

AD-A219 788

WRDC-TR-89-2119



ADVANCED FUEL PROPERTIES

ALLIED-SIGNAL

K. R. SQUIRE
C. L. VORRES
M. G. KOEHLER
J. A. ZAVATTARO
B. JUSKO
M. J. RYTHOR

UOP

J. T. DONNER
J. W. SOUZA

DTIC
ELECTE
MAR 28 1990
S D D

ALLIED-SIGNAL ENGINEERED MATERIALS RESEARCH CENTER
50 EAST ALGONQUIN RD.
DES PLAINES IL 60017-5016

UOP RESEARCH CENTER
50 EAST ALGONQUIN RD.
DES PLAINES IL 60017-5016

PHASE I INTERIM REPORT FOR THE PERIOD OCTOBER 1987 - MAY 1989

DECEMBER 1989

Approved for public release; distribution is unlimited.

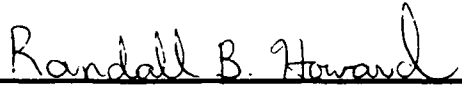
AERO PROPULSION AND POWER LABORATORY
WRIGHT RESEARCH AND DEVELOPMENT CENTER
AIR FORCE SYSTEMS COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OHIO 45433-6F93

NOTICE

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely Government-related procurement, the United States Government incurs no responsibility or any obligation whatsoever. The fact that the government may have formulated or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder, or any other person or corporation; or as conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

This report is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nations.

This technical report has been reviewed and is approved for publication.


RANDALL B. HOWARD, 1Lt, USAF
Fuels Branch
Fuels and Lubrication Division
Aero Propulsion and Power Laboratory


CHARLES L. DELANEY, Chief
Fuels Branch
Fuels and Lubrication Division
Aero Propulsion and Power Laboratory

FOR THE COMMANDER


BENITO P. BOTTERI, Chief
Fuels and Lubrication Division
Aero Propulsion and Power Laboratory

If your address has changed, if you wish to be removed from our mailing list, or if the addressee is no longer employed by your organization please notify WRDC/POSF, WPAFB, OH 45433-6563 to help us maintain a current mailing list.

Copies of this report should not be returned unless return is required by security considerations, contractual obligations, or notice on a specific document.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED			1b. RESTRICTIVE MARKINGS NONE		
2a. SECURITY CLASSIFICATION AUTHORITY N/A			3. DISTRIBUTION/AVAILABILITY OF REPORT APPROVED FOR PUBLIC RELEASE DISTRIBUTION IS UNLIMITED		
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE N/A					
4. PERFORMING ORGANIZATION REPORT NUMBER(S) N/A			5. MONITORING ORGANIZATION REPORT NUMBER(S) WRDC-TR-89-2119		
6a. NAME OF PERFORMING ORGANIZATION ALLIED SIGNAL EMRC		6b. OFFICE SYMBOL (if applicable)	7a. NAME OF MONITORING ORGANIZATION AEROPROPULSION & POWER LABORATORY(WRDC/POSF) WRIGHT RESEARCH DEVELOPMENT CENTER		
6c. ADDRESS (City, State, and ZIP Code) 50 EAST ALGONQUIN RD DES PLAINES IL 60017-5016			7b. ADDRESS (City, State, and ZIP Code) WRIGHT-PATTERSON AFB, OHIO 45433-6563		
8a. NAME OF FUNDING/SPONSORING ORGANIZATION FUELS BRANCH FUELS & LUBRICATION DIVISION		8b. OFFICE SYMBOL (if applicable) WRDC/POSF	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER F33615-87-C-2709		
8c. ADDRESS (City, State, and ZIP Code) WRIGHT-PATTERSON AFB OHIO 45433-6563			10. SOURCE OF FUNDING NUMBERS		
			PROGRAM ELEMENT NO. 62203F	PROJECT NO. 3048	TASK NO. 05
11. TITLE (Include Security Classification) ADVANCED FUEL PROPERTIES					
12. PERSONAL AUTHOR(S) J.T. DONNER, K.R. SQUIRE, C L. VORRES, R.J. SCHOFIELD					
13a. TYPE OF REPORT INTERIM		13b. TIME COVERED FROM 10/87 TO 5/89		14. DATE OF REPORT (Year, Month, Day) 1989 DECEMBER	
15. PAGE COUNT 135					
16. SUPPLEMENTARY NOTATION NONE					
17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number) FUEL PROPERTIES. PHYSICAL PROPERTIES, THERMODYNAMICS, PREDICTIONS, COMPUTATIONAL, AVIATION FUELS		
FIELD	GROUP	SUB-GROUP			
21	04	FUELS			
07	04	P.CHEMISTRY			
19. ABSTRACT (Continue on reverse if necessary and identify by block number) To develop fuels needed for the high performance aircraft of the future. the ability to design fuels based on the satisfaction of a set of needed properties is needed. This report covers the first phase of a project to develop software to accomplish this goal. Phase I covers the ability to determine numerous physical and thermochemical properties of pure-component organic fuel candidates based solely on molecular structure(s) of the molecules of interest. The result of Phase I is the software that obtains various physical and thermochemical properties for liquids, ideal gases, and real gases at various temperatures and pressures. Because of the number of methods that were found for predicting properties of molecules, a priority scheme was developed and utilizes the "best" method to predict the desired property. Additionally, the system has designed within it the ability to interface with an expert system to allow the system to better decide what method(s) should be used to predict each desired property.					
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS			21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED		
22a. NAME OF RESPONSIBLE INDIVIDUAL R.B. HOWARD, 1LT, USAF			22b. TELEPHONE (Include Area Code) (513)255-7423		22c. OFFICE SYMBOL WRDC/POSF

DD Form 1473, JUN 86

Previous editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE

UNCLASSIFIED

FOREWORD

This report is the result of and in satisfaction of Task 9.0 for Phase I of the Advanced Fuel Properties Project.

DISCLAIMER

This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor any agency thereof, nor any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process or service by trade name, mark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of the authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

TABLE OF CONTENTS

1.0	INTRODUCTION	1
2.0	OBJECTIVES	1
3.0	BACKGROUND	1
3.1	Aircraft Fuels.	1
3.2	Predictive Techniques.	1
3.3	Overall Approach.	2
3.4	Historical Perspective.	2
3.5	Interrelationships of Properties.	2
4.0	TASK REVIEW	5
4.1	Definition of Fuel Candidates	5
4.2	Definition of Properties	6
4.3	Data Base Development	14
4.3.1	Literature Review	14
4.3.2	Data Compilation	18
4.3.3	Obtaining Missing Data	28
4.4	Compilation, Evaluation, and Selection of Structure-Property Relationships	30
	Data Base Access Routines	33
	Methods for Structural Inputs	36
	Methods for Single Valued Properties	41
	Introduction to the Methods for Thermodynamic Properties	47
	Methods for Ideal Gases	49
	Methods for Residual Properties	54
	Methods for Real Gases	64
	Methods for Liquid Properties	67
	Methods for Phase Transitions	70
	Methods for Transport Properties	73
	Methods for Solid Properties	74

Methods for Mixtures	74
Methods for Error Tracking	75
4.5 New Model Development	76
4.6 Codification of Modeling Program	90
4.7 Model Testing and Verification	93
Summary of Results of Methods Testing	108
5.0 SOFTWARE DESIGN	112
5.1 Approach	112
5.2 Documentation	116
5.3 FORTRAN77 Standard Compliance	119
5.4 Error Handling	119
5.5 Graphical Input/Output	120
6.0 CONCLUSIONS	121
6.1 Technical Conclusions	121
6.2 Software Conclusions	121
REFERENCES	123

LIST OF FIGURES

Figure 3.5-1	P-T curves for a pure compound	4
Figure 4.3.2-1	Relational Data Base Diagram	20
Figure 4.4-1	API Prediction of Critical Properties and Tabulated Parameters	31
Figure 4.4-2	Prediction of densities and vapor pressures	32
Figure 4.4-3	Chemical Family Classification Scheme	42
Figure 4.4-4	Hydrocarbon Family (C and H only) Classification. . . .	43
Figure 4.4-5	Data Flow for Fluid Thermodynamic Calculations	48
Figure 4.4-6	Ideal Gas Enthalpy of Formation	50
Figure 4.4-7	Cubic Equations of State	56
Figure 4.4-8	Reduced pressure versus reduced volume for the Peng- Robinson equation.	58
Figure 4.4-9	Cubic Equation Subroutines.	59
Figure 4.5-1	Different types of molecular fragments used in graph theory	81
Figure 4.5-2	Plot of experimental normal boiling points (TNBP1) versus ${}^0\chi_p$ (XPA0) for the normal alkanes (Temperature in K)	82
Figure 4.5-3	Experimental (TNBP1) versus predicted normal boiling points where the predictions are based on ${}^1\chi_p$ (Temperatures in K)	83
Figure 4.5-4	Plot of experimental normal boiling points (TNBP1) versus $\ln({}^0\chi_p)$ (LN(XPA0)) for the normal alkanes . . .	83
Figure 4.5-5	Experimental (TNBP1) versus residual normal boiling points where the predictions are based on ${}^1\chi_p$ (Temperatures in K)	86
Figure 4.5-6	Experimental (TNBP1) versus relative errors for normal boiling points where the predictions are based on ${}^1\chi_p$ (Temperatures in K)	86
Figure 4.5-7	Experimental (TNBP1) versus predicted normal boiling points based on ${}^1\chi_p$, ${}^4\chi_p$, ${}^1K$, Shannon index, and ${}^0K$ (Temperatures in K)	87

Figure 4.5-8	Experimental boiling points (TNBP1) versus residuals, where the predictions are based on $^1\chi_p$, $^4\chi_p$, ^1K , Shannon index, and ^0K (Temperatures in K)	87
Figure 4.5-9	Experimental boiling points (TNBP1) versus relative errors; predictions are based on $^1\chi_p$, $^4\chi_p$, ^1K , Shannon index, and ^0K (Temperatures in K)	88
Figure 4.5-10	Experimental versus predicted for all of the hydrocarbon families combined (model basis in text) (Temperatures in K)	88
Figure 4.5-11	Experimental normal boiling points versus relative errors for all hydrocarbons (details in text and 1 of 15 points plotted) (Temperatures in K)	89
Figure 4.5-12	Experimental normal boiling points versus relative errors for all hydrocarbons (details in text and 1 of 15 points plotted) (Temperatures in K)	89
Figure 4.7-1	Percent error in molar volume calculated for CO_2 by using the Redlich-Kwong equation of state with standard parameters ¹³	95
Figure 4.7-2	Percent error in molar volume calculated for CO_2 by using the Redlich-Kwong equation of state with Soave parameters ¹³	96
Figure 4.7-3	Data base (TC1) and predicted (TC6) critical temperatures differences versus experimental critical temperatures	100
Figure 4.7-4	Fractional difference between data base and predicted critical temperatures versus data base critical temperatures	101
Figure 4.7-5	Data base (TC1) versus predicted (TC6) critical temperatures	102
Figure 4.7-6	Data base (TMPSP1) versus predicted (TMPSP2) melting temperatures at 1 atmosphere	104
Figure 4.7-7	Data base (TNBP1) versus predicted (TNBP2) normal boiling temperatures	105
Figure 4.7-8	Critical temperature mean fractional error by family (families 1 - 10)	106

Figure 4.7-9	Critical temperature mean fractional error by family (families 21 - 30)	107
Figure 4.7-10	Percent of predicted values having fractional errors less than 5 percent by property and method	109
Figure 4.7-11	Comparison of the number of data base property values and the number of cases where <u>both</u> data base and predicted values exist	110
Figure 5.1-1	Diagram of priority scheme for an example of obtaining the critical temperature (method name TC) for 10 compounds	113
Figure 5.1-2	Selections from the DEFAULTS.PRI file that is used by the priority scheme in the AFP system (Details in the text)	114
Figure 5.2-1	Programmer's reference sheet, front page	117
Figure 5.2-2	Programmer's reference sheet, back page	118

LIST OF TABLES

Table 4.1-1 AFP Data Base Family Counts	7
Table 4.2-1 Definition of Fuel Properties from RFP	9
Table 4.2-2 Fuel Properties in AFP System	10
Table 4.2-3 Examples of Counts for Experimental Properties	13
Table 4.4-1 Sources of the Methods for Single Valued Properties.	44
Table 4.4-2 Sources for the Methods for Ideal Gas Properties	52
Table 4.4-3 Summary of the Methods for Residual Properties	62
Table 4.4-4 Summary of Methods for Real Gas Properties	65
Table 4.4-5 Summary of Methods for Liquid Properties	68
Table 4.4-6 Comparison Between Peng-Robinson Predictions and Experiment .	71
Table 4.4-7 Summary of the Methods for Liquid-Gas Phase Transitions . . .	73
Table 4.4-8 Summary of Methods for Transport Properties	74
Table 4.4-9 Summary of Methods for Solids	75
Table 4.5-1 Simple Group Types for New Methods Development	77
Table 4.5-2 Graph Theory Parameters and Structural Descriptors for New Methods Development	78
Table 4.5-3 Transformed and Derived Parameters for New Methods Development	79
Table 4.5-4 General Normal Boiling Point Regression Results	85
Table 4.7-1 Distribution of the Number of Predictions at Various Fractional Errors	111

1.0 INTRODUCTION

The successful development and production of fuels needed for the high performance aircraft of the future must overcome two technical hurdles. The source of raw materials is changing from relatively light, paraffinic petroleum to hydrocarbons from other sources that may be much more aromatic and contain higher levels of contaminants. In addition, the performance specifications of the engine and fuel system may extend to regions beyond that attainable by today's fuels.

2.0 OBJECTIVES

The overall objective of this program is to develop a tool that will accurately predict the bulk fuel properties of a complex mixture of hydrocarbons and thereby aid in the design of fuels based on satisfying a set of specified fuel properties.

The objective of Phase I was to be able to predict desired physical and thermochemical properties of pure organic compounds based solely upon the knowledge of their molecular structures.

3.0 BACKGROUND

3.1 Aircraft Fuels. The changing quality of petroleum and the possible introduction of fuels derived from tar sand, oil shale, and coal will place new demands on analytical techniques and specification development. Fuels for future applications may require properties beyond those needed today. To solve these problems, a greater understanding of the relationship of fuel structure at the molecular level and the bulk fuel properties is needed.

3.2 Predictive Techniques. To cope with the complexity of current fuels and the large numbers of potential components of future fuels, the use of mathematical techniques is valuable in studying the structure-property

relationships of fuel components. Graph theory, group additivity, and multivariate statistics are all important tools.

The application of mathematical techniques depends on accurate experimental data. In addition, the broad base of current knowledge which has resulted in what is termed "empirical correlations" is also a valuable technique to augment the more fundamental approach.

3.3 Overall Approach. The strategy used was to make maximum use of available data, generate new data where needed, and use both theories based solely on structure and relationships derived from experiment to predict properties of single compounds. The technical approach for Phase I was designed to not only accomplish the objectives of Phase I, but also to consider how the objectives and results of Phase I logically fit into the overall project objectives. That is, as part of the Phase I objectives, one must always consider the ability to progress logically from single compound modeling to the final objective of computer-aided design of mixtures.

3.4 Historical Perspective. Predicting the physical properties of gases and liquids has long been a major goal of physical chemists. By the early 1950's, accurate structure-based theories had been developed for gas densities, thermodynamics, and transport properties¹; reliable experimental data for gases and liquids were also available from the American Petroleum Institute², National Bureau of Standards³ and JANAF Tables⁴; and the Hougen-Watson Tables permitted predictions of liquid and gas compressibilities and thermodynamic functions⁵. Since the 1950's, increasingly complex correlations for a wide range of properties have been developed^{6,7}, but they still use inputs of both experimental and structure-based data. Apparently now it is possible to predict most of the properties of gases and liquids using only their molecular structures.

3.5 Interrelationships of Properties. The phase diagrams (P-T curve shown in Figure 3.5-1) of all pure compounds have separate regions for solid, liquid,

vapor, and supercritical fluid phases. For temperatures greater than greater than the critical point (point C) or for pressures greater than the critical point with temperatures greater than those on the fus. curve, only a supercritical fluid phase is present. Curve 2-4 indicates where a supercooled liquid exists, and the line is dotted to indicate that this is a metastable phase. Comparisons of these P-T curves for several substances led to formulation of the law of corresponding states¹. According to this empirical law, if the temperature, pressure, and volume are scaled by the critical temperature (T_c), pressure (P_c), and volume (V_c), all substances obey the same equation of state.

Nearly all of the correlations available for the properties of real gases and liquids are based upon the law of corresponding states. For nonspherical and polar molecules, correction factors are also added into the property correlations to consider the shape of the molecules. The most widely used of these "structural parameters" are the acentric factor, ω , the Rackett parameter, Z_{RA} , and the COSTALD parameters, ω_{SRK} and V^* . Careful analysis of the API and AIChE methods for predicting the properties of pure liquids and gases^{6,7} shows that all the properties can be predicted given values of T_c , P_c , Z_{RA} , ω , ω_{SRK} , V^* , and two physical properties: the normal boiling point and liquid density at one temperature. (Note: these correlations also contain parameters which can be calculated directly from molecular structure.) The strategy taken was to develop highly accurate structure-based correlations for these eight key properties since they are used in many other predictive methods.

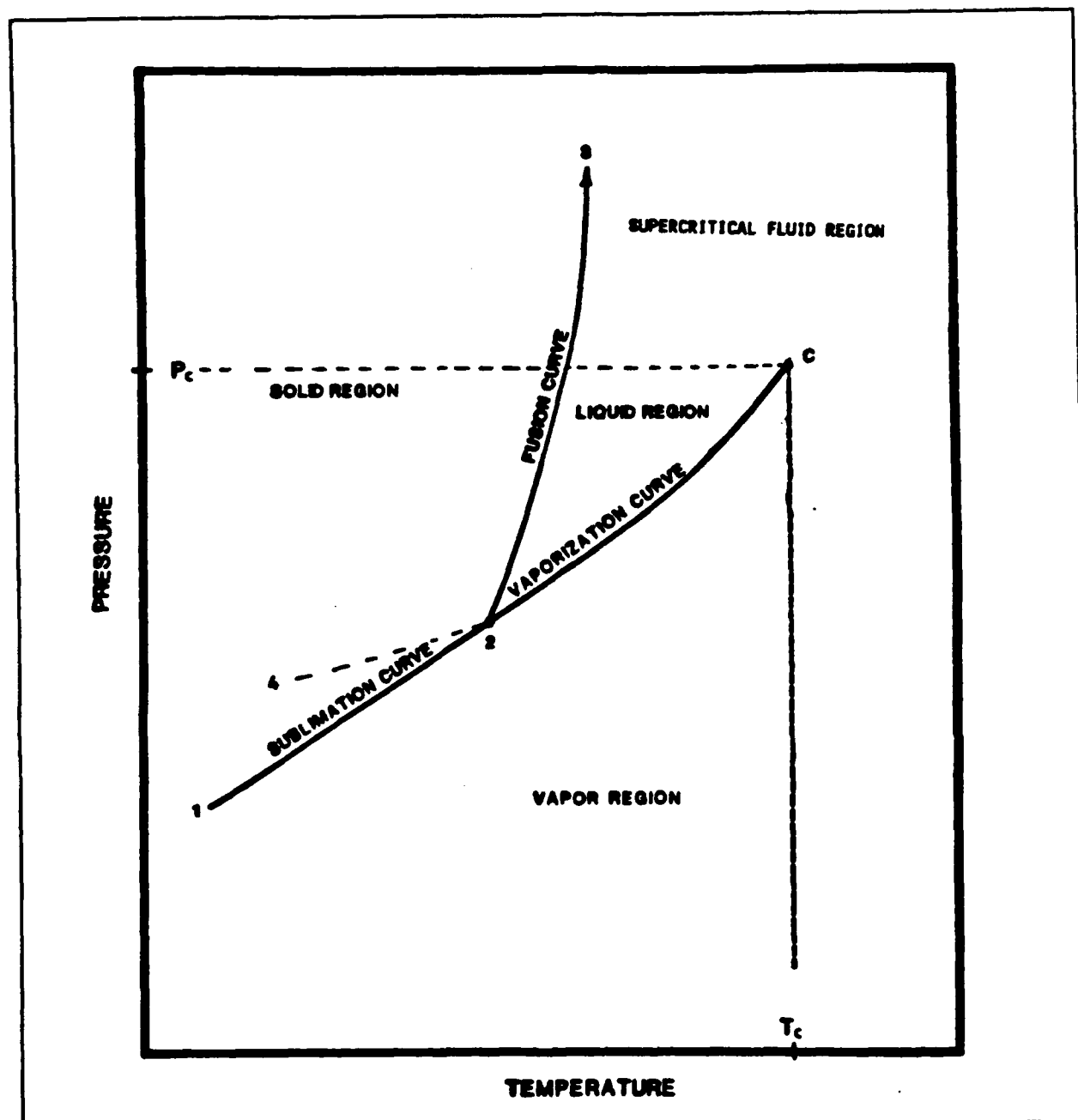


Figure 3.5-1 P-T curves for a pure compound

4.0 TASK REVIEW

This section reviews the Phase I work of the Advanced Fuel Properties (AFP) project by task as outlined in the original proposal. Each section defines the objectives for the task and describes what was actually completed during Phase I.

4.1 Definition of Fuel Candidates

Objective:

To define the types of hydrocarbon structures to be included in the data base and models of this project.

Work Completed:

The proposal listed aliphatics, olefinics, naphthenics, aromatics, and heteroatomics as the principal categories for fuel candidates. Data for all these compound classes have been assembled in the AFP data base. Originally, the proposal estimated we would build a data base containing 2,500 molecules, but the actual current total has reached 4,462. The hydrocarbon classes include normal alkanes through C₁₀₀, all branched alkane isomers up through C₁₂, cyclopentanes, cyclohexanes, other cycloalkanes, alpha-olefins, other olefins, diolefins, acetylenes, cycloalkenes, decalins, normal and branched alkylbenzenes, tetralins, indans, indenenes, diphenyls, biphenyls, other benzene derivatives such as styrenes, polyaromatics, and multicyclic compounds containing strained and saturated rings. These classes of compounds were chosen because of availability of good data and their presence in many fuel mixtures.

The nonhydrocarbon classes included some of the elements, normal and branched alcohols, aromatic alcohols, polyols, aldehydes and ketones, ethers, epoxides and peroxides, normal and branched carboxylic acids, aromatic carboxylic acids, anhydrides, various kinds of esters, halogenated compounds, amines and imines, nitriles, nitrates, polyfunctional compounds, a few phosphorous compounds, and aromatic rings containing oxygens and nitrogens. Some of these compounds occur in trace amounts in jet fuels derived from petroleum feedstocks but they are more prevalent in fuels derived from coal. However, most of them

were included in the data base because their structures will help definitize the structure based models.

The numbers of entries for each compound category are presented in Table 4.1-1. Nearly all the categories have several compounds with some, like branched alkanes, having hundreds. This large data set will be used to develop new correlative property prediction models based upon graph theory indices and group additivity counts (see section 4.5).

The compound classifications in Table 4.1-1 were made using the FAMLY subroutine (described in section 4.4 under the Structural Subroutines heading) which uses the SMILES strings (see Entry of Structural Data in section 4.3.1) for each compound. This process is straightforward for simple structures, but is open to interpretation when more than one functional group is present in a molecule. The details of how this classification works are described in section 4.4.

4.2 Definition of Properties

Objective:

To define a list of fuel properties to be modeled during the course of this project.

Work Completed:

The list of properties from the RFP (Request for Proposal) is presented in Table 4.2-1. This list has been extended during Phase I to include the single valued (non-temperature and pressure dependent), ideal gas, residual (the difference between real gas or liquid and ideal gas properties), real gas, liquid, liquid-gas transition, solid, and transport properties. These properties are given in Table 4.2-2.

Table 4.1-1
AFP Data Base Family Counts

<u>FAMILY #</u>	<u>NAME</u>	<u># OF COMPOUNDS</u>
1	n-PARAFFINS	101
2	METHYLALKANES	121
3	CYCLOALKANES	43
4	OTHER ALKANES	688
5	ALPHA-OLEFINS	101
6	OTHER ALKANES	164
7	DIOLEFINS	32
8	ALKYNES	89
9	N-ALKYLBENZENES	97
10	OTHER ALKYLBENZENES	85
11	OTHER MONOAROMATICS	40
12	OTHER POLYAROMATICS	568
13	MULTICYCLIC HYDROCARBON RINGS	24
15	ALDEHYDES	20
16	KETONES	65
17	N-ALCOHOLS	18
18	OTHER ALIPHATIC ALCOHOLS	32
19	AROMATIC ALCOHOLS	57
20	POLYOLS	37
21	N-ALIPHATIC ACIDS	19
22	OTHER ALIPHATIC ACIDS	33
23	AROMATIC CARBOXYLIC ACIDS	37
24	ANHYDRIDES	9
25	FORMATES & ACETATES	27
26	N-ALKYL ESTERS	50
27	UNSATURATED ALIPHATIC ESTERS	14
28	AROMATIC ESTERS	46
29	ESTERS	32
30	EPOXIDES & PEROXIDES	24
31	ALIPHATIC CHLORIDES	38
32	AROMATIC CHLORIDES	13
33	C,H,Br COMPOUNDS	14
34	C,H,I COMPOUNDS	6
35	C,H,F COMPOUNDS	19
36	C, MULTIHALOGEN	22
37	ALIPHATIC AMINES	25
38	AROMATIC AMINES	27
39	OTHER AMINES & IMINES	33
40	NITRILES	28
41	C,H,NO2 COMPOUNDS	33

Table 4.1-1 (cont.)
AFP Data Base Family Counts.

<u>FAMILY #</u>	<u>NAME</u>	<u># OF COMPOUNDS</u>
42	MULTIFUNCTIONAL C,H,N,O	85
43	C,H,S COMPOUNDS	310
44	POLYFUNCTIONAL C,H,O	79
45	POLYFUNCTIONAL C,H,O,N	0
46	POLYFUNCTIONAL C,H,O,S,Cl	31
47	POLYFUNCTIONAL C,H,O, HALIDES	26
48	POLYFUNCTIONAL C,H,O,N, HALIDES	11
54	ELEMENTS	38
100	DECALINS	29
101	TETRALINS	41
102	CYCLOOLEFINS	99
104	DIPHENYLS	78
105	BIPHENYLS	22
106	CYCLOPENTANES	115
107	CYCLOHEXANES	153
110	INDANS	181
111	INDENES	71
112	ALKYL RADICALS	18
114	MISCELLANEOUS	68
115	PHOSPHOROUS COMPOUNDS	5
116	NITROGEN AROMATIC RINGS	54
118	OLEFINS WITH>2 DOUBLE BONDS	2
119	OXYGEN AROMATIC RINGS	8
120	CHARGED SPECIES	6

Table 4.2-1
Definition of Fuel Properties from RFP

1. LIQUID DENSITY VERSUS TEMPERATURE
 2. VAPOR DENSITY VERSUS TEMPERATURE AND PRESSURE
 3. LIQUID VISCOSITY VERSUS TEMPERATURE
 4. VAPOR VISCOSITY VERSUS TEMPERATURE
 5. FREEZING POINT
 6. HEAT OF COMBUSTION
 7. VAPOR PRESSURE VERSUS TEMPERATURE
 8. FLASH POINTS
 9. AUTOIGNITION TEMPERATURES
 10. HEATS OF VAPORIZATION
 11. LIQUID HEAT CAPACITY AT CONSTANT PRESSURE
 12. LIQUID HEAT CAPACITY AT CONSTANT VOLUME
 13. GAS HEAT CAPACITY AT CONSTANT PRESSURE
 14. GAS HEAT CAPACITY AT CONSTANT VOLUME
 15. GAS THERMAL CONDUCTIVITY
 16. LIQUID THERMAL CONDUCTIVITY
 17. CRITICAL TEMPERATURE
 18. CRITICAL PRESSURE
 19. CRITICAL VOLUME
 20. BOILING POINT
 21. HEAT OF FUSION
-

Table 4.2-2
Fuel Properties in AFP System

Single Valued Properties:

x	*	1.	Critical Temperature
x	*	2.	Critical Pressure
x	*	3.	Critical Volume
x	*	4.	Critical Compressibility
x	*	5.	Acentric Factor
	*	6.	Characteristic Volumes
	*	7.	Soave-Redlich-Kwong Omega Parameters
x	*	8.	Rackett Parameters
x	*	9.	Normal Boiling Temperature
x	*	10.	Melting Temperature
x		11.	Liquid Molar Volume at 25 C
x		12.	Enthalpy of Formation at 25 C
x		13.	Gibbs Free Energy of Formation at 25 C
x		14.	Absolute Entropy at 25 C
x		15.	Standard Enthalpy of Combustion at 25 C
x		16.	Enthalpy of Fusion at the Melting Temperature
x		17.	Triple Point Temperature
x		18.	Triple Point Pressure
x		19.	Solubility Parameter at 25 C
x		20.	Dipole Moment
x		21.	Radius of Gyration
x		22.	Flash Point
x		23.	Lower Flammability Limit
x		24.	Upper Flammability Limit
x		25.	Autoignition Temperature

Table 4.2-2 (cont.)
Fuel Properties in AFP System

Ideal Gas Properties:

x	*	26.	Enthalpy of Formation at 298K
x	*	27.	Absolute Enthalpy at 298K
x	*	28.	Gibbs Free Energy of Formation at 298K
	*	29.	Enthalpy vs. Temperature
	*	30.	Absolute Entropy vs. Temperature
	*	31.	Gibbs Free Energy vs. Temperature
	*	32.	Helmholtz Free Energy vs. Temperature
	*	33.	Internal Energy vs. Temperature
x	*	34.	Isobaric Heat Capacity vs. Temperature
	*	35.	Isochoric Heat Capacity vs. Temperature
	*	36.	Enthalpy of Formation vs. Temperature
	*	37.	Gibbs Free Energy of Formation vs. Temperature
	*	38.	Formation Equilibrium Constant vs. Temperature

Residual Properties:

*	39.	Enthalpy vs. Temperature and Pressure
*	40.	Entropy vs. Temperature and Pressure
*	41.	Internal Energy vs. Temperature and Pressure
*	42.	Gibbs Free Energy vs. Temperature and Pressure
*	43.	Helmholtz Free Energy vs. Temperature and Pressure
*	44.	Isobaric Heat Capacity vs. Temperature and Pressure
*	45.	Isochoric Heat Capacity vs. Temperature and Pressure
*	46.	Fugacities vs. Temperature and Pressure

Real Gas Properties:

*	47.	Molar Volume vs. Temperature and Pressure
*	48.	Compressibility vs. Temperature and Pressure
x	49.	2nd Virial Coefficient vs. Temperature and Pressure
*	50.	Gas Density vs. Temperature and Pressure
*	51.	Enthalpy vs. Temperature and Pressure
*	52.	Entropy vs. Temperature and Pressure
*	53.	Internal Energy vs. Temperature and Pressure
*	54.	Gibbs Free Energy vs. Temperature and Pressure
*	55.	Helmholtz Free Energy vs. Temperature and Pressure
*	56.	Isobaric Heat Capacity vs. Temperature and Pressure
*	57.	Isochoric Heat Capacity vs. Temperature and Pressure
*	58.	Enthalpy of Formation vs. Temperature and Pressure
*	59.	Gibbs Free Energy of Formation vs. Temperature and Pressure
*	60.	Heat of Combustion vs. Temperature and Pressure

Table 4.2-2 (cont.)
Fuel Properties in AFP System

Liquid Properties:

x	*	61.	Saturated Molar Volumes vs. Temperature
	*	62.	Compressed Molar Volumes vs. Temperature and Pressure
	*	63.	Liquid Densities vs. Temperature and Pressure
	*	64.	Enthalpy vs. Temperature and Pressure
	*	65.	Entropy vs. Temperature and Pressure
	*	66.	Internal Energy vs. Temperature and Pressure
	*	67.	Gibbs Free Energy vs. Temperature and Pressure
	*	68.	Heimholtz Free Energy vs. Temperature and Pressure
x	*	69.	Isobaric Heat Capacity vs. Temperature and Pressure
	*	70.	Isochoric Heat Capacity vs. Temperature and Pressure
	*	71.	Enthalpy of Formation vs. Temperature and Pressure
	*	72.	Gibbs Free Energy vs. Temperature and Pressure
	*	73.	Heat of Combustion vs. Temperature and Pressure
x		74.	Surface Tension vs. Temperature and Pressure

Liquid-Gas Phase Transition Properties:

x	*	75.	Vapor Pressures vs. Temperature
x		76.	Boiling Point Correction
x		77.	Enthalpy of Vaporization vs. Temperature
		78.	Entropy of Vaporization vs. Temperature

Solid Properties:

x		79.	Solid Heat Capacity vs. Temperature
x		80.	Solid Density vs. Temperature

Transport Properties:

x		81.	Liquid Viscosity vs. Temperature and Pressure
x	*	82.	Vapor Viscosity vs. Temperature and Pressure
x		83.	Liquid Thermal Conductivity vs. Temperature and Pressure
x		84.	Vapor Thermal Conductivity vs. Temperature and Pressure

x	Indicates data present in AFP data base
*	Indicates predictive method programmed

The data base contains experimental data for the properties marked with x's. It does not contain experimental data for all the properties because many of them are interrelated. For example, the Gibbs free energies, internal energies, and Helmholtz free energies can all be calculated from the corresponding entropies and enthalpies. Most of the gas phase and residual thermodynamic properties are not stored in the data base because they can be calculated from equations of state. However, all these properties will be modeled because they are important in specialized areas of fuel science. Methods have already been programmed in Phase I for prediction of the properties marked with asterisks. These programs are described in section 4.4. Additional predictive methods will be incorporated into the AFP prediction system during Phase II. It is also likely that additional properties will be added to this list as we model the properties of fuel mixtures.

