

A COMPUTER ASSISTED STRUCTURAL ELUCIDATION FOR COAL MACROMOLECULES.

Jean-Loup Faulon, Patrick G. Hatcher,
and Kurt A. Wenzel.

Fuel Science Program,
The Pennsylvania State University,
University Park, PA 16802.

Keywords : coal structure, molecular modeling, computer aided structure elucidation.

INTRODUCTION

Study of the chemical structure of coal has been the subject of fundamental research for decades and we are now beginning to better understand this structure. It is well recognized that coal is chemically and physically a highly heterogeneous material mainly consisting of organic matter (macerals) and some inorganic materials (minerals). Organic matter which is the object of our interest is composed of carbon, hydrogen and oxygen with lesser amounts of sulfur and nitrogen. The amount, the distribution, and the chemical structure of various macerals in coals depend on the chemical nature of the original coal forming material and the conditions of coalification. Individual macerals may themselves be heterogeneous assemblages of sub-macerals with different chemical structural compositions. The nature of each phase progressively changes during coalification (with rank). For example it is well known that the carbon content of coal increases with its degree of maturation (1). In short, coals are very diverse, and even within any given coal type there is considerable heterogeneity.

Nonetheless, because of the strong link between structure and reactivity, there have been many attempts to model the molecular structure of various coals. For example, chemical structural models have been developed for vitrinite, a major component of coal, ranging in rank from lignite to subbituminous coal (2). For bituminous coals, the most widely accepted molecular models developed during the past 30 years have been the aromatic/hydroaromatic structures. These models are formed by fragments containing about three linked aromatic and/or hydroaromatic rings including appropriate numbers and types of heteroatoms. The fragments are interconnected by hydroaromatic, etheric or aliphatic linkages (3-6). These models have been built using available chemical and structural data on coal, without the aid of computers. They are certainly not unique and the same analytical data can produce similar but different models. Because of the highly heterogeneous character of coal these models are only intended to be representative structures, or average structures. The main limitation of these molecular structures are their two dimensional nature which necessarily ignores most of the interfragment interactions and three dimensional properties such as density and porosity.

Recently Spiro (7) and Carlson (8) have constructed 3D models of several of these structures. Using a space-filling model, Spiro identified several steric difficulties in the original structures and proposed some modifications. To determine the optimal structural conformations of the previous models (3-6), Carlson has used molecular mechanic and dynamic programs. Once minimized, some characteristics such as energy, density porosity can be systematically evaluated

for each model (8). An important result is that for all the models no porosity volumes were found since bituminous coals do clearly show substantial porosity.

The result of these studies have led us to reconsider the modelization of coal. Considering the complexity and the heterogeneity of this material, our intention is not to define another new general model but more specifically to establish a set of average 3D structures which can be correlated to different macerals at different rank levels. In this paper we have chosen to modelize only one maceral of high volatile bituminous coal : the vitrinite. The sample is specifically a coalified wood sample described by Hatcher et al. (9). Our method will involve use of chemistry and a computer. By chemistry we mean we will consider all the chemical analytical results already employed in the previous models but also new techniques which were not used in the past. By computer, we will see that it is possible to generate automatically 3D models from the analytical results and characterize these models in terms of statistics.

METHOD

From a chemical point of view, the task of determining the molecular structure is becoming simpler with the advent of many spectroscopic methods such as NMR, FTIR and flash pyrolysis/gas chromatography/mass spectrometry (py/gc/ms). ^{13}C NMR with cross polarization-magic angle spinning and dipolar dephasing can be used to analyze the overall distribution of various types of carbon structures present in coal such as aliphatic, oxygen-containing, aromatic and hydroaromatic structures (2). By using dipolar dephasing techniques, we can also identify protonated and non-protonated carbons (Table I). Information on the nature of pyrolysis products produced upon heating a coal particle can be obtained using py/gc/ms (2). Results of elemental analysis, and NMR are directly conducted on the coal macromolecule (9) and we can then define what we call a signature for this macromolecule. The signature is simply a series of numbers which count different types of atoms in a specific environment (Table II). The information obtained by py/gc/ms does not describe directly the macromolecule but degradation products or fragments of this macromolecule (Fig. 1). The information given by py/gc/ms is qualitative and defines the fragments which once connected together, in an appropriate amount, constitute the macromolecule. The problem is therefore to find the correct quantity of each fragment and each connection between fragments (interfragment bonds) to form a macromolecule which is consistent with all the structural information. Retrieval of molecular structure from a set of analytical results can be accomplished by an empirical method which consists of a repetitive trial-and-error process to find the correct structure. This method was probably used by the previous authors (3-6), but it is not an entirely satisfactory technique for at least two reasons :

