nature structural & molecular biology

Perfect seed pairing is not a generally reliable predictor for miRNA-target interactions

Dominic Didiano & Oliver Hobert

We use *Caenorhabditis elegans* to test proposed general rules for microRNA (miRNA)-target interactions. We show that G•U base pairing is tolerated in the 'seed' region of the *lsy-6* miRNA interaction with its *in vivo* target *cog-1*, and that 6- to 8-base-pair perfect seed pairing is not a generally reliable predictor for an interaction of *lsy-6* with a 3′ untranslated region (UTR). Rather, *lsy-6* can functionally interact with its target site only in specific 3′ UTR contexts. Our findings illustrate the difficulty of establishing generalizable rules of miRNA-target interactions.

Animal miRNAs constitute a large family of gene regulatory molecules that control the expression of target genes by binding their 3′ UTRs[1,2]. However, few *in vivo* targets of miRNAs are currently known[3,4]. Apart from a few cases where target genes have been identified by genetic loss-of-function approaches[1,2,5–7], attempts at identifying miRNA targets have largely focused on bioinformatic searches for target genes. These searches rely on the phylogenetically conserved complementarity of miRNAs to their presumptive target genes[8]. Experimental validations of predicted miRNA-target interactions have largely been conducted using cell culture transfection assays and *in vivo* over- and/or misexpression approaches (see, for example, refs. 9–11). Moreover, features of miRNA-target interactions have been deduced using synthetic miRNAs, synthetic 3′ UTR target constructs or both (see, for example, refs. 9,10,12). One theme that seems to be emerging from these approaches is that conserved perfect 6- to 8-base-pair (bp) matches at the 5′ end of predicted miRNA-target heteroduplexes ('seed' matches) are a reliable predictor of miRNA-target interactions[10,11,13–15]. The notion that any gene that contains a conserved seed match in its 3′ UTR to one of the many known animal miRNAs is likely to be regulated by that miRNA has led to the prediction of a large number of miRNA targets in animal genomes[10,11,13] and to global models concerning the evolution and function of miRNA-target interactions[16–18].

To reliably predict miRNA-target interactions, it is important to test whether rules of miRNA-target interactions, deduced through analysis of several test cases, are generally applicable to most or all miRNAs. A widely used validation strategy for predicted miRNA-target interactions assesses the effect of a heterologously expressed miRNA on a 3′ UTR fused to a reporter gene construct[9–12]. A shortcoming of these and other systems is that they rely on the ectopic, nonphysiological expression of an miRNA, which creates a situation in which two molecules with complementary surfaces may engage in a nonphysiological interaction. Analagous to the dependence of many regulatory interactions on cellular context, the expression of the 3′ UTR and the tested miRNA in a heterologous context may also lead to nonphysiological interactions because of an incorrect cofactor environment. Therefore, at least in theory, most previously described target validation strategies harbor the danger of incorrectly estimating the number of miRNA-target interactions occurring in a physiological context.

We have established a sensor system that relies on the presence of an endogenous miRNA, *lsy-6* (ref. 6). In this system, a green fluorescent protein (*gfp*)-based sensor construct is expressed under the control of a heterologous, cell type–specific promoter (*ceh-36^{prom}::gfp*) in four head neurons of the nematode *C. elegans* (**Fig. 1a**). These neurons include the ASEL and ASER neurons, two closely related, bilaterally symmetric neurons. Only the ASEL and not the ASER neuron expresses endogenous *lsy-6* miRNA, which therefore downregulates the expression of its target gene, *cog-1*, in ASEL but not ASER[6]. This regulation can be observed by fusing the 3′ UTR of the *lsy-6* target *cog-1* to the *ceh-36^{prom}::gfp* sensor construct. In contrast to control 3′ UTR fusions, the *cog-1*–3′ UTR fusion is effectively downregulated in ASEL but not in ASER (**Fig. 1b**,**e**,**f**). This downregulation depends on *lsy-6* (ref. 6) and on the *lsy-6*–binding site, which was inferred from two mutations, one that removes the *lsy-6*–complementary site[6] and one that introduces a point mutation in the seed region (**Fig. 1c**). This system therefore effectively monitors the ability of endogenous rather than mis- or overexpressed *lsy-6* to regulate its target in its normal cellular context. Moreover, the system is also internally controlled, as the expression levels of the reporter gene are compared in the ASEL and ASER neurons within a single animal.