Even for the properties which are stored in the data base, there are many gaps in the data set due to missing data. The actual numbers of experimental values are shown in Table 4.2-3 for the critical temperature, critical pressure, critical volume, critical compressibility, normal boiling point, melting point,

Table 4.2-3
Examples of Counts for Experimental Properties

T_c	- 1151 VALUES
P_c	- 1152 VALUES
V_c	- 1153 VALUES
Z_c	- 1171 VALUES
T_b	- 2409 VALUES
T_m	- 1893 VALUES
W	- 955 VALUES

and acentric factor. The most data are available for normal boiling points but even for this easily measured property over 2,000 compounds have missing values. These gaps in the literature are a major reason why it is so important to develop accurate methods to estimate the properties of fuels.

4.3 Data Base Development

Objective:

To develop a data base of experimentally measured properties for hydrocarbon fuels.

Development of the data base is divided into the three subtasks described in sections 4.3.1 through 4.3.3.

4.3.1 Literature Review

Objective:

To make a comprehensive and critical review of the scientific literature in order to identify and collect the most accurate experimental data and predictive methods for the properties of pure-component fuels.

Work Completed:

This task involved three parts: selection of data sources for the data base, selection of methods for entering and manipulating molecular structures, and selection of literature methods for the prediction of properties.

Selection of Data Sources:

All the property data were taken from critically evaluated data compilations from reliable sources. These included:

1. The American Institute of Chemical Engineers DIPPR Data base⁸
2. The National Institute for Petroleum and Energy Research Data base on C10 - C16 Molecules⁹
3. Texas A&M's Thermodynamic Research Center's Hydrocarbon Tables¹⁰
4. Texas A&M's Thermodynamic Research Center's Nonhydrocarbon Tables¹¹

5. The JANAF Thermodynamic Tables⁴
6. The National Bureau of Standards Thermodynamic Tables³

Since Allied-Signal is a corporate sponsor of the AIChE DIPPR project, access was available for the most recent data tape from them. Less complete data tapes are also available from the National Bureau of Standards, National Standards Reference Data System in Gaithersburg, MD. The Reference Data Office was also the source of the JANAF and NBS Thermodynamic Tables. The NIPER data base was provided by WRDC/POSF, Wright-Patterson AFB. Allied-Signal subcontracted with Dr. Kenneth Marsh, Director of Texas A&M's Thermodynamic Research Center, for a tape of the TRC Hydrocarbon Tables. The current version of the Advanced Fuel Properties Data Base contains data from DIPPR, NIPER, and the TRC Hydrocarbon Tables. The TRC Nonhydrocarbons, JANAF Tables, and NBS Tables are on the computer but have not yet been loaded into the data base because of time and budgetary constraints during Phase I. They will not be loaded during Phase II until we have completed our data base for fuel mixtures, again due to time and possible budgetary constraints.

Entry of Structural Data:

In addition to property data, the data base must contain the structure for each compound. After reviewing the literature, we selected SMILES (Simplified Molecular Input Line System) strings as the method for entering structural data. The AFP program incorporates the MedChem¹² software package for structural searching. The MedChem software uses SMILES strings as its method for structural input.

SMILES strings are computer readable strings of characters which describe a molecular structure as a 2-D representation where hydrogens are generally omitted. SMILES strings are easy to learn and are constructed using the following six basic rules:

1. Atoms are represented by their atomic symbols and are generally enclosed in square brackets when in the elemental state.

2. Single, double, triple, and aromatic bonds are represented by the symbols '-', '=', '#', and ':', respectively, with single and aromatic bonds being generally omitted.
3. Branches are specified by enclosures in parentheses.
4. Cyclic structures are represented by breaking one bond in each ring and identifying the atom on either side of the break with the same number.
5. Disconnected structures are written as individual structures separated by a '.'.
6. Atoms in an aromatic compound use lower case letters.

One of the drawbacks to SMILES strings is that optical isomers, and cis and trans isomers of double bonds and rings, cannot be distinguished. As the structure-based predictions become more sophisticated, methods of encoding these isomer structures will have to be addressed. At the moment, these isomers are distinguished by an isomer counter in the data base. Daylight Chemical Information Systems (the vendor for the MedChem software) is working on extensions to SMILES strings which will distinguish isomers. These modified SMILES strings will be based upon CONCORD strings and are expected near the end of 1989.

To enter SMILES strings for each of the compounds, lists of compound names for the TRC Hydrocarbon and Nonhydrocarbon, DIPPR, and NIPER data sets were obtained. Over 8,000 SMILES strings were written for these compounds.

In the simplest cases, i.e., methanol, ethanol, etc., SMILES strings were written directly from the name of the compound. As the compounds became more complex, chemical structures were first drawn on paper, or the structures were looked up in the CRC Handbook (Handbook of Chemistry and Physics, 52nd Edition) or other references.

One reliable method for finding obscure structures was to use a computer search. If the CAS (Chemical Abstracts Service) registry number was available for the compound, a computer search output included a line printer version of the

structure. SMILES strings were then written from these structures. This was the primary method of obtaining structures for the more complex inorganic compounds.

SMILES strings for the TRC compounds were entered from lists of compounds divided into families (See section 4.1). Files were generated using the VAX editor listing the SMILES string, formula, and ASID (Allied-Signal identification) number. The SMILES strings were then checked by reading the files into the MedChem software package UDRIVE which drew the structures from the SMILES strings. The structures were compared with the original drawings generated directly from the names. The UDRIVE software was also used to 'Uniquefy' the SMILES strings, i.e., rewrite the SMILES strings using a set of rules so that they would be unique for each compound. This procedure is useful because it speeds up structure based searching of SMILES strings.

DIPPR and NIPER SMILES strings were added to existing data files containing CAS registry number, compound name, and molecular formula. In many cases the molecular formulas were used to determine if the structures were correct.

Selection of Literature Models:

Numerous papers were reviewed during Phase I as part of the search for the best methods to predict fuel properties. Fortunately, the following four books, which include careful reviews of the literature up through 1987, were also found:

1. Reid, Prausnitz, and Poling. The Properties of Gases and Liquids.¹³
2. Edminster and Lee. Applied Hydrocarbon Thermodynamics.¹⁴
3. Danner and Daubert. Manual for Predicting Chemical Process Design Data from the AIChE.⁷
4. Technical Data Book - Petroleum Refining from the American Petroleum Institute.⁶

All of these books provided recommendations for predictive methods for various properties. The first book also contained quantitative comparisons of several of the methods. These books, taken as a collection, provided a very valuable guide to the enormous literature on the prediction of fuel properties and nearly

all of the methods programmed during Phase I were covered in one or more of these reviews.

4.3.2 Data Compilation

Objective:

To compile the fuel property data collected during Task 3.1 into a computer data base that can provide easy management, access, and analyses of the data of either structure or property based parameters.

Work Completed:

The data for all the measured values of all the pertinent properties of 4,462 fuel candidate chemical compounds have been compiled and stored in a data base on the Allied Signal EMRC VAX8600 computer.

The software tool used to manage the storage of these data is called a Data Base Management System (DBMS). The Digital Equipment Corporation (DEC) product VAX Rdb/VMS was the DBMS used for the Advanced Fuels Properties data base. It was chosen because it is a relational DBMS, it is marketed and supported by a reputable vendor, and it is one of the leaders in its field.

Relational Data Base Concepts:

The relational model of data storage offers several advantages over other data models:

1. The structure of the data base is easier to understand.
2. Data can be combined and compared in a wide variety of ways.
3. Relationships among data can be established dynamically.
4. The data base structure can be modified without necessarily rebuilding the entire data base.

Refer to Figure 4.3.2-1 for the following explanation of the concepts of the relational data model.

In a relational data base, data reside in two-dimensional data structures known as relations or tables. One or many relations may exist in a data base. Each relation is made up of rows and columns. The rows are called records and are a collection of fields (columns). Each record must be uniquely identified by one or more fields in the record. This concept is often referred to as the key.

Every record in a relation has the same set of fields in the same order as all the others. The width of the relation is fixed by the list of fields that comprise a record. The length of the table is limited only by the physical constraints of the system and can change at any time by adding to or deleting records from the table.

While each relation in a data base can be viewed as an independent entity, they can also be related to other relations by one or more common fields. When the relations are joined together by these common fields, they form a new larger "logical" relation containing all the information from both relations. For instance, if a relation X contains fields A, B, and C and relation Y contains fields A, D, and E, when they are joined the resulting relation would contain fields A, B, C, D, and E. It is in this simple operation that the real power of the relational data model resides.

Design of the AFP Data Base:

The goal of the AFP data base is to store all the measured values of all the pertinent properties of all the fuel candidates. Each measured value should carry with it an indication of quality, an indication of the source of the value, any references the data source might quote, and any notes or footnotes the measurement might carry.

NOMENCLATURE:
RELATION - RECORD - FIELD - KEY

**KEY - FIELD(S) WHICH
 UNIQUELY IDENTIFY
 A RECORD**

**FIELD OR
 COLUMN**

	A	B	C	DEF
1				
2				
3				
4				
5				

RELATION OR TABLE

Figure 4.3.2-1 Relational Data Base Diagram

Some of the problems this goal presents include:

1. Fuels are chemicals, and it is difficult to uniquely identify a chemical that will be valid, not only for existing chemicals, but also for new chemicals and mixtures.
2. The list of properties has grown, over the life of the project, from 39 to 114. It may yet increase again.
3. The number of measurements for a property may be 0, 1, or an unlimited number.
4. The number of references and notes for a measurement may be 0, 1, or an unlimited number.
5. Some of the data do not have an associated quality indicators.

The problem of the chemical's unique identity was overcome by using the MedChem SMILES string plus a secondary field that is a sequential counter of the number of nonunique occurrences of the SMILES string because of isomers. This solved the uniqueness problem but caused a potential disk storage problem because the SMILES string is currently a 240-byte character string, and the counter is a 4-byte integer. As the unique identifier (SMILES/counter), it would be carried through all relations in the data base that were related to the fuel candidate. Therefore a 4-byte integer field called the ASID (for Allied-Signal Identifier) was created to solve the disk problem. The ASID is a computer assigned number that is the sequential order of the fuel candidate's entry into the data base. It has no chemical meaning, but can be cross referenced to a SMILES string/isomer counter combination and thus a chemical. It saves 240 bytes of storage every time a unique fuel candidate ID is needed within the data base.

While either the SMILES string/isomer counter or the ASID each can uniquely identify a fuel candidate, neither is very practical for retrieving data because neither would be known to a chemist looking for information from the data base. For this reason, the COMPONENTS and SYNONYMS relations are in the data base. The

COMPONENTS relation contains many of the various methods the chemical industry has of identifying chemical compounds. The SYNONYMS relation contains all of the names, both formal and informal, by which a given compound is known.

The list of fields and their description for the COMPOUNDS relation:

ASID	The Allied-Signal Identifier
SMILES1	The first 60 characters of the 240-character SMILES string (Note: this field was partitioned to make it easier to display on a terminal)
SMILES2	The second 60 characters of the SMILES string
SMILES3	The third 60 characters of the SMILES string
SMILES4	The fourth 60 characters of the SMILES string
ISCOUNT	A sequential count of nonunique occurrences of SMILES strings caused by isomers
PSUID	The DIPPR unique identifier
NAMED	The chemical name as found in DIPPR
STRUCTD	The chemical structure as found in DIPPR
FORMULA	The chemical formula
FAMDCODE	The chemical family code as found in DIPPR
FAMKCODE	The chemical family code as assigned by the FAMILY routine
CASNUM	The Chemical Abstracts Services (CAS) chemical identifier
NAMEC	The chemical name as found in CAS
APIID	The American Petroleum Institute identifier for this chemical
NAMEA	The chemical name as found in the API tables
TRCID	The TRC identifier for this chemical
NAMET	The chemical name as found in the TRC tables
NIPERID	The NIPER identifier for this chemical
NAMEN1	The first 60 characters for the chemical name as found in NIPER
NAMEN2	The second 60 characters of the NIPER name
NAMEI	The chemical name according to IUPAC nomenclature rules

These are the fields in the COMPONENTS relation. There is one record in this relation for each fuel candidate in the data base.

The list of the fields in the SYNONYMS relation:

ASID	The Allied Signal Identifier
SYNONYM	A synonym for the chemical identified by this ASID

Only two fields appear in the SYNONYMS relation: one record for each name for each chemical in the data base, although there may be many records for any given ASID. There is usually at least one.

The problems with a loosely determined number of properties to be stored and having an undetermined number of measurements for each property was overcome by storing the measurements as records in a relation as opposed to storing them as fields in a record. Any number of measurements for any number of properties can be stored using this structure.

The relations ALLMSVP (ALL Measurements for Single Value Properties) and ALLMMVP (ALL Measurements for Multiple Value Properties) store all the property measurement values. The difference between the two is that ALLMMVP includes fields for the pressure and temperature at which the values were measured. ALLMSVP contains values for properties that are not dependent upon temperature and pressure.

The list of fields in the ALLMSVP relation:

ASID	The Allied-Signal Identifier
PROPCODE	The property code (see relation TABLE_PROPERTIES)
PROPCOUNT	A sequentially assigned counter for the number of measurements for this property for this ASID
PROPVALUE	The measurement value
DSRCECODE	A code indicating the source of the measurement (see relation TABLE_DATASOURCES)
DQUALCODE	Alphanumeric data quality indicator (carryover from DIPPR)
DQUALNUM	Numeric data quality indicator
DATEIS	Date this measurement was issued
DATEREV	Date this measurement was last revised

There is one record in the ALLMSVP relation for each measurement for each property for each ASID. If there is no measurement for a given property for a certain ASID, then there is no record in this relation with this particular ASID/PROPCODE combination. If there is only one measurement for a given ASID/PROPCODE combination, then PROPCOUNT will be 1. If there are five measurements for a given ASID/PROPCODE combination, then there will be five records each with a different PROPCOUNT and PROPCOUNT going from 1 to 5.

Relation ALLMMVP is identical to relation ALLMSVP except that relation ALLMMVP also contains the fields PROPTMP and PROPPRES, the temperature and pressure at which the measurement was performed.

Relations TABLE_PROPERTIES and TABLE_DATASOURCES are essentially look-up tables and contain the correct translation between the property code and the name of the property and also between the data source code and a text string describing the data source.

The problem of having multiple references, notes, and footnotes for a given measurement was overcome in much the same manner as the synonyms list. Relations ALLMSVP_XTRNLS and ALLMMVP_XTRNLS store the external references and footnotes for the ALLMSVP and ALLMMVP measurements respectively.

The list of fields in the ALLMSVP_XTRNLS relation:

ASID	The Allied Signal Identifier
PROPCODE	The property code
PROPCOUNT	The sequential counter for measurements (see ALLMSVP)
XTRNLCODE	An alphanumeric code to identify the reference/footnote
XTRNLTYPE	Code identifying this as a reference, footnote, or note

There is one record in this relation for each external reference for each measured value for each property for each ASID. The actual text for the reference/footnote/note is stored external to the data base in files associated with the data source. This relation merely contains the pointers to the text location.

With the relations mentioned above, all the data for the AFP project can be stored. There remains the problem of retrieval. When asking for any measurement for a given property, all measurements must be searched. And once a measurement is located and retrieved, it may not be a representative measurement, that is, the accuracy of any arbitrarily retrieved measurement is not known. To overcome these problems, the BESTMSVP and the BESTMMVP relations were added to the data base. The BESTMSVP relation contains the best measurements for each single value property for each ASID. The BESTMMVP relation contains the regression coefficients and a regression equation code for each of the multiple value properties for each ASID. These equations and coefficients are currently not stored in ALLMMVP.

The fields in the BESTMSVP relation:

ASID	The Allied Signal Identifier
BMV001	The best measured value for property code 1
BQN001	The numeric quality indicator for property code 1
BIX001	The cross reference back to the ALLMSVP relation for property code 1 - contains the value of PROPCOUNT
BMV002	The best measured value for property code 2
BQN002	The numeric quality indicator for property code 2
BIX002	The cross reference back to the ALLMSVP relation for property code 2 - contains the value of PROPCOUNT
etc.	

There is a field for the value, the quality, and the propcount for each of the single value properties. The single value properties currently have property codes 1-26, 42-68, 70, and 71.

There is a record in BESTMSVP for each fuel candidate in the data base. If there is no measurement for a given property in the ALLMSVP relation, then both the property value AND the cross reference back to the ALLMSVP relation will be zero. If the cross-reference field is nonzero, then the property value is actual. The quality indicator is the decimal fractional representation of the quality. For example, if a value is accurate to ± 5 percent, then the quality

indicator will be 0.05. Currently the AFP system only handles equal plus and minus errors.

The criteria used to load the BESTMSVP relation from the ALLMSVP relation are as follows:

1. Choose the measurement with the smallest nonzero quality indicator.
2. If there is more than one value with the same quality indicator, then choose by data source. The priority scheme is DIPPR, NIPER, and lastly TRC. This order was selected because the DIPPR data were selected by a committee of the American Institute of Chemical Engineers and contained error bars and references telling where the numbers came from. The NIPER data were collected in the last five years and also contained error information. The information in the TRC tables rarely included error bars or detailed references to where the values came from. However, the TRC tables have long been the standard reference source for thermodynamic data for the chemical and petroleum industries, have been updated regularly, and are generally considered to be reliable. In practice, the situation where the quality codes were the same for more than one value rarely occurred in building the database. This rule was, therefore, used only in a few dozen cases.
3. If more than one value has the same quality indicator and the same data source code, then keep the first one encountered. This situation only occurred in the DIPPR data where several experimental values were sometimes reported for the same property. By convention, the DIPPR committee stored the recommended value first in their data file and this rule picks it out. This rule was applied in very few cases.

The fields in the BESTMMVP relation:

ASID	The Allied Signal Identifier
PROPCODE	The property code
REQNCODE	The regression equation code
REQCOEFA	Coefficient A for the regression equation
REQCOEFB	Coefficient B for the regression equation
REQCOEFC	Coefficient C for the regression equation
REQCOEFD	Coefficient D for the regression equation
REQCOEFE	Coefficient E for the regression equation
REQCOEFF	Coefficient F for the regression equation
REQCOEFG	Coefficient G for the regression equation
REQCOEFH	Coefficient H for the regression equation
REQCOEFI	Coefficient I for the regression equation
REQCOEFJ	Coefficient J for the regression equation
REQTEMPU	The upper limit for valid temperature range
REQTEMPL	The lower limit for valid temperature range
REQPRESU	The upper limit for valid pressure range
REQPRESL	The lower range for valid pressure range
REQQCODE	The quality code
REQNUMCS	The number of coefficients actually used

There is one record in the BESTMMVP relation for each regression equation for each property for each ASID in the data base. If no regression equation has been fitted to the ALLMMVP data for a given ASID/property, then no record will exist in BESTMMVP for that ASID/property. If more than one regression equation has been fitted to the ALLMMVP data for a given ASID/property, then more than one record will exist in BESTMMVP for that ASID/property. There are no instances of multiple equations for a property as yet. Should that instance arise, the data base will be capable of handling it.

The quality code is alphanumeric in BESTMMVP. It uses the DIPPR interpretation for quality codes. These alphanumeric codes will be converted to numeric codes in the near future.

The final piece of the AFP data base is relation BTRMMVP. This relation contains temperature and pressure dependent data from TRC that had no regression

equation fitted to it. The description of each record is identical to the ALLMMVP record description. The data in BTRMMVP are not also contained in ALLMMVP.

4.3.3 Obtaining Missing Data

Objective:

To experimentally measure the properties of pure hydrocarbons which were not included in the literature data base but are judged to be important in determining structure-property relationships.

Work Completed:

Because of the enormous size of the literature data base, we did not feel that any critical data points were missing. To demonstrate the extent of the data base, thirty hydrocarbons were arbitrarily selected and the values for eighteen properties were requested for each compound. The compounds selected were:

Ethane	Toluene
Propylene	1,3-Dimethylbenzene
Butane	Ethylbenzene
Octane	m-Ethyltoluene
2-Methylpentane	Naphthalene
Neopentane	1-Ethyl-naphthalene
Cyclohexane	1,3-Dimethylnaphthalene
Methylcyclohexane	1-Ethyl-3-methylnaphthalene
Trans-1,3-dimethylcyclohexane	Trans-decahydronaphthalene
Ethylcyclohexane	1-Ethyl-cis-decahydronaphthalene
Trans-1-ethyl-3-methylcyclohexane	1,3-Dimethyldecahydronaphthalene
1-Ethyl-3methyldecahydronaphthalene	Vinylcyclohexane
2,2-Dimethylbutane	Cyclopentane
2,2-Dimethylpentane	Methylcyclopentane
2,2-Dimethylhexane	Benzene

The properties that were selected for testing and the number of compounds, out of the thirty listed above, for which values were retrieved are:

Triple Point Temperature	22
Triple Point Pressure	22
Liquid Molar Volume at 298K	22
Melting Point at Standard Pressure	23
Flash Point	19
Upper Flammability Limit	22
Lower Flammability Limit	22
Entropy at 298K for an Ideal Gas	22
Enthalpy of Formation at 298K for an Ideal Gas	22
Enthalpy of Formation at 298K for a Liquid	22
Enthalpy of Combustion at 298K	22
Critical Volume	22
Critical Temperature	22
Critical Pressure	22
Critical Compressibility	22
Normal Boiling Point	26
Autoignition Temperature	22
Acentric Factor	22

Three of the compounds (vinylcyclohexane, 1,3-dimethyldecahydronaphthalene, and 1-Ethyl-3-methyldecahydronaphthalene) were not in the data base. One compound (1-ethylnaphthalene) was found in the data base but did not have values for any of the test properties, four compounds (trans-1-ethyl-3-methylcyclohexane, 1-ethyl-3-methylnaphthalene, and 1-ethyl-cis-decahydronaphthalene) have only the normal boiling point, and one compound (1,3-dimethylnaphthalene) had only the normal boiling point and the melting point at standard pressure. This example demonstrates the extent of data that is available in the data base. Therefore, no work was done on this task during Phase I. As part of Phase II, we do anticipate collecting some experimental data because the literature on mixtures is substantially smaller than for pure component fuels.

4.4 Compilation, Evaluation, and Selection of Structure-Property Relationships

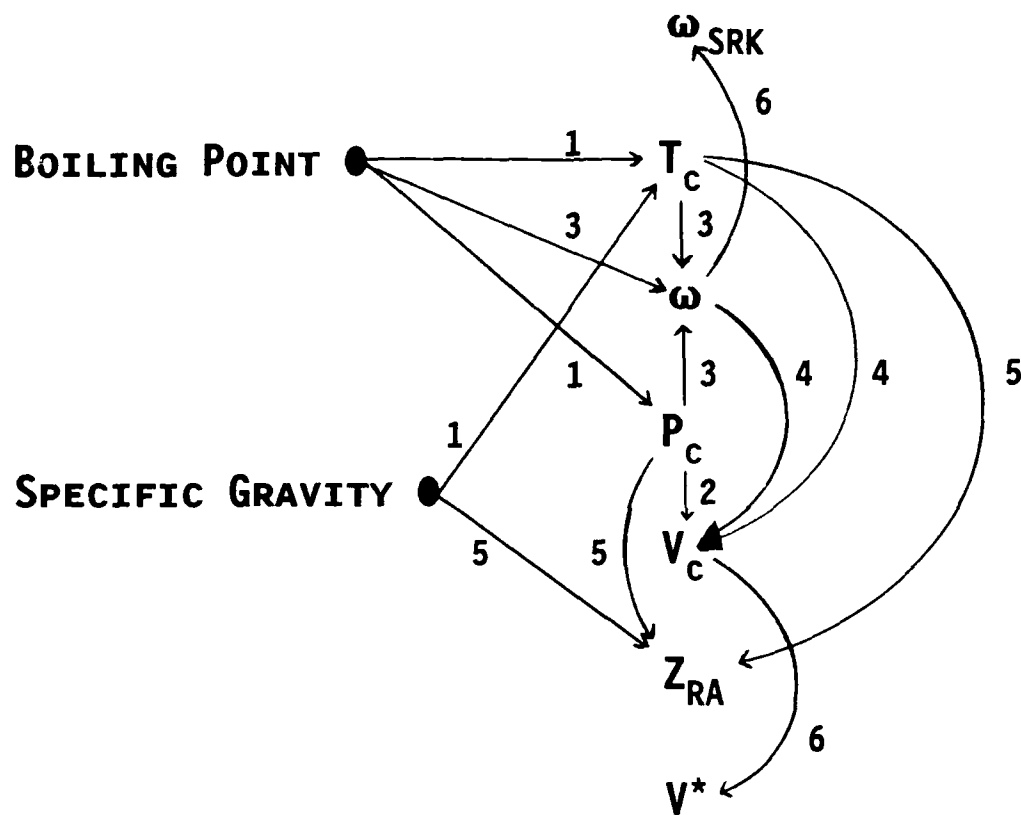
Objective:

To collect and assess known structure-property relationships for pure hydrocarbons in order to develop accurate structure based predictive methods for the properties listed in section 4.2.

Work Completed:

The methods, recommended by the American Petroleum Institute (API) and American Institute of Chemical Engineers (AIChE), for predicting the properties of small fuel molecules were carefully evaluated. It was found that all of these predictive methods were hierarchical and depended on only two experimental inputs: the normal boiling point and specific gravity at room temperature^{6,7}. Using these two experimental inputs, the critical temperature and pressure could be calculated followed by the acentric factor, critical volume, and various specialized parameters appearing in equations of state (Figure 4.4-1). Densities and thermodynamic properties were then calculated from the equations of state at any temperature and pressure (Figure 4.4-2).

Based upon this analysis, our strategy in developing structure based predictive methods has been to focus on the key single valued properties such as the normal boiling point, critical properties, and acentric factor and then program in established equations of state for the temperature and pressure dependence of properties. We have automated the user structural inputs required by many of the API and AIChE methods using the MedChem software and SMILES strings. We have also programmed several methods for many of the properties so that we could compare the accuracies of the various prediction schemes.



NUMBERS INDICATE THE LOGICAL SEQUENCE FOR PREDICTING PROPERTIES.

Figure 4.4-1 API Prediction of Critical Properties and Tabulated Parameters

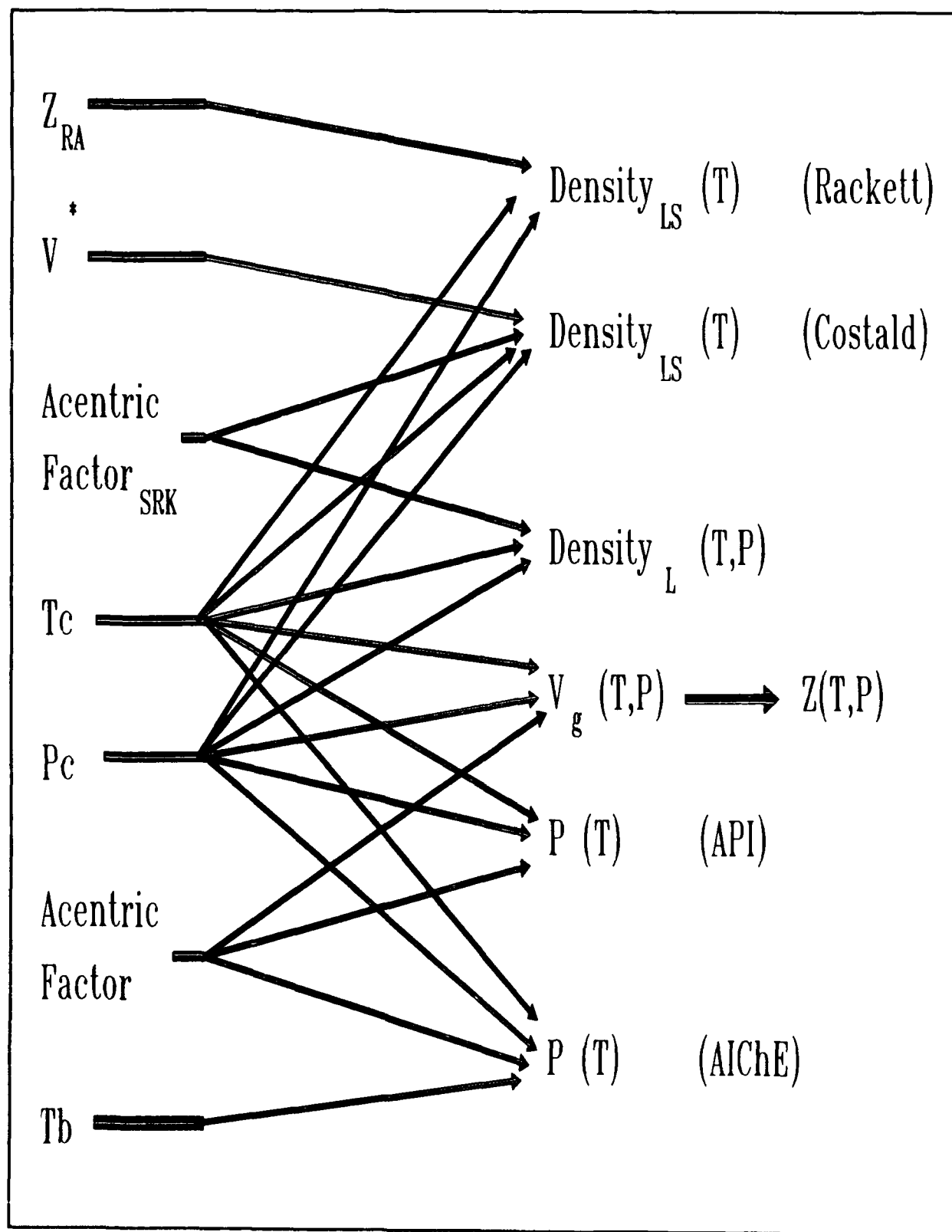


Figure 4.4-2 Prediction of densities and vapor pressures

The methods developed under this task are presented in the following subsections:

1. Data base Access Routines
2. Methods for Structural Inputs
3. Methods for Single Valued Properties
4. Introduction to the Methods for Thermodynamic Properties
5. Methods for Ideal Gases
6. Methods for Residual Properties
7. Methods for Real Gases
8. Methods for Liquids
9. Methods for Phase Transitions
10. Methods for Transport Properties
11. Methods for Solids
12. Methods for Mixtures
13. Methods for Error Tracking

Each method described in this section has been programmed as a separate subroutine which can be called independently. A discussion of how they compare with experiment is presented in section 4.7. The operation of the main program and user interface is discussed in section 5.

