

Review

Not peer-reviewed version

Differences and Similarities in Protein and Nucleic Acid Structures and Their Biological Interactions

[Tsutomu Arakawa](#)*, [Tajji Oyama](#), Tomoto Ura, Suguru Nishinami, Kentaro Shiraki, [Teruo Akuta](#)

Posted Date: 6 November 2025

doi: 10.20944/preprints202511.0385.v1

Keywords: peptide backbone;side chain;phosphate backbone;aromatic side chain;secondary structure;molecular interactions



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Review

Differences and Similarities in Protein and Nucleic Acid Structures and Their Biological Interactions

Tsutomu Arakawa ^{1,*}, Taiji Oyama ², Tomoto Ura ³, Suguru Nishinami ⁴, Kentaro Shiraki ³ and Teruo Akuta ^{2,5}

¹ Alliance Protein Laboratories, 13380 Pantera Road, San Diego, CA, USA

² Sales Division, JASCO Corporation, 2967-5 Ishikawa, Hachioji, Tokyo 192-8537, Japan

³ Institute of Pure and Applied Sciences, University of Tsukuba, Tsukuba, Ibaraki 305-8573, Japan

⁴ Institute for Genetic Medicine, Hokkaido University, Sapporo 060-0815, Japan

⁵ Research and Development Division, Kyokuto Pharmaceutical Industrial Co., Ltd, 3333-26, Aza-Asayama, Kamitezuna Tahahagi-shi, Ibaraki, Japan

* Correspondence: tarakawa2@aol.com

Abstract

Protein and nucleic acid play central roles in biology and pharmaceuticals. Both share a similar architecture made of a backbone and side chains. Protein has a peptide backbone and various side chains, whereas nucleic acid has a phosphate backbone and aromatic side chains. However, they are significantly different in the chemical properties of the backbone and side chains. The protein backbone is uncharged, while nucleic acid backbone is negatively charged. The protein side chains comprise widely different chemical properties, while the nucleic acid side chains comprise a uniform chemical property of aromatic bases. Such differences lead to fundamentally different folding and molecular interactions and co-solvent interactions, which are the focus of this review. In regular protein secondary structures, the peptide groups form polar hydrogen bonds, making the interior hydrophilic, while the side chains of different chemical properties are exposed on the outside and participate in molecular interactions and co-solvent interactions. On the contrary, hydrophobic/aromatic nucleobase side chains locate inside the typical double helix or quadruplex structures and charged phosphate groups of the nucleic acid backbone locate outside participating in molecular interactions. The nucleobases are also involved in molecular interactions, when exposed in breaks, hairpins, kinks and loops. These structural differences between protein and nucleic acid confer different interactions with commonly used co-solvents, such as denaturants, organic solvents and polymers.

Keywords: peptide backbone; side chain; phosphate backbone; aromatic side chain; secondary structure; molecular interactions

1. Introduction

Protein and nucleic acid are key players in biological systems and pharmaceuticals. Proteins and nucleic acids are currently two important biological macromolecules in research and biopharmaceutical sectors [1–8]. They are made of a similar architectural framework of a polymer backbone and side chains, which is folded into a secondary, tertiary and quaternary structure. When folded, proteins confer a variety of functions by binding to target molecules, e.g., substrates for enzymes, cell surface receptors and cellular signaling factors for signal transduction, nucleic acids for regulation of transcription and translation, and so on. Such bindings are in large part mediated by amino acid side chains, which are exposed to the protein surface of the folded structure, comprised of charged, hydrophilic, hydrophobic and aromatic groups. While peptide groups can contribute, when exposed, to molecular interactions, they mostly participate in hydrogen bonding to form regular secondary structures of α -helix and β -sheet. On the contrary, the side chains of nucleic acid

are in large part buried in double or other multimeric helices and thus do not participate in molecular interactions, except for the event of transcription and translation. When exposed as in, breaks, loops, hairpins and kinks of nucleic acids, however, the side chains do participate in molecular interactions, e.g., binding to proteins. The charged backbone of nucleic acids that are generally exposed play a major role in interacting with proteins that stabilize and regulate the folding and function of nucleic acids. This review focuses on the common structural framework shared by protein and nucleic acid and on the differences in folding pattern between them, leading to different molecular interactions and interactions with commonly used co-solvents.

2. Structure

Figure 1 shows the primary structure of protein and nucleic acid. Both structures are essentially made of a backbone and side chains. Protein is composed of peptide backbone groups flanked by carbon atoms, to which side chains (R) are attached. Similarly, nucleic acid is composed of phosphodiester linked backbone flanked by ribose, to which nucleobases (B) are attached. As is evident, protein and nucleic acid have a common architecture, a backbone and side chains. However, there are two critical differences between protein and nucleic acid. One is the charges on the backbone, and another is the chemistry of side chains.

The backbone of protein consisting of peptide groups bears no charges, meaning that the peptide groups can be easily accommodated into a folded structure without electrostatic repulsive interactions. The peptide groups form hydrogen bonds and are buried in the regular secondary structures of α -helix (blue) and β -sheet (green), although some peptide groups are exposed in the extended, disordered structures as depicted in Figure 2. The interior of the secondary structures is hydrophilic and segregated from the environments created by side chains or solvents. Such an incorporation of backbone into the secondary structure cannot occur with nucleic acids, in which the backbone phosphate groups bare negative charges and normally are exposed to interact with counter ions or other oppositely charged groups.

Side chains are also significantly different between protein and nucleic acid. The side chains of protein are composed of widely different chemical properties, as schematically shown by R1, R2, R3, R4, etc, in Figure 1. Tryptophan (Trp), tyrosine (Tyr) and phenylalanine (Phe) are aromatic and hydrophobic and participate in hydrophobic interactions and aromatic interactions as in π - π stacking and π -cation interactions. Leucine (Leu), isoleucine (Ile), valine (Val) and alanine (Ala) are hydrophobic, while threonine (Thr), serine (Ser), glutamine (Gln) and asparagine (Asn) are hydrophilic and normally exposed to solvents in the folded tertiary structure. Aspartic acid (Asp) and glutamic acid (Glu) are negatively charged and can contribute to electrostatic ion pair interactions and anion- π interactions [9]. Arginine (Arg), lysine (Lys) and histidine (His) carry a positive charge and can participate in electrostatic interactions with ions or oppositely charged groups. These positive charges, particularly, Arg, participate in key molecular interaction with aromatic compounds and aromatic side chains as π -cation interaction as described later [10–12]. Arg has a planer side chain structure that can form π - π stacking with aromatic groups [13–15].

Nucleic acid is different in chemical property of side chains from proteins. The side chains of nucleic acid are made of a highly similar chemical property. Namely, they are aromatic and hydrophobic, as is evident from their favorable (attractive) interaction with organic solvents, alkyl-urea and alkyl-amides, particularly with longer alkyl-chains [16,17]. The aromatic properties of nucleobases are also evident from their attractive interactions with aromatic arginine side chain, i.e., guanidinium group [18]. Thus, when protein and nucleic acid are folded into a secondary structure, the principal interaction mechanism is fundamentally different between them.

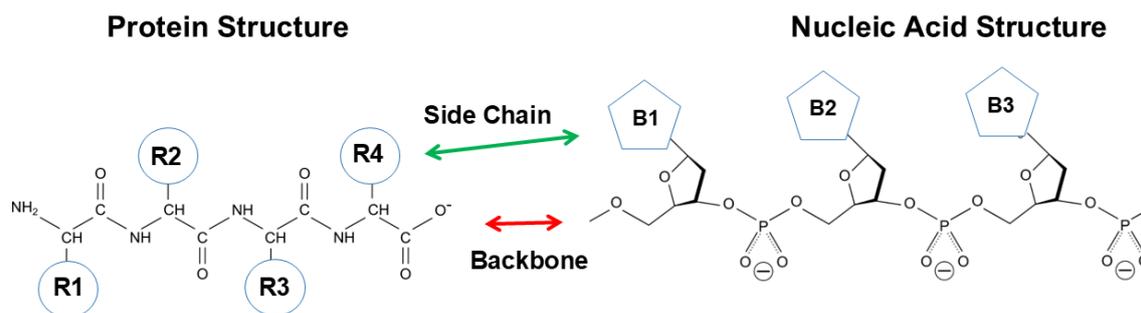


Figure 1. Comparison of protein and nucleic acid structures.

Major secondary structures of protein are α -helix and β -sheet, as depicted in Figure 2. Key factors in these protein structures are as follows. The forces that stabilize them are hydrogen bonding between peptide groups, i.e., no involvement of side chains. Thus, the inside of helix and sheet is hydrophilic, comprising hydrogen bonds. As described later, such hydrophilic hydrogen bonding is augmented in non-polar environments provided by hydrophobic side chains, organic solvents or binding of amphiphilic detergents [19–24]. The outside of these regular secondary structures of protein is made of side chains with widely different surface chemistries, as depicted in Figure 1 and 2. Depending on which side chains are exposed, the surface of the protein's secondary structures can be charged, hydrophilic, hydrophobic or aromatic and can participate in a variety of interactions responsible for formation of tertiary structure folding [9] or interactions with other molecules for formation of quaternary structures [25]. It can lead to self-association in oligomeric or assembled structures, such as oligomeric enzymes, collagen, F-actins, and microtubules [26]. Uneven distribution of these side chains can generate specificity in molecular interactions.

One of such examples of uneven side chain distribution is called "helical wheel", schematically shown in Figure 3 [27–29]. In α -helical structures, hydrophobic side chains are often distributed in one side of a helix (red line), while the other side is made of hydrophilic side chains (blue). This would make one side of helix too hydrophobic to be exposed to aqueous environments and hence tend to be buried in hydrophobic environments. This can be achieved by interacting with other helices with similar surface properties, when another helical wheel structure is available as seen in Figure 3, forming so-called α -helical bundles (right panel). This plays two important roles. Hydrophobic interaction between two or more helices mutually stabilizes each other as a single helix is not stable but can be stabilized by bundling or burying the helix into hydrophobic microenvironments, which destabilizes exposure of polar peptide groups. Exposure of peptide bonds to water may cause dissociation of inter-peptide hydrogen bonding and instead form hydrogen bonds with water molecules [30,31]: namely, hydrogen bonds between peptide groups are more stable in hydrophobic environments. Another important factor in helical wheel and α -helix bundling is the hydrophilicity of the other surface (2 blue lines), which would make the outside surface water-soluble. Namely, the overall folded structure becomes water soluble by burying non-polar groups inside.

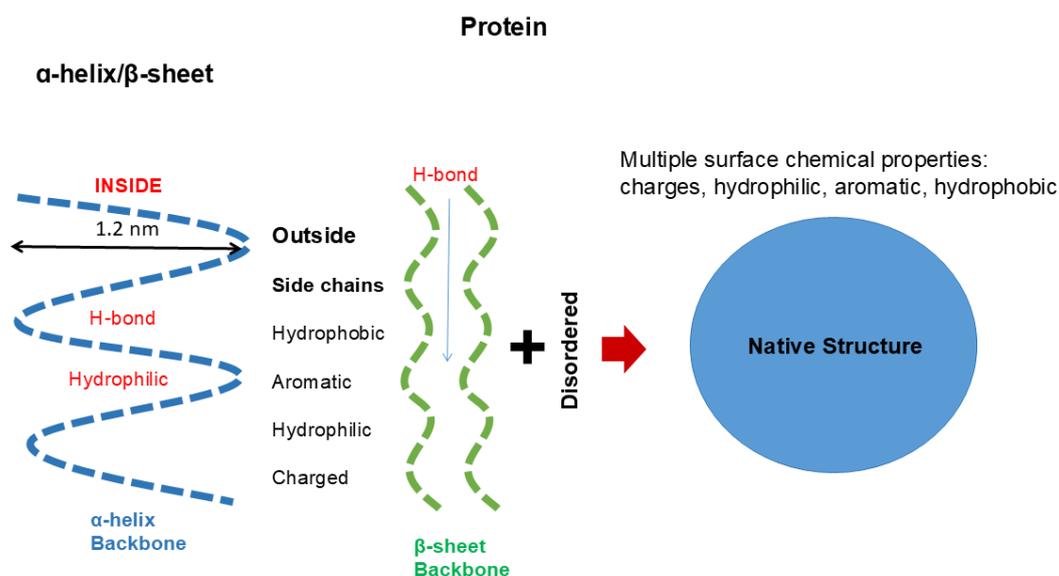


Figure 2. Secondary and tertiary structure of protein. A typical α -helix (blue) and β -sheet (green) structure is stabilized by hydrogen bonding between peptide groups, which render the interior of these structures hydrophilic. The exterior surface of α -helix and β -sheet are composed of amino acid side chains of different chemical properties. The disordered structures have both peptide groups and side chains exposed. Molecular interactions between these structures make the protein fold into a distinct globular structure.

