MOLECULAR PRESENTATIONS

Cubic Hydration and the Spontaneous Formation of Proteins

J. C. Collins, PhD

Dedicated to the late Professors Carl Djerassi and William S. Johnson of Stanford University.

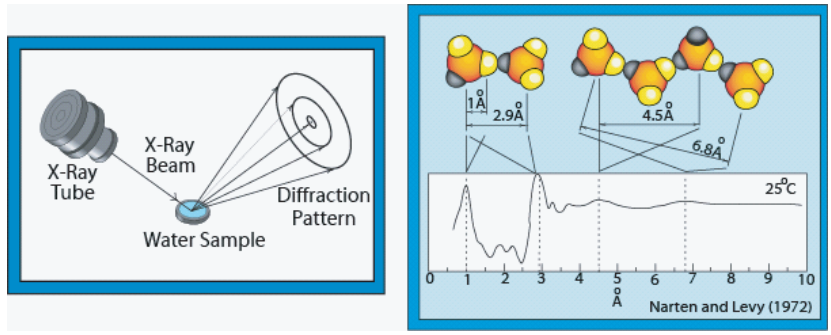
One of the most complex processes in the living cell is the spontaneous assembly of polypeptides into proteins. Water is believed to be involved, but, transitions are so rapid, that it has been difficult to define precisely how.

Recently, David Baker and Minkyung Baek reported the development of a computer program which stores structural information on thousands of proteins and uses it to derive the spatial structures for new proteins based solely on aminoacid sequences[1]. It was a momentous achievement, but did not address the water question.

Actually, it was in the 1950s when Christian Anfinsen performed the classical experiments which suggested that it is the sequence of aminoacids in polypeptides which define how they wrap spontantaneously into specific protein structures.[2] When urea was added to an aqueous solution of an enzymatic protein, it unwrapped to form the linear polypeptide, and then, with urea removed by dialysis, the polypeptide reassembled into the active enzyme. Based on the result, Christian and others concluded that urea must be binding to the protein to induce the unwrapping. However, subsequent studies demonstrated that urea does not bind to proteins - instead, it binds to water molecules[3] - it prevents freezing at 0°C and the formation of ice adjacent to lipid surfaces.

As a medicinal chemist in the 60s and 70s, I constructed permanent models of a number of neurotransmitters and hormones and found, to my surprise, that their dimensions corresponded to linear elements of hydration. When the study was extended to water-soluble enzymes, it appeared that their anhydrous cores corresponded to the spatial units in cubic ice. Since I was employed and the information was being used to develop drugs, it could not be published. However, when I retired in 1987 and the company was sold in 1988, I was able to publish *The Matrix of Life in 1991* with the proposal that it is transient linear elements of hydration in cubic ice conformations which guide polypeptides and biomolecules spontaneously into orderly assemblies.[4]

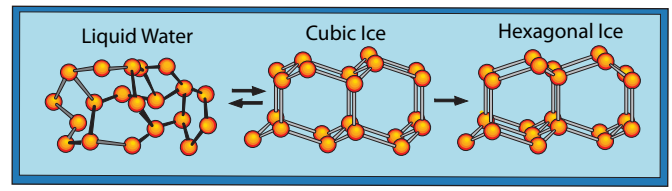
However, mine was not the first proposal of ice-like molecules in liquid water. In 1944, Erin Schrödinger, in his little book, *What is Life?*, proposed that there must be some sort of order in water to have moved natural molecules spontaneously from randomness to order to form the living cell.[5] In 1970, nuclear magnetic resonance studies revealed that water adjacent to proteins and nucleic acids exhibits the peaks of of ice as well as liquid water[6] and numerous subsequent studies revealed that water exhibits structuring properties around proteins [7] and on lipid surfaces.[8] In 1969, molecular orbital calculations suggested that the trimer, with three molecules bonded together in a linear element, might be the most stable structural-unit in liquid water[9] and, in 1984, that linear elements of five or six water molecules might form on lipid surfaces.[10] Experimentally, the first evidence of the trimer in liquid water came in 1972 when X-rays were deflected from the surface of liquid water.[11]

