

## **Polymer Property Modelling using Grid Technology for Design of Structured Products**

Kavitha Chelakara Satyanarayana,<sup>a</sup> Jens Abildskov,<sup>a</sup> Rafiqul Gani<sup>a</sup>

<sup>a</sup> CAPEC, Department of Chemical Engineering, Technical University of Denmark, DK-2800 Lyngby, Denmark

### **Abstract**

Development of group contribution plus models, for the prediction of polymer properties from the drawn chemical structure, which can be later used in computer aided molecular design (CAMD) for predicting the polymer structure from the given set of target properties is the goal of this work. In general, property prediction for a given polymer structure using group contribution models require that the structure can be fully represented by groups with well-defined contributions for that particular property. Frequently this can not be accomplished. To overcome this limitation a group contribution<sup>+</sup> approach is developed here, for polymers. With this approach the missing groups are created and their contributions are predicted using zero and first order connectivity indices. To minimize the time spent on computational aspects of model development, grid technology is used in this work. A comparison of the times required for generating the models in series and in parallel, using grid technology, is also presented

Property prediction methods for linear polymers are presented for the amorphous and crystalline volume, glass transition temperature, solubility parameter and refractive index.

Keywords: polymers, property prediction, grid technology, structured products

### **1. Introduction**

To satisfy consumer and regulatory demands, chemical and pharmaceutical industries continuously must focus on developing and producing improved/modified products. Previous trial-and-error methodologies to produce new products, with specific desired properties, are not feasible as these are capital intensive, time-consuming and difficult to automate. Computer-aided molecular design (CAMD) methods, can partially replace these in determining molecular structures matching specific sets of target properties. Extending CAMD techniques for identifying macromolecular (here polymers) structures having a set of target properties, requires new property models.

Developing new property models for predicting polymer properties based on the analysis of polymer structure is the main objective of this work.

Relationships between polymer repeat unit and polymer properties has been known in the literature for decades. Most notably, the group contribution methods of van Krevelen [1] and the connectivity index based methods by Bicerano [2]. The group contribution method of Van Krevelen is often quite accurate, but has a lot of less obvious features, like a large number of corrections to nearest neighbour interactions in the basic level of property prediction. For example, to enhance accuracy, the same group has different contributions depending on its position in the polymer repeat unit (side chain, main chain etc.). The groups defined by this method are relatively large fragments. These drawbacks limit the flexibility of this method and its usefulness in computer-aided polymer design algorithms.

The Marrero/Gani group contribution method [3] has advantages with respect to CAMD based polymer design. It has a large range of groups, classified as either first-order, second-order and third-order groups, and may be used to generate new polymer property prediction models as well as generation of polymer repeat unit structures. However, all group methods need the entire polymer repeat unit structure to be described by the available groups considered in that method, in order for the relevant property to be predicted. For example, if we consider the GC methods reported by Marrero and Gani [3], the polymer repeat unit structure shown in Figure 1 cannot be fully described and consequently its properties cannot be estimated.

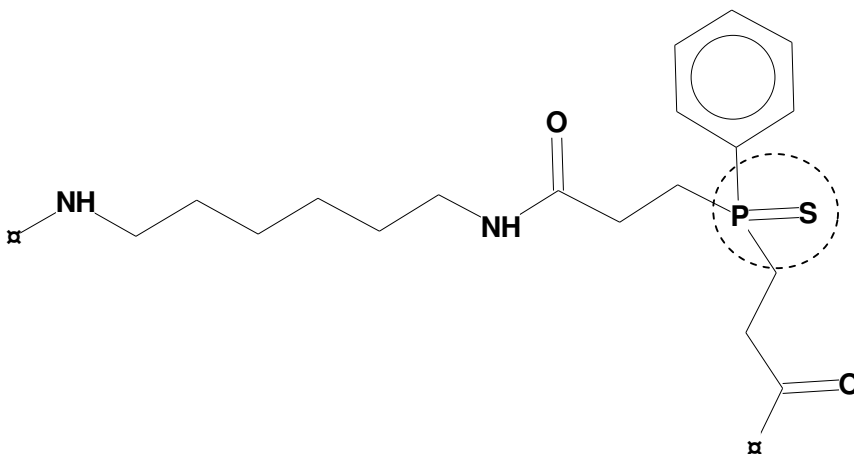


Fig. 1. Incomplete molecular structure representation – the circled group does not exist in the group tables for the Marrero/Gani method [3]. The squares denote the free ends of the repeat unit structure.

