THE JOURNAL OF PHYSICAL CHEMISTRY B
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCB

Article

Systematic Detection and Characterization of Hydrogen Bonding in Proteins via Local Vibrational Modes

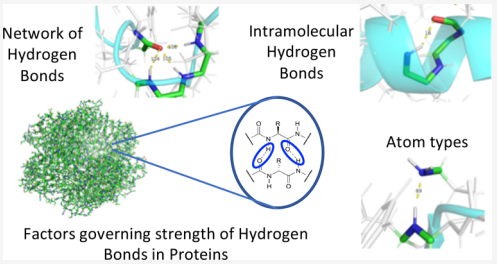
Niraj Verma, Yunwen Tao, and Elfi Kraka*

Cite This: *J. Phys. Chem. B* 2021, 125, 2551−2565

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: We introduce a new software, *Efficient Detection of Hydrogen Bonds* (EDHB), that systematically detects hydrogen bonds based on the nearest neighbors algorithm. EDHB classifies inter- and intramolecular hydrogen bonds as well as hydrogen bond networks. EDHB outperforms commonly used hydrogen bond detection methods in terms of speed of execution. An important additional feature of EDHB is that information from preceding quantum chemical studies (i.e., natural bond orbital analysis data and second energy derivative information) can be used to determine the electrostatic/covalent character of the hydrogen bonds and to calculate local-mode hydrogen bond force constants as a quantitative measure of their intrinsic strength. We applied EDHB to a diverse set of 163 proteins. We identified hydrogen bond networks forming intramolecular rings of different sizes as a common feature playing an important role for specific secondary structure orientations such as α-helixes and turns. However, these networks do not have a significant influence on the hydrogen bond strength. Our comprehensive local-mode analysis reveals the interesting result that the hydrogen bond angle is the governing factor determining the hydrogen bond strength in a protein. EDHB offers a broad range of application possibilities. In addition to proteins, EDHB can be generally used to detect and characterize hydrogen bonds in protein−ligand interactions, water clusters, and other systems where a hydrogen bond plays a critical role, as well as during molecular dynamics simulations. The program is freely available at https://github.com/ekraka/EDHB.



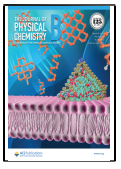
INTRODUCTION

Hydrogen bonds (HBs) play a crucial role in proteins forging enzyme reactions, protein folding, protein−ligand interactions, and other processes.[1−3] The early work of Pauling and Mirsky highlighted already in 1936 the importance of HBs in proteins suggesting a 5 kcal/mol stability from each HB.[4] In 1951, they discovered α-helix and β-pleated sheet as the most important conformational elements in a protein with α-helix HBs on the order of 8 kcal/mol.[5,6] Later that same year, they realized that an aqueous environment can lead to HBs of the order of 2 kcal/mol.[6] After almost 7 decades of intensive research, there are still some controversies about the strength and properties of HBs in a protein. Theoretical studies in the early 90s suggest that HBs do not contribute to thermodynamic stability as the overall energy balance of HBs is close to zero.[7] Campos et al.[8] suggested with a combined experimental and theoretical study that HBs in fact destabilize the native conformation of a protein. Recent studies have provided a better picture of HBs in proteins, although controversies about their strength still persist.[9]

The protein environment adds complexity to HBs' stability, and thus inclusion of all atoms is critical. Studies conducted on fragments of a protein structure[10] are limited by this crucial information. Pace et al.[3,11] have concluded that the polar residues buried inside the protein interior undergo stronger van der Waals forces in a HB. Further studies by Hubbard and Kamran Haider[1] have shown that the buried intramolecular HBs are stronger due to their electrostatic nature. The protein environment can induce an electric field on the order of 100 MV/cm.[12] Thus, to incorporate the protein environment, we utilized all atoms of a protein incorporating the complex environment.

A HB is influenced by a multitude of factors governing its stability. Intramolecular HBs are one such factor that plays a crucial role.[1,13−16] This fact was already highlighted by Takano et al.[17] in the 90s. Donor−acceptor pairs such as N−H···O, N−H···N, O−H···N, and O−H···O lead to HBs of different strengths due to the involvement of atoms with different electronegativities.[10,18] Another important aspect is the network of HBs,[19,20] where two or more HBs share the same acceptor.[21,22] A study by Ballesteros et al.[23] shows the implications of networks in guiding secondary structures in proteins. A study by Feldblum and Arkin[24] clearly showed the existence of various networks of HBs via both experimental and computational means. The influence of the protein environment and the various factors alter the strength of a

Received: December 22, 2020
Revised: February 21, 2021
Published: March 5, 2021



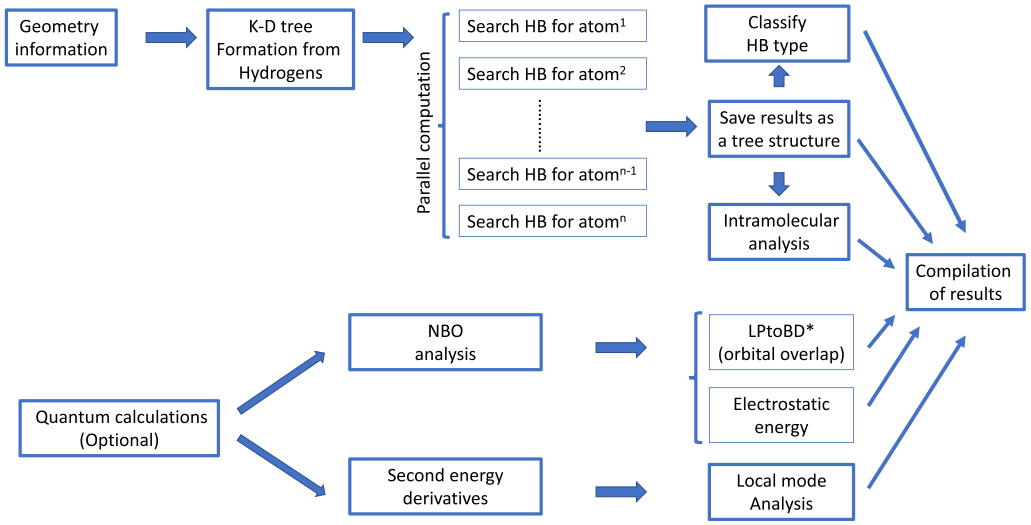


Figure 1. Architecture of EDHB. EDHB utilizes geometry information to construct a K-D tree that captures all hydrogen bonds and their nature in terms of type and inter- and intramolecular hydrogen bonds. Optionally, quantum calculations can be utilized to characterize covalent, electrostatic, and local-mode properties.

HB, and a critical assessment of its strength would bolster our understanding of HBs in a protein.

The strength of a chemical bond or weak chemical interaction such as hydrogen bonding has frequently been assessed via bond dissociation energies (BDE)s,[10,25−28] bond lengths,[29] or bond densities.[30] However, these measures are not directly related to the intrinsic bond strength as they depend on other quantities: the bond dissociation energy on the stabilization energies of the fragments, the bond length on the compressibility limit distance between the atoms, and the bond-stretching frequency on the atom masses.[31] Furthermore, BDEs are limited to intermolecular HBs and are not accessible for intramolecular HBs, which play an important role in proteins. Vibrational spectroscopy can serve in this situation as a perfect alternative provided we utilize *local vibrational modes* (LVM), which were originally introduced by Konkoli and Cremer.[32] A comprehensive review is provided by Kraka et al.[33] LVMs have turned out to be an excellent tool for quantifying the strength of any chemical bond or weak chemical interaction, including HBs.[19,34−44]

To systematically analyze the influencing factors of a HB, we developed a python-based program called *Efficient Detection of Hydrogen Bonding* (EDHB). EDHB can efficiently detect all HBs in a protein based on the geometry (nearest neighbor search[45]) and, in addition, classify them on the basis of secondary structure information, atom types involved, intramolecular ring size, and network of HBs. Furthermore, EDHB can be optionally utilized to explore and analyze HB properties such as LVMs and electrostatic/covalent contributions if data are available from a preceding quantum chemical calculation.

We note that the C−H group acts as a HB donor in proteins.[46] Weak polarization in a C−H bond leads the H atom to be slightly positively charged, creating possibilities for HB interactions. However, as shown by Hubbard and Kamran Haider,[1] C−H-based HBs have only a minor contribution to the overall stability of a protein structure. Likewise, N−H···S HBs also show minor implications for the protein structure. Therefore, in this work, we focused on strong HBs in proteins involving N and O as the donor or acceptor. Intramolecular HBs involving five-membered rings for N−H···O, C−H···O, and N−H···N type interactions are termed "unconventional" HBs.[47−49] In this work only conventional HBs were considered.

The paper is organized in the following way. First, we describe the architecture of EDHB and the computational tools utilized for the study. The information collected by EDHB is presented and discussed in the Results and Discussion section. Conclusions and a future outlook are provided in the Conclusions section.

METHODOLOGY

Architecture of EDHB. Figure 1 shows the architecture of EDHB. Based on the geometry of a protein given in atomic coordinates, EDHB forms a K-dimension tree (K-D tree[45]) from the coordinates of each hydrogen. For each HB donor (N, F, and O), the nearest hydrogen is searched which is characterized on the basis of the distance and angle. For larger systems (no. of atoms >50 000), the computation is done in parallel to reduce the computational time. A graph is formed and saved as a tree structure from which information such as HB type (a network of HBs) and the intramolecular ring size is extracted. Optionally, if data from quantum calculations are available from an NBO analysis and second energy derivatives, EDHB classifies orbital overlap energy (covalent character), electrostatic energies, and local-mode force constants (a measure of HB strength).

Detection of Hydrogen Bonds. HBs are detected by EDHB on the basis of geometry. Only three atoms involved as a donor (D)/acceptor (A) pair are considered so far, i.e., nitrogen, oxygen, and fluorine. Various references have used various criteria for HB length and angle.[1,50−52] Based on literature data, we constrained the distance between the hydrogen (H) and the A atom to 1.6−2.4 Å (Figure 2). The D−H ···A angle is constrained to a range of 90−180°. The K-D tree algorithm[45] was implemented to identify all possible HBs. The K-D tree is a space-partitioning data structure where space is divided into nonoverlapping regions until each point lies exactly in one of the regions. The K-D tree works like a binary tree[45] where every node is a *k*-dimensional point. The last node in each node is the leaf node which contains a point. Every other node generates two planes that divide the space

https://dx.doi.org/10.1021/acs.jpcb.0c11392
J. Phys. Chem. B 2021, 125, 2551−2565

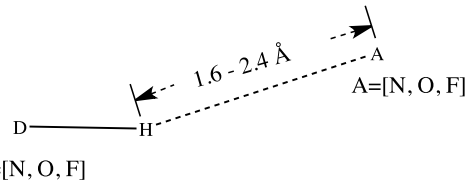


Figure 2. Criteria for the identification of hydrogen bonds. The donor, acceptor can be [N, O, F]. The angle between donor−hydrogen−acceptor is constrained to a range of 90−180°.

into two parts where in the left node will be all points lesser than and on the right will be points greater than the actual value of the division plane. For example (for 3D data), if the "x" axis is chosen for a particular split, all points with a smaller x value will appear on the left node, whereas the larger x values will appear on the right node. The same procedure is applied to the y axis and then to the z axis. This process is then repeated recursively until the leaf node is reached for every point. A representation for a two-dimensional (2D) data structure and its decomposition as a K-D tree is shown in Figure 3. The first split starts at (0, 0) and splits the x axis.

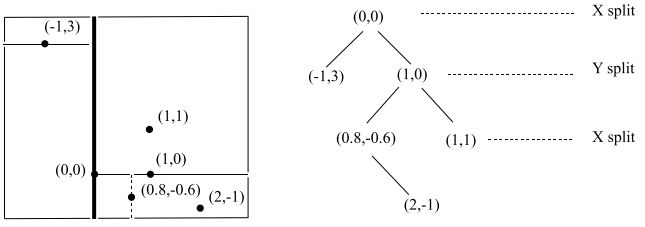


Figure 3. Iterative division of a 2D space to represent a K-D tree. The plots shown in the box are points in the 2D data structure. The K-D tree is shown (right) which is formed by alternative iterative division of the axis in the data.

(Note: the axis and splitting point are chosen by the "sliding midpoint" rule, which ensures that the cells do not all become long and thin. (0, 0) is chosen here for simplicity). All points with negative x values lie to the left of the node, and all points with positive x values to the right. A recursive split is performed for y followed by x and so on until the leaf node is reached. Utilizing this method, all HBs of a protein can be detected with at most $O(k\cdot\log(n))$ time complexity where k is the number of electronegative atoms (O, N, or F) and n is the number of hydrogen atoms.