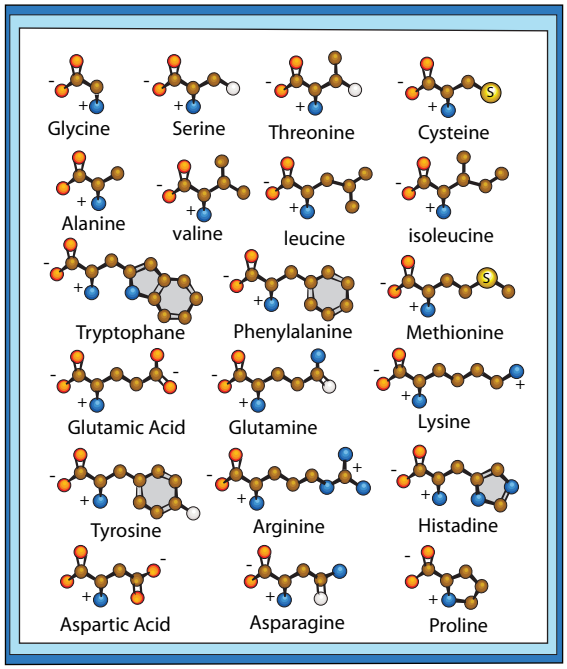


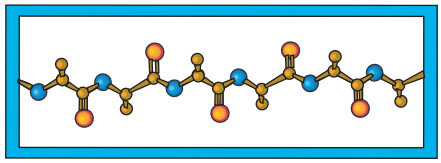
A major peak at 2.9 Angstroms in the X-ray pattern was identified as the mean distance between water molecules hydrogen-bonded together in the dynamic liquid state while two minor peaks at 4.5 and 6.8 Angstroms corresponded to the trimer and tetramer, bonded together at 2.76Å, the same as cubic ice.[12] Finally, in 2004, the trimer was identified in liquid water by bombardment with high-energy neutrons.[13]

However, it was not until 2010 that Dr. Sow-Hsin Chen, in a broad collaborative study, found that, as the native polypeptide of lysozyme enzyme wraps to form the functional protein, "water follows accurately all the protein behavior, detailing properly its structural and dynamical changes in this transition from native to denatured."[14] For the first time, high-resolution nuclear magnetic resonance had been used to follow the protons in surface water and provide experimental evidence that water initially covers all surfaces of polypeptides as they are released from ribosomes, but then, is spontaneously lost from lipid surfaces to form the anhydrous cores of proteins. A companion study reported that "tetrahedral order" is increased on "conserved helices and corresponding non-helical counter-parts" and that "longer helices exhibit significantly longer residence times."[15] Coupled with a 1970 thermodynamic study which indicated there is "an enthalpy-entropy compensation phenomenon which occurs in water as molecules approach the surface of proteins,"[16] and one in 2009 that water forms linear elements in cubic ice forms on the super-cooled surface of graphite,[17] it appeared that, indeed, water may form transient linear elements in cubic forms on non-hydrogen-bonding surfaces of polypeptides to provide direction, and then on leaving, provide the energy to form the orderly compact dehydrated cores of proteins.[8]

But, before we move further into protein assembly, we must understand more about water, aminoacids and polypeptides. Water molecules, with positive hydrogen atoms on two corners of an oxygen atom and two negatively-charged electron orbitals at the other two corners, form attachments at all kinds of angles and distances to form liquid water.[18] The bonds last about a million millionth of a second and continually form new attachments.[19] When cooled to 0°C, the molecules are too close together to form ice bonds and must be on a surface where atoms are in hexagonal positions, like those in ice, in order to freeze.[18] When ice-bonds form, the electron orbital of the oxygen atom in one water molecule overlaps the hydrogen atom in another water molecule to form a relatively ridged covalent-like linear bond.[20] As linear bonding proceeds, the ice produced is cubic,[12] but it rapidly isomerizes to the more common and stable hexagonal form with the hexagons in the planes over each other.[21]





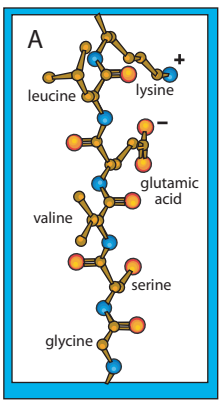


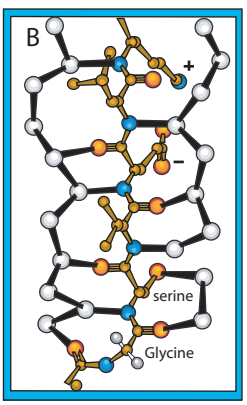
Turning now to polypeptides: they are formed by bonding aminoacids together to produce long chains with side-chains extending alternately above and below the plane of the chains. Each aminoacid, as shown on the left, has a different side-chain to regulate the liquid or ice-like bonding properties of water molecules above and below. the chain. Although figures are idealized, hopefully, they will provide a view of how the side-chains of the aminoacids interact with surface water to regulate assembly.

