

Strength of a bifurcated H bond

Esther S. Feldblum and Isaiah T. Arkin¹

Department of Biological Chemistry, The Alexander Silberman Institute of Life Sciences, The Hebrew University of Jerusalem, Edmund J. Safra Campus, Jerusalem 91904, Israel

Edited by Roger D. Kornberg, Stanford University School of Medicine, Stanford, CA, and approved February 5, 2014 (received for review October 22, 2013)

Macromolecules are characterized by their particular arrangement of H bonds. Many of these interactions involve a single donor and acceptor pair, such as the regular H-bonding pattern between carbonyl oxygens and amide H⁺s four residues apart in α -helices. The H-bonding potential of some acceptors, however, leads to the phenomenon of overcoordination between two donors and one acceptor. Herein, using isotope-edited Fourier transform infrared measurements and density functional theory (DFT) calculations, we measured the strength of such bifurcated H bonds in a transmembrane α -helix. Frequency shifts of the ¹³C=¹⁸O amide I mode were used as a reporter of the strength of the bifurcated H bond from a thiol and hydroxyl H⁺ at residue $i + 4$. DFT calculations yielded very similar frequency shifts and an energy of -2.6 and -3.4 kcal/mol for the thiol and hydroxyl bifurcated H bonds, respectively. The strength of the intrahelical bifurcated H bond is consistent with its prevalence in hydrophobic environments and is shown to significantly impact side-chain rotamer distribution.

protein structure | FTIR | membrane proteins

Hydrogen bonding is one of the characteristic features of biomacromolecular structures. In nucleic acids, H bonding enables specific and complementary base pairing between the two strands of the double helix (1). In proteins, the particular H-bonding pattern defines the typical secondary structure motifs. This enabled Pauling et al. and Eisenberg to propose the existence of α -helices and β -sheets even before these structures were determined experimentally (2–4).

In α -helices, a regular pattern of H bonds exists between an amide carbonyl of residue i (the acceptor) and the amide H⁺ donor of residue $i + 4$ ($C_i=O \cdots H-N_{i+4}$). Various techniques were used to measure the energetics of such canonical H bonds with results that vary, depending on the particular environment. The strength of isolated H bonds in the gas phase, or in a hydrophobic environment are in the range of 4–5 kcal/mol (or perhaps even higher), and in water, the value reduces substantially to 0.5–1.5 kcal/mol (5–14).

These conventional backbone H bonds normally comprise a single acceptor and donor pair. However, there are instances in which the H-bonding pattern may include more than one donor or acceptor. Such bonds are known as bifurcated H bonds and may describe an instance in which one H-bond donor is bound to two H-bond acceptors (e.g., $C=O \cdots H \cdots O=C$). Another possibility entails the opposite configuration, in which two H-bond donors are bound to a single H-bond acceptor (e.g., $N-H \cdots O \cdots H-N$). Such an acceptor is said to be overcoordinated, and an example in proteins was originally noted by Kendrew (15). The prevalence of such bifurcated H bonds in proteins has previously been analyzed (16, 17) and has even been implicated in bending the helices (18).

A particularly common example of a bifurcated H-bonding system takes place in α -helices between side-chain hydroxyl or thiol groups and a backbone amide carbonyl (17, 19). In particular, both the backbone amide H⁺ and the side-chain hydroxyl (or thiol) of residue $i + 4$ are H bonded to the backbone carbonyl oxygen of residue i ($N_{i+4}-H \cdots O_i \cdots H-O_{i+4}$). In terms of nomenclature, we refer to the “additional” H bond between the side-chain hydroxyl (or thiol) H⁺ of residue $i + 4$ and the

backbone C=O of residue i as the bifurcated H bond. The formation of the bifurcated H bond is only attained when the χ_1 side-chain dihedral angle is +gauche (17). Finally, Engelman and Steitz have suggested that the formation of this particular bifurcated H-bond pattern may facilitate the incorporation and accommodation of the mildly polar threonine and serine residues in the hydrophobic environment of the lipid bilayer (20). Therefore, the purpose of this study was to quantitatively examine bifurcated H bonds in transmembrane α -helices. Toward this end, we have taken both experimental and computational routes, arriving at self-consistent results.

Experimentally, we examined two independent transmembrane systems by isotope-edited Fourier transform infrared (FTIR) spectroscopy: the M2 protein from influenza A and the Severe Acute Respiratory Syndrome coronavirus (SARS) E protein. The homotetrameric M2 H⁺ channel from influenza A has been a subject of extensive characterization (21–28). Its structure consists of a four-helix bundle surrounding an aqueous pore that allows H⁺ permeation into the viral lumen. SARS-coronavirus E protein also serves as a suitable model system to investigate bifurcated H bonds because its transmembrane domain was shown to be largely α -helical (29, 30) and embedded in the lipid bilayer (29). Because its residues are located in the hydrophobic environment of the lipid bilayer, water molecules are not expected to substantially interfere with intramolecular H bonding.

