

МОДЕЛИРОВАНИЕ МЕХАНИЗМОВ РЕГУЛЯЦИИ ЭКСПРЕССИИ ГЕНА *EVEN-SKIPPED* У ДРОЗОФИЛЫ

А.Д. Матвеева¹, Дж. Ионидес¹, М.Г. Самсонова¹, Дж. Райнитц²

¹ Отдел компьютерной биологии, Центр перспективных исследований Санкт-Петербургского государственного политехнического университета, Россия, e-mail: anya@odd.bio.sunysb.edu; ²

Отдел прикладной математики и статистики, Центр генетики развития, Университет
Стони Брук, Нью-Йорк 11794-3600, США

Работа посвящена предсказанию организации 5'-проксимального регуляторного района гена *even-skipped* (*eve*) исходя из картин экспрессии репортерного конструкта *p1.7eve-lacZ*. Для предсказания была использована математическая модель, предложенная в работе Янсенса с соавторами (Janssens *et al.*, 2006). Численные эксперименты, выполненные в этой работе, показали, что наилучшее совпадение решений уравнений модели с экспериментальными данными дает набор из 34 сайтов связывания активаторов Bcd, Hb, Cad и репрессоров Kr, Gt, Kni и Tll. Нами показано, что решения уравнений модели можно улучшить, введя два дополнительных предположения о том, что в силу удаленности сайта N1 от кластера сайтов связывания Bcd фактор Hb, связывающийся с этим сайтом, действует как репрессор и о функциональной значимости сайта связывания транскрипционного фактора Sloppy-paired в исследуемом регуляторном районе гена *eve*.

Ключевые слова: дрозофила, гены сегментации, регуляция транскрипции, моделирование, синергия.

Введение

В основе детерминации сегментов у плодовой мушки дрозофилы на стадии бластодермы лежит точный контроль транскрипции генов сегментации во времени и пространстве. Программа транскрипционного контроля закодирована в регуляторных районах этих генов. Она определяет формирование сложных пространственных картин экспрессии генов сегментации, которые намечают границы будущих сегментов.

Большинство генов сегментации кодируют транскрипционные факторы (ТФ). Регуляторные районы этих генов содержат несколько энхансеров, каждый из которых отвечает за экспрессию на определенном этапе развития и в определенной области эмбриона. Например, на стадии бластодермы экспрессия гена *even-skipped* (*eve*) в виде 7 полос, расположенных перпендикулярно anteriорнопостериорной (переднезадней, А-Р) оси эмбриона, регулируется 4 энхансерами, каждый из которых контролирует экспрессию этого гена в 1 или 2 полосах (Levine, Tjian, 2003).

Наиболее хорошо изучены энхансеры 5'-проксимального регуляторного района гена *eve*, MSE2 и MSE3 (Minimal Stripe Elements) длиной 400 п.н. и 500 п.н. соответственно. Термин «Minimal» означает, что если эти фрагменты сделать меньше, то экспрессия либо значительно уменьшится, либо совсем прекратится. В составе репортерного конструкта *lacZ* энхансер MSE2 определяет формирование картины экспрессии, повторяющей вторую полосу экспрессии нативного гена *eve*, а MSE3 – третью и седьмую полосы.

Однако в эксперименте с конструктом, содержащим нативный ген *eve* без энхансера MSE2, все еще наблюдалась остаточная экспрессия в области второй полосы (Ludwig *et al.*, 2005), а при удалении MSE3 наблюдалась остаточная экспрессия в области седьмой полосы (Small *et al.*, 1992). Эти факты ставят под сомнение достаточность энхансеров MSE2 и MSE3 для регулирования экспрессии в области второй, третьей и седьмой полос и указывают на необходимость более тщательного исследования этих энхансеров и прилегающих к ним областей.