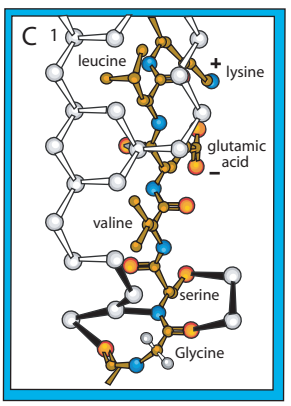
Glycine, with no side-chain to shield it, usually hydrogen-bonds directly with surface water, accepts units of energy from water and produces changes in direction of chains. Serine, with an oxygen on its alpha carbon, either fits the oxygen into the ice-like hexagonal patterning that is generated around the lipid regions of the chain, or forms a liquid-like bonds with surface water to disrupt hydration order and produce turns and wide hydrating loops to increase solubility.

On the other hand, the sulfur atom in cysteine, by not hydrogen-bonding readily with water, usually behaves like the aminoacids in Rows 2 and 3; all of which have side-chains which do not hydrogen-bond with adjacent water and reinforce cubic patterning. Aminoacids acids in Row 4 have extended side-chains, with hydrogens on alpha carbons which shield segments of the chain from binding with water but polar and ionic heads which either fit into ordering positions of water molecules, or disrupt order in adjacent hydration layers. Arginine has such a large, positively-charged end-group that it often, like the aminoacids in the bottom row, disrupts order ato produce changes in polypeptide direction. Proline is unique in that it does not permit normal polypeptide bonding with adjacent aminoacids - it always produces changes in chain direction.

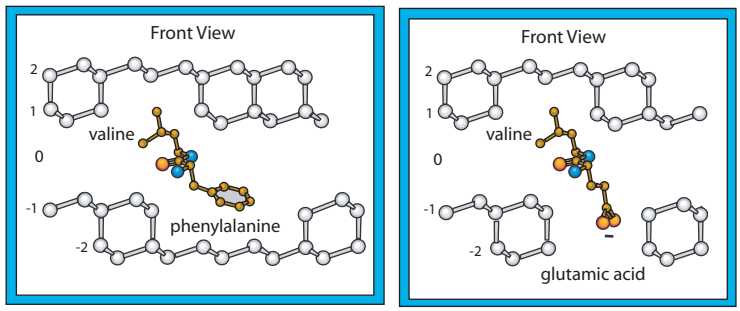
As a newly-formed polypeptide emerges from a ribosome, as in A, its oxygen and nitrogen atoms hydrogen-bond dynamically with water molecules in the plane of the chain, as in B. In C, linear elements of hydration (2.76 Angstroms between molecules) continually generate hexagonal patterning in layers above and below the ordering side-chains, while glycine and serine, as shown on the right above, disrupt hydration order by hydrogen-bonding directly with water above and below the chain.





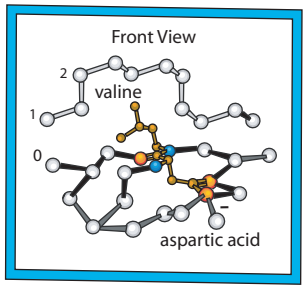


However, the drawing above is deceptive - at any instant, most water molecules will be in random positions - only occasionally, will order be in some units of water molecules, and propagate out to form linear elements in cubic probability positions in other locations. What is essential to realize is that bonding between water molecules is a thousands of times faster than the movements of molecules. Molecules must experience a flickering of hydration order around them - like blades of a fan which occupy space, but only part of the time.

