Biol. Chem., Vol. 391, pp. 767–770, July 2010 • Copyright © by Walter de Gruyter • Berlin • New York. DOI 10.1515/BC.2010.090

Minireview

Decoding transcription and microRNA-mediated translation control in *Drosophila* development

Ulrike Gaul

Gene Center, University of Munich, Feodor-Lynen-Straße 25, D-81377 Munich, Germany

e-mail: gaul@genzentrum.lmu.de

Abstract

The spatio-temporal regulation of gene expression lies at the heart of animal development. In this article we present an overview of our recent work to apply systems biological approaches to the study of transcription and microRNA-mediated translation control in *Drosophila* development. We have identified many new *cis*-regulatory elements within the segmentation gene network, a transcriptional hierarchy governing pattern formation along the antero-posterior axis of the embryo, and developed a novel thermodynamic model to predict their expression. A similar thermodynamic approach that takes into account the secondary structure of the target mRNA significantly improves the prediction of microRNA binding sites.

Keywords: *cis*-regulatory element; *Drosophila*; gene regulation; microRNA; segmentation; systems biology; thermodynamic modeling.

The establishment of complex patterns of gene expression lies at the heart of animal development as well as adult homeostasis. Deciphering the regulatory code that governs spatio-temporal gene expression has thus been the focus of much research over the past few years, resulting in a range of new experimental and computational approaches (Bonn and Furlong, 2008; Celniker et al., 2009; Jaeger, 2009; Kim et al., 2009). The issue is complex and comprises many problems – from finding *cis*-regulatory elements within the DNA or RNA sequence to understanding how they 'compute' expression. In our own recent studies, we have focused on two such recognition tasks – the interaction of transcription factors with *cis*-regulatory DNA and the interaction of microRNAs with their targets in the 3′ UTRs of mRNAs.

The biological paradigm we use to study transcription control is the establishment of the segmented body pattern of the fruitfly *Drosophila*, a process that is driven by a complex hierarchically organized gene network comprising approximately 60 genes (St Johnston and Nusslein-Volhard, 1992; Pankratz and Jäckle, 1993). Most of these genes encode transcription factors, whose binding preferences and protein distribution within the embryo are known. Combinatorial binding of factors to autonomously acting *cis*-regulatory elements determines the expression patterns, even with heterologous promoters and independently of the genomic environment (Arnosti, 2003).

Our first goal was to develop tools for detecting *cis*-elements that drive patterned expression. In collaboration with physicists and computer scientists, we developed algorithms that search for local clusters of binding sites for the participating transcription factors, based on a thermodynamic model that seeks an optimal binding of factors, represented by position weight matrices, to a given sequence window (Rajewsky et al., 2002; Sinha et al., 2004). These algorithms not only recover known *cis*-elements, but also predict novel ones with excellent success, providing us with a near complete repertoire of elements for further analysis (Schroeder et al., 2004; Schroeder et al., in preparation) (Figure 1).

The much greater challenge is to understand how these segmentation *cis*-elements work, i.e., how their binding site composition determines the resulting expression patterns. To this end, we built a mathematical model that takes as input the sequence of all *cis*-elements with experimentally determined patterns, the binding preferences of all factors, and the (relative) factor concentration at each position along the antero-posterior axis of the embryo (Segal et al., 2008). We assume that factor binding occurs under thermodynamic equilibrium conditions, with each factor binding to the DNA and contributing to transcription independently. The logistic function is used to integrate the different factor contributions, and the entire Boltzmann distribution of possible legal binding configurations is sampled. Free parameters, including absolute factor concentration and the expression contribution of each factor, are fitted to maximize agreement between measured and predicted patterns (Figure 2). This approach is conceptually straightforward, and unlike previous attempts at modeling the segmentation gene network (Albert and Othmer, 2003; Jaeger et al., 2004), it captures the mechanistic core of the process. Our model predicts the expression patterns of most known *cis*-elements well, with the exception of several elements for which little or no expression is predicted owing to a lack of transcriptional activation. Validation of the model was carried out by multiple means, including predicting expression of unseen *cis*-elements from a neighboring species (*Drosophila pseudoobscura*).