Data Base Access Routines

Single Valued Property Access Routines

All the single valued property access routines are identical in function. They access the BESTMSVP relation, count the number of records in the relation with the desired ASID, and, if there is a record for the ASID, retrieve the property value, quality indicator, and reference back to the ALLMSVP relation for the desired ASID.

Inputs to the routines are:

ASID	An integer array of the Allied-Signal identifiers
SMILES	A character array of SMILES strings
NCMPDS	An integer field indicating the number of ASID's in the ASID array

Outputs from the routines are:

VALUE	A real array of property values that have been retrieved
ERROR	A real array of the quality indicators for the property values
IER	An integer array of error codes

All the routines use the RDB\$INTERPRET function to send commands to Rdb/VMS and retrieve data from the data base. They all function as follows:

There is a DO loop that loops through the ASID array from element 1 to the NCMPDS element. Inside this loop:

RDB\$INTERPRET is used to count the records in relation BESTMSVP having the current ASID.

Error signals are put into the current element of IER if there is an Rdb error or if no records are found for this ASID.

The property value, the quality indicator, and the cross reference back to the ALLMSVP relation are retrieved using RDB\$INTERPRET.

Error signals are put into the current element of IER if there is an Rdb error or if both the property value and the cross reference value are zero. The latter indicates no value for this property.

The retrieved values are loaded into the output arrays.

Once the loop has finished, the routine is complete.

If IER is nonzero, then an error has occurred. Currently, the only possible error code suffixes are:

- 001 Indicates an Rdb error
- 501 Indicates no data for this property or ASID

Multiple Valued Property Access Routines

All the multiple valued property access routines are identical in function. They access the BESTMMVP relation, count the number of records in the relation with the desired ASID, and, if there is a record for the ASID, retrieve the property value regression equation and coefficients, quality indicator, and valid temperature and pressure ranges.

Inputs to the routines are:

- ASID An integer array of Allied-Signal identifiers
- SMILES A character array of SMILES strings
- NCMPDS An integer field indicating the number of ASID's in the ASID array

Outputs from the routines are:

- VALUE A two dimensional real array containing the equation code, the number of coefficients, the 10 coefficients, and the temperature and pressure limits for each ASID
- ERROR A real array of the quality indicators for the equation.
- IER An integer array of error codes

All the routines use the RDB\$INTERPRET function to send commands to Rdb/VMS and retrieve data from the data base. They all function as follows:

There is a DO loop that loops through the ASID array from element 1 to the NCMPDS element. Inside this loop:

RDB\$INTERPRET is used to count the records in relation BESTMMVP having the current ASID.

Error signals are put into the current element of IER if there is an Rdb error or if no records are found for this ASID.

The equation code, the number of coefficients, the 10 coefficients, the temperature and pressure limits, and the quality indicator are retrieved using RDB\$INTERPRET.

Error signals are put into the current element of IER if there is an Rdb error.

The retrieved values are loaded into the output arrays.

Once the loop has finished, the routine is complete.

If IER is nonzero, then an error has occurred. Currently, the only possible error code suffixes are:

001 Indicates an Rdb error

501 Indicates no data for this property or ASID

Methods for Structural Inputs

The structural methods in the AFP Property Prediction System are used to supply structural information to subroutines requiring group decompositions, atom counts, the Z number, and molecular formulas. They were also used to classify molecules into families (see Table 4.4-1) and to check the SMILES strings entered under Task 3.2. The structural methods are based upon MedChem software.

MedChem Software:

MedChem software is a system for the storage and retrieval of chemical information and structure. It is a product of Daylight Chemical Information Systems, Inc. Its capabilities include:

- * - Computer-readable chemical structure representation as a SMILES string.

- Graphical representation of SMILES strings.
- * - Substructure searching of SMILES strings using SMARTS strings
- THOR ("Thesaurus Oriented Retrieval") data base system provides MedChem's POMONA89, a 21,565-compound data base and the capability for the user to include additional chemical structures and information
- MERLIN routine for substructure searching of the compounds in a THOR data base

The capabilities marked with an asterisk are were determined to be useful and/or cost effective and therefore are the only capabilities used by the Advanced Fuel Properties system.

Substructure Searching:

The Advanced Fuel Properties system uses SMILES and SMARTS strings (fragments of SMILES strings representing pieces of molecules) to do substructure searching for chemical family classification and for property estimation, e.g., to search for Benson's groups in the estimation of the ideal gas heat of formation of a compound.

Substructure Searching Using GCL Files - Whenever possible, MedChem's GENIE Control Language, GCL, was used to do substructure searching. GCL is a command language that allows one to write a substructure search routine using SMARTS strings and execute the search on any SMILES string.

In the Advanced Fuel Properties system software, a GCL search is executed by calling the subroutine COUNT and passing the name of the GCL file to be executed. When a substructure search on a SMILES string is successful, the subroutine INCGRP is called to set the necessary variable. GCL file substructure

searching is used in the Benson's thermodynamic property estimation routine and other group decomposition routines.

Substructure Searching Using SMARTS Searching Directly - When GCL files could not be used (for example, when the type of search to be done required more decision-making or faster execution, a FORTRAN routine was preferred) direct substructure searching using SMARTS strings was done. This was accomplished by sending a SMARTS string along in a call to the subroutines FIND, SRCH, or COUNT. Each has different schemes for marking atoms as found in a SMILES string when matched by a SMARTS substructure. Subroutine FIND is used in the chemical family classification routines, subroutine SRCH is used in the atom-by-atom testing routines described below, and subroutine COUNT is used for multi-atom searching in the group decomposition routines.

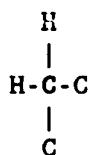
Group Decompositions:

One way of predicting properties from chemical structures is to break the structure into parts and sum the contribution of each of the parts to the property value. The Advanced Fuel Properties software utilizes two methods of group decomposition for property prediction: atomic groups and multi-atom groups.

Atomic Group Decompositions - In atomic group decompositions, the contribution to the property value is obtained by summing the contribution of one atom at a time. The contribution of each atom may or may not contain information about the hybridization or neighbors of that atom.

Example:

A Csp^3 carbon is one example of an atomic group that includes hybridization. An example of an atomic group definition that includes neighbors is



where the bolded carbon is the only atom counted for this group, the other atoms are only used to define the group.

Multi-Atom Group Decompositions - In multi-atom group decompositions, a group contains more than one atom and a given group may be contained within another group for which there is also a contribution. Therefore, a hierarchical search for groups, and a marking of atoms once a group has been found, is necessary in multi-atom group decompositions.

Example:

The search for the propyl group, $-\text{CH}_2\text{CH}_2\text{CH}_3$, must precede a search for a methyl, $-\text{CH}_3$, or an ethyl, $-\text{CH}_2\text{CH}_3$ group.

Benson's Group Additivity:

The Advanced Fuel Properties software uses a number of tables of group contributions for properties. One of the major tables is the one developed by S. W. Benson, published in his book Thermochemical Kinetics¹⁵. Benson's tables uses both atomic and multi-atom group decompositions in estimating the ideal gas entropy, enthalpy, and the heat capacity of a molecule.

Example: Calculation of the heat of formation of methylcyclohexane
using Benson's group additivity method

Group	Benson's notation	Contribution of Group to Heat of Formation	
-----	-----	-----	
Methyl	C-(H)3(C)	-10.20 kcal/mole	(atomic group)
Methylenes in ring	C-(H)2(C)2	5 x -4.93 kcal/mole	(atomic group)
Substituted ring carbon	C-(H)(C)3	-1.90 kcal/mole	(atomic group)
Ring correction	C1CCCCC1	0.00 kcal/mole	(multi-atom group)

		-36.75 kcal/mole (measured value	
		-36.99 kcal/mole)	

Atom-by-Atom Counting:

Once a SMILES string is initialized with the MedChem software, there is a great deal of information about the molecule in the MedChem arrays. Some of this information was used to determine certain properties of the molecule. The molecular weight (subroutine MW2), the Z number (ZNUMB), the number of carbons (CNUM), and the molecular formula (MOLCFM), for example, were determined by accessing the atomic number of each character in the SMILES string and the hydrogen count of the molecule.

Chemical Family Classification:

The Advanced Fuel Properties software uses a chemical family classification scheme to aid in property estimation and method development. The scheme, embodied in the subroutine FAMILY2 and used for Table 4.1-1 was based originally upon the classifications of chemical compounds used in the DIPPR and TRC data bases. New chemical families were created when it was found that the number of

compounds in a given family began to get too large and there was a chemically significant manner in which to subdivide the family.

The chemical family classification scheme is illustrated in Figure 4.4-3. A molecule is classified into a family by searching for a substructure within the molecule that characterizes the family. If the molecule contains the substructure, the search is completed. If not, another substructure search is done. This process continues until a family is found in which the molecule belongs.

The scheme is hierarchical. Therefore, a compound which contains two different functional groups may be classified into a family which only recognizes one of them as significant. Figure 4.4-4 illustrates the subdivision of the "Various hydrocarbon families" indicated in Figure 4.4-3. Again, this hydrocarbon family classification scheme was generally based upon the DIPPR and TRC chemical family schemes and will not classify a compound with more than one functional group in more than one family.

Methods for Single Valued Properties

The methods for the single valued properties are summarized in Table 4.4-1 (The convention for naming the method subroutines is described in section 5.1). For each property, the subroutines available for that property are listed along with a brief explanation of the method and a literature reference if it's appropriate. Many of these methods use group additivity along with experimental inputs to make their predictions. Some are simple correlations between one property and another such as method ZRA2 which calculates the Rackett parameter from the acentric factor.

Following each method is its priority for the priority system described in section 5. Data base lookup methods have the highest priority because they return experimental values. The rest of the methods are prioritized according to recommendations in the reviews^{6,7,13,14} listed in section 4.3.1 and the results of our own testing described in section 4.7. Methods which are not followed by

Molecules are classified into chemical families using substructure searching. A molecule will be a member of only one family. The hierarchical scheme for classifying molecules is as follows:

<u>MOLECULE CONTAINS:</u>	<u>FAMILY CLASSIFICATION</u>
1 atom or 2 identical atoms	Element family
Atoms other than H, C, N, O, S, P, or halogens	Miscellaneous
C and H only	Various hydrocarbon families
Phosphorous	Phosphorous family
Sulfur	Sulfur family
Halogen	Various halogen families
Nitrogen	Various nitrogen families
Oxygen	Various oxygen families

NOTE: Because of the fact that the scheme above is hierarchical and that each molecule belongs to only one chemical family, molecules have certain functional groups in common that may be placed in different chemical families.

EXAMPLE:	CCCN=O	----->	a nitrogen family
	CCCN=O	----->	a halogen family
	Cl		

Figure 4.4-3 Chemical Family Classification Scheme

<u>MOLECULE HAS:</u>	<u>FAMILY CLASSIFICATION</u>
TRIPLE BONDS	----> ALKYNES
ALIPHATIC RINGS	
DECALIN STRUCTURE	----> DECALIN FAMILY
Bi- OR Tri-CYCLIC RINGS	----> MULTICYCLIC
HYDROCARBON RINGS	
DOUBLE BONDS	----> CYCLOOLEFINS
CYCLOPENTANE STRUCTURE	----> CYCLOPENTANE FAMILY
CYCLOHEXANE STRUCTURE	----> CYCLOHEXANE FAMILY
OTHER	----> CYCLOALKANES FAMILY
DOUBLE BONDS	
1 DOUBLE BOND AND MORE THAN 1 METHYL	----> OTHER ALKENES FAMILY
1 DOUBLE BOND AND 1 METHYL GROUP	----> ALPHA-OLEFINS FAMILY
2 DOUBLE BONDS	----> DIOLEFINS
MORE THAN 2 DOUBLE BONDS	----> OLEFINS WITH > 2
DOUBLE	BONDS
METHANE	----> n-PARAFFINS
2 METHYL GROUPS	----> n-PARAFFINS
BRANCHING IN MOLECULE	
1 METHYL BRANCH	----> METHYLALKANES
MORE THAN 1 BRANCH	----> OTHER ALKANES
MORE THAN 6 AROMATIC CARBONS	
MORE THAN 2 FUSED RINGS	
ANTHRACENE STRUCTURE	----> ANTHRACENE FAMILY
PHENANTHRENE STRUCTURE	----> PHENANTHRENE FAMILY
OTHER	----> OTHER POLYAROMATICS
2 FUSED RINGS	
NAPHTHALENE STRUCTURE	----> NAPHTHALENE FAMILY
BIPHENYL RINGS	
1 BIPHENYL RING	----> BIPHENYL FAMILY
MORE THAN ONE BIPHENYL RING	----> OTHER POLYAROMATICS
PENDANT PHENYL RINGS	
2 PHENYL RINGS	----> DIPHENYL FAMILY
MORE THAN 2	----> OTHER POLYAROMATICS
6 AROMATIC CARBONS	
TETRALIN STRUCTURE	----> TETRALIN FAMILY
INDAN STRUCTURE	----> INDAN FAMILY
INDENE STRUCTURE	----> INDENE FAMILY
DOUBLE OR TRIPLE BONDS OR	----> OTHER MONOAROMATICS
ALIPHATIC RINGS	
BENZENE	----> n-ALKYL BENZENE
ONLY 1 METHYL GROUP	----> n-ALKYL BENZENE
MORE THAN 1 METHYL GROUP	----> ALKYL BENZENES

Figure 4.4-4 Hydrocarbon Family (C and H only) Classification

Table 4.4-1
Sources of the Methods for Single Valued Properties.

Critical Temperature:

- TC1 - data base lookup (priority 1)
- TC2 - Joback's method ¹⁶ (priority 2)
- TC3 - MW method ¹⁷ (priority 6)
- TC4 - Jalowka and Daubert's method ¹⁸ (priority 5)
- TC5 - Fedor's method ¹⁹ (priority 4)
- TC6 - AIChE 2C and API 4A1.1 ^{20, 21} (priority 3)

Critical Pressure:

- PC1 - data base lookup (priority 1)
- PC2 - Joback's method ¹⁶ (priority 2)
- PC3 - MW method ¹⁷ (priority 5)
- PC4 - Jalowka and Daubert's method ¹⁸ (priority 4)
- PC5 - AIChE 2F and API 4A1.1 ^{22, 21} (priority 3)

Critical Volume:

- VC1 - data base lookup (priority 1)
- VC2 - Joback's method ¹⁶ (priority 2)
- VC3 - MW method ¹⁷ (priority 4)
- VC4 - API 4A1.1 ²¹ (priority 3)

Critical Compressibility:

- ZC1 - data base lookup (priority 1)
- ZC2 - calculated from PC, VC, and TC (priority 2)
- ZC3 - from acentric factor ²³ (priority 3)

Acentric Factor:

- ACENF1 - data base lookup (priority 1)
- ACENF2 - Lee-Kesler ²⁴ (priority 2)
- ACENF3 - from PVAPS ²⁵ (priority 3)
- ACENF4 - Antoine Eq'n ²⁶ (priority 4)

Characteristic Volumes:

- VSTAR2 - substituted with VC ²⁷ (priority 2)
- VSTAR3 - correlation with omega-SRK ²⁸
- VSTAR4 - HBT method with liquid density at 25C ²⁹ (priority 1)

Soave-Redlich-Kwong Parameter:

- ACSRK2 - substituted with acentric factor ³⁰ (priority 1)
-

Table 4.4-1 (cont.)
Sources of the Methods for Single Valued Properties.

Rackett Parameter:

- ZRA1 - data base lookup (priority 1)
- ZRA2 - from acentric factor ³¹ (priority 3)
- ZRA3 - substituted with ZC ³² (priority 4)
- ZRA4 - calculated from liquid density at 25 C ³³ (priority 2)

Normal Boiling Point:

- TNBP1 - data base lookup (priority 1)
- TNBP2 - Joback's method ¹⁶ (priority 2)

Melting Temperature:

- TMPSP1 - data base lookup (priority 1)
- TMPSP2 - Joback's method ¹⁶ (priority 2)

Liquid Molar Volume at 25 C:

- LMV251 - data base lookup from DIPPR (priority 1)
- LMV252 - data base lookup from TRC (priority 2)

Enthalpy of Formation at 25 C:

- HF251 - data base lookup (priority 1)

Gibbs Free Energy of Formation at 25 C:

- GF251 - data base lookup (priority 1)

Absolute Entropy at 25 C:

- S251 - data base lookup (priority 1)

Standard Enthalpy of Combustion at 25 C:

- HC251 - data base lookup (priority 1)
-

Table 4.4-1 (cont.)
Sources of the Methods for Single Valued Properties.

Enthalpy of Fusion at T_m :

HFTMP1 - data base lookup (priority 1)

Triple Point Temperature:

TTP1 - data base lookup (priority 1)

Triple Point Pressure:

PTP1 - data base lookup (priority 1)

Solubility Parameter:

SP251 - data base lookup (priority 1)

Dipole Moment:

DM1 - data base lookup (priority 1)

Radius of Gyration:

RG1 - data base lookup (priority 1)

Flash Point:

FP1 - data base lookup (priority 1)

Upper Flammability Limit:

FLLW1 - data base lookup (priority 1)

Lower Flammability Limit:

FLUP1 - data base lookup (priority 1)

Autoignition Temperature:

TAI1 - data base lookup (priority 1)

a priority were still being debugged when this report was written during June 1989.

The calling sequences for the single value property routines are identical to those for single valued property data lookups. Inputs are:

ASID	An integer array of Allied-Signal identifiers
SMILES	A character array of SMILES strings
NCMPDS	An integer field indicating the number of ASID's in the ASID array

Outputs are:

VALUE	A real array of property values
ERROR	A real array of the quality indicators for the property values
IER	An integer array of error codes

Introduction to the Methods for Thermodynamic Properties

The data flow for calculations of thermodynamic properties of fluids is illustrated in Figure 4.4-5. Two groups of inputs are needed: (1) critical temperatures, critical pressures, and acentric factors are required for calculations of nonideal gas pressure effects using equations of state, and (2) ideal gas enthalpies of formation at 298K, ideal gas absolute entropies at 298K, and ideal gas heat capacities as a function of temperature are required to calculate ideal gas properties.

Using an equation of state, the gas and liquid molar volumes, densities, and compressibilities can be calculated from the first set of inputs. The molar volumes can then be used to calculate residual thermodynamic properties for either the gas or liquid phase. Directly from these residual thermodynamic properties, the properties associated with the liquid-gas phase transition such as boiling points, vapor pressures, and heats of vaporization can be calculated.

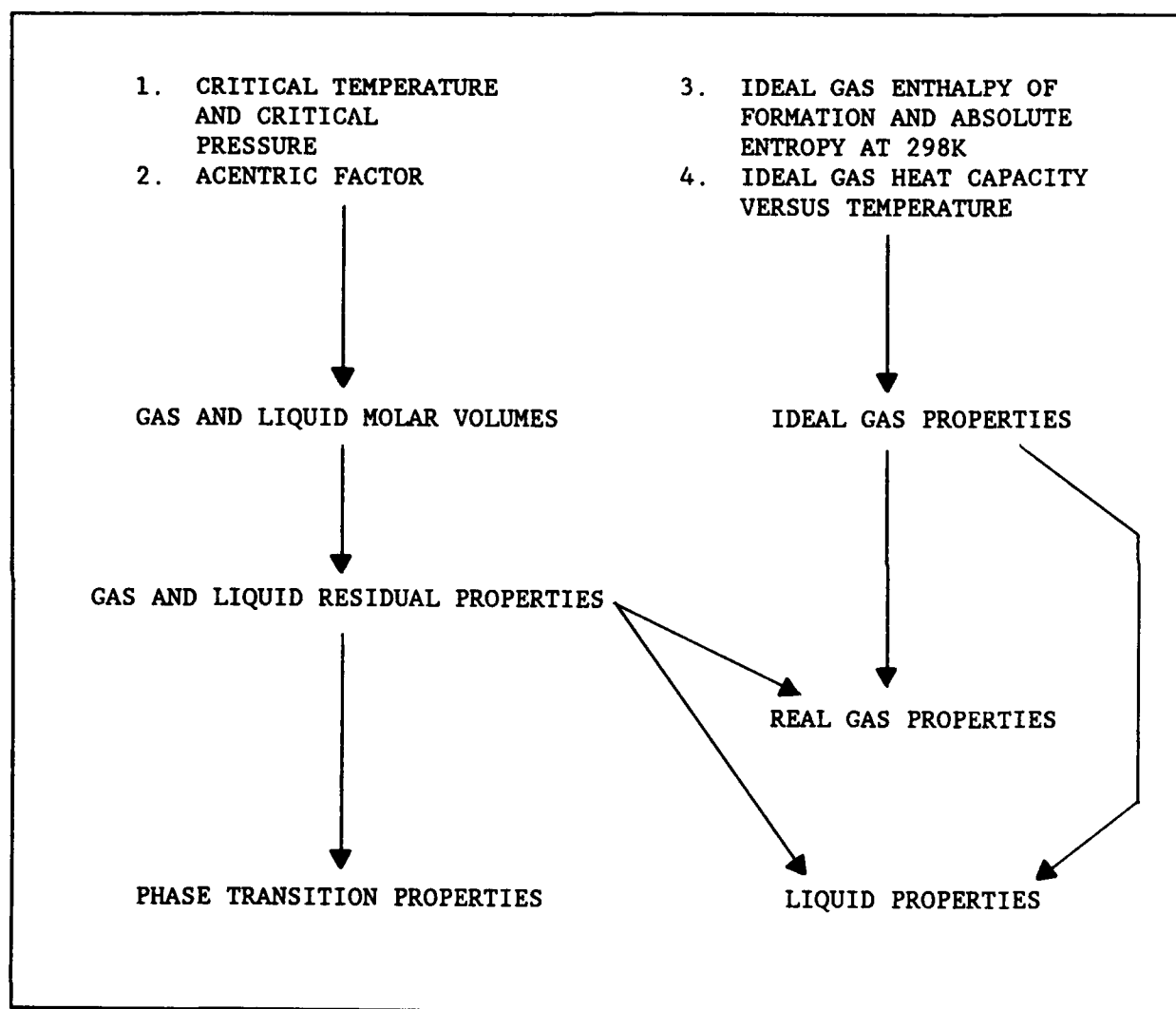


Figure 4.4-5 Data Flow for Fluid Thermodynamic Calculations

Starting with the second set of inputs, the ideal gas thermodynamic properties can be calculated by simply integrating the heat capacity for relative enthalpies and the heat capacity divided by temperature for relative entropies. The values of the enthalpy of formation and absolute entropy at 298K are used to calculate ideal gas enthalpies of formation and absolute entropies at any temperature.

By combining the ideal gas thermodynamic properties with the gas phase residual properties, the real gas properties can be calculated at any temperature and pressure. Similarly, the combination of ideal gas properties and the liquid phase residual properties gives liquid properties at any temperature and pressure.

Methods for Ideal Gases

The calculation of ideal gas thermodynamic properties is complicated by the great variety of ways in which temperature dependent heat capacities are stored in the literature. In the DIPPR data base alone two equations are used to describe ideal gas heat capacities:

$$C_p^{\text{ideal}} = A + B*T + C*T^2 + D*T^3 + E*T^4$$

$$C_p^{\text{ideal}} = A + B*(C/(\sinh(C/T)))^2 + D*(E/(T*\cosh(E/T)))^2$$

Benson's group additivity method for predicting ideal gas heat capacities produces values at temperatures of 300K, 400K, 500K, 600K, 800K, 1000K, and 1500K. Heat capacities at other temperatures are estimated by interpolating among these values. The ideal gas heat capacity data in the TRC tables are also tabulated at individual temperatures, but they are different than those from Benson's method. Thus, a set of ideal gas subroutines is required for every source of data or predicted values.

The relative enthalpy and absolute entropy of an ideal gas are calculated from the heat capacity using the following equations:

$$H^{\text{ideal}}(T) - H^{\text{ideal}}(298) = \int C_p^{\text{ideal}} dT$$

$$S^{\text{ideal}}(T) = S^{\text{ideal}}(298) + \int C_p^{\text{ideal}}/T dT$$

The relative enthalpy is not very useful for thermodynamic calculations, therefore the ideal gas enthalpy of formation is also calculated as described in Figure 4.4-6. This quantity is the enthalpy of reaction for the formation of a compound from its elements at standard conditions, i.e., the specified

$$\Delta H_f^0(T) = (H^0(T) - H^0(298))_{\text{compd}} - \sum \nu_i (H^0(T) - H^0(298))_{\text{elem}} + \Delta H_f^0(298)$$

where,

- * $\Delta H_f^0(298)$ is calculated using method HF25I (Table 4.4-1)
- * $(H^0(T) - H^0(298))_{\text{compd}}$ is calculated using method HID (page 50)
- * $(H^0(T) - H^0(298))_{\text{elem}}$ is calculated for all of the elements (based initially on Table 18 from TRC¹⁰)
- * ν_i are stoichiometric coefficients or atom counts

(Method ATMCNT does this automatically)

Figure 4.4-6 Ideal Gas Enthalpy of Formation

temperature and one atmosphere. The superscript circle is used to indicate that these are ideal gas standard state enthalpies. Real gas and liquid thermodynamic properties are used for the elements and are looked up from a data table. The stoichiometric coefficients, ν_i , are the counts for each atom in the compound and are automatically calculated in the AFP prediction system using the structural method ATMCNT.

The ideal gas predictive methods are summarized in Table 4.4-2. For most properties there are only two methods: one based upon the two DIPPR equations and one based upon Benson's method. A modification of the Benson's methods for using tabular heat capacity data from the TRC tables will be completed. The highest priority is given to the lookup of experimental data followed by Benson's group additivity method.

Many of these ideal gas thermodynamic properties are both temperature and pressure dependent since ideal gas entropies, Gibbs free energies, and Helmholtz free energies change with pressure. These effects are frequently forgotten when dealing with ideal gases but follow from the ideal gas equation of state, $P*V=R*T$. The pressure dependence for the ideal gas entropy is given by:

$$S^{ideal}(T,P) = S^{ideal}(T,1 \text{ atm.}) - R*\ln(P)$$

where the pressure, P , is given in atmospheres. The pressure dependence of other quantities can easily be calculated by substituting the pressure corrected entropy in their definitions.

Table 4.4-2
Sources for the Methods for Ideal Gas Properties

Enthalpy of Formation at 298 K:

HF25I1 - data base lookup (priority 1)
HF25I2 - Benson's method¹⁵ (priority 2)

Absolute Entropy at 298 K:

S25ID1 - data base lookup (priority 1)
S25ID2 - Benson's method¹⁵ (priority 2)

Gibbs Free Energy of Formation at 298 K:

GF25I1 - data base lookup (priority 1)
GF25I2 - calculated from HF25I1 and S25ID1 (priority 2)
GF25I3 - calculated from HF25I2 and S25ID2 (priority 3)

Enthalpy of Formation:

HFID2 - DIPPR data base equations (priority 1)
HFID3 - Benson's method¹⁵ (priority 2)

Gibbs Free Energy of Formation:

GFID2 - DIPPR data base equations (priority 1)
GFID3 - Benson's method¹⁵ (priority 2)

Formation Equilibrium Constant:

KID2 - calculated from GFID2 (priority 1)
KID3 - calculated from GFID3 (priority 2)

Table 4.4-2 (cont.)
Sources for the Methods for Ideal Gas Properties

Relative Enthalpy:

HID2 - calculated from DIPPR data (priority 1)
HID3 - Benson's method¹⁵ (priority 2)
HELM - data base lookup for the elements (special use)

Absolute Entropy:

SID2 - calculated from DIPPR data (priority 1)
SID3 - Benson's method¹⁵ (priority 2)
SELM - data base lookup for the elements (special use)

Gibbs Free Energy:

GID2 - calculated from HID2 and SID2 (priority 1)
GID3 - calculated from HID3 and SID3 (priority 2)
GELM - calculated from HELM and SELM (special use)

Helmholtz Free Energy:

AID2 - calculated from HID2 and SID2 (priority 1)
AID3 - calculated from HID3 and SID3 (priority 2)

Internal Free Energy:

UID2 - calculated from HID2 (priority 1)
UID3 - calculated from HID3 (priority 2)

Isobaric Heat Capacity:

CPID2 - data base lookup (priority 1)
CPID3 - Benson's method¹⁵ (priority 2)
CPELM - data base lookup for the elements (special use)

Isochoric Heat Capacity:

CVID2 - calculated from CPID2 (priority 1)
CVID3 - calculated from CPID3 (priority 2)

The calling sequences for the ideal gas property routines are more complicated than for the single value properties. Inputs are:

ASID	An integer array of Allied-Signal identifiers
SMILES	A character array of SMILES strings
NCMPDS	An integer field indicating the number of ASID's in the ASID array
TEMP	A real array of temperatures
NTEMP	An integer field indicating the length of the TEMP array
PRESS	A real array of pressures
NPRESS	An integer field indication the length of the PRESS array

Outputs are:

VALUE	A real 3D array of property values
ERROR	A real 3D array of the relative errors corresponding to the property values
IER	An integer 3D array of error codes

The new inputs and changed outputs reflect the fact that ideal gas properties are different for each compound, temperature, and pressure.

