Material discovery and formulation optimization using high throughput methods

Dominique Charmot

Résumé **Découverte de matériaux et formulations par criblage haut-débit**

Les méthodes de criblage haut-débit ont largement dépassé le domaine pharmaceutique et sont maintenant utilisées dans la chimie des grands intermédiaires et des spécialités. Symyx Technology Inc., pionnier en la matière, a développé cette science à un niveau tel qu'il est dorénavant possible de synthétiser, analyser et mesurer les performances de milliers de matériaux en un temps record. Des plates-formes de criblage permettent d'explorer les espaces de formulation dans des applications très variées, telles que les mélanges de polymères, pigments, émulsions, nanodispersions et formulation d'actifs pharmaceutiques. Contrairement à la thérapeutique humaine, la recherche haut-débit appliquée aux matériaux et formulations requiert tout un ensemble de méthodes de criblage pour décrire l'enveloppe de propriétés que constitue la valeur d'usage. Les procédures de criblage haut-débit sont devenues sophistiquées et égalent, sinon dépassent, la qualité des tests traditionnels. Ces ensembles intégrés de préparation et de criblage de formulations génèrent des bases de données que nul n'aurait pu suspecter il y a seulement quelques années, et constituent aujourd'hui un outil incontournable de la recherche et du développement.

Mots-clés **Formulation, criblage haut-débit, chimie combinatoire, polymères, pigment, émulsions, propriétés mécaniques.**

Key-words **Formulation, high throughput, combinatorial chemistry, polymers, pigment, emulsions, mechanical properties.**

In one original connotation, combinatorial chemistry referred to methods where pharmaceutical compounds were pooled together and screened for a specific activity at which point the signal was deconvoluted to identify the compound(s) of interest. When applied to polymer formulation discovery, the term *combinatorial* is improperly used since there is no deconvolution process taking place in hit identification. However, in contrast with the drug discovery approach, a property's space in formulation is always multidimensional and continuous in nature. One looks at combinations of materials, such as polymers, fillers, pigments, additives, processing aids, etc., and formulation process variables to achieve the best compromise in terms of performance and cost. Although extended in this context, the combinatorial term describes a set of high throughput synthesis, formulation, and screening techniques to search for optimal multicomponent formulations in a short period of time. Formulation is one of the most important aspects of material development. To quote the Association of Formulation Chemists (http://www.afc-us.org): « *With rare exceptions, e.g., as intermediates for the synthesis of other chemicals, molecules have no utility until they are formulated into products that have desirable properties. Once formulated, however, the end-products become extremely valuable: they become our foods, medicines, paints, coatings, cosmetics, packaging materials, adhesives, inks, crop protection agents, and many other functional items that are essential to modern life* ». With the high development cost of new chemical substances and the growing difficulties to register them in chemical inventories (with the notable exception of polymers), companies prefer to revert to formulation as a flexible and quick way to respond to the market demand or new or improved products. From the supplier viewpoint, a comprehensive understanding of the application properties of a new formulation additive is critical, not only to react immediately to the customer needs, but also to capture the widest coverage of intellectual property. From the user point of view, formulation skills and knowledge is a key factor of product differentiation in a fast moving market. Despite recent progress in formulation science and informatics, the day to day formulation work remains tedious and relatively cost inefficient. Molecular modeling today is still limited to model systems and is inadequate to predict properties of fully formulated systems with the required accuracy. On the other hand, statistical analysis and design of experiments (DOE) are enjoying a fast growing success among users. To develop this approach even further, there is a need to generate in quasi real time a large body of data that formulators can use to solve the problem at hand. In that respect, combinatorial tools are becoming ubiquitous. In the following, we will describe high throughput synthesis and characterization tools geared towards polymer materials, and discovery platforms related to formulation issues.

High throughput synthesis and characterization of polymers

Polymers are an essential component in formulations in general and are purposely taken as a backdrop in this paper for the sake of illustration. To some extent, polymer synthesis can be thought as a formulation problem where the building blocks are commercially available monomers and the process variables dictate the polymer macro and microstructure and its properties.

