

Fig. 1. Herman Wold, the pioneer of PLS, giving his lecture at the Frankfurt meeting (photo P. Geladi).

- "On the rationale of PLS modeling" by Fred Bookstein.
- "Soft modeling on multivariate images" by Paul Geladi, Kim Esbensen and Svante Wold.
- "Evaluation of school systems using partial least squares: an application in the analysis of open systems" by Richard Noonan.
- "Predicting classroom behaviour ratings by prenatal alcohol exposure: latent variable modeling and nonlinear scaling" by Paul Sampson, Ann Streissguth, Silvia Vega-Gonzalez, Helen Barr and Fred Bookstein.
- "Application of partial least

squares methodology to the behavioral teratology of alcohol" by Paul Sampson, Ann Streissguth, Helen Barr and Fred Bookstein.

An interesting aspect of the meeting was that different disciplines of science were drawn together to discuss their vision of PLS and its uses and to measure these visions against each other. The result was that many participants left the meeting with useful tips on how to use and modify PLS for problem definitions other than their usual ones. This will probably have a great influence on future developments.

All the lectures had written

hand-outs. It was decided that no proceedings volume of this symposium will be published. The authors of the papers were encouraged to publish their papers individually. More information can be obtained from the individual contributors via the organizing committee.

PAUL GELADI

*Research Group for Chemometrics,  
Umeå University, Umeå, Sweden*

## Computer Simulations in Protein Engineering and Drug Design, Amsterdam, The Netherlands, 20–23 April 1988

Computer simulations are becoming more and more important in crucial areas such as protein engineering

and drug design. This was the major conclusion of the symposium organised by mini-supercomputer producer

Alliant. Organiser Paul Weiner (Alliant's 'Computational Chemistry' person) had succeeded in attracting a number of outstanding speakers for the meeting.

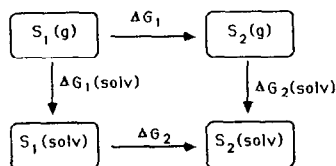
There were two major themes at the conference: structure elucidation of biomolecules and free-energy calculations. Both are aimed at ob-

taining more knowledge about the relation between spatial structure and the activity of such systems. Such knowledge is of the utmost importance for designing novel drugs or new or altered proteins ('site-directed mutagenesis' is the current buzzword here - replacing one amino acid in a protein or enzyme with the objective of drastically altering its activity, stability or specificity).

Information regarding the structure of proteins can be obtained due to the major progress made in the past two decades in the areas of both X-ray crystallography and NMR spectroscopy. Interestingly, state-of-the-art research combines experimental data and computer simulations (in particular, molecular dynamics) in the structural elucidation process.

### Free-energy calculations

Free-energy calculations may become of major importance in the study of the effects of, for example, point mutations in proteins and enzymes. A simple example (taken from the lecture by Peter Kollman, UCSF) serves to illustrate the issues. It is concerned with calculating the solvation free-energy difference of two compounds  $S_1$  and  $S_2$ :



The vertical arrows signify processes that are hard (or impossible) to simulate, but whose  $\Delta G$ -values can be determined experimentally. The horizontal arrows indicate non-physical processes that are relatively easy to simulate (for instance, 'transforming' a methyl group into a hydroxyl group). The following equation holds:

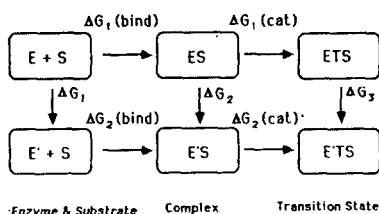
$$\Delta G_1 + \Delta G_2(\text{soln}) = \Delta G_2 + \Delta G_1(\text{soln})$$

From which it follows that:

$$\begin{aligned} \Delta \Delta G(\text{soln}) &= \Delta G_2(\text{soln}) - \Delta G_1(\text{soln}) \\ &= \Delta G_2 - \Delta G_1 \end{aligned}$$

Hence, the difference in  $\Delta G(\text{soln})$  can be computed indirectly and compared to the experimental value.

