

# Apport de la Méthodologie de la Recherche Expérimentale à un modèle de la température de transition vitreuse de polymères acryliques et méthacryliques

P. Camelio<sup>1,\*</sup>, B. Campisi<sup>2</sup>, A.M. Carro-Diaz<sup>3</sup>, V. Lazzeri, B. Waegell et R. Phan-Tan-Luu<sup>4</sup>

<sup>1</sup>Laboratoire de Stéréochimie, D.U. de Saint-Jérôme, Case 532, 13397 Marseille Cedex 20, France

<sup>2</sup>Dipartimento di Economia e Merceologia delle Risorse Naturali e della Produzione, Universtà di Trieste, V. Valerio 6, I-34127 Trieste, Italia

<sup>3</sup>Dpto. Química Analítica, Nutrición y Bromatología, Facultad de Química, Universidad de Santiago de Compostela, Avda de las Ciencias S/N 15706-Santiago de Compostela, Espagne

<sup>4</sup>Laboratoire de Méthodologie de la Recherche Expérimentale, D.U. de Saint-Jérôme, court D. 13397 Marseille Cedex 20, France

**D-optimal design has been used for constructing a experimental database required to set up a high quality model. We show that using a priori statistics criteria (i.e., before making the experiences) it is possible to select the minimum number of molecules required to set up the database. Using the same tools, and the previous results, it is possible to evaluate the quality of the predicted value for a set of new candidates and decide whether or not they have to be taken into account in the database.**

Depuis de nombreuses années, le laboratoire de Stéréochimie travaille au développement de monomères acryliques originaux [1-2] afin d'améliorer les propriétés de matériaux déjà existants en utilisant ces monomères comme comonomères ou bien en développant de nouveaux matériaux sur la base de ces monomères. Un des problèmes majeurs dans ce type de recherche est que l'on ne connaît pas à l'avance les propriétés macroscopiques des matériaux que l'on obtient à partir des monomères. Pour obtenir un polymère, il faut dans un premier temps synthétiser le monomère, puis le polymériser et enfin, on caractérise les matériaux obtenus. Si les propriétés recherchées sont obtenues, on peut passer à une phase de développement, sinon, il faut recommencer en modifiant la structure du monomère. Cette approche est lente car elle repose sur un processus itératif d'essais-erreurs. De plus, elle dépend essentiellement de l'expérience de l'homme de l'art et de son imagination. Pour rationaliser ce travail, l'utilisation d'un modèle pour une ou plusieurs propriétés basées sur la structure du monomère est d'une grande importance car on peut estimer la (les) propriété(s) rapidement d'un grand nombre de molécules en s'affranchissant dans un premier temps des problèmes liés à l'expérience. L'homme de l'art consacrera alors ses efforts sur les molécules qui présentent un intérêt potentiel au regard des résultats du modèle. C'est une démarche appliquée depuis un certain nombre d'années

en chimie pharmaceutique ou en agrochimie (pesticides, fongicides, ...) et qui tend à se développer en chimie des polymères.

C'est dans ce but que nous avons développé un modèle de prédiction de la température de transition vitreuse,  $T_g$ , pour les polymères acryliques et méthacryliques : le modèle EVM (pour Énergie, Volume et Masse) [3-7]. La  $T_g$  est une propriété importante en chimie des polymères, car on observe à son passage de brusques variations d'un grand nombre de propriétés physico-chimiques. Sa connaissance permet de définir une température limite d'utilisation ou de mise en œuvre du matériau. Ce modèle original combine des résultats de Mécanique et Dynamique Moléculaires (MM/MD) à une approche classique de corrélation structure-propriétés (QSPR : *Quantitative Structure Property Relationship*). C'est la physique du phénomène étudié qui a guidée le choix des descripteurs (Énergie de van der Waals, Énergie interne, Volume, Masse) et leur combinaison sous la forme d'une fonction de densité d'énergie. La forme générale du modèle EVM est la suivante :

$$T_g = A + B \times \frac{E_{vdw} \times M}{V} + C \times \frac{E_{int} \times M}{V} \quad (1)$$

où les termes  $E_{vdw}$  et  $E_{int}$  sont les termes d'énergie de van der Waals et d'énergie interne d'un segment de polymère dont la géométrie a été optimisée par MM/MD,  $V$  caractérise le volume représentatif d'un oligomère (comprenant 20 monomères) et  $M$  est la masse moléculaire de cet oligomère normalisée par le nombre de monomères le composant.

