Title

Kinematic Flexibility Analysis: Hydrogen Bonding Patterns Impart a Spatial Hierarchy of Protein Motion

Journal

Journal of Chemical Information and Modeling, 58(10)

ISSN

1549-9596

Authors

Budday, Dominik
Leyendecker, Sigrid
van den Bedem, Henry

Publication Date

2018-10-22

DOI

10.1021/acs.jcim.8b00267

Peer reviewed

HHS Public Access

Author manuscript

J Chem Inf Model. Author manuscript; available in PMC 2018 November 29.

Published in final edited form as:

J Chem Inf Model. 2018 October 22; 58(10): 2108–2122. doi:10.1021/acs.jcim.8b00267.

Kinematic Flexibility Analysis: Hydrogen Bonding Patterns Impart a Spatial Hierarchy of Protein Motion

Dominik Budday[*,†], **Sigrid Leyendecker**[†], and **Henry van den Bedem**[*,‡,¶]

[†]Chair of Applied Dynamics, University of Erlangen–Nuremberg, 91058 Erlangen, Germany

[‡]Biosciences Division, SLAC National Accelerator Laboratory, Stanford University, Menlo Park, California 94025, United States

[¶]Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, California 94158, United States

Abstract

Elastic network models (ENMs) and constraint-based, topological rigidity analysis are two distinct, coarse-grained approaches to study conformational flexibility of macromolecules. In the two decades since their introduction, both have contributed significantly to insights into protein molecular mechanisms and function. However, despite a shared purpose of these approaches, the topological nature of rigidity analysis, and thereby the absence of motion modes, has impeded a direct comparison. Here, we present an alternative, kinematic approach to rigidity analysis, which circumvents these drawbacks. We introduce a novel protein hydrogen bond network spectral decomposition, which provides an orthonormal basis for collective motions modulated by noncovalent interactions, analogous to the eigenspectrum of normal modes. The zero modes decompose proteins into rigid clusters identical to those from topological rigidity, while nonzero modes rank protein motions by their hydrogen bond collective energy penalty. Our kinematic flexibility analysis bridges topological rigidity theory and ENM, enabling a detailed analysis of motion modes obtained from both approaches. Analysis of a large, structurally diverse data set revealed that collectivity of protein motions, reported by the Shannon entropy, is significantly reduced for rigidity theory compared to normal mode approaches. Strikingly, kinematic flexibility analysis suggests that the hydrogen bonding network encodes a protein-fold specific, spatial hierarchy of motions, which goes nearly undetected in ENM. This hierarchy reveals distinct motion regimes that rationalize experimental and simulated protein stiffness variations. Kinematic motion modes highly correlate with reported crystallographic B factors and molecular dynamics simulations of adenylate kinase. A formal expression for changes in free energy derived from the spectral decomposition indicates that motions across nearly 40% of modes obey enthalpy–entropy compensation. Taken together, our results suggest that hydrogen bond networks have evolved to

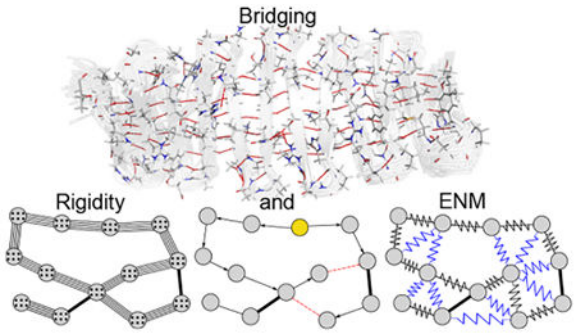
[*]**Corresponding Authors**: dominik.budday@fau.de (D.B.).; vdbedem@stanford.edu (H.v.d.B.).

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jcim.8b00267. Details on the various data sets used in this study, hydrogen bond identification and modeling via the Mayo potential, parameter settings for simulations with iMOD, and further implementation and parameter details on the conformation analysis of adenylate kinase with KGS and MD simulations (PDF)

The authors declare no competing financial interest.

modulate protein structure *and* dynamics, which can be efficiently probed by kinematic flexibility analysis.



INTRODUCTION

Coarse-grained modeling techniques can provide significant insight into the dynamic behavior and biological function of macromolecules. Elastic network models (ENMs[1]) and rigidity theory based flexibility analysis[2,3] are two well-known approaches that have been applied extensively to study molecular motion. Although they are based on physically distinct concepts, both aim to distinguish more flexible regions from compact ones in the molecule and to predict large-scale, functional motions.

ENM (Figure 1, right) approximate the potential energy function V via pairwise harmonic interactions $V = \sum_{ij} C_{ij}\left(\left|\mathbf{p}_{ij}\right| - \left|\mathbf{p}_{ij}^0\right|\right)^2$, where C_{ij} denotes the stiffness[1] or an exponentially fading weight[4] of the restraint between atoms i, j located at $\boldsymbol{p}_i, \boldsymbol{p}_j$, respectively, with $\mathbf{p}_{ij} = \mathbf{p}_j - \mathbf{p}_i$ and rest length $\left|\boldsymbol{p}_{ij}^0\right|$. The resulting spring-mass network can be formulated in terms of dihedral[5,6] or Cartesian[1,7–9] degrees of freedom (DoF) and analyzed with classical Hamiltonian mechanics. Diagonalization of the Hessian matrix leads to a spectral distribution of the dynamics in terms of eigenmodes with corresponding eigenfrequencies. The simplified energy function of ENM places the native structure at a global minimum, circumventing initial minimization necessary in traditional normal-mode analysis (NMA).[10] The motions corresponding to low-frequency modes are robust, often agree with those of more detailed models,[11,12] and can be functionally relevant.[13] Dihedral-based approaches often correlate better with experimentally observed conformational changes than Cartesian-based methods.[14,15] Other variants include Gaussian[16,17] or anisotropic[18] network models.

In rigidity theory, macromolecules are modeled as constraint graphs (Figure 1, left), with edges between interacting atoms (vertices) representing covalent and noncovalent bonds. Different types of constraint graphs, such as the bar-joint,[19,20] the body-bar,[21,22] or the equivalent body-bar-hinge graph[23,24] all share the concept of assigning a number of pebbles to atomic vertices as DoF, and a number of bars to each interaction as constraints. Typically, rotatable single-covalent bonds and hydrogen bonds retain a dihedral degree of freedom, while peptide and double covalent bonds are modeled rigid. Noncovalent interactions, such as hydrogen bonds, are added as constraints if their interaction strength exceeds a certain threshold. More constraints increasingly rigidify the graph. The remaining flexibility, i.e.,

the coordinated motion of internal DoF that obey constraints (*floppy modes* or *zero modes*)[19,25,26] can be determined via constraint counting through the pebble game algorithm.[3,19] Constraint counting on these graphs has been applied to examine conformational flexibility in macromolecules,[20,27–30] probe the effects of ligand binding,[31] estimate entropic measures,[32–34] and help shape perturbation strategies in complex motion planning algorithms.[35–37]

In constraint graphs, an energy threshold measuring the strength of a noncovalent interaction determines dichotomically whether to include the corresponding constraint, thereby changing the topology, but ignoring the collectivity of network effects. Monitoring changes in protein rigidity from diluting the constraint network by gradually raising the energy threshold has informed on thermostability[38–40] and the evolution of folding cores.[29,41] However, the topological and combinatorial nature of this approach denies explicit access to the kinematics of the underlying 3-dimensional molecular structure. Hence, two different substates, such as an open and a closed protein conformation, will have the same rigidity and flexibility properties, if they share an identical constraint network. Rigidity thus becomes dependent *only* on the energy threshold used, and ensemble-based approaches have to be employed to produce robust results.[28,42] Importantly, the exact motion vectors determined by the constraints (covalent and noncovalent bonds) remain unknown and can only be approximated by randomized perturbations[36] and iterative loop-closure algorithms such as ROCK[43] or FRODA.[44] While previous efforts tried to combine pebble game rigidity analysis with ENM,[45,46] these impediments have so far prevented a detailed comparison between the two methods.

Here, we overcome these limitations by using a kinematic approach to characterize flexibility and rigidity. Our approach explicitly provides basis vectors for motions corresponding to floppy modes, allowing us to analyze and compare motions directly from ENM and rigidity analysis. We previously established that topological rigidity and kinematic analysis yield an identical decomposition of the protein into rigid clusters for fully constraint compliant motions in nonsingular configurations.[47] Topological rigidity analysis fails for protein conformations corresponding to singular kinematic configurations, where kinematic analysis gives correct decompositions.[47] A singular conformation occurs when two constraints or DoF locally become linearly dependent, which can not be detected by simple constraint counting. More importantly, our kinematic analysis can provide a basis for protein motions when constraints are progressively relaxed, i.e., when non-covalent bond-lengths and -angles are collectively allowed to vary. These motion modes are inaccessible to topological rigidity analysis. Thus, our approach provides a framework to analytically examine the collective and universal effect of, for example, hydrogen bonding patterns on protein motion by constraint relaxation, instead of the common practice of omitting selected hydrogen bonds (constraints) entirely. In the remainder, we distinguish between *topological rigidity* as counting on constraint graphs via the pebble game and *kinematic flexibility* as our new method.

Our study makes three main contributions. First, we establish that the hydrogen bonding pattern of a protein structure, together with its rotatable covalent bonds, encodes a hierarchy of protein *motions*. These orthogonal motion “modes” are ranked by a *collective, weighted*

relaxation of the hydrogen bond network. By contrast, topological rigidity can only provide a hierarchy of rigid and flexible *regions* from ranking and removing *individual* constraints.[27,41] Second, we show that this motion spectrum is remarkably universal and conserved across a large data set of 183 diverse, monomeric high-resolution protein structures, but simultaneously reveals fold-specific differences across four large data sets separated by fold-types α, β, $\alpha + \beta$, and α/β. This fold-specific, spatial hierarchy suggests that hydrogen bond networks and folds are designed to modulate protein structure *and* dynamics. Its conserved nature allows more robust flexibility and rigidity analysis, insensitive to energy cut-offs on individual constraints. Third, we derive a formal, qualitative expression for a mode-specific free energy, which displays a near constant regime where increasing energetic cost is compensated by elevated entropy. Thus, our kinematic analysis provides a conceptual model system for enthalpy–entropy compensation (EEC), with a readily accessible interpretation of the controversial phenomenon. We compute and visualize kinematic flexibility modes for adenylate kinase. The modes highly and increasingly correlate with molecular dynamics simulations and crystallographic B factors when constraints are increasingly relaxed. Collectively, our results signify a deep connection between the structure of the hydrogen bond network, protein conformational dynamics, and functional motions. Our kinematic flexibility analysis is implemented in our KGS software. Source code and data to support our analyses are available from https://github.com/ExcitedStates.

KINEMATIC FLEXIBILITY ANALYSIS OF PROTEINS

We start by deriving an expression relating infinitesimally small rotations of dihedral angles to changes in the geometry of all hydrogen bonds collectively. We model macromolecules as a kinematic linkage, with groups of atoms as rigid body vertices and rotatable covalent bonds as directed edges with a single degree of freedom $q \in \mathbb{S}^1$ (Figure 1, center; Figure 2). The linear, branched topology of monomeric molecules can be represented with a single kinematic spanning tree, rooted at an arbitrarily selected vertex (yellow). Covalent double bonds, peptide bonds, and dihedrals amenable to planarity are modeled as rigid. Proline and aromatic rings are also modeled rigid, identical to standard topological approaches.[20,24] Groups of atoms without internal DoF are joined into rigid bodies. Dihedral angles of remaining single-covalent bonds are the DoF of the molecule, connecting two neighboring vertices (Figure 2).

In contrast to topological rigidity, we distinguish between covalent DoF and noncovalent constraints such as hydrogen bonds. Without constraints (red, Figure 1 center), our model corresponds to a serial, open kinematic chain, i.e., a kinematic linkage without closed loops. Noncovalent interactions form closed kinematic cycles that coordinate dihedral motion within two branches leaving from a common anchor vertex (Figure 2A), reducing the number of independent DoF. Multichain proteins can be treated within the same framework by connecting chains through interchain constraints.[48,49] Like topological rigidity theory, there are no interatomic *forces*. We encode hydrogen bonds as 5-fold holonomic constraints, i.e., constraints that only depend on the vector $\mathbf{q} = (q_1, \ldots, q_n)$, where n is the number of DoF. The 5-fold constraints permit a rotation ω_h about the bond axis but constrain all other relative motion. More precisely, the distance vector $\mathbf{f}_H - \mathbf{f}_A$ between the hydrogen (H) and acceptor (A) atom located at $\mathbf{f}_H$ and $\mathbf{f}_A$, the donor–hydrogen–acceptor (D–H–A) angle α =

$\alpha(\mathbf{f}_D, \mathbf{f}_H, \mathbf{f}_A)$, and the hydrogen–acceptor–base (H–A–AA) angle $\beta = \beta(\mathbf{f}_H, \mathbf{f}_A, \mathbf{f}_{AA})$ are constrained (Figure 2B).[47] This corresponds to a constraint with five bars in pebble game algorithms but explicitly specifies its kinematics. In our Kino-Geometric Sampling (KGS) software, salt bridges and disulfide bonds can be equally modeled with 5-fold constraints, but are not considered here. Likewise, other noncovalent interactions such as hydrophobic forces could be accommodated in the same framework through a 2-fold hydrophobic constraint. For a molecule with overall n dihedral DoF, we can distinguish between free DoF $\mathbf{q}_f \in \mathbb{T}^f$ that do not appear in any closed kinematic cycle and constrained, cycle DoF $\mathbf{q} \in \mathbb{T}^d$ where $f + d = n$. Free DoF can occur in side chains that do not accept or donate hydrogen bonds and are therefore not subject to kinematic constraints. They are not considered in the constraint analysis. The m hydrogen bonds define a constraint variety $\mathcal{Q}$, i.e., the roots of the algebraic equations $\mathbf{\Phi}$, on the cycle DoF

$$\mathcal{Q} = \left\{ \mathbf{q} \in \mathbb{T}^d \mid \mathbf{\Phi}(\mathbf{q}) = \mathbf{0} \in \mathbb{R}^{5m} \right\} \quad (1)$$

where

$$\mathbf{\Phi}(\mathbf{q}) = \left(\begin{bmatrix} \left(\mathbf{f}_{H,i}(\mathbf{q}) - \mathbf{f}_{A,i}(\mathbf{q})\right)^T, \alpha_i(\mathbf{q}), \beta_i(\mathbf{q}) \end{bmatrix}^T \right) - \left(\begin{bmatrix} \left(\mathbf{f}_{H,i}(\mathbf{q}^0) - \mathbf{f}_{A,i}(\mathbf{q}^0)\right)^T, \alpha_i(\mathbf{q}^0), \beta_i(\mathbf{q}^0) \end{bmatrix}^T \right) \quad (2)$$

where $[(\mathbf{f}_{H,i}(\mathbf{q}^0) - \mathbf{f}_{A,i}(\mathbf{q}^0))^T, \alpha_i(\mathbf{q}^0), \beta_i(\mathbf{q}^0)]^T$ denotes the geometry of hydrogen bond i, $i = 1, \ldots, m$ obtained from the input coordinate file with conformation $\mathbf{q}^0$.