A similar methodology can be applied to the computation of the effect of an amino acid substitution:



Kollman used an example in which  $E$  is subtilisin (a serine protease) and  $E'$  is the same enzyme with Asn 155 substituted by alanine. The  $\Delta \Delta G$ -values for substrate binding and catalysis of the formation of the transition state are now given by:

$$\begin{aligned} \Delta \Delta G(\text{bind}) &= \Delta G_2(\text{bind}) \\ &\quad - \Delta G_1(\text{bind}) \\ &= \Delta G_2 - \Delta G_1 \end{aligned}$$

$$\begin{aligned} \Delta \Delta G(\text{cat}) &= \Delta G_2(\text{cat}) - \Delta G_1(\text{cat}) \\ &= \Delta G_3 - \Delta G_2 \end{aligned}$$

David Beveridge (Wesleyan University) presented an introductory paper concerning the 'Free Energy Methods' phenomenon and Monte Pettitt (University of Houston) discussed some case studies of successful and less successful applications of the technique. Peter Kollman (UCSF) treated a number of examples from the work of his group (from which came the abovementioned work on

subtilisin). According to Kollman, progress has been very rapid in the past few years in the area of simulations of large systems in solution. This is due to the development of better methods for searching through conformational space (e.g. a combination of distance geometry and molecular dynamics), the combination of quantum mechanics and molecular mechanics force fields and the development of the free energy perturbation method (which enables free energy differences to be computed). Kollman stated that the results of  $\Delta \Delta G(\text{soln})$  computations are now comparable to experimental results. As a result, such computations might ultimately render a lot of experimental work superfluous. Evidently, as the reliability of the results of such computations increases, the necessity to actually mutate enzymes, for example, and carry out experiments with them (a non-trivial undertaking) decreases similarly.

### Structural determination methods

Naturally, for a better understanding of the relation between structure and (re-)activity/function/stability, a knowledge of the spatial structure of biomacromolecules is crucial. In a breathtaking presentation, Wim Hol (University of Groningen) sketched a number of recent developments that have extended the boundaries of protein X-ray crystallography. On the one hand, these are developments in the area of instrumentation (such as so-called area detectors) with, on the other hand, developments that accelerate the actual structure elucidation process (from electron densities to atomic coordinates). A very recent example of the latter category is the use of molecular dynamics in the structural refinement process.

The ultimate goal of all this work is to develop a fully automatic struc-

tural refinement procedure, i.e. one in which the human crystallographer does not have to intervene (using molecular graphics). At present it is already possible to attain  $R$  values below 25% without any human intervention (an  $R$  value is a measure which indicates how well (or poorly) a structure explains the experimental data).

In recent years, NMR spectroscopy has become a competitor to X-ray crystallography in the determination of the structures of biomolecules. A major advantage of this technique is that one obtains information regarding the structure of the system in solution (whereas crystallography yields information about the solid state). The majority of the proton chemical shifts can be assigned from a number of two-dimensional NMR spectra. Subsequently, distance information for pairs of protons may be obtained from several so-called NOESY spectra. Using a mathematical method called distance geometry, this distance information can be 'translated' into a number of structures which are (largely) consistent with the distance information. These structures can then be refined by means of molecular dynamics and energy minimisation. This results in a three-dimensional structure (or, rather, a 'family' of such structures).

A disadvantage of this method is that it is limited to relatively small molecules. A possible avenue to the alleviation of this problem may be offered by the use of three-dimensional NMR spectroscopy. Rob Kaptein (University of Utrecht) spoke about the first complete 3D NMR experiment (a so-called NOE-HOHAHA experiment) which was carried out at the end of 1987 on pike parvalbumine (a protein consisting of 108 amino acids). He showed a number of slides of the three-dimensional spectrum (the fourth dimension, intensity, is colour coded) which

were quite impressive. The overwhelming amount of data generated in such an experiment (in this example  $256 \times 256 \times 256 = 16,777,216$  data points, containing tens of thousands of peaks!) makes it imperative that the analysis of such spectra be automated.

Crystallography and NMR spectroscopy share the common property that they both are time consuming techniques. Hence the old dream of being able to predict protein structures without performing any experiments (the so-called protein folding problem). There were two presentations regarding this particular subject.

Suzan Holtzman (Molecular Architects Corp.) explained that she had not solved the protein folding problem, but it did not become clear what it was that she had actually done. There were many buzzwords during her talk (pattern recognition, fuzzy set theory), but how her structure prediction method works was not revealed to the audience.

Tom Blundell (Birkbeck College) gave an interesting overview of the methodology and software (COMPOSER) developed in his laboratory. With this package, protein structures can be predicted using analogies to structures stored in a database. A large number of rules is also used, which was the reason for labelling this approach 'knowledge-based modelling and design'.

