Products of the grg (Groucho-related Gene) Family Can Dimerize through the Amino-terminal Q Domain*

Madalena Pinto, and Corrinne G. Lobe†§

From the Molecular Biology Institute, Cancer Research Group and Departments of Biochemistry and Medical Sciences, McMaster University, Hamilton, Ontario L8S 4K1, Canada

The murine grg (Groucho-related gene) products are believed to interact with transcription factors and repress transcription, thereby regulating cell proliferation and differentiation. Most proteins in the grg family contain all of the domains found in the Drosophila Groucho protein, including the S/P (Ser-Pro-rich) domain required for interaction with transcription factors and the WD40 domain, which is thought to interact with other proteins. However, at least two grg proteins contain only the amino-terminal Q (glutamine-rich) domain. We examined whether the Q domain is used for dimerization between grg proteins, using the yeast two-hybrid system and binding assays with glutathione S-transferase fusion proteins. We found that grg proteins are able to dimerize through the Q domain and that dimerization requires a core of 50 amino acids. Surprisingly, the dimerization does not require the leucine zipper located within the Q domain.

The Groucho protein of Drosophila binds directly to basic helix-loop-helix proteins (bHLH)1 of the Hairy family (1). Together, Groucho and Hairy-like proteins repress transcription factors involved in cell determination, such as fushi tarazu, Enhancer of split, and Achaete-scute complex gene products (1). The consequence of Hairy/Groucho regulation is that cells are directed along one of two possible lineages (2). In this way, Groucho takes part in segmentation of the Drosophila embryo, neurogenesis, and sex determination (1).

The repression by Groucho and Hairy-like proteins requires the S/P (Ser-Pro-rich) domain of Groucho and the conserved carboxyl-terminal Trp-Arg-Pro-Trp (WRIPW) sequence of the Hairy-like transcription factor for direct interactions between the two proteins (1). In addition to the S/P domain, a WD40 motif was recognized in the carboxyl-terminal half of Groucho. The WD10 sequence is a loosely conserved repeat of 40 amino acids separated by a Trp-Asp dipeptide sequence (3). By comparison with the function of the WD10 repeat of other proteins, it is likely to be another domain in Groucho that is involved in protein interaction and may serve to recruit factors required for transcriptional repression (4, 5). The other sequence recognized in the Drosophila protein was the CcN motif, containing Cdc3 and casein kinase sites and a possible nuclear localization signal (6).

Homologues of the Drosophila Groucho gene have been isolated from mouse (Groucho-related genes, grg) (7–9), rat (Enhancer of split-related, Esp) (10), and human (transducin-like element, TLE) (11) cDNA libraries. Comparison of the predicted protein sequences revealed that most of the genes encode proteins with a conserved structural organization (11, 12). At the amino terminus, a glutamine-rich region (Q domain) is present, followed by a Gly/Pro-rich sequence (G/P region), and the CcN, S/P, and WD40 domains. Of these, the Q domain and WD40 domain have a high degree of conservation, suggesting that they play a highly conserved function. The S/P domain is less well conserved, implying that each Groucho family member binds preferentially to a particular Hairy-like transcription factor.

The families of Groucho-related genes in mouse, rat, and human also each include one gene that encodes a truncated homologue of Drosophila Groucho, with only the Q domain and part of the G/P sequence (7, 8, 10). In addition, we found that at least one of the mouse genes, grg3, encodes a long (Gr3g3a) and a short (Gr3g3b) Groucho-related protein (12). The lack of an S/P domain to interact with Hairy-related transcription factors and a WD40 domain to interact with other proteins suggested to us that the short proteins may act to regulate the activity of the longer proteins. This type of regulation is frequently observed for families of transcription factors, wherein a short protein down-regulates a full-length protein by dimerizing with it and acting as a dominant suppressor (13–19).

The consideration that the shorter Groucho homologues might act to regulate the longer Groucho homologues lead us to examine the Q domain for possible dimerization motifs. We and others noted that the Q domain contains a putative leucine zipper sequence (8). Alpha-helical projections of this sequence suggested that it might act in homodimerization among Groucho family members, because the charges that confer specificity of Leu zipper binding are arranged with a mirror image symmetry.2

We have tested the ability of the Groucho homologues to dimerize, using Gr5 and Gr3b with other members of the mouse Grg family. Our studies confirm that the short Groucho homologues can dimerize with the long proteins and localize a critical sequence within the Q domain for this activity. Surprisingly, dimerization does not require the leucine zipper.