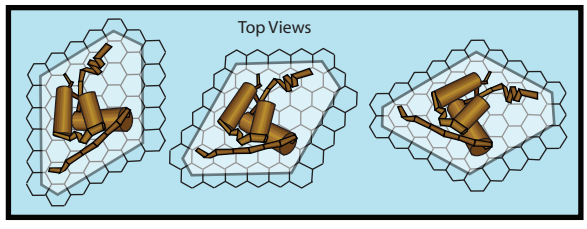


If we look at idealized front views of positions of ordered water molecules over time, we can see the order induced by lipid surfaces. Notice that the side-chains of valine and phenylalanine on the left, induce periodic ordering across four levels while the ionic end of the glutamic acid side-chain, in the figure on the right, supports limited order in level -1 but its ionic head, by hydrogen-bonding water molecules around it to delocalize its charge, disrupts local order in levels -1 and -2.

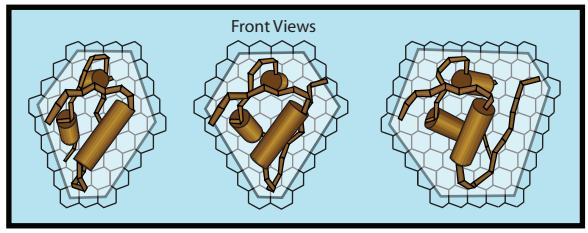
On the other hand, if lipid side-chains are on both sides of a segment of polypeptide, they have such a dominant tendancy to combine that, as the polypeptide enters an enlargement in the ribosomal tunnel, it forms a pre-coil "compact state" which then converts into the finished coils and protein when it contacts the water.[22] If peptides like glycine, serine or aspartic acid, are in every-other position on one side, linear elements of hydration form on the other side, and then on leaving, permit that side to combine with a lipid surface of the same quantized length. Of course, if small water-binding aminoacids are on both upper and lower surfaces, segments will twist and turn and form polar loops to increase mobility and solubility.

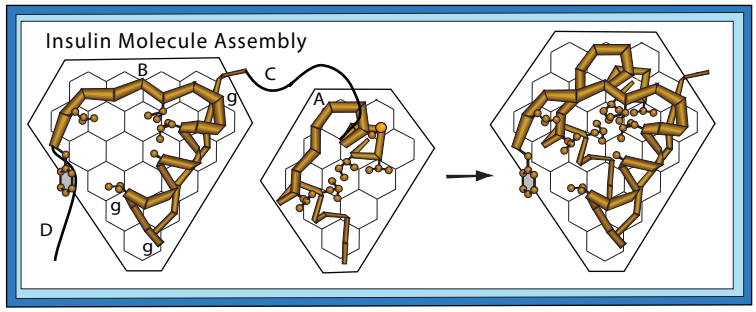


Although protein structures are presented in multiple orientations, hydration analyses must be performed on orientations which correspond most closely to those that reveal the role of linear elements of hydration in the assembly. For example in the analysis of insulin, the configuration reported by Hodgkin et al. in 1971,[22] was chosen with the xy plane of the insulin molecule positioned over the xy plane of the cubic lattice and the linear segments and major coils in line with linear elements in the cubic lattice.



Once that was done, the molecule and lattice could be viewed in multiple orientations perpendicular to the z axis. As shown on the right, a close spatial relationship is revealed in three cubic orientations.

