THE JOURNAL OF PHYSICAL CHEMISTRY B
Special Issue Preface
Cite This: J. Phys. Chem. B 2018, 122, 5263−5266
pubs.acs.org/JPCB

Autobiography of Ken A. Dill

I had a pretty normal upbringing, in 1950s Oklahoma. My family's means were modest. My Dad, Austin, had been raised in *Grapes of Wrath* dustbowl Oklahoma in the 1930s, by parents who were bankrupted in the Great Depression. As a kid, he often could not afford shoes, but he somehow acquired a donkey.

Ours was a loving and supportive family. My Mom Peggy, raised in Maplewood, New Jersey, was a cheery woman who loved to sing and play piano. My Dad, an engineer with the Southwestern Bell Telephone company, was our Boy Scout leader and a baseball coach. He taught us how to build and fix things. Like many kids of that era, we had lots of unsupervised free time. Much of mine was spent in sand-lot pick-up baseball. I was the first baseman on a little-league team that won the 1957 Oklahoma state championship. My brother Jim, a Harvard- and Princeton-trained computational chemist, now retired from CambridgeSoft, has been a lifelong companion.

My junior-high-school years were spent in Westfield, New Jersey, where a buddy Larry Hodes and I developed a hobby in electronics. I built transistor circuits, made a few bucks fixing TVs, and got a ham radio license. When we returned to Oklahoma, I went to Putnam City High School in Warr Acres. There, I worked on science-fair projects in math and physics, helped hugely by teachers Milton Boydstun and Don Blackerby. I did a math project on algebraic rings that reached the state level and caught the interest of science-fair judge Professor Roy Deal from Oklahoma State University. In an act of midwestern generosity, Deal showed me much fancier math than I knew, wrote me a paper, affixed my name to it as sole author, and got it published in the junior division of the Oklahoma Academy of Sciences (alas, still uncited...). It got me hooked on grown-up science. I had various summer jobs—in a print shop; painting houses; mowing lawns; as well as shoveling manure, bailing hay, and drinking cases of cokes at Dale Robertson's Horse Ranch to replenish buckets of sweat in the Oklahoma heat; and working in a rural hospital cleaning bedpans and fixing broken radios. I spent endless hours making electronics gizmos in a shop I built in our attic. I do not recommend summers in Oklahoma attics.

I entered MIT as an undergraduate in 1966. It had an unfamiliar but exciting intensity, partly resulting from the college-student soul-searching of the Viet Nam War years, and partly because MIT was the first community I met that was passionate about science and technology. By the time I finished an SB/SM program in Mechanical Engineering, I had acquired a reverence for the principles and methods of math and physics, but the problems I found most compelling were those of biology, particularly origins of life, thanks to a special-topics course with MIT's Alex Rich. So, in 1971 I joined the Biology PhD program at the University of California, San Diego—a university that was then less than a decade old—and did rotation projects with origins-of-life researchers Stanley Miller and Leslie Orgel.

Unsure of my abilities in experimental chemistry, and unclear how to find definitive answers to life's origins, I undertook my PhD research in biophysics with Bruno H Zimm, a pioneer of polymer physics, known for Rouse−Zimm theory of polymer dynamics and Zimm−Bragg theory of helix−coil transitions. His lab was a magnet for a kid who liked math and physics and machine shops. I revered his broad knowledge of physics, his ability to apply it to solving real-world problems, and his instincts for always knowing the simplest way to approach problems. In my PhD work, I built a gizmo for measuring sizes of DNA by viscoelasticity, and learned a little polymer statistical mechanics. Zimm's biophysics course introduced me to the mystery—then seen as a grand challenge—called the protein-folding problem.

Intrigued, I did postdoctoral work at Stanford (1979−1981) with polymer scientist Paul J. Flory. I hoped that Flory's lattice models could teach me how polymers are organized in tight spaces, like inside the cores of folded proteins. My two-year postdoc was spent on the warm-up problem of how lipid chains pack within the cores of micelles and bilayer membranes. Flory was larger-than-life—a Nobel Prizewinner who had a passion for science and a compassion for Russian dissident scientists, whose rights he supported. We lab members often helped. In 1984, he even offered himself up as a hostage (not accepted) to guarantee the return of Andrei Sakharov's ailing wife from medical care outside of Russia.