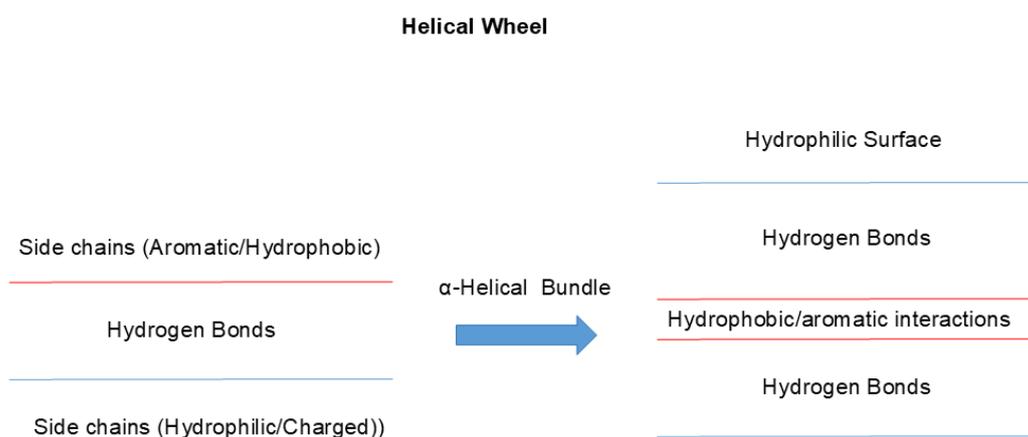


Figure 3. Architecture of helical wheels. Helix often assumes an amino acid orientation called “helical wheel”, in which one side of α -helix has a higher population of hydrophobic side chains and the other side has a higher population of hydrophilic side chains. Two or more α -helices bind together through hydrophobic surface, which would stabilize the helix. The other hydrophilic surfaces are exposed to aqueous solvent, which would make the helices more water-soluble.

This cannot happen in nucleic acid having very different backbones and side chain chemistries. However, protein-like secondary structure may be possible by attaching nucleobases to carbon atoms, which are flanked by peptide groups. Such a construct is called “Nucleopeptides”, which are composed of nucleobases inserted on a peptide backbone [32]. They exhibit interesting features due to their capacity to bind complementary single stranded RNA and DNA sequences via hydrogen-bonding between a nucleopeptide and a single stranded nucleic acid. As in nucleic acid double helix or quadruplex, base stacking in the nucleopeptide and the single nucleic acid strand can enforce hydrogen bonding. There appears no evidence of protein-like α -helix or β -sheet in nucleopeptides by CD measurements: namely, the nucleopeptide does not appear to form peptide bond-mediated hydrogen bonds and resultant protein-like helix or sheet.

What happens in nucleic acid structure is shown in Figure 4. It shows a typical double helix with nucleobase and ribose groups inside and negative charges of phosphodiester linkage outside. Its

inside is thus occupied by hydrophobic/aromatic nucleobases forming aromatic stackings and hydrogen bonds between the base pairs. Such structure renders the inside aromatic and hydrophobic and, when exposed, can be stabilized by non-polar organic solvents, different from peptide hydrogen bonds that will be destabilized by non-polar organic solvents. Because of bulky bases inside the helix, the distance between two strands is 2 nm (see Figure 4), larger than 1.2 nm (Figure 2) of protein helix diameter. It is important to note that the outer surface of nucleic acid double helix is very different from the outer surface of protein helix. The outer surface of nucleic acid double helix or quadruplex is highly charged and hydrophilic, while the protein's helix has the outer surface composed of variable chemistry, i.e., charged, hydrophilic, hydrophobic or aromatic.

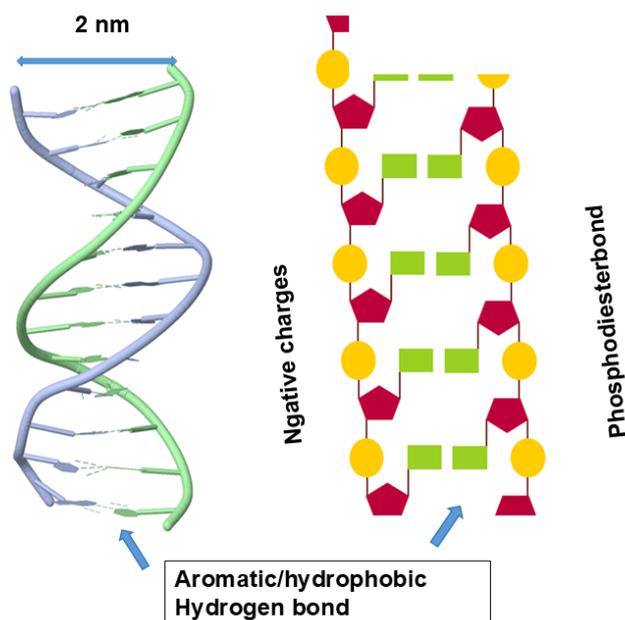


Figure 4. Secondary structure of double helix. The interior of the double helix is made of aromatic nucleobases, which would make the interior aromatic/hydrophobic. The outer surface has a constellation of negative charges arising from phosphate groups. Modified from [nucleic acid structure - Search Images](#).

Thus, formation of packed tertiary structure folding is very different between protein and nucleic acid. As described earlier, side chains of protein secondary structures are fully exposed and can interact with other side chains. Such inter-side chain interactions lead to packing of the secondary structures into a folded tertiary structure with a resultant globular shape. As shown in Figure 2, the exposed peptide groups as well as side chains of disordered secondary structure can also participate in formation of the folded structure. There are many key side chain interactions in the packed structure. Among them, charge pair between negatively and positively charged side chains is buried and stabilized by mutually neutralizing the charges. Such electrostatic interaction could be stronger in hydrophobic environment of protein's interior due to low dielectric constant [33]. Hydrophobic side chains are segregated from solvent water and stabilized by interacting with other hydrophobic side chains and form a core structure. In addition, aromatic interactions also play a key role in the folded structure by π - π interaction with other aromatic side chains and π -cation interaction with positively charged side chains, particularly arginine side chain that has a planer structure, as has been described earlier.

These interactions in protein's packed structure do not occur in nucleic acids. As described earlier, side chains of nucleic acid are buried inside of the double helix or quadruplex and outside surface has a constellation of negative charges, markedly different from the chemistry of the outer surface of protein's helix or sheet. Thus, folding of nucleic acid, if occurs, must accompany charge neutralization by counter ions to suppress charge repulsion of backbone phosphate groups or

binding to positively charged side chains from other compounds, e.g., polyamines, or proteins. Such nucleobase side chain-mediated interaction may occur through terminal bases, which are at least partially solvent-exposed. In addition, double strand helix of nucleic acid often has breaks, kinks, loops and hairpins, as depicted in Figure 5, which result in exposure to nucleobases. Flipping of nucleobases can also occur and exposes those nucleobases for interaction with other structures.

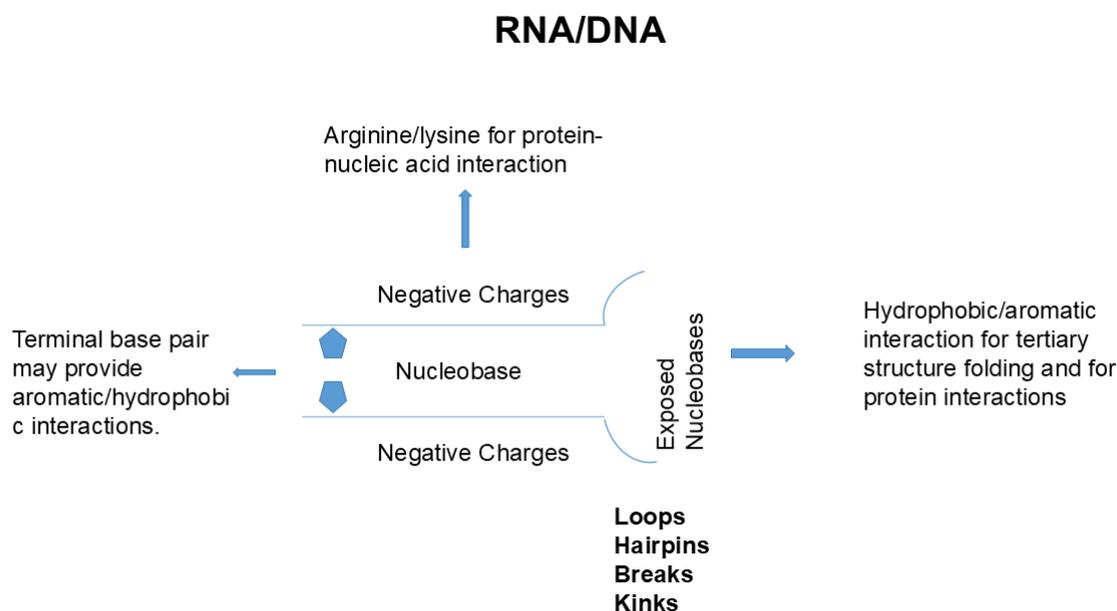


Figure 5. Schematic diagram of possible interactions between nucleic acids and proteins.

3. Interactions

Amino acid side chains play a key role in not only folding of tertiary structure but also intermolecular interactions, whether specific or non-specific. Proteins perform specific functions by binding to themselves or target proteins or nucleic acids or undergo non-specific aggregation. Nucleic acids also perform functions by binding to proteins. These molecular interactions are in vitro controlled by so-called “co-solvents” including salts, sugars, polyols and polymers, from which we can learn about the nature of molecular interactions.

3.1. Macromolecular Interactions

Amino acid side chains provide specific interactions within a protein molecule to generate a unique three-dimensional structure that is critical for their functions. In the folded structure, protein molecules have the solvent-exposed side chains of different chemical properties, as described earlier, which participate protein-protein interactions. Figure 6 shows an example of protein-protein interactions. Figure 6A shows two folded proteins (human G-protein subunit alpha; $G_{\alpha 3}$ and engineered regulator of G-protein signaling type 2 domain; RGS2) that form specific complex, a critical protein-protein interaction for signal transduction [34]. First, it is noted that both proteins are packed into a more or less globular shape, which generates rigid surface structure. These two proteins bind to each other at a specific site as seen in Figure 6, with tightly fit interface, with complimentary morphology. Many side chains, e.g., Glu65, Thr182, and Lys210 of $G_{\alpha 3}$ and Ser106, Lys191, Asp184, and Arg188 of RGS2, are involved to confer electrostatic interactions and hydrogen bonding, which generate specific and strong binding of these two proteins as shown in Figure 6A.

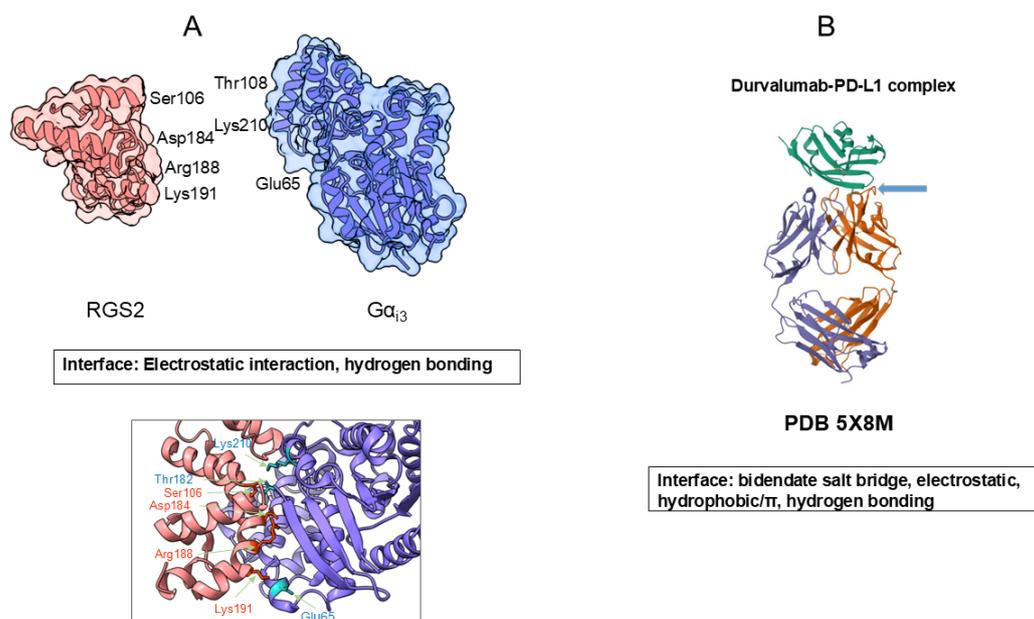


Figure 6. Model protein-protein interaction. A. Interaction between human G-protein subunit alpha, $G_{\alpha i3}$, and engineered regulator of G-protein signaling type 2 domain, RGS2. B. Antigen-antibody interaction between human antibody durvalumab and antigen PD-L1. Structure A was rendered from PDB entry 2V4Z using UCSF ChimeraX version 1.10.1, and structure B was based on PDB ID 5X8M.

Figure 6B shows the antigen-antibody interaction in which a human antibody durvalumab binds to an antigen, PD-L1 [35]. It is shown here that both the antibody and PD-L1 domains are also more or less folded into rigid tertiary structure. As shown, multiple side-chain contacts contribute to the interface, including bidentate salt bridges between PD-L1 Arg113 and the antibody heavy-chain Asp31, and an ion pair between PD-L1 Glu58 and light-chain residue Lys52. In addition, PD-L1 Asp26 forms an ion pair with the antibody light-chain Arg28, accompanied by several hydrophobic contacts. Notably, PD-L1 Arg125 forms an unusual hydrogen bond with the antibody peptide backbone. In the case of atezolizumab (not shown in Figure 6B), Tyr56 of PD-L1 engages in hydrophobic/ π - π interactions with heavy-chain Trp33, Trp50 and Ser57. Among these interactions, aromatic contacts are thought to play a central role in mediating strong and specific antigen-antibody binding [36–38].

