



# City Research Online

## City St George's, University of London

**Citation:** Pooler, G. W. (1986). Studies on drug molecules by theoretical methods, NMR and computer graphics. (Unpublished Doctoral thesis, The City University)

This is the accepted version of the paper.

This version of the publication may differ from the final published version. To cite this item please consult the publisher's version.

**Permanent repository link:** <https://openaccess.city.ac.uk/id/eprint/35455/>

**Copyright and Reuse:** Copyright and Moral Rights remain with the author(s) and/or copyright holders. Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge, unless otherwise indicated, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way. For full details of reuse please refer to [City Research Online policy](#).

## Contents

|       |                                                         |   |
|-------|---------------------------------------------------------|---|
| 1     | Preamble                                                | 3 |
| 1.1   | Introduction                                            | 3 |
| 1.2   | Structure of solvent molecular dynamics                 | 3 |
| 1.2.1 | The Lennard-Jones                                       | 3 |
| 1.2.2 | Interaction between ions, and electrostatic potentials  | 3 |
| 1.3   | Structure of some crystalline systems                   | 3 |
| 1.4   | Simulation of systems of many particles in solution     | 3 |
| 1.5   | Studies on drug molecules by theoretical methods,       | 3 |
| 1.5.1 | Simulation of systems of many particles in solution     | 3 |
| 1.5.2 | Summary of the simulation of many particles in solution | 3 |
| 2     | NMR and computer graphics.                              | 3 |
| 2.1   | Introduction                                            | 3 |
| 2.2   | Summary and                                             | 3 |
| 2.3   | Graham William Pooler                                   | 3 |
| 2.4   | Submitted for the degree of PhD                         | 3 |
| 2.5   | The Department of Physics, The City University.         | 3 |
| 2.6   | September, 1986.                                        | 3 |
| 3     | The simulation of many particles in solution            | 3 |
| 3.1   | Introduction                                            | 3 |
| 3.2   | Structure of solvent molecular dynamics                 | 3 |
| 3.2.1 | The Lennard-Jones                                       | 3 |
| 3.2.2 | Interaction between ions, and electrostatic potentials  | 3 |
| 3.3   | Structure of some crystalline systems                   | 3 |
| 3.4   | Simulation of systems of many particles in solution     | 3 |
| 3.5   | Studies on drug molecules by theoretical methods,       | 3 |
| 3.5.1 | Simulation of systems of many particles in solution     | 3 |
| 3.5.2 | Summary of the simulation of many particles in solution | 3 |
| 4     | NMR and computer graphics.                              | 3 |
| 4.1   | Introduction                                            | 3 |
| 4.2   | Summary and                                             | 3 |
| 4.3   | Graham William Pooler                                   | 3 |
| 4.4   | Submitted for the degree of PhD                         | 3 |
| 4.5   | The Department of Physics, The City University.         | 3 |
| 4.6   | September, 1986.                                        | 3 |



## IMAGING SERVICES NORTH

Boston Spa, Wetherby

West Yorkshire, LS23 7BQ

[www.bl.uk](http://www.bl.uk)

BEST COPY AVAILABLE.

VARIABLE PRINT QUALITY

## Contents.

|         |                                                                             |     |
|---------|-----------------------------------------------------------------------------|-----|
| 1       | Preamble                                                                    | 1   |
| 1.1     | Introduction                                                                | 1   |
| 1.2     | Receptors of unknown molecular structure                                    | 2   |
| 1.2.1   | The GABA <sub>A</sub> system                                                | 4   |
| 1.2.2   | Interaction between GABA <sub>A</sub> and benzodiazepine receptors          | 6   |
| 1.3     | Receptors of known molecular structure                                      | 9   |
| 1.4     | Conformational analysis of small molecules in solution                      | 9   |
| 1.4.1   | Experimental methods                                                        | 9   |
| 1.4.2   | Theoretical methods                                                         | 11  |
| 1.5     | Summary of the main aims of this work                                       | 12  |
| 2       | Preliminaries: theoretical calculations on small molecules in the gas-phase | 13  |
| 2.1     | Introduction                                                                | 13  |
| 2.2     | Survey and testing of theoretical methods                                   | 14  |
| 2.3     | Geometry optimisation                                                       | 27  |
| 2.3.1   | The effect of geometry optimisation on physical properties                  | 27  |
| 2.3.2   | Possible short cuts with geometry optimisation                              | 30  |
| 2.3.3   | Some possible pitfalls with geometry optimisation                           | 31  |
| 2.4     | Gas-phase calculations on GABA                                              | 32  |
| 2.5     | Gas-phase calculations on BIC, HBIC and MeBIC                               | 36  |
| 2.6     | Summary                                                                     | 41  |
|         | <u>Part 1: receptors of unknown molecular structure</u>                     | 43  |
| 3       | The conformations of GABA in solution                                       | 43  |
| 3.1     | Introduction                                                                | 43  |
| 3.2     | Spectral and dipole moment studies on GABA solution conformations           | 45  |
| 3.2.1   | Previous work on GABA conformational preferences                            | 45  |
| 3.2.2   | Variable-temperature study on the flexibility of GABA in solution           | 49  |
| 3.3     | Theoretical calculation of GABA solution conformations                      | 51  |
| 3.3.1   | The Supermolecule model                                                     | 51  |
| 3.3.2   | The SOLVEFF model                                                           | 56  |
| 3.3.3   | A Hybrid approach                                                           | 65  |
| 3.3.4   | Simulation methods                                                          | 69  |
| 3.4     | Conclusions                                                                 | 69  |
| 4       | The conformations of BIC, MeBIC and HBIC in solution                        | 72  |
| 4.1     | Introduction                                                                | 72  |
| 4.2     | The energy profile of BIC in solution                                       | 76  |
| 4.2.1   | Evidence for multiple BIC low-energy conformations in solution              | 77  |
| 4.2.1.1 | Theoretical calculation of averaged H1/H9 coupling constants for BIC        | 77  |
| 4.2.1.2 | Theoretical calculation of averaged chemical shifts for BIC                 | 83  |
| 4.2.1.3 | NOE spectroscopy on BIC                                                     | 89  |
| 4.2.2   | Barrier to interconversion between $\theta_1$ minima                        | 89  |
| 4.2.3   | The N-ring conformation ( $\theta_2$ )                                      | 93  |
| 4.3     | The energy profile of MeBIC in solution ( $\theta_1$ )                      | 95  |
| 4.3.1   | Evidence for only one minimum in solution                                   | 95  |
| 4.3.2   | Significance of minimum                                                     | 103 |

|         |                                                                                         |     |
|---------|-----------------------------------------------------------------------------------------|-----|
| 4.3.3   | The N-ring conformation (02)                                                            | 105 |
| 4.4     | HBIC - evidence for just one conformation in solution (01)                              | 106 |
| 4.5     | Experimental details                                                                    | 106 |
| 4.6     | Summary and discussion                                                                  | 109 |
| 5       | Structural comparisons between GABA, BIC and semi-rigid GABA analogues                  | 113 |
| 5.1     | Introduction                                                                            | 113 |
| 5.2     | Previous structural comparisons of GABA, BIC and GABA agonists                          | 113 |
| 5.3     | Distinction between agonist and antagonist structural requirements                      | 118 |
| 5.4     | Flexibility of the GABA <sub>A</sub> receptor                                           | 123 |
| 5.5     | Possible contributory reasons for BIC being an antagonist                               | 123 |
| 5.6     | Summary                                                                                 | 124 |
| 6       | Overview of Part 1                                                                      | 126 |
|         | <u>Part 2 - receptors of known molecular structure</u>                                  | 133 |
| 7       | Development and applications of IMDAC - Interactive Molecular Display And Calculation   | 133 |
| 7.1     | Introduction                                                                            | 133 |
| 7.2     | The molecular modelling system - description of the main features of IMDAC              | 139 |
| 7.2.1   | The basic system                                                                        | 139 |
| 7.2.2   | Visualising the active-site and other key features of a receptor molecule               | 141 |
| 7.2.2.1 | CLEFT (OP33)                                                                            | 141 |
| 7.2.2.2 | SURF - a surface display routine (OP34)                                                 | 143 |
| 7.2.2.3 | Close Contacts (CC - OP22)                                                              | 143 |
| 7.2.2.4 | Highlighting of key atoms and residues                                                  | 145 |
| 7.2.2.5 | SITE - isolation of active-site residues (OP30)                                         | 147 |
| 7.2.2.6 | Pseudo space-filling model (OP12)                                                       | 147 |
| 7.2.2.7 | Stereo and perspective models (OP27 and OP41)                                           | 150 |
| 7.2.3   | Docking of drugs to receptors (OP42)                                                    | 150 |
| 7.2.4   | Molecular editing (EditMol, OP28)                                                       | 152 |
| 7.2.4.1 | Deletion of molecular fragments (ED5 and ED6)                                           | 153 |
| 7.2.4.2 | Automatic addition of hydrogen atoms (ED9)                                              | 155 |
| 7.2.4.3 | Inversion of a chiral centre (ED10)                                                     | 157 |
| 7.2.4.4 | Split/concantonate molecules (ED7 and ED8)                                              | 157 |
| 7.2.4.5 | 'Help' information for atom addition                                                    | 157 |
| 7.2.5   | A frame of reference for rotation, translation and zoom transformations                 | 158 |
| 7.2.6   | Energy calculations (OP25, 26 and 40)                                                   | 160 |
| 7.2.7   | Interfaces to other programmes                                                          | 162 |
| 7.2.8   | Features specifically designed for overcoming the limitations of the available hardware | 164 |
| 7.2.9   | Producing hard-copy output                                                              | 166 |
| 7.3     | Applications of IMDAC                                                                   | 167 |
| 7.3.1   | Phospholipase-A2 - an enzyme for which only isolated receptor coordinates are known     | 167 |
| 7.3.2   | Thermolysin - an enzyme for which several drug/receptor structures are known            | 171 |
| 7.4     | Overview of Part 2                                                                      | 175 |

|      |                                                                                 |     |
|------|---------------------------------------------------------------------------------|-----|
|      | <u>Appendices</u>                                                               | 177 |
| A1   | Anomalous gaps in GABA $x_T$ distributions                                      | 177 |
| A2   | Additional NMR and mass spectral data                                           | 180 |
| A3   | Simultaneous equations for deriving BIC conformer populations                   | 207 |
| A4   | Major changes at ULCC during the course of this work                            | 208 |
| A5   | List of graphics installations (Univ. of North Carolina)                        | 209 |
| A6   | Programmes written during the course of this work                               | 214 |
| A6.1 | MNDODP (dual processor)                                                         | 214 |
| A6.2 | MM2PRE, MM2POST and FXY                                                         | 216 |
| A6.3 | JVIC and DELT (BIC coupling constants and chemical shifts)                      | 221 |
| A6.4 | Programmes for interpolating GABA conformational-surfaces                       | 223 |
| A6.5 | Principle IMDAC subroutines                                                     | 229 |
| A7   | Routines in MOLEC5 which required major modification for efficient use in IMDAC | 276 |
|      | References                                                                      | 280 |

#### List of Tables.

|      |                                                                                            |     |
|------|--------------------------------------------------------------------------------------------|-----|
| 2.1  | Comparison of MNDO, CNDO, MM2, ab initio and experimental conformations and dipole moments | 15  |
| 2.2  | Comparison of MNDO, CNDO, MM2, ab initio and experimental conformations and barriers       | 16  |
| 2.3  | Comparison of theoretical and experimental geometries                                      | 18  |
| 4.1  | Comparison of observed and calculated coupling constants for BIC                           | 80  |
| 4.2  | BIC proton coupling constants                                                              | 82  |
| 4.3  | MFC and ISO chemical shifts                                                                | 84  |
| 4.4  | Comparison of observed and calculated chemical shifts for BIC                              | 88  |
| 4.5  | Benzene-ring shielding contribution for H5, H1 H5' and H9 in BIC                           | 92  |
| 4.6  | Observed NOE effects for BIC and MeBIC                                                     | 92  |
| 4.7  | N-ring proton coupling constants for BIC at 296K in acetone                                | 94  |
| 4.8  | Counterion effects on MeBIC proton chemical shifts                                         | 96  |
| 4.9  | Effect of temperature on MeBIC chemical shifts                                             | 97  |
| 4.10 | Effect of temperature on HBIC chemical shifts                                              | 97  |
| 5.1  | Potencies of key GABA analogues                                                            | 115 |
| 7.1  | Parameters for 6-12 potentials                                                             | 156 |

#### List of illustrations.

|     |                                                                          |    |
|-----|--------------------------------------------------------------------------|----|
| 1.1 | GABA and BIC molecular structures                                        | 3  |
| 1.2 | Schematic representation of the GABA <sub>A</sub> supramolecular complex | 7  |
| 1.3 | Structural comparisons between GABA and BDZ's                            | 8  |
| 2.1 | Comparison of energy-surfaces calculated by various methods              | 15 |
| 2.2 | Isolated GABA energy-surface (standard geometry)                         | 33 |
| 2.3 | GABA geometryOptimised energy-surface (partial)                          | 35 |

|      |                                                                                                |      |
|------|------------------------------------------------------------------------------------------------|------|
| 2.4  | BIC and MeBIC energy-surfaces (MM2)                                                            | 38   |
| 2.5  | BIC energy-surface (MM1 and MNDO)                                                              | 39   |
| 2.6  | Truncated BIC molecules used in MNDO calculations                                              | 40   |
| 3.1  | GABA molecular structure                                                                       | 44   |
| 3.2  | GABA $x_T$ distributions by various methods                                                    | 47   |
| 3.3  | MNR spectra of GABA                                                                            | 49   |
| 3.4  | Comparison of GABA and 3-bromo-1-phenylpropane spectra                                         | 50   |
| 3.5  | GABA Supermolecule energy-surfaces                                                             | 53   |
| 3.6  | Water molecule orientation parameters                                                          | 54   |
| 3.7  | GABA $x_T$ distributions by SOLVEFF and Hybrid methods                                         | 60   |
| 3.8  | GABA SOLVEFF surfaces                                                                          | 61   |
| 3.9  | Hybrid GABA energy-surfaces                                                                    | 67   |
| 4.1  | BIC molecular structure                                                                        | 73   |
| 4.2  | Temperature-dependence of BIC chemical shifts                                                  | 78   |
| 4.3  | Vicinal proton coupling constants for morpholine                                               | 81   |
| 4.4  | MEC and ISO molecular structures                                                               | 84   |
| 4.5  | Theoretical shielding contributions for H6' and H8                                             | 85   |
| 4.6  | Schematic representation of the derivation of energy-differences for $\theta_1$ conformations  | 87   |
| 4.7  | BIC potential energy surface                                                                   | 90   |
| 4.8  | Position of the counterion in MeBIC                                                            | 99   |
| 4.9  | Spin decoupling on H1 and H9 for MeBIC                                                         | 101  |
| 4.10 | The effect of a counterion on MeBIC conformation                                               | 103  |
| 4.11 | Spin decoupling on H6' in HBIC                                                                 | 107  |
| 5.1  | Match of GABA and BIC                                                                          | 115  |
| 5.2  | Molecular comparisons by matching charge-centres                                               | 116  |
| 5.3  | Matching of the optimum Y-shaped arrangement of charge-centres with an analogue of low potency | 121  |
| 5.4  | Key partial atomic charges for GABA analogues                                                  | 122  |
| 7.1  | Brief description of the OPTS menu                                                             | 136  |
| 7.2  | Hierarchy of the main routines in IMDAC                                                        | 138  |
| 7.3  | Examples of the use of the CLEFT option                                                        | 142  |
| 7.4  | Examples of the use of CC on BAG                                                               | 144  |
| 7.5  | The drawing of only pre-selected residues                                                      | 146  |
| 7.6  | Difference between active-site and whole-receptor views                                        | 148  |
| 7.7  | Example of the space-filling model                                                             | 149  |
| 7.8  | Examples of the stereo model                                                                   | 150a |
| 7.9  | Example of molecular fragment deletion                                                         | 154  |
| 7.10 | Schematic representation of a proposed catalytic mechanism for PA2                             | 168  |
| 7.11 | phosphatidyl ethanolamine substrate                                                            | 169  |
| 7.12 | PA2 active-site region                                                                         | 170  |
| 7.13 | The structures of two TLN inhibitors for which coordinates of the bound drug are known         | 173  |
| 7.14 | Close Contacts around the benzene-ring in BAG                                                  | 174  |

### Acknowledgements.

To Professor Steward, whose helpful advice and encouragement throughout this project I cannot value too highly, thank you.

I am indebted to [REDACTED] (National Institute for Medical Research, London) and to [REDACTED] for helpful discussions and for help in obtaining some of the NMR spectra.

I wish to express thanks to [REDACTED] [REDACTED] [REDACTED] [REDACTED] [REDACTED] (Queen Mary College Chemistry Department) for 400MHz NMR spectra.

A basic molecular graphics was obtained from [REDACTED] [REDACTED]

To [REDACTED] [REDACTED] [REDACTED], thank you for all your help and support in many ways too numerous to mention here.

I am grateful to The Lord Dowding Fund for financial support.

### Restricted access.

Due to the confidential nature of some of the material contained in this work, restricted access has been placed on it for a period of two years.



## Abstract

Two types of study on drugs and drug receptors are covered in this work, (1) where the molecular structure of the receptor is unknown and (2) where molecular coordinates are known for the receptor. For the first the neurotransmitter  $\gamma$ -amino butyric acid (GABA) was examined. The GABA molecule is particularly troublesome because it is difficult to deal theoretically with a polar molecule in a polar solvent (aqueous - a commonly used approximation to the biophase), and GABA has too many internal rotations for accurate experimental treatment. We have, however, established using variable-temperature NMR that no particular conformational preferences exist for GABA in solution, without having to resort to any assumptions on the angles of the minima or their associated coupling constants. Of the theoretical methods for determining solution conformation, a 'continuum' model shows GABA to be essentially rigid, and a 'discrete' model gives several low-energy minima for GABA. We discuss these results and the problems associated with each method, and with a hybrid of the two methods. Gas-phase theoretical methods have also been examined. We found that reliable geometries and conformational energies can be obtained, provided allowance is made for the known shortcomings of each method.

For the competitive GABA antagonist bicuculline (BIC) we have used a novel NMR method, involving the observed temperature-dependence of BIC coupling constants and chemical shifts, to show that 3 low-energy conformations are present in solution with low barriers between them. The addition of an extra N-methyl group or proton to BIC, however, increases the hindrance to rotation so that only one minimum was observed by NMR in solution. From a knowledge of the accuracy of the NMR measurements we estimated a lower bound for the energy of the next minimum above the global minimum. This estimate was then used to show that the conformer observed for both N-methyl and protonated BIC in solution is the active conformation.

Previous comparisons between GABA and GABA agonists and antagonists are given, in which no distinction is made between agonist and antagonist structural requirements for activity. In the light of all our NMR results our own comparisons were then made from which separate requirements have been deduced.

For the second type of study, where receptor coordinates are known, we have developed an extensive molecular graphics system (IMDAC) for examining active-site clefts of receptors and the docking of drug molecules into receptors - with routines for examining in detail the space available for molecular modification and for intermolecular energy minimisation.

We have used a known receptor structure (thermolysin), as a model for a similar receptor of unknown molecular structure (enkephalinase), and have used IMDAC to model novel enkephalinase inhibitors based on existing thermolysin inhibitors. In addition, a novel possible mode of binding for an enkephalinase inhibitor has been determined.

### Abbreviations.

|          |                                                                      |
|----------|----------------------------------------------------------------------|
| BAG      | (2-benzyl-3-mercaptopropanoyl)-L-alanylglycinamide                   |
| BDZ      | benzodiazepine                                                       |
| BIC      | bicuculline                                                          |
| CC       | close intermolecular contacts (and subroutine for calculating these) |
| CI       | configuration interaction                                            |
| DMSO     | dimethyl sulphoxide                                                  |
| ENK      | enkephalinase                                                        |
| GABA     | $\gamma$ -aminobutyric acid                                          |
| GO       | geometry optimisation                                                |
| HBIC     | protonated bicuculline                                               |
| IGUV     | iso-guvacine                                                         |
| INIP     | iso-nipecotic acid                                                   |
| ISO      | 1,2,3,4-tetrahydro 6,7-dimethoxy isoquinoline                        |
| iso-THIP | 5,6,7,8-tetrahydroisoxazolo[3,4-c] pyridin-3-ol                      |
| MeBIC    | bicuculline methohalide(Cl or I)                                     |
| MEC      | meconin(6,7-dimethoxy phthalide)                                     |
| MM       | molecular mechanics                                                  |
| MUS      | muscimol                                                             |
| NMR      | nuclear magnetic resonance                                           |
| NOE      | nuclear Overhauser enhancement                                       |
| OP       | optimisation parameter (for solvent molecule orientation)            |
| PA2      | phthalide isoquinoline                                               |
| tcpu     | time taken on the central processor unit (computer time)             |
| THIP     | 4,5,6,7-tetrahydroisoxazolo[5,4-c] pyridin-3-ol                      |
| TLN      | thermolysin                                                          |
| QM       | quantum mechanics                                                    |
| $x_T$    | charge separation parameter (GABA)                                   |

### Special abbreviations used in Chapter 7.

The abbreviations ED, TR, MOD and OP followed by a number refer to the particular editing, translate, model type and main menu (OPTS) option numbers respectively.



# Studies on drug molecules by theoretical methods, NMR and computer graphics.

## 1 Preamble.

### 1.1 Introduction.

Many drug molecules are known to exert their biological function by binding to a macromolecular receptor. (Conforming with common practice we use the term 'drug' to include the endogenous ligand.) Molecular studies on the interaction of a drug with its biological receptor can be divided into two classes, (1) where the molecular structure of the receptor has not been determined, and (2) receptors of known molecular structure.

This work is therefore divided into two parts. In Part 1, where the molecular structure of the receptor is unknown, the structures of different drugs active at the same receptor site are examined and compared in an attempt to define the requirements for optimum binding to the receptor. Agonist and antagonist structural requirements are examined separately (largely ignored in the literature), and then compared in an attempt to find a structural distinction between agonist and antagonist action. In Part 2, where atomic coordinates of the receptor molecule are known, molecular graphics techniques are used for examining drugs within the 'active site cleft' of the receptor: with routines for finding possible active-site clefts, docking of drug molecules into an active site with conformational analysis of the docked drug, and calculating interaction energies and other important quantities eg close contacts between drug and receptor.

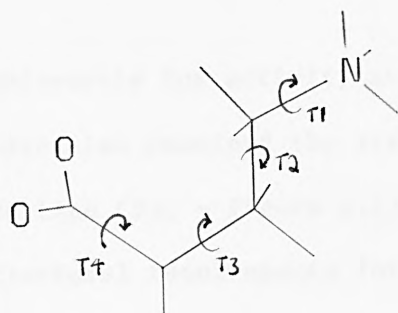
## 1.2 Receptors of unknown molecular structure.

The determination of the receptor-bound conformation of a drug would give a template for the design of new drugs with a similar spatial arrangement of key atoms or functional groups. Alas the endogenous drug is usually considerably flexible and the receptor-bound conformation is not readily obtained for drugs bound to receptors of unknown molecular structure. Many workers have therefore examined drug analogues of restricted flexibility, but high potency, which bind competitively to the same receptor site as the drug concerned. The minimum-energy conformation of such analogues in a (preferably aqueous) solution environment (ie unbound) should be representative of the 'active conformation'\* of the analogue. Comparison of the original drug with analogues in the active conformation will then yield the active conformation of the drug concerned (eg Richards, 1976). It is therefore essential to be able to determine accurately the flexibility and conformational modes of drug molecules in solution, to make valid any comparisons between the drug and its semi-rigid analogues. Some common experimental and theoretical means of examining the conformational behaviour of small molecules in solution are briefly outlined in section 1.4.

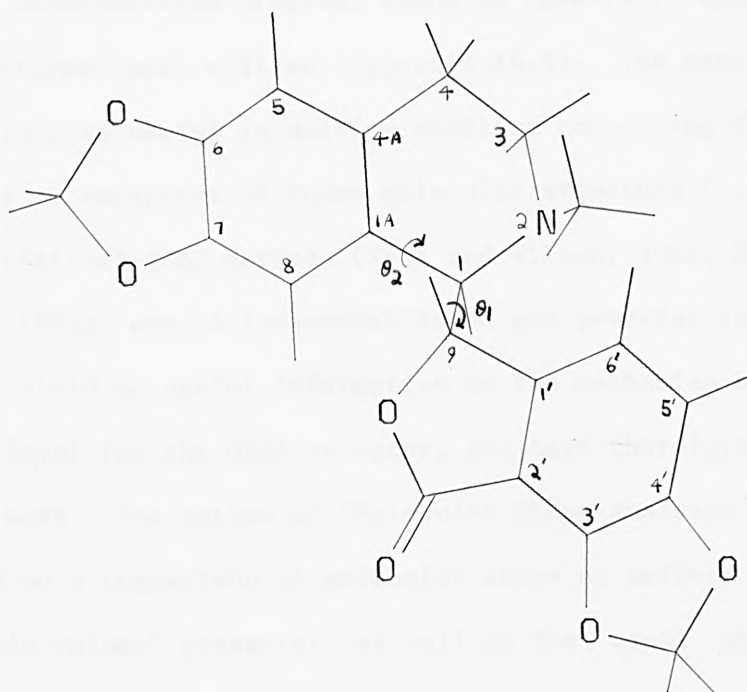
We have examined the flexibility and conformational energies of the inhibitory central nervous system transmitter  $\gamma$ -aminobutyric acid (GABA - Figure 1.1) and the use of semi-rigid analogues for determining the

---

\* A distinction can be made between drug 'recognition conformation' (prior to binding), 'bound conformation' and 'active conformation' (the one involved in eliciting a response from the receptor). A drug molecule is unlikely<sub>1</sub> to undergo a conformational change of more than a few kJ mol<sup>-1</sup> upon leaving the biophase and binding to a receptor (Lambrecht and Mutschler, 1974). Therefore for semi-rigid drugs with a highly populated global minimum, the recognition, bound and active conformations will be virtually identical. We therefore use the term 'active conformation', and only make a distinction when it is necessary (eg 5.4).



GABA



BIC

Figure 1.1 GABA and BIC molecular structures.

For GABA the 4 torsion angles  $T1 - T4$  are shown, and for BIC  $\theta1 = H1-C1-C9-H9$  (shown at  $180^\circ$ ) and  $\theta2 = C9-C1-C1-C4$  (shown at  $\approx 105^\circ$ ). The sign of the angle is defined according to the convention of Klyne and Prelog (1960) - ie a positive  $\theta1$  rotation is defined as a clockwise rotation of the phthalide group while viewing along the  $C9-C1$  bond.

structural requirements for activity at the GABA<sub>A</sub> receptor site (see below). We have also examined the structure of the competitive GABA antagonist bicuculline (BIC - Figure 1.1), with the aim of distinguishing between the structural requirements for GABA<sub>A</sub> agonists and antagonists.

The important structural features of different drug molecules could then be compared using molecular graphics. For this purpose a molecular-graphics superposition program, based on (general) three-point recognition features was written (Appendix A6.5). The same superposition algorithm is also useful in solving problems concerning the docking of drug molecules to receptors of known molecular structure (1.3 and 7.2.3).

Statistical QSAR methods (Free and Wilson, 1964, Martin, 1981, Hopfinger, 1985) are a long-established and powerful tool in drug design, but they yield no useful information on the mechanism of binding at the atomic level for the GABA receptor, and have therefore not been included in this work. The method of 'Molecular Shape Analysis' (Hopfinger, 1980) is based on a comparison of molecular shape as defined by a "common overlap steric volume" parameter, as well as the usual physicochemical and substructural features. Results obtained by applying this method to GABA (Walters and Hopfinger, 1984) are referred to in Chapter 5.

### 1.2.1 The GABA<sub>A</sub> system.

Postsynaptic receptors for GABA in the mammalian central nervous system have been classified as: 'A' - BIC-sensitive, baclofen-insensitive (Olsen, 1981), and 'B' - BIC-insensitive, baclofen-sensitive (Hill and Bowery, 1981). We are interested in the A (BIC sensitive) site and, as further subdivision may occur at this site (Krogsgaard-Larsen and Nielsen, 1984), it is important in any structural comparisons of GABA<sub>A</sub> analogues that the analogues being compared all act at, so far as is

known, precisely the same binding site. In this respect certain GABA analogues have not been used in our structural comparisons - for example, sulfonic acid GABA analogues, which are known to have a different mechanism of action to the analogues described later (Krogsgaard-Larsen et al, 1983).

Much experimental and theoretical work has been done in the past on determining the solution conformation of GABA, but with conflicting results - even as to whether GABA is flexible in solution (see 3.2 and 3.3). Using variable-temperature NMR spectroscopy, however, we have been able to establish that the GABA zwitterion has considerable flexibility **in solution**\* (3.2.2). The accuracy of two existing, opposing methods ('supermolecule' and 'SOLVEFF') for theoretically calculating solution conformations has also been examined (3.3).

Comparisons of GABA with the competitive antagonist BIC (Mohler and Okada, 1977a, Macksay and Ticku, 1984) have been included in many previous studies on GABA (eg Curtis et al, 1970, Andrews and Johnston, 1979), but with no consideration given to the possibility of separate agonist and antagonist structural requirements for activity. 'Partially folded' to 'fully extended' GABA conformations were deduced as the criterion for binding activity at the GABA<sub>A</sub> receptor (eg Krogsgaard-Larsen, 1983, Johnston and Allen, 1984), but in arriving at this the conformations which BIC adopts in solution had not been clearly established. Earlier NMR work (Andrews and Johnston, 1973, Shamma and St.Georgiev, 1974, Elango et al, 1982) showed  $\theta_1$  (Figure 1.1) to be hindered, but only

---

\* The GABA zwitterion, the predominant species in solution, has been shown using theoretical calculations to be quite rigid in the **gas-phase** (Warner and Steward, 1975, Pullman and Berthod, 1975). Solvation should, however, drastically alter this, though some theoretical and experimental methods still insist that GABA is quite rigid (though extended) in solution (see 3.2).



one conformational minimum was searched for. Because the conformational behaviour of BIC needed to be known more precisely we therefore undertook an extensive analysis of the low-energy conformations and barriers to internal rotation for BIC and some N-methyl salts of BIC (MeBIC) in various solvents (Chapter 4). This information was then used in a comparison with the structures of GABA and semi-rigid GABA analogues to postulate separate agonist and antagonist structural requirements for drugs active at the GABA<sub>A</sub> receptor (Chapter 5).

### 1.2.2 Interaction between GABA<sub>A</sub> and benzodiazepine receptors.

The GABA<sub>A</sub> receptor is usually found as part of a large supramolecular complex (molecular mass ca 210,000 - Chang and Barnard, 1982)) at which GABA, benzodiazepines (BDZs) and barbiturates are active at distinct but interacting sites (Guidotti et al, 1978, Martin, 1984 - see Figure 1.2). The complex is linked to a chloride ionophore in the subsynaptic membrane (Simmonds, 1984). Slight structural similarities exist between GABA and protonated BDZs (Figure 1.3), but they do not bind competitively (Squires and Braestrup, 1977, Mohler and Okada, 1977a). However, BDZs can enhance the binding and activity of GABA and GABA agonists (Guidotti et al, 1978, Toffano et al, 1978, Simmonds, 1980, Haefly et al, 1981), and GABA can enhance BDZ action in vitro (Tallman et al, 1978, Martin and Candy, 1978) and in vivo (Gallager et al, 1978). A postulated mechanism for this interaction, which may be important when considering conformational requirements for drugs binding to the GABA<sub>A</sub> site, is that BDZs elicit a conformational change in the GABA<sub>A</sub> receptor (Martin, 1984), somehow enhancing the affinity of GABA, with an increase in biological response. Other mechanisms have been proposed, including displacement of a modulator protein for the GABA receptor (GABA modulin -



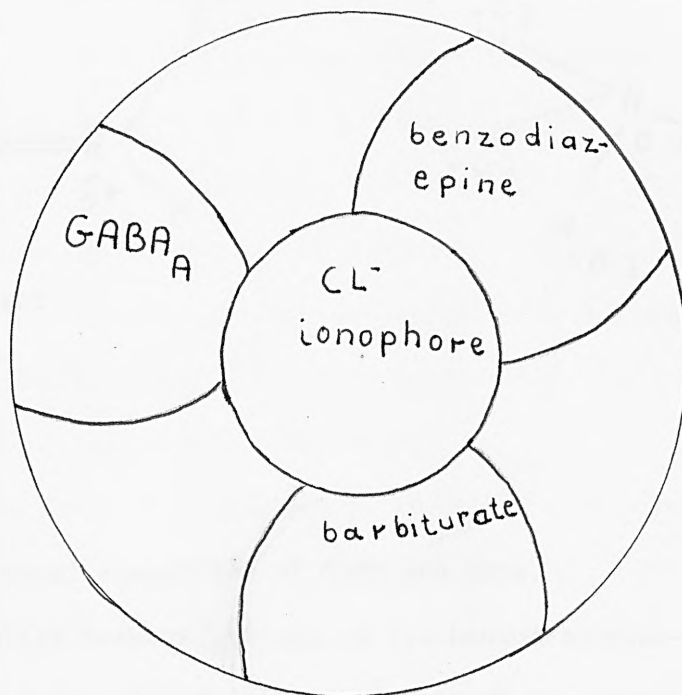


Figure 1.2 Schematic representation of the GABA<sub>A</sub> supramolecular complex.

(Taken from Defeudis, 1981. (abridged))

Guidotti, 1980, Toffano, 1983), and direct interaction at the chloride ionophore (Guidotti et al, 1978, Simmonds, 1980). As none of these mechanisms include direct interaction of BDZs at the GABA<sub>A</sub> binding site, and specific BDZ 'agonists' and 'antagonists' have not yet been very well classified (no definitive endogenous BDZ-receptor ligand has yet been positively identified (Costa and Guidotti, 1985)), BDZs are not further considered in this work.

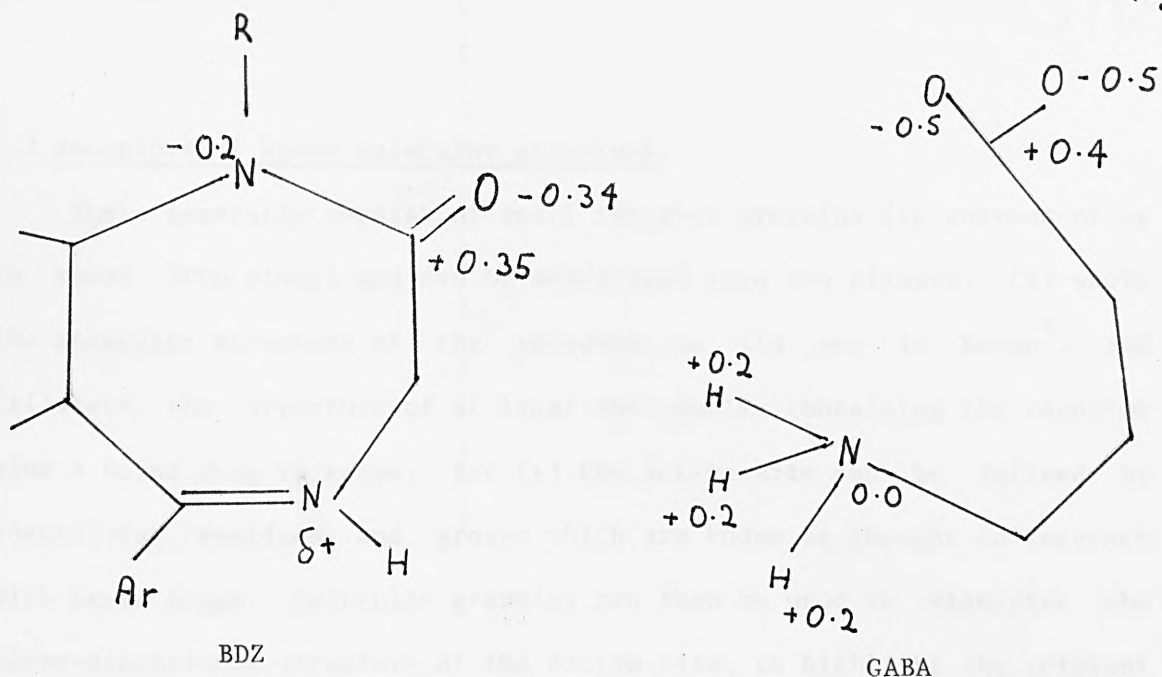


Figure 1.3 Structural comparison of GABA and BDZs.

The slight similarity between protonated 1,4-benzodiazepine-2-ones and GABA in a highly folded ('glycine') conformation is shown. The charge on the N in the BDZs is largely delocalised into the aromatic ring. (The BDZ charges are from Blair and Webb, 1977.)

Barbiturates potentiate GABA transmission by decreasing the rate of dissociation of GABA from its receptor (Johnston and Willow, 1982). Their binding site on the above supramolecular complex (Figure 1.2) has, however, been shown to be distinct from the GABA site (Johnston and Willow, 1982).

### 1.3 Receptors of known molecular structure.

These generally consist of small receptor proteins (eg enzymes of up to about 3000 atoms) and can be subdivided into two classes: (i) where the molecular structure of the receptor on its own is known\*, and (ii) where the structure of at least one complex containing the receptor plus a bound drug is known. For (i) the active site can be defined by identifying residues and groups which are known or thought to interact with known drugs. Molecular graphics can then be used to visualise the three-dimensional structure of the active site, to highlight the relevant residues and atoms (or groups), and to examine the cleft within the active site in which a drug could be placed.

For the second class, (ii) above, where the molecular structure of at least one drug/receptor complex is known, the active site is readily defined from the position of the bound drug. Potential new drugs can then be placed in the same or a similar position to the known drug, and be examined for possibilities of modification to improve interaction of the drug with the receptor.

### 1.4 Conformational analysis of small molecules in solution.

#### 1.4.1 Experimental methods:

##### (i) NMR

The most common is dynamic NMR spectroscopy (eg Sandstrom, 1982). Any system containing two or more conformational modes will be in a state of dynamic equilibrium between those modes. If the temperature of an NMR

\* Usually by X-ray diffraction, or by comparison with proteins of known crystal structure and with a similar residue sequence (eg Chotia et al, 1986). For proteins the crystal molecular structure is representative of the solution structure. (This is not generally true for small molecules, where crystal packing forces are relatively large.)

experiment is such that the rate of exchange is slow compared with the NMR time scale, then NMR peaks for the separate conformers will be observed at low temperature. And for fast exchange, at high temperature, averaged spectra are observed. Conformer populations can be readily and accurately obtained from spectra in the slow exchange limit, if the spin pattern is not too complicated (Sandstrom, 1982). (This is often done from a simple ratio of the peak intensities.) If a spectrum for separate conformers cannot be obtained, for example if the solvent freezes before separation occurs (see eg 4.1), then the conformer populations can be estimated from averaged coupling constant data (Pachler, 1964), but assumptions need to be made on the positions of the minima (usually classical gauche/trans). A better method, similar to this, but based on angles from gas-phase calculations and also incorporating chemical shifts, is described in Chapter 4.

Other NMR methods exist for molecules with low barriers, where certain special features of the molecule are exploited to determine conformational modes and populations (eg the 'J method' - Parr and Schaefer, 1980).

Several NMR methods exist for the evaluation of barrier heights (Sandstrom, 1982, Binsch and Kessler, 1980) which, to be used accurately, require parameters from the non-averaged low-temperature spectrum. Approximate barriers can be obtained, however, by estimating the values of these parameters (see 4.2.3).

#### (ii) Other experimental methods.

Other experimental methods are generally of limited use for solution studies. For example, infra red and Raman spectroscopy are limited by absorption of the solvent or by the complicated nature of spectra for

other than the simplest of molecules. The conformation-dependant signals are usually small and can easily be lost, although qualitative conformational information can still sometimes be obtained (eg Tanaka et al, 1978).

Ultrasonic relaxation techniques have been used on molecules in aqueous solution (eg Jordan et al, 1980) but are limited to simple systems of no more than two non-degenerate unimolecular relaxation processes (ie  $< 3$  conformational minima) with similar relaxation frequencies. For polar solutes and solvents the relaxation process is by no means unimolecular, making any results with polar molecules somewhat dubious.

Dipole moment measurements have also been used (eg Edward et al, 1973), but these give results for GABA which are contrary to more reliable NMR results (see 3.2).

#### 1.4.2 Theoretical methods.

Theoretical methods for the determination of conformational modes and energy barriers for small molecules in solution are generally based on two opposing models: a discrete approach based on short-range solute-solvent interactions (the 'supermolecule' method - Alagona et al, 1973, Pullman and Pullman, 1975), and a macroscopic 'continuum' approach based on the bulk electrostatic properties of the solute and solvent (SOLVEFF - Clarke, 1975, Sinanoglu, 1967, Beveridge, Radna and Kelly, 1974). The use of both of these, separately and combined, on GABA is discussed in Chapter 3. The theoretical methods (quantum mechanical and classical) on which these approaches are based were primarily designed for isolated molecules in the gas-phase, and Chapter 2 is therefore devoted to an analysis of the accuracy and reliability of theoretical gas-phase methods as applied to GABA and other small drug molecules.

'Simulation methods' such as 'Monte Carlo' or 'molecular dynamics' (Robson, 1982) are not directly applicable to this work. The reasons for this are briefly discussed in Chapter 3 (3.4).

#### 1.5 Summary of the main aims of this work.

1) The possibility of improving existing theoretical methods for the determination of conformations and barriers to internal-rotation for drug molecules in solution (Chapter 3).

2) The determination of the conformational behaviour of GABA and of BIC (and MeBIC etc) in solution (Chapters 3 and 4). These data are then used to gain insight into the active-conformations of GABA and BIC, and to determine whether any distinctions exist between the GABA agonist and antagonist structural requirements for activity (Chapter 5).

3) To obtain information on the bound conformations of drug molecules which act at receptors of known molecular structure (Chapter 7), by developing (7.2) and applying (7.3) computer graphics methods for molecular superposition, docking, 'cleft' searching and drug modification.



## 2 Preliminaries: theoretical calculations on small molecules in the gas-phase.

### 2.1 Introduction.

Although we are primarily interested in the conformation and flexibility of drug molecules in solution, especially aqueous as an approximation to the biophase, most theoretical methods for the calculation of conformational energies are based on methods for isolated molecules in the gas-phase, to which solvent effects are then added (3.3). The conformations adopted in the gas-phase are only likely to be similar to those in solution for molecules where steric hindrance is the major factor determining conformational energy, or without large bond moments (Abraham and Bretschneider, 1974, and see 4.2). It is important to examine the accuracy of gas-phase calculations themselves, because if they do not give good agreement with experimental data (where such comparison is possible), then adding solvent effects will almost certainly not produce reliable results.

In the past, the quantum-mechanical CNDO/2 method has been used by the City University Group for the calculation of conformational energies and charges for small drug molecules. This method was used because of its advantages in terms of speed and accuracy over other methods at that time (Borthwick, 1977). However, as several new and reputedly improved methods have since been developed it became necessary to review the situation and to test the new methods on GABA and on several small test molecules for which reliable experimental data were available.

Details of these gas-phase calculations and on BIC (and MeBIC) are reported here, with reference to the effects of geometry optimisation on GABA conformations, and the additional parameters required for MM2 calculations on BIC.

Included in this Chapter is an assessment of the need for geometry optimisation, particularly with regard to the calculation of conformational properties. The need for including configuration interaction is also considered.

## 2.2 Survey and comparison of theoretical methods.

### (i) Quantum mechanical (QM).

The great diversity of quantum mechanical methods now available to the modern computational chemist is at first rather alarming. Several reviews have appeared in the literature, covering the more common molecular orbital methods including: CNDO, INDO, PCILO, PRDDO, LINDO, MINDO, MNDO and ab initio (Halgren et al, 1978, Dewar and Ford, 1979, Huzinager, 1985, Schulz et al, 1985). The general conclusion drawn from these reviews is that for the calculation of relative molecular energies (eg conformation) and charges the MNDO method (Dewar and Thiel, 1977), when compared with experimental data, is usually more accurate than the other methods above (Dewar and Ford, 1979, Schulz et al, 1985), except for extended basis-set\* ab initio methods, which are prohibitively time-consuming for conformational analysis on the molecules of interest here. MNDO is comparable with the other semi-empirical methods in terms of computational speed (Table 2.1). (For semi-empirical methods, computer time ( $t_{\text{cpu}}$ )  $\approx kn^2$ ; where  $n$  is the number of molecular orbitals, and  $k$  is a constant. For ab initio  $t_{\text{cpu}} \approx kn^4$ .) We therefore undertook a series of test calculations on small molecules to compare the performance of MNDO with other methods in predicting conformational properties and dipole moments (charges), as compared with experiment.

\* Smaller and faster minimal basis-sets are known to be somewhat unreliable (Huzinager, 1985), and are not worth the large computer resources required since semi-empirical methods are, in general, as accurate as minimal basis-set calculations. It is also important that the correct basis set is chosen for the molecule in question (Huzinager, 1985).

For this, simple substituted ethanes and ethenes were chosen since there is only one rotation to be considered, accurate experimental data were available, and a wide range of molecular polarities could be covered. The molecules propanal and ethanoamide were also included since they are well covered in the literature (eg Nuffel et al, 1984, Neece, 1980). The results of the test calculations are summarised in Tables 2.1 - 2.3 and in Figure 2.1. From these it is evident that MNDO and extended basis-set ab initio methods (MM2 is discussed in the next section) do not always give the most accurate conformational minima and maxima (angles - Table 2.1, energies - Table 2.2, and see Figure 2.1), and dipole moments (Table 2.1). For comparisons between methods we refer to calculations with geometry optimisation (wherever possible), since more accurate conformational energies are obtained with geometry optimisation (Dewar and Ford, 1979). (The effects of geometry optimisation are discussed in detail in section 2.3.) In predicting molecular geometries we found (Table 2.3) that ab initio (and MM2) gave bond

Table 2.1 Comparison of MNDO, CNDO/2, MM2, extended basis-set ab initio and experimental conformations and dipole moments ( $\mu$ ).

| Method                  | 1,2-difluoroethane |      |           | propanal          |      |          |
|-------------------------|--------------------|------|-----------|-------------------|------|----------|
|                         | $\mu$ (Debye)      | tcpu | minima    | $\mu$ (Debye)     | tcpu | minima   |
| Experiment              | 2.24 <sup>h</sup>  | —    | 60°, 180° | 2.75 <sup>e</sup> | —    | 0°, 120° |
| MNDO (std)              | 1.7 (3.03)         | 0.19 | 180°, 60° | 2.0 (2.2)         | 0.17 | 100°     |
| MNDO+GO                 | 2.15(2.84)         | 1.6  | 62°, 180° | 2.35(2.33)        | 1.2  | 101°     |
| MNDO+GO+CI              | 2.04(2.84)         | 2.3  | 63°, 180° | 2.35(2.33)        | 2.3  | 100°, 0° |
| CNDO (std)              | 2.1 (2.78)         | 0.16 | 180°, 60° | 2.2 (2.4)         | 0.17 | 100°     |
| CNDO+GO                 | 1.1 (2.80)         | 1.4  | 64°, 180° | 2.8 (2.84)        | 3    | 100°     |
| PCILO(std)              | 1.8 (2.1)          | 0.03 | 60°, 180° | —                 | —    | —        |
| MM2+GO                  | —                  | 0.16 | 60°, 180° | —                 | 0.17 | 0°, 120° |
| ab initio<br>(4-31G+GO) | (2.29)             | 550  | 60°, 180° | 3.2               | —    | 0°, 120° |

GO=geometry optimisation, CI=configuration interaction, std=fixed standard geometry (Pople and Beveridge, 1970). Minima are in ascending order of energy. Theoretical dipole moments were calculated using a Boltzmann distribution to match the experimental temperature. The figure in brackets is the dipole at the experimental minimum-energy conformation. For references see Table 2.2.

Table 2.2: Comparison of MNDO, CNDO/2, MM2, extended basis-set ab initio and experimental conformational energies and barriers (kJ mol<sup>-1</sup>).

| Molecule                             | MNDO+GO<br>$\Delta E_{t-g}$ | $\Delta E^\#$ | $\Delta E_{t-g}$      | ab initio+GO<br>$\Delta E^\#$ | CNDO + GO<br>$\Delta E_{t-g}$ | MM2 + GO<br>$\Delta E_{t-g}$ | Experiment<br>$\Delta E_{t-g}$ | ΔG#                  |
|--------------------------------------|-----------------------------|---------------|-----------------------|-------------------------------|-------------------------------|------------------------------|--------------------------------|----------------------|
| CH <sub>2</sub> FCH <sub>2</sub> F   | 1.3                         | >6            | 0.3±0.3 <sup>a</sup>  | 12 <sup>a</sup>               | 0.45                          | 7.2                          | 3.6±3.6 <sup>a</sup>           | 9.9±2.4 <sup>a</sup> |
| CH <sub>2</sub> ClCH <sub>2</sub> Cl | -5.4                        | >7            | -8.0±0.2 <sup>a</sup> | -                             | -                             | -                            | -4.4±1 <sup>b</sup>            | -                    |
| CHCl <sub>2</sub> CH <sub>2</sub> Cl | -5.9                        | 18            | -                     | -                             | -                             | -                            | -10.9±1.2 <sup>b</sup>         | -                    |
| CHCl <sub>2</sub> CHCl <sub>2</sub>  | -0.3                        | 14            | -                     | -                             | -                             | -                            | 0±0.8 <sup>b</sup>             | -                    |
| CH <sub>2</sub> ClCH <sub>2</sub> F  | -2.8                        | 6             | -                     | -                             | -                             | -                            | -1.2±1 <sup>b</sup>            | -                    |
| CHF <sub>2</sub> CHF <sub>2</sub>    | 0.27                        | 6.5           | -                     | -                             | 0.25                          | 4.0                          | -4.8±0.4 <sup>b</sup>          | -                    |
| CH <sub>2</sub> ClCH <sub>2</sub> OH | -3.4                        | 7             | 4.1 <sup>c</sup>      | >20 <sup>c</sup>              | -                             | -                            | 5.0 <sup>c</sup>               | >ca20 <sup>c</sup>   |
| CH <sub>2</sub> FCH <sub>2</sub> OH  | 0.9                         | 6             | 2.6 <sup>c</sup>      | >20 <sup>c</sup>              | -0.5                          | 8.7                          | >5 <sup>c</sup>                | -                    |
| CHClCHCl                             | -4.2                        | -             | -1.0 <sup>d</sup>     | -                             | -                             | -                            | 3.0 <sup>d</sup>               | -                    |
| CHFCHF                               | 2.1                         | -             | 3.1 <sup>*d</sup>     | -                             | -0.5                          | -                            | 4.5 <sup>d</sup>               | -                    |
| CHONH <sub>2</sub>                   | -                           | 33            | -                     | 835 <sup>e</sup>              | -                             | -                            | -                              | 80±12 <sup>e</sup>   |
|                                      | -                           | 40 (+Cl)      | -                     | 66.5 <sup>f</sup>             | -                             | -                            | -                              | 70±10 <sup>f</sup>   |

The barriers are the lowest between the given minima (trans and gauche or cis). For abbreviations see Table 2.1 \* Obtained with a very elaborate (D106+\*\*) basis-set. Calculations with simpler basis-sets give the wrong conformer as the most stable.  
 \*\* Only one minimum was found at 80°

- a) Radom et al, 1985; b) a collection in Park et al, 1974; c) Murto et al, 1984; d) Hiramio and Miyajima, 1984; e) Radom and Riggs, 1980; f) Neece, 1980; g) Friesen and Hedberg, 1980, h) Nuffel et al 1984.

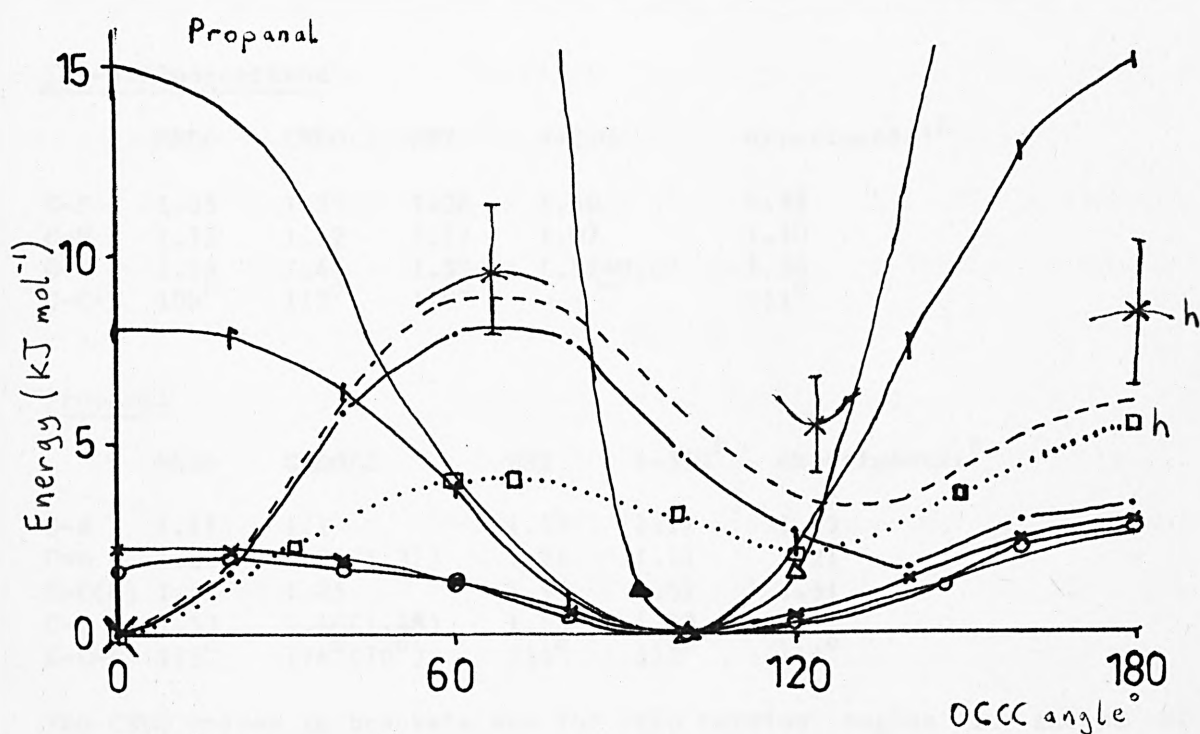
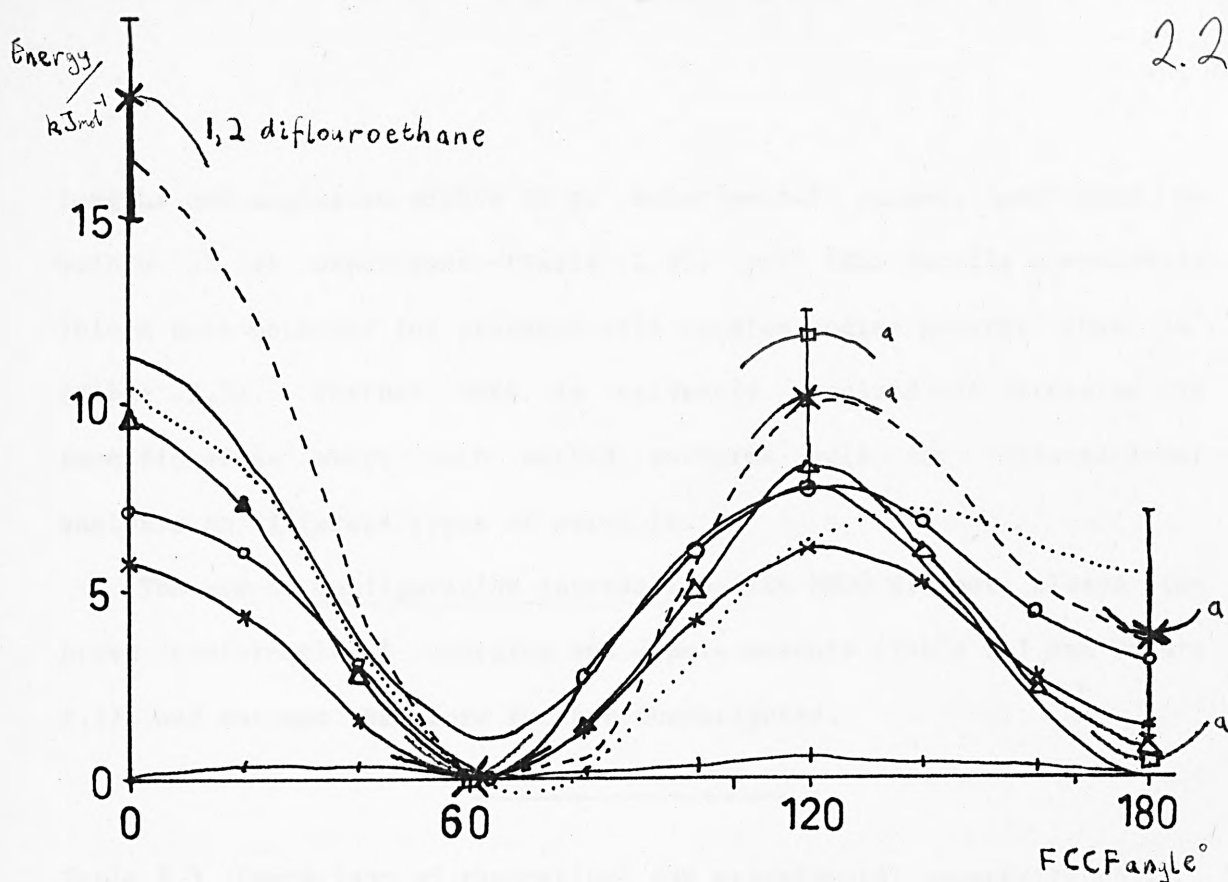


Figure 2.1. Energy surfaces calculated by various methods.

- |                                |                                  |
|--------------------------------|----------------------------------|
| — MNDO (standard geometry),    | xxx MNDO + GO                    |
| -o-o- MNDO + GO + CI           | —•—•— EHT (Kohler, 1971)         |
| - - - MM2 + GO                 | ..... PCILO (standard geom.)     |
| + + + CNDO (standard geometry) | -△-△- CNDO + GO                  |
| -□-□- ab initio                |                                  |
|                                | x experimental minima and maxima |
|                                | X                                |

For references see Table 2.2.



lengths and angles to within 2% of experimental values, and MNDO to within 3% of experiment (Table 2.3). With CNDO totally unrealistic values were obtained for propanal with torsion angles greater than  $40^\circ$  (Table 2.3). Further work is evidently required to determine the specific areas where each method performs well for conformational analysis on different types of molecule.

The use of configuration interaction with MNDO did not always improve conformational energies and dipole moments (Table 2.1 and Figure 2.1), and was not therefore further investigated.

Table 2.3 Comparison of theoretical and experimental geometries ( $\text{\AA}$ ).

1,2-difluoroethane

|       | MNDO         | CNDO/2       | MM2          | 4-31G           | experimental <sup>g</sup> |
|-------|--------------|--------------|--------------|-----------------|---------------------------|
| C-F   | 1.35         | 1.35         | 1.38         | 1.40            | 1.39                      |
| C-H   | 1.12         | 1.12         | 1.11         | 1.07            | 1.10                      |
| C-C   | 1.58         | 1.47         | 1.53         | 1.52 $\pm$ 0.03 | 1.50                      |
| C-C-F | 109 $^\circ$ | 113 $^\circ$ | 110 $^\circ$ | —               | 111 $^\circ$              |

Propanal

|        | MNDO         | CNDO/2                      | MM2          | 4-31G <sup>h</sup> | experimental <sup>h</sup> |
|--------|--------------|-----------------------------|--------------|--------------------|---------------------------|
| C-H    | 1.11         | 1.1                         | 1.15         | 1.12               | 1.13                      |
| C=O    | 1.22         | 1.26(1.31)                  | 1.21         | 1.21               | 1.21                      |
| C-C(O) | 1.53         | 1.45                        | 1.52         | 1.51               | 1.51                      |
| C-C    | 1.53         | 1.46(1.48)                  | 1.53         | 1.53               | 1.52                      |
| C-C-C  | 115 $^\circ$ | 116 $^\circ$ (70 $^\circ$ ) | 114 $^\circ$ | 112 $^\circ$       | 114 $^\circ$              |

The CNDO values in brackets are for CCCO torsion angles in excess of  $40^\circ$ . For references see Table 2.2.

Discussion of currently available QM methods.

Advantages of the MNDO method are:

1) In the present survey MNDO gave an average agreement with published experimental conformational energies within  $3\pm 3\text{kJ mol}^{-1}$ , often in better agreement than extended basis-set ab initio calculations (see Table



2.2). Other semi-empirical methods also usually gave reasonable agreement with experiment (eg for CNDO: ca  $4\text{kJ mol}^{-1}$  (Table 2.2)). However, the less elaborate methods (CNDO, INDO, MINDO and PCILO) are not generally as reliable as MNDO for reasons outlined below.

2) MNDO gives bond lengths and angles in good agreement with experiment (Dewar and Thiel, 1977; and see Table 2.3). In addition, MNDO tends not to produce absurd values such as those sometimes obtained by the simpler methods (see the CNDO and PCILO sections below).

3) Before Gaussian 82 (Collins et al, 1976) became available (at ULCC) in March 1985, and a new combined MNDO/CNDO/MNDOC/MINDO programme in mid 1985, MNDO was the only QM method available (except for some rather elaborate ab initio programmes, eg GAMESS (Dupuis et al, 1980)) with an efficient algorithm for geometry optimisation, and specifically definable geometry optimisation parameters for each bond variable.

4) An advantage of the MNDO programme available at ULCC is that it can be readily interfaced to other programmes (eg our SOLVEFF (3.3.2) or graphics (7.2.7) - see Appendix A6.1 and A6.5).

The MNDO method does have some known shortcomings:

1) It fails to account correctly for hydrogen-bonding (Dewar and Ford, 1979, Koller et al, 1985). In our calculations this only affected folded conformations of the isolated GABA molecule, which are of very high energy anyway. However, care is needed with the positioning of water molecules when using these to approximate the aqueous solution environment (see 3.3.1 - supermolecule). It is worth noting that other semi-empirical molecular orbital methods only fortuitously give correct results for H-bonding (Dewar and Ford, 1979). A modified MNDO method, MNDO/H (Burstein and Isaev, 1984), designed to overcome the problems with H-bonding, gives better results, but still cannot be relied upon (Koller et al, 1985).

2) Single-bond rotation barriers are underestimated (Dewar and Ford, 1979). We found (Figure 2.1) that this is generally true if geometry optimisation is used (a possible reason for this is given in 2.3.1 i), but with fixed geometry the rotation barriers tend to be rather erratic.

3) Inaccurate energies are reported for molecules with steric crowding - eg tertiary butyl groups are too unstable (Dewar and Ford, 1979, Dewar and Thiel, 1977). Steric crowding is a problem for most theoretical methods and, in this work, only affects GABA in high-energy folded conformations.

4) Calculations reported in the literature on molecules with N-O bonds give inconsistently inaccurate heats of formation, with an average error of  $60 \text{ kJ mol}^{-1}$  (Dewar and Thiel, 1977, Dewar and Ford, 1979). In our calculations on molecules containing N-O bonds (eg THIP, MUS) the N-O bonds were too short (as compared with experiment (Dewar and Thiel, 1977)) by ca  $0.1 \text{ \AA}$ . This problem was solved by calculating the N-O bond length for a fragment of the molecule concerned using a 4-31G basis-set ab initio method (Gaussian 82), and incorporating this fixed bond length into the MNDO calculations. The small increase in energy resulting from keeping the N-O bond fixed is unimportant since we only require molecular geometries and relative energies.

Other shortcomings of the MNDO method have been reported but are generally concerned with more unusual molecules (eg containing N-N bonds - Fos et al, 1985) and are therefore of no significance to this work.

• • • •

It is not wise to rely solely on one theoretical method (Borthwick, 1977, Gregory and Przybylska, 1978). The examples given in Figure 2.1 clearly demonstrate this for MNDO. For 1,2-difluoroethane the theoretical methods are in reasonable agreement in predicting conformational energies, but some of the methods, eg MNDO with fixed geometry, give the wrong conformer as the more stable one (Figure 2.1). For

propanal MNDO tends to give a maximum where a minimum should be at  $0^\circ$  (Figure 2.1). (The inclusion of configuration interaction with MNDO gives a small minimum at  $0^\circ$  (Figure 2.1 and Table 2.1), but it is not the expected global minimum, and since the computation time is roughly doubled, and configuration interaction does not always give improved results (Table 2.1), it has not been included in our calculations on drug molecules.)

We have therefore examined other semi-empirical methods for the possibility of using them alongside MNDO. (A classical mechanics method, MM2, is examined in the next section.)

The MINDO/3 method (Bingham, Dewar and Lo, 1975) has not been included in this work because it is generally inferior to MNDO in almost every respect (Dewar and Ford, 1979), and a newer method, MNDOC (Thiel, 1981), has also not been included because of the small number of atoms for which it has so far been parameterised.

The CNDO/2 method (Pople and Segal, 1966) had been widely used by the City University Group in the past as a reliable method, and is still in quite common use (eg Billes, 1986, Brakaspathy and Singh, 1986). And since it generally gives reasonable agreement with experimental results (Table 2.2) we have therefore continued to use it, mainly to enable new MNDO calculations to be compared with older CNDO calculations\* and, when applicable, ab initio. The CNDO method does have some quite serious faults (see next page):

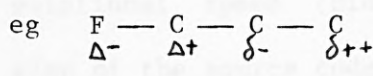
---

\* A much faster version of CNDO/2 became available later, on the Cray computer. This has the advantage of being incorporated within the MNDO programme (and also includes MNDOC and MINDO/3), making comparisons between methods easier. In addition, crude coordinates from, for example, the molecular editing routines in the graphics package (see 7.2.4) can be optimised rapidly with CNDO and then further refined using MNDO. (Although MM2 was later found to supercede CNDO in this respect - see iii below.)

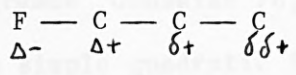
1) It fails almost completely on non-bonded lone-pair/lone-pair and lone-pair/pi-bond interactions (Gregory and Paddon-Row, 1976, Dewar and Ford, 1979), and only fortuitously appears to account for H-bonding (Dewar and Ford, 1979).

2) Agreement with experimental dipole moments and relative energies is poor (Halgren et al, 1978, Dewar and Ford, 1979).

3) CNDO/2 predicts alternating charges near a highly electronegative substituent:



This disagrees with the classical picture of gradually decreasing positive charge:



predicted by MNDO (Stowlow et al, 1981).

The classical view is further supported by calculations on the electrostatic component of the conformational enthalpy change (axial - equatorial) for 4-chloro - 1,1 bis (trifluoromethyl) cyclohexane. Calculations using CNDO/2 charges do not agree with experiment whereas calculations using classical type charges do (Stowlow et al, 1981).

4) Geometry optimisation can sometimes give alarmingly inaccurate structures, eg for propanal in 60° - 180° conformations the C-C-C angle tended to 70° instead of the expected (and MNDO calculated) value of ca 114°! (Table 2.3).

5) CNDO/2 fails to predict correctly the conformations of conjugated molecules or of molecules containing atoms in the second row of the periodic table (Veillard, 1975, Weller et al, 1975).

Further discussion of the CNDO method can be found in P. Borthwick's thesis (1977) and in Veillard (1975). The INDO method (Pople et al, 1967) has not been used because it is similar to CNDO in terms of accuracy (and often much worse (Gregory and Paddon-Row, 1976)), and requires slightly more computer time and memory than CNDO (Golebiewsky and Parczewski, 1974).

The non-SCF method, PCILO, is very popular due to its fast computational speed (Diner, Malrieu and Claverie, 1969). A modified version of the source code of the PCILO programme, QCPE 221, was available at the City University. We attempted to incorporate geometry optimisation into this programme using an algorithm based on the ab initio programme Gaussian 76, where bond length/angle increments are obtained by a simple quadratic fit to the 3 previous points in an iterative procedure. A serious problem arose in that bond lengths tended to infinity when optimised, especially for molecules with any steric crowding (eg the C1-C9 bond in BIC). In addition, when we attempted to optimise GABA/solvent molecule distances using PCILO (for a 'Supermolecule' solvent effect model - see 3.3.1) the solute-solvent separation also tended to infinity. To verify that it was the PCILO method that at fault, and not our added geometry optimisation, we performed calculations on a completely different polar system -  $\text{CH}_2\text{FCH}_2\text{F} \cdot \text{H}_2\text{O}$  - and obtained the same result. As we could find no simple solution to this problem and as PCILO gives generally less reliable conformational energies and is no\*

---

\* This is despite a ca 50% time-saving in conformational calculations made by our addition of code for transferring optimised polarisabilities from one conformation to the next. (This change had no effect on the final PCILO energies.)



faster than molecular mechanics methods, we decided to discontinue using PCILO. Additional reasons were that the somewhat artificial nature of the input molecular orbitals makes PCILO not well suited to molecules where delocalised bonding is important. Furthermore, the input for PCILO is very inefficient and unsuitable for interface to other programmes - even with modification of the source code, parts of which were inefficiently written and difficult to decipher. (Eg we replaced an unnecessarily complicated 14 line subroutine for simple vector multiplication with only 6 lines of code, which was then much clearer to understand.)