A meta-analysis of the model predictions reveals a number of important insights into the architecture of segmentation *cis*-elements and the underlying regulatory logic. Both strong and weak binding sites contribute to the total occupancy of the DNA by transcription factors – approximately one-half

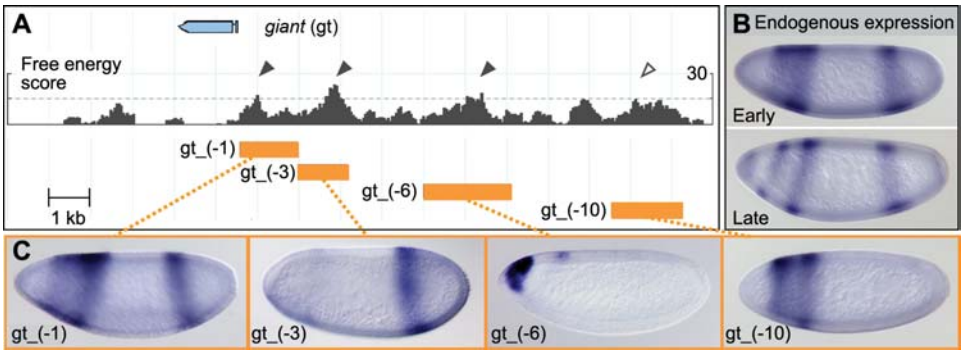

Figure 1 Identification of novel *cis*-regulatory elements through computational search for clusters of transcription factor binding sites. (A) Regulatory region of the *Drosophila* segmentation gene *giant*, with free energy score, which represents the local density and strength of binding sites for participating transcription factors, shown in dark gray and predicted *cis*-elements in orange. (C) When tested in reporter constructs, the newly predicted *cis*-elements nicely recapitulate all aspects of the endogenous gene expression (B). Modified from Schroeder et al. (2004).

is from high affinity, statistically overrepresented sites, the other half from lower affinity sites that occur no more frequently than is expected by chance. Sites for the same factor typically show clustering over a short range (0–200 bp); this is true for both activators and repressors. Such clustering facilitates cooperativity between sites, and the introduction of a (fitted) cooperativity parameter significantly improves the accuracy of our pattern predictions. By contrast, we