- 1- The process to build a structure is empirical, is accomplished through manual fitting, and can be time consuming for large molecules.
- 2- Generally, many structures can be built from the same analytical data, the reason one structure is chosen instead of another cannot be clearly defined. This is probably one of the reasons why the coal models cited previously are so different.

Many studies have been conducted in the 30 past years to resolve by computer the general problem of retrieving a structure from analytical data. The generic name for these studies is *computer-aided structure elucidation* (10). Unfortunately the techniques used in these studies are useful only for small molecules and, therefore, cannot be helpful for coal. Recently Faulon et al. (11) have developed a technique which is efficient for macromolecules such as coal. More precisely the data used by this technique are the quantitative chemical information regarding the macromolecule (a signature) and qualitative information concerning a set of fragments. The authors have shown that it is possible from a set of chemical data to compute the amount of fragments and the connections between fragments (interfragment bond) by resolving a linear equation (signature equation). Briefly, the quantitative data given by elemental analysis and NMR defines the signature of the macromolecule, the qualitative information defined by py/gc/ms identifies the molecular

fragments. The structure of each fragment is known from mass spectrometry, therefore it is possible to calculate a signature for each of them. It is also possible to determine a signature for each possible interfragment bond. The amount of fragment and interfragment bonds is obtained by solving the signature equation :

$$\text{signature of fragments} + \text{signature of interfragment bonds} = \text{macromolecular signature}$$

This equation is defined in a discrete space and can admit several solutions. Once resolved, the equation defines the pieces (fragment) and the relations between the pieces (interfragment bonds) of the unknown model which is at that point still an unresolved "puzzle". To "reconstruct the puzzle", a combinatorial program has been developed (12). The program computes first an estimation of the number of solutions : how many models can be built considering a set of pieces and the relation between pieces. When this number is reasonable, the program calculates automatically all the models which correspond to the analytical results. Otherwise the program builds randomly one or a fixed quantity of different models. In all cases the solutions are 3D structures coherent from a chemical point of view : the bonds lengths and the angles between bonds are consistent with the values defined in the literature. These structures can be read and displayed by molecular modeling softwares such as PCMODEL (Serena software), DISCOVER (Biosym) and can be submitted to energy minimization using programs of molecular mechanics like for example MMX (13). Once minimized, physical properties, such as energy, density, porosity can be evaluated.

RESULTS

The analytical results introduced in the program are described in the paper presented by Hatcher et al.(9). From a practical point of view the atomic information (elemental analysis and NMR data listed in Tables I and II) was stored in a file. The molecular fragments (Fig. 1) and the interfragments bonds (Fig. 2) were built using the molecular modeling software PCMODEL (Serena Software) and stored in a library.