Finally, the change at ULCC to Amdahl and Cray computers (see Appendix A4) meant that the version of PCILO on the CDC7600 would no longer work without extensive conversion to either Cray Fortran or to one of the Amdahl-supported compilers. As we were responsible for supporting PCILO at ULCC and no demand was voiced for PCILO (Altmann, 1984), we decided in view of all the above that the conversion was not worth the time needed to achieve it.

There has been a marked increase in the number of ab initio calculations appearing in the literature recently, making the use of ab initio methods now more common than semi-empirical methods. However, computers will need to advance considerably further before it would become feasible to perform extended basis-set ab initio calculations (with geometry optimisation) on a GABA conformational energy-surface - especially if a solvent is also considered (3.3).

In summary, the more elaborate QM methods MNDO and ab initio give conformational energies and barriers to within  $4\text{kJ mol}^{-1}$  of experiment, dipole moments to within 10% of experiment, and bond lengths and angles to within 2-3% of experiment (Tables 2.1 - 2.3 and Figure 2.1). Less

2.2

elaborate QM methods (MNDO and ab initio) give conformational energies and barriers to within  $4\text{kJ mol}^{-1}$  of experiment, dipole moments to within 10% of experiment, and bond lengths and angles to within 2-3% of experiment (Tables 2.1 - 2.3 and Figure 2.1). Less elaborate QM methods (eg CNDO and PCILO) can sometimes give totally unrealistic results. QM methods are time consuming for conformational analysis of molecules with several rotations.

(ii) Classical methods - molecular mechanics (MM).

The molecular mechanics (or force field) method offers a rapid and reasonably accurate (see Table 2.2 and 2.3) means of determining molecular structure and energies (Allinger, 1976, Burkert and Allinger, 1982). One of the most popular molecular mechanics programmes is Allinger's (1977) MM2 (and the earlier parameterisation MM1) which incorporates fast and efficient geometry optimisation and is therefore useful for rapid conformational analysis of small-medium size molecules which contain no delocalised bonding\* or unusual features for which parameters are not available and cannot readily be estimated.

For the first of these problems, if a molecule contains delocalised bonding which does not directly affect conformational energies (eg as in BIC), then a special aromatic atom type may be set up to account for the delocalisation (see 2.5 and Allinger, 1983). For the second problem, parameters are not always available for molecules with features which are at all unusual (eg THIP (5.3), KELA (7.3.2) - which contain N-O bonds). However, parameters may be approximated by comparison with parameters for similar atom/bond types. (The derivation of some parameters for BIC is described in 2.5.) In

\* A newer programme, MMP2, deals with delocalised bonding, but was not available for general release at the time this work was done.

addition, the number of published parameters is constantly expanding (eg Profeta and Allinger, 1985).

MM2 produces (Tables 2.1 and 2.3) reasonably accurate **positions** (torsion angles and bond lengths and angles) of conformational minima for molecules where steric factors strongly influence the conformational energy (eg BIC), and copes well with poor input molecular geometry. Molecular coordinates from the IMDAC molecular editing routine (7.2.4), which are too crude for SCF convergence using MNDO (or CNDO), usually converge readily after refinement using MM2.

### (iii) Quantum mechanics verses molecular mechanics methods.

A significant advantage that MM has over QM is the enormous difference in computer resources required, and since MM2 gives reasonably accurate molecular geometries, and copes well with poor input geometries, it is extremely useful for pre-refinement of crude molecular coordinates (eg produced using the IMDAC molecular editing routines - see 7.2.4) for input to eg MNDO. The technique has been used successfully on molecular coordinates which were so crude that they failed to give SCF convergence with MNDO or other QM methods (eg the drug analogues derived in Chapter 7 - see 7.3.2). The faster, less elaborate, QM methods (eg CNDO or PCILO) are by no means reliable in giving accurate geometries (see Table 2.3 and (i) above) and are therefore not as useful as MM2 for this purpose.

Overall, the more elaborate QM methods (eg MNDO or extended basis-set ab initio) are more generally applicable than MM methods, and should therefore be used whenever the molecule is small enough (< 50 atoms MNDO, < ca 20 atoms 4-31G ab initio - though  $t_{\text{cpu}}$  will be the limiting factor in conformational problems). However, MM2 performs very well for the molecules with no delocalised bonding, and for which parameters are available or can be readily calculated (see examples in Table 2.2 and Figure 2.1).

## 2.3 Geometry optimisation.

It is generally accepted that more accurate conformational energies and charges are obtained in theoretical calculations if molecular geometry is optimised with respect to energy (Dewar and Ford, 1979, Murto et al, 1984). However, geometry optimisation requires far more computer time than equivalent calculations using fixed geometry<sup>\*</sup>. It is particularly important to consider the effect of geometry optimisation when calculating conformational energies and angles, as the conformational energy is dependent on the optimised bond lengths and bond angles for which theoretical methods are often parameterised to give values in agreement with experimental data (eg MNDO - Dewar, 1983; MM1 and MM2 - Allinger, 1976). We have therefore included in our test calculations (2.2) an examination of the overall effect of geometry optimisation on conformational energy and charges. We have also examined ways of drastically reducing the prohibitive amount of computer time required for full analysis with geometry optimisation. Again MNDO was the method of choice, because geometry optimisation does not give good results with, eg CNDO and PCILO, and may well be detrimental (Weller et al, 1975, Borthwick, 1977, and see 2.2 (i)).

### 2.3.1 The effect of geometry optimisation on calculated physical properties.

Some generalisations on geometry optimisation are made here. More specific detail on the effects of geometry optimisation on GABA and BIC can be found in sections 2.4 and 2.5.

#### (i) Barriers.

<sup>\*</sup> When we examined the effects of geometry optimisation on GABA (using MNDO), we found a more than 30 fold increase in t<sub>cpu</sub> and for larger molecules the difference can be greater.

In general, internal-rotation energy-barriers calculated with geometry optimisation tend to be slightly low when compared with experimental gas-phase data (Borthwick, 1977, Dewar and Ford, 1979 - and see Table 2.2), but calculations employing fixed geometry tend to be less consistent (Figure 2.1). The former is partly due to the fact theoretical energy differences for single conformations are being compared with experimental free-energy values (see 2.3.3 iii). There is also a great difficulty in obtaining reliable experimental data (Bouma and Radom, 1978), but the correct trends can be seen to be reproduced in that more bulky substituents give rise to higher barriers (Table 2.2).

(ii) Conformational energy minima.

Intuitively, calculations with geometry optimisation should give more accurate conformational minima (angles and energies) than calculations with fixed geometry, since the internal geometry of a molecule would be expected to change with conformation. This is born-out in, for example, 1,2-difluoroethane where the correct global minimum is given with geometry optimisation, whereas calculations with fixed standard geometry (Pople and Beveridge, 1970), or experimental (Friesen and Hedberg, 1980) geometry give the wrong conformer (Figure 2.1). In general, for the molecules for which comparisons between fixed geometry and geometry-optimised conformations were available, considerably improved results were found with geometry optimisation (eg Figure 2.1). This will, of course, not always be true because of the approximate nature of the theoretical methods used, but MNDO performs much better than the less elaborate QM methods, where rather absurd geometries are sometimes obtained with geometry optimisation (eg CNDO/2 - where for propanal in conformations above  $60^\circ$  totally unrealistic geometries were



produced - Table 2.3; PCILO - see 2.2 i). The simpler methods were designed when computers were considerably less powerful, with the question of geometry optimisation not being considered - in contrast MNDO was designed with geometry optimisation specifically in mind (Dewar, 1983).

For molecules with single rotations reasonably accurate positions (angles) of minima can be obtained without geometry optimisation (Table 2.1). However, when more than one rotation is involved, the situation is more complicated, and an error in energy differences for one rotation will affect the positions of the minima for the other rotation(s). Therefore it is even more important to consider geometry optimisation when multiple rotation angles are involved, such as with GABA.

### (iii) Charges.

The effect of geometry optimisation on charges tends to be rather small. Calculated dipole moments changed by an average of ca 5% with geometry optimisation. For GABA the average change in charges with geometry optimisation (compared with fixed geometry) is 0.015eu, and the maximum change is 0.061eu. This result is not surprising since the charges on different functional groups have been found to be independent (+0.1eu) of the position of the group in a molecule (Borthwick and Steward, 1977). Dipole moments calculated with geometry optimisation are on average slightly closer to experimental values than 'fixed-geometry' dipole moments (Table 2.1). (Dipoles calculated with gas-phase experimental geometry are little different to those with standard geometry). In conclusion, it is only necessary to use geometry optimisation if accurate (better than  $\pm 0.1$ eu) charges are required, or if the input geometry is crude (which would lead to an inaccurately calculated dipole moment).

### 2.3.2 Possible short cuts with geometry optimisation.

For molecules without conformationally mobile ring-systems fixed geometry (experimental or standard (Pople and Beveridge, 1970)) can be used initially and then selected portions of the conformational energy surface - namely minima and maxima - re-calculated with geometry optimisation. If the conformations (angles and energies) calculated with geometry optimisation are significantly different from those with fixed geometry then the geometry-optimised surface should be further investigated.

Also the number of variables can be reduced by determining the parameters which have only a small conformation-dependence (on average  $< \text{ca } 0.005^\circ/\text{cycle}$  for bond lengths,  $< \text{ca } 2^\circ/\text{cycle}$  for bond angles and  $< \text{ca } 5^\circ/\text{cycle}$  for torsion angles) and then using fixed average values for these parameters. (This method could not be applied to GABA because of the lack of symmetry in most solution.) Care is needed in fixing parameters in this way, in that conformational energies must not be significantly altered by the artificial constraints. Alternatively, similar parameters can be set to be equal to one another. Eg benzene-ring bond lengths and angles could be set to be all equal or equal in pairs (see 2.5).

In addition, the geometry optimisation convergence criterion (EYEAD) could be lowered, since this makes little difference to **relative** (eg conformational) energies, and an appreciable saving in  $t_{\text{cpu}}$  can be achieved. However, care is needed in that energies for different conformations must be calculated with the same EYEAD value and, if EYEAD is lowered too much, local minima could become more of a problem (see 2.3.3 ii), since true convergence may not be reached. We have therefore kept EYEAD at the default value in our calculations on

GABA and BIC, since we wished to avoid problems with local minima on multidimensional energy-surfaces.

### 2.3.3 Some possible pitfalls with geometry optimisation.

#### (i) Chemical changes.

For example proton migration causing the GABA zwitterion to change to the more stable (in the gas-phase) non-zwitterion form. This problem was alleviated by the addition of bond-length constraints (see 2.4).

#### (ii) Local minima.

This is a well known problem (Stewart, 1985), particularly for molecules with several rotations. The simplest way of checking that minimisation has not become stuck in a local minimum, or a saddle point, is to repeat the calculation from different starting points (several if a lower energy is found) and take the conformation of lowest energy as the true minimum.

For GABA local minima were found not to be a problem with energy differences of  $< 0.1 \text{ kJ mol}^{-1}$  found between the same conformation calculated from different starting points. For BIC energy differences of up to  $2 \text{ kJ mol}^{-1}$  were found, and for MeBIC slightly higher differences. The differences are within the accuracy of the methods (MNDO and MM2) used (see 2.2) and are therefore not important.

#### (iii) Entropy.

The shape of the conformational well must be considered when comparing theoretical results with experimental. Unfortunately with the optimisation of many variables the shape of a multidimensional minimum can be exceedingly complex. Care is needed in checking for par-

ticularly shallow wells in the conformational-energy surface, since they could have a large entropy contribution. An approximate free-energy value could be obtained by a summation of points about the minima and (for barriers) maxima, but would require a much finer grid than the  $20^\circ$  one used in the test calculations.

Statistical methods exist for including entropy effects, but the methods (eg Monte Carlo - Finney, 1982, molecular dynamics - Gunsteren and Berendson, 1982) are very time consuming and are impracticable for the molecules of interest here. (Future advances in parallel computers may, however, change the situation (Counts, 1985).)

#### 2.4 Gas-phase calculations on GABA.

In previous theoretical calculations on the conformation of GABA, the CNDO/2 (Warner and Steward, 1975) or PCILO (Pullman and Berthod, 1975) methods were used with fixed bond-lengths and angles, showing GABA to be mainly folded in the gas-phase. We obtained a similarly folded GABA molecule with our fixed-geometry MNDO calculations (Figure 2.2). (EHT calculations (Kier and Truitt, 1970) show the isolated GABA molecule to be fully extended, which is clearly wrong!)

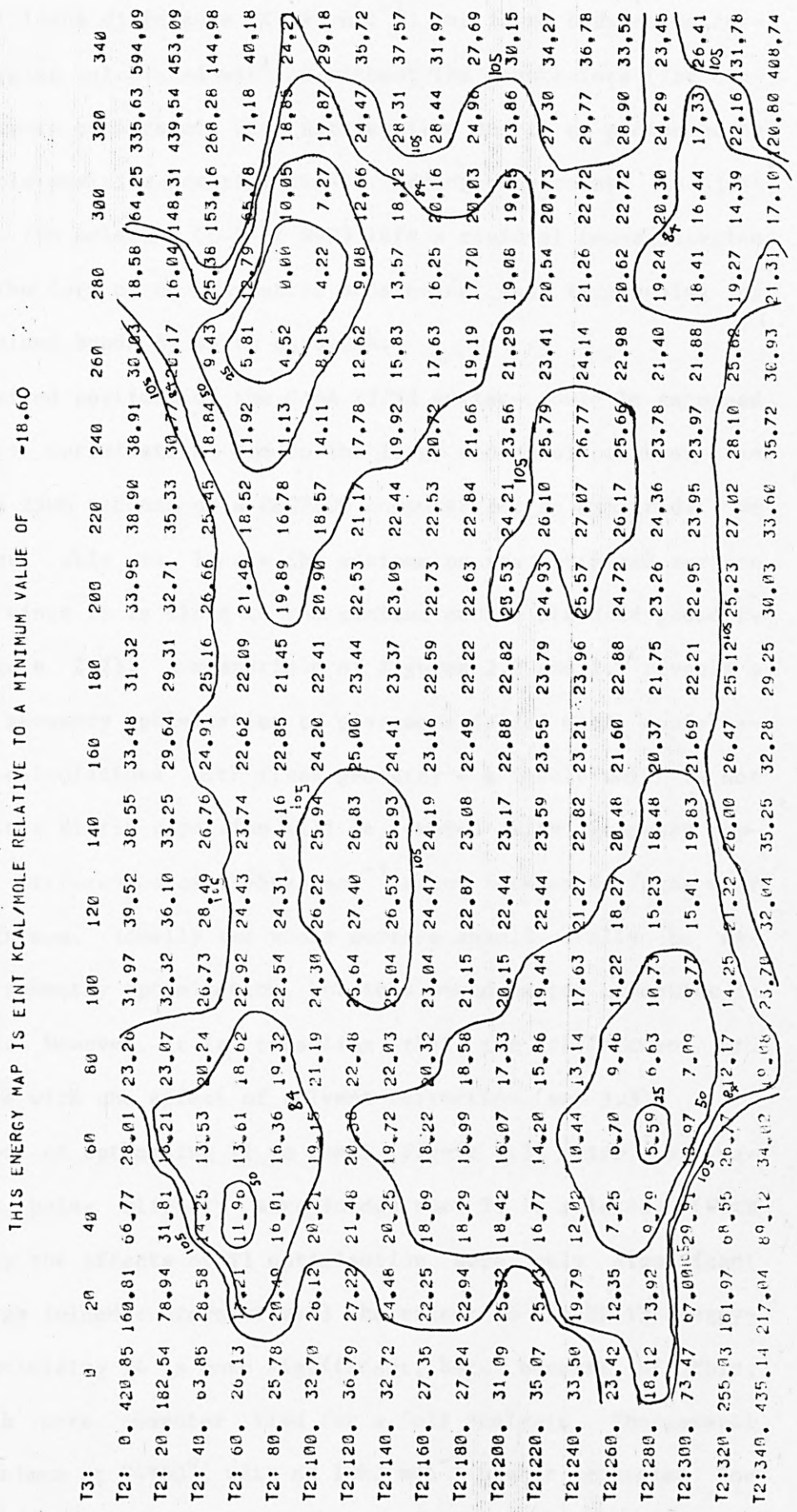
With the MNDO method we were able to examine the effect of geometry optimisation on conformational energies\*. This was hampered, however, by a tendency of the GABA zwitterion to change to the non-zwitterion form through proton migration, particularly in the more folded conformations. To prevent this, the simple geometrical constraints of keeping all the N-H bond lengths equal and all the C-O bond lengths equal were placed on the GABA molecule. The constraints were

---

\* Other methods are unsuited to this purpose for reasons already given.



T4=60°, T1 relaxed. Contours are in kJ mol<sup>-1</sup>.



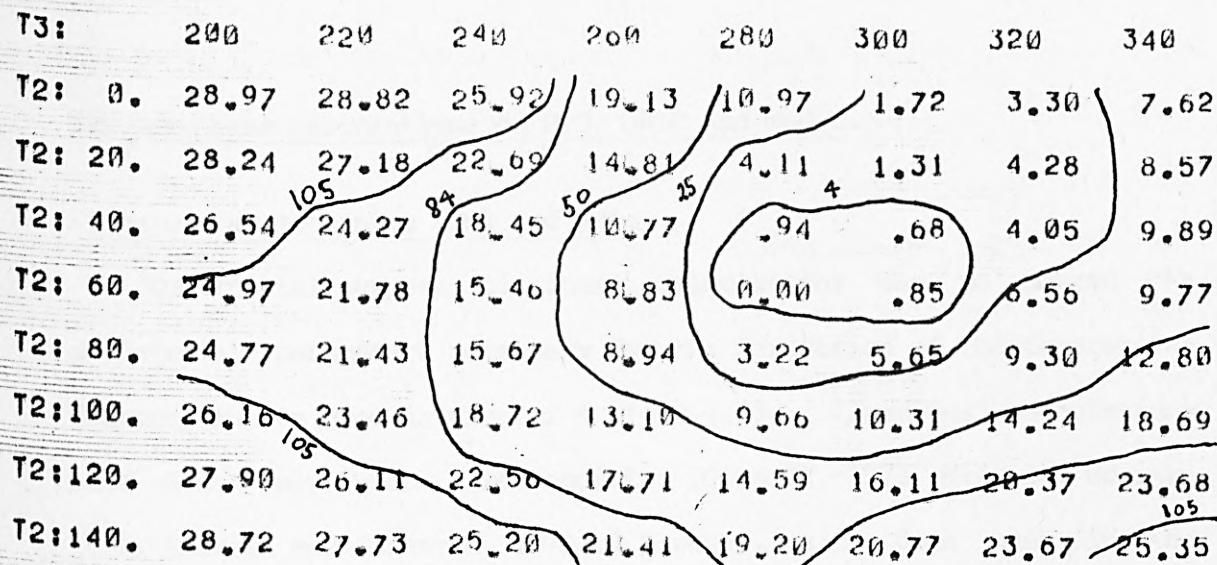


found to make little difference to the **zwitterion** energy from the fact that no significant difference ( $< 1\text{kJ mol}^{-1}$ ) was found between conformational energies calculated with or without the constraints (for conformations where constraints were not required due to no proton being in a favourable position for migration). Adding constraints to just one end of the molecule (C-O or N-H) left a residual non-zwitterion tendency in the form of an assymetric shortening and lengthening of the unconstrained bonds by up to ca  $0.08\text{\AA}$ .

Only limited portions of the GABA T2/T3 surface could be examined with geometry optimisation due to the large amount of computer time required - ca 2500 seconds on a CDC7600 computer for an  $8 \times 8$  grid. We were, however, able to locate the minimum on the optimised surface (Figure 2.3) since it is close to the minimum on the standard-geometry surface (Figure 2.2). A comparison of Figures 2.2 and 2.3 reveals a tendency for geometry optimisation to give more folded GABA conformations than calculations with fixed geometry - a result which is not surprising since steric repulsion will be lowered with geometry optimisation. Differences of  $13\text{--}50\text{kJ mol}^{-1}$  occur between surfaces near the global minimum. Ideally the whole surface should really be examined with geometry optimisation, but this would be too expensive in computer time. However, it is consoling that the differences are small compared with the effect of solvent correction (see 3.3).

The effect of optimising T1 is small (Figure 2.3) with, as expected, GABA being slightly more folded when T1 is relaxed. (With fixed geometry the effects of T1 optimisation were only significant for high-energy folded conformations.) The effect on the T2/T3 energy-surface of optimising T4 is more significant, but, because of this, required much more computer time for a full analysis. The general trend is a minimum at  $T4 \approx 60^\circ$ , with ca  $10\text{kJ mol}^{-1}$  higher energies for  $T4 = 40^\circ$  and  $T4 = 80^\circ$ .

a)



b)

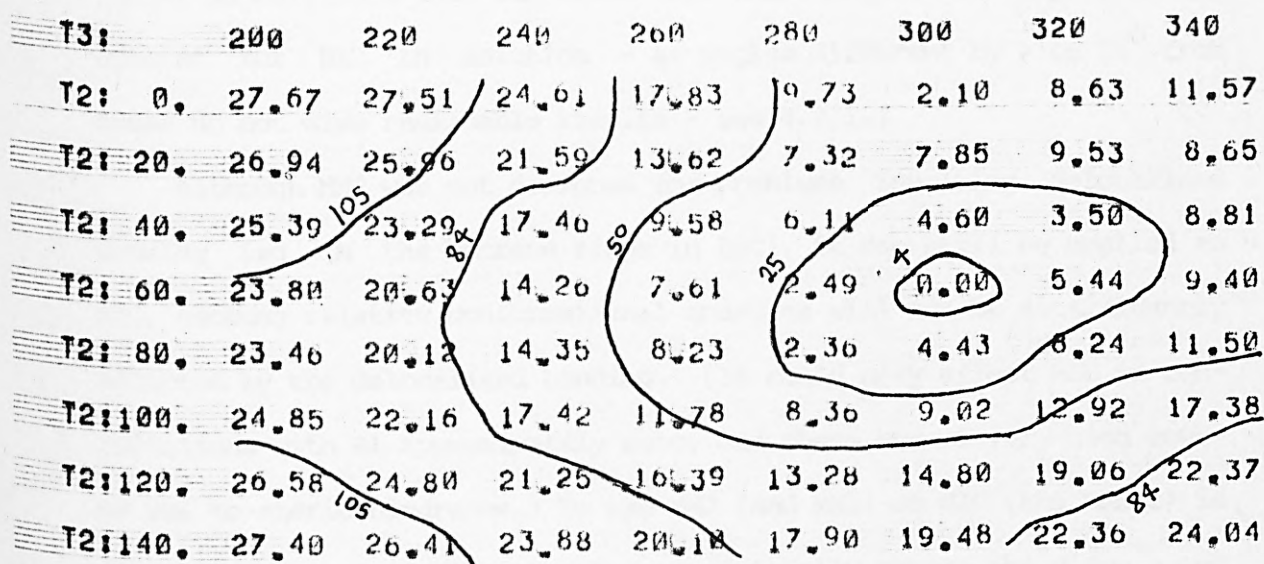


Figure 2.3 GABA geometry-optimised energy surfaces (partial).

a)  $T4=60^\circ$ ,  $T1$  relaxed.b)  $T4=60^\circ$ ,  $T1=180^\circ$ .(Contours are in  $\text{kJ mol}^{-1}$  and the other numbers in  $\text{kcal mol}^{-1}$ .)

## 2.5 Gas-phase calculations on BIC, HBIC and MeBIC.

### 1) Molecular mechanics (MM2 and MM1).

The initial purpose of these calculations was to obtain the geometrical parameters necessary for the derivation of the theoretical chemical shifts used in Chapter 4 (4.2.1.2). (Previous calculations were done only with fixed geometry (Gilardi, 1973, Kier and George, 1973, Andrews and Johnston, 1973).) However, as the data used in the chemical shift analysis was not accurate enough to derive both energies and angles empirically for the  $\Theta_1$  (Figure 1.1) conformational minima, angles from gas-phase calculations were required for the analysis. (Later calculations then confirmed that these angles are approximately correct for BIC in solution - as angles different by  $> ca\ 10^\circ$  from these do not give reasonable results - see 4.2.1.)

Although MM2 was not designed for problems involving delocalised bonding (eg in the benzene rings in BIC), it can still be applied to BIC, because relative conformational energies will not be significantly affected by the delocalised bonding. (It could only affect BIC in conformations with  $\Theta_1$  approximately zero, and these are of very high energy due to steric hindrance.) To use MM2 (and MM1) on BIC (and MeBIC) it was necessary to redefine atom type 3 (initially set to the C for a C=C bond) as 'aromatic' by changing the appropriate parameters for this atom type. The original calculations incorporated values derived from a comparison with existing parameters within MM2, with the C-C unstrained bond length set to  $1.39\text{\AA}$  and the benzene rings planar. Later calculations made use of parameters published by Allinger (1983), with no significant difference found between the results using the different parameter sets.

Some of the parameters for the lactone ring dihedral angles were also not provided in the MM2 programme. These parameters were set such that the ring would be rigid and almost planar (close to the crystal conformation). Initially values from the MM1 programme were used and then, by varying them slightly, the  $\Theta_1$  minima (and BIC bond lengths and angles) were shown to be not significantly dependent on the values of these parameters.

In addition, no parameters were available for the dioxolo rings in BIC. Since these rings have little effect on BIC conformations, the molecule was truncated by removing them. Later calculations, which employed approximate parameters for the ring dihedral angles (making the ring approximately planar - as in the crystal), gave results insignificantly different from the earlier calculations.

MM1 calculations on BIC (Snarey, 1982) give roughly the same angles for the position of the  $\Theta_1$  minima, but with slightly different energies (Figure 2.5).

Our MM2 calculations on the isolated MeBIC molecule show a sharply-bounded low-energy region for  $\Theta_1 = 265^\circ - 300^\circ$  (Figure 2.4). This result is in accord with our solution studies (4.3), in that for MeBIC with no attached counterion  $\Theta_1 = 290^\circ$ . The presence of a tightly-bound counterion changes the conformation to  $270^\circ$  (Figure 4.10). The minima at  $\Theta_1 = 60^\circ$  and  $\Theta_1 = 180^\circ$  are not found in solution (4.3).

All calculations on MeBIC made use of the more recent parameter set.

## 2) MNDO.

Due to a severe restriction on the maximum number of atomic orbitals for any molecule with the MNDO programme (CDC7600: 90, Cray:

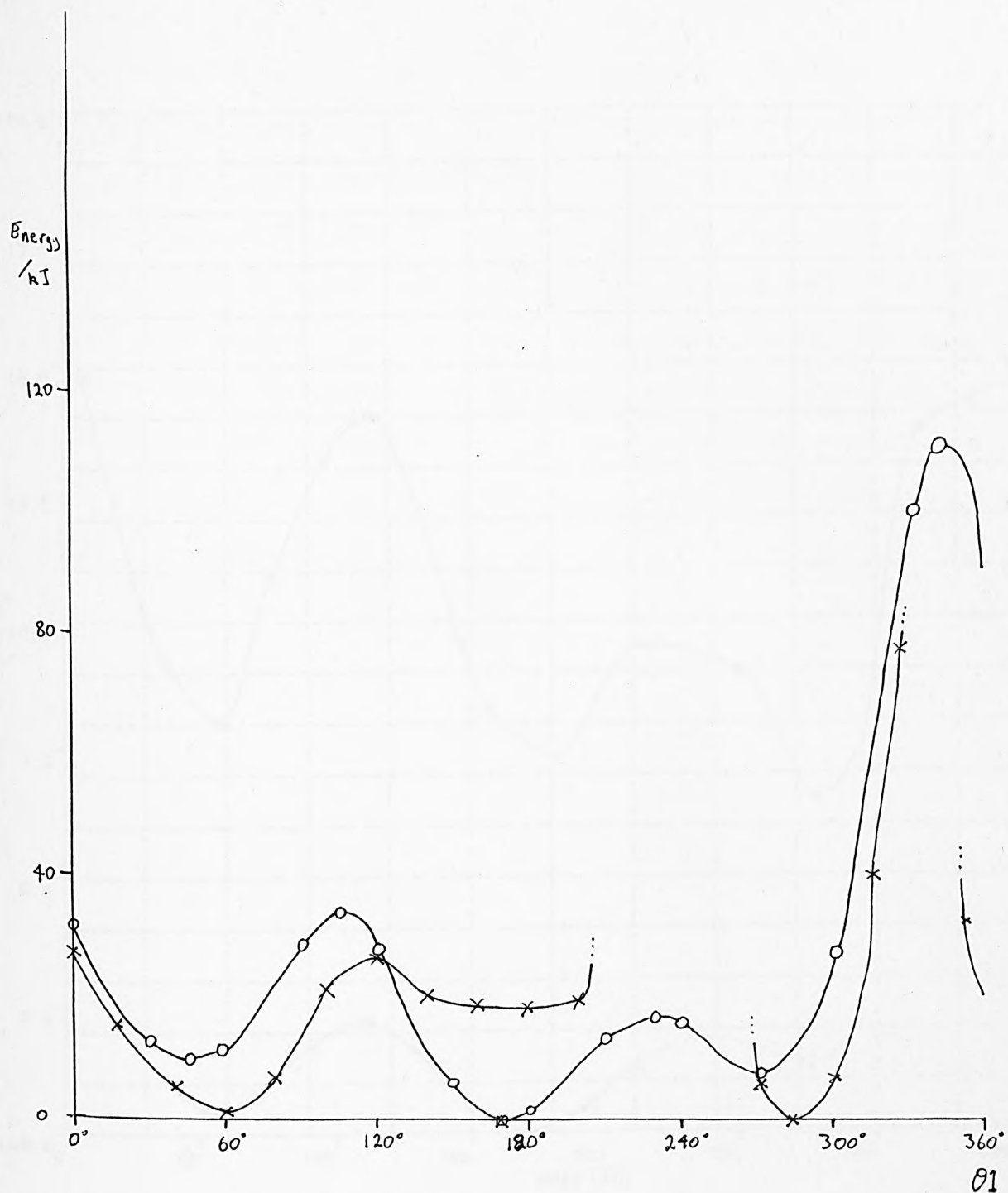


Figure 2.4 BIC and MeBIC energy-surfaces (MM2)

○ BIC,                      x MeBIC



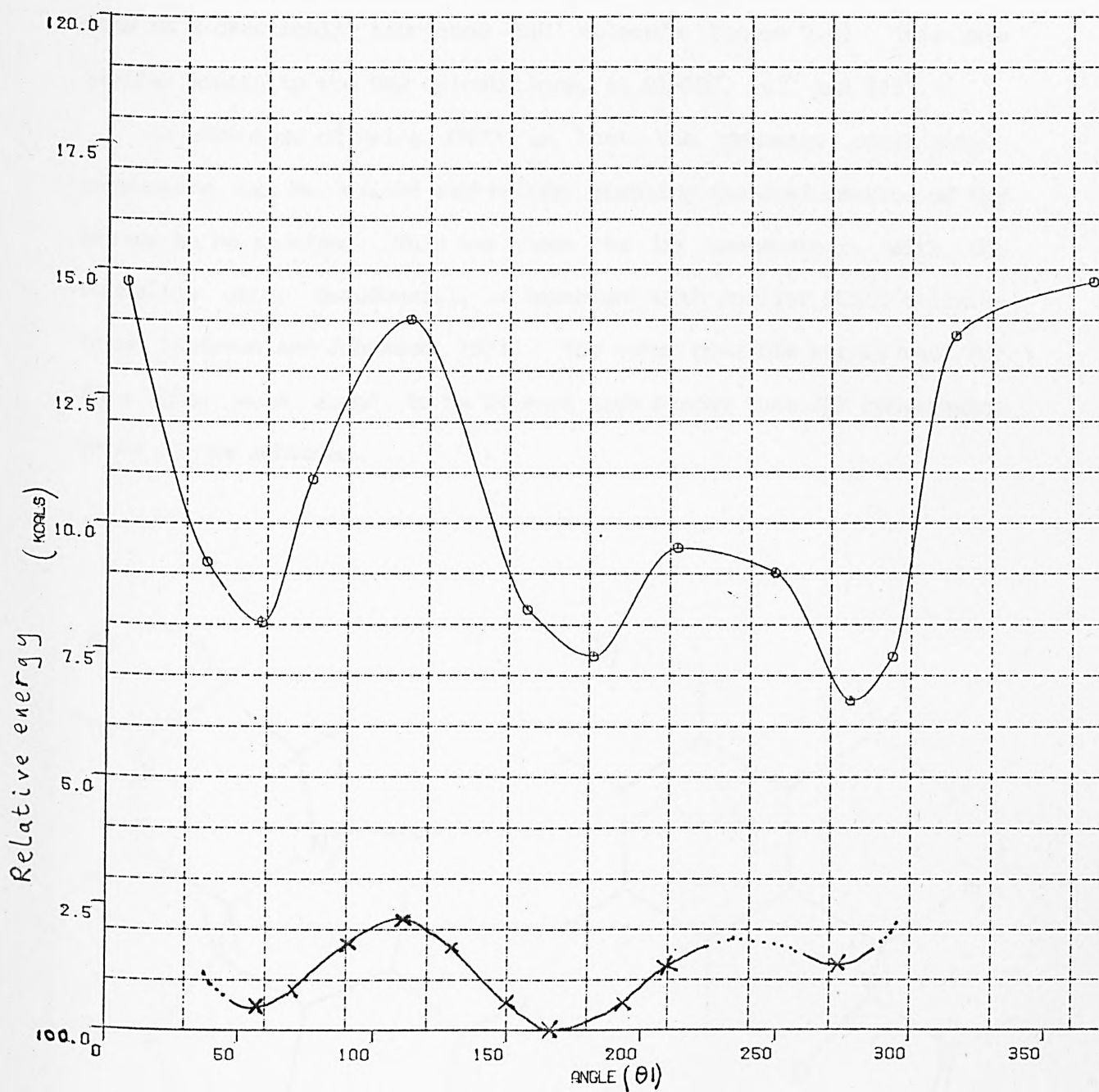


Figure 2.5. BIC  $\theta_1$  energy-surface (MM1 —○—; MNDO —x—).

(Snarey, 1982.)

initially 98, later (Feb. 1986) ca 120), calculations could only be done on a drastically truncated 'BIC' molecule (Figure 2.6). This gave similar minima to the MM2 calculations, at  $\theta_1=58^\circ$ ,  $165^\circ$  and  $273^\circ$ .

An advantage of using MNDO is that the geometry optimisation parameters can be defined explicitly, enabling the conformation of the N-ring to be examined. This was shown to be pseudochair, with the phthalide group pseudoaxial, in agreement with earlier PCILO calculations (Andrews and Johnston, 1973). The other possible structures for this ring were found to be of such high energy that SCF convergence could not be achieved.

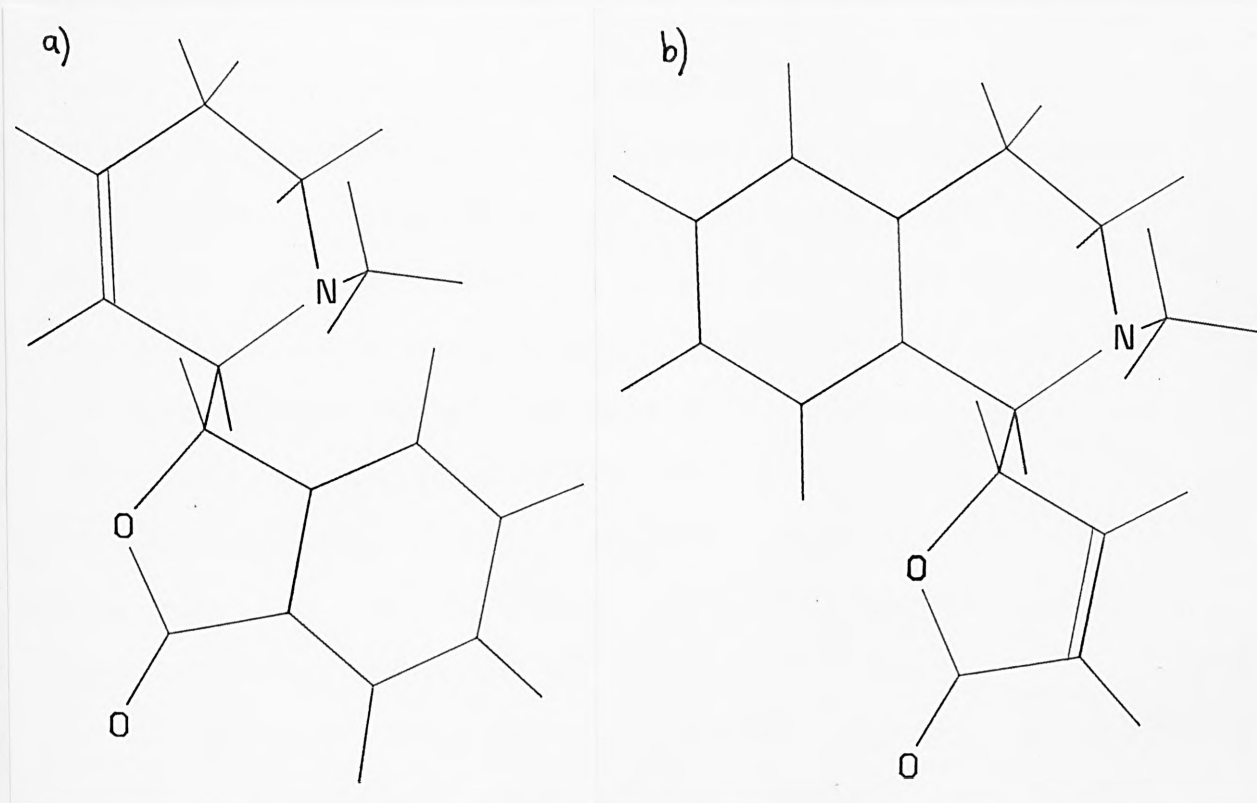


Figure 2.6. Truncated BIC molecules used in MNDO calculations. The energy-surface shown in Figure 2.5 is for molecule (b). Similar results were obtained with (a).

## 2.6 Summary.

(i) From a survey of the literature and from our gas-phase calculations on small test molecules the best currently available theoretical method for the calculation of relative molecular energies and charges (for the molecules of interest to us - eg GABA and BIC), in terms of speed and accuracy (as compared with available experimental data), is MNDO (2.2).

In our MNDO calculations (with geometry optimisation) on isolated molecules containing single rotations, the average departure from experimental conformational-energies was  $3 \pm 3 \text{ kJ mol}^{-1}$  (Table 2.2). MM2 also performs very well (Table 2.2), but parameters must be available for the molecule of interest, and delocalised bonding can only approximately be accounted for - which makes MM2 unsuited to GABA. These results are comparable with the variation in experimentally determined energies for the molecules used in the tests, and justify the use of theoretical methods where experimental data cannot be obtained.

However, care is needed in using theoretical methods because even the best of these can sometimes give results which are totally inaccurate, such as giving the wrong conformation as the most stable one (eg Figure 2.1, and see Radom et al, 1985), and is best dealt with by using more than one theoretical method and, whenever possible, reference to experimental data for similar molecules.

MM2 is also useful for pre-refinement of crude molecular coordinates which would either fail to converge or converge slowly with MNDO.

(ii) Geometry optimisation must be considered in conformational studies, even if only for small portions of the conformational-energy surface. If charges only are required then geometry optimisation is

not necessary, but is probably wise if the input geometry is crude or high accuracy is required (geometry optimisation makes on average  $< 0.05\text{eu}$  difference to atomic charges). A comparison of dipole moments calculated with and without geometry optimisation, and with experimental dipoles, shows that in general dipole moments calculated with geometry optimisation are slightly better than those calculated with fixed geometry (Table 2.1).

(iii) From MNDO calculations (with and without geometry optimisation) on the GABA zwitterion we found that it is highly folded in the gas-phase, with little flexibility (2.4). Our results are qualitatively similar to earlier results with fixed-geometry, and using the CNDO (Warner and Steward, 1975) and PCILO (Pullman and Berthod, 1975) methods. Differences of  $13 - 50 \text{ kJ mol}^{-1}$  were found between MNDO surfaces with fixed geometry and with geometry optimisation.

(iv) Our MM2 calculations on BIC yield 3 low-energy minima in the gas-phase at  $\theta_1=45^\circ, 170^\circ$  and  $270^\circ$ , with low energy barriers ( $7\text{kJ mol}^{-1}$  and  $10\text{kJ mol}^{-1}$ ) between the minima (2.5), and MNDO calculations on BIC fragments give roughly the same minima. Our MM2 calculations on MeBIC show 3 sharply-bounded conformations at  $\theta_1 \approx 60^\circ, 170^\circ$  and  $285^\circ$ . (The third minimum corresponds to that found in solution (see 4.3).)

## Part 1 - Receptors of unknown molecular structure.

### 3 The conformations of GABA in solution.

#### 3.1 Introduction.

In searching for the preferred conformations of GABA in solution there is conflict between the various experimental (3.2) and theoretical (3.3) approaches as to whether GABA is flexible in solution with equally populated conformational minima (eg the spectroscopic method of Tanaka et al, 1978, or the theoretical 'Supermolecule' method used by Pullman and Berthod (1975)), or fairly rigid in just one conformation (from experimental dipole moments: Edward, Farrell and Job, 1973, or from the theoretical 'SOLVEFF' method of Clarke (1976)). There is even disagreement between the theoretical calculations cited above, as to whether GABA is more flexible in solution than in the gas-phase (see 3.3). Theoretical calculations (Warner and Steward, 1975, Pullman and Berthod, 1975, and see 2.4) show that the GABA zwitterion is locked into highly folded conformations in the gas-phase<sup>\*</sup> due to attraction between the oppositely charged  $\text{COO}^-$  and  $\text{NH}_3^+$  groups (Figure 3.1). In polar solvents, the GABA zwitterion should become more extended and more flexible because the solvent molecules will interact with the charged groups and substantially reduce the electrostatic interaction between them: more low-energy conformational minima and/or lower barriers between minima should appear. (For GABA, a low-energy conformation is defined as any conformation within the GABA binding energy of the global minimum (ca  $40\text{kJ mol}^{-1}$  - calculated from a GABA  $K_D$  of  $335\pm 39\text{nm}$  (Lloyd et al,

<sup>\*</sup> With the exception of extended Huckel calculations which show GABA to be in the extended conformation (Kier and Truitt, 1970).



1977)), as any conformations of higher energy than this can be disregarded.)

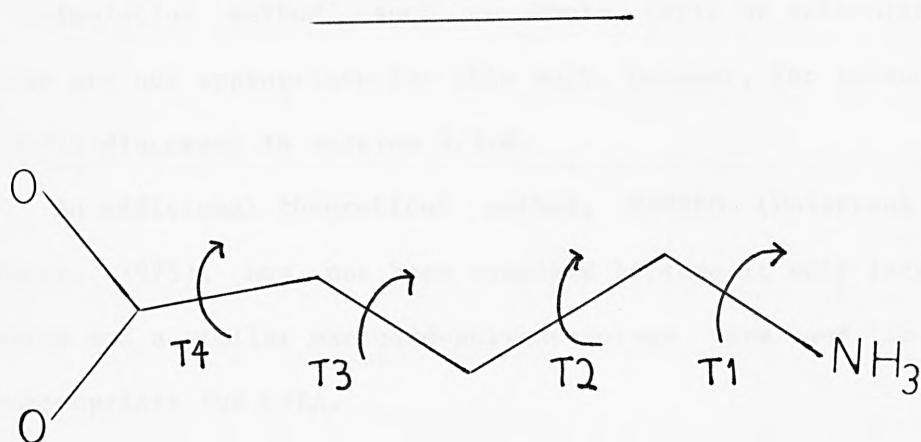


Figure 3.1. GABA molecular structure. The rotatable bonds are labelled T1-T4.

Because of the confusion over GABA flexibility we therefore studied (i) the flexibility of GABA in solution (using variable-temperature NMR) (3.2.2), and (ii) the accuracy of 2 opposing theoretical models for dealing with solvent effects - the 'Super-molecule' method (Pullman and Pullman, 1975), a discrete model in which close-range solute/solvent interactions are considered (3.3.1), and the SOLVEFF method (Clarke, 1976), a continuum-model which covers long-range solvent effects (3.3.2). A hybrid approach, with the advantage of including both short and long-range interactions (Beveridge and Schnuelle, 1974), was also briefly examined in an attempt to improve upon the conflicting results found by the other 2 methods (3.3.3).

Another type of theoretical approach to dealing with solvation is a 'simulation method' such as Monte Carlo or molecular dynamics. These are not appropriate for this work, however, for reasons that are briefly discussed in section 3.3.4.

An additional theoretical method, CAMSEQ (Weintraub and Hopfinger, 1975), has not been examined because it only incorporates a cavity and a similar excluded-solvent-volume term and is therefore inappropriate for GABA.

A measure of the flexibility of GABA is given by the  $\text{NH}_3^+/\text{COO}^-$  charge separation ( $x_T$ ) probability distribution and is used in Figure 3.2 to summarise results for various methods of determining GABA solution conformations. In addition, useful SAR correlations have been found using  $x_T$  probability distributions (Steward and Clarke, 1975), and  $x_T$  is later compared with a similar 'arrangement of charge centres' parameter in Chapter 5.

### 3.2 Spectral and dipole moment studies on GABA solution conformations.

#### 3.2.1 Previous work on GABA conformational preferences.

Results reported by Ham (1974) and by Tanaka and coworkers (1978) using NMR proton coupling constants show GABA to be flexible in aqueous solution with a series of conformational modes (Figure 3.2). These results, however, were based on averaged GABA NMR spectra at a single temperature and assumed classical gauche/trans conformers ( $T_2/T_3=180^\circ$ ,  $\pm 60^\circ$ ) and associated coupling constants. Theoretical calculations (Pullman and Berthod, 1975, Clarke, 1975 and see 3.3) show that the conformational minima are not necessarily close to the gauche/trans positions (see Figure 3.5). The above NMR results were additionally based on the assumption of the gauche conformers being of

equal energy, which was found not to be true for both Supermolecule and SOLVEFF calculated energies (see Figures 3.5 and 3.8).

Tanaka and coworkers (1978) also used Raman spectroscopy on GABA, in a nujol mull and in solution, to demonstrate that additional conformations of GABA are found in solution which are not present in the solid\*.

In contrast to these results dipole moment measurements (Edward, Farrell and Job, 1973) gave only extended conformations in solution. The possible cause of the discrepancy between results derived from dipole moment measurements and from NMR and Raman spectroscopy is discussed in section 3.3.2 i.

\* Angles T2 and T3 (Figure 3.1) were assigned 'trans-gauche' for the solid. The extra conformations in solution could not be assigned (Tanaka et al, 1978). In the crystal,  $T2=175^\circ$  and  $T3=-73^\circ$  (Steward, Player and Warner, 1973).

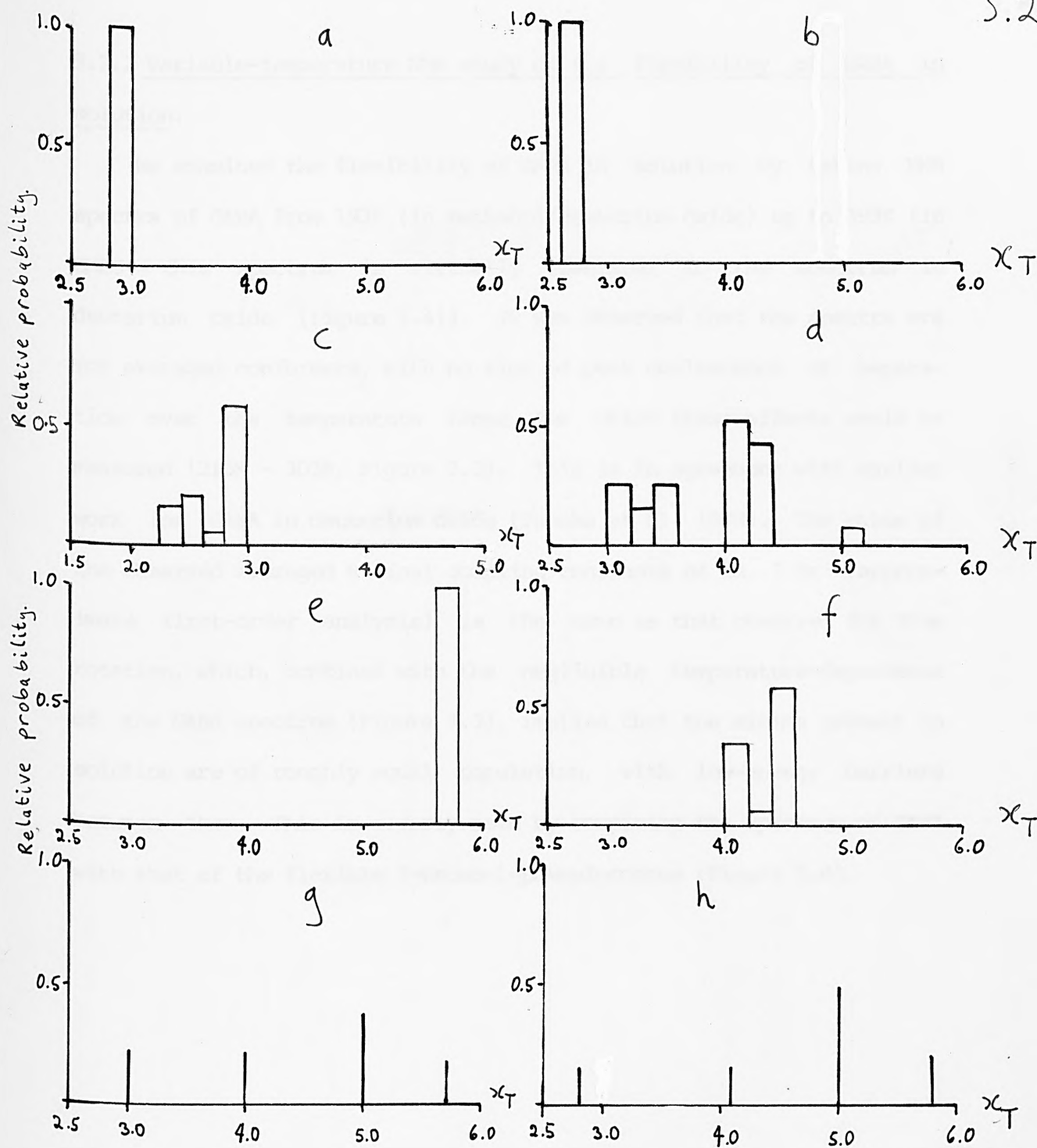


Figure 3.2. GABA  $x_T$  distributions by various methods.

Gas phase: a)CNDO (Warner and Steward, 1975), b)PCILO (Pullman and Berthod, 1975), c)MNDO+GO (partial surface). Theoretical solution methods: d)Supermolecule (Pullman and Berthod, 1975), e)SOLVEFF (MNDO+GO), f)Hybrid (partial surface). NMR: g)Tanaka et al (1978), h)Ham (1974).

### 3.2.2 Variable-temperature NMR study on the flexibility of GABA in solution.

We examined the flexibility of GABA in solution by taking NMR spectra of GABA from 193K (in methanol/deuterium oxide) up to 393K (in DMSO - this spectrum is virtually identical to the spectrum in deuterium oxide (Figure 3.4)). It was observed that the spectra are for averaged conformers, with no sign of peak coalescence or separation over the temperature range for which these effects could be measured (213K - 303K, Figure 3.3). This is in agreement with earlier work for GABA in deuterium oxide (Tanaka et al, 1978). The value of the observed averaged vicinal coupling constants of ca 7 Hz (approximate first-order analysis) is the same as that observed for free rotation, which, combined with the negligible temperature-dependence of the GABA spectrum (Figure 3.3), implies that the minima present in solution are of roughly equal population, with low-energy barriers between them. This is clearly seen by comparing the spectrum of GABA with that of the flexible 3-bromo-1-phenylpropane (Figure 3.4).



Figure 3.3 NMR spectrum of GABA in  $DMSO-d_6$  at 300K. The spectrum shows sharp, well-resolved peaks, indicating a rigid structure or slow rotation at this temperature.

The peak broadening in the GABA spectrum observed at low temperatures is not correlated with conformational effects but is due to the high viscosity of the solvent near the melting point. (See Figure 3.4 for spectra below 213K.)



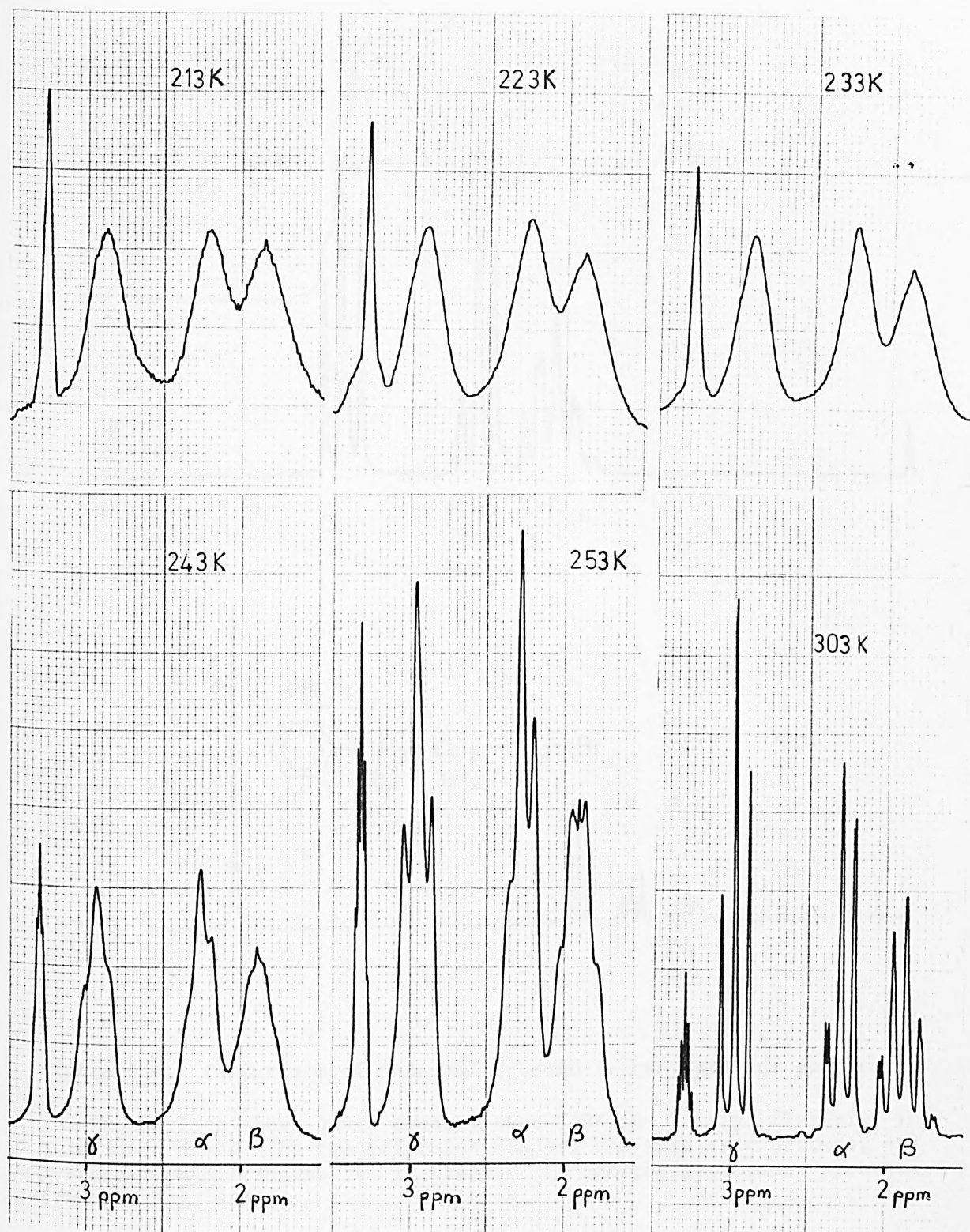


Figure 3.3 NMR spectrum of GABA in  $D_2O/CD_3OD$  at different temperatures.

The peak broadening in the GABA spectrum observed at low temperature is not consistent with exchange effects but is due to the high viscosity of the solvent near the melting point. (See Appendix A2 for spectra below 213K.)

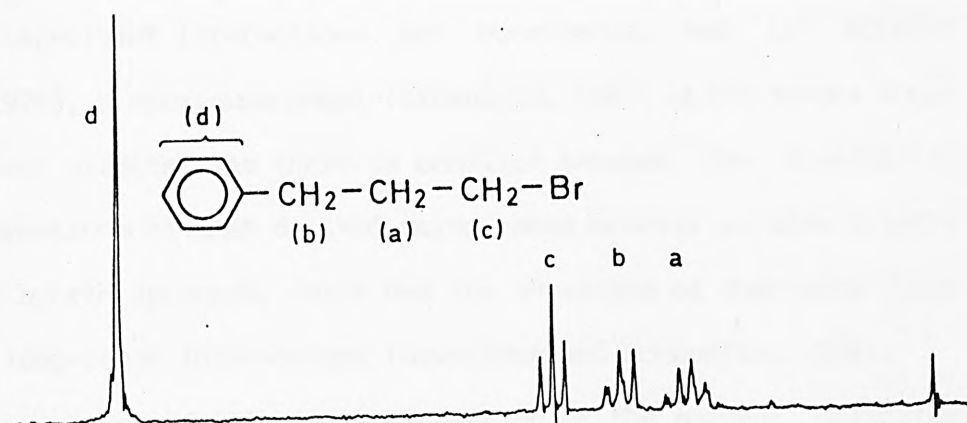
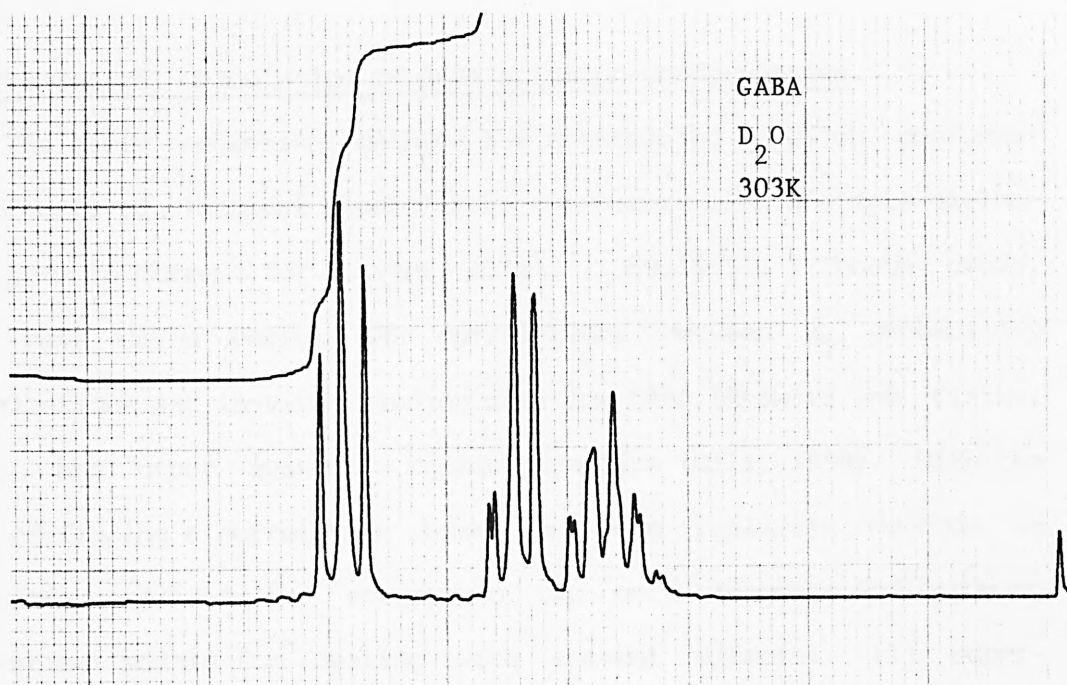


Figure 3.4. Comparison of GABA and 3-bromo-1-phenylpropane spectra.

This comparison shows quite clearly that GABA is fairly flexible in solution, since the two molecules display the same spin pattern, and 3-bromo-1-phenylpropane is flexible. (Bromo-phenylpropane spectrum from Gunther, 1980.)

### 3.3 Theoretical calculation of GABA solution conformations.

The early theoretical methods for determining solution conformations employed classical gauche/trans conformers with a simple dielectric term to correct for solvent (Gill, 1959,1965). Though crude, they have given useful SAR correlations between  $x_T$  probability distributions and in-vitro binding data for GABA (Steward and Clarke, 1975), and other systems - eg acetylcholine (Gill, 1959). With the hope of finding a correlation based on more realistic methods we therefore examined the accuracy of two more elaborate and opposing theoretical models for dealing with solvent effects: (1) Supermolecule (Pullman and Pullman, 1975), a discrete model in which close-range solute/solvent interactions are considered, and (2) SOLVEFF (Clarke, 1976), a continuum-model (Sinanoglu, 1967) which covers long-range solvent effects. As there is conflict between the flexibility and conformations of GABA derived using these models, we have briefly examined a hybrid approach, which has the advantage of combining both short and long-range interactions (Beveridge and Schnuelle, 1974).

The importance of allowing the position of the solvent molecules to vary with solute-molecule conformation and the excessive dominance of the electrostatic interaction term ( $E_{es}$ ) in SOLVEFF are included in this work as we found these parameters to have a greater effect on GABA conformation than the effect of combining the two approaches.

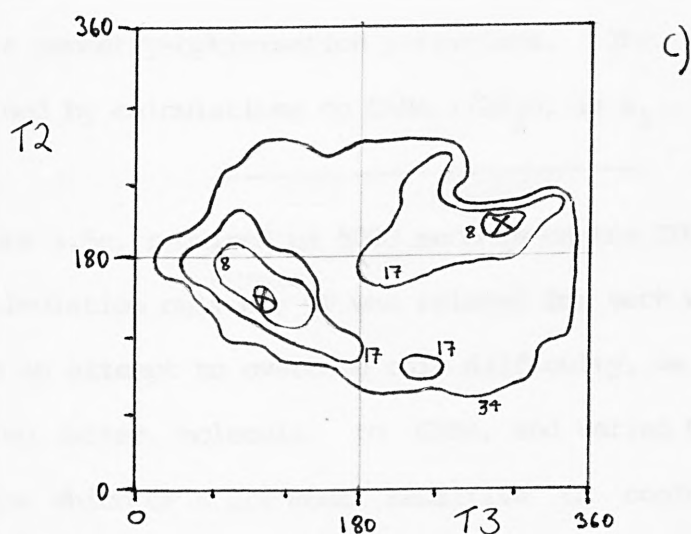
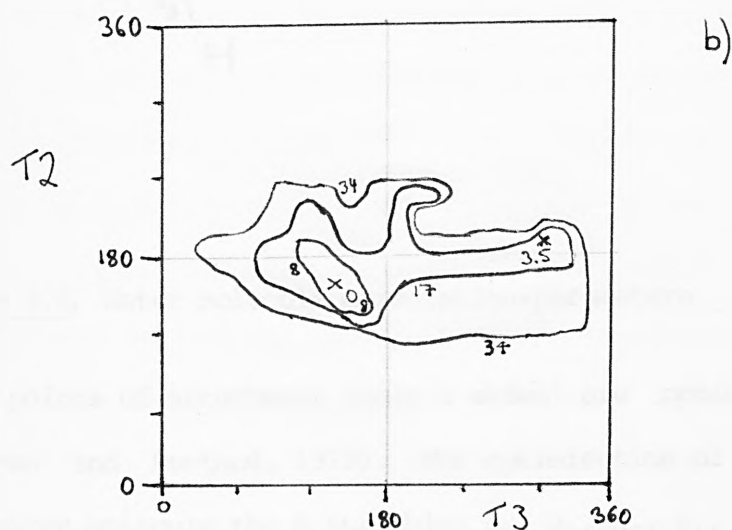
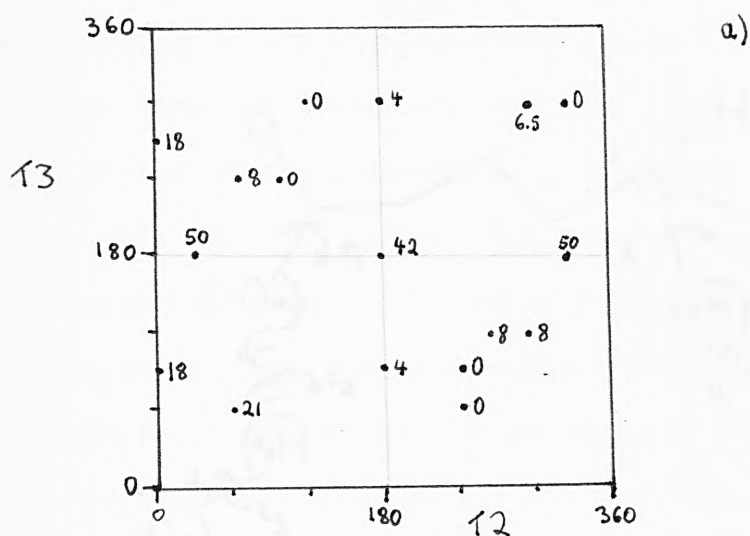
#### 3.3.1 The Supermolecule model.

With the Supermolecule model (Pullman and Pullman, 1975), close range solute-solvent interactions are accounted for by adding several solvent molecules to the isolated drug molecule to approximate the 'first hydration shell' (Pullman and Pullman, 1975). Quantum mechanical (or classical) calculations are then performed on the

resultant 'supermolecule'. A second and further hydration shells are not usually included because the supermolecule would become too large for practical computation and would not necessarily be any more accurate as the number of unknowns is also increased (Pullman, 1974).

For GABA (with six water molecules attached) Pullman and Berthod (1975) found several low-energy conformations (Figure 3.5) with a marked increase in flexibility over the gas-phase molecule, and a slight tendency towards more extended conformations (Figure 3.2): this is roughly in agreement with NMR and Raman spectroscopy (3.2). In that Supermolecule work the water molecules were placed in the optimum positions determined by minimal STO-3G basis-set ab initio calculations on GABA fragments (alkyl ammonium salts (Port and Pullman, 1973) and the formate ion (Port and Pullman, 1974)), but with the orientations and points of attachment of the water molecules not being allowed to change with changing GABA conformation. The effect of allowing the water molecule orientation parameters (OP's - see Figure 3.6) to vary with GABA conformation, though complicated (Beveridge et al, 1974), must be accounted for (see below).

In principle the theoretical method used for the conformational analysis should also be used for positioning the water molecules, with optimisation of the OP's for each conformation, with a check that the H-bond distances and angles remain within accepted (Kroon and Kanters, 1975) limits. One problem is that the 'GABA Supermolecule' with six waters represents a very large number of variables (six for each water molecule, excluding optimisation of the O-H bonds and angle in each water molecule), and would require a huge amount of computer time for a full conformational analysis - even without geometry optimisation on the GABA molecule itself. (The 18x18 conformational-energy grid shown



**Figure 3.5** GABA Supramolecule surfaces ( $\text{kJ mol}^{-1}$ ).

a) Using PCILO (Pullman and Berthod (1975)).

b) MNDO with fixed geometry.

c) MNDO with limited optimisation of water orientations.

(Minima are shown at • and ×.)



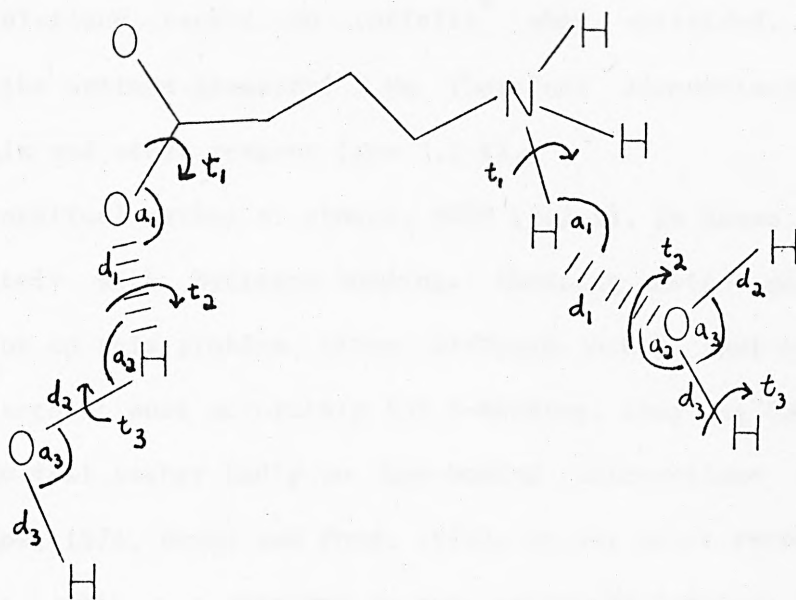


Figure 3.6. Water molecule orientation-parameters.

The 6 points of attachment (only 2 shown) are optimal for H-bonding (Pullman and Berthod, 1975). For optimisation of the orientation of each water molecule the 6 variables  $d_1$ ,  $a_1$ ,  $a_2$ ,  $t_1$ ,  $t_2$  and  $t_3$  are considered 'orientation parameters' (OP's) and the other 3 are solvent-molecule geometry-optimisation parameters. The most important OP, determined by calculations on GABA.1/2H<sub>2</sub>O, is  $a_1$ .

in Figure 3.5c, required ca 5000 seconds on the CDC7600 computer. In that calculation only one OP was relaxed for each water molecule.)