В настоящее время общепринятым подходом к идентификации регуляторных районов генов является *in silico* предсказание сайтов связывания ТФ с последующей экспериментальной проверкой. Однако в случае больших геномов эукариот наиболее распространенные методы предсказания, основанные на использовании позиционно-весовых матриц, не позволяют дискриминировать настоящие сайты связывания от фонового шума и поэтому дополняются поиском либо кластеров транскрипционных сайтов, либо эволюционно консервативных сайтов.

В последнее время был предложен новый подход, заключающийся в предсказании организации регуляторных районов эукариот исходя из картин экспрессии соответствующих генных конструктов (Janssens *et al.*, 2006; Segal *et al.*, 2008). В этой работе мы использовали такой подход с целью предсказания структурной организации, т. е. взаимного расположения и соотношения сайтов связывания ТФ и их возможных кооперативных эффектов, в регуляторном районе гена *eve* длиной 1,7 т.п.н. на основе количественных характеристик экспрессии репортерного конструкта p1.7eve-lacZ. Этот район содержит энхансер MSE2 с прилегающими регуляторными районами и запускает экспрессию репортерного гена в области второй полосы и отчасти в области седьмой полосы (Small *et al.*, 1992).

Материалы

Изображения картин экспрессии

В нашем распоряжении имелись 110 латерально ориентированных эмбрионов дрозофилы, трансформированных репортерным конструктом p1.7eve-lacZ, содержащим энхансер MSE2 (рис. 1). Все эмбрионы находятся в циклах деления ядер 13 и 14. У каждого из эмбрионов с помощью конфокальной микроскопии были получены изображения картин экспрессии: нативного гена *eve* на уровне



Рис. 1. Конструкт p1.7eve-lacZ.

белка, мРНК *lacZ* и гистонов. Для регистрации экспрессии гена *eve* использовали метод непрямого иммунофлуоресцентного маркирования, мРНК *lacZ* регистрировали методом гибридизации *in situ*, для выявления гистонов использовали PicoGreen (Molecular Probes) (Wu *et al.*, 2001).

Данные по концентрациям ТФ, взаимодействующих с энхансером MSE2

Данные по относительным концентрациям ТФ Bicoid (Bcd), Hunchback (Hb), Caudal (Cad), Kruppel (Kr), Giant (Gt), Knirps (Kni) и Tailless (Tll) в каждом временном классе и в каждом ядре эмбриона были получены из базы данных FlyEx (<http://urchin.spbcas.ru/flyex/>).

Методы

Получение количественных данных об экспрессии

Обработка изображений картин экспрессии гена *eve* и конструкта p1.7eve-lacZ, а также изображений гистонов осуществлялась с помощью пакета ProStack (Processing Stacks) по ранее разработанному алгоритму (Matveeva *et al.*, 2006). В процессе обработки изображения были приведены к стандартному виду и формату, а информация, заключенная в них, была переведена в текстовые файлы, содержащие значения относительных концентраций каждого из сканированных продуктов в каждом ядре эмбриона. В дальнейшем для построения одномерных картин экспрессии мы учитывали только ядра из центральной 10 %-й полосы, вырезанной в направлении А-Р оси эмбриона.

Принадлежность эмбрионов к циклу деления 13 или 14А определяли подсчетом числа ядер. Эмбрионы, принадлежащие циклу деления 14, дополнительно классифицировались по 8 временным классам, продолжительностью 6,5 мин развития каждый (Т1-Т8), на основе визуального анализа картин экспрессии гена *eve* (Myasnikova *et al.*, 2002). Одномерные данные внутри каждого временного класса регистрировали и усредняли (Myasnikova *et al.*, 2001).

Удаление неспецифического сигнала из количественных данных по экспрессии *lacZ*

было осуществлено по разработанной ранее методике (Myasnikova *et al.*, 2005). Для этого все картины экспрессии *lacZ* были сглажены квадратичными сплайнами (Myasnikova *et al.*, 2001).