Figure 7 shows a few examples of protein-nucleic acid interactions that play a central role in regulating and stabilizing nucleic acids. Here, the interactions of a few RNA binding proteins with three different structures of RNA are shown as examples. The global folding of proteins and nucleic acids appears to be different, As in Figure 6, these proteins are folded as a globular shape, while nucleic acids lack such a rigid folding, which in turn suggests a more flexible nature of nucleic acid folding than protein folding. Figure 7A shows binding of splicing factor 1 (SF1) with a single stranded RNA to generate spliceosome assembly [39]. In the SF1-RNA complex (PDB 1K1G), the 5'-terminal bases (UAC) are stabilized by the QUA2 helix (residues 239–249), while the 3'-terminal bases (UAAC) interact with the KH domain, particularly residues P159, R160, and G161 within the Gly-Pro-Arg-Gly loop, via hydrogen bonds and hydrophobic contacts. Lys and Arg in SF1 bind to the RNA backbone phosphate groups by electrostatic interactions. One of the exposed bases is deeply buried in hydrophobic cleft comprising a Gly-Pro-Arg-Gly motif and hydrophobic (Leu, Ile, Val)/aromatic (Tyr, Phe) amino acid residues. As summarized in Figure 7A, a few solvent-exposed aromatic nucleobases as well as phosphate backbones interact with the solvent-exposed amino acid side chains through electrostatic, hydrogen bonding and hydrophobic/aromatic interactions. Figure 7B shows a case of double stranded RNA, in which a RNA cleaving enzyme ADAR2 deaminase binds to a DNA double helix [40]. Glue488 of the deaminase intercalates into the RNA double helix and binds to the orphan (unpaired) base by hydrogen bond and enables the base to flip into the enzyme's active site. The enzyme's residue 454-477 that includes Arg455 forms a loop that inserts into the target RNA and

contacts the phosphate backbone. Such RNA binding renders the enzyme's loop ordered in structure and causes the base to flip. Multiple basic residues, e.g., Lys594 and those in the loop, form electrostatic interactions with the phosphate backbone as seen in Figure 7B.

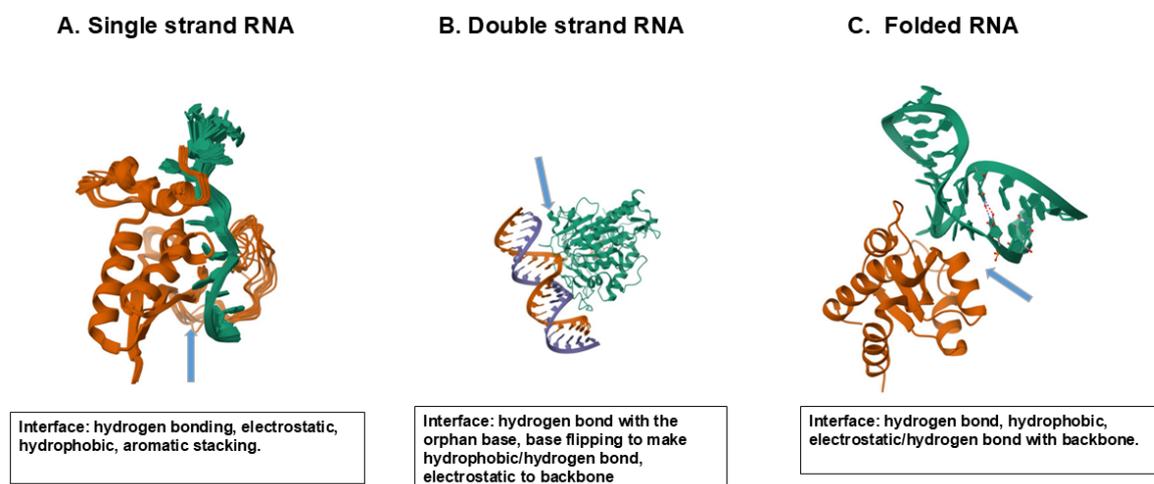


Figure 7. Model protein-nucleic acid interaction. (A) Protein binding to a single strand RNA. Splicing factor 1 (red). Single stranded RNA (green). (B) Protein binding to a double helix RNA. ADAR2 deaminase (green). DNA double Helix (red/blue). (C) Protein binding to a folded RNA. L7Ae (red). K-turn RNA structure (green). Structures A, B, and C are derived from PDB entries 1K1G, 5ED2, and 1RLG.

In Figure 7C is shown an example of binding of folded RNA to an RNA-binding protein L7Ae, a typical model of K-turn binding motif [41,42]. The L7Ae binds to the K-turn structure formed by two sheared base pairs and stabilizes the kinked RNA conformation as seen in Figure 7C, using various amino acid side chains. The Glu38 forms hydrogen bonds with the terminal guanine base. Arg95 interacts with the phosphate backbone and Ser94 hydrogen bonds with the RNA backbone. Val94 and Ile92 interact with the terminal guanine by hydrophobic interactions as summarized in Figure 7C.

Another example of protein-nucleic acid interaction is described in Figure 8 for quadruplex DNA, which is formed by a single stranded DNA, in which four bases stack over each other [43]. Three major binding sites on the quadruplex structure have been demonstrated, i.e., top-stacking [44], groove-binding [45] and loop-binding [46]. In top stacking, it has been shown that Arg and Lys of the quadruplex binding protein DHX36 bind to phosphate backbone via electrostatic and hydrogen bond interactions, while the aromatic amino acid of DHX36 has hydrophobic interaction or aromatic stacking with the terminal guanine base. In groove-binding, the telomeric end-binding protein binds to the quadruplex groove by Arg/Lys participating in electrostatic interactions with phosphate backbone and Tyr participating aromatic stacking with guanine base. In loop-binding, e.g., binding of zinc finger proteins, a basic His/Arg/Arg cluster binds to the outward-directed nucleotides, most likely through cation- π or π - π interactions.

In protein-protein and protein-nucleic acid interactions, arginine side chains can play a central role by engaging in various molecular interactions, including electrostatic interactions with charged groups and aromatic interactions with aromatic side chains. The mode of interactions of arginine side chains with other side chains has been well documented in protein structure. Among them, cation- π and π - π interactions play a significant role in protein folding and protein-protein interactions as described above. How does arginine contribute to the interactions of protein with nucleic acid? Different from proteins, nucleic acids are highly charged, and hence electrostatic force makes a larger contribution to the interaction of basic arginine side chain with nucleic acids than with the proteins. It has been shown that arginine residues participate in hydrogen bonding to the phosphate groups in nucleic acids [47]. This hydrogen bonding was often termed "arginine fork", as arginine side chain can form a specific network of hydrogen bonds with phosphate groups of RNA. In virus particles,

arginine side chains (more likely just a guanidinium group of Arg side chain) bridge with phosphate groups in the virus nucleic acids, while the aromatic side chains of the viral proteins engage in hydrophobic interactions with nucleobases and the aspartic acid side chains establish hydrogen bonds with sugar groups of RNA within the viral structure [48,49]. Arginine forks can participate in multiple hydrogen-bonding with phosphate groups and nucleobases to mediate nucleic acid binding to proteins [47,50,51].

Participation of hydrogen bonding between arginine side chain and phosphate groups has also been indicated in not only protein-nucleic acid interaction but also protein-protein interaction via phosphorylated serine residues on the proteins [52]. Such interaction of arginine side chains was also observed with polyphosphate compounds via bivalent hydrogen bonding, which cannot be achieved by lysine residues [53]. Arginine residues interact with nucleic acids via hydrogen bonding to phosphate backbone groups and nucleobases, and also via π -cation interaction with the nucleobases [54]. The ability of arginine to participate in multiple binding mechanisms for nucleic acids can lead to formation of stable protein-nucleic acid complexes. One of such binding mechanisms is aromatic/ π -cation interactions of arginine side chain with nucleobases.

Solubility measurement has been used to demonstrate interactions of nucleobases with amino acids and related compounds [18]. The above study showed a monotone increase in solubility with increasing arginine and GdnHCl concentration. Both arginine and GdnHCl co-solvents were effective in increasing the solubility of these bases, while glycine and NaCl either were ineffective or even decreased the solubility [18]. The results indicate that the interaction between arginine and nucleobases is not of electrostatic nature, as the electrostatic interactions can be replaced by NaCl and glycine. Between GdnHCl and arginine, arginine showed more favorable, attractive interactions with nucleobases. Arginine at 1 M was most effective on adenine (2.5-fold increase in solubility) followed by cytosine, guanine, thymine and uracil (1.3-fold increase), resulting in the same solubility order for these five nucleobases in 1 M arginine as in the pH 7 buffer. There appears to be no particular correlation with their structure parameters, e.g., the number of double bonds. The above observation is consistent with the reported favorable interaction of arginine residues (i.e., arginine side chain) with nucleic acids as described above.

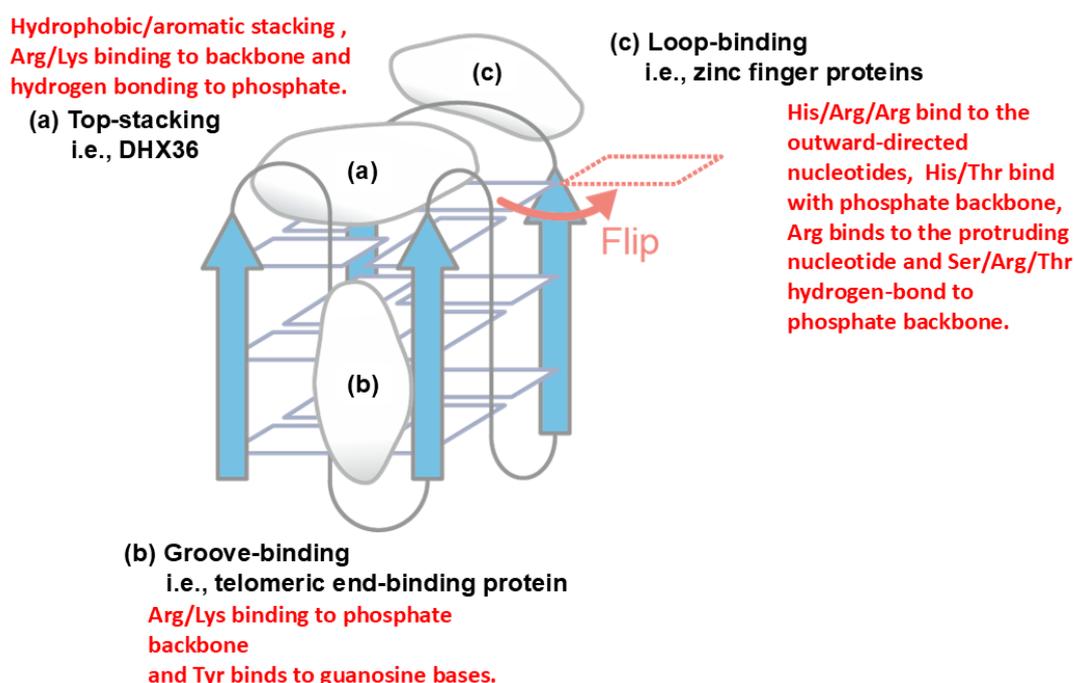


Figure 8. Potential interaction modes of a model quadruplex structure. (a) Top-stacking binding mode of DHX36. (b) Groove-binding mode of telomeric end-binding protein. (c) Loop-binding mode of zinc finger protein.

The illustration was originally created with reference to Ref. 43.

3.2. Biomolecular Condensates

The molecular complexes described above involve specific and strong interactions between proteins, between nucleic acids and between protein and nucleic acid. There are other types of biomolecular interaction that may occur in cells or *in vitro*. They are weak and transient molecular interactions, which may be augmented at high macromolecular concentrations (i.e., crowded environments) and often appear in cellular environments, forming a network structure termed “condensate”. When formed *in vitro*, such network structure can cause high viscosity, which leads to serious problems in pharmaceutical development, e.g., downstream process of protein purification and parenteral injection [55]. Recent studies have highlighted the crucial role of biomolecular condensates that regulate various cellular functions [56]. Proteins and nucleic acids are the principal molecular drivers of the condensates, and the mechanisms underlying their interactions are gradually being elucidated. In the context of condensate formation in cells under physiological environments, cooperative interactions, although weak and transient, between proteins and nucleic acids are involved. Under controlled *in vitro* settings, both proteins and nucleic acids are also individually capable of forming network structure, allowing detailed mechanistic analyses of condensate formation.

Condensates that are formed in cells or *in vitro* are detected by turbidity (light scattering) or hydrodynamic and thermodynamic measurements. Although their interrelation is not fully demonstrated, those condensates are often accompanied by optical observation as liquid droplets, termed “liquid-liquid phase separation”. Due to difference in refractive index between bulk solvent and condensate, those droplets can be detected by light microscopy. Those membrane-less cellular droplets/organelle include, but not limited to, nucleolus, speckles, paraspeckles, nuclear bodies, Cajal bodies, P-bodies and stress granules. Such phase separations are also observed *in vitro* under well-controlled conditions.

A common feature of both systems is their ability to form multivalent and dynamic interaction networks. Molecules with high structural flexibility and multiple interaction motifs are more prone to condensate formation. It should be noted that the multiple interactions described here differ from the multivalent interactions that confer binding avidity, as depicted in Figure 9. Figure 9A shows multivalent (here trivalent) interactions between two macromolecules, which confer binding avidity. Since these interactions occur on the same two molecules, each interaction adds binding energy to the dimeric assembly, leading to its strong interactions. On the contrary, as shown in Figure 9B, two molecules are bound by only a single interaction. However, each molecule offers multiple binding sites for other molecules, which lead to many molecules connected to each other, generating a network of macromolecular chains. This generates a network structure throughout the solution, a cause of high viscosity. When the network structure is separated from the bulk phase, it should lead to a higher macromolecular density and refractive index, resulting in appearance of light scattering and refractive index gradient between the bulk and the solution phase containing the network.

