AW.02: Advances in Macromolecular Phasing; Impact on Molecular Biology

The symposium opened with the award presentation to **Bi-Cheng Wang** by ACA President Marvin Hackert (*see p. 8*). In his keynote address *Resolution of Phase Ambiguity in Macromolecular Crystallography: 25 Years Later*. B.C. explained the problems of macromolecular phasing in the 1970-80s. The standard procedure for phasing macromolecular structures employed Perutz's multiple isomorphous replacement method (MIR) which required at least two heavy atom derivatives and the native data, used in combination, to resolve the phase ambiguity. At that time, BC was faced with the problem of having only one useful derivative out of over 100 attempts. Thus began an analysis of the single isomorphous replacement (SIR) phase ambiguity problem that eventually led to a successful protein structure analysis. A prime factor in resolving the single-derivative phase ambiguity was the invention of a simple algorithm applied to electron density maps, which distinguished protein electron density from solvent regions, thus defining the molecular envelope. This procedure made *solvent flattening* a routine and easily used process that *could locate and enhance the protein image, i. e. whatever is not solvent must be protein*. Using the enhanced protein image, improved phases are calculated which could then be processed through additional cycles of refinement in an iterative procedure. A similar phase ambiguity exists for single wavelength anomalous scattering phasing procedures (SAS) and can be treated in a similar manner. The entire SIR and SAS procedure was coded into an easy-to-use computer program and generally distributed to all who requested a copy during the mid-1980s.

From left: Wim Hol, Quan Hao, Wayne Hendrickson, John Rose, Emil Pai, Bi-Cheng Wang, Zbigniew Dauter and David Langs. ***Photo courtesy of Gary Newton.***

With regard to developments since 1990, B.C. talked about an idea he calls *Direct Crystallography* which makes use of anomalous scatterers that are naturally occuring in proteins, such as metals like iron or zinc or perhaps small substrates with built-in heavy atoms. One particularly interesting element is sulfur, present in nearly all proteins, with a small but measurable anomalous scattering component. If the anomalous differences from sulfur scattering can be measured accurately enough, then many native proteins can be solved without the need of a derivative, and from one data set. This idea galvanized efforts to measure data more accurately. The use of longer wavelength radiation at home lab sources and at synchrotrons for which the sulfur anomalous signal is substantially increased was one result. A new method under development at SER-CAT and UGA: *Signal Based Data Collection* utilizes real-time data monitoring and feedback of the anomalous signal strength during data collection, combining automation, robotics and multiple crystals. The program will sense when the current crystal has decayed to the point that the anomalous signal is unacceptable; cause a fresh crystal to be mounted automatically; and resume the data collection.

Wayne Hendrickson, HHMI Investigator, Columbia, and recipient of the first Patterson Award in 1981, spoke on the *Evolution of Phase Evaluation from MAD and SAD Measurements*. He recalled the use of anomalous dispersion measurements from sulfur only in the solution of the protein crambin, the first successful sulfur anomalous dispersion (SAD) structure. Later, the structure of lamprey hemoglobin was successfully solved with the use of the multiple wavelength anomalous dispersion (MAD) from iron, (1988). Soon after, the incorporation of selenium, with it's large anomalous signal, resulted in a highly successful MAD procedure. By 2000, the majority of *de novo* structures were solved using MAD procedures. In 2007, however, more structures were solved by SAD than MAD, probably because of improved data measurements and the need for only one data set. Wayne concluded by mentioning two new approaches: the incorporation of generically methionine-enriched Fab phasing vectors, and the use of very large metal clusters with "colossal" anomalous signals.

Zbigniew Dauter, APS, Argonne, talked about the *Wang Limit*. This term became widely used in the late 1980s and was applied to data sets in which only sulfur atoms provided the maximum anomalous signal. An error-free data set was calculated based on the anomalous scattering from two sulfur atoms (Bence-Jones protein Rhe, 114 amino acid residues). This simulation, by B. C. Wang, provided an example that suggested a macromolecular structure could be successfully phased even when the average anomalous scattering signal was as low as 0.6% of the average structure factor values. From B.C.'s 1985 paper: *"…we see that the best strategy for a macromolecular structure determination in the future is not to aim for more isomorphous derivatives but to concentrate on one 'good' derivative and to improve the quality of its SIR and SAS data, because, as long as one can accurately measure the data, a macromolecular structure can be determined from SIR or SAS data even when the occupancy of the heavy atom or the size of the anomalous scatterer is relatively small*. Zbigniew pointed out that a majority of novel structures today are solved by a Se-SAS approach, concluding that there is no such thing as the *Wang Limit;* everything depends on the diffraction data accuracy.

