Protein Science (1997), 6:2031–2034. Cambridge University Press. Printed in the USA.

RECOLLECTIONS

The problem was to find the problem

ROBERT L. BALDWIN

Biochemistry Department, Beckman Center, Stanford University, Stanford, California 94305-5307

In the summer of 1958, there was a highly unusual meeting on biophysics, sponsored by NIH, in Boulder, Colorado. About 120 participants agreed to stay the entire month. There were lectures in all areas of biophysics and some related fields. A stellar group of senior scientists gave the lectures, including some whose names still resonate today, such as Norman Davidson, Paul Doty, Bernard Katz, Walter Kauzmann, Arthur Kornberg, Cyrus Levinthal, Hans Neurath, Leo Szilard, and Bruno Zimm. A few junior scientists were invited, including David Davies, Matt Meselson, Nacho Tinoco, Tommie Thompson, and myself. The proceedings were published in *Reviews of Modern Physics* and in a book (Oncley, 1959), and the conference was intended to launch the new discipline of molecular biophysics. Molecular biology took off first.

Unknown to me, Arthur Kornberg's Department of Microbiology at Washington University had decided to hire a young physical biochemist when they moved the next year to Stanford to form a new Biochemistry Department. They were influenced by Howard Schachman, who had just spent a sabbatical year with them and introduced them to physical biochemistry. At the end of the Boulder meeting, Arthur Kornberg asked me if I would like to be one of those considered. I said yes and visited them in St. Louis before leaving for Copenhagen, where I would spend a sabbatical year in Linderstrøm-Lang's laboratory. When the offer came in midwinter, I said yes without ever having seen Stanford. I was ready to become a molecular biologist, I thought, and at Boulder I learned that Kornberg's department was one of the best.

After deciding to go to Stanford, what I needed was to find the right research problem. My previous work was in a very narrow field, developing new methods and applying the theory for using the analytical ultracentrifuge to study proteins. Improbable though this may seem today, in 1959 development of new methods for using the ultracentrifuge was a central topic in physical biochemistry. Although my specific training was narrow, I had studied with scientists who took very broad views of their subjects, Jack Williams, Bob Alberty, and Lou Gosting at Wisconsin, and Sandy Ogston at Oxford. Ken Van Holde was in the same laboratory at Wisconsin as I was, and he has described it recently (Van Holde, 1996).

The Stanford Biochemistry faculty had diverse backgrounds and research specialties, but in 1959 they had a common research focus, which was: how do the genes reproduce themselves and direct the cell to make RNA and proteins? Arthur Kornberg, Paul Berg, and Bob Lehman worked with enzymes: DNA and RNA polymerases, nucleases, and tRNA synthases. Dale Kaiser and Dave Hogness studied bacteriophage lambda as a molecular model for how a small genome goes through its life cycle. Dale had been trained in phage genetics and Dave had studied the synthesis of induced enzymes previously. Like Dave, Mel Cohn had studied the induced synthesis of β-galactosidase and, in 1959, he was studying the synthesis of specific antibodies in single cells. The prospect of working on problems like these was exciting, but where could I fit in?

Julius Adler and I used to go for long walks in the redwoods on Sunday afternoons and discuss possible research problems. Although Julius' specific interests were quite different from mine, we both enjoyed discussing future problems that molecular biology should solve, and hiking in the redwood forests near Stanford was a great venue for these discussions. Julius was in his third year as a postdoctoral fellow (two years with Arthur Kornberg followed by one with Dale Kaiser) and, in the next year, he would set up his own laboratory at Wisconsin. Julius had a list of four favorite major problems he wanted to tackle; bacterial chemotaxis finally headed his list. I was not so fortunate in having a clear sense of what to work on.

