





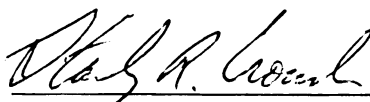
This is to certify that the  
dissertation entitled  
The Development of an Intelligent  
Automated High Performance Liquid  
Chromatographic System

presented by

Stephen Vance Medlin

has been accepted towards fulfillment  
of the requirements for

Ph.D. degree in Chemistry

  
Major professor

Date June 30, 1993



**PLACE IN RETURN BOX** to remove this checkout from your record.  
**TO AVOID FINES** return on or before date due.

| DATE DUE | DATE DUE | DATE DUE |
|----------|----------|----------|
| _____    | _____    | _____    |
| _____    | _____    | _____    |
| _____    | _____    | _____    |
| _____    | _____    | _____    |
| _____    | _____    | _____    |
| _____    | _____    | _____    |
| _____    | _____    | _____    |

**DEVELOPMENT OF AN INTELLIGENT AUTOMATED  
HIGH PERFORMANCE LIQUID CHROMATOGRAPHY  
SYSTEM**

**Volume I**

**By  
Stephen Vance Medlin**

**A DISSERTATION**

**Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of**

**DOCTOR OF PHILOSOPHY**

**Department of Chemistry**

**1993**



M

complex

appropria

automate

complexit

Th

expert sys

solute(s).

interaction

column, th

The

recommen

suggesting

success via

The

software. A

## **ABSTRACT**

# **DEVELOPMENT OF AN AUTOMATED, INTELLIGENT HIGH PERFORMANCE LIQUID CHROMATOGRAPHY SYSTEM**

**By**

**Stephen Vance Medlin**

**Methods development for high performance liquid chromatography (HPLC) is a complex process. An expert system has been created to assist in choosing an appropriate column, detector, and mobile phase. The HPLC instrument has been automated to provide the ability to perform column switching for samples of sufficient complexity that multiple columns may be required to separate the solutes.**

**The expert system uses a first principle approach to deduce a method(s). The expert system queries the user with respect to the chemical and physical nature of the solute(s). The expert system proposes a method based on the possible chemical interactions between analyte(s) and stationary phases. After recommendation of a column, the expert system recommends an appropriate detector and mobile phase.**

**The expert system uses confidence factors to rank the different recommendations. Validation studies are presented to demonstrate the first level of suggesting appropriate methods and providing a numerical ranking of the expected success via the use of confidence factors.**

**The HPLC instrument has been automated via a combination of hardware and software. A modular software system has been written which allows the user to create a**

method for

switching

achieved

selection v

Re:

consultation

perform th

ability to p

method for separating the sample. The software provides the ability to perform column switching based upon time criteria or the elution of sample. Column switching is achieved through the automation of two six-port valves and a six-position column selection valve.

Results are presented which demonstrate the ability to develop a method from consultation with the expert system and then programming the HPLC instrument to perform the experiment. Preliminary results are also presented which demonstrate the ability to perform reproducible and complex column switching.

*To Kelly, who always is there for me*

develop

support

J

From lat

prodded

instrume

compute

using a f

Er

"joys" of

about the

least I hav

won't ever

return the

I w

friendship

## Acknowledgements

Great discoveries and improvements invariably involve the cooperation of many minds. I may be given credit for having blazed the trail but when I look at the subsequent developments I feel the credit is due to others rather than to myself.

Alexander Graham Bell

I would like to Dr. Adrian Wade who introduced me to expert systems and who developed the prototype expert system. I would also like to thank Adrian for his support and help that he has always shown me.

Jim Ridge and I spent a lifetime, it seems, together in the Chemistry Building. From late night experiments at 4:00 a.m. to early breakfasts at Theios, Jim has helped, prodded, and supported my research. Jim's expertise with electronics and instrumentation helped get balky equipment to work. I am also thankful for Jim's 486 computer—writing a dissertation was made a little (just a little) more tolerable when using a fast computer. Thanks also for the dart, racquetball, and basketball games.

Eric Hemenway, the computer guru of the department, introduced me to the "joys" of assembly language. Eric was always there to help me debug my software, talk about the future of computers, or to play a game of one-on-one. Eric also donated (at least I haven't received a bill yet) his computer for me to use as the HPLC controller. I won't even mention all the reference books and materials that I borrowed from him (I did return them all, didn't I?).

I would like to thank several people who are my friends. In graduate school, friendship goes a long way to reducing stress. John Judge, D & D guru and all-around

good

what

pres

know

the p

supp

Kim

devel

send.

disser

Univer

I had n



good friend, was always there for me whenever I needed to talk, ask questions, or whatever. I also appreciate the friendship of the Crouch group members, past and present.

Of course, the project would never have occurred without the input from the knowledge acquisition team. Dr. Crouch, my research preceptor, was the team leader of the project and kept our project on track and focused. He also provided guidance and support of me as a graduate student. Dr. McGuffin was our resident HPLC expert. Drs. Kim Ratanathanawongs and Chris Evans provided additional input into the early development stages. Finally, Dr. Wade was the initial knowledge engineer.

Finally, I would like to thank my family members. My wife, Kelly, was a God-send. Not only did she support me emotionally, she proofread and edited my disseratation. My parents have been supportive during my tenure at Michigan State University. I think that they were as happy, if not more so, when I finally told them that I had my Ph. D.

Occasionally in life there are those moments of unutterable fulfillment which cannot be completely explained by those symbols called words. Their meanings can only be articulated by the inaudible language of the heart.

Martin Luther King, Jr.

LIST

LIST

CHAP

S

D

## TABLE OF CONTENTS

|                      |    |
|----------------------|----|
| LIST OF TABLES ..... | xv |
|----------------------|----|

|                       |      |
|-----------------------|------|
| LIST OF FIGURES ..... | xvii |
|-----------------------|------|

|                                                                                                                           |    |
|---------------------------------------------------------------------------------------------------------------------------|----|
| CHAPTER 1. MOTIVATION FOR THE DEVELOPMENT OF AN INTELLIGENT AUTOMATED HIGH-PERFORMANCE LIQUID CHROMATOGRAPHY SYSTEM ..... | 1  |
| Section 1.1. Introduction .....                                                                                           | 1  |
| Section 1.2. Use of Computers in Chromatography—Overview.....                                                             | 3  |
| Section 1.2.1. Prediction, Simulation, and Optimization .....                                                             | 3  |
| Section 1.2.2. Suggestion of Initial HPLC Conditions and Automation .....                                                 | 9  |
| Section 1.3. Expert System as an HPLC Tool—Overview .....                                                                 | 10 |
| Section 1.4. Required Information to Implement the Expert System ....                                                     | 11 |
| Section 1.5. Output from the Expert System.....                                                                           | 12 |
| Section 1.6. Automation as an HPLC Tool .....                                                                             | 14 |
| Section 1.7. Automating the HPLC .....                                                                                    | 16 |
| List of References.....                                                                                                   | 18 |

|                                                                                        |           |
|----------------------------------------------------------------------------------------|-----------|
| <b>CHAPTER 2. HISTORY AND THEORY.....</b>                                              | <b>20</b> |
| Section 2.1. Artificial Intelligence.....                                              | 20        |
| Section 2.2. Background on Expert Systems .....                                        | 23        |
| Section 2.3. Terminology .....                                                         | 26        |
| Section 2.3.1. Knowledge and Fact Bases .....                                          | 26        |
| Section 2.3.2. Inference Engine.....                                                   | 29        |
| Section 2.3.3. User Interface.....                                                     | 30        |
| Section 2.4. Historically Important Expert Systems.....                                | 31        |
| Section 2.5. Expert System Applications in Analytical Chemistry.....                   | 34        |
| Section 2.5.1. Non-HPLC Expert Systems.....                                            | 34        |
| Section 2.5.2. HPLC Expert Systems .....                                               | 35        |
| Section 2.6. Implementation of Expert Systems.....                                     | 38        |
| Section 2.6.1. Goal Determination .....                                                | 38        |
| Section 2.6.2. Knowledge Acquisition .....                                             | 39        |
| Section 2.6.2.1. Methods for Knowledge Acquisition.....                                | 39        |
| Section 2.6.2.2. Personnel and Roles.....                                              | 40        |
| Section 2.6.3. Encoding.....                                                           | 41        |
| Section 2.6.3.1. AI and Expert System Languages.....                                   | 41        |
| Section 2.6.3.2. Prototyping of the Expert System .....                                | 42        |
| Section 2.7. Confidence Factors .....                                                  | 42        |
| Section 2.7.1. Comparison of Probability and Possibility .....                         | 42        |
| Section 2.7.2. Fuzzy Logic .....                                                       | 44        |
| Section 2.7.3. Certainty Theory.....                                                   | 45        |
| Section 2.8. Conclusions of Expert Systems and Certainty Theory .....                  | 47        |
| Section 2.9. Multidimensional High Performance Liquid<br>Chromatography (MD-HPLC)..... | 49        |
| Section 2.9.1. Definition and Theory.....                                              | 49        |
| Section 2.9.2. Applications of MD-HPLC.....                                            | 51        |
| List of References.....                                                                | 55        |

CHA

CHA  
I  
CHAPT

S

S

## **CHAPTER 3. INSTRUMENTATION AND SOFTWARE..... 59**

|                                                                      |           |
|----------------------------------------------------------------------|-----------|
| <b>Section 3.1. Expert System .....</b>                              | <b>59</b> |
| Section 3.1.1. AI Languages Versus Commercial Shells.....            | 59        |
| Section 3.1.2. M1 Expert System Shell.....                           | 64        |
| Section 3.1.3. VP-Expert Expert System Shell .....                   | 65        |
| Section 3.1.4. Additional Software Support.....                      | 69        |
| <b>Section 3.2. HPLC Equipment and Supplies.....</b>                 | <b>69</b> |
| Section 3.2.1. Instrumentation.....                                  | 70        |
| Section 3.2.2 Solvents .....                                         | 70        |
| Section 3.2.3 Columns .....                                          | 70        |
| <b>Section 3.3. Computer Control and Automation.....</b>             | <b>71</b> |
| Section 3.3.1. Six Port Rheodyne Valves .....                        | 71        |
| Section 3.3.2. Column Selection Valve .....                          | 72        |
| Section 3.3.3. Valve Sensors .....                                   | 72        |
| Section 3.3.4. Pilot Valves .....                                    | 75        |
| Section 3.3.5. Actuators, Solenoids, and Relays .....                | 76        |
| Section 3.3.6 Computer Hardware to Control the HPLC Instrument ..... | 77        |
| Section 3.3.6.1 Prototype Board .....                                | 77        |
| Section 3.3.6.2 Programmable Parallel Interface Board .....          | 77        |
| Section 3.3.6.3 Timing Circuit .....                                 | 79        |
| Section 3.3.6.4 Data Acquisition Board .....                         | 80        |
| Section 3.3.7 Interface Box .....                                    | 80        |
| Section 3.3.8 Software .....                                         | 80        |
| <b>List of References.....</b>                                       | <b>84</b> |

## **CHAPTER 4. EXPERT SYSTEM: COLUMN SELECTION AND VALIDATION ..... 85**

|                                                                 |           |
|-----------------------------------------------------------------|-----------|
| <b>Section 4.1. Introduction .....</b>                          | <b>85</b> |
| <b>Section 4.2. Goals of Research.....</b>                      | <b>85</b> |
| Section 4.2.1. Off-Line Tool for HPLC Methods Development ..... | 85        |
| Section 4.2.1.1. End-User .....                                 | 86        |

|                  |                                                                     |     |
|------------------|---------------------------------------------------------------------|-----|
| Section 4.2.1.2. | Type of Application .....                                           | 87  |
| Section 4.2.1.3. | First Principle Approach .....                                      | 88  |
| Section 4.2.2.   | Tutorial and Training Tool .....                                    | 90  |
| Section 4.2.3.   | On-Line Tool for Automated Intelligent HPLC<br>Instrumentation..... | 91  |
| Section 4.3.     | Knowledge Acquisition.....                                          | 91  |
| Section 4.3.1.   | Personnel and Roles .....                                           | 92  |
| Section 4.3.2.   | Knowledge Acquisition Session Format.....                           | 92  |
| Section 4.3.3.   | Additional Sources of Information .....                             | 96  |
| Section 4.3.3.1. | Types of Data Used.....                                             | 96  |
| Section 4.4.     | Action Block .....                                                  | 99  |
| Section 4.4.1.   | Determination of Suitability .....                                  | 102 |
| Section 4.4.2.   | Determination of Sample Pre-Treatment .....                         | 106 |
| Section 4.4.3.   | Determination of Class of Separation .....                          | 106 |
| Section 4.4.4.   | Determination of Column and Mode.....                               | 108 |
| Section 4.4.4.1. | Physical Properties .....                                           | 109 |
| Section 4.4.4.2. | Chemical Properties.....                                            | 112 |
| Section 4.5.     | Rules .....                                                         | 116 |
| Section 4.5.1.   | Order of Rules.....                                                 | 116 |
| Section 4.5.2.   | Forward Chaining.....                                               | 119 |
| Section 4.5.3.   | Self-Determining Rules .....                                        | 121 |
| Section 4.6.     | User Interface.....                                                 | 122 |
| Section 4.7.     | Preliminary Validation of the M1 Prototype Expert<br>System .....   | 122 |
| Section 4.7.1.   | Estrogens Test Case—M1 Prototype .....                              | 124 |
| Section 4.7.2.   | Validation of B Vitamins—M1 Prototype .....                         | 129 |
| Section 4.7.3.   | Validation of a Hypothetical Case—M1<br>Prototype.....              | 131 |
| Section 4.8.     | Secondary Validation of the VP-Expert Expert<br>System .....        | 131 |
| Section 4.8.1.   | Estrogens Test Case—VP-Expert .....                                 | 133 |
| Section 4.8.2.   | B Vitamins Test Case—VP-Expert .....                                | 137 |
| Section 4.8.3.   | Hypothetical Test Case—VP-Expert.....                               | 140 |
| Section 4.8.4.   | Conclusions of Validation Studies .....                             | 145 |

CHA



|                         |                                                                            |     |
|-------------------------|----------------------------------------------------------------------------|-----|
| Section 4.9.            | Confidence Factors .....                                                   | 148 |
| Section 4.9.1.          | Motivations for Using Confidence Factors .....                             | 148 |
| Section 4.9.1.1.        | Fuzzy Input.....                                                           | 148 |
| Section 4.9.1.2.        | Appropriate Values for Confidence<br>Factors .....                         | 149 |
| Section 4.10.           | Ancillary Files—Text, Help Files, and Data Bases .....                     | 143 |
| List of References..... |                                                                            | 155 |
|                         |                                                                            |     |
| CHAPTER 5.              | EXPERT SYSTEM: DETECTOR AND MOBILE PHASE<br>SELECTION AND VALIDATION ..... | 157 |
|                         |                                                                            |     |
| Part 1:                 | Recommendation of Detector .....                                           | 158 |
|                         |                                                                            |     |
| Section 5.1.            | Overview of the Detector Module.....                                       | 159 |
| Section 5.1.1.          | Goals of the Detector Module.....                                          | 159 |
| Section 5.1.2.          | Detectors recommended by the Expert<br>System .....                        | 160 |
| Section 5.1.3.          | Format of the Presentation of the Different<br>Detectors.....              | 161 |
| Section 5.2.            | Confidence Factors Based on Concentration .....                            | 162 |
| Section 5.3.            | UV-Visible Detectors.....                                                  | 168 |
| Section 5.3.1.          | Overview.....                                                              | 168 |
| Section 5.3.2.          | Determination of Chromophores .....                                        | 169 |
| Section 5.3.3.          | Determination of Optimum Wavelength.....                                   | 172 |
| Section 5.3.4.          | Determination of Concentration Range .....                                 | 173 |
| Section 5.4.            | Fluorescence Detectors .....                                               | 174 |
| Section 5.4.1.          | Overview.....                                                              | 174 |
| Section 5.4.2.          | Determination of Fluorophores .....                                        | 174 |
| Section 5.4.3.          | Determination of Concentration Range and<br>CNF.....                       | 175 |
| Section 5.5.            | Electrochemical Detectors.....                                             | 177 |
| Section 5.5.1.          | Overview and Determination if Species are<br>Electrochemically Active..... | 177 |



CHAPTER

P

S

Se

Se

Se

Se

|                                                                                                    |         |
|----------------------------------------------------------------------------------------------------|---------|
| Section 5.12.1. Phthalate Validation Study.....                                                    | 207     |
| Section 5.12.2. Aromatic Validation Study.....                                                     | 209     |
| Part 4: Conclusions and Summary.....                                                               | 213     |
| Section 5.13. Summary of Validation Studies .....                                                  | 214     |
| List of References.....                                                                            | 216     |
| <br>CHAPTER 6. AUTOMATION OF THE HIGH-PERFORMANCE LIQUID<br>CHROMATOGRAPHY SYSTEM INSTRUMENT ..... | <br>217 |
| Section 6.1. Introduction .....                                                                    | 217     |
| Part 1: Hardware and Software Used for Automation.....                                             | 219     |
| Section 6.2. Initial Controller Software.....                                                      | 220     |
| Section 6.2.1. Objectives .....                                                                    | 220     |
| Section 6.2.2. Data Acquisition.....                                                               | 220     |
| Section 6.2.3. Control of Two Six Port Valves .....                                                | 227     |
| Section 6.2.4. Recycle Heartcut Chromatography .....                                               | 227     |
| Section 6.3. Second Level Controller Software—An Overview .....                                    | 230     |
| Section 6.3.1. SwitValv.Bas Program.....                                                           | 232     |
| Section 6.3.2. HPLC-Con.Bas Program .....                                                          | 239     |
| Section 6.3.3. LCMethod.Bas Program.....                                                           | 239     |
| Section 6.3.4. ProgSDS.Bas Program .....                                                           | 241     |
| Section 6.4. Assembly Language Routine .....                                                       | 242     |
| Section 6.5. Programmable Parallel Interface Prototype Board.....                                  | 243     |
| Section 6.6. Pneumatic Amplifier .....                                                             | 247     |
| Section 6.6.1. Motivation for Use .....                                                            | 247     |
| Section 6.6.2. Initial Software Control .....                                                      | 248     |
| Section 6.6.3. Timing Circuit.....                                                                 | 249     |

CHAP

APPE

|                         |                                                                                                                   |            |
|-------------------------|-------------------------------------------------------------------------------------------------------------------|------------|
| <b>Part 2:</b>          | <b>Preliminary Applications of the Controller<br/>Software for HPLC Separations .....</b>                         | <b>255</b> |
| <b>Section 6.7.</b>     | <b>Integration of the Expert System and HPLC Controller .....</b>                                                 | <b>256</b> |
| <b>Section 6.7.1.</b>   | <b>Determination of HPLC Controller Parameters .....</b>                                                          | <b>256</b> |
| <b>Section 6.7.1.1.</b> | <b>Peak Detection Thresholds.....</b>                                                                             | <b>256</b> |
| <b>Section 6.7.2.</b>   | <b>Validation Study of Benzaldehyde, Biphenyl, and<br/>Naphthalene .....</b>                                      | <b>269</b> |
| <b>Section 6.7.3.</b>   | <b>Validation Study of Benzoic Acid, p-Aminobenzoic Acid,<br/>2,4-Dinitrophenol, Anthracene, and Pyrene .....</b> | <b>270</b> |
| <b>Part 3:</b>          | <b>Conclusions Based on the Automation of the HPLC<br/>Instrument .....</b>                                       | <b>272</b> |
| <b>Section 6.8.</b>     | <b>Conclusions of the Automation of the HPLC Instrument .....</b>                                                 | <b>273</b> |
|                         | <b>List of References.....</b>                                                                                    | <b>275</b> |
| <b>CHAPTER 7.</b>       | <b>CONCLUSIONS AND FUTURE PROSPECTS .....</b>                                                                     | <b>276</b> |
| <b>Section 7.1.</b>     | <b>Introduction .....</b>                                                                                         | <b>276</b> |
| <b>Section 7.2.</b>     | <b>Conclusions .....</b>                                                                                          | <b>276</b> |
| <b>Section 7.3.</b>     | <b>Future Work.....</b>                                                                                           | <b>280</b> |
|                         | <b>List of References.....</b>                                                                                    | <b>282</b> |
| <b>APPENDIX A</b>       | <b>FLOW CHARTS FOR THE DETERMINATION OF AN<br/>APPROPRIATE COLUMN(S) .....</b>                                    | <b>283</b> |

TAB

3.1

3.2

3.3

3.4

3.5

4.1

4.2

4.3

4.4

4.

4.

4.

4.

4.

4.

5.

5.

5.

5.

## LIST OF TABLES

| TABLE |                                                                                                            | PAGE |
|-------|------------------------------------------------------------------------------------------------------------|------|
| 3.1   | Comparison of Shells Written from AI Languages versus Commercially-Available Shells                        | 61   |
| 3.2   | Comparison of Different Expert Systems                                                                     | 66   |
| 3.3   | Summary of Features of VP-Expert                                                                           | 68   |
| 3.4   | HPLC Columns and their Parameters Used in this Research                                                    | 71   |
| 3.5   | Digital Input/Output Lines Used From the PPI Chip                                                          | 81   |
| 4.1   | Excerpt from Knowledge Acquisition Meeting                                                                 | 95   |
| 4.2   | Action Block for First Module                                                                              | 101  |
| 4.3   | Action Block for Second Module                                                                             | 101  |
| 4.4   | Possible Sub-Modes that may be Recommended by the Expert System                                            | 110  |
| 4.5   | Determination of Sub-Modes                                                                                 | 117  |
| 4.6   | Dialog for the Estrogen Consultation                                                                       | 127  |
| 4.7   | Results of the Validation Study for the Estrogen Components Using the VP-Expert Version                    | 135  |
| 4.8   | Output from Validation Study of Hypothetical Case: VP-Expert Version                                       | 145  |
| 4.9   | Truth Threshold Values                                                                                     | 150  |
| 4.10  | Assignment of CNF for SEC based upon Competing Mechanisms                                                  | 151  |
| 5.1   | Detectors Recommended by the Expert System                                                                 | 162  |
| 5.2   | Determination of the Optimum Method to Calculate the CNF based on the CNFs of Two Different Concentrations | 170  |
| 5.3   | UV-Visible Chromophores Assigned by the Expert System                                                      | 172  |
| 5.4   | Structures of the Mobile Phase Databases used by HPLC Advisor                                              | 190  |



| <b>TABLE</b> |                                                                                    | <b>PAGE</b> |
|--------------|------------------------------------------------------------------------------------|-------------|
| 5.5          | Mobile Phase Modifiers for Normal-Phase Bonded Chromatography                      | 193         |
| 5.6          | Mobile Phase Modifiers for Reverse-Phase Bonded Chromatography                     | 195         |
| 5.7          | Recommended Mobile Phase Ions for Ion Exchange and Ion Chromatography              | 200         |
| 5.8          | Effects of Different Mobile Phase Parameters on IPC Separations                    | 204         |
| 5.9          | Physical and Chemical Properties of the Components of the Phthalate Test Sample    | 208         |
| 5.10         | Results of the Validation Study of the Phthalate Test Case                         | 209         |
| 5.11         | Physical and Chemical Properties of the Components of the Aromatic Test Sample     | 211         |
| 5.12         | Results of the Aromatic Validation Test                                            | 212         |
| 6.1          | Format for the Encoded Link List Array, LinkFile()                                 | 240         |
| 6.2          | Digital Input/Output Lines Needed for the Second Generation Controller             | 245         |
| 6.3          | Logic Table for the 74LS221 Dual Monostable                                        | 252         |
| 6.4          | Results from the Optimization of Switching Times and Pressures for the Pilot Valve | 254         |
| 6.5          | Experimental Parameters for the Analysis of Benzaldehyde and Biphenyl              | 260         |
| A.1          | Criteria Used to Recommend Normal-Phase Size Exclusion Chromatography              | 285         |
| A.2          | Criteria Used to Recommend Reverse-Phase Size Exclusion Chromatography             | 285         |

FIG

3

3

3

4.

4.

4.3

4.4

4.5

4.6a

4.6b

5.1

5.2

5.3

## LIST OF FIGURES

| FIGURE |                                                                                                          | PAGE |
|--------|----------------------------------------------------------------------------------------------------------|------|
| 3.1    | Several Configurations of the Two Rheodyne Valves that Can Be Used                                       | 74   |
| 3.2    | Sensor Connections For the Column Selection Valve                                                        | 75   |
| 3.3    | Pilot Valve Configuration                                                                                | 76   |
| 3.4    | Complete Valve Arrangement of the Two Rheodyne Valves, Column Selection Valve, and Pilot Valves          | 78   |
| 3.5    | 74LS221 Dual Monostable Timing Circuit                                                                   | 82   |
| 3.6    | Interface Box for the RTI 800/815 Board                                                                  | 83   |
| 4.1    | Flow Chart for Choosing an HPLC Column                                                                   | 100  |
| 4.2    | Example of Dialog that Determines the Separation Property                                                | 114  |
| 4.3    | Structures of the Conjugate Estrogens Used in the Validation Study                                       | 126  |
| 4.4    | Structure of Four B Vitamins: Niacin, Niacinamide, Riboflavin, and Pyridoxine HCl                        | 130  |
| 4.5    | Flow Chart for the Determination of Ion Chromatography Techniques: M1 Version                            | 141  |
| 4.6a   | Flow Chart for the Determination of Ion Chromatography Techniques for Small Molecules: VP-Expert Version | 142  |
| 4.6b   | Flow Chart for the Determination of Ion Chromatography Techniques for Large Molecules: VP-Expert Version | 143  |
| 5.1    | Trapezoidal Function for the Calculation of Concentration Confidence Factors                             | 166  |
| 5.2    | Different Concentration Values to Determine Optimum CNF <sub>Total</sub> Calculation                     | 168  |
| 5.3    | Flow Chart to Determine if a UV-Visible Detector is an Appropriate Choice                                | 171  |

FIG

5

5

5

5

5

5

5

5

5

5

5

5

5

5

6.

6.

6.

| <b>FIGURE</b> |                                                                                                       | <b>PAGE</b> |
|---------------|-------------------------------------------------------------------------------------------------------|-------------|
| 5.4           | Flow Chart to Determine if a Fluorescent Detector is an Appropriate Choice                            | 176         |
| 5.5           | Flow Chart to Determine if a Electrochemical Detector is an Appropriate Choice                        | 178         |
| 5.6           | Flow Chart to Determine if a Atomic Spectroscopy Detector is an Appropriate Choice                    | 180         |
| 5.7           | Flow Chart to Determine if a Refractive Index Detector is an Appropriate Choice                       | 183         |
| 5.8           | Flow Chart to Determine if a Conductivity Detector is an Appropriate Choice                           | 184         |
| 5.9           | Flow Chart to Determine if a Light Scattering Detector is an Appropriate Choice                       | 186         |
| 5.10          | Flow Chart to Determine if Mobile Phase Modifiers are Compatible with Different Detectors             | 191         |
| 5.11          | Flow Chart for the Determination of Mobile Phase Modifiers for Normal-Phase Bonded Chromatography     | 194         |
| 5.12          | Flow Chart for the Determination of Mobile Phase Modifiers for Reverse-Phase Bonded Chromatography    | 196         |
| 5.13          | Flow Chart for the Determination of Mobile Phase Modifiers for Normal-Phase Adsorption Chromatography | 197         |
| 5.14          | Flow Chart for the Determination of Mobile Phase Modifiers for Ion Exchange and Ion Chromatography    | 199         |
| 5.15          | Flow Chart for the Determination of Mobile Phase Modifiers for Size Exclusion Chromatography          | 202         |
| 5.16          | Flow Chart for the Determination of Mobile Phase Modifiers for Ion-Paired Chromatography              | 205         |
| 5.17          | Chromatographic Separation of Phthalate Sample                                                        | 210         |
| 6.1           | Envisioned System Involving Multiple Detectors and Valves Under Computer Control                      | 221         |
| 6.2           | Valve Configuration for Recycle Heartcut Chromatography                                               | 228         |
| 6.3           | Selectivity Enhancements using Recycle Heartcut Chromatography: Naphthalene and Biphenyl              | 231         |

F

6

6

6

6

6.

A.

A.2

A.3

A.4

| <b>FIGURE</b> |                                                                                        | <b>PAGE</b> |
|---------------|----------------------------------------------------------------------------------------|-------------|
| 6.4           | Overview of the Modules used in the Second Generation of the Controller Software       | 233         |
| 6.5           | Gate Circuit to Control Timer Counters 3 and 4                                         | 235         |
| 6.6           | Format of the Master File                                                              | 236         |
| 6.7           | Digital I/O Connections from the Programmable Parallel Interface                       | 246         |
| 6.8           | Schematic of Pilot Valve                                                               | 250         |
| 6.9           | 74LS221 Dual Monostable Circuit                                                        | 251         |
| 6.10          | Timing Scheme used to Switch the Column Selection Valve via the 74LS221 Timing Circuit | 253         |
| 6.11          | Biphenyl and Benzaldehyde Chromatogram Rate of 1 Acquisition/Second                    | 261         |
| 6.12A         | Biphenyl and Benzaldehyde Chromatogram Rate of 10 Acquisitions/Second                  | 262         |
| 6.12B         | Biphenyl and Benzaldehyde Chromatogram Rate of 10 Acquisitions/Second                  | 263         |
| 6.13          | Plot of Slope Versus Time for Rate of 1 Acquisition/Second                             | 265         |
| 6.14          | Plot of Slope Versus Time for Rate of 10 Acquisitions/Second                           | 266         |
| 6.15          | Inset of the Slope for the Rate of 10 Acquisitions/Second                              | 267         |
| 6.16          | Circuit to Provide a Gain of 5 to the HPLC Detector Output                             | 268         |
| 6.17          | Chromatogram of Benzaldehyde, Naphthalene, and Biphenyl                                | 271         |
| A.1           | Flow Chart for the Determination of Separation Property                                | 286         |
| A.2           | Flow Chart for the Determination of Mode for Normal Phase Adsorption (Part 1)          | 287         |
| A.3           | Flow Chart for the Determination of Mode for Normal Phase Adsorption (Part 2)          | 288         |
| A.4           | Flow Chart for the Recommendation of Packing Material                                  | 289         |

## Section

numbe

develo

lead to

develop

that ar

Automa

with pro

L

pathway

separatio

must be

few exper

(which is

should be



# **CHAPTER 1**

## **MOTIVATION FOR THE DEVELOPMENT OF AN INTELLIGENT AUTOMATED HIGH-PERFORMANCE LIQUID CHROMATOGRAPHY SYSTEM**

**[A computer is] like an Old Testament god, with a lot of rules and no mercy.**

**Joseph Campbell**

### **Section 1.1 Introduction**

Method development is a difficult and time-consuming process due to the number of parameters that must be optimized.<sup>1</sup> One approach to assist in methods development has been to automate parts of the procedure. The use of automation can lead to increases in productivity and speed and can decrease the complexity of methods development. Automation has been applied in particular to try to remove the bottleneck that arises from both sample preparation and from data collection and interpretation. Automation has been accomplished via robotics, auto samplers, data integrators, along with procedures such as simplex optimization and factor analysis, among others.

In the separation sciences, experienced chromatographers remove unproductive pathways from being considered when they choose appropriate parameters for a separation. The resultant minimized search field limits the total number of choices that must be considered and tested when doing method development. However, there are few experts available, and their expertise is not widely accessible. In economic terms (which is often the highest priority in the decision-making process), human experts should be solving the more difficult and non-routine problems. These experts are often

not a

exper

and a

kinds

type i

user i

will us

spectre

tool.

perform

develop

chroma

informa

system i

shell. V

DOS ope

literature

At

hardware

that conta

computer

program a

switching.

the analytic

not as interested in answering routine questions which come from non-experts. Human experts are not cost-effective when diverted from problem-solving to routine question and answer sessions with other chromatography users. There is a need to assist two kinds of non-expert high performance liquid chromatography (HPLC) users. The first type is the technician who routinely obtains methods from the HPLC expert. The second user is an expert (e.g., Ph.D. level) in some other area, such as mass spectrometry, who will use HPLC in conjunction with their field of expertise, such as tandem HPLC—mass spectrometry (LC-MS). Each type of user will have different expectations from the HPLC tool.

The research in this dissertation describes an intelligent, automated high performance liquid chromatography system. A medium sized expert system has been developed that assists analytical chemists in choosing viable high performance liquid chromatographic methods. The expert system's knowledge base is composed of information obtained from a human expert and from literature sources. The expert system is implemented using VP-Expert which is a commercially available expert system shell. VP-Expert can be used on machines of the PC XT class and higher under the MS-DOS operating system. The knowledge base has been validated via comparison with literature methods and from further consultations with the expert.

Automation of the HPLC instrument is accomplished via a combination of hardware and software. The automation provides the ability to generate a methods file that contains HPLC parameters (e.g., flow rates and mobile phase compositions). The computer then performs the experiment based on these method files. The automation program also provides a method to perform accurate and reproducible column switching. The next section describes some of the previous work using computers in the analytical chemical and HPLC fields.

Se

Earl

As s

of c

repla

HPLC

These

chroma

to provi

not be

instrume

S

Th

determina

mobile ph

schemes o

flow rate o

of stationar

separations

resolution, o

## **Section 1.2      Use of Computers in Chromatography — Overview**

Computers have played an important position in the development of HPLC. Early applications of computer—like tools were simple integrators and data collectors. As samples and instrumentation increased in complexity, a concomitant increase in use of computers has been seen. The experimental approach to optimization is being replaced by on-line, automated computer optimizations.<sup>2</sup> The use of computers in HPLC can fall into several broad categories:<sup>3</sup>

- acquisition and manipulation of chromatographic data
- suggestion of initial HPLC conditions
- prediction, simulation, and optimization of the HPLC results
- automation of chromatographic instrumentation

These areas will be described below to present a survey of computer usage.

Computers, as integrators, were first used to acquire and manipulate chromatographic data. These integrators had limited ability; their primary purpose was to provide peak area and retention times. The discussion of computer integrators will not be discussed in detail since integrators are routine in modern HPLC instrumentation.

### ***Section 1.2.1      Prediction, Simulation, and Optimization***

The next level of computer-HPLC interaction is a program that can assist in the determination and control of parameters that affect the HPLC separation. Since the mobile phase exerts a significant effect on the separation of a sample, optimization schemes often focus on the choice of solvents, the resultant solvent strength, and the flow rate of the mobile phase. Additional parameters that are optimized include choice of stationary phase and particle sizes, column dimensions, and detectors. Difficult separations require variation of these parameters to optimize either the analysis time, resolution, or some other criteria that is of interest to the user.<sup>2</sup>

the

def

the

(CR

fun

opti

wher

pair.

time

term

parti

by th

with

COF

COF/

chara

maxi

appro

mean

will p

revers

particu

analys

The objectives of optimization are to determine the best conditions to achieve the best separation. A systematic optimization strategy will need a mathematical definition of "best" in order to optimize the separation. Berridge<sup>2</sup> presents a review of the different Chromatographic Response and Chromatographic Optimization Functions (CRF and COF) which have been used to define the quality of a separation. These functions can be established such that they approach either zero or infinity as the optimum separation is reached. For example, Glajch's COF is:

$$\text{COF} = \sum A_i \ln(R_i/R_{id}) + B(t_m - t_n) \quad \text{Equation 1.1}$$

where  $R_i$  is the resolution for the  $i^{\text{th}}$  pair and  $R_{id}$  is the desired resolution for the  $i^{\text{th}}$  pair. The time terms,  $t_m$  and  $t_n$ , refer to the desired maximum analysis time and the time of the of the last detected peak, respectively. The A and B terms are weighting terms to allow the user to specify the importance of time and resolution for the particular COF. As the separation approaches the optimum separation (as determined by the  $R_{id}$  and  $t_m$  terms), Glajch's COF approaches zero and becomes more negative with a decrease in the quality of the separation. Optimization strategies can use this COF to approach the optimum by varying the parameters to maximize the COF. Other COF/CRF definitions can be used to allow optimization based on other desired characteristics (e.g., evenly spaced peaks in a chromatogram, determination of the maximum number of unknown components within a sample, cost).<sup>2</sup> The computerized approaches described below use these CRF/COF functions to provide a mathematical means for the computer to evaluate results and to predict a new set of parameters that will provide an improved CRF/COF.

Billiet and de Galan<sup>4</sup> have reviewed the optimization of the mobile phase for reverse-phase liquid chromatography and reverse-phase ion pair chromatography. In particular, they discuss approaches that utilize designs such as principal component analysis (PCA). This approach attempts to determine those factors that have the

str

flo

a c

cha

res

con

acce

reco

mut

min

sign

syste

that

comb

and t

havin

been

peaks

unkno

spectr

linear

spectr

betwee

differen



strongest effect on the optimization of the chromatography, such as column length, flow rate, and mobile phase composition. This method assumes that: (1) solutes have a enough spectral difference to differentiate between the solutes, (2) the spectral characteristics are constant, and (3) the logarithm of the capacity factor is linear with respect to the mixing rate of binary mobile phases.

The approach is first to perform the separation under gradient elution conditions and apply PCA to determine the number of components in a sample. PCA is accomplished by the following steps: (1) A multiwavelength detector (e.g., a photodiode) records the spectra of the eluting components. (2) The chromatogram is divided into mutually well separated peaks (i.e., clusters). (3) The spectra are normalized to minimize concentration effects. (4) PCA is performed on the spectra to create a list of significant components that describe the spectra and those which describe noise in the system. (5) Iterative Target Transformation Factor Analysis is applied to peak profiles that are spectrally impure. (6) The spectra of each peak is then compared with linear combinations of known spectra to determine if there is match between the unknown and the known spectra. If there is significant overlap in the comparisons, defined as having a correlation coefficient greater than 0.99, the component(s) in that peak have been identified. (7) Target Factor Analysis (TFA) is applied to identify unidentified peaks. TFA requires that test spectra be available to determine the identity of the unknown. TFA determines the identify by using PCA to obtain a number of abstract spectra that can be linearly combined to reconstruct the spectra of the unknown. The linear combination is used to reconstruct the unknown, resulting in an estimated spectra. TFA compares estimated and unknown spectra to determine if the difference between the two spectra is larger than the noise level. A match is obtained when the difference is smaller than the noise level.<sup>4</sup>

peak

select

donor

comb

hexan

separa

the so

ternary

the int

PCA de

were u

compor

chroma

optimu

used to

condition

noted (ba

T

based) to

number o

parameter

As

componen

chromatog

After the number of peaks has been determined in the gradient run, the required peak capacity needed to solve the separation problem can be predicted. A solvent selectivity triangle<sup>2</sup> composed of three mobile phases, one each from a group of proton donor, proton acceptor, and dipole-dipole solvent groups, is generated. Binary combinations of these three solvents are mixed with an appropriate diluent (e.g., n-hexane and water for normal- or reversed-phase chromatography, respectively) and the separations using these binary mixtures are performed on the HPLC. The logarithms of the solute capacity factors measured in the three binary mixtures are plotted over the ternary eluent composition and an optimum mobile phase can be selected based upon the interpolation of these runs using a linear retention model.<sup>4</sup>

Billiet and De Galan performed an optimization on a nine-component mixture. PCA determined the presence of the nine components, of which five of the components were unambiguously determined. TFA was applied and the remaining four components were recognized. This information was used with results from chromatographic runs using three different binary mixtures to predict an initial set of optimum conditions. The retention data obtained from the chromatographic data was used to refine the linear retention model and to predict a new set of optimum conditions. This process is repeated until no further improvement in the separation was noted (based on the CRF used by Billiet and De Galan).

This process is limited (besides the assumptions upon which the approach is based) to the optimization of only three or four parameters. This limit is due to the number of initial experiments that must be performed (binary combinations of the parameters must be performed) and due to the increase in computation time.

As the different binary mixtures are used, the peak order of the different components may change. This band crossing complicates the analysis of chromatograms, and therefore peak tracking algorithms are an important development

in co

(see

appr

math

on no

*DryLa*

metho

using

respon

maxim

will be

will be

ranked

the n-tu

be run.

experim

response

response

•

•

•

•

The disadv

•

in computer optimization. Lankmayr, et al.<sup>5</sup> presented an approach using fuzzy theory (see Chapter 2 of this thesis for an introduction to fuzzy theory) to track peaks. This approach enables the uncertainties in peak areas to be treated in a rigorous mathematical manner. Molnar, et al.<sup>6</sup> developed an approach for peak tracking based on normalized band areas. In particular, they used this approach in conjunction with *DryLab*, a simulation program that will be described below.

A second approach for optimization is simplex optimization. Whereas the above method requires mathematical analysis, the simplex approach empirically optimizes by using a hill-climbing method based upon ranked responses.<sup>7,8</sup> An appropriate response function is selected, which may be a function of resolution, peak shapes, and maximum retention times. The user then determines the number of variables,  $n$ , that will be optimized. The user runs  $(n+1)$  experiments to create the initial responses that will be used in the optimization. The response functions for the  $(n+1)$  experiments are ranked from worst to best, and the conditions for the worst point are reflected through the  $n$ -tuple figure of the other points to determine the next set of parameters that will be run. In this fashion results from previous experiments are used to define future experiments. The process is repeated and the simplex climbs (or descends, if the response function is designed to be minimized) the response surface to an optimum response. The advantages of simplex optimization over the regression methods are:

- simplex optimization is based upon an empirical method; it does not require that a model be created,
- any correlation between parameters is taken into consideration,
- there is no requirement for peak tracking,
- the calculations are simple and non-intensive.

The disadvantages are:

- the response function often converges at local optima rather than global optima,

Re

de

ret

wh

sol

stre

to:

wh

tem

of l

peal

dete

The

(e.g.,

Som

and

discu

have

optim

- no model is generated to provide fundamental information of the system,
- a large number of experiments may be required.

One of the most prominent simulation program is DryLab, marketed by LC Resources (Lafayette, CA). DryLab<sup>9,10</sup> uses the results from two gradient elutions to determine an empirical relationship between mobile phase composition and the retention times of solutes. Empirically, it has been shown that

$$\log k' = \log k_w - S\phi$$

where  $k'$  is the capacity factor of the solute,  $\phi$  is the volume fraction of the strong solvent in the mobile phase, and  $k_w$  and  $S$  are constants that are characteristic of the strong solvent and solute (i.e., fitting parameters). This relationship can be generalized to:

$$\log k' = A - Bx$$

where  $x$  is any of the common parameters used to adjust HPLC retention, such as temperature, pH, and mobile phase composition. This model assumes a linear response of  $\log k'$  to the different parameters. DryLab also does not assist with the problem of peak tracking; instead, it relies on the user being able to apply scientific reasoning in determining if peaks have switched.

After the gradients have been run, the software can derive values for  $K_w$  and  $S$ . The software calculates the effects of varying the mobile phase, column characteristics (e.g., particle size, column length) and shows the resultant simulated chromatogram. Some of the applications of DryLab to HPLC separations can be found in references 11 and 12.

There are many other techniques for optimization and simulation. Reference 12 discusses many other approaches that have been used. The ones in this dissertation have been presented as a representative sampling of the ability to use computers for optimization.

the c

a pr

cond

further

culm

system

work

section

develo

drug s

DryLab

the sug

through

separati

behavior

also assi

reverse-p

Pe

column-s

through d



### ***Section 1.2.2      Suggestion of Initial HPLC Conditions and Automation***

One of the more exciting and complete uses of computers has been to assist in the choice of the HPLC method. Whereas optimization provides a "best separation" for a previously chosen method, this second use of computers is to provide the starting conditions (e.g., the choice of column, detector, derivitization schemes) which may be further optimized using the previously described computer techniques. The culmination of the integration of computers and HPLC is the expert system.<sup>3</sup> Expert systems will not be discussed in this overview section since they are the focus of this work and are described in greater detail throughout the rest of this dissertation. This section briefly describes the use of computers to automate HPLC instruments.

Colgan and Pollard,<sup>13</sup> evaluated Perkin-Elmer's PESOS as a tool to assist in the development and auditing of HPLC methods for the analysis of starting materials for a drug synthesis. PESOS is a unified program that includes DryLab and LC Analyst Expert. DryLab assists in optimizing the mobile phase and the LC Analyst component will run the suggested DryLab experiment. The automation component of PESOS allows a larger throughput in the number of experiments that can be run.

ProDigest-LC<sup>14</sup> is a computer program that is used to develop methods for the separation and purification of peptides and proteins. The program predicts retention behavior and simulates the elution profiles of 2-50 amino acid residues. The program also assists in the choice of columns (chosen from size exclusion, cation exchange, and reverse-phase LC).

Perkin-Elmer<sup>15</sup> has also developed an HPLC instrument with an integrated column-switching accessory. This system allows for rapid analysis of multiple samples through different columns.

Section

Once

frequent

on heu

combin

resultin

system

engine

enginee

system.

(e.g., th

inference

engine v

of the in

compute

informa

will ask

detail in

T

supporti

advice.

Two gen

### **Section 1.3     Expert System as an HPLC Tool—Overview**

To develop expert systems one must glean relevant information from an expert. Once this information is encoded, it can then be disseminated more easily and frequently.

The information, often of heuristic nature (see Chapter 2 for more information on heuristics), is obtained during knowledge acquisition meetings with an expert and is combined with information obtained from other sources (e.g., literature, data bases).

Once the information is obtained, it is then encoded into a proper format. The resulting encoded information is called a knowledge or information base. The expert system developer or knowledge engineer must determine an appropriate inference engine (whether it is commercial software or written developed by the knowledge engineer) which will be able to apply the knowledge from the knowledge base in a systematic manner, to infer, to reach a conclusion of the goal(s) of the expert system (e.g., the appropriate HPLC column to use for a particular separation). The choice of inference engine will dictate the format of the information encoding.

The knowledge must be organized within the expert system so that the inference engine will be able to follow a logical pathway through the knowledge base. The order of the information has little significance with respect to the inference engine since the computer will be able to search the complete knowledge base to find the relevant information. However, the information should be arranged so that the expert system will ask the appropriate questions in the correct order. This is discussed in greater detail in Chapter 4.

The knowledge engineer develops a user-interface to query the user for supporting data and evidence to be combined with the knowledge base to develop advice. There are different methods to exchange information to and from the user. Two general methods are Dialogic and "Smart Forms." Dialogic, the more traditional

met

answ

stru

"Sma

The

field

exper

includ

and s

appea

3.

## Section

inform

chemic

support

separati

P

and the

concerni

high tem

technique

because t

of the sa

homogene

method, involves presenting questions to the user and having the user respond with answers. This method is easy to implement and allows the knowledge engineer to structure the flow of the questions to the user. The second format, which is called "Smart Forms" in the expert system used for this project, involves a graphical approach. The knowledge engineer creates a screen that will contain different questions (i.e., fields). The user then may select a particular field (via mouse or keyboard) and the expert system will present choices for the user in some form of dialog box which might include drop-down menus, option buttons, and so forth. The user chooses a response and selects another field to continue entering more information. Often other fields will appear based on input to previous fields. These are discussed in more detail in Chapter 3.

#### **Section 1.4 Required Information to Implement the Expert System**

The expert system described in this dissertation requires several types of information from the user. This information can be characterized as physical and chemical properties of the analyte and sample. Both types of information provide supporting evidence for the proposed solution(s) to the goal of determining an HPLC separation method.

Physical information allows conclusions to be made on the sample preparation and the separation method for that analyte in a particular matrix. Information concerning properties such as boiling point and physical stabilities (e.g., explosive at high temperatures or pressures) is used to help determine whether HPLC is a suitable technique. The physical nature of the analyte might eliminate the need for HPLC because the analyte might be directly detectable without a separation step. Descriptors of the sample, including the presence (or absence) of particulate matter or non-homogeneous layers, allow the expert system to suggest sample treatment. The expert

system

up is ne

determ

separate

groups,

system

informa

whether

T

required

in the ge

is some

answer o

detail in

to empha

will say Y

called Fu

uncertain

ramificati

**Section**

Th

several dif

system is able to suggest what part of the sample to analyze and whether sample clean-up is needed. Finally, solubility, molecular weight, and shape of the analytes are used to determine actual separation methods.

Chemical information is used predominately to provide a basis to resolve and separate the sample. The user must provide information on polarities, substituent groups, charge transfer characteristics, aromaticity, and so forth to allow the expert system to propose, by induction, a way to separate the components. In addition, this information can be used to determine whether a particular detector is required or whether a particular compound might irreversibly bind to a specific stationary phase.

The required information can be given in relative terms; that is, the user is not required to know actual polarities, boiling points, or so forth. Most questions are asked in the generalized form, "IS YOUR COMPOUND <PARAMETER>" (where PARAMETER is some physical or chemical characteristic). The user may simply provide a yes/no answer or may provide an answer weighted by confidence factors (described in greater detail in Chapters 2 and 4) to allow for either certainty/uncertainty on the user's part or to emphasis/de-emphasis an answer (e.g., "My compound is slightly polar, therefore, I will say YES with a weighting of only 70"). This "gray-scale" logic, or more accurately called Fuzzy Logic, allows the expert system to draw conclusions based on missing or uncertain data. The resulting conclusion(s) will also be weighted to reflect the ramifications of the fuzzy input.

## **Section 1.5      Output from the Expert System**

The present version of the expert system described in this dissertation provides several different results. Each module, or stage, may be by-passed by the user.

solution

that HP

(

(

These r

complex

(in part

which ca

(

provide

simple s

present

unsure

and par

1

informa

consult

informa

allow th

laborato

that lab

system



The first step is to determine the suitability of HPLC as a technique for the solution of a specific analytical problem. The approach is to assume the null hypothesis that HPLC is NOT a suitable technique since:

- (1) no separation is needed or
- (2) other techniques are either faster or simpler.

These two criteria were chosen with the idea that HPLC is both time intensive and complex. HPLC is recommended only if a separation is needed and if other techniques (in particular gas chromatography (GC) or supercritical fluid chromatography (SFC) which can provide faster results) are not suitable.

Once HPLC has been selected as an appropriate technique, the expert system will provide advice regarding sample preparation and treatment. This decision is fairly simple since the user should know whether particulates or inhomogeneous samples are present and if the analyte is present in the particulates or different layers. If the user is unsure of the location of the analyte, the expert system suggests sampling all the layers and particulate matter.

Finally, the expert system provides advice on appropriate columns based on information concerning the physical and chemical nature of the sample. During the consultation with the expert system, data bases are accessed to provide additional information, such as appropriate solvents and detectors. The use of data base will allow the expert system to make recommendations that can be tailored for a specific laboratory. In this fashion, generalized consultations can be made more specific for that laboratory (e.g., available columns can be stored in the data base and the expert system will then recommend specific columns rather than generic listings).

## Section

instrum

Addition

column-

extensiv

constitu

types of

develop

maintain

chromate

HPLC pa

classes a

injected

T

injection

followed

version c

sample c

collector

since less

advantage

seminal p

Th

modifiable

hardware s

## **Section 1.6      Automation as an HPLC Tool**

The second part of this research involves developing an automated HPLC instrument. The instrument assists the user in developing and running HPLC methods. Additionally, the final objective of this project was to develop a system to perform column-switching. Many samples are quite complex. Complex samples may require extensive sample preparation and method development to be able to separate all the constituents in a sample. Often, complex samples will also contain different classes or types of components. To completely separate all the components, it is necessary to develop a method that will provide separation of the components within a class and still maintain the initial fractionation of the classes. With traditional one-column chromatography, separation may require extensive time to optimize solvents and other HPLC parameters. Another approach is to fractionate the sample into the different classes and collect these fractions off-line as they elute. These fractions can then be re-injected onto more specialized or different columns to provide additional separation.

The approach used in this research was to develop this fractionation and re-injection scheme into a single system involving a preliminary "fractionating" column followed by selectively directing the fractions onto other columns to provide an on-line version of the manual system described above. This on-line approach minimizes sample consumption, decreases analysis time (since there is no need to have a fraction collector and then have to re-inject the samples), and increases accuracy and precision since less sample handling will be involved. Majors<sup>16</sup> recognized and presented the advantages of on-line separations as compared to an off-line approach in an early seminal paper.

The research presented in this dissertation provides a very flexible and modifiable system to perform the multi-column separations. The software and hardware system described in the subsequent chapters will allow the expert system to

sugge:

instrui

monito

based

compoi

critical

are re-d

develop

monitor

First, co

coordinat

able to c

solvent c

(at least

utilized

instrume

and repr

methods

Fourth,

simple c

method

develope

are used

linked, a

system i

suggest columns and mobile phases. The advice may then be run on the automated instrument to provide the desired intelligent, automated system. The software monitors the response from multiple detectors and control the valve configurations as based on the elution of samples. Complete automation of all the HPLC instrumental components is the first requisite in developing this intelligent instrument.

To achieve the benefits of on-line fractionation and further separation, it is critical to be able to monitor and to control accurately and precisely when the fractions are re-directed onto additional columns. An essential part of this research was to develop an automated HPLC instrument that could provide column switching and monitoring capabilities. Computer control of the instrument provides several benefits. First, computer-controlled valve switching is accurate and reproducible and can be coordinated with events such as a detected peak or a specific time. Second, the user is able to control many different events—such as valve changes, data acquisition, the HPLC solvent delivery system parameters such as mobile phase composition—simultaneously (at least very rapidly on the HPLC time scale). This allows very complex methods to be utilized without an increase in complexity to the user. Third, computer control of the instrumentation, allows the user to more easily audit the HPLC method for ruggedness and reproducibility since each HPLC event, such as a valve change, can be stored in a methods file along with a date/time stamp and a user-provided comment section. Fourth, the solvent delivery system used in this project (Spectra Physics 8700) is of simple design and can only contain seven different events per method file and ten method files total. This limit is easily reached when either a complex method is developed or when different columns (and therefore different parameters per column) are used in column-switching mode. While two or more of these method files may be linked, a person must be present to link the next file at an appropriate time when the system is not under computer control. Allowing the computer to program the methods

and to

addition

solvent

instrum

plans i

other p

and da

of the

develop

## Section

HPLC in

setup.

Microso

acquired

program

the stud

T

above to

prototype

valve (CS

were buil

allow rapi

and to monitor which method is being used allows unattended linkage of these files. In addition, the computer can store additional methods in memory and then program the solvent delivery system with new methods after old methods have been executed.

Finally, an automated system facilitates the development of an intelligent HPLC instrument that will be able to optimize the separation (i.e., time and resolution). Future plans include having the HPLC controller program return the results of the HPLC run to other programs (which might include the expert system, solvent optimization programs, and data manipulation programs). These programs can then assist in the interpretation of the run and provide feedback to the controller program for further methods development.

## **Section 1.7 Automating the HPLC**

Chapters 3 and 6 describe the equipment and circuits used to automate the HPLC instrument. A brief introduction is presented here as an overview of the HPLC setup. Initial automation of the instrument involved writing a controller program in Microsoft Quick Basic, version 4.0 (Microsoft Co. Redmond, WA). This program acquired data from the HPLC detector via a data acquisition board. In addition, the program controlled two six-port valves. The first version of the program was used for the study of recycle heartcut chromatography (RHC) and backflushing.

The second stage of automation was the extension of the software described above to control additional valves and a second interface board. An in-house built prototype board was developed to allow the program to control a six-column selection valve (CSV) and to program the HPLC solvent delivery system (SDS). Several circuits were built to interface the prototype board to both the SDS and the CSV. These circuits allow rapid and computer-independent control of the CSV.

T  
more det

into the s

T  
column

The soft

from a

re-direc

program

each f

as sol

is de

switc

and s

run

auto

a se

the

pro



The software was also extended to allow sequential data acquisition from two or more detectors connected to the RTI board. An assembly level routine was incorporated into the software to allow the scanning of a smaller (user-selected) number of channels.

The software is designed to control a series of experiments that would involve column switching. This is accomplished by designing an overall method (the "master"). The software assists the user in establishing the number of fractions that will be eluting from a preliminary column. The user then indicates which fraction, or peak(s), will be re-directed from the fractionating column to more selective columns. A master program might specify:

- A C8 column serves as the fractionating column.
- Peak one is switched to a C18 column for further separation of alkanes.
- Peak 2 is switched to a phenyl column to separate aromatics.
- The next 10 minutes of the run will be directed to waste (this might be a section of the chromatogram where there are no peaks of interest).

After programming the master method, the user then establishes a method for each fraction. It is at this stage where the user will program the HPLC parameters (such as solvent composition, flow rates, and so forth). At the appropriate times (when a peak is detected or the correct amount of time has elapsed), the controller software will switch valves, change HPLC parameters, by-pass columns to introduce a new solvent, and so forth.

Future development of the software will include automated optimization of the run via an appropriate algorithm (e.g., simplex or factor analysis). This further automation of the optimization stages could be accomplished by feeding the results to a second expert system that would evaluate the run and suggest further refinements to the method (or to use another method that the first expert system might have proposed).

## LIST OF REFERENCES

- 1 S.  
Pr  
D  
Li  
Pu
- 2 J.  
Wi
- 3 J.  
Hig  
De  
R.
- 4 H.
- 5 E.  
557
- 6 I. M.
- 7 A. C.
- 8 John
- 9 L. R.
- 10 J. W.
- 11 N. G.
- 12 Com  
Chro  
The
- 13 S. T.
- 14 C. T.  
Chro
- 15 A. T.

## LIST OF REFERENCES

- 1 S. S. Williams, J. F. Karnicky, J.-L. Excoffier, and S. R. Abbott, 'Expert System Program for Assistance in High Performance Liquid Chromatographic Method Development' in *Computer-Assisted Method Development for High-Performance Liquid Chromatography*, J.L. Glajch and L.R. Snyder, eds., Elsevier Science Publishers, 1989, 267.
- 2 J. C. Berridge, *Techniques for the Automated Optimization of HPLC Separations* Wiley-Interscience Pub. Chichester, Great Britian,1985, 1.
- 3 J. L. Glajch and L. R. Snyder, 'Computer-Assisted Method Development for High-Performance Liquid Chromatography' in *Computer-Assisted Method Development for High-Performance Liquid Chromatography*, J. L. Glajch and L. R. Snyder, eds., Elsevier Science Publishing Co., Inc., The Netherlands, 1990, ix.
- 4 H. A. H. Billiet and L. De Galan, *J. Chromatogr.* 1989, 485, 27.
- 5 E. P. Lankmayr, W. Wegscheider, and J. Daniel-Ivad, *J. Chromatogr.* 1989, 485, 557.
- 6 I. Molnar, R. Boysen, and P. Jekow, *J. Chromatogr.* 1989, 485, 569.
- 7 A. G. Wright, A. F. Fell, and J. C. Berridge, *Chromatographia* 1987, 424, 533-40.
- 8 John C. Berridge *J. Chromatog.* 1989, 485, 3.
- 9 L. R. Snyder, J. W. Dolan, and D. C. Lommen, *J. Chromatogr.* 1989, 485, 65.
- 10 J. W. Dolan, D. C. Lommen, and L. R. Snyder, *J. Chromatogr.* 1989, 485, 91.
- 11 N. G. Mellish, *LC-GC* 1991, 9(12), 845.
- 12 *Computer-Assisted Method Development for High-Performance Liquid Chromatography*, J. L. Glajch and L. R. Snyder, eds., Elsevier Science Pub. Co., The Netherlands, 1990, pp. 65, 91, 183, 421, 533, 569, 581, 597, 607, 657.
- 13 S. T. Colgan and E. B. Pollard *LC-GC* 1991, 9(11), 772.
- 14 C. T. Mant, T. W. Lorne Burke, N. E. Zhou, J. M. R. Parker, and R. S. Hodges, *J. Chromatogr.* 1990, 485, 365.
- 15 A. T. Rhys Williams *American Lab.* 1992, 24(14), 46.



- 16 R. Majors, *J. Chromatogr. Sci.*, 1980, 18, 571.

Section

as a su

driving

making

Intrinsic

modify

static in

Von N

demonst

manipul

Neuman

program

structure

## **CHAPTER 2 HISTORY AND THEORY**

Where a calculator on the ENIAC is equipped with 18,000 vacuum tubes and weighs 30 tons, computers in the future may have only 1,000 vacuum tubes and perhaps weigh 1 1/2 tons.

Popular Mechanics, March 1949

### **Section 2.1 Artificial Intelligence**

To understand expert systems, it is necessary to understand their development as a sub-field of artificial intelligence (AI). By understanding the history of AI, the driving factors and concerns that led to the creation of tools that aid in human-decision making can be understood.

AI has been defined as:<sup>1</sup>

"The scientific discipline which attempts to endow computers and computer-controlled machinery with the ability for actions, which, if carried out by a human being, would be thought to require intelligence"

Intrinsic within this definition, a computer with artificial intelligence should be able to modify itself in a dynamic manner to changing conditions; otherwise, the program is static in nature and not intelligent. A different approach is used in AI in conventional Von Neumann programs (e.g., FORTRAN and BASIC). Although researchers demonstrated the potential to use computers as efficient calculating devices to manipulate dimensioned, defined data structures, computers were not "intelligent." Von Neumann programs require complete data input; missing parameters would cause the program to abort prematurely. These languages were also unable to deal with abstract structures or concepts. Von Neumann programs are goal-oriented and have discrete



bound

numb

Some

be a p

solve p

accept

solution

values)

from o

machin

not hav

more v

program

playing

*symbolic*

develop

"

h

c

T

structure

Through

complex

intelligen

boundaries. The program is written to accept only pre-defined data such as facts and numbers. The program then applies fixed algorithms to the data to derive solutions. Some slight variation in the allowable range is achievable; however, the range still must be a pre-defined type.<sup>2,3</sup>

Contrasted with Von Neumann programs, AI programs exhibit the capability to solve problems that might be outside the original realm of the defined goal. They can accept incomplete or missing data and may still be able to arrive at a solution. This solution might be from many different domains (e.g., numerical, logical, or symbolic values). By dealing with both incomplete data and data domains which may be different from one application to the next, AI has the potential ability to learn.

As mentioned above, initial uses of computers were as high-speed computing machines.<sup>4</sup> The role of the computer as a symbol processor was ignored since they did not have the processing power to do symbolic manipulation. As computers became more wide-spread, scientists started creating "interesting programs" or "smart programs," as they were referred, that solved non-computational problems such as playing games, proving theorems, and so forth. Artificial intelligence must be able to do *symbolic programming*; that is, it must be able to deal with concepts. Simon, in 1981,<sup>5</sup> developed the *Physical Symbol Hypothesis* which states that:

- "a. Thoughts correspond to language.
- b. Language can be captured in symbols.
- c. By manipulating symbols on a computer, one can simulate the thinking process."

The first attempts at artificial intelligence involved manipulating computer structures called atoms. These atoms were binary representations of concepts. Through symbolic programming, these atoms would then be manipulated to form more complex structures, such as lists, sentences, and so forth, to emulate human intelligence.<sup>6</sup> By using atoms to build more complex units, symbolic programming can

be use

new ex

manip

action.

memor

progres

decreas

discove

and rep

later, t

role—th

requisit

comput

dissemin

concern

unable t

knowled

the exper

the infor

acquisiti

they were

Th

during th

the ideas

John McCa

be used with unpredictable dimensions, complex and/or mixed structures, and to create new expressions from previously defined atoms. Once the computer became a symbol manipulator, it now"...has the necessary and sufficient means for generating intelligent action."<sup>1</sup>

Although this hypothesis proposed more serious research in AI, limitations in memory and in the speed of the central processing unit (CPU) seemed to impede progress in developing AI programs. As CPU speed increased and memory cost decreased, researchers expected AI to come into fruition. However, researchers discovered that the principal obstruction in developing AI applications was in obtaining and representing the knowledge required to solve that problem. As will be discussed later, this obstacle, or bottleneck, led to the development of a new occupational role—that of the knowledge engineer. The knowledge engineer's role is to elucidate the requisite information and to encode efficiently this knowledge into an appropriate computer format. Human experts were often found to be unwilling or unable to disseminate the information.<sup>5,7,8</sup> Often this unwillingness came from the fear or concern that she/he would be replaced by the computer. At other times, the expert is unable to articulate the manner in which she/he approached the problem. The knowledge engineering field has developed methods<sup>7</sup> to extract this information from the expert. These methods involve developing appropriate questions that would extract the information from the expert. By developing reassuring and efficient knowledge acquisition sessions, the experts' fears were allayed and the experts were affirmed that they were a vital part of the process.<sup>7,8</sup>

The official inception of AI as an accepted discipline has been assigned to 1956 during the *Summer Research Project* at Dartmouth College.<sup>5</sup> This project consolidated the ideas of many different researchers regarding AI. It was during this period that John McCarthy coined the term "Artificial Intelligence."

Most fundamental research in AI has been in trying to mimic the human brain. One attempt has been to create computer models (in software) based on the human neural network.<sup>9</sup> Other approaches have involved hardware approximations of the human brain via interconnecting processors that simulate human neural connections.<sup>10</sup> One limitation is that the number of processors and connections needed to approach the number in the human brain is beyond both the current scientific and economic practicalities.

Presently, AI has developed several sub-classes beyond the original focus. These sub-classes include:

- *Machine Vision*, which attempts to recognize and identify different patterns,
- *Robotics*, which focuses on producing mechanical devices that are able to mimic human actions (e.g., movement),
- *Speech Processing*, which attempts to both speak and recognize human language,
- *Pattern Recognition*, which tries to arrange and organize objects into logical patterns,
- *Machine Learning*, which aims to increase incrementally the computer's knowledge by observing examples,
- *Expert Systems*, which provides advice on a particular application by the use of encoded information.

## **Section 2.2      Background on Expert Systems**

One particular sub-class of AI is *Expert Systems*. The field of expert systems is the branch of AI that is focused on general problem-solving. This field is fundamental in nature and application-oriented than the other AI sub-fields mentioned above. Expert Systems can be defined as programs that solve problems that seem to require human knowledge. It is important to have a concise and complete definition of expert systems.

