FOX01 Transcription Factor Folding Landscape Elucidates The Role of Disease Mutations

Dylan Novack[1], Lei Qian[2], Richard H.G. Baxter[2], Vincent A. Voelz[1].

[1]Department of Chemistry, Temple University, Philadelphia, PA, USA,

[2]Department of Medical Genetics and Molecular Biochemistry, Lewis Katz School of Medicine, Temple University, Philadelphia, PA, USA.

Introduction

Diffuse large B-cell lymphoma (DLBCL), a particularly aggressive cancer, is the most common Non-Hodgkin's lymphoma found among adult patients. Fatal if left untreated, approximately 40% of patients either have a treatment resistant form of the cancer or relapse after treatment. Mutants of the human forkhead transcription factor FOX01, have been found in DLBCL cell lines and associated with decreased survival in patients. The loss of function associated with these disease mutations is poorly understood. Thermal stability assays have shown that certain mutants destabilize the protein, while others have no effect or stabilize the protein.

Using GPU-accelerated *ab initio* folding simulations on the distributed computing platform Folding@home (F@H) we obtain information on the folding landscape of FOX01. In the absence of experimental folding kinetics, comparative contact order analysis suggested that FOX01 folds on the order of tens of microseconds, making GPU simulation a tractable approach. Understanding how mutations affect the regulatory function of this transcription factor will help elucidate the role of FOX01 in the development of DLBCL and may provide insight for the development of new therapeutics.

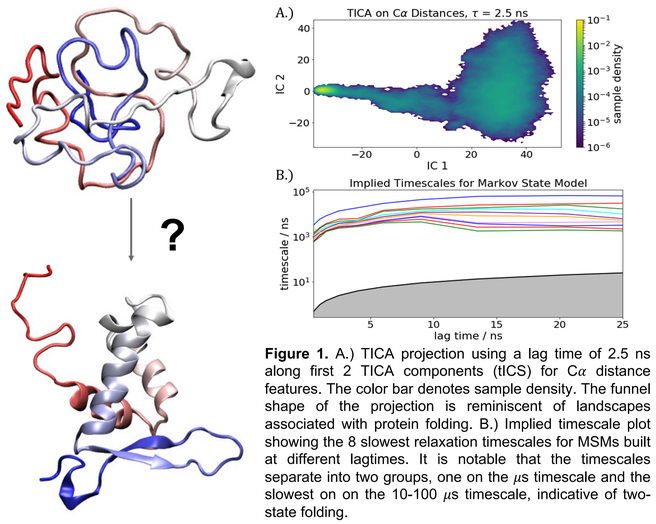

Figure 1. A.) TICA projection using a lag time of 2.5 ns along first 2 TICA components (tICS) for Cα distance features. The color bar denotes sample density. The funnel shape of the projection is reminiscent of landscapes associated with protein folding. B.) Implied timescale plot showing the 8 slowest relaxation timescales for MSMs built at different lagtimes. It is notable that the timescales separate into two groups, one on the μs timescale and the slowest on on the 10-100 μs timescale, indicative of two-state folding.

Methods

System Preparation

- The model of FOX01 was adapted from PDB 3CO6. Protonation states at pH 7.5 were determined using H++, and the AMBER14SB forcefield was used to build the topology.
- The protein was solvated in a cubic periodic box with TIP3P waters, with Na^+ and Cl^- counterions added at 0.1 M to neutralize the system, resulting in a system of ~64500 atoms.

Molecular Dynamics (MD) Simulations

- All simulations were performed using GPU-accelerated OpenMM, using stochastic (Langevin) integration with a 2 fs time step and PME electrostatics.
- To generate unfolding conformations: Multiple 60 ns NPT trajectories were performed at 300K, 400K, 450K, and 498K, respectively. Conformational clustering using a *k*-centers algorithm was used to identify a variety of folded and unfolded conformations.
- Production runs were performed using OpenMM on the Folding@home distributed computing platform, with all coordinates saved every 0.5 ns. Trajectories were initiated from 20 different starting structures at 375K, each in replicates of 500 clones with randomized initial velocities, totaling 10000 independent trajectories.

Markov Model Construction & Analysis

- Using PYEMMA 2.5.7, a subset of 497 trajectories trajectories was featurized using every eighth Cα pairwise distance for a total of 512 features.
- Time-lagged Independent Component Analysis (TICA) was performed using a lag-time (τ) of 2.5 ns. Using K-means clustering, 100 cluster centers were assigned to the TICA output.
- The MSM was built using a lag time of τ = 4.0 ns.