With the help of Ross Inman, whose Ph.D. work was in DNA physical chemistry, I plunged into this subject in 1959. Working together with Gerry Wake, who handled enzymes confidently, we embarked on a study of the replication of a synthetic DNA, the alternating dAT copolymer. Our work was made possible by the support of Arthur Kornberg, who gave us DNA polymerase 1 and the dAT copolymer, and who never interfered in our work. The dAT copolymer was known to be an exceptionally favorable template for replication by DNA polymerase 1, and we believed that the explanation lay in the unusual physico-chemical behavior of the dAT copolymer. Our first publications were on this subject. This was a truly exciting problem, and taking part in Arthur's research group meetings ("DNA Club") intensified the excitement.

Two spectacular experiments from this period had shown what physical biochemistry might contribute to molecular biology. The

Reprint requests to: Robert L. Baldwin, Biochemistry Department, Beckman Center, Stanford Medical Center, Stanford, California 94305-5307.

Robert L. ("Buzz") Baldwin received his undergraduate degree in chemistry in 1950 at the University of Wisconsin and went to the University of Oxford as a Rhodes Scholar, where he studied physical biochemistry and received his D. Phil. degree in 1954. After postdoctoral work in physical chemistry at Wisconsin from 1954 to 1955, he was briefly (1955–1959) in the Biochemistry Department at Wisconsin. In 1959, he joined the Biochemistry Department at Stanford and remained there. He served as Chairman from 1989 to 1994. He is a member of the National Academy of Sciences and of the American Academy of Arts and Sciences. He received the Stein and Moore Award of the Protein Society in 1992 and the Wheland Award in Chemistry in 1995.

The Baldwin laboratory in 1989 and their present locations. From left to right: Andy Robertson (University of Iowa), Chris Lee (rotation student) (Stanford University), Jayant Udgaonkar (Tata Institute, Bangalore), Katie Armstrong (IBM, Watson Research Center), Virginia Robbins (retired), David Schultz (seated) (University of California, San Diego), Rob Fairman (Bristol-Myers Squibb), Susan Marqusee (University of California, Berkeley), Fred Hughson (Princeton University), Doug Barrick (seated) (Johns Hopkins University), S. Padmanabhan (Instituto de Estructura, Madrid), Jannette Carey (Princeton University).

Meselson–Stahl experiment (1958) showed that DNA synthesis in *Escherichia coli* is semi-conservative, suggesting that the two parental strands of the DNA helix separate as the DNA is replicated. Their work was based on a new type of ultracentrifuge experiment, separating two DNAs by means of the difference between their buoyant densities, by allowing them to form sharp bands in an equilibrium CsCl density gradient. The Marmur–Doty experiment (Doty et al., 1960) showed that the complementary strands of the DNA helix not only come apart when the helix unwinds as it melts, but also that the complementary strands can be put back together again. These experiments provided a lodestar while I pondered possible research projects. I wanted to find a problem whose significance would justify spending the time needed to work out its individual parts. I remembered a seminar at Wisconsin given by Joe Hirschfelder, a chemistry professor, who was speaking 25 years after receiving his Ph.D. with Henry Eyring at Princeton. He had studied the theory of chemical reaction rates for his Ph.D. work, and this was also the subject of his seminar. He remarked that a Ph.D. adviser should consider carefully the problems he gives to his students because, if the problem is any good, the student will still be working on it 25 years later.

In the early 1960s, there were seminars and journal club discussions almost every day in the Biochemistry Department at Stanford, and molecular genetics figured prominently in them. Everyone else in the Department knew much more about genetics than I did. I resolved to take an early sabbatical and learn genetics. There were close connections between our Department and the Institut Pasteur. Dale Kaiser had been a postdoctoral fellow with François Jacob, and both Dave Hogness and Mel Cohn had studied with Jacques Monod. These connections smoothed a path for me to Paris, where I studied phage genetics with François Jacob. In the spring of 1963, before I left Stanford for my sabbatical, I listened to Manfred Eigen, who was visiting in the Stanford Chemistry Department, lecture on fast-reaction kinetics. His lectures were fascinating, and I got his permission to spend the summer of 1963 in his laboratory in Göttingen, where I began to study the kinetics of the helix-coil transition in a synthetic DNA, the dAT copolymer, working with his student Christof Spatz.