EXPERIMENTAL PROCEDURES

Gal-4-fusion Constructs—The plasmids (pGBT9, pGAD424, and pCL1) were obtained in the MatchmakerTM two-hybrid system (Clontech). To construct the in-frame GAL4(AD)-Gr3b fusion construct, the Gr3b cDNA (12) was digested with SmaI and PstI and cloned into

* This work was supported by a grant from the Medical Research Council of Canada. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† To whom correspondence should be addressed: Molecular Biology Institute, LSB 425, McMaster University, 1280 Main St. W., Hamilton, Ontario L8S 4K1, Canada. Tel.: 905-525-9140, ext. 27335; Fax: 905-521-2955.

§ Scholar of the Medical Research Council of Canada.

‡ To whom correspondence should be addressed: Molecular Biology Institute, LSB 425, McMaster University, 1280 Main St. W., Hamilton, Ontario L8S 4K1, Canada.

1 The abbreviations used are: bHLH, basic helix-loop-helix; GST, glutathione S-transferase; NTEN, NaCl-Tris-EDTA-Nonidet; PBS, phosphate-buffered saline.

2 M. Pinto and C. G. Lobe, unpublished data.
Dimerization of Grg Proteins

Results from the yeast two-hybrid assay to test Grg dimerization

TABLE I

| Transformation plasmid | Colony color |
|------------------------|--------------|
| pGBT9 (Gal-4 DNA binding domain only) | white   |
| pGAD24 (Gal-4 activation domain only) | white   |
| pGBT9 + pGAD24 | white   |
| pGBT9-mEspN | white   |
| pGAD24-mEspB | blue    |
| pGBT9-mEspN + pGAD24-mEspB | blue    |
| CL1 (complete Gal 4 protein) | blue    |

(5′-GACCGCATCAAGAATGAGTTCCAG-3′) to change the charged residues of lysine to glutamic acid and glutamic acid to lysine at amino acids 31 and 33, respectively. Leucine residues were mutated using oligonucleotides C5 (5′-GACGTAGCTTACAGCAACGCGAGCTG-3′) and C5L (5′-TACAGCGCGAAGCGTGC-3′) (Operon Technologies, Inc.), changing one leucine to a lysine and a second leucine to an arginine at amino acids 37 (CL32) and 44 (CL33), respectively. Potential positive colonies were identified based on change in antibiotic resistance from tetracycline-resistant to ampicillin-resistant colonies.

Sequence Analysis—All fusion constructs, deletion constructs, and site-directed mutagenesis clones were verified by sequencing using the T7 sequencing kit (Pharmacia Biotech Inc.).

GST-fusion Binding Assay—BL-21 (B. dcm ompT hsdS (rB mB gA) gal (DE3, Stratagene), a protease deficient strain of E. coli, was transformed with either pGEX2T, pGEX2T-Grg3b, or pGEX2T-Grg5 and grown overnight in TT medium with ampicillin (100 µg/ml). Five milliliters of the culture was added to 100 ml of medium containing ampicillin and grown at 37 °C to an A600 of 0.4–0.5 before adding 1 mM isopropyl-1-thio-β-D-galactopyranoside to induce expression of fusion proteins. The culture was incubated at 37 °C for 3 h. The bacteria was harvested and lysed by sonication in 10 ml of NTEN (20 mM Tris, pH 8.0, 100 mM NaCl, 1 mM EDTA, 0.5% Nonidet P-40, and 1 mM dithiothreitol). One ml of 10% Triton X-100 was added, the suspension was mixed, and insoluble material was removed by centrifugation. The supernatant was mixed with 1 ml of glutathione-Sepharose 4B beads (Pharmacia) and gently rocked for 30 min. The beads were washed twice with NTEN, followed by two washes with PBS (137 mM NaCl, 2.7 mM KCl, 5.4 mM Na2HPO4, 1.8 mM KH2PO4, pH 7.4) by adding 10 ml of solution, mixing, and centrifuging. The beads were stored in 10 ml of PBS.

Prior to in vitro transcription, plasmids were linearized by digestion with a restriction enzyme that recognized a site outside the coding sequence and produced a 5′-protruding-end on the plasmid. Capped RNA transcript was synthesized from the linearized plasmid using T3 RNA polymerase for all the deletion constructs and Mash2, and T7 RNA polymerase for the site-directed mutagenesis clones. Transcripts were subsequently used to synthesize 35S-labeled protein. Transcription and translation reactions were performed as instructed by the manufacturer (Promega) in the rabbit reticulocyte lysate system technical manual.

Equivalent amounts of immobilized fusion proteins were mixed with specific 35S-labeled reticulocyte lysate by gentle rocking at 4 °C for 1 h. The glutathione beads were washed twice with 1 ml of NTEN followed by 1 ml of PBS. Fusion and bound proteins were eluted with 30 µl of 50 mM Tris-Cl (pH 8.0), 15 mM reduced glutathione and analyzed by SDS-polyacrylamide gel electrophoresis and autoradiography.