Results

Folding Landscape of FOX01

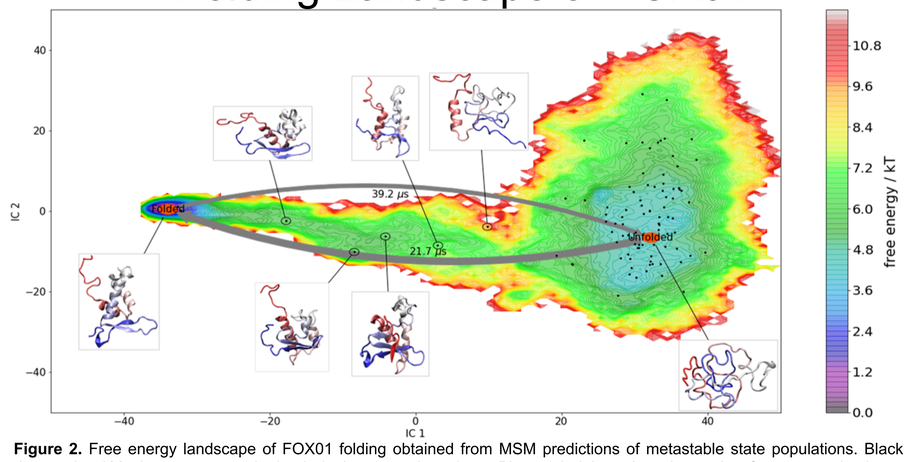

Figure 2. Free energy landscape of FOX01 folding obtained from MSM predictions of metastable state populations. Black dots denote the 100 cluster centers obtained via *k*-means clustering. Blue and violet colors denote low-free energy basins on the folding landscape. There are two such basins along the tIC1 axis, which suggest macroscopic "two-state" folding. The landscape was further coarse-grained using PCCA++, for the estimation of transition rates. Arrows show the effective folding and unfolding times, 21.7 μs and 39.2 μs, respectively Structures shown on the plot are MSM microstate centers representative of (from right to left) an unfolded state, 5 transitional states, and the folded state.

Differential Scanning Calorimetry of Selected Mutants

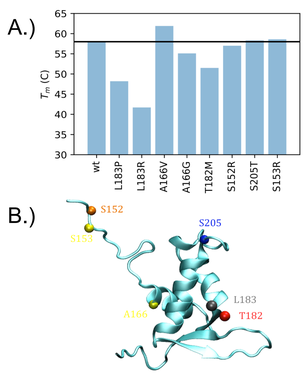

Figure 3. A.) DSC Tm data for wt FOX01 and eight mutants. The T_m for the wild-type protein is ~58º C. B.) Crystal structure (PDB 3CO6) of FOX01 with T_m mutation locations as colored spheres.

FEP probes how L183R affects FOX01 folding and DNA-binding

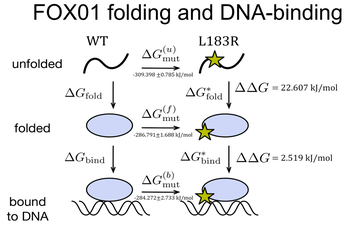

Figure 4. Thermodynamic cycle and results for alchemical FEP calculations of L183R. The unfolded state is modeled as T(L/R)S tripeptide.

Average SASA Values for Markov States

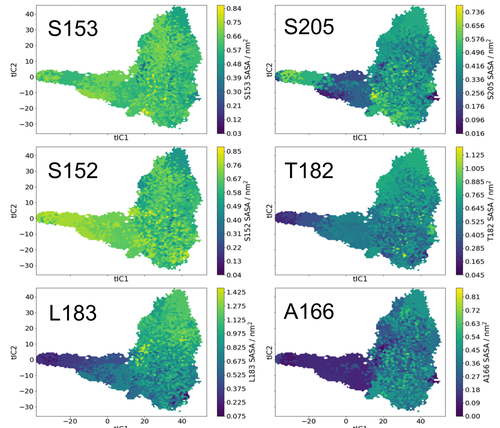

Figure 5. Average Solvent Accessible Surface Areas (SASA) of disease-associated mutations shown for each MSM state.

Δ Avg. SASA for Folded and Unfolded States

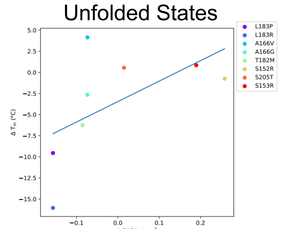

Figure 6. Change in experimental T_m as a function of the change in population weighted avg SASA value in nm^2.

