WEIGHTED TERNARY TREES AS DATA STRUCTURES FOR DENDRITIC MOLECULES

by

DANY I. DOUGHAN

Submitted in partial fulfillment of the requirements
for the Degree of Master of Science

Thesis Advisors: Dr. Ramzi Haraty
Dr. Ahmad Kabbani

Department of Computer Science
LEBANESE AMERICAN UNIVERSITY
December 2000

LEBANESE AMERICAN UNIVERSITY

GRADUATE STUDIES

We hereby approve the thesis of

Dany I. Doughan

Candidate for the Master of Science degree*.

[redacted]

Dr. Ramzi Haraty
Professor of Computer Science
Lebanese American University

[redacted]

Dr. Ahmad Kabbani
Professor of Chemistry
Lebanese American University

[redacted]

Dr. Issam Moghrabi
Professor of Computer Science
Beirut Arab University

Date: 1/1/2001

*We also certify that written approval has been obtained for any proprietary material contained therein.

WEIGHTED TERNARY TREES AS DATA STRUCTURES FOR DENDRITIC MOLECULES

ABSTRACT

by

DANY I. DOUGHAN

This thesis describes the construction of dendritic molecules using weighted ternary trees where each node is a record that represents the different physical, chemical, and quantum mechanical constants of a given atom and each weighted edge represents the number of bonds that exist between two adjacent atoms. This approach for representing such molecules offers flexibility as to reconfiguring the tree structure in case of addition of a new level or creation of a new generation in the dendrimer.

To my wonderful parents Chadia and Ibrahim and the love of my life Maya

ACKNOWLEDGMENTS

I would like to express my gratitude to Professor Ahmad Kabbani for providing me with exciting and stimulating projects.

My gratitude also goes to Professor Ramzi Haraty for his patience and help in refining this thesis.

My gratitude also goes to Professor Issam Moghrabi whose enthusiasm about computer science has launched my thesis at the Lebanese American University.

I would also like to thank Professors Ahmad Houri and Vatche Papazian who have always been a joy to talk to and bounce ideas around with.

I would like to thank my family, Chadia and Ibrahim, for their support and help all through my graduate career. I would also like to thank my best friend, Maya, for being such a great help whenever I needed her.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 2

CHAPTER 3

CHAPTER 1

INTRODUCTION

The representation of dendrimers is not a simple task and finding the right data structure for such complex entities demands a lot of care as to preserve the star-like nature of these molecules. Time consuming, implementation cumbersome, and slow running graphs are the data structures that are mostly used to represent chemical molecules including the dendritic ones [1], but weighted ternary trees offer a simple representation that is both convenient to work with and simple to trace.

Chapter 2 of this work discusses background issues related to weighted ternary trees and their structure from one side, and to dendrimers and their structure from another side. Chapter 3 discusses the nature of the data structure to be used. Chapter 4 analyzes the data structure. Chapter 5 provides the conclusions of this work.

CHAPTER 2

BACKGROUND

2.1 WEIGHTED TERNARY TREES

Weighted ternary trees are data structures that, like other trees, are made up of leaf and non-leaf nodes. In the pointer implementation, each non-leaf node is associated with, at most, three child pointers; on the other hand each leaf node is associated with three null pointers. Each edge in a weighted ternary tree is associated with a given value that provides it with a given weight that allows for future calculations on the edge itself. The weight of each edge is represented by a value that precedes the pointer representing the edge in each node (Figure 2.1).

