INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI®

University of Alberta

Molecular Modeling of Heavy Oil

by

Jeff M. Sheremata

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Master of Science

in

Chemical Engineering

Department of Chemical and Materials Engineering

Edmonton, Alberta

Spring 2001

0-612-60497-7

University of Alberta

Library Release Form

Name of Author: Jeff Sheremata

Title of Thesis: Molecular Modeling of Heavy Oil

Degree: Master of Science

Year this Degree Granted: 2001

Permission is hereby granted to the University of Alberta Library to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves all other publication and other rights in association with the copyright in the thesis, and except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatever without the author's prior written permission.

Jeff Sheremata

9 Heritage Blvd, St. Albert
Alberta, Canada, T8N-5P8
December 21, 2000

December 21, 2000

Fall 2000
University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled Molecular Modeling of Heavy Oil submitted by Jeff M. Sheremata in partial fulfillment of the requirements for the degree of Master of Science in Chemical Engineering.

W.C. McCaffrey

M.R. Gray

P. Fedorak

P.M. Fedorak

December 20, 2000

Abstract

Molecular kinetic modeling requires representations of the molecules in heavy oil feeds. By proposing a reaction network, the product composition can be predicted based upon feed representations. Previous molecular representations have consisted of molecules constructed from a limited number of chemical attributes including benzene, cyclohexane, and paraffins. Most previous representations have been overly simplified and unrealistic. The approach used in this study is to represent heavy oil fractions using a limited number of computer-generated molecules. This approach has been successfully applied to various hydrocracked Athabasca bitumen fractions. Molecular representations were generated using data from vapor pressure osmometry, nuclear magnetic resonance, and elemental analysis. The molecules in these new representations are composed of nineteen fundamental different chemical attributes that contain carbon, hydrogen, oxygen, sulphur, and nitrogen. Some chemical attributes including indoles and sulphoxides were used. Molecular representations from this study will assist the development of predictive heavy oil hydrocracking kinetic models.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

Abbreviation	Definition
NMR	Nuclear Magnetic Resonance Spectroscopy
^{1}H-NMR	Proton Nuclear Magnetic Resonance Spectroscopy
^{13}C-NMR	^{13}C Nuclear Magnetic Resonance Spectroscopy
SARA	Saturates, Aromatics, Resins, and Asphaltenes, Chromatography
US	Unit Sheets (sheets of aromatic rings)
ppm	Parts Per Million
RICO	Ruthenium Ion Catalyzed Oxidation
GC	Gas Chromatography
VPO	Vapor Pressure Osmometry
IBP	Initial Boiling Point
k	First Order Rate Constant
s	Stoichiometric Coefficient
CSTR	Continuous Stirred Tank Reactor
pdf	Probability Distribution Function

Chapter 1

Introduction

A very promising type of heavy oil hydroprocessing kinetic analysis is molecular based kinetics. Molecular kinetic models are based on molecular representations. Molecular representations are sets of molecules that are consistent with all available analytical data for a complex mixture. The tremendous chemical complexity of heavy oils makes molecular representations themselves models. Molecular based kinetic models contain detailed information about how different chemical components react. Such a model defines the type of chemical components that react by a given reaction, the products of each reaction, and the rate of each reaction. Once a molecular representation is made, a molecular based kinetic model reacts the molecules in the representation according to its details and rules. Since molecular based molecular kinetic models are based on actual chemistry they are capable of describing the reactivities associated with different feeds and catalysts.

Data from various analytical techniques are necessary to create molecular representations. Conventional analytical and organic chemistry identifies the structural components that compose heavy oil. Spectroscopy, including infra-red (IR) and nuclear magnetic resonance (NMR), is very important in quantifying the structural attributes that compose heavy oil. Atomic analysis is used to quantify the concentrations of the various atomic species present in heavy oil. To separate heavy oil into narrower boiling fractions that are easier to characterise, distillation is routinely used. The main challenge of molecular representations is integrating data from all of these experimental techniques. The data integration must yield individual molecules that are consistent with the experimental data and with the laws of chemistry and physics.

The overall goal of this thesis is to describe a new method of creating heavy oil molecular representations. Before the basis, fundamentals, and workings, of this method are explained, a detailed background of material relating to heavy oil molecular representations will be provided. To appreciate the complexity of molecular representations, one first must appreciate the complexity of the structure of heavy oil (Chapter 2). The origins of petroleum will first be explained. After this, the current knowledge about the structure of heavy oil and the analytical techniques that allowed this knowledge to be acquired will be documented. The historical progression of molecular representations will then be illustrated (Chapter 3). To demonstrate that molecular representations are useful, molecular based kinetic simulations will then be discussed.

Chapter 4 will describe in detail how heavy oil molecular representations are made. First the data that form the basis of the molecular representations in this thesis will be described. After this it will be shown how these data must be transformed in order to create molecular representations from it. At this point the algorithm that creates the heavy oil molecular representations will be explained. And finally, it will be demonstrated how this algorithm is used along with an optimisation routine to create heavy oil molecular representations.

Chapter 2

Chemical and Physical Analysis of Bitumen

Heavy oil is a very complex mixture of hydrocarbons. The complexity of the structural chemistry of heavy oil is responsible for most of the challenges associated with these feeds. The molecules found in heavy oils are essentially hydrocarbon chains containing aromatics, naphthenics, and aliphatics. Structures including fluorenes and porphyrins, also add to the complexity of heavy oil. To complicate matters, oxygen, sulphur, and nitrogen, are prevalent in heavy oil. Although these heteroatoms exist at low levels in heavy oil, they significantly influence both the structure and reactivity of heavy oil. Also, heavy oil is a very disperse mixture in terms of both molecular size and primary structure. Unprocessed heavy oil feeds consist of molecules with molecular weights ranging from 290 to 3220 g/mol (Gray *et al.*, 1992). It is estimated that a single heavy oil feed contains hundreds of thousands of components (Gray *et al.*, 1992).

2.1 Bitumen and Asphaltenes

As conventional light oil resources deplete, it is becoming necessary to utilize unconventional supplies. To meet the demand for petroleum, heavier sources including heavy oils and bitumens are being utilized. Bitumens are found in bituminous sands (oil sands). The Athabasca oil sand deposit in Alberta, is the worlds largest (Camp,1976). It is estimated that 830 billion barrels of oil is located in this reserve (Outrim and Evans, 1977).

Bitumens and heavy oils are defined by viscosity and density (Gray, 1994). The UNITAR definition of bitumen and heavy oil is found in Table 1. Bitumens and heavy oils have much higher viscosities and heteroatom compositions than conventional light crude oils (Table 2).

Table 1: UNITAR definitions of heavy oils and bitumens (Gray, 1994).

	Viscosity (mPa•s)	Density (g/cm^3)	API Gravity*
Heavy Oil	10^2-10^5	0.934-1.0	20-10
Bitumen	$>10^5$	>1.00	< 10

*API gravity=141.5/(specific gravity at 15.6°C) – 131.5

Table 2: Properties of light and heavy crude oils (Gray, 1994)

	Light Crude	Cold Lake	Athabasca	Morichal
API Gravity	38	10	9	4.9
Sulphur (wt%)	0.5	4.4	4.9	4.1
Nitrogen (wt%)	0.1	0.4	0.5	0.8
Metals (wppm)	22	220	280	863
Viscosity ($m^2/s*10^6$ at 40°C)	5	5000	7000	
Vacuum Resid 525°C+,(Liquid Vol %)	11	52	52	80

Many heavy oil characterization studies have focused on the asphaltenes. Asphaltenes are the highest molecular weight components present in heavy oil (Payzant *et al.*, 1991). Asphaltenes are soluble in benzene but insoluble in *n*-pentane. Asphaltenes comprise between 15-20% of Athabasca bitumen. Asphaltenes are characterized by high heteroatom contents. Alberta Athabasca asphaltenes contain sulphur (8-9%), oxygen (2-3%), and nitrogen (1%) (Ignasiak *et al.*, 1977).

2.2 Chemical Analysis Techniques

Various analytical techniques have contributed to our current knowledge of the chemical components present in heavy oil. ^{13}C-NMR, ^{1}H-NMR, gas chromatography, x-ray diffraction, IR, and distillation, are the techniques that have had the most significant impact in defining the chemical composition of heavy oil. To elucidate the complex structure of heavy oil very advanced chemical methods are required. Ruthenium ion-catalyzed oxidation (Strausz *et al.*, 1992) and pyrolysis (Payzant *et al.*, 1991) are two organic analytical techniques that have given tremendous insights into the structure of bitumen. Despite many recent advances in analytical and quantitative chemistry, a detailed analysis of bitumen is still not possible. For low molecular weight petroleum fractions such as naphthas, detailed characterization and composition analysis is possible (Ramadhan *et al., 1997*). However for heavier petroleum (or higher molecular weight) fractions such a detailed analysis is not possible. To date, it has been impossible to identify specific molecules that are found in heavy oil feeds. It is possible to determine various analytical properties of a bulk heavy petroleum fraction. To gain additional resolution, bulk heavy oil fractions are often analyzed. By using bulk chemical data, inferences and deductions must be made regarding the structure of individual molecules.

2.2.1 Elemental Composition

Bitumen is composed mainly of carbon and hydrogen. Carbon percentages in bitumen range from between 80% to 85% while hydrogen contents vary between 5% to 14% (Gray, 1994). Sulphur (5.14%), nitrogen (0.56%), and oxygen (1.17%) are undesirable heteroatoms that are also found in bitumen (Table 3). Iron, nickel, and vanadium are present at less than 300 ppm.

Table 3: Elemental composition of Athabasca bitumen (Gray, 1994).

Element	Composition
C*	83.0%
H*	10.1%
O	1.17%
N	0.56%
S	5.14%
Ni	150 ppm
Vn	290 ppm

*C and H values were determined using atomic mass balances, and the reported H/C ratio.

2.2.2 Ruthenium Ion Catalysed Oxidation

Ruthenium ion catalyzed oxidation (RICO) has had a very significant role in analyzing the structure of asphaltene. Although RICO was first proposed in the 1950s (Djerassi and Engle, 1953), it was not applicable to the analysis of petroleum until acetonitrile was found to be an ideal cosolvent (Carlson *et al.*, 1981). Acetonitrle prevents the precipitation of intermediates in the RICO reactions. This significantly improves the reproducibility and efficiency of this method. RICO almost completely oxidizes all aromatic carbons to carbon dioxide and polycarboxylic acids while not affecting the surrounding and connecting alkyl carbons. Alkyl attachments on aromatic rings are converted to carboxylic acids. Condensed aromatic species are oxidized to polycarboxylic acids. Benzene rings are oxidized to carbon dioxide. Alkyl chains that are attached to aromatic rings become alkanoic acids. It has been determined that the RICO catalyzed reactions are nearly quantitative (Strausz *et al.*, 1992). By using chemical standards, the RICO products are analysed using chromatography. The starting material, products, and stoichiometry, of many RICO reactions are known. Thus, identification and quantification of the endproducts of RICO reactions, allows insights into the structure of heavy oils.

RICO has also been used to analyze the condensation levels of Athabasca asphaltenes. When fused aromatic ring species such as naphthalene, phenanthrene, and chrysene, are oxidized by RICO, benzene polycarboxylic acids are formed (Strausz, 1988). Based upon the analysis of the types of benzene polycarboxylic acids produced, it is possible to deduce the degree of condensation of the original starting aromatic material. In general, the more fused rings an aromatic species has, the more carboxylic groups the resulting benzene carboxylic acid will have. Before RICO was further modified by Wang *et al.* (1997) it was not possible to identify higher order benzene polycarboxylic acids. This improvement made it possible to identify penta and hexa acids in addition to the di, tri, and tetra acids.

2.2.3 Pyrolysis

In addition to RICO, pyrolysis has also been an important tool in analyzing the structure of heavy oil. Originally, pyrolysis was used to simulate natural geological diagenesis (Rubinstein and Strausz, 1979a; Rubinstein *et al.*, 1979b). It was latter realized that pyrolysis would be a useful tool in identifying and analyzing asphaltene components. After pyrolysing asphaltene the resulting products are identified using gas chromatography and/or mass spectrometry (Payzant *et al.*, 1991).

2.2.4 NMR Analysis of Heavy Oil

Of all analytical techniques, NMR has had the greatest impact on heavy oil molecular representations. What makes NMR so powerful is that it can both identify specific components found in heavy oil, and it can analytically determine the concentrations of those components. A drawback of this technique is that interpreting the NMR spectra of heavy oils can be very challenging. There is substantial overlap in both the ^{13}C-NMR and ^{1}H-NMR spectra of heavy oil. To resolve the spectra into discrete individual components, mathematical optimisation is often necessary.

In the 1960s NMR became a widespread technique used to characterize heavy oil. Initially ^{1}H-NMR analysis was used both to calculate global physical parameters and relative proton concentrations. Physical parameters including aromaticity, the degree of substitution, and the atomic hydrogen to carbon ratio were calculated as early as 1960 (Brown *et al.*, 1960, a,b). The distributions of protons in Athabasca bitumen among aromatic, alpha, naphthenic, paraffinic methylene, and paraffinic methyl groups were calculated by Speight (1970, 1971) using ^{1}H-NMR.

The combination of ^{13}C-NMR and ^{1}H-NMR yields the most information regarding the various carbon types in heavy oil (Dickinson, 1980). Unfortunately, the NMR spectra of heavy oils contain many overlapping carbon signals and are often difficult to interpret. Before ^{13}C-NMR was used to analyze heavy oil, many assumptions had to be made regarding the aliphatic groups present in NMR. It was assumed that the H/C ratio of α groups attached to aromatic rings was the same as the ratio of the rest of the side chain. Another assumption was that the H/C ratio of the alkyl groups could in fact be measured using ^{1}H-NMR. A major assumption was that no aromatic rings were directly linked to intervening ring alkyl groups. These assumptions were very limiting in analysing heavy weight materials such as asphaltenes and residue fractions. When ^{13}C-NMR is used along with ^{1}H-NMR the last assumption is no longer necessary and the C/H of the aliphatic groups is measured directly from both spectra types, instead of estimating it directly from the ^{1}H-NMR spectra.

In the late 1970s ^{13}C-NMR was used for detailed peak analysis that was not possible with ^{1}H-NMR alone. Snape *et al.* (1979) assigned many peaks in the ^{13}C-NMR spectra to many different types of carbon. Snape *et al.* (1979) identified peaks corresponding to methylene groups in chains, α-methylene, β-methylene, methylene adjacent to S and O, and naphthenic methylene. The peaks corresponding to branched methyne and dibenzothiophene associated carbon were also identified. Gray and Thiel (1988) used the model compound farnesane to assist in the interpretation of ^{13}C-NMR spectra. Peaks in the ^{13}C-NMR spectra of Athabasca bitumens and gas oils were related with the peaks of specific carbons and subunits of farnesane. Peaks that did not correspond to farnesane peaks were determined from literature assignments. In total 19 discrete ^{13}C-NMR peaks were assigned.

2.2.5 Distillation and Liquid Chromatography

Distillation is commonly used in the analysis and characterization of heavy oil. More accurate structural information is obtained by characterizing relatively narrow boiling distillation fractions than entire heavy oils. In the distillation of heavy oil a common four fraction designation is (Gray, 1994): naphtha (IBP-177°C), middle distillate (177-343°C), gas oil (343-524°C), and residue (524°C+).

SARA is the simplest method used to chemically define a heavy oil sample. SARA is a column separation scheme that is widely used to characterize heavy petroleum fractions (Speight, 1991). SARA uses a series of columns to separate a petroleum feedstock into saturates, aromatics, resins, and asphaltenes. Usually the columns that are used in sequence are anion resin, cation resin, clay/$FeCl_3$, and silica-alumina. There are two major factors that contribute to the widespread use. It is very reproducible, and it separates asphaltenes as a group. For an Athabasca residue (524 °C+ fraction) a typical SARA analysis (Chung and Xu.) is saturates (5 wt%), aromatics (33 wt%), residues (30 wt%), and asphaltenes (32 wt%).

2.3 Asphaltene Structure

O.P. Strausz from the University of Alberta has had a major role in determining the structure of heavy oil by using an unique combination of organic chemistry and analytical analysis. Thermal degradation and RICO were all successfully used to identify the major structural blocks found in heavy oil (Strausz *et al.*, 1992). Through detailed analysis, Strausz *et al.* (1992) identified many aromatic, naphthenic, and aliphatic, constituents found in heavy oil. Extensive characterization indicates that many of the compounds found in heavy oils were derived from biomolecules including carotenoids, vitamin A, tricyclic terpanes, tricyclic terpenoid sulphides, pentacyclic terpenoids, and steranes. Strausz has also shown that compared to what was originally thought (Figure 1), a lower degree of aromatic saturation most likely exists in asphaltene (Figure 2). Strausz *et al.* (1997) also showed that Athabasca asphaltenes contain a large concentration of mono and di condensed aromatic molecules, and lower concentrations of tri, hexa, and penta, condensed aromatics.

The initial characterisation of heavy oil in the 1950's indicated that the predominant structures were aliphatic alkyl and aromatic. This was based upon NMR and x-ray diffraction studies. At this point, the relative abundance of naphthenic structures was underestimated in asphaltene. Yen *et al.* (1961) analyzed asphaltenes using x-ray diffraction. Based on their measurements they concluded that asphaltene is composed of an alternating series of *n*-alkyls and clusters of condensed aromatic disks. Dickie and Yen (1961) were later found to have overestimated the x-ray diffraction signal attributed to the aromatic discs, thus overestimating the degree of condensation.

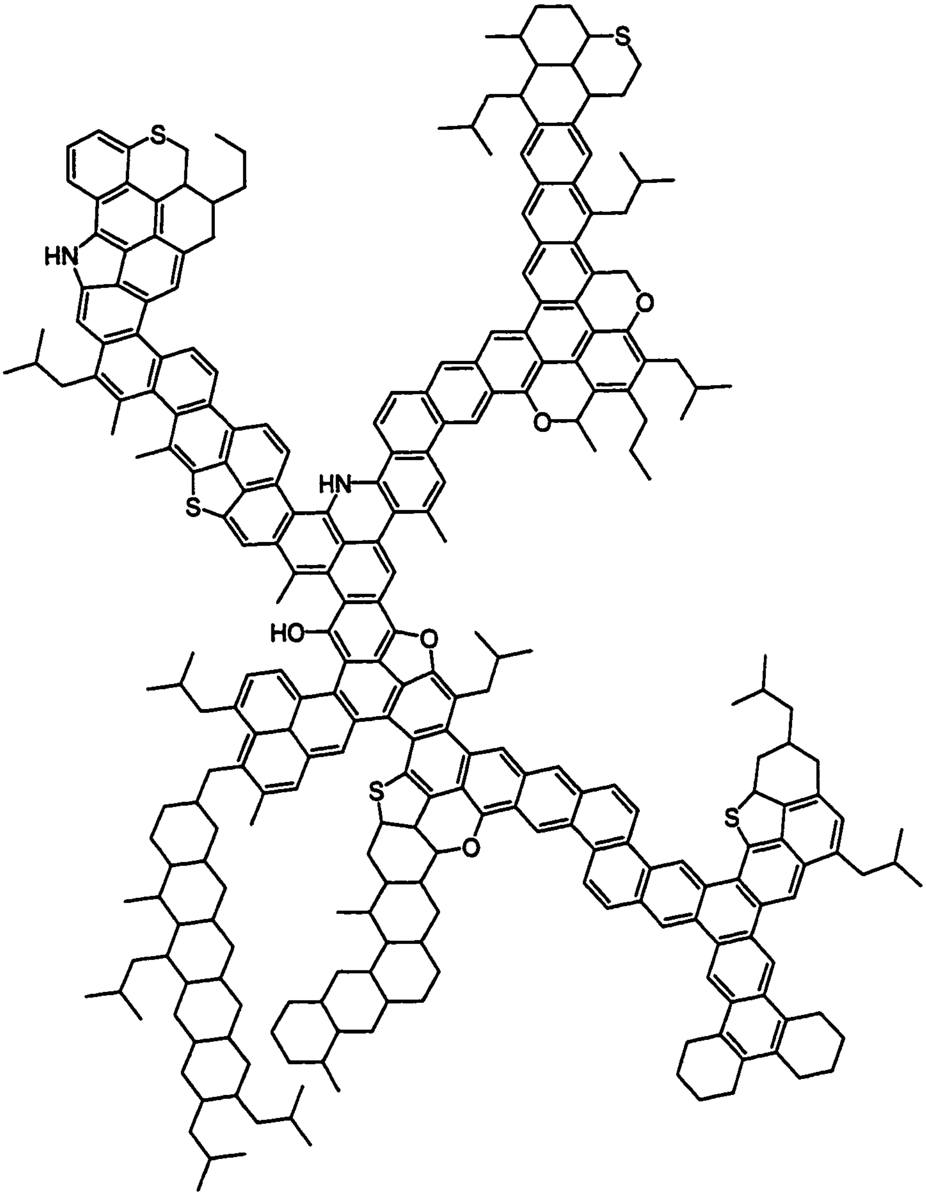


Figure 1: A heavy oil molecule that is inconsistent with current analytical information (Altamiran, 1986).

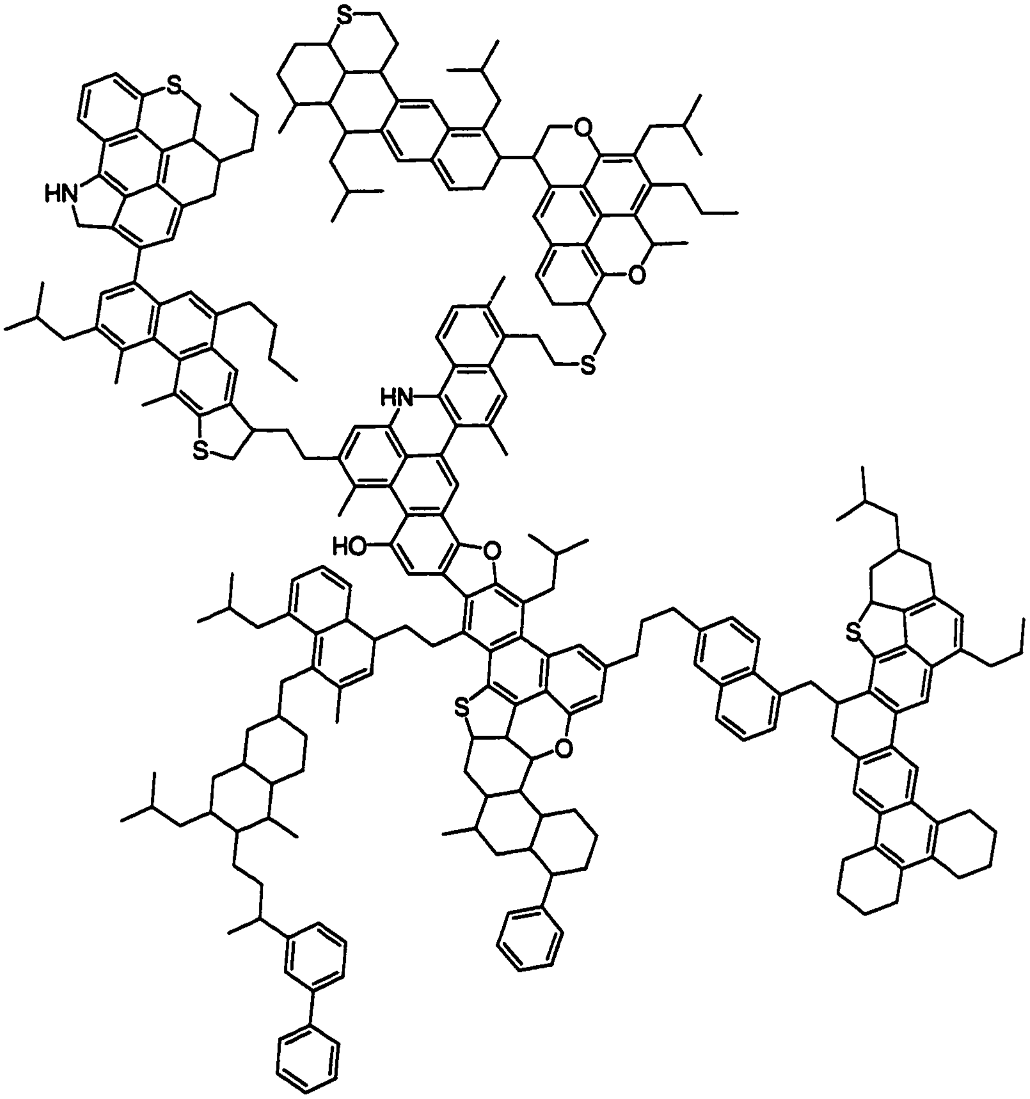
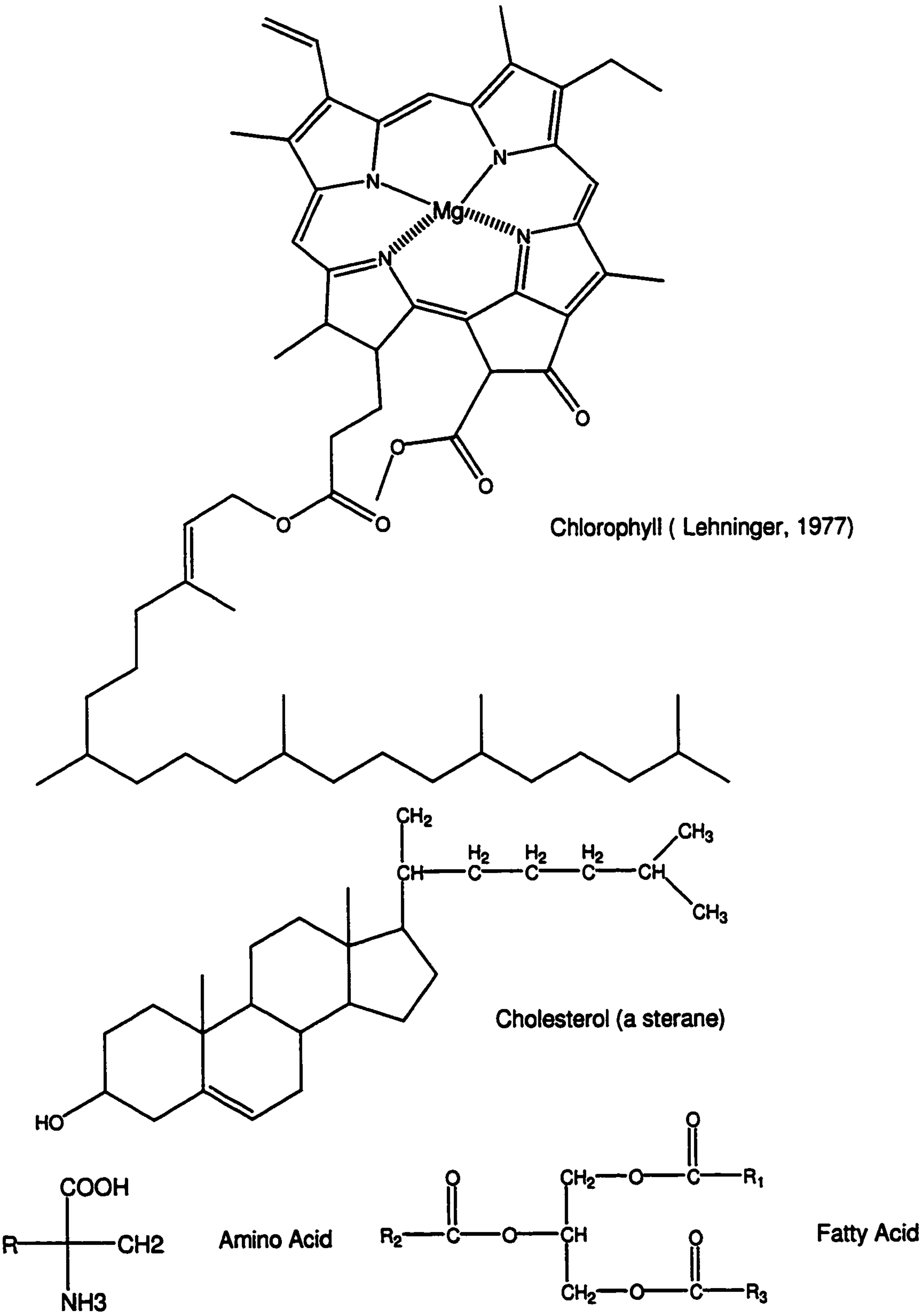


Figure 2: A heavy oil molecule that is consistent with current analytical information.

2.3.1 The Biological Origins of Petroleum

The complexity of the structural chemistry of petroleum is due to several factors. First of all, petroleum originates from biological material, predominately marine phytoplankton. Even the simplest cells that compose biological material are tremendously complex. Cells are composed of a variety of fats, lipids, proteins, sugars, and nucleic acids. Biodegregation and thermal reactions further decompose and react the biosource material to eventually create petroleum. When studying the structure of petroleum it is very important to take into account its biological origins. Insights can be made about the structure of petroleum by studying the structures present in plants. Many structural components present in the biosource material such as petrochlorophyls, amino acids, steranes, and fatty acids, are precursors to structures that are found in oil (Figure 3). Chlorophyls are structurally similar to porphyrins complexed with metals. The nitrogen present in oil is believed to have originated from amino acids. Sterols are believed to be the source of the aromatic rings present in oil. Fatty acids are the most probable source of the paraffinics.

