



Published in final edited form as:

Biochemistry. 2022 August 16; 61(16): 1669–1682. doi:10.1021/acs.biochem.2c00224.

Oncogenic mutations in the DNA-binding domain of FOXO1 that disrupt folding: quantitative insights from experiments and molecular simulations

Dylan Novack^{*,1}, Lei Qian^{*,2}, Gwyneth Acker², Vincent A. Voelz¹, Richard H. G. Baxter^{2,*}

¹Department of Chemistry, Temple University, Philadelphia, Pennsylvania 19122, USA.

²Department of Medical Genetics & Molecular Biochemistry, Lewis Katz School of Medicine at Temple University, 3440 North Broad Street, Philadelphia, PA, 19140, USA.

Abstract

FOXO1, a member of the family of winged-helix motif Forkhead box (FOX) transcription factors, is the most abundantly expressed FOXO member in mature B-cells. Sequencing of diffuse large B-cell lymphoma (DLBCL) tumors and cell lines identified specific mutations in the forkhead domain linked to loss of function. Differential scanning calorimetry and thermal shift assays were used to characterize how eight of these mutations affect the stability of the FOX domain. Mutations L183P and L183R were found to be particularly destabilizing. Electrophoresis mobility shift assays show these same mutations also disrupt FOXO1 binding to their canonical DNA sequences, suggesting the loss of function is due to destabilization of the folded structure. Computational modeling of the effects of mutations on FOXO1 folding was performed using alchemical free energy perturbation (FEP), and a Markov model of the entire folding reaction was constructed from massively parallel molecular simulations, which predicts folding pathways involving the late folding of helix α_3 . Although FEP is able to qualitatively predict the destabilization from L183 mutations, we find that a simple hydrophobic transfer model, combined with estimates of unfolded-state solvent accessible surface areas from molecular simulation, is able to more accurately predict changes in folding free energies due to mutations. These results suggest that atomic detail provided from simulation is important for accurate prediction of mutational effects on folding stability. Corresponding disease-associated mutations in other FOX family members support further experimental and computational studies of the folding mechanism of FOX domains.

Introduction

Forkhead box (FOX) transcription factors are a family of DNA-binding proteins containing a winged-helix motif, a variation of the helix-turn-helix (HTH) motif.¹ FOX proteins are

*These authors contributed equally

Supplemental Material

Supplemental Tables S1 and S2, and supplemental Figures S1–S9.

Supplemental Movies M1–M8.

Accession Codes

FOXO1: UniProtKB Q12778

conserved from *Drosophila* to humans with multiple roles in development and regulation, and mutations in these genes are associated with multiple pathologies including cancer.^{2,3} FOXO1 is a key factor in both insulin signaling and B-cell development, balancing apoptotic and survival signals that vary across tissues and in response to metabolic conditions.⁴

FOXO1 has an N-terminal region, a FOX DNA-binding domain, a nuclear localization signal (NLS), several phosphorylation sites, a nuclear export sequence (NES) and a C-terminal transactivation domain (TAD) (Figure 1A). Nuclear FOXO1 binds to regulatory sites and induces gene expression. Export to the cytoplasm, stimulated by AKT (Ser/Thr protein kinase) phosphorylation leads to ubiquitination and degradation.⁵ The DNA-binding domain has a compact three-helix fold, with the third helix ($\alpha 3$) sitting in the major groove of B-form DNA, and C-terminal β strands projecting along the axis of the DNA to contact one or both of the adjacent minor grooves.⁶

FOXO1 is the most abundantly expressed FOXO member in mature B-cells, where it targets genes in apoptosis, cell cycle and growth arrest.⁷ Tonic B-cell receptor (BCR) signaling stimulates the PI3K/AKT pathway via phosphorylated kinase BTK (pBTK) and helps release this brake by cytoplasmic localization of FOXO1. Diffuse large B-cell lymphoma (DLBCL), the most common non-Hodgkin lymphoma among adults,⁸ is fatal if untreated, with ~40% of patients either unresponsive or relapsing after front-line chemotherapy treatment.⁹ FOXO1 is frequently mutated in DLBCL, and oncogenic driver mutations are associated with cases refractory to, or relapsing after, treatment.¹⁰⁻¹²

Sequencing of DLBCL tumors identified two mutation hotspots in FOXO1, the N-terminus and the FKH domain.¹⁰ N-terminal mutations in FOXO1 prevent AKT phosphorylation leading to nuclear retention, and are also prevalent and oncogenic in GC-derived tumors such as Burkitt's Lymphoma.^{13,14} The effects of mutations in the FOX domain, however, have not been investigated. These mutations are spread throughout the folded domain and do not involve residues directly involved in DNA-binding, but mutations at T182 or L183 are adjacent to $\alpha 3$ and the major groove (Fig. 1B).^{6,10} We therefore sought to investigate whether these mutations have a global effect on FKH folding that would disrupt DNA-binding.

To address this question, we performed a joint experimental and computational study of eight oncogenic mutations identified in the FOX domain. Through differential scanning calorimetry and thermal shift assays, we show these mutations destabilize the FOX domain, and in turn disrupt FOXO1 binding to the insulin response element (IRE). Then, as test of state-of-the-art methods to computationally predict the effects of mutations on folding, we compare two simulation-based approaches: (1) an alchemical free energy perturbation (FEP) approach, and (2) Markov state models (MSMs) constructed from massively parallel *ab initio* folding simulations to characterize the folding mechanism of FOXO1. While both approaches reasonably rank-order the effects of destabilizing mutations, we find that changes in per-residue solvent accessible surface area (SASA) extracted from the MSMs, combined with a simple empirical hydrophobicity-based model of protein stability, makes superior quantitative predictions of the ΔG of mutation. These results suggest that detailed

information about folding intermediates and unfolded-state structure provided by atomistic simulations is important for accurate prediction of mutational effects on folding stability.

Results

Oncogenic mutations affect the stability of the FKH domain

We considered eight FKH mutations identified in DLBCL patients by Trinh et al. (2013),¹⁰ and two mutations identified in DLBCL cell lines (Figure 1B, Table 1). None of these mutations correspond to residues that directly coordinate DNA bases in canonical DNA recognition sequences such as the insulin response element IRE (Figure 1B, top) or DAF-16 binding elements DBE1 and DBE2.⁶ We therefore hypothesized these mutations may disrupt DNA binding by destabilizing the FKH domain.

To test this idea, we performed a thermal shift assay (TSA) for FKH mutants,¹⁵ using three independent protein batches, three technical replicates per batch (Table 1). According to these assays, the wild-type (wt) FKH domain had melting temperature $T_m = 50.5 \pm 0.8$ °C. FKH mutants S152R, S153R, K171E, S205N, and S205T had T_m 2 °C compared to wt, implying no significant effect on FKH structure. However, significant alterations in T_m were observed for A166G, A166V, T182M, L183P, and L183R.

In the case of A166, the site of multiple mutations in DLBCL patients, replacement of alanine with the more flexible glycine reduced the melting temperature to 46.4 ± 0.5 °C ($T_m = -4.2$ °C) whereas substitution with less flexible valine increases the melting temperature to 56.3 ± 0.5 °C ($T_m = 5.8$ °C). However, the most significant mutations were at positions T182 and L183. These residues lie at the N-terminus of helix 2 (α_2) in the FKH domain. These residues stabilize the overall fold by initiating the first turn of α_2 , and packing against helix 3 (α_3) which sits in the major groove of bound DNA. When bound to DNA, these residues make water-mediated contacts to the phosphate backbone along with residues 234–235 (Figure 1B).

The mutation T182M—identified in DLBCL cell line OCI Ly8—reduced the melting temperature to 46.4 ± 0.1 °C ($T_m = -4.2$ °C) equivalent to A166G. L183P was more destabilizing, reducing the melting temperature to 40.0 ± 0.1 °C ($T_m = -10.5$ °C). A second mutation at the same position L183R – identified in DLBCL cell line WSU-NHL – reduced the melting temperature to 33.6 ± 1.1 °C ($T_m = -16.9$ °C).

TSA is an indirect measure of protein folding as detection depends on dye binding, which may itself influence the melting temperature. Also, the thermal shift assay does not capture the full folding transition as fluorescence above the melting temperature is affected by other non-folding artifacts. Hence we validated our results with an independent measure of protein stability, differential scanning calorimetry (DSC). Two batches of FKH wt and single batches of 8 of the 10 mutants were analyzed using a non two-state folding model to derive melting temperature, the calorimetric and the van't Hoff enthalpy of unfolding (Table 1).

Melting temperatures determined by DSC were systematically higher by 7 ± 1 °C compared to the thermal shift assay. However, T_m from DSC closely matched that of the thermal

shift assay, with a mean unsigned error (MUE) of 1.3 °C and maximum difference of 2.2 °C (Figure S1). The change in enthalpy (ΔH) associated with unfolding were similar between FKH wt and mutants. The wild-type calorimetric enthalpy of unfolding was $H_{\text{cal}}(\text{wt}) = 47 \pm 1 \text{ kcal mol}^{-1}$, while the average value for mutants was $H_{\text{cal}}(\text{mutants}) = 47 \pm 8 \text{ kcal mol}^{-1}$. Similarly, the van't Hoff enthalpies of unfolding were $H_{\text{vH}}(\text{wt}) = 54 \pm 1 \text{ kcal mol}^{-1}$ vs. $H_{\text{vH}}(\text{mutants}) = 57 \pm 8 \text{ kcal mol}^{-1}$. In general, the van't Hoff change in enthalpy was higher than the calorimetric enthalpy change by $10 \pm 10 \text{ kcal mol}^{-1}$, suggesting that reductions in melting temperature of FKH mutants result from perturbations to the folded or unfolded states, but not a gross change in the folding landscape of the FKH domain.

Mutations that reduce FKH stability also reduce DNA binding

We expected that FKH mutations identified in DLBCL cell lines and patient specimens which significantly destabilize the FKH domain should also reduce binding to canonical DNA binding sequences such as the insulin response element (IRE) and Daf-16 binding element (DBE). We therefore performed electrophoresis mobility shift assay (EMSA) experiments to measure the binding of FKH wt, L183P and L183R to IRE and DBE2 oligonucleotides (Table 2). Experiments were performed using three independent protein batches, and bands quantified by ImageLab (BioRad). The results were fit to a two-state binding model to estimate the dissociation constant K_D .

Representative EMSA images clearly illustrate that L183P and L183R show progressively lower affinity for dsDNA (Figure 2). The affinity of FKH wt for IRE from three independent experiments was $K_D = 200 \pm 70 \text{ nM}$, ~2-fold weaker than previously reported.⁶ The affinity of FKH L183P for IRE was 2.5-fold weaker, $K_D = 460 \pm 100 \text{ nM}$. The affinity for L183R for IRE was significantly weaker, the dissociation constant was not measurable for the concentration tested, $K_D > 1 \mu\text{M}$.

The EMSA assay was repeated for dsDBE2 with single batches of FKH wt and FKH L183P. The affinity of FKH wt for DBE2 was $K_D = 90 \text{ nM}$, 10-fold weaker than previously reported, but again the affinity of FKH L183P was 2.5-fold weaker, $K_D = 230 \text{ nM}$. These results support the conclusion that FKH mutations which disrupt folding also disrupt binding to canonical target DNA sequences.

FEP calculations partially explain how FKH mutations disrupt folding

We next sought to more fully characterize structural mechanisms by which mutations destabilize the FOXO1 FKH domain, using alchemical free energy perturbation (FEP, see Materials and Methods). This method relies on atomistic molecular dynamics simulation to calculate the free energy cost of transforming a wild-type amino acid residue to a mutant one.¹⁶ The difference between this free energy cost for the folded state (calculated using the folded FKH domain), and the free energy cost for the unfolded state (calculated using a tripeptide model of the unfolded state), yields a prediction of the change in the folding free energy upon mutation, ΔG_{fold} (Figure 3A). The accuracy of this state-of-the-art method is expected to be 1.1–1.6 kcal mol⁻¹ (MUE) based on a number of recent benchmarks.^{17–19}

The trends seen in both FEP results and experimental T_m measurements agree (Pearson correlation coefficient of $r = -0.85$ for both TSA and DSC measurements, Figure 3B).