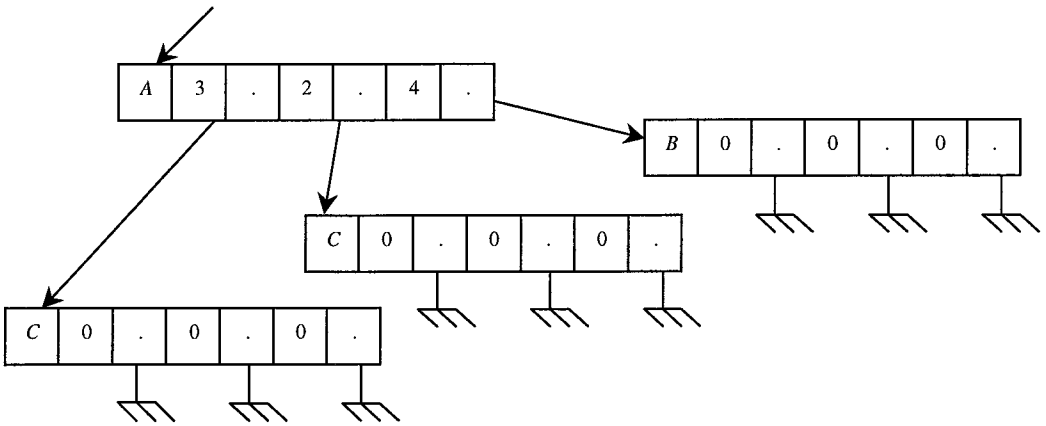

Figure 2.1 A typical weighted ternary tree

Since weighted ternary trees are not complete or essentially complete trees, then neither a specific order of nodes, a specific number of children, nor a conform tree depth are required. On the other hand, the two things that the tree has to preserve are the facts that each internal node cannot have more than three children and that each outgoing edge has to have a weight. Each internal node can have a left-child, a middle-child, and a right-child. A weighted ternary tree could take one of two forms by either having as root a dummy node or not having one at all.

2.2 DENDRIMERS

Dendritic macromolecules are hyperbranched polymers that emanate from a central core, have a defined number of generations and functional end groups, and are synthesized in a stepwise process by a repetitive reaction sequence.

The term dendrimer is given in reference to the branched structure adopted by these molecules and the name comes from the prefix dendri meaning treelike (from Greek dendron, tree) and the word polymer.

The first who reported a dendrimer synthesis was Fritz Vögtle in 1978, introducing the concept of cascade reaction by adding an amine to acrylonitrile. This permitted the attachment of the initial two arms or branches, subsequent reduction of the nitriles gave the desired diamine, which was subjected to the same synthetic sequence that yielded a second layer or generation (Jansen et al., 1995).

Research in dendrimers has flourished since the first major contributions of (Tomalia et al., 1990) and (Newkome et al., 1997, 1998). These molecule's topology results in a unique series of physical and chemical properties which, in turn, could be exploited in a number of diverse possible applications, such as nanoscale catalysis, immuno-diagnostic and NMR imaging (Genderen-van et al., 1994, 1995) and (Muijselaar et al., 1995) just to name a few. Dendrimers possess three distinguishing architectural components (Bosman et al., 1997a, 1998):

1- an initial core

2- interior layers (generations) composed of repeating units that are radially attached to the initiator core

3- exterior layer (terminal functionality) attached to the outermost interior generation.

All dendrimers have the following three properties (Bosman et al., 1997b):

1- regular size and shape

2- large number of readily accessible end groups, either nitrile or amine

3- possibility of end group modification in order to tailor properties as solubility, reactivity, toxicity, stability and glass transition temperature.