We used this system to ask whether two widely used rules that predict miRNA targets apply to *lsy-6* as well. The first was the 'G•U rule': largely on the basis of experiments with artificially engineered miRNA-target pairs, it has been postulated that G•U 'wobble' base pairs in the seed are detrimental for miRNA-target interactions[10,12]. As the *lsy-6*–*cog-1* heteroduplex contains multiple A-U base pairs, we were able to introduce several G•U wobbles at various positions in the seed region by altering solely the sequence of the sensor construct. The introduction of five individual G•U wobbles generated 3′ UTRs that are still efficiently downregulated by endogenous *lsy-6* in ASEL (**Fig. 2a**). 3′ UTRs that contain two G•U wobbles, either at positions 2 and 6 or at positions 6 and 8, are also still efficiently downregulated in ASEL (**Fig. 2a**). Together with the existence of a G•U wobble in the seed region of a functional *let-7*–binding site in the *lin-41* 3′ UTR[19], our results indicate that G•U wobbles may not be generally detrimental to miRNA-target interactions.

Department of Biochemistry and Molecular Biophysics, Columbia University Medical Center, Howard Hughes Medical Institute, 701 W. 168th Street, New York, New York 10032, USA. Correspondence should be addressed to O.H. (or38@columbia.edu).

Received 12 June; accepted 25 July; published online 20 August 2006; doi:10.1038/nsmb1138