Most importantly, FEP predicts the most destabilizing mutations to be L183P and L183R, in agreement with experiments, but for mutations that change the melting temperature by less than ± 5 °C, FEP often fails to predict over- vs. under-stabilization, particularly for FKH T182M and S205T.

To quantitatively evaluate the accuracy of the FEP predictions, we used the the DSC T_m values for each for each FKH variant to estimate experimental G_{fold} using a method adapted from Robertson, Murphy, and Rees (see Materials and Methods).^{20–22} This comparison reveals that poor quantitative agreement between FEP estimates of G_{fold} and experiment, especially for mutations that significantly decrease stability of the FOXO1 FKH domain (see Figure 7C), with a MUE of 2.11 kcal mol⁻¹ (RMSE 3.5 kcal mol⁻¹). Aside from protein force field issues (which have been found to be generally robust for relative FEP¹⁸), there are several reasons why this may be. One possibility is that the unfolded state is poorly approximated by the tripeptide simulation; by comparison, realistic protein unfolded states may be significantly more compact, with specific residual structure.^{23–25} Another reason may stem from the well-known challenges associated with charge-changing and proline transformations.^{18,26}

Massively parallel simulations and Markov models elucidate the folding mechanism of FOXO1 at atomic resolution

To better understand the folding of FOXO1 FKH domain at higher-resolution, and how oncogenic mutations might perturb this process, we next sought to construct a Markov model of the folding mechanism,^{27–31} from nearly ten thousand independent molecular simulation trajectories simulated on the Folding@home distributed computing platform.^{32,33} Unlike previous *ab initio* folding studies, which have focused on well-studied mini-proteins,^{34,35} the folding of FOXO1 has not yet been experimentally characterized.

Simulation trajectories were started from twenty different folded and unfolded conformations generated from high-temperature simulations (Figure S2, see Materials and Methods), with production runs performed at 375 K, which we expected to be close to the simulation melting temperature of FOXO1 FKH based on the results of simulation studies on ubiquitin³⁶ and other small two-state folders.³⁵ Simulating near the simulation melting temperature was employed to increase the likelihood of sampling folding and unfolding transitions. An aggregate ~6.8 ms of simulation data was collected from 9927 trajectories (average length 682 ns) generated using GPU-accelerated OpenMM³⁷ with the AMBER ff14SB force field³⁸ and TIP3P water.³⁹

Time-lagged independent component analysis (TICA) was used to project the trajectory data to components that best capture the slowest motions in the folding reaction.^{40,41} Trajectories were featurized using pairwise distances between every other backbone C α , and the time-lagged correlation matrix of these features was computed using lag time of 2.5 ns. Projection of the trajectory data onto the first two time-independent coordinates tIC1 and tIC2 shows a well-sampled funnel-shaped landscape, with the slowest motions along tIC1 corresponding to global folding/unfolding (Figure 4A).

Trajectory data projected to the four largest TICA components were clustered into 1000 states using a k -centers algorithm. The number of TICA components was chosen based on analysis of VAMP2 scores (see Materials and Methods). This enabled the construction of a Bayesian MSM, a method which reduces the statistical error of model estimation.⁴² The first relaxation eigenmode of this model shows population flowing from folded (+) to unfolded (-) conformations along tIC1, indicating that the protein is moderately unstable at the simulation temperature of 375 K. Although unfolding events are observed more frequently than folding events, many folding and refolding events are also observed (Figure S3 and Figure S4). Using PCCA+^{43,44} we clustered the individual microstates into three coarse-grained macrostates identified as unfolded (U), intermediate (I), and folded (N) (Figure 5). Mean passage times between macrostates are on the 5–20 μ s timescale (Table S1), with the mean first passage time of folding estimated to be 20.7 μ s.

To elucidate the main structural events in the folding pathway, the DSSP algorithm⁴⁵ was used to assign per-residue secondary structure to each macrostate and the populations of native secondary structure elements were calculated (Figure 5A). FOXO1 has three α -helices ($\alpha 1$, $\alpha 2$, $\alpha 3$), an α -turn (αT), and two β -sheets ($\beta 4$, $\beta 5$). In the unfolded state (U), the probability p of any of these structures to be correctly folded is less than 0.5. In the intermediate state (I), we find that there is a high probability ($p > 0.8$) that secondary structures are formed except for $\alpha 3$ ($0.2 < p < 0.3$). Only after the other structural elements are in place does $\alpha 3$ form in the native state (N). An analysis of interresidue contact frequencies for each macrostate reaches similar conclusions (Figure 5B).⁴⁶ Native contacts have low populations in U, but are mostly formed in I except for contacts with $\alpha 3$. In N all native contacts are formed. This is consistent with sequence-based predictions from PSIPRED of poor helix propensities for $\alpha 3$ (Figure S2), suggesting tertiary context is important for its formation.⁴⁷ This context-dependence may explain why none of the highly destabilizing mutations are located in $\alpha 3$, except S205, which is far from the major groove when bound to DNA.⁶

While the crystal structure of DNA-bound FOXO1 contains a Ca^{2+} ion that helps stabilize $\alpha 3$, the context-dependent folding of $\alpha 3$ is likely a general feature of FOXO domain structure and dynamics. NMR structural studies of four different FOXO domains in the absence of Ca^{2+} show differences in solution-state structure at αT and $\alpha 3$, but highly similar crystal structures when bound to DNA.⁴⁸

A hydrophobic transfer model accurately predicts thermostability changes of mutations from simulated changes in SASA

Simulations predict a compact denatured state for FOXO1 with populations of residual structure. We hypothesized that the structural detail and extensive statistical sampling provided by simulations could help us evaluate whether the tripeptide model used in the FEP studies is a sufficiently accurate approximation of the unfolded state.

The Shrake-Rupley algorithm⁴⁹ was used to calculate how the average solvent-accessible surface area (SASA) of residues S152, S153, A166, T182, L183, and S205 changes for each macrostate along the folding reaction (Figure 6). Interestingly, predicted changes in SASA were highly non-trivial. In many cases, the SASA of the native state (N) was comparable

(S152 and S153) or greater (S205) than the unfolded state (U). More importantly, for all residues, the average SASA in the tripeptide model of the unfolded state used for FEP was at least 2.8 times larger than in our simulations, suggesting that a more realistic model of SASA in the unfolded state might lead to more accurate predictions of how mutations change FOXO1 stability.

To test this idea, we constructed a model of how the free energy of folding $G(U \rightarrow N)$ depends on changes in the local hydrophobic environment, based on the empirical hydrophobic transfer model of Eisenberg et al.^{50–53} The change in the free energy of folding G_{fold} due to a mutation at residue position i , from amino acid residue r_i to s_i , is computed as

$$\Delta\Delta G_{\text{fold}} = [H(s_i) - H(r_i)](\Delta A_i / A_0(r_i)),$$

where $H(s_i)$ and $H(r_i)$ are the hydrophobicities of the mutant residue s_i and wild type residue r_i , respectively, A_i is the predicted change in SASA of (the wild type) residue i upon folding (U \rightarrow N), and $A_0(r_i)$ is the maximum solvent exposure of the amino acid residue at position i . Importantly, the only free parameter in this model is the input value A_i , which we compute directly from the molecular simulations (see Materials and Methods). All other parameters come directly from previously published work: the hydrophobicities H come from the consensus hydrophobicity scale of Eisenberg et al (1982),⁵¹ and the normalized maximum solvent exposures of amino acids A_0^i are taken from Tien et al.⁵³

For comparison, we also calculated G_{fold} estimates using the popular FoldX algorithm, a native structure-based empirical predictor of protein stability trained on a large corpus experimental data.⁵⁴

Figure 7 shows a side-by-side comparison of three different computational predictions of G_{fold} : from the hydrophobic transfer model (Figure 7A), from FoldX (Figure 7B), and from FEP calculations (Figure 7C). Of these, the hydrophobic transfer model agrees most accurately with the DSC results, with a RMSE of 0.25 kcal mol⁻¹, and a linear fit of slope 0.98. The FoldX predictions generally agree with results from DSC, but overestimate the magnitude of destabilizing mutations (RMSE = 0.87 kcal mol⁻¹). The FEP predictions are least accurate, severely overestimating the magnitude of destabilizing mutations with RMSE = 3.1 kcal mol⁻¹ and a linear fit slope of 5.09.

Interestingly, all three methods predict T182M to be stabilizing, despite the experimental finding that it is destabilizing ($G_{\text{fold}} = +0.42$ kcal mol⁻¹). This may be due to the context-dependent importance of threonine at that position in the native structure, where it serves both as a strand-pairing residue in a preferred beta-sheet conformation, and as a N-terminal capping group, whose side chain makes a backbone hydrogen bond to stabilize α_2 .

FEP calculations along the folding coordinate show a continuum of disruption to local folding

To further explore the mechanism by which mutations perturb the folding reaction, six representative structures of folding intermediates were taken along the folding coordinate tIC1, and used as fixed-backbone structural templates for FEP calculations of the free energy of mutation (Figure 8A). For most of these structures, helices $\alpha 1$ and $\alpha 2$ are folded or partially folded.

Figure 8B shows a plot of the predicted $G = G_i - G_U$ from FEP, where G_i is calculated using one of the structures along the folding coordinate tIC1, and G_U is taken from the tripeptide model of the unfolded-state. When $G_i = G_N$, G recapitulates the FEP predictions of G_{fold} , but for intermediate G_i , the value of G reports on the magnitude of the perturbation along the folding reaction. The results suggest that destabilizing mutations L183P and L183R may exert their effects differently along the folding reaction: L183P perturbations gradually increase as FOXO1 acquires native structure, while L183R appears to particularly destabilize the N state in the last step of folding.

Discussion

To the best of our knowledge, this work represents the first detailed investigation of FOX domain folding mechanism. The Markov model we have constructed makes testable predictions about the folding pathways and rates of this ‘winged helix’ motif that can be tested experimentally. Previous studies of Engrailed homeodomain (EnHD) and its homolog Pit1 demonstrate the malleable stability of the HTH motif, i.e. $\alpha 2$ - $\alpha 3$ of the FOX domain, resulting in a continuum of folding mechanisms from ultrafast three-state framework to apparent two-state nucleation-condensation model.⁵⁵ We speculate that similar malleability might be seen across the family of FOX domains, an intriguing area to explore in the future within the context of disease-related mutations.

A key result of this study is that two oncogenic point mutations at L183 dramatically destabilize the FOX domain, resulting in a FKH domain that is largely unfolded at physiological temperatures, and loss of function (DNA-binding).

Point mutations that destabilize the FOX domain have been previously reported. Mutation of paired residues within the β -turn ‘wing1’ of FOXO1 homolog FOXD3 to glycine and proline reduced the T_m by 3 °C and 5 °C, respectively.⁵⁶ A naturally occurring mutation in FOXD3, N173H associated with developmental ocular conditions, corresponds to FOXO1 K192; this occurs on the protein surface and is not expected to affect folding or DNA binding.⁵⁷ Two mutations in FOXC1, L130F and W152G also associated with developmental ocular conditions, correspond to conserved residues FOXO1 L217 and W237.^{58,59} FOXO1 L217 lies on helix $\alpha 3$ and W237 on the second β -strand of ‘wing1’, both within 4 Å of L183. Assays of both FOXC1 mutations in cell culture indicated reduced DNA-binding and formation of protein aggregates in the cytoplasm;^{58,59} cytoplasmic localization of FOXO1 is also observed in DLBCL specimens.^{60,61} Thermal analysis of six disease-causing mutants of FOXG1 all showed dramatic reductions in T_m , with T_m ranging

from -8 to -15 °C relative to wild-type. These included FOXG1 R230, corresponding to conserved FOXO1 R214 that—like L217—lies on helix $\alpha 3$ within 4 Å of L183.⁶² Taken together, there are three different residues with 4 Å of FOXO1 L183 with either in vivo and/or in vitro evidence of disrupted folding, suggesting a common conserved hydrophobic core involving the packing of helices $\alpha 2$ and $\alpha 3$. The loci of native-state contacts with these residues supports this idea (Table 4).

