Supplemental PWM File:

PWM learned by Zhao et al. (16) using BEEML trained on HT-SELEX data. See supplemental text file: *Zhao_Granas_Stormo_2009.Zif268_BEEML_SELEX_PWM.txt*

Supplemental Table 1:

This table shows the performance of each of the 45 GRaMS PWMs on the HT-SELEX data, as measured by the R^2 value of the counts predicted using each GRaMS PWM compared to the observed HT-SELEX nnnnnnGCGG counts. Each GRaMS PWM is also provided in 'vector' format, along with the estimated μ parameter. The R^2 value indicating the fit of each GRaMS model to its training dataset is also provided. Notice that GRaMS can learn predictive models that predict the HT-SELEX data well, even when the HT-B1H data is noisy (i.e. 4 hour, 0 IPTG data sets).

See supplemental table file: *supplemental_table_1.GRaMS_models_performance.xls*

FIGURE LEGENDS

Supplemental Figure 1. Simulated Ideal B1H data. (**A**) Energy versus log2(counts) curves for 5 different simulated data sets. Each curve represents a B1H experiment that was run for a different amount of time: 4, 8, 12, 18, 24 hours. (**B**) Energy versus growth rate curves for the same five data sets: 4, 8, 12, 18, 24 hours. Energies are given in units of K_bT. For these simulations, a μ value of 3 was used. All 4096 binding site alleles occurred once and each gave rise to a single simulated colony. Energies were assigned using the PWM of Zhao et al. (16).

Supplemental Figure 2. Plots of predicted energy values versus B1H growth rates (shifted so that the median is set to zero) for all 45 conditions. The PWM of Zhao et al. (16) was used to predict the energy of all 6mers. The plots are organized in a 5 column by 9 row grid. Each column corresponds to one of the 5 different time points (4, 8, 12, 18,24 hours), in that order, from left to right. Each row corresponds to one of the 9 IPTG and 3-AT combinations (50μM IPTG, 0.5 mM 3-AT; 10μM IPTG, 0.5 mM 3-AT; 0μM IPTG, 0.5 mM 3-AT; 50μM IPTG, 1 mM 3-AT; 10μM IPTG, 1 mM 3-AT; 0μM IPTG, 1 mM 3-AT; 50μM IPTG, 2 mM 3-AT; 10μM IPTG, 2 mM 3-AT; 0μM IPTG, 2 mM 3-AT), appearing in this order, from top to bottom.

Supplemental Figure 3. Sequence logos for all 45 PWMs produced using GRaMS on all of the HT-B1H datasets. The title of each sequence logo indicates (in order): the duration of each experiment in hours; the IPTG concentration (μM); the concentration of 3-AT (mM). All sequence logos were produced using in-house software, svgSeqLogo, written by RGC. The y-axis of each logo is in units of bits.

Supplemental Figure 4. Sequence logos for all 45 PWMs produced using BioProspector on all of the HT-B1H datasets. The title of each sequence logo indicates (in order): the duration of each experiment in hours; the IPTG concentration (μM); the concentration of 3-AT (mM). All sequence logos were produced using in-house software, svgSeqLogo, written by RGC. The y-axis of each logo is in units of bits.

Supplemental Figure 5. Scatter plots of R^2 values versus time showing the ability of BEEML, Log-Odds, BioProspector, and GRaMS PWMs to predict the nnnnnnGCGG SELEX counts. To obtain each R^2 value, the corresponding PWM was fixed, and BEEML was used to learn a μ parameter. The PWM and μ parameter were then used to predict the nnnnnnGCGG SELEX counts. The estimated counts were compared to the observed counts to obtain an R^2 value. Three figures are provided, one for each IPTG concentration: (**A**) 0 μM IPTG (**B**) 10 μM IPTG (**C**) 50 μM IPTG. For each time point, three points are shown per analysis method. Each of these three points corresponds to one of the different 3-AT concentrations: 0.5, 1, or 2mM.

Supplemental Figure 6. Sequence logo obtained using GRaMS on a data set obtained by performing B1H in liquid media (4hr, 50 μM IPTG, 5mM 3-AT). The PWM used to generate this logo was used to predict the nnnnnnGCGG subset of the SELEX data. As described in the results section, BEEML was used to fit a mu parameter and the resulting model was used to predict the counts of the nnnnnnGCGG sites obtained via SELEX. The R^2 value for this comparison was 0.90. Sequence logos were produced using in-house software, svgSeqLogo, written by RGC.