Jusqu'à présent, les coefficients  $A$ ,  $B$  et  $C$  du modèle avaient été estimés par régression linéaire multiple (MLR) pour une famille de polymères donnée (polyacrylates [6,7] ou polystyrènes [7-9]). Connaissant, pour cette famille de composés qui constitue notre matrice d'expériences ( $\xi_N$ ), les réponses expérimentales ( $Y$ ), nous pouvons déterminer, à partir de la matrice du modèle ( $X$ ), les estimateurs des coefficients du modèle ( $B$ ) (Fig. 1).

Une série de tests statistiques permet ensuite de voir quels sont les niveaux de signification des résultats obtenus par l'Analyse de la Variance et avec le test de Student  $t$  sur les coefficients). C'est la démarche *a posteriori*. Ce modèle donnait pleinement satisfaction, au regard des informations

\* E-mail: philippe.camelio@wanadoo.fr

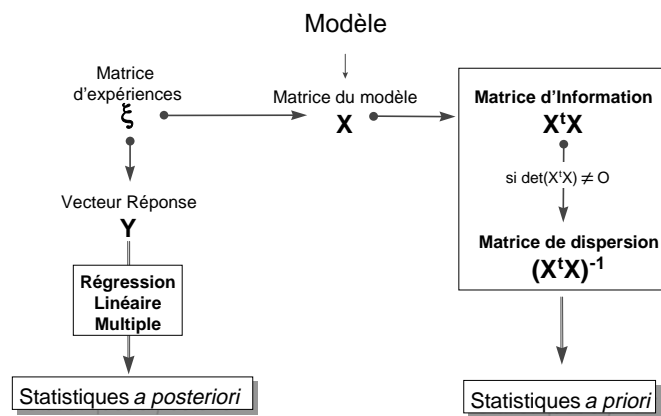


Figure 1. Les deux démarches de la Méthodologie de la Recherche Expérimentale.

recherchées. Néanmoins, les  $T_g$  utilisées pour la MLR étant pour une grande partie issues de la littérature, il n'est pas rare de trouver plusieurs valeurs de la  $T_g$  pour un même composé, jusqu'à 11 valeurs pour le poly(méthyl méthacrylate), et cela avec une dispersion importante (30 K), bien au-delà de l'erreur expérimentale. Cette multiplicité des réponses est liée aux nombreux facteurs dont dépend la  $T_g$  et qui ne sont pas pris en compte dans le modèle tels que : la stéréorégularité, la régiospécificité, la distribution des masses molaires moyennes, ou encore les techniques de caractérisation, ... Cette variabilité importante de la réponse expérimentale fait que l'incertitude sur les résultats obtenus par le modèle est plus importante en pratique que ne le laisserait supposer la variance du modèle.

Pour améliorer la qualité prévisionnelle du modèle (c'est-à-dire pour estimer les  $T_g$  avec la plus faible variance possible) il faut disposer de données expérimentales homogènes, connaître la  $T_g$  avec précision pour un certain nombre de polymères, ces derniers ayant été polymérisés et caractérisés dans des conditions bien précises. Pour constituer cette base de données expérimentales (ou matrice d'expériences), il faut répondre aux deux questions suivantes :

*Combien de polymères faut-il caractériser ? Lesquels ?*

### Développement

Compte tenu de la nature particulière du problème, nous ne pouvons pas utiliser les matrices d'expériences classiques. En effet, nous sommes dans un cas où nous avons un nombre fini de points (l'ensemble des monomères candidats), c'est-à-dire un ensemble non-connexe où chaque point objet est représenté par ses descripteurs ( $E_{vdw} \times Mw/V$  et  $E_{int} \times Mw/V$ ), le domaine expérimental est discret. L'objectif est de sélectionner un sous-ensemble de monomères (matrice d'expériences) qui va nous permettre de construire un modèle prédictif de bonne qualité, c'est-à-dire dont la variance sera la plus petite possible. Pour sélectionner ce sous-ensemble, nous avons utilisé une technique de construction de matrice d'expériences particulières, les

matrices optimale-D, et des critères de qualité disponibles *a priori* (avant expériences). Ce sont des outils et techniques développés dans le cadre de la Méthodologie de la Recherche Expérimentale (MRE) [10].