The destabilization of the FOXO1 fold by L183P and L183R has important translational significance. The loss of stability abrogates DNA-binding, implying loss-of-function (LOF) and oncogenicity due to the negation of canonical tumor suppressor function of FOXO1 in mature B cells. We note that: (i) although FOXO1 is one of the more commonly mutated genes in DLBCL, L183 mutations are overall extremely rare, (ii) cancer mutations are heterozygous, there should be a second wild-type allele to maintain FOXO1 function, and (iii) most of the 10 FOXO1 mutations tested had little to no effect on folding of the FOX domain. There are many other ways mutations in the FOX domain could affect FOXO1 function, and further cell-based assays are required. Nevertheless, we also note that FOXO1 cytoplasmic relocalization is relatively common in DLBCL cases, just as BCL2/MYC overexpression is more common than genetic translocation.⁶¹ FOXO1 L183 mutation could be a genetic instance of a more general phenotype of unfolding/misfolding of FOX domains, and this could result not only in LOF phenotypes but toxic gain-of-function (GOF) such as off-target or nonspecific DNA-binding and gene transcription. These observations, together with corresponding disease-associated mutations such as in FOXC1 and FOXG1,^{63,64} support further experimental and computational studies of the folding mechanism of FOX domains.

To better understand how disease-related mutations destabilize FOXO1, we performed both FEP calculations and massively parallel molecular simulations of the folding reaction. Were these calculations worth the expense? Yes, for several reasons. First, the destabilizing effect of L183P is not obvious from inspection of the native state alone; proline is not uncommon in the first turn of alpha helices, in fact L183 corresponds to the position of highest propensity (Ncap+1).⁶⁵ However, the restricted range of dihedral angles for proline clearly disrupt the folding of the HTH motif. Similarly, the destabilizing effect of arginine is not obvious from examination of the native state; the FOX domain is already highly basic with 8 arginines and 12 lysines in 115 residues facilitating favorable electrostatic interaction with the DNA backbone. However, the introduction of a bulky charged residue into the hydrophobic core of the HTH motif clearly destabilizes the HTH motif and hence the folding of the FOX domain. These observations of L183, together with the milder effect of FOXO1 T182M and identification of corresponding destabilizing mutations in FOXC1 and FOXG1, emphasize the value of modeling the folding landscape of FOXO1 in atomic detail.

Another reason to perform atomistic folding simulations is to accurately model the unfolded state. The AMBER14SB force field accurately predicts solution-state NMR observables,^{38,66} and can effectively model both folded and unfolded states.⁶⁷ While we found FEP predictions of the destabilizing effects of mutations to be qualitatively instructive, the hydrophobic transfer model is able to infer accurate *quantitative* predictions from

simulated unfolded states. One likely reason for the inaccuracy of FEP is the use of an unrealistic tripeptide model, which, as we have shown, severely overestimates the extent of solvent exposure in the unfolded state. On the other hand, our FEP studies along the folding coordinate (Figure 8) show that even if we assume that the unfolded state of FOXO1 resembles a compact and partially folded intermediate, FEP still overestimates the magnitude of destabilization.

While FEP remains state-of-the-art in terms of molecular modeling, “exaggerated” predictions of the effects of mutations are not uncommon. Gapsys et al. (2016) analyzed 119 mutations in the enzyme barnase by alchemical FEP, and found the range of calculated global G_{fold} to be -2.2 – 7.2 kcal mol⁻¹.¹⁸ Kucukkal et al., studying the effects of 10 Rett Syndrome mutations in MeCP2 MBD, note that accurately predicting the effects of charge changing mutations via FEP is particularly difficult.⁶⁸ Steinbrecher et al., in a large-scale validation of FEP+, provide additional evidence for this difficulty with charge changing mutations, as well as finding that FEP+ relative free energies tended to be more positive than experiment.¹⁷ They further elaborate on FEP+'s difficulty in accurately predicting the effects of Glu, Lys, and Arg mutations due to finite sampling error of sidechain electrostatic interactions in the unfolding reaction.

Another key result of our work is the success of a simple hydrophobic transfer model in predicting the effects of mutations on folding stability. Aside from the seminal work of Eisenberg et al. in characterizing hydrophobic moments of proteins, the key ingredient in this approach appears to be the use of simulated ensembles to accurately capture the solvent accessibility of unfolded protein states. Future work should explore the application of this method to a wider range of proteins, to better gauge its overall accuracy. Modification of the hydrophobic transfer model to include residue-wise secondary structure propensities might further improve its accuracy.

Conclusion

This work examined the destabilizing effects of ten oncogenic mutations found in the human FOXO1 FKH domain, as seen in DLBCL patients and cell lines. Mutations at L183 greatly reduce FKH stability, and also reduce binding to canonical DNA sequences IRE and DBE2, suggesting that the decrease in folding stability is responsible for the loss of function. To better understand the mechanism of FOXO1 FKH folding and how mutations perturb it, alchemical FEP calculations and massively parallel molecular simulations were performed, enabling the construction of a Markov model that predicts folding pathways, rates, and a mechanism in which the formation of helix α 3 is the final step. While FEP overestimates the magnitude of destabilization, a simple hydrophobic transfer model, used in conjunction with simulation-based estimates of solvent-accessible surface area, was shown to quantitatively predict changes in folding free energies from mutations, with high accuracy. These results make a strong case that more realistic, atomistic modeling of unfolded states is highly useful for studying the destabilizing effects of mutations. These observations, together with corresponding disease-associated mutations such as in FOXC1 and FOXG1,^{63,64} support further experimental and computational studies of the folding mechanism of FOX domains.

Materials and Methods

Sample Preparation

The FOXO1 Forkhead DNA-binding domain (FKH) was expressed and purified as previously described, with minor modifications.⁶⁹ A codon-optimized fragment encoding NdeI-TEV-FKH (151–265)-XhoI was subcloned into pET28. The construct was transformed into *E. coli* BL21 (DE3), grown in 2xYT medium with 30 µg/ml kanamycin. Protein expression was induced by 0.5 mM IPTG at 21 °C for 8–16 h. Cells were resuspended in 500 mM NaCl, 5 mM beta ME, 20 mM Imidazole, 10% glycerol, disrupted by sonication, clarified by centrifugation and applied to a HisTrap column (GE Healthcare). Protein was eluted with a 0–500 mM imidazole gradient.

His-tagged FKH was desalted into 200–300 mM NaCl, 20 mM HEPES pH 7.5, 0.5 mM EDTA, 1 mM DTT, 5–10% w/v glycerol, cut overnight with His-tagged TEV (1:40 w:w) at 10 °C, followed by passage over Ni-NTA (Qiagen) gravity column to remove TEV and uncut protein. Cleaved FKH protein was exchanged into 50–300 mM NaCl, 20 mM HEPES pH 7.5, 10% w/v glycerol and applied to a MonoS or Canto S column (GE Healthcare). Protein was eluted with a gradient to 1.0 M NaCl. For storage, FKH was concentrated, glycerol added to 25% w/v, flash-frozen in liquid N₂ and stored at –80 °C.

Biotinylated 32bp DNA oligos were synthesized (Eurofins) containing canonical FKH binding motifs IRE, DBE1 and DBE2 (See Table 1). Complementary oligos were dissolved at 100 µM each in 10 mM Tris-HCl, pH 7.5, 100 mM KCl and 1 mM EDTA and annealed in a PCR block with temperature gradient from 80°C to 20 °C over 60 min. Annealed dsDNA was purified by FPLC on a MonoQ 5/50 column (GE Healthcare) with 20 ml linear gradient 0.1–1.1 M KCl. Purified dsDNA was dialyzed against 20 mM Tris-HCl pH 8.0, 50 mM KCl, 5% w/v glycerol, 2 mM DTT, 0.2 mM EDTA, 2 mM MgCl₂.

Thermal Shift Assay (TSA)

FKH protein was exchanged into 20 mM Tris-HCl pH 8.0, 50 mM KCl, 5% w/v glycerol, 2 mM DTT, 0.2 mM EDTA, 2 mM MgCl₂, at 5 µM concentration and 2X Sypro Orange (Thermo Fisher). Aliquots of 50 µl were loaded into PCR strips and analyzed on a StepOne Plus™ qPCR instrument (Thermofisher) with temperature gradient 25–95 °C ramp, 1 °C increment, 1 min hold/increment. Data from the ROX channel was baseline corrected and fit to a 2-state model in Matlab.

Differential Scanning Calorimetry (DSC)

FKH protein was prepared in Tris-EDTA at 1 mg/ml (~7.5 µM). Samples were analyzed in a PEAQ-DSC instrument (Malvern) with temperature gradient 25–90 °C. Baseline corrected, buffer-subtracted data were analyzed by a non two-state model.

Electrophoretic Mobility Shift Assay (EMSA)

FKH at 5–1000 nM in 10 mM Tris-HCl pH 7.5, 50 mM KCl, 5% w/v glycerol, 1 mM DTT, 0.2 mM EDTA, 1 mM MgCl₂ was incubated with 1 nM biotinylated dsDNA in presence of 1 ng/µl poly-dI/dC for 30 min at room temperature. Samples were separated on 6% 0.5X

TBE polyacrylamide gels and analyzed using LightShift® Chemiluminescent EMSA Kit (ThermoFisher).

Free Energy Perturbation (FEP) Calculations

To model unfolded-state conformations, eight FKH tripeptides were prepared using UCSF Chimera 1.13.1. Alchemical topologies for wild type (wt, A state) and mutant (B state) were prepared using *pmx*.^{19,70} The tripeptide sequences originating from the FOXO1 FKH domain are T(L183R)S, K(S152R)S, S(S153R)S Y(A166G)D, Y(A166V)D, L(T182M)L, N(S205T)S, and T(L183P)S. Each tripeptide was solvated with TIP3P water in cubic periodic boxes ranging in volume from (3.15 nm)³ to (3.54 nm)³. Neutralizing K⁺ and Cl⁻ counterions were added at 50 mM, consistent with concentrations used in the TSA and EMSA assays. For charge-changing mutations, (L183R and S153R), a potassium ion was chosen to have charge +1 in the A state, and neutral charge in the B state.

The folded-state model of FOXO1 FKH(151–249)/Mg complex was adapted from PDB 3COA.⁶ Protonation states at pH 7.5 were determined using H++⁷¹ and the AMBER ff14SB force field was used to build the topology.³⁸ The calcium ion in 3COA was replaced with magnesium and the MCPB algorithm⁷² was used to build three coordination bonds with the carbonyl oxygens of Leu217, His220, and Phe223. Alchemical topologies were prepared using *pmx*,⁷⁰ with the assistance of in-house code (https://github.com/leiqiantemple/AlchemFEP_FOXO1). The final dual topology used the AMBER 14sbmut forcefield available in *pmx*. Each system was solvated in a cubic periodic boxes ranging in volume from (7.16 nm)³ to (7.25 nm)³, with TIP3P water and neutralizing K⁺ and Cl⁻ counterions added at 50 mM. Charge-changing mutations were addressed using the same strategy as the unfolded-state calculations. In addition, six FOXO1 FKH(151–249) conformations corresponding to folding intermediates were taken from the MSM (see Markov Model Construction and Analysis below) and used to prepare alchemical FEP models using a similar process, but with the magnesium-related MCPB steps omitted.