General infrastructure in high-speed synthesis and screening

For high throughput polymer synthesis, several 96 reactor blocks are used in parallel and operated in a batch mode. The library design is generated by Library Studio® [3] software that offers various options to create a plurality of polymer compositions across the library plate and has built-in functions to check for recipe consistency. Library Studio® manipulates any kind of variable, not necessarily reagents, and can quickly propose several strategies to map out not only a given chemical composition space, but also a formulation space. The library layout is then transferred to a second piece of software, Impressionist® [4] that uses this information to automatically dispense the individual components to each library element using solid and liquid handling robotics. The reaction is then carried out and the libraries transferred to the screening platforms. A typical throughput for this synthesis system is about 5,000 polymers per week per operator. For secondary screening, where more information or material is required or when more sophisticated processes are sought, a Semi Continuous Parallel Polymerization Reactor, SCPPR™, is used. This has about the same capability as a state-of-the-art fully automated bench reactor with the difference being that from 24 to 96 reactions are run in parallel [5]. Each reactor has mechanical agitation and temperature control, and is connected to up to 12 individual feed lines of fluid reagents. Both temperature and reagent feed profiles are individually programmed and executed under computer control. Gaseous reactants, *e.g.,* ethylene, can be used at a pressure of up to 1 500 psi. When real time polymerization kinetic data are needed, a solenoid valve system monitors the pressure of each vessel within a specified range by adjusting the flux of gas and thus measuring the uptake of ethylene. Lately, parallel Raman on line monitoring has been added as a feature to track down reaction kinetics. While the frame body of the SCPPR™ is stainless steel, the polymerization is actually carried out in glass vials, and both vials and stirrer paddles are disposed of to avoid the burden of cleaning and any risk of material contamination. These parallel reactors were designed and built to run polymer discovery programs but have been used at Symyx to prepare blends, pastes, and emulsions by a careful re-engineering of the agitation design.

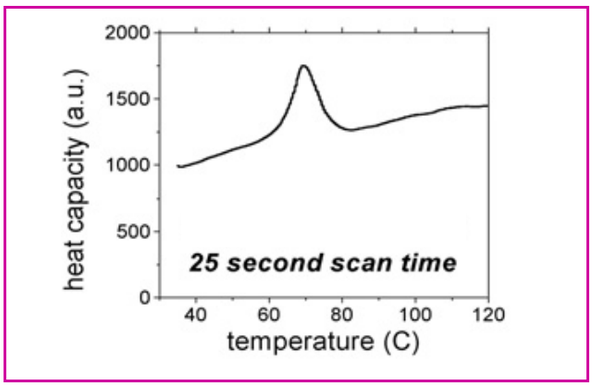

Figure 1 - Thermogram of a melting point of a polyethylene sample (Mw = 4000 g/mol) obtained on one of the 128 channels of the SAMMS™ instrument.

Screening of generic polymer properties

Molecular weight, glass temperature and chemical composition are measured by a set of tools: Rapid GPC™ [6], Rapid FT-IR™ and SAMMS™ [7] respectively. All of them operate in a fast serial mode and give analytical data at typical speeds of 0.5 to 4 minutes per sample. Rapid GPC™ provides molecular weight averages in a few seconds for homopolymers or up to a few minutes if the full molecular weight distribution is required or more complex samples are to be analyzed. The Rapid FT-IR™ system uses a commercial spectrometer hooked up to a robotic system that transfers polymer samples from the library plate to a specially prepared silicon wafer substrate on which the reflectance spectra are taken. SAMMS™, an acronym for Sensor Array for Modular Measurements System, is an array of electronically addressable sensors. It permits a number of different measurements depending on its configuration: for thermal analysis, a layout of microcalorimeters is micromachined on a silicon wafer or microfabricated on the polymer support confined in a high vacuum. The small size of the samples and the quasi absence of thermal loss to the environment allows one to scan in temperature at a much faster rate than that which is possible in a commercial DSC, as shown in *figure 1*. Interestingly, the high speed of analysis reduces neither the accuracy nor the resolution: the Tg's of a reference library of known polymer materials measured with SAMMS™ match well to the data acquired with a commercial DSC instrument. The sample preparation is critical and requires considerable attention to control the amount, thickness, and homogeneity of the deposited material, or the absence of residual solvent when solvent casting is used. Once those conditions are determined, the robotic environment is reliable enough to guarantee reproducibility and consistency within and between libraries. Moreover, these tools have such a high throughput that some of the wells on each plate can be sacrificed to run reference samples for internal calibration.

