Jane S. Richardson David C. Richardson

ANNUAL REVIEWS **Further**

Click here for quick links to Annual Reviews content online, including:

- Other articles in this volume
- Top cited articles
- Top downloaded articles
- Our comprehensive search

Doing Molecular Biophysics: Finding, Naming, and Picturing Signal Within Complexity

Jane S. Richardson and David C. Richardson

Department of Biochemistry, Duke University, Durham, North Carolina 27710; email: dcrjsr@kinemage.biochem.duke.edu

Annu. Rev. Biophys. 2013. 42:1–28

First published online as a Review in Advance on February 28, 2013

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

This article's doi: 10.1146/annurev-biophys-083012-130353

Copyright © 2013 by Annual Reviews. All rights reserved

Keywords

scientific biography, structural biology, molecular graphics, ribbon drawings, structure validation, all-atom contacts

Abstract

A macromolecular structure, as measured data or as a list of coordinates or even on-screen as a full atomic model, is an extremely complex and confusing object. The underlying rules of how it folds, moves, and interacts as a biological entity are even less evident or intuitive to the human mind. To do science on such molecules, or to relate them usefully to higher levels of biology, we need to start with a natural history that names their features in meaningful ways and with multiple representations (visual or algebraic) that show some aspect of their organizing principles. The two of us have jointly enjoyed a highly varied and engrossing career in biophysical research over nearly 50 years. Our frequent changes of emphasis are tied together by two threads: first, by finding the right names, visualizations, and methods to help both ourselves and others to better understand the 3D structures of protein and RNA molecules, and second, by redefining the boundary between signal and noise for complex data, in both directions—sometimes identifying and promoting real signal up out of what seemed just noise, and sometimes demoting apparent signal into noise or systematic error. Here we relate parts of our scientific and personal lives, including ups and downs, influences, anecdotes, and guiding principles such as the title theme.

Contents

EARLY INFLUENCES 2
EARLY PROTEIN CRYSTALLOGRAPHY 4
Staphylococcal Nuclease at MIT 4
Cu,Zn Superoxide Dismutase at Duke 6
STRUCTURAL INFORMATICS: FOLDS AND MOTIFS 8
REPRESENTING 3D STRUCTURE 11
Ribbons 11
Kinemages 11
PROTEIN DESIGN 13
HYDROGENS FOR BETTER STRUCTURES 14
All-Atom Contacts 14
MolProbity 15
RNA BACKBONE 17
CURRENT PREOCCUPATIONS: ENSEMBLES AND LOW RESOLUTION 18
PERSONAL BIOGRAPHIES 20
Family and School 20
Swarthmore and MIT 22
Hobbies 22
TAKE-HOME MESSAGES 23

EARLY INFLUENCES

Although Jane's education in philosophy sounds irrelevant, it has actually been very useful because it teaches an extreme level of critical thinking: identifying and questioning all assumptions, both what is considered "known" and your own assumptions as well. However, at least before the more exaggerated versions of postmodernism, it also encouraged the search for real meaning. That critical mindset is a great advantage for improving the chances that your scientific research will stand the test of time.

Being at Swarthmore and then MIT through the 1960s, we absorbed a very open ideology and what has turned out to be an overly optimistic view of human progress, but we haven't had to abandon it entirely. An important aspect that is still alive and well is the goal of open access to information, one root of which was the free software movement that started at MIT while we were there. Later, as it became more feasible to share software, we have embraced that idea wholeheartedly. All of our lab's programs are freely available, open source, and if possible multiplatform. As many of our papers as we can arrange are open access. Recently, Jane has become enamored of contributing both scientific and other open-license images to Wikimedia Commons (**http://commons.wikimedia.org/wiki/User:Dcrjsr**), where other people reuse them for unpredictable and interesting purposes. As part of being the current president of the Biophysical Society, she has started a WikiProject Biophysics to encourage society members and others to edit and improve biophysics-related Wikipedia articles (**http://en.wikipedia.org/wiki/Wikipedia:WikiProject_Biophysics**). Over the years we have done very well by emphasizing collaboration over competition, talking freely about results, and encouraging our students to carry their projects away with them. After all, there are many more

Figure 1

Top: Chris Anfinsen (courtesy of Richard Nowitz). Lower left: Fred Richards (courtesy of Sarah Richards). Lower right: Fred Brooks (courtesy of Jerry Markatos).

great directions to follow up than we can possibly do ourselves, and science is about the joint building of an actively evolving network of information.

Another influence from being at MIT was the chance to get a sneak preview of the first large RNA structure (Phe tRNA), which our friend Sung-Hou Kim solved in Alex Rich's lab there. It was very exciting to see the elegant and complex RNA backbone traced out by the clear phosphate peaks, even at 4 Å resolution. Later, when we saw a way that our methods could be useful for RNA 3D structure and found an interested graduate student (Laura Murray), we jumped at the chance and have found that RNA and protein structure act just differently enough to illuminate each other.

We were fortunate to have great role models and inspirations for our own research careers: Christian B. Anfinsen, Frederic M. Richards, and Frederick P. Brooks Jr. All three are notable as deep and innovative thinkers, for treating everyone well, and for thoroughly enjoying their work. Chris turned us on to the inexhaustible charms and complexities of protein 3D structure, suggested staphylococcal nuclease as a folding-relevant subject for crystallography, and solicited the long article for *Advances in Protein Chemistry* (53) that resulted in the development of ribbon drawings. In between MIT and Duke we spent part of a year at NIH, and admired Chris' daily one-on-one interactions in the lab: always supportive, but probing, and gently working in ideas that became the other person's own. In **Figure 1** he's shown holding a gold-plated Byron's bender wire model (70) that we gave him of our staphylococcal nuclease structure, with Jane's ribbon drawing of the TIM barrel (triose phosphate isomerase) in the background.

Both Chris Anfinsen and Fred Richards were visionary pioneers who saw through to the meaning of simple experiments and simple ideas for understanding the complexity of protein structure, folding, and functionality. Fred was originally a distant inspiration as running one of the first two groups in the United States to solve a protein structure (88). At NIH in 1969, Jane first built a brass model of staphylococcal nuclease by measuring into a stack of contoured glass sheets in Chris' lab, with nuclease experimentalists looking over her shoulder and changing what they did the next day. Then she built a model of γ-chymotrypsin in David Davies' lab, this time in a big Richards' box with its half-silvered mirror (46). Later, we interacted at meetings, absorbing Fred's informal style and his delight in provocatively worded lectures that drew attention to new, dogma-breaking ideas. In the 1980s we worked with him on a small committee that pushed successfully to require the deposition and release of coordinates from publicly supported macromolecular crystallography (47, 48), although for then we had to compromise on data deposition. Fred was intensely involved in research when at the lab, but every summer he took an entire month off for a sailing expedition. We've tried to emulate that with our backpacking and have found it an essential source of balance, perspective, and ideas for new directions. In **Figure 1** Fred is shown with his wife Sally on Mt. Washington, enjoying the benefits of their foul-weather sailing gear.

We spent a large fraction of our lives from the early 1970s to the early 1990s in Fred Brooks' computer graphics lab at the University of North Carolina (UNC). There we accomplished much of our own research in protein structure, acted as guinea pigs for in-depth testing of their software and hardware, and played happily with the science fiction–level gadgets that explored far-out new possibilities such as virtual reality displays, volume rendering, force feedback, fitting models into electron density, or tugging on atoms to move local structure with (more or less) physical realism (77). Some things worked splendidly and soon became widespread; some failed by being surprisingly unhelpful, making you sick, or whacking you in the chest (their gadgets never just fell apart); some of the most interesting were temporarily abandoned to await orders-of-magnitude increases in computer speed. We saw the intellectual fascination and personal satisfaction of doing methods development tuned to the real needs of users (whether or not they yet know what those needs are!) and then seeing the methods adopted (8). Fred taught us big principles about graphics and programming and gave us the courage to try it ourselves. In **Figure 1** he's shown in his office, with his characteristic wide smile.

EARLY PROTEIN CRYSTALLOGRAPHY

Staphylococcal Nuclease at MIT

After Chris Anfinsen's pivotal protein-folding work that used disulfide formation in ribonuclease A, he adopted staphylococcal nuclease as a second model system to study folding without disulfides. He visited MIT to try persuading Alex Rich to do its crystal structure, but Alex wanted to stick with nucleic acids (which indeed served him well, with Z-DNA and transfer RNA). But Al Cotton was intrigued with trying a protein structure and that led to Dave's thesis project.

First, of course, crystals are necessary. Ted Hazen (then our group's protein-chemist postdoc) and Dave explored many conditions one at a time, in the bottom of half-dram shell vials, in batch mode or by adding reagents slowly and swirling to watch for cloudiness in the schlieren lines that meant temporary precipitation. The magic precipitant turned out to be 2-methyl-2,4-pentanediol, which grew nice chunky square prisms in space group $P4_1$ (13). The fourfold screw axis lines up all the molecules the same way along that axis, and sometimes the crystal morphology showed that directionality with rhombic faces on one end and triangles on the other (5, 62). Data collection started on film, with Jane estimating spot intensities by eye, but soon we got a GE

4-circle diffractometer. Dave modified it from small molecule to protein use, having custom tantalum beam slits machined and learning how to line it up to thousandths of an inch by gently tapping the base with a large hammer. The crystals diffracted much better when grown with the inhibitor thymidine 3′,5′-diphosphate, a result that fits with tritium exchange done in Chris' lab showing much more rapid bulk exchange for apo than for inhibited nuclease. The inhibitor also enabled a closely isomorphous derivative by substitution with 5-iodouridine, synthesized by Jim Bier (4).

Developing our data reduction and analysis software was Dave's bailiwick, both learning to program (in Fortran) and reverse-engineering the algorithms from the two or three relevant papers a year—easy to keep up with, even if we needed to reinvent some of the crucial details. We were too young to have used Beevers-Lipson strips for calculating Fourier transforms manually, and Dave could even use other people's Fourier routines, but the rest was done from scratch. Our diffractometer, where Jane cherished and criticized each measurement, was controlled by punched paper tape, and its output, and Dave's programs, were on 80-column IBM punch cards—subject to the perils of dropped card decks and mangling by the mechanical card readers. They were ubiquitous and essential for many years, each little punch chad literally being an individual bit.

David Harker, when visiting MIT to present his ribonuclease A structure, pointed us to using the anomalous scattering data, and we became ardent converts to this example of pulling genuine information from what most people then considered hopelessly down in the noise. In 1967–1968 we obtained an electron density map at 4 Å resolution (5), which showed a nonuniform level of accuracy. The inhibitor, right at the best heavy-atom derivative site, looked even better than it should have at 4 Å: We could see density for oxygen positions on the phosphates and methyl and O substituents on the base ring. However, our attempt to trace the chain was entirely wrong. There was by then a very extensive body of chemical modification data available for the nuclease, such as which residues showed lowered reactivity on inhibitor binding, then interpreted as "burial" by the inhibitor. Our 4 Å chain tracing was made almost entirely consistent with what we thought all that data meant as well as with the map, but fortunately we realized it was very speculative and did not publish it. Later, it turned out that some residues apparently buried by the inhibitor are actually on the opposite side of the molecule, protected both by general tightening and by subtle conformational changes on binding.

