



Published in final edited form as:

J Chem Theory Comput. 2009 September 8; 5(9): 2486–2502. doi:10.1021/ct9001812.

Quantifying Correlations Between Allosteric Sites in Thermodynamic Ensembles

Christopher L. McClendon,

University of California San Francisco, Graduate Group in Biophysics and Department of Pharmaceutical Chemistry

Gregory Friedland,

California Institute for Quantitative Biosciences, University of California San Francisco, Graduate Group in Biophysics and Dept. of Bioengineering and Therapeutic Sciences; Sandia National Laboratories, Joint BioEnergy Institute; University of California San Francisco, Department of Pharmaceutical Chemistry

David L. Mobley,

University of California San Francisco, Department of Pharmaceutical Chemistry; University of New Orleans, Department of Chemistry; University of New Orleans, Department of Chemistry

Homeira Amirkhani, and

Sharif University of Technology, Department of Physics

Matthew P. Jacobson

University of California San Francisco, Pharmaceutical Chemistry

Abstract

Allostery describes altered protein function at one site due to a perturbation at another site. One mechanism of allostery involves correlated motions, which can occur even in the absence of substantial conformational change. We present a novel method, “MutInf”, to identify statistically significant correlated motions from equilibrium molecular dynamics simulations. Our approach analyzes both backbone and sidechain motions using internal coordinates to account for the gear-like twists that can take place even in the absence of the large conformational changes typical of traditional allosteric proteins. We quantify correlated motions using a mutual information metric, which we extend to incorporate data from multiple short simulations and to filter out correlations that are not statistically significant. Applying our approach to uncover mechanisms of cooperative small molecule binding in human interleukin-2, we identify clusters of correlated residues from 50 ns of molecular dynamics simulations. Interestingly, two of the clusters with the strongest correlations highlight known cooperative small-molecule binding sites and show substantial correlations between these sites. These cooperative binding sites on interleukin-2 are correlated not only through the hydrophobic core of the protein but also through a dynamic polar network of hydrogen bonding and electrostatic interactions. Since this approach identifies correlated conformations in an unbiased, statistically robust manner, it should be a useful tool for finding novel or “orphan” allosteric sites in proteins of biological and therapeutic importance.

Introduction

Originally, allosteric proteins were those where multiple subunits achieved cooperative binding through ligand-mediated shifts in conformational equilibria. Nowadays, allostery is broadly defined as any case in which an event at one site on a protein or complex impacts function, dynamics, or distribution of conformations of another site (for recent reviews see 1· 2). This broader definition includes single-domain proteins as well as proteins or complexes where

cooperativity occurs without substantial conformational change. Given this broader definition, it has been suggested that allostery is a property of many proteins³⁻⁵, but is only relevant when a localized event precipitates a change in function. Recently, there has been renewed interest in uncovering allosteric mechanisms of protein regulation and in discovering new allosteric sites, which are of significant interest in biological mechanisms of protein regulation and as novel sites for drug discovery⁶⁻⁸.

Typically, sites are identified as allosteric after mutational, structural, and thermodynamic characterization with allosteric protein, peptide, or small-molecule modulators, which are frequently found serendipitously⁹. As such, there has been much interest in computational approaches to identify novel allosteric sites. One of the most extensively used approaches has been the Statistical Coupling Analysis approach pioneered by Ranganathan and co-workers¹⁰⁻¹³, where pairs of residues that tend to be mutated together in multiple sequence alignments suggest coupling between protein sites. This approach has recently been used to engineer a novel allosteric network by combining predicted allosteric pathways from a light sensor and an enzyme¹⁴. However, this approach requires large multiple sequence alignments and the predicted couplings may or may not be relevant to particular proteins in the alignment¹⁵. Alternative methods to identify allosteric networks using sequence comparisons have also been described¹⁶.

Other computational methods to study allosteric mechanisms and identify potential sites for allosteric regulation focus on a protein's structure and dynamics. Cooper and Dryden showed that the free energy of cooperativity could be separated into two terms: one that accounts for changes in the protein's conformational distribution (i.e. by population shifts), and one that accounts for changes in the amplitudes and/or frequencies of protein vibrational motions. One approach to studying allostery is to focus on protein vibrations¹⁷⁻²⁰ around a static structure, often by a coarse-grained normal mode analysis²¹⁻²⁴, in which case allosteric effects of perturbations can be calculated analytically. However, these approaches are unable to capture the anharmonicity and multi-well nature of flexible degrees of freedom in proteins. Another approach is to infer groups of residues important for a given allosteric process by analyzing structures trapped in different conformations²⁵⁻²⁷.

Dynamical approaches to studying allostery generate an ensemble of structures and then analyze the ensemble using cross-correlations²⁸, contact correlations²⁹, principal components²⁹, or local unfolding correlations³⁰. One widely adopted approach uses a quasi-harmonic metric for correlations that assumes an "average" structure^{28, 29, 31-33}. This approximation may be appropriate for small backbone fluctuations but may not aptly describe conformational changes that involve basin-hopping, such as loop or side-chain motions. To overcome this quasi-harmonic limitation, Lange and Grubmuller^{34, 35} used a mutual information method to account for both quasi-harmonic and anharmonic correlations in atoms' motion in Cartesian space. Still other methods introduce mechanical perturbations and monitor the subsequent motions of residues^{36, 37}. The latter approach can detect substantial population shifts or structural changes following the induced local perturbations, as the added energy facilitates barrier crossing.

Our MutInf approach for identifying allosteric networks quantifies correlations between the conformations of residues in different sites. We use an entropy-based approach to analyze ensembles of protein conformers, such as those from molecular dynamics or Monte Carlo simulations. The method is applicable even in cases where conformational changes are subtle, e.g., when the coupling is mostly entropic in nature^{38, 39}. Unlike the approaches described above, our approach uses internal coordinates and focuses on dihedral angles, which are responsible for most low-frequency motions, in order to capture correlated changes in side chain rotamers, a highly anharmonic type of correlated motion. The most closely related

previously published method is a study that examined side-chain correlations using a mutual information metric and Monte Carlo simulations of side-chains⁴⁰ on a set of fixed protein backbones.

Our MutInf method builds upon and extends previous work by 1) directly connecting correlated conformations to the molecular configurational entropy, 2) incorporating more robust entropy estimators, 3) correcting for undersampling using data from multiple simulations, 4) testing statistical significance to filter out correlated motions that are not significant, and 5) analyzing both backbone and sidechain torsions, which are frequently coupled^{41, 42}. The theoretical underpinnings of our approach are described in detail in Methods. Briefly, we use second-order terms from the configurational entropy expansion, the mutual information⁴³, to identify pairs of residues with correlated conformations in an equilibrium ensemble. In calculating mutual information, it does not matter whether two residues move at the same time or whether one moves, and then the other; what counts is whether these residues' conformational distributions are correlated. In this work we use the terms correlated motions and correlated conformations interchangeably.

Because we look for correlated conformations in an unbiased, statistically robust manner, we believe that MutInf will be a useful tool in the discovery of novel, “orphan” allosteric sites, where endogenous protein or small molecule allosteric modulators have yet to be discovered. As a proof-of-principle, we used our approach to identify correlations between the conformations of protein residues lining two small-molecule binding sites in human interleukin-2 (IL-2). This single-domain protein exhibits cooperative ligand binding without substantial conformational change, and to date no follow-up work has been done to uncover the mechanism for this cooperativity. We discuss the rationale behind our approach and compare its strengths and weaknesses to those of other methods and then discuss the mathematical details of our method and our novel results on IL-2.

Methods

Theoretical Underpinnings of the Model

When an equilibrium ensemble of states is altered by small perturbations, the fluctuation–dissipation theorem relates equilibrium fluctuations to the system's response, which will be proportional to equilibrium pair-correlations of the degrees of freedom and linear in the applied perturbations. This linear response theory suggests that external forces, such as those due to ligand binding, cause the largest indirect changes in the degrees of freedom that are most correlated (at equilibrium) with those directly perturbed by the external forces. As has been previously noted²⁹, this also means that the response to small perturbations involves the same fluctuation pathways activated by random solvent collisions at equilibrium. Elastic network models have identified a correspondence between low-frequency normal modes and pathways used in several protein conformational changes^{22, 44}, suggesting that correlations observed in equilibrium simulations may propagate perturbations in structure and/or dynamics due to ligand or protein binding.

A perturbation at one site can couple to another site directly, through electrostatic or steric interactions, or indirectly, through solvent reordering, or through a network of residues with correlated conformations. When the conformation at one site depends on the conformation at another site, the sites' conformations are correlated. When the conformations are correlated, perturbations at one site can cause population shifts in conformations at other sites. Correlated conformations are then signals that can be used to identify allosteric mechanisms and predict new sites for allosteric inhibition by proteins or small molecules.

Our MutInf approach uses equilibrium molecular dynamics simulations to identify correlations in residues' conformations, from which functional coupling between sites is inferred.

Approaches such as ours that infer allostery from equilibrium simulations assume that the allosteric phenomena of interest (i.e. ligand binding, protein binding, protonation state changes, etc.) make perturbations to the energy landscape that are relatively small, i.e. at most a few kT. For example, proteins and ligand binders with fast on-rates will satisfy this assumption, while proteins and ligand binders with slow on-rates may not. Furthermore, equilibrium approaches that infer allostery assume that there are no large barriers to conformational changes required for allostery. If such barriers existed, they would prevent pairs of residues from sampling relevant correlated shifts in conformation when perturbations of interest are applied. Along these lines, these equilibrium approaches also assume that there is sufficient sampling along the degrees of freedom relevant to the allosteric phenomena, so that productive or “on-pathway” correlated motions can be observed.

To quantify correlations between residues' conformations from equilibrium simulations, we take advantage of a connection between information theory and thermodynamics. Inspired by the use of mutual information by Killian, Kravitz, and Gilson in calculating configurational entropies from conformational ensembles using internal coordinates⁴³, we use second-order terms from the configurational entropy expansion, i.e. the mutual information, to identify pairs of residues with correlated conformations. This approach directly and quantitatively connects correlations in conformation to configurational entropy. Using internal coordinates to calculate the mutual information has the twofold advantage of 1) capturing the rotameric, flipping, and gear-like nature of correlated side-chains; and 2) removing potentially spurious correlations that can arise due to structural alignment. The latter effect occurs because minimization of the r.m.s. error in aligning structures in Cartesian space can yield correlated displacements in many atoms' positions as some atoms are fit better than others. An overview of our approach is presented in Scheme 1.

In applying entropy and mutual information to studying allostery, we sought to obtain a measure of the statistical significance of our results and filter out noisy and artifactual correlations. To accomplish this, we extended established methods for calculating entropies and examining correlations via mutual information to handle finite sample sizes, to incorporate data from multiple simulations, to account for the variability between simulations, and to correct for the fact that multiple simulations do not, in practice, represent independent samples of the macromolecular ensemble.

Calculation of Mutual Information

The configurational space of a molecule can be described in a standard Cartesian coordinate system or in an internal coordinate system of bond lengths, bond angles, and torsion angles (BAT)⁴³. For proteins, key torsion angles include the ϕ , ψ , and ω torsion angles of the protein backbone and the χ torsion angles of the amino acid side-chains. In the present work, we consider only the ϕ , ψ , and heavy-atom χ torsion angles (only the first χ angle for proline) and neglect changes in bond lengths, bond angles and omega backbone torsion angles, as we believe that the dynamics of the first three are the most relevant to describing motions of biological importance⁴³.

Small sample sizes are notoriously challenging for entropy and mutual information-based approaches, so we use robust estimators and correct for bias using simulated data.

Configurational Entropy Expansion and Correlations Between Degrees of Freedom—We wish to quantify correlations between residues' torsions. Following the works of Matsuda⁴⁵ and of Killian, Kravitz, and Gilson⁴³, we connect correlated torsions to

thermodynamics using an expansion of the molecular configurational entropy into terms over single torsions, pairs of torsions, etc. The total torsional entropy is given by:

$$S_{conf} = \sum_i^n \int_0^{2\pi} p(\varphi) \ln p(\varphi) d\varphi - \sum_i^n \sum_j^n \int_0^{2\pi} \int_0^{2\pi} p(\varphi_1, \varphi_2) \ln \frac{p(\varphi_1, \varphi_2)}{p(\varphi_1)p(\varphi_2)} d\varphi_1 d\varphi_2 + \dots - \dots \quad (1)$$

where indices i and j are residues' torsions, and n is the number of torsions (φ , ψ , and χ torsion angles in the present work). The second-order term here represents a sum of the mutual information of each pair of torsions. The mutual information describes correlations between degrees of freedom, and gives a measure of how much information about one degree of freedom is gained by knowledge about another⁴⁶. Because the mutual information values are terms in the entropy, which is related to free energy, the mutual information in Eq. 1 is in units of kT . The mutual information has been a popular, distribution-free analysis method, and more recently has been used in the context of molecular conformational ensembles^{40, 43}.

As an example, consider the distributions of the χ_1 torsion angles for two side chains. For concreteness, we use an example from interleukin-2, which is described in greater detail below and involves two aromatic residues in close proximity (Fig. 1). The expected joint distribution of these torsion angles under the null hypothesis (Fig. 1A) of independence is merely the outer product of the marginal distributions. However, the joint distribution from the observed simulations (Fig. 1B) show that these torsion angles are correlated ($I = 0.203 kT$), because a cross-peak (indicated by a gray box) appears in the simulations that would not be expected if these torsions were independent.