Such a definition will allow one to consider the required aspects of expert systems. An expert system has been defined as:

...a program that solves problems competently at an expert level in a particular domain. The system uses a knowledge base, which is separate from the main part of the program (the part that does inferencing using the knowledge base). The program must be capable of dealing with incomplete knowledge and should be able to explain its reasoning and conclusions.<sup>11</sup>

Expert systems were not a specified objective by AI researchers. Expert systems arose from the realization that computer hardware and software limitations precluded solving extremely complex, global problems. Researchers discovered that solving small-domain problems was feasible. Since these types of problems required limited knowledge, as compared to global problems, these applications alleviated the problem of storing massive amounts of data. Also, problems in small domains lead to easier elucidation of information from the human experts. Knowledge engineers discovered that these small problems were often solved by a heuristic approach rather than using complex algorithmic methods (for example, choosing an appropriate tax-preparation method versus fluid dynamic calculations for the design of airplanes).

A heuristic approach is defined as

Of, relating to, or constituting an educational method in which learning takes place through discoveries that result from investigations made by the student.<sup>12</sup>

Heuristics are used either to solve problems directly or to minimize the search field that must be considered when choosing an appropriate answer or response. Heuristics are often based on empirical data and previous experiences of the expert. This approach does not preclude heuristics from being linked to fundamental or mathematical theory. Often the process of obtaining and encoding these heuristics will lead to fundamental research. Since the encoding requires that the heuristic information be put into some

form of computer format such as IF-THEN statements (see later), a link between the proper heuristic response and experimental data must be made. Thus, the knowledge engineer and human expert must discover or determine the fundamental reasons for which the heuristic approach works. For example, if a particular reagent seems to cause decomposition, this information must be encoded with emphasis on the experimental conditions under which this reaction is observed.. This thought process often leads to asking "why" types of questions rather than a lower level question of simply "when" or "what."

In particular, the development of expert systems has helped to formulate rules that describe chromatography. While the knowledge base for chromatography is still incomplete, the process of formulating these rules has led to an increase in the understanding of the science of chromatography. To be able to incorporate HPLC rules, one must understand the fundamentals of chromatography.

An example of incorporating HPLC rules would be when an expert states "I choose a cyano column for this particular set of conditions." This statement does not provide the type of information that can be encoded into computer format. The expert must be queried by the knowledge engineer to determine what conditions (and why) must exist to support a particular conclusion. If the engineer only obtains information concerning the conditions under which the decision is made, then the knowledge engineer has only developed a data base that links a particular column with a set condition. The objective is to develop a more generalized method that is able to extend the decision to other conditions where a particular column or method is appropriate. To do this, the engineer must obtain the "why" and not just the "what" or "when." These answers will allow the expert system to work in a situation that has not been "hardwired" into the knowledge base. Being able to draw inferences and extensions, the program appears "intelligent."

Extracting the heuristics, or knowledge acquisition, has often been the major bottleneck in developing expert systems.<sup>13</sup> Knowledge acquisition is "the transfer and transformation of potential problem-solving expertise from some knowledge source to a program."<sup>13</sup> Often human experts have not externalized their thought process or are reluctant to do so.<sup>8</sup> Knowledge acquisition meetings intrinsically have value other than simply obtaining the input for the knowledge base. One objective of developing an expert system is

"...to codify and thereby render explicitly the valuable knowledge that many experts possess ...."<sup>13</sup>

This often will lead to a more fundamental understanding of the problem area.<sup>14</sup> In addition, since expert systems systematically search ALL of the knowledge base for possible solutions to a problem (much faster than a human), they often exhibit a success rate approaching, and sometimes surpassing, that of human.<sup>5</sup>

## **Section 2.3 Terminology**

The definition of expert systems used previously<sup>11</sup> introduced several terms that need to be defined. An expert system is composed of three basic components:<sup>15,16</sup>

- the knowledge/fact base,
- the inference engine, and
- the man-machine interface.

These terms are described in greater detail with respect to the implementation of the expert system developed in this project (see Chapter 4).

### ***Section 2.3.1 Knowledge and Fact Bases***

The first consideration is the knowledge and/or fact base. Since knowledge is "... information about the world that allows an expert to make decisions,"<sup>5</sup> it is essential



that information is not only available to the computer but that the information can be expressed in the proper syntax and structure before an expert system can be built.

There are two different types of knowledge: facts and rules.<sup>5</sup> Facts are asserted as being true (or as having a particular value). The values of the facts do not change during a consultation with an expert system regardless of the dynamic nature of other information in the expert system. Some examples of facts might be:

### Example of Facts

**Maximum allowable pH = 12**  
**Maximum Pressure = 6000 psi**

Code 2.1

Facts are statements that usually are specific to the application . Facts are often transitory from one consultation to the next; although, during a particular consultation they are consistent). Facts are analogous to CONSTANTS in traditional programming languages and are the fundamental information in an expert system since the results of rules originate from the synergism of the facts to form conclusions. Facts can be obtained from the knowledge base, or supplied by the user during a consultation with the expert system.

The second type of knowledge, rules, contain the knowledge, or heuristics, of the application domain. Rules relate the facts that have been obtained with particular conclusions.. Rules may be stored in a form called production rules (e.g., *IF-THEN* format); although, this type of knowledge is not limited to IF-Then format. The term, RULES, will be used to denote information that uses facts to reach a particular goal or conclusion. The value of a rule is dependent upon the values of facts and other data; therefore, they are very dynamic in nature. The example that follows shows one rule that indirectly determines a separation mode for an HPLC method. Here, the rule is ascertaining whether a compound of interest is ionic, non-ionic, or is ionic but can be shifted to a non-ionic form by adjustment of the pH.

IF

TH

To

suc

min

logic

data

solut

acce

subse

is an

to ha

infe

an exp

### Example of Rule Structure

```
IF    Ionic = Yes AND  
      Required_pH_To_Neutralize < Min_pH OR  
      Required_pH_To_Neutralize > Max_pH OR  
      Detect_Nonionic_Form = No  
THEN Ionicity = Ionic
```

Code 2.2

To determine if the compound is ionic, the rule compares facts obtained from the user, such as the pH required to neutralize the compound, —to internal facts such as the minimum and maximum pH allowed for an HPLC instrument/experiment.

A special type of rule is called a *meta-rule*.<sup>5</sup> Meta-rules are used to control the logic flow in an expert system. The meta-rules might determine which particular database(s) to load based upon the user's input. Although meta-rules do not lead to the solution of the problem addressed by the expert system, they allow the expert system to access the correct information to solve the problem.

An example of a meta-rule is provided in Code 2.3. This rule determines the subsequent goals of the expert system. When the expert system determines that HPLC is an appropriate technique for the particular application and the user indicates a desire to have the expert system propose a method(s), this specific meta-rule will cause the inference engine to determine sample pre-treatment, column, and solvent.

Each type of rule has several different methods by which it can be represented in an expert system. Some of these different formats are discussed in Section 2.6.

### Example of a Meta-Rule

```
Rule Continue2
IF    HPLC_Suited = Yes AND
      Check_For_Methods = Yes THEN  Continue =
      Yes
      FIND Scuzz_Checked
      FIND Homogeneity_Checked
      FIND Rev_Phase_OK
      FIND Normal_Phase_OK
      FIND Mode
      FIND Column
      FIND Solvent;
```

Code 2.3

A unique feature of expert systems is the association of confidence factors (CNF) with facts and rules. Since expert systems must deal with uncertainty, they must be able to associate a numerical value to the certainty or validity of a fact or result. A CNF of 100 asserts absolute validity of the fact and a CNF of 0 states that a fact or result is absolutely false. All values between 0 and 100 indicate a graduation between absolute certainty and absolute falsity. As will be discussed later, confidence factors are distinctly different from statistical probabilities; rather, CNFs are a sub-field of fuzzy-logic.

#### **Section 2.3.2      *Inference Engine***

The second, and major, component of an expert system is the inference engine. An analogy can be made between the inference engine of expert systems and the compiler of traditional programming languages. The inference engine is the component that derives solutions to the goals of the expert systems based upon the linking, or *inference*, of all the rules, facts, and user input.

This linking is accomplished by two modes: forward chaining and backward chaining. In forward chaining, all rules are compared with the available facts to generate an initial set of conclusions. These conclusions are added to the fact base to

be used by the inference engine as it re-iterated through the knowledge base to generate a second set of conclusions. This process is iterated until a final, complete set of conclusions has been generated. At this point, the deduced conclusions are compared to the goals of the expert system to see if one, or more, of the conclusions is the solution to the objective. The forward chaining approach is often called "data-driven" since all the data are used to derive all the possible conclusions without any consideration for the goal that is being solved. Its principle use is in instrument control, planning, and scheduling.<sup>5</sup>

The second search strategy is backward chaining in which the inference engine first determines the goal(s) of the expert system. The inference engine searches through its knowledge base until it finds a rule that has a conclusion that matches the goal(s). Once such a rule is found, the inference engine attempts to determine the validity of the premise of the rule. If the premise is true, the rule "passes" or "fires." Backward chaining is goal-oriented And is used when only particular solutions are to be considered and not all the ramifications of the available data.

### ***Section 2.3.3      User Interface***

The primary function of the user interface is to obtain information from the user. Once the inference engine has induced a solution(s), the interface provides a method to present the results back to the user. The user interface also is responsible for the *Help* and *Explanation* components. The role of the *Help* facility is self-explanatory. The *Explanation* facility allows the user to query the system as to both WHY and HOW a solution was induced. These two features allow the user to determine the rationale behind a question that is being asked during a consultation. Once the user is aware of the rationale, she/he may supply an answer in the correct context of the

question. The *How* part of the facility not only allows the knowledge engineer to test and debug the system but also allows the user to learn from the system. The *Explanation* facility allows the knowledge engineer to extend further the knowledge base, and improves the user's confidence in the system when she/he can check the induction path of the conclusions.

## Section 2.4 Historically Important Expert Systems

The first expert system of note was the *DENDRAL* expert system that was developed in 1964 by Lindsay, et al.<sup>17</sup> NASA<sup>5,18</sup> provided funds to these researchers to develop a program that would interpret data obtained from mass spectra of Martian soil. The objective was to place both a mass spectrometer and the expert system onto a space vehicle.

Researchers limited *DENDRAL's* scope to the analysis of only simple monofunctional compounds. The program used a centroid (a specific atom or bond) and generated all possible structures by systematically arranging constituent radicals around the centroid. The predicted spectra were then compared to the experimental spectrogram. A match thus identified the compound. Obviously a near-infinite search area would be created from the generation of all possible structures around the centroid. The expert part of the system was the use of rules that limited the number of possibilities that would be generated to only those that were logically plausible. These rules were derived from heuristics supplied by human experts or from an associated program, *meta-DENDRAL*, which analyzed the mass spectra of classes of compounds and formulated probable fragmentation rules based upon the spectra.

Duffield, Schroll, and Bucks<sup>19</sup> validated *DENDRAL* and determined that general heuristics alone were not powerful enough to solve this complex problem and specific information from a mass spectrometry expert was needed. Although *DENDRAL* often

provided less information than the equivalent human expert,<sup>13,20</sup> *DENDRAL* was useful for searching intelligently through all possibilities and limiting the results to meaningful solutions. Since this search can be accomplished much faster on a computer than by a human, a powerful tool was provided for the human expert.

The next major development in AI and expert systems was the work done by Slage and co-workers at MIT during the mid 1960<sup>13,19</sup> to develop an expert system, called *MACSYMA*, that could evaluate symbolic mathematical expressions. This work demonstrated that the program must encode not only the mathematical rules, themselves, but also the meta-rules that dictate when and how these mathematical rules are applied. *MACSYMA*, using symbolic form, could solve over five hundred distinct mathematical operations such as integration, differentiation, and Taylor series.

Shortliffe at Stanford in the 1970s developed the *MYCIN* program.<sup>5,13,19</sup> *MYCIN* was one of the first rule-based systems and its creation provided a direction to the further development of expert systems; in particular, the use of expert system shells that are described below. *MYCIN* provided recommendations for appropriate anti microbial regimens for bacterial infections. The physician could input the symptoms, causal organisms, and the drug sensitivity of the patient into the *MYCIN* program. For very ill patients, antibiotics must often be prescribed before the exact organism responsible for the infection is identified. This incomplete, or "fuzzy," data is a prime candidate for a heuristic approach. However, the system needs to provide some type of reliability or confidence for the solutions so that the physician can evaluate and choose the appropriate course of action. This led to the development of confidence factors, which are discussed in detail later in this chapter.

*MYCIN* was validated in blind field tests and shown to provide correct recommendations,<sup>13</sup> but several factors led to its "failure." Shortliffe lost the knowledge source after the resident expert left the project. In addition, *MYCIN* was implemented

on a heavily used mainframe computer which precluded fast consultations. If the consultation could not be accomplished in a fairly short session, users were not as inclined to use the system. As discussed later, the need for fast expert systems and fast computers was a concern in the research project described in this dissertation.

Before *MYCIN*, each expert system had been created primarily as fundamental research. Each researcher would completely write the code for the research project without using any shared code from other AI projects. The next step in expert systems was to extend the existing expert system framework. In 1980, van Melle<sup>19</sup> developed *EMYCIN* (where the "E" stood for Empty) which consisted of the *MYCIN* inference mechanism with the knowledge base for the infectious diseases removed. This approach preserved the work already accomplished and facilitated the rapid development of new expert systems by re-using:

- the component that generated the solutions based upon the input (the inference engine), and
- the man-machine interface.

*EMYCIN* still maintained its ability to interact with the user to provide advice and to provide also a feedback to the user to indicate how such advice had been generated. A new developer only had to install the appropriate knowledge base for a new problem into the *EMYCIN* framework.

This ability to build onto existing systems led to the development of "Inference Engines" or "Shells", which led to the implementation of modular systems.<sup>20,21,22</sup> *EMYCIN* was used to develop *PUFF*,<sup>23</sup> an expert system that used fifty-five rules to diagnose pulmonary dysfunctions. This approach was used to develop other instant expert systems such as *SACON*,<sup>24</sup> a structure analysis program. The knowledge obtained during the construction of *EMYCIN* was used to develop *Guidon*,<sup>25</sup> an



instructional expert system that teaches medical students about infectious diseases. *Guidon* demonstrates the applicability of expert systems as tutorial tools.

## **Section 2.5 Expert System Applications in Analytical Chemistry**

### ***Section 2.5.1 Non-HPLC Expert Systems***

Expert system applications in analytical chemistry can be divided into two broad types.<sup>14,26</sup> The first type is systems to interpret complex experimental data, such as those resulting from nuclear magnetic resonance (NMR), infrared spectroscopy (IR), or mass spectrometry (MS), or combinations of these techniques. The *DENDRAL* project, as mentioned earlier, is an example of this approach. The second type is systems that guide the analytical chemist in methods development. The examples of expert systems for HPLC applications that are described later (and the system developed for this dissertation) will demonstrate this second approach. Presented below are several different expert systems that have been described in the literature. These examples of expert systems are presented to demonstrate several of the characteristics of expert systems.

Settle<sup>27</sup> has developed several expert systems to aid general chemistry students in analyzing their samples. The expert system accepts physical and chemical information such as melting points, litmus tests, and so forth, to ascertain the probable components. These types of expert system applications demonstrate the tutorial potential of expert systems.

Betteridge, et al.<sup>28</sup> developed a system to aid in the correct determination of sampling points for water content of various petroleum streams on a North Sea oil platform. This expert system was implemented to maintain their current expertise in sampling and moisture analysis and to train new experts.



### **Section 2.5.2      HPLC Expert Systems**

Expert systems in chromatography provide the user with a systematic tool for methods development. Researchers in this area have used two distinct approaches to construct tools to assist in methods development. The first approach is to use a limited application such as a particular class of pharmaceutical compounds or a particular aspect of methods development such as the detection of a class of compounds. This approach provides a narrow scope with a highly specific knowledge domain which ensures successful solutions in real-time.

Fell, et al.<sup>29</sup> have developed an expert system that optimizes the mobile phase in an HPLC system with a photodiode array detector. The expert system uses data obtained from an initial gradient elution experiment and the spectral profiles of the components obtained from the photodiode array detector to generate a model of the mobile phase-solute interactions. The expert system and a simplex implementation<sup>30</sup> together propose subsequent experiments to perform for optimization using the model.

Other specific applications of expert systems for HPLC include:

1. selecting methods for pharmaceutical analysis,<sup>31,32,33</sup> and steroids,<sup>34</sup>
2. troubleshooting HPLC assay methods,<sup>35</sup> and
3. developing intelligent spreadsheets<sup>36</sup> to evaluate the precision of HPLC methods.

The second approach is to assist in developing HPLC methods for generic systems and applications. These expert systems involve more of a first-principle approach and use generalized HPLC heuristics. Abbott and co-workers<sup>37,38</sup> have written an expert system that develops methods based on the physical and chemical natures of general classes of compounds (*ECAT* — **Expert Chromatographic Assistance Team**). *ECAT* is the project most similar to the research described in this thesis. As

such, more detail will be provided to draw comparisons and differences between *ECAT* and this work.

*ECAT* consists of four modules. The Column and Mobile Phase Module determines the packing material, column dimensions, mobile phase, and modifiers. The Method Optimization Module optimizes the method based on resolution and total analysis time. The Sample Preparation Module determines if sample pre-treatment is needed to enhance certain aspects of the method (separation, resolution, delectability, etc.). Finally the Factual Information Module contains information on specific chemicals and classes of chemicals. The facts are stored in predicate logic format, which is a different format than production rules (PREDICATE SUBJECT [PROPERTY-1] [PROPERTY-2]). An example of predicate logic used to store acid dissociation constants in the form  $pK_a$  values, as used in the *ECAT* system, is: (has- $pK_a$  benzoic-acid 4.2). After the program is provided information about the analyte class, it then applies forward chaining to develop a more general identification of the class and properties of the analyte. If the compound is unknown to the module or database, fundamental properties are determined by asking the user to provide the information directly, rather than by attempting to infer the information from supporting data. Normal-phase and hydrophobic interaction modes are never selected by the rules while  $C_{18}$  bonded phase is chosen for most reversed-phase separations. *ECAT* also suggests  $C_4$  columns for membrane proteins, cyano phase for small molecules, and strong cation- and anion exchange phases for the separation of ionic species. Only seven rules effect the selection of a separation mode while ninety other rules are used to determine column and mobile phase parameters. Additives—such as buffers, salts, and competing acids/bases—are specified after other parameters have been determined. This limited number of rules to determine the separation mode demonstrates the difference between *ECAT* and the expert system described in this dissertation. The first principle approach used in the

expert system described in this dissertation applies more rules to elicit chemical and physical properties of the analyte, and the recommended method is based upon these properties. Since *ECAT* classifies the analyte into different chemical families and recommends a method that is associated with the chemical family, *ECAT* requires less rules for this part of the expert system.

Peichang and Hongxin<sup>39</sup> have developed an expert system that also incorporates advanced algorithms to suggest columns, mobile phases, and other factors based upon user-provided compound classes. Their system also generates a representative chromatogram of the analyte class or compound, which can assist the chromatographer to decide if the inference mechanism proposed a valid method. Peichang and Hongxin's expert system still requires information to categorize the analyte. The user may provide (1) the molecular structure, (2) commercial name of the sample, or (3) the analyte class, such as phenylthiohydantoin (PTH)-amino acid. This expert system proposes the molecular interactions based on the classification of the analyte but does not use a first-principle approach.

Other applications of expert systems for HPLC include work by Fox and Tischler<sup>40</sup> on methods development. Kateman, et al.<sup>41</sup> have developed an expert system that helps to determine the ruggedness of HPLC methods before expensive methods validation of costly compounds.

Present work with expert systems includes developing rule-building expert systems<sup>21,42,43</sup> that help to alleviate the difficulties of knowledge acquisition and the subsequent encoding of this information into computer format. Expert systems are being developed that are able to access conventional computer languages<sup>44,45,46</sup> and programs such as databases and spreadsheets.<sup>5,39,47,48</sup> The synergism between conventional programs and expert systems can provide a complete system where the traditional weaknesses of expert systems (e.g., graphics and numerical computations)

are supplemented by external programs, such as databases and algorithmic-based programs. Databases and spreadsheets provide a means for concisely storing facts that can be updated rapidly and easily.

## **Section 2.6 Implementation of Expert Systems**

Many attempts to develop expert systems have failed because of inadequate considerations of the requirements to develop a successful system.<sup>7</sup> An overview of these steps is presented below,<sup>7,8</sup> and are described in greater detail in Chapters 3 and 4 of this thesis. In particular, these two chapters will discuss the steps that were taken during the development of the expert system described in this thesis.

### ***Section 2.6.1 Goal Determination***

The first consideration is to determine the appropriateness of the problem to an expert system. If it is not amenable to expert systems, the project will fail.

The project should **not** be solvable by an algorithmic approach since this type of problem is most efficiently solved by traditional, numerical programs. Rather, the problem should be heuristic or symbolic in nature. The problem must be relevant so that interest and support can be generated for the project.

Once a relevant problem has been determined, an expert(s) must be available **who** has sufficient knowledge to aid in solving that problem. The information needed **to** solve the problem must be fairly limited in scope. A heuristic that can be stated is:<sup>49,51</sup>

**IF** the problem can be solved by an approximate half-hour telephone conversation with an expert, **THEN** an expert system approach is possible.

An appropriate expert system is one that requires a descriptive analysis of the problem without on-site analysis (e.g., solvable by telephone). A heuristic solution must be stated briefly and relatively simply so that the solution can be transposed into computer format. A valid conclusion can be reached if enough rules are applied to eliminate the invalid conclusions. The elimination of invalid conclusions limits the scope of the problem so that a knowledge base can be generated from which to infer answers. As the number of possible solutions increases, the inference engine can become sluggish and results with low confidence limits can be returned.<sup>16,50,51</sup>

After an appropriate application is chosen but before the expert system is encoded, the goals and the end-user of the expert system must be clearly identified. An expert system written for the use of technicians will require a different amount of *a priori* knowledge than a similar expert system written as a tutorial tool for a student. If an expert system is encoded without consideration of the end-user, the system might be "... so foreign that no end user will understand or be able to use it."<sup>7</sup> Identification of the end user prior to encoding the expert system may also dictate the type of user-interface to use; that is, whether the expert system uses a graphical interface to represent instrumentation or a data input format that mimics forms used by a certain professional field (e.g., medical charts).

## **Section 2.6.2      Knowledge Acquisition**

### **Section 2.6.2.1      Methods for Knowledge Acquisition**

There are several different viewpoints on the best approach to use to obtain the information.<sup>8</sup> Some approaches use a single human expert to maintain consistency in the knowledge base while others use multiple experts to minimize personal biases and to obtain a broader range of information. However, the knowledge engineer may often have to arbitrate between two, perhaps conflicting, pieces of heuristics that may come

from the experts. The knowledge engineer may also use additional sources of information such as references, books, and so forth; although, the knowledge engineer must be familiar with the area of expertise so that he/she is able to decipher and apply the information obtained from alternative sources.

Most knowledge is gained by consultation during knowledge acquisition meetings. These meetings are normally attended by the expert, the knowledge engineer, a "champion", as well as several end-users.<sup>7</sup> It is typical to include a trainee knowledge engineer as a member of the knowledge acquisition team in order to increase the number of knowledge engineers. Since only a selected few universities, such as the University of Michigan and the University of Sweden, teach the theory behind knowledge-based systems, and new knowledge engineers must be developed using an apprenticeship approach.<sup>7</sup>

#### Section 2.6.2.2      Personnel and Roles

As alluded to in Chapter 1, the knowledge engineer is the bridge between the expert and the encoding of the information. Initially, knowledge engineers were usually computer scientists proficient in an AI language such as *LISP* since expert system tools were very rudimentary and required such a programmer. With the advance of expert system tools that facilitate expert system development, it is often better to have the knowledge engineer, who may now be the expert, who is familiar with the area of development. This facilitates the development of the expert system since the engineer is now familiar with the language and science of the intended application.

The champion's role is to be not only the supporter of the project but to provide a broad background in several disciplines to facilitate the incorporation of information into the expert system. The presence of one or two end-users insures that the developed system is usable. Finally, the trainee knowledge engineer will be the person



who completes the project and maintains it after helping to develop the expert system prototype.<sup>7</sup>

### ***Section 2.6.3      Encoding***

#### **Section 2.6.3.1      AI and Expert System Languages**

After the initial knowledge is obtained, the knowledge engineer must choose the appropriate method to implement the expert system. Expert systems may be implemented by using an Artificial Intelligence language (such as Prolog or LISP) or by employing a commercial shell to create the expert system. Detailed comparisons of shells versus AI languages have been made elsewhere.<sup>8,52</sup> In general, shells provide a consistent method to develop fast expert systems; however, shells are not as flexible or as optimizable as an application-specific (customized) shells developed from an AI language.

The next two steps are to choose an appropriate shell (if this is the direction taken) and to formalize the knowledge into the appropriate format. Different commercial shells will allow one to encode the information in different formats. There are three major ways to encode the information. The first method is to use some type of rule format such as production rules or predicate logic. The second and third types of knowledge representation are Frames And Objects. Frames and Objects allow representation of information that incorporates characteristics, or attributes, of that object. Each frame or object exhibit inheritance with higher/lower hierarchies of frames/objects. Further discussion of frames and objects is beyond the scope of this dissertation. (For further information on these types of knowledge representations, see reference 5, pages 48-57, 140-141, and all of Chapter 5.)

### Section 2.6.3.2 Prototyping of the Expert System

The final steps in the development of an expert system involve the creation and validation of a prototype based upon the initial data. After the prototype is generated, the system is modified until a robust system is generated. After this stage the expert system can now be used and the only steps left are the maintenance stages. During the usage of the expert system, new information might be incorporated into the system. Since the knowledge base is modular in nature (i.e., separate from the inference engine and user interface), it is fairly simple to incorporate new information into the system. If external knowledge bases have been created (e.g., databases, spreadsheets, and text files), this information can often be incorporated through a third party program (a database program, for example) without having to change the existing rules/frames of the expert system.

## Section 2.7 Confidence Factors

Expert systems are able to deal with facts that have varying degrees of confidence or validity that may be dependent upon a particular time of use or on the belief system of the different users. As such, inclusion of some numerical value must be considered that reflects the certainty or faith in these facts and the subsequent certainty of the conclusions based upon these facts.

### *Section 2.7.1 Comparison of Probability and Possibility*

Probability is based upon the relative frequency distribution of some event(s) over a long enough period to obtain a representative distribution. A formal definition of probability is:

The chance that a particular event will occur is equal to the number of ways the event can occur divided by the number of ways all possible events (including the particular event of interest) can occur.<sup>5</sup>

This classical, numerical approach to determine the certainty of an event is very appropriate for systems where the causality, or model, is well understood (e.g., gambling) or systems that have sufficient statistical records that exist such that the certainty of an event approaches the numerical probability of said event. If these criteria are met, probability theory is a powerful tool to assign a numerical value to the validity of the facts in an expert system.

Expert systems are usually developed in areas that are not amenable to numerical, algorithmic solutions. Often these applications involve heuristics that do not have explicit causalities, and/or applications that have non-formalized empiricisms. Other applications may lack sufficient records such that a statistical treatment of the data is not possible and therefore precludes using probability theory.

There are also two other concerns that have tempered the use of probability theory in expert systems. The first is that probability theory does not include the prior belief systems of the user. The second problem is that the certainty of an event decreases with supporting evidence. For example, :

**IF      CONDITION 1 AND  
          CONDITION 2  
THEN CONCLUSION**

If **CONDITION 1** and **CONDITION 2** have a probability of being true of 0.80 (that is,  $P(\text{Event})=0.80$ ), **CONCLUSION** has a certainty of 0.64. The **AND** function indicates that the two probabilities should be multiplied together:  $P(\text{CONCLUSION})=P(\text{CONDITION 1}) * P(\text{CONDITION 2})$ . If this analogy were extended to a rule with five conditions in the premise, each condition having a  $P(\text{Event})=0.80$ , then the conclusion would have a certainty of 0.33 ( $P(\text{Conclusion}) = (0.80)^5$ ) which means that the conclusion is false. This problem also occurs with **OR** situations (e.g., IF A OR B THEN

C). Probability theory does not allow the certainty of C to increase if both A and B are true.

### ***Section 2.7.2      Fuzzy Logic***

This problem has led to the development of Bayesian and other alternative approaches to deal with certainty.<sup>5,7</sup> The Bayesian definition of probability states:

A probability is a degree of belief held by a person about some hypothesis, event, or uncertain quantity.<sup>5</sup>

This definition allows subjective evaluation of a user's belief in the truth of a particular response and allows the user to incorporate the prior success of a solution along with the current evidence to deduce the result.<sup>5</sup> The extension of Bayesian logic has led to the development of fuzzy logic. Fuzzy logic is based on set theory; that is, the certainty of a particular event is defined as a membership function of a certain set of circumstances. Fuzzy logic allows vague terms to be used to describe such an event. An example of fuzzy logic would be the statement, "I am 80% certain that a Cyano column is the appropriate column." Probability theory would state that there is an 80% chance that a cyano column is the correct column and the uncertainty is based upon the error of the data used to derive that conclusion. In fuzzy logic, this would translate into cyano columns having a degree of membership in the "CORRECT COLUMN" set of 0.80. The uncertainty is an inherent component of the choice of a column and does not reflect the uncertainty of the data used to determine the correct column.<sup>5</sup> The sum of all the certainties in fuzzy logic does not have to add up to 1.00 as it must in probability theory. This allows a subjective evaluation of fitness for every event regardless of other events.

Fuzzy logic can be stated in mathematical form as follows:<sup>5,53</sup>

1.      The intersection of events A and B yield a conclusion with certainty of

$$P(\text{Conclusion}) = \text{MIN} [P(A), P(B)] \quad \text{Equation 2.1}$$

Thus, in a rule which states: **IF A AND B THEN C**, the confidence in C would be limited by the minimum value of the certainty in premise A and B.

2. The union of events A and B yields a conclusion with certainty of

$$P(\text{Conclusion}) = \text{MAX}[P(A), P(B)] \quad \text{Equation 2.2}$$

In logical syntax, **IF A OR B THEN C**, the conclusion, C, would have a certainty equal to the maximum certainty of A or B. Since the certainty of C depends only on either A or B being true, the premise with the highest degree of support for the conclusion determines the certainty of the conclusion.

### **Section 2.7.3      Certainty Theory**

These two rules of fuzzy logic insure that a conclusion's certainty does not approach zero as the number of premises increases. However, fuzzy logic does not provide for an increase in certainty in conclusion C if both A and B are true for the union of A and B. The **OR** statement is mathematically handled by Equation 2.2 which limits the certainty of the conclusion to the most reliable premise. The solution to this problem is an extension of the logic to form a sub-division of fuzzy logic, called certainty theory. The mathematics of certainty theory are presented below. Certainty theory also incorporates Equations 2.1 and 2.2.

1. For a rule in the format where **CNF a** and **CNF c** signify the confidence of the premise and rule, respectively: **IF A CNF a THEN C CNF c**, the certainty of the conclusion is the product of the certainties of the premise and the rule:

$$\text{CNF}(\text{Conclusion}) = \text{CNF c} * \text{CNF a} \quad \text{Equation 2.3}$$

2. Simple logical **AND** statements are used as in Equation 2.1, so that the confidence factor is the minimum confidence factor of all premises.
3. For logic combinations involving the **OR** function:

$$\text{CNF}_{\text{Total}} = \text{CNF}_1 + \text{CNF}_2 - \text{CNF}_1 * \text{CNF}_2 \quad \text{Equation 2.4}$$

where  $CNF_1$  and  $CNF_2$  are the individual components of the hypothesis and are calculated via Equation 2.3 above. For example,

IF        A CNF a OR  
          B CNF b  
THEN C CNF c

**CNF<sub>1</sub> and CNF<sub>2</sub> would be the certainty of the conclusion is the rule was written into two separate rules with separate premises A and B:**

IF A CNF a  
THEN C CNF c

|      |         |
|------|---------|
| IF   | B CNF b |
| THEN | C CNF c |

The values for CNF<sub>1</sub> and CNF<sub>2</sub> would be determined from  $P(\text{CNF}_1) = (\text{CNF a}) * (\text{CNF c})$  and  $P(\text{CNF}_2) = (\text{CNF b}) * (\text{CNF c})$ . The subtraction term accounts for the fact that the CNF<sub>1</sub> and CNF<sub>2</sub> are being added.<sup>5</sup> The total confidence factor can be larger than the individual confidence factors to allow for an increase in certainty when both parts of the supporting evidence are true.

4. For complex logic combinations involving AND and OR, the rule is re-written into two, or more, parts involving the AND. The CNFs of the two rules are calculated using Equations 2.2 and 2.1 above. Then the two CNFs are combined using the OR function of Equation 2.4 above. A calculation demonstrating this definition is given below for a rule that contains two conclusions:

**IF            A CNF 80 AND  
              B CNF 70 OR  
              C CNF 90  
THEN       D CNF 90  
              E CNF 70;**

**First, this rule has to be re-written into two components separated at the OR function:**

## Part 1

## Part 2

|                                                                                                                                  |                                                                                                                                  |
|----------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------|
| <p><b>IF</b>      A CNF 80 and<br/>                   B CNF 70<br/> <b>Then</b>    D CNF 90<br/>                   E CNF 70;</p> | <p><b>IF</b>      A CNF 80 and<br/>                   C CNF 90<br/> <b>Then</b>    D CNF 90<br/>                   E CNF 70;</p> |
|----------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------|

**Next, the appropriate equations are applied Part 1 and 2 (and converting the percent confidences into decimal equivalents):**

**Part 1****Part 2**

|                                        |                                        |
|----------------------------------------|----------------------------------------|
| $P(D) = \text{MIN}(0.80, 0.70) * 0.90$ | $P(D) = \text{MIN}(0.80, 0.90) * 0.90$ |
| $P(E) = \text{MIN}(0.80, 0.70) * 0.70$ | $P(E) = \text{MIN}(0.80, 0.90) * 0.70$ |
| $P(D) = 0.70 * 0.90 = 0.63$            | $P(D) = 0.80 * 0.90 = 0.72$            |
| $P(E) = 0.70 * 0.70 = 0.49$            | $P(E) = 0.80 * 0.70 = 0.56$            |

The next step is to recombine the two parts above into the original OR form and then calculate the final certainty using Equation 2.4.

**Rule****Calculation**

|                                                      |                                                               |
|------------------------------------------------------|---------------------------------------------------------------|
| IF     Part1 CNF 0.63 OR<br>Part2 CNF 0.72<br>THEN D | $P(D) = 0.63 + 0.72 - (0.63 * 0.72)$<br>$P(D) = 0.896 = 89.6$ |
| IF     Part1 CNF 0.49 OR<br>Part2 CNF 0.56<br>THEN E | $P(E) = 0.49 + 0.56 - (0.49 * 0.56)$<br>$P(E) = 0.776 = 77.6$ |

This example demonstrates that when both premises **A** and **B** are true, the resulting certainty of conclusions **D** and **E** is higher than the individual CNFs of **A** and **B**. Certainty theory provides the desired response of an increase in certainty as more data support the conclusions.

## **Section 2.8     Conclusions of Expert Systems and Certainty Theory**

The previous sections described the development of expert systems as a sub-field of AI. Expert systems provide a means to incorporate human expertise into computer format so that the information can be disseminated and used by a larger number of people. Expert systems use heuristics that are typically encoded into rule format and an inference engine applies either forward or backward chaining (or a combination of the two mechanisms) to propose a solution for some goal. Several

representative expert systems are described to demonstrate the capabilities of expert systems to assist in the solution of complex problems. Typically, expert systems are the most successful for small-domain areas where the information required to solve the goal is limited in nature and can be stated in concise, heuristic format. The expert systems that are typically used for HPLC applications are either specific to a small problem, such as the choice of method for the analysis of pharmaceutical compounds, or provide solutions for a larger, more generalized, domain. The expert systems described in the literature do not apply a first-principle approach. Rather, they typically categorize the sample into different classes of compounds, and then propose the method that has been associated with that class of compound. While this allows for very specific advice to be provided, the expert system described in this dissertation uses a first-principle approach (as described in Chapter 4). The expert system uses the sample's chemical and physical properties to propose an HPLC method. This approach allows the expert system to propose a method for any compound, regardless of its type or class, provided the user knows some chemical or physical property of the sample.

Confidence factors have been described in this chapter. Confidence factors provide a means for the user to indicate the reliability of the information that he/she provides to the expert system. The expert system can then use certainty theory (or some other form of fuzzy logic) to determine the certainty of the proposed methods. Most of the expert systems in the HPLC literature do not utilize confidence factors. Confidence factors have been incorporated into the expert system of this dissertation. As demonstrated in Chapter 4, confidence factors allow the expert system to provide a ranked response. The user may then select the appropriate method by considering the confidence factors of the different proposals.

The envisioned system of this research is an intelligent, automated HPLC instrument which has the capabilities to perform separations that may necessitate the



use of column switching. Column switching is a technique in which eluting components from a first column are re-directed onto a second column. Column switching provides the means to separate very complex samples by using two, or more columns, that are selective for different classes of compounds in the complex matrix.

The next section of this chapter describes the theoretical considerations of multidimensional HPLC. While the research described in this dissertation does not use multidimensional, an automated system has been developed to provide for this capability. It is envisioned that the expert system described in this dissertation will be expanded to provide a method(s) for multidimensional separations. The automation, via software and hardware, provides a means to automate the separation. The combination of automation and expert system provides the first generation of an intelligent, automated instrument.

## **Section 2.9      Multidimensional High Performance Liquid Chromatography (MD-HPLC)**

A simple heuristic can be postulated that relates society, analytical problems, and complexity: An increasingly complex, technological world creates complex problems that can be solved only by powerful (and hence, complex) tools and techniques. In a fairly recent survey of liquid chromatography experts,<sup>54</sup> Joseph Glajch states,

With the increasing complexity of samples encountered, especially in biological and environmental areas, the additional efficiency of newer LC columns will never be sufficient. Multidimensional approaches will lead to overall better resolution in shorter time.

### ***Section 2.9.1      Definition and Theory***

This thesis will use an extended version of J. C. Giddings's definition of multidimensionality.<sup>55</sup> Giddings defines multidimensional separations as separations

that must experience two or more orthogonal separative displacements. The separation must be developed such that the resolution achieved by the first displacement is not degraded by the subsequent displacement(s). This criterion eliminates purely tandem arrangements of columns since the resolution achieved by the first column can be completely eliminated by the second column. Multidimensionality can be sub-divide into so called "hyphenated" techniques (e.g., GC-MS, LC-NMR), and multi-modal techniques where one particular instrument is used in various modes (e.g., different columns). This will allow techniques such as column-switching and recycle chromatography to be included in this discussion.

Multidimensional separations are implemented by way of column switching. In column switching,<sup>55,56,57</sup> a small cut,  $\Delta t$ , from a primary column is placed onto secondary columns. If  $\Delta t$  is sufficiently small, the resolution achieved on the first column is not degraded by the second column. The cut from primary to secondary columns can be done either on- or off-line. Majors<sup>58</sup> reviews the advantages and disadvantages of on- and off-line coupling. With the advent of computers, improved precision and reproducibility is achieved which benefits on-line techniques. This leads to a decrease in sample loss and analysis time. Other articles<sup>59,60,61,62</sup> demonstrate the importance and necessity of automation when using multidimensional chromatography by incorporating the word *automation* into the title.

The motivation for the development of the multidimensional techniques comes from several correlating factors.<sup>54</sup> Statistical treatment concludes that, without *a priori* information, there are approximately five to six million possible compounds that every technique must be able to differentiate to provide comprehensive qualitative analysis. Judicious application of knowledge about the sample can minimize the sample set. The peak capacity,<sup>56,63</sup>  $\Theta$ , for a single mode of separation is given by:

$$\Theta = 1 + N^{\frac{1}{2}} \frac{\ln(1 + k'_n)}{m} \quad \text{Equation 2.5}$$

where N is the number of theoretical plates,  $k'_n$  is the capacity factor for the last member of a series of elutes, and  $m = 4 (4\sigma)$  is unit resolution. For a series of consecutive columns with independent, orthogonal modes, Equation 2.5 reduces to:

$$\Theta = (\Theta_1 \Theta_2 \Theta_3 \dots \Theta_n) \quad \text{Equation 2.6}$$

For a series of n identical columns Equation 2.5 reduces to:

$$\Theta = n^{\frac{1}{2}} \Theta_n \quad \text{Equation 2.7}$$

Comparison of Equations 2.6 and 2.7 shows the exponential increase in peak capacity that arises from column (mode) switching. Snyder<sup>64</sup> presented a detailed comparison of normal elution; coupled columns; and solvent-, flow-, and temperature-programming in liquid chromatography. He concluded that coupled columns have advantages for a large class of separations, particularly those involving repetitive, high efficiency liquid chromatograms. Snyder also discussed the benefits of coupled columns to provide sharply increased resolution at specific points in the chromatogram.

### ***Section 2.9.2 Applications of MD-HPLC***

Some of the initial applications of multidimensional chromatography have been in the analysis of polycyclic aromatic compounds (PAC) in complex matrices. Increasing concern over environmental and safety impacts has led to detailed study on PAC's. The determination of PAC's has proven to be a difficult problem. Typically there is a large number of components with widely varying chemical properties, and hence widely varying capacity factors, in a dirty sample. Sonnefeld and co-workers<sup>65</sup> described the on-line coupling of both a normal phase system and a reverse phase system for the analysis of organic compounds in shale oil. The sample was passed through a normal

phase aminosilane column, and the desired fraction was switched onto a diamine column for on-column concentration. The concentrator column (i.e., the diamine column) was then placed in-line with an analytical ODS column and the sample focused by gradient elution. The on-line method gave comparable results to their previous sequential HPLC method but offered several advantages. The on-line system was readily automated, minimized sample loss, and yielded lower detection limits.

Pearson and Gharfeh<sup>66</sup> used three columns—a cyano and two amino-cyano columns—to achieve group-type determination of hydrocarbon types in crude oil residues. The cyano column retained the resins while the aminocyano columns separated the aromatics and saturates. The strongly retained components (the resins) were backflushed off the cyano column to decrease analysis time and to improve peak shape. Without the multiple columns, Pearson and Gharfeh were unable to resolve the three fractions. The analysis was rapid, clean, and readily automated.

Bartle, et al.<sup>62</sup> used an on-line multidimensional liquid chromatography instrument coupled to a capillary gas chromatography instrument to isolate and detect tract level PAC's in diesel exhaust particulate samples. The LC used an aminosilane and a silica microbore column to clean the sample and fractionate the aromatics. By backflushing the LC columns, the aromatics were then directly introduced into the GC with minimal LC solvent (i.e., no dilution) and no intermediate sample pre-treatment was needed. They reported better sensitivity, reproducibility, and ease of operation as compared to the equivalent off-line, non-coupled, systems.

Cortes and co-workers<sup>60</sup> demonstrated the characterization of polymers by the combination of size exclusion chromatography (SEC) and pyrolysis gas chromatography. SEC provided molecular weight information whereas the pyrolysis GC provided polymer information. This application is applicable to other nonvolatile samples, such as bacteria, proteins, and complex carbohydrates.

Several papers describe using multidimensional chromatography for the determination of drugs in plasma. Sample preparation has become the limiting step in the determination of biological samples. Takahagi, Horiguchi, and co-workers<sup>67</sup> have developed a column switching technique that combines aqueous gel chromatography along with reverse phase chromatography using an ODS column. This combination allowed direct injection of plasma without the traditional pre-treatment, such as deproteinization or extraction. This automated technique linked sample pre-treatment, concentration, and chromatographic separation into one process.

Other references describe applications using multidimensional chromatography and specific detectors such as fluorimeters<sup>68</sup> and mass spectrometers<sup>69</sup> to provide complex separations coupled with "information-rich" detectors to speciate plasma samples.

Multidimensional chromatography is also finding applicability in food determinations. Stringent labeling laws and consumer awareness have created an impetus to provide detailed information on food samples and processing, and on feed stocks for animals.

Majors<sup>58,70</sup> determined the concentration of malathion on tomato plants via gel permeation chromatography (GPC) and reverse phase chromatography. The tomato fronds were macerated, extracted, and dissolved in THF. The resulting extract was injected onto a GPC column, and the resulting malathion fraction was heart cut onto a reverse phase column with good recovery and resolution. Majors<sup>58</sup> also analyzed a protein food supplement for the concentration of B vitamins. The aqueous size exclusion chromatogram exhibited broad, multi-banded peaks. The appropriate band was then switched onto a reverse phase column. The composite chromatogram exhibited resolved vitamins from the excipients. The SEC column combined with

selective cutting prevented proteins from passing onto the RP column and being irreversibly retained.

Sterigmatocystin is a recognized carcinogen and causes death to dairy cattle. Previous determinations were unsuccessful due to interfering compounds in the feed stock. Lepom<sup>71</sup> developed a multidimensional method to analyze for sterigmatocystin in feed stocks at ppb levels without interference. Lepom also discovered that his analytical column life was extended significantly by using his multidimensional method.

Additional food samples analyzed by combination multidimensional chromatography and/or hyphenated techniques include caffeine by HPLC,<sup>72</sup> pear sugars and acids by GC-MS,<sup>73</sup> rum flavors via Gas-Liquid Chromatography-MS,<sup>74</sup> and trace metals in natural waters.<sup>61</sup> Majors<sup>58</sup> also lists approximately thirteen other multidimensional applications for food samples, and a recent review<sup>75</sup> on multidimensional liquid chromatography lists additional references.

## LIST OF REFERENCES

## LIST OF REFERENCES

- 1 N. Graham; *Artificial Intelligence--Making Machines 'Think'* 1979, TAB Books Inc.
- 2 C. G. Enke, A.P. Wade, P. T. Palmer, K. J. Hart, *Anal. Chem.*, 1987, 59, 1363A.
- 3 *Anal. Chem.*, 1988, 60, 1534.
- 4 S. S. Williams, J. F. Karnicky, J.-L. Excoffier, and S. R. Abbott, Expert System Program for Assistance in High Performance Liquid Chromatographic Method Development in *Computer-Assisted Method Development for High-Performance Liquid Chromatography*, J.L. Glajch and L.R. Snyder, editors, Elsevier Science Publishers, 1989, 267.
- 5 K. Parsaye, M. Chignell; *Expert Systems for Experts*, John Wiley & Sons, Inc. 1988.
- 6 P.-H. Winston; *LISP*, Addison-Wesley Publishing Company, 1986.
- 7 J. R. Walters and N. R. Nielsen, *Crafting Knowledge-Based Systems Expert Systems Made Realistic*, John Wiley & Sons, Inc., 1988, Chapter 5.
- 8 A.P. Wade, S.R. Crouch, and D. Betteridge, *TrAC*, 1988, 7(10), 358.
- 9 D. H. Ackley G. E. Hinton, and T. J. Sejnowski, *Cognitive Science*, 1985, 9, 147.
- 10 J. J. Hopfield, 'Neural Networks and Physical Systems with Emergent Collective Computational Abilities,' *Proceedings of the National Academy of Sciences, USA*, 1982, 2554.
- 11 A. Lane, *PC AI*, 1989, 3(6), 20.
- 12 *The American Heritage Dictionary*, software release 1.1, Houghton Mifflin Co., 1992.
- 13 P. Jackson; *Introduction to Expert Systems*, Addison-Wesley Publishing Co. Wokingham, England, 1986.
- 14 T.P. Bridge, M.H. Williams, and A.F. Fell, *Chem. Bri.* 1987, 23(11), 1085.
- 15 D. H. Smith, 'Artificial Intelligence: The Technology of Expert Systems' in *Artificial Intelligence Applications in Chemistry*, T. H. Pierce, Bruce A. Hohne, eds. 1986, 1.
- 16 F.A. Settle and M.A. Pleva, *American Lab.*, 1988, 20(10), 64-79.



- 17 J. McCarthy; *LISP--Notes on its Past and Future*, Conference Record of the 1980 LISP Conference, Stanford University, August 25-27, 1980.
- 18 N. A.B. Gray; *Computer Assisted Structure Elucidation*, 1986, Wiley-Interscience, 166.
- 19 Barr and E. A. Feigenbaum, *The Handbook of AI*, William Kaufmann, Inc., 1982.
- 20 P.-H. Winston; *Artificial Intelligence*, Addison-Wesley Publishing Co. 1986.
- 21 M. P. Derde, L. Buydens, C. Guns, D. L. Marrat, and P. K. Hopke, *Anal. Chem.* 1987, 59, 1868.
- 22 *TI Users Guide*, Tecknowledge, Inc., Palo Alto, CA 1984.
- 23 E. A. Feigenbaum, *Proceedings of the 5<sup>th</sup> International Joint Conference on Artificial Intelligence*, 1977, 1014.
- 24 J. S. Bennet and R. S. Englemore, *Proceedings of the 6<sup>th</sup> International Joint conference on Artificial Intelligence*, 1977, 1014.
- 25 W. J. Clancey, The Epistemology of a Rule-Based Expert System—A Framework for Explanation in Artificial Intelligence, 1983, 20, 215.
- 26 B. J. Wythoff and S. A. Tomellini, *Research and Development*, 1989, 31(4), 52.
- 27 F. A. settle, Jr., *J. Chem. Ed.*, 1987, 64(4) 340.
- 28 D. Betteride, R. Mackison, C. M. Mottershead, A. F. Taylor, and A. P. Wade, *Anal. Chem.* 1988, 60, 1534.
- 29 T. P. Bridge, M. H. Williams, G.G. R. Seaton, and A. F. Fell, *Chromatogr.* 1987, 24, 691.
- 30 A. G. Wright, A. F. Fell, and J. C. Berridge, *Chromatogr.* 1987, 24, 533.
- 31 M. De Smet, A. Peeters, L. Buydens, and D. L. Massart, *J. Chromatogr.* 1988, 457, 25.
- 32 R. Hindriks, F. Maris, J. Vink, A. Peeters, M. De Smet, D. L. Massart, and L. Buydens, *J. Chromatogr.*, 1989, 485, 255.
- 33 Marina De Smet, Greet Musch, Anne Peeters, Lutgarde Buydens, and Desire L. Massart, *J. Chromatogr.* 1989, 485, 237.
- 34 H. Gunasingham, B. Srinivasan, A. L. Ananda, *Anal. Chim. Acta*, 1986, 182, 183.
- 35 K. Tsuji and K. M. Jensins, *J. Chromatogr.* 1989, 485, 297.
- 36 M. Mulholland, J. A. Van Leeuwen, and B. Vandeginste, *Anal. Chim. Acta*, 1989, 223(1), 183.

- 37 R. Bach, J. Karnicky, and S. Abbott, 'An Expert System for High Performance Liquid Chromatography Methods Development' in *Artificial Intelligence Applications in Chemistry*, Thomas Pierce, ed., ACS Symposiums Series; 306, 1986, 278.
- 38 S.S. Williams J.F. Karnicky, J.-L. Excoffier, and S.R. Abbott, 'Expert System Program for Assistance in High-Performance Liquid Chromatographic Method Development,' in *Computer-Assisted Method Development for High-Performance Liquid Chromatography*, J.L. Glajch and L. R. Snyder, eds. Elsevier, 1990, 267.
- 39 L. Peichang and H. Hongxin, *J. Chromatogr.* 1988, 452, 175.
- 40 M. A. Tischler and E. A. Fox, *Comput. Chem.* 1987, 11(4), 235.
- 41 J. A. Van Leeuwen, B. G. M. Vandeginste, and G. Kateman, *Anal. Chim. Acta*, 1990, 228, 145.
- 42 P. de B. Harrington and K. J. Voorhees, *Anal. Chem.* 1990, 62, 729.
- 43 P. de B. Harrington, T. E. Street, K. J. Voorhees, F. R. di Brozolo, and R. W. Odom, *Anal. Chem.* 1989, 61, 715.
- 44 K. Pedersen, *AI Expert* 1988, 3(5), 26.
- 45 D. Merritt, *PC AI*, 1988, 2(3), 51.
- 46 M. C. Paxton, *AI Expert* 1988, 3(10), 55.
- 47 K. Parsaye, M. Chignell, S. Khoshafian, and H. Wong, *Intelligent Databases*, John Wiley and Sons, Inc., New York 1989, pps. 164, 440.
- 48 K. Parsaye, M. Chignell, S. Khoshafian, and H. Wong, *AI Expert* 1990, 62, 38.
- 49 D. P. Webb and E. D. Salin *Intelligent Instruments and Computers* 1991(issued 1992), 9(5), 185.
- 50 *M! Sample Knowledge Systems*, Teknowledge Inc., Palo Alto, Ca., 1984.
- 51 *M! Reference Manual*, Teknowledge Inc., Palo Alto, Ca., 1984.
- 52 T. J. Laffey, *Byte*, 1991, 16(1), 259.
- 53 L. A. Zadeh, *AI Expert* 1992, 7(3), 23-7.
- 54 R. E. Majors, *LC-GC*, 1987, 5(6), 454.
- 55 J. C. Giddings, *J. High Resolut. Chromatogr. Chromatogr. Comm.* 1987, 10, 319.
- 56 D. H. Freeman, *Anal. Chem.* 1981, 53, 2.
- 57 Poole and Schuette, *Contemporary Practice of Chromatography*, Elsevier Science Pub., Netherlands, 1984, 334.

- 58 R. Majors, *J. Chromatogr. Sci.* 1980, 18, 571.
- 59 M. M. Bushey and J. W. Jorgenson, *Anal. Chem.* 1990, 62, 161.
- 60 H. J. Cortes, G. L. Jewett, C. D. Pfeiffer, S. Martin, and C. Smith, *Anal. Chem.* 1989, 61, 961.
- 61 Y. Liu and J. D. Ingle, Jr., *Anal. Chem.* 1989, 61, 525.
- 62 I. L. Davies, M. W. Raynor, P. T. Williams, G. E. Andrews, and K. D. Bartle, *Anal. Chem.* 1987, 59, 2579.
- 63 J. C. Giddings, *Anal. Chem.* 1964, 39, 1027.
- 64 L. R. Snyder, *J. Chromatogr. Sci.* 1970, 8, 692.
- 65 W. J. Sonnefeld, W. H. zoller, W. E. May, and S. A. Wise, *Anal. Chem.* 1982, 54, 723.
- 66 C. D. Pearson and S. G. Gharfeh, *Anal. Chem.* 1986, 58, 307.
- 67 H. Takahagi, K. Inoue, and M. Horiguchi, *J. Liq. Chromatogr.* 1986, 362, 369.
- 68 C. A. Palmerini, M. G. Cantelmi, A. Minelli, and C. Fini, *J. Chromatogr.* 1987, 417, 378.
- 69 M. S. Lant and J. Oxford, *J. Chromatogr.* 1987, 394, 223.
- 70 E. L. Johnson, R. Gloo, and R. E. Majors, *J. Chromatogr.* 1978, 149, 571.
- 71 P. Lepom, *J. Chromatogr.* 1986, 354, 518.
- 72 J. Van Duijn and G. H. D. van Der Stegen, *J. Chromatogr.* 1979, 179, 199.
- 73 I. Akhavan, R. E. Wrolstad, and D. G. Richardson, *J. Chromatogr.* 1980, 190, 452.
- 74 H. M. Liebich, W. A. Koenig, and E. Bayer, *J. Chromatogr. Sci.* 1970, 8, 527.
- 75 H. G. Barth, W. E. Barber, C. H. Lockmuller, R. E. Majors, and F. E. Regnier, *Anal. Chem.* 1988, 60, 407R.

## **CHAPTER 3 INSTRUMENTATION AND SOFTWARE**

**Man has received direct from God only one instrument wherewith to know himself and to know his relation to the universe--he has no other--and that instrument is reason.**

**Leo Tolstoi**

A combination of instrumentation, software, and electrical circuits is used in this dissertation. A high performance liquid chromatography instrument is interfaced to a computer by both a commercial analog/digital interface board and a prototype board, built in-house, which contained a programmable parallel interface chip. This chapter describes both the hardware (computer, HPLC instrument, circuits, and supporting peripherals) and the software (expert system and other software) used in this research.

### **Section 3.1     Expert System**

#### ***Section 3.1.1     AI Languages Versus Commercial Shells***

As mentioned in Chapter 2, there are two approaches that can be used when developing an expert system. The first approach is to write the shell, containing the inference engine and man-machine interface, from some other lower-level computer language such as C or Prolog. The AI language approach provides faster execution speeds and better customizability of the expert system for the particular application. This approach does require more time and knowledge on the part of the knowledge engineer. The language approach requires that the knowledge engineer be proficient in

both programming techniques and in the skills of the knowledge engineer position itself. While this computer proficiency has been required in the past when commercial shells were unavailable, knowledge engineers are now expected to have skills with the application itself.<sup>1</sup> This familiarity with the area of the application allows for a more rapid implementation of the expert system and a decrease in training time for the knowledge engineer, who now could also be the expert, that is needed to become knowledgeable in the language and required skills of the application.

To provide the knowledge engineer with more time to become familiar with the knowledge domain of the application, expert system tools, such as commercial shells, must be developed that allow him/her to create the expert system with minimal programming. Most commercial shells contain a framework that incorporates advanced database and inference techniques, and provides sophisticated user interfaces.<sup>2</sup> The knowledge engineer can rapidly develop and distribute the expert system can be more quickly prototyped. Additionally, other expert systems developed from the same shell will be similar in design. This is a very important aspect to consider in industry where a concern of management is the cost required to train employees to use different computer tools. While shells are a powerful tool, the resultant expert system based on a commercial shell is often less extendible and is slower than an expert system based on an AI language.

Wade, Crouch, and Betteridge have compared both AI languages and commercial shells.<sup>3</sup> Laffey<sup>4</sup> has also evaluated traditional shells as applied to systems that monitor and control complex instruments and other systems in real-time. A summary of their conclusions is presented in Table 3.1; however, some of the comparisons have changed as the commercial shells have evolved.

**Table 3.1**  
**Comparison of Shells Written from an AI Languages versus**  
**Commercially-Available Shells**

| Remark | AI Languages                                                                              | Shells                                                                                                 |
|--------|-------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------|
|        | Hand-crafted by local programmer. <sup>†</sup>                                            | Commercial System. <sup>†</sup>                                                                        |
|        | Source code available and modifiable. <sup>†</sup>                                        | Source code usually not available. <sup>†</sup>                                                        |
| 1      | Command set and knowledge formalism are extensible. <sup>†</sup>                          | Limited to supplied commands. <sup>‡</sup>                                                             |
| 2      | Support of sound, color, and graphics limited only by programmer. <sup>†</sup>            | Support of graphics, mouse, etc. routinely added. <sup>‡</sup>                                         |
|        | Slow implementation of expert system. <sup>†</sup>                                        | Fast development of the system. <sup>†</sup>                                                           |
| 3      | Flexible/modifiable user interface. <sup>†</sup>                                          | More limited interface, but some control offered to engineer. <sup>‡</sup>                             |
|        | User interface/error checking routines must be written. <sup>†</sup>                      | User interface usually included. Error checking possible. Consistent interface. <sup>†</sup>           |
|        | No external support for system. <sup>†</sup>                                              | System supported by vendor. <sup>†</sup>                                                               |
| 4      | Routines to access of external programs must be written. Limited only by programmer.      | Intrinsic support of external programs. Limited to the supplied calls of the command set. <sup>‡</sup> |
|        | Fast execution, amenable to real-time applications. <sup>□</sup>                          | Slow, not optimized for particular applications. <sup>□</sup>                                          |
|        | Input from non-human sources possible. Allows monitoring/control of systems. <sup>□</sup> | Not often implemented. Limits implementation for real-time applications. <sup>□</sup>                  |

Legend

- † from Wade, Crouch, and Betteridge<sup>3</sup>  
‡ includes proposed revisions by the author to Wade, Crouch, and Betteridge<sup>3</sup>  
□ from Laffey<sup>4</sup>

Discussion of the suggested changes is based on the shell used in this thesis, namely VP-Expert, versions 2.0 and 3.0 (WordTech Systems, Inc., Orinda, CA). Some comparisons to other systems will be mentioned briefly.

1. Although users are limited to the supplied commands, the command set for commercial shells has become quite rich and diverse. Current commands that

are implemented in VP-Expert include: forward/backward chaining, loop constructions (FOR, WHILETRUE, WHILEKNOWN), file manipulation (text, database, spreadsheet), graphics, and string handling, among others. Extra features are available via external program (see point 4).

2. VP-Expert has supported mouse control, color, and graphics since version 1.1 (1988). Later implementations allow the creation of dynamic images which are pre-defined graphical constructs that are dynamically linked to the value of a particular variable. The graphical object is updated as the value of the variable changes.<sup>5</sup> Examples of dynamic images would be gauges, buttons, and graphs. These dynamic images can be combined to create visual representations of instrumentation, present results graphically, such as chromatograms, and so forth. One application of dynamic images would be to create a diagram of an HPLC instrument in which the user would be able to select different points (e.g., injector site) on the diagram to direct the expert system to provide diagnostic information concerning that area.

In addition to dynamic images, VP-Expert version 3.0 can read graphic files stored in PCX format from some device, such as a hard disk, and display it on screen. This capability would allow representative chromatograms to be displayed. Table 3.2 provides a list of the graphic formats that other expert system shells are able to import.

3. VP-Expert supports the creation of "smart forms," display screens that resemble regular paper forms. Integral to these smart forms are graphical constructs called "Formfields." Formfields are rectangular boxes containing the value of an associated variable. A user may select the formfield and a pull-down menu

appears from which the user may select a legal response to the variable. Smart forms may be used conjunction with the other graphical dynamic images mentioned above to allow the knowledge engineer to have control of the format of the interface. The user supplies answers to all the answers on the screen and eventually selects an "ALL DONE" button to enter this information into the computer. The user may answer many questions at once and edit the answers before submitting them to the expert system. Via a smart form, a knowledge engineer can create an interface that mimics forms with which the user might already be familiar (e.g., medical diagnostic forms). Smarforms do require that the engineer write more complex code to create visually appealing forms. The knowledge engineer may also use simultaneously traditional question-response format (dialogic) to generate a wide variety of user-interfaces.

4. VP-Expert supports direct calls to Lotus- and DBase-compatible spreadsheets and databases, respectively. VP-Expert also may directly invoke external executable files (\*.EXE, \*.COM, and \*.BAT); limited only by the available computer memory. This feature allows the knowledge engineer to write, or use pre-existing, supporting software in more appropriate languages (e.g., data manipulation in a language such as C or BASIC).

Table 3.2, adapted from Tway and Riedel,<sup>6</sup> presents some of the features of different expert systems and AI languages that are commercially available. The first two columns give the software name and company. The Hypertext, and Graphics, and Data Access columns provides information on whether these features, and format, are directly supported in the package. Prices listed in parenthesis indicate additional costs that are required to obtain that specific feature. This table lists the different features



that are available and shows that VP-Expert is an appropriate choice for developing small- to medium-sized expert systems.

Based upon the above considerations in Table 3.1 and the subsequent discussion, the decision was made to implement the expert system described in this dissertation using one of the commercially available shells. This decision allowed the project to be developed by chemistry specialists who otherwise did not have the experience and/or expertise in developing a complex expert system shell. The prototype expert system was implemented using M-1 and then adapted to VP-Expert.

### ***Section 3.1.2 M1 Expert System Shell***

The prototype HPLC expert system was developed using the commercial shell, M1, version 1.0 (Teknowledge, Inc., Palo Alto, CA). M1 was developed initially for use on mainframe computers and was subsequently ported to the PC environment, which could have caused some of its limitations described below.<sup>1</sup> The initial version of M1 was very expensive (\$10,000) and had slow execution speeds due to being encoded in the Prolog language. Because of limited memory, the knowledge base was restricted to a maximum of 250 rules. M1 was also unable to support all of the available memory, such as expanded or extended memory, of a PC. Version 2.0 was re-coded in the C programming language to improve speed and to minimize memory constraints such that the knowledge base was increased to 1000 rules.

With the limit of 1000 rules, the knowledge base of the prototype still exceeded the available space. The space limitations necessitated breaking the knowledge base into several parts which required that all important parameters (variables, intermediate results, and so forth) were preserved between calls from one knowledge base file to the next file. These adaptations created additional overhead in execution time. The user interface was also found to be quite difficult to use. Finally, direct communication with

databases, spreadsheets, and executable programs was not possible, although an extension, or hook, in the M1 system allowed programmers to write interfaces in C to communicate with these programs. Since multiple calls to external programs were envisioned in the final version of this expert system, the decision was made to transfer the expert system to a program that supported these calls intrinsically. so that faster, and easier, calls could be made.

### ***Section 3.1.3      VP-Expert Expert System Shell***

Subsequent development of the expert system described in this dissertation was accomplished using VP-Expert (PaperBack Software, Berkeley, CA). Preliminary prototyping was accomplished using version 1.0 with a majority of the expert system completed using version 2.0. The final version uses version 3.0 (WordTech Systems, Inc. Orinda, CA). Many of the advanced features of VP-Expert were not implemented. These capabilities will be discussed in Chapter 7 with respect to the future development of the system.

One goal of the work described in this dissertation was to create a system that could be run by the widest variety of PC-compatible systems. Many shells are still quite expensive and/or require expensive computer hardware and software, which would minimize the use of such a system and in academia, where the expert system was envisioned to be primarily used. As mentioned in Chapters 1 and 4, one objective of this system is for the expert system to serve as a teaching tool.

