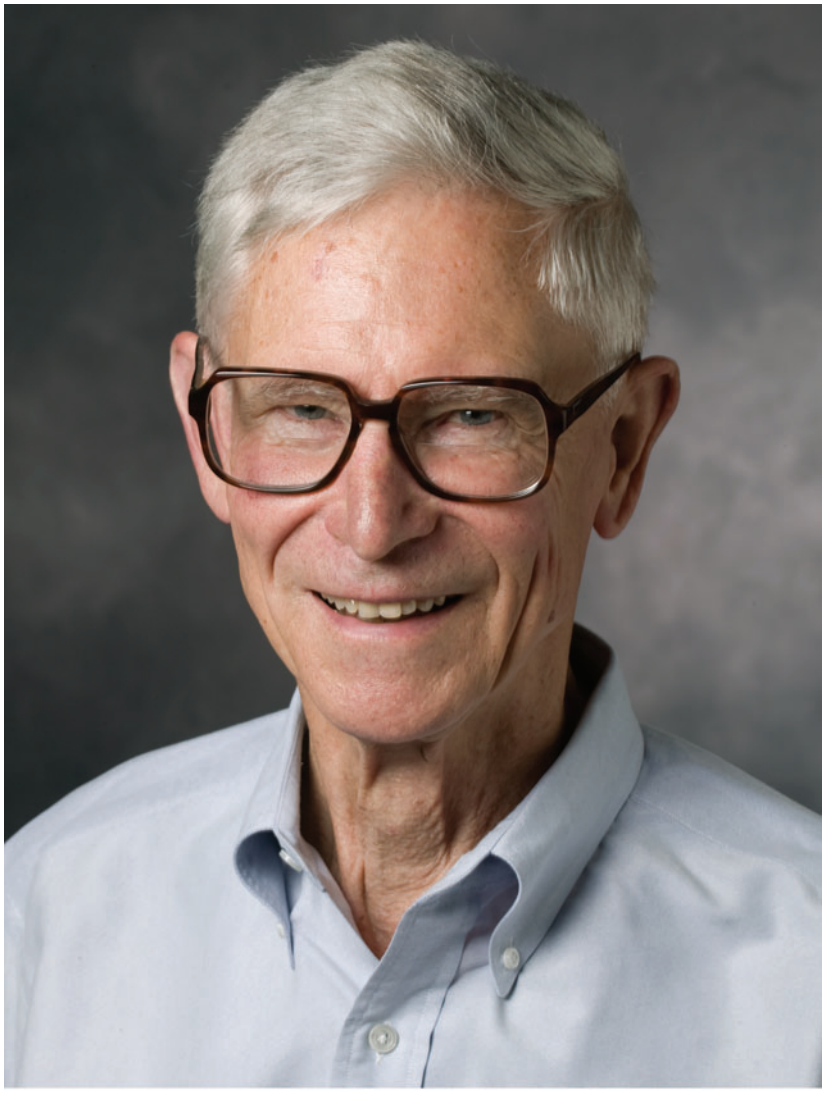

Robert L. Baldwin

# The Search for Folding Intermediates and the Mechanism of Protein Folding

Robert L. Baldwin

Biochemistry Department, Beckman Center, Stanford University Medical Center, Stanford, California 94305; e-mail: baldwinb@stanford.edu

Annu. Rev. Biophys. 2008. 37:1–21

First published online as a Review in Advance on November 21, 2007

The *Annual Review of Biophysics* is online at biophys.annualreviews.org

This article's doi:
10.1146/annurev.biophys.37.032807.125948



1936-122X/08/0609-0001$20.00

## Key Words

protein folding intermediates, peptide helices, hierarchic mechanism

## Abstract

My research began with theory and methods for ultracentrifugal studies of proteins, first at the University of Wisconsin, Madison, with Bob Alberty and Jack Williams, then at Oxford University with A.G. ("Sandy") Ogston, and finally back at Wisconsin with Williams and Lou Gosting. In 1959 I joined Arthur Kornberg's Biochemistry Department at Stanford University. Our first work was physical studies of DNA replication and then DNA physical chemistry, and DNA studies ended with the energetics of DNA twisting. In 1971 we began to search for protein folding intermediates by fast-reaction methods. We found the slow-folding and fast-folding forms of unfolded ribonuclease A, which led to the understanding that proline isomerization is sometimes part of the folding process. Using hydrogen exchange as a probe, we found the rapid formation of secondary structure during folding and used this to provide an NMR pulse labeling method for determining structures of folding intermediates. Our studies of peptide helices provided basic helix-coil parameters, also evidence for hierarchic folding, and further indicated that peptide hydrogen bonds are important in the energetics of folding.

## Contents



## PROLOGUE

My father, Ira Baldwin, was a scientist himself. He remarked once that the scientific achievements of a scientist often depend on his having been in the right place at the right time. I feel lucky indeed to have been in the Stanford Biochemistry Department, founded by Arthur Kornberg, when the new science of molecular biology was getting into gear. Several people have said that someone should write the story of this remarkable department. My part was to be a biophysicist in a department of enzymologists and molecular geneticists. Andy Robertson, when he was a postdoc with me, said it was like being a child in a candy store.

## GETTING STARTED IN SCIENCE

I grew up in Madison, Wisconsin, home to the University of Wisconsin, where Dad was a professor of bacteriology. He loved science and he wanted his children to appreciate the scientific method. He taught us not to accept received knowledge simply because a teacher or a book told us it was so. He wanted us to find things out for ourselves and be prepared to defend our views. This was fine at the university high school (Wisconsin High) I attended in Madison, where the focus was on teaching students to think for themselves. But in my junior year (1944) I attended the high school in Frederick, Maryland, where Dad was attached to Camp Detrick during World War II, and I was nearly thrown out of Frederick High for asking impertinent questions.

I've always been known as Buzz Baldwin because there were four Roberts in my first-grade class and Buzz served to identify me. My sister Frances, who was three years older, had trouble saying Brother when we were small. She first said Bruzzer, then Bruz, and finally Buzz. I've heard this is the usual origin of the nickname Buzz.

At the age of 13, in 1940, I caught tuberculosis and spent a year on complete bed rest because that was the only cure at a time when there were no drugs effective against tb. During high school, I was not allowed to play games or even run, because exercise could crack the calcified lesion in my lung before my body had reabsorbed the lesion. When the five years of no serious exercise were up and after one semester of university, I was drafted into the Army (January 1946) into the last group drafted at the end of World War II. Being in the Army made me appreciate university life considerably more when I returned in 1947 than when I started in the fall of 1945.

I began the University of Wisconsin with a major in prelaw because I liked to debate (I was on the college team), and I had only vague ideas about possible careers. I didn't care much for my prelaw required course, and after the first semester, I dropped prelaw and became a chemistry major while I cast around for a subject that might point my way to a career. In my junior year I took Physical Chemistry from Paul Bender, who was an inspiring lecturer, and I found it fascinating. Then Jack Williams, who was a friend of our family and a professor of physical chemistry, asked if I would do research in his lab that summer. I jumped at the chance. I worked

under the guidance of Mike Wales, a senior postdoc, who was testing a method for determining molecular weight distributions of flexible polymers from sedimentation equilibrium data.

I stayed on for senior thesis research in Williams' lab. Bob Alberty, who had been a PhD student with Williams, was an assistant professor in the Chemistry Department and his group held joint group meetings with Williams' group. The group meetings introduced me to a range of interesting problems. By spring I had an idea that produced both a thesis and a publication (9). It was a proposal for eliminating the effect of diffusion on the shape of a moving boundary in electrophoresis, based on extrapolation against the reciprocal of time to infinite time. This gave a sensitive test for differences in electrophoretic mobility among the protein molecules in the moving boundary. By reversing the electric field after a specified time, Bob Alberty had demonstrated earlier that mobility differences do exist in the antibody-containing $\gamma$-globulin fraction from blood plasma.

I was headed to Oxford University for my PhD (or, in Oxford parlance, D.Phil.) study. Mother asked me to apply for a Rhodes Scholarship and I thought I should humor her, because there was no risk I would get one. I was in no sense an athlete: my case of tb had not allowed me to play sports during high school. However, the Rhodes Selection Committee neglected to ask me about my (nonexistent) athletic interests. Our friends were as surprised as I was. When the selection results were announced, our next-door neighbor—instead of congratulating me—asked "but Buzz, how did YOU get a Rhodes Scholarship?"

Jack Williams solved the problem of what I should study at Oxford. He had earlier pointed out that, if I could remove the effect of diffusion from a moving boundary in electrophoresis, I could do the same thing for a sedimenting boundary in the ultracentrifuge. This might be more interesting because it would allow making a highly sensitive test for heterogeneity in mass among the protein molecules. Fred Sanger had not yet determined the first amino acid sequence of a protein, and the possibility that proteins might be heterogeneous in various properties, including size, was still being debated in 1950. Oxford had one of the world's eight oil-turbine ultracentrifuges and the University of Wisconsin had another, which was in Williams' charge; the Spinco ultracentrifuge was not yet in wide use. A.G. ("Sandy") Ogston, a brilliant biophysicist, was in charge of the Oxford oil-turbine instrument. Williams wrote to Ogston, who accepted me as a D.Phil. student in the Biochemistry Department. Although I would not get a PhD in physical chemistry as I first planned, I would nevertheless study under one of Oxford's most remarkable scientists.