Formally differentiating with respect to time yields a linear relationship between instantaneous changes in the DoF ("velocities") and corresponding changes in the constraints:

$$\frac{d\mathbf{\Phi}}{dt} = \mathbf{J}\dot{\mathbf{q}} \begin{cases} = \mathbf{0}, & \text{if } \dot{\mathbf{q}} \text{ constraint observing} \\ \neq \mathbf{0}, & \text{if } \dot{\mathbf{q}} \text{ constraint perturbing} \end{cases} \quad (3)$$

$\mathbf{J} \in \mathbb{R}^{5m \times d}$ is known as the constraint Jacobian matrix. Equation 3 characterizes two disjoint, orthogonal subspaces corresponding to velocities that observe constraints and those that perturb constraints. In practice, derivatives of forward kinematic atom positions $\mathbf{f}$ with respect to DoF q_k can be efficiently calculated using the cross-product

$$\frac{\partial \mathbf{f}}{\partial q_k} = \mathbf{r}_k \times (\mathbf{f} - \mathbf{p}_{k-1}) \quad (4)$$

for the rotatable covalent bond between atoms located at $\mathbf{p}_{k-1}$, $\mathbf{p}_k$ with axis of rotation $\mathbf{r}_k = (\mathbf{p}_k - \mathbf{p}_{k-1})/\|\mathbf{p}_k - \mathbf{p}_{k-1}\|$ (Figure 2C).

Constraint observing velocities lie in the nullspace null $(\mathbf{J}(\mathbf{q})) = \left\{\dot{\mathbf{q}} \in \mathbb{T}^d : \mathbf{J}\dot{\mathbf{q}} = \mathbf{0}\right\}$. This matrix subspace emerges from linearly dependent rows (constraints) or columns (DoF) in **J** and is generally of dimension $d - r$, with r as the rank of **J**. In contrast, velocities that do not comply with constraints lie in the range of **J**. Note that $r \leq p = \min(d, 5m)$ due to the rectangular shape of **J**. The singular value decomposition (SVD)[50]

$$\mathbf{JV} = \mathbf{U}\,\mathbf{\Sigma} \quad (5)$$

Where $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_{5m}] \in \mathbb{R}^{5m \times 5m}$ and $\mathbf{V}[\mathbf{v}_1, \ldots, \mathbf{v}_d] = \in \mathbb{R}^{d \times d}$, provides orthonormal bases for the range and nullspace of **J**. The rectangular matrix $\sum = \mathrm{diag}(\sigma_1, \ldots, \sigma_p) \in \mathbb{R}^{5m \times d}$ contains the singular values $\boldsymbol{\sigma}$ on the diagonal, where $\sigma_1 \geq \ldots \geq \sigma_r > \sigma_{r+1} = \ldots = \sigma_p = 0$, and corresponding $\mathbf{u}_i$ and $\mathbf{v}_i$ are the ith left and right singular vector, respectively. Let

$$\begin{aligned} \mathrm{range}\left(\mathbf{J}^{\mathrm{T}}\right) &= \mathrm{span}\left\{\mathbf{v}_1, \ldots, \mathbf{v}_r\right\} : \mathbf{R} \\ \mathrm{null}\left(\mathbf{J}\right) &= \mathrm{span}\left\{\mathbf{v}_{r+1}, \ldots, \mathbf{v}_d\right\} : \mathbf{N} \end{aligned} \quad (6)$$

Then, any $\dot{\mathbf{q}} = [\mathbf{R}, \mathbf{N}][\nu_R^{\mathrm{T}}, \nu_N^{\mathrm{T}}]^{\mathrm{T}}$ can be expressed with $\boldsymbol{\nu}_i$ the proportion of velocity pointing along vector $\mathbf{v}_i$, respectively. For a broader introduction to these linear algebra concepts we refer the reader to the literature.[50]

Since $\mathbf{u}_i$ and $\mathbf{v}_i$ are orthonormal vectors, σ_i encodes nonorthogonality between **J** and **V**; the singular value is the magnitude of constraint perturbation or constraint relaxation dictated by moving along right singular (velocity) vector $\mathbf{v}_i$. In molecular terms, right singular vectors $\mathbf{v}_i$ have the dimension of the number of degrees of freedom d (rotatable covalent bonds): $\mathbf{v}_i$ denotes a set of coupled motions in dihedral space. Left singular vectors $\mathbf{u}_i$ have the dimension of the number of hydrogen bond constraints: each element of $\mathbf{u}_i$ represents either a change in length (x, y, or z) or a change in angle (α, β) of a hydrogen bond. Thus, the expression $\mathbf{J}\mathbf{v}_i = \sigma_i \mathbf{u}_i$ relates a change in molecular conformation $\mathbf{v}_i$ to a change in the geometry $\mathbf{u}_i$ of all hydrogen bonds. If $\sigma_i = 0$, the vector of coupled motions $\mathbf{v}_i$ is in the nullspace of **J** and no change in the hydrogen bond geometry is observed. For $\sigma_i > 0$, coupled motions $\mathbf{v}_i$ result in progressively greater changes in the geometry $\mathbf{u}_i$ of all hydrogen bonds as σ increases ("constraint relaxation"). The singular value $\sigma_i > 0$ denotes the overall magnitude of constraint relaxation in the left singular vector $\mathbf{u}_i$, $i = 1, \ldots, r$ when moving along right singular vector $\mathbf{v}_i$, $i = 1, \ldots, r$.

It follows that null(**J**) spans all constraint-observing motions. As shown previously, these modes are identical to floppy modes from topological rigidity approaches,[47] i.e., they are the only remaining internal DoF if no constraint relaxation is permitted. Without constraint relaxation, dihedrals not part of these floppy modes are rigidified and merge adjacent rigid bodies, leading to a rigid cluster decomposition of the molecule, which is a central result from topological rigidity analysis. Note that free DoF $\mathbf{q}_f$ always comply with constraints, as they are outside kinematic cycles. The nullspace is highly sensitive to the set of noncovalent

constraints and careful tuning of energy thresholds, e.g., for the inclusion of hydrogen bonds in protein and RNA, is crucial to prevent over-rigidification.[30]

Constraint relaxation and kinematic flexibility modes.

The range **R** provides a spectrum of motions ranked by increasing constraint relaxation. We term those motions *kinematic flexibility modes*. They are not accessible in topological rigidity analysis, since they do not respect constraints but can potentially inform on hydrogen bonds that need to adapt their configuration during functional rearrangements or exchange with solvent. Thus, the full range of protein conformational dynamics encoded by the hydrogen bonding pattern is accessible through orthonormal bases spanning both matrix subspaces. We therefore refer to the SVD (5) of the constraint Jacobian obtained from the hydrogen bonding pattern as the *hydrogen bond network spectral decomposition*. While ENMs provide a spectrum in terms of eigenfrequencies and corresponding eigenmodes, our spectral variable σ denotes the norm of constraint relaxation necessary to access the corresponding kinematic flexibility mode.

RESULTS

We first provide a graphical interpretation of our kinematic flexibility analysis on two secondary structure example systems, before we perform a more theoretical, large-scale analysis of fold-specific kinematic motion spectra found in the protein universe. A subsequent computational study on adenylate kinase (ADK) demonstrates how hierarchical kinematic motion modes efficiently probe protein conformational dynamics.

Kinematic Flexibility Modes of α-Helices and β-Sheets.

We analyzed how the hydrogen bond network spectral decomposition imposes a range of collective motions on proteins. To visualize motions across the spectrum and their effect on hydrogen bonds, we analyzed an α-helix and a β-sheet in detail. The top panels of Figure 3 show the perturbation matrix **P** resulting from the matrix product **JV**, summed over the five constraints per hydrogen bond

$$P_{i,d-j-1} = \sqrt{\sum_{k=5i}^{5i+4} \left(J_{kl} V_{lj}\right)^2}, i = 0, \ldots, m-1, j = 0, \ldots, d-1 \quad (7)$$

P graphically displays eq 5, visualizing the magnitude of collective perturbations to hydrogen bond geometries when constraints are relaxed in successive kinematic flexibility modes. Rows correspond to individual hydrogen bonds ordered by first appearance along the protein sequence, irrespective of donor or acceptor, while columns correspond to kinematic flexibility modes $\mathbf{v} \in \mathbf{V}$ ranked by singular values, increasing from left to right. The predicted perturbation of individual hydrogen bonds (rows) for instantaneous motions along each kinematic flexibility mode (columns), i.e., values in **P**, is color-coded, increasing from dark to light. Note that the magnitude of the *collective* perturbation over all hydrogen bonds i in each column j, i.e., $\sqrt{\sum_i P_{ij}^2}$, equals the singular value σ_j associated with this motion

mode. For the α-helix (A), we observe a striking, alternating pattern of constraint perturbation, suggesting motions of the dihedral DoF are modulated by concerted hydrogen bond fluctuations. We observe a similar, but weaker pattern for the β-sheet (B).

We then projected the conformational change $\Delta_{q,j}$ for degree of freedom j corresponding to selected kinematic flexibility modes $\mathbf{v}_i$ onto the helix and sheet (Figure 3, bottom panels). This conformational change follows from taking a small step δ along the direction of each right singular vector

$$\mathbf{\Delta}_{\mathbf{q},i} = \delta \mathbf{v}_i \quad (8)$$

The vectors $\mathbf{v}_i$ are all unit length, and therefore, $\|\mathbf{\Delta}_{\mathrm{q},i}\| = \delta$, $i = 1, \ldots, d$.

The left-most, purple-colored set of modes in the matrix correspond to local fluctuations (e.g., A11, B5, indices label column numbers), affecting only DoF close to individual hydrogen bonds shown as dots. As the mode number increases from left to right, DoF are increasingly engaged. For example, in superhelical twisting (e.g., A38) and straight bending (e.g., A46) compensating motions of DoF, indicated by alternating blue–red patterns, mediate moderate constraint perturbation. By contrast, high constraint perturbations found in compression/tension (A54/A55) are a result of changes in DoF that reinforce each other. Interestingly, unraveling (e.g., A19) demonstrates significantly less overall perturbation than compression/tension (A54/A55), indicating an energetically favorable mode for helix dissociation. Similarly, the structure of β-sheets permits twisting and bending motions (e.g., B11), while shear (e.g., B33) and especially widening/narrowing (e.g., B38) requires substantial distortion of hydrogen bonds. These findings agree with molecular dynamics simulations, which revealed high strength of β-sheet proteins in response to shear loading, with an elastic modulus of about 240 pN/Å.[51] Also, the simulation results predicted an initial linear elastic regime during tensile loading of α-helical protein domains. Beyond the elastic regime the helix uncoils, releasing one turn at a time. It is worth noting that there are no nullspace floppy modes along the backbone in either structure, i.e., topological rigidity analysis predicts only flexibility of dangling side-chains, but full backbone rigidity.

Besides the graphical interpretation, these results hint at three important characteristics of the motion spectrum. First, the pattern of hydrogen bond perturbations suggest distinct motion regimes, separated by allowed constraint relaxation and collectivity, i.e., how much of the structure is engaged in each mode. Second, motions hierarchically ranked by constraint relaxation seem to correlate significantly with energy required to perturb a structure. Third, the hydrogen bonding networks encode fold-specific motions. We examine these attributes quantitatively in the next section.

Kinematic Flexibility Modes Hierarchically Rank Collective Hydrogen Bond Energy Perturbation.

To establish a predictive relationship between singular values and the *collective* variation in protein hydrogen bond geometry and energy, we monitored geometry and energy changes while performing conformational changes via (8). Changes in hydrogen bond geometry

follow from $\|\mathbf{\Phi}(\mathbf{\Delta}_{q,i})\|$ (1), while energies $\|\mathbf{\Delta}\,\mathbf{E}_{HB}(\mathbf{\Delta}_{q,i})\|$ are evaluated using the Mayo energy-potential[52] for each individual hydrogen bond

$$E_{HB} = w_0\left\{5\left(\frac{l_0}{l}\right)^{12} - 6\left(\frac{l_0}{l}\right)^{10}\right\} f(\alpha,\beta,\gamma) \quad (9)$$

with well-depth w_0 = 8 kcal/mol, equilibrium distance l_0 = 2.8 Å, hydrogen bond donor–acceptor distance $l = \|\mathbf{f}_D - \mathbf{f}_A\|$, and angular terms $f(\alpha, \beta, \gamma)$ (Figure 2B) that depend on the hybridization state of donor and acceptor (Supporting Information (SI), eq 1). Values of E_{HB} then range between −8 kcal/mol for strong and 0 for weak hydrogen bonds.

We illustrate our findings using the crystal structure of 189-residue human DJ-1 protein (PDB ID 1p5f). Figure 4A shows monotonically increasing singular values (blue) which rank-order the kinematic motion modes along the horizontal axis. The magnitudes of the corresponding collective geometric perturbations (magenta) and hydrogen bond energies (red) are shown for $\delta = 10^{-5}$. The three curves clearly follow similar trends. Strikingly, while our kinematic flexibility analysis is informed by only individual, geometric hydrogen bond information, singular values of the hydrogen bond network spectral decomposition rank-order protein motions hierarchically by collective (constraint) geometric relaxation and associated energy penalty of the entire constraint network. This suggests that the network is designed to selectively favor certain directions of collective motion in conformation space over other motions. Except for the kinematic flexibility modes at both ends of the spectrum, changes in constraint geometry and energy follow a log–linear regime. Modes outside the log–linear regime may correspond to nonfunctional protein dynamics. For example, modes with the largest penalties correspond to unfolding motions where tertiary and secondary elements lose structure, as graphically displayed in bottom panels of Figure 3. Modes in the nullspace, i.e., floppy modes to the left of the indicated nullspace limit, carry no geometric penalty.

