RECOLLECTIONS

The first years of the Protein Data Bank*¹

EDGAR F. MEYER

Biographics Laboratory, Department of Biochemistry and Biophysics, Texas A&M University,
College Station, Texas 77843-2128

It is a challenge to paint portraits of yesterday with today's colors, especially when we really want to paint tomorrow's portraits, but let's try. In describing the origin of the Protein Data Bank (PDB, Bernstein et al., 1977), it is essential to consider the state of science and technology (especially protein crystallography and computing). Describing computing with decks of punched cards or using Beevers–Lipson strips to calculate structure factors is like using the vocabulary of the electronics age to describe the function of a bronze-age artifact.

The conception and development of the Protein Data Bank was directly linked to its database, crystallographically determined protein structures,² and to the availability of facile graphics hardware and software. Structural studies of nucleic acids and oligosaccharides and analysis of proteins by NMR came some years later. A

crucial component for the utilization of this database and later to the advancement of the crystallographic method was the development of interactive computer graphics³ (Hamilton, 1970). Three technologies contributed to this advancement. First was the Calcomp plotter; second, the monochromatic (black and white) cathode ray tube (CRT) display, which appeared about the same time. CRT graphics existed at first only at very few locations (e.g., Project MAC at MIT) because ad hoc software had to be written for ad hoc hardware. Plotter graphics had the greatest initial impact on structural studies for two reasons: an affordable Calcomp plotter and the mature software package, ORTEP (Johnson, 1965),⁴ still in use today! The third and most durable technology was/is color raster graphics.

The birth place and cradle of interactive graphics was Project MAC at MIT. Prof. Cyrus Levinthal (1965, 1966) and his group (Langridge & Macewan, 1965) developed software to study known molecular structures and to try to predict novel structures; MAD or FORTRAN programs were written for the "kludge." As the 1971 proceeding of the Cold Spring Harbor Symposium illustrates (Barry & North, 1972; Yankeelov & Coggins, 1972), affordable ad hoc displays were slow to appear. The Electronics Systems Laboratory (Marshall and Barry) at Washington University, St. Louis, began to develop original hardware and software, as did Tony North's group at Leeds (Barry & North, 1972). In the early 1970s, Richard Feldmann had assembled a graphics resource at the NIH and also had begun to gather atomic coordinates of proteins. His efforts included a search procedure for the Crystal Data Centre files, development of 3D teaching tools, modeling efforts, and serving the NIH community. The graphics laboratory at Princeton University, run for a time by Bob Langridge and Todd Wipke (cf. Langridge, 1988), used a PDP-10 and early Evans & Sutherland monochromatic display (the "LDS-1"), costing ca. \$700,000. In the early

Reprint requests to: Edgar F. Meyer, Biographics Laboratory, Department of Biochemistry and Biophysics, Texas A&M University, College Station, Texas 77843-2128; e-mail: meyer@bioch.tamu.edu.

Edgar F. Meyer obtained his B.S. in chemistry from North Texas State University and his Ph.D. in chemistry from the University of Texas, under the direction of Prof. Stanley H. Simonsen. He then was a postdoctoral fellow in the laboratory of Prof. Jack Dunitz at the Swiss Institute of Technology (E.T.H.), Zürich, and subsequently at M.I.T., in the laboratories of Prof. Alexander Rich and Prof. Cyrus Levinthal, where he was introduced to the "Kludge" at Project MAC. He spent his academic career thus far at Texas A&M University, first studying small molecule structures (metalloporphyrins and bile pigments) and later switching to protein-ligand complexes, especially porcine pancreatic elastase and more recently, the collagenolytic isozyme, Ht-d, from the hemorrhagic venom of the western diamondback rattlesnake, *Crotalus atrox*.

He was a Visiting Scientist at the Brookhaven National Laboratory from 1968–1974 and has been a Visiting Scientist at the Max-Planck Institut für Biochemie, Martinsried, since 1979. He has been a Visiting Professor at the University of Florence, Italy, and the E.T.H., Zürich, Switzerland. Coming from the flats of Texas, he spends his scientific life looking at molecules in 3D, using or devising programs and methods, as necessary, and directing a small but productive research group balanced between computational and experimental methods.

*This paper is dedicated to the memory of Dr. Stanley H. Simonsen, Professor of Chemistry, The University of Texas (1918–1996).

¹Presented in an abbreviated form at the XVII Congress of the International Union of Crystallography, Seattle, Washington, August, 1996.

²By 1970, solved protein structures included myoglobin, hemoglobin, lysozyme, ribonuclease, carboxypeptidase, α -chymotrypsin, papain, and subtilisin.

³Just as books printed before 1500 are considered the *incunabula*, interactive molecular graphics before 1970 may likewise be so considered.

⁴At the 1965 ACA meeting in Gatlinburg, Tennessee, Carroll Johnson offered to prepare a stereo slide, with thermal ellipsoids, from the coordinates of structures being presented. I used this opportunity to present the structure of the corrin nucleus of vitamin B₁₂, synthesized by the Eschenmoser group and determined in Jack Dunitz's laboratory at the E.T.H., Zürich (Bertelet et al., 1964).

