(MAPKK) and ERK (MAPK), which in turn generate downstream signals. Currently the understanding of the pathway target genes is very limited — it is only known that they have AP-1 and Elk-1 binding sites[3] in their promoter regions, and there is very little known about what transcription factors (TF) regulate the pathway itself. Transcriptional analysis of pathways can be a complicated task, because the functional groups of genes may not be clear and often overlap, i.e. a single gene may play multiple roles.

In this work we made the first attempt to determine candidates for regulatory elements of genes involved in the MAP kinase G-protein coupled pathway and compare the results with an ubiquitously expressed dataset of ribosomal genes. Also we tried to gain insight into the relationships of genes in PPI pairs in terms of co-regulation.

1.2. *Comparative genomic analysis*

Confidence in identification of *cis*-elements (motifs) in the regulatory regions of higher eukaryotic genomes cannot be achieved based on a single match to a score matrix, but rather requires some supporting evidence. One source of such evidence is the predicted occurrence of the same *cis*-elements in the orthologous promoters of other species. Stretches of conserved nucleotide sequences in the promoters of orthologous genes give a strong suggestion of their functional importance, possibly for regulating expression. Comprehensive studies in this direction have been performed on various data.[4,5] When investigating transcription regulation, we have to consider both the comparatively short promoter regions in the proximity of TSSs, where the formation of the RNA polymerase II initiation complex occurs, as well as distantly located enhancer regions interacting with one or more promoter sites[6,7] of neighboring genes and whose distance from the TSS can exceed 100 kb.

Nevertheless many motif finding tools have been constructed so far and few of them are appropriate for the analysis of several genomes simultaneously,[8,9] it seems that the problem is not satisfactory solved in a practical sense by any of them for the comparative analysis of higher eukaryotes. The difficulties are that orthologous motifs can only partially match the consensus pattern for they belong to and their occurrence and locations exhibit significant irregularities.[6] In an attempt to overcome the above problems we have developed an original comparative genomic analysis tool *SHOE* (in preparation), which maintains three genomes simultaneously at the present (human, mouse and rat) and searches for the reliable orthologous motifs candidates.

2. Data

The MAPK experimental PPI dataset used contains 105 interacting pairs, involving 68 genes. The genes are classified into four groups by their functional role in the pathway: 25 Receptors (R), 15 MAP core genes (M), 6 Receptor-MAP core genes (RM), 7 Effectors (E), 14 genes with unknown function in the pathway (U) and 1

gene in a group of others (O), which are the scaffold genes, having the fixation role for the interacting proteins in the pathway. The general view of the pathway is depicted in Fig. 1, and the number of genes for which the alignment across three species was obtained is pointed to by arrows (discussed in results).

The ribosomal dataset includes 58 genes. Orthologous genes for mouse and rat were obtained from Mammalian orthology database[11] and transcriptional start sites (TSSs) for human and mouse verified in DBTSS[10] for available genes. Regulatory regions were defined from +200 base pairs downstream of the TSSs up to the beginning of the adjacent upstream gene. The adjacent gene regions were taken from DBTSS and Refseq.[12] Gene regions were defined from either the TSS (when available) or 5'UTR to the 3'UTR.

3. Method

3.1. *Co-expression from microarray data*

Due to the fact that genes with similar transcriptional profiles might be influenced by the same transcriptional mechanism we calculated *Pearson correlation coefficients* for all gene pairs in the datasets of MAPK pathway genes (MAPK: 2278 pairs) and ribosomal genes (RIB: 1653 pairs). Calculations were done based on 112 "eigen cell" synthetic expression profiles available from the Cell Montage web site (http://cellmontage.cbrc.jp/). The Eigen cell profiles were obtained by principal component analysis of 1714 normal human cell profiles, selected from profiles in the Gene Expression Omnibus Database.[13] The distribution of gene pairs with various correlation coefficients is slightly different between the two datasets, as it is shown in Fig. 2 below. This may be due to the fact that ribosomal genes, composing small and large ribosomal subunits, are closer functionally than genes involved in the whole MAPK pathway. For further analysis we used pairs with correlation

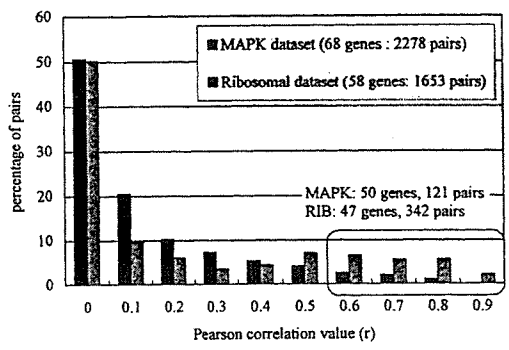

Fig. 2. Distribution of *co*-regulated gene pairs in the MAPK and ribosomal (RIB) datasets for various correlation coefficients (r). It is evident that the ribosomal dataset includes more highly *co*-regulated pairs (324 pairs) than the MAPK dataset (121 pairs).

coefficient $r \geq 0.6$. In general, the issue of determining and appropriate threshold is complicated.[14]

3.2. *Transcriptional analysis with SHOE*

3.2.1. *Aligning orthologous promoters*

Each human gene regulatory region from highly *co*-regulated groups was aligned to its orthologous partners (mouse and rat) by the SSEARCH local alignment program, and in the cases where the same human promoter region was successfully aligned to mouse and rat promoters, all three respective regions were extracted and realigned with ClustalW.

To evaluate the degree to which an observed region in an alignment of the three species should contribute to our belief that it is part of a conserved region, we adopted an estimate of the likelihood ratio of observing the region in an alignment of orthologous promoters versus that of observing the region in an alignment of unrelated promoters. To obtain this estimate we repeatedly aligned orthologous promoters triplets and randomly grouped promoter triplets and observed the frequency of each possible alignment column. We refer to the obtained alignment column frequencies as the *good alignment* and *random alignment* frequencies respectively.

Using those column *frequency tables* we define the multiple alignment score (***MAscore***) as:

$$MAscore = \log_{10} \frac{\Pr/g1 * \Pr/g2 * \Pr/g3 \cdots \Pr/gn}{\Pr/r1 * \Pr/r2 * \Pr/r3 \cdots \Pr/rn} = \log_{10} \frac{\prod_m \mathrm{pr}(c|good\_alignment)}{\prod_m \mathrm{pr}(c|random\_alignment)},$$

where c is the probability of pattern in each column is calculated using patterns frequencies from *good alignment* and *random alignment* tables and m is the length of the alignment region, which for our application is equal to the motif length m because we apply this score as a supporting score to the motifs score (discussed below).

3.2.2. *Matching with known motifs*

Position Specific Scoring Matrices (PSSMs) of transcription factors binding sites (TFBSs) from Transfac[15] (version 8.4) for human, mouse and rat were applied to the regions in human found conserved in other two genomes and PSSM score (***PMscore*** here) is calculated for each position in the region by

$$PMscore = \sum_{i=1}^{m} \log_{10} \frac{count_{xi} + pseudocount_{xi}}{\sum_{x=A,T,G,C} count_{xi} + \sum_{x=A,T,G,C} pseudocount_{xi}},$$

where $pseudocount = 1$ and m is a motif length. The sum of *MAscore* and *PMscore* is used then for the selection of the binding site candidates.

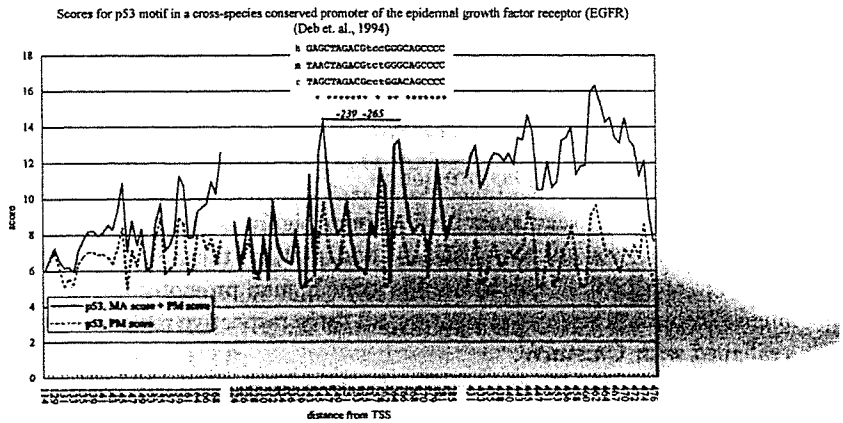

Fig. 3. Distribution of scores for p53 motif in a promoter of the epidermal growth factor receptor (EGFR). The solid line represents the sum of *MAscore* and *PMscore*. We observe the increase of both scores around the position of the motif location: −239 −265 nt. from the TSS. Three alignments blocks are shown: −114 to −168 nt., −224 to −285 nt. and −431 to −476 nt. from TSSs. The disruptions of lines reveal the borders of aligned blocks.

3.2.3. *Defining threshold of the motifs significance*

How high must *MAscore* and *PMscore* be to distinguish the correct motif candidates from the wrong ones? To determine the cutoff threshold we selected several experimentally verified examples of human motifs and calculated the scores for samples found conserved across three species. Two examples are demonstrated in Figs. 3 and 4.

Looking at the two examples in Fig. 3 and Fig. 4 we notice that the sum of scores is comparatively high at the positions where experimentally determined motifs are located. This is a reason to emphasize the importance of the *MAscore* as a powerful supporting evidence for reliable identification of binding sites. Upon inspection of several well characterized transcription binding sites (some not shown) we adopted a threshold of 10 for the sum of scores. Motif length does influence the score and thus the threshold could be adjusted for length, however a single threshold seemed to be sufficient over the limited range of motif lengths (5–15, average 9) used in this study.

