Published in final edited form as:
J Bioinform Comput Biol. 2014 April ; 12(2): 1441009. doi:10.1142/S0219720014410091.

In silico evolution of the *hunchback* gene indicates redundancy in cis-regulatory organization and spatial gene expression

Elizaveta A. Zagrijchuk[*,§], **Marat A. Sabirov**[*,¶], **David M. Holloway**[†,||], and **Alexander V. Spirov**[*,‡,**]

[*]Lab Modeling of Evolution, I.M. Sechenov Institute of Evolutionary Physiology & Biochemistry, Russian Academy of Sciences, Thorez Pr. 44, St.-Petersburg, 2194223, Russia

[†]Mathematics Department, British Columbia Institute of Technology, 3700 Willingdon Ave. Burnaby, B.C. V5G 3H2, Canada

[‡]Computer Science and CEWIT, SUNY Stony Brook, Stony Brook, 1500 Stony Brook, Road, Stony Brook, 11794 NY, USA

Abstract

Biological development depends on the coordinated expression of genes in time and space. Developmental genes have extensive cis-regulatory regions which control their expression. These regions are organized in a modular manner, with different modules controlling expression at different times and locations. Both how modularity evolved and what function it serves are open questions. We present a computational model for the cis-regulation of the *hunchback* (*hb*) gene in the fruit fly (*Drosophila*). We simulate evolution (using an evolutionary computation approach from computer science) to find the optimal cis-regulatory arrangements for fitting experimental *hb* expression patterns. We find that the cis-regulatory region tends to readily evolve modularity. These cis-regulatory modules (CRMs) do not tend to control single spatial domains, but show a multi-CRM/multi-domain correspondence. We find that the CRM-domain correspondence seen in *Drosophila* evolves with a high probability in our model, supporting the biological relevance of the approach. The partial redundancy resulting from multi-CRM control may confer some biological robustness against corruption of regulatory sequences. The technique developed on *hb* could readily be applied to other multi-CRM developmental genes.

Keywords

Gene regulation; evolutionary computations; evolution *in silico*

[§]Zagriychuk.Elisaveta@yandex.ru
[¶]sabirov@iephb.nw.ru
[||]David_Holloway@bcit.ca
[**]Alexander.Spirov@stonybrook.edu

1. Introduction

The development of biological organisms requires the coordinated expression of numerous genes. Cis-regulatory regions of genes are critical in maintaining expression at the right levels, positions and times. Developmental processes such as segmentation of the body or formation of the limbs can involve the coordinated expression of hundreds of genes. Gene products can act as transcription factors (TFs) affecting the spatial and temporal expression of target genes, giving rise to these gene regulatory networks (GRNs).[1–3]

Cis-regulatory organization in the targets can strongly modulate the effect of TFs. Mutual co-activation, co-repression or quenching can depend on the relative proximity of bound TFs. Medium-range TF clustering can support highly synergistic transcriptional effects. At longer range, cis-regulatory modules (CRMs) tend to be present in many genes. These can be regulatory regions of hundreds to thousands of base pairs (bp), separated by much longer sequences unrelated to the target gene of interest. CRMs can operate semi-autonomously, with distinct CRMs controlling particular temporal or spatial aspects of a gene's expression. This has been well-documented in a number of cases of body segmentation and neural development in the fruit fly, *Drosophila*,[4–8] and it appears that many, if not most, developmental genes are regulated by multiple CRMs.[9]

How these long range organizational structures evolved and what function they serve are still very open questions. Increasingly, studies are finding multiple CRMs which appear to control overlapping or even identical spatio-temporal expression patterns.[10–13] Such redundancy may play a part in developmental robustness, such that an organism could survive partial corruption of its regulatory sequences (e.g. Refs. 14 and 15). Or, the redundancy may help to buffer development in extreme environments.[16] For instance, removal of one redundant CRM has been shown to produce more developmental defects in a fluctuating temperature environment.[17,18] It has also been suggested that synergy between CRMs is used to interpret broad upstream gradients of TFs.[19,20] CRM synergy has been shown during regulation of the *brk*[21] and *slp1*[22] genes in *Drosophila*, but it is likely that there are a number of mechanisms for integrating the control of multiple CRMs.

Multiple CRMs may have arisen from duplications [e.g. Ref. 23], such as is documented for coding regions of genes. This has the potential for disrupting the gene regulation controlled by the original CRM, but formation of multiple CRMs also has the potential for adding functionality while retaining the regulatory capabilities of the original CRM.

In this work, we focus on the *hunchback* (*hb*) gene in *Drosophila*, one of the first zygotic genes expressed in early body segmentation. This gene has been very well characterized, from its DNA structure, including the coding regions and several distinct CRMs, to high resolution data on the spatial and temporal expression of its mRNA and protein. Recent analysis of the *hb* regulatory region (RedFly database: http://redfly.ccr.buffalo.edu/search.php) indicates numerous CRMs. Four CRMs have been studied extensively with respect to spatial expression patterns: The proximal and distal enhancers and the oogenesis element have long been characterized; a shadow element has been found more recently[19,24] (Fig. 1). Recent Chip-Seq data reinforce the importance of the proximal, distal, and shadow enhancers in early segmentation.[24] For further information, see the Web resource HOX pro[25,26] (http://www.iephb.nw.ru/hoxpro/hunchback.html).

We develop a model for the transcription of *hb* mRNA which depends on concentrations of its known TFs and on the relative location and strength of the TF binding sites (BSs) (the model is different from other published models of *hb*[27–30]). Strength is modified by neighboring TFs, through short-range activation (cooperativity)[31] or repression (quenching).[32] BS locations and strengths are altered with an evolutionary computation (EC) approach, simulating evolution of the *hb* cis-regulatory region (see also Refs. 33–35). Spacer sites in the cis-regulatory sequence allow for the evolution of distinct CRMs. We use the model not to reproduce the fine-scale locations of the BSs, but to study how the longer-range organization of multiple CRMs for a single gene might evolve and function.

Our computations produce thousands of cis-regulatory sequences which generate the experimental *hb* mRNA expression patterns. These represent potential cis-architectures for solving the developmental patterning problem. The results fall into a very few distinct classes, which include the architecture seen in *Drosophila hb*. We find that multi-CRM architecture evolves readily, as does the multi-CRM to multiple-domain correspondence seen biologically. In particular, we find solutions with a CRM controlling the posterior stripes of the expression pattern, corresponding to the *Drosophila* distal “stripe element”.[24] Extraction of these biological features with our by multiple trajectories). We model *hb* cis-regulation, but the approach should be generally applicable to spatial gene expression controlled from multiple CRMs.

2. Methods and Approaches

The model starts from a calculation of transcriptional strength, with transcription rate depending on binding strength and the location of the BSs (Sec. 2.1, Fig. 2). Transcriptional strength is used as input to a reaction-diffusion model for *hb* transcription, decay and transport (Sec. 2.2). Solutions of this model (*hb* mRNA concentration versus spatial coordinate) are evaluated against experimental data for *hb* patterns (Sec. 2.3, Fig. 3). The TF BS strengths and locations in the *hb* cis-regulatory region are altered by simulated evolution (Sec. 2.3, Fig. 4). The presence of nonTF spacers allows separate CRMs to evolve. Multiple generations of evolution produce optimized cis-regulatory sequences for fitting the *hb* expression pattern. This generates a set of potential cis-architectures for *hb*. Characterizing classes within these, and understanding where the biological solution falls, sheds light on the evolutionary constraints of a developmental GRN at the cis-regulatory level. For a broader overview of this evolution *in silico* or evolutionary design of GRNs approach, please see Ref. 33.