Fig. 1. Out shopping for a 3D graphics system. Dr. Bob Sparks, Tom Workman (President of Syntex Analytical), Neville Crooke (a salesman), and Dr. Walter Hamilton (left to right) are pictured during our visit to Syntex Analytical in 1972.

1970s, commercial graphics hardware was so expensive and computer memory so limiting that ad hoc programs were necessary and systems were not exportable—this did not change until the second generation Vector General display and PDP-11 led to the development of program FRODO (Jones, 1978) at Martinsried by Alwyn Jones in the late 1970s. Cost of hardware was the main bottleneck that limited progress with these early efforts, all of which used monochromatic displays.

While a member of Levinthal's group at MIT, I wrote a program to display small molecule structures, but I had to generate the bond list by hand. After taking a faculty position at Texas A&M in 1967, my first program (UMZUG) used atomic radii and interatomic distance criteria to create FROM and RING connectivity lists similar to the descriptors used by Chemical Abstracts (Morgan, 1965). This procedure was automatic for well-refined small molecules, but the atomic positional errors of unrefined protein structures required occasional manual editing of the connectivity files.

In 1968, the third visualization technology, color raster graphics, was developed and became the predecessor of the graphics workstation. The high-energy physics laboratory at Brookhaven National Laboratory (BNL) developed the BRAD (Brookhaven Raster Display; Ophir et al., 1969) system, which drew digitized images of Hough–Powell bubble chamber photographs on a black and white TV monitor equipped with a keyboard and trackball pointer (the first mouse!). Operators at two stations worked two 8-h shifts, moving the track ball pointer along orthogonal projections of the paths of subatomic particles; the 3D trajectories were calculated off-line to try to identify elusive high-energy events. The developers of the BRAD system perceived the potential of the BRAD system for 3D graphics and linked the two display channels to the red and green circuits of a Philco color television set.

In 1968, Dr. Walter Hamilton (cf. Fig. 1; also see Addendum), in the Chemistry Department at Brookhaven, recognized the potential of the BRAD display for 3D molecular graphics and invited me to spend a summer writing a program to draw 3D pictures of molecules; my family did not see much of me that summer because the computer was free only on the third shift, from midnight to dawn. At the end of the summer, program DISPLAY (Meyer, 1968),

consisting of ca. 2,000 cards,⁵ could draw, rotate, zoom, and label molecular models in 3D. Because program DISPLAY could draw 3D models with up to 512 atoms, it could draw the trace of C_α positions of a protein or it could depict an enlarged view of a region of interest of a protein. Figure 2 shows an example of a stereo model drawn by this system.⁶

During my second visit to Brookhaven, in the summer of 1969, I wrote program LINK (Meyer, 1970b) to display small molecule structures and started using program DISPLAY to study protein structures. Benno Schoenborn provided me with an “unofficial” set of myoglobin coordinates, and the fun began. Punched cards containing the C_α and heme atom coordinates were extracted manually and input to the program; revealing the first color display of a model⁷ of the structure (Meyer, 1971). Upon visual inspection, one of the N atoms of the heme macrocycle was clearly nonplanar⁸ (Fig. 1) due to an error in one of the atomic coordinates provided to me.

Beginning with the summer of 1970, I began to solicit coordinates from various groups, e.g., ribonuclease-S (Richards et al.,

⁵This was not Project MAC; the rest of the world used punched cards or paper tape. These functions are now available at the flick of a command, knob, or icon.

⁶The full paper (Meyer, 1970a) only appeared two years later, first because a reviewer insisted that perspective be included—a feature totally ignored today—and then because the paper was somehow misplaced in moving the editorial offices of *Acta Crystallographica*.

⁷Remember that most protein structures then were made of Kendrew models (brass rods; 1.25 cm = 1 Å). Brass models took up a cubic meter of laboratory space and were dynamically unstable. Also, several developers of early graphics systems were essentially blind in one eye, so not every program pursued the 3D properties of this new medium.

⁸Because the first version of DISPLAY did not yet have moveable atoms, what should have taken 1 min at the graphics terminal took a month in batch mode, back in Texas, to add a subroutine for a least-squares plane calculation—and discover that a key punch mistake in the third significant figure (a 7 in place of a 6, as I recall) caused a 0.5 Å perpendicular displacement of an heme N atom. This alerted me to the need of an error-checking routine, as did the distance-dependent definition of bonding; refinement procedures in the 1970s were in their infancy, resulting in occasional occurrences of missing or excessive display of lines = bonds, and thus another strong visual cue (valuable even today).

Fig. 2. Reproduction of an original (1970) color transparency from the Philco TV monitor using the BRAD system to create a 3D image. Program SEARCH on the CDC 6600 used the “scoop” subroutine to load the myoglobin coordinates from the PDB, center on the Fe atom, and extract all atoms within 10 Å. A card deck of the extracted atomic coordinates was loaded into program DISPLAY and an orientation obtained by inputting Eulerian angles from the BRAD teletype. Red/green spectacles were used to obtain a 3D view, which showed that a heme N atom was misplaced. The BRAD system required ca. 1 min to create such a view, which then could be queried interactively (distances, angles, atom type, etc.) by a track-ball pointer. Circles of various radii were used to indicate non-C atoms.