Next, we examined if the change in internal (hydrogen) bond energy could be predicted from the spectral decomposition. Note that the norm of constraint penalty and the singular values σ_i, when taking a small step along a right singular vector are related by a scale factor $c_V = \delta = 10^{-5}$, i.e., $\|\mathbf{J}\mathbf{\Delta}_{q,i}\| = c_V\sigma_i$ (combining (5) and (8)). Fitting the singular values curve (blue) to the magnitude of geometric penalties of hydrogen bonds $\|\mathbf{\Phi}(\mathbf{\Delta}_{q,i})\| = c_G\sigma_i$ (1), we found a scale factor of $c_G = 10^{(-5.02\pm0.05)} \approx \delta$, indicating that linearization has negligible effects at small step sizes. For the highly nonlinear energy function ((9) and SI eq 1), we found $c_E = 10^{(-4.51\pm0.17)}$ kcal/mol ≈ 3.09 δ kcal/mol. Hence, for a sufficiently small step size δ, the hydrogen bond network spectral decomposition predicts the collective hydrogen bond energy penalty along mode i by $\|\mathbf{\Delta}\boldsymbol{E}_{HB}(\mathbf{\Delta}_{q,i})\| = c_E\sigma_i$ for this example structure. For floppy modes inside the nullspace, the difference between singular value prediction and energetic cost is larger. This can be explained by sp2–sp2 hybridized hydrogen bonds that observe small energetic changes due to a torsional term related to angle γ in the hydrogen bond energy function (SI), which is not present in the geometric constraint formulation (Figure

2B). The torsional term becomes more relevant if other terms such as distance and angles remain small, which is the case inside the nullspace.

This hierarchy of protein motions is remarkably conserved in the protein universe. We compiled a diverse benchmark data set of 183 high-resolution, nonredundant crystal structures ranging in size from 30 to 555 amino acids in length from the PDB. We observed a strong linear relationship between the number of cycle DoF d and the number of constraints $5m$ in the structures (Figure 4B). Table 1 evaluates statistical properties of the data set, with the hydrogen bond energy (9) summed up over all hydrogen bonds. The fraction of hydrogen bonds per residue agrees well with the literature.[53]

Similar to our single example crystal structure of human DJ-1 protein, singular values spanned many orders of magnitude. We also observed a large range in the number and distribution of vanishing singular values across the crystal structures, a median of 34 nullspace floppy modes, with a maximum of 148 and one (engineered) protein with zero floppy modes (PDB ID 5eca). While most structures have slightly more constraints than cycle DoF, i.e., $5m > d$, it appears that a fairly constant fraction of constraints is linearly independent, leading to a linear increase in the number of floppy modes over protein size (Figure 4B). This was also observed previously.[33] We therefore analyzed kinematic flexibility modes corresponding to non-vanishing singular values, i.e., outside of the nullspace, for all 183 crystal structures (Figure 4C). For comparison we normalized modes, and we grouped modes into 50 bins per structure. Repeating our analysis above, we obtained a surprisingly universal law

$$\left(\mathbf{E}_{\mathrm{HB}}^{\mathbf{q}}\right)' = \frac{1}{\delta}\left(\| \Delta \mathbf{E}_{\mathrm{HB}}\left(\vdots \atop \mathbf{\Delta}_{\mathbf{q},i}\right) \| \right) = \bar{c}_E \boldsymbol{\sigma} \text{ with } \bar{c}_E \approx 3.24 \text{ kcal / mol} \quad (10)$$

from fitting the mean curves, where $\bar{c}_E$ is *independent* of mode number, with error on the same order as in the single structure before. This suggests that protein hydrogen bonding patterns impart a distribution of orthogonal, coordinated motions on the DoF. Kinematic flexibility modes identify preferred directions of deformation for the protein in (hydrogen bond) internal energy landscapes. Note that $\left(\mathbf{E}_{\mathrm{HB}}^{\mathbf{q}}\right)'$ reports on the *collective* energy change over all hydrogen bonds for individual modes; motions along kinematic flexibility modes i do not necessarily increase all hydrogen bond energies. Instead, for each direction, fluctuations in hydrogen bond energies can provide compensatory mechanisms, i.e., many could marginally reduce in energy to allow a handful to significantly increase. Initial constrained minimization of the data set ensures starting structures near a local minimum of hydrogen bond energy.

Spectral Distribution of the Hydrogen Bond Network.

To obtain a physical understanding of the distribution of modes across protein structure, we performed a full spectral decomposition and computed probability distributions for singular values on our benchmark data set. The size distribution of protein structures is reflected in the SVD of associated constraint Jacobian matrices **J**. It is well-known mathematically that

the maximum singular values monotonically increase by adding a column (a degree of freedom), while the smallest nonzero singular values decrease.[50] Physically, a larger protein with more DoF can be interpreted by a larger lever arm that allows a larger maximum constraint perturbation, or simultaneously provides more motion capabilities to avoid perturbation. Adding a row (constraint) will increase the minimum singular value, i.e., there are more constraints to be perturbed, and the protein is rigidified. Overall, the spectral range becomes size-dependent (Figure 4B), which can also be seen in the widened distributions near the spectral limits in Figure 4C. Hence, to compare mode densities across differently sized proteins, we normalized singular values of each structure by their maximum σ_1 to obtain $\sigma \in [0, 1]$ for all modes across all structures and grouped them by σ into ten bins per decimal power.

Singular values for the 183 single-chain proteins span several orders of magnitude and show well-conserved, sharp peaks near integer exponents (Figure 5A). The two most common kinematic flexibility modes occur at $\sigma \approx 10^{-2}$ and $\sigma \approx 10^{-3}$. The peak at lowest σ values (Figure 5A, far left) represents floppy modes, which are the only modes considered in topological rigidity analysis. Their density is least conserved across the spectrum and shows a linear increase over protein size (Figure 4B). Thus, while normalizing by σ_1 helps spectral comparison in the regime $\sigma > 0$, the variability in the number of floppy modes where $\sigma = 0$ remains. Note that we pick a threshold $\sigma_N = 10^{-12}$ to numerically identify vanishing singular values. All modes with $\sigma \leq \sigma_N$ are floppy modes and represented by this peak. There is a clear gap of several orders of magnitude between nonzero and vanishing singular values, which allows correct and simple identification of floppy modes.[47]

The spectral distribution relates directly to the stiffness of the protein. Formally, this follows from defining a cumulative perturbation

$$p_c = \int_\sigma p(\bar{\sigma})\bar{\sigma}\, \mathrm{d}\bar{\sigma} \quad (11)$$

for perturbation-specific probability densities $\rho(\sigma)$ following the spectral distribution, assuming individual modes are enabled at equal probability. Thus, proteins enriched in high-perturbation modes require overall more energy to access their motion modes, rendering them stiffer. Discrete jumps between spectral peaks suggest that modes are distributed across different stiffness regimes. Interestingly, atomic force microscopy (AFM) on single antibody proteins also found two distinct elastic regions with an ~4-fold increase in stiffness between a low- and high-strain regime, before plastic deformation sets in.[54] Again, these experimental findings agree well with our perturbation analysis.

The geometry of hydrogen bonds in our analysis critically depends on accurately placed hydrogen atoms in structure preparation. To exclude the possibility of bias at that stage we repeated our analysis using a smaller set of 34 structures from the PDB with experimentally determined hydrogen atom positions from neutron diffraction experiments (SI). Its spectral signature, i.e., peak locations and heights, was virtually identical, validating the results from our initial data set of 183 high-resolution single-chain structures for subsequent analysis.

Furthermore, spectral analysis of singular values from a set of random matrices with the same dimensions as the original protein data set produced a completely different distribution, while a set of random matrices with the same sparsity pattern led to similar distributions (SI Figure S1). A control study with the same number of 5-fold constraints between randomly selected atoms instead of real hydrogen bonds maintained some of the features but generally showed less conservation and higher stiffness. This clearly demonstrates that the hydrogen bonding network, together with the covalent kinematic structure of the protein, stores this fold-specific dynamic information.

Motions from the Hydrogen Bonding Pattern Are Spatially Distributed.

Besides distinct perturbation levels, we observed various levels of mode engagement, i.e., how much of the molecule is involved in a specific motion mode. To measure this collectivity s of mode $\mathbf{v}_i$, we compute the exponential of the Shannon entropy of its squared components[14,55]

$$s_i = \frac{1}{d}\exp\left(-\sum_{j=1}^{d}\kappa_{i,j}\log(\kappa_{i,j})\right) \quad (12)$$

where $\kappa_{i,j} = v_{i,j}^2 / \sum_{j=1}^{d} v_{i,j}^2$. This normalization of $\mathbf{v}$ is trivial for singular vectors, since they are unit length by definition, but nontrivial for eigenmodes from ENM or NMA. The second normalization by the number of modes d (equals the length of $\mathbf{v}$) reduces size differences across protein structures, which allows $s_i \in [0, 1]$ to be interpreted as the fraction of significant contributors in motion mode i.

Figure 5B plots collectivity computed over modes with similar singular values, grouped into ten bins per decimal power. Nullspace floppy modes with zero perturbation at the lower end of the spectrum show a relatively low collectivity compared to medium or high-perturbation modes. From panels A and B together, we observe that the most collective motion modes are also most abundant. Collectivity, and thus entropy, follows a near-exponential regime between 10^{-7} and 10^{-3}, showing a similar trend as energetic perturbations in Figure 4C. Finally, modes with strongest constraint perturbation are less collective than medium-perturbation modes.

Protein Fold Classes Have a Unique H-Bond Network Spectral Signature.

To analyze protein-fold specific differences encoded by the hydrogen bond pattern, we examined the hydrogen bond spectrum and collectivity of kinematic flexibility modes for four separate data sets of α-only, β-only, $\alpha + \beta$, and α/β proteins, ranging from 655 to 1051 structures (SI). Figure 5 shows the mean curves for the spectrum (C) and collectivity (D), corresponding to the black mean curves of the original mixed-fold data set in parts A and B, respectively. Percentiles follow similar patterns as for the previous data set and are omitted for visibility. While the location of spectral peaks is conserved across folds, the peak heights show clear differences. The class of α-only folds has more modes at higher singular values, yielding them stiffer to perturbations (11). In other words, the constraint network requires

more relaxation to access these modes. By contrast, β-only proteins are enriched in smaller spectral modes, rendering them more flexible, while mixed folds $\alpha + \beta$ and α/β show intermediate spectra. The density of zero-perturbation floppy modes with vanishing singular values is highly size-dependent and less informative regarding fold content. Nonetheless, our results predict more floppy modes in α-helical folds than β-folds, which can most likely be attributed to less interhelical connectivity compared to inter-β-strand connectivity.

Collectivity, computed as before (12), is generally higher for β-folds than α-folds, and intermediate for mixed folds. Interestingly, medium-perturbation modes in β-dominated folds are more collective than the highest-perturbation modes, indicating localized hinges, e.g. in β-turns, that are able to significantly unfold the protein (cf. Figure 3B40). Note that floppy modes with zero-perturbation, the only accessible modes from topological rigidity analysis, show relatively small and very similar collectivity across folds, rendering them more localized compared to medium-perturbation modes and again less informative regarding fold content.

To further evaluate predictive capabilities of our method, we analyzed the spectral distribution of a set of four hyperstable, designed peptides,[56] each with an NMR bundle of 20 distinct structures (PDB codes 2nd2, 2nd3, 5jhi, 5ji4; details in the SI). Their spectrum shows a clear shift toward modes with higher relative perturbation compared to the high-resolution data set (Figure 6) and an ~3-fold increase in p_c (11), identifying their designed constraint pattern as a key contributor to increased stability. Collectivity of modes shows trends as the other data sets.

Free Energy of Modes.

The hydrogen bond energy perturbations predicted by the magnitude of singular values can be formally combined with the conformational entropy contributions encoded by collectivity into a dimensionless expression for free-energy changes ΔF related to each mode i

$$\mathbf{\Delta} F_i = \sigma_i - c_T s_i \quad (13)$$

with a dimensionless temperature factor c_T. For normalized singular values and normalized collectivities, the range of ΔF is between −1 and +1 when $c_T = 1$. Clearly, (13) is a formal abstraction, since only hydrogen bond energy contributes to internal energy. Surprisingly, though, our free-energy changes demonstrate how enthalpic and entropic contributions compensate each other in the overall spectrum of conformational motion. Figure 7 depicts ΔF over normalized motion modes computed from all 183 structures in the high-resolution data set. The red curve averages over individual modes grouped into 100 bins. ΔF roughly levels for about 40% of modes (Figure 7, modes between dashed lines) with most favorable entropy (collectivity) and medium enthalpic cost. Modes to the right of the interval have unfavorably high enthalpic cost; they correspond to unfolding (high-perturbation modes in Figure 3). Modes to the left of the interval are more localized, with smaller enthalpic cost, but simultaneously less entropic benefit. Nullspace floppy modes are excluded in the graph; their enthalpic cost is zero, as encoded by vanishing singular values. Thus, associated free

energy is identical to their collectivity as depicted in Figure 5B, which is on average $\Delta F = -0.05$ and thus turns out less favorable than the free energy plateau at $\Delta F \approx -0.2$.

Comparison with iMOD.

We compared the spectrum and collectivity of motions obtained from kinematic flexibility analysis with normal modes from iMOD.[4] iMOD is a versatile ENM-based tool to study normal modes of macromolecules in internal coordinates, i.e., dihedral angles. Although iMOD's mass matrix is based on a full atom representation, flexibility is limited to main-chain and χ_1 angles, while our model maintains full side-chain flexibility. We carried out iMOD simulations with the command line settings $-n$ 10000 to force computation of all normal modes, $-x$ to enable the χ_1 degree of freedom, and otherwise default parameters. Comparison of reduced ENM models to full NMA previously revealed that agreement of low-frequency modes is conserved but that higher-frequency modes can differ significantly.[57] Figure 8 shows the iMOD spectrum of eigenfrequencies (A) and collectivity of eigenmodes (B) for our 183 single-chain protein data set. Although the spectral variables in ENM and our approach are different, this comparison sheds light on general characteristics of normal modes, rigidity-theory based and kinematic flexibility modes. Similarly to our spectrum, eigenfrequency distributions are broadly conserved across structures, confirming results from previous NMA and ENM analysis.[11,57] Modes with low to medium eigenfrequency are most abundant. Similar to kinematic flexibility, the most abundant modes are also most collective. Higher-frequency motions with $\omega > 1000$ cm^{-1} cannot be obtained with this coarse-grained model and require full NMA. Floppy (no-relaxation) modes in topological rigidity and our kinematic flexibility analysis correspond to zero-frequency modes in ENM,[27] i.e., modes with vanishing energetic cost. However, zero-frequency modes in ENM are often considered an artifact, as the network breaks down into multiple independent ones. These modes are often avoided by increasing the number of weak interactions, for example by increasing cutoff distances.[13] Floppy motions from topological rigidity therefore cannot be directly accessed with ENM.