Supplemental Figure 7. Sequence logo showing the PWM of Zhao et al. (16). Sequence logos were produced using in-house software, svgSeqLogo, written by RGC. The y-axis of each logo is in units of bits.
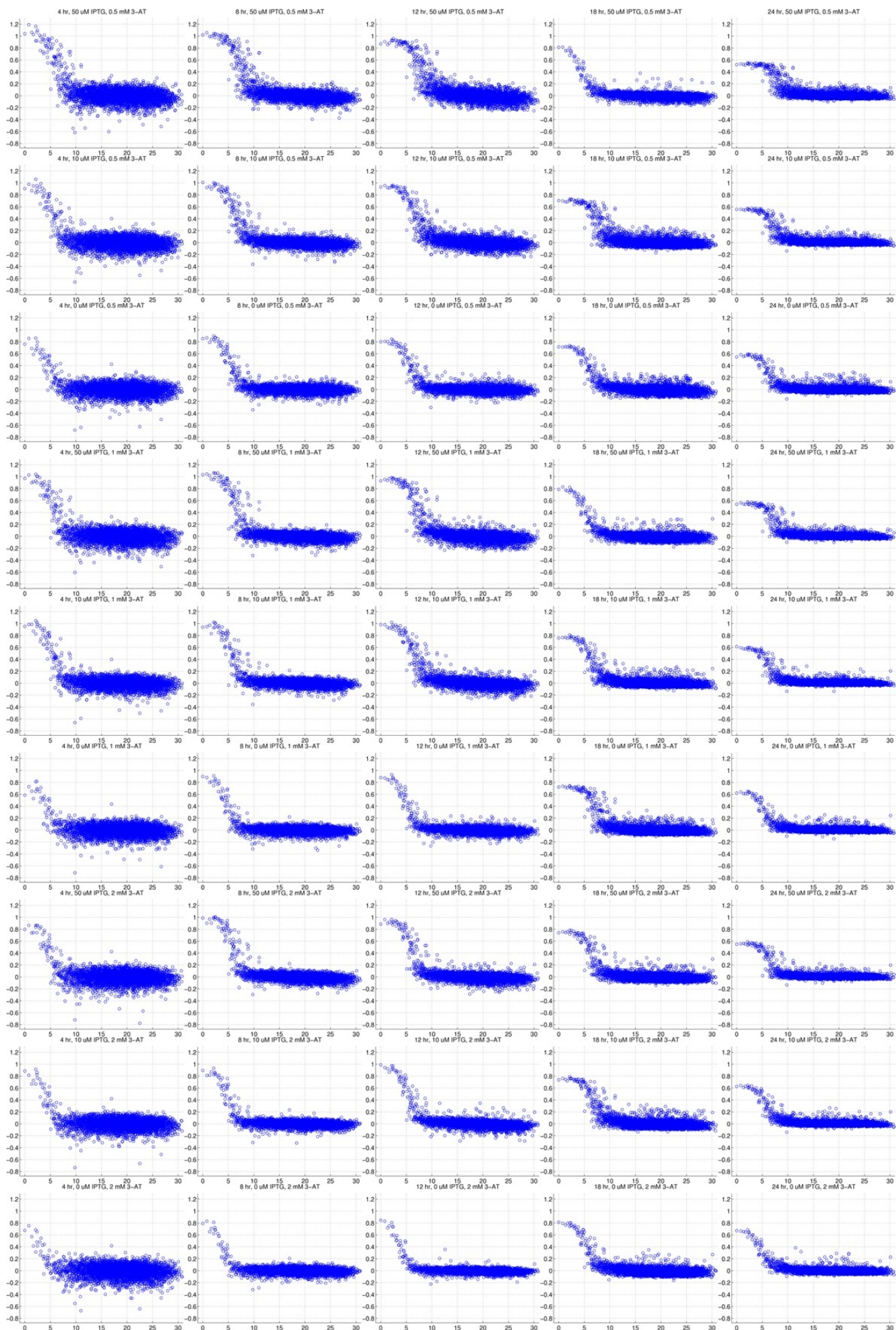