Minimization, equilibration, and production molecular dynamics for the FEP calculations were performed on Temple University's Owlsnest HPC cluster using the GROMACS 2016.3 simulation package.⁷³ Energy minimization was performed until all forces were less than 1000 kJ mol⁻¹ nm⁻¹. For equilibration and production, stochastic (Langevin) integration was performed using explicit TIP3P water solvent, with a time step of 1 fs and friction coefficient 1 ps⁻¹. PME electrostatics were used with fourier spacing 0.12 and PME order 4. Long-range dispersion correction was used. A total of 30 alchemical intermediates were used to interpolate from the A topology to B topology, using potential energy function $U = (1-\lambda)U_A + \lambda U_B$, where $0 \leq \lambda \leq 1$ is the interpolation parameter. The λ values used were (0.000, 0.002, 0.004, 0.006, 0.008, 0.01, 0.02, 0.03, 0.04, 0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.45, 0.50, 0.55, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90, 0.95, 0.975, 1.00). For each alchemical intermediate, equilibration at 300 K was performed in the NVT ensemble for 100 ps, and then in the NPT ensemble at 1 bar for 100 ps using the Parrinello-Rahman barostat with compressibility constant 4.5×10^{-5} and time constant 0.5 ps. Production runs of each model were performed for 8 ns in the NPT ensemble. For each simulation performed using a given value λ_j , the values of $U(\lambda_j)$ and $U_{ij} = U(\lambda_j) - U(\lambda_i)$ were written to

file every 0.5 ps, for all intermediates j . Free energy estimation was performed with the Multistate Bennett Acceptance Ratio (MBAR) method,⁷⁴ using the alchemical-analysis⁷⁵ and pymbar⁷⁴ Python packages. Time-reversed convergence plots for G estimates were used to identify and discard the non-equilibrated regions of the trajectories (Figure S6). These results show that most simulations meet conditions of equilibrium and convergence in the last 4 ns, with the possible exception of folded-state S152R. Therefore, we removed the first 4 ns of each 8-ns trajectory before analysis. The overlap matrix, whose elements O_{ij} quantifies the overlap between the distributions of U_{ij} ,⁷⁵ shows sufficient overlap for accurate free energy estimates (Figure S7).

Massively Parallel Folding Simulations

The model of FOXO1 was adapted from PDB 3CO6.⁶ Protonation states at pH 7.5 were determined using H++,⁷¹ and the AMBER14SB force field was used to build the topology.³⁸ The protein was placed in a cubic periodic box of volume (8.2698 nm)³ and solvated with TIP3P water,³⁹ with Joung-Cheatham neutralizing Na⁺ and Cl⁻ counterions added at 0.1 M,^{76,77} resulting in a system of ~64500 atoms. Molecular dynamics simulations were performed using the GPU-accelerated OpenMM (CUDA platform),³⁷ using stochastic (Langevin) integration with a 2 fs time step, friction coefficient of 1 ps⁻¹ and PME electrostatics⁷⁸ with cutoff 1.0 nm and tolerance 0.005. To generate a range of folded and unfolded conformations, several 60-ns NPT simulations were performed at 300 K, 400 K, 450 K, and 498 K, respectively. Conformational clustering was performed using a k -centers algorithm to identify representative conformations.⁷⁹ Production runs were performed at 375 K on the Folding@home distributed computing platform,^{32,33} with all coordinates saved every 0.5 ns. Five hundred trajectories each with randomized initial velocities were initiated from 20 different starting structures, resulting in 10,000 total independent simulations. An aggregate of ~6.8 ms of simulation data was collected.

Markov Model Construction and Analysis

Using PyEMMA 2.5.7,^{80,81} all 9927 obtained trajectories were featurized using the pairwise distance for every other C α in residues 5 to 95 for a total of 1036 features. Previous work has shown that pairwise distance information can provide excellent featurization of reversible folding trajectory data for the purpose of constructing Markov models.⁸² Dimensionality reduction was performed using time-lagged Independent Component Analysis (TICA) with a lag-time $\tau_{\text{TICA}} = 2.5$ ns. A k -means algorithm was used to conformationally cluster in the low-dimensional TICA projections to define MSM microstates. The VAMP2 score was used to assess models constructed with different numbers of microstates (Figure S8) and different low-rank TICA projections. Based on this score, and a general desire for computational tractability, we selected an MSM with 1000 microstates, clustered after projecting to the four largest TICA components, which adequately capture the slowest conformational motions (Figure S9). MSM implied timescales were calculated as $t_j = -\tau / \ln \mu_j$, where μ_j are the largest nonstationary eigenvalues of the microstate transition matrix. The slowest timescale was stable after 5.0 ns. A Bayesian MSM with a lag time of $\tau = 5.0$ ns was constructed using these microstate definitions.⁴² The microstates were assigned to macrostates (U, I and N) using the PCCA+ algorithm.^{43,44} This coarse-grained three-state model passed the Chapman-Kolmogorov test (Figure S10).

Mean first passage times from macrostates A to B were calculated using PyEMMA, as the self-consistent expectation value E_A of the time T_B to reach a state in B from a state in A, $E_A[T_B] = \sum_{z \in A} (\pi_a E_a[T_B]) / \sum_{z \in A} \pi_z$, where π_a are the equilibrium populations of state a .

All trajectory analysis was performed through a series of in-house Python 3.7.6 scripts utilizing MDTraj⁸³ 1.9.5 and Numpy 1.18.1 libraries, unless otherwise stated. Visualization was performed using Matplotlib 3.3.4 and VMD 1.9.3. Per-residue secondary structure probabilities for each macrostate were calculated via the DSSP algorithm, by comparing the frequency of trajectory snapshot DSSP assignments to the DSSP assignments of the crystal structure. Secondary structural elements were defined by the following residue ranges: $\alpha 1$ (165–175), $\alpha 2$ (183–193), αT (195–197), $\alpha 3$ (203–219), $\beta 4$ (223–226), $\beta 5$ (236–239).

The frequency of native contacts for each macrostate was calculated using the sigmoidal function $(1 + \exp(\beta(d_{ij} - \lambda d_0)))^{-1}$ to count contacts, where d_{ij} are the distances between atoms i and j , $\beta = 50 \text{ nm}^{-1}$, $\lambda = 1.8$, $d_0 = 0.45 \text{ nm}$, as described by Best et al. (2013).⁴⁶ We adapted this method to coarse-grain the contact description as residue-wise rather than atomic, such that if an atomic contact is present between two residues, those residues are considered in contact. The contact frequencies for each macrostate were calculated as a population-weighted average across all microstates assigned to the macrostate.

The Shrake-Rupley algorithm as implemented in MDTraj was used to calculate the solvent-accessible surface area (SASA) for 20 trajectory samples chosen at random from each microstate.⁴⁹ The average residue-wise SASA for each macrostate was then calculated as a population-weighted average across all microstates assigned to the macrostate, where the microstate populations come from the MSM.

A hydrophobic transfer model of G_{fold}

To make quantitative predictions about the about the effects of mutations on protein stability, we developed a transfer free energy model of G_{fold} based on the work of Eisenberg et al. on environmental hydrophobicity.^{50,51} In this model, the free energy of a protein is given in terms of residue hydrophobicities $H(r_i)$ for each residue r_i at sequence position i , and the environmental hydrophobicity of each residue, M_i .

$$G = \sum_i H(r_i) M_i.$$

The hydrophobicities are per-residue free energies, representing the free energy of transfer from a hydrophobic to a hydrophilic phase (specifically, the “consensus hydrophobicities” compiled by Eisenberg et al.⁵¹). The environmental hydrophobicities are assumed to depend on solvent exposure, such that $M_i = A(r_i)/A_0(r_i) - 1/2$, where $A(r_i)$ is the conformation-dependent solvent exposure of residue r_i , and $A_0(r_i)$ is the maximum possible solvent exposure of residue r_i . In this way, the values of M_i range from a fully aqueous environment of $M = +1/2$, to an environment of full burial in the interior of a protein, $M = -1/2$.

According to this model, the free energy of folding is

$$\begin{aligned}
 \Delta G_{\text{fold}} &= G(\text{N}) - G(\text{U}) \\
 &= \sum_i H(r_i) \left[A^{\text{N}}(r_i)/A_0(r_i) - 1/2 \right] - \sum_i H(r_i) \left[A^{\text{U}}(r_i)/A_0(r_i) - 1/2 \right] \\
 &= \sum_i H(r_i) \left(A^{\text{N}}(r_i) - A^{\text{U}}(r_i) \right) / A_0(r_i) \\
 &= \sum_i H(r_i) (\Delta A_i / A_0(r_i)).
 \end{aligned}$$

Here, A_i denotes the change in the solvent exposure of residue i upon folding. For a single-point mutation at position i from residue r_i (wild type) to s_i (mutant), the change in the free energy of folding, $G_{\text{fold}}(\text{U} \rightarrow \text{N})$, depends *only* on residue i ,

$$\Delta \Delta G_{\text{fold}} = \Delta G_{\text{fold}}(\text{mutant}) - \Delta G_{\text{fold}}(\text{wt}) = [H(s_i) - H(r_i)] (\Delta A_i / A_0(r_i)).$$

Thus, the model posits that G_{fold} is the change in the hydrophobicity upon mutation ($H(r_i \rightarrow s_i)$), multiplied by the fractional change $A_i/A_0(r_i)$ in the solvent exposure upon folding. In practice, we obtain values of $A_0(r_i)$ from the normalized maximal accessible surface area scale of Tien et al.⁵³ Consensus hydrophobicities and A_0 parameters used in our model are listed in Table S2.

Estimation of folding free energies from T_m

To estimate the change in the folding free energy G_{fold} from the experimental values of T_m , we assume the standard model of protein stability temperature-dependence,²⁰

$$\Delta G_{\text{fold}} = [\Delta H_{\text{fold}} + \Delta C_p(T - T_m)] - T[(\Delta H_{\text{fold}}/T_m) + \Delta C_p \ln(T/T_m)].$$

We use the work of Robertson and Murphy²¹ to estimate H_{fold} and C_p from the number of residues N ,

$$\Delta H_{\text{fold}} = N \times (-0.698 \text{ kcal mol}^{-1})$$

$$\Delta C_p = N \times (-0.0139 \text{ kcal mol}^{-1} \text{ K}^{-1})$$

Using these values with $N=100$, we estimate G_{fold} as

$$\begin{aligned}
 \Delta \Delta G_{\text{fold}}(T) &= \Delta G_{\text{fold}}(\text{mut}) - \Delta G_{\text{fold}}(\text{wt}) \\
 &= T \left[\Delta H_{\text{fold}} \left(T_m^{-1}(\text{wt}) - T_m^{-1}(\text{mut}) \right) + \Delta C_p \ln(T_m(\text{mut})/T_m(\text{wt})) \right] - \Delta C_p (T_m(\text{mut}) - T_m(\text{wt})).
 \end{aligned}$$

Results in this manuscript are reported at $T = 298.15 \text{ K}$.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

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

Data Availability

All simulation trajectory data and analysis scripts are available via the Open Source Framework at <https://osf.io/t7h5b/>.

