

Appendix*

Jean-Pierre Bretauière (1946-2008) and the early days of multivariate statistics in electron microscopy.

Marin van Heel,

Imperial College London



Fig A-1: Jean-Pierre Bretauière in front of a Paris poster – photo: Jan Galligan, Albany 1979

Jean-Pierre Bretauière had a very fundamental influence in the early days of multivariate statistical analysis of the images of single particles in electron microscopy, an influence that was not very well documented in terms of co-authorships or in other written documents. Let me explain how he was involved in the “intelligent averaging” of specific views of single particle images. Averaging of a great many “unit cells” of molecules arranged as a helical assembly [De Rosier & Moore 1970] or as a two-dimensional crystal [Unwin & Henderson 1975], had at the time made quite an impression in the field and many electron microscopy groups worldwide thus started image processing projects. The first ideas of applying the concept of averaging to the images of individual macromolecules had been formulated [Saxton & Frank 1977] and implemented in Owen Saxton’s early IMPROC software system. It was not yet clear at the time how best to find the 3D structure of a macromolecular complex based on individual two-dimensional projection images of the molecules. Would single-particle tomography such as proposed by the group of Hoppe [Hoppe *et al.* 1974] be the way towards, or did one first have to find all the different views presented by a macromolecular complex? Around 1980, we were still a decade away from answering such questions.

***Appendix to:** Van Heel M, Portugal R & Schatz M: **Multivariate Statistical Analysis in Single Particle (Cryo) Electron Microscopy.** In: "An electronic text book: Electron microscopy in Life Science", 3D-EM Network of Excellence, Editors: A. Verkley and E. Orlova (2009).

Since 1976, I had been employed as a project scientist in the electron microscopy group of Professor Ernst van Bruggen of the Biochemistry Department of the University of Groningen, The Netherlands. Erni van Bruggen had been most impressed by the early “averaging” successes of the group around Aaron Klug in Cambridge and wanted to incorporate those techniques in his research. It was my task to develop image processing software for data collected either with the transmission electron microscope (TEM) or the Scanning version of that instrument (a “STEM”). In the fall of 1979, I went to visit Albany, for a 6-week period, to work with Joachim Frank on automatically finding sub-populations of images in a heterogeneous population as discussed in the main paper. We had met two years earlier at an EMBO image processing course in Basel and we had been corresponding ever since. We both agreed that when averaging images of individual macromolecules, the presence of a mixture of different views was a serious and important complication: a challenging problem in need of a solution. I travelled to Albany taking with me two different sets of images in which different views could already be visually distinguished. One set consisted of images of worm hemoglobin (two different preparations: one with a clearly visible extra subunit in the middle and one without) and the other dataset of arthropod 4x6 hemocyanins where the difference between “flip” and “flop” could be directly seen (**Fig A-2**).

What followed were many nights in the windowless catacombs of the New York State Department of Health (Division Laboratories and Research, now the Wadsworth Centre) beneath the Empire State Plaza. During these late sessions I met Jean-Pierre Bretauière (“JP”) while working at the same PDP-11 45 computer (**Fig A-3/A-4/A-5**). We first talked about other issues than science, such as his personal ties to Paris (**Fig A-1**) where I had spent almost all of 1971 as a student. We also discussed our actual work and I explained the first attempts Joachim and I were making at separating subsets of images from the overall population of molecular images. Jean-Pierre explained what he was doing at the State Department of Health: quality control of medical laboratories in the state of New York. The Health Department would send out identical blood samples to many different medical laboratories. The results of the various chemical tests from the different labs were then compared using a French multivariate statistical technique called “l’analyse des correspondances” [Bretauière, 1981]. At the time I had never heard of multivariate statistical analysis (“MSA”), let alone of Benzécri’s “correspondence analysis” [Benzécri, 1973-1980].