Discussion

Massively parallel ab initio folding trajectories reveal a funnel-like folding landscape for FOXO1. TICA projection of C_α distance features (**Figure 1 A.**) reveals two low-free energy basins along IC1, indicating two-state folding. The gap in MSM implied timescales also indicate two-state folding, with relaxation on the 10-100 μs time scale, and a predicted folding time of~20 μs at 375 K. **Figure 2** is a visualization of the protein folding free energy landscape. The free energy basin with high tIC1 values represents the unfolded state. As the reaction progress to the left, along tIC1, the N-terminal helix appears to form. Progressing further, the second helix forms, along with the C-terminal beta sheet. Finally, the third helix forms to reach the native folded state. Using an inverse mean first passage time (MFPT) approach, the rates of folding and unfolding were calculated to be 21.7 μs and 39.2 μs respectively.

Alchemical free energy perturbation (FEP) calculations were performed for the L183R mutation using Gromacs/2016.3. The protein force field was ffamber99sb-star-ildn , and the metal-coordination site on FOXO1 was parameterized with MCPB.py (Li et al.) from Ambertools17.6. Alchemical topologies were built with the *pmx* webserver (Gapsys et al.) Each of 21 lambda values (0.0, 0.05, … 1.0) was simulated for 1 ns.

Several disease-related mutations likely perturb FOXO1 folding, rather than disrupt DNA binding. The FEP results show a ΔΔG of folding of 22.61 kJ/mol, and ΔΔG of DNA binding of 2.519 kJ/mol, suggesting disruption of folding is likely responsible for the loss of function (**Figure 4**). This is consistent with melting temperatures (T_m) measured by DSC, which show that A166G, L183P, and L183R mutations decrease the T_m by ~ 3 C, ~10 C, and ~17 C, respectively. L183P and L183R mutants destabilize the protein the most out of those tested. As a simple preliminary analysis, we used the MSM of FOXO1 folding to calculate the change in solvent-accessible surface area (ΔSASA) upon folding (**Figure 5**) for all mutant residues (**Figure 3 B**). Comparison of ΔSASA to the experimentally measured T_m show reasonable correlation, suggesting that mutants that disrupt residues in hydrophobic cores are more destabilizing (**Figure 6**). While A166G is slightly destabilizing, A166V actually stabilizes the protein, increasing the T_m by ~ 3 C. A166 is solvent-accessible only in the unfolded state, so a mutation to a bulkier hydrophobic residue further reward burial of A166V in the native state.

Conclusions

- Ab-initio folding of a ~20μs folding protein.
- The MSM model predicts folding may be precipitated by the formation of the third helix going into the beta sheet motif.
- Inverse MFPT calculation of effective rates yielded rates of 21.7 μs and 39.2 μs, respectively, for unfolding and folding at 375 K.
- FEP shows that the L183R mutation disrupts the folding by ΔΔG = 22.607 kJ/mol
- SASA analysis shows that L183 is buried in the hydrophobic core in the folded state, but fairly solvent accessible in both unfolded and transition states.
- L183R mutation likely disrupts the hydrophobic interactions at this locus due to hydrophilicity, L183P mutation
- Contrasting effects of A166V and A166G mutations are likely due to the relative hydrophobicity of the mutant residues, either stabilizing or disrupting the hydrophobic core, respectively.

Citations

Bremt M.M., et. al. *Structure* (2008) 16, 1407–1416
Zwanzig R. W. J. Chem. Phys. **22**, 1420 (1954)
Abraham M. J. et. al. *SoftwareX* (2015) 1–2:19–25
Shrake, A; Rupley, JA. (1973) J Mol Biol 79 (2): 351–71.
Eastman P,. et al. *PLoS Comput. Biol.* (2017) 13:e1005659
Hoel, P G et. al. 1972. Intro. to Stochastic Processes
Maier J.A. , et al.*J. Chem. Theory Comput.* (2015) 11, 3696-3713.
Anandakrishnan R., et al. *Nucleic Acids Res.*(2012) 40(W1):W537-541.
Scherer M.K, J. Chem. Theory Comput. 2015, 11, 11, 5525-5542
Gapsys, Vytautas, and Bert L. de Groot. J. Chem. Inf. Model. 57(2) 109–14.

Acknowledgements

We thank the participants of Folding@home, without whom this work would not be possible. D.N. and V.A.V. were supported was supported by NIH 1R01GM123296. This work was supported in part by the National Institute of General Medical Sciences (1R01GM114358) of the National Institutes of Health to RHGB.Temple HPC resources (Owlsnest) were supported by NSF CNS-162506 and US Army Research Laboratory W911NF-16-2-0189. Computing resource CB2RR is supported by NIH S10-OD020095.