In an attempt to overcome this difficulty, we added first one and then two water molecules to GABA, and varied GABA conformation to determine which OP's are most sensitive to conformational changes. However, in attempting this two problems arose:

(i) Using PCILO on GABA plus one or two attached water molecules the GABA-water distance tended to infinity\* when optimised. This is clearly not the optimum geometry! We therefore discontinued using PCILO for this and other reasons (see 2.2 i).

(ii) Our theoretical method of choice, MNDO (2.2 i), is known not to deal adequately with hydrogen bonding. Choosing another method was not a solution to this problem, since although other semi-empirical methods may account more accurately for H-bonding, they (eg CNDO/2 and INDO) tend to fail rather badly on non-bonded interactions (Gregory and Paddon-Row, 1976, Dewar and Ford, 1979), or for other reasons (see 2.2 i). This problem was overcome to some extent by keeping all the O...H H-bond distances ( $d_1$  - see Figure 3.6) fixed at the average value of  $1.9\text{\AA}$  (Kroon and Kanters, 1975). In our calculations with 1 and 2 attached water molecules the relaxed value of  $d_1$  was ca  $2.9\text{\AA}$  for O...HN and ca  $3.4\text{\AA}$  for H...OC, with higher values (up to  $+1\text{\AA}$ ) found for high-energy crowded conformations. (MNDO predicts very long H-bonds because it has a tendency to make H-bonds much too weak (Dewar and Thiel, 1977).) More importantly we found that by far the most conformation-sensitive OP is the GABA-water H-bond angle ( $a_1$ ), which varies from ca  $90^\circ$  to  $180^\circ$  on changing GABA conformation. A surface with  $a_1$  relaxed for each water molecule is given in Figure 3.5c, and is somewhat flatter (GABA more flexible) than the fixed-geometry surface (Figure 3.5b). This qualitatively describes the effect of optimising just one OP (for each water), and shows that the fixed-geometry and partially-optimised Supermolecule surfaces are at best only semi-quantitative.

\* To verify that it was the PCILO method that was at fault, and not our added geometry optimisation (see 2.2 i), we performed calculations on the completely different polar system - 1,2 difluoroethane/water - and obtained the same result.

One advantage of the Supermolecule method is that the tendency towards the GABA non-zwitterion form observed on optimising GABA geometry (this effects the isolated molecule (see 2.4) and therefore also the total SOLVEFF conformational energies - see 3.3.2) no longer exists because the zwitterion is stabilised by the presence of the water molecule shell. However, geometry optimisation on GABA within the GABA supermolecule is somewhat impracticable!

### 3.3.2 The SOLVEFF model.

A continuum model, where the solute molecule is considered as lying in a spherical or spheroidal cavity within a dielectric solvent-continuum was first applied to conformational problems by Sinanoglu (1967) and was taken by Clarke (1976) as a basis for his solvent effect programme SOLVEFF. (The bulk dielectric effects for continuum models are based on the original calculations of Onsager (1936).) The total energy for each conformation of the solute molecule is obtained by adding solvent correction terms to the isolated-molecule energy

$$(E_{\text{isol}}): E_{\text{tot}} = E_{\text{isol}} + E_{\text{es}} + E_{\text{dis}} + E_{\text{cav}}.$$

For highly polar molecules such as GABA  $E_{\text{es}}$  is by far the most dominant term, and is therefore described in detail below. For the other, much smaller, cavity and dispersion terms, a detailed discussion can be found in Clarke's thesis (1976).

On applying SOLVEFF to GABA only extended conformations were found (Figure 3.2), which is at variance with the experimental results derived from NMR and Raman spectroscopy (3.2). In an attempt to improve upon Clarke's results we therefore examined the following:

- (i) ways of reducing the excessive dominance of the electrostatic interaction term ( $E_{\text{es}}$ ).

(ii) A comparison of the sphere and spheroid models for the whole GABA conformational energy surface (all quoted SOLVEFF results were derived using the more realistic spheroid model, unless otherwise stated.)

(iii) The use of MNDO instead of CNDO/2 for calculating the isolated GABA molecule energies and dipole moments.

(iv) the effect of geometry optimisation on GABA conformational energies (most of the calculations were with fixed standard (Pople and Beveridge, 1970) geometry since geometry optimisation has little effect on SOLVEFF energies). The effects of T1 and T4 relaxation on T2/T3 conformational energy surfaces are included in this section.

(v) Possible extensions to the SOLVEFF software - a modified version of Clarke's SOLVEFF programme was used for all our SOLVEFF work.

(vi) The advantages of SOLVEFF over Supermolecule.

(i) The dominance of the  $E_{es}$  term in SOLVEFF.

By far the most dominant term in SOLVEFF is  $E_{es}$ , the electrostatic interaction term, which for the spheroid model can be expressed (Clarke, 1976):

$$E_{es} = -90.21 m_0^2 F(A) / (ab^2) \text{ kJ mol}^{-1}$$

where (Buckingham, 1953a):

$$F(A) = A(1-A)(e-1)(1+(n^2-1)A)^2(e-(e-1)A) / (e+(n^2-e)A)^2$$

and  $m_0$  is the isolated molecule dipole moment,  $n$  is the refractive index of the solute ( $n^2=2.5$ , Clarke, 1976, Beveridge et al, 1974),  $e$  is the dielectric constant of the solvent,  $a$  is the semi-axis of the spheroid,  $b$  is the radius of the spheroid at its equator and  $A$  is an internal field factor.

For highly polar solvents  $e \gg 1$  and  $e \gg n^2$ , which leads to the much simpler expression for  $F(A)$  of:

$$F(A) \approx A(1+(n^2-1)A)^2 = A(1+1.5A)^2 *$$

As the variation of  $A$  (a function of the  $a/b$  ratio) with GABA conformation is small (from ca 1.5 to ca 2.5\*\* for the extremes of conformation) compared with the large variation of GABA dipole moment (ca 4 Debye - ca 28 Debye), then  $E_{es}$  for GABA is approximately proportional to minus the square of the dipole moment, which is a sharply varying function of GABA conformation. (MNDO calculations give a GABA dipole moment of from ca 5 Debye in folded conformations up to a maximum at extended conformations of ca 28 Debye.) For the sphere model this dependence of  $E_{es}$  on  $m_o^2$  is even stronger as there are no correcting internal field factors with this model.

The problem is that for GABA in extended conformations the N-H and C-O bond dipoles combine to give a high net dipole, but in folded GABA conformations the dipoles are aligned roughly opposite to one another giving a falsely low net dipole. (The internal field factor is designed to account for this, but appears to be somewhat inadequate.) A single molecular dipole moment is therefore inadequate for describing the polarity of a molecule such as GABA. Buckingham's original equations (1953a, 1953b) were designed for polar molecules in solvents of **low polarity**. The theory works for such cases since the above simplified expressions for  $F(A)$  no longer apply (since  $e \gg 1$  is no longer true).

\* We verified that  $E_{es}$  is independent of  $e$  for high  $e$  values by calculating  $E_{es}$  with  $e=20$  and with  $e=80$  and observing little difference between the  $E_{es}$  energies.

\*\* Calculated by applying the maximum variation of the  $a/b$  ratio for GABA to the equations (Clarke, 1976, Osborne, 1945) for the components of  $A$  parallel and perpendicular to the spheroid axis.



A model is required which more accurately allows for the changes in solute molecule polarity with conformation, taking into account that for GABA two dipoles are involved and not just an average value. The fact is that microscopic effects, at least for the solute molecule, need to be somehow correctly taken into account. A hybrid SOLVEFF/Supermolecule model (3.3.3) appeared at first to have solved the problem, with an observed overall lowering of the GABA 'super-molecule' average dipole moment. However, since extra variables - the water molecule bond dipoles - are being added, this simply makes the problem far more complicated! (see 3.3.3).

In view of all the above it is not surprising that the conformation of GABA derived from experimental dipole moment measurements (Edward et al, 1973) is similar to that derived by SOLVEFF, since they are both based on the same theoretical arguments of Buckingham (1953).

(ii) The sphere verses the spheroid model.

The more realistic model is the solute molecule lying in a spheroidal cavity which changes shape with the conformation of the molecule. We found no significant difference in GABA flexibility between the two models (Figure 3.7). The small difference between the flexibility predicted using the sphere and spheroid models supports the above arguments for  $E_{es} \approx -km_0^2$ . The only significant difference between the two models is that the position of the T4 global minimum is shifted from  $T4=90^\circ$  (sphere) to  $T4=60^\circ$  (spheroid). This difference is not surprising in view of the different treatment of GABA geometry.

(iii) MNDO compared with the earlier use of CNDO/2.

The SOLVEFF conformational energy surfaces for GABA using the CNDO and MNDO methods (2.2) are qualitatively very similar, each

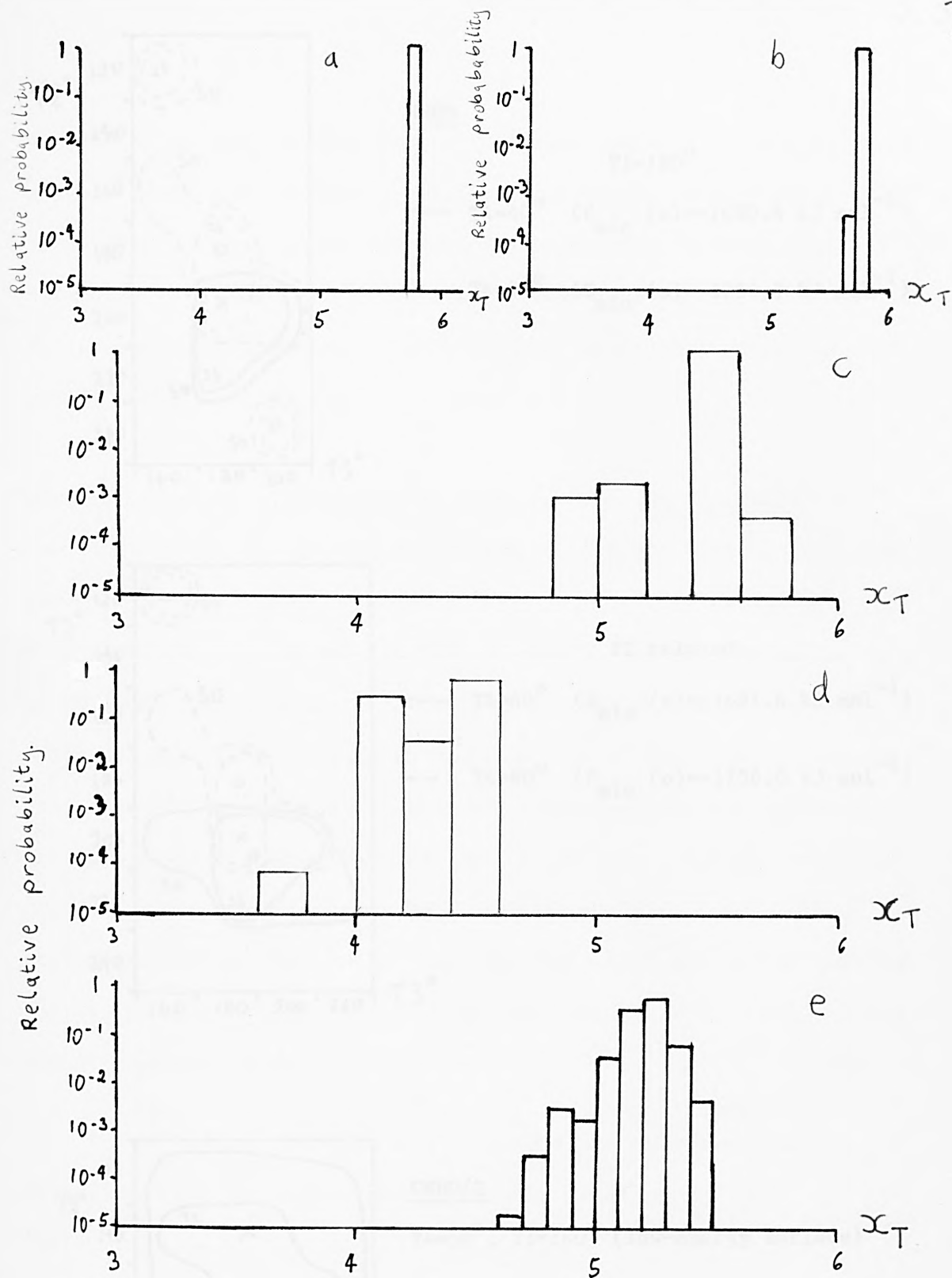
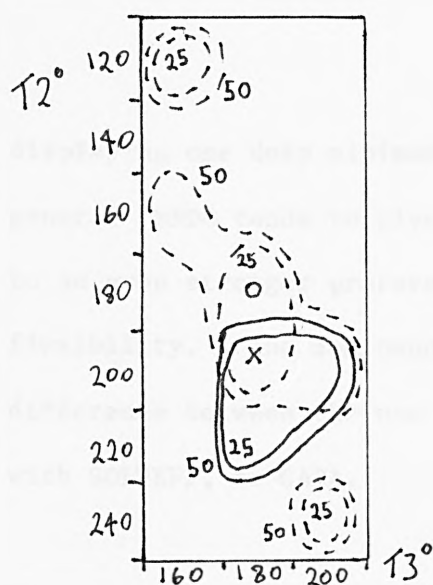


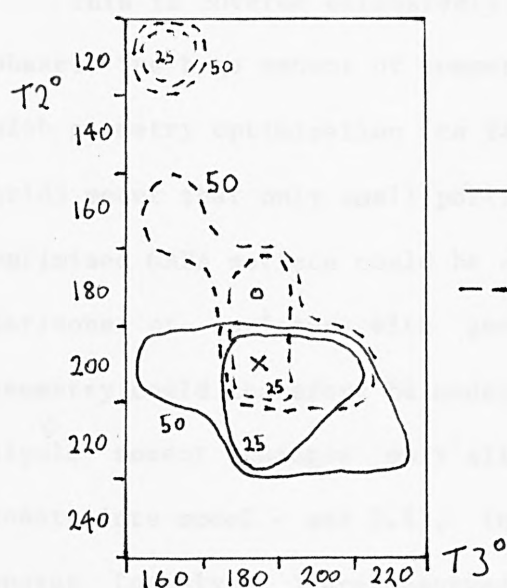
Figure 3.7. GABA  $x_T$  distributions by SOLVEFF and Hybrid methods.

a) Spheroidal cavity (MNDO), b) spherical cavity (MNDO),  
 c) Hybrid,  $T_4=60^\circ$  (with  $a_1$  relaxed for each water molecule),  
 d) Hybrid,  $T_4=40^\circ$  (fixed geometry), e) spheroid, CNDO (Clarke,  
 1975) - interpolated (see Appendix A1).

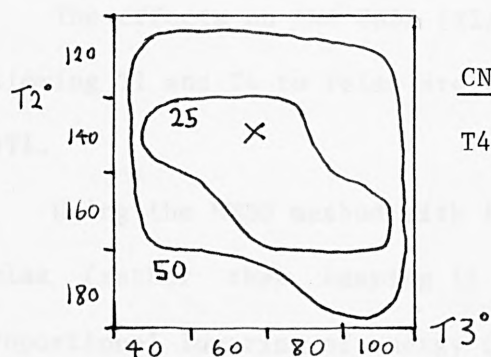
A logarithmic population scale is used to highlight small changes in  $x_T$  distribution too small to be seen with a linear scale.

MINDO

T1=180°

— T4=40° ( $E_{\min}(x) = -1690.8 \text{ kJ mol}^{-1}$ )--- T4=80° ( $E_{\min}(o) = -1754.3 \text{ kJ mol}^{-1}$ )

T1 relaxed

— T4=40° ( $E_{\min}(x) = -1691.6 \text{ kJ mol}^{-1}$ )--- T4=80° ( $E_{\min}(o) = -1756.0 \text{ kJ mol}^{-1}$ )CNDO/2

T4=60°, T1=180° (low-energy surface)

Figure 3.8. GABA SOLVEFF surfaces.

The energy for all other parts of the surfaces was greater than  $50 \text{ kJ mol}^{-1}$ . This is in contrast to the experimental NMR results, which show GABA to be flexible (3.2.1).

displaying one deep minimum but at different angles (Figure 3.8). In general MNDO tends to give slightly higher dipole moments which leads to an even stronger preference for extended conformations and less flexibility. The dominance of the  $E_{es}$  term (see (i) above) makes any difference between the use of MNDO or CNDO negligible when applied, with SOLVEFF, to GABA.

(iv) The effect of geometry optimisation on GABA conformations.

This is covered extensively in Chapter 2.4 for GABA in the gas-phase. The huge amount of computer time required for the calculations with geometry optimisation (ca 2400 seconds on the CDC7600 for an 8x8 grid) meant that only small portions around the minima of the geometry optimised GABA surface could be examined. Only semi-quantative comparisons of surfaces with geometry optimisation and with fixed geometry could therefore be made. We found that because the GABA dipole moment changes only slightly with geometry optimisation (two constraints model - see 2.4), insignificant differences in SOLVEFF energy (namely  $E_{es}$ ) are observed between surfaces with fixed geometry and with geometry optimisation.

The effects on the GABA (T2/T3) conformational energy-surface of allowing T1 and T4 to relax are considered separately:

a) T1.

Using the MNDO method with fixed geometry and allowing T1 to relax (rather than keeping it fixed at  $180^\circ$ ) gave only a very small proportional lowering of energy (with and without SOLVEFF), and had no significant effect on the positions of the conformational minima. Within  $50\text{kJ mol}^{-1}$  of the global minimum T1 optimisation makes  $< \text{ca } 2.5\text{kJ mol}^{-1}$  difference (Figure 3.8). Using different starting

values for T1 had no effect on this result. (In contrast, for the parts of the GABA surface examined with geometry optimisation (with 2 constraints - see 2.4), energy differences of up to  $27\text{kJ mol}^{-1}$  were observed. This is due to differences in  $E_{\text{int}}$ .) Only in high energy, highly folded conformations did T1 vary significantly from  $180^\circ$ . In comparison, with the Supermolecule method (Pullman and Berthod, 1975) a global minimum at  $195^\circ$  was found, with a lowering of overall energy by  $4\text{kJ mol}^{-1}$  on changing T1 from  $180^\circ$  to  $195^\circ$  and significant changes in the positions of conformational minima. This difference in results for the 2 methods is not surprising since with SOLVEFF the GABA dipole moment is independent of T1, and with Supermolecule T1 variation will effect the steric energy of the GABA supermolecule.

#### b)T4.

Variation of T4 had a pronounced effect on the GABA conformation energy surface, with a  $85\text{kJ mol}^{-1}$  difference between the  $T4=90^\circ$  low-energy surface and the  $T4=40^\circ$  high-energy surface - due largely to differences in  $E_{\text{es}}$  ( $78\text{kJ mol}^{-1}$ ). This result is unexpected since  $E_{\text{es}}$  should be independent of T4 (the T4 dependence is not an artifact of MNDO because Clarke (1981) found the same result using CNDO). A probable explanation is the effect of the internal field factor on  $E_{\text{es}}$  since the GABA dipole moment is virtually independent of T4. In contrast, with Supermolecule (Pullman and Berthod, 1975), T4 was found to have negligible effect on energies. In our limited Supermolecule calculations (using MNDO instead of PCILO) we found the  $T4=60^\circ$  energy-surface to be  $10.5\text{kJ mol}^{-1}$  lower than with  $T4=40^\circ$ . (Pullman and Berthod (1975) only examined the  $T4=0^\circ$  and  $T4=90^\circ$  surfaces.)



(v) Extensions to the SOLVEFF software.

The extra software and modifications to SOLVEFF which were required to enable SOLVEFF to be used in combination with MNDO (including geometry optimisation), and for improving the accuracy of SOLVEFF energies, are briefly described here.

a) MNDODP - a MNDO/SOLVEFF interface.

As the main input required for the SOLVEFF programme consists of just the molecular coordinates, the isolated molecule (or Super-molecule) energy and dipole moment for each conformation, SOLVEFF is independent of the programme which produces this data. We therefore wrote MNDODP, an interface between MNDO and SOLVEFF which enabled the calculation and comparison of MNDO (isolated molecule or Super-molecule) and SOLVEFF energy surfaces for two rotation angles in just one computer run (Appendix A6.5). The SOLVEFF programme had to be slightly modified for use with this interface, because SOLVEFF had been originally written for interface to a CNDO/2 programme (AP562STEW).

b) Modifications to the SOLVEFF programme.

An important modification to SOLVEFF was the lowering of the convergence criterion (DFX) from  $10^{-4}$  to  $10^{-5}$ , which improved the accuracy of resultant energies to better than  $0.2 \text{ kJ mol}^{-1}$ , with only a marginal increase in  $t_{\text{cpu}}$ . Clarke (1981) had found  $0.5 \text{ kJ mol}^{-1}$  energy differences between symmetric conformations. In addition, for a summation of  $x_{\text{T}}$ s over a  $18 \times 18$  point energy-grid the cumulative errors introduced by inadequate convergence can be quite large. As an additional measure, in ensuring that rounding-off errors are insignificant, an extra decimal place of accuracy was added to the input and output energies.

Other modifications were made to enable SOLVEFF to be used on small sections of an energy surface (eg for geometry optimisation), and with different grid increments. It is important to be able to use SOLVEFF with other than (the original)  $20^\circ$  grids, because anomalous gaps appear on mapping a  $20^\circ$  energy grid onto a  $0.1\text{\AA} \times_T$  distribution. (See Appendix A1 for why the gaps appeared in Clarke's (1976) results and how they can be avoided.)

A further modification, which greatly simplified the use of SOLVEFF, was an extensive conversion to free-format data input. In addition, allowance had to be made for any data points which were missing on the isolated surface by setting them to 999kcal/mol. This was necessary for the GABA Hybrid model where SCF convergence could often not be achieved for very high energy conformations.

(vi) The advantages of SOLVEFF over Supermolecule.

The main advantage of SOLVEFF is that it is computationally very rapid (a  $18 \times 18$   $T_2/T_3$  grid required on average  $7.8 \pm 0.3$  seconds for GABA). When applied to GABA SOLVEFF is several orders of magnitude faster than Supermolecule. A second advantage is that SOLVEFF is simple to use - there is no solvent molecule positioning/orientation to consider such as with Supermolecule. If the problems arising from the use of molecular dipoles (see (i) above) could therefore somehow be resolved, (eg by allowing for 2 dipole moments, or simply by using a lower dielectric constant) SOLVEFF would have very great potential.

### 3.3.3 A hybrid approach and tests on solvent effect algorithms.

Due to their complexity hybrid approaches are seldom used for practicable work on solution conformations (Burch et al, 1976). However, only a small extension to SOLVEFF was required to convert

MNDO Supermolecule energies into hybrid energies, since the dipole moments were available for each conformation.

We therefore examined a combined Supermolecule/SOLVEFF (Hybrid) approach by placing up to six water molecules around the GABA molecule to simulate close range solvent effects, and then adding long range effects with SOLVEFF. This gave results which were intermediate between the two methods ( $x_T$  - Figure 3.2, energy surface Figure 3.9), but closer to the extended conformation of SOLVEFF.

A problem with the Hybrid method is that it is very sensitive to the orientation of the water-molecule dipoles. This is clearly seen by comparing the  $T_4=60^\circ$  and  $T_4=40^\circ$  surfaces (Figure 3.9a), where a combination of water bond-dipoles at the  $T_4=40^\circ/T_2=220^\circ/T_3=220^\circ$  conformation leads to a considerable lowering of  $E_{tot}$  for this conformation - giving a lower minimum  $E_{tot}$  for the  $T_4=40^\circ$  fixed-geometry surface than for the  $60^\circ$  surface with OP optimisation. In addition, the change in  $E_{cav}$  is also much greater (isolated GABA ca 30 kJ, Hybrid ca 170 kJ). With the Hybrid method it is therefore even more important to consider low-energy permutations of the water-molecule orientations - which alas, is not only beyond the means of current computers, but also goes beyond the accuracy of the methods used (MNDO and SOLVEFF).

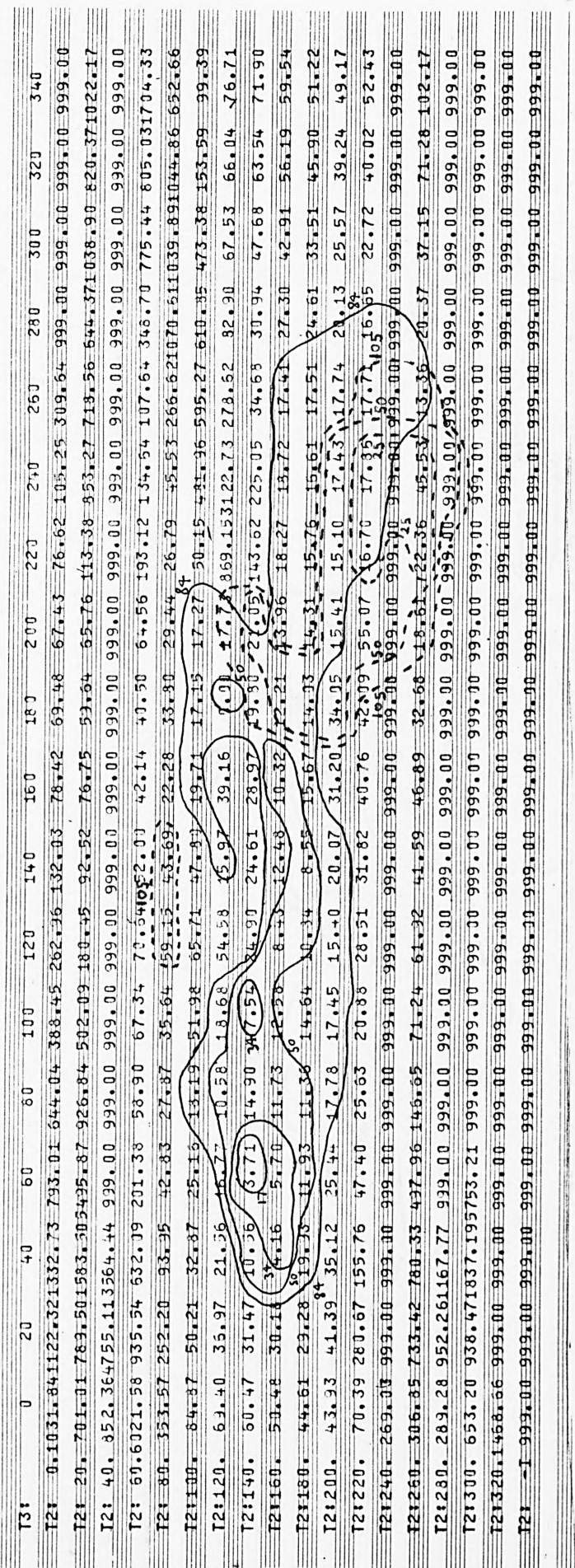
The Hybrid results appear to be more reasonable than those of SOLVEFF (due to a substantially reduced dipole moment of the GABA 'supermolecule'), but the addition of the O-H bond dipoles only adds to the existing errors in the SOLVEFF model (3.2.2 i), and the errors in the Supermolecule model are probably due more to neglect of optimisation of the orientation parameters with GABA conformation than the lack of long-range effects.

Figure 3.9a. Hybrid GABA energy-surface (MNDO, standard geometry).

Solid contours (and numbers): T4=60°, T1=180°.

Dashed contours: T4=40°, T1=180° (minimum energy surface, E<sub>min</sub> = -407.2 kcal mol<sup>-1</sup>).  
(Contours are in kJ mol<sup>-1</sup>.)

THIS ENERGY MAP IS ETOI KCAL/MOLE RELATIVE TO A MINIMUM VALUE OF -375.2









### 3.3.4 Simulation methods.

The main advantage of statistical simulation methods such as Monte Carlo (Finney, 1982) and Molecular Dynamics (Gunsteren and Berendsen, 1982) is that entropy can be accurately calculated. The methods, however, were designed to give **averaged** thermodynamic properties and are therefore not directly applicable to the present problem. It is conceivable that a full solvent effect simulation could be performed for each (solute) molecular conformation over an angular grid, but this would be extremely expensive in terms of computer time, especially for molecules such as GABA with several rotation angles.

Also, the methods require inter-molecular potential functions which are difficult to calculate accurately, especially for polar solvents and solutes. Small variations in potential functions have been shown to lead to large possible changes in predicted structure (Goodfellow, Finney and Barnes, 1982).

It is worth noting that with the ever increasing power of computers simulation methods will be used more and more, particularly for full simulations of drugs docked to receptors of known molecular structure (Chapter 7). Results reported so far (Van Gunsteren, 1986) for such simulations are apparently far superior to any results using just energy minimisation and solvent effect models (Bush and Halgren, 1986).

### 3.4 Conclusions.

We have established quite clearly from our variable-temperature NMR work that GABA is flexible in solution, with multiple minima and low energy-barriers between the minima. Other workers (Ham, 1974, Tanaka et al, 1978) had found by a single-temperature NMR method that

multiple GABA conformations are present in solution, but in attempting to find the populations of the minima they had to make assumptions about the angles of the minima and the associated coupling constants (3.2.1).

Of the theoretical models we examined, Supermolecule shows GABA to be flexible and SOLVEFF shows GABA to be essentially inflexible with only extended conformations present in solution. This deduction of only extended conformations with SOLVEFF is due to the dominance of the  $E_{es}$  term which, for polar molecules such as GABA, is roughly proportional to the square of the molecular dipole moment. For GABA in folded conformations the N-H and C-O bond dipoles are aligned roughly opposite one another giving a low net dipole moment, masking the fact that the molecule is still very polar. In extended conformations the dipoles combine. The effect of the  $E_{es}$  term overshadows the slight improvement obtained by using a spheroid rather than a sphere to represent the solute molecule.

With Supermolecule the problem is that the positions and energies of the conformational minima are dependent on the siting and orientation of the water molecules comprising the first hydration shell. Unless the optimisation of the orientation parameters (Figure 3.6) is fully examined the resultant energy-surface is at best only semi-quantitative.

Combining the two models does not solve the problem of the water molecule positioning in the first hydration shell, and although the Hybrid results appear to be more reasonable than those of SOLVEFF (due to a substantially reduced dipole moment of the GABA 'supermolecule'), the addition of the O-H bond dipoles only adds to the existing errors in the SOLVEFF model (3.2.2 i). The errors in the Supermolecule model

are due more to neglect of optimisation of the orientation parameters with GABA conformation than the lack of long-range effects.

Further support for GABA being flexible in the biophase comes from the ability of GABA to adopt the required conformation at the various types of GABA receptor (eg GABA<sub>A</sub> relatively extended or GABA<sub>B</sub> partially folded (Johnston, 1984 and see 5.3)).

## 4 The conformations of BIC, MeBIC and HBIC in solution.

### 4.1 Introduction

The main conformational features of BIC (Figure 4.1) and its salts - BIC methohalides (MeBIC) and protonated BIC (HBIC) - are the H-C1-C9-H torsion angle ( $\theta_1$ ), and the N-ring. For the N-ring the C9-C1-C1A-C4A dihedral angle ( $\theta_2$ ) is used to define the relative position of the phthalide group with respect to this ring (Figure 4.1). For a comparison of the structures of BIC (and salts) and GABA (Chapter 5), it is essential first to establish the values of the  $\theta_1$  and  $\theta_2$  conformational minima (angles and populations) present in solution, and the energy barriers between those minima.

The background and outline of our methods for finding this conformational information for BIC, HBIC and MeBIC are given in (i) and (ii) below. We found that BIC is fairly flexible in solution, with 3 low-energy conformations (4.2), whereas HBIC and MeBIC are more rigid with only 1 conformation (4.3 and 4.4). Details on the methods used to determine conformation for these two situations are given in sections 4.2 and 4.3.

#### (i) BIC.

The conformation of BIC has been established in the solid state by X-ray diffraction (Gilardi, 1973, Gorinsky and Moss, 1973), with  $\theta_1=172^\circ$ ,  $\theta_2=105^\circ$  (pseudoaxial) and the N-ring pseudochair. For BIC in solution, a value of  $\theta_1 \approx 50^\circ$  has previously been found using NMR, by applying the Karplus equation (Karplus, 1959) to the H1-H9 coupling constant (Andrews and Johnston, 1973), and by comparing the chemical shifts of key protons in various phthalide isoquinoline alkaloids (Elango et al, 1982). With both of these NMR procedures only a single minimum was searched for. However, three minima, all within

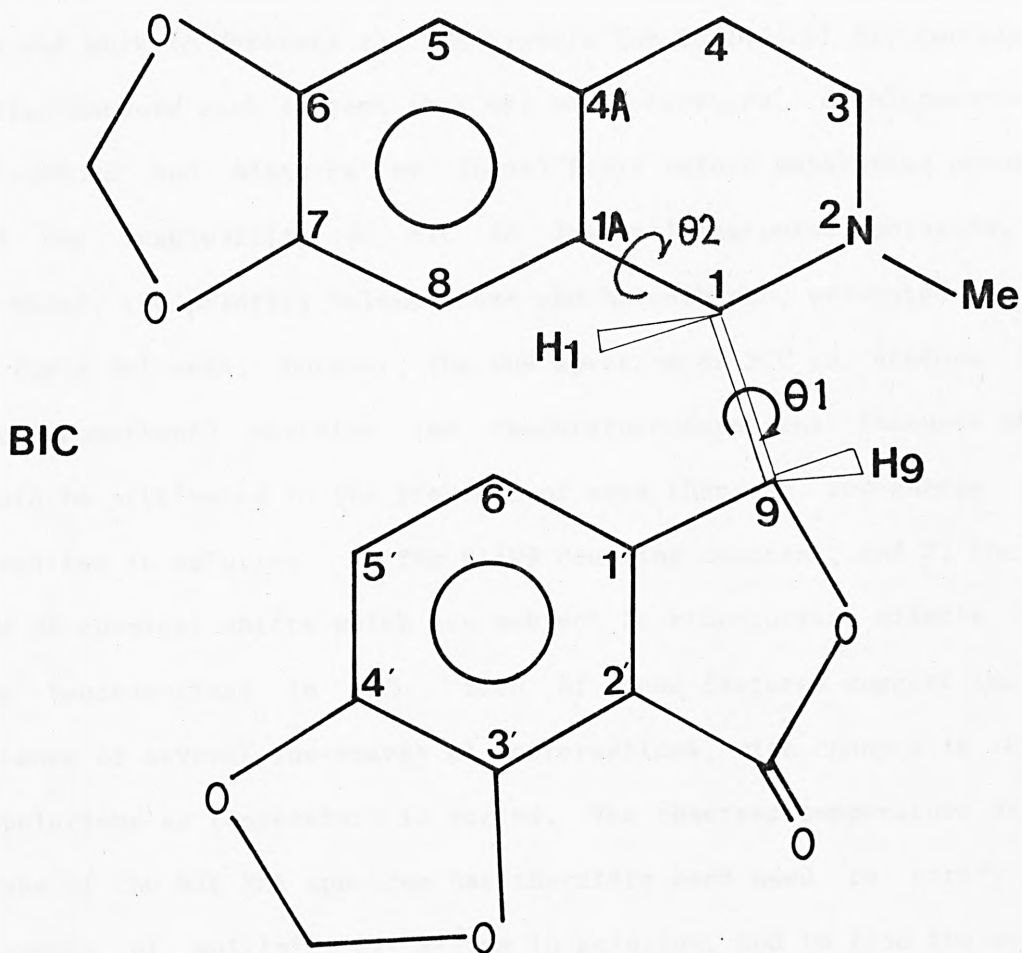


Figure 4.1. BIC molecular structure.

$10\text{kJ mol}^{-1}$  of each other, are found from gas-phase theoretical calculations (Figure 2.4), suggesting that more than one low-energy conformation may well exist in solution.

Variable temperature NMR spectroscopy is commonly used to identify and distinguish between different conformers which exist in solution. Normally this is only possible if the barriers between them are  $20\text{--}120\text{ kJmol}^{-1}$  (Lambert et al, 1981), depending on how soluble the substance is in solvents of low melting point. Attempts in the



present work to separate the NMR signals for individual BIC conformers failed because each solvent that was used (acetone, dichloromethane, chloroform and mixtures of these) froze before separation occurred. And the insolubility of BIC in low melting-point solvents, eg methanol, low polarity halomethanes and haloethanes, prevented the use of these solvents. However, the NMR spectrum of BIC in acetone (and dichloromethane) contains two temperature-dependent features which could be attributed to the presence of more than one low-energy conformation in solution: 1) The H1/H9 coupling constant, and 2) the H6' and H8 chemical shifts which are subject to ring-current effects from the benzene-rings in BIC. Both of these features suggest the existence of several low-energy  $\theta_1$  conformations, with changes in their populations as temperature is varied. The observed temperature dependence of the BIC NMR spectrum has therefore been used to verify the presence of multiple  $\theta_1$  minima in solution, and to find the energy differences between these minima. Nuclear Overhauser enhancement (NOE) difference spectroscopy gives additional support for the presence of at least two of these minima.

An upper limit for the barrier to interconversion between  $\theta_1$  conformational minima in BIC was estimated from the H6' and H8 peak broadening observed at low temperature.

The N-ring coupling constants combined with NOE difference spectroscopy have been used to verify that in solution the N-ring is pseudochair with the phthalide group pseudoaxial, as in the gas-phase (Andrews and Johnston, 1973) and the crystal (Gilardi, 1973, Gorinsky and Moss, 1973). Determination of the N-ring conformation was only possible for BIC in acetone, because in the other systems the nitrogen-ring proton coupling constants were too far from first-order to apply a similar analysis, even at 400MHz.

4.1

Another method for finding conformational information for molecules with low rotation barriers is the 'J-method' (Parr and Schaefer, 1980). The method makes use of long-range couplings for benzene derivatives containing side chains whose rotation is characterised by a twofold barrier (Parr and Schaefer, 1980), and is therefore not applicable to BIC.

(ii) MeBIC and HBIC.

For both MeBIC and HBIC in solution possible  $\theta_1$  ranges of  $70^\circ - 110^\circ$  and  $250^\circ - 290^\circ$  have been reported (Andrews and Johnston, 1973). These were derived, using the Karplus equation (Karplus, 1959), from the H1/H9 coupling constant of  $< \text{ca } 1\text{Hz}$  for  $\text{MeBIC}^+\text{Cl}^-$  in deuterium oxide. Subsequently, on the basis of their theoretical gas-phase PCILO calculations, Andrews and Johnston (1973) chose  $250^\circ - 290^\circ$  as the low-energy conformation present in solution. The global minimum in the calculations did not, however, fall within the above range and too coarse a grid was used to yield quantitative rotation barriers. (In addition their assignment of 5.15ppm to H9 for MeBIC in deuterium oxide is not in agreement with our assignments based on spin decoupling and NOE effects - see 4.3.)

For an experimental method of narrowing down the above  $\theta_1$  ranges to a single value (only a single conformation is present in solution, though the angle is slightly different for different solvents - see below), we therefore turned to chemical shifts with the hope that conformational information could be derived from the benzene-ring shielding effects (as described earlier for BIC). However, for MeBIC in non-dissociating solvents (eg acetone) certain key chemical shifts (H1, H6', Me(ax) and Me(eq)) were found to be strongly affected by ion-pairing of MeBIC with the halide counterion. We therefore used

two different halide salts of MeBIC (I and Cl) and different types of solvent (acetone and deuterium oxide), to determine the effect of the counterion on chemical shift and conformation. The position of the H8 chemical shift (which was shown to be unaffected by the counterion), NOE difference spectroscopy, and the counterion effects themselves could then be used to narrow  $\theta_1$  down to  $255^\circ$ - $290^\circ$ . An accurate value for  $\theta_1$  was then determined by applying the Karplus parameters derived for BIC to the H1/H9 coupling constant. The HBIC conformation was determined from the shielding of the H8 and H6' chemical shifts, and from the H1/H9 coupling constant.

As calculations in the gas-phase indicate the presence of more than one low-energy conformation for MeBIC and HBIC (see 2.5), we used the invariance of the H1/H9 coupling constant to show that just one conformation exists for both MeBIC and HBIC in solution. (Though for MeBIC the exact  $\theta_1$  value varies slightly with solvent.)

A lower bound for the energy of the next conformation above the global minimum was then estimated from a knowledge of the accuracy which the H1/H9 coupling constant was measured (4.3.2).

A comparison of the conformational behaviour of BIC, HBIC and MeBIC was then made to determine whether a positive nitrogen-region is essential for GABA<sub>A</sub> antagonist activity (4.6 iii). This is because pharmacological evidence, based on binding study data, is apparently inconclusive (Kardos et al, 1984).

#### 4.2 The energy profile of BIC in solution.

The observed temperature-dependence of the H6' and H8 chemical shifts (Figure 4.2) - which are subject to ring-current effects and

are therefore conformation dependent, and the H1/H9 coupling constant have been used to determine  $\theta_1$  conformer populations (4.2.1). An upper limit to the barrier between these minima was then estimated from the H6' and H8 peak broadening at low temperature (4.2.2). For  $\theta_2$  the N-ring proton coupling constants and NOE effects were used to determine the N-ring conformation (4.2.3).

#### 4.2.1 Evidence for multiple BIC solution minima ( $\theta_1$ ).

Gas-phase geometry-optimised MM2 calculations (2.5) gave three minima at  $\theta_1 \approx 45^\circ$ ,  $170^\circ$  and  $270^\circ$  (Figure 2.4), and MNDO calculations on a truncated BIC molecule gave similar minima (see 2.5). The MM2  $\theta_1$  angles have been used to calculate (i) the theoretical H1/H9 coupling constant, and (ii) the H6' and H8 chemical shifts, for comparison with experimental data at each temperature. The good fit found using **both** (i) and (ii) implies that the same  $\theta_1$  minima (angles) are present in solution as in the gas-phase. (Though the populations are not necessarily the same.)

##### 4.2.1.1 Theoretical calculation of the averaged H1/H9 coupling constant in BIC.

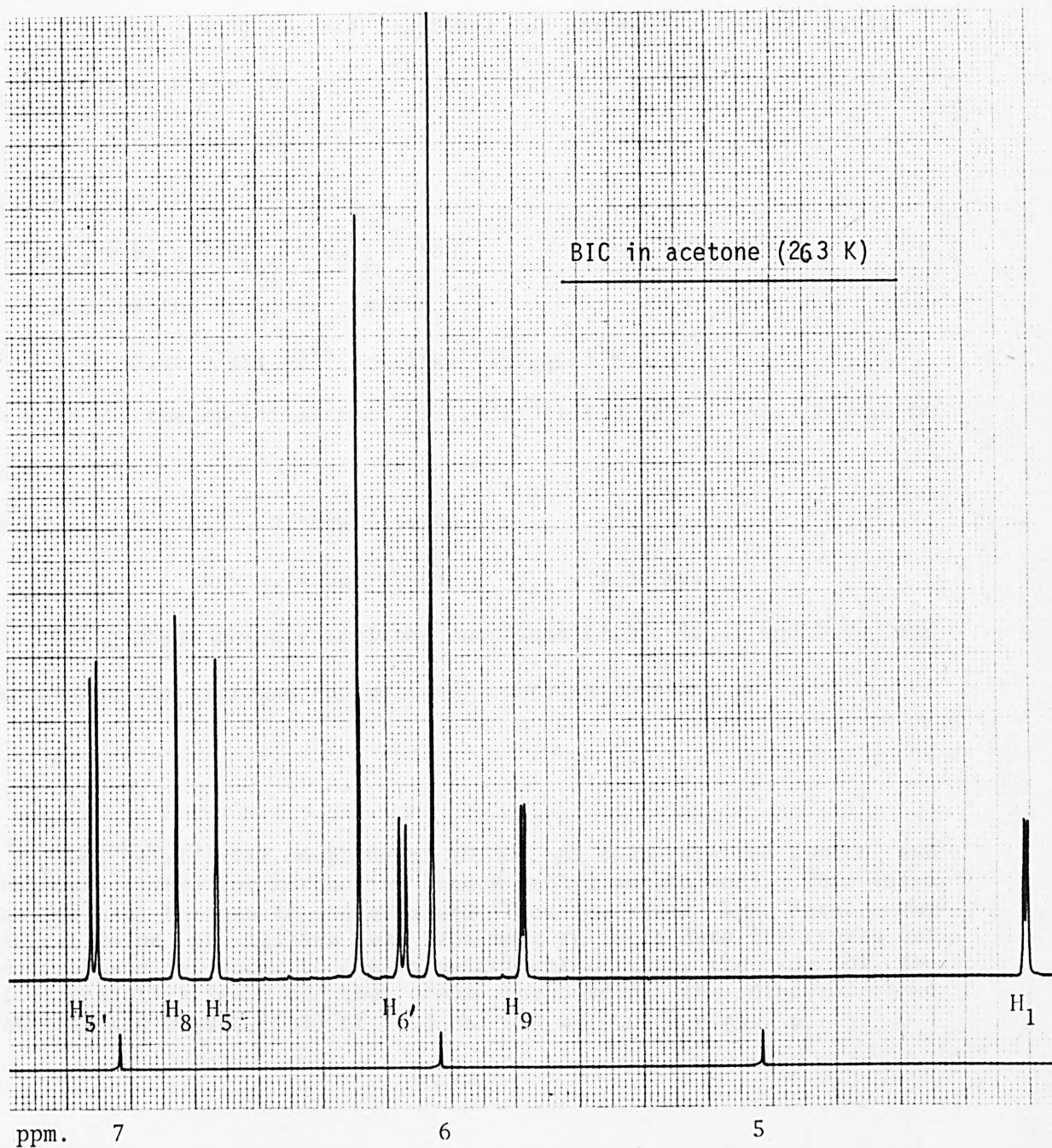
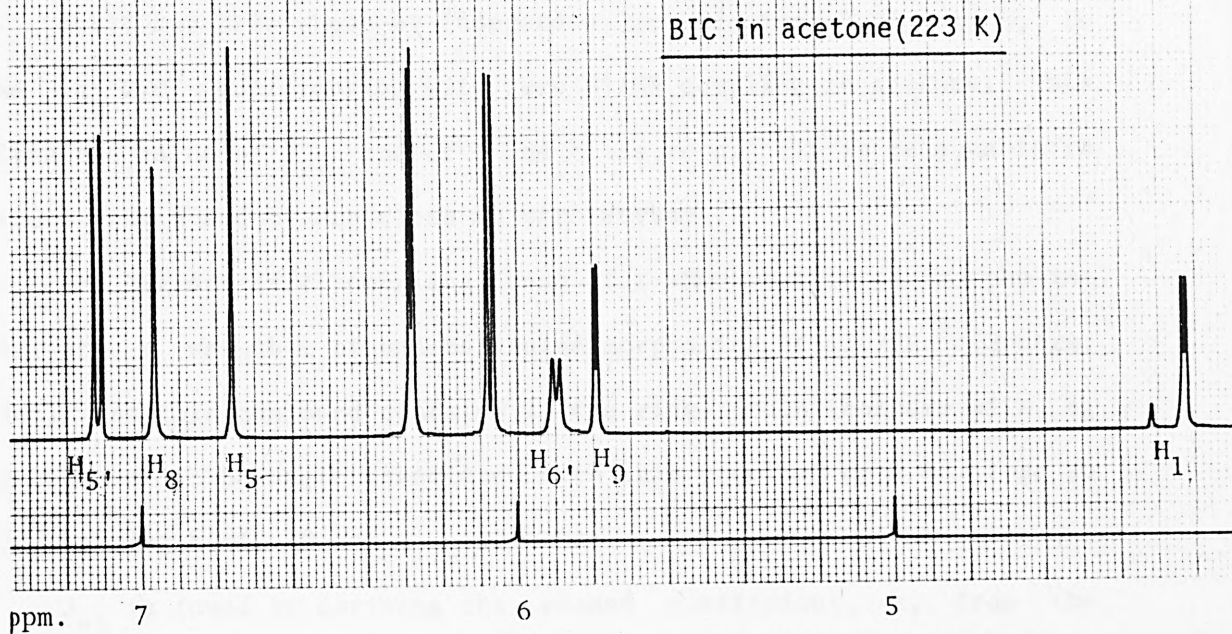
As the effect on vicinal couplings of substituents of similar electronegativity is negligible (Abraham and Gatti, 1969), Me-substituted morpholine derivatives can be used to obtain values for the coupling associated with each  $\theta_1$  angle.

Firstly, the required coefficients for the Karplus equation must be derived:

$$\begin{aligned} J_{\theta_1} &= k' \cos^2 \theta_1 + b' & 90^\circ < \theta_1 < 270^\circ \\ J_{\theta_1} &= k \cos^2 \theta_1 + b & -90^\circ < \theta_1 < 90^\circ \end{aligned}$$



Figure 4.2 Temperature-dependence of BIC chemical shifts.





MeBIC has been shown by independent methods (see 4.3) to be in the ca  $270^\circ$  conformation with zero H1/H9 coupling in acetone. This gives the value of  $J_{270}$  and of  $b$  and  $b'$  as zero. (It is reasonable to assume that  $b$  and  $b'$  are close to zero anyway.)

The highest axial coupling found for Me-substituted morpholine derivatives was that of unsubstituted morpholine at  $12.79\text{Hz}^*$  (Spragg, 1968). As this compound contains little ring-strain (compare with eg cyclohexane), the angle associated with this coupling must be close to  $180^\circ$ , giving  $k'=12.79$  and  $J_{170}=12.41\text{Hz}$ .

$J_{45}$  is found by deriving the second coefficient,  $k$ , from the other (Spragg, 1968) vicinal couplings for morpholine (Figure 4.2):

$$J_{1,3} = 1.1 = k \cos^2 \theta_{1,3}$$

$$J_{2,3} = 1.3 = k \cos^2 \theta_{2,3}$$

$$J_{1,4} = 2.8 = k \cos^2 \theta_{1,4}$$

$$\text{and } \theta_{1,3} + \theta_{2,3} + \theta_{1,4} = \theta_{2,4} \approx 180^\circ.$$

Eliminating  $\theta_{1,3}$ ,  $\theta_{2,3}$  and  $\theta_{1,4}$ :

$$k^2 - 5.2k + 2.46 - 2k(1.43)^{1/2}(1-1.1/k)^{1/2}(1-1.3/k)^{1/2} = 0$$

As  $k \gg 1.3$  the approximation:  $(1-1.1/k)^{1/2}(1-1.3/k)^{1/2} \approx 1 - 1.2/k$  can be used,

$$\text{giving: } k^2 - 7.592k + 5.730 = 0$$

$$k = 6.74 \text{ Hz} \quad \text{and} \quad J_{45} = 3.37 \text{ Hz}.$$

Using these three calculated couplings (3.37, 12.41 and 0.0), and expressing the averaged (normalised) coupling constant as:

\* A lower axial/axial coupling of 10.3Hz has been reported (Smith and Shoulders, 1969) for morpholine (and N-methyl morpholine). This value was not used because it was obtained from averaged NMR data. The higher value of 12.79Hz was from more reliable data for frozen out spectra, and is closer to the value determined using the sum of the electronegativities of the atoms surrounding the dihedral bond (Abraham and Gatti, 1969) of 12.3 Hz. ( $J_t = 18.07 - 0.88 e_i$  .)

$$J_{\text{calc}} = (J_{45} + J_{170} \exp(-G_{170}/RT) + J_{270} \exp(-G_{270}/RT)) / n$$

a computer program (JVIC - see A6.3) was written to produce values of  $J_{\text{calc}}$  which were then matched with  $J_{\text{obs}}$  at each temperature (Table 4.1). The conformational energy differences (from the  $45^\circ$  conformation) are those which produce the best match between  $J_{\text{calc}}$  and  $J_{\text{obs}}$ . (A second solution, with higher energies, was found but was rejected since it disagrees with the results derived from chemical shifts - see 4.2.1.2.) Simultaneous equations were not used to determine these energies because of the non-linear nature of several of the terms involved. (See Appendix A3 for a more detailed explanation.)

Table 4.1. Comparison of observed and calculated coupling constants for BIC.

| Solvent                  | $G_{170}$                  | $G_{270}$                  | $J_{\text{calc}}$ (Hz) |      | $J_{\text{obs}}$ (Hz) |      |
|--------------------------|----------------------------|----------------------------|------------------------|------|-----------------------|------|
|                          | $\pm 1 \text{ kJmol}^{-1}$ | $\pm 1 \text{ kJmol}^{-1}$ | 248K                   | 317K | 248K                  | 317K |
| $\text{CD}_2\text{Cl}_2$ | 2.8                        | 0.8                        | 4.00                   | 4.30 | 3.9                   | 4.3  |
| acetone                  | 4.1                        | 1.4                        | 3.86                   | 4.06 | 3.85                  | 4.1  |
| $\text{CD}_2\text{Cl}_2$ | 1.5                        | 0.7                        | 3.94                   | 4.20 | 3.9                   | 4.3  |
| acetone                  | 2.1                        | 0.8                        | 3.86                   | 4.10 | 3.85                  | 4.1  |

The upper results were obtained with  $k=8.8$  and  $k'=12.4$ , and the lower results with  $k=12.79$  and  $k'=6.74$  (see text).

The value of  $k=6.74$  seemed rather low compared with  $k'=12.79$ . We therefore examined the above approximation,  $\theta_{2,4}=180^\circ$ , by using the expression  $\theta_{2,4}=180^\circ + e$ , where  $e$  is a perturbation in the diaxial angle of up to  $10^\circ$ . (Higher angles could be considered, but the mathematics becomes much more complicated and the same end result will be achieved). Repeating the above calculations with  $e=+10^\circ$  ( $\theta_{2,4}=190^\circ$ )

gave an increased  $k$  value of 8Hz and  $k'=13\text{Hz}$ . To check if this change was justified we performed gas-phase calculations (MNDO and MM2) on N-methyl morpholine and morpholine (with full geometry optimisation and with 'm' symmetry), and obtained  $\theta_{2,4} \approx 160^\circ$  (ie  $e=-20^\circ$ ). The difference in angle could be explained by differences in solution and gas-phase conformations (and morpholine is slightly more polar than BIC), but was not very helpful! We therefore also calculated values for N-methyl morpholine (Spragg, 1968), finding  $k=8.8$  and  $k'=12.4$  (with  $\theta_{2,4}=180^\circ$ ), and applied these to the H1/H9 coupling constant calculations. The results ( $\theta$ 1 energies) are given in Table 2.1 and are not significantly different from those with  $k=6.74$ .

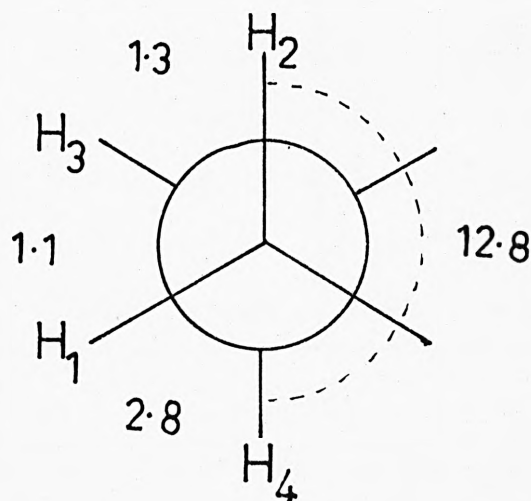


Figure 4.3. Vicinal proton coupling constants for morpholine (Hz)  
(Spragg, 1968).

Table 4.2. Bicuculline proton chemical shifts.

| Proton             | acetone solvent |      |                  | dichloromethane solvent |      |      |                             |
|--------------------|-----------------|------|------------------|-------------------------|------|------|-----------------------------|
|                    | 213K            | 296K | $\Delta\delta_T$ | 193K                    | 213K | 296K | $\Delta\delta_{T(213-296)}$ |
| H4e                | 1.93            | 2.20 | 0.27             | 2.00                    | 2.04 | 2.22 | 0.22                        |
| H3e                | 2.35            | 2.51 | 0.16             | -                       | 2.42 | 2.52 | 0.10                        |
| Me                 | 2.47            | 2.54 | 0.07             | 2.45                    | 2.47 | 2.54 | 0.07                        |
| H4a                | 2.48            | 2.61 | 0.13             | -                       | 2.49 | 2.60 | 0.11                        |
| H3a                | 2.62            | 2.80 | 0.18             | -                       | 2.67 | 2.78 | 0.11                        |
| H1                 | 4.23            | 4.14 | -0.09            | 4.13                    | 4.12 | 4.04 | -0.08                       |
| H9                 | 5.79            | 5.71 | -0.08            | 5.56                    | 5.55 | 5.55 | 0.00                        |
| H6'                | 5.82            | 6.22 | 0.40             | -                       | 5.89 | 6.19 | 0.30                        |
| OCH <sub>2</sub> O | 6.08            | 6.00 | -0.08            | 5.96                    | 5.95 | 5.93 | -0.02                       |
| OCH <sub>2</sub> O | 6.27            | 6.22 | -0.05            | 6.18                    | 6.17 | 6.16 | -0.0                        |
| H5                 | 6.78            | 6.67 | -0.11            | 6.58                    | 6.58 | 6.60 | 0.02                        |
| H8                 | 7.01            | 6.75 | -0.26            | -                       | 6.49 | 6.51 | 0.02                        |
| H5'                | 7.12            | 7.05 | -0.07            | 6.88                    | 6.89 | 6.93 | 0.04                        |

Proton assignments are in agreement with those of Elango et al (1982).

The observed 1Hz coupling between H6' and H9 is common in phthalide derivatives (Safe and Moir, 1964).

#### 4.2.1.2 Theoretical calculation of averaged chemical shifts in BIC.

The temperature-dependence of the H6' and H8 chemical shifts in acetone (Table 4.2) was found to be mainly due to the proximity of the benzene rings in BIC and thence primarily dependent on  $\theta_1$ . This was established by NMR examination of the separate phthalide and isoquinoline halves of BIC, using 6,7-dimethoxy phthalide (MEC) for H6' and 6,7-dimethoxy 1,2,3,4-tetrahydro isoquinoline (ISO) for H8 (Figure 4.2). All the aromatic protons in these were found to have negligible temperature-dependence (Table 4.3). Furthermore, as the H6' and H8 chemical shifts were also found to be independent of concentration<sup>\*</sup>, the temperature-dependence of the BIC chemical shifts would seem to be due mainly to changes in the populations of the  $\theta_1$  conformational modes with temperature. The benzene ring shielding contribution was calculated for each of the three  $\theta_1$  angles used in the H1/H9 coupling constant calculations, using MM2 optimised geometry<sup>\*\*</sup> and published shielding contribution tables (Emsley et al, 1965) based on the Johnson-Bovey (1958) equation. This equation has been well tested (eg Perkins and Wuthrich, 1979) and has not been improved upon by later quantum mechanical models (Haigh and Mallion, 1972 and 1980). Computer programs were written to convert the lengthy MM2 output geometry into the format required for use with the tables (see Appendix A6).

The energy differences between minima were derived in a similar way to those using the H1/H9 coupling constant data (4.2.1.1), with

<sup>\*</sup> Spectra recorded in saturated acetone are little different to those in dilute solution and spectra recorded in dichloromethane at different concentrations are virtually identical.

<sup>\*\*</sup> Initially, coordinates based on crystal bond lengths and angles were used, which gave qualitatively the same results (Figure 4.5).



ISO

MEC

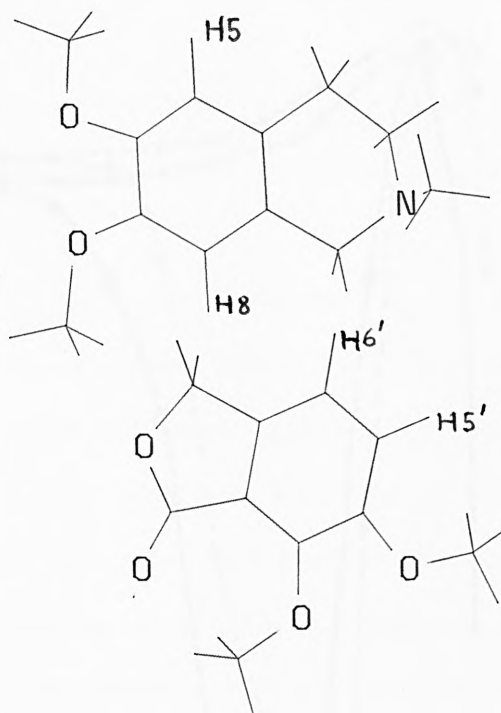


Figure 4.4. ISO and MEC molecular structures.

These molecules were used to determine values for the H6' and H8 chemical shifts without the influence of the second benzene ring. A comparison of several dioxolo and dimethoxy benzene derivatives showed that the use of these (more readily available) dimethoxy compounds makes less than 0.1ppm difference to the benzene-ring proton chemical shifts.

Table 4.3. ISO and MEC chemical shifts (ppm).

|     | BIC    | acetone |       | dichloromethane |       |
|-----|--------|---------|-------|-----------------|-------|
|     |        | 254K    | 296K  | 213K            | 296K  |
| ISO | proton |         |       |                 |       |
|     | H8     | 6.616   | 6.655 | 6.616           | 6.588 |
|     | H5     | 6.578   | 6.615 | 6.551           | 6.528 |
| MEC | H6'    | 7.245   | 7.192 | 7.241           | 7.124 |
|     | H5'    | 7.410   | 7.371 | 7.314           | 7.249 |

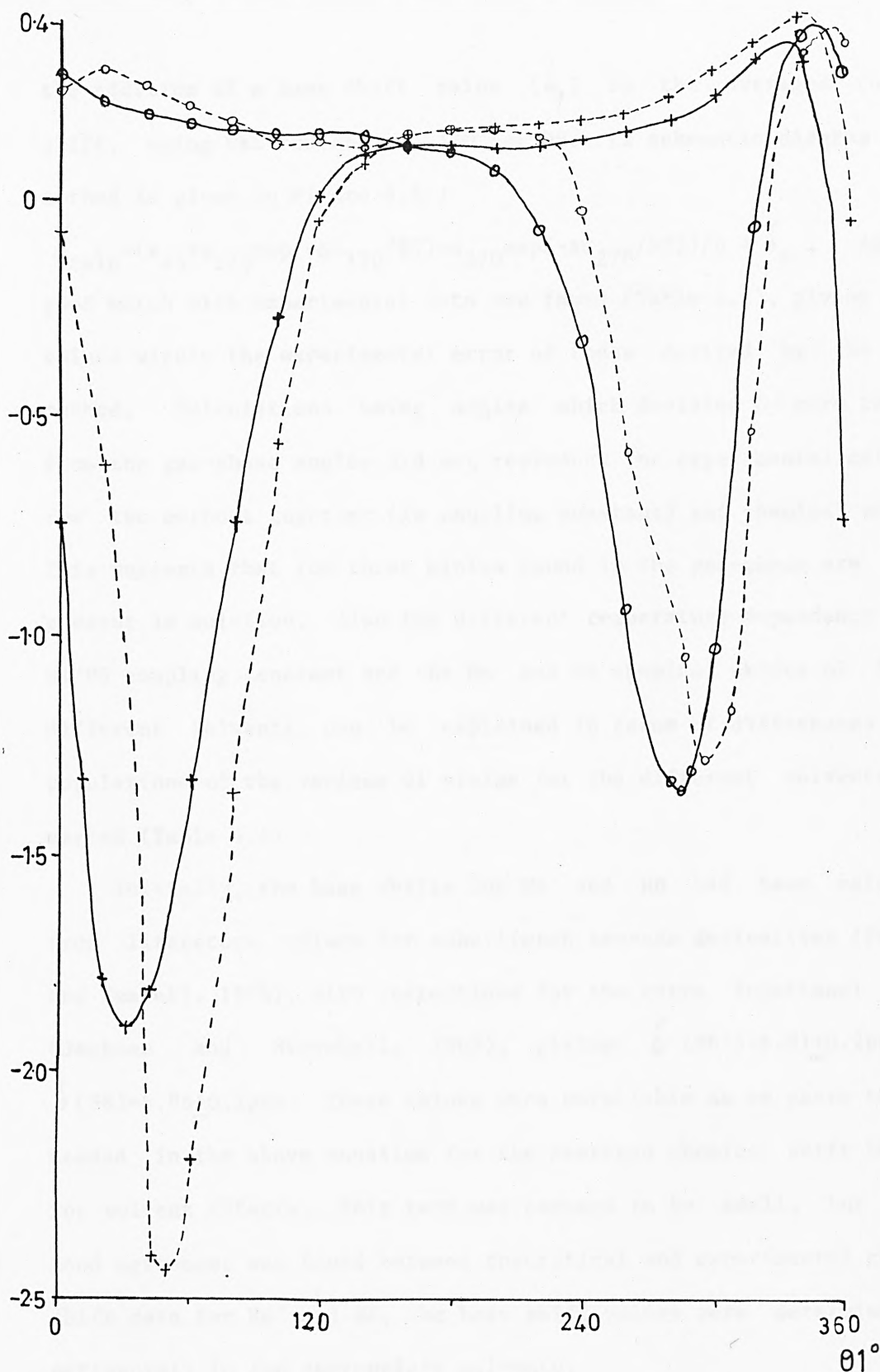


Figure 4.5. Theoretical shielding contribution for H6' and H8 in BIC. Two of the proposed solution angles are in regions of extreme shielding, which would account for the large temperature-dependence of the chemical shifts of these protons. (The third angle ( $\theta_1 \approx 170^\circ$ ) is important in the coupling constant calculations, because of the large coupling for this angle.) (+)H6'; (o)H8; solid lines MM2 geometry; dashed lines crystal geometry.

the addition of a base shift value ( $\delta_0$ ) to the averaged chemical shift, using MEC for H6' and ISO for H8. (A schematic diagram of the method is given in Figure 4.6.)

$\delta_{\text{calc}} = (s_{45} + s_{170} \exp(-\Delta G_{170}/RT) + s_{270} \exp(-\Delta G_{270}/RT)) / n + \delta_0$ . Again a good match with experimental data was found (Table 4.4), giving energy values within the experimental error of those derived by the first method. Calculations using angles which deviated by more than  $20^\circ$  from the gas-phase angles did not reproduce the experimental data for the two methods together (ie coupling constants and chemical shifts). This suggests that the three minima found in the gas-phase are indeed present in solution. Also the different temperature-dependence of the H1/H9 coupling constant and the H6' and H8 chemical shifts of BIC in different solvents can be explained in terms of differences in the populations of the various  $\theta_1$  minima for the different solvents concerned (Table 4.4).

Initially, the base shifts for H6' and H8 had been calculated from literature values for substituted benzene derivatives (Pouchert and Cambell, 1974), with corrections for the extra functional groups (Jackman and Sternhell, 1969), giving:  $\delta(\text{H6}') = 6.81 \pm 0.2 \text{ ppm}$  and  $\delta(\text{H8}) = 6.86 \pm 0.2 \text{ ppm}$ . These values were unreliable as an extra term was needed in the above equation for the averaged chemical shift to allow for solvent effects. This term was assumed to be small, but as no good agreement was found between theoretical and experimental chemical shift data for H6' and H8, the base shift values were determined experimentally in the appropriate solvents.

Further support for the chemical shifts of these protons being conformation-dependent comes from work (Elango et al, 1982) on a series of related phthalide isoquinolines. Here, the compounds, which

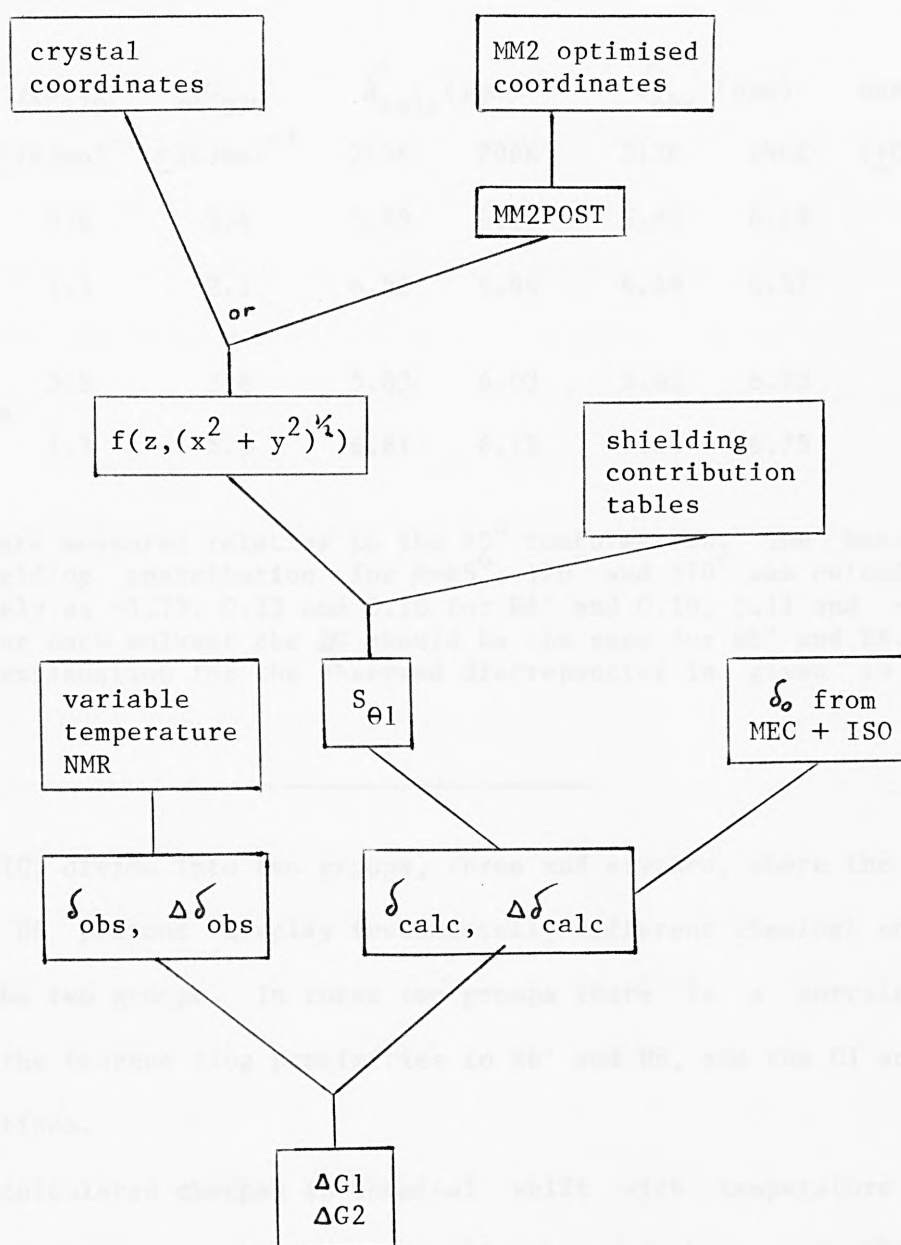


Figure 4.6. Schematic representation of the derivation of energy differences for  $\theta_1$  conformers.

