Domain architecture of a *Caenorhabditis elegans* AKAP suggests a novel AKAP function

Sanna Herrgård[a,b], Per Jambeck[a], Susan S. Taylor[c], Shankar Subramaniam[a,c,*]

[a] *Department of Bioengineering, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA*
[b] *Laboratory of Bioprocess Engineering, Helsinki University of Technology, P.O. Box 6100, 02015 Helsinki, Finland*
[c] *Department of Chemistry and Biochemistry, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA*

Received 4 October 2000; accepted 20 October 2000

First published online 21 November 2000

Edited by Shmuel Shaltiel

Abstract A-kinase anchoring proteins (AKAPs) are adapter proteins that are involved in directing cAMP-dependent protein kinase and some other signaling enzymes to certain intracellular locations. In this study, we investigate the domain architecture of an AKAP from *Caenorhabditis elegans* ($AKAP_{CE}$). We show that $AKAP_{CE}$ shares two domains with the Smad anchor for receptor activation, a FYVE-finger and a transforming growth factor β (TGFβ) receptor binding domain, suggesting that $AKAP_{CE}$ may interact with a receptor belonging to the TGFβ receptor family. This predicted novel AKAP function supports the recent view of AKAPs as adapter proteins that can be involved in various signaling pathways.

Key words: A-kinase anchoring protein; Domain architecture; Transforming growth factor β signaling pathway; Smad anchor for receptor activation; *Caenorhabditis elegans*

1. Introduction

A-kinase anchoring proteins (AKAPs) are a family of proteins involved in targeting cAMP-dependent protein kinase (PKA) to a variety of intracellular compartments [1]. Although structurally diverse, the AKAPs have a modular structure and share an α-helical domain corresponding to the PKA-binding site. The model for the function of AKAPs has evolved in the past years significantly [2]. According to the initial AKAP model, AKAPs were proteins consisting of a PKA-binding domain and a targeting domain. It was shown that different AKAPs were targeted to different cellular locations, e.g. to mitochondria, centrosomes, actin cytoskeleton, the Golgi, microtubules, plasma membrane, vesicles, endoplasmic reticulum, dendrites and nuclear membrane (reviewed in [1,3,4]). Along with the discovery that some AKAPs were able to bind not only PKA, but also other signaling enzymes, it became apparent that AKAPs have the ability to function as scaffolding proteins. Further, the latest studies have shown that some AKAPs act as bridges in multiunit signaling complexes [5].

In this study, we use sequence-based bioinformatics methods to investigate the domain architecture of a *Caenorhabditis elegans* AKAP ($AKAP_{CE}$) [6,7], and we propose a novel function for $AKAP_{CE}$ as a transforming growth factor β (TGFβ) receptor binding protein. This study is also an example of the ways in which sequence-based bioinformatics methods can be applied to identify novel domains in multi-domain signaling proteins, and to formulate experimentally testable hypotheses about their function.

2. Materials and methods

2.1. Sequences used in this study

The GenBank identifiers of the sequences used in this study are: AKAP-*C. elegans*, g3258651; SARA-*Homo sapiens*, g4759060; SARA-*Xenopus laevis*, g4092769; SARA-*Drosophila melanogaster*, g7595827; KIAA0305-*H. sapiens*, g2224551.

2.2. Methods

PSI-BLAST [8] was used for searching the NCBI non-redundant protein sequence database for sequence similarities. PSI-BLAST performs several iterative BLAST searches, using a position-specific score matrix generated from the significant hits found in the previous run. Pairwise BLAST [8,9] was used for finding local similarities in protein sequences and for generating local alignments. Other sequence alignments were generated with ClustalW [10,11] and shaded by T_EXshade [12]. Gene structure predictions were performed with GENSCAN [13].

The program MEME [14] was used to discover conserved sequence motifs in a set of sequences, and Meta-MEME [15] and MAST [16] were used to detect homologs sharing these motifs. MEME uses expectation maximization to identify motifs in a set of DNA or protein sequences [14]. Meta-MEME combines these motifs into a motif-based linear hidden Markov model (HMM), which is used to search for homologs sharing the motifs [15]. MAST is an alternative tool that can be used to search for homologs sharing a set of motifs [16]. MAST compares all sequences in a sequence database to a set of motifs, and calculates the corresponding *P*-values for matches.

