Journal of chemical & engineering data

Article

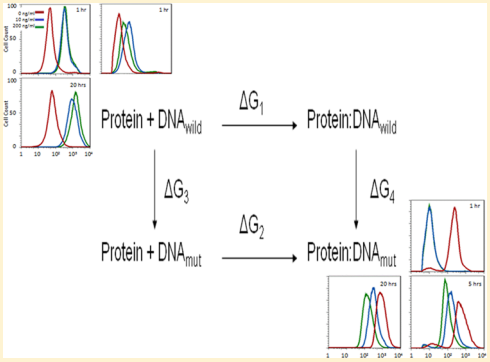
pubs.acs.org/jced

Investigation of Changes in Tetracycline Repressor Binding upon Mutations in the Tetracycline Operator

Dan S. Bolintineanu,[†] Katherine Volzing,[‡] Victor Vivcharuk,[§] Abdallah Sayyed-Ahmad,[∥] Poonam Srivastava,[⊥] and Yiannis N. Kaznessis*

Department of Chemical Engineering and Materials Science, University of Minnesota, Minneapolis, Minnesota, United States

ABSTRACT: The tetracycline operon is an important gene network component, commonly used in synthetic biology applications because of its switch-like character. At the heart of this system is the highly specific interaction of the tet repressor protein (TetR) with its cognate DNA sequence (tetO). TetR binding on tetO practically stops expression of genes downstream of tetO by excluding RNA polymerase from binding the promoter and initiating transcription. Mutating the tetO sequence alters the strength of TetR−tetO binding and thus provides a tool to synthetic biologists to manipulate gene expression levels. We employ molecular dynamics (MD) simulations coupled with the free energy perturbation method to investigate the binding affinity of TetR to different tetO mutants. We also carry out in vivo tests in *Escherichia coli* for a series of promoters based on these mutants. We obtain reasonable agreement between experimental green fluorescent protein (GFP) repression levels and binding free energy differences computed from molecular simulations. In all cases, the wild-type tetO sequence yields the strongest TetR binding, which is observed both experimentally, in terms of GFP levels, and in simulation, in terms of free energy changes. Two of the four tetO mutants we tested yield relatively strong binding, whereas the other two mutants tend to be significantly weaker. The clustering and relative ranking of this subset of tetO mutants is generally consistent between our own experimental data, previous experiments with different systems and the free energy changes computed from our simulations. Overall, this work offers insights into an important synthetic biological system and demonstrates the potential, as well as limitations of molecular simulations to quantitatively explain biologically relevant behavior.



1. INTRODUCTION

Synthetic biology has emerged as a distinct discipline based on a rational, bottom-up approach to the study and engineering of biological systems.[1,2] A key factor contributing to the ongoing success of synthetic biology is the ability to combine gene network elements in a modular fashion to create complex architectures with tailored functionalities.

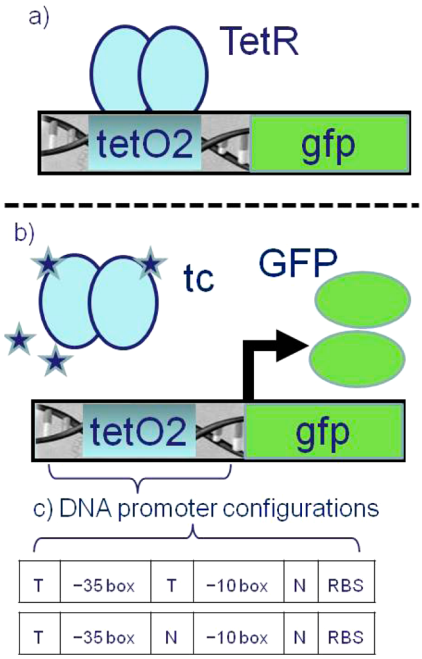
The tetracycline (tet) operon is one such gene network element that has found widespread use in synthetic biology applications. The tet operon and modifications thereof offer the ability to create a plethora of novel biological devices in bacteria and control gene expression levels simply by adding tetracycline to the bacterial environment. This level of control affords for the development of sophisticated biological circuits and devices, with applications ranging from the production of high-value therapeutics to biofuel synthesis and the development of biological sensors.

The natural function of the tet operon is to regulate the production of TetA protein, which confers resistance to the antibiotic tetracycline (Tc) in Gram-negative bacteria. The tet operon consists of two key DNA sequences: the *tetO* DNA operator, to which the tet repressor protein (TetR) binds with high specificity, and the *tetA* gene, which encodes the TetA protein that eliminates tetracycline from the cell via active transport. The *tetO* sequence is located immediately upstream of the *tetA* gene promoter region. In the absence of Tc, a dimer of the TetR protein is bound to the *tetO* operator, which prevents the binding of RNA polymerase and, hence, the transcription of the *tetA* gene. Upon the addition of Tc, a conformational change results in TetR that causes it to unbind from *tetO*. This, in turn, allows RNA polymerase to bind and transcribe the *tetA* gene. This system has been described and investigated in greater detail in earlier work.[3−7] In synthetic biology applications, the *tetA* gene can be replaced with any other gene, the expression of which can then be modulated by adjusting the concentration of Tc.

In the present work, we focus on the binding of the uninduced TetR protein to *tetO2*, a 19-base-pair-long DNA operator segment to which TetR binds with high specificity (schematic in Figure 1). In particular, we are interested in the effect of point mutations in the *tetO2* sequence on the protein−DNA binding affinity and the resulting downstream protein expression. Point mutations of the *tetO2* sequence can be carried out with ease and, thus, represent an important strategy for the fine control of gene expression levels in synthetic biological systems.[5]

Special Issue: Modeling and Simulation of Real Systems

Received: March 7, 2014
Accepted: August 1, 2014
Published: August 11, 2014



d) Promoter components

tetO2 (T)	TCCCTATCAGTGATAGAGA
−35 box	TTGACA
(N)	GTGTCACAAAGACTCAC
−10 box	GATACT
RBS	AAGAAACCGGTTC

e) Operator mutants

WT	TCCCTATCAGTGATAGAGA
M3	TCCCTA**A**CAGTG**T**TAGAGA
M4	TCCCT**T**TCAGTGA**A**AGAGA
M5	TCCC**A**ATCAGTGAT**T**GAGA
M6	TCC**G**TATCAGTGATA**C**AGA

Figure 1. (a) Schematic of TetR binding the DNA promoter at the *tetO2* site in the absence of tetracycline (Tc). (b) When Tc is added, TetR unbinds the promoter. RNA polymerase (not shown) can then initiate transcription of the gfp gene downstream, resulting in GFP protein production. (c) Two promoters were used in the experiments, TTN with two *tetO2* sites, and TNN with two intronic fungal DNA sequences (these fungal DNA sequences will not be bound by *E. coli* proteins). The −35 and −10 sequences are DNA sequences recognized by *E. coli* RNA polymerase. The ribosomal binding site (RBS) is recognized by *E. coli* ribosomes during translation. (d) Actual DNA sequences of all components. (e) *tetO2* mutant sequences.