4. Results

4.1. *Regulatory motifs in the MAPK and ribosomal datasets*

4.1.1. *Calculation of Motif Conservation Score*

The common motif candidates with the sum of scores exceeding the defined threshold in highly *co*-regulated pairs of genes were selected in both datasets. From 121 pairs (50 genes) of the MAPK dataset with $r \geq 0.6$ (Fig. 2) only 37 pairs (39 genes)

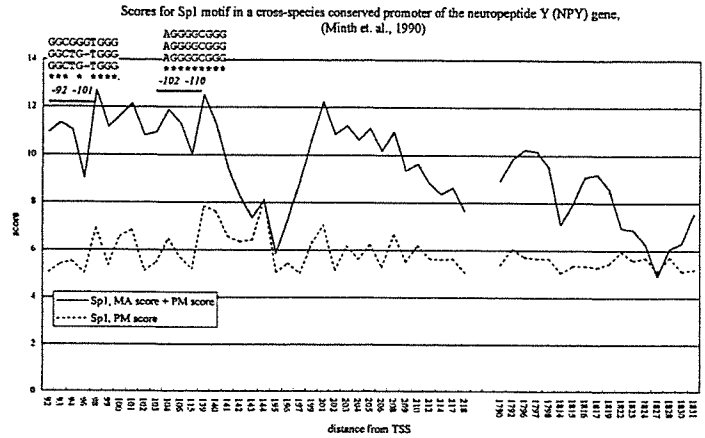

Fig. 4. Distribution of scores for Sp1 motifs in a promoter of the neuropeptipe Y gene (NPY). The solid line represents the sum of *MAscore* and *PMscore*. We observe the increase of both scores around the positions of the motif locations: -92 $-101\,nt.$ and -102 $-110\,nt.$ from TSSs. Two alignments blocks are shown: -92 to $-218\,nt.$ and -1790 to $-183\,1nt.$ from the TSSs. The motifs are aligned in order: human-mouse-rat. The disruptions of lines reveal the borders of aligned blocks.

were determined to have common motifs above the threshold in their regulatory regions. In the ribosomal dataset with 324 highly *co*-regulated pairs (40 genes), 160 pairs (40 genes) had common motifs. For this analysis we excluded candidate sites with a distance greater than 100 kb from the TSS.

As a result of the above selection 33 transcription factors (TFs) were found common for MAPK and ribosomal datasets, 32 factors were unique for MAPK dataset and 1 TF was unique for ribosomal dataset. We refer to these motifs as motifs from the "first round of selection". To estimate the specificity of motifs belonging to both datasets we calculated the Motif Conservation Score (MCS)[5] of a motif $m$by comparing its frequencies in the MAPK and ribosomal datasets (p) by

$$\mathrm{MCS} = (\mathrm{K} - \mathrm{Np_0})/[\mathrm{Np_0}(1 - \mathrm{p_0})]^{1/2}$$

where $\mathbf{K}$ — number of pairs with common motif in MAPK, $\mathbf{N}$ — total number of pairs in MAPK, $\mathbf{p_0}$ is a frequency of the same motif of in ribosomal dataset. MCSs with the respective TFs are shown in Table 3 and Table 4.

4.1.2. *Re-estimation of the motifs significance*

To gain more confidence in the regulatory roles of the transcription factors (TFs), whose motifs were found in the promoters of genes in pairs, we also calculated the correlation coefficient of expression between those TFs and the genes they potentially regulate. If the correlation coefficient of the TF and gene was greater than 0.6, we keep the motif as a reliable candidate binding site of the respective

TF. The motifs which successfully passed this second round of selection are listed in Table 1 and Table 2 with the highly *co*-regulated genes pairs. For the ribosomal dataset 50 representative pairs are shown. The last column in the Table 1 shows the function of genes included in pairs and, for example, "R-R" points to the Receptor-Receptor pair and "R-E" points to the "Receptor-Effector" pair. Table 2 includes the same information about the ribosomal genes with the exception of functional characteristics. Genes from large ribosomal subunit genes have "L" in their ID (RPL), and genes from small ribosomal subunit genes have "S" (RPS).

The final result of our screening yielded 19 candidate regulating TFs for the MAPK dataset 12 for the ribosomal dataset. The functions of the TFs in both groups are listed in Table 3 and Table 4.

5. Discussion

The results of our analysis showed that a completely different set of TFs, except three TFs: PPARG, LUN-1, NF-Y, were obtained for MAPK (Table 3) and the ribosomal datasets (Table 4). Although the number of TFs decreased markedly in the second round of selection, we note that 10 out of 19 MAPK TFs were unique to the MAPK dataset (mark "M") even after the first round of selection. Four TFs: GR, STAT5A, OLF-1, VDR have comparatively high MCS, indicating that they are more prevalent in the regulatory regions of MAPK rather than in ribosomal regulatory regions. Two other TFs: PPARG and LUN-1 with non-significant MCS values in the first round of selection, were found to be highly *co*-regulated with genes in both datasets, and only NF-Y TF finally showed a preference for the ribosomal dataset, which is in contradiction with the results of the first round of selection. This result generally supports the correctness of our approach, but we have to sharpen the method determining false positive motifs.

Looking at the function of TFs in Table 3 we noticed an interesting tendency that 6 out of 19 MAPK TFs are receptors. This might be somehow related to the fact that 26 out of 37 highly *co*-regulated pairs include receptor genes (11 R-M pairs, 9 R-R pairs, 3 RM-R pairs and 3 RM-M pairs). All but three highly *co*-regulated pairs show reasonable functional relationships, and it sufficiently meets the idea that functionally related genes influenced by similar regulatory mechanism. In these three pairs both partners are known and they are Receptor-Effector pairs, which relationships are difficult to be explained as they might have distant locations in the cell. High *co*-regulation of receptor genes and MAP-core genes combined with the comparatively strong conservation of their regulatory regions over three species tells us about the evolutionary importance of these genes for the functionality of the G-protein coupled MAPK pathway. Because these genes are on the top of the pathway (Fig. 1) they may have an indispensable role for the adaptation of the signal from the outside of the cell and passing it to the downstream target genes.

Table 1. *Co*-regulated pairs (37 pairs, 39 genes) of MAPK dataset with motifs conserved in the promoters of genes across three species and whose transcription factors were found to be *co*-expressed with genes in pairs are shown. In column "FUNCTION" the functional category of the genes in the pathway are listed. For example, "R-R" points to the Receptor-Receptor pair and "R-E" points to the "Receptor-Effector" pair. UNG means the Unique Number of Genes possessing motif and PCM — number of Pairs with the respective Common Motif (single copy per pair is counted). In the last two rows average distance from TSSs and standard deviations (SD) of distances are shown.

Co- regulated pairs of MAPK genes (r≥0.6) and binding sites, which transcription factors are found in a high *co* -regulation (r≥0.6) with genes in pairs																					FUNCTION
AP2B1-BRAF	GR		PPARG																		R-M
AP2B1-CBL	GR		PPARG																		R-R
ARAF-SHC1																					R-R
CBLB-CRK	GR			LUN-1	STAT1	AP-4	NF-kappaB			AhR											R-R
CBLB-GRB2	GR			LUN-1	STAT1	AP-4	NF-kappaB	C/EBPgamma	IRF-1					Max					VDR			R-R
CBL-BRAF	GR		PPARG								Olf-1											R-M
CD2AP-CBLB	GR				STAT1																	R-R
CD2AP-GRB2					STAT1																	R-R
DNM1-EIF5A		STAT5A																				R-E
DNM1-EPS15L1		STAT5A																		c-Myb		R-R
DNM2-MAP2K1	GR	STAT5A	PPARG																			R-M
GNB1-BRAF	GR		PPARG								Olf-1		ER		HSF1	Evi-1	CREB					R-M
GNB1-CBL	GR		PPARG								Olf-1											R-R
GRB2-CRK	GR			LUN-1	STAT1	AP-4	NF-kappaB			AhR		NF-Y										R-R
GRB2-MAP2K1		STAT5A																				R-M
GRB2-MAPBPIP			PPARG	LUN-1	STAT1	AP-4	NF-kappaB	C/EBPgamma	IRF-1	AhR		NF-Y	ER	Max	HSF1				VDR			R-M
GRB2-RABGGTB		STAT5A			STAT1			C/EBPgamma	IRF-1													R-O
MAP2K1-CBLB	GR	STAT5A																				M-R
MAP2K1-CRK	GR	STAT5A																				M-R
MAP2K1-YWHAG	GR	STAT5A																				M-RM
MAP3K7-BRAF	GR		PPARG	LUN-1							Olf-1		ER		HSF1	Evi-1	CREB	RORalpha2			FOXP1	M-M
MAPK1-RPS6KA5		STAT5A				AP-4	NF-kappaB			AhR				Max			RORalpha2					M-E
RABGGTB-CBLB	GR	STAT5A						C/EBPgamma	IRF-1													O-R
RABGGTB-RPS6KA5	GR	STAT5A						C/EBPgamma	IRF-1													O-E
RPS6KA2-MGC3234			PPARG	LUN-1																		M-U
RPS6KA5-CRK	GR			LUN-1		AP-4	NF-kappaB			AhR		NF-Y										M-R
RPS6KA5-CBLB						AP-4				AhR												E-R
RPS6KA5-GRB2						AP-4		C/EBPgamma		AhR										VDR		E-R
RPS6KA5-MAP2K1	GR	STAT5A																				M-M
RPS6KA5-YWHAG	GR	STAT5A																				M-RM
SH3KBP1-LYST			PPARG																			R-M
SH3KBP1-RPS6KA2			PPARG	LUN-1																		R-M
YWHAG-CBLB	GR	STAT5A																				RM-R
YWHAG-CRK	GR	STAT5A																				RM-R
YWHAG-GRB2	GR	STAT5A																				RM-R
YWHAG-MAPBPIP	GR	STAT5A	PPARG																			RM-M
YWHAG-RABGGTB	GR	STAT5A																				RM-O
UNG/PCM	(14/24)	(13/18)	(14/12)	(8/8)	(6/7)	(5/8)	(5/6)	(5/6)	(5/5)	(7/5)	(4/4)	(4/3)	(5/3)	(5/3)	(5/3)	(3/2)	(3/2)	(4/2)	(3/3)	(2/1)	(2/1)	
Aver. distance from TSS	22564	9290	9794	8460	29001	4437	6107	9615	8952	875	13185	6975	6193	5568	4516	13749	9477	10710	6182.75	3670	16236	
SD	±40221.9	±8745.8	±7654	±8436.9	±64116.9	±8352.8	±7931	±10514.1	±10910	±1311	±4600	±9046	±5559.4	±7614.6	±6306.6	±6065.7	±4431	±7134	±8745	±4613.1	±2346.9	

Table 2. *Co*-regulated pairs (50 pairs, 40 genes) of the Ribosomal dataset with motifs conserved in the promoters of genes across three species and whose transcription factors were found to be highly *co*-expressed with genes in pairs are shown. UNG — means the Unique Number of Genes possessing Motif and PCM — number of Pairs with the respective Common Motif (single copy per pair is counted). In the last two rows average distance from TSSs and standard deviations (SD) are shown. Genes, which ID includes "L", belong to large ribosomal subunit and genes with "S" in their ID belong to small ribosomal subunit.