Software environment

Software plays a major role in tying together and synchronizing the various pieces of hardware, as much as, in managing the flow of information generated along the different phases of the workflow. Symyx has developed a whole suite of informatic tools incorporated in the Renaissance™ package: from library design to execution control of screening protocols, processing data in readable formats, storing, handling data with a central database, visualization and data mining. Still the amount of data can be quickly overwhelming. A typical program that produces about 500 materials/formulations a week can easily generate several thousands of data points and then the data mining can become challenging. What was formely rewarding for the scientist, *i.e.,* extracting information from a limited dataset and providing guidance for the following set of experiments, can be a daunting task when not properly handled. Several commercial software packages now provide ways of sorting out data and performing statistical analysis and are didactic enough to be good decision aids. Along the same vein, DOE and SAR techniques are being used more and more to help sample the experimental space and build knowledge models in a relatively short period of time. It is not uncommon to tackle a formulation problem with 5 different functional ingredients, each being selected from an archive of 50 elements. Add to this 5 concentration levels

and 3 formulation process variables (*e.g.,* temperature, shear mixing, order of addition) and one ends up with as many as 37,500 formulated materials, and more likely 50,000 if one includes duplicates for reproducibility. Assuming that we screen for three formulation attributes then we are talking about 150,000 data points! Clearly the library design has to be handled differently whether the purpose of the study is to build up a knowledge database, regardless of any performance criteria (*i.e.,* formulation tool), or whether a specific target is contemplated. In the latter case the use of DOE and data modeling maybe useful.

Selection of formulation platforms

Preformulation of pharmaceutical actives

In drug discovery, small compound archives are screened for a particular biological target. Once lead compounds have been identified, a comprehensive and costly program is then initiated to bring the drug candidate to the market place. This includes toxicology studies, clinical trials, and manufacturing and formulation processes. The latter encompasses the identification of crystal polymorphs and the solubilization profile of the active as well as the mode of delivery (oral, IV, etc.). The early identification of all the crystal polymorphs is important since usually the most stable form will be selected for commercial development and FDA registration. Interestingly, while the vast majority of drug companies have heavily invested in combinatorial chemistry and highspeed biological assays to fuel the discovery effort, relatively few, if any, have brought drug formulation into a high throughput regime. Symyx has identified this as a bottleneck in the time to market as well as in the lifecycle management of the drug; we thus created a complete infrastructure to accelerate and systematize drug pre-formulation research and development. In a nutshell, the platforms consist of a set of tools geared towards the selection of salts and crystallization conditions, coupled with several characterization techniques of crystalline solids. The crystallization platform executes in parallel a set of typical chemical engineering operations such as blending, solubilization, filtration, crystallization, decanting, and centrifugation in miniaturized devices. This micro chemical plant occupies the deck of a robotic system of modest dimensions (1.5 meters long) with a throughput of several hundred reactions/crystallizations per day. A grid of solvents and counter ions, various cooling and evaporation rates are predefined to create thousands of process conditions to form crystals which are then analyzed in a secondary screening by their birefringence, melting points, low angle X-ray scattering, and Raman spectra. Log P values (partition coefficients of the drug between water and 1-octanol) that characterize the hydrophilicity of the active are measured by high speed HPLC. The entire approach is integrated and is carried out in a fraction of the time and using a fraction of the material needed for a conventional polymorph study. *Figure 2* gives a snapshot of the birefringence pattern of a drug under various crystallization conditions.