Computational Time for EDHB. To validate the performance of the K-D tree, we compared the results with classical methods of detecting HBs using naive nested loop and vectorized approaches. The nested loop is formed by running three nested loops, one for the donor, one for the hydrogen, and one for the acceptor. The vectorized approach is implemented and utilized by most of the currently available software that detects HBs such as MDTraj.[54] Table 1 shows a comparison of the timings between the nested loop, vectorized, and K-D tree approach (an exhaustive comparison of 50 different proteins is provided as Table S2). All computations were run on an Intel(R) Xeon(R) CPU E5-2680 v4 @ 2.40 GHz with a 64-bit op-mode. Geometry information for the proteins investigated was taken from the Protein Data Bank[53] for proteins characterized by NMR solution. Proteins that took more than 30 min for the nested approach were not considered and are shown by "−". For a fair comparison, the K-D tree, vectorized, and nested approachs all were not parallelized.

When comparing the timings for even a small protein such as 1UAO with just 138 atoms, the K-D tree is almost 1000 times faster compared to nested and 4 times faster compared to vectorized, respectively. As the number of atoms increases, the performance of the nested approach highly degrades. On the other hand, the K-D tree is able to perform the same computation in less than a second. For the protein with PDB ID 2BI5 (2189 atoms), the K-D tree approach is almost 10^6 times faster than the nested approach. For proteins as large as 52 000 atoms, the K-D tree is able to perform the computation in just 0.2 s compared to the vectorized method which takes 309 s (1755 times faster).

Intramolecular Hydrogen Bond Search. The intramolecular analysis is performed via breadth first search[55] (BFS) in EDHB. BFS is an algorithm searching path in a network of connections. The algorithm first traverses to all the neighbors in the first layer (Figure 4) while keeping track of path and history via backtracking, significantly reducing the computational time. The process is repeated recursively until the goal state is found.

As an example, Figure 4 shows the representation of the seven-membered intramolecular hydrogen-bonded (N−H···O type) ring. To calculate the size of the ring, the initial and the goal state are provided, i.e., O11 and H1, respectively. BFS searches for the H1 layer by layer, and when the goal state is reached it returns the path (O11−C10−N9−C6−C3−N2−H1) from which the ring size can be computed.

Table 1. Timing Comparison of EDHB

PDB ID	nested (s)	vectorized (s)	KD tree (s)	HBs	Natoms	ratio (nested)	ratio (vectorized)
1UAO	0.47	0.002	0.000	6	138	1 177	4
1HD6	23.47	0.020	0.001	39	526	16 552	14
2LDZ	364.51	0.079	0.006	54	977	64 801	14
2EOT	267.94	0.105	0.003	48	1 205	82 847	33
2M20	1269.26	0.294	0.005	96	2 046	247 221	57
2BL5	1817.29	0.379	0.006	149	2 189	295 929	68
5JTN	–	12.602	0.033	696	11 068	–	378
2KU1	–	71.505	0.080	1 869	25 935	–	894
4BY9	–	308.782	0.176	3 877	52 476	–	1755

[a]The protein geometries were taken from the Protein Data Bank (PDB).[53] Nested, vectorized, and K-D-tree refer to the time taken by nested, vectorized, and K-D-tree approaches, respectively. HBs refer to the number of HBs detected. Natoms refers to the number of atoms in the protein. Ratio refers to the ratio of time taken by the nested/vectorized approach to the K-D-tree approach.

https://dx.doi.org/10.1021/acs.jpcb.0c11392
J. Phys. Chem. B 2021, 125, 2551−2565

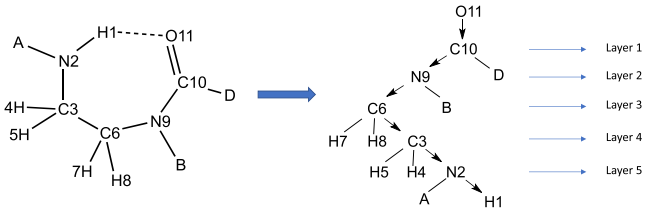


Figure 4. Representation of a breadth first search (BFS). On the left is a seven-membered intramolecular hydrogen bonding. On the right is shown the process of how BFS traverses layer by layer in a network of connections to reach the goal state (H1) from the initial state (O11).

Hydrogen Bond Network Classification. The strength of a HB is influenced by the presence/absence of other HBs that share the same D or A.[20] Figure 5 shows a hydrogen-bonding network. The type of HB network can be defined by the coordination number of the D, A, or hydrogen(s) from the D as a string

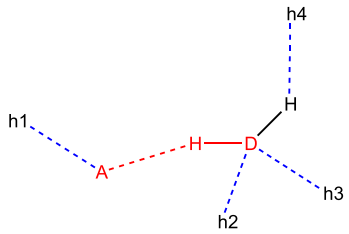


Figure 5. Network of a HB. The D−H···A shown in red is the targeted HB. The dashed line in blue represents possible HBs near the targeted HB.

$$HB_{\text{type}} \equiv A_c - D_c - DH_c \tag{1}$$

where A_c, D_c, and DH_c represent the number of HBs from A excluding targeted HB, number of HB(s) from D, and number of HB(s) from the covalently bonded hydrogen(s) to D, respectively. Following this procedure, the HB network type shown in Figure 5 is 1−2−1.

Hydrogen Bond Characterization. EDHB characterizes HBs based on two properties: vibrational spectroscopy and NBO analysis data. Vibrational spectroscopy data are utilized when the second energy derivative for a molecule is available to calculate local-mode force constants k^a, obtained from the pure local stretching vibration of a HB. The resultant force constants provide the strength of a HB. The electrostatic and covalent contributions are estimated based on results obtained from the NBO analysis.

Local-Mode Force Constants. The vibrational spectra of a molecule hold enormous information about the structure and properties of the molecular interactions. We utilize vibrational spectra to obtain the relative strengths of HBs. We first perform a normal-mode analysis by the basic equation of vibrational spectroscopy formulated as

$$\mathbf{f}^x\mathbf{L} = \mathbf{ML}\Lambda \tag{2}$$

where $\mathbf{L}$ collects the vibrational eigenvectors $\mathbf{l}_\mu$ and the diagonal eigenvalue matrix Λ contains N_{vib} vibrational eigenvalues $4\pi^2c^2\nu_\mu^2$ (with μ = 1, ..., N_{vib} and $N_{\text{vib}} = 3N - K$; K = 6 for nonlinear molecules and 5 for linear molecules) and K zero eigenvalues corresponding to translations and rotations of the molecule (ν_μ represents the harmonic vibrational frequencies given in cm^{-1} and c is the speed of light), $\mathbf{M}$ is the diagonal mass matrix, and $\mathbf{f}^x$ is the force constant matrix expressed in Cartesian coordinates. The $(3N \times 3N)$ dimensional $\mathbf{L}$ matrix collects the N_{vib} normal-vibrational-mode vectors $\mathbf{l}_\mu$ and K-mode vectors corresponding to translations and rotations.

Expressing eq 2 in internal coordinates $\mathbf{q}$ leads to the Wilson GF formalism with[56,57]

$$\mathbf{F}^q\mathbf{D} = \mathbf{G}^{-1}\mathbf{D}\Lambda \tag{3}$$

where

$$\mathbf{G} = \mathbf{B}\mathbf{M}^{-1}\mathbf{B}^\dagger \tag{4}$$

$$\mathbf{D} = \mathbf{BL} \tag{5}$$

$\mathbf{F}^q$ is the force constant matrix expressed in terms of internal coordinates $\mathbf{q}$. The $(N_{\text{vib}} \times N_{\text{vib}})$ dimensional $\mathbf{D}$ matrix represents each normal-mode vector $\mathbf{d}_\mu$. The rectangular $(N_{\text{vib}} \times 3N)$ dimensional $\mathbf{B}$ matrix contains the first derivatives of the internal coordinates with regard to the Cartesian coordinates, thus connecting both coordinate systems, and the $(N_{\text{vib}} \times N_{\text{vib}})$ dimensional matrix $\mathbf{G}$ is the Wilson $\mathbf{G}$ matrix..[56,57]

In terms of internal coordinates, the Lagrange equations can be formulated as

$$L(\mathbf{q}, \dot{\mathbf{q}}) = \frac{1}{2}\dot{\mathbf{q}}^\dagger\mathbf{G}^{-1}\dot{\mathbf{q}} - \frac{1}{2}\mathbf{q}^\dagger\mathbf{F}^q\mathbf{q} \tag{6}$$

where decoupling of the second part (potential energy) is done by the diagonalization of $\mathbf{F}^q$. However, the first part (kinetic energy) remains coupled due to the matrix $\mathbf{G}$. Konkoli and Cremer[32] solved the problem of mass coupling by introducing the concept of local vibrational modes. The mass-decoupled Euler−Lagrange equation was solved by setting all the atomic masses to zero except those of the molecular fragment (e.g., bond, angle, or dihedral, etc.) carrying out a localized vibration. As shown in their original work, the change in the local displacement of a specific internal coordinate is equivalent to an adiabatic relaxation of the molecule.[32] For any molecular fragment associated with an internal coordinate $\mathbf{q}_\mu$, the corresponding local mode vector $\mathbf{a}_\mu$ is given by

$$\mathbf{a}_\mu = \frac{\mathbf{K}^{-1}\mathbf{d}_\mu^\dagger}{\mathbf{d}_\mu\mathbf{K}^{-1}\mathbf{d}_\mu^\dagger} \tag{7}$$

where $\mathbf{d}_\mu$ is a row vector of matrix $\mathbf{D}$. Matrix $\mathbf{K}$ is the diagonal matrix of force constants k_Q, expressed in normal coordinates Q_μ with

$$\mathbf{F}^Q = \mathbf{K} = \mathbf{L}^\dagger\mathbf{f}^x\mathbf{L} \tag{8}$$

resulting from the Wilson GF formalism shown in eqs 3 and 4.

The local mode force constant k_μ^a corresponding to local mode $\mathbf{a}_\mu$ is obtained by

$$k_\mu^a = \mathbf{a}_\mu^\dagger\mathbf{K}\mathbf{a}_\mu \tag{9}$$

Local vibrational modes have been successfully applied to quantify weak chemical interactions such as hydrogen bonding,[19,34−38] halogen bonding,[39−41] pnicogen bonding,[42−44] chalcogen bonding,[58] and tetrel bonding[59] and to derive new chemical descriptors such as a new aromaticity index[60−62] or a generalized Tolman electronic parameter,[63−65] as well as for the derivation of a generalized Badger Rule[66] and several others new concepts.[66−73]

https://dx.doi.org/10.1021/acs.jpcb.0c11392
J. Phys. Chem. B 2021, 125, 2551−2565

Covalent Characteristics of Hydrogen Bonds. The covalent contribution to the HBs is reinforced by the two electron exchange integrals, a phenomenon specific to electrons where they are assumed to be at multiple places at the same time. These contributions stabilize the overall molecular energies leading to stronger bonds with increasing covalent contributions. We utilized NBO analysis to obtain the overlap energy of the oxygen lone pair n_p with the empty antibonding σ^* orbital of donor hydrogen bond (LPtoBD*) shown in Figure 6. This results in a concomitant polarization of the donor hydrogen bond which is a direct consequence of Bent's rule[74] and can be observed in various bonds such as C−H, Si−H, S−H, N−H, etc.

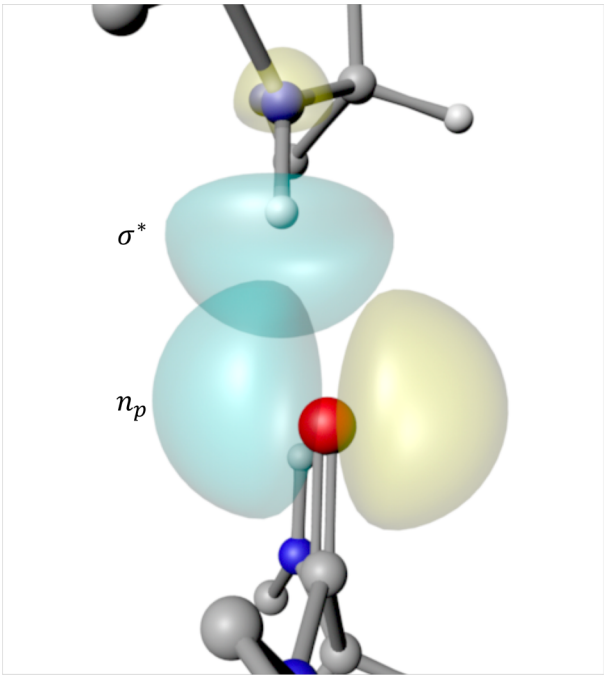


Figure 6. Overlap of the lone pair n_p of the acceptor to the σ^* of the donor hydrogen bond. The overlap energy estimates the covalent characteristics of a hydrogen bond.