Petroleum is regarded as a gaseous, liquid, or solid mixture of hydrocarbons and/or hydrocarbon derivatives that naturally occur in the earth (Speight, 1980). Lighter hydrocarbons compose gaseous petroleum, while liquid hydrocarbons, waxes, dissolved gases, and bitumenous materials form liquid petroleum. Heavier hydrocarbons are found in solid and semisolid petroleum. Petroleum results from a multistep process that involves bacterial degradation of animal and plant deposits on seabeds. The first stage in the production of petroleum mostly plant and some animals are deposited at the bottom of seabeds (Speight, 1980). This layer of 'biosediment' is covered with a layer of sediment. This top layer creates the anaerobic conditions necessary for bacterial degradation. Eventually the top layer becomes impermeable and prevents the leakage of petroleum from the deposit. When anaerobic conditions are established, bacteria decompose the biomaterial.



Chlorophyll (Lehninger, 1977)

Cholesterol (a sterane)

Amino Acid

Fatty Acid

Figure 3: The biological precursors of many of the components found in petroleum.

Bacterial decomposition plays a major role in petroleum formation. The marine biomass is converted by bacterial decomposition and syntesis of new organic compounds by enzymatic transformations (Speight, 1980). Decarboxylations, deaminations, cyclizations, hydrogenations, and isomerizations are all thought to be enzyme catalysed reactions that have roles in petroleum transformation. If bacteria transformation has a major role in the production of oil and bacterial biomass is not found in the final petroleum product, it is only logical to conclude that bacterial biomass is incorporated into petroleum. Carbohydrates and most of the proteins are converted either to gases or water-soluble material, and diffuse into seawater (Speight, 1980). Fats, waxes, and stable materials (rosins, cholesterol, etc.) remain in the deposit. In the second stage of petroleum formation, high temperature and pressure result in the liberation of carbon dioxide and water from chemical groups containing carboxyl groups and hydroxy acids, respectively (Speight, 1980). This results in the formation of a bitumenous (very heavy) residue. The bacterial degradation products chemically react to form petroleum. Many chemical reactions are believed to occur during the second stage. Decarboxylations, deaminations, cyclizations, and hydrogenations, all lead to the formation of petroleum (Speight, 1980). The high pressure and temperature result in light cracking, producing petroleum with a high olefin content.

While marine phytoplankton is the prominent source material for petroleum formation, local deposit conditions may also favour the deposition of marine algae, larger marine animals, and even material from terrestrial sources (Speight, 1980). The nature and type of material in sediment influence the characteristics of the petroleum that results. Over time, the sediment covering the biolayer becomes thicker, and the pressure inside the biolayer increases. As the pressure in the biolayer increases, the water content decreases from an initial 70-80% to 10% or less. Bacterial biochemical hydrogenation plays two important roles in the biomaterial degradation. Firstly, it removes carbon dioxide from fatty acids to form paraffins (hydrocarbons), and secondly it reduces sulphate to hydrogen sulphide.

A wide variety of biomolecules serve as the source of petroleum (Speight, 1980). Pigments, fatty acids, sterols, and amino acids, are all believed to be incorporated into petroleum. Porphyrins are nitrogen containing products of chlorophylls (a magnesium porphyrin), hemes (iron porphyrin), and corrinoids (cobalt porphyrin). Porphyrins are found in petroleum usually complexed with vanadium. Alternatively, they are sometimes complexed with nickel, iron, or copper. Fats are mixtures of various fatty acids complexed with glycerides. Hydrocarbons can be formed from fatty acids when carboxyl groups are removed from them. Fats constitute a large weight percentage of aquatic organisms and are believed to be the largest contributing molecule type to petroleum formation.

Carbohydrates are composed of sugar molecules. They can either be composed of individual units such as glucose and sucrose or they can be polymers such as cellulose and glycogen. In plants, the predominant carbohydrate is cellulose. It is believed that most carbohydrate material found in biolayers was not incorporated into petroleum. Most carbohydrates were either metabolised by bacteria or lost to seawater because of their high water solubility (Speight, 1980).

Proteins are long polymers of amino acids. Chemically, the proteins are very reactive. Usually, proteins are readily hydrolysed by bacteria to yield water soluble degradation products. However, when many proteins loose their structural organisation (denaturation) they become insoluble in water. Thus, denatured protein is very likely preserved in biolayers and contributes to the formation of nitrogen compounds in the petroleum (Speight, 1980).

2.3.2 Heavy Oil Aliphatics

Aliphatic groups have a significant role in the structure of heavy oils. The naphthenics and paraffinics collectively form the aliphatics. Paraffinics (Figure 4) are long chain alkanes that either branch off aromatic clusters or connect them together. The breakage of paraffinic C-C bonds during upgrading is an important molecular weight decreasing reaction. The naphthenics (Figure 4) are saturated ring systems that originated from the biosource material (Strausz *et al.*, 1999). Hence they are called biomarkers.

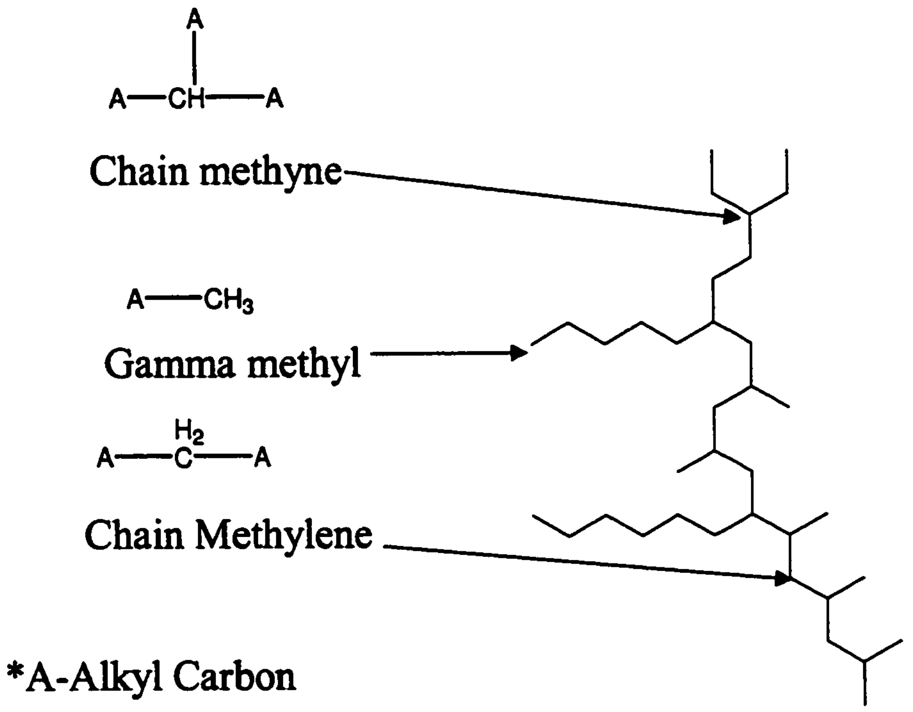


Figure 4: Paraffinic carbons that are found in heavy oils (Gray *et al.*, 1992).

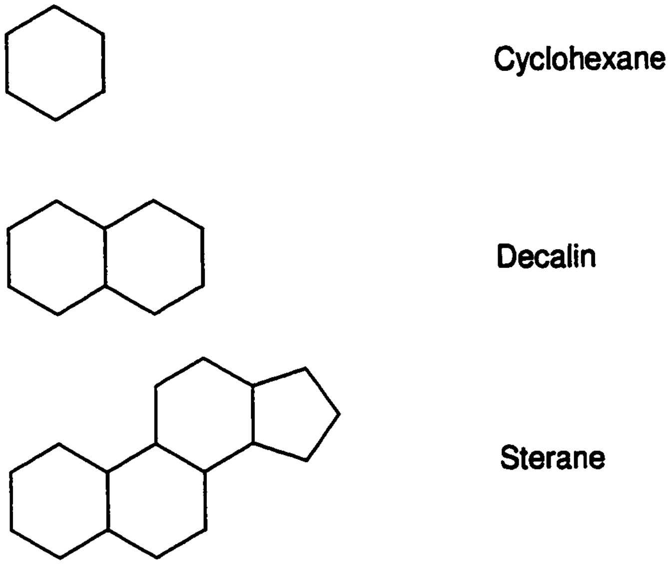


Figure 5: Some naphthenic components found in heavy oil (Gray, 1992).

Mojelsky *et al.* (1991) used RICO to determine the concentrations and distributions of alkyl chains attached to single aromatic rings, and of methylene bridges connected between aromatic rings. Using gas chromatography (GC) analysis on Athabasca asphaltene, alkyl chains having lengths up to 32 carbons were observed. The average chain length was 9.18 carbons while the most common chain length was 1 (α). For the methylene bridges between aromatic rings, bridges as long as 24 carbons were observed while the average length was 3.82 carbons.

The sidechains associated with aromatics in Athabasca maltenes were studied by Mojelsky *et al.* (1986). Using pyrolysis at 300°C, it was found that most alkane chains had lengths between C_{15} and C_{17}. Results from RICO indicated that the alkane chains had lengths between 12 and 16 carbons long. It was concluded that the alkane distribution produced by pyrolyis does not reflect the actual distribution of alkanes connected to aromatic cores.

RICO has assisted in identifying many naphthenes and biomarkers present in asphaltenes. Isoprenoids, cheilanthanes, hopanes, steranes, and pregnanes, have all been identified using RICO along with standard compounds (Strausz *et al.*, 1999). Peng *et al.* (1999) identified hopanes, steranes, pregnanes, and cheilanthanes, in a Chinese salinine lake asphaltene using RICO.

2.3.3 Heavy Oil Aromatics

Pyrolysis studies have identified many different types of aromatic compounds present in Athabasca asphaltene. It has identified sulphur containing compounds, including thiolanes, thiophenes, and benzo[b]thiophenes (Payzant *et al.*, 1991). Payzant *et al.* (1991) concluded that Athabasca asphaltene is 0.2% fluorene based upon their pyrolysis study. Alkylpyridines, quinolines, *n*-alcohols, *n*-alkanoic acids, and *n*-alkylamides (Strausz *et al.*, 1999) have been identified in Athabasca asphaltene using pyrolysis, GC, and mass spectrometry.

2.3.3.1 Aromatic Condensation Levels In Bitumen

Even to this day, the degree of aromatic condensation in heavy oils is a controversial topic. Early evidence in the 1960s (Yen *et al.*, 1961) indicated that aromatic clusters in heavy petroleum fractions are predominately highly condensed multi-ring aromatic systems. In the late 1970s experimental evidence started to indicate that highly condensed aromatic systems are not very prevalent in unprocessed heavy oil fractions (Payzant *et al.*, 1979). By the late 1990s this was supported by many different physical (El-Mohamed *et al.*, 1986) and chemical (Strausz *et al.*, 1997) analyses. Despite extensive experimental evidence supporting low condensation levels, many researches still believe that high aromatic condensation levels are present in heavy oil.

2.3.3.1.1 X-ray Analysis

The first study of the structure of heavy oil using modern analytical methods was that of Yen *et al.* (1961). Yen used x-ray diffraction studies to gain insights into the structure of asphaltene. This work significantly influenced many other researchers studying the structure of heavy oil. Based predominately on NMR and x-ray diffraction data, Yen *et al.* (1961) concluded that the aromatic structures found in asphaltenes are very highly condensed. Even though these conclusions are inconsistent with those from more modern experimental methods, it is important to discuss the contributions of Yen as they are still extensively employed in the interpretation of analytical data.

In their first published study, Yen *et al.* (1961) group studied eight different asphaltenes. X-ray diffraction generates quantitative intensity curves. Structural related features can be calculated from the shape and position of the peaks that are generated by an x-ray analysis. The x-ray diffraction patterns of asphaltene very closely resembles the pattern that is produced by a 1:1 mixture of polyethylene and carbon black. Carbon black contains condensed aromatic sheets while polyethylene is representative of the carbon chains that are found in heavy oil. The diffraction analysis of the asphaltenes allowed the aromaticity (fraction of carbon being aromatic) to be calculated. The aromaticities were all between 0.26 and 0.53. Using the Scherrer crystallite size formula, the average diameter of asphaltene aromatic clusters was determined to be 14 Angstrom.

Yen and Erdman (1963) were interested in determining the shape and arrangement of the aromatic clusters found in asphaltene. The ratio of the maximal potential hydrogen substitution to aromatic carbon content (calculated from NMR) was plotted against the inverse of the asphaltene layer diameter (H_I/C_A versus $1/L_a$). Layer diameter was determined using x-ray diffraction. Points corresponding to nine different asphaltenes were plotted alongside two theoretical plots corresponding aromatic rings arranged in circular and linear arrangements. It was concluded that for a Burgan asphaltene, the average aromatic structures resembled 3',5',1,10-tetracenyl (a 7 ring aromatic system, consisting of two stacked aromatic layers), or 2',7',1,8 anthryl (a 6 ring aromatic system, consisting of two stacked aromatic layers). For a Lagunillas asphaltene, the average aromatic system was found to be closest to nine naphtho-2',7',1,13-ovalene (a pericondensed 10 ring system). These average structural assignments were based upon the comparison of the ratio of the maximal potential hydrogen substitution to aromatic carbon content. Based upon both the average aromatic structure assignments and the H_l/C_A versus $1/L_a$ plot, it was concluded that aromatic sheets are predominant in asphaltenes.

Yen and Dickie (1962) further examined the condensation levels in asphaltenes. More specifically they examined the compactness of asphaltenes. Compactness (F/C_A) is defined as the ratio of fused aromatic carbon atoms (F) to the total number of aromatic carbon atoms (C_A). F in turn represents the number of carbon atoms that are connected to three other carbons. A correlation was developed that related compactness to aromaticity, ring condensation index, and C_A. Aromaticity was determined using x-ray diffraction, C_A was approximated using x-ray diffraction data, and the ring condensation index was calculated using the densimetric method. The compactness of seven different asphaltenes was analyzed. Yen and Dickie (1962) stated that positive values of compactess, as calculated using their correlation, correspond to pericondensed aromatic systems, while negative values correspond to either linked (aromatic rings only connected by biphenyl bridges) or kata-condensed aromatic species (a single fused layer of aromatic rings). For the asphaltenes the compactness values ranged from 0.5 to –0.085. Six out of the seven asphaltene samples had positive compactness values. Based upon the results in this study it was concluded that aromatic systems found in asphaltenes are highly condensed. The correlation is used to calculate compactness involves considerable error. Yen and Dickie (1962) indicate that it is very difficult to interpret x-ray diffraction patterns. Also, the correlation used to calculate ring compactness was overly simplified. The molar contraction value was assumed to be independent of H/C and aromaticity when in fact a function of these two parameters.

2.3.3.1.2 Condensation Levels Determined Using RICO.

When Mojelsky *et al.* (1992) analyzed the benzene polycarboxylic acids that were produced from Athabasca asphaltene by RICO, no benzene penta and hexa -carboxylic acids were detected. Based upon this observation, it was concluded that pericyclic aromatic rings are not present in Athabasca asphaltene. However when Strausz *et al.* (1997) analysed Athabasca asphaltene with Wang's modified RICO, both benzene penta and hexa carboxylic acids were observed in low concentrations (4.3 and 1.5% of all benzene polycarboxylic acids). Benzene di, tri, and tetra-carboxylic acids were detected at much higher concentrations, 36.2%, 32.6%, and 25.4%, respectively. It has been shown experimentally that aromatic species containing four fused rings such as acephenanthrylene and triphenylene will yield benzene penta and hexa-carboxylic acids when reacted with RICO. Pericondensed aromatic fused systems containing five or more fused rings such as dibenzopyrene and cornene can also yield penta and hexacids. Strauz observed that penta and hexacids are produced at levels 3 to 4 times lower compared to what would be expected if each asphaltene molecule contained at least one pericondensed aromatic system. Strausz *et al.* (1997) further concluded that "pericyclic aromatic moieties have no significant role in the average asphaltene molecule".

According to Strausz, Wang *et al.* (1997) mistakenly claimed that the aromatic structures of Guado asphaltene are mainly pericondensed. Using Wang's published data, Strausz calculated that one out of every 7.62 Guado asphaltene molecules has one pericyclic system. This calculation assumed that all of the hexacid came from pericondensed systems. Small ring systems including triphenylene and perylene, in all probability produce a significant portion of the hexacid.

2.3.3.1.3 Additional Support For Low Aromatic Condensation Levels In Asphaltenes

Results from field-ionization mass spectrometry (Payzant *et al.*, 1979) indicate that mono to hexa-cyclic aromatic compounds are found in Lyodminister heavy oil. The mono and di-cyclic compounds are most common while the tri to hexacyclic compounds are present at much lower concentrations.

El-Mohamed *et al.* (1986) studied the diamagnetic properties of various heavy petroleum products. Included in their study were asphaltenes, asphalts, vacuum residues, and deasphalted oils. Using the Faraday method, the magnetic susceptibilities of each of these petroleum products was determined. By using Pascal's additivity method, a parameter defined as d was then calculated for each fraction. The chemical properties of a molecule significantly influence its d value. The d values calculated for 11 heavy petroleum fractions ranged from –0.21 to +1.47. Large pericondensed systems such as those proposed by Yen (1972), -0.93, have d values that are negative (anthracene: -0.122, phenanthrene: -0.158). Conversely, saturated naphthenic rings have positive d values. El-Mohamed *et al.* (1986) reasoned that if the fractions in their study all contained large pericondensed aromatic clusters they would also have to contain unsaturated naphthenic rings in more than a 10/4 ratio. Using ^{13}C-NMR the carbon aromaticity (% aromatic carbon/% total carbon) for most of the fractions was found to be approximately 50 %. It was concluded that large condensed aromatic systems are not found in the fractions that were studied. El-Mohamed *et al.* (1986) further stated that the aromatic systems in heavy oil fractions "must contain small aromatic systems probably connected by short aliphatic bridges, and linked also with polyhydrogenated rings whose condensation must be weak for the lighter fractions and higher for the heavier fractions".

2.3.3.1.4 The Effect of Thermoprocessing on Aromatic Condensation Levels In Petroleum

Evidence exists that the aromatic saturation level of petroleum increases when it is hydrocracked. Sullivan *et al.* (1989) used a very novel experimental design to determine if aromatic condensation increased during oil hydroprocessing. They used a three stage pilot plant to hydroprocess a Middle Eastern atmospheric residuum. The first stage simulated a desulphurization process. Product from this stage was distilled. The distillation produced three cuts, a vacuum gas oil (345-540°C), a SOR fraction (start of run, 345°C, top), and an EOR (end of run, 540+°C, bottom). The vacuum gas oil that was produced was then hydrotreated to remove nitrogen and sulphur. The product was then distilled and the bottoms were hydrotreated once again and the product was distilled. The bottoms from the second hydrotreater were recycled to enter the second hydrotreater.

The aromatic condensation level in the vacuum gas oil was analysed using a consortium of analytical techniques. High performance liquid chromatography, UV/vis diode-array detection, and field-ionization mass spectrometry were all utilised. In both the SOR and EOR fractions leaving the first distillation tower the concentration of mono and di condensed aromatic compounds was found to be high (27.50 wt% and 28.46 wt %, respectively for the mono-aromatics; and 13.04 and 14.35 wt% for the diaromatics). The concentrations of tri (2.862 vs 1.198 wt%), tetra (0.729 vs 0.187 wt%), and penta and higher (0.380 vs 0.134 wt%), aromatics was found to be substantially higher in the higher boiling EOR fraction compared to the lighter boiling SOR fraction. When the VGO (vacuum gas oil) was cracked in the first stage the total aromatics were reduced from 47 to 14 wt %. The second stage of cracking reduced this further to about 2 wt%. It was found that hydrotreating basically had no affect on the concentration of tetra-aromatics. However, it was found that the concentrations or hexa and hepta-aromatics increased rapidly while those of mono, di, tri, and penta-aromatics decreased. These observations support the postulate that hydroprocessing increases the degree of aromatic condensation in petroleum.

Poly-aromatic hydrocarbon (PAH) build-up depends on many variables. These include feedstock composition, flow regime, operation severity, and catalyst type. Zeolite catalysts have a greater propensity to form PAHs than amorphous catalysts. In general, PAH formation during hydrotreating is unwanted, as PAHs are difficult to crack.

2.3.4 Heteroatoms and Metals

Although the heteroatoms are present in low concentrations in heavy oil they significantly influence the reactivity and processability during upgrading. S, N, O, and V, are found in higher concentrations in the higher boiling point fractions (Gray, 1994). On a weight basis, high molecular weight fractions such as residues contain more heteroatoms than lower molecular weight fractions such as naphthas.

2.3.4.1 Sulphur

Of all the heteroatoms, sulphur is present at the highest concentrations (usually greater than 5%). Sulphur is present in many forms in heavy oil including primarily thiophenes, thioethers, thiols, and sulphoxides (Figure 6). It is undesirable to have sulphur in the end product that results after a heavy oil feed is upgraded. However, the initial presence of sulphur in the structure of bitumen is highly advantageous from an upgrading perspective. Asphaltenes have the general structure, [core]-S-[core]-S-[core]-S-[core] Peng *et al.* (1997). The core segments are held together by thermally labile thioether bonds. During upgrading these bonds tend to break, thus decreasing the molecular weight of the heavy asphaltene material. This is a major goal of heavy oil upgrading.

Peng *et al.* (1997) used Ni_2B reduction to study the contribution of sulphur to the overall asphaltene structure. Ni_2B specifically cleaves carbon-sulphur bonds. It was found that when HMA (high molar mass asphaltene) was reduced by Ni_2B it's molecular mass decreased 4-fold. This is due to the specific break-up of sulphide linkages. This large decrease in molecular mass is the result of a 3.3% decrease in the total content of the asphaltene. Based upon the 4-fold reduction in molar mass of HMA following Ni_2B reduction, it was concluded that HMA has a general structure: [core]-S-[core]-S-[core]-S-[core]. Upon reduction with Ni_2B this would yield 4 cores and $3H_2S$. Conversely the molar mass of LMA (low molar mass asphaltene) was found not to change with Ni_2B reduction. It was concluded that the core segments in LMA are not held together by sulphide linkages.

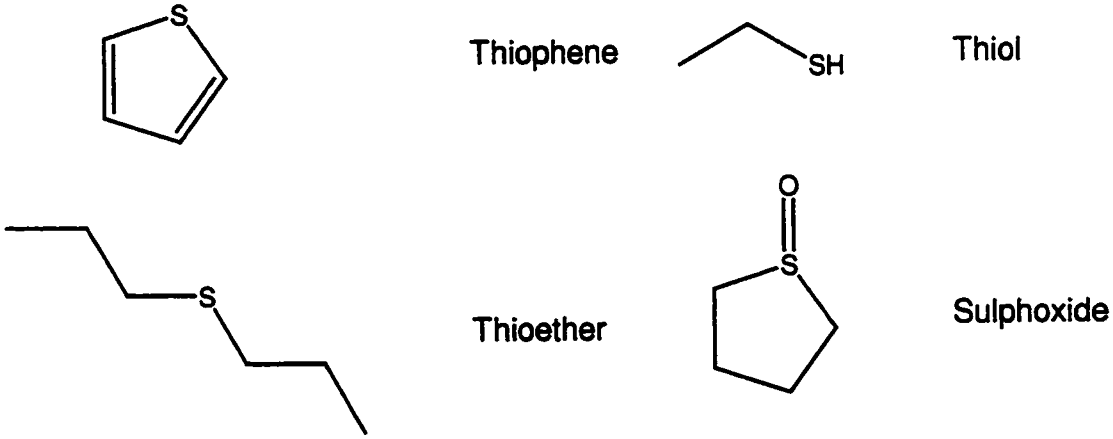


Figure 6: Sulphur-containing petroleum structures (Gray 1994).

2.3.4.2 Nitrogen and Oxygen

Nitrogen is present in two forms in heavy oil (Figure 7), as nonbasic pyrole derivatives, and as basic derivatives of pyridine (Gray, 1994). Nitrogen containing structures in heavy oils are difficult to remove during upgrading. Oxygen is found in aromatics such as furan, ketones, carboxylic acids, and phenols (Figure 8) (Gray, 1994).

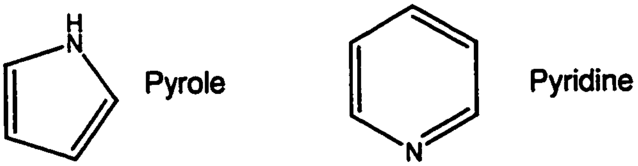


Figure 7: Nitrogen-containing petroleum structures (Gray, 1994).

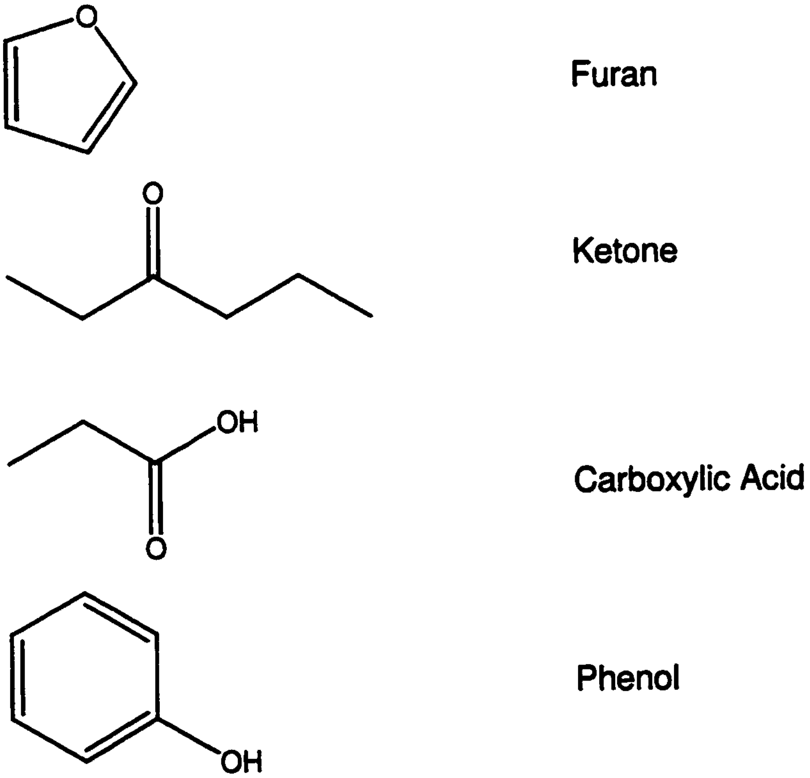


Figure 8: Oxygen-containing petroleum structures (Gray, 1994).

2.3.4.3 Metals

In heavy oil, metals occur in two forms, porphyrin metals, and nonporphyrin metals (Gray, 1994). The porphyrin metals (Ni or V) are complexed with a large aromatic pyrole cluster (Figure 9). Less information is available on the nonporphyrin metals although they are generally associated with polar groups.

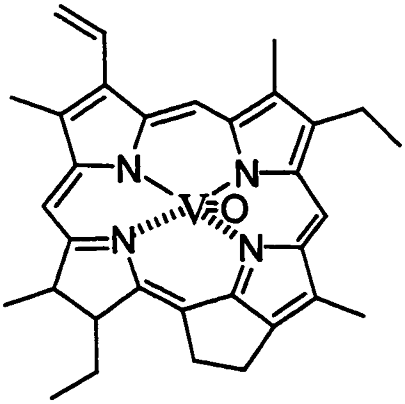


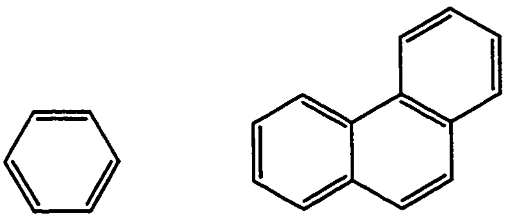
Figure 9: A vanadyl petroporphyrin.

2.4 Heavy Oil Chemical Structural Attributes

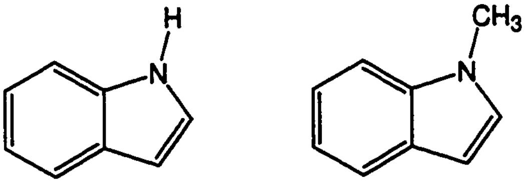
Gray *et al.* (1992) calculated the relative concentrations of different aromatic (Figure 10), aliphatic (Figure 11), and heteroatomic (Figure 12), functional groups present in various catalytic hydrocracked bitumen fractions. Four different heavy oil feeds, Athabasca, Cold Lake, Lloydminster, and Peace River, were hydrocracked using various catalysts. Using these feeds, four different catalysts were used: Ni/Mo on alumina, demetallization, alumina support (1/20" trilobe), and Mo on alumina support (1/20" trilobe). The product from each catalytic cracking run was analysed using distillation, vapour pressure osmometry (VPO), ^{13}C-NMR, ^{1}H-NMR, IR, potentiometric titration, and elemental analysis. The relative concentrations (mol functional group/100 g of oil sample) were determined (Table 4) either directly from NMR spectra, from differential analysis using either atomic or spectral region balances, or directly using IR and potentiometric titration.

Table 4: How the relative concentrations of various chemical functional groups present in various catalytic hydrocracked heavy oil fractions were calculated (Gray *et al.*, 1992).