It has been shown that the liquid droplet formed around the condensates are often spherical or globular, not fully expected from the possible irregular shape of condensates. This may be explained by dynamic and fluid nature of the liquid droplets. Since the condensate creates higher density and refractive index, it generates interface between the droplet and bulk and hence the interfacial tension. If the condensate and droplet is not globular, as shown in Figure 9C, the interfacial tension and hence the surface free energy is greater than those expected of the globular shape. When the interior of the droplet is fluid and not viscous, and the molecular interactions stabilizing the condensate are weak, flexible and transient, the network structure trends to rearrange to become more globular, as shown in Figure 9D, thereby reduce the surface free energy. If the droplet phase is viscous and the molecular assembly in the condensates is rigid, the network structure will not rearrange, and the droplet will stay non-globular.

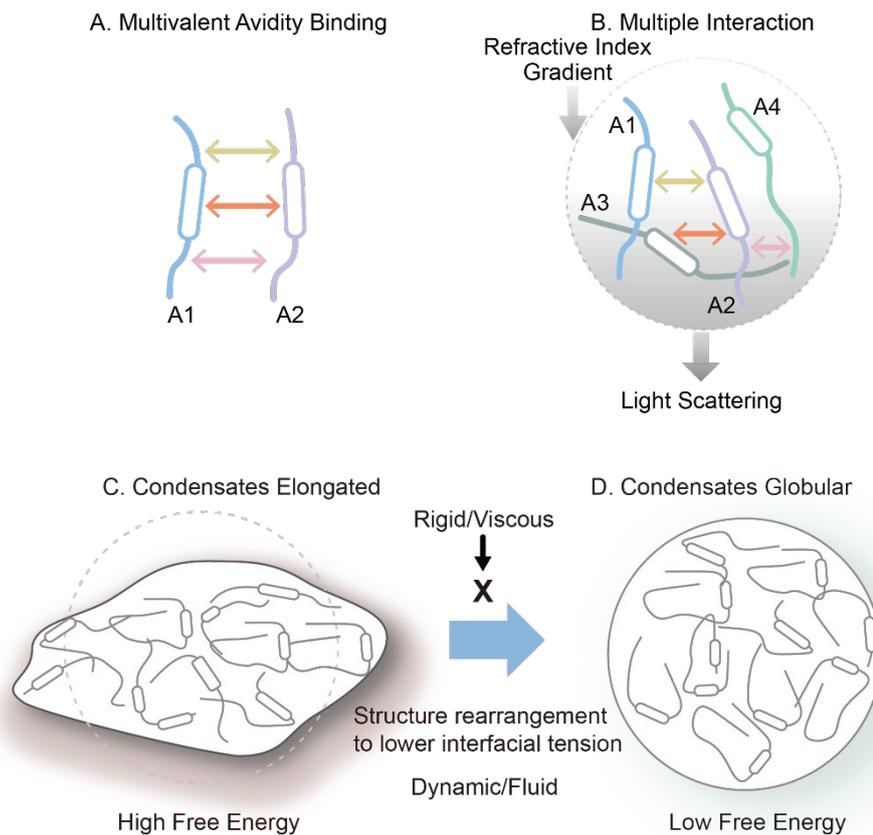


Figure 9. Schematic illustration of macromolecular assembly and condensate formation. Multivalent avidity binding mode between two flexible molecules (A). Multiple interaction mode between four flexible molecules (B). Conversion of non-globular condensate/liquid droplets (C) to globular condensate (D). This is an original diagram drawn by us.

In proteins, intrinsically disordered regions (IDRs) and low-complexity regions (LCRs) provide multiple weak interaction sites—such as RGG motif [57] and aromatic residues [58]—that drive reversible condensation. Similarly, in nucleic acids, flexible single-stranded RNAs exhibit higher phase-separation propensity than rigid double helices [59]. In this sense, proteins and nucleic acids are similar in terms of the structure that drives formation of condensates and liquid droplets. Although the negatively charged phosphate backbone may seem to hinder self-assembly, base pairing and base stacking between nucleobases can facilitate condensation [60]. Furthermore, planar higher-order structures such as guanine quadruplexes (G4) can generate hydrophobic π -stacking interfaces that serve as nucleation sites for condensation [61]. In cells, such homotypic interactions are complemented by heterotypic interactions between proteins and nucleic acids, which play a central role in driving condensates. Positively the solvent-exposed charged side chains of proteins, such as arginine and lysine, electrostatically interact with the negatively charged phosphate backbone, while π - π and π -cation interactions between aromatic and basic residues in proteins and nucleobases exposed in single strand or loops, breaks, hairpins and kinks of double and quadruplex nucleic acids, stabilize the condensates [62]. These cooperative interactions give rise to Protein-RNA condensates (RNP condensates), forming the basis of membrane less cellular structures described above. Higher-order nucleic acid structures—including G4s, and hairpins [63]—also function as structural elements that modulate condensate formation. While the importance of higher-order nucleic acid structures in regulating biomolecular condensates has been increasingly recognized, their precise roles in condensate regulation remain an intriguing subject for future investigation.

The RGG motif forms condensates not only with RNA but also with folded DNA sequences such as the promoter region of the oncogene c-Myc and the G₄C₂ repeat of the C9orf72 gene, which has

been implicated in amyotrophic lateral sclerosis (ALS) [64,65]. Moreover, not only flexible peptides but also chromatin-associated proteins containing both well-folded domains and intrinsically disordered regions can form condensates with the KRAS promoter G4 [66]. These condensates between nucleic acids and proteins require the formation of folded G4 structure. A 5-methylcytosine modification of DNA has been shown to influence condensate formation, an effect that arises from its ability to modulate the topology of G4 structures [65]. The formation of G4 creates planar G-quartet surfaces, multiple loops, and flanking regions, which together provide an ideal platform for multiple interactions with proteins [67,68]. Furthermore, a G4 formation in the BCL3 promoter has been shown to promote transcriptional condensate formation, whereas the G4 in the transcript exerts negative feedback on condensate formation, suggesting that DNA and RNA G-quadruplexes mutually influence condensate dynamics [69]. It appears that the G4 structure is unique in that, although folded, it can drive condensate formation.

3.3. Co-Solvent Interaction

This review focuses on the similarities and differences in structure between protein and nucleic acid, which result in different folding and molecular interactions. Why then should this review include co-solvent interactions with protein and nucleic acid? Nowadays, co-solvents are widely used *in vitro* for manipulating aqueous solutions containing protein and nucleic acid, e.g., in purification, chromatography, long-term storage, formulation and solubilization. Understanding how the co-solvents interact with these macromolecules should help design their optimal applications. However, earlier studies on co-solvent interactions with proteins were aimed at understanding the nature of protein folding. Namely, co-solvent interaction pattern was used to define hydrophobicity scale of amino acid side chains and peptide backbones, which was then correlated with their preference for water or internal hydrophobic core [70,71]. Thus, introduction of earlier studies on co-solvent interaction should shed light on the differences in protein and nucleic acid folding.

3.3.1. Organic Solvent

There is wealth of information on the interaction between protein side chains and organic solvents [70,72–74]. It was shown that a highly polar glycine, which has negligible side chain contribution to solvent interaction, drastically loses aqueous solubility with increasing concentration of organic solvents, including ethanol, dioxane, ethylene glycol and dimethyl sulfoxide, indicating that the interactions are repulsive between the polar glycine and the organic solvents. It is noted that although glycine has terminal amino and carboxyl groups, it is a constituent of peptide backbone structure. Hydrophilic groups of protein side chains, e.g., glutamine and asparagine, also have repulsive interactions with the organic solvents. Namely, these hydrophilic side chains and glycine dislike non-polar environments and like aqueous environments. The histidine side chain showed negligible interactions with the organic solvents, which may be ascribed to a balance between the repulsive hydrophilic interaction and attractive aromatic interaction of histidine side chain with the organic solvents. There are no interaction data available between charged side chains and organic solvents. However, one can speculate from the glycine data that charged amino acid side chains (i.e., Asp, Glu, Lys and Arg) have repulsive (unfavorable) interactions with the organic solvents, consistent with repulsive interaction of ionic electrolytes with organic solvent [75]. More importantly, polar peptide groups, which constitute the polypeptide backbone structure, showed repulsive interactions with organic solvents, e.g., ethanol, dioxane and ethylene glycol [70,73]. This means that the peptide groups will be segregated from contact with organic solvents and favor inter-peptide interactions or interactions with other polar groups. Nozaki and Tanford also did extensive measurements of the interactions between aromatic/hydrophobic groups of protein side chains, i.e., Trp, Tyr, Phe, Ile, Leu, Val and Ala, and organic solvents [70,73]. Organic solvents showed attractive interactions with these hydrophobic and aromatic groups.

The interactions of nucleic acid constituents with organic solvents have not been studied as extensively as with those of proteins. The mechanism of phosphate backbone interaction with organic

solvents should be the same as the mechanism of protein charges and hence repulsive. The repulsive interactions between organic solvents and phosphate backbones forming charge clusters (see Figure 4) could even be stronger than separated charged amino acid side chains. The question is whether nucleobases are similarly aromatic to the aromatic side chains of proteins. To our knowledge, there seem to be few data for the interactions between aromatic nucleobases and organic solvents, although we may expect them to be attractive, as seen with the attractive interactions between organic solvents and aromatic side chains of the proteins. Solubility measurements showed a monotone increase in solubility of adenine in the presence of several organic solvents, including *N,N*-dimethylformamide, *N*-methyl pyrrolidone, propylene glycol and dimethyl sulfoxide [76] and uracil in the presence of ethanol below ~50 % [77]. Thus, these heteronuclear aromatic nucleobases showed favorable, attractive interactions with non-polar organic solvents, which are like the aromatic amino acid side chains in proteins.

How do organic solvents affect protein and nucleic acid structure and what do we learn from the effects of organic solvents on their structure? Figure 10 schematically shows the effects of organic solvents on protein structure. Organic solvents induce denaturation of the proteins, which can be explained by organic solvents binding to (stabilizing) hydrophobic core comprising mostly hydrophobic amino acid side chains [74,78,79]. This is consistent with the established hydrophobic scale of side chains, as those side chains are segregated from aqueous environments and buried in the interior of the protein molecules. Upon unfolding, these side chains are stabilized by organic solvents by hydrophobic interactions. How do organic solvents affect secondary structures? Two factors need to be considered in understanding the effects of organic solvents on protein secondary structure. First factor is the consequence of denaturation (unfolding of tertiary structure), which is brought by the organic solvents. This eliminates interactions between secondary structures, i.e., inter-helical or inter-sheet interactions. For example, it was shown that hydrophobic side chain interactions of helical wheel stabilize helices forming bundles, which will be eliminated during denaturation by organic solvents. Thus, this factor should result in unfolding (loss) of secondary structure. However, the second factor operates against unfolding effects of organic solvents on secondary structure. Namely, the interaction between polar peptide group and organic solvent is repulsive, which favors segregation of peptide groups from non-polar environments of organic solvents. Thus, this second factor by organic solvents should stabilize the secondary structure as shown in Figure 10. Which of two factors dominate over the other depends on non-polar nature of the organic solvents used and the strength of hydrophobic interactions between the secondary structures. In fact, a number of organic solvents and detergents that provide non-polar environments for peptides have shown enhanced helical formation [80–83].

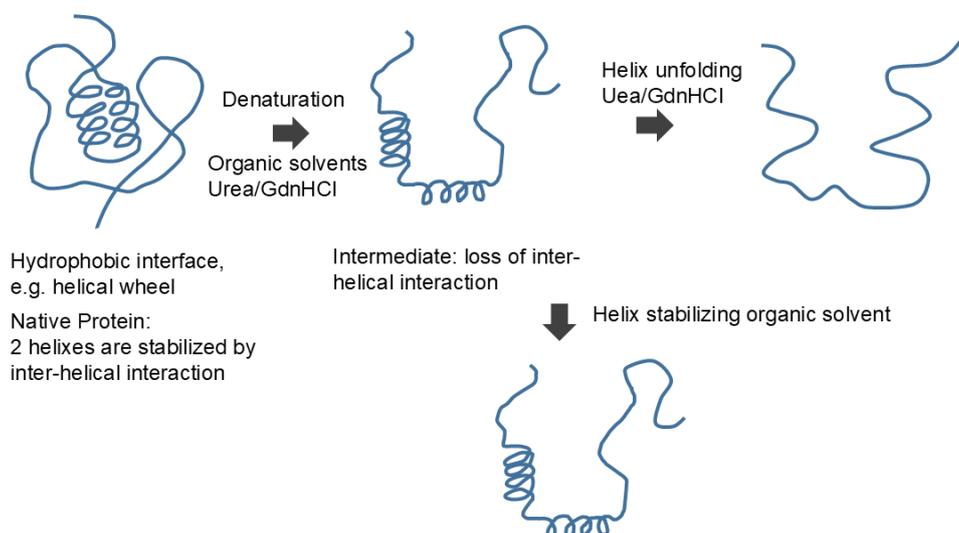


Figure 10. Schematic diagram of protein denaturation. Native protein is unfolded by organic solvent or denaturants, which may form an intermediate structure, which can cause loss of interactions between regular

secondary structures. Denaturants can further unfold the secondary structures, while organic solvent may stabilize the secondary structures. This is an original diagram drawn by us.

The effects of organic solvents on nucleic acid secondary structure may be more straightforward. The organic solvents most likely melt the secondary structure of the nucleic acid by favorable (attractive) interactions with nucleobases [84–86], although hydrogen bonding between base pairs may be destabilized by organic solvents, as in hydrogen bonding of peptide groups. If melting occurs, then strong electrostatic repulsion between phosphate charges would cause expansion of the nucleic acid structure as shown as a more extended structure in Figure 11. A lower dielectric constant of organic solvents would augment the charge repulsion. Thus, we can derive from the effects of organic solvents that the protein secondary structure is made of hydrophilic (organic solvent repulsive) interactions of peptide groups, while the nucleic acid secondary structure is made of hydrophobic (organic solvent attractive) interactions of nucleobases.