2.1. Model for hb CRMs

The model incorporates TFs known to regulate *hb*: the maternal activators Bicoid (Bcd) and Caudal (Cad); and the gap repressors Krüppel (Kr), Giant (Gt), Knirps (Kni), Huckebein (Hkb), Tailless (Tll), and the head gap factor Empty spiracles (Ems). BSs for these factors in the *hb* cis-regulatory region were coded as shown in Fig. 2. Each BS is represented by two characters on the string representing the *hb* regulatory region: The first identifies the TF (the letter representations of these are shown in Fig. 2; computations used an octal representation); the second represents the binding affinity of the site (set to 1 for all BSs in this study). Transcription depends on both the TF binding strengths and their relative positions (modeling TF cross-effects). The activation strength of the cis-regulatory region is summed according to the number and strength of activator sites, the activator concentrations and co-activation effects; as well as the number, strength and quenching radii of the repressor sites, and the repressor concentrations [Fig. 2(c)]. Activation strength for each (*i*th) BS is calculated according to

$$S_i=a_iA_i+\alpha_i\left(\sum_{k=1}^{m}a_kA_k\right)-\sum_{j=1}^{l}r_jR_j, \quad (1)$$

where A_i is the local concentration of the activator with strength a_i; A_k is the local concentration of the *k*th co-activator (with co-activation coefficient α_i and strength a_k), summed over m neighboring activating BSs; and R_j is the local concentration of the *j*th repressor (with strength r_j), summed over l neighboring repressing BSs.

The cis-regulatory region was modeled as a string of length 104. Preliminary computations showed this length to be appropriate for up to three CRMs (shorter strings took longer to find good solutions; longer strings produced more redundant CRMs). The string was divided into modules if BS sequences longer than a minimum threshold MinCRM were delineated by spacers longer than the quenching radius (RadCRM). That is, CRMs were defined as sub-sequences that were independent of short-range regulation from other CRMs. Appropriate values on this string length were MinCRM = 5 and RadCRM = 3 (though larger values also worked). A maximum of three CRMs were allowed per string.

The model balances treating biologically relevant regulatory aspects, such as TF co-action, with speed of solution, in order to perform evolutionary scale simulations. For instance, rather than modeling absolute distances (in nucleotides), with an associated high computational cost, TF interactions are modeled for neighboring BSs on the string. Results from evolutionary surveys with this simplified model could lead to testing of particular cases with finer-scale (but slower) models of transcriptional regulation (e.g. the thermodynamic models[36]). In general, we find our results to be fairly robust to changes in parameters. For example, individual BS affinity is a tuneable parameter, but we find *hb* patterning can be modeled with this affinity set to 1.

2.2. Additive and selective CRM strengths

A partial differential equation (PDE) was solved for the *hb* mRNA expression pattern resulting from regulation, transcription, decay and diffusion. The strength of TF binding in the *hb* cis-regulatory region [Eq. (1)] is a factor in the overall transcription rate in the PDE. Two approaches were used for incorporating TF strengths: selective CRM action

$$\partial C/\partial t=D\partial^2C/\partial x^2+R\sigma\left(\sum S_i'-h\right)-\lambda C \quad (2)$$

and additive CRM action

$$\partial C/\partial t = D\partial^2 C/\partial x^2 + R\sigma\left[\left(\sum S_i^1 - h\right) + \left(\sum S_j^2 - h\right) + \cdots\right] - \lambda C, \quad (3)$$

where *C* is *hb* mRNA concentration, activator strength *S* is summed over *n* activator BSs in the CRMs, *D* is a diffusion coefficient, *h* represents regulatory input from ubiquitous factors, and λ is a decay coefficient. Equations (2) and (3) were solved by Euler forward-differencing in 100 cells representing the anterior-posterior (AP) axis of a fly embryo. In additive CRM action, each CRM contributes to the transcription rate (typically in equal, 1/3, measures). In selective CRM action, the strongest CRM (in a given cell) is the only one contributing to the transcription rate. σ(*x*) is a sigmoid regulation-expression function

$$\sigma(x) = \sqrt{x^2/(1+x^2)}. \quad (4)$$

2.3. Evolutionary simulations of the CRMs

A set of initial parameters [for Eqs. (2) and (3)] was chosen for each evolutionary experiment. Data on the spatial distributions of the TFs [Fig. 3(a)] were used as input to the model.

Evolution of *hb* CRMs was simulated with an EC algorithm, shown schematically in Fig. 4. Random candidate BS sequences comprised “individuals” in the initial population (36,000–54,000 individuals; Fig. 4, upper left). Individuals were evaluated according to the closeness of fit (squared differences) between the *C* pattern from Eqs. (2) or (3) and the data [Figs. 3(b) and 3(c); Fig. 4, lower left]. The best-fitting 33.33% of individuals were retained for the next generation of the evolutionary algorithm (Fig. 4, lower right). In each new generation, 20% of individuals underwent crossover operations and 80% underwent point mutation to alter the sequences (altering BSs, their positions, and the number of active CRMs; Fig. 4 upper right). Good fits to the data typically took between 100,000 and 4,000,000 generations (cycles around Fig. 4), depending on the CRM approach used (additive or selective).

Goodness of fit was evaluated over the entire AP axis, divided into four expression sub-domains observable in the mRNA data Figs. 3(b) and 3(c): domain 1, the broad far-anterior peak; sub-domain 2, a fine stripe or shoulder just posterior of domain 1; domain 3, the mid-embryo PS4 (parasegment 4) stripe; and domain 4, the posterior peak.

2.4. Model variations

We tested several variations on the model, including: (i) additive versus selective CRM action (see Sec. 2.2); (ii) whether or not *hb* self-effects were allowed (whether Hb protein acted as a self-TF or not); (iii) whether mature or immature posterior *hb* domains were fit [Figs. 3(b) and 3(c); (iv) whether Hkb, Tll or Ems BSs were included in the model, and if Ems BSs were present, whether they acted as activators or repressors. On the last item, initial results indicated that Hkb and Tll are not crucial (especially with selective CRM action), while Ems and whether it acts as an activator or repressor is important for fitting immature [Fig. 3(b)] or mature *hb* patterns [Fig. 3(c)], respectively.

3. Results

3.1. Additive CRM action

With additive CRM action, all CRMs in the *hb* regulatory region actively contribute to the transcription rate. Biologically, this corresponds to all CRMs being available for TF binding and either interacting directly with the promoter (region of the transcription start site) or cooperating with other CRMs. We found this mode of regulation to have the richest variety of biologically interesting features.

Of special interest was that for particular ranges of the R, *h* and λ parameters, [Eqs. (2) and (3); especially R = [79, 109], h = [0.1, 1.1] and λ = 1] solutions commonly included a CRM controlling the PS4 and posterior stripes [Fig. 5(a), CRM2; Fig. 5(b), CRM3).

This is analogous to the biological “stripe element”[24] and indicates that the evolutionary processes used in our model can efficiently capture the evolutionary trajectories of real genes.

3.1.1. Dependence on overall regulation strength, R—How is the overall level of expression controlled when there are redundant CRMs, i.e. with similar expression patterns? For example, if one is deleted (as in a mutant), does the overall level of expression go down (as is seen with Kr and Kni mutants[37]), or is it compensated?

In the model, the overall expression level depends on the *R* parameter. To test the effect of expression level on CRM number, we ran a series of evolutionary simulations varying *R* from 59 to 109, with λ(= 1) and *h*(= 1.1) held constant. Above R = 109, no solutions could be found. Over this range of *R*’s, we see five dominant classes of outcomes (Fig. 6). As R is raised, the proportion of solutions in different classes changes, with different dominant patterns at different R values. Below R = 69, outcomes are dominated by a set of identical patterns, which tend to express in all domains and not show the distinct anterior (classical and shadow CRMs) and posterior (stripe CRM) expression seen in *Drosophila*.