References

- (1). Brennan RG (1993) The winged-helix DNA-binding motif: Another helix-turn-helix takeoff. *Cell* 74, 773–776. [PubMed: 8374950]
- (2). Herman L, Todeschini A-L, and Veitia RA (2021) Forkhead Transcription Factors in Health and Disease. *Trends in Genetics* 37, 460–475. [PubMed: 33303287]
- (3). Dai S, Qu L, Li J, and Chen Y (2021) Toward a mechanistic understanding of DNA binding by forkhead transcription factors and its perturbation by pathogenic mutations. *Nucleic Acids Research* 49, 10235–10249. [PubMed: 34551426]
- (4). Coomans de Brachène A, and Demoulin J-B (2016) FOXO transcription factors in cancer development and therapy. *Cell. Mol. Life Sci.* 73, 1159–1172. [PubMed: 26686861]
- (5). Rena G, Guo S, Cichy SC, Unterman TG, and Cohen P (1999) Phosphorylation of the Transcription Factor Forkhead Family Member FKHR by Protein Kinase B. *Journal of Biological Chemistry* 274, 17179–17183. [PubMed: 10358075]
- (6). Brent MM, Anand R, and Marmorstein R (2008) Structural Basis for DNA Recognition by FoxO1 and Its Regulation by Posttranslational Modification. *Structure* 16, 1407–1416. [PubMed: 18786403]
- (7). Ushmorov A, and Wirth T (2018) FOXO in B-cell lymphopoiesis and B cell neoplasia. *Seminars in Cancer Biology* 50, 132–141. [PubMed: 28774833]
- (8). Teras LR, DeSantis CE, Cerhan JR, Morton LM, Jemal A, and Flowers CR (2016) 2016 US lymphoid malignancy statistics by World Health Organization subtypes: 2016 US Lymphoid Malignancy Statistics by World Health Organization Subtypes. *CA: A Cancer Journal for Clinicians* 66, 443–459. [PubMed: 27618563]
- (9). Miyazaki K (2016) Treatment of Diffuse Large B-Cell Lymphoma 56, 79–88.
- (10). Trinh DL, Scott DW, Morin RD, Mendez-Lago M, An J, Jones SJM, Mungall AJ, Zhao Y, Schein J, Steidl C, Connors JM, Gascoyne RD, and Marra MA (2013) Analysis of FOXO1 mutations in diffuse large B-cell lymphoma. *Blood* 121, 3666–3674. [PubMed: 23460611]
- (11). Morin RD, Assouline S, Alcaide M, Mohajeri A, Johnston RL, Chong L, Grewal J, Yu S, Fornika D, Bushell K, Nielsen TH, Petrogiannis-Halios T, Crump M, Tosikyan A, Grande BM, MacDonald D, Rousseau C, Bayat M, Sesques P, Froment R, Albuquerque M, Monczak Y, Oros KK, Greenwood C, Riazalhosseini Y, Arseneault M, Camlioglu E, Constantin A, Pan-Hammarstrom Q, Peng R, Mann KK, and Johnson NA (2016) Genetic Landscapes of Relapsed and Refractory Diffuse Large B-Cell Lymphomas. *Clinical Cancer Research* 22, 2290–2300. [PubMed: 26647218]
- (12). Reddy A, Zhang J, Davis NS, Moffitt AB, Love CL, Waldrop A, Leppa S, Pasanen A, Meriranta L, Karjalainen-Lindsberg M-L, Nørgaard P, Pedersen M, Gang AO, Høgdall E, Heavican TB, Lone W, Iqbal J, Qin Q, Li G, Kim SY, Healy J, Richards KL, Fedoriw Y, Bernal-Mizrachi L, Koff JL, Staton AD, Flowers CR, Paltiel O, Goldschmidt N, Calaminici M, Clear A, Gribben J, Nguyen E, Czader MB, Ondrejka SL, Collie A, Hsi ED, Tse E, Au-Yeung RKH, Kwong Y-L, Srivastava G, Choi WWL, Evens AM, Pilichowska M, Sengar M, Reddy N, Li S, Chadburn A, Gordon LI, Jaffe ES, Levy S, Rempel R, Tzeng T, Happ LE, Dave T, Rajagopalan D, Datta

- J, Dunson DB, and Dave SS (2017) Genetic and Functional Drivers of Diffuse Large B Cell Lymphoma. *Cell* 171, 481–494.e15. [PubMed: 28985567]
- (13). Kabrani E, Chu VT, Tasouri E, Sommermann T, Baßler K, Ulas T, Zenz T, Bullinger L, Schultze JL, Rajewsky K, and Sander S (2018) Nuclear FOXO1 promotes lymphomagenesis in germinal center B cells. *Blood* 132, 2670–2683. [PubMed: 30333121]
- (14). Zhou P, Blain AE, Newman AM, Zaka M, Chagaluka G, Adlar FR, Offor UT, Broadbent C, Chaytor L, Whitehead A, Hall A, O'Connor H, Van Noorden S, Lampert I, Bailey S, Molyneux E, Bacon CM, Bomken S, and Rand V (2019) Sporadic and endemic Burkitt lymphoma have frequent FOXO1 mutations but distinct hotspots in the AKT recognition motif. *Blood Advances* 3, 2118–2127. [PubMed: 31300419]
- (15). Huynh K, and Partch CL (2015) Analysis of Protein Stability and Ligand Interactions by Thermal Shift Assay. *Current Protocols in Protein Science* 79.
- (16). Aldeghi M, de Groot BL, and Gapsys V (2019) Accurate Calculation of Free Energy Changes upon Amino Acid Mutation, in *Computational Methods in Protein Evolution* (Sikosek T, Ed.), pp 19–47. Springer New York, New York, NY.
- (17). Steinbrecher T, Zhu C, Wang L, Abel R, Negron C, Pearlman D, Feyfant E, Duan J, and Sherman W (2017) Predicting the Effect of Amino Acid Single-Point Mutations on Protein Stability—Large-Scale Validation of MD-Based Relative Free Energy Calculations. *Journal of Molecular Biology* 429, 948–963. [PubMed: 27964946]
- (18). Gapsys V, Michielssens S, Seeliger D, and de Groot BL (2016) Accurate and Rigorous Prediction of the Changes in Protein Free Energies in a Large-Scale Mutation Scan. *Angew. Chem. Int. Ed.* 55, 7364–7368.
- (19). Seeliger D, and de Groot BL (2010) Protein Thermostability Calculations Using Alchemical Free Energy Simulations. *Biophysical Journal* 98, 2309–2316. [PubMed: 20483340]
- (20). Johnson CM (2013) Differential scanning calorimetry as a tool for protein folding and stability. *Archives of Biochemistry and Biophysics* 531, 100–109. [PubMed: 23022410]
- (21). Robertson AD, and Murphy KP (1997) Protein Structure and the Energetics of Protein Stability. *Chem. Rev.* 97, 1251–1268. [PubMed: 11851450]
- (22). Rees DC, and Robertson AD (2001) Some thermodynamic implications for the thermostability of proteins. *Protein Sci.* 10, 1187–1194. [PubMed: 11369857]
- (23). Peran I, Holehouse AS, Carrico IS, Pappu RV, Bilsel O, and Raleigh DP (2019) Unfolded states under folding conditions accommodate sequence-specific conformational preferences with random coil-like dimensions. *Proc Natl Acad Sci USA* 116, 12301–12310. [PubMed: 31167941]
- (24). Cho J-H, Meng W, Sato S, Kim EY, Schindelin H, and Raleigh DP (2014) Energetically significant networks of coupled interactions within an unfolded protein. *Proceedings of the National Academy of Sciences* 111, 12079–12084.
- (25). Voelz VA, Singh VR, Wedemeyer WJ, Lapidus LJ, and Pande VS (2010) Unfolded-State Dynamics and Structure of Protein L Characterized by Simulation and Experiment. *J. Am. Chem. Soc.* 132, 4702–4709. [PubMed: 20218718]
- (26). Gapsys V, and de Groot BL (2017) pmx Webserver: A User Friendly Interface for Alchemistry. *J. Chem. Inf. Model.* 57, 109–114. [PubMed: 28181802]
- (27). Bowman GR, Huang X, and Pande VS (2009) Using generalized ensemble simulations and Markov state models to identify conformational states. *Methods* 49, 197–201. [PubMed: 19410002]
- (28). Noé F, and Fischer S (2008) Transition networks for modeling the kinetics of conformational change in macromolecules. *Current Opinion in Structural Biology* 18, 154–162. [PubMed: 18378442]
- (29). Bowman GR, Beauchamp KA, Boxer G, and Pande VS (2009) Progress and challenges in the automated construction of Markov state models for full protein systems. *The Journal of Chemical Physics* 131, 124101. [PubMed: 19791846]
- (30). Pande VS, Beauchamp K, and Bowman GR (2010) Everything you wanted to know about Markov State Models but were afraid to ask. *Methods* 52, 99–105. [PubMed: 20570730]
- (31). Husic BE, and Pande VS (2018) Markov State Models: From an Art to a Science. *J. Am. Chem. Soc.* 140, 2386–2396. [PubMed: 29323881]

- (32). Shirts M (2000) COMPUTING: Screen Savers of the World Unite! *Science* 290, 1903–1904. [PubMed: 17742054]
- (33). Zimmerman MI, Porter JR, Ward MD, Singh S, Vithani N, Meller A, Mallimadugula UL, Kuhn CE, Borowsky JH, Wiewiora RP, Hurley MFD, Harbison AM, Fogarty CA, Coffland JE, Fadda E, Voelz VA, Chodera JD, and Bowman GR (2021) SARS-CoV-2 simulations go exascale to predict dramatic spike opening and cryptic pockets across the proteome. *Nat. Chem.*
- (34). Voelz VA, Bowman GR, Beauchamp K, and Pande VS (2010) Molecular Simulation of *ab Initio* Protein Folding for a Millisecond Folder NTL9(1–39). *J. Am. Chem. Soc.* 132, 1526–1528. [PubMed: 20070076]
- (35). Lindorff-Larsen K, Piana S, Dror RO, and Shaw DE (2011) How Fast-Folding Proteins Fold. *Science* 334, 517–520. [PubMed: 22034434]
- (36). Piana S, Lindorff-Larsen K, and Shaw DE (2013) Atomic-level description of ubiquitin folding. *Proceedings of the National Academy of Sciences* 110, 5915–5920.
- (37). Eastman P, Swails J, Chodera JD, McGibbon RT, Zhao Y, Beauchamp KA, Wang L-P, Simmonett AC, Harrigan MP, Stern CD, Wiewiora RP, Brooks BR, and Pande VS (2017) OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Comput Biol* (Gentleman, R., Ed.) 13, e1005659.
- (38). Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, and Simmerling C (2015) ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* 11, 3696–3713. [PubMed: 26574453]
- (39). Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, and Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* 79, 926–935.
- (40). Schwantes CR, and Pande VS (2013) Improvements in Markov State Model Construction Reveal Many Non-Native Interactions in the Folding of NTL9. *J. Chem. Theory Comput.* 9, 2000–2009. [PubMed: 23750122]
- (41). Pérez-Hernández G, Paul F, Giorgino T, De Fabritiis G, and Noé F (2013) Identification of slow molecular order parameters for Markov model construction. *The Journal of Chemical Physics* 139, 015102. [PubMed: 23822324]
- (42). Trendelkamp-Schroer B, Wu H, Paul F, and Noé F (2015) Estimation and uncertainty of reversible Markov models. *The Journal of Chemical Physics* 143, 174101. [PubMed: 26547152]
- (43). Röblitz S, and Weber M (2013) Fuzzy spectral clustering by PCCA+: application to Markov state models and data classification. *Adv Data Anal Classif* 7, 147–179.
- (44). Noé F, Wu H, Prinz J-H, and Plattner N (2013) Projected and hidden Markov models for calculating kinetics and metastable states of complex molecules. *The Journal of Chemical Physics* 139, 184114. [PubMed: 24320261]
- (45). Kabsch W, and Sander C (1983) Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22, 2577–2637. [PubMed: 6667333]
- (46). Best RB, Hummer G, and Eaton WA (2013) Native contacts determine protein folding mechanisms in atomistic simulations. *Proceedings of the National Academy of Sciences* 110, 17874–17879.
- (47). Buchan DWA, and Jones DT (2019) The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Research* 47, W402–W407. [PubMed: 31251384]
- (48). Psenakova K, Kohoutova K, Obsilova V, Ausserlechner M, Veverka V, and Obsil T (2019) Forkhead Domains of FOXO Transcription Factors Differ in both Overall Conformation and Dynamics. *Cells* 8, 966.
- (49). Shrake A, and Rupley JA (1973) Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *Journal of Molecular Biology* 79, 351–371. [PubMed: 4760134]
- (50). Eisenberg D, Wilcox W, and McLachlan AD (1986) Hydrophobicity and amphiphilicity in protein structure. *J. Cell. Biochem.* 31, 11–17. [PubMed: 3722276]
- (51). Eisenberg D, Weiss RM, Terwilliger TC, and Wilcox W (1982) Hydrophobic moments and protein structure. *Faraday Symp. Chem. Soc.* 17, 109.