Methods for Residual Properties

The AFP property prediction system will eventually calculate residual thermodynamic properties using any one of the following equations of state:

1. Ideal Gas Equation: $P * V = R * T$
2. Second Virial Equation: $P = R * T * ((1/V) + (B/V^2))$ where
B is a function of temperature that must be retrieved or predicted.

3. Peng-Robinson Equation: $P = (R * T)/(V - b) - a/(V * (V + b) + b * (V - b))$ where a and b are given in Figure 4.4-7.
4. Redlich-Kwong Equation: $P = (R * T)/(V - b) - a/(V * (V + b))$ where a and b are given in Figure 4.4-7.
5. Soave Equation: $P = (R * T)/(V - b) - a/(V * (V + b))$ where a and b are given in Figure 4.4-7.
6. Van der Waals Equation: $P = (R * T)/(V - b) - a/V^2$ where a and b are given in Figure 4.4-7.
7. Lee-Kesler Equation ³⁴
8. Starling-Han Equation ³⁵
9. TRC Hydrocarbon Table j Equations ³⁶

The first six equations of state have been fully implemented during Phase I. Work will be done on equations 7-9 during Phase II as deemed necessary for the completion of Phase II.

Experimental values for the second virial equation are available in the DIPPR data base and have been used to calculate gas phase residual properties. Second virial coefficients are not valid for the liquid phase so these methods have not been programmed.

Equations 3-6 are all cubic equations in the molar volume as shown in Figure 4.4-7 ¹³. They can be solved using the algebraic solution for cubic equations ³⁷ or by iterative root solvers ³⁸. We have used the algebraic solution to test for the presence of 1, 2, or 3 real roots but found the actual roots using the iterative solver since this method could also be used for the more complex equations, 7-9.

EQUATION	u	w	$P = \frac{RT}{V - b} - \frac{a}{V^2 + ubV + wb^2}$	$\frac{b}{a}$
PENG-ROBINSON	2	-1	$\frac{0.07780 RT_c}{P_c}$	$\frac{0.45724 R^2 T_c^2}{P_c} [1 + f_\omega(1 - t_r^{1/2})]^2$
REDLICH-KWONG	1	0	$\frac{0.08664 RT_c}{P_c}$	$\frac{0.42748 R^2 T_c^{2.5}}{P_c T_c^{0.5}}$
SOAVE	1	0	$\frac{0.08664 RT_c}{P_c}$	$\frac{0.42748 R^2 T_c^2}{P_c} [1 + f_\omega(1 - t_r^{1/2})]^2$
VAN DER WAALS	0	0	$\frac{RT_c}{8 P_c}$	$\frac{27 R^2 T_c^2}{64 P_c}$
General			$t_r = T / T_c$	
PENG-ROBINSON:			$f_\omega = 0.37464 + 1.54226\omega - 0.26992\omega^2$	
SOAVE:			$f_\omega = 0.48 + 1.574\omega - 0.176\omega^2$	

Figure 4.4-7 Cubic Equations of State

Cubic equations of state are tricky to solve because there are non-physical and multiple roots which must be trapped out of the calculations. This is illustrated in Figure 4.4-8 which is a plot of the Peng-Robinson equation¹⁴. For temperatures above the critical point (e.g., $T_r^* = 2.7$), the high molar volume (low density) root (point C on Figure 4.4-8) is clearly the physical solution for the supercritical fluid. However, for reduced pressures greater than P_3 , there are two nonphysical roots (points A and B on Figure 4.4-8) at very small or negative molar volumes. Similarly, for temperatures below the critical point (e.g., $T_r = 0.9$), there are three regions of concern. For reduced pressures below P_1 , there is only one large molar volume root which is the pure gas phase. Between reduced pressures P_1 and P_2 , there are three roots: the largest is the vapor phase, the middle is nonphysical, and the smallest is the liquid. This is the two-phase region of the phase diagram. For reduced pressures greater than P_2 , only one small molar volume root exists for the liquid phase. The AFP software finds all of these roots and correctly assigns them for each case.

All equations of state are difficult to solve because different types of data are available in different problems. Sometimes, thermodynamic properties are calculated given values of temperature and pressure. At other times, the inputs might be temperature and molar volume or pressure and molar volume. These three cases are illustrated in Figure 4.4-9, the flow diagram for the cubic equation subroutines in the AFP software.

In the first column, T and P are known, but the molar volumes for the gas and liquid phases, V_g and V_l need to be calculated. In the second column, a molar volume and T is known, but the molar volume of the other phase (if it exists) and the pressure are desired. Finally, in the third column, a molar volume and P is known, but molar volume of the other phase (if it exists) and the temperature are desired.

* The reduced temperature and reduced pressure (t_r and p_r) are defined as the temperature and pressure divided by the critical temperature and critical pressure, respectively.

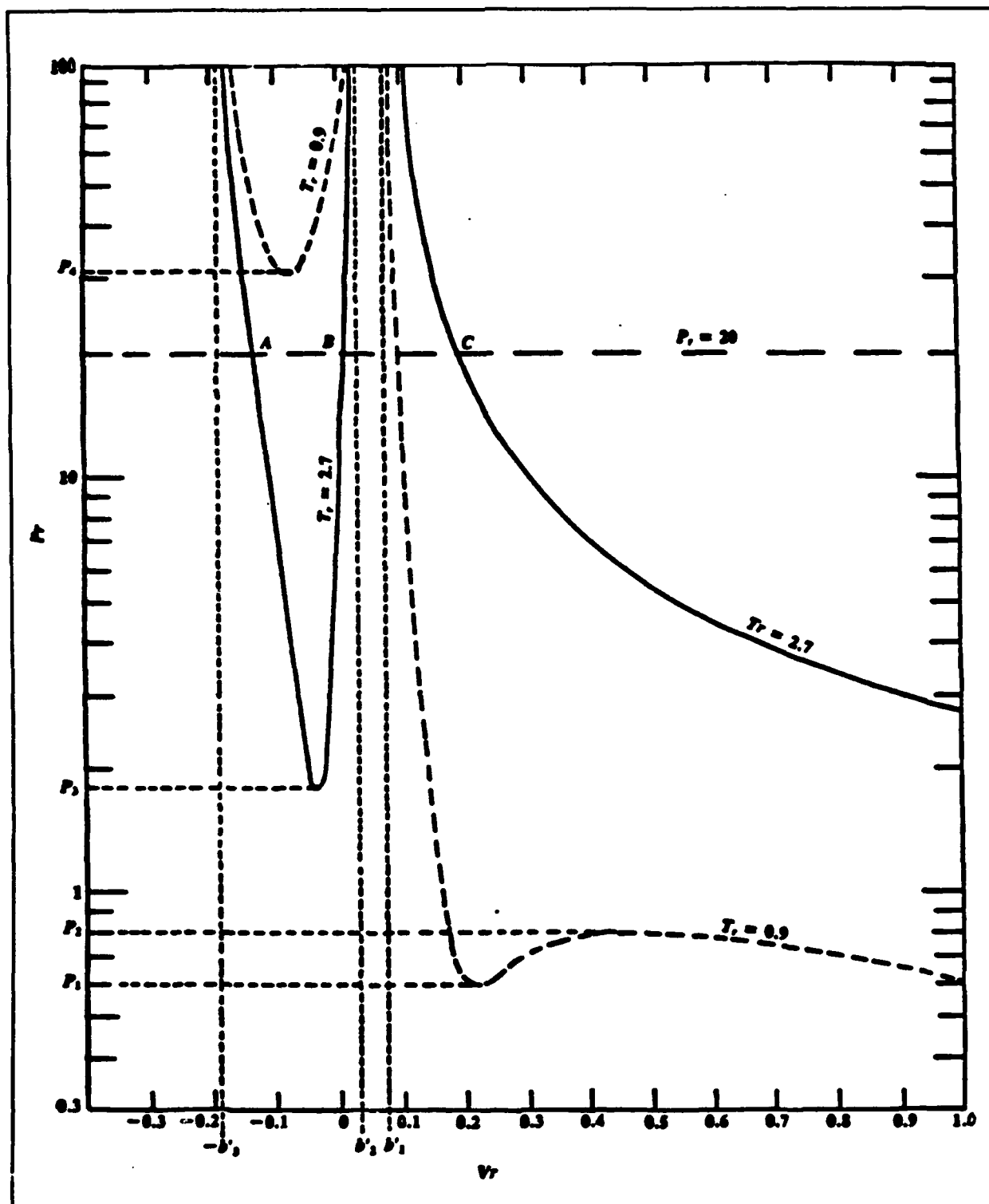


Figure 4.4-8 Reduced pressure versus reduced volume for the Peng-Robinson equation.

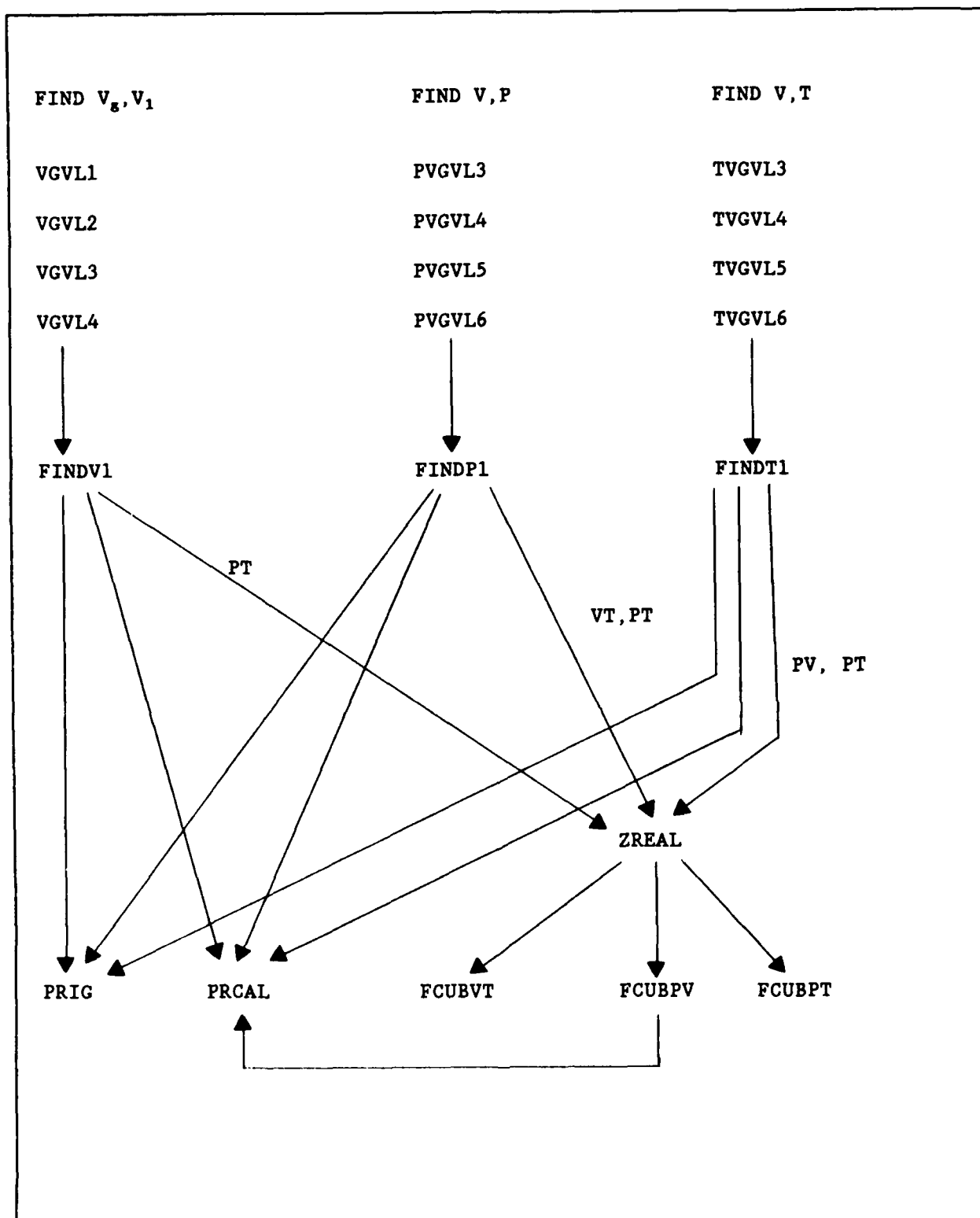


Figure 4.4-9 Cubic Equation Subroutines

In all three cases the equation of state is solved differently. In the first case, the cubic equation is solved for molar volumes by calling the root finder ZREAL with values of P and T and the function FCUBPT. In the second case, first the equation of state for pressure is solved by calling the root finder ZREAL with values of V and T and the function FCUBVT. Next, the missing molar volume is found by calling ZREAL with values of P and T and the function FCUBPT. In the third case, first solve the equation of state for temperature by calling the root finder ZREAL with values of P and V and the function FCUBPV. And next, the missing molar volume is found by calling ZREAL with values of P and T and the function FCUBPT.

Once the values of P, T, V_g , and V_l are known, we can calculate residual properties by inserting either gas or liquid phase molar volumes into the appropriate equations for the given equation of state. These equations are tabulated in Reid, Prausnitz, and Poling¹³ and Edminster and Lee¹⁴. They are derived by inserting the equations of state listed above into the following thermodynamic relations:

$$\text{Residual Enthalpy: } H^{res} = \int_0^P (V - T * (\delta V / \delta T)_P) dP$$

$$\text{Residual Entropy: } S^{res} = \int_0^P ((R/P) - (\delta V / \delta T)_P) dP$$

$$\text{Residual Internal Energy: } U^{res} = (R * T) - (P * V) +$$

$$\int_0^P (V - T * (\delta V / \delta T)_P) dP$$

$$\text{Residual Gibbs Free Energy: } G^{res} = \int_0^P (V - (R * T) / P) dP$$

Residual Helmholtz Free Energy: $A^{res} = (R * T) - (P * V) +$

$$\int_0^P (V - (R * T)/P) dP$$

Residual Isobaric Heat Capacity: $C_p^{res} = - \int_0^P (T * (\delta^2 V / \delta T^2)_P) dP$

Residual Isochoric Heat Capacity: $C_v^{res} = \int_{\infty}^V (\delta^2 P / \delta T^2)_V dV$

Fugacities: $\ln(f/P) = G^{res} / (R * T)$

The methods for residual properties are summarized in Table 4.4-4. For most of the properties, there are five methods corresponding to equations of state numbered 2 through 6 in the section on "Methods for Residual Properties" on 54. Residual properties are zero for ideal gases by definition.

The calling sequences for the residual property routines are slightly more complicated than for the ideal gas properties. Inputs are:

ASID	An integer array of Allied-Signal identifiers
SMILES	A character array of SMILES strings
NCMPDS	An integer field indicating the number of ASID's in the ASID array
TEMP	A real array of temperatures
NTEMP	An integer field indicating the length of the TEMP array
PRESS	A real array of pressures
NPRESS	An integer field indicating the length of the PRESS array

Table 4.4-3
Summary of the Methods for Residual Properties

Enthalpy:

- HRES2 - 2nd virial equation (priority 4)
- HRES3 - Peng-Robinson equation (priority 1)
- HRES4 - Van der Waals equation (priority 5)
- HRES5 - Redlich-Kwong equation (priority 3)
- HRES6 - Soave equation (priority 2)

Entropy:

- SRES2 - 2nd virial equation (priority 4)
- SRES3 - Peng-Robinson equation (priority 1)
- SRES4 - Van der Waals equation (priority 5)
- SRES5 - Redlich-Kwong equation (priority 3)
- SRES6 - Soave equation (priority 2)

Internal Energy:

- URES2 - 2nd virial equation
- URES3 - Peng-Robinson equation (priority 1)
- URES4 - Van der Waals equation (priority 4)
- URES5 - Redlich-Kwong equation (priority 3)
- URES6 - Soave equation (priority 2)

Gibbs Free Energy:

- GRES2 - 2nd virial equation (priority 4)
- GRES3 - Peng-Robinson equation (priority 1)
- GRES4 - Van der Waals equation (priority 5)
- GRES5 - Redlich-Kwong equation (priority 3)
- GRES6 - Soave equation (priority 2)

Helmholtz Free Energy:

- ARES2 - 2nd virial equation (priority 4)
 - ARES3 - Peng-Robinson equation (priority 1)
 - ARES4 - Van der Waals equation (priority 5)
 - ARES5 - Redlich-Kwong equation (priority 3)
 - ARES6 - Soave equation (priority 2)
-

Table 4.4-3 (cont.)
Summary of Methods for Residual Properties

Isobaric Heat Capacity:

CPRES2 - 2nd virial equation (priority 4)
CPRES3 - Peng-Robinson equation (priority 1)
CPRES4 - Van der Waals equation (priority 5)
CPRES5 - Redlich-Kwong equation (priority 3)
CPRES6 - Soave equation (priority 2)

Isochoric Heat Capacity:

CVRES2 - 2nd virial equation
CVRES3 - Peng-Robinson equation (priority 1)
CVRES4 - Van der Waals equation (priority 4)
CVRES5 - Redlich-Kwong equation (priority 3)
CVRES6 - Soave equation (priority 2)

Fugacities:

FUGAC2 - 2nd virial equation (priority 4)
FUGAC3 - Peng-Robinson equation (priority 1)
FUGAC4 - Van der Waals equation (priority 5)
FUGAC5 - Redlich-Kwong equation (priority 3)
FUGAC6 - Soave equation (priority 2)

STATE	An integer field indicating the state for calculation; 1 is for gases, 2 is for liquids, 3 is for solids
-------	---

Outputs are:

VALUE	A real 3D array of property values
ERROR	A real 3D array of the relative errors for the property values
IER	An integer 3D array of error codes

The same subroutine will be used for all the phases.

Methods for Real Gases

Given the methods for ideal gas and residual thermodynamic properties, a large number of methods for real gases can be constructed. These methods are summarized in Table 4.4-4. There are currently ten methods for most thermodynamic properties because there are two choices for the ideal gas properties (see Table 4.4-2) and five choices for the residual properties (see Table 4.4-3). The real gas thermodynamic methods do not have priorities of their own because they call priority level routines for the ideal gas and residual contributions.

The molar heat of combustion method uses the structural method ATMCNT to find the number of elements in the compound and then uses HFRG methods to calculate the change in enthalpy during the combustion reaction. Methods for heats of combustion per unit mass and unit volume will be completed by the time this report issues.

The calling sequences for these methods are the same as for the ideal gas properties.

Inputs are:

ASID	An integer array of Allied-Signal identifiers
SMILES	A character array of SMILES strings
NCMPDS	An integer field indicating the number of ASID's in the ASID array
TEMP	A real array of temperatures
NTEMP	An integer field indicating the length of the TEMP array
PRESS	A real array of pressures
NPRESS	An integer field indication the length of the PRESS array

Table 4.4-4
Summary of Methods for Real Gas Properties

Molar Volume:

- MVOLG2 - 2nd virial equation (priority 4)
- MVOLG3 - Peng-Robinson equation (priority 1)
- MVOLG4 - ideal gas equation (priority 6)
- MVOLG5 - Van der Waals equation (priority 5)
- MVOLG6 - Redlich-Kwong equation (priority 3)
- MVOLG7 - Soave equation (priority 2)

Compressibility:

- CMPR2 - 2nd virial equation (priority 4)
- CMPR3 - Peng-Robinson equation (priority 1)
- CMPR4 - ideal gas equation (priority 6)
- CMPR5 - Van der Waals equation (priority 5)
- CMPR6 - Redlich-Kwong equation (priority 3)
- CMPR7 - Soave equation (priority 2)

2nd Virial Coefficient:

- BRG2 - data base lookup (priority 1)
- BRG4 - ideal gas equation (priority 3)
- BRG5 - Van der Waals equation (priority 2)

Density:

- RHORG2 - 2nd virial equation (priority 4)
- RHORG3 - Peng-Robinson equation (priority 1)
- RHORG4 - ideal gas equation (priority 6)
- RHORG5 - Van der Waals equation (priority 5)
- RHORG6 - Redlich-Kwong equation (priority 3)
- RHORG7 - Soave equation (priority 2)

Table 4.4-4 (cont.)
Summary of Methods for Real Gas Properties

Enthalpy:

HRG - 10 methods (2 ideal gases * 5 residuals)

Entropy:

SRG - 10 methods (2 ideal gases * 5 residuals)

Internal Energy:

URG - 10 methods (2 ideal gases * 5 residuals)

Gibbs Free Energy:

GRG - 10 methods (2 ideal gases * 5 residuals)

Helmholtz Free Energy:

ARG - 10 methods (2 ideal gases * 5 residuals)

Isobaric Heat Capacity:

CPRG - 10 methods (2 ideal gases * 5 residuals)

Isochoric Heat Capacity:

CVRG - 10 methods (2 ideal gases * 5 residuals)

Enthalpy of Formation:

HFRG - 10 methods (2 ideal gases * 5 residuals)

Gibbs Free Energy of Formation:

GFRG - 10 methods (2 ideal gases * 5 residuals)

Molar Heat of Combustion:

HCRG - 10 methods (2 ideal gases * 5 residuals)

Outputs are:

VALUE	A real 3D array of property values
ERROR	A real 3D array of the relative errors for the property values
IER	An integer 3D array of error codes

Methods for Liquid Properties

The ideal gas and second virial coefficient equations of state are only valid for the gas phase. Thus, by combining the various ideal gas and residual property methods to obtain liquid property methods, one obtains eight liquid thermodynamic methods. This is due to having two ideal gas methods and four equations of state (Peng-Robinson, Redlich-Kwong, Soave, and van der Waals). Since DIPPR and the TRC tables contained experimental data for liquid heat capacities, liquid phase thermodynamics could be calculated by integrating over the heat capacities. Methods listed in Table 4.4-5 without priorities have not been implemented; if deemed technically necessary they will be programmed in Phase II.

Because of the enormous amount of work required to use equations of state to calculate liquid properties, several researchers have developed equations that are used for liquids only. The Rackett³⁹ and Hankinson-Brost-Thomson⁴⁰ equations have currently been programmed for liquid densities at saturation pressure. These equations are:

$$\text{Rackett Equation}^{39}: V_s = ((R * T_c)/P_c) * Z_{RA}^{(1+(1-Tr)^{2/7})}$$

$$\text{Hankinson-Brost-Thomson}^{40}: V_s = V^* * V_R^{(0)} * (1. - \omega_{SRK} * V_R^{(\delta)})$$

where $V_R^{(0)}$ and $V_R^{(\delta)}$ are known functions of $T_r = T/T_c$.

Table 4.4-5
Summary of Methods for Liquid Properties

Enthalpy:

HLQ2 - 8 methods (2 ideal gases * 4 residuals)
HLQ4 - integration of data base heat capacities

Entropy:

SLQ2 - 8 methods (2 ideal gases * 4 residuals)
SLQ4 - integration of data base heat capacities

Internal Energy:

ULQ2 - 8 methods (2 ideal gases * 4 residuals)
ULQ4 - integration of data base heat capacities

Gibbs Free Energy:

GLQ2 - 8 methods (2 ideal gases * 4 residuals)
GLQ4 - integration of data base heat capacities

Helmholtz Free Energy:

ALQ2 - 8 methods (2 ideal gases * 4 residuals)
ALQ4 - integration of data base heat capacities

Isobaric Heat Capacity:

CPLQ2 - 8 methods (2 ideal gases * 4 residuals)
CPLQ4 - integration of data base heat capacities

Isochoric Heat Capacity:

CVLQ2 - 8 methods (2 ideal gases * 4 residuals)
CVLQ4 - integration of data base heat capacities

Enthalpy of Formation:

HFLQ2 - 8 methods (2 ideal gases * 4 residuals)

Gibbs Free Energy of Formation:

GFLQ2 - 8 methods (2 ideal gases * 4 residuals)

Molar Heat of Combustion:

HCLQ2 - 8 methods (2 ideal gases * 4 residuals)

Table 4.4-5 (cont.)
Summary of Methods for Liquid Properties

Saturated Molar Volumes:

MVLQS2 - Hankinson-Brost-Thomson equation ⁴¹ (priority 2)
MVLQS3 - Rackett Equation ⁴⁰ (priority 3)
MVLQS4 - data base lookup (priority 1)
MVLQS5 - Peng-Robinson equation
MVLQS6 - Soave equation

Compressed Molar Volumes:

MVLQ2 - Tait-HBT ⁴² (priority 1)
MVLQ3 - Density Correlation ⁴³ (priority 2)
MVLQ4 - Peng-Robinson equation
MVLQ5 - Soave equation

Densities:

RHOLQ2 - calculated from MVLQ3 (priority 3)
RHOLQ3 - calculated from MVLQ2 (priority 2)
RHOLQ4 - calculated from MVLQS (priority 1)
RHOLQ5 - Peng-Robinson equation
RHOLQ6 - Soave equation

For the liquid densities under compression, the Tait-Hankinson-Brost-Thomson⁴¹ and Density Correlation⁴² equations:

Tait-HBT⁴¹: $V = V_s * (1 - c * \ln((\beta + P)/(\beta + P_{vap})))$ where c is a known function of ω_{SRK} and β is a function of T_r and ω_{SRK} .

Density Correlation⁴²: $V_1 = V_2 * C_2 / C_1$ where the correlation coefficients are known functions of the reduced temperature and pressure.

The methods for liquids are summarized in Table 4.4-5. The methods that use ideal gas and residual properties to calculate liquid thermodynamic properties do not have priorities because they call priority subroutines for the ideal gas and residual properties. The methods based upon integrating the experimental heat capacities currently do not have priorities assigned since they have not been completed. However, they will be completed soon and at that time priorities will be assigned to them.

The calling sequences for these methods are the same as for the ideal gas properties.

Inputs are:

ASID	An integer array of Allied-Signal identifiers
SMILES	A character array of SMILES strings
NCMPDS	An integer field indicating the number of ASID's in the ASID array
TEMP	A real array of temperatures
NTEMP	An integer field indicating the length of the TEMP array
PRESS	A real array of pressures
NPRESS	An integer field indication the length of the PRESS array

Outputs are:

VALUE	A real 3D array of property values
ERROR	A real 3D array of the quality indicators for the property values
IER	An integer 3D array of error codes

Methods for Phase Transitions

The properties of the liquid-gas phase transition may be calculated by comparing the fugacities for each of the two phases. These fugacities are equal at the equilibrium phase transition temperature and pressure. To calculate the boiling point at a given pressure, the temperature is varied until the fugacities are equal. The temperature where they are equal is the boiling point. Similarly, to calculate the vapor pressure at a given temperature, the pressures

of the two phases are varied until the fugacities are equal. The resulting pressure is the vapor pressure. Once the equilibrium temperature and pressure are known, the enthalpy and entropy of vaporization can be calculated by subtracting the residual enthalpy and entropy of the liquid phase from the residual enthalpy and entropy of the gas phase.

Since these calculations require an accurate equation of state, only the Peng-Robinson, Soave, Lee-Kesler, Starling-Han and experimental equations of state will be implemented. An example of these calculations using the Peng-Robinson equation is presented in Table 4.4-6 for methylcyclohexane.

Table 4.4-6
Comparison Between Peng-Robinson Predictions and Experiment

	<u>Peng-Robinson Equation</u>	<u>Experimental Values</u>	<u>Units</u>
$V_1(298)$	1.2422×10^{-1}	1.2818×10^{-1}	m^3/kmol
$V_1(293)$	1.2365×10^{-1}	1.2762×10^{-1}	m^3/kmol
T_b	3.7396×10^2	3.7408×10^2	K
$\Delta H_{\text{vap}}(T_b, 1 \text{ atm})$	3.105×10^7	3.11×10^7	J/kmol
$\Delta S_{\text{vap}}(T_b, 1 \text{ atm})$	8.302×10^4	8.322×10^4	J/kmol·K
P_b	6.563×10^3	6.133×10^3	Pa
$\Delta H_{\text{vap}}(298, P_b)$	3.436×10^7	3.536×10^7	J/kmol·K
$\Delta S_{\text{vap}}(298, P_b)$	1.152×10^5	1.186×10^5	J/kmol·K
$S_{\text{gas}}(298, 1 \text{ atm})$	3.425×10^5	3.433×10^5	J/kmol·K
$S_{\text{liq}}(298, 1 \text{ atm})$	2.509×10^5	2.479×10^5	J/kmol·K

The Peng-Robinson equation is quite accurate for calculations of the molar liquid volumes at 298K and 293K, boiling point at 1 atm, the enthalpy and entropy of vaporization at the boiling point and 1 atm, the vapor pressure at 298K, the enthalpy and entropy of vaporization at 298K and the vapor pressure (at 298K), and for the entropies of the gas and liquid phases at 298K and 1 atm.

In addition to the equation of state methods, there are specialized methods for calculating phase transition properties. So far only the Riedel⁴³ and Lee-Kesler⁴⁴ methods have been programmed. During Phase II more of these methods (such as the Two Reference Fluid Equation⁴⁵ will be programmed since these extend the mixture capabilities of the system. During Phase II the system will be extended to handle phase transition methods which include vapor-liquid equilibria of mixtures (distillations) and liquid-liquid equilibria of mixtures (solubilities). The liquid-gas phase transition methods are summarized in Table 4.4-7.

The calling sequences for the phase transition methods vary from one method to another because some properties are only temperature dependent such as vapor pressures while others are both temperature and pressure dependent. The calling sequences for the vapor pressure methods do not contain the PRESS and NPRESS variables found in the real gas and liquid methods. The calling sequences for the enthalpy and entropy of vaporization are the same as for the real gas and liquid methods.

Table 4.4-7
Summary of the Methods for Liquid-Gas Phase Transitions

Vapor Pressures:

PVAPS2 - data base lookup (priority 1)
PVAPS3 - Riedel's method ⁴³ (priority 2)
PVAPS4 - Lee-Kesler method ⁴⁴ (priority 3)
PVAPS5 - Two Reference Fluid Method⁴⁵
PVAPS6 - Peng-Robinson equation
PVAPS7 - Soave equation

Boiling Point Correction:

DTDPB1 - data base lookup (priority 1)
DTDPB3 - Peng-Robinson equation
DTDPB4 - Soave equation

Enthalpy of Vaporization:

HVSAT2 - data base lookup (priority 1)
HVSAT3 - Peng-Robinson equation
HVSAT4 - Soave equation

Entropy of Vaporization:

SVSAT3 - Peng-Robinson equation
SVSAT4 - Soave equation

Methods for Transport Properties

During Phase I, the level of effort was not as high for transport properties as for the thermodynamic properties. This emphasis was an outgrowth of the strategy to focus on the single value properties and then the thermodynamic properties. An API method⁴⁶ and an AIChE method⁴⁷ for the viscosity of gases have been programmed. The methods currently available for transport properties are summarized in Table 4.4-8.

The calling sequences for these properties are the same as for thermodynamic properties of real gases or liquids.