In principle, the free energy of binding between TetR and *tetO2* is the thermodynamic property that quantifies the strength of the interaction. We set out to compute the free energy differences in TetR:*tetO2* binding as a function of several different point mutations. Using simulations, we aim to examine whether computed free energy values can correlate to biologically relevant behavior. We have constructed and tested a series of in vivo promoter sequences in *E. coli* as previously described.[4,6] Promoter sequences are segments of DNA found immediately upstream of the gene one wishes to control, in this case, green fluorescent protein (GFP). RNA polymerase binds to DNA regions located 35 and 10 base pairs upstream of the transcriptional start site (−35 and −10 positions). In our designs, there are three possible locations for operator sites that can facilitate control of gene expression: immediately upstream of the −35 position, between the −35 and −10 positions, and immediately downstream of the −10 position (see Figure 1). In each location, we can insert a DNA operator sequence that binds to a repressor protein; in this case, we insert the *tetO2* sequence, which is recognized with high specificity by TetR. The design of the promoter is such that if the TetR protein dimer is bound to any of the operator sites, RNA polymerase cannot bind to the promoter region due to steric hindrance caused by TetR, and expression of GFP is repressed. In previous work, we have used a combination of operator sites for two different protein repressors to create a biological AND gate;[6] in the present work, we only use the *tetO2* operator sequence (or its mutants, discussed below), either in the first position only (denoted TNN) or in both the first and second positions (denoted TTN). In the remaining positions, we insert an intronic sequence from the cc Okayama strain fungus (this is the "N" in TNN and TTN). This ensures that no proteins constitutively expressed in *E. coli* will bind to these regions.

TetR is constitutively expressed in the DH5αPro cells that we use; therefore, GFP expression is repressed in the absence of tetracycline. We thus expect the affinity of binding between TetR and different *tetO2* mutants to correlate to GFP repression levels. These promoter designs are based on earlier promoters for a biological AND gate system, which in turn were based on the modular transcriptional unit design of Lutz and Bujard.[7]

The mutants that we selected were based on earlier work by Sizemore et al.,[5] which tested the changes in protein repression levels resulting from all possible single-point mutations in both the *tetO1* and *tetO2* palindromic sequences. We elected to work exclusively with the *tetO2* operator, as it binds TetR more strongly[8] and has a much stronger effect on the overall behavior of the natural tet operon system.[4] Due to the high computational cost of the simulations carried out herein, we limited our choice of mutants to only four point mutations of the *tetO2* operator. We selected these to achieve a broad range of binding affinities and repression levels, as indicated by the data of Sizemore et al.[5]

The four mutants are designated M3, M4, M5, and M6, based on the position of the point mutations in the DNA sequence (with position 0 corresponding to the center of the near-palindromic sequence of *tetO2*; see sequences in Figure 1). In all cases, the mutations consist of simply swapping the appropriate base pairs between the leading and complementary strands. This leads to a total of four base pairs being mutated in each double-stranded DNA segment because each mutation is carried out on both sides of the palindromic center position (e.g., both the +3 and −3 positions in M3) and in both the leading and complementary strands (e.g., A → T, T → A in M3). The sequences of the leading strands of all mutants are shown in Figure 1.

In all cases, we are only interested in the binding of the uninduced TetR protein to each *tetO2* mutant. However, mutations in the *tetO2* DNA sequence will affect the strength

dx.doi.org/10.1021/je500225x | J. Chem. Eng. Data 2014, 59, 3167−3176

of the unrepressed promoter (i.e., the binding of RNA polymerase to the promoter) in addition to the binding of the TetR protein. As such, GFP levels for different promoters and mutants cannot be directly compared to computed TetR:*tetO2* binding affinities but rather must be normalized to account for changes in the strength of the free promoter. In the work of Sizemore et al.,[5] this was accomplished by comparing results to bacterial strains that contained the desired promoters but had the TetR gene knocked out, so that they did not contain any TetR protein. The percent change in the expression levels between the strains that contain TetR and those that do not thus yields a quantitative measure that is only reflective of the binding affinity between TetR and tetO.

In the present work, we achieve the same effect by comparing the expression levels for different mutants in the absence of anhydrous tetracycline (aTc) to those in the presence of very high aTc concentrations. In the latter case, there is a large excess of aTc relative to the amount of TetR protein, which results in complete induction of TetR, thereby freeing the promoter region and leading to fully unrepressed expression of GFP. We thus have a valid measure for comparing the free energies of binding of the uninduced TetR protein to *tetO2* and experimentally measured GFP expression levels.

2. METHODS

2.1. Experimental Methods. *Synthetic Promoter Synthesis.* Functional synthetic modules of the designed promoters were constructed using standard molecular biology techniques.[9] All of the promoter/operator sequences, transcriptional start sites and ribosome binding sites were obtained from previously published sequences.[10] For the present work, we designed a total of ten constructs, corresponding to the TNN and TTN topologies discussed earlier, where the *tetO2* sequence (denoted as T) is either the wild-type sequence or one of the four mutant sequences shown in Figure 1.

The synthetic, hybrid promoters were synthesized using splicing by overlap extension polymerase chain reaction (SOEing PCR).[11] Forward and reverse primers were designed for this two step process such that there was a 20 bp overlap, and the forward primer for all the constructs was kept the same. An initial template was generated with two, 111 bp, overlapping synthetic oligos corresponding to the sequence of each desired promoter. A total of 50 pmol of forward and reverse primers with 5 units of Taq DNA polymerase were combined and incubated at 72 °C for 65 min, hybridizing the two oligos. The hybridized DNA was column purified using a Qiagen PCR purification kit and the secondary nested PCR was completed. Products of the nested PCR were used as the template for final amplification of the synthetic promoters using external primers that corresponded to the terminal 20 bp of each sequence. The final 203 bp amplicons, the desired promoter sequences, were gel purified. Each promoter was cloned into pGLOW-TOPO (Invitrogen 12567-020) upstream of the green fluorescence protein reporter gene (*gfp*) and transformed into chemically competent Top10 (LacI−, TetR−) *E. coli* (Invitrogen, C404010) by heat shock at 42 °C for 45 s. Transformants were screened by GFP expression. Positive clones were cultured in Luria broth (LB) media with 100 μg mL^{-1} ampicillin (Amp) at 37 °C and 220 rpm, plasmids were isolated and promoter integrity was confirmed by DNA sequencing.

Culture Growth Conditions. All synthetic promoters were characterized in DH5αPro (LacI+, TetR+) *E. coli* cells over a range of aTc concentrations: 0 ng mL^{-1}, 1 ng mL^{-1}, 10 ng mL^{-1}, 50 ng mL^{-1}, 100 ng mL^{-1}, and 200 ng mL^{-1}. The GFP expression of all cultures, at 37 °C and 200 rpm, was monitored over 24 h. Cultures were diluted with fresh inducer media throughout the experiment to maintain them in mid logarithmic growth, $0.1 \leq$ OD600 ≤ 0.6 by spectrophotometry. At 3 h, 6 h, and 9 h, 200 μL cells per culture, approximately 10^5 cells, were isolated for analysis by flow cytometry. Cells were fixed with 4% paraformaldehyde (PFA) for 30 min at room temperature, washed with ice cold 1× phosphate buffered saline (PBS), resuspended in 500 μL 1× PBS and stored at 4 °C.