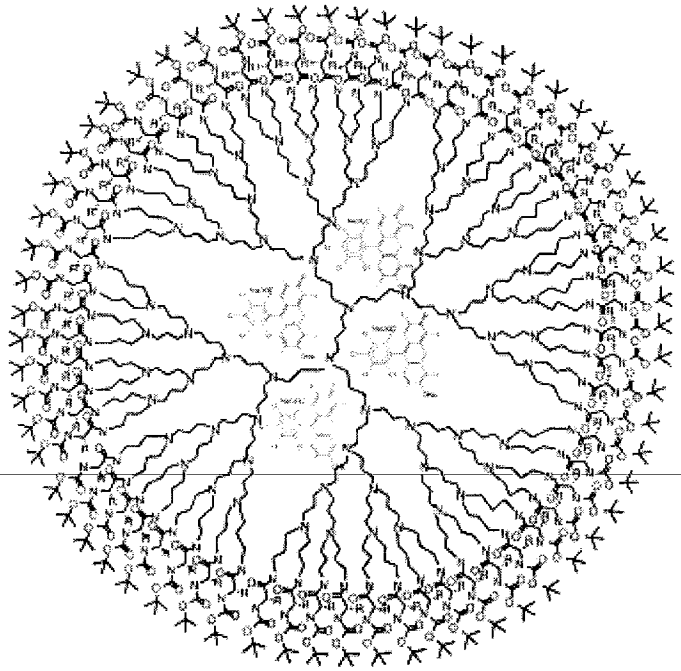

Figure 2.2 A typical dendritic molecule

Source: Meijer et al., Polym. Mater. Sci. Eng., 73, 123 (1995)

CHAPTER 3

THE DATA STRUCTURE

To represent a dendritic molecule, a rooted weighted ternary tree could be used. The choice of a ternary tree comes from the fact that a given atom could have in addition to its "parent" bond at most three other "children" bonds. The root is a node that contains no data but a set of pointers to two of the initial core starting atoms that generate the molecule. The root has two children, either the central atoms in case the core chain is made up of an even number of atoms, or the central atom and a dummy node, which connects via an edge of weight zero to one of the central atom's adjacent atoms in the chain in case the core chain is made up of an odd number of atoms.

Each node is a record that holds the different physical, chemical, and microscopic values of a specific atom. A node also holds weights of integer type corresponding to the different children bonds where the integer value corresponds to the number of inter-atomic bonds that exist between the node

and a specific child. A value of "1" for example means that the parent node is attached to the child by a single bond, whereas a "2" means that there is a double bond in between them. Thus, the weights of edges of internal nodes can have values ranging from zero up to three. For each internal node, the total weight of all child edges should be equal to three except for the leaf nodes where it is zero, and for the root where it might range from zero (empty tree) to six (triply bonded starting pair of atoms).

Dynamic allocation and binding implement the data structure to allow for the continuous growth and expansion of dendritic molecules and the possible emergence of new generations (Figure 3.1).

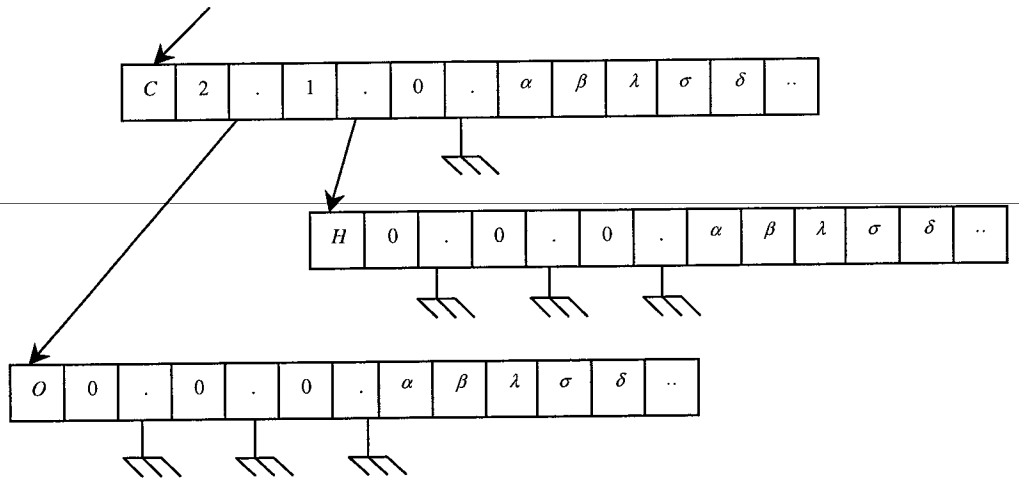

Figure 3.1 A weighted ternary tree representing an external radical • COH of a dendritic molecule

Figure 3.1 shows the •COH external radical of a given dendritic molecule where the C=O double bond and the C-H single bond are represented respectively by the integer value "2" of the weight on the edge going from C to O and the integer value "1" of the weight on the edge going from C to H.