Organic pigment optimization

Organic pigments are obtained through multistep organic synthesis and, for reasons briefly outlined above, pigment manufacturers are reluctant to develop new chemical substances while materials with novel color properties can be developed through smart formulations of existing

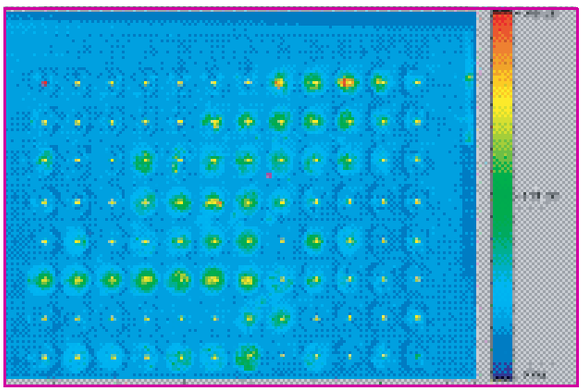

Figure 2 - Color enhanced birefringence image of a crystallisation microtiter plate of a commercial Non Steroid Non Inflamatory Drug.

compounds. For instance, certain pigments mixed in a particular stœchiometric ratio can form co-crystals, whose color characteristics are not the mere additive combination of the spectra of starting pigments but reflect the new crystalline lattice of the co-crystal. In a joint project with Ciba Chemicals, Symyx undertook a discovery program aimed at novel pigment combinations selected from a 20 element repertory of known pigments, and examined their properties in paint formulations (essentially color coordinates and light fastness under weathering conditions). A full workflow has been developed which includes:
- paint formulation,
- paint patch generation,
- reflection spectroscopy,
- color calculations,
- X-rays diffraction acquisition and analysis,
- fade, migration, luminescence.

The pigments are prepared by a wet solution process and milled to submicron size. X-ray diffraction data are acquired by a rapid transmission mode using a 6 kW/mm^2 copper anode able to record 96 spectra in approximately 2 hours. Once formulated with a solvent borne binder in a parallel blender, the paint samples are then drawn down with a specially designed knife to an approximately 1 mil. thick pigmented film arrayed on a polyester substrate (see *figure 3*). An integrating sphere mounted on an X-Y arm robot scans the paint library and collects the reflectance spectra. The software then compares the actual spectra of the pigment composition with the spectra calculated from a simple combination, and automatically signals pigment compositions which depart from the additivity rule. The same process is applied with the high throughput X-ray diffraction system. This high throughput approach has led to the identification of novel pigment compositions in a timeframe about 10 times shorter than in a conventional « one at a time » method. This infrastructure also serves other purposes such as the characterization of new binders and additives in paint formulations.

Nanodispersions and emulsions

Dispersion and emulsification are two basic techniques broadly used in modern formulation. Even though the colloid science has greatly evolved during the last decades, this remains mostly based on know-how and trial and error, and makes it an ideal case study for high throughput

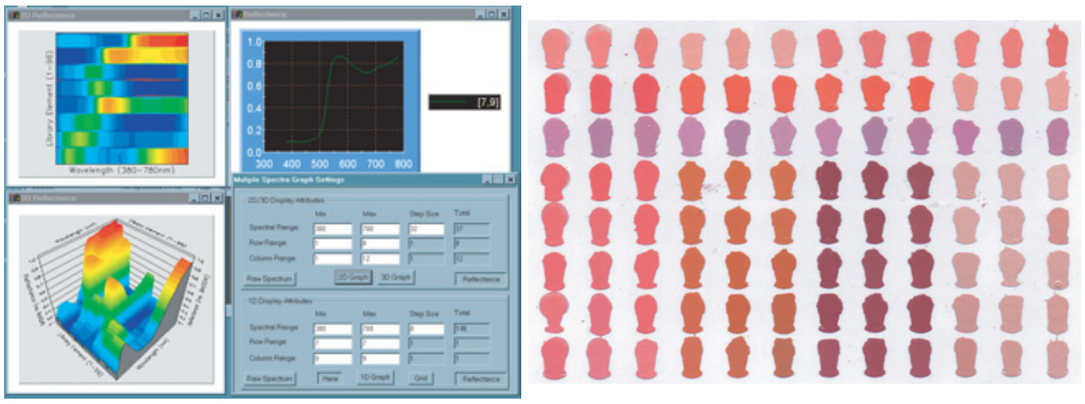

Figure 3 - Right: pigmented paint library. Left: high throughput spectral analysis and pigment color visualization.

