Feeding, egg-laying, dauer arrest and reproductive lifespan assays

The rate of pharyngeal pumping at 20 °C was scored by counting pharynx terminal bulb contractions[12]. The age of eggs in the uterus was estimated by counting embryonic cells in each egg. For dauer arrest, eggs laid overnight at 20°C were shifted to 15 °C, 20 °C, or 26°C, and scored for dauer arrest at 48 hours (26 °C), 72 hours (20 °C) and 110 hours (15 °C). The number of animals scored represents two trials of each genotype and three independent experiments. Visualized by DIC optics, the *tph-1(mg280)* dauers had dauer alae (17/18, full dauer alae; and 1/18, partial alae) and constricted pharynx (17/18, although some of them pumped occasionally in a 2 min interval (3/18). Both *daf-7(e1372);Ex tph-1(+)* and *daf-7(e1372);Ex tph-1::GFP* animals do not form dauers at 15 °C. Number of animals scored in dauer arrest assays at 20 °C: wild type, 5,442; *tph-1(mg280)*, 2362; *tph-1(mg280)* with serotonin, 2,514; *tph-1(mg280);daf-16(mgDf50)*, 3,537; *tph-1(mg280);daf-3(mgDf90)*, 2,423. At 15 °C: wild type, 1,059; *daf-7(e1372)*, 1,743; *tph-1(mg280)*, 3,099; *daf-7(1372);tph-1(mg280)*, 1,514. At 26 °C: wild type, 1,797; *daf-7(1372)*, 901; *tph-1(mg280)*, 3,462; *daf-7(e1372);Ex tph-1(+)*, 863; *daf-7(e1372);Ex tph-1::GFP*, 1,601. The error bars represent s.e.m. of the scores from three independent experiments. For reproductive lifespan assays, animals at 20 °C were transferred every day after the L4 stage to a fresh plate seeded with bacteria and progeny during that 24-hour period were counted. Most animals lived days after they ceased reproduction. Many *tph-1(mg280)* animals died as the consequence of progeny hatching internally (12/29). Thus, their reproductive lifespans may be underestimated. Average progeny laid per animal in reproductive lifespan experiment: wild type, 287 (*n*=22); *tph-1(mg280)*, 215 (*n*=29); *tph-1(mg280);Ex tph-1(+)*, 296 (*n*=10); *daf-16(mgDf50);tph-1(mg280)*, 199 (*n*=17); *daf-16(mgDf50)*, 226 (*n*=15); *egl-1(n986)*, 144 (*n*=18). The data are the summary of 2–4 independent sets of experiments. For fat staining, animals grown at 20°C were stained with Sudan Black as previously described[14]. All behavioural and metabolism anlayses used animals fed *E. coli* OP50.