The day I arrived in Paris in the fall of 1963 and went to the Institut Pasteur, I was told that Monsieur Monod wanted to see me. He had just finished a first draft of what would become the Monod–Wyman–Changeux (MWC) paper (1965) on a model for allosteric proteins. He had chosen hemoglobin as an archetypal allosteric protein, even though its allosteric effector, bisphosphoglycerate, had not yet been discovered by Ruth and Reinhold Benesch. As I read that first draft, I was quite skeptical. Here was Jacques Monod, a molecular biologist without training in physical chemistry, proposing a solution to the problem of the heme–heme interaction in hemoglobin, which had been studied in depth by such eminent physical chemists as Linus Pauling and Jeffries Wyman. As the year passed and new papers came out on allosteric enzymes, I read successive drafts of the MWC paper and my appreciation of the MWC model grew. A key point was the demonstration by Max Perutz that oxy- and deoxy-hemoglobin have quite different quaternary structures; this information was not available to Pauling and Wyman

when they studied the mechanism of the heme–heme interaction in hemoglobin.

The style of research in Paris was quite different from the style at Stanford. In Stanford, the style was to choose a fundamental problem in biology that could be expressed in biochemical terms, and then solve it step by step using classic biochemical approaches (isolate, purify, study mechanism). In Paris, after identifying an important problem, the style was to propose various models that might provide the solution, and then devise experiments (relying on genetics as well as biochemistry) to distinguish between these models. Watching Jacob and Monod propose and criticize models raised my level of scientific excitement to the point where it began to interfere with my sleep. I took to rock climbing in the forest of Fontainebleau on Sundays to drive scientific questions from my mind for one day out of the week. At the end of my year in Paris, Jacques Monod suggested I join him in learning more about allosteric enzymes. Much as I admired his contribution, I didn't want to do this. He had found his problem (he had a well-developed instinct for doing this) and I wanted to find my problem.

In 1968, I listened to a seminar by Cy Levinthal at Stanford; it must have been the first talk I heard on the protein folding problem. It was entitled "How to fold gracefully." He discussed what came to be known as the Levinthal paradox: an unfolded protein does not have the time needed to explore all possible conformations, and yet apparently it folds rapidly to the single most stable conformation. Paul Flory, who was at the seminar, commented to me "so there must be folding intermediates." The thought percolated in my mind that it would be important to detect and characterize folding intermediates. Their structures could reveal stages in the folding process, and analyzing the interactions that stabilize folding intermediates should help to understand how the amino acid sequence determines the 3D structure of a protein.

In 1968, and still today, the standard model for the equilibrium folding reactions of single-domain proteins was and is the 2-state model U $\rightleftharpoons$ N (U = unfolded, N = native) of Lumry and coworkers (1966) and of Tanford (1968). In the 2-state model, folding intermediates are not detectable. This point probably helped to motivate Levinthal's calculation of how long folding would require if there really were no folding intermediates. In 1970, I was listening to a talk on two-state folding at a Biopolymers Gordon Conference when the thought struck me: fast kinetics should be the way to detect folding intermediates. Immo Scheffler, Elliot Elson, and I had been studying the folding and unfolding behavior of $d(AT)_n$ oligonucleotides, whose repeating ATAT... sequence allows them to form hairpin helices. The presence of folding intermediates is much more evident in experiments based on fast kinetics than in equilibrium studies of these folding reactions.

I persuaded Tian Tsong, who had come to the laboratory recently, to take up this problem. He made temperature-jump measurements of the thermal unfolding transition of ribonuclease A, and found unmistakable evidence for a kinetic intermediate in the process, appearing in the millisecond time range (Tsong et al., 1971). In independent work, Atsushi Ikai and Charles Tanford found kinetic intermediates in a stopped-flow study of the GdmCl-induced unfolding transition of cytochrome *c* (Ikai & Tanford, 1971). I concluded that finding out the nature of these folding and unfolding intermediates could be the long-term research problem I was seeking. Roger Pain likes to remind me that I thought the problem would be solved in 10 years.