Un des postulats de base de la MRE est qu'« il n'y a aucune information dans le résultat d'une expérience et que toute l'information est contenue dans les conditions expérimentales » [10]. Dans notre cas, le résultat d'une expérience correspond à une  $T_g$  expérimentale et les conditions expérimentales correspondent aux deux variables de l'équation (1).

### Facteurs d'inflation

Dans une première étape, il faut vérifier que l'ensemble des points que l'on propose est cohérent pour le modèle postulé, c'est-à-dire savoir si l'information apportée par les molécules est suffisante. Pour cela, le calcul des facteurs d'inflation est effectué.

Le facteur d'inflation d'un coefficient  $b_i$  (estimateur de  $\hat{\beta}_i$ ) est l' $i^{\text{ème}}$  élément diagonal de la matrice inverse de la matrice de corrélation [11-13].

$$f(\hat{\beta}_i) = c^{ij} \sum_j (x_{ij} - \bar{x}_j)^2 \quad (2)$$

Ce facteur d'inflation est une mesure de l'indépendance des coefficients. Dans le cas où la matrice d'expériences est optimale (matrice d'information  $X'X$  diagonale), les facteurs d'inflation sont tous égaux à l'unité. Une valeur supérieure à 1 indique que la matrice perd de son optimalité et on admet que lorsqu'un facteur d'inflation a une valeur supérieure à 4 ou 7 (suivant les auteurs), l'information apportée par la matrice d'expériences n'est plus de qualité suffisante pour le modèle postulé. Pour cela, les facteurs d'inflations ont été calculés et sont égaux à 1,37 pour les deux descripteurs. Puisque les facteurs d'inflation sont voisins de 1, le modèle EVM peut être étudié à l'intérieur du domaine expérimental composé par les 33 molécules (Fig. 2).

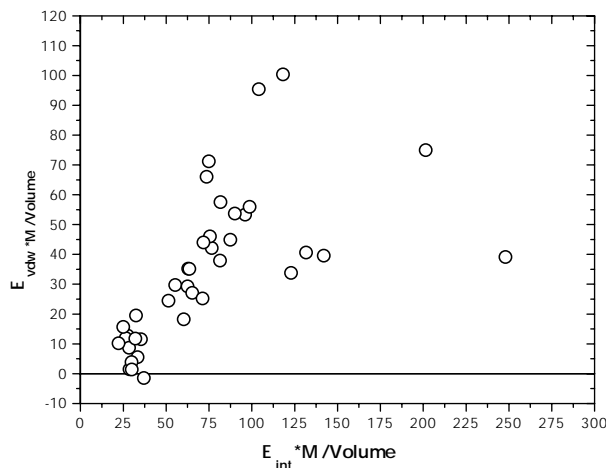


Figure 2. Représentation du domaine expérimental.

### Construction des matrices optimale-D

L'étape suivante consiste en la sélection de monomères pour constituer la base de données. Il en faut au minimum 3 puisque le modèle comprend trois coefficients ( $A$ ,  $B$ ,  $C$ ), et nous avons considéré que le nombre de 25 était un maximum. Sur la base d'un algorithme d'échange, nous avons construit les matrices optimale-D pour un nombre d'expérience donné (matrice d'expériences  $X$  qui maximise le déterminant de la matrice d'information  $X'X$ ) [14-15]. Plus il y a de molécules dans la matrice d'expériences, plus celle-ci contient d'informations. Ceci est représenté par la figure 3 qui montre l'évolution du déterminant de la matrice d'information normée en fonction du nombre de points candidats la constituant.

Lors de cette étape, les éléments de la matrice d'expériences, c'est-à-dire la base de données, ont été choisis pour un nombre d'expériences fini,  $N$ . Il faut maintenant choisir quelle est la matrice qui apporte une information suffisante, tout en ayant le plus petit nombre de composés.

### Critères de choix

Deux critères *a priori* sont alors utilisés pour sélectionner la matrice d'expériences recherchée : le critère-A et la fonction de variance maximale  $d_{\max}$  [14].