Поиск сайтов связывания

Поиск сайтов связывания ТФ в регуляторном районе конструкта *p1.7eve-lacZ* осуществлен при помощи программы *patser-v3d* (Hertz, Stormo, 1999) с использованием позиционно-весовых матриц (ПВМ) (Berman *et al.*, 2002; Rajewsky *et al.*, 2002).

Модель транскрипционного контроля экспрессии гена *eve*

Модель транскрипционного контроля основана на предположении, что инициация транскрипции – ферментативный процесс, катализируемый адаптерами транскрипции. В ходе катализа с увеличением общего числа связавшихся адаптеров уменьшается активационный энергетический барьер по закону Аррениуса и экспоненциально увеличивается скорость транскрипции. Для активации адаптера требуется $1/C_a$ активирующих ТФ, где величина C_a характеризует активационную эффективность. Посадке ТФ на ДНК могут помешать две причины: конкурентное связывание и репрессия близкого радиуса действия.

Вероятность связывания ТФ a с сайтом i с координатами $[m_i, n_i]$ относительно старта транскрипции записывается в виде:

$$f_{i[m_i, n_i; a]} = \frac{K_{i[m_i, n_i; a]} v^a}{1 + K_{i[m_i, n_i; a]} v^a}, \quad (1)$$

где v^a – концентрация ТФ a , а K_i – константа связывания с сайтом i . Константа связывания определенного ТФ с сайтом, обладающим наивысшей оценкой по ПВМ, является свободным параметром, а для последующих сайтов определяется из соотношения ПВМ оценки данного сайта к наивысшей оценке.

Если сайт связывания ТФ a , $i[m_i, n_i; a]$ перекрывается с сайтом связывания ТФ b , $j[m_j, n_j; b]$, возникает конкурентное связывание, и вероятность связывания ТФ a с сайтом i запишется в виде:

$$f_{i[m_i, n_i; a]} = \frac{K_{i[m_i, n_i; a]} v^a}{1 + K_{i[m_i, n_i; a]} v^a + K_{j[m_j, n_j; b]} v^b}. \quad (2)$$

Если ТФ a является активатором, то обозначаем f как $f_{i[m_i, n_i; a]}^A$, если репрессором близкого радиуса действия (РБД), то как $f_{i[m_i, n_i; a]}^Q$. При-сутствие РБД влияет на связывание активаторов со своими сайтами, поэтому полная вероятность связывания ТФ a с сайтом i с учетом действия РБД записывается в виде:

$$F_{i[m_i, n_i; a]}^A = f_{i[m_i, n_i; a]}^A \prod_k (1 - q(d_{ik}) E_b f_{k[m_k, n_k; b]}^Q), \quad (3)$$

где $d_{ik} = m_i - m_k$ – расстояние между сайтами и $q(d_{ik})$ – эмпирическая функция, равная 1 при $d_{ik} < 100$ п.н., 0 при $d_{ik} > 150$ п.н., а при d_{ik} в интервале от 100 до 150 п.н. $q(d_{ik})$ находят с помощью линейной интерполяции между 0 и 1. E_b – эффективность репрессии.

Количество сайтов для посадки адаптера выражается как:

$$N = \sum_a C_a \sum_i F_{i[m_i, n_i; a]}^A. \quad (4)$$

Тогда количество адаптеров, связавшихся через активаторы с этими сайтами, запишется:

$$M = f^{AF} N. \quad (5)$$

В бластодерме дрозофилы концентрация адаптеров фиксирована, поэтому величину f^{AF} считают фиксированным параметром модели.

Таким образом, по закону Аррениуса скорость транскрипции можно выразить как:

$$d[mRNA]/dt = R_0 \exp(-(\Theta - QM)), \quad \text{при } QM < \Theta \quad (6)$$

$$d[mRNA]/dt = R_0, \quad \text{при } QM \geq \Theta,$$

где свободный параметр Θ определяет максимальное значение активационного энергетического барьера в отсутствии активаторов, Q – фиксированная величина, на которую уменьшается величина энергетического барьера с присоединением каждого из адаптеров, свободный параметр R_0 – максимальная скорость транскрипции (Janssens *et al.*, 2006). Рассчитываемая скорость транскрипции пропорциональна концентрации мРНК, поскольку ее время жизни невелико (6 мин) по сравнению с рассматриваемыми в модели интервалами времени (Reinitz *et al.*, 2003).