The 2 Å map that solved the structure (4) was obtained in 1969 from two heavy-atom derivatives using both isomorphous and anomalous differences. Tracing the chain in that very clear map was a real delight, and it began our lifelong education in what is protein-like: how amino acids, secondary structures, motifs, and ligand binding really look and how they prefer to arrange themselves. With good data and phases, the well-ordered parts were quite unambiguous, even with no refinement or solvent flattening (and of course with no model bias), so that we saw what real signal looks like for protein structure.

Staphylococcal nuclease has a five-strand, twisted β-barrel and three α-helices (**Figure 2*a***), one of which was in the plane of our stacked-up map sheets and gave us the first match to the sequence. We still use fitting that helix from a mini-map slab as a student exercise in our macromolecular structure class. The electron density for the helix is shown in **Figure 2*b***; see if you can work out which sidechain shapes are the Pro and the Met in the sequence YGPEASAFTKKMVENAKK. The coordinates were deposited to the Protein Data Bank (PDB) in 1973 as 1SNS (obsoleted in 1982 by 2SNS). The 1SNS structure factors were deposited in 2007, entered from printout since we had lost all electronic forms of them—a good argument for now doing it right away.

The first hundred of the 149 residues (the β-barrel, with the inhibitor-binding site and our favorite helix linking strands 3 and 4) form a compact core, which Murzin (41) made the founding member of the oligonucleotide/oligosaccharide-binding fold (OB-fold); his SCOP database currently lists the OB-fold as occurring in 16 different protein superfamilies (42). Five residues

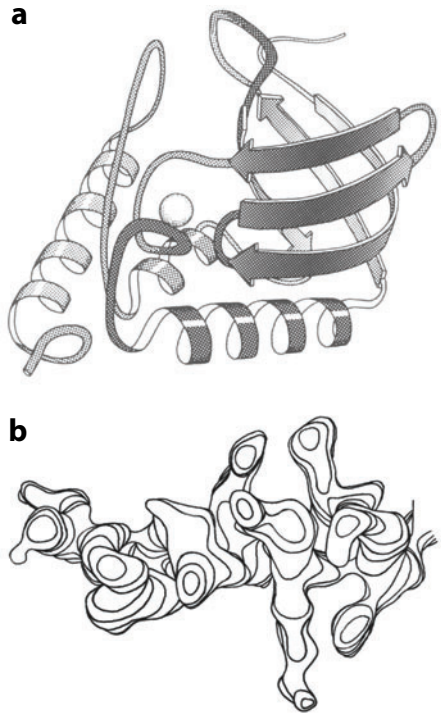

Figure 2

Staphylococcal nuclease. (*a*) Ribbon drawing of the backbone fold and the bound Ca^{2+}. (*b*) Electron density for the horizontal helix in panel *a*; the solvent side is below the helix and the hydrophobic core is above the helix.

at the N terminus and seven at the C terminus were invisible in the 2 Å map and were cleavable with very little effect. The K45-K53 loop near the active site was partially disordered (30–50% of full density), and trypsin cleavage at its tip reduced activity by tenfold. Apparently our description of those regions was the first crystal structure report that discussed intrinsic disorder in proteins (81; K. Dunker, personal communication), now the subject of a whole field of study (e.g., 86). These observations sensitized us to issues of alternate conformations, motion and ensembles that we have pursued again more recently (e.g., 7, 15). Anfinsen's desire for the nuclease structure as a model system for studying protein folding was borne out both by work in his lab and later by very extensive mutational, structural, energetic, and computational folding studies at Johns Hopkins (e.g., 73).

Cu,Zn Superoxide Dismutase at Duke

During our time at NIH in 1969–1970, we commuted to set up our lab space at Duke. Dave was a starting Assistant Professor in the Department of Biochemistry and Jane had an odd but nice position in the Department of Anatomy. We canvassed the faculty for interesting crystallographic projects, and the one that grew crystals was bovine Cu,Zn superoxide dismutase (SOD), which was the Biochemistry "departmental enzyme." Irwin Fridovich and Joe McCord had just recently identified its function as an efficient scavenger of the damaging $O_2^{-\cdot}$ superoxide radical (38), and they went on to study its enzymology, variants, and biomedical implications very productively

(23). Bob Hill determined its amino acid sequence (75), Raj (K.V. Rajagopalan) studied its metal sites (6), and we determined its crystal structure.

Quite nice monoclinic crystals of SOD grew early on, blue from one direction and green from the others, reflecting the spectroscopy and orientation of the Cu site. Our bottleneck was that the cell dimensions meant there were four nonequivalent chains in the asymmetric unit (two biological dimers), and heavy-atom derivatives had many binding sites at varying occupancies. A viable solution for the phases was finally obtained by Ken Thomas in a classic graduate student coup: Against our advice, he located and refined numerous low-occupancy sites in addition to the major ones. On a heavy-atom Fourier map, those sites outlined the four chains almost as well as a negative-stain EM image (see figure 3 in Reference 79).

At 3 Å resolution the chain was traceable, but only barely—it relied heavily on comparing the density for all four chains to confirm the signal. Dave had to leave to give a talk at an American Crystallographic Association meeting, and Jane carried the mini-map around with her for days and worked on the tracing obsessively, with our infant son Robert cradled next to it, or showed students what she was doing rather than teaching them the course material. Just before Dave's talk he phoned her and got the scoop on the latest β-strands that had emerged to define the fold. Figure 1 of Richardson et al. (68) is a stereo photograph of the mini-map with narrow, color-coded tape stuck on the layers to follow the Cα backbone trace. We assigned chain IDs as O, Y, B, and G from the colors of the tape (O first, because the orange chain turned out to be the best ordered). The Cα coordinates were first published as a printed table (67) and then deposited to the PDB as 1SOD, so at that point we had contributed two of the first 20 or so distinct protein structures in the database (62).

Data was then collected out to 2 Å resolution and a complete model was built on the GRIP-75 system in the UNC computer graphics lab, where the SOD structure was a driving problem for methods development (62). Refinement for protein structures was still quite new, slow, and expensive. Because we couldn't afford such long runs at the Duke computer center, our students Libby Getzoff and John Tainer did the refinement by traveling the country and using night-shift time at our friends' labs who had dedicated machines. After refinement, the 2SOD coordinates (78) obsoleted 1SOD. The active site has copper ligands in a distorted square plane and tetrahedral coordination for the zinc, with the surprise that they share a histidine ligand (**Figure 3**). Every

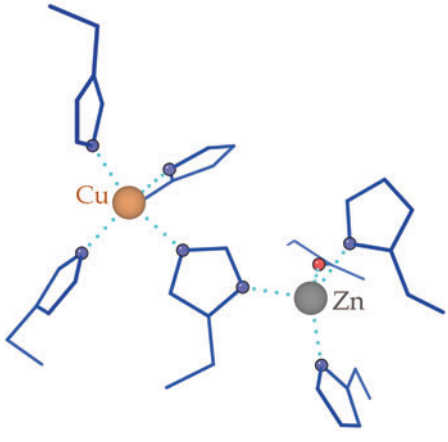

Figure 3

Cu,Zn superoxide dismutase: active-site metal ligands, including shared histidine.

metal ligand sidechain is held stably in place by a second-sphere H-bond, and the entire site is cradled between the outer side of the β-barrel and the well-ordered but nonrepetitive structure of two long loops. The SOD structure taught us additional things about 3D organization in protein structure, such as the convex curvature that a fully H-bonded Lβ Gly could produce in an otherwise regular β-sheet, the prevalence of sidechain-to-backbone H-bonding that can form nonrepetitive loops just as compact and stable as secondary structures, and the electrostatic guidance by strategically positioned charge pairs that could pull in the $O_2^{-\cdot}$ substrate for catalysis at near diffusion-limited rate (26).

STRUCTURAL INFORMATICS: FOLDS AND MOTIFS

Macromolecular crystallography gives the eventual stupendous reward of beautiful, reliable, and highly detailed information more extensive than you can fully assimilate in many years of analysis. But for structures that are difficult for their era, it may take many years to get there, and in the meantime you have essentially no interpretable intermediate results. At Duke, while working on the SOD structure, we amused and motivated ourselves by looking for patterns in the few dozen then-known protein structures. This sort of thing is now called structural bioinformatics, but had no name at the time.

Our first major result was the finding that essentially all crossover connections between parallel β-strands are righthanded (51), as diagrammed in **Figure 4*a***. This is the major organizing principle in determining the fold patterns for parallel β structures, such as the TIM barrel and the Rossmann fold of nucleotide-binding proteins, and is now referred to as a supersecondary structure. As happened a number of times, Janet Thornton's lab made the same discovery independently at almost the same time (76), in a very friendly example of shared preoccupations. We have to admit that their name of "β-α-β loop" has statistically won out over our "crossover connection," in spite of the fact that there is not always a helix in the connection. A more local β motif we discovered and named is the β-bulge (56), a common irregularity in β-sheet that has two residues opposite one between a pair of backbone H-bonds, as shown schematically in **Figure 4*b*** for the classic

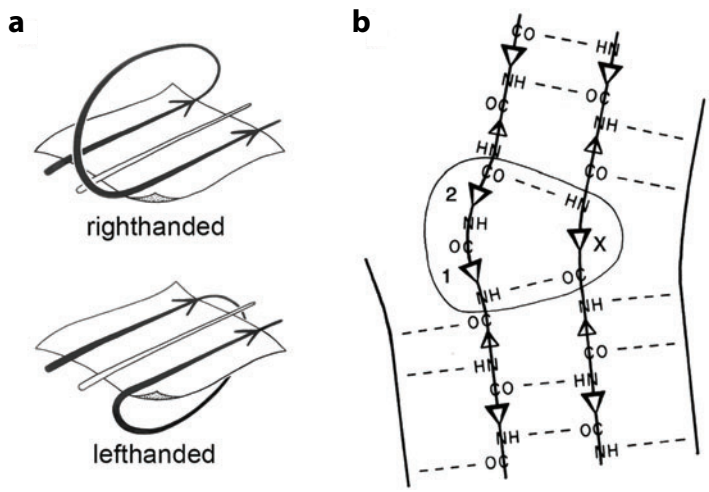

Figure 4

Motifs in β structure. (*a*) Handedness of the crossover connections that join parallel β-strands; righthanded is overwhelmingly preferred. (*b*) Diagram of the classic β-bulge.