**Table 3.2**  
**Comparison of Different Expert Systems**  
 Table adapted from Tway and Riedel<sup>6</sup>

| Software                 | Company               | DOS Cost | Windows Cost | Hyper-text         | Graphics                          | Data Access          |
|--------------------------|-----------------------|----------|--------------|--------------------|-----------------------------------|----------------------|
| 1 <sup>st</sup> Class HT | AI Corp.              | 495      | NA           | Y                  | 1                                 | A,D,L                |
| Adept                    | Symbolic              | NA       | 695          | N                  | 2                                 | A                    |
| Eclipse                  | Haley Enterprise      | 79       | 199          | N                  | N                                 | A,D                  |
| EXSYS Pro                | EXSYS Corp.           | 995      | 1295         | Y                  | 4                                 | A,E                  |
| Goldworks III            | Gold Hill Computer    | NA       | 7900         | Y                  | 2,3                               | A,D,L                |
| GURU                     | mdbs                  | 7000     | NA           | N                  | N                                 | A,D,E,L              |
| IBIS                     | Intelligence Mfg. Co. | 250      | NA           | Y                  | 4                                 | A,D,L<br>(\$950 DOS) |
| Knowledge Pro            | Knowledge Garden      | 195      | 249          | Y<br>(\$179 )      | 2,4<br>(\$179 DOS)<br>(\$249 Win) | A,D                  |
| Level 5 Object           | Information Builders  | 295      | 495          | N (DOS)<br>Y (Win) | 2,4                               | A,D                  |
| M.4                      | Cimflex Teknowledge   | NA       | 995          | NA                 | NA                                | A,D                  |
| Mentor                   | Icarus Corp.          | 495      | NA           | N                  | 4                                 | A,D,L                |
| Nexpert Object           | Neuron Data Corp.     | 5000     | 5000         | N                  | 2                                 | A,D,E,I,L,O          |
| Object Vision            | Borland               | NA       | 99           | N                  | 5                                 | A,D,E,L,P,Q          |
| VP-Expert                | Wordtech Sysems       | 349      | NA           | Y                  | 4                                 | A,D,L                |
| Smalltalk\V              | Digital Corp.         | 125      | 500          | N                  | NA(DOS)<br>2(Win)                 | A                    |
| Star Sapphire LISP       | Sapiens Software      | 100      | NA           | N                  | NA                                | A                    |
| GCLISP                   | Gold Hill             | NA       | 1995         | Y                  | 2                                 | A,D,L                |
| Arity Prolog             | Arity Corp.           | 650      | 350          | N                  | N                                 | A,D,L                |
| PDC Prolog               | Prolog Devel. Center  | 299      | 599          | Y<br>(\$429)       | 1                                 | A,D,L<br>(\$125)     |

| Legend           |                    |                                                                                                                             |
|------------------|--------------------|-----------------------------------------------------------------------------------------------------------------------------|
| <b>Graphics</b>  | <b>Data Access</b> | <b>Other</b>                                                                                                                |
| 1 Screen Capture | A ASCII            | Extra costs for the different functions are given in parenthesis under the columns where extra modules/versions are needed. |
| 2 BMP Format     | D dBASE            |                                                                                                                             |
| 3 TIFF Format    | E Excel            |                                                                                                                             |
| 4 PCX Format     | I Ingres           |                                                                                                                             |
| 5 Clipboard      | L Lotus            |                                                                                                                             |
|                  | O Oracle           |                                                                                                                             |
|                  | P Paradox          |                                                                                                                             |
|                  | Q QuatroPro        |                                                                                                                             |

VP-Expert is compiled in Microsoft C version 6.0 (Redmond, WA) which allows fast execution. VP-Expert allows mixed modes of both forward and backward chaining. Rules are encoded in production rule format (IF THEN format) but do not support advanced features such as objects and frames. The program directly supports spreadsheets (WK1 and WKS format) and databases (dBase II, III, III+ format). The program has an integrated editing and debugging facility to assist in generating correct syntax. In addition, the program generates both traces and reports to show the induction path that was taken. A trace is an output that shows the different rules that were used by the inference engine to obtain the values for the goals specified in the expert system. A report is a file that contains the explanatory text contained in the **BECAUSE** clauses of every rule that was invoked. The report feature facilitates in debugging the system and in using the expert system as a tutorial tool. Reviews of VP-Expert provide additional descriptions and criteria that were used in choosing this shell.<sup>7,8,9</sup> Presented below in Table 3.3 are the requirements and features of VP-Expert.

Some of the features in Table 3.3 deserve special mention. While it is possible to use other editors (e.g., word processors), an integrated editor allows the inference engine to check the syntax of the code and show where an error occurs. While often overlooked, showing the position of an error can be an essential feature for the novice knowledge engineer. VP-Expert presents the trace mode in both a text and a graphics mode. The trace mode is helpful, but it is not as elaborate or "user-friendly" as some of the more expensive/complex expert systems (particularly the Windows-based ones).

**Table 3.3**  
**Summary of Features of VP-Expert**

| <b><u>Item</u></b>        | <b><u>Requirement/Feature</u></b>                                                                                                                                                                                            |
|---------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Computer</b>           | <b>IBM PC, XT, or higher</b>                                                                                                                                                                                                 |
| <b>RAM</b>                | <ul style="list-style-type: none"> <li>• 512 K</li> <li>• Extended and Expanded Memory Supported</li> </ul>                                                                                                                  |
| <b>Graphics Supported</b> | <b>Hercules, CGA, EGA, VGA</b>                                                                                                                                                                                               |
| <b>Mouse</b>              | <b>Supported (Required for Graphics operations)</b>                                                                                                                                                                          |
| <b>Inference Engine</b>   | <b>Forward/Backward Chaining</b>                                                                                                                                                                                             |
| <b>External Programs</b>  | <ul style="list-style-type: none"> <li>• Executable Programs (.EXE,.COM,.BAT)</li> <li>• Database (dBase II, III, III+ compatible)</li> <li>• Spreadsheet (WKS format)</li> <li>• Text and HyperText Files</li> </ul>        |
| <b>Graphics</b>           | <ul style="list-style-type: none"> <li>• Dynamic images (meters, gauges, buttons)</li> <li>• Graphic Primitives</li> <li>• PCX format files</li> </ul>                                                                       |
| <b>Debugging</b>          | <ul style="list-style-type: none"> <li>• Integrated Editor with Syntax checking</li> <li>• Trace Mode (Text/Graphics)</li> <li>• "WHAT IF" features</li> <li>• WHY and HOW supported</li> <li>• Report Generation</li> </ul> |

The WHY and HOW functions were described in Chapter 2. The "WHAT IF" capability allows the user to change the value of one of the variables after the consultation is completed. The obvious utility of this procedure is to allow the user to correct an invalid response that the user provided during the consultation. The more intriguing application is to use this function as a learning tool. The user can change the values of different variables and can observe the effect the changes have on the results

of the consultation. The user can deduce the importance and interaction of different facts. This observation allows the expert system to serve as a tutorial tool.

#### ***Section 3.1.4      Additional Software Support***

Efficient expert systems are generated when dynamic information is stored in a format that allows rapid and easy editing of the data. Expert systems are also the most powerful when they can access the large body of the information that may be already available. To implement this idea, most of the facts for the project described in this dissertation are stored in databases and text files. It was decided that the most efficient storage of information in this case would be via a database program, VP-Info (PaperBack Software, Berkeley, CA).

Text files were created using Word 5.0 (Microsoft Corp., Redmond, WA) and stored in ASCII format. Some of these text files were then converted into hypertext format. Hypertext format allows cross-referencing of words via INDEXER.COM, an ancillary program included with VP-Expert. This format allows comprehensive help files and explanatory information to be available without storing text inside the knowledge base itself.

### **Section 3.2      HPLC Equipment and Supplies**

This section describes the HPLC instrumentation used in this research. In addition to the prerequisite pump, injector valve, column, and detector, several additional valves were used.

### **Section 3.2.1      Instrumentation**

A Spectra Physics 8700 Ternary Solvent Delivery System (Freemont, CA) was used for this research. The solvent delivery system can be controlled either manually or by computer (further details are given in Chapter 6).

For the recycle heartcut experiments, a Chromatronix Model 225 Mixed Wavelength Absorbance Detector (Berkeley, CA) was connected to a Chromatronix Model 220 Absorbance Controller (Berkeley, CA) to provide detectability at wavelengths  $\lambda_1$ ,  $\lambda_2$ ,  $\lambda_1 - \lambda_2$ , and  $\lambda_2 - \lambda_1$  (where  $\lambda_1$  and  $\lambda_2$  are 254 and 280 nm, respectively). Output was connected to the computer via an in-house designed interface connector (*vide infra*) and also connected to a Heath Schlumberger Model EU-200-01 Potentiometric Amplifier.

Additional HPLC experiments were performed using a System Gold Programmable Detector, Module 166 (Beckman Instruments, Arlington Heights, IL). The detector is connected to a NEC PC-8300 computer controller (NEC Electronics, Inc., 7501 Foothills Blvd., Roseville, CA) via a System Gold Analog Interface, Module 406 (Beckman Instruments, Arlington Heights, IL).

### **Section 3.2.2      Solvents**

Solvents in this research were spectrophotometric grade from various sources. All solvents were vacuum filtered with a Millipore filtration system using Sartorius #20006-47-N 0.45  $\mu\text{m}$  nylon filter paper (Göttingen, Germany). Tap water was passed through a Corning Mega-Pure D2 System containing organic and ionic filter cartridges and then re-distilled in glass.

### **Section 3.2.3      Columns**

Different columns are available via the column selection valve. The ones that are used for the different experiments are listed in Table 3.4.

### Section 3.3 Computer Control and Automation

A PC-Designs FD-1000 XT-class computer was initially used to control the HPLC system. The XT computer was upgraded to a 12MHz AT class computer (80286). The computer was used to control the positions of two six-port valves and a column selection valves. The positions of the valves were read from a set of valve sensor devices. The computer was also used to program and monitor the HPLC solvent delivery system.

**Table 3.4**  
**HPLC Columns and their Parameters Used**  
**in this Research**

| Manufacturer     | Trade Name     | Packing Material           | Length (cm)<br>Diameter (mm) | Particle Size<br>( $\mu\text{m}$ ) |
|------------------|----------------|----------------------------|------------------------------|------------------------------------|
| Waters           | $\mu$ BondaPak | Phenyl                     | 25 x 4.6                     | 5                                  |
| In House         |                | Pellicular NH <sub>2</sub> | 10 x 4.6                     | 5                                  |
| Alltech          | Spherisorb     | NH <sub>2</sub>            | 25 x 4.6                     | 5                                  |
|                  | Zorbax         | C8                         | 25 x 4.6                     | 5                                  |
| Applied Sciences |                | C18                        | 25 x 4.6                     | 5                                  |
|                  | Zorbax         | C18                        | 25 x 4.6                     | 5                                  |

#### ***Section 3.3.1 Six Port Rheodyne Valves***

Two Rheodyne Model 7040 and 7010 six-port valves (Cotati, CA) were used in conjunction with the column selection valve (CSV) to provide the ability to bypass different columns. One possible configuration used is presented below in Figures 3.1 (A-C). In the first configuration (Figure 3.1A), the flow is directed through the primary column and out to waste. When the valve positions are changed to the configuration seen in Figure 3.1B, the eluate can be collected in the collection loop to be recycled back



through

column, a

A

loop was

utilize a

Load to

acquisiti

S

A

(Housto

sequent

two air

valve w

360 deg

accesse

The tim

two pul

informa

signal

counte

through a second column (if desired) connected to the CSV, through the primary column, and eventually to waste.

A Rheodyne Model 7125 six port injection valve (Cotati, CA) with a 20  $\mu$ L sample loop was used. The injection valve is not under computer control; however, it did utilize a reed valve assembly to send a TTL signal when the valve was turned from the Load to Inject position. The TTL signal was used to initiate the controller and data acquisition software.

### ***Section 3.3.2      Column Selection Valve***

A pneumatically-actuated Valco model ACST6UW multiple position valve (Houston, TX) was used to provide column selection capability. This valve allows sequential (counter-clockwise only) access to one of six columns. This valve requires two air pulses for every valve position change. The first air pulse ratchets, or arms, the valve while the second air pulse then changes the valve. The valve is designed so that a 360 degree translation will involve twelve column switches; that is, each column will be accessed twice within 360 degrees.

Two TTL signals from the computer are sent to a timing circuit (Section 3.3.6.1). The timing circuit then sends two TTL pulses of controlled duration to the CSV. These two pulses arm and change the CSV.

### ***Section 3.3.3      Valve Sensors***

All valves used in this research are connected to a switch to provide status information to the computer. The Rheodyne valves uses a reed switch to provide a TTL signal to the computer to determine whether the valve was clockwise (CW) or counterclockwise (CCW). Magnets attached to the rotors of the valves are used to

make el

position

and con

changes

points.

go thro

initial p

then co

positio

seven v

inputs

from t

residu

signals

increa

reliabi

make electrical contact of the switches when the valves are switched from CCW to CW position.

The column selection valve (*vide infra*) included a standoff containing a bushing and contacts (see Figure 3.2). As the valve position changes, the contact rotates and changes the electrical connection between the initial point and the different contact points. Two points are provided for each valve position since a 360 degree rotation will go through two complete sets of valve changes. A five volt signal is connected to the initial point and twelve wires are connected to the contact points. The first six lines are then connected to the seventh through twelfth lines so that each equivalent valve position (recall that two valve changes/360 degrees) is tied together (e.g., lines one and seven were connected). The six lines are then connected to one group of the digital inputs of the Programmable Parallel Interface (PPI) board (*vide infra*).

Pull-down resistors (47 K $\Omega$ ) on the negative side of the contact (the side farthest from the computer/TTL signal) are connected to ground. The resistors insured that no residual signal would appear on the sensor/relay due to induction and that the applied signals from the computer would be dissipated to ground much faster. This led to an increase in response time (~2 seconds to less than milliseconds) and an improvement in reliability.

Pump

Pump

Pump

Figure 3.1A: Through Primary Column to Waste

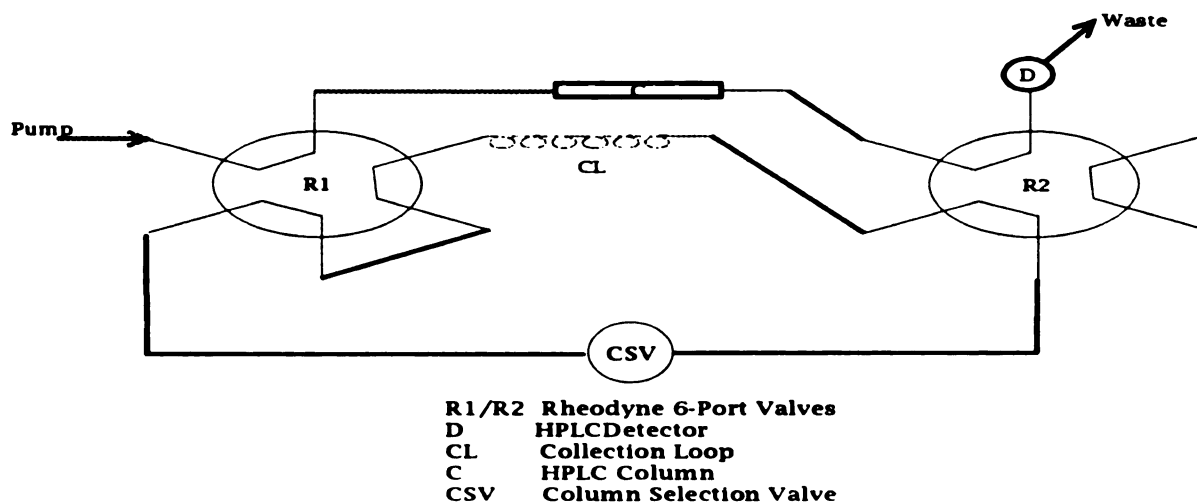


Figure 3.1B: Collection of Sample

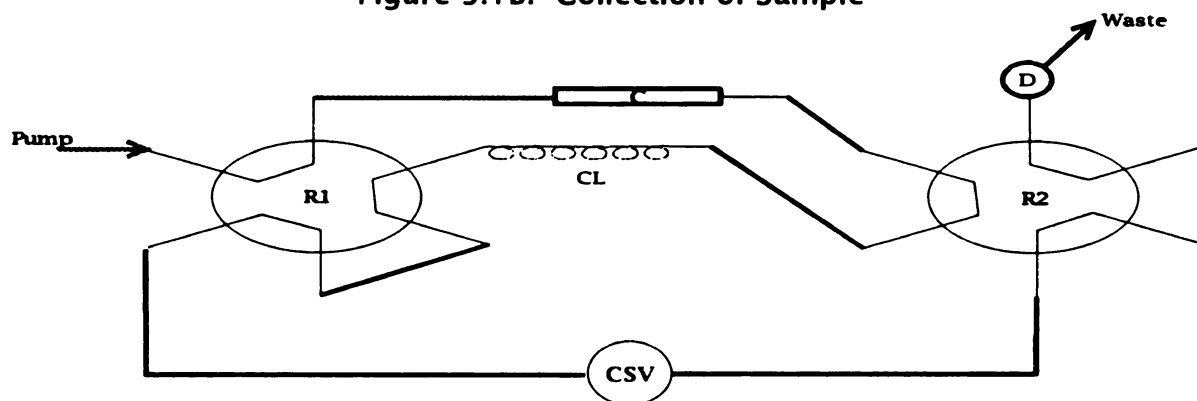
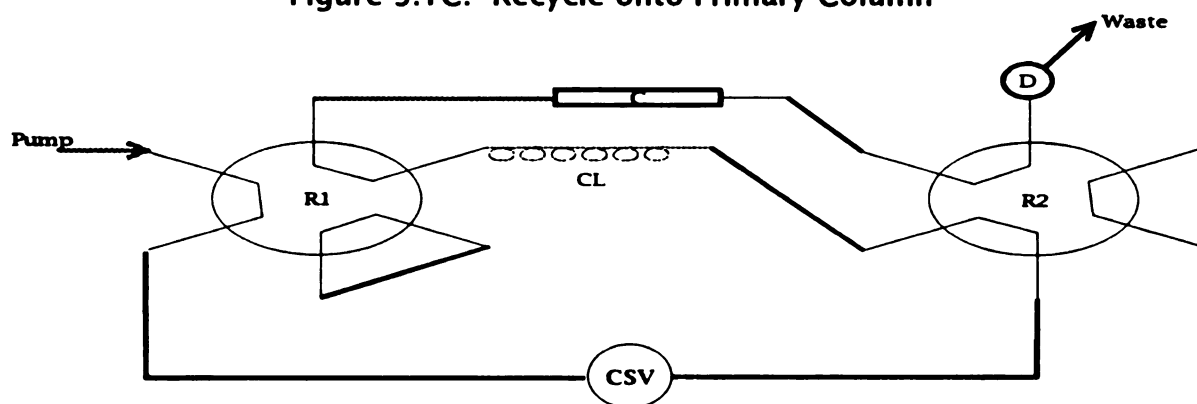
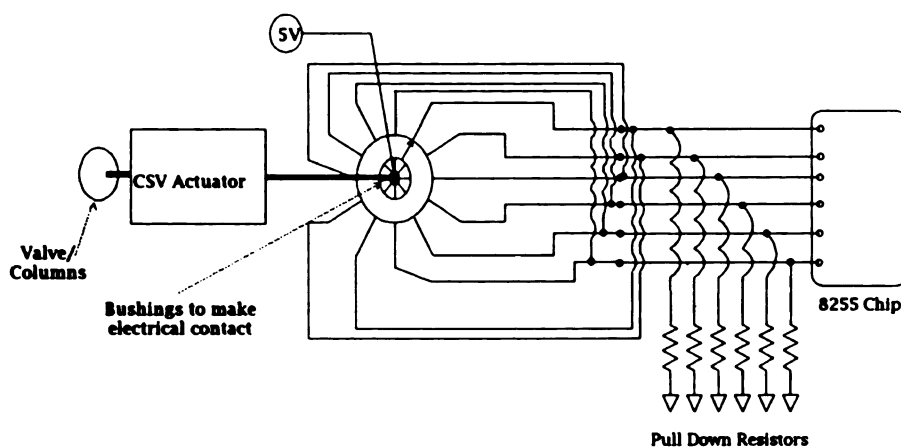


Figure 3.1C: Recycle onto Primary Column



**Figure 3.1**  
**Several Configurations of the Two Rheodyne**  
**Valves that Can Be Used**

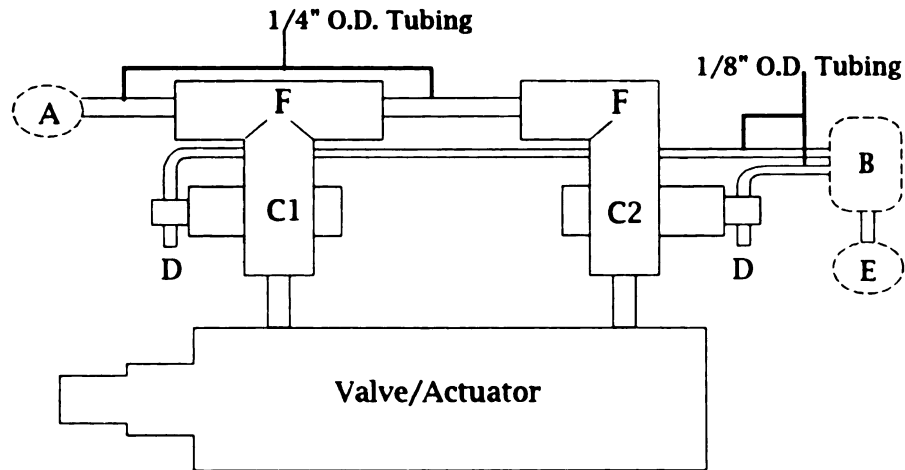


**Figure 3.2**  
**Sensor Connections For the Column Selection Valve**

#### ***Section 3.3.4 Pilot Valves***

Every time a valve switch occurs, a pressure shock wave is created.<sup>10,11</sup> This shock wave can create disturbances in the flow pattern, physical stress on the packing material, and changes in the refractive index of the solvent and the concomitant change in the detector response. Harvey and Stearns<sup>10,11</sup> have demonstrated the pressure drop associated with different valve switching times. In addition, slow valve changes may potentially split a chromatographic peak between two, or more, columns. Thus, a faster valve switches will minimize the effect of these stresses that are imposed on the system.

Since there are occasions to require up to five consecutive valve changes (e.g., switching from position 1 to position 6 on the CSV), a Valco High Speed Switching Accessory Kit (Houston, TX), composed of two Humphrey Products 125A Pilot Valves (Kalamazoo, MI), is used to accelerate the valve changes. Figure 3.3 shows the configuration of the pilot valves.



| Legend       |                           |    |          |
|--------------|---------------------------|----|----------|
| A            | Gas Tank for Pilot Valves | C2 | Pilot    |
| Valve (Fire) |                           | D  | Quick    |
| B            | Actuator                  | E  | Gas Tank |
| Exhaust      |                           |    |          |
| C1           | Pilot Valve (Arm)         |    |          |
| for Actuator |                           |    |          |

**Figure 3.3**  
**Pilot Valve Configuration**

### ***Section 3.3.5 Actuators, Solenoids, and Relays***

The valves described above are all air actuated. The two six-port Rheodyne valves are connected to two Schrader Model 41-1115 pneumatic actuators (Wake Forest, NC). The actuators are controlled by sending a High TTL signal from the RTI-800/815 interface board (*vide infra*). Since the actuators are active LOW, the RTI 800/815 interface board inverts the HIGH TTL signal to a LOW to maintain compatibility with the actuators.

The CSV was controlled by sending HIGH TTL signals to two Opto-22 1781-0A5S single point I/O optically isolated relays (Kendall Electronics, Lansing, MI). The two



relays

8615 b

describ

relays (

the HI

acquisi

built pr

Interfac

Decodin

board f

support

chips to

sources.

A

the CSV.

program

(SDS) of

three gro

groups o

relays then powered a four way, three position (center exhaust) Schrader Model 52000 8615 bellows solenoid (Wake Forest, NC) which directed the air pulses to the pilot valves described above. The two Rheodyne valves were controlled by two 611-3 Teledyne relays (Hawthorne, CA). Figure 3.4 shows the complete valve arrangement.

### ***Section 3.3.6 Computer Hardware to Control the HPLC Instrument***

The next section describes the two interface boards that were used to automate the HPLC instrument. The first board is a commercial multifunction data acquisition/interface board from Analog Devices. The second board was an in-house built prototype board from JDR Devices configured with an 8255 Programmable Parallel Interface (PPI) device. These two different boards are described below.

#### **Section 3.3.6.1          Prototype Board**

A JDR Microdevices Model IBM-PR2 Wirewrap Prototype Card with Input/Output Decoding for XT class computers (Los Gatos, CA) was used to develop a computer board for input and output to different peripherals. This board provides decoding support for eight different devices to be addressed through the computer bus. The chips to support the address and multiplexing logic were purchased from various sources. The primary chip on this board is the PPI device.

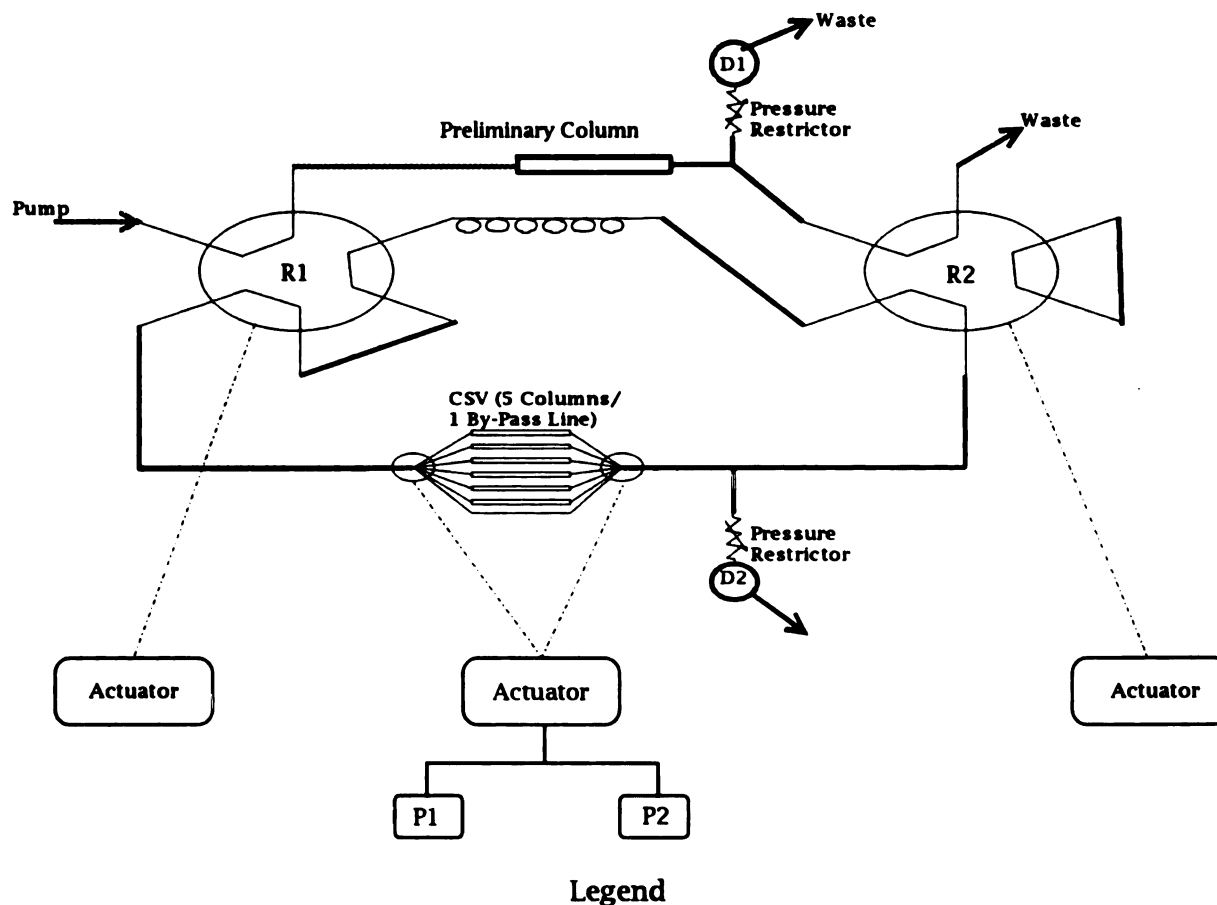
#### **Section 3.3.6.2          Programmable Parallel Interface Board**

An 8255-5A PPI chip was used to control both the solvent delivery system and the CSV. The PPI allows up to 24 digital input/output lines to be used. 2. The PPI is programmable and provides sufficient digital lines to control solvent delivery system (SDS) of the HPLC and the column selection valve (CSV). The 24 lines are gathered into three groups: Groups A and B each are 8 bits wide, while Group C is split into two groups of 4 bits. The chip may be programmed in three different modes of operation.

Pump

Ac

and are  
connect  
signal w  
status c



**Figure 3.4**  
**Complete Valve Arrangement of the Two Rheodyne**  
**Valves, Column Selection Valve, and Pilot Valves**

The PPI is configured so that six lines from Group A were configured as INPUT and are used as the sensors for the position of the CSV. A seventh line from Group A is connected to the Flow Ready signal of the SDS. The SDS sends out a TTL High to Low signal when a programmed flow rate has been reached. This signal is used to sense the status of the SDS. The software pauses data acquisition and re-directs the eluant

through  
indicat

Two o  
connec  
Group  
the RT

connec  
from a  
receive  
slow a  
to ensu  
proces  
latter s

provid

widths  
upon th  
and  $R_2$   
turn, g  
 $R_{2,ext}$   
width :

through a series of by-pass plumbing lines until the signal of the Flow Ready line indicates that the flow is ready.

Six lines from Group B are configured as OUTPUT and used to control the SDS. Two of the lower four bits from Group C are programmed as OUTPUT and are connected to the two Opto-22's which controls the CSV. A third bit from the lower Group C is connected to a 74LS00 chip (a NAND chip). The NAND chip is used to gate the RTI 800 board. This allows the software to pause the data acquisition.

Finally, one of the upper four bits from Group C are setup as INPUT and connected to the SDS as an acknowledgment line (ACK). The SDS sends out a TTL pulse from a monostable<sup>12</sup> chip to indicate that a control word from a computer had been received. Unfortunately, the XT-class computer that was interfaced to the SDS was too slow and proved unreliable to read the TTL pulse. As such, a software pause was used to ensure that the computer sent out control words to the SDS at an appropriate rate to process the command. An AT-class computer was connected to the system during the latter stages of this research and would have been able to read the ACK signal.

Details of programming and using this chip are given in Chapter 6. Table 3.5 provides a synopsis of the digital lines used from the PPI.

### Section 3.3.6.3 Timing Circuit

A 74LS221 dual monostable multivibrator circuit was built to control pulse widths sent to the Opto-22 relays. The circuit generates controlled pulse widths based upon the selection of external capacitors ( $C_{1,ext}$  and  $C_{2,ext}$ ) and external resistors ( $R_{1,ext}$  and  $R_{2,ext}$ ). The 74LS221 is triggered by appropriate TTL signals from the PPI, which, in turn, generates specified pulses to the relay. Variable resistors were used for  $R_{1,ext}$  and  $R_{2,ext}$  to allow the pulse widths to be tuned to the optimum values. Based upon pulse width measurements obtained from an oscilloscope and the measured values of the

resistan

determ

Figure

circuit

Board (

In addi

position

Addres

storage

elsewh

HPLC p

between

Microsc

located

resistances, the capacitances were determined. Equation 3.1 can then be used to determine the pulse widths for different values of R.

$$t_w = R_{ext} C_{ext} \ln(2) \quad \text{Equation 3.1}$$

Figure 3.5 provides the circuit that was built. Discussion on the application of this circuit is presented in Chapter 6.

#### **Section 3.3.6.4      Data Acquisition Board**

An Analog Devices RTI 800/815 Multifunction Analog/Digital Input/Output Board (Norwood, MA) was used to acquire data from the UV-Visible absorbance detector. In addition, this board controlled the two six-port Rheodyne valves and monitored the position of the injection valve. The RTI board contains an 8237A Direct Memory Address controller and a AM9513 counter-timer chip for fast data acquisition and storage. Detailed descriptions of the RTI board and the AM9513 chip are found elsewhere.<sup>13,14</sup>

#### **Section 3.3.7      Interface Box**

An in-house interface box was made to connect the computer to the different HPLC peripherals. Figure 3.6 shows the interface box which provides the connections between the RTI board and the HPLC instrumentation.

#### **Section 3.3.8      Software**

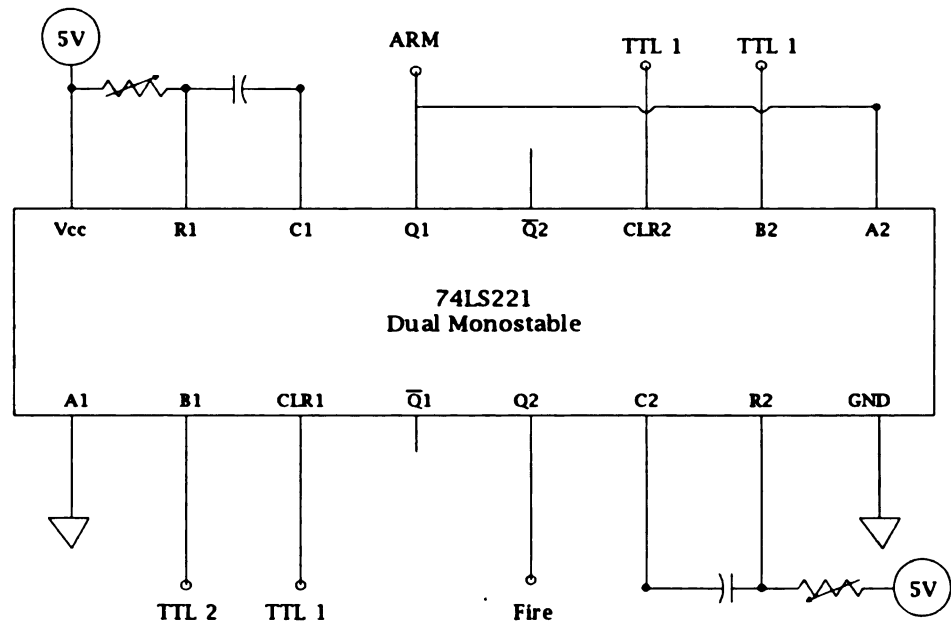
The software to control the HPLC equipment and to acquire data was written in Microsoft QuickBasic version 4.0 (Redmond, WA). Further discussion of the software is located in Chapter 6.



|      |
|------|
| Grou |
| A    |
|      |
|      |
|      |
|      |
|      |
|      |
|      |
| B    |
|      |
|      |
|      |
|      |
|      |
| C Lc |
|      |
|      |
| C U  |
|      |
|      |
|      |

**Table 3.5**  
**Digital Input/Output Lines Used**  
**From the PPI Chip**

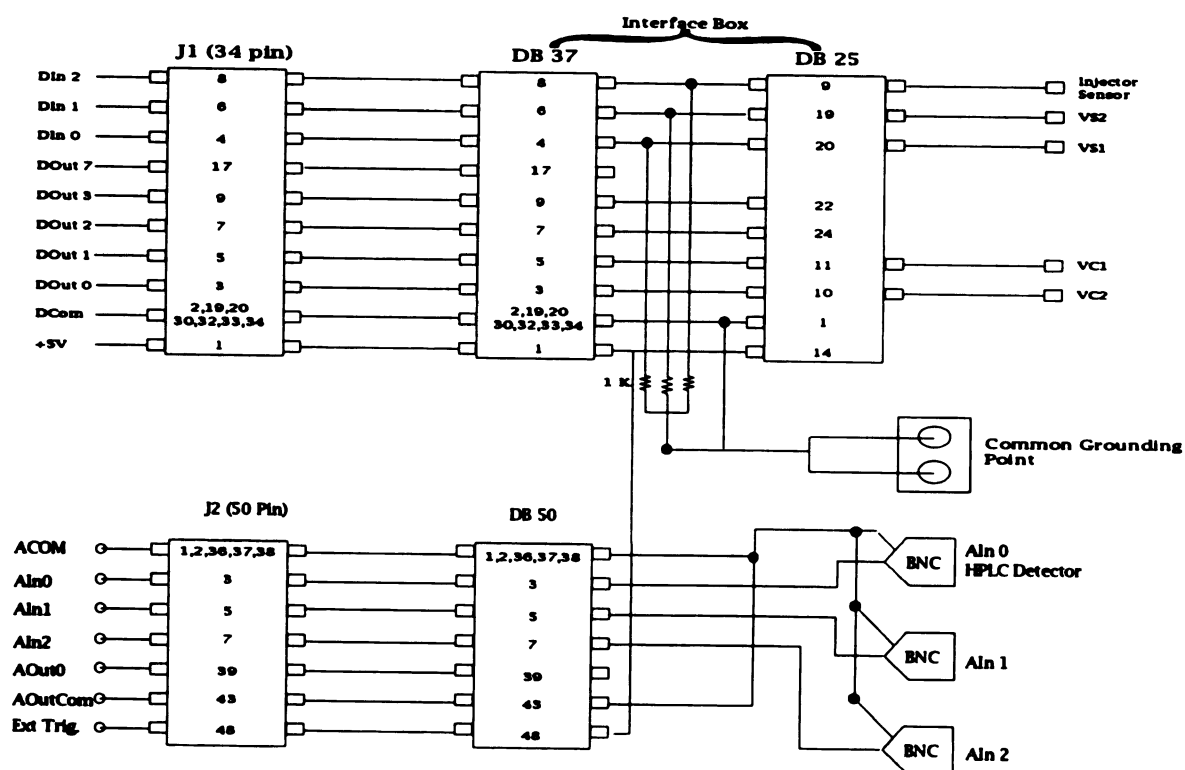
| Group   | Bit | Input/Output | Connection            |
|---------|-----|--------------|-----------------------|
| A       | 0   | I            | Column 1 sensor (CSV) |
|         | 1   | I            | Column 2 sensor (CSV) |
|         | 2   | I            | Column 3 sensor (CSV) |
|         | 3   | I            | Column 4 sensor (CSV) |
|         | 4   | I            | Column 5 sensor (CSV) |
|         | 5   | I            | Column 6 sensor (CSV) |
|         | 6   | I            | Flow Ready (SDS)      |
|         | 7   | I            | not used currently    |
| B       | 0   | O            | SDS Control           |
|         | 1   | O            | SDS Control           |
|         | 2   | O            | SDS Control           |
|         | 3   | O            | SDS Control           |
|         | 4   | O            | SDS Control           |
|         | 5   | O            | SDS Control           |
|         | 6   | O            | not used currently    |
|         | 7   | O            | not used currently    |
| C Lower | 0   | O            | Opto 22 (CSV control) |
|         | 1   | O            | Opto 22 (CSV control) |
|         | 2   | O            | Gate Control (74LS00) |
|         | 3   | O            | not used currently    |
| C Upper | 4   | I            | Acknowledge (SDS)     |
|         | 5   | I            | not used currently    |
|         | 6   | I            | not used currently    |
|         | 7   | I            | not used currently    |



**Figure 3.5**  
**74LS221 Dual Monostable Timing Circuit**

Din 2  
Din 1  
Din 0  
DOut 7  
DOut 3  
DOut 2  
DOut 1  
DOut 0  
DCoin  
+5V

ACOM  
AIn0  
AIn1  
AIn2  
AOut0  
AOutCom  
Ext Trig



**Figure 3.6**  
**Interface Box for the RTI 800/815 Board**

## LIST OF REFERENCES

1 F  
2 S  
S  
C

3

4

5

6

7

8

9

10

11

12

13

14

## LIST OF REFERENCES

- 1 F.A. Settle, Jr. and M.A. Pleva, *Amer. Lab.* 1988, 20(10), 64-79.
- 2 S. A. Tomellini, B. J. Wythoff, and H. B. Woodruff, 'Developing Knowledge-Based Systems' in *Expert System Applications in Chemistry*, B. A. Hohne and T. H. Pierce, eds. American Chemical Society, Washington, D.C., 1989, 236.
- 3 A.P. Wade, S.R. Crouch, and D. Betteridge, *TrAC*, 1988, 7(10), 358.
- 4 Thomas J. Laffey, 'The Real-Time Expert' in *Byte* 1991, 16(1), 259.
- 5 Anne Moose, Terry Schussler, and Dan Shafer, *VP-Expert Rule-Based Expert System Development Tool*, 1989, WordTech Systems, Inc., Orinda, CA, 7.
- 6 L. E. Tway and W. R. Riedel, *PC AI* 1992, 6(4), 52
- 7 Don Shafer, *PC AI*, 1988, 2(3), 54.
- 8 Joan Stoddard, *AI Expert*, 1988, October
- 9 J. Schmuller and Y. Mi, *PC AI*, 1990, 4(3), 52.
- 10 M. C. Harvey and S. D. Stearns, *Anal. Chem.* 1984, 56, 837.
- 11 M. Harvey and S. D. Stearns, 'HPLC Sample Injection and Column Switching' in *Liquid Chromatography in Environmental Analysis*, James F. Lawrence, ed., The Humana Press, 1983, 301-40.
- 12 P. Wiegand, Ph. D. Thesis, Michigan State University, East Lansing, Michigan, 1985.
- 13 *RTI-800/815 User's Manual*, Analog Devices, 1500 Space Park Drive, Santa Clara, CA, 1986.
- 14 *AM9513A System Timing Controller Handbook*, Advanced Micro Devices, Inc., 901 Thompson Place, P. O. Box 453, Sunnyvale, CA, 1986.



EXP

Yo  
ev

## Section

T  
of obtai  
describe  
appropri  
column  
present  
of a Hi  
separat

## Section

describ  
these g

advice

## **CHAPTER 4**

### **EXPERT SYSTEM: COLUMN SELECTION AND VALIDATION**

You think you know when you learn, are more sure when you can write,  
even more when you can teach, but certain when you can program.

Alan J. Perlis

#### **Section 4.1 Introduction**

This chapter describes the dynamics of expert systems. The general processes of obtaining, encoding, and validation of the expert system (called HPLC Advisor) described in this dissertation is presented. HPLC Advisor provides recommendations of appropriate column(s), detector(s), and mobile phases. This chapter presents the column selection module and shows validation studies of this module. Chapter 5 presents the detector and mobile phase modules. Chapter 6 describes the automation of a High Performance Liquid Chromatography system and shows some of resulting separations based upon the recommendations of HPLC Advisor.

#### **Section 4.2 Goals of Research**

Three distinctive objectives for the development of the expert system were described in Chapter 1 and are discussed in more detail in this section. In many ways, these goals determined the final expert system.

##### ***Section 4.2.1 Off-Line Tool for HPLC Methods Development***

The main goal of the project was to develop a tool that would provide intelligent advice for a user. It was decided that personal computers were not powerful enough to

be able to run simultaneously the expert system and the data acquisition/computer control program. Later developments of operating systems such as Microsoft Windows version 3.0 (and higher) have enabled multitasking to be feasible. However, since the expert system and the HPLC controller are separate programs both of the programs can be optimized separately. The expert system can be used independently as a tutorial tool (vide infra). In making a viable expert system, there are several things to consider: 1) the end-user, 2) the type of application (universal or specific), and 3) the level of expertise that the expert system uses to solve the problem. These criteria are discussed below.

#### Section 4.2.1.1      End-User

As mentioned in previous chapters, determination of the end-user is critical. This determination dictates the type of information that must be encoded in the expert system, and the level of knowledge that must be expected from the user. We envisioned two different end-users for this system: a technician who is familiar with some aspects of HPLC, and non-HPLC expert. The latter is an expert in another field (e.g., mass spectrometry) who might, on occasion, have a need to use HPLC. The technician would be familiar with HPLC terminology, and the second type of user would have a good background knowledge of fundamental chemistry. As such, both principal users would be able to answer the questions asked of them by the expert system with little need to obtain answers from outside sources. This expert system provides initial advice which the technician can use to initiate the development of new methods. The expert system will also allow the end-user to better obtain preliminary information about the sample and appropriate method(s) to separate the components when he/she approaches the HPLC expert. By being better informed, the meeting between HPLC technician (or expert) and the end-user should be more informative and productive.

The type of end-user necessitates that the language of the expert system be on a fairly high level; that is, a familiarity with common scientific terminology is expected. However, the expert system does provide additional help than an HPLC expert might require. For example, an initial goal of the system is to determine the suitability of HPLC for a particular separation. A dialog is held between the expert system and the user to determine whether a separation is even needed and/or if a different technique might be better suited (e.g., gas chromatography, GC, or supercritical fluid chromatography, SC). While this information might be essential for the non-HPLC expert, it can be quite time-consuming and aggravating for the HPLC expert who would be using the system. As such, there are stages in the consultation where the user can bypass particular sections, such as the HPLC suitability determination. This feature allows the system to be used by both expert and non-expert alike.

#### Section 4.2.1.2      Type of Application

A review of expert systems that have been used in analytical chemistry shows that there are typically two types : general and specific applications (see references in chapter 2). These two types have specific advantages that were considered when developing the expert system described in this dissertation.

Specific applications are very domain-specific; they provide advice for a sub-area of the application. Examples in the chromatography area include expert systems that:

- determine a method for pharmaceuticals that contain one to five active compounds,<sup>1</sup>
- optimize parameters for systems using a photodiode array detector,<sup>2</sup> and
- determine conditions that will improve the ruggedness of an HPLC method.<sup>3</sup>

These specialized expert systems allow for concise and thorough knowledge bases. Because the applications are specialized, the assumptions of the knowledge and experience of the end-user can be more stringent. Specialized applications typically

allow for fast and small systems to be developed. One obvious disadvantage is that the system is limited in its applicability.

The second type of expert system is the more generalized system. These systems usually attempt to establish a general method or a solution for a broad range of applications. Such systems allow the system to be utilized by more and varied users. However, these applications typically require a larger knowledge/data base and often are not as accurate as the more limited approach. Examples of these systems are mentioned in Chapter 2.

The second approach is used for this project to provide a tool for general methods development and provide only limited advice for esoteric methods. For example, methodology for chiral separations is very complex and is almost a separate area of expertise in itself. When the expert system determines that a separation should be made on the basis of chirality, the user is informed and an initial recommendation for a column may be provided. However, the advice for these types of samples is much more limited than the advice for more routine samples.

#### Section 4.2.1.3 First Principle Approach

Once the user and the application were determined, it was then necessary to determine how to implement this expert system. There are as many ways to implement such an expert system as there are different HPLC experts. Several of the references in Chapter 2 have used different approaches that deserve mentioning here.

In the ECAT project, Abbott et al.<sup>4,5</sup> uses a database approach to propose a method. This approach is demonstrated in a sample session. When developing a method for phthalates, the expert system first searches its knowledge base to see if a method already exists. If the expert system determines that such a method is not available, it then queries the user to see if the phthalate class has been stored in a

slightly different fashion (e.g., a glossary of entries and keywords). If the user indicates that the sample is a new class of compounds, the expert system asks for the sample size to determine column geometries and detector criteria. ECAT then queries for general molecular weight information and size, obtains a specific name of the compound(s), and determines the acid-base characteristic of the compound class. From this approach of categorizing the compound into a particular class, ECAT develops the method.

In contrast to ECAT, the expert system in this project utilizes a first principles approach. The user must answer more questions about the chemical and physical nature of the compound. This approach allows the method development to be based upon the chemical interactions of the analyte with the column. The present approach requires many more rules to determine the method of separation.

Most of the different expert systems presented do not present the use of confidence factors as part of the system. The expert system in this project must be able to solve a wide variety of problems, but it is not optimized for a particular sub-set of chromatography applications. So that the user might apply external criteria, by either supplying additional information or by consulting with an HPLC expert, it was decided that confidence factors (CNF) must be an integral part of the system. By presenting all of the methods that pass a truth threshold (a value established by the knowledge engineer that determines the minimum value of the CNF before a rule is considered to be true), the user is able to select alternative methods based on other criteria that might not be incorporated in the expert system (e.g., the best column has been loaned to another research group and is currently unavailable). Based on the CNF of different solutions, the user is able to evaluate the reliability of the proposed method(s).

ECAT never selects methods based on normal phase and hydrophobic interactions modes and is biased towards  $C_{18}$  columns.<sup>4</sup> The expert system in this

dissertation is more generalized in the early stages of the methods development. The opportunity to pursue several different lines of inference is provided before eliminating any competitive techniques.

#### **Section 4.2.2      *Tutorial and Training Tool***

Expert systems have been predicted to become ubiquitous and transparent.<sup>6</sup> Many systems will become embedded in different instruments to provide system diagnostics/monitoring and to serve as on-line documentation.<sup>6</sup> Another use for an expert system will be as a tutorial and training tool. Expert systems become a very powerful teaching tool when they are linked with simulators.<sup>7</sup> Schaefer et al.<sup>8</sup> proposes that a developed expert system can be converted into a computer-based training program. With such in mind, Schaefer and co-workers developed ACQUIRE-ITS, a program that converts a completed knowledge base into an intelligent training system (ITS). Two modes of training are possible, *learn* and *explore*. The learn mode presents a series of lessons to the student and evaluates the progress of the student, applying expert system techniques to assist in the learning process. The second mode, the explore mode, allows the student to inspect the knowledge base and to run the student's own scenario to compare the results from the expert system with the student's predicted answer.

While the *learn* mode is beyond the scope of the research in this project, the *explore* mode was determined to be a suitable application for the expert system in this dissertation. To achieve this goal, detailed text files were incorporated into the expert system to provide feedback to the user. In general, to achieve the sub-goal of using this expert system as a training tool, much more explicit help and text was provided. Incorporation of this expert system, besides using a tool such as ACQUIRE-ITS to convert it into a intelligent training system, into a course could be achieved by giving students

different systems to predict a method and then comparing their results to those obtained from the expert system.

### ***Section 4.2.3      On-Line Tool for Automated Intelligent HPLC Instrumentation***

The final objective of this research is to develop a complete intelligent HPLC instrument in which a method (and possible alternatives) would be developed during a consultation with an expert system. The controller program would perform the proposed method and evaluate the results of the HPLC run (with respect to resolution, peak shape, and so forth) and consult with another expert system to further optimize the method.

While this objective is envisioned to be possible, it was deemed not feasible at the current time—predominately due to computer requirements. The DOS environment is only a single-tasking environment and is too constrained in terms of memory to make possible a complete system that would have different programs running simultaneously. Since VP-Expert is a DOS program, the current status of this project is that the expert system is run off-line from the controller software (albeit on the same computer as the one that controls the HPLC instrumentation). The methods generated from the consultation with the expert system are then manually entered into the controller program. Feedback from the HPLC run is not passed back into the expert system for further optimization; however, such a feature is possible with appropriate additions to the knowledge bases.

## **Section 4.3      Knowledge Acquisition**

Two different information sources were used to develop the expert system in this dissertation. The preliminary information was obtained from the resident HPLC



expert during knowledge acquisition meetings and then supplemented with information obtained from both the chromatography literature and from HPLC manufacturers' literature. The manufacturers' literature was used primarily in the develop of the ancillary data bases to provide the user with information about the different columns that were available.

#### ***Section 4.3.1      Personnel and Roles***

The group composed to create this expert system consisted of the chromatography expert, the champion, knowledge engineer, assistant knowledge engineer, and two users. The chromatography expert was Dr. V. L. McGuffin who provided the heuristics that were encoded into this expert system. The champion, Dr. S. R. Crouch, provided additional support and experience. Dr. A. P. Wade, a postdoctoral associate in this research group, served as the knowledge engineer and provided experience and guidance in the early stages of the project. Dr. Wade also trained this researcher to continue with the project after his departure. The two users assisted in testing and evaluating the initial prototype; in particular, they evaluated the user-friendliness and lucidity of the system. Finally, this researcher served as the assistant knowledge engineer in the early stages of the project and has continued the project in conjunction with the automation of the HPLC instrumentation.

#### ***Section 4.3.2      Knowledge Acquisition Session Format***

Dessert<sup>9</sup> provides a case study in elucidating information during a knowledge acquisition meeting. A similar approach for the expert system in this dissertation was used. These meetings consisted initially of the knowledge acquisition team meeting at a scheduled time for approximately 1 to 1.5 hours. These preliminary meetings took the format of brain-storming sessions in which general perceptions of the intended expert

system were discussed. It was in these meetings that the goals and end-users of the expert system were formulated. As the focus of the expert system was generated, these meetings then took on the additional role of eliciting the HPLC heuristics that were incorporated into the expert system.

These knowledge acquisition meetings were recorded on audio tape and then were transcribed into minutes so that the information could be extracted and placed into the knowledge base of the expert system. Initially, one of the secretaries in the chemistry department transcribed the minutes from the meeting. Much of the information was lost when the secretary attempted to not only understand the vernacular of HPLC, but to also decipher the important information from residual environmental noises (e.g., coughs, chair scrapings). To improve the knowledge acquisition process, it was decided to use a chemist (in particular, this researcher) to transcribe the audio recording of the meetings. In this fashion, deductive reasoning could be used to extrapolate the information from background clutter. To prevent bias, a strong attempt was made to perform a verbatim transcription with no editing at this stage. To insure accurate transcriptions, the tapes and minutes were checked by the other knowledge engineer.

The third stage of the meetings involved the analysis of the expert system. As the heuristics were encoded, incompleteness and misunderstanding of the information were assessed by the knowledge engineers. Questions based on the analysis of the information were then compiled and included with a copy of the transcripts. The transcripts were then distributed to the acquisition team before the next meeting to evaluate the information. During the subsequent meetings, these questions were then discussed and solved. In Table 4.1, excerpts from one of the acquisition meetings are shown. Two general types of questions were presented. The first type of question involves elucidation of material that might not have been understood from a prior

meeting. These usually take the form of "Please explain this statement you made at the previous knowledge acquisition meeting." The second type is attempts to obtain a clearer understanding of the heuristics. As mentioned in Chapter 2, experts often do not consider or vocalize the thought processes that were used to derive their advice. This second type of question attempts to direct the expert at providing more succinct and clarified reasons behind the advice. The expert is directed to define the heuristics and discuss the interactions and ramifications with other heuristics. This questioning will often lead to the development of further heuristics. The knowledge engineer should interactively ask three questions of the expert:

1. What do you do next?
2. What does that mean?
3. Why do you do that?

The last stage of the meetings involved evaluation of the prototype expert system to determine if the system was concise and logical. The members made notes while running the system and reported these concerns at the next meeting. These concerns were then addressed as either limitations of the software or were corrected in the next version of the expert system.

**Table 4.1**  
**Excerpt from Knowledge Acquisition Meeting**

**Meeting No.3 - 19th February 1986**

**Those Present;**

**S: Stan Crouch      V: Vicki McGuffin      A: Adrian Wade      C: Christine Evans**  
**K: Kim Ratanathanawongs   M: Stephen Medlin**

**Aim of Meeting: To determine answers to the above attached questions and to determine corrections/errors to the expert system in its present condition.**

**QUESTIONS**

- 1. What solvents should / should not be used with phenyl columns ?**
- 2. You suggest use of gas chromatography in preference to HPLC whenever possible. Would you go as far as to suggest use of solid liquid chromatography whenever geometric isomers are present ?**
- 3. What system should be used for "polarity only" separations ? (Silica Gel ? - Is this the same as "silica"? - Why not alumina too ?)**

---

**V: Another stationary phase that has been used for geometric isomers are some of the charge transfer complexes. They selectively form complexes with certain of the geometric isomers which otherwise might not be electronically suited to bind to the charge transfer phase and that is another approach.**

**A: What mode of HPLC would you term that?**

**C: Obtuse!**

**V: Obtuse isn't bad. How do you want it divided? In other words which categories do you want to use?**

**A: Is it reverse or normal phase for a start? Like...**

**V: It can be either one because the interaction in this case is an electronic interaction. And so you could do it in normal or reverse phase like cyano. Cyano phase is an electronic interaction.**

**A: Would you... Now in terms of molecular size--large or small?**

**V: Could be either one but probably smaller because the stationary phase that you are dealing with has one group that has been designed to interact with this molecule and it might react with a part of that molecule but I would say in general it would be small.**

**A: And which solvents would you use?**

---

### ***Section 4.3.3 Additional Sources of Information***

As a forewarning, knowledge engineers must not assume that they can omit the human expert and go to the literature directly for the purpose of developing the expert system. Most of the literature is, by nature, fairly terse. Many of the heuristics that were applied by the authors of scientific papers in developing a particular method are not explained in the literature. Therefore, it is can be difficult to develop an expert system based solely on the literature; such a system could become simply a data base of methods. While the data base would have utility in itself, it is desirable to have the expert system be able to extend the principles used in the literature to different applications. In this fashion, the expert system appears to be intelligent and can provide recommendations for applications that might be different from the original. This type of expert system can only be developed when the heuristics of the original method can be obtained and encoded to allow the expert system to apply first principles in extending the recommendations to other areas.

#### **Section 4.3.3.1 Types of Data Used**

The first type of data used was different journal articles to determine some of the different uses for size exclusion chromatography. The prototype expert system was based upon a Waters Chromatography flow chart<sup>10</sup> (Figure 4.1) which provided a logic chart to determine which type of column to use when developing a method. Figure 4.1, adapted from Waters, does not include the different packing materials, such as amine or phenyl columns. The first stage is to classify the solutes based upon their average molecular mass, which provides information on the required pore sizes of the stationary phase. The solubility of the sample will indicate the type of chromatography (normal- versus reversed-phase) that will be used. In the Waters Chart, the ionic nature

of the compound(s) is used to discriminate between columns based on separation of ionic species and the remainder of the columns.

This chart discriminated against both size exclusion chromatography (SEC) and also many forms of normal phase chromatography. Snyder and Kirkland's flow chart<sup>11</sup> emphasizes using SEC as a both preliminary sample clean-up technique and also a method to fractionate the sample. As the final objective of this research includes column switching, rules based on principles of SEC were incorporated into the expert system. In addition, Figure 4.1 does not include any additional criteria that is used to choose between the different recommendations (e.g., Reversed-Phase with Ionization Control, Reversed-Phase IPC, IEC, or Ion Chromatography for small ionic compounds). Each of these species will be more appropriate for different species than the other. Finally, this figure does not indicate any criteria that is used to choose the different packing materials (e.g., phenyl, amine based stationary phases).

The prototype expert system also did not differentiate between the different methods based on ionic properties. The prototype expert system suggested that the pH be changed to shift the ions into the neutral phase, if possible, or that generic ion chromatography be used. Literature on ionic separations was read and incorporated into the expert system. This allowed a more detailed recommendation on when to use ion chromatography; to be used even when the solute might be shifted into the neutral phase. In particular, if there are no other separation modes available, or if the solute(s) are highly ionic (or multiply ionic), it might be advantageous to use one of the ionic techniques such as Ion-Pair Chromatography (IPC), Ion Exchange (IEC), or Ion chromatography (IC). The knowledge base was expanded to provide rules to determine which of the ionic techniques would be appropriate (based upon the relative charge, size, and type of ion). These rules then are kept in consideration and are weighted with the appropriate range of confidences. This allows the expert system to consider all

methods throughout the complete consultation. Elimination of methods is only done at the end of the consultation when criteria such as not having a cation exchange column or not having a detector that responds to ions precludes using ion chromatography. (As an aside, this approach of eliminating only at the end of the consultation is also used for the other modes of separation.)

There are several different specialty columns that are available, such as columns that separate specifically proteins, carbohydrates, and chiral compounds. Since these are fairly specific in nature to the manufacturers, literature from HPLC companies was used to provide several rules dealing with these columns. The user is informed about the presence of such columns and asked to do further reading to use these methods. It must be recalled that one goal of this expert system is to remain as generalized as possible. Thus, the expert system does know of these columns and knows when to recommend them, but the advice on their use is kept limited. The reasoning behind this decision is that a complete expert system could be developed based on a single one of these areas of chromatography, in particular, the separation of geometric and optical isomers. Therefore, the information on these columns is limited in nature to avoid an excessively long and slow consultation with the user.

Finally typical column parameters such as the average number of theoretical plates, maximum operating pressures and pHs, and so on, have been obtained from the manufacturers literature of representative companies and placed into a data base. The motivation behind this was to be able to suggest different columns that might solve a particular problem where the user does not have that specific column in his/her laboratory. In this fashion, the user can be directed to at least one company to determine whether the purchase of a particular column might be warranted. In addition, the user may then provide the parameters of the columns that are currently

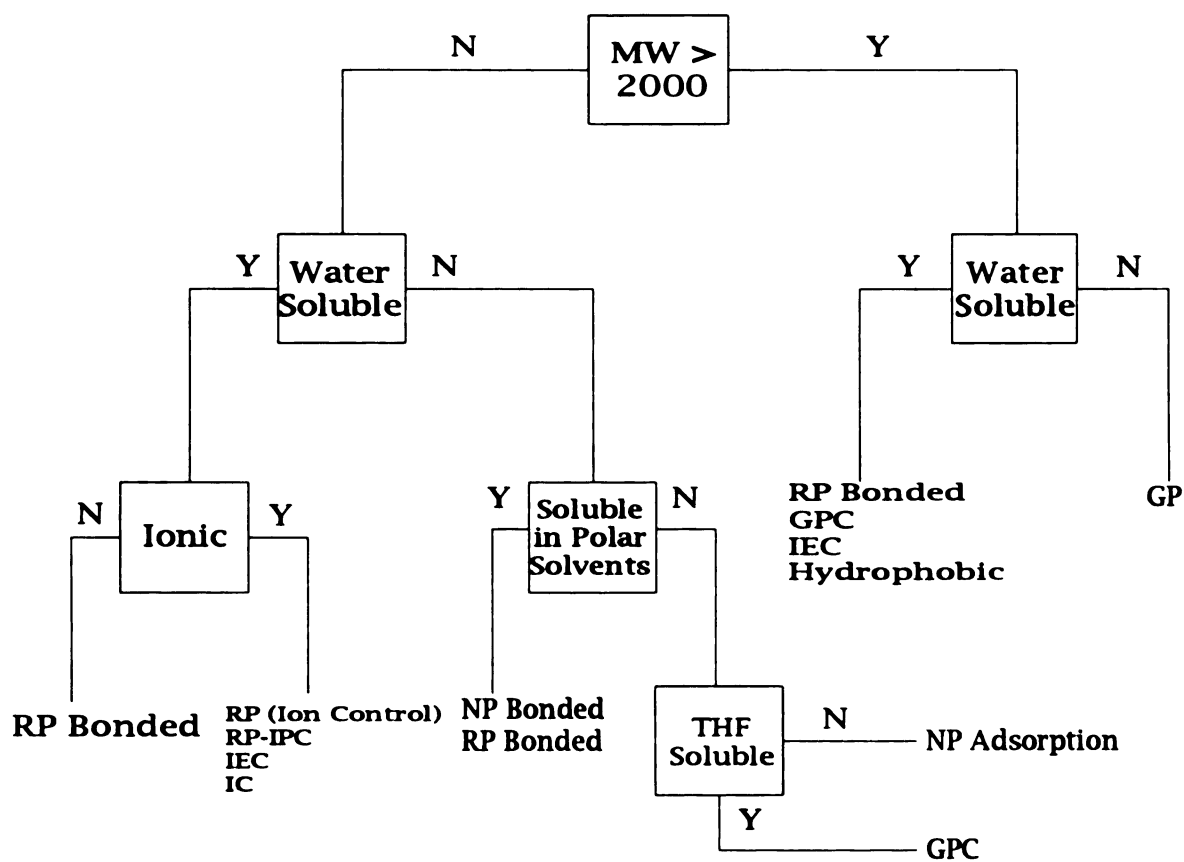
available in the laboratory. This allows the expert system to provide much more detailed and tailored recommendations for that laboratory.

## **Section 4.4     Action Block**

VP-Expert is written such that the highest level of code is the **ACTION BLOCK**. The Action Block is where the objectives of the expert system are specified. The Action Block also contains the code that establishes the user-interface, any graphical constructs, and so forth. The rest of the knowledge base consists primarily of the rules and facts. An abridged copy of the Action Blocks used in this expert system is presented in Tables 4.2 and 4.3. (A complete listing of the knowledge base is available in a separate document and can be obtained by contacting either the laboratory or the author of this dissertation.) The following sections will reference this section of code to describe the specific goals and approach of this expert system.

The earlier versions of the expert system were broken into two different modules, due to VP-Expert's limitation in the utilization of available memory. With the current version of VP-Expert, both modules could be combined into one larger knowledge base; however both modules are maintained separately to improve the performance of the expert system. The two modules were divided logically into the first module, which determines the separation mode (e.g., size exclusion chromatography) , and the second module, which accesses the data base to present particular columns, solvents, detectors, and so forth. In this fashion, the second module serves primarily as the user-interface for the recommendations. This will allow different presentations to be used by changing only the second module.





|     |                               |        |     |                             |
|-----|-------------------------------|--------|-----|-----------------------------|
| RP  | Reversed-Phase                | Legend | NP  | Normal-Phase                |
| IPC | Ion Paired Chromatography     |        | IEC | Ion Exchange Chromatography |
| GPC | Gel Permeation Chromatography |        |     |                             |

**Figure 4.1**  
**Flow Chart for Choosing an HPLC Column**

**Table 4.2**  
**Action Block for First Module**

**ACTIONS**

```

Max_pH_% = 8
Min_pH_% = 2
  <lines deleted>
FIND HPLC_Suited
FIND Scuzz_Checked
FIND Homogeneity_Checked
FIND Continue
Find Lo_pH_%
FIND Hi_pH_%
SAVEFACTS cont_lcx

```

**Table 4.3**  
**Action Block for Second Module**

**ACTIONS**

```

Loadfacts cont_lcx
Display "Module 2 Loaded"
Loop_Count = 1
WHILETRUE  Loop_Count <= (Maxcounter) AND
           Mode <> Not_Checked OR
           Mode <> UNKNOWN
    THEN  DISPLAY "Advice results are #{Loop_Count} {Advice[Loop_Count]}"
          Loop_Count = (Loop_Count + 1)
    END
SORT Mode
display "The values of mode are: {#Mode}"
The values of packing are: {Packing_TYPE}"
Pop Mode,Mode_tmp
WHILETRUE  Mode_tmp <> Unknown and
           Mode_tmp <> Not_Suited and
           Mode_Tmp <> Not_Checked
    THEN  DISPLAY "The columns chosen for {#Mode_TMP}:"
          FIND Column
          RESET Mode_tmp
          Pop Mode,Mode_tmp
    END
Savefacts cont-lc2;

```

In Table 4.2, the first several lines are facts for the expert system. For example, the maximum allowable pH is set to 8 (**Max\_pH\_% = 8**). This value is used in conjunction with other rules to determine whether the pH required to neutralize the ionic species is outside of the operating parameters. This fact is therefore used also in determining the appropriate column for ionic species. These values are easily changeable and allow the total knowledge base to be edited by changing the facts. Other facts (not shown) are used to improve the readability of the expert system by using mnemonics rather than numbers (e.g., for pointers in arrays).