Table 4.4-8
Summary of Methods for Transport Properties

Liquid Viscosity:

NVL\1 - data base lookup (priority 1)

Vapor Viscosity:

NUVAP2 - API method 11B1-6 ⁴⁶ (priority 2)

NUVAP3 - AIChE method 8A ⁴⁷ (priority 3)

NUVAP4 - data base lookup (priority 1)

Liquid Thermal Conductivity:

LTC1 - data base lookup (priority 1)

Vapor Thermal Conductivity:

VAPTC1 - data base lookup (priority 1)

Methods for Solid Properties

There are experimental data in the DIPPR data base and TRC Hydrocarbon tables for the heat capacity and density of solids. Only the data base lookup methods have been programmed. If During Phase II it is determined that thermodynamic properties of solids are necessary, a method which integrates the heat capacity equation will be programmed. This work will be a relatively low priority because, for aviation applications, fuels must be fluids. The methods for solids are summarized in Table 4.4-9. Their calling sequence is the same as for real gases or liquids.

Methods for Mixtures

Equation of state methods for pure components are easily transferred over to mixtures using mixing laws for the parameters in the equations. For the cubic

Table 4.4-9
Summary of Methods for Solids

Solid Heat Capacity:

CFS1 - data base lookup (priority 1)

Solid Density:

RHOS1 - data base lookup (priority 1)

equations of state, these mixing rules are:

$$a_m = \sum_i \sum_j x_i * x_j * (a_i * a_j)^{1/2} * (1 - k_{ij})$$

$$b_m = \sum_i x_i * b_i$$

where the x_i 's are mole fractions of components, the a_i 's and b_i 's are the pure component coefficients, and the k_{ij} 's are binary interaction parameters. These equations have already been programmed into the Phase I equation of state methods, therefore the system will be able to calculate mixture thermodynamic properties as soon as there are predictive methods for the binary interaction parameters. During Phase II, the whole phase fugacity calculations will be extended to predictions of the fugacities of individual components in the mixture. These fugacities will be used to solve vapor-liquid and liquid-liquid equilibrium problems.

Methods for Error Tracking

In order to assess the accuracy of the predictive methods, experimental and known predictive errors were propagated through the subroutines programmed during Phase I. This method of propagating errors tends to over-estimate errors because errors that should cancel but occur in different subroutines are added rather than subtracted. To overcome this problem, relative errors for the properties are starting to be determined by numerically calculating partial derivatives of

the answers with respect to the input experimental data. For example, if the experimental errors were in T_c and P_c , the absolute error in the resulting thermodynamic functions would be:

$$\text{error in } f = \left| \delta f / \delta T_c \right| * (\text{error in } T_c) + \left| \delta f / \delta P_c \right| * (\text{error in } P_c).$$

These error calculations have been programmed for densities and will be extended to other properties during Phase II.

4.5 New Model Development

Objective:

To develop new property predictive methods based solely on the structure of the molecule of interest.

Work Completed:

The development of new models for the prediction of physical and thermochemical properties was based upon the use of simple groups, graph theory parameters, and structural descriptors such as the number of non-hydrogen atoms in the molecule and the molecular weight. This approach was chosen since it is well known that each of these types of parameters can be used to predict properties. The intent was to allow statistical analysis to provide the best choice of graph theory parameters, group contributions, structural descriptors, or any combination to be used for the new method.

Three types of groups were defined: zero order groups (atoms), zero order groups with hybridization, and bonds. The groups considered for each class are shown in Table 4.5-1.

The graph theory parameters that were used are shown in Table 4.5-2. This table also lists the structural descriptors used in the new methods development. The graph theory parameters chosen were done so because of available software

Table 4.5-1
Simple Group Types for New Methods Development

Zero Order Groups

C	N	O	P	S	F
Cl	Br	I	H	D	T

Zero Order Groups with Hybridization

F	#C	C	-C	-C-	-C<
Csp3	Csp2	-C=	Csp	Carom	N
-N	-N-	Nsp3	>N<	#N(-)-	Nsp
Narom	Narom-H	Nsp2	-N(-)-	-N	Osp3
Osp2	Oarom	-Cl	-Cl=	-Cl(-)-	
-Cl(-)(-)-		-Br	-Br(-)-	-I	-I(-)-
-I(-)(-)-		-S-	-S	-S<	Sarom
-S(-)(-)-		S(-1)	-P(-)(-)-	-P<	P
H	D	T			

Bonds

C-C	C-Carom	C=C	C#C	C-N	C=N
C#N	C-O	C=O	C-Cl	C-Br	C-F
C-I	C-S	C-S	C-P	C-P	Carom-H
Carom-Carom		Carom:Carom	Carom-N	Carom=N	
Carom-Narom		Carom:Narom	Carom-O	Carom=C	
Carom=O		Carom-Oarom	Carom:Oarom	Carom-Cl	
Carom-Br		Carom-F	Carom-I	Carom-S	
Carom-P		N-N	N=N	N#N	
N-Narom		N-O	N=O	N-Cl	N-Br
N-F	N-I	N-S	N-P	Narom-H	
Narom-Narom		Narom:Narom	Narom-O	Narom-Oarom	
Narom:Oarom		Narom-Cl	Narom-Br	Narom-F	
Narom-I		Narom-S	Narom-P	O-O	O-Cl
O-Cl	O-Br	O=Br			
O-F	O=F	O-I	O=I	O-S	O=S
O-P	O=P	Oarom-H	Cl-H	Cl-Cl	Cl-Br
Cl-F	Cl-I	Cl-S	Cl=S	Cl-P	Cl=P
Br-H	Br-Br	Br-F	Br-I	Br-S	Br=S
Br-P	Br=P	F-H	F-F	F-I	F-S
F-P	I-H	I-I	I-S	I=S	I-P
I-P	S-S	S=S	S-P	S=P	
Sarom:Carom		P-P	P=P	C-H	N-HP-H
S-H	O-H	C-D	c-D	D-D	D-H
D-O	D-S	D-N	D-T	T-T	

*D and T are deuterium and tritium, respectively.

(MOLCONN2⁴⁸) that can automatically derive these parameters from a SMILES string. Other types of parameters were also considered. These parameters consisted of transformed or derived variables. For example, the square root of the first order path connectivity index was considered; examples of some of these "transformed" or derived parameters are shown in Table 4.5-3. An explanation of why these parameters were considered will be discussed later in the report (See 81).

Table 4.5-2
Graph Theory Parameters and Structural Descriptors
for New Methods Development

Connectivity Indices

<u>Symbol</u>	<u>Name</u>	<u>Order</u>
$^{0-20}X_p$	Simple Path	0 - 20
$^{0-20}X_p^v$	Valence Path	0 - 20
^3X_c	Simple Cluster	3
$^3X_c^v$	Valence Cluster	3
$^4X_{pc}$	Simple Path/Cluster	4
$^4X_{cc}^v$	Valence Path/Cluster	4
$^{0-20}X_{ch}$	Simple Chain	3 - 20
$^{0-20}X_{ph}^v$	Valence Chain	3 - 20

Other Parameters

Total Topological State Index
 Wiener Number
 Total Wiener
 Shannon Index
 Kappa Zero Index
 Kappa Simple Indices (first to third order)
 Kappa Indices (first to third order)

Structural Descriptors

Atom Count, nonhydrogen
 Atom Count, hydrogen
 Molecular Weight

At this point, it is necessary to define the various graph theory parameters. The connectivity indices are based the encoding of structural information according to the connectivity of the nonhydrogen atoms in a molecule. The first parameter that needs to be define is δ . Delta is the number of nonhydrogen atoms attached to the atom of interest. Therefore, a methylene ($-\text{CH}_2-$) group has a δ of 2.

The simple path connectivity χ_p of order m is given by

$${}^m\chi_p = \sum_{i=1}^{N_s} \left(\prod_{k=1}^{m+1} (\delta_k)^{-0.5} \right)$$

Table 4.5-3
Transformed and Derived Parameters
for New Methods Development

Transformed Parameters

Reciprocal Zero-Order Path Index
Reciprocal First-Order Path Index
Reciprocal Second-Order Path Index
Reciprocal Atom Count
Reciprocal Molecular Weight

Derived Parameters

Sum Zeroth-Order Simple and Valence Connectivity Indices
Sum First-Order Simple and Valence Connectivity Indices
Sum Second-Order Simple and Valence Connectivity Indices
Difference Zeroth-Order Simple and Valence Connectivity Indices
Difference First-Order Simple and Valence Connectivity Indices
Difference Second-Order Simple and Valence Connectivity Indices

where N_s is the number of paths of order m , and δ_j are the delta values for the $m+1$ atoms in the path. As an example, the first order (simple) path connectivity index for di thyl ether would be given by

$$\begin{aligned} {}^1\chi_p &= (1*2)^{-0.5} + (2*2)^{-0.5} + (2*2)^{-0.5} + (2*1)^{-0.5} \\ &= 2.414 \end{aligned}$$

However, it should be noted that this value would be the same for n-pentane. To account for heteroatoms and multiply bonded carbon atoms, one can use what is referred to as the valence connectivity indices and are designated by a trailing superscript, v. The difference between the simple connectivity indices and the valence indices is in the definition of δ . Whereas for the simple indices, δ is just the number of nonhydrogen atoms attached to the atom of interest, for the valence indices, δ^v (the valence delta) replaces δ . δ^v is defined as

$$\delta^v = Z^v - h$$

where Z^v is the number of valence electrons in the atom of interest and h is the number of hydrogens on that atom. The effect that this would have on our previous example of diethyl ether, for the first-order valence path connectivity index, is given by

$$\begin{aligned} {}^1\chi_p &= (1*2)^{-0.5} + (2*6)^{-0.5} + (6*2)^{-0.5} + (2*1)^{-0.5} \\ &= 1.992 \end{aligned}$$

This example, together with the previous example shows the value of the valence connectivity indices and why they were considered in the new methods development.

Additionally other types of molecular fragments can be considered as being part of a molecule and were also considered as part of the graph theory based new method development. These fragments are most easily depicted as:

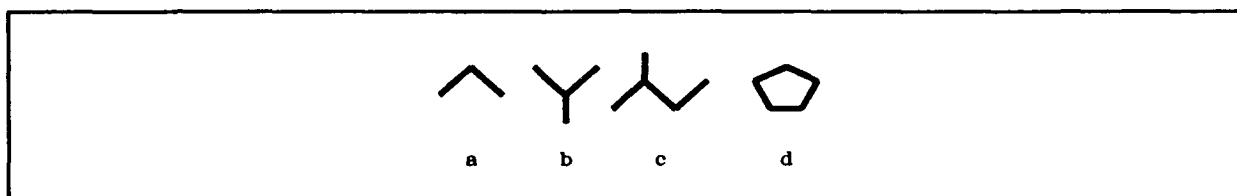


Figure 4.5-1 Different types of molecular fragments used in graph theory

These fragments are referred to as a path (a), cluster (b), path/cluster (c), and a chain (d).

The topological state index encodes information (using the valence delta described above) about each atom in a molecule and how it relates to all of the paths in the molecule in which it is involved. The total topological state index is the sum of all of the topological state indices for each atom in the molecule. The molecular shape indices (Kappa and Simple Kappa indices) are included to account for properties that may have a molecular shape dependence. The Kappa indices are based on the number of one, two, and three bond fragments in a molecule, relative to the minimum and maximum number of fragments possible for real or hypothetical molecules having the same number of atoms as the molecule of interest.

The overall concept is to encode structural information into a series of parameters that in some combinations will provide property predictive capabilities. Further, more detailed information on the various graph theory parameters, graph theory in general, and graph theory in property prediction can be found in several books^{49,50,51}.

The initial approach was to find a new predictive method for the normal boiling point using all of the experimental data in the AFP data base. The procedure employed was to use the SAS⁵² procedures such as STEPWISE and RSQUARED. The experimental data and all of the parameters mentioned above were provided to the statistical analysis software to determine what linear combinations of the various parameters would provide a good method for the prediction of properties.

As an example, the first property that was investigated was the normal boiling point (designated TNBP). One of the initial efforts undertaken was to plot each of the parameters versus the normal boiling points for each family. An example plot is shown in Figure 4.5-2, where the normal boiling points for the normal alkanes are plotted versus ${}^0\chi_p$. It is quite evident that the relationship is nonlinear. Figure 4.5-3 shows an example of predictions of normal boiling points based on ${}^1\chi_p$. The results of these plots indicated that not only should

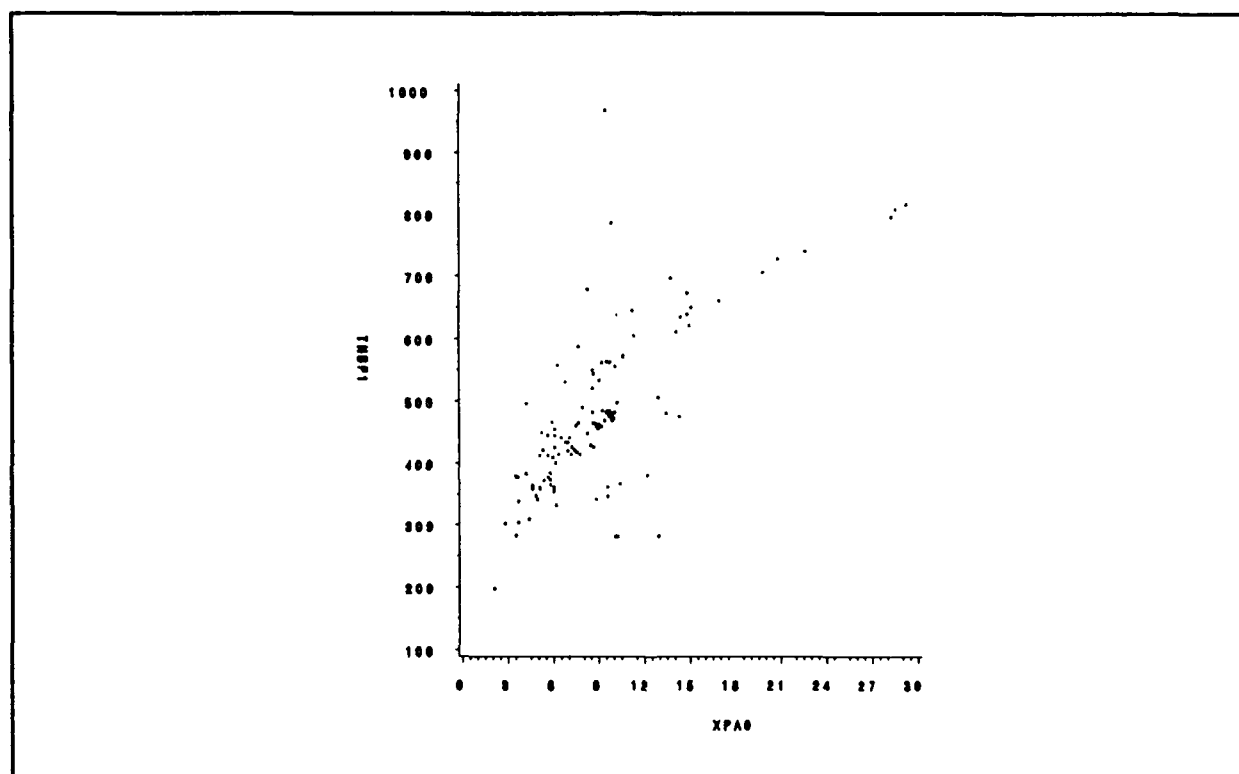


Figure 4.5-2 Plot of experimental normal boiling points (TNBP1) versus ${}^0\chi_p$ (XPA0) for the normal alkanes (Temperature in K)

the various parameters be examined, but also transforms of the parameters. Two such transforms were examined, the square root and the natural logarithm of the parameter. A plot of the latter transform is plotted in Figure 4.5-4 and shows how this helped to linearize the relationship between the normal boiling point and ${}^0\chi_p$.

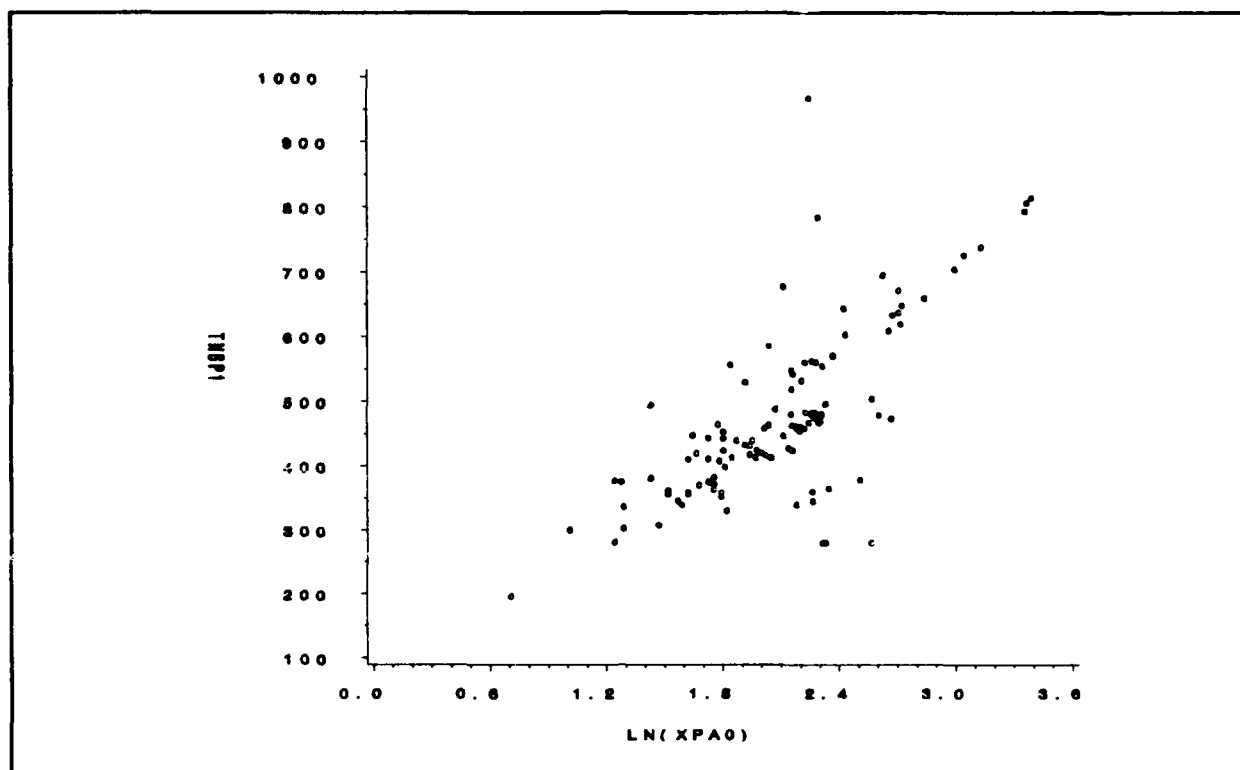


Figure 4.5-3 Plot of experimental normal boiling points (TNBP1) versus $\ln(^0\chi_p)$ (LN(XPA0)) for the normal alkanes

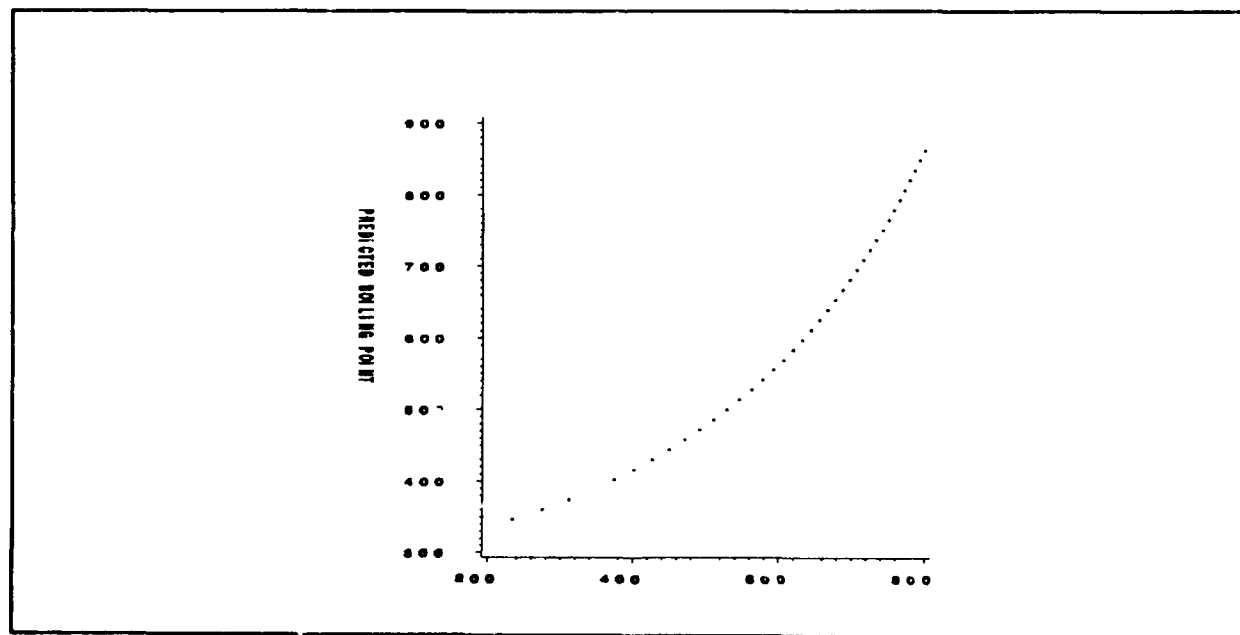


Figure 4.5-4 Experimental (TNBP1) versus predicted normal boiling points where the predictions are based on $^1\chi_p$ (Temperatures in K)

The initial attempt to find a single equation which would predict the normal boiling point of any arbitrary compound, based on the couple of thousand normal boiling points in the AFP data base, did not prove successful. The results up to an eight-parameter fit are given in Table 4.5-4. As is evident even with an eight-parameter fit, the R^2 just got above 0.8. Also, the last several increases in R^2 with the number of parameters used in the regression shows that increasing the number of parameters will not cause the R^2 to increase significantly. Therefore, it is clear that one equation can not be used to predict the normal boiling points of all classes of compounds.

The next strategy was to do the regression by classes of compounds. (The class of each compound was determined by the AFP family classification routine discussed earlier). For the normal alkanes, Figures 4.5-5, 6, and 7 show the predicted values based on the first order simple path connectivity index ($^1\chi_p$), the residuals (experimental values minus predicted values), and the relative errors (residuals divided by the experimental values) versus the experimental normal boiling points, respectively. These plots show that even for the normal alkanes over the range of compounds that are in the AFP data base, that the normal boiling points cannot be predicted accurately using only one parameter. However, going to the best (as determined by the SAS⁵² STEPWISE procedure) five-parameter fit shows a good correlation between experimental and predicted values, Figure 4.5-7. This can especially be seen in the residual and relative error plots, Figures 4.5-8 and 9, respectively. Doing the regressions by family and examining the two-parameter regressions, one finds that the R^2 's are between 0.91 and 0.9997, except for 12 families of compounds accounting for only 448 compounds. One know that it would be possible to obtain good predictive capabilities if each family of compounds was analyzed individually, but the question arose as to whether some of the families could be combined to reduce the very large number of equations (there are over 70 families of compounds in the AFP system).

The initial attempt at combining families was to combine all of the hydrocarbon families (compounds containing only carbon and hydrogen) and do the regression analysis. The graphical results of the experimental versus predicted, residuals, and relative errors are shown in Figures 4.5-10, 11, and 12. The

Table 4.5-4
General Normal Boiling Point Regression Results

<u>R²</u>	<u>Variables Used*</u>							<u>Resulting</u>
	MW							0.696
	¹ X _P	WWT						0.726
	¹ X _P	⁹ X _{CH}	WWT					0.751
	¹ X _P	⁹ X _{CH}	⁰ K	WWT				0.766
	¹ X _P	⁶ X _{CH}	⁶ X _P ^v	WWT	C-Carom			0.793
	¹ X _P	⁶ X _{CH}	⁶ X _P ^v	⁰ K	WWT	C-Carom		0.799
	¹ X _P	⁶ X _{CH}	¹⁰ X _{CH}	⁶ X _P ^v	⁰ K	WWT	C-Carom	0.803
	¹ X _P	¹¹ X _P	⁶ X _{CH}	¹⁰ X _{CH}	⁶ X _P ^v	⁰ K	WWT C-Carom	0.807

*The definition of the variables is:

MW	Molecular weight
¹ X _P	First-order simple path connectivity index
WWT	Total Wiener number
⁹ X _{CH}	Ninth-order simple chain connectivity index
⁰ K	Zeroth-order simple kappa shape index
⁶ X _{CH}	Sixth-order simple chain connectivity index
⁶ X _P ^v	Sixth-order valence path connectivity index
C-Carom	Number of aliphatic carbon - aromatic carbon bonds
¹⁰ X _{CH}	Tenth-order simple chain connectivity index
¹¹ X _P	Eleventh-order simple path connectivity index

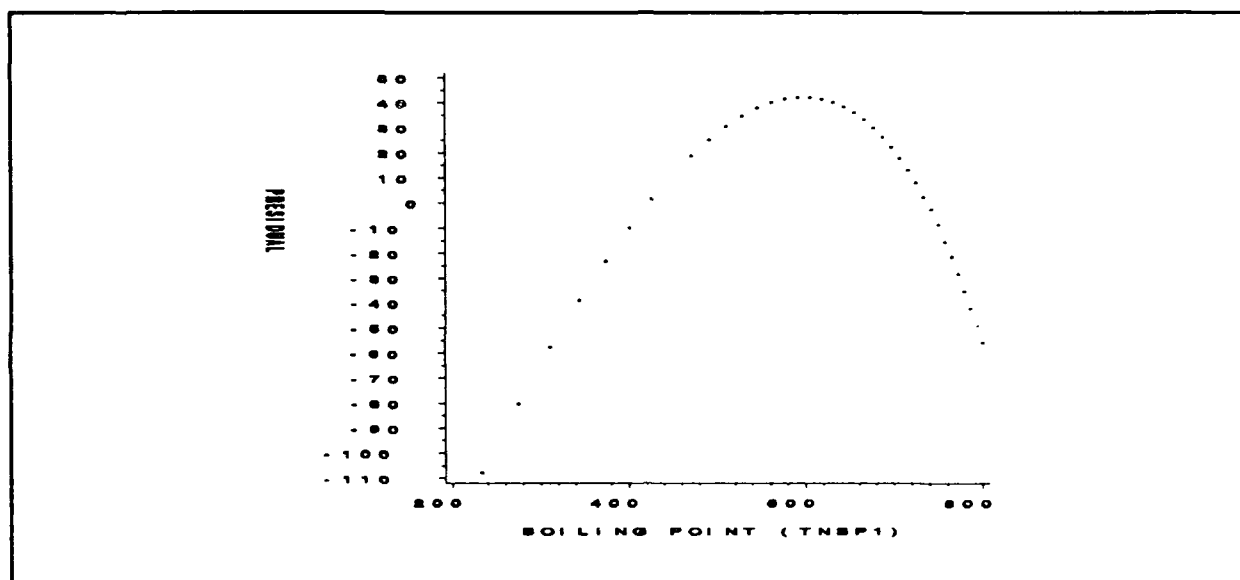


Figure 4.5-5 Experimental (TNBP1) versus residual normal boiling points where the predictions are based on ${}^1\chi_p$ (Temperatures in K)

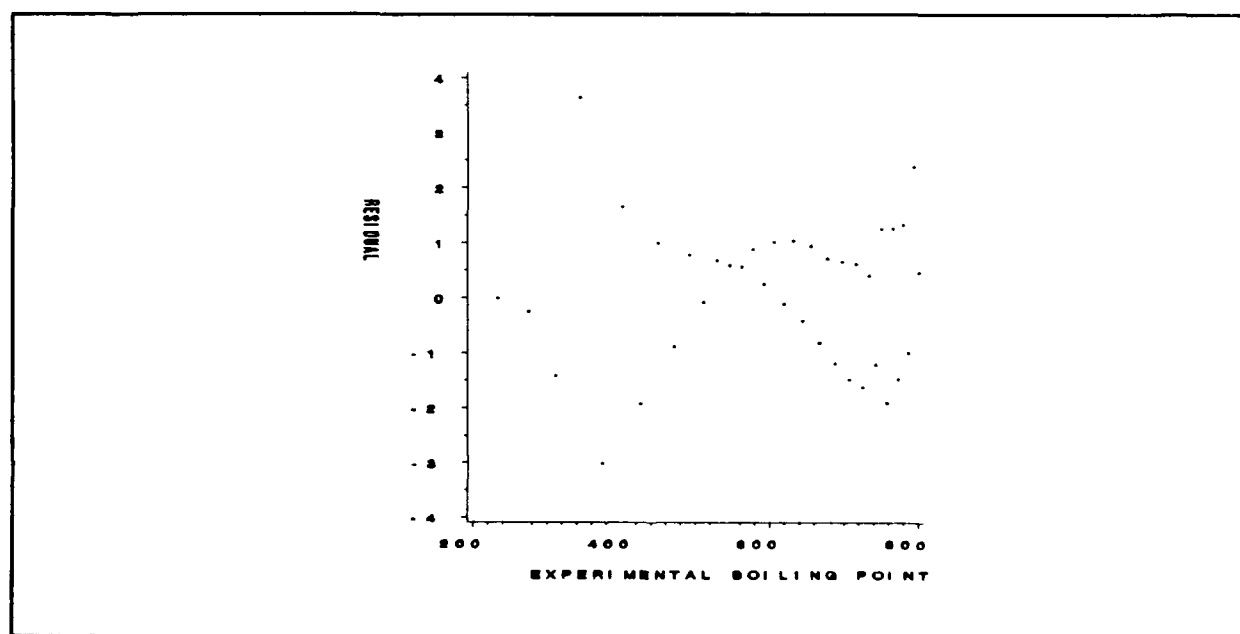


Figure 4.5-6 Experimental boiling points (TNBP1) versus residuals, where the predictions are based on ${}^1\chi_p$, ${}^4\chi_p$, ${}^1K$, Shannon index, and ${}^0K$ (Temperatures in K)