In M2, the stretching vibrations of C=O groups with bifurcated H bonds were compared with those without bifurcated H bonds at different locations in the same peptide sequence (see Fig. 14). This experimental setup is used to prove the occurrence of bifurcation and to give a range for experimental measurements that may be influenced by location-dependent factors, such as environmental polarity. In contrast, the E protein system was investigated at a single residue location. The C=O of Phe26 is normally H bonded to the backbone amide H⁺ of Thr30, as well as to the hydroxyl side chain of Thr30. We then mutate Thr30 to other naturally occurring amino acids to investigate this

Significance

Hydrogen bonding is one of the characteristic features of macromolecules. Therefore, it is not surprising that extensive research has gone into understanding the energetics and importance of hydrogen bonding. All of these studies focused on hydrogen bonds that involve a single donor and acceptor pair. Yet the H-bonding potential of many acceptors leads to the phenomenon of overcoordination between two donors and one acceptor. Here we have used both experimental approaches and computational analyses to measure the strength of such bifurcated hydrogen bonds and show that their energy is between 60% and 50% of canonical H bonds. Finally, we show how the energetics of bifurcated H bonds directly impact amino acid side-chain structure.

Author contributions: E.S.F. and I.T.A. designed research; E.S.F. performed research; E.S.F. and I.T.A. analyzed data; and E.S.F. and I.T.A. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

¹To whom correspondence should be addressed. E-mail: arkin@huji.ac.il.

bifurcation. This experimental setup allows us to investigate the strength of bifurcated H bonds in a location-independent manner. See Fig. 2A for schematic representations of the bifurcated H bonds in cysteine, serine, and threonine residues.

In parallel, a computational approach using density functional theory (DFT) was used to examine a bifurcated H-bonding configuration. We performed energetic and vibrational frequency analyses comparing systems in which a carbonyl group was H bonded to donors capable of donating one or two H⁺s. Such calculations enabled us to calculate the strength of the additional H bond along with the predicted vibrational frequency shift in its presence. Two factors validated the accuracy of these calculations: (i) The similarity between the predicted shift in vibrational frequency and that which was measured experimentally. (ii) The close calculation of the strength of the canonical H bond.

Finally, the impact of bifurcated H bonds on protein structure is analyzed by examining side-chain dihedral angles. We show a clear correlation between rotamer preference and solvent exposure. Rotamers that facilitate bifurcated H-bond formation are preferred in solvent-excluded regions despite steric preferences.

Results and Discussion

To investigate and measure the strength of bifurcated H bonds, we use both experimental and computational approaches, yielding empirical and theoretical data. We start by describing the experimental results based on isotope-edited FTIR spectroscopy and then show that DFT calculations yield very similar results that allow us to compute the energy of the bifurcated H bond. Finally, using statistical analysis of side-chain rotamer distribution, we show the impact of the bifurcated H bond on protein structure.

Experimental Analysis. Vibrational spectroscopy, a tool that is particularly sensitive to H bonding, was used to examine bifurcated H bonds. The amide I vibrational mode of the C=O group to which the two protons are bound, was used as a reporter of the bifurcated H-bond strength. Specifically, the vibrational frequency of a C=O group that is bound to one H⁺ is expected to change when it is bound to two H⁺s (31). The acceptor carbonyl group has a different equilibrium bond length when it is involved in a bifurcated H bond compared with a single canonical H bond. This in turn affects the dipole moment of that bond, which causes the quantum energy levels of the C=O bond vibration to differ in the two cases (bifurcated and canonical). These differing energy levels imply that the C=O group will absorb different wavelengths of light. Thus, an FTIR shift is seen in the amide I stretch.

An individual amide group can be examined in the background of the entire protein–lipid background using isotope editing with the ¹³C=¹⁸O label (32, 33). Subsequently, the extent of the isotope-edited shift can be used to gauge the strength of a particular additional H bond. Later we used DFT calculations to compute the frequency shift and compared it to that measured by FTIR.