Co -regulated pairs of Ribosomal genes ($r \geq 0.6$) and binding sites, which transcription factors are found in a high *co* -regulation ($r \geq 0.6$) with gene in pairs

RPL13-RPL35	AIRE	Brachyury	NRSF	HEN1	p53	PAX-3	NF-Y				LUN-1	
RPL13-RPL39	AIRE	Brachyury	NRSF	HEN1	p53	PAX-3						
RPL13-RPL6	AIRE	Brachyury		HEN1			NF-Y					
RPL13-RPL8	AIRE	Brachyury	NRSF		p53	PAX-3			PPARG		LUN-1	HTF
RPL13-RPL9	AIRE	Brachyury	NRSF	HEN1	p53	PAX-3						
RPL13-RPS13	AIRE	Brachyury					NF-Y	GFI-1				
RPL15-RPL13	AIRE	Brachyury	NRSF	HEN1		PAX-3	NF-Y	GFI-1				
RPL15-RPL19	AIRE	Brachyury	NRSF	HEN1			NF-Y					
RPL15-RPL28	AIRE	Brachyury	NRSF	HEN1								
RPL15-RPL35	AIRE	Brachyury	NRSF	HEN1		PAX-3	NF-Y					
RPL15-RPL39	AIRE	Brachyury	NRSF	HEN1		PAX-3						
RPL15-RPL4	AIRE	Brachyury	NRSF	HEN1				GFI-1				
RPL15-RPL8	AIRE	Brachyury	NRSF			PAX-3						
RPL15-RPS13	AIRE	Brachyury					NF-Y	GFI-1				
RPL15-RPS27A	AIRE	Brachyury	NRSF	HEN1		PAX-3	NF-Y					
RPL19-RPL37	AIRE	Brachyury	NRSF				NF-Y					
RPL19-RPL8	AIRE	Brachyury	NRSF		p53							
RPL19-RPS15	AIRE	Brachyury	NRSF	HEN1	p53							
RPL19-RPS23	AIRE	Brachyury		HEN1	p53							
RPL19-RPS27A	AIRE	Brachyury	NRSF	HEN1			NF-Y					
RPL28-RPL37	AIRE	Brachyury	NRSF									
RPL28-RPL6	AIRE	Brachyury		HEN1								
RPL28-RPS23	AIRE	Brachyury		HEN1	p53				PPARG			
RPL30-RPL13	AIRE		NRSF	HEN1								
RPL30-RPL6	AIRE			HEN1								
RPL35-RPL37	AIRE	Brachyury	NRSF			PAX-3	NF-Y					
RPL35-RPS23	AIRE	Brachyury			p53							
RPL37-RPL39	AIRE	Brachyury	NRSF			PAX-3						
RPL39-RPL6	AIRE	Brachyury		HEN1								
RPL4-RPL6	AIRE	Brachyury		HEN1								
RPL4-RPS15	AIRE	Brachyury	NRSF	HEN1	p53			GFI-1		Arnt		
RPL4-RPS23	AIRE	Brachyury		HEN1	p53			GFI-1				
RPL4-RPS27A	AIRE	Brachyury	NRSF	HEN1	p53			GFI-1		Arnt		
RPL8_RPS27A	AIRE	Brachyury	NRSF		p53	PAX-3				Arnt		
RPL8-RPL39	AIRE	Brachyury	NRSF		p53	PAX-3				Arnt		
RPL8-RPL9	AIRE	Brachyury	NRSF		p53	PAX-3						
RPL8-RPS15	AIRE	Brachyury	NRSF		p53					Arnt		
RPL9-RPL37	AIRE	Brachyury	NRSF			PAX-3						
RPL9-RPL6	AIRE	Brachyury		HEN1								
RPL9-RPS15	AIRE	Brachyury	NRSF	HEN1	p53			GFI-1				
RPL9-RPS23	AIRE	Brachyury		HEN1	p53			GFI-1				
RPS13-RPL37	AIRE	Brachyury					NF-Y					
RPS13-RPL37	AIRE	Brachyury	NRSF			PAX-3	NF-Y					
RPS13-RPS15	AIRE	Brachyury		HEN1	p53			GFI-1				
RPS15-RPS23	AIRE	Brachyury										
RPS15-RPL37	AIRE	Brachyury	NRSF									
RPS15-UBA52	AIRE	Brachyury	NRSF									
RPS18-RPL31									PPARG		LUN-1	
RPS23-RPS24	AIRE	Brachyury		HEN1	p53			GFI-1				
UBA52-RPS27A	AIRE	Brachyury	NRSF				NF-Y					
UNG/PCM	(18/50)	(15/47)	(12/32)	(15/28)	(13/20)	(10/16)	(8/14)	(8/11)	(7/3)	(5/4)	(5/)	(2/1)
Aver. distance from TSS	4883	4451	4605	4891	1741	3288	2202	1679	1012	1404	2351	122
SD	±6991.2	±6467.2	±7056	±6729.4	±3175.2	±6173.4	±2753.5	±3460.1	±1656.9	±1381.5	±3198.7	±35.4

Although the more precise biological interpretation of the TFs found for the MAPK and ribosomal datasets still remains to be done, we believe that our results contribute to the general understanding of the regulatory mechanisms of MAPK G-protein coupling pathway related genes in the comparison with the housekeeping ribosomal dataset. In this study we used *co*-regulated pairs without distinguishing between *up*-regulated and *down*-regulated pairs and this is a subject for further analysis. One another purpose of this work was to investigate the relationship between

Table 3. 19 transcription factors (TFs) found *co-regulated* ($r \geq 0.6$) with genes in pairs of MAPK dataset and whose motifs were found to have significantly high scores in regulatory regions of the respective genes. Underlined TFs were found in the ribosomal dataset also. Alphabetic symbols "M" and "B" in the columns "FUNCTION" shows the attachment of the motif in the first round of selection to MAPK unique TFs group (M) or to group of TFs shared by both datasets (B), and the MCSs for the latter case are also shown. Asterisks "*" near TFs in Table 1 means that TFs are supposed somehow to be related to the MAPK pathway, according to the published PubMed resources.

TF	UNG	Function MAPK dataset
GR	14	Glucocorticoid receptor. (B: 5.5)
STAT5A*	13	Signal transducer and activator of transcription 5A. Phosphorylated by the receptor associated kinases. (B:4.2)
PRARG	14	Peroxisome proliferator-activated receptor. (B:1.2)
LUN-1	8	Topoisomerase I binding, argininie/serine-rich. (B: 0.42)
STAT1*	6	Signal transducer and activator of transcription 1, phosphorylated by the receptor associated kinases. (M)
AP-4	5	Transcription factor AP-4 (activating enhancer binding protein 4). (M)
NF-kappaB*	5	Nuclear factor KB, activated by serine kinases. (M)
C/ERBgamma	5	CCAAT/enhancer binding protein (C/EBP), gamma. (M)
IRF-1*	5	Interferon regulatory factor 1. Responsible for the induced coexpression of NF-kappa B activator 1 (Act1). (M)
AHR	4	Aryl hydrocarbon receptor. MEK1 facilitates transcriptional activation while targeting the AHR for degradation. (B:1.71)
OLF-1	4	Early B cell factor (B: 3)
NF-Y	4	CCAAT-binding transcription factor. (B: 188)
ER	5	Estrogen receptor 1. Nuclear IRS-1 interacts with ER and induces its transcriptional activity. (M)
MAX*	5	MYC associated factor X. (M)
HSF1	5	Heat shock transcription factor 1. (M)
Evi-1	3	Ecotropic viral integration site 1. (M)
CREB*	3	cAMP responsive element binding protein. Phosphorylated by several protein kinases. (B: −0.85)
RORalpha2	4	RAR-related orphan receptor A, NRI subfamily of nuclear hormone receptors. (M)
VDR*	3	Nuclear hormone receptor for vitamin D3. Enhance STAT1-mediated transcription. Involved in p38 MAPK pathway (B:2.2).

Table 4. Transcription factors (TFs) found *co-regulated* ($r \geq 0.6$) with genes in pairs of the ribosomal dataset and whose motifs were found to have significantly high scores in regulatory regions of the respective genes. Underlined TFs were also found in the MAPK dataset. "B" indicates TFs found in both datasets in the first round of selection and the number next to it shows the MCS with respect to the MAPK dataset.

TF	UNG	Function Ribosomal dataset
AIER	18	Autoimmune regulator with unknown function (B: −0.05)
Brachyury	15	Embryonic nuclear transcription factor (B:0.27)
NFSF	12	RE-1 silencing transcription factor (B: −0.37)
HENI	15	Putative TF. Known to be involved in turmogenesis. (B: 0.91)
P53	13	Turmor protein p53. (B: −1.05)
PAX-3	10	Paired box gene 3. (B: −0.5)
NF-Y	8	CCAAT-binding transcription factor. (B: 1.88)
GFI-1	8	Growth factor independent (B: 6.21)
PPARG	6	Peroxisome proliferator-activated receptor. (B: 0.41)
ARNT	5	Aryl hydrocarbon receptor nuclear translocator. (B: 1.79)
LUN-1	5	Topoisomerase binding arginine/serine rich protein (B: 0.41)
HTF	2	Hepatocarcinogenesis-related transcription factor (B: 1.83)

co-regulation and protein-protein interactions. As we found only three highly *co*-regulated protein interacting pairs, we have to conclude that in the case of the MAPK G-protein coupling pathway, the transcriptional co-regulation of interacting proteins is not readily apparent.

References

1. Chang L, Kartin M, Mammalian MAP kinase signaling cascades, *Nature* **410**:37–40, 2001.
2. Hunter T, Signaling — 2000 and Beyond, *Cell* **100**:113–127, 2000.
3. Brivanlou A, Darnell J, Signal transduction and the control of gene expression, *Science* **295**:813–818, 2002.
4. Suzuki Y, Yamashita R, Shirota M, *et al.*, Sequence comparison of human and mouse genes reveals a homologous block structure in the promoter regions, *Genome Res* **14**:1711–1718, 2004.
5. Xie X, Lu J, Kulbokas EJ, *et al.*, Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals, *Nature* **434**:338–345, 2005.
6. Serizawa S, Miyamichi K, Nakatani H, *et al.*, Negative feedback regulation ensures the one receptor — One olfactory Neuron rule in mouse, *Science* **19**:2088–2094, 2003.
7. Muller M, Hagstrom K, Gyurkovics H, *et al.*, The Mcp element from the Drosophila bithoral complex mediates long-distance Regulatory Interactions, *Genetics* **153**:1333–1356, 1999.