## D.Phil. STUDY AT OXFORD

Even if they were graduate students, Rhodes Scholars were allowed to live in college, an experience otherwise reserved for undergraduates. Although Trinity College let me live in college, it always referred to me as an undergraduate. Life in the Biochemistry Department was quite different from that in the United States. There were no group meetings, seminars, or graduate courses. In fact, Sandy Ogston did not have a group. There were four or five research students, Sandy's own assistant, and sometimes a postdoc or other visitor, and Sandy. The Oxford system for producing scientists at that time was the sink-or-swim method. D.Phil. students were told to have their research projects planned before they chose their research advisors.

I soon found that the shapes of protein boundaries in the ultracentrifuge are affected by more than diffusion and possible heterogeneity in size. The dependence of the protein's sedimentation coefficient on its own concentration causes a major boundary sharpening effect. Eventually I found a solution to this problem (see below), and meanwhile I played with other ultracentrifuge problems,

such as finding a method for measuring the sedimentation coefficients of molecules too small to form moving boundaries (4). I was able to turn in a D.Phil. thesis within the three years allotted by the Rhodes Trust.

The Oxford D.Phil. exam, as practiced then, was unlike any other I have seen. Four years for turning in a thesis was the limit and three years was customary. The student paid 65 pounds from his own pocket (a significant sum in 1953), which went to the external examiner. At the exam the student and the examiner were closeted for 1 to 2 hours while the examiner probed weak points in the thesis. A verdict of "Failed" was not uncommon. My examiner was John Philpot, inventor of the schlieren optical system used to observe moving boundaries in sedimentation and electrophoresis.

Happily I passed and headed back to Wisconsin to be a postdoc in Jack Williams' lab. I would be in the Physical Chemistry division, although I had not yet taken any courses in Physical Chemistry beyond the undergraduate introductory course. I considered instead taking a second PhD in Physical Chemistry, but Williams counseled against it. "Study hard, Buzz, and it will work out all right."

## UNIVERSITY OF WISCONSIN

Dad used to say that the happiest days of a scientist's life are his postdoc days—no exams, no responsibilities, just research and learning new things. Certainly I have happy memories of my postdoctoral time. First, Lou Gosting gave me an education in how to do physicochemical research. Lou was one of Jack Williams' most cherished PhD students and he came back as an assistant professor of chemistry after being a postdoc first with Lewis Longsworth and then Lars Onsager. Bob Alberty had moved on to become Dean of the Graduate School. Other figures in the Williams-Gosting group were Hiroshi Fujita, a theoretical physical chemist on leave from Japan; Ken Van Holde, who had taken his PhD with Williams and then returned as a postdoc; and Peter Dunlop, Lou Gosting's student. During a postdoctoral year at Yale studying with Onsager, Lou read the papers of Willard Gibbs and adopted Gibbs' methods for solving problems in equilibrium thermodynamics. At Wisconsin Lou organized an informal seminar at which the group read and discussed Gibbs' papers, and this was a great experience. Tragically, Lou died at an early age a few years later.

At Oxford I had worked out a theory, based on moments of the boundary curve, for separating the contributions to boundary spreading from diffusion, heterogeneity in the sedimentation coefficient $s$, and the dependence of $s$ on protein concentration $c$ (5). On a trip home to Madison in 1952, I discussed the problem with Hiroshi Fujita, who then produced a solution to the differential equation for the case when the sedimenting molecule has a linear dependence of $s$ on $c$ (23). I used Fujita's equation to analyze the boundary spreading of bovine plasma albumin (6), whose diffusion coefficient was known accurately. The results gave the known diffusion coefficient of albumin and showed that the protein was homogeneous in $s$. Fujita's equation gave a more direct and simpler method of testing protein homogeneity than my own method of moments.

In 1955 I was unexpectedly offered an assistant professor position split between the departments of Biochemistry and Dairy and Food Industries. The University wanted someone to develop a research program focused on milk proteins, since Wisconsin was a dairy state. I started to study casein, which is central to the problem of understanding the physicochemical behavior of milk and how it is affected by $Ca^{2+}$. R.G. ("Gerry") Wake, a postdoc, joined me in this project.

Ken Van Holde and I discovered we both had begun work on the same problem, how to reach sedimentation equilibrium rapidly in order to measure protein molecular weights, and we agreed to collaborate. Ken was interested in measuring the transient period before sedimentation equilibrium is reached, and I

was interested in getting accurate molecular weights when the height of the solution column is short. A short column is needed because the time required to reach equilibrium is proportional to the square of the height of the solution column. Our method (84) was soon widely used, and sedimentation equilibrium became the method of choice for determining the molecular weights of proteins. With the older procedure of using a full ultracentrifuge cell, two weeks were needed to reach equilibrium with a small protein like ribonuclease A, whereas an overnight experiment sufficed in a short-column experiment.

In the summer of 1958, the National Institutes of Health (NIH) sponsored a month-long conference on "Biophysical Science" at the University of Colorado, Boulder. The organizers insisted that the participants stay for the entire conference. The talks were given by such senior figures in biophysics as Bob Alberty, Sir Bernard Katz, Norman Davidson, Paul Doty, Walter Kauzmann, John Kendrew, Cyrus Levinthal, and Bruno Zimm. Leo Szilard was there. The talks were published both as a book (58) and in *Reviews of Modern Physics*, intended to launch the new field of biophysics. A few young scientists, including David Davies, Matt Meselson, Tommy Thompson, Nacho Tinoco, and myself, were present. The scientific excitement was intense. Kendrew described his work on the myoglobin structure: Clearly a new era was opening in which protein structures would be determined by X-ray crystallography. Near the end of the meeting, Arthur Kornberg told me his department would move from Washington University to Stanford to form a new biochemistry department, and they planned to add a physical biochemist. I said I would be glad to come to St. Louis for an interview.

Earlier I had decided to take a six-month research leave in order to study with Kaj Linderstrøm-Lang at the Carlsberg Laboratory in Copenhagen. John Schellman and Bill Harrington convinced me that great things were happening in understanding how proteins fold and that the Carlsberg Laboratory was the center of the action. Before I arrived, Linderstrøm-Lang fell ill and was in the hospital. He was seriously diabetic and a minor operation had upset his insulin therapy. He died shortly after I returned to Wisconsin at the end of March 1959. Martin Ottesen, who became his successor, kept the laboratory running. I studied Linderstrøm-Lang's papers on hydrogen exchange with great interest and talked often with Sigurd Nielsen, who was developing infrared spectroscopy as a new method of measuring hydrogen-deuterium exchange in proteins.

Halfway through my stay in Copenhagen, I got a letter from Arthur Kornberg offering me a job. I had never seen Stanford and I don't think I even asked what my salary would be, I just said yes. I knew from the Boulder Biophysics Conference that molecular biology was going to be a major new field of science and I wanted to take part.

## STANFORD AND DNA

In June 1959, when I came to Stanford, I was more than ready to give up ultracentrifuge research. But how was I going to make a sudden transition to become a molecular biologist? In Copenhagen Bob Warner made the interesting suggestion that I should repeat the Meselson-Stahl experiment (55) with enzymatically synthesized DNA by using Arthur Kornberg's DNA polymerase I. Meselson and Stahl had found that hybrid $^{15}N$-$^{14}N$ DNA is formed by *Escherichia coli* after the synthetic growth medium containing $^{15}N$-$NH_4Cl$ is suddenly changed to $^{14}N$-$NH_4Cl$. Their results strongly suggested that DNA replication in *E. coli* is semiconservative, meaning that a newly synthesized DNA strand remains base-paired with the parental strand from which it is copied. However, they found they could separate the $^{15}N$ and $^{14}N$ subunits of the hybrid DNA by denaturing the DNA, and at that time this result was a puzzle (see below) that cast doubt on the semiconservative model, as they emphasized.

**poly dAT:** synthetic DNA copolymer with repeating sequence ATAT...

**BrU:** 5-bromouracil

In 1959 the Stanford Biochemistry Department had many surprising features, most of them carefully planned by Arthur Kornberg. He gave the same close attention to the design of the department that he gave to his own research. The most surprising features of the department were that general laboratory space was shared by everyone, so that students and postdocs from the various research groups were mixed together in common laboratories, and research grants were also shared. Each faculty member was supposed to bring in the amount of research funds he spent, but close accounting was not then required. The outstanding feature of the department was the strong spirit of cooperation among the faculty. They knew in detail each other's research and they shared rare chemicals, including enzymes, and all major instruments.

The focus of the department was, How do the genes reproduce themselves and direct the synthesis of proteins? Arthur Kornberg, Paul Berg, and Bob Lehman worked with enzymes, DNA and RNA polymerases, tRNA synthases, and nucleases, while Dave Hogness and Dale Kaiser worked with bacteriophage lambda as a model for how a small genome goes through its life cycle. Mel Cohn studied the synthesis of specific antibodies in single cells. When Mel Cohn left for the Salk Institute, George Stark and Lubert Stryer joined the department in 1963 and brought new strength in chemistry and biophysics.