1. Sanders-Bush, E. & Mayer, S. E. in *Goodman & Gilman's Pharmacological Basis of Therapeutics* 9th edn (eds Hardman, J. G. *et al.*) 249–263 (McGraw-Hill, New York, 1996).
2. Leibowitz, S. F. & Alexander, J. T. Hypothalamic serotonin in control of eating behaviour, meal size, and body weight. *Biol. Psychiatry* **44**, 851–864 (1998).
3. Loer, C. M., Davidson, B. D. & McKerrow, J. M. A phenylalanine hydroxylase gene from the nematode *C. elegans* is expressed in the hypodermis. *J. Neurogenet.* **13**, 157–180 (1999).
4. D'Sa, C. M., Arthur, R. E. Jr & Kuhn, D. M. Expression and deletion mutagenesis of tryptophan hydroxylase proteins: delineation of the enzyme catalytic core. *Neurochem.* **67**, 917–926 (1996).
5. Rand, J. B. & Nonet, M. L. *Synaptic Transmission in C. elegans II* (eds Riddle, D. L. *et al.*) 611–643 (Cold Spring Harbor Laboratory Press, New York, 1997).
6. Jansen, G., Hazendonk, E., Thijssen, K. L. & Plasterk, R. H. Reverse genetics by chemical mutagenesis in *C. elegans. Nature Genet.* **17**, 119–121 (1997).
7. Horvitz, H. R., Chalfie, M., Trent, C., Sulston, J. E. & Evans, P. D. Serotonin and octopamine in the nematode *C. elegans. Science* **216**, 1012–1014 (1982).
8. Avery, L. & Horvitz, H. R. Effects of starvation and neuroactive drugs on feeding in *C. elegans. J. Exp. Zool.* **253**, 263–270 (1990).
9. Sulston, J., Dew, M. & Brenner, S. Dopaminergic neurons in the nematode *Caenorhabditis elegans. J. Comp. Neuro.* **163**, 215–226 (1975).
10. Weinshenker, D, Garriga, G. & Thomas, J. H. Genetic and pharmacological analysis of neurotransmitters controlling egg laying in *C. elegans. J. Neurosci* **15**, 6975–6985 (1995).
11. Waggoner, L. E., Zhou, G. T., Schafer, R. W. & Schafer, W. R. Control of alternative behavioural states by serotonin in *Caenorhabditis elegans. Neuron* **21**, 203–214 (1998).
12. Duerr, J. S. *et al.* The *cat-1* gene of *Caenorhabditis elegans* encodes a vesicular monoamine transporter required for specific monoamine-dependent behaviours. *J. Neurosci.* **19**, 72–84 (1999).
13. Riddle, D. L. *Genetic and Environmental Regulation of Dauer Larva Development in* C. elegans II. (eds Riddle, D. L. *et al.*) 739–768 (Cold Spring Harbor Laboratory Press, New York, 1997).
14. Kimura, K. D., Tissenbaum, H. A., Liu, Y. & Ruvkun, G. *daf-2*, an insulin receptor-like gene that regulates longevity and diapause in *C. elegans. Science* **277**, 942–946 (1997).
15. Ren, P. F. *et al.* Control of *C. elegans* larval development by neuronal expression of a TGF-β homolog. *Science* **274**, 1389–1391 (1996).
16. Schackwitz, W. S., Inoue, T. & Thomas, J. H. Chemosensory neurons function in parallel to mediate a pheromone response in *C. elegans. Neuron* **17**, 719–728 (1996).
17. Bargmann, C. Neurobiology of the *C. elegans* genome. *Science* **282**, 2028–2033 (1998).
18. Zwaal, R. R., Mendel, J. E., Sternberg, P. W. & Plasterk, R. H. A. Two neuronal G proteins are involved in chemosensation of the *C. elegans* dauer-inducing pheromone. *Genetics* **145**, 715–727 (1997).
19. Ogg, S. *et al.* The Fork head transcription factor DAF-16 transduces insulin-like metabolic and longevity signals in *C. elegans. Nature* **389**, 994–999 (1997).
20. Patterson, G. I., Koweek, A., Wong, A., Liu, Y. & Ruvkun, G. The DAF-3 Smad protein antagonizes TGF-β-related receptor signaling in the *C. elegans* dauer pathway. *Genes Dev.* **11**, 2679–2690 (1997).
21. Kenyon, C., Chang, J., Gensch, E., Rudner, A. & Tabtiang, R. A. *C. elegans* mutant that lives twice as long as wild type. *Nature* **366**, 461–464 (1993).
22. Gems, D. *et al.* Two pleiotropic classes of *daf-2* mutation affect larval arrest, adult behaviour, reproduction and longevity in *Caenorhabditis elegans. Genetics* **150**, 129–155 (1998).
23. Conrad, B. & Horvitz, H. R. The *C. elegans* protein EGL-1 is required for programmed cell death and interacts with the Bcl-2-like protein CED-9. *Cell* **93**, 519–529 (1998).
24. Duret, L. Gues, N., Peitsch, M. & Bairoch, A. New insulin-like proteins with atypical disulfide bond pattern characterized in *C. elegans* by comparative sequence analysis and homology modeling. *Genome Res.* **8**, 348–353 (1998).
25. Peschke, E., Peschke, D., Hammer, T. & Csernus, V. Influence of melatonin and serotonin on glucose-stimulated insulin release from perifused rat pancreatic islets in vitro. *J. Pineal Res.* **23**, 156–163 (1997).
26. Breum, L., Bjerre, U., Bak, J. F., Jacobsen, S. & Astrup, A. Long-term effects of fluoxetine on glycemic control in obese patients with non-insulin-dependent diabetes mellitus or glucose intolerance: influence on muscle glycogen synthase and insulin receptor kinase activity. *Metabolism* **44**,1570–1576 (1995).
27. Avery, L. The genetics of feeding in *Caenorhabditis elegans. Genetics* **133**, 897–917 (1993).
28. Lakowski, B. & Hekimi, S. The genetics of caloric restriction in *Caenorhabditis elegans. Proc. Natl Acad. Sci. USA* **95**,13091–13096 (1998).
29. Nonogaki, K., Strack, A. M., Dallman, M. F. & Tecott, L. H. Leptin-independent hyperphagia and type 2 diabetes in mice with a mutated serotonin 5-HT2C receptor gene. *Nature Med.* **4**, 1152–1156 (1998).
30. Barzilai, A., Kennedy, T. E., Sweatt, J. D. & Kandel, E. R. 5-HT modulates protein synthesis and the expression of specific proteins during long-term facilitation in Aplysia sensory neurons. *Neuron* **2**, 1577–1586 (1989).