Table 4.4. Comparison of observed and calculated chemical shifts for BIC.

|                          | $\Delta G_{170}$      | $\Delta G_{270}$      | $\delta_{\text{calc}}$ (ppm) |      | $\delta_{\text{obs}}$ (ppm) |      | base shift<br>(+0.03ppm) |
|--------------------------|-----------------------|-----------------------|------------------------------|------|-----------------------------|------|--------------------------|
|                          | $\text{+2kJmol}^{-1}$ | $\text{+2kJmol}^{-1}$ | 213K                         | 296K | 213K                        | 296K |                          |
| H6'                      | 3.6                   | 3.4                   | 5.89                         | 6.10 | 5.89                        | 6.19 | 7.18                     |
| $\text{CD}_2\text{Cl}_2$ |                       |                       |                              |      |                             |      |                          |
| H8                       | 1.5                   | 2.1                   | 6.52                         | 6.46 | 6.49                        | 6.51 | 6.60                     |
| H6'                      | 3.8                   | 3.8                   | 5.83                         | 6.03 | 5.82                        | 6.22 | 7.23                     |
| acetone                  |                       |                       |                              |      |                             |      |                          |
| H8                       | 1.5                   | 5.9                   | 6.81                         | 6.75 | 7.01                        | 6.75 | 6.63                     |

Energies are measured relative to the  $45^\circ$  conformation. The benzene-ring shielding contribution for  $\theta=45^\circ$ ,  $170^\circ$  and  $270^\circ$  was calculated respectively as -1.73, 0.12 and 0.16 for H6' and 0.19, 0.11 and -1.20 for H8. For each solvent the  $\Delta G$  should be the same for H6' and H8. A possible explanation for the observed discrepancies is given in the text.

---

include BIC, divide into two groups, threo and erythro, where the same H6' and H8 protons display fundamentally different chemical shifts between the two groups. In these two groups there is a correlation between the benzene ring proximities to H6' and H8, and the C1 and C9 configurations.

Our calculated **changes** in chemical shift with temperature are slightly smaller than the experimentally observed changes. Small perturbations to the calculated shielding contributions in accord with observed small discrepancies in the Johnson-Bovey equation (Mallion 1971, Haigh and Mallion, 1980), and allowing for the effect of the carbonyl group only slightly increased the magnitude of the calculated chemical shift changes. A more likely cause of the discrepancies is the overall effect of thermal vibrations on chemical shift. A similar explanation has been given (Frigerio et al, 1982) for similarly low



theoretical chemical-shift changes calculated for benzyl-piperazine-2,5-diones.

The benzene-ring shielding contribution for the H5, H5', H1 and H9 chemical shifts has also been calculated (Table 4.5), but the experimental chemical shift changes with temperature for these protons are too small for a meaningful comparison with theoretical values. The H4e temperature coefficient is relatively large (Table 4.2), but the theoretical value of the shielding contribution for this proton is susceptible to small changes in the molecular geometry, and is therefore difficult to calculate accurately. (This also applies to the other N-ring protons.)

#### 4.2.1.3 NOE difference spectroscopy on BIC.

Additional evidence for multiple BIC solution minima is found from NOE difference spectroscopy. For an NOE effect to be observed between two protons, a low-energy conformation must exist such that the two protons are in close proximity to one another. NOE effects observed for H8/H9 and H6'/H3a (Table 4.6) require ca 1% or more (Prazeres, 1982) of BIC to be in the conformation range  $\theta_1=40^\circ - 140^\circ$ , and the effect for H9/H3a likewise indicates another  $\theta_1$  range of  $160^\circ - 240^\circ$ . Two of the gas-phase minima fall within these ranges. The third minimum ( $\theta_1=270^\circ$ ) cannot be verified by NOE because of the lack of conveniently located protons.

#### 4.2.2 Barrier to internal conversion between $\theta_1$ minima.

Information on energy minima is incomplete without knowledge of the barrier separating them. This barrier is commonly calculated from the temperature ( $T_c$ ) at which coalesced NMR peaks separate and the chemical shift separation ( $\Delta\nu$ ) of the individual conformers involved.

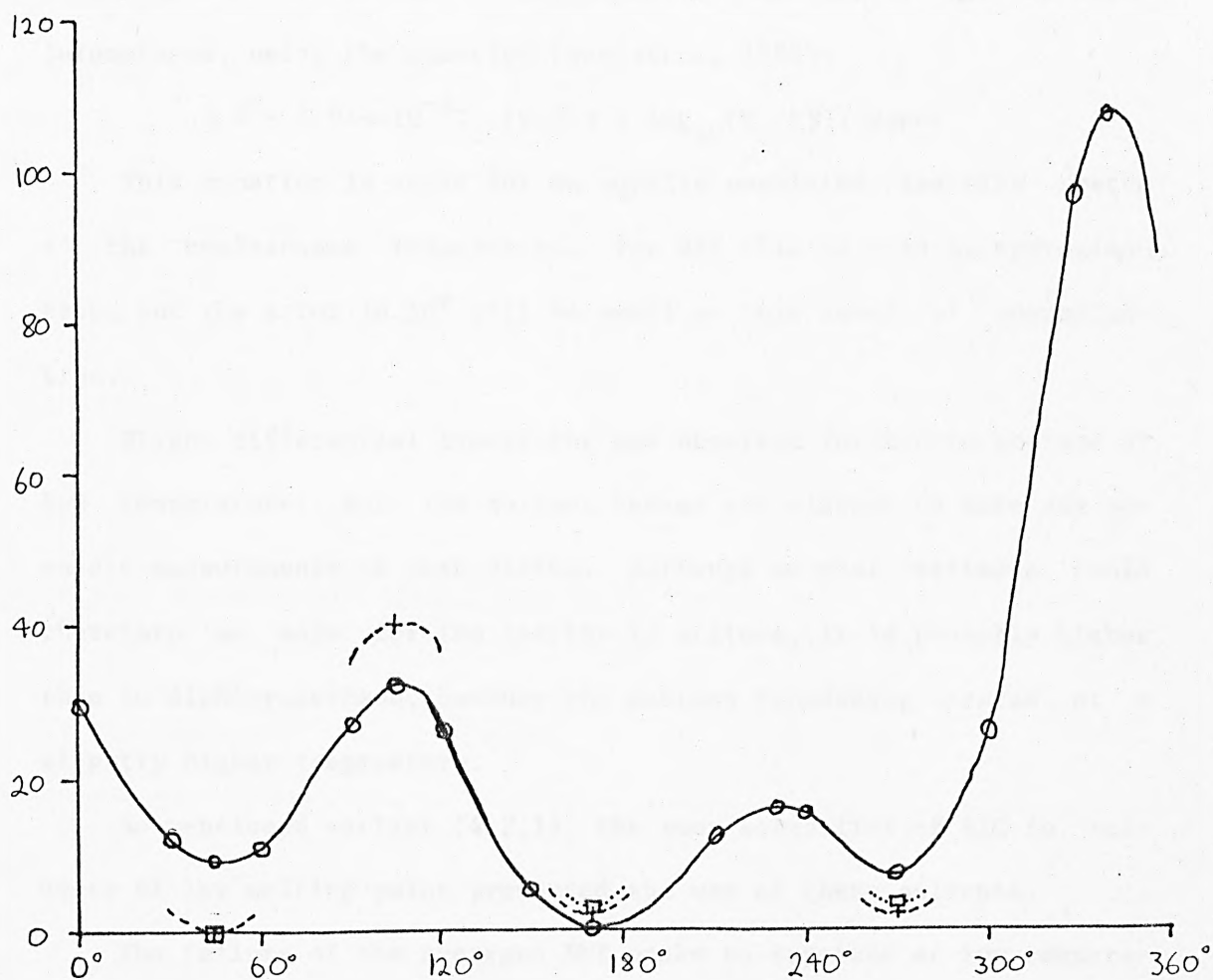
Energy/kJmol<sup>-1</sup>

Figure 4.7. BIC potential energy surface, —○— MM2 gas-phase calculations; --+-- dichloromethane solvent; ...□... acetone solvent. The latter two are average values.

Below 240K the H6' and H8 peaks of BIC began to show significant differential broadening. At 193K the solvent (dichloromethane) froze and the H6' peak was so broad as to have virtually disappeared. Assuming that 193K is approximately the coalescence temperature, and estimating  $\Delta\nu$  as 400Hz, from the theoretical chemical shift difference, an upper limit of ca 40 kJ mol<sup>-1</sup> is obtained for the barrier ( $\Delta G^\ddagger$ ) in dichloromethane, using the equation (Sandstrom, 1982):

$$\Delta G^\ddagger = 1.914 \times 10^{-2} T_c (9.972 + \log_{10}(T_c/\Delta\nu)) \text{ kJmol}^{-1}$$

This equation is valid for an equally populated two-site system at the coalescence temperature. For BIC this is only an approximation, but the error in  $\Delta G^\ddagger$  will be small at this level of approximation.

Slight differential broadening was observed for BIC in acetone at low temperature, but the solvent became too viscous to make any accurate measurements of peak widths. Although no real estimate could therefore be made for the barrier in acetone, it is probably higher than in dichloromethane, because the acetone broadening occurred at a slightly higher temperature.

As mentioned earlier (4.2.1), the poor solubility of BIC in solvents of low melting-point prevented the use of these solvents.

The failure of the averaged NMR peaks to separate at low temperature leaves the  $\theta_1$  value for the estimated 40 kJ mol<sup>-1</sup> barrier undetermined by this method. Molecular mechanics (MM2) calculations (Figure 2.4) give barriers for  $\theta_1$  at ca 110° and 230° of ca 30 kJ mol<sup>-1</sup>, corresponding roughly to the 40 kJ mol<sup>-1</sup> barrier found for solution (Figure 4.7).

Table 4.5

Benzene-ring shielding contribution for H5, H1, H5' and H9.

|     | 01   | 20   | 25   | 30   | 35    | 40   | 45    | 50   | 55   | 60   | 65   | 70   | 75   |
|-----|------|------|------|------|-------|------|-------|------|------|------|------|------|------|
| H5  | -.09 | -.08 | -.05 | -.03 | -.02  | -.00 | +.01  | +.01 | +.03 | +.04 | +.04 | +.04 | +.04 |
| H1  | +.22 | +.22 | +.22 | +.22 | +.22  | +.22 | +.215 | +.23 | +.21 | +.21 | +.21 | +.21 | +.21 |
| H5' | -.46 | -.44 | -.42 | -.40 | -.37  | -.35 | -.31  | -.30 | -.25 | -.23 | -.20 | -.17 | -.17 |
| H9  | +.26 | +.27 | +.27 | +.28 | +.28  | +.29 | +.28  | +.26 | +.28 | +.27 | +.27 | +.27 | +.26 |
|     | 155  | 160  | 165  | 170  | 175   | 180  | 185   | 190  | 195  | 200  | 205  |      |      |
| H5  | 0.   | 0.   | 0.   | 0.   | 0.    | +.01 | 0.    | 0    | 0    | 0    | 0    |      |      |
| H1  | +.07 | +.07 | +.05 | +.03 | +.03  | +.04 | -.01  | -.02 | -.03 | -.03 | -.04 |      |      |
| H5' | 0.   | 0.   | 0.   | 0.   | 0.    | +.01 | 0.    | 0    | 0    | 0    | 0    |      |      |
| H9  | -.09 | -.07 | -.06 | -.05 | -.005 | +.02 | +.04  | +.06 | +.07 | +.09 | +.10 |      |      |
|     | 245  | 250  | 255  | 260  | 265   | 270  | 275   | 280  | 285  | 290  |      |      |      |
| H5  | +0.  | +0.  | +0.  | 0    | 0     | 0    | -.01  | -.02 | -.03 | -.05 |      |      |      |
| H1  | +.14 | +.19 | +.22 | +.24 | +.26  | +.28 | +.29  | +.31 | +.31 | +.32 |      |      |      |
| H5' | +.04 | +.05 | +.06 | +.06 | +.05  | +.06 | +.06  | +.08 | +.08 | +.08 |      |      |      |
| H9  | +.15 | +.15 | +.16 | +.17 | +.17  | +.18 | +.17  | +.18 | +.18 | +.18 |      |      |      |

Table 4.6: Observed NOE effects for BIC & MeBIC.

|                 | BIC in<br>acetone          | BIC in<br>CD <sub>2</sub> Cl <sub>2</sub> | MeBIC<br>in D <sub>2</sub> O |
|-----------------|----------------------------|-------------------------------------------|------------------------------|
| H6 <sup>1</sup> | H5 <sup>1</sup> , (H1, H9) | -                                         | -                            |
| H5 <sup>1</sup> | H6 <sup>1</sup>            | -                                         | -                            |
| H5              | (H4)                       | H4, (H1)                                  | -                            |
| H9              | H8, H1, H6 <sup>1</sup>    | H8, H6 <sup>1</sup> , H1, H3a             | Me(eq), H1, H6 <sup>1</sup>  |
| H1              | Me, H8, H9                 | -                                         | H8, H6 <sup>1</sup> , Me(ax) |
| H8              | (H6 <sup>1</sup> )H1, H9   | H9, H1                                    | -                            |
| H3a             | -                          | H9, H1, H4e, H6 <sup>1</sup>              | -                            |
| Me(eq)          | -                          | -                                         | H3, H9, H1, Me(ax)           |
| Me(ax)          | -                          | -                                         | Me(eq), H1, (H5)             |
| H4e             | -                          | H5, H3, H4, (H6 <sup>1</sup> )            | -                            |

Effects in brackets are either small or possible 'knock-on' effects.

The high barrier at  $\theta_1=340^\circ$  found in the gas-phase molecular mechanics calculations is due to internal steric interactions and would not be significantly affected by solvent. Inter-minima conversion is of course possible via much lower barriers at other  $\theta_1$  values (ca 30 kJ mol<sup>-1</sup> in the gas phase, and the  $\approx 40$  kJ mol<sup>-1</sup> estimate found in solution). High-temperature spectra were reproduced after lowering and then raising the temperature again, demonstrating experimentally that inter-minima conversion takes place readily in solution.

The NMR results suggest that the solvents used have little effect on these barriers.

#### 4.2.3 The nitrogen-ring conformation ( $\theta_2$ - see p73).

With the assumption that the N-methyl group is above the plane of the molecule, away from the phthalide group, there are four theoretically possible conformations of this ring. The absence of exchange effects for the N-ring protons demonstrates that the N-ring is effectively locked into one conformation due to the bulky phthalide substituent. In comparison 2-methyl tetrahydro isoquinoline is conformationally mobile (Katritzky et al, 1981). And the NMR spectrum for the isoquinoline derivative ISO is for averaged ring conformations, even at low temperature (Appendix A2).

Evidence that this ring is in the expected (Andrews and Johnston, 1973) pseudochair conformation (with the phthalide group pseudoaxial) was obtained by the following rationale.

Of the observed nitrogen-ring proton couplings (Table 4.7) those of 15.6Hz and 11.6Hz are geminal, as they are too large to be vicinal couplings. This leaves either 6.6Hz or 7.7Hz for the axial coupling as the other couplings are too small to be axial.



An NOE effect is observed for H5/ $\delta_{2.61}$ , implying that  $\delta_{2.61}$  is H4. With 6.6Hz as the axial coupling this proton will be axial leading to the assignments given in the Table. And the NOE effects observed for H9/H3a and H6'/H3a require H3a and the phthalide ring to be axial, which is only possible with the nitrogen-ring pseudochair. (The axial coupling of 6.6Hz is small due to an expected ring distortion. Our MM2 calculations (2.5) showed the ax/ax torsion angle to be ca 200°.

For the alternative choice, viz 7.7Hz as the axial coupling, no agreement is found with the NOE data for H9/ $\delta_{2.80}$  (H3a) and H6'/ $\delta_{2.80}$  (H3a).

Table 4.7. N-ring proton coupling constants for BIC at 296K in acetone.

| Proton | coupling constant (Hz) |      |       | chemical shift (ppm) |
|--------|------------------------|------|-------|----------------------|
| H4e    | 4.5                    | 7.65 | 15.65 | 2.20                 |
| H3e    | 4.05                   | 7.7  | 11.7  | 2.51                 |
| H4a    | 4.2                    | 6.45 | 15.65 | 2.61                 |
| H3a    | 4.6                    | 6.6  | 11.3  | 2.80                 |

The lack of any NOE effects for H6'/Me and H9/Me suggests the absence of any other low-energy conformations of the nitrogen-ring in solution. This is in agreement with earlier PCILO calculations for the gas-phase molecule (Andrews and Johnston, 1973), and with the crystal conformation (Gilardi, 1973, Gorinsky and Moss, 1973). Calculation of the N-ring proton couplings using the CNDO/2 (and INDO) methods (CNINDO74 - Dobosh and Osland, 1975) gave results in very poor agreement with experiment (ax/ax=17.1Hz, ax/eq=5.3Hz, gem=3.1Hz).

In  $\alpha$ -narcotine, a phthalide isoquinoline alkaloid with a methoxy group at C8, the sofa conformation is found in the crystal (Moss and Watson, 1984).

#### 4.3 The energy profile of MeBIC in solution.

For MeBIC the small temperature-dependence of the NMR spectrum implies that just one conformation is present in solution. A counterion effect on chemical shifts (Table 4.8) was observed and once the effect was allowed for the position of the H8 chemical shift and the H1/H9 coupling constant were used, with support from NOE, to identify the conformation (4.3.1). A lower bound for the energy of the next low-energy conformation was then estimated from a knowledge of the accuracy which the H1/H9 coupling constant has been measured (4.3.2).

##### 4.3.1 Evidence for only one minimum in solution ( $\theta$ 1).

For the chloride and iodide salts of MeBIC in acetone, in DMSO, and in deuterium oxide<sup>\*</sup>, the H1/H9 coupling constant was found to be invariant with temperature, implying that there is just one low-energy conformation of MeBIC in solution. (The slight solvent dependence of this coupling is dealt with later in this section.) Also no pattern of temperature-dependence of chemical shifts could be found which was consistent with conformational changes with temperature (compare Table 4.9 with BIC data in Table 4.2). The following experimental observations are used to show that this conformation is in the 255° - 290° range (and not the 70° - 110° range mentioned earlier):

---

\* At high temperature (363K) the aqueous solution turned slightly yellow. However, the infra-red spectrum indicated no significant signs of decomposition with the gamma-lactone peak dominant at 1780 cm<sup>-1</sup> and no spurious absorptions (see Edwards and Handa (1961) for an infra-red comparison.)

Table 4.8.

4.3.1

Counterion effects on Me BIC proton chemical shifts (ppm)

| <u>chloride counterion (299K)</u> |         |                  |                            |                                                                          |
|-----------------------------------|---------|------------------|----------------------------|--------------------------------------------------------------------------|
| Proton                            | acetone | +1 drop<br>water | 50/50<br>acetone/<br>water | $\Delta\delta_{\text{solv}}$<br>(acetone $\rightarrow$ 50/50<br>mixture) |
| H1                                | 5.96    | 5.68*            | 5.40                       | -0.56                                                                    |
| H9                                | 6.68    | 6.71             | 6.67                       | -0.01                                                                    |
| H6'                               | 7.80    | 7.60*            | 7.54                       | -0.26                                                                    |
| H8                                | 6.85    | 6.85             | 6.8                        | -0.05                                                                    |
| H5                                | 5.81    | 5.80             | 5.73                       | -0.08                                                                    |
| H5'                               | 7.43    | 7.43             | 7.41                       | -0.02                                                                    |

| <u>iodide counterion (323K)</u> |         |                  |                   |                 |                                                               |                                                                   |
|---------------------------------|---------|------------------|-------------------|-----------------|---------------------------------------------------------------|-------------------------------------------------------------------|
| Proton                          | acetone | +1 drop<br>water | +2 drops<br>water | water<br>(323K) | $\Delta\delta_{\text{solv}}$<br>(acetone $\rightarrow$ water) | $\Delta\delta_{\text{ion}}$<br>(in acetone)<br>I $\rightarrow$ Cl |
| H1                              | 6.46    | 5.56*            | 5.41              | 5.01            | -1.45                                                         | -0.50                                                             |
| H9                              | 6.63    | 6.64             | 6.63              | 6.55            | -0.08                                                         | 0.05                                                              |
| H6'                             | 8.19    | 7.67*            | 7.59              | 7.36            | -0.83                                                         | -0.39                                                             |
| H8                              | 6.78    | 6.79             | 6.80              | 6.81            | 0.03                                                          | -0.07                                                             |
| H5                              | 5.83    | 5.75             | 5.74              | 5.67            | -0.16                                                         | -0.02                                                             |
| H5'                             | 7.34    | 7.36             | 7.38              | 7.38            | 0.04                                                          | 0.09                                                              |

\*The chemical shift change must be due more to a counterion effect than the change in solvent, as the greatest shift change is on dissociation when one drop of water is added.

Table 4.9

Effect of temperature on MeBIC chemical shifts (ppm).

|        | acetone | acetone | water | water |
|--------|---------|---------|-------|-------|
| proton | 208K    | 323K    | 299K  | 363K  |
| H1     | 6.24    | 6.46    | 5.05  | 4.95  |
| H9     | 6.72    | 6.63    | 6.56  | 6.52  |
| H6'    | 8.04    | 8.19    | 7.35  | 7.37  |
| H5     | 6.93    | 6.78    | 6.78  | 6.82  |
| H8     | 5.69    | 5.83    | 5.66  | 5.68  |
| H5'    | 7.49    | 7.34    | 7.36  | 7.40  |

Table 4.10

Effect of temperature on HBIC chemical shifts (ppm) in D2O/DCI.

| proton | 287K | 296K  | 323K  |
|--------|------|-------|-------|
| H1     | 5.1  | 5.1   | 5.035 |
| H9     | 6.26 | 6.259 | 6.23  |
| H6'    | 7.16 | 7.158 | 7.13  |
| H5     | 6.82 | 6.828 | 6.82  |
| H8     | 5.91 | 5.914 | 5.91  |
| H5'    | 7.35 | 7.35  | 7.34  |

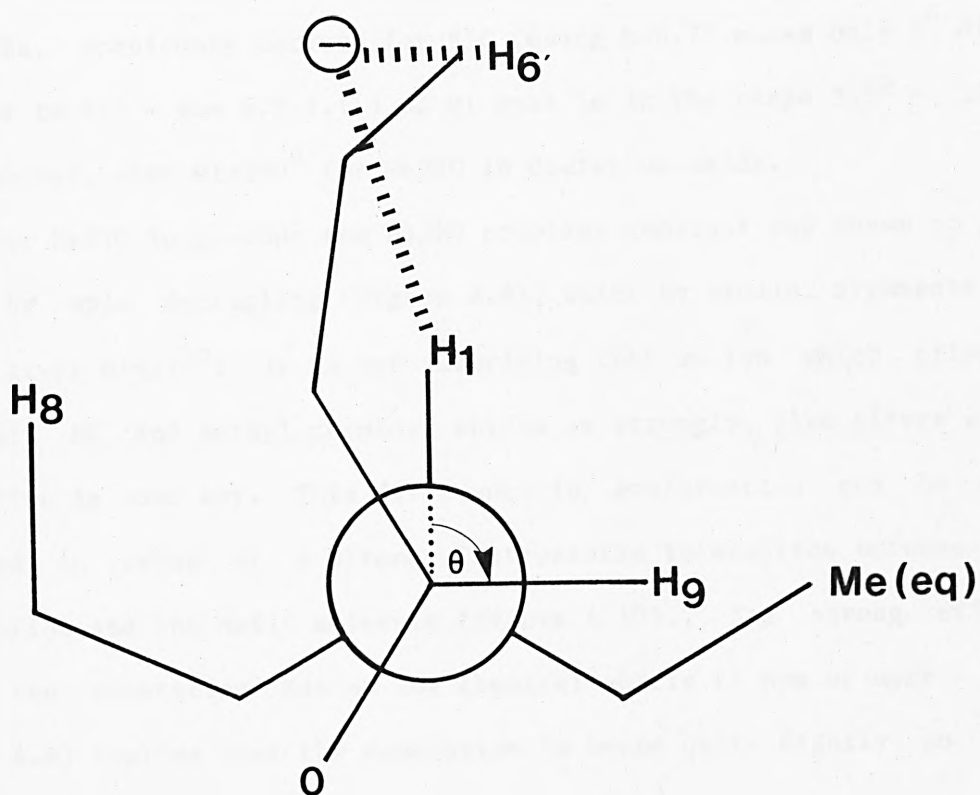
(i) The H8 chemical shift is changed from its base shift value (4.2.1.2) in a way which is consistent with benzene-ring shielding for  $\theta_1$  in the range  $255^\circ - 305^\circ$  (-0.8 ppm in acetone, -1.0 ppm in deuterium oxide). The same type of shielding effect has previously been reported for other phthalide isoquinolines (Elango et al, 1982).) The position of the H6' chemical shift cannot be used because it is counterion dependent (Table 4.8), whereas the H8 peak is not significantly affected.

(ii) The observed nuclear Overhauser effects between  $\text{Me}_{\text{eq}}$  and H9, and between H1 and H9 require  $\theta_1$  in the range  $240^\circ - 310^\circ$ .

The above arguments apply equally to both acetone and deuterium oxide solvents, and define a small range of values for  $\theta_1$ . The counterion effect can be used to determine the position of the counterion and to narrow this range down further:

(iii) In non-dissociating solvents, the chemical shifts of H1, H6' and the two methyl groups are dependent on the counterion ( $\text{Cl}^-$  or  $\text{I}^-$ ) of MeBIC and on temperature and solvent (Table 4.8). In the dissociating solvent, deuterium oxide, no significant temperature-dependence was found, implying that the counterion was primarily responsible for the above observed effects on chemical shifts. It is readily apparent from a Drieding model that only with  $\theta_1$  in the  $210^\circ - 290^\circ$  range can the counterion be close enough to affect all these proton shifts, and no others (Figure 4.8).





**Figure 4.8** Position of the counterion in MeBIC.

$\theta_1 \approx -90^\circ$  gives the best fit for the ion to be close to H6' and H1, but not as close to H8, H9 or the equatorial methyl group.

An accurate value for  $\theta_1$  could then be obtained from the H1/H9 coupling constant. For MeBIC in deuterium oxide this coupling was observed to be 1.0Hz, measured for the H9 peak at 6.56ppm (299K). (This peak was significantly broader than the other peaks, which could be due to some effect of the iodide ion on this chemical shift.) Application of the Karplus equation to this 1 Hz coupling yields four possible  $\theta_1$  values of:  $70^\circ$ ,  $110^\circ$ ,  $250^\circ$  or  $290^\circ$ . (This is with  $k=8.8$  Hz, previously derived for BIC (using  $k=6.74$  makes only  $2^\circ$  difference to  $\theta_1$ ) - see 4.2.1.1.) As  $\theta_1$  must be in the range  $255^\circ - 290^\circ$  (see above), then  $\theta_1=290^\circ$  for MeBIC in deuterium oxide.

For MeBIC in acetone the H1/H9 coupling constant was shown to be zero by spin decoupling (Figure 4.9), which by similar arguments to above gives  $\theta_1=270^\circ$ . It is not surprising that an ion which affects the H1, H6' and methyl chemical shifts so strongly, also alters conformation in some way. This difference in conformation can be explained in terms of a strong electrostatic interaction between the counterion and the MeBIC molecule (Figure 4.10). The strong effect that the counterion has on key chemical shifts (1 ppm or more - see Table 4.8) implies that the counterion is bound quite tightly to the MeBIC molecule in non-dissociating solvents.

A possible alternative explanation for the observed temperature-independence of the MeBIC NMR spectrum (excluding counterion effects) is that two low-energy conformations exist with high barriers between them (see Figure 2.4 for comparison with the gas-phase). This is unlikely, however, since H8 clearly displays shielding associated only with  $\theta_1=270^\circ-290^\circ$ . In addition, the solution conformation would not be expected to be as close to the gas-phase conformation(s) as found with BIC, because the quaternary nitrogen of MeBIC makes it somewhat more polar.

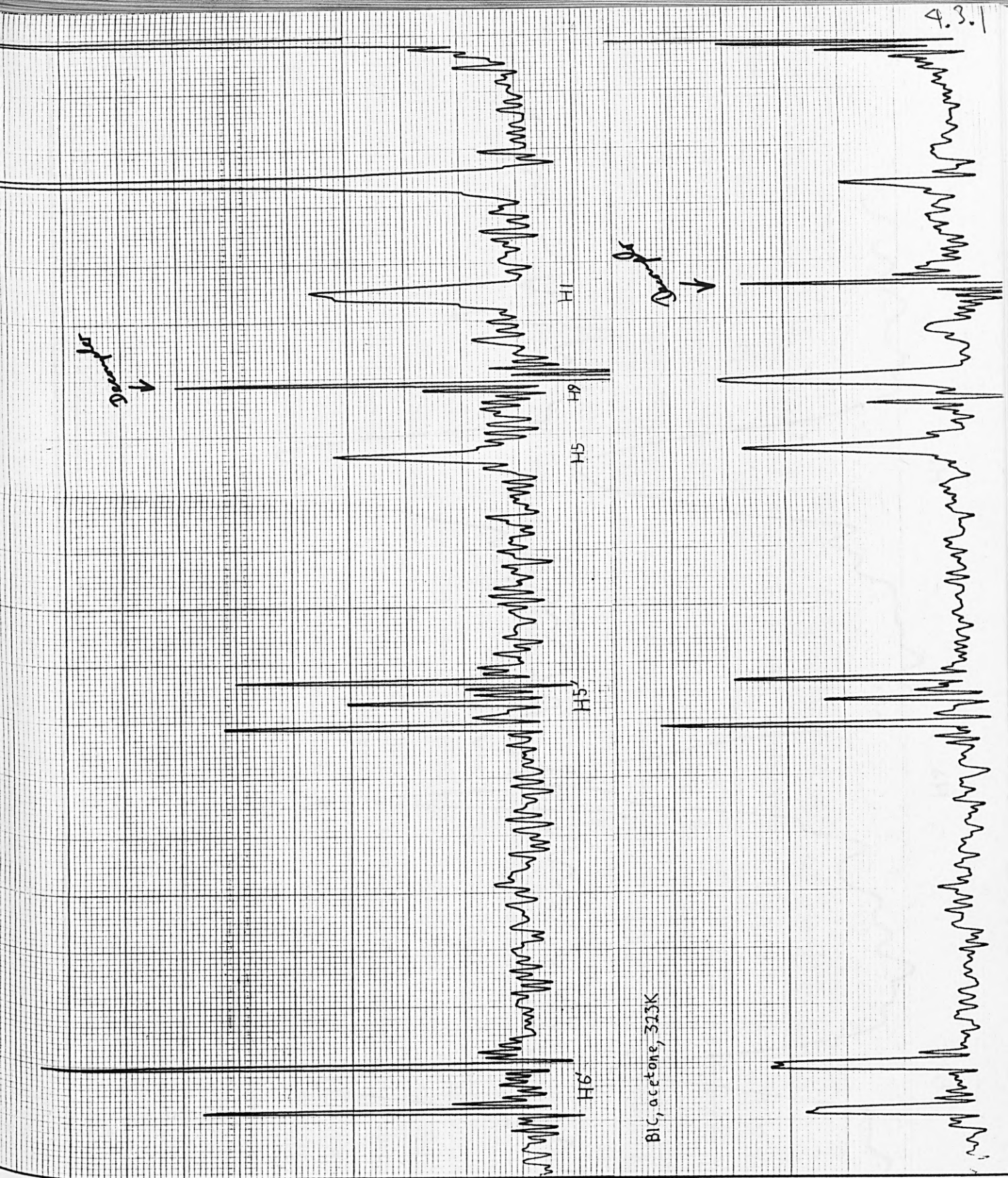


Figure 4.9. Spin decoupling on H1 and H9 for MeBIC in acetone.

Decoupling at H9 has no effect on the peak height or half-width of H1 showing that they are not coupled (see next page).



4.3.1

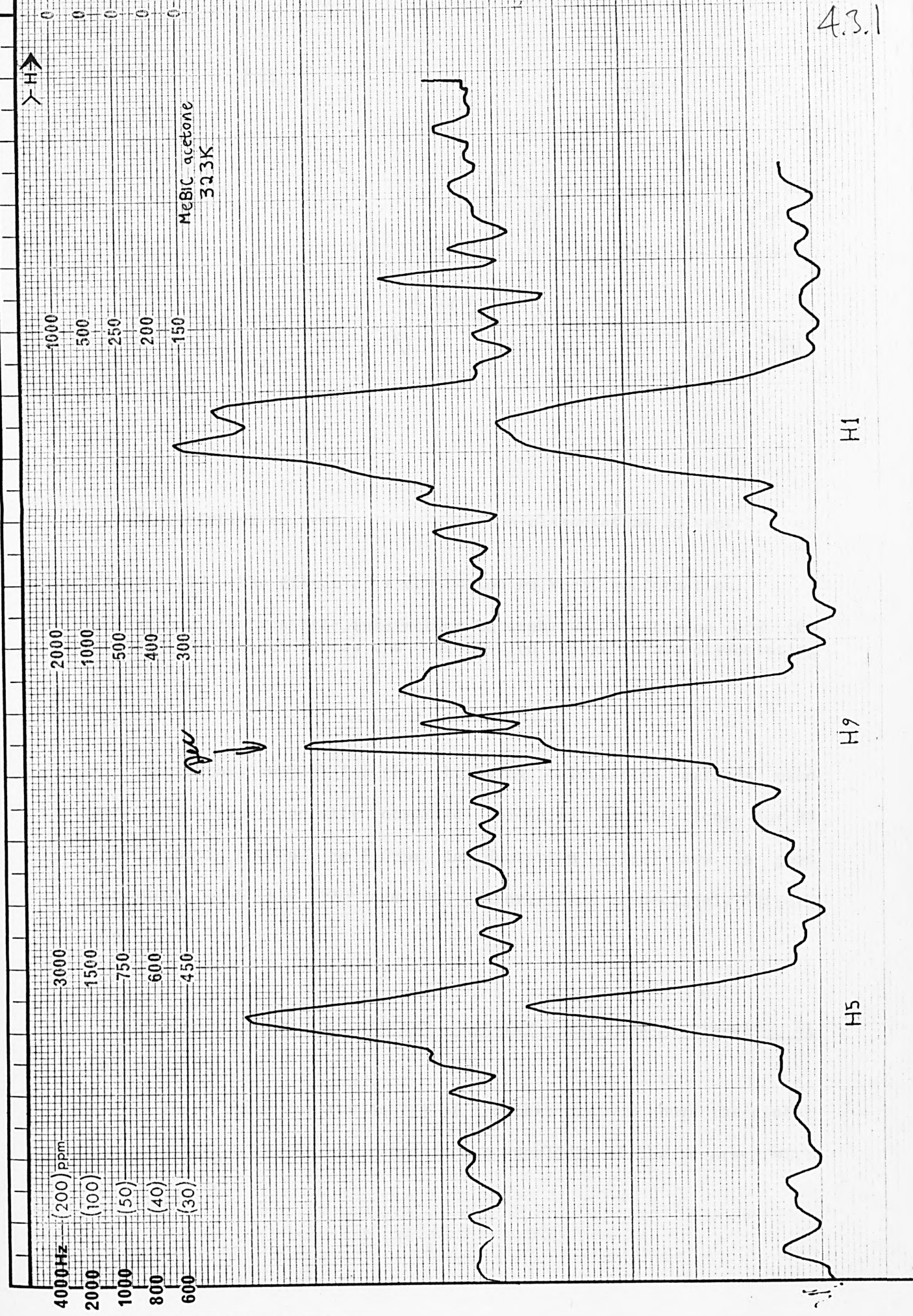


Figure 4.9 (continued).

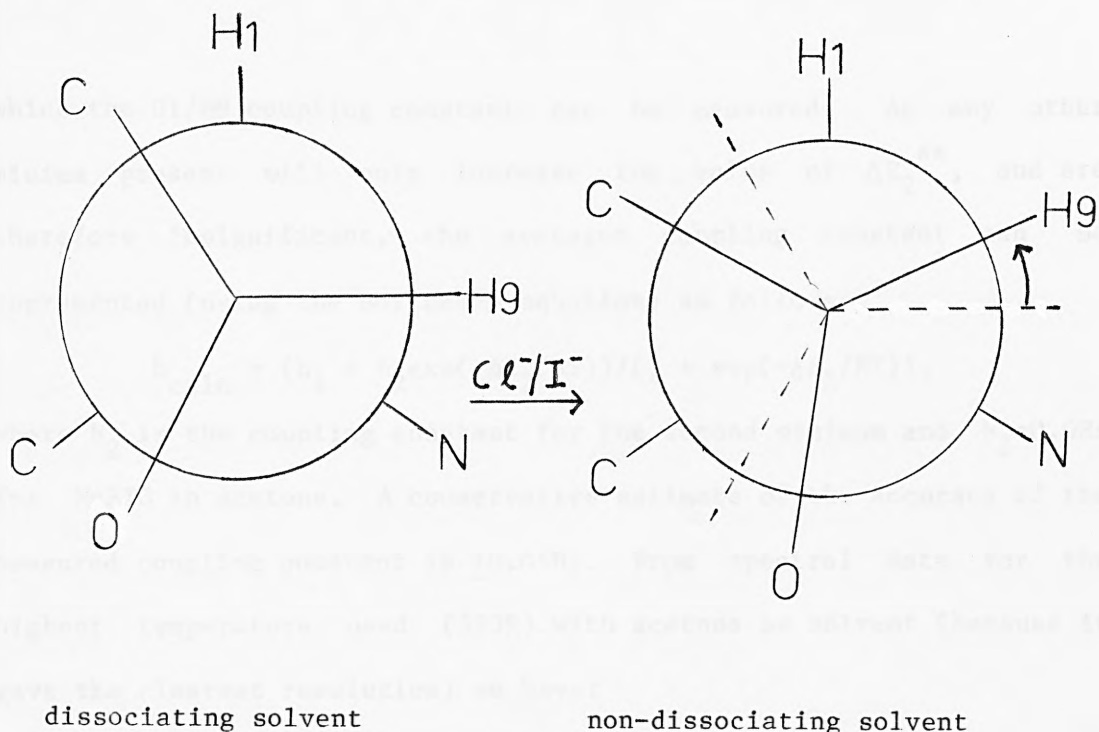


Figure 4.10. The effect of a counterion on MeBIC conformation.

The strong electrostatic interaction between the counterion and the quaternary nitrogen forces the conformation change in non-dissociating solvents.

---

As further support for the shielding calculation methods used here, the experimentally observed shielding contributions for H8 (ca 0.8ppm in acetone, ca 1.0ppm in deuterium oxide) are in good agreement with the  $270^\circ$  and  $290^\circ$  conformations respectively (Figure 4.2).

#### 4.3.2 Significance of minimum (energy of next highest $\theta_1$ conformation).

A lower bound for the energy of the next minimum above the global minimum ( $\Delta E_2$ ) can be estimated from a knowledge of the accuracy with



which the H1/H9 coupling constant can be measured. As any other minima present will only increase the value of  $\Delta E_2^{**}$ , and are therefore insignificant, the averaged coupling constant can be represented (using the Boltzmann equation) as follows:

$$h_{\text{calc}} = (h_1 + h_2 \exp(-\Delta E_2/RT)) / (1 + \exp(-\Delta E_2/RT)).$$

Where  $h_2$  is the coupling constant for the second minimum and  $h_1=0.0\text{Hz}$  for MeBIC in acetone. A conservative estimate of the accuracy of the measured coupling constant is  $\pm 0.05\text{Hz}$ . From spectral data for the highest temperature used (323K) with acetone as solvent (because it gave the clearest resolution) we have:

$$h_{\text{max}}(323) = (h_2 \exp(-\Delta E_2/2701)) / (1 + \exp(-\Delta E_2/2701)) < 0.05\text{Hz}$$

$$\Delta E_2 > -2701 \ln(0.05/(h_2 - 0.05)) \text{ kJmol}^{-1}.$$

Gas-phase MM2 calculations on MeBIC (2.5), with geometry optimisation, give a second minimum at around  $180^\circ$ . From the Karplus parameters derived for BIC (4.2.1.1), this yields an approximate value for  $h_2$  of  $12\text{Hz}$  and  $\Delta E_2 > \text{ca } 11\text{kJmol}^{-1}$ . Although this is less than the BIC (and MeBIC) binding energy of  $30\text{--}40\text{kJmol}^{-1}$  (calculated by applying the Boltzmann equation to the MeBIC( $\text{I}^-$ ) dissociation constant ( $K_D$ ) of  $380\pm 20\text{nM}$  (Mohler and Okada, 1977b)), the fact that no other PIQ GABA antagonists with a different conformational minimum\*, and of greater

\*\* The averaged coupling constant for a system of 3 minima can be represented by:  $h_{\text{calc}} = (h_1 p_1 + h_2 p_2 + h_3 p_3) / (p_1 + p_2 + p_3) < h_{\text{acc}}$ . Where  $h_i$  and  $p_i$  are the couplings and populations respectively for each minimum (set  $p_1=1$  and  $p_1 > p_2$  etc), and  $h_{\text{acc}}$  is the accuracy at which the coupling was measured. Since  $h_1=0.0$  (see below), by rearrangement we have:  $p_2 < (h_{\text{acc}} - h_3 p_3) / h_2$ .

As all the terms in the above equation are positive, the  $h_3 p_3$  term will only lower the derived value of  $p_2$  and therefore increase  $\Delta E_2$ .

\* ie a different N and COO arrangement due to changing the C1/C9 configuration and/or changing substituents at the key H8 or H6' positions. Note that changing the 1S configuration has a much greater effect on potency than changing the 9R configuration (Enna et al, 1977). This is because changing the configuration at C9 does not alter the arrangement of N - O - O charge centres (see Chapter 5).

or equal potency to BIC, have been found reinforces the argument for a  $\theta_1$  of  $270^\circ - 290^\circ$  being the active conformation of MeBIC.

The high magnitude of the counterion effect on the H6' and H1 chemical shifts ( $\Delta\delta_{\text{solv}}$  and  $\Delta\delta_{\text{ion}}$ ) and the comparatively small temperature-dependence of these chemical shifts further imply that in non-dissociating solvents, the counterion remains tightly bound to MeBIC, even at high temperature. However, no (additional) quantitative estimate of  $E^2$  could be made from these effects because the information required for determining conformer populations is lost in the 'noise' from other, unknown temperature-dependent factors.

In addition, it is apparent from examination of a Drieding model that the counterion can gain the maximum number of favourable interactions for  $\theta_1$  in the  $250^\circ - 290^\circ$  range. This will help to stabilise the  $270^\circ$  conformation in acetone, since other possible positions that the counterion would be forced to adopt in different conformations are of higher energy. The  $20^\circ$  conformation change for MeBIC in acetone and deuterium oxide solvents can be explained by the different interaction with the counterion in the different solvents (Figure 4.10).

The fact that the strong electrostatic interaction with a counterion has so small an effect on conformation (ca  $20^\circ$ ) is further evidence that MeBIC is relatively rigid in solution.

#### 4.3.3 The nitrogen-ring conformation ( $\theta_2$ - see p73).

As the N-ring proton coupling constants are not first-order in any of the solvents used, an analysis of the same kind used for BIC (4.2.3) could not be undertaken. However, as the extra methyl group will enhance the rigidity of this ring (Katritsky et al, 1981) it can

be reasonably assumed that the conformation is the same as for BIC. (These same arguments also apply to HBIC.)

#### 4.4 HBIC - evidence for just one conformation in solution ( $\theta_1$ ).

The spectrum of HBIC in deuterium oxide much more closely resembles the MeBIC (in deuterium oxide and acetone) spectra than the spectra for BIC in acetone (compare Table 4.10 with Tables 4.2 and 4.9), even with allowance made for the difference in solvent. (The free base, BIC, was found to be not soluble enough in deuterium oxide to obtain NMR spectra, even on warming the solvent and using 400MHz fourier transform spectroscopy). The positions of the shielded H8 (-0.7ppm) and unshielded H6' (-0.02ppm) chemical shifts imply that  $\theta_1 \approx 270^\circ$ . The H1/H9 coupling constant was shown to be zero by spin decoupling (Figure 4.11), implying that  $\theta_1 = 270^\circ$  for HBIC in deuterium oxide, which is similar to the MeBIC result (4.3).

#### 4.5 Experimental details.

FT-NMR spectra were obtained using the Intercollegiate Research Service at Queen Mary College, London, on a Bruker WH-400 MHz spectrometer. (A few spectra were also recorded using the City University 60 and 100 MHz spectrometers and the University of Sheffield 400 MHz service.)

Solution concentrations used were from  $10^{-3}M$  to saturated. Low concentrations were used to minimise the possible effect of solute - solute interactions on chemical shift, and several solvents were used in order that the effect of solvent on conformation could be examined. Acetone-D6 was chosen as a solvent because of its good solubility and miscibility properties, the clear resolution of spectra taken in it,

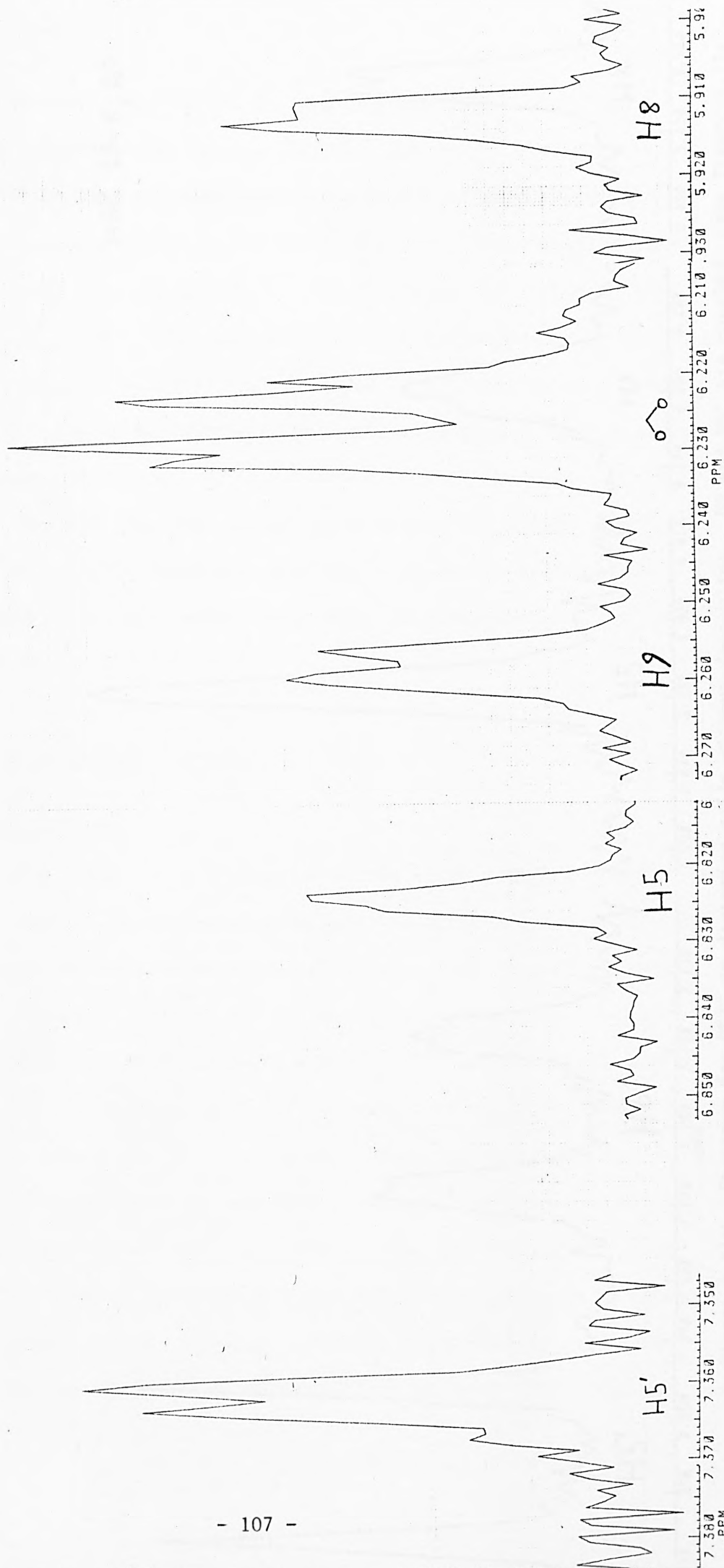


Figure 4.11a. Spin decoupling at H6' (7.16ppm) for HBIC in deuterium oxide. The apparent splitting in the H9 peak is due to noise in the spectrum (cf H5' and H8, and Figure 4.11b).

HBIC, 296K, D<sub>2</sub>O

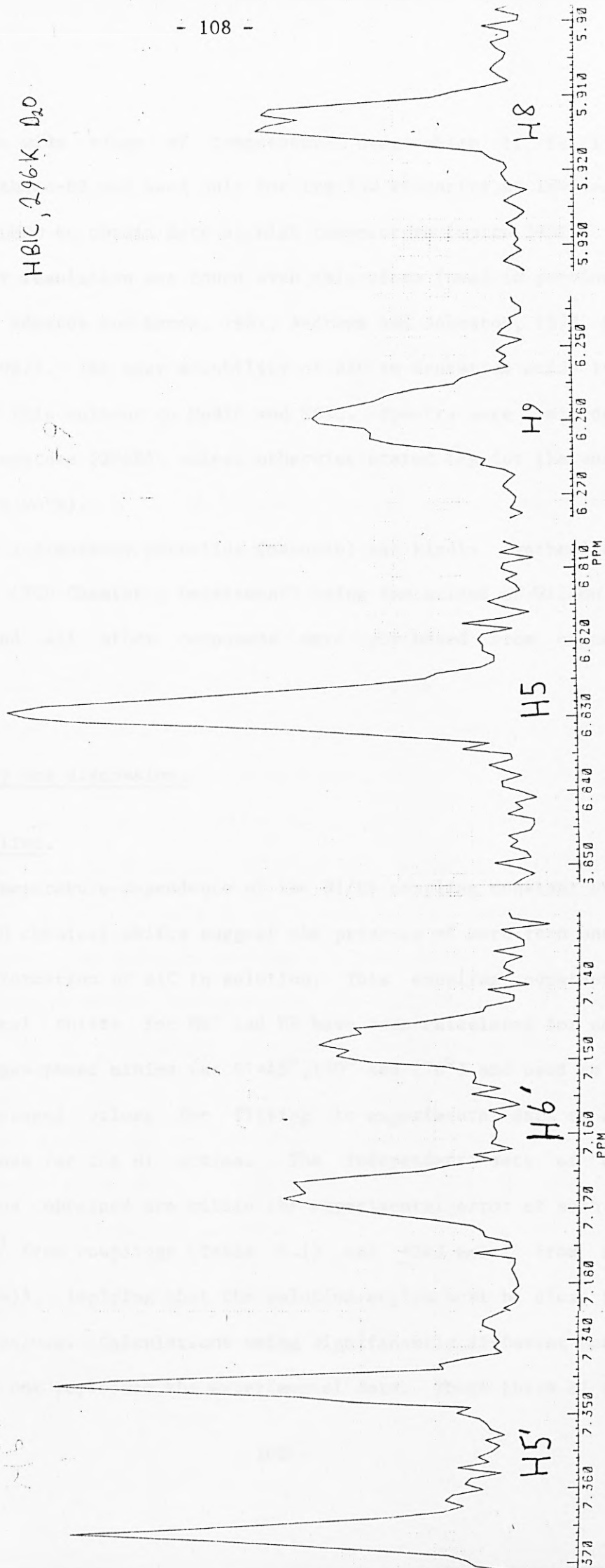


Figure 4.11b. Spectrum for HBIC in deuterium oxide without decoupling. H9 and H6' are coupled (see Figure 4.11a).



and for the wide range of temperature over which it is liquid. Dichloromethane-D<sub>2</sub> was used only for its low viscosity at low temperature, and DMSO to obtain data at high temperature (up to 393K). Relatively poor resolution was found with chloroform (used in previous BIC NMR work - Edwards and Handa, 1961, Andrews and Johnston, 1973, Elango et al, 1982). The poor solubility of BIC in deuterium oxide limited the use of this solvent to MeBIC and HBIC. Spectra were recorded at room temperature (296K), unless otherwise stated (eg for the variable temperature work).

The 6,7-dimethoxy phthalide (meconin) was kindly synthesized by Mr T. Rose (TCU Chemistry Department) using the method of Wilson et al (1951), and all other compounds were purchased from commercial sources.

#### 4.6 Summary and discussion.

##### (i) Bicuculline.

The temperature-dependence of the H<sub>1</sub>/H<sub>9</sub> coupling constant and the H<sub>6'</sub> and H<sub>8</sub> chemical shifts suggest the presence of more than one low-energy conformation of BIC in solution. This coupling constant and the chemical shifts for H<sub>6'</sub> and H<sub>8</sub> have been calculated for each of the three gas-phase minima (at  $\theta_1=45^\circ, 170^\circ$  and  $270^\circ$ ) and used to calculate averaged values for fitting to experimental data to obtain energy values for the  $\theta_1$  minima. The independent sets of energy values thus obtained are within the experimental error of each other ( $\pm 1 \text{ kJ mol}^{-1}$  from couplings (Table 4.1) and  $\pm 2 \text{ kJ mol}^{-1}$  from shifts (Table 4.4)), implying that the solution angles must be close to the gas-phase values. Calculations using significantly different sets of angles do not reproduce the experimental data. These three  $\theta_1$  values

are all found in the condensed state:  $50^{\circ}$  previously found in solution (Andrews and Johnston, 1973, Elango et al, 1982),  $172^{\circ}$  in the crystal (Gilardi, 1973, Gorinsky and Moss, 1973), and ca  $270^{\circ}$  for MeBIC salts in solution (4.3). All the evidence suggests a  $\theta_1$  energy profile for BIC in solution similar to the gas-phase energy-surface (Figure 4.7). This suggests that the hindrance to rotation for BIC is mostly steric, which is supported by the somewhat greater hindrance to rotation observed in HBIC and MeBIC.

NOE studies give further evidence for the presence of more than one low-energy conformation in solution.

An estimate of the barrier to inter-conversion between  $\theta_1$  minima was made, based on the temperature nearest to coalescence of the BIC NMR spectrum. This gave an upper limit to the barrier of ca  $40 \text{ kJ mol}^{-1}$  for BIC in dichloromethane solution. Corresponding roughly to the energy of one average hydrogen-bond, it can be considered to be relatively small. The net effect is therefore relatively free oscillation about one broad shallow well and restricted only in the range  $\theta_1 \approx 310^{\circ} - 360^{\circ}$ , this being the steric barrier readily seen with a Drieding model, and found by gas-phase calculations.

With regard to the nitrogen-ring, the conformation of this was confirmed to be pseudochair in solution, as in the crystal, from NOE studies combined with an examination of the nitrogen-ring coupling constants. It is also the conformation found from gas-phase calculations, though a different conformation is found in the related phthalide isoquinoline,  $\alpha$ -narcotine.

#### (ii) Bicuculline methohalides and protonated bicuculline.

The lack of temperature-dependence of the  $H_1/H_9$  coupling constant for the chloride and iodide MeBIC salts, in all solvents used, implies

that there is just one low-energy  $\theta_1$  conformation of these salts in solution, with the next highest conformation at least  $11\text{kJmol}^{-1}$  higher in energy. In aqueous solution this conformation is  $290^\circ$  and, with acetone as solvent, a  $\theta_1$  value of  $270^\circ$  is found. This difference in conformation is explained in terms of an electrostatic interaction between the counterion and the MeBIC molecule (Figure 4.8).

HBIC also appears to be in just one conformation at  $\theta_1 \approx 270^\circ$ , from a comparison of HBIC with MeBIC and BIC NMR spectra.

For the N-ring conformation ( $\theta_2$ ) the N-ring proton coupling constants for HBIC and MeBIC were not first-order in any of the solvents used and  $\theta_2$  could not therefore be determined.

### (iii) Biological significance of these results.

It can now be deduced that the active conformation of MeBIC and HBIC must be  $\theta_1 \approx 270^\circ - 290^\circ$ , because any other solution conformations are of at least  $11\text{kJmol}^{-1}$  higher energy. And, although this energy is possibly less than the MeBIC binding energy of  $30-40\text{kJmol}^{-1}$ , the fact that no other PIQ GABA antagonists with a different conformational minimum and of greater or equal potency to BIC have been found (see eg Enna et al, 1977, Kardos et al, 1984) reinforces this argument.

In an attempt to show that a positive nitrogen-region is not essential for antagonist activity, Kardos and coworkers (1984) have used the apparently opposing effects that N-methylation has on the PIQs BIC and adlumidine (ADD) binding data (BIC binds ca 6 times more strongly than MeBIC, and ADD binds ca 15 times less strongly than MeADD (Kardos et al, 1984)). The N-methyl derivative of BIC should be ca 10 times more potent than the base compound due to only ca 10% of BIC (and ADD) being in the protonated form at physiological pH (Kardos et al, 1984). However, *in vivo* data (Johnston, 1972) for MeBIC and BIC show MeBIC to

be more potent than BIC (as with ADD and MeADD). An active conformation of  $\theta \approx 270^\circ - 290^\circ$  gives further support to BIC being in the protonated form (rather than the free base) when binding to the receptor, because  $< \text{ca } 20\%$  of BIC is in the required conformation (Figure 4.7).

In addition, the lack of positive charge around the N of BIC (base) is incompatible with the positively-charged N-region in GABA (see Figure 5.4), whereas the charges in HBIC and MeBIC are similar to those of GABA (Steward et al, 1975).

## 5 Structural comparisons between GABA, BIC and semi-rigid GABA analogues.

### 5.1 Introduction.

In this Chapter we consider the structural requirements for drugs active at the GABA<sub>A</sub> receptor. We shall see that separate requirements are found for GABA agonists and antagonists (5.3). This explains previous discrepancies found (Krogsgaard-Larsen et al, 1978, Andrews and Johnston, 1979, Galli et al, 1980, and see 5.2) between GABA conformations derived by comparing the structures of GABA with GABA agonists and with the antagonist BIC.

In addition, the flexibility of the GABA<sub>A</sub> receptor in relation to drug flexibility (5.4), and factors which possibly contribute to the antagonist action of BIC are also discussed (5.5).

### 5.2 Previous structural comparisons of GABA, BIC and GABA agonists.

Many previous comparisons of the molecular structure of GABA with BIC (and MeBIC) and GABA analogues have been made in an attempt to discover the structural requirements for activity at the GABA<sub>A</sub> receptor\* (eg Curtis et al, 1970, Beart et al, 1971, Andrews and Johnston, 1979, Sytinskii, 1978, Krogsgaard-Larsen et al, 1983), though in doing so no consideration was given to why BIC is an antagonist rather than an agonist. Mohler and Okada (1977b, 1978) have deduced the existence of different agonist and antagonist conformational states of

---

\* It is now generally accepted that a negatively charged region (eg COO<sup>-</sup> in GABA or O-C=O in BIC) and a positively-charged nitrogen-region (which may be delocalised, but at the cost of lowering activity - see the examples in Table 5.1) are essential for GABA agonist and antagonist activity (Johnston, 1976, Krogsgaard-Larsen et al, 1983 - see also 4.6 iii on N<sup>+</sup> as an antagonist requirement). It is the required spatial arrangement of these charged groups (and others - see below) which we now consider.



the GABA receptor. The evidence, however, was based on the effect of monovalent anions on bicuculline methohalide (MeBIC) binding and could alternatively be explained by the existence of a "multiplicity of GABA receptors" (Andrews and Johnston, 1979). And, although these results were reported before GABA 'A' and 'B' sites were established (Hill and Bowery, 1981, Olsen, 1981), the GABA<sub>A</sub> (BIC sensitive) site could possibly be further sub-divided (Krogsgaard-Larsen and Nielsen, 1984). In none of this work, however, has any distinction been made between the structural requirements for agonist and antagonist drug molecules.

When Curtis and coworkers originally compared Dreiding models of BIC and GABA (Curtis et al, 1970) they found an isosteric match of the  $N^+$  and  $COO^-$  of a relatively extended GABA molecule, with the  $N^+$  and lactone  $C-C=O$  of BIC (Figure 5.1). The possibility of exact congruence of the nitrogen and  $COO$  charge centres in both molecules was later suggested (Steward et al, 1971) (Figure 5.2a), but Curtis and coworkers (Beart et al, 1971) ruled this out because it would involve a GABA molecule too folded for congruence with the semi-rigid GABA agonists muscimol (MUS) and 4-amino tetrolic acid (4ATA). Again these comparisons did not take into account the fact that GABA is an agonist and BIC an antagonist.

Figure 5.1 Match of the N, COO group and carbon chain of GABA (left) with the N, C=O group and part of the carbon skeleton of BIC (based on Curtis et al, 1970). The only way that the carbon skeletons can match is with the CH<sub>2</sub>CO groups aligned. (Hydrogen atoms have been excluded for clarity.)

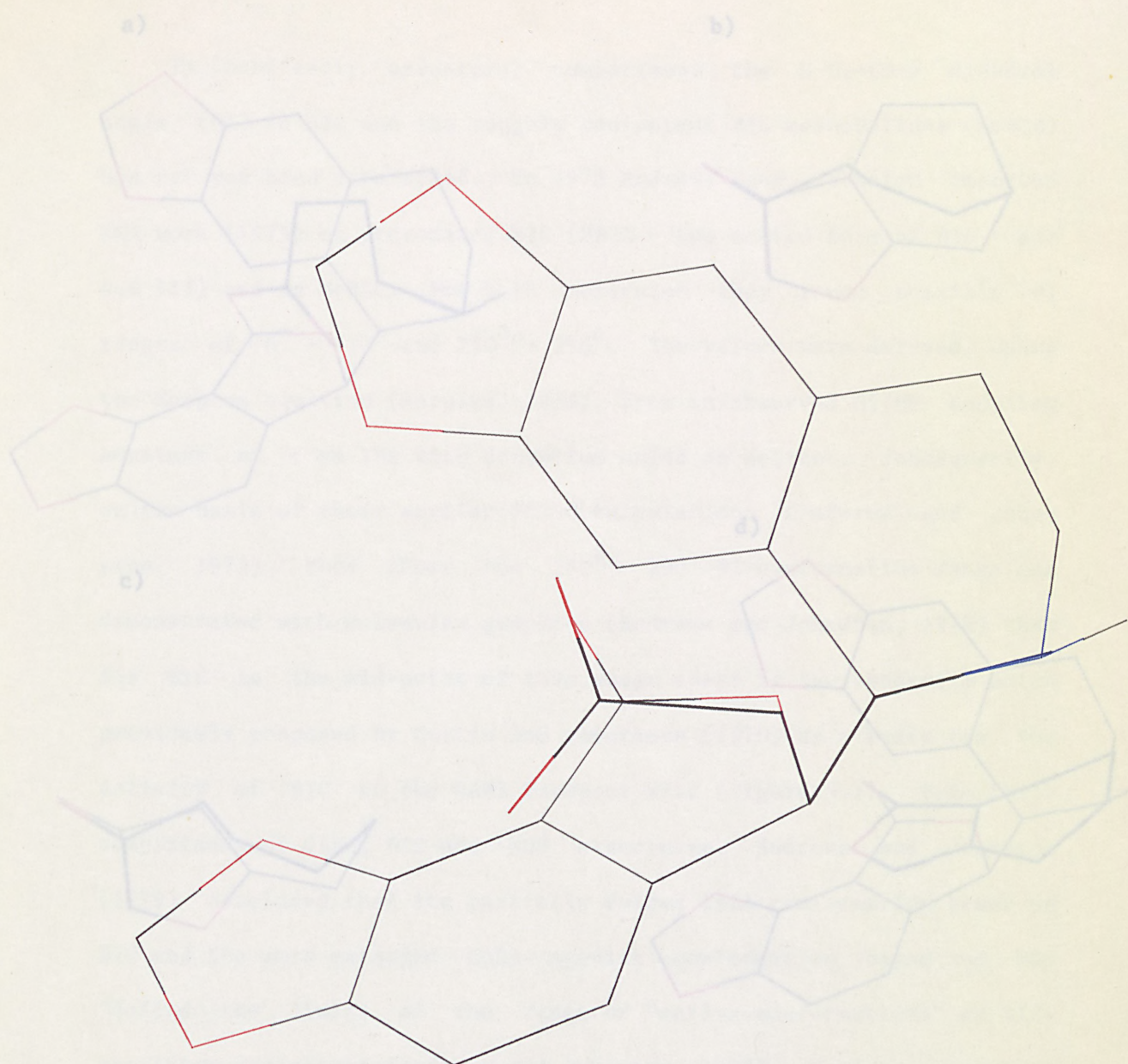


Figure 5.2. Molecular comparisons by matching charge centres.

Figure 5.1 Match of the N, COO group and carbon chain of GABA (bold) with the N, C-C=O group and part of the carbon skeleton of BIC (based on Curtis et al, 1970). The only way that the carbon skeletons can match is with the COO/CCO groups matched. (Hydrogen atoms have been excluded for clarity.)



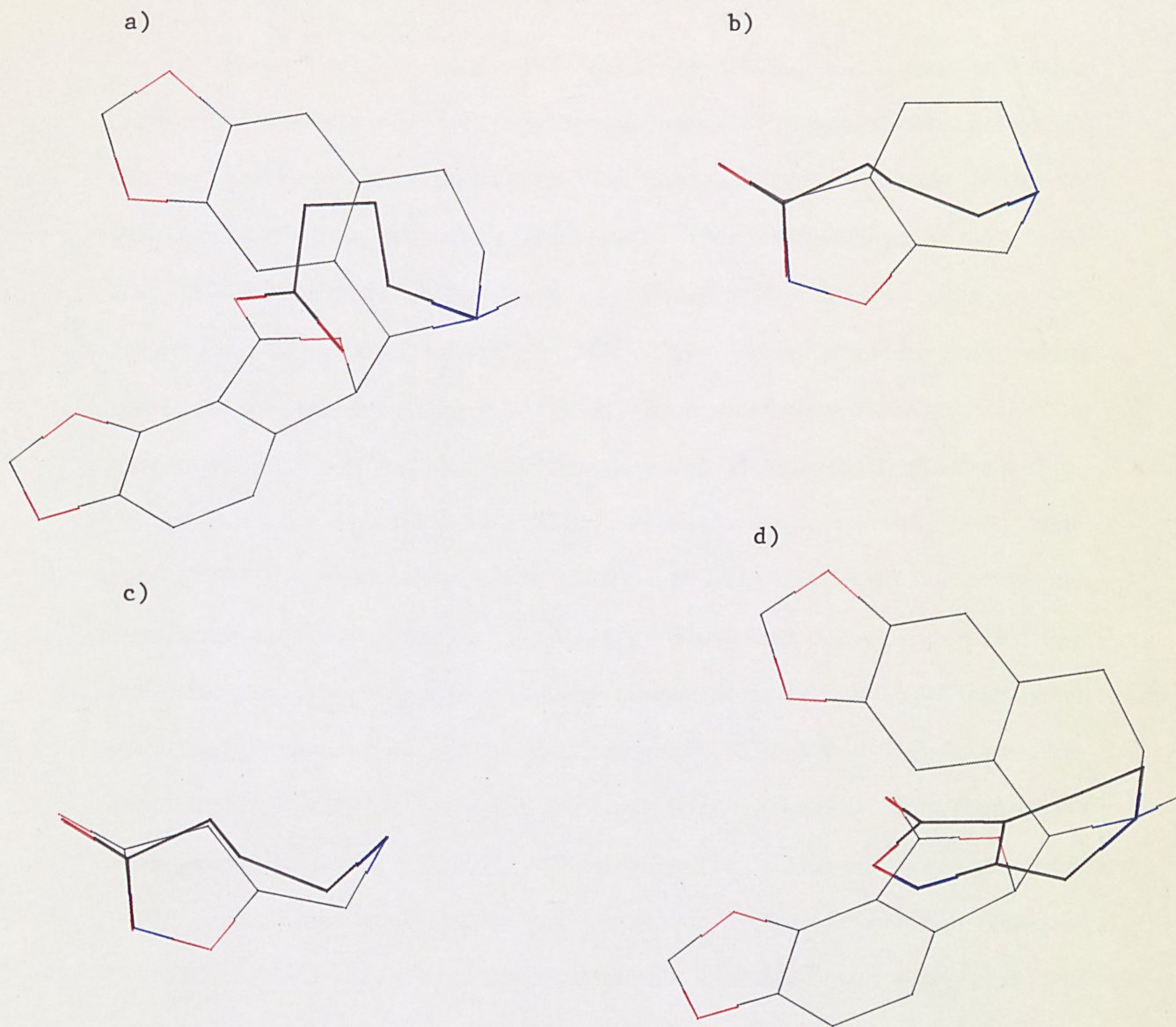


Figure 5.2 Molecular comparisons by matching charge centres.

