ACCOUNTS of chemical research

Commentary

pubs.acs.org/accounts

Characterizing the *Conformationome*: Toward a Structural Understanding of the Proteome

Published as part of the Accounts of Chemical Research special issue "Holy Grails in Chemistry".

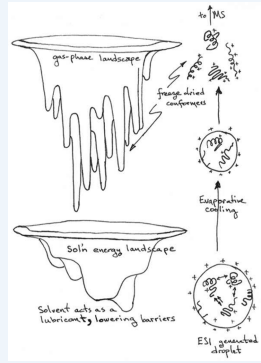
David E. Clemmer,*[,†] David H. Russell,*[,‡] and Evan R. Williams*[,§]

[†]Department of Chemistry, Indiana University, Bloomington, Indiana 47405, United States

[‡]Department of Chemistry, Texas A&M University, College Station, Texas 77843, United States

[§]Department of Chemistry, University of California, Berkeley 94720, United States

ABSTRACT: While non-native protein conformations such as folding intermediates are rarely observed in solution such species are often stabilized as gaseous ions during electrospray ionization for mass spectrometry. This opens the possibility of large scale efforts to capture information about many non-native structures such as folding intermediates or malformed conformations having deleterious effects: studies of the *conformationome*.



INTRODUCTION

One early outcome of genome sequencing is the realization that there are far fewer protein-coding genes than were expected for complex living organisms, only about 20000 in humans. And genomes are relatively static. Thus, a fundamental understanding of the complex changes that occur over an organism's lifecycle requires the characterization of genome products: proteins, the functional machinery of cells. It is estimated that one of the $\sim 10^{13}$ to 10^{14} cells (human, bacterial, and other) that makes each of us up expresses $\sim 10^4$ to 10^5 different proteins over its lifetime. And, significant percentages of proteins are made and eliminated each day, such that on a monthly basis (roughly) we are renewed, at least in terms of most of our proteins. Variations in protein abundances, structures, and functions alter cells and ultimately organisms, for example, from formation of networks of new tissues to the accumulation of malfunction leading to disease states.

Understanding the protein machinery of cells is a *holy grail*. This requires that we address the questions of *which proteins are present and how much of each at different points across an organism's lifespan: a characterization of the proteome.* We also need to know the role of post-translational modifications: *Where, on each protein, are modifications located? What signals each modification and regulates its accuracy? How does each modification affect structure and function?* The complexity of these questions expands as we seek to understand *who interacts with whom (the interactome),* scaling to a nearly inconceivable level since the number of possible associations increases exponentially with the number of interacting molecules. Complex machinery such as the ribosome involves many proteins (and RNA), intricately assembled into functional components. Because of the expansiveness of these questions and the experimental complexity of sorting through such systems using tedious measurements, one might seek to rephrase the problem. Perhaps the real issue is the conformational and functional changes of proteins within the entire system: networks of individuals participating in an intricate choreography of dance within the larger structure of the cell, sampling large numbers of configurations as they weave themselves into what are often viewed to be relatively static assemblies. Perhaps, this structural understanding (glimpses of the conformations of every protein, the outcome of sequences and concentrations, modifications and complex stoichiometries, snapshots of the distributions of folds and interactions responsible for the architecture of the complete proteome, *the conformationome*) is the real key to understanding the essence of proteins.

Such an understanding may allow for control of structure and reactivity. For example, in long-lived proteins, "misfolding" may lead to aggregation and fibril formation and is implicated in a number of human degenerative conditions, such as Alzheimer's disease. Insight into the structures of transient intermediates may present possibilities to intercede such that desirable conformations and reactivities are favored.

