RETROSPECTIVE

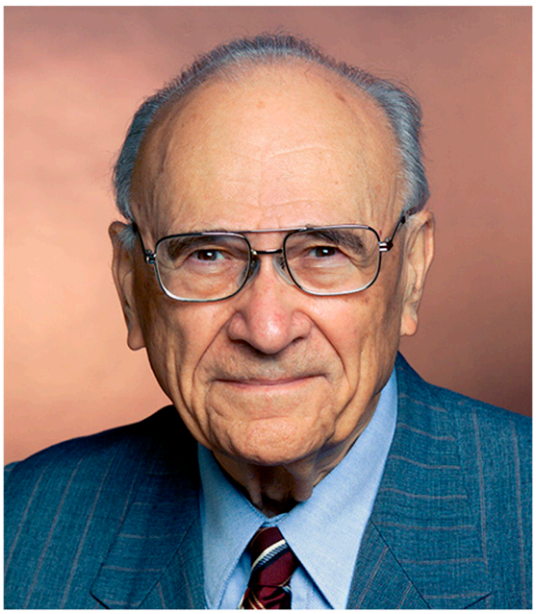
Harold A. Scheraga (10/18/1921–8/1/2020): A pioneering scientist who laid the foundations of protein science in the 20th century

Hagai Meirovitch[a] and Ivet Bahar[a,1]

Harold Abraham Scheraga, an eminent Professor of chemistry and biophysics at Cornell University for 73 years, died on August 1, 2020 at the age of 98. Scheraga (known to his colleagues as Harold) has been a pioneer in the general field of macromolecules (polymers, proteins, DNA, and so forth). He started his own research in 1947 (when proteins were viewed just as ellipsoids of colloidal assemblies of amino acids), contributing vigorously to the progress in this field until his death. Unlike most scientists, his studies encompassed both experimental and theoretical–computational disciplines. He targeted this wide research area together with more than 400 researchers he trained as students, postdoctorates, and research associates, leading to an enormous output of around 1,400 papers. This exceptional activity is also reflected by the more than 50 honors Harold received from universities and companies around the world. He became an elected fellow of the American Association for the Advancement of Science (1966), an elected member of the US National Academy of Sciences (1966), and the American Advancement of Art and Sciences (1967).

Harold was a highly popular lecturer around the globe, and served on national and international committees, scientific advisory boards, and editorial boards of a large number of scientific journals. While his 2014 curriculum vitae includes a list of 85 such activities, we would like to mention those reflecting Harold's great support of the State of Israel. In 1963, he spent a year at the Weizmann Institute with family, as a Guggenheim Fellow and a Fulbright Research Scholar; in 1972 to 1978 he held a Visiting Professor position in the Biophysics Department at the same institute and served on the Board of Governors of the Weizmann Institute from 1970 to 1997, becoming an Emeritus member in 1997.

Harold's unusual activity was driven by his strong passion and enthusiasm for science, supported by a sharp mind, phenomenal memory, encyclopedic knowledge, and high efficiency. These traits were manifested in the Monday group seminars by his mastery in the details, and ability to interconnect different phenomena. Along with his ambitious scientific life, Harold was a friendly person—a mensch—who maintained close ties with many of his former students and colleagues, generously providing advice and support when needed. Harold liked to socialize; he and his beloved wife, Miriam (who died 7 months before him), were very hospitable hosts in their home. He also loved to travel, but despite his frequent trips he was able to maintain an unusual productivity. The trick was using the flying time to work intensively on his manuscripts, probably with the same efficiency as in his office.



Harold A. Scheraga. Image credit: Cornell University, licensed under CC BY-NC-ND.