(a)HBIC/GABA, (b)THIP/GABA, (c)MUS/GABA, (d)iso-THIP/HBIC. For charges see Figure 5.4. (Hydrogen atoms have been excluded for clarity.)

In those early structural comparisons the H-C1-C9-H dihedral angle ( $\theta_1$ ) in BIC and the roughly equipotent BIC methohalides (MeBIC) had not yet been determined. In 1973 Andrews and Johnston reported NMR work (1973) on protonated BIC (HBIC - the active form of BIC - see 4.6 iii) and on MeBIC. For both molecules they found possible  $\theta_1$  ranges of  $70^\circ - 110^\circ$  and  $250^\circ - 290^\circ$ . The values were derived, using the Karplus equation (Karplus, 1959), from an observed H1/H9 coupling constant of  $< \text{ca } 1\text{Hz}$  with deuterium oxide as solvent. Subsequently, on the basis of their earlier PCILO calculations (Andrews and Johnston, 1973), they chose the  $250^\circ - 290^\circ$   $\theta_1$  conformation range and demonstrated with molecular graphics (Andrews and Johnston, 1979) that for BIC in the mid-point of this range there is the isosteric match previously proposed by Curtis and coworkers (1970) as a basis for the activity of BIC at the GABA receptor site (Figure 5.1). From their comparisons of GABA, BIC and MUS structures, Andrews and Johnston (1979) concluded that the partially folded GABA conformation based on BIC and the more extended GABA agonist conformation based on MUS 'define the limits of the range of "active-conformations" at BIC-sensitive receptors' (Andrews and Johnston, 1979). Walters and Hopfinger (1984) have obtained a similar result to the above 'GABA MUS conformation' by comparing GABA with agonists only (ie not including BIC), and using a rather elaborate 'Molecular Shape Analysis' approach (Weintraub and Hopfinger, 1975, Potenzone et al, 1977).

In addition, a "tridentate electrostatic interaction between GABA agonists and GABA receptors" has been suggested (Galli et al, 1980, Krogsgaard-Larsen et al, 1978), which involved a charge centre match and does not include the GABA skeleton, as suggested earlier for GABA and BIC (Steward et al, 1971).



For BIC we have now shown experimentally, using variable-temperature NMR, nuclear Overhauser enhancement (NOE) and different solvents, that  $\theta_1=270^\circ-290^\circ$  is the active conformation. (This was deduced from the single conformations of MeBIC and HBIC found in solution, and was the main aim of Chapter 4 - see 4.6 iii.)

If the structures of GABA, MeBIC, HBIC, and semi-rigid GABA agonists are now compared using the above N and C-C=O match (Curtis et al, 1970) no particular distinction is made between agonist and antagonist structural requirements for activity at the GABA<sub>A</sub> receptor (compare Figures 5.2 and 5.2b). However, if one reverts to the match of N and COO charge centres in both molecules, which has been suggested both for agonists (eg Curtis et al, 1970, Krogsgaard-Larsen et al, 1978, Galli et al, 1980), and for antagonists (Steward et al, 1971, 1975), **separate** agonist and antagonist structural requirements can now be clearly identified (5.3).

### 5.3 Distinction between agonist and antagonist structural requirements.

Comparing the charge centres in GABA with those in the potent GABA<sub>A</sub> agonists 4,5,6,7-tetrahydro isoxazolo [5,4-c] pyridin-3-ol (THIP) and iso-guvacine (IGUV) yields a relatively extended GABA molecule (as previously found - eg Curtis et al, 1970, Andrews and Johnston, 1979, Galli et al, 1980, Krogsgaard-Larsen et al, 1983, Nicholson, Suckling and Iverson, 1979) and a 'Y-shaped'\* arrangement of

\* Note that the Y-shape criterion we specify is more specific than earlier (eg tridentate - Galli et al, 1980) criteria of other workers. The stricter definition explains the apparently anomalous low potency of piperidine-3-acetic acid (Galli et al, 1980) because it does not satisfy the optimum Y-shaped arrangement in either 'flat' conformation (Figure 5.3).

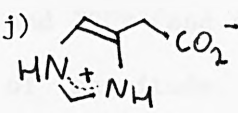
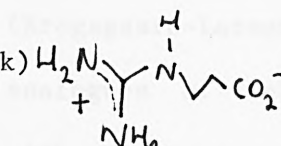


charge centres (Figure 5.2). The restricted rotational freedom in these molecules allows little flexibility in this arrangement (rotation about the carboxylate group in IGUV leaves the arrangement unaltered), suggesting that this is the optimum arrangement of charge centres for agonist interaction with the GABA<sub>A</sub> receptor. (The agonist iso-nipecotinic acid (INIP), with a rotatable carboxylate group, is slightly less potent than THIP and IGUV, implying that the position of the nitrogen atom is more important than that of the COO oxygens.)

For BIC in the active conformation ( $\theta_1=270^\circ-290^\circ$  - see 4.6 iii), matching the COO in GABA with the C-C=O in BIC, as suggested by Curtis and coworkers (1970), necessitates a relatively extended GABA molecule (Figure 5.1) similar to the match with THIP and IGUV. However, by reverting to our suggestion that the COO match is pharmacologically more significant (Steward et al, 1971, 1975), congruence with a more folded GABA molecule is found, with an approximately linear arrangement of charge centres (Figure 5.2). Furthermore the same arrangement is found in the GABA antagonists iso-THIP (Figures 5.2d and 5.4) and iso-THAZ (Table 5.1). The fact that none of these display any agonist activity is consistent with this being the antagonist arrangement of charge centres since they cannot adopt the suggested GABA agonist arrangement. (Note that the reversal of the N and O atoms in THIP and iso-THIP leads to the different arrangements of charge centres and thus the different roles.)

The GABA<sub>A</sub> agonist MUS used in previous structural comparisons can also adopt a Y-shaped arrangement of charge centres in a 'flat' conformation (Figure 5.2) which is slightly removed from the gas-phase minimum-energy conformation of MUS (Andrews and Johnston, 1979, Armstrong, Breckenridge and Suckling, 1982). However, MUS is more potent than THIP and IGUV,

Table 5.1. Potencies of key GABA analogues.

- a) GABA,  $IC_{50}=33nM$  (Krogsgaard-Larsen et al<sup>\*</sup>, 1983)
- b) THIP,  $IC_{50}=13nM$  (Krogsgaard-Larsen et al<sup>\*</sup>, 1983)
- c) IGUV,  $IC_{50}=37nM$  (Krogsgaard-Larsen et al<sup>\*</sup>, 1983)
- d) MUS,  $IC_{50}=6nM$  (Krogsgaard-Larsen et al<sup>\*</sup>, 1983)
- e) HBIC,  $IC_{50}=170nM$  (Kardos et al, 1983)  $IC_{50}=5000nM$  (Arnt and Krogsgaard-Larsen, 1979)
- f) 4,5-TAZA,  $IC_{50}=15000nM$  (Walters and Hopfinger<sup>\*</sup>, 1984)
- g) iso-THIP,  $IC_{50}=83000nM$  (Arnt and Krogsgaard-Larsen, 1979)
- h) iso-THAZ,  $IC_{50}=15000nM$  (Arnt and Krogsgaard-Larsen, 1979)
- i) [4,3]-THIP,  $IC_{50}=72000nM$  (Walters and Hopfinger<sup>\*</sup>, 1984)
- j)  imidazole acetic acid,  $IC_{50}=100nM$   
(Krogsgaard-Larsen et al<sup>\*</sup>, 1983)
- k)  guanidino propionic acid,  $IC_{50}=3600nM$   
(Krogsgaard-Larsen et al<sup>\*</sup>, 1983)

The structures for (a) to (i) are given in Figure 5.4. \*A collection of data.



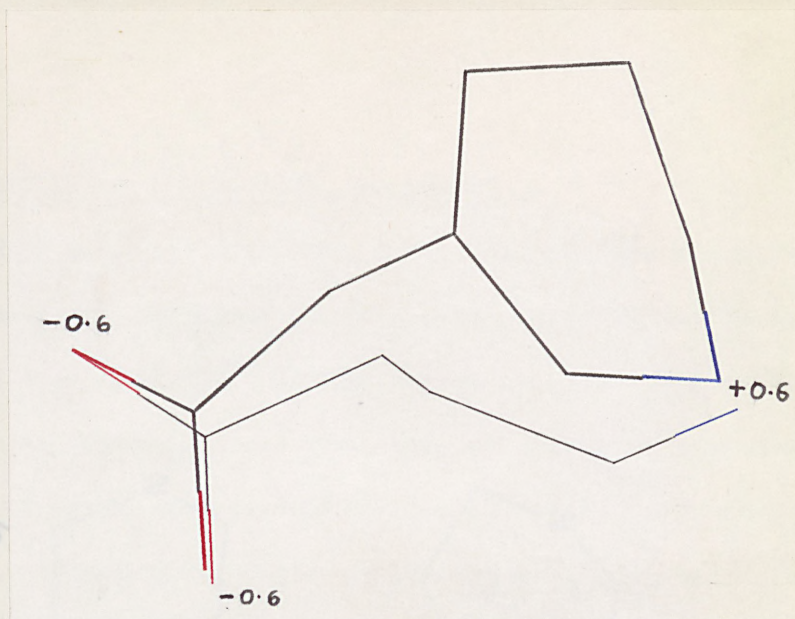


Figure 5.3. Matching of the optimum Y-shaped arrangement of charge centres with an analogue of low potency (piperidine-3-acetic acid).

which are locked into what seems to be the optimum arrangement of charge centres. A possible explanation of the lower potency of THIP and IGUV (and similar semi-rigid agonists) is the roughly two orders of magnitude loss of activity due to alkyl substitution at N (Krogsgaard-Larsen, Jakobson and Falch, 1983). Semi-rigid GABA analogues of substantially reduced potency are seen to be associated with an arrangement of charge centres significantly different from the optimum Y-shaped arrangement (see examples in Figure 5.3). This is a similar finding to the reduced potency found for GABA analogues with a charge separation ( $x_T$  - see 3.1) significantly different from the optimum (Steward and Clarke, 1975). It is clear, however, that the arrangement of charge centres is a more specific SAR parameter since GABA agonists and antagonists are not clearly differentiated using  $x_T$ .

These new results and the finding of Y-shaped and linear arrangements of charge centres offer a structural basis for the distinction between GABA<sub>A</sub> agonists and antagonists. And the GABA conformation that matches BIC may well be inactive, or even antagonist! (Compare this GABA conformation with eg iso-MUS and iso-THIP (Figure 5.2).)

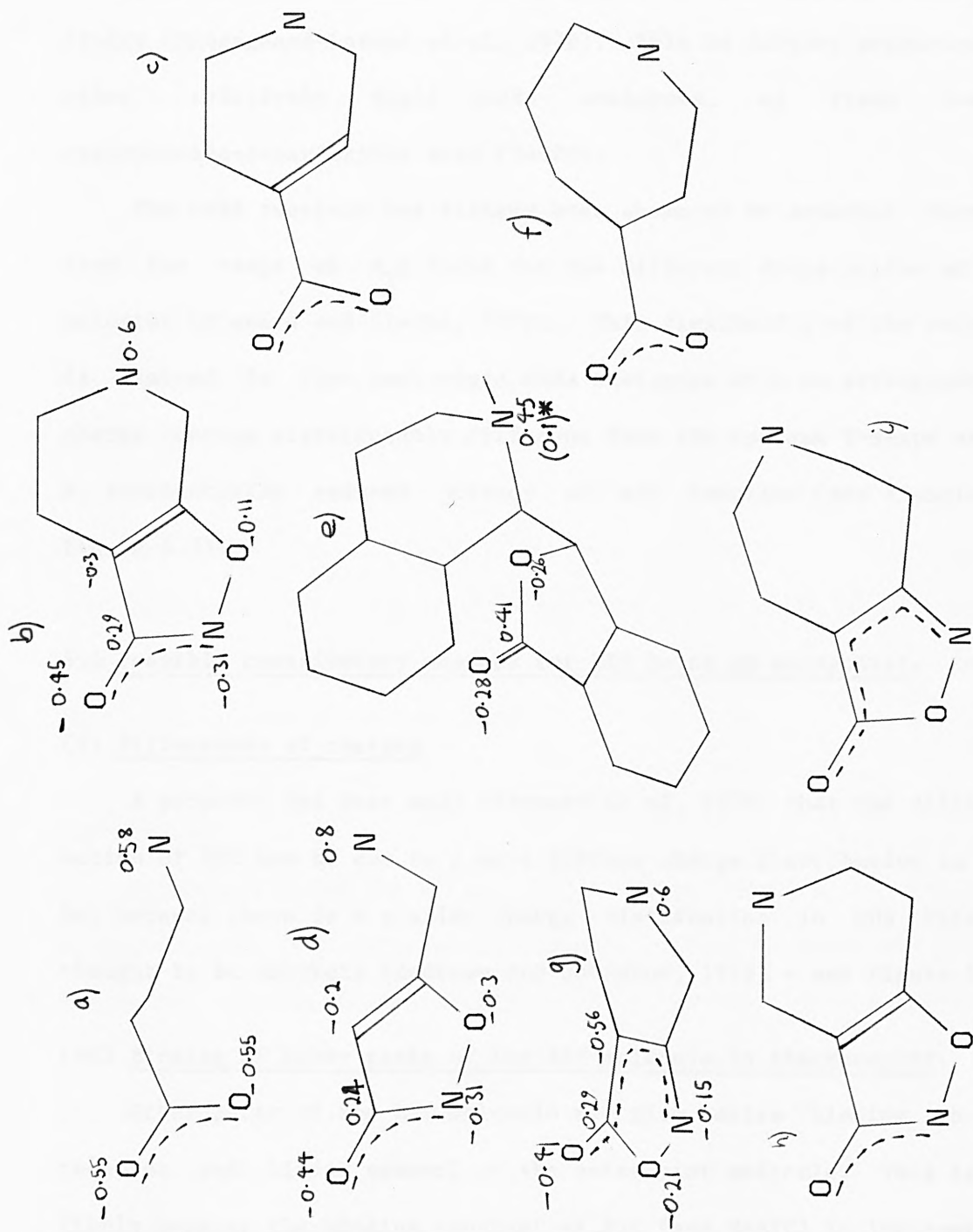


Figure 5.4. Partial atomic charges for key GABA analogues (see Table 5.1 for names and potency).  
 The charge shown on the positive N atom includes the charges on adjacent atoms. \*For BIC.



#### 5.4 Flexibility of the GABA<sub>A</sub> receptor.

THIP is so rigid that it cannot undergo any significant conformation change of the traditional 'bound conformation' to 'active conformation' type implying that any change is with the receptor. The flexibility of GABA is not therefore an essential feature for GABA<sub>A</sub> activity (Krogsgaard-Larsen et al, 1978). This is further supported by other relatively rigid GABA analogues, eg trans 3-amino cyclopentane-1-carboxylic acid (3ACPC).

The GABA receptor has already been shown to be somewhat flexible from the range of  $x_T$ s found for the different drugs active at this receptor (Steward and Clarke, 1975). This flexibility of the receptor is limited in that semi-rigid GABA analogues with an arrangement of charge centres significantly different from the optimum Y-shape are of a substantially reduced potency or are inactive (see examples in Figure 5.3).

#### 5.5 Possible contributory reasons for BIC being an antagonist.

##### (i) Diffuseness of charges.

A proposal has been made (Steward et al, 1975) that the different action of BIC may be due to a more diffuse charge distribution in BIC, but because there is a similar charge distribution in MUS this is thought to be unlikely (Andrews and Johnston, 1979) - see Figure 5.4.

##### (ii) Binding of other parts of the BIC molecule to the receptor.

Other parts of the BIC molecule may give extra binding to the receptor and hinder removal of the antagonist molecule. This is unlikely because the binding constant of BIC (and MeBIC) is low compared with that of GABA and GABA agonists (see Table 5.1). The binding constant would be expected to increase if extra binding were involved.



(iii) The steric bulk of the BIC molecule.

The bulk of the BIC molecule may block the required postulated conformational change in the receptor needed to elicit the agonist response (Steward et al, 1975). The stereospecificity of PIQs (Enna et al, 1977, Kardos et al, 1984) supports this possibility. This would help to explain why iso-THIP and iso-THAZ are only weak antagonists: they are much smaller molecules than BIC and would not block the receptor as effectively as BIC can.

Note also that for PIQ-based GABA antagonists the effects of N-methylation on potency are small (Kardos et 1984) compared with the large lowering of potency observed (Krogsgaard-Larsen et al, 1983) for N-methylation of GABA agonists. This can be explained by a more crowded nitrogen being associated with antagonist activity. Support for this explanation comes from the fact that MeBIC is more potent *in vivo* than BIC (Johnston, 1972) - due to a more effective blocking of the receptor. MeBIC is less potent than BIC *in vitro* (Kardos et al, 1984), which is probably due to a slightly worse fit at the receptor (MeBIC being more bulky than HBIC).

## 5.6 Summary.

Superposition of the  $N^+$  and  $COO^-$  charge centres of the GABA molecule with the corresponding charge centres in MeBIC and HBIC in the active conformation ( $\theta = 270^\circ - 290^\circ$ ) reveals an approximately linear arrangement of charge centres. The same arrangement is also found in the weak GABA<sub>A</sub> antagonists iso-THIP and iso-THAZ. A similar comparison of GABA with potent semi-rigid GABA<sub>A</sub> agonists yields a Y-shaped arrangement of charge centres. This clear difference may offer a structural basis for the distinction between GABA<sub>A</sub> agonist and

antagonist action, and shows that the arrangement of charge-centres is a more specific SAR parameter than the previously used charge separation,  $x_T$  (Steward and Clarke, 1976).

Another possible contributory factor to the antagonism of BIC is the bulk of the BIC molecule blocking any conformation change in the receptor, hindering the agonist response (Steward et al, 1975). This is supported by the lower activity of the smaller antagonists iso-THIP and iso-THAZ.

From the different effects that N-methylation has on GABA antagonist and agonist action, and from the different ratios of BIC/MeBIC in-vivo/in-vitro activity, we have deduced that steric hindrance of the N-region in particular is associated with GABA antagonism.

## 6 Overview of Part 1.

From a survey of the literature and from our gas-phase calculations on small test molecules the best currently available theoretical method for the calculation of relative molecular energies and charges, in terms of speed, accuracy (as compared with available experimental data) and versatility, is MNDO (2.2). The classical mechanics method, MM2, also performs very well, but is limited in that parameters must be available for the molecule of interest, and delocalised bonding can only approximately be accounted for (which makes MM2 unsuited to GABA). In our MNDO and MM2 calculations on isolated molecules containing single rotations, the average departure from experimental conformational-energies and internal-rotation barriers was ca  $3\text{kJ mol}^{-1}$  (Table 2.2), and this is comparable with the variation in experimentally determined energies for the molecules used in testing the methods (and with the value found in the literature ( $3\text{ kJ mol}^{-1}$  - Dewar and Ford, 1977) for MNDO-calculated barriers). This justifies the use of theoretical methods where experimental data cannot be obtained. A  $3\text{kJmol}^{-1}$  error in conformational energies is well within the ca  $25\text{kJmol}^{-1}$  limit on the expected conformational change involved when a drug leaves the biophase and binds to a receptor (Lambrecht and Mutschler, 1974, Clarke, 1976).

However, care is needed in using theoretical methods because even the best of these can sometimes give results which are totally inaccurate, such as giving the wrong conformer as the most stable one (eg Figure 2.1, and see Radom et al, 1985), and is best dealt with by using more than one theoretical method and, whenever possible, reference to experimental data for similar molecules.

Of the theoretical methods for determining solution conformation the continuum 'SOLVEFF' model shows the GABA molecule to be essen-

tially rigid in an extended conformation (3.3.2); this is in contrast to our experimental NMR results which show GABA to have no particular conformational preferences in solution. The SOLVEFF energy for polar molecules such as GABA is, however, dominated by an electrostatic interaction energy term which is roughly proportional to the square of the molecular dipole-moment (3.3.2 i). For GABA in extended conformations the N-H and C-O bond dipoles combine to give a high net dipole, but in folded GABA conformations the dipoles are aligned roughly opposite to one another giving a falsely low net dipole. A single dipole moment is therefore really inadequate in describing the polarity of a molecule such as GABA. The deduction of the sole existence of extended conformations from experimental dipole moment measurements (Edward et al, 1973) can be explained in the same way (3.3.2).

The discrete 'Supermolecule' model of Pullman and Pullman (1975) showed GABA (Pullman and Berthod, 1975) to be flexible with several low-energy minima and low barriers between them (3.3.1). The model is based on a solvent shell of added water molecules, these having been placed in the optimum positions determined by minimal STO-3G basis-set *ab initio* calculations on GABA fragments (alkyl ammonium salts (Port and Pullman, 1973) and the formate ion (Port and Pullman, 1974)). However, the orientations and points of attachment of the water molecules were not allowed to change with changes in the conformation of the molecule they are attached to. The water molecule orientations should really be allowed to change with solute molecule conformation (Beveridge and Schnuelle, 1974), though this is rather impracticable due to the enormous number of variables involved. Pullman and Berthod (1975) applied the model to GABA (using PCILO) and obtained

6

conformational-energy surfaces that for the above reasons can only be regarded as, at best, semi-quantitative (3.3.1). Their results do, however, suggest that GABA is flexible (with multiple low-energy minima and low barriers between them) - as now established experimentally in our work reported here. In addition, our Supermolecule calculations on GABA (using MNDO) showed a slight increase in GABA flexibility when partial relaxation of the water molecule orientations was considered (3.3.1).

One would expect a hybrid of the Supermolecule and SOLVEFF models to be more realistic, but in practice only a compounding of the errors of each method is observed (3.3.3).

The available theoretical methods for determining the solution conformations of polar molecules in polar solvents are therefore not very reliable.

To gain insight into the structural requirements for activity at a receptor where the endogenous 'drug' molecule is flexible, potent semi-rigid drug analogues are commonly used. For GABA, many analogues have already been studied, some of which are restricted to essentially only one conformation (eg THIP) or rotation (eg MUS). The ring structures and lower polarity of these analogues allow little possibility of strong solvent interactions affecting conformation, such as with GABA (Chapter 3), and gas-phase structures will be closely representative of those in solution. MNDO is known to give reliable molecular geometries (Dewar and Ford, 1977 - and see Table 2.3) and was therefore used with some confidence to calculate the structures of such analogues for comparison with the endogenous (GABA) molecule concerned.



We have established experimentally, using variable-temperature NMR, that GABA is indeed considerably flexible in solution. This is contrary to the SOLVEFF (3.3.2) and dipole moment results (3.2.2) cited above. Other workers (Ham, 1974, Tanaka et al, 1978) had found by NMR that multiple GABA conformations are present in solution, and had attempted to find the populations of the minima, but had to make assumptions on the angles of the minima and the coupling constants associated with the angles (3.2.1).

Knowing that GABA is flexible permitted us to compare its conformational range with the structure of semi-rigid GABA agonists and antagonists. From the comparisons separate structural requirements for GABA agonists and antagonists were derived.

In the earliest comparisons of the molecular structure of GABA with BIC (and MeBIC) and GABA analogues (Curtis et al, 1970) the assumption was indeed made that GABA is flexible, but with no distinction made between structural requirements for agonist and antagonist drug molecules. When Curtis and coworkers originally compared Dreiding models of BIC and GABA (Curtis et al, 1970) they found an isosteric match of the  $N^+$  and  $COO^-$  of a relatively extended GABA molecule, with the  $N^+$  and lactone  $C-C=O$  of BIC (Figure 5.1). The possibility of exact congruence of the nitrogen and  $COO$  charge centres in both molecules was later suggested (Steward et al, 1971) (Figure 5.2a), but was ruled out (Beart et al, 1971) because it would involve a GABA molecule too folded for congruence with semi-rigid GABA agonists. Again these comparisons did not take into account the fact that an agonist is being compared with antagonists.

In those early structural comparisons the  $H-C1-C9-H$  dihedral angle ( $\theta_1$  - Figure 4.1) in BIC and the roughly equipotent BIC

methohalides (MeBIC) had not yet been determined. From NMR work on protonated BIC (HBIC - the active form of BIC - see below) and MeBIC, Andrews and Johnston (1973) found possible  $\theta_1$  ranges of  $70^\circ - 110^\circ$  and  $250^\circ - 290^\circ$ . Subsequently, on the basis of their PCILO calculations (Andrews and Johnston, 1973), they chose the  $250^\circ - 290^\circ$   $\theta_1$  conformation range and demonstrated (Andrews and Johnston, 1979) that for BIC in the mid-point of this range there is the isosteric match previously proposed by Curtis and coworkers (1970) as a basis for the activity of BIC at the GABA receptor site (Figure 5.1). From their comparisons of the structures of GABA, BIC and the agonist MUS, Andrews and Johnston (1979) concluded that the partially folded GABA conformation based on BIC and the more extended GABA agonist conformation based on MUS 'define the limits of the range of "active-conformations" at BIC-sensitive receptors' (Andrews and Johnston, 1979).

The MNR methods we have used here exploited the conformation-dependence of the key H6' and H8 chemical shifts, the H1/H9 coupling constant, and ion-pairing effects on MeBIC, and showed that indeed only one conformation exists for both MeBIC (4.3) and HBIC (4.4) at  $\theta_1 = 270^\circ - 290^\circ$  (Pooler and Steward, 1986b). Furthermore, we have shown that for BIC three conformations can be found, at  $\theta_1 \approx 45^\circ$ ,  $170^\circ$  and  $270^\circ$  and have estimated their conformational energy differences to be  $1 - 5 \text{ kJmol}^{-1}$  (4.2, Pooler and Steward, 1986a). (Previous workers (Andrews and Johnston, 1973, Elango et al, 1982) only found one conformation.) We have also estimated a lower bound for the energy of the next low-energy conformation of MeBIC, from a knowledge of the accuracy that the H1/H9 coupling constant could be measured, as  $11 \text{ kJmol}^{-1}$  (4.3.2). Although this is less than the BIC (and MeBIC) binding energy of  $30-40 \text{ kJmol}^{-1}$ , the fact that no other PIQ GABA an-

6

tagonists with a different conformational minimum and of greater or equal potency to BIC have been found reinforces the argument for a  $\theta$  of  $270^\circ - 290^\circ$  being the active conformation of MeBIC. An active conformation of  $\theta \approx 270^\circ - 290^\circ$  gives support for BIC being in the protonated form (rather than the free base) when binding to the receptor, because  $< c. 20\%$  of BIC is in the required conformation (4.2.1).

If the structures of GABA, MeBIC, HBIC, and GABA agonists are compared using the above N and C-C=O match (Curtis et al, 1970) no particular distinction is made between agonist and antagonist structural requirements for activity. However, if one reverts to the suggested (Steward et al, 1971, Galli et al, 1980) match of N and COO charge centres in both molecules **separate** agonist and antagonist structural requirements can now be identified (5.3).

By matching the charge centres in GABA with those in potent semi-rigid GABA agonists we found that a 'Y-shaped' arrangement of charge centres is associated with agonist activity (Figure 5.2), and that GABA analogues with an arrangement of charge centres significantly different from the optimum Y-shape are of considerably reduced potency (eg Figure 5.3). This criterion is stricter than those from earlier models (Galli et al, 1980, Walters and Hopfinger, 1984) and explains the apparently low potency of certain analogues (Galli et al, 1980) which can now be seen not to satisfy our proposed Y-shape requirement (5.3).

Matching the charge centres of GABA with those in HBIC and MeBIC in the active conformation gives an approximately linear arrangement of charge centres. The same arrangement is also found in the weak GABA antagonists iso-THIP and iso-THAZ (Figure 5.4). The lack of agonist activity in these drugs is consistent with this being the an-

tagonist arrangement of charge centres since they cannot adopt the suggested agonist arrangement.

This clear difference between agonist and antagonist arrangements of charge centres now offers a structural basis for distinction between GABA<sub>A</sub> agonists and antagonists.

A possible contributory factor to the antagonism of BIC is the bulk of the BIC molecule blocking any conformation change in the receptor, hindering the agonist response (Steward et al, 1975). This is supported by the lower activity of the smaller antagonists iso-THIP and iso-THAZ. From the different effects that N-methylation has on GABA antagonist and agonist action, and from the different ratios of BIC/MeBIC in-vivo/in-vitro activity, we have deduced that steric hindrance of the N-region in particular is associated with GABA antagonism.

## Part 2: Receptors of known molecular structure

### 7 Development and applications of IMDAC (Interactive Molecular Display And Calculation).

#### 7.1 Introduction.

On setting out to explore the possibilities of using computer graphics in examining enzyme/substrate interactions we found that certain desirable software features were not commercially available. We have therefore developed IMDAC<sup>\*</sup> - a molecular modelling system with special features for cleft searching (CLEFT), examining interactions between a docked drug and a receptor (Close Contacts (CC) - special aspects), and an internal reference frame for coordinate transformations.

A great profusion of molecular graphics packages are available for the visualisation of the three-dimensional (3D) structure of even quite large (protein) molecules. (Some examples are ChemGraf, Sybil, Frodo, MIDAS, Mogli, Insight and Gramps - fairly comprehensive lists of available graphics software can be found in Appendix A5 and Morffew (1984).) In this Chapter we present features of IMDAC which have been specifically designed to gain the maximum amount of structural information from known receptor coordinates, within the limitations of the graphics hardware currently available at the City University (see 7.2.8 for hardware capabilities). We have aimed at developing completely new graphics features in IMDAC, and not just to reproduce those found in other systems. Some of the features of IMDAC cannot

---

\* Initially the basic graphics package MOLEC5 (Islam, 1984) was obtained and substantially modified, and many new and more powerful features added. Details of the capabilities of the original software and of the basic modifications required can be found in 7.2.1 and A7.



therefore be found in any of the above systems - at least not in the form found in IMDAC.

If the molecular structure of a receptor with a bound drug is known, then the residues constituting the 'active site' (binding site) of the receptor are readily identified from the position of the bound drug. Otherwise the active site can be defined by identifying the residues and groups of the receptor which are thought to interact with known drugs, and finding the position for these drugs which gives the best interaction between the drug and these groups and the rest of the active-site. Due to the lack of acceptable available software at the time (mid 1984), we developed the routines (7.2.2): CLEFT, for finding possible active site clefts within a specified volume of the receptor; SURF for examining the surface of a cleft/active site (not quite working correctly yet) and CC for determining the specific amount of space (lack of Close Contacts) around a drug molecule positioned within the active site. All the above features, most of which were designed to overcome the difficulties of visualising 3D using a 2D static display, come under the visualisation heading (7.2.2), under which stereo and space-filling models are also included.

We have also developed the routines DOCK and CONF (6.2.3) for placing a drug within an active site (with optimisation of the fit and/or intermolecular energy), and EN and MINIM (originally from MOLEC5 - see 7.2.6 and A7) for calculating and minimising the interaction energy between drug and receptor. Plus routines for molecular editing (EditMol - 7.2.4), and for interface to MNDO and other programmes (eg MNDOIN and MM2IN - see 7.2.7). An additional special feature of IMDAC is a frame of reference for all rotations and translations. This is described in detail in section 7.2.5.

A full menu, covering all IMDAC options, is described briefly in Figure 7.1. Section 7.2 of this Chapter is a detailed description of the main features of IMDAC and section 7.3 covers applications of IMDAC to the PA2 and TLN systems.

We have applied IMDAC to two different kinds of problem: (i) the phospholipase-A2 (PA2) system - an enzyme for which only isolated receptor coordinates have been determined, and (ii) thermolysin (TLN) - an enzyme for which molecular structures of several drug/receptor crystal complexes are known. The PA2 section covers the fitting of a substrate into the PA2 active-site, and is only very brief because PA2 was used mainly for development work. We have used TLN as a model for encephalase (ENK), a similar enzyme to TLN but of unknown molecular structure. By starting with coordinates for known TLN inhibitors we have modelled novel analogues which display significant ENK activity (Palfreyman, 1986). In addition we have found a novel possible mode of binding for a new (Palfreyman, 1985) ENK inhibitor. However, the above TLN work was carried out in collaboration with a pharmaceutical company and, since it is of a confidential nature, we have been requested not to report details of those findings. We can, however, still show the basic principles of applying IMDAC to a known inhibitor and examining the possibilities of modifying the inhibitor for better interaction with the receptor (in this case TLN).

Figure 7.1 Brief explanation of the OPTS menu.

- 1/2 MOL1 (receptor or drug)/MOL2 (drug) on/off  
(Right hand number = model number.)
- 3/4/5 View x-y,y-z,z-x planes
- 6/7/8 Rotate about x/y/z axis
  - 9 Change highlighting (for residues and/or atoms)
  - 10 Translate molecule(s): 1=backwards, 2=down, 3=left, 4=right,  
5=up, 6=forwards, 7=point 1 - point 2, 8=atom 1 - atom 2
  - 11 Draw on/off
  - 12 Space-filling model
  - 13 Zoom
  - 14 Produce file for drawing on quality plotter (Benson despooler)
- 15/16 Torsion angle rotate/calculate
- 17 Save data (coordinates or rotation/translation matrices)
- 18 Window z-axis
- 19 Change energy parameters
- 20/21 Calculate distance/angle
  - 22 CC - calculate and display Close intermolecular Contacts for  
selected MOL2 atoms (Figure 7.4). A CC is drawn when the  
interatomic distance is less than the sum of the Van der Waal  
radii plus a sensitivity coefficient (SENS). There is also an  
option for calculating and displaying interaction energies.
  - 23 Identify and label atoms
  - 24 Change model: 1,2,3=white, black, red; 4,5=colour code; 6=depth-  
cue; 7,8=depth-cue + colour code; 9,10=highlighted atoms only.
  - 25 Calculate energy (6-12, electrostatic, H-bond). More accurate  
energy calculations can be performed by transferring coordinates  
to the ULCC Cray or Amdahl computer.
  - 26 Write-out individual atom potentials to a data file.
  - 27 Draw stereo-perspective view.

(continued on next page)

Figure 7.1 (continued)

- 28 Enter molecular editing menu: make/break bond, add/delete atom(s), delete fragment, split/join molecules, invert chiral centre.
- 29 Redefine colours.
- 30 Define active site for a drug/receptor as all residues within  $r_A$  of a docked drug.
- 31/32 Prepare data file suitable for direct input to MM2/MNDO programme.
- 33 Search for clefts within the enzyme using a spherical probe. Draw all spheres which contain no MOL1 atoms for the chosen region of space.
- 34 Display colour-coded dot surface for a receptor.
- 35 Return to the READAT menu for replacement of both molecules or just MOL2. The absolute coordinates of MOL1 are invariant to all transformations (rotation, translation and zooming) and all transformations are stored, enabling the direct replacement of one molecule with another in the same coordinate system.
- 36 Draw Van der Waal radii (atom colour coded).
- 37 Erase whole screen or just MOL2.
- 38 No-erase on/off
- 39 Draw specific residues in a preselected colour.
- 40 Minimise intermolecular energy (with fixed geometry) - for MOL2 docked into MOL1. (Internal geometry optimisation can be optimised by transferring coordinates to the Cray.)
- 41 Draw stereo pair (without perspective).
- 42 Molecular superposition - used for docking drug molecules by adding dummy atoms to the MOL1 atoms thought to be interacting directly with MOL2. Torsion angle optimisation can be included.
- 43 Calculate charge separation ( $x_T$  - see Chapter 3).
- 44 Rock - a rough impression of 3D is obtained by rapidly switching alternate superimposed stereo-images.
- 45 Temporary return to the operating system (Primos).





## 7.2 The molecular modelling system - description of the main features of IMDAC.

The full menu, covering all IMDAC options, is described briefly in Figure 7.1. Details of the more important features are included in this section. For clarity, option numbers for the main operational menu, OPTS, are prefixed 'OP', molecular editing options 'ED', translation options 'TR', and model numbers 'MOD'. (Eg the Close Contacts (CC) option in OPTS is OP22.) A simplified diagram of the layout of the main routines in IMDAC can be found in Figure 7.2.

### 7.2.1 The basic system (and essential changes required).

Initially the source code for a basic graphics package, MOLEC5 (Islam, 1984), was obtained (Islam, 1984) and converted from DEC graphics to GINO-F graphics for the Prime 550 computer at the City University. (This was the best available programme at the time which fitted in with existing graphics facilities at City.) The programme contained routines for: molecular rotation and translation; some limited molecular editing (bond/atom addition/deletion); a simple grey-shade depth-cue; energy calculations - including inter-molecular energy minimisation (all for small molecules only - up to 235 atoms); distance, angle and simple close contact calculations; plus various redundant routines and several bugs.

Substantial modification was required for use with proteins, with many new features added to change MOLEC5 into the much more extensive graphics system IMDAC. The primary modification consisted of a massive speeding up of the bonding and energy calculation routines (see A7 for details). The features described in this Chapter are our own subsequent additions to the basic system (unless otherwise stated).

IMDAC is built around two main menus - in subroutines READAT and OPTS - with smaller menus within OPTS (eg for changing models or molecular editing). READAT is used for reading-in molecular data (coordinates, charges etc) and can be easily returned-to from the operational menu, OPTS, for input of new data (eg replacing one of the molecules).

One restriction of IMDAC is that the second molecule read-in should be only quite small (up to 100 atoms - though more atoms can be input, but most options will then not work). This restriction actually makes IMDAC easier to use because the receptor molecule can only be MOL1, with the drug as MOL2. This restriction fits in with the frame of reference for molecular rotation and translation, and the fixing of the absolute coordinates of MOL1 (see 7.2.5). (If necessary, similar proteins may be compared by reading them in separately and then displaying them on the screen together.)

## 7.2.2 Visualising the active site and other key features of a receptor molecule.

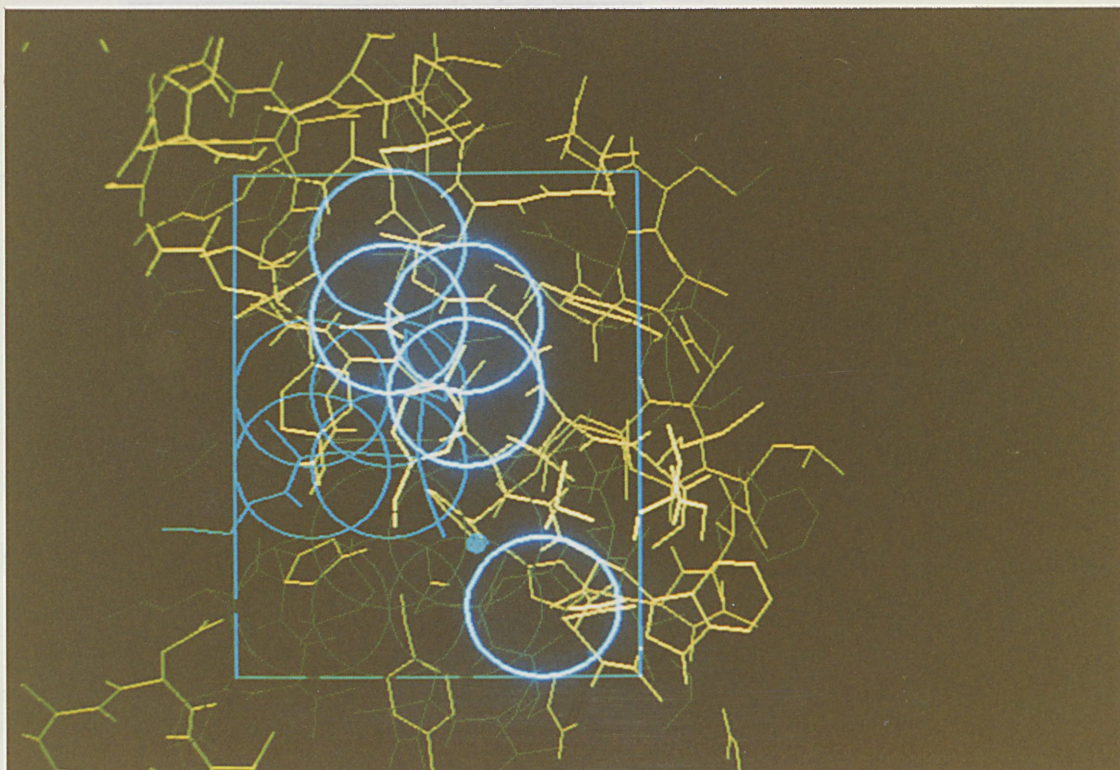
### 7.2.2.1 CLEFT (OP33).

It is important that any holes within the receptor molecule which are large enough to contain the drug (and could therefore be potential active sites) can be readily visualised. For this purpose we designed the programme CLEFT, which makes use of a spherical probe (of any desired radius) to find volumes of space inside the receptor molecule volumes of space within which there are no atoms present. (Waters of hydration can be included or excluded.) All such spheres are then displayed, with depth cueing, for a chosen volume of the receptor molecule (Appendix A6.5), with any large clefts clearly visible as a heavy concentration of spheres.

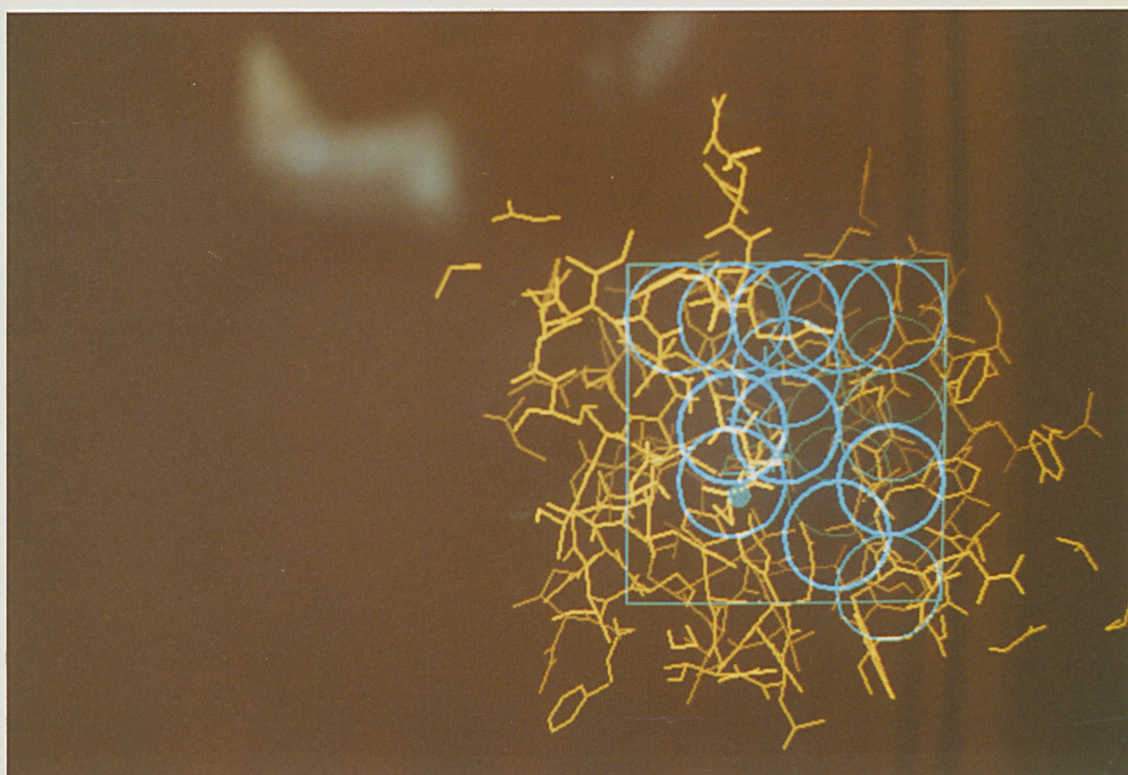
By carefully selecting different values for the radius of the spherical probe (RSPH - usually between 2.5 and 3.5 $\text{\AA}$ ), search increment (RINC) and depth of space searched, the shape of the cleft can be clearly visualised (Figure 7.3a). A very poor choice of RSPH and the wrong view will not give a clear picture (Figure 7.3b).



Figure 7.3. Examples of the use of the CLEFT option (OP33).



a) The shape of the TLN active-site is shown by the depth-cued circles (RSPH=RINC=2.5Å (default)).



b) This view shows that some experimentation is needed with CLEFT to obtain a clear picture of the cleft.



#### 7.2.2.2 SURF - a surface display routine (OP34).

The routine SURF was designed to rapidly calculate and display selected portions of the surface of a receptor as a series of dots, which are colour-coded for hydrophobic (red) and hydrophilic (blue) regions. The data represent the visible surface of the Van der Waals spheres of the atoms on the surface of the protein, and should make it easier to see actual clefts within the receptor. The advantage of this routine over similar surface representation routines (Conolly, 1983, 1985) would be that small portions of the protein surface could be viewed rapidly for different views of the receptor.

However, this is a recent addition (June, 1986) to IMDAC and is not that well tried and tested yet. Unfortunately, since the programme still contains a few bugs we have not yet been able to obtain a reasonable picture of a protein surface.

#### 7.2.2.3 Close Contacts (CC - OP22).

With this option all close contacts (CC's) between selected drug (MOL2) atoms and the receptor (MOL1) can be calculated and displayed, subject to the following condition:

$$\text{CC if } R_{i,j} < \text{RVDW}_i + \text{RVDW}_j + \text{SENS.}$$

Where SENS is a term added to the sum of Van de Waal radii for tuning the calculation and display of CC's. (SENS should be set low initially (ca -1) and then increased to find more CC's.) An energy term (6-12 plus electrostatic) can also be calculated for each CC and colour coded for favourable/disfavourable interactions. For crudely docked drugs, calculating energies in this way is much faster than using the energy routine EN (7.2.6), because the calculation is limited to just the poorly positioned atoms. The molecule may then be



manipulated manually to a better position and the calculation rapidly repeated. (Note that the energy calculated in CC is for very close intermolecular interactions only and is only meant to be used as a rough guide.)

CC is a very powerful tool for exploring the space around a docked drug for possibilities of modifying the drug for better interaction with the receptor. CC was used extensively in our work on modelling ENK inhibitors and an example of the use of CC is given in Figure 7.4 (see also 7.3.2).

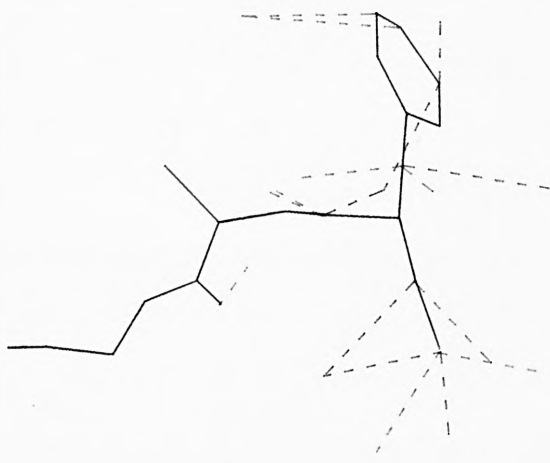


Figure 7.4. Example of the use of CC on BAG (see 7.3.2).

The CC's shown can be used in making modifications to the drug by varying SENS for selected MOL2 atoms (see also Figure 7. ).

The atom and residue numbers of the close MOL1 atoms are printed on the screen (or data file) and can be drawn using DrawRes (OP39).

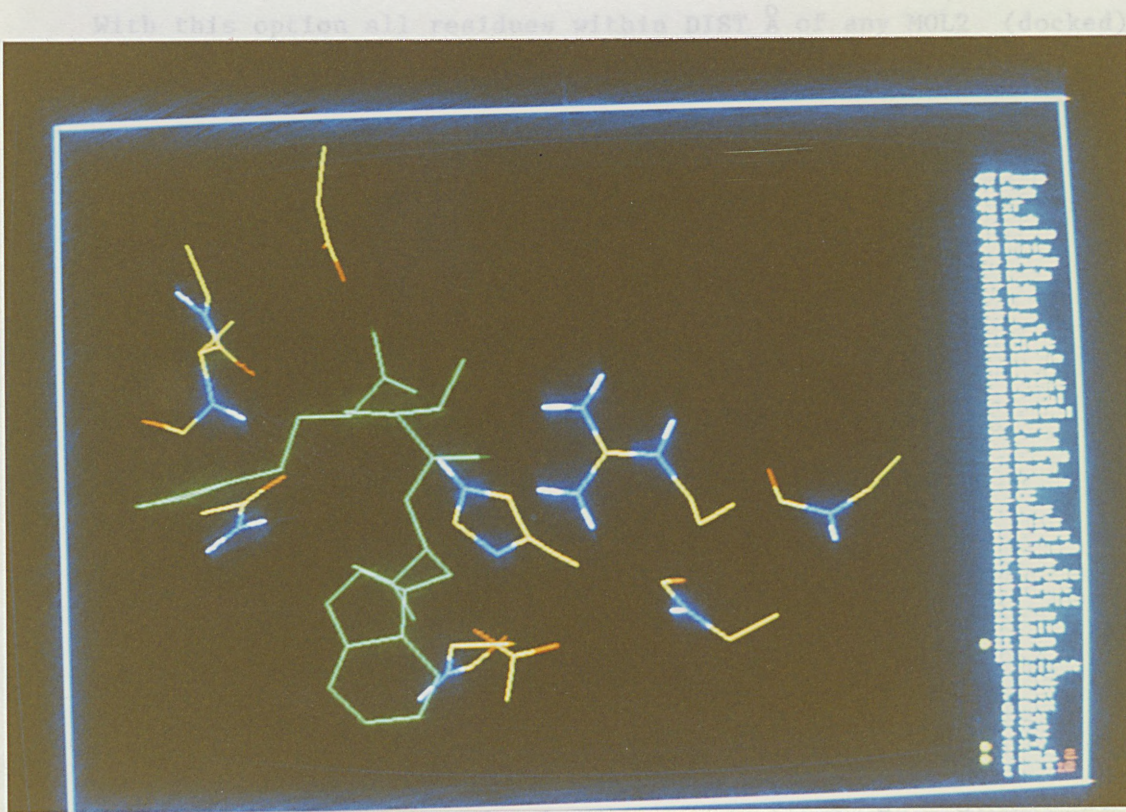
#### 7.2.2.4 Highlighting of key atoms and residues.

As the process of drawing a large molecule on a static raster display is extremely slow (eg up to 30 seconds for TLN), we devised two models (MOD9/10) such that only predefined atoms/residues ( $SPE_i = .TRUE.$ ) are drawn. The drawing of up to ca 8 residues with MOD9 (depth cued) or MOD10 (colour coded - Figure 7.5) requires less than one second (real time), and thereby enables the rapid selection of different views of the molecule. An additional advantage is that important features are now clearly visible, without the obscuring effect of the rest of the molecule. The other residues may then be drawn either individually using OP39, DrawRes, or for the whole molecule, with depth-cueing to give a 3D effect, while maintaining the highlighted residues (MOD7 and MOD8). The highlighting switch, SPE, can be readily changed for atoms or whole residues using the Hilight option (OP9).

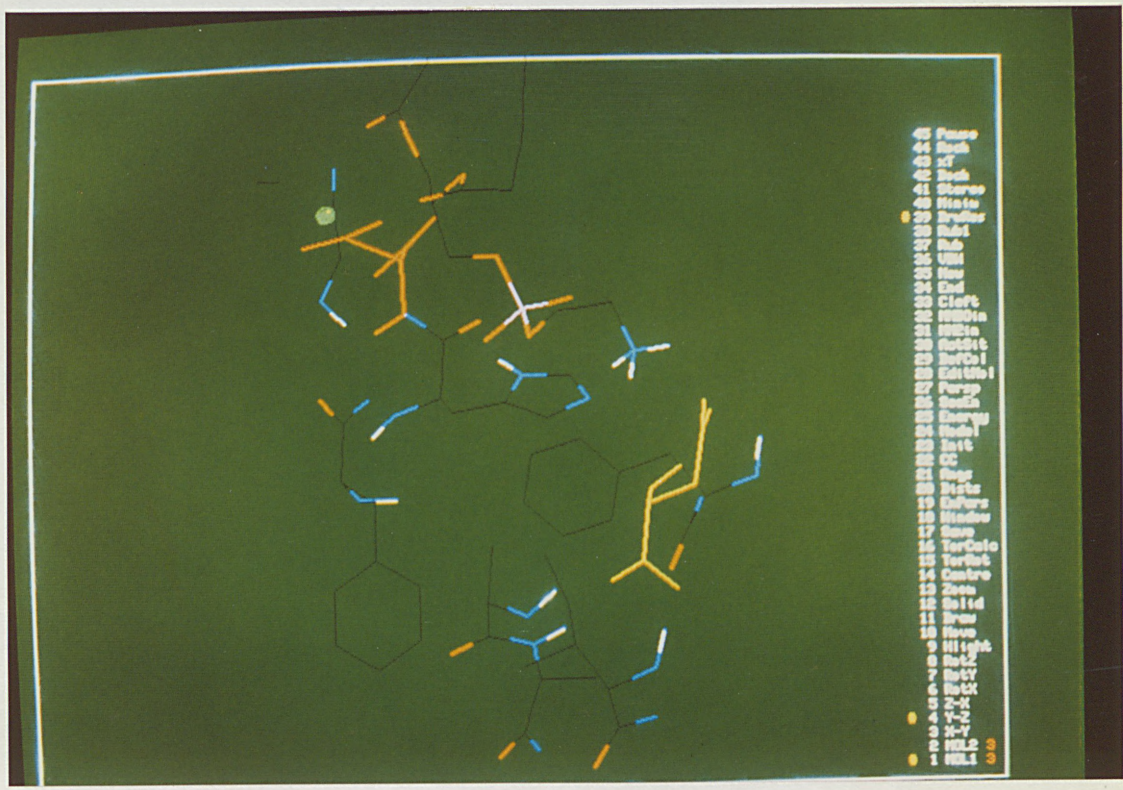
The DrawRes option is also useful when combined with CC (OP22 - see above) for pinpointing neighbouring residues. (Residues are drawn by number - either by the order in which the residue was read in (input positive number), or by the order in which the residue occurs in the original protein (input negative number) - useful for when dealing with active-site residues only.)



Figure 7.5. The drawing of only pre-selected residues.



a) TLN + bound inhibitor (CLT - see 7.3.2). simple approximation for the



b) PA2 + docked inhibitor (PETH - see 7.3.1).



#### 7.2.2.5 SITE - isolation of active-site residues (OP30).

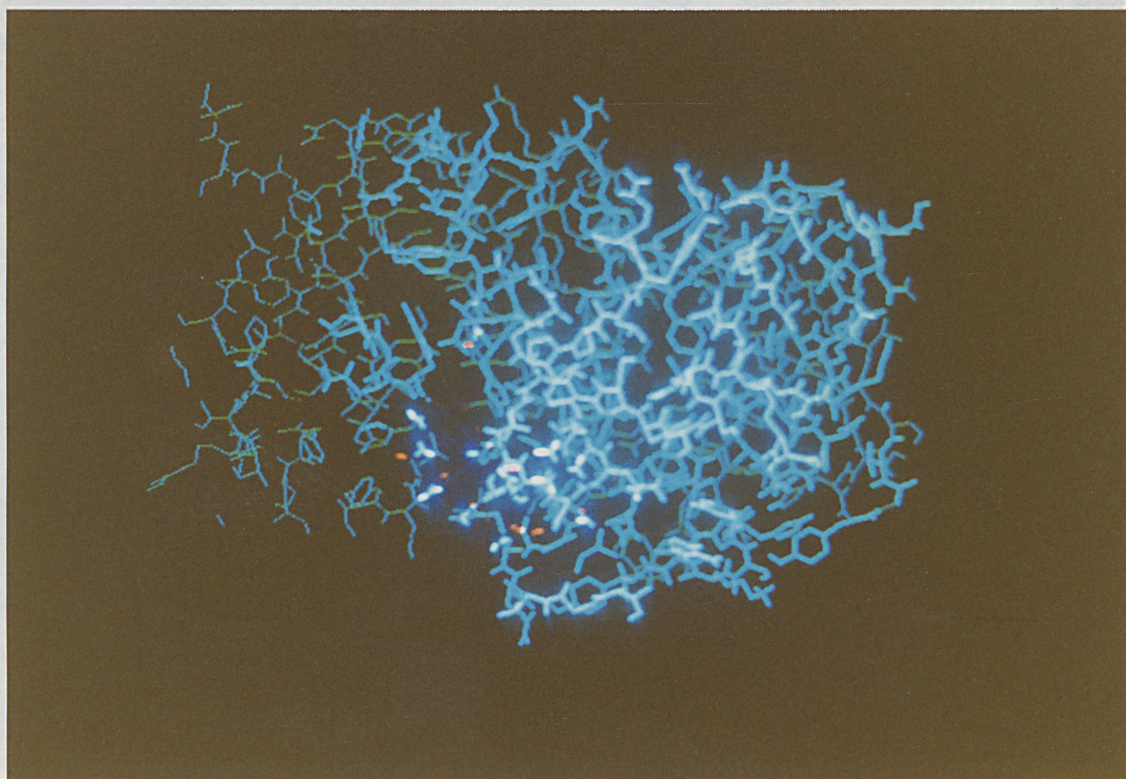
With this option all residues within DIST  $\overset{\circ}{\text{A}}$  of any MOL2 (docked) atom are labeled ( $\text{ACTSIT}_i = \text{.TRUE.}$ ) and can be treated separately. For a protein of ca 3000 atoms (eg TLN) a DIST of  $9\overset{\circ}{\text{A}}$  reduces the number of atoms to ca 800, which greatly speeds up the bonding, energy and CC calculations, and drawing, and gives a much clearer view of the active site (Figure 7.6).

#### 7.2.2.6 Pseudo space-filling model (OP12).

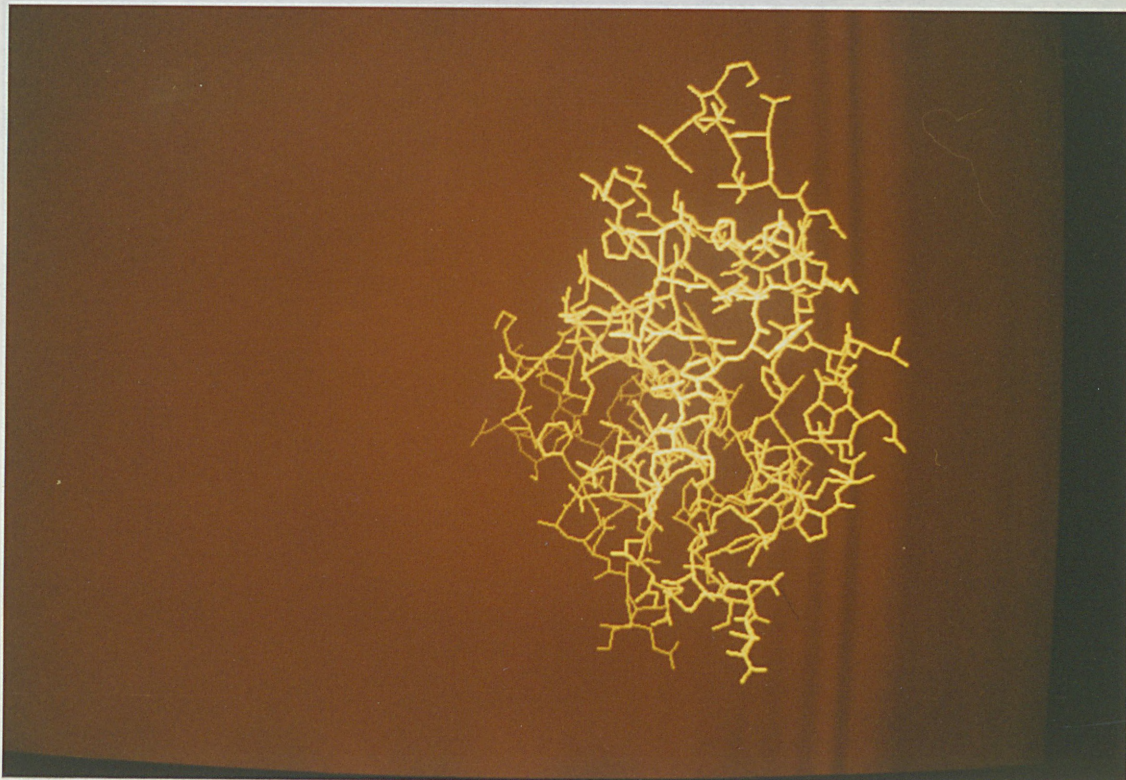
The limitation of being able to use only 15 colours/shades at once makes the production of realistic, depth-cued, space-filling models (with or without atom colour coding) very difficult. However, by carefully defining these colours (7 for spheres, 7 for depth cue), an impression of space-filling may be obtained (see eg Figure 7.7). Because of this limitation we only use a simple approximation for the atom-intersections of drawing atoms as a series of concentric circles, shaded to represent spheres (of Van de Waal radius), with the atoms furthest away being drawn first (Figure 7.7). When a circle is drawn on top of part of another sphere, the overwritten section is no longer visible.



Figure 7.6. Difference between active-site and whole receptor views.



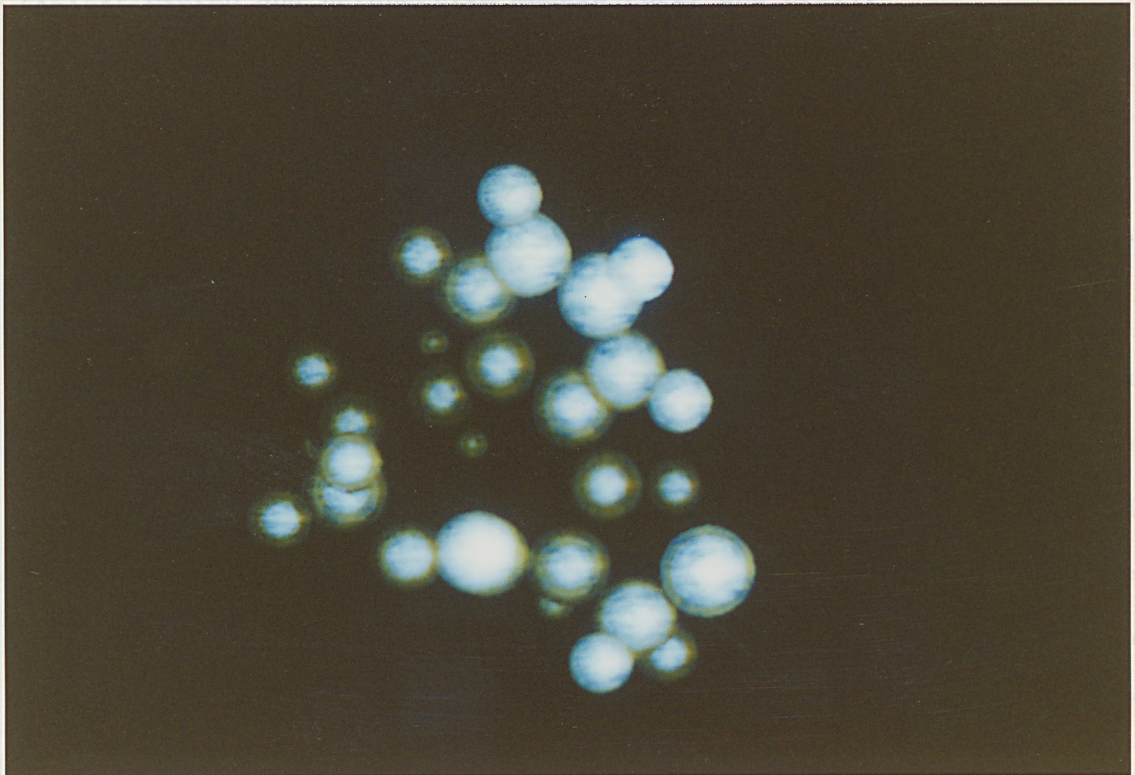
a) Whole TLN receptor (the colour-coded residues cannot be distinguished).



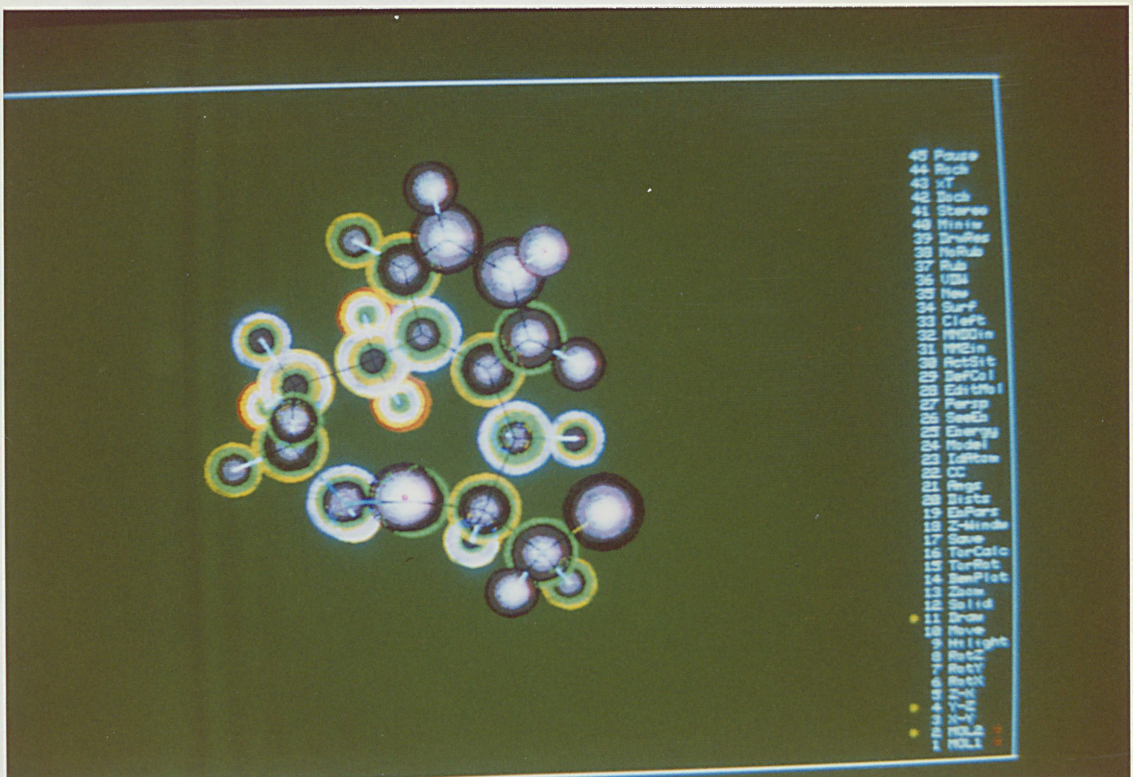
b) TLN active-region only (more detail will be seen by highlighting).



Figure 7.7. Example of space-filling model.



a) The spheres appear to be too small because the outer circles are too dark to be picked up by the camera.



b) This is the same picture but with the colours re-defined. An impression of space-filling is obtained even with this crude model.



#### 7.2.2.7 Stereo and perspective models (OP27 and OP41).

A stereo image can be displayed by drawing left and right eye views on the screen, separated such that the images may be superimposed, either 'by eye' or by using special spectacles. The default image separation may readily be changed for easier superposition of the two images.