Received: October 31, 2016
Published: March 21, 2017

DOI: 10.1021/acs.accounts.6b00548
Acc. Chem. Res. 2017, 50, 556−560

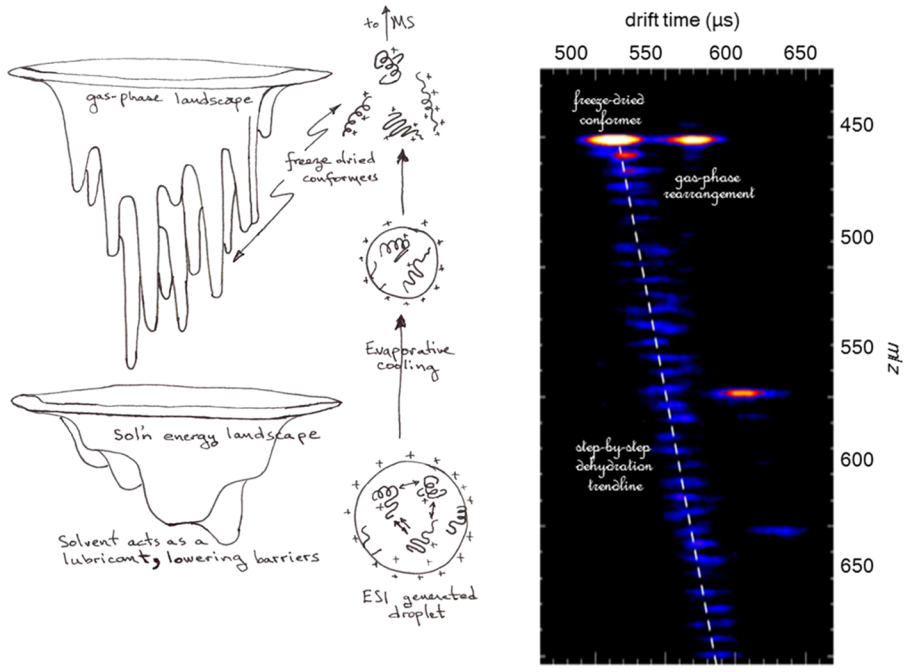


Figure 1. Hypothetical drawings illustrating protein structures emerging from solution into the gas phase by the process of ESI. In solution, solvent acts as a lubricant, lowering barriers for interconversion of conformations. Evaporative cooling, as droplets shrink, traps metastable structures on a new free energy landscape associated with the solvent free ion.

MASS SPECTROMETRY WITH "SOFT IONIZATION" FOR HIGH-THROUGHPUT PROTEIN ANALYSIS

As presented above, any measurement that encompasses all aspects of such complexity seems intractable. Thus, new measurement technologies are critical for advancing this holy grail in even the most rudimentary way. However, spectacular advances in cellular biology, computational and informatics approaches, theoretical models and protein simulations, and other chemical-labeling and analytical technologies are making it possible to begin thinking about taking the first baby steps toward such characterization. As an example, "soft ionization" techniques for mass spectrometry (MS), such as electrospray ionization (ESI),[1] have progressed rapidly and are transforming protein analysis. It is now routine to determine molecular weights of proteins to within fractions of a Dalton, and fragmentation methods for delineating sequences as well as positions and types of post-translational modifications have proliferated. Strategies for quantitation are available. Stoichiometries of individual proteins in complexes can be determined. The relative concentrations of species, as they are up- and down-regulated in cells, can be measured with great fidelity, with zeptomolar detection limits possible, that is, hundreds of molecules. Combinations of MS with separations and informatics-interpretation techniques make it possible to follow changes in expression levels, modifications, and abundances for thousands of proteins. While far from a solved problem, spectacular advances have been made.