GFP Quantification Using Flow Cytometry. The GFP expression of individual cells was measured by flow cytometry using a FACScalibur (BD Biosciences) flow cytometer. A total of 100 000 cells were investigated per sample with excitation at λ_{ex} = 488 nm and subsequent fluorescence detection at $\lambda_{em} = 530 \pm 30$ nm. The cytometry data was collected using CellQuest (BD Biosciences) and analyzed using FlowJo (Tree Star) software. Each sample's healthy cell population was selected by first removing erroneous events (due to electronic noise) that fell below a minimum emission at $\lambda_{em} = 530 \pm 30$ nm and, second, removing events that fell outside of the characteristic side-scatter and forward-scatter range for single *E. coli* cells. The differential GFP expression of the selected cells was analyzed and compared across samples.

2.2. Molecular dynamics simulations. In order to mimic the experimental systems in our work as closely as possible, we created a homology model structure for the TetR protein dimer corresponding to the TetR variant constitutively expressed in DH5αPro cells. A BLAST[12] search of the PDB database[13] revealed that our protein is 95% similar to the variant of Luckner et al.[14] (PDB ID 2NS7). However, the structure of this protein was resolved for a monomer in free solution, rather than a dimer in a protein:DNA complex. Additionally, the crystal structure of the uninduced TetR dimer bound to *tetO1* was resolved by Orth et al.[15] (PDB ID 1QPI). Our protein has a 70% sequence similarity to the TetR variant of Orth et al. We therefore took the following approach to creating a homology model for our protein: using the Schrödinger Prime software,[16] we threaded the sequence of our protein onto the monomer structure of Luckner et al.[14] (PDB ID 2NS7). Next, this monomer structure was duplicated and aligned to the TetR monomers in the TetR dimer:*tetO1* crystal structure of Orth et al.[15] (PDB ID 1QPI) to create a homology model relevant to our systems. Finally, base pair point mutations were made at appropriate locations to turn *tetO1* into *tetO2*. The resulting complex is shown in Figure 2.

Because the primary sequence similarity between the protein in our systems and the TetR variant from both Luckner et al.[14] and Orth et al. [15] is quite high, we are confident in the quality of the resulting homology model; the few differences that do exist are primarily in regions far from the DNA binding site, so we can safely surmise that the relevant portions of the protein structure are accurately captured.

The TetR:*tetO2* complex resulting from the homology model was simulated for ∼125 ns in fully atomistic explicit solvent in order to equilibrate the structure prior to free energy perturbation calculations. The TetR:*tetO2* structure was solvated in a cubic simulation box with an initial side length of 98 Å, containing almost 27 000 TIP3P[17] water molecules. An additional 47 sodium counterions were added to balance the net negative charge of the protein:DNA complex, as well as an additional 53 Na^+ and 53 Cl^- ions to simulate a solution with an ionic strength of 0.1 M.[18] The system contained a total of 88 289 atoms. We also simulated the *tetO2* DNA segment without the

dx.doi.org/10.1021/je500225x | J. Chem. Eng. Data 2014, 59, 3167−3176

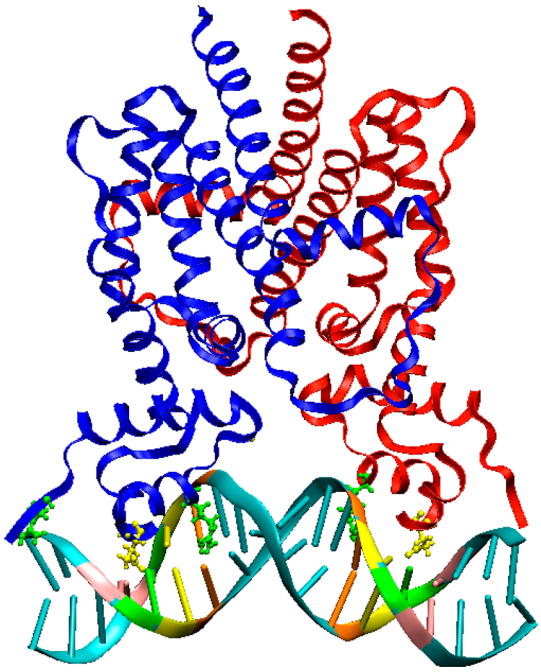


Figure 2. TetR:*tetO2* wild-type complex obtained from homology modeling based on PDB ID's 2NS7 and 1QPI after equilibration. The TetR monomers are shown in red and blue. Arginine and glutamine residues on the TetR protein near the binding site are shown as green and yellow ball-and-stick representations, respectively. Mutation positions on the *tetO2* sequence are color-coded as follows: 3, orange; 4, yellow; 5, green; 6, pink.

TetR protein in order to complete the thermodynamic cycle discussed in the free energy methods section below. In this case, the *tetO2* structure as well as surrounding solvent were extracted from the latter part of the TetR:*tetO2* simulation and resolvated in a cubic simulation box with an initial side length of 95 Å; appropriate counterions and salt molecules were added, and a similar equilibration strategy was implemented.

The construction of both systems was carried out using the CHARMM software package, version c33b1.[18] Prior to the production simulations, the solvated systems were minimized for 5000 steps and gradually heated to 310 K, with the heavy atoms of the protein and DNA restrained to their initial locations. The restraints were gradually removed over an additional several nanoseconds of equilibration and completely removed for the production runs.

All minimization, equilibration and production runs were carried out with the NAMD software package version 2.7b2.[19] We employed the CHARMM 27 force field[20] with CMAP corrections.[21] All simulations were carried out in the constant temperature and pressure (NPT) ensemble, which is implemented in NAMD using the Nosé–Hoover–Langevin[22] piston method. In all cases, the pressure was set to 1 atm, the pressure piston period was set to 200 fs and the piston decay was set to 100 fs. Simulation box dimensions were scaled isotropically to avoid box shape distortions that could be a source of inconsistencies among different mutants (see below). Electrostatic interactions were modeled using the particle mesh Ewald summation technique,[23] with a fast Fourier transform grid with a spacing of approximately 1 grid point per Ångstrom, whereas van der Waals interactions were smoothly switched off between 8 Å and 11 Å. All bonds involving hydrogen atoms were constrained using the SHAKE algorithm,[24] which allowed for an integration time step of 2 fs.

In order to prevent fraying of the ends of the *tetO2* DNA segment, we applied additional restraints along the complementary base pair hydrogen bonds of the terminal base pairs in all cases. In this manner, we more closely approximate a biological system, in which TetR binds to a *tetO2* sequence that belongs to a much longer DNA strand, rather than an isolated 19-bp segment. In all cases, the relevant pairings are Thy, N3:Ade, N1 and Thy,O4: Ade, N6. Harmonic restraints are applied between these atom pairings, with spring constants of 10 kcal mol^{-1} $Å^{-2}$, and a set distance of 3.0 Å. This effectively ensures that the Watson–Crick hydrogen bond pairing in the terminal base pairs is always maintained. These restraints were active at all times in all simulations.

2.3. Mutant Simulations and Free Energy Perturbation Calculations. We used the final structure from the simulation of the wild-type TetR:*tetO2* as a starting structure for alchemical transformation/free energy perturbation (FEP) simulations to study the effects of *tetO2* mutations. In particular, we are interested in computing the differences in binding affinities between TetR and different *tetO2* mutants that results from such mutations. In the scheme shown in Figure 3, $tetO2_{wild}$ and $tetO2_{mut}$ represent the wild-type and a mutated variant of the *tetO2* operator site (where *mut* is one of M3, M4, M5, and M6; see Figure 1).