A second feature of VP-Expert is the use of **PLURAL** variables. A plural variable is similar to a memory stack in that multiple values may be assigned to the variable. In this fashion, one or more results may be placed into the variable and multiple conclusions can then be drawn. While all values of the plural variable are simultaneously available, it is necessary to **POP** the plural variable when manipulating a single value of the variable.

#### ***Section 4.4.1      Determination of Suitability***

The first goal of the expert system is to determine whether the problem is amenable to an HPLC solution. HPLC can be an expensive technique to employ—both in terms of cost, time, and complexity. If the sample can be separated or analyzed without using HPLC, the expert system suggests that the user seek an alternative technique.

The first series of questions determines if a separation is required via four questions. These questions ask if the sample could be assayed by a physical technique such as a visual determination, weighings, and so forth. The expert system may suggest that a separation is needed when the user knows only the motive for a potential separation. For example, if the user must purify large quantities of a sample in which

he/she knows the components, then a separation is needed (i.e., Prep-Scale chromatography).

An example is shown in Code 4.1. This code example demonstrates two different expert system principles. The first principle is the ability of the expert system to deal with incomplete or missing data. In this code segment, if the user has indicated that he or she does not know if a direct analysis is possible (**Direct\_Analysis = Unknown**), the expert system can apply additional information to determine whether to proceed with the consultation. The second principle is the use of confidence factors (e.g., CNF 90). The use of a confidence factor reflects the lack of information regarding direct analysis of the sample.

### Code Example to Determine if Separation Needed

**Rule Sep\_Needed3**

**IF      Direct\_Analysis = Unknown OR  
          Direct\_Analysis = No AND  
          Separation\_Reason = Analytical OR  
          Separation\_Reason = Preparative  
 THEN Separation\_Needed = Yes CNF 90;**

Code 4.1

If the sample must be separated, the expert system uses twenty-four rules to ascertain the immediate chemical and physical characteristics of the sample that preclude or necessitate the use of HPLC. If the sample is unstable (labile or reactive compounds such as explosives or large, fragile proteins). The expert system will assist the user in determining the "values" to these questions by asking the user additional questions to elucidate information and by presenting a list of labile components that typically must be separated by HPLC rather than GC. The expert system also determines if the boiling point of the sample is relatively high (greater than 100°C). A sample that has a low boiling point is better suited to GC or SFC.

The final step needed in determining the suitability of HPLC is to determine if a detector is available and capable of responding to the analytes. Eighteen questions are used to determine the answer to this question. Codes 4.2 and 4.3 demonstrate the determination if the detector chosen by the user is suitable for the sample. Code 4.2 is the "master" rule that is used to determine the overall suitability of HPLC. Within this rule are two sub-goals, **Detector\_Suitable** and **Correct\_Conc\_Range**, which must be determined before assigning a value to **HPLC\_Detector\_Suitable**. The first sub-goal assists the user in choosing a detector suitable while the second goal insures that the detector will respond to the concentration range of the analyte.

Code 4.3 determines if a detector is available for ionic compounds. This rule will assign "Yes" to **Detector\_Suitable** only if the analyte is ionic (and cannot be shifted into its neutral form via an appropriate change in the pH) and the species can be detected either by an ionic-sensitive detector or by a detector that can detect the non-ionic form. The BECAUSE part of the rule is used by the expert system (upon request by the user) to generate a report of the session. This report will contain the text of the BECAUSE clauses for every rule that passed. In this fashion, the user can then determine the induction path of the consultation. During the consultation, the user can either query the inference engine to the reasoning behind a question ("WHY") or request information on how a value was assigned to a variable such as **Detector\_Suitable** ("HOW"). The inference engine will respond with the text from the BECAUSE clause. If a rule does not have a BECAUSE clause, the inference engine only displays the rule that is being asked or the rule that assigned the value to the variable.

Code 4.3 again shows the interrelated nature of data. The chemical nature of the analytes (e.g., ionic) is used to solve different problems. In this case, it is used to determine the suitability of HPLC with regards to detectability and to choose a column. The inference engine can make inductions based on data obtained from earlier

questions and apply them to additional goals. A second example of this is the earlier determination of whether GC or SFC should be precluded. If the user needs assistance in determining if the sample is stable and the user chooses proteins from this list (or as an answer to other questions asked by the expert system), the rules are written such that this information is stored and used in determining an appropriate column and separation method. By using results from previous goals, the expert system exhibits faster consultations and appears to be intelligent by applying data used in the determination of one goal (e.g., determining if HPLC is suitable) to the determination of an appropriate column.

At any point during the consultation, the user may choose to by-pass this module or to override the suggestion from the expert system by two different approaches. The first approach explicitly asks if the user wants to ascertain the suitability of HPLC to the problem. Since the end-user for this expert system will be both novice and semi-expert users, the user is given the choice to have the expert system ignore the HPLC suitability determination. The second method explicitly asks if the user knows of any reason NOT to use GC or SFC. This method allows the user to force the consultation to conclude that HPLC is suitable.

### Code to Determine if Detector is Suitable

```
Rule Detector_Suitable1
  IF      Detector_Suitable = Yes AND
         Correct_Conc_Range = Yes
  THEN HPLC_Detector_Suitable = Yes
  ELSE  HPLC_Detector_Suitable = No;
```

Code 4.

### Code to Determine Detector For Ionic Compounds

```
RULE Detector_Suitable5
  IF      Ionicity = Ionic AND
         Detect_Ions = Yes OR
         Detect_Direct = Yes
  THEN Detector_Suitable = Yes
  ELSE  Detector_Suitable = No
  BECAUSE "
  If your solute is ionic and you cannot shift the solute into its
  nonionic form then you must be able to detect the ionic form.";
```

Code 4.

### ***Section 4.4.2      Determination of Sample Pre-Treatment***

The next stage of the consultation determines if any sample pre-treatment is required. The current implementation of the expert system uses twenty-one rules to provide assistance. The expert system determines if there are obvious problems with the sample that would cause either system failure (e.g., obstructing the column or scoring the instrument components). The system assists in determining which layer of a bi-, tri-, multilayer to sample. The current implementation also suggests simple steps such as varying the pH to shift the species of interest from one layer to the desired layer. Additional rules prompt the user to determine if changing parameters such as pH will cause degradation or loss of speciation of the analytes.

### ***Section 4.4.3      Determination of Class of Separation***

There are two major types of classes of chromatographic separations, normal- and reverse-phase. Simply, normal-phase chromatography uses mobile phases that are less polar than the stationary phase and reverse-phase uses the opposite arrangement. While it might seem fairly simple to determine the type based on the solubility of the analytes, the choice of separation type will dictate the choice of column to use. For solutes are only soluble in polar or non-polar mobile phases, the choice is simplified for the user. For the solutes that are soluble to some extent in both classes of mobile phases, it becomes necessary to choose the most appropriate one. Reverse-phase (RP-HPLC) is the more common mode of operation, but there are many normal phase modes (NP-HPLC) which provide unique resolving powers.

For the case in which the solute is soluble in only one type of mobile phase, the expert system chooses that type of chromatography (i.e., normal-phase). For the other cases, the expert system uses solubility information in conjunction with additional rules to choose an appropriate column. Either method is not excluded at the early

stages of the consultation, and the final choice of the values of Phase (Normal versus Reverse) do not exclude any column until the final data are input into the computer. Then the information on the mode is combined with other information to make the final recommendation. A total of sixteen rules are used to determine the phase. These rules use both solubility data and separation properties to determine the appropriate phase.

Code 4.4 provides an example of a rule that determines if reversed phase is appropriate. In this example, for solutes that differ by both polarity and size and there is no reason to preclude using reverse-phase, reverse phase is appropriate. The confidence factor is lowered to represent the possibility of using normal phase chromatography for polar components. The individual modes of RP-LC (e.g., cyano-based RP-LC) are assigned additional confidence factors at a later stage during the consultation to reflect a discrimination between the modes.

Two different types of data are used to determine the mode. The first type involves information of the differences between the components of the sample (the Separation\_Property variable). This information has been obtained from twenty-seven earlier rules. These data will be combined with the mode (reversed versus normal phase) and additional rules and data to determine the correct column in the next stage of the consultation.

The second information type is represented by the Preclude\_Rev\_Phase (and the Preclude\_Normal\_Phase) variable. The values of these variables are assigned based on incompatibilities and solubilities. The expert system focuses on the solubility aspect when assigning a value to these two variables. It is envisioned that the future version of this expert system will use this variable to determine the compatibility between different columns and mobile phases when implementing multidimensional chromatography via column switching.



### Code to Determine the Appropriate Phase

**RULE Phase11**

```

IF      Separation_Property = Polarity AND
        Separation_Property = Size AND
        Preclude_Rev_Phase = No
THEN    Rev_Phase_OK = Yes CNF 80;

```

Code 4.4

#### ***Section 4.4.4      Determination of Column and Mode***

As the final version of this expert system is envisioned as a tool to assist in the development of methods for samples which might require column switching, the choice of columns is the principal objective for this expert system. There are two sub-goals that are solved in the process of choosing a column. The first sub-goal is the determination of mode. In the vernacular of the expert system in this project, mode refers to the separation mechanism and incorporates the results from the assignment of normal- and reverse-phase techniques. The modes that are assigned in this expert system are given in Table 4.4. The mode name is given in the first column to serve as a glossary when discussing the knowledge base code. The second column gives a description of the mode, and the third column gives the additional sub-modes that are ascertained by the expert system.

Both implicit and explicit information is used in deriving a recommendation for the column. Implicit information results from data that were used to solve earlier objectives. For example, one rule used in the determination of HPLC as being suitable determines if an HPLC detector is available for the detection of ionic species. This data is then applied to determine the choice of column. There are also twenty-seven explicit rules that are used to choose between normal and reversed phase chromatography. These rules also assist in the choice of mode.

There are also forty-rules that are used to specifically select the correct mode. Several physical and chemical characteristics were used during this stage and each is presented below.

#### **Section 4.4.4.1      Physical Properties**

The first characteristics are the physical size and shape of the analyte(s). One unique approach taken by this expert system is the use of confidence factors to describe the priority of size as a discriminating factor among the components of the sample. In particular, this expert system attempts to assign a confidence factor that describes both the size distribution of the analytes and the presence or absence of competing separation effects, such as polarity.

The distribution of sizes is determined using forward-chaining rules. A forward-chaining rule is implemented in VP-Expert by the use of WHENEVER rules. When the value(s) of the premise(s) change, WHENEVER rules are activated. WHENEVER rules are analogous to computer interrupts in that they take over immediate control of the program. In this fashion, the logic flow of the expert system can be changed to reflect the appearance of new data. Codes 4.5 and 4.6 are two of the twelve rules that establish the distribution of size and calculate a confidence factor based upon the distribution.

The rule in Code 4.5 is activated whenever the size distribution (Wt\_Range\_%) lies between 0.10 and 0.15. This rule assigns to the variable, Solute\_Size\_Distribution, the value of Wide\_Distribution with a confidence factor equal to the value of Dist\_CNF\_%. Both Wt\_Range\_% and Dist\_CNF\_% are assigned by a previous rule, as seen in Code 4.6.

**Table 4.4**  
**Possible Sub-Modes that may be**  
**Recommended by the Expert System**

| <b>Name of Mode</b> | <b>Description</b>                     | <b>Packing Materials</b>                              |
|---------------------|----------------------------------------|-------------------------------------------------------|
| NP_Ads              | Normal-Phase Adsorption                | Silica<br>Alumina                                     |
| NP_Bonded           | Bonded Normal-Phase                    | Amine<br>Cyano<br>Phenyl                              |
| Geometric/Optical   | Geometric and optical isomers          | Various materials                                     |
| NP_SEC              | Normal-Phase Size Exclusion            | Various materials based on size of analytes           |
| NP_IPC              | Normal-Phase Ion-Paired Chromatography | Special materials designed for NP-IPC                 |
| Rev_Bonded          | Bonded Reverse Phase                   | Alkyl-Silanes (e.g., C18)<br>Amine<br>Cyano<br>Phenyl |
| Rev_SEC             | Reversed-Phase Size Exclusion          | Various pore size polymers/gels                       |
| Rev_IEC             | Ion-Exchange Chromatography            | Anion and Cation Exchange Resins                      |
| Rev_IPC             | Reversed-Phase Ion Pair                | C <sub>18</sub> columns<br>Specialty columns          |
| Rev_Ion_Chrom       | Ion Chromatography                     | Anion and Cation Exchange Resins                      |
| Peptide_Rev_Bonded  | Speciality columns                     | Peptide Specialty Column                              |
| Glycophase_CPG      | Speciality columns                     | polyaminoacids/peptides specialty columns             |
| RP_Bonded_Phosphate | Speciality columns                     | polyaminoacids/peptides specialty columns             |

Whenever the expert system determines that size is a discriminating property, Whenever Size\_Distribution1 is activated. The knowledge base is set up such that multiple rules might determine that the separation can be based upon size; therefore the Size\_Fired variable insures that this rule will only fire one time (because Size\_Fired is assigned the value of Yes after the first time that the rule is accessed). Upon activation of this rule, the weight distribution is calculated.

According to Snyder and Kirkland,<sup>11</sup> size exclusion chromatography is most effective for components that have distributions greater than 15% and is least effective for components with a distribution under 10%. Therefore, Dist\_CNF\_% is calculated to provide a confidence of 100% for components with a distribution greater than 15% and to give a confidence of 80% for a distribution of 10%. As is discussed in Section 4.4.9.1.2 (Page 149), confidence factors must be chosen such that the confidence is not dropped to zero as it propagates through the knowledge base. With this in mind, a distribution of 10% was given the relatively high confidence value of 80%. This confidence factors prevents size exclusion from being rejected at too early of a stage in the consultation.

This distribution confidence factor is also combined with additional rules to determine the selection based on size. These rules are discussed in greater detail in Section 4.7 and 4.8, validation of the M1 prototype and the VP-Expert based expert systems.

### Code to Calculate of Size Distribution

```
WHENEVER Size_Distribution5
IF    Wt_Range_% < 0.15 AND
      Wt_Range_% > 0.10
THEN Solute_Size_Distribution=Wide_Distribution CNF Dist_CNF_%
```

Code 4.

### Code to Calcule Confidence Factors for Size Distribution

```

WHENEVER Size_Distribution1
  IF      Separation_Property = Size AND
          Size_Fired = No
  THEN Wt_Range_% = ((Hi_Wt_% - Low_Wt_%)/Hi_Wt_%)
          Dist_CNF_% = (400 * Wt_Range_% + 40)
          Wt_Avg_% = ((Hi_Wt_% + Lo_Wt_%)/2
          Size_Fired = Yes;
  
```

Code 4.

#### Section 4.4.4.2

#### Chemical Properties

The significant property used to determine the appropriate mode and column is the chemical nature of the solutes. The chemical nature of the solute determines the interaction between the solute and the mobile and stationary phases.

There are several different sets of rules that are used in determining the appropriate column. The first set of rules is labeled with the rule name of Separation\_Property and ascertain the chemical distinction(s) of the different solutes. For example, a rule determines if there is a distinction in the polarity of the components of the sample. There are two approaches taken in solving for the variable, Separation\_Property. The first approach allows the user to select directly the different property(ies) in a single question. This approach allows the user to access more quickly the recommendations from expert system without having to answer additional questions. If the user selects SEEK from the choices offered, the second approach uses additional rules and questions to determine these properties. This second approach provides supporting text to assist the user in choosing the appropriate properties. Codes 4.7 and 4.8 demonstrate these two types of approaches while Figure 4.2 provides an example of a dialog that the user will see if the hypotheses to these rules are true. The text in the two parentheses are comments to the reader and are not part of the knowledge base. The POP function provides the ability to sequentially access the multiple values that may have been assigned to the PLURAL variable

Separation\_Assessment. If the user chooses UNKNOWN to the question, rule Separation\_Property2 will force the user to choose SEEK. Once the user selects SEEK, additional rules assist the user in determining a response to this question. The rule in Code 4.8 provides an example of rules that may be used to obtain additional information from the user. There are approximately forty-three rules that are used to establish the separation property(ies). These properties include the ionic nature, acidity/basicity, the presence of different functional groups, and so forth. Many of these chemical properties are discussed in further detail below in the choice of Mode and Column.

### Code to Select Chemical Properties Directly from Menu

```

RULE Separation_Property1
  IF Separation_Assessment <> UNKNOWN AND
    Separation_Assessment <> Seek
  THEN Separation_Property =(Separation_Assessment)
    POP Separation_Assessment,Separation_Property
    WHILETRUE Separation_Assessment <> Unknown THEN
      POP Separation_Assessment,Separation_Property
    END
    Stop_Sep_Prop_Cycle = Yes;

RULE Separation_Property2
  IF Separation_Assessment = UNKNOWN AND
    Stop_Sep_Prop_Cycle = No
  THEN Separation_Property = Temp
    RESET Separation_Property
    RESET Separation_Assessment
    DISPLAY
    "If you are unsure of the separation property, Please select 'SEEK'
    rather than entering a question mark (i.e., UNKNOWN)."
    FIND Separation_Property;
```

Code 4.

### Code to Assist in the Selection of Chemical Properties

```

RULE Separation_Property14
  IF Separation_Assessment = Seek AND
    Differing_Optical_Activity = Yes
  THEN Separation_Property = Optical_Activity
    DISPLAY "
    Separation based on optical activity is sometimes possible,
    but is very difficult. An alternative might be to use Liquid
    Solid Chromatography."
```

Code 4.8

What is your assessment of likely separation mechanisms—Choose ONE or More of the following responses. If you are not sure, selecting SEEK will make the system help you determine an answer(s).

Does your sample differ in:

|             |                   |                   |
|-------------|-------------------|-------------------|
| Size        | Polarity          | Optical Activity  |
| Aromaticity | Diff Substituents | Geometric Isomers |
| Seek        |                   |                   |

(The user would select one or more of these choices using the cursor. The user may also enter a certainty for each individual choice. Assume in this case that the user chose only SEEK. The rest of the dialog would continue, eventually asking...)

.....

Is there a difference in the optical activity of your solute species (e.g., are you trying to separate optical isomers?)

Yes                      No

(The user would once again select either Yes or No or choose Unknown as the response. Further questions would continue.)

**Figure 4.2**  
**Example of Dialog that Determines the Separation Property**

The second set of rules used in determining a correct column and mode are the rules used to determine the use of reversed- versus normal-phase chromatography. The chemical and physical information contained in the choice of the phase is combined with the rules used to determine explicitly the separation property(ies).

Once the expert system has ascertained the properties that differentiate the solutes, the next stage is to provide more detailed advice. Twenty-one MODE rules in the knowledge base attempt to further classify the mode into different sub-fields that are listed in Table 4.4 (page 110). These modes are determined by the separation properties and by the presence or absence of particular compounds. A description of the sub-modes and how they are determined is given in Table 4.5. While not comprehensive, Table 4.5 does provide an overview of the sub-modes that are assigned.

Not mentioned in this table are the specialty modes that are determined (e.g., special columns optimized for chiral, protein, carbohydrate samples).

The first column in Table 4.5 provides the different sub-modes that will be determined. The second column gives different conditions that must be met before the sub-mode is assigned. This column is structured so that upper level entries indicate a concise response or determination from the user or expert system. An example is the Cyano entry. If the user indicates that the solutes have a net dipole moment, the expert system will assign CYANO to the variable PACKING. Another rule will assist the user if he/she replies with UNKNOWN to the question regarding the presence of dipole moments. In this case, the expert system will determine if either ketones, aldehydes, or nitro groups are present (by explicitly requesting for this information from the user or previous dialog). If these functional groups are present, the expert system will decide that dipole interactions are present. These secondary conditions are indicated within the second column in Boolean Logic format (e.g., AROMATIC WITH (DIFFERENT SUBSTITUENTS OR AMINES OR NITRO GROUPS OR PHENOLS)). Ionic modes are not mentioned in this table since the assignment of ionic modes is discussed in detail in the validation studies (Sections 4.7 and 4.8).

The third column indicates conditions under which will the choice of that packing or sub-mode is prohibited. These exclusions typically involve solutes which can irreversibly bind, or exhibit undesirable HPLC characteristics such as long retention times or peak tailing, with the packing material. Each mode has associated with it a variable which indicates if that packing material should be excluded (e.g., exclude amine columns if carboxylic acids are present). If a value has not been explicitly assigned to the PACKING variable and the packing material has not been determined to be inappropriate from other rules, the expert system will still suggest these packing materials, with a resulting decrease in the CNF. Several examples of these are indicated



in Table 4.5 with the entry, \*Don't Exclude. Table 4.5 lists the confidence factors that will be associated with the packing material if the required and excluded conditions are met.

## **Section 4.5 Rules**

This section describes a few special aspects of the rules used in the expert system. The major focus of this section is on self-determining rules that. Self-determining rules are defined as those which are able to predict answers based upon data obtained during the consultation with the expert system. This section also describes the importance of the order of rules within the knowledge base and the implementation and use of forward-chaining rules.

Flow charts and discussion of the logic used to recommend an appropriate column(s) are presented in Appendix A: *Flow Charts for the Determination of an Appropriate Column(s)*. The remainder of this chapter describes the recommendation of a column in more general terms. Please refer to Appendix A for more detailed explanation of the inference processes.

### ***Section 4.5.1 Order of Rules***

The random arrangement of the information and rules in the expert system will not usually affect the end result. There are three considerations that dictate why the knowledge base should be composed in a structured format.

**Table 4.5**  
**Determination of Sub-Modes**

| <b>Sub-Mode</b>     | <b>Required Conditions</b>                                                                      | <b>Excluded Conditions</b>               | <b>CNF</b> |
|---------------------|-------------------------------------------------------------------------------------------------|------------------------------------------|------------|
| <b>Alkyl-Silane</b> | Neutrals<br>Homologous Series<br>Ion-Paired Species                                             |                                          |            |
| <b>Cyano</b>        | Dipoles<br>Ketones or<br>Aldehydes or<br>Nitro groups                                           |                                          | 100        |
| <b>Phenyl</b>       | Charge Transfer<br>Aromatic with<br>(Substituents or<br>Amine or<br>Nitro groups or<br>Phenols) |                                          |            |
| <b>Amine</b>        | Proton Donor<br>Thiols or<br>Phenols<br>Aliphatic Alcohols                                      | Ketones<br>Aldehydes<br>Carboxylic acids | 100        |
| <b>Amine</b>        | Unknown or<br>Not a Proton Donor and<br>Aromatic or<br>Protein or<br>Nitro Groups               | Ketones<br>Aldehydes<br>Carboxylic acids | 70         |
| <b>Alumina</b>      | Aromatic<br>Net Dipole Moments                                                                  | Carboxylic acids<br>Phenols              | 100        |
| <b>Alumina</b>      | *Don't Exclude                                                                                  | Carboxylic acids<br>Phenols              | 50         |
| <b>silica</b>       | Geometric or<br>Steric Isomers                                                                  | Strongly ionic                           | 100        |
| <b>Silica</b>       | *Don't Exclude                                                                                  |                                          | 90         |
| <b>Ionic</b>        | Ionic Groups                                                                                    | pK <sub>a</sub> Values                   |            |

The first consideration is the ability to edit and maintain the knowledge base. When the knowledge base is organized such that the rules associated with a particular goal or aspect of the goal are arranged together, the knowledge engineer will be able to see the interrelationship between the rules. It will also be easier to edit these rules when new information becomes available. Ideally, the knowledge base should be completely modular by having each goal be self-contained within its own knowledge base. However, this approach requires that the inference engine perform additional work since the inference engine must save the state of all variables, load the new knowledge base, re-load the values of the variables, and continue processing. While this time would be negligible on modern computers, the computers used in this research are slower machines. Therefore, the modularity is obtained by grouping similar rules together and providing the rules with consistent names (e.g., Sep\_Property1, Sep\_Property2).

The second consideration in arranging the knowledge base is to minimize the time that the inference engine dedicates to searching the knowledge base. Since the inference engine will sequentially read the knowledge base to find rules that have hypotheses matching the goal (for backward chaining mechanisms), the inference engine will find the rules more rapidly when the rules are arranged in contiguous fashion.

The third consideration in the arrangement of the rules is to provide a clear and logical dialog session with the user. If the knowledge base is randomly arranged, the user will often be asked questions that appear to be non-sequiturs due to the sequential search procedure. An early version of this expert system asked the user the following series of questions:

|                                                                                                                                                                                                                       |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>Is the solute concentrated enough for your detector?<br/>Can you dilute the sample to match the concentration range of your detector?<br/>Is solute concentration so high that you will saturate the detector?</p> |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

These questions are asked to determine whether the solute concentration is within the linear operating range of the chosen detector. The above series of questions do not follow a logical thought process. The first rule determined if the solute concentration was greater than the limit of detection for the detector. The second rule tried to determine if the sample could be diluted if it was too concentrated. The third rule determined if the sample was too concentrated. While the expert system still obtained the correct recommendation, the user may be perplexed when asked the second question prior to being asked the third question. By changing the order of the second and third rules the inference engine first determined if the sample was too concentrated before asking if dilution is possible. The dialog session between user and expert system is now more logical, and the user will realize the correct context behind a particular question.

The rules in this knowledge base have been arranged in a logical order as much as possible, but the arrangement cannot be completely enforced in this expert system since much of the information is interrelated. That is, the data for one goal can be the solution to another goal and the inference engine can be forced, at times, to ask questions that might seem illogical. The solution to this problem is to write the questions such that they convey the reasoning behind the question.

#### ***Section 4.5.2 Forward Chaining***

Chapters 2 and 3 of this dissertation described the two types of inference mechanisms that are available. Most of the rules described in this chapter have involved backward-chaining rules. This section describes the use of forward chaining rules.

The principal use of forward chaining is in the determination of the separation mode. There are two different ways that forward chaining can be initiated. An implicit forward-chaining mechanism is the FIND <VARIABLE> clause. When the hypothesis is true for a rule containing a FIND clause in the conclusion, the inference engine is forced to determine a value for <VARIABLE>. An example of this mechanism can be seen in Code 4.9. If the user requests advice for a method, the rule will force the expert system to determine the separation property, class, mode, and packing type. This meta-rule re-directs the expert system through a different logic path.

VP-Expert implements explicit forward-chaining by the WHENEVER command. An example of this rule can be seen in Code 4.5 on page 111. This rule will execute automatically whenever the conditions within the premise become true. In this case, whenever a molecular weight range between 10 and 15% has been assigned, the rule will determine that a wide molecular weight distribution exists. This conclusion will then be used in determining if size exclusion chromatography (SEC) is a viable technique. The "WHENEVER" type of rule provides what is known in the AI vernacular as a "demon;" that is, the inference engine will monitor the status of variables within the premise and will pre-empt the execution of other rules whenever the variables become true. The demon will then change the logic path that the inference engine will follow.

Forward chaining is used in this expert system to facilitate the determination of size as a separating factor. SEC is a powerful technique, particularly when there is a large distribution of sizes and there are no, or few, competing separation mechanisms occurring. Without forward-chaining rules, the expert system would always have to ask the user to supply values for the molecular weight distribution. SEC can be postulated based upon data supplied directly by the user (i.e., the user has indicated that size is an appropriate separation mechanism) or can be inferred from other information. For example, the expert system asks the user if the solutes comprise a homologous series.

If the user indicates that the solutes do form a homologous series that increase in molecular weight or size, the expert system will infer that SEC might be a useful technique even if the user did not originally choose size. Without the forward-chaining rules, the expert system does not have a mechanism to determine the distribution of the sizes.

### Forward Chaining Code Via Find Statement

**RULE Continue2**

```
IF   HPLC_Suited = Yes AND
     Check_For_Methods = Yes
THEN Continue = Yes
     FIND Separation_Property
     FIND Rev_Phase_Ok
     FIND Normal_Phase_OK
     FIND Mode
     FIND Packing_Type;
```

Code 4.9

### **Section 4.5.3      *Self-Determining Rules***

The expert system should be able to infer as much of the solution as possible as information became available. By having the rules be "self-determining," the expert system appears to be intelligent. The user does not have to re-answer questions that were determinable from previous data entries. Without this ability, the user may become irritated when he/she has to enter answers more than once. With self-determining rules, the number of questions asked of the user is decreased. The speed of the expert system is increased due to a decreased number of questions asked and the decrease in the response time of the user to read, think, and provide a response to the questions.

There are several places where self-determining rules are implemented. The first case is where the expert system determines the significance of size in choosing a separation mechanism. The expert system can ask relevant questions to determine the

size distribution and calculate a confidence factor for the certainty of success of using SEC.

Many of these self-determining rules occur during the determination of functional groups that might be present. The expert system is able to infer if there are dipole, charge transfer, or hydrogen bonding interactions present. Throughout the rest of the chapter, self-determining rules will be described in addition to the examples cited above.

## **Section 4.6      User Interface**

The user interface for the expert system based upon VP-Expert is provides the user with a menu of legal responses to questions. These choices are selected with the cursor keys in order to minimize typing and insure accurate entry of data. Display text is written to provide contextual information that would answer both HOW and WHY questions that the user may require.

Data base access is accomplished by a similar menu system. The expert system accesses the data base with the proper syntax and reports the responses back to the user who may select from the entries. The expert system also uses a data base to provide advice that is specialized for that particular laboratory. The expert system has the capability to automate entry of data into this data base, but this feature is not implemented. As such, the user must modify directly the data base to add information such as available columns and their properties (e.g., dimensions, maximum pressure).

## **Section 4.7      Preliminary Validation of the M1 Prototype Expert System**

Validation studies provide a method to determine if the expert system provides the correct conclusions.<sup>12</sup> The validation studies demonstrate that:

- (1) particular sections of the logic are perform as expected (e.g., confidence factors are calculated correctly).
- (2) functions of the expert system operate as expected (e.g., menu choices are correct and are visually appealing).
- (3) representative problems are treated correctly.
- (4) difficult problems can be solved.
- (5) that the expert system is robust and can provide advice for "trick" problems.

The first two objectives were monitored repetitively during the development of the expert system in this dissertation. As new information and modules were added, they were tested to ascertain that the expert system was functioning correctly. The results from these interim tests are not presented in this dissertation. The last three objectives are demonstrated in the following validation studies.

There are several different ways to develop test cases. These include:<sup>12</sup>

- (1) **Historical Data:** This test case allows the knowledge engineer to select cases which exhibit specific properties. In addition, the knowledge engineer can evaluate the solution(s) proposed by the expert system to the accepted, or expected, method.
- (2) **Collected Data:** This approach uses current research as the test case to allow optimization of the expert system for the expected applications for which the expert system will be used.
- (3) **Generated Data:** Artificial test cases are created to test particular aspects of the expert system. These cases may contain conflicting or missing data to test properties such as the robustness and completeness of the knowledge base.

The validation studies for this expert system use mixtures of all three types of test cases. The major test cases use historical data obtained from literature. In addition, the authors of the literature used for these test cases typically provided a reasoning behind their choice of method. One validation study based on generated data is presented to demonstrate the ability to handle a case that is difficult and contains conflicting data. The collected data approach involves validation studies which were then run on the HPLC instrument, as described in the Chapter 6 of this dissertation.



The first validation studies were performed using the prototype expert system that was based on the M1 shell. These validation studies demonstrate the functionality and limitations of the prototype. The test cases are described and the data used as input for the expert system is discussed. Results from the consultation are presented for each individual study. Discussion of these results is presented in Section 4.8 together with the results for the validation studies using the VP-Expert version of the expert system.

#### ***Section 4.7.1      Estrogens Test Case—M1 Prototype***

The first test case described here involves the separation of three water-soluble estrogens, equilenin sodium sulfate, equilin sodium sulfate, and estrone sodium sulfate. The structures of these compounds are given in Figure 4.3. This validation study was used because the development of the accepted method was detailed and provided the data that could be used as input for the expert system. The accepted methods<sup>13,14</sup> acknowledged of the ionic nature of the compounds; however, the pH of the three estrogens is easily adjusted so that the three estrogen conjugates are neutral. There are significant differences in the aromaticity and polarity of each (each component has decreasing amounts of conjugation). The differences in the charge transfer characteristics and polarities of the three estrogens may be used as appropriate separation mechanisms and a phenyl column would separate the three estrogens. In this method development, a phenyl column was not readily available, and a method based upon reversed-phase chromatography was chosen. A methanol/water mobile phase was suggested and phosphoric acid was added to maintain neutrality of the estrogen conjugates.

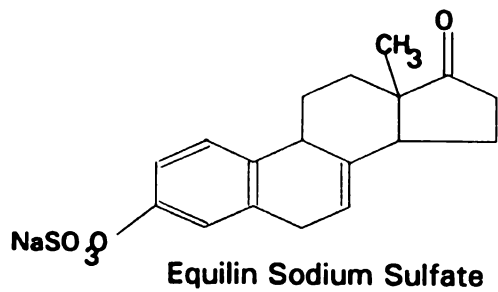
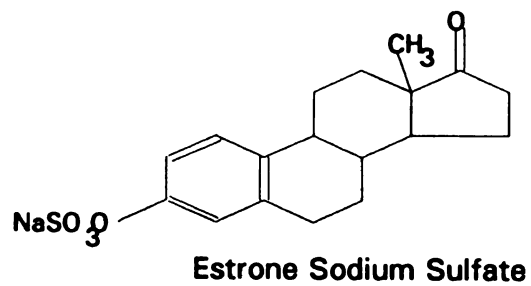
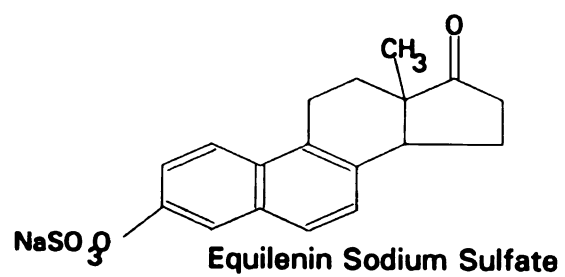
The input to the expert system for this validation study is as follows:

- water soluble

- aromatic molecules
- a molecular weight less than 2000 g/mole
- ketone functional groups present
- large organic ionic species (more than 5 carbons) with small inorganic ions also present
- the  $pK_a$ s of the species indicate that they could be neutralized at a pH within the typical operating range of HPLC

The abbreviated dialog for this consultation can be seen in Table 4.6. The response from the user is prefaced with a >> prompt. The expert system recommended:

- reversed-phase ion pair chromatography (IPC)
- reversed-phase with an alkyl silane packing material
- reversed-phase with a cyano packing material



**Figure 4.3**  
**Structures of the Conjugate Estrogens**  
**Used in the Validation Study**

## Table 4.6

### Dialog for the Estrogen Consultation

Is the molecular weight of the solute large (i.e. >2000) or small (i.e. <=2000)?

>> small

Is the solute soluble in organic solvent(s) ?

>> U

Error 222. Responses that begin with a capital letter are not allowed. Try retyping the entry in lower case or enclose it within 'quotes'. Your response must be chosen from the following:

yes, no.

—

Now consider the ionic character of your solute. In the normal pH range for HPLC (pH 2 - 8) is your solute

l - ionic

n - nonionic or

l,n - ionic, but could be made nonionic by appropriate  
pH or other change ?

(Note: If you have the choice, and your solute molecules can be made non-ionic, then you should make them so - separation of ions is a much more difficult problem)

>> l,n

Assessment of likely separation mechanisms. Enter one or more of the following responses. If you do not know the answer, or do not understand the terms, type "u." for "unknown. The components you wish to separate

s - differ in size only

p - differ in polarity only

b - differ in both polarity and size;

o - differ in optical activity

a - have aromatic character

d - are aromatic and have different substituents

g - contain geometric isomers

>> u

Does the solute contain large organic ions ?

>> y

What pH would be required to make solute species neutral ?

>> acid

Error 90. A non-numeric term is found where a number should be. Check for arithmetic that mixes numbers with atoms.

Is the solute soluble in water mixed with methanol, tetrahydrofuran or acetonitrile ?

>> y

Does the solute contain inorganic ions ?

>> y

Does the solute contain very small organic ions (i.e. less than 3 - 4 carbon atoms) ?

>> n

Does the solute contain

p - peptides,

s - small\_proteins,

o - oligonucleotides or

n - none of these ?

## Table 4.6 (continued) Dialog for the Estrogen Consultation

>> n

Does your solute contain:

c - carboxylic acids,

p - phenols,

k - ketones,

n - nitro\_groups,

t - thiols or

a - aliphatic\_alcohols ?

>> k

Do you have, or could you obtain, a column with packing: alkyl\_silane ?

>> y

Does your solute have the capability for dipole interaction with the column packing ? (this depends on what polar groups are present - If it has ketone, aldehyde, or nitro groups present, or is aromatic then it probably does).

>> y

Do you have, or could you obtain, a column with packing: cyano ?

>> y

Does your solute have the capability of charge transfer with a charge-donating or charge-accepting stationary phase (e.g. aromatics with several electron donating substituents would have this capability).

>> u

Recommended separation mode: reverse\_phase\_paired\_ion

If the required pH for solute neutrality <pH2 or >pH8 then you MUST use ion pairing reagents e.g. TMA, TBA, sulfonate. Ion pairing MAY also be used within the pH 2-8 range.

**Suggested packing: alkyl\_silane**

This packing separates primarily on the basis of size, but less polar (overall polarity) solutes will be retained longer.

Recommended separation mode: reverse\_phase\_paired\_ion

**Suggested packing: cyano**

This packing separates on the basis of dipole interaction with the cyano nitrogen lone pair of electrons. As such it is better at dipole interaction than phenyl packing. Its efficacy depends on the polar group present on the solute. It is suitable for separating (e.g.) ketones, aldehydes, aromatics and nitro compounds.

Recommended separation mode: reverse\_phase\_bonded

**Suggested packing: alkyl\_silane**

Recommended separation mode: reverse\_phase\_bonded

**Suggested packing: cyano**

### **Section 4.7.2      Validation of B Vitamins—M1 Prototype**

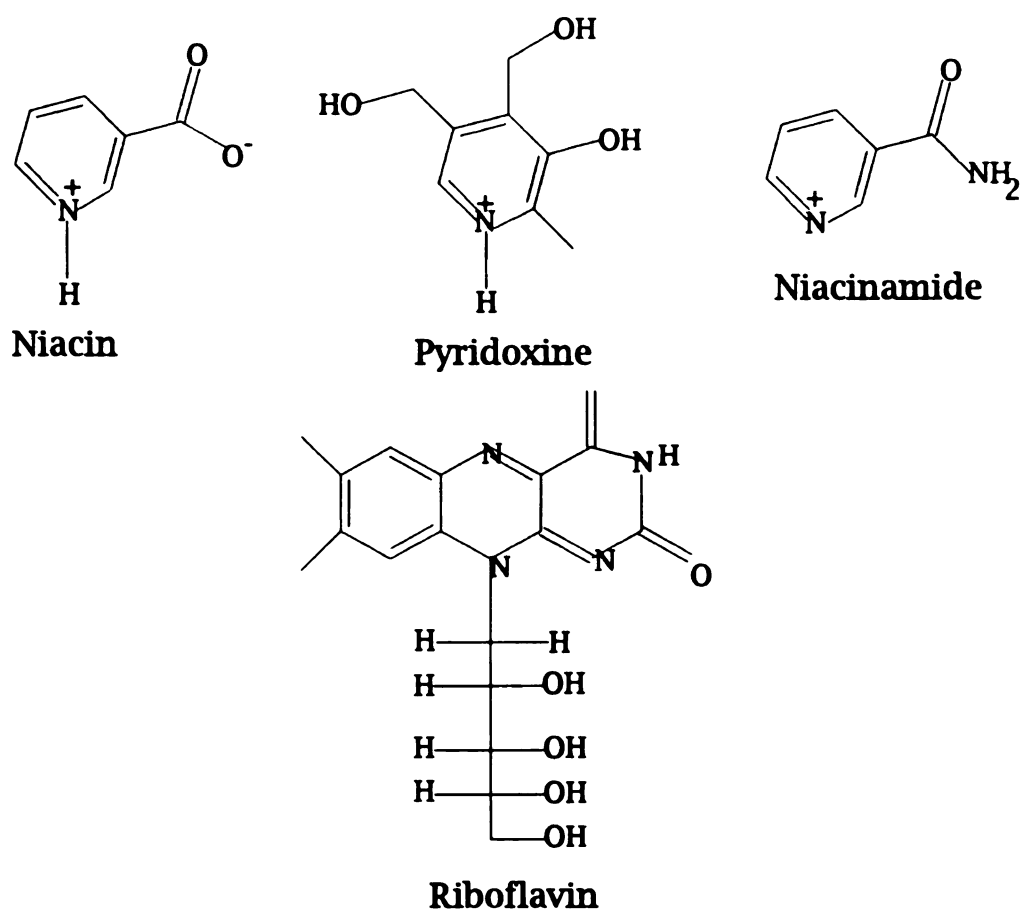
The next validation study was a sample containing four water soluble B vitamins, niacin, niacinamide, riboflavin, and pyridoxine HCl (see Figure 4.4).

Accepted methods<sup>15</sup> emphasize the ionic nature of the vitamins. As such, the accepted methods involve ion exchange (IEC) or ion-paired (IPC) chromatography. Reverse-phase chromatography using ionization control was also recommended as an appropriate method. An Aminex A-25 amine column (typically used for anion exchange chromatography) at a pH of 10 was also recommended.<sup>15</sup> At a pH of 10, the vitamins are non-ionic and the probable separation mechanism is via partition chromatography between the amine packing material and the polar functional groups of the vitamins. One objective of this validation study was to test additional rules which had been included to differentiate between the various ion chromatography techniques.

The data for this validation study is as follows:

- soluble in water
- molecular weights less than 2000 g/mole
- contain ketone groups
- large organic ionic species (more than 3-4 carbon atoms) with inorganic ions present
- the species can be neutralized at a pH within the typical operating range of HPLC

The expert system proposed initially only reversed-phase ion pair chromatography. Subsequent validations with the same test case provided ion-exchange chromatography as a recommendation only when the expert system was informed that the pH required to shift the solutes into their neutral form was excessively high.



**Figure 4.4**  
**Structure of Four B Vitamins:**  
**Niacin, Niacinamide, Riboflavin, and Pyridoxine HCl**

### **Section 4.7.3      *Validation of a Hypothetical Case—M1 Prototype***

The final validation study presented here is a hypothetical case that included a very complex mixture with various properties. This test case contains what might appear to be conflicting characteristics. However, this is an attempt to ascertain how well the expert system would be able to treat more difficult cases. The characteristics of this sample are:

- soluble in both aqueous and organic solvents
- molecular weights greater than 2000 g/mole
- the sample is a homologous series in which the species increase in size and polarity.
- strongly ionic species with multiple ionic sites present.
- A pH of 7 is required to shift the species into the neutral forms.

The recommendations from the M1 prototype expert system provided only two suggestions: normal-phase SEC and reverse-phase alkyl-silane.

## **Section 4.8      Secondary Validation of VP-Expert Expert System**

This section discusses the results of the preliminary validation of the expert system. These discussions are described in context of the secondary validation of the expert system that is based on the VP-Expert shell. These secondary validation studies are performed using the same test cases described above.

The prototype expert system attempts to minimize or remove any recommendation that involves any type of ion chromatography. This approach provides inadequate recommendations for samples in which the ionic natures of the eluents can provide an appropriate separation mechanism.



The M1 version does not provide any additional rules to help determine the presence or absence charge transfer characteristics and dipolar interactions whenever the user responds with an unknown reply.

The M1 version also does not provide a ranking of the proposed solutions to allow the user to determine which of the suggestions is the most likely to succeed. The recommendations are returned only in the order that they are assigned in the knowledge base.

Finally, several problems with the user interface were discovered. The M1 version did not provide the user with the acceptable responses for a question unless an invalid response is entered. At that point the expert system responded with an error message and supplied the correct choices. The expert system also required that the responses be entered with correct spelling and with all lower-case letters. In addition the expert system allowed a response from the user of ACID when asked for the numeric value of a pH that would result in a neutralization of the ionic species.

The objectives arising from this validation study were to improve the second generation of the expert system such that the expert system provides:

- (1) a more user-friendly environment
- (2) error checking on data provided by the user
- (3) ability to predict interactions when the user responds UNKNOWN
- (4) a ranking of the recommendations with some numerical value to connote the certainty of the recommendations
- (5) an increase in the logic pathways from which the expert system may predict possible separation mechanisms

The VP-Expert shell was chosen to replace M1 as the inference engine because VP-Expert is more intrinsically user-friendly. Every question in the VP-Expert knowledge base has an associated CHOICES command. The CHOICES command presents the allowed values for the question, and the user need only use the cursor keys to select the

answer(s). Variables that have a numerical value may be controlled by a VP-Expert command, RANGE, which allows only numerical entries between specified minimum and maximum values. The result of these additions to the knowledge base is an increase in valid entries and a decrease in the complexity of using the expert system.

The knowledge base was also expanded to allow the inference engine to handle unknown values and to prompt the user with further questions. The knowledge base contains additional rules which predict the values for variables based upon supporting evidence (i.e., the self-determining rules mentioned earlier). For example, rules were added which predict a dipolar interaction based upon the presence of ketone groups and to predict charge transfer interactions due to the presence of aromatic rings with substituents. One example of this approach is seen in Code 4.10 that provides an recommendation based upon previously entered data. An additional objective was to make the expert system easier to use by providing additional help by incorporating detailed text via the DISPLAY command and also by the BECAUSE command (which allows the user to ask the expert system WHY and HOW).

The addition of a ranking mechanism to provide an assurance factor to the user was accomplished by the inclusion of confidence factors. This addition to the expert system is discussed in Section 4.9.

Finally, the increase in allowed separation mechanisms is described in Section 4.9 and in the following validation studies.

#### ***Section 4.8.1      Estrogens Test Case—VP-Expert***

The estrogen test was chosen to test the ability of the expert system to treat a reversed-phase separation. Solute characteristics for the VP-Expert system were similar to the M1 version with several differences. The molecules were described as organic ions containing ketone groups.

## Code to Determe Dipole Interactions

### Rule Packing7A

```

If      Charge_Transfer = Unknown AND
        Aromatic_Solute = Yes AND
        Separation_Property = Differring_Substituents AND
        Functionalities = Amine_Groups OR
        Functionalities = Nitro_Groups OR
        Functionalities = Alcohols OR
        Functionalities = Phenols
Then Charge_Transfer = Yes
    Display "

```

Code 4.10

You have an aromatic with charge transfer groups present on the ring. The functional groups have a lone pair of electrons which are donating to the aromatic ring. This creates a compound which will have charge transfer characteristics."

BECAUSE "

We are trying to determine if the solute has charge transfer characteristics. If the user does not know the answer, then we will base the answer on whether there is an aromatic ring with electron donating substituents present.";

The molecules were also described as containing a distinctive size difference between the species. This descriptions was an intentional mistake to determine if error correction rules added to the knowledge base would query the user sufficiently to determine if there was a size difference in the solutes. The goal was to add rules that employ forward-chaining inference mechanisms that would automatically be accessed when the user suggests that size is a discriminating factor. These rules would then ascertain the molecular weight range of the species and override the user's input if the difference was not sufficiently significant to insure a separation based on size alone.

A UNKNOWN response was given when the expert system asked if the solutes have the capability of dipolar and charge transfer interactions with the stationary phase. This response tested the addition of the rules similar to Code 4.10. The output for this test case is seen in Table 4.7.

The inclusion of confidence factors delineates the recommended separation modes into different levels of certainty. IPC was ranked at a lower confidence factor (70) since the species were could be neutralized with the appropriate pH. Therefore, the

expert system inferred that a separation could be based upon the ionic nature of the solutes, but one of the other techniques would be better suited.

**Table 4.7**  
**Results of the Validation Study for the**  
**Estrogen Components Using the VP-Expert Version**

| Mode                                   | Column           | Confidence Factor |
|----------------------------------------|------------------|-------------------|
| Reverse-phase with Ionization Control  | RP-18 Spheri-5   | 100               |
| Reversed-phase Phenyl                  | Bonda Pak        | 100               |
| Reversed-phase Cyano                   | Bonda Pak        | 100               |
| Reversed-phase ion pair Chromatography | none in database | 70                |

The expert system also decided against using SEC. This decision is based on the size distribution, average molecular weight, and the presence or absence of alternative separation methods. These rules calculate a confidence factor to provide a numerical descriptor to the ability to separate based on size, as discussed in Section 4.4.4.1 on page 109. The average mass (defined as small if less than 2000 g/mole) is also calculated during this process. SEC is most efficient for large molecules since small molecules require the use of gel permeation stationary phases. The second set of rules determines if there are alternative separation properties that are present. If the solutes have additional properties (e.g., polarity, ionic), these rules will decrease the certainty of using SEC from 100% CNF (size is the sole discriminating factor) to 50% (every other mode is present). An example of this rule is provided in Code 4.11. If the sample is labile, these rules increase the confidence factor since SEC is a gentle technique. These

rules are discussed in further detail in Section 4.9, page 148. The expert system applied all of these rules and determined that:

- (1) The difference between large and small molecules was too small to provide a possible separation mechanism.
- (2) The average molecular weight was small.
- (3) Dipole, charge transfer, and ionic characteristics are present.
- (4) Combining the above three conclusions, size was discounted as a factor.

Unknown values for dipole and charge transfer were handled by the extra rules to assign the proper value for these variables.

The expert system provided the accepted responses of Reverse-Phase alkyl-silane with ionization control, and a reversed-phase phenyl column. In addition, the expert system suggested a cyano column. IPC was recommended as a possible alternative.

#### Code that Determines the Priority for Size Discrimination

|                                                                                                                                                                                                                                                              |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>RULE Size_Consideration4</b><br><b>IF</b> Separation_Property = Size AND<br>Labile = No AND<br>Separation_Property <> Polarity AND<br>Separation_Property <> Aromaticity AND<br>Ionicity = NonIonic<br><b>THEN</b> Size_Consideration = Size_Only CNF 85; |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

Code 4.11

The expert system also suggests particular brand names of columns to use. The expert system incorporates a data base that maintains a listing of representative columns. This database contains, among its different fields, information on whether this column is available in the laboratory of the specific user, column geometries, and so forth. This database allows the expert system to make recommendations with consideration of the particular abilities of the laboratory. The expert system will suggest only available columns, if so directed, and will also suggest the appropriate size of column (e.g., preparatory versus analytical scale) depending upon the particular

application as determined in an earlier stage which assists the user in deciding if HPLC is a suitable technique.

Not provided in Table 4.7 is the explanatory text that is given at the end of the consultation. After a method has been proposed, the expert system will access a hypertext file and display information that explains why that method was chosen. A short portion of the file is provided below in Code 4.12. The databases and hypertext files are described in more detail in Section 4.10 of this chapter.

### **Hypertext Entry to Provide a Description for the Selected Method**

.pa  
Rev\_Cyano

Code 4.12

---

**Column to use: Reverse Phase Cyano Column**

**This column will separate based on partitioning between the mobile phase and the cyano stationary phase. The cyano group will interact in several ways with solutes. First, it is a moderately polar group (less than the amine) so it will interact with aldehydes, ketones, and halides. It also has a dipole which will then interact with compounds containing dipoles. It is selective for aromatics that have electron donating substituents.**

### **Section 4.8.2      *B Vitamins Test Case—VP-Expert***

The estrogen validation study demonstrated the importance of having the expert system consider all of the possible modes of separation. The prototype minimized ion chromatography techniques. One objective in the vitamins validation study was to test the inclusion of additional rules to delineate between the different ion chromatography techniques. The accepted method for these B vitamins included cation exchange and ion paired chromatography. Figure 4.4 on Page 130 gives the structures of the four vitamins. One can see the different ionic sites which are present on the vitamins.

There are several different modes that can be used for the chromatographic separation of ionic species. Ion-exchange chromatography (IEC) is typically used to separate strong organic acids and bases. The retention mechanism involves ion exchange between the sample ions and the mobile counter-ion of the packing material. Conductivity detectors are typically used to monitor the IEC effluents. IEC packing materials are not as efficient, less reproducible, and less stable, particularly for anion exchange, as other techniques.<sup>11</sup> In addition, the large background signal resulting from the mobile phase electrolyte can decrease the signal-to-noise ratio of the conductivity detector.

Ion Chromatography (IC) is a specialized form of IEC that usually employs conductimetric detection. The technique uses a combination of IEC resins to remove the large background signal from the electrolyte to allow the use of the conductimetric detector. The sample passes through a separation column where the ions are resolved based primarily on ionic density. The separated ions then pass through a stripper column where the eluant is neutralized and the sample elutes in a background of deionized water. This procedure allows IC to be used for trace analysis.

Ion pair chromatography (IPC) is another chromatographic technique to separate ionic species. In IPC a counter-ion is added to the mobile phase to form an ion pair with the analyte(s). IPC can be performed in both normal-phase and reversed-phase; however, the reversed-phase mode will be discussed here. The stationary phase typically consists of an alkyl-silane packing material such as a C<sub>8</sub> or C<sub>18</sub> bonded phase. For normal-phase operation, a porous silica support containing a mechanically held aqueous solution of buffer and counter-ion is used for the stationary phase.<sup>11</sup> In reversed-phase operation, the ion pair will be retained on the non-polar column while the free ion forms partition into the mobile phase. Therefore, the partition coefficient is a function of equilibrium constant for the ion pair formation. IPC is particularly

suitable for polar samples such as bases, species with multiple, ionizable groups, and for species in which the pH to neutralize the ionic species may cause degradation of the HPLC column packing material. Specialty columns exist for IPC which contain a water-immiscible ion exchanger which serves as the counter ion.

The prototype expert system did not differentiate between these different methods to chromatographically separate ionic species. As demonstrated in the M1 validation study, ion separation techniques were only recommended when the required pH to neutralize the acid forms was excessively acidic or basic.

Figures 4.5, 4.6a, and 4.6b provide a logic flow chart for the M1 and VP-Expert expert systems, respectively. The number of different paths that the inference engine can access to reach a mode to separate ions has been increased in the VP-Expert version (HPLC Advisor). Because backward-chaining is used to determine the mode and column, HPLC Advisor can attempt every induction path and can recall those that were successful. The numbers following the different modes indicate the CNF that are assigned to that mode via that particular path. The modes without a number have a CNF of 100%. The values for these CNF are described later in Section 4.9

The characteristics of the B vitamins used as the input for HPLC ADVISOR were similar to the M1 validation case with the addition that the solutes were strongly ionic and contained alcohol and amine functional groups. HPLC ADVISOR recommended:

- ion-pair chromatography CNF 100
- cation exchange CNF 100
- reversed-phase amine CNF 80

The recommendation to use an amine column was surprising since this recommendation was not one of the original accepted methods. An additional literature search found that other researchers have used an amine column for the separation of B vitamins. The recommendation of an amine column is based upon the polar nature of



the B vitamins and dipolar interactions of the functional groups present with the amine stationary phase.

### ***Section 4.8.3 Hypothetical Test Case—VP-Expert***

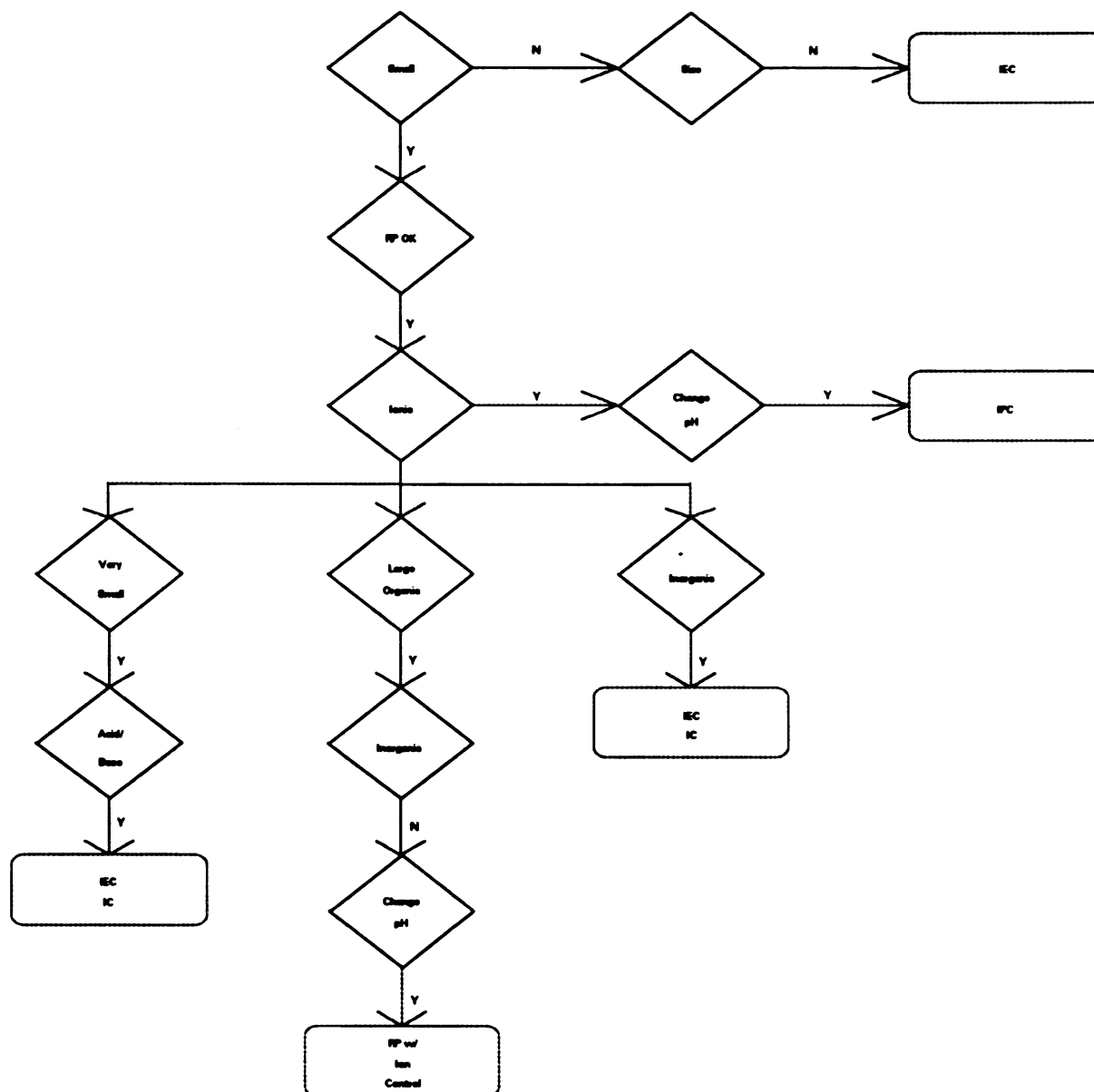
The objective of the hypothetical test case is to provide a very complex case and see if the expert system is able to provide all possible solutions with appropriate CNF.

As a reminder, the case involved:

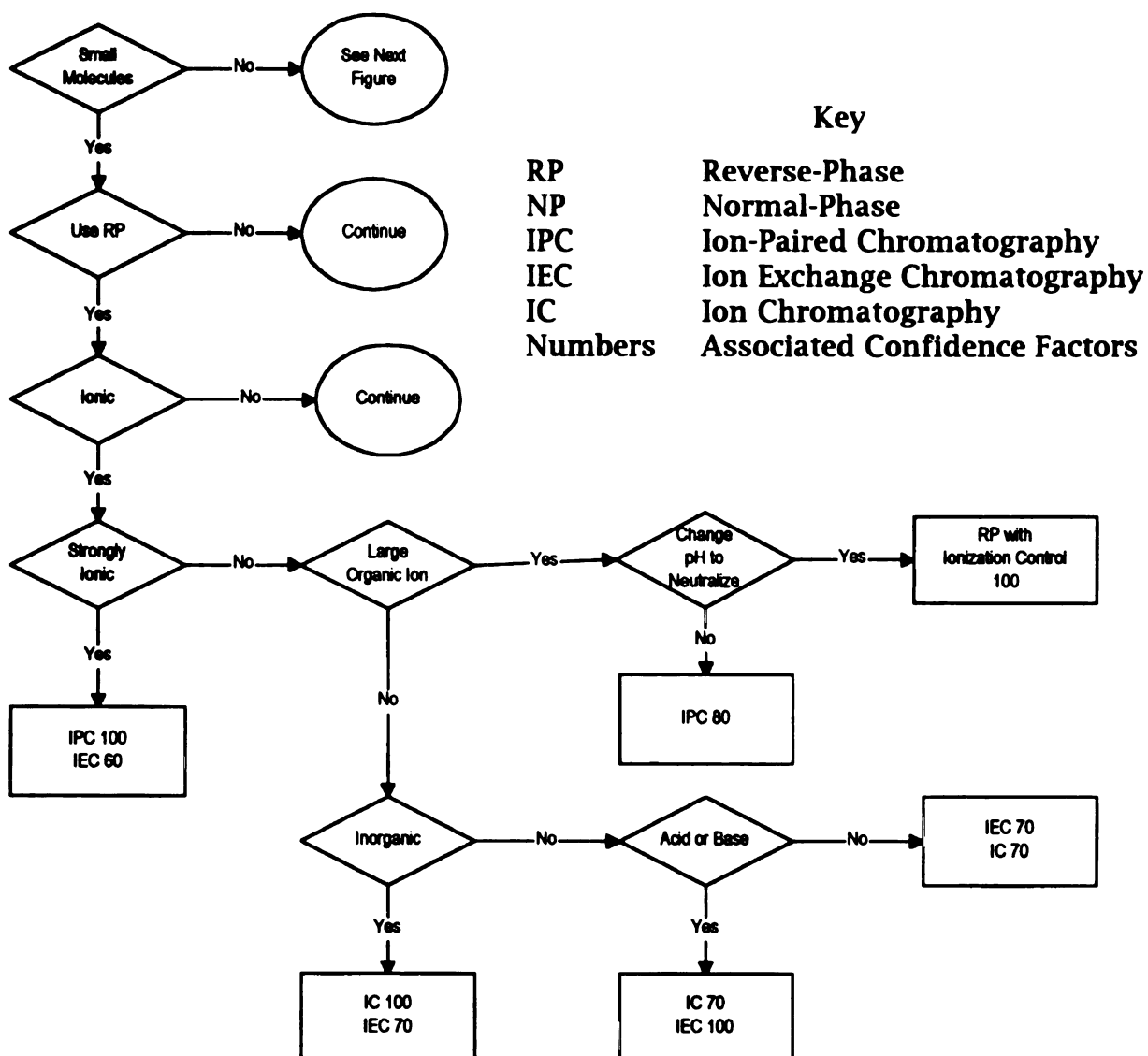
- (1) Aqueous and Organic Soluble components
- (2) Large molecules (>5000 Daltons)
- (3) Homologous Series in which both size and polarity increase for each member of the series. Difference of 12% between low and high molecular weight species.
- (4) Strongly Ionic with multiple ionic sites. A pH of 7 will change the species into the non-ionic forms

Points 3 and 4 indicate conditions that the M1 version does not have rules with which to assist.

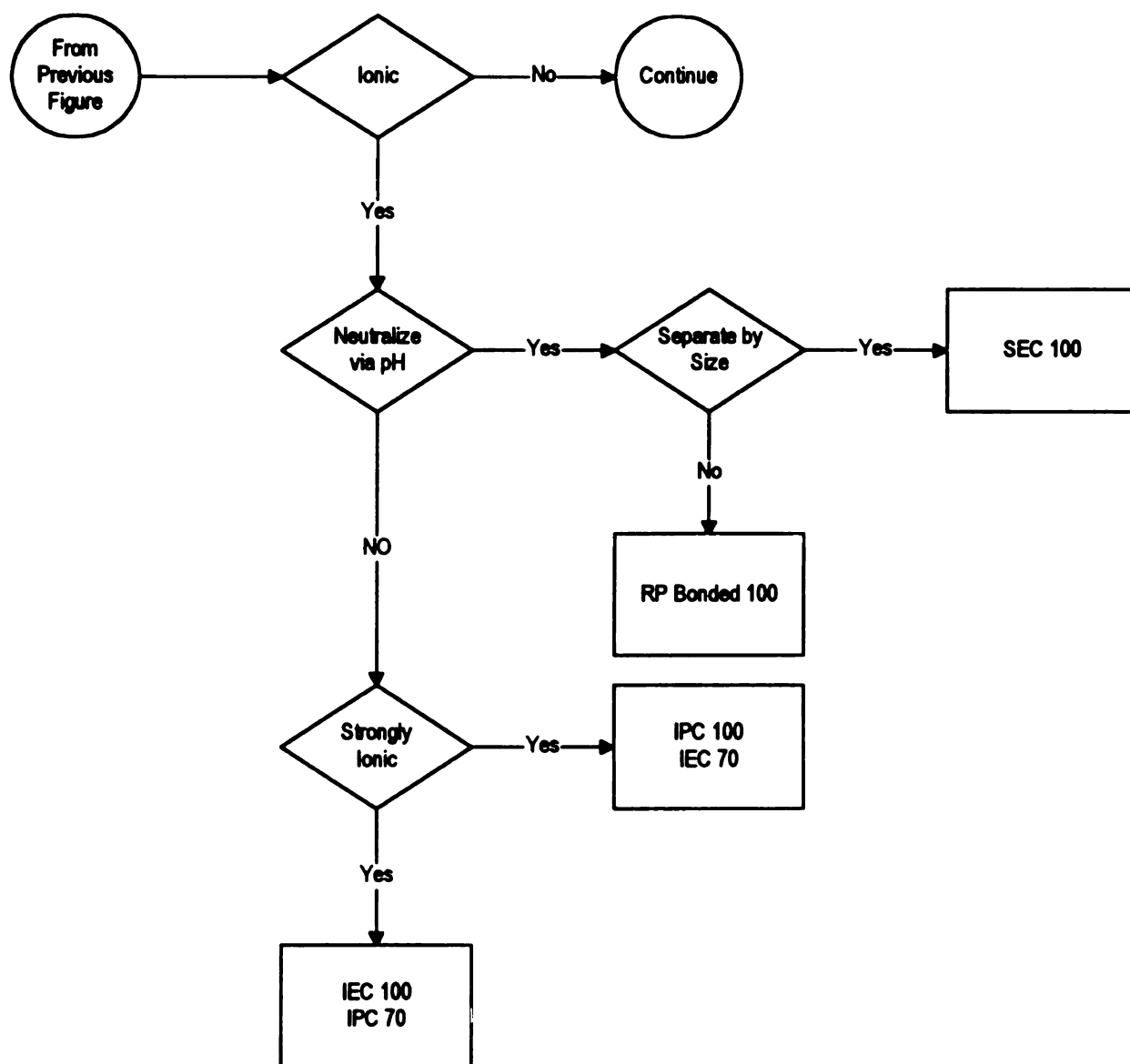
This test case determines the behavior of the expert system on several points. The first is to see if a proper response for size will be determined. As mentioned earlier, HPLC ADVISOR will determine that SEC is a viable solution only if the species have a large weight distribution and there are few or no competing techniques. In this test case, the distribution is fairly large, but there are polarity and ionic characteristics that compete. SEC is a complex and expensive technique (columns are expensive and fragile). Typical SEC columns also have low capacity factors which indicates that they will not be able to resolve a large number of components (recall that the 10% difference criteria applies not only to the highest and lowest molecular weights, but also to adjacent molecules in the size distribution profile). Therefore, if other techniques will do the job, the expert system should weight those responses higher.