1972). I also received a printed listing of C_{α} positions of subtilisin (Wright et al., 1969) from Joe Kraut’s laboratory at UCSD. Because these structures were determined in the U.S., the coordinates were measured in inches,⁹ whereas the myoglobin coordinates were in centimeters. Looking ahead, I knew that Cotton’s group at MIT was working on the structure of staphylococcal nuclease (Cotton et al., 1972); I anticipated that the MIT structure would be reported in the local system of measure, the Smoot.¹⁰ I subsequently wrote program PROIN (Meyer, 1974), which accepted a variety of formats and dimensions and was independent of crystal class.

To facilitate the automatic processing of data, a dictionary of pointers to group and atom names and aqueous characteristics (charge, hydrophobic, etc.) was assembled. John Coggins, a postdoctoral associate in Biology with Elliott Shaw at Brookhaven and now a Professor of Biochemistry at the University of Glasgow, contributed to the fundamental design of the dictionary. Numeric pointers were packed in binary as a coded word for ease of extraction (and especially to save CPU memory). John was enormously helpful in making informed decisions for setting up the protocols for input and storage. He taught me some biochemistry and I taught him to be a key-punch operator.

⁹Structures then were built of Kendrew models as a mirror image of electron density in a Richards’ Box (Richards, 1968) and atomic positions were measured individually with a traveling telescope.

¹⁰An MIT fraternity used a pledge, Oliver Smoot, to measure the length of Harvard Bridge from Boston to Cambridge over the Charles River; the distance is 364.4 Smoots and 1 ear (or 2,165 feet).

Programmers today may not appreciate that the largest CPU memory (magnetic cores) available in 1970 was 32K words for storage of program + data. So program PROIN was written to pack nonstructural information in binary into control words that could be searched rapidly with shifting and logical masking operations requiring one CPU cycle, in order to retrieve residue number or type, atom type, etc. Each atom was stored as a punched card image in the 80-column format, which is still maintained as an ASCII image by the PDB, requiring 20 A4 words (or 10 A8 words on the CDC 6600), whereas PROIN achieved a fourfold saving by compressing this to five words of CPU memory for internal storage and retrieval (three words stored the x , y , z atomic coordinates and two coded binary words described the atom type, residue name, and sequence number). Walter suggested that the coordinates be orthogonalized¹¹ with the crystallographic c axis unique for display purposes. Coordinates were stored in orthogonal Å units; transformation matrices were placed in the header. Thus was the PDB begun.

This technology provided the capability to store and display the structures of a growing list of magnificent macromolecules. But diplomacy was needed to convince the crystallographic community of its merits, not of the utility of the PDB as an archival depository, but of the necessity of distributing the coordinates. Years, if not graduate-student lifetimes, were expended determining these first structures. Refinement methods in 1970 were crude and generally not available; the structures were imprecise and subject to misinterpretation. The REMARK feature of the PDB was therefore introduced to permit authors to identify regions requiring special attention—or neglect. In order to establish the PDB, acceptance by the crystallographic community was necessary, requiring a pilgrimage in 1970 to the Medical Research Council (MRC) laboratory and Crystal Data Centre (CDC) in Cambridge. One result of this exchange was a concession that coordinates of protein structures would be stored in the same format as the small molecule CDC database (with a redundant ATOM label at the beginning of each card), retaining the now-arcane counting number at the end. But the idea of a PDB was accepted by Professors Perutz, Blow, Kennard, Diamond, and colleagues in Cambridge.

To this point, all available display terminals (with the exception of the BRAD system) were monochromatic and very expensive. At the American Crystallographic Association meeting in Ottawa in 1971,¹² Syntex Analytical demonstrated a novel 3D color (512 × 512) raster display with ingenious use of 4,096 words of CPU memory (Data General Nova with a blazing 1-ms cycle time) and a disc with eight read heads, which could store programs, atomic information, and two complete, two-color stereo images, years before color raster graphics became commercially available (Table 1). This prototype (Willoughby et al., 1974) was delivered subsequently to my laboratory in Texas but, thanks primarily to

¹¹I recall distinctly the resistance I encountered from some members of the crystallographic community for wanting to use internal fractional coordinates so that the program could apply symmetry operators; is it disrespectful to refer to that time as the bronze age of protein crystallography? To the contrary, it is a tribute to the heroic efforts of these pioneers, who obtained dramatic results with “bronze-age” tools.

¹²At this same meeting, Picker X-ray demonstrated the first area-detector-like device, a TV display of a few diffraction spots. Bob Sparks remarked to me that although the Picker group didn’t know how to use this information (and shortly afterward Picker dropped their single-crystal division), Bob’s newly developed auto-indexing algorithm would make it possible to index these reflections.