Given the limitations on the number of reliable data on polymer systems, one can not expect to generate Marrero/Gani group tables covering large ranges of groups without missing entries. Thus, such a method may not be very widely applicable.

Wide range of application due to general applicability is one advantage of the method of Bicerano, although these methods frequently incorporated measured data for one property in the estimation of another. For example the estimation of glass transition temperatures with the method of Bicerano requires knowledge of the solubility parameter of the polymer. Nevertheless, the method of Bicerano, obviates the need for group contributions and the method can in principle handle all polymers constructed from nine elements (C, H, O, N, F, Si, S, Cl and Br).

To retain the advantages of group contribution methods and to overcome the limitation of their missing contributions, combined group contribution / connectivity index (GC – CI) methods or group contribution<sup>+</sup> have been proposed by Gani et al [4], with emphasis on solvents and other low-molecular weight substances. The underlying idea is that missing groups are automatically created with the CI-based method, and their corresponding contributions are estimated.

This paper presents the development of a GC-CI framework for polymer property prediction. The property of a given polymer repeat unit is predicted using a group contribution method based on the group definitions of Marrero/Gani [3], if the entire molecule can be represented by the Marrero/Gani [3] groups and all groups have values. If not, the missing group is generated and the missing group contribution is predicted using a connectivity index based model, in a way resembling that of Harper et al [4]. The group is added to the GC model, which finally forms the GC-CI model.

## 2. Polymer Property Prediction Models

### 2.1 Models

Inductive learning process, where the experimental data is collected and a model is derived from the data set and properties are predicted, is used for developing models. The essentials of GC-CI framework is described in the section below:

#### 2.1.1 Marrero/Gani Group Contribution Model

The Marrero/Gani group contribution method is usually written as follows:

$$f(X) = \sum_i N_i C_i + w \sum_j M_j D_j + z \sum_k O_k E_k \quad (1)$$

where,  $C_i$  is the contribution of the first-order group of type- $i$  that occurs  $N_i$  times,  $D_j$  and  $E_k$  are the contributions of the second-order group of type- $j$  and the third-order group of type- $k$ , that occurs  $M_j$  and  $O_k$  times, respectively. In the first level of estimation, the constants  $w$  and  $z$  are assigned zero values because only first-order groups are employed. In the second level, the constant  $w$  is unity and  $z$  is zero, because only first- and second-order groups are involved. In the third level, both  $w$  and  $z$  are unity. The left-hand side of Eq. (1) is a simple function of  $f(X)$  of the property,  $X$ . The selection of this function has been based on the following criteria:

(a) The function has to achieve additivity in the contributions  $C_i$ ,  $D_j$  and  $E_k$

- (b) It has to exhibit the best possible fit of the experimental data  
 (c) It should provide good extrapolating capability and therefore, a wide applicability  
 Thus the procedure is the same as in the case of low molecular weight substances, except that here a set of Marrero/Gani [3] groups describes polymer repeat unit structures.

### 2.1.2 Atom Connectivity Index Model

Polymer repeat unit structures can also be given by an atomic representation for connectivity index based models, as written by Gani et al [4]:

$$f(Y) = \sum_i (a_i A_i) + b({}^v\chi^0) + 2c({}^v\chi^1) + d \quad (2)$$

where  $Y$  is the polymer property to estimate,  $A_i$  is the number of atom  $i$  occurring in the molecular structure  $a_i$  times.  ${}^v\chi^0$  and  ${}^v\chi^1$  is the zeroth- (atom) and first-order (bond) connectivity index, respectively, as described by Kier and Hall [5].  $b$ ,  $c$  and  $d$  are adjustable parameters. Such models have very few adjustable parameters:  $a_i$  (one per atom  $i$ ),  $b$ ,  $c$  and  $d$ . Thus such models cannot be expected to give high precision predictions of  $Y$  since, with only a few parameters, a large set of compounds will be represented. But to overcome the missing groups/ group parameters for polymer property prediction, this model can be used together with GC model.