The fields titled α, β, λ, σ, δ, ... represent the different physical, chemical, and microscopic values that each atom or pair of atoms should hold. For example α, β, and λ might represent the bond length between the parent atom and the left-child, middle-child, and right-child respectively.

For each record or atom, the δ field might hold the estimated ^{1}HMR shift (Figure 3.2) and the ^{13}CMR shift (Figure 3.3) values of that given atom.

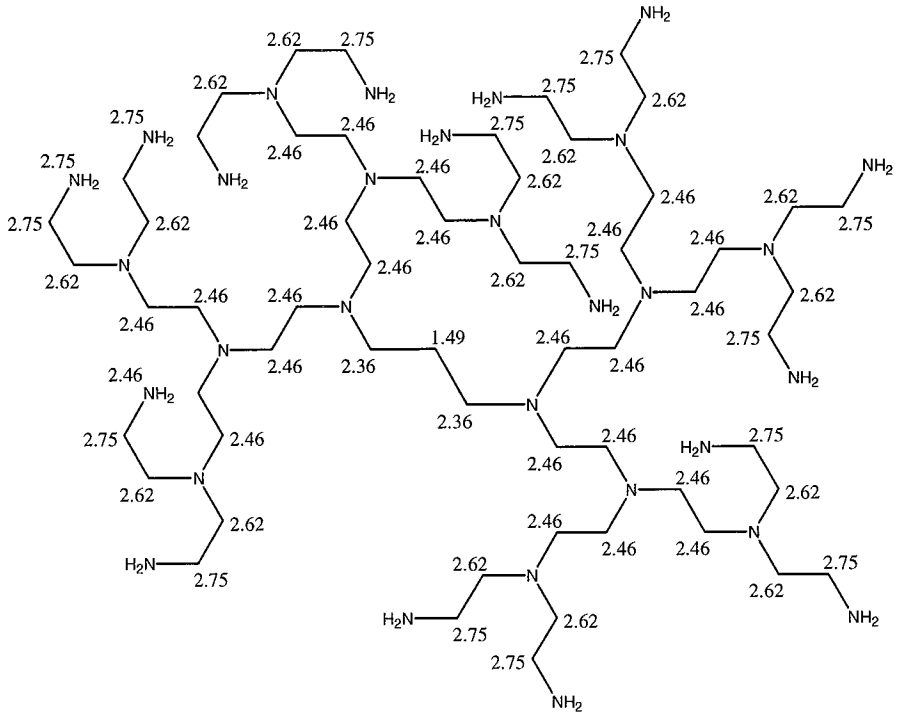

Figure 3.2 The estimated ^{1}HMR shifts for the dendritic molecule in Figure 2.2 at the third generation level