Functional Group	Quantification Method
Benzene	Mass balance on aromatic carbon
Phenanthrene	^{1}H-NMR
Biphenyl Bridge	Balance on aromatic substitution
α-methyl	^{13}C-NMR, ^{1}H-NMR
α-methylene	^{13}C-NMR, ^{1}H-NMR
β-methyl	^{13}C-NMR, ^{1}H-NMR
γ-methyl	^{13}C-NMR, ^{1}H-NMR
Chain methylene	^{13}C-NMR, ^{1}H-NMR
Chain methyne	^{13}C-NMR, ^{1}H-NMR
Naphthenic methyne	^{13}C-NMR, ^{1}H-NMR
Naphthenic methyl	^{13}C-NMR, ^{1}H-NMR
Naphthenic methylene	^{13}C-NMR, ^{1}H-NMR
Benzothiophene	^{13}C-NMR
Sulphoxide	IR and potentiometric titration
Thioether	^{13}C-NMR
Indole	^{13}C-NMR
Quinoline	^{13}C-NMR
Amide	IR and potentiometric titration
N-substituted Indole	Balance on nitrogen
Benzofuran	Balance on oxygen
Carboxylic acid	IR and potentiometric titration
Ketone	IR and potentiometric titration
Aromatic hydroxyl	IR and potentiometric titration



Benzene Phenanthrene

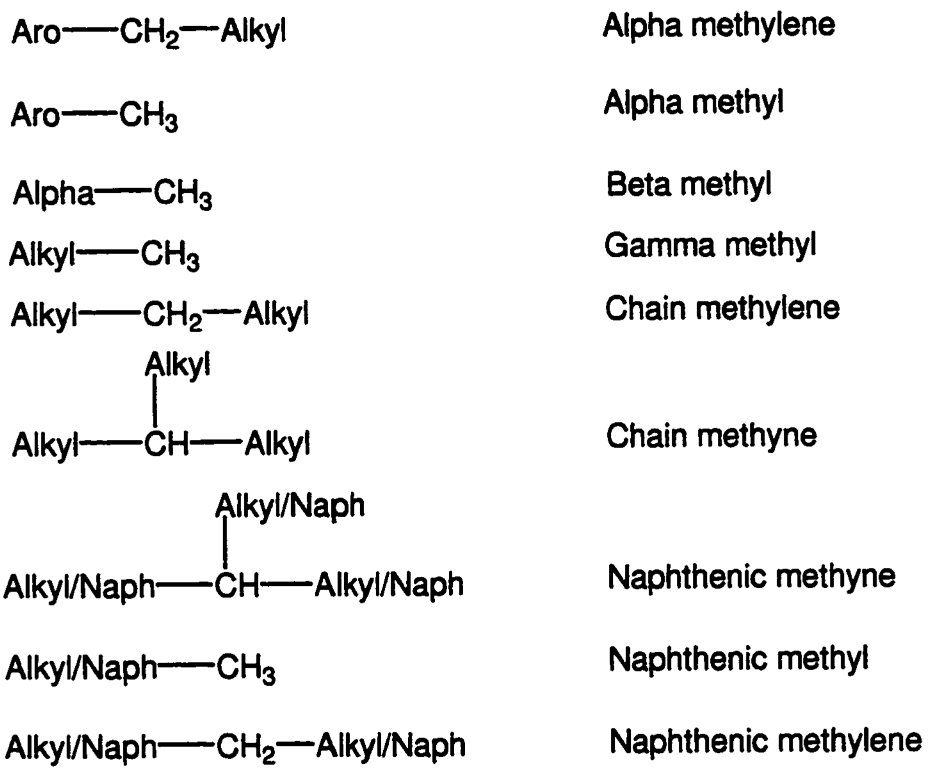


Indole N-substituded Indole

Benzofuran Benzothiophene

Quinoline Biphenyl Bridge

Figure 10: Aromatic functional groups used in the structural analysis of heavy oil catalytically hydroprocessed fractions (Gray *et al.*, 1992).



$\mathrm{Aro{-}CH_2{-}Alkyl}$	Alpha methylene
$\mathrm{Aro{-}CH_3}$	Alpha methyl
$\mathrm{Alpha{-}CH_3}$	Beta methyl
$\mathrm{Alkyl{-}CH_3}$	Gamma methyl
$\mathrm{Alkyl{-}CH_2{-}Alkyl}$	Chain methylene
$\mathrm{Alkyl{-}\overset{\overset{Alkyl}{\vert}}{CH}{-}Alkyl}$	Chain methyne
$\mathrm{Alkyl/Naph{-}\overset{\overset{Alkyl/Naph}{\vert}}{CH}{-}Alkyl/Naph}$	Naphthenic methyne
$\mathrm{Alkyl/Naph{-}CH_3}$	Naphthenic methyl
$\mathrm{Alkyl/Naph{-}CH_2{-}Alkyl/Naph}$	Naphthenic methylene

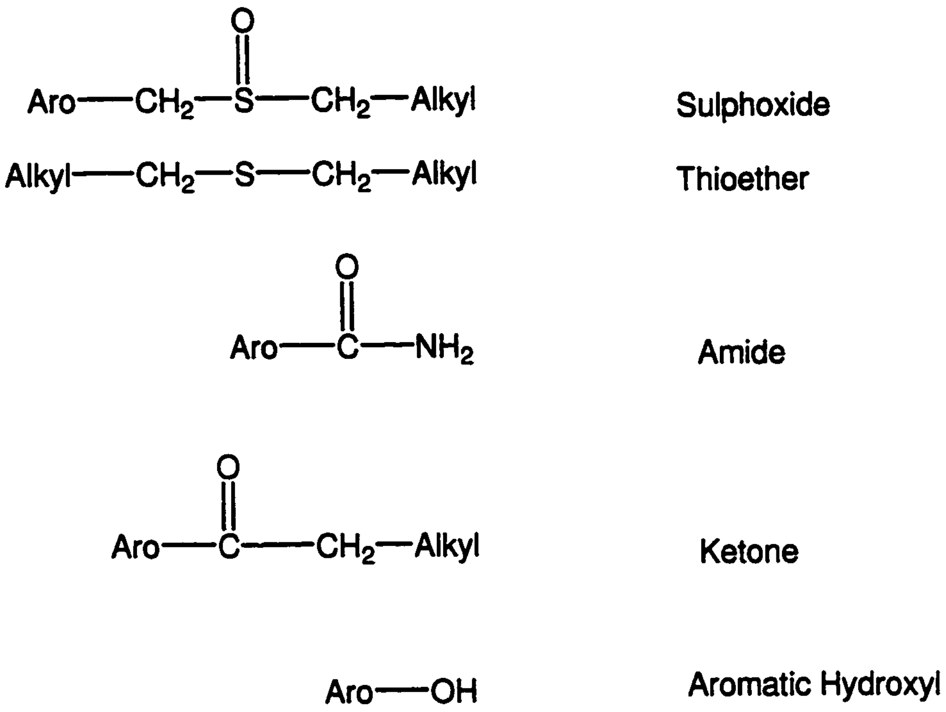
Aro-Aromatic carbon

Alkyl-Alkyl carbon

Alpha-Alpha carbon

Alkyl/Naph=Alkyl or naphthenic carbon

Figure 11: Aliphatic groups used in the structural analysis of heavy oil catalytically hydroprocessed fractions (Gray *et al.*, 1992).



$\mathrm{Aro{-}CH_2{-}\overset{\overset{\large O}{\|}}{S}{-}CH_2{-}Alkyl}$ Sulphoxide

$\mathrm{Alkyl{-}CH_2{-}S{-}CH_2{-}Alkyl}$ Thioether

$\mathrm{Aro{-}\overset{\overset{\large O}{\|}}{C}{-}NH_2}$ Amide

$\mathrm{Aro{-}\overset{\overset{\large O}{\|}}{C}{-}CH_2{-}Alkyl}$ Ketone

$\mathrm{Aro{-}OH}$ Aromatic Hydroxyl

Aro-Aromatic carbon
Alkyl-Alkyl carbon

Figure 12: Heteroatomic groups used in the structural analysis of heavy oil catalytically hydroprocessed fractions (Gray *et al.*, 1992).

In total 17 different catalytic hydrocracking runs were performed. For each of these reactions the resulting product was separated into four fractions using distillation. These four fractions are respectively, naphtha (IBP-176°C, IBP-initial boiling point), middle distillate (176-343°C), gas oil (343-524°C), and residue (524°C+). VPO determined the average molecular weight (g/mol) for each subfraction, while the detailed structural analysis determined the relative concentration of each of the structural attributes (mol structural attribute/100 g of heavy oil fraction).

Chapter 3

Modeling of Bitumen Structure and Reactions

There are several approaches to modeling bitumen structure and reactivity. Molecular based kinetics is a promising branch of kinetics that uses computer simulations to analyse chemical reactivity at the molecular level (Joshi *et al.*, 1998). The advantage of molecular based kinetics is that a single kinetic scheme can potentially be applied to many different reaction feeds and processing conditions. The kinetic predecessor to molecular based kinetics is lumped kinetics (Ayasse *et al.*, 1997). Lumped kinetic models are usually restricted to specific feeds, catalysts, and processing conditions. Unprocessed heavy oil feeds consist of molecules with molecular weights ranging from 290 to 3220 g/mol (Gray *et al.*, 1992). To apply molecular based kinetics to the study of heavy oil, heavy oil must be represented at the molecular level. Thus, there is a need for molecular representations (Khorasheh *et al.*, 1998). Molecular representations are sets of simulated molecules that have bulk properties identical to the corresponding experimentally determined ones. The structural complexity of heavy oil makes it impossible and impractical to identify and analyse individual molecules. However, it is possible to analyse the bulk properties of heavy oil. Molecular weight distributions, carbon distributions, SARA fractionation, and elemental analysis all yield collective bulk information about the structure of heavy oil. Thus, using only bulk experimental data, molecular representations create a series of unique simulated molecules.

3.1 Kinetics

3.11 Lumped Modeling

A mathematical model that describes the correlation between reactivity and reactor conditions (i.e. flow rates, reactor size, pressure, temperature) is of tremendous benefit to both academia and industry. In industry, such a model can help in the design of controllers and reactors. Strategies that can eliminate the formation of unwanted byproducts, and reduce unwanted reactions with process equipment, could also be formulated with such a model. The lumped kinetic model is currently the best model in describing the hydrocracking of heavy oil. Lumped kinetic models divide complex feeds, into lumped fractions. The interconversion of each fraction into each other fraction is usually described using first order kinetics. There are many different ways to define lumps. Boiling point, SARA, or structural based lumps can all be used in lumped kinetic models. Despite their simplicity, mass lumped kinetic models are usually limited in their usefulness. Most lumped models are only applicable to a given feed and catalyst for a given set of process conditions. If a different catalyst or feed is used, the model predictions are inaccurate.

Ayasse *et al.* (1997) used a first order lumped kinetic scheme to describe the hydrocracking of Athabasca bitumen. In the kinetic scheme (Figure 13) proposed by Ayasse *et al.* (1997) lumps were created for the residue and for the gas oil. Three different models were created. In the first, the residue was modelled as a single lump while the gas oil was modelled as two lumps (a product and a feed lump). The second model consisted of two residue lumps (a hard and a soft) and a single gas oil lump. The third model contained a single lump for both the residue and the gas oil. Product and feed gas oil lumps were allowed to be converted into three other lumped groups, middle distillate, naphtha, and gases. Residue was allowed to be converted into middle distillate, naphtha, and gas oil lumps. The middle distillate, naphtha, and gas lumps were assumed to be unreactive.

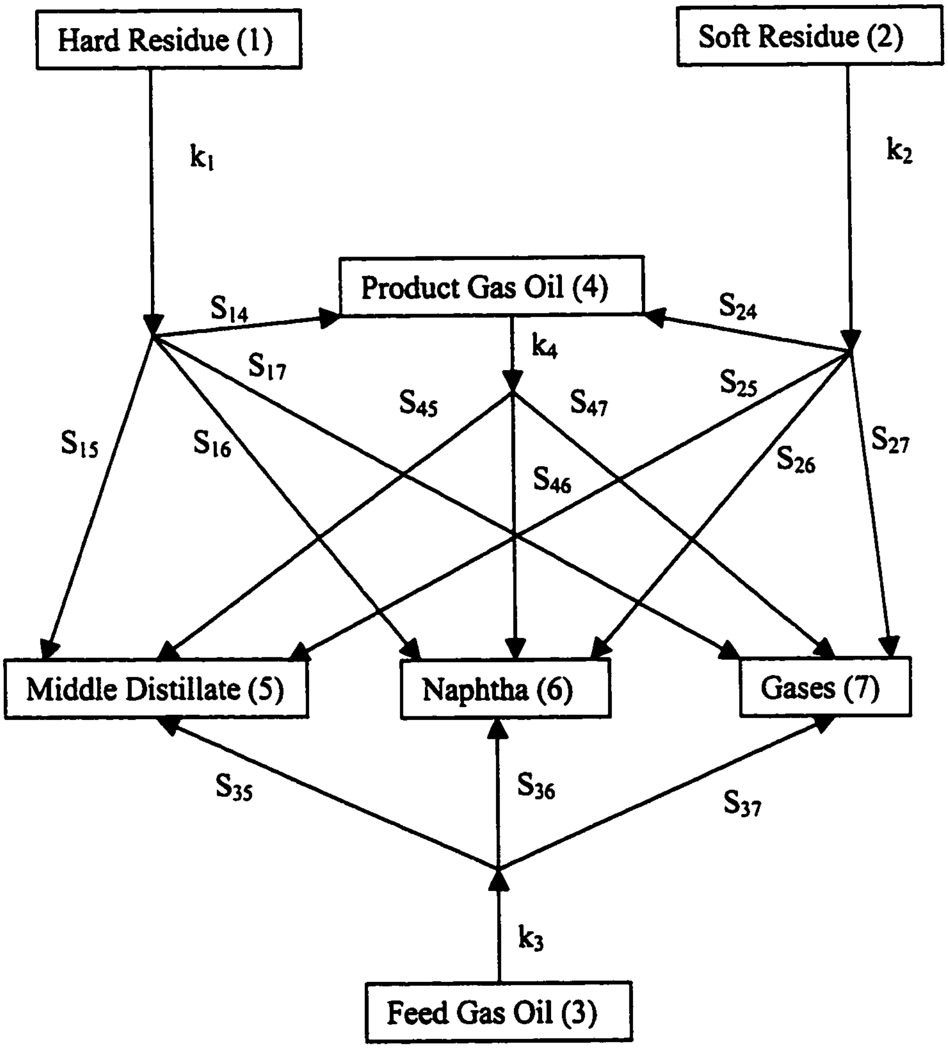


Figure 13: A mass lumped kinetic scheme used to describe the hydrocraking of bitumen (Ayasse *et al.*, 1997). Each lumped fraction is denoted by a number i. k_i represents the first order rate constant associated with lump i. S_{ij} represents a stoichiometric coefficient (describes the selectivity of lump i to be converted to lump j).

A single first order rate coefficient was associated with each gas oil and residue lump. Cracking of a lump was defined as "its disappearance to form lighter species". Cracking reactions can give multiple products, but the products are constrained because all cracked precursor appears as product. As a result a proposed kinetic network contains a maximum of four independent cracking reactions. A selectivity coefficient was used to describe the fraction of a lump that forms middle distillate, naphtha, gases, or gas oil.

Athabasca bitumen was cracked on NiO on γ-alumina catalyst at 430°C. A liquid sample was collected from each reactor stage and distilled to determine the relative compositions of the naphtha, middle distillate, gas oil, and residue fractions. CSTR (continuous stirred tank reactor) mass balances were done for the concentration of each mass lumped fraction. For each model, a nonlinear least squares optimisation routine optimised all rate constants and selectivities.

The most accurate kinetic model described by Ayasse *et al.* (1997) was a single residue, two gas oil lumped model. This model had ten independent parameters (three rate constants and seven selectivities). Data from both single and multi pass experiments were used to optimise the parameters in this model. In the single residue lump, two gas oil lumped model the middle distillate yield was over-predicted and the naphtha yield was under-predicted in experiments that used bitumen as a feed. In the multipass experiments this model fitted the multipass data very well except for the fourth and final stage, which had the least experimental reliability.

The stoichiometric coefficients calculated in the three models were found to be significantly model dependent. This indicates that they should not be interpreted in terms of chemistry. In all models, the gas oil fractions were found to be less reactive than the residue fractions. This is consistent with the experimental observation that reactivity is proportional to fraction molecular weight. Two models indicated that the gas oil fraction did not produce naphtha.

All of the models indicated that more light ends were produced from the gas oils than from the residues. These observations are all consistent with experimental findings.

3.1.2 Molecular Kinetic Simulation

A very promising successor of lumped based bitumen kinetic models is the molecular based kinetic model. Molecular based models are based on fundamental chemical and physical principles. Because these models are fundamentally based, they are very versatile. They can describe the reactivity of a variety of oil feeds under various reaction conditions. Typically molecular based kinetic simulations require a molecular representation of an oil feed as a starting input. Molecular based simulations react the molecules in the feed representation according to a predefined kinetic framework.

The first documented heavy oil molecular kinetic simulation utilised asphaltene representations (Neurock *et al.*, 1989). Psuedo-first order rate constants were determined for alkyl side chain thermolysis, aromatic hydrogenation, and naphthenic ring dehydrogenation. p-electron density was calculated for each aromatic core in the asphaltene molecules using both the ZINDO (Zerner's modified intermediate neglect of differential orbital overlap) and Huckel's method. For the hydrogenation of the aromatic cores, the rate constants were determined using a logarithmic correlation that related the first order hydrogenation rate constant to electron density. The first order rate constants were used to calculate the relative probabilities of reaction. Reactions were allowed to happen randomly according to their relative probabilities. The simulation was carried out simulating 10,000 s of reaction time at 400°C. The hydroprocessing reactions were only allowed to occur for a limited time. Due to hydrogenation, the molecular weight/unit sheet initially increased. Over time, due to aliphatic cracking, the molecular weight/unit sheet decreased. This evidence indicated that at shorter times a faster rate of hydrogenation initially dominated, while at later times the cracking of aliphatic chains was dominant.

Joshi *et al.* (1998) used a molecular representation algorithm in creating a fluid catalytic cracking model. A molecular representation for a gas oil that contained 34 different molecules each with a corresponding mole fraction. Each molecule was represented by a connectivity matrix that described the connectivity of each atom in a given molecule. Nine different reactions were allowed to occur: proteolytic cleavage to hydrocarbons, proteolytic cleavage to hydrogen, hydride abstraction, β-scission, hydride or methyl shift, PCP isomeration, proteonation or deprotonation, coking, and catalyst deactivation. Associated with each of these reactions was a transformation matrix. Each transformation matrix acted on a modified connection matrix, converting a reactant into a product. A modified connection matrix only contains connection information from atoms involved in a reaction.

For each type of reaction, a detailed set of rules was made that specifies the types of molecules that could react by the reaction. Both coking and catalyst deactivation were allowed to occur. The Evans Polyani relationship was used to calculate the rate constant for a given reaction. Literature values were used for the Arrhenius factors. The activation energies were a model parameter. The heat of reaction for each reaction was independently calculated. The rate of coke formation was assumed to be constant and the degree of catalyst deactivation was assumed to be proportional to the total amount of coke produced. Constraints were applied to the number of allowable isomerization, methyl shift, and hydride shift reactions allowed. The maximum carbon number for a carbenium ion that can abstract a hydride ion was also constrained. These constraints stopped the reaction simulation at some point. Without them, the simulation would have run indefinitely.

Each kinetic simulation was allowed to run to completion. Nine reaction activation energies, along with a catalyst dependent parameter, a coking rate catalyst, and a catalyst deactivation parameter were optimised. After all these parameters were optimised, a very good agreement between the predicted and experimental compositions of various cracked gas oil components was observed. In total, 450 molecular species were produced that resulted from 5500 different reactions.

3.2 Molecular Representations

The structure of heavy oil has been characterised and studied using many different analytical techniques. NMR, VPO, mass spectroscopy, and elemental analysis, have all yielded insights into the structure of heavy oil. However, very few studies have linked together the analytical information derived from these techniques to create an overall global representation of structures present in heavy oil. Heavy oil molecular representations require analytical information from a wide range of analytical experimental techniques. Molecular representations must correctly model many parameters including molecular mass based on VPO, carbon type concentrations from NMR, and elemental concentrations. Heavy oil molecular representations are unique in that they represent the final collective result from many different experimental analyses, but represent the starting point for molecular based heavy oil kinetics. The input for molecular based heavy oil kinetic simulators is heavy oil molecular representations. Also by providing insights into the overall structure of heavy oil, more will be known about the reactivity of heavy oil.

All petroleum molecular representations have a common objective. This objective is to visually and/or mathematically represent a petroleum mixture in a manner that is consistent with available analytical data. To accomplish this, many routes can be taken. The simplest molecular representation type is the lumped single molecule approach (Yen, 1972). This involves creating a single molecule to represent an entire fraction. A more complicated but more accurate approach is the molecular set approach. This involves creating a series of molecules to represent a single fraction. The molecular set approach usually involves Monte Carlo simulation.

All molecular representations require a framework on which to represent molecules. The simplest representation framework is the lumped attribute method. When the lumped method is used, each molecule is represented as a specific number of attributes (Campbell and Klein, 1997). The confirmation and structure are not specified but the number of each attribute is defined. A much more specific approach is the matrix approach (Khorasheh *et al.*, 1998). By using a series of matrices, the entire structure of a molecule is mapped out atom by atom. A hybrid of the lumped and matrix approach is the vector approach (Quann and Jaffe, 1992). Similar to the lumped approach, a petroleum molecule is represented using a series of chemical attributes. The attributes are all linked together by a series of vectors and/or matrices.

3.2.1 Coal Molecular Representations

Coal molecular representations were completed long before heavy oil molecular representations were first attempted. At the time that coal molecular representations were first attempted, very little heavy oil analytical information existed. In general, heavy oil molecules have molecular weights that are much larger than coal molecules. Creating larger molecules requires more computing power and more computer memory. It is arguable that in the mid 1970s when coal molecular representations were first being created, that computers were not powerful enough to create heavy oil molecular representations. The coal molecular representations showed that it was indeed possible to create molecular representations for complex hydrocarbon fractions.

3.2.1.1 Graphical Based Coal Molecular Representations

A graphical approach was used by Bartle *et al.* (1975) to create average molecular representations for coal subfractions. A coal supercritical-gas extract was further separated by solvent extraction, silica-gel and gel-permeation chromatography. The soluble fractions were analysed using a variety of analytical techniques including ^{1}H-NMR, ^{13}C-NMR, and infrared mass spectrometry. The ^{1}H-NMR and ^{13}C-NMR spectral data allowed three important average aromatic molecular content values to be calculated. These are respectively number of ring joining carbon atoms (C_J), number of aromatic ring carbons (C_R), and the number of ring hydrogen atoms in the unsubstituted aromatic nucleus (H_R). For each soluble fraction and for a number of model representative aromatic compounds, these three parameters were plotted in three dimensions. The representative molecules (Figure 14) contained various numbers of aromatic and naphthenic rings. These compounds did not contain substituted groups or heteroatoms. For each fraction the representative structure with C_J, C_R ,and H_R, values closest to the average analytically calculated values for the fraction, was determined. As would be expected, exact matches were not observed. Average molecules were created for each fraction by incorporating oxygen, NH, sulphur groups, and alkyl chains. For high molecular weight fractions, a single average molecule was adequate to represent the required structural features. However, for low molecular weight fractions a series of molecules best represented the fraction.

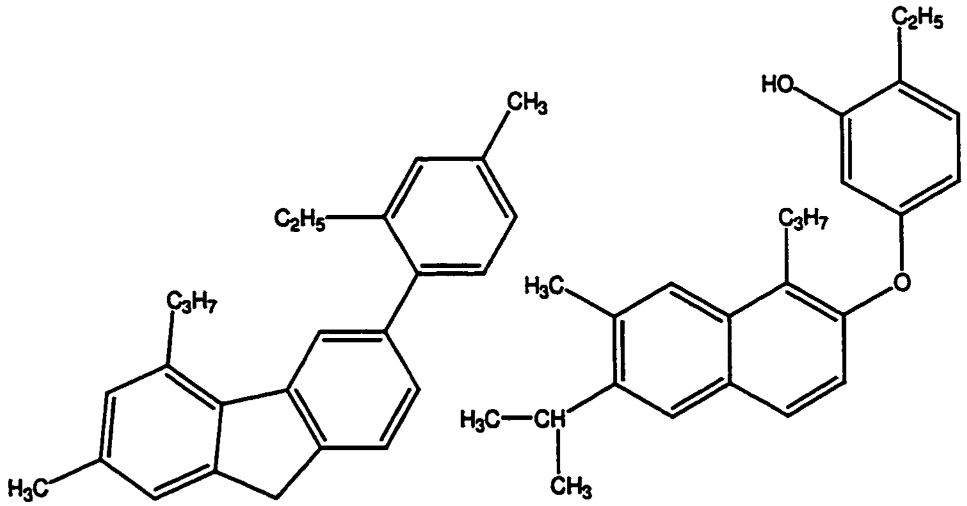


Figure 14: Representative molecules from a supercritical-gas extract of coal at 350°C (Bartle *et al.*, 1975).

3.2.1.2 A Computer Assisted Coal Molecular Representation

More detailed and complex coal molecular representations were created by Oka *et al.* (1976) who created their molecular representations using a computer algorithm. This algorithm, termed CAMSC (computer assisted molecular structure construction), utilized data from NMR, elemental analysis, and molecular weight data. This representation was the first molecular based representation created by a computer algorithm. Seven experimentally determined parameters were necessary to run this algorithm. These 7 parameters are α hydrogen (hydrogen on carbon α to aromatic rings) content, β hydrogen (hydrogen on carbon β to aromatic rings) content, aromatic hydrogen content, aromatic carbon content, aliphatic hydrogen content, phenolic oxygen content, and ether oxygen content. The parameters were either calculated from the analysis of isolated and purified compounds or from coal fractions. The framework for the CAMSC algorithm is 10 aromatic cluster groups, 11 naphthenic rings, and 13 aliphatic carbon groups (24 attributes in total). Permutations of the 24 attributes form individual coal molecules. CAMSC determines all molecules (permutations of the 24 attributes) that are consistent with the input parameters. This algorithm was used both to propose structures (Figure 15) for a purified compound found in a coal extract and to create an average molecule representative of a coal distillate.

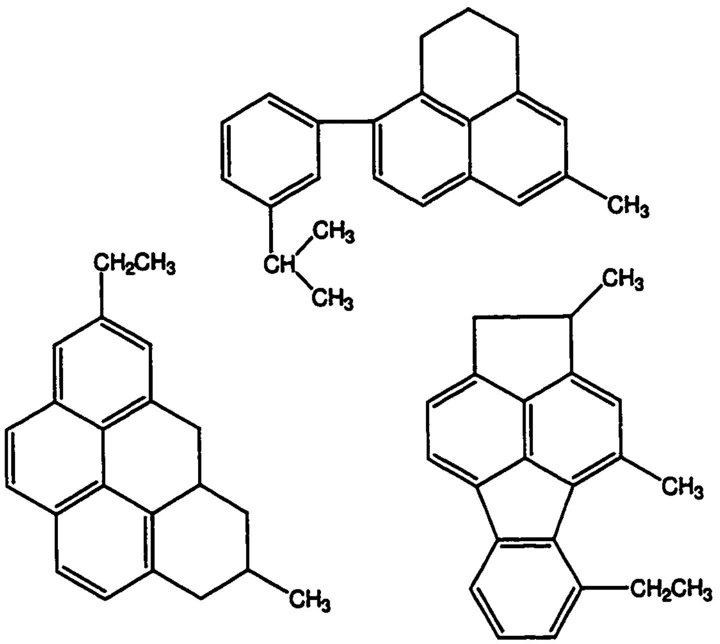


Figure 15: Representative coal structures created by the CAMSC algorithm (Oka *et al.*, 1976).

3.2.2 Structural Group Analysis

The structural analysis of heavy oil began in the late 1950's with the work of Montgomery and Boyd (1959). Before their studies very little was known about the structure of Athabasca bitumen. Essentially, it was only known that Athabasca bitumen contained aromatic, paraffinic, and naphthenic, secondary structures. The relative abundance of each carbon type in bitumen was not known in the late 1950's. Using very limited data, Montgomery and Boyd (1959) used mathematical analysis to calculate the relative distribution of various carbon contents in Athabasca bitumen. Montgomery and Boyd (1959) in essence created the first heavy oil structural model.