A HMM of the PKA-binding domain was generated using HMMer [17]. Starting from an aligned set of sequences, HMMer builds a profile HMM that can be used to search for similar sequences. The difference between the motif HMM generated by Meta-MEME and the profile HMM generated by HMMer is that the motif model is based on modeling ungapped blocks of sequence motifs, whereas the profile model allows insertions and deletions anywhere in the target sequence [18].

The InterPro [19] and SMART [20] databases were searched to identify functional sites and domains in $AKAP_{CE}$. InterPro combines several databases that contain annotations for protein families, domains, motifs and functional sites [19]. SMART is a database of domains that exist in signaling proteins [20].

*Corresponding author. Fax: (1)-858-822 3752.
E-mail: shankar@ucsd.edu

Abbreviations: AKAP, A-kinase anchoring protein; $AKAP_{CE}$, *Caenorhabditis elegans* AKAP; HMM, hidden Markov model; PKA, cAMP-dependent protein kinase; SARA, Smad anchor for receptor activation; TGFβ, transforming growth factor β

0014-5793/00/$20.00
PII: S0014-5793(00)02201-8

Secondary structure predictions were performed using Jpred[2] [21] and PHD [22]. Jpred[2] combines predictions from six different secondary structure prediction algorithms that all use evolutionary information obtained from homologous sequences [21]. One of the secondary structure prediction methods included in Jpred[2] is PHD, a neural network-based method that has been reported to predict protein secondary structure at better than 72% accuracy [22].

3. Results

Searching the NCBI non-redundant protein sequence database using PSI-BLAST reveals four significant sequence homologs (*E*-values $< 10^{-25}$) for $AKAP_{CE}$ in the initial search: SARA (Smad anchor for receptor activation) from *X. laevis*, *H. sapiens* and *D. melanogaster*, and an uncharacterized protein from *H. sapiens* (KIAA0305). Our sequence analysis shows that $AKAP_{CE}$ shares two domains with SARA, which is involved in anchoring Smad2 and Smad3 to the TGFβ receptor [23,24]. One domain, the FYVE zinc-finger, has been previously characterized [25–29]. The other, a novel ~530 residue domain, has been found to mediate the interaction between SARA and the TGFβ receptor [23]. We call this domain the TGFβ receptor binding domain. Fig. 1 presents the domain architecture of human and *Xenopus* SARA, and our predictions for the domain architecture of *Drosophila* SARA, $AKAP_{CE}$ and KIAA0305.

The FYVE-finger resembles a double zinc-finger domain and binds specifically to intracellular membranes which contain phosphatidylinositol-3-phosphate [25–27]. We find that $AKAP_{CE}$ and its homologs (the SARA proteins and KIAA0305) have well-conserved FYVE-finger domains (Fig. 2a). However, in $AKAP_{CE}$, the second of the eight conserved cysteine residues (which coordinate two Zn^{2+} ions in FYVE-finger proteins [28]) is replaced by a gap, and an extra cysteine is located six residues upstream, at position 544.

The interaction between SARA and the TGFβ receptor is mediated by regions located at the C-terminal region of the protein [23]. The C-terminal region, which is shown in Fig. 2b, is well conserved in $AKAP_{CE}$ (*E*-value against the TGFβ receptor binding domain in human SARA is 7×10^{-28}) and in KIAA0305 (*E*-value is 10^{-171}). In addition, secondary structure predictions of the $AKAP_{CE}$ and its homologs are in agreement for this domain. The domain has several well-conserved sequence motifs, indicating that this ~530 amino acid domain is functionally important. We observe two highly conserved regions within the domain (see Fig. 2b), implying some functional significance for these regions. No other sequence homologs could be detected for this domain by using motif-based sequence search programs Meta-MEME and MAST, and the domain was not detected by InterPro (release 1.0) or SMART.

Residues 236–255 have been shown experimentally to contain the PKA-binding site in $AKAP_{CE}$ [6], and it is postulated that an amphipathic α-helix is likely involved in the binding. So far, no PKA-binding site has been identified in SARA. In order to investigate whether SARA contains a possible PKA-binding site, a profile HMM was generated using 17 known PKA-binding sites from 15 different AKAPs. The model is unable to identify any statistically significant putative PKA-binding site in SARA.