Параметры модели находили параллельным методом численного отжига Лама (Lam, Delosme, 1988).

Результаты

С использованием ПВМ в регуляторном районе гена *eve* длиной 1.7 т.п.н. было найдено большое количество сайтов. Наибольшим сходством с консенсусами, описанными соответствующими весовыми матрицами, обладают 34 сайта связывания активаторов Bcd, Hb, Cad и репрессоров Kr, Gt, Kni и TII (рис. 2).

Как было показано в работе Янссенс с авторами (Janssens *et al.*, 2006), именно этот набор сайтов дает хорошее совпадение решений уравнений модели с экспериментальными данными. Среднеквадратичное отклонение решений уравнений от экспериментальных данных (rms) составило 9,20.

Коактивация

Качество модели можно улучшить, если помимо информации об относительной локализации сайтов связывания и их аффинности к соответствующим ТФ учитывать возможность возникновения кооперативных эффектов. Экспериментальные данные указывают на существование синергетического взаимодействия между транскрипционными факторами Bcd и Hb в энхансере MSE2 (Simpson-Brose *et al.*, 1994). В MSE2, содержащем несколько сайтов связывания активатора Bcd (рис. 2, B1-B5), фактор Hb проявляет себя в роли активатора транскрипции, тем самым усиливая активационную функцию Bcd. В соседнем энхансере MSE3, не содержащем сайты связывания Bcd, фактор Hb оказывает репрессирующее воздействие на транскрипцию (Small *et al.*, 1991).

Мы предположили, что в регуляторном районе конструкта p1.7*eve-lacZ* степень синергии между факторами Bcd и Hb напрямую зависит от расстояния между их сайтами связывания на ДНК. Таким образом, в нашей модели было учтено, что на фактор Hb, ассоциированный с сайтом, максимально удаленным от кластера сайтов Bcd (рис. 2, H1), синергетическое воздействие не распространяется, и этот фактор действует как репрессор.

Среднеквадратичное отклонение картин экспрессии конструкта, рассчитанных по такой модели, от экспериментальных данных составило $rms = 8,66$. По сравнению с моделью, опубликованной в работе Янссенс (Janssens *et al.*, 2006), решения уравнений новой модели лучше описывают формирование области экспрессии в районе седьмой полосы нативного гена *eve* в ранних временных классах T1 и T2, а также в поздних временных классах T5 и T6. Более того, решения модели очень хорошо согласуются с экспериментальными данными во временном классе C13.

Лучшее согласие решений уравнений новой модели с экспериментальными данными указывает на правомочность гипотезы о том, что в силу удаленности сайта H1 от кластера сайтов связывания Bcd фактор Hb, связывающийся с этим сайтом, действует как репрессор.

Влияние репрессора Sloppy-paired1

На ранних стадиях, предшествующих циклу 14A, ген *sloppy-paired1* (*slp1*) экспрессируется в виде непрерывной области в антериорной части эмбриона. Действуя в роли репрессора внутри этой области и в ее окрестностях, Slp1 предотвращает эктопическую активацию некоторых энхансеров генов pair-rule (в том числе энхансера MSE2 гена *eve*), тем самым напрямую влияя на формирование границ полос экспрессии этих генов (Andrioli *et al.*, 2002).