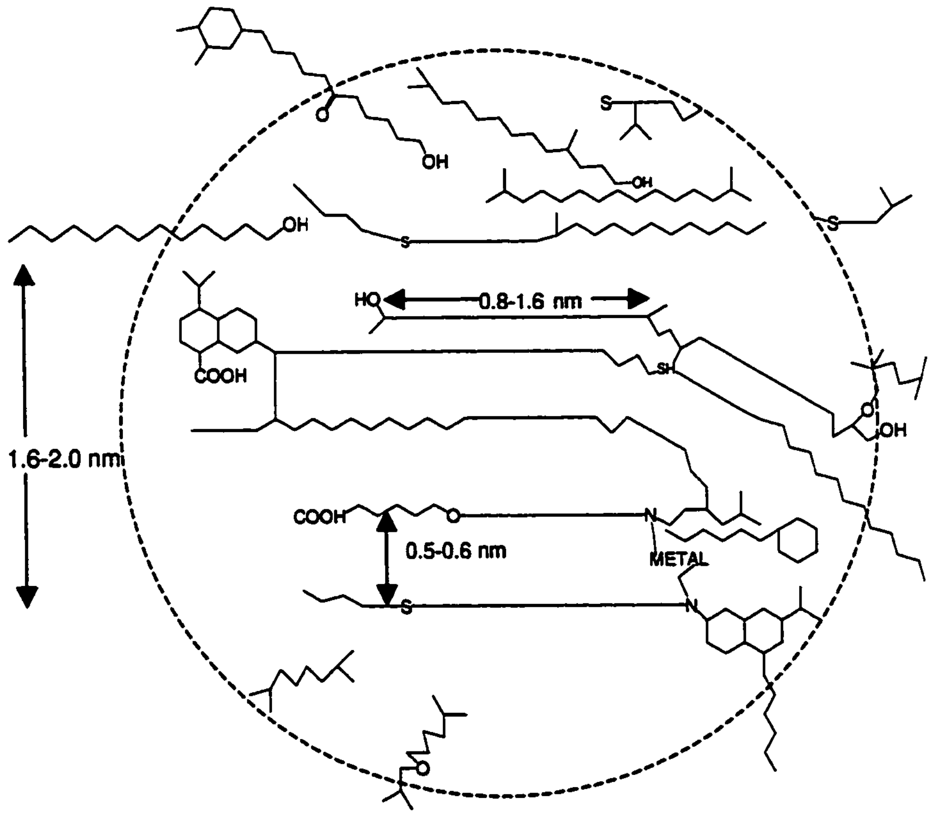
Boyd and Montgomery (1962) and Sawatzky *et al.* (1967) used molar volume and refraction data to analyse the structures present in Athabasca asphaltene. At the time of these studies, NMR analysis of heavy oils was rather crude and not very reliable for the quantification of specific chemical groups. Seventy model hydrocarbons were used to formulate correlations (Montgomery and Boyd, 1962) that related molar volume and refraction to five different carbon contents. The five carbon types analysed by Montgomery and Boyd were aliphatic carbons (C1), naphthenic carbons not involved in ring junctions (C2), naphthenic ring junction carbons (C3), non-ring junction aromatic carbons (C4), and ring junction aromatic carbons (C5). In total there were five equations (two correlations, three balances), with five unknown carbon content values. The carbon content ranges were determined by solving the equations simultaneously using a computer. Thus, by determining the molar volume and refraction of asphaltene, the contents of these various carbon types can be determined. Three balance equations were created, a total carbon balance, a total hydrogen balance, and an aromatic carbon balance. However, at the time of these studies (Sawatzky *et al.*, 1967) it was not possible to estimate the overall aromatic content accurately. Because of this, the five equations were solved for various aromaticity fractions. The fraction of aromatic carbon was fixed at various arbitrary values. For each value the five simultaneous equations were solved. In an earlier study (Boyd and Montgomery, 1962) instead of using arbitrary values of aromaticity, four different procedures were used to estimate the aromaticity of Athabasca bitumen. These methods were IR absorption, the densimetric method, the densimetric-molecular weight method, and the graphical-densimetric method. No agreement existed among the results from the different methods. The aromaticity (fraction of aromatic carbon) ranged between 0.20 and 0.60. Only aromatic content values that resulted in all other carbon content values being positive were considered. The maximum and minimum carbon content values (these form a range) were found within the resulting solution set. The following carbon content ranges were determined by Sawatzky *et al.* (1967); C1: 33.8-34.2%, C2: 0-6.5%, C3: 10.3-0%, C4: 45.2-41.2%; C5: 11.2-18%. The degree of aromatic condensation reported by this study is consistent with that determined by NMR analysis in a much latter study (Ignasiak *et al.*, 1977). The literature (Ignasiak *et al.*, 1977) reports that even though Montgomery and Boyd had limited data and used rather crude techniques, they produced very good results.

3.2.3 Heavy Oil Molecular Representations

3.2.3.1 Schematic Representations

It was not until the late 1980s that mathematical representations of heavy oil fractions were created. This was mainly due to the difficulty associated with developing quantitative methods that were capable of analysing heavy oil. Before the development of mathematical representations, schematic diagrams were used to represent the structures and molecules present in heavy oil. The schematic representations were mainly single molecules that were constructed to be consistent with current analytical information. Based predominantly on x-ray diffraction data, Yen proposed numerous schematic representations of asphaltenes that represented the secondary and tertiary structures associated with asphaltenes (Figure 16).

Based upon NMR, x-ray diffraction, and ring compactness data, Yen (1972) proposed an average model structure for Langunillas oil (Figure 17). This model structure had an empirical formula $C_{74}H_{87}NS_2O$. The average structure was constructed so that it was consistent with data from proton NMR and compactness studies. The α, aromatic, naphthneic, methylene, and methyl, proton contents of the structure were very close to the experimentally determined values. This asphaltene molecular representation is a very historically significant molecular representation. Most importantly, it was the first asphaltene molecular representation based upon analytical data. Secondly, it summarises about 13 years of pioneering analytical work in a single molecule. Even in the 1990s this representation and its derivatives were used to graphically illustrate the molecules that compose asphaltenes.



Zig-zag lines represent saturated sheets or naphthenic rings; the straight lines represent the edge of flat sheets of condensed aromatic rings

Figure 16: A schematic representation of an asphaltene micelle (Yen, 1974).

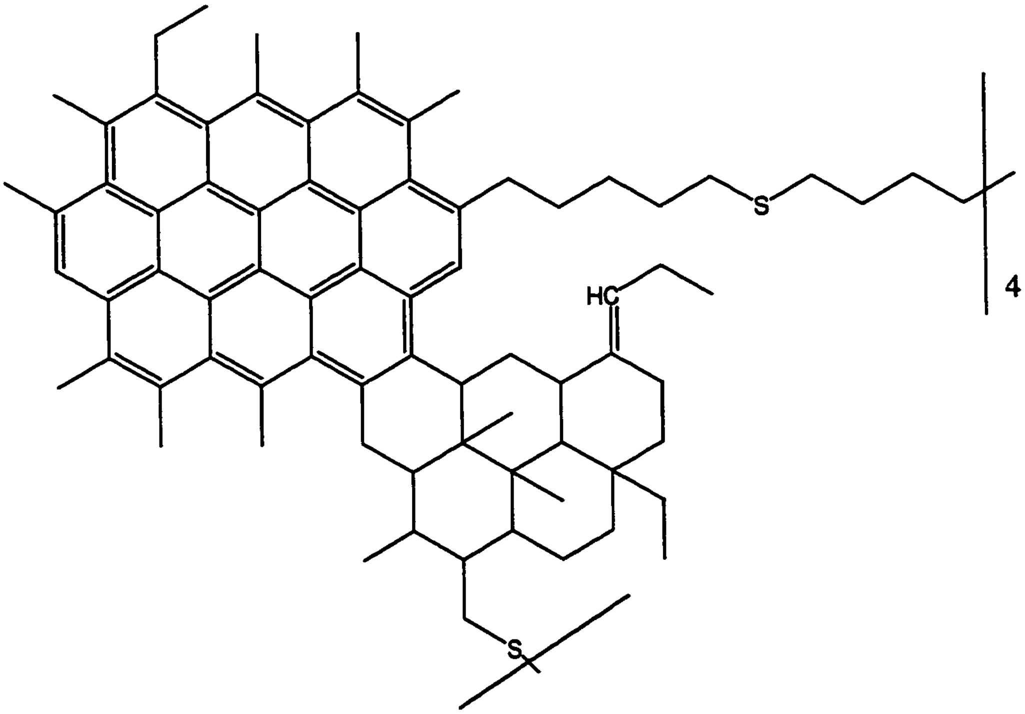


Figure 17: Asphaltene "tile" representation of Yen (1972).

In the late 1970s (Ignasiuk *et al.*, 1977) it became evident that pericondensed aromatic rings are not prevalent in asphaltenes. Ignasiuk *et al.* proposed an average model asphaltene molecule that was consistent with ^{13}C-NMR, ^{1}H-NMR, and elemental analysis, data (Figure 18). This average molecule was an octamer. Each repeating unit had a sulphide linkage, an alkyl chain, three naphthenic rings, and a naphthene ring system with a hydroxyl group. The analytical data that Ignasiuk *et al.* (1977) used did not support the possibility that asphaltene aromatics are pericondensed.

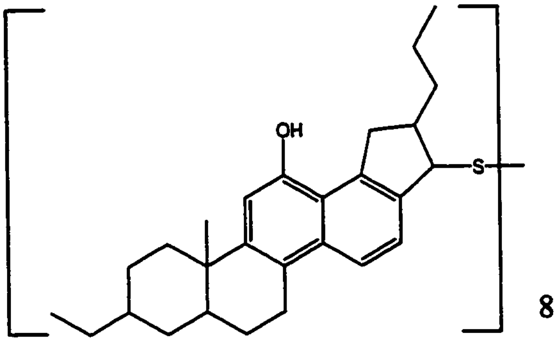


Figure 18: A schematic asphaltene representation that does not contain pericondensed aromatic rings (Ignasiuk *et al.*, 1977).

Despite the initial work of Strausz, Speight and Moschopedis (1979) described asphaltene molecules as condensed polynuclear systems with alkyl side chains (Figure 19). In 1979 it was generally believed that asphaltenes contained between 6 and 15 aromatic rings distributed between one and three highly condensed aromatic cores. This was predominately due to the work of Yen. For 12 years, Yen published a series of papers documenting the pericondensed nature of the aromatic systems present in asphaltene. It was not until the late 1970s that any other research group began challenging the pericondensed nature of asphaltene aromatics. The Yen pericondensed asphaltene model has had a significant influence on heavy oil molecular representations. By the late 1980s RICO and other experimental techniques indicated that pericondensed aromatics are not common in asphaltenes. Despite this, the majority of heavy oil and asphaltene molecular representations in the 1980s and 1990s contain pericondensed aromatic structures.

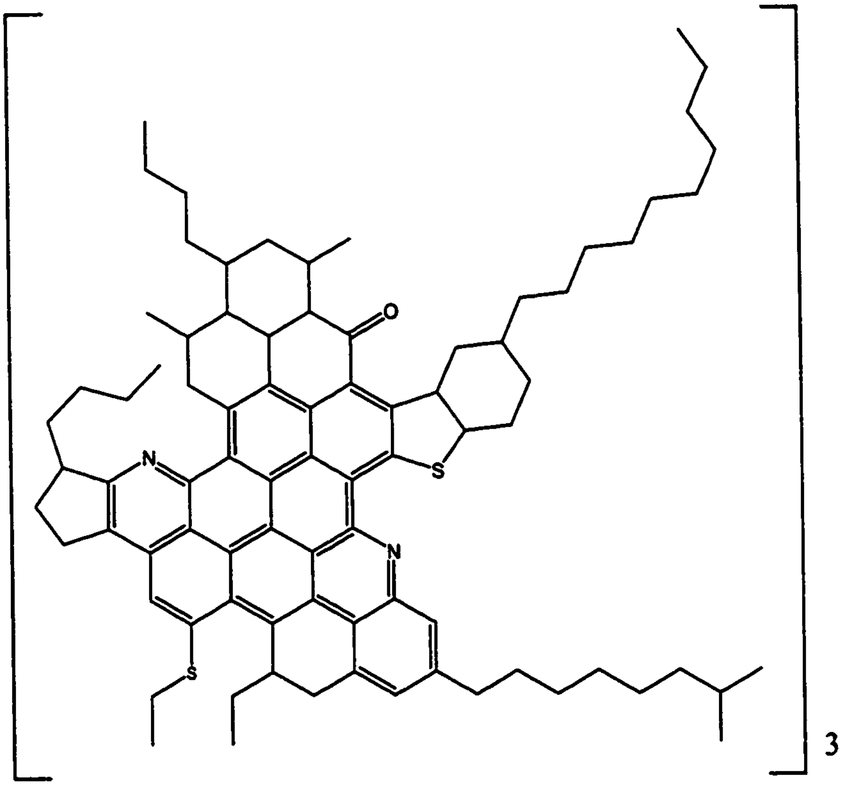


Figure 19: A schematic representation of a Venezuelan crude oil asphaltene (Speight and Moschopedis, 1979).

3.2.3.2 Mathematically Based Representations

The first mathematically based heavy oil molecular representation was proposed by Neurock *et al.* (1989). This molecular representation (Figure 20) was based on data from ^{1}H-NMR, elemental analysis, and VPO. Using balance equations, many structural parameters were calculated for an unidentified asphaltene feed. These parameters included aromatic ring content, naphthenic ring content, number of sidechains per molecule, length of sidechains, number of fused rings in an unit sheet, number of unit sheets (a sheet of many aromatic rings) per molecule, and the degree of substitution on aromatic and naphthenic rings. Normal probability distributions were fitted to each of these attributes. Asphaltene molecules were constructed by the sequential random sampling of each probability distribution. Constraints were applied to each distribution so that negative values were not sampled. Ten thousand molecules were constructed to represent an asphaltene feed. The resulting molecules were later used in a molecular kinetic simulation of hydrogenation, dehydrogenation, and thermolysis reactions.

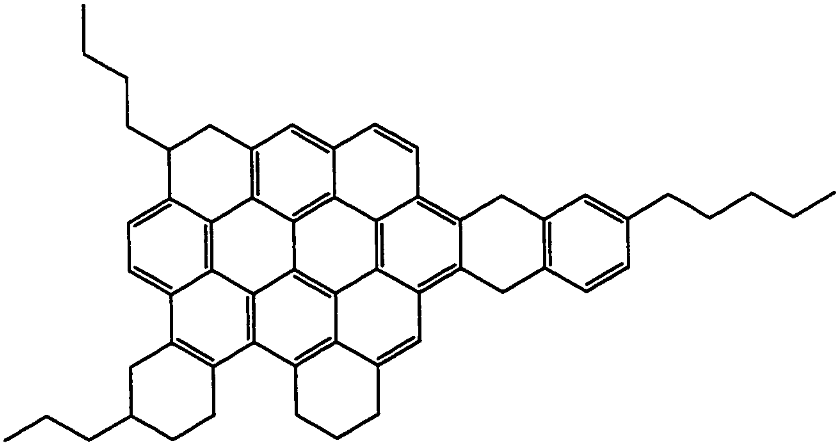
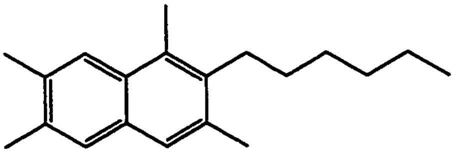


Figure 20: An asphaltene molecule created by using a Monte Carlo based algorithm (Neurock *et al.*, 1988).

3.2.3.2.1 Vector Based Representations

Quann and Jaffe (1992) created a vector-based framework to represent hydrocarbon molecules (Figure 21). Although their framework is only applicable to molecules containing single aromatic clusters, it is very efficient in describing small molecules. A single molecule is represented by a row vector. Each column in a vector corresponds to a specific chemical attribute. The attributes used by Quann and Jaffe (1992) included benzene rings, naphthenic rings, thiophenic rings, and paraffinic carbons. A structural vector also indicates the number of paraffinic chains and branch points. Although a structural vector specifies the exact chemical composition of a molecule, the exact chemical arrangement within a molecule is not specified. Many different possible atomic arrangements might be consistent with a given structural vector.

It is very straightforward to implement chemical reactions with the vector molecules created by Quann and Jaffe (1992). Aromatic saturation, aromatic opening, paraffin isomerization, and paraffin cracking, reactions can all be applied to molecules represented by structural vector. For a molecule to participate in a given reaction it must meet certain structural criteria. These criteria specify certain chemical properties that a molecule must have for it to be allowed to participate in a given reaction. When a reaction is allowed to occur, a reaction vector is added to the structural vector. The addition of the reaction vector changes the number of each attribute either participating or being formed in the reaction.



A6	A4	A2	N6	N5	N4	N3	N2	N1	R	br	me	IH	AA	NS	RS	AN	NN	RN	NO	RO	KO
1	1	0	0	0	0	0	0	0	10	0	4	0	0	0	0	0	0	0	0	0	0

Figure 21: A petroleum molecule represented using a single structural vector (Quann and Jaffe, 1992). A6 represents a six carbon aromatic ring. A4 represents a four carbon aromatic ring implement that is attached to A6. R represents the total number of alkyl carbons associated with the aromatic or aliphatic compounds.

3.2.3.2.2 Monte Carlo Based Representations

Neurock *et al.* (1994) created a Monte Carlo based algorithm to create molecular representations for a California asphaltene and a Kern River heavy oil (Figure 22). Probability distributions were separately applied to structural attributes that represented each oil. For the California asphaltene these attributes were number of US per molecule, number of aromatic rings per US, number of naphthenic rings per US, degree of aromatic carbon substitution, degree of naphthenic carbon substitution, and alkyl chain length. The Kern River heavy oil was represented as having two components, an asphaltene component, and a volatile component. The asphaltenes were represented by the same structural attributes that were used for the California asphaltene. The volatile component was further expressed in terms of separate paraffin, naphthenic, and aromatic molecules. Paraffins were only described by paraffin carbon number. Naphthenics were described by the number of naphthenic rings/naphthene molecule, percent of naphthene carbons subsituted, and substituted carbon chain length. The aromatic molecules were described by the number of aromatic rings/aromatic molecule, number of naphthene substituents/aromatic cluster, percent of aromatic carbons substituted, and substituted carbon chain length.

For both the California asphaltene and a Kern River heavy oil, either normal or gamma distributions were applied to each attribute. Probability distribution functions (pdf) were applied so that 98% of each attribute fell between maximum and minimum values reported in the literature. For both oils, 10,000 molecules were created. For the California asphaltene the average predicted molecular weight (2770-normal pdf, 2726-gamma pdf) only deviated slightly from the experimentally determined value (2683). A simulated distillation curve was created based upon the heavy oil representation. A very good fit was observed between the experimental and simulated heavy oil boiling curves (Neurock *et al.*, 1994).

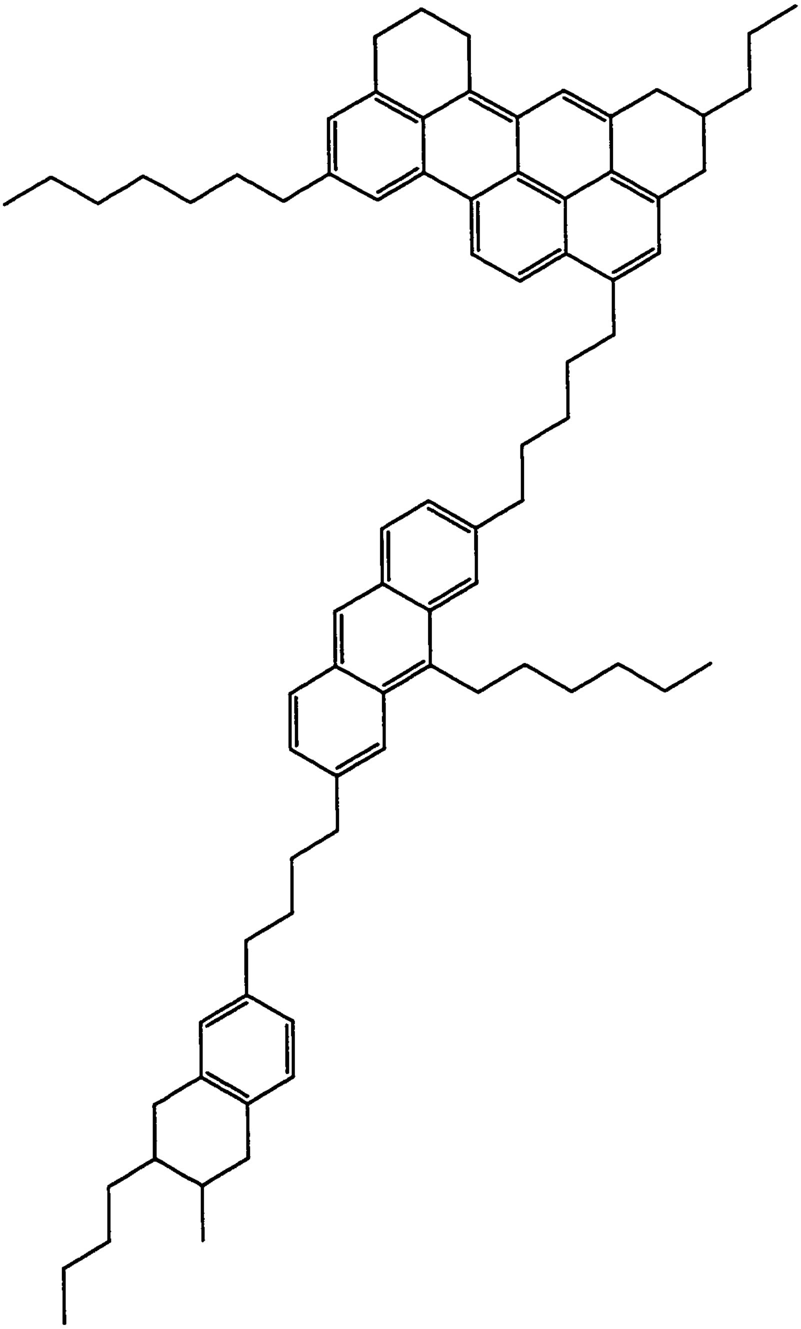


Figure 22: An asphaltene molecule created by using a stochastic algorithm (Neurock *et al.*, 1994).

Campbell and Klein (1997) were the first to create molecular representations that were based on nonunique sets of molecules (Figure 23). They viewed heavy oil molecules as a collection of molecular attributes. Aromatics were described by their number of benzene, thiophenic, and naphthenic rings; and number and length of sidechains. Parraffins were described by chain length. Naphthenes were described by sidechain number and length, and ring number. Asphaltenes were described by the number of US they contained. A gamma distribution was applied to each attribute. Campbell and Klein (1997) indicate that the gamma distribution is very effective in approximating the boiling distribution curves of petroleum residues. Since a strong correlation exists between boiling point and carbon number of a molecule, it was therefore reasonable to model the carbon number distribution of an oil fraction with the gamma distribution. Attributes such as aromatic rings, are also implicitly related to carbon number, and should also be modelled well by the gamma distribution. A least squares objective function was created that took into account a wide range of experimental data. Molecular weight, C and H concentrations, SARA separation, α and aromatic proton concentration, and boiling distribution were included in the objective function. For each property, the difference between the analytical value and the predicted value from a set of molecules was calculated. The difference for each property was divided by the standard deviation of the analytically measured property (a weighting factor), and the result squared. For a given set of molecules, the value of the objective function was the sum of the squared weighted differences of all properties.

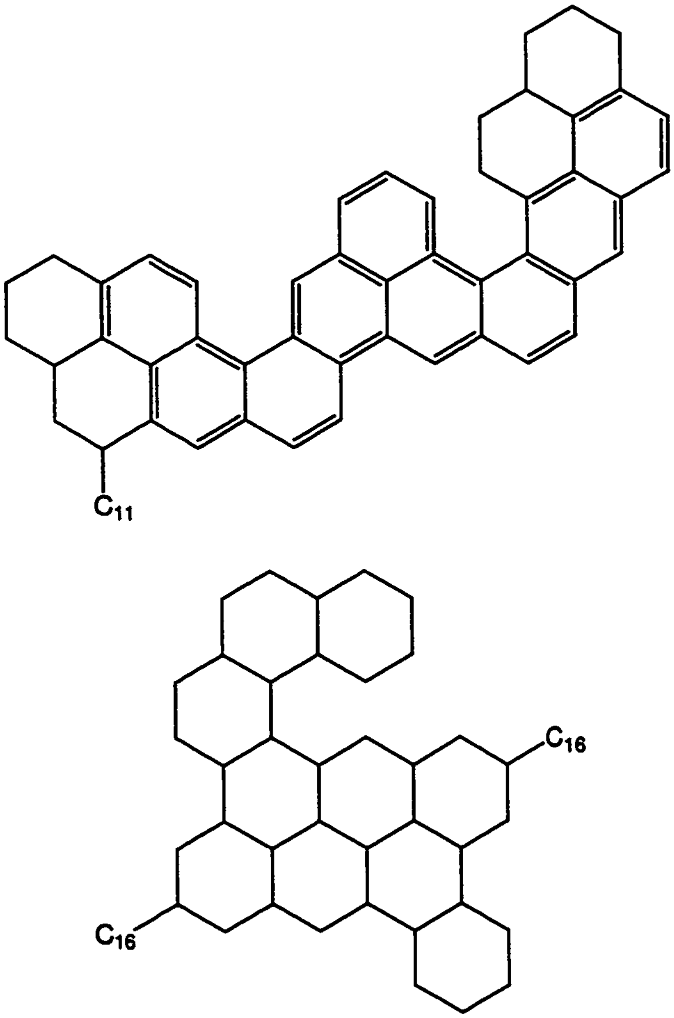


Figure 23: Two representative molecules from a Maya Monte Carlo representation proposed by Campbell and Klein (1997).

By using a global annealing technique, the optimal gamma distribution parameters for each attribute were found. This involved an iterative approach where, 100,000 molecules would be produced, the objective function calculated for the current set of molecules, then the parameters for each gamma distribution were changed. This continued until the objective function was minimised. In order to pursue future reaction studies, molecular representations containing a small number of molecules were desired. The gamma distribution for each attribute was divided into two (one node), three (two nodes), or four (three nodes). Each region of each distribution was randomly sampled. This is known as quadrature sampling. This insures that each pdf is sampled over a wide range. For each type of molecule (aromatic, naphthenic, paraffinic, asphaltene) the resulting molecules are formed by all combinations of the attribute values. Naphthenics are composed of three attribute values. If each naphthenic attribute were divided into two sampling regions, 8 different molecules would result.

Klein applied his quadrature-Monte Carlo algorithm to an Arabian light residue, a Maya residue, and a naphtha. For each oil fraction about 40 molecules were generated. A mole fraction was assigned to each molecule in each fraction. Using an optimisation routine, each mole fraction was adjusted to minimise the objective function for each oil fraction. For the two residues the predicted boiling point distributions and the H/C ratios matched well with the experimental data. Conversely, large deviations were observed for average molecular weight, the ^{1}H-NMR data, and sulphur content.

The final naphtha representation contained approximately 20 molecules (about half the initial molecules were assigned "0" mole fractions). Standard experimental techniques have identified many molecules contained in naphtha. Klein compared the molecules that his algorithm predicted to those that have been identified in naphtha. Most of the molecules that his algorithm generated have been identified in naphtha. However, his predicted paraffin distribution differed significantly with the experimentally determined distribution. Except for the paraffin fraction, the quadrature representation very closely resembled the experimentally determined composition of each PINA (paraffins, isoparaffins, naphthenics, and aromatics) fraction. The simulated boiling curve associated with the naphtha computer generated molecules closely resembled the experimental boiling point curve.

3.3 A New Type of Structural Analysis And Its Application To Molecular Representations

By the mid 1980's ^{13}C-NMR, ^{1}H-NMR, and elemental analysis, had become standard techniques, in the analysis of heavy oil. These techniques are able to quantify the concentrations of specific atoms or carbon groups in a hydrocarbon sample. However, they are unable to determine the concentrations of secondary groups or the distribution within a specific carbon type. For example, the aromaticity of a heavy oil can be determined, but the concentration of benzene rings or aromatic bridges within the oil was unknown. Also, the naphthenic content of a heavy oil could be determined, but the distribution among the various naphthenic carbon types was unknown. Knowledge of the relative distribution of various primary and secondary structure elements would be very helpful in creating chemical representations of heavy oils. Using molar volume and refraction data, Montgomery and Boyd (1962) were able to use structural analysis to gain insights into the primary structure of heavy oil. With the vastly improved computers, analytical techniques, and numerical methods, that existed in the 1980's a new type of structural analysis became possible.

Khorasheh *et al.* (1998) created a molecular representation for a synthetic Alberta crude oil with an average molecular weight of 370 (Figure 24). Their molecular representation is based upon chemical attribute concentrations that were determined using structural group analysis (SGA). SGA involves representing an oil fraction using a limited number of chemical attributes including aromatic rings, naphthenic rings, aliphatic chains, and heteroatomic structures. Khorashesh *et al.* created linear constraint balance equations based on ^{1}H data and an objective function based on ^{13}C data. The concentrations of various chemical attributes were optimised to minimise the objective function while satisfying the linear constraints.

Khorasheh *et al.* (1998) represented each molecule using a connectivity matrix and two adjuvant arrays (atom identification and permittable connection). Each connectivity matrix documented the connection of all atoms in a molecule. The SGA attribute concentrations were multiplied by the average molecular weight of the oil to express them in terms of mol attribute/avg. molecule. The structural groups were randomly chosen and connected to each other (when permissible) in the order in which they were chosen. Construction of a molecule was completed when either no more connections were possible or when all of the attributes had been accounted for. If the molecular weight of a molecule did not fall within a certain range, it was rejected. 600 individual molecules were generated for the synthetic Alberta crude oil. Averaged over the 600 molecules an average error of 5.6% was observed between the SGA calculated concentrations and the predicted attribute concentrations from the molecule set.