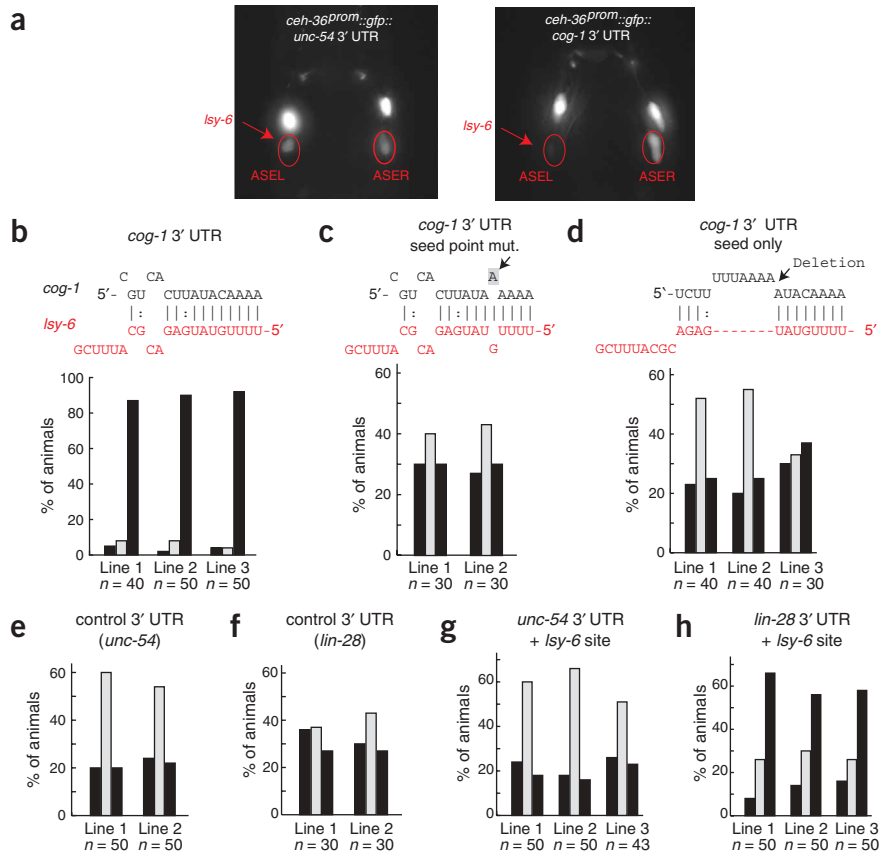

Figure 1 Responsiveness of various 3′ UTRs to *lsy-6*. (**a**) Representative animals that express a sensor with either a control 3′ UTR or the *cog-1* 3′ UTR. The other two cells that express *gfp* are the AWCL and AWCR neurons, which serve as internal controls to eliminate mosaic animals. (**b**–**h**) Quantification of sensor data, as indicated. Each group of three bars represents a single transgenic line. Left bar, greater *gfp* intensity in ASEL than ASER (ASEL > ASER); center bar, ASEL = ASER; right bar, ASER > ASEL. Scoring of the *gfp* signal is very sensitive, as it picks up substantial noise in animals that express control 3′ UTRs (*unc-54* 3′ UTR, *lin-28* 3′ UTR, *cog-1* seed-mutated 3′ UTR), reflected by a substantial number of animals with slight differences in *gfp* intensity between the two neurons, both ASEL > ASER and ASER > ASEL. To eliminate this noise, regulation of a 3′ UTR has to be assessed by comparing numbers of ASEL > ASER and ASER > ASEL animals. No substantial difference indicates the 3′ UTR is not regulated by endogenous, ASEL-expressed *lsy-6*; a substantially lower number of ASEL > ASER animals than ASER > ASEL (for example, *cog-1* wild-type 3′ UTR) indicates the 3′ UTR is regulated by *lsy-6*. Where regulation of the 3′ UTR occurs, the difference between *gfp* signals in ASEL versus ASER is also greater (substantial *gfp* signal in ASER, but little to none in ASEL), whereas in unregulated 3′ UTRs, differences between *gfp* signals in ASEL versus ASER are minor (data not shown). **Supplementary Methods** and **Supplementary Table 1** online have details on scoring and DNA constructs.

The second rule we examined was the 'seed rule': 3′ UTRs that contain a perfect 6- to 8-bp match to the 5′ end of an miRNA (usually starting at position –1 of the miRNA) are generally expected to be regulated by this miRNA. We tested the general applicability of this rule in a variety of ways. First, making use of published prediction algorithms, we searched the *C. elegans* genome for 3′ UTRs that contain seed matches to *lsy-6* (**Supplementary Methods** online)[8]. We tested 13 putative target 3′ UTRs that contain 6- to 12-bp seed matches to *lsy-6* (**Fig. 2b** and **Supplementary Fig. 1** online). Each of these sites is phylogenetically conserved (**Supplementary Fig. 2** online). In contrast to our positive control (*cog-1* 3′ UTR, a genetically validated target of *lsy-6*), none of these 13 3′ UTRs are efficiently downregulated by *lsy-6* in our sensor system (**Fig. 2b** and **Supplementary Fig. 1**). The predicted *lsy-6* targets may be regulated by *lsy-6* in some specific cellular contexts, but they are not regulated as efficiently by *lsy-6* in the ASEL neuron as *cog-1* is regulated.

Next, we asked whether the seed region of the *lsy-6* target site within the *cog-1* 3′ UTR, which we found to be required for *lsy-6*–mediated downregulation (**Fig. 1c**), is sufficient to confer downregulation of *cog-1*. The result of a 21-bp deletion in the *cog-1* 3′ UTR that disrupts the 3′ end pairing of *lsy-6* to its site, but maintains its seed match, reveals that the seed region alone is not sufficient to confer efficient downregulation (**Fig. 1d**).

We furthermore tested whether the entire *lsy-6* site is sufficient to confer *lsy-6*–mediated downregulation. Fusing the site into a heterologous 3′ UTR, the *unc-54* 3′ UTR, we found that this synthetic 3′ UTR is not downregulated by *lsy-6* (**Fig. 1e**,**g**). The presence of a *lsy-6* site is therefore not sufficient to confer the efficient downregulation of a 3′ UTR. However, when fused into the 3′ UTR of a gene, *lin-28*, that is normally regulated by miRNAs[20] but shares no overt similarity with the *cog-1* 3′ UTR, we found that the *lsy-6* site confers downregulation of the sensor (**Fig. 1f**,**h**). We conclude that features independent of the *lsy-6* site are required for miRNA-mediated control of a 3′ UTR. The 3′ UTR context dependence of the *lsy-6* site is reminiscent of the observation that *let-7* miRNA sites confer downregulation of the *lin-41* 3′ UTR only in a specific sequence context that is independent of the miRNA-complementary sites[19].