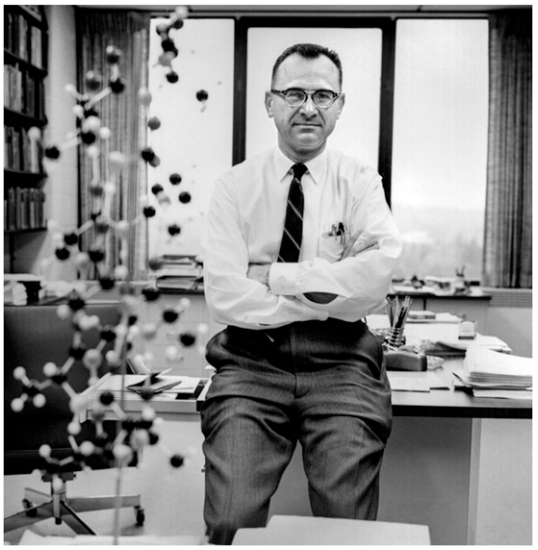
[a]Department of Computational and Systems Biology, School of Medicine, University of Pittsburgh, Pittsburgh, PA 15213

Author contributions: H.M. and I.B. wrote the paper.

The authors declare no competing interest.

[1]To whom correspondence may be addressed. Email: bahar@pitt.edu.

Published February 5, 2021.



Harold Scheraga as a young professor at Cornell University. Image credit: Cornell University, licensed under CC BY-NC-ND.

Harold was born on October 18, 1921 in Brooklyn, New York, where he attended high school and the City College of New York (CCNY). Exposure to physics at CCNY attracted him to the field of physical chemistry. He was graduated from CCNY in 1941 with a Bachelor's degree and moved to Duke University for doctoral studies. Harold's thesis, on the Kerr effect in small molecules, was completed under the supervision of Paul Gross. During his years at Duke, he also devoted time to World War II projects. After graduation, in 1946, he spent a year as a postdoctoral fellow at Harvard Medical School with John Edsall, where he first began working on proteins.

In 1947 Harold applied for a position in the Chemistry Department at Cornell University. He was interviewed by the Department Chairman, Peter Debye, and accepted an Instructor position the same day. He advanced rapidly, becoming an Assistant Professor in 1950, Associate Professor in 1953, and a Full Professor in 1958. In 1965 he was appointed to an endowed chair, the Todd Professorship of Chemistry. From 1960 to 1967 Harold served as Chair of the Department, presiding on the construction of an additional building—the S. T. Olin Chemistry Research Laboratory—and led the expansion of the department's research areas into molecular biology and materials science. Harold built a strong department through both senior and junior hires, which became one of the top 10 chemistry departments in the nation. His management skills were exceptional. In 1992, he became the George W. and Grace L. Todd Professor Emeritus, a position he held until his passing; during this 38-year period between 1992 and 2020 he did not slow down his scientific work; in fact, some of his papers are still in preparation, under review, or accepted for publication.

While the research output of the Scheraga group has been enormous, equally important was the depth. Harold sought to understand the fundamental interactions that dictate 1) the folding of a polypeptide chain in water into the three-dimensional structure of a native protein and 2) the reactivity of a protein (e.g., as an enzyme) with other small and large molecules. The earlier experimental work involved genetic engineering and hydrodynamic (e.g., sedimentation and viscosity), spectroscopic (Raman, infrared, fluorescence, NMR, electron spin resonance, UV absorption, circular dichroism, and optical rotatory dispersion), immunochemical, and other physicochemical measurements. These methods were applied to proteins, synthetic polymers of amino acids, and model compounds, in studies of protein-folding pathways. These techniques were also used to study the mechanism of action of thrombin on fibrinogen (an important reaction in blood clotting).

The early theoretical studies (from 1962 to 1970) include a series of very influential papers on the structure of water and hydrophobic bonding in proteins. Initially, his models were developed for the thermodynamic properties of liquid water, hydrocarbons, and liquid deuterium oxide in water. Investigations were also conducted on the influence of water structure and hydrophobic interactions on the strength of hydrogen bonds in the side chains of proteins, and on the contribution of hydrophobic bonds to the thermal stability of proteins. These models were further developed by him extensively. In parallel, Harold developed experimental methods that enable the measurement of distances between specific sites. These methods were applied systematically over an extended period of time to the protein ribonuclease A, before the resolution of its structure by X-ray crystallography.