In practice, we compute the mutual information, I , between two degrees of freedom as the difference between the self-entropies and the joint entropy, using the relation, $I = S(1)+S(2) - S(1,2)$, and a corrected histogram entropy estimate⁴⁷ over adaptive partitions⁴⁶:

$$I = \sum_{i=1}^r \frac{n_i}{N} (\ln N - \Psi(n_i) - \frac{(-1)^{n_i}}{n_i+1}) + \sum_{j=1}^s \frac{n_j}{N} (\ln N - \Psi(n_j) - \frac{(-1)^{n_j}}{n_j+1}) - \sum_{j=1}^{rs} \frac{n_{ij}}{N} (\ln N - \Psi(n_{ij}) - \frac{(-1)^{n_{ij}}}{n_{ij}+1}) \quad (2)$$

where r and s are the number of marginal bins, n_i , n_j , and n_{ij} are the histogram counts, N is the number of data points, and Ψ is the digamma function. Adaptive partitions make efficient use of discrete bins, preserve correlations between variables, and normalize each joint distribution to a reference distribution in which marginal counts are as uniform as discretization allows⁴⁶. In this work we used 24 bins per dimension. Furthermore, adaptive partitions enable accurate mutual information values to be calculated whether torsional motions are large or small. Note also that we account for the two-fold symmetry in the χ_2 angle of Asp, Phe, and Tyr and in the χ_3 angle of Glu.

The histogram entropy estimator above assumes that histograms are populated by a Poisson process ($n_{ij} \ll N$) and so is especially appropriate for sparse joint histograms. It also implicitly includes finite bin and data size corrections used in other discrete entropy estimators⁴⁸. As a statistic for examining correlations between variables, the mutual information (with corrections

discussed below) is far more robust against small sample sizes than the χ^2 statistic, which assumes $n_{ij} \geq 5$. While the nearest-neighbor approach⁴⁹ could have been used instead to compute these integrals, it can require $N \approx 50,000$ data points⁵⁰ or more to yield a converged estimate of the mutual information for pairs of torsions. Nearest-neighbor approaches are accurate for very large datasets but have biases under finite sample sizes that depend on the topology of conformational space sampled in simulations⁵⁰. As our goal was to extend mutual information calculations to handle smaller sample sizes, we chose instead to use adaptive partitioning in combination with the corrected histogram mutual information estimate above (Eq.2), so that each pair of degrees of freedom could be compared against the same empirically-generated reference distribution and evaluated for significance.

Correction for Nonzero Mutual Information in Independent Datasets—In a number of applications using mutual information, it has been found that samples of two variables that are independent can yield nonzero mutual information in calculations^{46, 51-53}. We empirically observe the same in simulated data (data not shown), and this is not surprising because errors in estimates of the true mutual information are a consequence of finite samples. To correct for this, one approach is to create P permutations of the original data, so that the marginal probability distributions remain the same while correlations between the data are scrambled. One can use these permutations to establish a test of significance of the observed mutual information with a null hypothesis of independence versus a one-sided alternative. The approximate p-value is then the percentage of mutual information values from different permutations that are greater than the observed mutual information from the original data^{46, 53}. Also, the average mutual information for the permuted data, the “independent information”, can be subtracted from the observed mutual information to yield the “excess mutual information”, a more reliable estimate of the true mutual information^{51, 52}.

When adaptive partitioning is not used (and hence the marginal densities are not normalized), permutation approaches are inefficient in sampling the distribution of the mutual information under the null hypothesis, because permuted values are likely to fall into bins overrepresented in the marginal densities; adaptive partitioning fixes this inefficiency by normalizing marginal densities without altering correlations between variables. One can apply the permutation approach above to nearest-neighbor estimates as well, as these also will have bias due to finite sample sizes. For example, a combined K-fold resampling/permutation test was found to be useful in conjunction with nearest-neighbor mutual information estimates to discriminate between relevant and independent features in feature selection⁵³. A major drawback to the permutation approach is that it is computationally demanding in processing time and in memory. Moreover, permutations introduce random error because not all $N!$ permutations can be made⁵³.

Instead, since adaptive partitioning is used in this work, we noted that the same distribution of the “independent information” is appropriate for all pairs of degrees of freedom. The distribution of the mutual information for independent variables for given data size N and number of marginal bins r and s has not yet been analytically solved, though in some cases can be empirically fit⁵². However, because an analytical, parametric approach is not available, we perform Monte Carlo sampling to obtain the reference distribution of the “independent information” for all pairs of torsions. With adaptive partitioning, the marginal counts are nearly uniform and in any case are equivalent for different pairs of torsions. Thus, all pairs of torsions will have the same distribution in histogram bin space under the null hypothesis of independence. To construct the reference distribution for a pair of independent torsions, we first make a copy of the marginal distributions for a given pair of torsions (it doesn't matter which pair we choose). Then, we choose ordered pairs of bin indices at random from these marginal distributions and place them into a 2-D histogram without replacement. The mutual information is calculated according to Eq. 2 above, and this procedure is repeated 1000 times

to create a distribution of the mutual information under the null hypothesis of independence for the given number of datapoints N and number of bins r .

We use this distribution of “independent information” for a significance test of observed mutual information values, and we subtract the average “independent information” from the observed mutual information to yield the “excess” mutual information; this filters out insignificant mutual information values and corrects for finite sample size bias. Because this analysis empirically generates a distribution under the null hypothesis, the false positive rate for keeping a nonzero mutual information value for torsions that are truly independent is α , the significance level (in our case, 0.01). This false positive rate will be further reduced by consideration of the alternative hypothesis.

Bayesian Filter to Remove False Positives—Most approaches that filter mutual information values using tests of statistical significance do so according to whether the null hypothesis of independence can be rejected using descriptive statistics. One disadvantage of these approaches is that they do not consider the distribution of the mutual information under the alternative hypothesis. In Bayesian statistics, the mutual information is a random variable with a distribution, and the probability that the mutual information is greater than a given value can be calculated. Approximations to the distribution of the mutual information have been described that account for uncertainties in the estimates of the probability density functions⁵⁴. The first two central moments of the distribution, the expectation $E[I]$ and variance $\text{Var}[I]$ of the true information given the data and prior, are given as follows:

$$E[I] = \frac{1}{N} \sum_{i=1}^r \sum_{j=1}^s n_{ij} (\Psi(n_{ij}+1) - \Psi(n_i+1) - \Psi(n_j+1) + \Psi(N+1)) \quad (3)$$

$$\text{Var}[I] = \left| \frac{K-J^2}{N+1} + \frac{M+(r-1)(s-1)(\frac{1}{2}-J)-Q}{(n+1)(n+2)} \right| + O(n^{-3}),$$

$$J = \sum_{ij} \frac{n_{ij}}{N} \ln \frac{n_{ij}N}{n_i n_j}, K = \sum_{ij} \frac{n_{ij}}{N} \left(\ln \frac{n_{ij}N}{n_i n_j} \right)^2, M = \sum_{ij} \left(\frac{1}{n_{ij}} - \frac{1}{n_i} - \frac{1}{n_j} + \frac{1}{N} \right) n_{ij} \ln \frac{n_{ij}N}{n_i n_j}, Q = 1 - \sum_{ij} \frac{n_{ij}^2}{n_i n_j}, \quad (4)$$

where $n_{ij} = n_{ij}(\text{observed}) + n_{ij}(\text{virtual})$, and the virtual counts come from a noninformative Dirichlet prior ($n_{ij}=1$ for the uniform prior, which was used in this work). Approximations for the leading order terms for the third and fourth central moments have been reported⁵⁴, and could be used in an Edgeworth expansion to approximate the distribution, but for robustness we chose instead to simply use a Gaussian with the above mean and variance, which fit reasonably well to simulated data in a model system⁵⁴. We then use this approximate distribution of the mutual information to calculate $P(I < E[I_{\text{ind}}])$, the probability that the true mutual information is below that expected for independent torsions (calculated using Eq. 3 averaged over 1000 simulated independent datasets). Pairs of torsions with $P(I < E[I_{\text{ind}}]) > \alpha$ are not significant and are discarded.

Corrections to the Mutual Information Accounting for Incomplete Sampling—To obtain accurate entropies and mutual information values up to second order, simulations must be run many times longer than the slowest autocorrelation and pair correlation times, and data points should represent independent observations. Due to limited computing power, this is rarely the case, and molecules in simulations carry with them some memory of their initial states. For example, consider a salt bridge. Salt bridges can form strong electrostatic interactions, and hence it can take a long time to sample their full conformational space (long

autocorrelation time) and even longer to sample all populated pairwise conformations (long pair correlation times). Thus, a salt bridge may retain some memory of its initial conformation, which will fade away on the timescale of the pair decorrelation time (approximately). In practice, we decided to use data from multiple simulations to penalize this kind of undersampling in a novel way.

First, we first aggregate the counts for two degrees of freedom from a set of simulations (sample ensembles) of size n_{sims} , and calculate the mutual information for all the simulations taken together. Intuitively, two torsions in different simulations should not be correlated, as they should sample their probability distributions independently. Any non-zero (excess) mutual information between these torsions is a measure of conformational undersampling bias that we can subtract from the mutual information between the torsions for the set of simulations. To correct the mutual information for artifactual correlations due to incomplete sampling, we calculate the excess mutual information and then subtract the average excess mutual information between two degrees of freedom in different pairs of simulations (when it is positive):

$$I_{i,j}^{sims} = I_{i,j}^{sims} - \langle I^{ind}(N, r) \rangle - \left(\frac{n_{sims}}{2} \right)^{-1} \sum_{k=1}^{n_{sims}} \sum_{l \neq k}^{n_{sims}} I_{i,j}^{k,l} - \langle I^{ind}(N', r') \rangle \quad (5)$$

Here i and j correspond to the different torsions, l and k are the indices of the pairs of different simulations, and I^{ind} is calculated for a pair of simulations just as the independent information is calculated for a set of simulations using the Monte Carlo recipe above, except that values of $\langle I^{ind} \rangle$ lower than the standard deviation of I^{ind} are zeroed to reduce noise from this term. For the mutual information between torsions in different simulations, we use half as many bins ($r' = r/2$), because the number of datapoints N' for the histograms is smaller than the total number of datapoints from all simulations, N ($N' = N/n_{sims}$). Significant mutual information values are those that have passed the significance test vs. the null hypothesis, the Bayesian filtering using the alternative hypothesis, and whose corrected excess mutual information (eq. 5) is greater than zero.

When we consider the mutual information between pairs of residues, we take the sum of the mutual information between pairs of residues' torsions:

$$I_{residues}(i, j) = \sum_{\substack{k=\varphi, \phi, \chi' s \\ \text{residue}(i \neq j)}} \sum_{\substack{l=\varphi, \phi, \chi' s \\ \text{residue}(j)}} I_{k,l}^{sims} \quad (6)$$

This may overestimate the total mutual information between two residues, as we neglect the higher-order terms in Eq. 1. Inclusion of statistically significant higher order terms (which would require more data points) would further increase the accuracy of the calculated mutual information between pairs of residues. Nonetheless, our results below show that our robust use of second-order terms is a powerful means to identify residues with correlated conformations.

Molecular Dynamics Simulations

Molecular dynamics simulations on interleukin-2 (IL-2), alone and in complex with three ligands, were performed using GROMACS 3.355, 56 and the AMBER-99 ϕ forcefield57. Loops missing atomic coordinates, such as residues 1-5, 75-76, and 99-102 in apo IL-2, were

closed using loop prediction via the Protein Local Optimization Program (PLOP58). Protonation states of histidine side-chains at pH 7 were given by MCCE59, 60: we modelled His16 as positively charged (residue name “HIP”) and His55 and His79 as ϵ -protonated. Two ligand-bound forms of IL-2 were prepared, with either Ro26-4550 (amino(3-(2-(1-methoxy-1-oxo-3-(4-(phenylethynyl)phenyl)propan-2-ylamino)-2-oxoethyl)piperidin-1-yl)methaniminium)^{61, 62} bound to the competitive IL-2R α site (PDB:1M48), or compound 7c (1-(3,4-dihydro-1H-pyrido[3,4-b]indol-2(9H)-yl)-2-methoxyethanone) bound to the allosteric site⁶³. Compound 7c was built from PDB:1NBP by modifying the covalent compound 7t in the crystal structure to the noncovalent compound 7c in Maestro (Schrodinger, 2007), then using PLOP loop prediction to optimize the loop from residue 29 to 33, and to fill in missing residues between residues 73 to 78, in each case simultaneously optimizing side-chains within 12 Å of the given loop region. These ligands were parametrized for MD by GAFF⁶⁴ and assigned AM1-BCC charges^{65, 66}. Each apo protein or complex was energy minimized in explicit solvent using GROMACS, and five copies of each of the three constructs (apo, competitive site bound, and allosteric site bound) were equilibrated at 300K (with different random seeds) using constant volume for 10 ps and using a constant pressure of 1 atm for 100 ps using the Berendsen barostat⁶⁷, with hydrogens constrained using the Lincs algorithm⁶⁸. Equilibration of each simulation was followed by a 10 ns production run, with snapshots of the atomic coordinates recorded every 1 ps. Actual lengths of individual simulations ranged between 9.6ns and 11ns for technical reasons. RMSDs of the two compound binding sites over the simulations showed that all of the five simulations per apo or small-molecule bound construct were stable (Supplemental Figure 1).

Ensemble Docking

We clustered MD snapshots from the IL-2 simulations with allosteric compound bound according to the coordinates of residues in the competitive site using QT clustering⁶⁹ as provided in GROMACS (“g_cluster -method gromos”). Then, to each cluster representative we docked the Roche competitive site ligand Ro26-4550⁶¹, which binds cooperatively with the allosteric ligand. This ensemble docking was performed using the XGlide cross-docking script (Schrodinger, 2007, Script Center XGlide v. 1.1.2.6, mmshare v. 16109, using inner and outer grid box lengths of 10Å and 18Å, respectively)⁷⁰.