A new postdoctoral fellow, Ross Inman, joined Gerry Wake and me in testing whether DNA polymerase I follows the semiconservative model of DNA replication. Ross had done his PhD at the University of Adelaide in DNA physical chemistry, which was lucky because my own knowledge was limited. We chose the synthetic DNA copolymer poly dAT, with the sequence ATAT...., as the primer-template for the polymerase. Ross set out to make physically formed hybrid molecules between dAT and its analog dABrU, in which 5-bromouracil replaces thymine. This was an audacious project in 1959 because the prevailing view then was that the complementary strands of the DNA helix remain entangled after denaturation, and we ought not to be able to make hybrid DNA. Light-scattering data said that the DNA molecular weight does not change after thermal denaturation (63). However, Meselson and Stahl had shown that the labeled subunits of their hybrid DNA do in fact come apart after denaturation (55), although the $^{15}N$ and $^{14}N$ subunits might be duplex DNA molecules containing two entangled strands, not the strands themselves.

Gerry Wake set out to make enzymatically synthesized poly dABrU and to study its relation to the poly dAT primer-template at early stages of reaction (85). Arthur Kornberg gave us DNA polymerase, poly dAT, and deoxynucleoside triphosphates without interfering in our research plans or wanting his name on our papers. Ross succeeded in making physically formed hybrid DNA, starting from poly dAT and poly dABrU (37), and by 1960 Julius Marmur and Paul Doty had shown that hybrid DNA could be made by annealing denatured strands of natural DNA. Ross found that in low salt, where the poly dAT melting transition is sharp, poly dABrU has a melting temperature ($T_m$) nine degrees higher than that of poly dAT (38). The hybrid DNA polymer made by annealing poly dAT with poly dABrU has hybrid base pairs that melt in an intermediate temperature zone (37). Using this analysis, Gerry Wake found that the newly synthesized dABrU remains base-paired to the parental dAT (85). Thus, pol I DNA replication is semiconservative in the Stent and Delbrück sense. Today it is known that RNA polymerase, like DNA polymerase, copies the DNA base sequence by complementary base-pairing but nevertheless leaves the DNA helix intact afterward. Thus, nature might have chosen a different mechanism for DNA replication.

An unexpected result from our dAT-dABrU studies was that these synthetic DNA helices melt from the helix ends (85) and not from "bubbles" (regions with open base pairs in the helix interior). Because newly

synthesized dABrU forms hybrid base pairs with the dAT primer-template, and these base pairs melt with a $T_m$ three degrees higher than that of dAT base pairs (37), the newly synthesized dABrU seals the ends of the poly dAT helices and the template dAT cannot melt until the higher $T_m$ of the hybrid is reached.

It was important to learn whether poly dAT has the canonical B-form helical structure. In 1961 I took a three-month research leave to go to David Davies' lab at NIH to answer this question. In addition to David, the small NIH Laboratory of Molecular Biology then also included Marty Gellert and Sid Bernhard. To my relief, the $Li^+$ salt of poly dAT does give the regular B-form fiber pattern (21), indicating that poly dAT exists in the B-form helix in aqueous solution. When $Na^+$ or $NH_4^+$ ion is present, fibers of poly dAT frequently give a new "D"-form structure (21), the structure of which is still not entirely understood.

I felt that the lessons learned from studying hybrid base pairs formed between poly dAT and poly dABrU ought to provide a basis for deciding whether *E. coli* DNA replication is indeed semiconservative. In 1962 the proposal of Cavalieri and coworkers (16) was still being debated; they argued that the $^{15}N$ and $^{14}N$ subunits of the Meselson-Stahl hybrid DNA are two-stranded DNA molecules whose entangled strands never come apart. Eric Shooter came from University College London on sabbatical and agreed to work on this project. The alkaline melting transition of DNA, caused by the ionization of thymine, changes sensitively when 5-bromouracil is substituted for thymine because of their different $pK_a$ values, and the alkaline transition can be monitored by banding DNA in a CsCl gradient. The hybrid density *E. coli* DNA, containing 5-bromouracil in one subunit, contained hybrid base pairs (10). This experiment basically ended the argument about whether DNA replication in *E. coli* is semiconservative.

By 1963 I was acutely conscious that I knew no genetics. The Stanford biochemistry seminars, which took place almost daily, often involved genetic arguments. François Jacob kindly allowed me to spend a sabbatical year (1963–1964) with him at the Institut Pasteur in Paris, where I would learn phage genetics. As it turned out, I had close interactions in Paris with Jacques Monod as well as with François Jacob, because Jacques was writing what would become the Monod-Wyman-Changeux (MWC) paper (56) and he wanted my perspective as a physical chemist. I read 6 drafts before the paper was published. I describe my conversations with Jacques in a book of recollections about him (8).

Jacques gave me the following advice, which I took to heart: "The way to do interesting research, Buzz, is to look for a paradox—but you must be sure it is a real paradox." The contrast between the Monod & Jacob style of research, based on deciding between competing models, and the Stanford style could not have been more striking. Arthur Kornberg once told me "It's all right to have these models, Buzz—everyone does—but don't talk about them."

The summer of 1963 was spent in Manfred Eigen's laboratory in Göttingen. Manfred was turning his attention from inorganic chemistry and proton-transfer reactions to molecular biology. Don Crothers, George Hess, Kasper Kirschner, and Gerhard Schwarz were all studying fast-reaction problems in molecular biology. My stay in Göttingen made a deep impression on me. A basic maxim of the Eigen laboratory was "always make kinetic measurements in the time range where you suspect a fast step might occur, and never rule the step out on the basis of slow kinetic results." I would remember that maxim later when we began to study protein folding.

John Schellman, like me, spent the winter of 1963–1964 on sabbatical in Paris, and I joined John and his wife, Charlotte, for a skiing holiday in Austria that Christmas. For years afterward, we took skiing vacations together and discussed science in the chairlift.

On my return to Stanford, a new student, Immo Scheffler, was waiting for me.

**Helix-coil transition:** Peptide helix forms by the mechanism proposed in the ZB theory and the LR theory. Folding begins with an unfavorable nucleation step and proceeds with a favorable propagation step

**Folding intermediate (I):** structural folding intermediates direct the folding process and are observable (populated) in favorable cases

(To learn who was in my lab in a given period, see the roster below.) Immo wanted to study the conformational changes that occur when a tRNA is loaded with its cognate amino acid. The stem and loop structures of tRNAs were just becoming known. I persuaded Immo to start by studying a simple model system for forming hairpin helices. A recent student, Elliot Elson, had learned how to prepare $d(TA)_n$ oligonucleotides of varying sizes ($n = 7$ to 19) by pancreatic DNase digestion of poly d(AT). The ATAT. . . oligonucleotides can form both monomeric hairpin and dimeric straight-chain helices.

Immo prepared substantial amounts of the d(TA) oligonucleotides and measured the heat-induced helix-coil transitions, which are reversible. He also measured the kinetics and equilibria of interconversion between the two types of helices, hairpins and dimers (70). Elliot supplied a theoretical analysis based on DNA statistical mechanics. Our aim had changed to deconstructing the physical behavior of poly dAT by studying oligonucleotide helices. Longer oligonucleotides (at least 16 TA pairs) could be circularized using DNA ligase. The circular molecules form double hairpin helices with two open loops and their helix-coil curves give detailed information about the DNA loop-weighting function for small loops (71). These studies provided a wealth of model compound data (7) for extracting basic parameters of the theory of DNA helix-coil transitions.

Our DNA work ended in 1983 with a study of closing short (~240 bp) double-stranded DNA molecules into covalently closed circles (78). My aim was to analyze very tight DNA loops similar to the ones found in nucleosomes. A student, David Shore, used *Eco* RI restriction endonuclease to cleave the DNA from phage ϕx174 into fragments of widely varying lengths and he then put *Eco*RI linkers on the DNA fragments. The *Eco*RI cohesive ends allow either the full-length DNA to cyclize or half molecules (with one blunt end) to form dimers. T4 DNA ligase was used to close covalently the molecules joined by cohesive ends. Comparison of the rates of covalent cyclization and dimer formation gives the cyclization probability, termed the j-factor. Large DNA fragments (>500 base pairs) gave j-factors expected from earlier analyses, but short DNA fragments (<500 base pairs) gave j-factors that varied erratically with DNA length (78). Various explanations seemed possible: For example, there might be special sequences that allow kinking of the DNA helix.

David gambled that the correct explanation is DNA twisting, i.e., that the DNA helix must be twisted to give an integral number of helix repeats when the cohesive ends are covalently joined. The size range for which DNA twisting should be energetically important was not known then. He made a series of DNA fragments with lengths that vary stepwise from 237 to 254 base pairs and found that their j-factors oscillate up and down with a period equal to the pitch of the DNA helix. We had found the DNA size range in which twisting is energetically important, and our results gave a firm value for the twisting force constant (78). Bruno Zimm's counsel was very helpful.

## PROTEIN FOLDING INTERMEDIATES AND PROLINE ISOMERIZATION

In 1968 Cy Levinthal gave a seminar at Stanford entitled "How to Fold Gracefully." Cy was using a computer to examine possible folding pathways of cytochrome (cyt) *c*, and at first he hoped to find the most plausible pathway by inspection. But then he made his famous estimate (46) of the number of possible backbone conformations. He concluded it would take longer than the age of the universe to fold by a random search of all possible backbone conformations. Paul Flory was sitting next to me at the seminar and he leaned over to whisper, "So there must be folding intermediates."