It is not known whether $AKAP_{CE}$ or KIAA0305 binds Smads. We identified 22 different proteins that all bind Smads (most of them are reviewed by Wrana [30]). Using pairwise BLAST and comparative secondary structure predictions, we conclude that the Smad-binding domains are not well conserved at their sequence or secondary structure level. We were able to identify a putative Smad-binding site in KIAA0305 (residues 814–865, *E*-value 4×10^{-5}), which is similar to the Smad-binding site identified in SARA. No potential Smad-binding site could be identified in $AKAP_{CE}$.

4. Discussion

During the past years, the amount of available DNA and protein sequence data has been growing exponentially. This rapid growth renders it difficult to determine the function of each gene or protein experimentally. In order to facilitate a faster annotation process of unknown genes and proteins, a wide variety of sequence analysis methods have been developed, and are now widely used and accepted.

In this study, we have used sequence-based bioinformatics methods to understand the functional role of $AKAP_{CE}$. $AKAP_{CE}$ is an example of an adapter protein that consists

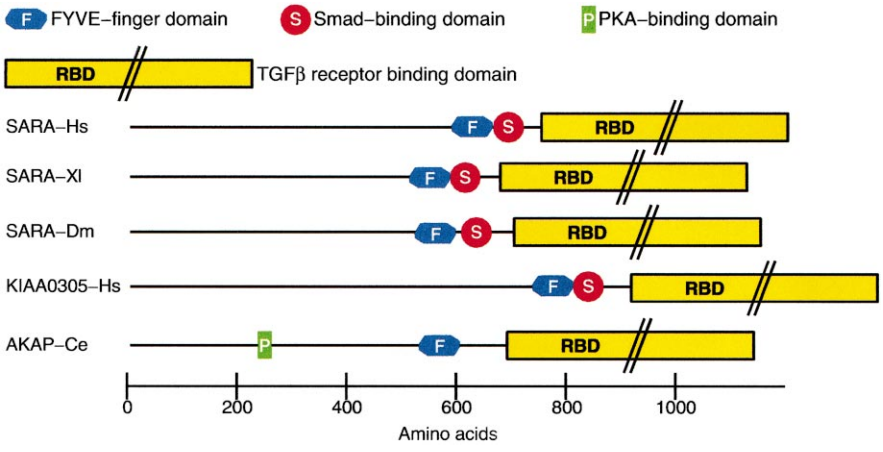

Fig. 1. Domain architecture of the $AKAP_{CE}$ homologs. Abbreviations of the species used: Hs, *H. sapiens*; Xl, *X. laevis*; Dm, *D. melanogaster*; Ce, *C. elegans*. The presence of a Smad-binding domain in KIAA0305 is uncertain. The extent of the TGFβ receptor binding domain has been truncated.