The signature equation (11) has been applied for the previous information and all the molecular structures containing a number of carbon atoms between 140 and 180 has been searched. The signature equation found 7 solutions listed in Table III. The best structure with the least error is structure number 6 (molecular formula $C_{178}H_{164}O_8$). This structure is composed of the following fragments: 2 toluenes, 3 C2 benzenes, 1 C3 benzene, 1 o-cresol, 3 m- and p-cresols, 3 C2 phenols, 1 C4 phenol, 3 alkylnaphthalenes, 3 propyls, 9 ethyls, and 7 methyls. The interfragment bonds are 6 benzene-benzenes, 8 benzene-phenols, 10 benzene-toluenes, 5 toluene-toluenes, 7 ethylbenzene-ethylbenzenes, 3 propyl-benzenes, 9 ethyl-toluenes, and 7 methyl-ethylbenzenes. The fragments and inter-fragments bonds are the structures listed in Fig. 1 and 2. Benzene and phenol which appear in the py/gc/ms results are not present in the solution but they can be obtained by pyrolysis from any alkyl benzene or alkyl phenol. C3 phenol is also not present in the solutions. This compound is formed by connecting cresol to ethyl or ethylphenol to methyl. In the same way, alkyldibenzofuran is formed by bonding alkyl benzene and alkyl phenol with a benzene-phenol bond and a benzene-benzene bond.

The combinatorial program presented in (12) has been applied to determine how many models can be built for the structure number 6. This structure is composed of 36 fragments. Considering the structure of the fragments and the global percentage of hydrogen, each fragment must be attached to the rest of the molecule by an average of 3.05 bonds. The total number of bonding sites is therefore $36 \times 3.05 = 110$; this number is also equal to the number of interfragment bonds (55) multiplied by 2. The first bonding site cannot be connected to itself, and there is $110 - 1$ or 109 possible connections. After this operation, $110 - 2$ or 108 bonding sites remain. For the same reason, there is $108 - 1$ or 107 possibilities to connect the second bonding site, 105 for the third, 103 for the fourth, and so on. Consequently the maximum number of structures that can be built is $109 \times 107 \times 105 \times 103 \dots \times 1$: this number is greater than 10^{86} ! However most of the

structures are identical and most of them do not respect the classical chemical constraints and the quantity of interfragment bonds calculated by the signature equation. By avoiding the redundancies and respecting all the analytical constraints, the program found around 1,000,000 possible structures. All these structures are composed of the same fragments and inter-fragments bonds but are different. Fig. 3 shows one of these structures randomly generated by the program. This structure has been subjected to an energy minimization to find a conformationally correct form (MMX program 13). Fig. 4 is the 3D representation of the structure after minimization. All the fragments analyzed by py/gc/ms can be retrieved in this solution, especially C3 benzene and alkyldibenzofuran which were not present in the initial data but obtained by combination of other fragments.

CONCLUSION

The conclusion which can be made from the above results is that for coal the number of models is too large to build all of them. Therefore, in such cases the following question has to be considered : *how many and which structures must be selected and built ?*

The theory of sampling assumes that when the size of a finite population is known, it is possible to extract a subset (a sample) of this population which, in terms of probability, is a good representation. Many techniques of sample design are possible : random sampling, stratified sampling, and sampling with unequal probability of selection (14). The most simple technique which can be applied as a first approximation for our problem is the simple random sampling without replacement (SRSWOR). With this technique it is possible to define an optimal size for a sample (Mean Square Error Method : 15) and extrapolate the mean sample value of a certain characteristic to the whole population. Clearly, according to the theory of sampling, it will be possible to define a sample of average structures which represents the studied coal. It will be also possible to evaluate on this sample certain characteristic such as the energy, the density, the porosity, and to extrapolate these characteristics to the whole population of coal models.

The computer program developed for this study is built in such a way that any new experimental result can be introduced, at any time, without modification of the program. The same program will be useful in the future, when from new experimental data, new models will be needed. Furthermore, we can point out that the program can be applied for any other bituminous coal maceral and any other molecule studied in coal science such as lignite, subbituminous coal, or anthracite. Because different ranks of coal can be modeled, the program can be helpful in understanding the process of coalification. In fact the method is based on a original study made for the kerogen macromolecule (11), and is a general system of structure elucidation efficient for any unknown molecular structure. Therefore, it is realistic to consider applications in organic geochemistry, fuel science and petroleum science such as modelization of humic acids, peat, asphaltenes, crude oil, jet fuel, etc.