RESULTS

Yeast Two-Hybrid Interaction Analysis—We first used the yeast two-hybrid system to detect possible interactions of the short Grg proteins through the Q domain. The yeast two-hybrid assay is a genetic test used to detect protein interactions in vivo (20). It is based on the fact that transcription factors consist of two separable functional domains, the DNA binding domain and the transcription activation domain. Both components are required to activate transcription. In the yeast two-hybrid syst...
FIG. 1. Binding assays with GST fusion proteins confirm the ability of Grg3b and Grg5 to dimerize. In vitro translated proteins were used for binding assays with GST or GST fusion proteins, adsorbed to glutathione-Sepharose beads. 

A, one-tenth of each of the in vitro translated proteins used for binding was run in lanes 1–4 to visualize the protein products. Lanes 5 through 16 show binding of the translated proteins to three different amounts of GST beads. B, binding of the translated proteins to GST-Grg3b beads. C, binding of the translated proteins to GST-Grg5 beads. Mash2 is a negative control that does not bind Grg proteins.
tem used here, the sequences for the two functional domains of GAL4 have been cloned into two different expression vectors. The pGBT9 vector contains the sequence for the GAL4 DNA binding domain (BD) and the pGAD424 vector contains the sequence for the GAL4 activation domain (AD). By fusing sequences of known proteins (X, Y) downstream and in-frame to the GAL4 sequences, co-transformations with these two expression vectors allows for detection of interactions between X and Y combinations of proteins. Interactions result in β-galactosidase expression that, upon substrate (X-gal) addition, turns colonies with interacting proteins blue, while colonies with non-interacting proteins remain white.

We placed the Grg3b sequence in-frame to the GAL4(AD) of pGAD424 and the Grg5 sequence into GAL4(BD) of pGBT9. β-Galactosidase activity was assayed, and the results are shown in Table I. Yeast transformed with pGBT9, pGAD424, or both resulted in no β-galactosidase activity as indicated by white colonies. Yeast transformed with either the pGBT9-Grg5 or pGAD424-Grg3b fusion constructs also resulted in white colonies. However, yeast co-transformed with both fusion constructs in blue colonies, indicating an interaction between Grg3b and Grg5. The positive control plasmid, CL1, encodes the complete GAL4 protein and also produced blue colonies upon β-galactosidase staining.

Binding Assays with GST Fusion Proteins—The results of the yeast two-hybrid assay were verified with binding assays using GST fusion proteins. We used pGEX2T for inducible expression of Grg3b and Grg5 as GST fusion proteins in bacteria. Bacterial extracts were mixed with glutathione-Sepharose beads to adsorb the fusion protein. The beads with the adsorbed GST proteins were subsequently incubated with radiolabeled in vitro translated Grg1, Grg3b, and Grg5. Translated products that were retained by the GST-Sepharose beads were visualized by gel electrophoresis and autoradiography. As a control, Mash2, which is a basic helix-loop-helix protein that does not bind with Groucho, was used. The first four lanes in Fig. 1A show one-tenth of the translation products. The remainder of the gel in Fig. 1A shows that none of the in vitro translated proteins were retained by the GST protein alone. However, translated Grg1, Grg3b, and Grg5 did bind to beads carrying the GST-Grg3b (Fig. 1B) and GST-Grg5 (Fig. 1C) fusion proteins, whereas Mash2 did not. This again demonstrated that the Grg proteins can dimerize and further showed that the short proteins can bind both to long and to short proteins.

Deletion Analysis to Delineate the Minimal Dimerization Motif—To delineate the necessary sequences for dimerization, we constructed a series of deletion constructs of the Grg5 cDNA template used for in vitro transcription (Fig. 2A). These deletions made progressive truncations from the amino terminus.
and from the carboxyl terminus of the Grg5 protein product. Radiolabeled proteins were synthesized in vitro and incubated with the sepharose beads carrying GST, GST-Grg3b, or GST-Grg5 fusion proteins. Bound proteins were again visualized by gel fractionation and autoradiography (Fig. 2B). One-tenth of the translation products were also visualized (Fig. 2B, lanes 1–7 in the left panel) to determine translation efficiency.

As for the full-length Grg5, no binding was observed when proteins were mixed with GST beads (Fig. 2B, lanes 8–14 in the left gel). However, for the GST-Grg3b and GST-Grg5 beads, binding of the full-length Grg5 and some of the deletion products was observed (Fig. 2B, right panel). We found that removal of 35 amino acids of the carboxyl-terminal region (N162) did not decrease the amount of bound protein. However, further deletion into the Q domain, removing 94 amino acids (N103), significantly reduced the ability of the truncated protein to bind with Grg3b or Grg5. Further deletion from the carboxyl terminus, removing much of the Q domain but leaving the leucine zipper intact (N50), drastically reduced the ability of the truncated protein to dimerize. A deletion of 50 amino acids from the amino terminus, removing the leucine zipper (C147), also significantly reduced the amount of dimerized protein and deletion of 103 amino acids from the amino terminus eliminated dimerization (C94). Deletions of both the amino-terminal 50 amino acids and the carboxyl-terminal 94 amino acids (50/103) resulted in loss of all binding. These results indicate that the critical sequence for dimerization is located between amino acids 50 and 103, but this sequence is insufficient and requires additional amino-terminal or carboxyl-terminal sequence.