Evidence for stabilizing selection in a eukaryotic enhancer element

Michael Z. Ludwig*, **Casey Bergman***, **Nipam H. Patel**†
& Martin Kreitman*

* *Department of Ecology and Evolution, University of Chicago, 1101 E. 57th Street, Chicago, Illinois 60637, USA*
† *Department of Organismal Biology and Anatomy, and Howard Hughes Medical Institute, MC1028, N-101, 5841 South Maryland Avenue, Chicago, Illinois 60637, USA*

Eukaryotic gene expression is mediated by compact *cis*-regulatory modules, or enhancers, which are bound by specific sets of transcription factors[1]. The combinatorial interaction of these bound transcription factors determines time- and tissue-specific gene activation or repression. The *even-skipped* stripe 2 element controls the expression of the second transverse stripe of *even-skipped* messenger RNA in *Drosophila melanogaster* embryos, and is one of the best characterized eukaryotic enhancers[2–4]. Although *even-skipped* stripe 2 expression is strongly conserved in *Drosophila*, the stripe 2 element itself has undergone considerable evolutionary change in its binding-site sequences and the spacing between them. We have investigated this apparent contradiction, and here we show that two chimaeric enhancers, constructed by swapping the 5′ and 3′ halves of the native stripe 2 elements of two species, no longer drive expression of a reporter gene in the wild-type pattern. Sequence differences between species have functional consequences, therefore, but they are masked by other co-evolved differences. On the basis of these results, we present a model for the evolution of eukaryotic regulatory sequences.

Multiple binding sites for each of four transcription factors, the activators bicoid (bcd) and hunchback (hb), and the repressors Kruppel (Kr) and giant (gt), have been physically localized to the stripe 2 element (S2E) (ref. 5). Genetic and experimental evidence indicates that concentration gradients of bicoid and hunchback can activate *eve* in a broad domain, whereas the more localized expression of the S2E is determined by the repressors giant anteriorly and Kruppel posteriorly[5–7]. Experimental elimination, addition or augmentation of both repressor and activation sites produces predictable changes in reporter-gene expression, and provides evidence for the mechanisms of enhancer function[5–7]. This model emphasizes the functional importance of binding-site sequences, as well as the

number and spatial configuration of binding sites within an element[8].

In a comparison of 13 species, none of the 16 surveyed *D. melanogaster* binding sites is completely conserved (Fig. 1a). Most experimentally verified binding sites have accumulated point substitutions, and three are recognizable in only a subset of taxa. Each S2E also differs in the spacing between binding sites[9,10]. Despite these differences, our previous experiments with reporter constructs of native S2Es of four species showed that each enhancer drives reporter-gene expression at the identical time and location in early *D. melanogaster* blastoderm embryos[9] (see also Fig. 1).

Does this mean that the substitutional differences between species are functionally inconsequential? To answer this question, we

focused our investigation on comparison of native and chimaeric stripe 2 elements of *D. melanogaster* and *D. pseudoobscura*, whose most recent common ancestor occurred 40–60 million years ago[11]. These species were chosen for this analysis because our comparative binding-site prediction method (see Methods) indicated potentially important differences in the S2Es, including the absence of the bcd-3 site, the presence of a new Kr site and reductions in likelihood potentials for bcd-4, bcd-2, Kr-4 and hb-3 sites in *D. pseudoobscura* relative to *D. melanogaster*.