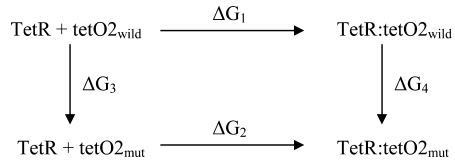


Figure 3. Thermodynamic cycle of alchemical transformations used in the free energy calculations.

We are interested in computing the binding free energy difference between the top and bottom association processes, $\Delta G_1 - \Delta G_2$. Because the association process itself involves protein and DNA conformational changes that are slow to converge on the time scales of typical MD simulations, we do not attempt to compute ΔG_1 or ΔG_2 directly. Instead, we employ an alternate approach, in which the DNA is computationally mutated from the wild-type structure to the mutant structure. This defines two nonphysical but highly useful processes in the thermodynamic cycle shown in Figure 3. The first of these is the transformation of the wild-type *tetO2* DNA segment to a particular mutant in the absence of the TetR protein, which entails a free energy change ΔG_3. The second process involves carrying out a similar transformation, where the associated TetR:$tetO2_{wild}$ complex is transformed to the mutant complex TetR:$tetO2_{mut}$, resulting in a free energy change ΔG_4. Because free energy is a state function, we can calculate the free energy difference that we are interested in as follows:

$$\Delta\Delta G_{binding} = \Delta G_1 - \Delta G_2 = \Delta G_3 - \Delta G_4 \quad (1)$$

In order to compute the free energy differences ΔG_3 and ΔG_4, we employ the alchemical free energy perturbation technique, the theory of which dates back more than 60 years.[25] Briefly, the free energy difference between two states *a* and *b* of a system can be calculated according to

$$\Delta A_{a\to b} = -k_B T \ln\langle \exp(-(H_b - H_a)/k_B T)\rangle_a \quad (2)$$

dx.doi.org/10.1021/je500225x | J. Chem. Eng. Data 2014, 59, 3167–3176

where k_B is Boltzmann's constant, T is the temperature and H_a and H_b are the Hamiltonians of the two states of the system (total kinetic and potential energies; evaluated based on the system conformation and the simulation force field). The $\langle ... \rangle_a$ brackets indicate an ensemble average over state a. If the two states a and b are sufficiently similar, the ensemble average can be estimated with a relatively short MD simulation. In the case of mutating several residues on a DNA segment, this condition is not met. To circumvent this problem, we define a series of intermediate states between the wild-type and mutant DNA, associated with a coupling parameter λ.[26] In the dual-topology approach for free energy perturbation (FEP) calculations implemented in NAMD, λ scales the interaction parameters of the atoms that are being created, whereas $(1 - \lambda)$ scales the atoms that are being destroyed. The combined Hamiltonian is therefore given by

$$H_\lambda = H_0 + \lambda H_b + (1 - \lambda)H_a \quad (3)$$

Here, H_0 represents the system Hamiltonian excluding all atoms being mutated, whereas H_a and H_b are the Hamiltonians of the systems corresponding to the initial and final mutation states, respectively. As λ is increased from 0 to 1, the alchemical mutation is completed. Typically, λ is increased over many intervals between 0 and 1, and the total free energy change is calculated as the sum of the free energy difference at each interval i

$$\Delta A_{a\to b} = -k_B T \sum_{i=1}^{N} \ln\langle \exp(-(H_{\lambda_{i+1}} - H_{\lambda_i})/k_B T)\rangle_{\lambda_i} \quad (4)$$

Here, N is the total number of intervals. A more detailed description of the underlying theory is given by Chipot and Pohorille.[27]

We implemented the free energy perturbation method using the appropriate features in the NAMD software version 2.7b1. We have taken advantage of the recently implemented soft-core van der Waals radius-shifting coefficient, which improves the convergence characteristics of FEP calculations and avoids extremely high energies resulting from the creation and annihilation of different atoms in the system (often referred to as "end-point catastrophes"). The van der Waals shifting coefficient was set to a value of 5, and the electrostatic interactions for the atoms being mutated were introduced at values of $\lambda > 0.5$. In all cases, the mutation process was divided into approximately 100 windows, with values of λ incremented by 0.01 between 0.01 and 0.99, and increasingly smaller intervals near the end points ($\lambda = 10^{-6}$, 10^{-4}, 10^{-3}, $5{\cdot}10^{-3}$, 0.01 and 0.99, 0.995, 0.999, 0.9999, 1).

Each window consisted of a total of 400 ps of MD simulation, of which the first 120 ps were treated as equilibration time and discarded from the ensemble averaging in eq 2. Simulations at different values of λ were carried out sequentially, so that each window was started from the final coordinates and velocities of the previous window. This helps to ensure faster equilibration and convergence of the relevant energy values at each window. For both the free *tetO2* mutations (ΔG_3) and the TetR:*tetO2* complex simulations (ΔG_4), the velocities at the end of the wild-type simulations were used to start the FEP calculations for all mutants. All MD simulation parameters were the same as those of the wild-type simulations discussed above.

3. RESULTS AND DISCUSSION

3.1. Experimental Results. For our present purposes, we are interested in GFP expression levels for all the mutants and the wild-type *tetO2* sequences in both types of promoters (TNN and TTN). The fluorescence values represent the mean of 100 000 fluorescence events from cultures extracted after 9 h of incubation, with no aTc as well as with high levels of aTc. The mean fluorescence levels were then used to compute the percent repression for each case, where we have assumed that the maximum aTc concentration corresponds to the operator site being completely unbound, while the case of zero aTc corresponds to maximum repression for a particular construct. In all cases, the mean fluorescence level is the average of all fluorescence events from the FACS assay, as well as the average of two or three separate experiments for the same construct. The repressor strength is calculated according to

$$\%\ \text{repression} = \frac{\text{mean GFP at max. aTc} - \text{mean GFP at 0 aTc}}{\text{mean GFP at max. aTc}} \times 100\% \quad (5)$$

As discussed in section 2.1, this calculation is necessary in order to normalize the results of the different systems to a common basis, both due to the effects of *tetO2* mutations on RNA polymerase binding, as well as possible variations in FACS calibration levels among experiments. Figure 4 summarizes the results for all mutants at all time points.

As expected given the complexity of these systems, there are some notable variations in the relative repression levels of the different mutants. As there is no clear basis for selecting a single time point or construct as a definitive measure of repression

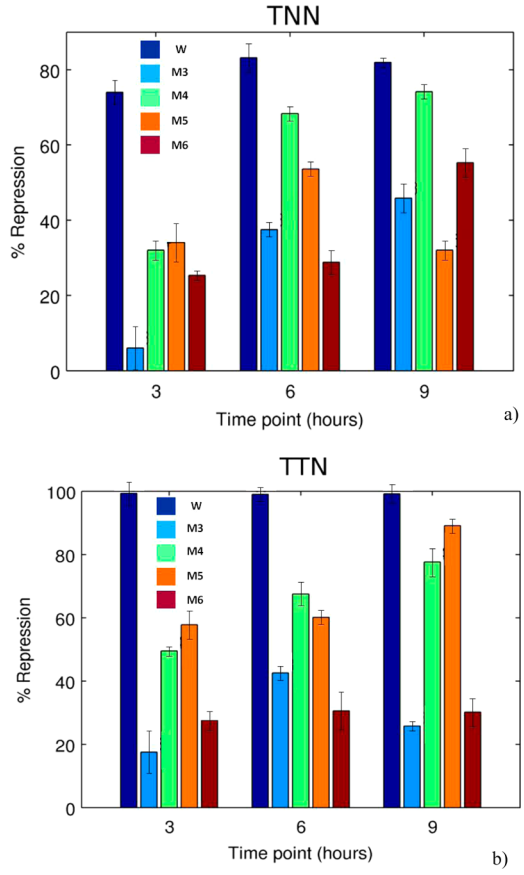


Figure 4. Experimental results for TNN (a) and TTN (b) systems. Data are shown for all four mutants at three time points. Higher repression levels indicate stronger binding of the uninduced TetR protein. Error bars represent the variance over 100 000 flow cytometry measurements, in experiments conducted two or three times.