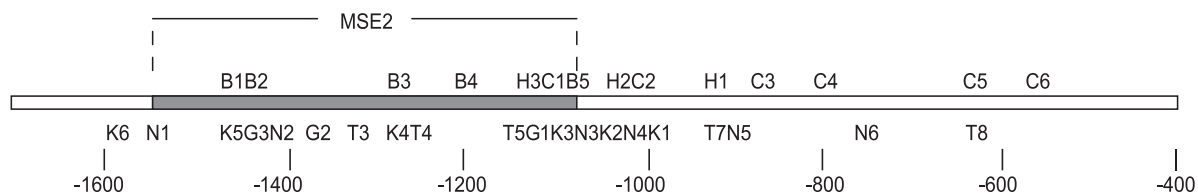


Рис. 2. Схема расположения сайтов связывания ТФ Bcd (B), Hb (H), Cad (C), Kr (K), Gt (G), Kni (N) и TII (T) в 5'-проксимальном регуляторном районе гена *eve* длиной 1724 п.н.

В энхансере MSE2 была найдена высококонсервативная последовательность (GTTT)₄, являющаяся специфическим сайтом связывания репрессора Slp1 (Andrioli *et al.*, 2002). Так как этот сайт расположен рядом с сайтами связывания активатора Bcd, Андриоли с соавторами (Andrioli *et al.*, 2002) предположили, что связывание Slp1 может мешать присоединению активаторов, что объясняет высокую чувствительность энхансера MSE2 к действию Slp1.

Для того чтобы проверить это предположение, результирующая модель была дополнена сайтом связывания Slp1 с координатами от -1221 до -1205.

Среднеквадратичное отклонение картин экспрессии конструкта, рассчитанное по такой модели, от экспериментальных данных составило 8,48.

По сравнению с моделью, учитывающей принцип коактивации, решения уравнений новой модели лучше описывают формирование области экспрессии в районе второй полосы нативного гена *eve* во временном классе T4. Также наблюдается небольшое улучшение в предсказании уровня экспрессии в седьмой полосе во временном классе T6 (рис. 3).

Лучшее согласие решений уравнений модели с экспериментальными данными указывает на

правильность предположения о том, что сайт связывания Slp1 с координатами от -1221 до -1205 важен для контроля экспрессии конструкта *p1.7eve-lacZ*, а значит и для правильной экспрессии гена *eve*.

Выводы

Данная работа посвящена предсказанию организации 5'-проксимального регуляторного района гена *eve* исходя из картин экспрессии репортерного конструкта *p1.7eve-lacZ*. Этот район содержит энхансер MSE2 с прилегающими регуляторными районами и запускает экспрессию репортерного гена в области второй полосы и отчасти в области седьмой полосы (Small *et al.*, 1992).

Для предсказания была использована математическая модель, предложенная в работе Янссенс с соавторами (Janssens *et al.*, 2006). Численные эксперименты, выполненные в этой работе, показали, что наилучшее совпадение решений уравнений модели с экспериментальными данными дает набор из 34 сайтов связывания активаторов Bcd, Hb, Cad и репрессоров Kr, Gt, Kni и Tll. Эти сайты обладают наибольшим сходством с консенсусами, описанными соответствующими весовыми матрицами.