**Figure 4.5**  
**Flow Chart for the Determination of Ion**  
**Chromatography Techniques:**  
**M1 Version**



**Figure 4.6a**  
**Flow Chart for the Determination of Ion Chromatography Techniques**  
**for Small Molecules:**  
**VP-Expert Version**



### Key

|         |                               |
|---------|-------------------------------|
| RP      | Reverse-Phase                 |
| IPC     | Ion-Paired Chromatography     |
| IEC     | Ion Exchange Chromatography   |
| SEC     | Size Exclusion Chromatography |
| Numbers | Associated Confidence Factors |

**Figure 4.6b**  
**Flow Chart for the Determination of Ion Chromatography Techniques**  
**for Large Molecules: VP-Expert Version**

The test case also tests the rules which deal with ion chromatography. The species are large and have strong multiple ionic sites. Typically a very large pH extreme (basic or acidic) would be required to neutralize each ionic site; therefore, necessitating the use of IPC or other ionic technique. This case still tests the treatment of a very ionic molecule which can easily be neutralized. While this seems illogical, recall that these types of cases are meant to test the knowledge base for trick samples. As such, the expected answer should be Reverse phase with ion control, perhaps followed by IPC. IEC will not likely be a recommendation since IEC is typically used for smaller molecules which can easily exchange with the counter ion of the resin. If the species is are very ionic, these components are retained for very long periods. Also large molecules often are unable to penetrate the resin material to exchange with the counter ions, and the sample may not adequately resolved.

The output for this test case is given in Table 4.8. There is a difference of 10 between the CNF of Normal Phase and Reverse phase bonded. These CNF are meant to be only an indicator of the certainty of the recommendation. With this in mind, the CNF of HPLC ADVISOR were set up in increments to represent a certainty of success (please see Section 4.9). In this implementation, a range of 100 to 90 indicates the best choice(s). The test case did not provide enough information to allow the 10% difference to be significant. Therefore, this difference is only the result of the propagation of different CNF throughout the knowledge base. Reverse phase was also slightly lower since there are competing techniques that might serve as well (SEC, IPC). This served to slightly decrease the resultant CNF. However, the first two recommendations are the expected ones.

The next recommendation is IPC with a CNF of 70. This value of the CNF connotes a good choice (usually a second choice if the first choices are not available). SEC and IEC were third with individual CNF of 60. This value of a CNF suggests (as

programmed into the knowledge base) a choice that is not likely to provide a good separation or might require extensive optimization to achieve separation and resolution. These are still viable techniques since they have a CNF that is above the TruthThreshold value of 50.

**Table 4.8**  
**Output from Validation Study of the Hypothetical Case:**  
**VP-Expert Version**

| Recommend Mode                  | Confidence Factor |
|---------------------------------|-------------------|
| Large Pore Normal Phase Bonded  | 100               |
| Large Pore Reverse Phase Bonded | 90                |
| Large Pore Reverse Phase IPC    | 70                |
| Large Pore Reverse Phase SEC    | 60                |
| Large Pore IEC                  | 60                |

#### ***Section 4.8.4      Conclusions of Validation Studies***

These test cases were chosen to test representative components of the expert system/knowledge base. Additional validation studies were done to test different sub-modes of the system. While these are not presented in this thesis, they were fairly successful. Typically, the test cases often demonstrated holes in the knowledge base or syntax errors. A general conclusion can be made that the expert system will provide the optimum method as one of the first couple of proposed solutions. Additional modes will be recommended that have a logic to them.

The evolution of the expert system from M1 to HPLC ADVISOR has seen some important improvements in the knowledge base and inference engine. HPLC ADVISOR

provides a faster and more efficient inference engine. Even with the addition of new rules, the three knowledge bases from M1 were condensed into two knowledge bases (HPLC ADVISOR, version 3.0 allowed the consolidation into one knowledge base, but this was not done as mentioned earlier). HPLC ADVISOR is approximately two times faster than M1. Absolute speed comparisons are difficult to make since the two inference engines require different knowledge base formats, and the knowledge bases themselves quickly evolved into two different systems.

The evolution of the knowledge base itself is even more important than the improvements that occurred when switching from M1 to HPLC ADVISOR. Self-predictive rules allow for better treatment of UNKNOWN responses from the user and also allow the expert system to use data already stored during the consultation. These self-predictive rules encompass:

- the determination of labile components (with the increase in the CNF for SEC as a possible mode since SEC is the gentlest technique);
- prediction of proteins, dipoles, charge transfer, and so forth;
- determination of the presence of homologous series and the appropriate methods to use for them;
- calculation of SEC based on molecular weight distribution, average MW, and competing modes;
- determine the significance and relevance of particulate matter in the choice of both sample pre-treatment and stationary phase particle diameters (dirty compounds should not be used on very small analytical columns due to the potential for clogging).

Additional rules also "fine tune" the recommendations to provide more assistance for the different separation modes. During the development, the definition of whom the end-user would be was redefined. It was decided that the expert system

should provide help for those who are less versed with HPLC. Because of this it is necessary to provide additional help in selecting appropriate answers. This was accomplished by the inclusion of more descriptive questions, BECAUSE text (HOW and WHY facility), and hypertext files to give the user some fundamental information on how the recommended modes operate. The inclusion of data bases and text files also allow the expert system to provide detailed information that is tailored for that specific laboratory. Future expansion of the data bases will provide even more power to the expert system.

Finally, the inclusion of confidence factors provide a measure of reliability to the expert system. For expert systems to be used, the user must be able to trust the recommendations. Inclusion of the CNF allay fears that the methods are reliable. The user may also "hedge" his/her responses to questions when unsure and see a resulting decrease in certainty. This can encourage the user to seek additional information when a proposed method has low CNF. These CNF can also decrease the usage of the expert system if the CNF do not reflect the real probability of success. This double-edge sword requires that the expert system be monitored and refined as it is used. Therefore, the existence of a thesis describing the expert system should not indicate that the expert system is now completely stable and omniscient.

With the emphasis on CNF throughout this chapter, the final section of this chapter will discuss additional information regarding the incorporation of CNF. This section will be brief since much discussion of certainty theory and its application in this expert system has been mentioned in all of the previous chapters.



## **Section 4.9      Confidence Factors**

Confidence factors were not initially used in the prototype of the expert system. Once the expert system was transferred to VP-Expert, it was decided to implement confidence factors for several reasons. Confidence factors provide:

- Applications of fuzzy logic to the data that is input during the consultation.
- Condensation of multiple rules into fewer rules.
- Consideration of all possibilities throughout the consultation.
- Internal method to determine goals.
- Ranked presentation of all possibilities.
- A numerical descriptor of the reliability of the advice.

These motivations for using confidence factors are presented below. Following that will be a description of how confidence factors were implemented in this project.

### ***Section 4.9.1      Motivations for Using Confidence Factors***

#### **Section 4.9.1.1      Fuzzy Input**

It is necessary that the user be able to indicate a reliability to the answers that he/she supplies to the queries from the expert system. VP-Expert allows the user to supply confidences to every input. This is accomplished by hitting the HOME key at any input prompt. Confidence factors are implemented in VP-Expert using a scale from 0 to 100 (where 100 is absolutely true and 0 is absolutely false). The demarcation between true and false is established by the VP-Expert command, TRUTHTHRESH. A value of 50 was used for TruthThresh (the default value in version 1.1 was 70). This indicates that any rule which has a total confidence of greater than or equal to 50 will pass and the conclusion (the "Then" part of the rule) will be executed.

#### Section 4.9.1.2      Appropriate Values for Confidence Factors

Though fuzzy logic indicates that confidences should not decrease as supporting evidence is added (see Chapter 2), initial development of the expert system demonstrated that assigning too low of initial confidences caused rules to not fire when they were expected to do so. This problem arose from the way that confidence factors are calculated (*vide infra*). The confidence of the conclusion is based upon the confidence of the conclusion multiplied by the confidence of the premises. Some rules had conclusions that were weighted with lower confidence factors so that several rules could be combined into one. When the premise of a rule became fairly low, the resulting conclusion would end up having a confidence that would be below the truth threshold. Any further rules that used the value of this conclusion would no longer pass. Experimentally, a scale was established to maintain consistent "truths" as seen in Table 4.9. The ranges are subjective and are established so that the confidences of test cases would be accurate. That is, test cases with known methods were fed into the system and the confidences of the rules would be changed so that reasonable and consistent confidence factors would be generated.

Confidence factors have been implemented for all of the different modes that HPLC ADVISOR recommends. The most extensive of use of CNF has been for the determination of SEC as a possible solution to a given consultation. The addition of rules for SEC also required that a weighting factor to be applied to the recommendation since SEC depends only on size differences among the sample components. While other modes such as polarity are allowable, the predictive nature of the SEC technique—where the largest elute elutes first—requires that size be the only separation factor. Since components may have additional differences, it is necessary that a weighted recommendation (via confidence factors) of SEC be included to account for the presence or absence of other modes of separation.

The method of implementing CNF for SEC has all ready been described. Presented in Table 4.10 is the assignment of different CNF depending on different modes of separation being present. As different effects appear, SEC becomes less desirable. This may be due to a different mode being easier, faster, and better than SEC (e.g., charge transfer effects of particular substituents on an aromatic ring) or because these different effects will cause the SEC mechanism to not predominate (e.g., SEC separates based on size and shape; polar components may have a different partition effect with the stationary phase and the result is that the largest analytes may not elute first).

**Table 4.9**  
**Truth Threshold Values**

| <b>Range</b> | <b>Subjective Evaluation of Input</b>                                                                                                              | <b>Subjective Evaluation of Output</b>                                                                                                                                      |
|--------------|----------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 100 - 90     | A known fact for the user                                                                                                                          | First or Best Choice of action                                                                                                                                              |
| 90 - 75      | The user is certain of the information. The user would only obtain confirmation of the information if she/he checked with a more knowledge source. | Good alternative. Will provide the required separation if the first choice is not available.                                                                                |
| 75 - 50      | The use is unsure of the validity of the information. The user needs to confirm the validity of this data.                                         | The recommendation is questionable. Additional information is needed (i.e., further experimentation). Use this method if all other choices are not available or do not work |
| 50 - 0       | The data is invalid.                                                                                                                               | Method is unlikely to succeed. Most consultations will not present any solutions below 50 (or the Truth threshold value).                                                   |

**Table 4.10**  
**Assignment of CNF for SEC based upon**  
**Competing Mechanisms**

| Size | Polar | Arom | Sub | Geo | Opt | Ion | Lab | CNF |
|------|-------|------|-----|-----|-----|-----|-----|-----|
| Y    | N     | N    | N   | N   | N   | N   | N   | 100 |
| Y    | N     | N    | N   | —   | —   | N   | Y   | 90  |
| Y    | N     | N    | N   | Y   | Y   | N   | N   | 85  |
| Y    | N     | N    | Y   | Y   | Y   | N   | N   | 80  |
| Y    | N     | N    | Y   | Y   | Y   | C   | N   | 75  |
| Y    | N     | Y    | Y   | Y   | Y   | N   | Y   | 70  |
| Y    | N     | Y    | Y   | Y   | Y   | I   | N   | 65  |
| Y    | Y     | Y    | Y   | Y   | Y   | N   | Y   | 60  |
| Y    | Y     | Y    | Y   | Y   | Y   | I   | Y   | 50  |

**Legend and Remarks:**

| <u>Heading</u> | <u>Description</u> | <u>Comment</u>                                                                              |
|----------------|--------------------|---------------------------------------------------------------------------------------------|
| Size           | Size               | In this set up, all entries has some size characteristics.                                  |
| POL            | Polar              | Is there a difference in the polarity of the samples.                                       |
| AROM           | Aromatic           | Sample contains aromaticity.                                                                |
| SUB            | Substituents       | Substituents are present which will lead to a recommendation to use packings such as cyano. |

### Legend and Remarks (continued)

| <u>Heading</u> | <u>Description</u> | <u>Comment</u>                                                                                                                                                                                                                                                                          |
|----------------|--------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| GEO            | Geometric          | Presence of geometric isomers. Typically, there will not be any size differences for geometric isomers; however, there could be samples which have large MW differences and have geometric isomers at the same time. You will note that this column does not affect the CNF's very much |
| OPT            | Optical            | Optical isomers present. Same comments that apply for geometric isomers apply here also.                                                                                                                                                                                                |
| ION            | Ionic              | N: Non-ionic. C: Ionic but the pH can be adjusted to leave the species in their neutral form. I: Ionic.                                                                                                                                                                                 |
| LAB            | Labile             | Are there labile compounds present. This serves to raise the CNF since SEC is a gentle technique. Does not denature proteins, for example.                                                                                                                                              |

While this table is not complete, it does cover most of the different possibilities. One limitation to this approach is that the knowledge base must contain a rule for each combination in the table. This lead to a many rules which had slight variations in them. One possibility is to use some of new features of HPLC ADVISOR, version 3.0. Two new commands, GETCNF, and MODCNF, allow the confidence factors of different variables to be obtained, mathematically modified, and re-assigned to the variables. One could determine the CNF that has been assigned to the different modes and then perform a weighted subtraction of these CNF from the current CNF of SEC. A weighted subtraction is necessary since different competing modes will over-ride SEC differently (e.g., Polar effects have the largest effect on the mechanism involved in SEC).

## **Section 4.10 Ancillary Files—Text, Help Files, and Data Bases**

New methods also apply more than knowledge base--use fact bases, text bases, etc. This allows information to be stored in more appropriate formats (more condensed, easier to change, more dynamic). General rules (heuristics) which are global information, more static in nature, can be encoded in the knowledge base in the format of IF-THEN rules. These form the skeleton of the knowledge base. Then on top of that exists the data bases (usually in the form of data bases or spreadsheets). These data bases would be created from a separate program (e.g., Lotus, DBASE). They would incorporate dynamic information that might be more local in nature. For example, a data base could hold information concerning the HPLC columns that are available. The expert system would then use this information to generate a solution (method) that would be optimized for that lab location. There would be no need to suggest a particular column if it were not available. Or better yet, to suggest with a higher confidence the best column while giving a solution that would work with the columns that are available. This would allow the user to decide whether a one-time separation could be accomplished on the columns that he/she has or whether a particular column might be a good investment. The incorporation of confidence factors will also allow the user to decide if the available column will suffice.

Other supporting files are text files. These incorporate help screens when the user requests a more "wordy" explanation or a question in a less terse fashion. The help screens would also include reference material or perhaps experimental procedures (e.g., pre-/post-column derivitization schemes). By separating these text files outside of the knowledge base, the knowledge base becomes much smaller. This allows the inference machine (and hence the expert system itself) to be much faster. One prime consideration in developing an expert system that will be a useful tool is that the whole process of developing a method should be around 30 minutes or less in time. Much

longer than this and the user becomes impatient with the process and is less likely to use the tool. By having a smaller knowledge base, the knowledge engineer will be able to maintain the system much easier. The text files can then be put together by more productive tools such as word processors. Since they will not create "bugs", they can be put into the system at a later date, can be changed much easier, and user-friendly features such as correct spellings/grammar via spelling and grammar checkers can be used to make the text "easier" on the eyes.

## **LIST OF REFERENCES**



## LIST OF REFERENCES

- 1 M. De Smet, A. Peeters, L. Buydens, and D.L. Massart, *J. Chromatogr.*, 1988, 457, 25.
- 2 T.P. Bridge, M.H. Williams, G.G.R. Seaton, and A.F. Fell, *Chromatogr.* 1987, 24, 691.
- 3 J. A. Van Leeuwen, B. G. M. Vandeginste, and G. Kateman, *Anal. Chim. Acta* 1990, 228, 145.
- 4 S. S. Williams, J. F. Karnicky, J.-L. Excoffier, and S. R. Abbott 'Expert System Program for Assistance in High-Performance Liquid Chromatographic Method Development' in *Computer-Assisted Method Development for High-Performance Liquid Chromatography* J. L. Glajch and L. R. Snyder, eds. Elsevier Science Pub., Amsterdam, The Netherlands, 1990, 267.
- 5 R. Bach, J. Karnicky, and S. Abbott, in *Artificial Intelligence Applications in Chemistry*, T. H. Pierce and B. A. Hohne, eds., Wiley-Interscience, New York, 1988.
- 6 E. D. Salin and P. H. Winston *Anal. Chem.* 1992, 64, 49A.
- 7 C. Cachet, D. Cabrol, and A. Milleliri *J. Chem. Ed.* 1989, 66(7), 553.
- 8 B. Schaefer, R. Side, and I. Morrison *PC AI* 1992, 6(4), 20.
- 9 P.E. Dessert *AI Expert* 1991, 6(12), 36.
- 10 *Waters Sourcebook of Chromatography*, Millipore Corporation, Milford, MA, 1992.
- 11 L. R. Snyder and J. J. Kirkland, *Introduction to Modern Liquid Chromatography*, John Wiley and Sons, Inc., New York, 1979.
- 12 John R. Walters and Norman R. Nielsen *Crafting Knowledge-Based Systems (Expert Systems Made Realistic)* Wiley-Interscience Pub., New York 1988, 121-43.
- 13 Lecture and slide notes. Separations Course, Michigan State University, East Lansing, Michigan.
- 14 G. Schmidt 'Analysis of Estrogens by HPLC' in *Steroid Analysis by HPLC*, M.P. Kautsky, ed. Chromatographic Science Series Vol. 16 Marcel Dekker, NY (1981) 145.

- 15 **Modern Chromatographic Analysis of Vitamins, A. P. De Leenheer, W. E. Lambert, and H. J. Nelis, eds. Marcel Dekker, Inc. New York (1992).**

MICHIGAN STATE UNIV. LIBRARIES



31293010192338



**PLACE IN RETURN BOX** to remove this checkout from your record.  
**TO AVOID FINES** return on or before date due.

| DATE DUE | DATE DUE | DATE DUE |
|----------|----------|----------|
| _____    | _____    | _____    |
| _____    | _____    | _____    |
| _____    | _____    | _____    |
| _____    | _____    | _____    |
| _____    | _____    | _____    |
| _____    | _____    | _____    |
| _____    | _____    | _____    |

**DEVELOPMENT OF AN INTELLIGENT AUTOMATED  
HIGH PERFORMANCE LIQUID CHROMATOGRAPHY  
SYSTEM**

**Volume II**

**By  
Stephen Vance Medlin**

**A DISSERTATION**

**Submitted to  
Michigan State University  
in partial fulfillment of the requirements  
for the degree of**

**DOCTOR OF PHILOSOPHY**

**Department of Chemistry**

**1993**

recon

logic

and c

phase

mobili

## **CHAPTER 5**

### **EXPERT SYSTEM: DETECTOR AND MOBILE PHASE SELECTION AND VALIDATION**

As soon as we started programming, we found to our surprise that it wasn't as easy to get programs right as we had thought. Debugging had to be discovered. I can remember the exact instant when I realized that a large part of my life from then on was going to be spent in finding mistakes in my own programs.

Maurice Wilkes discovers debugging, 1949

This chapter is divided into four parts. The first part details the recommendations of an appropriate detector. The second part describes the rules and logic used to recommend mobile phase based upon the recommendation of a column and detector. The third part presents validation studies for both detector and mobile phase. Conclusions based on the development and validation of the detector and mobile phase modules are presented in the final part.



## **PART 1: RECOMMENDATION OF DETECTOR**

**Science without religion is lame; religion without science is blind.**

**Albert Einstein**

Se

ap

th

co

fu

ele

de

de

sa

tha

de

alr

wo

hav

Wh

obr

rem

elu

san

ald

## **Section 5.1      Overview of the Detector Module**

The previous sections have described the selection and recommendation of an appropriate column. This section describes the expert system module which assists in the selection of an appropriate detector. The information used to propose a suitable column is also utilized for the selection of the detector. Databases that contain functional groups properties such as molar absorptivities, quantum efficiencies, and electrochemical potentials are integrated into the inference process. The expert system determines confidence factors based on concentration of solutes and detector response.

### ***Section 5.1.1      Goals of the Detector Module***

The principle objective of the detector module is to recommend an appropriate detector for the quantitative and qualitative analysis of the solutes in a particular sample. Several secondary objectives were set to achieve the main objective.

The first objective was to have the expert system incorporate the information that was used to recommend the column into the detector selection process. It was determined that much of the information that is required to recommend a detector is already obtained during the selection of the column. It was reasoned that the user would prefer to supply most of the information early during the consultation and then have the expert system use these data during the detector recommendation stage. While additional questions are still required, most of the required information was obtained by editing and expanding the questions used earlier in the consultation. The remaining information that was needed for the detector module consisted primarily of elucidating the chemical nature of the solutes. For example, if the user indicates that a sample contains aldehydes, the detector module queries the user to determine if the aldehydes are conjugated with carbon-carbon double or triple bonds. This elucidation

may be used to determine optimum wavelengths and molar absorptivities for UV-Visible detectors.

The expert system should propose confidence factors that use fundamental solute information in addition to the subjective input from the user. In our approach, the expert system uses concentration-response functions, such as Beer's Law, that are used to indicate acceptable concentration ranges for each detector. The expert system generates confidence factors that reflect the extent that the solute concentrations are within the concentration range of the detector.

Finally, we want the expert system to remain heuristically-based. This goal allows the expert system to make recommendations for any chemical systems—provided the user has chemical and physical information about the solutes. To achieve this goal, the expert system (HPLC Advisor) uses functional group information. For example, it will recommend specific wavelengths for UV-Visible detectors based upon the general molar absorptivities of different functional groups. While these generalized functional group parameters are not as accurate as tabulated values for specific compounds, HPLC Advisor will still be able to make recommendations for solutes that are not specifically in the relevant databases and knowledge bases.

### ***Section 5.1.2 Detectors recommended by the Expert System***

The first step in developing the detector module was to determine which of the different HPLC detectors would be recommended by the expert system. Table 5.1 lists these detectors. Atomic emission, absorption, and fluorescence spectroscopic detectors are grouped into the comprehensive category of Atomic. Amperometric and voltaic electrochemical detection modes are also grouped into the single category of EChem. It

was decided that the intended audience would not have the knowledge required to subclassify these categories.

The seven detectors in Table 5.1 were chosen for several reasons. They are representative detectors from both bulk- and specific-property classes. The expert system attempts to recommend a specific detector, provided that the solutes are all active for that detector; otherwise, the expert system is still able to recommend one of the bulk detectors. For example, UV-Vis is recommended only if all of the solutes absorb UV-Visible radiation. If there are solutes which do not absorb light to a significant extent, the expert system is still able to recommend one of the bulk detectors (provided the solutes meet the criteria for the bulk detectors).

The set of detectors were chosen to provide detection capabilities for all of the different columns that may be recommended by HPLC Advisor. These criteria led to the inclusion of Atomic detectors to provide detection of metal species (i.e., Ion Exchange columns), Scatter for large molecules (SEC), and so on.

Two additional criteria used to select the set of detectors that would be recommended are the availability of the detector and the amount of information needed to recommend the detector. It was decided to not include mass spectrometry (MS) detectors due to the cost, availability, and complexity of MS instruments. Some of the obsolete detectors, such as moving wire, were excluded as were detectors that are still in the research stage and thus not commercially available.

### ***Section 5.1.3      Format of the Presentation of the Different Detectors***

The following section describes an overview of the methodology to calculate confidence factors. The specifics of these calculations for each detector are given in the relevant section for that detector.

1

a

P

d

V

f

S

d

b

ac

op

cc

de

**Table 5.1**  
**Detectors Recommended by the Expert System**

| Detector            | Abbreviation | Detector Class |
|---------------------|--------------|----------------|
| Refractive Index    | RI           | Bulk           |
| UV-Visible          | UV-Vis       | Specific       |
| Fluorescence        | Fluor        | Specific       |
| Atomic Spectroscopy | Atomic       | Specific       |
| Conductivity        | Conduct      | Bulk           |
| Electrochemical     | EChem        | Specific       |
| Light Scatter       | Scatter      | Bulk           |

The different detectors that are recommended have both similar and unique aspects. As such, each detector is presented separately. A flow chart of the decision processes that the expert system uses to recommend the detector is presented and described. The databases used for each detector are also presented. The section on UV-Visible detectors is the most detailed; subsequent descriptions describe any differences from the approach used for the selection of UV-Visible detectors.

## **Section 5.2      Confidence Factors Based on Concentration**

Confidence factors for column selection are empirically-based. The choice of detector depends upon several factors. First, all solutes must be detectable (i.e., active) by a particular detector. Once HPLC Advisor has determined that the species are all active for that detector, it then determines if the solute concentrations are in the linear operating range. It was decided to apply a more fundamental determination of confidence factors based on this concentration requirement.

The first decision was to determine an appropriate response function for each detector, if possible. While some functions were obvious (e.g., Beer's Law), additional

cr

av

th

co

Vl

sh

ap

u

p

sl

o

co

tr

d

u

ti

T

d

b

v

r

a

v

(



criteria were considered. First, the required parameters for the function must be available to the expert system via a database or a user-supplied response. Therefore, the parameters must be well documented and should be information a junior/senior college chemistry student would know. Due to the limited mathematical feature set of VP-Expert, the functions should not be mathematically intensive. Finally, the functions should use parameters based on functional group properties to maintain a heuristic approach. These considerations led to simplification of many of the functions that are used. Each function is detailed in the following sections.

Once an appropriate response function was determined, it was correlated to a particular confidence factor (CNF). The CNF should have several features. The CNF should have a value of 100 when the concentrations of the solutes are within the optimum concentration range of the detector. The CNF should then decrease as the concentrations vary from this optimum range until a CNF of 50 (or the value of the truth threshold) when the concentrations are beyond the linear operating range of the detector. Rather than using an average solute concentration, it was decided to use the upper and lower concentrations to determine the CNF. The user may not know all of the concentrations of the solutes, but may have a estimation of the concentration range. The most and least concentrated solutes provide worst case scenarios to insure that the detector will be able to detect all of the solutes. Finally, the two different CNF values based on the most and least concentrated solutes must be combined into a total CNF value for all of the solutes.

A trapezoidal function was chosen where the vertices of the function are related to the absolute maximum and minimum concentrations, and typical minimum and maximum concentrations. These absolute concentration points define the point at which the CNF is 50. The absolute maximum and minimum concentrations ( $C_{\text{Max}}$  and  $C_{\text{Min}}$ ) are determined by either accessing a database to obtain the values or by

calculating values based on the minimum and maximum response values ( $R_{\text{Min}}$  and  $R_{\text{Max}}$ ) of the detector and on solute functional group information. The typical minimum and maximum concentrations, obtained from a database, establish the plateau region that defines a concentration range in which the CNF has a value of 100. A linear equation is derived between the  $C_{\text{Min}}$  to typical minimum concentration points and a second linear equation is derived for the corresponding maximum points to give linear equations ( $\Delta_1$  and  $\Delta_2$ ). These equations are used to calculate a CNF for the low and high concentrations of the solutes. Figure 5.1 shows a pictorial representation of the CNF versus concentration relationship. This approach allows the expert system to apply a CNF adjustment for species that are near the detection limit or near the non-linear concentration range. CNF values that are below 50 or above 100 are adjusted to 50 and 100.

After the expert system has determined CNF values for the two outlier concentrations, the expert system combines these two values ( $\text{CNF}_L$  and  $\text{CNF}_H$ ) into a total CNF ( $\text{CNF}_{\text{Total}}$ ). Several different methods can be used to determine  $\text{CNF}_{\text{Total}}$ . The first approach is to average the low ( $C_L$ ) and high ( $C_H$ ) concentrations and then use the relevant  $\Delta_1$  or  $\Delta_2$  equation to calculate  $\text{CNF}_{\text{Total}}$  (Equation 5.1). A second approach is to average the  $\text{CNF}_L$  and  $\text{CNF}_H$  (Equation 5.2). Finally,  $\text{CNF}_L$  and  $\text{CNF}_H$  can be combined using the method that VP-Expert combines CNF values as show in Equation 5.3.

To determine empirically which method will provide the desired  $\text{CNF}_{\text{Total}}$ , a series of calculations was performed comparing the three approaches above. Five different concentrations were chosen for the lowest concentrated species, denoted very low (VL), low (L), medium (M), high (H), and very high (VH). The values of VL and VH are 10% of the threshold concentrations that give a CNF of 50 and 100, respectively. L is 33% and H is 66% of the concentration range that delineates the linear (e.g.,  $\Delta_1$ )

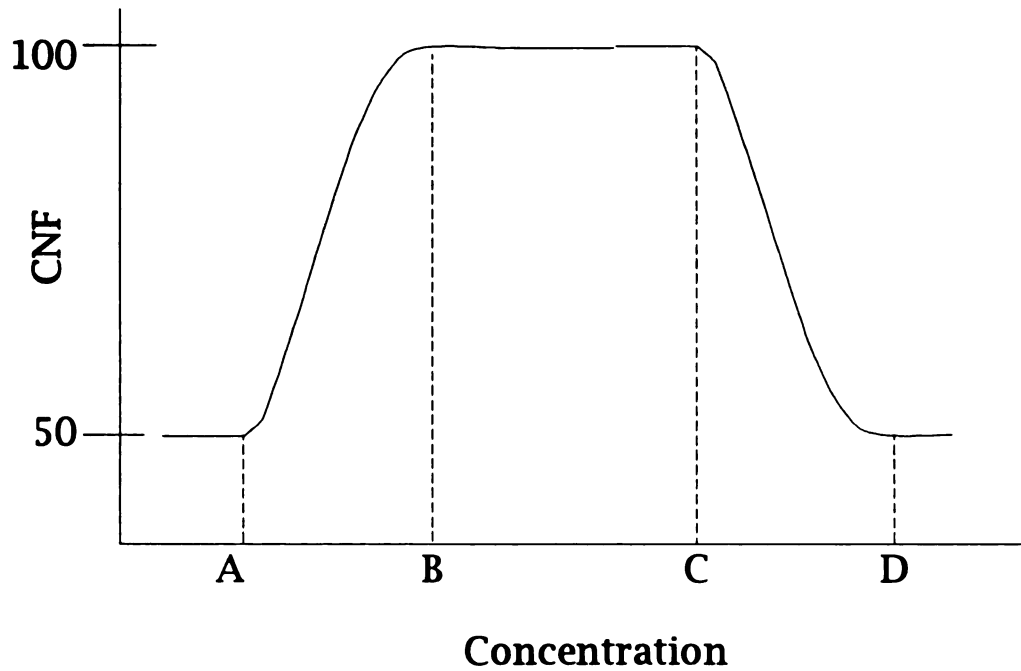
increase from a CNF of 50 to 100. M is 50% of this concentration range. The same set of concentrations were chosen for the most concentrated species. The set of concentrations can be seen graphically in Figure 5.2.

$$\text{CNF}_{\text{Total}} = \Delta_1 \text{ or } \Delta_2 \text{ using } [\text{Average}(C_L, C_H)] \quad \text{Equation 5.1}$$

$$\text{CNF}_{\text{Total}} = \text{Average}(\text{CNF}_L, \text{CNF}_H) \quad \text{Equation 5.2}$$

$$\text{CNF}_{\text{Total}} = \text{CNF}_L + \text{CNF}_H - (\text{CNF}_L * \text{CNF}_H)/100 \quad \text{Equation 5.3}$$

The extreme values of VL and VH test the effect of having compounds whose concentrations have the most significant effect on  $\text{CNF}_{\text{Total}}$ . L, M, and H are typical cases where the concentration is important. Combinations of VL to VH for both the  $C_L$  and  $C_H$  were used to determine which method (Equations 5.1-5.3) would provide the desired  $\text{CNF}_{\text{Total}}$ . The concentration values used for the thresholds (Absolute Minimum/Maximum and Typical Minimum/Maximum) are 1, 10, 100, and 10000  $\mu\text{M}$ . The results of these calculations are presented in Table 5.2. Methods 1-3 refer to  $\text{CNF}_{\text{Total}}$  calculations based on Equations 5.1-5.3.



- A Absolute Minimum Concentration,  $C_{\text{Min}}$
- B Typical Minimum Concentration
- C Typical Maximum Concentration
- D Absolute Maximum Concentration,  $C_{\text{Max}}$

**Figure 5.1**  
**Trapezoidal Function for the Calculation of Concentration**  
**Confidence Factors**

The conclusion was that Equation 5.2 (an average of  $CNF_L$  and  $CNF_H$ ) provides the desired effect. Equations 5.1 and 5.3 give too large a value for  $CNF_{Total}$  when  $C_L$  or  $C_H$  is either VH or VL. One disadvantage to Equation 5.2 is that it does not consider the distribution of all of the concentrations. In the case where VL or VH is an outlier and the rest of the concentrations of the solutes are within the linear range of the detector, methods 1 and 3 may reflect the more desired condition in which one concentration would not lower the total CNF. Three points were made to counter this argument: (1) Typically, concentrations are going to be more evenly distributed between the high and low concentrations, (2) The user may not know the concentrations of all the species to calculate an average C, and (3) If the  $C_L$  or  $C_H$  species is of interest, then the CNF must reflect the desire to detect all species of interest—thus, an outlier should significantly decrease  $CNF_{Total}$ .

The expert system also applies a concentration threshold. If  $C_L$  or  $C_H$  is below or above the threshold, the expert system indicates that the concentration range is inappropriate for the specified detector. If  $CNF_L$  or  $CNF_H$  has a value below the threshold, but the other outlier has a CNF of ~100%, the  $CNF_{Total}$  may be greater than 50%. The threshold limit insures that this detector should not be used (provided that dilution/pre-concentration is not an option).

S

ab

ty

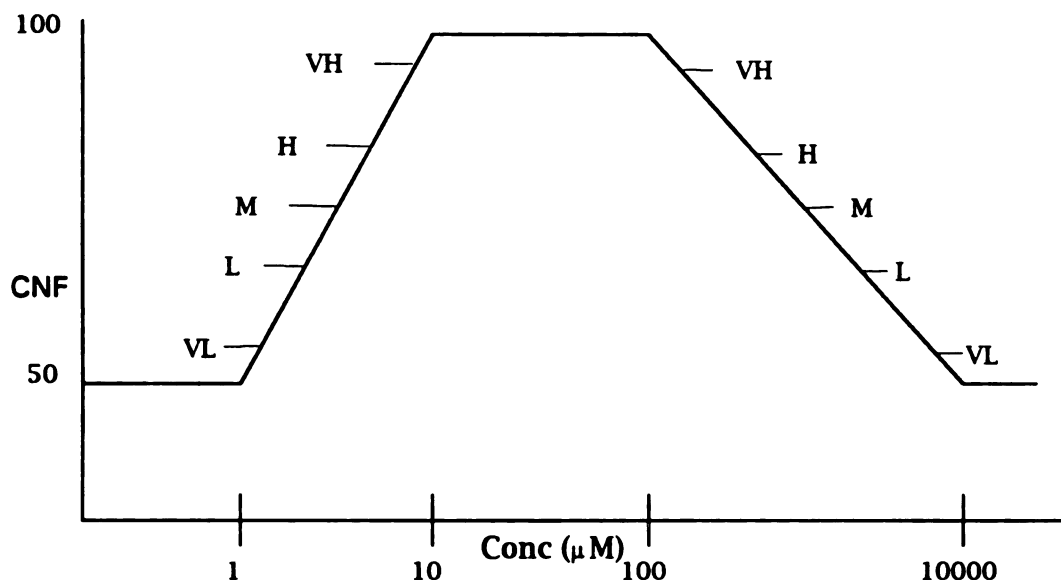
to

re

w

co

is



**Figure 5.2**  
**Different Concentration Values to Determine**  
**Optimum  $CNF_{Total}$  Calculation**

## **Section 5.3     UV-Visible Detectors**

### ***Section 5.3.1     Overview***

This section describes the criteria that must be met before the expert system is able to recommend UV-Visible detectors. HPLC Advisor first determines the different types of UV-Visible functional group chromophores that are present and uses a database to obtain molar absorptivities,  $\epsilon$ , for each chromophore at six different wavelength regions from 190 nm to >325 nm. The expert system determines an optimum wavelength and uses the  $\epsilon$  values for that wavelength to calculate the CNF and concentration thresholds. Figure 5.3 provides an abbreviated flow chart of the logic that is used to determine if UV-Visible is an appropriate detector.

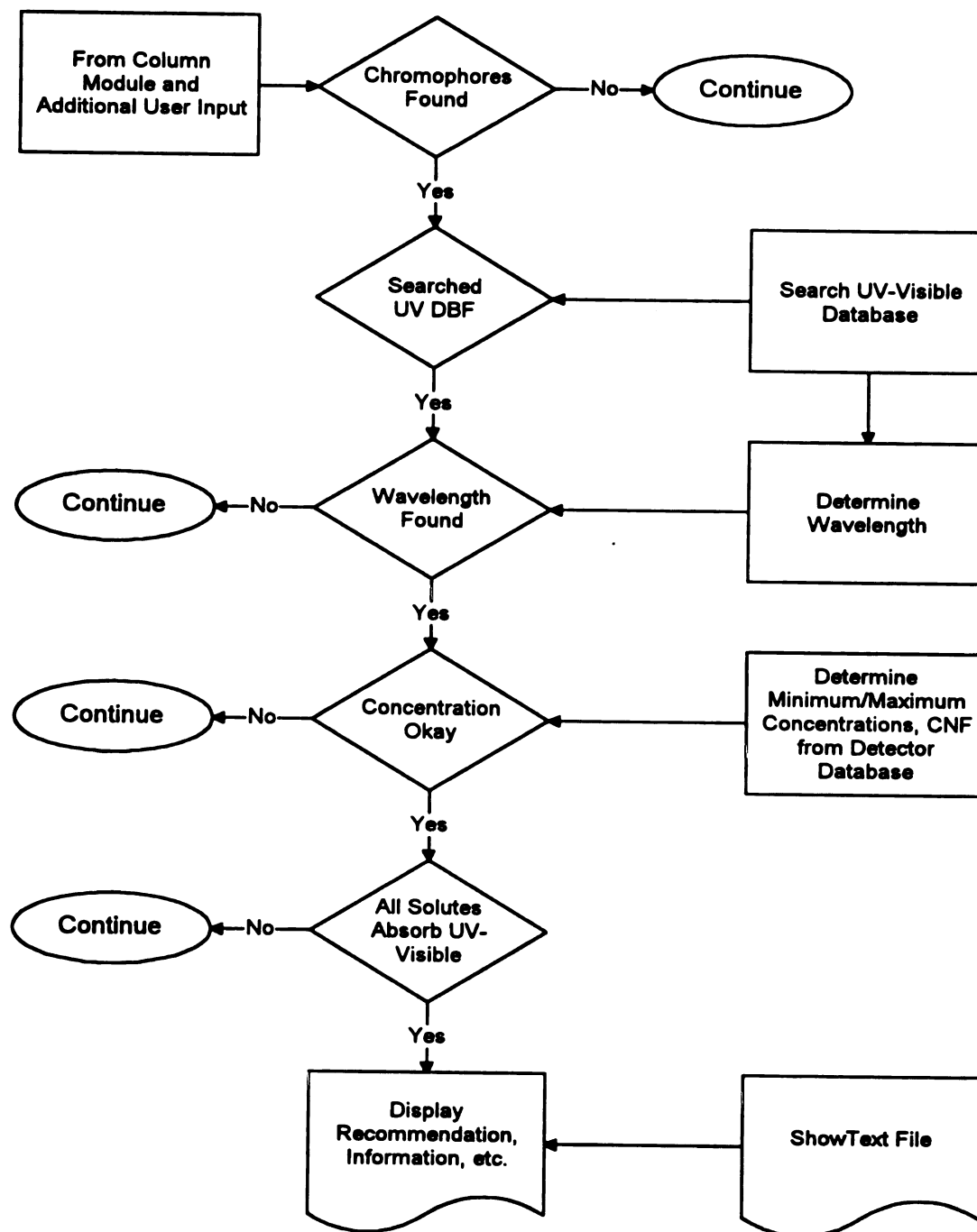
### ***Section 5.3.2      Determination of Chromophores***

During the column selection process the expert system determines the different functional groups present on the solutes. The expert system uses this information and asks additional questions to define further the different functional groups that are present. For example, if the user indicated that the sample contains aliphatic aldehydes, the expert system will query the user to determine if these aldehydes are conjugated with carbon/carbon double or triple bonds. The molar absorptivities of conjugated aldehydes are different than that of unconjugated aldehydes. The possible functional groups that the expert system can assign as UV-Visible chromophores are listed in Table 5.3. The aliphatic and aromatic forms are listed along with the number of conjugated forms (if any) of the aliphatic form of the chromophore. HPLC Advisor also insures that all of the solutes absorb radiation.



**Table 5.2**  
**Determination of the Optimum Method to Calculate the CNF based**  
**on the CNFs of Two Different Concentrations**

| Concentration Combinations | C <sub>L</sub> | C <sub>H</sub> | CNF <sub>L</sub> | CNF <sub>H</sub> | Method 1 | Method 2 | Method 3 |
|----------------------------|----------------|----------------|------------------|------------------|----------|----------|----------|
| VL, VL                     | 1.1            | 9000           | 50.5             | 54.6             | 77.6     | 52.6     | 77.5     |
| VL, L                      | 1.1            | 6633           | 50.5             | 66.7             | 83.6     | 58.6     | 83.5     |
| VL, M                      | 1.1            | 5050           | 50.5             | 74.8             | 87.6     | 62.6     | 87.5     |
| VL, H                      | 1.1            | 3467           | 50.5             | 82.8             | 91.7     | 66.7     | 91.5     |
| VL, VH                     | 1.1            | 110            | 50.5             | 99.9             | 100.0    | 75.2     | 100.0    |
| L, VL                      | 3              | 9000           | 61.1             | 54.6             | 77.6     | 57.8     | 82.3     |
| L, L                       | 3              | 6633           | 61.1             | 66.7             | 83.6     | 63.9     | 87.0     |
| L, M                       | 3              | 5050           | 61.1             | 74.8             | 87.6     | 67.9     | 90.2     |
| L, H                       | 3              | 3467           | 61.1             | 82.8             | 91.7     | 72.0     | 93.3     |
| L, VH                      | 3              | 110            | 61.1             | 99.9             | 100.0    | 80.5     | 100.0    |
| M, VL                      | 5              | 9000           | 72.2             | 54.6             | 77.5     | 63.4     | 87.4     |
| M, L                       | 5              | 6633           | 72.2             | 66.7             | 83.6     | 69.4     | 90.7     |
| M, M                       | 5              | 5050           | 72.2             | 74.8             | 87.6     | 73.5     | 93.0     |
| M, H                       | 5              | 3467           | 72.2             | 82.8             | 91.7     | 77.5     | 95.2     |
| M, VH                      | 5              | 110            | 72.2             | 99.9             | 100.0    | 86.1     | 100.0    |
| H, VL                      | 7              | 9000           | 83.3             | 54.6             | 77.5     | 69.0     | 92.4     |
| H, L                       | 7              | 6633           | 83.3             | 66.7             | 83.6     | 75.0     | 94.4     |
| H, M                       | 7              | 5050           | 83.3             | 74.8             | 87.6     | 79.0     | 95.8     |
| H, H                       | 7              | 3467           | 83.3             | 82.8             | 91.7     | 83.1     | 97.1     |
| H, VH                      | 7              | 110            | 83.3             | 99.9             | 100.0    | 91.6     | 100.0    |
| VH, VL                     | 9.9            | 9000           | 99.4             | 54.6             | 77.5     | 77.0     | 99.7     |
| VH, L                      | 9.9            | 6633           | 99.4             | 66.7             | 83.6     | 83.1     | 99.8     |
| VH, M                      | 9.9            | 5050           | 99.4             | 74.8             | 87.6     | 87.1     | 99.9     |
| VH, H                      | 9.9            | 3467           | 99.4             | 82.8             | 91.6     | 91.1     | 99.9     |
| VH, VH                     | 9.9            | 110            | 99.4             | 99.9             | 100.0    | 99.7     | 100.0    |



**Figure 5.3**  
**Flow Chart to Determine if a UV-Visible Detector**  
**is an Appropriate Choice**

**Table 5.3**  
**UV-Visible Chromophores Assigned by**  
**the Expert System**

| Aliphatic Form   | Aromatic Form       | Conjugated Aliphatic Form |
|------------------|---------------------|---------------------------|
| Carboxylic Acids | Carboxylic Acids    | 2 to >5 Conjugated        |
| -NA-             | Benzene             | -NA-                      |
| -NA-             | 2 to >5 Fused Rings | -NA-                      |
| Alcohols         | Alcohols            | -NA-                      |
| Aldehydes        | Aldehydes           | 2 to >5 Conjugated        |
| Ketones          | Ketones             | 2 to >5 Conjugated        |
| Amines           | Amines              | -NA-                      |
| Cyano            | Cyano               | -NA-                      |
| Alkanes          | -NA-                | -NA-                      |
| Alkenes          | -NA-                | 2 to >5 Conjugation       |
| Nitro Groups     | Nitro Groups        | -NA-                      |
| Thiols           | Thiols              | -NA-                      |

### ***Section 5.3.3      Determination of Optimum Wavelength***

Once HPLC Advisor has determined the chromophores present, it accesses the UV\_VIS.DBF database which contains molar absorptivities. The spectral region from 190 to 325 nm is divided into six wavelength regions. Each region is 30 nm wide and centered at 190, 220, 250, 280, 310, and >325 nm. Each chromophore has a molar absorptivity value,  $\epsilon$ , for each wavelength region. The expert system obtains the molar absorptivities of each chromophore and applies an empirical optimization function, EOF (for  $\epsilon$  Optimization Function) to each wavelength region (six EOF values are calculated). The EOF is given in Equation 5.4. A threshold  $\epsilon$  ( $\epsilon_{\text{Threshold}}$ ) determines a minimum molar absorptivity that will give a significant absorption. Any  $\epsilon$  values below the

threshold are incremented by 1 above  $\epsilon_{\text{Threshold}}$  so that all EOF values are positive and defined. In the calculation of EOF, smaller  $\epsilon$  values contribute more significantly than large  $\epsilon$  values. The EOF values reflect, therefore, the importance of the more weakly absorbing chromophores in determining optimum wavelengths in which all of the chromophores absorb. After an EOF is calculated for each wavelength region, the expert system sorts the EOF values from optimum (minimum EOF) to worst (maximum EOF). HPLC Advisor then obtains the minimum and maximum  $\epsilon$  value for each (sorted) wavelength region and determines the absolute minimum and maximum concentrations that can be analyzed with a particular detector (described below).

$$\text{EOF} = \sum_1^n \left( \frac{1}{(\epsilon - \epsilon_{\text{Threshold}})} \right) \quad \text{Equation 5.4}$$

$n = \text{Number of Chromophores}$

#### **Section 5.3.4      Determination of Concentration Range**

The expert system determines if the concentration range of the solutes of a sample are within the dynamic range of the detector. It calculates a CNF value that reflects the certainty that the solutes will be detected by a UV-Visible detector.

The first step in calculating the minimum and maximum concentrations is to obtain detector parameters from the DETECTOR.DBF file. This database contains information regarding  $R_{\text{Min}}$  and  $R_{\text{Max}}$  of detectors, typical minimum and maximum concentrations, and manufacturer's information (e.g., model name, company). Beer's Law ( $A = b\epsilon C$  with a unit pathlength assumed) is used to calculate  $C_{\text{Min}}$  and  $C_{\text{Max}}$ . The absorbance,  $A$ , is either  $R_{\text{Min}}$  and  $R_{\text{Max}}$  and the molar absorptivities are the minimum and maximum  $\epsilon$  values in the wavelength region. The expert system compares the

solute concentrations,  $C_L$  and  $C_H$ , to  $C_{Min}$  and  $C_{Max}$ . If a wavelength region has  $\epsilon$  values that give  $C_L$  and  $C_H$  values between  $C_{Min}$  and  $C_{Max}$ , HPLC Advisor recommends UV-Visible detectors.

## **Section 5.4 Fluorescence Detectors**

### ***Section 5.4.1 Overview***

Fluorescence detectors are highly selective and sensitive detectors. Before recommending fluorescent detectors, the expert system must determine if all species are fluorescent and if the concentrations of the solutes are within the dynamic range of the detector. The expert system uses two databases for this determination. The first database, FLUOR.DBF, contains the same chromophores that are listed in UV\_VIS.DBF and has fields that indicate if the functional groups are fluorescent and the approximate quantum yield of the group. Quantum yields are dependent upon pH, mobile phase, and other parameters; therefore, the quantum yields are average values for the functional groups. The second database, DETECTOR.DBF, provides instrumental figures of merit for different fluorescence detectors. Figure 5.4 shows an abbreviated pictorial representation of the logic used to assign a fluorescence detector.

### ***Section 5.4.2 Determination of Fluorophores***

HPLC Advisor uses the UV-Vis chromophores information for the fluorescence detector rules. A database, FLUOR.DBF, contains a field that indicates if the functional groups are fluorescent. The expert system determines if the sample contains other non-fluorescent species, such as proteins and metals. Since fluorescence is dependent on the mobile phase and pH, HPLC Advisor stores any functional groups/solutes that are determined to be non-fluorescent. After checking the concentration ranges for the

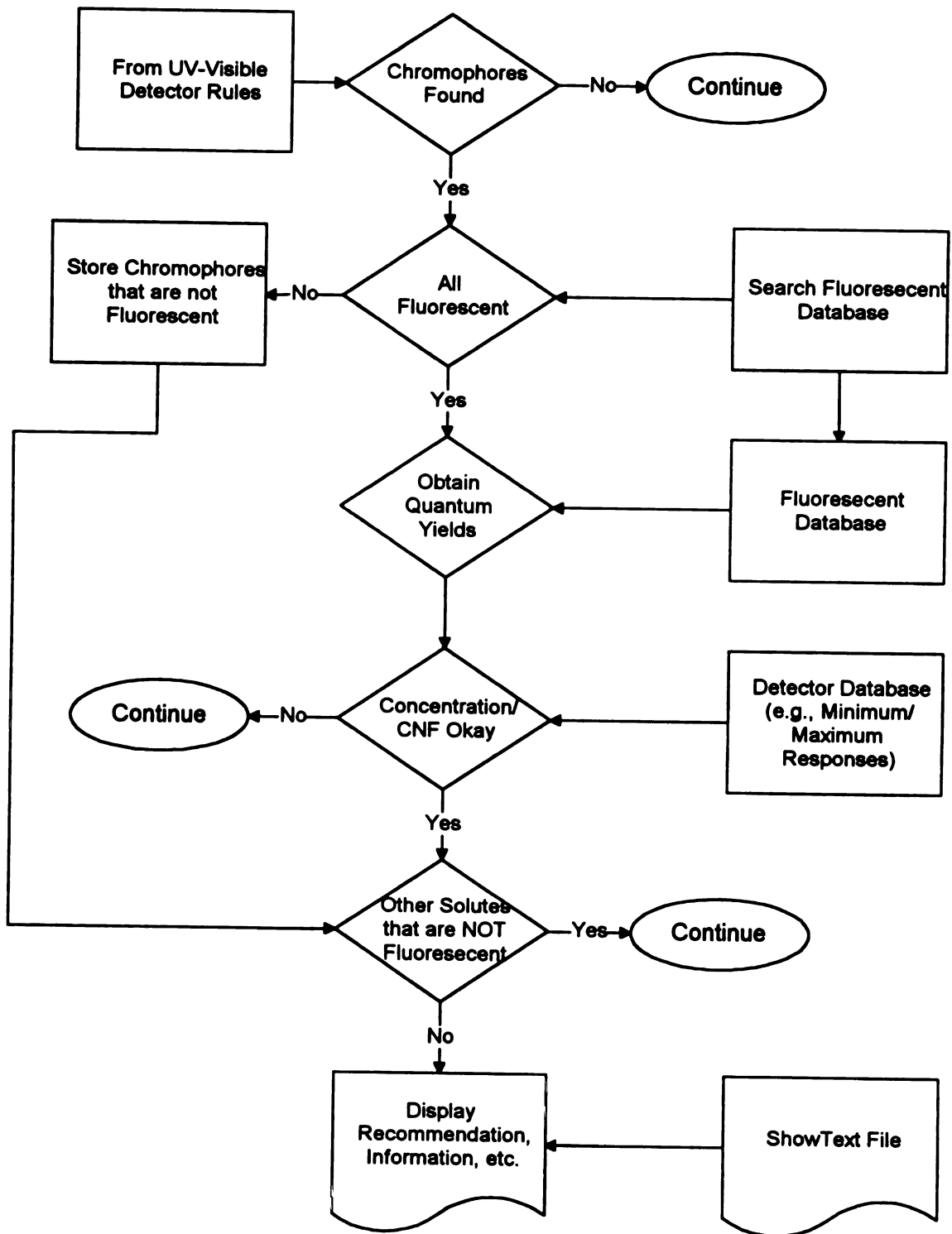
fluorescent species, HPLC Advisor reports which species are non-fluorescent. This allows the user to decide on whether to precede with using fluorescence detectors based on additional information not available to the expert system.

### ***Section 5.4.3      Determination of Concentration Range and CNF***

Two different equations<sup>1,2</sup> are used to relate minimum/maximum concentration with the detector response. Equations 5.5 and 5.6 are used to calculate  $C_{\text{Min}}$  and  $C_{\text{Max}}$ . In these equations,  $\epsilon_{\text{Min}}$  and  $\epsilon_{\text{Max}}$  are the minimum and maximum molar absorptivities determined by the UV-VIS rules, assuming that the optimum wavelength is used for the fluorescent absorption wavelength.  $Y_{\text{Min}}$  is the minimum quantum yield.  $C_{\text{Min}}$  is multiplied by 10 to insure that good precision is obtained (see reference 1, page 10).

$$C_{\text{Min}} = \left( \frac{2 \times 10^{-8}}{\epsilon_{\text{Min}} \times Y_{\text{Min}}} \right) \times 10 \quad \text{Equation 5.5}$$

$$C_{\text{Max}} = \frac{0.05}{\epsilon_{\text{Max}}} \quad \text{Equation 5.6}$$



**Figure 5.4**  
**Flow Chart to Determine if a Fluorescent Detector**  
**is an Appropriate Choice**

## Section 5.5 Electrochemical Detectors

### ***Section 5.5.1 Overview and Determination if Species are Electrochemically Active***

Figure 5.5 shows the logic tree used to assign electrochemical detectors. Much of the logic is similar to the detectors described previously. HPLC Advisor uses a database, ECHEM.DBF, which contains reduction potentials of functional groups. This database also indicates if the functional group undergoes reduction or oxidation. Most HPLC electrochemical detectors are used with reversed-phase solvents.<sup>3</sup> The expert system does not exclude electrochemical detection if reversed-phase solvents are not used, but it does provide additional advice to inform the user of the necessity to add a supporting electrolyte to the mobile phase.

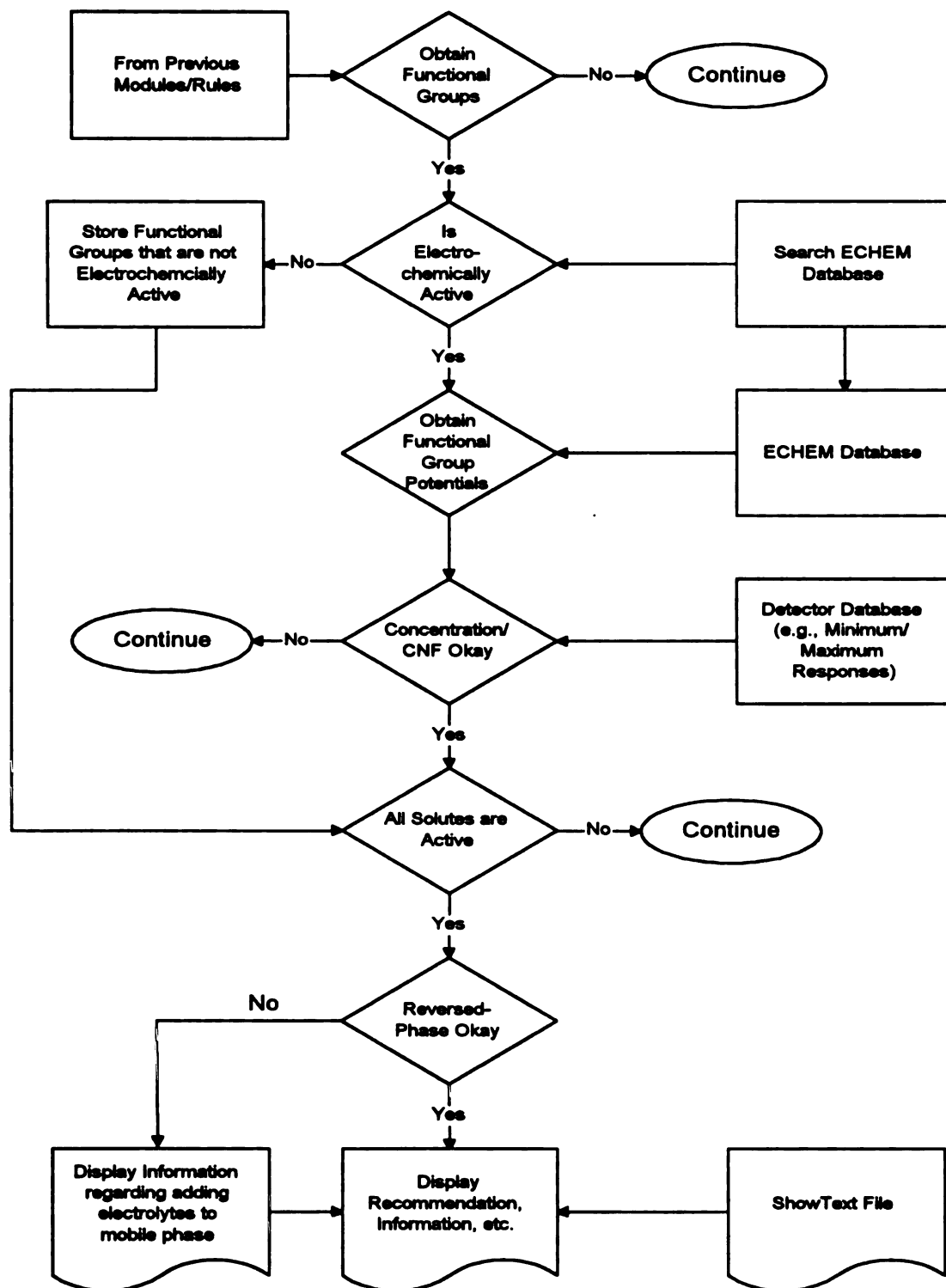
### ***Section 5.5.2 Determination of Concentration Range***

The expert system uses Faraday's Law (Equation 5.7) to develop a relationship between concentration and response. Rearrangement and incorporation of flow rates and instrumental efficiencies give Equation 5.8. To simplify this equation, a flow rate of 5.0 mL/min and a two electron transfer are assumed. This should provide a reasonable estimate of  $C_{Min}$ . The maximum concentration is obtained from the electrochemical database. The efficiency term (5%) is the reduction/oxidation conversion efficiency of most electrochemical detectors.<sup>3</sup>

$$i = nF \left( \frac{dn}{dt} \right) \quad \text{Equation 5.7}$$

$$C_{Min} = \frac{i_{Min}}{96500 \times \text{Flow} \times n} \times \frac{1}{(\text{efficiency})} \quad \text{Equation 5.8}$$





**Figure 5.5**  
**Flow Chart to Determine if an Electrochemical Detector**  
**is an Appropriate Choice**

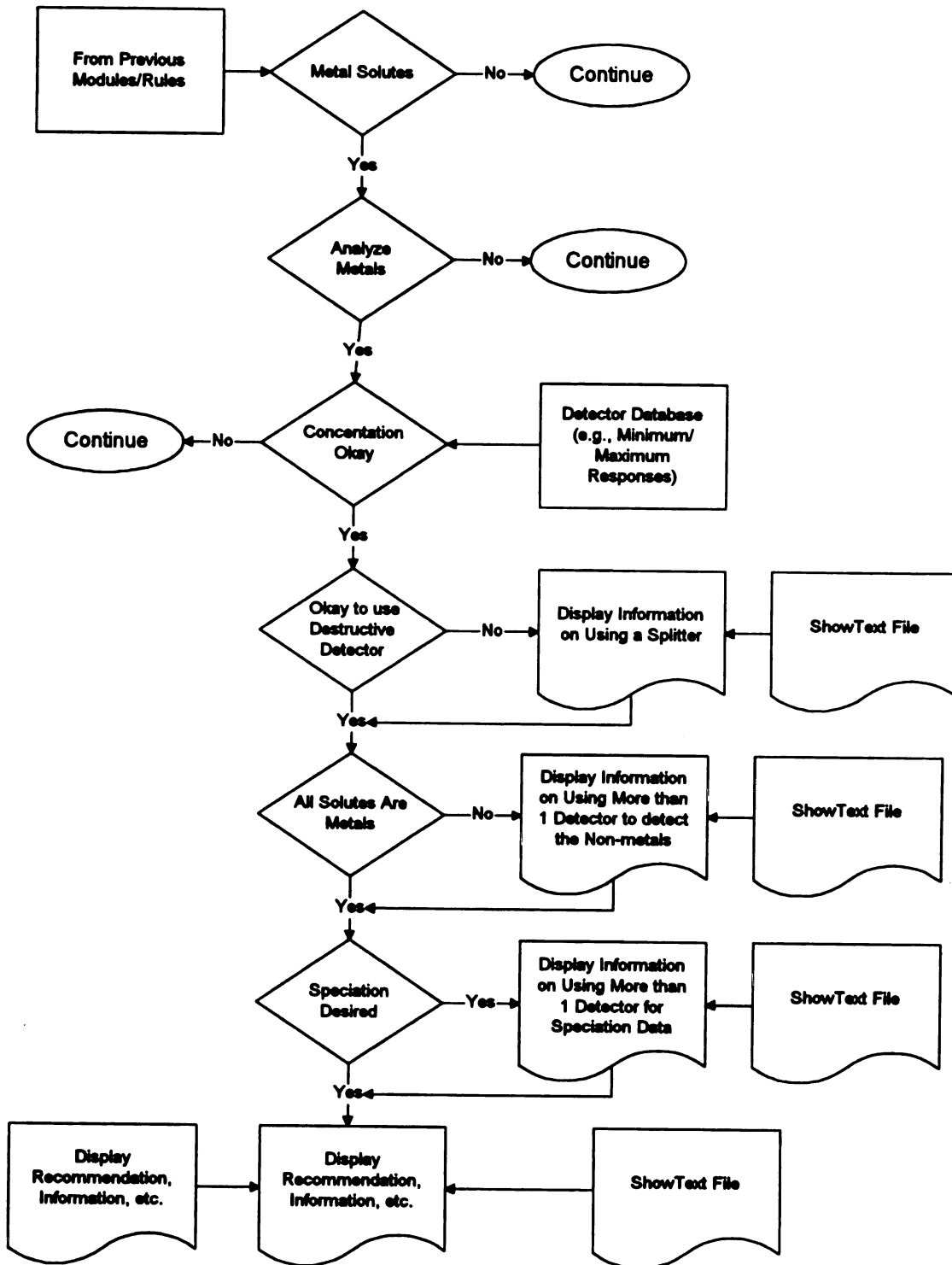
## **Section 5.6 Atomic Spectroscopy Detectors**

### ***Section 5.6.1 Overview***

Atomic spectroscopy was chosen to provide a detector for metal-containing solutes. If the sample contains metals, the expert system asks the user questions to determine if a second detector is required. A second detector is necessary if there are other non-metal solutes that the user wants to detect. If the user has indicated that the sample needs to be recovered, the expert system provides an explanation of the use of splitters when using a destructive detector such as an atomic spectroscopy detector.

### ***Section 5.6.2 Determination of Concentration Range***

The detection limits for metals by atomic spectroscopic techniques depend upon the metal to be determined and the type (emission, absorption, or fluorescence) of atomic technique that is used. Additionally, the response function for these techniques depends on the sample matrix, nebulizer efficiencies, and so forth. It was decided that rather than using a response function to determine  $C_{\text{Min}}$  and  $C_{\text{Max}}$ , we would use a typical minimum/maximum concentration range for atomic spectroscopic detectors. These values are obtained from the DETECTOR.DBF database file.



**Figure 5.6**  
**Flow Chart to Determine if an Atomic Spectroscopy**  
**Detector is an Appropriate Choice**

## Section 5.7 Refractive Index Detectors

Refractive Index (RI) detectors provide almost universal detection, but with low sensitivity. The expert system first obtains concentration ranges from DETECTOR.DBF and uses these concentration values to determine if the concentrations of the solutes are within the operating range of the RI detector. Due to the lack of sensitivity of RI detectors, the confidence factor for RI detectors will be increased if the expert system was unable to propose other, more sensitive, detectors. The logic path that the expert system uses in assigning RI detection is shown in Figure 5.7.