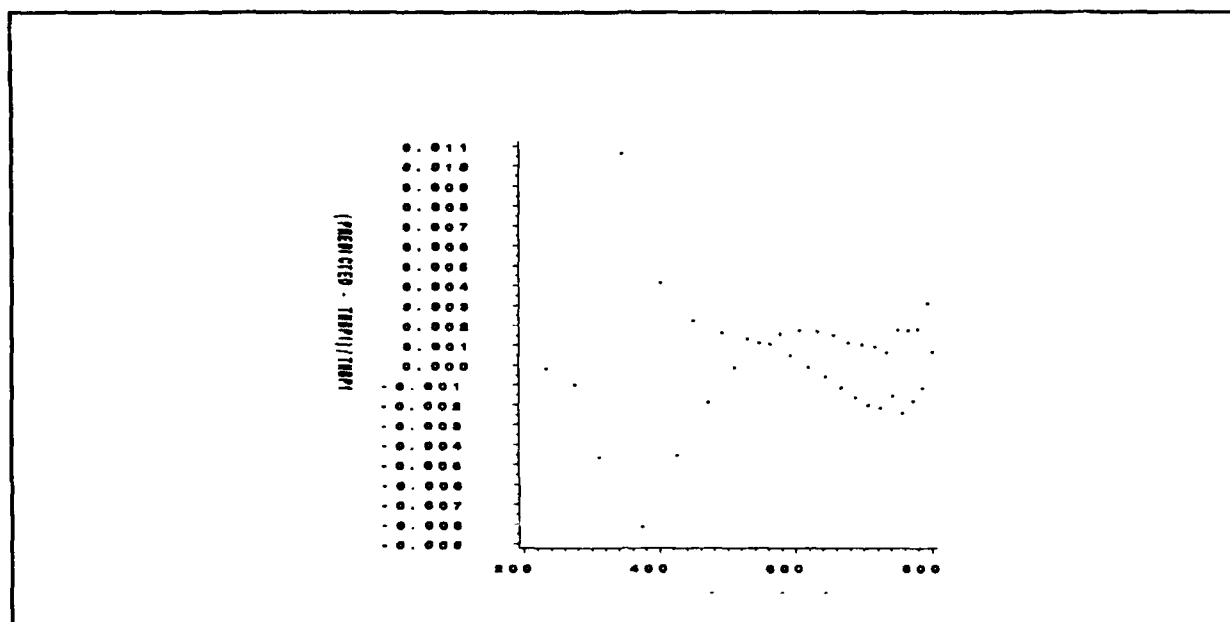


Figure 4.5-7 Experimental boiling points (TNBP1) versus relative errors; predictions are based on $^1\chi_p$, $^4\chi_p$, ^1K , Shannon index, and ^0K (Temperatures in K)

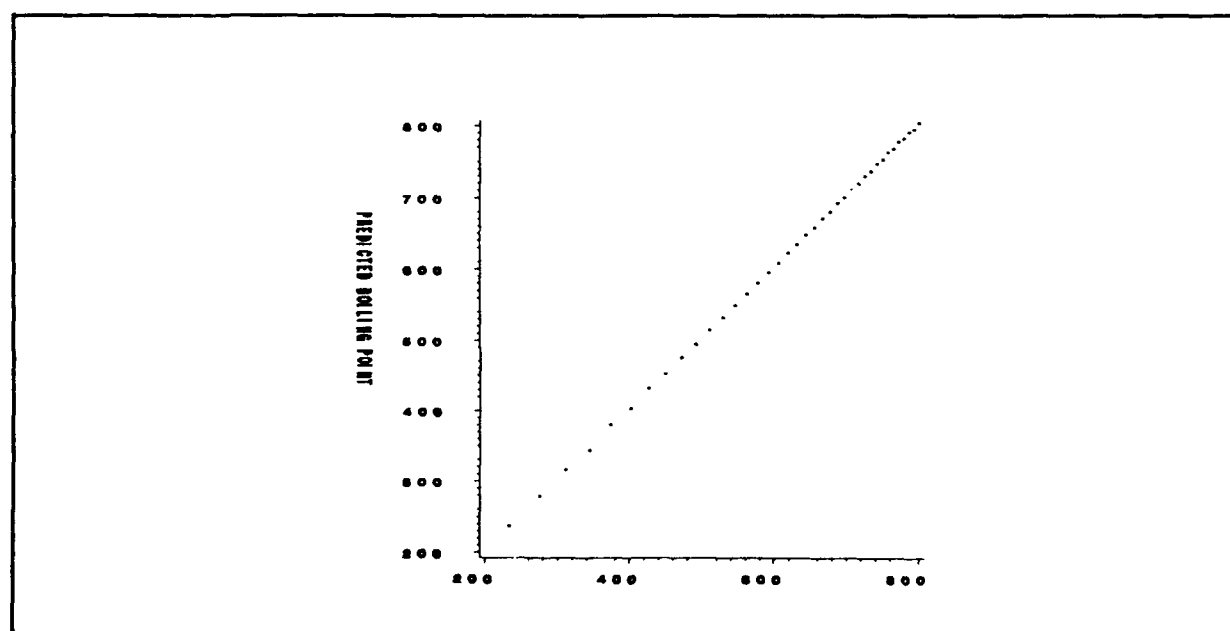


Figure 4.5-8 Experimental (TNBP1) versus predicted normal boiling points based on $^1\chi_p$, $^4\chi_p$, ^1K , Shannon index, and ^0K (Temperatures in K)

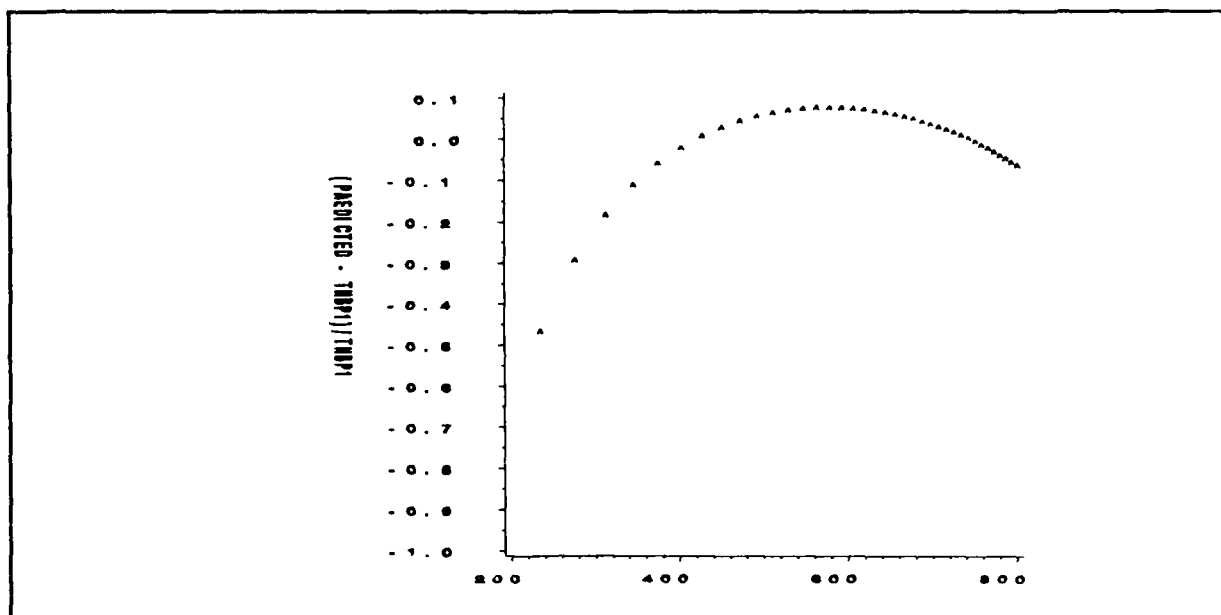


Figure 4.5-9 Experimental (TNBP1) versus relative errors for normal boiling points where the predictions are based on $^1\chi_p$ (Temperatures in K)

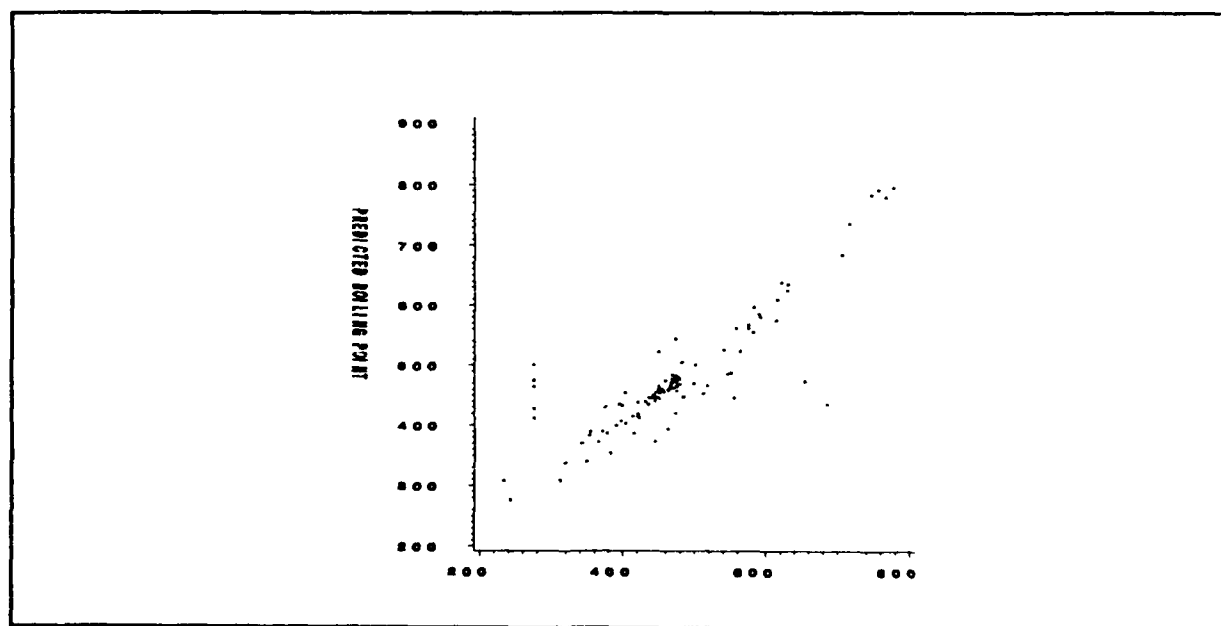


Figure 4.5-10 Experimental versus predicted for all of the hydrocarbon families combined (model basis in text) (Temperatures in K)

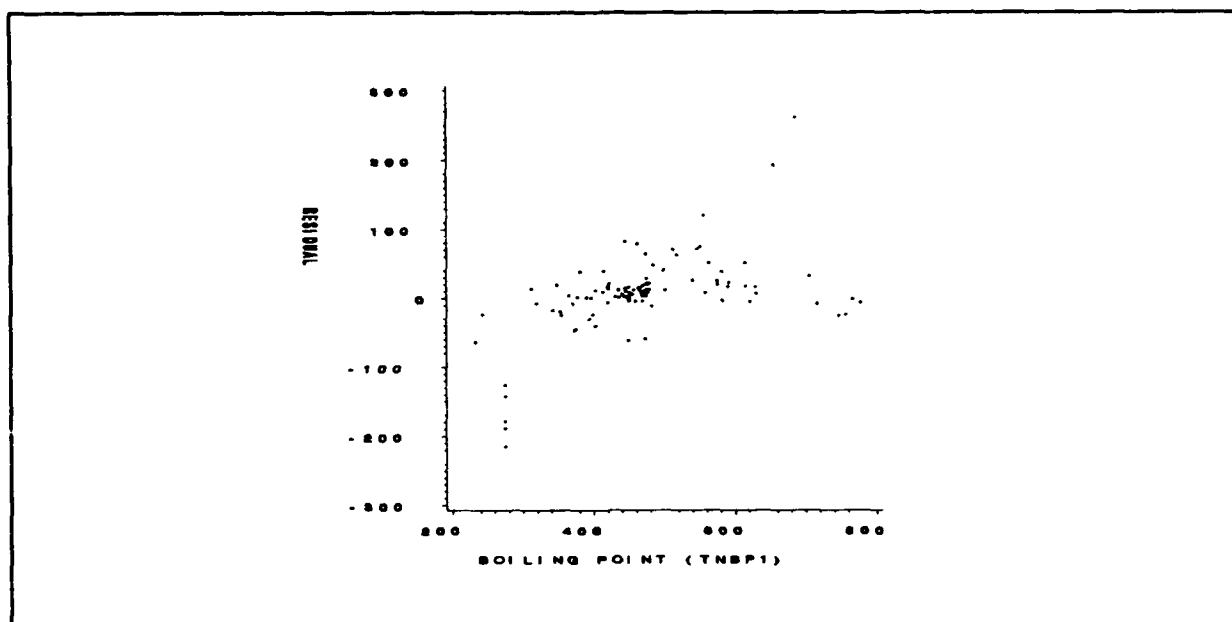


Figure 4.5-11 Experimental normal boiling points versus relative errors for all hydrocarbons (details in text and 1 of 15 points plotted) (Temperatures in K)

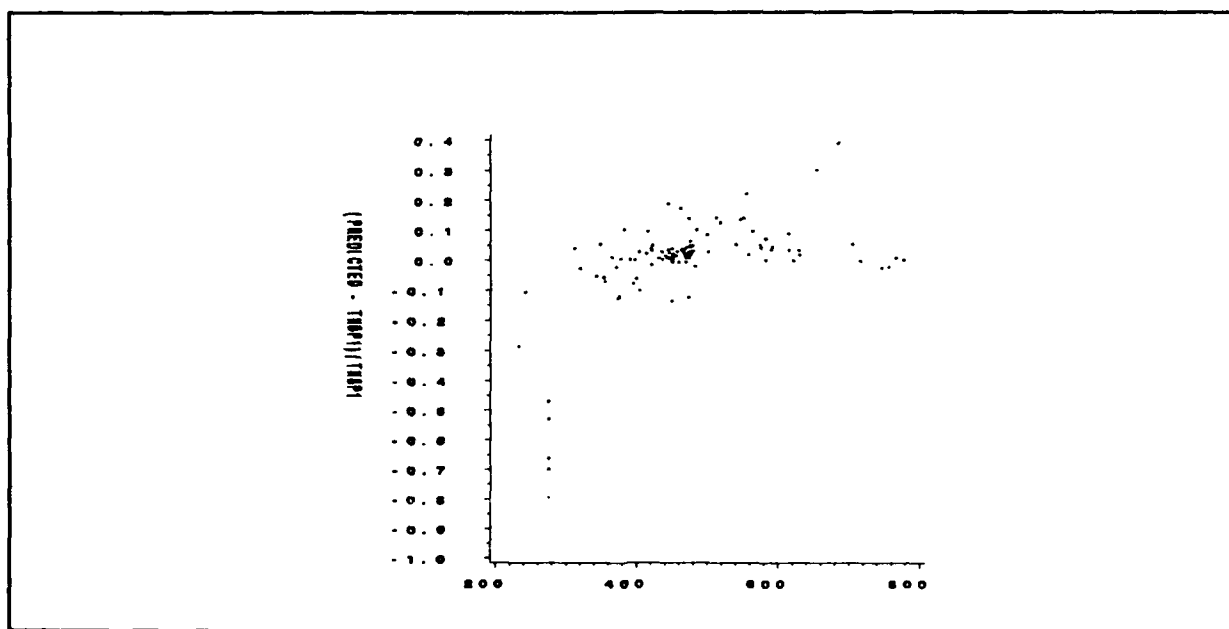


Figure 4.5-12 Experimental normal boiling points versus relative errors for all hydrocarbons (details in text and 1 of 15 points plotted) (Temperatures in K)

parameters used in the model for these plots are $^1\chi_p$, $^6\chi_{ch}$, $^6\chi_p^v$, the total Wiener number, and the number of aliphatic carbons attached to aromatic carbons. These plots clearly show that even looking at all of the hydrocarbons together does not work well. By examining the results from each family, it was found that some of the families of compounds were "well behaved." These 16 families were then grouped together for regression analysis, and they accounted for over 1,200 compounds. The R^2 's for the one through five variable fits were, 0.971, 0.979, 0.980, 0.984, and 0.985. This shows that some combinations of families are possible so that the minimal number of equations to predict a property for any compound is achievable.

The concerns that we have for minimizing the number of equations and the number of variables used in the equations results from the anticipated necessity of having not only good predictive methods, but also methods that are not unreasonably complicated. This is desirable, not only from a conceptual standpoint but also is prudent when consideration is made for the inversion process that is to occur in Phase III of this project.

4.6 Codification of Modeling Program

Objective:

To develop the necessary code to make the AFP system an integrated user friendly system.

Work completed:

The AFP system is based on a series of menus that allows the user to select the various input and output options that are available. The main menu currently allows the user to select from options for:

- Single compound information
- Multiple compound information
- Find compounds with specific properties
- Mixture information

Exit program

At the present, only the first two and the last option are functional. The other options are for future capabilities.

Selecting either of the first two options brings up a menu with the following selections:

- Select compounds
- Property information
- Temperature(s) selection
- Pressure(s) selection
- Choose output units
- Output results
- Return to previous menu

It should be mentioned that for all but the main menu, the last selection is always to return to previous menu (further menu listings will not include this item), which in this case would take the user back to the main menu.

The methods available for compound selection include the input of a molecular structure using the SMILES notation described earlier, input of the name of the compound of interest, or supplying the name of a file that contains a list of SMILES strings, compound names, and/or ASID numbers.

In the case of the SMILES string entry, the SMILES string can be put in exactly, or only a partial SMILES string or strings need be entered. In the latter case, the system will search the entire AFP data base for compounds that contain the group(s) designated. Two examples show how the searching works. First, to find all compounds that contain a benzene ring, the SMILES string that would be input would be c1ccccc1*, the asterisk designating that this is to be a subgroup search as opposed to wanting to choose benzene. The other example is to specify "all" compounds (in the AFP data base) that have both a cyclohexane ring and an ethyl group, the SMILES input would be C1CCCCC1,C-[C&H3]. In this case, the comma designates that the search is to be for all molecules that

contain the specified groups. In either case, after the computer has found all of the compounds that meet the search criteria, the user is allowed to select the compound(s) of interest.

The search by compound name is similar to the SMILES entry in that either an exact match can be sought or wildcards can be utilized. In either case, all of the names available in the AFP data base will be searched for matches. The syntax for the wildcard name searches is that anywhere that an asterisk is located, any number of characters may be substituted. For example, to find all compounds that have dimethylcyclohexane in their name, the search name would be entered as *DIMETHYLCYCLOHEXANE*. This method of searching is not trivial since the matching must be exact where there are no asterisks, including the possibility for spaces and/or hyphens. All searches are done in uppercase so misses due to mismatched cases is not a problem.

Property selection is accomplished by providing the user with a long scrollable list of all of the available properties in alphabetical order. Each property can be selected or deselected by putting the cursor on the property of interest and pressing enter. A property is designated as selected when an asterisk appears on the far right side of the screen across from the property description. The asterisk disappears upon deselection. As many or as few properties as desired can be selected.

Currently, temperature and pressure selection can be done by entering individual values or the number of values, the initial value, and an incremental value. The current limit on the number of temperatures and pressures is 100 for each.

The output of the AFP system defaults to SI units; however, the user can choose to change this to one of many different units that are available for each property. Again, the selection of the units is done by presenting the user with a menu showing the possible output units available. For example, for the density of a liquid, the possible units that could be selected are:

kg/m³,
g/cm³,
oz/in³,
lb/in³,
lb/ft³, or
lb/gal.

The method that was developed to handle the different units was designed for flexibility; therefore, the effort necessary to add new units (up to a maximum of 10 different units, current limit) is minimal.

Currently the results can be output immediately to the computer screen or they can be sent to a file for later viewing or printing. The output first lists the compounds selected and assigns them a number. Following this, the system outputs the property values for all temperature and pressure combinations for each compound for every property selected.

The ability to output the results graphically is currently available as an external routine. This will be implemented in Phase II as it is determined what the appropriate format or formats should be for mixtures and pure compounds to avoid any duplication of effort.

4.7 Model Testing and Verification

SAS⁵², a statistical software package on the VAX was used to test theoretical calculations compared to experimental values in the data base of the single valued properties.

Testing of temperature and pressure dependent properties is difficult for the following reasons and was not done at this time:

1. Errors change with temperature and pressure relative to the compound's critical temperature. The closer the sampling temperature and pressure are to the compound's critical temperature and pressure the greater the error.

2. Errors become smaller as the pressure approaches zero because the gas approaches ideal gas behavior.

Figures 4.7-1 and 4.7-2 show plots of error in molar volume for CO_2 , calculated using two different methods. Figure 4.7-1 was calculated using the Redlich-Kwong equation of state using standard parameters. Figure 4.7-2 was calculated using Redlich-Kwong equation of state with Soave parameters. It would be very difficult to evaluate which method is better based on a single point from either graph.

The single valued properties of critical temperature, critical pressure, critical volume, critical compressibility, boiling point, acentric factor, melting point, and molecular weight were evaluated.

The correct evaluation of temperature and pressure dependent properties would be contour plots, however these calculations are beyond the scope of this project.

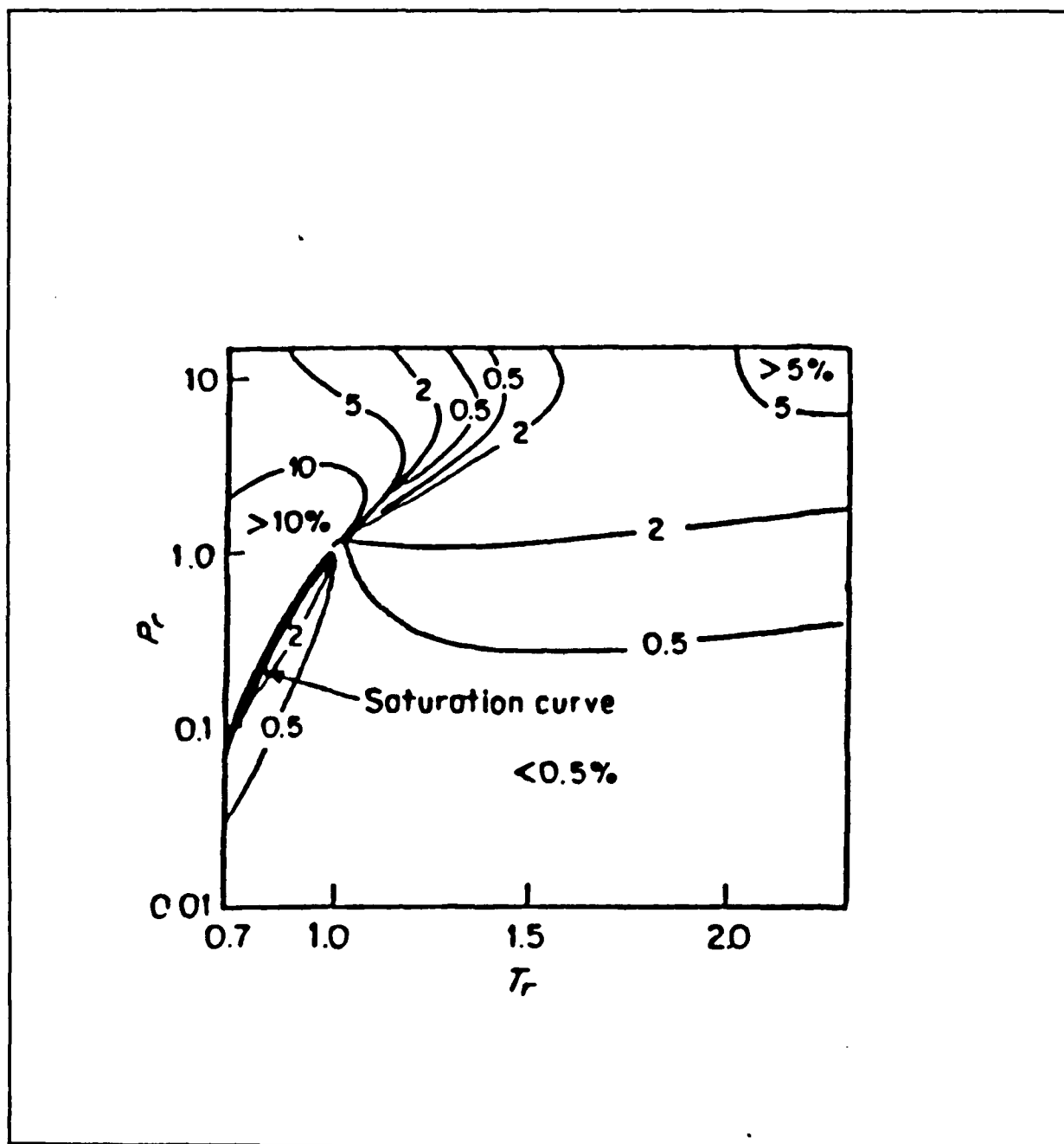


Figure 4.7-1 Percent error in molar volume calculated for CO_2 by using the Redlich-Kwong equation of state with standard parameters¹³

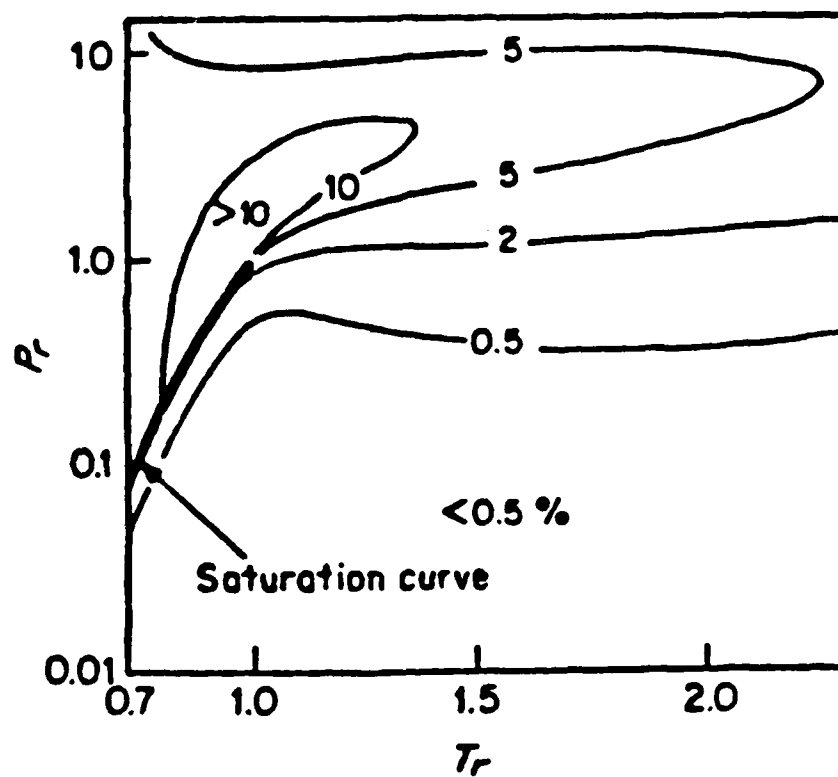


Figure 4.7-2 Percent error in molar volume calculated for CO_2 by using the Redlich-Kwong equation of state with Soave parameters¹³

A SAS program was written to read the RESULTS.DAT file from the USEMTH program. The USEMTH program retrieves data from the data base or calculates the property for a specified method. The user enters the compounds by ASID or by family number. An example of the output file produced by USEMTH is:

METHOD	ASID	FAMILY	TEMP	PRESS	VALUE	ERROR	SUBERROR
TC1	4	1	.	.	0.36982E+03	0.1E-01	0
TC1	5	1	.	.	0.42518E+03	0.1E-01	0
TC1	6	1	.	.	0.46970E+03	0.1E-01	0
TC1	7	1	.	.	0.50743E+03	0.1E-01	0
TC1	201	15	.	.	0.64000E+03	0.5E+01	0
TC1	202	15	.	.	0.65700E+03	0.5E+01	0
TC1	203	15	.	.	0.67200E+03	0.5E+01	0
TC1	204	15	.	.	0.68500E+03	0.5E+01	0
TC2	4	1	.	.	0.36846E+03	0.0E+00	0
TC2	5	1	.	.	0.42381E+03	0.1E-01	0
TC2	6	1	.	.	0.46943E+03	0.1E-01	0
TC2	7	1	.	.	0.50771E+03	0.1E-01	0
TC2	201	15	.	.	0.64317E+03	0.1E-01	0
TC2	202	15	.	.	0.66074E+03	0.1E-01	0
TC2	203	15	.	.	0.67577E+03	0.1E-01	0
TC2	204	15	.	.	0.68973E+03	0.1E-01	0

Files were generated requesting data for all 4,464 compounds in the data base for TC1, TC2, TC3, TC4, TC5, TC6, PC1, PC2, PC3, PC4, VC1, VC2, VC3, VC4, ZC1, ZC2, ZC3, TNBP1, TNBP2, ACENF1, ACENF2, MW1, MW2, TMPSP1, and TMPSP2. Each of these files contained 4,464 lines. Files with the experimental values files (TC1, PC1, VC1, ZC1, TNBP1, ACENF1, MW1, and TMPSP1) were then merged with the files generated by predictive methods, i.e., TC1 with TC2, TC1 with TC3, TC1 with TC4, etc. making each file 8,928 lines long. Although the files with the experimental values contained 4,464 lines, some of the experimental values were missing and the SUBERROR shown above would indicate that no experimental value is available for that ASID.

The first step for comparing the experimental and theoretical data using SAS was to read in the values for all of the compounds and remove any invalid data (values where the subroutine error was not equal to zero). The initial sort done by SAS cut the length of the file in half by sorting the original file by ASID, resulting in a new file containing columns like this:

ASID	FAMILY	TC2	TC1
4	1	0.36846E+03	0.36952E+03
5	1	0.42381E+03	0.42518E+03
6	1	0.46943E+03	0.46970E+03
7	1	0.50771E+03	0.50743E+03
201	15	0.64317E+03	0.64000E+03
202	15	0.66074E+03	0.65700E+03
203	15	0.67577E+03	0.67200E+03
204	15	0.68973E+03	0.68500E+03

The difference between the experimental and theoretical value was calculated in SAS, thus creating a new variable, DIFF. The fractional error (FRAC), is calculated by dividing DIFF by the experimental value. The root mean square deviation (RMSD) was calculated and used as an error bar indicator. The equation defining RMSD is given by:

$$RMSD = \sqrt{\sigma^2 \left(\frac{N-1}{N} \right) + (\bar{X}_{ERR})^2}$$

where σ^2 is the variance between the predicted and data base values, and $\bar{X}_{ERR}$ is the mean of the difference between the predicted and data base values. This latter term would be essentially zero for a normal distribution of errors. Therefore, since the differences between the predicted and the data base values does not follow a normal distribution, the use of RMSD was recommended in place of root mean error*. The error bar was made by adding the RMSD to (upper point HIGH) and subtracting (lower point, LOW) the RMSD from the theoretical value

*The RMSD was recommended by J. D. Nelligan, an applied mathematician at Allied-Signal.

(middle point, e.g. TC2). These values were then put into a file such as:

ASID	FAMILY	TC2	TC1	DIFF	FRAC	RMSD	HIGH	LOW
4	1	368.46	369.82	-1.36	-.0036	0.984	369.44	367.47
5	1	423.81	425.18	-1.37	-.0032	0.984	424.79	422.82
6	1	469.43	469.70	-0.27	-.0005	0.984	470.79	468.44
7	1	507.71	507.43	0.28	0.0005	0.984	508.69	506.72
201	15	643.17	640.00	3.17	0.0049	3.893	647.06	639.27
202	15	660.74	657.00	3.75	0.0056	3.893	664.63	656.84
203	15	675.77	672.00	3.77	0.0056	3.893	679.66	671.87
204	15	689.73	685.00	4.73	0.0069	3.893	693.62	685.83

These three points (TC2, HIGH, LOW), were plotted against the experimental value (e.g. TC1) so as to give a visual representation of the accuracy of the predictive ability of the method.