As stated above, two transmembrane peptides were used: influenza M2 and SARS coronavirus E protein. In the M2 peptide, the isotope-edited amide I modes of 10 different amino acids were examined. As shown in Fig. 1, all isotope-edited amide I modes, with the exception of Val27, were centered around $1,597 \pm 4$ cm⁻¹. The small variation in peak position is important to note as it gives an uncertainty range that might be explained due to differences in environment or local structure of the resonating C=O (34, 35). For example, a carbonyl in the M2 channel might be facing the lipid bilayer, protein–protein interface, or the aqueous lumen. In contrast, Val27 exhibited a significantly shifted peak to lower energies, at 1,581 cm⁻¹. It is noteworthy that only Val27 has a hydroxyl residue, four amino acids to its C-terminal direction, which allows for bifurcated H-bond formation. Hence, the potential to form a bifurcated

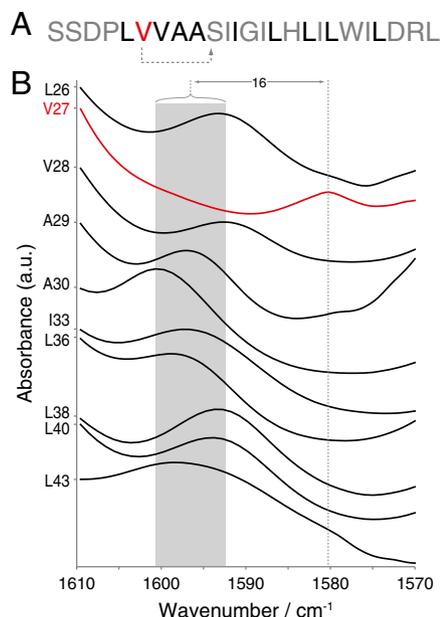


Fig. 1. (A) Sequence of the M2 peptide used in the analysis. Residues in black or red were isotopically labeled with ¹³C=¹⁸O. Val27 is in red because it is the only residue in the sequence that has a hydroxyl residue four residues in its C-terminal direction, as indicated by the dotted arrow. Note that numbering is according to the sequence of the full length protein. (B) FTIR spectra in the region of the isotope-edited amide I peak of 10 M2 transmembrane peptides in hydrated lipid bilayers (35). The different peptides are labeled with a ¹³C=¹⁸O at the position indicated on the ordinate. The shaded region represents the peak center range for all peptides except for the peptide labeled at Val27. The wavenumber shift (cm⁻¹) between the Val27 peak and the average of all of the other peptides is indicated.

H bond was consistent with a significant shift of the vibrational frequency of the C=O group.

To investigate the strength of bifurcated H bonds within a constant polar environment, we made use of the SARS coronavirus E protein. We synthesized four different peptides, each with a ¹³C=¹⁸O label at residue Phe26, but with different amino acids at location 30. This allowed the precise wavenumber of the amide I vibrational mode of Phe26 to be used as a reporter. Specifically, the peptides were: Thr30, Val30, Cys30, and Ser30. All four residues occur in natural variants of the SARS E protein and are therefore not expected to affect protein structure.

In Fig. 2B we present the FTIR spectra of the four different SARS E peptides in the region between 1,570 and 1,610 cm⁻¹. Each of the peptides exhibits a well-pronounced peak, which corresponds to the isotope-edited, ¹³C=¹⁸O Phe26 amide I mode (33). However, the precise location of the isotope-edited peak differs between the different peptides. In other words, the wavenumber of the amide I mode of Phe26 depends on which residue is located four amino acids away in its C-terminal direction (residue 30 in the sequence). The peptide with valine at position 30 had a Phe26 amide I peak at 1,598 cm⁻¹, and a cysteine at residue 30 resulted in a Phe26 peak at 1,594 cm⁻¹. The hydroxyl residues at position 30 produced an even larger shift, whereby threonine and serine resulted in Phe26 amide I peaks at 1,584 and 1,583 cm⁻¹, respectively. Hence, if the aliphatic residue is used as a baseline, a thiol at position *i* + 4 produced a shift of 5 cm⁻¹ and a hydroxyl yielded a shift of 14–15 cm⁻¹. The hydroxyl shift in SARS E of 14–15 cm⁻¹ is very similar to the 16 cm⁻¹ shift obtained in the entirely different M2 protein and falls within the ± 4 location-dependent variation range.

Based on the above two systems, we can surmise that H-bond bifurcation to a hydroxyl group results in amide I shifts

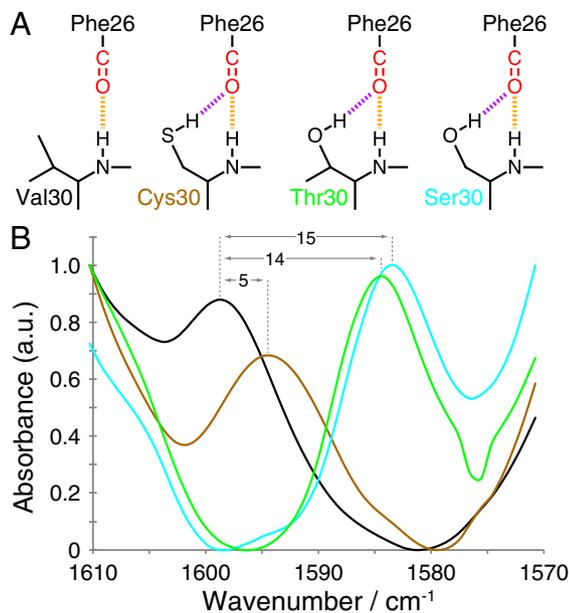


Fig. 2. (A) Schematic structures of the different H-bonding configurations in the four different peptides in the region of residue number Phe26 and 30 that were used for FTIR experimental measurements. The isotopically (¹³C and ¹⁸O) labeled carbonyl group of Phe26 is depicted in red. The main-chain (canonical) and bifurcated H bonds are shown in orange and purple, respectively. (B) FTIR spectra in the region of the isotope-edited amide I peak of four SARS E peptides in hydrated lipid bilayers. The different peptides are: Val30 (black), Cys30 (brown), Thr30 (green), and Ser30 (cyan). The wavenumber shifts relative to the peptide with a valine at position 30 are indicated.

of $14\text{--}16 \pm 4 \text{ cm}^{-1}$. H bonding the less-polar thiol group results in a $5 \pm 4 \text{ cm}^{-1}$ shift.

Computational Analysis. To obtain a quantitative measure of the strength of bifurcated H bonds we turned to DFT calculations. We constructed a system resembling a small portion of an α -helical turn with canonical and bifurcated H bonds.

The simulations consisted of *N*-Methylacetamide, paired with four different acetamide derivatives. The amide carbonyl of *N*-Methylacetamide served as the H-bond acceptor, and the amide H⁺ of the acetamide derivatives was the H-bond donor. Together, the two molecules represented a minimal canonical α -helix H-bonding configuration. Derivatives of acetamide mimicked amino acid side chains that were used to quantify the bifurcated H-bonding system. Valine was mimicked by *N*-Isobutylacetamide, which lacks the potential of forming a bifurcated H bond (Fig. 3A). Serine was mimicked by *N*-(2-Hydroxyethyl)acetamide that

can donate a hydroxyl bifurcating H bond depending on the dihedral angle (Figs. 3B and D). Cysteine was mimicked by *N*-(2-Thioethyl)acetamide that can donate a thiolic bifurcating H bond depending on the dihedral angle (Figs. 3C and E).

We first sought to corroborate our FTIR results by theoretical DFT calculations of the shift in vibrational frequency upon bifurcated H-bond formation. Next, we validate the accuracy of our DFT-determined bifurcated H-bond energetic results by calculating the strength of the canonical H bond, which is known from previous studies (9).

Vibrational frequency analysis. Calculation of the anharmonic frequencies of the different systems shown in Fig. 3 enabled us to derive the shifts due to bifurcated H-bond formation. In particular, we focused on the isotopically labeled ¹³C=¹⁸O of *N*-Methylacetamide, which is equivalent to the protein's amide I mode (red group in Fig. 3). The values listed in Table 1 show that the additional bifurcated H bond causes a shift to lower frequencies. Specifically, a shift of -8 cm^{-1} was obtained when *N*-(2-Thioethyl)acetamide (cysteine mimic) was used as the H-bond donor relative to the aliphatic *N*-Isobutylacetamide (valine mimic). The hydroxyl group in *N*-(2-Hydroxyethyl)acetamide (serine mimic) resulted in an even larger shift of -15 cm^{-1} .

The above values are in remarkable agreement to those obtained experimentally (Figs. 1 and 2). The experimental measurement of the shift resulting from a hydroxyl bifurcating H bond was $14\text{--}16 \text{ cm}^{-1}$, which is identical to the DFT-calculated results using *N*-(2-Hydroxyethyl)acetamide. The experimentally measured shift due to the thiolic bifurcating H bond was 5 cm^{-1} compared with 8 cm^{-1} obtained from the DFT frequency calculations. The minor difference between the experimental and calculated vibrational frequency shifts may be attributed to local environmental variations that cause peak shifts of up to 4 cm^{-1} as shown in Fig. 1.

Canonical H-Bond energy. The strength of the canonical H bond could be estimated by examining a system of *N*-Methylacetamide and *N*-Isobutylacetamide as a function of distance between the two molecules (Fig. 3A). In such a system, the energetic difference due to separation should be directly related to the energy of the single, canonical H bond. As listed in Table 1, separating the two molecules and thereby breaking the single H bond resulted in increasing the energy of the entire system by 5.7 kcal/mol. This value is in very good agreement with previous calculations from Honig and coworkers (9). In their study, the authors measured the dimerization energy of *N*-Methylacetamide in liquid alkane and obtained a value for the amide H bond of 5.3 kcal/mol.

Bifurcated H-bond energy. Above we have shown our computational approach to be accurate in estimating the energy of the canonical, amide H bond. We can now compute the energy of the bifurcated H bond using the same method. The energy obtained upon forming a bifurcated H-bonding system is mainly dependent on two components: (i) a potential change due to the particular dihedral angle of the donor amino acid side chain, and

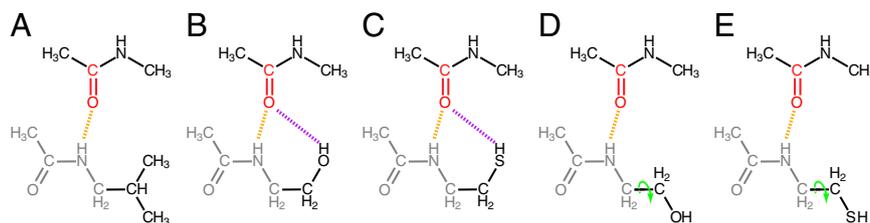


Fig. 3. Structures of the systems used in the DFT calculations. All calculations contained *N*-Methylacetamide as the H-bond acceptor (top molecule) with different potential H-bond donors as the bottom molecule: (A) *N*-Isobutylacetamide (valine mimic), (B) *N*-(2-Hydroxyethyl)acetamide (serine mimic), and (C) *N*-(2-Thioethyl)acetamide (cysteine mimic). Systems D and E are equivalent to B and C, respectively, but in rotameric configuration (+gauche) that does not facilitate bifurcated H bonding (angle rotation in green). The *N*-Methylacetamide backbone in the donors is shown in gray. The main-chain (canonical) and bifurcated H bonds are shown in orange and purple, respectively. The isotopically (¹³C and ¹⁸O) labeled carbonyl group of *N*-Methylacetamide is depicted in red.

Table 1. Summary of the DFT calculations shown in Fig. 3

System	H-bond acceptor	H-bond donor	Rotamer	$\Delta G_{\text{proximity}}$ (kcal/mol)	$\Delta \bar{\nu}$ (cm ⁻¹)
A	<i>N</i> -Methylacetamide	<i>N</i> -Isobutylacetamide	–gauche	–5.7	0
B	<i>N</i> -Methylacetamide	<i>N</i> -(2-Hydroxyethyl)acetamide	–gauche	–10.0	–15
C	<i>N</i> -Methylacetamide	<i>N</i> -(2-Thioethyl)acetamide	–gauche	–8.2	–8
D	<i>N</i> -Methylacetamide	<i>N</i> -(2-Hydroxyethyl)acetamide	+gauche	–6.6	ND
E	<i>N</i> -Methylacetamide	<i>N</i> -(2-Thioethyl)acetamide	+gauche	–5.6	ND

The different systems and their potential H bonds are depicted in Fig. 3; $\Delta G_{\text{proximity}}$ is the difference between the energies obtained from the calculations when the two molecules were in close proximity or far apart, reflecting the affect of proximity upon the energy; and $\Delta \bar{\nu}$ is the calculated wavenumber of the isotope edited C=O group of *N*-Methylacetamide relative to system in which the H-bond donor is *N*-Isobutylacetamide. Wavenumber shifts were not determined (ND) for systems D and E.

(ii) the contribution of the actual H bond. Therefore, we first determined the energetic difference in the H-bond donor when shifting between the dihedral angle that enables bifurcation (–gauche) and one that does not enable it (+gauche). This could be achieved by comparing systems in Fig. 3 *B* and *D* when the molecules were far apart from one another (i.e., no H bonding was present). Results show that the energetic difference between the –gauche and +gauche rotamers in *N*-(2-Hydroxyethyl)acetamide and *N*-(2-Thioethyl)acetamide are 2.2 and 0.47 kcal/mol, respectively.

With the knowledge of the energetics of the dihedral angle component at hand, we could now proceed to estimate the energy of the hydroxylic bifurcated H bond as follows. This was done by comparing the energies of systems in Fig. 3 *B* and *D*. In other words, these two identical systems differ only in the dihedral angle of *N*-(2-Hydroxyethyl)acetamide (the H-bond donor): –gauche in the case of a bifurcated H-bonding configuration (Fig. 3*B*) and +gauche in the case of only a single, canonical H bond (Fig. 3*D*). As shown in Table 1 the energy difference between the two systems is –1.2 kcal/mol. When we subtract from this value the energy that accounts for the dihedral angle component (2.2 kcal/mol), we obtain that the strength of the hydroxylic bifurcated H bond is –3.4 kcal/mol. A similar analysis comparing systems in Fig. 3 *C* and *E*, which differ in the dihedral angle of *N*-(2-Thioethyl)acetamide as the H-bond donor yielded an energy for the thiolic bifurcated H bond of –2.6 kcal/mol.

Comparison between bifurcated and canonical H bonds. The above results show that the energy of the hydroxylic bifurcated H bond is somewhat smaller relative to the canonical H bond found in α -helices: 3.4 versus 5.7 kcal/mol, respectively. The thiolic H⁺, being less polar, results in an even weaker bifurcated H bond of 2.6 kcal/mol. We note that the bifurcated H bond does not necessarily weaken the canonical H bond due the fact that it does not compete with it. However, the energy of all H bonds depends greatly on their surroundings, stronger in apolar conditions and appreciably weaker in polar environments, such as water. Therefore, we expect that the formation of bifurcated H bonds would be favored in hydrophobic conditions, such as solvent inaccessible areas in water soluble proteins, or transmembrane proteins, as previously suggested (20). We proceed to show this effect in a statistical analysis of proteins by taking into account water accessibility.

Impact of H-Bond Bifurcation on Structure. A manifestation of the importance of the bifurcated hydroxylic H bond may be found by close examination of protein structure. Toward this end, we analyzed the side-chain rotamer preference of serine. Threonine was not examined due to its biased side-chain preference, resulting from side-chain steric hindrance (19, 36). We focused on the behavior of the χ_1 dihedral as a function of secondary structure and exposure in a nonredundant database of water-soluble proteins (37).

The preferred χ_1 rotamer of serine in nonhelical structures is +gauche (19, 36). Similarly, in our calculations the difference

between the +gauche and –gauche rotamers of *N*-(2-Hydroxyethyl)acetamide was 2.2 kcal/mol (Table 1). Moreover, this statistical preference does not depend on the solvent accessibility of the residue (Fig. 4, *Lower*). In stark contrast, in helices, the preferred χ_1 is shown to be dependent on solvent accessibility (Fig. 4, *Upper*): When the residue is exposed to the aqueous environment, its preferred χ_1 rotamer is +gauche. However, when the residue is not exposed to the aqueous environment, it prefers a χ_1 rotamer of –gauche.

The above phenomenon may be explained by realizing that the –gauche χ_1 rotamer facilitates bifurcated H bonding between the hydroxylic side chain and the carbonyl of the *i*–4 residue. Therefore, when the residue is shielded from the solvent, a rotamer is preferred that enables the bifurcated H bonding of the polar side chain. However, when the residue faces an aqueous environment, water may satisfy the H-bonding potential of both the hydroxylic side-chain and the main-chain carbonyl group, regardless of the side-chain rotamer.

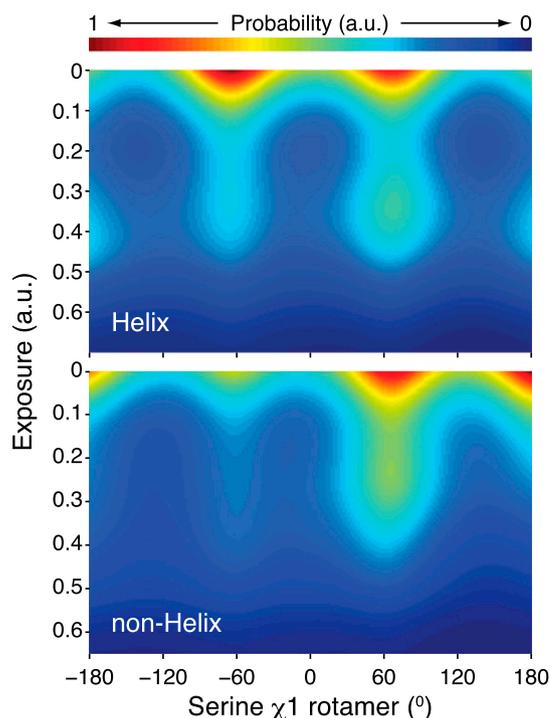


Fig. 4. Distribution of serine χ_1 rotamers as a function of side-chain exposure in (*Upper*) helices and (*Lower*) nonhelical elements. The data set was a nonhomologous representation of all solved water-soluble proteins (37). Exposure was calculated as a ratio between the exposure of the specific serine and the maximum exposure.

Chamberlain and Bowie have also pointed to the same phenomenon when examining membrane proteins (38). In their analysis, the authors found that the $-gauche \chi_1$ rotamer is more prevalent in residues that face the lipid bilayer in comparison with those in the protein core. Hence, the strength of the bifurcated H bond is demonstrated by its ability to influence the side-chain rotamer. Namely, a side-chain rotamer is selected despite steric preferences when a bifurcated H bond is energetically favorable.

Materials and Methods

Experimental Procedures. Peptide synthesis. Two different peptide systems were examined in the current study: influenza A M2 H⁺ channel and the SARS E protein. The procedures for sample preparation and FTIR data collection for the M2 peptides were described in detail previously (35).

Four different SARS peptides were used in the current study, each encompassing the hydrophobic domain of the SARS-coronavirus E protein and corresponding to residues Glu7 to Arg38 (29). The only distinguishing feature between the different peptides was the residue at position 30 of the original sequence: Thr30, Val30, Cys30, or Ser30. Note that all of the four residues are found in natural variants of the protein at this position. The synthesis of the peptides made use of solid-phase *N*-(9-fluorenyl methoxycarbonyl) chemistry. The $1-^{13}\text{C}=^{18}\text{O}$ labels were introduced as labeled amino acid precursors during the synthesis. The procedure of the synthesis of the $1-^{13}\text{C}=^{18}\text{O}$ labels is described elsewhere in detail (32, 33).

Peptide purification. The purification of the peptides was achieved using a 20-mL Jupiter 5 C4-300 Å high-performance liquid chromatography column (Phenomex); 2 mg of crude synthesis were dissolved in 2 mL of trifluoroacetic acid and injected into the column, which was pre-equilibrated with 80% H₂O, 8% (vol/vol) acetonitrile and 12% (vol/vol) isopropanol. Finally, elution was attained using a linear gradient to a final solvent composition of 60% isopropanol and 40% acetonitrile. Trifluoroacetic acid was present in all solvents (0.1%).

Membrane reconstitution. Membrane reconstitution was achieved using organic solvent cosolubilization. Peptide (ca. 1 mg) and 10 mg of lipid (1,2-dimyristoyl-sn-glycero-3-phosphocholine, Avanti polar lipids, Alabaster Al) were dissolved in 2 mL of 1,1,1,3,3,3-hexafluoro-2-propanol (Merck). The solution was then warmed to 37° for 30 min. Rotoevaporation, followed by desiccation, was used to remove traces of the organic solvent. Finally, liposomes were prepared by dissolving the sample in 2 mL of water.

FTIR spectra collection. FTIR spectra were collected in attenuated total internal reflection geometry, using a 25-reflection accessory from Graseby Specac. The spectrometer used was a Nicolet Magna 560 spectrometer, which was equipped with a high-sensitivity liquid nitrogen-cooled mercury cadmium telluride detector.

Also, 400 μL of sample, which contained ca. 0.5 mg/mL protein and 5 mg/mL lipid, were deposited onto a trapezoidal Ge internal reflection element ($50 \times 2 \times 20$ mm). Bulk solvent was removed under a stream of CO₂ and water-depleted air. The spectrometer was purged with CO₂ and water-depleted air. Then 1,000 interferograms were collected for each sample and averaged. Spectra were processed with 1-point zero filling and Happ-Genzel apodization.

DFT Calculations. All calculations were carried out with the Q-chem software package to determine energy values for the bifurcated H bond (39). Geometric optimizations, along with the self-consistent field (SCF) electronic energy and the vibrational frequency calculations were determined for all four systems indicated in Fig. 3 using aug-cc-pVDZ basis functions with

a B3LYP exchange-correlation functional. The aug-cc-pVDZ basis set is an augmented double-zeta correlation-consistent basis set, which includes diffuse *s* and *p* functions on H atoms and diffuse *s*, *p*, and *d* functions on C, N, and O atoms (40, 41). The B3LYP exchange correlation functional includes the correlation functional derived by Lee et al. (42) with Becke's exchange correction (43). The dielectric constant was set at 4 to mimic a membrane environment.

Geometric models. All molecules shown in Fig. 3 were constructed using the molecule building subroutine in vmd: molefactory (44). Initially, the structures of the molecules were superimposed onto a canonical helix: *N*-Methylacetamide, the H-bond acceptor, was superposed on the backbone of residue *i*; and the acetamide derivatives, as H-bond donors, were superposed on the backbone of residue *i* + 4. The dihedral angle of the acetamide derivatives was set according to statistical preferences: +*gauche* or $-gauche$.

SCF energy calculations. For energetic calculations of both canonical and side-chain H bonds within the bifurcated H-bond systems the hydroxyl or thiol group of *N*-(2-Hydroxyethyl)acetamide or *N*-(2-Thioethyl)acetamide, respectively, were initially allowed freedom of movement to undergo geometric optimization to determine the bonded structures (Fig. 3 *B* and *C*). The non-bonded structures were created by rotating the dihedral angle of the bonded structures from $-gauche$ to +*gauche* (Fig. 3 *D* and *E*). Finally, simulations were also conducted when the H-bond donor and acceptor were separated by 1,000 Å to discount any effect of H bonding between them.