## Section 5.8 Conductivity Detectors

### *Section 5.8.1 Overview*

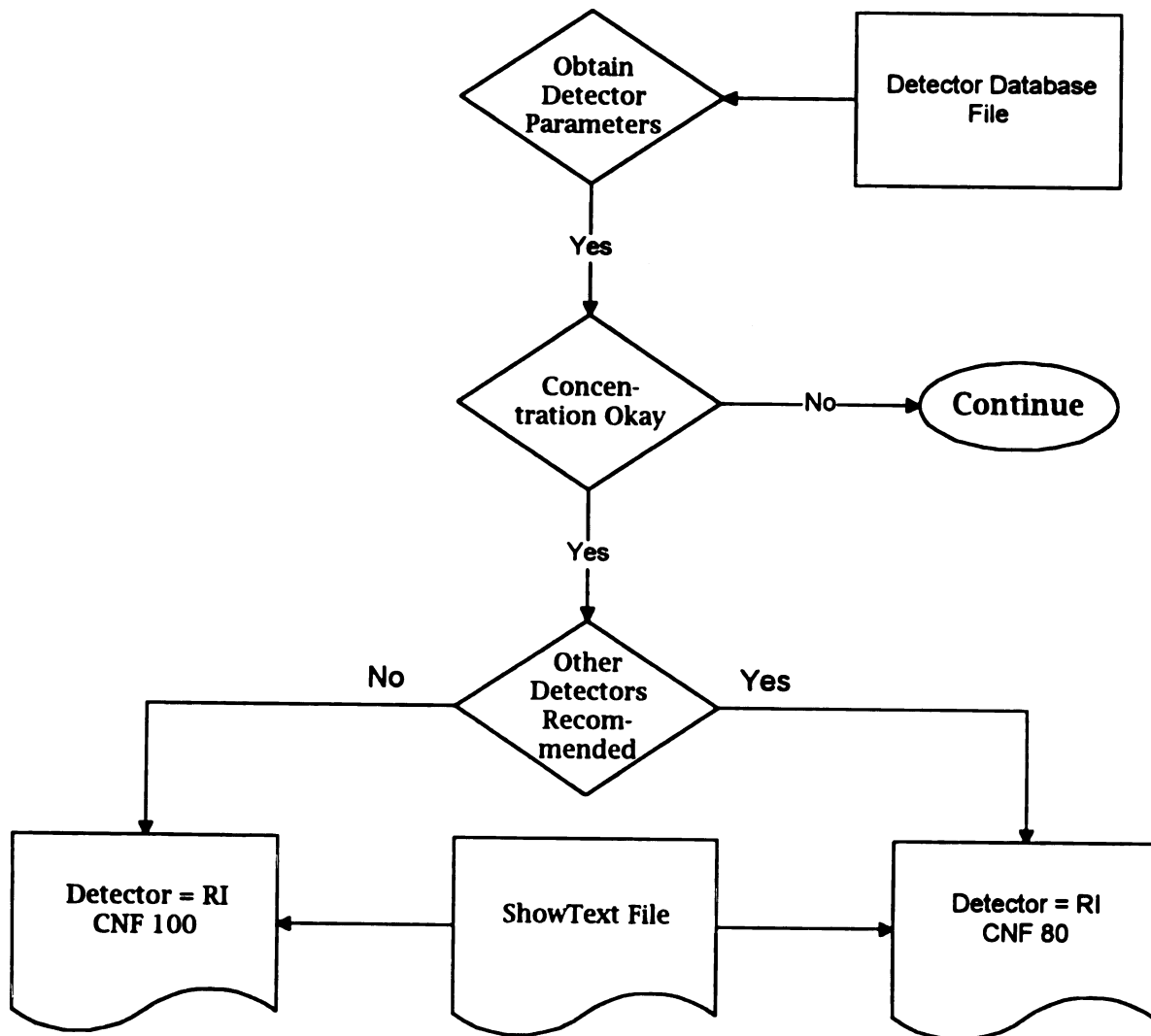
Conductivity detectors will be considered if there are ionic solutes. A database, CONDUCT.DBF, contains different ionic compounds and the respective equivalent ionic conductances,  $\lambda$ . The  $\lambda$  values are used to calculate  $C_{\text{Min}}$ , as described below. The logic required to assign conductivity detectors is shown in Figure 5.8.

### *Section 5.8.2 Determination of Concentration Range*

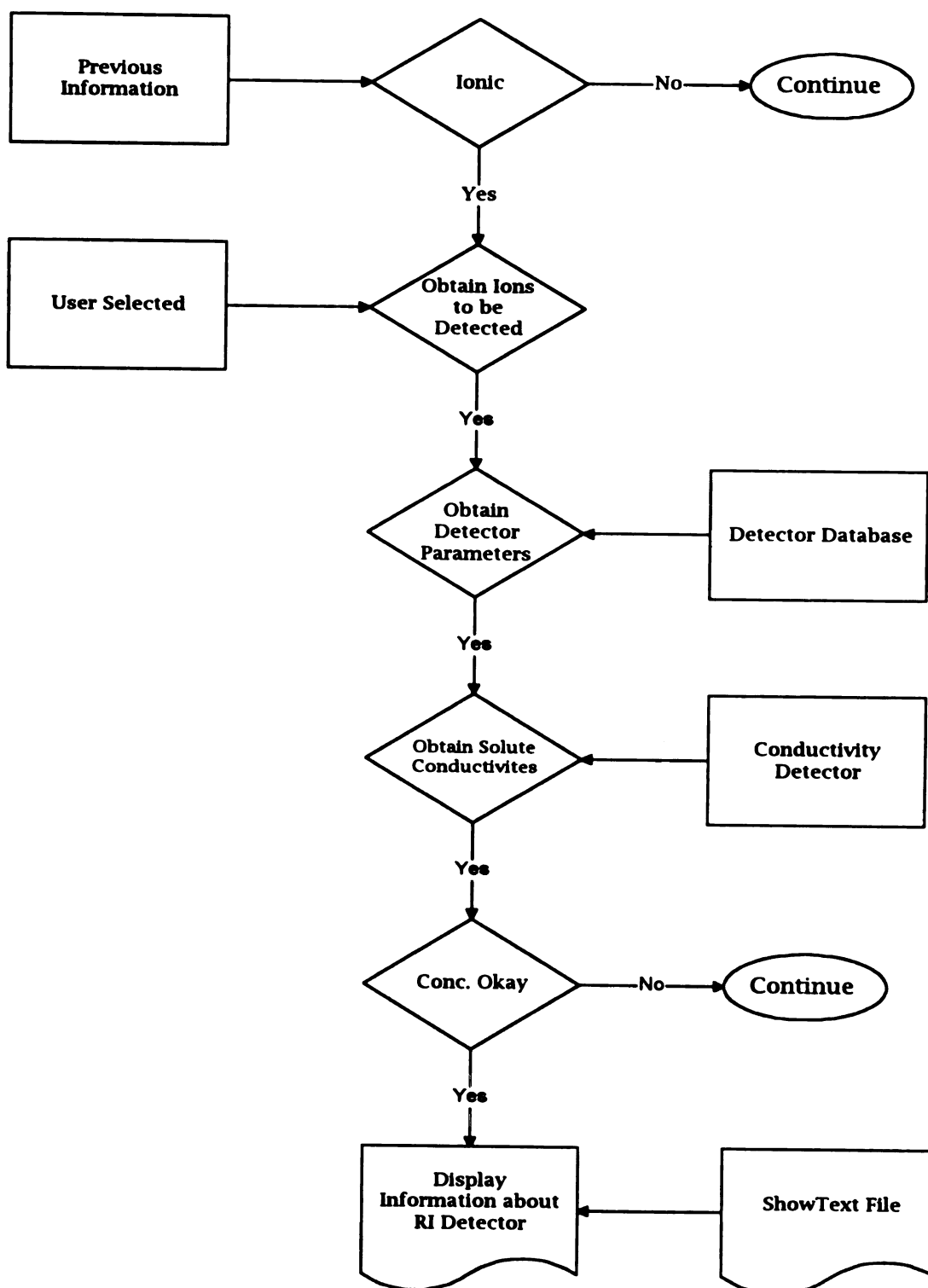
The expert system queries the user to determine what ionic species are present. If the ionic species are not present in the conductivity detector, the expert system uses average of the  $\lambda$  values (anionic or cationic). The response function that relates concentration and detector response is given by Equation 5.9.<sup>4</sup> In this equation  $\Delta G$  is the change in conductance and is obtained by using the value of the minimum response field from the DETECTOR.DBF database.  $K$  is the cell constant. A value of 33 for  $K$  (a typical value for commercial detectors)<sup>4</sup> is used for this expert system.  $\lambda_s$  is the equivalent conductance of the sample ion and  $\lambda_e$  is the equivalent conductance for the

eluant counter-ion. The expert system uses the maximum  $\lambda$  for the  $\lambda_s$  value in Equation 5.9. For simplification purposes,  $\text{OH}^-$  is assumed to be the eluant counter-ion.

$$C_{\text{Min}} = \frac{(\Delta G)1000K}{(\lambda_s - \lambda_e)} \quad \text{Equation 5.9}$$



**Figure 5.7**  
**Flow Chart to Determine if a Refractive Index**  
**Detector is an Appropriate Choice**

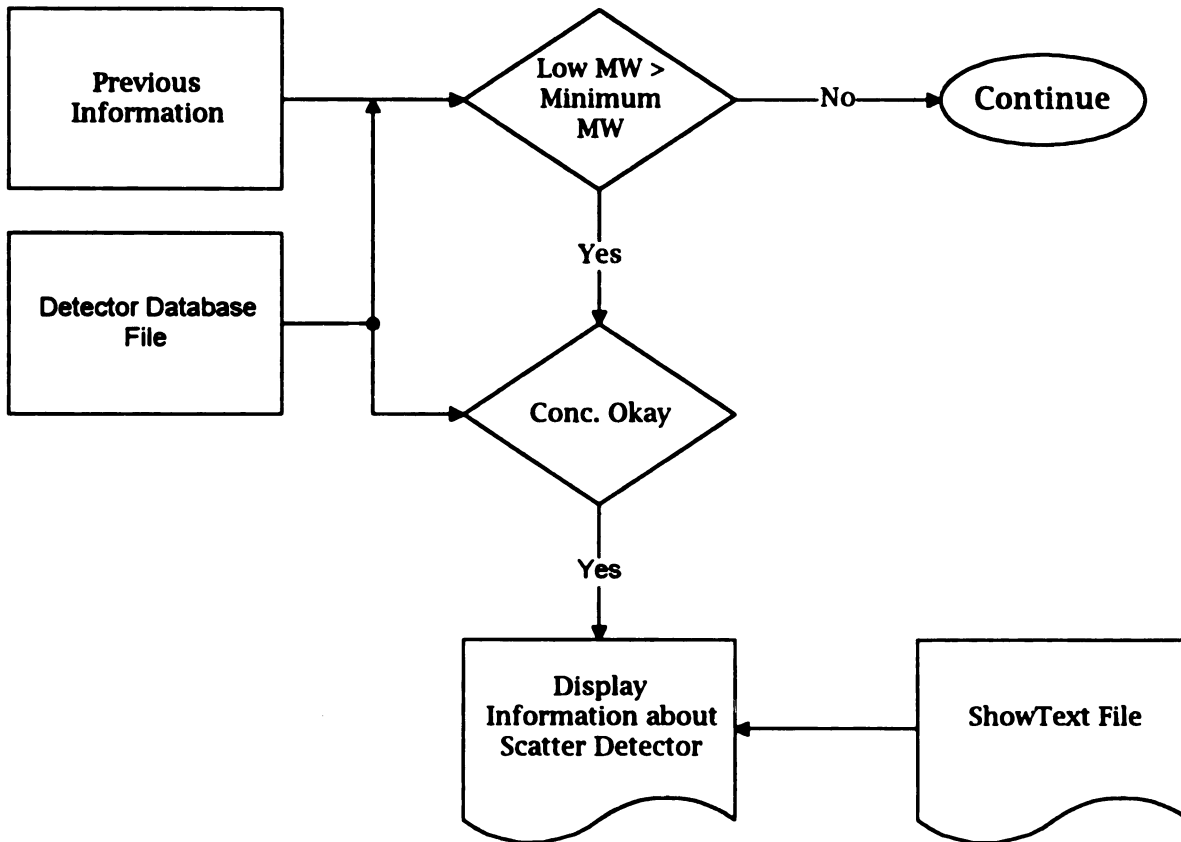


**Figure 5.8**  
**Flow Chart to Determine if a Conductivity**  
**Detector is an Appropriate Choice**

## Section 5.9     Light Scattering Detector

Light scattering detectors require that the solutes have a sufficient physical size to scatter significantly the incident radiation. While the response function depends upon the size of the solutes, molecular weight is a good size indicator. Scattering response functions are quite complex and incorporate parameters that, among others, describe the arrangement (e.g., coiled or linear) of the solutes.<sup>5</sup> The expert system first determines a minimum molecular weight based on  $R = KCM$ , where  $K$  is a complex parameter that incorporates such parameters as shape,  $R$  is the detector response,  $C$  is the concentration, and  $M$  is the molecular weight. This function is used to calculate a minimum molecular weight required for efficient scattering. If the smallest molecular weight (determined in the column module) is larger than the minimum value, the expert system compares the solute concentrations to the minimum/maximum concentrations that are obtained from the DETECTOR.DBF file. Figure 5.9 shows the logic used to assign scattering detectors.





**Figure 5.9**  
**Flow Chart to Determine if a Light Scattering**  
**Detector is an Appropriate Choice**

## **PART 2: RECOMMENDATION OF MOBILE PHASE**

The more efficient computers become at inducing new knowledge, the more widely that knowledge will be applied, even in matters of life and death. It is essential that such knowledge be open to inspection. This means that designers of learning systems have a public duty to use comprehensible description languages — even if that means sacrificing performance. Otherwise we run the risk of generating truly “unknowable knowledge.”

**Richard Forsyth**  
*Machine Learning for Expert Systems*

## **Section 5.10 Overview of Mobile Phase Selection**

To design a successful HPLC separation, a correct mobile phase must be matched to the choice of column and sample.<sup>6</sup> It was decided that HPLC Advisor should recommend a suitable mobile phase based upon the choice of column and solute(s) properties. HPLC Advisor will recommend only initial mobile phases and the user can apply additional techniques for solvent optimization.<sup>7,8</sup>

### ***Section 5.10.1 Goals of the Mobile Phase Module***

The principal goal of HPLC Advisor is to recommend both the main solvent and possible mobile phase modifiers. Possible chemical interactions between the solute(s) and mobile phase, such as dipolar and proton-donating characteristics, are used to determine the solvent modifiers recommended by HPLC Advisor.

A secondary goal is to minimize the number of questions asked of the user. Most of the relevant information is obtained in the column and detector modules. The current version of the mobile phase module contains 34 mobile phase rules and does not require any additional input from the user. The rules used for recommending the solvent(s) are described below.

After a mobile phase has been recommended, HPLC Advisor determines if the initially-recommended solvents are compatible with the recommended detector(s). If a solvent is incompatible, HPLC Advisor recommends an alternative solvent that is compatible with the detector.

### ***Section 5.10.2 Databases Used by the Mobile Phase Module***

HPLC Advisor maintains relevant mobile phase information in several databases. This allows the information to be stored in an efficient format separate from the

knowledge base. Table 5.4 provides a listing of the different databases and the fields/information contained in the databases. (The exact field names are not given; rather, an expanded form of the fields is given in the table.)

The principal database used by HPLC Advisor is the mobile phase database, MP\_DBF.DBF. The different solvents are listed with their UV-Visible cutoff wavelengths. The SOLUBLE field indicates if the solvent is miscible with water, hydrocarbons (hexane), or both. The FLUORESCENT field indicates if the solvent is fluorescent. The SOLVENT GROUP field is Snyder and Kirkland's solvent group number.<sup>6</sup> The REFRACTIVE INDEX field is provided for future use.

The SEC\_MP database contains solvents typically used for Size Exclusion Chromatography (SEC). All of the fields in SEC\_MP provide the same functionality as MP\_DBF.

The last three databases are used for normal-phase adsorption columns. Adsorption energies of different functional groups (FUNCTIONAL GROUP) on both Silica and Alumina is contained in the database, ABS\_HV. This database is used by the column module of HPLC Advisor. The two databases, SILICA and ALUMINA, contain adsorption energies of different solvents on the respective absorbent. The SOLUBLE fields contain miscibility information.

### ***Section 5.10.3      Detector Compatibility Rules***

After the initial mobile phase is recommended, HPLC Advisor determines if these mobile phases are compatible with the different detectors recommended. Also, HPLC Advisor determines if the solvents are compatible with only UV-Visible and fluorescence detectors. The databases do not contain mobile phases that undergo electrochemical reactions; therefore, the expert system does not perform a compatibility

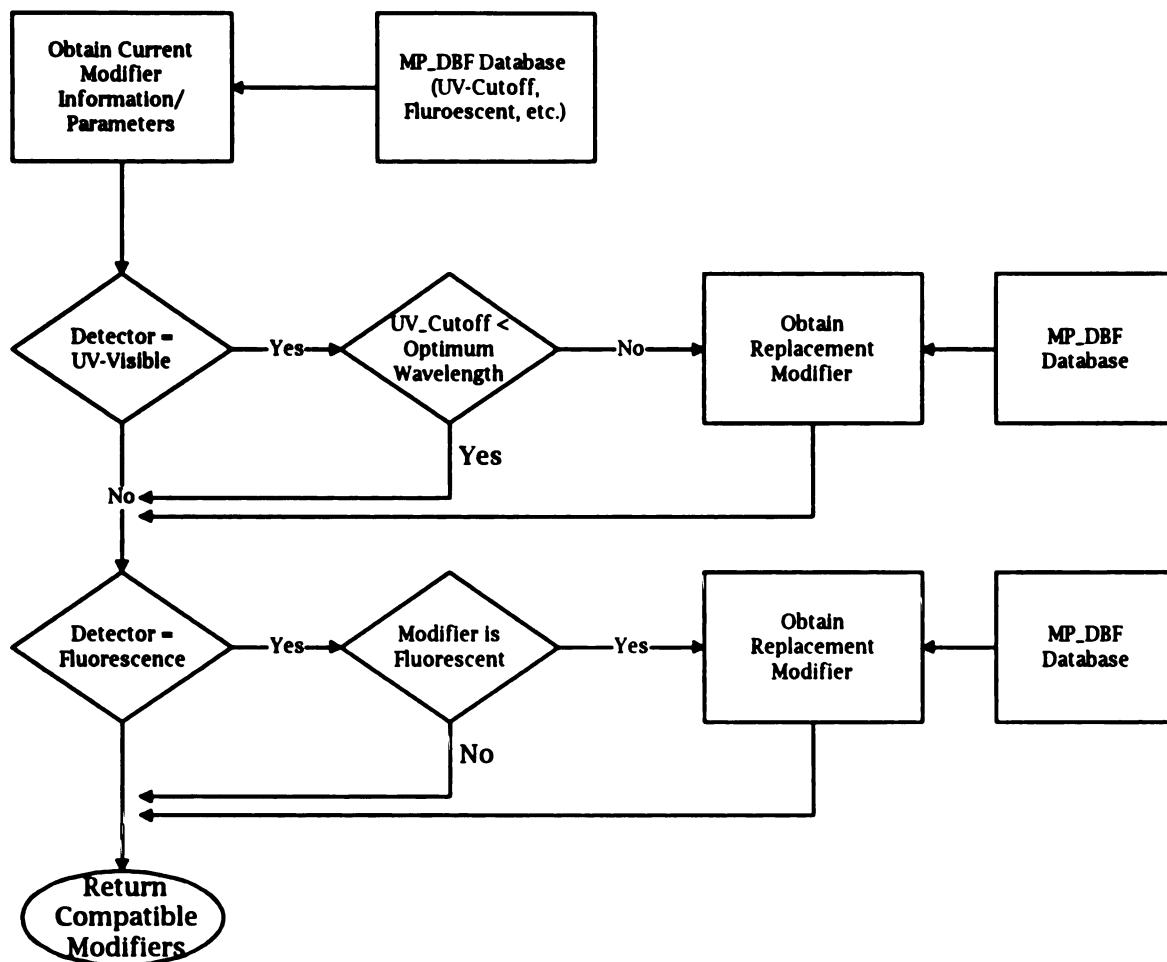
check with electrochemical detectors. The logic used in determining compatibility is presented below and shown pictorially in Figure 5.10.

**Table 5.4**  
**Structures of the Mobile Phase Databases used**  
**by HPLC Advisor**

| File    | Field 1           | Field 2   | Field 3 | Field 4     | Field 5          | Field 6       |
|---------|-------------------|-----------|---------|-------------|------------------|---------------|
| MP_DBF  | Solvent           | UV Cutoff | Soluble | Fluorescent | Refractive Index | Solvent Group |
| SEC_MP  | Solvent           | UV Cutoff | Soluble | Fluorescent | Refractive Index |               |
| Silica  | Adsorption Energy | Soluble   |         |             |                  |               |
| Alumina | Adsorption Energy | Soluble   |         |             |                  |               |
| Abs_HV  | Functional Group  | Alumina   | Silica  |             |                  |               |

Refractive index (RI) and light scattering detectors measure differences in transmission of light through the solvent. The main requirement of RI and light scatter detectors is that the solvents do not absorb a significant portion of the incident radiation. Any adsorption is adjusted such that the only the change in light intensity is due to refraction or scattering of the incident radiation. The solvents in MP\_DBF are compatible with this criterion.

Next, HPLC Advisor obtains the relevant solvent properties from the MP\_DBF database. For each recommended detector, HPLC Advisor invokes a specific rule that determines if the mobile phase(s) is compatible with the recommended detector.



**Figure 5.10**  
**Flow Chart to Determine if Mobile Phase Modifiers are**  
**Compatible with Different Detectors**

If a UV-Visible detector is recommended, HPLC Advisor compares the cutoff wavelength,  $\lambda_{\text{cutoff}}$ , for the modifier with the optimum wavelength,  $\lambda_{\text{opt}}$  that the detector module recommended. If  $\lambda_{\text{cutoff}}$  is greater than  $(\lambda_{\text{opt}} - 15)$  nm, HPLC Advisor obtains an alternative modifier from MP\_DBF that is from the same solvent group but has a  $\lambda_{\text{cutoff}}$  that is at least 15 nm less than  $\lambda_{\text{opt}}$ . Either the initial recommended solvent or the alternative is presented as a viable mobile phase.

If a fluorescence detector is recommended, HPLC Advisor applies similar rules to obtain a mobile phase that is not fluorescent. Each modifier is checked for compatibility with the recommended detectors. HPLC Advisor reports the recommended mobile phase to the user.

## **Section 5.11 Selection of Mobile Phases for the Different Separation Modes**

### ***Section 5.11.1 Overview***

As discussed above, HPLC Advisor recommends mobile phases based upon the particular separation mode (e.g., Normal-Phase Adsorption) and the properties of the solutes with the stationary phase (e.g., dipolar interactions). Recommendations of solvents for different columns and separation modes is based upon information from Snyder and Kirkland.<sup>6</sup> The following sections describe the rules used to recommend particular solvents. Abbreviated flow charts are presented to clarify the decision processes.

From zero to three different mobile phase modifiers can be recommended by the expert system. These recommendations are based upon the presence of dipolar, proton donating, or proton accepting characteristics of the solutes. HPLC Advisor reports all possible modifiers and indicates that the user can use a combination of the main

solvent and zero to all of the recommended modifiers. The user must perform experiments and use solvent optimization routines, such as those described in Chapter 2, to determine which of the modifiers provides the desired separation.

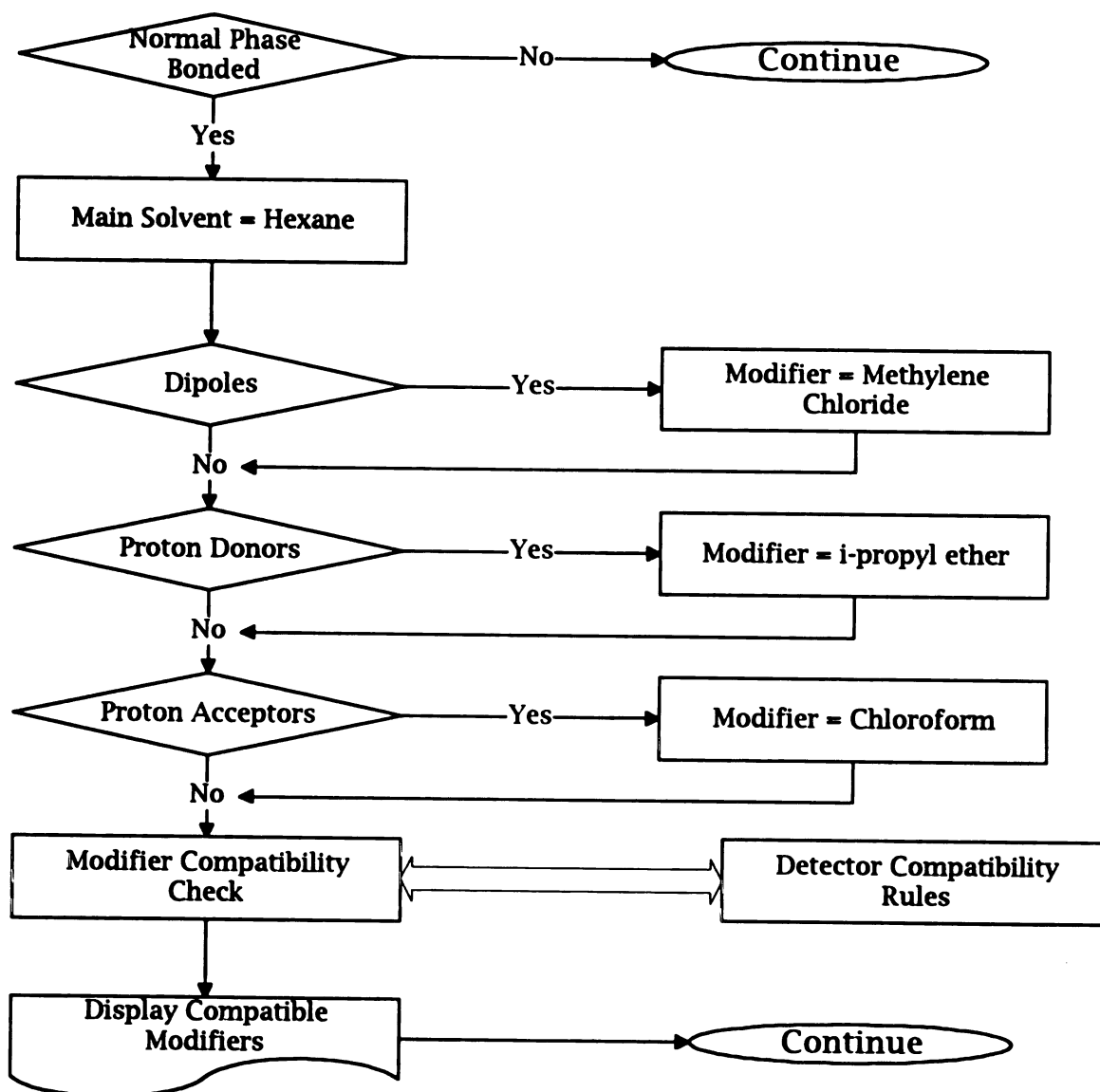
### ***Section 5.11.2      Mobile Phase Recommendations for Normal-Phase Bonded Chromatography***

For normal-phase bonded chromatography, HPLC Advisor recommends a hydrocarbon, hexane in particular, as the main solvent. Depending upon the characteristics of the solutes present, the expert system recommends different polar solvent modifiers. For example, proton-accepting solvents are recommended for solutes that are proton-donors. Table 5.5 lists the different solvent modifiers and solvent group numbers that are recommended. Figure 5.11 shows the logic flow chart for normal-phase bonded chromatography.

**Table 5.5**  
**Mobile Phase Modifiers for Normal-Phase**  
**Bonded Chromatography**

| <b>Solute-Stationary Phase Interactions</b> | <b>Modifier</b>           | <b>Solvent Group</b> |
|---------------------------------------------|---------------------------|----------------------|
| <b>Dipole</b>                               | <b>Methylene Chloride</b> | <b>5</b>             |
| <b>Proton Donor</b>                         | <b>i-Propyl Ether</b>     | <b>1</b>             |
| <b>Proton Acceptor</b>                      | <b>Chloroform</b>         | <b>8</b>             |





**Figure 5.11**  
**Flow Chart for the Determination of Mobile Phase Modifiers for**  
**Normal-Phase Bonded Chromatography**

### ***Section 5.11.3      Mobile Phase Recommendations for Reverse-Phase Bonded Chromatography***

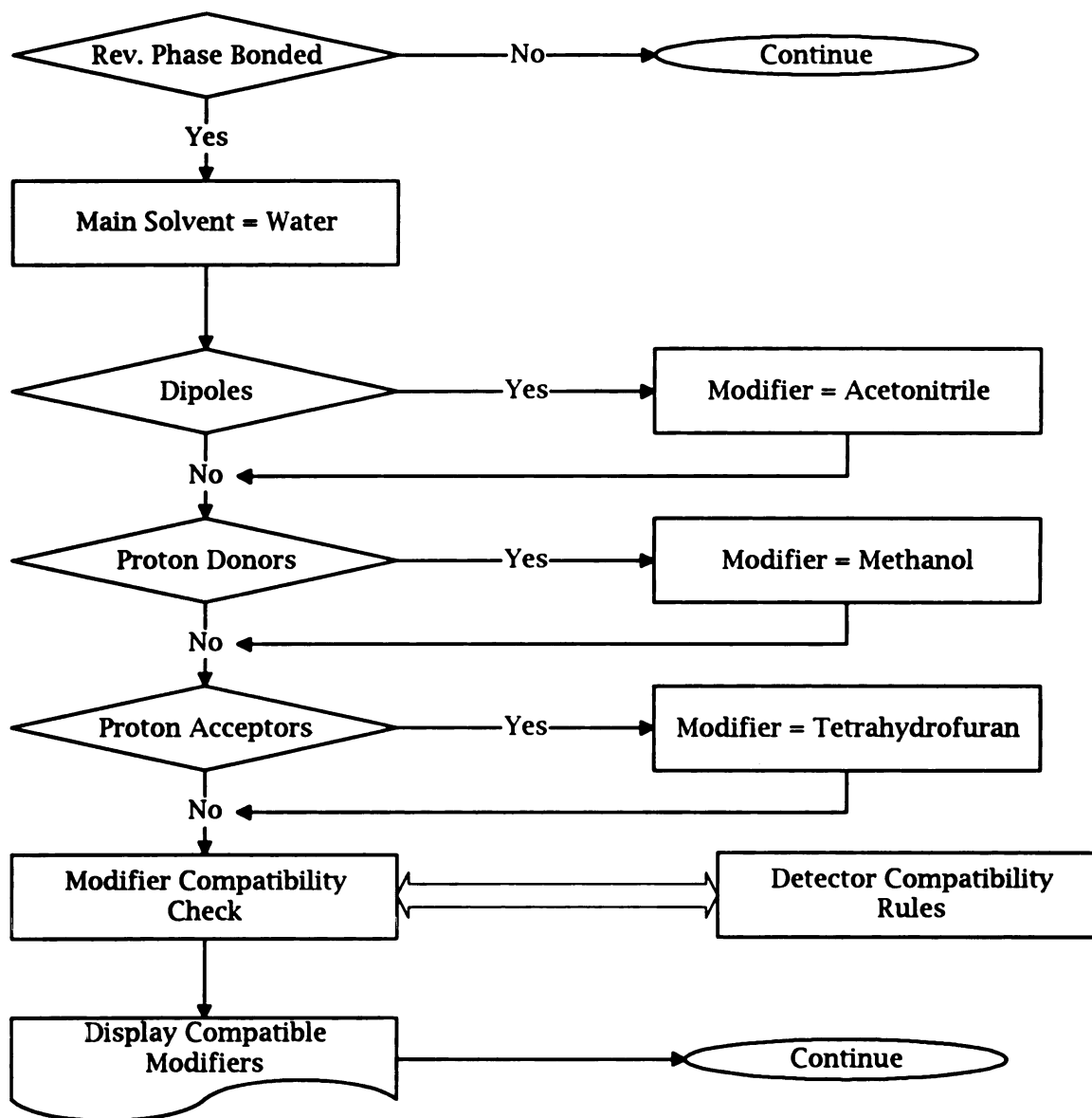
For reverse-phase bonded chromatography, HPLC Advisor recommends water as the main solvent. Depending upon the types of solutes present, the expert system recommends different miscible organic modifiers. Table 5.6 lists the possible recommended solvent modifiers. Figure 5.12 shows the logic flow chart for reverse-phase bonded chromatography.

**Table 5.6**  
**Mobile Phase Modifiers for Reverse-Phase Bonded Chromatography**

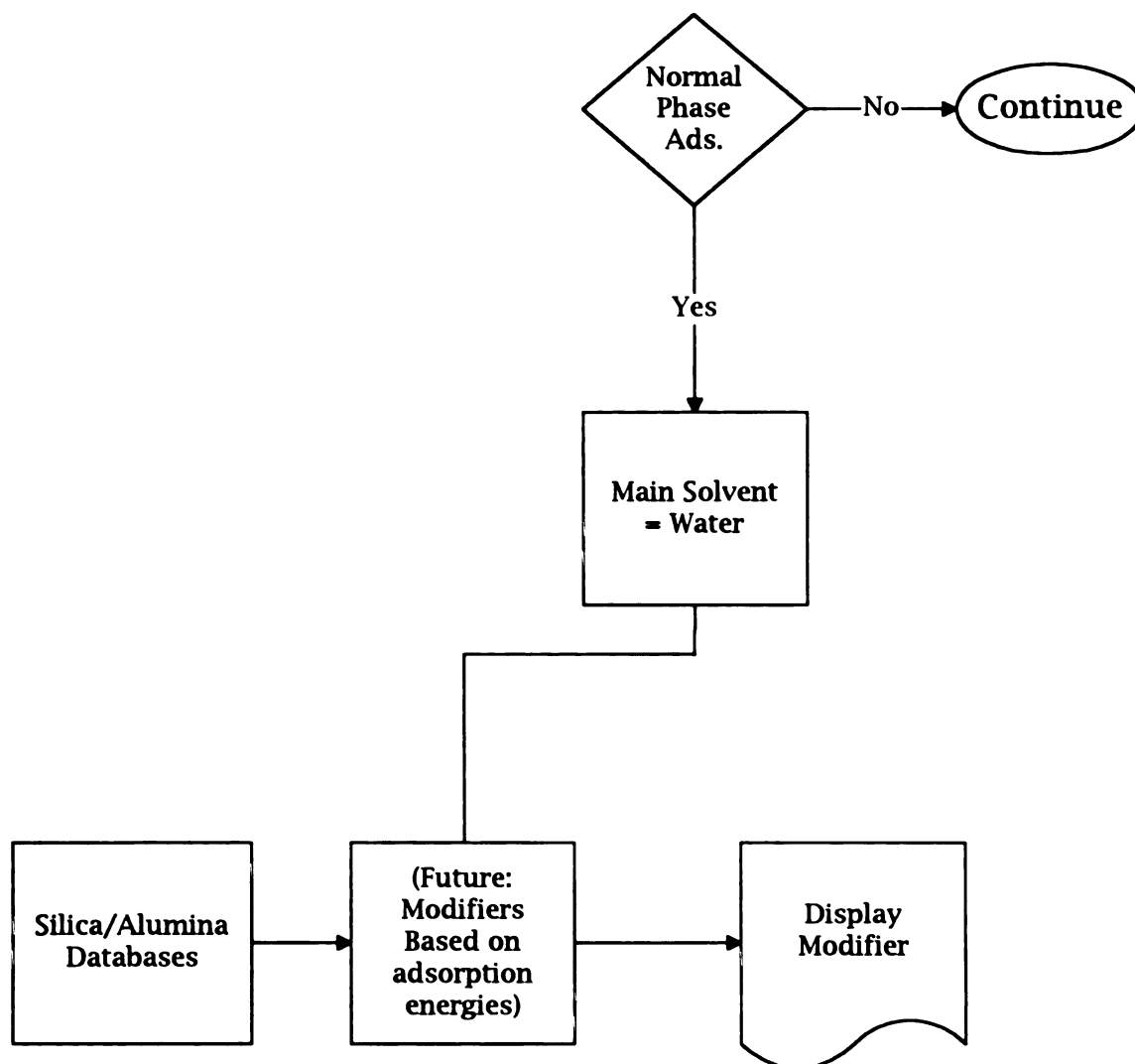
| Solute-Stationary Phase Interactions | Modifier        | Solvent Group |
|--------------------------------------|-----------------|---------------|
| Dipole                               | Acetonitrile    | 6             |
| Proton Donor                         | Tetrahydrofuran | 3             |
| Proton Acceptor                      | Methanol        | 2             |

### ***Section 5.11.4      Mobile Phase Recommendations for Normal-Phase Adsorption Chromatography***

For normal-phase adsorption chromatography, HPLC Advisor recommends hexane as the main solvent. HPLC Advisor informs the user to select a modifier which dissolves the sample and has an adsorption energy for the stationary phase (either silica or alumina) that is slightly less than the largest adsorption energy of your solute(s). This version of HPLC Advisor does not recommend the modifier because the column selection module uses heuristics to recommend silica versus alumina. Figure 5.13 shows the logic flow chart for normal-phase adsorption chromatography.



**Figure 5.12**  
**Flow Chart for the Determination of Mobile Phase Modifiers for**  
**Reverse-Phase Bonded Chromatography**



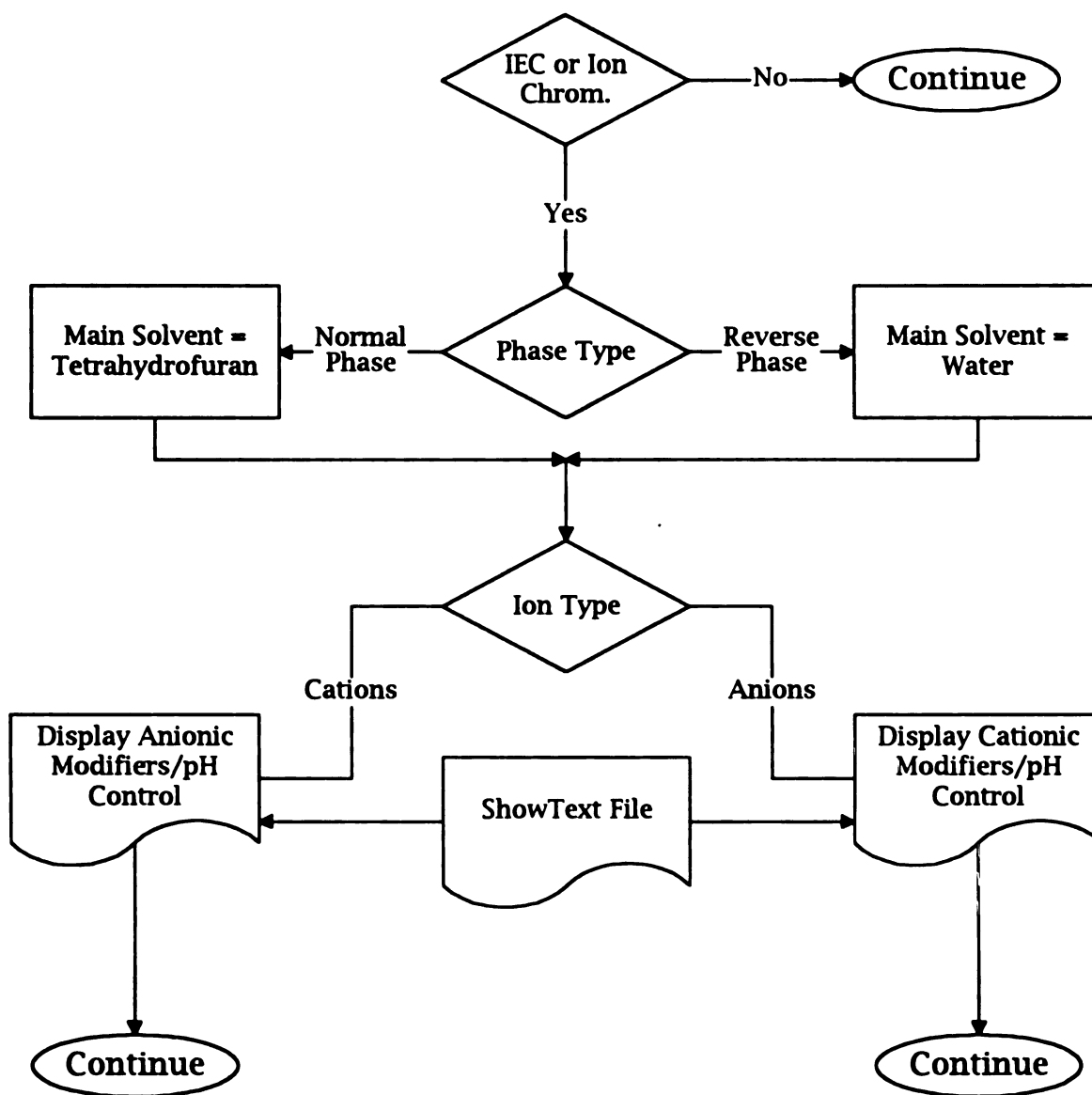
**Figure 5.13**  
**Flow Chart for the Determination of Mobile Phase Modifiers for**  
**Normal-Phase Adsorption Chromatography**

### **Section 5.11.5      *Mobile Phase Recommendations for Ion Exchange and Ion Chromatography***

For ion exchange (IEC) and ion chromatography (IC), HPLC Advisor recommends water and tetrahydrofuran (THF) as the main solvent for reverse-phase and normal-phase separations respectively. These two solvents dissolve the ionic solutes. Figure 5.14 shows the logic flow chart for these two separation modes. The expert system then recommends appropriate cationic or anionic counter-ions to add to the mobile phase. The separation mechanism for IEC and IC is ion exchange between the ionic solutes and the solvent ions with the ionic functional groups attached to the stationary phase. The equilibrium for anion and cation exchange are depicted in Equations 5.10 and 5.11, where X is the solute ion, R is the stationary phase ionic functional group, and Y is the solvent ion.



HPLC Advisor lists, from weakest to strongest solvating strength, the different choices of the solvent counter-ion, Y. Weaker solvent counter-ions give longer retention times. The user is informed to use initially the weakest mobile phase ion. If the desired separation is not achieved, the user is recommended to use the next stronger solvent counter-ion. Typical mobile phase ion concentrations are 0.001 to 0.5 Molar. Table 5.7 lists the different solvent ions that are recommended.



**Figure 5.14**  
**Flow Chart for the Determination of Mobile Phase**  
**Modifiers for Ion Exchange and Ion Chromatography**

**Table 5.7**  
**Recommended Mobile Phase Ions for**  
**Ion Exchange and Ion Chromatography**

| Mobile Phase Strength | Mobile Phase Anions | Mobile Phase Cations |
|-----------------------|---------------------|----------------------|
| Weakest               | Hydroxyl            | Lithium              |
|                       | Acetate             | Hydrogen             |
|                       | Formate             | Sodium               |
|                       | Chloride            | Ammonium             |
|                       | Thiocynate          | Potassium            |
|                       | Nitrate             | Magnesium            |
|                       | Sulfate             | Zinc                 |
|                       |                     | Copper               |
|                       |                     | Calcium              |
| Strongest             |                     | Barium               |

HPLC Advisor informs the user of the relationship between sample retention and pH. As pH changes, the equilibria in Equations 5.10 and 5.11 are shifted. For anion separations, the retention increases as pH increases since the basic solution will cause ionization of the solute(s). For cation separations, the retention increases as pH decreases.

#### ***Section 5.11.6      Mobile Phase Recommendations for Size Exclusion Chromatography***

Mobile phases in size exclusion chromatography (SEC) are not adjusted; rather, an appropriate mobile phase dissolves all of the solutes. SEC solvents have low viscosities at the elevated temperatures often used for the separation. In addition, very

polar solvents are not recommended for SEC columns composed of polystyrene packing materials.<sup>6</sup> Figure 5.15 shows the flow chart of the logic used to recommend a SEC mobile phase. HPLC Advisor uses a database that contains mobile phases typically compatible with SEC columns.<sup>6</sup> This database contains both water and hexane miscible solvents. HPLC Advisor chooses an appropriate solvent based on solute solubility and detector compatibility.

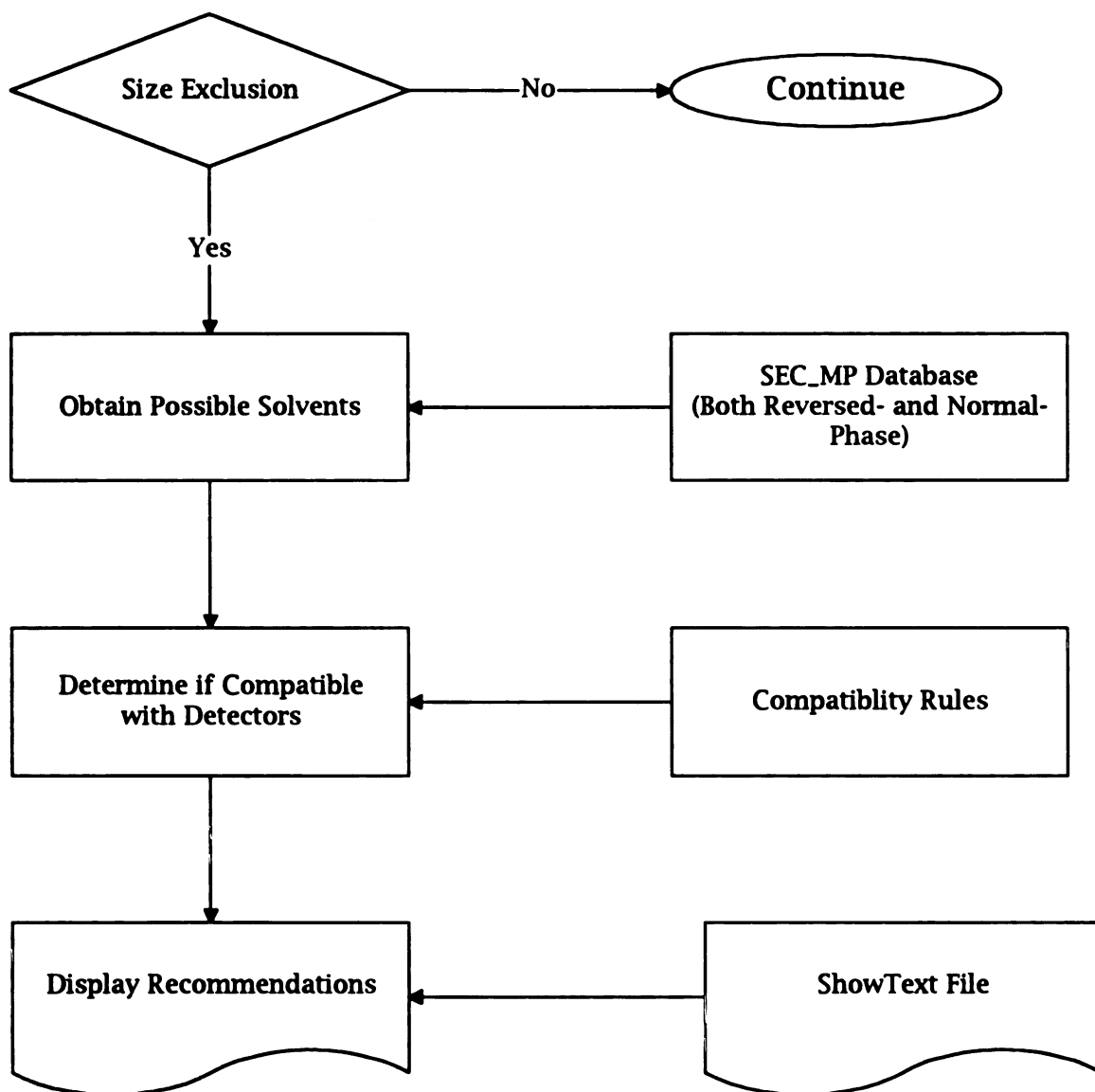
The expert system also suggests an appropriate ionic strength and mobile phase pH. Salts (soluble in the mobile phase) are added to the mobile phase to minimize adsorption effects that can occur on silica-based SEC columns. Normally, the ionic strength is between 0.05 and 0.1 Molar. Since the exact effect of mobile phase ionic strength on the retention times is not predictable, HPLC Advisor recommends that the user experimentally determine the optimum ionic strength. A pH of 2-8 is typically used to minimize degradation of the stationary phase material.<sup>6</sup>

#### ***Section 5.11.7      Mobile Phase Recommendations for Ion-Pair Chromatography***

Mobile phases for Ion-Pair Chromatography (IPC) contain a main solvent to which a counter-ion of opposite charge to the solutes is added. The counter-ion forms an ion pair with the ionic solutes. For reverse-phase chromatography, the ion pair has an increased affinity for the stationary phase, while the ion pair has an increased affinity for the mobile phase in normal-phase operation.

The first step is the recommendation of the appropriate main solvent. HPLC Advisor recommends pentanol and water for normal- and reverse-phase operations, respectively. These two solvents provide suitable solubility of the ion-pairs and are compatible with the respective column packing materials. Other alcohols have been used for Normal-Phase IPC.<sup>6</sup>





**Figure 5.15**  
**Flow Chart for the Determination of Mobile**  
**Phase Modifiers for Size Exclusion Chromatography**

Organic modifiers can also be added to the mobile phase to control the separation. HPLC Advisor uses the same rules for normal-phase and reverse-phase bonded chromatography to recommend suitable organic modifiers.

HPLC Advisor also displays information that assists the user in optimizing the separation. The separation is controlled via the addition of organic modifiers, changes in the counter-ion, and pH. For the distribution of an anionic analyte ion,  $R^-$ , between the aqueous (i.e., ionic) and ion pair phases (Equation 5.12), the capacity factor,  $k'$ , is given by Equations 5.13 and 5.14.<sup>6</sup>  $E$  is an extraction constant (Equation 5.15). In these equations,  $X^+$  is the counter-ion,  $V_s$  and  $V_m$  are the stationary phase and mobile phase volumes, respectively,  $C^+$  and  $C^-$  are the cationic/anionic counter-ion concentrations, and  $n$  is the valence of the counter-ion. Based upon these equilibria, an increase in the counter-ion concentration increases the capacity factor for reverse-phase IPC and decreases the capacity factor for normal-phase IPC.



$$k' = \frac{V_s}{V_m} E(C^+)^n \quad \text{(Anionic samples) Equation 5.13}$$

$$k' = \frac{V_s}{V_m} E(C^-)^n \quad \text{(Cationic samples) Equation 5.14}$$

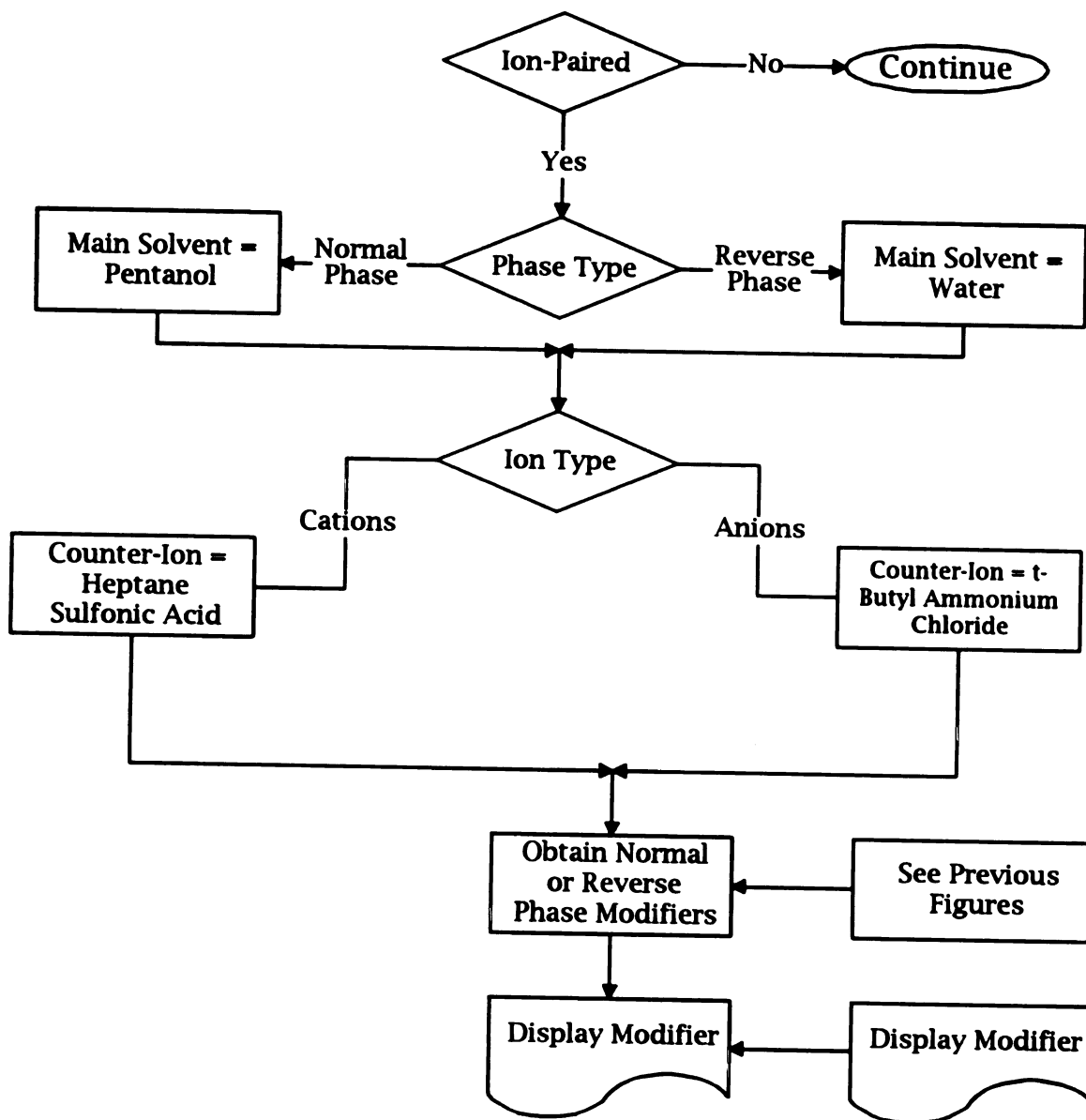
$$E = \frac{(R^- X^+)_{org}}{(R^-)_{aq}(X^+)_{aq}} \quad \text{Equation 5.15}$$

HPLC Advisor provides additional information regarding the effect of different parameters on the separation of a sample. Some of these effects are listed in Table 5.8. Figure 16 shows a pictorial version of the logic used to recommend mobile phases for IPC columns.

**Table 5.8**  
**Effects of Different Mobile Phase Parameters on IPC Separations**  
**(RP = Reverse Phase NP = Normal Phase)**

| <u>Variable</u>                         | <u>Effect on k'</u>            | <u>Effect on Solvent Selectivity</u> |
|-----------------------------------------|--------------------------------|--------------------------------------|
| <b>Counter Ion</b>                      |                                |                                      |
| Increase Concentration                  | Increase (RP)<br>Decrease (NP) | No Effect<br>No Effect               |
| Change<br>(Increase -CH <sub>2</sub> -) | Increase (RP)<br>Decrease (NP) |                                      |
| <b>Increase pH</b>                      | Decrease (RP)<br>Increase (NP) | Large effect<br>Large Effect         |
| <b>Organic Modifier</b>                 |                                |                                      |
| Increase Concentration                  | Decrease (RP)<br>Increase (NP) | Small Effect<br>Small Effect         |

HPLC Advisor suggests a counter-ion concentration range of 0.005 to 0.05 Molar. This concentration range minimizes micelle formation. For example, the critical micelle concentration, CMC, of dodecyl sulfate anion is 0.05 Molar.



**Figure 5.16**  
**Flow Chart for the Determination of Mobile Phase**  
**Modifiers for Ion-Paired Chromatography**

### **PART 3: VALIDATION STUDIES OF DETECTOR AND MOBILE PHASE**

**On two occasions I have been asked [by members of parliament!], 'Pray, Mr. Babbage, if you put into the machine wrong figures, will the right answers come out?' I am not able rightly to apprehend the kind of confusion of ideas that could provoke such a question.**

**Charles Babbage**

## **Section 5.12 Validation Studies of the Detector and Mobile Phase Modules**

Several validation studies are presented below that determine if HPLC Advisor is able to recommend an appropriate detector(s) and mobile phase. Since the detector and mobile phase modules require the information and conclusions that are obtained from the first module, the column recommendations are also presented. Since previous information in this dissertation has described the process of validation studies, these validation studies are presented briefly here.

### ***Section 5.12.1 Phthalate Validation Study***

The first validation case that is presented here is a mixture of phthlates that is Alltech's standard test mixture for normal phase columns. This test case was chosen since a representative chromatogram was included with the columns. The mixture contained three compounds: Toluene, Diethyl Phthalate, and Dimethyl Phthalate. Physical and chemical characteristics of these three compounds are given in Table 5.9. This information, in addition to information presented below, was used as the input data for the validation study of the column, detector, and mobile phase modules of HPLC Advisor.

The version of HPLC Advisor used for the preliminary validation of this sample did not have aromatic esters as a possible selection. Since the UV-Visible spectra of esters are similar to the parent carboxylic acid, the molecules were described as aromatic carboxylic acids. As described below, this input led to the recommendation of columns such as ion exchange. To minimize the significance of ionic properties, HPLC Advisor was informed that the species could be neutralized at a pH of 7.00. The species were also described as having differences in polarity and contained different substituents on the aromatic ring (with a CNF of 70 to indicate that the two phthlates

contained similar functional groups). Additional input to HPLC Advisor described other relevant information (e.g., the absence of fused aromatic rings).

**Table 5.9**  
**Physical and Chemical Properties of the Components**  
**of the Phthalate Test Sample**

| Property          | Toluene | Diethyl Phthalate | Dimethyl Phthalate |
|-------------------|---------|-------------------|--------------------|
| Molecular Weight  | 92.14   | 222.24            | 151.11             |
| Boiling Point     | 111 °C  | 298 °C            | 284 °C             |
| Solubility        |         |                   |                    |
| Water             | 0.05%   | Insoluble         | 0.43%              |
| Alcohol           | Soluble | Soluble           | Soluble            |
| Benzene           | Soluble | Soluble           | Soluble            |
| Hexane            | Soluble | Moderate          | Soluble            |
| Concentration (M) | 0.0109  | 0.0044            | 0.00662            |

Recommendations from HPLC Advisor are presented in Table 5.10. HPLC Advisor did not recommend fluorescent detectors since esters are not fluorescent. Scatter detectors also were not recommended due to the small molecular weight of the solutes. Electrochemical detectors were recommended with a confidence factor below the truth threshold. The version of HPLC Advisor used for this validation study did not provide esters as a functional group option for the user to select. As such, aromatic carboxylic acids were selected to indicate that esters are present. Since carboxylic acids may form irreversible Schiff Bases with the amine groups, HPLC Advisor did not recommend the amine column. The next version of HPLC Advisor included esters as a possible selection. Esters have similar molar absorptivities as the parent carboxylic acids;<sup>9</sup> therefore, the modifications to the knowledge base of HPLC Advisor only

required that the expert system use the molar absorptivity data of the corresponding (aliphatic or aromatic) carboxylic acid.

**Table 5.10**  
**Results of the Validation Study of the Phthalate**  
**Test Case**

| Column             | Detector/<br>$\lambda$ Recommendations                                                            | Mobile Phase             |
|--------------------|---------------------------------------------------------------------------------------------------|--------------------------|
| Normal Phase Cyano | UV-Visible CNF 100<br>175 - 205<br>205 - 235<br>265 - 295<br>295 - 325<br>$\geq 325$<br>235 - 265 | Hexane<br>i-propyl ether |
| Normal Phase Amine | Refractive Index CNF 65                                                                           | Hexane<br>i-propyl ether |

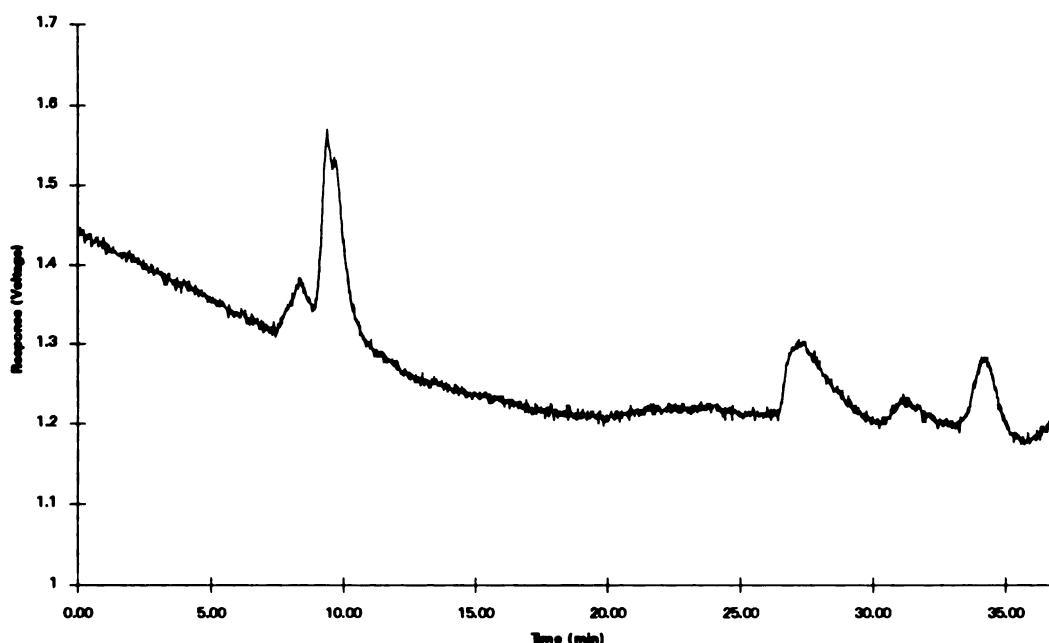
Figure 5.17 shows the resulting separation of the phthalate sample using an Alltech Econosil amine column (25 cm x 4.6 mm) operated in normal-phase mode using 1.00 mL/min hexane as the mobile phase. A UV-Visible detector at 250 nm and 0.005 absorption units full scale was used. Performance of the column was similar to the test mixtures chromatogram provided by the manufacturer.

### ***Section 5.12.2      Aromatic Validation Study***

The second validation case that is presented here is a mixture containing Benzaldehyde, Benzoic Acid, and p-Aminobenzoic Acid. Physical and chemical characteristics of these three compounds are given in Table 5.11. This information, in



addition to information presented below, was used as the input data for the validation study of the column, detector, and mobile phase modules of HPLC Advisor.



**Figure 5.17**  
**Chromatographic Separation of Phthalate Sample**

In addition to the data in Table 5.11, input for HPLC Advisor indicated that the solutes were aromatic aldehydes, amines, and carboxylic acids and that the aromatic rings contain different functional groups. A pH of ~2 is required to neutralize the carboxylic acids ( $pK_a$  of 4.189). The solutes were described as being soluble in water (CNF of 80) and hexane.

**Table 5.11**  
**Physical and Chemical Properties of the Components**  
**of the Aromatic Test Sample**

| Property          | Benzaldehyde   | p-Aminobenzoic Acid | Benzoic Acid   |
|-------------------|----------------|---------------------|----------------|
| Molecular Weight  | 106.12         | 137.14              | 122.12         |
| Boiling Point     | 178 °C         | >187 °C             | 284 °C         |
| Solubility        |                |                     |                |
| Water             | 0.29%          | 0.6%                | 2.9%           |
| Alcohol           | Soluble        | Soluble             | Soluble        |
| Benzene           | Soluble        | Slightly Soluble    | Soluble        |
| Hexane            | Soluble        | Soluble             | Soluble        |
| Concentration (M) | $\sim 10^{-4}$ | $\sim 10^{-4}$      | $\sim 10^{-4}$ |
| pK <sub>a</sub>   | -NA-           | >4.2                | 4.189          |

HPLC Advisor determined that the solutes have possible dipolar interactions with stationary phases and also has proton-donating groups. HPLC Advisor indicated that amine columns are used when there are differences in polarity of the solutes; however, HPLC Advisor indicated that Schiff Bases may form with the carboxylic acids. Additional recommendations from HPLC Advisor are indicated in Table 5.12.

**Table 5.12**  
**Results of the Aromatic Validation Test**

| Column                                                         | Mobile Phase                      | Detector                        |
|----------------------------------------------------------------|-----------------------------------|---------------------------------|
| Reverse-Phase Bonded CNF 93<br>Phenyl<br>Cyano<br>Alkyl Silane | Water<br>Methanol<br>Acetonitrile | UV-Visible CNF 100<br>RI CNF 68 |
| Normal-Phase Bonded CNF 89<br>Phenyl<br>Cyano                  | Hexane<br>i-Propyl Ether          |                                 |
| Aromatic CNF 80                                                | Column Specific                   |                                 |
| Reverse-Phase Anion Exchange CNF 72                            | Water                             |                                 |

## **PART 4: CONCLUSIONS AND SUMMARY**

**If you follow reason far enough it always leads to conclusions that are contrary to reason.**

**Samuel Butler**

### Section 5.13 Summary of Validation Studies

During the validation studies of HPLC Advisor, several limitations of VP-Expert (version 3.0) became apparent. The major problem is that VP-Expert does not support completely scientific notation. While it is possible to input scientific notation, any further calculations based on the input is converted to floating point representation. In addition VP-Expert has a precision of 1 part/million. Any value smaller than  $10^{-6}$  is displayed as zero and any value smaller than  $10^{-10}$  is truncated to zero. Since many of the concentration values (both solute concentrations and the calculated minimum, maximum, and typical concentration values used to calculate the concentration CNF) are smaller than these limits, VP-Expert rounded the numbers to zero. This prevented HPLC Advisor from determining correct concentration CNF values in that HPLC Advisor reported the confidence factor as either 100 or 50. A temporary solution has been to multiply all concentrations by some multiplier (in this case,  $10^6$ ) to convert all concentration values into micromolar.