Le critère-A, aussi appelé critère de A-optimalité (Average optimal) représente la qualité globale des estimateurs des coefficients.

On établit que la matrice de variance-covariance  $[\hat{\beta}]$  peut s'écrire sous la forme suivante :

$$\text{var}[\hat{\beta}] = (X'X)^{-1}\sigma^2 \quad (3)$$

où  $(X'X)^{-1}$  est la matrice de dispersion qui ne dépend que des expériences qui constituent la matrice  $X$ , et  $\sigma^2$  la variance expérimentale qui est indépendante des résultats expérimentaux. La variance d'un estimateur,  $\hat{\beta}_i$ , est obtenue en multipliant la variance de l'erreur expérimentale  $\sigma^2$  par le terme diagonal de la matrice de dispersion ( $c^{ii}$ ). Si l'on fait la somme des variances des estimateurs, on obtient l'équation suivante :

$$\sum_i \text{var}[\hat{\beta}_i] = \sum_i c^{ii}\sigma^2 = \text{trace} (X'X)^{-1}\sigma^2 \quad (4)$$

On constate alors que la somme des variances des estimateurs, c'est-à-dire leur qualité globale, peut être calculée uniquement à partir de la trace de la matrice de dispersion. Elle est donc indépendante de la valeur du résultat des expériences. Minimiser la trace de la matrice de dispersion revient alors à améliorer la qualité globale du modèle par une meilleure estimation des coefficients du modèle.

La fonction de variance maximale,  $d_{\max}$ , correspond à la plus grande valeur des fonctions de variance calculées pour l'ensemble des points proposés. La variance de la réponse

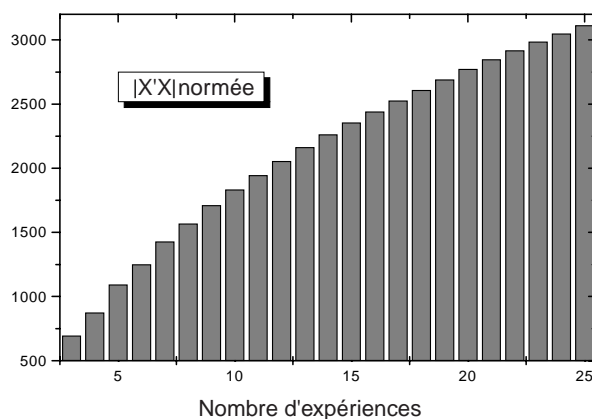


Figure 3. Variation du déterminant normé de la matrice d'information en fonction du nombre d'expériences.

prévisionnelle pour le point considéré  $i$  (représenté par le vecteur  $x_i$ ),  $\hat{y}_i$ , est égale à :

$$\text{var}(\hat{y}_i) = x_i'(X'X)^{-1}x_i\sigma^2 = d_i\sigma^2 \quad (5)$$

$d_i$  est appelé fonction de variance du point  $i$ . On constate que la variance de la prévision au point  $i$  ne dépend que du point  $i$  considéré, de la variance expérimentale et des éléments de la matrice de dispersion, c'est-à-dire de la matrice d'expériences et de la forme du modèle.

La figure 4 représente les variations de la trace de la matrice de dispersion et de la fonction de variance maximale pour les matrices optimale-D choisies pour un nombre d'expériences donné. On constate que la qualité des matrices augmente avec le nombre d'expériences. Ceci est en accord avec le fait que plus il y a de points dans la matrice d'expériences, plus celle-ci contient d'information. Il faut donc choisir une matrice optimale-D en fonction d'une qualité que l'on va juger suffisante, par rapport à l'information que l'on recherche, tout en gardant à l'esprit que l'on cherche à

