Supporting Information

Law et al. 10.1073/pnas.1405831111

SI Text

Simulation Model and Details. To study the allosteric binding involved in the mixed-lineage leukemia–kinase-inducible domain interacting (MLL-KIX)-c-Myb ternary complex, we used the sequence-flavored Gō-like model developed by Karanicolas and Brooks (1, 2). In this model, each protein residue is represented as a single bead with a mass equal to its corresponding amino acid, centered at its Cα atom, and connected to neighboring residues along the protein backbone via virtual bonds. The potential energy function, V, consists of both bonded and nonbonded terms:

$$V = V_{\text{bond}} + V_{\text{angle}} + V_{\text{dihedral}} + V^{\text{nonnative}}_{\text{nonbonded}} + V^{\text{native}}_{\text{nonbonded}}, \quad \text{[S1]}$$

where V_{bond} and V_{angle} are harmonic potentials with equilibrium values set to those found in the NMR structure (3). V_{dihedral} is a statistical potential based on probability distributions obtained from Ramachandran plots for each of the 20 × 20 amino acid pairs and, thus, provides additional sequence-specific information while remaining independent of the protein topology and avoiding locally driven folding that results from directly incorporating native dihedral angles into V_{dihedral}. Native nonbonded terms mediating the interaction between residue pairs, i and j, are modeled using a 12-10-6 Lennard-Jones–type potential whose interaction strength is proportional to the statistical potential reported for the residue types of i and j by Miyazawa and Jernigan (4). All nonnative interactions, defined by residue pairs with all side-chain heavy atoms separated by more than 4.5 Å and that do not form backbone hydrogen bonds, were subject to repulsive interactions. We can also separate $V^{\text{native}}_{\text{nonbonded}}$ into its intramolecular and intermolecular components as follows:

$$V^{\text{native}}_{\text{nonbonded}} = \alpha V^{\text{native}}_{\text{intramolecular}} + \beta V^{\text{native}}_{\text{intermolecular}}, \quad \text{[S2]}$$

where α and β are scaling factors used for renormalizing the intramolecular and intermolecular interaction energies, respectively (see below). A detailed description of this Gō-like model can be found in refs. 1, 2, and 5.

The cumulative production simulation time for each simulation model is 900 μs (i.e., each set consists of 60 independent × 15-μs-long simulations) and with at least 70 c-Myb binding events and 150 MLL binding events being observed in total. With the exception of KIX-free, which was simulated in a 60 Å per edge cubic volume, each system was simulated with a 120 Å per edge cubic volume using a periodic volume to account for the finite concentration of species in the simulation, e.g., MLL + KIX, etc. All simulations were performed at 300 K using Langevin dynamics with a 0.1 ps^{-1} friction coefficient along with a 15-fs simulation time step. The cutoff for nonbonded interactions was set to 25 Å, and virtual bonds were constrained using SHAKE (6).

Calibrating the Gō-Like Model. In accordance with the calibration protocol described in ref. 5, the intramolecular interaction energies of c-Myb and MLL were rescaled by systematically altering α in Eq. **S2** above so as to match the experimentally reported helical content. The intermolecular interaction energies between KIX-c-Myb and MLL-KIX were also recalibrated by tuning β so as to match the experimentally reported binary dissociation constants. The intramolecular interaction energies for KIX were unaltered (i.e., $\alpha_{KIX} = 1.0$). All simulations of the MLL-KIX-c-Myb ternary complex use only the tuned α and β parameters from the monomer and binary systems, respectively. No further adjustments are made. The 1.5-μs-long simulations were initially used to calibrate α. The percent helicity, measured as the fraction of the native intramolecular contacts formed, was found to be identical even with 10 times more sampling. Simulations for calibrating β (eventually used in the production simulations) were 15 μs in length and repeated 30 times to provide an accurate estimate of the experimentally reported dissociation constants.

When the native Gō-like model was used ($\alpha = 1.0$), the fraction of helical content was ∼70% and ∼80% for c-Myb and MLL, respectively (Fig. S1 *A* and *B*). Therefore, the intramolecular interaction energies were reduced by rescaling α to match the helical content reported in the literature (∼25–30% for c-Myb) (7–9) and as predicted empirically by AGADIR (∼1–5% for MLL) (10–13). To achieve this, the final intramolecular interaction energies were scaled down to $\alpha_{c\text{-}Myb} = 0.45$ and $\alpha_{MLL} = 0.05$ for c-Myb and MLL, respectively (Fig. S1 *A* and *B*). In contrast, the intermolecular interaction energies for the KIX-c-Myb and MLL-KIX binary complexes were increased to $\beta_{KIX\text{-}c\text{-}Myb} = 1.2$ and $\beta_{MLL\text{-}KIX} = 1.55$, respectively, to match the reported binary dissociation constants (14) (Fig. S1 *C* and *D* and Table 1 in the main text).