The two native S2E sequences S2E(m) and S2E(p), where m and p refer to *D. melanogaster* and *D. pseudoobscura* respectively, were placed in a reporter-gene construct that also included an internal control for position effect[9,12]. These two native constructs expressed
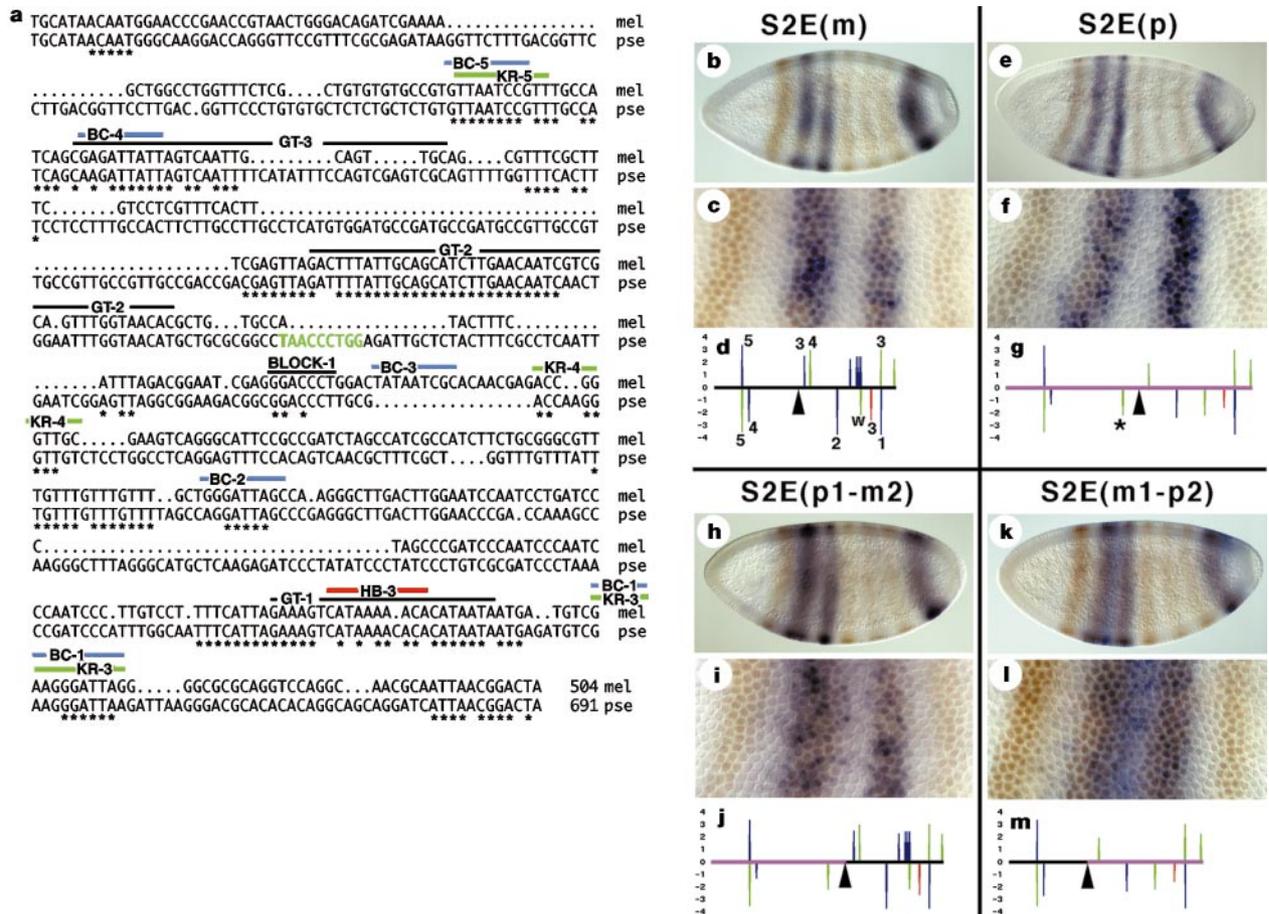

Figure 1 Expression of native and chimaeric *eve* stripe 2 elements from *D. melanogaster* and *D. pseudoobscura*. **a**, Alignment of *eve* stripe 2 enhancer regions in *D. melanogaster* (mel) and *D. pseudoobscura* (pse). Dots indicate gaps in aligned sequences. The binding sites in *D. melanogaster* for the transcription factors, bicoid (BC, blue), hunchback (HB, red), Kruppel (KR, green) and giant (GT, black), are shown above the sequence. The conservative 'Block 1' was used to create two complementary chimaeric enhancers, designated S2E(m1-p2) and S2E(p1-m2), where (*x-y*) corresponds to species *x* distal segment connected to species *y* proximal segment. Stars denote conserved nucleotides of 13 *Drosophila* species (*D. melanogaster*, *D. simulans*, *D. mauritiana*, *D. sechellia*, *D. erecta*, *D. orena*, *D. yakuba*, *D. teissieri*, *D. takahashii*, *D. ananassae*, *D. pseudoobscura*, *D. virilis*, and *D. picticornis*). The predicted *D. pseudoobscura* Kr site is indicated in green text. **b**–**m**, Comparison of *lacZ* mRNA expression driven by natural stripe 2 enhancers from *D. melanogaster* and *D. pseudoobscura* with that driven by chimaeric stripe 2 enhancers S2E(m1-p2) and S2E(p1-m2). Even-skipped protein (brown) and *lacZ* mRNA (purple) are simultaneously detected in the embryos transformed with the stripe 2 enhancers and *D. melanogaster* stripe 3 + 7 enhancer–*lacZ* gene fusion. Embryos were selected to be at the same time point in development by choosing ones in which the native *eve* protein stripes 3 and 7 and the *lacZ* stripes coincided, and by observing the extent of cellularization. **b**–**d**, Stripe 2 enhancer from *D. melanogaster*; **e**–**g**, stripe 2 enhancer