Floppy modes from topological rigidity analysis are less collective, i.e., correspond to more local motions compared to low-frequency normal modes or medium constraint relaxation kinematic flexibility modes (compare Figure 5B to Figure 8B). This signifies that topological rigidity theory based methods tend to overestimate molecular rigidity and study conformational flexibility based on more local motions than ENM or NMA.

When we compared fold-specific spectra (Figure 8C) and collectivity (Figure 8D), we observed several important differences between the methods. For example, the location and height of the main spectral peak in iMOD is indistinguishable for α-only and β-only and shifted slightly to higher frequencies for mixed folds, while kinematic flexibility analysis shows stronger differences in peak heights across fold types. Interestingly, α-folds show slightly increased density at very low frequencies in iMOD, predicting lower stiffness compared to other folds, in contrast to other established methods. Our kinematic flexibility analysis, detailed NMA,[10,57] and analysis of force–displacement curves from MD simulations[51,58] all predict that α-folds are stiffer. However, iMOD predicts lower collectivity in α-folds, similar to our approach.

A more detailed comparison between the methods, e.g., analyzing direct mode overlap, remains difficult, since iMOD is limited to main-chain and χ_1 dihedral motion, while our kinematic model maintains full side-chain flexibility, thus demanding lower-dimensional projections. Other ENM or NMA tools are mostly formulated in Cartesian coordinates or are unavailable for other users, making a comparison even more difficult. Nevertheless, our results reveal similarities and differences important for users of ENM, topological rigidity, and our new kinematic analysis.

Kinematic Flexibility Modes of Adenylate Kinase.

To illustrate the molecular motions encoded by the hydrogen bond spectral decomposition, we analyzed adenylate kinase (ADK). ADK is a ubiquitous phosphotransferase that plays a central role in maintaining adenosine triphosphate (ATP) levels in the cell.[59] It is a model system that has been extensively studied, both experimentally and computationally.[60–69] ADK consists of a large CORE domain and two binding domains, the LID domain that binds ATP, and the NMP domain (Figure 9A). The two binding domains open and close during its catalytic cycle.

We examined how removing individual constraints using $E_{\mathrm{HB,cut}}$ and collectively relaxing the constraint network using σ_{cut} modulates flexibility of ADK. We used our KGS software framework to randomly sample conformations from kinematic flexibility modes, using a rapidly exploring random tree (RRT,[68,70] SI; conformational ensembles available at https://doi.org/10.5281/zenodo.1315003). We explored the effects of different constraint network relaxations, corresponding to singular value thresholds $\sigma_{\mathrm{cut}} = \{10^{-10}, 10^{-4}, 10^{-3}, 10^{-2}\}$. At each conformation $\mathbf{q}$, we determined the subspace $\overline{\mathbf{V}}_q := \mathrm{span}\{\mathbf{v}_i, \ldots, \mathbf{v}_d\}$ such that $\sigma_i \leq \sigma_{\mathrm{cut}}$, i.e., we identify the subspace of collective motions in dihedral space that limit the magnitude of constraint relaxation up to the level dictated by singular value σ_{cut}. For each trial random step $\boldsymbol{\delta}_q \in \mathbb{T}^d$ generated by the RRT, we obtained a step $\boldsymbol{\Delta}_q \in \overline{\mathbf{V}}_q$ consistent with the specified constraint relaxation σ_{cut} by projecting $\boldsymbol{\delta}_q$ onto the subspace $\overline{\mathbf{V}}_q$

$$\boldsymbol{\Delta}_q = \overline{\mathbf{V}}_q \overline{\mathbf{V}}_q^{\mathrm{T}} \boldsymbol{\delta}_q \quad (14)$$

Free DoF $\mathbf{q}_f$ are unaffected by the projection and can be perturbed randomly. We ensured individual entries in $\boldsymbol{\Delta}_q \leq 0.05$ rad. A new molecular conformation was obtained as $\mathbf{q} + \boldsymbol{\Delta}_q$ and accepted if there is no steric overlap between pairwise atoms (van der Waals radii were scaled to 0.75). KGS uses spatial hashing to quickly identify neighboring atoms for efficient collision detection.

We added hydrogens and performed energy minimization of the open conformation of ADK (PDB ID 4ake[64]) with Maestro and identified hydrogen bonds with KGS at four different $E_{\mathrm{HB,cut}}$. For each $E_{\mathrm{HB,cut}}$, KGS randomly generated 2000 samples at the four different thresholds σ_{cut}. Note that $\sigma_{\mathrm{cut}} = 10^{-10}$ corresponds to motions encoded by topological rigidity analysis, which permits no constraint relaxation. The normalized dimensions of sampling space $\dim\left(\overline{\mathbf{V}}_q^0\right) \Big/ \dim\left(\mathbf{V}_q^0\right)$ computed at the initial conformation $\mathbf{q}^0$ indicate a strong

dependence on $E_{HB,cut}$, while differences vanish at $\sigma_{cut} \approx 10^{-2}$ (Figure 9B), demonstrating increased robustness of the geometric hierarchy of motions and to some extent invariance with respect to constraint input (Figure 9C). In other words, while the number of floppy modes with zero relaxation dramatically depends on the exact value of $E_{HB,cut}$, the number of modes for relaxation at $\sigma_{cut} \approx 10^{-2}$ is almost invariant to $E_{HB,cut}$. The graphs of Figure 9B simultaneously represent the cumulative distribution function of the spectrum of singular values. Finally, we computed per residue root mean squared fluctuations (RMSF, Figure 9D) and synthetic B factors (Figure 9E; SI) from heavy atom coordinates and compared them to RMSF from a 100 ns MD simulation (SI) and experimentally reported isotropic B factors in the PDB file.

Correlations are relatively weak for the (near) absence of constraint relaxation $\sigma_{cut} \leq 10^{-4}$. Only when all hydrogen bonds weaker than $E_{HB,cut} = -3.0$ kcal/mol were excluded did motion modes from topological rigidity reproduce functional motions (Figure 9D/E, top two panels). By contrast, motions with elevated constraint relaxation dramatically improve correlations even when the weakest hydrogen bonds are included (Pearson $r_{P,MD} \geq 0.82$; $r_{P,Bfac} \geq 0.7$). Thus, compared to using modes corresponding to those from topological rigidity analysis ($\sigma_{cut} \leq 10^{-10}$), slightly relaxing the constraint network consistently samples functional motions, near-independent of $E_{HB,cut}$. RMSF correlations between the KGS ensemble with constraint relaxation and experimental B factors are comparable to those of the MD ensemble ($r_{P,MDvsBfac} = 0.75$), underscoring the strength of our kinematic approach (Figure 9A). Interestingly, in view of free energy of modes (13), correlations appeared to improve when constraint relaxation was sufficiently high to access motion modes within the lowest free energy regime (cf. Figure 7), but sufficiently low to prevent unfolding motions to the right of this regime. Thresholds $\sigma_{cut} > 10^{-2}$ overstrain the constraint network and unfold the protein, leading to high rejection rates that made it impossible to generate 2000 samples in adequate time (additional analysis provided in SI, Figure S2). The computational complexity of instantaneous flexibility analysis is dominated by the SVD of $\mathbf{J}$, $\mathcal{O}\left(d^3\right)$ worst case. Instantaneous flexibility analysis completes in one to three seconds. Computation times to generate the 2000 samples ranged from 90 min to 14 h, increasing with higher σ_{cut} as the protein increasingly unfolds resulting in rejections due to steric overlap. Several improvements in numerical algorithms could improve speed, such as updating the SVD rather than resolving the full matrix, replacing the SVD with QR, or GPU-based implementations.

Overall, our new hierarchy of motion modes (Figure 9B) based on the SVD of the constraint Jacobian matrix demonstrates increased robustness toward constraint topology and performance when sampling functional motions, compared to pebble game rigidity, which requires careful tuning of the constraint topology via $E_{HB,cut}$ to capture meaningful protein dynamics.

DISCUSSION

Our new, kinematic approach to rigidity analysis treats hydrogen bonds as a geometric constraint network. Compared to topological rigidity, which admits no constraint relaxation, it extends analyses to constraint relaxing motions, providing a full spectral decomposition of

motion modes ranked by their collective perturbations of the hydrogen bond network. While analyses of the structural dynamics of proteins often implicitly assume that weak hydrogen bonds disrupt first, our approach does not require such assumptions. Instead, the network imparts a hierarchy of motions and stiffness regimes on the protein, which modulate the conformational response. The hydrogen bond spectral decomposition is (1) highly conserved in the protein universe and (2) reveals key fold-specific differences.

Kinematic flexibility analysis indicated that zero-perturbation floppy modes, the motions obtained from topological rigidity analysis, are more localized than low-frequency modes from ENM, suggesting an overly rigidified representation in topological rigidity or alternatively, overconnectivity in ENM. Therefore, strict rigidity theory based methods rely on exceedingly local motions to estimate conformational flexibility,[20,27,28] ligand binding,[31] entropy,[33,34] or thermo-stability[38–40] compared to ENM or NMA. Nonetheless, they show convincing agreement with experimental data, if energy cutoffs to determine constraints are carefully chosen. Moreover, kinematic floppy modes that observe hydrogen bond constraints have guided conformational transitions[36,68,71] and revealed coordinated loop motions,[72] often more successfully than normal mode based methods.[49,68]

Analyses of motions beyond the floppy modes, i.e., in the kinematic flexibility regime, revealed several unexpected insights. First, we found that hydrogen bond networks determine structure *and* modulate structural dynamics. This could have important implications for *de novo* protein design and folding[73] or hydrogels,[74] where recent attention is focused on designing hydrogen bonding patterns to create stable interfaces[75] and mediate specificity.[76,77] For example, our procedure revealed a clear shift toward stiffer modes for designed, hyperstable peptides.[56] While hydrogen-bond guided designs are often successful structurally, i.e., the crystal and predicted structure are near identical, it remains a challenge to design dynamic, functional proteins.[78] Our procedure could predict motion modes of hundreds of designed proteins and hydrogen bond networks in minutes.

Second, the conserved hierarchy and collectivity of protein motions revealed distinct motion regimes, which are often observed in experiments. For example, AFM on single antibody proteins[54] uncovered two elastic motion regimes, separated by a near 4-fold increase in stiffness, and a regime of plastic deformation. Our spectral decomposition structurally rationalizes the distribution of these motions under different strains. Low strain triggers low-perturbation modes that are low in collectivity and mostly locally perturb the structure. Higher strain engages highly collective modes toward stiffer motion regimes, explaining elevated stiffness in the second elastic regime measured with AFM. Interestingly, our EEC model suggests that a fraction of the energetic cost could be entropically balanced, rendering the deformation elastic and reversible.[54,79] Any strain beyond the elastic regime leads to plastic deformation, indicated by a sharp increase in hydrogen bond energy perturbation. Motion modes in this regime likely completely unfold the protein.

Third, the fold-specific differences we observe in our perturbation analysis suggest distinct functional roles of secondary structure, confirming previous simulations.[51,58] We found more modes at higher energy perturbation in α-helices than β-sheets, yielding helices stiffer. This is consistent with experimental data from low-frequency Raman spectroscopy[80] and

eigenfrequencies from detailed NMA.[10,57] Detailed NMA captures additional differences between folds in high-frequency regimes such as the amide I, II, or III bands.[57] These higher-frequency differences can also be detected experimentally with infrared spectroscopy and have been used to determine secondary structure content.[81,82] Simplified elastic network analysis using iMOD failed to identify this shift toward stiffer modes, highlighting a limitation in the simplified force-fields of most ENM. Our analysis also predicted that β-sheets are highly resistant to shear motions, which is confirmed by MD simulations.[51] Simulations on the mechanical response of silk crystalline units under shear loading showed high rupture forces due to efficient force distribution in the β-sheet structures.[83] Buehler and Keten further found an initial linear elastic regime during tensile loading of α-helical proteins, after which the helix unravels turn by turn.[51] Our analysis also predicted unraveling as the favored mode of helix dissociation.

We demonstrated how collective constraint relaxation improves conformational sampling in adenylate kinase (ADK). Previous work in topological rigidity analysis required careful tuning of energy cutoffs for individual constraints or employed ensemble-based rigidity analysis to correctly predict flexibility, such as the distance constraint model (DCM)[42] or constraint network analysis (CNA).[28] We eliminate this drawback by specifying admissible constraint relaxation on the entire network, such that weaker areas are automatically perturbed more. Our procedure consistently led to higher correlations with MD and crystallographic B factors. However, over-relaxation of constraints led to exceedingly numerous and severe steric clashes, suggesting a relatively sparse hydrogen bond network (≈0.7 h-bond/residue) guided protein motion on a hard-sphere potential “energy” landscape. The conserved nature of the associated hydrogen bond network spectral decomposition, compared to high fluctuations in the number of floppy modes without constraint relaxation, underlines the robustness of our approach.

Our method can provide quick insight into how rigid clusters, collective motions, and stiffnesses shift as constraints are added, increasingly relaxed, or fully removed. While we found convincing agreement with experiment and simulation, important limitations remain. Imperfect correlations between the ADK crystal structure and kinematic flexibility analysis could also be related to factors like crystal packing or idiosyncratic artifacts introduced by flash-cooling. Anticorrelated RMSF between kinematic flexibility analysis and MD simulations on ADK near residue 190 (Figure 9D) could be related to disruption of a hydrogen bond in the simulations. This suggests that our method can be improved by dynamically removing h-bonds if their strain exceeds a threshold. Here, we considered constraint relaxation of hydrogen bonds only to provide a direct comparison to the large body of work in topological rigidity theory, ignoring other noncovalent interactions. Bond lengths and angles are fixed and only torsional DoF contribute to molecular motion. While hydrogen bonds are important determinants of protein structure and dynamics, hydrophobic effects, electrostatics, solvent, or protein–protein interactions also modulate structural dynamics. The effects of these interactions could be explored with our method. For example, the role of hydration layers on protein dynamics is complex and remains poorly understood.[84,85] Neutron diffraction can reveal the position and orientation of hydrogen atoms in waters. Adding water-mediated hydrogen bonds can help understand how protein conformational fluctuations on the surface are linked to the motions of water, and propagate into protein

cavities or internal motions.[86] Our new, kinematic flexibility analysis is a versatile method, bridging topological rigidity and ENM. By way of a novel spectral decomposition of protein hydrogen bonding patterns, it provides explicit access to collective motions and free energy of modes, signifying that hydrogen bonds store intrinsic, fold-specific functional motions. These quantitative models and insights can help improve *de novo* protein design and folding, help to understand mechanobiology probed by AFM or single-molecule fluorescence resonance energy transfer (smFRET) at the molecular level or help interpret experimental data such as hydrogen–deuterium exchange.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

The authors gratefully acknowledge financial support to D.B. from the Deutsche Telekom Stiftung. H.v.d.B. is supported by NIH GM123159.