dx.doi.org/10.1021/je500225x | J. Chem. Eng. Data 2014, 59, 3167−3176

strength, we present the complete data and restrict our interpretation to general trends. We note that the wild-type *tetO2* sequence always results in the strongest repression. Furthermore, with only one exception, mutants M4 and M5 are stronger repressors than mutants M3 and M6 (i.e., M4, M5 > M3, M6). This is consistent with the data of Sizemore et al., even though these authors used different bacterial strains, different experimental conditions, and a different promoter consisting of both *tetO1* and *tetO2*, as well as a slightly different variant of the TetR protein.

Previous work has shown that the *tetO2* operator binds tetR more strongly[8] and tends to have a much stronger effect on the overall behavior of the tet operon than *tetO1*.[4] This may partially explain why our experimental results, which are based entirely on *tetO2*, are in good agreement with those of Sizemore et al., even though these authors used a combination of *tetO1* and *tetO2*.

3.2. Free Energy Perturbation Results. The results of the FEP simulations described in section 2.3 above are summarized in Table 1. A detailed discussion of the errors associated with these data is deferred to section 3.3.

Table 1. Summary of All $\Delta\Delta G$ Values Obtained from FEP Simulations[a]

mutant	ΔG_3	ΔG_4	$\Delta G_1 - \Delta G_2 = \Delta G_3 - \Delta G_4$
	kcal mol^{-1}	kcal mol^{-1}	kcal mol^{-1}
wild	0	0	0
M3	1.61 ± 0.18	−0.99 ± 0.17	2.60 ± 0.25
M4	−1.15 ± 0.17	−3.05 ± 0.17	1.90 ± 0.24
M5	2.08 ± 0.18	0.78 ± 0.16	1.30 ± 0.24
M6	8.35 ± 0.15	2.25 ± 0.15	6.10 ± 0.21

[a]The errors are indicative of sampling errors only, as discussed in section 3.3. ΔG_3 is the free energy change associated with the mutation of the solvated wild-type *tetO2* DNA sequence to the mutated form, whereas ΔG_4 is the free energy associated with mutation of the wild type TetR:*tetO2* complex to the mutated complex.

In order to provide a more straightforward comparison to experimental data, we have estimated the absolute free energies of binding of all four mutants. The free energy of binding $\Delta G_{wild} = -kT \log K_{as}$ of the wild-type *tetO2* sequence to TetR was estimated to be −12.3 kcal mol^{-1} using the correlation given by Kedracka-Krok et al.[8]

$$\log K_{as} = 5.6 - 3.1\log[Na^+] \quad (6)$$

Here, K_{as} is the equilibrium constant for TetR:*tetO2* binding, and $[Na^+]$ is the sodium concentration, in M. In applying eq 6, we set the sodium chloride concentration to 0.1 M, the same concentration used in our simulations. Using the $\Delta\Delta G$ values given in Table 1 above, the free energies of binding ΔG_x of the four mutant operators were also estimated (where the subscript x indicates one of M3, M4, M5, or M6). Finally, values of $\Delta G_x/\Delta G_{wild} \times 100\%$ are plotted in Figure 5, which allows for a clear comparison to the percent repression data discussed earlier.

The key features common to all of the experimental data discussed above are also found in our simulation data. In particular, the wild-type sequence shows the strongest affinity for the TetR repressor protein, whereas the M4 and M5 mutants are stronger than the M3 and M6 mutants. The match is not quantitatively exact, but this is not surprising, at least when considering these two factors: (a) the challenge of computing differences in free energies of binding upon mutation and (b) the inferential nature of the experimental measurements.

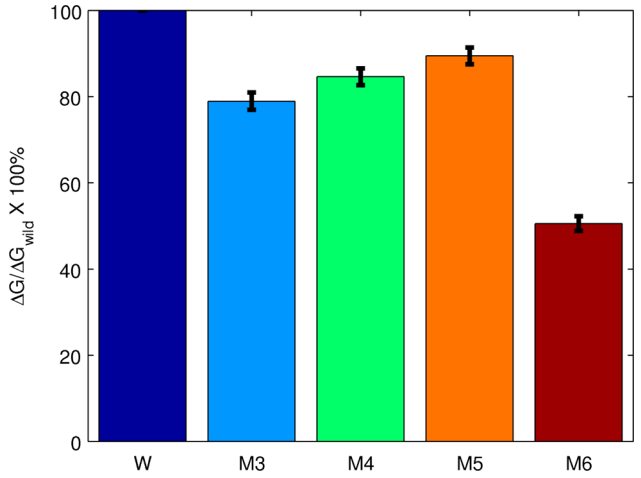


Figure 5. Computed free energies for all mutants. Error bars indicate only sampling error. The results are reported as $G_x/\Delta G_{wild} \times 100\%$, where ΔG_x is the free energy of binding for a particular mutant (x = M3, M4, M5, M6).

The simulations only address the binding of the TetR protein to the *tetO2* sequence, and span time scales of tens of nanoseconds. In the biological setting, the GFP repression levels that are ultimately measured are influenced by many other factors, including the binding of RNA polymerase, subsequent transcription and translation steps, as well as cell culture growth, all of which take place on time scales of minutes to hours. Considering these and many other differences between the simulation and experimental systems, our simulation results are surprisingly successful at predicting the relative ranking of the binding affinity and repressor strength of various *tetO2* sequences. Table 2 summarizes all the data presented above in terms of the ranking of different *tetO2* sequences according to repression strength or TetR:*tetO2* binding affinities.

3.3. Error Analysis of MD and FEP Simulations. Because all free energy calculation methods are known to suffer from convergence and sampling errors, it behooves us to address these issues as they pertain to the present work. We discuss sources of errors related to the equilibration of our systems, sampling during the FEP simulations and convergence of the reported free energy values.

As mentioned earlier, the structure of the wild-type TetR:*tetO2* complex was obtained using homology modeling and equilibrated during approximately 125 ns of MD simulation prior to the FEP calculations. Figure 6 shows the plot of the root-mean-square deviation of the protein and DNA backbone atoms from the initial structure. We only show data for the TetR:*tetO2* complex simulations, since structural equilibration is not an issue in the case of *tetO2*-only simulations in comparison to the protein:DNA case.