## 3. GRID Computing

Grid computing enables higher throughput computing by taking advantage of many networked computers to model a virtual computer architecture capable of distributing process execution across a parallel infrastructure [6]. Here we use 'OfficeGRID' [7] of MESH-technologies. With 'OfficeGRID', the user can submit the jobs from command line. It also supports the working of different platforms simultaneously. Moreover, it provides good flexibility as the machines can join and leave the grid freely and the machines can accept one or more jobs concurrently. 'OfficeGRID' also includes advanced error handling like detecting node failure, automatic resubmission of jobs sent to the node, and network failures - even failures during large network transfers - are dynamically corrected.

Three 'client' computers are connected to one 'master' computer for this work. The job is submitted to the master computer, which then looks for the freely available clients and assigns the job to it. Once the job is completed the client sends back the completed job to the master, which finally gives back the results.

## 4. Methodology

Contributions of the groups/atoms and constants are determined by fitting the model to the experimental data. That is, by adjusting parameter values to minimize a sum of squared residuals (differences between experimental and estimated property values).

The Levenberg/Marquardt [8] minimization method is used for solving the least squares problem.

Using GRID technology property models and parameter estimation can be generated in parallel for different properties. This not only reduces the time of generating the models by using the freely available computer resources. To evaluate the grid technology, the regression code for estimating parameters for Eq (1) and Eq (2) were executed in series and the total time taken for execution of twelve runs was 198 seconds. These codes were also executed such that the minimizations were made parallelly using 'OfficeGRID'. The total time taken for these twelve runs execution was just 70 seconds. This reduced the time for model prediction considerably. Even though, the model development is not so time consuming at this moment, in course of time, these models will be extended for predicting polymer properties for polymer structure with several side chains (repeated at some frequency, etc.), where the computations could be really massive and the grid technology will be needed to develop the necessary models for a large range of properties within a reasonable period of time.

Table 1 GC and CI based property models

Properties	GC-based model
Glassy Amorphous Volume ( $V_g$ )	$M_w/\rho_g - 1.0644 = \sum_i N_i V_{g1i} + w \sum_j M_j V_{g2j} + z \sum_k O_k V_{g3k}$
Rubbery amorphous Volume ( $V_r$ )	$M_w/\rho_r + 1.1152 = \sum_i N_i V_{r1i} + w \sum_j M_j V_{r2j} + z \sum_k O_k V_{r3k}$
Amorphous Volume ( $V_a$ )	$M_w/\rho_v + 0.9412 = \sum_i N_i V_{a1i} + w \sum_j M_j V_{a2j} + z \sum_k O_k V_{a3k}$
Crystalline Volume ( $V_c$ )	$M_w/\rho_c + 1.149 = \sum_i N_i V_{c1i} + w \sum_j M_j V_{c2j} + z \sum_k O_k V_{c3k}$
Glass transition temperature ( $T_g$ )	$T_g \cdot M_w + 1236.52 = \sum_i N_i Y_{g1i} + w \sum_j M_j Y_{g2j} + z \sum_k O_k Y_{g3k}$
Solubility parameter ( $\delta$ )	$\delta^2 V_a - 3274.698 = \sum_i N_i H_{coh1i} + w \sum_j M_j H_{coh2j} + z \sum_k O_k H_{coh3k}$
Refractive Index (n)	$R^d - 0.7258 = \sum_i N_i R1i + w \sum_j M_j R2j + z \sum_k O_k R3k$
CI-based model	
Glassy Amorphous Volume ( $V_g$ )	$M_w/\rho_g = \sum_i a_i A_{gi} + 6.2466(\chi^0) - 5.4192(\chi^1) + 2.3368$
Rubbery amorphous Volume ( $V_r$ )	$M_w/\rho_r = \sum_i a_i A_{ri} - 3.3445(\chi^0) - 1.4868(\chi^1) + 0.38925$
Amorphous Volume ( $V_a$ )	$M_w/\rho_a = \sum_i a_i A_{vi} + 4.5318(\chi^0) - 1.5182(\chi^1) - 1.15$
Crystalline Volume ( $V_c$ )	$M_w/\rho_c = \sum_i a_i A_{ci} - 1.4664(\chi^0) + 1.9719(\chi^1) - 1.81$
Glass transition temperature ( $T_g$ )	$T_g \cdot M_w = \sum_i a_i A_{gi} + 3935.242(\chi^0) - 11047.9(\chi^1) + 770.402$
Solubility parameter ( $\delta$ )	$\delta^2 V_a = \sum_i a_i A_{cohi} - 9730.9(\chi^0) - 2161.73(\chi^1) + 138.3326$
Refractive Index (n)	$R^d = \sum_i a_i A_{Ri} + 0.0997(\chi^0) + 0.3081(\chi^1) + 0.1606$