8. Polouliakh N, Takagi T, Nakai K, MELINA: Motif extraction from promoter regions of potentially co-regulated genes, *Bioinformatics* **19**(3):423–424, 2003.
9. Corcoran DL, Feingold E, Benos PV, FOOTER: A web tool for finding mammalian DNA regulatory regions using phylogenetic footprinting, *Nucleic. Acids Res* **33**: W442–W446, 2005.
10. Blake JA, Richardson JE, Bult CJ *et al.*, MGD: The mouse genome database, *Nucleic Acids Res* **31**:193–195, 2003.
11. Suzuki Y, Yamashita R, Nakai K, Sugano S, DBTSS: DataBase of human transcriptional start sites and full-length cDNAs, *Nucleic Acids Res* **30**:328–331, 2002.
12. Pruitt K, Tatusova T, Maglott D, NCBI reference sequence (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins, *Nucleic Acids Res* **33**:D501–D504, 2005.
13. Barret T, Suzek TO, Troup DB *et al.*, NCBI GEO: Mining millions of expression profiles — database and tools, *Nucleic Acids Res* **33**:D562–D566, 2005.
14. Sandberg R, Ernberg I, The molecular portrait of in vitro growth by meta-analysis of gene expression profiles, *Genome Biol* **6**(8):R65, 2005.
15. Wingender E, Dietze P, Karas H, Knüppel R, TRANSFAC®: A database on transcription factors and their DNA binding sites, *Nucleic Acids Res* **24**:238–241, 1996.

Natalia Polouliakh received her Ph.D. degree at the Faculty of Science from the State Ochanomizu University of Tokyo in 2003. She was working on the problem of evaluation of the performance of motif discovery tools for several years, and her current interests are comparative genomic analysis of regulatory regions in higher eukaryotes. Currently, she is working as a postdoctoral fellow in Computational Biology Research Center, Tokyo.

Tohru Natsume received his Ph.D. at the faculty of Medical Science from the Kyoto University in 1993. He has been leading the Protein Network Team from 2001 at JBRC and a visiting professor at Department of Molecular Biology, Institute of Molecular and Cellular Biosciences of the University of Tokyo and at the Proteomic Center of the Medical Institute of Bio-regulation of the Kyushu University. His research interests include large-scale analyses of protein expression patterns and protein interactions and development of high-speed nanoscale proteomic analysis system.

Hajime Harada received his Bachelor of Engineering in Computer Science, from Tokyo University of Agriculture and Technology, Japan, and his M.S. degree in Integrative Environmental Sciences, from University of Tsukuba, Japan, in 2000 and 2002, respectively. From April 2005, he is working as a technical staff in Computational Biology Research Center, Tokyo.

Wataru Fujibuchi received his Ph.D. degree at the Department of Biophysics from Kyoto University in 1998. From 1999–2002 he worked as an invited researcher at the NCBI, USA. Now he is hired as a Research Scientist of National Institute of Advanced Industrial Science and has a current position of Visiting Associate Professor, at the Research Institute of IT-Bio, Waseda University. He is an author of Cell Montage database. His research interests include sequence analysis of promoter functions, microarray data analysis, prediction of genetic, networks from microarray, integrative analysis of cell features.

Paul Horton received a bachelor degree in Molecular Biology from the University of Washington, a master degree in Biophysics from Kyoto University, and a Ph.D. (1997) in Computer Science from UC Berkeley. He joined CBRC in 2003, where he leads the sequence analysis team. His research projects include developing the WoLF PSORT protein localization prediction tool, algorithms for motif discovery in genomic sequences, and algorithms for rapid profile search from microarray databases.

Developmental Cell 11, 69–79, July, 2006 ©2006 Elsevier Inc. DOI 10.1016/j.devcel.2006.04.019

XGAP, an ArfGAP, Is Required for Polarized Localization of PAR Proteins and Cell Polarity in *Xenopus* Gastrulation

Junko Hyodo-Miura,[1] Takamasa S. Yamamoto,[1] Akiko C. Hyodo,[1] Shun-Ichiro Iemura,[2] Morioh Kusakabe,[3] Eisuke Nishida,[3] Tohru Natsume,[2] and Naoto Ueno[1,4,*]

[1] Department of Developmental Biology
National Institute for Basic Biology
Okazaki 444-8585
Japan

[2] Biological Information Research Center
National Institute of Advanced Industrial Science and Technology
Tokyo 135-0064
Japan

[3] Department of Cell and Developmental Biology
Graduate School of Biostudies
Kyoto University
Sakyo-ku, Kyoto 606-8502
Japan

[4] Department of Molecular Biomechanics
The Graduate University for Advanced Studies
Okazaki 444-8585
Japan

Summary

To dissect the molecular mechanisms underlying convergent extension (CE), a prominent set of cell movements during *Xenopus* gastrulation, we performed a functional expression screen and identified a GTPase-activating protein for ADP ribosylation factors (ArfGAP), which we termed XGAP. We demonstrated that XGAP is required to confine or restrict the cellular protrusive activity to the mediolateral ends of cells, where XGAP is normally localized, and therefore for the proper intercalation of cells participating in CE. We also demonstrated that a C-terminal conserved domain of XGAP, but not its GAP activity, is required and sufficient for this intracellular localization and function. We further showed that XGAP physically interacts with the known polarity proteins 14-3-3ε, aPKC, and PAR-6 and directs them to the mediolateral ends of dorsal mesoderm cells during gastrulation. We propose that XGAP controls CE through the restriction and maintenance of partitioning-defective (PAR) proteins in the regions that harbor protrusive activity.

Introduction

Xenopus gastrulation, which involves the dynamic rearrangement of the three germ layers, occurs primarily in the dorsal embryo and results from the orchestration of various cell movements. Recently, convergent extension (CE), one of the cell-movement processes in gastrulation, has drawn considerable attention (Keller, 2002; Wallingford et al., 2002). CE consists of two major cellular events: polarization and intercalation. The polarization of the dorsal mesodermal cells, which align and intercalate mediolaterally, is organized according to the intrinsic axis polarity of the embryo (Ninomiya et al., 2004). The intercalating cells are driven by mediolaterally polarized protrusive activity; they appear as spindle-shaped cells (Keller, 2002). These movements narrow (converge) and elongate (extend) the tissues along the body axis, eventually pushing the head away from the tail, and produce a driving force for the gastrulation movements (Keller, 2002). These movements appear to be driven by internal forces independent of other tissues and external substrates, given that Keller explants dissected from the dorsal region, including the presumptive notochordal, somatic, and neural tissues, can converge and extend in a culture dish (Keller and Danilchik, 1988).

The signaling pathway regulating gastrulation in zebrafish and *Xenopus* is thought to be similar to that establishing planar cell polarity (PCP) in the *Drosophila* wing, ommatidia, and notal bristles, called the Wnt/PCP pathway (Heisenberg et al., 2000; Mlodzik, 2002; Tada and Smith, 2000; Ueno and Greene, 2003). In *Xenopus*, the loss of function of PCP components results in gastrulation defects as a consequence of the aberrant mediolateral alignment of dorsal mesodermal cells and the resulting insufficient intercalation (Kuhl, 2002; Takeuchi et al., 2003; Ueno and Greene, 2003; Wallingford et al., 2000; Yamanaka et al., 2002; Ohkawara et al., 2003). Although these reports show that the Wnt/PCP pathway plays a major role in the regulation of CE and identified the molecules essential for CE, it remains to be learned when the polarization of dorsal mesoderm cells is initiated and how the cell polarity is actually established by those molecules.

Recently, another clue as to the cell polarity during gastrulation was obtained from functional analyses of polarity proteins such as the partitioning defective (PAR) proteins; like PCP, PAR proteins are required for gastrulation in *Xenopus* (Kusakabe and Nishida, 2004). These proteins, including aPKC, PAR-3, PAR-6, 14-3-3/PAR-5, and PAR-1, constitute an essential system for regulating a variety of cellular processes related to cell polarity, including asymmetric cell division in the fertilized *C. elegans* egg and *Drosophila* neuroblast (Ohno, 2001; Pellettieri and Seydoux, 2002; Betschinger and Knoblich, 2004), the apical-basal polarity of epithelial cells (Macara, 2004; Suzuki et al., 2004), and neuronal polarity (Nishimura et al., 2004). In *Xenopus*, the polarity-inducing kinase PAR-1 controls gastrulation in cooperation with 14-3-3/PAR-5 and aPKC (Kusakabe and Nishida, 2004). It is possible that a similar system is used to establish the bipolarity of the dorsal mesodermal cells that participate in gastrulation cell movements in *Xenopus*, although it is still unclear whether these proteins are distributed in a polarized manner in the cells during gastrulation, and the mechanism by which cell polarity is triggered and established prior to intercalation and the coordination of cell movements has remained to be investigated.

To address the mechanisms regulating gastrulation, we performed a functional screen and identified

*Correspondence: nueno@nibb.ac.jp

a gene encoding an ArfGAP, which we called XGAP, and found that it controls the morphological cell movements of gastrulation. Our data suggest that polarization along the mediolateral axis in gastrulation cell movements employs a regulatory system involving PAR proteins that is common to the apical/basal polarization in epithelial cells and that this XGAP function is a prerequisite for PAR proteins to function in establishing cell polarity.

Results

Isolation of *XL221p20* as a Gene Essential for Gastrulation

We first constructed a cDNA library with mRNAs from elongating DMZ explants (Keller explants; see Experimental Procedures). Using this library, we performed a functional screen to identify candidates for gastrulation-regulating genes. The mRNAs synthesized from pooled clones were injected into the two dorsal blastomeres of four-cell-stage embryos. The mRNA pools that caused a shortened AP axis, *spina bifida*, and/or failure to elongate in Keller explants were considered positive and rescreened to isolate a single clone. After the screening of 1,535 cDNAs, we found 105 genes, approximately 6.8% of all the screened clones, were positive by the above criteria.