Table 1. Brief history of macromolecular graphics

1964	First interactive macromolecular display (the "Kluge," Project MAC, MIT).
1965	Program ORTEP (Carroll Johnson); stereo plots with thermal ellipsoids.
1968	First use of color raster graphics, program DISPLAY (E.M.): Brookhaven National Laboratory.
1970	Syntex color raster system (<i>J Appl Cryst</i> 7:430–434): TV monitor + Data General computer: 4,096 words of memory.
1971	3D display of density for ligand fitting (Barry & North).
1975	Staphylococcal nuclease fit to density (Cotton, Hazen, Legg), program FIT (<i>Science</i> 190:1047–1053).
1976	PS2 graphics terminal from Evans & Sutherland. Monoclinic Lysozyme (Hogle & Sundaralingam).
1977	Sea snake venom toxin protein fit to density at UNC Petsko et al. (<i>Science</i> 197:1378–1381).
	Arabinose binding protein (Quioco, Gilliland, Newcomer).
1978	Program FRODO (A. Jones, <i>J Appl Cryst</i> 11:268–272). General purpose molecular modeling and refinement program—Jan Hermans' refinement algorithm. MPS color graphics terminal from Evans & Sutherland.
1981	BUILDER (Diamond; "Computational Crystallography," D. Sayer, ed., Clarendon Press, Oxford).
	GRAMPS (O'Donnel & Olson, <i>J Comp Graph</i> 15:133–142).
1982	GRINCH (Williams & Brooks).
1985	FRODO ported to the Silicon Graphics workstation (Oakley, UCSD).
	GRANNY (Connolly & Olson, <i>Comp & Chem</i> 9:1–6).
	GRID (Goodford, <i>J Med Chem</i> 28:849).
1987	Rice FRODO for the E&S PS300 (Pflugrath, Quioco, Saper).
1988	MIDAS (<i>J Mol Graph</i> 6:13–27).
	Program O (Alwyn Jones).
	Program CHAIN (Quioco & Sack).
1990	Program PRONTO (Laczowski & Swanson).

company leadership, the display was not developed commercially, despite Bob Spark's genius at creating cutting-edge technology.¹³ At that time, a few sensitive crystallographers were too easily offended by the jagged lines representing atomic bonds; the anti-aliasing algorithm had not yet been invented.

By 1970, available atomic coordinates of proteins were collected and stored with program PROIN in a uniform format at Brookhaven. They received little general use because of limited access to graphics displays; about all one could do was admire endless lists of atomic coordinates or bend brass rods to look like petrified earth worms. Therefore, the program SEARCH (Meyer, 1971) was written in FORTRAN to access the PDB and provide coordinates for off-line display. It could select a given protein from the PDB and extract atomic coordinates on the basis of atom type, residue type, or sequence range. Because 2,500 atoms would overwhelm the BRAD system,¹⁴ a novel feature permitted the user to

select a position or residue and extract (scoop) all atoms within a given radius. Today, the ease of access of structural data with World Wide Web search engines and the facility of current display software make these pioneering efforts to store and retrieve structural data appear crude, but one need only compare the illustrations in a biochemistry textbook of 1970 with those of today to see the dramatic transformation that has occurred in our lifetime.

Because of some personality conflicts at BNL, initially there wasn't the necessary harmony between some in the Biology and Chemistry Departments at Brookhaven, but recently the PDB moved to the Biology Department. There was very little protein research (i.e., no protein crystallography) being pursued in the Chemistry Department, so the PDB, like some other databases, suffered initially from a lack of local critical users and the necessary feedback they provide. Obviously, these difficulties have been overcome, but it is interesting to trace the contours of history and speculate about alternative trajectories. Even though the PDB has always been housed at Brookhaven, for a time during the early years, the only paid employee, Frances Bernstein, who has been with the PDB for almost 25 years, was a consultant for Texas A&M University, thanks to NSF funding of the CRYSTNET project. Her husband, Herbert, was extremely helpful in teaching elementary computer science to the first programmer, including the transition away from punched cards as text editors became available.

In perspective, it is revealing to see how far protein crystallography had progressed by 1971, which has been arbitrarily chosen as the beginning of the PDB, 25 years later. The Cold Spring Harbor Symposium¹⁵ of 1971 represented most of the then-current work, especially in terms of structures still being solved, over and above those already solved at that time. Earlier in 1971, the meeting of the American Crystallographic Association in North Carolina had an ad hoc session where the need for a PDB and some of the requirements of the user community were discussed. The second Alpbach meeting in early 1972 gave me the first opportunity to present the PDB informally to the European crystallographic community.

After the August 1971 ACA meeting in Iowa, Walter Hamilton and his family visited relatives in Waco, Texas, and spent the weekend with us in College Station. For some time, theoretical physicists (Professors John Nuttall and Ron Bryan) at Texas A&M, with the growing disapproval of the university administration, used a leased telephone line to Austin to compute on the CDC 6600 at the University of Texas, which was more user-friendly and computationally far superior to the "iron" from IBM here at Texas A&M. By administrative mandate, this usage was terminated that summer. But the phone company was a little slow in removing the teletype and leased line to the Physics Department. Also, Walter was exceptionally busy during my six-week visit to BNL that summer. So, on the first Monday, September 6, 1972, we went to the A&M Physics Department and used the teletype to connect to

¹³The price, ca. \$30,000, was far less than commercial monochromatic systems, and commensurate with the price of workstations today, giving an endpoint to the market projection of the linear increase of computational power over time at constant price. Although the Syntex leadership neglected to exploit this breakthrough, Bob Sparks is well known for his vision and unparalleled ability to develop computer programs for the crystallographic community.