### Criticisms

Apart from all exciting success stories, there were also a number of sobering presentations. 1987 Nobel Prize winner Jean-Marie Lehn (Université Louis Pasteur and Collège de France), while reviewing a fascinating area of chemistry (see also: J.M. Lehn, *Angewandte Chemie, International Edition in English*, 27 (1988) 89-112), emphasised the importance

of traditional models on a number of occasions: "I want to be able to hold a model in my hands".

Wilfred van Gunsteren (University of Groningen) discussed a number of problems related to computer simulations. In the first place, there is the problem of searching through conformational space. This can be done in a number of ways: systematically (impractical on present day computers), probabilistically (so-called Monte Carlo methods) or deterministically (molecular dynamics). Distance geometry may also be used, but this does not yield a proper Boltzmann-sampling of the conformational space. In the second place, one may use experimental information in order to reduce the conformational space. Both direct (proton-proton distances or structure factor amplitudes) and indirect information can be used. The use of structures from a database is an example of the latter, but Van Gunsteren calls this 'a poor man's device'. In the third place, there are problems related to the force fields. A molecule's energy is considered to consist of a number of terms, for instance

$$E_{\text{total}} = E_{\text{bonds}} + E_{\text{angles}} + E_{\text{dihedrals}} \\ + E_{\text{non-bonded}} + E_{\text{H-bonds}}$$

The set of these terms and the parameters governing them is called the force field. As will be obvious, a large number of choices has to be made:

- what is the atomic model (are protons considered separately, or are united atoms, such as '-CH<sub>3</sub>' employed?);
- which terms occur in the expression for the energy?
- how are the parameters to be chosen (fitting to the results of ab-initio calculations or to experimental data?).

In addition, these choices are never absolute, but depend on the application of the force field. At the end of his 'practitioner's critique', Van Gunsteren concluded (much to the relief of those present) that computer simulations are indeed useful. They promote a better understanding of the relation between microscopic and macroscopic phenomena, can be useful in the interpretation of experimental results, may yield semi-

quantitative estimates of some quantities, and may be used for extrapolation to extreme conditions. With respect to the computational chemist's instrumentation, Van Gunsteren foresees a bright future for parallel computers.

**Acknowledgements**

I would like to thank Alliant Computer for enabling me to attend

the conference, and R. Kaptein and J.A.C. Rullmann for useful suggestions regarding the text of this report.

GERARD J. KLEYWEGT  
*Department of NMR Spectroscopy,  
 University of Utrecht,  
 Padualaan 8, 3584 CH Utrecht,  
 The Netherlands*

**CAC 88. 4th International Conference on Chemometrics in Analytical Chemistry, Amsterdam, The Netherlands, 18-20 May 1988**

A wide range of topics was covered at CAC 88, which was located in a picturesque theatre in the heart of the city. Specific applications as well as new theoretical approaches were presented by over 130 participants, most of whom were based either in Western Europe or in the U.S.A.

Principal component analysis was a subject very much at the center of interest. Several speakers gave examples of the pros and cons of the technique or compared its results with those of others, e.g. multiple linear regression. Chemometrics as a discipline was brought into the picture as usual, as some argued that this is still not generally accepted as a distinct profession in its own right. Talks

about applications to food chemistry and testing of pharmaceuticals clarified the need to improve the image of chemometrics with regulatory authorities.

In the field of expert systems several reported vast progression and excellent results. Indeed, one might be inclined to believe that this technique will eventually make chemometricians superfluous. That prospect should not, however, worry current workers on this topic, for the applications were still largely restricted to quite small tasks.

Outlier detection may not be the most profound of techniques in chemometrics, at least it is one that can be understood by those without extended knowledge of mathematics.

A number of applications was presented, and in an interesting presentation by Rousseeuw a new, more robust technique was introduced, based on taking the least median of squares.

On the subject of artificial intelligence some general lectures were given which seemed hardly to belong in the specific field of chemometrics. A highly interesting presentation was given by Holland on the subject of genetic algorithms in artificial intelligence. His results seemed to indicate that almost all techniques can basically be derived from random actions, as long as good evaluation algorithms exist. Hardly anybody would dare to predict the implications this may have on the subject of artificial intelligence.

E.M.M. ROSENDAAL  
*University of Amsterdam,  
 Amsterdam, The Netherlands,*