Electrostatic Characteristics of Hydrogen Bonds. The electrostatic interaction is caused by the induction of charges in the molecules and can be relatively estimated for a HB as

$$E_c = \frac{q_1^* q_2}{r} \tag{10}$$

where E_c is the electrostatic energy, q_1 and q_2 are the atomic charges of A and H, respectively, and r is the distance between the two atoms.[75] We choose NBO charges as they are more reliable than Mulliken charges[76] since they operate on the basis of electron density. Localized natural atomic orbitals were used to describe the computed electron densities and therefore include bond polarization.

Computational Details. Geometry optimization and normal-mode analysis were carried out in the gas phase using Gaussian 16.[77,78] Eight smaller proteins, 1K43, 1L2Y, 1R4G, 1S4A, 2EVQ, 1UAO, 2O0S, and 5E61, shown in Figure 7 were optimized with the hybrid B3LYP functional.[79−82] High accuracy paired with relatively low computational costs has made B3LYP a popular DFT choice. Kuhn et al.[83] and Bartlett et al.[84] successfully used this functional for the description of small molecules and peptides. We further employed empirical dispersion correction[85,86] as suggested by Fadda et al.[87] to include polarization effects that can significantly impact hydrogen bond stability. Based on the findings of Fadda et al.[87] that B3LYP paired with the 6-31G(d,p) basis[88−94] provides a good description of short alanine peptides, we also used the 6-31G(d,p) in our study. For all geometry optimizations, conversion criteria to a maximum step size of 0.01 au and a RMS force of 0.0017 au were applied. Protein 1R4G was cut, and only the area of major interest was taken. Some side chains were removed from proteins 1K43, 1S4A, and 5E61 for faster calculation (details and geometry information are provided in the Supporting Information as pages S23−S59).

To obtain statistically relevant information in addition, a set of 155 smaller proteins was optimized, and second energy derivatives were calculated with the semiempirical DFT method GFN2-xTB.[95] The GFN2-xTB is significantly faster than B3LYP with a low compromise regarding accuracy. A similar distribution of second-order properties such as vibrational constants for GFN2-xTB compared to B3LYP was observed, which is described in the Results and Discussion section. The similarity between the results of the two methods led to our choice of GFN2-xTB as the DFT method for the statistical analysis. The 155 proteins were selected based on the small size, single chain, and NMR solution as characterizing techniques from the Protein Data Bank.[53] Hydrogen atoms were added utilizing Leap (a tool from Amber[96]), structures were minimized for 1000 steps in Amber,[96] and the overall charges were calculated from Leap prior to the GFN2-xTB calculations. The PDB IDs for all 155 proteins are provided in the Supporting Information on page S2.

The local-mode force constants for each HB were calculated using the LModeA package.[33,97] A description of the corresponding local vibrational mode theory is given below. LPtoBD* and atomic charges were obtained from NBO[98] analysis utilizing optimized geometries with the B3LYP[79−82] level of theory including empirical dispersion correction[85,86] along with the 6-31G(d,p)[88−94] basis set.

RESULTS AND DISCUSSION

This section is organized into subsections providing results and discussion by analyzing patterns for a general overview of HB properties, intramolecular HBs, HB type based on donor−acceptor pairs, networks of HBs, and the governing factors for the HB strength in proteins. We start by comparing predicted HB properties from the two DFT methods to get statistically relevant information. We observed similar patterns of HB properties from the two DFT methods. Thus, we provide a detailed characterization of HB properties from a higher level of DFT method (B3LYP) followed by statistical results from a lower level of DFT method (GFN2-xTB).

Figure 7 shows the 8 proteins investigated with B3LYP, and the PDB IDs for 155 proteins investigated with GFN2-xTB are provided on pages S2−S10. The HBs detected by EDHB are shown in yellow. Each HB is given a specific ID for discussion (larger images of proteins are shown in the Supporting Information as Figures S1−S8). These proteins accommodate a variety of secondary structures which include a majority of HB types observed in proteins.

To generalize the outcomes from GFN2-xTB and B3LYP, we focused on the distribution of HB length and k^a for the eight proteins optimized as shown in Figure 8. The HBs that

https://dx.doi.org/10.1021/acs.jpcb.0c11392
J. Phys. Chem. B 2021, 125, 2551−2565

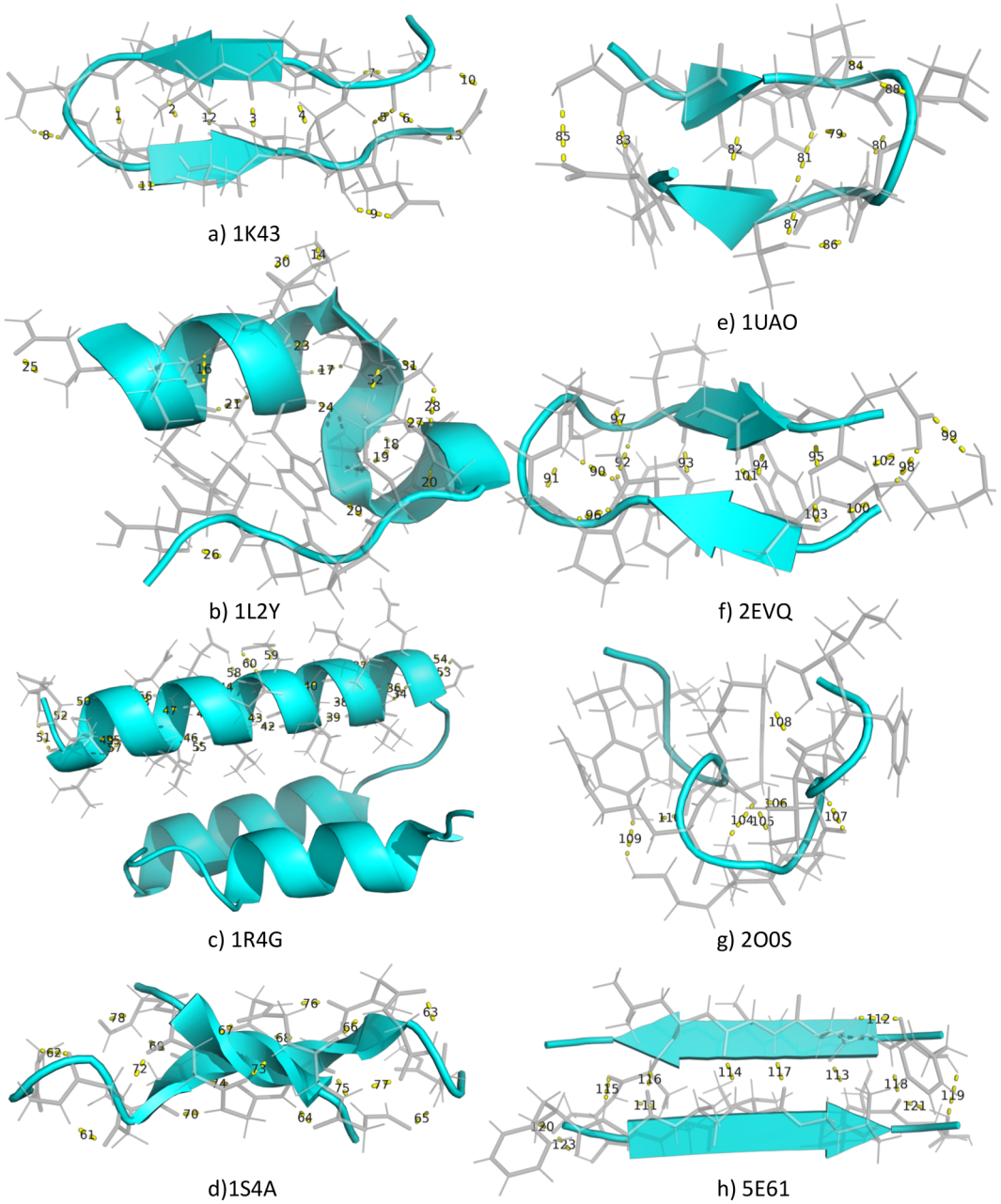


Figure 7. Eight proteins investigated. Each number is shown as a unique ID associated with each HB. (a) 1K43 is a beta hairpin type structure and contains 13 HBs (1−13). (b) A trp-cage motif 1L2Y has α helix and 19 HBs (14−32). (c) The Sendai virus protein 1R4G was cut to keep the ideal α-helix. Other atoms were removed for the convenience of running ab initio calculations. Twenty-eight HBs were detected for this protein (33−60). (d) A double-strand β-helix of a D,L-alternating oligonorleucine 1S4A has 18 HBs (61−78). (e) A designed protein Chignolin 1UAO is a β-sheet type structure and looks more like a hairpin with 10 HBs (79−88). (f) A 12-residue beta hairpin 2EVQ with 15 HBs (89−103). (g) A designed peptide 2O0S is a coil type structure and has 7 HBs (104−110), and the amyloid-forming peptide 5E61 with a near ideal β-strand has 13 HBs (111−123).

were absent in either one of the methods were removed. The GFN2-xTB and B3LYP methods differ, and therefore the optimized geometries are not exactly the same. This is reflected by Figure 8b, where the median of the HB length for B3LYP (1.95 Å) is slightly higher than that for GFN2-xTB (1.92 Å). However, the overall distribution remains the same and is also observed for k^a as shown in Figure 8a. The comparison of LPtoBD*, HB angle, and electrostatic energy is shown in Figure S9 where we observe similar distributions. Based on these results, we conclude that statistically relevant information can be obtained from the GFN2-xTB method.

We analyzed quantitatively the number of different HBs present in 155 proteins based on secondary structures shown in Table 2. Note that EDHB uses data obtained from Dictionary of Protein Secondary Structure (DSSP)[99] for secondary structure classification if the PDB ID of the protein is available. The proteins circumvent a variety of secondary structures with α, turn, bend, and extended strand composing 20.4, 9.3, 8.6, and 6.7%, respectively. Less common HBs such as 3_{10} helix and isolated β-bridge are 2.2 and 0.7%, respectively. The secondary structures with no backbone classification are

https://dx.doi.org/10.1021/acs.jpcb.0c11392
J. Phys. Chem. B 2021, 125, 2551−2565

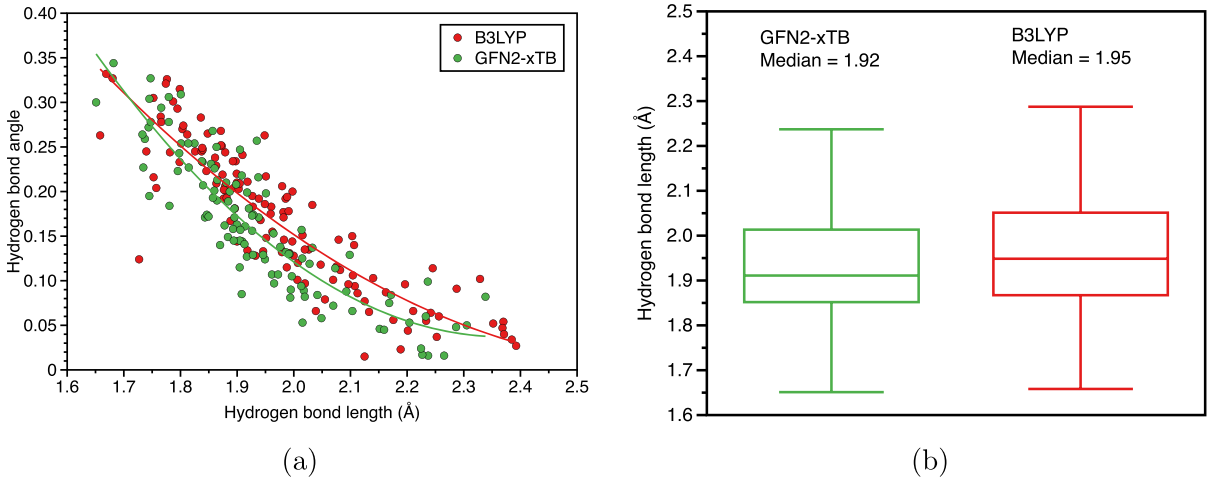


Figure 8. Comparison of B3LYP and GFN2-xTB on eight small proteins. Hydrogen bonds that do not exist in either one of the methods were removed. (a) Plot of k_a vs HB length. The line shows the corresponding quadratic fit. We observe a similar distribution of k^a values from either method. (b) Box plot for the HB length. The median distribution of GFN2-xTB is slightly higher than that of B3LYP.