- (52). Eisenberg D, Schwarz E, Komaromy M, and Wall R (1984) Analysis of membrane and surface protein sequences with the hydrophobic moment plot. *Journal of Molecular Biology* 179, 125–142. [PubMed: 6502707]
- (53). Tien MZ, Meyer AG, Sydykova DK, Spielman SJ, and Wilke CO (2013) Maximum allowed solvent accessibilities of residues in proteins. *PLoS one* 8, e80635. [PubMed: 24278298]
- (54). Delgado J, Radusky LG, Cianferoni D, and Serrano L (2019) FoldX 5.0: working with RNA, small molecules and a new graphical interface. *Bioinformatics (Valencia, A., Ed.)* 35, 4168–4169.
- (55). Banachewicz W, Religa TL, Schaeffer RD, Daggett V, and Fersht AR (2011) Malleability of folding intermediates in the homeodomain superfamily. *Proceedings of the National Academy of Sciences* 108, 5596–5601.
- (56). Yan H, and Liao X (2003) Amino Acid Substitutions in a Long Flexible Sequence Influence Thermodynamics and Internal Dynamic Properties of Winged Helix Protein Genesis and Its DNA Complex. *Biophysical Journal* 85, 3248–3254. [PubMed: 14581224]
- (57). Kloss BAV, Reis LM, Brémond-Gignac D, Glaser T, and Semina EV (2012) Analysis of FOXD3 sequence variation in human ocular disease. *Molecular Vision* 10.
- (58). Ito YA, Footz TK, Murphy TC, Courtens W, and Walter MA (2007) Analyses of a Novel L130F Missense Mutation in FOXC1. *ARCH OPHTHALMOL* 125, 8.
- (59). Ito YA, Footz TK, Berry FB, Mirzayans F, Yu M, Khan AO, and Walter MA (2009) Severe Molecular Defects of a Novel *FOXC1 W152G* Mutation Result in Aniridia. *Invest. Ophthalmol. Vis. Sci.* 50, 3573. [PubMed: 19279310]
- (60). Bogusz AM, Baxter RHG, Currie T, Sinha P, Sohani AR, Kutok JL, and Rodig SJ (2012) Quantitative Immunofluorescence Reveals the Signature of Active B-cell Receptor Signaling in Diffuse Large B-cell Lymphoma. *Clinical Cancer Research* 18, 6122–6135. [PubMed: 22966017]
- (61). Bogusz AM, Kovach AE, Le LP, Feng D, Baxter RH, and Sohani AR (2017) Diffuse large B-cell lymphoma with concurrent high MYC and BCL2 expression shows evidence of active B-cell receptor signaling by quantitative immunofluorescence. *PLoS One* 12, e0172364. [PubMed: 28212447]
- (62). Dai S, Li J, Zhang H, Chen X, Guo M, Chen Z, and Chen Y (2020) Structural Basis for DNA Recognition by FOXC1 and the Characterization of Disease-causing FOXC1 Mutations. *Journal of Molecular Biology* 432, 6146–6156. [PubMed: 33058871]
- (63). Elian FA, Yan E, and Walter MA (2018) FOXC1, the new player in the cancer sandbox. *Oncotarget* 9, 8165–8178. [PubMed: 29487724]
- (64). Vegas N, Cavallin M, Maillard C, Boddaert N, Toulouse J, Schaefer E, Lerman-Sagie T, Lev D, Magalie B, Moutton S, Haan E, Isidor B, Heron D, Milh M, Rondeau S, Michot C, Valence S, Wagner S, Hully M, Mignot C, Masurel A, Datta A, Odent S, Nizon M, Lazaro L, Vincent M, Cogné B, Guerrot AM, Arpin S, Pedespan JM, Caubel I, Pontier B, Troude B, Rivier F, Philippe C, Bienvenu T, Spitz M-A, Bery A, and Bahi-Buisson N (2018) Delineating *FOXC1* syndrome: From congenital microcephaly to hyperkinetic encephalopathy. *Neurol Genet* 4, e281. [PubMed: 30533527]
- (65). MacArthur MW, and Thornton JM (1991) Influence of proline residues on protein conformation. *Journal of Molecular Biology* 218, 397–412. [PubMed: 2010917]
- (66). Petrovi D, Wang X, and Strodel B (2018) How accurately do force fields represent protein side chain ensembles? *Proteins* 86, 935–944. [PubMed: 29790608]
- (67). Ruan H, Yu C, Niu X, Zhang W, Liu H, Chen L, Xiong R, Sun Q, Jin C, Liu Y, and Lai L (2021) Computational strategy for intrinsically disordered protein ligand design leads to the discovery of p53 transactivation domain I binding compounds that activate the p53 pathway. *Chem. Sci.* 12, 3004–3016.
- (68). Kucukkal TG, Yang Y, Uvarov O, Cao W, and Alexov E (2015) Impact of Rett Syndrome Mutations on MeCP2 MBD Stability. *Biochemistry* 54, 6357–6368. [PubMed: 26418480]
- (69). Ha Y, Li J, Chen Y, Chen L, and Chen Z (2017) Expression and purification of FOXO1 DNA binding domain and its DNA properties. *Zhong Nan Xue Xue Bao Yi Xue Ban* 42, 1–7.

- (70). Gapsys V, Michielssens S, Seeliger D, and de Groot BL (2015) pmx: Automated protein structure and topology generation for alchemical perturbations. *J. Comput. Chem.* 36, 348–354. [PubMed: 25487359]
- (71). Anandakrishnan R, Aguilar B, and Onufriev AV (2012) H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Research* 40, W537–W541. [PubMed: 22570416]
- (72). Li P, and Merz KM (2016) MCPB.py: A Python Based Metal Center Parameter Builder. *J. Chem. Inf. Model.* 56, 599–604. [PubMed: 26913476]
- (73). Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, and Lindahl E (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1–2, 19–25.
- (74). Shirts MR, and Chodera JD (2008) Statistically optimal analysis of samples from multiple equilibrium states. *The Journal of Chemical Physics* 129, 124105. [PubMed: 19045004]
- (75). Klimovich PV, Shirts MR, and Mobley DL (2015) Guidelines for the analysis of free energy calculations. *J Comput Aided Mol Des* 29, 397–411. [PubMed: 25808134]
- (76). Joung IS, and Cheatham TE (2008) Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B* 112, 9020–9041. [PubMed: 18593145]
- (77). Joung IS, and Cheatham TE (2009) Molecular Dynamics Simulations of the Dynamic and Energetic Properties of Alkali and Halide Ions Using Water-Model-Specific Ion Parameters. *J. Phys. Chem. B* 113, 13279–13290. [PubMed: 19757835]
- (78). Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, and Pedersen LG (1995) A smooth particle mesh Ewald method. *The Journal of Chemical Physics* 103, 8577–8593.
- (79). Gonzalez TF (1985) Clustering to minimize the maximum intercluster distance. *Theoretical Computer Science* 38, 293–306.
- (80). Scherer MK, Trendelkamp-Schroer B, Paul F, Pérez-Hernández G, Hoffmann M, Plattner N, Wehmeyer C, Prinz J-H, and Noé F (2015) PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *J. Chem. Theory Comput.* 11, 5525–5542. [PubMed: 26574340]
- (81). Wehmeyer C, Scherer MK, Hempel T, Husic BE, Olsson S, and Noé F (2019) Introduction to Markov state modeling with the PyEMMA software [Article v1.0]. *LiveCoMS* 1.
- (82). Scherer MK, Husic BE, Hoffmann M, Paul F, Wu H, and Noé F (2019) Variational Selection of Features for Molecular Kinetics. *J. Chem. Phys.* 150, 194108. [PubMed: 31117766]
- (83). McGibbon RT, Beauchamp KA, Harrigan MP, Klein C, Swails JM, Hernández CX, Schwantes CR, Wang L-P, Lane TJ, and Pande VS (2015) MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophysical Journal* 109, 1528–1532. [PubMed: 26488642]

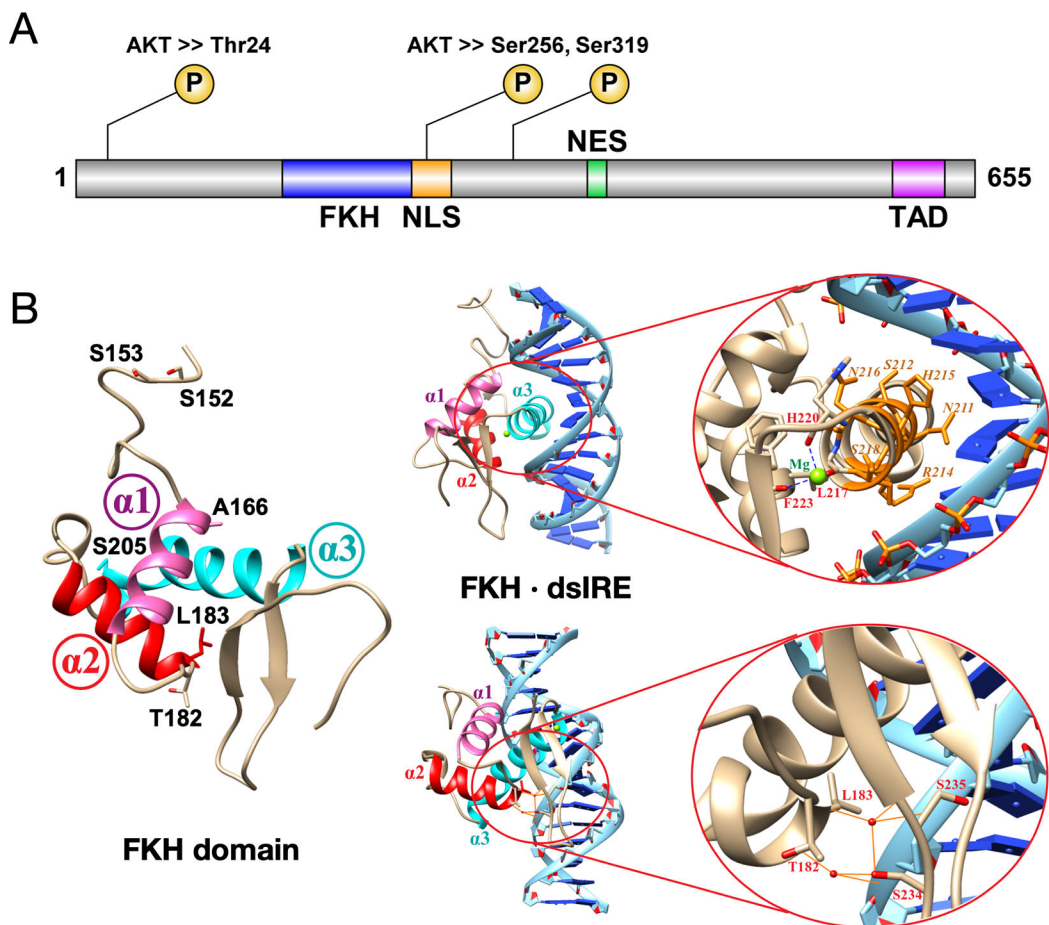


Figure 1.

(A) FOXO1 functional domains are shown with three AKT-mediated phosphorylation sites for inactivation. FKH is forkhead DNA-binding domain; NLS is nuclear localization sequence/signal; NES is nuclear export sequence/signal; TAD is transactivation domain. (B) The FKH domain (PDB 3COA) has three helices labeled $\alpha 1$ (pink), $\alpha 2$ (red) and $\alpha 3$ (cyan). The six residues where all eight mutations are found are labeled; none are located in helix $\alpha 3$, which makes contacts with binding sequence IRE. Residues T182, L183, S234 and S235 make key hydrogen bonds with the DNA phosphate backbone, mediated by water molecules (red spheres).