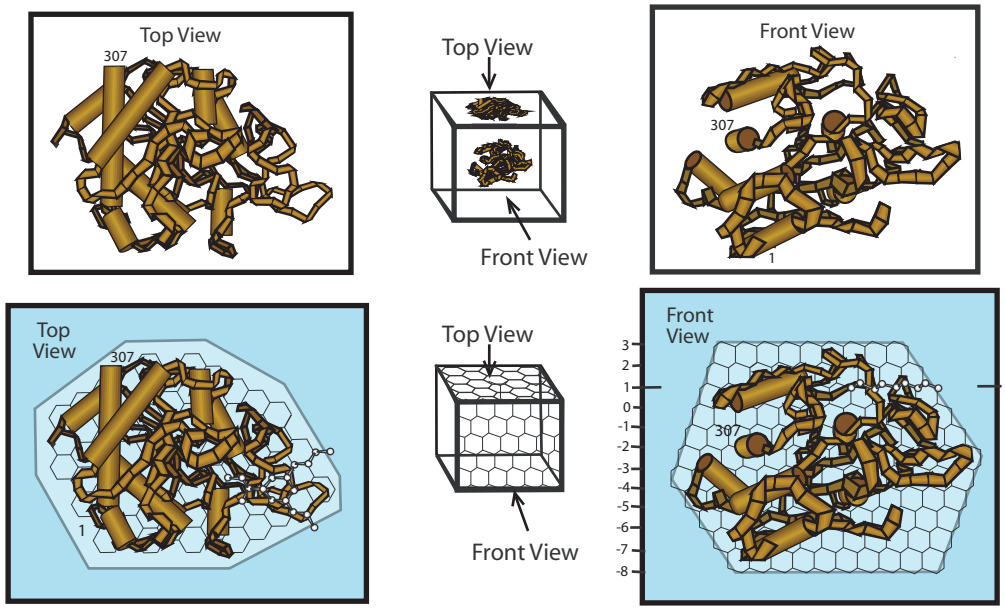




When analyzed as illustrated in the central Front View, the central section of B is composed of so many hydration-ordering peptides, that it rapidly forms a coil between the two glycines,[25] with a short linear segment down to a third glycine. A linear segment then passes up the lipid back side of the coil with the release of unstable ordered water from both surfaces. At the top of B, the linear segment with lipid on the back side, curves over the top of the coil to position the aromatic ring of phenylalanine next to valine on the coil.

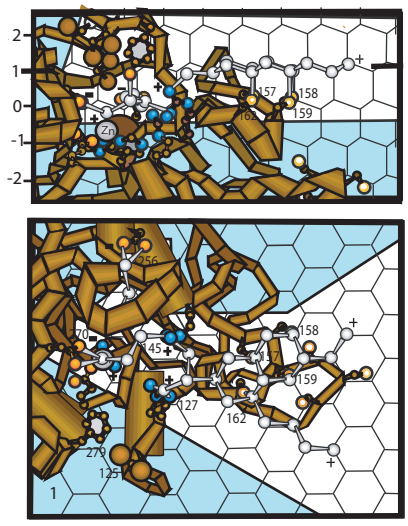
The A-unit, composed of two short coils and a short linear segment with hydration-ordering front side, binds to the lipid back side of B to form the finished protein. Segment C is filled with glycines and prolines to provide the mobility required to carry A into B. Segment D forms a number of coils to transport the insulin molecule through membranes. C an D are removed enzymatically to release insulin into the blood-stream. With only one hydration-ordering surface left on the coil (on the lower right-hand side), insulin binds that side into binding sites in membrane receptors to displace cubically-ordered water, alter shape of the membrane and increase the uptake of glucose into cells.[26]

Just as most of the outer-surface of the insulin molecule is covered by order-disrupting ionic and polar peptides to increase stability and solubility, most surfaces of water-soluble enzymes are covered by order-disrupting peptides; only small regions around reaction sites are surrounded by ordering peptides to direct substrates into the sites and bind them tightly for catalytic reactions. For example, in the analysis of the carboxypeptidase A enzyme,[27] the molecule was oriented so that the long 307 coil and most of the shorter coils were parallel to to linear elements in the cubic lattice.



On the left, the upper surface of the molecule in the crystalline lattice is viewed over the regular hexagonal pattern of water molecules in the cubic lattice, 2.25Å apart in the plane. Notice in the front view with the cubic lattice behind it, that the coils and linear segments present an entirely different appearance: it is as if they are floating on the planes of ice-like water molecules which are numbered to define position. However, it must be remembered - although the cubic lattice is illustrated as a complete unit, only small linear segments are present at any instant as the polypeptide is folding. Lattices presented behind the protein are simply composites of all the linear elements which were propagated as unstable units which left to permit the lipid regions to bind together.

With the xy coordinates of the molecule and cubic lattice overlapped and the upper surface of the molecule at level designated as 2 in the front view, four of the oxygens of serine peptides at positions 157, 158, 159 and 162 are in spatial locations at level 0 to hydrogen-bond with transient linear elements of hydration which tunnel positive charges out from the reaction site, with the zinc ion at the center to direct negatively-charged ends of polypeptide chains into site for cleavage.[28]



Of critical importance, based on multiple analyses,[29] is the fact that *once cubic hydration patterning is established around primary structural units in a polypeptide, it is maintained as a quantized pattern, not only to direct assembly, but enzymatic activity and lipid-binding with other proteins.*[4] Ionic and hydrogen bonding between proteins and other molecules may be at multiple angles but binding between lipid surfaces to form anhydrous cores appears to be guided by cubic patterning.