Table 2. Statistics for Secondary Structure

secondary structures	percentage of HBs	number of HBs
side chain	31.7	674
α	20.4	435
backbone	20.1	427
turn	9.3	207
bend	8.6	182
extended strand	6.7	143
3_{10} helix	2.2	46
isolated β-bridge	0.7	14
total		2128

termed as backbone and involve 20.1% of all HBs. The side chains involve 31.1% of all HBs.

General Trends of Hydrogen Bonds in Proteins. The estimated contribution of an individual HB, in general, is in the range of 2−8 kcal/mol depending on factors such as geometry and environment. It has been suggested that the strength of a HB in a protein depends on the HB length.[29] However, we observe that the strength can vary depending on the environment of the HB. Furthermore, shorter bonds are not always stronger bonds.[68] The environment creates an electric field that perturbs the electrostatic and covalent interactions of the HB. Thus, we analyzed the covalent character by LPtoBD*

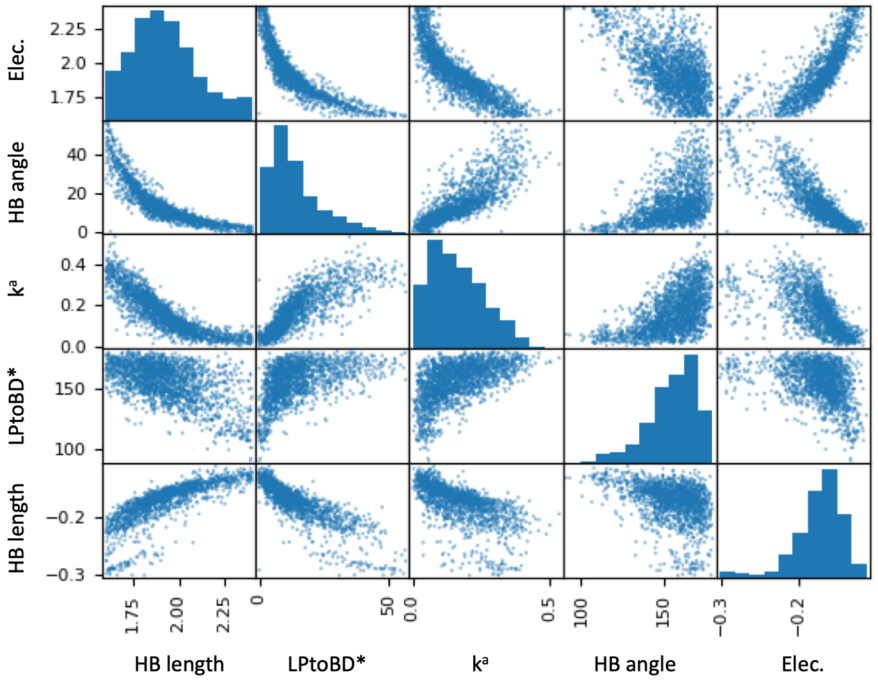


Figure 9. Scatter matrix plot for all continuous data from EDHB results on 155 proteins. HB length, LPtoBD*, k^a, HB angle, and Elec. are the hydrogen bond length (in Å), LPtoBD* (in kcal/mol), k^a for stretching vibration of HB (in mdyn/Å), the angle created by acceptor, hydrogen, and donor, respectively, and the electrostatic energy (in e^2/Å). The diagonal elements of the matrix show the histogram of the distributions. Each of the off-diagonal elements is the plot of x vs y where x, y are the HB properties named as xlabel and ylabel, respectively.

https://dx.doi.org/10.1021/acs.jpcb.0c11392
J. Phys. Chem. B 2021, 125, 2551−2565

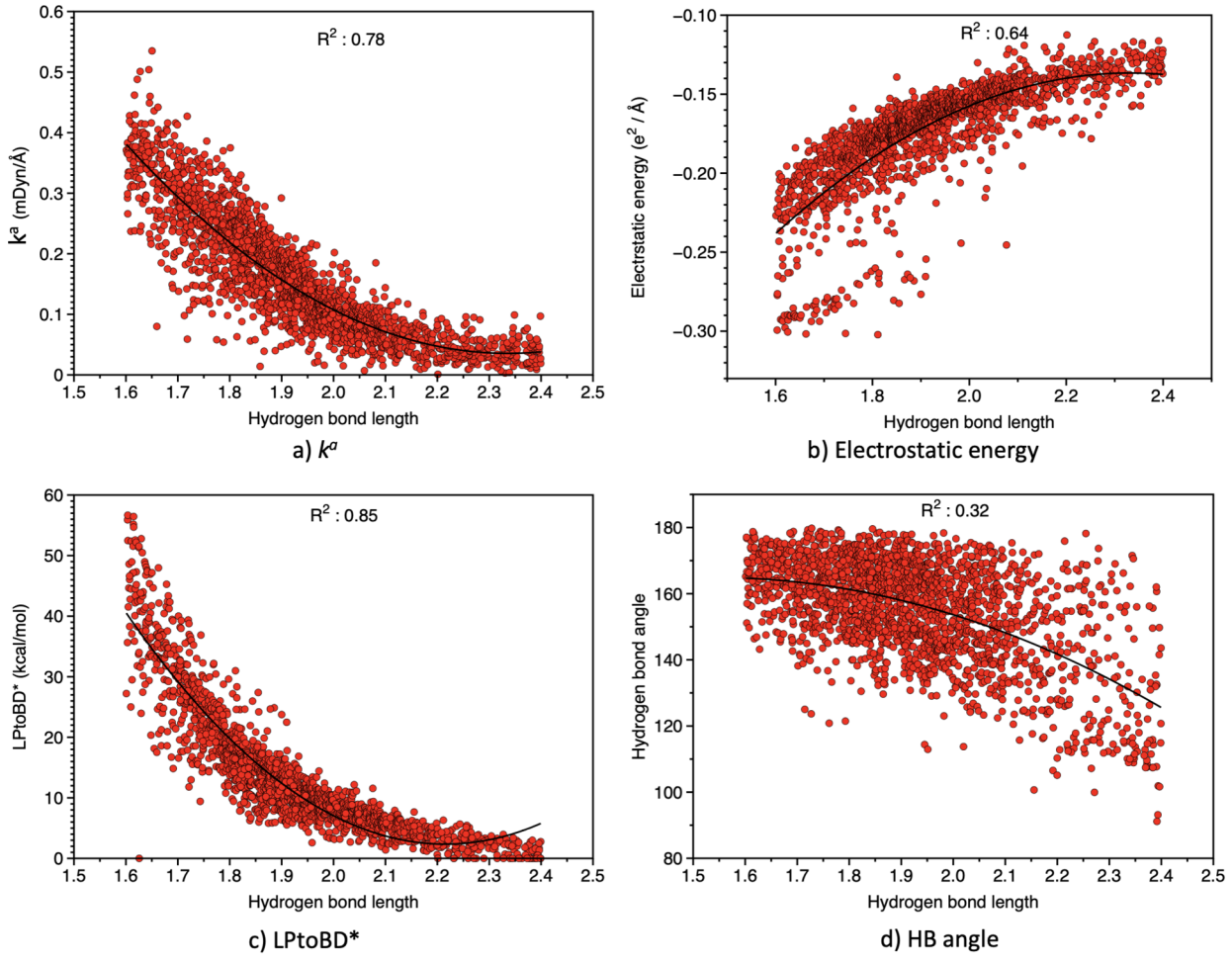


Figure 10. General trends for hydrogen bond properties in proteins relative to hydrogen bond length. The black line shows a quadratic fit; R^2 documents the quality of the fit.

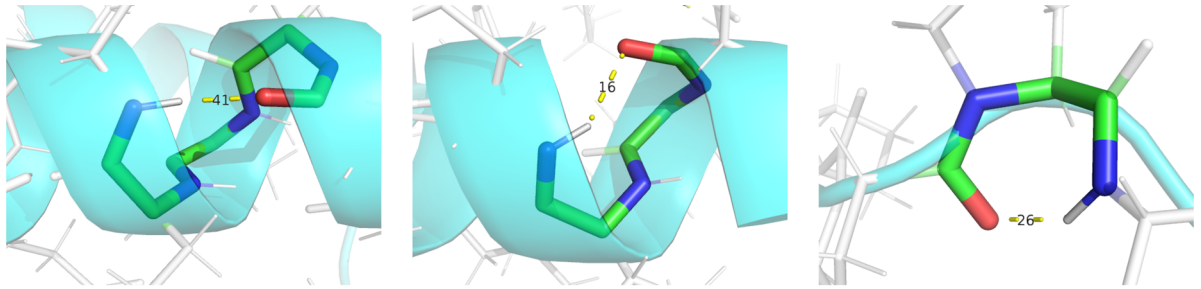


Figure 11. Three most commonly observed intramolecular HBs in proteins. The letter I represents intramolecular HB, and the corresponding number represents the size of the ring. The numbers denote the ID for each HB corresponding to Figure 7.

and the electrostatic character based on the atomic charges of the H and A atoms and the geometry of the HB.

Figure 9 shows a scatter matrix plot for all the continuous data extracted from EDHB: HB length, LPtoBD*, k^a, HB angle, and electrostatic energy. The HB length is majorly distributed around 1.8−2.1 Å, LPtoBD* around 10−20 kcal/mol, k^a around 0.1−0.3 mdyn/Å, HB angle around 150−170 °, and electrostatic energy around −0.2 to −0.3 e^2/Å. We observe a quadratic decay for LPtoBD*, k^a, HB angle, and electrostatic energy relative to HB length. The LPtoBD* and electrostatic energies follow a somewhat linear relationship relative to k^a. However, the HB angle relative to k^a has a larger noise created by geometry constraints in a protein structure. We observe similar patterns when comparing HB properties relative to the HB length or HB properties relative to k^a, LPtoBD*, or electrostatic energy. Thus, we confined ourselves to look into HB properties relative to HB length.

Figure 10 shows the general trends observed in the analysis of 155 proteins. The k^a and LPtoBD* decrease as the HB length increases following an inverse quadratic relationship. The electrostatic energy and HB angle follow similar patterns. We observe large deviations in electrostatic energies as different HB types such as N−H···O, N−H···N, O−H···O, and O−H···N have significant charge differences due to

https://dx.doi.org/10.1021/acs.jpcb.0c11392
J. Phys. Chem. B 2021, 125, 2551−2565

Table 3. Intramolecular HBs

ID	bond	HB length (Å)	elec. (e^2/Å)[a]	HB angle (deg)	k^a (mdyn/Å)	intra.[b]	LPtoBD* (kcal/mol)	chain[c]	HB network
41	N−H···O	1.90	−0.17	157	0.21	I(13)	13.0	b−b	0−0−0
16	N−H···O	2.13	−0.14	143	0.07	I(10)	1.48	b−b	0−0−0
26	N−H···O	2.00	−0.16	146	0.14	I(7)	6.16	b−b	0−0−0

[a]Elec. refers to the electrostatic energies. [b]Intra. refers to intramolecular HB type with ring size enclosed in the bracket. [c]Chain is composed of a donor−acceptor pair as either from s (side chain) or b (backbone).