experimentation and combinatorial methodologies. A suite of tools has been built to formulate, disperse, and emulsify using various processes, and also characterize the final dispersion or emulsion as is or after being submitted to diverse stresses. Each workflow is different and customized to the particular material and application target in view. Below is a panorama of high throughput formulation processes and characterization screens in place that may tackle a whole set of formulation needs in fields as diverse as paints, cosmetics, lubricants, and agrochemicals to name a few.

Formulation design and execution
- Library studio®, Impressionist™ software,
- High throughput robotics liquid and solid dispensing.

Processing
- Solid solution redispersion,
- Uncontrolled precipitation,
- Emulsification (high shear homogeneizer, ultrasonic probe),
- Controlled precipitation,
- Milling,
- Drying.

Screening
- Solids content,
- Thin layer chromatography (TLC),
- pH,
- Thermal analysis,
- Molecular weight analysis,
- Spectroscopic analysis (FT-IR, Raman),
- Composition analysis,
- Image analysis,
- Viscosity,
- Stability (vs. shear, electrolyte…),
- Particle size analysis.

Particle sizing is an important aspect and several techniques have been developed. For very high throughput and dispersions with particle size in the submicron range, a dynamic light scattering instrument has been built that probes the upper surface of the liquid samples (see *figure 4*). The accuracy and precision of particle size for standards (polystyrene calibrated latex) across a microtiter plate is better than 5% for a 8 sec. time of analysis per sample, including data processing and reporting. When the full particle size distribution is needed, commercial instruments such as laser diffraction and capillary hydrodynamic fractionation particle sizers are automated with Symyx's software and robotic platforms. Two examples illustrate the capability of this infrastructure: water nanodispersions and emulsions.
These water nanodispersions include an aqueous vehicle and nanoparticles made of a small molecule active stabilized

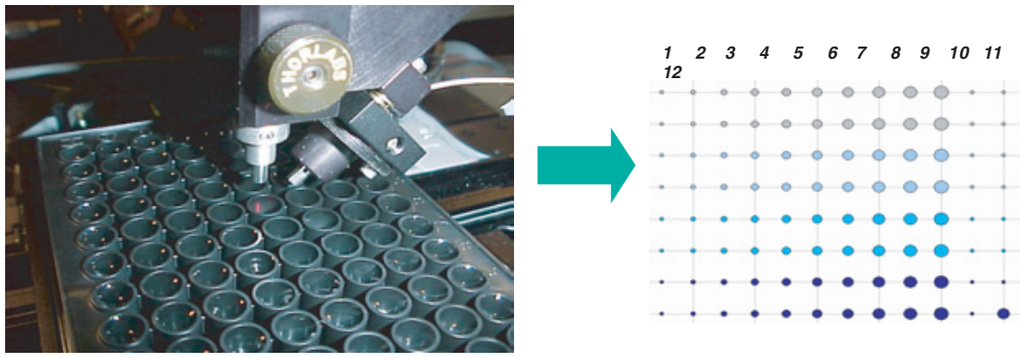

Figure 4 - Left: high throughput dynamic light scattering instrument. Right: particle size mapping of a 96 elements library (the spot size is proportional to the hydrodynamic diameter of the particles).

by a synthetic polymer; the set of specifications includes particle size range, minimum of active percent solids in dispersion, and a good colloidal stability under specified conditions. The process is based on a solid solution route wherein a polymer is intimately mixed with the active and solvent and then reduced to a solid solution by parallel drying, followed by redispersion in water. The main variable is the composition and architecture design of the polymer used as a dispersing aid: the combinatorial space is composed of a set on monomers (up to 100, selected among hydrophobic, hydrophilic, charged, neutral, acidic, basic, etc.) and the workflow in this particular configuration generates and screens 15,000 formulations per week. Given narrowly defined criteria, a limited polymer compositional space accommodates the delicate balance of solubility/interaction between the solvent and the active to provide a good dispersibility. This appears quite clearly in *figure 5* where the hits are materialized in a pseudo quaternary space domain of polymer composition ABCD.