By 1973, a cloud hung over the problem, although it was an interesting cloud. Jean-Renaud Garel had found that the major kinetic intermediate in the refolding of RNase A was a second unfolded form, a fast-folding form (U_F), which could fold as much as 100 times faster than the better-populated slow-folding form U_S (Garel & Baldwin, 1973). A different cloud, which might prove to be the same cloud, hung over the cytochrome *c* results: Ikai and Tanford (1971) had found that a major kinetic folding intermediate was an off-pathway intermediate. Their analysis assumed only one unfolded form, and perhaps their results could be explained by two unfolded forms. I visited Duke University in 1974 and was shocked to learn that Charles Tanford was leaving the protein folding problem in order to study membranes and membrane proteins. His careful studies of equilibrium unfolding reactions and of GdmCl-denatured proteins (Tanford, 1968, 1970) had laid the groundwork for experimental studies of the protein folding problem.

My own interest in the folding problem went back to 1950–1953 when I was a Ph.D. student at Oxford (the Oxford term is D. Phil. student). My supervisor, A.G. (Sandy) Ogston, had received preprints from Walter Kauzmann of his kinetic studies of the unfolding reactions of proteins, and Sandy handed them to me to read. The focus was on showing that protein denaturation is in fact protein unfolding, and on investigating the factors that determine the rate of unfolding (see Simpson & Kauzmann, 1953). I found these papers extremely interesting, and my interest in folding was increased later by reading John Schellman's (1955) paper on a model for the stability and unfolding behavior of α-helices in solution. John and I became close friends, and for many years we took annual ski trips (one year in Oregon, the next in California) during which we discussed scientific problems as we rode in the ski lifts. Much of the pleasure in working on any scientific problem derives from the interactions with colleagues who are doing related work. In the 1970s, the scientists interested in the protein folding problem were a small and friendly group, who typically saw each other once a year in a symposium on protein folding at some larger international meeting. I would often see at these meetings Chris Anfinsen, Cyrus Chothia, Tom Creighton, Alexey Finkelstein, Nobuhiro Go, Michel Golberg, Jan Hermans, Jr., Rainer Jaenicke, Joel Janin, Martin Karplus, Jon King, Michael Levitt, Andrew McLachlan, Roger Pain, Max Perutz, Peter Privalov, Oleg Ptitsyn, Fred Richards, Jane Richardson, John Schellman, Harold Scheraga, Julian Sturtevant, Akiyoshi Wada, Don Wetlaufer, and Kurt Wüthrich.

The dilemma posed by the two unfolded forms of RNase A became more acute when John Brandts and coworkers (1975) proposed isomerization about proline bonds in the unfolded protein as the cause. This was an immediately plausible proposal: proline isomerization was known to be slow and the fraction *cis* of a proline peptide bond, unlike ordinary peptide bonds, was often substantial. *Cis* proline peptide bonds appear frequently in the X-ray structures of proteins and RNase A has two *cis* proline bonds. Brandts and coworkers (1975) proposed, moreover, that proline isomerization could account for all kinetic intermediates seen in the folding and unfolding reactions of small proteins.

Thus, we needed to find out if proline isomerization was the correct explanation for the two unfolded forms of RNase A and, if so, how to disentangle structural folding intermediates from kinetic intermediates caused by proline isomerization.

In 1978, Franz Schmid showed that the $U_F \rightleftharpoons U_S$ reaction in unfolded RNase A is catalyzed by strong acid in the manner expected for proline isomerization (Schmid & Baldwin, 1978). The experiment is technically difficult because strong acid is needed. It left no doubt that isomerization about proline or other peptide

bonds is the explanation for the $U_F \rightleftharpoons U_S$ reaction in unfolded RNase A. Resolution of this problem also showed us how to look for structural folding intermediates. Proline isomerization becomes very slow at low temperatures and we needed only to test for rapid partial folding near 0 °C of the slow-folding U_S form of RNase A, which has at least one non-native proline isomer. This experiment gave a very surprising answer (Cook et al., 1979): not only does rapid partial folding of the U_S form occur before proline isomerization, but the product is a native-like folding intermediate (I_N) that, remarkably, binds the specific inhibitor 2′CMP and even, as Franz Schmid later showed, has RNase catalytic activity.