Jean-Pierre was immediately convinced that his “AFC” program (for: “Analyse Factorielle des Correspondances”) would be able to help solve our problem and handle our “massive” amount of image data. The AFC program was a modified version of the correspondence analysis program “CORAN” by Jean Paul Benzécri [Benzécri, 1973/1980]. What was “massive” in this context was a very relative and very dated concept: at the time we had less than 64 images of 64x64 pixels = 4096 pixels each; that is a total data set size of less than one megabyte (each image being 16 Kbyte in floating point format). An uncompressed picture from even the cheapest digital camera today is much larger than one megabyte! However, it was characteristic for those early days in computing, that the 16-bit PDP-11 computer would only address 2^{16} bytes = 64 Kbytes. It was thus difficult to fit any “major” computational problem in the central memory of such a computer! The critical memory requirement in our case was the “core” needed for the square variance-covariance matrix (see main article) which, for 64 images would already occupy 16 Kbyte of the precious available address space.

As we sat together with Joachim the following morning to discuss the matter, we were all convinced that this was very much worth trying. Joachim wrote a small conversion program to allow JP's AFC program to read the aligned images produced by the SPIDER program. The result of this first correspondence analysis of single particles was an instant success!



Fig A-2: Wadsworth Computer Centre, 1979. Visible on the image monitor (controlled by the 24x80 characters VDU) is a gallery of 4x6 arthropod hemocyanin images used for the first correspondence analysis of EM images. Already in this gallery one manages to directly see the difference between the “flip” and “flop” versions of the 4x6 half molecules of the *Limulus polyphemus* hemocyanin – photo: Marin van Heel, Albany 1979.

The first trial was on a giant annelid hemoglobin data set extracted from images of hemoglobins from two different species. The giant annelid hemoglobin of *Oenone fulgida* visibly had extra density in the centre [Van Bruggen & Weber, 1974] compared to the well known giant hemoglobin of the common earthworm *Lumbricus terrestris* which rather had a “hole” in the centre of its double-ring geometry. The separation of the two species, artificially mixed for this analysis, was so obvious that we immediately moved to a single data set with real internal variations. We thus moved to the second data set I had brought from Groningen, negative-stain images of *Limulus polyphemus* 4x6-meric hemocyanins (a dissociation product of the full 8x6-meric native hemocyanin).

On the correspondence analysis maps of these images four clear groups (“classes”) were visible. These reflected the “flip and flop” versions of the 4x6 oligomer where the right hand 2x6 part can be shifted up or down with respect to the left-hand side 2x6-mer. This behaviour had already been anticipated based on visual inspection of the images (**Fig A-2**). However, that effect would have led to a subdivision into two groups rather than four. The other effect, which I had not anticipated, split both the “flip” and the “flop” groups in two, namely that of “rocking”. The four hexamers constituting the 4x6-mer were apparently not in one plane

which led to two stable “rocking” positions where 3 out of 4 hexamers would be in direct contact with the carbon support film. This explanation, however, was not so straightforward because the alignment algorithm used at that time would itself introduce a 180° ambiguity in the in-plane rotational alignment. Both the shape of the molecule and the flaws of the alignment scheme thus led to the same effect. This issue was eventually resolved after confusing argumentations with referees [van Heel & Frank 1980; van Heel & Frank 1981]. The books that JP recommended [Benzécri 1973/1980]; and especially [Lebart *et al.* 1977] became my “bibles” for years to come (English versions of these books have since appeared making that literature more accessible).

In these early MSA results, we would print out two-dimensional maps of the positions of the images on the first factorial co-ordinates and then draw a circle around a cluster of images, call that a “class” and then sum the members of that class for further interpretation. This visual classification, after eigenvector-eigenvalue data reduction, obviously had its shortcomings and was impossible to perform when more than ~3 factorial axes were to be taken into account. It was JP who also pointed me in the direction of automatic classification and the wealth of literature on the subject [van Heel, Bretaudière & Frank, 1982], which then led to the introduction of automatic Hierarchical Ascendant Classification, “HCA”, in electron microscopy [van Heel, 1984].