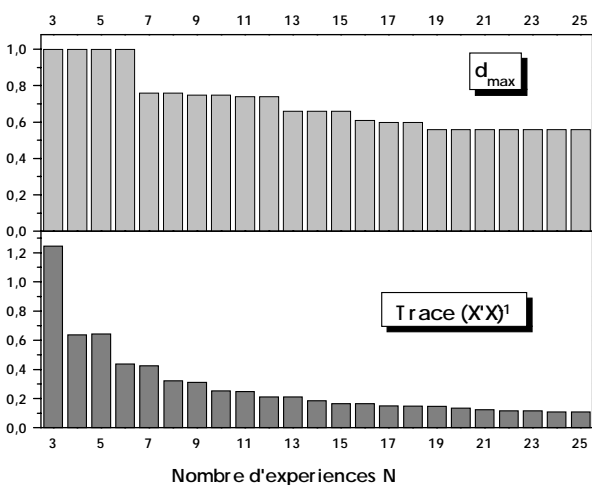


Figure 4. Évolution de la trace de la matrice de dispersion et de la fonction de variance maximale avec le nombre d'expériences (molécules).

construire la base de données expérimentales contenant le moins de molécules possibles.

Nous avons choisi la matrice à 7 candidats (Fig. 5), car elle satisfait à nos *desiderata*, à savoir, une bonne qualité prévisionnelle ( $d_{\max}$  inférieur à 1) et des estimateurs des coefficients ayant une faible valeur de la trace de la matrice de dispersion. Si l'on considère des matrices comprenant plus d'expériences, on n'observe pas d'amélioration significative de la qualité des résultats.

### Étude de nouvelles molécules

Une série de 52 monomères pour lesquels le laboratoire porte un intérêt particulier a ensuite été étudié dans le but de prédire la  $T_g$  des polymères correspondants. La majorité des molécules n'a pas été synthétisée et l'objectif est de pouvoir, grâce au modèle, choisir les molécules potentiellement intéressantes. Mais avant toute prédiction, il faut se poser la question suivante :

#### Le modèle construit à partir de la matrice à 7 expériences permet-il une prédiction de bonne qualité (faible fonction de variance) pour ces nouvelles molécules ?

Pour cela, il suffit de représenter, dans le domaine expérimental, les courbes d'isovariance (à la variance expérimentale près), calculées à partir de la matrice d'expériences choisie. Puis on projette la série de composés dans cet espace (Fig. 7). On constate alors que les prévisions pour ces composés sont de bonne qualité (fonction de variance toujours inférieure à 1). En d'autres termes, les 7 molécules choisies précédemment apportent une information suffisante pour avoir un modèle prédictif de qualité pour cette série.

### Conclusion

La nécessité de disposer de données expérimentales de qualité, afin d'établir des modèles de corrélation structure-propriété, pose le délicat problème de leur constitution. La Méthodologie de la Recherche Expérimentale permet d'apporter une réponse efficace et simple à mettre en œuvre pour choisir les molécules candidates à la constitution d'une base de données expérimentales. La limite principale de ce type d'approche est qu'elle suppose l'existence d'un modèle ; dans le cas présent, le modèle a été développé en amont [3-7]. Actuellement, la constitution de cette base de données expérimentales est en cours de réalisation (synthèse des monomères, polymérisation et caractérisation des 7 polymères choisis).

La possibilité de prendre en compte des facteurs extérieurs au modèle, comme le coût de la synthèse, dans la phase de construction des matrices optimale-D (fonction de pondération introduite par D. Mathieu [14]) permet d'apporter d'autres solutions.

Enfin, s'il est nécessaire de disposer d'un ensemble de points pour construire un modèle (matrice optimale-D), il faut de plus disposer de points-tests qui vont permettre de valider le modèle dans le domaine expérimental considéré.

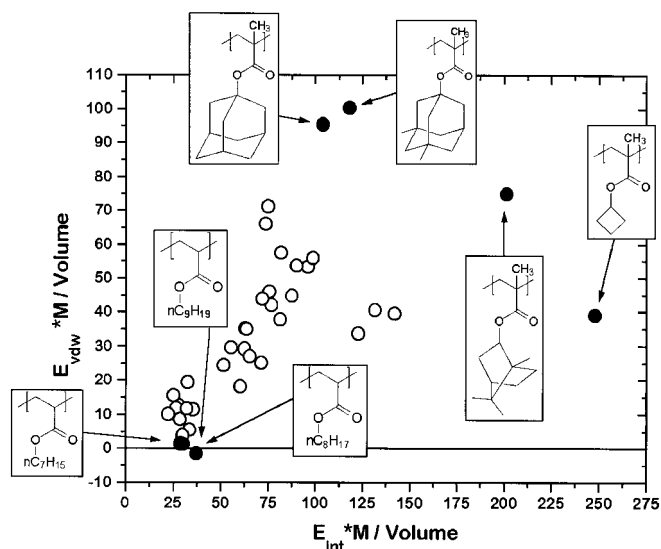


Figure 5. Matrice d'expériences optimale-D à 7 molécules.

