quantitative $^1H$ NMR spectroscopy based on comparison of the integral intensity of selected ligand signals with the trimethylsilyl peak of 2,2-dimethyl-2-silapentane-5-sulfonate (DSS), which was added to the samples as an internal standard.

**Determination of dissociation constants from NMR spectroscopy data**: Ligand-binding-induced changes of protein chemical shifts encode information on the population of the free and ligand-bound states of the protein. Under fast exchange conditions, an affected $^1H$–$^{15}N$ correlation peak in the HSQC spectrum moves gradually from the free to the bound state position upon titration with increasing amounts of ligand, and the observed chemical shifts are population-weighted averages over the shifts in the bound and free states. The total ligand and protein concentrations along with the fraction of complex at each titration step can be fitted to a simple binding isotherm; this yields the dissociation constant $K_d$ if the binding mode is compatible with a local one-to-one complex.[33] The composite chemical shift perturbation, $\Delta\delta$, that takes both $^1H$ and $^{15}N$ chemical shift changes of protein backbone amides into account was employed in the fitting procedure [Eq. (1)]:

$$\Delta\delta = \sqrt{[(\Delta\delta\,^1\mathrm{HN})^2 + (\Delta\delta\,^{15}\mathrm{N}/5)^2]} \quad (1)$$

**Molecular docking**: The molecular docking routine that was implemented in the software package GOLD v3.0.1 (Cambridge Crystallographic Data Centre)[34] was used to search the surface of the three-dimensional structure of GABARAP for low-energy indole-

  

binding sites. Ligands were docked within a radius of 10 Å around the $C_\alpha$ position of Y49 by using the GoldScore scoring function and default parameters.

**J-surface localization**: The software Jsurf (kindly provided by G. Moyna of USP in Philadelphia, USA and M. A. McCoy of Schering–Plough Research Institute) was used to derive a J-surface from HN chemical shift changes of GABARAP that were induced by the addition of IAA (28 mM). Atomic coordinates of GABARAP were taken from the crystal structure (PDB accession code 1 gnu).[12] Standard parameters were used, and the eleven amide protons of GABARAP that shifted by more than 0.05 ppm upon addition of IAA were considered in the calculation.

**Surface plasmon resonance**: Experiments were performed on a Biacore X instrument (GE Healthcare). GABARAP was amine-coupled to a CM5 sensor chip. Peptide was dissolved at various concentrations (100 nM to 1 mM) in buffer (10 mM HEPES, 150 mM NaCl, 3 mM EDTA, 0.005% SP20, pH 7.4) and aliquots (60 μL) were injected at a flow rate of 30 $\mu L\,min^{-1}$ at 21.5 °C. The BIAevaluation software package was used for data analysis.

**Sequence alignment**: Protein amino acid sequence alignment was performed by using ClustalW.[35]

## Acknowledgements

*This study was supported by a research grant from the Deutsche Forschungsgemeinschaft (DFG) to D.W. (Wi1472/5). We thank Prof. Antti Poso (University of Kuopio, Finland) for stimulating discussions and Olga Dietz for technical assistance. Prof. Guillermo Moyna of USP in Philadelphia and Dr. Mark McCoy of Schering–Plough Research Institute kindly provided the software Jsurf.*