Fig A-3: Wadsworth Computer Centre, 1979. The PDP-11/45 “minicomputer” – photo: Marin van Heel, Albany 1979

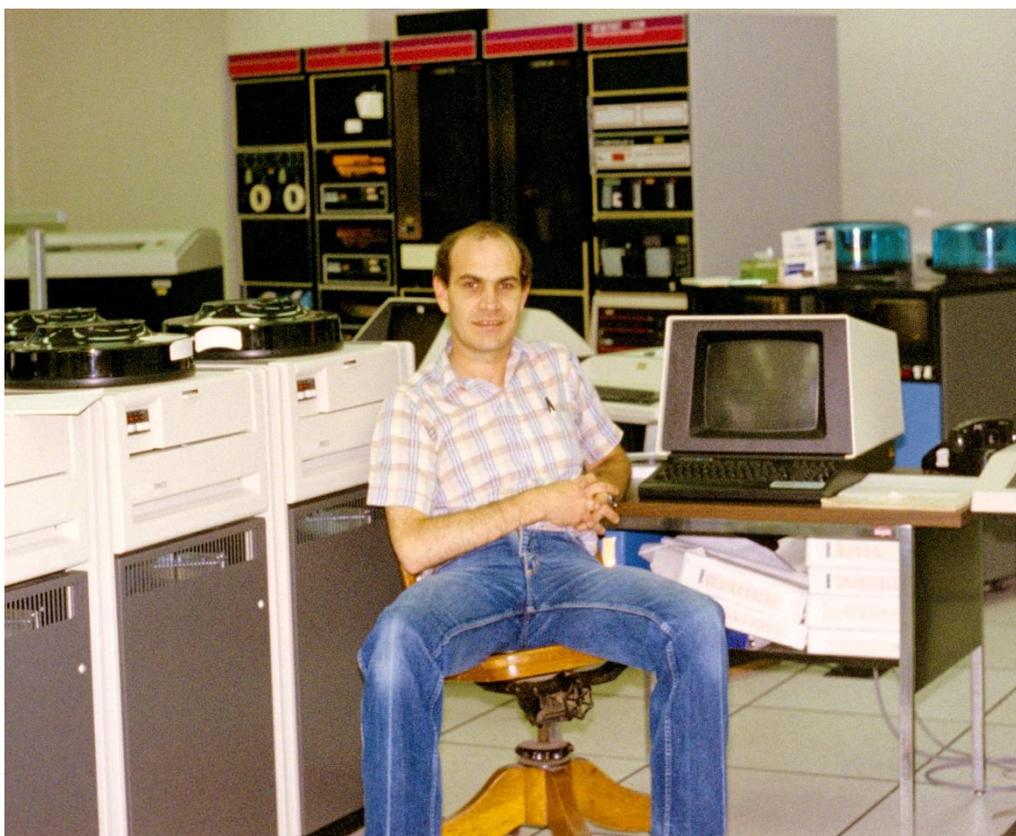


Fig A-4: Jean-Pierre Bretauière in the Wadsworth Computer Centre – photo: Robert Rej, Albany 1981.

Jean-Pierre was a wild guy - bursting with energy, full of ideas and projects, always burning the candle at both ends. I have vivid memories of, after a late evening session at work and an even longer night in a bar, stumbling out of the bar roaring with laughter along with JP and Vicky, Joachim's programmer who later became JP's wife. JP was always exuberant in his love for life and all its adventures. Jean-Pierre, having seen the success of correspondence analysis in single-particle electron microscopy, became directly interested in EM image processing. He changed fields and moved to the University of Texas Medical Center in Houston, Texas, where he developed the SUPRIM image processing system for electron microscopy [Schroeter & Bretauière, 1996]. In Houston, being JP, he was proud to declare himself more Texan than the Texans: he bought a ten-gallon hat, a huge Lincoln with "BRETO" vanity plates, and a large gun to defend his home and family. I still vividly remember the self-satisfied smile on his face when he told me this...