A more biological “stripe” element becomes significant for R = 79 and above. For R = [79, 89], the 3rd class of outcomes [Fig. 6(c)], the closest to the biological classical/shadow and stripe patterns, is most common. CRMs controlling one or two domains are more common at high *R*; 3-domain CRMs are more common at low *R* (Fig. 7). As *R* increases, it takes longer to fit the data profile (Table 1). This may correspond to the overall higher expression at high *R*, causing difficulty in finding solutions with moderate or low expression levels. (For example, at R = 109 the model maximum is twice that for the *hb* data [Figs. 3(b) and 3(c)]. High *R* values may bias the evolutionary search from picking up CRMs that all redundantly produce complete (or nearly complete) profiles (since each such CRM would express at 1/3 maximum, on average), favoring CRMs that control pairs or single domains.

3.2. Selective CRM action

With selective CRM action, only one of the CRMs is active in controlling the transcription rate. This occurs biologically if some CRMs are not available for TF binding or interaction with the promoter (e.g. they are tightly wrapped in the histone), or if there is competition between CRMs for interaction with the promoter (e.g. there is some steric hindrance for multiple CRM interaction with the promoter.

We computed 860 runs with a basic selective CRM action model: fitting to early (immature) posterior *hb*; no *hb* self-activation; eight input TFs (Bcd, Cad, Tll, Ems, Kr, Gt, Kni, Hkb); and Ems as a repressor. All runs successfully fit the data. Other variations with selective CRM action had similar outcomes.

Most successful solutions had either 2 or 3 CRMs (with roughly equal occurrence). Single CRM solutions producing all *hb* domains were much more rare (3–4%). Cases with one-to-one domain-CRM correspondence were very rare: 3–4% of solutions showed a single CRM controlling domain 1; and there were only two cases of a single CRM controlling domain 3 (see e.g. Sec. 3.1, *R* = 109). The two CRM solutions tend to show a recurrent motif, with one CRM expressing in the anterior domains 1/2/3 (see e.g. the classical and shadow enhancers[25]) and the other CRM expressing in domains 1/3/4 [Fig. 8(a)]. With three CRMs there is more diversity of outcomes, including a common motif in which two of the CRMs control domains 1/3, and the other CRM can be expressed in all domains [Fig. 8(B)].

Tests with other versions of the model indicate that the simpler the model (less TF BSs), the faster the evolutionary search (Table 2). Selective CRM action produces substantially faster computational searches than additive action (see e.g. Tables 1 and 2), but the results are less relevant to the *Drosophila* CRM-domain relation.

4. Discussion

Nearly all simulations, with all tested versions, returned successful fits to the data (with a very few exceptions). Most good solutions belonged to one or a few dominant classes with respect to CRM-domain correspondence.

For additive CRM action, all three allowed CRMs usually evolved. We initially thought this might reflect the particular *R* (production rate) and λ (degradation rate) parameters chosen: *R* and λ were initially set such that each CRM produced about 1/3 of the total required *hb* level. But we found that even with doubled *R*, the solutions still tended to have multiple CRMs, indicating a more general effect. For selective CRM action, we see outcomes with one, two, or three CRMs. Cases with one-to-one domain-CRM correspondence were generally rare: A few percent of solutions showed a single CRM controlling PS4 (domain 3) or the posterior domain (4) only. At some parameter limits (e.g. high *R*), we observed an increase in the formation of one-to-one correspondence. But in general each CRM tended to contribute to several domains, with each domain controlled by multiple CRMs.

4.1. Stripe element evolution is robust

The main biological conclusion from the evolution *in silico* is that the stripe element, i.e. a CRM driving expression in the PS4 and posterior peaks (domains 3 and 4), can readily evolve. This was seen in particular for the basic model with additive CRM action, over a broad range of parameters. In light of the simplicity of the gene regulation and the evolutionary processes in our model, this suggests that stripe element evolution is quite robust and reproducible, i.e. it is represented by many evolutionary trajectories and is relatively insensitive to the details of the evolutionary process. We plan to test the generality of these findings on other types of models for multiple CRM control in other genes.[38–40]

4.2. Redundant CRMs evolve readily

Across the different versions of our evolutionary simulations, partial or even nearly complete redundancy of CRMs (multiple CRMs having the same expression patterns) appears to be common. For instance, at $R = [39, 59]$ ($h = 0.1, 1.1$) the dominant case displayed nearly complete redundancy of CRM-domain correspondence, with all three CRMs controlling complete or nearly-complete pattern [Fig. 6(A)]; and several percent of the cases have two redundant CRMs controlling anterior patterning [domains 1/2/3; Fig. 5(b)].

$R = 69$($h = 0.1; 1.1$) was the richest case for redundant solutions, with more than 2/3 of all solutions having either two or three CRMs controlling complete or nearly-complete patterns, and about 1/10 of solutions having two CRMs controlling anterior (domain 1/2/3) patterning.

Redundancy appears lower at higher R. For $R = [79, 99]$($h = 0.1; 1.1$) we observed only a few percent of solutions with two CRMs redundantly controlling domains 1/2/3. At $R = 99$, we see about 5% of cases with two (not three) redundant CRMs controlling a complete or nearly-complete pattern.

Overall, the simulations indicate that evolution of multiple CRM control of the domains occurs readily, and that the CRM-domain correspondence can be repeated by multiple CRMs. We have focused on the three *Drosophila* CRMs studied with respect to segmentation patterning. The RedFly database (http://redfly.ccr.buffalo.edu/search.php) predicts many more *hb* CRMs (though with uncharacterized functionality) — our simulations indicate that such high redundancy is readily evolvable. This redundancy may contribute to robustness by making particular sequences nonessential for proper gene expression. Even with the simple model for the evolution and regulation of *hb*, the redundancy of regulatory elements is common. This likely reflects a general evolutionary trend which could be observed across developmental genes.

4.3. TFs active in the stripe element

The TFs found in our computed “stripe element” CRMs correspond to the TFBSs found experimentally. We generally find the canonical gap genes (Gt, Kni, Kr) strongly represented[24,41] as repressors, and maternal factors Bcd and Cad well represented as activators.[41,42] In many solutions, Hkb appears as a repressor, Tll as an activator, and Hb appears as a self-effector; these factors have all been reported in the stripe element.[24,41]

TF binding in the model stripe element suggests that a minimum of two activators is needed. We observe that the activators tend to pair or form small clusters — short-range homotypic clusters which are related to the cooperative interaction of the sites.[43] Repressor binding agrees with experimental conclusions, that domain 3 (PS4) is repressed by Kr and domain 4 (posterior peak) is repressed by Kni and Gt.[24] If Hb binding is allowed in the computations, it tends to be seen in good solutions, corroborating the self-binding observed experimentally.[44,45]

In general, TF cooperativity and co-action is an active area of discussion in *Drosophila* segmentation. We hope that our computed CRMs might aid in guiding experimental elucidation of these effects.

4.4. Conclusion

i. Partial or even nearly-complete redundancy of CRMs, one of the key features of the regulatory design for many developmental genes, is typical and common across the various evolutionary simulations presented here for *hb*, one of the most studied *Drosophila* segmentation genes.

ii. The key CRM responsible for mature PS4 and posterior domain formation (the "stripe element") evolves easily and reproducibly at a broad range of model parameters. CRMs controlling pairs of domains have also been reported in the pair-rule class of segmentation genes,[46] with similar design to the *hb* stripe element.[24] This suggests a general applicability of our evolutionary approach to study the cis-architecture of developmental genes.

iii. Our computations suggest that the evolution of GRNs with one-to-one CRM-domain correspondence is very rare. If evolution of multiple CRMs is possible (as with *hb*, biologically), it is much more likely to have domains which depend on several CRMs (with each CRM contributing to multiple domains).

Acknowledgments

The authors would like to thank the anonymous reviewers for their helpful and constructive comments that greatly contributed to improving the final version of the paper. This work is supported by U.S. NIH grant R01-GM072022 and The Russian Foundation for Basic Research grant 13-04-02137.