XL221p20 was one candidate whose overexpression caused gastrulation defects. In the screening, 1 ng of *XL221p20* mRNA caused a short trunk or *spina bifida* in approximately 27% of the injected embryos (Figures 1A and 1B), while it had little effect on the elongation of Keller explants (see below).

To test whether the overexpression of *XL221p20* affected mesodermal differentiation, we examined the expression of *Xbra* and *Xwnt11* as mesodermal marker genes and *goosecoid* (*gsc*) as a dorsal marker gene by whole-mount in situ hybridization (WISH). The expression of all three marker genes was unaffected by the overexpression of *XL221p20* mRNA, even though the embryo exhibited delayed involution (Figure 1C, upper panels versus lower). This result indicated that the gastrulation defect caused by *XL221p20* overexpression was not a secondary consequence of the inhibition of mesodermal differentiation.

The spatial expression pattern of *XL221p20* during embryogenesis by WISH showed that at the onset of gastrulation, *XL221p20* transcripts were dorsally enriched (Figures 1Da versus 1Dg). In the hemisection at this stage, *XL221p20* mRNA was detected in the involuting marginal zone and in the ectoderm adjacent to the mesodermal layer (Figures 1Dd and 1De). At the neurula stage, *XL221p20* was expressed intensely at the dorsal midline (Figures 1Db and 1Dc versus 1Dh and 1Di, respectively), in the dorsal half of the notochord, where cell intercalation takes place during gastrulation (Figure 1Df). These patterns indicated that *XL221p20* is expressed where dynamic cell movements occur during gastrulation.

XL221p20 Encodes a GAP Belonging to the ArfGAP3 Family

The deduced primary structure of *XL221p20* revealed that it encodes a GTPase-activating protein (GAP) composed of 524 amino acids; thus, we named it XGAP.

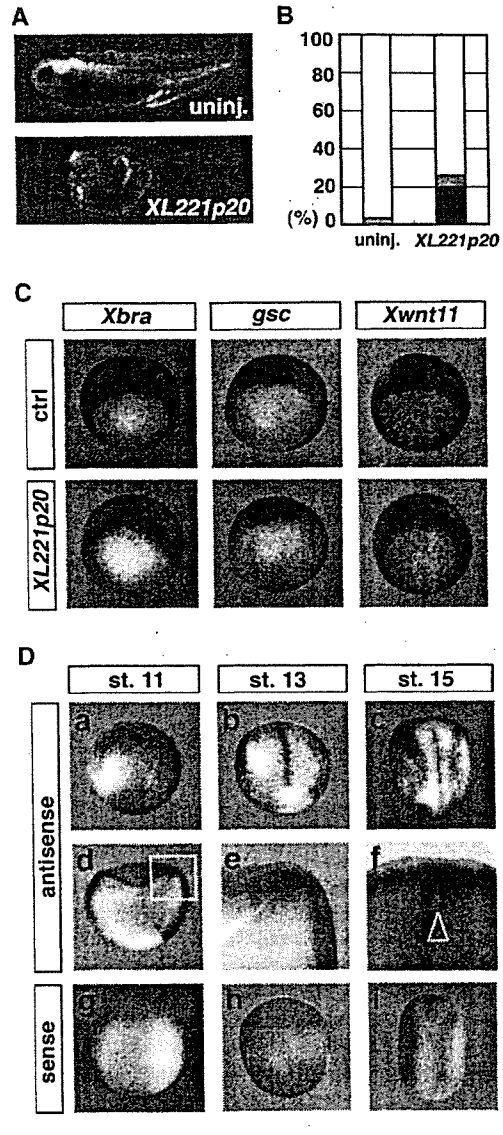

Figure 1. Identification of *XL221p20* as a Molecule that Plays a Role in Gastrulation, by Expression Cloning

(A and B) Overexpression of *XL221p20* causes gastrulation defects. The percentages of normal (open), short trunk (shaded), and *spina bifida* (solid) embryos are indicated in (B) (n = 60).

(C) Mesodermal specification is unaffected by *XL221p20* overexpression. The expression of *goosecoid* (gsc), *Xenopus brachyury* (*Xbra*), and *Xenopus Wnt11* (*Xwnt11*) was detected by WISH at stage 11 (dorsal-vegetal view). Both control (ctrl; upper panels) and *XL221p20*-overexpressing (lower panels) embryos were given injections of β-gal as a lineage tracer (red).

(D) *XL221p20* is expressed in the axial mesoderm and neuroectoderm in gastrulae. The *XL221p20*-expressing region was detected by WISH at stage 11 (Da, Dd, De, and Dg), stage 13 (Db and Dh), and stage 15 (Dc, Df, and Di). An enlarged view of the hemisection image ([Dd], square) is shown in (De). The triangle in (Df) points to the notochord. No signal was detected by the sense probe (Dg–Di).

Because *XL221p20* included the full-length XGAP cDNA, its overexpression was thought to cause a gain of XGAP function in the screen (Figures S1A and S1B). XGAP is closely related to the GAPs for a small G protein family, the ADP-ribosylation factors (ARFs) (Casanova, 2003;

Randazzo et al., 2000). Unlike several well-known ArfGAPs, including GIT, ACAP, and ASAP, all of which possess other functional domains, such as PH and SH3 domains (Randazzo and Hirsch, 2004), the only known domain we found in XGAP was a GAP catalytic domain, which was most similar to that of ArfGAP3 (Liu et al., 2001) (ARG3; Figure S1A). The predicted amino acid sequences of XGAP and xARG3 have 96% identity in the GAP domain (amino acid residues 1–126) and 79% identity over the entire coding region. In addition to the GAP domain, we found another highly conserved domain (amino acid residues 410–524) with no evident homology with known motifs, in the C terminus of XGAP and the ARG3 from *Xenopus*, human, mouse, and zebrafish (Figures S1A and S1B).

To determine whether XGAP and xARG3 are coregulated pseudoalleles and functionally equivalent, we performed reverse transcription PCR (RT-PCR) to examine their temporal expression patterns with specific primer sets for each. XGAP was maternally provided and expressed throughout early embryogenesis, whereas xARG3 expression began later in embryogenesis, after gastrulation was initiated (Figure S1C). This temporal difference indicated that XGAP and xARG3 are independently regulated.

XGAP-Mo Perturbs Gastrulation Cell Movements without Affecting Cell Fates

To investigate the in vivo function of XGAP, we attempted to deplete XGAP with an antisense morpholino oligonucleotide (XGAP-Mo) specific to XGAP. First, we confirmed that XGAP-Mo inhibited the translation of XGAP mRNA specifically. XGAP-Mo, but not the control Mo (Ctrl Mo), blocked the protein production of a Venus-tagged XGAP construct, including the original 5′UTR (utrXGAP-vn). In contrast, the translation of resXGAP, a rescuing construct that had several mutations in the target site of XGAP-Mo, was not affected (Figure 2A).

To inhibit XGAP synthesis, the Mo was injected into the two dorsal blastomeres of a four-cell-stage embryo. It did not affect involution at the onset of gastrulation but delayed epiboly and closure of the blastopore (data not shown). WISH revealed that the dorsal mesodermal region expressing *chd* expanded laterally, the movement of the prechordal plate marked by *gsc*-positive cells was attenuated and shifted posteriorly, and the notochord marker *Xnot* was split into two domains at the end of gastrulation (Figure S2). Consequently, XGAP-Mo caused *spina bifida* and short-trunk embryos in a dose-dependent manner (Figures 2B and 2C). These gastrulation-defective phenotypes could be partially rescued by the coexpression of resXGAP (Figures 2B and 2C and Figure S4). This result suggested that XGAP is required for normal gastrulation.

To confirm that XGAP-Mo did not change the cell fates, we examined the expression of several marker genes in DMZ explants by RT-PCR. The expression of mesodermal markers (*Xbra*, *Xwnt11*, and *myf5*), dorsal mesodermal markers (*gsc*, *chd*), an anterior neural marker (*otx2*), and a ventral marker (*bmp4*) was not affected by XGAP-Mo compared with the Ctrl Mo (Figure 2D and data not shown). We also examined the subsequent mesodermal tissue differentiation, by immunohistochemistry. Despite the change in overall morphology, the notochord and somites differentiated normally in the XGAP-Mo-injected embryos (Figure 2E). These data indicated that XGAP is essential for the control of gastrulation, particularly for cell movements rather than for cell specification.

XGAP-Mo Inhibits the Intercalation of Dorsal Marginal Cells

Many molecules with a role in gastrulation cell movements cause defects in whole embryos and impair the elongation of Keller explants by both the gain and loss of their function. However, in our initial screen, a gain of XGAP function did not seem to appreciably affect the elongation of Keller explants. To examine the effect of XGAP mRNA and XGAP-Mo on the elongation of Keller explants in more detail, we carefully observed the elongating explants over time. The elongation speed of explants from XGAP-Mo-injected embryos was severely retarded, and as a result, their final average length/width ratio was approximately 50% that of uninjected controls (Figures 2F and 2G). Interestingly, the overexpression of XGAP slightly increased the initial rate of elongation (Figures 2F and 2G). In retrospect, we think that the accelerated intercalation caused by the overexpression disrupted the normal gastrulation process in the initial screen (Figure 1A). The results from these experiments suggested that XGAP has an important role in the efficiency of tissue elongation along the AP axis.

We further examined the effect of XGAP-Mo at the cellular level. To observe the intercalating cells in Keller explants, one of two distinguishable lineage tracers, Venus and RFP, which carried sequences that rendered them membrane bound, was injected into each of the two dorsal blastomeres of four-cell-stage embryos. The progress of intercalation was calculated by the proportion of cells positive for one fluorescent marker that were surrounded by cells with the other marker, which indicates the extent of the cell mixture of two laterally originated group of cells after intercalation. In the control embryos, the spindle-shaped morphology and vigorous intercalation of dorsal mesodermal cells were observed (Figure 2H; 43.5% ± 12.3%). The XGAP mRNA-injected side of the explants also showed cell intercalation, to a similar extent as the control side (data not shown; 40.7% ± 18.9%). However, the XGAP-Mo-injected side showed impaired cell intercalation and a disruption of the isodiametric cell morphology (Figure 2H; 21.3% ± 9.4%). The length/width ratio of the XGAP-Mo-injected cells was approximately 50% that of the control cells (Figure 2I). These results strongly suggested that XGAP is required for normal cell intercalation because it enables cells to form the spindle-shaped morphology, which is one of the main events associated with the CE movements in the DMZ.