Although all of the intramolecular and intermolecular native contacts were derived from the NMR ternary complex, it is important to reiterate that our recalibration process only targets experimental observables from the monomeric and binary systems. In other words, no additional tuning was made to influence cooperative binding ($\alpha_{c\text{-}Myb} = 0.45$ and $\alpha_{MLL} = 0.05$ were used when scaling β, but none of the scaling parameters was tuned to match the ternary dissociation constants). Thus, one would not expect changes in the binding affinities for the ternary complex unless allostery was at play and could be adequately captured by our coarse-grained model. Additionally, unlike traditional all-atom simulations where only a single binding/unbinding event is often observed, we recorded multiple binding/unbinding events from long, continuous trajectories and so any biases caused by the starting structure are minimized as a result of increased statistics.

Fraction of Bound and Unbound States. If the Cα-Cα distance between any given native interaction (or, equivalently, native contact) was within 1 Å of its native contact distance (found in the NMR structure), then the native contact was considered formed. We defined a bound state as having at least one native intermolecular contact being formed or a configuration that had last visited a bound state (without first visiting an unbound state) and with at least one native intermolecular contact within 25 Å (corresponding to the cutoff distance for nonbonded interactions). This helps to remove any bias in rebinding coming from long-range interactions and ensures that binding events are uncorrelated. Conversely, an unbound state was defined as having all native intermolecular contact distances greater than 25 Å or a configuration that had last visited an unbound state and had no native intermolecular contacts. Together, these definitions provide a clear delineation between the bound and unbound states.

The dissociation constants, K_{d}, were calculated from the following:

$$K_{\text{d}} = \frac{1,660}{V_0} \cdot \frac{P^2_{\text{unbound}}}{1 - P_{\text{unbound}}}, \quad \text{[S3]}$$

where V_0 is the box volume in units of $Å^3$, 1,660 converts the concentration from units of molecules per $Å^3$ to units of moles per liter, and P_{unbound} is the fraction of unbound states.

Contact Appearance Order. To assess whether or not the binding of one peptide changes the order in which the native intermolecular contacts are formed between a second peptide and KIX, we adopted the contact appearance order metric originally developed to study protein folding (15). Here, we record the order in which each native intermolecular contact is formed during N independent c-Myb or MLL binding events. Then, we calculate the probability of each native intermolecular contact being formed at a given order in time.

Kinetic Rates. By treating the ligand binding process as a simple two-state process with either bound or unbound states, we estimated the kinetic rate constants, k_{off} and k_{on}, by calculating the mean first passage time (MFPT) for the off and on reactions:

$$k_{off} = \frac{1}{\mathrm{MFPT}_{off}}; \qquad k_{on} = \frac{1}{\mathrm{MFPT}_{on} \cdot \left(\frac{1,660}{V_0} \cdot P_{unbound} \right)}, \qquad \textbf{[S4]}$$

where k_{off} has units of seconds^{-1} and k_{on} is dependent on the concentration (or fraction of unbound states) and so has units of molar^{-1}·seconds^{-1}.

Principal Component Analysis. The overall translational and rotational motion was removed by fitting each simulation snapshot to the average KIX structure. Then, a symmetric covariance matrix, C, of the positional deviations was constructed as follows:

$$C = \left\langle (r(t) - \langle r \rangle)(r(t) - \langle r \rangle)^T \right\rangle, \qquad \textbf{[S5]}$$

where $< >$ denotes an ensemble average and $r(t)$ is a $3N$-dimensional vector of x, y, and z coordinates for all N atoms at some simulation time, t. C can then be diagonalized by an orthogonal coordinate transformation to obtain the mean square fluctuations (eigenvalues) along each principal component/mode (eigenvectors) of the system. When the eigenvectors are sorted by their corresponding eigenvalues in decreasing order, the total fluctuations of the system can often be described by the first few lowest frequency modes.

In Silico Mutations. Six C-terminal residues (Glu663, Glu666, Lys667, Ser670, Arg671, and Leu672) of the α_3 helix, which make no intermolecular native contacts with either c-Myb or MLL, were chosen for mutational studies to increase the helicity in this region. The KIX sequence was mutated (to either Ala, Leu, Arg, Met, Lys, Asn, Glu, Ile, Trp, or Ser) at each of the six target positions and the increase in the percent helicity was predicted using AGADIR (10–13) (Table S1).