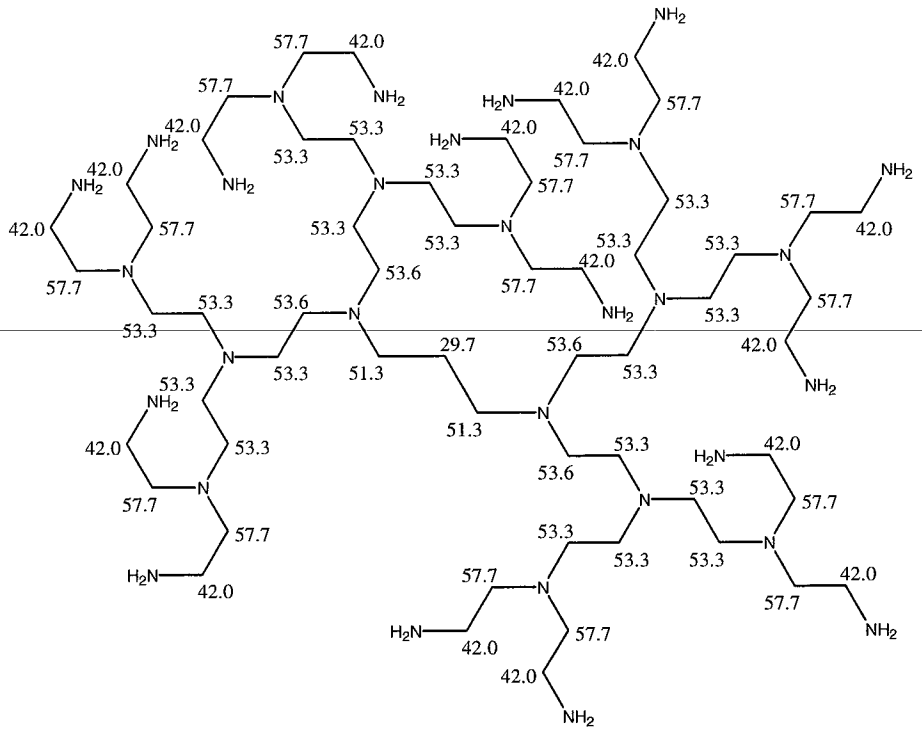

Figure 3.3 The estimated ^{13}CMR shifts for the dendritic molecule in Figure 2.2 at the third generation level

CHAPTER 4

ANALYSIS

Graphs are the data structures that are used to model chemical molecules, but as it is known these structures are very irregular, and hard to traverse and analyze. It is also almost impossible to relate the size of a graph to the number of nodes (or atoms) of which it is made up.

Trees, on the other hand, can relate the size to the number of nodes. The maximum number of nodes, in a ternary tree with depth h, can be given by $\Sigma^{h}_{k=0} 3^{k}$. By mathematical induction, this can be proved to be equal to $(3^{h+1} - 1) / 2$.

Proof 4.1:

The basis step:

A ternary tree of depth $h = 0$ consists of just a single node. And indeed $(3^{0+1} - 1) / 2 = 1$.

The induction:

Assume that the number of nodes of a ternary tree of depth n is $(3^{h+1} - 1) / 2$. The number of leaf nodes is 3^h, and each of these leaves grows 3 new leaves to produce the ternary tree of depth $h + 1$. Thus, the number of leaves of the tree of depth $h + 1$ is given by:

$$\sum_{k=0}^{h+1} 3^k = \sum_{k=0}^{h} 3^k + 3^{h+1}$$

Using the induction hypothesis, this would be equal to:

$$(3^{h+1} - 1) / 2 + 3^{h+1} = (3^{h+1} - 1) / 2 + 2*3^{h+1} / 2 = (3^{h+2} - 1) / 2$$

Therefore: $\Sigma^{h}_{k=0} 3^k = (3^{h+1} - 1) / 2.$

This result relating the depth of the tree to the maximum number of nodes (or atoms in this case) does not provide a good estimate of the number of atoms in a dendritic molecule as compared to its height. Assume the dendritic molecule of Figure 2.2 at the third generation level, the tree would have a depth of (2 + 3 * 4) = 14, and by using the result of proof 4.1 where

$\Sigma^{h}_{k=0} 3^{k} = (3^{h+1} - 1) / 2$ the tree would have at most 7174453 nodes while it actually only has 239. This discrepancy in numbers comes from the fact that not all atoms in the dendrimer branch, only Nitrogen atoms do so (de Brabander-van et al., 1993). That is why one cannot predict the number of atoms in a tree modeled dendritic molecule by knowing the depth of the tree.