¹⁴It took some 4 h to read the paper tapes containing the coordinates of α -chymotrypsin at 110 Baud = teletype speed (2,500 atoms $\times$ 80 ch/

atom $\times$ 8 bits/ch/110 bits/s = 4 h). A T3 Internet connection (45 Mbits/s) is 400,000 times faster and even a 28-kbit FAX/modem is 250 times faster than standard teletype I/O. But in retrospect, 110 Baud (ca. 10 characters/s) is about the response speed occasionally obtained during daytime use of the World Wide Web. Recently, on sabbatical leave in Zürich, I could access the PDB at 1 PM (7 AM in New York) at 61 bytes/s. Ibs = Infobahnstau! So 110 Baud is not as primitive as it seems.

¹⁵The PDB was not represented in the published proceedings, although it had been described clearly in published papers (Meyer, 1970a, 1970b, 1971), because my department gave me the opportunity just then to present my first lectures in undergraduate biochemistry, an offer I could not refuse.

Fig. 3. The tetrahedral simplex is used to illustrate the four necessary components of structural biochemistry available by 1972, thanks to remote access to the PDB and facile 3D graphics. The CRYSNET consortium funded by the NSF linked Brookhaven, Helen Berman's laboratory at the Institute for Cancer Research (Philadelphia), and Texas A&M by telephone lines and 4K modems. Modern developments like the World Wide Web now make these components readily available to users around the world. A contemporary version of the simplex would replace "crystallography" with "structure determination" in order to include structures from NMR and electron microscopy.

the CDC 6600 at BNL. Selecting myoglobin from the PDB with the Fe atom at the center, program SEARCH scooped out the atomic coordinates of ca. 50 heme atoms and nearby amino acids, which were printed out at 110 Baud over a distance of ca. 3,000 km. This was the genesis of the fourth apex (Fig. 3) of our technology, networking (+crystallography, graphics, the PDB). The demonstration lasted less than 30 min, but set in motion the first practical use of networking for crystallography (Meyer et al., 1974), which brought together a critical mass of programmers (Fig. 4), funds for graphics displays, and distributed use of the BNL suite of

Fig. 4. The Vector General (monochromatic) graphics terminal of the CRYSNET system is used to demonstrate graphics software being developed at Texas A&M. David Klunk (a graduate student, seated) is demonstrating a preliminary version of program FIT to Dr. Tom Koetzle, E.F.M., Drs. Herbert Bernstein and Tom Willoughby (left to right; missing are Drs. Larry Andrews and Carl Morimoto); the photo was taken at Brookhaven the summer of 1973.

crystallographic software and support for the PDB.¹⁶ This was the first use of networking and remote, non-textual information retrieval in the chemical and life-sciences.

As an occasional contributor to the PDB, and likewise as an avid user, I am pleased to see the wealth of information that has been amassed under one roof. Recent search engines (e.g., the 3DB browser by Jaime Prilusky [pers. comm.]) and the powerful SWISS-MODEL procedure (Peitsch & Jongeneel, 1993) made it possible for an undergraduate student, Kendra Worick, to predict on the basis of sequence similarities the structure of an homologous matrix metalloproteinase (Laczkowsky et al., 1996), which compared favorably with known structures in the PDB, even though stabilizing metal ions (Zn, Ca) were not included in the prediction algorithm.

In addition to founding the PDB, I have been able to use it to make some original contributions:

1. Foreseeing the impact structural data and interactive graphics would make on the field of structure-based drug design, I visited Prof. E.J. Ariens in Nijmegen, the Netherlands; he invited me to write a review chapter (Meyer, 1980) to summarize the current efforts and point the way for subsequent developments.¹⁷
2. A curious pattern of highly resolved internal water molecules in our first high-resolution (1.65 Å) elastase structure (3EST) lead to a survey of structures. The original authors apparently neglected to trace out H-bonding chains radiating from buried active sites; this striking structural motif (Meyer, 1988, 1992) evolved from a curiosity into a fundamental mechanistic principle for proton conduction. Ready access to PDB files made it possible to look for confirming examples as well as counter-examples. A chain of H-bonds, often mediated by internal water molecules, is now an established structural/functional motif in enzymes (cf. Baciou & Michel, 1995; Kandori et al., 1995). Likewise, the necessity of low-energy dissipation of bound water molecules for ligand binding and product release was resolved from a series of high-resolution structures by observing a tunnel or "back-door" through which water could pass in/out of the core of the enzyme. These two mechanistic roles could be enunciated thanks to access to the apices of the structural simplex (Fig. 3), each of which played a crucial role.
3. Most recently, a survey of structures led to a review of surprising binding modes, showing the limitations of rigid models for structure-based drug design (Meyer et al., 1995). What motifs and surprises still await us?