Results and Discussion

We applied our MutInf approach to elucidate the mechanism of small molecule binding cooperativity in human interleukin-2. Little is known about how binding of ligand at the IL-2R α binding site enables binding of a small molecule fragment to a cryptic allosteric site. These ligands bind at least 6.9 Å apart at their closest approach in the predicted ternary complex. Crystal structures of complexes of interleukin-2 with small molecules bound to different sites did not show substantial structural changes at the other sites (maximum C α RMSD 0.88 Å at the allosteric site for apo PDB:1M47 and competitive-bound PDB:1M48). We therefore hypothesized that allostery and cooperativity in IL-2 arises largely from changes in dynamics and subtle population shifts rather than a major change in the preferred backbone conformation^{28, 38, 39}.

We used molecular dynamics to study correlated motions at the atomic level on a picosecond-nanosecond timescale, and used our MutInf approach to analyze sets of 10 nanosecond trajectories of human interleukin-2, alone and in complex with different small molecule binding partners. Our goal was to show that MutInf can identify significant correlated conformations for functionally-important residues in simulations whose lengths and recording frequencies are typical of those in the current literature.

Mutual Information From Molecular Dynamics Identifies Significant Long-Range Correlations

We first analyzed whether an unbiased, whole-protein analysis of correlations between residues in interleukin-2 would be able to identify cooperative sites and the correlations between them from the apo simulations alone. For each pair of residues, we calculate the mutual information as per (Eq. 6) between all pairs of ϕ , ψ , and χ torsion angles for our apo simulations of interleukin-2. The mutual information is reported in units of kT, because of the relationship between mutual information and entropy (Eq. 1).

When we plotted the statistically significant mutual information between pairs of residues' torsions in IL-2, we found that only a small subset of residue pairs are highly correlated, while many are only marginally correlated (Fig. 2A). A substantial part of the present work involved incorporating more robust entropy estimators for calculating the mutual information and filtering out insignificant correlations with the help of empirical or approximated distributions under the null and alternative hypotheses. So, as a control, we plotted the unfiltered mutual information between residues' torsions in Fig. 2B. Protocols with and without statistical filtering showed correlation between residues in the loops after helix 1 and between helices 2 and 3 (Fig. 2A and 2B, red boxes on the diagonal and off the diagonal, respectively, and Fig. 2C, red residues). Our statistical filtering, however, highlights these and removes background noise; only a subset of the pairwise correlations between residues in these regions make statistically significant contributions to the conformational entropy. Moreover, our MutInf approach identified significant correlations between residues in the loop region between helices 3 and 4 and residues in other regions, in particular residues in the floppy N-terminal tail (Ser6, Thr7) and the beginning of the N-terminal helix (11-15), residues in the loop between helices 2 and 3, and residues in the C-terminal helix (Fig. 2A, blue boxes, and Fig. 2C, blue residues). The loop between helices 2 and 3 displays significant variability in the different crystal structures of IL-2, and is at least partially disordered in most structures, indicating both that it is flexible and that it can adopt at least several conformations. Residues showing significant correlations near the C-terminus include two residues in the loop before the C-helix (Glu100 and Thr101), and residues along the C-helix (Arg120, Ile128, and Leu132, proximal to IL-2's negatively-charged C-terminus).

We compared our MutInf method to previously-reported methods for identifying correlated motions, in particular the Gaussian Network Model (GNM) approach of Bahar and colleagues⁷¹ and the Cartesian mutual information method of Lange and Grubmüller³⁴. Both our method and the GNM method suggested correlation between residues in the loop after helix 1 and residues in the loop between helices 2 and 3 (red boxes on the diagonal and off the diagonal, respectively, Fig. 2A, 3A). We found that our approach highlighted strong correlations and gave low background noise, while the $C\alpha$ cross-correlation matrix using a GNM (with 10 Å $C\alpha$ - $C\alpha$ cutoff) gave a noisier pattern of correlations, as did Lange and Grubmüller's mutual information method applied to the Cartesian coordinates of $C\alpha$ atoms (Fig. 3B).

Our identification of significant long-range correlations led us to investigate the distribution of correlations between residue pairs as a function of distance between the residues' $C\alpha$ atoms (Fig. 4). As would be expected, the number of weak correlations decreases as the distance between residues increases. However, residues separated by substantial distances have correlations of 0.4 kT or more just as often as residues separated by short distances.

Looking more closely, we see that there are strong couplings between pairs of distant residues (Fig. 5). Here, we highlight correlations of magnitude greater than kT between distant residues, namely those with alpha carbon separations of more than 5 Å. Again, we observe strong couplings between the N-terminus of helix A, Tyr31 at the C-terminal end of helix A, and the

adaptive loop region (74, 76-78), as well as between Gln74 and Glu100, and between Glu100 and Ile128 near IL-2's C-terminus. It is not surprising that many of these residues are polar, as molecular dynamics simulations include terms for long-ranged electrostatic and ion-dipole interactions. Gaussian Network Models, on the other hand, do not model such sequence-dependent, long-range interactions, and so it is not surprising that the correlations between distant residues are typically weak. Electrostatic interactions can be both directly and indirectly responsible for long-range correlations in residues' conformations; directly, through Coulomb's law, and indirectly, through a dynamic network of charged and hydrogen-bonding polar residues, and through altering the first-shell water structure around the protein (in simulations with explicit solvent). Unlike charge-dipole or dipole-dipole interactions, where the effective range decreases through averaging over orientations⁷², charge-charge interactions retain their long-range nature even when averaged over orientations.

Other factors that can give rise to correlated conformations include hydrophobic packing and rigid-body motions of semi-rigid secondary structure elements, such as α -helices. We note that our approach does not typically show strong correlations within semi-rigid elements such as α -helices or the central hydrophobic core of the four-helix bundle. Two possible reasons for this are because mutual information values are not normalized quantities and because higher-order correlations are not captured. As the maximum mutual information between two residues is the minimum of their self-entropies (flexibilities), residues that have higher self-entropies (more flexible) can exhibit a greater magnitude of coupling with other residues. This behavior is thermodynamically appropriate because the un-normalized mutual information is related to the configurational entropy (Eq. 1), and not a normalized quantity. Furthermore, this behavior is consistent with thermodynamic considerations in intrinsically disordered proteins, where disorder in one or both domains serves to optimize allosteric coupling between the sites⁷³. Such allosteric coupling does not require a network of interactions linking the sites. In the present study on interleukin-2, flexible residues at either end of a helix showed couplings $> kT$ in some cases while intervening helical residues did not (Fig. 5), because mutual information values, unlike correlation coefficients, are not normalized for residues' self-fluctuations. We note that even normalized pairwise correlations (i.e. Fig. 3A) do not include the higher-order correlations that would be expected within semi-rigid elements, and while the Gaussian Normal Mode results (with default cutoffs) show weak couplings between the N-terminus of helix A and Tyr31 at the C-terminal end of helix A, and between the adaptive loop region (74, 76-78) and Glu100, these are not visibly distinct features (Fig. 3A). In any case, as the goal of our approach is to identify correlations between the conformations of functional sites on a protein, it is not necessary to identify all of the residues that indirectly mediate such correlations, though this is an area for future work. We will later focus on coupling between two small molecule binding sites in interleukin-2, which are physically connected by flexible loops and sidechains, rather than semi-rigid secondary structure elements.

Hierarchical Clustering Identifies Dynamic “Hot-Spots” In Interleukin-2

For a site to be suitable for allosteric inhibitor design, it must be both (1) allosteric, causing shape or flexibility changes at other sites, and (2) druggable, having the right shape and hydrophobicity for drug-like small molecules⁷⁴. Here, our goal was to predict which sites were most likely to alter structure or dynamics at known functional sites upon perturbation (by ligand binding, for example). To search for such sites, we wanted to identify groups of residues responsible for correlations between functional sites.

In biological networks such as protein-protein interaction networks, functional connections between various proteins are preferentially mediated by “hubs” that interact with a greater than average number of partners⁷⁵. Similarly, functional connections between various protein sites are thought to be mediated by “hub” residues or clusters of residues^{25, 27}. We hypothesized

that clusters of residues correlated to many other residues, i.e. “dynamical hotspots”, could be potential sites or mediators for allosteric modulation of other sites. To find such “dynamical hotspots”, we performed a hierarchical clustering of the matrix of mutual information values between residues, in analogy to the analysis of microarray data, using the “heatmap” function in R (<http://www.r-project.org/>). We used a Euclidean distance metric so that residues showing similar patterns of correlations with other residues are clustered together. Interestingly, one cluster of residues emerged with the strongest correlations within cluster members and the strongest correlations to other residues, and was previously found to be an adaptive region that could bind a number of small-molecule fragments as measured by Tethering experiments⁶². Residues in this cluster are colored red in Fig. 6 and mostly reside in the flexible loop between helices 2 and 3, with two in the N-terminal floppy tail and one in the flexible C-terminal loop. Because the mutual information between two torsions is less than either of their self-entropies, it is not surprising that flexible residues often have high mutual information with other residues. This red cluster constitutes a “dynamic hotspot”, as it is highly correlated to other clusters of residues. Furthermore, as this red cluster is correlated to the blue cluster containing the IL-2R α inhibitor binding site, our method predicts the red cluster to be a candidate region for allosteric modulation of the IL-2R α site. Two similar clusters can also be seen when mutual information values from subsets of the five simulations are block-averaged (Supplemental Figure 2). However, our approach does not yet predict whether such a site would be druggable by small-molecule allosteric modulators or contain “hotspots” of affinity for protein-protein interactions.

Chemical Shift Perturbations Upon Binding Corroborate Predicted Correlated Motions

While direct experimental methods to identify correlated motions by NMR are limited, one can use chemical shift perturbations and changes in side-chain order parameters of residues outside a binding pocket to identify population shifts in residues within or proximal to allosteric sites that accompany ligand binding. A study by scientists at Roche identified IL-2 residues showing backbone and side-chain amide chemical shift perturbations upon binding IL-2R α receptor or IL-2R α -competitive small molecule⁶¹. For example, Tyr31, Gln74 and Ser75 (in the strong cluster colored “red” in Fig. 6) show strong ¹⁵N/¹H chemical shift perturbations following competitive ligand binding, though these residues are not in the competitive binding site. While ring current effects from the ligand’s biphenyl group could contribute to these shift perturbations, they are qualitatively consistent with our prediction that these residues’ conformations are correlated to the conformation of the IL-2R α binding site. Furthermore, several residues or regions distal from the IL-2/IL-2R α interface identified by our approach as highly correlated to the “blue” cluster (encompassing many residues in the IL-2/IL-2R α interface, PDB: 1Z9276) showed substantial chemical shift perturbations upon IL-2R α binding (Fig. 7). Unfortunately, resonance overlap restricted the analysis of chemical shift perturbations, notably in most of the flexible loop in the red cluster, so we cannot test our prediction that one would see many perturbations in the flexible loop. It is important to note that none of the nine residues showing insignificant chemical shift perturbations (Asn26, Thr37, Met104, Cys105, Tyr107, Thr113, Ile122)⁶¹ appeared in the red or blue highly correlated clusters.

Communication Between Cooperative Compound Binding Sites Involves a Polar, Solvent-Exposed Network and a Greasy Core

In the previous sections we used a global description of pairwise couplings to identify putative allosteric sites from correlations between clusters of residues. Presently, we apply our method to study the mechanism of coupling between two given allosteric sites. From our matrix of pairwise correlations between residues in apo-IL2 and representative structures from the conformational ensemble, we could infer a structural mechanism by which the two small molecule binding sites might be coupled. In Fig. 8 we show matrix elements corresponding to

residues in the IL-2R α -competitive site and residues in the allosteric site, along with representative conformations of these residues. These representative conformers were picked by clustering the MD snapshots according to the RMSD of residues in the “red” cluster that belonged to the highly flexible loop or were proximal to the N-terminus.

Thermodynamically, the two sites are coupled directly by the off-diagonal gray matrix elements in Fig. 8 in the box denoting “Correlations Between Sites”. These two sites may also be coupled by higher-order terms involving other residues, which our pairwise analysis does not address (save for the hierarchical clustering which uses patterns of correlation rather than the correlations themselves). From the representative conformations for these residues in Fig. 8, the two sites appear to be coupled via a polar network on the protein surface and a greasy core. Two residues are common to both binding sites, namely Lys35 and Met39. The side-chain of Met39, for example, can directly interact with both of the cooperative small molecules, and will be discussed in more detail later. A number of polar side-chains pointing toward the solvent form a network of hydrogen-bonding and electrostatic interactions. In particular, Gln74 (dark green lines) samples a wide swath of conformations, sometimes hydrogen bonding with basic residues in the competitive site (Lys35, dark blue, and Arg38, light blue), and other times hydrogen bonding with basic Arg81 (gray) near the allosteric site. Also, correlations between residues in the greasy core connect hydrophobic surfaces of both compound binding sites; when bound, the allosteric or competitive small molecule would be contiguous with this hydrophobic network. Notably, the matrix elements showing “Correlations Between Sites” indicate that the conformations of residues in the polar network and greasy core are coupled. Thus, a more accurate mechanism for the coupling would be that the two sites are coupled via a polar network, a greasy core, and crosstalk between these elements.