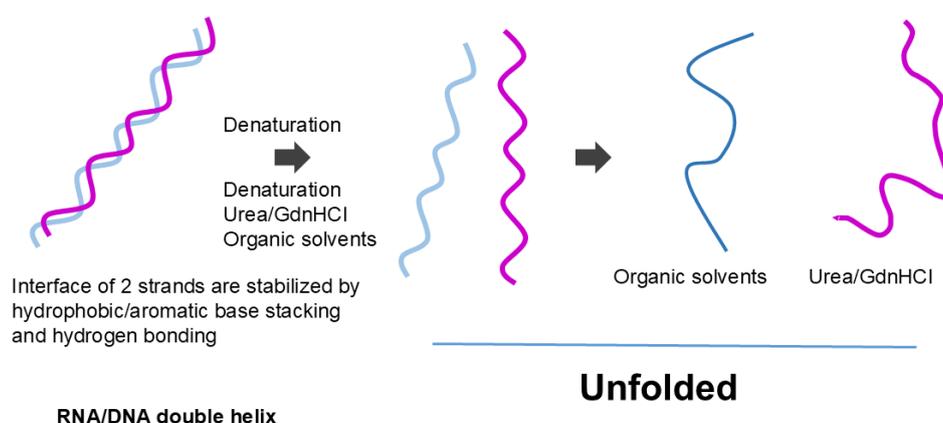


Figure 11. Unfolding of nucleic acid helix. Organic solvents and denaturant destabilize the helix. While organic solvents cause expansion of the dissociated strand, denaturants may allow local base pair formation. This is an original diagram drawn by us.

Regarding manipulation of protein and nucleic acid, differential precipitation is a fundamental process of purification [74,84,87–89]. Ethanol precipitation of protein and nucleic acid is a classical technology for fractionation and concentration. The mechanism of ethanol-induced precipitation is often ascribed to dehydration of bound water by proteins or nucleic acids [90–92]. However, we believe that the repulsive interactions of organic solvents are the main cause of enhanced aggregation/association of proteins and nucleic acids. Repulsive interactions, which are principally energetically (thermodynamically) unfavorable, decrease inversely with the surface area: namely, the surface area decreases from the unfolded to folded and then aggregated proteins (or nucleic acids). This means that organic solvents stabilize more compact state, namely in the order of unfolded < folded < aggregated, provided that only repulsive interaction is involved in the effects of organic solvents. This is of course not correct as described earlier, when organic solvents interact favorably with non-polar/aromatic side chains and nucleobases. This mechanism for organic solvents is true also for any excluded (repulsive) co-solvents, e.g., polymers. Thus, organic solvents (particularly ethanol) have been used to bring precipitation of native proteins and nucleic acids. With nucleic acid precipitation, presence of counter ions to reduce charge repulsion between nucleic acids in the precipitates is essential for organic solvents (that increase electrostatic repulsion due to reduced dielectric constant) to induce nucleic acid precipitation. This is the major difference in organic solvent precipitation between protein and nucleic acid. Due to abundant solvent-exposed charges, precipitation of nucleic acids by organic solvent requires charge neutralization by ions.

3.3.2. Denaturants

The mechanisms by which denaturants such as urea and guanidine hydrochloride (GdnHCl) unfold proteins, and nucleic acids may provide insights into the fundamental forces that drive their folding [93]. The denaturants are used to study the conformational transitions and stability of proteins and nucleic acids [79]. Another important application of denaturants is to solubilize denatured proteins, such as those expressed as inclusion bodies [94–96]. The solubilized proteins are then refolded to generate functional proteins for research and pharmaceutical applications. The expression and refolding technologies played a crucial role in producing proteins and advances in development of research reagents and biopharmaceutical products, which were nearly impossible before invention of recombinant technology.

How do the denaturants interact with proteins and nucleic acids? The interactions of urea and GdnHCl with amino acid side chains and peptide backbone were reported in pioneering studies [97,98]. The glycine solubility either changed little or slightly decreased with increasing concentration of urea and GdnHCl, meaning that the interactions are neutral (no preference for polar glycine or water) or slightly repulsive between glycine and these denaturants and hence are not thermodynamically favorable. Namely, there appeared to be no affinity of the denaturants for dipolar glycine, different markedly from the strong repulsive interactions of organic solvents with glycine. Regarding hydrophilic threonine and glutamine, the interactions are attractive, but marginally small. In contrast, urea and GdnHCl showed strong attractive interactions with aromatic groups, like organic solvents. Urea has been shown to form π - π stacking interactions with aromatic groups of proteins [99,100]. GdnHCl and urea also favorably interact with the peptide groups, showing binding equilibrium with peptide groups in proteins [97,98,101]. The solubility of peptide groups increased with increasing denaturant concentration, clearly demonstrating attractive interactions of peptide bonds with these denaturants, again markedly different from repulsive interactions of organic solvents with peptide groups [97,98]. The attractive interactions of the denaturants with the peptide bonds were ascribed to bivalent hydrogen bonds [97,98].

We are not aware of attractive (or repulsive) interactions of urea and GdnHCl with nucleobases or phosphate backbones in nucleic acids. However, like the interactions of urea with aromatic side chains of proteins, urea can also confer π - π stacking with nucleobases: namely, urea confers attractive interactions with nucleobases [102]. Thus, the denaturants also can have affinity for aromatic nucleobases that are buried in the helical structures of nucleic acids (Figure 4). In addition, those co-solvents that favorably interact with the peptide bonds might interact similarly with the phosphate backbones in nucleic acids. For example, urea and GdnHCl that bind to the peptide bonds by bivalent hydrogen bonds can bind to the phosphate groups similarly [97,103], although these denaturants can interact with side chains of both proteins and nucleic acids. Urea can cause denaturation of nucleic acids by virtue of its binding to nucleobases through hydrogen bonding and aromatic/hydrophobic interactions. These attractive denaturant interactions with peptide groups stabilize the exposed peptide groups, upon unfolding, can lead to extended structures as shown in Figure 10. GdnHCl can stabilize backbone charges of nucleic acids, different from the effects of organic solvents, and hence may not cause fully extended structure of nucleic acid as shown in Figure 11.

3.3.3. Polymers

Polyethylene glycol is perhaps the most frequently used polymer for fractionation and purification of proteins. The principal mechanism of the effects of polymers on protein precipitation and binding to inert surface is steric hindrance, namely exclusion of polymers from the macromolecular surface, called “excluded volume effects” [104–108], “depletion effects” [109], or “macromolecular crowding [110,111]”. No matter how the effects are called, the effects of polymers arise from their inertness to the macromolecular surface and the large molecular size. Because of high aqueous solubility, availability of different sizes and modifications and cost, PEG is widely used for different applications. However, PEG is not necessarily inert, an often-cited misconception of “inertness of PEG”, against proteins and nucleic acids. PEG has been shown to reduce the surface

tension of water, independent of the size (from the molecular weight of 200 to 1000), suggesting that its effects on water structure are determined by the PEG monomeric unit [72]. The degree of surface tension depression by the PEG is greater than typical organic solvents of dimethyl sulfoxide and methyl-pentenediol [72]. Hirano et al [112] studied the interaction of PEG 4000 and 20,000 with amino acids in detail and derived thermodynamic parameters of PEG interaction with the amino acid side chains, leading to a conclusion that “PEG behaves like a weak organic solvent” as indicated in the paper title and thus a conclusion that PEG is not definitely inert. Interestingly, PEG 20,000 resulted in phase separation of glycine at the saturation concentration of glycine, one phase containing only glycine (glycine-rich phase) and another phase containing both PEG 20,000 and glycine (PEG-rich phase). Namely, PEG is enriched, and glycine is depleted in PEG-rich phase, while in glycine-rich phase, PEG is depleted. To summarize this observation, glycine and PEG 20,000 repel each other. Whether this is due to excluded volume effects of PEG 20,000 or other mechanism cannot be answered from this experiment only. However, it has been shown that organic solvents, including ethanol [70], dioxane [70], ethylene glycol [73,112] and DMSO [74] decrease the solubility of glycine, indicating unfavorable interaction of glycine with these organic solvents. Thus, it is possible that PEG is acting on glycine as an organic solvent. In fact, Hirano et al [112] showed that PEG 400 and PEG 4000 equally decrease the glycine solubility, even more effectively than ethylene glycol on the weight concentration basis. The interaction energy of amino acid side chains with PEG was calculated from the solubility difference between glycine and amino acids, which requires the solubility data of glycine. Because of phase separation of aqueous glycine solution into PEG-rich and glycine-rich phases with PEG 20,000, the interaction energy of side chains could not be obtained for PEG 20,000 (which requires the solubility data of glycine in PEG 20,000). Nevertheless, PEG 20,000 as well as PEG 4000 showed highly favorable interaction with tryptophan amino acid. This interaction has a potential contribution of unfavorable (repulsive) interaction of terminal amino and carboxyl groups of tryptophan (blue circle), depicted in Figure 12, as speculated from the repulsive interaction of PEG 20,000 and glycine. The favorable (attractive) interaction between PEG and indole structure (red circle) overwhelms the unfavorable contribution of overall excluded volume effect and charge-PEG interactions (blue circle). In fact, when the unfavorable contribution of the terminal interactions, not only tryptophan but also phenylalanine, leucine and isoleucine showed favorable interaction with PEG 4000: this calculation was possible for PEG 4000, which caused no phase separation of glycine and allowed determination of glycine solubility measurement. This observation that PEG 4000 caused no phase separation is interesting. If organic solvent-like behavior is a sole factor, this PEG may be expected to cause phase separation. Thus, the reason for phase separation of PEG 20,000 may be due to both organic solvent-like property and excluded volume effect, the former being identical for PEG 4000 and 20,000 and the latter differing between these two PEGs.

This organic solvent-like property and excluded volume effect operate on the effects of PEG on protein and nucleic acid stability. PEG has been shown to decrease the stability of proteins to a very limited extent, which was explained by a balance between excluded volume effect and PEG binding [106]. Such PEG binding clearly indicates that hydrophobic groups are buried inside the protein molecule and stabilized by hydrophobic PEG upon protein unfolding by heating. PEG, particularly small PEGs, resulted in significant reduction of melting of DNA [114], which can be explained by hydrophobic binding of PEG to the melted DNA and in turn means that PEG binding requires exposure of non-polar nucleobases. On the contrary, PEG binding to protein occurs to the hydrophobic side chains buried in the tertiary structure folding, not requiring melting of helical structure of the protein. A large PEG 8000 slightly increased melting temperature (stabilized) of 30-mer DNA duplex [113], indicating that a large size PEG resulted in overwhelming excluded volume effects on a large DNA duplex and meaning “inertness” of such a large PEG on the large DNA.

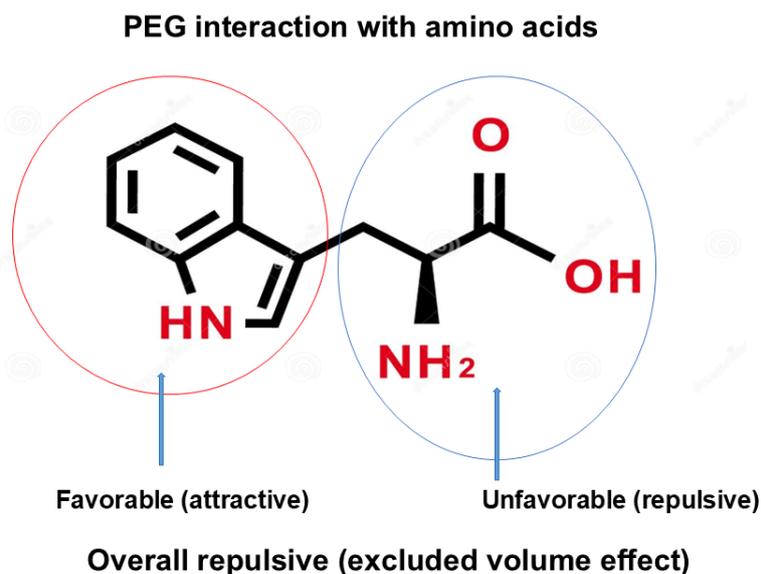


Figure 12. Schematic diagram of polymer interaction with tryptophan. Mode of PEG interaction with Trp terminal groups (blue circle) and side chain (red circle). This is an original diagram drawn by us.

A similar effect of PEG was observed with quadruplex DNA helix. As depicted in Figure 13, there is small stabilization of quadruplex structure by the PEG [113–115]. For the quadruplex helix, the excluded volume effect of PEG overwhelmed the aromatic-hydrophobic interactions between PEG and nucleobases, leading to stabilization of the quadruplex structure. The excluded volume effect of PEG on the quadruplex stabilization is schematically depicted in Figure 13, where the quadruplex structure is shown as a box (left) and its unfolded structure is shown as a rectangular box (right). PEG depicted as a ellipsoid is excluded at a greater extent for the unfolded DNA with a larger surface area than the folded quadruplex with reduced surface area, which would stabilize the folded structure. Alternatively, dehydration mechanism was proposed as a major mechanism of stabilization of the quadruplex structure. This mechanism is proposed in a classical work of Kauzmann as hydrophobic hydration of non-polar surface, exposure of which is unstable in aqueous environments and reduced upon association of the non-polar surface [116]. This mechanism is equivalent to the excluded volume effects, which lead to loss of excess water, upon the DNA folding, which arises from exclusion of polymeric PEG. However, it should be noted that hydrophobic PEG may bind to the unfolded DNA by hydrophobic interaction between the exposed nucleobases and hydrophobic moiety of the PEG.