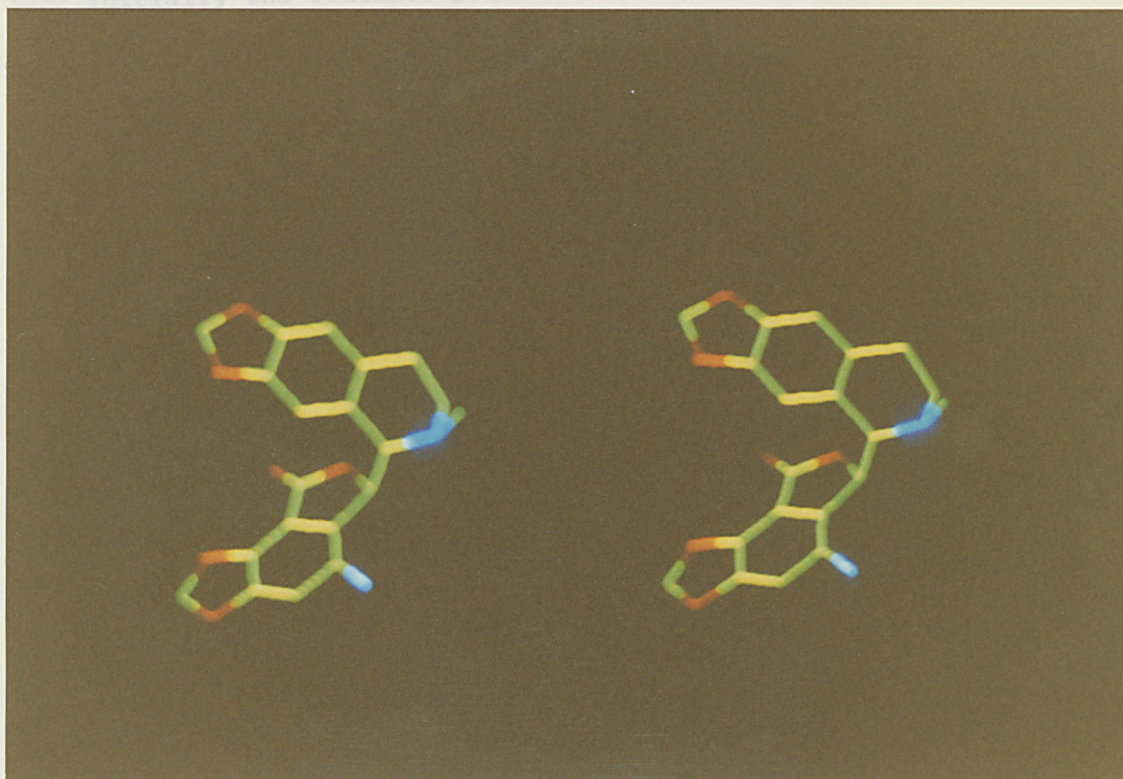
With OP41 the left and right stereo images are produced by a rotation (default  $4^{\circ}$ ) about the vertical axis (Figure 7.8). In producing a stereo image it is vitally important that the required rotation and translation are performed in the correct order (ie move left and draw; move right; rotate; draw; rotate back; move back to centre), otherwise the molecule will not be in the same place after drawing.

A stereo image drawn using OP27 is produced as two perspective views for an observer positioned ca 30cm away from the front of the graphics screen. We found little difference between the images produced with or without perspective. A much better representation of 3D can be obtained with dynamic parallax (Diamond et al, 1982), but alas is not possible without real-time translation and rotation. An alternative dynamic stereo view can be obtained using OP44 - Rock (see 7.2.8 i).

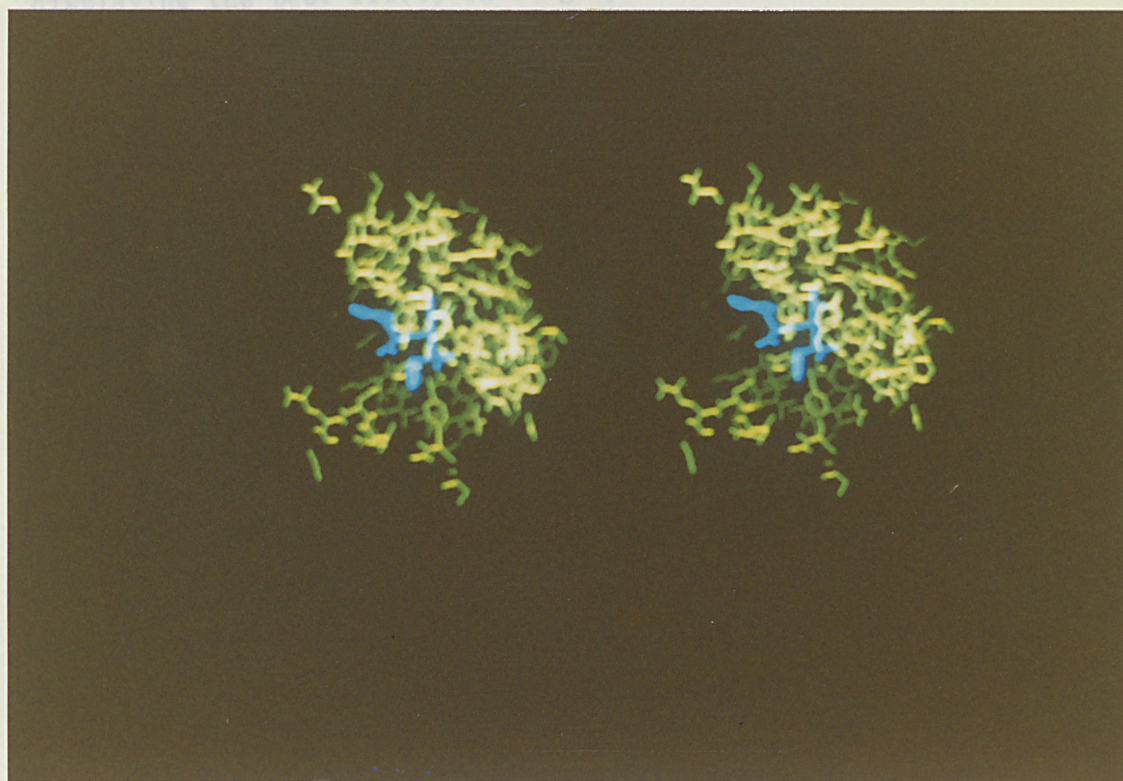
#### 7.2.3 Docking of drugs to receptors (OP42).

Although many routines exist for molecular superposition and docking (eg DOCKER (Busetta et al, 1983), ChemGraf (Davies, 1985)), because of its importance in making IMDAC complete we include our own version here.

Figure 7.8. Examples of the stereo model (OP41).



a) BIC,  $\theta_1=270^\circ$ . (The yellow colour is an artifact of the photograph.)



b) TLN active-site with a bound drug (CLT - see Figure 7.13).

\*  $FIT^2$  is compared with  $FITMAX^2$  since this avoids the calculation of square roots, which is slow.



Initially the routines DOCK and CONF were designed for the comparison of two drugs by superposition of atom pairs (DOCK), with an option for varying up to 3 conformations to obtain the best fit (CONF). This same process can also be used for placing new drugs into a receptor, either by superposition onto an already docked (bound) drug, or by superimposing the drug onto dummy atoms attached to key atoms within the active site of the receptor.

The measure of the fit (FIT) is given by the sum of the squares of the differences between atom pair separations:

$$\text{FIT}^2 = \sum_{j=1}^3 \left( \sum_{\substack{i=1 \\ k \neq j=1}}^3 (x_{j,i} - x_{k,i})_{\text{enz}}^2 - \sum_i (x_{j,i} - x_{k,i})_{\text{drug}}^2 \right)$$

We can calculate FIT very rapidly while varying up to 3 torsion angles, and reject any FIT greater than a preset maximum (FITMAX<sup>\*</sup>), displaying the best fit(s) on a graphics terminal (Appendix A6.5). The inter- and/or intra-molecular (for MOL2) energy can be calculated simultaneously and all high energy conformations rejected. The MOL2 intramolecular energy can be calculated quickly, and high-energy isolated-molecule conformations sifted out. (The low-energy conformations of MOL2 should then be checked using a more accurate method - eg MNDO or MM2.)

The calculation of intermolecular energies in CONF has not yet been finalised (since it requires a large amount of computer time). The principle is:

(i) Rotate angles to the first conformation to be examined and DOCK the drug (so that the intermolecular energy will be calculated correctly).

\* FIT<sup>2</sup> is compared with FITMAX<sup>2</sup> since this avoids the calculation of square roots, which is slow.

(ii) Continue rotations in CONF and rapidly reject all conformations of poor fit or very high energy.

(iii) Store all conformations which are within the specified FIT and energy requirements, for inspection by the user.

The position of a drug docked using DOCK and CONF may then be refined by:

(i) Calculating CC's for selected MOL2 atoms and making small manual translations, rotations etc. CC has an option for very rapid calculation of part of the intermolecular energy (7.2.2.3).

(ii) MINIM - a simple intermolecular-energy minimising routine which translates a rigid MOL2 (ie with no conformation variation) to the optimum position (7.2.6).

(iii) The coordinates can be transferred to the Cray computer for full geometry optimisation using the package EMP (see 7.2.6).

#### 7.2.4 Molecular editing (EditMol, OP28).

The original add/delete atom/bond features in the MOLEC5 routines FRAG and HYDRO (atom positioning) have been used as a basis for the molecular editing features of IMDAC. Much re-writing of the original code was necessary in order to apply the routines in IMDAC. The changes were: the positioning of added atoms immediately after the atom of attachment (instead of at the end), adjustments for auto H-addition (such as the storing of added atoms in array HX and addition of the logical variable AUTO in appropriate places), separating out the 'delete atom' part of FRAG into a single subroutine (DELATM) - this greatly facilitated the deletion of fragments (see below) and clarified the FORTRAN code (as several backward GOTO statements could then be removed), and a general speeding up and tidying of the HYDRO vector manipulation routines and FRAG code.



The features described below are subsequent additions to the basic features.

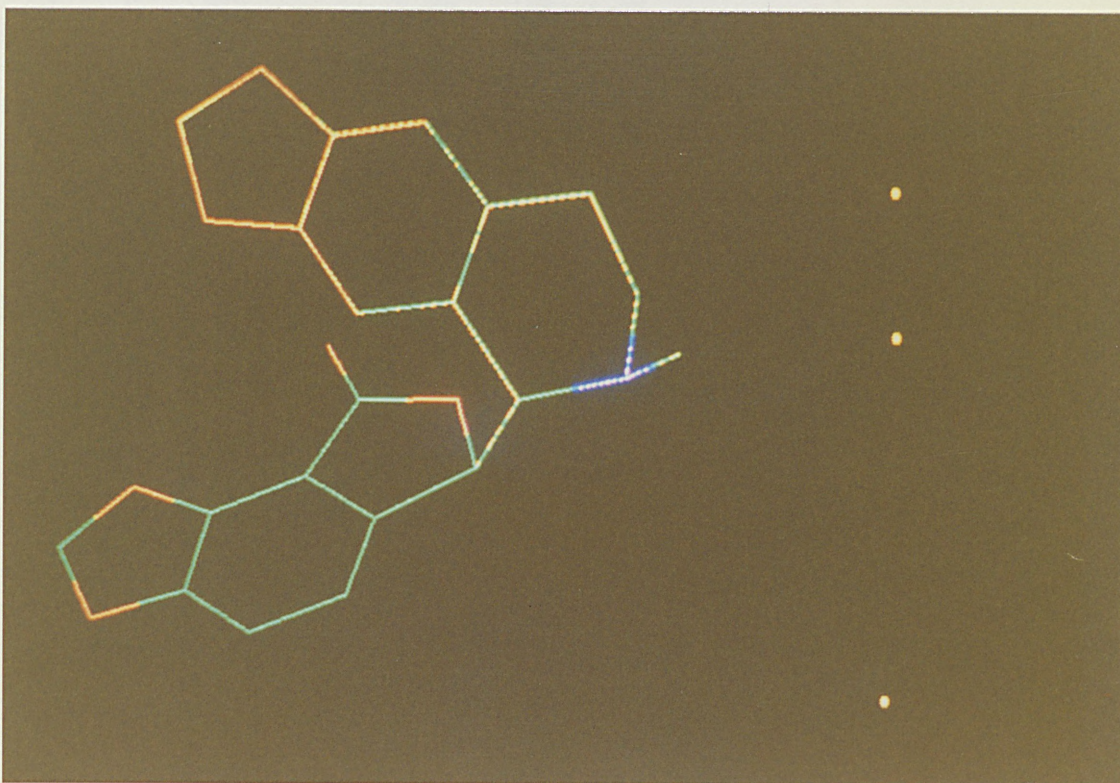
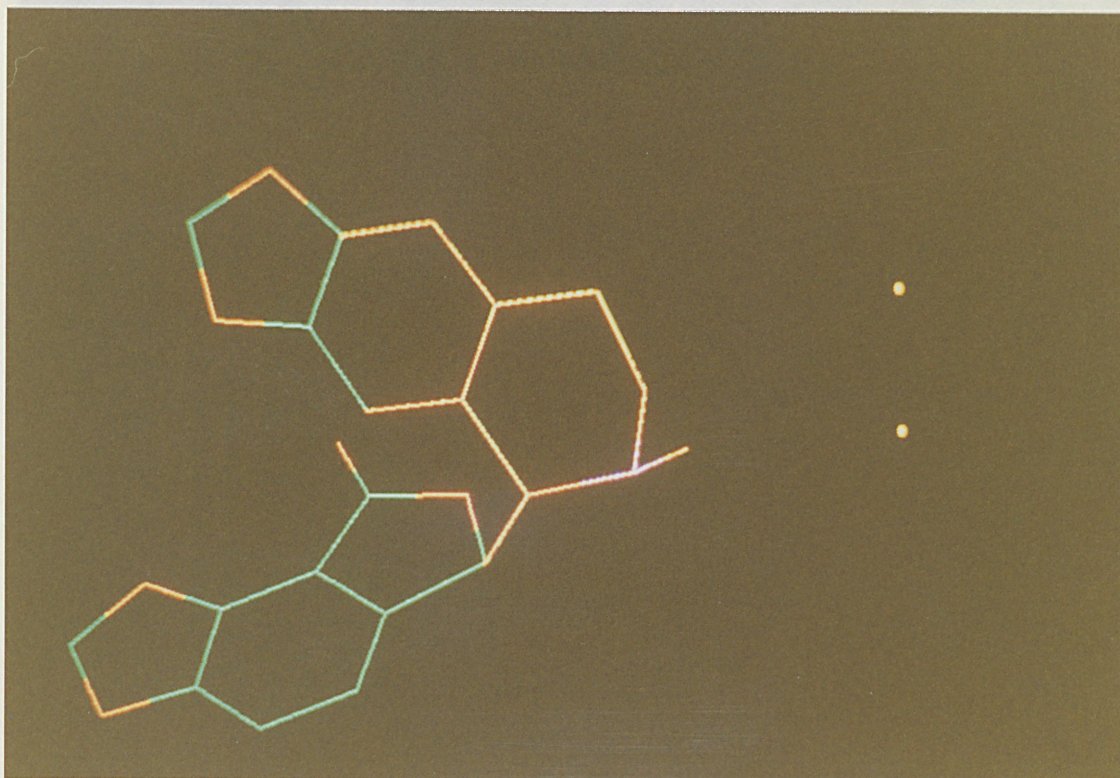
#### 7.2.4.1 Deletion of molecular fragments (ED5 and ED6).

A routine (ED5) for fragment deletion was taken from MOLEC5 and debugged\*. With this routine the input of 2 atoms is prompted for, and all atoms which are connected to the second input atom, and up to (but excluding) the first atom are deleted. As this is not always desirable (especially for ring structures, where the entire molecule could be deleted), we therefore developed an alternative method of deleting fragments (ED6) such that a chain of connecting atoms between one input atom and a second are deleted inclusively. A connecting chain chosen by the programme is drawn on the graphics screen in red. If that particular route is not desired then other routes may be surveyed rapidly and the desired fragment deleted. (The lowest numbered route is always chosen first.) For complicated molecules this process may need to be repeated on different fragments to obtain the desired result. If the 2 input atoms are not connected an error message is given. An example of the use of ED6 on the BIC molecule is given in Figure 7.9.

\* The original routine did not work at all, but became stuck in an infinite loop.



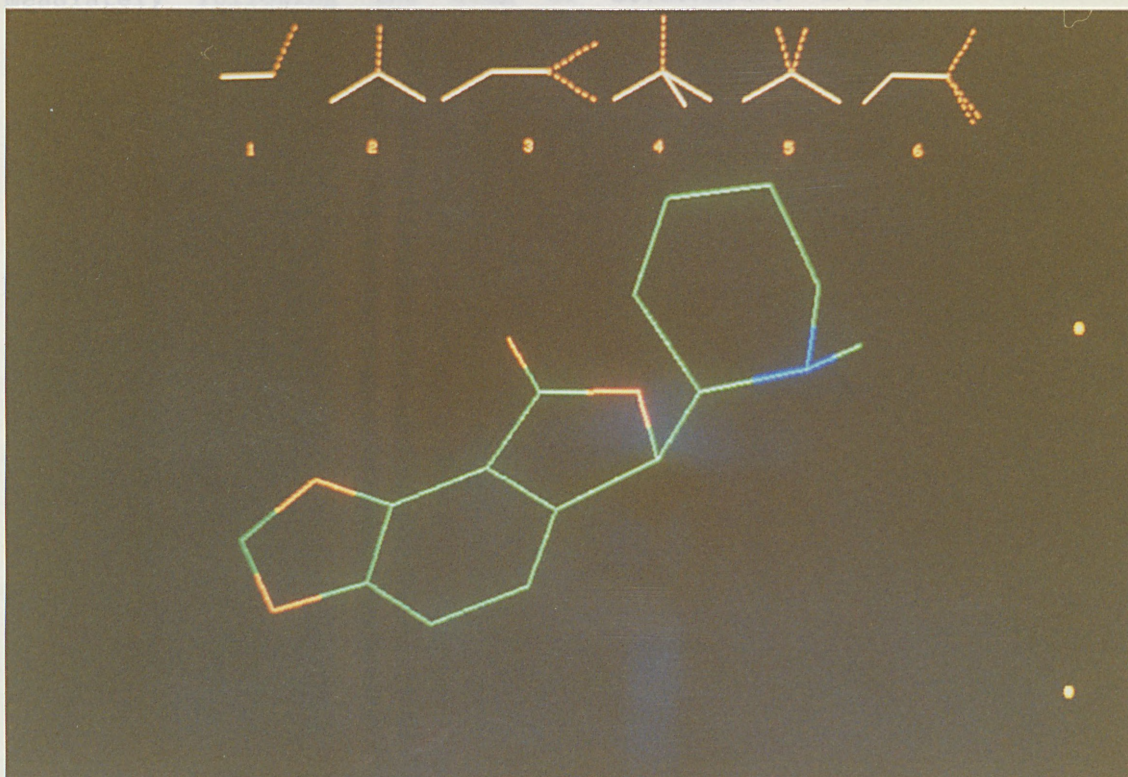
Figure 7.9a. Example of molecular fragment deletion (ED6). (see Figure 7.9b). The two pictures below show 2 paths between C5 and C8 in BIC chosen by IMDAC for possible fragment. The latter fragment was subsequently deleted (see Figure 7.9b).



\*\* The atom types need to be exact, in order that the correct connectivity is established (eg  $sp^2$  hybridised atoms will be different to  $sp^3$ .)



Figure 7.9b. BIC fragment after deletion of a fragment (see Figure 7.9a). 'Help' information for atom addition is displayed at the top of the picture (see 7.2.4.5).



#### 7.2.4.2 Automatic addition of hydrogen atoms (ED9).

The routine ADDH has been designed for the rapid addition of hydrogens to large molecules with the regular sequence of atoms in residues found in the Brookhaven data bank (see also EDIT and EDPC - 7.2.7 ii). For molecules with no regular sequence of atoms, but with correctly defined atom types (ITYPE - see Table 7.1), the routine SMLADH was devised<sup>\*\*</sup>.

The routine HPOSN was then written to enable these routines to function rapidly, with all the added hydrogens being stored (in HX and NATCHH) and then inserted after all atoms have been examined for H-addition possibilities.

<sup>\*\*</sup> The atom types need to be exact, in order that the correct connectivity is established (eg  $sp^2$  hybridised atoms will be different to  $sp^3$ .)

For all atom additions the added atoms are always ordered immediately following the atom of attachment. If most of the atom types are set correctly, then SMLADH may still be used and the offending atoms modified manually (using ED3 and ED4).

Table 7.1. PARAMETERS FOR 6-12 POTENTIAL

P--POLARIZIBILITY

N--EFFECTIVE NO OF ELECTRONS

R--VD VAAL RADII [ANGSTROMS]

| TYPE | SPECIES          | P    | N    | R    |
|------|------------------|------|------|------|
| 1    | H                | 0.42 | 0.85 | 1.2  |
| 2    | C (SP2)          | 1.30 | 5.2  | 1.7  |
| 3    | C (SP3)          | 0.93 | 5.2  | 1.7  |
| 4    | CH (ALIPHATIC)   | 1.35 | 6.0  | 1.95 |
| 5    | CH2 (ALIPHATIC)  | 1.77 | 7.0  | 1.95 |
| 6    | CH3 (ALIPHATIC)  | 2.17 | 8.0  | 1.95 |
| 7    | CH (AROMATIC)    | 2.07 | 6.0  | 1.90 |
| 8    | N (SP2-AMIDE)    | 1.15 | 6.0  | 1.55 |
| 9    | N (SP3)          | 0.87 | 6.0  | 1.55 |
| 10   | NH3+(AMINE)      | 2.13 | 9.0  | 1.75 |
| 11   | N+ (IMIDAZOLE)   | 2.03 | 6.0  | 1.65 |
| 12   | O (SP)           | 0.84 | 7.0  | 1.52 |
| 13   | O (SP2)          | 0.59 | 7.0  | 1.52 |
| 14   | O- (CARBOXYL)    | 2.14 | 7.0  | 1.60 |
| 15   | S (SINGLE BONDS) | 0.34 | 16.0 | 1.80 |
| 16   | S (DOUBLE BONDS) | 0.50 | 14.8 | 1.8  |
| 17   | P                | 3.45 | 14.2 | 1.8  |

#### REFERENCES FOR PARAMETERS

##### ATOM TYPE VARIABLE REF

|                                                                          |        |                                                      |
|--------------------------------------------------------------------------|--------|------------------------------------------------------|
| 3,9                                                                      | P      | OLSEN W.K. BIOPOLYMERS 12,1787(1973)                 |
| 16                                                                       | P      | LINDEBERG K.G., ET AL ACTA.CRYST. B33,2165(1977)     |
| 17                                                                       | P      | THORNTON J.M., BAYLEY P.M. J. BIOCHEM. 149 585(1975) |
| 14                                                                       | P=1.47 | " " "                                                |
| ALL OTHER P'S FROM GIBSON C., SCHERAGA H.A. Proc.Nat.Ac.Sc. 58,421(1967) |        |                                                      |
| ALL N'S FROM SCOTT R.A., SCHERAGA H.A. J.CHEM.PHYS. 42,2209(1965)        |        |                                                      |
| ALL R'S FROM BONDI A. J.PHYS.CHEM. 68,441(1964)                          |        |                                                      |

(Original Table from Islam, 1984)



#### 7.2.4.3 Inversion of a chiral centre (ED10).

A chiral centre may be inverted by finding two of the strings of atoms connected to the chiral atom (using SERCH - A6.5), and placing them into the appropriate arrays for torsion-angle rotation (NATOMS and JCON), with exclusion of the chiral atom (not inverted). Inversion is then achieved by swapping the two strings using the routine TORROT. The result is unpredictable if 2 (or more) of the 4 atoms connected directly to the chiral atom are joined by a ring structure. This can be circumvented, however, by breaking, and later re-joining, one of the ring bonds.

#### 7.2.4.4 Split/concatonate molecules (ED7 and ED8).

These options for splitting molecules (or fragments) are very useful for building new molecules by fragment addition. A molecule can therefore be read in, cut down (and/or added) to the desired fragment, and then joined to another molecule using ED8. The translation options (OP10: TR7 and TR8) for moving a molecule from one point to another point, or atom to atom, are very useful for aligning up the fragment in exactly the desired orientation. The split option (ED7) is also useful for changing the orientation of one part of a molecule to the rest.

#### 7.2.4.5 'Help' information for atom addition.

Six types of atom addition exist in IMDAC (all originally taken from MOLEC5) corresponding to the 6 possible kinds of bonding which can occur in organic compounds - excluding pentacoordination, which is

\* This could readily be included, but would increase the overall memory and execution time required for some routines (eg BONDS and FRAG).

rare, and is not covered by IMDAC\*. When the atom addition option (ED3) is selected, an 'addition type' (IGTYPE, numbered 1-6) is prompted for. As these numbers are meaningless to the general user, a 'help' option was therefore added. When IGTYPE is prompted for, if anything other than 1-6 is input, then the 6 possible types of addition are drawn at the top of the graphics screen (eg see Figure 7.9b).

#### 7.2.5 A frame of reference for rotation, translation and zoom transformations.

To enable the viewing of an unlimited number of molecules on the screen concurrently the absolute coordinates of MOL1 (usually the receptor) are maintained upon all coordinate transformations (rotation, translation and zooming), and all transformations affecting MOL1 are stored. Each new molecule read in can be placed in exactly the correct position on the screen, even if MOL1 has been moved. (This is assuming that all molecules read in are in the same coordinate system as MOL1 initially. If not they can be readily docked/superimposed using DOCK (7.2.3).)

The maintaining of this reference frame required changing the way in which the screen coordinates are stored and drawn, as well as storing all the rotation/translation information for transformations affecting MOL1.

The matrices ASTR(3,3) and FSTR(3) are used store the rotations and translations of MOL1 and are updated as follows:

(i) **For rotation:**  $\hat{X}_1 = \underline{A} \cdot \hat{X}_0 + (\hat{F} - \underline{A} \cdot \hat{F})$

Where  $\hat{F}$  = point about which rotation occurs;  $\hat{X}_0$  and  $\hat{X}_1$  are the original and rotated coordinates; and  $\underline{A}$  (taken from MOLEC5) is the rotation matrix.

Therefore ASTR and FSTR are updated:

$$\text{ASTR}_{\text{new}} = \underline{A} \cdot \text{ASTR}_{\text{old}}$$

$$\text{FSTR}_{\text{new}} = \underline{A}(\text{FSTR}_{\text{old}} - \hat{F}) + \hat{F}.$$

(Initially ASTR and FSTR are set to the null transformations of the unit matrix and zero vector respectively.)

(ii) **Translation:**

For FSTR a translation is simply a shifting of the coordinate system and is updated accordingly. ASTR is independent of all translations and requires no updating.

Two sets of coordinates are stored for each molecule - absolute (X) and screen (DUMX) coordinates. The screen coordinates are used for drawing and calculations, and the absolute coordinates are used as a fixed reference frame. (A 3x3 matrix could be used (in a similar way to ASTR) instead of having to store absolute coordinates for both sets of molecules, but this would require many changes to the IMDAC software.) If only MOL1 is moved (rotate/translate) on the screen, then the absolute coordinates of MOL1 are kept fixed and MOL2 moved in the opposite direction. The absolute coordinates are unchanged when both molecules are moved.

(iii) **Zooming.**

To make FSTR and ASTR independent of the zoom transformation only the scale factor (SCAL(1)) is altered, and the zooming performed immediately prior to drawing using the transformation:

$$\text{DUMX}_1 = \text{DUMX}_0 \cdot \text{SCAL}(1) + 128.$$

(The +128 is to place the origin at the centre of the screen.) The opposite transformation is used for taking coordinates from the screen (eg in IDATOM - see 7.2.8).

With ASTR and FSTR it is now possible to read in new data for both molecules without losing any transformations that have been performed on the old data. ASTR and FSTR can also be saved for later runs of IMDAC.

The reference frame was shown to be working correctly by reading in coordinates for MOL1, subjecting MOL1 to various transformations, and then reading in the original MOL1 coordinates as MOL2. For all combinations of rotate/translate/zoom the two molecules superimposed exactly. (ASTR and FSTR are updated to allow for the non-commutative nature of the rotation and translation operations.)

Other graphics systems may be able to deal with more than two molecules, but do not as a rule have a reference frame for transformations. An advantage of this reference frame is that the data for all molecules stored on disk are automatically in the same coordinate system, once they have been positioned (7.2.3) correctly and saved by IMDAC.

#### 7.2.6 Energy calculations (OP25, 26 and 40).

All the energy routines in IMDAC were originally taken from MOLEC5, but required substantial modification for use with proteins (see Appendix A7). Options exist for calculating hard 6-12 potentials, electrostatic potentials and explicit inclusion of hydrogen bonding. (These options are set in the energy setting-up routine, SETUPE, which can be called with OP19.) Although many of the usual molecular mechanics terms are not included, the excluded terms (bond stretching and twisting potentials) are small and, with the exception of torsional potentials - which can readily be allowed for, do not greatly effect relative energies for calculations on molecules with



fixed geometry (fixed bond lengths and angles, and fixed protein torsion angles). The energy routines in IMDAC were designed as a rapid first approximation. Separate routines were designed for interface (7.2.7 iii) to an existing, and more elaborate, programme (EMP - see below) for accurate energies (of drugs docked within a protein), and refinement of molecular structures.

Activation of OP26 (SeeEn) during energy calculations sends the sum of potentials for individual atoms to a data file: VLJ/ for 6-12, VSTAT/ for electrostatic and VHBOND/ for H-bonds, for inspection later. This is useful for locating the offending atoms in structures of high energy. (For intermolecular potentials the CC option (OP22 - see 7.2.2.3) complements OP26.)

OP40 (MINIM) can be used for minimising intermolecular energy, and works by optimising the cartesian coordinates of a rigid drug molecule docked (7.2.3) within a receptor. The programme utilises a minimising routine from the Harwell library (Islam, 1985), and requires first derivatives of the energy. These are readily obtained by differentiating the 6-12 and electrostatic potentials. A special parameter (DFN) exists for dealing with problems in obtaining satisfactory convergence. DFN is a multiple of the intermolecular energy for that particular cycle. If the default value of 1.0 fails to give reasonable convergence, then a value between -1 and 1 may work better.

We tested MINIM by attempting to optimise the position of a drug for which bound coordinates are known (BAG - see 7.3.2) and found a change in energy from +9.8kcal/mol for the original coordinates, to -27.74kcal/mol for the optimised coordinates. This relatively small lowering of energy shows that the potentials used are reasonably accurate in determining the position of a docked drug.

The whole coordinates can be sent to the Cray computer and a more elaborate Energy Minimising Programme (EMP - Haneef, 1985) used to further refine the structure of both drug and receptor. In practice we found the extra refinement obtained using EMP to be small (see 7.3.1), which supports the accuracy of the optimised drug position obtained with MINIM.

#### 7.2.7 Interfaces to other programmes.

##### (i) MNDOIN and MM2IN (OP32 and OP31).

With these routines the cartesian coordinates (X) and molecular connectivity (ICON and NCON) are converted into internal coordinates suitable for MNDO (or similar QM programme) input, or into the connectivity required for MM2 input. The advantages of these interfaces are an enormous saving in time (these routines take only a few seconds compared with the 1 hour+ required for manual input of molecules up to ca 50 atoms), and the lack of errors in the data produced - manual data preparation of this kind is prone to errors which may not be detected very quickly. For MNDO input, all geometry optimisation variables are switched on by default and can easily be switched off by changing a '1' parameter to a '0'. Symmetry can only be added manually, but this is straightforward and is not usually necessary because most drug molecules contain little symmetry of value.

Note that the programme we have been using for MM2 input is a much simpler version of MM2IN, which we wrote as a pre-processing programme for MM2, and is on the Amdahl computer at ULCC (MM2PRE - see Appendix A6.2). This is because the IMDAC version of MM2IN is not currently working due to some bugs in the programme. An improved algorithm, which should correct the bugs, has been written, but has not

yet been implemented, since, though not as convenient to use, the Am-dahl version works adequately.

(ii) Atom type definition and atomic charge routines - EDIT and EDPC.

These programmes are separate from IMDAC and make use of the regular order of atomic coordinates (within residues) for proteins stored on the Brookhaven Data Bank (Bernstein et al, 1977), and define the atom type, ITYPE (EDIT), and atomic charge, PC (EDPC), for each atom. The required input is standard Brookhaven coordinates for EDIT, or EDIT output for EDPC. Output is in the standard format used by IMDAC. (Other formats are also accepted by IMDAC - see below.)

(iii) Non-standard input file formats in IMDAC.

Options exist (in READAT) for reading-in atomic coordinates directly from the Brookhaven data bank (BROK), or output coordinates from the an energy-minimisation programme (EMP - Haneef, 1985) on the Cray computer (EMP). Alternatively, coordinates may be read-in in free-format (NSTD), with the only condition being that the first 4 characters specify the atom name followed by the 3 atomic coordinates (x,y and z). This option is useful when using eg MM2 output coordinates.

With the EMP option the atom types are set by conversion from the EMP atom types (with the exception of atom types 3,7,8,11,14 and 16 (see Table 7.1) which need to be set manually, due to a different definition of atom type in EMP). For the other two options, BROK and NSTD, ITYPE is set using the first character of the atom name (H,C,N,O,S,P or, for others, ITYPE is set to 20 - a dummy atom value) with the default that the atom is saturated. The exact ITYPE values

may then be set in BONDS from the calculated connectivity for that type of atom. (Note that this only works correctly for molecules with all hydrogen present - either explicitly or implicitly (ITYPE=3-5 - see Table 7.1).) It is possible to calculate the correct connectivity in BONDS when the atom types are not defined exactly, because the bonding radii are dependent only on the kind of atom (eg H, C, N etc), and not on the precise value of ITYPE. (See Table 7.1 for ITYPE definitions.) ITYPE only needs to be defined precisely for the energy (7.2.6) and close contact calculations (7.2.2.3), and automatic addition of hydrogens (7.2.4).

Any attempt to use the wrong format will result in an error message to that end, and the title and first line of the data file will be printed on the screen to facilitate correction of the error.

### 7.2.8 Features specifically designed for overcoming the limitations of available hardware.

Some features which come under this heading have already been described - eg DrawRes and Hilight (7.2.2.4). In this section we focus on certain features which were designed to overcome the slow drawing speed (no buffered, refresh display - as with eg the Evans and Sutherland PS300 series graphics devices), and lack of real-time rotation with the Prime 550/Sigma S5660 graphics setup. (Real-time rotation is very useful in visualising the 3D nature of an image on a 2D screen.)

#### (i) Rock (OP44).

A rough impression of 3D is obtained by rapidly viewing alternate superimposed stereoimages. The rapid alternation is achieved by



drawing the molecule(s) twice, using different 'pen' numbers (colours). Then one group of 'pens' is set to the current colours (which are dependent on the model number - MOD1-MOD10) while the other group is set to the background colour (ie invisible). The situation is then reversed, and repeated as many times as desired.

(ii) Options for drawing only limited sections of a molecule.

The option for drawing residues individually (DrawRes, OP39) has already been mentioned (7.2.2.4), and is useful for examining details of groups within (or close to) the active-site - particularly in conjunction with the CC option (7.2.2.3), and the IdentAtm option (OP23), by which the residues of the receptor close to the docked drug can be found.

An option also exists for cutting down the number of atoms drawn to only a specified range of atoms. This can be useful if only one section (or single residue) of a large molecule needs to be examined from different views. The range is prompted for on exiting OP39 (DrawRes), with default values set to the residue last drawn.

Another option for cutting out unwanted information is a window on the z-axis (OP18). With OP18 activated a slice through the molecule can be viewed and important detail more clearly visualised.

In addition to the above features, because it is not always easy to see whether one atom is further back (or closer) than another\*, the value of the z-coordinate is included in the information given in the identify/label atom option (OP23).

---

\* Real-time rotation with depth-cueing would solve the problem, but, as mentioned earlier, is not possible with the hardware currently available.

### 7.2.9 Producing hard-copy output.

The routines BENSON and BENDRW were written for producing drawings of journal quality directly from IMDAC (OP14). Within these routines options exist for: labelling specific atoms (eg Figures 1.1 and 7.14), atom colour coding (C, N and O; eg Figure 5.2), and the enhancement of MOL2 (eg Figure 5.2).

A separate option (DefCol, OP29) also exists in IMDAC for re-defining any colour, in order that a copy of whatever is on the screen may be sent to the (black and white) hardcopy device (Tetronix 4631/2), or for taking quality photographs directly from the screen (eg with a black background instead of the default green - compare Figures 7.5a and b).

### 7.3 Applications of IMDAC.

#### 7.3.1 Phospholypase-A2 - an enzyme for which only isolated receptor coordinates are known.

##### (i) Background.

The action of the esterolytic enzyme PA2 is to specifically cleave the 2-acyl linkage of phosphoglycerides in a calcium-dependent reaction (Dijkstra, Drenth and Kalk, 1981). The products of the reaction depend on the substrate, but often result in the formation of leukotrienes, which are associated with coronary thrombosis. This and other malevalent actions of PA2 have prompted a search for potent, specific inhibitors of PA2 (Withnall, 1984).

A mechanism for the ester cleavage by PA2 involving HIS48, GLY30 and the calcium ion has been proposed (Verheij et al, 1980), and is given in Figure 7.10. We have used these key groups to position a phosphatidyl ethanolamine substrate (PETH - Figure 7.11) into the active site of PA2 and performed subsequent energy minimisation. Because PA2 was used mainly for development work on IMDAC we only give results for the fitting of known substrates into the PA2 active site. (Many useful results were, however, obtained for TLN - see 7.3.2.)

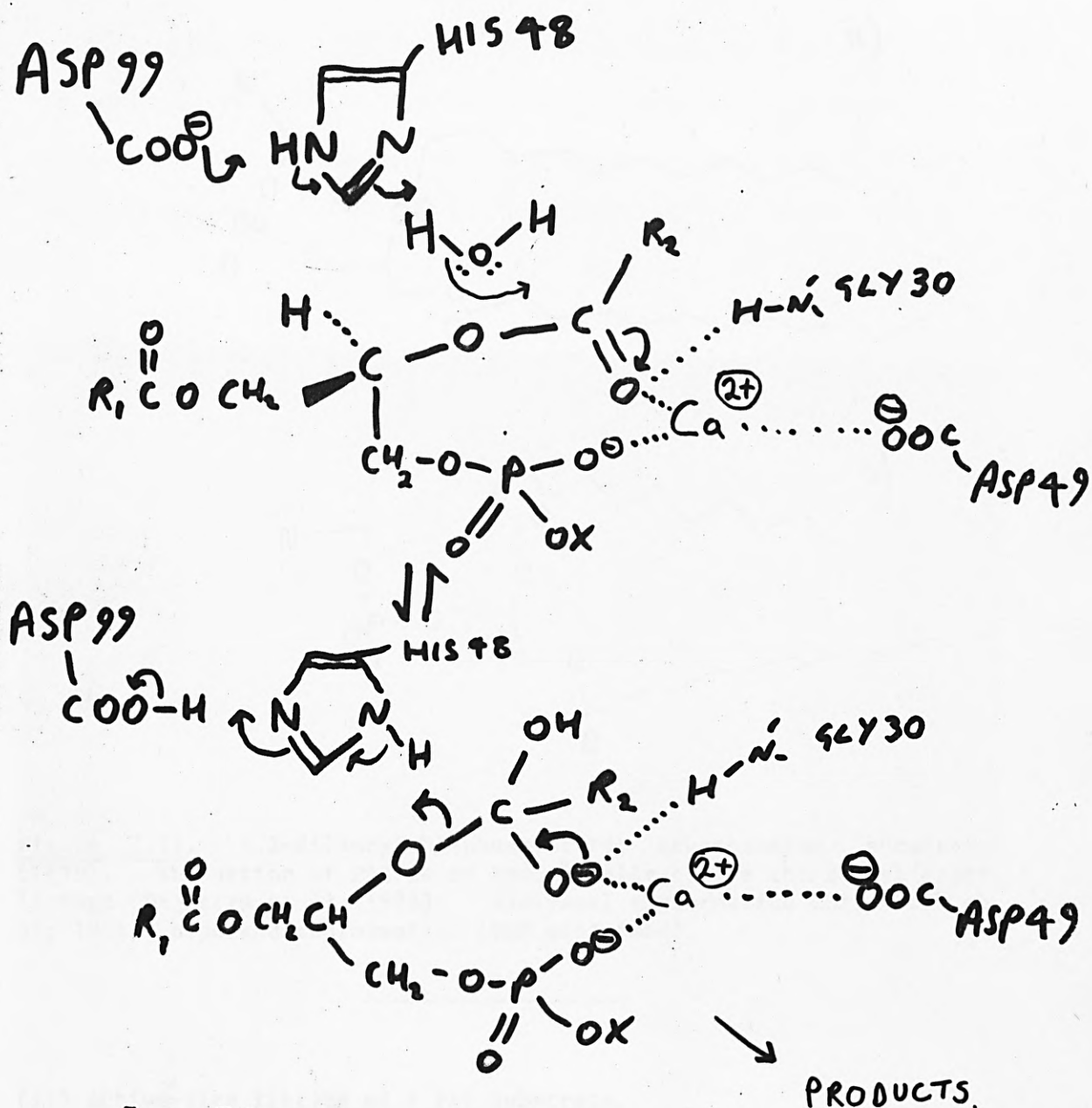
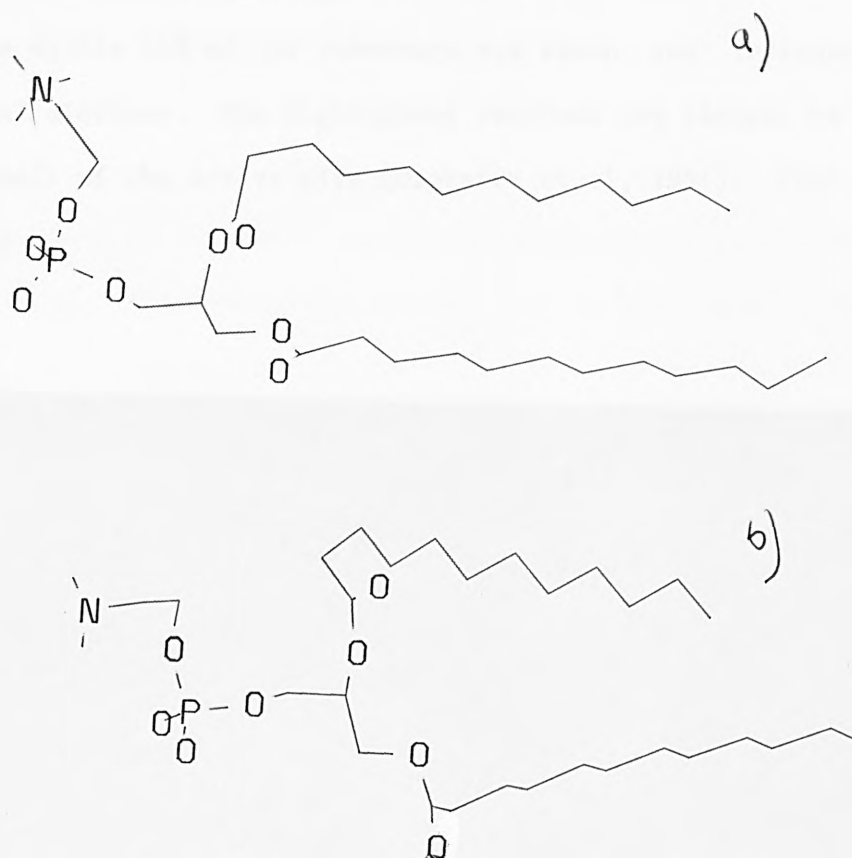


Figure 7.10.

SCHEMATIC REPRESENTATION OF A PROPOSED CATALYTIC MECHANISM FOR  $P. A_2$





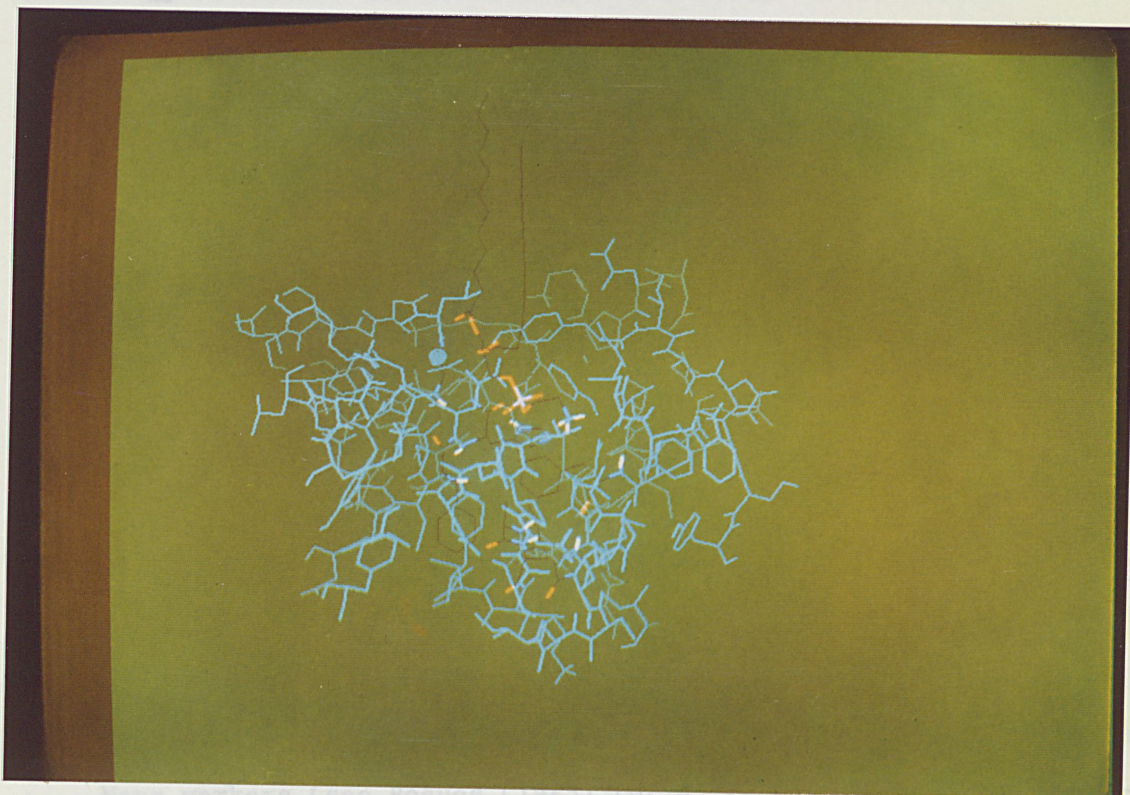
**Figure 7.11.** 1,2-dilauryl-DL-phosphatidylethanolamine substrate (PETH). The action of PA2 is to specifically cleave the 2-acyl ester linkage (Dijkstra et al, 1981). a) crystal conformation (Hitchcock et al, 1974), b) docked conformation (EMP minimised).

#### (ii) Active-site fitting of a PA2 substrate.

We initially used the N-H group of HIS48 and the COO group of GLY30, which are reputed (Verheij et al, 1980) to be directly involved in ester cleavage (Figure 7.10,) to position (using DOCK) a phosphatidylethanolamine (PETH) substrate (Figure 7.11) into the active site of PA2. The PA2 active-site cleft is large and clearly visible (Figure 7.12), but had to be located by highlighting HIS48 and GLY30 and rotating PA2 to the desired view, because the cleft-searching option (CLEFT- see 7.2.2.1) had not been developed at that

Figure 7.12 PA2 active-site region (OP30).

Only residues within 12Å of the substrate are shown and included in the energy calculations. The highlighted residues are thought to constitute the wall of the active site (Dijkstra et al, 1981). (See also Figure 7.5b.)



tempted to find inhibitors of enkephalinase (ENK, a zinc metallo peptidase) with the aim of finding a non-addictive analgesic agent. The molecular structure of ENK has not yet been determined but the enzyme TLN, for which the molecular structure is known, is so similar to ENK that the active site of TLN can be used as a model for drugs acting at the ENK active-site (Palfreyman, 1985). TLN has already been used as a model (Henguever et al, 1984) in successfully designing inhibitors of angiotensin converting enzyme, another zinc metallo peptidase.



time. We were able to show later using CLEFT that there are no other clefts in PA2 which are large enough to be considered as active sites. The position of PETH was refined by using the CC option (7.2.2.3) and by rotating and translating PETH manually to a position for which the number and magnitude of disfavourable close contacts was minimised. Once the intermolecular energy was brought down to a reasonable value ( $<ca 10^7$  kJ!), the energy minimising routine MININ (7.2.6) was used to lower the intermolecular energy to only a few kJ. On transferring the coords of PA2 and PETH docked in this position for energy minimisation using EMP (Haneef, 1985 - see 7.2.6) the resultant optimised PETH coords displayed a kink in one of the aliphatic chains (Figure 7.11b), a feature which has been shown to be an absolute requirement for fitting of PETH (Dijkstra et al, 1981). This result has at least shown that we were on the right lines!

### 7.3.2 TLN - an enzyme for which several drug/receptor structures are known - a model for other zinc metallo peptidases.

#### (i) Background.

Many workers (see Roques (1985) for a recent review) have attempted to find inhibitors of enkephalinase (ENK, a zinc metallo peptidase) with the aim of finding a non-addictive analgesic agent. The molecular structure of ENK has not yet been determined but the enzyme TLN, for which the molecular structure is known, is so similar to ENK that the active site of TLN can be used as a model for drugs acting at the ENK active-site (Palfreyman, 1985). TLN has already been used as a model (Hanguaver et al, 1984) in successfully designing inhibitors of angiotensin converting enzyme, another zinc metallo peptidase.

(ii) Modelling of ENK inhibitors using the TLN active site.

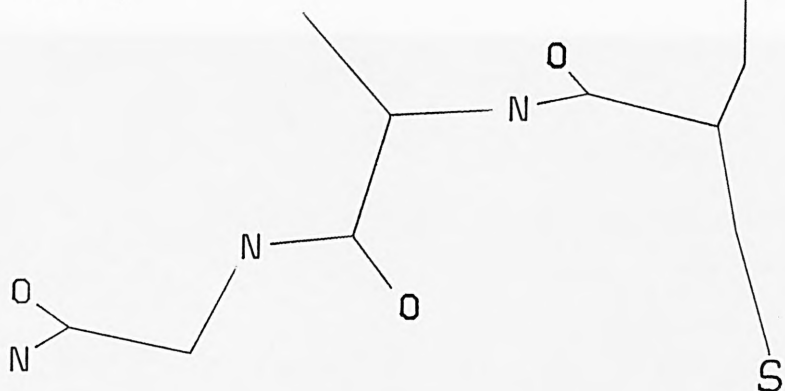
By starting with coordinates for known TLN inhibitors we have modelled novel analogues which have been found to display significant ENK activity (Palfreyman, 1986). In addition we have found a novel possible mode of binding for a new (Palfreyman, 1985) ENK inhibitor. This work, however, was undertaken in collaboration with a major pharmaceutical company, and since it is of a confidential nature, we have been requested not to report details of our main findings. We can, however, still show the basic principles of applying IMDAC to a known inhibitor and show how the possibilities of modifying the inhibitor for better interaction with the receptor (in this case TLN) could be examined using IMDAC.

a) (2-benzyl-3-mercaptopropanoyl)-L-alanylglycinamide (BAG) - type ENK inhibitors.

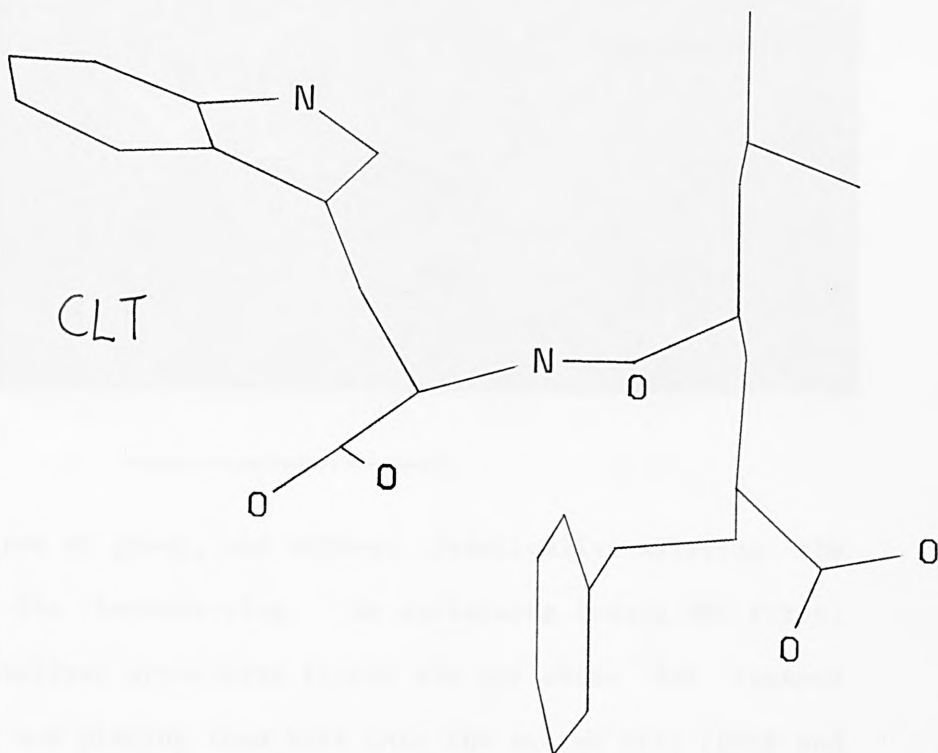
By starting from BAG (Figure 7.13) a TLN inhibitor for which coordinates of the bound drug are available<sup>\*</sup>, we could examine the space around BAG for possibilities of modifying the drug, and thereby improve its interaction with the receptor (TLN and ENK). We used CC (7.2.2.3) to examine the space around whole molecule (Figure 7.4) initially, and then on the benzene-ring in particular (Figure 7.14), for the possibilities of adding extra atoms or groups. It was evident from the CC's (Figure 7.14) that there is no room for adding atoms at the meta and para positions, but at the ortho position there is plenty of space for

\* We input BAG coordinates manually from the paper by Monzingo and Matthews (1982), and TLN coordinates from the Brookhaven Data Bank (Bernstein et al, 1977). The low intermolecular energy found for BAG positioned in this way showed that no significant errors had been introduced by the mixing of coordinates. (No significant difference was found between BAG positioned into any of the four different TLN coordinates available on the data bank - 3/4/5/7TLN).



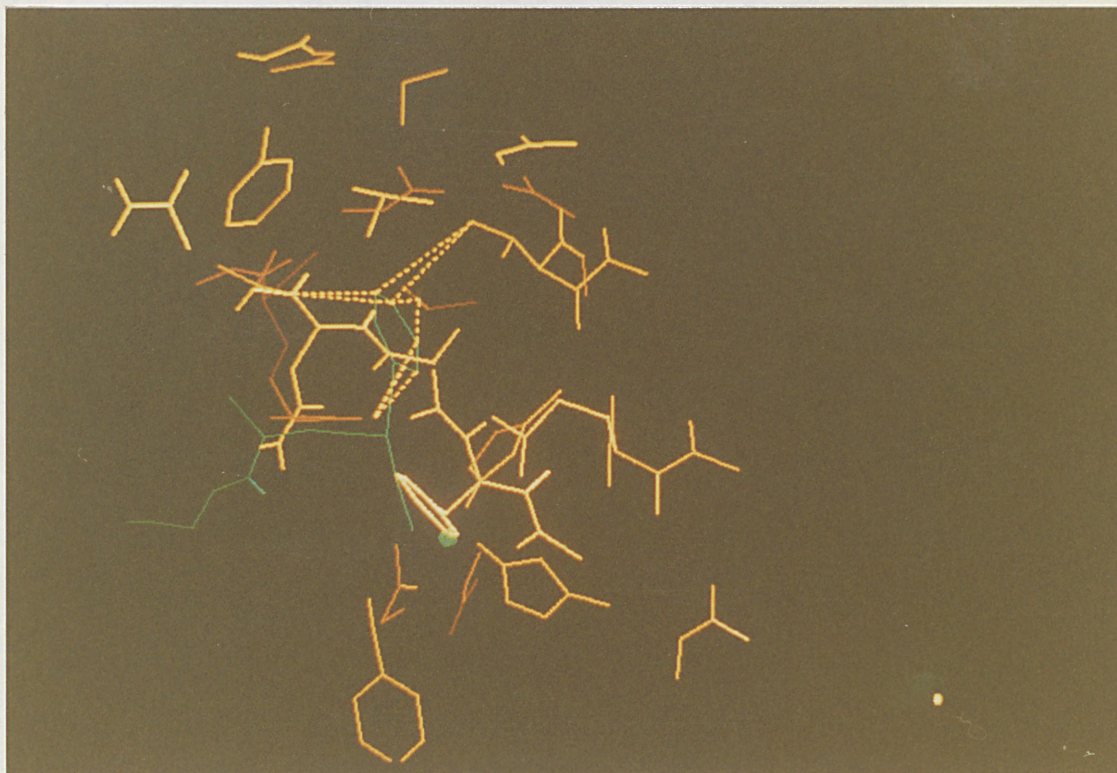


CLT



- 173 -

74  
Figure 7.14. Close contacts around the benzene ring in BAG. The CC's shown were calculated with high sensitivity (SENS=1.0Å) and give an indication of how much room there is near a particular atom for modifying the drug. SENS can be further increased and the atoms for which no CC's are shown here further investigated.



adding an extra atom or group, and without drastically altering the conformation of the benzene-ring. By optimising (using MM2 first, then MNDO) the resultant structures (these are not shown for reasons already given), and placing them back into the active site (DOCK and MINM) we obtained a ca 200 kJ mol<sup>-1</sup> lower intermolecular interaction energy than for BAG (optimised in the same way), which implies that the new drugs will be of greater potency than BAG, as was indeed found (Palfreyman, 1986). (Although only qualitative deductions tend to be made on drug potency from energy calculations (Bush, 1986), an energy difference of the order of 200 kJ mol<sup>-1</sup> cannot be ignored!)



b) Inhibitors with zinc-chelating groups other than sulphur.

The sulphur ligand, common in TLN inhibitors, is by no means the ideal ligand due to its susceptibility to hydrolysis in the body. We have therefore examined the possibility of replacing sulphur with a different zinc-chelating group, hydroxamic acid (found in several ENK-active drugs (Fournie-Zaluski et al, 1985)). We used BAG (and some of our modelled drugs) as base structures, and found that simply swapping the S atom for the larger acid group does not work since the acid is then much too close to the zinc ion - even with optimisation of the drug torsion angles. However, when the central chiral atom is inverted (an easy task with IMDAC -see 7.2.4.3) a good overall fit was obtained. However, hydroxamic acid derivatives have been found not to be stereospecific in ENK activity (Fournie-Zaluski et al, 1985). This shows that care is needed in using TLN as a model for other receptors. Support for our findings above comes, however, from the fact that hydroxamic acid derivatives are stereospecific for aminopeptidase activity (Fournie-Zaluski et al, 1985) - another zinc metallo peptidase.

7.4 Overview of Part 2.

We have developed an extensive molecular graphics system, IMDAC, for examining the key features of drug and receptor molecules - both visually and mathematically.

Possible active-sites (binding sites) in a receptor molecule can be found, and the shape of the active-site cleft visualised using the CLEFT-searching routine. Specific active-site residues can be highlighted, either by drawing individually (DrawRes) or by setting a switch for each residue (HiLight).

Drug molecules can be docked (superimposed) into the active-site using the DOCK routine, with primary optimisation of the molecular fit by the routine CONF. The coordinates of the docked drug molecule may be optimised further using the energy-minimising routine (MINIM). (Additional optimisation, if desired, may be obtained by transferring drug and receptor coordinates to the Cray computer and optimising with EMP (Haneef, 1985).) Furthermore, the routine CC (Close intermolecular Contacts) may be used as a guide for manual adjustment of the drug orientation or optimisation of drug torsion angles (to fit the receptor more efficiently).

Another important use of the CC option is for examining the space around a docked (or bound) drug for possibilities of modifying the drug. For this there are molecular modification routines in IMDAC, with which atoms/bonds/fragments may be added/deleted, hydrogen atoms may be added automatically (even with proteins), and chiral centres can be inverted. The resultant structures can be refined using interfaces to QM (and MM) programmes.

We have successfully applied IMDAC to model drug analogues which have been found (Palfreyman, 1986) to be of significant potency, and to determine a novel possible mode of binding for an ENK inhibitor (the receptor for which we have used TLN as a model). The basic principles of this work have been given, with examples of application of the techniques on a known drug, BAG. In addition, we have shown that the replacement of the zinc-chelating sulphur ligand in BAG (and BAG analogues) with a hydroxamic acid ligand can result in the opposite (and increased) stereospecificity of the drug.



## Appendix A1

### Anomalous gaps in GABA $x_T$ distributions.

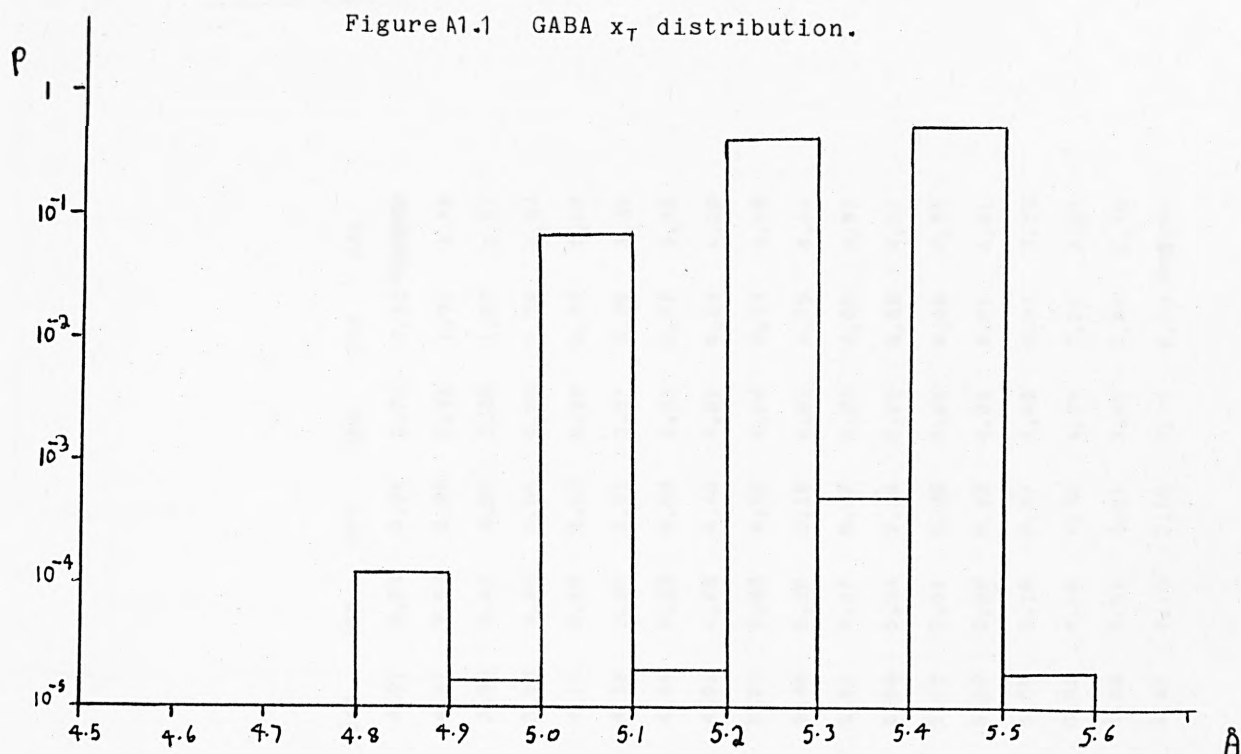
If a GABA potential energy surface at  $20^\circ$  increments is mapped onto a  $0.1\text{\AA}$  (or less) sampled  $x_T$  distribution anomalous gaps will almost certainly appear in the  $x_T$  distribution (see example in Figure A1.1a). This is because a  $20^\circ$  change in either central torsion angle,  $T_2$  or  $T_3$ , will produce an up to  $0.57\text{\AA}$  change in  $x_T$ ! (Table A1.1). We therefore wrote a programme for interpolating a  $20^\circ$  energy grid into a  $5^\circ$  grid, using a cubic spline function (NAG routine E01ADF) fitted to 2 dimensions (INTERP - see A6.5). The result is a smoother distribution with no anomalous gaps (Figure A1.1b). Note that even though  $x_T$  can still change by more than  $0.1\text{\AA}$  with a  $5^\circ$  grid, there are no other gaps. This is because the larger  $x_T$  changes occur at high-energy folded conformations, which do not contribute to the  $x_T$  distribution.

Since the interpolation programme was somewhat time consuming\* and the calculated energies are not accurate enough to justify such an elaborate treatment, a compromise of mapping  $20^\circ$  grids onto  $0.2\text{\AA}$   $x_T$  distributions was used for most of our results (especially the SOLVEFF results, where the  $x_T$  angle-dependence is small).

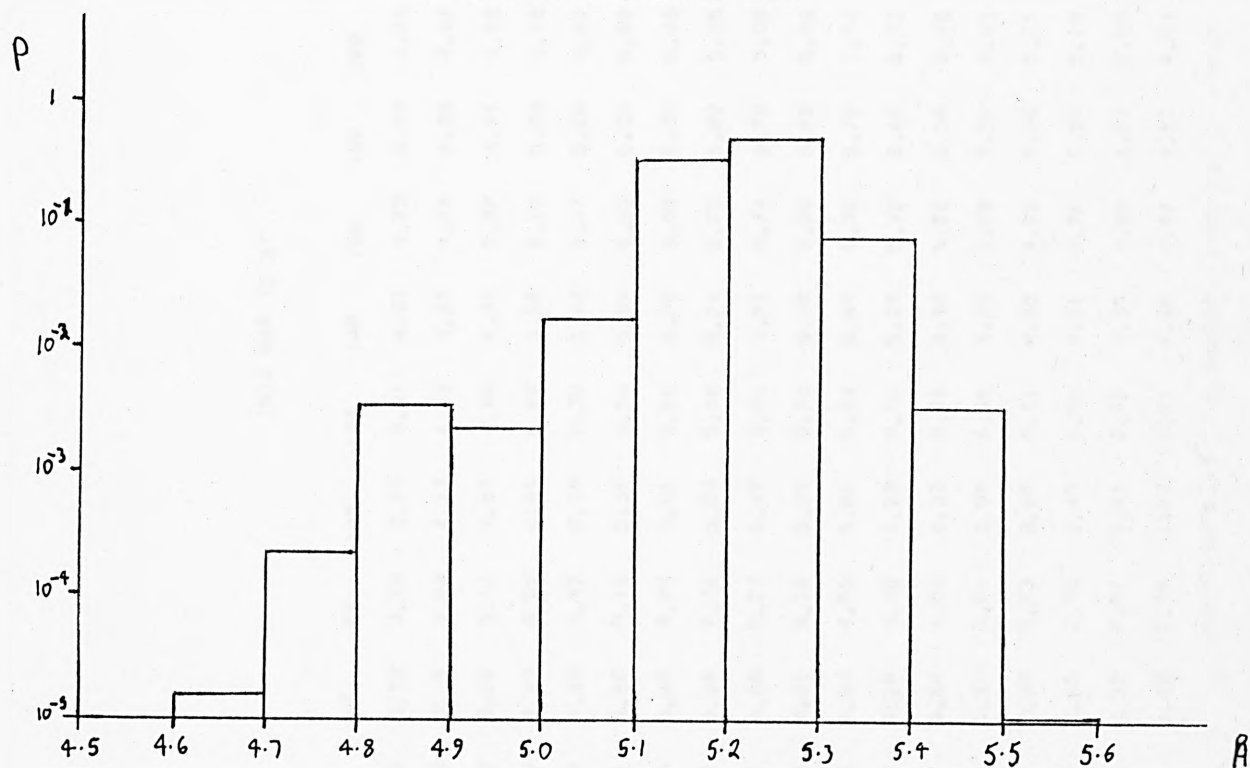
---

\* The interpolation programme was written for the Honeywell computer at City, because it was the only computer at the time with reasonable interactive access. Since direct file transfer between ULCC computers and the Honeywell has only recently become possible, all the data for the interpolation programme had to be transferred manually.

Figure A1.1 GABA  $x_T$  distribution.



a) Without interpolation.



b) With interpolation.