from *D. pseudoobscura*; **h**–**j**, chimaeric S2E(p1-m2) enhancer; **k**–**m**, chimaeric S2E(m1-p2) enhancer. **b**, **e**, **h**, **k**, Sagittal focus; **e**, **f**, **i**, **l**, higher magnification in superficial focus of stripes 2 and 3 from the embryos in **b**, **e**, **h**, **k**. **d**, **g**, **j**, **m**, Schematic presentations of the natural and chimaeric stripe 2 enhancers and binding-site likelihood prediction of the enhancer constructs. Solid black lines indicate the *D. melanogaster* enhancer sequence. Solid pink lines indicate the *D. pseudoobscura* enhancer sequences. Positive and negative values in the plots correspond to putative binding sites on the plus strand and minus strand, respectively. The ordinate represents the log-likelihood value of a given sequence under the PWM model of binding-site usage relative to random base usage. Statistical-mechanic considerations predict a correlation between likelihood ratio scores and binding affinity[19]. The abscissa represents position in the construct and shows differences in the total construct lengths. Plotted are peaks with $\text{Log}(L) \geq 2$ and peaks otherwise located in conserved blocks. Green, *Kruppel*; blue, *bicoid*; red, *hunchback*. Experimentally verified binding sites in *D. melanogaster* are labelled in **a**. The cluster of high likelihood bcd peaks in **d** are not detected by DNase footprinting[6], but one of them has experimentally mutated by one nucleotide to a functional site[7]. A Kr site at the 3′ end of the construct may also be functional because it is conserved between species. The conserved 'Block 1' is indicated by the triangle.

in transformed embryos at a time point and location along the A–P axis that is indistinguishable from native *eve* expression (Fig. 1b, c, e, f). This result indicates either that the mutational differences between these S2Es are functionally unimportant (to the level of resolution allowed by the experimental procedure), or that the changes, if functional, balance one another in such a way as to produce no net functional change in expression.

We predicted that any chimaeric enhancer that contains the conserved sequences of the S2E would have wild-type function if these sequences only are important for *eve* stripe 2 expression. However, if substitutional differences between lineages are functional then it should be possible to create chimaeric sequences that disrupt wild-type function. To test these alternatives, we created two complementary chimaeric enhancers in which the distal or the proximal parts of *D. melanogaster* S2E were substituted with corresponding parts of the *D. pseudoobscura* enhancer (Fig. 1d, g, j, m). The break lies within a conservative block of seven bases midway through the S2E (Fig. 1a, Block 1), which allowed us to create a 'seamless' connection between homologous parts in the chimaeric constructs. The chimaeric enhancers are designated S2E(m1-p2) and S2E(p1-m2), where (*x-y*) corresponds to species *x* distal (5′) segment connected to species *y* proximal (3′) segment.

In separate experiments involving multiple independent transformants of each chimaeric construct, we found consistent evidence for a posterior shift of about the width of two cells in reporter-gene expression in the S2E(m1-p2) construct (Fig. 1k, l). In some embryos, and in some locations along the circumference of the stripe, this posterior expression appears to involve an expansion of the stripe rather than a simple shift. These two phenotypic defects, posterior shifting and expansion, suggest both a change in the sensitivity to the bcd/hb activation gradients, and a reduction in the overall effectiveness of Kr repression of the posterior stripe margin. The absence of a putative *D. pseudoobscura*-specific Kr site (Fig. 1g, asterisk) in the S2E(m1-p2) chimaera, on the basis of binding-site prediction, provides a testable hypothesis to explain the posterior shift in reporter-gene expression. The complementary construct, S2E(p1-m2), does not show the same defect. Instead, it appears to be defective in that it leads to subtle expansion of both the anterior and posterior borders of the stripe compared with native *eve* expression (Fig. 1h, i).

To explain these results, we propose that stabilizing selection has maintained phenotypic constancy for *eve* expression but has allowed mutational turnover of functionally important sites. Stripe 2 expression can be viewed as a quantitative character in which stabilizing selection has preserved *eve* expression to a specific band of cells and a particular time in embryogenesis. Mutational changes in the element, including some base substitutions within binding sites and short insertions or deletions in the spacer regions between binding sites, are proposed to have only weak functional (and therefore selective) effects. Theoretical models of phenotypic traits under stabilizing selection show that nearly neutral variation (that is, slightly deleterious and advantageous mutations) reaches fixation at an appreciable rate by the process of genetic drift[13–15]. The model implies that each species lineage will differ by many functionally compensatory mutations. Such a pattern of substitution, we predict, will be a common theme in *cis*-regulatory evolution.