**Mutagenesis of the Putative Leucine Zipper**—To further examine whether the leucine zipper plays any role in dimerization activity, we used the N103 Grg5 cDNA, which contains the leucine zipper and the critical core of sequence required for dimerization. The leucine zipper of N103 was disrupted in two different ways using site-directed mutagenesis (Fig. 3A). In one case, the charged amino acids that flank the aliphatic face of the putative alpha helix were changed (Lys31→Glu, Glu33→Lys), to possibly disrupt the dimerization specificity of the protein (N103-CHG). In the second mutant (N103-LEU), two of the leucines were changed to charged amino acids

![Fig. 3. The leucine zipper is not required for Grg dimerization. A, schematic of the changes made to N50-Grg5. Set 1 changes charged residues that are potentially involved in dimerization specificity. Set 2 changes leucine residues to charged residues. B, one-tenth of the in vitro translated proteins, as shown in A, were run in lanes 1–4. The following lanes show the results of incubation of the translated proteins with GST or GST fusion proteins adsorbed to glutathione-Sepharose beads. Proteins that were retained, and therefore capable of dimerizing with the GST fusion protein, were visualized by gel electrophoresis and autoradiography. Mash2 is unable to bind Groucho and serves as a negative control.](Image)
(Leu37→Lys; Leu44→Arg). Mash2 was again used as a negative control. Each of the translation products was run on the gel, as previously, to examine translation efficiency (Fig. 3, lanes 1–4).

The translated proteins were mixed with glutathione-Sepharose beads carrying GST, GST-Grg3b, or GST-Grg5. No binding was observed with GST beads (Fig. 3, lanes 5–8). The N103 product was able to bind to GST-Grg3b and GST-Grg5, as noted above (Fig. 3, lanes 9 and 13), but Mash2 was not (Fig. 3, lanes 12 and 16). Changing either the charged amino acids or the Leu zippers had no effect on the level of protein heterodimerization by N103 to Grg3 (Fig. 3, lanes 10 and 11) but did reduce the ability to form homodimers between N103 (derived from Grg5) and Grg5 (Fig. 3, lanes 14 and 15). Thus, the leucine zipper does not appear to be critical to dimerization of Grg proteins but may strengthen the interactions between homodimers of Grg5.

**DISCUSSION**

We have demonstrated that the Grg proteins are able to dimerize through the amino-terminal Q domain. This domain is highly conserved in all of the mouse Grg family members (12), as well as in other vertebrate and invertebrate species (11); therefore, we expect that the ability to dimerize is common to all of the Groucho homologues.

Grg dimerization requires a core sequence within the Q domain, plus flanking sequence from either side. The additional sequence may be required for the correct structural conformation of the dimerization domain, or there may be additional dimerization sequences in both flanking regions that are required for maximal binding. Mutagenesis of the leucine zipper indicated that, although it may confer stronger binding of the Grg5 protein, it was not the additional sequence required on the amino-terminal side of the core sequence. Therefore, in contrast to our prediction, the leucine zipper does not appear to play a role in Grg dimerization to other Grg proteins. One possibility is that it instead interacts with the basic helix-loop-helix transcription factors with which it is likely to interact, namely the Hes (Hairy/Enhancer of split homologous) (21–23) proteins.

The potential interaction between Grg proteins suggests that the short Grg proteins may act to modify activity of the long Grg proteins. At the present time, we have no indication whether this would be a positive or negative regulation. However, the function of the long Grg proteins probably requires binding to Hes and DNA, by analogy to the Drosophila proteins (24); therefore, it will be important to determine if tertiary complexes form between the short and long Grg proteins and Hes. It will also be interesting to examine if both forms of the Grg protein interact with Hes while it is bound to DNA and also the effect on transcriptional regulation by the Hes proteins.

In Drosophila, Groucho is part of a complex network of transcriptional regulatory proteins. This includes basic helix-loop-helix transcription factors that are involved in cell determination, partners that lack a basic domain and act as negative regulators, the Hairy-related proteins, which also act as negative regulators, and Groucho, which is required for repression by Hairy-related proteins (6, 25, 26). The complex regulation of these transcription factors may exist because they are involved in controlling cell proliferation and differentiation, and therefore, inappropriate function would have serious consequences for the host. The existence of two types of Groucho-related proteins and their ability to dimerize suggest that there is yet an additional level of transcriptional regulation in mammalian systems.