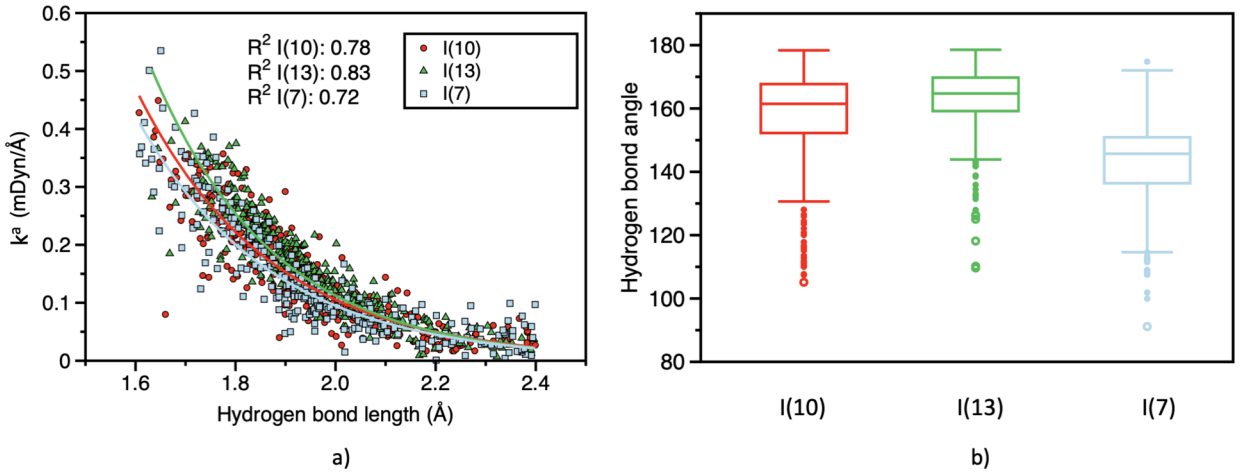


Figure 12. Statistics of three major intramolecular HBs for (a) k^a and (b) HB angle. The lines in (a) shows the quadratic fit curve for the corresponding intramolecular hydrogen bonds. The plots show that the hydrogen bond stability follows the order I(13) > I(10) > I(7) with respect to the same HB length.

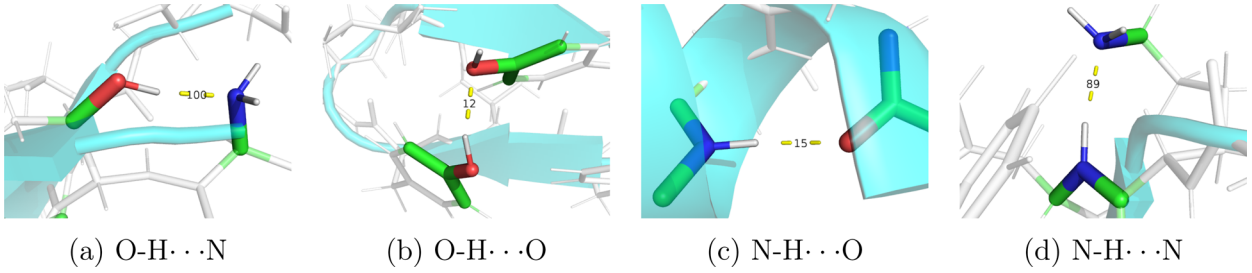


Figure 13. Various HB types based on the donor−acceptor pairs. The numbers denote the HB ID corresponding to Figure 7.

variations in the electronegativity for N and O atoms. The N−H···N and O−H···N HBs have a high electrostatic character accompanied by a short hydrogen bond length. In general, N as the acceptor increases the electrostatic character. When acting as the donor, O, being more electronegative compared to N, makes the corresponding H more positively charged. This results in a strong electrostatic interaction between the H and the acceptor. Thus, we observe an island of data points that correspond to N−H···N and O−H···N type hydrogen bonds as shown in Figure 10b. The HB angle follows a similar pattern; however, it has significant noise that might be caused due to the geometric constraints of a protein structure.

Intramolecular Hydrogen Bonds. Figure 11 shows the three most commonly observed intramolecular HBs, the properties of which are summarized in Table 3 and are frequently observed in the backbone. Figure 11a shows I(13) (intramolecular HB with a ring size of 13 atoms) that forms a HB observed in the α-helix. The orientation of the four consecutive amino acids (R_n−R_{n+4}, where R_n is the residue at the nth position in the sequence of the protein) makes the HB almost linear, significantly increasing the LPtoBD*. Figure 11b shows I(10) formed in between R_n−R_{n+3}. A study by Perutz[100] in the 1950s confirmed the presence of I(10) in 3_{10} helixes. However, we observe I(10) over all backbone types, majorly in α-helixes and in smaller amounts from β-sheets and coil. I(10) has a larger deviation in the HB angle depending on its orientation, exhibiting variation in electrostatic energy, LPtoBD*, and HB angle. Figure 11c shows I(7) formed in between R_n−R_{n+2}. Contrary to I(13) and I(10), I(7) is not just confined to the backbone but extends to side chains as well.

Further investigation from the 155 proteins provided statistically relevant data. I(10) was the most commonly observed intramolecular HB constituting 18.6% of all intramolecular HBs, followed by I(13) (16.1%) and I(7) (13.0%). The majority of I(13) is observed in α-helices constituting 43.7% of all I(13). The percentage of other intramolecular HBs contributing more than 3% is provided in Table S3.

Figure 12a shows k^a for I(10), I(13), and I(7) relative to HB lengths. I(13) have a higher mean of HB angle as shown in Figure 12b leading to an increase in the LPtoBD* and therefore k^a. This is reflected by the exponential fit curve that tends to be consistently higher than the exponential fit for I(10) and I(7). We observe I(7) have lower k^a values relative to HB length due to the smaller HB angle (Figure 12b),

https://dx.doi.org/10.1021/acs.jpcb.0c11392
J. Phys. Chem. B 2021, 125, 2551−2565

Table 4. Various HB Types Based on Donor−Acceptor Pairs

ID	bond	HB length (Å)	elec. (e^2/Å)[a]	HB angle (deg)	k^a (mdyn/Å)	intra.[b]	LPtoBD* (kcal/mol)	chain[c]	HB network
89	N−H ···N	2.01	−0.21	167.0	0.12	I(9)	17.25	s−b	0−0−0
15	N−H···O	1.96	−0.16	168.3	0.16	I(10)	10.15	b−b	0−0−1
12	O−H···O	1.86	−0.20	155.7	0.21	I(29)	15.04	b−s	0−0−0
100	O−H···N	1.95	−0.25	166.4	0.15	I(8)	19.6	b−s	0−1−1

[a]Elec. refers to the electrostatic energies. [b]Intra. refers to intramolecular HB type with ring size enclosed in the bracket. [c]Chain is composed of a donor−acceptor pair as either from s (side chain) or b (backbone).

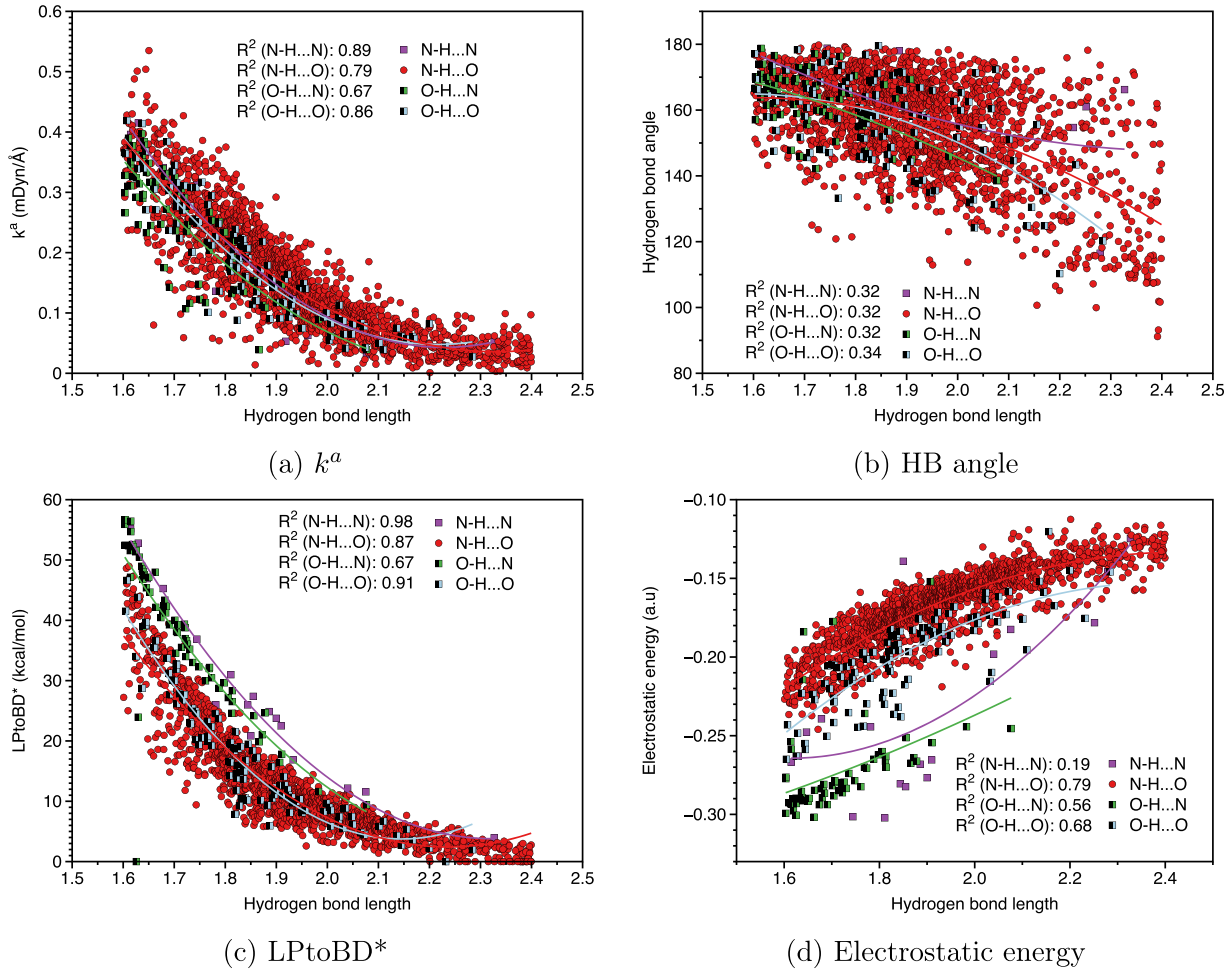


Figure 14. Statistics for HB properties based on donor−acceptor pairs. The black line shows a quadratic fit, and R^2 documents the quality of the fit.

decreasing the LPtoBD*. We further observed I(7) have a near-linear relationship for HB angle and HB length as shown in Figure S10b. The HB angle tends to decrease as the HB length increases. The HB angle for I(10) varies significantly and follows no pattern relative to HB length as shown in Figure S10. This might be the reason we observe I(10) in between I(13) and I(7).

Hydrogen Bond Type. Figure 13 and Table 4 show the HBs based on donor−acceptor pair and their computed properties, respectively. The N−H···N and O−H···N HBs have a larger electrostatic character as O is more electronegative compared to N acting as a donor and therefore makes the corresponding H more positively charged. The LPtoBD* is also larger for O compared to N as the acceptor. The k^a values however do not follow such a pattern. This can be attributed to the way k^a is computed that considers not just the atoms involved in a HB but the environment as well, which has a varying level of effect on the electronic distribution near a HB.

The N−H ···O type is the leading HBs covering 89.1% of all HBs observed from the analysis of 155 proteins. The HBs where both the donor and acceptor are formed by the backbone of the protein cover around 68% of all N−H···O HBs while the side chain covers 32%, signifying a larger contribution to protein stability coming from backbone HBs. The results agree with Pace et al.[3] who showed 65% of HB contribution from the backbone and 35% from the side chains. Further, HBs with donor−acceptor pairs both from the backbone contribute around 12.6% for α-helix−α-helix and 8.03% from extended strand−extended strand, suggesting higher stability from the N−H ···O α-helix type HBs.[3] The second major contribution comes from O−H ···O HBs covering 6.3% of all HBs. All the O−H···O HBs are from side chains where the O−H functional group majorly involves

https://dx.doi.org/10.1021/acs.jpcb.0c11392
J. Phys. Chem. B 2021, 125, 2551−2565