REFERENCES

(1). Tirion MM Large Amplitude Elastic Motions in Proteins from a Single-Parameter, Atomic Analysis. Phys. Rev. Lett 1996, 77, 1905. [PubMed: 10063201]

(2). Laman G On Graphs and Rigidity of Plane Skeletal Structures. J. Eng. Math 1970, 4, 331–340.

(3). Jacobs DJ; Thorpe MF Generic Rigidity Percolation: The Pebble Game. Phys. Rev. Lett 1995, 75, 4051. [PubMed: 10059802]

(4). Lopéz-Blanco JR; Garzón JI; Chacón P Imod: Multipurpose Normal Mode Analysis in Internal Coordinates. Bioinformatics 2011, 27, 2843–2850. [PubMed: 21873636]

(5). Noguti T; Go N σA Method of Rapid Calculation of a Second Derivative Matrix of Conformational Energy for Large Molecules. J. Phys. Soc. Jpn 1983, 52, 3685–3690.

(6). Tirion MM; ben Avraham D Atomic Torsional Modal Analysis for High-Resolution Proteins. Phys. Rev. E 2015, 91, 032712.

(7). Durand P; Trinquier G; Sanejouand Y-H A New Approach for Determining Low-Frequency Normal Modes in Macromolecules. Biopolymers 1994, 34, 759–771.

(8). Tama F; Gadea FX; Marques O; Sanejouand Y-H Building-Block Approach for Determining Low-Frequency Normal Modes of Macromolecules. Proteins: Struct., Funct., Genet 2000, 41, 1–7. [PubMed: 10944387]

(9). Suhre K; Sanejouand Y-H Elnemo: A Normal Mode Web Server for Protein Movement Analysis and the Generation of Templates for Molecular Replacement. Nucleic Acids Res 2004, 32, W610–W614. [PubMed: 15215461]

(10). Levitt M; Sander C; Stern PS Protein Normal-Mode Dynamics: Trypsin Inhibitor, Crambin, Ribonuclease and Lysozyme. J. Mol. Biol 1985, 181, 423–447. [PubMed: 2580101]

(11). Na H; Song G Conventional NMA As a Better Standard for Evaluating Elastic Network Models. Proteins: Struct., Funct., Genet 2015, 83, 259–267. [PubMed: 25471641]

(12). Na H; Jernigan RL; Song G Bridging Between NMA and Elastic Network Models: Preserving All-Atom Accuracy in Coarse-Grained Models. PLoS Comput. Biol 2015, 11, 1–22.

(13). Nicolay S; Sanejouand Y-H Functional Modes of Proteins Are Among the Most Robust. Phys. Rev. Lett 2006, 96, 078104. [PubMed: 16606146]

(14). Mendez R; Bastolla U Torsional Network Model: Normal Modes in Torsion Angle Space Better Correlate with Conformation Changes in Proteins. Phys. Rev. Lett 2010, 104, 228103. [PubMed: 20867208]

(15). Bray JK; Weiss DR; Levitt M Optimized Torsion-Angle Normal Modes Reproduce Conformational Changes More Accurately Than Cartesian Modes. Biophys. J 2011, 101, 2966–2969. [PubMed: 22208195]

(16). Bahar I; Atilgan AR; Erman B Direct Evaluation of Thermal Fluctuations in Proteins Using a Single-Parameter Harmonic Potential. Folding Des 1997, 2, 173–181.

(17). Bahar I; Wallqvist A; Covell D; Jernigan R Correlation between Native-State Hydrogen Exchange and Cooperative Residue Fluctuations from a Simple Model. Biochemistry 1998, 37, 1067–1075. [PubMed: 9454598]

(18). Atilgan A; Durell S; Jernigan R; Demirel M; Keskin O; Bahar I Anisotropy of Fluctuation Dynamics of Proteins with an Elastic Network Model. Biophys. J 2001, 80, 505–515. [PubMed: 11159421]

(19). Jacobs DJ; Hendrickson B An Algorithm for Two-Dimensional Rigidity Percolation: The Pebble Game. J. Comput. Phys 1997, 137, 346–365.

(20). Thorpe M; Lei M; Rader A; Jacobs DJ; Kuhn LA Protein Flexibility and Dynamics Using Constraint Theory. J. Mol. Graphics Modell 2001, 19, 60–69.

(21). Jacobs DJ Generic Rigidity in Three-Dimensional Bond-Bending Networks. J. Phys. A: Math. Gen 1998, 31, 6653.

(22). Whiteley W Counting Out to the Flexibility of Molecules. Phys. Biol 2005, 2, S116. [PubMed: 16280617]

(23). Katoh N; Tanigawa S.-i. A Proof of the Molecular Conjecture. Proceedings of the 25th annual symposium on Computational geometry 2009, 296–305.

(24). Fox N; Jagodzinski F; Li Y; Streinu I KINARI-Web: A Server for Protein Rigidity Analysis. Nucleic Acids Res 2011, 39, W177–W183. [PubMed: 21693559]

(25). Cai Y; Thorpe M Floppy Modes in Network Glasses. Phys. Rev. B: Condens. Matter Mater. Phys 1989, 40, 10535.

(26). Chubynsky M; Thorpe M Algorithms for Three-Dimensional Rigidity Analysis and a First-Order Percolation Transition. Phys. Rev. E 2007, 76, 041135.

(27). Jacobs DJ; Rader AJ; Kuhn LA; Thorpe MF Protein Flexibility Predictions Using Graph Theory. Proteins: Struct., Funct., Genet 2001, 44, 150–165. [PubMed: 11391777]

(28). Pfleger C; Radestock S; Schmidt E; Gohlke H Global and Local Indices for Characterizing Biomolecular Flexibility and Rigidity. J. Comput. Chem 2013, 34, 220–233. [PubMed: 23007873]

(29). Rader A; Hespenheide BM; Kuhn LA; Thorpe MF Protein Unfolding: Rigidity Lost. Proc. Natl. Acad. Sci. U. S. A 2002, 99, 3540–3545. [PubMed: 11891336]

(30). Fulle S; Gohlke H Analyzing the Flexibility of RNA Structures by Constraint Counting. Biophys. J 2008, 94, 4202–4219. [PubMed: 18281388]

(31). Raschka S; Bemister-Buffington J; Kuhn LA Detecting the Native Ligand Orientation by Interfacial Rigidity: Siteinterlock. Proteins: Struct., Funct., Genet 2016, 84, 1888–1901. [PubMed: 27699847]

(32). Vorov OK; Livesay DR; Jacobs DJ Nonadditivity in Conformational Entropy upon Molecular Rigidification Reveals a Universal Mechanism Affecting Folding Cooperativity. Biophys. J 2011, 100, 1129–38. [PubMed: 21320459]

(33). Gohlke H; Ben-Shalom IY; Kopitz H; Pfeiffer-Marek S; Baringhaus K-H Rigidity Theory-Based Approximation of Vibrational Entropy Changes upon Binding to Biomolecules. J. Chem. Theory Comput 2017, 13, 1495–1502. [PubMed: 28355485]

(34). Ben-Shalom IY; Pfeiffer-Marek S; Baringhaus K-H; Gohlke H Efficient Approximation of Ligand Rotational and Translational Entropy Changes upon Binding for Use in MM-PBSA Calculations. J. Chem. Inf. Model 2017, 57, 170–189. [PubMed: 27996253]

(35). Abella JR; Moll M; Kavraki LE Maintaining and Enhancing Diversity of Sampled Protein Conformations in Robotics-Inspired Methods. J. Comput. Biol 2018, 25, 3–20. [PubMed: 29035572]

(36). Thomas S; Tang X; Tapia L; Amato NM Simulating Protein Motions with Rigidity Analysis. J. Comput. Biol 2007, 14, 839–855. [PubMed: 17691897]

(37). Luo D; Haspel N Multi-Resolution Rigidity-Based Sampling of Protein Conformational Paths Proceedings of the International Conference on Bioinformatics, Computational Biology and Biomedical Informatics 2013, 786.

(38). Rader A Thermostability in Rubredoxin and Its Relationship to Mechanical Rigidity. Phys. Biol 2010, 7, 016002.

(39). Radestock S; Gohlke H Protein Rigidity and Thermophilic Adaptation. Proteins: Struct., Funct., Genet 2011, 79, 1089–1108. [PubMed: 21246632]

(40). Rathi PC; Fulton A; Jaeger K-E; Gohlke H Application of Rigidity Theory to the Thermostabilization of Lipase A from Bacillus Subtilis. PLoS Comput. Biol 2016, 12, 1–21.

(41). Hespenheide BM; Rader A; Thorpe MF; Kuhn LA Identifying Protein Folding Cores from the Evolution of Flexible Regions during Unfolding. J. Mol. Graphics Modell 2002, 21, 195–207.

(42). Livesay DR; Huynh DH; Dallakyan S; Jacobs DJ Hydrogen Bond Networks Determine Emergent Mechanical and Thermodynamic Properties across a Protein Family. Chem. Cent. J 2008, 2, 17. [PubMed: 18700034]

(43). Lei M; Zavodszky MI; Kuhn LA; Thorpe M Sampling Protein Conformations and Pathways. J. Comput. Chem 2004, 25, 1133–1148. [PubMed: 15116357]

(44). Wells S; Menor S; Hespenheide B; Thorpe M Constrained Geometric Simulation of Diffusive Motion in Proteins. Phys. Biol 2005, 2, S127. [PubMed: 16280618]

(45). Ahmed A; Gohlke H Multiscale Modeling of Macromolecular Conformational Changes Combining Concepts from Rigidity and Elastic. Proteins: Struct., Funct., Genet 2006, 63, 1038–1051. [PubMed: 16493629]

(46). Jimenez-Roldan JE; Freedman RB; Römer RA; Wells SA Rapid Simulation of Protein Motion: Merging Flexibility, Rigidity and Normal Mode Analyses. Phys. Biol 2012, 9, 016008 [PubMed: 22313618]

(47). Budday D; Leyendecker S; van den Bedem H Geometric Analysis Characterizes Molecular Rigidity in Generic and Non-Generic Protein Configurations. J. Mech. Phys. Solids 2015, 83, 36–47. [PubMed: 26213417]

(48). Pachov DV; van den Bedem H Nullspace Sampling with Holonomic Constraints Reveals Molecular Mechanisms of Protein Gαs. PLoS Comput. Biol 2015, 11, 1–24.

(49). Pachov DV; Fonseca R; Arnol D; Bernauer J; van den Bedem H Coupled Motions in β2AR:Gαs Conformational Ensembles. J. Chem. Theory Comput 2016, 12, 946–956. [PubMed: 26756780]

(50). Golub GH; Van Loan CF Matrix Computations; JHU Press, 2012; Vol. 3.

(51). Buehler MJ; Keten S Elasticity, Strength and Resilience: A Comparative Study on Mechanical Signatures of α-helix, β-sheet and Tropocollagen Domains. Nano Res 2008, 1, 63–71.

(52). Dahiyat BI; Benjamin Gordon D; Mayo SL Automated Design of the Surface Positions of Protein Helices. Protein Sci 1997, 6, 1333–1337. [PubMed: 9194194]

(53). Glyakina AV; Bogatyreva NS; Galzitskaya OV Accessible Surfaces of Beta Proteins Increase with Increasing Protein Molecular Mass More Rapidly Than Those of Other Proteins. PLoS One 2011, 6, 1–10.

(54). Perrino AP; Garcia R How Soft Is a Single Protein? The Stress-Strain Curve of Antibody Pentamers with 5 pN and 50 pm Resolutions. Nanoscale 2016, 8, 9151–9158. [PubMed: 26732032]

(55). Brüschweiler R Collective Protein Dynamics and Nuclear Spin Relaxation. J. Chem. Phys 1995, 102, 3396–3403.

(56). Bhardwaj G; Mulligan VK; Bahl CD; Gilmore JM; Harvey PJ; Cheneval O; Buchko GW; Pulavarti SV; Kaas Q; Eletsky A; Huang P-S; Johnsen WA; Greisen PJ; Rocklin GJ; Song Y; Linsky TW; Watkins A; Rettie SA; Xu X; Carter LP; Bonneau R; Olson JM; Coutsias E; Correnti CE; Szyperski T; Craik DJ; Baker D Accurate De Novo Design of Hyperstable Constrained Peptides. Nature 2016, 538, 329–335. [PubMed: 27626386]

(57). Na H; Song G; ben Avraham D Universality of Vibrational Spectra of Globular Proteins. Phys. Biol 2016, 13, 016008. [PubMed: 26907186]

(58). Keten S; Buehler MJ Geometric Confinement Governs the Rupture Strength of H-Bond Assemblies at a Critical Length Scale. Nano Lett 2008, 8, 743–748. [PubMed: 18269263]

(59). Dzeja P; Terzic A Adenylate Kinase and Amp Signaling Networks: Metabolic Monitoring, Signal Communication and Body Energy Sensing. Int. J. Mol. Sci 2009, 10, 1729–1772. [PubMed: 19468337]

(60). Schulz GE; Müller CW; Diederichs K Induced-Fit Movements in Adenylate Kinases. J. Mol. Biol 1990, 213, 627–630. [PubMed: 2162964]

(61). Müller CW; Schulz GE Structure of the Complex between Adenylate Kinase from Escherichia Coli and the Inhibitor Ap5A Refined at 1.9 Å Resolution: A Model for a Catalytic Transition State. J. Mol. Biol 1992, 224, 159–177. [PubMed: 1548697]

(62). Gerstein M; Schulz G; Chothia C Domain Closure in Adenylate Kinase: Joints on Either Side of Two Helices Close like Neighboring Fingers. J. Mol. Biol 1993, 229, 494–501. [PubMed: 8429559]

(63). Vonrhein C; Schlauderer GJ; Schulz GE Movie of the Structural Changes during a Catalytic Cycle of Nucleoside Monophosphate Kinases. Structure 1995, 3, 483–490. [PubMed: 7663945]

(64). Müller C; Schlauderer G; Reinstein J; Schulz GE Adenylate Kinase Motions during Catalysis: An Energetic Counterweight Balancing Substrate Binding. Structure 1996, 4, 147–156. [PubMed: 8805521]

(65). Beckstein O; Denning EJ; Perilla JR; Woolf TB Zipping and Unzipping of Adenylate Kinase: Atomistic Insights into the Ensemble of Open ↔ Closed Transitions. J. Mol. Biol 2009, 394, 160–176. [PubMed: 19751742]

(66). Seyler SL; Beckstein O Sampling Large Conformational Transitions: Adenylate Kinase As a Testing Ground. Mol. Simul 2014, 40, 855.