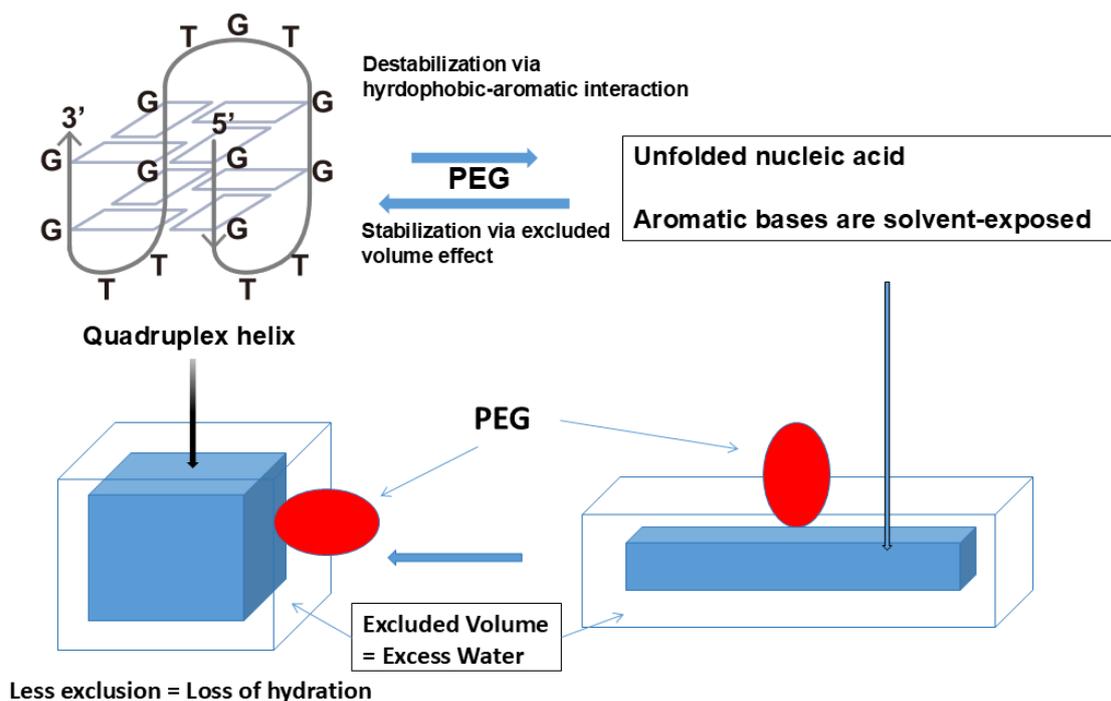


Figure 13. Schematic diagram of PEG interaction with quadruplex. While aromatic/hydrophobic interaction of PEG with nucleobases destabilize the helix, excluded volume effects stabilize the helix. Hydration of nucleic acid is due to excluded volume effects and decreases with formation of quadruplex structure from single strand structure. This is an original diagram drawn by us.

4. Conclusion

We here described similarities and differences in structure and interactions between proteins and nucleic acids. Their similarity in molecular architecture, i.e., backbone and side chains, leads to one common structure, helix. Among the side chains, both protein and nucleic acid have aromatic groups, which play a central role in interactions. A major difference is their location. While aromatic side chains of protein are excluded from participating in secondary structures, nucleobases participate in helix formation in nucleic acid. Thus, the side chains in proteins are exposed to confer molecular interactions for folding or protein-protein and protein-nucleic acid interaction, while those of nucleic acids require their exposure by strand breaks, loops, hairpins or kinks. This difference may be simply due to their physical location or different chemical properties. Both aromatic groups are similar in terms of hydrophobic properties in the sense that they interact favorably (attractively) with organic solvents and in terms of aromatic nature in the sense that they participate in π - π interaction and cation (or anion)- π interaction. Particularly, they show favorable interaction with arginine side chain. These interactions are thermodynamic parameters associated with the free energy changes of interactions and do not give any information on the physical interaction mechanism, e.g., the mechanism of aromatic-organic solvent interaction or aromatic-arginine interaction. In addition, we do not know the physical properties of aromatic-aromatic interaction between the side chains, which competes with aromatic-solvent interaction in thermodynamic measurement. It would of great interest to characterize this physical mechanism of aromatic-aromatic interaction and aromatic-solvent interaction side-by-side for nucleobases and amino acid side chains. Difference, if any, may give insight into why they play different roles in participating in secondary structure formation. Based on the knowledge obtained so far from co-solvent interactions, the aromatic properties of nucleic acid and protein side chains are similar, while their location in protein and nucleic acid structures is the main difference.

The structure of nucleopeptide may also give insight into the aromatic properties of nucleobases. When nucleobases are inserted into the peptide backbone, the structure does not appear to adopt a

protein-type secondary structure. Instead, it forms base pairs with a complementary single strand nucleic acid, which may be due to the nature of the nucleobases. A question is whether a polypeptide comprising only aromatic amino acid side chains, e.g., polyphenylalanine or polytyrosine, would adopt a protein-type secondary structure or instead form aromatic stacking interactions that generate a different secondary structure.

Co-solvent interactions were used to draw understanding the properties of macromolecular interactions. It should be noted that co-solvents are used to stabilize proteins and nucleic acids or suppress their aggregations in pharmaceutical formulation development. Sucrose and trehalose are frequently used to enhance protein stability. Arginine has been shown to suppress protein aggregation during storage. Detergents are used to suppress surface-mediated aggregation. Artificial intelligence (AI) may be able to assist an optimal co-solvent to be used based on knowledge of the mechanisms of protein or nucleic acid aggregation and co-solvent interactions with these macromolecules.

Author Contributions: Tsutomu Arakawa, TO and Teruo Akuta contributed to the acquisition of literature and creation of Figures for the work and wrote the paper. TU, SN and KS approved the final manuscript. These authors have participated extensively in the work and have agreed to be accountable for all aspects of the work.

Funding: This review article received no external funding.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: I gratefully acknowledge the support provided to conduct the review.

Conflicts of Interest: Tsutomu Arakawa was previously affiliated with the for-profit company Alliance Protein Laboratories. T.O. is an employee of the for-profit company JASCO Corporation. Teruo Akuta is an employee of the for-profit company Kyokuto Pharmaceutical Industrial Co., Ltd. T. U. and K.S. are staff members at University of Tsukuba and S.N. is a postdoctoral researcher at Hokkaido University. These authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

| | |
|--------|-------------------------|
| CD | circular dichroism |
| SF1 | splicing factor 1 |
| GdnHCl | guanidine hydrochloride |
| PEG | polyethylene glycol |

References

1. Agyei, D.; Ahmed, I.; Akram, Z.; Iqbal, H.M.N.; Danquah, M.K. Protein and peptide biopharmaceuticals: an overview. *Protein Pept. Lett.* **2017**, *24*, 94–101. doi:10.2174/0929866523666161222150444.
2. Kesik-Brodacka, M. Progress in biopharmaceutical development. *Biotechnol. Appl. Biochem.* **2018**, *65*, 306–322. doi:10.1002/bab.1617.
3. Mulligan, M.J.; Lyke, K.E.; Kitchin, N.; Absalon, J.; Gurtman, A.; Lockhart, S.; Neuzil, K.; Raabe, V.; Bailey, R.; Swanson, K.A. et al. Phase I/II study of COVID-19 RNA vaccine BNT162b1 in adults. *Nature* **2020**, *586*, 589–593. doi:10.1038/s41586-020-2639-4.
4. Zhu, G.; Chen, X. Aptamer-based targeted therapy. *Adv. Drug Deliv. Rev.* **2018**, *134*, 65–78. doi:10.1016/j.addr.2018.08.005.
5. Crooke, S.T.; Liang, X.H.; Baker, B.F.; Crooke, R.M. Antisense technology: A review. *J. Biol. Chem.* **2021**, *296*, 100416. doi:10.1016/j.jbc.2021.100416.

6. Rosa, S.S.; Prazeres, D.M.F.; Azevedo, A.M.; Marques, M.P.C. mRNA vaccines manufacturing: Challenges and bottlenecks. *Vaccine* **2021**, *39*, 2190–2200. doi:10.1016/j.vaccine.2021.03.038.
7. Kis, Z.; Kontoravdi, C.; Shattock, R.; Shah, N. Resources, production scales and time required for producing RNA vaccines for the global pandemic demand. *Vaccines (Basel)* **2021**, *9*, 3. doi:10.3390/vaccines9010003.
8. Sanyal, G.; Särnefält, A.; Kumar, A. Considerations for bioanalytical characterization and batch release of COVID-19 vaccines. *NPJ Vaccines* **2021**, *6*, 1–9. doi:10.1038/s41541-021-00317-4.
9. Lucas, X.; Bauzá, A.; Frontera, A.; Quiñonero, D. A thorough anion- π interaction study in biomolecules: on the importance of cooperativity effects. *Chem. Sci.* **2016**, *7*, 1038–1050. doi:10.1039/c5sc01386k.
10. Dougherty, D.A. The cation- π interaction. *Acc. Chem. Res.* **2013**, *46*, 885–893. doi:10.1021/ar300265y.
11. Cole, C.C.; Misiura, M.; Hulgan, S.A.H.; Peterson, C.M.; Williams, J.W., 3rd; Kolomeisky, A.B. Cation- π interactions and their role in assembling collagen triple helices. *Biomacromolecules* **2022**, *23*, 4645–4654. doi:10.1021/acs.biomac.2c00856.
12. Gallivan, J.P.; Dougherty, D.A. Cation- π interactions in structural biology. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9459–9464. doi:10.1073/pnas.96.17.9459.
13. Ferrari, L.; Stucchi, R.; Konstantoulea, K.; van de Kamp, G.; Kos, R.; Geerts, W.; van der Geer, R.M.J.; Schreurs, M.A.H.; van Dam, B.H.C.; Schutten, L.H.E. et al. Arginine π -stacking drives binding to fibrils of the Alzheimer protein Tau. *Nat. Commun.* **2020**, *11*, 571. doi:10.1038/s41467-019-13745-7.
14. Carter-Fenk, K.; Liu, M.; Pujal, L.; Loipersberger, M.; Tsanai, M.; Vernon, R.M.; Johnson, E.R.; Sherrill, C.D. The energetic origins of P- π contacts in proteins. *J. Am. Chem. Soc.* **2023**, *145*, 24836–24851. doi:10.1021/jacs.3c09198.
15. Cumpstey, I.; Salomonsson, E.; Sundin, A.; Leffler, H.; Nilsson, U.J. Studies of arginine-arene interactions through synthesis and evaluation of a series of galectin-binding aromatic lactose esters. *ChemBioChem* **2007**, *8*, 1389–1398. doi:10.1002/cbic.200700040.
16. Herskovits, T.T.; Bowen, J.J. Solution studies of the nucleic acid bases and related compounds. Solubility in aqueous urea and amide solutions. *Biochemistry* **1974**, *13*, 5474–5483. doi:10.1021/bi00724a004.
17. Herskovits, T.T.; Harrington, J.P. Solution studies of the nucleic acid bases and related model compounds. Solubility in aqueous alcohol and glycol solutions. *Biochemistry* **1972**, *11*, 4800–4811. doi:10.1021/bi00775a025.
18. Hirano, A.; Tokunaga, H.; Tokunaga, M.; Arakawa, T.; Shiraki, K. The solubility of nucleobases in aqueous arginine solutions. *Arch. Biochem. Biophys.* **2010**, *497*, 90–96. doi:10.1016/j.abb.2010.03.009.
19. Dufour, E.; Haertlé, T. Alcohol-induced changes of β -lactoglobulin-retinol-binding stoichiometry. *Protein Eng.* **1990**, *4*, 185–190. doi:10.1093/protein/4.2.185.
20. Shiraki, K.; Nishikawa, K.; Goto, Y. Trifluoroethanol-induced stabilization of the α -helical structure of β -lactoglobulin: implication for non-hierarchical protein folding. *J. Mol. Biol.* **1995**, *245*, 180–194. doi:10.1006/jmbi.1994.0015.
21. Ferreon, A.C.M.; Deniz, A.A. Alpha-synuclein multistate folding thermodynamics: implications for protein misfolding and aggregation. *Biochemistry* **2007**, *46*, 4499–4509. doi:10.1021/bi602461y.
22. Cao, Y.; Wang, D.; Zhou, P.; Zhao, Y.; Sun, Y.; Wang, J. Influence of conventional surfactants on the self-assembly of a bola-type amphiphilic peptide. *Langmuir* **2017**, *33*, 5446–5455. doi:10.1021/acs.langmuir.7b00405.
23. Arakawa, T.; Tokunaga, M.; Kita, Y.; Niikura, T.; Baker, R.W.; Reimer, J.M.; Vaughn, D.F.; Roberts, C.J.; Joshi, P.K. Structure analysis of proteins and peptides by difference circular dichroism spectroscopy. *Protein J.* **2021**, *40*, 867–875. doi:10.1007/s10930-021-10024-7.
24. Ahari, D.; Sahil, K.; Kaushal, S.; Sharma, A.; Rangan, L.; Swaminathan, R. Structural transitions of dehydrin in response to temperature, the presence of trifluoroethanol and sodium dodecyl sulfate, and its protective role in heat and cold stress. *Biochemistry* **2025**, *64*, 3045–3062. doi:10.1021/acs.biochem.5c00222.
25. Yang, J.F.; Wang, F.; Wang, M.Y.; Wang, D.; Zhou, Z.S.; Hao, G.F.; Liu, Y.; Chen, X.; Zhao, Q. CIPDB: A biological structure databank for studying cation and π interactions. *Drug Discov. Today* **2023**, *28*, 103546. doi:10.1016/j.drudis.2023.103546.
26. Chiang, C.H.; Horng, J.C. Cation- π interaction induced folding of AAB-type collagen heterotrimers. *J. Phys. Chem. B* **2016**, *120*, 1205–1211. doi:10.1021/acs.jpcc.5b11189.