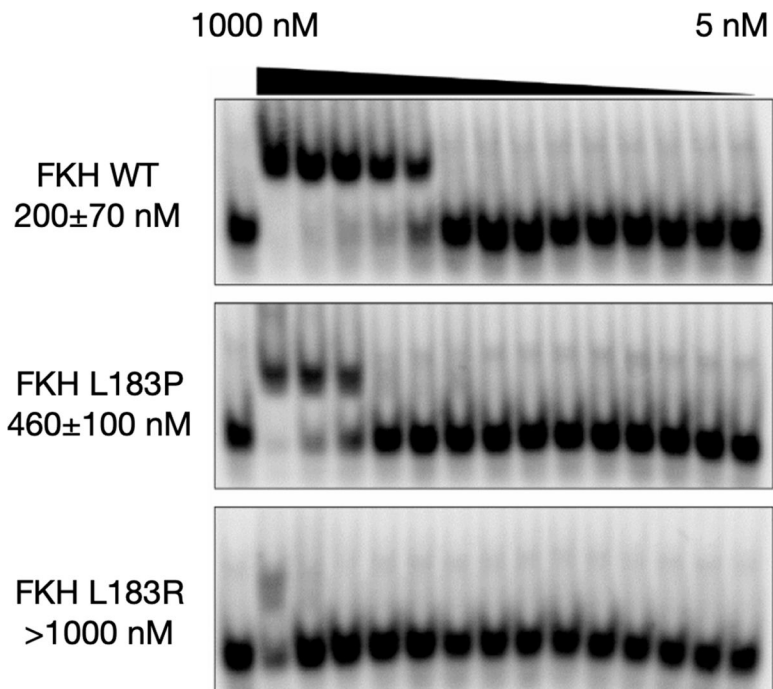


Figure 2. EMSA analysis of FKH binding to the insulin response element (IRE). Representative blots of FKH wt, FKH L183P and FKH L183R serial dilution from 1 μ M (1000 nM) to 5 nM in complex with 1 nM dsIRE. Each blot is labeled with K_D estimates in nM.

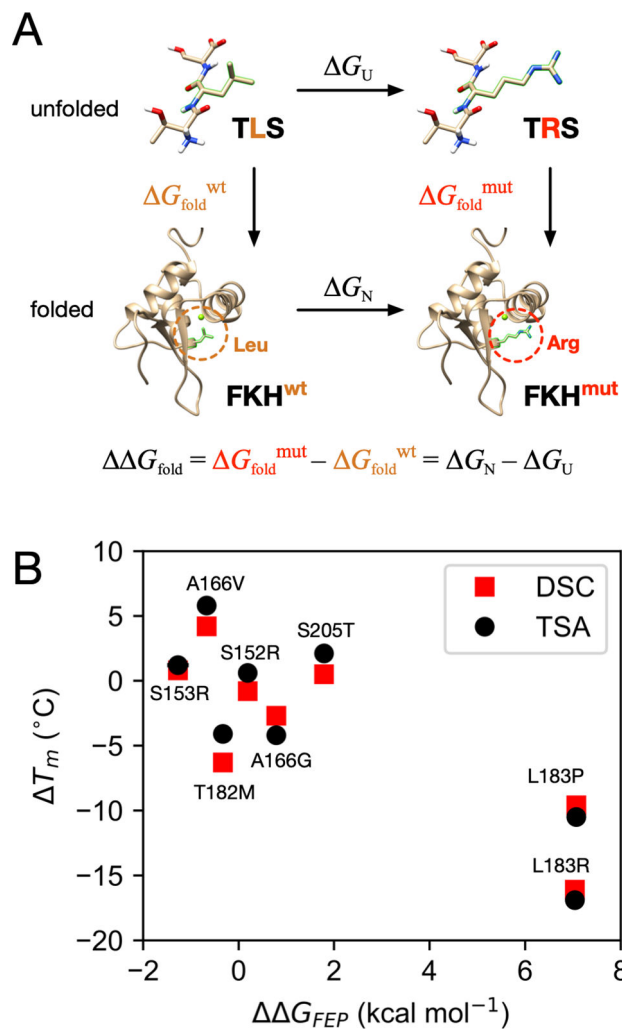
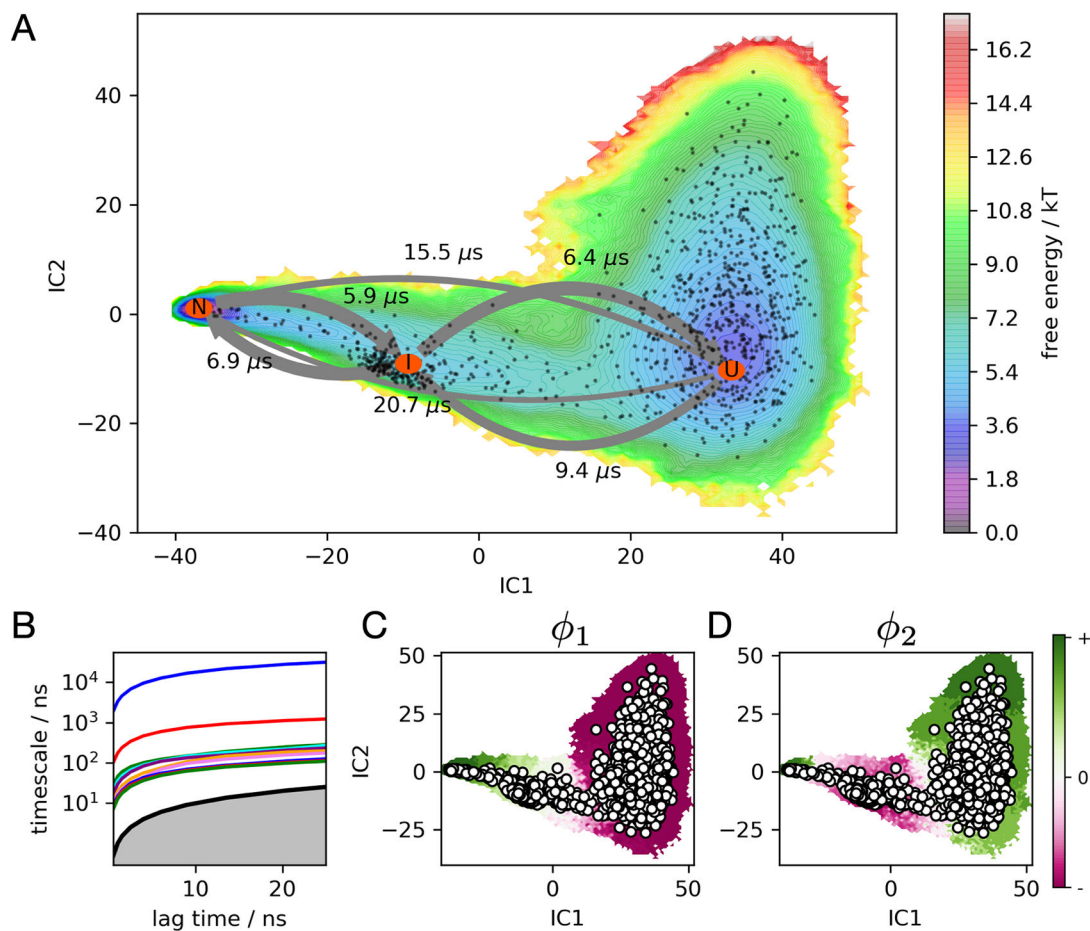


Figure 3. (A) A thermodynamic cycle illustrates how FEP is used to calculate G_{fold} for eight FOXO1 FKH mutants, with the L183R mutant shown as an example. (B) Correlation of G_{fold} values calculated using FEP (G_{FEP}) with T_m from TSA (black) and DSC (red). Uncertainty estimates for G_{FEP} are smaller than the marker width.

**Figure 4.**

(A) FOXO1 folding free energy landscape at 375 K calculated from the MSM, projected to the two largest TICA components. MSM macrostates U, I, and N are labeled with orange ovals, connected by arrows whose widths represent transition frequencies, labeled with mean first passage times. Black dots represent MSM microstates. (B) An implied timescale plot shows the ten slowest timescales as a function of the MSM lag time. (C and D) The first and second relaxation eigenmodes from the MSM, colored by sign structure. The first eigenmode shows population flux from N to U at 375 K, while the second eigenmode shows flux into I from U and N.

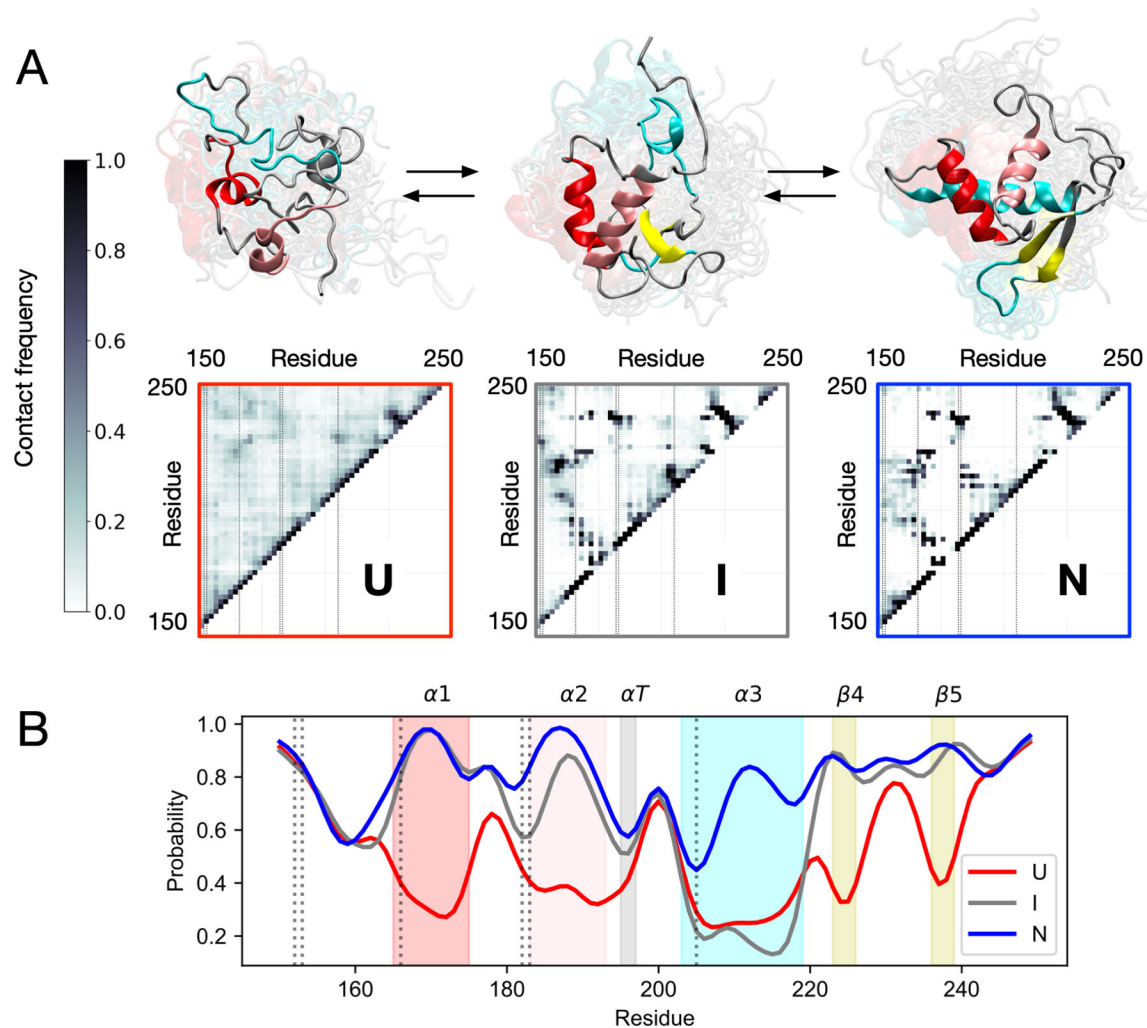


Figure 5. Simulations predict helix $\alpha 3$ formation to be the final step in folding. (A) Interresidue contact frequencies and representative structures of MSM macrostates U, I and N. Vertical lines on the contact maps denote mutant positions. (B) Per-residue populations of native secondary structure for each macrostate. Shaded bands mark native secondary structures $\alpha 1$, $\alpha 2$, αT , $\alpha 3$, $\beta 4$ and $\beta 5$, with vertical lines denoting mutant positions.