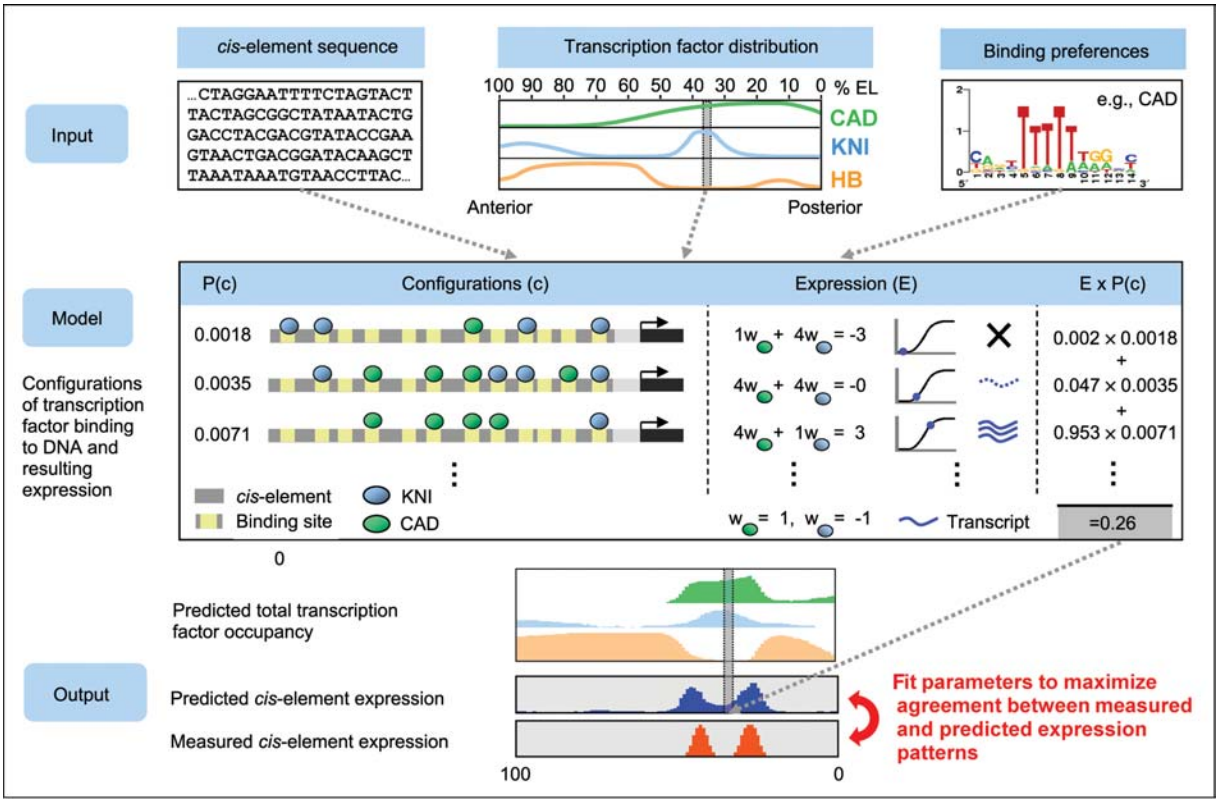

Figure 2 Thermodynamic modeling of *cis*-element expression in the *Drosophila* segmentation gene network: schematic overview of input, output and modeling approach.
For a given *cis*-element (here the element driving *even-skipped* stripes 4+6), expression at each position along the antero-posterior axis is modeled as the sum over the expression contribution (E) of each sterically possible configuration of transcription factors on the sequence (c), weighted by the probability of the configuration [P(c)], which in turn is computed from the local factor concentration, its binding preferences as represented by position weight matrices, and the *cis*-element sequence. A small number of free parameters, including the expression contribution of a given factor (w), are fit to maximize the agreement between measured and predicted expression patterns, based on a total of 44 segmentation *cis*-elements. Modified from Segal et al. (2008).

observe no heterotypic clustering of sites, suggesting that the factors do act independently from one another. These features – the presence of multiple binding sites of varying strength, self-cooperativity, and independence between factors to maximize use of the combinatorial space – explain how a small number of broadly expressed factors is able to generate the many distinct outputs that are necessary to define precise positions along the antero-posterior axis of the embryo.

Among the different mechanisms of post-transcriptional control, microRNAs represent a particularly interesting and important one. These small, genomically encoded RNAs are processed into single stranded 21–23mers and incorporated into a RNP complex (microRISC), which binds to sites primarily within 3′ UTRs to induce translational repression (Filipowicz et al., 2008). The microRNA target sites show imperfect sequence complementarity to the microRNA, with a strong match to the 5′ region ('seed') and pairing of varying extent at the 3′ end. Computational target predictions are very sensitive to the exact pairing rules and predict large numbers of potential targets within the transcriptome, which are typically reduced by introducing various types of filters, especially evolutionary conservation (Rajewsky, 2006). However, this approach is problematic, because the interaction between microRNAs and their targets is often not well conserved.

We therefore asked whether the accessibility of the target site for microRNA binding, as determined by the secondary structure of the target mRNA, might have an impact on the efficacy of microRNA-mediated repression, and could thus serve to improve computational predictions (Kertesz et al., 2007). To test this idea, we established a simple experimental paradigm, *Drosophila* S2 tissue culture cells and their endogenously expressed microRNAs, and assayed translational repression using a dual luciferase assay. Using synthetic oligonucleotides, we engineered a series of 3′ UTR constructs in which the sequence surrounding a given target site, but not the target site itself, was mutated to alter site accessibility. We found that reductions in accessibility indeed lead to proportional decreases in repression. Encouraged by these