The goal of these simulations was to produce a stable, equilibrated structure as a starting point for FEP calculations so that no major conformational changes of the protein or DNA would take place that could strongly affect the FEP results. As shown by the RMSD plots, the structure appears to have stabilized well before the end of the simulations. Furthermore, with the possible exception of M4, none of the other mutants show any significant drift in the RMSD, indicating that no major structural changes are taking place. The small magnitude of the

dx.doi.org/10.1021/je500225x | J. Chem. Eng. Data 2014, 59, 3167−3176

Table 2. Summary of All Simulation and Experimental Data by Rank for the Four Mutants Tested[a]

rank	Sizemore et al. β-galactosidase	Sizemore et al. galactokinase	present work experiments TNN (3/6/9 h)	present work experiments TTN (3/6/9 h)	FEP simulations
1	W	W	**W/W/W**	**W/W/W**	**W**
2	M5	M5	**M5/M4/M4**	**M5/M4/M5**	**M5**
3	M4	M4 (=M5)	**M4/M5/M6**	**M4/M5/M4**	**M4**
4	M6	M6	**M6/M3/M3**	**M6/M3/M6**	**M3**
5	M3	M3	**M3/M6/M5**	**M3/M6/M3**	**M6**

[a]Mutant sequences are shown in greater detail in Figure 1.

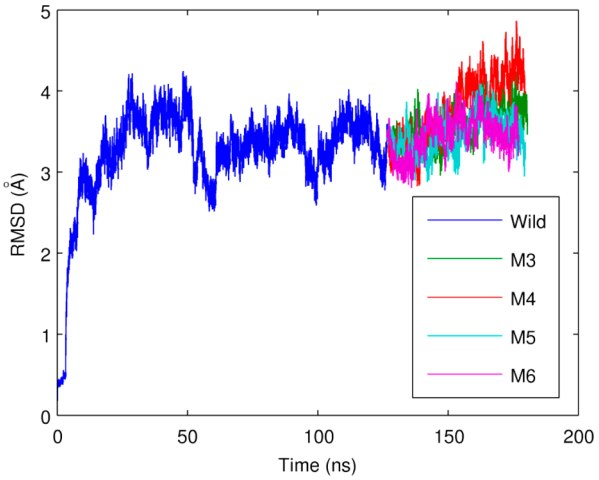


Figure 6. Root mean squared deviation of the protein and DNA backbone atoms from the starting structure for the wild-type equilibration simulation (blue; no FEP) as well as the mutants during FEP simulations. In all cases, best-fit alignments of the structures were carried out prior to calculation of RMSDs.

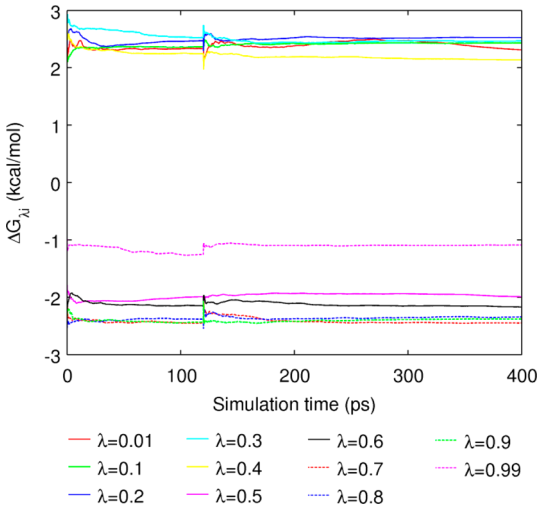


Figure 7. Plots of the cumulative average of ΔG_{λ_i} for several representative windows from the M3 TetR:*tetO2* simulation system during FEP simulations. Similar plots are obtained for all other mutants as well as the *tetO2*-only simulations.

drift in the case of M4 is likely within the equilibrium fluctuation range and does not pose a major concern for FEP results. The final coordinates and velocities from the wild-type MD simulation were used to start the FEP simulations for all four mutants. In this manner, we also avoid perturbing our systems by reheating, which could introduce a source of inconsistency among the different mutants.

We now turn our attention to the sampling and convergence errors from the free energy calculations. This is a topic that has received a great deal of attention and has recently been thoroughly reviewed in the context of FEP by Pohorille et al.[28] We argue that the simulation window length we have selected does indeed lead to converged free energy values, we show that the number and frequency of our FEP simulation windows provide sufficient overlap in the probability densities of adjoining intervals, and we provide an estimate of the statistical error of our computed free energy differences.

The ensemble average free energy for a particular window is given by

$$\Delta G_{\lambda_i} = -k_B T \ln\langle \exp(-(H_{\lambda_{i+1}} - H_{\lambda_i})/k_B T)\rangle_{\lambda_i} \quad (7)$$

The overall free energy change for a particular process where λ increases from 0 to 1 is simply the sum of individual free energies, as discussed in section 2.3 (see eq 4). In Figure 7, we show cumulative averages of ΔG_{λ_i} values for several representative simulation windows as a function of simulation time in each interval.

Although slight drifts in the plots in Figure 7 are occasionally observed, the results generally show excellent convergence. The discontinuity at approximately 120 ps corresponds to the end of the equilibration period, where the cumulative average is effectively reset. The final data reported (Table 1) are based on the last 280 ps of simulation time for all windows.

Several sophisticated techniques for the estimation of sampling and bias errors in free energy calculations have been developed.[27−30] In many of these methods, forward sampling as well as reverse sampling are required to obtain the most accurate estimates of sampling error; however, it is not clear that the forward and reverse transformations share the same convergence characteristics, and the resulting error estimate may not be relevant.[31] In the present work, we therefore restrict ourselves to methods that only require forward sampling. In particular, we show that the probability distributions of $\Delta U_{\lambda_i} = H_{\lambda_{i+1}} - H_{\lambda_i}$ are sufficiently narrow and closely approximated by Gaussian distributions, both of which are characteristics typical of situations where states λ_i and λ_{i+1} are sufficiently similar for good sampling during FEP simulations. For a more detailed discussion of these ideas, the interested reader is referred to refs 27, 28, and 30.

As shown in Figure 8, the probability distributions $P(\Delta U_{\lambda_i})$ are closely Gaussian, with standard deviations $\sigma_{\Delta U}$ on the order of 0.2 kcal mol^{-1}, or approximately 0.33 $k_B T$. The recommended maximum widths of $P(\Delta U_{\lambda_i})$ in FEP simulations are on the order of 1 to 2 $k_B T$, so we feel confident even without an analysis of the overlap between forward and reverse simulations that our results correspond to well-sampled distributions.

dx.doi.org/10.1021/je500225x | J. Chem. Eng. Data 2014, 59, 3167−3176

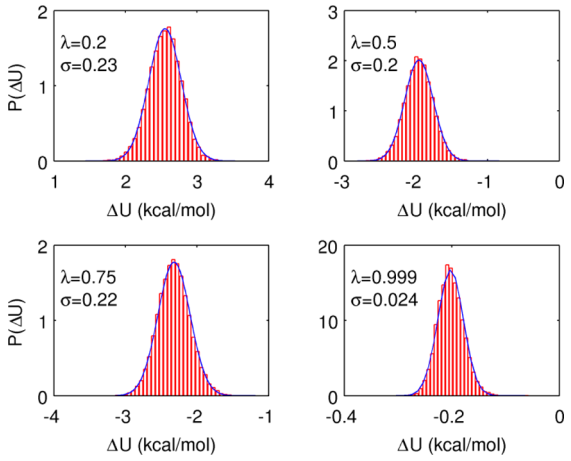


Figure 8. Density of states for representative simulation windows. Histograms from the simulations are shown in red, whereas corresponding Gaussian distributions are shown in blue. The vast majority of FEP windows across all mutants and λ-values show nearly perfect Gaussian distributions, indicative of well-sampled distributions with high overlap between adjacent states.