The prevailing view then was that protein folding is a two-state process and folding intermediates are not detectable. In 1970

I attended a Gordon Conference at which one speaker said that tRNA folding is a two-state process, according to the standard calorimetric criterion for detecting intermediates. Applying this calorimetric criterion to protein folding was a main reason why people believed that intermediates were not detectable. I knew that fast-reaction studies were revealing intermediates in the tRNA folding process and I resolved to try fast-reaction methods with protein folding.

Tian Tsong, a new postdoc, was interested in this problem, and we had the Eigen-DeMaeyer temperature (T)-jump apparatus, whose deadtime is a short 10 μs. We could look for unfolding intermediates formed between $10^{-5}$ and $10^{-1}$ s. Tian's 1971 T-jump experiments on RNase A unfolding showed that at least one intermediate is present (81). Tian next found a similar result with chymotrypsinogen A (80), which, like RNase A, was considered a paradigm for two-state protein folding. Our results clashed with those of Fritz Pohl, who had analyzed the unfolding kinetics of both proteins with a slow T-jump apparatus (61) and who found good agreement with the two-state model. But Fritz' instrument had a long deadtime (~1 s) and he missed our fast-unfolding reactions.

Atsushi Ikai and Charles Tanford independently found folding intermediates in 1971 in the unfolding/refolding reactions of cyt *c* (36). Their analysis was based on a quantitative treatment relating kinetic amplitudes to relaxation times. They concluded there is at least one off-pathway intermediate in refolding, but their test assumes that only one unfolded form is present (see below). Tanford's laboratory did not follow up on this discovery of folding intermediates, since they had started a new research program of characterizing membrane proteins.

My initial idea about searching for unfolding intermediates assumed that folding probably starts with a nucleation reaction, much like the Zimm-Bragg model for peptide helix formation. But we soon found a fast reaction in RNase A refolding that has a large amplitude (~20%). This is not expected if folding is rate-limited by a nucleation process. In 1973 Jean-Renaud Garel, a new postdoc, set out to learn if the fast reaction yields a partly folded intermediate. He decided to monitor refolding by the absorbance change that accompanies binding of a specific inhibitor, 2′CMP. Inhibitor binding to native RNase A (N) is a fast reaction, and we expected binding to serve as a probe for the formation of N. The startling result of Jean-Renaud's experiment was that both the fast- and slow-refolding reactions yield N (native form) (24), even when the initial protein is completely unfolded by denaturing it in 6 M guanidinium chloride (25). We concluded there must be two different forms of unfolded RNase A, a fast-folding form ($U_f$) and a slow-folding form ($U_s$), and we began work on finding out why this is so.

In 1975 John Brandts and coworkers (14) proposed a proline isomerization model for the $U_f = U_s$ reaction of RNase A. In this model, the direct unfolding reaction is $N \rightarrow U_f$, and $U_f$ has the correct proline isomers of N, while the slow $U_f = U_s$ reaction represents the slow *trans* → *cis* or *cis* → *trans* isomerization of the proline residues. Native RNase A has two *trans* and two *cis* proline residues. This model was very attractive, but in 1975, when site-directed mutagenesis still lay in the future, it was difficult to test. Built into the model was the additional assumption that there are no populated structural folding intermediates. This assumption, which proved to be wrong, complicated matters because it meant that the kinetic properties of the $U_s \rightarrow N$ refolding reaction should be strictly predictable from those of proline isomerization. Barry Nall's (57) data on the refolding kinetics of RNase A gave a much more complex picture.

We concluded that the properties of the $U_f = U_s$ reaction must be determined in experiments made when RNase A remains unfolded. A key property of proline isomerization is catalysis by acid, but only by very strong acid. In 1978 Franz Schmid, a new postdoc, tested for this property by making refolding

**T-jump:** temperature jump

**RNase A:** ribonuclease A from bovine pancreas

**cyt *c*:** cytochrome *c* (oxidized form) from horse heart

**N:** native form

**U:** unfolded form

**$U_f$:** fast-folding form of an unfolded protein

**$U_s$:** slow-folding form of an unfolded protein

**Alternative folding pathways:** Folding occurs by the fastest route when alternative pathways are available

**S-peptide:** 19-residue peptides from the N terminus of RNase A

assays for the $[U_f]/[U_s]$ ratio at different times after adding $HClO_4$ (72). His results demonstrated unequivocally that the $U_f = U_s$ reaction is indeed proline isomerization.

Franz's results also suggested an experiment for directly testing whether structural intermediates can be populated in refolding. At 10°C the kinetics of the $U_s \rightarrow N$ refolding reaction are quite slow. Usually it is possible to speed up the formation of folded structure by adding a stabilizing salt such as $(NH_4)_2SO_4$, and this salt has only a small effect on the rate of proline isomerization in model compounds. When the nonnative proline isomer of a $U_s$ species reverts to its native isomer during refolding, the conversion can be observed by a coupled unfolding-refolding assay (20). These experiments demonstrate that when one major slow-folding species of unfolded RNase A refolds, it forms a highly folded, native-like intermediate ($I_N$) before the nonnative proline isomer reverts to the native one. The intermediate $I_N$ binds 2′CMP (20), a specific inhibitor of native RNase A.

These experiments proved that a structural folding intermediate can be populated during refolding (20). They also showed that folding can proceed via alternative pathways: either by the direct $U_f \rightarrow N$ pathway or by the $U_s \rightarrow I_N \rightarrow N$ pathway, in which folding begins in a slow-folding species. These experiments also support Levinthal's argument (47) that folding proceeds by the fastest route if alternative pathways are available. See the modern review of alternative folding pathways by Jayant Udgaonkar (82) in this volume.

In 1976 a student, Paul Hagerman, chose a different route for analyzing the unfolding/refolding reactions of RNase A (27). His approach resembled the quantitative analysis of Ikai and Tanford but differed in specifying that there are two unfolded forms in addition to the native protein N. Paul found an unfolding intermediate (I), presumably the same one seen earlier by Tian Tsong in unfolding at pH 1.3, and he extended his three-state analysis to include it. Paul measured the unfolding and refolding kinetics of RNase A in the same final conditions at temperatures throughout the thermal unfolding transition at pH 3, and he compared the predicted kinetic amplitudes with the observed values (27). His success in fitting the observed kinetic amplitudes was impressive.

Paul found that the kinetic amplitude of species I increases strikingly with temperature and that it has the tyrosine absorbance and enthalpy of an unfolded species. He next built a stopped-flow mix-and-quench apparatus in order to populate species I by unfolding and then measure its refolding kinetics. Franz Schmid did the refolding experiment after Paul graduated; Franz found that species I does refold more rapidly than $U_f$ (28). Later, in 1994 the use of site-directed mutants in Harold Scheraga's lab showed that when N unfolds, the direct product is not $U_f$ but instead a very fast-folding form, $U_{vf}$ (30). It is only a minor species (~6%) in RNase A when unfolding is at equilibrium; almost certainly $U_{vf}$ is the same as Paul's species I.

Arlene Blum, a new postdoc, outlined a plan for monitoring the refolding of RNase A by 1D $^1$H-NMR in real time. She planned to observe the chemical shifts of the histidine ring resonances of the four His residues. I pointed out that conditions where folding is slow enough to observe by real-time NMR are exactly the conditions where folding intermediates are not populated, but it's not easy to discourage Arlene when she has a plan. We agreed on a down jump in temperature (from 45°C to 10°C) at pH 2. I was astonished and Arlene was triumphant when in 1978 the chemical shift of His12 indicated the transient formation of a folding intermediate (13). Then I remembered the 1971 paper of Brown & Klee (15), which argues that peptides containing residues 3–12 of RNase A (the site of the N-terminal helix) show partial helix formation at temperatures near 0°C. Sure enough, the chemical shift of His12 in S-peptide matches the chemical shift of the transient folding intermediate of RNase A (13). So, Arlene's experiment

strongly supported the claim of Brown & Klee (15) that partial helix formation can be observed in S-peptide. This was an exciting result for us because it opened a new research area, the mechanism of helix formation in peptides.

## MECHANISM OF PEPTIDE HELIX FORMATION

In 1979 peptide helices were supposed to be unstable in water, especially because water can make H-bonds to the peptide CO and NH groups. It was possible to study peptide helices in favorable organic solvents, such as 2-chloroethanol or 2,2,2-trifluoroethanol, but not in water. Or one could use the hydroxyalkyl-L-glutamine amino acids, discovered by Arieh Berger and coworkers (51), which do form the α-helix in water. Berger attributed this ability to a hydrophobic interaction made by the hydroxyalkyl side chain with the helix backbone (51). Harold Scheraga used helical peptides formed by these nonnatural amino acids as hosts for determining "host-guest" helix propensities of the 20 natural amino acids, and his results indicated that the 20 natural amino acids should not form the α-helix in water. Papers both from Scheraga's lab on peptide fragments from myoglobin and from Chris Anfinsen's lab on a peptide fragment from staphylococcal nuclease failed to find any helix formation.