But identification and quantitation are only a part of the story. Soft ionization makes it possible to address questions about protein conformation and dynamics, for ions in the gas phase as well as species in solution. Chemical labeling experiments, such as isotopic H−D exchange or cross-linking, provide insight about contact regions within or between proteins. In solution, the conformations that a protein accesses as it transitions between states, that is, "the protein folding problem", is exceptionally difficult to study. With few exceptions, experimental measurements show that populations of structures interconvert between folded (F) states and unfolded (U) states through cooperative, two-state F ↔ U transitions in a manner leaving traditional solution spectroscopies insensitive to pathways and intermediates. Thus, the step-by-step motions by which proteins fold is largely now addressed through computational or single molecule approaches, with few experimental benchmarks regarding non-native states against which to test theory. Interestingly, a vast range of structures can often be observed when solvent is removed, some closely resembling the native solution fold and other non-native states. At first, non-native structures in the gas phase were viewed somewhat as a curiosities, unnatural forms of naked polypeptide chains upon exposure to the harsh environment of a vacuum. However, it is becoming clear that many gas-phase ions preserve aspects of structure that was imposed by solution and thus provide an important glimpse into populations of non-native structures in solution. Once isolated, many powerful physical and chemical strategies can be used to assess conformation. Ion mobility spectrometry (IMS) for measuring collisional cross sections provides information about ion shapes and has emerged as a powerful structural tool. IMS-MS measurements can directly characterize complex mixtures of different sequences and conformations in a high-throughput manner with little sample, a necessary first step to efficiently characterize the *conformationome*. Of course characterization of shapes by mobility measurements is a crude determination of structure and benefits tremendously when configurations are determined by traditional structural methods, such as crystallography and NMR, which can reveal atomic details or other complementary techniques such as cryo-electron microscopy or circular dichroism.

Figure 1 illustrates how information about different conformations in solution can be preserved during the process of creating gaseous ions by ESI. Consider the hypothetical

DOI: 10.1021/acs.accounts.6b00548
Acc. Chem. Res. 2017, 50, 556−560

energy landscapes showing relative stabilities of different structures in solution and upon emergence into the gas phase. In the presence of a lubricating solvent such as water, small energetic barriers separate different conformers resulting in rapid interconversion between structures, making them hard to differentiate experimentally. The process of producing gaseous ions from electrospray droplets is accompanied by evaporative cooling, which slows or even stops interconversion between states. The resulting gaseous ions do not have the exact same structures as they did in solution but can retain a "memory" of their solution structures. These ions can be isolated and preserved for long times and information about the antecedent solution structures can be obtained by characterizing the anhydrous conformations.

The final step-by-step losses of individual solvent molecules associated with formation of "naked" ions by ESI have been captured by cryogenic IMS-MS measurements.[2] An example data set (Figure 1) for the nonapeptide substance P reveals two mechanisms for forming gas-phase conformations. Compact structures emerge from a clear trendline of associated varying degrees of hydration; a more extended structure is independent of solvent. These studies illustrate the general observation that the conformations of gaseous ions produced by ESI are often metastable, kinetically trapped species, "frozen out" as a result of evaporative cooling as ions are formed.

SOLUTION FOLDING MECHANISMS, PATHWAYS, AND ENERGY LANDSCAPES FROM MONITORING IONS

While often there are similarities in the structures of dehydrated ions with the antecedent solution populations, typically dehydrated states reveal many new types of conformations that are not discernible in solution. This difference makes it possible to follow structural transitions in detail. Consider the small model peptide, polyproline-13 (Pro13). The 12 peptide bonds linking the 13 proline residues can be configured in either a *cis*-orientation, leading to the right-handed PPI helix favored in nonpolar solvents such as propanol, or a *trans*-orientation for the left-handed PPII helix favored in water. Although this simple transition has been known for ~60 years, the inability to follow ensembles of intermediates in solution by conventional measurements prohibits determination of the folding mechanism and pathway. At first glance, it is surprising that such a small homopolymer presents such a challenge; however, the simplest direct pathway for complete PPI(*cis*) → PPII(*trans*) conversion could involve any combination of 12! (479,001,600) unique folding pathways. It turns out that it is possible to follow the solution transition in detail by monitoring the gas-phase ions. When the all *cis*-PPI form of Pro13 (in propanol) is diluted into water, it is possible to trap intermediates that exist in this solution as gaseous ions.[3] Additionally, theoretical calculations of intermediate structures are simplified for *in vacuo* species, making it possible to characterize the detailed *cis*–*trans* configurations of intermediate structures in solution based on calculations of ion shapes. The IMS measurements in Figure 2 show evidence for six intermediate states during the PPI → PPII transition, and fits to the kinetics data require a sequential folding pathway. The measured cross sections require that the transition initiates from the N-terminal side of the peptide, flipping two peptide bonds (A → B), and then a third (B → C). From intermediate C, a *cis* → *trans* conversion occurs near the midpoint of the peptide before proceeding to the all-trans PPII product.