The weighted ternary tree representation of dendritic molecules offers the following advantages over the traditional graph representation:

- Being rooted, trees can point very easily to the initial core (the atom or atoms that generate the molecule), unlike graphs that often do not have any one vertex singled out as special (the root).
- Unlike graphs, trees do not require adjacency information since the structure itself of trees takes care of these details.
- Assigning the different ^{1}HMR shift values to the different Hydrogen atoms is real easy since it is enough to do the calculations and assign the values to only one branch. This is allowed chemically because we only need to look at the adjacent Carbon's connected Hydrogen atoms (the parent and the child of the node). Then we simply traverse the tree and copy the shift

value at each depth to all Hydrogen atoms at the same depth in all other branches (Figure 3.2). This is feasible because a dendritic molecule is spherically symmetrical and the way we build the weighted ternary tree preserves this nature. For example all Hydrogen atoms at depth 11 have a chemical shift δ value of 2.75 ppm (parts per million).

- The procedure that assigns the different ^{1}HMR shift values to the different Hydrogen atoms can be used to allocate the different ^{13}CMR shift values to the different Carbon atoms (Figure 3.3). For example all Carbon atoms at depth 11 have a chemical shift δ value of 42.0 ppm.
- Adding a copy of a tree's sub-section is also a very simple task using the pointer based implementation. All one needs to do is to find the start of the sub-section, traverse it and copy it into a temporary sub-tree, search in the tree for the node onto which the sub-section is to be copied, and then add the temporary sub-tree to the tree at the desired node (applied at the addition of new generations that are structurally identical).
- Deleting a tree's sub-section is an even simpler task using the pointer based implementation since all one needs to do is to find

the start of the sub-section and dispose of it (applied when functional groups get replaced by new ones or by new generations).

- Graphs can be traversed in a time that is in the order of the number of vertices n and the number of edges e ($T_g(n) \in O(n + e)$) by either choosing a depth-first approach or a breadth-first approach. Trees, on the other hand, can be traversed in a time that is in the exact order of the number of nodes n ($T_t(n) \in \Theta(n)$), thus trees can access and process data and information stored in them in a shorter time.

A simple proof showing that $T_t(n) \in \Theta(n)$ is given in Proof 4.2.

Proof 4.2:

Suppose that visiting a node takes a time in O(1), that is the time required is bounded above by some constant c. Without loss of generality we may suppose that $c \geq T(0)$. Suppose further that we are to explore a tree containing n nodes, $n > 0$;

one of these nodes is the root, so if g of them lie in the left sub-tree, then there are $n - g - 1$ in the right sub-tree. Then

$$T(n) \leq \max (T(g) + T(n - g - 1) + c), n > 0$$

This is true whatever the order in which the left and right sub-trees and the root are explored. We prove by constructive induction that $T(n) \leq an = b$, where a and b are appropriate constants, as yet unknown. If we choose $b \geq c$ the hypothesis is true for $n = 0$, because $c \geq T(0)$. For the induction step, let $n > 0$ and suppose the hypothesis is true for all m, $0 \leq m \leq n - 1$. Then

$$\begin{aligned} T(n) &\leq \max (T(g) + T(n - g - 1) + c) \\ &\leq \max (ag + b + a(n - g - 1) + b + c) \\ &\leq an + 3b - a \end{aligned}$$

Hence, provided we choose $a \geq 2b$ we have $T(n) \leq an + b$, so the hypotesis is also true for $m = n$. This proves that $T(n) \leq an + b$ for every $n \geq 0$, and therefore $T(n) \in O(n)$.

On the other hand it is clear that $T(n) \in \Omega(n)$ since each of the n nodes is visited.

Since $T(n) \in O(n)$ and $T(n) \in \Omega(n)$ therefore $T(n) \in \Theta(n)$.

The major disandvantage that trees have when compared to graphs is the fact that the latter can support cyclic compounds and rings which the former by definition cannot. This does not cause serious problems because spanning trees could be generated from these graphs or sub-graphs as described by Nikolic et al.1990 and incorporated into the main dendritic molecule tree structure.