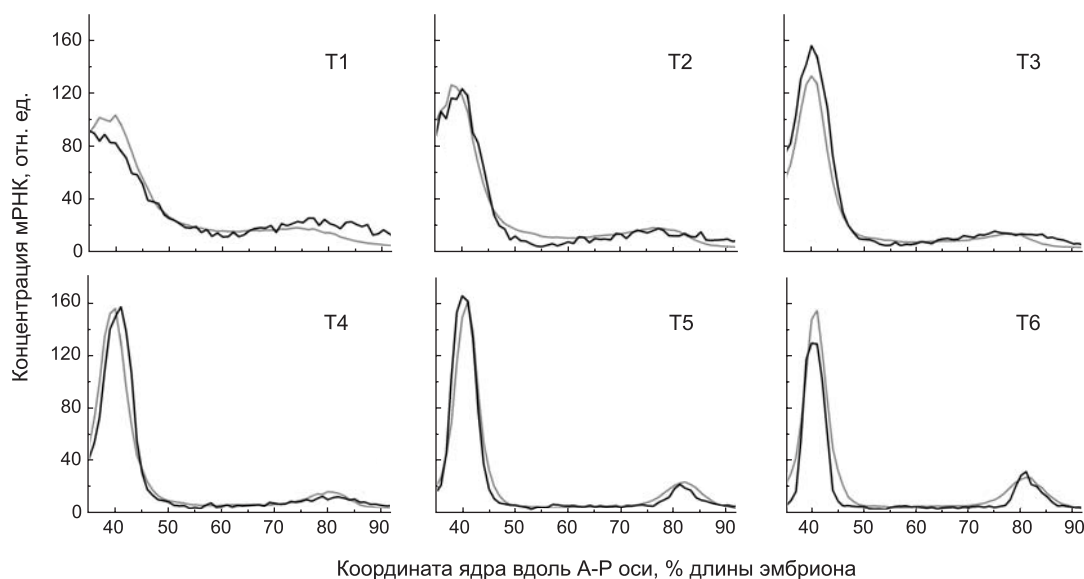


Рис. 3. Картины экспрессии конструкта *p1.7eve-lacZ* во временных классах T1 – T6, построенные по уравнениям модели (серый цвет) и соответствующие экспериментальные данные (черный цвет).

Нами показано, что решения уравнений модели можно улучшить, введя два дополнительных предположения о том, что в силу удаленности сайта H1 от кластера сайтов связывания Vcd фактор Hb, связывающийся с этим сайтом, действует как репрессор (1), и о функциональной значимости сайта связывания Slp1 с координатами от -1221 до -1205 (2). Лучшее согласие решений уравнений модели с экспериментальными данными указывает на правомерность таких гипотез и целесообразность их экспериментальной проверки.

Следует подчеркнуть, что хорошая предсказательная способность модели может быть дополнительно улучшена путем учета нуклеосомной организации 5'-проксимального района, а также путем дополнения модели математическим описанием механизмов коактивации, ко-репрессии и кооперативного взаимодействия.

В конечном счете эти теоретические исследования в сочетании с направленным экспериментом для проверки сформулированных гипотез позволят лучше понять принципы организации регуляторных районов генов у эукариот.

Работа поддержана грантом RR07801 Института здоровья США; программой поддержки грантов GAP Американского фонда гражданских исследований и развития, грант RUB1-1578-ST-05; грантом NWO-RFBR 047.011.2004.013; грантами РФФИ 08-01-00315-а и 08-04-00712-а.