¹⁶The day-to-day details of the PDB were later cared for by a postdoctoral associate in Walter's lab, Tom Koetzle, who continued after Walter's death to direct the PDB for some 20 years. Both distance and my status as a visiting scientist at Brookhaven prevented me from assuming a more active role in the continuing development of the PDB, which was assuming a life of its own. Ironically, the disadvantage of distance was a crucial factor in this first demonstration of crystallographic networking—who would have noticed if I had demonstrated program SEARCH at Brookhaven? But 3,000 km away, from Texas! Also, 1971 was about the time the ARPA network was getting started, showing the further relevance of these efforts.

¹⁷From my perspective, the rate-limiting step for introduction of structure-based drug design was the advent of color graphics, not in the laboratory, but in the board room. Tradition-minded organic chemists were difficult to convince that one could develop novel and useful compounds from playing around with computer models.

Taking some liberty, a heuristic linear relationship appears to persist that equates the number of solved macromolecular structures with the number of graphics workstations in use around the world. This equation may be expanded, on the one side, by prolific groups and the eagerness of NMR studies to present not 1, but 99 structures of a given macromolecule and, on the other side, by the explosion of Macintosh and PC displays with enhanced graphics capability. However, from my perspective, most displays still lack an essential component, stereo filters for emphatic 3D viewing—the difference is comparable to viewing a postcard of Niagara Falls versus a video versus virtual reality versus Niagara Falls itself. One is easily underwhelmed by high-resolution information that is mostly discarded in 2D images as colorful low-resolution worm-like representations (similar to calendar graphics or elevator music).

Even though the PDB is growing exponentially, I predict that the balance will shift in the direction of the number of 3D graphics workstations. Imagine surfing over to channel 512 where a Sussman surrogate entertains questions in real-time about biological macromolecules with instantaneous display of answers, like: "Show me membrane-bound structures with reduced forms of disulfide linkages, and conformational shifts that occur relative to the oxidized forms." "What is the sequence of events that occurs upon binding of an allosteric effector?" "Which structures have two aromatic groups within 4 Å of each other, and how do they pack?" By extending database technology, other questions can be asked, such as: "give me the structure of the RNA polymerase with the tightest binding of rifampin" or "give me examples of drug molecules which bind backward (Meyer et al., 1995) and how do their structures differ from 'forwards-binding drugs'" or "which structural features of hemoglobin make some mutants superior for athletes at high altitude" or "what mutations¹⁸ of mitochondrial proteins are unique to Olympic athletes?" Of course, the display will be a flat screen with a 3D view. It will cost the same as today's TV set, it will contain your phone, FAX, family albums, it will be linked to the libraries of the world, with simultaneous translations for cross-document searches (e.g., the metaphoric uses of "black" among Arabic and Venetian poets).

What will be left to the imagination? Well, the video screen and mouse interface are artificial; we need direct cranial hookups. With increased demand, we'll need more responsive and reliable networks to avoid cyber-boredom. Expert systems will be developed that successfully duplicate the neural networks of the experts; for all the efforts needed to develop this system, it will be antiquated in a generation—and a creator will publish a 25-year retrospective of how virtual neural networks began, between midnight and dawn, when the world slumbered at the close the electronic age. But such efforts to gaze into the future usually miss the mark, when viewed

historically, yet we try, with the limited means at our disposal, which means that much indeed will be left to the imagination. This challenge is given further perspective by comparing current progress and prospects with those of 50 years ago, as enunciated by a much more enlightened seer, Vannevar Bush (Bush, 1945); the World Wide Web has the possibility to meet his challenge,¹⁹ except for 3D (and higher dimensional) data. It is questionable whether software (e.g., virtual reality modeling language or VRML) can solve an essential hardware problem.

Conclusion

A visitor to the Accademia in Florence can see magnificent images that emerged from blocks of marble at the hands of Michelangelo. By analogy, the noncrystallographer can capture the vision that a crystallographer has when admiring a rigorously shaped crystal before exploring the marvelous structure hidden within. So the PDB is our museum, with models of molecules reflecting the wonders of nature and complex shapes that may be as old as life itself. With the aid of interactive graphics and networking, the PDB makes these images readily available. What wonders still remain hidden as we build, compare, and extend our database?

Beyond the initial use of color raster graphics for molecular modeling, what of this recollection is unique? We know of individuals who have made greater contributions to the life sciences using one or another method discussed here. Perhaps it is the ability to forge a link between crystallography, a database, graphics, and a network to span computational and experimental sciences and thereby make the results conveniently available that will be viewed retrospectively as an original contribution?

With broad brushstrokes and subdued pigments, a vision emerges from a union of computational and experimental science that appeals directly to our most powerful input device, our eyes. Although many of the events related here appear arcane, one can only speculate how limited our current accomplishments may appear to the next generation. But we know where the life sciences would be without them.