Ligand Binding At Allosteric Site Causes Rotamer Population Shifts That Promote Competitive Site Inhibitor Binding

The preceding analysis suggests that the two sites are connected via correlated motions, and that this could explain the observed allostery. To directly test our hypothesis that the experimentally observed cooperativity between the sites involves subtle population shifts in residues exhibiting correlated motions, we performed additional simulations with a competitive or an allosteric inhibitor bound, and asked whether simulations with the allosteric inhibitor bound would cause population shifts in the competitive site similar to those observed in simulations with the competitive inhibitor. Comparing the crystal structures of apo IL-2 (PDB: 1M47) to competitive-site-inhibited IL-2 (PDB: 1M48), we note that the motion of two side chains, Met39 and Phe42, opened up a binding groove for the ligand that was closed in the apo structure.

We then asked whether population shifts caused by allosteric ligand binding would help open up the competitive site for competitive ligand binding. To address this question, we examined side-chain torsion angle distributions for Phe42 and Met39 in simulations with compound at either site and compared these to distributions from the apo simulations (using histogram bins of 6 degrees and 10 ps time intervals). The populations of side chain dihedrals angles in Phe42 and Met39 show substantial differences between the apo protein and the protein bound with competitive and allosteric inhibitors (Fig. 9, Fig. 10). Interestingly, the populations observed for the allosteric compound-bound protein are intermediate between apo IL-2 and competitive-site-bound IL-2. Phe42, a hot-spot residue critical for ligand binding and protein binding, adopts a different χ_1 rotamer for ligand binding than it does for protein binding, which is more similar to the apo rotamer (Fig. 9). The χ_1 rotamer selected by ligand in competitive-site-bound simulations (100% population) was more populated (89%) in allosteric-bound simulations than in apo simulations (39%), showing a population shift caused by allosteric compound binding.

We predict that Met39 is an important mediator of binding cooperativity because its conformation is correlated to that of Phe42 and because it shows χ_1 and χ_2 population shifts upon allosteric compound binding in the same direction as population shifts from the apo to competitive inhibitor-bound distributions. In crystal structures, Met39 adopts similar conformations in complexes with competitive inhibitor or allosteric inhibitor that both differ from the conformation in the apo structure. Met39 is in an “up” conformation in the apo structure, packed against hot-spot residue Phe42. In competitive inhibitor-bound structures, the side-chain of this Met moves down to slightly enlarge the pocket for a ligand aromatic ring, while in the allosteric-bound structure, the Met sidechain moves down to interact weakly with the ligand and fill in part of the hydrophobic pocket opened to accommodate the ligand (Fig. 10). Interestingly, this Met39 is not critical for a high-affinity competitive ligand to bind at the competitive site⁷⁸, presumably because mutating it to alanine would simply make that hydrophobic pocket a little larger. However, it is currently unknown whether this residue is required for allosteric ligand binding or for the binding cooperativity, as we predict. Our calculations indicate that cross-talk contributing to cooperativity involves not only the greasy core (of which Met39 and Phe42 are a part), but also the loop between helices 3 and 4 and a polar network involving a number of basic residues on the protein surface. This hypotheses could be further tested by conservative mutations of residues such as Gln74, which is not part of either ligand's binding site; Lys35, whose alkyl tail but not polar head contacts compound 1 in the crystal structure; or Met39, whose alanine mutation only shows a slight effect on competitive ligand binding⁷⁹

Conformational Selection *In Silico* by Allosteric Ligand

The preceding analysis suggests that binding at the allosteric site can positively modulate binding at the competitive site through changing the distribution of side chain rotamers. To more directly assess the relationship between these relatively subtle changes and ligand binding, we have performed small-molecule docking against snapshots from the MD simulations (Fig. 11). Although the scores from docking to MD snapshots cannot be interpreted as accurate binding affinities, we use them as way of qualitatively assessing whether the conformation of the site is appropriate for binding the competitive inhibitor. Because this site consists of many flexible side-chains, RMSD of the competitive binding site residues to the complexed crystal structure was not an appropriate measure of whether the competitive site's conformation was favorable for competitive ligand binding. We found that the best-scoring docked pose roughly superimposes with the crystal ligand (2.8 Å RMSD without fitting, 1.4 Å RMSD with rotational/translational fitting, Fig. 11C). The cluster of MD snapshots with the best-scoring docked ligand represented 0.033% of total snapshots. Thus, our relatively short molecular dynamics simulations sampled conformers suitable for binding competitive site ligand at 300K in the presence of allosteric ligand, enabling us to create a model of the ternary complex (Fig. 11C).

Conclusions

We have reported novel improvements to mutual information calculations that make them robust enough for relatively short molecular dynamics simulations and have applied our MutInf method to interrogate the mechanism of small molecule binding cooperativity in human interleukin-2. We found better separation of signal from noise in our matrix of correlations between pairs of torsions than in similar matrices that examined backbone C_α correlations. We identified not only local correlations in sequence and in distance space but also long-range correlations. Clustering the matrix of mutual information between residues, we identified a few clusters whose residues showed strong patterns of correlations. Two of these clusters highlighted key functional sites, namely the IL-2R α -competitive protein interface/inhibitor binding site and a highly flexible loop that has to move to reveal a cryptic binding pocket for

the allosteric ligand. Furthermore, we found that the conformations of a number of pairs of residues in these two functional sites were strongly correlated.

As MutInf identified known cooperative binding or functional sites within the top clusters and correlations between them, we believe that this approach will be useful in identifying novel allosteric sites for proteins or for small molecules. For example, we predict potential allostery between the flexible loop surrounding the allosteric compound binding site and the N-terminus of helix 1, the loop region around Glu100, and the C-terminus. Our prediction is further supported by the observation that all of these regions showed significant backbone NMR chemical shift perturbations upon binding of the IL-2R α receptor⁶¹. However, the biological roles of these regions are not clear. The C-terminus of IL-2 interacts weakly with the γ_c receptor^{75, 80, 81} ($K_D \sim 0.7$ mM) and independently of IL-2R α binding. NMR chemical shift perturbations and isoelectric point changes upon addition of methionine to IL-2's N-terminus suggested a potential interaction between the N- and C-termini of apo IL-2 in solution⁸². Intriguingly, Thr3 is a site on human IL-2 that is variably glycosylated⁸³; in mice, the N-terminus is longer and displays substantial sequence and glycosylation pattern variability, which in turn impacts IL-2's function in Type I diabetes in mouse models^{84, 85}. An important caveat is that correlated motions are necessary but not sufficient for biologically-relevant allostery.

Though our statistical filtering enables us to find significant correlations in relatively short simulations, in applications where accurate total conformational entropy is desired, multiple longer simulations may be required to obtain absolute total entropy values that all converge to the same value, and higher-order terms might be needed. Nonetheless, our approach is useful in determining which residues or groups of residues show correlated conformations and which residues may mediate crosstalk between functional sites. One caveat is that MutInf focuses on coupled residue conformations rather than vibrations, and so we may not efficiently capture the role of semi-rigid elements in mediating correlations between more flexible sites. Though we did not look at motions faster than 1 ps, these are likely not as critical for ligand binding cooperativity in interleukin-2, where the residues linking the sites are primarily in flexible loops and have flexible side-chains. In general, however, such faster-timescale motions can help mediate cooperativity between more flexible sites, and so future work is needed to properly account for these in our approach.

Our calculations suggest that small molecule binding cooperativity in human interleukin-2 involves subtle population shifts and correlated conformations of two binding pockets coupled through a greasy core and a solvent-exposed polar network. New biophysical techniques to directly measure correlated motions by NMR would be useful in testing our predictions about correlated motions that couple allosteric sites.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors acknowledge J. Wells, M. Gilson, J. Gross, M. Arkin, H. Li, I. Kuntz, M. Kelly, J. Gross, and R. Abel for helpful discussions. The authors also acknowledge M. Hutter for assistance with his Bayesian approach. GF thanks T. Kortemme (UCSF) for support, and DLM and HA thank Ken Dill (UCSF) for support. CLM thanks J. Nilmeier for inspiration and gratefully acknowledges funding for this work from the UCSF integrated Program in Quantitative Biology fellowship and the UCSF Cancer Research Coordinating Committee. MPJ is a consultant to Schrodinger LLC.

References

1. Tsai CJ, del Sol A, Nussinov R. Allostery: Absence of a Change in Shape Does Not Imply that Allostery Is Not at Play. *J Mol Biol* 2008;378(1):1–11. [PubMed: 18353365]
2. Cui Q, Karplus M. Allostery and cooperativity revisited. *Protein Sci* 2008;17(8):1295–1307. [PubMed: 18560010]
3. Lindsley JE, Rutter J. Whence cometh the allosterome? *Proc Natl Acad Sci U S A* 2006;103(28):10533–10535. [PubMed: 16818878]
4. Gunasekaran K, Ma B, Nussinov R. Is allostery an intrinsic property of all dynamic proteins? *Proteins: Struct, Funct, Bioinf* 2004;57(3):433–443.
5. Kuriyan J, Eisenberg D. The origin of protein interactions and allostery in colocalization. *Nature* 2007;450(7172):983–990. [PubMed: 18075577]
6. Hardy JA, Wells JA. Searching for new allosteric sites in enzymes. *Curr Opin Struct Biol* 2004;14(6):706–715. [PubMed: 15582395]
7. Jeffrey Conn P, Christopoulos A, Lindsley CW. Allosteric modulators of GPCRs: a novel approach for the treatment of CNS disorders. *Nat Rev Drug Discov* 2009;8(1):41–54. [PubMed: 19116626]
8. Zhang J, Yang PL, Gray NS. Targeting cancer with small molecule kinase inhibitors. *Nat Rev Cancer* 2009;9(1):28–39. [PubMed: 19104514]
9. Hardy JA, Wells JA. Searching for new allosteric sites in enzymes. *Curr Opin Struct Biol* 2004;14(6):706–715. [PubMed: 15582395]
10. Shulman AI, Larson C, Mangelsdorf DJ, Ranganathan R. Structural determinants of allosteric ligand activation in RXR heterodimers. *Cell* 2004;116(3):417–429. [PubMed: 15016376]
11. Suel GM, Lockless SW, Wall MA, Ranganathan R. Evolutionarily conserved networks of residues mediate allosteric communication in proteins. *Nat Struct Biol* 2003;10(1):59–69. [PubMed: 12483203]
12. Lockless SW, Ranganathan R. Evolutionarily conserved pathways of energetic connectivity in protein families. *Science* 1999;286(5438):295–299. [PubMed: 10514373]
13. Hatley ME, Lockless SW, Gibson SK, Gilman AG, Ranganathan R. Allosteric determinants in guanine nucleotide-binding proteins. *Proc Natl Acad Sci USA* 2003;100(24):14445–14450. [PubMed: 14623969]
14. Lee J, Natarajan M, Nashine VC, Socolich M, Vo T, Russ WP, Benkovic SJ, Ranganathan R. Surface sites for engineering allosteric control in proteins. *Science* 2008;322(5900):438–442. [PubMed: 18927392]
15. Fuentes EJ, Gilmore SA, Mauldin RV, Lee AL. Evaluation of Energetic and Dynamic Coupling Networks in a PDZ Domain Protein. *J Mol Biol* 2006;364(3):337–351. [PubMed: 17011581]
16. Page MJ, Carrell CJ, Di Cera E. Engineering Protein Allostery: 1.05 Å Resolution Structure and Enzymatic Properties of a Na⁺-activated Trypsin. *J Mol Biol* 2008;378(3):666–672. [PubMed: 18377928]
17. Ming D, Wall ME. Quantifying allosteric effects in proteins. *Proteins: Struct, Funct, Bioinf* 2005;59(4):697–707.
18. Hawkins RJ, McLeish TCB. Coarse-Grained Model Of Entropic Allostery. *Phys Rev Lett* 2004;93(9):098104. [PubMed: 15447145]
19. Ming D, Wall ME. Allostery in a Coarse-Grained Model of Protein Dynamics. *Phys Rev Lett* 2005;95(19):198103. [PubMed: 16384030]
20. Hawkins RJ, McLeish TCB. Dynamic allostery of protein alpha helical coiled-coils. *J Royal Soc Interface* 2006;3(6):125–138.
21. Zhang D, McCammon JA. The Association of Tetrameric Acetylcholinesterase with ColQ Tail: A Block Normal Mode Analysis. *PLoS Comput Biol* 2005;1(6):e62. [PubMed: 16299589]
22. Chennubhotla C, Yang Z, Bahar I. Coupling between global dynamics and signal transduction pathways: a mechanism of allostery for chaperonin GroEL. *Mol BioSyst* 2008;4(4):287–292. [PubMed: 18354781]
23. Ming, Dengming; W, ME. Quantifying allosteric effects in proteins. *Proteins: Struct, Funct, Bioinf* 2005;59(4):697–707.