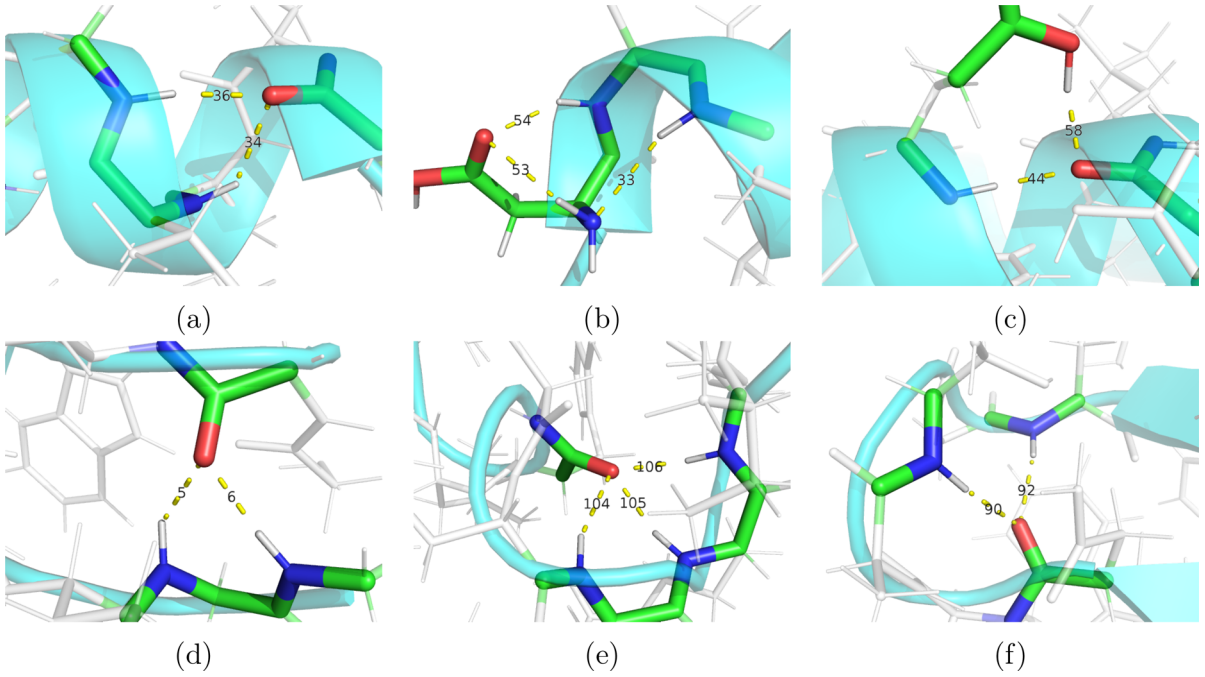


Figure 15. Various scenarios of HBs leading to various HB networks derived from eq 1. The network 1−0−0 is shown in four different plots: (a), (c), (d) and (f). The network 1−1−0 is shown as (b), and the network 2−0−0 is shown as (e).

Table 5. Various HB Properties Based on the Network of a HB

no.	ID	bond	HB length (Å)	elec. (e^2/Å)[a]	HB angle (deg)	k^a (mdyn/Å)	intra.[b]	LPtoBD* (kcal/mol)	chain[c]	HB network
1	34	N−H ···O	2.19	−0.14	135.5	0.02	I(10)	0.9	b−b	1−0−0
	36	N−H···O	2.04	−0.15	164.8	0.07	I(13)	9.07	b−b	1−0−0
2	33	N−H···N	2.39	−0.17	154.2	0.03	I(8)	4.44	b−b	0−0−1
	53	N−H···O	2.25	−0.11	126.5	0.11	I(6)	1.59	s−b	1−1−0
	54	N−H···O	1.9	−0.15	149.5	0.22	I(7)	12.05	s−b	1−0−0
3	44	N−H···O	2	−0.17	158.5	0.2	I(13)	8.93	b−b	1−0−0
	58	O−H···O	1.77	−0.21	167.2	0.28	I(18)	12.15	b−s	1−0−0
4	5	N−H···O	2.2	−0.15	128.8	0.04	I(34)	1.35	b−b	1−0−0
	6	N−H···O	1.83	−0.18	174.8	0.25	I(37)	21.58	b−b	1−0−0
5	104	N−H···O	2.29	−0.14	160	0.09	I(10)	0.97	b−b	2−0−0
	105	N−H···O	1.95	−0.16	165.7	0.26	I(13)	10.83	b−b	2−0−0
	106	N−H···O	2.02	−0.16	150.7	0.14	I(16)	6.06	b−b	2−0−0
6	90	N−H···O	2.14	−0.15	162.6	0.1	I(10)	3.01	b−b	1−0−0
	92	N−H···O	2.1	−0.15	163.4	0.15	I(16)	6.28	b−b	1−0−0

[a]Elec. refers to the electrostatic energies. [b]Intra. refers to intramolecular HB type with ring size enclosed in the bracket. [c]Chain is composed of a donor−acceptor pair as either from s (side chain) or b (backbone).

the residues SER (44.2%), THR (40.0%), and TYR (8.6%) as donors. The contributions from O−H···N and N−H···N are 3.4% and 1.1% respectively. The small fraction of O−H···N HBs is majorly dominated by side chain HBs (90.8%) compared to backbone HBs (9.2%). However, the small fraction of N−H···N bonds are either from the side chains (44.0%) or the backbone HBs (56.0%) signifying almost equal contribution from both the side chain and the backbone.

Figure 14 shows the distribution of k^a, HB angle, LPtoBD*, and electrostatic energy relative to the HB length for the proteins investigated with GFN2-xTB. In general, O−H ···N HBs have the lowest quadratic fit curve signifying lower HB strength relative to HB length when compared with the quadratic fit curve for other HBs. However, the HB strength of N−H···O, N−H···N, and O−H···O does not seem to have significant differences. The distribution in the HB angle has a similar pattern for all the HBs. The LPtoBD* has a higher quadratic fit curve for N−H···N and O−H···N HBs relative to HB length and so does the electrostatic energy. The O−H···O and N−H···O HBs do not have significant differences. However, N−H···O HBs have the lowest electrostatic energy compared to other HB types.

Hydrogen Bond Networks. HBs in general are comprised of single acceptor and donor pair. However, there are instances where two or more HBs share the same acceptor forming a network of HBs. These HBs have been analyzed,[21,22] and their implications for the bending of helices have been discussed.[23] The most common example can be found in α-helices where a side chain interacts with the backbone amide carbonyl group.[22] However, these HBs are also commonly observed at the start or the end of β-sheets. A study by Feldblum and Arkin[24] clearly shows the existence of HB networks with both experimental and computational means and highlighted that

https://dx.doi.org/10.1021/acs.jpcb.0c11392
J. Phys. Chem. B 2021, 125, 2551−2565

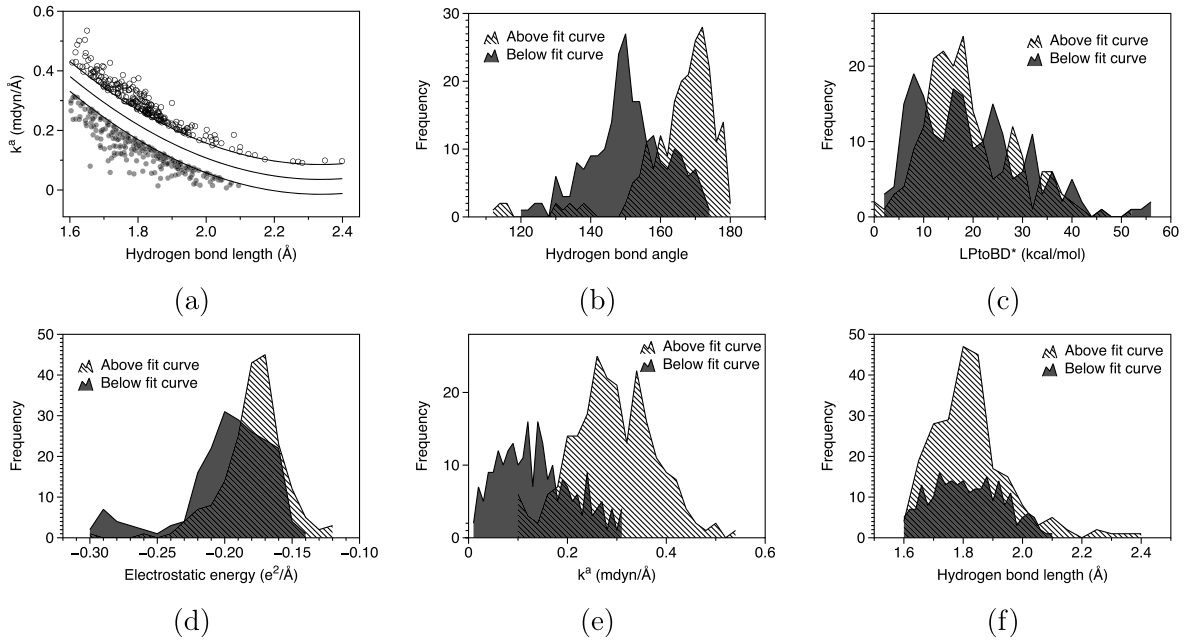


Figure 16. Quadratic fit curve for all HBs relative to HB length is shifted by ±0.05 to obtain two fit curves. The two-shifted fit curve characterizes the strong and weak HBs for relatively the same HB length: (a) shifted fit curves of HBs based on k^a denoting strong (above fit curve) and weak HBs (below fit curve). (b–f) histograms for each HB property generated for these HBs.

such HBs do not necessarily decrease the strength of the canonical HB (the HB with the shortest HB length).

Figure 15 and Table 5 show six different networks of HBs in different scenarios. Figure 15a, 15d, and 15f show a 1−0−0 (eq 1) network of HBs in α-helix, extended strand, and turn, respectively. The HB angle for the two HBs varies from 128° to 165°. The HB with a higher angle has higher LPtoBD* signifying more covalent characteristics. The electrostatic energy for both HBs remains almost equivalent. However, the covalent contribution to the linear HB makes it stronger as accepted by k^a values. A similar setup with an additional HB from the donor of a canonical HB (1−1−0) is shown in Figure 15b. Although the angle is lower compared to 1−0−0 as shown in Figure 15a, there is a significant LPtoBD*. This is a consequence of the extended electron density causing a higher concentration of electrons for the canonical HB. We observe a raise in covalent energy for the canonical HB and thus in its strength. The electrostatic energies do not have significant differences. Figure 15c shows the network 1−0−0 with one of the HB as O−H···O from a side chain. The LPtoBD* is similar to 1−1−0 as shown in Figure 15b. The electronegative nature of O atom increases the electrostatic energy. The overall contribution of electrostatic and covalent energies leads to similar k^a values. Figure 15e shows three HBs with the same acceptor as network 2−0−0. Despite the HB length of 1.95 Å for the canonical HB, LPtoBD* is relatively high (10.83 kcal/mol) which contributes to its high k^a value of 0.26 mdyn/Å. The elongated electron density of the three HBs stabilizes the energy and therefore guides the protein to form a unique secondary structure.

As stated by Feldblum and Arkin,[24] the network of HBs does not necessarily decrease the strength of canonical HBs. We thus observe no pattern in the strength among various networks of HBs. The relationship of k^a, LPtoBD*, HB angle, and electrostatic energy is shown in Figure S11, where the quadratic fit curve relative to the HB length overlap for various networks.

What Governs Hydrogen Bond Stability in Proteins?

In this section, we detail the statistical results obtained for properties of a HB based on the angle and amino acids, in order to investigate what drives strong HBs in proteins. Figure 16a shows the quadratic fit curve (middle) for all HBs relative to the HB length. We shifted the curve to ±0.05 to obtain two fit curves. The two-shifted fit curve characterizes the strong and weak HBs (based on k^a values) for the same HB length. Thus, the points above the +0.05-shifted fit curve should correspond to strong HBs while those below the −0.05-shifted curve should correspond to weak HBs relative to the HB length. Figure 16b–f show the histogram for the two data points: above the fit curve and below the fit curve for the HB angle, LPtoBD*, electrostatic energy, k^a, and HB length. Except for k^a and HB angle, other HB properties tend to overlap, denoting no significant differences. The HB angle thus plays a strong role in distinguishing strong and weak HBs where we observe a clear distinction for above fit curve points and below fit curve points.

We further analyzed the amino acids involved in hydrogen bonding. ARG is the major donor for HBs contributing to 24% of all HBs followed by LYS (13%). Both have N in the side chain to act as an HB donor. Similarly, they act as major acceptors contributing to 10% and 9% for ARG and LYS, respectively. The contribution of other residues is shown in Table S4. Table S5 shows the contribution of amino acids in terms of a network of HBs and intramolecular HBs. However, no intuitive relationship was found that clearly distinguishes amino acids based on HB properties.