In conclusion, the presence of a well-matched miRNA target site in a 3′ UTR alone is not a generally reliable predictor for the mRNA indeed being a target of this miRNA. Many predicted miRNA-target interactions, particularly if they involve phylogenetically conserved miRNA target sites, may occur under some circumstances, but our study suggests that the coexpression of a miRNA and a predicted, seed-matched target does not guarantee a functional interaction. A similar concept applies to transcription factors: a predicted DNA-binding site in a genomic DNA sequence is not a reliable predictor for a functional interaction of the site with its cognate transcription factor (see, for example, ref. 21). Rather, cofactors and the chromatin state–dependent accessibility of the binding site determine whether the site is occupied by its cognate transcription factor. In keeping with this analogy, additional factors, such as sequence-dependent 3′ UTR accessibility and/or specific RNA- or protein-based cofactors, may be major determinants of 3′ UTR responsiveness to a seed-matched miRNA. As currently available miRNA-target prediction algorithms do not take these issues into account, it is conceivable that many proposed miRNA-target interactions occur only in very specific contexts. Hence, the general applicability of proposed

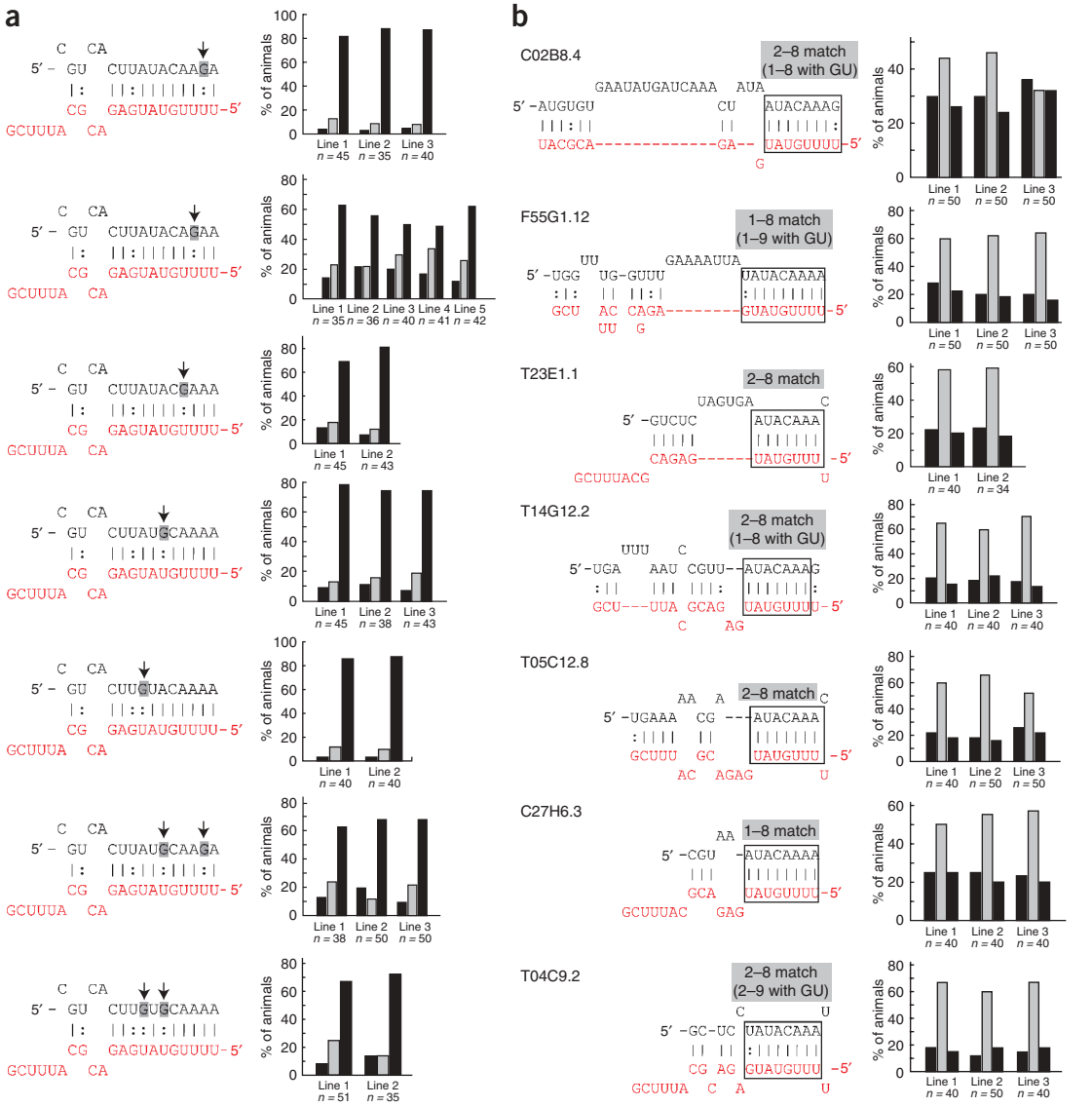

Figure 2 *lsy-6* responsiveness of G•U-wobbled 3′ UTRs and predicted *lsy-6* target 3′ UTRs, represented as in **Figure 1**. (**a**) Compared to the wild-type control (**Fig. 1b**), *cog-1* 3′ UTRs with G•U wobbles are still efficiently downregulated in ASEL. (**b**) Predicted genomic 3′ UTRs with matches (boxed) to the *lsy-6* miRNA (red). Structures of microRNA-target heteroduplexes were predicted with RNAhybrid[22]. None of the sensors showed downregulation of *gfp* in ASEL comparable to the *cog-1* positive control (**Fig. 1a**); rather, all appear similar to unregulated control 3′ UTRs (**Fig. 1e**,**f**). Data for an additional set of six predicted target 3′ UTRs is shown in **Supplementary Figure 1**.

concepts of miRNA function that are based on the genome-wide prediction of seed-based miRNA-target interactions[18] needs to be assessed cautiously.

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

ACKNOWLEDGMENTS