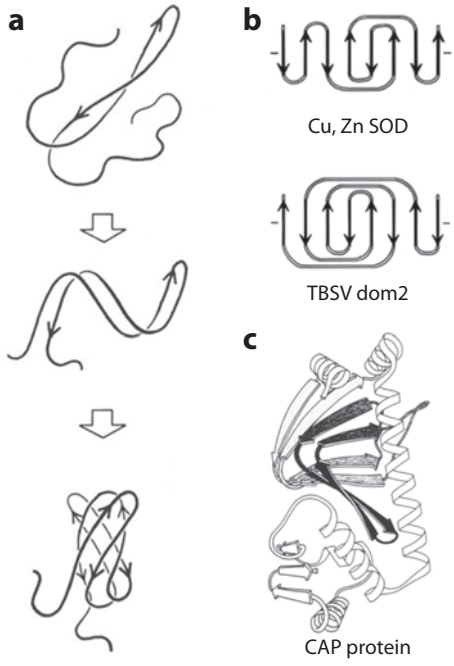

Figure 5

The Greek key motif. (*a*) Proposed folding pathway from a long, twisted, two-strand β-ribbon curling around to form the handed topology of a β-barrel. (*b*) Topology diagrams of a Greek key (*top*) and a jellyroll (*bottom*) β-barrel. (*c*) Highlighted β-ribbon that forms the jellyroll topology in the β-barrel of CAP protein.

antiparallel type. Bulges accentuate the twist of the sheet and are presumably a mechanism for adapting without disaster to a single-residue insertion or deletion within β structure.

The next development was precipitated by an accidental meeting at the door before a regional crystallography meeting, when we and David Davies were each carrying wire models of our proteins—SOD for us and an immunoglobulin domain for him. We each stopped dead, said "That really looks familiar," and then sat down inside to do a proper comparison of the resemblance we had not seen from the published figures. The two structures indeed had the same crisscrossing topology of connections between the β-strands in their antiparallel β-barrel folds. After further cogitation, we realized that the arrangement matched the Greek key design common on Greek vases, hotel bathmats, etc. We therefore named this fold the Greek key β-barrel, which has turned out to be one of the most common antiparallel β folds in spite of containing non-nearest-neighbor connections. At the time, we did a simple-minded calculation of how probable it would be to get this match among topologies by chance and concluded that immunoglobulins and SOD were evolutionarily related (64). Later, we realized there was an organizing factor in protein folding that made the Greek key a highly favored arrangement (52), because a long pair of β-strands that fold up together will produce the Greek key topology as a natural consequence, as illustrated in **Figure 5*a***. An extension of the Greek key wrapping to more strands, such as the virus domain topology in **Figure 5*b*** or the CAP ribbon in **Figure 5*c***, is called a jellyroll fold, for obvious reasons.

We have often gained understanding from metaphors, such as the Greek key. The most important example probably was origami as a metaphor for protein folding (60), which led to the realization that paper needed 90° folds in order to make truly 3D objects and thus that we should

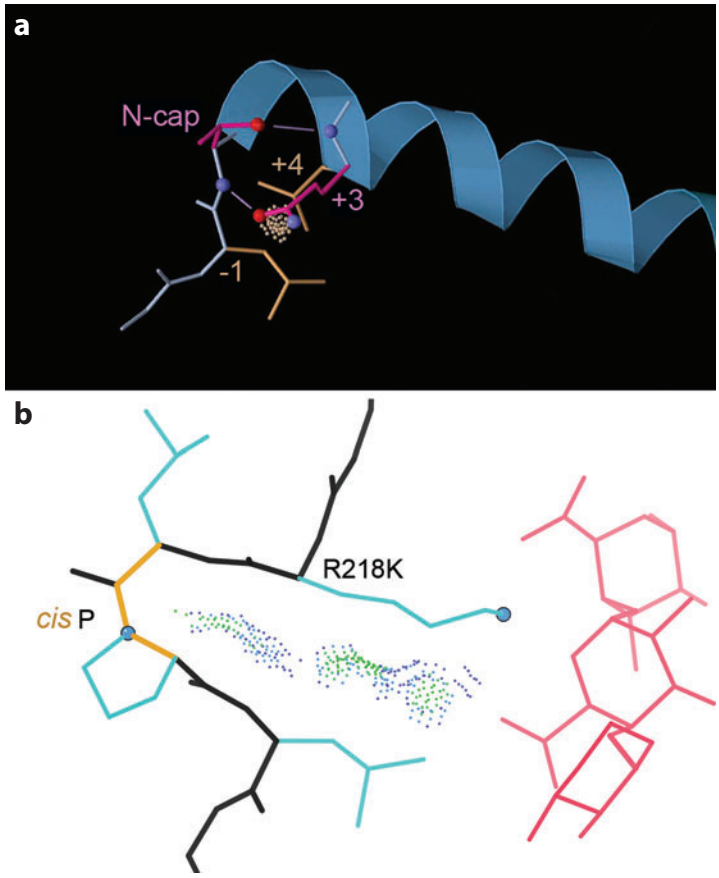

Figure 6

(*a*) Example of a classic helix Ncap motif (1lmb Ser 3), with Ncap and cap-box H-bonds and Ncap-1 to Ncap+4 hydrophobic contact (*beige dots*). (*b*) Example of a *cis*-Pro touch-turn (2ewe 220), positioning the active-site Arg to cleave a cell wall polysaccharide.

look for 90° motifs at the ends of secondary structures. What we found was the helix Ncap and Ccap, the local arrangements that specify and stabilize the ends of α-helices, with the cap residue half in and half out of the helix and its flanking peptides perpendicular to each other (58). That was another example of friendly back-to-back publication, next to George Rose's complementary analysis of helix ends (43); this time our terminology of helix caps won out. A classic Ncap like the one in **Figure 6*a*** has a Ser/Thr/Asn/Asp sidechain oxygen H-bonded to the backbone NH of cap+3, a Gln/Glu at cap+3 that H-bonds to the backbone NH of the Ncap residue, and a hydrophobic contact between cap-1 and cap+4. A classic Ccap, following Charlotte Schellman (72), consists of an Lα Gly making two backbone H-bonds in reverse order, to turn the polypeptide chain outward, plus a hydrophobic contact of Ccap+1 and Ccap-4 sidechains. The concept was so apt and timely that helix caps entered the vocabulary immediately, and almost no one remembers where the term came from—that's real success! The capping motifs are used in prediction and design and in analysis of stability and folding (e.g., 30).

Other motifs that we were quite taken with at the time but that didn't catch on very generally were the Tyr corner that stabilizes many Greek key connections (28), the helix lap-joint showing

that the DNA-binding helix-turn-helix and the Ca-binding EF hand are sequence and charge reversals of each other (59), and the Alacoil, an especially tight antiparallel interaction of two helices that depends on strategic placement of small Ala or Ser residues (25). With Duncan McRee, we even got to trying whether one could recognize protein folds by inspection of the diffraction pattern. The answer was yes in a few especially favorable cases, such as a four-helix bundle in a simple unit cell, or the tilted ring of helices in a TIM barrel, which produces strong reflections in a two-armed spiral-galaxy pattern but spiral in all three dimensions.

Most recently we described the *cis*-Pro touch-turn (82), where the *cis* Pro allows the two flanking peptides to stack tightly on their planar faces rather than making an H-bond. It is notable for joining the γ-turn in what is probably a general class of motifs that are somewhat energetically unfavorable and therefore quite rare, but that are perhaps even more biologically important than common motifs because nearly every occurrence is at a functional catalytic or binding site where its unusual arrangement is needed. **Figure 6*b*** shows the *cis*-Pro touch-turn in the β-helix that positions a critical Arg in the active site of pectate lyase and related enzymes (71). Of course, finding this sort of rare motif depends on quality-filtering the database to prevent errors from swamping the small signal.

REPRESENTING 3D STRUCTURE

Ribbons

In 1979 Chris Anfinsen, as one of the editors of *Advances in Protein Chemistry*, persuaded Jane to write a review called "The Anatomy and Taxonomy of Protein Structures" (53; now out of print but available in annotated form at **http://kinemage.biochem.duke.edu/teaching/anatax**), which ended up very productively occupying two years of her time and nearly that much of Dave's as well. It's hard to imagine getting away with that on NIH grant funding these days, unfortunately. About half of it was a review and the other half was new, including the development of ribbon drawings to illustrate, compare, and classify all the then-known 75 or so distinct protein domain structures. Ironically, although many people consider drawings or computer graphics images of structures to be optional extras rather than real science, the ribbons are still Jane's contribution that is most known and valued.

There had been several earlier individual schematics in rather similar styles, but the 1981 ribbons adopted a set of conventions for representing helices, β-strands, and loops which are not actually consistent with each other but look as though they are (54, 55) and have proven effective at conveying the primary essentials of a 3D protein fold. They are our most iconic example of showing the signal in complexity. This is illustrated in **Figure 7** by a comparison, at the same view and scale, of an all-atom stick-figure model and the ribbon drawing of a Cu,Zn SOD subunit.

These days, even Jane is happy to be able to produce ribbon diagrams with computer graphics rather than by hand. After trials of several alternative algorithms such as trapezoidal peptide planes, helix cylinders (which work very well for large helical proteins), tighter helical spirals, or less smoothed β-strands, most computer-drawn ribbons have converged on a look very close to the originals.

Kinemages

Because we've always thought of macromolecules in very visual terms, we started using molecular displays on the computer as soon as that was feasible. After haunting the UNC computer graphics lab and commuting to Richard Feldmann's setup at NIH and Bob Langridge's at UCSF for their

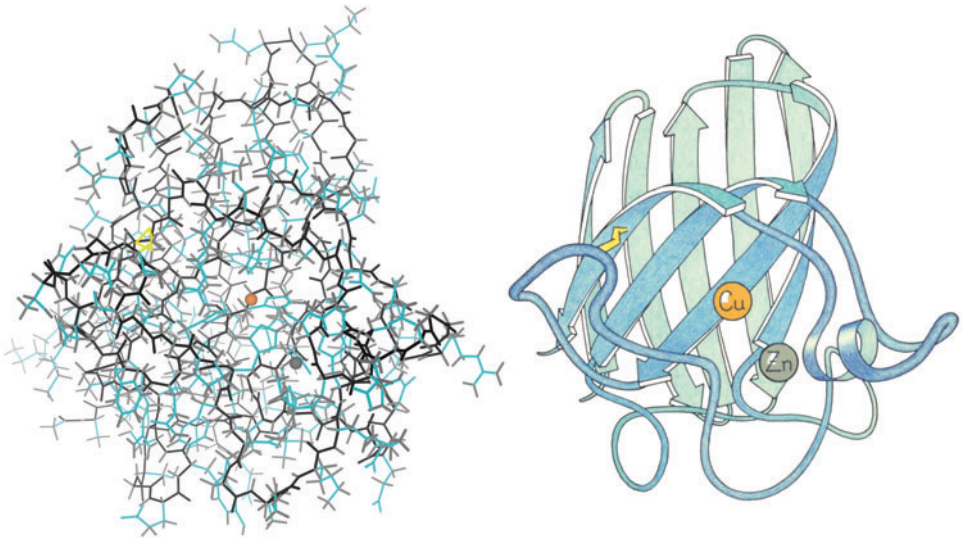

Figure 7

How a ribbon shows the signal within macromolecular complexity. Left: All-atom stick-figure graphics of the Cu,Zn superoxide dismutase (SOD) subunit (2sod) (main chain in black, sidechains in cyan, and hydrogens in gray). Right: Ribbon drawing of the Cu,Zn SOD subunit, with Cu in peach and Zn in gray (53).

pioneering early graphics, in the 1980s we managed to spring for our own Evans & Sutherland display and lecture hall projector, and Dave programmed stick-figure and multistrand ribbon graphics in its function-network language. As the hardware gradually got smaller and more affordable, we went through several generations and in 1990 realized that the new small Macs were capable of smoothly moving about 500 vectors in real time, which is just enough to show meaningful macromolecular graphics if they are thoughtfully constructed. That was also when the journal *Protein Science* was being launched, and Dave undertook to develop a graphics system for its diskette supplement. The resulting kinemage graphics and Mage display program (49), and the contemporaneous Rasmol software, were the first widely distributed macromolecular graphics on personal computers.