Biographies

Elizaveta A. Zagrijchuk was educated at Herzen State Pedagogical University, Saint-Petersburg (Russia). Her main field of expertise is environmental engineering. Her field of interests include genetic algorithms, evolutionary computation, computer modeling in biology, systems biology.

Marat A. Sabirov M.Sc., Junior Researcher, is interested in population dynamics modeling. After receiving his grade of specialist of Ecology in Saint-Petersburg State University he finished M.Sc. Program “Biodiversity and Nature Conservation” in the University. Within a framework of his Ph.D. project in Sechenov Institute of Evolutionary Physiology and Biochemistry (Saint-Petersburg) he is developing a data-driven model of population dynamics. Field of interests also includes evolutionary optimization, artificial life and artificial intelligence.

David M. Holloway’s research is on how the biochemical and gene regulatory networks that determine spatial positioning of cell types stay robust to external variability (in temperature, maternal dosage, size, for example) and to the noise intrinsic to chemical reactions. He is also interested in the interplay between chemical patterning and tissue growth, particularly in plant morphogenesis. Holloway’s work is supported by NSERC (Canada) and the NIH (US). He has a Ph. D. in physical chemistry from the University of British Columbia (1995) and has been in the Mathematics Department at the B.C. Institute of Technology since 1998.

Alexander V. Spirov is a research assistant professor in the Department of Computer Science at the State University of New York at Stony Brook and a senior researcher in the Sechenov Institute of Evolutionary Physiology and Biochemistry, St.-Petersburg, Russia. Dr. Spirov received M.S. degree in molecular biology in 1978 from the St.-Petersburg State University, St.-Petersburg, Russia. He received his Ph.D. in the area of biometrics in 1987 from the Irkutsk State University, Irkutsk, Russia. His research interests are in computational biology and bioinformatics, Web databases, data mining, evolutionary computations, animates, artificial life and evolutionary biology.

References

1. Levine M, Davidson EH. Gene regulatory networks for development. Proc Natl Acad Sci. 2005; 102:4936–4942. [PubMed: 15788537]
2. Erwin DH, Davidson EH. The evolution of hierarchical gene regulatory networks. Nature Rev Gen. 2009; 10:141–148.
3. Davidson EH. Emerging properties of animal gene regulatory networks. Nature. 2010; 468:911–920. [PubMed: 21164479]
4. Kuzin A, Kundu M, Ekatomatis A, Brody T, Odenwald WF. Conserved sequence block clustering and flanking inter-cluster flexibility delineate enhancers that regulate nerfin-1 expression during *Drosophila* CNS development. Gene Expr Patterns. 2009; 9:65–72. [PubMed: 19056518]
5. Fujioka M, Jaynes JB. Regulation of a duplicated locus: *Drosophila* sloppy paired is replete with functionally overlapping enhancers. Dev Biol. 2012; 362:309–319. [PubMed: 22178246]
6. Kuzin A, Kundu M, Ross J, Koizumi K, Brody T, et al. The cis-regulatory dynamics of the *Drosophila* CNS determinant castor are controlled by multiple sub-pattern enhancers. Gene Expr Patterns. 2012; 12:261–272. [PubMed: 22691242]
7. Bejerano G, Siepel AC, Kent WJ, Haussler D. Computational screening of conserved genomic DNA in search of functional noncoding elements. Nat Methods. 2005; 2:535–545. [PubMed: 16170870]
8. Kundu M, Kuzin A, Lin T-Y, Lee C-H, Brody T, et al. Cis-regulatory complexity within a large non-coding region in the *Drosophila* genome. PLoS ONE. 2013; 8:e60137. [PubMed: 23613719]
9. Hong JW, Hendrix DA, Levine MS. Shadow enhancers as a source of evolutionary novelty. Science. 2008; 321:1314. [PubMed: 18772429]
10. Werner T, Hammer A, Wahlbuhl M, Bosl MR, Wegner M. Multiple conserved regulatory elements with overlapping functions determine Sox10 expression in mouse embryogenesis. Nucl Acids Res. 2007; 35:6526–6538. [PubMed: 17897962]

11. Zeitlinger J, Zinzen RP, Stark A, Kellis M, Zhang H, Young RA, Levine M. Whole-genome ChIP-chip analysis of Dorsal, Twist, and Snail suggests integration of diverse patterning processes in the *Drosophila* embryo. Genes Dev. 2007; 21:385–390. [PubMed: 17322397]

12. Jeong Y, El-Jaick K, Roessler E, Muenke M, Epstein DJ. A functional screen for sonic hedgehog regulatory elements across a 1Mb interval identifies long-range ventral fore-brain enhancers. Development. 2006; 133:761–772. [PubMed: 16407397]

13. Spitz F, Furlong EM. Transcription factors: From enhancer binding to developmental control. Nat Rev Genet. 2012; 13:613–626. [PubMed: 22868264]

14. Frankel N, Erezyilmaz DF, McGregor AP, Wang S, Payre F, Stern DL. Phenotypic robustness conferred by apparently redundant transcriptional enhancers. Nature. 2010; 466 (7305):490–493. [PubMed: 20512118]

15. Cretekos W, Eric DG, et al. Regulatory divergence modifies limb length between mammals. Genes Dev. 2008; 22:141–151. [PubMed: 18198333]

16. Barolo S. Shadow enhancers: Frequently asked questions about distributed cis-regulatory information and enhancer redundancy. Bioessays. 2012; 34:135–141. [PubMed: 22083793]

17. Frankel N, Davis GK, Vargas D, Wang S, Payre F, Stern DL. Phenotypic robustness conferred by apparently redundant transcriptional enhancers. Nature. 2010; 466:490–493. [PubMed: 20512118]

18. Perry MW, Boettiger AN, Bothma JP, Levine M. Shadow enhancers foster robustness of *Drosophila* gastrulation. Curr Biol. 2010; 20:1562–1567. [PubMed: 20797865]

19. Perry MW, Boettiger AN, Levine M. Multiple enhancers ensure precision of gap gene-expression patterns in the *Drosophila* embryo. Proc Natl Acad Sci USA. 2011; 108:13570–13575. [PubMed: 21825127]

20. Garnett AT, Square TA, Medeiros DM. BMP, Wnt and FGF signals are integrated through evolutionarily conserved enhancers to achieve robust expression of Pax3 and Zic genes at the zebrafish neural plate border. Development. 2012; 139:4220–4231. [PubMed: 23034628]

21. Yao L-C, Phin S, Cho J, Rushlow C, Arora K, Warrior R. Multiple modular promoter elements drive graded brinker expression in response to the Dpp morphogen gradient. Development. 2008; 135:2183–2192. [PubMed: 18506030]

22. Prazak L, Fujioka M, Gergen JP. Non-additive interactions involving two distinct elements mediate sloppy-paired regulation by pair-rule transcription factors. Dev Biol. 2010; 344:1048–1059. [PubMed: 20435028]

23. Jeong S, Rebeiz M, Andolfatto P, Werner T, True J, Carroll SB. The evolution of gene regulation underlies a morphological difference between two *Drosophila* sister species. Cell. 2008; 132:783–793. [PubMed: 18329365]

24. Perry MW, Bothma JP, Luu RD, Levine M. Precision of Hunchback expression in the *Drosophila* embryo. Curr Biol. 2012; 22:1–6. [PubMed: 22197242]

25. Spirov AV, Bowler T, Reinitz J. HOX Pro: A specialized data base for clusters and networks of homeobox genes. Nucl Acids Res. 2000; 28:337–340. [PubMed: 10592267]

26. Spirov AV, Borovsky M, Spirova OA. HOX Pro DB: The functional genomics of hox ensembles. Nucl Acids Res. 2002; 30:351–353. [PubMed: 11752335]