1. Karanicolas J, Brooks CL, 3rd (2002) The origins of asymmetry in the folding transition states of protein L and protein G. *Protein Sci* 11(10):2351–2361.
2. Karanicolas J, Brooks CL, 3rd (2003) Improved Gō-like models demonstrate the robustness of protein folding mechanisms towards non-native interactions. *J Mol Biol* 334(2):309–325.
3. De Guzman RN, Goto NK, Dyson HJ, Wright PE (2006) Structural basis for cooperative transcription factor binding to the CBP coactivator. *J Mol Biol* 355(5):1005–1013.
4. Miyazawa S, Jernigan RL (1996) Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. *J Mol Biol* 256(3):623–644.
5. Ganguly D, Chen JH (2011) Topology-based modeling of intrinsically disordered proteins: Balancing intrinsic folding and intermolecular interactions. *Proteins* 79(4): 1251–1266.
6. Ryckaert JP, Ciccotti G, Berendsen HJC (1977) Numerical-integration of Cartesian equations of motion of a system with constraints—molecular dynamics of *n*-alkanes. *J Comput Phys* 23(3):327–341.
7. Zor T, Mayr BM, Dyson HJ, Montminy MR, Wright PE (2002) Roles of phosphorylation and helix propensity in the binding of the KIX domain of CREB-binding protein by constitutive (c-Myb) and inducible (CREB) activators. *J Biol Chem* 277(44): 42241–42248.
8. Parker D, et al. (1999) Role of secondary structure in discrimination between constitutive and inducible activators. *Mol Cell Biol* 19(8):5601–5607.
9. Shammas SL, Travis AJ, Clarke J (2013) Remarkably fast coupled folding and binding of the intrinsically disordered transactivation domain of cMyb to CBP KIX. *J Phys Chem B* 117(42):13346–13356.
10. Muñoz V, Serrano L (1994) Elucidating the folding problem of helical peptides using empirical parameters. *Nat Struct Biol* 1(6):399–409.
11. Muñoz V, Serrano L (1995) Elucidating the folding problem of helical peptides using empirical parameters. II. Helix macrodipole effects and rational modification of the helical content of natural peptides. *J Mol Biol* 245(3):275–296.
12. Muñoz V, Serrano L (1995) Elucidating the folding problem of helical peptides using empirical parameters. III. Temperature and pH dependence. *J Mol Biol* 245(3): 297–308.
13. Muñoz V, Serrano L (1997) Development of the multiple sequence approximation within the AGADIR model of α-helix formation: Comparison with Zimm-Bragg and Lifson-Roig formalisms. *Biopolymers* 41(5):495–509.
14. Goto NK, Zor T, Martinez-Yamout M, Dyson HJ, Wright PE (2002) Cooperativity in transcription factor binding to the coactivator CREB-binding protein (CBP). The mixed lineage leukemia protein (MLL) activation domain binds to an allosteric site on the KIX domain. *J Biol Chem* 277(45):43168–43174.
15. Gin BC, Garrahan JP, Geissler PL (2009) The limited role of nonnative contacts in the folding pathways of a lattice protein. *J Mol Biol* 392(5):1303–1314.