Upon moving in 1981 to become an assistant professor in Chemistry at the University of Florida, Gainesville, I found some mental space to think about protein folding. This continued when I moved in 1983 to a faculty position in Pharmaceutical Chemistry at the University of California, San Francisco, where I stayed until 2010. UCSF was a uniquely rich and supportive research environment of terrific colleagues and students. UCSF's Pharmacy Dean Jere Goyan had assembled a team that became founders of computational biology: Bob Langridge, an originator of molecular graphics; Tack Kuntz, originator of DOCK, distance geometry, and other methods of biomolecule discovery; Peter Kollman, originator of the physics-based molecular simulation software called AMBER; and Tom James, a pioneer in NMR approaches to structural biology. This was the perfect ecosystem for my group's interests in protein folding.

In 1979, I met Jolanda Schreurs—the woman who would become my wife—at a seminar at Stanford. She was a PhD student in Pharmacology when I was with Flory. We've been happily married since 1984. She tolerates my equations lying around the house, probably because such litter was familiar from her Dad, a PhD chemical physicist at Corning Glass, whose evenings were spent working theoretical astrophysics problems. Because of her scientific breadth and engagement, Jolanda has always been a close partner in my scientific life. One of my greatest joys in life has been our tight-knit family with our two sons, Tyler and Ryan, now a PhD nanoengineer and physical chemist, respectively. We had many unforgettable

Special Issue: Ken A. Dill Festschrift

Published: May 31, 2018

Downloaded via STONY BROOK UNIV SUNY on June 22, 2018 at 20:00:33 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

ACS Publications © 2018 American Chemical Society

DOI: 10.1021/acs.jpcb.8b02469
J. Phys. Chem. B 2018, 122, 5263−5266

adventures backpacking in the Sierras, in Boy Scouts, building gizmos in our garage, and traveling as a family.

Now, to keep this essay more a scientific autobiography than a review, I'll focus on my personal perceptions and the questions that drove our research group, simplifying here and there.

A protein molecule is a nanoscale piece of string. On one hand, it can flop and twist into many open shapes. On the other hand, in the cell, a protein collapses down into a specific compact ball-of-string shape, different for different amino-acid sequences, and with each one performing a different function in biology. *How does a protein find its right shape so quickly through random physical processes?* This has been called the "protein folding problem". It was seen as a needle-in-a-haystack problem because it was not clear how the protein could find its biological (native) structure so reliably and quickly by searching through the astronomically large "space" of all its possible floppy-string shapes. But this perspective seemed wrong to me. Many problems of material science, when viewed at the microscopic scale of statistical mechanics, resembled needles in haystacks. Despite astronomical searches, large entropies, and infinitesimal probabilities, many were solvable. (Think about crystallizing salt from salty water. If every molecule concentrates by 10-fold in that process, then crystallization by random actions is improbable by (1/10) raised to the power of Avogadro's number!) Our first question was "How big is the protein-folding haystack?" Polymer theories estimated the numbers as $z^N \sim 10^{60}$ conformations, where z is the number of rotational states for each peptide bond, roughly $z \sim 4$, and N is the number of peptide bonds (100, for a small protein). I saw the key to this needle/haystack problem as the Flory−Huggins polymer concept of "excluded volume": a chain cannot physically pass through itself. In a metaphorical "spaghetti" of polymer chains, 10^{60} is the number of states accessible to one "noodle of wet spaghetti" alone free in space. But in a pot dense with other spaghetti, one noodle's conformational space is reduced to $(z/e)^N$, a reduction of 43 orders of magnitude! Flory's spaghetti pot contained multiple chains. My spaghetti pot contained just one chain—a folding protein—no part of which could pass through any other. Excluded volume meant that the protein-folding haystack was not so big after all.

But still, the critical questions were: *How does the protein avoid searching most of the haystack? And, how does it find the needle so fast?* The prevailing views were that hydrogen bonding was a main driver of protein structure; that hydrophobic interactions were nonspecific "glue", that just kept the protein balled-up rather than holding it into a particular shape; and that protein structures resulted from many types of small forces, not a single dominant one. However, hydrogen bonds were known to be weak. And I felt that hydrogen bonds could not be structure-determining because different sequences can make roughly the same hydrogen bonds. Hydrogen bonds are mostly in the protein backbone, where all sequences are the same. Sequence-structure relations must come from the patterning of the amino-acid side-chains. So, I tried a different model, namely, that hydrophobic interactions might themselves be dominant and the origin of the folding code. I modeled proteins as having a binary HP code (H hydrophobic; P polar), wrote a Flory-like polymer collapse theory, and published it in 1985 in *Biochemistry*. I argued that proteins fold on funnel-shaped energy landscapes and that the hydrophobic effect—rather than hydrogen bonding—is the dominant component of the folding code.