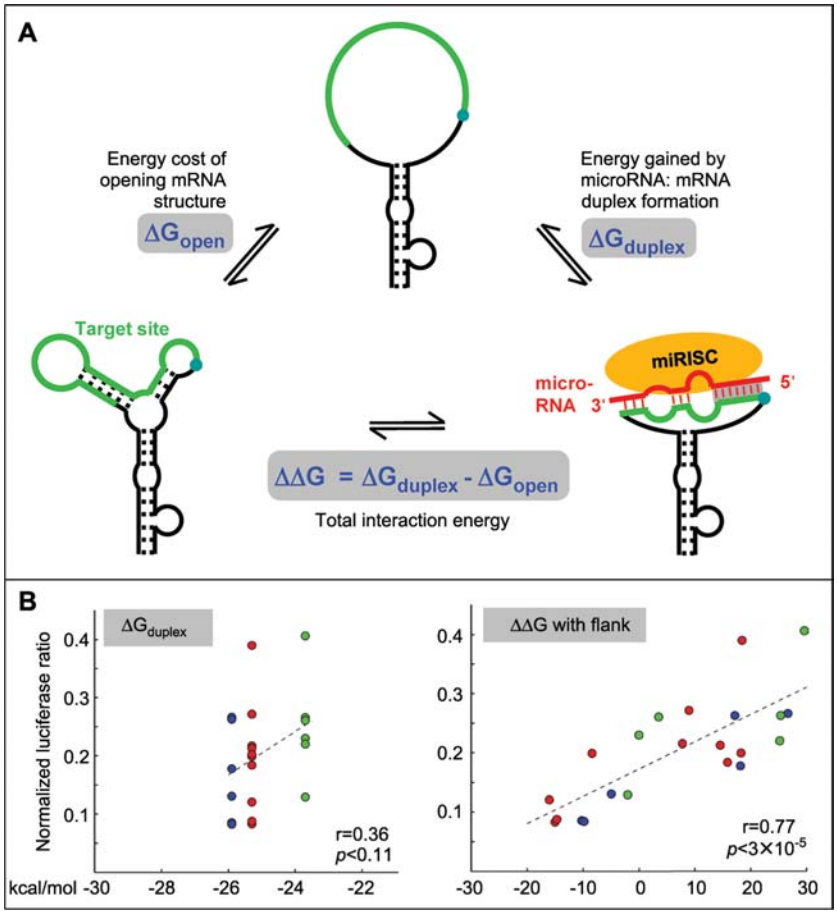

Figure 3 Thermodynamic modeling of microRNA-mRNA interactions.
(A) Schematic overview of the approach: the total interaction energy $\Delta\Delta G$ is computed as the difference between the energy gained by the formation of the microRNA-mRNA duplex (ΔG_{duplex}) and the energy lost by unpairing the target site nucleotides to make them accessible for microRNA binding (ΔG_{open}). (B) Scatterplots of experimentally measured expression levels (*y*-axis) vs. computed ΔG_{duplex} and $\Delta\Delta G$ scores, respectively (*x*-axis), for three sets of engineered 3′ UTR constructs that differ in the accessibility of the microRNA target site but not in the sequence of the site itself, and thus have identical ΔG_{duplex} scores. Regression analysis shows that $\Delta\Delta G$ is a very good predictor of observed expression values; correlation coefficient *r* and *p*-values as shown. Modified from Kertesz et al. (2007).

results, we modeled the interaction between microRNA and mRNA by computing not only the free energy gained by the duplex formation (ΔG_{duplex}), but also the energetic cost of opening the mRNA structure, using the Vienna RNA package (Hofacker, 2003). The difference between these two values, $\Delta\Delta G$, performed much better in predicting experimental outcomes for a large set of artificially altered target sites, particularly if we assume that additional bases flanking the site need to be unpaired to allow access for the large microRISC complex (Figure 3). Applied to the entire genome, we find that microRNA complementary sites are significantly overrepresented in highly accessible regions of the genome. Subsequent work suggests that an algorithm based on this approach recovers functionally important microRNA targets that are missed by other methods (Iovino et al., 2009).