Kinemage graphics are unusual in having an intermediate display list (between the coordinate file and the on-screen display) that can be stored externally. The kinemage format is plain text, with very simple markup designed to be written and edited by people as well as by computers. The display characteristics were tuned to optimize smooth movement, depth perception, color palette, one-click identification, and a shallow learning curve to 3D molecular understanding—but not especially for making 2D presentation images. The first uses were as interactive additions to journal papers, lectures, and student exercises (49, 50), but kinemages rapidly became essential to our own research and accumulated many additional functionalities. **Figure 8** shows a ribbon representation from Mage and the creator of Mage and kinemages, "Dave the Mage." Since 2003, kinemages can be displayed online by KiNG (Kinemage Next Generation) software, written in Java by Ian Davis (17) and continued by Vincent Chen (11) as students in the Richardson lab. KiNG also enables model rebuilding with "backrub" motions (15) and interactive validation feedback such as the all-atom contacts described below, and it is available within the GUI for the Phenix crystallographic system (1). Kinemages have even been ported to virtual reality in order for Jeremy Block (7) to immersively study the relationship of NMR experimental data to NMR model ensembles.

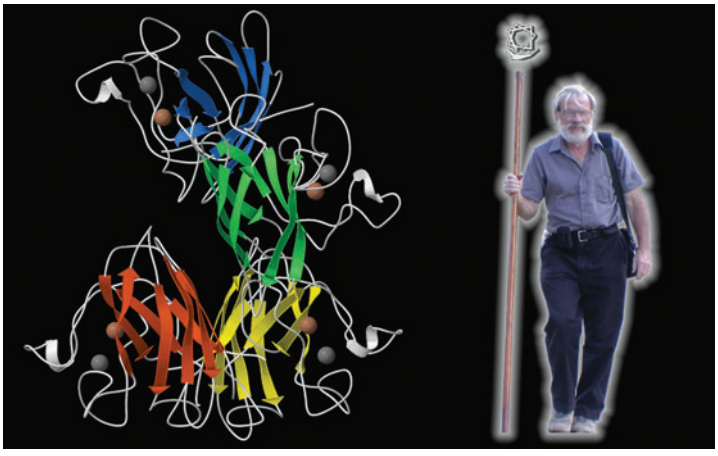

Figure 8

Left: Ribbon representation from the Mage program of the two Cu,Zn SOD dimers in the asymmetric unit of the 2sod crystal structure; rendered in Raster3D (39). Right: "Dave the Mage," creator of kinemages and Mage, with the program icon topping his hiking staff (courtesy of Ian Davis).

PROTEIN DESIGN

After many years of looking at and analyzing protein structure, we felt perhaps it was time to test our understanding by trying to design de novo an amino acid sequence that would fold up into a specified 3D structure. That was an exciting but rather scary idea in the early 1980s when we and Bruce Erickson started collaborating on Betabellin, a twofold β-sandwich with a native-like sequence (22, 57) (**Figure 9*a***), and David Eisenberg and Bill DeGrado started collaborating on α-helical peptides with reduced sequence that assembled into a bundle (20). Both were designed by simple sequence/structure motif statistics and constructed by peptide synthesis, as computational model building and gene synthesis were still in their infancy. Betabellin clearly formed β structure,

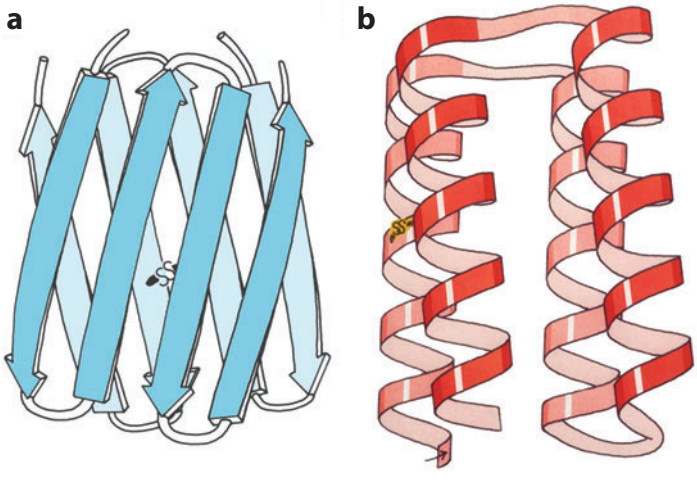

Figure 9

Early de novo design models. (*a*) Betabellin, a β-sandwich dimer. (*b*) Felix, a left-turning four-helix bundle with an SS bond to diagnose the topology.

but it had quite low solubility. It was improved substantially by replacing the unusual chemical linker with a disulfide between the sheets and by using D-amino acids to match the β-hairpin twist rules recently discovered (74), but it still behaved more like a molten globule than a native protein (65). Soon molecular biology enabled DeGrado to make a single-chain helix bundle that was very stable (45), and enabled us, with the expert help of Michael Hecht, to design and make a left-turning four-helix bundle named Felix (**Figure 9*b***) that achieved the intended topology (27). However, both designs still failed at the well-ordered specificity of native protein structures. From aspects that worked, and painfully but instructively from what failed to work, we all learned important new lessons such as the importance of negative design (27) and that specificity of structure is more difficult than stability (65). Later, we realized that a key reason for the difficulty and insolubility of β-sheet designs was the negative-design need for irregularity in edge β-strands to avoid aggregation (61, 65).

Protein design has progressed greatly since the early days, with more knowledge and the essential help of large-scale computations and high-throughput protein production. Redesign on a known backbone is quite successful (e.g., 9, 14), and de novo design can now achieve the correct fold with native-like, crystallizable order fairly often, following on from the landmark Top7 design (33). These days we are involved collaboratively with the Baker lab's Rosetta (35, 80) and with the Donald lab's Osprey through joint students Swati Jain and Daniel Keedy pursuing design of protein/RNA interfaces and of protein backbone, respectively (24, 32).

Even now, the design endeavor still retains some of the feel and the risk of verbal magic. By an arcane process one comes up with an incantation in the language of one-letter code, dramatically intoned to the molecular biologist: MPEVAENFQQCLERWAKLSVGGELAHMANQAAEAILKGGNEAQLKNAQAMHEAMKTRKYSEQLAQEFAHCAYKARASQ !!! (Felix; you can listen to the word at User:Dcrjsr on Wikimedia Commons; **http://commons.wikimedia.org/wiki/File:Felix_sequence_incantation.ogg**). If you're talented and lucky, and pronounce it exactly right, you might get your desire; if not, the earth might open up and swallow you and your grant.

HYDROGENS FOR BETTER STRUCTURES

All-Atom Contacts

Among the roadblocks we saw to protein design (65), the one that seemed most addressable was the lack of a suitably complete and detailed metric for analyzing "goodness of fit" between the residues and atoms inside a well-folded protein. Our exposure to early protein NMR (87) and to interactively steered molecular modeling (77) had convinced us that paying careful attention to all the hydrogen atoms would be essential, because an H intervenes between nearly all neighboring pairs of other atoms, in nonpolar contacts as well as in H-bonds. More than five years of development, spearheaded by Mike Word, seconded by Simon Lovell, and involving near-daily discussions among all of us in the lab, produced the new method of all-atom contact analysis (84, 85). The Reduce program adds and optimizes all explicit hydrogen atoms by identifying and optimizing complete local networks of H-bonds and van der Waals contacts. That necessitates consideration of 180° NQH flips of sidechain amides and His rings, which are very often fit backward because of the similar electron density of N versus O or N versus C atoms, but which can be very reliably corrected. However, allowing methyls to rotate proved counterproductive, because they actually stay within 10–15° of staggered and most of their clashes are due to slightly misplaced methyl carbons, which can then adjust in refinement if the hydrogens are considered. Similarly, we diagnose protonation states only by their H-bonding and van der Waals contacts

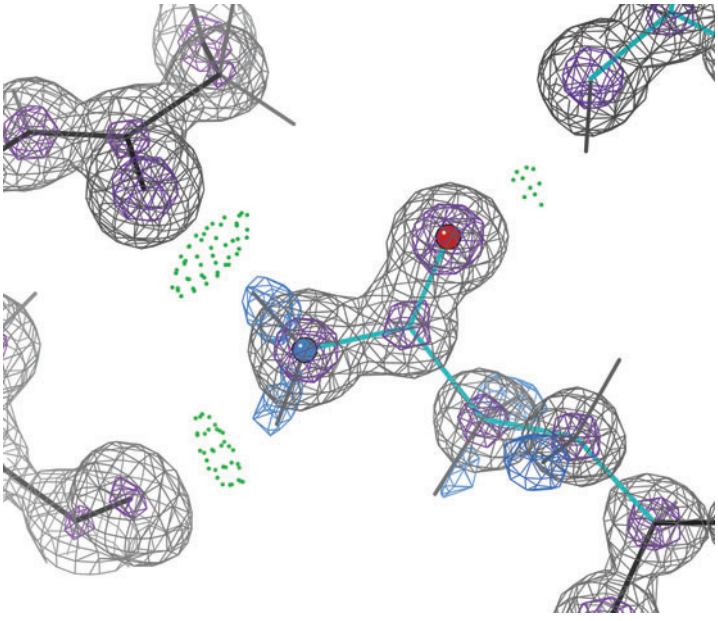

Figure 10

All-atom contacts and the orientation of a Gln sidechain (2wfj 123). Reduce's assignment is confirmed by high-resolution electron density that is stronger for the O atom than for the N atom, by difference peaks (*blue*) for the NH_2 hydrogens, and by H-bonds to surrounding O atoms.

with neighbors, not by electrostatic calculations. The resulting system is quite accurate as judged by atomic-resolution electron density and H difference peaks (**Figure 10**) and by residual dipolar coupling NMR data (29). The original brute-force enumeration of the local networks was hugely sped up using dynamic programming techniques by our computer science collaborators in Jack Snoeyink's lab at UNC (16), allowing routine use even for very large structures.