Our results have implications for understanding changes in gene expression and morphological evolution. If our model of enhancer evolution is correct, then weakly selected mutations in regulatory elements will be present in natural populations and will be available for directional change by natural selection. Consistent with this prediction, two traits in *Drosophila* that differ between species but that are likely to be under stabilizing selection within a species—abdominal bristle number and wing morphology—are reported to have significant associations between quantitative trait variation and nucleotide polymorphism in noncoding regions of candidate loci[16,17]. But as our results indicate, selection can maintain func-

tional conservation of gene expression for long periods of evolutionary time despite binding site turnover, which may make it difficult to identify homologous elements in different species groups by sequence comparison alone. □

Methods

Binding-site prediction

Samples of *Drosophila* DNase 1 footprinted sequences were compiled from the literature for bcd, hb and Kr. Only five sites have been established for gt and thus a position weight matrix (PWM) was not built for this factor. We used the predictive update version of the Gibbs sampler to produce an unbiased local alignment and corresponding binding-site usage matrix[18]. We computed the likelihood ratio $Log(L) = \Sigma_j Log[P_{ij}/P_i]$ for all overlapping windows, where P_{ij} is the frequency of base i at position j of the binding-site usage matrix, P_i is the frequency of base i in the background nucleotide composition, and summation is over the width of the binding site. For bcd we used 53 sites from 8 targets; for Kr we used 42 sites from 10 targets; and for hb we used 96 sites from 15 targets (available on request from C.B.). This method successfully identifies all of the experimentally verified binding sites for these three factors (Fig. 1).

Cloning, P-element-mediated transformation and whole-mount *in situ* hybridization

Initial plasmids containing *eve* stripe 2 enhancer regions from *D. melanogaster* and *D. pseudoobscura*, and P-element transformation vector, cloning, amplification, sequencing, P-transformation and whole-mount *in situ* hybridization, analysis of enhancer stripe 2 expression using a reporter gene, and alignment of DNA sequences have been described[9].

Construction of transgenes

The enhancer regions containing the *eve* stripe 2 element for *D. melanogaster* and *D. pseudoobscura* species, and enhancer halves used for making chimaeric constructs were obtained by PCR from the corresponding plasmids[9]. The fragment containing the S2E(m) element was amplified using primers ME Asp + (5′-aaaaggtacctgcataacaatggaacccga-3′) and ME Pst – (5′-aaaactgcagtagtccgttaattgcgtt-3′). The fragment containing the S2E(p) element was amplified using primers PSE Asp + (5′-aaaaggtacctgcataacaatgggcaagga-3′) and PSE Pst – (5′-aaaactgcagtagtccgttaatgatcct-3′). The fragment S2E(m1) was amplified using primers ME Asp + and ME 1 bl – (5′-agggtccctcgattccgtct-3′). The fragment S2E(m2) was amplified using primers ME 2 bl + (5′-ggactataatcgcacaacga-3′), and ME Pst–. The fragment S2E(p1) was amplified using primers PSE Asp+ and PSE 1 bl – (5′-agggtccgccgtcttccgcct-3′). The fragment S2E(p2) was amplified using primers PSE 2 bl + (5′-tgcgaccaagggttgtctcct-3′) and PSE Pst –. The chimaeric enhancer S2(m1-p2) was prepared by blunt ligation of S2E(m1) and S2E(p2), and re-amplification with the primers ME Asp+ and PSE Pst–. The chimaeric enhancer S2(p1-m2) was prepared by blunt ligation of S2E(p1) and S2E(m2), and re-amplification with the primers PSE Asp+ and ME Pst–. The primers contained the restriction sites for Asp718 and *Pst*1, respectively, at their 5′ and 3′ ends. After digestion with these enzymes, the PCR fragments were cloned into Asp718 and *Pst*1 sites of the plasmid. Entirely correct inserts were identified for further use by sequencing independent clones. The spacing between the 'experimental' stripe 2 element and the 'control' stripe 3 + 7 element is sufficient to assure independent function of the two enhancers[12]. The stripe 3 + 7 *lacZ* expression allows us to control for position effects in expression due to the site of insertion in the genome of different constructs, and it also allows us to precisely control for the time point in development when expression patterns in individual embryos are being compared[9].