Early efforts were supported by Brookhaven National Laboratory and the US National Science Foundation. The Robert A. Welch Foundation (A-328) has provided sustaining support for 27 years.²⁰ Dr. Robert Spinrad (Xerox Corp.) provided supplemental information about the BRAD system. My wife, Catarina, has been a sustaining inspiration for 31 years and I have benefited greatly from 20+ years of faithful collaboration and innovative algorithm development by Dr. Stan Swanson, as well as a number of capable students and postdoctoral associates. Drs. Erik Meyer and Rosemarie Swanson and Mrs. Francis Bernstein provided useful editorial suggestions. Walter's close friend, Sidney Abrahams, and the obituary provided by BNL, were used to compose the addendum on Walter Hamilton.

The seed was planted in Stan Simonsen's laboratory in Austin, where I grew my first crystals, measured my first data, and wrote my first programs. It took root in Jack Dunitz's laboratory in Zürich and prospered at MIT. It blossomed when Walter invited a

¹⁸The universally revered cockroach will be the common host for testing structure-based mutations, which will be synthesized by robots and injected and tested in safe facilities. The Protein Society will sponsor the Cockroach Olympics: marathon running, underwater swimming, track and field, and aerial events; *Science* will choose the Cockroach of the Year for its athletic and scientific prowess. High-school science fairs will never be the same. TV sit coms will let viewers diagnose and cure the physical or psychic diseases created by frantic producers; holograms of structures of mutant proteins will be collected and swapped by children's fan clubs. The sequel to "Dungeons and Dragons" will be "Molecules and Mutants"; instead of memorizing metabolic pathways, medical students will learn to search the web; biochemists will be designing computational links between databases to fill the gaps, and computer scientists will be designing new algorithms to create optimal pathways in *n*-dimensional space to create totally new drugs.

¹⁹"The difficulty seems to be, not so much that we publish unduly in view of the extent and variety of present-day interests, but rather that publication has been extended far beyond our present ability to make real use of the record."

²⁰Research in my laboratory has progressed even without 20+ years of NIH support, thanks mostly to very capable and dedicated colleagues.

fellow Texan and his family to spend a few summers on Long Island. It continues to bear fruit.

Addendum

Dr. Walter C. Hamilton, Senior Chemist and Deputy Chairman of the Chemistry Department at Brookhaven National Laboratory, became one of the world's leading crystallographers in his thirties as a result of his development of the neutron diffraction facilities at Brookhaven, his outstanding neutron investigations of magnetic and ferroelectric systems, and his neutron diffraction studies of H-bonding in water and amino acids, as well as by his books on crystallographic statistics, hydrogen bonding, and symmetry.

Born in Austin, Texas, February 16, 1931, Walter received a B.S. degree from Oklahoma State University (then, Oklahoma A&M); the following year he attended the Eidgenössische Technische Hochschule in Zurich, Switzerland as a Swiss Government Fellow. He received his Ph.D. from Cal Tech with Verner Schomaker. He was an NSF Postdoctoral Fellow at the Mathematical Institute of Oxford University in England with Prof. C.A. Coulson. He then spent his professional career at Brookhaven, building an impressive crystallographic group and taking advantage of the automation that was then developing actively to use 4-circle diffractometers for the measurement of precise X-ray and neutron data. His neutron diffraction facility was one of the best in the world in the 1960s. He was thus able to take full advantage of other uses of computers, which were then becoming available (in particular, Digital's PDP-8 at the low end and the CDC 6600 at the high end) to the scientific community. His brilliant work led to his appointment as Deputy Chair of the Chemistry Department at Brookhaven in 1968. He authored more than 100 scientific papers dealing with crystal and molecular structure determinations. He was author or co-author of three books: *Statistics in Physical Science* (1964), *Hydrogen Bonding in Solids* (1968), and *Symmetry* (1972). In 1966, he was appointed volume editor and member of the International Tables Commission of the IUCr.

Walter served from 1964, until his death from cancer at age 41, on the USA National Committee for Crystallography; he had been a delegate to several IUCr General Assemblies, was serving actively on several National Research Council committees, and had organized the symposium "Computational Needs and Resources in Crystallography," published posthumously as an influential report by the National Academy of Sciences. He developed an extensive set of crystallographic programs for small-molecule structure analysis, which became a central part of our CRYSTNET project.