In a second example, two phase emulsions are optimized through the combination of several surfactant packages and phase to phase ratios and screened for several attributes. One key performance criterion is to minimize the settling upon standing at different temperatures after a given period of time. It is also critical to accurately reduce in a ml. volume the emulsification conditions that prevail in large scale in order to produce the same particle size distribution. This was made possible by using a modified version of the SCPPR™ described previously. The conventional approach to estimate the emulsion stability is to visually inspect the emulsion settling pattern and rank it according to a predetermined scale. The method is converted in a fully automated protocol whereby an image of each dispersion-containing vial is captured and analyzed by Symyx proprietary form recognition software to quantify the settling pattern. As this is often the case in combinatorial chemistry, a validation phase is necessary to insure that the high throughput screen output is, if not identical, at least statistically related to the conventional measurement output. This was indeed the case for this study where the emulsion libraries were made and screened and confronted with a reduced set of data acquired through conventional characterization techniques.

Bulk formulation of polymers

Polymer blends are used in daily life in the forms of thermoplastics for food packaging, engineering plastics in textiles and sportswear, and rubbers in tires, carpets, and hot melt adhesives, to name a few. These industries rely heavily in the art of compounding to offer various property profiles using commercial polymer grades and a range of additives. Applying a combinatorial approach to polymer compounding makes sense as long as one masters some important aspects such as blending and mechanical property testing. The former has been solved at Symyx by pursuing different blending options. The latter has resulted several methods and tools being conceived and built that are now routinely used in discovery programs aimed at the development of novel thermoplastic elastomers. A parallel Dynamic Mechanical Thermal Analysis (DMTA) [8] instrument measures the elastic and loss moduli versus temperature under controlled relative humidity of 96 polymer samples at a time. These samples are robotically deposited on a polyimide membrane with a controlled thickness monitored by laser profilometry. The plate supporting the membrane is translated in the Z direction and brought in contact with an array of 96 probe sensors. A metal pin mounted on the force sensor is touching on the reverse side of the membrane and records the stiffness on a static or harmonic mode. The set up is, by and large, equivalent to a three point fixture mechanical testing system and provides elastic and loss moduli in a three decade frequency range. A typical plot of modulus vs. temperature is given in *figure 6* and witnesses of the remarkable sensitivity of the instrument, comparable with a state-of-the art bench DMTA. Interestingly, this device can serve different purposes by switching to other configurations and allowing for tack and strength to failure measurements as well. This mechanical screen was instrumental in the development of a new generation of hydrophilic thermoplastic elastomers using living free radical polymerization [9]. Melt rheology is also screened in a parallel manner with a concept that utilizes a micromachined wafer with embedded microsensors whose details are given in *figure 7*. Each polymer sample is sandwiched between a fixed plate (sensor) and a moving plate and the force needed to keep the fixed plate stationary as the moving plate oscillates is recorded. This measures fluid moduli as functions of frequency, strain, time, and environmental characteristics (*e.g.,* temperature) and has an operating range in the frequency domain and a moduli range comparable to those of conventional instruments (*figure 8*). Both instruments operate in a truly parallel mode that is particularly valuable for low frequency measurements, as opposed to conventional methods operating in sequence which are extremely time consuming.