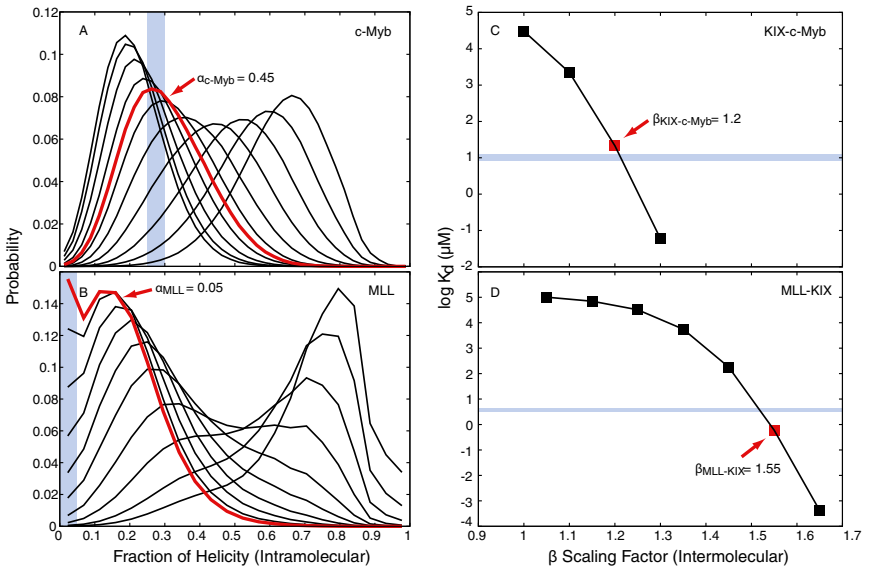

Fig. S1. Calibration of (*A* and *B*) the helical content (α increases from *Left* to *Right*) and (*C* and *D*) the binding affinities (K_d). The final simulated scaling factors are indicated by red arrows and the target ranges are shown as blue bars [c-Myb: ~25–30% helicity (7–9); MLL: ~1–5% helicity (10–13); KIX-c-Myb: $K_d = 10 \pm 2$ μM (14); MLL-KIX: $K_d = 2.8 \pm 0.4$ μM (14)].

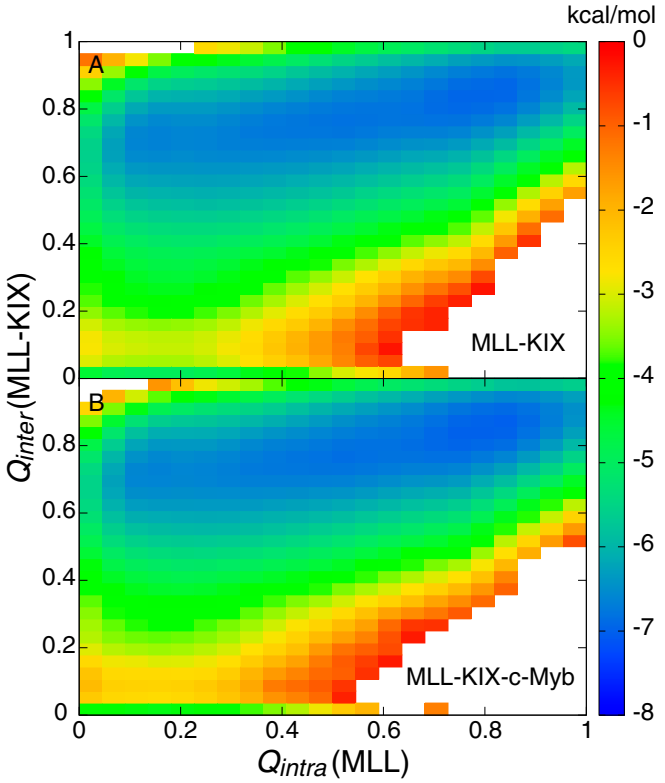

Fig. S2. Two-dimensional free-energy surfaces as a function of MLL folding and binding in (*A*) the absence of c-Myb and (*B*) the presence of c-Myb. Q_{intra} is the fraction of native intramolecular contacts in MLL, whereas Q_{inter} is the fraction of native intermolecular contacts formed between MLL-KIX.