XGAP Is Required to Confine the Protrusions to the Two Tips of the Polarized Cells in the DMZ

It is known that the mediolateral intercalation of mesodermal cells is driven by mediolaterally polarized protrusive activity and that the Wnt/PCP pathway regulates cortical actin dynamics (Wallingford et al., 2000; Iioka et al., 2004). To test whether XGAP affects the induction of the protrusions before the orientation, XGAP mRNA or XGAP-Mo was injected into the animal pole with Xwnt11

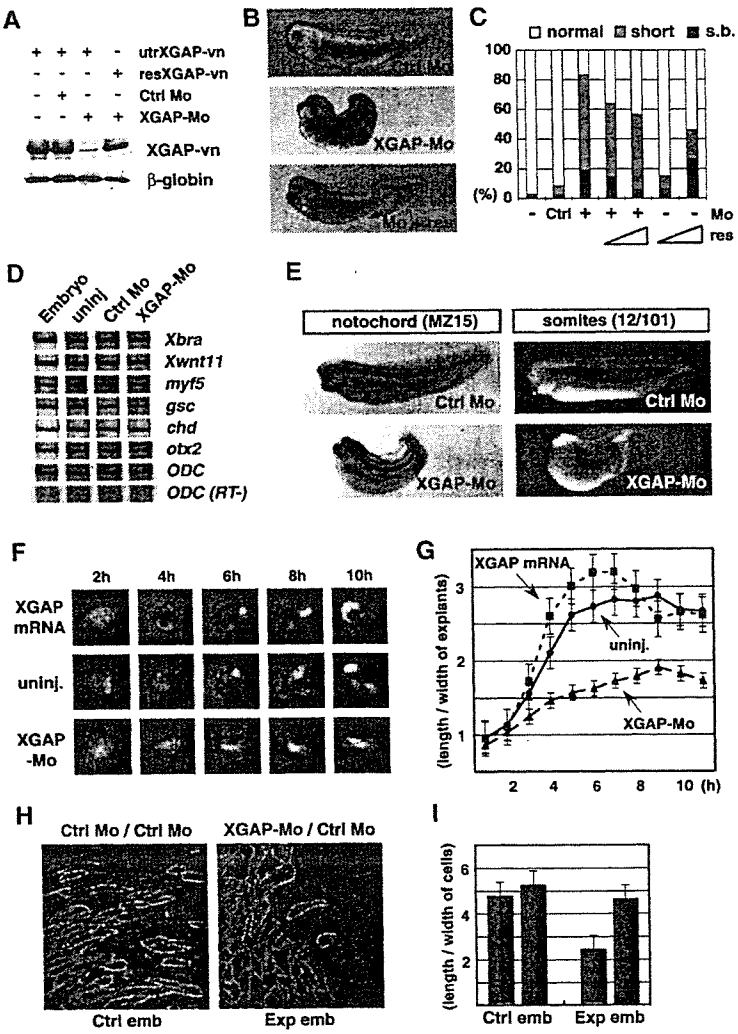

Figure 2. XGAP Is Required for the Intercalation of Dorsal Mesoderm Cells Undergoing Gastrulation

(A) A morpholino-antisense oligonucleotide (Mo) for XGAP specifically inhibits the translation of XGAP mRNA. XGAP mRNA, constructed to generate a fusion protein with a variant of YFP (Venus, vn) (Nagai et al., 2002), was injected with or without each Mo into the animal pole of two-cell-stage embryos. XGAP-vn protein was detected by Western blotting with an anti-GFP polyclonal antibody. Flag-tagged β-globin was detected as the loading control.

(B and C) XGAP-Mo causes gastrulation defects. The external appearance was examined at stage 35 ([B] and [C]; n = 120, gray and black indicate a short trunk and *spina bifida*, respectively). The gastrulation-defective phenotypes caused by the Mo were partially rescued by the coinjection of resXGAP (100 pg, 200 pg) in a dose-dependent fashion (C). In the absence of XGAP-Mo, resXGAP manifested the effects of XGAP overexpression, consistent with our initial observation (C).

(D) Cell fates are unaffected by XGAP-Mo. Whole embryos or DMZ explants that had received injections of Ctrl Mo or XGAP-Mo were harvested at the midgastrula stage and then analyzed by RT-PCR with primer sets for *Xbra*, *Xwnt11*, *myf5*, *gsc*, *chd*, *otx2*, and *ODC*.

(E) Mesodermal differentiation is unaffected by XGAP-Mo. Each Ctrl Mo- or XGAP-Mo-injected embryo was harvested at stage 30, and the notochord and somites were immunostained with the MZ15 and 12/101 antibodies, respectively.

(F and G) XGAP-Mo attenuates the elongation of Keller explants. Keller explants from embryos given injections of XGAP mRNA (upper panels in [F], squares in [G]) or XGAP-Mo (lower panels in [F], triangles in [G]) or that did not receive the injections (uninj; the second row in [F], circles in [G]) were excised at stage 10.5, and the explants were cultured and their elongation monitored for 12 hr. Photographs were taken every 2 hr (F), and the length and width were then measured to determine the time course of elongation ([G], n = 20). The error bars represent standard error.

(H) XGAP-Mo inhibits the intercalation in the dorsal mesoderm. XGAP-Mo or Ctrl Mo was coinjected with Venus with a membrane bound sequence (mb-Venus) into one of the two dorsal blastomeres at the four-cell stage. Ctrl Mo and RFP with a membrane bound sequence (mb-RFP) were coinjected into the other dorsal blastomere. We repeated this experiment more than three times.

(I) XGAP-Mo disrupts the isodiametric cell morphology. Each cell in the intercalation experiment (H) was measured for length along the ML axis and width along the AP axis, and the length/width ratio was calculated (n = 25). The error bars represent standard error.

and Xfz7 mRNA. However, neither XGAP mRNA nor XGAP-Mo affected the protrusive formation induced by Xwnt11 and Xfz7 in the animal cap cells (Figure S3).

Next, to examine the protrusive behavior of the cells, membrane bound RFP was injected into one of the two dorsal blastomeres and observed by time-lapse recording. In the controls, the labeled cells were polarized by active protrusions that repeated extension and withdrawal at the mediolateral ends during time-lapse recording (Figure 3A). In contrast, the polarity of the active protrusions was disrupted in the experimental cells into which XGAP-Mo with Venus had been injected, even though the surrounding normal cells made the labeled cells elongate mediolaterally by wedging between them (Figure 3A). This result showed that the polarization of the protrusions is cell autonomous. The number of active protrusions in the mediolateral (ML) and anteroposterior (AL) regions was then counted to assess the orientation of the protrusions (Figure 3B). The orientation of the active protrusions was random in the XGAP-Mo-injected explants, while 80% of the protrusions were in the mediolateral region in the controls (Figure 3C). The defect in orientation caused by XGAP-Mo was suppressed by the coinjection of resXGAP (Figures 3A and 3C). These data suggested that XGAP is required not for the formation of active protrusions but for confining them to the mediolateral sides of cells during intercalation.

To investigate the subcellular localization of XGAP in the intercalating cells, we next expressed Venus-fused XGAP (vn-XGAP) to target the DMZ cells. vn-XGAP partially rescued the gastrulation-defective phenotype caused by XGAP-Mo (Figure S4) and therefore was a functional construct. Interestingly, this rescuing

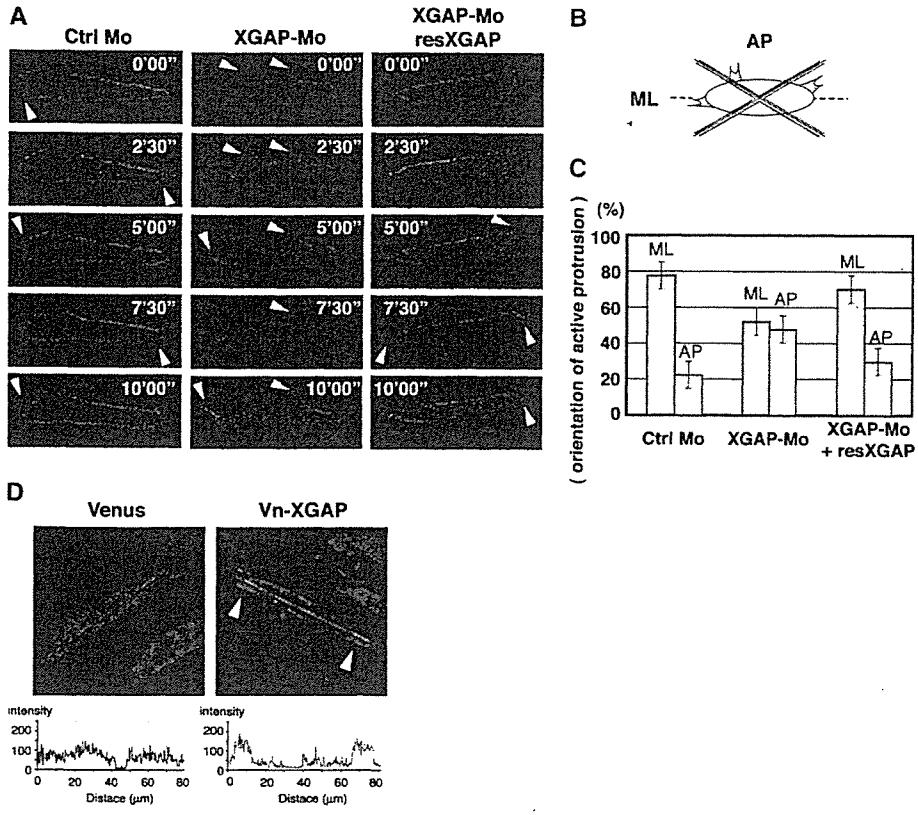

Figure 3. XGAP Confines the Active Protrusions to the Tips of the Spindle-Shaped Cells

(A) XGAP-Mo disrupts the orientation of the active protrusions. mb-RFP was injected into one dorsal blastomere of 4-cell-stage embryos with Ctrl Mo (left panels), XGAP-Mo (middle panels), or XGAP-Mo and resXGAP (right panels). The active protrusions (arrowheads) of the mb-RFP expressing cells, which were surrounded by normal (unlabeled) cells, were observed by time-lapse recording by using confocal microscopy.