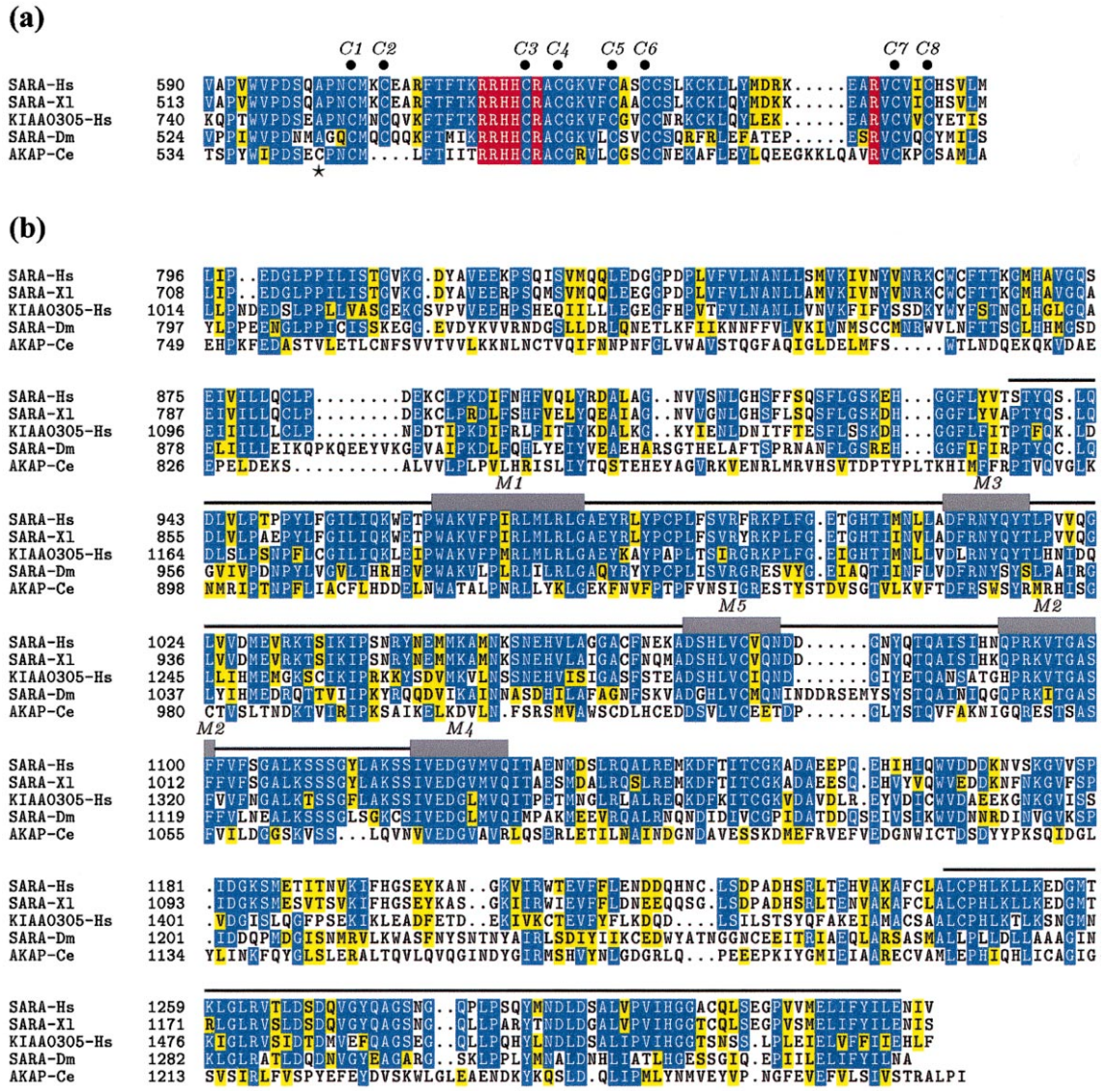

Fig. 2. (a) Sequence alignment of the FYVE-finger domain in $AKAP_{CE}$ homologs. The conserved cysteines [28] are identified by C1 to C8, and residues forming the conserved basic patch [28,29] are colored red. The extra cysteine observed at position 544 in $AKAP_{CE}$ is marked by a star. (b) Sequence alignment of the TGFβ receptor binding domain. The five most significant motifs generated by MEME are identified by M1 to M5 in the order of statistical significance, and the two most conserved regions are labeled with a bar above the figure. For species abbreviations, see Fig. 1.

of multiple domains, each domain having its own structure and function. When annotating the function of a multi-domain protein, the function of each individual domain should be identified, if possible. The function of a multi-domain protein is the sum of the functions of all of its domains, and each individual domain can reveal a new functionality for the protein.

We find that one caveat in analyzing the architecture of any multi-domain protein is related to predicting gene structure. Errors in identifying intron/exon boundaries can affect the apparent organization of domains. For example, in our study, the original SARA homolog that we identified in *Drosophila* (GenBank identifier g7291295) seemed to lack a few C-terminal residues of the FYVE-finger domain and part of the Smad-binding domain. When the gene structure was re-predicted using GENSCAN, we observed that the resulting translated sequence did contain the missing parts of the FYVE-finger and Smad-binding domains.

Our sequence analysis shows that $AKAP_{CE}$ shares two common domains with SARA, the FYVE-finger domain and the TGFβ receptor binding domain. The existence of these two domains in $AKAP_{CE}$ leads us to propose that $AKAP_{CE}$ binds to a receptor belonging to the TGFβ receptor family. Fig. 1 shows that the order and extent of the FYVE-finger domain and the TGFβ receptor binding domain are similar in $AKAP_{CE}$ and in SARA. The ~65 residue FYVE-finger domain is located around residue 600 and the TGFβ receptor binding domain is located near the C-terminus.

The structure and function of the FYVE-finger domain has been well established, thus the key functional residues and the boundaries of this domain are known [25–29,31,32]. We observe a four residue gap around the second conserved cysteine

in the $AKAP_{CE}$ FYVE-finger domain. A multiple sequence alignment of 17 FYVE-finger domains presented by Kutateladze et al. [33] shows three sequences that have two or three residue long gaps around the first conserved cysteine, indicating that gaps are tolerated in this area.