When we began work in 1979, it was difficult to get funding for any peptide helix studies and a close friend warned me against starting on this problem. But both Andrzej Bierzynski, a new postdoc, and Peter Kim, a new student, were enthusiastic about tackling it. Two results indicated we were on the right track (12). First, side chain resonances of several residues—not just His12—change in helix-forming conditions. Second, we found that the curve of helix content versus pH (measured by circular dichroism) is bell-shaped and indicates that the charged side chains of His12 and also a Glu residue (C-peptide has both Glu2 and Glu9) are needed for helix formation. Thus, one or more helix-stabilizing side chain interactions should explain the unexpected stability of this peptide helix. With the help of John Stewart and Eunice York, we set out to characterize the helix-stabilizing interactions by making and studying new peptides.

Peter found by NMR that there is a helix-stop signal in S-peptide (45): the helix ends at His12 in S-peptide just as it does in RNase A. The same result was found independently in Professor Manuel Rico's laboratory (Instituto de Estructura de la Materia) in Madrid. Finding a helix-stop signal was startling because a calculation made with host-guest helix propensities indicated that the helix should propagate for many residues after side chain interactions nucleate a helix. The discovery of the stop signal showed we had to change our thinking about helix formation in peptides derived from proteins.

The picture that emerged from these studies (76) is the familiar one today, that this helix is an autonomous folding unit and the mechanisms of forming local structures in proteins can be analyzed in peptide studies. The same helix-stabilizing interactions, a $Glu2^{-}...Arg10^{+}$ salt bridge and a $Phe8...His12^{+}$ amino-aromatic interaction, that were found in the C-peptide helix can also be seen in the 1971 X-ray structure of RNase S (76).

Finding a helix that is an autonomous folding unit gave strong support to the hierarchic model of folding. The concept of hierarchic folding had been discussed earlier in Oleg Ptitsyn's 1973 model of the folding process (62), and in the 1976 diffusion-collision model of Martin Karplus and David Weaver (41), and in the 1979 analysis by George Rose (69) of domains and subdomains in proteins. But in 1987 belief in two-state folding was still very strong and many workers believed in nucleation-limited folding. Our finding, first of a helix-stop signal and then of an autonomously folding helix, influenced the protein folding community to seriously consider hierarchic models of folding. See the modern

**C-peptide:** 13-residue peptide from the N terminus of RNase A

**Hierarchic folding:** begins with forming local structures, proceeds with forming molten globule intermediates (with defined secondary structures), and forms the native structure by locking side chains together

review of protein folding mechanisms by Ken Dill and colleagues (22) in this volume.

**Charge-helix dipole interaction:** A charged group close to one end of a peptide helix interacts with the helix macrodipole and affects helix stability, either favorably or unfavorably

The search for helix-stabilizing interactions in the C-peptide helix had an unexpected by-product. Kevin Shoemaker found in 1987 that the charge-helix dipole interaction makes an important contribution to helix stability in certain variants of C-peptide (77). This interaction had long been suspected of being an important factor affecting helix stability because Glu and Asp occur preferentially near the N termini of protein helices, whereas Lys, Arg, and His occur preferentially at the C termini. Evidence favoring a charge-helix dipole interaction in peptides had been found in 1982 by Prof. Tatsuo Ooi at Kyoto University (35), but he disclaimed this interpretation until 1989, when he and his coworkers analyzed a second, complementary set of peptides (79).

In 1989 Susan Marqusee found that alanine peptides can form a stable helix in water (54). Further work showed that of the 20 natural amino acids, only alanine has this property (18, 67). Her result changed our way of thinking about peptide helices and peptide H-bonds. The helix backbone itself must be stable in water because alanine has only a methyl group for a side chain. Consequently, the energetic importance of peptide H-bonds in protein folding must have been underestimated in earlier studies.

It was important to confirm that peptide H-bonds are helix-stabilizing by accurately determining the enthalpy of alanine helix formation. When Der-Hang Chin, a visiting professor from National Changhua University in Taiwan, came to my lab, we arranged a collaboration with George Makhatadze to do this. In 2002 his titration calorimetry experiments gave the enthalpy change as $-0.9 \pm 0.1$ kcal per mol per residue (50) and this value was confirmed by Andrzej Bierzynski from the Polish Academy of Sciences in Warsaw (26), who had worked out this peptide system for initiating helix formation by binding $La^{3+}$.

After Susan Marqusee's discovery of alanine helix formation in water, it became possible to analyze whether the alanine peptide system obeys helix-coil theory. The theory had been worked out by Zimm and Bragg (ZB theory) in 1959 (86) and by Lifson and Roig (LR theory) (48) in 1961. In 1991 Marty Scholtz found very good agreement with helix-coil theory when he fitted helix-coil transition curves for a set of repeating-sequence peptide helices; their lengths varied from 14 to 50 residues (74). He found the helix nucleation constant ($\sigma = 0.003$) to be almost $10^3$ times smaller than the helix propagation constant.

The ZB theory applies strictly only to homopeptides because the helix propagation parameter $s$ refers to the peptide H-bond and not to the amino acid residue. We wanted to study alanine-based peptides that contain a single glycine residue at different positions in the helix, and we expected that a theory adapted to heteropeptides would be required. On John Schellman's advice, we turned to the LR theory because its propagation parameter $w$ refers to the amino acid residue. In 1991 John worked out a computer adaptation of the LR theory for heteropeptides, and we applied it to Avi Chakrabartty's data for alanine peptides that contain a single glycine residue (19). The results demonstrated a large difference (~100-fold) between the helix propensities of Gly and Ala and also showed that the ends of these peptide helices are strongly frayed.

In 1992 Carol Rohl developed a second method for determining the helix-coil parameters for alanine peptide helices, based on NMR determination of the kinetics of $^1H$-$^2H$ exchange (68). Her method gave essentially the same numbers as measuring helix unfolding curves by circular dichroism, and the agreement increased our confidence in both methods. In 1994 Avi Chakrabartty and Tanja Kortemme made 58 peptides and used them to determine the helix propensities of all 20 amino acids (18). When the alanine peptide helix is used as a host for determining helix propensities, it gives absolute values, whereas other peptide hosts (with the exception of the hydroxyalkyl-L-glutamine peptides) give only propensities relative to a reference amino acid.

In 1993 Avi Chakrabartty and Andy Doig made the surprising finding that the N-terminal interface residue of a peptide helix has a special energetic role in peptide helix formation (17). The role of the N-terminal interface residue is parallel to that of the N-cap residue in protein helices, as judged by the relative frequencies of different amino acids at the N-cap position in protein helices. Later, in 1996, Carol and Avi redetermined the helix propensities of all the amino acids after taking account of the N-cap effect, and they also gave values for the N-cap propensities. In addition, they determined these parameters for a helix-stabilizing solvent, 40% 2,2,2-trifluoroethanol (67).

To find out why the helix propensities measured in an alanine peptide host are different from the host-guest propensities measured in a hydroxyalkyl-L-glutamine host, S. Padmanabhan (Padhu) studied the helix-stabilizing properties of hydroxybutyl-L-glutamine in an alanine peptide host. He found there is an energetic penalty for replacing a hydroxyalkyl residue with a natural amino acid (60) because, as deduced earlier by Berger and colleagues (51), the hydroxyalkyl moiety makes a helix-stabilizing interaction with the helix backbone.

The alanine peptide host is well-suited to determining the energetics of helix-stabilizing interactions such as salt bridges and pairwise nonpolar interactions. Marty Scholtz made a completely uncharged alanine peptide helix by using glutamine as the solubilizing residue (75). Bea Huyghues-Despointes then used this system to measure the energetics of the charge-helix dipole interaction when a single charged $Asp^-$ residue is placed at different positions in a neutral helix, and we compared the results with a simple theory (34). Bea also measured the energetics of the (i,i+4) H-bond formed between the side chain amide of Gln (residue i) and the $COO^-$ group of Asp (residue i+4) (33). Bea analyzed the stereochemistry of the H-bond by NMR. Padhu measured several pairwise nonpolar side chain interactions and found that the interaction is stronger when one side chain, such as Tyr, Phe, or Val, has limited flexibility (59).

**N-cap:** interface residue at N-terminal end of an $\alpha$-helix

## MECHANISMS OF FORMING FOLDING INTERMEDIATES

For a long time I wanted to develop a pulse labeling method, based on $^1H$-$^2H$ exchange and $^1H$-NMR, for determining the secondary structures and stabilities of transient folding intermediates. Linderstrøm-Lang foresaw in the 1950s that hydrogen exchange would someday be used to characterize protein reactions that involve changes in conformation. In 1979 Franz Schmid tested the concept by looking for an early folding intermediate of RNase A whose peptide NH protons are protected from $^3H$-$^1H$ exchange (73). A competition between folding and exchange-out of $^3H$ label is started at the time folding begins: Protection against exchange-out occurs only if a folding intermediate is formed rapidly. Franz found the important result that peptide NH protons are protected from exchange early in folding, before the tertiary structure is formed (73).

The NMR–pulse labeling method was applied to the folding of RNase A in 1988 (83) by Jayant Udgaonkar, a postdoc. Only a limited set of the NH protons in the β-sheet of RNase A was available for study then, and Jayant found that all of them are rapidly protected from exchange early during refolding. Andy Robertson made resonance assignments of RNase A (64) in our lab while working in collaboration with Harold Scheraga's group, after we discovered they were also making the assignments.