Below are our lab efforts that followed. I am deeply grateful for support from the NIH and NSF, and for the many outstanding students and colleagues who have powered these projects. I'm sorry that because of the threads I've chosen to describe below, I necessarily have to leave out many other interesting ones and key collaborators.

What are the forces that drive protein structures? Since we were already counting conformations in the HP mean-field model (thus, estimating chain entropies), and sorting them by hydrophobic energies, we could immediately compute folding free energies, as functions of charges, solvation, temperature, and other properties and aggregation properties (Alonso, Stigter, Ghosh, Fields, Zhou).

But, was our mean-field approximation of chain entropy accurate enough? To check, we published in 1989 what we called an "exact HP lattice model", where we exhaustively enumerated every conformation of short chains by computer (Lau, Chan, Fiebig, Bromberg, Miller, Thomas). Yes, the mean-field approximation gave the right shape of the density of states (Chan, Ghosh).

Exact modeling also opened new doors. For one thing, it explained the "folding code". Our better counting now showed that excluded volume reduced the conformational search by all 60 orders of magnitude (for a chain of 100 monomers), not just 43. It meant that even a model that represented polymers at low resolution, with sequences drawn only from a binary alphabet of monomer types based on hydrophobicity, could explain folding to a single unique native structure, without the need to treat the atomistic details. A key experimental proof of the binary HP-code hypothesis came in 1993, when Michael Hecht's group at Princeton published in *Science* an elegant experiment of partially randomized HP-encoded sequences.

Exact modeling also gave us the tool we needed for origins of life questions. *What is the probability that a prebiotic random sequence might fold and be protein-like?* This was another needle-in-a-haystack problem. We found that a given folded protein shape is encodable within a huge number of different sequences, indicating that random sequences are often quite protein-like. In this case, the haystack was more than 100 orders of magnitude smaller than previously estimated, implying that the chemical origins of life were not as impossibly unlikely as was thought (Lau, Chan). Bob Sauer, with Wendell Lim and others in his group at MIT, published in *Nature* in 1989 and in *Science* in 1990 experimental proof that large numbers of random sequences do indeed fold like native proteins.

For another thing, the HP lattice model had sufficiently controllable complexity such that it has been useful to computer scientists who develop global optimization methods and quantum computers.

What was the first biomolecule? It is often argued to be RNA, but HP modeling shows instead that it might have been proteins (Guseva). Unlike RNA, random HP chains have the capability for self-elongation, self-catalysis, and spontaneous transitioning toward informational sequences. Our forays into origins have led us into occasional tangles with the Intelligent Design community (https://www.youtube.com/watch?v=fOtP7HEuDYA).

Next, we asked: *Why are proteins so special?* As a "proteins guy", I hold the modest opinion that what mainly distinguishes between a walking talking living being and a rock is proteins. Arguably, the cell owes its thousands of biochemical functions to its many protein folds. *Is foldability in proteins inherent in*

DOI: 10.1021/acs.jpcb.8b02469
J. Phys. Chem. B 2018, 122, 5263−5266

some special aspect of the chemistry of peptides? Our folding-code ideas were telling us that the specialness of proteins was just a result of two things: (1) that proteins have multiple monomer types, solvent-disliking (H) and solvent-liking (P), and (2) that those monomers could be strung together in specific orders along a chain. Chemists making nonbiological polymers had not been able to string monomers into specific sequences. My insightful wife, who was working at Chiron Corp in the 1990s, introduced me to organic polymer chemist Ron Zuckermann. In 1992, Zuckermann et al. published the technology to make a new class of polymer called *peptoids*. In a first for polymer synthesis, Ron was able to string together specific sequences, including H and P monomers. When Chiron lost interest in peptoids as small-molecule drugs, Ron and I schemed instead to make long-chain peptoids to test folding principles. I think these were the first nonbiological molecules found to fold to have secondary and tertiary structures, like proteins have. The peptoid story has now grown into a bigger enterprise that develops new polymeric materials, called *foldamers*, partly powered by former students and postdocs (Zuckermann, Kirshenbaum, Barron, Lee).