Another substantial project during the 1960s of the 20th century dealt with the helix-coil transition in biopolymers, summarized in the 800-page book, *Theory of the Helix-Coil Transition in Biophysics*, by Poland and Scheraga (1). The use of potential energy functions for searching the protein's conformational space had just emerged at that time. Typically, the potential energy surface of a protein is decorated by a tremendous number of local energy minima, where the global energy minimum structure is identified with the native (folded) structure. However, finding the global energy minimum has been an unsolved problem in global optimization; Harold referred to this challenge as "the multiple-minima problem in protein folding." To address this problem, he developed the potential energy program (force field), ECEPP (empirical conformational energy program for peptides), which became the main engine of research in his group and the basis of hundreds of papers. ECEPP was enhanced several times over the course of the years and the effect of water was added implicitly.

While other force fields are also available, ECEPP relies on a much smaller number of variables (dihedral angles), with bond lengths and bond angles kept fixed; this makes it a very convenient tool for protein-folding studies. Indeed, the Scheraga group developed numerous stochastic methodologies for global

optimization, based on the energy criterion, which proved to be significantly more efficient than methods based on a random search. We mention here only the Monte Carlo minimization method of Li and Scheraga that has found wide application and has been borrowed by other fields of science. Due to its efficiency, ECEPP has been widely used by various groups as a tool for developing simulation techniques (such as replica exchange and the multicanonical method).

It will be impossible to review here all of Harold's contributions. Still, it is important to emphasize the highly original research of his group that may have been largely overlooked. The global optimization methods mentioned above are based on the energy, while the correct criterion of stability is the free energy; that is, the entropic effect should be included as well. The Scheraga group (mainly with N. and M. Gō) was the first to address this issue by calculating the harmonic entropy as applied to a large set of peptides. These researchers also developed the first procedure for handling a cyclic peptide, by designing a technique for changing its conformation systematically. Other important theoretical papers from this team elaborate on the question of how to treat a protein structure described by internal coordinates within the framework of statistical mechanics.

Protein science has become a highly active well-established field with extensive information on the structure and function of proteins based at the atomic level. A breakthrough in this field shook the structural biophysics world 2 weeks ago: AlphaFold, a deep-learning–based computing system has been declared to have solved the 60-year-old "protein-folding" challenge. At the heart of this success is the training of the AlphaFold artificial intelligence system using 170,000 structures deposited in the Protein Data Bank, in addition to sequence data. Data-driven studies are now interwoven in our tools for discovery in a diversity of disciplines. We have come a long way since the 1975 data-driven study of Tanaka and Scheraga applied to the protein-folding problem (2, 3). This was the first time the concept of learning from data on protein structures (approximately two dozen resolved then) was introduced to the field to extract the so-called "knowledge-based parameters" for assisting in protein-folding simulations. While it was then premature to expect limited data and methods to solve the problem, this is one of several studies that Harold conceptualized ahead of his time and may have been overlooked among the sheer volume of work he contributed to the scientific community. A closer look at some of the earlier work of Harold Scheraga, where protein science was in its infancy, portrays him as a pioneer with monumental contributions, who helped lay the foundations of this field.

1 D. Poland, H. A. Scheraga, *Theory of the Helix-Coil Transition in Biophysics* (Academic Press, 1970).

2 S. Tanaka, H. A. Scheraga, Medium- and long-range interaction parameters between amino acids for predicting three-dimensional structures of proteins. *Macromolecules* **9**, 945–950 (1976).

3 S. Tanaka, H. A. Scheraga, Model of protein folding: Inclusion of short-, medium-, and long-range interactions. *Proc. Natl. Acad. Sci. U.S.A.* **72**, 3802–3806 (1975).

https://doi.org/10.1073/pnas.2026796118