(B and C) XGAP is required to confine the active protrusions to the mediolateral sides of the cells. The positions of the active protrusions were determined, and the numbers in the mediolateral (ML; blue) or anteroposterior (AP; orange) region were counted. The active protrusions were mediolaterally biased in the control cells (77.6% ML and 22.4% AP; 27 cells/7 explants) but were randomly distributed in the XGAP-Mo-injected cells (52.2% ML and 47.8% AP; 48 cells/10 explants). This randomization was reversed by the coinjection of resXGAP mRNA (70.1% ML and 29.9% AP; 28 cells/11 explants). Differences were statistically significant ($p < 0.01$) by ANOVA. The error bars represent standard error.

(D) vn-XGAP accumulates at the tips of spindle-shaped cells in the DMZ. DMZ explants from embryos that had been given an injection of a low concentration of Venus or vn-XGAP mRNA, which caused no phenotype, were dissected at stage 10.5 and observed by confocal microscopy. The arrowheads point to both tips of a spindle-shaped cell. The lower panels show the intensity of Venus. The magenta bar indicates the x axis and scanning was from left to right.

vn-XGAP protein was detected in the perimembrane region and very often accumulated at the tips of spindle-shaped cells undergoing intercalation (Figure 3D). Moreover, this localization of vn-XGAP was specific to the DMZ, and it was seen neither in the VMZ nor in animal caps (data not shown). These data indicated that XGAP is localized to bipolar positions in the dorsal mesodermal cells and acts to confine the active protrusions to both ends.

The C Terminus of XGAP Is Necessary and Sufficient for the Regulation of Gastrulation Movements

We next asked whether GAP activity is essential for the gastrulation cell movements. To examine the importance of the GAP domain in rescuing the Mo phenotype, we constructed several different XGAP deletion mutants (Figure 4A). Surprisingly, the GAP domain-deleted construct (XGAP-ΔN; 127–524AA), which lacks GAP activity (described in Discussion), reversed the Mo-induced phenotype in whole embryos in a dose-dependent manner (Figure 4C). This result indicated that the GAP activity of XGAP is not essential for accelerating the gastrulation cell movements. In contrast, the C-terminally deleted mutants (XGAP-N; 1–126AA and XGAP-ΔC; 1–408AA) failed to rescue the Mo phenotype (Figures 4B and 4D). In addition, the C-terminal domain of XGAP (XGAP-C; 409–524AA) alone could rescue the Mo phenotype (Figure 4E). These findings demonstrated that the highly conserved C-terminal region of XGAP is required and sufficient for XGAP's function in gastrulation.

To investigate the relationship between the function of XGAP in gastrulation and its subcellular localization, we used a Venus-tagged version of each construct. The constructs that could restore the Mo phenotype, XGAP-ΔN and XGAP-C, accumulated at the tips of spindle-shaped cells like wild-type XGAP (Figures 4G and 4I), while the localization of the C-terminal-deleted mutants, XGAP-N and XGAP-ΔC, was diffuse (Figures 4F

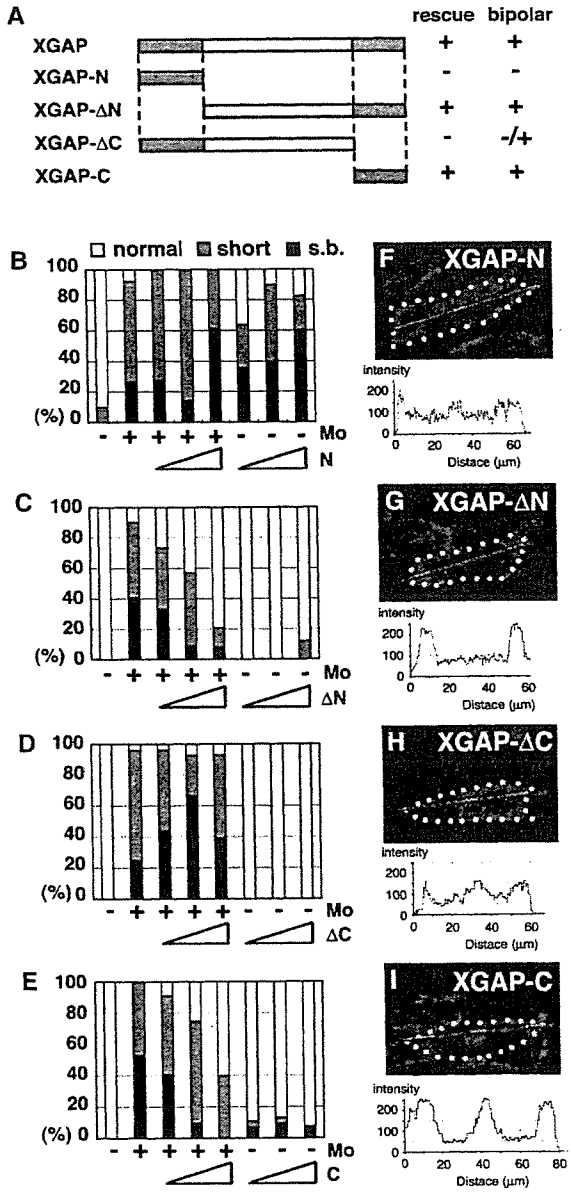

Figure 4. The Conserved C-Terminal Region, but Not the GAP Domain, Is Essential for XGAP to Control Gastrulation Cell Movements

(A) Truncated constructs of XGAP and a summary of their Mo phenotype-suppressing activity and intracellular localization. The orange box and green box indicate the GAP domain and conserved region, respectively.

(B–E) Suppression of the XGAP-Mo phenotype by coinjection of each truncated mRNA. XGAP-ΔN and XGAP-C, but not XGAP-N and XGAP-ΔC, suppressed the gastrulation-defective phenotype caused by XGAP-Mo, in a dose-dependent manner (100 pg, 200 pg, or 500 pg; n = 70). XGAP-ΔC appeared to be a null allele.

(F–I) Intracellular localization of each truncated construct of XGAP in the DMZ. The mRNA for each Venus-tagged construct was injected into the dorsal blastomeres at the four-cell stage at a low dose (100 pg) that induced no phenotype. XGAP-ΔN and XGAP-C, but not XGAP-N or XGAP-ΔC, accumulated at the tips of the spindle-shaped cells. XGAP-C was also detected in the nucleus. White dots indicate the cell shape. The lower panels show the intensity of Venus. The magenta bar indicates the x axis and scanning was from left to right.

and 4H). Together, these data indicated that the C-terminal region of XGAP is important for both its bipolar subcellular localization and its function in cell intercalation.

Extra-GAP Domain of XGAP Interacts with PAR Protein

Our next interest was in how XGAP functions to establish cell polarity in the absence of GAP activity. To address this question, we sought to identify molecules that physically interact with XGAP. We screened for XGAP-interacting proteins by pull-down assay followed by mass-spectrometry analysis with XGAP's closest counterpart, human ARG3 (hARG3). Flag-tagged hARG3 or hARG3 lacking the GAP domain (hARG3-ΔN) were expressed in HEK293T cells and pulled down by an anti-Flag antibody. In this screen, 14-3-3ε that is related to PAR-5 was found to interact with both proteins, which suggested that 14-3-3ε binds to hARG3 through the extra-GAP domain. We also confirmed this interaction with GST-tagged XGAP and Flag-tagged *Xenopus* 14-3-3ε (Figure 5A).

To examine the functional consequence of the interaction between XGAP and 14-3-3ε, we observed the intracellular localization of the RFP-tagged 14-3-3ε in *Xenopus* embryos. We observed that 14-3-3ε accumulated weakly at the perimembrane of the spindle-shaped cells in the DMZ, while it was uniformly distributed within animal cap cells (Figure 5B and data not shown). Next, Venus-tagged XGAP was coexpressed with RFP-14-3-3ε to investigate the effect of XGAP on its distribution. Consistently, either XGAP or XGAP-ΔN, but not XGAP-ΔC, colocalized with 14-3-3ε and enhanced the polarized localization of 14-3-3ε (Figure 5B). This result suggests that XGAP functions to confine 14-3-3ε to the mediolateral ends of the cells through XGAP's C-terminal region.

aPKC Phosphorylates XGAP and Enhances Its Binding to 14-3-3

According to recent reports, the partitioning defective (PAR) proteins, which include 14-3-3/PAR-5, PAR-6, aPKC, and PAR-1, are required for the gastrulation movements (Kusakabe and Nishida, 2004; Ossipova et al., 2005). The relationship between XGAP and 14-3-3ε raised the possibility that the action of XGAP correlates closely with that of the PAR proteins in the establishment of polarity in the cells undergoing gastrulation. To investigate whether XGAP is involved in the PAR protein functions, we first performed immunoprecipitation assays to test for physical interactions between the XGAP and PAR proteins. We found that xPAR-6 or xPKCλ, but not xPAR-1 or rPAR-3, was pulled down with XGAP (Figure 6A and data not shown), indicating that XGAP binds at least to the xPAR-6 and xPKCλ.

To examine whether XGAP is phosphorylated by aPKC, GST-XGAP was coexpressed with Flag-xPKCλ and pulled down by glutathione-Sepharose beads. In the presence of PKCλ, the amount of 14-3-3ε precipitated with XGAP increased significantly (Figure 6B). The increased level of XGAP phosphorylation at Ser residue(s) detected by an anti-phospho-serine antibody and the 14-3-3ε bound to XGAP were cancelled by the coexpression of PKCλKE, a dominant-negative form of PKCλ (Figure 6B). These data indicate that XGAP is

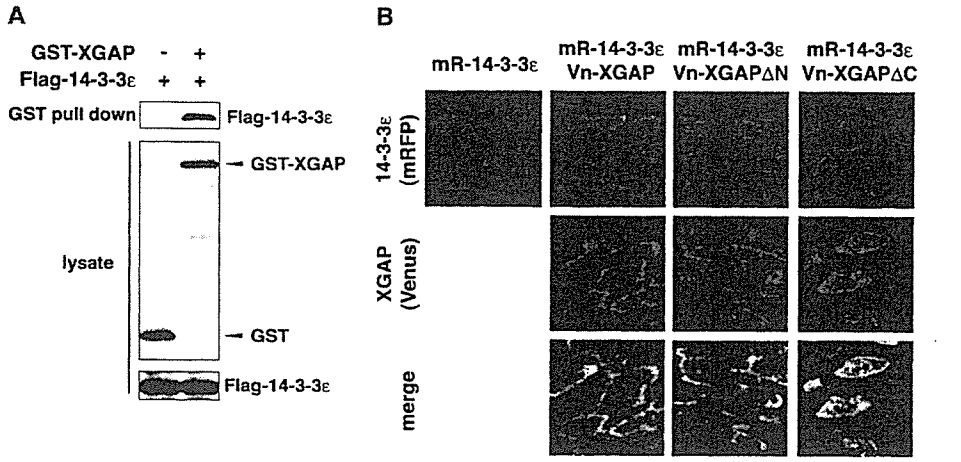

Figure 5. XGAP Interacts with 14-3-3/PAR-5

(A) 14-3-3ε binds to XGAP. Flag-tagged 14-3-3ε and GST or GST-tagged XGAP were expressed in HeLa cells, and the lysate was pulled down by glutathione-Sepharose beads.