Once all H atoms are present, the all-atom contacts are calculated, quantified, and visualized by Probe as small paired patches of dot surface: Hot pink spikes make steric clashes look appropriately disturbing, while the soft green and blue dots of good van der Waals and H-bonds signal comfortably interdigitated packing. Residues with bad steric clashes (>0.4 Å overlap) typically have outliers by other criteria as well, and they are candidates for demotion from signal down into noise.

The all-atom contacts were indeed very effective at analyzing the detailed fit inside folded proteins, and by then nearly all protein designers had added explicit hydrogens. However, we discovered that there were at least a few impossible things in nearly all the crystal structures, which impeded such analysis. It turned out that our new method was very powerful for fixing as well as diagnosing the problems, and thus we became hooked on a rather peculiar, quite unpopular, and nearly unfundable obsession with improving the accuracy of all the models in the PDB. It works best, of course, on the new ones and is most popular with new crystallographers, so its use grows with considerable momentum.

MolProbity

We began our quest for model improvement by working with the SouthEast Structural Genomics Collaboratory on about 30 of their high-throughput pipeline structures, showing that it was feasible to lower the bad clashes and the Ramachandran and rotamer outliers by factors of 5 or 10

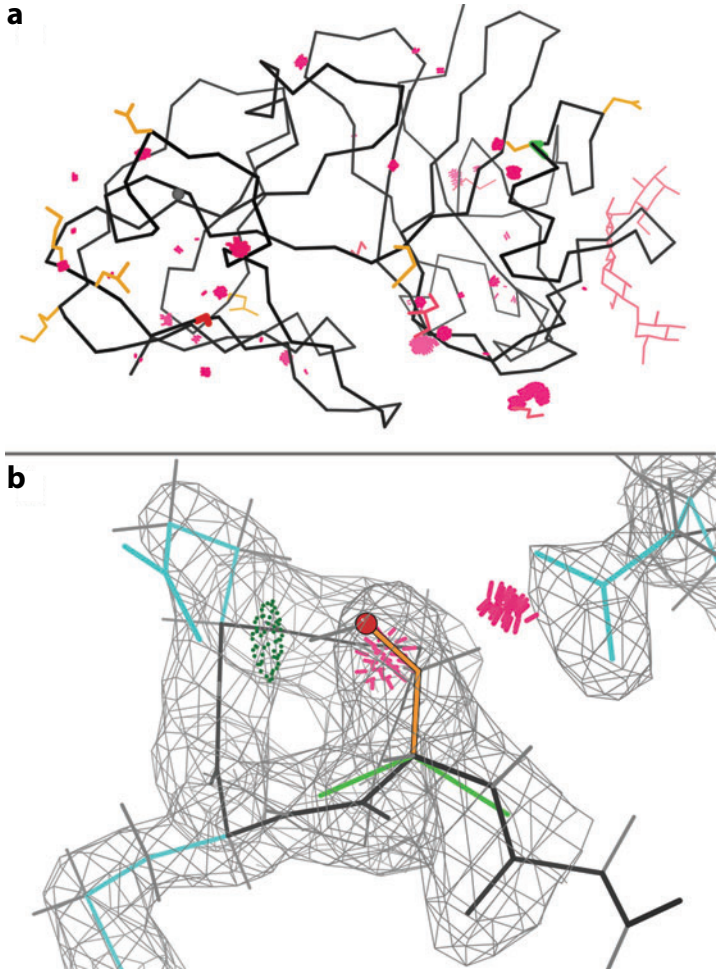

Figure 11

MolProbity 3D graphics output for 4fe9, an above-average recent structure at 2 Å resolution. (*a*) Multicriterion kinemage overview of the C-terminal two domains, with Cα backbone, bound carbohydrate (*pink*), bad steric clashes (*hot pink spikes*), and rotamer outliers (*gold*). (*b*) Close-up of a correctable clash, rotamer, and Ramachandran outlier: The Ser 341 Ncap residue, fit backward into good but ambiguous electron density.

while maintaining or improving R and Rfree (3). We learned a lot about what strategies worked at what resolutions, and we got further ideas such as diagnosing bound ions and problems at chain termini (e.g., 1LPL versus 1TOV).

During his lab rotation with us in 2002, Ian Davis created the initial MolProbity website to help others diagnose and correct local problems in their protein crystal structures using all-atom contacts (17). From the start, NQH flips were automatically corrected, but with 3D graphics comparisons in KiNG which let users see the evidence and choose whether to accept the changes. The service was gradually improved in convenience and completeness, with updated versions of rotamer (37), Ramachandran (36), and geometry criteria, percentile scores, and expansion to evaluations of NMR and RNA structures (10, 16). **Figure 11*a*** shows a multicriterion kinemage

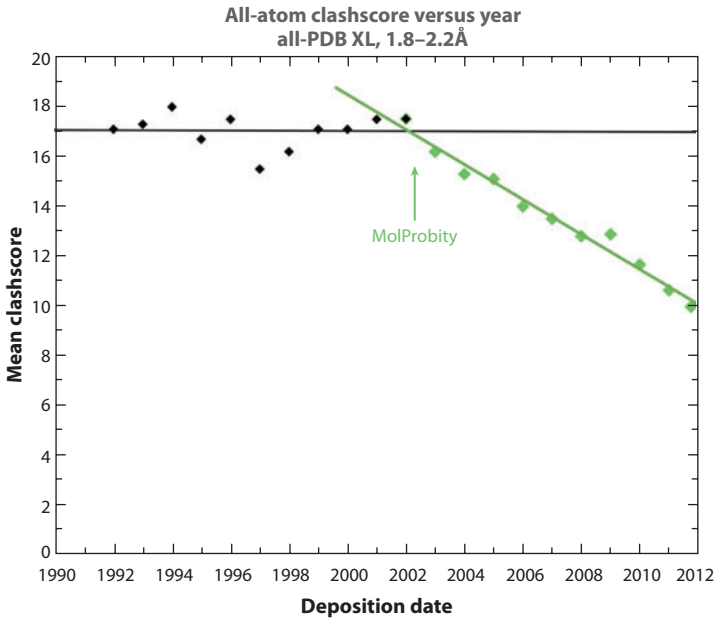

Figure 12

The effect of MolProbity validation on improving clashscores of worldwide depositions to the PDB.

from MolProbity for a recent, above-average quality structure at 2 Å resolution, and **Figure 11*b*** shows a close-up of a correctable misfit Ser that would not have been obvious with earlier tools. Jeff Headd helped implement automated correction of such problem rotamers in general, first using Coot (21) with Reduce and Probe and then, along with Pavel Afonine, in Phenix refinement.

Usage has grown exponentially, among both structural biologists (about 80%) and end users of the structures (the other 20%). Many of the MolProbity criteria are included in the recommendations of the X-ray Validation Task Force (44) now being implemented by the wwPDB (worldwide Protein Data Bank). We are most surprised and gratified by the fact that new PDB depositions worldwide have steadily improved since 2002, NQH flips by about 45% and clashscores by 40%, as plotted in **Figure 12**.

We have gradually been developing what we think of as "The Zen of Model Anomalies," a guide to interpretation for the end user of macromolecular structures and a philosophy of practice for the structural biologist:

- Consider each outlier and correct most.
- Treasure the valid, meaningful few.
- Don't fret over a small inscrutable remainder.

RNA BACKBONE

With the background of our interest in the tRNAs and the early hint that hydrogen contacts would be very revealing for nucleic acids (84), we were delighted when Laura Murray chose to pursue work on RNA structure in our lab. We spent a couple of years getting up to speed on what was already known (mostly about base interactions) and applying our all-atom contacts to diagnose which parts of the backbone had problems and which were reliable. She was then able to show that RNA backbone has a rotameric character (40), with distinct combinations of its many

torsion angles. She defined the "suite" division of the backbone that goes from sugar to sugar (as contrasted with the traditional nucleotide division from phosphate to phosphate), needed because the torsion angles are much more highly correlated within the suite. That is surely one of the most adroit puns in the scientific literature. Working with a group of labs in the RNA Ontology Consortium, we jointly came up with a nomenclature and a set of 54 backbone conformers that fit with everyone's different analyses (66). The social aspects of achieving such a community consensus were even trickier than the considerable scientific issues: For instance, a key point was choosing a new nomenclature (two-character number-letter, with !! for outliers) different from what any of us had previously used.

Our lab also developed a system for diagnosing RNA ribose pucker from the best-seen features, the phosphates and the bases. That pucker analysis, the backbone conformers, and the all-atom clashes guided our manual corrections of RNA models and led collaboratively to a series of increasingly effective automated software: RNABC (83), RCrane (31), and Erraser (12). **Figure 13** shows an example of a problem area in RNA backbone corrected by Erraser, from our collaborators in the Das lab at Stanford, which combines MolProbity diagnosis, Rosetta sampling and remodeling, and Phenix refinement [see also the review by Adams et al. (2) in this volume]. We've set ourselves a quixotic quest of producing one reliably accurate reference model for each of the major ribosome structures. As well as being of great biological importance, they constitute two-thirds of the data for RNA structural bioinformatics. Laura, Jane, Gary Kapral, and Vincent Chen have made a good start (19), and we are encouraged that with the new computational tools linked into Phenix (12, 18) we might actually achieve MolProbity-satisfactory models for ribosomes within our lifetime. Studying RNA in addition to protein structure has been very educational as well as appealing. They share most characteristics, such as complex tertiary structure, highly specific binding, and catalysis (as opposed to DNA, which is biologically central and structurally somewhat boring), but are different enough to illuminate one's unrecognized assumptions.

CURRENT PREOCCUPATIONS: ENSEMBLES AND LOW RESOLUTION

MolProbity has already proven effective across the mid-resolution range. With the help of the other Phenix development teams, we are currently working to extend similar benefits to the more difficult cases of RNA (above), multiple conformations at very high resolution, and especially to resolutions poorer than ~3 Å, which are typical for the exciting structures of big complexes and molecular machines. We have reanalyzed the bond length and van der Waals parameters for H atoms, which produces somewhat more accurate results, consistency between MolProbity and Phenix, and slightly lower clashscores for everyone! The huge expansion in deposited structures has enabled our new MySQL database of nearly 2 million residues in the "Top8000" proteins chains with better than 2 Å resolution, MolProbity score <2.0, geometry outliers <5%, and best in their 70% homology cluster. That wealth lets us quality-filter more rigorously, divide into more categories, and/or try additional dimensions (63), and we are enjoying the exploration of new relationships and motif clusters that have emerged.