Supplemental Figure 8. Plot showing the ability to predict the training data, as measured by R^2, of all GRaMs models, as a function of time. Points show the average R^2 value per time point. Error bars indicate the range of the nine different IPTG and 3-AT concentrations for each time point.
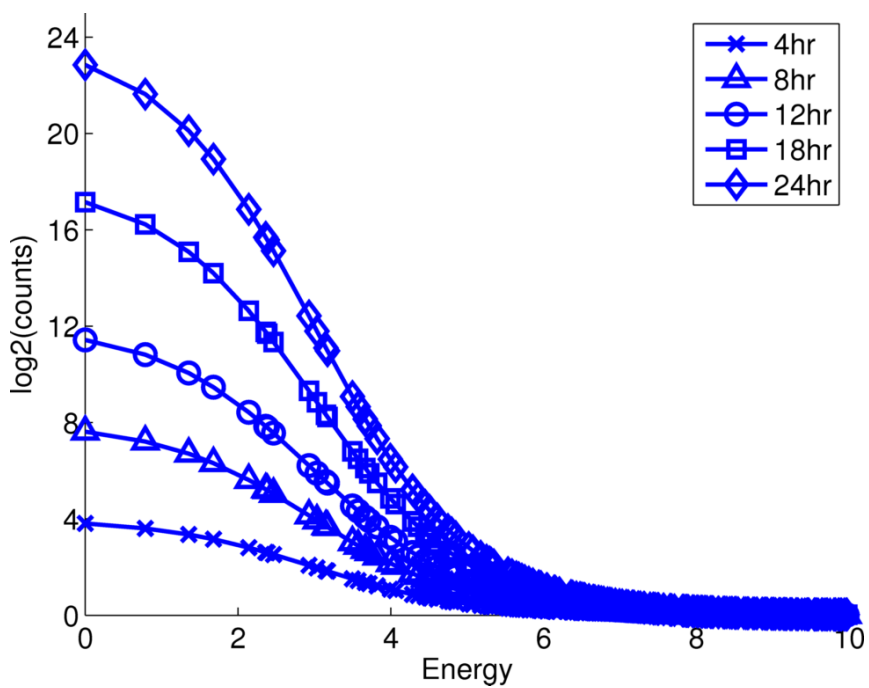
Supplemental Figure 1A
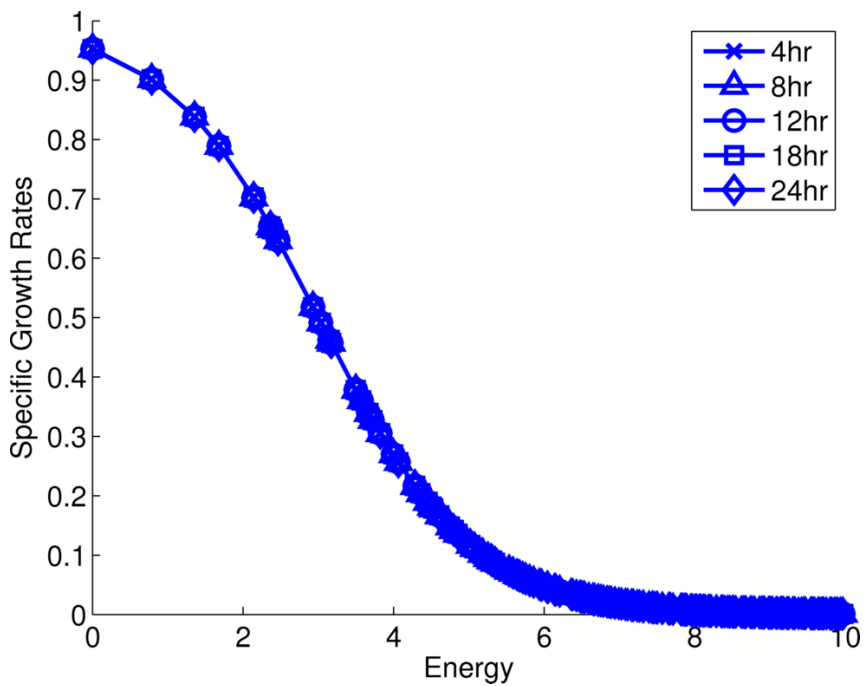
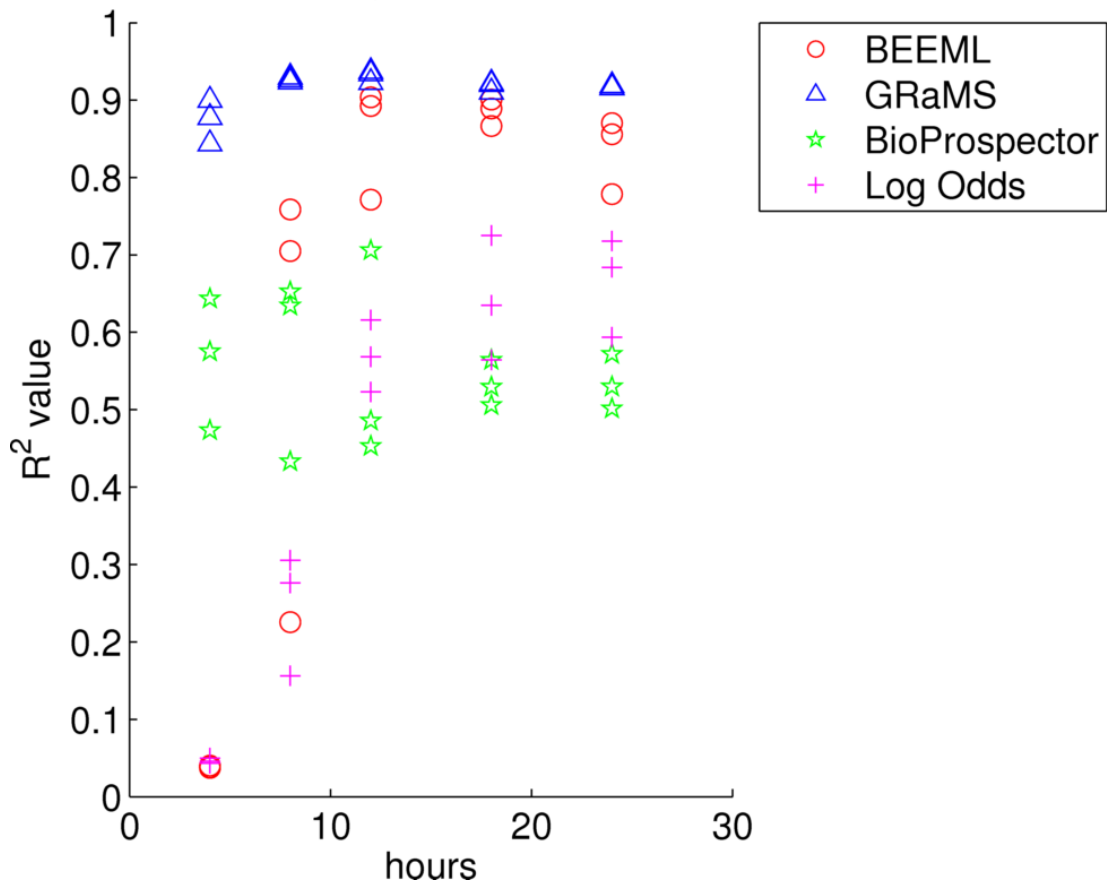Supplemental Figure 1B