CONCLUSIONS

In this work, we developed a new software, *Efficient Detection of Hydrogen Bonds* (EDHB), that systematically detects HBs based on a nearest neighbors algorithm. EDHB classifies inter-

https://dx.doi.org/10.1021/acs.jpcb.0c11392
J. Phys. Chem. B 2021, 125, 2551−2565

and intramolecular HBs as well as HB networks. We could show that EDHB outperforms other commonly used methods to detect HBs in terms of speed of execution. An important feature of EDHB is that information from preceding quantum chemical studies (i.e., NBO data and second energy derivative information) can be used to derive the electrostatic/covalent character of the HBs and to calculate local-mode force constants as a quantitative measure of the intrinsic HB strength via the local-mode analysis program LModeA.[33]

Utilizing EDHB, we analyzed various HB properties in a protein system for a diverse set of 163 proteins. General trends of HB strengths follow an inverse quadratic relationship with the HB length as observed by local-mode force constants, electrostatic energies, and covalent characteristics. The most common intramolecular HB is observed to be I(13) followed by I(10) and I(7). In general, the strength of I(13) is higher than that of I(10) followed by I(7). Analyzing various HBs based on donor−acceptor pairs highlights higher electrostatic and covalent characters for N−H···N and O−H···N type HBs. Networks of HBs are common in protein structures and play an important role in specific secondary structure orientations such as α-helixes and turns; however, they do not significantly influence the HB strength. A comprehensive analysis of the distribution of local-mode force constants comes to the interesting conclusion that the HB angle is the governing factor determining the HB strength.

All EDHB scripts are provided at https://github.com/ekraka/EDHB. We further note that a linux binary file is also available for quick implementation on high-performance computers (HPC) and servers. The applicability of EDHB in discerning patterns of hydrogen bonding is extendable but not limited to protein−ligand interaction, protein folding, protein−protein interaction, molecular dynamics, water, ice clusters, and improving force fields for proteins. EDHB can be extended to any molecular system provided the geometric information is known, reaching out to a broad computational chemistry audience.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcb.0c11392.

Coordinates of the optimized geometry for the eight proteins; PDB IDs for 155 proteins that were optimized via GFN2-xTB; enlarged pictures of the eight proteins; EDHB time comparison; EDHB output for the eight proteins; various statistical plots made from the output from the GFN2-xTB method (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Elfi Kraka − *Department of Chemistry, Southern Methodist University, Dallas, Texas, United States;* orcid.org/0000-0002-9658-5626; Email: ekraka@gmail.com

Authors

Niraj Verma − *Department of Chemistry, Southern Methodist University, Dallas, Texas, United States*

Yunwen Tao − *Department of Chemistry, Southern Methodist University, Dallas, Texas, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcb.0c11392

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank SMU for its generous supercomputer resources. This work was financially supported by the National Science Foundation Grants CHE 1464906 and Dean's Research Council Fellowship, Southern Methodist University.

■ REFERENCES

(1) Hubbard, R. E.; Kamran Haider, M. *Encyclopedia of Life Sciences*; John Wiley & Sons, Ltd.: Chichester, U.K., 2010.

(2) Thapa, B.; Raghavachari, K. Energy Decomposition Analysis of Protein-Ligand Interactions Using Molecules-in-Molecules Fragmentation-Based Method. *J. Chem. Inf. Model.* **2019**, *59*, 3474−3484.

(3) Pace, C. N.; Fu, H.; Fryar, K. L.; Landua, J.; Trevino, S. R.; Schell, D.; Thurlkill, R. L.; Imura, S.; Scholtz, J. M.; Gajiwala, K.; et al. Contribution of Hydrogen Bonds to Protein Stability. *Protein Sci.* **2014**, *23*, 652−661.

(4) Mirsky, A. E.; Pauling, L. On the Structure of Native, Denatured, and Coagulated Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **1936**, *22*, 439−447.

(5) Pauling, L.; Corey, R. B.; Branson, H. R. The Structure of Proteins: Two Hydrogen-Bonded Helical Configurations of the Polypeptide Chain. *Proc. Natl. Acad. Sci. U. S. A.* **1951**, *37*, 205−211.

(6) Pauling, L.; Corey, R. B. Configurations of Polypeptide Chains With Favored Orientations Around Single Bonds: Two New Pleated Sheets. *Proc. Natl. Acad. Sci. U. S. A.* **1951**, *37*, 729−740.

(7) Sippl, M. J.; Ortner, M.; Jaritz, M.; Lackner, P.; Flöckner, H. Helmholtz Free Energies of Atom Pair Interactions in Proteins. *Folding Des.* **1996**, *1*, 289−298.

(8) Campos, L. A.; Cuesta-López, S.; López-Llano, J.; Falo, F.; Sancho, J. A Double-Deletion Method to Quantifying Incremental Binding Energies in Proteins from Experiment: Example of a Destabilizing Hydrogen Bonding Pair. *Biophys. J.* **2005**, *88*, 1311−1321.

(9) Herschlag, D.; Pinney, M. M. Hydrogen Bonds: Simple after All? *Biochemistry* **2018**, *57*, 3338−3352.

(10) Scheiner, S. Contributions of NH···O and CH···O Hydrogen Bonds to the Stability of β-Sheets in Proteins. *J. Phys. Chem. B* **2006**, *110*, 18670−18679.

(11) Pace, C. N.; Scholtz, J. M.; Grimsley, G. R. Forces Stabilizing Proteins. *FEBS Lett.* **2014**, *588*, 2177−2184.

(12) Fried, S. D.; Boxer, S. G. Measuring Electric Fields and Noncovalent Interactions Using the Vibrational Stark Effect. *Acc. Chem. Res.* **2015**, *48*, 998−1006.

(13) Sticke, D. F.; Presta, L. G.; Dill, K. A.; Rose, G. D. Hydrogen Bonding in Globular Proteins. *J. Mol. Biol.* **1992**, *226*, 1143−1159.

(14) Alber, T.; Dao-Pin, S.; Wilson, K.; Wozniak, J. A.; Cook, S. P.; Matthews, B. W. Contributions of Hydrogen Bonds of Thr 157 to the Thermodynamic Stability of Phage T4 Lysozyme. *Nature* **1987**, *330*, 41−46.

(15) Green, S. M.; Meeker, A. K.; Shortle, D. Contributions of the Polar, Uncharged Amino Acids to the Stability of Staphylococcal Nuclease: Evidence for Mutational Effects on the free Energy of the Denatured State. *Biochemistry* **1992**, *31*, 5717−5728.

(16) Serrano, L.; Kellis, J. T.; Cann, P.; Matouschek, A.; Fersht, A. R. The Folding of an Enzyme. *J. Mol. Biol.* **1992**, *224*, 783−804.

(17) Takano, K.; Yamagata, Y.; Funahashi, J.; Hioki, Y.; Kuramitsu, S.; Yutani, K. Contribution of Intra- and Intermolecular Hydrogen Bonds to the Conformational Stability of Human Lysozyme. *Biochemistry* **1999**, *38*, 12698−12708.

(18) Sarkhel, S.; Desiraju, G. R. N-H···O, O-H···O, and C-H···O Hydrogen Bonds in Protein-Ligand Complexes: Strong and Weak Interactions in Molecular Recognition. *Proteins: Struct., Funct., Genet.* **2004**, *54*, 247−259.

(19) Tao, Y.; Zou, W.; Jia, J.; Li, W.; Cremer, D. Different Ways of Hydrogen Bonding in Water - Why Does Warm Water Freeze Faster than Cold Water? *J. Chem. Theory Comput.* **2017**, *13*, 55−76.
(20) Tao, Y.; Zou, W.; Kraka, E. Strengthening of Hydrogen Bonding with the Push-Pull Effect. *Chem. Phys. Lett.* **2017**, *685*, 251−258.
(21) Baker, E.; Hubbard, R. Hydrogen Bonding in Globular Proteins. *Prog. Biophys. Mol. Biol.* **1984**, *44*, 97−179.
(22) Gray, T.; Matthews, B. Intrahelical Hydrogen Bonding of Serine, Threonine and Cysteine Residues within α-Helices and its Relevance to Membrane-bound Proteins. *J. Mol. Biol.* **1984**, *175*, 75−81.
(23) Ballesteros, J. A.; Deupi, X.; Olivella, M.; Haaksma, E. E.; Pardo, L. Serine and Threonine Residues Bend α-Helices in the χ1 = g- Conformation. *Biophys. J.* **2000**, *79*, 2754−2760.
(24) Feldblum, E. S.; Arkin, I. T. Strength of a Bifurcated H bond. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 4085−4090.
(25) Alkorta, I.; Elguero, J.; Frontera, A. Not Only Hydrogen Bonds: Other Noncovalent Interactions. *Crystals* **2020**, *10*, 180.
(26) Summers, T. J.; Daniel, B. P.; Cheng, Q.; DeYonker, N. J. Quantifying Protein Contact Networks through Interaction Energies. *J. Chem. Inf. Model.* **2019**, *59*, 5034−5044.
(27) Lucarini, M.; Pedrielli, P.; Pedulli, G. F.; Cabiddu, S.; Fattuoni, C. Bond Dissociation Energies of O-H Bonds in Substituted Phenols from Equilibration Studies. *J. Org. Chem.* **1996**, *61*, 9259−9263.
(28) Deepak, R. N. V. K.; Sankararamakrishnan, R. Unconventional N-H···N Hydrogen Bonds Involving Proline Backbone Nitrogen in Protein Structures. *Biophys. J.* **2016**, *110*, 1967−1979.
(29) Remer, L. C.; Jensen, J. H. Toward a General Theory of Hydrogen Bonding: The Short, Strong Hydrogen Bond [HOH···OH]. *J. Phys. Chem. A* **2000**, *104*, 9266−9275.
(30) Klein, R. A. Electron Density Topological Analysis of Hydrogen Bonding in Glucopyranose and Hydrated Glucopyranose. *J. Am. Chem. Soc.* **2002**, *124*, 13931−13937.
(31) Cremer, D.; Kraka, E. From Molecular Vibrations to Bonding, Chemical Reactions, and Reaction Mechanism. *Curr. Org. Chem.* **2010**, *14*, 1524−1560.
(32) Konkoli, Z.; Cremer, D. A New Way of Analyzing Vibrational Spectra. I. Derivation of Adiabatic Internal Modes. *Int. J. Quantum Chem.* **1998**, *67*, 1−9.
(33) Kraka, E.; Zou, W.; Tao, Y. Decoding Chemical Information from Vibrational Spectroscopy Data: Local Vibrational Mode Theory. *WIREs: Comput. Mol. Sci.* **2020**, *10* (5), No. e1480.
(34) Kraka, E.; Freindorf, M.; Cremer, D. Chiral Discrimination by Vibrational Spectroscopy Utilizing Local Modes. *Chirality* **2013**, *25*, 185−196.
(35) Zhang, X.; Dai, H.; Yan, H.; Zou, W.; Cremer, D. B-H□□□π Interaction: A New Type of Nonclassical Hydrogen Bonding. *J. Am. Chem. Soc.* **2016**, *138*, 4334−4337.
(36) Kalescky, R.; Kraka, E.; Cremer, D. Local Vibrational Modes of the Formic Acid Dimer - The Strength of the Double Hydrogen Bond. *Mol. Phys.* **2013**, *111*, 1497−1510.
(37) Kalescky, R.; Zou, W.; Kraka, E.; Cremer, D. Local Vibrational Modes of the Water Dimer - Comparison of Theory and Experiment. *Chem. Phys. Lett.* **2012**, *554*, 243−247.
(38) Freindorf, M.; Kraka, E.; Cremer, D. A Comprehensive Analysis of Hydrogen Bond Interactions Based on Local Vibrational Modes. *Int. J. Quantum Chem.* **2012**, *112*, 3174−3187.
(39) Oliveira, V.; Kraka, E.; Cremer, D. The Intrinsic Strength of the Halogen Bond: Electrostatic and Covalent Contributions Described by Coupled Cluster Theory. *Phys. Chem. Chem. Phys.* **2016**, *18*, 33031−33046.
(40) Oliveira, V.; Kraka, E.; Cremer, D. Quantitative Assessment of Halogen Bonding Utilizing Vibrational Spectroscopy. *Inorg. Chem.* **2017**, *56*, 488−502.
(41) Oliveira, V.; Cremer, D. Transition from Metal-ligand Bonding to Halogen Bonding Involving a Metal as Halogen Acceptor a Study of Cu, Ag, Au, Pt, and Hg complexes. *Chem. Phys. Lett.* **2017**, *681*, 56−63.
(42) Setiawan, D.; Kraka, E.; Cremer, D. Strength of the Pnicogen Bond in Complexes Involving Group Va Elements N, P, and As. *J. Phys. Chem. A* **2015**, *119*, 1642−1656.
(43) Setiawan, D.; Kraka, E.; Cremer, D. Description of Pnicogen Bonding with the help of Vibrational Spectroscopy - The Missing Link Between Theory and Experiment. *Chem. Phys. Lett.* **2014**, *614*, 136−142.
(44) Setiawan, D.; Cremer, D. Super-pnicogen Bonding in the Radical Anion of the Fluorophosphine Dimer. *Chem. Phys. Lett.* **2016**, *662*, 182−187.
(45) Samet, H. *Fault Detection and Diagnosis in Industrial Systems*, 1st ed.; Addison-Wesley: Boston, MA, 1990.
(46) Derewenda, Z. S.; Lee, L.; Derewenda, U. The Occurence of C-H ··· O Hydrogen Bonds in Proteins. *J. Mol. Biol.* **1995**, *252*, 248−262.
(47) Adhikary, R.; Zimmermann, J.; Liu, J.; Forrest, R. P.; Janicki, T. D.; Dawson, P. E.; Corcelli, S. A.; Romesberg, F. E. Evidence of an Unusual N-H···N Hydrogen Bond in Proteins. *J. Am. Chem. Soc.* **2014**, *136*, 13474−13477.
(48) Newberry, R. W.; Raines, R. T. A Prevalent Intraresidue Hydrogen Bond Stabilizes Proteins. *Nat. Chem. Biol.* **2016**, *12*, 1084−1088.
(49) Mejía, S.; Hernández-Pérez, J. M.; Sandoval-Lira, J.; Sartillo-Piscil, F. Looking Inside the Intramolecular C-H···O Hydrogen Bond in Lactams Derived from a-Methylbenzylamine. *Molecules* **2017**, *22*, 361.
(50) Jeffrey, G. A. *An Introduction to Hydrogen Bonding*; Oxford University Press: Oxford, U.K., 1997; p 303.
(51) Langkilde, A.; Kristensen, S. M.; Lo Leggio, L.; Mølgaard, A.; Jensen, J. H.; Houk, A. R.; Navarro Poulsen, J. C.; Kauppinen, S.; Larsen, S. Short Strong Hydrogen Bonds in Proteins: A Case Study of Rhamnogalacturonan Acetylesterase. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2008**, *D64*, 851−63.
(52) Harris, T. K.; Mildvan, A. S. High-Precision Measurement of Hydrogen Bond Lengths in Proteins by Nuclear Magnetic Resonance Methods. *Proteins: Struct., Funct., Genet.* **1999**, *35*, 275−282.
(53) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(54) McGibbon, R. T.; Beauchamp, K. A.; Harrigan, M. P.; Klein, C.; Swails, J. M.; Hernández, C. X.; Schwantes, C. R.; Wang, L.-P.; Lane, T. J.; Pande, V. S. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* **2015**, *109*, 1528−1532.
(55) Cormen, T. H.; Leiserson, C. E.; Rivest, R. L.; Stein, C. *Introduction to Algorithms*, 2nd ed.; MIT Press: Cambridge, MA, 2001; pp 531−539.
(56) Wilson, E. B.; Decius, J. C.; Cross, P. C. *Molecular Vibrations*; McGraw-Hill: New York; 1955.
(57) Wilson, E. B., Jr. A Method of Obtaining the Expanded Secular Equation for the Vibration Frequencies of a Molecule. *J. Chem. Phys.* **1939**, *7*, 1047.
(58) Oliveira, V.; Cremer, D.; Kraka, E. The Many Facets of Chalcogen Bonding: Described by Vibrational Spectroscopy. *J. Phys. Chem. A* **2017**, *121*, 6845−6862.
(59) Sethio, D.; Oliveira, V.; Kraka, E. Quantitative Assessment of Tetrel Bonding Utilizing Vibrational Spectroscopy. *Molecules* **2018**, *23*, 2763.
(60) Kalescky, R.; Kraka, E.; Cremer, D. Description of Aromaticity with the Help of Vibrational Spectroscopy: Anthracene and Phenanthrene. *J. Phys. Chem. A* **2014**, *118*, 223−237.
(61) Setiawan, D.; Kraka, E.; Cremer, D. Quantitative Assessment of Aromaticity and Antiaromaticity Utilizing Vibrational Spectroscopy. *J. Org. Chem.* **2016**, *81*, 9669−9686.
(62) Li, Y.; Oliveira, V.; Tang, C.; Cremer, D.; Liu, C.; Ma, J. The Peculiar Role of the Au3 Unit in Au_m Clusters: σ-Aromaticity of the $Au5Zn^+$ Ion. *Inorg. Chem.* **2017**, *56*, 5793−5803.
(63) Cremer, D.; Kraka, E. Generalization of the Tolman Electronic Parameter: The Metal-Ligand Electronic Parameter and the Intrinsic