27. Papatsenko D, Levine MS. Dual regulation by the Hunchback gradient in the Drosophila embryo. PNAS. 2008; 105(8):2901–2906. [PubMed: 18287046]

28. Manu SS, Spirov AV, Gursky VV, Janssens H, et al. Canalization of gene expression in the Drosophila blastoderm by gap gene cross regulation. PLoS Biol. 2009; 7(3):e1000049. [PubMed: 19750121]

29. Papatsenko D, Levine M. The Drosophila gap gene network is composed of two parallel toggle switches. PLoS ONE. 2011; 6(7):e21145. [PubMed: 21747931]

30. Holloway DM, Lopes FJP, da Fontoura Costa L, Travençolo B, Golyandina N, et al. Gene expression noise in spatial patterning: Hunchback promoter structure affects noise amplitude and distribution in Drosophila segmentation. PLoS Comput Biol. 2011; 7(2):e1001069. [PubMed: 21304932]

31. Ma XG, Yuan D, Diepold K, Scarborough T, Ma J. The Drosophila morphogenetic protein Bicoid binds DNA cooperatively. Development. 1996; 122:1195–1206. [PubMed: 8620846]

32. Hewitt GF, Strunk B, Margulies C, Priputin T, Wang XD, Amey R, Pabst B, Kosman D, Reinitz J, Arnosti DN. Transcriptional repression by the Drosophila giant protein: Cis element positioning provides an alternative means of interpreting an effector gradient. Development. 1999; 126:1201–1210. [PubMed: 10021339]

33. Spirov A, Holloway D. Using evolutionary computations to understand the design and evolution of gene and cell regulatory networks. Methods. 2013; 62:39–55. [PubMed: 23726941]

34. Spirov, AV.; Holloway, DM. New approaches to designing genes by evolution in the computer. In: Roeva, O., editor. Real-World Applications of Genetic Algorithms. InTech; 2012. p. 235-260.Available from: http://www.intechopen.com/books/real-world-applications-of-genetic-algorithms/new-approaches-to-designing-genes-by-evolution-in-the-computer

35. Spirov AV, Holloway DM. Evolution in silico of genes with multiple regulatory modules on the example of the *Drosophila* segmentation gene *hunchback*. IEEE Sympos Computational Intelligence and Computational Biology, CIBCB 2012. 2012:244–251.

36. Dresch JM, Liu X, Arnosti DN, Ay A. Thermodynamic modeling of transcription: Sensitivity analysis differentiates biological mechanism from mathematical model-induced effects. BMC Syst Biol. 2010; 4:142. [PubMed: 20969803]

37. Surkova S, Golubkova E, Manu, Panok L, Mamon L, Reinitz J, Samsonova M. Quantitative dynamics and increased variability of segmentation gene expression in the Drosophila Krüppel and knirps mutants. Dev Biol. 2013; 376:99–112. [PubMed: 23333947]

38. Janssens H, Hou S, Jaeger J, Kim AR, Myasnikova E, Sharp D, Reinitz J. Quantitative and predictive model of transcriptional control of the *Drosophila melanogaster* even skipped gene. Nature Gen. 2006; 38:1159–1165.

39. Martinez CA, Barr K, Kim AR, Reinitz J. A synthetic biology approach to the development of transcriptional regulatory models and custom enhancer design. Methods. 2013; 62:91–98. [PubMed: 23732772]

40. He X, Samee MAH, Blatti C, Sinha S. Thermodynamics-based models of transcriptional regulation by enhancers: The roles of synergistic activation, cooperative binding and short-range repression. PLoS Comput Biol. 2010; 6:e1000935. [PubMed: 20862354]

41. Li XY, MacArthur S, Bourgon R, Nix D, Pollard DA, Iyer VN, Hechmer A, Simirenko L, Stapleton M, Luengo Hendriks CL, et al. Transcription factors bind thousands of active and inactive regions in the Drosophila blastoderm. PLoS Biol. 2008; 6:e27. [PubMed: 18271625]

42. Berman BP, Nibu Y, Pfeiffer BD, Tomancak P, Celniker SE, Levine M, Rubin GM, Eisen MB. Exploiting transcription factor binding site clustering to identify cis-regulatory modules involved in pattern formation in the Drosophila genome. Proc Natl Acad Sci USA. 2002; 99(2):757–762. [PubMed: 11805330]

43. He X, Duque TS, Sinha S. Evolutionary origins of transcription factor binding site clusters. Mol Biol Evol. 2012; 29:1059–1070. [PubMed: 22075113]

44. Treisman J, Desplan C. The products of the Drosophila gap genes Hunchback and krüppel bind to the Hunchback promoters. Nature. 1989; 341:335–337. [PubMed: 2797150]

45. Margolis JS, Borowsky ML, Steingrimsson E, Shim GW, Lengyel JA, et al. Posterior stripe expression of Hunchback is driven from 2 promoters by a common enhancer element. Development. 1995; 121:3067–3077. [PubMed: 7555732]

46. Schroeder MD, Greer C, Gaul U. How to make stripes: Deciphering the transition from non-periodic to periodic patterns in Drosophila segmentation. Development. 2011; 138:3067–3078. [PubMed: 21693522]

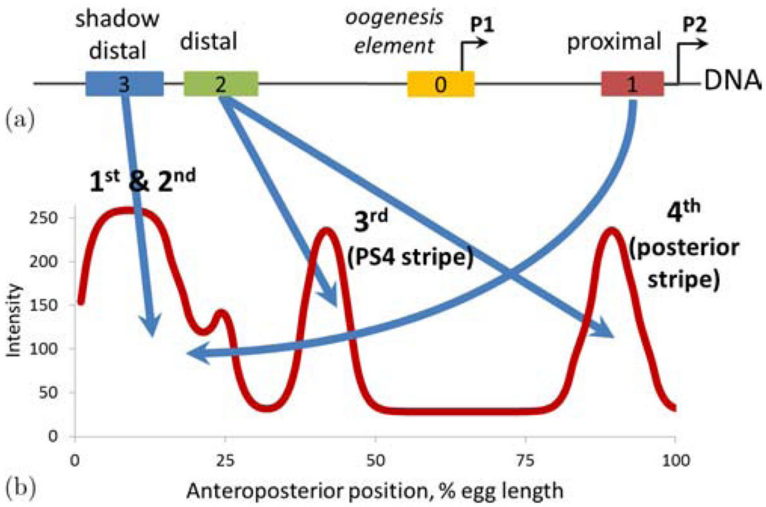


Fig. 1.

Organization, pattern and regulation of the *Drosophila* segmentation gene *hb*. (a) Organization of the *hb* regulatory region, with four separate autonomous regulatory elements (CRMs). (b) Mature *hb* expression pattern with four sub-domains in the early fruit fly embryo: One-dimensional spatial expression profile of fluorescence intensity (proportional to mRNA concentration) plotted along the main head-to-tail embryo axis (data from BID DB, BDTNP: http://bdtnp.lbl.gov/Fly-Net/bioimaging.jsp). Blue arrows indicate that each CRM is expressed in distinct sub-domains.