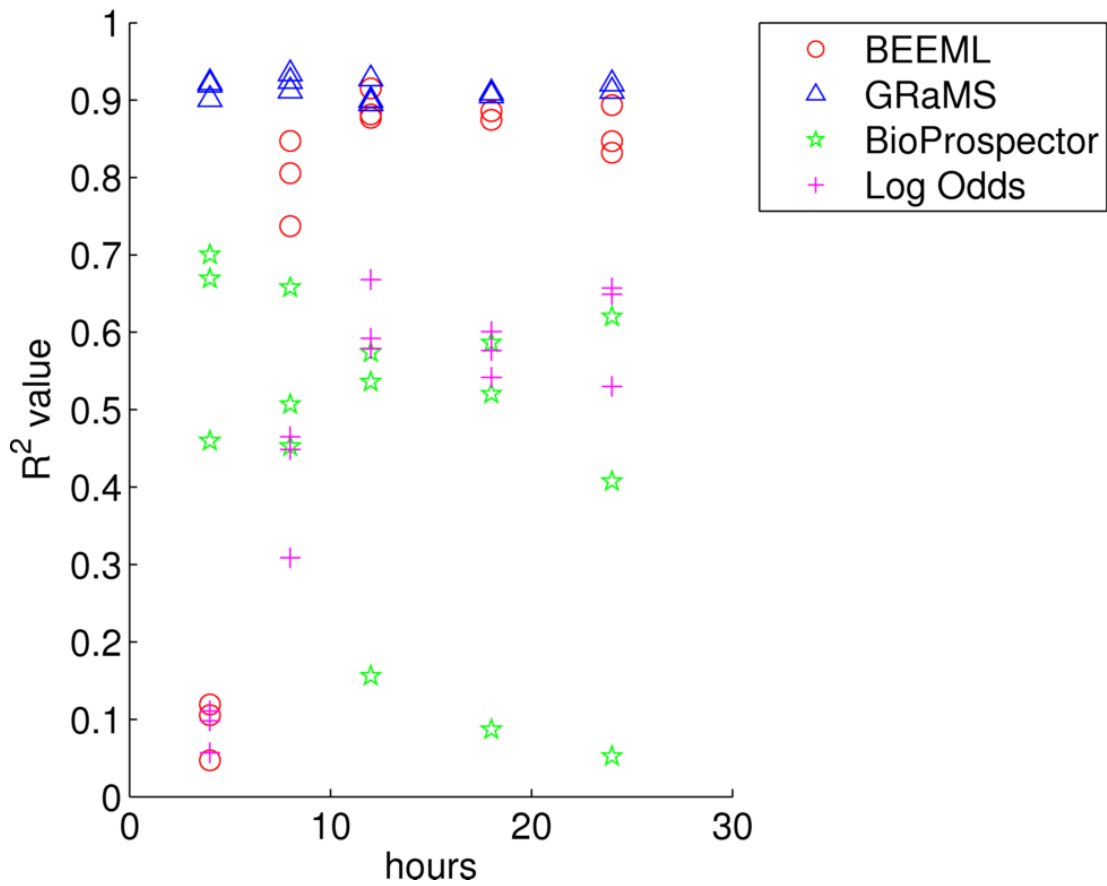Supplemental Figure 2
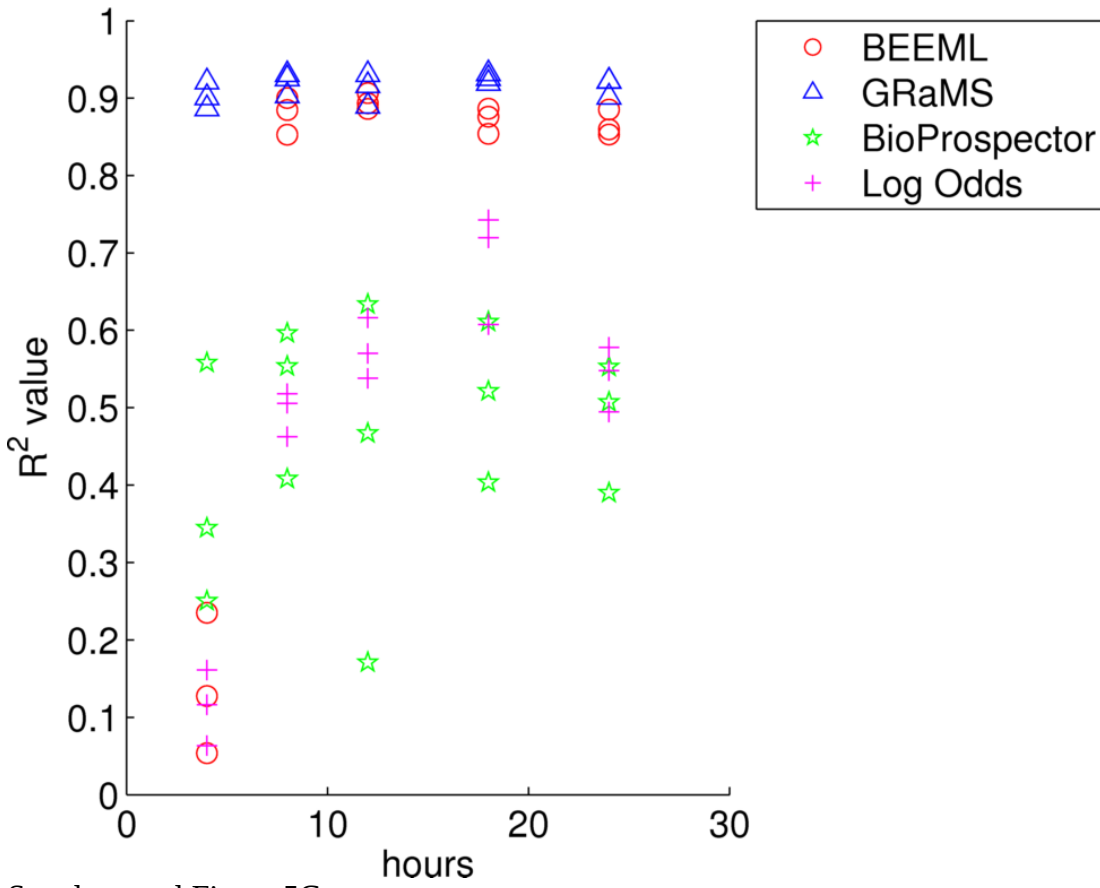
Supplemental Figure 3