1. Arnone, M. I. & Davidson, E. H. The hardwiring of development: organization and function of genomic regulatory systems. *Development* **124**, 1851–1864 (1997).
2. Frasch, M. & Levine, M. Complementary patterns of *even-skipped* and *fushi tarazu* expression involve their differential regulation by a common set of segmentation genes in *Drosophila*. *Genes Dev.* **1**, 981–995 (1987).
3. Goto, T. P., Macdonald, P. & Maniatis, T. Early and late periodic patterns of *even-skipped* expression are controlled by distinct regulatory elements that respond to different spatial cues. *Cell* **57**, 413–422 (1989).
4. Harding, K., Hoey, T., Warrior, R. & Levine, M. Autoregulatory and gap response elements of the *even-skipped* promoter of *Drosophila*. *EMBO J.* **8**, 1205–1212 (1989).
5. Stanojevic, D., Small, S. & Levine, M. Regulation of a segmentation stripe by overlapping activators and repressors in the *Drosophila* embryo. *Science* **254**, 1385–1387 (1991).
6. Small, S., Blair, A. & Levine, M. Regulation of *even-skipped* stripe 2 in the *Drosophila* embryo. *EMBO J.* **11**, 4047–4057 (1992).
7. Arnosti, D., Barolo, S., Levine, M. & Small, S. The eve stripe 2 enhancer employs multiple modes of transcriptional synergy. *Development* **122**, 205–214 (1996).
8. Hewitt, G. F. *et al.* Transcriptional repression by the *Drosophila* giant protein: *cis* element positioning provides an alternative means of interpreting an effector gradient. *Development* **126**, 1201–1210 (1999).
9. Ludwig, M. Z., Patel, N. H. & Kreitman, M. Functional analysis of eve stripe 2 enhancer evolution in *Drosophila*: rules governing conservation and change. *Development* **125**, 949–958 (1998).
10. Ludwig, M. Z. & Kreitman, M. Evolutionary dynamics of the enhancer region of *even-skipped* in *Drosophila*. *Mol. Biol. Evol.* **12**, 1002–1011 (1995).
11. Beverly, S. M. & Wilson, A. C. Molecular evolution in *Drosophila* and the higher Diptera II. A time scale for fly evolution. *J. Mol. Evol.* **21**, 1–13 (1984).
12. Small, S., Arnosti, D. N. & Levine, M. Spacing ensures autonomous expression of different stripe enhancers in the *even-skipped* promoter. *Development* **119**, 767–772 (1993).

13. Kimura, M. Possibility of extensive neutral evolution under stabilizing selection with special reference to non-random usage of synonymous codons. *Proc. Natl Acad Sci. USA* **78**, 5773–5777 (1981).

14. Kimura, M. *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, Cambridge, 1983).

15. Ohta, T. & Tachida, H. Theoretical study of near neutrality. I. Heterozygosity and rate of mutant substitution. *Genetics* **126**, 219–229 (1990).

16. Long, A. D., Lyman, R. F., Langley, C. H. & Mackay, T. F. Two sites in the Delta gene region contribute to naturally occurring variation in bristle number in *Drosophila melanogaster. Genetics* **149**, 999–1017 (1998).

17. Gibson, G. & Hogness, D. S. Effect of polymorphism in the *Drosophila* regulatory gene Ultrabithorax on homeotic stability. *Science* **271**, 200–203 (1996).

18. Lawrence, C. E. *et al.* Detecting subtle sequence signals: a Gibbs sampling strategy for multiple alignment. *Science* **262**, 208–214 (1993).

19. Berg, O. G. & von Hippel, P. H. Selection of DNA binding sites by regulatory proteins. Statistical–mechanical theory and application to operators and promoters. *J. Mol. Biol.* **193**, 723–750 (1987).

Structure of human guanylate-binding protein 1 representing a unique class of GTP-binding proteins

Balaji Prakash*, Gerrit J. K. Praefcke*, Louis Renault, Alfred Wittinghofer & Christian Herrmann

Max-Planck-Institut für Molekulare Physiologie, Otto-Hahn-Strasse 11, 44227 Dortmund, Germany

* These authors contributed equally to this work

Interferon-γ is an immunomodulatory substance that induces the expression of many genes to orchestrate a cellular response and establish the antiviral state of the cell. Among the most abundant antiviral proteins induced by interferon-γ are guanylate-binding proteins such as GBP1 and GBP2 (refs 1, 2). These are large GTP-binding proteins of relative molecular mass 67,000 with a high-turnover GTPase activity[3] and an antiviral effect[4]. Here we have determined the crystal structure of full-length human GBP1 to 1.8 Å resolution. The amino-terminal 278 residues constitute a modified G domain with a number of insertions compared to the canonical Ras structure, and the carboxy-terminal part is an extended helical domain with unique features. From the structure and biochemical experiments reported here, GBP1 appears to belong to the group of large GTP-binding proteins that includes Mx and dynamin, the common property of which is the ability to undergo oligomerization with a high concentration-dependent GTPase activity[5].