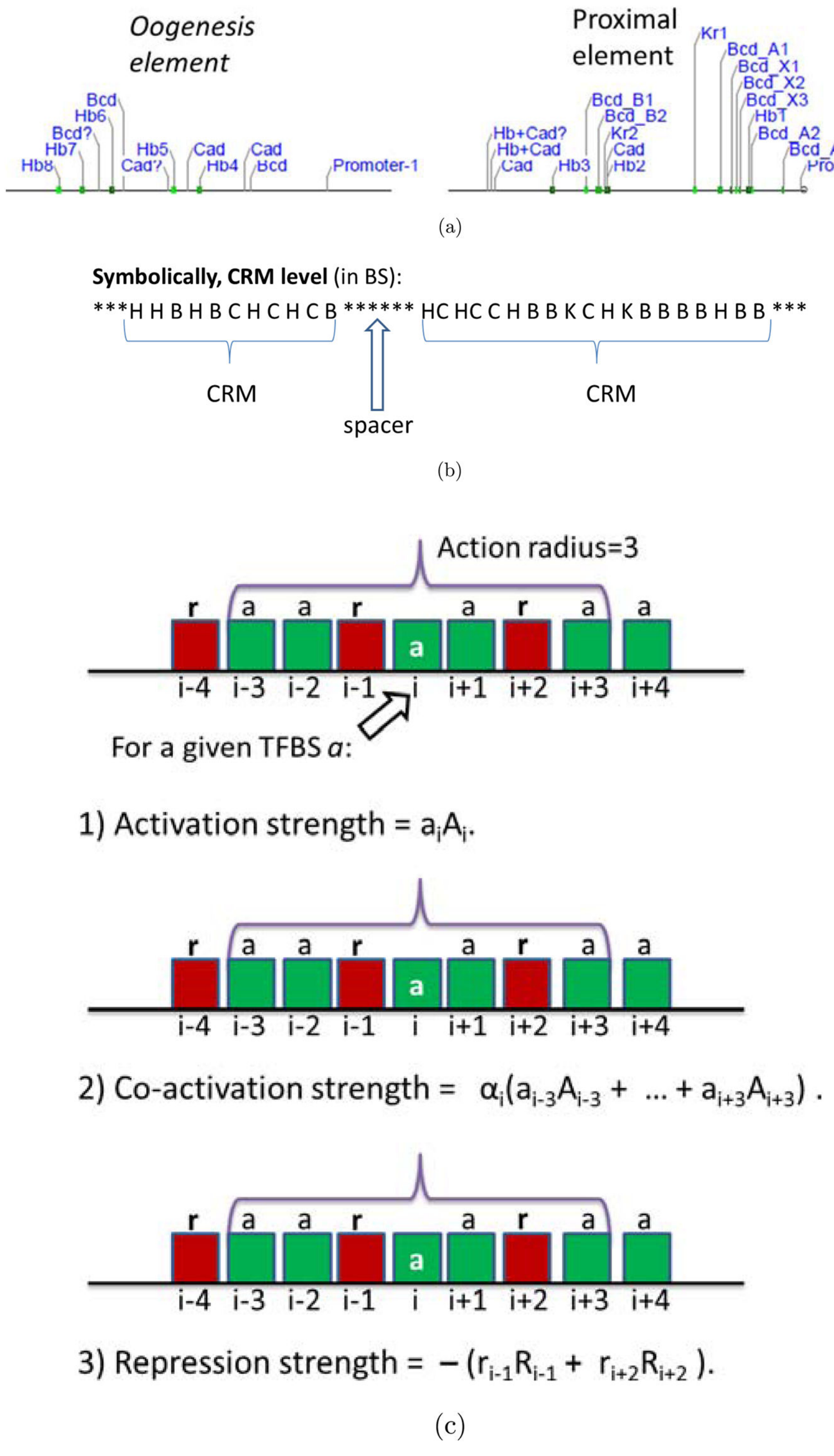
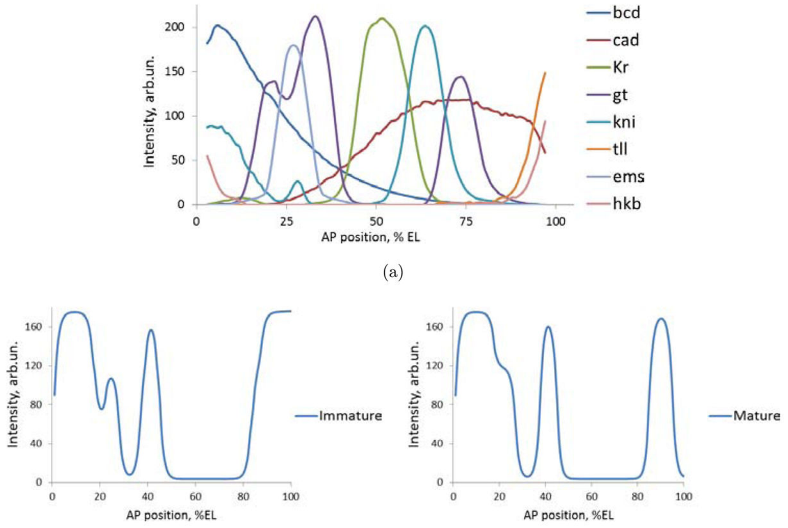

Oogenesis element
Hb8
Hb7
Bcd?
Hb6
Bcd
Hb5
Cad?
Cad
Hb4
Cad
Bcd
Promoter-1
Proximal element
Hb+Cad?
Hb+Cad
Cad
Hb3
Bcd_B1
Bcd_B2
Kr2
Cad
Hb2
Kr1
Bcd_A1
Bcd_X1
Bcd_X2
Bcd_X3
HbT
Bcd_A2
Bcd_A3
Promoter-2
(a)
Symbolically, CRM level (in BS):
H H B H B C H C H C B *** HC HC C H B B K C H K B B B B H B B ***
CRM
spacer
CRM
(b)
Action radius=3
r a a r a a r a a
i-4 i-3 i-2 i-1 i i+1 i+2 i+3 i+4
For a given TFBS a:
1) Activation strength = a_iA_i.
r a a r a a r a a
i-4 i-3 i-2 i-1 i i+1 i+2 i+3 i+4
2) Co-activation strength = $\alpha_i(a_{i-3}A_{i-3} + \ldots + a_{i+3}A_{i+3})$.
r a a r a a r a a
i-4 i-3 i-2 i-1 i i+1 i+2 i+3 i+4
3) Repression strength = $-(r_{i-1}R_{i-1} + r_{i+2}R_{i+2})$.
(c)

Fig. 2.

Representation of *hb* regulatory region. (a) Schematic of two of the four *hb* CRMs, showing BSs for specific TFs (visualized with *Genamics Expression* software). (b) Representation of the TFBS (Bcd – B, Cad – C, Ems – E, Gt – G, Hb – H, Kr – K, Kni – N, Tll – T, etc.) as strings of characters, neglecting distance along the DNA between BSs. Asterisks denote spacers (nonBS DNA) separating CRMs. (c) Strength of an activating BS is calculated by a 3-step algorithm which sums activation (including co-activation) and repression (quenching) strengths dependent on neighboring TFs (a short action radius of three BSs is used): (1) local activation strength is tallied; (2) neighboring activation is added; (3) neighboring repression is added.



(a)

(b)

(c)

Fig. 3.

Data used for evolutionary model fitting. (a) Experimental spatial patterns of hb regulatory TFs in mid nuclear cleavage (NC) cycle 14, from FlyEx (urchin.spbcas.ru/flyex). Vertical axis: fluorescence intensity, proportional to protein concentration; horizontal axis, AP position, in percent egg length (%EL) Early, immature (b) and late, mature (c) NC14 AP "target" profile used to test goodness of fit for the evolutionary computations (ECs) (derived from averaged *hb* mRNA data from BID BDTNP [http://bdtnp.lbl.gov/Fly-Net/bidatlas.jsp]).