Table A1.1 GABA  $x_T$ 's

THIS MAP IS XT

| T3:     | 0    | 20   | 40   | 60   | 80   | 100  | 120  | 140  | 160  | 180  | 200  | 220  | 240  | 260  | 280  | 300  | 320  | 340              |
|---------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------------------|
| T2: 0.  | 1.48 | 1.68 | 2.15 | 2.72 | 3.28 | 3.79 | 4.21 | 4.53 | 4.73 | 4.79 | 4.73 | 4.53 | 4.21 | 3.79 | 3.28 | 2.72 | 2.15 | <del>1.48</del>  |
| T2: 20. | 1.60 | 2.04 | 2.58 | 3.15 | 3.68 | 4.13 | 4.49 | 4.73 | 4.84 | 4.82 | 4.67 | 4.40 | 4.01 | 3.53 | 2.99 | 2.42 | 1.91 | 1.60             |
| T2: 40. | 2.10 | 2.55 | 3.09 | 3.63 | 4.11 | 4.51 | 4.80 | 4.96 | 5.00 | 4.91 | 4.69 | 4.36 | 3.92 | 3.41 | 2.87 | 2.35 | 1.97 | 1.87             |
| T2: 60. | 2.63 | 3.09 | 3.59 | 4.09 | 4.52 | 4.86 | 5.09 | 5.20 | 5.16 | 5.04 | 4.78 | 4.41 | 3.96 | 3.45 | 2.95 | 2.53 | 2.30 | 2.34             |
| T2: 80. | 3.17 | 3.59 | 4.04 | 4.49 | 4.87 | 5.16 | 5.35 | 5.42 | 5.37 | 5.20 | 4.92 | 4.55 | 4.11 | 3.65 | 3.23 | 2.90 | 2.77 | 2.87             |
| T2:100. | 3.65 | 4.01 | 4.41 | 4.80 | 5.13 | 5.38 | 5.54 | 5.59 | 5.53 | 5.36 | 5.09 | 4.75 | 4.36 | 3.96 | 3.61 | 3.37 | 3.29 | 3.39             |
| T2:120. | 4.05 | 4.35 | 4.68 | 5.00 | 5.29 | 5.51 | 5.65 | 5.70 | 5.65 | 5.50 | 5.28 | 4.99 | 4.66 | 4.33 | 4.04 | 3.84 | 3.77 | 3.85             |
| T2:140. | 4.36 | 4.57 | 4.82 | 5.09 | 5.34 | 5.54 | 5.68 | 5.74 | 5.72 | 5.62 | 5.45 | 5.22 | 4.96 | 4.70 | 4.46 | 4.28 | 4.20 | 4.22             |
| T2:160. | 4.54 | 4.67 | 4.85 | 5.06 | 5.27 | 5.47 | 5.61 | 5.71 | 5.74 | 5.70 | 5.59 | 5.44 | 5.24 | 5.03 | 4.82 | 4.64 | 4.53 | 4.49             |
| T2:180. | 4.60 | 4.64 | 4.75 | 4.91 | 5.10 | 5.29 | 5.46 | 5.60 | 5.69 | 5.72 | 5.69 | 5.60 | 5.46 | 5.29 | 5.10 | 4.91 | 4.75 | 4.64             |
| T2:200. | 4.54 | 4.49 | 4.53 | 4.64 | 4.82 | 5.03 | 5.24 | 5.44 | 5.59 | 5.70 | 5.74 | 5.71 | 5.61 | 5.47 | 5.27 | 5.06 | 4.85 | 4.67             |
| T2:220. | 4.36 | 4.22 | 4.20 | 4.28 | 4.46 | 4.70 | 4.96 | 5.22 | 5.45 | 5.62 | 5.72 | 5.74 | 5.68 | 5.54 | 5.34 | 5.09 | 4.82 | 4.57             |
| T2:240. | 4.05 | 3.85 | 3.77 | 3.84 | 4.04 | 4.33 | 4.66 | 4.99 | 5.28 | 5.50 | 5.65 | 5.70 | 5.65 | 5.51 | 5.29 | 5.00 | 4.68 | 4.35             |
| T2:260. | 3.65 | 3.39 | 3.29 | 3.37 | 3.61 | 3.96 | 4.36 | 4.75 | 5.09 | 5.36 | 5.53 | 5.59 | 5.54 | 5.38 | 5.13 | 4.80 | 4.41 | 4.01             |
| T2:280. | 3.17 | 2.87 | 2.77 | 2.90 | 3.23 | 3.65 | 4.11 | 4.55 | 4.92 | 5.20 | 5.37 | 5.42 | 5.35 | 5.16 | 4.87 | 4.49 | 4.04 | 3.59             |
| T2:300. | 2.63 | 2.34 | 2.30 | 2.53 | 2.95 | 3.45 | 3.90 | 4.41 | 4.78 | 5.04 | 5.18 | 5.20 | 5.09 | 4.86 | 4.52 | 4.09 | 3.59 | 3.09             |
| T2:320. | 2.10 | 1.87 | 1.97 | 2.35 | 2.87 | 3.41 | 3.92 | 4.36 | 4.69 | 4.91 | 5.00 | 4.96 | 4.80 | 4.51 | 4.11 | 3.63 | 3.09 | 2.55             |
| T2:340. | 1.00 | 1.00 | 1.01 | 2.42 | 2.09 | 3.53 | 4.01 | 4.40 | 4.67 | 4.82 | 4.84 | 4.73 | 4.49 | 4.13 | 3.68 | 3.15 | 2.58 | <del>99.00</del> |

SUM OF EXPS IS 1.352380 FWHM IS -15.2

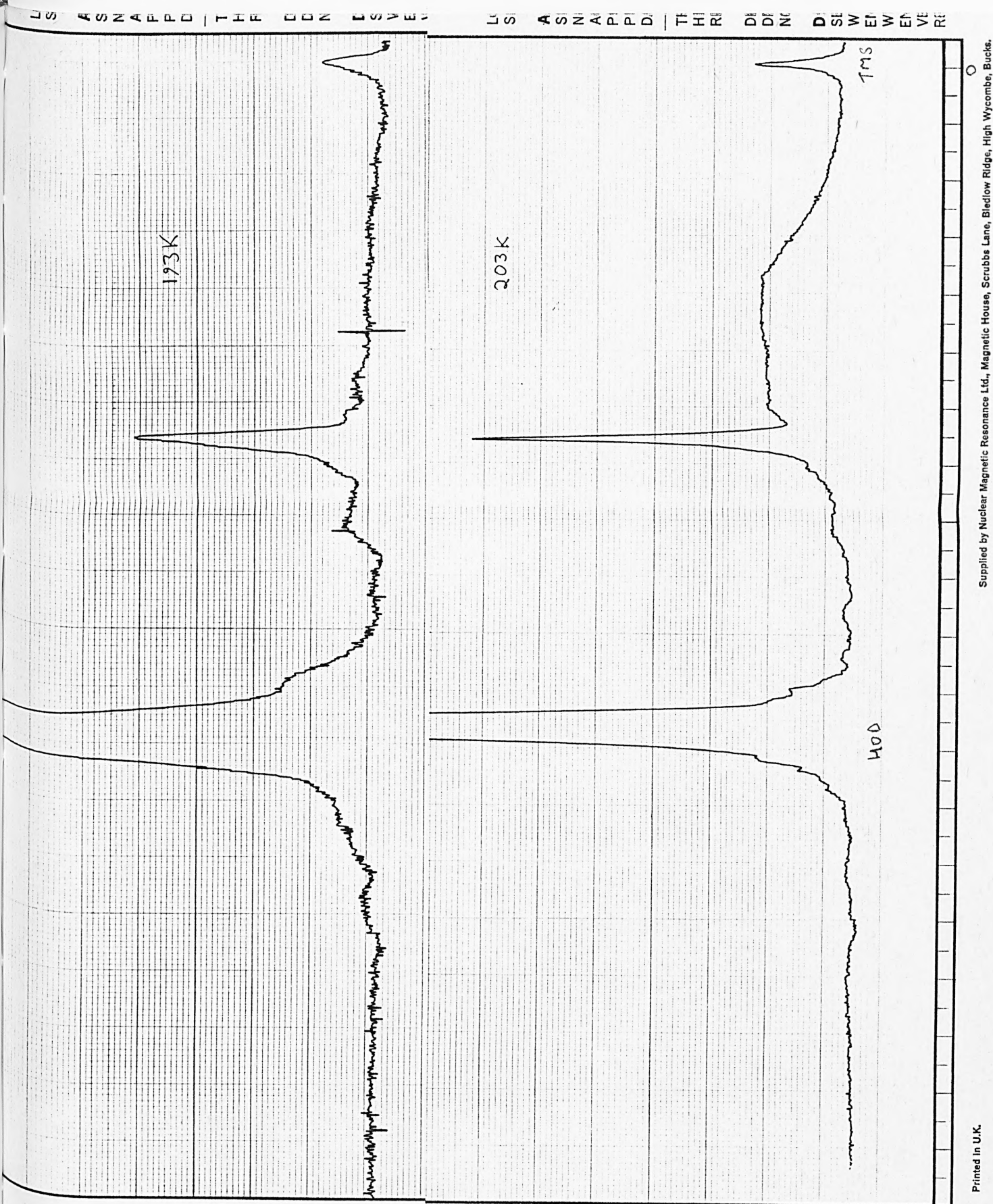
Appendix A2. Additional NMR and mass spectral data.

- 1) CABA NMR spectra at low temperature (solvent froze)
- 2) BIC NMR data in acetone and dichloromethane
- 3) MEC and ISO NMR and mass spectral data
- 4) BIC N-ring proton data
- 5) MeBIC data:
  - (i) acetone - spin decoupling, NOE, variable temperature
  - (ii) deuterium oxide, NOE
  - (iii) DMSO spectra
- 6) HBIC data (in deuterium oxide) - variable temperature, spin decoupling

Figure A2.1 CABA NMR spectra at low temperature (solvent frozen)



Figure A2.1 GABA NMR spectra at low temperature (solvent frozen)



| Proton             | 223K     | 233K      | 248K      | 263K                   | 296K              |
|--------------------|----------|-----------|-----------|------------------------|-------------------|
| H4 e               | 1.97(m)  | 2.01(m)   | 2.06(m)   | 2.12(m)                | 2.194(0)          |
| H3 e               | 2.37(m)  | 2.392(sp) | 2.433(sp) | 2.462(sp)              | 2.510(sp)         |
| Me                 | 2.481(s) | 2.496(s)  | 2.508(s)  | 2.522(s)               | 2.545(s)          |
| H4 a               | 2.51(m)  | 2.53(m)   | 2.543(m)  | 2.566(0)               | 2.605(0)          |
| H3 a               | 2.645(m) | 2.671(m)  | 2.703(sp) | 2.737(sp)              | 2.795(sp)         |
| H1                 | 4.215(d) | 4.207(d)  | 4.185(d)  | 4.168(d)               | 4.140(d) 4.0*     |
| H9                 | 5.775(d) | 5.768(q)  | 5.747(q)  | 5.729(q)               | 5.71(q) 4.0, 1.0* |
| H6'                | 5.882(d) | 5.946(d)  | 6.03 (d)  | 6.107(d)               | 6.22(q) 7.9, 1.0* |
| OCH <sub>2</sub> O | 6.058(q) | 6.051(q)  | 6.028(q)  | 6.013(s)               | 6.00(s)           |
| OCH <sub>2</sub> O | 6.269(q) | 6.266(q)  | 6.253(s)  | 6.243(s)               | 6.22(q)           |
| H5                 | 6.750(s) | 6.736(s)  | 6.707(s)  | 6.689(s)               | 6.67(s)           |
| H8                 | 6.955(s) | 6.923(s)  | 6.863(s)  | 6.815(s)               | 6.75(s)           |
| H5'                | 7.106(d) | 7.100(d)  | 7.082(d)  | 7.070(d) <sup>11</sup> | 6.05(d) 7.9*      |

Table A2.1 Additional NMR data for BIC in acetone at 400 MHz.  
 Chemical shifts are in PPM, with acetone as reference at 2.086.  
 S = Singlet, d = doublet, q = quartet, sp = septet, 0 = Octet,  
 M = multiplet, \* coupling constant in Hz.

| Proton             | 223K | 233K  | 263K  | 306K     |
|--------------------|------|-------|-------|----------|
| H4e                | 2.06 | 2.08  | 2.14  | 2.24(m)  |
| H3e                | 2.43 | 2.44  | 2.48  | 2.53(m)  |
| Me                 | 2.48 | 2.49  | 2.51  | 2.55(s)  |
| H4a                | 2.50 | 2.51  | 2.55  | 2.61(m)  |
| H3a                | 2.68 | 2.695 | 2.74  | 2.81(m)  |
| H1                 | 4.10 | 4.09  | 4.06  | 4.03(d)  |
| H9                 | 5.55 | 5.55  | 5.55  | 5.55(d)  |
| OCH <sub>2</sub> O | 5.95 | 5.945 | 5.94  | 5.93(s)  |
| OCH <sub>2</sub> O | 6.17 | 6.17  | 6.17  | 6.16(s)  |
| H6'                | 5.9  | 5.98  | 6.09  | 6.23(d)  |
| H5                 | 6.58 | 6.58  | 6.595 | 6.60(s)  |
| H8                 | 6.50 | 6.50  | 6.51  | 6.515(s) |
| H5'                | 6.90 | 6.90  | 6.92  | 6.93(d)  |

Table A2.2 Additional NMR data for BIC in dichloromethane at 400 MHz. Chemical shifts are in ppm with dichloromethane as reference at 5.33 ppm.

s = singlet, d = doublet, m = multiplet. Other spectra, taken at a different concentration, gave chemical shifts within 0.01 ppm of these.

Figure A2.2.

MEC NMR spectrum (254K, acetone)

9/20.004  
1.00  
1.00 = 1/2

MEC  
254K  
acetone

9/20.004  
MEC  
254K

| #    | CURSOR | FREQUENCY | PPM    | INTENSITY |
|------|--------|-----------|--------|-----------|
| 1    | 2866   | 3710.248  | 7.4184 | 5.728     |
| 2    | 2883   | 3701.993  | 7.4019 | 6.806     |
| 3    | 3034   | 3627.674  | 7.2533 | 4.542     |
| 4    | 3051   | 3619.394  | 7.2368 | 3.699     |
| 5    | 5123   | 2599.199  | 5.1970 | 25.966    |
| 6    | 6454   | 1944.167  | 3.8873 | 39.470    |
| 7    | 6494   | 1924.538  | 3.8480 | 56.761    |
| 8    | 6507   | 1918.332  | 3.8356 | 4.328     |
| 9    | 7253   | 1550.772  | 3.1007 | 1.871     |
| 10   | 7286   | 1534.975  | 3.0691 | 6.255     |
| 11   | 8334   | 1018.909  | 2.0373 | 6.389     |
| 12   | 8361   | 1005.594  | 2.0106 | 23.573    |
| 13   | 8365   | 1003.514  | 2.0065 | 35.691    |
| 14   | 8370   | 1001.272  | 2.0020 | 45.589    |
| 15   | 8374   | 999.186   | 1.9978 | 32.127    |
| 16   | 8379   | 996.935   | 1.9933 | 17.636    |
| 17   | 10404  | .101      | .0002  | 12.535    |
|      |        | .269 .1   | .100   | 10.232    |
|      |        | PPM/CM=   | F1=    |           |
| 2.EP |        | 4.232     | 2.1    |           |
| F2=  |        | 1.600     | 1.60A  |           |
| F2=  |        |           | 1.     |           |



Figure A2.3. ISO NMR spectrum (254K, acetone).

JF6P00.005  
 MIN. INTENSITY = 2.112 MAXV = 0.00000 PP CONSTANT = 1.00000  
 INTENS. LEVEL = 2.112 NOISE = .00188 SENS. LEVEL = .00753  
 F1 = 5113.17 HZ = 10.2236 PPM F2 = -2943.47 HZ = -5.8854 PPM  
 FREQ. DIFF. = 8066.64 HZ

| #  | CURSOR | FREQUENCY | PPM     | INTENSITY |
|----|--------|-----------|---------|-----------|
| 1  | 3681   | 3308.967  | 6.6162  | 4.851     |
| 2  | 3721   | 3289.681  | 6.5776  | 4.748     |
| 3  | 6657   | 1844.293  | 3.6876  | 20.349    |
| 4  | 6668   | 1838.833  | 3.6767  | 19.910    |
| 5  | 6834   | 1757.272  | 3.5136  | 8.275     |
| 6  | 7229   | 1562.572  | 3.1243  | 6.025     |
| 7  | 7266   | 1544.746  | 3.0887  | 2.802     |
| 8  | 7859   | 1350.903  | 2.7011  | 13.584    |
| 9  | 8335   | 1018.578  | 2.0366  | 8.671     |
| 10 | 8370   | 1000.998  | 2.0015  | 11.573    |
| 11 | 10404  | -1.174    | -0.0003 | 5.845     |

ISO  
 254K  
 acetone

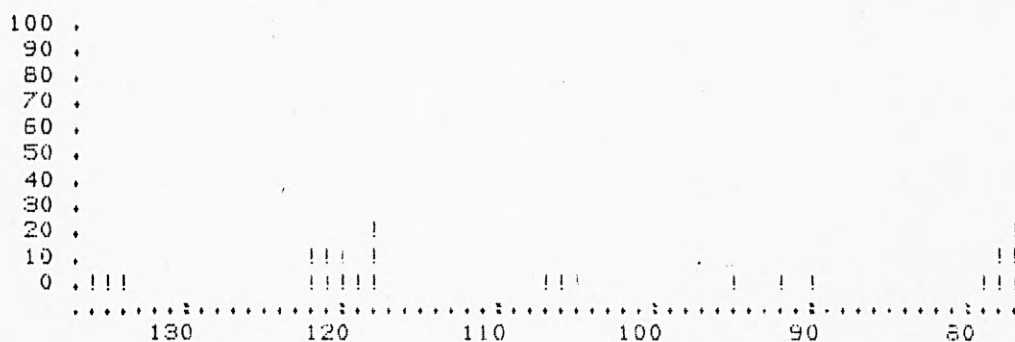
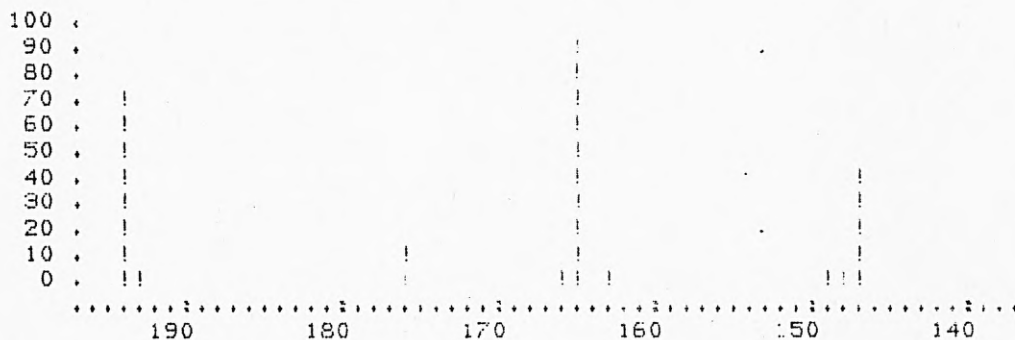
DS-50 MASS INTENSITY REPORT  
DPO:GP1.MS  
SCAN: 14, 5/20/85 14:27

Table A2.3.

MEC

mass spectrum.

IONISATION: EI  
NO. PEAKS: 171  
BASE/NREF INT: 36782./ 36782.  
TIC: 394856.  
MASS RANGE: 50.0197 - 196.0738  
RETN TIME/MISC: 1: 4/ 19/ 0/ 0



THE CITY UNIVERSITY CHEMISTRY DEPARTMENT

MS30/DS50S MASS SPECTROMETRY SYSTEM

RELEASE 10.0

G.POOLER MEC

DPO:GP1.MS  
SCAN: 14, 5/20/85 14:27

- 186 -

IONISATION: EI  
NO. PEAKS: 171  
BASE/NREF INT: 36782./ 36782.  
TIC: 394856.

| MEASURED<br>MASS | % INT.<br>BASE |
|------------------|----------------|
|------------------|----------------|

|          |         |
|----------|---------|
| 196.0738 | 1.1     |
| 195.0673 | 3.4     |
| 194.0596 | 87.5 ←  |
| 193.0524 | 13.1    |
| 191.0368 | 0.6     |
| 180.0443 | 1.6     |
| 179.0382 | 4.4     |
| 178.0510 | 0.9     |
| 177.0472 | 4.2     |
| 176.0492 | 23.4    |
| 175.0392 | 0.8     |
| 167.0626 | 1.1     |
| 166.0621 | 10.1    |
| 165.0558 | 100.0 — |
| 164.0443 | 2.3     |
| 163.0404 | 13.5    |
| 162.0304 | 1.8     |
| 161.0249 | 4.3     |
| 159.8205 | 3.2     |
| 158.0382 | 0.2     |
| 152.0472 | 0.7     |
| 151.0458 | 7.9     |
| 150.0397 | 9.4     |
| 149.0366 | 11.8    |
| 148.0535 | 13.2    |
| 147.0455 | 50.7    |
| 146.0381 | 1.0     |
| 137.0427 | 1.6     |
| 136.0363 | 10.3    |
| 135.0400 | 14.2    |
| 134.0372 | 16.8    |
| 133.0296 | 5.0     |
| 132.0361 | 0.7     |
| 131.0490 | 0.6     |
| 130.0508 | 0.4     |
| 124.0586 | 0.6     |
| 123.0479 | 7.2     |
| 122.0402 | 25.1    |
| 121.0382 | 28.4    |
| 120.0457 | 25.4    |
| 119.0353 | 16.1    |
| 118.0438 | 31.2    |
| 117.0387 | 0.7     |
| 116.0304 | 1.3     |
| 109.0643 | 2.6     |
| 108.0320 | 5.3     |
| 107.0283 | 18.5    |
| 106.0283 | 10.1    |
| 105.0395 | 13.6    |
| 104.0323 | 4.5     |
| 103.0264 | 1.1     |
| 97.0918  | 0.3     |
| 96.0500  | 0.9     |
| 95.0503  | 14.0    |
| 94.0399  | 2.9     |

| MEASURED<br>MASS | % INT.<br>BASE |
|------------------|----------------|
|------------------|----------------|

|         |      |
|---------|------|
| 93.0339 | 9.7  |
| 92.0258 | 15.7 |
| 91.0404 | 9.3  |
| 90.0429 | 15.3 |
| 89.0327 | 7.9  |
| 88.0287 | 0.4  |
| 87.0222 | 0.5  |
| 86.0146 | 0.4  |
| 84.9399 | 1.3  |
| 83.0638 | 0.6  |
| 82.9436 | 2.1  |
| 82.0213 | 0.2  |
| 81.0372 | 1.5  |
| 80.0263 | 7.5  |
| 79.0279 | 15.9 |
| 78.0268 | 24.6 |
| 77.0292 | 37.0 |
| 76.0281 | 12.0 |
| 75.0212 | 9.0  |
| 74.0144 | 5.6  |
| 73.0117 | 1.6  |
| 71.0881 | 0.3  |
| 70.0777 | 0.2  |
| 69.0266 | 2.5  |
| 68.0242 | 2.4  |
| 67.0377 | 4.4  |
| 66.0242 | 9.1  |
| 65.0375 | 15.5 |
| 64.0326 | 15.3 |
| 63.0259 | 23.9 |
| 62.0185 | 12.0 |
| 61.0120 | 6.1  |
| 60.0126 | 1.2  |
| 59.0150 | 2.1  |
| 57.0730 | 0.9  |
| 56.0643 | 0.5  |
| 56.0233 | 0.2  |
| 55.0294 | 7.1  |
| 54.0204 | 1.9  |
| 53.0134 | 15.6 |
| 52.0335 | 12.1 |
| 51.0278 | 27.5 |
| 50.0197 | 25.4 |

DPO:GP1.MS

SCAN: 14, 5/20/85 14:27

IONISATION: EI

NO. PEAKS: 171

BASE/NREF INT: 36792./ 36792.

TIC: 394856.

MASS RANGE: 50.0197 - 195.0736

RETN TIME/MISC: 1: 4/ 19/ 0/ 0

MEC mass  
spectral data

| C  | C13 | H  | N | O | DEV  | MEAS MASS | #PTS | %INT   |
|----|-----|----|---|---|------|-----------|------|--------|
| 20 | 1   |    | 2 | 6 |      |           |      |        |
| 13 | 0   | 9  | 1 | 1 | -1.1 | 195.0673  | 29   | 9.42   |
| 12 | 1   | 8  | 1 | 1 | 3.4  |           |      |        |
| 10 | 0   | 11 | 0 | 4 | 1.6  |           |      |        |
| 13 | 0   | 8  | 1 | 1 | -1.0 | 194.0596  | 63   | 87.52  |
| 12 | 1   | 7  | 1 | 1 | 3.5  |           |      |        |
| 10 | 0   | 10 | 0 | 4 | 1.7  |           |      |        |
| 13 | 0   | 7  | 1 | 1 | -0.3 | 193.0524  | 31   | 13.09  |
| 12 | 1   | 6  | 1 | 1 | 4.1  |           |      |        |
| 8  | 1   | 8  | 2 | 3 | -4.4 |           |      |        |
| 10 | 0   | 9  | 0 | 4 | 2.3  |           |      |        |
| 13 | 0   | 6  | 1 | 0 | -0.3 | 176.0492  | 49   | 29.44  |
| 12 | 1   | 5  | 1 | 0 | 3.7  |           |      |        |
| 8  | 1   | 7  | 2 | 2 | -4.9 |           |      |        |
| 10 | 0   | 8  | 0 | 3 | 1.9  |           |      |        |
| 5  | 1   | 9  | 1 | 5 | -2.2 |           |      |        |
| 12 | 0   | 8  | 1 | 0 | -3.6 | 166.0621  | 29   | 10.11  |
| 11 | 1   | 7  | 1 | 0 | 0.9  |           |      |        |
| 9  | 0   | 10 | 0 | 3 | -0.9 |           |      |        |
| 8  | 1   | 9  | 0 | 3 | 3.6  |           |      |        |
| 4  | 1   | 11 | 1 | 5 | -5.0 |           |      |        |
| 12 | 0   | 7  | 1 | 0 | -2.0 | 165.0559  | 63   | 100.00 |
| 11 | 1   | 6  | 1 | 0 | 2.4  |           |      |        |
| 9  | 0   | 9  | 0 | 3 | 0.6  |           |      |        |
| 4  | 1   | 10 | 1 | 5 | -3.4 |           |      |        |
| 12 | 0   | 5  | 1 | 0 | -1.3 | 163.0404  | 31   | 13.55  |
| 11 | 1   | 4  | 1 | 0 | 2.7  |           |      |        |
| 9  | 0   | 7  | 0 | 3 | 0.9  |           |      |        |
| 4  | 1   | 8  | 1 | 5 | -3.2 |           |      |        |
| 11 | 1   | 6  | 0 | 0 | -4.5 | 151.0459  | 25   | 7.90   |
| 11 | 0   | 5  | 1 | 0 | 3.6  |           |      |        |
| 7  | 0   | 7  | 2 | 2 | -4.9 |           |      |        |
| 6  | 1   | 6  | 2 | 2 | -0.5 |           |      |        |
| 4  | 0   | 9  | 1 | 5 | -2.2 |           |      |        |
| 3  | 1   | 8  | 1 | 5 | 2.2  |           |      |        |
| 9  | 0   | 7  | 0 | 2 | 0.9  | 147.0455  | 44   | 50.73  |
| 4  | 1   | 8  | 1 | 4 | -3.2 |           |      |        |
| 4  | 0   | 7  | 2 | 4 | 4.9  |           |      |        |
| 5  | 1   | 5  | 2 | 1 | -3.3 | 122.0402  | 35   | 25.13  |
| 7  | 0   | 6  | 0 | 2 | 3.5  |           |      |        |
| 2  | 1   | 7  | 1 | 4 | -0.6 |           |      |        |
| 6  | 0   | 5  | 2 | 1 | -2.0 | 121.0382  | 40   | 28.44  |
| 5  | 1   | 4  | 2 | 1 | 2.5  |           |      |        |
| 3  | 0   | 7  | 1 | 4 | 0.7  |           |      |        |
| 7  | 0   | 6  | 1 | 1 | 0.7  | 120.0457  | 44   | 25.43  |
| 2  | 1   | 7  | 2 | 3 | -3.3 |           |      |        |
| 4  | 0   | 8  | 0 | 4 | 3.4  |           |      |        |
| 6  | 1   | 5  | 2 | 0 | -4.3 | 118.0439  | 53   | 31.24  |
| 8  | 0   | 6  | 0 | 1 | 2.0  |           |      |        |
| 3  | 1   | 7  | 1 | 3 | -2.1 |           |      |        |
| 4  | 1   | 3  | 1 | 0 | -3.1 | 78.0268   | 43   | 24.58  |
| 2  | 0   | 6  | 0 | 3 | -4.9 |           |      |        |
| 1  | 1   | 5  | 0 | 3 | -0.4 |           |      |        |
| 5  | 0   | 3  | 1 | 0 | 2.6  | 77.0292   | 43   | 36.95  |
| 0  | 1   | 4  | 2 | 2 | -1.5 |           |      |        |
| 5  | 0   | 3  | 0 | 0 | 2.4  | 63.0259   | 35   | 23.95  |
| 0  | 1   | 4  | 1 | 2 | -1.7 | - 188 -   |      |        |



Figure A2.4a. BIC N-ring protons (NMR spectrum - acetone)

BIC  
acetone  
296K  
N-ring protons

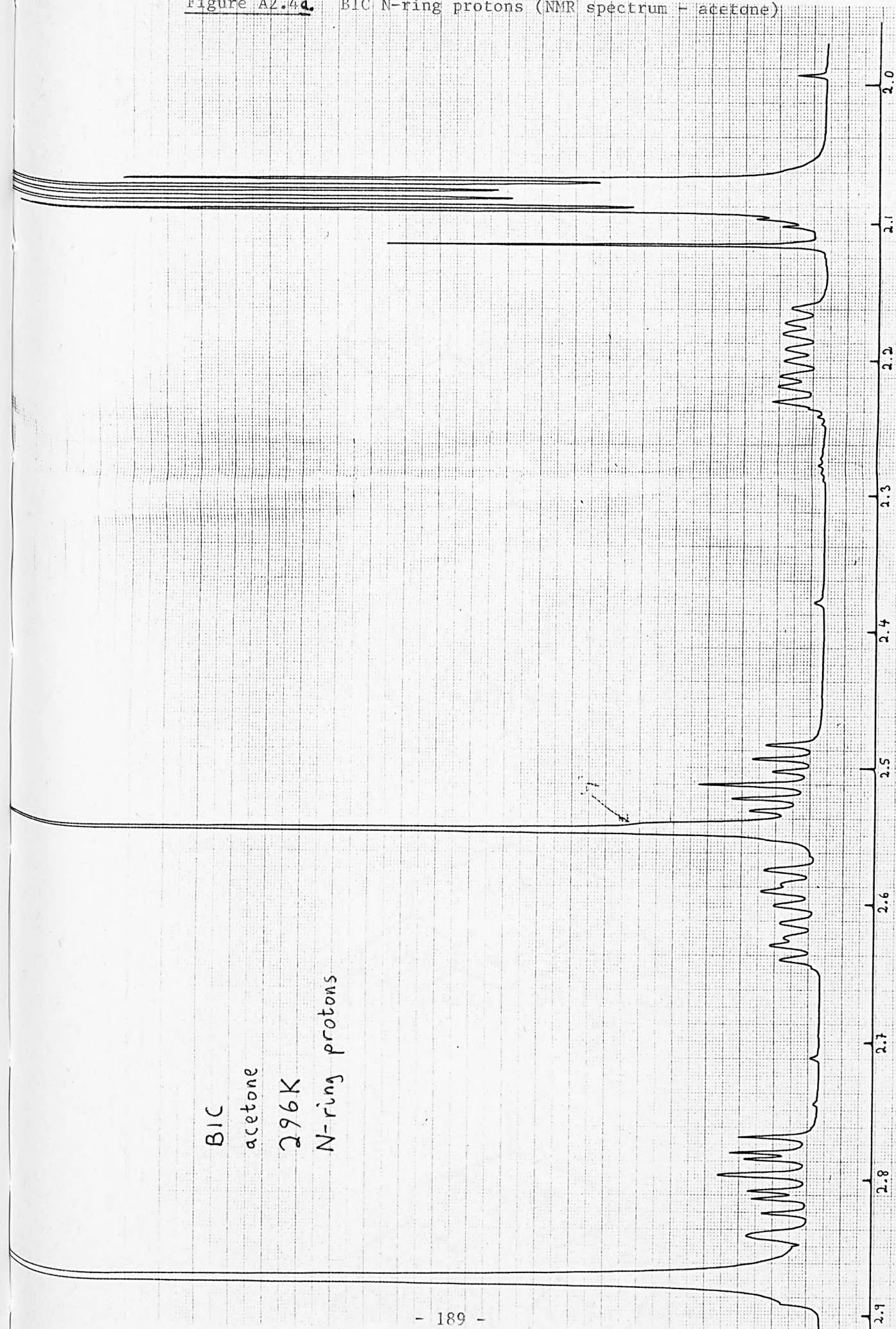
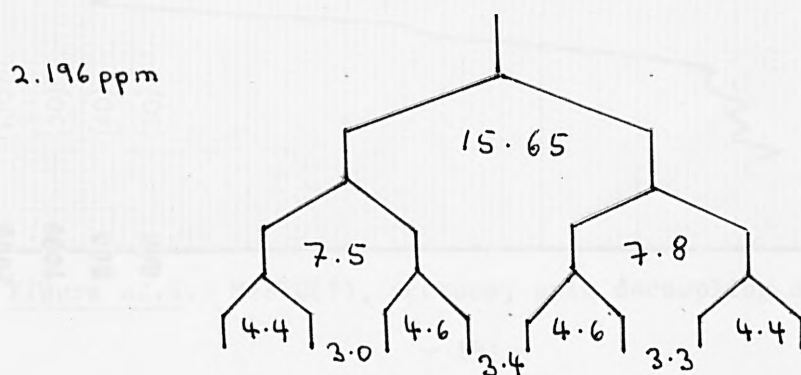
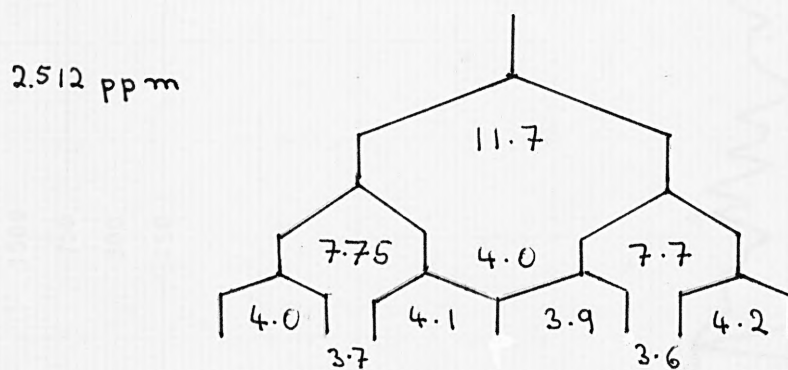
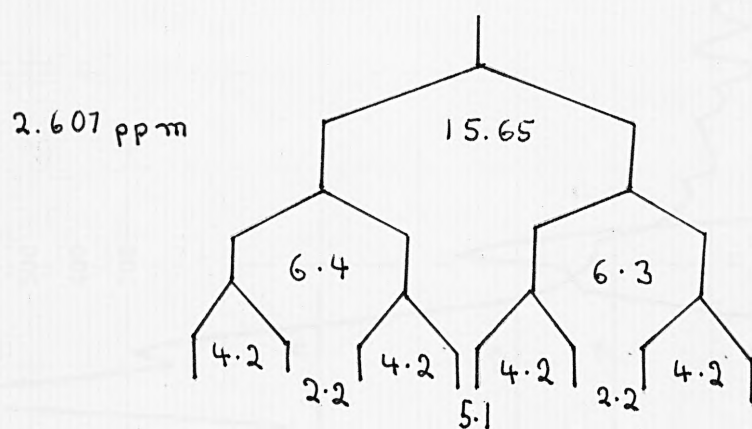
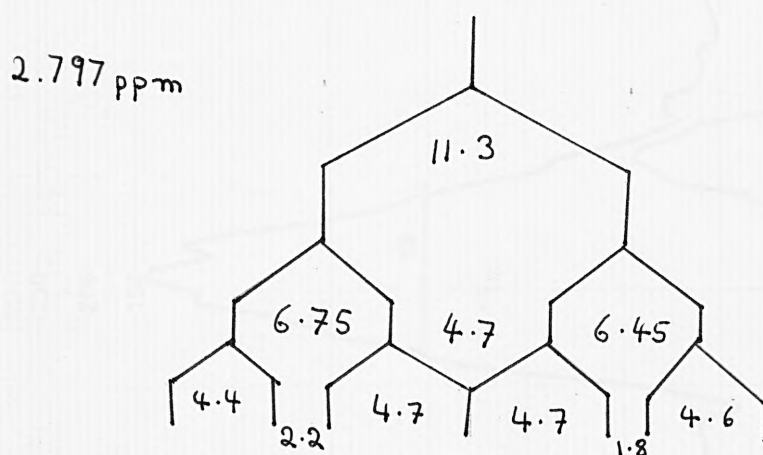


Figure A2.4b. BIC N-ring proton coupling constants.



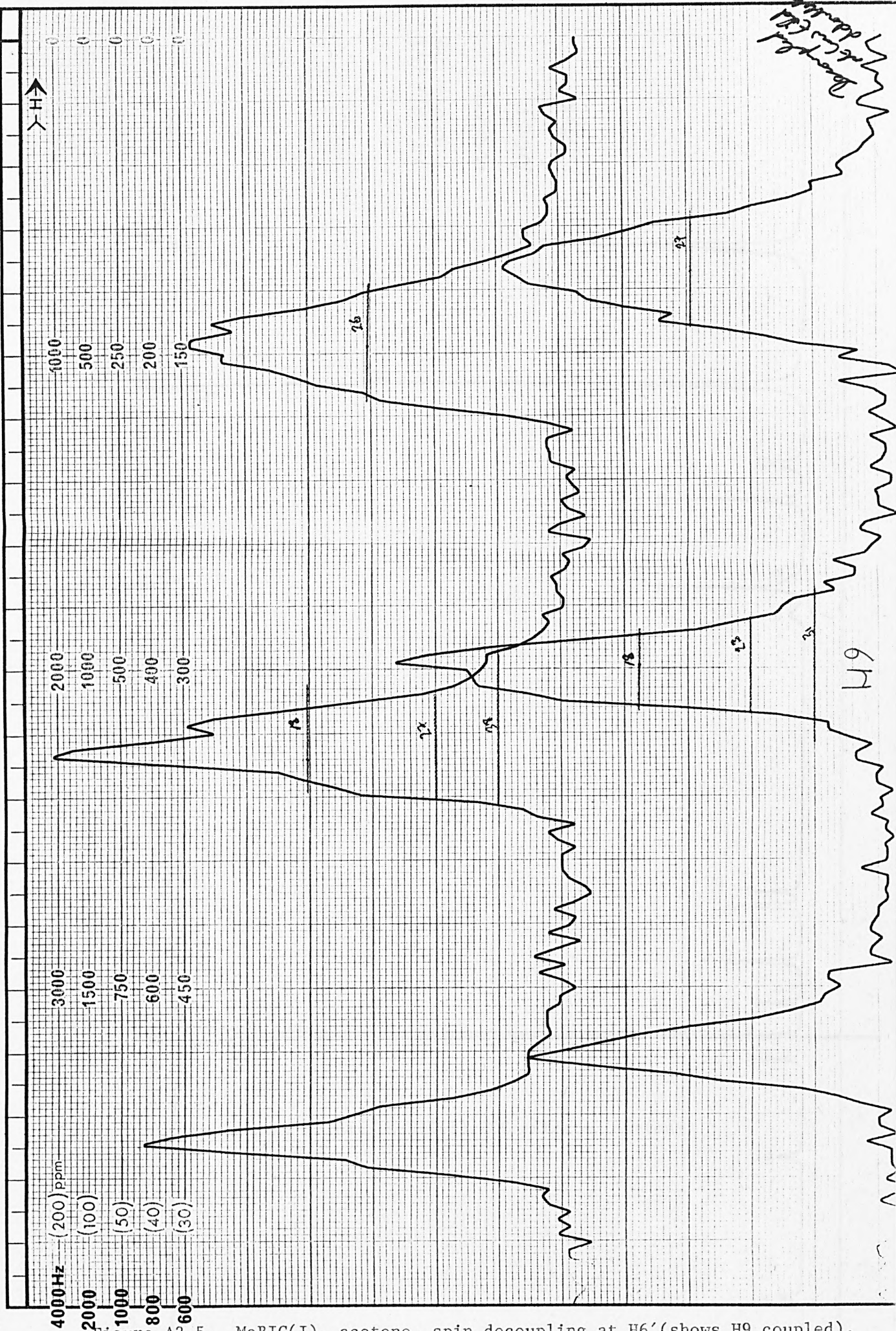


Figure A2.5. MeBIC(I), acetone, spin decoupling at H6' (shows H9 coupled).



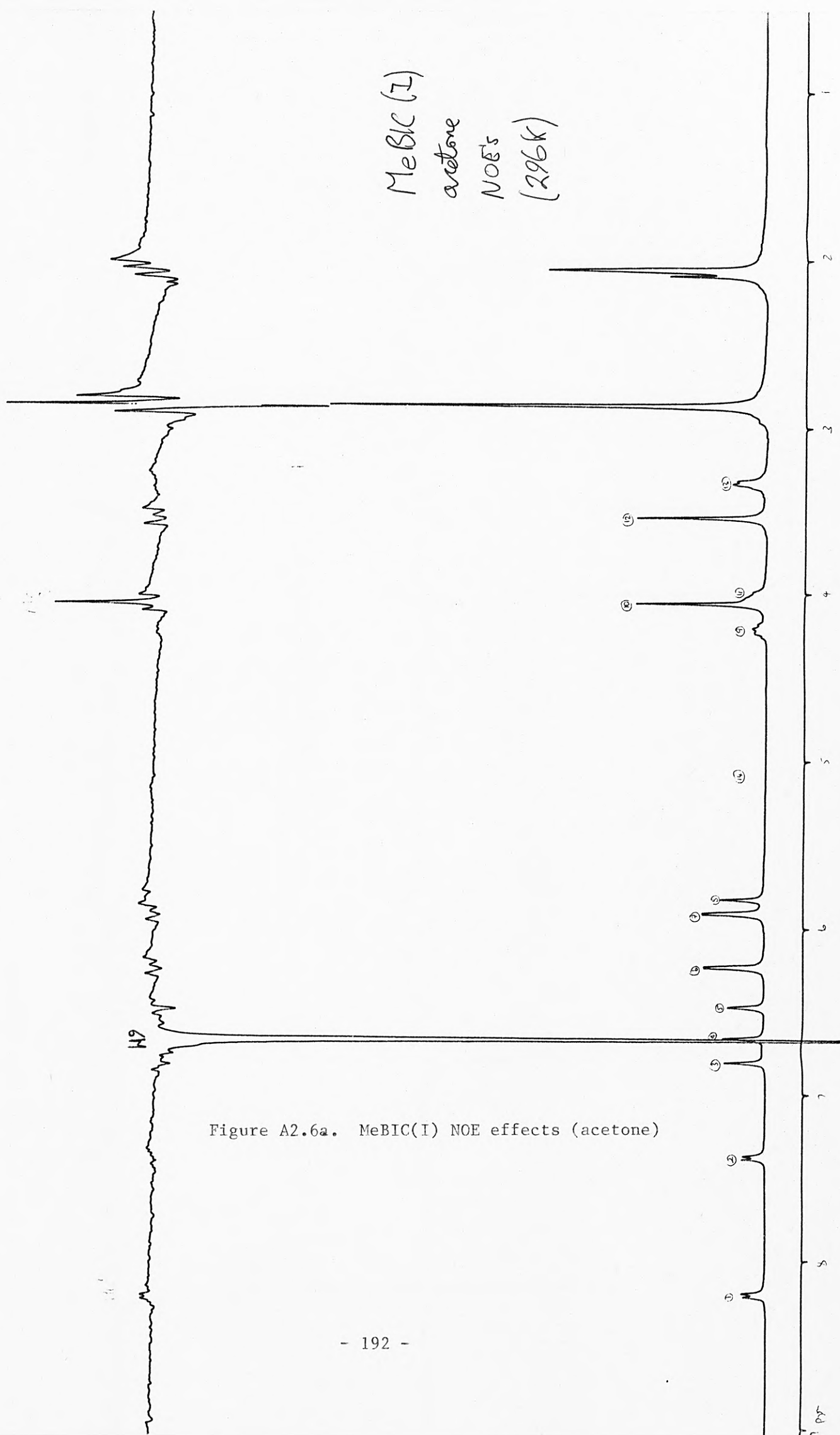
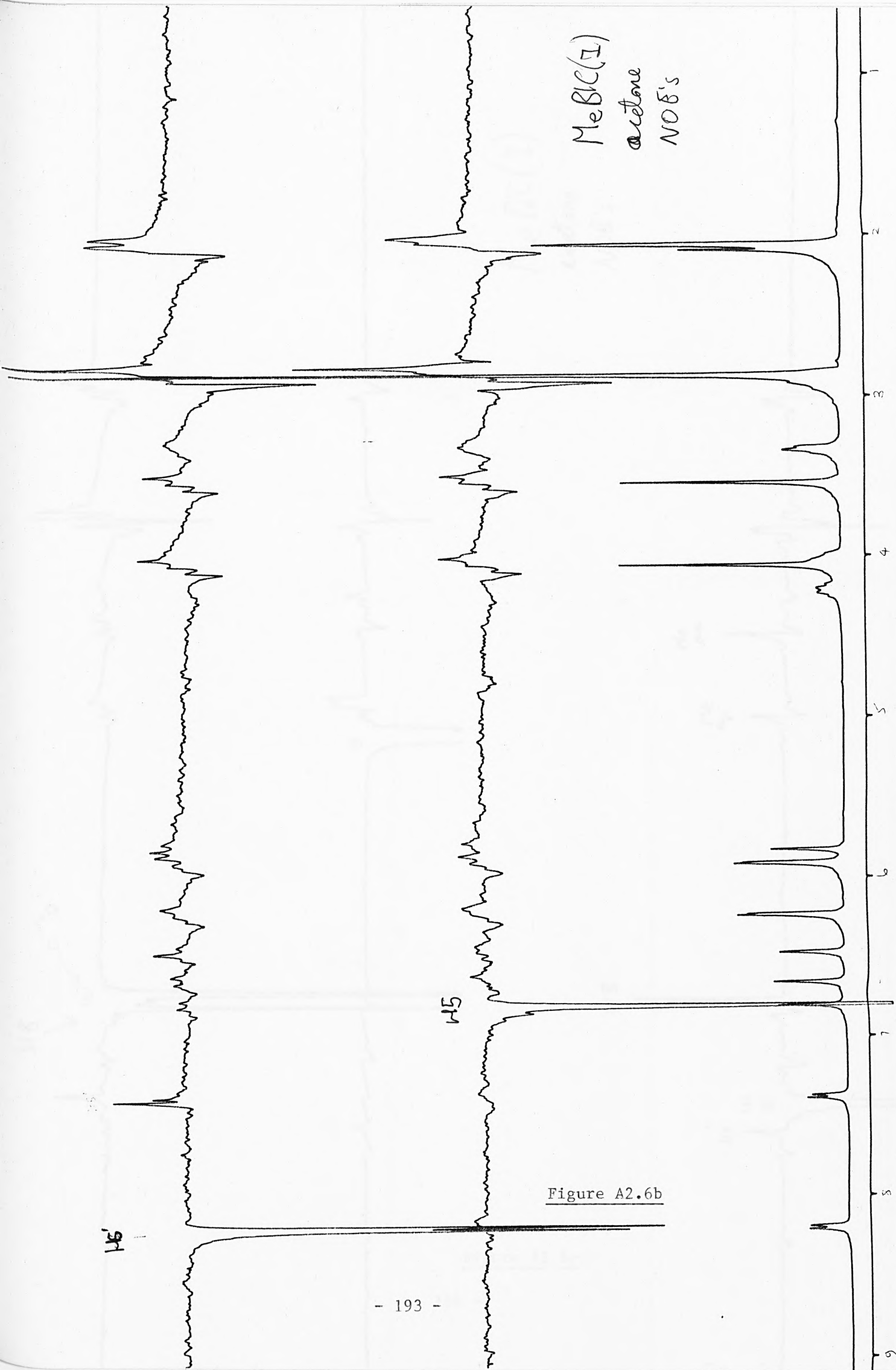


Figure A2.6a. MeBIC(I) NOE effects (acetone)





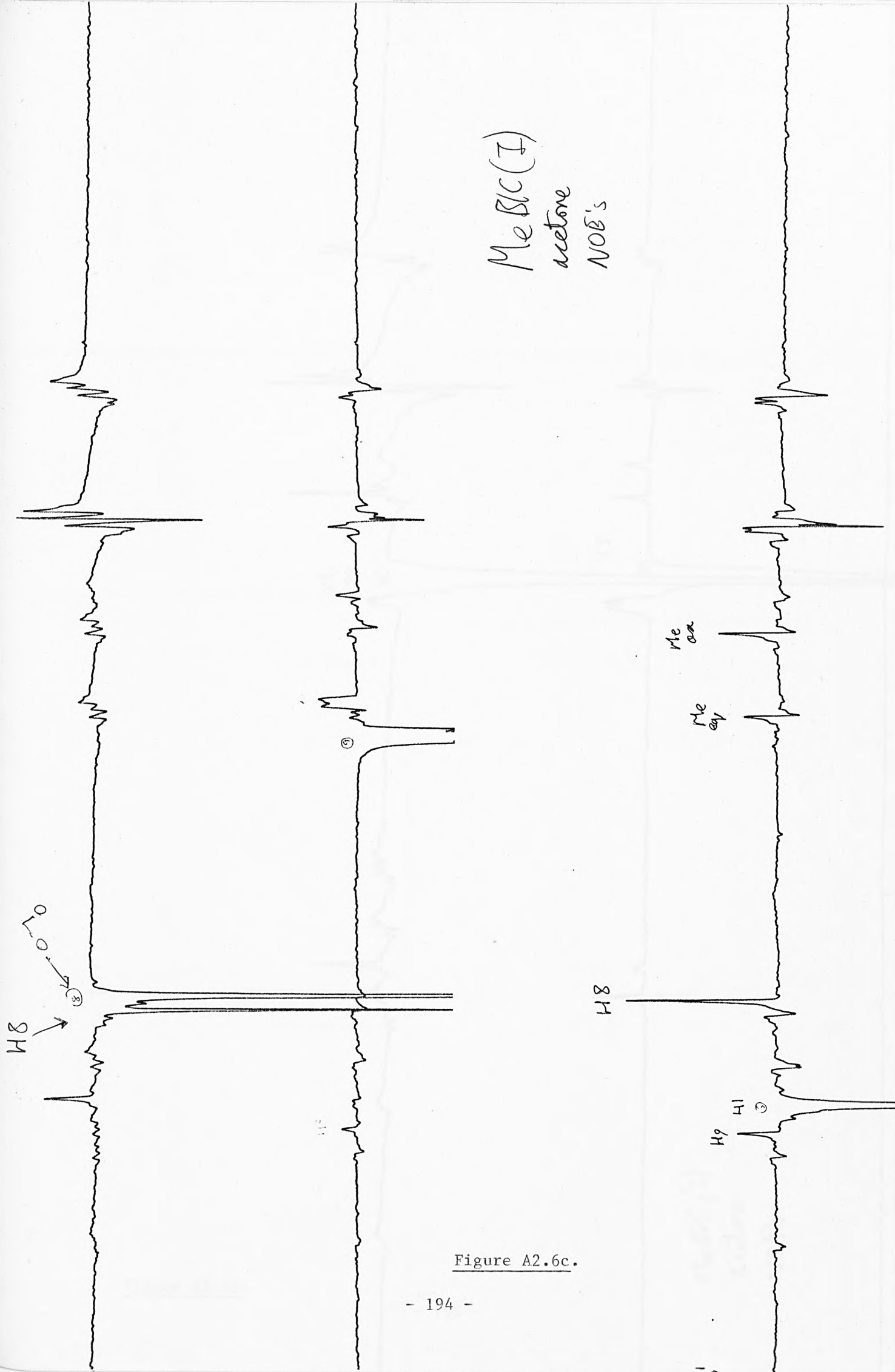


Figure A2.6c.

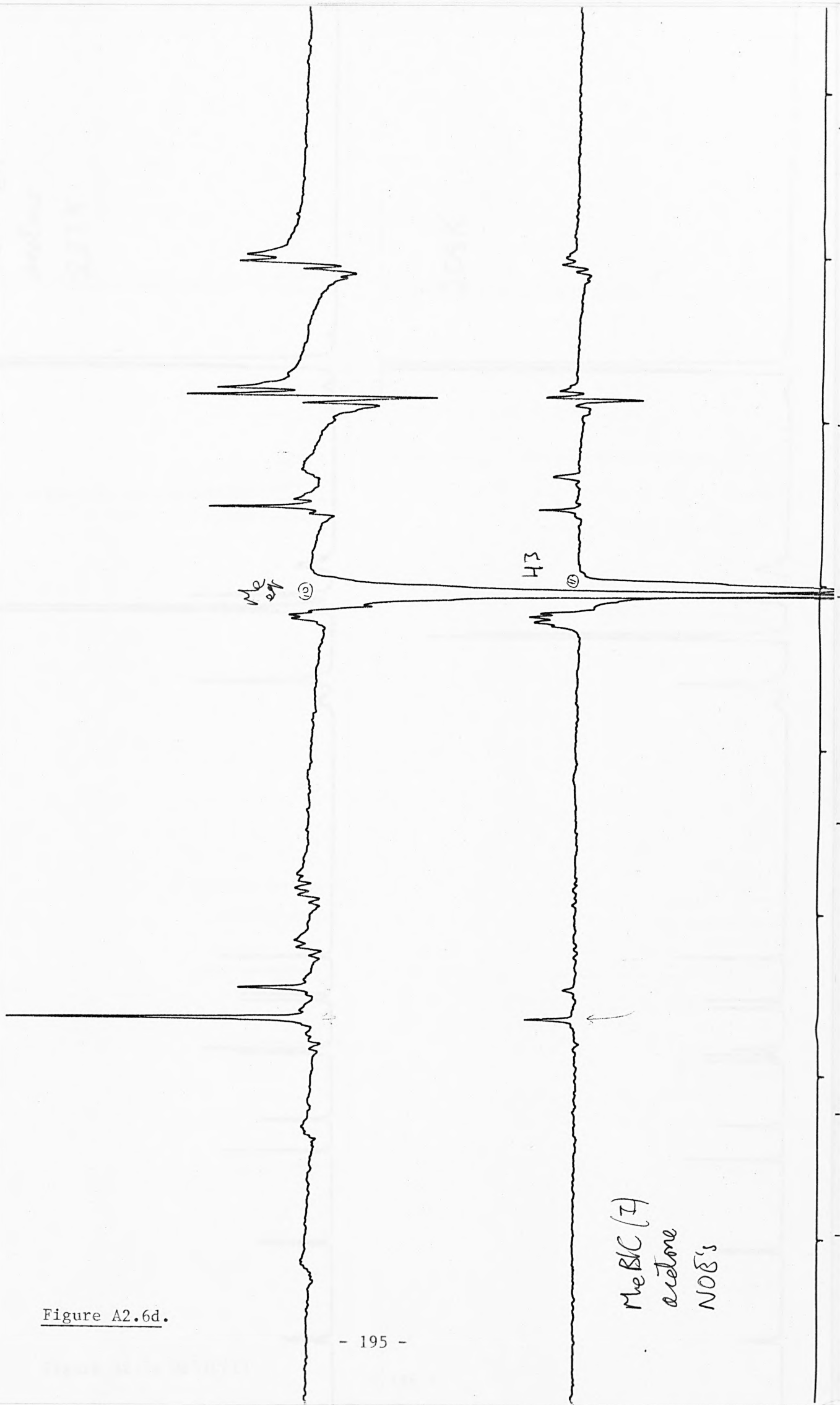


Figure A2.6d.

MeBIC (I)  
acetone  
223K

208K

Figure A2.7a MeBIC(I)



MeBIC(I)

acetone

248 K

273 K

Figure A2.7b MeBIC(I)

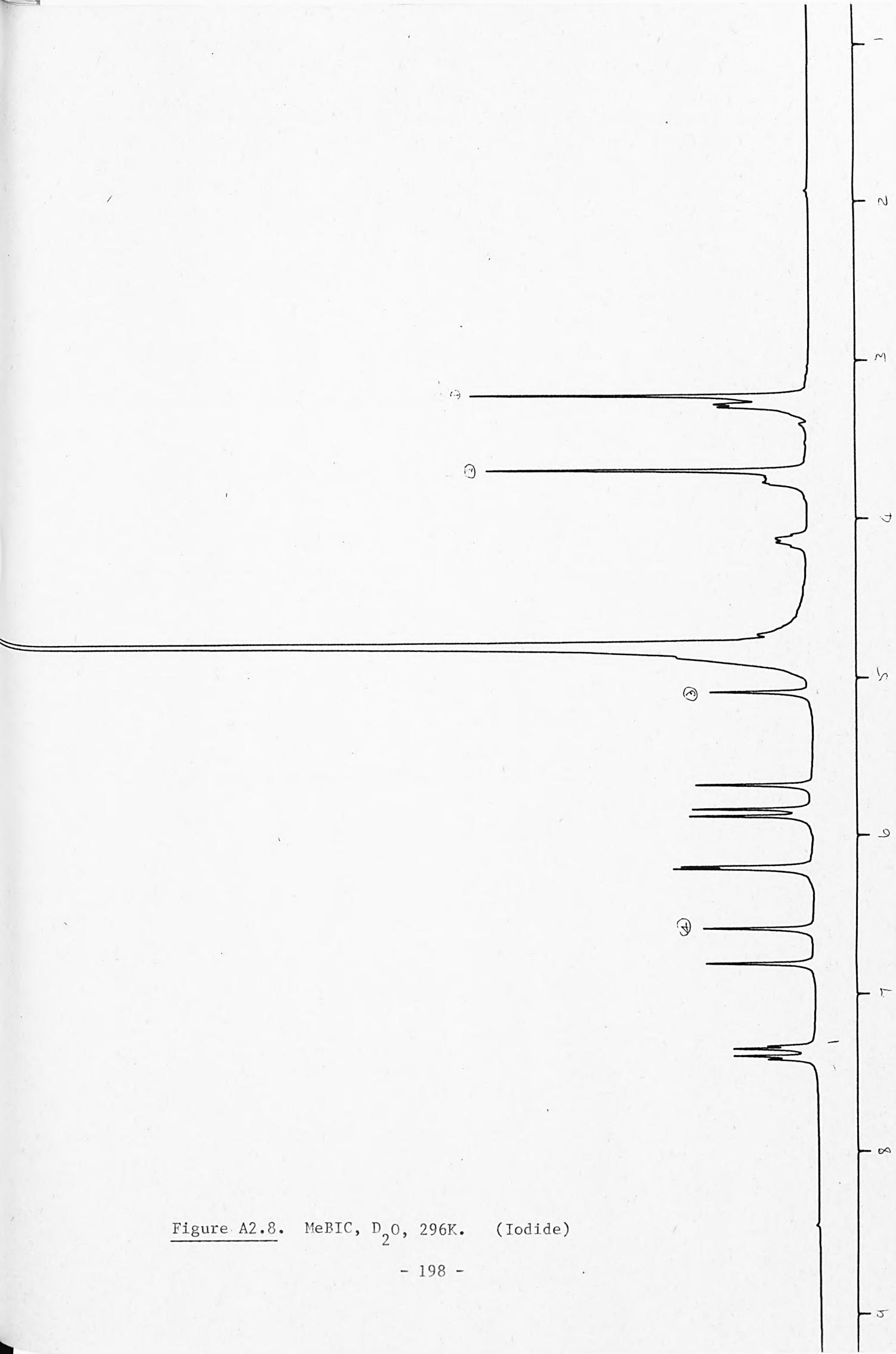


Figure A2.8. MeBIC, D<sub>2</sub>O, 296K. (Iodide)

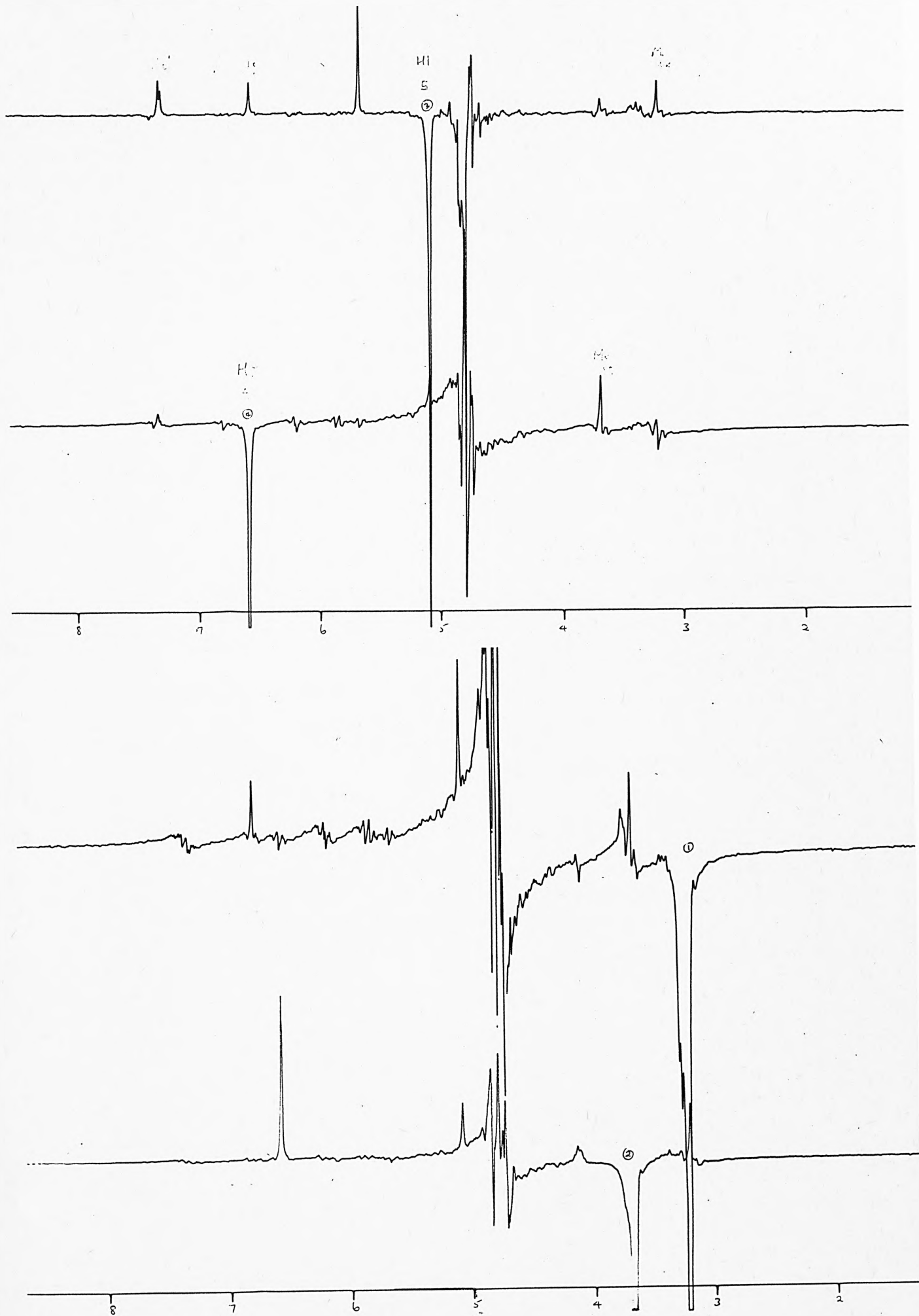


Figure A2.9. MeBIC(I) NOE effects ( $D_2O$ ).

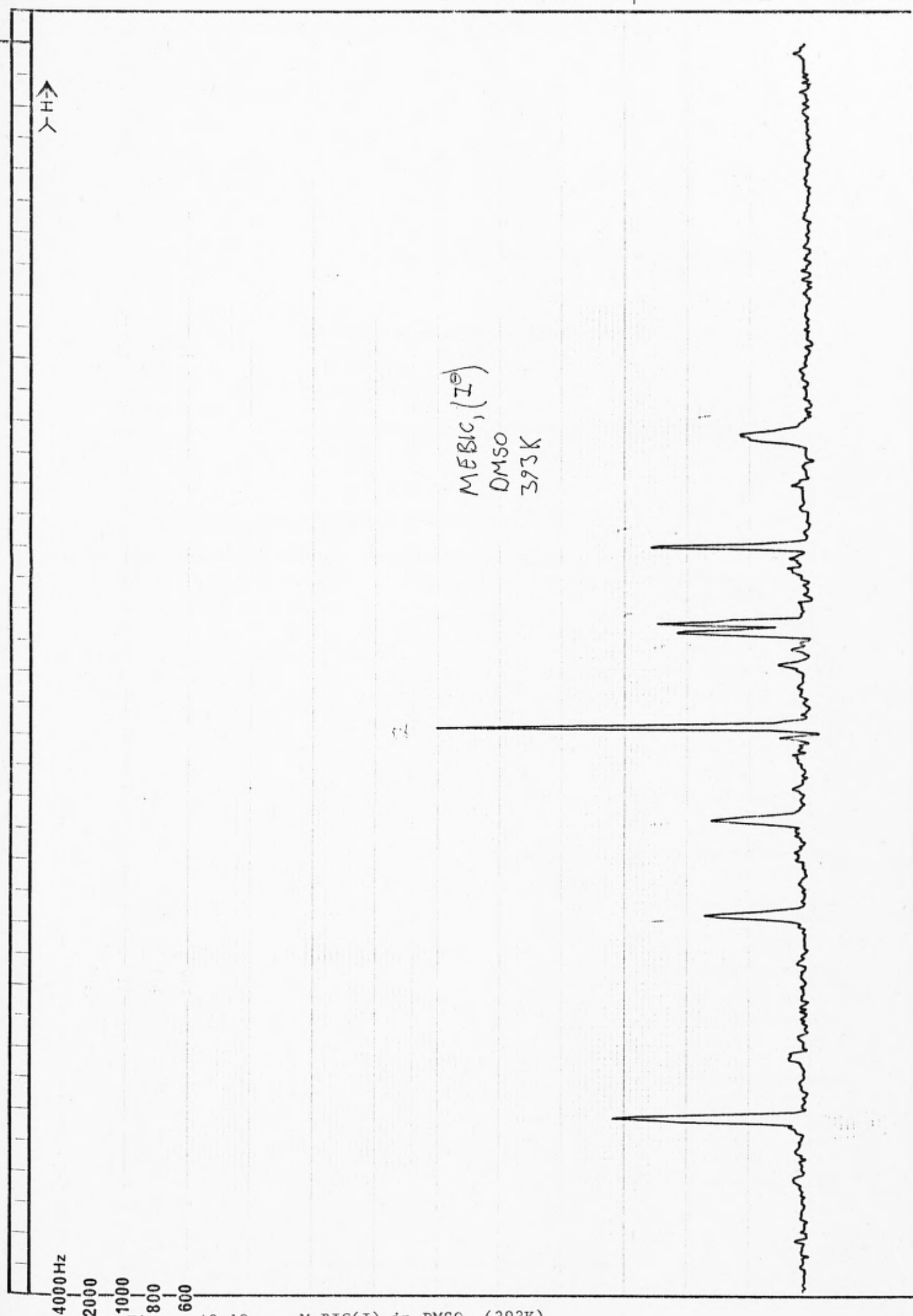


Figure A2.10a. MeBIC(I) in DMSO (393K)



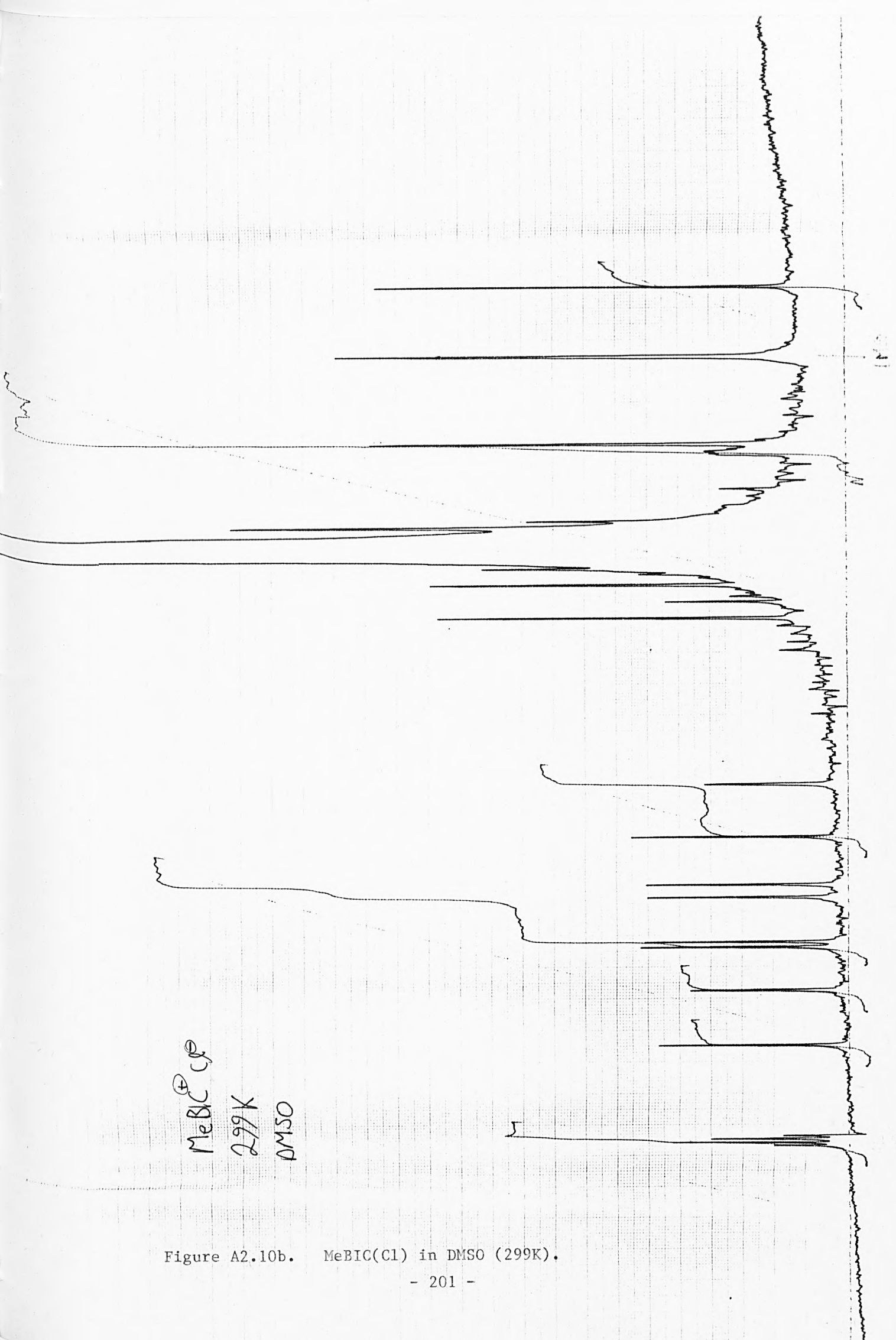


Figure A2.10b. MeBIC(Cl) in DMSO (299K).