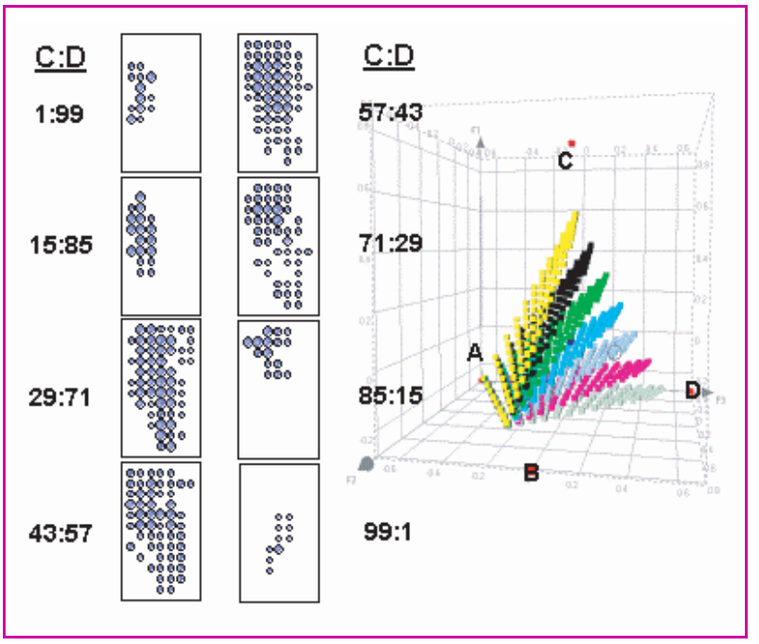

Figure 5 - Hit representation of polymer dispersants for nanodispersions.
The polymer are prepared by free radical polymerization of monomers A, B, C and D. Left: particle sizes of nanodispersion as a function of polymer composition A/B for a given ratio C/D (the spot size is inversely proportional to the hydrodynamic diameter of the particles). Right: library design in a pseudo quaternary representation.

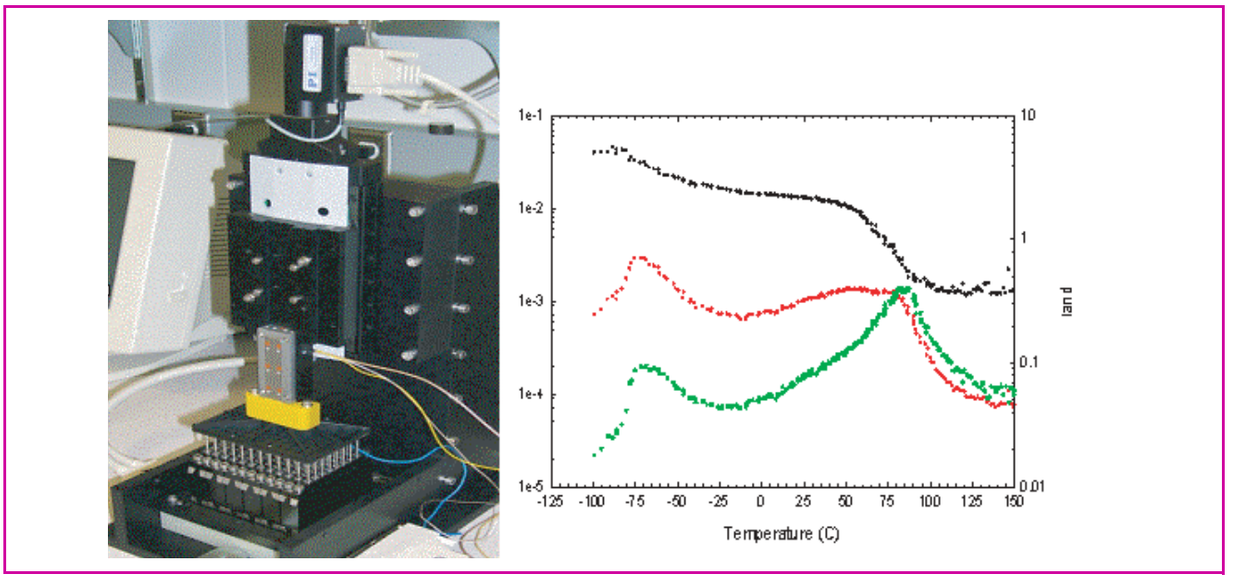

Figure 6 - Left: 96 channels Parallel Dynamic Mechanical Analysis instrument. Left: DMA spectrum of a thermoplastic elastomer material averaged out of the 96 library elements.