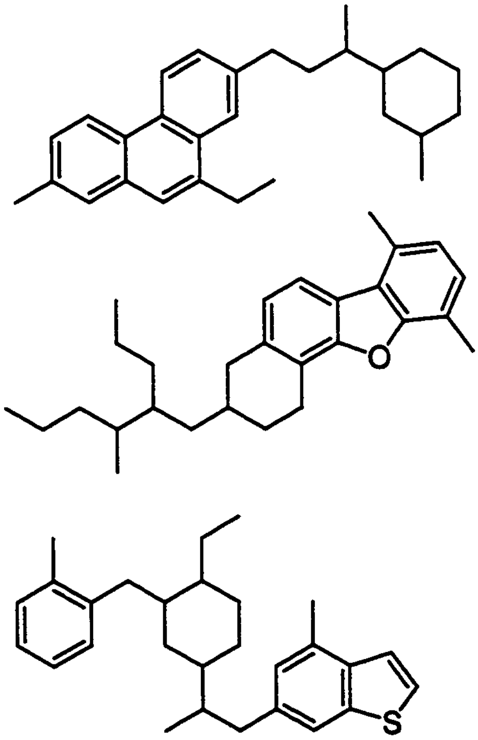


Figure 24: Molecules from a gas oil molecular representation (Khorasheh *et al.*, 1998).

^1H and ^{13}C-NMR form the foundation that SGA procedure is based on (Khorasheh *et al.*, 1987). Specific carbon types including chain methylene, chain methyne, alpha methyl, aromatic, ketones, and naphthenics, are assigned to specific peaks in both types of spectra. A digitizer was used to quantify the area corresponding to each peak. It is assumed that the variance associated with each NMR band is proportional to the area of the peak. The peak areas are used to assign weights to each NMR peak. These relative weightings appear in a quadratic objective function based on the ^{13}C-NMR carbon balance. An adjuvant ^1H balance equation and constraint equations were also created. One constraint specified that the concentrations of all attributes must be positive. Another constraint specified that chain terminating carbon groups could not exceed the concentration of chain forming carbon groups. The last constraint ensured that the number of aromatic sites was at least equal to the number of aromatic substituents. The ^{13}C-NMR objective function was minimised using Wolfe's algorithm for quadratic programming. Minimisation was achieved by adjusting the concentrations of all the proposed structural attributes.

Both the accuracy and the reproducibility of SGA were tested. Mixtures of known organic compounds were created and analysed using SGA. Structures in these compounds included aromatics, saturates, and aliphatics. The predicted SGA concentrations of the various attribute concentrations were reported to be on average within 20% of known concentrations. This error was due to uncertainties and error in interpreting the NMR spectra. SGA analysis was carried out on three separate solubility fractions (saturates, aromatics, and resins) that formed an Alberta gas oil. It was found that on average the attribute concentrations calculated from the three separate fractions were within 20% of the concentrations determined from the SGA analysis of the whole oil.

3.4 Molecular Dynamics

Molecular dynamics is a relatively new and exciting field that is becoming important in analysing the structures found in heavy oils. Molecular dynamics applies energy and force balances to each atom in a molecule. The spatial positions of atoms in a molecule are gradually adjusted so that the energy associated with the entire molecule is minimised. The final result of a molecular mechanics simulation is a molecule with a minimal energy three dimensional conformation. Molecular dynamics simulations can also be used to show how molecules interact. Inter and intra-molecular forces that help associate and stabilise molecules such as Π-Π bonding, hydrogen bonding, and van der Walls interactions, can all be identified using molecular dynamics.

Murgich *et al.* (1999) used molecular mechanics to model an Athabasca asphaltene molecule in three dimensions. A single Athabasca asphaltene molecule having the formula $C_{412}H_{509}O_9N_7S_{17}$ and molecular weight of 6239 g/mol was used. This average molecule was consistent with the analytical work of Otto Strausz. The entire molecule was first divided into nine fragments. Each fragment was then subjected to an energy minimisation routine that adjusted the confirmation of each fragment. After the energy associated with each fragment was minimised, all the fragments were connected. The energy of the entire molecule was then minimised. The resulting molecule was a highly folded molecule with a very complex globular three-dimensional structure. The energies of interaction with toluene, *n*-octane, and three resins, of each of the nine subunits were determined. The lowest energy values were for toluene and *n*-octane. The highest energy values were for polar resins. The resins showed higher affinities for the external cavities of the asphaltene molecule than for the two solvents. It was concluded that the resins have a selective affinity for some of the external sites on the asphaltene. This might explain why resins from one oil are sometimes unable to solubilise asphaltenes from other oils.

3.5 The Challenges of Molecular Representations

Thus it is apparent that in order to pursue molecular kinetic based heavy oil hydrocracking models, heavy oil molecular representations first must be created. The complexity of the structural chemistry of heavy oil makes creating molecular representations very challenging and time consuming. First of all, a wide variety of analytical data is needed to create molecular representations of heavy oil fractions. Elemental analysis is necessary to determine the concentrations of various atoms (C,H,O,N,S) present in heavy oil. In order to create a molecular representation, it is critical to know the average molecular weight of the oil fraction being represented. Thus, VPO data is necessary to create molecular representations. Ideally, molecular weight distribution data should be incorporated into a molecular representation. By correctly modeling the molecular weight distribution of heavy oil, the reactivity of the heavy oil will be more precisely modelled. This is important if kinetic work is going to be done using the molecular representation. At the core of a molecular representation is the concentration data of the various structural attributes that appear in the representation. Typically the concentrations of the attributes appearing in molecular representations are determined using a combination of ^{13}C-NMR, ^{1}H-NMR, and IR. It is both very expensive and very time consuming to generate all of the analytical data necessary to create molecular representations.

Once the data are generated, it is still not easy to create molecular representations. An algorithm is generally created that creates heavy oil molecules that are consistent with the current analytical data. The main complexity is that the molecules that are created must both be chemically possible and physically feasible. Once a set of molecules is created, the relative composition of each molecule is adjusted so that the predicted analytical properties of the set are consistent with experimental data. Thus, optimisation algorithms must be used to create molecular representations. Optimisation algorithms require a great deal of computing power and time. It was not until the

1990s that computers powerful enough to run molecular representation optimisation algorithms existed. The creation of molecular representations involves both detailed analytical analysis and challenging modeling and programming. The potential benefits of molecular representations to both industry and academia make the challenges associated with the creation of these representations worthwhile.

Chapter 4

Creating Molecular Representations

The purpose of this section is to document how molecular representations were created for four Athabasca catalytic hydroprocessed fractions. First the data that these representations were created from will be presented. After this it will be explained how this data is transformed to a form that allows probability distributions to be applied. At this point an algorithm that creates heavy oil molecules will be described. After all of this, the optimisation procedure that creates sets of molecules consistent with the starting analytical data will be described. The individual representations for each of the four fractions will then be presented and discussed.

4.1 Data

The data used (Table 5) in this study came from the Gray *et al.* (1992) study. Only data from a medium conversion product that resulted from hydrocraking Athabasca bitumen over a Ni/Mo catalyst was used. This particular data set was used for two particular reasons. First of all, no coke formation occurred during this hydrocracking run. Secondly, the overwhelming majority of all previous heavy oil structural studies have focused on Athabasca bitumen. All of the data in Table 5 was used to create separate molecular molecular representations for each of three distillation fractions middle distillate (176-343°C), gas oil (343-524°C), and residue (524°C+). A generic residue composed of all of the structural attributes determined in the Gray *et al.* (1992) study is found in Figure 25.

Table 5: The data used to create molecular representations for Athabasca bitumen middle distillate (176-343°C), gas oil (343-524°C), and residue (524°C+) catalytic hydroprocessed fractions.

Fraction	Middle Distillate	Gas Oil	Residue
Distillation Yield (wt%)	19.6	38.1	33.7
Elemental Analysis (wt%)			
C	85.71	86.6	84.53
H	12.78	10.94	8.88
O	0.33	0.5	1.67
N	0.37	0.38	0.98
S	0.81	0.159	3.94
Aromatic Structures (mol/100g)			
Benzene	0.146	0.178	0.109
Phenanthrene	0	0.018	0.021
Biphenyl bridge	0.067	0.115	0.499
Aliphatic Structures			
Alpha methylene	0.459	0.571	0.433
Alpha methylene	0.053	0.133	0.1
Beta methyl	0.101	0.118	0.102
Gamma methyl	0.624	0.375	0.269
Chain methylene	1.037	0.641	0.691
Chain methyne	0.184	0.105	0.075
Naphthenic methyl	0.785	0.532	0.273
Naphthenic methylene	1.633	1.3	1.176
Naphthenic methyne	0.93	1.282	0.833
Heteroatomic Structures (mol/100g)			
Benzothiophene	0.015	0.046	0.093
Aromatic sulphoxide	0.004	0.003	0.013
Aliphatic thioether	0.006	0	0.017
Benzofuran	0.012	0.028	0.09
Aromatic hydroxyl	0.002	0	0
Aromatic ketone	0.002	0	0.002
Indole	0.001	0.008	0.014
N-substituted indole	0.018	0.018	0.057
Total alpha carbon (mol/100g)	0.511	0.704	0.532
Total paraffins (mol/100g)	1.221	0.746	0.766
Total naphthenes (mol/100g)	3.348	3.114	2.281
Total aromatics (molC/100g)	1.305	2.135	2.976
Mol weight by VPO (g/mol)	200	365	1160

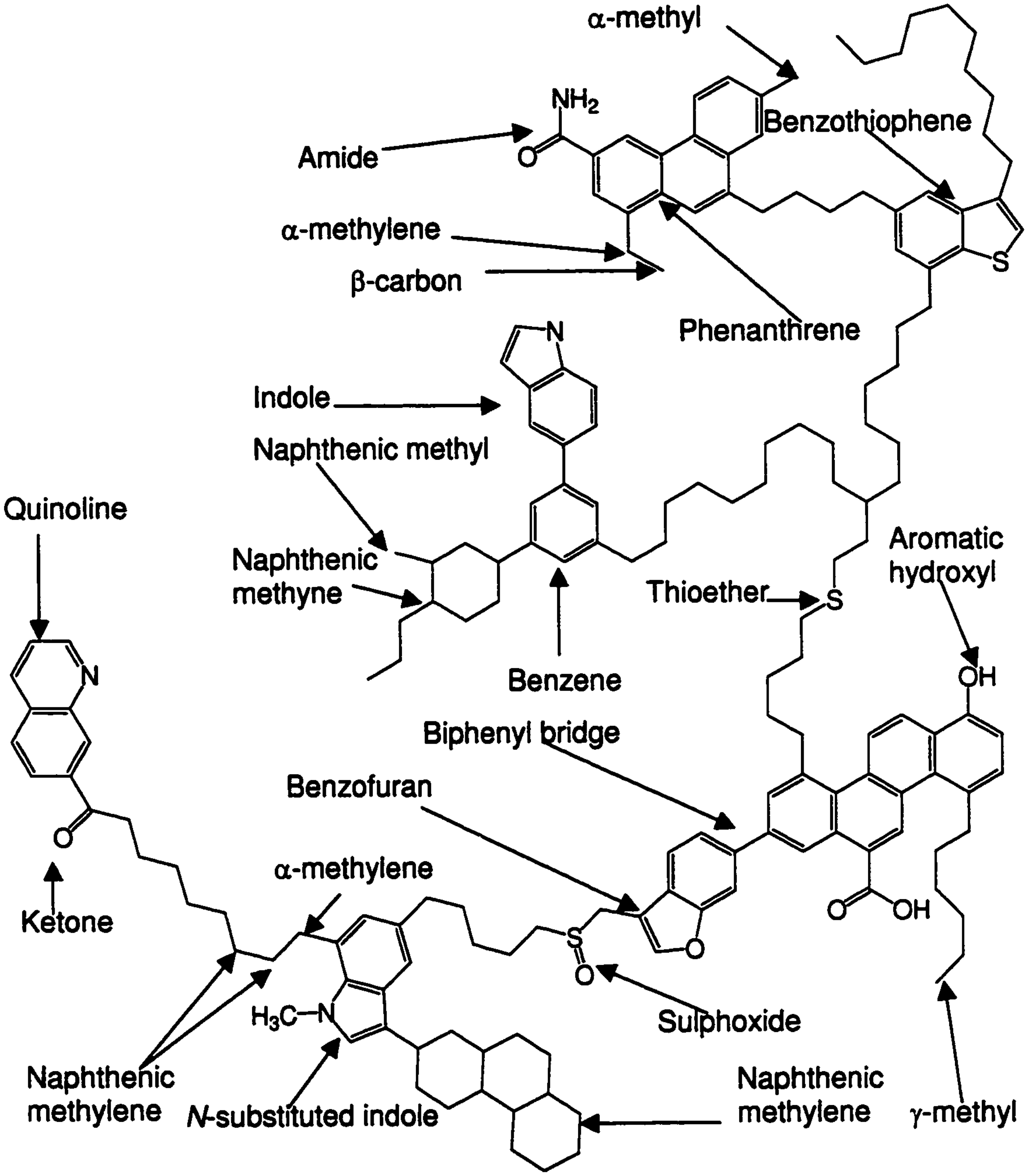


Figure 25: A generic residue molecule composed of all the structural attributes determined in the Gray *et al.* (1992) study.

4.2 Method

The overall objective of this study was to create individual molecular representations (Figure 26) for the four hydrocracked Athabasca bitumen fractions. To accomplish this, two general algorithms were developed one for the residue fraction and gas oil, and a separate one for the naphtha and middle distillate. Each algorithm created heavy oil molecules from the chemical structural attributes determined by Gray *et al.* (1992). Probability distributions were applied to each attribute and randomly sampled. The distributions were sampled cocurrently and concurrently to generate sets of attributes that allowed chemically and physically feasible molecules to be constructed. For each fraction, a large number of molecules were generated. An optimisation algorithm was used to create a representative set for each fraction. Mole fractions for each molecule were optimised. Also, the number of molecules that compose each fraction was reduced by the optimisation program to the minimum number necessary for the representation to be consistent with the analytical data.

4.2.1 Software

All of the molecular representations presented in this thesis were created in Matlab 5.3. The Stats toolbox was used to generate random numbers and for probability distributions. The Optimisation toolbox was used to optimise the mole fractions of each molecule in the final representations. The Microsoft Excel 97 Solver was used to determine the concentration of naphthenic rings. ChemDraw Pro 5.0 was used to draw the molecules in each molecular representation. Cerius 2 was used for the molecular dynamics analysis of proposed bitumen structures.

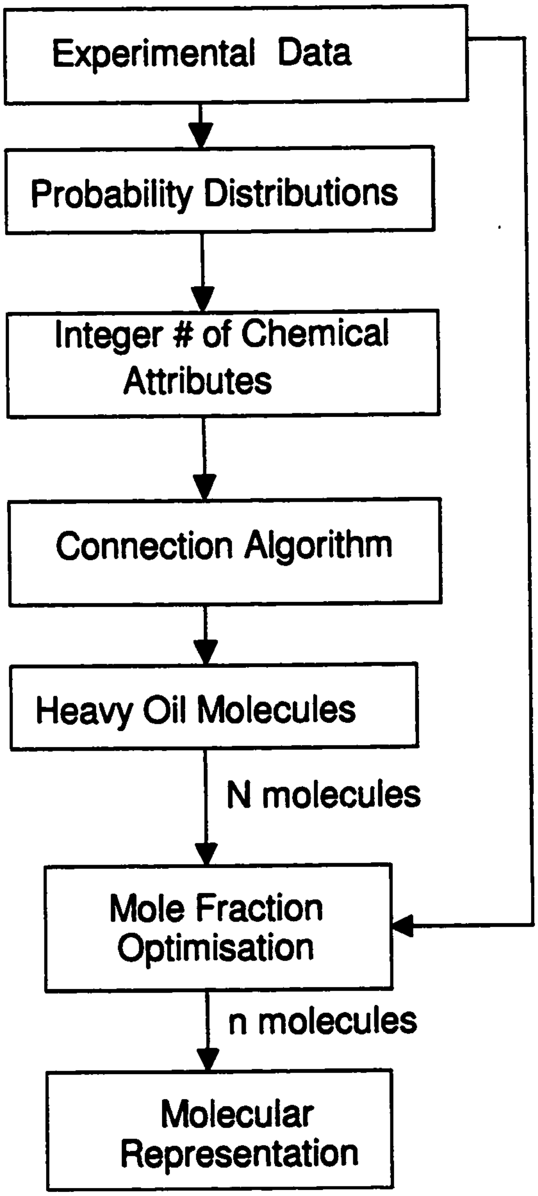


Figure 26: Flow diagram of the strategy used to create molecular representations. N and n are integers. N>n.

4.2.2 Chemical and Physical Feasibility

If a molecule is chemically feasible (Figure 27), all atoms of the molecule are bound together in such a way so that the valency of each atom is obeyed. If a chemically feasible molecule is at a steady state, any radical sites are stabilised, and all non-radical atoms do not have any vacant binding orbitals. From this point on, all references to chemically feasible molecules, are to chemically feasible steady state molecules. A molecule that is physically feasible is a chemically feasible molecule that is stable and does not disintegrate as a result of steric hindrance or repulsion.

Carbon only has four binding orbitals. The molecule to the left is chemically feasable while the one to the right is not chemically feasable.

Figure 27: Chemical feasibility involved when creating molecular representations.

4.2.3 Application of Probability Distributions

Probability distributions are used in molecular modeling for several reasons. Most importantly, they generate random samples that have a mean and/or variance that is consistent with set conditions. Thus, when applied to molecular modeling, probability distributions allow random sets of chemical attributes to be generated. The mean and/or variance of the randomly generated chemical attributes is/are consistent with current available experimental data.

It was assumed that each heavy oil molecule contains a variable number of each of these structural attributes. A given attribute may be very common on one molecule but non-existent on another. The application of probability distributions to the molecular concentration of each structural attribute allows the variation in molecular structure to be modelled.

For each structural attribute in each fraction, if the concentration of the attribute is multiplied by the average molecular weight of the oil fraction, the average concentration of the structural attribute per molecule is expressed. Molecular concentration refers to the concentration of each chemical attribute having the units mol structural attribute/mol of heavy oil.

4.2.3.1 Molecular Concentration

By expressing the concentration of each structural attribute in terms of mol structural attribute/mol heavy oil (Figure 28), it is possible to apply a probability distribution to each attribute. Under certain conditions and with certain distributions, the mean of repeated random sampling from such a distribution will correspond to the number of mol structural attribute/mol heavy oil (the molecular concentration).

$$(\textit{concentration, volume basis})(\textit{avg. MW}) = (\textit{concentration, molecular basis})$$

$$\left(\frac{\textit{mol structural attribute}}{\textit{g heavy oil}}\right)\left(\frac{\textit{g heavy oil}}{\textit{mol heavy oil}}\right)=\left(\frac{\textit{mol structural attribute}}{\textit{mol heavy oil}}\right)$$

Figure 28: Deriving molecular concentrations. In order to apply pdfs to each attribute, the concentration of each attribute must be expressed in terms of mol attribute/mol heavy oil.

4.2.3.2 The γ Distribution

The gamma distribution has been frequently used in modeling various distributions associated with heavy oil properties (Campbell and Klein, 1997). This distribution is very versatile, it can resemble other distributions such as the delta, normal, and exponential. For molecular modeling, the γ-distribution is ideal to work with because it does not generate negative numbers. When dealing with numbers of chemical attributes, negative numbers have no physical significance. It is impossible to have -2 benzene rings/molecule. When applying a probability distribution to a molecular concentration it is very important to apply one that is bounded between 0 and $+\infty$.

The typical gamma distribution is found in Figure 29. α and β are parameters that influence the shape of the distribution. X represents a value that the distribution is calculated at for a given α and β. p represents the value of the distribution for a given x, α, and β. p is confined between 0 and 1. When β=1, the gamma distribution becomes a standard gamma distribution (Figure 30). With the standard gamma distribution, the arithmetic mean of all values of x associated with all values of p between 0 and 1 is α. When a standard gamma distribution is created for the molecular concentration of a structural attribute, α is assigned to be the magnitude of the molecular concentration (mol structural attribute/mol of heavy oil). When repeated random numbers (probabilities) between 0 and 1 are generated and the molecular concentration for each random number is calculated, the mean of all of these numbers will correspond to α, the average molecular concentration for the structural attribute.

$$p = f(x,\alpha) = \frac{x^{\alpha-1}e^{-x}}{\Gamma(\alpha)}$$

Figure 29: A typical gamma distribution.

$$p = f(x,\alpha,\beta) = \frac{x^{\alpha-1} e^{-\frac{x}{\beta}}}{\beta^{\alpha}\Gamma(\alpha)}$$

Figure 30: A standard gamma distribution with mean α.

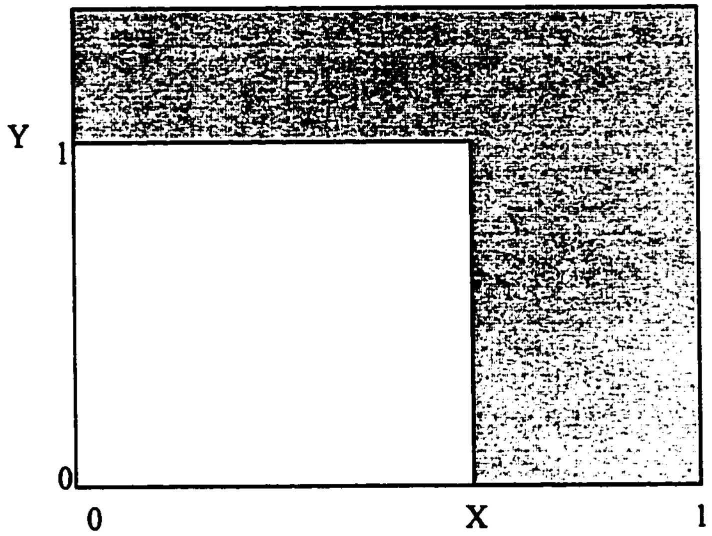
It is currently not possible to measure the distribution of structural attributes among heavy oil molecules. To accomplish this, the isolation of single molecular species would be necessary. The approach taken in this study to create heavy oil representations requires probability distributions to be applied to each chemical attribute. Because no chemical attribute distribution data exists, it was necessary to assume a distribution shape for each chemical attribute. For those chemical attributes that had molecular concentrations less than 1 attribute/molecule, step functions were applied. In most cases the molecular concentrations were much less than one. In such cases it would be expected that most molecules would either contain only one of these attributes or none at all. For those chemical attributes with molecular concentrations greater than 1 attribute/molecule, standard gamma distributions were applied (α=molecular concentration, β=1). For these chemical attributes, the gamma distribution was chosen for three main reasons, its a very versatile distribution, its unable to generate negative numbers, and it has previously been successfully used in the creation of molecular representations. Any distribution that does not generate negative numbers could have been used. However, the gamma distribution is the most versatile of all the probability distributions. It can resemble the normal, delta, and exponential distributions (Campbell and Klein, 1997). If future experimental evidence indicates that one of these types of distributions is better suited for heavy oil molecular modeling, the gamma distributions present in the algorithm developed in this study can easily be modified (by changing α and β). Campbell and Klein (1997) successfully applied the gamma distribution to various chemical attributes in their molecular modeling study. They justified using gamma distributions in that gamma distributions are very effective in describing heavy boiling fractions. Their logic was that a strong correlation exists between a molecules carbon number and boiling point. Furthermore they stated that "many attributes will be implicitly related to the carbon number so the attribute pdfs should be reasonably well modelled by the gamma distribution as well" (Campbell and Klein, 1997).

4.2.3.3 The Step Function

Another useful distribution in molecular modeling is the step function (Figure 31). The step function has 2 variables x, a number between 0 and 1, and α, the mean. The step function has a value of 1 for all values of x less than α, and a value of 0 for all values of x greater than α. Similar to the standard gamma distribution, the arithmetic mean of all values the step function obtains for all values of x between 0 and 1 is α.

4.2.3.4 Random Sampling

When a distribution is randomly sampled, a bounded random number (in this study between 0 and 1) is generated and supplied to a probability distribution. In the case of the cumulative γ distribution, the random number represents probability (the y variable), while for the step function, it represents the x variable, which in turn directly correlates to probability. For the step function (Figure 31) if the random number is less than the predetermined distribution mean, the sampled output is 1 attribute/molecule, otherwise the output is 0. In the case of the cumulative γ distribution, the resulting sample is the number of attributes/molecule consistent with the generated random probability. Usually a non-integer number of attributes/molecule are sampled from the gamma distributions. As a result, the sampled number of attributes/molecule is rounded to the nearest whole number.



X=Distribution Mean
Y=Distribution Output
R=A random number
0≤R≤1
0≤X≤1
IF 0<R≤X, Y=1
IF X<R≤1, Y=0

Figure 31 : The step function.

4.2.3.5 Concurrent Sampling

For all given product fractions, pdfs are applied to all of the chemical attributes that compose the entire subfraction. When all pdfs are sampled at the same time (i.e. concurrently) occasionally the resulting combination of sampled attributes can be linked together to create a chemically and physically feasible molecule. Such molecules contain the exact number of each chemical attribute that was sampled when all of the probability distributions were sampled concurrently. To generate heavy oil molecules in this manner is very time consuming and tedious. Some attribute combinations obviously are not possible to link together in a way that creates physically and chemically feasible molecules. Certain attributes (especially α carbon) may not have been sampled, while other attributes may be in excess (such as β carbon). It is also sometimes difficult to identify concurrent samplings that have produced chemically and physically feasible molecules. Sometimes attributes have to be reconnected many times in order to identify samplings that have produced chemically and physically feasible molecules.

4.2.3.6 Cocurrent Sampling

A much better sampling method that can be used to generate chemically and physically feasible molecules is cocurrent sampling. With cocurrent sampling, some chemical attributes are sampled only after other attributes are sampled. When certain constraints are violated, some structural attribute pdfs are resampled. Resampling continues until the number of structural attributes sampled, allows a feasible molecule to be constructed. After certain attributes are sampled, maximum and/or minimum constraints are put on the samplings from certain distributions that have not been sampled yet. For example, if all of the distributions corresponding to aromatic ring systems are sampled, and only a single benzene ring is sampled, then the maximum number of α carbons that can be placed on the benzene ring is six. To increase the likelihood that the resulting molecule will be physically feasible, no more than three substituted groups are allowed to be attached to the benzene ring. Therefore a maximum of three α methylene carbons would be permitted to be attached to the benzene ring. In such a situation, a constraint would be applied to the α methylene probability distribution. This constraint would specify that if the number of α methylenes sampled exceeded three; α methylene would be resampled until the number sampled was less than three.

4.2.4 Storage of Molecules

It was desired to represent heavy oil molecules in the most computationally efficient way possible. A vector approach was taken that is similar to that of Quann and Jaffe (1991). However, due to the large molecular weights of heavy oil molecules, more than one structural vector was necessary to represent most heavy oil molecules. Thus, a structural matrix framework was used to represent heavy oil molecules. The structural organisation of a molecule was described using an adjuvant connection matrix.

Two major matrices, a structural matrix, and a connection matrix represent each heavy oil molecule. Each column in a structural matrix (Figure 32) corresponds to a certain chemical attribute. A structural matrix consists of 4 different sections. Each section of a structural matrix corresponds to a section of rows of the matrix. A row corresponds to a chemical building block of the molecule. The first few rows correspond to substituted aromatic clusters. An aromatic cluster contains a maximum of four aromatic rings. Rows corresponding to aromatic clusters indicate which aromatic attributes are present in a cluster, as well as which substituted groups (α-carbon, β-carbon, naphthenic groups) are attached to the cluster. Each entry in the structural matrix indicates how many chemical attributes are present in the subsection, building block, that it is entered in. Each aromatic cluster and each individual paraffinic carbon, either chain methylene or methyne, is represented as a separate row. All γ carbons are lumped together in the very last row.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
Aromatic Clusters	1	0	0	0	0	1	0	1	0	0	1	2	0	0	0	0	0
Chain Methylene	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
Chain Methylene	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
Chain Methyne	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
γ-carbon	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2

Column	Structural Attribute
A	Benzene
B	Phenanthrene
C	Benzothiophene
D	Benzofuran
E	Indole
F	N-indole
G	Alkylcyclohexane
H	Alkyldecalin
I	Sterane
J	Short naphthenic chains
K	α-Methyl
L	α-Methylene
M	β-Methyl
N	Sulphoxide
O	Chain Methylene
P	Chain Methyne
Q	γ-Methyl

Figure 32: The structural matrix.

$$\begin{array}{c|cccccc}
 & \mathbf{Aro} & \mathbf{CM} & \mathbf{CM} & \mathbf{Cm}\ \gamma & & \gamma \\
 & 1 & 2 & 3 & 4 & 5 & 6 \\
\mathbf{Aro1} & 0 & 0 & 0 & 1 & 0 & 0 \\
\mathbf{CM2} & 0 & 0 & 1 & 0 & 1 & 0 \\
\mathbf{CM3} & 0 & 1 & 0 & 1 & 0 & 0 \\
\mathbf{Cm4} & 1 & 0 & 1 & 0 & 0 & 1 \\
\gamma\mathbf{5} & 0 & 1 & 0 & 0 & 0 & 0 \\
\gamma\mathbf{6} & 0 & 0 & 0 & 1 & 0 & 0
\end{array}$$

Aro-Aromatic Cluster; CM-Chain Methylene
Cm-Chain methyne; γ-γ carbon

Figure 33: The connection matrix.

