У высших эукариот типичные сайты связывания сравнительно коротки, а их мотивы «вырождены», т.е. допускают неточные совпадения. Это контрастирует с четко определенными сайтами, характерными, например, для эндонуклеаз рестрикции. Способность факторов транскрипции узнавать сайты с различной аффинностью и конкурировать за связывание похожих участков ДНК делает возможной гибкую количественную регуляцию экспрессии. Вырожденность и сравнительно малая длина сайтов связывания (10-20 пар оснований) ограничивает применение стандартных методов для поиска гомологичных участков последовательностей [Cliften и др., 2001] на основе множественного локального выравнивания [Thompson, Gibson, Higgins, 2002]. Тем не менее, задача идентификации мотива, т.е. паттерна, описывающего локальную область сходства между последовательностями сайтов связывания, часто формулируется именно как построение множественного безделеционного локального выравнивания [Das, Dai, 2007]. Следует сразу сказать, что для ДНК-мотивов сайтов связывания факторов транскрипции обычно совместно рассматривают мотив и его обратно-комплементарное отражение (ориентацию), поскольку большинство факторов транскрипции осуществляют регуляцию через связывание именно двуцепочечной ДНК.

При рассмотрении множественного локального выравнивания достаточно естественным выглядит базовое предположение, что высоко консервативные позиции выравнивания (с предпочтительным содержанием конкретных нуклеотидов) соответствуют позициям сайтов связывания, которые являются более важными для ДНК-белкового узнавания. Фиксация конкретных наиболее консервативных нуклеотидов дает простейшие модели представления мотива как паттерна на четырехбуквенном нуклеотидном алфавите {A,C,G,T}: последовательности в нотации IUPAC 7 с допустимыми заменами [Day, McMorris, 1992] и регулярные выражения [Myers, Miller, 1989]. Такие модели легко редуцируются до ограниченного списка допустимых подстрок (слов), могут быть построены по считанным экспериментально определенным сайтам связывания, но только грубо, в бинарной форме «да/нет», описывают эффект от возможных замен нуклеотидов. Тем не менее, для слабо изученных или чрезвычайно хорошо выраженных сайтов связывания консенсусные модели до сих пор используются: например, консенсусы представлены в популярных коллекциях мотивов связывания для факторов транскрипции растений [Hehl, Bülow, 2014; Yilmaz и др., 2011]. Естественное развитие консенсусных моделей состоит в том, чтобы явно учесть количественную информацию о нахождении различных нуклеотидов в конкретной позиции выравнивания различных сайтов связывания для одного фактора транскрипции. Замечательно, что эта простая идея оказалось достаточной для создания базовой модели мотива – позиционно-весовой матрицы, которая не просто описывает множество вырожденных подстрок-сайтов, но и позволяет количественно оценить энергию ДНК-белкового взаимодействия [Berg, Hippel von, 1987; Stormo, Schneider, Gold, 1986].