27. Schiffer, M.; Edmundson, A.B. Use of helical wheels to represent the structures of proteins and to identify segments with helical potential. *Biophys. J.* **1967**, *7*, 121–135. doi:10.1016/S0006-3495(67)86579-2.
28. Brindley, M.A.; Suter, R.; Schestak, I.; Kiss, G.; Wright, E.R.; Plemper, R.K. A stabilized headless measles virus attachment protein stalk efficiently triggers membrane fusion. *J. Virol.* **2013**, *87*, 11693–11703. doi:10.1128/JVI.01945-13.
29. Phuong, H.B.T.; Phuong, N.T.; Bui, L.M.; Thi, H.P.; Tran, T.T.P.; Quoc, T.N.; Nguyen, L.T.; Le, P.; Nguyen, D.T. Optimizing amphipathic antimicrobial peptides via helical wheel rotation. *ChemMedChem* **2025**, *e202500316*. doi:10.1002/cmcd.202500316.
30. López, J.C.; Sánchez, R.; Blanco, S.; Alonso, J.L. Microsolvation of 2-azetidione: a model for the peptide group–water interactions. *Phys. Chem. Chem. Phys.* **2015**, *17*, 2054–2066. doi:10.1039/c4cp04577g.
31. Scoppola, E.; Sodo, A.; McLain, S.E.; Ricci, M.A.; Bruni, L. Water–peptide site-specific interactions: a structural study on the hydration of glutathione. *Biophys. J.* **2014**, *106*, 1701–1709. doi:10.1016/j.bpj.2014.01.046.
32. Scognamiglio, P.L.; Platella, C.; Napolitano, E.; Musumeci, D.; Roviello, G.N. From prebiotic chemistry to supramolecular biomedical materials: exploring the properties of self-assembling nucleobase-containing peptides. *Molecules* **2021**, *26*, 3558. doi:10.3390/molecules26123558.
33. Lund, M.; Jönsson, B.; Woodward, C.E. Implications of a high dielectric constant in proteins. *J. Chem. Phys.* **2007**, *126*, 225103. doi:10.1063/1.2741543.
34. Kimple, A.J.; Soundararajan, M.; Hutsell, S.Q.; Roos, A.K.; Urban, D.J.; Setola, V.; Sucic, M.B.; Tesmer, K.K. Structural determinants of G-protein alpha subunit selectivity by regulator of G-protein signaling 2 (RGS2). *J. Biol. Chem.* **2009**, *284*, 19402–19411. doi:10.1074/jbc.M109.024711.
35. Lee, H.T.; Lee, J.Y.; Lim, H.; Lee, S.H.; Moon, Y.J.; Pyo, H.J.; Kim, K.S.; Kim, D.H. Molecular mechanism of PD-1/PD-L1 blockade via anti-PD-L1 antibodies atezolizumab and durvalumab. *Sci. Rep.* **2017**, *7*, 5532. doi:10.1038/s41598-017-06002-8.
36. Tsumoto, K.; Ogasahara, K.; Ueda, Y.; Watanabe, K.; Yutani, K.; Kumagai, I. Role of Tyr residues in the contact region of anti-lysozyme monoclonal antibody HyHEL-10 for antigen binding. *J. Biol. Chem.* **1995**, *270*, 18551–18557. doi:10.1074/jbc.270.31.18551.
37. Tsumoto, K.; Yokota, A.; Tanaka, Y.; Ui, M.; Tsumuraya, T.; Fujii, I.; Kato, M.; Kobayashi, S. Critical contribution of aromatic rings to specific recognition of polyether rings: the case of ciguatoxin CTX3C-ABC and its specific antibody 1C49. *J. Biol. Chem.* **2008**, *283*, 12259–12266. doi:10.1074/jbc.M710553200.
38. Akiba, H.; Tsumoto, K. Thermodynamics of antibody–antigen interaction revealed by mutation analysis of antibody variable regions. *J. Biochem.* **2015**, *158*, 1–11. doi:10.1093/jb/mvv049.
39. Liu, Z.; Luyten, I.; Bottomley, M.J.; Messias, A.C.; Houngninou-Molango, S.; Sprangers, R.; Lam, R.F.; de Almeida, M.S.; Abrahams, J.P.; Steinmetz, M.O. Structural basis for recognition of the intron branch site RNA by splicing factor 1. *Science* **2001**, *294*, 1098–1102. doi:10.1126/science.1064719.
40. Matthews, M.M.; Thomas, J.M.; Zheng, Y.; Tran, K.; Phelps, K.J.; Scott, A.I.; Lam, R.F.; Roberts, C.; Zimmerman, M.I.; Vithani, N.; Griffith, D.; Wagoner, J.A.; Bowman, G.R.; Hall, K.B.; Soranno, A.; Holehouse, A.S. Structures of human ADAR2 bound to dsRNA reveal base-flipping mechanism and basis for site selectivity. *Nat. Struct. Mol. Biol.* **2016**, *23*, 426–433. doi:10.1038/nsmb.3203.
41. Hama, T.; Ferré-D’Amaré, A.R. Structure of protein L7Ae bound to a K-turn derived from an archaeal box H/ACA sRNA at 1.8 Å resolution. *Structure* **2004**, *12*, 893–903. doi:10.1016/j.str.2004.03.015.
42. Moore, T.; Zhang, Y.; Fenley, M.O.; Li, H. Molecular basis of box C/D RNA–protein interactions: cocrystal structure of archaeal L7Ae and a box C/D RNA. *Structure* **2004**, *12*, 807–818. doi:10.1016/j.str.2004.02.033.
43. Meier-Stephenson, V. G4-quadruplex-binding proteins: review and insights into selectivity. *Biophys. Rev.* **2022**, *14*, 635–654. doi:10.1007/s12551-022-00952-8.
44. Chen, M.C.; Tippana, R.; Demeshkina, N.A.; Murat, P.; Balasubramanian, S.; Myong, S.; Yu, H.; Gross, J.T.; Zhao, K. Structural basis of G-quadruplex unfolding by the DEAH/RHA helicase DHX36. *Nature* **2018**, *558*, 465–469. doi:10.1038/s41586-018-0209-9.
45. Horvath, M.P.; Schultz, S.C. DNA G-quartets in a 1.86 Å resolution structure of an *Oxytricha nova* telomeric protein–DNA complex. *J. Mol. Biol.* **2001**, *310*, 367–377. doi:10.1006/jmbi.2001.4766.

46. Ladame, S.; Schouten, J.A.; Roldan, J.; Redman, J.E.; Neidle, S.; Balasubramanian, S. Exploring the recognition of quadruplex DNA by an engineered Cys₂-His₂ zinc finger protein. *Biochemistry* **2006**, *45*, 1393–1399. doi:10.1021/bi050229x.
47. Chavali, S.S.; Cavender, C.E.; Mathews, D.H.; Wedekind, J.E. Arginine forks are a widespread motif to recognize phosphate backbones and guanine nucleobases in the RNA major groove. *J. Am. Chem. Soc.* **2020**, *142*, 19835–19839. doi:10.1021/jacs.0c09689.
48. MacDougall, G.; Anderton, R.S.; Trimble, A.; Mastaglia, F.L.; Knuckey, N.W.; Meloni, B.P. Poly-arginine-18 (R18) confers neuroprotection through glutamate receptor modulation, intracellular calcium reduction, and preservation of mitochondrial function. *Molecules* **2020**, *25*, 2977. doi:10.3390/molecules25132977.
49. Rodriguez-Casado, A.; Molina, M.; Carmona, P. Core protein–nucleic acid interactions in hepatitis C virus as revealed by Raman and circular dichroism spectroscopy. *Appl. Spectrosc.* **2007**, *61*, 1219–1224. doi:10.1366/000370207782597139.
50. Ham, S.; Lee, C. Exploring the impact of nucleic acids on protein stability in bacterial cell lysate. *Biochim. Biophys. Acta* **2023**, *1867*, 130445. doi:10.1016/j.bbagen.2023.130445.
51. Cubuk, J.; Alston, J.J.; Incicco, J.J.; Singh, S.; Struchell-Brereton, M.D.; Ward, M.D.; Zimmerman, M.I.; Vithani, N.; Griffith, D.; Wagoner, J.; Bowman, G.R.; Hall, K.B.; Soranno, A.; Holehouse, A.S. The SARS-CoV-2 nucleocapsid protein is dynamic, disordered, and phase separates with RNA. *Nat. Commun.* **2021**, *12*, 1936. doi:10.1038/s41467-021-21953-3.
52. Vos, J.D.; Aguilar, P.P.; Köppl, C.; Fischer, A.; Grünwald-Gruber, C.; Dürkop, M.; Becker, S.; Fuchs, R.; Müller, S.; Schneider, T.M. Production of full-length SARS-CoV-2 nucleocapsid protein from *Escherichia coli* optimized by native hydrophobic interaction chromatography hyphenated to multi-angle scattering detection. *Talanta* **2021**, *235*, 122691. doi:10.1016/j.talanta.2021.122691.
53. Carlson, C.R.; Asfaha, J.B.; Ghent, C.M.; Howard, C.J.; Hartooni, N.; Safari, M.; Gerber, R.A.; Gardner, D.A.; Williamson, E.M. Phosphoregulation of phase separation by the SARS-CoV-2 N protein suggests a biophysical basis for its dual functions. *Mol. Cell* **2020**, *80*, 1092–1103. doi:10.1016/j.molcel.2020.11.025.
54. Lu, S.; Ye, Q.; Singh, D.; Cao, Y.; Diedrich, J.K.; Yates, J.R., 3rd; Guttman, S.H.; Smith, R.R. The SARS-CoV-2 nucleocapsid phosphoprotein forms mutually exclusive condensates with RNA and the membrane-associated M protein. *Nat. Commun.* **2021**, *12*, 502. doi:10.1038/s41467-020-20768-y.
55. Laue, T.M.; Shire, S.J. The molecular interaction process. *J. Pharm. Sci.* **2019**, *307*, 1. doi:10.1016/j.xphs.2019.10.045.
56. Banani, S.F.; Lee, H.O.; Hyman, A.A.; Rosen, M.K. Biomolecular condensates: Organizers of cellular biochemistry. *Nat. Rev. Mol. Cell Biol.* **2017**, *18*, 285–298. doi:10.1038/nrm.2017.7.
57. Garabedian, M.V.; Su, Z.; Dabdoub, J.; Tong, M.; Deiters, A.; Hammer, D.A.; Good, M.C. Protein condensate formation via controlled multimerization of intrinsically disordered sequences. *Biochemistry* **2022**, *61*, 2237–2246. doi:10.1021/acs.biochem.2c00250.
58. Martin, E.W.; Holehouse, A.S.; Peran, I.; Farag, M.; Incicco, J.J.; Bremer, A.; Grace, C.R.; Soranno, A.; Pappu, R.V.; Mittag, T. Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science* **2020**, *367*, 694–699. doi:10.1126/science.aaw8653.
59. Viereggs, J.R.; Lueckheide, M.; Marciel, A.; Leon, D.M.; Tirrell, K.L. Oligonucleotide–peptide complexes: Phase control by hybridization. *J. Am. Chem. Soc.* **2018**, *140*, 1632–1638. doi:10.1021/jacs.7b03567.
60. Wadsworth, G.; Srinivasan, S.; Lai, L.; Datta, M.; Gopalan, V.; Banerjee, P.R. RNA-driven phase transitions in biomolecular condensates. *Mol. Cell* **2024**, *85*, 3670–3682. doi:10.1016/j.molcel.2024.09.005.
61. Zhang, Y.; You, R.; Zhao, W.; Zhou, J.; Chen, J. G-quadruplex structures trigger RNA phase separation. *Nucleic Acids Res.* **2019**, *47*, 11746–11754. doi:10.1093/nar/gkz978.
62. Tauber, D.; Tauber, G.; Parker, R. Mechanisms and regulation of RNA condensation in RNP granule formation. *Trends Biochem. Sci.* **2020**, *45*, 764–778. doi:10.1016/j.tibs.2020.05.002.
63. Maity, H.; Nguyen, H.; Hori, N.; Thirumalai, D. Odd–even disparity in the population of slipped hairpins in RNA repeat sequences with implications for phase separation. *Proc. Natl. Acad. Sci. U.S.A.* **2023**, *120*, e2301409120. doi:10.1073/pnas.2301409120.
64. Shil, S.; Tsuruta, M.; Kawauchi, K.; Miyoshi, D. Factors Affecting Liquid-Liquid Phase Separation of RGG Peptides with DNA G-Quadruplex. *ChemMedChem.* **2025**, *20*, e202400460. doi: 10.1002/cmdc.202400460