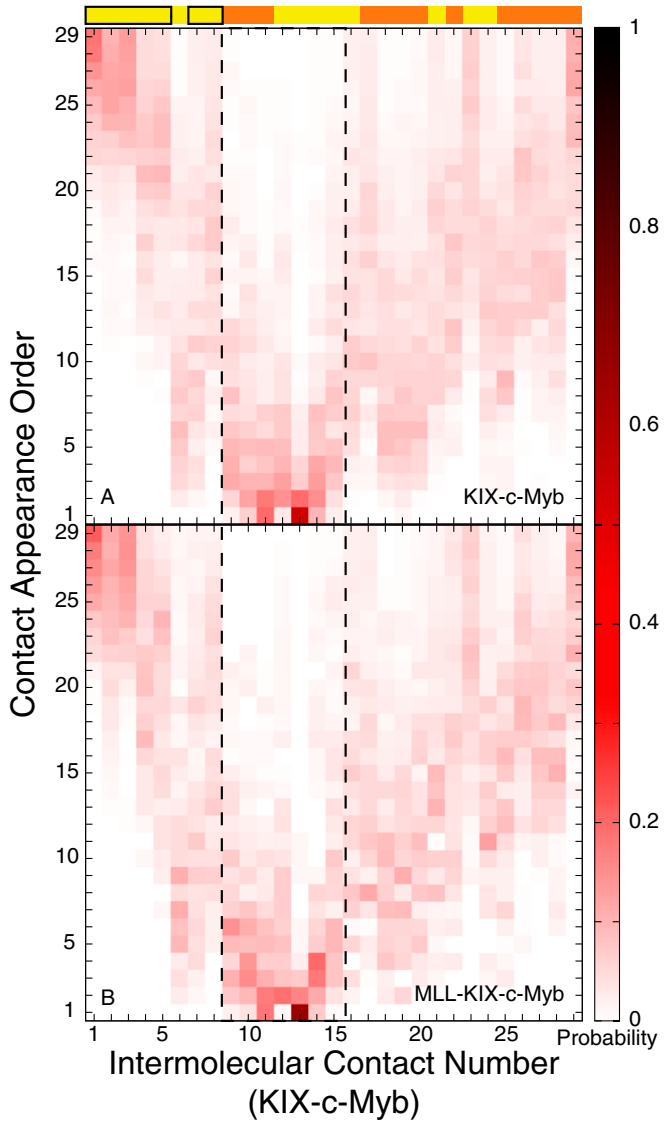

Fig. S3. Contact appearance order (CAO) for c-Myb binding in (*A*) the absence of MLL and (*B*) the presence of MLL. The horizontal axis corresponds to native intermolecular contacts made between KIX-c-Myb, and low, medium, and high contact numbers correspond to contacts made by the N terminus, middle, and C terminus of c-Myb, respectively. The order in which contacts are formed between KIX-c-Myb is shown along the vertical axis and can be thought of as a time axis. The white squares have a zero probability, and the dark red squares have the highest probability of being the *n*th contact formed. The colored bar above the figure shows where c-Myb contacts KIX with orange and yellow squares corresponding to helices α_1 and α_3, respectively. The C terminus of α_3 is denoted by solid black boxes. Intermolecular native contacts made by Leu302 are enclosed by a dashed box (contact numbers 9–15).

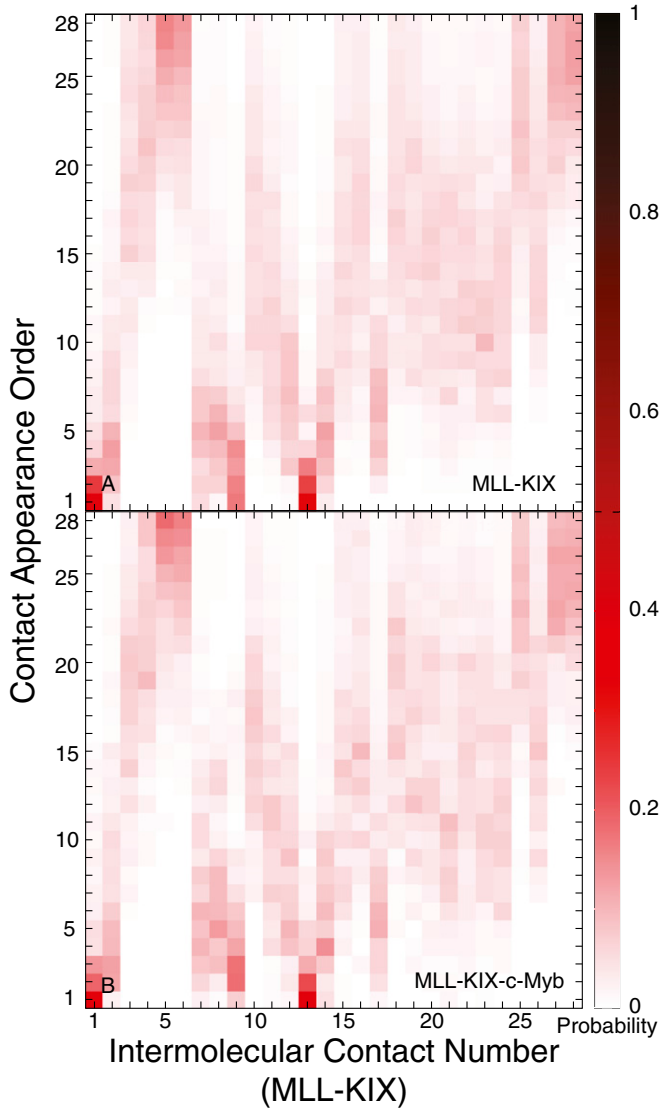

Fig. S4. Contact appearance order (CAO) for MLL binding in (*A*) the absence of c-Myb and (*B*) the presence of c-Myb. The horizontal axis corresponds to native intermolecular contacts made between MLL-KIX, and low, medium, and high contact numbers correspond to contacts made by the N terminus, middle, and C terminus of MLL, respectively. The order in which contacts are formed between MLL-KIX is shown along the vertical axis and can be thought of as a time axis. The white squares have a zero probability, and the dark red squares have the highest probability of being the *n*th contact formed.