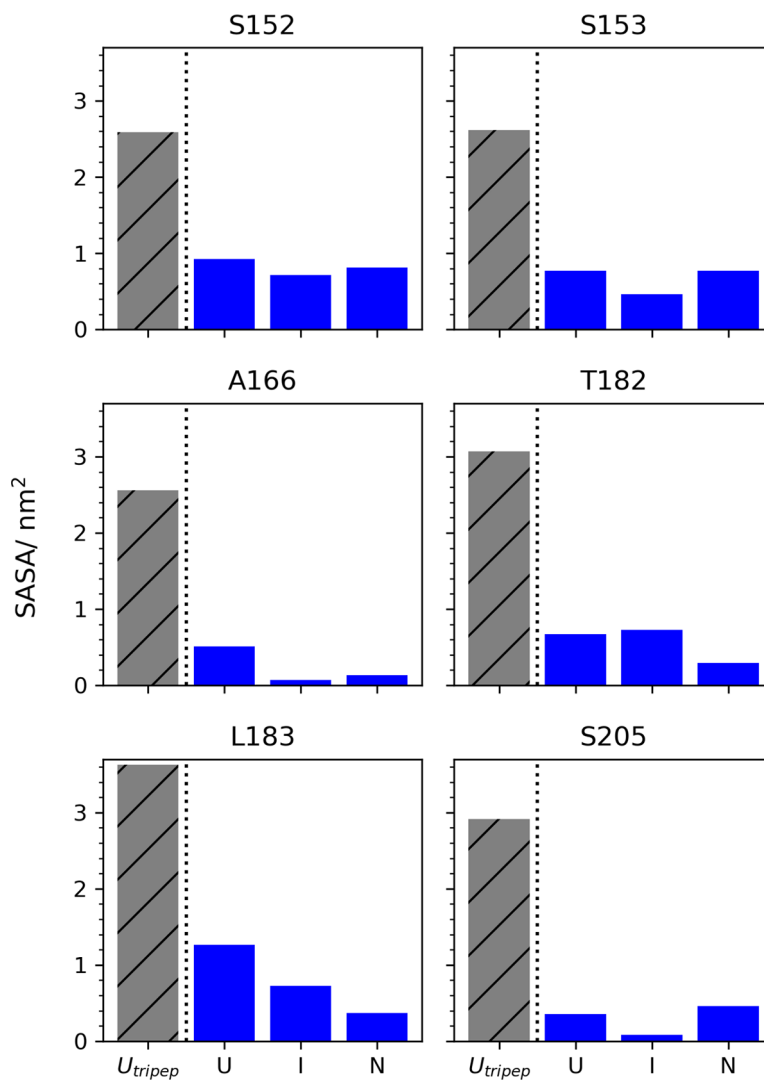


Figure 6. Tripeptide models of the unfolded state consistently overestimate per-residue solvent-accessible surface area (SASAs) compared to molecular simulations. For each mutant position (S152, S153, A166, T182, L183 and S205), bar graphs show average SASA calculated from tripeptide FEP simulations (patterned gray), versus average SASA of MSM macrostates U, I, and N (blue).

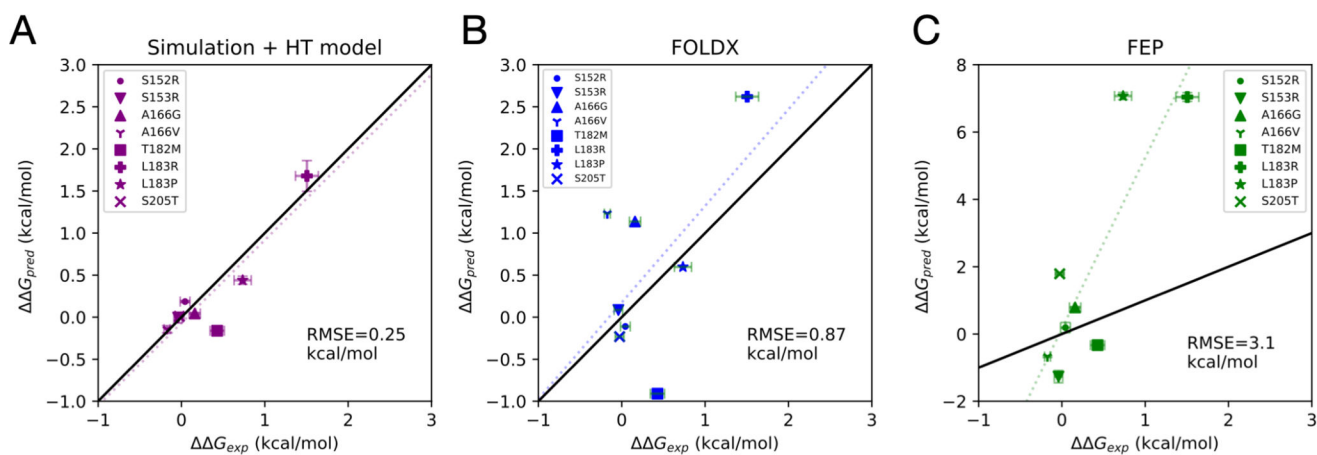


Figure 7.

Comparison of estimated experimental folding free energy changes upon mutation ΔG_{exp} , and predicted estimates ΔG_{pred} from three different computational methods: (A) a combined molecular simulation and hydrophobic transfer (HT) model, (B) predictions from the FoldX algorithm, and (C) alchemical FEP estimation. In each plot, the dotted line is the least-squares linear fit, and the root mean squared error (RMSE).

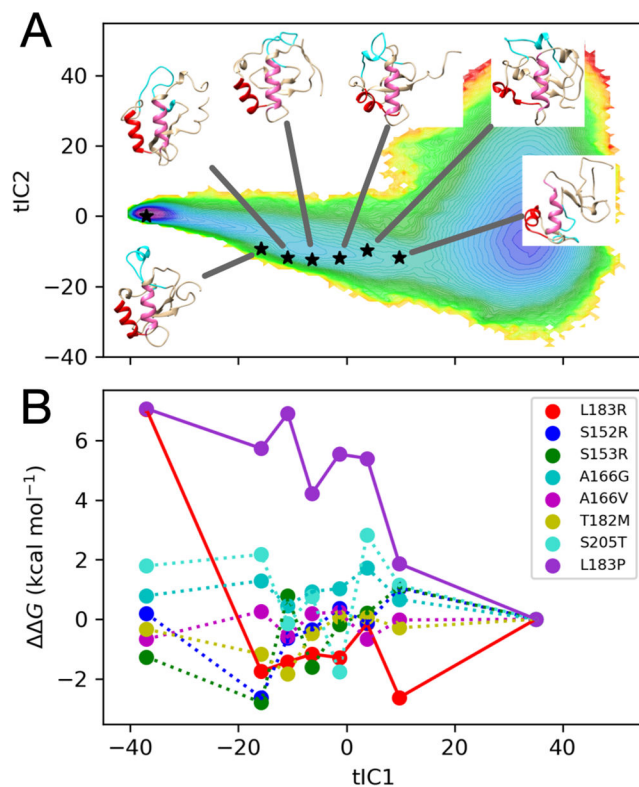


Figure 8.

(A) Six folding intermediates i along the tIC1 folding coordinate were chosen as templates for FEP estimation of $G = G_i - G_U$, where the tripeptide model of the unfolded state U is used as a reference. (B) Estimates of G along tIC1 suggest that perturbations from mutations are dependent on the folding reaction coordinate, particularly for L183R and L183P (solid lines).

Table 1.

Melting temperatures and enthalpies of unfolding measured by TSA and DSC experiments.

| FKH construct | TSA | | DSC ^c | | H_{cal} (kcal/mol) | H_{vH} (kcal/mol) | H_{vH-cal} (kcal/mol) |
|--------------------|------------|-------|------------------|-------|----------------------|---------------------|-------------------------|
| | T_m (°C) | T_m | T_m (°C) | T_m | | | |
| WT | 50.5±0.8 | - | 57.8 ± 0.3 | - | 47 ± 1 | 54 ± 1 | 7 |
| S152R | 51.1±0.1 | 0.6 | 57.0 ± 1.0 | -0.8 | 51 ± 1 | 58 ± 1 | 7 |
| S153R | 51.8±0.1 | 1.2 | 58.6 ± 1.0 | 0.8 | 52 ± 1 | 60 ± 1 | 8 |
| A166G | 46.4±0.5 | -4.2 | 55.1 ± 1.0 | -2.7 | 44 ± 1 | 54 ± 1 | 10 |
| A166V | 56.3±0.9 | 5.8 | 61.9 ± 1.0 | 4.2 | 47 ± 1 | 68 ± 1 | 21 |
| K171E | 52.5±0.8 | 2.0 | - | - | - | - | |
| T182M ^a | 46.4±0.1 | -4.1 | 51.5 ± 1.0 | -6.3 | 40 ± 1 | 62 ± 1 | 22 |
| L183R ^b | 33.6±1.1 | -16.9 | 41.7 ± 1.0 | -16.1 | 50 ± 1 | 42 ± 1 | -8 |
| L183P | 40.0±0.6 | -10.5 | 48.2 ± 1.0 | -9.6 | 31 ± 1 | 53 ± 1 | 22 |
| S205N | 52.4±1.1 | 1.9 | - | - | - | - | |
| S205T | 52.7±1.7 | 2.1 | 58.3 ± 1.0 | 0.5 | 58 ± 1 | 59 ± 1 | 1 |

^a Mutation in OCI-Ly8 cell line.^b Mutation in WSU-NHL cell line.^c Wild-type measurements were performed in triplicate (± 0.3 °C), while single measurements were made for mutants (± 1.0 °C)

Table 2.

DNA oligonucleotide sequences used for FKH binding studies.

| Oligo | Sequences ^a |
|--------|----------------------------------|
| IRE-F | TAAGCTAGTGGTTTGTGCTTGCTAGCAAT |
| IRE-R | ATTGCTAGCAAGCAAAACAAACCACTAGCTTA |
| DBE1-F | TAAGCTAGTGGTTGTTACCTTGCTAGCAAT |
| DBE1-R | ATTGCTAGCAAGGTAACAAACCACTAGCTTA |
| DBE2-F | TAAGCTAGTCTGTTTACATTTGCTAGCAAT |
| DBE2-R | ATTGCTAGCAAAATGTAACAAGACTAGCTTA |

^aAll oligos 5'-biotinylated.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3.

Estimated G_{fold} values for FOXO1 from DSC melting temperatures, and predicted G_{fold} values from three different computational methods: a hydrophobic transfer model (G_{HT}), the FoldX algorithm (G_{FoldX}), and free energy perturbation (FEP, G_{FEP}). All values are reported in kcal/mol.

| Mutation | G_{fold} (DSC) ^a | G_{HT} | G_{FoldX} | G_{FEP} |
|----------|--------------------------------------|-----------------|--------------------|------------------|
| S152R | 0.04 ± 0.06 | 0.185 ± 0.01 | -0.110 | 0.19 ± 0.15 |
| S153R | -0.04 ± 0.05 | -0.004 ± 0.06 | 0.085 | -1.27 ± 0.18 |
| A166G | 0.16 ± 0.07 | 0.044 ± 0.01 | 1.137 | 0.79 ± 0.02 |
| A166V | -0.17 ± 0.04 | -0.143 ± 0.04 | 1.231 | -0.67 ± 0.03 |
| T182M | 0.43 ± 0.09 | -0.161 ± 0.03 | -0.911 | -0.33 ± 0.06 |
| L183R | 1.51 ± 0.14 | 1.679 ± 0.18 | 2.621 | 7.04 ± 0.13 |
| L183P | 0.73 ± 0.10 | 0.44 ± 0.05 | 0.597 | 7.07 ± 0.11 |
| S205T | -0.03 ± 0.05 | 0.009 ± 0.002 | -0.231 | 1.79 ± 0.03 |

^a calculated from the theory of Robertson and Murphy, using $N_{\text{res}}=100$ (see Materials and Methods).

Table 4.

The location of native-state contacts for residues with selected oncogenic mutations.

| Residue | Crystal Contacts |
|---------|--|
| A166 | T170, N216, H220, F223 |
| T182 | I186, N228, S234, S235, M236, W237 |
| L183 | T187, I213, R214, L217, S234, S235, W237 |
| S205 | D199, K200, G201, W209 |

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript