
Motohiro Nishio

The CHPI Institute, 794-7-910 Izusan, Atami, Shizuoka, Japan
*E-mail: dionisio@tim.hi-ho.ne.jp

(Received December 25, 2016; accepted April 28, 2017; published online August 24, 2017)

Abstract

Widespread arguments of the hydrogen bond are criticized with respect to theories of protein folding and interactions of proteins with their specific ligands. Contrary to popular belief, by no means does the hydrogen bond play an important role in determining the conformation of proteins and the interactions of proteins with specific substrates. Stereotypical thinking on the hydrogen bond constitutes a myth in modern science and seriously restricts the success of structure-based drug design.

Key Words: conformation of proteins, folding, interaction of proteins with substrates, structure-based drug design

Area of Interest: In silico drug discovery
1. Introduction

There are many myths in the world: Greek and Roman myths, Northern European myths, Polynesian myths, Japanese myths, and so on. Here, the author defines the myth as a stereotypic thought restricting our rational ideas on present-day science. The hydrogen bond, in itself, is not a myth. However, arguments concerning the hydrogen bond (H-bond) constitute a myth, undoubtedly.

2. The hydrogen bond

The H-bond, along with the van der Waals and electrostatic forces [1, 2] is one of the most general interactions taking place between molecules.

Linus Pauling, in his book “The Nature of the Chemical Bond,” defined the H-bond as an attractive force arising from the Coulombic interaction between positively charged hydrogen (O-H or N-H) and a lone pair of electrons of O or N [3]. This definition, however, can be more comprehensively understood by perusing recent reviews [4]. Pauling’s definition has gradually become inadequate, over the years, by the continual progress made in chemistry, and it has been replaced by Pimentel and McClellan’s modern and more general expression [5], which involves weaker H-bonds (OH/π, CH/O, CH/π, etc.). In 1989, Hirota and Nishio classified the H-bond into four categories [6], namely 1) conventional H-bond, hard acid (HA)/hard base (HB) combinations, 2) XH/π (X: O, N: HA/soft base (SB) H-bond, 3) CH/Y (Y: O, N: SA/HB) H-bond [7], and 4) CH/π (SA/SB) H-bond [8]. In 2009, we published a review with a more comprehensive explanation [9]. From 2005 to 2009, there have been repeated discussions by many crystallographers, spectroscopists, and theoretical chemists on the H-bond and, finally, in 2011 an IUPAC Report on the subject was published [10]. A short article by G. Desiraju is easier to read and I recommend it [11].

Before discussing the hydrogen-bond-myth, it seems relevant to call attention to the characteristics of the H-bond. Compared to the covalent bond, the energy of the H-bond is much weaker. An important and general characteristic is its directionality. Another characteristic of the H-bond is that many interactions work together simultaneously.

In polar or protic solvents, the energy of Coulombic interactions becomes much smaller, as compared to the case in a nonpolar environment. In thinking about the function of the H-bond, one should not overlook the fact that the H-bond does not work efficiently in polar media.

The most important characteristic of the H-bond is that solvent water interferes in interactions between molecules. Arguments regarding the 3D-structure formation (folding) of proteins and interactions of proteins with their ligands forget this vital point [12]. In the process of protein folding, for instance, the H-bond never works. It is crucial to consider nonpolar forces, such as van der Waals, or the CH/π H-bond, functioning between side chains of nonpolar amino-acid residues.

3. The driving force of the 3D structure-formation of a protein

The α-Helix [13] and β-sheets [14] are important components of protein secondary structures. Pauling has shown that the α-helix and β-sheets are sustained by H-bonds between C=O and NH in
the main-chain of proteins. Remember, however, that the sequence of the main-chain bears no information about what residue chooses its partner within the same chain [15]. In other words, a residue does not know its partner, at all. Such necessary information is found in the environment around the residue. The conformation of a protein molecule changes, little by little, and finally converges into a structure having minimum energy. In other words, only side-chain groups of amino-acid residues of the main-chain bear necessary information. Thus, an N-H hydrogen seeks another C=O nearby to it, and if a new NHCO/CONH combination is energetically more feasible, swapping occurs. Thus, the minimum energy conformation is determined by the total structure of the protein. Pairs of NHCO/CONH found in an existing α-helical structure are the results of energetic swapping, and are by no means the cause\(^1\). The same argument is valid for the H-bond formation of β-sheets. To illustrate how this occurs, here I present Anfinsen’s work [16-18] regarding the folding of ribonuclease (RNase), although the 3D structure of this protein is not maintained by NHCO/CONH H-bonds, but rather by –S–S– covalent bonds.

In the 1950’s, Anfinsen developed what he called the "thermodynamic hypothesis" of protein folding to explain the native 3D structure of ribonuclease (RNase). This is a small protein of 124 amino-acid residues with eight cysteines and four -S–S– bonds. Anfinsen unfolded RNase by reductive cleavage of the 4 disulfide bonds (4 S–S→8 SH) and observed that the protein’s 3D structure was destroyed and its enzymatic activity was completely lost. This result occurred because a protein’s enzymatic activity is closely associated with its 3D structure. When the reduced protein with its eight SH groups is carefully oxidized under mild conditions, refolding occurs to give back the original conformation of RNase, and the enzymatic activity is recovered. But how has this occurred? Anfinsen hypothesized that the natural conformation returned because this particular shape is the most stable thermodynamically and is maintained in the intracellular environment. In other words, the native conformation is determined by the totality of its interatomic interactions and, hence, by the amino-acid sequence in a given environment. An SH group does not know which SH is the most suitable among 105 (7 x 5 x 3) possibilities. Here, the strategy is always the same. The minimum energy conformation is determined by the total structure of the protein. Under Anfinsen’s experimental conditions, the 3D structure remained similar to the original one. An SH group will bind another SH, located close to it.

The Anfinsen conclusion, however, is limited to small proteins such as RNase, and the experimental conditions are rather strict [19]. In view of these restrictions, modifications of his theory have been presented [20-22]. For peptides or small proteins of about 8 to 80 residues, MD simulation by the use of high-speed supercomputers has given results that are acceptable, at most [23]\(^2\). I do not think, however, that the approaches presently being used would give satisfactory results. All necessary information for the 3D structure-formation of a protein is included in its sequence. In order to reproduce the original conformation, therefore, one must correctly estimate all the interactions occurring between the side-chain groups and between each side-chain and the other parts of the protein. To better reproduce the conformation, however, there is another (statistical) approach [24]. It is reported that crowded and aromatic side-chains are abundantly present in the starting position of the folding [25]\(^3\).

\(^1\) Es gibt keinen gefährlicheren Irrthum als die Folge mit der Ursache zu verwechseln (Friedrich Nietzsche, Götzen-Dämmerung). Translated into Japanese by T. Hara, Chikuma Gakugei Bunko, The complete works of F. Nietzsche, Vol. 14, p. 56.

\(^2\) 0.6–65 μs, RMSD 0.5–49 Å

\(^3\) This may suggest the importance of the CH/π H-bond.
4. To what extent is water necessary for the 3D structure-formation?

Next, we consider how water participates in the 3D structure-formation. In crystal structures deposited in the Protein Data Bank, we often see many water molecules within proteins. We have to ask, however, is there any evidence that the water molecules were present in the physiological environment? I think the answer is “NO.” These water molecules must have been involved in the final stage of the crystal formation. In other words, the presence of water molecules in the protein crystal structure shows the results and does not mean that they were present at these positions during the crystallization process.

As a thought experiment, let us imagine the binding process of a receptor with its specific ligand. The binding occurs under physiological conditions (water). Do you think that the H-bond does work? Of course it does not. Why? Because the interaction of the receptor with its ligand is occurring in water. Nonpolar forces such as van der Waals, or CH/π interactions are involved, instead.

I imagine, rather, that a ligand, located remotely from the receptor moves very violently. By chance, then, both participants would come together. Some interactions occur between parts of the ligand and the protein, and loosely bonded units are formed by nonpolar forces. And, then, what will occur next? The parts will move away and separate, or remain close together for a moment (complex formation), and then repeat the rearrangement to seek another complex having smaller free energy. Finally, a ligand/protein complex having the smallest free energy might be obtained. Thus, a specific ligand is tightly bound within the pocket of the receptor. I believe that, except for at the last instant, there is no room here for hydrogen bonds to play a role.

5. Rational drug design

Presently, the random screening method does not play a major role in the development of drugs. The QSAR (quantitative structure-activity relationship) or Hansch-Fujita method is more useful. The SBDD (structure-based drug design) method is the most commonly used. In SBDD, we use computer programs, purchased from software-houses. An all-around useful program does not exist. Drug designers, therefore, must use all possible techniques, by trial and error, in order to reach the final result. Two successful outcomes are cited in the references [26, 27]. The next part of this review presents contributions by Ozawa, Ema, Watanabe, and their coworkers, to show how intermolecular interactions, including weak H-bonds, are playing effective roles in the binding of specific substrates with proteins.

In 2008, Ozawa et al. reported that the CH/π H-bond was found in complexes of the β2 adrenergic receptor and suggested that CH/π hydrogen bonds play key roles in ligand recognition and conversion between the active and inactive states [28]. The abstract (edited) of this paper is given below.

The interaction energy was calculated, by the ab initio FMO method, for complexes between LCK protein and four inhibitors (staurosporine, BMS compound 2, and our compounds 3 and 4). In every case a number of CH/π hydrogen bonds have been disclosed in the so-called adenine pocket. In complexes of 2, 3, and 4, CH/π and NH/π hydrogen bonds have been observed in another pocket.

4 http://www.rcsb.
5 The part of quotation has been permitted by the publisher.
In view of the above results, the aniline ring of 3 was replaced by 2, 6-dimethyl aniline to increase the potency for LCK kinase. A 10-fold increase in the potency has been achieved for 4 over 3. We suggest that the concept of weak hydrogen bonds is useful in the rational design of drugs.

Ozawa, Okazaki, and Kitaura reported the importance of the CH/π H-Bond in recognition of the proline-recognition domains by an ab initio fragment molecular orbital (FMO) study [29]. The abstract of this paper is given below.

We examined CH/π hydrogen bonds in protein/ligand complexes involving at least one proline residue using the ab initio fragment molecular orbital (FMO) method and the program CHPI. FMO calculations were carried out at the Hartree–Fock (HF)/6-31G*, HF/6-31G**, second-order Möller–Plesset perturbation (MP2)/6-31G*, and MP2/6-31G** levels for three Src homology 3 (SH3) domains and five proline-recognition domains (PRDs) complexed with their corresponding ligand peptides. PRDs use a conserved set of aromatic residues to recognize proline-rich sequences of specific ligands. Many CH/π hydrogen bonds were identified in these complexes. CH/π hydrogen bonds occurred, in particular, in the central part of the proline-rich motifs. Our results suggest that CH/π hydrogen bonds are important in the recognition of SH3 and PRDs by their ligand peptides and play a vital role in the signal transduction system. Combined use of the FMO method and CHPI analysis is a valuable tool for the study of protein/protein and protein/ligand interactions and may be useful in rational drug design.

Ema and coworkers reported on the design of lipase for improving catalytic activity and selectivity toward poor substrates [30]. The abstract (edited) of this paper is given below.

Secondary alcohols having bulky substituents on both sides of the OH group are inherently poor substrates for most lipases. In view of this, we designed a lipase to create a variant with improved enzymatic characteristics. The I287F/I290A double mutant showed a high conversion and a high E value (>200) for a poor substrate for which the wild-type enzyme showed a low conversion and a low E value. This enhancement of catalytic activity and enantioselectivity of the variant resulted from the cooperative action of two mutations: Phe287 contributed to both enhancement of the (R)-enantiomer reactivity and suppression of the (S)-enantiomer reactivity, while Ala290 created a space to facilitate the acylation of the (R)-enantiomer. The kinetic constants indicated that the mutations effectively altered the transition state. Substrate mapping analysis strongly suggested that the CH/π interaction partly enhanced the (R)-enantiomer reactivity, the estimated energy of the CH/π interaction being −0.4 kcal mol⁻¹.

Watanabe et al. reported the fragment molecular orbital (FMO) calculations of a complex of Oseltamivir or Tamiflu (an anti-influenza drug: inhibitor of neuraminidase) [31]. From that they found an important role of the CH/π H-bond in stabilizing the drug-enzyme complex.

In 2014, Nishio and coworkers wrote a review providing evidence that the nonpolar CH/π H-bond plays an important role in biochemistry [32]. The abstract of this review is given below.

Evidence that weak hydrogen bonds play an enormously important role in chemistry and biochemistry has now accumulated to an extent that the rigid classical concept of hydrogen bonds formulated by Pauling needs to be seriously revised and extended. The concept of a more
generalized hydrogen bond definition is indispensable for understanding the folding mechanisms of proteins. The CH/π hydrogen bond, a weak molecular force occurring between a soft acid CH and a soft base π-electron system, among all is one of the most important and plays a functional role in defining the conformation and stability of 3D structures as well as in many molecular recognition events. This concept is also valuable in structure-based drug design efforts. Despite their frequent occurrence in organic molecules and bio-molecules, the importance of CH/π hydrogen bonds is still largely unknown to many chemists and biochemists. Here we present a review that deals with the evidence, nature, characteristics and consequences of the CH/π hydrogen bond in biological macromolecules (proteins, nucleic acids, lipids and polysaccharides). It is hoped that the present Perspective will show the importance of CH/π hydrogen bonds and stimulate interest in the interactions of biological macromolecules, one of the most fascinating fields in bioorganic chemistry. Implication of this concept is enormous and valuable in the scientific community.

In a book published by Wiley-VCH in 2001, Pharmacokinetic Optimization in Drug Design, we see a chapter “Hydrogen Bonding: The Last Mystery in Drug Design?” written by Kubinyi [33]. The conclusion of this chapter is reproduced below 9.

Significant progress has been achieved in the derivation of scoring functions for docking and de novo design of ligands. However, considering the examples presented here, it is not surprising that these affinity estimations still lack sufficient precision, especially with respect to the influence of hydrogen bonding on ligand affinities. Despite all attempts to arrive at a better understanding of the role of water and of hydrogen bonds in biological systems, we are far from a satisfactory situation.

For rational drug design, all the necessary tools are available: we can generate meaningful 3D-structures of ligands from scratch, we can convert them into multiple low-energy conformations, we can calculate their steric, electrostatic, and lipophilic properties, and we can dock them in a flexible manner into their binding sites. We can even construct ligands in a combinatorial manner, but in the very end we fail because we do not have a sufficient understanding of all the individual enthalpy and entropy terms that are involved in desolvation, hydrogen-bond formation, and hydrophobic interactions. The challenge for the current decade will be to further understand and explain the mystery of hydrogen bonding. However, it might take much longer than ten years until we will arrive at reliable scoring functions. Then, the way to tailor-made ligands seems to be straightforward.

The author of this chapter seems to think about the drug-design according to the flawed notion that an H-bond works in water. My impression is that by such a way of thinking, drug design will never give satisfactory results. This is not a Mystery, but a result of pursuing such an approach. In other words, this is a natural consequence of the hydrogen-bond-myth; the failure of drug design carried out with the hydrogen-bond-myth and the hydrophobic-force myth is apparent.

6. Conclusion

Contrary to widespread dogma, by no means does the H-bond (conventional) play a role in determining the conformation of proteins and the interactions of proteins with specific ligands. The

---

9 The part of quotation has been permitted by the publisher.
10 Such a force does not exist. See “Part 2. The Hydrophobic-Force-Myth”.

90
reason is simple and clear. The H-bond does not work in physiological environments (water), at all. Nonpolar weak interactions work, instead.

Stereotypical thinking on the H-bond is a myth in modern science and seriously restricts the progress of protein sciences, including structure-based drug design.

Note added in proof

Very recently, Ozawa and coworkers provided evidence that the classical H-bond does not work, almost at all, in the binding of streptavidin with biotin; there, the CH/π H-bond plays the most important role [34].

References

[14] Pauling, L. The Nature of the Chemical Bond, Cornell University Press: Ithaca, NY, 1939; Fig. 12-19 and Fig. 12-20.
The importance of CH/π interactions: manipulation of the transition state.