A connection matrix (Figure 33) indicates the order in which the aromatic clusters, paraffinic carbons, and γ carbons found in the structural matrix, are connected to each other. Each aromatic cluster, paraffinic carbon, and γ carbon is given its own row in connection matrix. For the aromatic clusters and chain methylenes, the chemical attribute in each ith row in the connection matrix corresponds directly to chemical attribute in the ith row in the structural matrix (Figure 32). A separate row is created for each γ carbon in the connection matrix. In a connection matrix the chemical attribute in each jth row corresponds to the chemical attribute in each jth column. A 1 entry e(x,y) in the connection matrix indicates a connection between the chemical attribute in row (column) x with the attribute in column (row) y. A 0 entry indicates no connection. Because of the fact that each chemical attribute in each jth row corresponds to the chemical attribute in each jth column, the connection matrix is symmetrical about its (n,n) diagonal (the diagonal progressing from (1,1) to (n,n) were n is the number of rows in the matrix). The connection matrix is used along with the corresponding structural matrix to directly construct each heavy oil molecule (Figure 34).

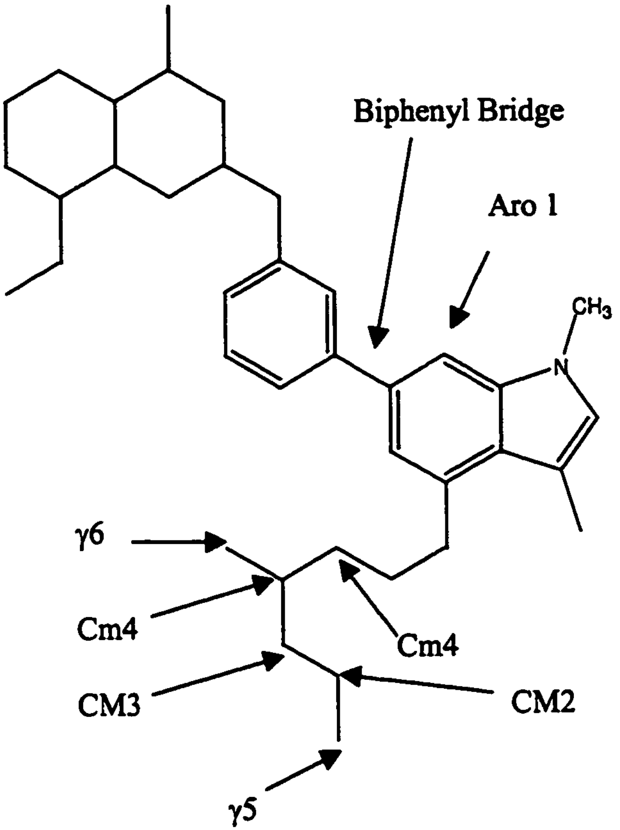


Figure 34: A heavy oil molecule created from connection and structural matrixes.

4.2.5 Two Unique Cases, Fused Aromatic Connections and Thio Ethers

When an aromatic cluster contains two aromatic attributes linked together by a bridgehead connection, an additional structural vector or matrix is necessary. This additional vector or matrix indicates which aromatic groups in a molecule are linked together via a bridge head connection. When a molecule contains more than two aromatic clusters that contain fused aromatic linkages, a matrix indicates the locations of the linkages. If the benzene and N-substituted indole groups in the molecule in Figure 34 were linked together (Figure 35) a vector [1 1], [aromatic cluster # that contains bridgehead connections, number of connections] would indicate that the aromatic groups in the first aromatic cluster are linked together by a fused connection. It was assumed that only clusters that contain benzene attributes could contain bridgehead connections. For two sets of molecular representations only allowed fused aromatic connections to exist between benzene and benzene attributes, and benzene and phenanthrene attributes. In another case benzene was allowed to be connected by fused connections to all other aromatic ring attributes. For aromatic clusters that contain two benzenes and one two membered aromatic ring attribute (i.e. benzofuran) the number of benzene attributes connected in a fused connection is indicated in the third column of the vector.

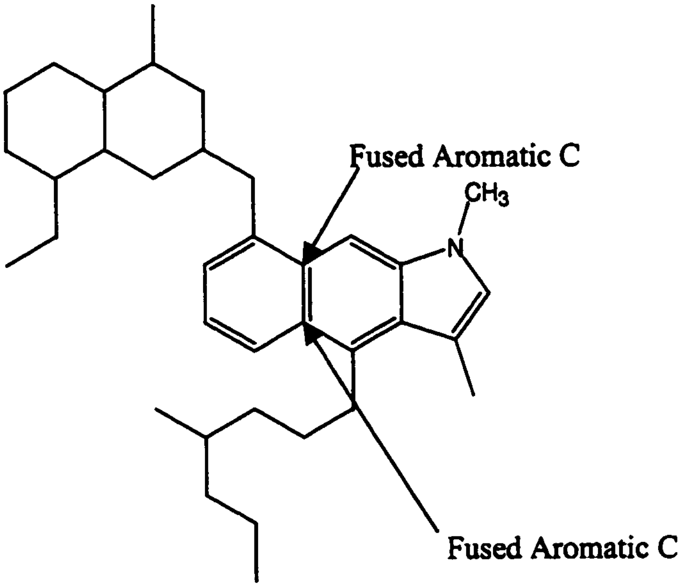


Figure 35: A heavy oil molecule containing two aromatic groups linked together by a fused connection.

When a heavy oil molecule contains a thioether (Figure 37) another additional structural vector is necessary. This additional vector/matrix indicates parrafinic carbons that aliphatic thioethers are located between. When a molecule contains more than two thioethers, a matrix indicates the locations of the linkages. If the molecule in Figure 35 contained a thioether between chain methylene 3 and chain methylene 2, this would be indicated by the vector [3 2 1].

Vectors that indicate thioether position have the format [paraffinic C #, paraffinic C #, 1].

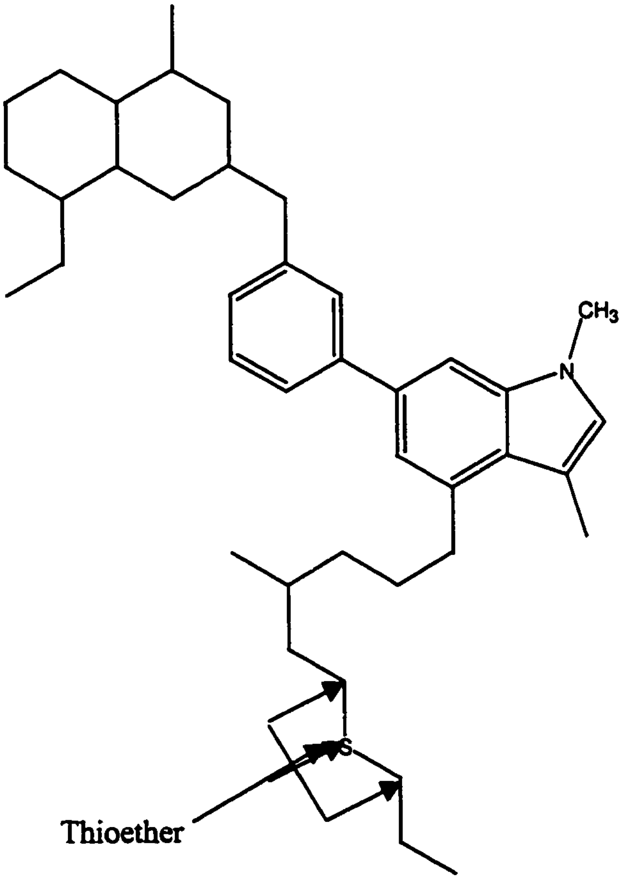


Figure 36: A heavy oil molecule containing a thioether.

4.2.6 Naphthenic Carbon

Besides being found at the start of carbon chains branching off aromatic rings, naphthenic carbons are also found in heterocyclic fully saturated (no double bounds) rings. Some common naphthenic rings (Figure 37) found in heavy oils include alkyl cyclohexane, alkyl decalin, and sterane.

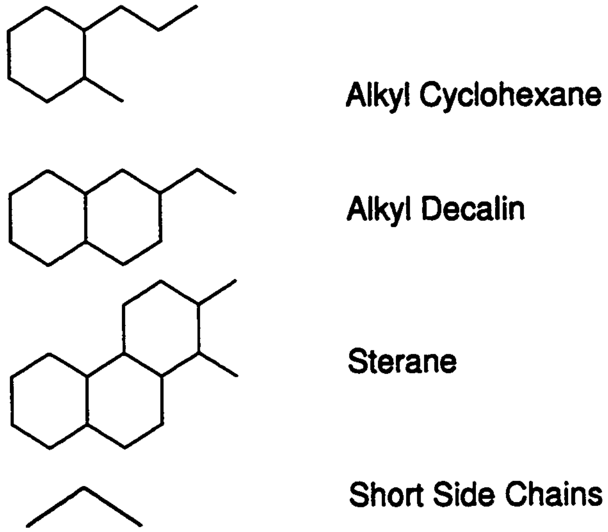


Figure 37: Common naphthenic containing secondary structures found in heavy oils.

When modeling heavy oil molecules, to be consistent with analytical information, naphthenic carbon must be incorporated both at the start of aromatic substituted carbon chains and in heterocylic ring systems. Naphthenic containing secondary structures refer to the heterocyclic rings and to short four carbon side chains that contain two naphthenic methylenes, a single γ carbon, and a single α carbon. A least squares optimisation routine can determine the molecular concentrations of the naphthenic containing secondary structures that are consistent with the molecular concentration data for a given heavy oil subfraction. The molecular concentration of each napthenic secondary structure is initially guessed. For the arbitrarily initial ring concentrations, the molecular concentrations of naphthenic methyl, methylene, and methyne carbons are determined for the heavy oil fraction. For each naphthenic carbon type, the square of the difference between the predicted and experimentally determined molecular concentration is determined. An optimisation routine is used to optimise the molecular concentrations of the naphthenic secondary structures. This routine minimises the sum of the squares of the difference between experimental and predicted molecular concentrations for each naphthenic carbon subtype. At this point, the naphthenic carbons at the start of aromatic substituted carbon side chains have been neglected. Using the optimised molecular concentrations of naphthenic structures, a large number (at least one hundred) of heavy oil molecules are made. The average number of α methylenes associated with paraffinic chains are calculated. It is assumed that each α methylene associated with paraffinic chains is first connected to two naphthenic methylenes. The average number of napthenic methylenes associated with paraffinic chains is calculated per molecule. This average value is subtracted from the molecular concentration of naphthenic methylene to yield the molecular concentration of naphthenic methylenes present in naphthenic secondary structures. Using this new concentration, the optimisation procedure is repeated to determine the molecular concentration of each naphthenic containing secondary structure.

4.2.7 α-Carbons and Aromatic Clusters

α-carbons are important in that they link together aromatic groups with substituted groups including naphthenic rings and β-carbons. Aromatic clusters are defined as groups of aromatic attributes and attached naphthenic groups, β-methyl, and α carbons. The aromatic rings within an aromatic cluster can either be connected together with biphenyl bridges (Figure 34) or with bridgehead connections (Fig. 35).

4.2.8 The Heavy Oil Molecule Connection Algorithm

An algorithm that generates chemically and physically feasible heavy oil molecules was created. This algorithm uses both cocurrent and concurrent random sampling. In this algorithm probability distribution functions were assigned to the molecular concentrations of each chemical attribute. The molecular concentrations of naphthenic containing secondary structures were determined using the methodology outlined in section 4.2.6. First, random sampling determines the number of attributes found in a given molecule. Sampling is performed with some constraints that assure that only physically and chemically feasible molecules will be generated. After sampling, a connection algorithm connects paraffinic carbons to aromatic groups, assuring that all atoms in the molecule are connected together. This is a comprehensive summary of the algorithm.

1. Aromatic attributes are sampled.
2. If no aromatic attributes are sampled, the aromatic attributes are sampled until at least one is generated.
3. Each aromatic attribute is randomly put into an aromatic cluster group. Aromatic clusters containing four aromatic rings are favoured. Each aromatic cluster has at least one aromatic ring and no more than four.
4. Naphthenic attributes; alkylcyclohexane, alkyldecalin, sterane, four carbon chain, are randomly sampled.
5. The napthenic groups that were generated are randomly placed in aromatic clusters. Each aromatic cluster has a maximum of two attached naphthenic attributes.
6. α-methyl, α-methylene, β-methyl, and aromatic sulphoxides, are randomly sampled. Each group that is generated is then randomly assigned to an aromatic cluster.
7. If necessary, the number of α methylenes per aromatic cluster is adjusted so that for each aromatic cluster:
   a. total naphthenics + total β methyl < total α methylene + total aromatic sulphoxides.
   b. total α methylene + total aromatic sulphoxides >0.
   c. no more than 3 α methylenes and aromatic sulphoxides not associated with a naphthenic attribute or β methyl, are associated with the cluster.
8. If necessary, the number of naphthenics, α-methyls, and β-methyls, are decreased for each aromatic cluster so that each aromatic cluster contains

no more than (2n + 1) substituted groups (n=number of aromatic rings in the cluster). This prevents an extreme degree of aromatic substitution and thus maximises the probability that each aromatic cluster and thus every heavy oil molecule will be physically in addition to chemically feasible.

9. If more than two aromatic clusters were created, it is assured, if necessary, by adjusting the number of α methylenes per aromatic cluster, that for n aromatic clusters, (n-2) clusters each contain at least 2 α methylenes not associated with a β methyl or naphthenic group, and the other 2 aromatic clusters each contain at least one α methylene not associated with a β methyl or naphthenic group. This assures that all the aromatic clusters will be able to be connected together via paraffinc carbons.
10. Paraffinic carbons; chain methyne and chain methylene, are sampled.
11. The valency of each aromatic cluster, chain methylene, and chain methyne is determined. For each aromatic cluster valency is defined as:

$$v_{AR,i} = \sum \alpha\ \textit{methylene} + \sum \textit{aromatic sulphoxides} - \sum \beta\ \textit{methyl} - \sum \textit{naphthenic attributes}$$
$$- \sum \textit{attached chain methylenes} - \sum \textit{attached chain methynes}$$

For each chain methylene valency is defined as:

$$v_{CM,j} = 2 - \sum \textit{attached aromatic clusters} - \sum \textit{attached chain methylenes}$$
$$- \sum \textit{attached chain methynes}$$

For each chain methyne valency is defined as:

$$v_{CMY,k} = 2 - \sum \textit{attached aromatic clusters} - \sum \textit{attached chain methylenes}$$
$$- \sum \textit{attached chain methynes}$$

12. When all the aromatic cluster, chain methynes, and chain methylenes, are connected together, the sum of all the valencies must be:

$$v_{\min} = \sum_{i=1}^{I} v_{AR,i} + K - (I-1) * 2$$

I=total number of aromatic clusters.
J=total number of chain methylenes.
K=total number of chain methynes.

A connection matrix of size (I+J+K)*(I+J+K) is created having 0 as every entry. Each row and column in this matrix corresponds to one of the aromatic clusters, chain methylenes, or chain methynes, that were generated. In succession, two of the generated chemical attributes are randomly chosen. If both attributes have at least one valency, and have not already been connected to other attributes, the two attributes will be connected together, provided that they both are not aromatic clusters. If each attribute has already been connected to another attribute, no connection will be allowed. This prevents paraffinic carbon from forming naphthenic secondary structures. When two attributes, A and B, are allowed to connect, the valency of each group is decreased by one, and the entries (A,B) and (B,A) in the connection matrix become 1. Pairs of attributes are randomly chosen and connected until the sum of the valencies of all aromatic clusters, chain methynes, and chain methylenes is $v_{\min}$.

13. All α-methylenes not associated with β-methyls, or naphthenic groups, or connected to chain methylenes, or chain methynes, are eliminated.
14. Aliphatic thioethers are randomly sampled. If any thioethers are generated, they are randomly placed between two paraffinic carbons.
15. Two different models for fused aromatic connections were used. In the first model, fused aromatic connections were only allowed between two benzenes or a benzene and a phenanthrene. In the second model, fused aromatic connections were allowed between a single benzene group and any other aromatic attribute.

    In both cases, whenever a fused aromatic connection was allowed to exist, a 50:50 probability was assumed in determining whether an aromatic connection was a fused or a biphenyl connection. Fused aromatic connections are then randomly assigned to aromatic clusters that contain groups of attributes that are allowed to connect by bridgehead connections.
16. The total number of substituted groups, fused aromatic connections, and biphenyl connections is determined, as well as the number of α-methylenes connected to paraffinic carbons.

4.3 Molecular Representations For Athabasca Heavy Oil Catalytic Hydroprocessed Distillation Fractions

A heavy oil molecular representation is a set of molecules that is consistent with all available analytical data for the heavy oil they represent. A molecule set can either include a unique copy of each molecule, or alternatively it can contain a relative fraction assigned to each molecule. When mole fractions are assigned to each molecule, each mole fraction can be adjusted so that the overall properties of the molecules are consistent with analytical data.

Molecular representations were created for the gas oil, middle distillate, and residue fractions, that resulted when Athabasca heavy oil was hydrocracked using Ni/Mo on alumina catalyst. Fifty molecules of each fraction were made. For each subfraction, an algorithm was created that calculated for a set of fifty molecules and corresponding fifty mole fractions, ten properties. These properties were the carbon content (mol C/100 g) of aromatics, paraffinics, naphthenics, and α-carbons; the percent mass content of C, H, O, N, S; and the average molecular weight of the mixture. Each property calculation algorithm (an objective function) also calculated the sum of the absolute differences between the analytically determined values of these properties and the average properties calculated from the molecule set.

To create a molecular representation for a heavy oil fraction, the mole fractions of each molecule must be adjusted. Each mole fraction is adjusted so that the sum of the absolute differences between the predicted and analytically determined property values is minimised. Thus the problem of creating heavy oil molecular representations is an optimisation problem. To optimise each molecular mole fraction an objective function is created. It was found that the objective function that yields the best optimisations is based on the absolute difference between the average property value calculated from a molecule set and the analytically determined value. Alternatively the square of the difference could be used, this method did not to work as well as the absolute method. The objective function (Figure 38) calculates the sum of the weighted normalised differences for all properties. Each normalised absolute difference is multiplied by a weighting factor. The larger the weighting factor, the smaller the deviation between the predicted and experimental properties. When a weighting factor of one is assigned to all properties, some predicted properties deviate very slightly from the experimentally determined properties, while others have very large deviations. Thus the more a property tends to deviate, the larger the weighting factor assigned to it will be. For the first optimisation all weighting factors are assigned to be one. Each molecule has an associated number within this vector. The calculation algorithm normalises all elements within this vector so that the normalised sum of all elements in it is one. Associated with a set of molecules is a nonnormalised vector. Using an optimisation routine, the elements in the nonnormalized vector are optimised, so that the value of the objective function calculated using the corresponding mole fractions is minimised. For all representations, the optimised mole fraction associated with a molecule is 0. Thus, the optimisation procedure also optimises the number of molecules in the representation. After the first optimisation, an iterative process is done were the weighting factors are changed by looking at the results of previous optimisations. Ideally, the percent deviation for each property should be less than 10 %.

It was desired to represent heavy oil molecules in the most computationally efficient way possible. A vector approach was taken that is similar to that of Quann and Jaffe (1992). However, due to the large molecular weights of heavy oil molecules, more than one structural vector was necessary to represent most heavy oil molecules. Thus, a structural matrix framework was used to represent heavy oil molecules. The structural organisation of a molecule was described using an adjuvant connection matrix.

$$
\begin{aligned}
objF = {} & a\left|molW\ analyt - molW\ pred\right| / molW\ analyt + b\left|aromatic\ C\ analyt - aromatic\ C\ pred\right| / aromatic\ C\ analyt \\
& + c\left|naphth\ C\ analyt - naphth\ C\ pred\right| / naphth\ C\ analyt + d\left|paraff\ C\ analyt - paraff\ C\ pred\right| / paraff\ C\ analyt \\
& + e\left|\alpha\ C\ analyt - \alpha\ C\ pred\right| / \alpha\ C\ analyt + f\left|\%\ C\ analyt - \%\ C\ pred\right| / \%\ C\ analyt \\
& + g\left|\%\ H\ analyt - \%\ H\ pred\right| / \%\ H\ analyt + h\left|\%\ O\ analyt - \%\ O\ pred\right| / \%\ O\ analyt \\
& + i\left|\%\ N\ analyt - \%\ N\ pred\right| / \%\ N\ analyt + j\left|\%\ S\ analyt - \%\ S\ pred\right| / \%\ S\ analyt
\end{aligned}
$$

Figure 38: The objective function used to optimise the mole fractions of each molecule in a heavy oil molecular representation. Each letter a,b,c,..j, represents a weighting factor.

4.3.1 Residue Representations

The residue representation (Table 6) contained 8 molecules. Randomness is evident in the residue representation. Some molecules contain nitrogen, others do not. Some molecules contain sulphur, others do not. No two molecules have the same molecular weights. In this representation, two molecules compose 50% of the representation (Table 6).

Table 6: Molecular representation for an Athabasca residue. Only benzene rings were allowed in polyaromatic clusters.

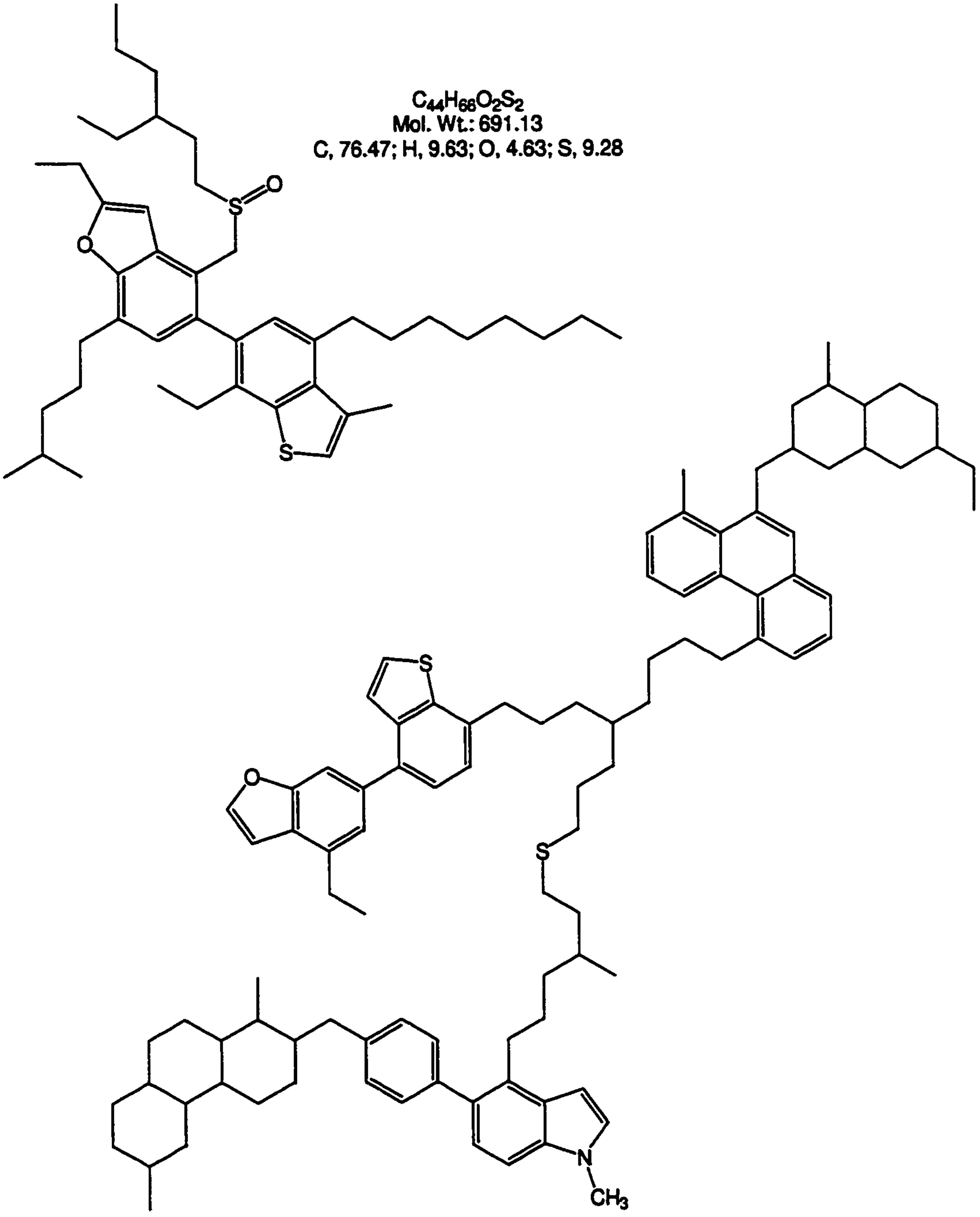
Molecule Number	% of representation (mole basis)	MW (g/mol)	C	H	O	N	S
1	6.64	866.4	61	87	0	1	1
2	0.41	733.2	50	72	0	2	1
3	24.68	1383.2	96	135	1	1	2
4	25.86	691.1	44	66	2	0	2
5	0.23	657.1	48	66	0	1	0
6	18.45	1273.0	91	117	1	1	1
7	3.81	1449.3	107	135	0	2	0
8	19.91	1245.0	88	125	1	1	1

It is very difficult to achieve a balance between α-carbon and aromatic carbon in all of the molecular representations, including those of the residue. In general α-carbon content is proportional to the aromatic carbon content. The optimisations generally either result in the predicted α-carbon content being very close to the experimentally determined value, or for the aromatic content value being very close to the experimentally determined value. It is very rare for both aromatic carbon content and α-carbon content to be very close to their experimentally determined values. It was found that when the naphthenic, aromatic, and paraffinic, carbon contents are all very close to their experimental values, α-carbon is generally in over abundance. Thus, the objective functions must carefully balance α-carbon and aromatic carbon.

The properties of the first representation matched the experimentally determined properties very closely (Table 7). Some of the molecules from the first representation are found in Figure 39.

Table 7: Predicted versus experimental properties of the Athabasca residue representation. Only benzene rings were allowed to be in polyaromatic clusters.

Property	Predicted	Experimental	% deviation
avg molecular weight (g/mol)	1120	1120	0
avg total naphthenics (mol C/100 g)	2.45	2.281	7.19
avg total aromatic (mol C/100 g)	2.68	2.976	9.93
avg total alpha (mol C/100 g)	0.58	0.532	9.91
avg total paraffinic (mol C/100 g)	0.81	0.746	8.45
avg % C	83.46	84.53	1.27
avg % H	9.73	8.88	9.53
avg % O	1.64	1.67	1.83
avg % N	0.98	0.98	0.00
avg % S	4.19	3.94	6.44
total deviation			54.53



$C_{44}H_{68}O_2S_2$
Mol. Wt.: 691.13
C, 76.47; H, 9.63; O, 4.63; S, 9.28

$C_{97}H_{123}NOS_2$
Mol. Wt.: 1383.15
C, 84.23; H, 8.96; N, 1.01; O, 1.16; S, 4.64

Figure 39: Some molecules found in the Athabasca residue representation. Only benzene rings were allowed to be in polyaromatic clusters.

4.3.2 Gas Oil Representations

The gas oil fraction had the most precise molecular representation. Three representations were done for the gas oil fraction. Benzene groups were allowed to fuse with other benzene groups and phenanthrene, along with di condensed aromatic clusters containing oxygen (benzofuran), nitrogen (indole, N-indole), or sulphur (benzothiphene).

Three representations were completed for the gas oil fraction. The first representation (Tables 8 and 10) did not contain a simulated molecular weight distribution. In the second (Tables 9 and 11) and third representations (Figure 40) simulated molecular weight distributions were also considered in the model. A standard cumulative gamma function was applied to the average molecular weight of the gas oil. In the second distribution, three distribution critical points were determined, a 5% cumulative point, a 50% cumulative point, and a 95% cumulative point. A cumulative point indicates the percent of all molecules that have molecular weights less than a given molecular weight. Thus, since the simulated molecular weight distribution had three critical points, it was divided into four sections; <5% (MW%<334 g/mol), 5-50% (334≤MW%<365 g/mol), 50-95% (365≤MW%<397 g/mol), and >95% (MW%>397 g/mol). Fifty molecules were used as an input into the second representation.

The simulated distribution for the third distribution used nine critical points (Table 12). Thus it was divided into ten discrete regions. One hundred molecules were used as an input into the third representation. The objective function for the second and third representations incorporated the cumulative points of the simulated molecular weight distribution.

Table 8: Molecular representation for the first Athabasca gas oil representation. Benzene rings were allowed to be in polyaromatic clusters with aromatic rings containing oxygen, nitrogen, and sulphur.

Molecule Number	% of representation (mole basis)	MW (g/mol)	C	H	O	N	S
1	11.35	394.64	28	42	1	0	0
2	0.68	382.67	28	46	0	0	0
3	0.22	476.8	33	48	0	0	1
4	36.5	340.59	25	40	0	0	0
5	4.11	368.64	27	44	0	0	0
6	4.9	354.62	26	42	0	0	0
7	17.89	350.56	24	30	0	0	1
8	14.34	372.68	27	48	0	0	0
9	9.9	435.73	31	49	0	1	0

Table 9: Molecular representation for the second Athabasca gas oil representation. Benzene rings were allowed to be in polyaromatic clusters with aromatic rings containing oxygen, nitrogen, and sulphur. A simulated molecular weight distribution containing three critical points were used to create this molecular representation.