In his book in 1944, Erwin Schrödinger also proposed that there might be a law, other than thermodynamics, to guide natural molecules spontaneously from randomness toward order.[4] Based on the above analysis of the spontaneous folding and assembly of polypeptides to produce proteins, cubic spatial patterning, may be that "Law." However, at the present time, there is no acceptance or even recognition that cubic hydration patterning may play a role in biomolecular functions and it may be a number of years before it is recognized as important.

The Lumry and Rajender study provided thermodynamic evidence that water forms ordered units on surfaces;[16] the Chen study, that water forms on all surfaces of polypeptides before they form proteins;[14] the Zewail study that surface water forms linear elements and cubic structures on non-polar and poly-ionic surfaces;[17] and, rational interpretions have been developed for five proteins based on the Transient Linear Hydration Hypothesis,[29] and yet, there is no recognition that even linear hydration may occur on lipid surfaces.

Since it is extremely difficult to calculate energy changes in water-bonding,[30] it may be that we will have to wait for advanced computer science to provide positive or negative evidence for the hypothesis and its value in interpreting the role of water in the living cell.

References

1. D. Baker and M. Baek, *Science* **373** (6552): 262-263 (2021). Protein structure prediction is now easier, faster. See also, D. Baek et al., *Science* **373**(6557): 871-876 (2021). Accurate prediction of protein structures and interactions using a three-track neural network.

2. C. B. Anfinsen, E. Haber, M. Sela and F. H. White Jr., *Proc. Natl. Acad. Sci. USA* **47**(9): 1309-1314 (1961). The kinetics of formation of native ribonuclease during oxidation of the reduced peptide chain. Also, C. B. Anfinsen, *Science* **181**: 223-230 (1973).Principles which Govern the Folding of Protein Chains.

3. S. Funkner, M. Havonith and G. Schwaab, *J. Phys. Chem. B* **116**(45): 13374-13380 (2012). Urea, a Structure Breaker? Answers from THz Absorbtion Spectroscopy

4. J. C. Collins, *The Matrix of Life* (Molecular Presentations, 1991). See also, *Biomolecular Evolution from Water to the Molecules of Life* (Molecular Presentations, 2013). For more information, see www.cubichydration.com.

5. E. Schrödinger, *What is Life? with Mind and Matter.* (Cambridge University Press, 1944 and 1967).

6. D. E, Woessner and B. S. Snowden, *Ann. NY Acad. Sci.* **204:** 113-124 (1973). A pulsed NMR study of the dynamics and ordering of water molecules at interfacial systems.

7. F. Mallamace et al., *J. Phys. Chem. B*, **115**(48): 1480-1494 (2011). A Possible Role of Water in the Protein Folding Process. See also, M. Gadzala, D. Dutak, B. Kalinowska, Z. Baster, M. Brylinski, L. Konieczhy, M. Banach and I. Roterman, *J. of Molecular Graphics cular Graphics and Modelling* **87**:117 (2019). The Aqueous Environment as an Active Participant in the Protein Folding Process.

. 8. Chandler, *Nature* **437**: 640 (2005). Interfaces and the driving force of hydrophobic assembly. See also, A. J. Patel, et al. , *J. Phys. Chem.B* **116**(8): 2498-2503 (2012). Sitting on the Edge: How Biomolecules use Hydrophobicity to Tune Their Interactions and Functions.

9. R. Hoyland and L. B. Kier, *Chim. Acta.* **15**: 1-11 (1969). Molecular orbital calculations for hydrogen-bonded forms of water. Also, J. Del Bene and J. A. Pople, *J. Chem. Phys.* **52**: 48-61(1970). Theory of Molecular Interactions: Molecular Orbital Studies of Water.

10 C. Y. Lee, J. A. McCammon and P. J. Rossky, *J. Chem. Phys.* **80**(9): 4448 (1984). The structure of liquid water at extended hydrophobic surfaces.