Another puzzle was the so-called *Levinthal paradox. Is there a folding mechanism?* Is there a general sequence of events by which a protein reaches its native structure—rapidly, without searching the whole space, and for any protein molecule, no matter what its final "endstate" native structure is? The funnel idea alone was not sufficient. Funnels had only explained that there are many high-entropy floppy conformations and few compact low-entropy states. In the 1990s, we were finding fruitful the idea that proteins follow a "divide and conquer" strategy ("local metastability first, global stability later"), which we called *Zipping and Assembly* (Chan, Fiebig, Ozkan, Ghosh, Weikl, Voelz, Hockenmaier, Rollins). The chain first explores different twists and bends and turns and helices, in localized regions of the chain, finding metastable states of them, and then the chain randomly bumps those pieces together, leading to increasingly larger assemblies and more native-like structure. Still, we needed numbers. *Could we compute the right folding speeds, across many different proteins, based on known physical interaction energies, and within a framework of the randomness and thermodynamics required of a physical mechanism?* Yes, Geoff Rollins captured this in his *Foldon Assembly Mechanism*.

During the 1990s, our growing theory community found it tough to convince biochemists of the value of polymer theory. Since the 1940s, there had been an intellectual wall between polymer science, with its focus on disorder, random flights, and conformational ensembles, and protein science, with its focus on precisely detailed atomistic structures and their role in biological functions. We encountered resistance of three types: (1) Our modeling was criticized for leaving out the atom-level details. However, as Mark Kac and others had said before, the value of modeling is in learning from *making choices*, not from *making mirror images of reality in a computer*. To model the quack of a duck, it is not helpful to model the feathers. Some problems (that entail entropies or the nature of conformational or sequence spaces) are not governed by details, but by global features. Our questions required that we capture those properties. Also, ironically, (2) we theorists looked at the kinetic routes *more microscopically* than the experimentalists could then measure, and our two communities were developing conflicting terminologies. For us, *pathways* were ensembles of microscopic processes; for them, pathways were coarser—whatever you could observe in experiments. (3) Theorists and experimentalists were prisoners of our different cultures. We theorists grew animated about folding funnels in the 1980s. Experimentalists only started referring to folding funnels as the "new view" of protein folding a decade later. Why the long delay? I believe it was simply that it took us theorists awhile to recognize the communication value of making simple cartoon pictures of funnels. The idea those pictures represent—statistical mechanical densities of states—had seemed dry as dust, and were hard to appreciate by those outside our own circle.

During the early 2000s, as a past-president of the Biophysical Society, I did some science-policy work in Washington on the side. In conjunction with the Biophysical Society (including society officers Ro Kampman and Ellen Weiss), my colleague Mary Barkley (Professor of Chemistry, Case-Western Reserve University) and I assembled a "Bridging the Sciences" coalition of basic research societies, ultimately representing more than 200,000 scientists. We went on a mission to the US Congress and federal agencies. Assisting us was John E. Porter, a former US congressman from Illinois, salt-of-the-earth human being, and extraordinary public servant. We felt the US government was no longer adequately funding *deep innovation*, as it had during the Cold War and Space Race in the 1950s and 1960s. By deep innovation, we meant something different from normal innovation, which our federal agencies—NIH, NSF, DOE—have supported very effectively. We meant research that was curiosity-driven, not payoff-driven, with outcomes that were more unexpected than expected, with funding focused more on the capabilities of individuals than the expected payoffs of projects, and where our mission agencies could reach beyond their defined missions to *explore the unknown*, rather than to *exploit the known*, for solving tomorrow's problems, not today's. Our work helped to initiate some new grant programs and raised some agency awareness. But, far more should be done.

The 1990s saw advances in two fields of protein computational physics: (i) native-structure prediction, and (ii) atomistically detailed physical molecular dynamics simulations of proteins. The ultimate goal of item (ii) has been to harness computing to understand how proteins undergo their biological actions, and to speed up drug discovery, using the principles of physics. But the computational expense has been too high; there are inaccuracies in the physics, and scaling is not good enough to handle big actions or big proteins. This led our group into two efforts at tool-building for atomistic physical simulations (below), as well as some drug discovery modeling (Shoichet, Rocklin, Mobley, Chodera, Morrone, Perez, Simmerling, MacCallum, Liu).