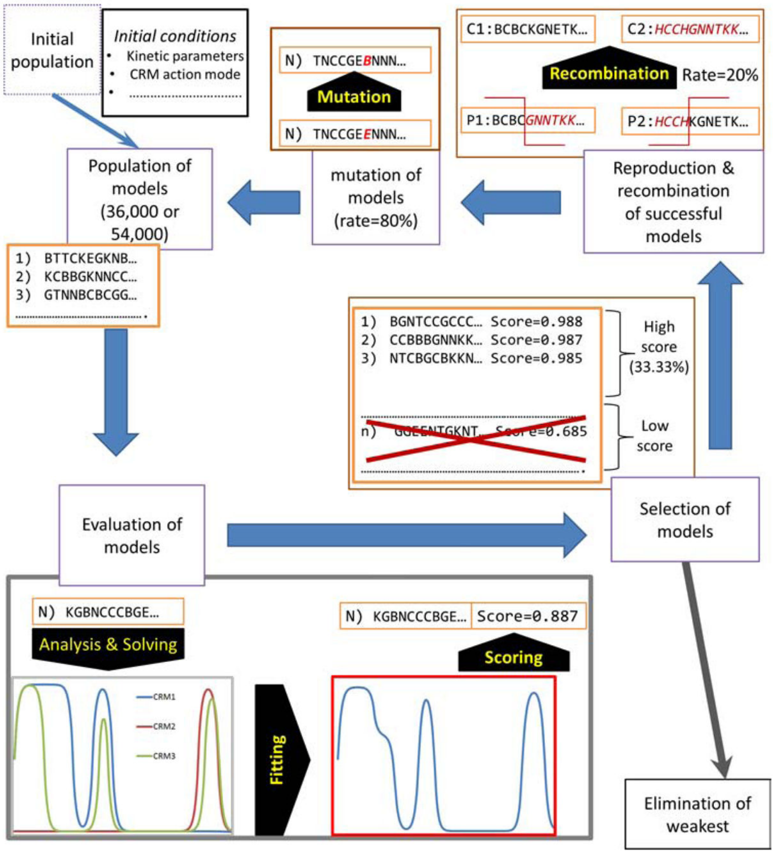


Fig. 4.

A cycle of the evolutionary simulation. Top left: Initial conditions are applied to an initial population of individuals (strings of random BS sequences). These sequences are used to solve spatial patterns according to Eqs. (1)–(3) (lower left). Computed patterns are scored for fit to experimental data (lower center). Lowest scoring individuals are selected out of the population (lower right). The remaining individuals undergo mutation and crossover (upper right), establishing the population for the next generation (iteration of the evolutionary cycle).

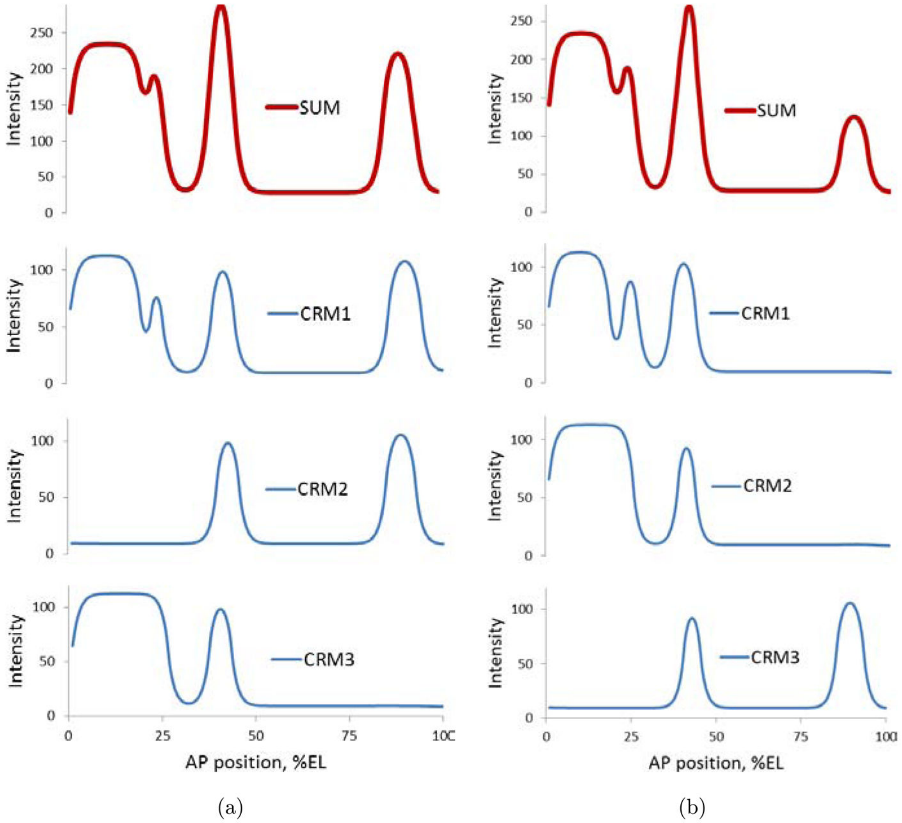


Fig. 5.

Evolutionary simulation results with multiple *hb* CRMs. (a) Predominant expression patterns for each CRM at low R. (b) A case with clearer AP distinction between the CRMs, as seen biologically.

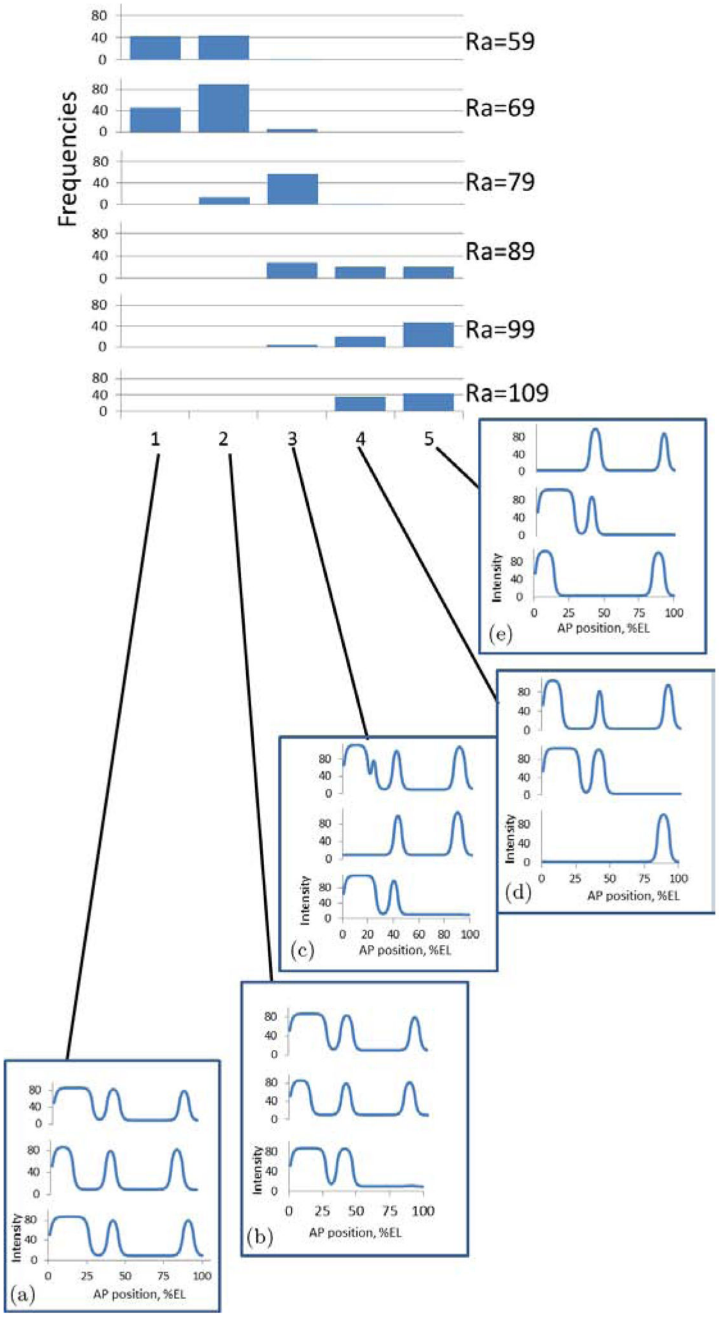


Fig. 6.