cont'd, next page

Emil Pai, U.Toronto, talked about direct sulfur SAS phasing, emphasizing the use of longer wavelength radiation, especially CrKα (2.29Å), where the anomalous signal from sulfur is twice that for CuKα. He also presented some practical tips on selecting crystals suitable for sulfur phasing (no split spots, no ice rings, no high mosaicity) and on data collection (high redundancy, increased signal/noise ratio). He noted that solving the Patterson function in lower space groups might help by displaying symmetry relationships. In his feasibility study of structures that failed to solve by a Se-Met approach, he has to date solved solved 10 of 23 structures using the native sulfur anomalous signal alone. His message: *the higher the I/σ(I), the better the chances for a structure solution – there is no substitute for signal!*

David Langs, Hauptman-Woodward Inst., recalled numerous Pittsburgh Diffraction Conferences and one particular Direct Methods Workshop in Buffalo in 1984. B.C. presented his new solvent flattening procedure, supported by calculations on the VAX 11/780 computer in Buffalo. His work certainly inspired the development of the real space component for phase refinement used in HWI's Shake and Bake minimal function program. B.C. also indirectly influenced the Buffalo group to refocus their efforts from "large" small molecules to the determination of macromolecular heavy atom substructures. David went on to describe a new SnB-based phasing method that can identify large subgroups of phases having a lower mean phase error than the rest, an idea that came to him while preparing his talk for this symposium. For the first time individual phases can be directly determined outside the trial-and-error construct of refining phases to fit the probabilistic constraints of the three phase invariant relationships.

Quan Hao, Cornell, followed with a talk jointly prepared by Quan and **Hai-fu Fan** and colleagues, Inst. Physics, Beijing, China. A central feature of the phasing process is the OASIS program, developed in Fan's group. This program implements a direct method for solving the phase ambiguity, originally proposed in 1965. More recently, OASIS has been improved by the inclusion of a dual-space iterative procedure including solvent flattening, model building and structure refinement and is included in the CCP4 program package. Quan listed numerous examples of successful phasings on a variety of proteins with various combinations of OASIS, SOLVE/RESOLVE, ARP/wARP and DM. One particularly impressive example was a protein TTHA1634 with 1206 residues, a resolution of 2.1Å, 22 sulfur atoms per ASU and a Bijvoet ratio of 0.55% that was solved by OASIS and DM. Another impressive example was a model completion procedure using OASIS starting from a partial (20%) model produced by molecular replacement.

Wim Hol, U. Washington, presented an intriguing talk: *Molecular Machines, Tropical Diseases and the Power of Llamas*. As is well known in the macromolecular crystallographic community, Wim has spent many years studying the structural biochemistry of global diseases. Wim focused on two components of the Type II Secretion System (T2SS), a sophisticated "machine" that secretes nasty toxins like cholera toxin and homologs. The first example was the structure determination of protein peri-D from enterotoxigenic *E. coli*. Initial attempts to solve the structure produced crystals which diffracted to no more than 6-7Å. This difficulty led to the use of llama antibodies in search of an answer to the problem. Unlike normal antibodies, camelid (llamas and camels, etc.) antibodies have a single-domain 15 kDa antigen binding fragment resembling a 2.4 nm x 4 nm prolate ellipsoid, here termed a *nanobody*. Briefly, the process involved, in close collaboration with the group of Jan Steyaert in Brussels: injection of llamas with the protein of interest; waiting for the immune system to work; extracting lymphocytes from the blood and nanobody DNA from the lymphocytes; expressing nanobodies in *E. coli*; and binding the antibody to the protein (in this example, a SeMet derivative). This effort was successful in producing a structure of the nanobody-peri-D complex. Another example, the EpsI-EpsJ heterodimer complex from *Vibrio vulnificus*, required species variation, truncations and SER mutations to produce crystals which diffracted to 2.1Å. The llama antibody route again proved successful in this case: crystals of the IJ-nanobody complex appeared within days. The "IJ" pseudopilin heterodimer has a remarkable shape: a "hand" waiting for partner(s). Wim's final slide showed the many structures of T2SS proteins which have been solved in his group by SIR or SAS procedures.

John Rose, UGA, reviewed early *de novo* structure determinations using Wang's ISAS technique. Bovine neurophysin (isolated from the posterior pituitary gland) complexed with an iodinated Phe-Tyr dipeptide was successfully solved with iodine-SAS in 1987, and the next year the structure of ferrochelatase, which converts protoporphyrin to heme, was solved by iron-SAS.The iodine and iron atoms furnished a strong anomalous signal using a CuKα home source. A quest for a synchrotron beamline with stable x-radiation generation at longer wavelengths -so that the sulfur anomalous signal would be enhanced.- was successful in 1999 when the structure of protein obelin was solved using sulfur SAD with data measured at 1.74Å at IMCA-CAT at APS. In conclusion John showed 26 structures in the current PDB solved by sulfur SAD.

Gary Newton and John Rose

AV and registration desk crew (they also took most of the photos of speakers): In back, from left: Max Trent, Jonathan Page, Nicholas Sanjines, Will Zhou; in front: Brenda Dougan, Julia Abbott, Michelle Minton.