REFERENCES

1. Van Krevelen, D. W., *Coal*, Elsevier, Amsterdam, 1961.
2. Hatcher, P. G., *Advances in Org. Geochem.*, **16**, 959, 1990.
3. Given, P. H., *Fuel*, **39**,147, 1960.
4. Wiser, W. H., *Proc. of the Electric Power Research Institute Conf. on Coal Catalysis*, 1973.
5. Solomon, P.R., in *New Approaches in Coal Chemistry*, Am. Chem. Soc. Symp. Series No. **169**, 61, 1981.
6. Shinn, J. H., *Fuel*, **63**, 1187, 1981.
7. Spiro, C. L. *Fuel*, **60**, 1121, 1981.
8. Carlson, G. A., Granoff B., in *Coal Science II*, Advances in Coal Sciences series, 159 1990.
9. Hatcher, P. G., Faulon, J. L., Wenzel, K. A., Cody, G. D., *this volume*, 1992.

10. Smith, D. H., *Computer-Assisted Structure Elucidation*, Am. Chem. Soc. Symp. Series No. 54, 1977.
11. Faulon, J. L., Vandenbroucke, M., Drappier, J. M., Behar, F., and Romero, M., *Advances in Org. Geochem.*, **16**, 4-6, 1990.
12. Faulon, J. L., *Prediction Elucidation and molecular modeling : Algorithm and Application in Organic Geochemistry*, PhD Thesis, Ecole des Mines, Paris, 1991.
13. Alinger, N. L., *J. Am. Chem. Soc.*, **99**, 8127, 1977.
14. Dalenius, T., in *handbook of statistic 6 - Sampling*, Elsevier, Amsterdam, 1988.
15. Rao, J. N. K., in *handbook of statistic 6 - Sampling*, Elsevier, Amsterdam, 1988.

ACKNOWLEDGMENTS

Financial support for this study was provided by the U. S. Department of Energy, Sandia National Laboratories under contracts DE-AC04-76DP00789 and 12-5543. We also acknowledge financial support from DOE contract DE-AC22-91PC91042 from the Pittsburgh Energy Technology Center. We thank Jackie M. Bortiatynski for assistance in assembling the data and manuscript.

Table I. The atomic information. These values are taken from Hatcher et al.(9).

ELEMENTAL ANALYSIS			¹³ C NMR			
	WEIGHT %	for 100 C		Value 1	Value 2	AVERAGE
C	85.8	100	fa	0.64	0.61	0.63
H	6.5	91.2 ± 1	Aro-O / Aro	0.11	0.13	0.12
O	5.9	5.2 ± 1	Aro-H / Aro	0.44		0.44
			CH3 / Ali	0.33	0.28	0.30

Table II. The macromolecular signature. The notations used by the signature are the potential types defines in Biosym softwares : h_ represents an hydrogen atom, c_ an aliphatic carbon, cp an aromatic carbon, and o_ an oxygen atom. The values are given for 100 carbon atoms, h_ = 91.2 means that the ratio H/C is equal to 0.912, c_(c_h_h_h_) = 9.2 means that there is 9.2% of carbon which is aliphatic, linked to one other aliphatic carbon and three hydrogen atoms.

SIGNATURE

	MIN	MAX	AVERAGE
h_	90.2	92.2	91.2
o_	4.2	6.2	5.2
c_	34.2	39.8	37.0
cp	60.2	67.8	64.0
c_(c_h_h_h_)	11.1	13.3	12.2
cp(cp cp h_)	25.3	29.9	27.6
cp(cp cp o_)	6.6	8.9	7.7

Table III. The solutions of the signature equation. The error listed is the difference between the model and the analytical results

solution number	molecular formula	error for 100 C
1	C ₁₅₂ H ₁₄₀ O ₇	0.92
2	C ₁₆₁ H ₁₄₈ O ₈	0.91
3	C ₁₆₂ H ₁₄₈ O ₈	0.83
4	C ₁₇₀ H ₁₅₆ O ₉	0.94
5	C ₁₇₁ H ₁₅₆ O ₉	0.86
6	C ₁₇₈ H ₁₆₄ O ₈	0.82 (-)
7	C ₁₈₀ H ₁₆₄ O ₁₀	0.97