References

- Baciou L, Michel H. 1995. Interruption of the water chain in the reaction-center from rhodobacter-sphaeroides reduces the rates of the proton uptake and of the 2nd electron-transfer to Q(B). *Biochem* 34:7967-7972.
- Barry CD, North ACT. 1972. The use of a computer-controlled display system in the study of molecular conformations. *Cold Spring Harbor Symp Quant Biol* 36:577-584.
- Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M. 1977. The Protein Data Bank: A computer-based archival file for macromolecular structures. *J Mol Biol* 112:535-542.
- Bertele E, Boos H, Dunitz JD, Elsinger F, Eschenmoser A, Felner I, Gribo HP, Gschwend H, Meyer EF, Pesaro M, Scheffold R. 1964. Ein Synthetischer Zugang zum Corrinssystem. *Angewandte Chemie* 76:393.
- Bush V. 1945. As we may think. *The Atlantic Monthly* 176:101-108; cf. also <http://www.isg.sfu.ca/~duchier/misc/vbush/vbush.txt>.
- Cotton FA, Bier CJ, Day VW, Hazen EE Jr, Larsen S. 1972. Some aspects of the structure of staphylococcal nuclease. *Cold Spring Harbor Symp Quant Biol* 36:243-255.
- Hamilton WC. 1970. The revolution in crystallography. *Science* 169:133-141.
- Johnson CK. 1965. ORTEP. *Oak Ridge National Laboratory Report*. ORNL-3794.
- Jones TA. 1978. A graphics model building and refinement system for macromolecules. *J Appl Crystallogr* 11:268-272.
- Kandori H, Yamazaki Y, Sasan J, Needleman R, Lanyi JK, Maeda A. 1995. Water-mediated proton transfer in proteins: An FTIR study of bacteriorhodopsin. *J Am Chem Soc* 117:2118-2119.
- Laczowsky A, Meyer EF, Swanson SM, Worick RK. 1996. Molecular modeling and rational drug design—A perspective. In: Mahmoudian M, ed. *Proceedings, 12th Iranian congress of Physiology and Pharmacology*. Forthcoming.
- Langridge R. 1988. The early days of molecular graphics. In: Stezowski JJ, Huang JL, Shao MC, eds. *Molecular structure: Chemical reactivity and biological activity*. Oxford: Oxford University Press. pp 583-588.
- Langridge R, Macewan AW. 1965. *IBM scientific symposium on computer aided experimentation*. Yorktown Heights, New York: IBM. p 305.
- Levinthal C. 1965. *IBM scientific symposium on computer aided experimentation*. Yorktown Heights, New York: IBM. p 315.
- Levinthal C. 1966. Molecular model-building by computer. *Scientific American* 214:42-52.
- Meyer EF. 1968. *Annual Report, Brookhaven National Laboratory*. p 81.
- Meyer EF. 1970a. Three-dimensional graphical models of molecules and a time-slicing computer. *J Appl Crystallogr* 3:392-395.
- Meyer EF. 1970b. Towards an automatic, three-dimensional display of structural data. *J Chem Doc* 10:85-86.
- Meyer EF. 1971. Interactive computer display for the three-dimensional study of macromolecular structures. *Nature* 232:255-257.
- Meyer EF. 1974. Storage and retrieval of macromolecular structural data. *Biopolymers* 13:419-422.
- Meyer EF. 1980. Interactive graphics in medicinal chemistry. In: Ariens EJ, ed. *Drug Design, vol IX*. New York: Academic Press. pp 267-289.
- Meyer EF. 1988. A structure: function study of receptor+substrate interactions derived from high-resolution X-ray crystallography. In: Stezowski J, Huang JL, Shao MC, eds. *Molecular structure: Chemical reactivity and biological activity*. Oxford: Oxford University Press. pp 179-188.
- Meyer EF. 1992. Internal water molecules and H-bonding in biological macromolecules: A review of structural features with functional implications. *Protein Sci* 1:1543-1562.
- Meyer EF, Botos I, Scapozza L, Zhang D. 1995. Backwards binding and other structural surprises. *Perspectives in Drug Discovery and Design* 3:168-195.
- Meyer EF, Morimoto CN, Villareal J, Berman HM, Carrell HL, Stodola RF, Koetzle TF, Andrews LC, Bernstein FC, Bernstein HL. 1974. CRYSTNET, a crystallographic computing network with interactive graphics display. *Fed Proc* 33:2399-2402.
- Morgan HL. 1965. The generation of a unique machine description for chemical structures—A technique developed at chemical abstracts service. *J Chem Doc* 5:107-117.
- Ophir D, Shepherd BJ, Spinrad RJ. 1969. Three-dimensional computer display. *Comm ACM* 12:309-310.
- Peitsch MC, Jongeneel CV. 1993. A 3D model for the CD40 ligand predicts that it is a compact trimer similar to the tumor necrosis factors. *Int Immun* 5:233-238.
- Richards FM. 1968. The matching of physical models to three-dimensional electron-density maps: A simple optical device. *J Mol Biol* 37:225-230.
- Richards FM, Wyckoff HW, Carlson WD, Allewell NM, Lee B, Mitsui Y. 1972. Protein structure, ribonuclease-S and nucleotide interactions. *Cold Spring Harbor Symp Quant Biol* 36:35-43.
- Willoughby TV, Morimoto CN, Sparks RA, Meyer EF. 1974. Mini-computer control of a stereo graphics display. *Appl Crystallogr* 7:430-434.
- Wright CS, Alden RA, Kraut J. 1969. Structure of subtilisin BPN' at 2.5 Å resolution. *Nature* 221:235-242.
- Yankeelov JA Jr, Coggins JR. 1972. Construction of space-filling models of proteins using dihedral angles. *Cold Spring Harbor Symp Quant Biol* 36:585-587.