We thank Q. Chen for expert technical assistance in generating transgenic strains, N. Rajewsky for stimulating discussions, I. Rigoutsos for suggestions about target predictions and R. Mann, F. Slack, I. Greenwald, P. Sengupta, I. Rigoutsos and N. Rajewsky for comments on the manuscript. This work was funded, in part, by US National Institutes of Health grants 1R03HD050334-01 and 5R01NS050266-02, a predoctoral US National Institutes of Health fellowship to D.D. and the Howard Hughes Medical Institute.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Published online at http://www.nature.com/nsmb/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Lee, R.C., Feinbaum, R.L. & Ambros, V. *Cell* **75**, 843–854 (1993).
2. Wightman, B., Ha, I. & Ruvkun, G. *Cell* **75**, 855–862 (1993).
3. Ambros, V. *Nature* **431**, 350–355 (2004).
4. Carthew, R.W. *Curr. Opin. Genet. Dev.* **16**, 203–208 (2006).
5. Ambros, V. & Horvitz, H.R. *Science* **226**, 409–416 (1984).
6. Johnston, R.J. & Hobert, O. *Nature* **426**, 845–849 (2003).
7. Slack, F.J. *et al. Mol. Cell* **5**, 659–669 (2000).
8. Rajewsky, N. *Nat. Genet.* **38** Suppl. 1, S8–S13 (2006).
9. Lewis, B.P., Shih, I.H., Jones-Rhoades, M.W., Bartel, D.P. & Burge, C.B. *Cell* **115**, 787–798 (2003).
10. Brennecke, J., Stark, A., Russell, R.B. & Cohen, S.M. *PLoS Biol.* **3**, e85 (2005).
11. Krek, A. *et al. Nat. Genet.* **37**, 495–500 (2005).
12. Doench, J.G. & Sharp, P.A. *Genes Dev.* **18**, 504–511 (2004).
13. Lewis, B.P., Burge, C.B. & Bartel, D.P. *Cell* **120**, 15–20 (2005).
14. Lim, L.P. *et al. Nature* **433**, 769–773 (2005).
15. Jackson, A.L. *et al. RNA* **12**, 1179–1187 (2006).
16. Stark, A., Brennecke, J., Bushati, N., Russell, R.B. & Cohen, S.M. *Cell* **123**, 1133–1146 (2005).
17. Farh, K.K. *et al. Science* **310**, 1817–1821 (2005).
18. Hornstein, E. & Shomron, N. *Nat. Genet.* **38** Suppl. 1, S20–S24 (2006).
19. Vella, M.C., Choi, E.Y., Lin, S.Y., Reinert, K. & Slack, F.J. *Genes Dev.* **18**, 132–137 (2004).
20. Moss, E.G., Lee, R.C. & Ambros, V. *Cell* **88**, 637–646 (1997).
21. Cha-Molstad, H., Keller, D.M., Yochum, G.S., Impey, S. & Goodman, R.H. *Proc. Natl. Acad. Sci. USA* **101**, 13572–13577 (2004).
22. Rehmsmeier, M., Steffen, P., Hochsmann, M. & Giegerich, R. *RNA* **10**, 1507–1517 (2004).