Atomic resolution is in most respects the crystallographer's dream—gloriously clear maps with each atom distinct in all the well-ordered regions, and even difference peaks for many of the hydrogens. But in other ways it can be a nightmare, because you see too much—especially hard-to-disentangle overlapping density for alternative conformations in the not-so-well-ordered regions. There are several layers of difficulties. It is very hard to identify multiple models that are each physically reasonable and that jointly account for the density. Much more complete sampling of the possibilities is almost certainly key for this step. When alternates near each other interact, the diffraction data do not directly tell you which ones go together, so we are developing tools that

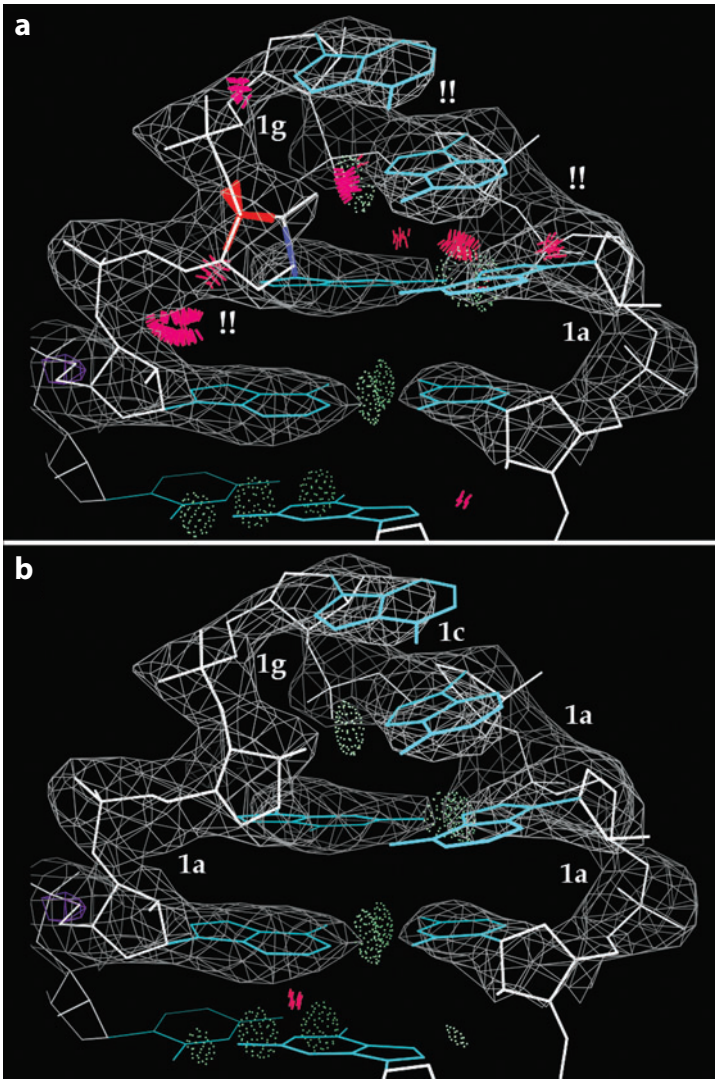

Figure 13

Automated correction of a GNRA tetraloop in the 2gis riboswitch. (*a*) As deposited, with clashes, bond angle, and backbone rotamer (!!) outliers. (*b*) After Erraser.

use all-atom contacts and relative occupancies to make each model (alt A, alt B, etc.) internally consistent. At the final step of refinement and output, current protocols return to a single model too soon at each end of the alternate regions, forcing quite distorted geometry not supported by the data (**Figure 14**). This problem is definitely correctable if we can persuade other people to adopt new procedures. Lindsay Deis has gotten the lab back into collecting our own high-resolution crystallographic data, and she and Dan Keedy plan to unify bottom-up crystallographic approaches of handling just a few alternate conformations with top-down treatments that generate very large conformational ensembles (34, 69; P. Gros, personal communication), in both cases aiming for an optimal size of ensemble that collectively fits the data.

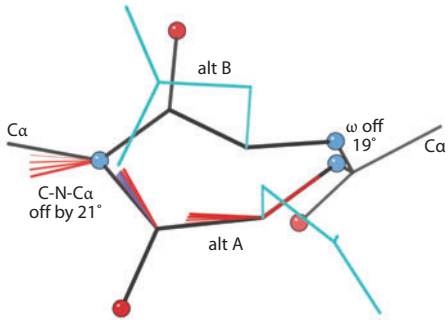

Figure 14

A typical case of alternate conformations ended too soon (1w0n Asn 42), producing unnecessary bad bond lengths, angles, and dihedrals.

The problems at high resolution are not simple to deal with, but they mainly amount to very complex bookkeeping. The problems at low resolution are more fundamental and less well understood. Not only is there much less experimental data, but past a resolution of ~3 Å there seem to be confusing patterns in the electron density that mislead both people and programs into systematic errors such as scrunching all sidechain atoms down into the small nubbins of density, connecting across H-bonds rather than along backbone, and fitting peptides in impossible orientations. We also find that parts of the sequence fit out of register are disturbingly common. After trying many possible leads, our low-resolution team of Christopher Williams and Bradley Hintze is concentrating on a new parameter space called CaBLAM. It uses a pair of Cα virtual dihedral angles to diagnose what secondary structure or motif is likely to be the real answer (**Figure 15**, left) and a CO-CO virtual dihedral to show where the fitting is incorrect (**Figure 15**, right) [see also the review by Adams et al. (2) in this volume]. In most cases, CaBLAM is proving remarkably effective at pulling reliable signal out of quite poor models.

One final preoccupation is what we think of as putting ourselves out of business. Crystallographers have now become quite dependent on MolProbity, but no single lab-maintained website is dependably stable over time. Therefore, we are working to get our methodology stolen or adapted into as many other systems as we can. Most notably, it is now tightly integrated into Phenix (1, 2).

PERSONAL BIOGRAPHIES

Family and School

David Claude Richardson grew up in rural Delaware County, Pennsylvania, near Philadelphia, and Jane Shelby Richardson grew up in Teaneck, New Jersey, near New York City, so they each had the combined advantages of woods and streams to explore on one hand and science and art museums on the other. Their fathers were a doctor in general practice and at Elwyn School (Claude E. Richardson) and an engineer in early color television at NBC (Robert E. Shelby), respectively, and their mothers a professional artist (Anne C. Richardson) and a teacher of English composition and literature (Marian E. Shelby). Each of them had an older sister. Rose

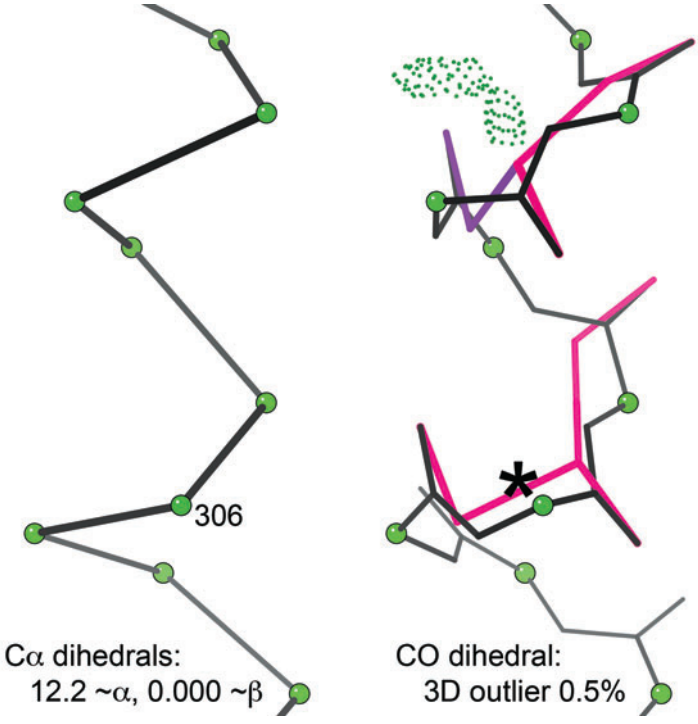

Figure 15

CaBLAM markup in distorted helix (2o01, 3.4 Å). Left: Cα trace. Right: Full backbone outlier CO dihedrals in pink (worst) and purple. Cα dihedrals score as highly α-like for residues 303–310, despite five of the eight COs pointed backward or at 90° (numerical scores given for Ile 306 as an example).

Richardson Olver was the first female professor at Amherst College, married to John W. Olver, US Congressman from western Massachusetts since 1991. Barbara Shelby Merello served in the US Information Service in Brazil, Costa Rica, Spain, Argentina, and Peru, translated novels from the Portuguese, and married Agustin Merello, a futurist from Argentina. Dave and Jane have two children: Robert D. Richardson is an engineer at Lockheed, living in Oakland, California, and Claudia J. Richardson is the bread baker at Fearrington House, living in Durham, North Carolina.

Dave collected minerals and explored the local landscape throughout school, rewarded with the changing seasons, the behavior patterns of frogs and beef cattle, and the charms and complexities of crystals. He got a head start in chemistry from exhibits at the Franklin Institute and from helping Swarthmore's Professor Gil Haight set up his theatrical general public lectures in chemistry at the Wagner Free Institute of Science in downtown Philadelphia. Jane was an amateur astronomer, acquiring a love of natural history, observation, and instrumentation from counting meteors, chasing eclipses, grinding and polishing a 6″ telescope mirror and machining its mount, and plotting naked-eye observations of Sputnik. She still watches for Perseid meteors on hiking trips. Jane's high school crowd helped her count meteors and track Sputnik. Dave's high school crowd loved to talk about science but got rained out when they first tried to spot Sputnik. Both families took long car trips to scenic parts of the country, the Richardsons toured

Europe, and Jane spent a summer in Rio de Janeiro with her sister. Dave developed his general talent for fixing things on projects with his father (later applied to lab equipment and software debugging). Jane learned writing from her mother and math and carpentry from her father, but unfortunately did not pick up even a hint of her sister's linguistic talent.

Swarthmore and MIT

Dave and Jane met as freshmen at Swarthmore College, when Dave deliberately started bringing his lunch to the physics (and science fiction) library where Jane hung out. Both of us loved the intellectual environment and beautiful surroundings of Swarthmore. Jane switched from physics and astronomy to philosophy, with a minor in math and physics. She thrived in the intense discussions, all-day labs, and week-long schedule of the honors program. However, a probably undesirable side effect was that she learned how to stay up all night researching and writing an honors paper and then to present it fairly coherently before collapsing the next day. Dave especially enjoyed the senior-year revisit of freshman inorganic chemistry, comparative anatomy, and biology classes, with tales of a sloth migrating slowly around Professor Enders' shoulders and taking a swipe at his formal dinner plate every few minutes, and a physiology lab that led to a senior-year chemistry presentation on implications of the exciting, brand-new crystal structures of myoglobin and hemoglobin.

After graduating in 1962 we went together up to Boston for graduate school, Dave at MIT and Jane at Harvard, and were married on Groundhog Day 1963. Dave joined Al Cotton's lab to do structural inorganic chemistry but happily ended up doing a protein crystal structure—at least the nuclease had calcium bound at the active site! It turned out that the excellent philosophy department at Harvard specialized in things very different from Jane's interests, and she decided to leave with an MA after a year. (She got a great deal out of courses in the botany department, meanwhile.) After trying some other options, she joined Dave's project as a technician, and the combination was so synergistic that we've worked together ever since. The only real frustration in such a joint professional career is that it's very hard for others to accept that two people can truly complement one another with different but equally essential contributions to their joint productivity and innovation.