(67). Li D; Liu MS; Ji B Mapping the Dynamics Landscape of Conformational Transitions in Enzyme: The Adenylate Kinase Case. Biophys. J 2015, 109, 647–660. [PubMed: 26244746]

(68). Budday D; Fonseca R; Leyendecker S; van den Bedem H Frustration-Guided Motion Planning Reveals Conformational Transitions in. Proteins: Struct., Funct., Genet 2017, 85, 1795–1807. [PubMed: 28597937]

(69). Aviram HY; Pirchi M; Mazal H; Barak Y; Riven I; Haran G Direct Observation of Ultrafast Large-Scale Dynamics of an Enzyme under Turnover Conditions. Proc. Natl. Acad. Sci. U. S. A 2018, 115, 3243. [PubMed: 29531052]

(70). LaValle SM Rapidly-exploring Random Trees: A New Tool for Path Planning Iowa State University, 1998.

(71). Héliou A; Budday D; Fonseca R; van den Bedem H, Fast, Clash-Free RNA Conformational Morphing Using Molecular Junctions. Bioinformatics 2017, 33, 2114–2122. [PubMed: 28334257]

(72). Fonseca R; Budday D; van den Bedem H Collision-Free Poisson Motion Planning in Ultra High-Dimensional Molecular Conformation Spaces. J. Comput. Chem 2018, 39, 711–720. [PubMed: 29315667]

(73). Lee S; Wang C; Liu H; Xiong J; Jiji R; Hong X; Yan X; Chen Z; Hammel M; Wang Y; Dai S; Wang J; Jiang C; Zhang G Hydrogen Bonds Are a Primary Driving Force for De Novo Protein Folding. Acta Crystallogr 2017, D73, 955–969.

(74). Zhang J; Wang N; Liu W; Zhao X; Lu W Intermolecular Hydrogen Bonding Strategy to Fabricate Mechanically Strong Hydrogels with High Elasticity and Fatigue Resistance. Soft Matter 2013, 9, 6331–6337.

(75). Stranges PB; Kuhlman B A Comparison of Successful and Failed Protein Interface Designs Highlights the Challenges of Designing Buried Hydrogen Bonds. Protein Sci 2013, 22, 74–82. [PubMed: 23139141]

(76). Huang P-S; Boyken SE; Baker D The Coming of Age of De Novo Protein Design. Nature 2016, 537, 320–327. [PubMed: 27629638]

(77). Boyken SE; Chen Z; Groves B; Langan RA; Oberdorfer G; Ford A; Gilmore JM; Xu C; DiMaio F; Pereira JH; Sankaran B; Seelig G; Zwart PH; Baker D De Novo Design of Protein Homo-Oligomers with Modular Hydrogen-Bond Network-Mediated Specificity. Science 2016, 352, 680–687. [PubMed: 27151862]

(78). van den Bedem H; Fraser JS Integrative, Dynamic Structural Biology at Atomic Resolution - It's about Time. Nat. Methods 2015, 12, 307–318. [PubMed: 25825836]

(79). Buehler MJ; Wong SY Entropic Elasticity Controls Nanomechanics of Single Tropocollagen Molecules. Biophys. J 2007, 93, 37–43. [PubMed: 17434941]

(80). Painter P; Mosher L; Rhoads C Low-Frequency Modes in the Raman Spectra of Proteins. Biopolymers 1982, 21, 1469–1472. [PubMed: 7115900]

(81). Cai S; Singh BR A Distinct Utility of the Amide Iii Infrared Band for Secondary Structure Estimation of Aqueous Protein Solutions Using Partial Least Squares Methods. Biochemistry 2004, 43, 2541–2549. [PubMed: 14992591]

(82). Yang H; Yang S; Kong J; Dong A; Yu S Obtaining Information about Protein Secondary Structures in Aqueous Solution Using Fourier Transform IR Spectroscopy. Nat. Protoc 2015, 10, 382–396. [PubMed: 25654756]

(83). Xiao S; Stacklies W; Cetinkaya M; Markert B; Gräter F Mechanical Response of Silk Crystalline Units from Force-Distribution Analysis. Biophys. J 2009, 96, 3997–4005. [PubMed: 19450471]

(84). Gavrilov Y; Leuchter JD; Levy Y On the Coupling between the Dynamics of Protein and Water. Phys. Chem. Chem. Phys 2017, 19, 8243–8257. [PubMed: 28277584]

(85). Johnson ME; Malardier-Jugroot C; Murarka RK; Head-Gordon T Hydration Water Dynamics near Biological Interfaces. J. Phys. Chem. B 2009, 113, 4082–4092. [PubMed: 19425247]

(86). Frauenfelder H; Chen G; Berendzen J; Fenimore PW; Jansson H; McMahon BH; Stroe IR; Swenson J; Young RD A Unified Model of Protein Dynamics. Proc. Natl. Acad. Sci. U. S. A 2009, 106, 5129–5134. [PubMed: 19251640]

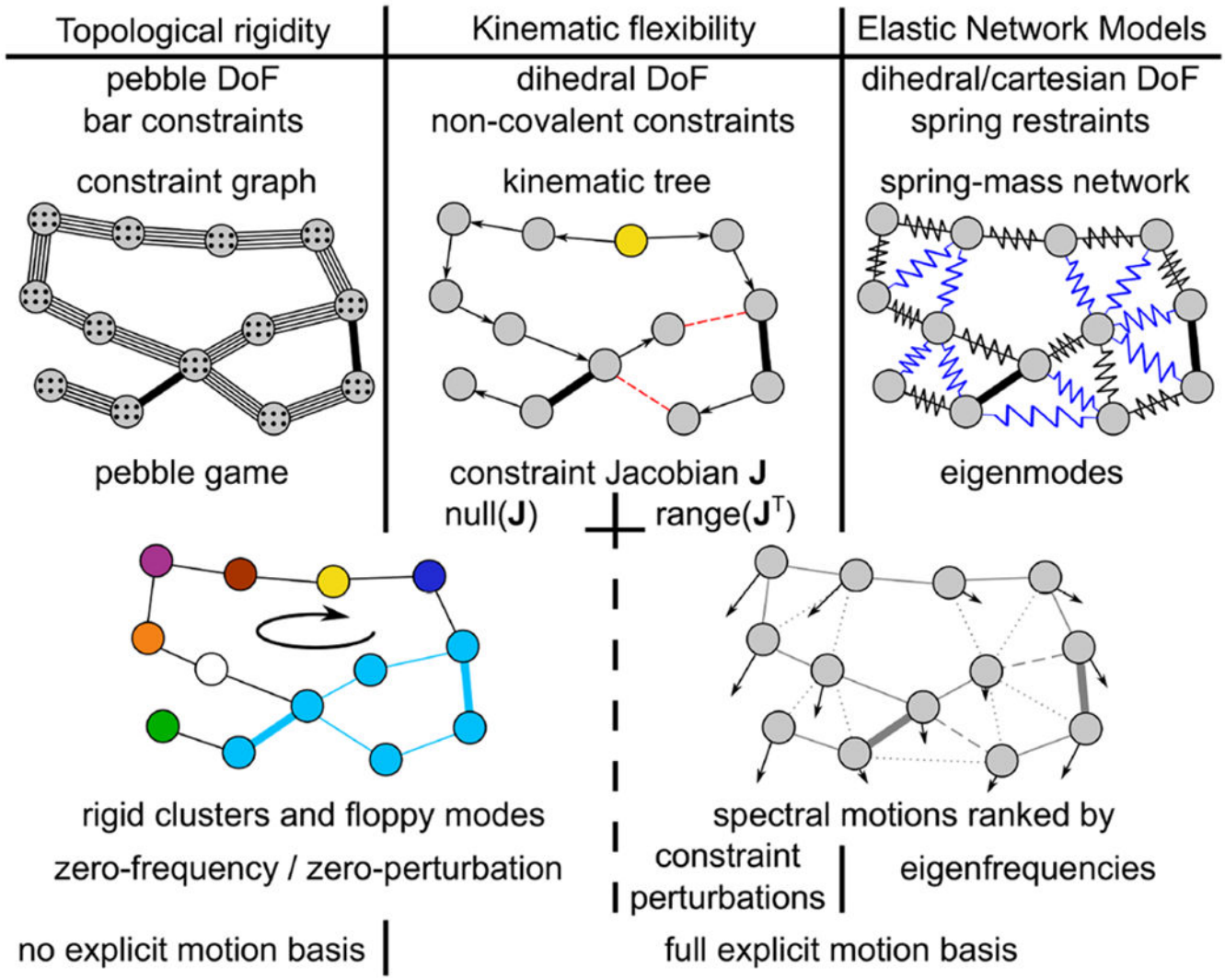


Figure 1.

Coarse-grained protein modeling via topological rigidity analysis (left) and elastic network models (ENMs, right). The pebble game represents proteins as constraint graphs (top left), with pebble DoF (small dots) and bar constraints (thin lines are single constraints, thick lines are rigid links with six constraints) connecting atomic vertices (spheres). It decomposes a protein into rigid clusters of atoms (lower left, individually colored), quantifying the number of internal “floppy modes” (circular arrow), without an explicit motion basis. By contrast, ENM obtains an explicit motion basis (arrows lower right) corresponding to eigenmodes of a spring–mass network (top right), with covalent (black) and noncovalent (blue) one-dimensional spring restraints. Our kinematic flexibility approach (center) models proteins as kinematic spanning trees, with a root vertex (yellow), dihedral DoF (arrows between vertices), and noncovalent constraints (red dashed lines), for example, hydrogen bonds. It combines features from topological rigidity and ENM, providing explicit motion modes from a spectral decomposition of the constraint Jacobian matrix.

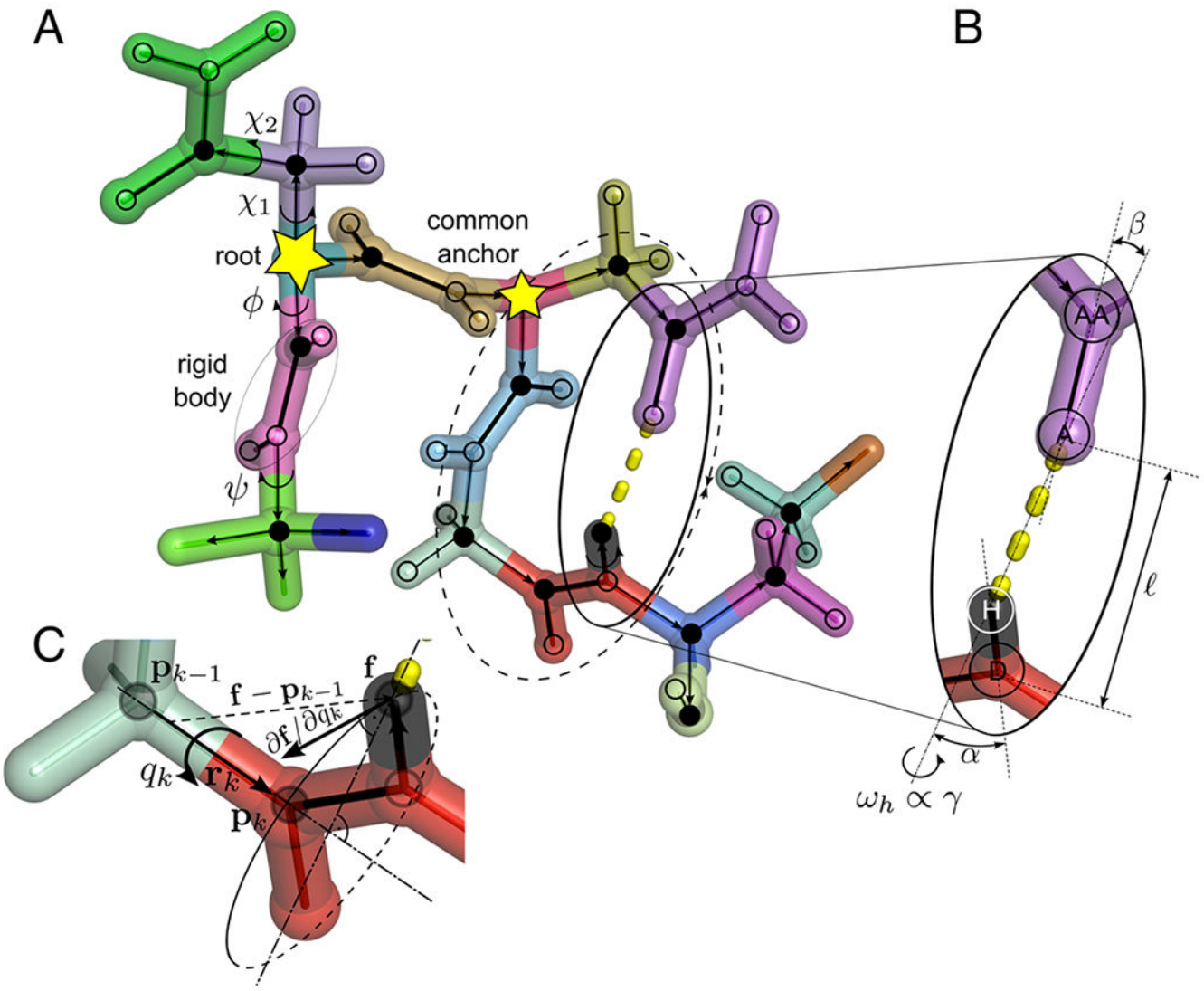


Figure 2.

Kinematic representation of a protein. (A) Directed, kinematic spanning tree of a protein fragment. Edges (arrows) represent rotatable bonds, and groups of atoms in the same color represent the rigid body vertices. Atoms (circles) that are connected via nonrotatable (single/double) covalent bonds (thin/bold lines) are merged into a single rigid body vertex. Starting from the root vertex (large yellow star), each vertex is visited by a directed edge from its parent vertex (arrows to filled circles). Hydrogen bonds constrain two branches leaving from a common anchor (small yellow star) at their end effectors to form closed kinematic cycles. (B) Constraint parametrization. Hydrogen bond distance $\mathbf{f}_H - \mathbf{f}_A$ and angles α and β are formulated as constraints, which permits a rotation ω_h around the hydrogen bond axis. (C) Partial derivative $\partial\mathbf{f}/\partial q_k$ required for the constraint Jacobian matrix is the cross-product $\mathbf{r}_k \times (\mathbf{f} - \mathbf{p}_{k-1})$.