The TGFβ receptor binding domain has not been characterized earlier, and the extent of this domain and the details of the interactions between the receptor and SARA are largely unknown. We observe that the sequence similarity between SARA, KIAA0305 and $AKAP_{CE}$ is statistically highly significant for this domain. Two highly conserved regions and several conserved sequence motifs were identified, suggesting structurally or functionally important regions. The fully conserved residues in the sequence motifs would be excellent candidates for mutational studies to characterize the interaction between the domain and the receptor.

There are two domains in SARA and in $AKAP_{CE}$ that differ, the PKA-binding domain in $AKAP_{CE}$ and the Smad-binding domain in SARA. The PKA-binding domain identified in $AKAP_{CE}$ is located in the N-terminal part of the protein. So far, no domains have been identified in the N-terminal part of SARA, which covers 500 or 600 residues, depending on species. Thus, it is highly probable that the N-terminal part of SARA contains one or more domains. Our attempt to use a HMM to identify a putative PKA-binding domain in SARA did not yield any statistically significant hits. However, given that the PKA-binding domain is short ($\sim$20 residues) and shows only a limited number of conserved residues, no definite conclusions can be drawn at this point about the presence of a PKA-binding domain. Recent studies also show that, in addition to the α-helical PKA-binding domain most often found in AKAPs, other types of PKA-binding domains do exist. Diviani et al. [34] report that pericentrin binds PKA through a binding domain that is structurally different from the one traditionally observed in AKAPs.

The second differing domain is the Smad-binding domain, which is located immediately after the FYVE-finger domain in SARA. The structure of the SARA Smad-binding domain bound to the MH2-domain of Smad2 has been determined [24], indicating that the domain has a fold of its own. Our study of 22 Smad-binding proteins shows that the sequence similarity between the Smad-binding domains is generally low, and there are probably many different types of Smad-binding domains. Thus, our finding that $AKAP_{CE}$ does not have a region with a high sequence similarity to any known Smad-binding domain does not guarantee that $AKAP_{CE}$ would not bind Smads.

In summary, we have identified putative FYVE-finger and TGFβ receptor binding domains in $AKAP_{CE}$, suggesting that $AKAP_{CE}$ may interact with the TGFβ signaling pathway by recruiting PKA to a receptor belonging to the TGFβ receptor family. Our findings are supported by recent experiments, which show that PKA is stimulated by three TGFβ superfamily members: TGFβ1, activin and bone morphogenetic protein 2 [35–37]. Although the TGFβ signaling pathway is known to interact with several other signaling pathways, no experimental information exists about the possible role of AKAPs in TGFβ signaling. Thus, the interaction proposed in this study is novel. This study also demonstrates that AKAPs are adapter proteins that control specificity and selectivity within various signaling pathways.

Acknowledgements: The authors are grateful to Dr. Eric Beitz for his kind help with the $T_EX$shade program. Professor Daniel Donoghue is acknowledged for useful discussions. The authors thank the NSF (Grant KDI 98-73384) for financial support. S.H. was supported by grants from Academy of Finland, Foundation of Technology (Finland) and Helsinki University of Technology Foundation for Financial Aid.