CHAPTER 5

CONCLUSIONS

This thesis describes an alternative way to construct dendritic molecules using weighted ternary trees rather than graphs. Each node in the tree is a record that represents the different variables and constants needed to bring the molecule to life. Each edge has a weight that represents the number of bonds that exist between two adjacent atoms. Trees are convenient to work with when it comes to reconfiguring the dendritic molecule by removing atoms or functional groups from it, or adding ones to it, or simply doing changes to it. Weighted ternary trees are also very convenient for locating and working with the molecule's initial core and exterior layer being respectively the top of the tree (the root area) and the bottom of the tree (the leaves area). Finally, trees access and process data stored in them in a time that is in the exact order of the number of nodes n $\Theta(n)$, not in the order of the number of vertices n and edges e $O(n + e)$, which is the case of graphs.

REFERENCES

Bosman A.W., Janssen R.A.J., Meijer E.W., 1997a. Five generations of Nitroxyl-Functionalized Dendrimers. *Macromolecules, 30, 3606-11.*

Bosman A.W., Schenning A.P.H.J., Janssen R.A.J., Meijer E.W., 1997b. Probing the Structure and Dynamics of End-Group Functionalized Poly(propylene imine) Dendrimers. *ACS PMSE Prepr., 77, 80-81.*

Bosman A.W., Bruining M.J., Kooijman H., Spek A.L., Janssen R.A.J., Meijer E.W., 1998. Concerning the localization of end groups in dendrimers. *J. Am. Chem. Soc., 120 (33), 8547-8.*

de Brabander-van den Berg E.M.M., Meijer E.W., 1993. Poly(propylene imine) Dendrimers: Large Scale Synthesis by Heterogeneously Catalysed Hydrogenations. *Angewandte Chemie, International Edition in English, 32(9), 1308-11.*

Genderen-van M.H.P, Baars M.W.P.L., van Hest J.C.M., de Brabander-van den Berg E.M.M., Meijer E.W., 1994. Observing individual generations in poly(propyleneimine) dendrimers with natural abundance ^{15}N NMR spectroscopy. *Recl. Trav. Chim. Pays-Bas, 113, 573-4.*

Genderen-van M.H.P., Baars M.W.P.L., Elissen-Román C., de Brabander-van den Berg E.M.M., Meijer E.W., 1995. Natural Abundance ^{15}N NMR Spectroscopic Investigations of Poly(propylene imine) Dendrimers. *ACS PMSE Prepr., 73, 336-7.*

Jansen J.F.G.A., de Brabander-van den Berg E.M.M, Meijer E.W., 1995. The Dendritic Box and Bengal Rose. *ACS PMSE Prepr., 73, 123-4.*

Muijselaar P.G.H.M., Claessens H.A., Cramers C.A., Jansen J.F.G.A., Meijer E.W., de Brabander-van den Berg E.M.M., van der Wal S., 1995. Dendrimers as a Pseudo-Stationary Phases in Electrokinetic Chromatography. *J. High. Resolut. Chromatogr., 18, 121-3.*

Nikolic S., Trinajstic N., Knop J.V., Muller W.R., Szymanski K., 1990. On the Concept of the Weighted Spanning Tree of Dualist. *Journal of Math. Chemistry, 4, 357-310.*

Newkome G.R., Weis C.D., Moorefield C.N., 1997. A useful dendritic building block: di-tert-butyl 4-{(2-tert-butoxycarbonyl)ethyl}-4-isocyanato-1,7-heptadicarboxylate. *Tetrahedron Lett., 38(40), 7053-6.*

Newkome G.R., Weis C.D., Moorefield C.N., Baker G.R., Childs B.J., Epperson J., 1998. Isocyanate-based dendritic building blocks: combinatorial tier construction and macromolecular property modification. *Angewandte Chemie, International Edition in English, 37(3), 307-310.*

Tomalia D.A., Naylor A.M., Goddard III W.A., 1990. Starburst dendrimers: molecular-level control of size, shape, surface chemistry, topology, and flexibility from atoms to macroscopic matter. *Angewandte Chemie, International Edition in English, 29, 138-175.*