24. Chennubhotla C, Bahar I. Markov propagation of allosteric effects in biomolecular systems: application to GroEL-GroES. *Mol Syst Biol* 2006;2
25. Daily MD, Upadhyaya TJ, Gray JJ. Contact rearrangements form coupled networks from local motions in allosteric proteins. *Proteins: Struct, Funct, Bioinf* 2008;71(1):455–466.
26. Daily MD, Gray JJ. Local motions in a benchmark of allosteric proteins. *Proteins: Struct, Funct, Bioinf* 2007;67(2):385–399.
27. Daily MD, Gray JJ. Allosteric communication occurs via networks of tertiary and quaternary motions in proteins. *PLoS Comput Biol* 2009;5(2):e1000293. [PubMed: 19229311]
28. Li L, Uversky VN, Dunker AK, Meroueh SO. A Computational Investigation of Allostery in the Catabolite Activator Protein. *J Am Chem Soc* 2007;129(50):15668–15676. [PubMed: 18041838]
29. Bradley MJ, Chivers PT, Baker NA. Molecular Dynamics Simulation of the Escherichia coli NikR Protein: Equilibrium Conformational Fluctuations Reveal Interdomain Allosteric Communication Pathways. *J Mol Biol* 2008;378(5):1155–1173. [PubMed: 18433769]
30. Sayar K, Ugur O, Liu T, Hilser V, Onaran O. Exploring allosteric coupling in the alpha-subunit of Heterotrimeric G proteins using evolutionary and ensemble-based approaches. *BMC Struct Biol* 2008;8(1):23. [PubMed: 18454845]
31. Horstink LM, Abseher R, Nilges M, Hilbers CW. Functionally important correlated motions in the single-stranded DNA-binding protein encoded by filamentous phage Pf3. *J Mol Biol* 1999;287(3):569–577. [PubMed: 10092460]
32. Liu J, Nussinov R. Allosteric effects in the marginally stable von Hippel-Lindau tumor suppressor protein and allostery-based rescue mutant design. *Proc Natl Acad Sci U S A* 2008;105(3):901–906. [PubMed: 18195360]
33. Watney JB, Hammes-Schiffer S. Comparison of Coupled Motions in Escherichia coli and Bacillus subtilis Dihydrofolate Reductase. *J Phys Chem B* 2006;110(20):10130–10138. [PubMed: 16706474]
34. Lange OF, Grubmüller H. Generalized correlation for biomolecular dynamics. *Proteins: Struct, Funct, Bioinf* 2006;62(4):1053–1061.
35. Lange OF, Grubmüller H, de Groot BL. Molecular Dynamics Simulations of Protein G Challenge NMR-Derived Correlated Backbone Motions. *Angew Chem Intl Ed* 2005;44(22):3394–3399.
36. Ming D, Cohn J, Wall M. Fast dynamics perturbation analysis for prediction of protein functional sites. *BMC Struct Biol* 2008;8(1):5. [PubMed: 18234095]
37. Ho BK, Agard DA. Probing the Flexibility of Large Conformational Changes in Protein Structures through Local Perturbations. *PLoS Comput Biol* 2009;5(4):e1000343. [PubMed: 19343225]
38. Cooper A, Dryden DTF. Allostery without conformational change. *Eur Biophys J* 1984;11(2):103–109. [PubMed: 6544679]
39. Popovych N, Sun S, Ebricht RH, Kalodimos CG. Dynamically driven protein allostery. *Nat Struct Mol Biol* 2006;13(9):831–838. [PubMed: 16906160]
40. Lenaerts T, Ferkinghoff-Borg J, Stricher F, Serrano L, Schymkowitz J, Rousseau F. Quantifying information transfer by protein domains: Analysis of the Fyn SH2 domain structure. *BMC Struct Biol* 2008;8(1):43. [PubMed: 18842137]
41. Smith CA, Kortemme T. Backrub-Like Backbone Simulation Recapitulates Natural Protein Conformational Variability and Improves Mutant Side-Chain Prediction. *J Mol Biol* 2008;380(4):742–756. [PubMed: 18547585]
42. Friedland GD, Linares AJ, Smith CA, Kortemme T. A Simple Model of Backbone Flexibility Improves Modeling of Side-chain Conformational Variability. *J Mol Biol* 2008;380(4):757–774. [PubMed: 18547586]
43. Killian BJ, Kravitz JY, Gilson MK. Extraction of configurational entropy from molecular simulations via an expansion approximation. *J Chem Phys* 2007;127(2):024107–024116. [PubMed: 17640119]
44. Zheng W, Brooks BR, Thirumalai D. Low-frequency normal modes that describe allosteric transitions in biological nanomachines are robust to sequence variations. *Proc Natl Acad Sci U S A* 2006;103(20):7664–7669. [PubMed: 16682636]
45. Matsuda H. Physical nature of higher-order mutual information: Intrinsic correlations and frustration. *Phys Rev E* 2000;62(3):3096.

46. Steuer R, Kurths J, Daub CO, Weise J, Selbig J. The mutual information: Detecting and evaluating dependencies between variables. *Bioinformatics* 2002;18(suppl2):S231–240. [PubMed: 12386007]
47. Grassberger P. Finite sample corrections to entropy and dimension estimates. *Phys Lett A* 1988;128(67):369–373.
48. Roulston MS. Estimating the errors on measured entropy and mutual information. *Physica D* 1999;125(34):285–294.
49. Hnizdo V, Tan J, Killian BJ, Gilson MK. Efficient calculation of configurational entropy from molecular simulations by combining the mutual-information expansion and nearest-neighbor methods. *J Comput Chem* 2008;29(10):1605–1614. [PubMed: 18293293]
50. Hnizdo V, Darian E, Fedorowicz A, Demchuk E, Li S, Singh H. Nearest-neighbor nonparametric method for estimating the configurational entropy of complex molecules. *J Comput Chem* 2007;28(3):655–668. [PubMed: 17195154]
51. Karchin R, Kelly L, Sali A. Improving functional annotation of non-synonymous SNPs with information theory. *Pac Symp Biocomput* 2005:397–408. [PubMed: 15759645]
52. Shackelford, George; K, K. Contact prediction using mutual information and neural nets. *Proteins: Struct, Funct, Bioinf* 2007;69(S8):159–164.
53. Francois D, Rossi F, Wertz V, Verleysen M. Resampling methods for parameter-free and robust feature selection with mutual information. *Neurocomputing* 2007;70(79):1276–1288.
54. Hutter M, Zaffalon M. Distribution of mutual information from complete and incomplete data. *Comp Stat & Data Anal* 2005;48(3):633–657.
55. Lindahl E, Hess B, van der Spoel D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J Mol Model* 2001;7(8):306–317.
56. Spoel DVD, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC. GROMACS: Fast, flexible, and free. *J Comput Chem* 2005;26(16):1701–1718. [PubMed: 16211538]
57. Sorin EJ, Pande VS. Exploring the Helix-Coil Transition via All-Atom Equilibrium Ensemble Simulations. *Biophys J* 2005;88(4):2472–2493. [PubMed: 15665128]
58. Jacobson MP, Pincus DL, Rapp CS, Day TJF, Honig B, Shaw DE, Friesner RA. A hierarchical approach to all-atom protein loop prediction. *Proteins: Struct, Funct, and Bioinf* 2004;55(2):351–367.
59. Alexov EG, Gunner MR. Incorporating protein conformational flexibility into the calculation of pH-dependent protein properties. *Biophys J* 1997;72(5):2075–2093. [PubMed: 9129810]
60. Georgescu RE, Alexov EG, Gunner MR. Combining Conformational Flexibility and Continuum Electrostatics for Calculating pKas in Proteins. *Biophys J* 2002;83(4):1731–1748. [PubMed: 12324397]
61. Emerson SD, Palermo R, Liu CM, Tilley JW, Chen L, Danho W, Madison VS, Greeley DN, Ju G, Fry DC. NMR characterization of interleukin-2 in complexes with the IL-2R α receptor component, and with low molecular weight compounds that inhibit the IL-2/IL-R α interaction. *Protein Sci* 2003;12(4):811–822. [PubMed: 12649439]
62. Arkin MR, Randal M, DeLano WL, Hyde J, Luong TN, Oslob JD, Raphael DR, Taylor L, Wang J, McDowell RS, Wells JA, Braisted AC. Binding of small molecules to an adaptive protein-protein interface. *Proc Natl Acad Sci U S A* 2003;100(4):1603–1608. [PubMed: 12582206]
63. Hyde J, Braisted AC, Randal M, Arkin MR. Discovery and characterization of cooperative ligand binding in the adaptive region of interleukin-2. *Biochemistry (Mosc)* 2003;42(21):6475–6483.
64. Junmei W, Romain MW, James WC, Peter AK, David AC. Development and testing of a general amber force field. *J Comput Chem* 2004;25(9):1157–1174. [PubMed: 15116359]
65. Jakalian A, Bush BL, Jack DB, Bayly CI. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. *J Comput Chem* 2000;21:132.
66. Jakalian A, Jack DB, Bayly CI. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J Comput Chem* 2002;23:1623. [PubMed: 12395429]
67. Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR. Molecular dynamics with coupling to an external bath. *J Chem Phys* 1984;81(8):3684–3690.
68. Berk H, Henk B, Herman JCB, Johannes GEMF. LINCS: A linear constraint solver for molecular simulations. *J Comput Chem* 1997;18(12):1463–1472.

69. Daura X, Gademann K, Jaun B, Seebach D, Gunsteren WFv, Mark AE. Peptide Folding: When Simulation Meets Experiment. *Angew Chem Intl Ed* 1999;38(12):236–240.
70. Friesner RA, Murphy RB, Repasky MP, Frye LL, Greenwood JR, Halgren TA, Sanschagrin PC, Mainz DT. Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J Med Chem* 2006;49(21):6177–6196. [PubMed: 17034125]
71. Yang LW, Rader AJ, Liu X, Jursa CJ, Chen SC, Karimi HA, Bahar I. oGNM: online computation of structural dynamics using the Gaussian Network Model. *Nucl Acids Res* 2006;34(suppl2):W24–31. [PubMed: 16845002]
72. Dill, K.; Bromberg, S. *Molecular Driving Forces: Statistical Thermodynamics in Chemistry & Biology*. Garland Science; 2002.
73. Hilser VJ, Thompson EB. Intrinsic disorder as a mechanism to optimize allosteric coupling in proteins. *Proc Natl Acad Sci U S A* 2007;104(20):8311–8315. [PubMed: 17494761]
74. Rekharsky MV, Mori T, Yang C, Ko YH, Selvapalam N, Kim H, Sobransingh D, Kaifer AE, Liu S, Isaacs L, Chen W, Moghaddam S, Gilson MK, Kim K, Inoue Y. A synthetic host-guest system achieves avidin-biotin affinity by overcoming enthalpy entropy compensation. *Proc Natl Acad Sci USA* 2007;104(52):20737–20742. [PubMed: 18093926]
75. Wang H, Kakaradov B, Collins SR, Karotki L, Fiedler D, Shales M, Shokat KM, Walther T, Krogan NJ, Koller D. A complex-based reconstruction of the *S. cerevisiae* interactome. *Mol Cell Proteomics* 2009:M800490–MCP800200.
76. Rickert M, Wang X, Boulanger MJ, Goriatheva N, Garcia KC. The structure of interleukin-2 complexed with its alpha receptor. *Science* 2005;308(5727):1477–1480. [PubMed: 15933202]
77. Fruchterman TMJ, Reingold EM. Graph drawing by force-directed placement. *Softw Pract Exper* 1991;21(11):1129–1164.
78. Thanos CD, DeLano WL, Wells JA. Hot-spot mimicry of a cytokine receptor by a small molecule. *Proc Natl Acad Sci USA* 2006;103(42):15422–15427. [PubMed: 17032757]
79. Thanos CD, DeLano WL, Wells JA. Hot-spot mimicry of a cytokine receptor by a small molecule. *Proc Natl Acad Sci USA* 2006;103(42):15422–15427. [PubMed: 17032757]
80. Liparoto SF, Myszka DG, Wu Z, Goldstein B, Laue TM, Ciardelli TL. Analysis of the Role of the Interleukin-2 Receptor Gamma Chain in Ligand Binding. *Biochemistry* 2002;41(8):2543–2551. [PubMed: 11851401]
81. Stauber DJ, Debler EW, Horton PA, Smith KA, Wilson IA. Crystal structure of the IL-2 signaling complex: paradigm for a heterotrimeric cytokine receptor. *Proc Natl Acad Sci U S A* 2006;103(8):2788–2793. [PubMed: 16477002]
82. Endo S, Yamamoto Y, Sugawara T, Nishimura O, Fujino M. The Additional Methionine Residue at the N-Terminus of Bacterially Expressed Human Interleukin-2 Affects the Interaction between the N- and C-Termini. *Biochemistry* 2001;40(4):914–919. [PubMed: 11170412]
83. Robb RJ, Smith KA. Heterogeneity of human T-cell growth factor(s) due to variable glycosylation. *Mol Immunol* 1981;18(12):1087–1094. [PubMed: 6977702]
84. Podolin PL, Wilusz MB, Cubbon RM, Pajvani U, Lord CJ, Todd JA, Peterson LB, Wicker LS, Lyons PA. Differential glycosylation of interleukin 2, the molecular basis for the NOD Idd3 type 1 diabetes gene? *Cytokine* 2000;12(5):477–482. [PubMed: 10857762]
85. Sgouroudis E, Albanese A, Piccirillo CA. Impact of Protective IL-2 Allelic Variants on CD4+Foxp3 + Regulatory T Cell Function In Situ and Resistance to Autoimmune Diabetes in NOD Mice. *J Immunol* 2008;181(9):6283–6292. [PubMed: 18941219]

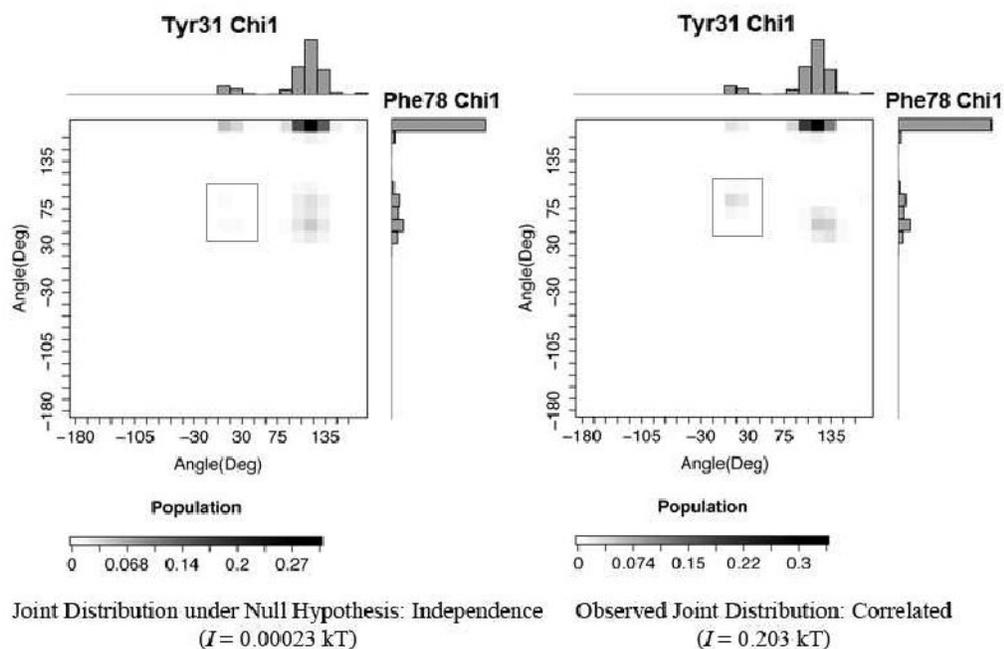


Fig. 1. Joint distributions of correlated torsions are different from what would be expected if they were independent. (A) Distributions of two χ_1 torsion angles are shown along with the joint distribution expected if they were independent, i.e. the product of the marginal probabilities. (B) Distributions of the same two χ_1 torsion angles are shown along with the observed joint distribution from molecular dynamics simulations. Grey boxes highlight a cross-peak with substantial height in the observed simulations (B) but with negligible height under the null hypothesis of independence (A).