References

[1] Colledge, M. and Scott, J.D. (1999) Trends Cell Biol. 9, 216–221.
[2] Dodge, K. and Scott, J.D. (2000) FEBS Lett. 476, 58–61.
[3] Scott, J.D., Dell'Acqua, M.L., Fraser, I.D.C., Tavalin, S.J. and Lester, L.B. (2000) Adv. Pharmacol. 47, 175–207.
[4] Edwards, A.S. and Scott, J.D. (2000) Curr. Opin. Cell Biol. 12, 217–221.
[5] Fraser, I.D.C., Cong, M., Kim, J., Rollins, E.N., Daaka, Y., Lefkowitz, R.J. and Scott, J.D. (2000) Curr. Biol. 10, 409–412.
[6] Angelo, R. and Rubin, C.S. (1998) J. Biol. Chem. 273, 14633–14643.
[7] Angelo, R.G. and Rubin, C.S. (2000) J. Biol. Chem. 275, 4351–4362.
[8] Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Nucleic Acids Res. 25, 3389–3402.
[9] Tatusova, T.A. and Madden, T.L. (1999) FEMS Microbiol. Lett. 174, 247–250.
[10] Higgins, D.G., Bleasby, A.J. and Fuchs, R. (1992) Comput. Appl. Biosci. 8, 189–191.
[11] Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) Nucleic Acids Res. 22, 4673–4680.
[12] Beitz, E. (2000) Bioinformatics 16, 135–139.
[13] Burge, C. and Karlin, S. (1997) J. Mol. Biol. 268, 78–94.
[14] Bailey, T.L. and Elkan, C. (1995) in: Proceedings of the Third International Conference on Intelligent Systems for Molecular Biology (Rawlings, C., Clark, D., Altman, R., Hunter, L.C. and Rawlings, L.C., Eds.), pp. 21–29, AAAI Press, Menlo Park.
[15] Grundy, W.N., Bailey, T.L., Elkan, C.P. and Baker, M.E. (1997) Comput. Appl. Biosci. 13, 397–406.
[16] Bailey, T.L. and Gribskov, M. (1998) Bioinformatics 14, 48–54.
[17] Eddy, S.R. (1995) in: Proceedings of the Third International Conference on Intelligent Systems for Molecular Biology (Rawlings, C., Clark, D., Altman, R., Hunter, L.C. and Rawlings, L.C., Eds.), pp. 114–120, AAAI Press, Menlo Park.
[18] Eddy, S.R. (1998) Bioinformatics 14, 755–763.
[19] Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Birney, E., Bucher, P., Codani, J.-J., Corpet, F., Croning, M.D.R., Durbin, R., Etzold, T., Fleischmann, W., Gouzy, J., Hermjakob, H., Jonassen, I., Kahn, D., Kanapin, A., Schneider, R., Servant, F. and Zdobnov, E. (2000) CCP11 Newsletter Issue 10 (1 March 2000).
[20] Schultz, J., Milpetz, F., Bork, P. and Ponting, C.P. (1998) Proc. Natl. Acad. Sci. USA 95, 5857–5864.
[21] Cuff, J.A., Clamp, M.E., Siddiqui, A.S., Finlay, M. and Barton, G.J. (1998) Bioinformatics 14, 892–893.
[22] Rost, B. (1996) Methods Enzymol. 266, 525–539.
[23] Tsukazaki, T., Chiang, T.A., Davison, A.F., Attisano, L. and Wrana, J.L. (1998) Cell 95, 779–791.
[24] Wu, G., Chen, Y.-G., Ozdamar, B., Gyuricza, C.A., Chong, P.A., Wrana, J.L., Massagué, J. and Shi, Y. (2000) Science 287, 92–97.
[25] Burd, C.G. and Emr, S.D. (1998) Mol. Cell 2, 157–162.
[26] Gaullier, J.-M., Simonsen, A., D'Arrigo, A., Bremnes, B., Stenmark, H. and Aasland, R. (1998) Nature 394, 432–433.
[27] Patki, V., Lawe, D.C., Corvera, S., Virbasius, J.V. and Chawla, A. (1998) Nature 394, 433–434.
[28] Stenmark, H., Aasland, R., Toh, B.-H. and D'Arrigo, A. (1996) J. Biol. Chem. 271, 24048–24054.
[29] Stenmark, H. and Aasland, R. (1999) J. Cell Sci. 112, 4175–4183.
[30] Wrana, J.L. (2000) Science's STKE (14 March 2000).
[31] Misra, S. and Hurley, J.H. (1999) Cell 97, 657–666.
[32] Mao, Y., Nickitenko, A., Duan, X., Lloyd, T.E., Wu, M.N., Bellen, H. and Quiocho, F.A. (2000) Cell 100, 447–456.

[33] Kutateladze, T.G., Ogburn, K.D., Watson, W.T., de Beer, T., Emr, S.D., Burd, C.G. and Overduin, M. (1999) Mol. Cell 3, 805–811.
[34] Diviani, D., Langeberg, L.K., Doxsey, S.J. and Scott, J.D. (2000) Curr. Biol. 10, 417–420.
[35] Wang, L., Zhu, Y. and Sharma, K. (1998) J. Biol. Chem. 273, 8522–8527.
[36] Joore, J., van de Water, S., Betist, M., van den Eijnden-van Raaij, A. and Zivkovic, D. (1998) Mech. Dev. 79, 5–15.
[37] Gupta, I.R., Piscione, T.D., Grisaru, S., Phan, T., Macias-Silva, M., Zhou, X., Whiteside, C., Wrana, J.L. and Rosenblum, N.D. (1999) J. Biol. Chem. 274, 26305–26314.