(B) XGAP directs the localization of 14-3-3ε through XGAP's C-terminal region. The mRNA of mRFP-tagged 14-3-3ε with or without the mRNA of each truncated construct of Venus-tagged XGAP was coexpressed in the DMZ.

phosphorylated on at least its Ser residue(s) by aPKC, which enhances its binding to 14-3-3.

XGAP and PAR Proteins Are Mutually Required for the Localization to the Mediolateral Ends of Cells

We then observed the intracellular localization of the PAR proteins in dorsal marginal cells. We found that Venus-tagged xPAR-6 and xPKCλ accumulated at both tips of the bipolar cells in the DMZ (Figure 7A), although we could not examine xPAR-1 localization because of its toxicity (data not shown). To further investigate the functional relationship between XGAP and xPAR-6 or xPKCλ, we examined the intracellular localization of xPAR-6 and xPKCλ in the presence or absence of XGAP. XGAP-Mo disrupted the polarized localization of both xPAR-6 and xPKCλ and this aberrant localization was restored by coinjecting resXGAP (Figure 7A). This result indicated that XGAP is required for the mediolateral localization of the xPAR-6 and xPKCλ proteins.

We also observed the Vn-XGAP localization in the cells from which PAR-6 or xPKCλ activity is depleted, to test whether the polarized localization of XGAP requires these PAR proteins. As a result, not only the spindle-shape cell morphology but also the mediolateral localization of Vn-XGAP was inhibited by PAR-6 Mo or xPKCλKE, a dominant negative form of xPKC (Figure 7B). In addition, the speckle-like localization of XGAP in PKCλKE injected cells indicates that PKC may be required for the membrane localization of XGAP. Taking these data together, we propose that XGAP forms a complex with PAR proteins and acts cooperatively with them to polarize the cells undergoing gastrulation cell movement.

Discussion

A GTPase-Independent Function of XGAP Is Required for Gastrulation

XGAP is a member of a GAP subfamily, the ArfGAPs, which target small G proteins. The ArfGAPs comprise a large family of proteins named for their ability to induce the hydrolysis to GDP of GTP bound to Arf (Randazzo and Hirsch, 2004). One obvious possibility was that XGAP regulates gastrulation cell movements by controlling ARF activity. An unexpected finding in this study, however, was that the C-terminal conserved region of XGAP, but not its N-terminal region that includes the GAP catalytic domain, was necessary and sufficient to

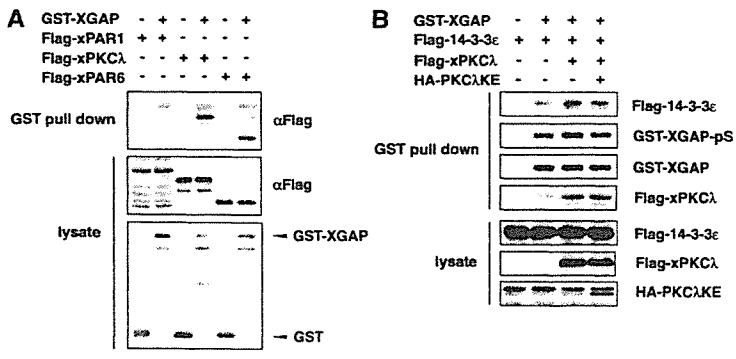

Figure 6. XGAP Interacts with PAR-6 and aPKC

(A) XGAP binds to xPAR-6 or xPKCλ. GST-tagged XGAP was expressed with Flag-xPAR-1, Flag-xPKCλ, Flag-xPAR-6, or Myc-rPAR-3 in HeLa cells.

(B) aPKC phosphorylates XGAP and facilitates the binding of XGAP with 14-3-3. Phosphorylated XGAP was detected by an anti-phospho-serine antibody. xPKCλKE is a dominant-negative form of xPKCλ described previously (Nakaya et al., 2000).

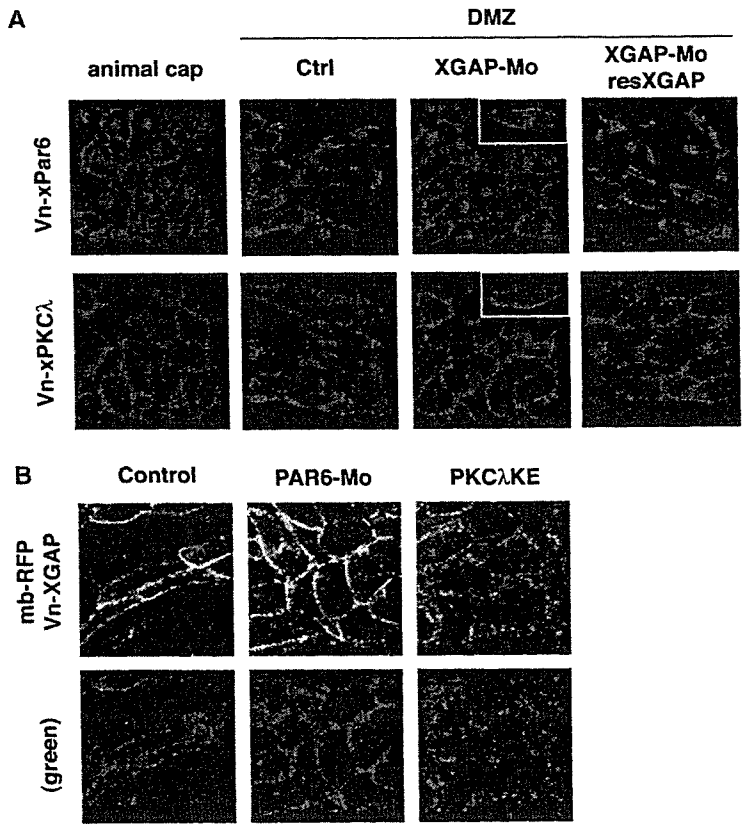

Figure 7. XGAP and PAR Proteins Are Mutually Required for the Intracellular Localization

(A) XGAP regulates the localization of xPAR-6 and xPKCλ proteins. Venus-tagged xPAR-6 or xPKCλ was coinjected with or without XGAP-Mo into two blastomeres to destine them to the animal cap or DMZ. The XGAP-MO phenotype in DMZ was rescued by resXGAP. The horizontal plane is the mediolateral axis. The insets show the disruption of their polarized localization even in the XGAP-Mo-injected cells with spindle-shape morphology like Figure 3A.

(B) xPAR-6 and xPKCλ are required for the polarized localization of XGAP. Venus-tagged XGAP was coinjected with xPAR-6 Mo or xPKCλKE and fated to DMZ. The membrane bound RFP was also injected to mark the cell membrane. The horizontal plane is the mediolateral axis.

promote gastrulation cell movement, suggesting that the N-terminal region is dispensable, at least for controlling gastrulation. In addition, our observation that wild-type XGAP exhibited a weaker effect in the Mo-rescuing experiments and on cell motility in a wound-healing assay than XGAP-ΔN (Figure S5) suggested that the N-terminal region of XGAP might function as a negative regulatory domain for the C-terminal region. These results demonstrated that XGAP may have independent functions through its GAP domain and C-terminal conserved region.

We also tested whether XGAP has a GAP activity. ARFs are classified into three groups based on their structure, and thus we used ARF1, ARF4, and ARF6 as a representative of each group and tested XGAP's GAP activity for each potential substrate. In our experiments, XGAP strongly bound ARF6, but not ARF1 or ARF4, through its GAP domain and activated ARF6's GTPase hydrolysis activity (Figures S6A–S6C), even though XGAP's sequence is most similar to ARG3, which has a GAP activity for ARF1 in vitro (Liu et al., 2001). These data indicated that XGAP may be a specific GAP for ARF6 in *Xenopus*.

Moreover, in a preliminary experiment, the overexpression of both constitutively active and dominant-negative forms of ARF6 (ARF6-Q67L and ARF6-T27N, respectively) caused gastrulation-defective phenotypes, such as neural-tube closure defects, without affecting cell fates (data not shown). These results indicated that ARF6 might also be involved in the regulation of gastrulation cell movements or later neural tube closure. According to our spatio-temporal analysis of ARF6 expression by WISH, however, the transcripts were almost ubiquitous except in the axial mesoderm region and notochord, where XGAP transcripts are detected (Figure S6D). Thus, we conclude that XGAP might cooperate with ARF6 during gastrulation, but even if it does, the cooperation occurs in the ectoderm, not in the intercalating mesoderm.

The Mechanism of XGAP Function and Its Cooperative Action with PAR Proteins

In this report, we found a correlation between XGAP's ability to rescue the XGAP-Mo phenotype and its localization to the tips of the spindle-shaped cells in the DMZ (Figure 4). Therefore, the polarized subcellular localization of the XGAP protein seems to be critical for normal morphogenetic cell movements. We also demonstrated that XGAP has a role in the mediolateral localization of the PAR proteins 14-3-3ε, xPAR-6, and xPKCλ in the DMZ during gastrulation, and the localization coincides with the lamellipodial protrusive activity that pulls cells between one another (Keller, 2002). Since we also found that depletion of one of the polarity proteins, PAR-6, by its specific MO or a dominant-negative form of aPKC disrupted the spindle-shape morphology of DMZ cells and the bipolar localization of Vn-XGAP (Figure 7B), we concluded that XGAP and PAR proteins function in a mutually dependent fashion. Based on our findings that XGAP is required to confine the active protrusions to the mediolateral ends of the cells and that the modification of XGAP by aPKC may be required for