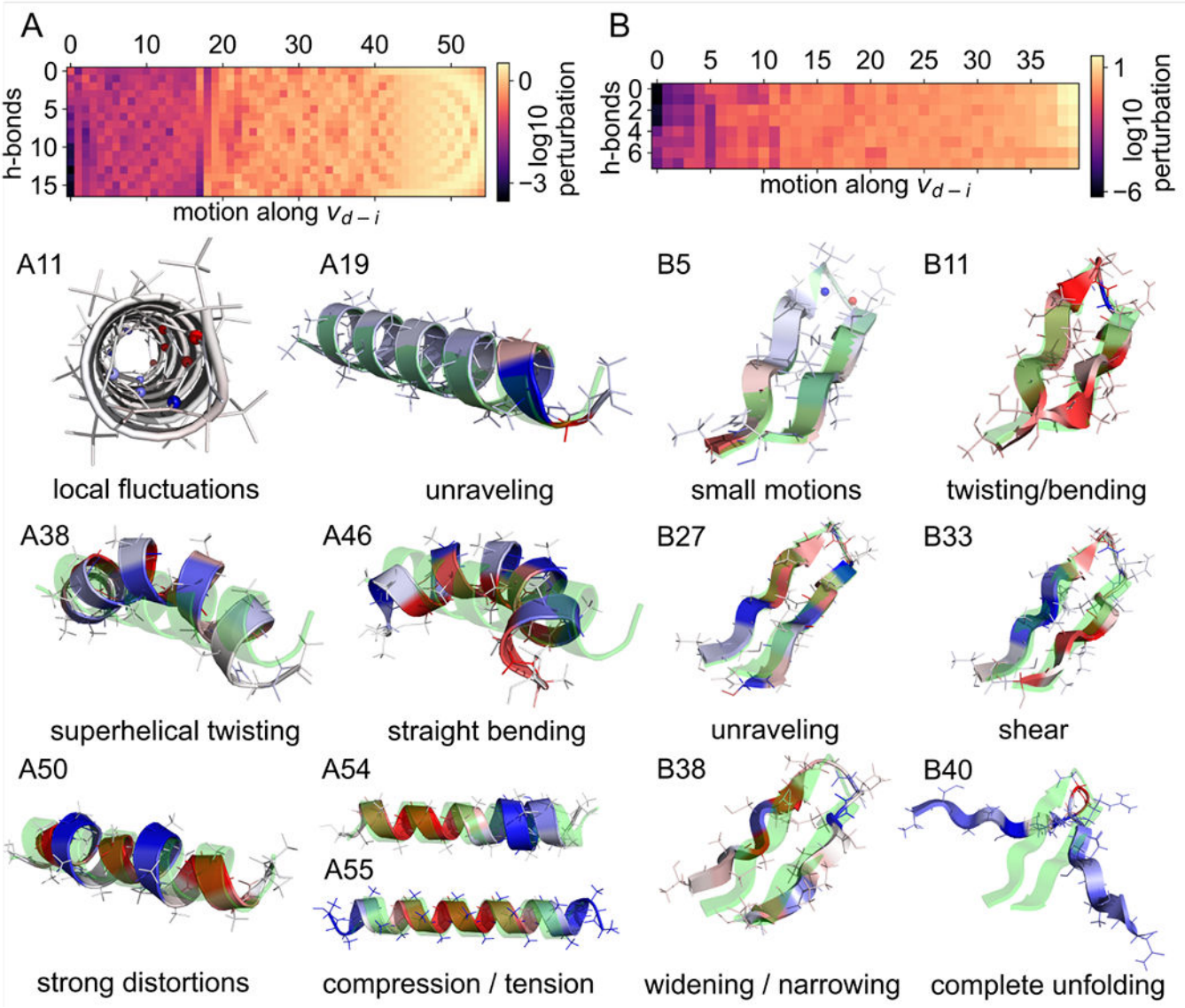


Figure 3.

Hierarchical constraint relaxation in an α-helix (A panels) and an antiparallel β-sheet (B panels). The two top panels plot the matrix **P**, with individual hydrogen bond (rows) perturbations associated with instantaneous motions along singular vectors (kinematic flexibility modes, columns), increasing from purple to yellow. Modes are ranked by collective constraint perturbation $\boldsymbol{\sigma}$, increasing from left to right. Bottom panels depict snapshots from a finite step along selected, representative kinematic flexibility modes (column indices given). The motion amplitude is exaggerated (step size $\delta = 1$) for visualization purposes. Color coding indicates increasing change in DoF, from blue to red. The original conformation is shown in green.

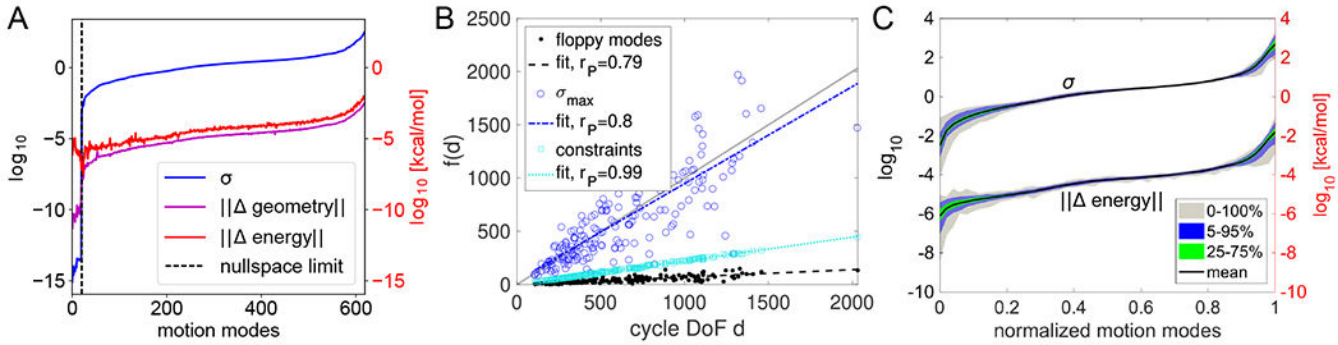


Figure 4.

Analysis of the hydrogen bond network spectral decomposition for human DJ-1 protein (A, PDB ID 1p5f) with $d = 620$ and a large, diverse data set with 183 high-resolution single-chain proteins (B,C). (A) The singular values $\boldsymbol{\sigma}$ predict collective kinematic and energetic hydrogen bond perturbations associated with each motion mode. Modes (horizontal axis) are plotted corresponding to increasing $\boldsymbol{\sigma}$. Geometry and energy changes are computed from small steps (step size $\delta = 10^{-5}$) along kinematic flexibility modes **v**. (B) The number of hydrogen bond constraints, remaining floppy modes, as well as the magnitude of the largest singular value vary fairly linearly (correlation coefficient r_p) with the number of DoF d in kinematic cycles, which serves as indicator for protein size. (C) The ranking of collective changes in hydrogen bond energy by kinematic flexibility modes and the near log–linear motion regime are remarkably conserved among all proteins. Motion modes are limited to **R**, and indices are normalized by r to remove size-related differences in the diverse data set.

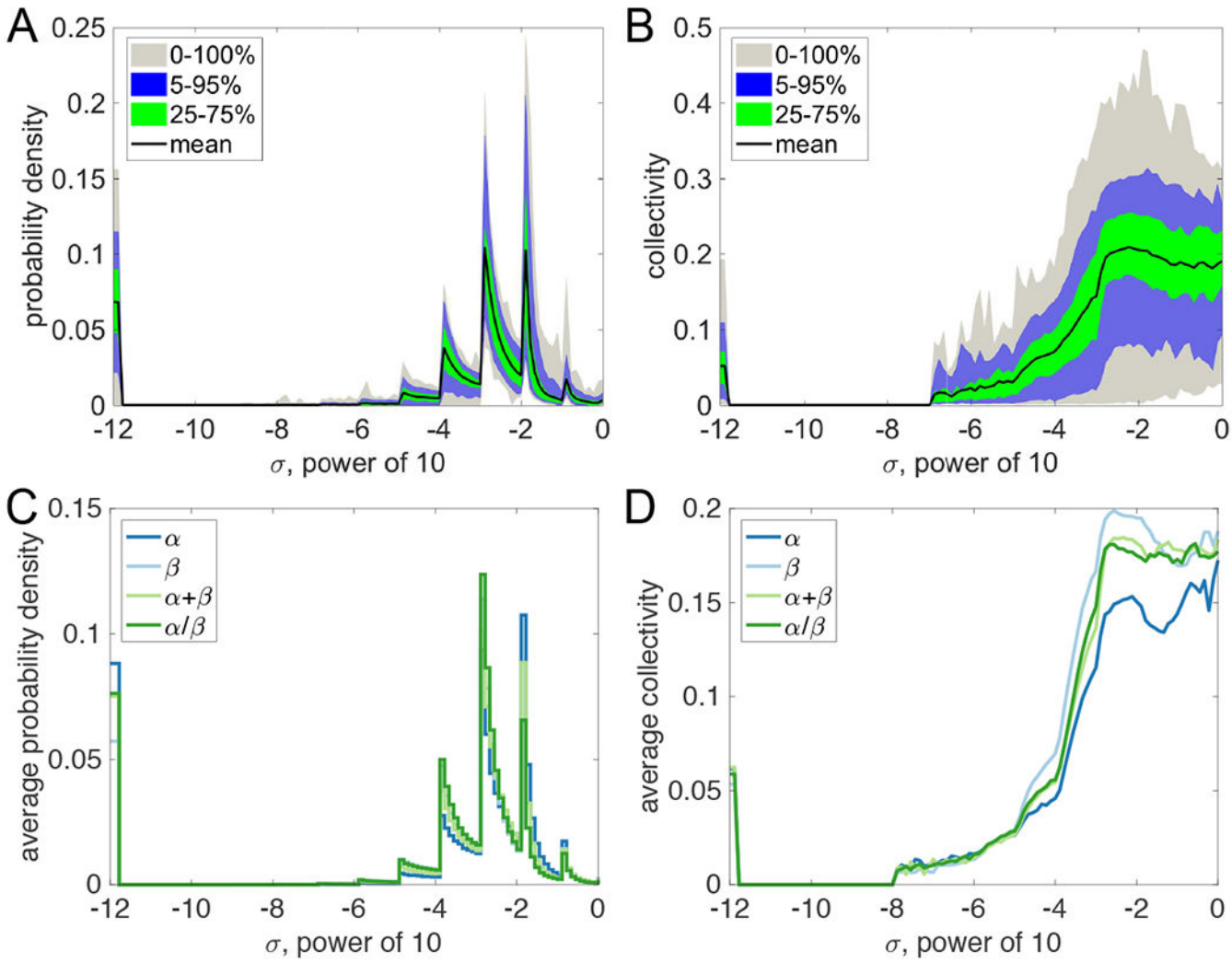


Figure 5.

Kinematic flexibility analysis with hydrogen bond constraints reveals fold-specific motions in proteins. (A) Spectrum of normalized singular values for 183 single-chain proteins, with striking, well-conserved peaks. The nullspace dimension (peak at lowest σ) highly varies with protein size. (B) Collectivity of motion as defined from Shannon entropy. Floppy modes with zero perturbation (lowest σ) show relatively small collectivity compared to modes at higher perturbations. Over a medium-perturbation range, collectivity increases with increasing singular values. (C) Fold-specific, average spectrum for classes of α-only, β-only, $\alpha + \beta$, and α/β proteins (stair representation only for better visibility). Peak locations are well conserved across folds, while peak heights are shifted to lower σ from α-only to β-only. Floppy modes (lowest σ) vary rather with protein size than fold. (D) Average collectivity as in part B for fold-specific data sets. Again, floppy modes with zero perturbation show relatively small collectivity; they are indistinguishable across folds. Throughout the remaining spectrum, β-folds are generally more collective.

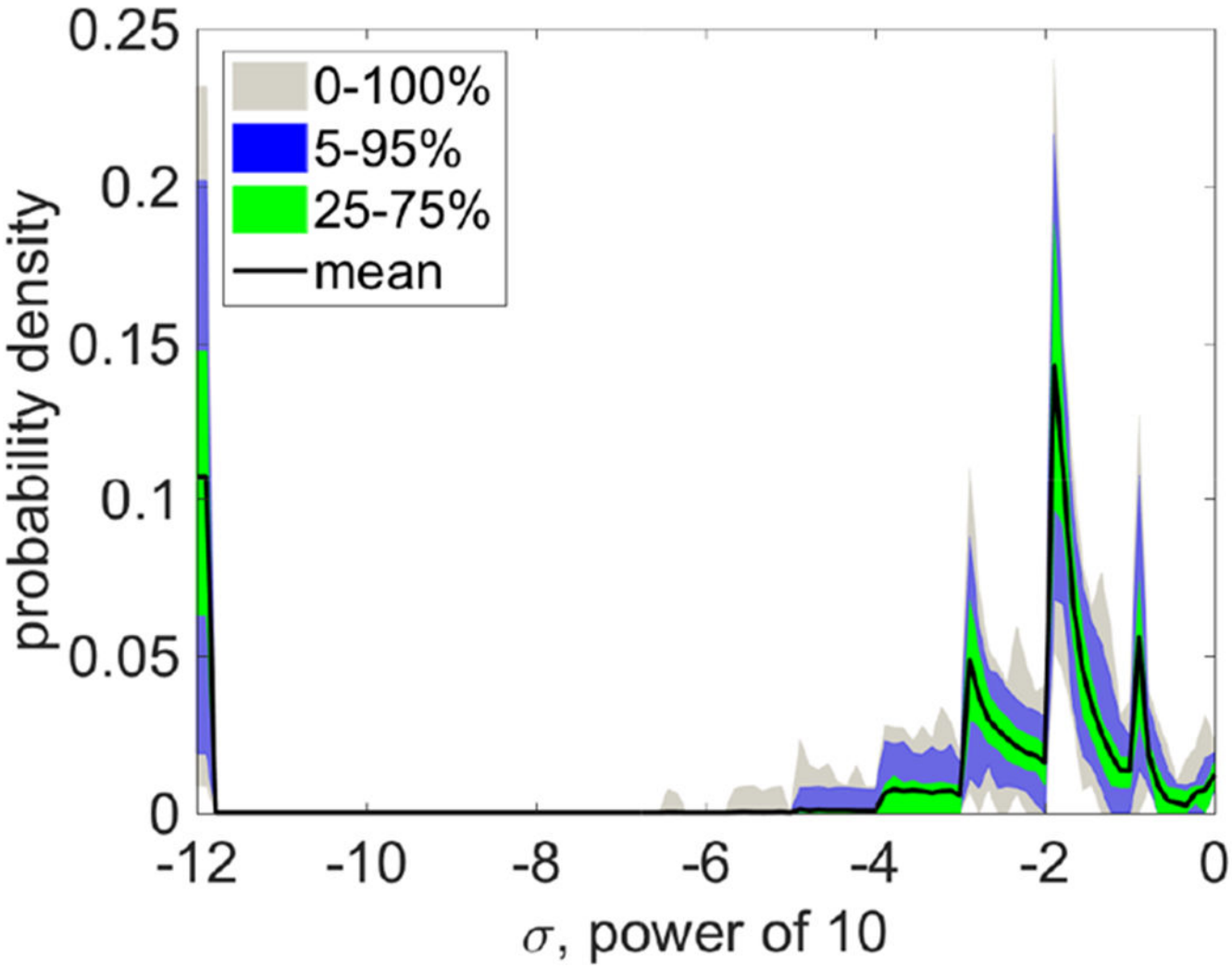


Figure 6.