Supplemental Figure 4

Supplemental Figure 5A
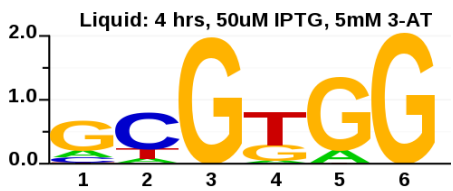

Supplemental Figure 5B
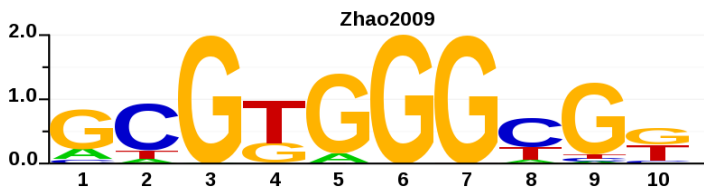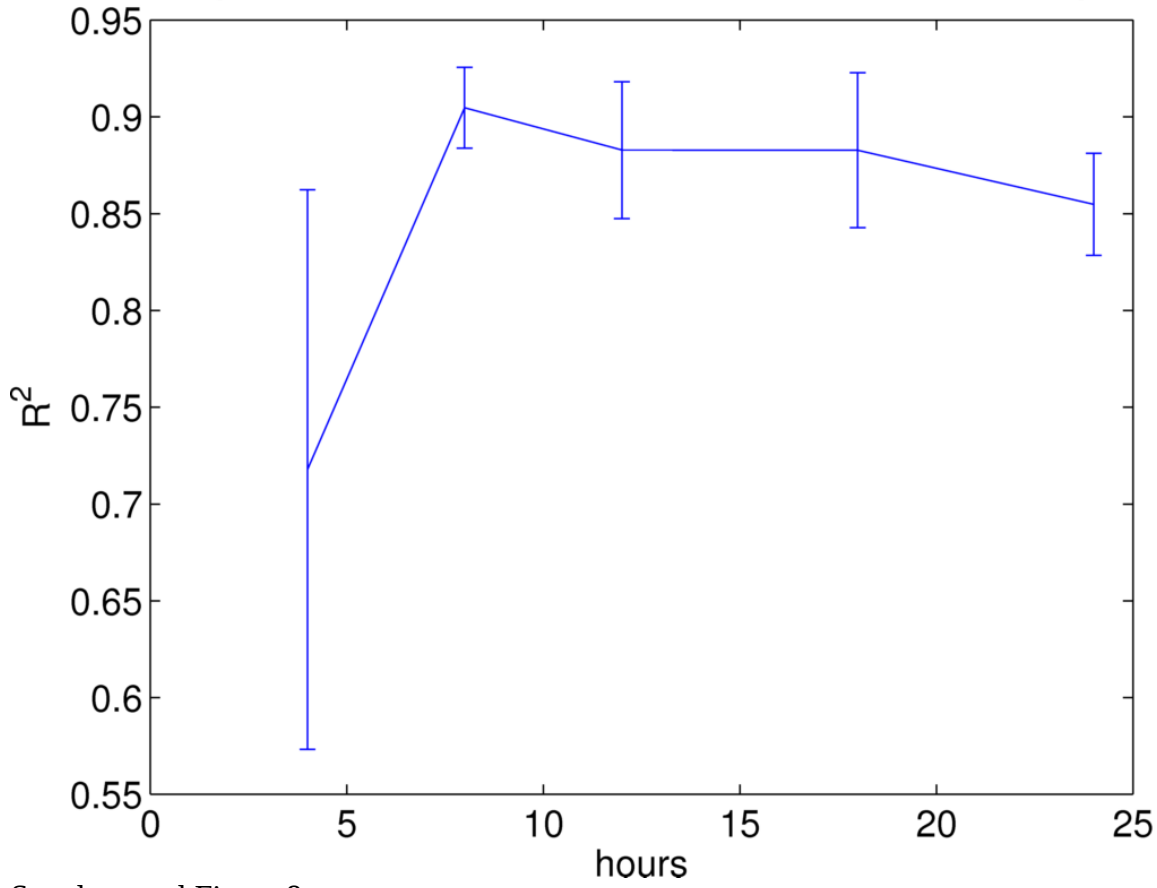
Supplemental Figure 5C

Supplemental Figure 6

Supplemental Figure 7

Average GRaMS Model Performance On HT−B1H Training Data

Supplemental Figure 8