I breathed a deep sigh of relief. We had found a folding intermediate that clearly was a structural intermediate and this problem would last many years. After the discovery of I_N, progress in characterizing the structures and stability of folding intermediates had to wait for the development of 2D NMR methods and their application to proteins by Ernst and Wüthrich. Then, stopped-flow studies of the folding process could be monitored by the hydrogen exchange reactions of individual peptide NH protons, making use of Linderstrøm-Lang's visionary plan of employing hydrogen exchange to probe conformational reactions in proteins. Other approaches to the study of protein folding pathways had appeared by the 1980s. Equilibrium folding intermediates were discovered in the 1970s by Wong and Tanford and by Kuwajima, and their general significance as molten globule intermediates was recognized by Ptitsyn. In 1974, Tom Creighton showed how to isolate intermediates on the pathway of disulfide bond formation.

As I write this, the reality of detectable folding intermediates is widely accepted, but their role in guiding folding pathways is under dispute. Folding intermediates may become populated only when a misfolding barrier slows down or blocks the normal folding process, as suggested by the cytochrome *c* experiments of Sosnick and Englander and the lysozyme experiments of Rothwarf and Scheraga. Simulations and theories of the folding process demonstrate the importance of having an adequate physical model and, starting in particular with work by Wolynes and by Shakhnovich and Karplus, these studies suggest that folding may not follow unique pathways. A search is now under way for a unifying model of the folding process that will satisfy both theorists and experimentalists. As part of this search, methods are being developed for examining very early events in the folding process. The problem of comprehending the nature of the folding process is still very much an open problem.

Acknowledgment

I thank Anne Baldwin for her comments.

References

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*:4953–4963.

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*:6157–6161.

Doty P, Marmur J, Eigner J, Schildkraut C. 1960. Strand separation and specific recombination in DNA: Physical chemical studies. *Proc Natl Acad Sci USA 46*:461–476.

Garel JR, Baldwin RL. 1973. Both the fast and slow refolding reactions of ribonuclease A yield native enzyme. *Proc Natl Acad Sci USA 70*:3347–3351.

Ikai A, Tanford C. 1971. Kinetic evidence for incorrectly folded intermediate states in the refolding of denatured proteins. *Nature 230*:100–102.

Lumry R, Biltonen R, Brandts JF. 1966. Validity of the "two-state" hypothesis for conformational transitions of proteins. *Biopolymers 4*:917–944.

Meselson M, Stahl FW. 1958. The replication of DNA in *Escherichia coli*. *Proc Natl Acad Sci USA 44*:671–682.

Monod J, Wyman J, Changeux JP. 1965. On the nature of allosteric transitions: A plausible model. *J Mol Biol 12*:88–118.

Oncley JL, ed. 1959. *Biological science—A study program*. New York: Wiley & Sons.

Schellman JA. 1955. The stability of hydrogen-bonded peptide structures in aqueous solution. *CR Trav Lab Carlsberg, Ser Chim 29*:230–259.

Schmid FX, Baldwin RL. 1978. Acid catalysis of the formation of the slow-folding species of RNase A: Evidence that the reaction is proline isomerization. *Proc Natl Acad Sci USA 75*:4764–4768.

Simpson RB, Kauzmann W. 1953. The kinetics of protein denaturation. I. The behavior of the optical rotation of ovalbumin in urea solutions. *J Am Chem Soc 75*:5139–5152.

Tanford C. 1968. Protein denaturation. Parts A, B. *Adv Protein Chem 23*:121–282.

Tanford C. 1970. Protein denaturation. Part C. *Adv Protein Chem 24*:1–95.

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*:2712–2715.

Van Holde KE. 1996. A random walk amid the macromolecules. *Protein Sci 5*:792–796.