Our 1988 paper on results of the pulse labeling method was published side by side with a corresponding study of cyt *c* by Heinrich Roder, Gülnur Elöve, and Walter Englander (65), and this happened as follows. In 1979 I attended a small symposium on protein folding organized by Kurt Wüthrich at the ETH, where I discussed with Kurt's student, Heinrich Roder, my plan for determining

**apoMb:** apomyoglobin from sperm whale

structures of transient folding intermediates. Heinrich tried the method on BPTI but failed to get clear evidence for any folding intermediate (66). We agreed to see if he could come to my lab for postdoctoral work, but when the time came, the NIH NMR facility at Stanford had been shut down and I temporarily lost access to an NMR instrument. Heinrich went instead to Walter Englander's laboratory to study the folding of cyt *c*.

When the pulse labeling results with RNase A proved to be unexpectedly complex, our lab began in 1990 to study a pH 4 equilibrium folding intermediate of apoMb. The intermediate had been discovered in Peter Privalov's laboratory. Fred Hughson, a student, took up the problem with great verve. The NMR resonance assignments for Mb and apoMb had not yet been published but we knew they were being made in Peter Wright's laboratory, and Peter agreed to collaborate with us. The structural results were fascinating (32): A folded subdomain of Mb is present that contains both the A helix (from the N-terminal end of Mb) and the G and H helices (from the C-terminal end). This result clearly gave strong support to the hierarchic model of folding, but because the A and G,H helices come from opposite ends of the polypeptide chain, folding of the subdomain cannot be the assembly of one contiguous region.

Although the A, G, and H helices are folded in the intermediate, they are not very stable. The protection factors of the peptide NH protons are in the range of 10–100 (32), whereas they can be greater than $10^8$ in native Mb. In 1995 Stewart Loh found a second, more stable form of the apoMb intermediate by adding trichloroacetate (49), which gives a structure in which the B helix is partly stable in addition to more stable A, G, and H helices (49).

Both apoMb intermediates are typical molten globule intermediates in which tertiary structure is ill-defined, although secondary structure is fairly well-defined, as judged by hydrogen exchange and circular dichroism. Our studies of the folding mechanism focused chiefly on the first intermediate.

We sought to answer two questions: Which interactions stabilize it, and how does it fold? In 1991 Fred Hughson and Doug Barrick used site-directed mutagenesis to test for stabilizing interactions and they found that the intermediate is loosely packed. Not only do small-to-large mutations of nonpolar side chains fail to destabilize the intermediate as they destabilize N, they actually tend to stabilize the intermediate, suggesting that the hydrophobic interaction is a main source of its stability (31). In 1993 Doug made a thorough survey of the stability of the intermediate as a function of pH and [urea] (11). He found that the simple three-state (U = I = N) model fits the unfolding curves of I and N (measured by circular dichroism) essentially quantitatively from pH 2 to pH 8 and from 0 M to 7.6 M urea.

A prime objective was to find whether the intermediate folds cooperatively. Two 1996 studies, by Michael Kay (42) and by Marc Jamin (39), found that it does. Their results indicate that packing is strongly involved in the folding process, as one might guess from the fact that the A and G,H helices come from opposite ends of the chain. Michael analyzed the equilibrium unfolding curve by the classical procedure of asking whether the unfolding curves measured by probes of secondary structure (circular dichroism) and tertiary structure (fluorescence of partly buried Trp residues) are superimposable, and he found they are (42). Marc found that the kinetics of unfolding and refolding fit the two-state model (39). Later he also found an additional, more structured intermediate at a further stage in folding (40). In 1997 Yongzhang Luo made single mutations to glycine or proline at solvent-exposed positions in either the A or G helix; Gly and Pro are strongly helix-destabilizing residues. The question is: Does the mutation destabilize only the single helix that contains it, or are all three helices destabilized? Yongzhang found that all three helices fold together cooperatively (53), although

decreasing the stability of the intermediate can cause a loss of cooperativity (53).

By 1995, the pulse labeling method had been applied to characterizing refolding intermediates in several laboratories. Thomas Kiefhaber, a new postdoc, wanted to find out if hydrogen exchange could detect any unfolding intermediates in proteins whose equilibrium unfolding curves fit the two-state model. The prevailing view then was that there are no detectable structural intermediates in unfolding, although intermediates in refolding experiments are often observable. We chose RNase A unfolding at pH 8 and pH 9, conditions in which hydrogen exchange follows the EX1 mechanism and the exchange rate gives the rate of unfolding. Thomas obtained a simple and clear answer (43). The entire β-sheet of RNase A opens and becomes accessible to exchange in a single cooperative step, and the unfolding rate measured by hydrogen exchange agrees with the rate measured by optical probes of tertiary structure when proline isomerization is taken into account (43). His results fitted the prevailing view that there are no detectable unfolding intermediates in this class of proteins and gave a much more detailed picture of the unfolding process.

Thomas then decided to use 1D $^{1}$H-NMR in real time to determine if side chains become free to rotate before the rate-limiting step in unfolding is reached. His hydrogen exchange results (43) had shown that water does not enter the β-sheet before the rate-limiting step in unfolding is reached, but there might be a dry molten globule intermediate. Using the resolved methyl resonance line of Val63, he found that its side chain does become free to rotate before the rate-limiting step in unfolding is reached, and so there does appear to be a dry molten globule type of unfolding intermediate (44). Later that same year (1995), an elegant study by Hoeltzli & Frieden (29) of the unfolding of *E. coli* dihydrofolate reductase, using $^{19}$F-labeled Trp residues, confirmed the existence of this type of unfolding intermediate.

## PEPTIDE SOLVATION

When I retired and closed my lab in 1998, it wasn't clear what I would do next. Rainer Jaenicke gave me the sensible advice "just wait, something will turn up." Meanwhile there were several papers waiting to be written. In one of them, Peizhi (Peter) Luo had measured the temperature dependences of the helix propensities of five nonpolar amino acids (52). He found that the helix propensities are substantially enthalpy driven, in contrast to the prevailing view that helix propensity is determined by the loss in side chain conformational entropy that occurs on helix formation. When the paper came out, I got an email from Franc Avbelj in Ljubljana, Slovenia, saying "Look, I predicted this." I read his paper carefully, and a correspondence followed that led to a collaboration.

**Electrostatic solvation:** the interaction between water and the peptide NH and CO groups analyzed by treating water as a continuum solvent

**ESF:** electrostatic solvation free energy

Franc's thesis is that the backbone propensities of the amino acids are determined by backbone electrostatics: both by the dipole-dipole interactions made between the NH and CO groups in the peptide backbone and by the solvation of these dipolar groups. The electrostatic energy and the electrostatic solvation are calculated by standard electrostatic algorithms and both are equally important. The algorithms give finite difference solutions to the Poisson-Boltzmann equation, and they give the electrostatic solvation free energy (ESF) as well as the electrostatic energy.

ESF calculations show that water still interacts substantially (–2.5 kcal mol$^{-1}$) with a peptide group after peptide H-bonds are made, either in an α-helix (3) or in a β-hairpin (1). This favorable interaction of H-bonded peptide groups with water helps to explain the surprising stability of molten globule folding intermediates. Evidence that ESF is important in determining peptide backbone conformation comes from a linear correlation between ESF and β-structure propensities (1). The important role of side chains in determining ESF, via limiting the access of solvent to peptide NH groups, is shown by the

linear correlation between ESF and the blocking effect of a nonpolar side chain on the hydrogen exchange rate of an adjacent peptide NH (2). ESF data explain the neighboring residue effect: One group of amino acids (the "L" group, FHITVWY) has large effects on the backbone conformations of neighboring residues in short peptides, compared with other amino acids (2). The neighboring residue effect was observed in Chris Dobson's lab by NMR measurements of $^3J_{HN\alpha}$ coupling constants.

## ACKNOWLEDGMENTS

I try to make clear that students and postdocs contributed in a major way to planning the research described here as well as carrying it out. I'm sorry there isn't space to describe everyone's work.

I discuss our work on the mechanism of protein folding, a subject that is being actively investigated today, and I emphasize that our results support the hierarchic mechanism, as defined here. I say little about the important contributions made by others because this is not a review, and I apologize to the other workers.

In addition to students and postdocs, friends have had a major influence on my research and I wish to acknowledge them: in particular, Julius Adler, David Davies, Ken Dill, Walter Englander, Paul Flory, Marty Gellert, Bob Harris, Jan Hermans, Jr., Neville Kallenbach, Kasper Kirschner, B.-K. Lee, George Makhatadze, Fred Richards, George Rose, John Schellman, Peter von Hippel, and Bruno Zimm.

I thank Anne Baldwin and George Rose for their suggestions concerning this account.

When thorny issues connected with my research work arose, I regularly turned to my wife, Anne, and she always gave the right advice. I cannot thank her enough.

I gratefully acknowledge funding by the Rhodes Trust, the United States National Institutes of Health, and the National Science Foundation.