Molecule Number	% of representation (mole basis)	MW (g/mol)	C	H	O	N	S
1	11.41	394.64	28	42	1	0	0
2	2.5	382.67	28	46	0	0	0
3	26.9	340.59	25	40	0	0	0
4	14.55	368.64	27	44	0	0	0
5	18.1	350.56	24	30	0	0	1
6	11.45	372.68	27	48	0	0	0
7	5	304.47	23	28	0	0	0
8	9.9	435.73	31	49	0	1	0

Unlike the residue molecules, all the molecules in the gas oil representations contain a single heteroatom. The first gas oil representation (Table 8) required more molecules than the residue representation (9 versus 8). The second (Table 9) and third gas oil representations required eight and twenty one molecules respectively. Some of the molecules from the first representation are found in Figure 41. It is of interest that the second gas oil representation contained fewer molecules than the one that did not have the with the simulated molecular weight distribution. It would be expected that the more stringent the optimisation criteria, the more molecules it would take to create a molecular representation. Only minimal deviation was observed between the predicted properties of the first gas oil representation (Table 10) and the experimental analytical properties. For the first gas oil representation with the simulated molecular weight distribution (Table 11), no properties had significant deviation. In the third gas oil representation (Table 12) only aromatic carbon had significant deviation (13.70%). Fifty molecules were supplied to the optimisation routines to create the first and second representations. To satisfy the simulated distribution critical points, it was necessary to supply one hundred molecules into the optimisation routine of the third gas oil representation. The total deviations of all structural properties (all properties except those related to the simulated molecular weight distributions) of the three gas oil representations (7.49%, 8.80%, and 25.19%, were all less than the total deviation for the residue representation (54.53%) (Table 7). With the second gas oil representation, it is interesting that even though a simulated molecular weight distribution is imposed, compared to the representation with no simulated weight distribution, the total error increased only by 1.2%. Overall the molecular weight distribution of the third gas oil representation matched quite well with that of the standard cumulative gamma function (Figure 40). The most significant deviation was observed for cumulative fractions below 0.10 and above 0.90. Cumulative fractions below 0.10 and above 0.90 correspond to the two extreme points used to construct the simulated molecular weight distribution curve. It would be expected that if more points were included beyond these levels, that this deviation would be less severe.

Table 10: Predicted versus experimental properties of the first Athabasca gas oil representation. Eight molecules composed this representation.

Property	Predicted	Experimental	% deviation
avg molecular weight (g/mol)	365.00	365	0.00
avg total naphthenics (mol C/100 g)	3.21	3.114	3.10
avg total aromatic (mol C/100 g)	2.16	2.135	1.05
avg total alpha (mol C/100 g)	0.70	0.704	0.00
avg total paraffinic (mol C/100 g)	0.75	0.746	0.00
avg % C	86.27	86.6	0.38
avg % H	11.26	10.94	2.94
avg % O	0.50	0.5	0.02
avg % N	0.38	0.38	0.00
avg % S	1.59	1.59	0.00
total deviation			7.49

Table 11: Predicted versus experimental properties of the second Athabasca gas oil representation. A simulated molecular weight distribution containing three critical points was used to create this molecular representation. Eight molecules composed this representation.

Property	Predicted	Experimental	% deviation
avg molecular weight (g/mol)	365.00	365	0.00
avg total naphthenics (mol C/100 g)	3.13	3.114	0.55
avg total aromatic (mol C/100 g)	2.27	2.135	6.18
avg total alpha (mol C/100 g)	0.70	0.704	0.00
avg total paraffinic (mol C/100 g)	0.75	0.746	0.00
avg % C	86.39	86.6	0.24
avg % H	11.14	10.94	1.83
avg % O	0.50	0.5	0.00
avg % N	0.38	0.38	0.00
avg % S	1.59	1.59	0.00
total deviation of structural properties			8.80
% MW <334g/mol	0.05	0.05	0
334g/mol≤% MW <365g/mol	0.45	0.45	0.02
365g/mol≤% MW <397g/mol	0.40	0.45	10.89
% MW>397g/mol	0.10	0.05	98

Table 12: Predicted versus experimental properties of the third Athabasca gas oil representation. A simulated molecular weight distribution containing nine critical points was used to create this molecular representation. Twenty one molecules composed this representation. One hundred molecules were used as an input to the optimisation routine associated with this representation.

Property	Predicted	Experimental	% deviation
avg molecular weight (g/mol)	374.60	365	2.63
avg total naphthenics (mol C/100 g)	2.88	3.114	7.66
avg total aromatic (mol C/100 g)	2.43	2.135	13.70
avg total alpha (mol C/100 g)	0.71	0.704	0.67
avg total paraffinic (mol C/100 g)	0.75	0.746	0.12
avg % C	86.57	86.6	0.04
avg % H	10.96	10.94	0.22
avg % O	0.50	0.5	0.00
avg % N	0.38	0.38	0.03
avg % S	1.59	1.59	0.13
total deviation of structural properties			25.19
% MW <340.7 g/mol	0.15	0.1	49.50
340.7 g/mol<% MW <348.8 g/mol	0.00	0.1	100.00
348.8 g/mol<% MW <354.7 g/mol	0.11	0.1	9.40
354.7 g/mol<% MW <359.9 g/mol	0.00	0.1	100.00
359.9 g/mol<% MW <364.7 g/mol	0.10	0.1	0.70
364.7 g/mol<% MW <369.5 g/mol	0.10	0.1	2.00
369.5 g/mol<% MW <374.8 g/mol	0.11	0.1	10.90
374.8 g/mol<% MW <381.0 g/mol	0.10	0.1	0.70
381.0 g/mol<% MW <389.7 g/mol	0.10	0.1	0.50
% MW>389.7 g/mol	0.23	0.1	125.60

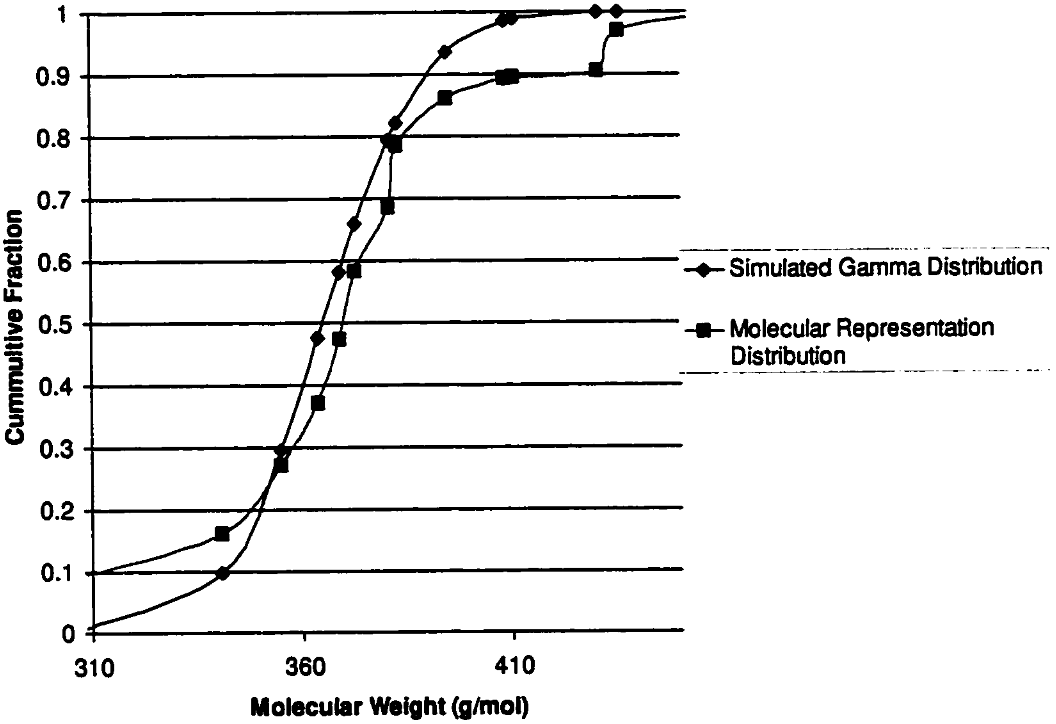
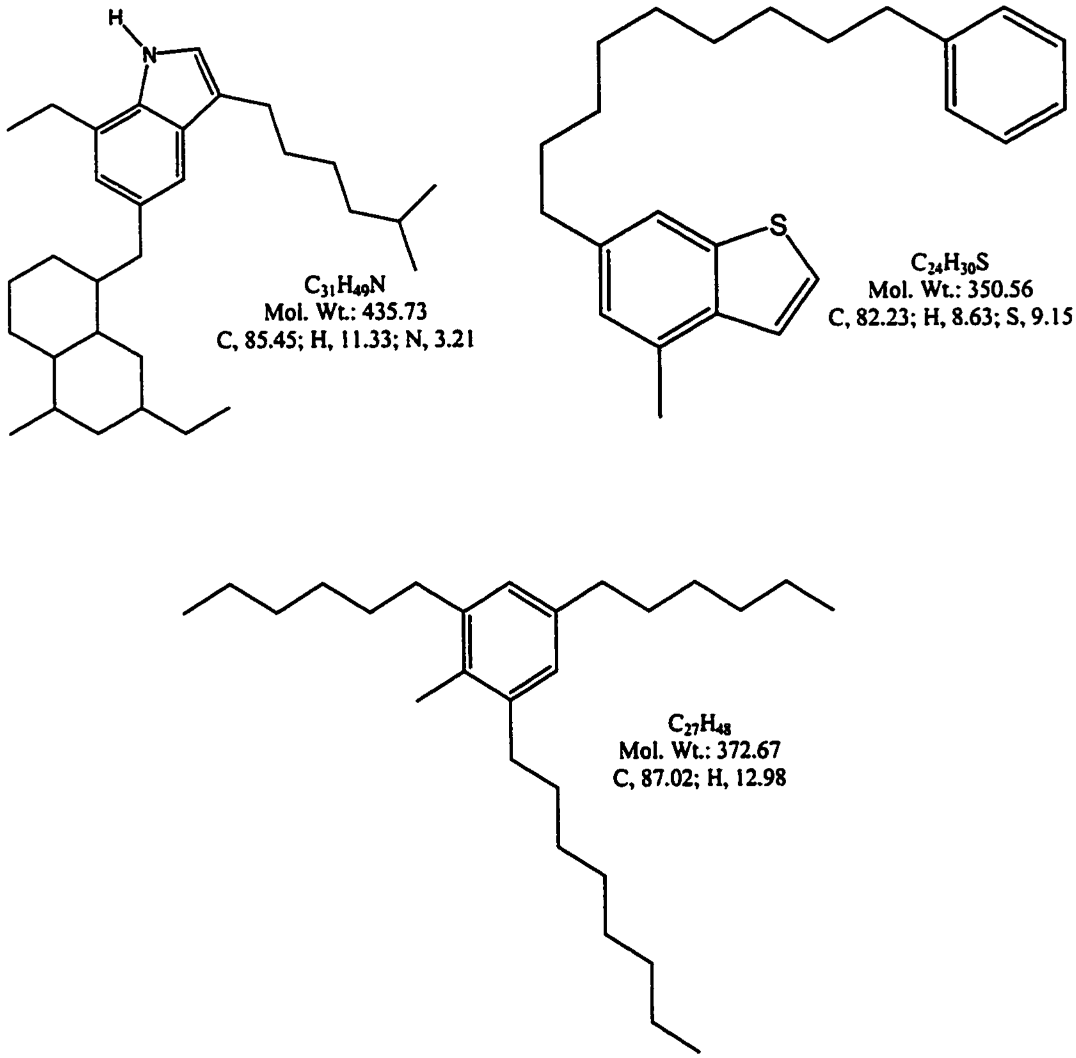


Figure 40: Simulated molecular weight distribution for the third gas oil representation. The simulated distribution curve was divided into ten discrete regions. Twenty one molecules were necessary to represent this fraction.



$C_{31}H_{49}N$
Mol. Wt.: 435.73
C, 85.45; H, 11.33; N, 3.21

$C_{24}H_{30}S$
Mol. Wt.: 350.56
C, 82.23; H, 8.63; S, 9.15

$C_{27}H_{48}$
Mol. Wt.: 372.67
C, 87.02; H, 12.98

Figure 41: Some molecules found in the first Athabasca gas oil representation.

4.3.3 Middle Distillate Representation

An assumption that was made when creating the residue and gas oil representations was that all of the molecules in the representations contained at least one aromatic group. For the middle distillate fraction, each individual aromatic attribute (benzene, phenanthrene, indole, N-indole, benzofuran, and benzothiophene) on average contains 6.5 aromatic carbons. However, the gas oil fraction contains an average of 2.61 aromatic carbons for every middle distillate molecule. Based upon these data, only 40% of the middle distillate molecules contain an aromatic group. It is therefore evident, that a large percentage of middle distillate molecules do not contain any aromatic attributes. To account for this, both naphthenic molecules (alkyl decalin and sterane) and aromatic containing molecules generated by gas oil molecule making algorithm were supplied to the gas oil optimisation routine. Aromatic clusters did not occur in the middle distillate representation due to their low frequency of appearance.

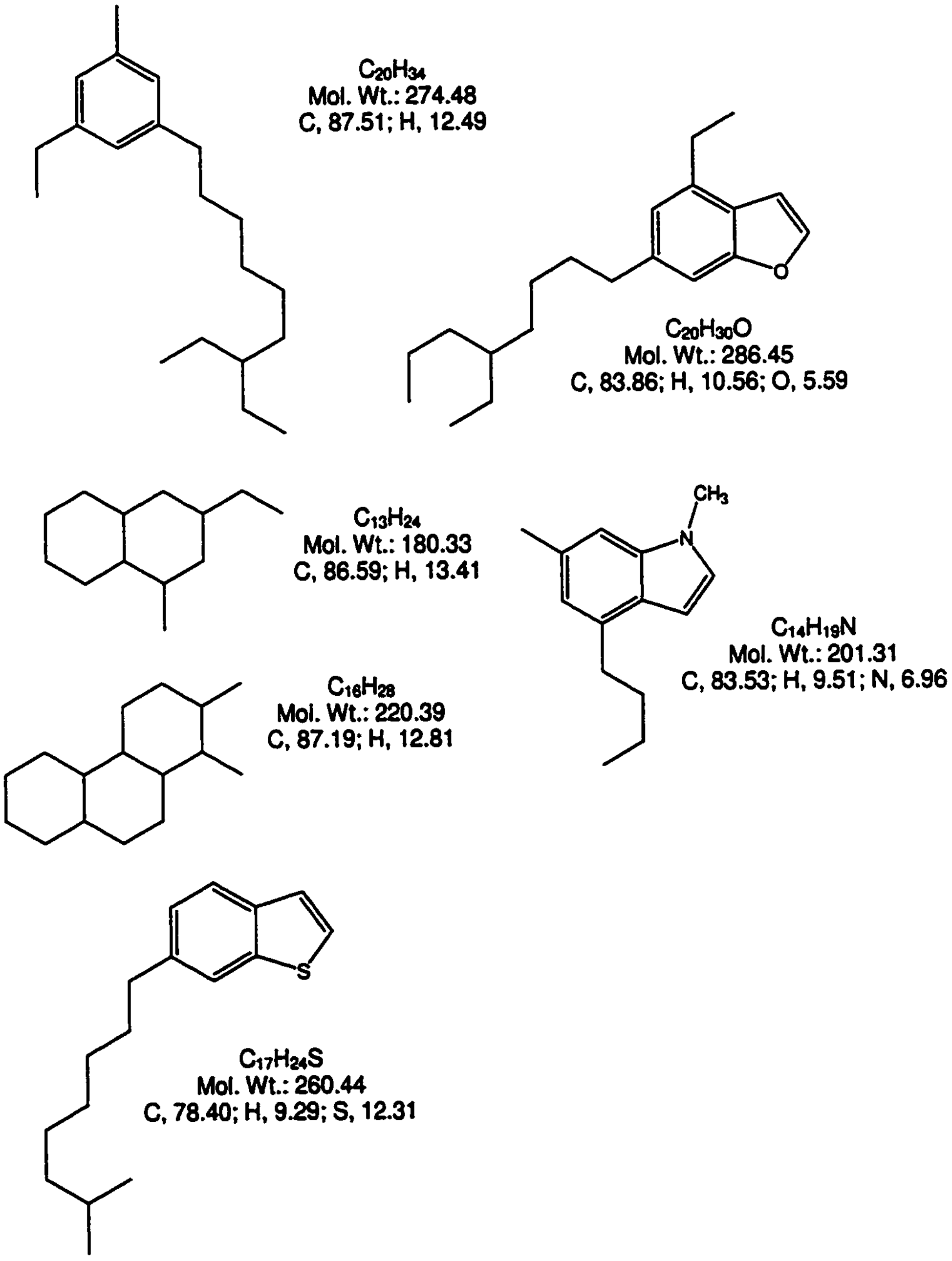
More molecules (sixteen) were needed for the middle distillate representation (Table 13) compared to the numbers needed to create the residue representation (8). A 11.94% deviation relative to the experimentally determined molecular weight was observed (Table 14). The only other property with significant deviation was naphthenic carbon (11.51% deviation). All other property deviations were minimal. A few molecules from the gas oil representation are found in Figure 42.

Table 13: Molecular representation for the Athabasca middle distillate.

Molecule Number	% of representation (mole basis)	MW (g/mol)	C	H	O	N	S
1	27.3	180.32	13	24	0	0	0
2	14.6	220.40	16	28	0	0	0
3	8.7	262.48	19	34	0	0	0
4	0.2	288.52	21	36	0	0	0
5	2.1	290.53	21	38	0	0	0
6	0.1	260.46	19	32	0	0	0
7	0.2	229.36	16	23	0	1	0
8	3.2	276.50	20	36	0	0	0
9	2.3	201.31	14	19	0	1	0
10	3.2	276.50	20	36	0	0	0
11	3.4	215.34	15	21	0	1	0
12	2.1	232.41	17	28	0	0	0
13	0.2	318.59	23	42	0	0	0
14	21.9	206.37	15	26	0	0	0
15	4.6	260.42	18	28	1	0	0
16	5.7	340.57	23	32	0	0	1

Table 14: Predicted versus experimental properties of the Athabasca middle distillate representation.

Property	Predicted	Experimental	% deviation
avg molecular weight (g/mol)	223.88	200	11.94
avg total naphthenics (mol C/100 g)	3.73	3.348	11.51
avg total aromatic (mol C/100 g)	1.31	1.305	0.13
avg total alpha (mol C/100 g)	0.51	0.511	0.02
avg total paraffinic (mol C/100 g)	1.22	1.221	0.18
avg % C	86.09	85.71	0.44
avg % H	12.40	12.78	2.98
avg % O	0.33	0.33	0.58
avg % N	0.37	0.37	0.11
avg % S	0.81	0.81	0.22
Total			28.11



$C_{20}H_{34}$
Mol. Wt.: 274.48
C, 87.51; H, 12.49

$C_{20}H_{30}O$
Mol. Wt.: 286.45
C, 83.86; H, 10.56; O, 5.59

$C_{13}H_{24}$
Mol. Wt.: 180.33
C, 86.59; H, 13.41

$C_{14}H_{19}N$
Mol. Wt.: 201.31
C, 83.53; H, 9.51; N, 6.96

$C_{16}H_{28}$
Mol. Wt.: 220.39
C, 87.19; H, 12.81

$C_{17}H_{24}S$
Mol. Wt.: 260.44
C, 78.40; H, 9.29; S, 12.31

Figure 42: Molecules found in the Athabasca middle distillate representation.

4.4 Molecular Mechanics

The chemical and physical feasibility of the molecules generated in this study was tested using molecular dynamics. Cerius 2 running on a Silicon Graphics Octane workstation was used to analyse the proposed bitumen structures. Cerius 2 uses molecular dynamics analysis to arrange the arrangement of a molecule into a minimal energy conformation (Figure 43). Once the minimal energy conformation of a molecule is determined, Cerius 2 calculates the total energy associated with the molecule. The total energy calculation includes both potential and kinetic energies. It includes bond stretch, angle, coulombic, and other energies. The total energy is expressed relative to the relaxed unstrained molecule (or molecules) at 0 K.

A residue molecule with a molecular weight of 1156 g/mol was analysed. All molecular mechanics analysis was calculated at 500 K. When this molecule is placed in a simulated vacuum (Figure 43) Cerius 2 calculated the minimised energy to be 454.9 Kcal/mol. A few simulations were done with two molecules in vacuum. When the two molecules are placed in close proximity and in certain orientations they can potentially associate. In one simulation (Figures 44 and 45), after the molecular dynamics simulation was run to convergence, the two molecules were very tightly associated (Figure 44). In other simulations the two molecules dissipated into the simulated vacuum and were far apart at the end of the simulation. The total energy of the associated cluster was calculated to be 889.5 kcal/mol. It would be expected that the total energy of the two associated molecules would be less than twice that of a single unassociated molecule. The two molecules associate to minimise the total energy of the system (in the case of this simulation, the two molecules). For an asphaltene molecule with a molecular weight of 6239 g/mol Murgich *et al.* (1999) calculated a minimised energy of approximately 3100 Kcal/mol at 300K. Thus the energies of the systems in this study seam reasonable. For the case where the two asphaltene molecules associated together, it was observed that aromatic rings of one molecules would stack directly above those of another (Figure 46). This is indicative of Π-Π bonding occurring between the two molecules. Stacking of aromatic rings of the same molecule was also observed. This is indicative of intermolecular Π-Π bonding.

The low total energies associated with the residue molecules supports the notion that the molecules produced in this study are both chemically and physically feasible. The molecular mechanics simulations offer evidence that bitumen molecules are held together by Π-Π bonding. However, further simulations would have to be carried out in order to verify this, as both molecule-molecule association and Π-Π bonding were observed in a single simulation.

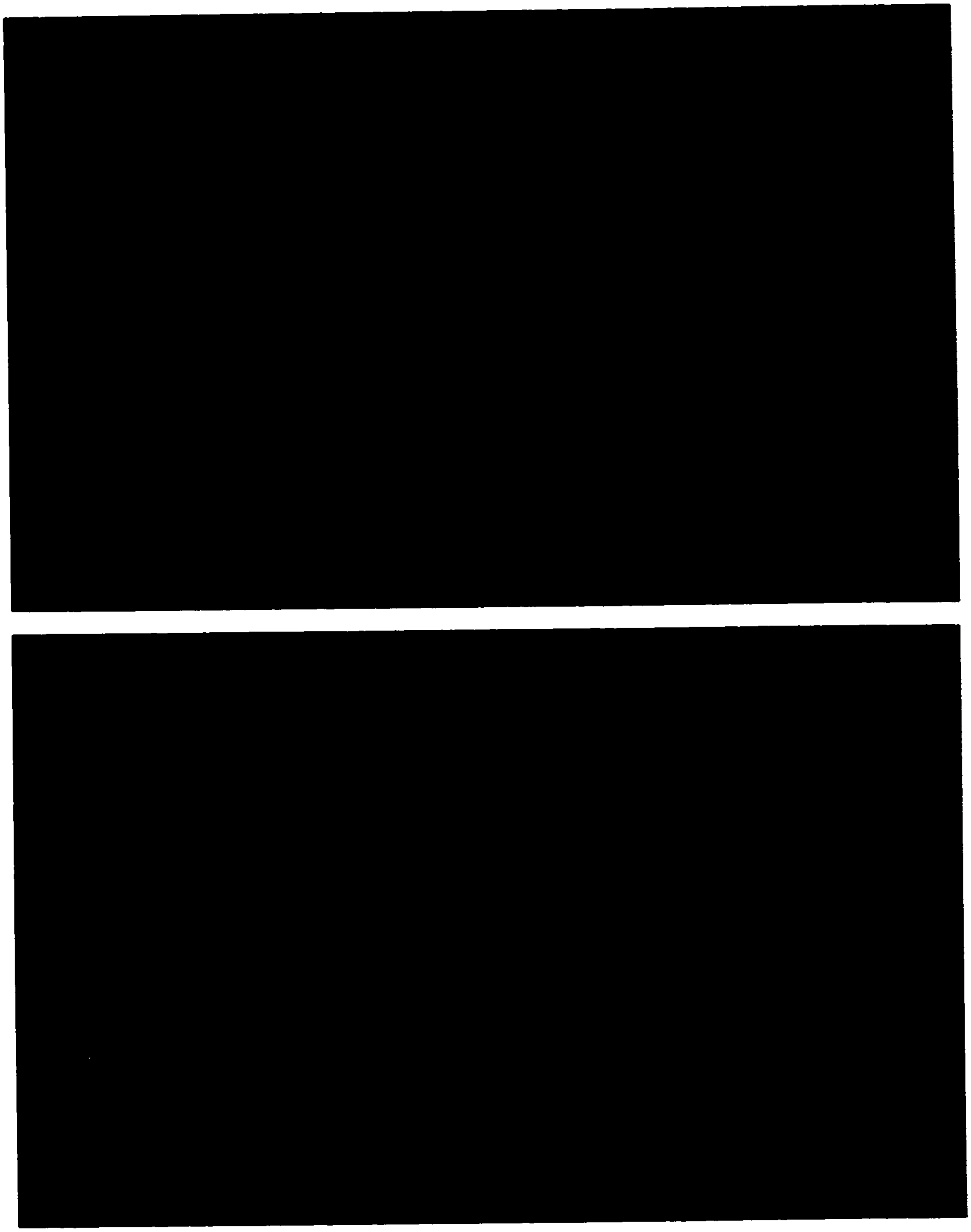


Figure 43: Energy minimisation. The heavy oil molecule in the top panel is fully extended and is not in a minimal energy state. On the other hand the molecule in the bottom panel is in a minimal energy state and is very compact. Special attention should be given to the phenanthrene ring system stacked above the *N*-substituted indole ring system. This is indicative of Π-Π bonding.

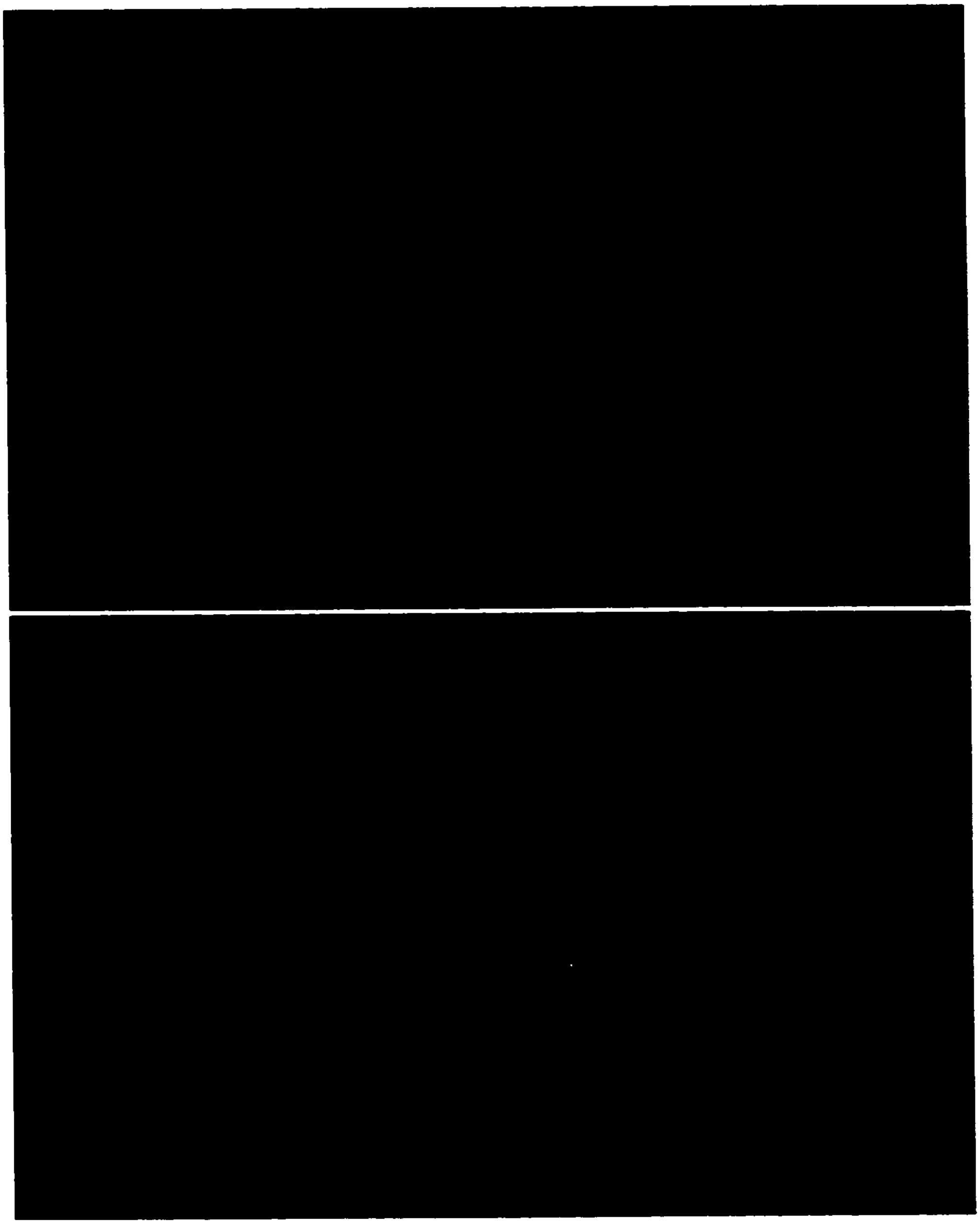


Figure 44: Association of heavy oil molecules in a simulated vacuum. The top half of the figure illustrates the initial placement and conformation of two heavy oil molecules. The bottom panel illustrates the minimum energy conformation of the two molecules. The two molecules in the lowest energy conformations are very tightly associated and very compact. The stacking of aromatic rings is indicative of Π-Π bonding (see Figure 46 for a close up of the aromatic ring stacking).