VP-Expert also does not transfer significant figures correctly when mathematically manipulating these numbers. For example, if the value of some input, X, is equal to  $1 \times 10^{-8}$  and some interim variable, Y, has a value that is based on X (such as  $Y = X * 10$ ), then VP-Expert will report that the value of Y is zero (since X is converted to floating point first and VP-Expert truncates the value of  $1 \times 10^{-7}$  to zero).

VP-Expert also has problems with memory management. Once a knowledge base exceeds approximately 70 KB in size, many of the features of VP-Expert became unstable and lead to the system crashing. This necessitated disabling these features (e.g., REPORT feature, TRACES, and so forth). Since VP-Expert is intended to be used as a teaching tool, the lack of these features minimizes the ability of the student to see the logic used in assigning a recommendation. Of course, this exacerbated the problem of debugging the expert system.

In general, HPLC Advisor is able to recommend a logical detector and mobile phase. Quite a few problems arose when HPLC Advisor would access appropriate databases to determine detector and mobile phase parameters. Most of these limitations are based on the above described mathematical limitations of VP-Expert. Validation studies during the development of HPLC Advisor demonstrate that HPLC Advisor is able to recommend an appropriate detector and to recommend a CNF value based on the concentrations of the solutes. These intermediate validation studies used very large concentrations such that the mathematical limitations described above were not a consideration. The logic of the knowledge bases is consistent

Further development of this expert system should be accomplished using either a newer version of VP-Expert (when it becomes available) or ported to a newer operating system and inference engine. The best recourse would be to use one of the newer Windows-based expert systems. This would allow full usage of available memory, accurate manipulation of numbers, and better support of modern databases and spreadsheets.

## LIST OF REFERENCES

## LIST OF REFERENCES

- 1 J. D. Ingle and S. R. Crouch, *Spectrochemical Analysis*, Prentice Hall, New Jersey, 1988, 460.
- 2 D. A. Skoog, *Principles of Instrumental Analysis* 3<sup>rd</sup> ed., Saunders College Publishing, New York, 1985, 235.
- 3 P. T. Kissinger 'Electrochemical Detectors' in *Liquid Chromatography Detectors*, T. M. Vickrey, ed., Vol. 23, Marcel Dekker, Inc., N.Y., 1983, 125.
- 4 E. S. Yeung, Chapter 9 of *Detectors for Liquid Chromatography*, Wiley-Interscience, N.Y., vol. 89, 1986, 331.
- 5 P. J. Wyatt, *Anal. Chim. Acta.*, 272, 1993, 1.
- 6 L. R. Snyder and J. J. Kirkland, *Introduction to Modern Liquid Chromatography*, 2<sup>nd</sup> ed., Wiley-Interscience, N.Y., 1979.
- 7 J. C. Berridge Techniques for the Automated Optimization of HPLC Separations, John Wiley and Sons, New York (1985).
- 8 L. R. Snyder, J. W. Dolan, and D. C. Lommen, *J. Chromatogr.*, 485 (1989) 65.
- 9 R. M. Silverstein, G. C. Bassler, and T. C. Morrill, *Spectrometric Identification of Organic Compounds*, 4<sup>th</sup> ed., John Wiley and Sons, New York 1981.



## **CHAPTER 6**

# **AUTOMATION OF THE HIGH-PERFORMANCE LIQUID CHROMATOGRAPHY SYSTEM INSTRUMENT**

**A tool is but the extension of a man's hand, and a machine is but a complex tool. He that invents a machine augments the power of man and the well-being of mankind.**

**Henry Ward Beecher**

### **Section 6.1 Introduction**

Instrumentation has become increasingly more complex and, in some cases, more difficult to use. Along with an increase in the complexity of instrumentation, the number of technically trained people to operate instruments has decreased. These two factors are among those that have led to the development of automated instrumentation. The next evolutionary phase of instrumentation is towards the introduction of intelligence into automated instruments.<sup>1,2,3</sup> Intelligent instrumentation can monitor and diagnose itself and can alleviate the complexity of using the instrument by assisting the user in choosing the appropriate course(s) of action(s).

This chapter describes the initial stage in automating the High Performance Liquid Chromatography (HPLC) instrumentation. The intelligence component exists in the form of the stand-alone expert system described in the Chapters 4 and 5 of this dissertation. Automation has been accomplished by creating software and hardware to assist the user in operating an HPLC instrument. Much of the work described in this chapter is in the prototype stage; however, it does show the utility of automation.

column

either

Select

primi

immi:

and s

prelin

descri

instru

Most of the focus of this part of the project has been the automation of the column selection part of the expert system. The detectors used in this research are either under independent computer control or are not amenable to automation. Selection of mobile phase requires manual manipulation such as filtering, sparging, priming the pump, and using intermediate mobile phases when a new mobile phase is immiscible with a previous mobile phase.

This chapter is divided into three parts. The first part describes the hardware and software used to automate the HPLC instrument. The second part describes preliminary applications of the controller software for HPLC separations. The final part describes conclusions reached during the development of the automated HPLC instrument.

## **PART 1: HARDWARE AND SOFTWARE USED FOR AUTOMATION**

**To me, there is something superbly symbolic in the fact that an astronaut, sent up as assistant to a series of computers, found that he worked more accurately and more intelligently than they. Inside the capsule, man is still in charge.**

**Adlai E Stevenson**

## **Section 6.2     Initial Controller Software**

### ***Section 6.2.1     Objectives***

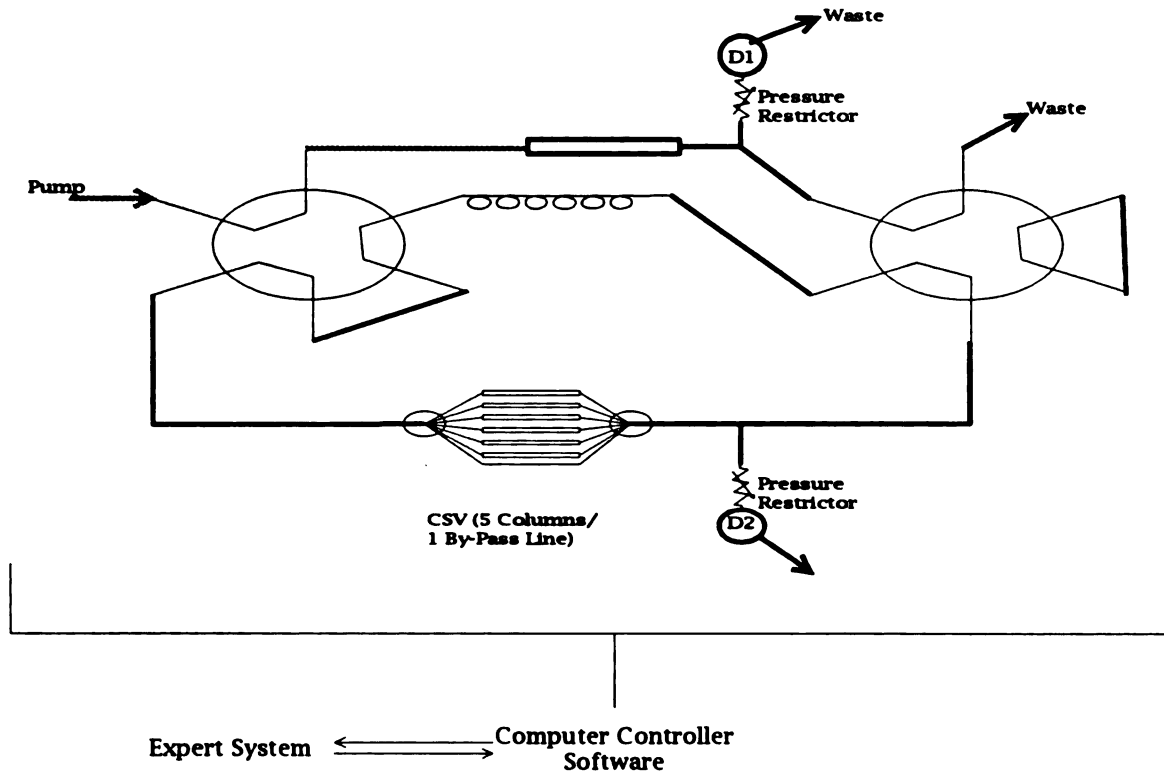
The initial objectives of the software that was developed to control the HPLC instrument were relatively simple. The controller program acquired data from the HPLC UV-Visible detector and provided simple control of two six-port valves. The control of the valves allowed the study of the application of recycle heartcut chromatography (RHC).<sup>4</sup>

It was envisioned that the software would develop into a controller program that would monitor multiple detectors, change multiple valve configurations, program the HPLC solvent delivery system, and so forth, as seen in Figure 6.1. An expert system, as described in Chapters 4 and 5, would assist in developing a method that the controller would implement. Because the software would be controlling multiple components, the data acquisition process should rely as little as possible on the central processing unit (CPU).

These two objectives are described in the following two sections. Data from the RHC experiments are also presented to demonstrate the ability of the software to switch valves.

### ***Section 6.2.2     Data Acquisition***

Data acquisition techniques typically involve the following steps. After some event initiates the acquisition process, the acquisition hardware samples and digitizes the sample. The hardware signals to the software (or other hardware) that an analog to digital conversion has occurred. The software and hardware must read and store the



**Figure 6.1**  
**Envisioned System Involving Multiple Detectors and**  
**Valves Under Computer Control**

digitized value in an appropriate memory location. The acquisition hardware is re-set and waits for the next event that will re-initiate the acquisition process. The event that initiates the acquisition can be a software command that instructs a signal to be acquired. More often, some repetitive signal, such as a high precision clock, is used to acquire the data at some fixed frequency.

Data acquisition can be sub-divided into three types: software driven, interrupt driven, and direct memory address driven. These sub-divisions are based on the manner in which the acquisition hardware signals the end of a conversion and on the method that is used to move the digitized signal to memory.

In software driven acquisition, an event, such as a software command or a clock signal, initiates the data acquisition (*vide infra*). The software monitors the status of the analog to digital conversion (ADC). Once the conversion is completed, the software reads the digitized value and stores it in an appropriate memory location, such as Random Access Memory (RAM) or on the hard disk. Software driven acquisition is simple to implement and allows the acquisition of large amounts of data (limited only by the available amount of RAM or hard disk space). The central processing unit (CPU), however, must devote valuable clock cycles to send out the appropriate controls to the data acquisition board and store the signals. During these acquisition times, the CPU cannot process other events (e.g., valve changes, monitoring the status of the HPLC run). Software driven acquisition is also the slowest method to obtain data.

Interrupt-driven data acquisition is a second approach that can be taken. The RTI 800/815 data acquisition board used in this work (see Chapter 3) can be programmed to cause an Interrupt Request (IRQ) whenever a conversion has occurred. An IRQ service routine can be written to read the digitized value from the RTI board, convert the value from two 8 byte values (a high and low byte representation, *vide infra*) into a number, and store the data in the desired memory location. With interrupt-driven data acquisition, the CPU does not have to poll continuously the status of the analog to digital converter; the CPU is involved with the acquisition process only when the conversion has occurred. This process results in faster acquisitions, and a more efficient utilization of the CPU for other processes such as valve changes. However, the CPU is still involved in the process of handling the IRQ service and storing the data. While IRQ service was not used for the acquisition, an IRQ service was written to acquire sequentially data from multiple detectors. This IRQ service is described in Section 6.4.

After considering the advantages and disadvantages of the two data acquisition approaches mentioned, the decision was made to use the third approach, direct memory

addressing (DMA). Typically, DMA is used for high speed acquisition. While the acquisition rates used in acquiring typical HPLC data do not necessarily require the speed advantages of DMA, DMA provides a second advantage. The two previous techniques (software and interrupt driven acquisition) require that the CPU be involved in the acquisition process. The controller software described in this chapter already requires fairly intensive CPU usage. The CPU must monitor the status of the run, display screen information, determine if a peak has been detected at one of the detectors, and switch valves. This is a significant load on the computers that are used to control the HPLC (a 10 MHz 8088 XT class machine, and a 12 MHz AT class computer). DMA, as described below, does not require CPU involvement in the acquisition and transfer process.

DMA is a technique in which an 8237A DMA controller chip (or similar chip) controls the transfer of data. The DMA writes the digitized data directly from the RTI board to a specified memory location. Because the data transfer runs transparently, the CPU is able to process other instructions. The net effect is a higher data acquisition rate and a decreased CPU load. The current limitation is that the DMA can only be programmed to transfer a maximum of 64 kilobytes (KB) of data before the DMA has to be re-programmed. This limitation is due to two causes. XT-class personal computers (8088 CPUs) use only 16 bit addressing which limits the addressing to only 64KB. Newer DMA chips (found in AT class machines and higher) contain 24 bit, and higher, addressing capabilities. Even with a larger address space, DMA is still limited to 64KB due to the programming language used to write the controller software. QuickBasic (as most other BASIC compilers) has a 64KB limitation on the size of variable arrays. Expansion for larger data collection sets would require porting the controller software to a newer language, such as C. Even with the 64KB limitation, DMA still provides for



up to 54.6 minutes of continuous data acquisition, assuming an acquisition rate of 1 data point per 0.1 seconds (as seen in Equation 6.1):

$$\frac{1 \text{ DataPoint}}{1 \text{ Integer}} * \frac{1 \text{ Integer}}{2 \text{ Bytes}} * 65536 \text{ Bytes} * \frac{0.1 \text{ sec}}{\text{DataPoint}} = 3276.8 \text{ sec} \quad \text{Equation 6.1}$$

$$3276.8 \text{ sec} = 54.6 \text{ min}$$

Since this time is of sufficient duration for most HPLC experiment, it was decided that extraordinary measures were unnecessary. For chromatographic separations that require more than 55 minutes, the user may select an acquisition rate of 1/second which will extend the maximum run time to 546 minutes (9 hours). A second approach would be to acquire data at the maximum rate desired. Once the 64KB limit is reached, the controller software can pause the chromatographic experiment, transfer the acquired data to a hard drive (or floppy), re-program the DMA for further data acquisition, and continue the experiment. Provisions for stopping both the HPLC instrument and RTI board are already implemented.

Two major steps must occur when DMA is used. The first step is to apply a pacing signal to the ADC component of the RTI board. This pacing signal provides a synchronized pulse that initiates an analog to digital conversion. The RTI board contains a AM9513 timer/counter chip which provides five 16-bit counters of which three are software controllable. Timer Counter 4 (TCN4) is programmed to provide the pacing for the ADC. This is accomplished by taking the desired acquisition rate (in microseconds) and converting it into a range and multiplier. The 9513 contains a 1 MHz oscillator that is divided into 5 sub-frequencies of 1 MHz, 100, 10, and 1 KHz, and 100 Hz (F1, F2, F3, F4, and F5, respectively). The desired rate is successively divided by F1-F5 until the remainder (the multiplier) is smaller than 32767 (the number, 32767, is due to maximum value that can be assigned to a variable of INTEGER type). Since the

precision of pacing is  $\pm 1$  count, this approach insures that the largest value will be used to give the smallest relative precision. The determination for range and multiplier is calculated from:

$$\text{AcquisitionRate} = \text{Multiplier} * 10^{(\text{Range} - 1)} \quad \text{Equation 6.2}$$

In this equation, the Range corresponds to F1-F5 (e.g., a Range of 1 corresponds to F1).

For a desired acquisition rate of 2 seconds (.5 Hz):

$$\text{Multiplier} = \frac{\text{Acq Rate (ms)}}{10^{(\text{Range} - 1)}} \quad \text{Equation 6.3}$$

$$\text{For Range} = 1 \quad \text{Multiplier} = \frac{2 * 10^6}{10^{(0)}} = 2 * 10^6 > 32767$$

$$\text{For Range} = 2 \quad \text{Multiplier} = \frac{2 * 10^6}{10^{(1)}} = 2 * 10^5 > 32767$$

$$\text{For Range} = 3 \quad \text{Multiplier} = \frac{2 * 10^6}{10^2} = 2 * 10^4 < 32767$$

$$\text{Multiplier} = 2 * 10^4, \text{Range} = 3 = \text{F3} = 10 \text{ KHz}$$

$$\text{RSD} = \frac{1}{2 * 10^4} * 100 = 0.005\%$$

The multiplier ( $2 * 10^4$ , in the above example) is then loaded into the LOAD register of TCN4. The source for TCN4 is given by the range, Range=3=10 KHz. TCN4 will then count repetitively down from the value stored in the LOAD register, where each count is triggered by the TTL from the sub-frequency. Upon reaching terminal count (in this case, 1), TCN4 will generate a terminal count (TC) pulse. The TC pulse triggers an analog to digital conversion at the desired acquisition rate.

The second step in using DMA is to set-up the RTI board and the DMA controller. Programming the RTI board simply requires putting a 1 in the appropriate bit of the RTI Status register. The RTI board will then send the appropriate interrupt to

the DMA controller upon the end of each digitization. The harder part is to program the DMA itself. The DMA requires two different types of information. The first places the DMA controller into the correct operation mode. The software developed here used DMA channel 1. Channel 1 is configured as:

|                       |                                                                                                |
|-----------------------|------------------------------------------------------------------------------------------------|
| <b>WRITE mode</b>     | Instructs the DMA to write to the memory address.                                              |
| <b>Autoinitialize</b> | Upon reaching terminal count, the DMA will reload the number of data acquisitions to be taken. |
| <b>Increment</b>      | The Memory Address is automatically incremented upon each DMA transfer.                        |

This mode is programmed by sending out Hex 45 to port address 0B (Hex). The second step in programming the DMA is to indicate the memory location where the transfers are to take place. The controller software reserves, by way of the dimension array command, a block of memory that will be used for the data from the detector(s). The 20 bit address of the variable is broken into a 4 bit page address. The remaining 16 bits are broken into two 8-bit high and low offsets within the page address. The page address and the high/low offsets are stored in the page and base address registers of the DMA. At this point, the DMA knows the starting address for the memory transfers. The remaining step is to program the COUNT register of the DMA with the number of transfers that are to take place. The number of transfers is a function of the number of data points (length of chromatographic run times the data acquisition rate) and the type of data that will be transferred. The software transfers integer values which require two bytes per value. Therefore, the COUNT register will contain a value equal to:

$$\text{COUNTS} = 2 * (\text{Data Run Length}) * (\text{Data Acquisition Rate}) \quad \text{Equation 6.4}$$

During the data acquisition, the controller software accesses the CURRENT register of the DMA. The value that is stored in this register contains the current number of data points that has been transferred. This value is used to determine if a valve change is needed.

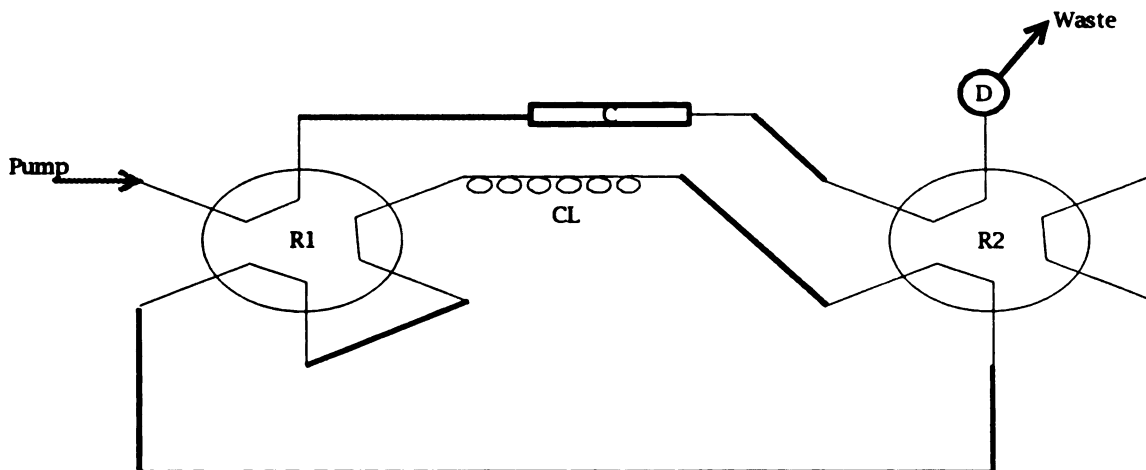
### **Section 6.2.3      *Control of Two Six Port Valves***

The software also controls two 6-port Rheodyne valves. These valves were initially used to perform recycle heartcut chromatography (RHC). Figure 6.2 provides the valve configuration used for RHC. The two valves are controlled by sending a TTL signal through the digital output lines of the RTI board. The software monitors the position of the valves through the digital input lines of the RTI board. A five volt signal is connected to a reed switch. When the valve is in the clockwise position, a magnet mounted on the body of the actuator moves over the reed switch, which closes the reed switch. The 5V signal then appears as a TTL high on the appropriate digital input line. A similar switch is mounted on the injection valve. The injector is not automated; however, the controller software monitors the status of the injection valve. When the reed switch is closed (inject position), the software notes that an injection has been made, and the software starts the data acquisition process.

### **Section 6.2.4      *Recycle Heartcut Chromatography***

Recycle heartcut chromatography (RHC) is a technique that attempts to maximize the separation of two or more components. When two peaks elute with very similar retention times, several procedures can be taken to improve the resolution. Solvent optimization or a new mobile phase can be chosen to change the capacity factor(s),  $k$ , of the closely eluting peaks. Likewise, a column with a different selectivity factor,  $\alpha$ , may resolve the components. Improving the capacity factor or selectivity factor will improve resolution,<sup>5</sup>  $R_s$ , according to equation 6.5:

$$R_s = \left(\frac{1}{4}\right)(\alpha - 1)\sqrt{N}\left(\frac{k'}{1 + k'}\right) \quad \text{Equation 6.5}$$



R1/R2 Rheodyne 6-Port Valves  
 D HPLC Detector  
 CL Collection Loop  
 C HPLC Column

**Figure 6.2**  
**Valve Configuration for Recycle Heartcut**  
**Chromatography**

A third approach to improving the resolution is to increase  $N$ , the number of theoretical plates. One way to increase  $N$  is to recycle the unresolved components through the same column (e.g., RHC).<sup>6,8</sup> RHC assumes that if the non-resolved components are extracted (heart cut) and re-directed back onto the column, they will experience further separation. By recycling the components back onto the column, the effective column length is increased. There are several advantages to this approach.<sup>7</sup> The pressure drop across the column is minimized compared to using an actual longer column. A single column has the same characteristics that allow for easier optimization. Finally, by selective heartcutting, components that are not of interest can be removed.

There are several disadvantages to RHC. The first is that multiple passes may be required to resolve the components. Each recycle requires that the heartcut plug be stored while the remaining components elute. After multiple recycles, the volume of the recycle plug equals or exceeds the dead volume of the column. At this point any further recycles would require that a recycle be started before the complete elution of the recycle plug.<sup>4</sup> The presence of ever longer tubings to store the cut increases the dispersion. The result is an eventual decrease in detector signal as the sample is dispersed. A second limitation is that if the unresolved peaks are complex, the resolving power of the column may never be large enough to separate completely the components. At this point, a complete change in the chromatography is required.

RHC is implemented by changing the valve configurations for R1 and R2 (see Figure 6.2). At the beginning of the chromatographic experiment, both R1 and R2 are in the clockwise configuration (CW).

When the components of interest start to elute, R1 is left in the CW configuration while R2 is placed into the counter-clockwise position (CCW). This directs the flow from the column through the collection loop (CL in Figure 6.2). The collection loop is a coiled tube of a particular volume needed to collect a certain elution volume. Dispersion is minimized by using a coiled CL tube.<sup>4</sup> The coiled tube breaks up, via the Coriolis effect, the laminar profile that leads to longitudinal dispersion.

After the components of interest have been collected, R1 and R2 are changed back to the CW configurations to allow the continued elution of the remainder of the sample. After a sufficient time to elute the rest of the sample, R1 is placed in the CCW configuration. The flow is then directed through the collection loop and back onto the column.

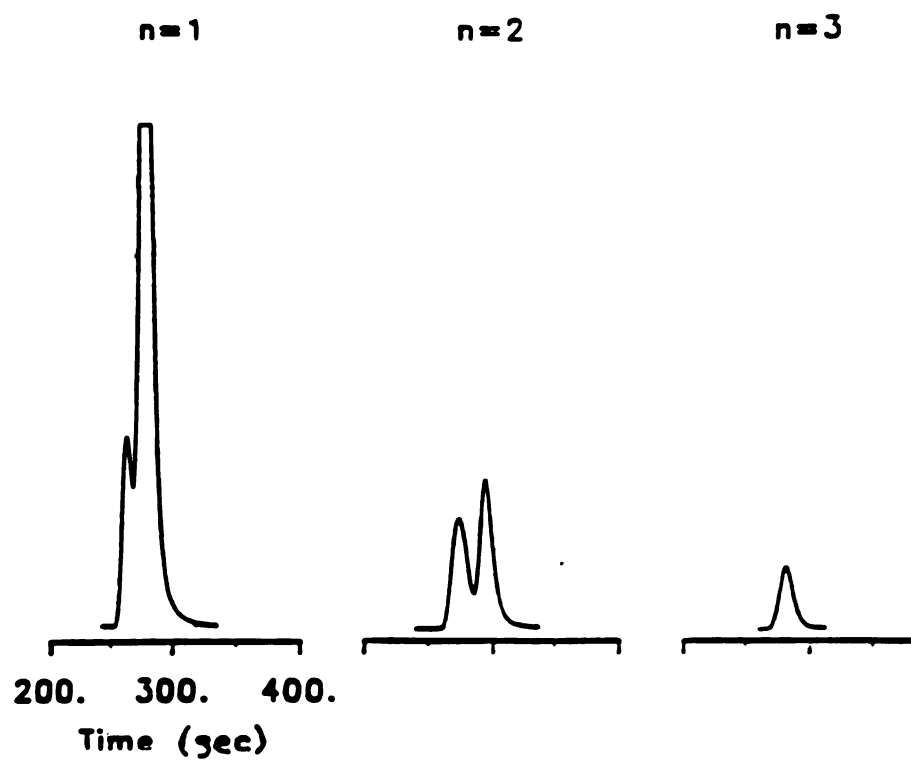
The separation enhancement of a mixture of naphthalene (250  $\mu\text{M}$ ) and biphenyl (200  $\mu\text{M}$ ) using RHC is shown in Figure 6.3.<sup>4</sup> The separation was accomplished on a

Dupont 4.6 mm x 25 cm Zorbax ODS column using isocratic elution with 92.5% methanol at 1 ml/min. After the first pass, the two components are not completely resolved. The higher response of the detector to the biphenyl peak has made it difficult to quantify the naphthalene peak. A heartcut was made to discriminate against the biphenyl peak, as seen in the second pass. Resolution has been increased and the response to the biphenyl peak does not predominate the chromatogram. A second cut was made to provide additional selectivity enhancement towards the naphthalene peak. The results of this cut are seen for the third pass. At this point the remaining component is predominately naphthalene.

The inclusion of this chromatogram demonstrates the potential for using automation. A reproducible method of automated control over valve switching is required to obtain good precision.<sup>4</sup> Without the automation of the instrument, it would have been difficult to switch the valves quickly enough to allow this technique to work. The controller software written by the author switched valves based on time only. The software was modified to allow the operator to change the valves at his/her discretion to perform RHC. The current version of the software controller allows switching based upon the peak detection. This development would improve the automated RHC technique such that the heartcutting could be more reproducible.

### **Section 6.3      Second Level Controller Software—An Overview**

The next stage of development of the controller software incorporated significant changes. The objectives in developing this second generation of the controller program are:



**Figure 6.3**  
**Selectivity Enhancements using Recycle Heartcut Chromatography:**  
**Naphthalene and Biphenyl**



- (1) Program methods for the solvent delivery system
- (2) Monitor the solvent delivery system
- (3) Control the column selection valve and the two Rheodyne valves
- (4) Provide support for multiple detectors

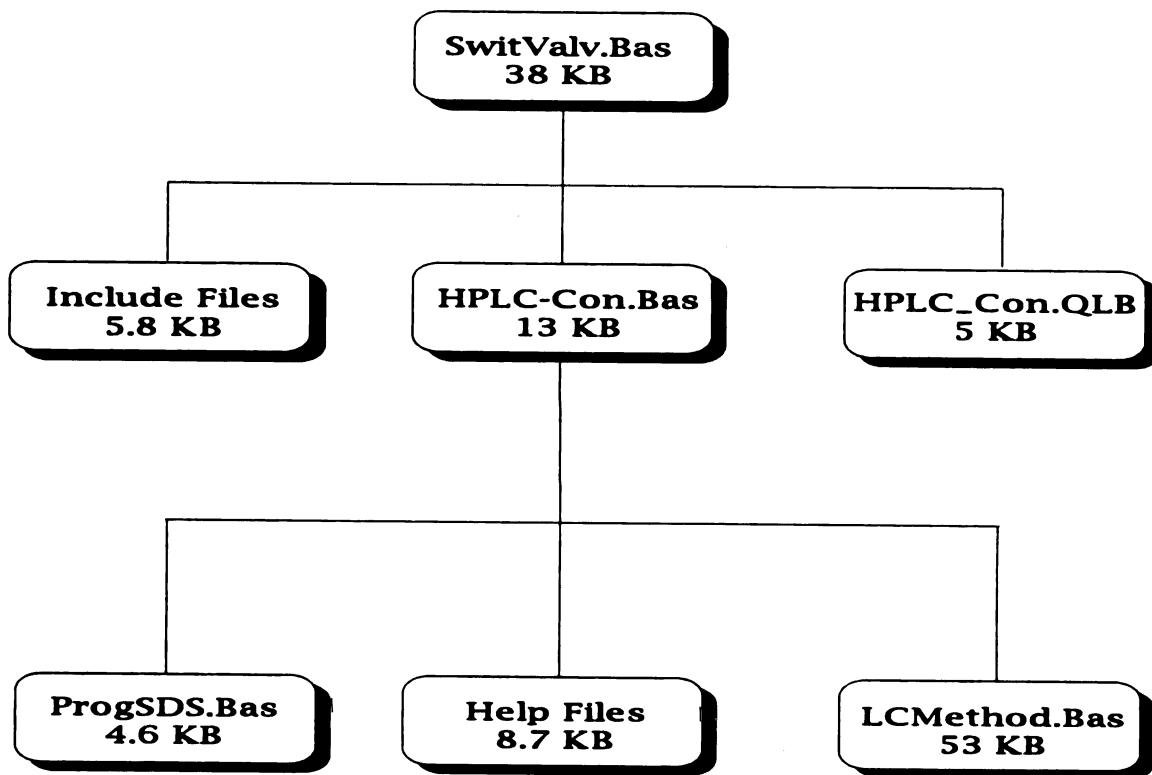
A block diagram of the software is provided below in Figure 6.4. Provided with each module is the approximate size, in Kilobytes, of each module. Modular programming allows independent development of each component of the software. QuickBasic also limits the size of a single module to 64 Kilobytes (KB). The total code for these modules would not fit into a single 64 KB code segment. Each module will be briefly discussed.

### ***Section 6.3.1 SwitValv.Bas Program***

The main module is the SwitValv.Bas program. SwitValv calls the other modules to set up HPLC run parameters and valve configurations. SwitValv first initializes the DMA and obtains the length of chromatographic experiment and the data acquisition rate. After programming the DMA, SwitValv programs the three timer counters, TCN1, TCN2, and TCN4.

TCN4 has already been described. TCN1 and TCN2 are used to provide accurate and precise timing of the chromatographic experiment. The source for TCN1 is F4, which has a frequency of 1 KHz. This allows TCN1 to count in milliseconds. When TCN1 reaches terminal count (65536) it generates a terminal count pulse, resets itself, and continues counting. This TTL pulse from TCN1 serves as the source for TCN2. Together, TCN1 and TCN2 allow the software to count for a maximum of 49.7 days.

TCN1 and TCN2 are used for several vital purposes. These two counters provide a high precision clock for accurate and reproducible valve switches which may be programmed to change at a particular time during the run. The run time is also displayed on the screen to provide feedback for the user.



**Figure 6.4**  
**Overview of the Modules used in the Second**  
**Generation of the Controller Software**

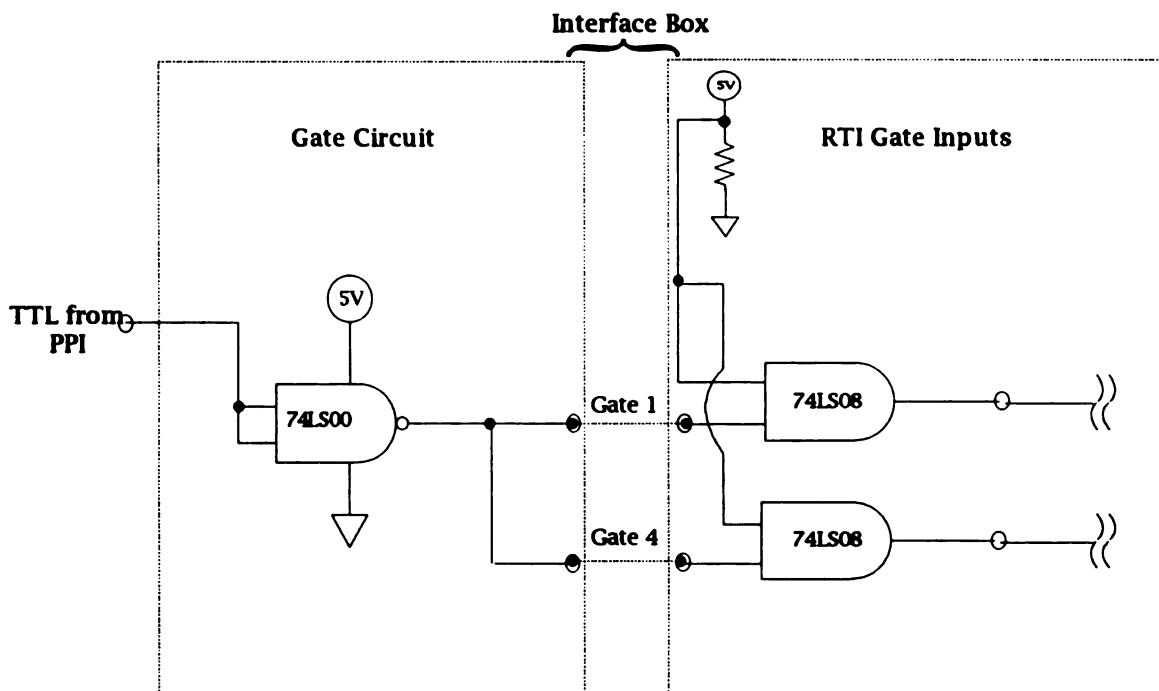
The use of column switching requires that the HPLC be placed in a stand-by mode whenever a valve switch is made. This pause may be due to a change in mobile phase or flow rate. Therefore, a gate must be applied to TCN1 and TCN4 to stop the data acquisition and clock while the controller program allows the columns, mobile phase, and flow rates to equilibrate. The gate is provided by sending a TTL High to a gate circuit composed of a 74LS00 chip (Quad NAND gate). The TTL signal is sent from one of the digital outputs of the programmable parallel interface chip (vide infra). The output from the chip is then sent through the gate input for TCN1 and TCN4. TCN2 does not need to be gated since the source for TCN2 is the terminal count pulse from TCN1. The circuit for this gate is shown in Figure 6.5.

When a High signal is applied to the NAND gate, the output from the chip is a logic level Low. The 5V internal signal from the RTI board is pulled to a logic Low. Since TCN1 and TCN4 are programmed to be Active High gated, the counters are stopped.

SwitValv next obtains the Master File (which must already exist before running the program). This Master File is an ASCII text file that contains the desired valve configurations for the Rheodyne valves and the column selection valve. The format of this Master file is provided in Figure 6.6. Up to ten lines of description can be included. The next line must contain, as the first characters, <END COMMENTS>. Following this entry is a line which contains the total number of different valve configurations that will be used during the run. The next line is simply a header line to provide readability to the file. The remaining lines of the file contain the values for the fields (tab, comma, or space delimited format).

The Event field contains a code that specifies when that valve configuration will be implemented. For example, TT specifies that after a specified total run time the valve configuration will be changed. The valve can be changed at a particular time or

after a certain number of peaks have been detected. "ELAPSED" refers to the time and number of peaks during a particular valve configuration; the time and number of peaks detected are re-set to zero after every valve change. The total time and number of peaks detected is kept and can be used as selection criteria.



**Figure 6.5**  
**Gate Circuit to Control**  
**Timer Counters 3 and 4**

The Value field is the value to use for the event. For a timed event, the value field will be time (in minutes). For a switch based on a peak count, Value contains the number of peaks for which to look. The valve change can be based upon the emergence of a peak or the total number of peaks that have occurred. Switching based upon a peak emergence allows the controller software to cut a particular peak(s) onto a second

column. Switching after a certain number of peaks have been detected can be used to insure that all the peaks have been eluted from the column.

The Detector field indicates which detector output is used to monitor when the valve change is based upon peak detection. This value is needed since different valve configurations will place certain detectors on- or off-line. The Detector field also provides additional flexibility for the user of the software since different detectors can be monitored.

Finally, the last three fields are the positions for the three different valves that are used. The first two fields are for the Rheodyne valves where a 1 in the fields indicates that the valve(s) will be placed in a clockwise position and a 0 refers to the counter-clockwise position. For the column selection valve (CSV), the legal value is an integer is between 0 and 5, inclusive, which indicates one of the six different flow paths through this valve. The CSV has been described in Chapter 3.

```

This is an example Master File. There are
three lines of comments in this file.
Date: 10/28/92      Author: Stephen Medlin
<END Comments>
3
Event      Value      Detector      V1           V2           CSV
SPE         2           1             1             1             4
EPE         2           2             0             1             2
TT          14.3        1             1             1             3

```

### Legend

|     |                                                         |
|-----|---------------------------------------------------------|
| TT  | Switch after a total run time                           |
| TE  | Switch after an elapsed time on that column             |
| SPT | Switch at the start of peak (Total count)               |
| EPT | Switch at end of peak (Total Count)                     |
| SPE | Switch at start of Peak (Elapsed)                       |
| EPE | Switch at end of Peak (Elapsed)                         |
| Dn  | Use Detector n for the peak counting (0..NumChannels-1) |

**Figure 6.6**  
**Format of the Master File**

SwitValv passes control to the HPLC-Con.Bas module. HPLC-Con creates HPLC parameter files for each valve configuration. This module is described below. HPLC-Con passes back two sets of information. The first is a two dimensional array that contains the HPLC parameters for every valve configuration. Each element of the array is a complex variable type of the form:

LCRun(Column\_Number, Event).Time.Flow.PerA.PerB.PerC.MaxP

where:

1. Column\_Number is an index to represent a particular valve set-up
2. Event is an index that indicates a different point of the run for that configuration (e.g., the different points during a gradient run where the HPLC changes the mobile phase or flow rate).
3. Time indicates the elapsed run time when the HPLC solvent delivery system will be changing the conditions.
4. Flow provides the flow rate for that point in the elution.
5. PerA, PerB, and PerC are the mobile phase compositions (in percentages) for up to three different solvents.
6. MaxP indicates the maximum pressure limit to which a particular column will be exposed

The second parameter that is returned to SwitValv is a two dimensional array, LINKFILE. This array is an encoded linked list that indicates which HPLC parameter file is associated with each valve configuration. The Spectra Physics SP8700 solvent delivery system (SDS) can store a maximum of ten parameter files, with each file containing only seven lines of programming. While the SDS allows parameter files to be linked, it is easy to use all of the parameter files when performing a large number of column switches. The automated instrument is also able to link the appropriate files via the LINKFILE array. The values of LINKFILE indicate which SP8700 files are associated with the method or the particular valve configurations. The first dimension of LINKFILE indicates the order of the valve configuration. The second dimension contains the encoded link list information. Table 6.1 gives the format of the information stored in this array.

After all the methods have been programmed and control returns to SwitValv, SwitValv runs the HPLC instrument. The program continuously provides an updated screen display indicating the run time, the next event for the next valve change, the current digital value from the detector, and which detector is being scanned.

SwitValv also determines if a component is eluting by performing a linear regression upon the last seven data points that were acquired. When a positive slope greater than some user-defined threshold is observed and the average of the seven data points is larger than a second threshold, the computer indicates the beginning of a peak. The program determines the end of the peak when two criteria are met. First, a peak must have been observed (a positive slope change). Then, a negative slope must be observed; followed by a slope and average value that is less than the two relevant thresholds (i.e., the peak has returned to baseline).

If a valve change should be initiated (based on time or peak), SwitValv moves the valves into the appropriate positions. The Rheodyne valves are switched by sending out the appropriate TTL signal via the RTI board. The CSV is moved by sending out the appropriate number of TTL signals via the 8255 Prototype board. After a delay time to allow the valves to switch, the program determines the status of the valves to make sure that they are in the desired position. If they are not, the program will try up to two more times to switch the valves before calling an error routine.

After every switch, the program will store the time and peak count when the valves were switched. If a solvent parameter (Flow or mobile phase composition) changes for the new valve/column configuration, the program sends out a TTL signal to gate off the DMA and timers. The program switches the Rheodyne and CSV valves to bypass mode (the CSV has simple tubing as one of the 6 flow paths so that the flow will not go through any of the columns on the CSV), and changes the mobile phase and flow rate to the new values. The SDS has a TTL line that goes Low when the desired flow

rate has been met. After a user-defined time to purge the plumbing of the old mobile phase with the new, the program sends out a Low to the gate to re-start data acquisition.

The steps of monitoring, displaying, and changing valves are iterated until several conditions are met. The first condition is that all the data points have been collected and the DMA controller changes the status bit of the RTI board to indicate that Terminal Count has been reached. The second condition is a user-initiated abort which is accomplished by the user hitting a capital S during the run. The program stores the acquired data to that point. The program closes any open files and stops the run.

### ***Section 6.3.2      HPLC-Con.Bas Program***

This module is the interface between the master module, SwitValv, and the modules that assist in creating and programming the SP8700 method files. HPLC-Con calls two sub-modules, LCMMethod.Bas and ProgSDS.Bas. The program returns the programmed methods in the array, LCRun, and the linked list in LinkFile.

### ***Section 6.3.3      LCMMethod.BAS Program***

The purpose of this module is to obtain a SP8700 parameter file containing valid entries. The module either reads in a method file that has been previously created (assumed to be created via LCMMethod and therefore containing valid entries) or assists in creating the method. The program is user-friendly and provides error checking, error messages, and help files to assist the user in developing an SP8700 parameter file.

The program allows the user to enter the time for a particular change in the operating conditions to be made (e.g., developing the different parts of a gradient elution profile). The times are entered as elapsed times rather than as total time, which the SP8700 eventually requires. The program also allows the user to enter a time to



hold the elution profile. After the user has entered the time, the program obtains valid entries for the mobile phase composition, flow rate, and maximum operating pressure. The program converts the elapsed times into cumulative times. After entry of the method, the program displays the entered method in a scrollable manner and allows the operator to edit the method.

**Table 6.1**  
**Format for the Encoded Link List Array, LinkFile()**

| Array Indices<br>(m,n) | Value of<br>n Index | Explanation                                                                                                                                                                                                                                                                                                                                                                             |
|------------------------|---------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| (m,1)                  | -1                  | No open SP8700 files are available. SwitValv module must then monitor when a method has been run and program the complete method into the SDS.                                                                                                                                                                                                                                          |
| (m,n)                  | 0 .. 9              | The number Indicates the order of the SP8700 file. For example, if LinkFile(3,2)= 4, then the second linked file for the third valve configuration is SP8700 file number 4.                                                                                                                                                                                                             |
| (m,n>0)                | -1                  | Last linked file. The method was completely programmed into the SP8700 SDS.                                                                                                                                                                                                                                                                                                             |
| (m,n>0)                | >100                | A partial file was loaded into the SP8700. The most significant digit indicates that a partial file was loaded while the remaining digits indicate which was the last line programmed into the SDS. For example, if LinkFile(2,6)=123, line #23 of the method was the last line programmed. When an SP8700 file becomes available, lines 24 to the end must be programmed into the SDS. |

Once the method has been programmed, the program allows the method to be written to a file, along with appropriate comments, headers, dates, and times. This approach allows methods to be generated and stored for later recall and reuse.

LCMethod returns the parameter file to the HPLC-Con module that combines the parameter file with the other parameter files for the different valve changes.

This module was written to improve the user-friendliness of the HPLC instrument. The program assures that the times are calculated correctly and that the different parameters are entered correctly. The program provides fairly detailed help in the form of scrolled text files. The first time through the module, the program displays the help files. Subsequent uses of the module remove the help files so that the user may quickly enter the parameters. By saving the method to a file, the productivity of the operator can be improved since complex methods can be re-called and re-used.

#### ***Section 6.3.4      ProgSDS.Bas Program***

The final module in the HPLC controller program is ProgSDS. This module is called from HPLC-Con.Bas after each method file has been generated (or read in from a file) via the LCMethod.Bas module.

This module has two principal objectives. The program first sends out the parameters to the SP8700 SDS. The program can send out a command as fast as 0.2 seconds per command for rapid and error free programming of the SDS. A wait state of 0.2 was imposed since the program is able to send out commands faster than the SP8700 is able to accept.

The commands were sent via six digital lines from the 8255 Prototype board. The ability of a computer to serve as a keypad emulator for the SP8700 has been described previously.<sup>8,9</sup>

This module also monitored the availability of the SP8700 files and reported which files were in use or vacant to the HPLC-Con module. The module also encodes the LinkFile array to indicate the status of programming the methods into the SDS.

## **Section 6.4     Assembly Language Routine**

The final software that was written to automate the HPLC is an assembly language program which serves as an interrupt request (IRQ) service routine. One goal of the controller software is that it be able to monitor multiple detectors. A detector is placed after the preliminary column while a second detector is placed after the CSV. The presence of these two, or more (up to 16), detectors allows the program to monitor the elution from two different columns. To monitor two different detectors, the RTI and DMA controller must be able to read the detector signals at two, or more, different analog input channels. The RTI 800/815 board can be programmed to scan sequentially the different analog channel inputs; however, it scans all sixteen channels.

To overcome the inflexibility of the RTI board, an interrupt procedure was written that sequentially increment the multiplexer register (the multiplexer register determines which analog channel is read and digitized). After reading all n-detectors, the assembly language program wraps around to the first channel.

The RTI board is programmed to generate an Interrupt on IRQ 2 (or IRQ 9 for AT class machines and above) when an analog to digital conversion is completed. The first part of the assembly program is the loader. Whenever the operator uses more than one detector, SwitValv calls the loader. The loader stores and replaces the current interrupt service routine for IRQ 2/9 with the new routine. When the acquisition is completed, SwitValv calls a second procedure to restore the original service routine.

The IRQ routine is called after each analog to digital conversion. The program reads the current value of the multiplexer, increments it by one, and wraps back to the first channel if the multiplexer was reading the final (nth) detector. The program then re-programs the multiplexer register of the RTI board.

By scanning the channels, SwitValv must de-convolute the data when saving the data or when determining if a component is eluting. This is a simple process of taking

the current data point (obtained from the DMA current word register) and dividing it by the number of detectors. The modulus of the division indicates which detector is being used. To obtain the current data from a particular detector, the modulus and desired channel number are subtracted from the current word. The modulus is subtracted repetitively to obtain the latest seven data points from the desired detector. These data points are used in the linear regression determination of slope. For example, assume three detectors are being used, the current data point is the 100<sup>th</sup>, and information from the second detector is desired.

$$100 \text{ Mod } 3 = 1$$

The value of 1 indicates that the 100th data point came from the second detector (the first detector is labeled as detector 0 on the RTI board). Therefore the last seven data points for the second detector are:

$$\begin{aligned} 100 - 0 - 2(0) &= 100 \\ 100 - 0 - 2(1) &= 98 \\ 100 - 0 - 2(2) &= 96 \end{aligned}$$

and so forth. In the calculations, zero is the modulus, the value of two indicates that the data is to be taken from the second detector, and the multiplier of 2 (e.g., 0, 1, 2, ..., 6) provides the seven successive data points.

## **Section 6.5 Programmable Parallel Interface Prototype Board**

The previous section described the software that was written to provide an automated instrument. This section describes the hardware used to provide additional automation for the HPLC instrument. In particular, this section describes a prototype card containing an 8255-5 Programmable Parallel Interface chip.

The initial controller program used the digital input/output lines on the Analog Devices RTI 800/815 data acquisition board to control the different devices. At that time, the capabilities of the board were sufficient to provide all the digital input/output

(I/O) lines that were needed. The RTI board has 8 input and 8 output lines that can be used. Of these, two outputs were used to control the two Rheodyne valves and three input lines were used to monitor the position of the two Rheodyne valves and the injection valve.

The second generation of the controller program exceeded the number of digital I/O lines that were available. Table 6.2 describes the digital lines that are needed. The entries marked with an asterisk are already provided by the RTI board. Since the controller program now requires seventeen additional I/O lines, in addition to the five lines already in use, the PPI prototype board was developed.

A Wirewrap Prototype Card, Model IBM-PR2, was purchased from JDR Microdevices (Los Gatos, CA). This board contains the address encoding hardware wired on the board. The board requires only that the support chips be soldered onto the board. These support chips enable addressing of both the board and up to eight peripheral chips on the board. Communication with the board is accomplished by setting address lines  $A_8 - A_5$ , via a DIP switch, to an unused port address ( $A_9$  is hardwired to a 1). Address lines  $A_4 - A_2$  are used to select one of the eight possible peripheral chips.  $A_1$  and  $A_0$  are used as control words for the PPI chip. Port address 340 (hexadecimal) is currently used by the Prototype board.

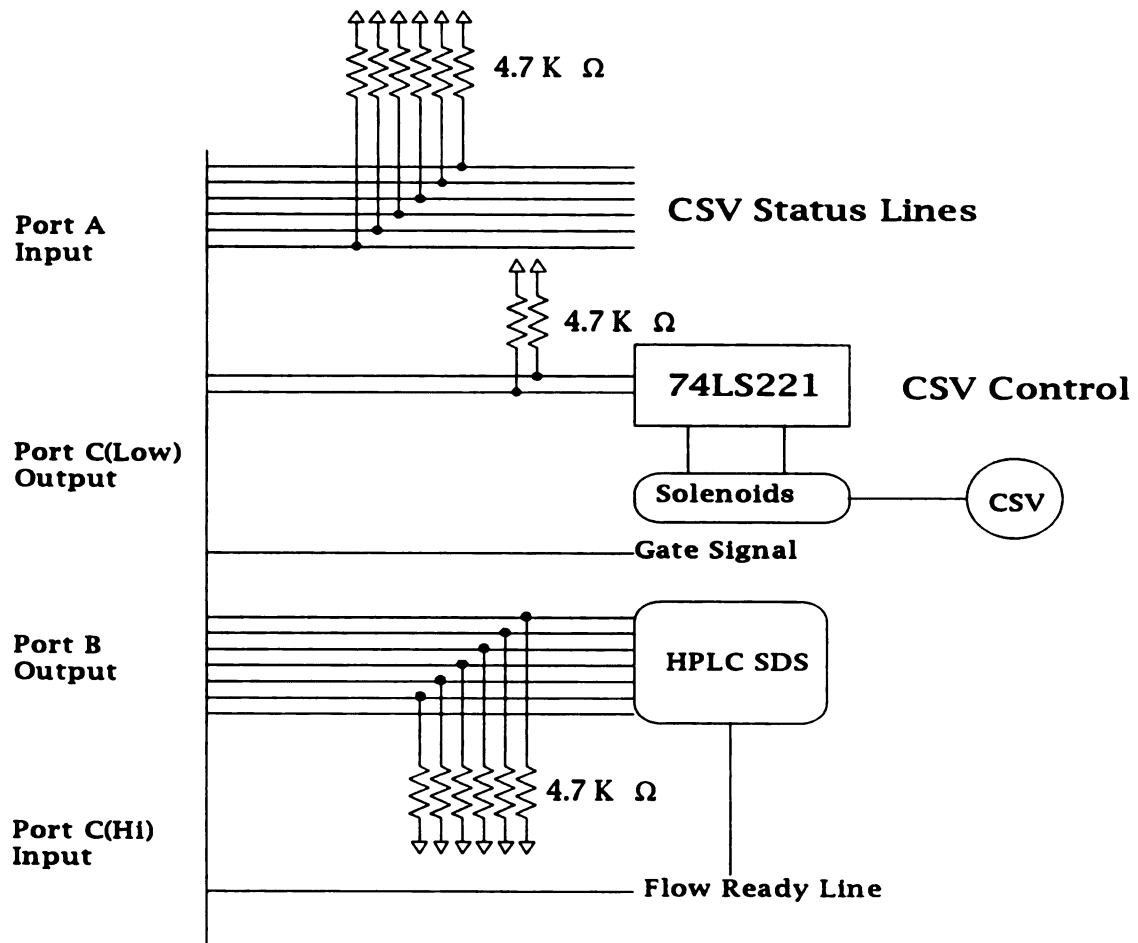
The PPI chip (8255A-5) is a 40 pin, VLSI (very large scale integrated) chip which provides twenty-four programmable TTL I/O lines with very good DC driving capability.<sup>10,11</sup> The Inputs are not latched while the output lines are latched. The PPI has three modes of operation. In Mode 0, groups A and B are 8 bits wide while group C is divided into two sub-groups,  $C_{High}$  and  $C_{Low}$ , each 4 bits wide. A, B, and the two C groups can be independently programmed to be either inputs or outputs. In Mode 1 Groups A and B are either inputs or outputs and Group C is used as a control for handshaking. Mode 2 uses one 8 bit wide bi-directional port and one 5-bit control port.

Group A and B can be programmed in any one of the three modes independent of the other. The PPI can read the ports faster than 250 nanoseconds. The flexibility of the PPI allows for a wide possible use to control devices.

**Table 6.2**  
**Digital Input/Output Lines Needed**  
**for the Second Generation Controller**

| Use                              | Number of I/O Lines Required | Input/Output |
|----------------------------------|------------------------------|--------------|
| SDS Control                      | 7                            | Output       |
| CSV Control                      | 2                            | Output       |
| CSV Status                       | 6                            | Input        |
| Flow Ready Status Line           | 1                            | Input        |
| Gate Control for Timer- Counters | 1                            | Output       |
| *Control Rheodyne Valves         | 2                            | Output       |
| *Rheodyne/Injector Status        | 3                            | Input        |

In this research project, the PPI is programmed to Mode 0 to provide the largest number of I/O lines. Group A is programmed as Input and is connected to the CSV to monitor the position of the valve. Group B is set to Output mode and is connected to the SDS to allow the controller software to program the SDS. Group  $C_{Low}$  operates in Output mode. Two of the four bits are connected to the 74LS221 monostable circuit (vide infra) which controls the CSV. One other bit is connected to the gate circuit. Group  $C_{High}$  is set up in input mode and is used to monitor the Flow Ready signal from the SDS. These connections are shown in Figure 6.7.



**Figure 6.7**  
**Digital I/O Connections from the**  
**Programmable Parallel Interface**

During testing of the circuitry, it was determined that pull down resistors were required. The high impedance of the PPI acted as a capacitor which caused erratic and false signals to appear on the digital lines. Applying a momentary high to one of the CSV status lines took approximately two seconds before the voltage dropped back to a low signal. When the computer was first turned on, residual voltages appeared on the CSV control lines, which caused the CSV immediately to fire. The 74LS221 circuit would work only 80% of the time since the TTL signals that were applied to the circuit would not go to a Low after the signal was removed. The solution was to put a current limiting resistor tied to ground on each I/O line. This pull down resistor provides a discharge path and allows the voltage on the I/O line to return to zero volts (ground) shortly after the TTL High signal is removed.

A second problem that occurred was inconsistent valve switching due to ground loop problems. The problem stemmed from the computer, HPLC, and valves not being at the same common voltage. A single common point was made on the interface box that connects the RTI board to the HPLC components and a common line was brought from every component to this common point. The inconsistency problem was resolved by this procedure.

## **Section 6.6      Pneumatic Amplifier**

### ***Section 6.6.1      Motivation for Use***

During a valve change, the flow of the mobile phase is perturbed while the valve moves from one position to the next. The perturbation can create a shock wave that travels down the column. This pulse can shorten the lifetime of the column; therefore, it is desirable to decrease the transition time between positions of the valve to minimize the pressure wave. For column switching, valve changes have the undesirable potential



to cause eluants to be split between two, or more, columns if the valve switching is too slow. Harvey and Stearns<sup>12</sup> presented a method to increase the speed of valve switching by using a pilot valve. A pilot valve puts a higher flow of gas to an actuator to amplifying pneumatically the air pulse used to actuate the valve. After discussion with Harvey,<sup>13</sup> it was decided that a pilot valve would be appropriate in conjunction with the CSV used in this dissertation.

Figure 6.8 provides a schematic of the pilot valve. A TTL signal is applied to the Opto-20 optically isolated relay that applies a 110V signal to the solenoid. The solenoid then sends an air pulse to the pilot valves. The pilot valves are in turn actuated and open the valve to the high pressure gas. The CSV requires two different events to initiate a valve change. First, the CSV must be ratcheted, or armed. After being armed, a second air pulse moves the valve.

### ***Section 6.6.2 Initial Software Control***

Initially the CSV and pilot valve assembly were controlled by sending out two TTL signals to the two Opto-22's connected to the ARM and FIRE ports of the CSV. Software pauses were required between arming and firing the valve. These pauses were accomplished by putting a dummy loop structure into the code:

```
FOR X=1 TO 1000
  X=X
NEXT X
```

The dummy structure created a 0.123 second transition time between positions of the CSV. The transition time was deemed to be unacceptable since a total transition time to switch from position 1 to 6 would require 0.615 seconds. Also, while the CPU is occupied with executing the dummy loop, a peak may not be detected or a valve change may be missed. One final concern is that the reproducibility of valve switching is decreased since the time between switching is now determined by the CPU. If a

different computer is used, the code must be changed to a new loop time. Also, the time required to iterate the above loop varies even on the same computer. This is due to the Interrupt Request service routine and DMA controller taking control of the CPU and/or the data bus (even ambient temperature affects the CPU speed).

The next development was to create a hardware circuit that would generate very fast and reproducible TTL pulses. Using hardware to control the pulses makes the valve switching independent of the computer used. This circuit is described in the next section.

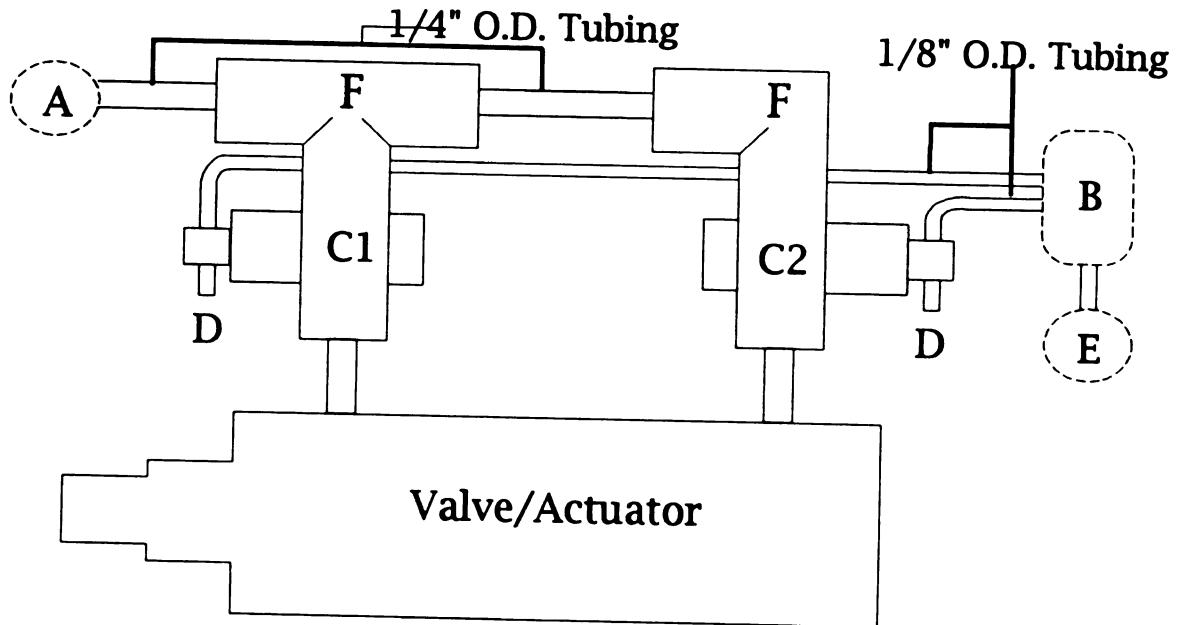
### ***Section 6.6.3      Timing Circuit***

A timing circuit using a 74LS221 dual monostable was made. This circuit is given in Figure 6.9. The two inputs, TTL 1 and TTL 2, are supplied from the prototype board. The output, Q1, is used to ARM, or ratchet, the CSV and as an input to the second monostable. Output Q2 is used to FIRE, or move, the CSV. The width of the peak generated can be calculated from

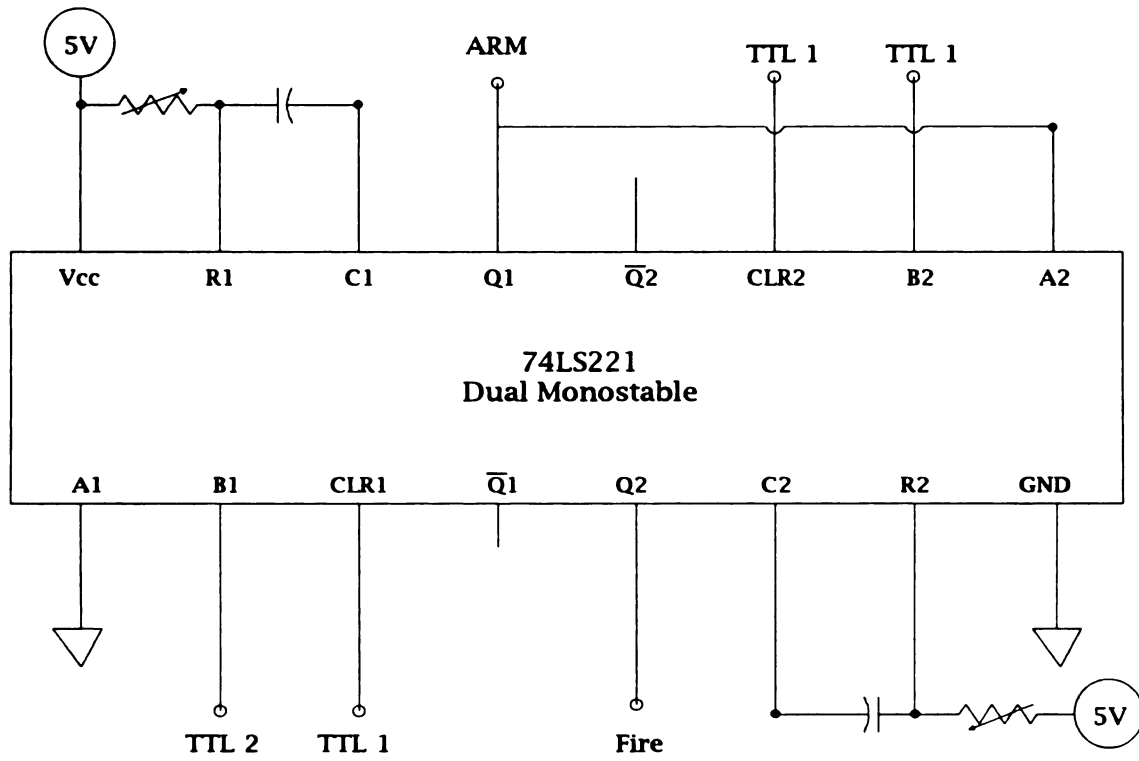
$$t_w = R_{ext}C_{ext}\ln 2 \quad \text{Equation 6.6}$$

where  $R_{ext}$  and  $C_{ext}$  are the external resistance and capacitance for each monostable.

Table 6.3 provides the truth table for the monostable. Figure 6.10 gives the timing scheme that is used to generate two pulses. The falling edge of Q1 is used to trigger the output from Q2. In this fashion, two successive peaks are generated so that the CSV has time to ARM and then immediately FIRE.