65. Tsuruta, M.; Shil, S.; Taniguchi, S.; Kawauchi, K.; Miyoshi, D. The role of cytosine methylation in regulating the topology and liquid-liquid phase separation of DNA G-quadruplexes. *Chem Sci.* **2025**, *16*, 4213-4225. doi: 10.1039/d4sc06959e.
66. Amato, J.; Madanayake, T.W.; Iaccarino, N.; Novellino, E.; Randazzo, A.; Hurley, L.H.; Pagano, B. HMGB1 binds to the KRAS promoter G-quadruplex: a new player in oncogene transcriptional regulation? *Chem Commun (Camb).* **2018**, *54*, 9442-9445. doi: 10.1039/c8cc03614d.
67. Chen, L.; Dickerhoff, J.; Zheng, K.W.; Erramilli, S.; Feng, H.; Wu, G.; Onel, B.; Chen, Y.; Wang, K.B.; Carver, M. et al. Structural basis for nucleolin recognition of MYC promoter G-quadruplex. *Science.* **2025**, *388*, eadr1752. doi: 10.1126/science.adr1752.
68. Sahoo, B.R.; Deng, X.; Wong, E.L.; Clark, N.; Yang, H.J.; Kovalenko, A.; Subramanian, V.; Guzman, B.B.; Harris, S.E.; Dehury, B. et al. Visualization of liquid-liquid phase transitions using a tiny G-quadruplex binding protein. *Nat Commun.* **2025**, *16*, 8578. doi: 10.1038/s41467-025-63597-7.
69. Yoo, W.; Song, Y.W.; Bansal, V.; Kim, K.K. G-quadruplex-dependent transcriptional regulation by molecular condensation in the Bcl3 promoter. *Nucleic Acids Res.* **2025**, *53*, gkaf827. doi: 10.1093/nar/gkaf827.
70. Nozaki, Y.; Tanford, C. The solubility of amino acids and two glycine peptides in aqueous ethanol and dioxane solutions. Establishment of a hydrophobicity scale. *J. Biol. Chem.* **1971**, *246*, 2211-2217. doi:10.1016/S0021-9258(19)77210-X.
71. Trikulenko, A.V. Thermodynamic and stereochemical parameters to evaluate stability of hydrophobic cores of globular proteins. *Biochemistry (Mosc.)* **1998**, *63*, 1290-1293. <https://pubmed.ncbi.nlm.nih.gov/9864468/>.
72. Kita, Y.; Arakawa, T.; Lin, T.Y.; Timasheff, S.N. Contribution of the surface free energy perturbation to protein-solvent interactions. *Biochemistry* **1994**, *33*, 15178-15189. doi:10.1021/bi00254a029.
73. Nozaki, Y.; Tanford, C. The solubility of amino acids and related compounds in aqueous ethylene glycol solutions. *J. Biol. Chem.* **1965**, *240*, 3568-3575. doi:10.1016/S0021-9258(18)97181-4.
74. Arakawa, T.; Kita, Y.; Timasheff, S.N. Protein precipitation and denaturation by dimethyl sulfoxide. *Biophys. Chem.* **2007**, *131*, 62-70. doi:10.1016/j.bpc.2007.09.004.
75. Pitts, E.P.; Timasheff, S.N. Interaction of ribonuclease A with aqueous 2-methyl-2,4-pentanediol at pH 5.8. *Biochemistry* **1978**, *17*, 615-623. doi:10.1021/bi00597a009.
76. Xu, R.; Zheng, M.; Farajtabar, A.; Zhao, H. Solubility modelling and preferential solvation of adenine in solvent mixtures of N,N-dimethylformamide, N-methyl pyrrolidone, propylene glycol and dimethyl sulfoxide plus water. *J. Chem. Thermodyn.* **2018**, *125*, 225-234. doi:10.1016/j.jct.2018.05.030.
77. Zhou, Y.; Han, D.; Tao, T.; Zhang, S.; Wang, J.; Gong, J.; Li, L.; Sun, H. Solubility measurement, thermodynamic correlation and molecular simulations of uracil in alcohol + water binary solvents at 233.15-318.15 K. *J. Mol. Liq.* **2020**, *318*, 114259. doi:10.1016/j.molliq.2020.114259.
78. Herskovits, T.T. Conformation of proteins and polypeptides. I. Extension of the solvent perturbation technique of difference spectroscopy to the study of proteins and polypeptides in organic solvents. *J. Biol. Chem.* **1965**, *240*, 628-638. [https://www.jbc.org/article/S0021-9258\(17\)45221-5/pdf](https://www.jbc.org/article/S0021-9258(17)45221-5/pdf).
79. Alonso, D.O.; Dill, K.A. Solvent denaturation and stabilization of globular proteins. *Biochemistry* **1991**, *30*, 5947-5985. doi:10.1021/bi00238a023.
80. Khan, J.M.; Malik, A.; Sharma, P.; Fatima, S. Anionic surfactants cause dual conformational changes in insulin. *Int. J. Biol. Macromol.* **2023**, *247*, 125790. doi:10.1016/j.ijbiomac.2023.125790.
81. Luo, P.; Baldwin, R.L. Mechanism of helix induction by trifluoroethanol: a framework for extrapolating the helix-forming properties of peptides from trifluoroethanol/water mixtures back to water. *Biochemistry* **1997**, *36*, 8413-8421. doi:10.1021/bi9707133.
82. Kumari, N.K.P.; Jagannadham, M.V. Deciphering the molecular structure of cryptolepain in organic solvents. *Biochimie* **2011**, *94*, 310-317. doi:10.1016/j.biochi.2011.07.017.
83. Srisailam, S.; Kumar, T.K.; Simathi, T.; Yu, C. Influence of backbone conformation on protein aggregation. *J. Am. Chem. Soc.* **2002**, *124*, 1884-1888. doi:10.1021/ja012070r.
84. Sinanoglu, O.; Abdunur, S. Effect of water and other solvents on the structure of biopolymers. *Fed. Proc.* **1965**, *24*, S12-S23. <https://documentsdelivered.com/source/032/901/032901967.php>.

85. Geiduschek, E.P.; Herskovits, T.T. Nonaqueous solutions of DNA. Reversible and irreversible denaturation in methanol. *Arch. Biochem. Biophys.* **1961**, *95*, 114–129. doi:10.1016/0003-9861(61)90116-3.
86. Herskovits, T.T.; Singer, S.J.; Geiduschek, E.P. Nonaqueous solutions of DNA. Denaturation in methanol and ethanol. *Arch. Biochem. Biophys.* **1961**, *94*, 99–114. doi:10.1016/0003-9861(61)90016-9.
87. Yoshizawa, S.; Arakawa, T.; Shiraki, K. Dependence of ethanol effects on protein charges. *Int. J. Biol. Macromol.* **2014**, *68*, 169–172. doi:10.1016/j.ijbiomac.2014.04.041.
88. Yoshikawa, H.; Hirano, A.; Arakawa, T.; Shiraki, K. Mechanistic insights into protein precipitation by alcohol. *Int. J. Biol. Macromol.* **2012**, *50*, 865–871. doi:10.1016/j.ijbiomac.2011.11.005.
89. Yoshikawa, H.; Hirano, A.; Arakawa, T.; Shiraki, K. Effects of alcohol on the solubility and structure of native and disulfide-modified bovine serum albumin. *Int. J. Biol. Macromol.* **2012**, *50*, 1286–1291. doi:10.1016/j.ijbiomac.2012.03.014.
90. van Oss, C.J. On the mechanism of the cold ethanol precipitation method of plasma protein fractionation. *J. Protein Chem.* **1989**, *8*, 661–668. doi:10.1007/BF01025606.
91. Klemm, W.R. Dehydration: a new alcohol theory. *Alcohol* **1990**, *7*, 49–59. doi:10.1016/0741-8329(90)90060-p.
92. Peng, L.; Dai, H.; Wang, H.; Zhu, H.; Ma, L.; Yu, Y.; Zhang, X.; Li, M. Effect of different dehydration methods on the properties of gelatin films. *Food Chem.* **2022**, *374*, 131814. doi:10.1016/j.foodchem.2021.131814.
93. Herskovits, T.T. Nonaqueous solutions of DNA; denaturation by urea and its methyl derivatives. *Biochemistry* **1963**, *2*, 335–340. doi:10.1021/bi00902a027.
94. Misawa, S.; Kumagai, I. Refolding of therapeutic proteins produced in *Escherichia coli* as inclusion bodies. *Biopolymers* **1999**, *51*, 297–307. doi:10.1002/(SICI)1097-0282(1999)51:4<297::AID-BIP5>3.0.CO;2-I.
95. Singh, S.M.; Panda, A.K. Solubilization and refolding of bacterial inclusion body proteins. *J. Biosci. Bioeng.* **2005**, *99*, 303–310. doi:10.1263/jbb.99.303.
96. Upadhyay, V.; Singh, A.; Panda, A.K. Purification of recombinant ovalbumin from inclusion bodies of *Escherichia coli*. *Protein Expr. Purif.* **2016**, *117*, 52–58. doi:10.1016/j.pep.2015.09.015.
97. Nozaki, Y.; Tanford, C. The solubility of amino acids and related compounds in aqueous urea solutions. *J. Biol. Chem.* **1963**, *238*, 4074–4081. doi:10.1016/S0021-9258(18)51830-5.
98. Nozaki, Y.; Tanford, C. The solubility of amino acids, diglycine, and triglycine in aqueous guanidine hydrochloride solutions. *J. Biol. Chem.* **1970**, *245*, 1648–1652. doi:10.1016/S0021-9258(19)77141-5.
99. Goyal, S.; Chattopadhyay, A.; Kasavajhala, K.; Priyakumar, U.D. Role of urea-aromatic stacking interactions in stabilizing the aromatic residues of the protein in urea-induced denatured state. *J. Am. Chem. Soc.* **2017**, *139*, 14931–14946. doi:10.1021/jacs.7b05463.
100. Raghunathan, S.; Jaganade, T.; Priyakumar, U.D. Urea-aromatic interactions in biology. *Biophys. Rev.* **2020**, *12*, 65–84. doi:10.1007/s12551-020-00620-9.
101. Lee, J.C.; Timasheff, S.N. Partial specific volumes and interactions with solvent components of proteins in guanidine hydrochloride. *Biochemistry* **1974**, *13*, 257–265. doi:10.1021/bi00699a005.
102. Alodia, N.; Jaganade, T.; Priyakumar, U.D. Quantum mechanical investigation of the nature of nucleobase-urea stacking interaction, a crucial driving force in RNA unfolding in aqueous urea. *J. Chem. Sci.* **2018**, *130*, 158. doi:10.1007/s12039-018-1563-8.
103. Sarkar, S.; Singh, P.C. Sequence specific hydrogen bond with DNA with denaturants affects its stability: Spectroscopic and simulation studies. *Biochim. Biophys. Acta* **2021**, *1865*, 129735. doi:10.1016/j.bbagen.2020.129735.
104. Shkel, I.A.; Knowles, B.; Record, M.T., Jr. Separating chemical and excluded volume interactions of polyethylene glycols with native proteins: comparison with PEG effects on DNA helix formation. *Biopolymers* **2015**, *103*, 517–527. doi:10.1002/bip.22662.
105. Arakawa, T.; Timasheff, S.N. Mechanism of poly(ethylene glycol) interaction with proteins. *Biochemistry* **1985**, *24*, 6756–6762. doi:10.1021/bi00345a005.
106. Lee, L.L.; Lee, J.C. Thermal stability of proteins in the presence of poly(ethylene glycols). *Biochemistry* **1987**, *26*, 7813–7819. doi:10.1021/bi00398a042.
107. Lee, J.C.; Lee, L.L. Preferential solvent interactions between proteins and polyethylene glycols. *J. Biol. Chem.* **1981**, *256*, 625–631. doi:10.1016/S0021-9258(19)70019-2.

108. Schellman, J.A. Protein stability in mixed solvents: a balance of contact interaction and excluded volume. *Biophys. J.* **2003**, *85*, 108–125. doi:10.1016/S0006-3495(03)74459-2.
109. Santos, A.; de Haro, M.L.; Fiumara, G.; Saija, F. The effective colloid interaction in the Asakura-Oosawa model. Assessment of non-pairwise terms from virial expansion. *J. Chem. Phys.* **2015**, *142*, 224903. doi:10.1063/1.4922031.
110. Minton, A.P. Macromolecular crowding. *Curr. Biol.* **2006**, *16*, R269–R271. doi:10.1016/j.cub.2006.03.047.
111. Rivas, G.; Minton, A.P. Macromolecular crowding in vitro, in vivo and in between. *Trends Biochem. Sci.* **2016**, *41*, 970–981. doi:10.1016/j.tibs.2016.08.013.
112. Hirano, A.; Shiraki, K.; Arakawa, T. Polyethylene glycol behaves like weak organic solvent. *Biopolymers* **2012**, *97*, 117–122. doi:10.1002/bip.21708.
113. Nakano, S.; Karimata, H.; Ohmichi, T.; Kawakami, J.; Sugiyama, N. The effect of molecular crowding with nucleotide length and cosolute structure on DNA duplex stability. *J. Am. Chem. Soc.* **2004**, *126*, 14330–14331. doi:10.1021/ja0463029.
114. Miyoshi, D.; Nakao, A.; Sugimoto, N. Molecular crowding regulates the structural switch of the DNA G-quadruplex. *Biochemistry* **2002**, *41*, 15017–15024. doi:10.1021/bi020412f.
115. Miyoshi, D.; Karimata, H.; Sugimoto, N. Hydration regulates thermodynamics of G-quadruplex formation under molecular crowding conditions. *J. Am. Chem. Soc.* **2006**, *128*, 7957–7963. doi:10.1021/ja061267m.
116. Baldwin, R.L.; Rose, G.D. How the hydrophobic factor drives protein folding. *Proc. Natl. Acad. Sci. U.S.A.* **2016**, *113*, 12462–12466. doi:10.1073/pnas.1610541113.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.