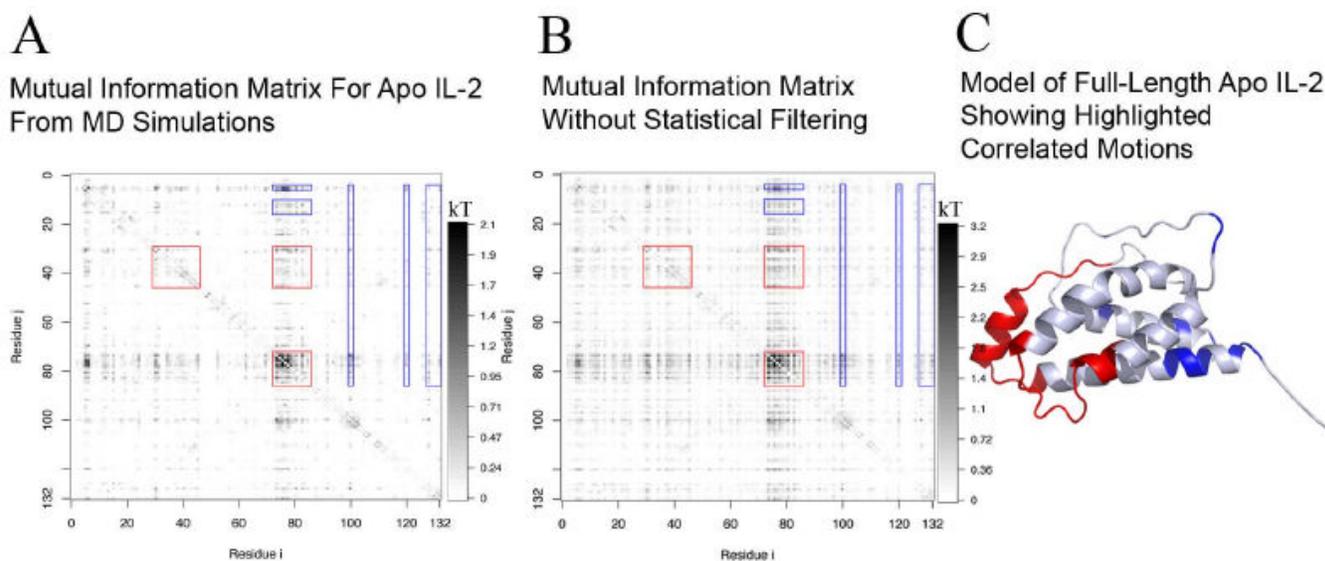


Fig. 2.

Mutual Information captures significant correlations between residues in human interleukin-2. (A) Mutual information between residues' torsions computed using the present approach, with statistical filtering as detailed in Methods. (B) Same as in A but without any of the aforementioned statistical corrections. (C) The model of full-length human interleukin-2 used in the apo simulations, based on the crystal structure of apo IL-2 (PDB: 1M47). Residues surrounded by red boxes in A are colored red, while residues correlated to these that are surrounded by blue boxes in A are colored blue.

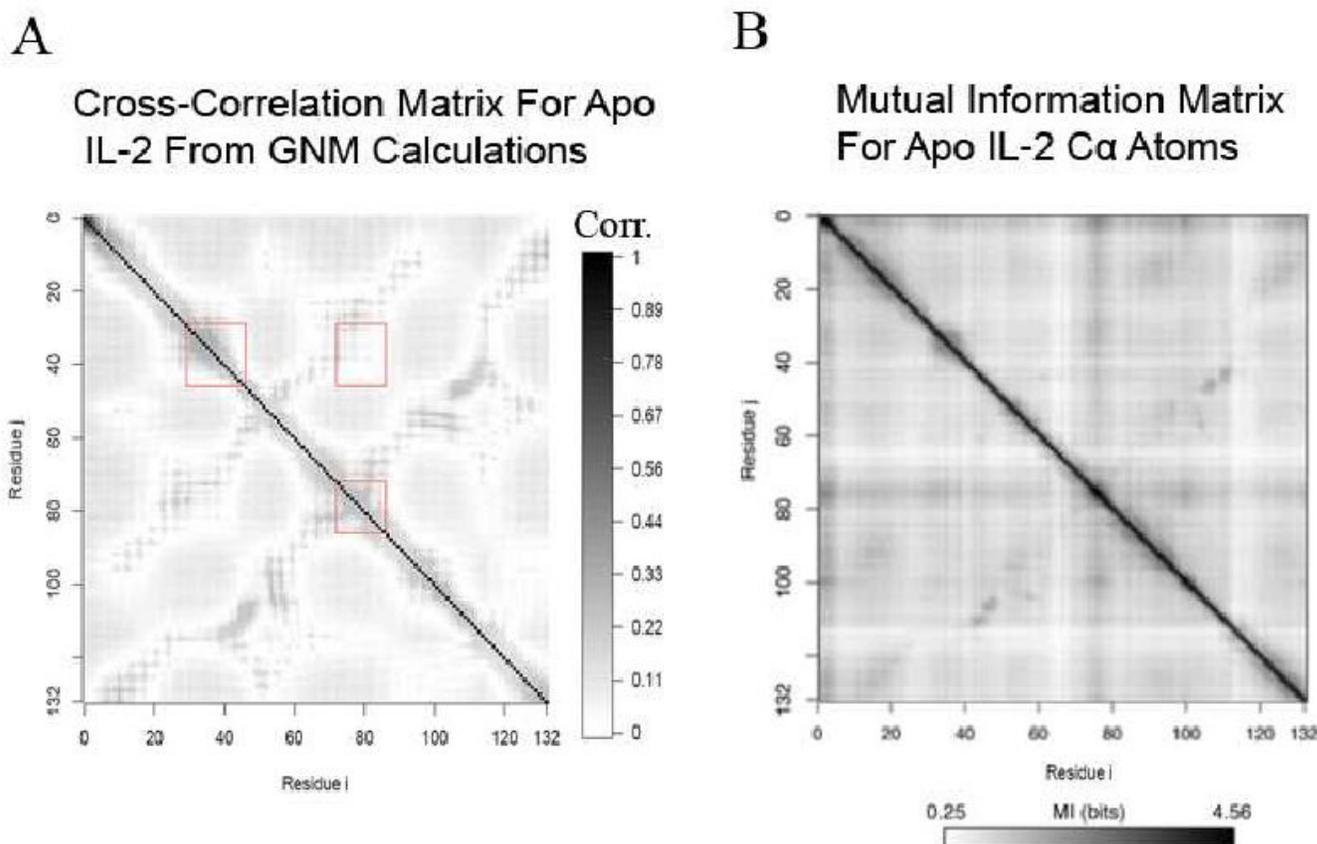


Fig. 3. Comparison of pairwise, dynamical correlations between residues computed by alternative methods. (A) Absolute value of the cross-correlation matrix computed using the Gaussian Network Model. (B) Mutual Information between residues' C α Cartesian coordinates using the approach of Lange and Grubmüller.

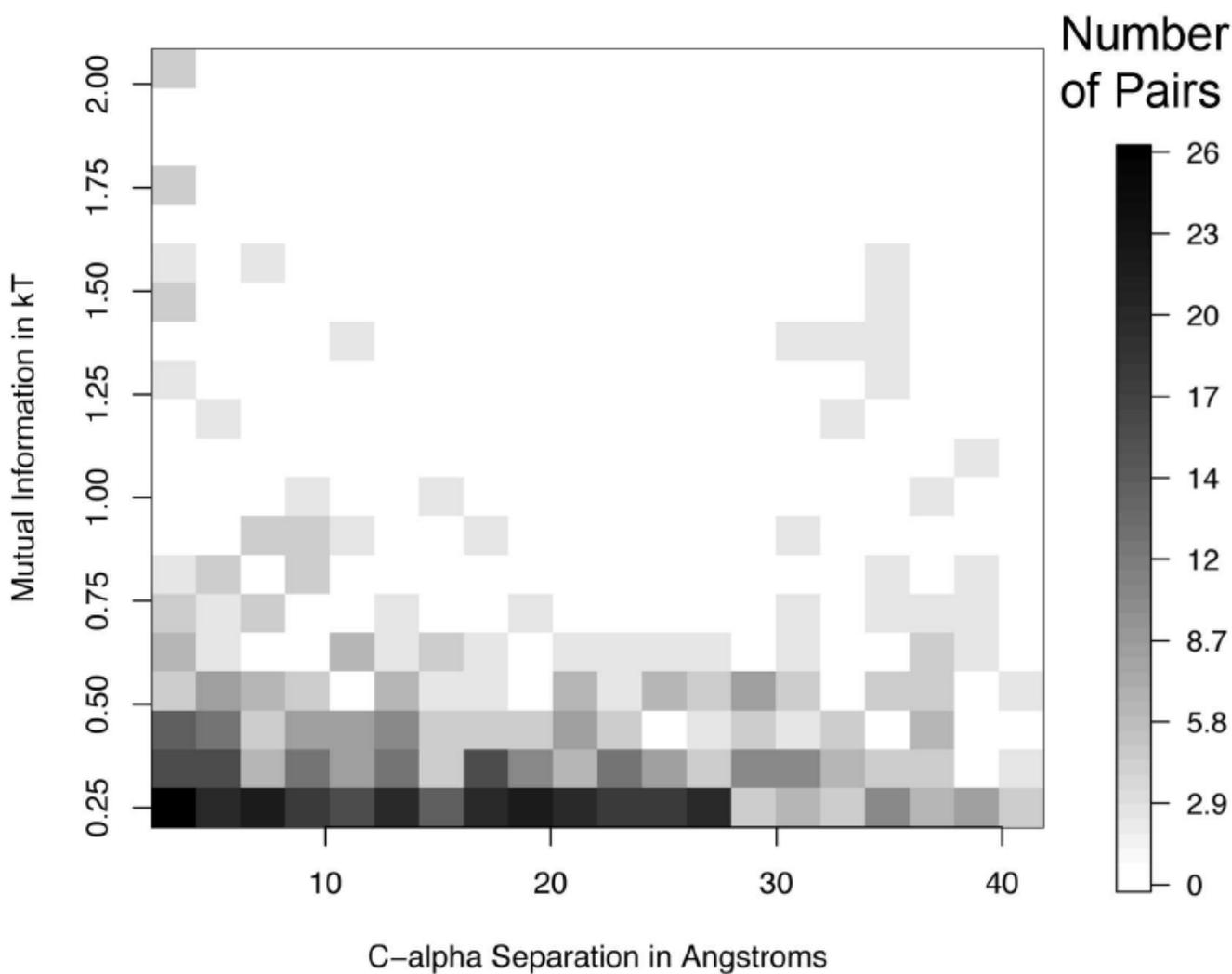


Fig. 4.

Most of the significant correlations are between distant residues. A 2-D histogram showing the number of pairs of correlated residues vs. C α separation and mutual information value shows that the number and strength of correlated pairs decreases only modestly with distance. For clarity, only those pairs of correlations with a mutual information greater than 0.25 kT are shown.