11. A. H. Narten and H. A. Levy, *Water - A Comprehensive Treatise* pp. 311-332 (Plenum Press 1972). Surface of Liquid Water: Scattering of X-rays.

12. E. Mayer and A. Hallbrucker, *Nature* **317**: 601 (1987). Cubic ice form of liquid water. See also, A. K. Soper, *Science* **297**: 1288 (2002). Water and Ice.

13. Y. Zubavicus and M. Grunze, *Science* **304**: 974-976 (2004). New Insights into the Structure of water with Ultrafast Probes. See also, N. Calder, www.slac.stanford-structureofwater.edu.html.

14. F. Mallamace, C. Cosaro, D. Mallamace, P. Baglion, H.E. Stanley and Sow-Hsin Chen,*J. Phys. Chem.* B **115**(48): 14280-14294(2011). A Possible Role of Water in the Protein Folding Process.

15. N. B. Jee and P. Biswas, *J. Phys. Chem.* B 115 (42): 12257-12265 (2011). Local Order and Mobility of Water Molecules around Ambivalent Helices.

16. R. Lumry and S. Rajender, *Biopolymers* **9**: 1125-1227 (1970). Enthalpy-Entropy Compensation Phenomena in Water Solutions of Proteins and Small Molecules: A Ubiquitous Property of Water.

References

17.. D-S. Yang and A. H. Zewail, *Proc. Natl. Acad. Sci. USA* **106**(11): 4122-4126 (2009). Ordered water structure at hydrophobic graphite interfaces observed by 4D ultrafast electron crystallography. See also, C.-Y. Ruan, V. A. Lobastov, F. Vigliotti, S. Chen, A. H. Zewail, *Science* **304**:80 (2004), Ultrafast Electron Crystallography of Interfacial Water.

18. F. Franks (Ed.) *Water- A Comprehensive Treatise*. (Plenum, 1972).

19. A. Tokmakoff, *Science* **317**: 54 (2007). Shining light on the rapidly-evolving structure of water.

20. P.G. Kosolik and I.M. Svishchev, *Science* **265**: 1219 (1994). Spatial Structure in Liquid Water.

21. E. D. Isaacs, et al., *Physical Rev. Letters* 82(3): 600 (1999). Covalency of the Hydrogen Bond in ice. A Direct X-Ray Measurement. See also, B. Dereka, Q. Yu, N.H.C. Lewis, W.G. Carpenter, J.M. Bowman and A. Tokmakoff, *Science* **371**(6525): 160 (2021). Crossover from hydrogen to chemical

22. B. Kamb, *Structural Chemistry and Molecular Biology* pp. 507-542. (Freeman, 1968). Ice polymorphism and the structure of water.

23. W. Holtkamp, G. Kokic, M. Jagar, J. Mittelstaet, A. A. Komar and M. Rodnina, *Science* **350**:(6264): 1104-1107 (2015). Cotranslational protein folding on the ribosome monitored in real time.

24. T. L. Blundell, J. F. Cutfield, S. M. Cutfield, E. K. Dodson, G. G. Dodson, G. G. Hodgkin, D. A. Mercola and M. Vijayan, *Nature* **231**: 506 (1971). Atomic positions in rhombahedral 2-zinc insulin crystals.

25. K. Lindorff-Larsen, P. Stefano, R. O. Dror and D. E. Shaw, *Science* **334**: 517-520 (2011). How Fast-Folding Proteins Fold.

26. C. W. Ward and M. C. Lawrance, *BioEssays* **31**(4): 422 (2009). Ligand-Induced activation of the insulin receptor.

27. F. A. Quiocho and W. N. Lipscomb, *Adv. in Protein Chem*. **25**: 1 (1971). Carboxypeptidase A.

28. W. N. Lipscomb, *Ann. Rev. Biochem*. **52**: *Rev. Biochem*. **52**: 17 (1983). Structure and Catalysis of Carboxypeptidase A.

29. Check out www.cubichydration.com for linear hydration analyses of four water-soluble enzymatic proteins.

30. A. T. Hagler and H. A. Scheraga, *Ann. N.Y. Acad. Sci.* **204**: 51-75 (1973). Current Status of the Water-Structure Problem; Application *to* Proteins.