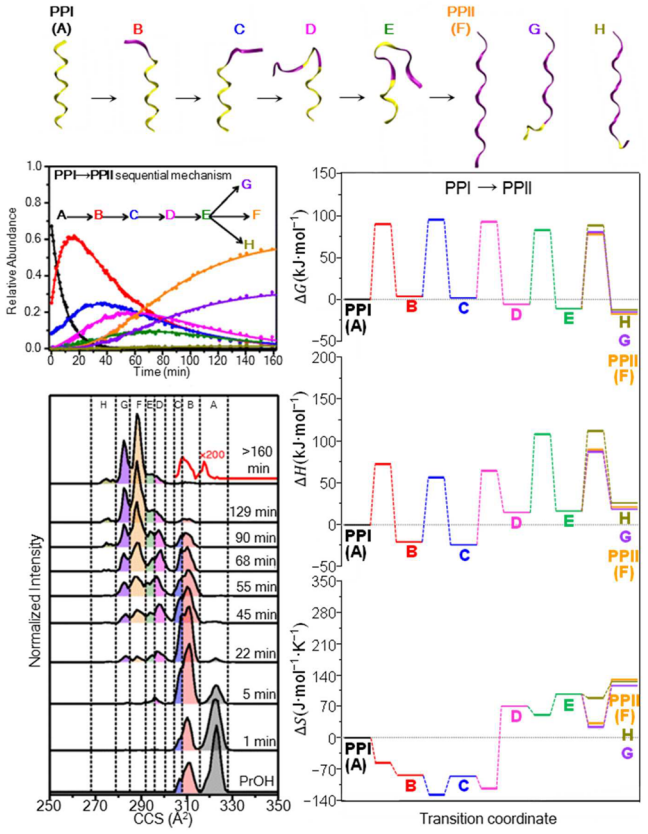


Figure 2. Ion mobility, cross section distributions for $[Pro13 + 2H]^{2+}$ showing the transition from PPI → PPII in 10:88:2 1-propanol/H_2O/acetic acid (v/v/v) (left, bottom) at 23 °C. Conformer regions are shown as different colors: A (black), B (red), C (blue), D (magenta), E (green), F (orange), G (violet), and H (olive). The initial distribution (from 1-propanol) is shown as the bottom panel. The inset (red) shows a blow up of low-abundance species present at equilibrium. Dashed lines delineate the regions for each conformer used to extract corresponding kinetics plots (left, top). The right panel shows thermodynamic landscapes (Gibbs free energy, top; enthalpy, middle; and, entropy, bottom) and structures for proposed intermediates as determined through molecular dynamics simulations. In the proposed structures, *cis* regions are shown as yellow ribbons, and the *trans* are shown as purple ribbons.

Measurements of the kinetics and equilibria at different temperatures provide details about the folding pathway within the landscape, revealing not only how the transition occurs but also why. Values of ΔH show that early steps are exothermic; presumably carbonyl groups along the backbone of early intermediates are stabilized by new water hydrogen bonds. This interaction is entropically disfavored, limiting the direct N → C-terminal pathway to the first few steps. The system gets around the entropic restriction by a *cis* → *trans* flip near the oligomer's midpoint and uses entropy to overcome the subsequent endothermicity to complete the folding transition.