## ROSTER OF STUDENTS AND POSTDOCTORAL FELLOWS

**1960s**

**PhD students**: James J. Butzow, Robert C. Costello, Elliot L. Elson, Frank E. LaBar, Immo E. Scheffler

**Postdoctoral fellows and visitors**: Anne Bodner, Peter J. Dunlop, Magdalena Fikus, Alex Fritsch, Martin Gellert, Michel E. Goldberg, Ross B. Inman, Tadao Kotaka, Jean-Bernard LePecq, Peter M. McPhie, M. Thomas Record Jr., Eric M. Shooter, Hanns-Christof Spatz, Rolf Sternglanz, R.G. Wake

**1970s**

**PhD students**: Scott W. Emmons, Paul J. Hagerman, Barry T. Nall, J. Andrew Ridge, David Shore, Jonathan Widom

**Postdoctoral fellows and visitors**: Alan J.D. Bellett, Arlene D. Blum, Heinz G. Busse, Claus Christiansen, Kem H. Cook, Jean-Renaud Garel, Martin Gellert, Magali Jullien, Alexander M. Labhardt, C. Robert Matthews, Gabriele Milanesi, Franz X. Schmid, Alan A. Schreier, Charles Seiter, Grayson Snyder, John O. Thomas, Tian Y. Tsong

**1980s**

**PhD students**: Robert Fairman, Frederick M. Hughson, Peter S. Kim, Susan Marqusee, David A. Schultz, Kevin R. Shoemaker

**Postdoctoral fellows and visitors**: Andrzej Bierzynski, David N. Brems, Jannette Carey, George O. Gbenle, Homme Hellinga, Neville R. Kallenbach, Kunihiro Kuwajima, David

Loftus, Jörg Langowski, Colin Mitchinson, John J. Osterhout Jr., Andrew D. Robertson, Karen G. Strehlow, Jayant B. Udgaonkar

**1990s**

**PhD students:** Katherine M. Armstrong, Doug Barrick, Beatrice M.P. Huyghues-Despointes, Michael S. Kay, Douglas V. Laurents, Carol A. Rohl

**Postdoctoral fellows and visitors:** Avijit Chakrabartty, Andrew J. Doig, Wayne Fiori, Bernhard Geierstanger, Jonathan M. Goldberg, Marc Jamin, Thomas Kiefhaber, Tanja Kortemme, Stewart N. Loh, Peizhi (Peter) Luo, Yongzhang Luo, Stephen L. Mayo, S. Padmanabhan, Carlos H.I. Ramos, J. Martin Scholtz

**2000s**

**Visitor:** Der-Hang Chin



## DISCLOSURE STATEMENT

The author is not aware of any biases that might be perceived as affecting the objectivity of this review.

## LITERATURE CITED

1. Avbelj F, Baldwin RL. 2002. Role of backbone solvation in determining thermodynamic β-propensities of the amino acids. *Proc. Natl. Acad. Sci. USA* 99:1309–13
2. Avbelj F, Baldwin RL. 2004. Origin of the neighboring residue effect on peptide backbone conformation. *Proc. Natl. Acad. Sci. USA* 101:10967–72
3. Avbelj F, Luo P, Baldwin RL. 2000. Energetics of the interaction between water and the helical peptide group and its role in determining helix propensities. *Proc. Natl. Acad. Sci. USA* 97:10786–91
4. Baldwin RL. 1953. Sedimentation coefficients of small molecules: methods of measurement based on the refractive-index gradient curve. The sedimentation coefficient of polyglucose A. *Biochem. J.* 55:644–48
5. Baldwin RL. 1957. Boundary spreading in sedimentation velocity experiments. 4. Measurement of the standard deviation of a sedimentation coefficient distribution: application to bovine albumin and β-lactoglobulin. *Biochem. J.* 65:490–502
6. Baldwin RL. 1957. Boundary spreading in sedimentation velocity experiments. 5. Measurement of the diffusion coefficient of bovine albumin by Fujita's equation. *Biochem. J.* 65:503–12
7. Baldwin RL. 1971. Experimental tests of the theory of deoxyribonucleic acid melting with d(T-A) oligomers. *Acc. Chem. Res.* 4:265–72
8. Baldwin RL. 1979. Discussions about proteins. In *Origins of Molecular Biology*, ed. A Lwoff, A Ullmann, pp. 203–207. New York: Academic
9. Baldwin RL, Laughton PM, Alberty RA. 1951. Homogeneity and the electrophoretic behavior of some proteins. III. A general method for the determination of mobility distributions. *J. Phys. Colloid Chem.* 55:111–25
10. Baldwin RL, Shooter EM. 1963. The alkaline transition of BU-containing DNA and its bearing on the replication of DNA. *J. Mol. Biol.* 7:511–26
11. Barrick D, Baldwin RL. 1993. Three-state analysis of sperm whale apomyoglobin folding. *Biochemistry* 32:3790–96

12. Bierzynski A, Kim PS, Baldwin RL. 1982. A salt bridge stabilizes the helix formed by the isolated C-peptide of RNase A. *Proc. Natl. Acad. Sci. USA* 79:2470–74
13. Blum AD, Smallcombe SH, Baldwin RL. 1978. Nuclear magnetic resonance evidence for a structural intermediate at an early stage in the refolding of ribonuclease A. *J. Mol. Biol.* 118:305–16
14. Brandts JF, Halvorson HR, Brennan M. 1975. Consideration of the possibility that the slow step in protein denaturation reactions is due to cis-trans isomerism of proline residues. *Biochemistry* 14:4953–63
15. Brown JE, Klee WA. 1971. Helix-coil transition of the isolated amino terminus of ribonuclease. *Biochemistry* 10:470–76
16. Cavalieri LF, Rosenberg BH, Deutsch JF. 1959. The subunit of deoxyribonucleic acid. *Biochem. Biophys. Res. Commun.* 1:124–28
17. Chakrabartty A, Doig AJ, Baldwin RL. 1993. Helix capping propensities in peptides parallel those in proteins. *Proc. Natl. Acad. Sci. USA* 90:11332–36
18. Chakrabartty A, Kortemme T, Baldwin RL. 1994. Helix propensities of the amino acids measured in alanine-based peptides without helix-stabilizing side-chain interactions. *Protein Sci.* 3:843–52
19. Chakrabartty A, Schellman JA, Baldwin RL. 1991. Large differences in the helix propensities of alanine and glycine. *Nature* 351:586–88
20. Cook KH, Schmid FX, Baldwin RL. 1979. Role of proline isomerization in folding of ribonuclease A at low temperatures. *Proc. Natl. Acad. Sci. USA* 76:6157–61
21. Davies DR, Baldwin RL. 1963. X-ray studies of two synthetic DNA copolymers. *J. Mol. Biol.* 6:251–55
22. Dill KA, Ozkan SB, Shell MS, Weikl TR. 2008. The protein folding problem. *Annu. Rev. Biophys.* 37:In press
23. Fujita H. 1956. Effects of a concentration dependence of the sedimentation coefficient in velocity ultracentrifugation. *J. Chem. Phys.* 24:1084–90
24. Garel JR, Baldwin RL. 1973. Both the fast and slow refolding reactions of ribonuclease A yield native enzyme. *Proc. Natl. Acad. Sci. USA* 70:3347–51
25. Garel JR, Nall BT, Baldwin RL. 1976. Guanidine-unfolded state of ribonuclease A contains both fast- and slow-refolding species. *Proc. Natl. Acad. Sci. USA* 73:1853–57
26. Goch G, Maciejczyk M, Oleszczuk M, Stachowiak D, Malicka J, Bierzynski A. 1993. Experimental investigation of initial steps of helix propagation in model peptides. *Biochemistry* 42:6840–47
27. Hagerman PJ, Baldwin RL. 1976. A quantitative treatment of the folding transition of ribonuclease A. *Biochemistry* 15:1462–73
28. Hagerman PJ, Schmid FX, Baldwin RL. 1979. Refolding behavior of a kinetic intermediate observed in low pH unfolding of ribonuclease A. *Biochemistry* 18:293–97
29. Hoeltzli S, Frieden C. 1995. Stopped-flow NMR spectroscopy: real-time unfolding studies of 6–$^{19}$F-tryptophan-labeled *Escherichia coli* dihydrofolate reductase. *Proc. Natl. Acad. Sci. USA* 92:9318–22
30. Houry WA, Rothwarf DM, Scheraga HA. 1994. A very fast phase in the refolding of disulfide-intact ribonuclease A: implications for the refolding and unfolding pathway. *Biochemistry* 33:2516–30
31. Hughson FM, Barrick D, Baldwin RL. 1991. Probing the stability of a partly folded apomyoglobin intermediate by site-directed mutagenesis. *Biochemistry* 30:4113–18
32. Hughson FM, Wright PE, Baldwin RL. 1990. Structural characterization of a partly folded apomyoglobin intermediate. *Science* 249:1544–48