Overall, our studies show that equilibrium thermodynamics provides a satisfactory description of key mechanisms in transcription and microRNA-mediated translation control. Our findings also demonstrate that combining computation and ensemble-level experimentation is a potent approach to unravel the molecular underpinnings of regulatory systems in development.

References

Albert, R. and Othmer, H.G. (2003). The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in *Drosophila melanogaster*. J. Theor. Biol. *223*, 1–18.

Arnosti, D.N. (2003). Analysis and function of transcriptional regulatory elements: insights from *Drosophila*. Annu. Rev. Entomol. *48*, 579–602.

Bonn, S. and Furlong, E.E. (2008). cis-Regulatory networks during development: a view of *Drosophila*. Curr. Opin. Genet. Dev. *18*, 513–520.

Celniker, S.E., Dillon, L.A., Gerstein, M.B., Gunsalus, K.C., Henikoff, S., Karpen, G.H., Kellis, M., Lai, E.C., Lieb, J.D., MacAlpine, D.M., et al. (2009). Unlocking the secrets of the genome. Nature *459*, 927–930.

Filipowicz, W., Bhattacharyya, S.N., and Sonenberg, N. (2008). Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? Nat. Rev. Genet. *9*, 102–114.

Hofacker, I.L. (2003). Vienna RNA secondary structure server. Nucleic Acids Res. *31*, 3429–3431.

Iovino, N., Pane, A., and Gaul, U. (2009). miR-184 has multiple roles in *Drosophila* female germline development. Dev. Cell *17*, 123–133.

Jaeger, J. (2009). Modelling the *Drosophila* embryo. Mol. Biosyst. *5*, 1549–1568.

Jaeger, J., Surkova, S., Blagov, M., Janssens, H., Kosman, D., Kozlov, K.N., Manu, Myasnikova, E., Vanario-Alonso, C.E., Samsonova, M., et al. (2004). Dynamic control of positional information in the early *Drosophila* embryo. Nature *430*, 368–371.

Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U., and Segal, E. (2007). The role of site accessibility in microRNA target recognition. Nat. Genet. *39*, 1278–1284.

Kim, H.D., Shay, T., O'Shea, E.K., and Regev, A. (2009). Transcriptional regulatory circuits: predicting numbers from alphabets. Science *325*, 429–432.

Pankratz, M. and Jäckle, H. (1993). Blastoderm segmentation. In: The Development of *Drosophila melanogaster*, M. Bate and A. Martinez Arias, eds. (Cold Spring Harbor, NY, USA: Cold Spring Harbor Laboratory Press), pp. 467–516.

Rajewsky, N. (2006). microRNA target predictions in animals. Nat. Genet. *38 (Suppl.)*, S8–S13.

Rajewsky, N., Vergassola, M., Gaul, U., and Siggia, E.D. (2002). Computational detection of genomic cis-regulatory modules applied to body patterning in the early *Drosophila* embryo. BMC Bioinformatics *3*, 30.

Schroeder, M.D., Pearce, M., Fak, J., Fan, H., Unnerstall, U., Emberly, E., Rajewsky, N., Siggia, E.D., and Gaul, U. (2004). Transcriptional control in the segmentation gene network of *Drosophila*. PLoS Biol. *2*, E271.

Segal, E., Raveh-Sadka, T., Schroeder, M., Unnerstall, U., and Gaul, U. (2008). Predicting expression patterns from regulatory sequence in *Drosophila* segmentation. Nature *451*, 535–540.

Sinha, S., Schroeder, M.D., Unnerstall, U., Gaul, U., and Siggia, E.D. (2004). Cross-species comparison significantly improves genome-wide prediction of cis-regulatory modules in *Drosophila*. BMC Bioinformatics *5*, 129.

St Johnston, D. and Nusslein-Volhard, C. (1992). The origin of pattern and polarity in the *Drosophila* embryo. Cell *68*, 201–219.

Received January 26, 2010; accepted April 20, 2010