Figures 4.7-3, 4.7-4, and 4.7-5 illustrate the three types of plots made for each comparison. Every point represents one compound, and four to six families of compounds were plotted on the same graph. All three figures are plots of families 1, 2, 3, and 4 (n-paraffins, methylalkanes, cycloalkanes, and other alkanes, respectively). The method for calculating the critical temperatures was method TC6 (Ambrose method, a group additive method parameterized using boiling point; recommended by AIChE and API). Figure 4.7-3 was plotted with the difference between TC6 value and TC1 value (data base, i.e., experimental) on the Y axis and TC1 on the X axis. In Figure 4.7-3, family 1, the n-paraffins, has a very small, several degrees Kelvin, deviation in temperature for compounds with critical temperatures less than 700K. For compounds with critical temperatures above 700K, the difference is approximately 5 degrees. Family 4 on the other hand, has compounds whose critical temperatures gather around 600-650K for the experimental values; however, the theoretical predictions vary over a much wider range than did those for family 1. Figure 4.7-4, is the fractional error plotted against experimental TC1 values. This plot readily shows that although the spread on family 4 appears rather large, the fractional error as a percent is still quite small, on the order of 4 percent.

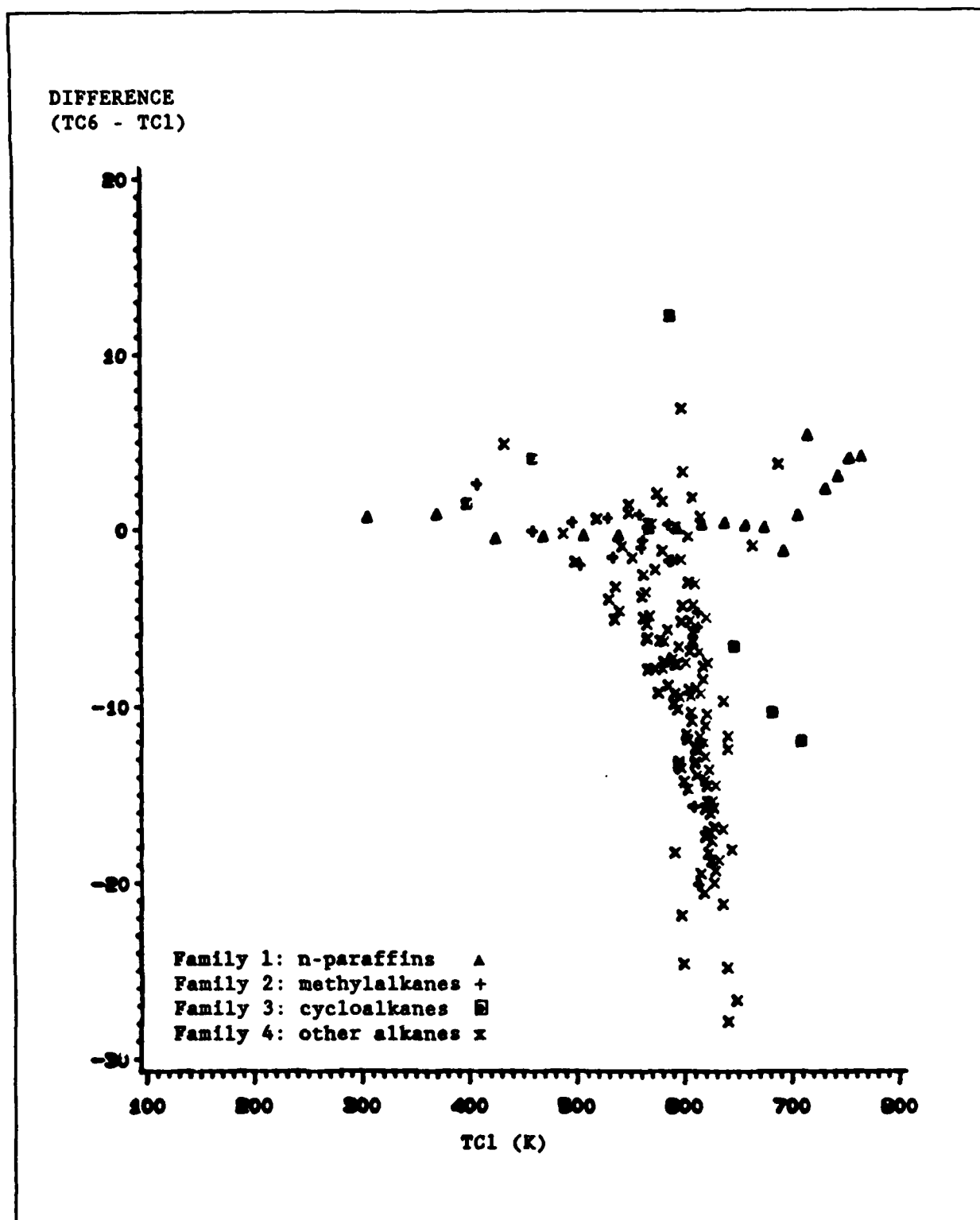


Figure 4.7-3 Data base (TC1) and predicted (TC6) critical temperatures differences versus experimental critical temperatures

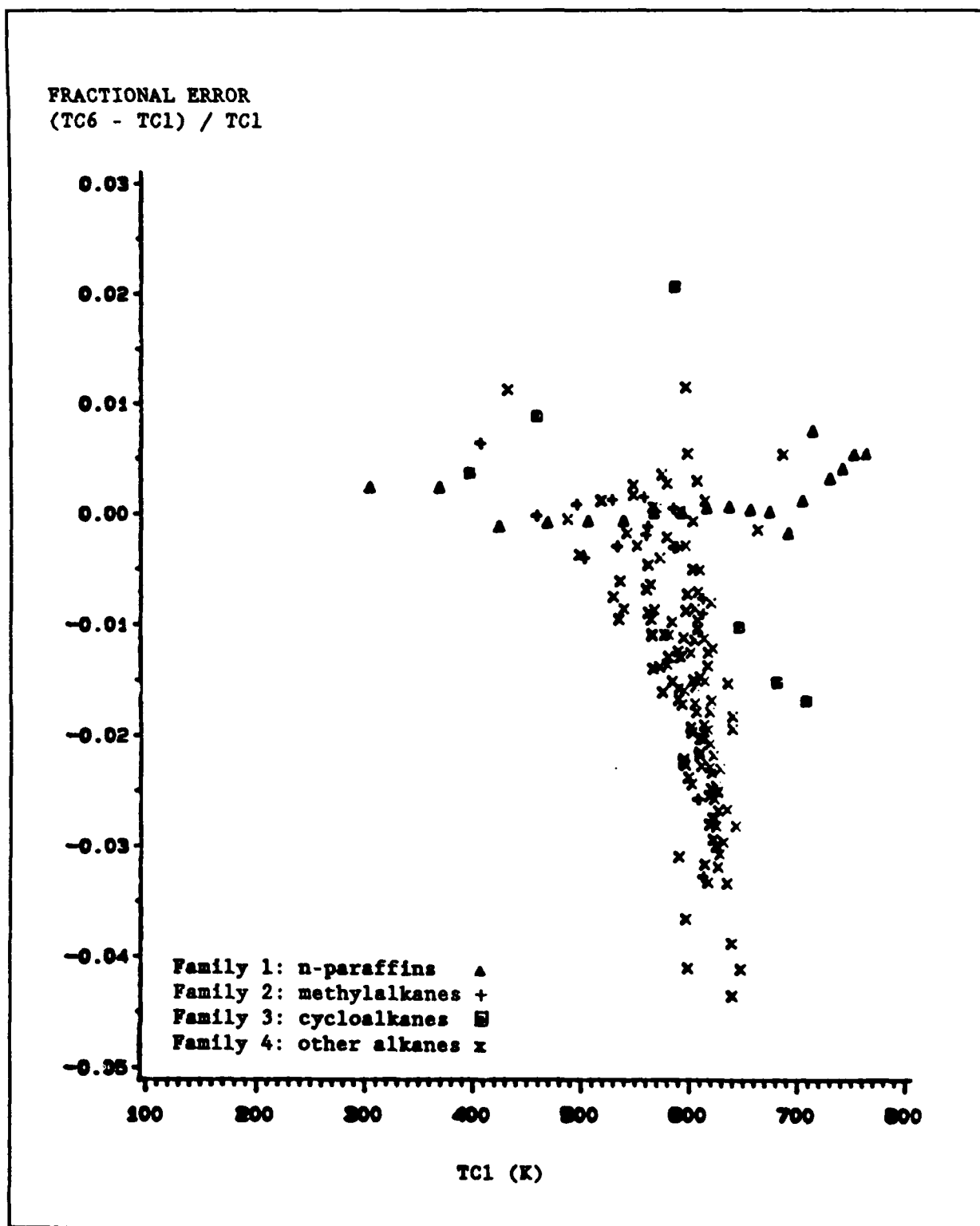


Figure 4.7-4 Fractional difference between data base and predicted critical temperatures versus data base critical temperatures

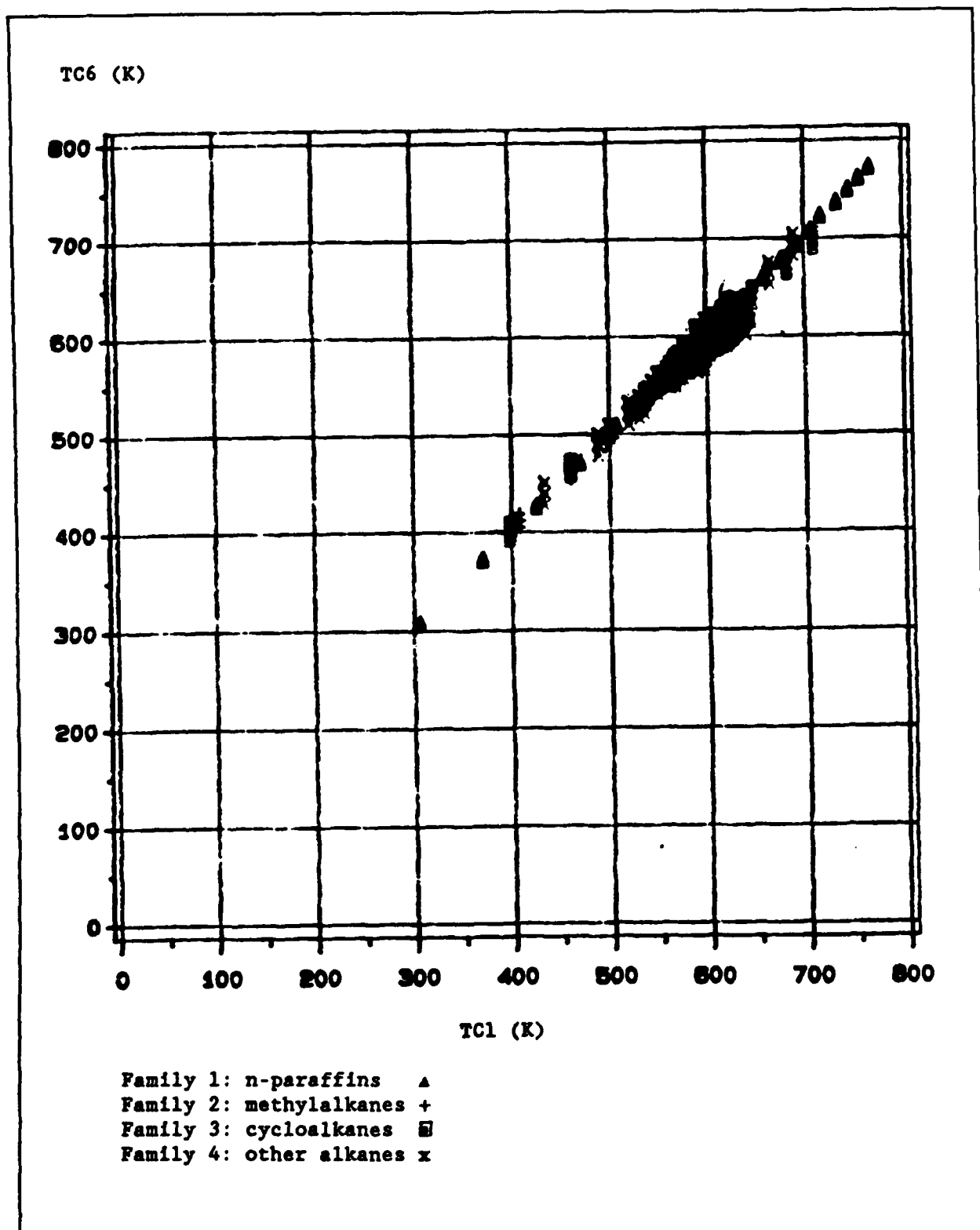


Figure 4.7-5 Data base (TC1) versus predicted (TC6) critical temperatures

Figure 4.7-5 is a plot of TC6 vs. TC1 with the RMSD error bars. These points appear in a straight line, and the error bars are so small as to not be visible. Therefore, we concluded from these plots that for these four families TC6 is a relatively accurate method of calculating critical temperature, to within a few percent or less.

Figure 4.7-6 is a plot of melting point at 1 atmosphere predicted (TMPSP2) plotted against experiment (TMPSP1). Families 1-4 were also used in this plot. The error bars are much larger than observed in the TC6 plot and the correlation is also nonlinear. Note also that the error bars for family 1 appear to decrease with increased experimental temperature values.

Figure 4.7-7 is a similar plot to Figure 4.7-6 where predicted boiling points are plotted against experimental values. Families 1-3 were used for this plot. Note again that the error bars decrease with temperature of the experimental value.

SAS was also used to calculate the mean fractional errors by family for the different methods of calculating critical temperature (TC2, TC3, TC4, and TC5). The results are shown in a bar graph, Figure 4.7-8. The results show that for families 1 and 2, TC5 appears to be the best method to calculate critical temperature; mean fractional errors of less than 1 percent. TC2 on the other hand is the best of these methods for family 3. Figure 4.7-9 shows a similar bar graph as the previous figure except for families 21-30. Families 21-23 have extremely mean high fractional errors on the order of 25 percent and TC3 would be a better method of calculating critical temperature for families 21 and 22, but not for 23 because there are no values present. This analysis is important because it gives us rules for which method is best for which families of compounds. This is being incorporated into a priority scheme, thereby making the program an expert system.

Figure 4.7-10 shows the percent of available compounds for the various methods which have fractional errors of less than 5 percent. A few of the

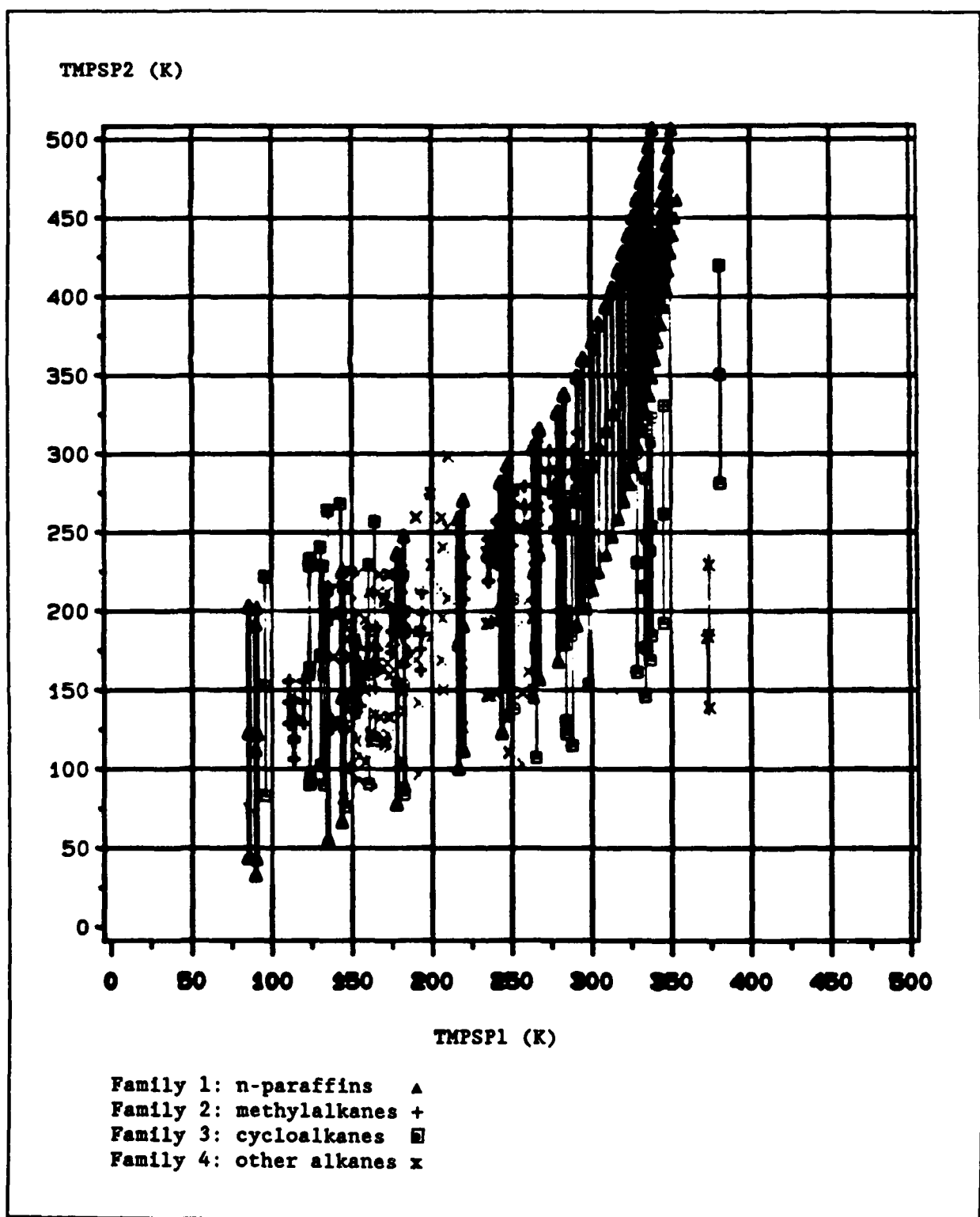
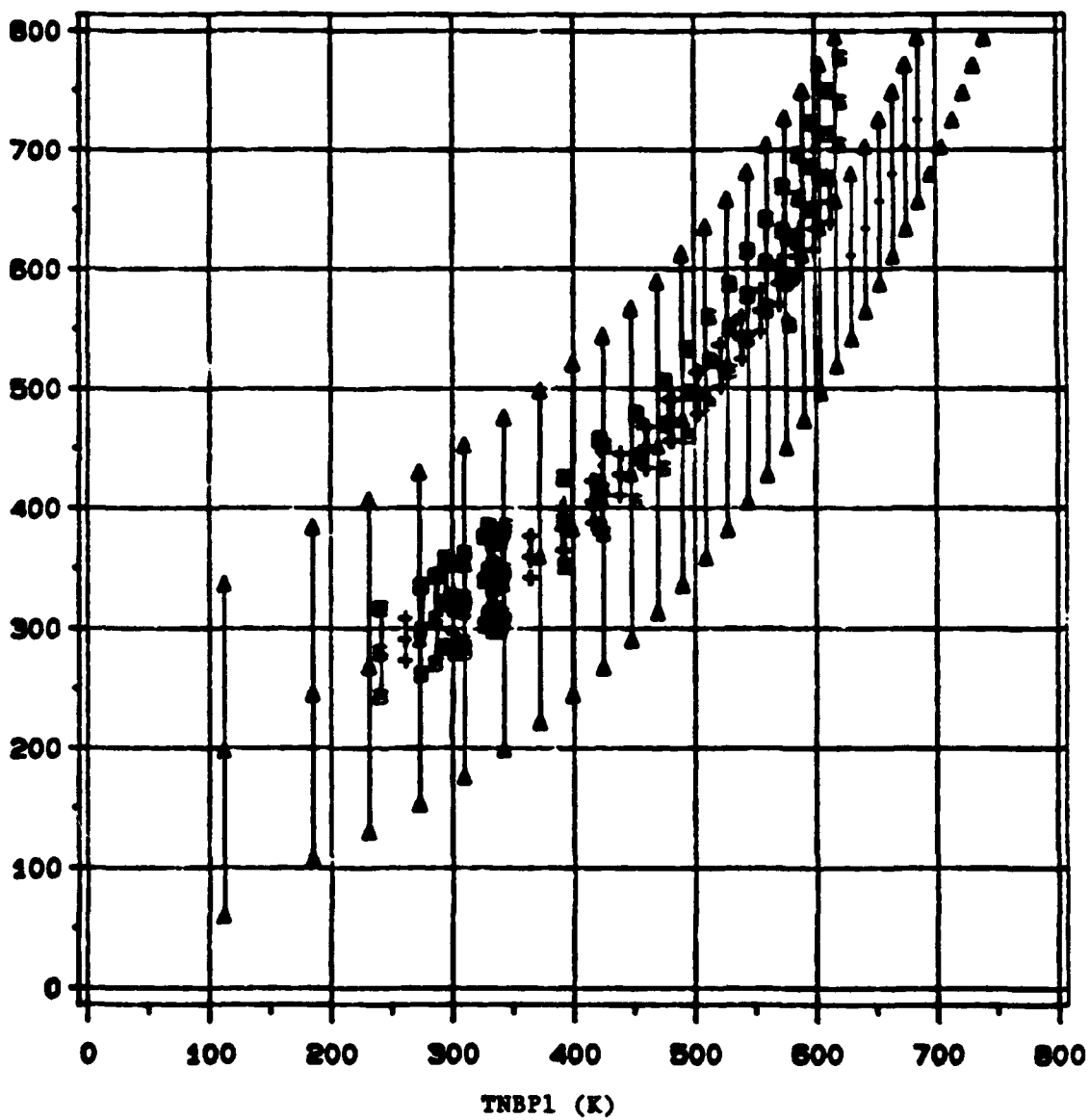


Figure 4.7-6 Data base (TMPSP1) versus predicted (TMPSP2) melting temperatures at 1 atmosphere

TNBP2 (K)



Family 1: n-paraffins Δ
Family 2: methylalkanes +
Family 3: cycloalkanes $\square$

Figure 4.7-7 Data base (TNBP1) versus predicted (TNBP2) normal boiling temperatures

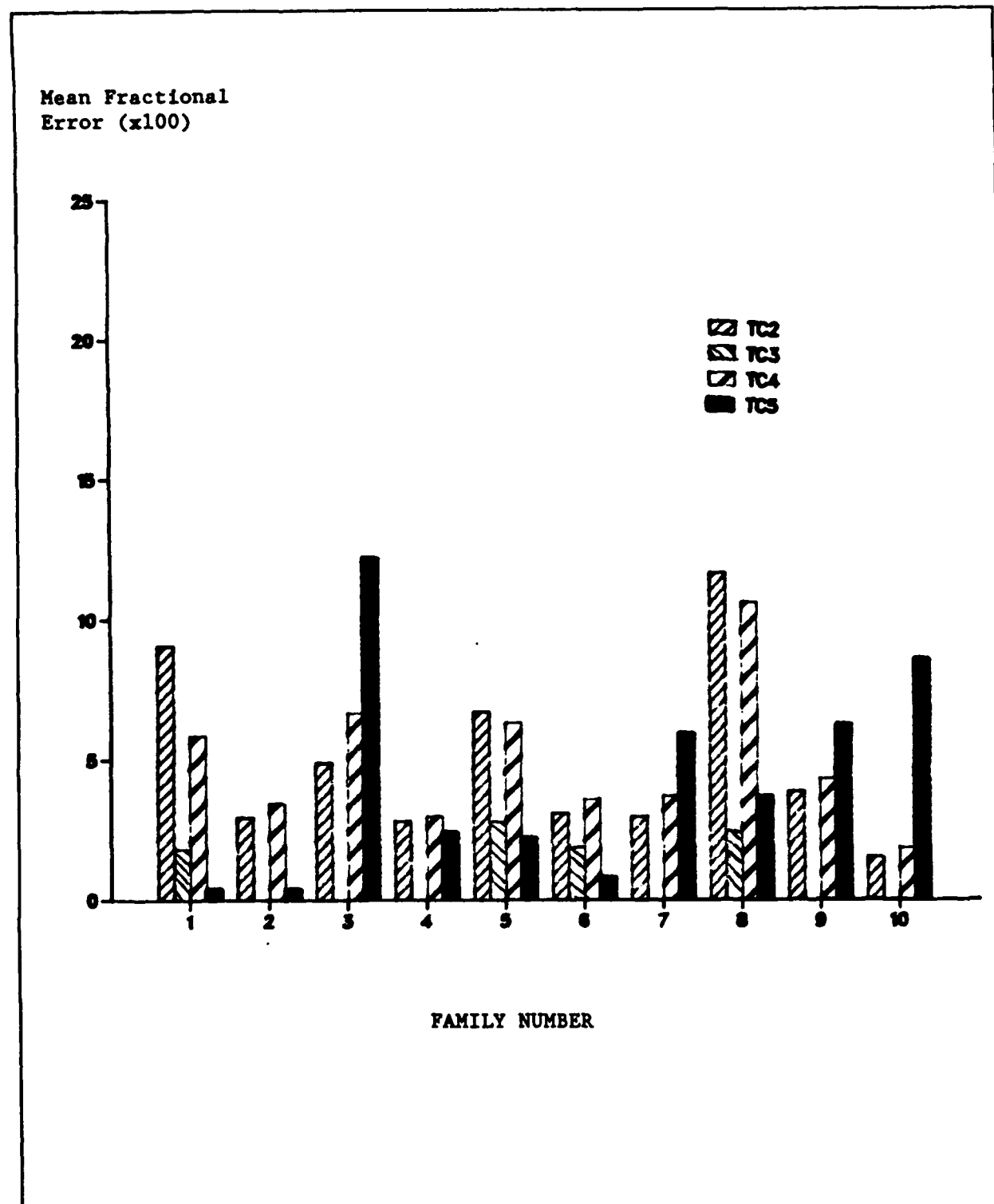


Figure 4.7-8 Critical temperature mean fractional error by family (families 1 - 10)

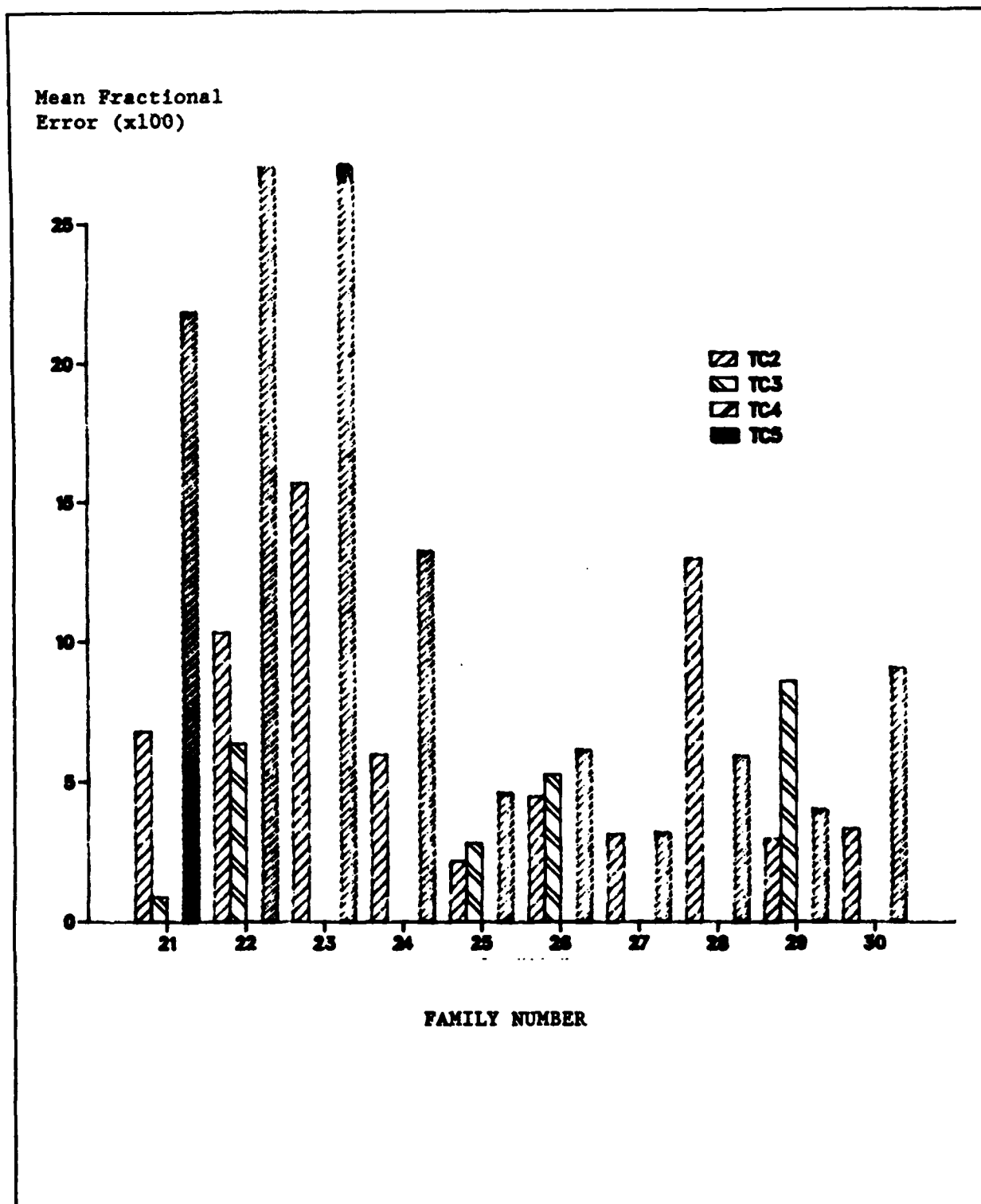


Figure 4.7-9 Critical temperature mean fractional error by family (families 21 - 30)

methods, such as melting point at 1 atmosphere, have many values with fractional errors higher than 5 percent, but several have at least 90 percent of the values with less than 5 percent error.

Figure 4.7-11 illustrates the numbers of ASID's for which experimental data exist in the data base and which can be calculated using the various methods. Many of the methods will calculate values for over 4,000 compounds.

Table 4.7-1 lists the method abbreviation in the first column. The number following the letters indicates if it is an experimental value (1) or which method was used to make the calculation (2, 3, ...). TC is critical temperature, PC critical pressure, VC critical volume, ZC critical compressibility, TNBP normal boiling point, ACENF acentric factor, and TMPSP melting point at 1 atmosphere. The second column lists the number of compounds in the data base or for how many the compounds the predictive method could handle. The third column shows how many values were available where both theory and experiment had values. The percent of the total values represented by the matches (column 3/column 2 X 100) is listed in the fourth column. The last columns break down the number of compounds by their fractional errors.

Summary of Results of Methods Testing

It was observed by examining the mean fractional errors by family, that some methods are better than others for specific families. Therefore, it is difficult to say which method is best for all compounds. Many of the calculations only apply to certain families of compounds.

The critical temperature data show that for methods TC3 and TC4 only 314 and 367 compounds were covered respectively in the matching routine. While methods TC2, TC5 and TC6 had "matches" of more than 1,000 compounds each. Method TC5 has the highest number of compounds with fractional error greater than 20 percent.