**Figure 6.8**  
**Schematic of Pilot Valve**



**Figure 6.9**  
**74LS221 Dual Monostable Circuit**

**Table 6.3**  
**Logic Table for the 74LS221 Dual Monostable**

| CLR                                                                               | A Input                                                                           | B Input                                                                           | Q Output                                                                           | $\bar{Q}$ Output                                                                    |
|-----------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|-----------------------------------------------------------------------------------|------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|
| L                                                                                 | —                                                                                 | —                                                                                 | L                                                                                  | H                                                                                   |
| —                                                                                 | —                                                                                 | L                                                                                 | L                                                                                  | H                                                                                   |
| —                                                                                 | H                                                                                 | —                                                                                 | L                                                                                  | H                                                                                   |
|  | L                                                                                 | H                                                                                 |  |  |
| H                                                                                 |  | H                                                                                 |  |  |
| H                                                                                 | L                                                                                 |  |  |  |

 Rising Edge

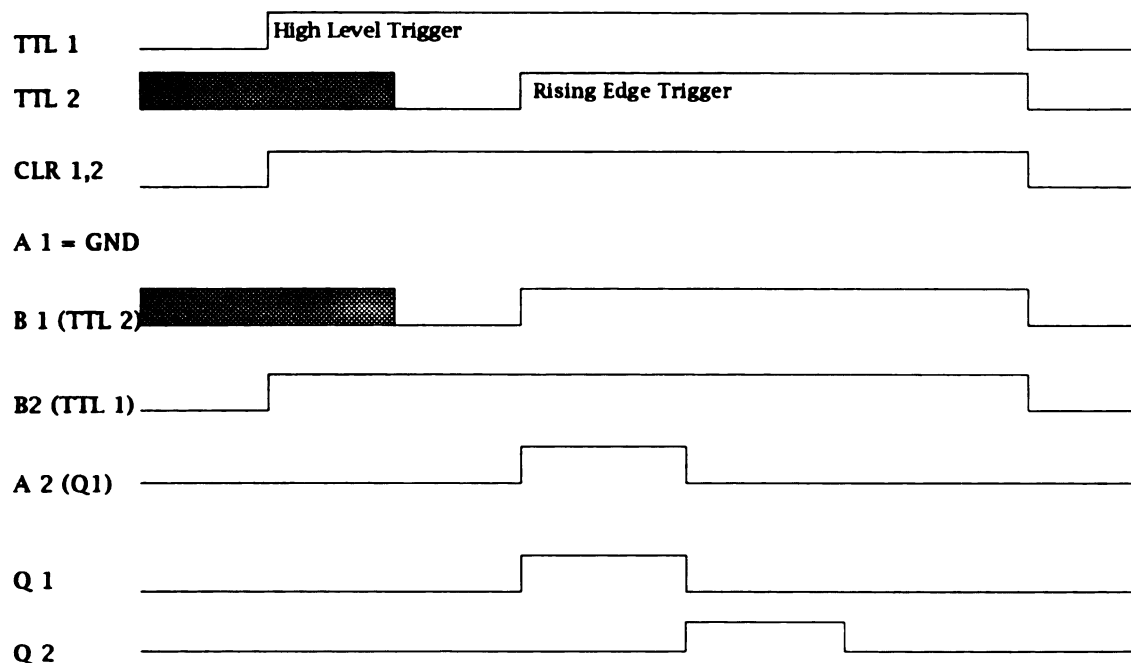
 Falling Edge

— Does Not Matter

A series of experiments was done to determine the fastest possible switching time. Different resistors were chosen to give smaller pulse widths for ARM and FIRE, and at least twenty-four valve switches were made for that combination of pulse widths. When the CSV started not to make the valve switches, the gas pressure to the pilot valves was increase or the ratio of ARM to FIRE pulse width was changed. The results of these experiments are presented in Table 6.4. All times are in milliseconds and gas tank pressures are in psi.  $P_{\text{solenoid}}$  and  $P_{\text{pilot}}$  refer to pressures (and gas types) used for the solenoid and the pilot valves, respectively.

Not included in this table are a series of experiments in which  $t_{\text{arm}}$  was much smaller than  $t_{\text{fire}}$  since the arming time was determined to require at least equal to, or longer, time than the firing time; otherwise, the valve will not switch at all.





**Figure 6.10**  
**Timing Scheme used to Switch the Column Selection Valve via**  
**the 74LS221 Timing Circuit**

Pressure and gas choice affect significantly the switching times. Once the solenoids are fired, a higher pressure allows the air pulse to traverse more quickly from the gas tank to the pilot valve and CSV. Switching to helium for both the solenoid and the pilot valve also improved switching times. This follows from Graham's Law that states that gas velocity is inversely related to the square root of the molecular mass of the gas. He should be approximately 2.6 times faster than nitrogen. This speed increase can be seen in the drop from a minimum switching time of approximately 50 milliseconds to approximately 35 milliseconds. Unfortunately, the maximum pressure is 150 psi (rated at 100 psi) which limits the fastest switching time to approximately 35 milliseconds. This limitation gives a maximum transition time (switching from position 0 to 5) of 175 milliseconds.

**Table 6.4**  
**Results from the Optimization of Switching Times and Pressures for**  
**the Pilot Valve**

| $P_{\text{solenoid}}$ | $P_{\text{pilot}}$ | $t_{\text{arm}}$ | $t_{\text{fire}}$ | $t_{\text{total}}$ | Remarks                     |
|-----------------------|--------------------|------------------|-------------------|--------------------|-----------------------------|
| 50 (N <sub>2</sub> )  | 50 (He)            | 230              | 40                | 270                | Works                       |
| 50 (N <sub>2</sub> )  | 50 (He)            | 70               | 40                | 110                | Works                       |
| 50 (N <sub>2</sub> )  | 50 (He)            | 23               | 40                | 63                 | Failed                      |
| 50 (N <sub>2</sub> )  | 50 (He)            | 70               | 18                | 88                 | Failed                      |
| 85 (N <sub>2</sub> )  | 90 (He)            | 32               | 18                | 50                 | Worked                      |
| 100 (N <sub>2</sub> ) | 100 (He)           | 23               | 18                | 41                 | Worked only 50% of the time |
| 100 (He)              | 100 (He)           | 22.7             | 23                | 45.3               | Worked 94% of the time      |
| 120 (He)              | 120 (He)           | 17               | 19.4              | 36.8               | 80% success                 |

Private communication with Dr. Harvey provided information that this research was the first to have the CSV switch this fast. Harvey suggested that a potential improvement of 20 to 30% could be achieved by re-boring the pilot valves. A reducer bushing which connects the pilot valves to the CSV is the main constriction for the gas flow when operating the CSV at this high of a speed. The CSV has not been re-bored at this time, but it is possible to drop the maximum transition time to 120-140 milliseconds.



## **PART 2: PRELIMINARY APPLICATIONS OF THE CONTROLLER SOFTWARE FOR HPLC SEPARATIONS**

**The most dangerous thing in the world is to try to leap a chasm in two jumps.**

**William Lloyd George**

**The only time you don't fail is the last time you try anything--and it works.**

**William Strong**

## **Section 6.7      Integration of the Expert System and HPLC Controller**

This section describes two different sets of experiments that were performed using the HPLC controller software. The first set of experiments demonstrates the first level of obtaining an intelligent automated HPLC instrument. Consultation with the expert system proposed several methods for the separation of the different samples. The proposed methods are then programmed into the HPLC instrument and the controller software ran the experiments. These experiments demonstrate the ability of the expert system to propose a method(s) for the separation of samples.

The second set of experiments demonstrates the ability of the software to perform column switching based upon the peak detection. Preliminary studies are shown that demonstrate the ability of the controller software to perform reproducible valve changing. These experiments do not demonstrate multidimensional chromatography, but do demonstrate the potential for multidimensional chromatography via the extension of the expert system (see Chapter 7) and the current HPLC controller software.

This section describes the first generation of expert system integration with the HPLC controller software. The recommendations of the expert system are manually programmed into the controller software and the controller program performs the experiment. Two additional validation studies are presented to show the integration of HPLC Advisor with the controller software. The validation studies incorporated only the column selection module of the expert system.

### ***Section 6.7.1      Determination of HPLC Controller Parameters***

#### **Section 6.7.1.1      Peak Detection Thresholds**

The controller software contains a peak detection routine to determine the elution of components. A peak determination occurs based upon two criteria. The first

criterion uses a seven point linear regression routine to determine the slope of the output from the detector. The slope of the linear regression is given by Equation 6.7.<sup>14</sup>

$$b = \frac{\sum_i \{(x_i - \bar{x})(y_i - \bar{y})\}}{\sum_i (x_i - \bar{x})^2} \quad \text{Equation 6.7}$$

The absolute value of the slope is not important; only the relative slope is important. Since the seven data points are taken at regular increments, the values for  $x_i$  can be given as integer values from 0 to 6. These values for  $x_i$  can be substituted into Equation 6.7, and the relative linear regression equation can be mathematically reduced to the form given in Equation 6.8. This form is computationally simpler and requires less CPU time to be calculated. The value of the calculated slope is compared to some user-defined slope threshold value. Once the calculated slope exceeds this threshold, the first criterion for peak detection has been met.

The second criterion that must be met is that the average of the seven data points must be larger than a second height threshold. This threshold insures that a noisy signal that may have therefore a slope greater than the slope threshold will not be erroneously counted as a peak.

$$b = 3(y_6 - y_0) + 2(y_5 - y_1) + (y_4 - y_2) \quad \text{Equation 6.8}$$

The complete algorithm for the determination of a peak requires three events. The first event is that the calculated slope and average is larger than the respective thresholds. This first event indicates the start of a peak. The second event requires that a negative slope is calculated which has a value less than the negative value of the slope threshold. This event occurs at, or slightly after, the crest of the peak has

occurred, and is used in conjunction with the third event to determine the end of the peak. The third event occurs when a negative slope has occurred (as determined by the second event) and the calculated slope is less than the threshold. The algorithm has been written such that the detection of a start of a peak is disabled until the complete sequence of events 2 and 3 have occurred. Without this ability, the algorithm would indicate a start of a peak for every point during the upslope of the peak. This scheme forces a start, crest, and end of the peak to occur. This approach is used also to provide additional flexibility for the HPLC controller software. The software is written to allow the user to switch columns and valves at the beginning or end of the elution of a component of the sample.

A series of experiments was conducted to determine appropriate threshold values for the slope and peak height. The sample contained benzaldehyde and biphenyl and the HPLC conditions used during the separation are given in Table 6.5. Two different data acquisition rates were used. The data for one set of experiments were collected at 1 data acquisition/second, and the second set was collected at 10 acquisitions/second. The resulting chromatograms are shown in Figures 6.11, 6.12A, and 6.12B. (The chromatogram taken at 10 acquisitions/second had to be made into two parts to allow importation into an Excel spreadsheet for data interpretation.)

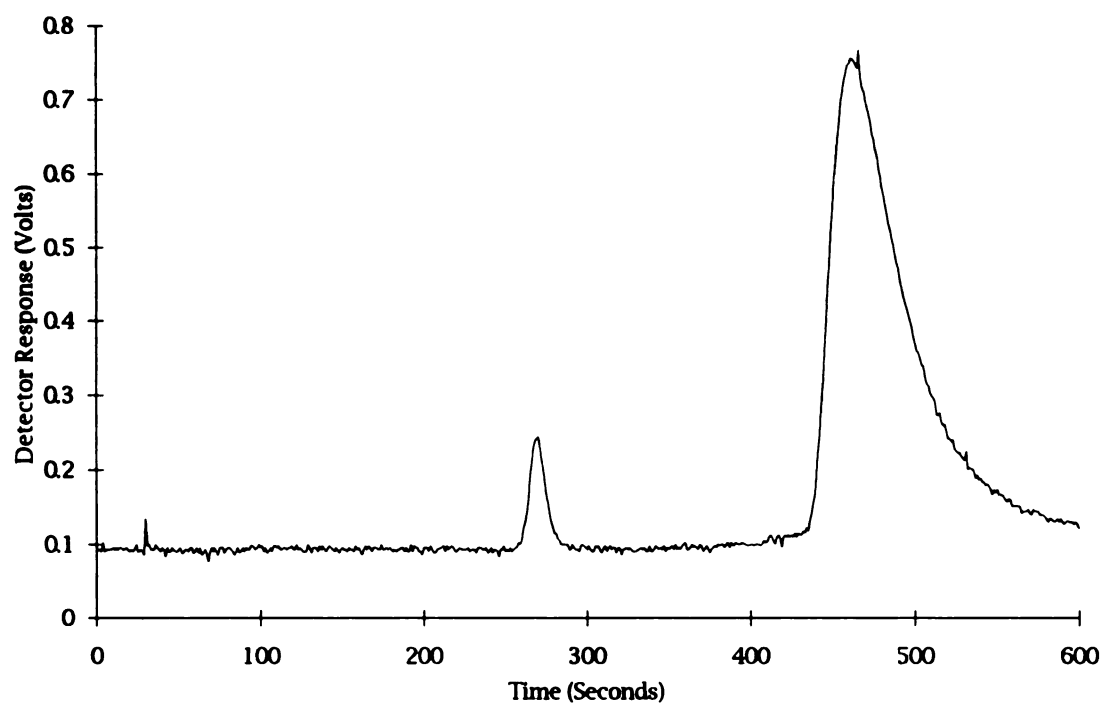
Figures 6.13 and 6.14 show plots of the calculated slopes versus time. It was noted that the peak detection algorithm was unable to reproducibly detect a peak for the experiments taken at the faster acquisition rates. Figure 6.13, for the acquisition rate of 1/second, shows that the slopes of the two peaks are sharp and distinctive. The baseline slopes are small and a slope of 0.005 is at least two times the slope of the baseline.

Figure 6.14 provides a similar plot of the slope for the data collected at 10 acquisitions/second. The first peak, which is a sharper peak, was correctly detected,

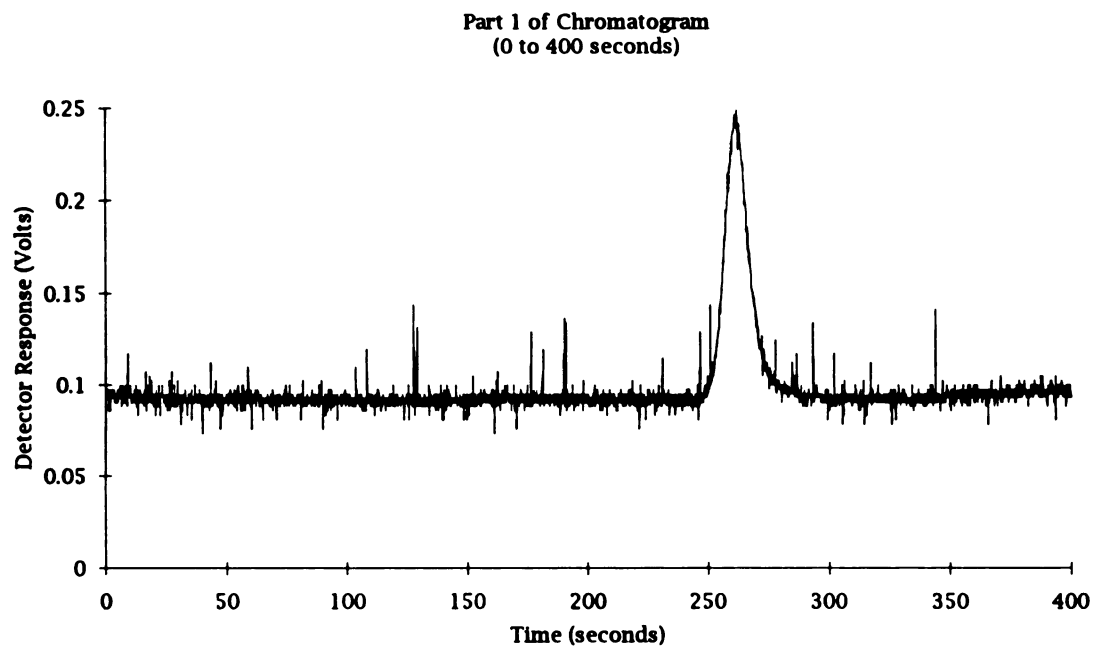
but broader peaks were not detected or not detected early in time. Figure 6.14 shows the slope from 400 seconds to the end of the experiment. In this case, the baseline is noisier and provides erroneous peak reporting. Since there are more points per second, Figure 6.15 shows an expansion from 440 to 450 seconds. This figure demonstrates that the acquisition rate is faster than the changes in the signal from the detector (the effect of the acquisition rate is described in more detail below). The effects of small changes in the voltage of the detector also cause difficulty in peak detection, as can be seen at the points indicated by the arrows. Here, the slope experiences a significant decrease in the slope, which may be below the threshold. For even broader or more gradual peaks of other data, the algorithm missed the peak all together. If it is desired to acquire data at rates faster than 0.1/second, the peak algorithm must be modified to either use more than seven data points in the regression or to average seven points and take the linear regression of seven average points. As discussed in the next section, an acquisition rate of 1/second is sufficient for most experiments and the above modifications are not needed.

**Table 6.5**  
**Experimental Parameters for the Analysis of Benzaldehyd**  
**and Biphenyl**

| Parameter                       | Value                   |
|---------------------------------|-------------------------|
| <b>Column</b>                   | Dupont Zorbax C18       |
| Length (cm)                     | 25                      |
| Diameter (mm)                   | 5.6                     |
| Particle Size ( $\mu\text{m}$ ) | 5                       |
| <b>Mobile Phase</b>             | 100% Methanol           |
| <b>Flow Rate (mL/min)</b>       | 0.50                    |
| <b>Detector</b>                 |                         |
| Type                            | Beckman Gold UV-Visible |
| Wavelength (nm)                 | 254                     |
| Range (AUFS)                    | 0.050                   |
| <b>Peak Thresholds</b>          |                         |
| Slope                           | 0.005                   |
| Peak Height                     | 0.100                   |



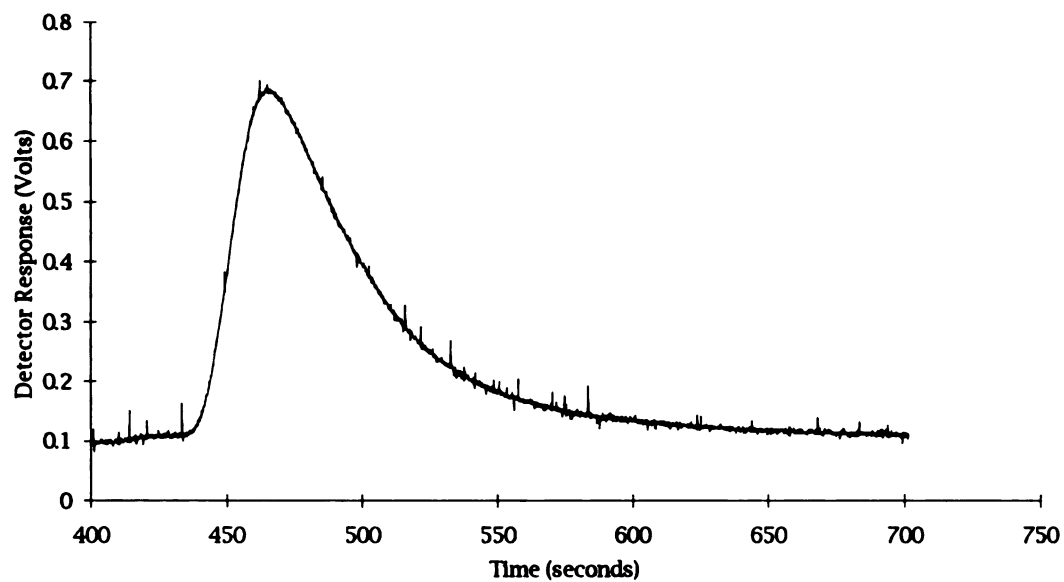
**Figure 6.11**  
**Biphenyl and Benzaldehyde Chromatogram**  
**Rate of 1 Acquisition/Second**



**Figure 6.12A**  
**Biphenyl and Benzaldehyde Chromatogram**  
**Rate of 10 Acquisitions/Second**

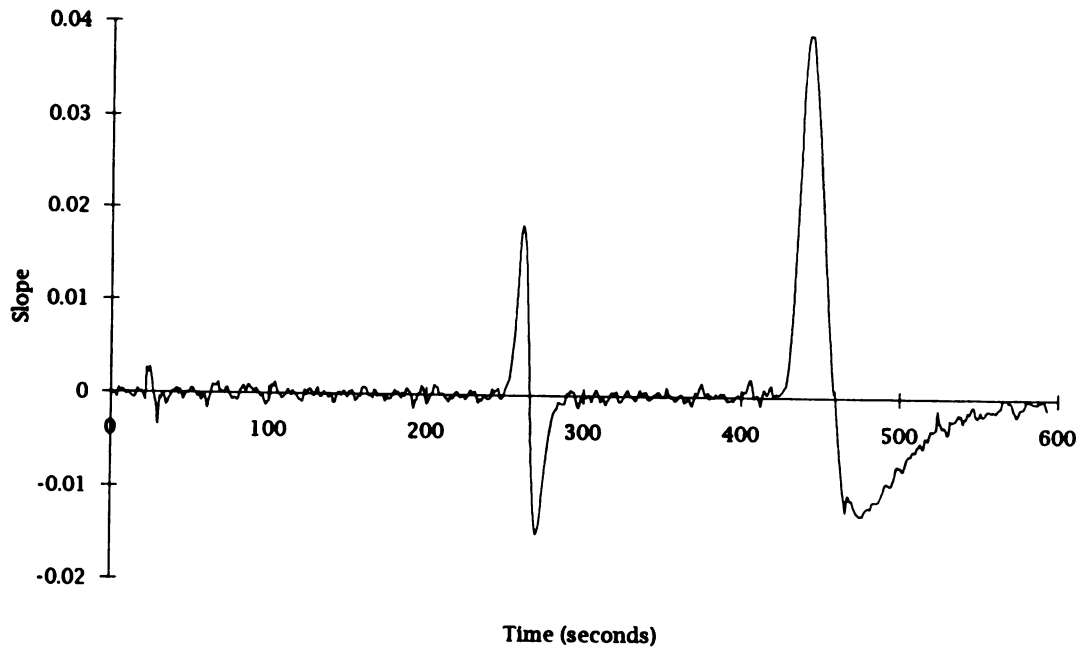


Part 2 of Chromatogram  
(400 - 700 seconds)

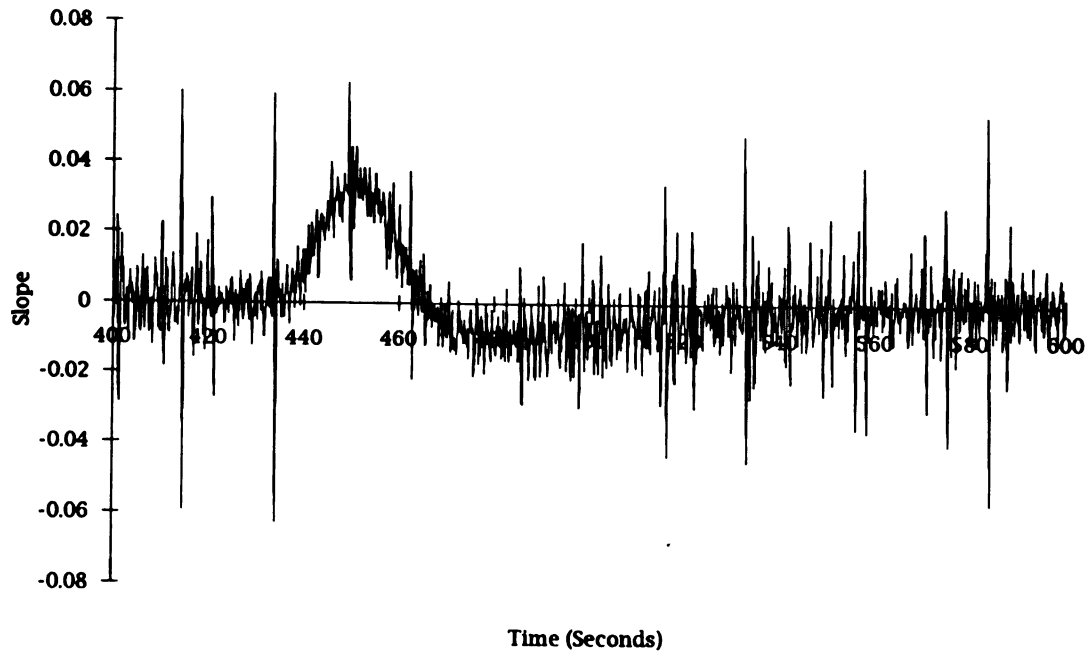


**Figure 6.12B**  
**Biphenyl and Benzaldehyde Chromatogram**  
**Rate of 10 Acquisitions/Second**

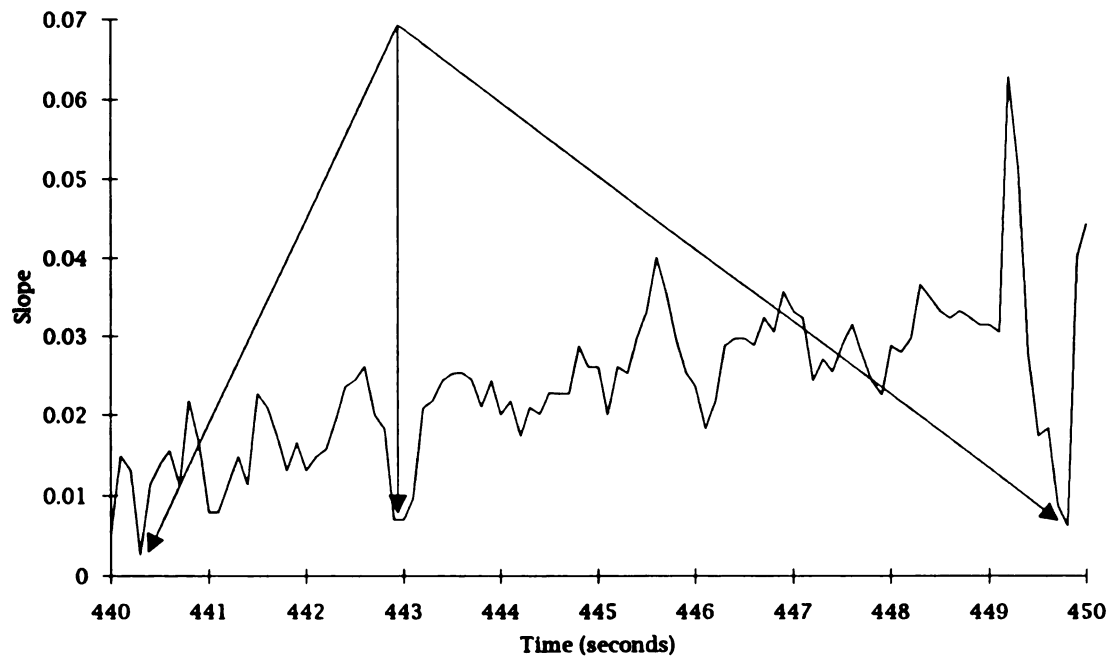
A second modification was made to the system after comparing the detector response and the digitized values stored by the computer. The Beckman System Gold UV-Visible Detector has an output of  $\pm 1$  Volt Full Scale, while the RTI 800/815 data acquisition board has a minimum full scale of  $\pm 5$  Volts full scale. This non-equality between the detector and RTI board caused an 80% decrease in the resolution of the analog to digital conversion. To improve the resolution requires that the full 12 bits of the RTI board be used. This improvement is achieved by taking the output from the detector and amplifying the output to 5 Volts through a LF351 Op-Amp. The schematic for the amplification circuit is given in Figure 6.16. The amplification, in addition to increasing the precision of the data, improves the peak detection routine since the slopes will be multiplied by a factor of 5 which allows the slopes of gradual changes to be amplified.



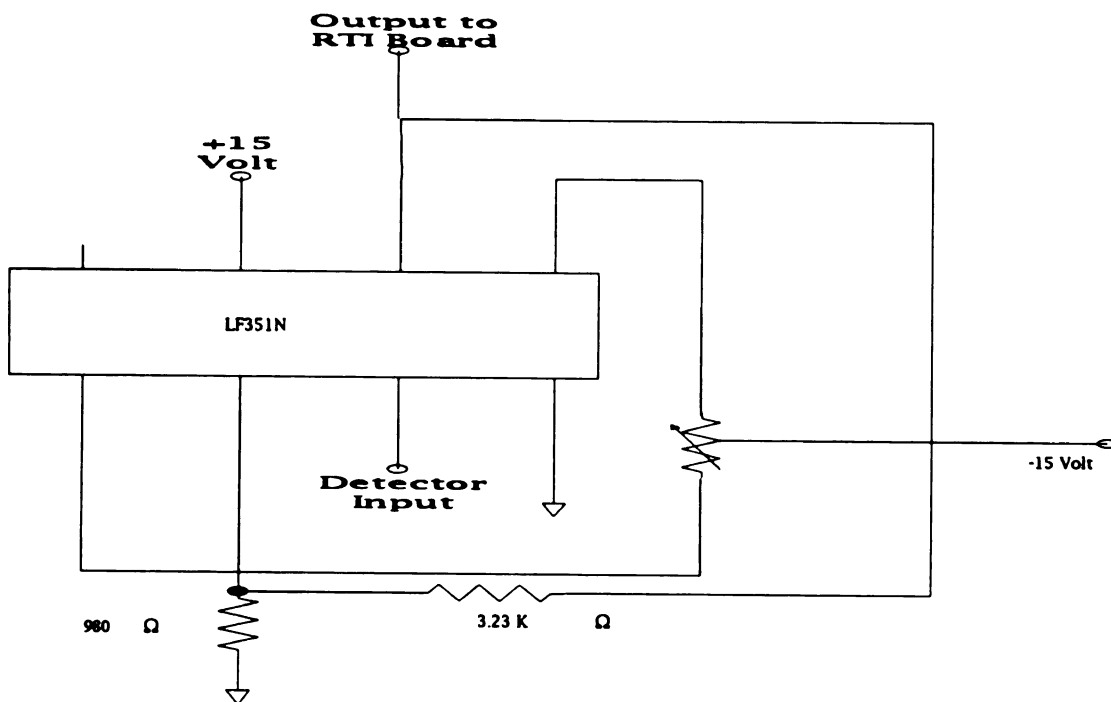
**Figure 6.13**  
**Plot of Slope Versus Time for**  
**Rate of 1 Acquisition/Second**



**Figure 6.14**  
**Plot of Slope Versus Time for**  
**Rate of 10 Acquisitions/Second**



**Figure 6.15**  
**Inset of the Slope for the**  
**Rate of 10 Acquisitions/Second**



**Figure 6.16**  
**Circuit to Provide a Gain of 5**  
**to the HPLC Detector Output**

### **Section 6.7.2      *Validation Study of Benzaldehyde, Biphenyl, and Naphthalene***

A sample containing  $1.31 \times 10^{-3}$  M Benzaldehyde,  $3.83 \times 10^{-3}$  M Naphthalene, and  $8.33 \times 10^{-4}$  M Biphenyl was made. The sample was described as aromatic molecules containing different substituents as part of the ring. The molecules were also described as having different polarities with a confidence of 80%. The 80 CNF was used to subjectively described that only one of the molecules, benzaldehyde, was polar while the other two species have similar polarities. The molecules were described as containing aldehydes, were small (less than 2000 g/mole), and were soluble in hexane and methanol/water. Finally, the expert system was allowed to determine that the species had a dipole moment and exhibited charge transfer characteristics.

The expert system recommended

- |                                                        |         |
|--------------------------------------------------------|---------|
| • Reversed-phase with an alkyl silane stationary phase | CNF 100 |
| • Reversed-phase with cyano packing material           | CNF 100 |
| • Normal-phase adsorption                              | CNF 97  |
| • Normal phase bonded with cyano packing material      | CNF 92  |
| • A specialty column for aromatics                     | CNF 80  |

The sample was then injected, and the data acquired using the HPLC controller software. The arrow indicates a shoulder that is difficult to see due to the digitization of the data. The shoulder should be the naphthalene, which elutes with approximately the same elution time as the biphenyl (refer to the recycle heartcut chromatography experiment earlier in this chapter). Unfortunately, the available columns in this laboratory were old and did not provide a very good separation. Different conditions and columns were used, but these columns also were old and unable to provide additional separations.

However, this data does show the utility of using the expert system in conjunction with the controller software to develop an automated, intelligent instrument. One additional validation study is presented below. While unable to separate the mixtures on the available columns, a separation on these species (vide infra) was demonstrated in Snyder and Kirkland.<sup>5</sup>

**Section 6.7.3      *Validation Study of Benzoic Acid, p-Aminobenzoic acid, 2,4-Dinitrophenol, Anthracene, and Pyrene***

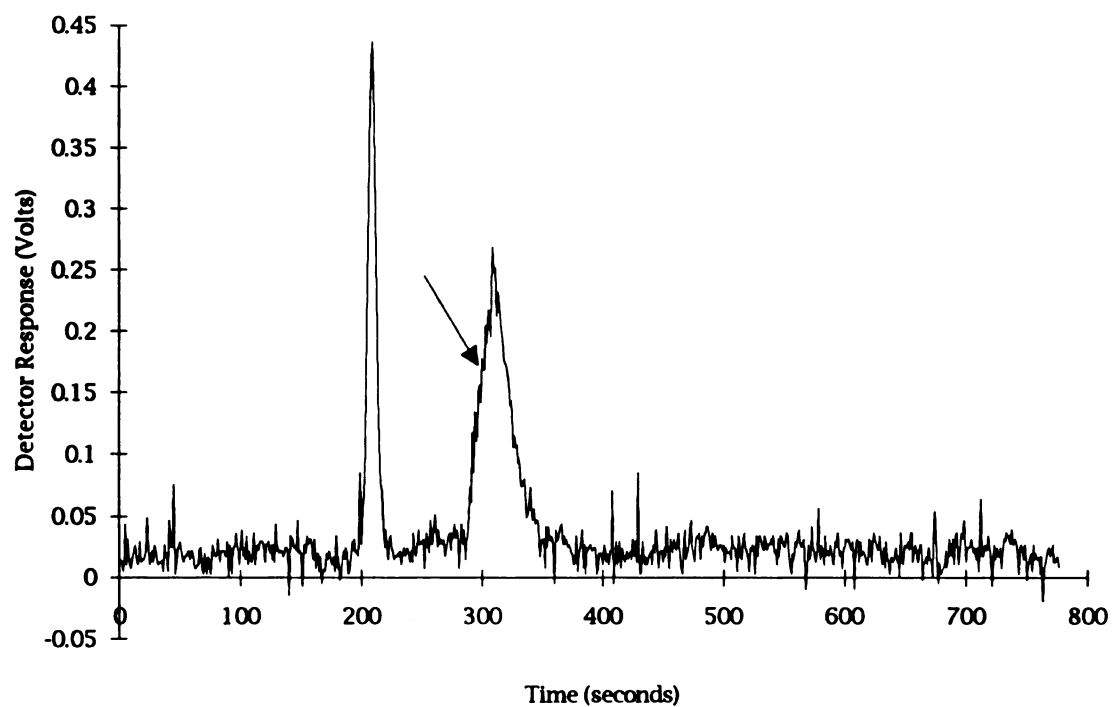
A validation study of the expert system was accomplished using a mixture of benzoic acid, p-aminobenzoic acid, 2,4-dinitrophenol, anthracene, and pyrene. Input for the expert system was:

The species are soluble in organic, aqueous/organic solvents. The species are small (MW<2000) and organic (acid/base) ionic species, but a pH of 4, CNF 80, would neutralize the two acids in the sample. The species were described as aromatic species that have different substituents (carboxylic acids, phenols, nitro groups, and amine groups) on the aromatic ring. The components in the sample are also described as having different polarities.

The expert system was able to predict dipole interactions based upon the types of functional groups present. In addition, the expert system also predicted that the species should form charge transfer complexes with the stationary phases based upon the presence of the aromatic ring with electron-withdrawing groups (e.g., the nitro groups).

The expert system predicted the use of reversed-phase cyano columns CNF 95, reversed-phase alkyl silane columns (e.g., C<sub>18</sub>) CNF 95, reversed-phase anion exchange chromatography CNF 80, and a specialty aromatic column CNF 80.





**Figure 6.17**  
**Chromatogram of Benzaldehyde, Naphthalene,**  
**and Biphenyl**

### **PART 3: CONCLUSIONS BASED ON THE AUTOMATION OF THE HPLC INSTRUMENT**

**I have but one lamp by which my feet are guided; and that is the lamp of  
experience. I know of no way of judging the future but by the past.**

**Patrick Henry**

## **Section 6.8 Conclusions of the Automation of the HPLC Instrument**

This chapter describes the software and hardware that was created to allow for the automation of the HPLC instrument. The software is fairly extensible and provides a user-friendly interface. The framework is in place for the next development of the instrumentation; that is, an intelligent automated instrument. The expert system can generate advice that can be stored in an ASCII file. A simple addition to the controller software can be made to read in this advice file, which would contain column and solvent advice. The operator may generate an LCMethod file for each valve configuration that will be used or the controller software can use previously written method files. The software can monitor the run and switch at the appropriate times. One limitation is that the injector valve is not under computer control which necessitates that the operator load and inject the sample.

The hardware provides a fast and reliable way to switch valves. The PPI still has several digital lines remaining that could be used to control chart recorders, autoinjectors, and so forth. Additional PPI chips can be added onto the prototype board to extend the number of control lines that could be available.

One limitation to the software is that it is written in QuickBasic. This limits the number of data points that can be taken to 32KB. One simple approach would be to stop the HPLC instrument, stop the data acquisition via the gate control, and save the 32KB of data. The program could then continue acquiring data since the DMA is set up for autoinitialization. A second solution would be to port the code to either the C program language or to use VisualBasic (Microsoft's Basic language for the Windows environment). C would provide faster control over the instrumentation while

**VisualBasic would allow the controller system to access other programs (e.g., the expert system, data manipulation software, or optimization programs).**

## LIST OF REFERENCES

## LIST OF REFERENCES

- 1 D. P. Webb and E. D. Salin, *Intelligent Instruments and Computers* 1991 (issued 1992), 9(5), 185-93.
- 2 A. P. Wade and S. R. Crouch, *Spectroscopy*, 1988, 3, 24.
- 3 S. R. Crouch, *Chem. Intell. Lab. System* 1990, 8, 259.
- 4 P. D. Wentzell, *Ph. D. Thesis* Michigan State University, East Lansing, MI, 1987.
- 5 L. R. Snyder and J. J. Kirkland, *Introduction to Modern Liquid Chromatography*, John Wiley and Sons, Inc., New York, 1979.
- 6 S. K. Ratanathanawongs, *Ph.D. Thesis*, Michigan State University, East Lansing, Michigan, 1986.
- 7 M. Martin, F. Verillon, C. Eon, and G. Guiochon, *J. Chromatogr.* 1976, 125, 17.
- 8 P. Wiegand, *Ph. D. Dissertation*, Michigan State University, East Lansing, Michigan, 1985
- 9 P. M. Wiegand and S. R. Crouch, *Talanta*, 1985, 32, 37.
- 10 L. C. Eggebrecht, *Interfacing to the IBM Personal Computer*, Howard W. Sams and Co., Inc., Indianapolis, IN 1989.
- 11 P. F. Goldsbrough, *Microcomputer Interfacing with the 8255 PPI Chip*, Howard W. Sams, and Co., Inc., Indianapolis, IN, 1980.
- 12 M. C. Harvey and S. D. Stearns, *Anal. Chem.* 1984, 56, 837.
- 13 M. C. Harvey, *Personal Communication*.
- 14 J. C. Miller and J. N. Miller, *Statistics for Analytical Chemistry*, Ellis Horwood, Ltd. New York, 1984, 90.

We  
exp

Ex

## Section

T

intelligen

intelligen

approac

mobile

particul

recomm

the val

conclus

## Section

from o

## **CHAPTER 7**

### **CONCLUSIONS AND FUTURE PROSPECTS**

**We know nothing of what will happen in future, but by the analogy of experience.**

**Abraham Lincoln**

**Experience is a comb that life gives you after you lose your hair.**

**Judith Stern**

#### **Section 7.1 Introduction**

The preceding chapters have described the development of a first generation, intelligent, automated, high performance liquid chromatography instrument. The intelligence aspect has been provided by an expert system that applies a first principle approach to provide a user with a recommendation for column type, detector, and mobile phase. The expert system also provides advice on the suitability of HPLC for a particular sample, sample pre-treatment suggestions, and information on the recommended method. Through software and hardware, total control over the HPLC, the valves, and the data acquisition system is achieved. This chapter summarizes the conclusions from the previous chapters and presents suggestions for future work.

#### **Section 7.2 Conclusions**

In Chapter 4, the expert system was presented. This expert system is different from other HPLC expert systems. The expert system described here uses first principles



to provide

chemical

principles

literature,

systems

pharmac

a method

recommen

unknown

the exper

advice o

makeup

based u

water s

addition

the only

and/or

technic

The ex

approa

a samp

the int

to rec

major

to provide advice on chromatographic methods. Recommendations are based on the chemical and physical interactions between solute and stationary phase. The first principles approach, which is distinctive from the HPLC expert systems described in the literature, allows the expert system to deal with any type of sample. Other expert systems provide advice for a small sub-set of the chromatographic field (e.g., pharmaceuticals) or place compounds into categories (e.g., phthalates) and recommend a method for phthalates. While these types of expert systems provide very detailed recommendations, they are limited in their ability to provide recommendations for unknown samples or samples outside their domain. The goal for the development of the expert system described in this dissertation was to create a tool that could provide advice on almost any problem—provided the user knows something about the chemical makeup or the physical properties of the sample. Often educated guesses can be made based upon where the sample originated. For example, a river water sample would be water soluble and may contain inorganic ions or acid or bases. A database and additional rules could be developed that would provide assistance for samples in which the only known information is the source of the samples.

This expert system does not provide detailed advice on the more esoteric and/or specific methods such as chiral separations and affinity chromatography. These techniques typically use packing materials that are selective for specific compounds. The expert system described here can provide initial advice, but when using these approaches the user must find the column that will separate the specific components in a sample.

Chapter 5 presents the detector and mobile phase modules. These modules use the information obtained during the column recommendation phase and uses this data to recommend appropriate detector(s) and mobile phases. HPLC Advisor uses two major criteria to select a detector. First, the species must be detectable by the detector

and the concentration range of the solutes must be within the linear operating range of the detector. After recommendation of a column and detector, HPLC Advisor suggests a mobile phase which is compatible with both column and detector. Mobile phase modifiers are recommended based upon solute-solvent interactions. Validation studies were shown that demonstrates the ability of HPLC Advisor to provide an appropriate column, detector, and a starting mobile phase,.

A second aspect unique to this expert system is the application of confidence factors. While confidence factors themselves are not original, the HPLC expert systems in the literature do not use confidence factors. The expert system described here uses confidence factors in several unique ways. Confidence factors provide a means of ranking different columns which may be valuable to the user. Providing several possible methods allows the user to select among the possibilities. The user has a higher level of trust in the proposed methods since each has a numerical descriptor of their reliability.

Self-calculating confidence factors (CNF) allow the expert system to apply intelligence to solve the problem, and the use of self-calculating CNF is predominately illustrated in the case of selecting size as a separation mechanism. The expert system is able to determine if the components have a large enough size differential to warrant size being considered and assigns a confidence factor to reflect the importance of size. The expert system also weights the size differential with the presence or absence of competing mechanisms.

A third type of CNF used by HPLC Advisor are the concentration CNF values that are determined for the different detector. These CNF values are more fundamental in nature and minimize the empirical or subjective nature of CNF. The concentration CNF values are directly related to fundamental relationships between concentration and response. In addition, HPLC Advisor uses the particular characteristics of the available

(or g

is in

recc

dea

self

exp

opt

rec

so

da

ch

el

ar

b

a

c

a

n

r

(or generic values) of the detector to tailor the recommendation for the equipment that is in the laboratory of the user.

Validation studies were presented to show that the expert system can provide recommendations for different types of samples. The expert system has some ability to deal with uncertainties by asking further questions to elucidate the problem or by using self-determining rules to provide an answer based upon information available to the expert system. Validation studies have shown that the expert system provides the optimum method as one of the first few proposed solutions. Additional solutions are recommended which provide logical alternatives.

Chapter 6 presents preliminary work on automating the HPLC instrument. The software and hardware are capable of programming methods into the HPLC, acquiring data from multiple detectors, and controlling three different valves. The software can change the valve configurations based on a time criterion or the occurrence of a sample eluting. These capabilities allow the software to perform complex column switching and re-programming of the solvent delivery system for every valve change.

Preliminary results are presented that show the advice from the expert system being performed on the automated instrument. The controller software selects the appropriate column and programs the HPLC. The software monitors the run, collects data, and presents an audited report at the conclusion of the run.

Preliminary results showing the ability to switch columns based on the elution of a peak are also presented. The studies show that the software can provide accurate and reproducible switching. These results demonstrate that the software can be used for the analysis of complex samples that may require column switching.

Sec

be f

con

the

be

the

fu

co

co

e

s

w

-

c

## Section 7.3     Future Work

There are several aspects of both the automation and the expert system that can be further developed. The first objective is to have the expert system on-line with the controller software. In this fashion, a true intelligent automated system, as described at the beginning of Chapter 4, will be possible. Currently, the expert system results must be entered manually into the controller software. An on-line system will be able to feed the results directly to the software, which can provide feedback to the expert system for further optimization.

The knowledge base, while expanded significantly during this research, still contains areas in which further expansion is desirable. Solvent optimization protocols could be added to assist the user in selecting optimum mobile phases. The approach envisioned would not perform the optimization (a technique better suited to specific software created for that purpose, such as DryLab or factor analysis programs), but would assist the user in providing the correct responses to an optimization program. The rules that should be added would propose solvents based on detector and column compatibilities. The solvents would then be used by the external optimization programs.

The use of databases in this expert system can be extended. Additional databases could be incorporated to provide information on chemical and physical properties (e.g., electronegativities) of the analyte(s).

With the development of Microsoft Windows and other graphical user interface systems (GUI's), many new options have been created for the knowledge engineer. For example, *Level5 Object*, version 2.5, (Information Builders, Inc., New York, NY)<sup>1</sup> obtains memory management and graphical abilities through API (Application Program Interface) calls to Windows 3.x. This allows knowledge engineers to use effectively icons, display screens, and other tools to create visually appealing and intuitive

interfaces. One exciting capability is the ability to use the Dynamic Data Exchange (DDE) and Object Linking and Embedding (OLE) capabilities of Windows. This will allow seamless integration of data and communication with other programs. In this fashion, expert systems will be able to access any Windows spreadsheets or data bases. Once data are exchanged through DDE, the expert system can use OLE for data manipulation by data acquisition and control programs, statistical packages and so forth. With DDE and OLE, the expert system can be extended beyond its intrinsic capabilities by accessing the features of other programs. As mentioned in previous chapters, VP-Expert is deficient in memory management and numerical capabilities. Many of the features of VP-Expert version 3.0 had to be removed or disabled to maintain an operating program. The detector module of HPLC Advisor uses much more intensive "number crunching" than for which VP-Expert seems to be designed. Further development of the expert system should exploit the strengths of another expert system shell (or a newer version of VP-Expert when/if it becomes available). Developments of graphical operating systems demonstrate the advantages of moving to a non-DOS based expert system shell.

Final development of the expert system would introduce an additional "master" module to provide advice on multidimensional and column switching chromatography. The module would have two specific goals: (1) determine fractions or classes of compounds in a complex sample and (2) determine incompatibility problems between different columns, solutes, and mobile phases. The master module would propose ways to sub-divide the sample into fractions and then call the current expert system (slave) to develop a method(s) to separate the fractions. The master module would repetitively call the slave expert system to propose methods to resolve components within the fractions. The master module would apply compatibility rules to develop a master separation method. This would complete the process of creating an intelligent, automated, (pseudo)multidimensional HPLC instrument.



## LIST OF REFERENCES

## LIST OF REFERENCES

- <sup>1</sup> Don Barker, *PC AI*, 1992, *July/August*, 44-7.

## APPENDIX

## APPENDIX A

### Flow Charts for the Determination of an Appropriate Column(s)

The recommendation of a column involves three major steps. The first step determines the separation property. The separation property indicates the possible chemical and physical differences between the solutes that may provide a means to separate the solutes. The next step is to determine the type of chromatographic separation (i.e., normal-phase and reverse-phase). Finally, the expert system determines the possible packing materials (e.g., amine or silica functional groups). The packing material is based on the chemical and physical properties of the solutes.

This appendix provides flow charts that describe the recommendation of an appropriate column. The logic paths used by the expert system (HPLC Advisor) are quite complex and inter-related. In addition, HPLC Advisor provides many rules that assist the user in selecting an appropriate response. Due to the complexity of the inference mechanisms, a simplified view of the logic paths are presented. Some of the inference paths are presented in tabular form instead.

Figure A.1 provides a simplified flow chart for the determination of separation property. Information for ionic separations are not included in this figure. Figures 4.6a and 4.6b (Chapter 4) contain detailed flow charts that describe the assignment of

columns for ionic separations. As mentioned above, additional rules to predict or assist the user in answering the questions contained in this flow chart.

Figures A.2 and A.3 show the logic used to assign normal-phase adsorption modes. This logic is similar to the rules used in assigning reverse-phase columns. Whenever HPLC Advisor determines that SEPARATION PROPERTY equals NEUTRAL, reverse-phase bonded chromatography is assigned. Tables A.1 and A.2 list the different conditions that are considered when assigning size exclusion (SEC) as a mode for normal-phase (NP) and reverse-phase (RP) separations. HPLC Advisor recommends either normal-phase adsorption (NP Ads) or reverse phase bonded (RP Bond) chromatography as a possible alternative to size exclusion. The confidence factors (CNF) for both SEC and the alternative are given in Tables A.1 and A.2.

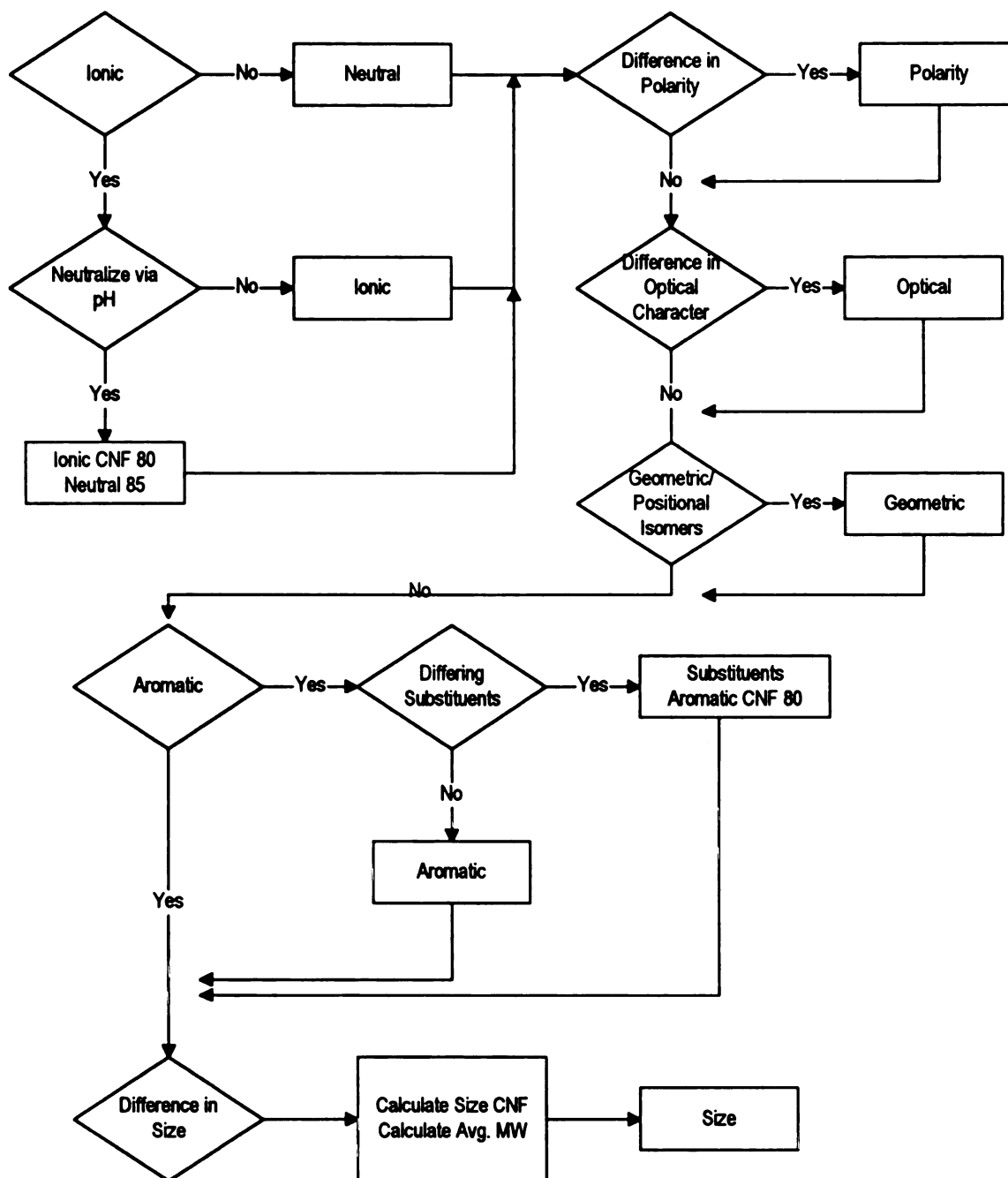
Figure A.4 is a flow chart of the logic used to determine the appropriate packing material. Not detailed in this figure is the recommendation of ionic packing materials (i.e., cation or anion exchange resins). These are detailed in Chapter 4. Adsorption chromatography packing materials (silica or alumina) are recommended equally provided that there are no solutes which have excessive binding energy with the packing material (e.g., carboxylic acids on alumina).

**Table A.1**  
**Criteria Used to Recommend Normal-Phase**  
**Size Exclusion Chromatography**

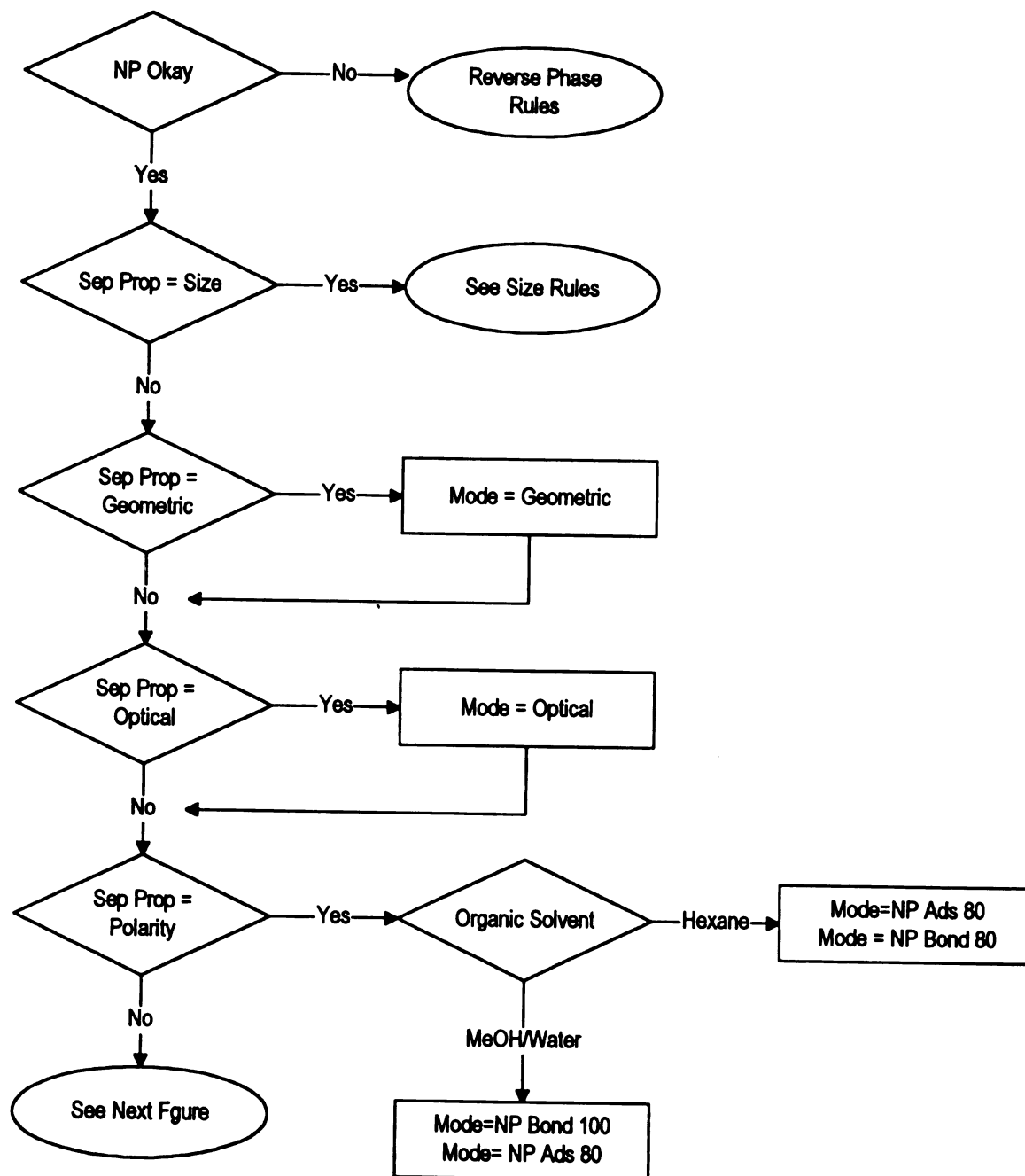
| <b>Molecular Weight / MW Distribution</b> | <b>NP SEC CNF</b> | <b>NP Ads CNF</b> |
|-------------------------------------------|-------------------|-------------------|
| > 5000 AND Wide Distribution              | 100               | 60                |
| >5000 OR Wide Distribution                | 90                | 70                |
| 2000-5000 AND Wide Distribution           | 85                | 80                |
| 2000-5000 OR Narrow Distribution          | 70                | 80                |
| <2000 AND Wide Distribution               | 80                | 85                |
| <2000 AND Narrow Distribution             | 50                | 100               |

**Table A.2**  
**Criteria Used to Recommend Reverse-Phase**  
**Size Exclusion Chromatography**

| <b>Molecular Weight / MW Distribution</b> | <b>RP SEC CNF</b> | <b>RP Bond CNF</b> |
|-------------------------------------------|-------------------|--------------------|
| > 5000 AND Wide Distribution              | 100               | 70                 |
| >5000 OR Wide Distribution                | 90                | 75                 |
| 2000-5000 AND Wide Distribution           | 80                | 80                 |
| 2000-5000 OR Narrow Distribution          | 70                | 90                 |
| <2000 AND Wide Distribution               | 80                | 85                 |
| <2000 AND Narrow Distribution             | 50                | 100                |

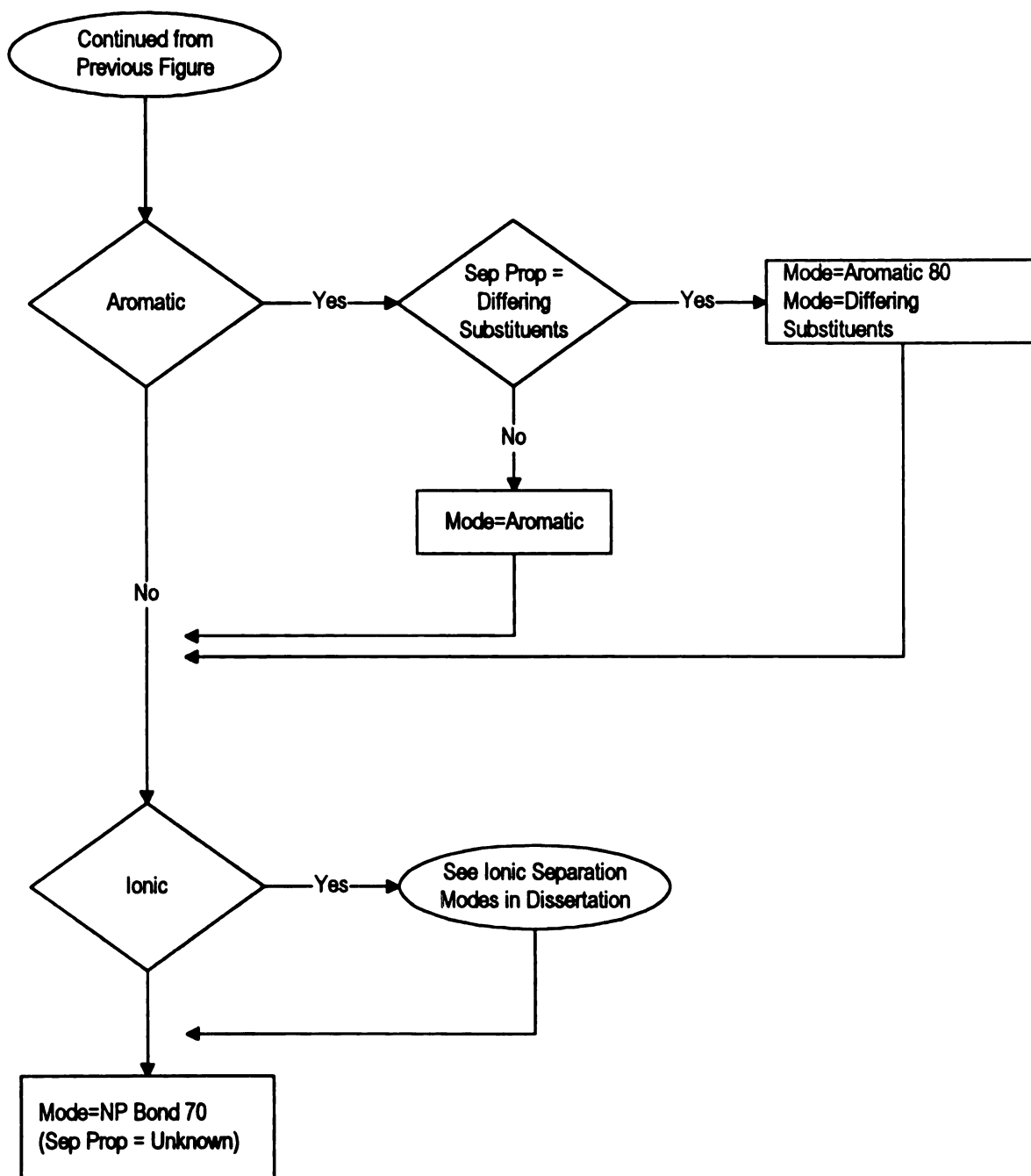


**Figure A.1**  
**Flow Chart for the Determination of Separation Property**

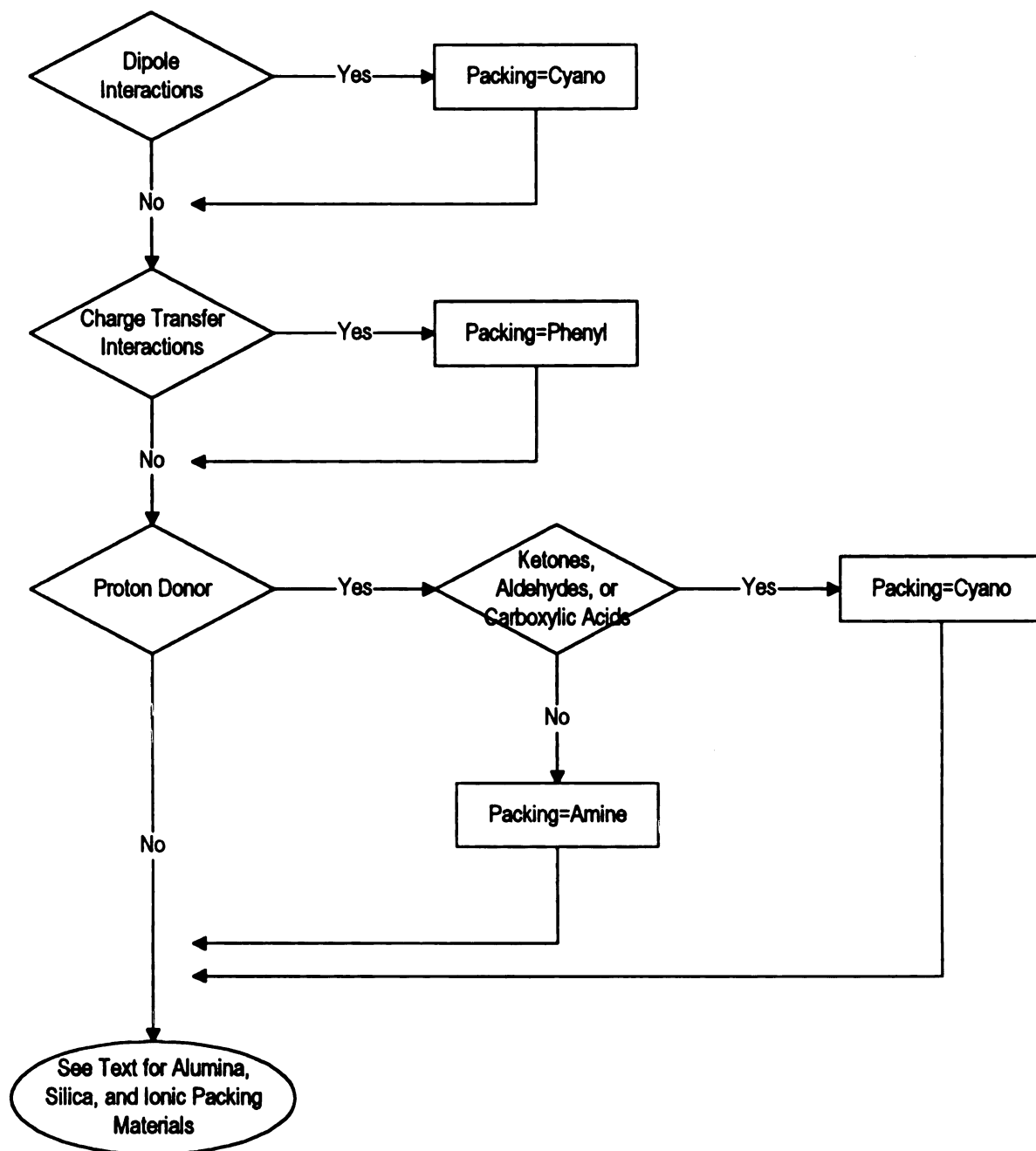


**Figure A.2**  
**Flow Chart for the Determination of Mode for**  
**Normal Phase Adsorption (Part 1)**





**Figure A.3**  
**Flow Chart for the Determination of Mode for**  
**Normal Phase Adsorption (Part 2)**



**Figure A.4**  
**Flow Chart for the Recommendation of Packing Material**

MICHIGAN STATE UNIV. LIBRARIES



31293010192320