Hobbies

Aside from macromolecular structure, our most consistent and engrossing hobbies have been backpacking, house building, travel, and photography. One of the great side benefits of working in science is that it is thoroughly international. As well as (nowadays) easily collaborating with colleagues anywhere, we've gotten the chance to visit many fascinating places such as Russia, China (**Figure 16**), India, Australia, and much of Europe. On the ends of meetings, we almost always take a few days to explore.

We have (literally) built our own house twice, first north of Durham on the bank above the Little River, and then, after that was taken by eminent domain for a reservoir, on an enclave in Duke Forest (the Duke Forestry School's research forest) partway between Durham and Chapel Hill. The geodesic-dome half of the house looks like an icosahedral virus, and the cantilevered-octagon treehouse half (**Figure 16**) looks like a T-even virus from the side. North Carolina, as well as the Sierras, has many elegant flowers to photograph, such as the magnolia just outside our lab (**Figure 16**).

Nearly every year since 1967 we've gone backpacking on or near the John Muir Trail in the Sierra Nevada. As well as the exercise, the gorgeous scenery, and the wildflowers, we also

Figure 16

Top left: The Great Wall of China at Badaling. Bottom left: The Richardson's octagonal treehouse in Duke Forest. Bottom right: Magnolia bloom at Duke. Top right: Steelhead Lake and Conness Glacier, Hoover Wilderness, California. By J.S. Richardson and D.C. Richardson.

find those trips essential for the perspective and the new ideas we can get after the first week or so. **Figure 16** (top right) shows a classic High Sierra view past whitebark pines, in the Hoover Wilderness north of Tioga Pass. We have a cabin at 7,000 ft on the east side of the Sierras, shared with quail and foxes, Jeffrey and pinyon pines, sagebrush, and stream orchids. It's a perfect basecamp for backpacking and day hikes and has wonderful views from the porch to help inspire our writing, programming, and molecule-gazing.

TAKE-HOME MESSAGES

Protein and RNA structures are inherently elegant.
Collaboration is better than competition.
The model is not the molecule (**Figure 17**).
Science is fun, especially the details.

We believe in working on what looks most intriguing and productive, enjoy methods development, and prefer projects that other people don't yet see as both important and possible, so that they might not be done if we didn't do them. Therefore we've reinvented our research emphasis many times: from protein crystallography to structural bioinformatics, to protein design, to molecular representation and computer graphics, to lowest-level details of atomic packing, to RNA backbone and ribosomes, to model validation and improvement, to crystallography again. However, within that rather random-looking trajectory we think there is a consistent signal of unifying themes: fascination with the determinants of macromolecular 3D structure, and

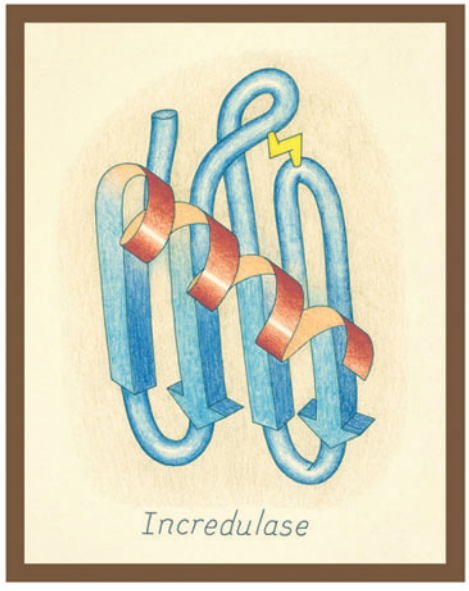

Figure 17

The model is not the molecule, but it should be "protein-like" (or RNA-like, . . .).

seeking new understanding largely by redefining the boundary between signal and noise, in either direction.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

Our work has been supported over the years almost entirely by research grants from NIGMS, NIH, including over 30 years on R01-GM15000 and currently GM073919, GM073930, GM088674, and P01-GM063210 (Phenix). We treasure our many excellent collaborators and the wonderful set of people who have come through our lab: students, postdocs, and long-term support staff like Lizbeth Videau, expert on data analysis, *cis* prolines, and culinary issues; Michael Prisant, expert on chemical physics, Unix, and historical scientific literature; and Bryan Arendall, expert on database and web support, crystallography, and philosophical questions. We thank Rita Lo for permission to use the frontispiece photo. **Figures 2*b***, **6**, **7*a***, **10–13**, and **15** were prepared for this chapter. **Figures 1** (lower left), **2*a***, **3**, **5*c***, **7*b***, **8**, **9**, **14**, **16**, and **17** are available on Wikimedia Commons. In limited space we have followed a multithreaded path, managing to mention only some of the people very important to us. Thank you all!

LITERATURE CITED

1. Adams PD, Afonine PV, Bunkóczi G, Chen VB, Davis IW, et al. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* 66:213–21

2. Adams P, Baker D, Brunger AT, Das R, DiMaio F, et al. 2013. Advances, interactions, and future developments in the CNS, Phenix, and Rosetta structural biology software systems. *Annu. Rev. Biophys.* 42:265–87
3. Arendall WB III, Tempel W, Richardson JS, Zhou W, Wang S, et al. 2005. A test of enhancing model accuracy in high-throughput crystallography. *J. Struct. Funct. Genomics* 6:1–11
4. Arnone AA, Bier CJ, Cotton FA, Day VW, Hazen EE Jr, et al. 1971. A high resolution structure of an inhibitor complex of the extracellular nuclease of *Staphylococcus aureus*. I. Experimental procedures and chain tracing. *J. Biol. Chem*. 246:2303–16
5. Arnone AA, Bier CJ, Cotton FA, Hazen EE Jr, Richardson DC, Richardson JS. 1969. The extracellular nuclease of *Staphylococcus aureus*: structures of the native enzyme and an inhibitor complex at 4 Å resolution. *Proc. Natl. Acad. Sci. USA* 64:420–27
6. Beem KM, Richardson DC, Rajagopalan KV. 1977. Metal sites of Cu,Zn superoxide dismutase. *Biochemistry* 16:1930–36
7. Block JN, Zielinski DJ, Chen VB, Davis IW, Vinson EC, et al. 2009. KinImmerse: macromolecular VR for NMR ensembles. *Source Code Biol. Med.* 4:3
8. Brooks FP Jr. 1977. The computer scientist as toolsmith: studies in interactive computer graphics. In *Information Processing* 77, *Proceedings of IFIP Congress* 77, ed. B Gilchrist, pp. 625–34. Toronto: North Holland
9. Chen C-Y, Georgiev I, Anderson AC, Donald BR. 2009. Computational structure-based redesign of enzyme activity. *Proc. Natl. Acad. Sci. USA* 106:3764–69
10. Chen VB, Arendall WB III, Headd JJ, Keedy DA, Immormino RM, et al. 2010. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* 66:12–21
11. Chen VB, Davis IW, Richardson DC. 2009. KiNG (Kinemage, Next Generation): a versatile interactive molecular and scientific visualization program. *Protein Sci.* 18:2403–9
12. Chou F-C, Sripakdeevong P, Dibrov SM, Hermann T, Das R. 2013. Correcting pervasive errors in RNA crystallography through enumerative structure prediction. *Nat. Methods* 10:74–76
13. Cotton FA, Hazen EE Jr, Richardson DC. 1966. Crystalline extracellular nuclease of *Staphylococcus aureus*. *J. Biol. Chem*. 241:4389–90
14. Dahiyat BI, Sarisky CA, Mayo SL. 1997. De novo protein design: towards fully automated sequence selection. *J. Mol. Biol.* 273:789–96
15. Davis IW, Arendall WB III, Richardson JS, Richardson DC. 2006. The backrub motion: How protein backbone shrugs when a sidechain dances. *Structure* 14:265–74
16. Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, et al. 2007. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* 35:W375–83
17. Davis IW, Murray LW, Richardson JS, Richardson DC. 2004. MolProbity: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res.* 32:W615–19
18. DiMaio F, Tyka M, Baker M, Chiu W, Baker D. 2009. Refinement of protein structures into low-resolution density maps using Rosetta. *J. Mol. Biol.* 392:181–90
19. Dunkle JA, Wang L, Feldman MB, Pulk A, Chen VB, et al. 2011. Structures of the bacterial ribosome in classical and hybrid states of tRNA binding. *Science* 332:981–84
20. Eisenberg D, Wilcox W, Eshita SM, Pryciak PM, Ho SP, DeGrado WF. 1986. The design, synthesis and crystallization of an alpha helical peptide. *Proteins: Struct. Funct. Genet.* 1:16–22
21. Emsley P, Lohkamp B, Scott WG, Cowtan K. 2010. Features and development of Coot. *Acta Crystallogr. D* 66:486–501
22. Erickson BW, Daniels SB, Reddy PA, Unson CG, Richardson JS, Richardson DC. 1986. Betabellin: an engineered protein. In *Computer Graphics and Molecular Modeling*, ed. R Fletterick, M Zoller, pp. 53–57. New York: Cold Spring Harb. Lab.
23. Fridovich I. 2001. Reflections of a fortunate biochemist. *J. Biol. Chem.* 276:28629–36
24. Gainza P, Roberts KE, Georgiev I, Lilien RH, Keedy DA, et al. 2013. OSPREY: protein design with ensembles, flexibility, and provable algorithms. *Methods Enzymol.* 523:87–107
25. Gernert KM, Surles MC, LaBean TH, Richardson JS, Richardson DC. 1995. The Alacoil: a very tight antiparallel coiled-coil of α-helices. *Protein Sci.* 4:2252–60