Vibrational frequency calculations. Infrared vibration spectra information was determined from the Fourier transform of the bond dipole moment autocorrelation function. The dipole moment of the isotope edited carbonyl bond is $\vec{\mu}(t)$. The autocorrelation function for this bond after perturbation by an external electric field is $\langle \vec{\mu}(0) \cdot \vec{\mu}(t) \rangle$. Finally, the vibrational spectrum, $I(\omega)$ is proportional to the Fourier transform of the autocorrelation function of the bond dipole (see ref. 45 and examples in refs. 46, 47):

$$I(\omega) \propto \frac{\omega}{2\pi} \int_{-\infty}^{\infty} dt e^{-i\omega t} \langle \vec{\mu}(0) \cdot \vec{\mu}(t) \rangle.$$

The amide carbonyl acting as the H-bond acceptor was labeled with $^{13}\text{C}=^{18}\text{O}$ to mimic the experimental frequency system (see red group in Fig. 3). Geometrical optimization of the $^{13}\text{C}=^{18}\text{O}$ along with *N*-(2-Hydroxyethyl)acetamide side-chain atoms was allowed; however, all other atoms were constrained to the known transmembrane α -helical structure.

Bioinformatic Analyses. Database creation. A database of nonhomologous protein structures was retrieved from PDBselect (37), using the following parameters: 25%, Nov 2012, and $n\text{sigma} = 3.5$. The database was then purged from any membrane protein using a list of all solved membrane protein structures from Protein Data Bank of Transmembrane Proteins (48).

Rotamer analysis. Side-chain rotamer analysis for every one of the water-soluble proteins in the database was undertaken using a Tcl script in vmd (44). The solvent accessible surface area was calculated using vmd, whereby only the side-chain atoms of the residue were used in the computation. The maximal accessibility ratio was calculated by dividing the solvent accessible surface area by the area of the residue detached from the protein.

ACKNOWLEDGMENTS. The authors wish to thank Prof. Roi Baer and Dr. Ester Livshits from the Institute of Chemistry at the Hebrew University of Jerusalem for their advice and help in the DFT calculations. This work was supported in part by grants from the Binational Science Foundation (2008035) and the Israeli Science Foundation (1581/08). I.T.A. is the Arthur Lejwa Professor of Structural Biochemistry at the Hebrew University of Jerusalem.

- Watson JD, Crick FH (1953) Molecular structure of nucleic acids; A structure for deoxyribose nucleic acid. *Nature* 171(4356):737–738.
- Pauling L, Corey RB (1951) The pleated sheet, a new layer configuration of polypeptide chains. *Proc Natl Acad Sci USA* 37(5):251–256.
- Pauling L, Corey RB, Branson HR (1951) The structure of proteins; Two hydrogen-bonded helical configurations of the polypeptide chain. *Proc Natl Acad Sci USA* 37(4):205–211.
- Eisenberg D (2003) The discovery of the alpha-helix and beta-sheet, the principal structural features of proteins. *Proc Natl Acad Sci USA* 100(20):11207–11210.
- Fersht AR, et al. (1985) Hydrogen bonding and biological specificity analyzed by protein engineering. *Nature* 314(6008):235–238.
- Williams DH, Searle MS, Mackay JP, Gerhard U, Maplestone RA (1993) Toward an estimation of binding constants in aqueous solution: Studies of associations of vancomycin group antibiotics. *Proc Natl Acad Sci USA* 90(4):1172–1178.
- Klotz IM (1993) Solvent water and protein behavior: View through a retro-scope. *Protein Sci* 2(11):1992–1999.
- Mirsky AE, Pauling L (1936) On the structure of native, denatured, and coagulated proteins. *Proc Natl Acad Sci USA* 22(7):439–447.
- Ben-Tal N, et al. (1997) Free energy of amide hydrogen bond formation in vacuum, in water, and in liquid alkane solution. *J Phys Chem B* 101(3):450–457.
- Avbelj F, Luo P, Baldwin RL (2000) Energetics of the interaction between water and the helical peptide group and its role in determining helix propensities. *Proc Natl Acad Sci USA* 97(20):10786–10791.
- Mitchell JBO, Price SL (1990) The nature of the n-ho=c hydrogen bond: An intermolecular perturbation theory study of the formamide/formaldehyde complex. *J Comput Chem* 11(10):1217–1233.
- Sheu SY, Yang DY, Selzle HL, Schlag EW (2003) Energetics of hydrogen bonds in peptides. *Proc Natl Acad Sci USA* 100(22):12683–12687.
- Ben-Tal N, Ben-Shaul A, Nicholls A, Honig B (1996) Free-energy determinants of alpha-helix insertion into lipid bilayers. *Biophys J* 70(4):1803–1812.
- Sneddon SF, Tobias DJ, Brooks CL, 3rd (1989) Thermodynamics of amide hydrogen bond formation in polar and apolar solvents. *J Mol Biol* 209(4):817–820.