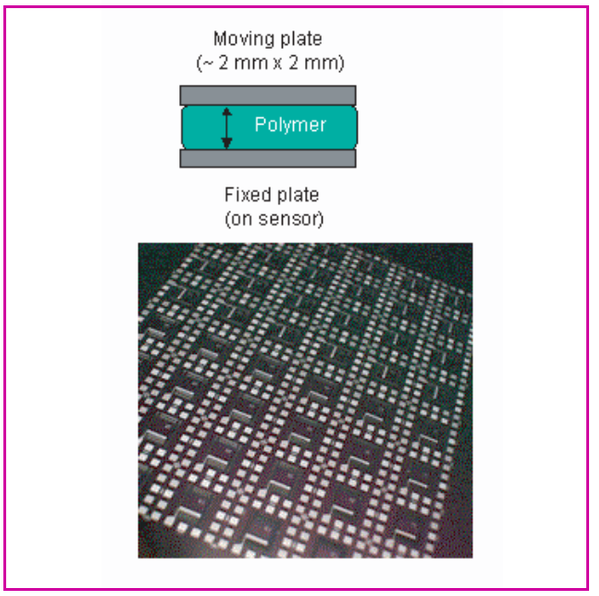

Figure 7 - Prototype of a 48 channels parallel rheometer.

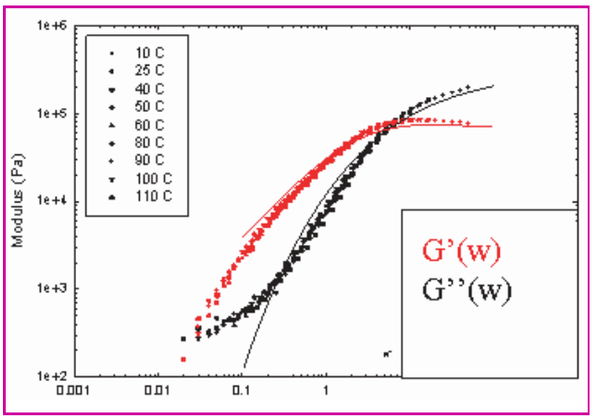

Figure 8 - Rheology master curves of polymer samples obtained on the parallel rheometer.

Acknowledgments

I would like to thank all the people at Symyx who have conceived and built this infrastructure and thus given the substance to this paper, in particular Dr. Eric Carlson, Dr. Han Ting Chang, Dr. Damian Hajduk, Dr. Oleg Kolosov, Dr. Sigrid Kubler, Dr. Paul Mansky, Dr. Miroslav Petro, Dr. Adam Safir and Dr. Tom McWaid.

References

[1] Engstrom J.R., Weinberg W.H., AIChE J., 38 (17), **2000**, *461*, p. 2.
[2] Jandeleit B., Shaefer D.J., Power T.S., Turner H.W., Weinberg W.H., *Angew. Chem.*, **1999**, *111*, p. 2648.
[3] Lacy S., *PCT Int. Appl.*, **2000**, WO 00/23931.
[4] Rust W., *PCT Int. Appl.*, **2000**, WO 00/67086.
[5] Nielsen R.B., Safir A., Tiede R., McWaid T., Vanerden L., *PCT Int. Appl.*, **2001**, WO 01/93998.
[6] Petro M., Safir A.L., Nielsen R.B., *Polym. Prepr.* (Am. Chem. Soc., Div. Polym. Chem.), **1999**, *40 (2)*, p. 702.
[7] Mansky P., Bennett J., *PCT Int. Appl.*, **2000**, WO 00/36410.
[8] Hajduk D.A., EP 1158290.
[9] Hajduk D.A., Abstracts of Papers, 222nd ACS National Meeting, Chicago (IL, USA), August 26-30, **2001**, Poly-338.

Dominique Charmot
est Directeur de recherches des polymères chez Symyx Technologies Inc.*

* Symyx Technologies Inc., Santa Clara, CA 95051, États-Unis.
E-mail : dcharmot@symyx.com
http://www.symyx.com