Guanylate-binding proteins (GBP1 and 2) were originally identified as proteins from an extract of human fibroblasts treated with interferons, γ-interferon being the most effective, that bind to agarose-bound GMP, GDP and GTP[1,2]. Smaller guanylate-binding proteins of relative molecular mass 47,000 ($M_r = 47K$)[6] are also induced by γ-interferon, whereas α- and β-interferon induce antiviral GTP-binding Mx proteins[7]. Human (h)GBP1 is expressed to mediate an antiviral effect against vesicular stomatitis virus and encephalomyocarditis virus[4]. The biochemical properties of GBPs are clearly different from those of Ras-like and heterotrimeric GTP-binding proteins. They bind guanine nucleotides with low affinity (micromolar range), are stable in their absence and have a high turnover GTPase[3,8]. In addition to binding GDP/GTP, they have the unique ability to bind GMP with equal affinity and hydrolyse GTP not only to GDP but also to GMP[3]. As a first step towards under-

standing the biochemistry and biology of GBPs, we have determined the three-dimensional structure of hGBP1. Furthermore, we show nucleotide-dependent oligomerization of hGBP1 and concentration dependence of its GTPase reaction rate.

We determined the structure of full-length, histidine(His$_6$)-tagged hGBP1 in the absence of nucleotide to 1.8 Å. The final model comprises residues 6–583 and 341 water molecules (Fig. 1a, b), with some poorly defined, probably mobile loops (dashed lines). The crystallographic data are summarized in Table 1. The structure can be divided into a compact, globular α,β-domain (6–278), which we term the LG (Large G) domain, and an elongated, purely α-helical domain. The domains are connected by a short intermediate region consisting of one α-helix and a short two-stranded β-sheet. The connecting region is not an independent domain; it is packed, via helix α6, against the β1/α1 region of the LG domain, away from the presumed nucleotide-binding site (see below). It could be involved in stabilizing the relative location of the two domains against each other. The helical domain is composed of seven helices, which extend 90 Å away from the LG domain.

The LG domain of GBPs contains the conserved sequence elements of GTP-binding proteins with modifications. Originally the N/TKxD motif was believed to be absent in the GBPs[2]. An Asp-Asn mutation can produce a change in specificity from guanine to xanthine nucleotides in many GTP-binding proteins such as EF-Tu[9]. As an Asp-Asn mutation in the [181]TLRD[184] motif of hGBP1 behaves similarly, it is postulated that Asp 184 should bind the guanine base through a bidentate hydrogen bond[8]. It was thus expected that GBP1 would contain the Ras G domain or a variation thereof. The structure shows that the 278-residue globular domain largely resembles the canonical architecture of Ras (~170 residues), allowing for additions and insertions (labelled I). Superimposing the structures of hGBP1 and Ras-GDP (Fig. 2a) gives a root mean square (r.m.s) deviation of 1.1 Å for 112 common Cα atoms. The LG domain consists of an eight-stranded β-sheet with six parallel and two anti-parallel strands surrounded by nine helices, whereas Ras contains six β-strands and five helices (Fig. 1b). Using the secondary structure elements of Ras as the basis for the comparison and retaining the corresponding numbering, the additional elements are β0, α0 and β−1 on the N-terminal side of the sheet, and helices α3′ (I3) and α4′ (I4) (Fig. 1b). Apart from a short insertion (I1) in switch I, there are comparatively long loop insertions between the [97]DxxG[100] motif and α2 (I2), and between β6 and α5 (I5) of the canonical G domain, respectively (Figs 1 and 2). Whereas I1, I2 and I5 could be involved in nucleotide binding, I3 and I4, on the opposite side of the protein, apparently mediate contact with the C-terminal helix.

As GBP is stable in the absence of nucleotide, whereas Ras-like and Gα GTP-binding proteins are not, it was of interest to investigate the effect of the absence of nucleotide on the structure. As all P-loop-containing proteins[10] bind the β/γ-phosphate of the nucleotide in a similar manner, and as the role of Asp 184 in binding the guanine base is similar to that of the Asp of the canonical N/TKxD motif, we can locate the nucleotide-binding site of hGBP1 using the RasGDP–hGBP1 overlay (Fig. 2b). From this comparison we can also see that, although part of the binding site is more accessible to the solvent than in Ras–nucleotide complexes, part of the polypeptide chain is in a position that interferes with nucleotide binding. Perhaps owing to the absence of nucleotide, the polypeptide chain around the binding site is mobile, as no electron density is visible for residues 69–72 (I1) in the region analogous to switch I, residues 190–193 close to the [181]TLRD[184] motif and residues 244–257 in I5, close to the SAK/L motif, which is conserved only in the Ras family and is absent in GBPs.

The (phosphate-binding) P loop[10], residues 45–52, adopts a structure different from that of the Ras–nucleotide complexes. The invariant lysine residue of the P loop does not interact with the main-chain carbonyls for stabilization. Instead, in hGBP1 the