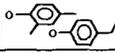
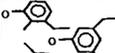
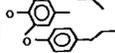
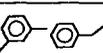
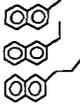
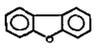
NAME	WEIGHT %	CHOSEN STRUCTURE	NAME	WEIGHT %	CHOSEN STRUCTURE
methyl ethyl propyl			C2 phenol	23.0	
benzene	3.3		C3 phenol	11.0	
toluene	4.8		C4 phenol	4.8	
C2 benzene	5.2		alkyl naphthalene	17.0	
C3 benzene	3.9				
phenol	3.6				
o cresol	5.7				
m and p cresol	13.0		alkyl dibenzo furans	5.2	

Fig. 1. The molecular information. These fragments are results of flash pyrolysis/gas chromatography/mass spectrometry. The weight percentages and the different isomers for the chosen structures are discussed in Haicher et al.(9). The fragments are considered by the program only as qualitative information. Once connected together in an appropriate amount these fragments form the unknown macromolecule.

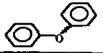
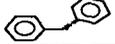
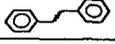
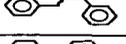
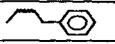
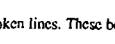
BOND	STRUCTURE
benzene-benzene	
benzene-phenol	
benzene-toluene	
toluene-toluene	
ethylbenzene-ethylbenzene	
propyl-benzene	
ethyl-toluene	
methyl-ethylbenzene	

Fig. 2. The inter-fragment bonds are shown as broken lines. These bonds represent the different ways to connect the fragments listed in Fig. 2.

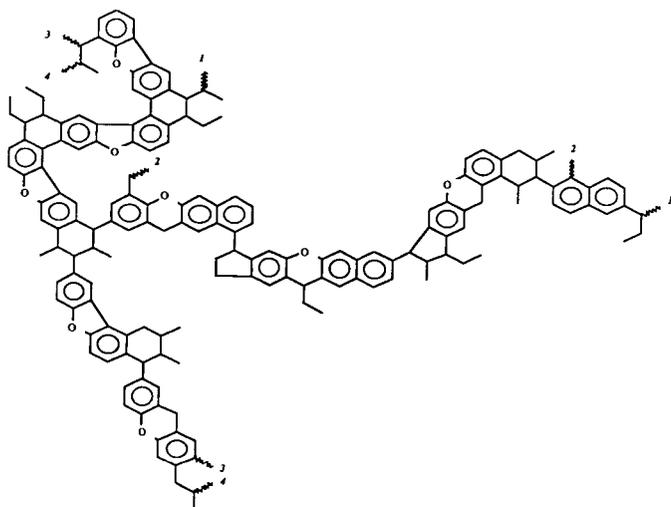


Fig. 3. One possible 2D representation for vitrinite from high volatile bituminous coal. Some other models are possible. An other possibility is given in Hatcher et al. (9). The number indicates the bonding sites (e.g. 1 -> 1).

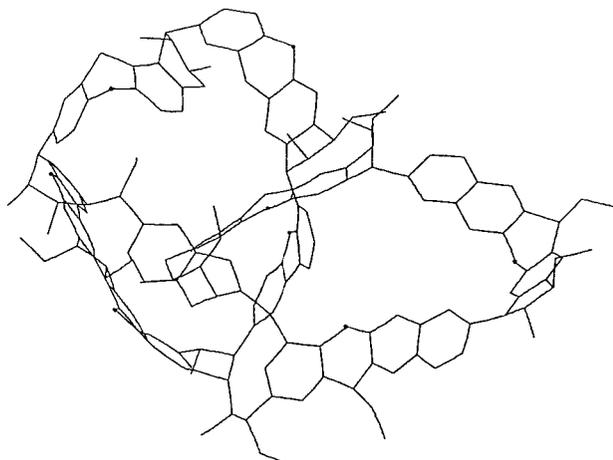


Fig. 4. One possible 3D structure for vitrinite from high volatile bituminous coal. This structure is the 3D representation of Fig. 3.