Frequencies of the main classes of successful solutions change as *R* is varied. Histograms (top) of common dominant classes, illustrated in A–E. As *R* is increased, solutions shift from completely redundant (Class A, all three CRMs each control all domains), to increasing diversity (with a prevalence of CRMs controlling pairs of domains), and the emergence of single-domain CRMs (Class D).

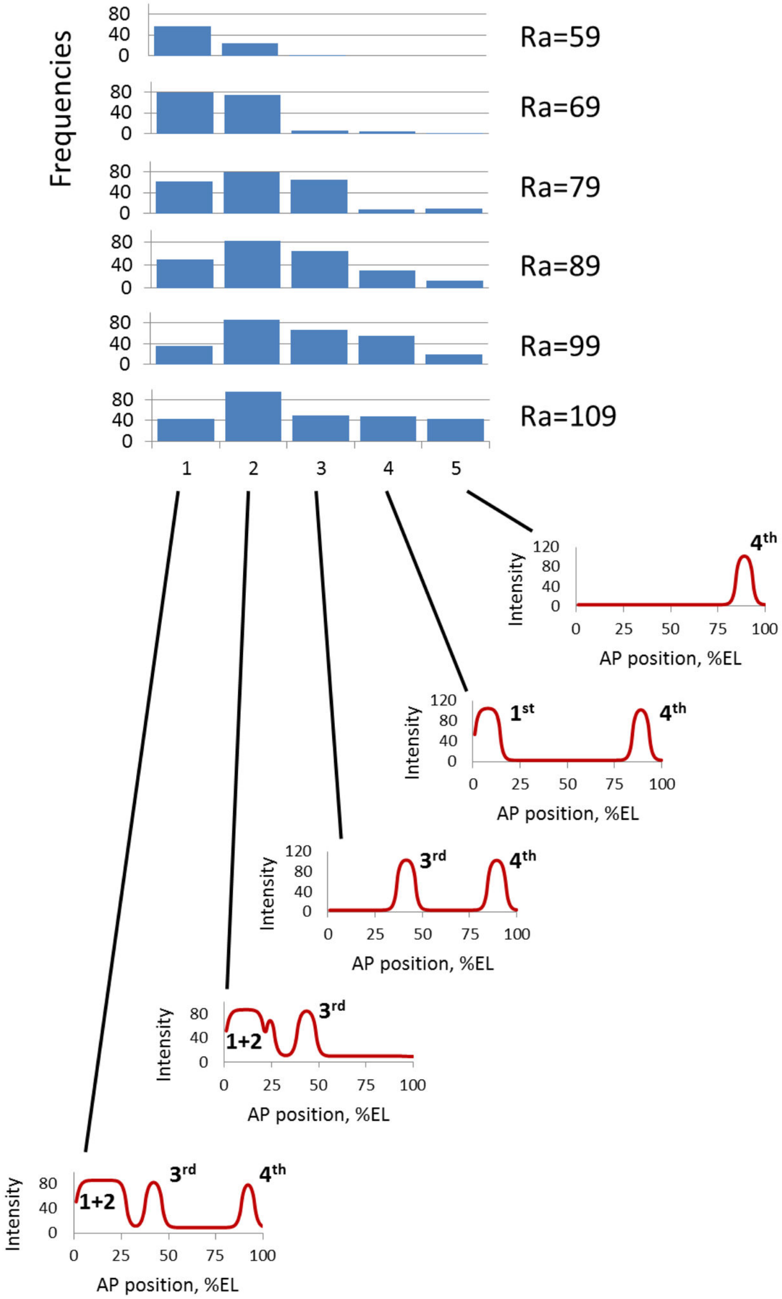
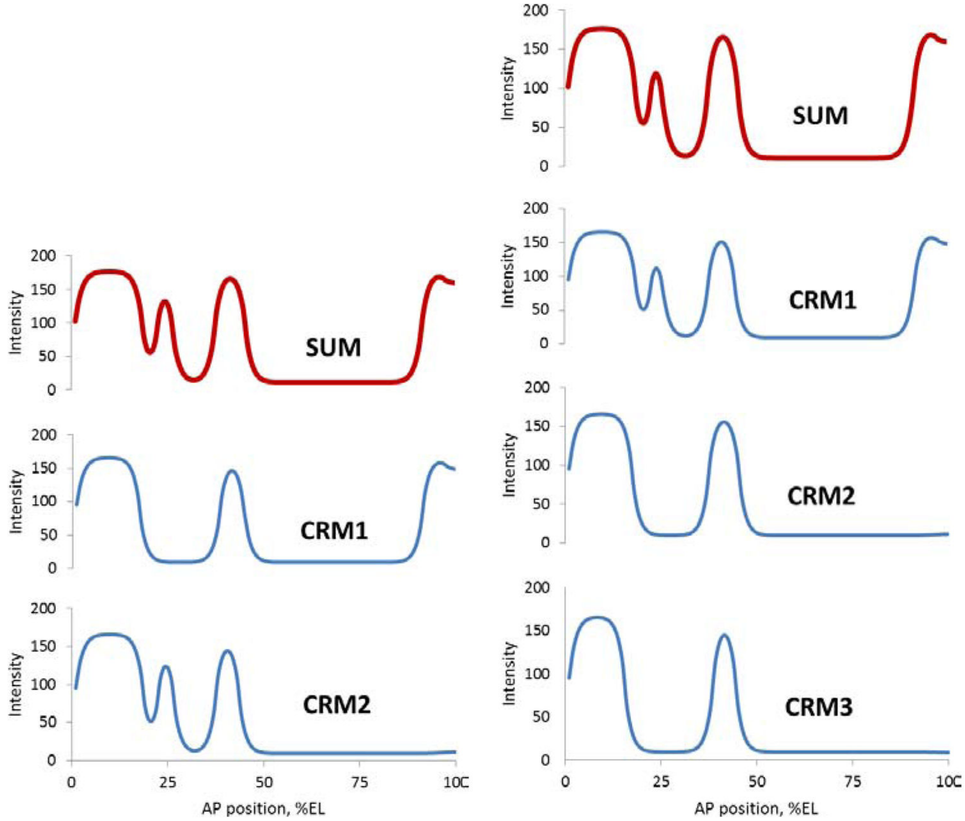


Fig. 7.

Frequencies of expression patterns driven by single CRMs change as *R* is varied. Histograms (top) of common single-CRM expression patterns, shown below in red. Low *R* favors Pattern 1, with a CRM controlling all domains; high *R* favors Pattern 5, a single-domain CRM. At intermediate *R* values, CRMs controlling pairs of domains are common, with the stripe element (Pattern 3) dominant for *R* = [79, 99].



(a)

(b)

Fig. 8.

Selective CRM action. Common outcomes for *hb* expression controlled by (a) two CRMs, and (b) three CRMs.

Table 1

Efficacy of evolutionary search for additive CRM action (mean number of solutions tested before success).

R	Efficacy
59	647, 085.7 ± 67, 360.3
69	1, 346, 888.7 ± 170, 803.0
79	1, 737, 514.3 ± 212, 023.8
89	1, 877, 451.8 ± 229, 023.4
99	2, 059, 735.7 ± 249, 811.2

Table 2

Efficacy of evolutionary searches with selective CRM action (mean number of solutions tested before success).

Evolutionary test	Efficacy
No *hb* self-activation, 7 TFs, repressive Ems	80, 171.5 ± 34, 969.3
No *hb* self-activation, 8 TFs, repressive Ems	146, 365.8 ± 45, 401.6
No *hb* self-activation, 9 TFs, repressive Ems	238, 863.7 ± 65, 308.7
No *hb* self-activation, 9 TFs, activating Ems	1, 451, 041.0 ± 218, 555.9
hb self-activation, 9 TFs, repressive Ems	1, 208, 207.8 ± 141, 947.8