Finally, we analyze the sampling errors associated with the estimator of the free energy in eq 2. These are the errors reported in Table 1. On the basis of the computed autocorrelation functions of each time series of ΔU_{λ_i} (data not shown), we have estimated a generous upper bound for the correlation time of these data to be approximately 20 ps. As such, we apply a block averaging technique to the equilibrated portion of each ΔU_{λ_i} time series, where the block size is 20 ps. The averages of these blocks are then used as uncorrelated, independent variables to compute the variance of ΔU_{λ_i} at each λ_i value, according to a standard first-order delta method[28]

$$\sigma^2_{\Delta G_{\lambda_i}} = \frac{\langle \exp(-2\beta\Delta U)_{\lambda_i} \rangle}{N_0\beta^2 \langle \exp(-\beta\Delta U) \rangle_{\lambda_i}} - \frac{1}{N_0\beta} \quad (8)$$

The variance of the mean for the reported free energy values is then calculated as the sum of the variances at all λ_i values: $\sigma^2_{\Delta G} = \sum_\lambda \sigma^2_{\Delta G_{\lambda i}}$. The resulting standard deviation is reported as the error in all the values in Table 1, as well as the error bars in Figure 6. Note that this only represents statistical sampling error, and says nothing of biasing errors due to, for example, poor convergence or force field errors.

3.4. Analysis of MD Simulations: Counterion Release, Salt Bridges, and Hydrogen Bonding. In addition to free energies of binding as computed through FEP calculations, we also attempt to explain our results in terms of specific biophysical interactions. Our analysis here focuses on quantifiable differences among the mutants that can explain the observed differences in binding affinities. A more general analysis of the binding mechanism between TetR and *tetO2* is beyond the scope of the present work and has been previously investigated both experimentally[15] and via MD simulations.[32]

Mascotti and Lohman[33] found that a major contribution to the binding free energy associated with most protein−nucleic acid complexes is the increase in entropy due to counterion release from the nucleic acid. We therefore report a brief analysis of counterion release and related electrostatic interactions based on our MD simulations. In particular, we focus on the counterion release, salt bridge formation, and hydrogen bond formation involved in TetR:*tetO2* binding.

First, we investigate the formation of ionic bridges between the TetR protein and the *tetO2* DNA segment. We distinguish two types of ionic bridges: direct contact pairs (CP) and Na^+ ion separated pairs (ISP). CP ionic bridges form between cationic arginine or lysine residues and negatively charged phosphate oxygens of the DNA. We analyze the mean number of such ionic bridges between TetR and *tetO2* in the case of the wild-type *tetO2* sequence as well as the four mutants. In the case of the wild-type the analysis was performed on the MD simulation carried out for equilibration purposes; in the case of the mutants, additional simulations of at least 10 ns were carried out for each mutant after the FEP alchemical transformations (λ = 1 state).

A CP ionic bridge is considered to be formed when any arginine guanidinium group ($R{-}NHC(NH_2)_2^+$) is located within 3 Å of any of the anionic DNA phosphate oxygens. ISP ionic bridges are defined by the number of Na^+ counterions bound to both protein and DNA within 3 Å of the negatively charged oxygen of Glu^- residues and any of the anionic DNA oxygens. The results are summarized in Table 3, along with other results from the analysis of the MD simulations.

Table 3. Analysis of Various Interactions between TetR and *tetO2* from All Simulations[a]

number of contacts	CP	ion release	ISP	H bonds	ΔG	$\Delta\Delta G$
					kcal mol^{-1}	kcal mol^{-1}
W	2.8	2.5	2.8	19.4	**53.1**	**0**
M3	2.1	0.5	2.3	17.5	**43.0**	**10.1**
M4	1.9	1.2	1.7	16.1	**40.7**	**12.4**
M5	2.3	2	2.1	20.8	**52.7**	**0.38**
M6	2.9	1.1	3.4	18.4	**49.2**	**3.9**
energy (kcal mol^{-1}) per contact	**4**	**1.8**	**0.2**	**1.9**		
reference for energy value	35	36	37	38		

[a]Table headings refer to direct contact pairs (CP), ion separated pairs (ISP), and hydrogen-bonds (H-bonds) defined such that the donor and acceptor atom are closer than 2.4 Å.

We find that the tetR protein when bound to *tetO2* decreases the number of counterions bound to the DNA. For our system, we define the existence of the bound state for counterions as any Na^+ ion found within 6 Å of the DNA. The resulting averages are shown in Table 3. We find that on average, 16 Na^+ counterions are bound to the wild-type DNA in the absence of the protein. In the last row of the table, we present an approximate estimate of the free energy gain associated with counterion release, CP formation, ISP contacts, and hydrogen bonds for different mutants based on various references listed in the table.

We also analyze the hydrogen bonds between tetR and *tetO2*. We identify a hydrogen bond when a pair of hydrogen bond donor and acceptor atoms are closer than 2.4 Å,[34] regardless of the donor−hydrogen−acceptor angle. The numbers of hydrogen bonds averaged over the last 2 ns of all simulations are presented in Table 3. The most common hydrogen bonds form between Lys-48, Thr-26/27 and various oxygen atoms of DNA. The energy of such hydrogen bonds is approximately 1.9 kcal mol^{-1}.

Clearly, the overall energy estimates based on these data are not in agreement with the FEP simulation results. This is not surprising, considering the extremely approximate nature of this

dx.doi.org/10.1021/je500225x | J. Chem. Eng. Data 2014, 59, 3167−3176

analysis, and the many other factors not considered (e.g., water molecules, long-range forces, protein and DNA conformational entropies, etc.). The counterion release data appear to suggest that a larger number of counterions being released leads to stronger binding and is fairly consistent with both the FEP data and the experimental data, at least with regards to overall mutant ranking. Aside from the counterion release data, we do not observe any correlation between any particular type of contact, suggesting that there is not one dominant mechanism for binding, but rather that multiple coupled phenomena are involved. In particular, we note that hydrogen bonding appears to be a poor indicator of TetR:*tetO2* interactions, at least with regards to the differences observed among the mutants tested here.

4. SUMMARY AND CONCLUSIONS

The combined modeling and experimental work presented herein demonstrate the potential of using rigorous molecular simulations to study binding free energies that dictate biological behavior. The binding of TetR to *tetO2* is a key step in the function of the tet operon and has potential applications in any synthetic biological system that requires gene expression to be controlled by the addition or removal of a small molecule (in this case tetracycline). Although we have not investigated the case of the induced TetR protein, it seems unlikely that mutations in *tetO2* would cause any notable changes in the extremely weak binding of induced TetR to *tetO2*; certainly these changes would be overshadowed by the effects of *tetO2* mutations on the binding of the uninduced protein. As such, we believe we have isolated a key biophysical process in the operation and modulation of *tetO2*.

We have focused our analysis of the simulations on the free energy results. Molecular dynamics simulations also provide a wealth of information on structural properties and atomistic interactions, some of which we have also briefly presented. There does not appear to be a specific molecular part, mechanism of interaction, or binding that can explain the differences observed among the mutants that we tested. It therefore appears that the more computationally intensive free energy simulations are indeed necessary because the binding differences are likely a result of a complex combination of several competing physical phenomena.