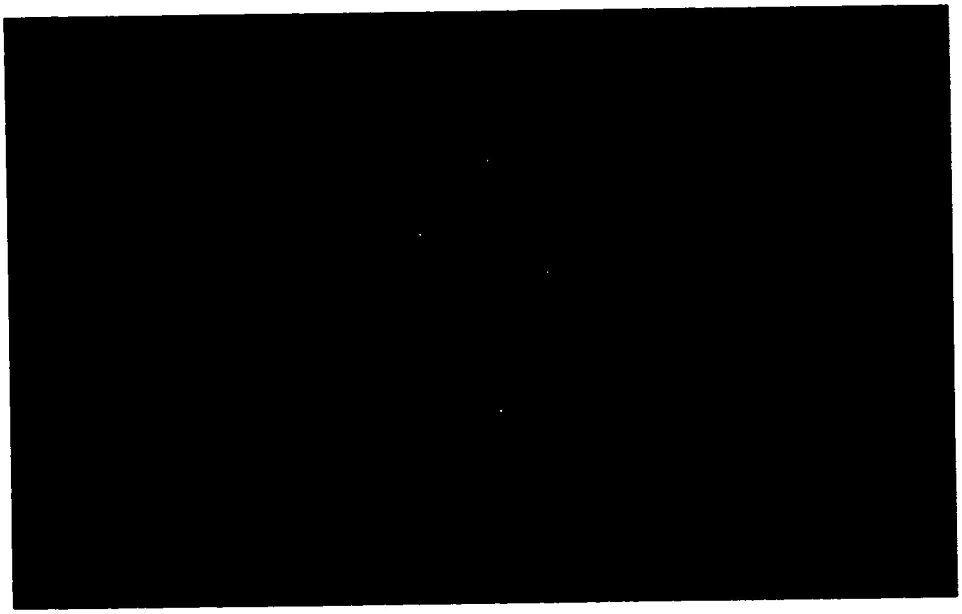


Figure 45: Ball and stick representation. This is an alternative representation of the two asphaltene molecules in their minimal energy conformations.

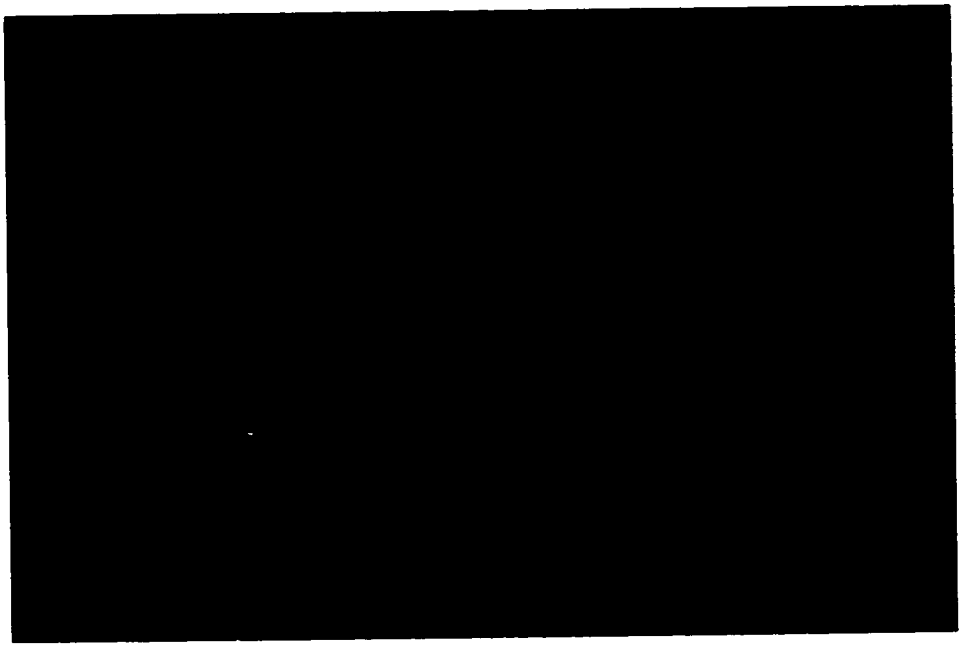


Figure 46: The close association of aromatic ring systems. The close proximity of the central phenanthrene cluster to *N*-substituted indole ring systems from both the same and the separate molecule is indicative of both intramolecular and intermolecular Π-Π bonding.

4.5 Discussion

Many inferences can be drawn from this study. It is possible to represent heavy oil fractions using a limited number of molecules. Molecular representations for three separate heavy oil fractions were created. The properties of the residue, gas oil, and middle distillate, were consistent with the experimentally determined values. It is also suggested that the molecular representations for higher boiling point fractions require fewer molecules than those for lower boiling point fractions. Only eight molecules were required to represent the residue fraction while 16 were needed to represent the middle distillate. This study also makes it evident that molecular dynamics is valuable in analysing heavy oil structures. Using molecules produced in this study, molecular dynamics analysis provided insights into the three dimensional structure of heavy oil molecules. It also provided insights into both heavy oil intramolecular and intermolecular association. In general, this study makes it very clear that to create heavy oil molecular representations, both detailed experimentation and computational modeling is required.

The vector-connection matrix approach used to represent molecules in this study is very efficient. Rather than taking into account all atoms of a molecule, only attributes of structural importance are considered. An alternative to the vector-connection matrix method is the n by n matrix method. The "n by n" method uses a single matrix to describe the connectivity of each non-hydrogen atom in a molecule to every other atom. Thus if a molecule contains n non-hydrogen atoms, the matrix that describes the molecule will contain n*n total entries. Most gas oil molecules contain at least 27 non-hydrogen atoms while most residue molecules contain at least 80 non-hydrogen atoms. The molecule in Figure 34 is represented by the structural and connectivity matrixes found in Figures 32 and 33 respectively. 121 matrix entries were necessary to define the molecule illustrated in Figure 34. It could be argued that the matrices found in Figures 32 and 33 do not specify the exact molecular arrangement. An additional connection matrix could be added to define the exact connection of each chemical attribute. The benzene ring and N-indole ring system could be given three potential connection sites each. Single connection sites would be given to the α-methyl, and the 2 α-methylenes. In total a nine by nine adjuvant connection matrix would be required to specify the exact connection of the molecule in Figure 34. Thus, a total of 202 matrix entries would be necessary to define the molecule. Using the n*n matrix methodology, a 39*39 atom connectivity matrix (1521 matrix entries) would be required to define an equivalent molecule. Thus using the vector approach in this thesis, which is similar to that of Quann and Jaffe (1991), 1300 fewer matrix entries would be required to represent the heavy oil molecule found in Figure 34.

For molecular kinetic simulations, it is very important to represent each molecule efficiently. Molecular kinetic simulations are very computationally demanding. Computationally efficient representations require less memory for storage. By freeing up memory, the time it takes for the kinetic simulation to finish is reduced. Also, it would be much simpler to create a kinetic simulator when efficient representations are used. Using the n by n connection matrix methodology, a 100 carbon residue molecule would require 10,000 matrix inputs. Thus, a kinetic simulation that involves 10,000 molecules would require 100,000,000 matrix inputs! The efficient representation of molecules proposed in this thesis will promote and ease the development of future heavy oil molecular kinetic simulations.

Based upon this study it is evident that heavy oil can be represented using a limited number of structures. Thirty-three molecules were necessary to define all four hydroprocessed fractions (middle distillate, gas oil, residue). The representations in this study indicate that an inverse correlation exists between the average molecular weight of a fraction and the number of molecules it takes to represent it. The optimisation routines used in this study assigned mole fractions of 0 to the majority of the molecules. Thus, the optimisation routines create a representation with the minimum number of molecules necessary to represent a fraction. For the three fractions, the smaller the average molecular weight of the fraction the more molecules it took to represent the fraction (Figure 47). This observation would be expected. Larger molecules contain more chemical structural attributes per molecule than smaller molecules. For example, some residue molecules contain oxygen, nitrogen, and sulphur. A single residue may contain 15 different structural attributes. Conversely, ten middle distillate molecules may be necessary to achieve 15 different structural attributes. Most molecules in the residue representation contain at least two different heteroatoms. For the gas oil and middle distillate fractions no molecules contained two different heteroatoms. The optimisation algorithm would naturally need more different molecules to create representations for the fractions that have fewer attributes per molecule.

For heavy oil molecules as molecular weight increases, the complexity of modeling the molecules increases as well. With very large molecules, chemical feasibility becomes a very important issue. Quite often the probability distribution functions of certain chemical attributes have to be resampled by the molecule creation algorithm to maintain chemical and physical feasibility. Pdf resampling sometimes causes deviations of the predicted properties from the experimental properties. With smaller molecules, a much larger percentage of all chemical attribute pdf samplings result in physically and chemically feasible molecules without resampling. The smaller molecular weight molecules contain fewer structural attributes/molecule. As a result the optimisations associated with them are more complex compared to the larger molecules. When the optimisation and resampling issues are taken into account, it would be predicted that the fraction representations with intermediate molecular weights (middle distillate and gas oil) would be more consistent with the analytical data than the very heavy (residue) and very light (naphtha) representations. This is what was observed in this study (Figure 48).

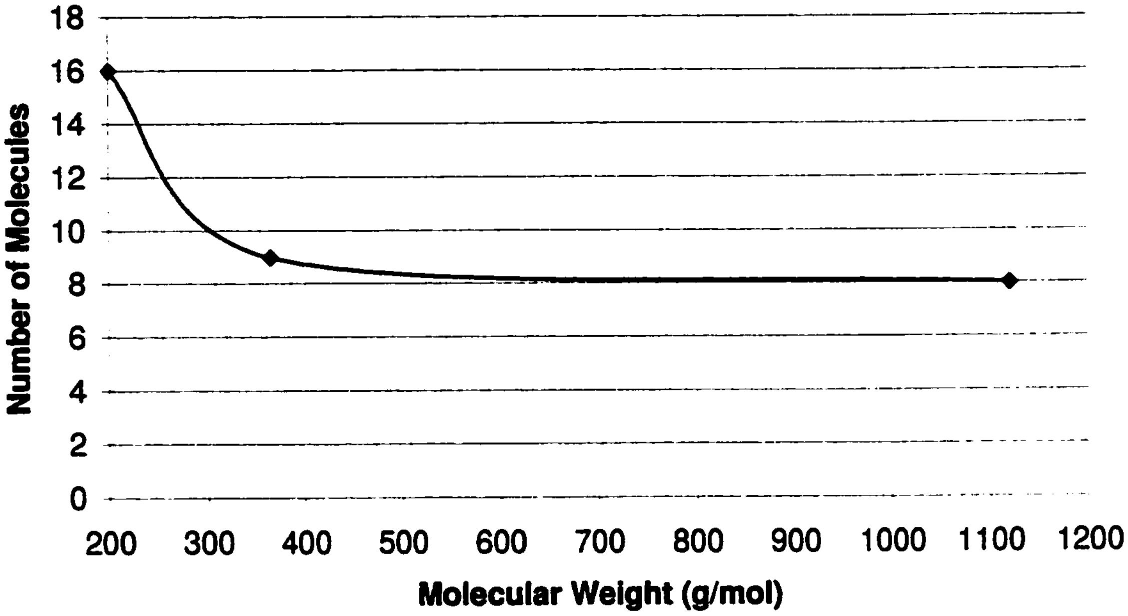


Figure 47: The relationship between the number of molecules it takes to represent a fraction and the molecular weight of the fraction.

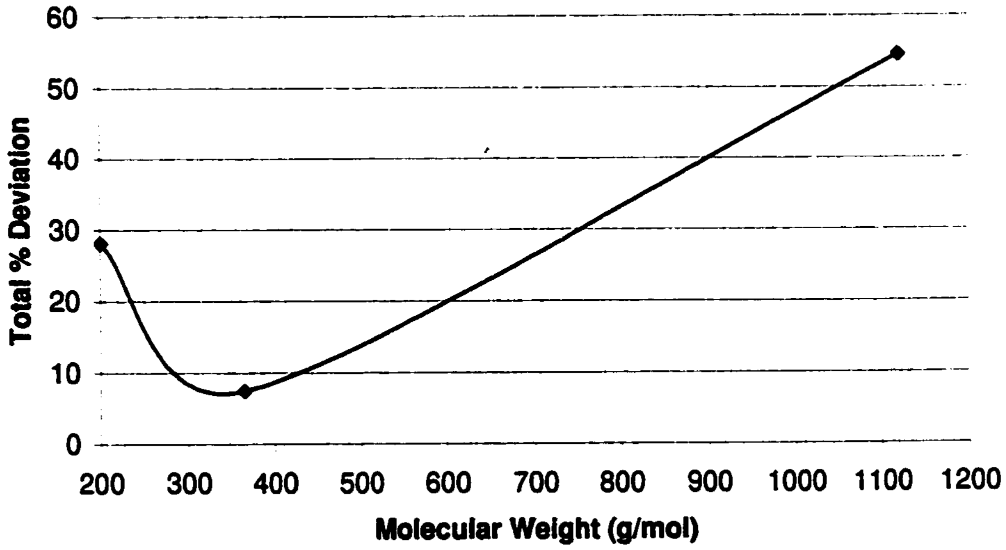


Figure 48: The relationship between total percent deviation and molecular weight.

The heavy oil molecules created in this study were both chemically and physically feasible. A wide variety of analytical data was used to create the molecular representations in this study. VPO, elemental analysis, ^{13}C-NMR, and ^{1}H-NMR, were all utilised. By using a variety of analytical data, the chemical composition, structure, and thus reactivity, would be modelled more precisely. Various sampling constraints and connection strategies were used to assure that all molecules were chemically feasible. The molecular mechanics analysis indicates that the generated molecules were also physically feasible. No other molecular representation study to date has used molecular mechanics to analyse physical feasibility. With notable exceptions such as the Khorahesh *et al.* (1998) representation, most previous representations have been composed of unrealistic chemical attributes. When using very simple chemical attributes, almost all arrangements of the resulting molecules are chemically feasible.

Based upon analytical data available at the time, Yen (1977) postulated that intermolecular association has a major role in asphaltene micelle association (Figure 16). Such intermolecular association could be the result of a combination of hydrogen bonding, Π-Π bonding, hydrophobic interactions, and van der Waals interactions. In one molecular dynamics simulation, two molecules in their final spatial orientations were very tightly associated (Figures 44 and 45). Although no firm conclusions can be made based upon this observation, it offers evidence that the heavy oil molecules generated in this study have physically realistic properties.

The gas oil representation documented by Khorashesh *et al.* (1998) is the most detailed and analytically consistent heavy oil representation published to date. Their representation is based upon ^{1}H-NMR, ^{13}C-NMR, and elemental analysis data. Fourteen different structural attributes are the building blocks of this representation. The only two chemical inaccuracies in this representation is that nitrogen was only found in amide side groups and that no structural attributes contained oxygen. No aromatic attributes (such as indole or N-indole) contained nitrogen. Similar to the gas oil representation proposed in this thesis, no aromatic clusters were pericondensed. In general, the overall structure of the molecules found in Khorashesh *et al.* (1998) were very similar to those found in the gas oil representation presented in this thesis.

The petroleum representations proposed Campbell and Klein (1997) are based on limited data and inconsistent with current petroleum analytical information. Only ^{1}H-NMR was used in their study. As a result, the attributes that create the representations in their model were limited to overly simplified attributes including aromatic ring number, number of sidechains, length of sidechains, etc. The building attributes of the Campell and Klein (1997) representation only include benzene rings, thiophenes, cyclohexanes, paraffinic methylenes, and thioether sulphur. Paraffinic branching is not accounted for (all paraffinic chains are linear). The position of sulphur within thioether containing chains is arbitrary. The only two aromatic attributes appearing in the Campbell and Klein (1997) representations are benzene rings with or without sulphur. Conversely, benzene, phenanthrene, benzothiophene, benzofuran, indole, and N-substituted indole, were used in the study presented in this thesis. All the molecules in the Campbell and Klein (1997) study contain a single aromatic (usually pericondensed) core. Thus, even though Campbell and Klein (1997) used a population of molecules to represent a petroleum fraction, the structure of the molecules are lumped around a central aromatic core. Conversely, in the study documented in this thesis, aromatic attributes and thioethers are distributed randomly throughout molecules. The simplicity of the Campbell and Klein (1997) molecules makes chemical and physical feasibility an unimportant issue. However, the large number of attributes involved, and the sampling strategies used, in the study in this thesis makes it necessary to use chemical and physical constraints when creating molecules. The only advantage of the Campbell and Klein study over the one in this thesis is that it incorporates molecular weight distributions determined from simulated distillations. In summary, the molecular representations of Campbell and Klein (1997) are chemically very simplistic compared to those presented in this thesis.

It is evident that to create molecular representations both detailed experimental and computational work is required. The quality of molecular representations is directly dependent upon the quality of analytical data. As technology evolves, and analytical techniques improve, so will the quality of molecular representations. It is possible that simulations based on molecular representations will make certain costly experiments unnecessary. Molecular representations will ultimately help experimentalists design better experiments and assist in the interpretation of analytical, structural, and reactivity data.

Chapter 5
Conclusions

A series of algorithms and optimisation routines were developed that create chemically and physically feasible representations for bitumen fractions. The representations created are chemically more realistic and diverse than those previously published in the literature. Compared to other bitumen representations, a more efficient matrix and vector framework was used to represent each molecule. Based upon the results of simulations using this algorithm it was determined that there is an inverse relationship between molecular weight and the number of molecules it takes to represent a fraction. Compared to middle distillate (light) and residue (heavy) representations, intermediate molecular weight fractions (gas oil) had representations that were more consistent with experimental data.

Chapter 6

Bibliography

Altamirano, *et al.* (1986). *IMP Bulletin.*

Ayasse, A. Nagaishi, H. Chan, E. Gray, M. (1997). Lumped kinetics of hydrocracking of bitumen, *Fuel*, 76, 1025-1033.

Bartle, K.D. Martin, T.G. Williams, D.F. (1975). Chemical nature of a supercritical-gas extract of coal at 350°C, *Fuel*, 54, 226-235.

Boyd, M.L. Montgomery, D.S. (1962). Structural group analysis of the asphaltene and resin components of the Athabasca bitumen, *Fuel*, 41, 335-350.

Brown, J.K. Ladner, W.R. Sheppard, N. (1960)a. A study of the hydrogen distribution in coal-like materials by high-resolution nuclear magnetic resonance spectroscopy I-The measurement and interpretation of the spectra, *Fuel*, 39, 79-86.

Brown, J.K. Ladner, W.R. Sheppard, N. (1960)b. A study of the hydrogen distribution in coal-like materials by high-resolution nuclear magnetic resonance spectroscopy II-A comparison with infra-red measurement and the conversion to carbon structure, *Fuel*, 39, 87-96.

Camp,F.W. (1976). *The Tar Sands of Alberta, Canada,* Colorado: Cameron Engineers Inc.

Campbell, D. Klein, M. (1997). Construction of a molecular representation of a complex feedstock by Monte Carlo and quadrature methods. *Applied Catalysis: A General*, 160, 41-54.

Carbognani *et al.* (1992) *INTEVEP S.A. Tech. Rept.*

Carlson, P.H. J. Katsuki, T. Martin, V.S. Sharpless, K.B. (1981). A greatly improved procedure for ruthenium tetraoxide catalysed oxidations of orgainc compounds. *Journal of Organic Chemistry,* 46, 3936-3938.

Champagne, P.J. Manolakis, E. Ternan, M. (1985). Molecular weight distribution of Athabasca bitumen. *Fuel*, 64, 423-425.

Chung, K.H. Xu, C.M. Narrow-cut characterization reveals process chemistry. From a paper submitted to *Fuel*: special issue of "Heavy-Oil Upgrading, Renewable Energy and Catalysis".

Dickie, J.P. Yen, T. F. (1967). Macrostructures of the asphaltic fractions by various instrumental methods. *Analytical Chemistry*, 39, 1847-1852.

Dickinson, E.M. (1980). Structural comparison of petroleum fractions using proton and 13C n.m.r. spectroscopy, *Fuel*, 59, 290-294.

Djerassi, C.J. and Engle, R. R. (1953). Oxidations with ruthenium tetroxide, *Journal of the American Chemical Society*, 75, 3838.

El-Mohamed, S. Achard, M.F. Hardouin, F. Gasparoux, H. (1986). Correlations between diamagnetic properties and structural characters of asphaltenes and other heavy petroleum products. *Fuel*, 65, 1501-1504.

Gray, M. R. Krzywicki, A.K. Wanke, S.E. (1992). Chemical transformation during resid upgrading: catalytic and thermal. *Report to CANMET*, SSC # 06SQ.23440-7-9011.

Gray, M.R. (1994). *Upgrading Petroleum Residues and Heavy Oils.* New York: Marcel Dekker, Inc.

Ignasiak, T. Kemp-Jones, A.V. Strausz, O.P. (1977). The molecular structure of the carbon-sulphur bonds by radical ion electron transfer reactions, *Journal of Organic Chemistry*, 42, 312-320.

Joshi, P. Iyer, S. Klein, M. (1998). Automated mechanistic kinetic modeling of gas oil catalytic cracking, *Reviews in Process Chemistry and Engineering*, 1, 111-140.

Khorasheh, F. Gray, M.R. Dalla Lana, I.G. (1987). Structural analysis of Alberta heavy gas oils, *Fuel*, 66, 505-511.

Khorasheh, F. Khaledi, R. Gray, M.R. (1998). Computer generation of representative molecules for heavy hydrocarbon mixtures, *Fuel*, 77, 247-253.

Lehninger, A.L. (1977). *Biochemistry.* New York : Worth Publishers, Inc.

Milan, L. Chu, Y. Ruo, T. Strausz, O.P. (1977). Chemical composition of Athabasca bitumen, *Fuel*, 56, 369-381.

Mojelsky, T.W. Montgomery, D.S. Strausz, O.P. (1986). The side chains associated with the undistillable aromatic and resin components of Athabasca bitumen. *AOSTRA Journal of Research*, 2, 177-184.

Mojelsky, T.W. Ignasiak, T.M. Frakman, Z. McIntyre, D.D. Lown, E. M. (1992). Montgomery, D. S. Strausz, O.P. Structural features of Alberta oil sand bitumen and heavy oil asphaltenes. *Energy and Fuels*, 6, 83-96.

Montgomery, D.S. Boyd, M.L. (1959). New method of hydrocarbon structural group analysis, *Analytical Chemistry*, 31, 1290-1298.

Murakami, K, Okumura, M. Yamamoto M. Sanada, Y. (1996). Structural analysis of mesophase pitch with high-resolution, high-temperature ^{13}C-NMR, *Carbon*, 34, 187-192.

Murgich, J. Abanero, J. A. Strausz, O. P. (1999). Molecular recognition in aggregates formed by asphaltene and resin molecules from the Athabasca oil sand, *Energy and Fuels*, 13 278-286.

Nellensteyn, F.J. (1933). Theoretical aspect of the relation of bitumen to solid matter. *Proceedings/World Petroleum Congress*, 2, 616-618.

Neurock, M. Libanati, C. Klein, M.T. (1989). Modeling asphaltene reaction pathways: intrinsic chemistry. *AIChE Symposium Series*. 273(85). 7-14.

Neurock, M. Nigam, A. Trauth, D. Klein, M.T. (1994). Molecular representation of complex hydrocarbon feedstocks through efficient characterization and stochastic algorithms, *Chemical Engineering Science*, 49, 4153-4176.

Oeng, P. Fu, J. Sheng, G. Morales-Izquirdo, A. Lown, E. M. Strausz, O.P. (1999). Ruthenium-ions-catalysed oxidation of an immature asphaltene: structural features and biomarker distribution, *Energy and Fuels*, 13, 266-277.

Oka, M. Chang, H.C. Gavalas, G. R. (1977). Computer-assisted molecular structure construction for coal derived compounds, *Fuel*, 56, 3-8.

Outrim, C.P. Evans, R.G. (1977). Alberta's oil sands reserves and their evaluation, *The Oil Sands of Canada-Venezuela*, CIM Special Issue, 17, Redford, D.A. Winestock, A.G. Eds.

Payzant, J.D. Rubinstein, I. Hogg, A.M. Strausz, O.P. (1979). Field-ionization mass spectrometry: application to geochemical analysis, *Geochimica et Cosmochimica Acta*, 43, 1187-1193.

Payzant, J.D. Lown, E.M. Strausz, O.P. (1991). Structural units of Athabasca asphaltene: the aromatics with a linear carbon framework, *Energy and Fuels*, 5, 445-453.

Peng, P. Morales-Izquirdo, A. Hogg, A. Strausz, O.P. (1997). Molecular structure of Athabasca asphaltene: Sulfide, ether, and ester linkages, *Energy and Fuels*, 11, 1171-1187.

Quann, R.J. Jaffe, S.S. (1992). Structure-oriented lumping: Describing the chemistry of complex hydrocarbon mixtures, *Industrial and Engineering Chemical Research*, 31 2483-2497.

Ramadhan, O.M. Burker, R.A. Thanoon, A.A. (1997). A study of aromatic hydrocarbons in Qaiyarah naphtha fraction 50-165°C using gas chromatocraphy/mass spectroscopy, *Petroleum Science and Technology*, 15 (1&2), 127-135.

Rubinstein, I. Spyckerelle, C. Strausz, O. P. (1979)a. Pyrolysis of asphaltenes: a source of geochemical information, *Geochimica et Cosmochimica Acta*, 43, 1-6.

Rubinstein, I. Strausz, O. P. (1979)b. Thermal treatment of the Athabasca oil sand bitumen and its component parts, *Geochimica et Cosmochimica Acta*, 43, 1887-1893.

Snape, C.E. Ladner, W.R. (1979). Survey of carbon-13 chemical shifts in aromatic hydrocarbons and its application to coal-derived materials, *Analytical Chemistry*, 51, 2189-2198.

Speight, J.G. (1970). A structural investigation of the constituents of Athabasca bitumen by proton magnetic resonance spectroscopy, *Fuel*, 49, 76-90.

Speight, J.G. (1971). Structural analysis of Athabasca asphaltenes by proton magnetic resonance spectroscopy, *Fuel*, 50, 102-112.

Speight, J.G. Moschopedis, S.E. (1979). Some observations on the molecular "nature" of petroleum asphaltenes. American Chemical Society Division of Petroleum Chemistry Preprints. 24(4). 910-918.

Speight, J.G. (1980). *The Chemistry and Technology of Petroleum,* New York: Marcek Dekker, Inc.

Strausz, O.P. Jha, K.N. Montgomery, D.S. (1977). Chemical composition of gases in Athabasca bitumen and in low-temperature thermolysis of oil sand, asphaltene and maltene, *Fuel*, 56, 113-120.

Strausz, O.P. (1989). Structural studies on resids: Correlation between structure and reactivity, *AIChE Symposium Series*, 85, 1-6.

Strausz, O.P. Mojelsky, T. Lown, E. (1992). The molecular structure of asphaltene: an unfolding story, *Fuel*, 71, 1355-1363.

Strausz, O.P. Mojelsky, T.W. Faraji, F. Lown, E.M. (1999). Additional structural details on Athabasca asphaltene and their ramifications, *Energy and Fuels*, 13, 207-227.

Sullivan, R. Boduszynski, M. Fetzer, J. (1989). Molecular transformations in hydrotreating and hydrocracking, *Energy and Fuels*, 3, 603-612.

Suzuki, T. Itoh, M. Takegami, Y. Watanbe, Y. (1982). Chemical structure of tar-sand bitumens by 13C and 1H n.m.r. spectroscopic methods, *Fuel*, 61, 402-408.

Sawatzky, H. Boyd, M.L. Montgomery, D.S. (1967). The application of hydrocarbon structural group analyses to Mildred Lake and Abasand (Alberta) asphaltenes and their reduction products, *Journal of the Institute of Petroleum*, 53, 162-168.

Yen, F. Erdman, J.G. Pollack, S.S. (1961). Investigation of the structure of petroleum asphaltenes by x-ray diffraction, *Analytical Chemistry*, 33, 1587-1594.

Yen, T.F. Erdman, G. (1962). Investigation of the structure of petroleum asphaltenes and related substances by proton nuclear magnetic resonance, *American Chemical Society Division of Petroleum Chemistry Preprints*, 7, 99-111.

Yen, T.F. Dickie, J.P. (1968). The compactness of the aromatic systems in petroleum asphaltics, *Journal of the Institute of Petroleum*, 54, 50-53.

Yen, T.F. (1972). Present status of the structure of petroleum heavy ends and its significance to various technical applications, *American Chemical Society Division of Petroleum Chemistry Preprints, 17, F102-F110.*

Wang, Z. Liang, W. Que, O. Qian, J. (1997). Structural characterization of Gudao asphaltene by ruthenium ion catalyzed oxidation, *Petroleum Science Technology*, 15, 559-577.