Литература

- Andrioli L.P., Vasisht V., Theodosopoulou E. *et al.* Anterior repression of a *Drosophila* stripe enhancer requires three position-specific mechanisms // *Development*. 2002. V. 129. P. 4931–4940.
- Berman B.P., Nibu Y., Pfeiffer B.D. *et al.* Exploiting transcription factor binding site clustering to identify *cis*-regulatory modules involved in pattern formation in the *Drosophila* genome // *PNAS*. 2002. V. 99. P. 757–762.
- Hertz G.Z., Stormo G.D. Identifying DNA and protein patterns with statistically significant alignments of multiple sequences // *Bioinformatics*. 1999. V. 15. P. 563–577.
- Janssens H., Hou S., Jaeger J. *et al.* Quantitative and predictive model of transcriptional control of the *Drosophila melanogaster even-skipped* gene // *Nature Genet*. 2006. V. 38. P. 1159–1165.
- Lam J., Delosme J.M. An efficient simulated annealing schedule: Derivation // Technical Report 8816 (Yale Electrical Engineering Department, New Haven, Connecticut). 1988.
- Lam J., Delosme J.M. An efficient simulated annealing schedule: Implementation and evaluation // Technical Report 8817 (Yale Electrical Engineering Department, New Haven, Connecticut). 1988.
- Levine M., Tjian R. Transcription regulation and animal diversity // *Nature*. 2003. V. 424. P. 147–150.
- Ludwig M.Z., Palsson A., Alekseeva E. *et al.* Functional evolution of a *cis*-regulatory module // *PLoS Biol*. 2005. V. 3. I. 4. e93.
- Matveeva A., Kozlov K., Samsonova M. Methodology for building of complex workflows with PROSTAK package and iSIMBioS // Proc. of the 5th Intern. Conf. on Bioinformatics of Genome Regulation and Structure (BGRS'2006). Novosibirsk, 16–22 July 2004. Novosibirsk: Inst. Cytol. Genet., 2006. V. 2. P. 204–208.
- Myasnikova E., Samsonova M., Kosman D., Reinitz J. Removal of background signal from *in situ* data on the expression of segmentation genes in *Drosophila* // *Dev. Genes. Evol.* 2005. V. 215. P. 320–326.
- Myasnikova E., Samsonova A., Kozlov K. *et al.* Registration of the expression pattern of *Drosophila* segmentation genes by two independent methods // *Bioinformatics*. 2001. V. 17. P. 3–12.
- Myasnikova E., Samsonova A., Samsonova M., Reinitz J. Support vector regression applied to the determination of the developmental age of a *Drosophila* embryo from its segmentation gene expression patterns // *Bioinformatics*. 2002. V. 18. P. 87–95.
- Rajewsky N., Vergassola M., Gaul U., Siggia E.D. Computational detection of genomic *cis*-regulatory modules applied to body patterning in the early *Drosophila* embryo // *BMC Bioinformatics*. 2002. V. 3. Article N. 30.
- Reinitz J., Hou S., Sharp D. Transcriptional control in *Drosophila* // *ComplexUs*. 2003. V. 1. P. 54–64.
- Segal E., Raveh-Sadka T., Schroeder M. *et al.* Predicting expression patterns from regulatory sequence in *Drosophila* segmentation // *Nature*. 2008. V. 451. P. 535–540.
- Simpson-Brose M., Treisman J., Desplan C. Synergy between the hunchback and bicoid morphogens is required for anterior patterning in *Drosophila* // *Cell*. 1994. V. 78. P. 855–865.
- Small S., Blair A., Levine M. Regulation of *even-skipped* stripe 2 in the *Drosophila* embryo // *EMBO J*. 1992. V. 11. P. 4047–4057.
- Small S., Kraut R., Hoey T. *et al.* Transcriptional regulation of a pair-rule stripe in *Drosophila* // *Genes and Development*. 1991. V. 5. P. 827–839.
- Wu X., Vasisht V., Kosman D. *et al.* Thoracic patterning by the *Drosophila* gap gene *hunchback* // *Developmental Biol.* 2001. V. 237. P. 79–92.

MODELING THE EXPRESSION OF THE DROSOPHILA EVEN-SKIPPED (EVE) GENE

A.D. Matveeva¹, J.M.C. Ionides¹, J. Reinitz², M.G. Samsonova¹

¹Department of Computational Biology, Center of Advanced Studies, St. Petersburg State Polytechnical University, St. Petersburg, Russia, e-mail: anya@odd.bio.sunysb.edu; ²Department of Applied Mathematics and Statistics, and Center for Developmental Genetics, Stony Brook University, Stony Brook, NY, USA

Summary

In this work we use mathematical model introduced in Janssens *et al.* (2006) to predict the structural organization of the 5'-proximal regulatory region of the *even-skipped* (*eve*) gene from expression pattern of the p1.7*eve-lac* construct. Numerical experiments of Janssens *et al.* (2006) have shown that a set of 34 binding sites, in which Bcd, Hb, Cad act as activators and Kr, Gt, Kni, Tll act as repressors gives the best correspondence between the model solutions and experimental data.

We have shown that two additional assumptions can substantially improve the model accuracy. The first assumption is that Hb bound to H1 site acts as a repressor, because this site is located far away from the cluster of Bcd sites. The second one is about the functional importance of the Sloppy-paired binding site, which is found in the regulatory region of the *eve* gene.