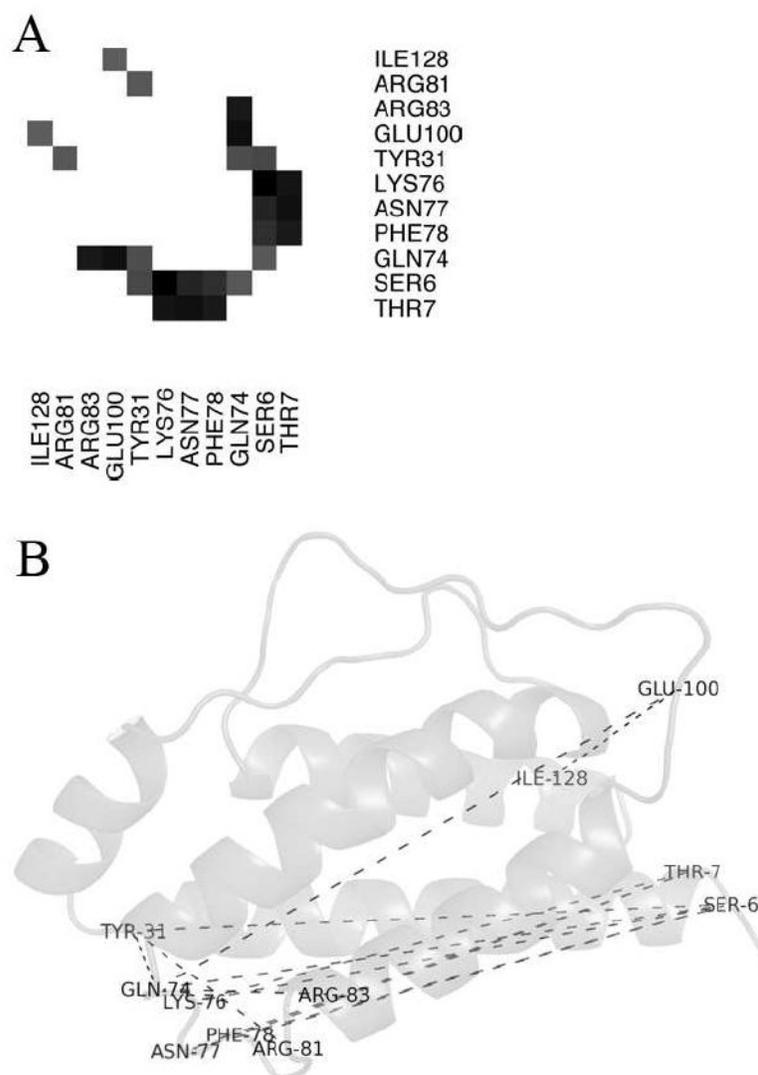


Fig. 5. Several distant residues are highly correlated. (A) Correlations greater than kT are shown only for IL-2 residues whose alpha carbons are separated by more than 5 Å. (B) Dashes connecting each pair of these correlated residues show long-range correlations across the length of the helical bundle.

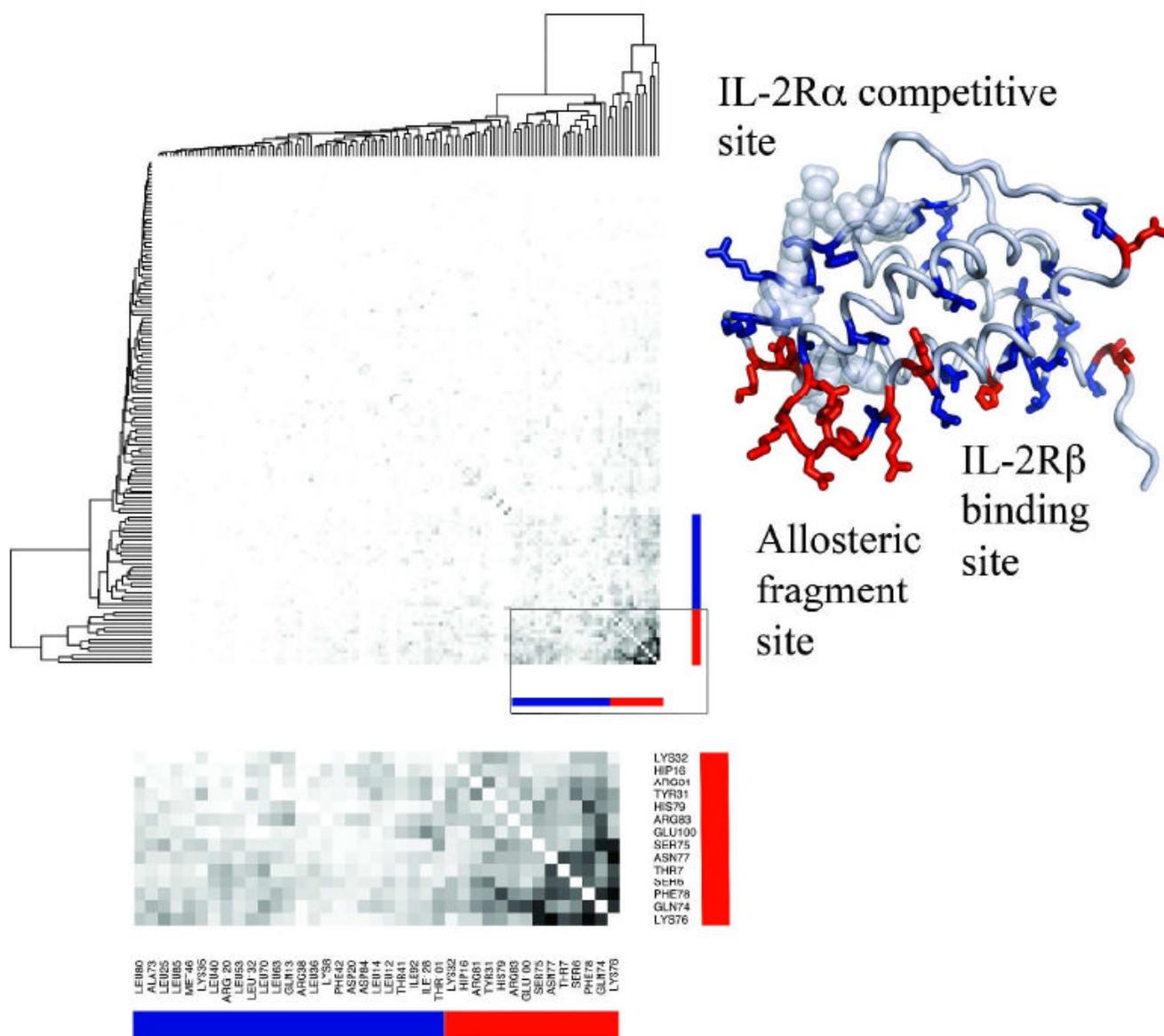


Fig. 6.

Hierarchical clustering of significant mutual information values identifies allosteric sites. A hierarchically-clustered heatmap shows clusters (top left) of residues with similar patterns of mutual information across IL-2 residues. A close-up view highlights numerous significant mutual information values between pairs of residues in two different clusters, red and blue. These red and blue clusters are highlighted in a model of IL-2's ternary complex (right). The strongest cluster (red sticks) chiefly involves a loop enclosing the allosteric fragment's binding site, and this cluster is correlated to a cluster (blue sticks) containing two protein binding sites, the IL-2R α -receptor-binding/IL-2R α -inhibitor-binding site and the IL-2R β -binding site. The two compound binding sites and the two protein-binding sites are directly correlated through the hydrophobic core (in the blue cluster), through a highly flexible loop (in the red cluster), and crosstalk between these elements, seen in a close-up view of the matrix (bottom).

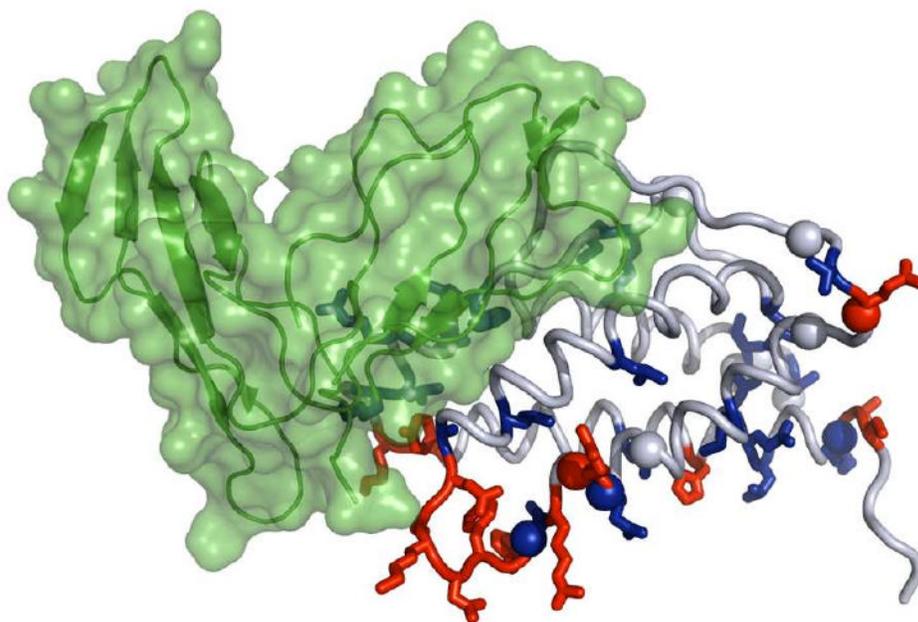
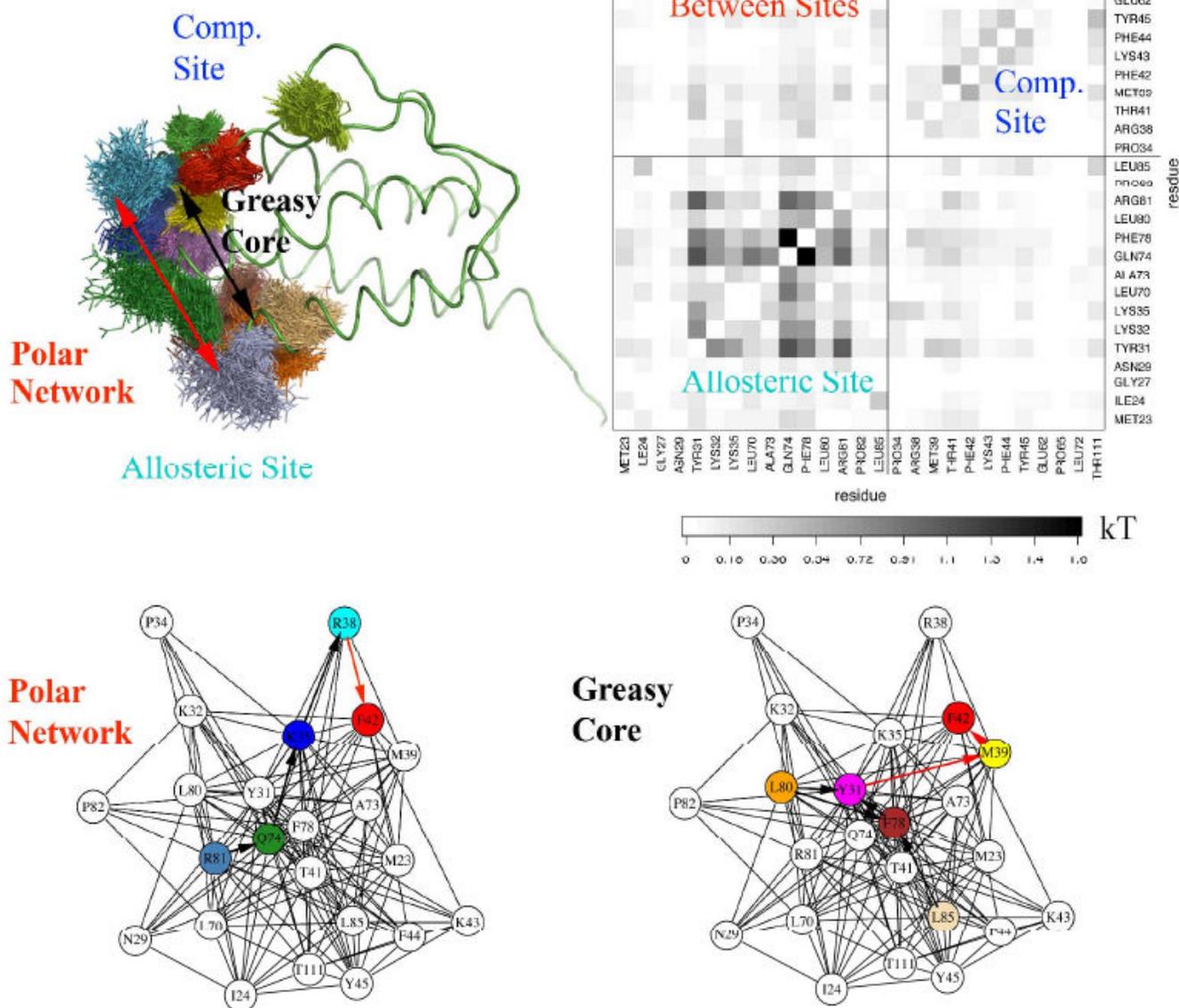


Fig. 7.

Predicted couplings are consistent with regions perturbed upon IL-2R α binding. Regions distant from the IL-2R α receptor binding site that show substantial backbone chemical shift perturbations upon IL-2R α binding⁶¹ roughly correspond to regions with residues whose conformations are correlated with residues in the IL-2R α binding site (predominantly residues in the “blue” cluster in Fig. 6). Amides on IL-2 whose resonances shifted by more than three linewidths upon IL-2R α binding or fell below 7% of the original intensity are shown as spheres. Residues from the “red and “blue” clusters shown in Fig. 6 are colored accordingly. IL-2 is as shown in cartoon and sticks as in Fig. 6, while IL-2R α is shown in green.

**Fig. 8.**

Direct, pairwise correlations couple residues in the IL-2R α -competitive site (at the IL-2:IL-2R α interface) to residues in the allosteric fragment-binding site (near the IL-2:IL-2R β interface).