FAST FOLDING IN DROPLETS

ESI droplets naturally confine proteins in miniscule (atto- to femtoliter) volumes. This confinement makes it possible to study extremely fast conformational changes, such as folding or unfolding transitions induced upon rapid (~1 μs) solvent exchange. Figure 3 shows a measurement for the ~12.4 kDa protein cytochrome *c*.[4] In this case, the acid denatured form of

DOI: 10.1021/acs.accounts.6b00548
Acc. Chem. Res. 2017, 50, 556–560

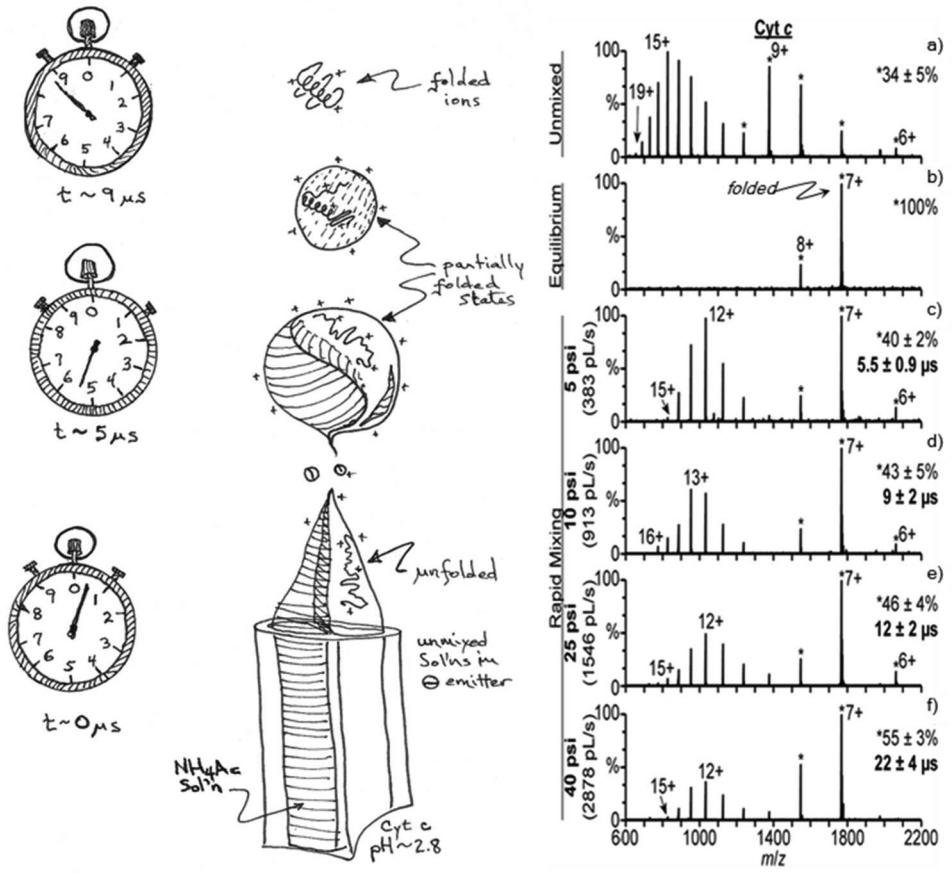


Figure 3. (left) Illustration of the fast mixing of two solutions using a dual-fluid θ-tip emitter for producing ESI droplets. The U → F transition is monitored from changes the abundances of high charge state (U) and low charge state (F) ions. Mass spectra correspond to (a) an acidified aqueous cytochrome *c* (cyt *c*) solution (pH = 2.8), (b) the acidified cyt *c* solution mixed with an ammonium acetate solution producing a pH jump to 4.4, and (c-f) spectra upon sampling different mixing times by varying the ESI flow rates. The signals centered around the +12 charge state are associated with acid denatured cyt *c*, and those centered at +7 are associated with the folded protein.

the protein (pH = 2.8) is mixed in a 1:1 ratio with an ammonium acetate solution at the tip of the ESI emitter. The pH jump (2.8 → 4.4) associated with formation of the mixed droplet initiates the U → F transition, which is monitored by measuring the ESI charge state distributions. By changing the flow rate of the solution through the capillary, it is possible to control droplet lifetimes, enabling the extent of folding, prior to trapping the reaction products as gaseous ions, to be monitored as a function of reaction times as short as 1 μs.