Spectrum of normalized singular values for four designed hyperstable, constrained peptides, each consisting of 20 NMR structures. The spectrum is shifted to higher perturbation modes relative to the high-resolution data set.

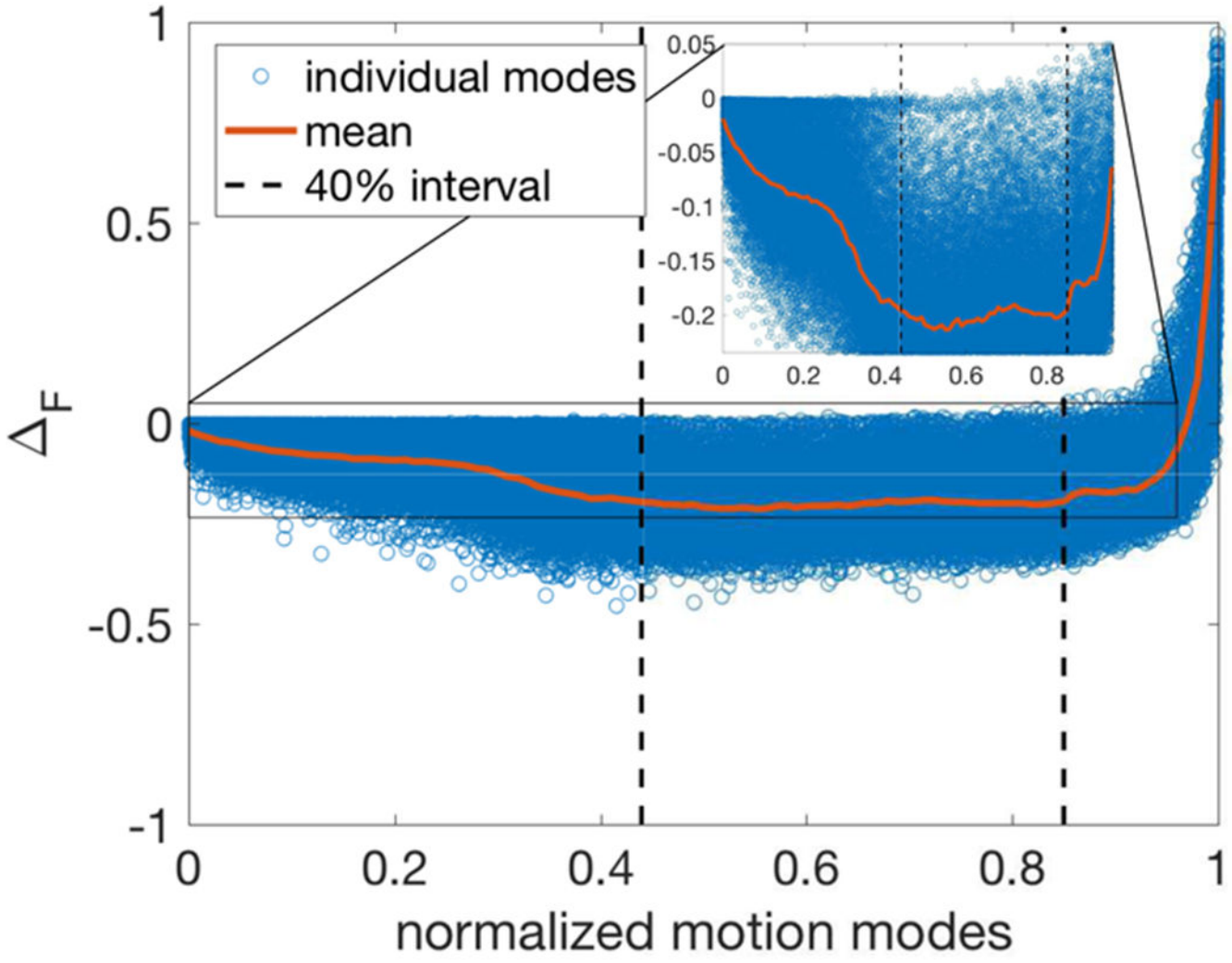


Figure 7.

Dimensionless free energy of modes demonstrates entropy–enthalpy compensation encoded by the hydrogen bonding pattern. A 40% interval of motion modes indicates a near constant free-energy level for highly collective modes at acceptable enthalpic cost (dashed lines, inset enlarges the blackbox). Motions to the right correspond to unfolding, while motions to the left are more localized, with smaller entropic benefit.

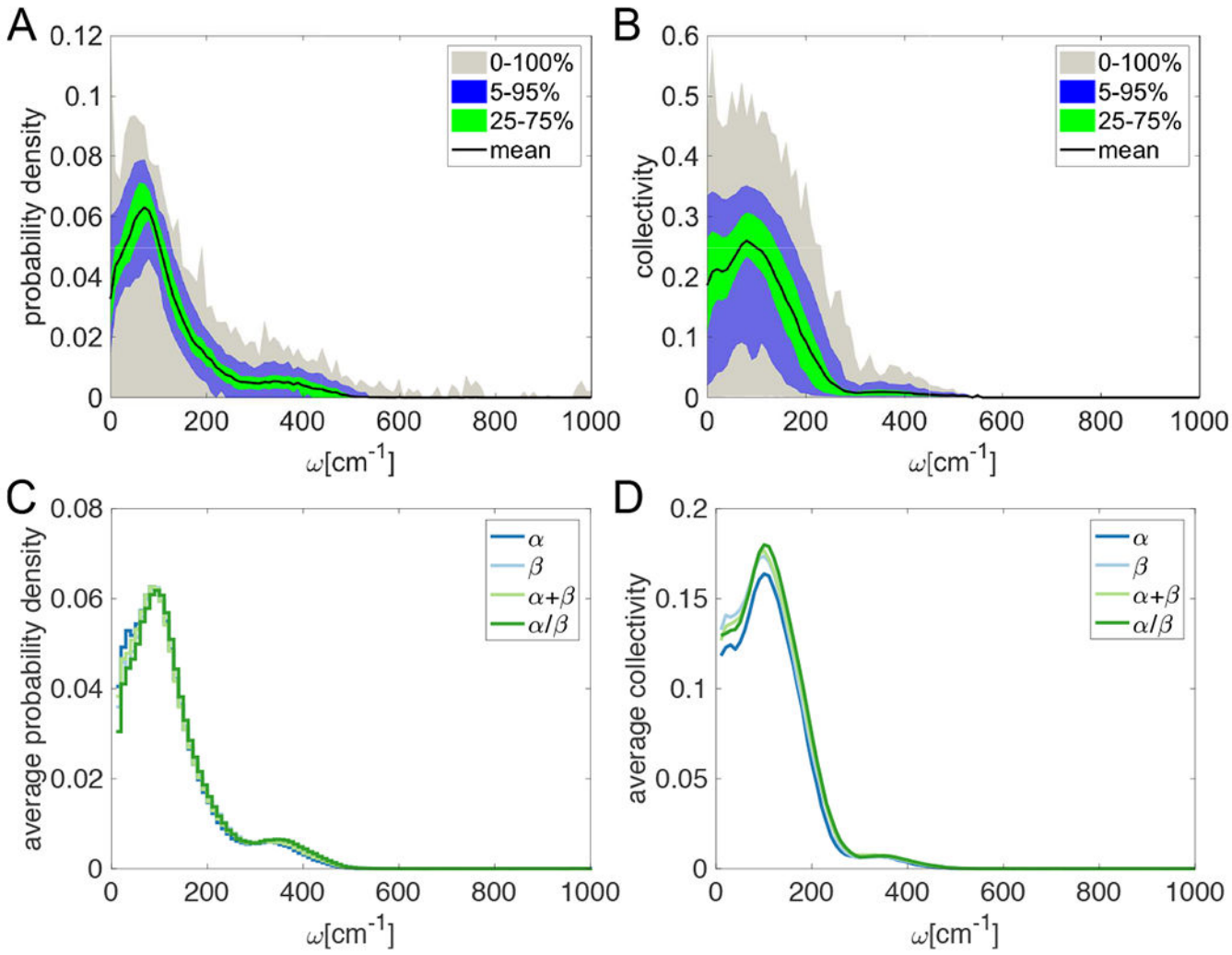


Figure 8.

Eigenspectrum and collectivity for normal modes computed with iMOD. (A) Eigenfrequency spectra of the data set with 183 single-chain proteins. (B) Collectivity of eigenmodes. (C) Eigenfrequency spectra of four fold-specific data sets. (D) Fold-specific collectivity of eigenmodes.

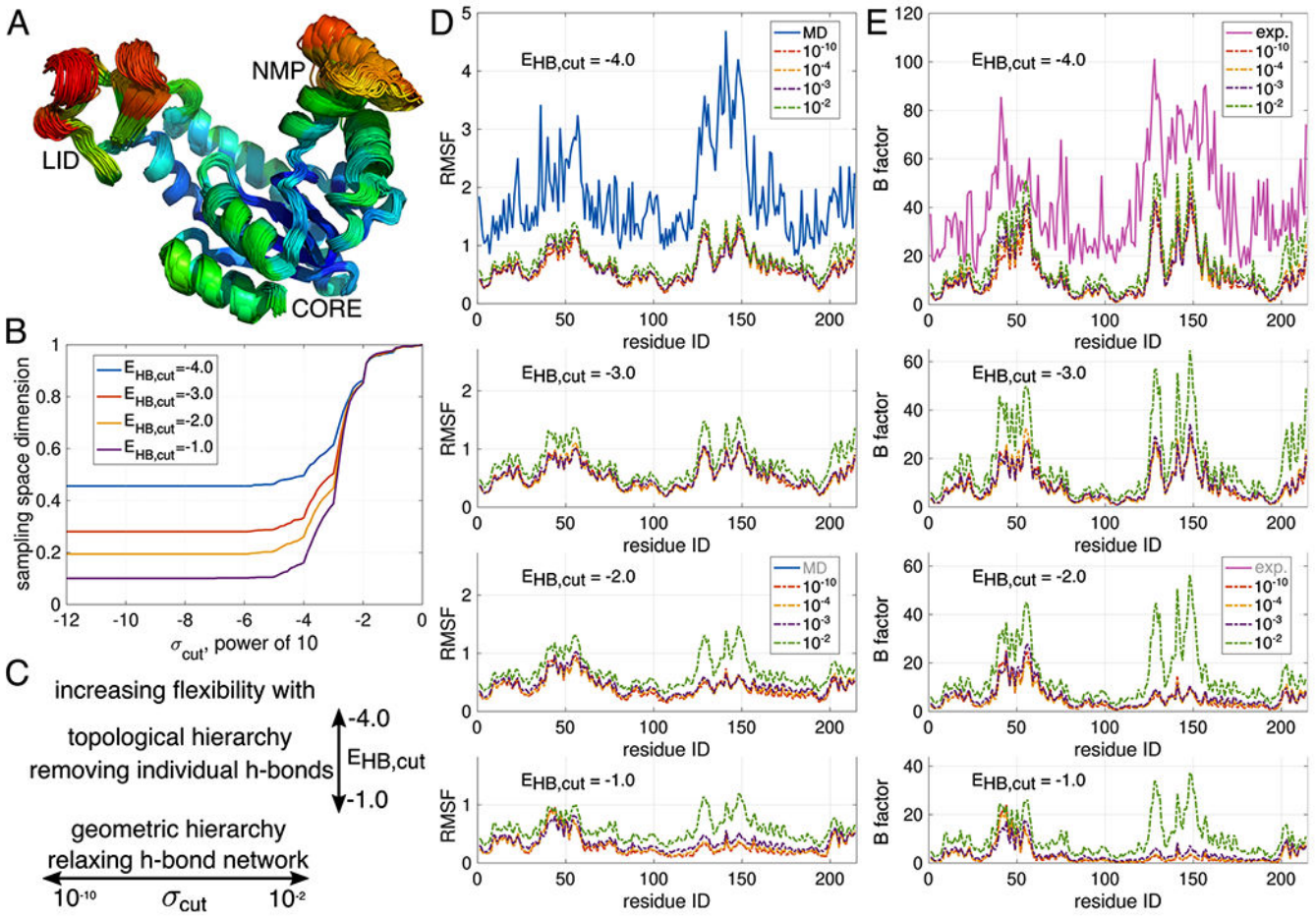


Figure 9.

Molecular motions of adenylate kinase (ADK). (A) ADK opens and closes the LID and NMP domains during its catalytic cycle. The conformational ensemble colored by RMSF from blue to red shows 100 substates from sampling ADK with $E_{HB,cut} = -2.0$ kcal/mol and $\sigma_{cut} = 10^{-2}$, starting from the open conformation (PDB ID 4ake[64]). (B) Normalized dimensions of sampling space $\dim\left(\vec{V}_q^0\right) / \dim\left(V_q^0\right)$ show significant dependence on $E_{HB,cut}$, while differences vanish at $\sigma_{cut} \approx 10^{-2}$. (C) Hierarchies to tune protein flexibility by constraint removal (topological, $E_{HB,cut}$) and network relaxation (geometric, σ_{cut}). (D, E) Correlations of KGS motions with RMSF (Å) from MD (D) and experimental B factors (Å^2) reported in the PDB file (E) are consistently high for medium constraint relaxation $\sigma_{cut} = 10^{-2}$ and robust toward constraint input. The previous topological hierarchy without relaxation ($\sigma_{cut} = 10^{-10}$) highly depends on constraint input and only achieves good correlations for strong $E_{HB,cut}$.

Table 1.

Statistics of the High-Resolution Dataset with 183 Single-Chain Proteins[a]

	no. of residues	m	d	$d-r$	h-bond energy	σ_1
median	163	116	516	34	−422.94 kcal/mol	485.67
min	30	23	107	0	−1506.81 kcal/mol	82.94
max	555	445	2032	148	−59.88 kcal/mol	1966.29

[a] m is the number of hydrogen bonds, d is the number of cycle DoFs, $d-r$ is the nullity of the constraint matrix, h-bond energy is the total energy of all hydrogen bonds, and σ_1 is the largest singular value.