26. Getzoff ED, Tainer JA, Weiner PK, Kollman PA, Richardson JS, Richardson DC. 1983. Electrostatic recognition between superoxide and Cu,Zn superoxide dismutase. *Nature* 306:287–90
27. Hecht HM, Ogden RM, Richardson JS, Richardson DC. 1990. De novo design, expression, and characterization of Felix: a 4-helix bundle protein of native-like sequence. *Science* 249:884–91
28. Hemmingsen JM, Gernert KM, Richardson JS, Richardson DC. 1994. The tyrosine corner: a feature of most Greek key β-barrel proteins. *Protein Sci.* 3:1927–37
29. Higman VA, Boyd J, Smith LJ, Redfield C. 2004. Asparagine and glutamine side-chain conformation in solution and crystal: a comparison for hen egg-white lysozyme using residual dipolar couplings. *J. Biomol. NMR* 30:327–46
30. Kapp GT, Richardson JS, Oas TG. 2004. Kinetic role of helix caps in protein folding is context-dependent. *Biochemistry* 43:3814–23
31. Keating KS, Pyle AM. 2012. RCrane: semi-automated RNA model building. *Acta Crystallogr. D* 68:985–95
32. Keedy DA, Georgiev I, Triplett EB, Donald BR, Richardson DC, Richardson JS. 2012. The role of local backrub motions in evolved and designed mutations. *PLoS Comput. Biol.* 8:e1002629
33. Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D. 2003. Design of a novel globular protein fold with atomic-level accuracy. *Science* 302:1364–68
34. Lange PT, Ng H-L, Fraser JS, Corn JE, Echols N, et al. 2010. Automated electron-density sampling reveals widespread conformational polymorphism in proteins. *Protein Sci.* 19:1420–31
35. Leaver-Fay A, O'Meara MJ, Tyka M, Jacak R, Song Y, et al. 2013. Scientific benchmarks for updating the Rosetta energy function. *Methods Enzymol.* 523:109–43
36. Lovell SC, Davis IW, Arendall WB III, de Bakker PIW, Word JM, et al. 2003. Structure validation by Cα geometry: ϕ,ψ and Cβ deviation. *Proteins: Struct. Funct. Genet.* 50:437–50
37. Lovell SC, Word JM, Richardson JS, Richardson DC. 2000. The penultimate rotamer library. *Proteins: Struct. Funct. Genet.* 40:389–408
38. McCord JM, Fridovich I. 1969. Superoxide dismutase. An enzymatic function for erythrocuprein (hemocuprein). *J. Biol. Chem.* 244:6049–55
39. Merritt EA, Bacon DJ. 1997. Raster3D photorealistic molecular graphics. *Methods Enzymol.* 277:505–24
40. Murray LW, Arendall WB III, Richardson DC, Richardson JS. 2003. RNA backbone is rotameric. *Proc. Natl. Acad. Sci. USA* 100:13904–9
41. Murzin AG. 1993. OB(oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences. *EMBO J.* 12:861–67
42. Murzin AG. Fold: OB-fold category, in the SCOP (Structural Classification Of Proteins) database. **http://scop.mrc-lmb.cam.ac.uk/scop/data/scop.b.c.hb.html**
43. Presta LG, Rose GD. 1988. Helix signals in proteins. *Science* 240:1632–41
44. Read RJ, Adams PD, Arendall WB III, Brunger AT, Emsley P, et al. 2011. A new generation of crystallographic validation tools for the Protein Data Bank. *Structure* 19:1395–412
45. Regan L, DeGrado WF. 1988. Characterization of a helical protein designed from first principles. *Science* 241:976–78
46. Richards FM. 1968. The matching of physical models to three-dimensional electron-density maps: a simple optical device. *J. Mol. Biol.* 37:225–30
47. Richards FM. 1988. Public access to X-ray diffraction data. *J. Comput. Aided Mol. Des.* 2:3–4
48. Richards FM. 1997. Whatever happened to the fun? An autobiographical investigation. *Annu. Rev. Biophys. Biomol. Struct.* 26:1–25
49. Richardson DC, Richardson JS. 1992. The kinemage: a tool for scientific illustration. *Protein Sci.* 1:3–9
50. Richardson DC, Richardson JS. 2002. Teaching molecular 3-D literacy. *Biochem. Mol. Biol. Educ.* 30:21–26
51. Richardson JS. 1976. Handedness of crossover connections in beta sheets. *Proc. Natl. Acad. Sci. USA* 73:2619–23
52. Richardson JS. 1977. Beta sheet topology and the relatedness of proteins. *Nature* 268:495–500
53. Richardson JS. 1981. The anatomy and taxonomy of protein structures. *Adv. Protein Chem.* 34:167–339
54. Richardson JS. 1985. Schematic drawings of protein structures. *Methods Enzymol.* 115:341–58
55. Richardson JS. 2000. Early ribbon drawings of proteins. *Nat. Struct. Biol.* 7:624–25
56. Richardson JS, Getzoff ED, Richardson DC. 1978. The beta bulge: a common small unit of non-repetitive protein structure. *Proc. Natl. Acad. Sci. USA* 75:2574–78

57. Richardson JS, Richardson DC. 1987. Some design principles: Betabellin. In *Protein Engineering*, ed. D Oxender, CF Fox, pp. 149–63. New York: Liss
58. Richardson JS, Richardson DC. 1988. Amino acid preferences for specific locations at the ends of alpha helices. *Science* 240:1648–52
59. Richardson JS, Richardson DC. 1988. Helix lap-joints as ion-binding sites: DNA-binding motifs and Ca-binding E-F hands are related by charge and sequence reversal. *Proteins: Struct. Funct. Genet.* 4:229–39
60. Richardson JS, Richardson DC. 1989. Protein origami. In *Protein Folding: Deciphering the Second Half of the Genetic Code*, ed. L Geirasch, J King, pp. 5–17, 304–9. Washington, DC: Amer. Assoc. Adv. Sci.
61. Richardson JS, Richardson DC. 2002. Natural β-sheet proteins use negative design to avoid edge-to-edge aggregation. *Proc. Natl. Acad. Sci. USA* 99:2754–59
62. Richardson JS, Richardson DC. 2013. Studying and polishing the PDB's macromolecules. *Biopolymers* 99:170–82
63. Richardson JS, Richardson DC. 2013. The Plot thickens: more data, more dimensions, more uses. In *Biomolecular Forms and Functions: A Celebration of 50 Years of the Ramachandran Map*, ed. M Bansal, N Srinivasan, pp. 46–61. Singapore: World Scientific
64. Richardson JS, Richardson DC, Thomas KA, Silverton EW, Davies DR. 1976. Similarity of three-dimensional structure between the immunoglobulin domain and the Cu,Zn superoxide dismutase subunit. *J. Mol. Biol.* 102:221–35
65. Richardson JS, Richardson DC, Tweedy NB, Gernert KM, Quinn TP, et al. 1992. Looking at proteins: representations, folding, packing, and design. Biophysical Society National Lecture, 1992. *Biophys. J.* 63:1186–209
66. Richardson JS, Schneider B, Murray LW, Kapral GJ, Immormino RM, et al. 2008. RNA backbone: consensus all-angle conformers and modular string nomenclature. *RNA* 14:465–81
67. Richardson JS, Thomas KA, Richardson DC. 1975. Alpha-carbon coordinates for bovine Cu,Zn superoxide dismutase. *Biochem. Biophys. Res. Commun.* 63:986–92
68. Richardson JS, Thomas KA, Rubin BH, Richardson DC. 1975. Crystal structure of bovine Cu,Zn superoxide dismutase at 3 Å resolution: chain tracing and metal ligands. *Proc. Natl. Acad. Sci. USA* 72:1349–53
69. Richter B, Gsponer J, Varnai P, Salvatella X, Vendruscolo M. 2007. The MUMO (minimal under-restraining minimal over-restraining) method for the determination of native state ensembles of proteins. *J. Biomol. NMR* 37:117–35
70. Rubin BH, Richardson JS. 1972. The simple construction of protein alpha-carbon models. *Biopolymers* 11:2381–85
71. Scavetta RD, Herron SR, Hotchkiss AT, Kita N, Keen NT, et al. 1999. Structure of a plant cell wall fragment complexed to pectate lyase C. *Plant Cell* 11:1081–92
72. Schellman C. 1980. The alpha-L conformation at the ends of helices. In *Protein Folding*, ed. R Jaenicke, pp. 53–61. New York: Elsevier/North Holland
73. Shortle D. 1992. Mutational studies of protein structures and their stabilities. *Q. Rev. Biophys.* 25:205–50
74. Sibanda BL, Thornton JM. 1985. β-hairpin families in globular proteins. *Nature* 316:170–74
75. Steinman HM, Naik VR, Abernethy JL, Hill RL. 1974. Bovine erythrocyte superoxide dismutase. Complete amino acid sequence. *J. Biol. Chem.* 249:7326–38
76. Sternberg MJE, Thornton JM. 1977. On the conformation of proteins: the handedness of the connection between parallel beta-strands. *J. Mol. Biol.* 110:269–83
77. Surles MC, Richardson JS, Richardson DC, Brooks FP Jr. 1994. Sculpting proteins interactively: continual energy minimization embedded in a graphical modeling system. *Protein Sci.* 3:198–210
78. Tainer JA, Getzoff ED, Beem KM, Richardson JS, Richardson DC. 1982. The determination and analysis of the 2 Å structure of copper, zinc superoxide dismutase. *J. Mol. Biol.* 160:181–217
79. Thomas KA, Rubin BH, Bier CJ, Richardson JS, Richardson DC. 1974. The crystal structure of bovine Cu^{2+},Zn^{2+} superoxide dismutase at 5.5-Å resolution. *J. Biol. Chem.* 249:5677–83
80. Tyka MD, Keedy DA, Andre I, DiMaio F, Song Y, et al. 2011. Alternate states of proteins revealed by detailed energy landscape mapping. *J. Mol. Biol.* 405:607–18
81. Uversky VN, Dunker AK. 2012. Multiparametric analysis of disordered proteins: looking at intrinsic disorder through compound eyes. *Anal. Chem.* 84:2096–104

82. Videau LL, Arendall BW III, Richardson JS. 2004. The *cis* Pro touch-turn: a rare motif preferred at functional sites. *Proteins: Struct. Funct. Genet.* 56:298–309
83. Wang X, Kapral G, Murray L, Richardson D, Richardson J, Snoeyink J. 2008. RNABC: forward kinematics to reduce all-atom clashes in RNA backbone. *J. Math. Biol.* 56:253–78
84. Word JM, Lovell SC, LaBean TH, Zalis ME, Presley BK, et al. 1999. Visualizing and quantitating molecular goodness-of-fit: small-probe contact dots with explicit hydrogen atoms. *J. Mol. Biol.* 285:1711–33
85. Word JM, Lovell SC, Richardson JS, Richardson DC. 1999. Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. *J. Mol. Biol.* 285:1735–47
86. Wright PE, Dyson HJ. 1999. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.* 293:321–31
87. Wuthrich K. 1986. *NMR of Proteins and Nucleic Acids*. New York: Wiley
88. Wyckoff HW, Hardman KD, Allewell NM, Inagami T, Johnson LN, Richards FM. 1967. The structure of ribonuclease-S at 3.5 Å resolution. *J. Biol. Chem.* 242:3984–88

RELATED RESOURCES

Duke Biochemistry Department. **http://www.biochem.duke.edu**
MolProbity structure validation. **http://molprobity.biochem.duke.edu**
Phenix: Python-based Hierarchical Environment for Integrated Xtallography. **http://phenix-online.org**
Richardson laboratory home page. **http://kinemage.biochem.duke.edu**
RNA Ontology Consortium (ROC). **http://roc.bgsu.edu**
Structural Biology and Biophysics graduate program at Duke. **http://sbb.duke.edu**
TIM ribbon drawing as English Wikipedia picture-of-the-day for November 19, 2009. **http://en.wikipedia.org/wiki/Wikipedia:Picture_of_the_day/November_2009**
User:Dcrjsr (Jane Richardson) and User:DaveTheMage (Dave Richardson) on Wikimedia Commons, image galleries and audio. **http://commons.wikimedia.org/wiki/User:Dcrjsr**; **http://commons.wikimedia.org/wiki/User:DavetheMage**
Wikipedia article on kinemage graphics. **http://en.wikipedia.org/wiki/Kinemage**
Wikipedia article on ribbon drawings. **http://en.wikipedia.org/wiki/Ribbon_diagram**