We work on better and faster computer modeling of water. Proteins fold into shapes by moving water molecules out of the way. You cannot understand protein physics unless you understand water. And, each case involves many water molecules, so computers that are simulating proteins spend lots of time just simulating the waters. We wanted new models that would be both accurate and fast to compute. But, water is a complicated liquid because it makes molecular cages. Before we could make better models, we needed to know the following: *How does its "caginess" drive water's structuring and energies when it forms blankets ("solvation shells") around other molecules, such as proteins?* The problem is that, in cages, water's positions are coupled to its orientations, and these states are too rare and fleeting in molecular dynamics simulations to characterize adequately. So, over 20 years, we have built models and sampling approaches that could focus on these small

DOI: 10.1021/acs.jpcb.8b02469
J. Phys. Chem. B 2018, 122, 5263−5266

populations, and their corresponding entropies, heat capacities, and free energies (Silverstein, Haymet, Southall, Truskett, Chorny, Fennell, Kehoe, Mobley, Chodera, Li, plus a longstanding partnership with Prof Vojko Vlachy in Slovenia, and his associates Hribar-Lee, Urbič, Kalyuzhnyi, Luksic, Kastelic, and others). For different subquestions, we made various toy models (such as the "Mercedes-Benz model", where each 2D molecule has 3 arms like the car logo); we applied different statistical mechanical tricks and approximations, and we did computer simulations of detailed models in some cases. These studies led to a new model (SEA Water) (Fennell, Kehoe, Li) that is relatively fast and accurate, although not yet available in standard software packages.

We work on algorithms to tame the exponential explosion of conformational sampling. Even though proteins can search their haystacks rapidly, computers cannot, without clever algorithms. We have worked on several methods (Phillips, Rosen, Yue, Seok, Chodera, Mobley, Coutsias, Shell, MacCallum, Perez, Brini, Morrone, Robertson, Nassar). I am currently enthusiastic about MELD, developed by Justin MacCallum and Alberto Perez, a method for accelerating molecular dynamics by "melding" it with external knowledge or insights that are not required to be precise or definitive.

In 2010, I moved to Stony Brook University to establish the Laufer Center for Physical and Quantitative Biology, in cooperation with University President Sam Stanley and donors Henry and Marsha Laufer, whose support, engagement, enthusiasm, and scientific and operational freedom I have deeply appreciated. We are building a research collaboration around the math and physics needed to better understand biomolecular and cellular actions and evolution. It is a work in progress, so I just indicate below our current motivations.

Imagine someday having a framework of the complex dynamics of cells and evolution that is as powerful as the Second Law is today for material equilibria and kinetics (Ghosh, Presse, Hazoglou, Peterson, Dixit, Stock, Lee, Ge, Agozzino). Unlike dead stuff, cells are born and die, seek food, duplicate, compete, collaborate, attack, escape, remember, and improve. They can even rewire their own operating rules. *How are purpose-like cell mechanisms and evolution encoded in physical chemistry and random actions? And, what is the role of protein folding in cell health?* Proteins have their "day jobs" (their individual biological functions). Proteins also have their "health" (folded or not, aggregated or not). Metaphorically, cities function not only because city workers perform day jobs, but also because those workers are healthy and have the infrastructure they need. Aging and disease and cellular action mechanisms depend on both the biology and the physics of proteins (Maitra, Santra, de Graff, Szenk, Schmit, Ghosh, Peterson, Wagoner, Weistuch).

I've had a lifelong affinity for statistical mechanics, with its central focus on entropies and variational principles, and its capacity to express how astronomical searches and infinitesimal probabilities manifest on ordinary time scales in real life. I've liked simplicity and analytical theory. We've tried to share these enthusiasms in textbooks: *Molecular Driving Forces* with Sarina Bromberg, and *Protein Actions* with Ivet Bahar and Bob Jernigan. I owe special thanks to Sarina, who, over many years, continues to help me and my group with our writings and communication, and art and style.

Over the years, I've been greatly privileged to share in the pleasures of the academic enterprise—the freedom of inquiry, the joy of discovery, the engagement with like-minded people seeking truth and knowledge, the fun of teaching what we learn, the eye-opening new narratives we learn from other scientists, and the delight in watching my students and postdocs grow their own careers. Research is humbling because there is right and wrong in nature, so many of our expectations are wrong, and distinguishing right from wrong takes hard work and discipline. But, what a great ride!

Ken A. Dill

DOI: 10.1021/acs.jpcb.8b02469
J. Phys. Chem. B 2018, 122, 5263−5266