## 5. Results and Discussion

The main idea behind the development of new, simple and accurate models is to predict polymer properties using methods, which are convenient for computer-aided polymer design algorithms. Table 2 gives the summary of the regression statistics in terms of number of data points used for each property model, the correlation coefficient, standard deviation, average absolute error, property function  $f(Y)$  and the units for each property.

The data points for  $V_g$ ,  $V_r$ ,  $V_c$ , and  $\delta$  in Table 2 are used from reference 1. Using the pseudo data from reference 2, for cohesive energy, the number of data points can be increased for solubility parameter model, but they are not stated in Table 2. While having  $V_a$  in Table 2, the models for  $V_g$  and  $V_r$  may not be necessary.

**Table 2 Correlation Statistics**

A. Marrero/ Gani based Group Contribution (GC) method correlation statistics							
	Properties						
	$V_g$	$V_r$	$V_a$	$V_c$	$T_g$	$\delta$	n
Data points	58	35	144	43	222	37	151
Correlation <sup>a</sup> , $R^2$	0.9995	0.9993	0.9991	0.9995	0.9374	0.9615	0.986
AAE <sup>b</sup>	0.665528	1.048	1.3821	0.565608	12.58	0.320405	0.00478
ARE <sup>c</sup>	0.530%	0.919%	1.1675%	1.11%	4.24%	1.66%	0.318%
f(Y)	$M_w/\rho_g - V_{g0}$	$M_w/\rho_r - V_{r0}$	$M_w/\rho_a - V_{a0}$	$M_w/\rho_c - V_{c0}$	$T_g - M_w - Y_{g0}$	$\Delta^2 V_a - E_{coh0}$	$R^d - R_0$
Units	cm <sup>3</sup> /mol	cm <sup>3</sup> /mol	cm <sup>3</sup> /mol	cm <sup>3</sup> /mol	K	J <sup>1/2</sup> /cm <sup>3/2</sup>	-
B. Connectivity Index (CI) method correlation statistics							
	Properties						
	$V_g$	$V_r$	$V_a$	$V_c$	$T_g$	$\delta$	n
Data points	58	35	144	43	222	37	151
Correlation <sup>a</sup> , $R^2$	0.9968	0.9988	0.9974	0.992	0.6685	0.7006	0.8488
AAE <sup>b</sup>	1.846876	1.534785	2.8187	2.964982	35.884	1.5973	0.01589
ARE <sup>c</sup>	1.45 %	1.53%	2.362%	4.56%	12.14%	8.474%	1.049%
f(Y)	$M_w/\rho_g - V_{g0}$	$M_w/\rho_r - V_{r0}$	$M_w/\rho_a - V_{a0}$	$M_w/\rho_c - V_{c0}$	$T_g - M_w - Y_{g0}$	$\delta^2 V_a - E_{coh0}$	$R^d - R_0$
Units	cm <sup>3</sup> /mol	cm <sup>3</sup> /mol	cm <sup>3</sup> /mol	cm <sup>3</sup> /mol	K	J <sup>1/2</sup> /cm <sup>3/2</sup>	-

<sup>a</sup>correlaton coefficient  $R^2 = 1 - (\sum (X_{\text{expt}} - X_{\text{est}})^2) / \sum (X_{\text{expt}} - \bar{X}_{\text{expt}})^2$

<sup>b</sup>  $AAE = 1/N \sum |X_{\text{est}} - X_{\text{expt}}|$

<sup>c</sup>  $ARE(\%) = 1/N \sum |(X_{\text{est}} - X_{\text{expt}}) / X_{\text{expt}}| \times 100$

N is the number of data points,  $X_{\text{est}}$  is the estimated property of X,  $X_{\text{expt}}$  is the experimental property value of X and  $\bar{X}_{\text{expt}}$  is the mean value of experimental property value of X.