**Keywords:** GABARAP · indole-binding pocket · molecular modeling · molecular recognition · NMR spectroscopy

[1] B. J. Mayer, P. K. Jackson, D. Baltimore, *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 627–631.
[2] M. A. Lemmon, J. E. Ladbury, *Biochemistry* **1994**, *33*, 5070–5076.
[3] J. M. Bradshaw, V. Mitaxov, G. Waksman, *J. Mol. Biol.* **1999**, *293*, 971–985.
[4] R. L. Macdonald, R. W. Olsen, *Annu. Rev. Neurosci.* **1994**, *17*, 569–602.
[5] H. Wang, F. K. Bedford, N. J. Brandon, S. J. Moss, R. W. Olsen, *Nature* **1999**, *397*, 69–72.
[6] J. E. Coyle, D. B. Nikolov, *Neuroscientist* **2003**, *9*, 205–216.
[7] M. Kneussel, S. Haverkamp, J. C. Fuhrmann, H. Wang, H. Wassle, R. W. Olsen, H. Betz, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 8594–8599.
[8] H. Wang, R. W. Olsen, *J. Neurochem.* **2000**, *75*, 644–655.
[9] Z. W. Chen, R. W. Olsen, *J. Neurochem.* **2007**, *100*, 279–294.
[10] S. Vernier-Magnin, S. Muller, M. Sallot, J. Radom, J. F. Musard, P. Adami, P. Dulieu, J. P. Remy-Martin, M. Jouvenot, A. Fraichard, *Biochem. Biophys. Res. Commun.* **2001**, *284*, 118–125.
[11] V. N. Bavro, M. Sola, A. Bracher, M. Kneussel, H. Betz, W. Weissenhorn, *EMBO Rep.* **2002**, *3*, 183–189.
[12] D. Knight, R. Harris, M. S. McAlister, J. P. Phelan, S. Geddes, S. J. Moss, P. B. Driscoll, N. H. Keep, *J. Biol. Chem.* **2002**, *277*, 5556–5561.
[13] J. E. Coyle, S. Qamar, K. R. Rajashankar, D. B. Nikolov, *Neuron* **2002**, *33*, 63–74.
[14] Y. Paz, Z. Elazar, D. Fass, *J. Biol. Chem.* **2000**, *275*, 25445–25450.
[15] K. Sugawara, N. N. Suzuki, Y. Fujioka, N. Mizushima, Y. Ohsumi, F. Inagaki, *Genes Cells* **2004**, *9*, 611–618.
[16] T. Stangler, L. M. Mayr, D. Willbold, *J. Biol. Chem.* **2002**, *277*, 13363–13366.
[17] J. Mohrlüder, T. Stangler, Y. Hoffmann, K. Wiesehan, A. Mataruga, D. Willbold, *FEBS J.* **2007**, *274*, 5543–5555.
[18] J. Mohrlüder, Y. Hoffmann, T. Stangler, K. Hänel, D. Willbold, *Biochemistry* **2007**, *46*, 14537–14543.
[19] M. Mayer, B. Meyer, *Angew. Chem.* **1999**, *111*, 1902–1906; *Angew. Chem. Int. Ed.* **1999**, *38*, 1784–1788.
[20] B. Meyer, T. Peters, *Angew. Chem.* **2003**, *115*, 890–918; *Angew. Chem. Int. Ed.* **2003**, *42*, 864–890.
[21] M. Görlach, M. Wittekind, R. A. Beckman, L. Mueller, G. Dreyfuss, *EMBO J.* **1992**, *11*, 3289–3295.
[22] Y. Chen, J. Reizer, M. H. Saier, Jr., W. J. Fairbrother, P. E. Wright, *Biochemistry* **1993**, *32*, 32–37.
[23] E. R. Zuiderweg, *Biochemistry* **2002**, *41*, 1–7.
[24] E. Kleinpeter, S. Klod, A. Koch, *J. Mol. Struct.: THEOCHEM* **2007**, *811*, 45–60.
[25] M. A. McCoy, D. F. Wyss, *J. Biomol. NMR* **2000**, *18*, 189–198.
[26] G. Moyna, R. J. Zauhar, H. J. Williams, R. J. Nachman, A. I. Scott, *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 702–709.
[27] M. A. McCoy, D. F. Wyss, *J. Am. Chem. Soc.* **2002**, *124*, 11758–11763.
[28] C. Landolt-Marticorena, K. A. Williams, C. M. Deber, R. A. Reithmeier, *J. Mol. Biol.* **1993**, *229*, 602–608.
[29] S. Jones, J. M. Thornton, *J. Mol. Biol.* **1997**, *272*, 121–132.
[30] D. A. Dougherty, *Science* **1996**, *271*, 163–168.
[31] W. M. Yau, W. C. Wimley, K. Gawrisch, S. H. White, *Biochemistry* **1998**, *37*, 14713–14718.
[32] T. Stangler, L. M. Mayr, A. J. Dingley, C. Luge, D. Willbold, *J. Biomol. NMR* **2001**, *21*, 183–184.
[33] P. Roesch, W. Klaus, M. Auer, R. S. Goody, *Biochemistry* **1989**, *28*, 4318–4325.
[34] G. Jones, P. Willett, R. C. Glen, A. R. Leach, R. Taylor, *J. Mol. Biol.* **1997**, *267*, 727–748.
[35] J. D. Thompson, D. G. Higgins, T. J. Gibson, *Nucleic Acids Res.* **1994**, *22*, 4673–4680.
[36] A. P. Hinck, M. A. Markus, S. Huang, S. Grzesiek, I. Kustonovich, D. E. Draper, D. A. Torchia, *J. Mol. Biol.* **1997**, *274*, 101–113.

Received: February 21, 2008
Published online on June 20, 2008

  