(Top left) Apo IL-2 is shown in green ribbon while representative conformations of residues showing strong correlations within and between these sites are shown with lines. Overlap between clouds of residues' conformations suggest steric coupling, particularly in the greasy core, from Leu80 (orange) and Leu85 (tan) to Phe78 (brown), to Tyr31 (magenta), to Met39 (yellow), and to Phe42 (red). (Top right) A subset of the full matrix of pairwise correlations reveals direct correlations between residues in the two sites, with the labeled boxes showing correlations within the allosteric site, within the competitive site, and between these two sites. (Bottom) A force-directed network diagram⁷⁷ for residues in these sites filtered for correlations

of at least 0.05kT shows Phe78, Tyr31, Gln74, and Arg81 as central “hub” residues mediating correlations between the sites.

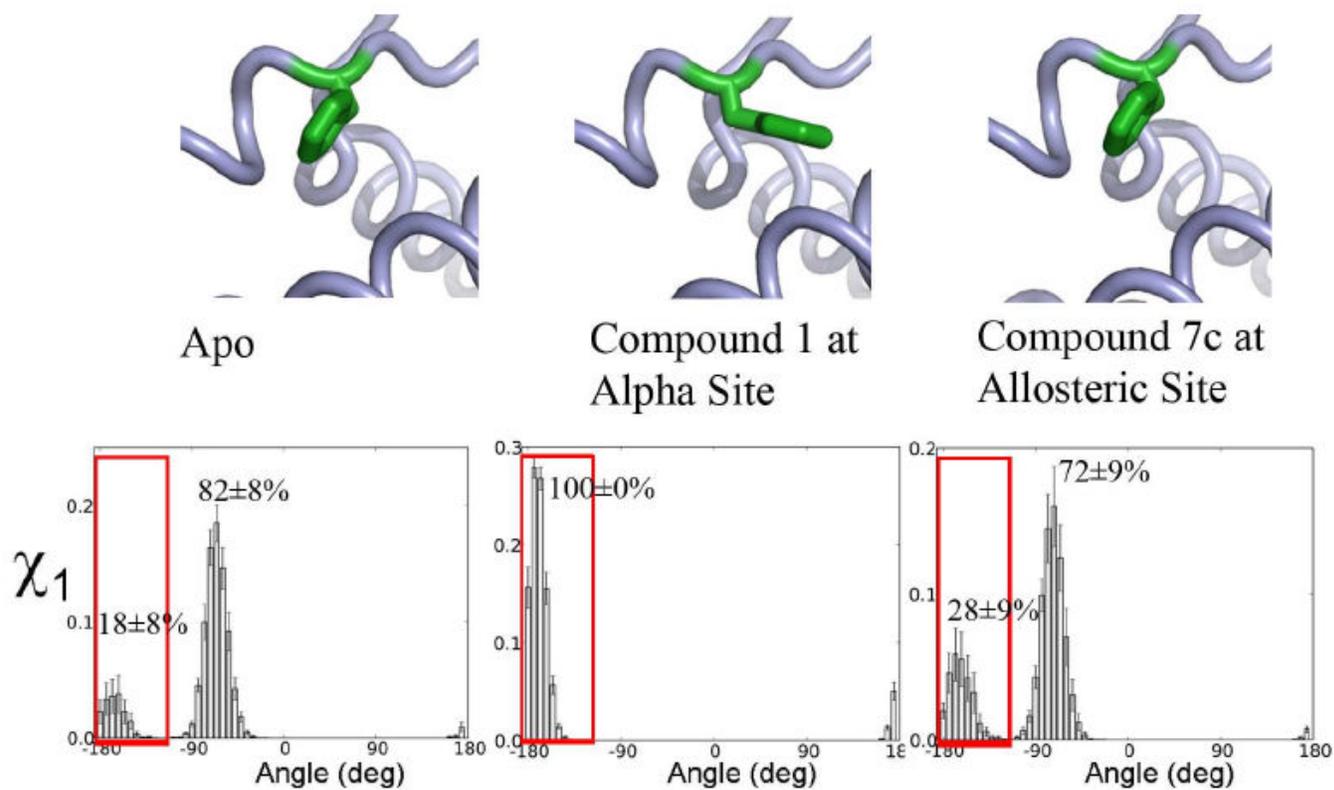


Fig. 9. Compound binding to the allosteric site causes a population shift in the conformation of hot-spot residue Phe42 that favors binding compound at the IL-2R α -competitive site. (Top) Conformations of Phe42 in apo and compound-bound crystal structures (PDB IDs 1M47, 1M48, and 1NBP, resp.). (Bottom) Histograms of Phe42's χ_1 angle from MD simulations. Red boxes highlight the χ_1 population selected by ligand binding at the competitive site.

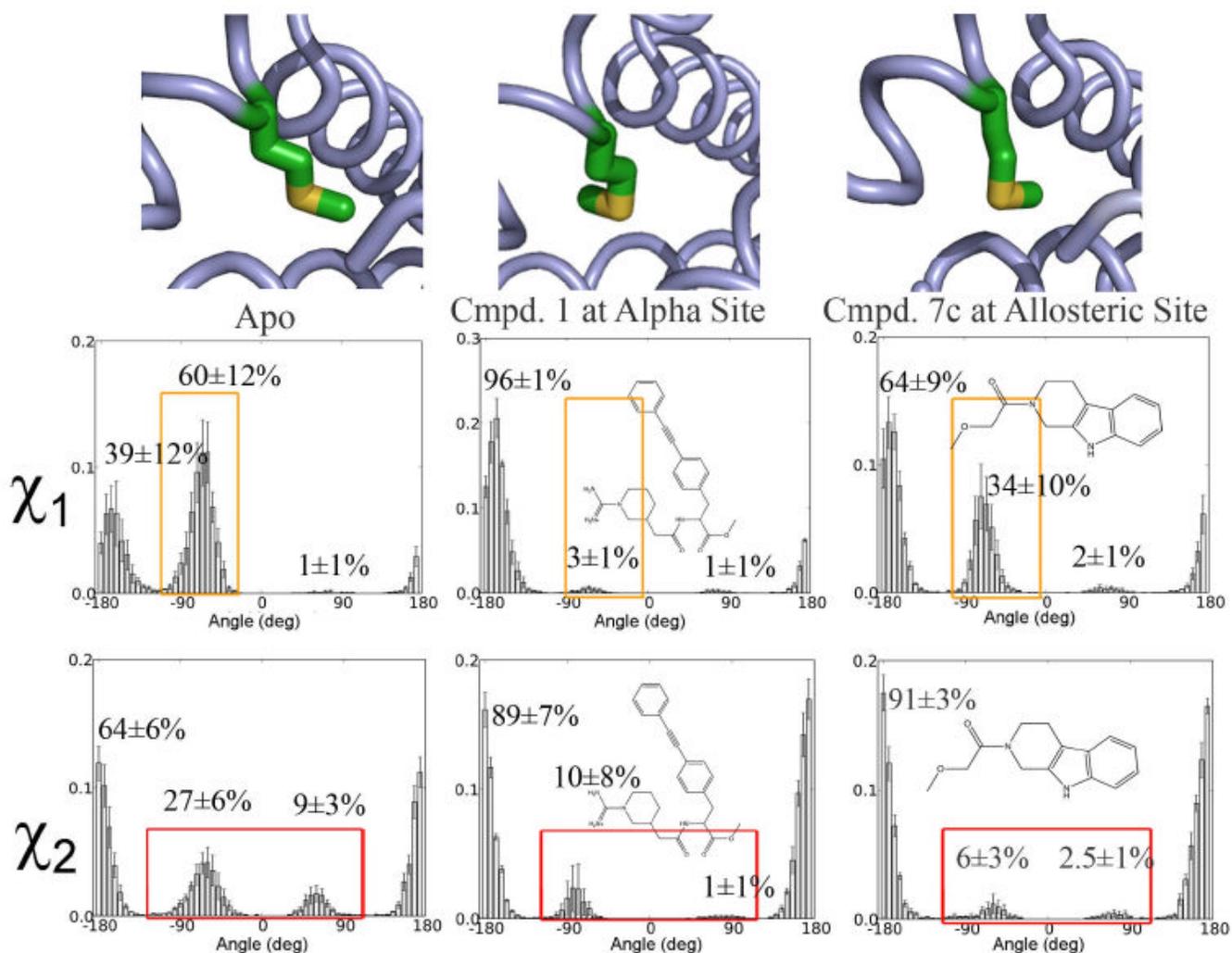


Fig. 10. Compound binding to the IL-2R α site or to the allosteric site selects conformations of Met39 favorable for binding compound at the other site. (Top) Conformations of Met39 in apo and compound-bound crystal structures (PDB IDs 1M47, 1M48, and 1NBP, resp.). (Bottom) Histograms of Met39's χ_1 and χ_2 angles from MD simulations. Orange boxes in χ_1 and red boxes in χ_2 highlight populations suppressed in ligand-bound simulations.

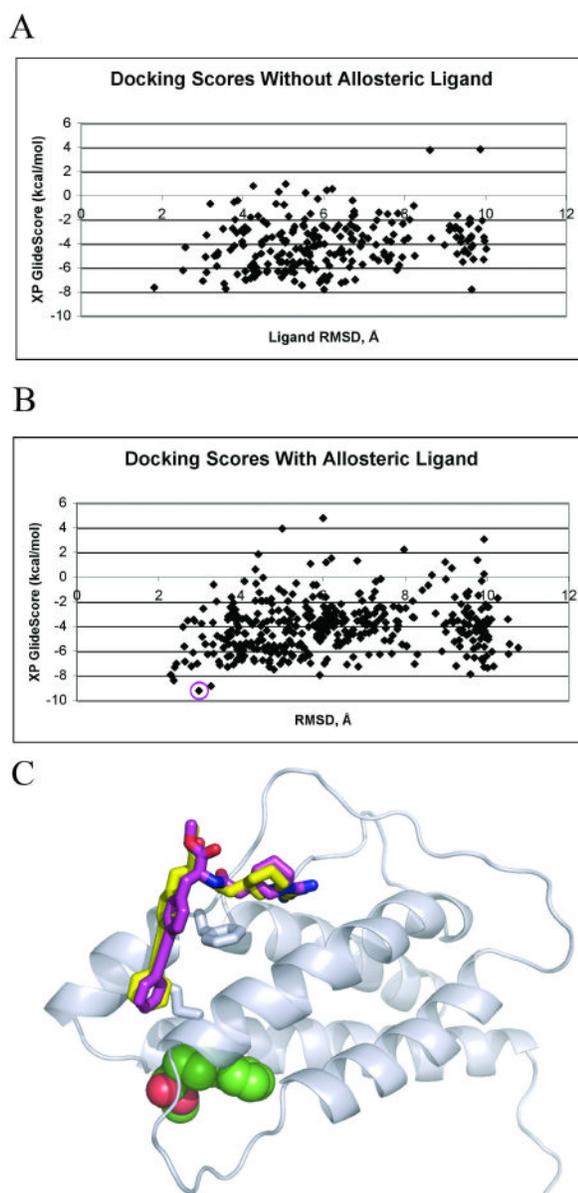
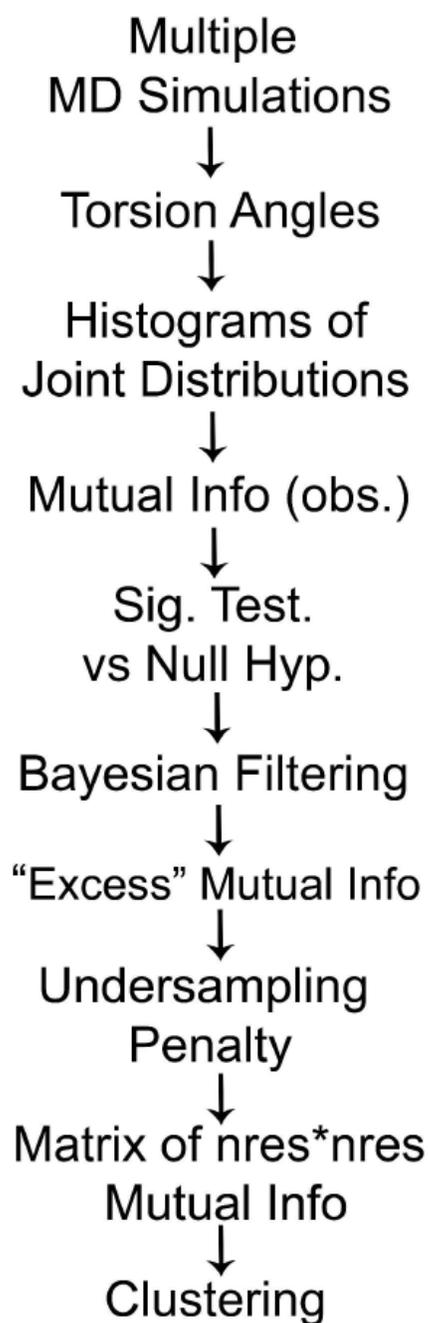


Fig. 11. Docking using Glide XP selects a holo-like conformation from an MD ensemble. (A, B) Plots of docking score vs. ligand RMSD to the forcefield-minimized co-crystal conformation show that the best-scoring docked poses from simulations with (B) but not without (A) allosteric compound bound had relatively low RMSD values. The best-scoring pose is circled. (C) Molecular dynamics snapshot from a simulation of IL-2 with bound allosteric fragment is shown with docked (yellow) vs. superimposed X-ray (magenta) conformations of a micromolar small molecule inhibitor of IL-2R α binding. Though the absolute RMSD for this ligand is 2.9Å (1.6Å RMSD after fitting), it has a binding mode very similar to that of the crystal ligand.

**Scheme 1.**

A schematic of the MutInf approach for identifying correlated residue conformations shows how the observed mutual information is statistically filtered and corrected before being summed over residues pairs. The resulting matrix is then clustered as in a microarray experiment in order to identify groups of residues showing similar patterns of correlations.