The last time we spoke, already many years ago, was about the time he left Houston to return to France. We spoke of his health issues and about his plans to leave the single-particle EM field - again. JP was never afraid of rigorous decisions. To this day, I am sorry we did not include JP as a co-author of the first papers that resulted [van Heel & Frank 1980; van Heel & Frank 1981]. He was offered co-authorship but refused. Maybe Joachim and I simply did not push him hard enough. His co-authorship would have reflected his essential early contributions to the ideas in this field. The only printed evidence of our early collaboration which remains is the abstract for the European electron microscopy meeting in Hamburg in 1982 mentioned above [Van Heel, Bretauière & Frank, 1982].

The beautiful portrait of JP by Jan Galligan (**Fig A-1**), which so clearly documents his *joie de vivre*, resurfaced in my archives a few years ago. After scanning it, I searched the web to find JP's current email address hoping to put a smile on his face. Unfortunately I failed to find a current address. Only recently, after receiving the sad message of Jean-Pierre's death, did I learn he had moved to Madagascar, after selling his very successful French computer business – Brett Computers – moving away from western society's rat-race. Some say that scientists are dry and boring people... There is no doubt that with JP no longer around, the world of science has become a slightly more boring small universe.

I am grateful to Jan Galligan and Robert Rej for letting me to use their pictures, and for their support and constructive suggestions.

References

Benzécri J.-P. **L'Analyse des Données Vol 2** (1973-1980) Dunod Paris.

Bretaudière JP, Dumont G, Rej R, Bailly : **Suitability of control materials. General principles and methods of investigation.** *Clin. Chem*, **27/6** (1981) 798-805,

DeRosier DJ, Moore PB: **Reconstruction of three-dimensional images from electron micrographs of structures with helical symmetry.** *J. Mol. Biol.* **52:** (1970) 355-369.

Hoppe W, Gaßmann J, Hunsmann N, Schramm HJ, Sturm M: **Three-dimensional reconstruction of individual negatively stained yeast fatty acid synthetase molecules from tilt-series in the electron microscope.** *Hoppe-Seyler's Z. Physiol. Chem.* 1974, **355:** 1483-1487.

Lebart L, Morineau A, et Tabard N, **Techniques de la Description Statistique** (1977)

Saxton WO, Frank J: **Motif detection in quantum noise-limited electron micrographs by cross-correlation.** *Ultramicroscopy* **2** (1977) 219-227.

Schroeter JP & Bretaudière JP: **SUPRIM: Easily modified image processing software:** *J. Struct. Biol.* **116** (1996) 131-137.

Unwin PNT & Henderson R: **Molecular structure determination by electron microscopy of unstained crystalline specimens.** *J. Mol. Biol.* **94** (1975) 425-440.

Van Bruggen EFJ, Weber RE: **Erythrocrucorin with anomalous quaternary structure from the polychaete *Oenone fulgida*.** *Biochim. Biophys. Acta* **359** (1974) 210-214.

Van Heel M, Frank J: **Classification of particles in noisy electron micrographs using correspondence analysis**, in: *Pattern Recognition in Practice*, Ed. E.S. Gelsema and L.N. Kanal, North Holland (1980) 235-243.

Van Heel M, Frank J: **Use of multivariate statistics in analyzing the images of biological macromolecules,** *Ultramicroscopy* **6** (1981) 187-194.

Van Heel M, Bretaudière JP, and Frank J: **Classification and Multi-Reference Alignment of Images of Macromolecules**, in: *Proc. 10th Intern. Congress on Electron Microscopy, Hamburg, 1982, Vol. 1*, pp. 563-564.

Van Heel M: **Multivariate Statistical Classification of Noisy Images (Randomly Oriented Biological Macromolecules),** *Ultramicroscopy* **13** (1984) 165-183.