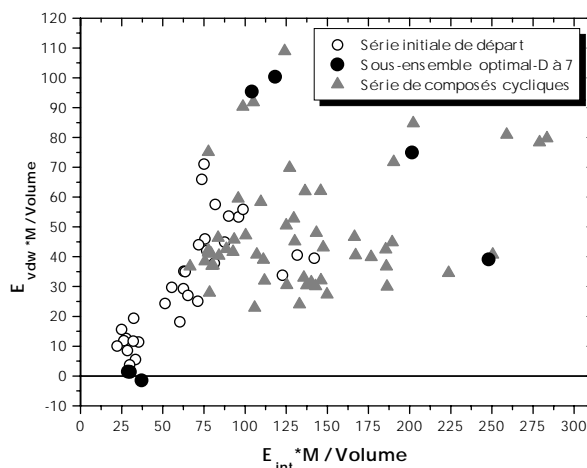


Figure 6. Nouvelle famille de monomères étudiée (cf. Tab. II).

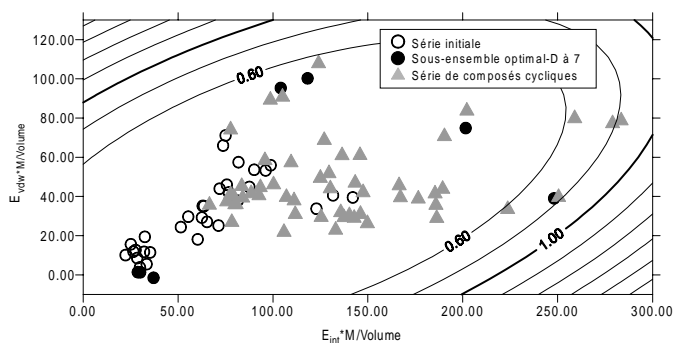


Figure 7. Courbes d'isovariance projetées dans le domaine expérimental.

On utilise pour cela les matrices uniformes [11]. Une étude qui prend en compte ces deux aspects est en cours de réalisation.

### Références

1. Ferret, N. Thèse de doctorat en sciences, Université d'Aix-Marseille III, 1992.
2. Augier-Dimino, E. Thèse de doctorat en sciences, Université d'Aix-Marseille III, 1997.
3. Camelio, P. Corrélations Structure Propriété de polymères amorphes, D.E.A.SM2B, Université d'Aix-Marseille III, 1994.
4. Camelio, P. et al. *Polym. Prep. (Am. Chem. Soc., Div. Polym. Chem.)* **1995**, 36(1), 661.
5. Camelio, P. et al. *J. Polym. Sci. Part A: Polym. Chem.* **1997**, 35, 2579.
6. Cypcar, C. C. et al. *Macromolécules* **1996**, 29, 8954.
7. Camelio, P. Thèse de doctorat en sciences, Université d'Aix-Marseille III, Déc. 1997.
8. Camelio, P. et al. *Polym. Prep. (Am. Chem. Soc., Div. Polym. Chem.)* **1996**, 37(2), 252.
9. Camelio, P. et al. *Macromolécules* 1998, à paraître.
10. Mathieu, D.; Phan-Tan-Luu, R. *Récents Progrès en Génie des Procédés* **1995**, 36(9), 1.
11. Sergent, M. Thèse de doctorat en Sciences, Université d'Aix-Marseille III, 1989.
12. Carlson, R. *Design and optimization in organic synthesis*, Elsevier Amsterdam, 1992.
13. Fedorov, V. V. *Theory of optimal experiments*, Academic Press, New-York, 1972.
14. Mathieu, D. Thèse de doctorat en sciences, Université d'Aix-Marseille III, 1981.
15. Mathieu, D. et al. *Bull. Soc. Chim. Belg.* **1980**, 89(4), 267.