■ OUTLOOK

While MS-based techniques for sampling non-native structures, folding transitions, and thermodynamics are at an early stage, these methods show promise for examining complex structural systems. High-throughput, low-detection limit measurements can be conducted over wide ranges of concentrations. It is possible to simultaneously characterize conformations and structural transitions on mixtures of species in highly controlled environments—to the extreme of understanding the influence of a single solvent molecule upon conformational changes. In some sense, characterization of the *conformationome* has already begun as IMS-MS proteomics measurements, which naturally yield cross sections for peptide sequences. One can imagine extending this approach to unimolecular folding transitions to identify sequences that undergo interesting structural changes. New spectroscopies and methods for chemical labeling and fragmenting specific conformations show great promise for high-throughput conformational characterization, and hyphenated approaches will undoubtedly improve the ability to discern structural details. Still, these approaches are primitive. In order to understand the *conformationome,* one desires structures for vast numbers of non-native sequences, modifications, and complexes at atomically detailed resolutions, which is hard to imagine at this point by MS based methods alone. Other complementary structural methods and molecular biochemistry approaches will surely be needed. Should vast information of this type exist, the details may prove prohibitively complex for simple interpretation. Some clues about non-native complexity might be drawn from the ~10^5 high-resolution *native* structures currently available in the Protein Data Bank. These structures fall into only ~1200 folds. Perhaps, key non-native states within the *conformationome* will also fall into a relatively small number of unique families. Non-native structural families for some sequences might even overlap with native families for other sequences. In this case, we would be left with a relatively small number of ways to manipulate structures: lowering the energy of preferable structures, either through drug or gene therapies, in order to gently avoid malformed states with deleterious functions or to increase the throughput of a particularly favorable metabolic pathway. Certainly, large experimental data sets describing the structures and properties of non-native conformations would be of great value to theoretical efforts to

DOI: 10.1021/acs.accounts.6b00548
Acc. Chem. Res. 2017, 50, 556−560

understand pathways and function. Similarly, understanding how the protein environment can affect structure or even what changes occur in the gas-phase itself could provide useful information about how cellular structures, such as membranes or confinement in vesicles, can affect protein conformations and their interactions with other proteins or other biopolymers.

AUTHOR INFORMATION

Corresponding Authors

*David E. Clemmer. E-mail: clemmer@indiana.edu.
*David H. Russell. E-mail: Russell@chem.tamu.edu.
*Evan R. Williams. E-mail: erw@berkeley.edu.

ORCID

David E. Clemmer: 0000-0003-4039-1360
David H. Russell: 0000-0003-0830-3914
Evan R. Williams: 0000-0002-1733-3018

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank many students, colleagues, and reviewers for stimulating discussions.

REFERENCES

(1) Fenn, J. B.; Mann, M.; Meng, C. K.; Wong, S. F.; Whitehouse, C. M. Electrospray Ionization for Mass-Spectrometry of Large Biomolecules. *Science* **1989**, *246*, 64−71.

(2) Silveira, J. A.; Fort, K. L.; Pierson, N. A.; Clemmer, D. E.; Russell, D. H.; et al. From Solution to the Gas Phase: Stepwise Dehydration and Kinetic Trapping of Substance P Reveals the Origin of Peptide Conformations. *J. Am. Chem. Soc.* **2013**, *135*, 19147−19153.

(3) Shi, L.; Holliday, A. E.; Glover, M. S.; Ewing, M. A.; Russell, D. H.; Clemmer, D. E. Ion Mobility-Mass Spectrometry Reveals the Energetics of Intermediates that Guide Polyproline Folding. *J. Am. Soc. Mass Spectrom.* **2016**, *27*, 22−30.

(4) Mortensen, D. N.; Williams, E. R. Ultrafast (1 μs) Mixing and Fast Protein Folding in Nanodrops Monitored by Mass Spectrometry. *J. Am. Chem. Soc.* **2016**, *138*, 3453−3460.

DOI: 10.1021/acs.accounts.6b00548
Acc. Chem. Res. 2017, 50, 556−560