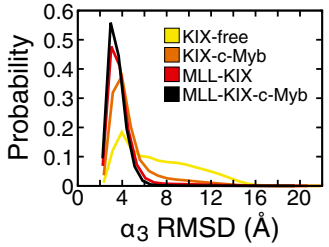

Fig. S5. Helicity of the α_3 helix. The rmsd is calculated with respect to the fully helical α_3 in the NMR structure and rmsd values of more than ~5 Å correspond to a significant loss of secondary structure.

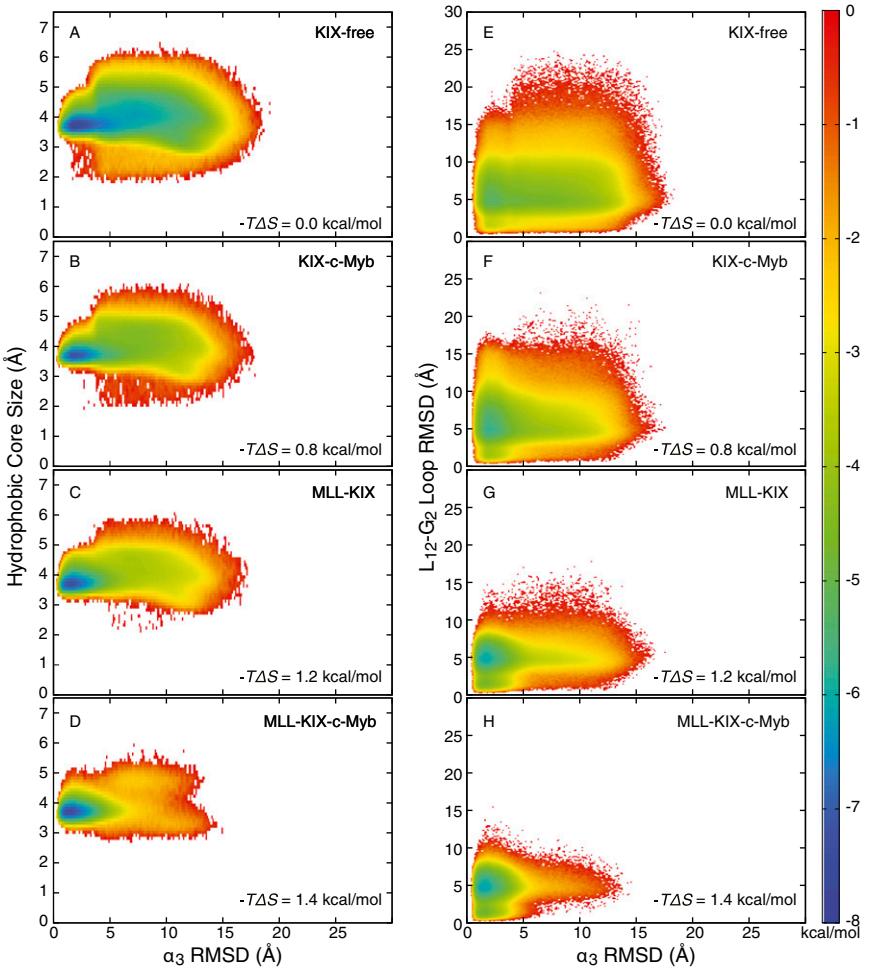

Fig. S6. Two-dimensional free-energy surfaces as a function of the α_3 rmsd and (*A–D*) the hydrophobic core size or (*E–H*) the L_{12}-G_2 loop rmsd. All rmsd values were calculated relative to the NMR structure, which had a hydrophobic core size of ~3.6 Å. All ΔS were calculated with respect to KIX-free at 300 K (*Methods* in the main text).

Table S1. KIX point mutations predicted to increase the helicity of α_3

Point mutation	Percent helicity increase*
K667L	4.4 (1.5)
K667R	9.2 (3.1)
S670L	3.5 (1.2)
S670R	9.5 (3.2)
S670K	7.5 (2.5)
S670N	12.7 (4.2)

*Calculated from α_3 (residues 646–672) and, in parentheses, from the full length of KIX (residues 586–672).