https://dx.doi.org/10.1021/acs.jpcb.0c11392
J. Phys. Chem. B 2021, 125, 2551−2565

Strength of the Metal-Ligand Bond. *Dalton Transac.* **2017**, *46*, 8323−8338.
(64) Kalescky, R.; Kraka, E.; Cremer, D. New Approach to Tolman's Electronic Parameter Based on Local Vibrational Modes. *Inorg. Chem.* **2014**, *53*, 478−495.
(65) Setiawan, D.; Kalescky, R.; Kraka, E.; Cremer, D. Direct Measure of Metal-Ligand Bonding Replacing the Tolman Electronic Parameter. *Inorg. Chem.* **2016**, *55*, 2332−2344.
(66) Kraka, E.; Larsson, J. A.; Cremer, D. In *Computational Spectroscopy: Methods, Experiments and Applications*; Grunenberg, J., Ed.; Wiley: New York, 2010.
(67) Kalescky, R.; Kraka, E.; Cremer, D. Identification of the Strongest Bonds in Chemistry. *J. Phys. Chem. A* **2013**, *117*, 8981−8995.
(68) Kraka, E.; Setiawan, D.; Cremer, D. Re-evaluation of the Bond Length - Bond Strength Rule: The Stronger Bond is not Always the Shorter Bond. *J. Comput. Chem.* **2016**, *37*, 130−142.
(69) Setiawan, D.; Kraka, E.; Cremer, D. Hidden Bond Anomalies: The Peculiar Case of the Fluorinated Amine Chalcogenides. *J. Phys. Chem. A* **2015**, *119*, 9541−9556.
(70) Humason, A.; Zou, W.; Cremer, D. 11,11-Dimethyl-1,6-methano[10]annulene-An Annulene with an Ultralong CC Bond or a Fluxional Molecule? *J. Phys. Chem. A* **2015**, *119*, 1666−1682.
(71) Kalescky, R.; Zou, W.; Kraka, E.; Cremer, D. Quantitative Assessment of the Multiplicity of Carbon-Halogen Bonds: Carbenium and Halonium Ions with F, Cl, Br, and I. *J. Phys. Chem. A* **2014**, *118*, 1948−1963.
(72) Kraka, E.; Cremer, D. Characterization of CF Bonds with Multiple-Bond Character: Bond Lengths, Stretching Force Constants, and Bond Dissociation Energies. *ChemPhysChem* **2009**, *10*, 686−698.
(73) Verma, N.; Tao, Y.; Marcial, B. L.; Kraka, E. Correlation Between Molecular Acidity (pKa) and Vibrational Spectroscopy. *J. Mol. Model.* **2019**, *25*, 48.
(74) Bent, H. A. An Appraisal of Valance-bond Structures and Hybridization in Compounds of the First-row Elements. *Chem. Rev.* **1961**, *61*, 275−311.
(75) Halliday, D.; Resnick, R.; Walker, J. *Fundamentals of Physics*, 5th ed.; Wiley: New York, 1997.
(76) Mulliken, R. S. Electronic Population Analysis on LCAO-MO Molecular Wave Functions. I. *J. Chem. Phys.* **1955**, *23*, 1833−1840.
(77) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; et al. *Gaussian 16*, revision A.03; Gaussian Inc.: Wallingford, CT, 2016.
(78) Dennington, R.; Keith, T. A.; Millam, J. M. *Gauss View*, version 5.0.9; Semichem Inc.: Shawnee Mission, KS, 2016.
(79) Vosko, S. H.; Wilk, L.; Nusair, M. Accurate Spin-dependent Electron Liquid Correlation Energies for Local Spin Density Calculations: A Critical Analysis. *Can. J. Phys.* **1980**, *58*, 1200−1211.
(80) Frisch, M. J. *Gaussian 09*, revision A.1; Gaussian, Inc.; Wallingford, CT, 2009. (b) Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623.
(81) Lee, C.; Yang, W.; Parr, R. G. Development of the Colle-Salvetti Correlation-energy Formula into a Functional of the Electron Density. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1988**, *37*, 785−789.
(82) Becke, A. D. Density-functional Thermochemistry. III.. The Role of Exact Exchange. *J. Chem. Phys.* **1993**, *98*, 5648−5652.
(83) Kuhn, B.; Mohr, P.; Stahl, M. Intramolecular Hydrogen Bonding in Medicinal Chemistry. *J. Med. Chem.* **2010**, *53*, 2601−2611.
(84) Bartlett, G. J.; Newberry, R. W.; VanVeller, B.; Raines, R. T.; Woolfson, D. N. Interplay of Hydrogen Bonds and n-pi* Interactions in Proteins. *J. Am. Chem. Soc.* **2013**, *135*, 18682−18688.
(85) Chai, J.-D.; Head-Gordon, M. Long-range Corrected Hybrid Density Functionals with Damped Atom-atom Dispersion Corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615.
(86) Chai, J.-D.; Head-Gordon, M. Systematic Optimization of Long-range Corrected Hybrid Density Functionals. *J. Chem. Phys.* **2008**, *128*, 084106.
(87) Fadda, E.; Woods, R. J. Contribution of the Empirical Dispersion Correction on the Conformation of Short Alanine Peptides Obtained by Gas-phase QM Calculations. *Can. J. Chem.* **2013**, *91*, 859−865.
(88) Rassolov, V. A.; Ratner, M. A.; Pople, J. A.; Redfern, P. C.; Curtiss, L. A. 6-31G* Basis Set for Third-row Atoms. *J. Comput. Chem.* **2001**, *22*, 976−984.
(89) Rassolov, V. A.; Pople, J. A.; Ratner, M. A.; Windus, T. L. 6 - 31G□ basis set for atoms K through Zn. *J. Chem. Phys.* **1998**, *109*, 1223−1229.
(90) Binning, R. C.; Curtiss, L. A. Compact contracted basis sets for third-row atoms: Ga-Kr. *J. Comput. Chem.* **1990**, *11*, 1206−1216.
(91) Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; DeFrees, D. J.; Pople, J. A. Self-Consistent Molecular Orbital Methods. XXIII. A Polarization-type Basis set for Second-Row Elements. *J. Chem. Phys.* **1982**, *77*, 3654−3665.
(92) Hariharan, P. C.; Pople, J. A. The Influence of Polarization Functions on Molecular Orbital Hydrogenation Energies. *Theoretica Chimica Acta* **1973**, *28*, 213−222.
(93) Hehre, W. J.; Ditchfield, R.; Pople, J. A. Self-Consistent Molecular Orbital Methods. XII.. Further Extensions of Gaussian-Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1972**, *56*, 2257−2261.
(94) Ditchfield, R.; Hehre, W. J.; Pople, J. A. Self-Consistent Molecular-Orbital Methods. IX.. An Extended Gaussian-Type Basis for Molecular-Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1971**, *54*, 724−728.
(95) Bannwarth, C.; Caldeweyher, E.; Ehlert, S.; Hansen, A.; Pracht, P.; Seibert, J.; Spicher, S.; Grimme, S. Extended Tight-Binding Quantum Chemistry Methods. *WIREs Comput. Mol. Sci.* **2020**, *11* (2), e1493.
(96) Case, D. A.; Belfon, K.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E., III; et al. *AMBER 2018*; University of California, San Francisco, CA, 2018.
(97) Zou, W.; Tao, Y.; Freindorf, M.; Makoś, M. Z.; Verma, N.; Kraka, E. *Local Vibrational Mode Analysis (LModeA)*; Computational and Theoretical Chemistry Group (CATCO), Southern Methodist University: Dallas, TX, 2020.
(98) Pounds, A. J. Valency and Bonding: A Natural Bond Orbital Donor-Acceptor Perspective (Frank Weinhold and Clark Landis). *J. Chem. Educ.* **2007**, *84*, 43.
(99) Kabsch, W.; Sander, C. Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-bonded and Geometrical Features. *Biopolymers* **1983**, *22*, 2577−2637.
(100) Perutz, M. F. New X-Ray Evidence on the Configuration of Polypeptide Chains: Polypeptide Chains in Poly-γ-benzyl-L-glutamate, Keratin and Hæmoglobin. *Nature* **1951**, *167*, 1053−1054.

https://dx.doi.org/10.1021/acs.jpcb.0c11392
J. Phys. Chem. B 2021, 125, 2551−2565