Although free energy perturbation results cannot predict absolute GFP expression levels, they appear to be capable of predicting relative rankings of *tetO2* mutants. We have analyzed the possible sources of errors in our FEP simulations, and we believe that the results are not spurious with regards to sampling or convergence bias. Considering the complex nature of these systems as well as the inconsistencies in ranking even among different experiments, we believe the results presented make a strong argument for the use of molecular simulations in synthetic biology. Future improvements in molecular force fields and increasingly more powerful computers in the near future hold the promise of enabling the use of free energy simulations for the quantitative prediction and design of gene repressor protein systems as well as other biomolecular applications.

AUTHOR INFORMATION

Corresponding Author

*E-mail: yiannis@umn.edu. Phone: 612-624-4197 Fax: 612-626-7246. Address: 253 Amundson Hall 421 Washington Ave SE Minneapolis, MN 55455, United States.

Present Addresses

[†]Currently at Sandia National Laboratories.
[‡]Currently at Princeton University.
[§]Currently at National Research Council Canada.
[∥]Currently at Birzeit University.
[⊥]Currently at University of Chicago.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by a grant from the NIH (GM 070989). It was also partially supported by the National Center for Supercomputing Applications under grant TG-MCA04N033 and utilized the Kraken and Abe resources. Computational support from the Minnesota Supercomputing Institute (MSI) and the University of Minnesota Digital Technology Center (DTC) is also gratefully acknowledged.

REFERENCES

(1) Benner, S. A.; Sismour, A. M. Synthetic biology. *Nat. Rev. Genet.* **2005**, *6*, 533−543.
(2) Kaznessis, Y. Models for Synthetic Biology. *BMC Syst. Biol.* **2009**, *1*, 47.
(3) Sotiropoulos, V.; Kaznessis, Y. N. Synthetic tetracycline-inducible regulatory networks: computer-aided design of dynamic phenotypes. *BMC Syst. Biol.* **2007**, *1*, 7.
(4) Weeding, E.; Houle, J.; Kaznessis, Y. SynBioSS designer: a web-based tool for the automated generation of kinetic models for synthetic biological constructs. *Briefings Bioinf.* **2010**, *11*, 394−402.
(5) Sizemore, C.; Wissmann, A.; Gulland, U.; Hillen, W. Quantitative analysis of Tn10 Tet repressor binding to a complete set of tet operator mutants. *Nucleic Acids Res.* **1990**, *18*, 2875−2880.
(6) Ramalingam, K. I.; Tomshine, J. R.; Maynard, J. A.; Kaznessis, Y. N. Forward engineering of synthetic bio-logical AND gates. *Biochem. Eng. J.* **2009**, *47*, 38−47.
(7) Lutz, R.; Bujard, H. Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res.* **1997**, *25*, 1203−1210.
(8) Kedracka-Krok, S.; Wasylewski, Z. Kinetics and equilibrium studies of Tet repressor-operator interaction. *J. Protein Chem.* **1999**, *18*, 117−125.
(9) Sambrook, J.; Fritsch, E. F.; Maniatis, T. *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 1989.
(10) Cox, R. S., 3rd; Surette, M. G.; Elowitz, M. B. Programming gene expression with combinatorial promoters. *Mol. Syst. Biol.* **2007**, *3*, 145.
(11) Horton, R. M.; Cai, Z. L.; Ho, S. N.; Pease, L. R. Gene splicing by overlap extension: tailor-made genes using the polymerase chain reaction. *Biotechniques* **1990**, *8*, 528−535.
(12) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403−410.
(13) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(14) Luckner, S. R.; Klotzsche, M.; Berens, C.; Hillen, W.; Muller, Y. A. How an agonist peptide mimics the antibiotic tetracycline to induce Tet-repressor. *J. Mol. Biol.* **2007**, *368*, 780−790.
(15) Orth, P.; Schnappinger, D.; Hillen, W.; Saenger, W.; Hinrichs, W. Structural basis of gene regulation by the tetracycline inducible Tet repressor-operator system. *Nat. Struct. Biol.* **2000**, *7*, 215−219.
(16) *Schrödinger Release 2011-1*, Schrödinger, LLC: New York, NY, 2011.
(17) Jorgensen, W. L. Quantum and statistical mechanical studies of liquids. 10. Transferable intermolecular potential functions for water, alcohols, and ethers. Application to liquid water. *J. Am. Chem. Soc.* **1981**, *103*, 335−340.

(18) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187−217.
(19) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781−1802.
(20) MacKerell, J.; A, D.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(21) MacKerell, A. D., Jr.; Feig, M.; Brooks, C. L., III Improved treatment of the protein backbone in empirical force fields. *J. Am. Chem. Soc.* **2004**, *126*, 698−699.
(22) Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys.* **1995**, *103*, 4613−4621.
(23) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577−93.
(24) Ryckaert, J. P.; Ciccotti, G.; Berendesen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.
(25) Zwanzig, R. W. High-Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *J. Chem. Phys.* **1954**, *22*, 1420.
(26) Beveridge, D. L.; DiCapua, F. M. Free energy via molecular simulation: applications to chemical and biomolecular systems. *Annu. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 431−492.
(27) Chipot, C.; Pohorille, A. *Free Energy Calculations*; Springer: Berlin, 2007.
(28) Pohorille, A.; Jarzynski, C.; Chipot, C. Good practices in free-energy calculations. *J. Phys. Chem. B* **2010**, *114*, 10235−10253.
(29) Shirts, M. R.; Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **2008**, *129* (12), 124105.
(30) Chodera, J. D.; Mobley, D. L.; Shirts, M. R.; Dixon, R. W.; Branson, K.; Pande, V. S. Alchemical free energy methods for drug discovery: progress and challenges. *Curr. Opin. Struc. Biol.* **2011**, *21*, 150−160.
(31) Chipot, C.; Rozanska, X.; Dixit, S. B. Can. free energy calculations be fast and accurate at the same time? Binding of low-affinity, non-peptide inhibitors to the SH2 domain of the src protein. *J. Comput.-Aided Mol. Des.* **2005**, *19*, 765−770.
(32) Aleksandrov, A.; Schuldt, L.; Hinrichs, W.; Simonson, T. Tet repressor induction by tetracycline: a molecular dynamics, continuum electrostatics, and crystallographic study. *J. Mol. Biol.* **2008**, *378*, 898−912.
(33) Mascotti, D. P.; Lohman, T. M. P Thermodynamic extent of counterion release upon binding oligolysines to single-stranded nucleic acids. *Natl. Acad. Sci. U. S. A.* **1990**, *87*, 3142−3146.
(34) De Loof, H.; Nilsson, L.; Rigler, R. Molecular Dynamics Simulation of Galanin in Aqueous and Nonaqueous Solution. *J. Am. Chem. Soc.* **1992**, *114*, 4028−4035.
(35) Anderson, D. E.; Becktel, W. J.; Dahlquist, F. W. pH-induced denaturation of proteins: a single salt bridge contributes 3−5 kcal/mol to the free energy of folding of T4 lysozyme. *Biochemistry* **1990**, *29*, 2403−2408.
(36) Briscoe, W. H.; Attard, P. Counterion-only electric double layer: a constrained entropy approach. *J. Chem. Phys.* **2002**, *117*, 5452−5464.
(37) Bloomfield, V. A. DNA condensation by multivalent cations. *Biopolymers* **1997**, *44*, 269−282.
(38) Jeffrey, G. *An Introduction to Hydrogen Bonding*; Oxford University Press: New York, 1997.