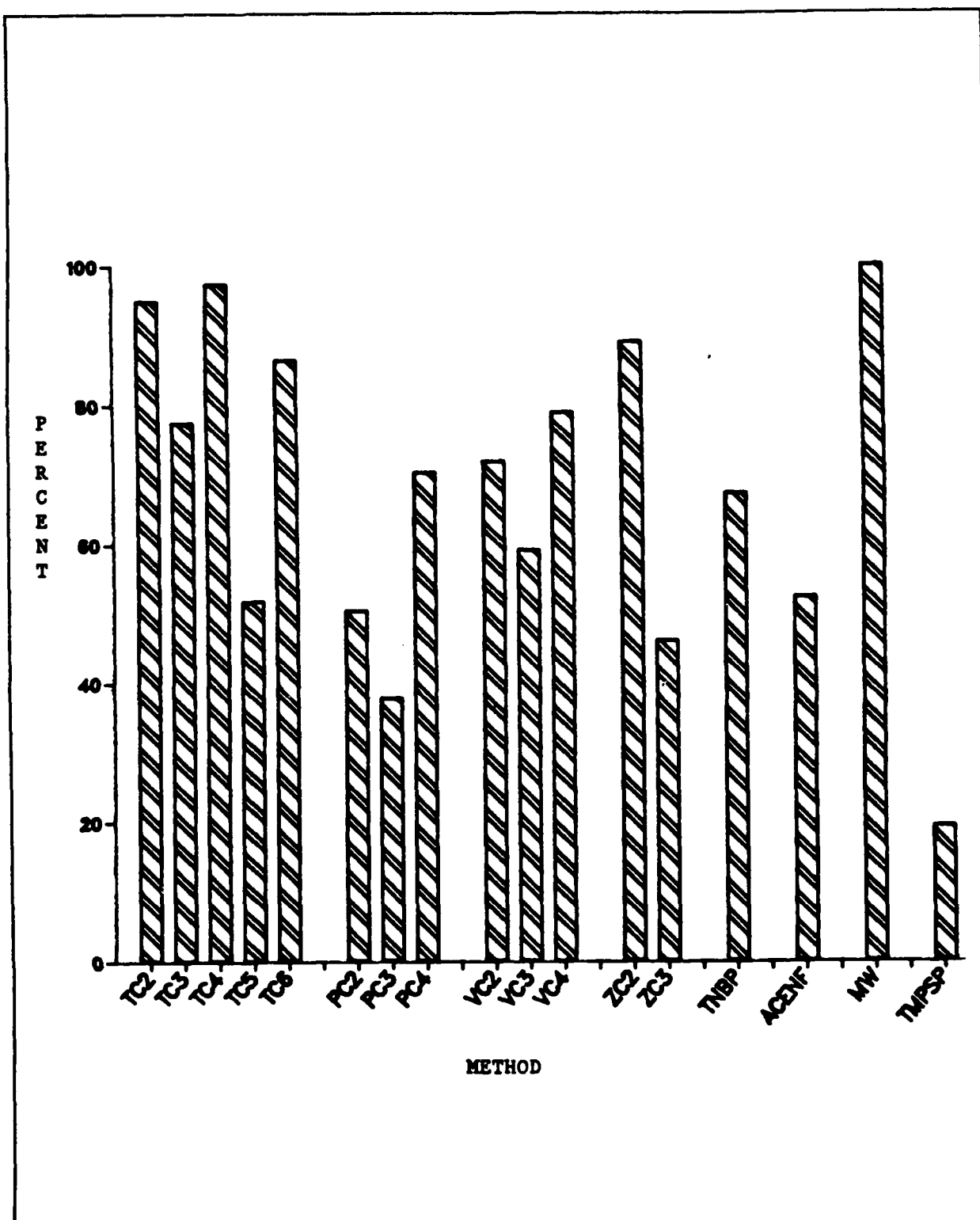


Figure 4.7-10 Percent of predicted values having fractional errors less than 5 percent by property and method

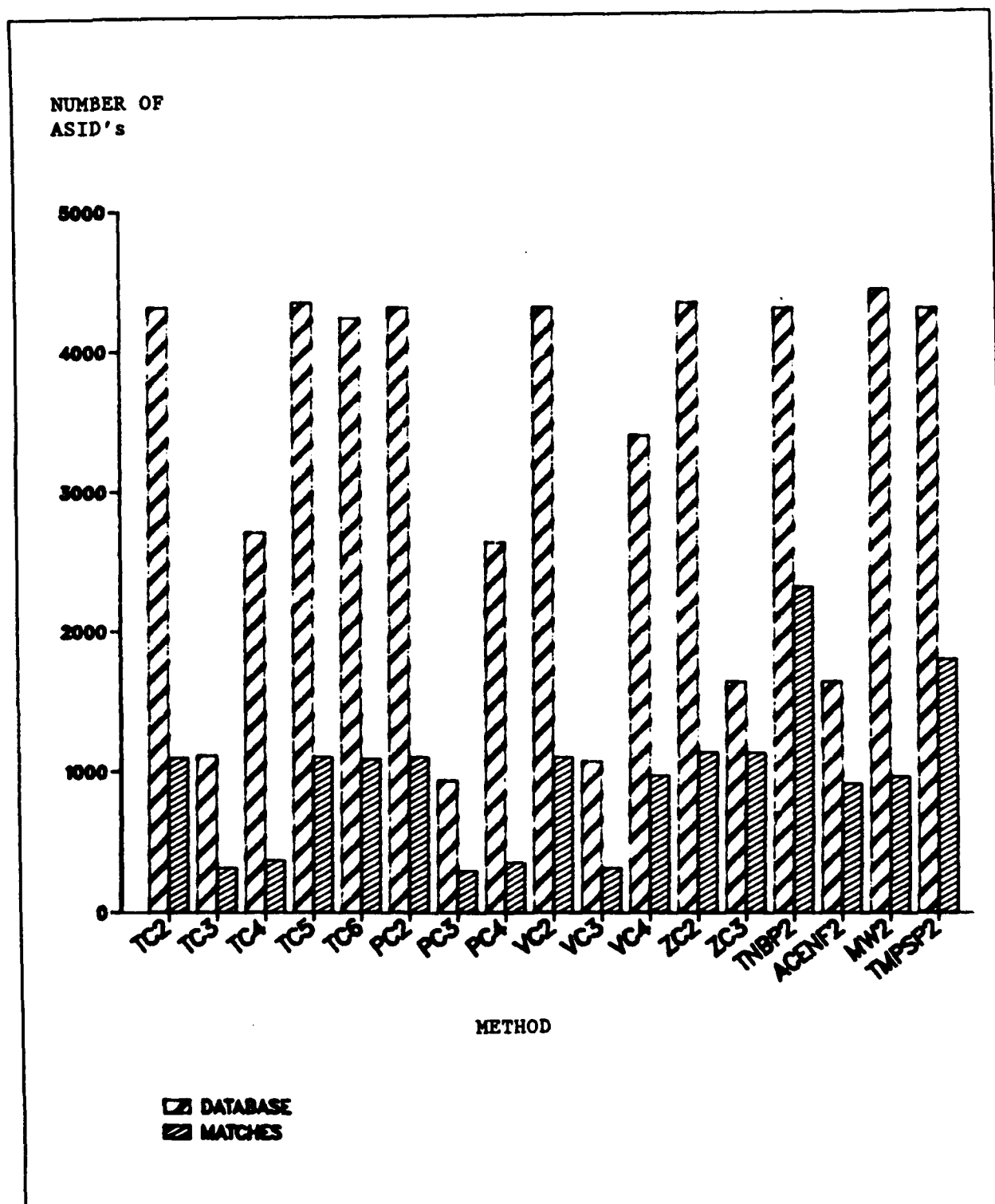


Figure 4.7-11 Comparison of the number of data base property values and the number of cases where both data base and predicted values exist

Table 4.7-1
Distribution of the Number of Predictions at Various Fractional Errors

Experiment vs. Theory														
Method	# of values in the database (or calculated) are present	# ASIDs where values for both theory & expt	# values Fractional Error (FE) FE<5%			# values Fractional Error (FE) 5%<FE<10%			# values Fractional Error (FE) 10%<FE<20%			# values Fractional Error (FE) FE>20%		
			#	z	N/A	#	z	N/A	#	z	N/A	#	z	N/A
TC1	1151	N/A	1095	25.36	N/A	1038	94.79	35	3.20	9	0.82	13	1.19	N/A
TC2	4318		314	28.26		243	77.39	41	13.06	18	5.73	12	3.82	
TC3	1111		367	13.54		357	97.28	8	2.18	0	0	2	0.54	
TC4	2710		1102	25.31		569	51.63	181	16.42	132	11.98	220	19.96	
TC5	4354		1088	25.64		939	86.31	95	8.738	50	4.60	4	0.37	
TC6	4244													
PC1	1152	N/A	1100	25.47	N/A	553	50.27	252	22.91	183	16.64	112	10.18	N/A
PC2	4318		292	31.30		110	37.67	58	19.86	31	10.62	93	31.85	
PC3	933		352	13.34		247	70.17	72	20.45	22	6.25	11	3.13	
PC4	2639													
VC1	1153	N/A	1101	25.50	N/A	791	71.84	171	15.53	73	6.63	66	6.00	N/A
VC2	4318		312	29.16		184	58.97	74	23.72	31	9.94	23	7.37	
VC3	1070		970	28.50		765	78.87	131	13.51	64	6.60	10	1.03	
VC4	3404													
ZC1	1171	N/A	1137	26.12	N/A	1011	88.92	66	5.80	35	3.08	25	2.20	N/A
ZC2	4353		1133	69.09		521	45.98	200	17.65	220	19.42	192	16.95	
ZC3	1640													
TNBP1	2409	N/A	2320	53.73	N/A	1558	67.16	256	11.03	164	7.07	342	14.74	N/A
TNBP2	4318													
ACENF1	955	N/A	915	55.72	N/A	478	52.24	243	26.55	83	9.07	111	12.13	N/A
ACENF2	1542													
MW1	964	N/A	964	21.65	N/A	963	99.90	0	0	0	0	1	0.10	N/A
MW2	4453													
TMFSP1	1893	N/A	1802	41.73	N/A	346	19.20	309	17.15	488	27.08	659	36.57	N/A
TMFSP2	4318													

Critical pressure method PC2 has the highest number of compounds in the "matches" category, 1,100. Compared to PC3, PC2 has fewer compounds with fractional errors of greater than 20 percent. Method PC4 has 71 percent of its compounds with less than a 5 percent fractional error, but there are only 352 compounds in the match category available for comparisons.

Critical volume method VC2 has the largest number of compounds which were compared, over 1,100. Method VC4 is the method with the highest percent of compounds having the smallest fractional error.

Critical compressibility methods ZC2 and ZC3 have similar numbers of compounds in the match category, approximately 1,100 each. Method ZC2 had almost 90 percent of the compounds with fractional errors less than 5 percent, while ZC3 had less than 50 percent samples in the less than 5 percent fractional error category.

All of these results will be used in the design and implementation of the priority system for determining which method should be used to predict a property for any given molecular structure.

5.0 SOFTWARE DESIGN

5.1 Approach

The Advanced Fuel Properties System is based on a user friendly menu driven concept. Additionally, the software has the capability to have available on-line help to explain the operation of each menu. The system was designed for ease of use, expandability, modification, and incorporation into the Phase II and III software. The concept was to construct modular software routines that would do a minimal number of tasks, with the target being one task for each routine. This approach was successful and is one of the things that makes the AFP software system easy to expand, modify, and incorporate into the latter phases of the project.

Since it was evident in the early stages of the project that several methods would be necessary to obtain a desired property, a method that would

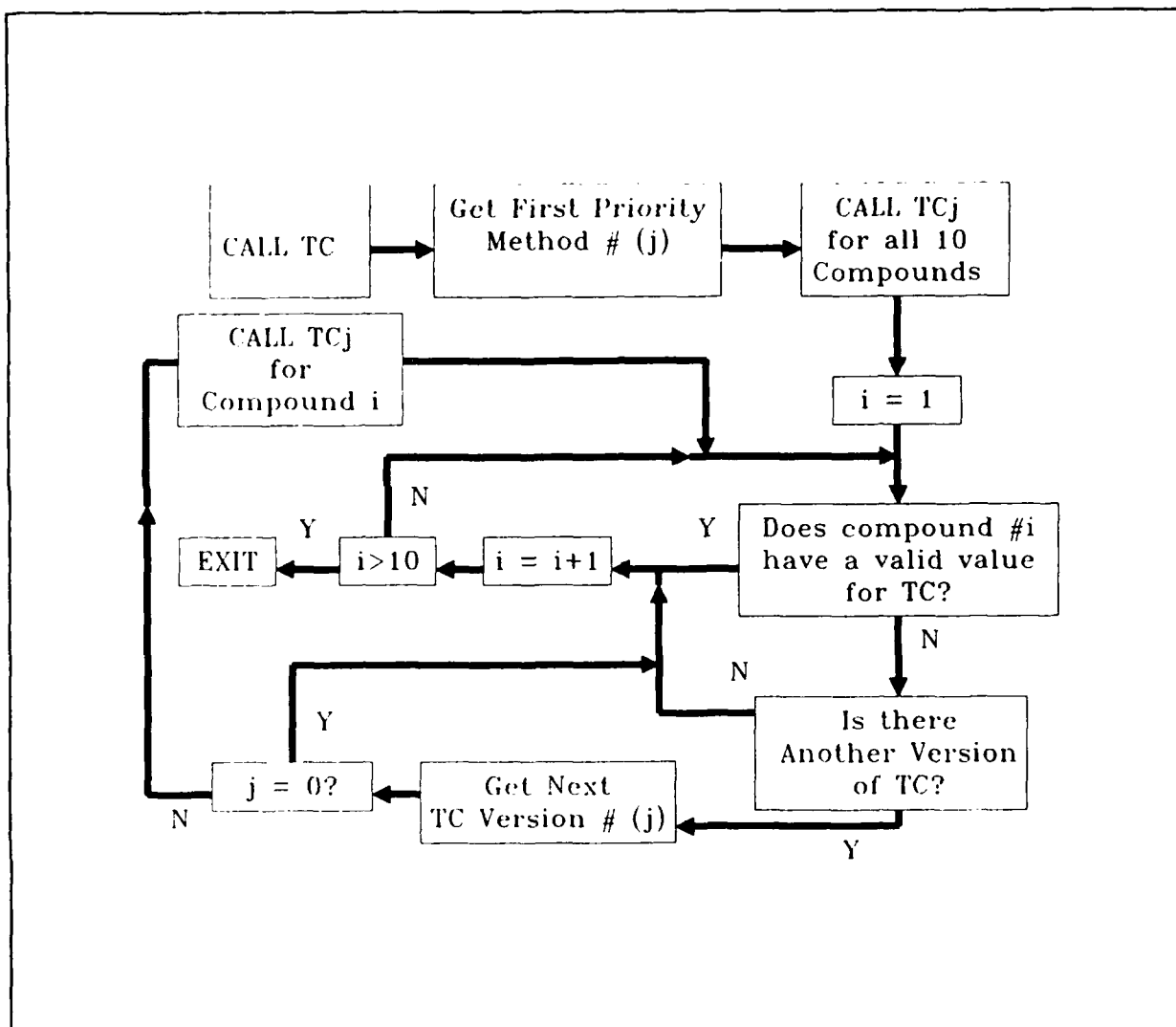


Figure 5.1-1 Diagram of priority scheme for an example of obtaining the critical temperature (method name TC) for 10 compounds

afford the necessary flexibility and expandability was needed. This concept developed into the priority scheme approach. The basis for the priority scheme is a driver routine (a FORTRAN subroutine) for each property. This routine will call each of the different methods available to obtain a good property value for each compound. The current system is diagrammed in Figure 5.1-1, using as an example the request to obtain the critical temperature (method name TC) for 10 compounds. This is currently accomplished by having a file on disk (named DEFAULTS.PRI) that is accessed and contains the order in which each different

method should be called. This order depends on the accuracy of the property values provided by each method. An example of part of the DEFAULTS.PRI file is shown in Figure 5.1-2.

6	TC	7	F	F	6	1	2	6	5	4	3
7	PC	8	F	F	5	1	2	5	4	3	
10	ACENF	11	F	F	4	1	2	3	4		
13	TMPSP	18	F	F	2	1	2				
33	FUGAC	56	T	T	5	3	6	5	2	4	
34	GF25I	57	F	F	3	1	2	3			
35	CMPR	58	T	T	6	3	7	6	2	5	4
40	NUVAP	64	T	F	3	4	3	2			

Figure 5.1-2 Selections from the DEFAULTS.PRI file that is used by the priority scheme in the AFP system (Details in the text)

The priority scheme also had to be flexible so that if the user wanted to use specific methods, the default priority scheme could be overridden. This is accomplished by allowing the user to select the priority order for any property. After modifying the priority scheme, the user can save the customized priority scheme in a user file. This saved priority scheme file can later be reloaded into the AFP system (overriding the default priority scheme) so that a user can use the same customized priority scheme at different times with minimal effort.

The priority scheme was also designed with sufficient flexibility such that an expert system could be added. The expert system would be able to modify the priority scheme during program operation. Although this is not implemented at this time, the extension would not be difficult because of the design of the current AFP priority scheme.

To keep control of the hundreds of software routines that were necessary to accomplish this modular concept, a systematic subroutine naming convention was implemented. Two naming conventions were used in the project. The first was used to name routines that generated values for properties and the second was for utility routines. The convention used for the former was to use up to five characters that described the property. The first character(s) relating to the type of property (T for temperature, H for enthalpy, S for entropy, NU for

viscosity, etc.), and the remainder of the characters to specify the exact property. For example, TNBP would be the normal boiling point; T for temperature and NBP for normal boiling point and HF25I would be enthalpy (H) of formation (F) at 298.15K (25 for 25C) for an ideal gas (I). This naming scheme is how the priority routines are named. The names of the various routines that obtain the property values via different methods are named the same except for a trailing character. Specifically, a 1 is added to any routine that is used to look up a value from the data base, 2 through 9 and A through Y are available for any alternate methods by which the property can be obtained. Trailing zeros and Z's are reserved for special usage.

The naming convention for utility routines is much simpler in that each the names of the routines are chosen to be descriptive of its function. This is adhered to as much is allowable within the FORTRAN77 standard of six character names.

In addition to a systematic routine naming convention, every utility routine and base property determination routine is assigned a unique method number. The base method number is modified for the various property determination routines by adding 10,000 times the value of the trailing to the base method number. For example, the method number for CPID (heat capacity at constant pressure for an ideal gas) is 81; therefore, the method number for CPID3 would be $3 \times 10000 + 81$ or 30081. This number is used in a number of ways, but the most important usage is in the reporting of subroutine errors. A subroutine error is when the subroutine is requested to do an illegal function, it cannot obtain a value for what has been requested, or any other type of error that may occur. As the result of any error condition, the routine will pass back an error code. This error code has embedded within its method number and a code for the error condition so that the calling routine can handle the condition properly. Additionally, if this error code is output for a requested value, one can determine what subroutine caused the error condition. The encoding scheme is to take the full method number multiplied by 1,000 and add the error code to obtain the resulting full error code. In addition, if the error condition is actually only a warning, the error code will be negated.

The final aspect of the approach taken was that the main program or its utility routines (later references to the main program will imply reference to the main program's utility routines also) should be the only routines that interact with the user. The main program should also control selection of properties, compounds, units, and options.

5.2 Documentation

Documentation of the various routines is done in two ways. The first is by having the programmer complete a "programmer's reference sheet," an example of which is shown in Figures 5.2-1 and 5.2-2. The purpose of this form is to:

1. describe the purpose of the routine
2. provide enough information that another programmer will know how to use the routine
3. describe the input and output variable names and define their type
4. provide a reference as to where the method originated (if applicable)
5. specify how the routine is called
6. list any routines that are called by the routine
7. specify the programmer
8. provide space for any other comments the programmer feels should be specified

The other method of documentation is the in-line documentation in each routine. At a minimum, this should include the purpose of the routine, the method number, and the programmer's name.

Subroutine Name: _____

Purpose: _____

<u>Input</u>	<u>Description</u>	<u>Variable Name</u>	<u>Variable Type</u>
1.	_____	_____	_____
2.	_____	_____	_____
3.	_____	_____	_____
4.	_____	_____	_____
5.	_____	_____	_____
6.	_____	_____	_____
7.	_____	_____	_____
8.	_____	_____	_____
9.	_____	_____	_____
10.	_____	_____	_____

<u>Output</u>	<u>Description</u>	<u>Variable Name</u>	<u>Variable Type</u>
1.	_____	_____	_____
2.	_____	_____	_____
3.	_____	_____	_____
4.	_____	_____	_____

<u>Regression Coefficients</u>	<u>Variable Name</u>	<u>Variable Type</u>
1.	_____	_____
2.	_____	_____
3.	_____	_____
4.	_____	_____
5.	_____	_____
6.	_____	_____
7.	_____	_____
8.	_____	_____
9.	_____	_____
10.	_____	_____

Reference

Calling Method

Figure 5.2-1 Programmer's reference sheet, front page

This image shows a single sheet of white paper with horizontal blue or grey ruling lines, typical of notebook paper. The lines are evenly spaced and run across the width of the page. There is no handwriting or other markings on the paper.

Figure 5.2-2 Programmer's reference sheet, back page

5.3 FORTRAN77 Standard Compliance

The software has been written in general to the FORTRAN77 standard. Those places where this was not done are listed below, along with a brief explanation of the reasons and consequences for the action.

1. The use of INCLUDE statements is probably the largest deviation from the FORTRAN77 standard. This was done as a method for efficient code generation. The INCLUDE statement allows common code to be kept in a file and then included into the individual routines at compile time. This was done to keep items such as array size declarations in one place so that if they need to be changed during program development, the change need only be done to one file before recompilation with the changes. Also this deviation is not difficult to correct before the delivery of the final version.
2. The use of nonstandard subroutine and variable names was done only where it was necessary to accommodate the requirements of commercial software packages.
3. The use of VAX extensions to FORTRAN77 standards was also only done when required by commercial software packages being used by the AFP system.
4. The final nonconformity to FORTRAN77 was in the use of the READONLY parameter in OPEN statements. This was necessary because some of the data files that the AFP system needs are stored in an area that is read only to most of the users, and even though the files are only going to be read, the VAX requires that the files be opened with the READONLY qualifier.

5.4 Error Handling

One of the important aspects of predictive methods that is frequently bypassed is the reporting of estimated errors for the predicted values. The initial method that was implemented to handle the generation of errors was to report the larger of the method error or a propagated error. The former would

be determined by the error reported in the literature or as determined by checking the accuracy of the method by comparing the predicted values with the experimental data in the AFP data base. The other method would be to properly propagate the errors from the parameters or properties used in the prediction. However, this was found to severely overestimate the errors in some cases because of the nested nature of the AFP system. What is partially implemented now, and will be the method of choice will be to numerically determine the errors by modifying the basic parameters used by the method of interest to determine how sensitive the current predictive method is to each parameter.

An additional benefit to doing the error determination by this method, rather than propagating errors is that it will also provide the capability to do sensitivity analyses. The results of the sensitivity analyses provides information as to where further effort is necessary due to a high sensitivity in the accuracy of certain parameters.

5.5 Graphical Input/Output

Currently, the output of data in a graphical format is not part of the AFP system software but is done external to the main program. The plans are to add this capability but were not done in Phase I since the formats for graphical output should allow for mixtures. Therefore, the graphical display of results will be added in Phase II. However, a significant amount of the work was accomplished in Phase I for the external graphical display capability.

The ability to input molecular structures via a graphical user interface was delayed during Phase I due to the high cost of the initially considered commercial molecular graphics package. This became even a greater consideration when it was discovered that the software pack MedChem was soon to introduce its own molecular graphics input package. This would be a much more cost effective path since the MedChem software package is already an integral part of the AFP system. Therefore, until MedChem releases its molecular graphics input package and it can be examined for quality and ease of use, no further effort on graphical input is planned. An additional plus for the MedChem graphical input choice is the availability of source code; the package that was initially

considered is supplied only as an executable, making it much more difficult to interface with the AFP system and impossible within FORTRAN77 standards.

6.0 CONCLUSIONS

This section summarizes the primary technical and software development aspects of the project.

6.1 Technical Conclusions

One of the major accomplishments of the project had was the development of a very large, computer accessible data base of physical and thermochemical properties. A large number of property prediction methods have been selected, programmed, tested, and integrated into a user friendly property determination system.

One of the key emphases taken in the project was to carefully consider the implementation of all aspects of Phase I and how they will need to be integrated into Phases II and III. In this regard, during some of the programming effort, it was prudent to add in the capability to handle mixtures initially, rather than waiting for Phase II and going back and modifying the software (although some of the routines have this capability, it is not yet fully implemented.)

6.2 Software Conclusions

The AFP determination system is a menu based user friendly system that can provide the user with values for a large number of properties and compounds. A priority scheme was developed which attempts to provide the best value available for the compound(s) of interest. The priority scheme was designed to be flexible so that as the system progresses more towards an expert system, the system itself can modify the method priority scheme "on-the-fly."

The system was designed to easily accommodate new methods to calculate new properties, as well as to handle new ways to supplement existing methods, and possibly to better handle certain classes of compounds.

An effort was made early to incorporate the reporting of an estimated relative error for every value reported. This was implemented for a number of routines; however, it was found that in numerous cases the method by which the error was being determined (propagating errors from the parameters needed in the calculation) proved to significantly overestimate the error. We have begun implementing a different (numerically based) method for determining the relative errors for the predicted property values. One additional benefit that this method will have, is that it will provide us with the capability to do a sensitivity analysis on the parameters needed for each property prediction method.

REFERENCES

1. Hirschfelder, J.O., Curtiss, C.F., and Bird, R.B., "Molecular Theory of Gases and Liquids," John Wiley, New York, (1964).
2. American Petroleum Institute Research Project 44, Carnegie Press, Pittsburgh, PA (1953).
3. NBS Chemical Thermodynamics Database, Available from NBS Standard Reference Databases, Gaithersburg, MD 20899 (1983).
4. NBS Standard Reference Database 13: JANAF Thermochemical Tables, 1985. Available from NBS Standard Reference Databases, Gaithersburg, MD 20899 (1985).
5. Hougen, O.A., Watson, K.M., Ragatz, R.A., "Chemical Process Principles Charts," 3rd Ed. (1964).
6. "Technical Data Book Petroleum Refining," 4th Ed., American Petroleum Institute, 1983. (Revised 1985).
7. "Manual for Predicting Chemical Process Design Data," Danner, R.P. and Daubert, T.E., Editors, American Institute of Chemical Engineers, 1984.
8. DIPPR Database, Available from NBS Standard Reference Databases, Gaithersburg, MD 20899 (1987).
9. Steele, W. V., Archer, D. G., Chirico, R. D., and Stube, M. M., "Thermodynamics of Materials in the Range C_{10} - C_{16} Data Base Description, Uses, and Future Work Recommendations," NTIS Number: NIPER-333, March, 1988.
10. "TRC Thermodynamic Tables, Hydrocarbons," Thermodynamic Research Center, Texas Engineering Experimental Station, Texas A&M University System, 1986.
11. "TRC Thermodynamic Tables, Non-hydrocarbons", Thermodynamic Research Center, Texas Engineering Experimental Station, Texas A&M University System, 1986.
12. "MedChem Software Manual," Release 3.54, Daylight Chemical Information Systems, Inc., Claremont, CA, 1989.
13. Reid, R. C., Prausnitz, J. M., and Poling, B. E., "The Properties of Gases and Liquids," 4th Edition, McGraw-Hill, New York, 1987.
14. Edminster, W. C. and Lee, B. I., "Applied Hydrocarbon Thermodynamics," Vol. 1, 2nd Edition, Gulf Publishing, Houston, TX, 1984.
15. Benson, S.W., "Thermochemical Kinetics," Wiley, New York, 1976.

16. Reid, Prausnitz, and Poling (Ref. 13), p. 12.
17. Mathur, Ibrahim, Kuloor. Chem. Eng., (3/24/69).
18. Jalowka, J. W. and Daubert, T. E. Ind. Eng. Chem. Process Des. Dev., 25, 139-142 (1986).
19. Reid, Prausnitz, and Poling (Ref. 13), p. 14.
20. AIChE Manual (Ref.7), p. 2C-1.
21. API Manual (Ref. 6), p. 4-13.
22. AIChE (Ref. 7), p. 2F-1.
23. Reid, Prausnitz, and Poling (Ref. 13), eq. 2-3.5, p. 24.
24. Reid, Prausnitz, and Poling (Ref. 13), eq. 2-3.4, p. 23.
25. Reid, Prausnitz, and Poling (Ref. 13), eq. 2-3.1, p. 23.
26. Reid, Prausnitz, and Poling (Ref. 13), eq. 2-3.3, p. 23.
27. Reid, Prausnitz, and Poling (Ref. 13), p. 56.
28. Reid, Prausnitz, and Poling (Ref. 13), eq. 3-11.4, p. 56.
29. Reid, Prausnitz, and Poling (Ref. 13), eq. 3-11.4, p. 55.
30. Reid, Prausnitz, and Poling (Ref. 13), p. 56.
31. Reid, Prausnitz, and Poling (Ref. 13), eq. 3-11.10, p. 67 .
32. API Manual (Ref. 6), p. 6-29.
33. Reid, Prausnitz, and Poling (Ref. 13), eq. 3-11.9, p. 67.
34. Reid, Prausnitz, and Poling (Ref. 13), p. 47.
35. Edminster and Lee (Ref. 14), p. 62-63.
36. TRC Hydrocarbon Tables (Ref. 10), Table j.
37. Abramowitz, Milton and Stegun, I. A., "Handbook of Mathematical Functions," Dover, New York, NY, 1964, p. 17.
38. "IMSL Library FORTRAN Subroutines for Mathematics and Statistics Manual," IMSL, Houston, TX, 1987.
39. Reid, Prausnitz, and Poling (Ref. 13), p. 67.
40. Reid, Prausnitz, and Poling (Ref. 13), p. 55.

41. Reid, Prausnitz, and Poling (Ref. 13), eq. 3-11.5, p. 66.
42. API (Ref. 6), eq. 6A2.22, p. 6-47.
43. AIChE (Ref. 7), p. 3A-1.
44. Reid, Prausnitz, and Poling (Ref. 13), eq. 7-2.6 - 7-2.8, p. 207.
45. Reid, Prausnitz, and Poling (Ref. 13), eq. 7-5.1 & 7.5-2, p. 212.
46. API (Ref. 6), method 11B1.6, p. 11-63.
47. AIChE (Ref. 7), p. 8A.
48. Hall, L. H., "MOLCONN2, A User's Guide," Hall Associates Consulting, Quincy, MA (1988).
49. Kier, L. B. and Hall, H. H., "Molecular Connectivity in Structure-Activity Analysis," Research Studies Press Ltd., Letchworth, Hertfordshire, England, (1986).
50. Trinajstić, Nenad, "Chemical Graph Theory," Volumes I and II, CRC Press, Boca Raton, Florida (1983).
51. Kier, L. B. and Hall, L. H., "Molecular Connectivity in Chemistry and Drug Design," Academic Press, New York (1976).
52. SAS Statistical Analysis Package, SAS Institute Inc., Cary, North Carolina (1985).