<sup>d</sup> R is the molar refraction, having additive property, and refractive index is calculated from R as  $\sqrt{(2R+V)/(V-R)}$ .

### 5.1 Comparison of the GC model with Bicerano's and Van Krevelen's models

The predictions for the glass transition temperature and the solubility parameter have been compared with the group contribution method of Van Krevelen, as all the necessary information required for this comparison could be found only for these properties. Glass transition temperatures predicted by Van Krevelen are accurate to within 20 K for 80 % of around 600 polymers. The Marrero/Gani [3] GC-based method predicts glass transition temperatures with the accuracy of 82 % within 20K for a data set of 222 polymers. Regarding solubility parameter, the Van Krevelen method is in the order of 10 % accuracy (with respect to average relative error) while the present GC-based method given here is 1.66 % accurate. Comparison between the method of Bicerano [2], the new GC methods and the CI based method is given in Table 3.

Table 4 Comparison of accuracy between the Bicerano [2], GC and CI based methods

Property	Bicerano		GC		CI	
	SD	Correlation Coefficient, R <sup>2</sup>	SD	Correlation Coefficient, R <sup>2</sup>	SD	Correlation Coefficient, R <sup>2</sup>
$\rho_a$ (g/cm <sup>3</sup> )	0.0354	0.9872	0.02240	0.989	0.03648	0.9708
T <sub>g</sub> (K)	24.651	0.9749	19.52	0.948	45.9371	0.6685
E <sub>coh</sub> (J/mol)	1646.7	0.9988	1926.282	0.989	5758.332	0.9016
N	0.0157	0.9770	0.008146	0.986	0.02776	0.8488

Standard deviation =  $\sqrt{\sum (X_{est} - X_{expt})^2 / N}$ , where N is the number of data points, X<sub>est</sub> is the estimated property of X, X<sub>expt</sub> is the experimental property value of X

## 6. Conclusions

The predictions of the present models (GC/CI) look promising considering the standard deviation and average relative error. Additionally, the GC-CI model facilitates the use of property prediction models when a chemical structure cannot be represented by Marrero/Gani groups or certain groups representing the structure do not have the contribution to that particular property. Thus this integration of models enhances the wide usage of this property prediction model. But, as the data points were limited in number, more data points will be collected and these models will be validated for those data points in the future. As soon as the property models for other properties like permeability, dielectric constant are generated, polymer design software can be developed for finding polymer structures with desired properties. ‘OfficeGRID’ proved very efficient. But, at this point, the model development is not so time consuming. In time, these models will be extended for predicting polymer properties for polymer structure with several side chains (repeated at some frequency, etc.), where the computations could be massive and grid technology will be needed to solve these massive computations in a small time period.

## References

1. D.W. Van Krevelen: Properties of Polymers, Their Estimation and Correlation With Chemical Structure. Elsevier, 1990.
2. J. Bicerano: Prediction of Polymer Properties. Marcel Dekker Inc., 2002.
3. J. Marrero, R. Gani, Fluid Phase Equilibria, 183-184 (2001) 183-208.
4. R. Gani, P.M. Harper, M. Hostrup, Ind. Eng. Chem. Res., 44 (2001) 7262-7269.
5. B. Kier, H.L. Hall: Molecular connectivity in Structure Activity Analysis. John Wiley & Sons. New York, 1986.
6. <http://www.meshtechnologies.com/products.php?id=OfficeGRID>
7. [http://en.wikipedia.org/wiki/Grid\\_computing](http://en.wikipedia.org/wiki/Grid_computing)
8. R. Fletcher: Practical Methods of Optimization. John Wiley & Sons, 2000.