33. Huyghues-Despointes BMP, Klingler TM, Baldwin RL. 1995. Measuring the strength of side-chain hydrogen bonds in peptide helices: the Gln•Asp (i,i+4) interaction. *Biochemistry* 34:13267–71
34. Huyghues-Despointes BMP, Scholtz JM, Baldwin RL. 1993. Effect of a single aspartate on helix stability at different positions in a neutral alanine-based peptide. *Protein Sci.* 2:1604–11
35. Ihara S, Ooi T, Takahashi S. 1982. Effects of salts on the nonequivalent stability of the α-helices of isomeric block copolypeptides. *Biopolymers* 21:131–45
36. Ikai A, Tanford C. 1971. Kinetic evidence for incorrectly folded intermediate states in the refolding of denatured proteins. *Nature* 230:100–2
37. Inman RB, Baldwin RL. 1962. Formation of hybrid molecules from two alternating DNA copolymers. *J. Mol. Biol.* 5:185–200
38. Inman RB, Baldwin RL. 1962. Helix-random coil transitions in synthetic DNAs of alternating sequence. *J. Mol. Biol.* 5:172–84
39. Jamin M, Baldwin RL. 1996. Refolding and unfolding kinetics of the equilibrium folding intermediate of apomyoglobin. *Nat. Struct. Biol.* 3:613–18
40. Jamin M, Baldwin RL. 1998. Two forms of the pH 4 folding intermediate of apomyoglobin. *J. Mol. Biol.* 276:491–504
41. Karplus M, Weaver DL. 1976. Protein-folding dynamics. *Nature* 260:404–6
42. Kay MS, Baldwin RL. 1996. Packing interactions in the apomyoglobin folding intermediate. *Nat. Struct. Biol.* 3:439–45
43. Kiefhaber T, Baldwin RL. 1995. Kinetics of hydrogen bond breakage in the process of unfolding of ribonuclease A measured by pulsed hydrogen exchange. *Proc. Natl. Acad. Sci. USA* 92:2657–61
44. Kiefhaber T, Labhardt AM, Baldwin RL. 1995. Direct NMR evidence for an intermediate preceding the rate-limiting step in the unfolding of ribonuclease A. *Nature* 375:513–15
45. Kim PS, Baldwin RL. 1984. A helix stop signal in the isolated S-peptide of ribonuclease A. *Nature* 307:329–34
46. Levinthal C. 1968. Are there protein folding pathways? *J. Chim. Phys.* 65:44–45
47. Levinthal C. 1969. How to fold graciously. In *Mössbauer Spectroscopy in Biological Systems*, ed. P Debrunner, JCM Tsibris, E Münck, pp. 22–24. Urbana: Univ. Ill. Press
48. Lifson S, Roig A. 1961. On the theory of helix-coil transition in polypeptides. *J. Chem. Phys.* 34:517–28
49. Loh SN, Kay MS, Baldwin RL. 1995. Structure and stability of a second molten globule intermediate in the apomyoglobin folding pathway. *Proc. Natl. Acad. Sci. USA* 92:5446–50
50. Lopez MM, Chin D-H, Baldwin RL, Makhatadze GI. 2002. The enthalpy of the alanine peptide helix measured by isothermal titration calorimetry using metal-binding to induce helix formation. *Proc. Natl. Acad. Sci. USA* 99:1298–302
51. Lotan N, Yaron A, Berger A. 1966. The stabilization of the α-helix in aqueous solution by hydrophobic side-chain interaction. *Biopolymers* 4:365–68
52. Luo P, Baldwin RL. 1999. Interaction between water and polar groups of the helix backbone: an important determinant of helix propensities. *Proc. Natl. Acad. Sci. USA* 96:4930–35
53. Luo Y, Kay MS, Baldwin RL. 1997. Cooperativity of folding of the apomyoglobin pH 4 folding intermediate studied by glycine and proline mutations. *Nat. Struct. Biol.* 4:925–30
54. Marqusee S, Robbins VH, Baldwin RL. 1989. Unusually stable helix formation in alanine-based peptides. *Proc. Natl. Acad. Sci. USA* 86:5286–90
55. Meselson M, Stahl FW. 1958. The replication of DNA in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 44:671–82

56. Monod J, Wyman J, Changeux J-P. 1965. On the nature of allosteric transitions: a plausible model. *J. Mol. Biol.* 12:88–118
57. Nall BT, Garel J-R, Baldwin RL. 1978. Test of the extended two-state model for the kinetic intermediates observed in the folding transition of ribonuclease A. *J. Mol. Biol.* 118:317–30
58. Oncley JL, ed. 1959. *Biophysical Science—A Study Program*. New York: Wiley
59. Padmanabhan S, Baldwin RL. 1994. Tests for helix-stabilizing interactions between various nonpolar side-chains in alanine-based peptides. *Protein Sci.* 3:1992–97
60. Padmanabhan S, York EJ, Gera L, Stewart JM, Baldwin RL. 1994. Helix-forming tendencies of amino acids in short (hydroxybutyl)-L-glutamine peptides: an evaluation of the contradictory results from host-guest studies and short alanine-based peptides. *Biochemistry* 33:8604–9
61. Pohl FM. 1969. On the kinetics of structural transition I of some pancreatic proteins. *FEBS Lett.* 3:60–64
62. Ptitsyn OB. 1973. The stepwise mechanism of protein self-organization. *Dokl. Nauk. SSSR* 210:1213–15
63. Rice SA, Doty P. 1957. The thermal denaturation of deoxyribose nucleic acid. *J. Am. Chem. Soc.* 79:3937–47
64. Robertson AD, Purisima EO, Eastman MA, Scheraga HA. 1989. Proton NMR assignments and regular backbone structure of bovine pancreatic ribonuclease A in aqueous solution. *Biochemistry* 28:5930–38
65. Roder H, Elöve G, Englander SW. 1988. Structural characterization of folding intermediates in cytochrome *c* by H-exchange labelling and proton NMR. *Nature* 335:700–4
66. Roder H, Wüthrich K. 1986. Protein folding kinetics by combined use of rapid mixing techniques and NMR observation of individual amide protons. *Proteins Struct. Funct. Genet.* 1:34–42
67. Rohl CA, Chakrabartty A, Baldwin RL. 1996. Helix propagation and N-cap propensities of the amino acids measured in alanine-based peptides in 40 volume percent trifluoroethanol. *Protein Sci.* 5:2623–37
68. Rohl CA, Scholtz JM, York EJ, Stewart JM, Baldwin RL. 1992. Kinetics of amide proton exchange in helical peptides of varying chain lengths. Interpretation by the Lifson-Roig theory. *Biochemistry* 31:1263–69
69. Rose GD. 1979. Hierarchic organization of domains in globular proteins. *J. Mol. Biol.* 134:447–70
70. Scheffler IE, Elson EL, Baldwin RL. 1968. Helix formation by dAT oligomers. I. Hairpin and straight-chain helices. *J. Mol. Biol.* 36:291–304
71. Scheffler IE, Elson EL, Baldwin RL. 1970. Helix formation by d(TA) oligomers. II. Analysis of the helix-coil transitions of linear and circular oligomers. *J. Mol. Biol.* 48:145–71
72. Schmid FX, Baldwin RL. 1978. Acid catalysis of the formation of the slow-folding species of ribonuclease A: evidence that the reaction is proline isomerization. *Proc. Natl. Acad. Sci. USA* 75:4764–68
73. Schmid FX, Baldwin RL. 1979. Detection of an early intermediate in the folding of ribonuclease A by protection of amide protons against exchange. *J. Mol. Biol.* 135:199–215
74. Scholtz JM, Qian H, York EJ, Stewart JM, Baldwin RL. 1991. Parameters of helix-coil transition theory for alanine-based peptides of varying chain lengths in water. *Biopolymers* 31:1463–70
75. Scholtz JM, York EJ, Stewart JM, Baldwin RL. 1991. A neutral, water-soluble, α-helical peptide: the effect of ionic strength on the helix-coil equilibrium. *J. Am. Chem. Soc.* 113:5102–4

76. Shoemaker KR, Fairman R, Kim PS, York EJ, Stewart JM, Baldwin RL. 1987. The C-peptide helix from ribonuclease A considered as an autonomous folding unit. *Cold Spring Harbor Symp. Quant. Biol.* LII:391–98
77. Shoemaker KR, Kim PS, York EJ, Stewart JM, Baldwin RL. 1987. Tests of the helix dipole model for the stabilization of α-helices. *Nature* 326:563–67
78. Shore D, Baldwin RL. 1983. Energetics of DNA twisting. I. Relation between twist and cyclization probability. *J. Mol. Biol.* 170:957–81
79. Takahashi S, Kim E-H, Hibino T, Ooi T. 1989. Comparison of alpha-helix stability in peptides having a negatively or positively charged residue block attached either to the N- or C-terminus of an alpha-helix. The electrostatic contribution and anisotropic stability of the alpha-helix. *Biopolymers* 28:995–1009
80. Tsong TY, Baldwin RL. 1972. Kinetic evidence for intermediate states in the unfolding of chymotrypsinogen A. *J. Mol. Biol.* 69:145–48
81. Tsong TY, Baldwin RL, Elson EL. 1971. The sequential unfolding of ribonuclease A: detection of a fast initial phase in the kinetics of unfolding. *Proc. Natl. Acad. Sci. USA* 68:2712–15
82. Udgaonkar J. 2008. Multiple routes and structural heterogeneity in protein folding. *Annu. Rev. Biophys.* 37:489–510
83. Udgaonkar JB, Baldwin RL. 1988. NMR evidence for an early framework intermediate on the folding pathway of ribonuclease A. *Nature* 335:694–99
84. Van Holde KE, Baldwin RL. 1958. Rapid attainment of sedimentation equilibrium. *J. Phys. Chem.* 62:734–43
85. Wake RG, Baldwin RL. 1962. Physical studies on the replication of DNA in vitro. *J. Mol. Biol.* 5:201–16
86. Zimm BH, Bragg JK. 1959. Theory of the phase transition between helix and random coil in polypeptide chains. *J. Chem. Phys.* 31:476–85

# Contents

Annual Review of Biophysics
Volume 37, 2008