267K (20% D<sub>2</sub>O)

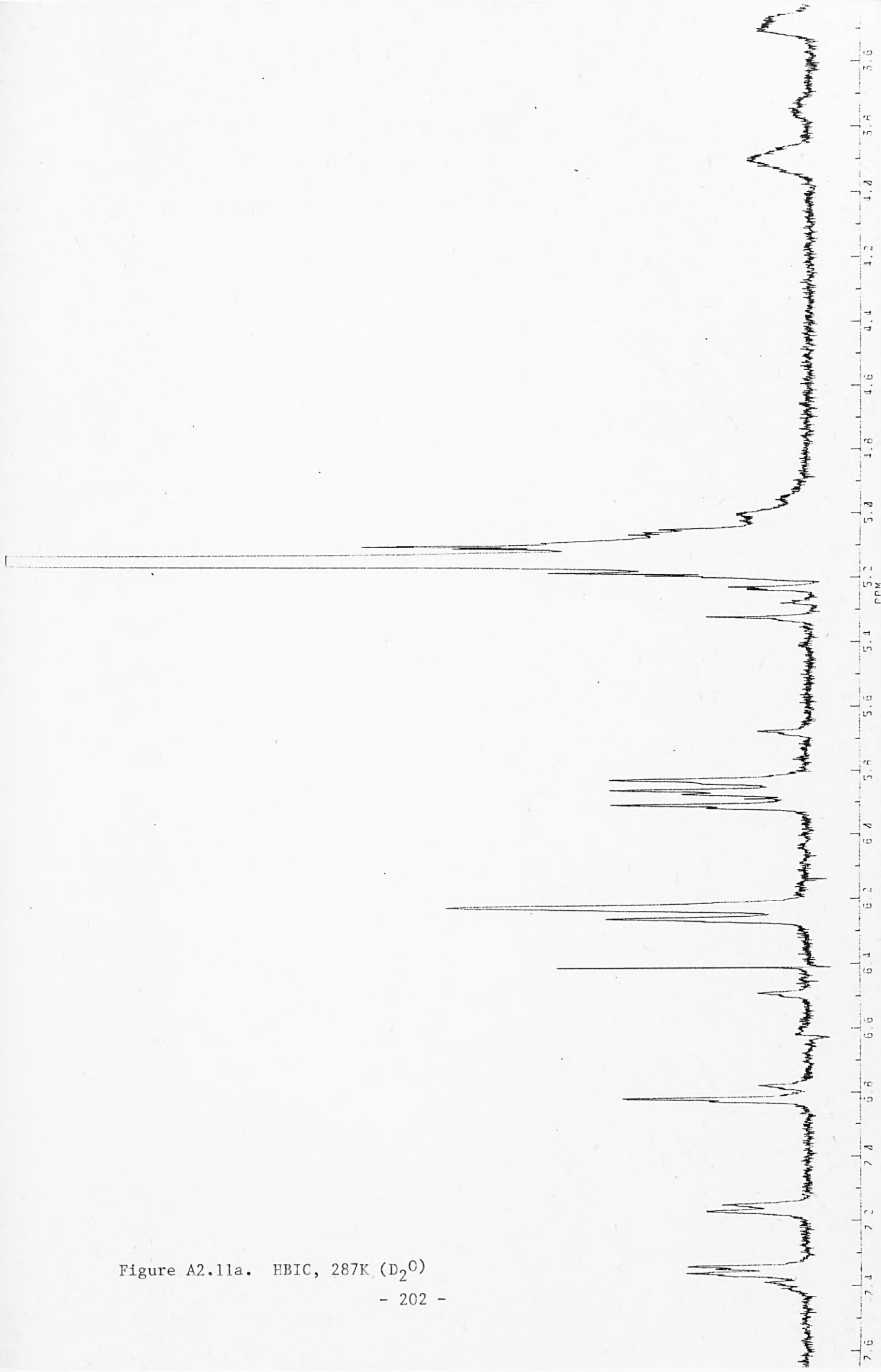


Figure A2.11a. HBIC, 287K (D<sub>2</sub>O)

JUL 24 001  
DATE 24-7-86

|       |          |
|-------|----------|
| SF    | 400.130  |
| SY    | 0.0      |
| 01    | 7600.000 |
| SI    | 32768    |
| TD    | 32768    |
| SW    | 4807.692 |
| HZ/PT | .293     |

| PW   | RD   | AG    | RG | NS  | TE  |
|------|------|-------|----|-----|-----|
| 4.00 | 0.00 | 3.408 | 10 | 652 | 297 |

FW 02 DP  
6100 6633.000  
7H P0

|        |         |
|--------|---------|
| LB     | 100     |
| RGB    | 300     |
| CCX    | 41.00   |
| CY     | 0.0     |
| F1     | 8.0000P |
| F2     | 8.0000P |
| Hz/CM  | 80.023  |
| PPM/CM | 200     |
| SR     | 5305.61 |

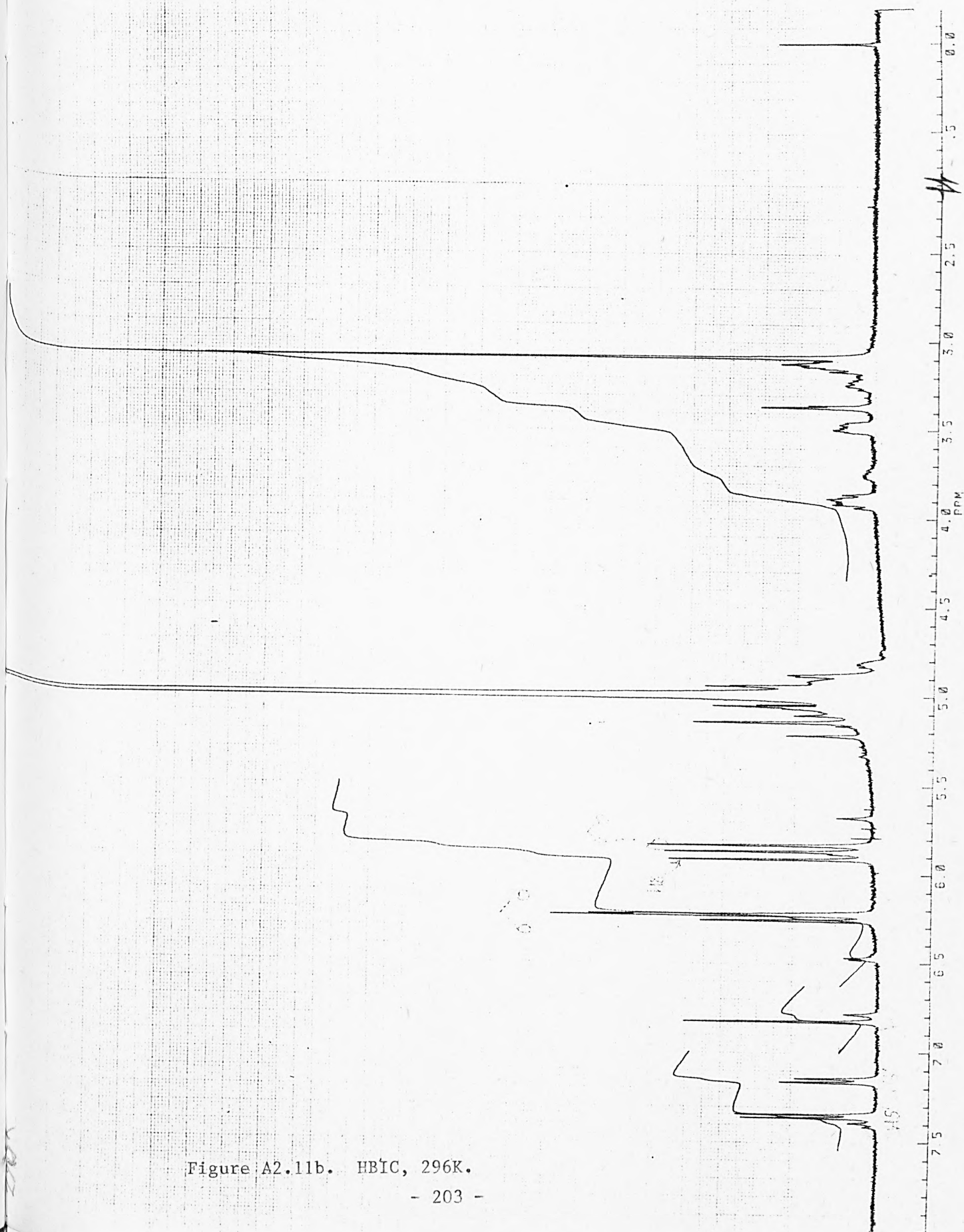
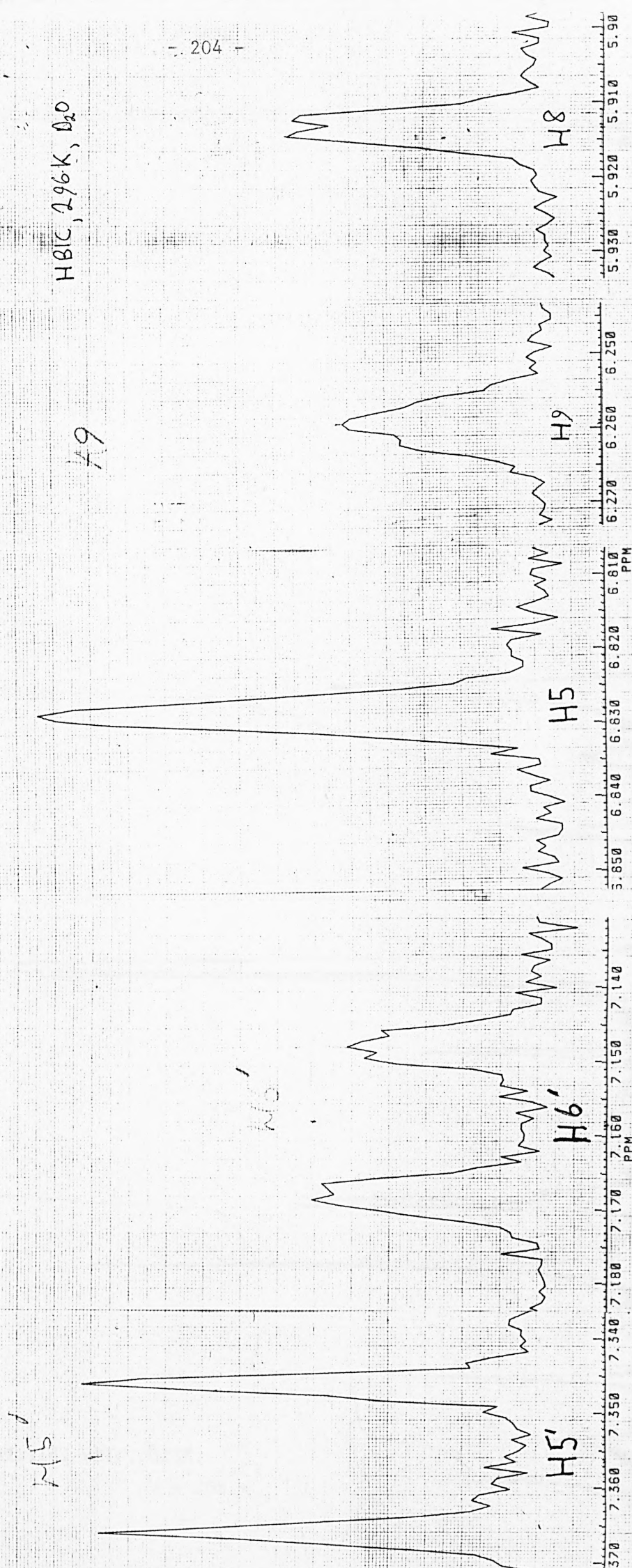


Figure A2.11b. HBIC, 296K.

Figure A2.11c. HBIC, 296K (expanded)





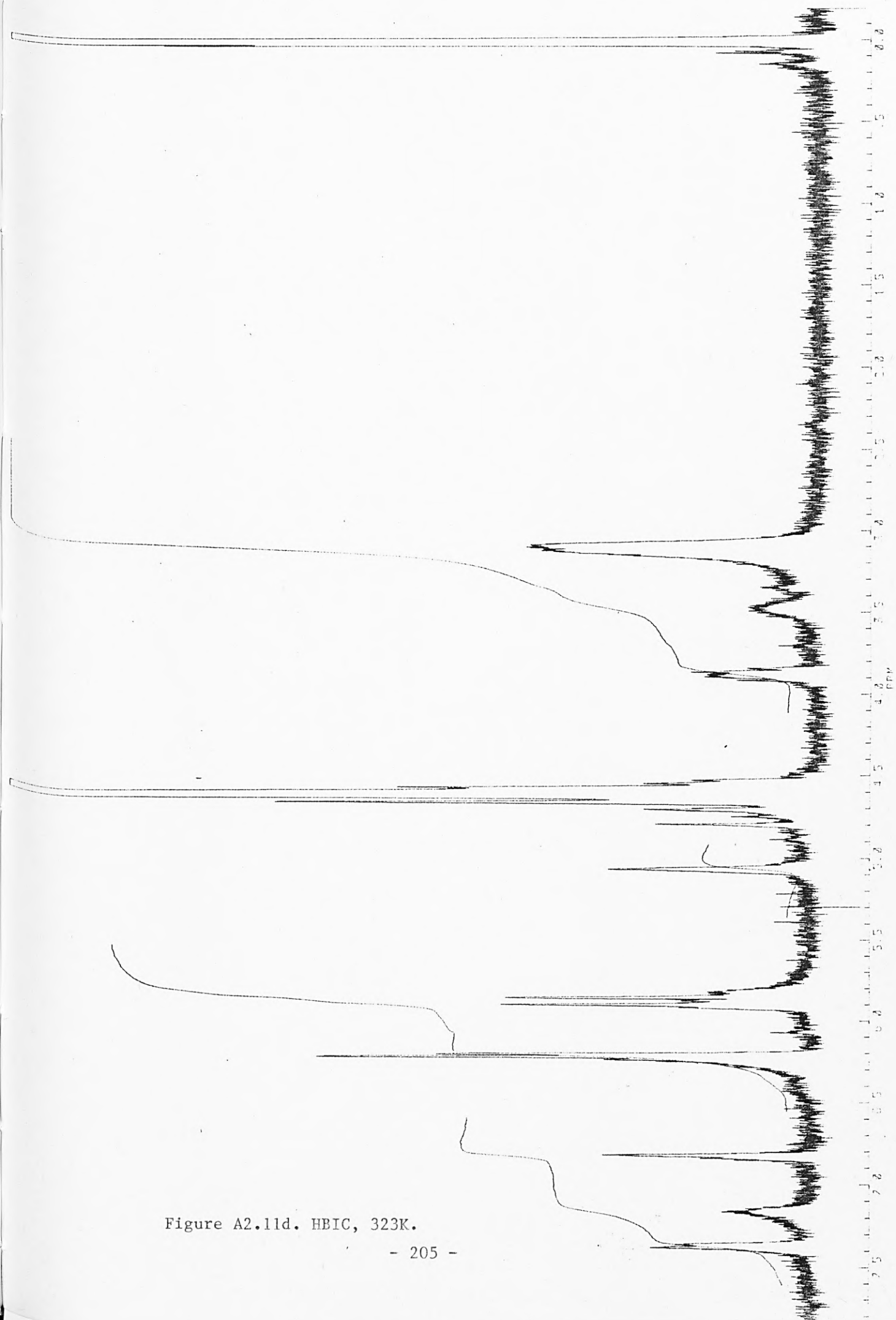


Figure A2.11d. HBIC, 323K.

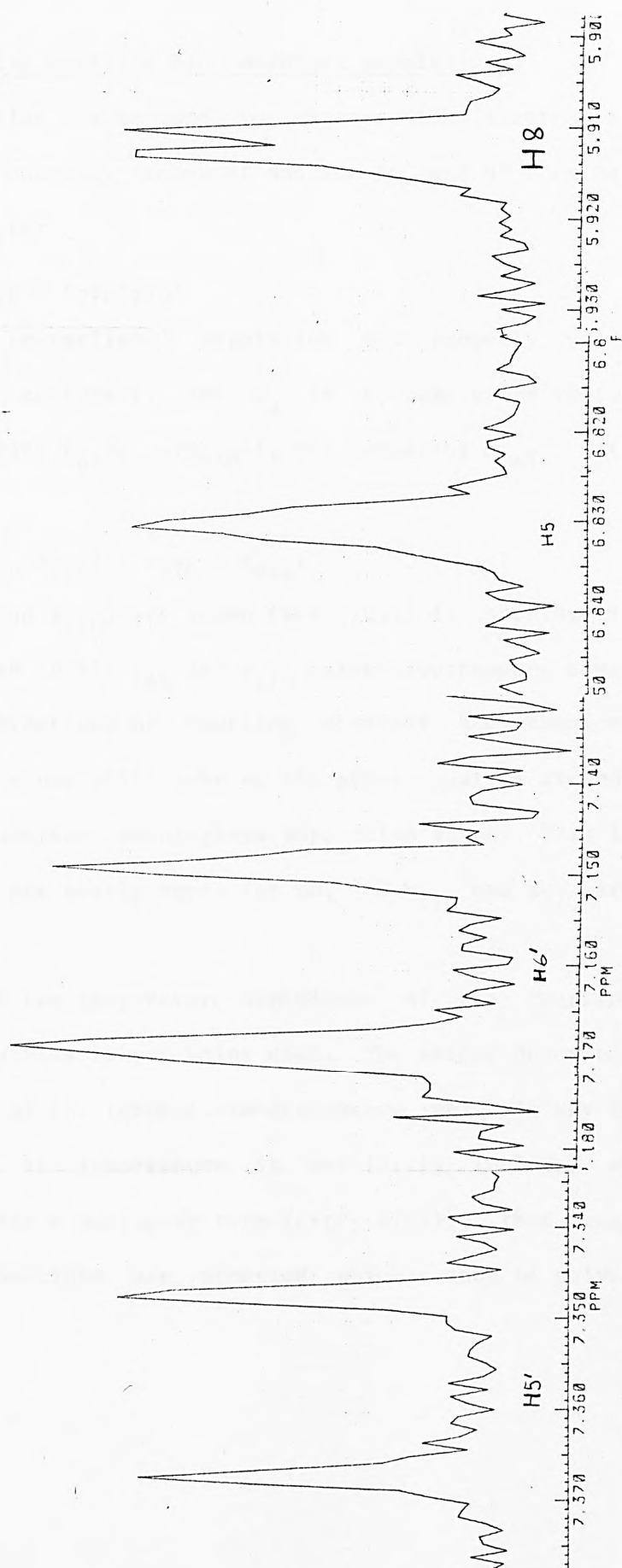


Figure A2.11e. HBIC in D<sub>2</sub>O, decoupled at H<sub>9</sub> (6.26ppm).

### Appendix A3.

#### Simultaneous equations for deriving BIC conformer populations.

The following equation can be used to express the theoretical value for the H1/H9 coupling constant and the H6' and H8 chemical shifts at each temperature:

$$x_{\text{calc}} = p_{45}x_{45} + p_{170}x_{170} + p_{270}x_{270}.$$

Where  $p_i$  and  $x_i$  are the (normalised) population and property value respectively for each conformer, and  $x_0$  is a base value (0 for coupling constants). Using  $p_{45}+p_{170}+p_{270}=1$ , and comparing  $x_{\text{calc}}$  with  $x_{\text{obs}}$  gives:

$$p_{45}(x_{45}-x_{170}) + p_{170}(x_{170}-x_{270}) + x_{270} = x_{\text{obs}}.$$

Now, since  $(x_{45}, x_{170}$  and  $x_{270})$  are known (see 4.2.1) it should, in principle, be possible to obtain  $p_{45}$  and  $p_{170}$  using simultaneous equations for different combinations of coupling constant and chemical shift. However, the values which make up the above equation are not exact and tend to give somewhat meaningless population values. This is because  $x_{45}$  and  $x_{170}$  are nearly equal for H8, and  $x_{170}$  and  $x_{270}$  are nearly equal for H6'.

The problem is that the temperature-dependence of the coupling constants and chemical shifts is not being used. The method described in 4.2.1 makes full use of the temperature-dependence, which is why it was used. Note that if temperature is explicitly included by replacing the  $p_i$  term with a Boltzmann term ( $\exp(-G/RT)$ ), then non-linear simultaneous equations are produced which cannot be solved analytically.

#### Appendix A4. Major changes at ULCC during the course of this work.

These included the withdrawal (31.1.84) of the CDC computers on which most of our programmes were based. This necessitated the conversion of these programmes to work either on the Cray or Amdahl computers which had been introduced earlier. Also, for more than a year the programmes MNDO and MM2 were not available as they were undergoing conversion.

In comparing computation times the Cray is ca 5 times faster than the CDC7600, which is roughly as fast as the Amdahl. (And our data processing programmes (see A6.5) associated with these programmes could not be converted until MNDO and MM2 were available.)



**Appendix 5: The University of North Carolina list of known Molecular Computer Graphics Installations: Equipment Summary, Contact persons, and Addresses (pp. 209- 213) has been removed for copyright reasons**

Appendix A6      Programmes written during the course of this work.

A6.1 MNDODP

-a pre and post processor for direct interface between MNDO and SOLVEFF. This routine enabled GABA calculations with 2 rotations to be performed in one run. (The MNDO programme itself was only designed for one rotation).

```
JOB,US=GPAP571,JN=MNDODP,T=160,MFL=150000.
CFT,OFF=CT,L=0.
ASSIGN,DN=MNDGIN,A=FT05.
ASSIGN,DN=MNDGOT,A=FT06.
ASSIGN,DN=DPIN,A=FT15.
ASSIGN,DN=COOR,A=FT18.
ASSIGN,DN=DPOUT,A=FT16.
COPYD,I= IN,O=INPT.
REWIND,DN=INPT.
COPYF,I=INPT,O=COOR.
COPYD,I=INPT,O=DPIN.
ASSIGN,DN=SOLVIN,A=FT20.
ACCESS,DN=MNDO,ID=KZCCHEM.
REWIND,DN=COOR:DPIN.
COPYF,I=COOR,O=MNDGIN.
REWIND,DN=MNDGIN.
MNDO.
LDR.
MNDO.
LDR.
MNDO.
LDR.
EXIT,U.
REWIND,DN=DPOUT.
COPYF,I=DPOUT,O= OUT.
SKIPR,DN=MNDGOT,NR=-120.
COPYF,I=MNDGOT,O= OUT.
SAVE,DN=SOLVIN,ID=CPAP571.
/EOF
      PROGRAM MNDODP(TAPE15,TAPE16,TAPE6,TAPE18,TAPE5,TAPE20,TAPE21)
      CHARACTER*8 A(10),B(16),T
      REAL MX,MY,MZ
C     MNDO PREPROCESSOR - GABA
      REWIND 18
      REWIND 5
      READ(15,2) U
      2 FORMAT(F10.3)
C     IF NO MORE DATA POSTPROCESS ONLY
      IF(U.EQ.999) GOTO 99
      READ(18,1)A
      WRITE(5,1)A
      READ(18,1)(A(J),J=1,5)
      WRITE(5,4)(A(J),J=1,5),-U
      4 FORMAT(5A8,' T2= ',F5.0)
```

```

C      DO 10 I=2,6
        READ(18,1) A
        1  FORMAT(10A8)
        10 WRITE(5,1) A
C
        READ(18,*) I1,R,I2,S,I3,DUM,I4,I5,I6,I7
        WRITE(5,3) I1,R,I2,S,I3,U,I4,I5,I6,I7
        3  FORMAT(3(I2,1X,F7.2,1X),1X,4I4)
C
        20 READ(18,1,END=21) A
        WRITE(5,1) A
        GOTO 20
        21 CONTINUE
        REWIND 5
C
        99 CONTINUE
C      POSTPROCESS MNDO OUTPUT (TAPE6)
        REWIND 6
C
        100 READ(6,11,END=1000) T
        11  FORMAT(5X,A8)
        IF(T.NE.'SUM      ') GOTO 100
        BACKSPACE 6
        READ(6,12) MX,MY,MZ,DIP
        12  FORMAT(17X,4(F12.5))
C
        110 READ(6,13) T
        13  FORMAT(6X,A8)
        IF(T.NE.'EAT OF F') GOTO 110
        BACKSPACE 6
        READ(6,14) EN
        14  FORMAT(30X,F12.6)
C
        120 READ(6,15) T
        15  FORMAT(1X,A8)
        IF(T.NE.'      4 ') GOTO 120
        BACKSPACE 6
        READ(6,16) T3,T2,T4,T1
        16  FORMAT(44X,F9.4,/,44X,F9.4,/,44X,F9.4,/,44X,F9.4)
        T1=-T1
        T2=-T2
        T3=-T3
        T4=-T4
C      WRITE OUT ENERGIES AND DPOLES TO TAPE20 (FOR SOLV. INPUT) AND TO OUTP
        WRITE(20) T2,T3,EN
        WRITE(20) MX,MY,MZ
        WRITE(16,17) T2,T3,T4,EN,MX,MY,MZ,DIP,T1
        17  FORMAT(1X,'T2=',F6.2,2X,'T3=',F6.2,2X,'T4=',F6.2,2X,'EN=',
          \ F12.6,4X,'DIPOLE:',4F8.4,8X,'T1=',F9.3)
        WRITE(21,19) EN
        19  FORMAT(/,' HEAT OF FORMATION=',F9.4)
C      DO 130 I=1,7
C      130 BACKSPACE 6
C      DO 140 I=1,19
C      READ(6,18) B
C      18  FORMAT(13A10)
C      WRITE(21,18) B

```

```

C 140 CONTINUE
      GOTO 100
1000 CONTINUE
      REWIND 6
      STOP
      END
/EOF
molecular coordinates etc.
/EOF

```

A similar programme (CNDOP) was written for shortening the amount of output from the QM programme CNIND074 (QCPE 281). These CNDO routines are no longer used, however, since CNDO is now available within the MNDO programme.

#### A6.2 MM2PRE, MM2POST and FXY.

MM2PRE simplifies the input required for the MM2 programme on the Amdahl. The advantage of this programme is that coordinates produced by any programme (eg graphics or QM) can be added directly to the MM2 input data. This is not possible with normal MM2 input because of the unusual format required. (Note that molecular connectivity needs to be input with this programme.) MM2POST was used for processing the MM2 output for BIC and MeBIC, the processed data then being used by FXY to give reference coordinates for benzene ring shielding contributions (4.2.1).

```

C  MM2 input programme
      CHARACTER*4 TITLE(15),IC
      DIMENSION X(3,100),ITYPE(100),Q(100),ICON(20,16),
      * JAT(80),KAT(80),M(4),F(3)
C1*
      IC='C1 '
C  TITLE & NATS etc
      READ(5,110)TITLE
110  FORMAT(15A4)
      READ(5,*,END=1000)NATS,IPR,NRST,INIT,NCONST,TMAX
      WRITE(8,1)TITLE,NATS,IPR,NRST,INIT,NCONST,TMAX
1  FORMAT(15A4,I5,I2,I3,I2,I3,F5.2)
      WRITE(6,99)(TITLE(I),I=1,12)
99  FORMAT(' MM2 INPUT PROG.  FILE TITLE: ',12A4)
C
C2*
      IC='C2* '

```



```

      READ(5,*,END=1000)NCON,NBUT,NATCH,NSYM,NX,NROT,LAB,NDC,NCALC,
*   IHFM,MVDW,NDRIVE
      WRITE(8,2)NCON,NBUT,NATCH,NSYM,NX,NROT,LAB,NDC,NCALC,
*   IHFM,MVDW,NDRIVE
2   FORMAT(I5,15X,10I5,5X,I5)

C
C2B*
      IC='C2B*'
      IF(NDC.EQ.3)THEN
        READ(5,*,END=1000)(Q(I),I=1,NATS)
        WRITE(8,22)(Q(I),I=1,NATS)
22   FORMAT(8F10.5)
      ENDIF

C
C3*
      IC='C3*'
C   REMEMBER TO ADD EXTRA COMMAS AT END OF LINES
      DO 30 I=1,NCON
        READ(5,*,END=1000)(ICON(I,J),J=1,10)
30   WRITE(8,3)(ICON(I,J),J=1,10)
      3   FORMAT(16I5)

C
C4*
      IC='C4*'
      READ(5,*,END=1000)(JAT(I),KAT(I),I=1,NATCH)
      WRITE(8,4)(JAT(I),KAT(I),I=1,NATCH)
4   FORMAT(16I5)

C
C5*
      IC='C5*'
      DO 20 I=1,NATS
20   READ(5,*,END=1000)(X(J,I),J=1,3)
      READ(5,*,END=1000)(ITYPE(I),I=1,NATS)
      WRITE(8,5)((X(J,I),J=1,3),ITYPE(I),I=1,NATS)
5   FORMAT(2(3F10.5,I5,5X))

C
C9*
      IC='C9*'
C   REDEFINED CONSTANTS
      READ(5,*,END=1000)NT,NS,NV,NB,DIEL
      WRITE(8,9)NT,NS,NV,NB,DIEL
9   FORMAT(4I5,20X,F5.2)

C
      IF(NT.NE.0)THEN
        DO 40 I=1,NT
          READ(5,*,END=1000)M,F
40   WRITE(8,91)M,F
91   FORMAT(4I5,3F10.5)
      ENDIF

C
      IF(NS.NE.0)THEN
        DO 50 I=1,NS
          READ(5,*,END=1000)I1,K,F
50   WRITE(8,92)I1,K,F
92   FORMAT(2I5,3F10.5)
      ENDIF

C
      IF(NV.NE.0)THEN

```

```

        DO 55 I=1,NV
        READ(5,*,END=1000)I1,I2,F
55     CONTINUE
        ENDIF
C
        IF(NB.NE.0)THEN
            DO 60 I=1,NB
                READ(5,*,END=1000)I1,I2,I3,B,T,K
                WRITE(8,94)I1,I2,I3,B,T,K
94         FORMAT(3I5,5X,2F10.5,I5)
60     CONTINUE
        ENDIF
C
C10*
        IC='C10*'
        IF(NDRIVE.EQ.0)STOP
        READ(5,*,END=1000)M,F
        WRITE(8,10)M,F
10     FORMAT(4I5,5X,3F5.1)
        STOP
1000  WRITE(6,1010)IC
1010  FORMAT(' End of file.      TXT=',A4)
        STOP

C    MM2POST - MM2 OUTPUT EDITING PROGRAM - SINGLE DRIVER
        CHARACTER*8 IN
        REWIND 12

C    SET UP CONSTANTS
        I=-1
        IO=0
        I1=1
        I2=2
        I3=3
        I4=4
        I5=5
        I6=6
        I7=7
        I9=9
        I99=99
        R=1.39
        A90=90.

C    INITIALISE MNDO INPUT
        WRITE(5,99) I
99     FORMAT(I2)
C
        10 READ(12,101,END=100) IN
101    FORMAT(9X,A8)
        IF(IN.NE.'1)- C( 2') GOTO 10
        BACKSPACE 6
        READ(12,102)R12,R114,R119,R29,R218
102    FORMAT(3(19X,F11.5,/),/,19X,F11.5,/,19X,F11.5,16(/))
        READ(12,103)R811,R838,R910,R1028,R1113,R1127,R1314,R1429,R1830
103    FORMAT(3(19X,F11.5,/),/,2(19X,F11.5,/),19X,F11.5,///,19X,F11.5,
        \//,19X,F11.5,7(/),19X,F11.5)
C

```

```

CALL A(8H1)- C(14,A214)
CALL A(8H1)- H(19,A219)
CALL A(8H1)- H(19,A1419)
CALL A(8H2)- C( 9,A19)
CALL A(8H2)- C(18,A118)
CALL A(8H8)- H(38,A1138)
CALL A(8H8)- H(38,A1138)
CALL A(8H9)- C(10,A210)
CALL A(8H0)- H(28,A928)
CALL A(8H1)- C(13,A813)
CALL A(8H1)- H(27,A1327)
CALL A(8H1)- H(27,A1327)
CALL A(8H3)- C(14,A1114)
CALL A(8H3)- C(17,A1417)
CALL A(8H3)- C(17,A1417)
CALL A(8H4)- C(13,A113)
CALL A(8H4)- H(29,A129)
CALL A(8H4)- H(29,A1329)
CALL A(8H8)- H(30,A230)

```

C

```

CALL T(8H 9) C(10,T110)
CALL T(8H18) H(30,T130)
CALL T(8H13) C(11,T111)
CALL T(8H13) C(17,T117)
CALL T(8H14) C(13,T213)
CALL T(8H10) H(28,T228)
CALL T(8H13) C(14,T814)
CALL T(8H13) C(14,T814)
CALL T(8H 1) C(14,T914)
CALL T(8H14) H(29,T1129)
CALL T(8H 8) H(38,T1338)
CALL T(8H 1) H(19,T1319)
CALL T(8H 2) C(18,T1418)
CALL T(8H11) H(27,T1427)
CALL T(8H14) H(29,TH)
CALL T(8H14) H(29,TH)
CALL T(8H14) H(29,TH)

```

```

40 READ(12,107) IN
107 FORMAT(17X,A10)
   IF(IN.NE.10HC ENERGY I) GOTO 40
   BACKSPACE 6
   READ(12,109) EN
109 FORMAT(30X,F10.5)

```

C

```

WRITE(36,*) TH,T914,EN

```

C

C PREPARE MNDO INPUT: (1)FR1B

```

WRITE(5,202) IO,I99
202 FORMAT(I2,/,I2)
WRITE(5,*)I6,R,IO
A1417=A1417+60.
WRITE(5,*)I6,R1314,I1,A1417,IO
WRITE(5,*)I6,R114,IO,A113,IO,T117,IO,I3,I2,I1
WRITE(5,*)I6,R12,IO,A214,IO,T213,IO,I4,I3,I2
WRITE(5,*)I6,R29,IO,A19,IO,T914,IO,I5,I4,I3
WRITE(5,*)I6,R910,IO,A210,IO,T110,IO,I6,I5,I4
WRITE(5,*)I1,R1028,IO,A928,IO,T228,IO,I7,I6,I5
WRITE(5,*)I6,R218,IO,A118,IO,T1418,IO,I5,I4,I3

```

```

WRITE(5,*)I1,R1830,I0,A230,I0,T130,I0,I9,I5,I4
WRITE(5,*)I1,R119,I0,A1419,I0,T1319,I0,I4,I3,I2
WRITE(5,*)I0,I0,I0,I0,I0,I0,I0,I0,I0,I0
C
(2)FR2B
C
WRITE(5,202) I0,I99
R12=R12+1.39
WRITE(5,*)I6,R12,I0
WRITE(5,*)I99,R12,I0,A90,I0
WRITE(5,*)I6,R114,I0,A214,I1,T914,I0,I2,I1,I3
WRITE(5,*)I6,R1314,I0,A113,I0,T213,I0,I4,I2,I1
WRITE(5,*)I6,R1113,I0,A1114,I0,T111,I0,I5,I4,I2
WRITE(5,*)I6,R811,I0,A813,I0,T814,I0,I6,I5,I4
WRITE(5,*)I1,R838,I0,A1138,I0,T1338,I0,I7,I6,I5
WRITE(5,*)I1,R1127,I0,A1327,I0,T1427,I0,I6,I5,I4
WRITE(5,*)I1,R1429,I0,A1329,I0,T1129,I0,I4,I5,I6
WRITE(5,*)I0,I0,I0,I0,I0,I0,I0,I0,I0,I0
C
GOTO 10
C
100 CONTINUE
C FINALISE MNDO INPUT
WRITE(5,111)I99
111 FORMAT(I2)
REWIND 5
STOP
SUBROUTINE A(H,X)
CHARACTER*8 T,H
10 READ(12,1) T
IF(T.NE.H) GOTO 10
BACKSPACE 6
READ(12,2) X
RETURN
1 FORMAT(16X,A8)
2 FORMAT(25X,F8.4)
END
SUBROUTINE T(H,X)
CHARACTER*8 T1,H
10 READ(12,1) T1
IF(T1.NE.H) GOTO 10
READ(12,2) X
RETURN
1 FORMAT(19X,A8)
2 FORMAT(28X,F8.4)
END
C
C FXY - FIND X,Y AND Z FOR THE RELEVANT PROTONS AND
C PRINT SQRT(X**2+Y**2) AND Z FOR EACH ANGLE
C (IN BENZENE-RING COORDINATES)
CHARACTER*8 T
REWIND 6
REWIND 36
C
10 READ(36,*,END=100) TH,T2,EN
WRITE(13,9) TH,T2,EN
9 FORMAT(/,30X,10H*** THETA=,F9.3,4H ***,10X,3HT2=,F9.3,8X,3HEN=

```



```

        ,F10.4)
C
C   FR1B
C
20 READ(6,2) T
   2 FORMAT(2X,A8)
     IF(T.NE.'        6') GOTO 20
     CALL CALC(2HH5)
     READ(6,3)
   3 FORMAT(1X)
     CALL CALC(2HH8)
     CALL CALC(2HH1)
C
C   FR2B
C
     WRITE(13,3)
30 READ(6,2) T
     IF(T.NE.'10H        5') GOTO 30
     CALL CALC(3HH5')
     CALL CALC(3HH6')
     CALL CALC(2HH9)
     GOTO 10
C
100 CONTINUE
     WRITE(13,4)
   4 FORMAT(' END#')
     REWIND 13
     STOP
     END
     SUBROUTINE CALC(H)
     READ(6,1) X,Y,Z
   1 FORMAT(34X,3(6X,F14.7))
     X=X/1.39
     Y=Y/1.39
     Z=Z/1.39
     FXY=SQRT(X*X+Y*Y)
     WRITE(13,2) X,Y,Z,FX,Y,H
   2 FORMAT(1X,2HX=,F7.4,2X,2HY=,F7.4,10X,2HZ=,F7.4,5X,4HFX,Y=,F7.4,
     \3X,A3)
     RETURN
     END

```

### A6.3 JVIC and DELT.

These programmes were used to calculate average theoretical coupling constants and chemical shifts, which were then used in obtaining BIC conformer energy differences (see 4.2.1). Required input: coupling constant or chemical shift for each of the 3 BIC minima, plus a base shift value for DELT.

```

DIMENSION S(3),T(2),DD(2),DELT(2)
PROGRAM JVIC
DATA STP/4HSTOP/
DATA (T(I),I=1,2)/247,317/
D=0
1 PRINT, "Proton?"
READ(5,9)PROT
9 FORMAT(A4)
IF(PROT.EQ.STP) GOTO 300
PRINT, "J values?"
READ, S
WRITE(6,5)PROT,S
5 FORMAT(1X,A4,3F6.3)
DO 200 IG1=2,12
DG1=IG1*.1
DO 200 IG2=1,14
DG2=IG2*.1
WRITE(6,11)DG1,DG2
11 FORMAT(1X,4HDG1=,F5.2,6H DG2=,F5.2)
DO 100 I=1,2
P180=EXP(-DG1*1000/(2.*T(I)))
P270=EXP(-DG2*1000/(2.*T(I)))
SUM=1+P180+P270
P180=P180/SUM
P270=P270/SUM
DELT(I)=(1-P180-P270)*S(1)+P180*S(2)+P270*S(3)
DD(I)=DELT(I)-D
D=DELT(I)
WRITE(6,20)T(I),DELT(I),DD(I)
20 FORMAT(2H T,F4.0,3H S=,F7.4,3H D=,F7.4)
100 CONTINUE
DIFF=DELT(2)-DELT(1)
WRITE(6,30)DIFF
30 FORMAT(' DIFF= ',F7.4)
200 CONTINUE
GOTO 1
300 WRITE(6,15)
15 FORMAT(4H END)
STOP
END

```

```

PROGRAM DELT
DIMENSION S(3),T(6),DD(6),DELT(6)
DATA STP/4HSTOP/
DATA (T(I),I=1,6)/213,223,233,248,263,296/
D=0
1 PRINT, "Proton?"
READ(5,9)PROT
9 FORMAT(A4)
IF(PROT.EQ.STP) GOTO 300
PRINT, "S values and base shift?"
READ, S,S0
WRITE(6,5)PROT,S,S0
5 FORMAT(1X,A4,3F6.3,/' Base shift= ',F6.3)
DO 200 IG1=2,12
DG1=IG1*.1

```

```

DO 200 IG2=1,14
DG2=IG2*.1
WRITE(6,11)DG1,DG2
11 FORMAT(1X,4H DG1=,F5.2,6H DG2=,F5.2)
DO 100 I=1,6
P180=EXP(-DG1*1000/(2.*T(I)))
P270=EXP(-DG2*1000/(2.*T(I)))
SUM=1+P180+P270
P180=P180/SUM
P270=P270/SUM
DELT(I)=(1-P180-P270)*S(1)+P180*S(2)+P270*S(3)+S0
DD(I)=DELT(I)-D
D=DELT(I)
WRITE(6,20)T(I),DELT(I),DD(I)
20 FORMAT(2H T,F4.0,3H S=,F7.4,3H D=,F7.4)
100 CONTINUE
DIFF=DELT(6)-DELT(1)
WRITE(6,30)DIFF
30 FORMAT(' DIFF=',F7.4)
200 CONTINUE
GOTO 1
300 WRITE(6,15)
15 FORMAT(4H END)
STOP
END

```

#### A6.4 INTERP (versions 1 and 2) and XT.

The first of the two-dimensional interpolation programmes uses a *Newton-Raphson* interpolation method, and was superseded by the second which makes use of a NAG routine (E01ADF) for interpolation with a cubic spline function. We found that a cubic spline function gives better interpolated energies than a higher order bicubic spline function (used by Clarke 1980), since energy differences start to converge beyond 3rd order. The programme XT was used to compute GABA  $x_T$  distributions for interpolated surfaces.

```

10C THIS PROGRAM INTERPOLATES P.E. SURFACES (version 1)
20 INTEGER ANGSA,B,C,D
25 REAL INTI,INTJ
30 COMMON EN(73,73)
40 DIMENSION ANGSA(73)
50C READ IN INDICES : A,B ARE X LIMITS & C,D ARE Y LIMITS
60 READ, A,B,C,D
70C READ IN ENERGIES
80 DO 20 I=A,B,4
90 20 READ,(EN(I,J),J=C,D,4)
100C
110C INTERPOLATE Y'S ALONG THE LINES X=A TO B (STEP 4)

```

```

130 DO 30 I=A,B,4
140 DO 30 J=C+1,D-3,4
150 EN(I,J)=INTJ(I,J,D,C)
160 EN(I,J+1)=INTJ(I,J+1,D,C)
170 EN(I,J+2)=INTJ(I,J+2,D,C)
180 30 CONTINUE
190C
200C INTERPOLATE X'S ALONG THE LINES Y=C TO D
210C --MISSING OUT KNOWN VALUES
230 DO 40 J=C,D
240 DO 40 I=A,B
250 IF(I-1.EQ.((I-1)/4)*4) GO TO 40
260 EN(I,J)=INTI(I,J,B,A)
270 40 CONTINUE
280C
290C PRINT FINAL VALUES
295 IF(D-C.GT.17)CALL MATOUT(A,B,C,D)
300 DO 50 J=C,D
310 50 ANGS(J)=(J-1)*5
320 WRITE(6,60)(ANGS(J),J=C,D)
330 60 FORMAT(6X,21I6)
340C
350 DO 70 I=A,B
360 K=(I-1)*5
370 70 WRITE(6,80)K,(EN(I,J),J=C,D)
380 80 FORMAT(1X,I4,3X,21(F5.2,1X))
390 STOP
400 END
410 REAL FUNCTION INTI(I,J,N,C)
415 INTEGER C
420 DIMENSION EN2(73)
430 COMMON EN(73,73)
440C
450 DO 5 K=C,N,4
460 5 EN2(K)=EN(K,J)
470C
480 DO 10 K=C,N-4,4
490 DO 10 L=K+4,N,4
500 10 EN2(L)=((I-K)*EN2(L)-(I-L)*EN2(K))/(L-K)
510 INTI=EN2(N)
520 RETURN
530 END
540 REAL FUNCTION INTJ(I,J,N,C)
545 INTEGER C
550 DIMENSION EN2(73)
560 COMMON EN(73,73)
570C
580 DO 5 K=C,N,4
590 5 EN2(K)=EN(I,K)
600C
610 DO 10 K=C,N-4,4
620 DO 10 L=K+4,N,4
630 10 EN2(L)=((J-K)*EN2(L)-(J-L)*EN2(K))/(L-K)
640 INTJ=EN2(N)
650 RETURN
660 END
670 SUBROUTINE MATOUT(C,D,A,B)
680 COMMON EN(73,73)

```



```

690 INTEGER A,B,C,D, IANG(73)
700C
710 M=16+A
720 IC=0
730C**
740 9 IF(IC*17.GT.B) STOP
745 IF(M.GT.B)M=B
750 DO 5 I=A,M
760 5 IANG(I)=(I-1)*5
770 WRITE(6,101)(IANG(I),I=A,M)
780C
790 DO 10 I=C,D
800 I1=(I-1)*5
810 WRITE(6,102)I1,(EN(I,J),J=A,M)
820 10 CONTINUE
830C
840 IC=IC+1
850 A=A+17
860 M=M+17
880 GOTO 9
890C**
900 101 FORMAT(1H0,5X,17I7,/)
910 102 FORMAT(1X,I4,3X,17(F6.2,1X))
920 END

```

```

1C Calculate XT distribution for interpolated surfaces
010 COMMON I11,I21
020 DIMENSION EN(40,40),IX(40,40),XP(32),X(32)
030C
33 DO 10 I=1,32
36 10 XP(I)=0.0
040 READ, NUM
45 SUM=0.0
050 DO 1000 IC=1,NUM
060C
070 READ, I11,I12,I21,I22
080 I1=(I12-I11)*0.2+1
090 I2=(I22-I21)*0.2+1
93 I11=I11*0.2+1
96 I21=I21*0.2+1
100C
110 DO 20 I=1,I1
120 20 READ, (EN(I,J),J=1,I2)
130C
140C FIND PROBABILITY FOR EACH XT
150 DO 200 I=1,I1
160 DO 200 J=1,I2
170 IX(I,J)=0
180 IF(EN(I,J).GT.12.) GOTO 200
190 IX(I,J)=IXT(I,J)
200 EN(I,J)=EXP(-(EN(I,J)+2.12)*1.62)
205 SUM=SUM+EN(I,J)
210 200 IF(EN(I,J).GT.12.) EN(I,J)=0.0
220C
230C GROUP THE XT'S IN TERMS OF PROBABILITIES
240 DO 300 I=1,I1
250 DO 300 J=1,I2

```

```

260 I3=IX(I,J)-26
265 IF(EN(I,J).EQ.0.0) GO TO 300
270 IF(I3.LE.0 .OR. I3.GE.33) GOTO 250
280 XP(I3)=XP(I3)+EN(I,J)
290 GOTO 300
300 250 PRINT, "XT OUT OF RANGE, =",IX(I,J)
310 300 CONTINUE
320C
330 1000 CONTINUE
340C
350C PRINT RESULTS
360 DO 400 I=1,32
365 XP(I)=XP(I)/SUM
370 400 X(I)=(I+26.)/10.0
380C
390 DO 500 I=1,25,8
400 WRITE(6,1)( X(J),J=I,I+7)
410 WRITE(6,2)(XP(J),J=I,I+7)
420 500 CONTINUE
430 1 FORMAT(1X,"XT=",8F11.1)
440 2 FORMAT(1X,"P= ",4X,8(E10.4,1X))
450 STOP
460 END
470 FUNCTION IXT(I,J)
480 COMMON I11,I21
490 DATA A/1.54/,B/1.81/,C/2.17/,R,S,T/3*1.9106232/,P/0.0872664/
500 T2=(I-2+I11)*P
510 T3=(J-2+I21)*P
520 C1=-COS(R)
530 S1=SIN(R)
540 S2=SIN(S)
550 S3=SIN(T)
560 X1=C*S3*COS(T3)
570 Y1=C*S3*SIN(T3)
580 Z1=-A+C*COS(T)
590 X2=B*C1*S2*COS(T2)+A*S1-B*S1*COS(S)
600 Y2=-B*S2*SIN(T2)+A*C1-B*C1*COS(S)
610 Z2=-B*S2*S1*COS(T2)+A*C1-B*C1*COS(S)
620 IXT=((X2-X1)2+(Y2-Y1)2+(Z2-Z1)2geometry optimisation.5)*10.0
630 RETURN
640 END

```

```

100 THIS PROGRAM INTERPOLATES P.E. SURFACES
20 LOGICAL LT
30 INTEGER ANG$A,B,C,D,ST
40 DOUBLE PRECISION A1(49),EN,EN2(49),D1(49),W(49)
50 COMMON EN(53,53),A,B,C,D
60 DIMENSION ANG$(73)
510
62 READ, NUM
64 DO 1000 ICT=1,NUM
660
70 LT=.FALSE.
800 READ IN INDICES : A,B ARE X LIMITS & C,D ARE Y LIMITS
90 READ, A,B,C,D
94 WRITE(5,97) A,B,C,D
97 97 FORMAT(4I4)
1000 CONVERT FROM DEGREES TO INTERNAL COORD'S
110 A=(A*.2)+1
120 B=(B*.2)+1
130 C=(C*.2)+1
140 D=(D*.2)+1
150 IF(C.LT.0) LT=.TRUE.
160 IF(LT) ST=C
170 IF(LT) C=1
180 IF(LT) D=D-ST+1
190 IF(A.LT.0) GOTO 300
2000 READ IN ENERGIES
210 DO 20 I=A,B,4
220 20 READ,(EN(I,J),J=C,D,4)
2300
2400 INTERPOLATE Y'S ALONG THE LINES X=A TO B (STEP 4)
250 N=(D-C)*0.25+1
260 IG=N+1
270 CALL INTJ(N,IG,EN2,A1,D1,W)
2800
2900 INTERPOLATE X'S ALONG THE LINES Y=C TO D
3000 --MISSING OUT KNOWN VALUES
310 N=(B-A)*0.25+1
320 IG=N+1
330 CALL INTI(N,IG,EN2,A1,D1,W)
3400
3500 PRINT FINAL VALUES
360 CALL MATOUT(LT,ST)
365 1000 CONTINUE
367 STOP
3700
380 300 PRINT, "A CANNOT BE < 0. PUT C < 0."
390 STOP
400 END
410 SUBROUTINE INTI(N,IG,EN2,A1,D1,W)
420 DOUBLE PRECISION EN2(N),EN
430 DOUBLE PRECISION A1(N),W(IG),D1(IG),X,Y
440 INTEGER C,D,A,B
450 COMMON EN(53,53),A,B,C,D
4600
470 I1=A-4
480 DO 10 J=C,D
490 DO 5 K=1,N
500 I2=(K*4)+I1
510 A1(K)=I2
520 5 EN2(K)=EN(I2,J)
5300
540 DO 10 K=A+1,B-3,4

```

```

550 DO 10 I=K,K+2
560 X=I
570 CALL ED1ADF(N,X,A1,EN2,W,D1,IG,Y)
580 EN(I,J)=Y
590 10 CONTINUE
600 RETURN
610 END
620 SUBROUTINE INTJ(N,IG,EN2,A1,D1,W)
630 INTEGER C,D,A,B
640 DOUBLE PRECISION EN2(N),A1(N),EN,W(IG),D1(IG),X,Y
650 COMMON EN(53,53),A,B,C,D
660C
670 I1=C-4
680 DO 10 I=A,B,4
690 DO 5 K=1,W
700 I2=4*K+I1
710 A1(K)=I2
720 5 EN2(K)=EN(I,I2)
730C
740 DO 10 K=C+1,D-3,4
750 DO 10 J=K,K+2
760 X=J
770 CALL ED1ADF(N,X,A1,EN2,W,D1,IG,Y)
780 EN(I,J)=Y
790 10 CONTINUE
800 RETURN
810 END
820 SUBROUTINE MATOUT(LT,IR)
830 DOUBLE PRECISION EN
840 COMMON EN(53,53),C,D,A,B
850 INTEGER A,B,C,D,IANG(73)
860 LOGICAL LT
870C
880 IF(LT) A=IR
890 IF(LT) B=B+IR-1
990C
1000 DO 10 I=C,D
1020 WRITE(6,102)(EN(I,J),J=A,B)
1030 10 CONTINUE
1040 RETURN
1110 102 FORMAT(4D(F5.2,1X))
1120 END

```



A6.5 Principle IMDAC subroutines.

Below is a list of the main routines in IMDAC in alphabetical order of the files they are in on the Prime. To save space all the COMMON blocks at the beginning of each subroutine have omitted (since the main blocks are given at the start of IMDAC), as well as other non-essential information (eg END and assignment statements), and code of a trivial nature (eg the array-clearing code in READAT). In addition, routines which have already been (or will be - see A7) described in detail have been omitted.

```

PROGRAM IMDAC
CHARACTER IAT*4,TITLE*4,RES*3
LOGICAL INEW,DBG,MOL1,MOL2,JUST2,DONE,BOND,EN,EXCL1,MOVD,NOSCL
* ,KEEP,THICK,ASET,RPT
COMMON/CL1/NT1,NT2,X(3,3410),ITYPE(3410),PC(3410)
* /ATS/IAT(3410),RES(3410)
./BEN/NTMS(3410),THICK,KEEP,RPT
* /DUM/DUMX(3,3410)
* /MOL/MOL1,MOL2,JUST2,EN,EXCL1
./RTR/MOVD,NOSCL,ASET
* /AV1/F1,F2,F3,F4,F5,F6
* /AV2/G1,G2,G3,G4,G5,G6
* /VIEW/IVEW(50),DBG
* /SCL/SCAL(50)
* /NMOL/INEW
COMMON/TIT/TITLE(2,20)
COMMON/EN2/POLEF(20,20)
* /OLD/DONE
* /RSS/IRES(700),NRES,NRES2
C
C Programme for Interactive Molecular Display and Calculation
OPEN(12,FILE='POLEF')
DO 111 I=1,20
111 READ(12,*)(POLEF(I,J),J=1,20)
CLOSE(12)
NT1=0
MOVD=.FALSE.
KEEP=.FALSE.
ASET=.TRUE.
JUST2=.FALSE.
INEW=.FALSE.
CALL TTYP
10 DO 200 I=2,50
SCAL(I)=0.
200 IVEW(I)=0
250 CALL READAT(BOND)
IF(DONE) GOTO 20
IF(BOND)CALL BONDS
IF(.NOT.(JUST2.OR.NOSCL))CALL ZSCALE

```

```

20 IF(.NOT.JUST2)CALL SETUPE
   CALL OPTS
   IF(JUST2.OR..NOT.ASET)GOTO 250
   IF(INEW) GOTO 10
   CALL DEVEND
   STOP
   END
   SUBROUTINE ANS(L)
   LOGICAL L
   CHARACTER*1 A,YES
   DATA YES/'Y'/
C
   READ(1,1) A
1  FORMAT(A1)
   L=(A.EQ.YES)
   RETURN

   SUBROUTINE BENSON
   SCAL(1)=SCAL(1)*0.8
   IF(.NOT.KEEP)THEN
     CALL CHAMOD
5    WRITE(1,10)
10   FORMAT(' File name: ', )
     READ(1,20)FILE
20   FORMAT(A12)
     IF(FILE.EQ.' ')FILE='PLOT1'/_
     IF(FILE.EQ.'Q')RETURNE
     CALL DEVEND
     THICK=MOL2
     IF(MOL2)WRITE(1,30)
30   FORMAT(' Enhance MOL2? ', )/_
     IF(MOL2)CALL ANS(THICK)
     IF(MOL2.AND.NT1.GT.100)THEN/_
40     WRITE(1,40)
        FORMAT(' Include Close Contacts?')
        CALL ANS(CCS)
     ENDIF
     WRITE(1,50)
50   FORMAT(' Draw twice?')E
     CALL ANS(RPT)
     WRITE(1,60)
60   FORMAT(' Keep plot file open?')'
     CALL ANS(KEEP)
     NODRAW=KEEP
     NORUB=KEEP
     OPEN(16,FILE=FILE,ERR=5)
     CALL B1302
     CALL DEVPAP(880.,330.,0)
     CALL PENSEL(0,0.,0)
     CALL WINDOW(3)
     BEN=.TRUE.
   ELSE
C   Note that 'APPEND' is none standard'
     OPEN(16,FILE=FILE,STATUS='APPEND')')
     ENDIF
80  IF(MOL1)CALL BENDRW(1)
     IF(MOL2)CALL BENDRW(2)

```

BENDRW

```

IF(THICK)THEN
  IVEW(14)=1
  D=0-2
  CALL LRUD(D,F)
  CALL BENDRW(2)
  D=-0-2
  CALL LRUD(D,F)
  CALL BENDRW(2)
  IVEW(14)=2
  CALL LRUD(D,F)
  CALL BENDRW(2)
  D=0-2
  CALL LRUD(D,F)
  CALL BENDRW(2)
ENDIF
IF(CCS)CALL CC(CCS)
IF(RPT.AND..NOT.RESET)THEN
  IVEW(14)=1
  D=330.0
  CALL LRUD(D,F)
  RESET=.TRUE.
  GOTO 80
ELSEIF(RESET)THEN
  IVEW(14)=1
  D=-330.0
  CALL LRUD(D,F)
ENDIF
SCAL(1)=SCAL(1)*1.25
BEN=.FALSE.
IF(KEEP)CLOSE(16)
IF(KEEP)RETURN
CALL DEVEND
IF(ITT.EQ.0)CALL BS5660E
IF(ITT.EQ.1)CALL S5600
IF(ITT.EQ.2)THEN
  CALL T4010
  CALL UNITS(0.52)
ENDIF
IF(ITT.EQ.3)CALL S5660
CALL WINDOW(3)
CALL ERRMAX(500)
IF(ITT.EQ.0.OR.ITT.EQ.3)CALL SETCOL(')
RETURN
SUBROUTINE BENDRW(NMOL)E
IF(NMOL.EQ.1)THEN
  N1=IVEW(11)
  N2=IVEW(12)
  IF(N2.GT.NT1)N2=NT1
ELSE
  N1=NT1+1
  N2=NT1+NT2
ENDIF
C
YES=.TRUE.
CALL CHAMOD
IF(N2-N1.GT.300)WRITE(1,13)D/
IF(N2-N1.GT.300)CALL ANS(YES)/
13 FORMAT(' >300 ATOMS, DO YOU WISH TO CONTINUE? ', )

```

```

      IF(.NOT.YES)RETURN
      R=SCAL(1)*0.2+0.5
      IF(R.LT.1.0) R=1.0
C
      IF(IVEW(NMOL).NE.4.AND.IVEW(NMOL).NE.10)THEN
        DO 100 I=N1,N2
          IF(IVEW(NMOL).EQ.5)CALL CIR3(I,R)E.1
          IF(LAB(I))CALL LABEL(I)
          DO 100 J=1,NCON(I)
            L=ICON(J,I)
            IF(L.GT.I)CALL LINE3(L,I,.FALSE.)E.1
100      CONTINUE
        ELSE
C          DO 200 I=N1,N2
C          IF(ITYPE(I).EQ.1.OR.ITYPE(I).GT.14)LAB(I)=.TRUE.<
C 200      CONTINUE
          DO 300 I=N1,N2
            DO 300 J=1,NCON(I)
              L=ICON(J,I)
              IF(L.GT.I)CALL LINE3(L,I,.TRUE.)
300      CONTINUE
        ENDIF
        RETURN
        SUBROUTINE CC(BEN)
          EQUIVALENCE (IP1,IVEW(3)),(IP2,IVEW(4)),(IP3,IVEW(5)).
          ICC=0
          V612=0.0
          VSTAT=0.0
          VIJ=0.0
          EN=(IVEW(40).NE.0)
          STAT=(IVEW(33).NE.0)
          CALL CHAMOD
          IF(I2.NE.0)THEN
            WRITE(1,102)
102      FORMAT(' KEEP SAME MOL2 ATOMS? ', )
            CALL ANS(YES)
            IF(YES)GOTO 105
          ENDIF
C      HOW MANY MOL2 ATOMS INCLUDED IN SEARCH?
          IF(.NOT.BEN)THEN
            1  WRITE(1,2)
            2  FORMAT(' NO. OF MOL2 ATOMS TO BE SEARCHED (1-20 OR ALL(0)): ', )
              READ(1,3,ERR=1)I2
            3  FORMAT(I6)
          ELSE
            I2=0
            CALL PENSEL(0,0.,0)
          ENDIF
C
          IF(I2.LT.1.OR.I2.GT.20.OR.I2.GT.NDRUG)THEN
            I2=NDRUG
            DO 110 I=1,I2
110      IMOL2(I)=N+I
          ELSE
            WRITE(1,4)
            4  FORMAT(' INPUT ATOMS FOR MOL2')
              ICNT=0
              CALL IDATOM(IMOL2,ICNT,I2,N+1,N+NDRUG)

```



CC

```

        IF(ICNT.NE.I2)RETURN
        CALL PENSEL(5,0.,0)
C      CALL CHAMOD
        ENDIF
C
C      WHERE DO WE LIST 'CONTACTS'D/
105  WRITE(IO1,9)
      9  FORMAT(' Output File [Default - Terminal]: ', )
        READ(IO1,6) FILE
        IF(FILE.EQ.'Q')RETURN
        IF(FILE.EQ.SPACE) IO2=1E
        IF(FILE.NE.SPACE) OPEN(IO2,FILE=FILE,STATUS='NEW',ERR=105)
C
        WRITE(1,5)
      5  FORMAT(' Find H-bonds? ', )/
        CALL ANS(HB)
        IF(.NOT.HB)THEN
99      WRITE(1,101)
101     FORMAT(' Sensitivity, A (lower No. = more sensitive): ', )
        READ(1,7,ERR=99) SENSE
        IF(SENS.GT.3.) SENS=3.E
C      SET FIRST CUT-OFF - BASED ON BIGGEST RHRD
        COARSE=((3.9-SENS))*2E
C      PICK UP VAN DER WAAL RADII (SOFT OR HARD)
C      ARE WE LOOKING FOR HYDROGEN BONDS'
        ISFT=(IVEW(31).EQ.1)
C
C      IF NO ENERGY CALC. THEN PICK UP HARD RADII
C      IF ENERGY CALC. PICK ACCORDINGLY&
        IF(ISFT)THEN
C
C      SOFT
        OPEN(21,FILE='VDW1')E
        READ(21,6) TEXT
        READ(21,6) TEXT
      6  FORMAT(A4)
        DO 50 I=1,11
        READ(21,6) TEXT
        READ(21,7) (RSFT(I,J),J=1,11)'
50      CONTINUE
        DO 70 I=1,11
        DO 60 J=1,11
        IF(RSFT(I,J).LT.0.00000001) RSFT(I,J)=RSFT(J,I)<
60      CONTINUE
70      CONTINUE
      7  FORMAT(11F6.4)
        CLOSE(21)
C      ENDIF
        WRITE(IO2,19)
19      FORMAT(' Atom RES',6X,'atom no.',7X,'Calc.',2X,'Theor.',4X,
      * 'Diff. V6-12',5X,'VSTAT',/)&
        ELSE
          COARSE=3.9*3.9
80      WRITE(1,10)
10      FORMAT(' RMIN,RMAX: ', )E
        READ(1,11,ERR=80) ARAD,BRADD/
11      FORMAT(2F4.2)

```

```

      IF(BRAD.LT.0.00000001) BRAD=6.25
      ARAD=ARAD*ARAD
      WRITE(IO2,18) BRAD
18   FORMAT(/' Possible Hydrogen Bonds [H...X < ',F10.3,'A ]',
      ENDIF
      CALL DASHED(1,3.,1.,1.)E
C
C
      DO 500 I=1,N
      IF(.NOT.ACTSIT(I))GOTO 500
      K=ITYPE(I)
      DO 400 JJ=1,I2
C
      J=IMOL2(JJ)
      L=ITYPE(J)
C
      BNAME=IAT(J)
      CNAME=IAT(I)
      N1=NBONDS+1
      N2=NBONDS+MBONDS
      DO 303 IHB=N1,N2
      IF(IGOTHB.EQ.2) GO TO 306
      IF(IA(IHB).NE.I.AND.IB(IHB).NE.I) GO TO 3033
      IF(IA(IHB).EQ.I) IDONOR=IB(IHB)%
      IF(IB(IHB).EQ.I) IDONOR=IA(IHB)%
C
C      IF THIS NOT ELECTRONEGATIVE THEN SKIPT
      IF(ITYPE(IDONOR).LE.7) GO TO 400&
      ANAME=IAT(IDONOR)
      GO TO 304
306 IF(IA(IHB).NE.J.AND.IB(IHB).NE.J) GO TO 3033
      IF(IA(IHB).EQ.J) IDONOR=IB(IHB)%
      IF(IB(IHB).EQ.J) IDONOR=IA(IHB)%
C
      IF THIS NOT ELECTRONEGATIVE THEN SKIPT
      IF(ITYPE(IDONOR).LE.7) GO TO 400&
      ANAME=IAT(IDONOR)
      GO TO 304
C 303 CONTINUE
304 RB=RIJ
      RA=0.0
      RC=0.0
      DO 305 IHB=1,3
      IF(IGOTHB.EQ.2) GO TO 307
      RA=RA+(X(IHB,IDONOR)-X(IHB,I))**2'
      RC=RC+(X(IHB,IDONOR)-X(IHB,J))**2'
      GO TO 305
307 RA=RA+(X(IHB,IDONOR)-X(IHB,J))**2'
      RC=RC+(X(IHB,IDONOR)-X(IHB,I))**2'
305 CONTINUE
      IF(RA.LT.0.00000001.OR.RC.LT.0.00000001)THENN
      WRITE(IO2,22) ANAME,BNAME,CNAME,RA,RB,RC
22   FORMAT(' Atoms coincide: ',A8,3H- ,A8,5H... ,A8,2X,3F7.3)
      GOTO 400
      ENDIF
      RA=SQRT(RA)
      RC=SQRT(RC)
      ANG=(RA*RA+RB*RB-RC*RC)/(2.0*RA*RB) ,

```

CC

```

    ANG=ACOS(ANG)*180.0/3.141592D/
    WRITE(IO2,21) ANAME,BNAME,CNAME,RA,RB,RC,ANGG
21  FORMAT(1X,A4,3H- ,A4,5H... ,A4,2X,3F7.3,2X,F6.1)
    ICC=ICC+1
C
    ELSE
        RIJ=0.0
        DO 130 M=1,3
            RIJ=RIJ+((X(M,I)-X(M,J))*(X(M,I)-X(M,J)))
            IF(RIJ.GT.COARSE)GOTO 400D/
130    CONTINUE
        IF(RIJ.GT.0.00000001)R=SQRT(RIJ)
C
C    COMPARE WITH THEORETICALE
    RTHEOR=RHRD(K)+RHRD(L)E
C
    IF(R.GT.RTHEOR-SENS) GO TO 400&
    IF(EN)THEN
        IF(R.GT.0.0001)THENE
            CALL ENIJ(K,L,R,RIJ)
            V612=V612+RIJ
            IF(STAT)VIJ=332.0*PC(I)*PC(J)/RM
            IF(BEN)THEN
                CALL PENSEL(2,0.,0)D/
                IF(RIJ+VIJ.GT.0.0)CALL PENSEL(1,0.,0)
            ELSE
                CALL PENSEL(5,0.,0)D/
                IF(RIJ+VIJ.LT.0.0)CALL PENSEL(15,0.,0)
            ENDIF
            IF(STAT)VSTAT=VSTAT+VIJ/
        ELSE
            RIJ=99999999.
        ENDIF
    ELSE
        RIJ=0.
    ENDIF
    ICC=ICC+1
    DIFF=RTHEOR-R
    CALL CHAMOD
    WRITE(IO2,13) IAT(I),MRES(I),IAT(J),J,R,RTHEOR,DIFF,RIJ,VIJ
13  FORMAT(1X,A4,I3,6X,A4,I4,2X,F10.3,2X,F5.3,4X,F6.3,2F9.3)
    ENDIF
C
    CALL LINE(I,J)
C
C 400 CONTINUE
C 500 CONTINUE
C
    IF(FILE.NE.SPACE)CLOSE(IO2)D/
    CALL DASHED(0,3.,2.,1.)E
    WRITE(IO1,170) ICC
170  FORMAT(' No. of contacts/bonds=',I6)2X
    IF(EN)WRITE(1,171)V612,VSTATD/
171  FORMAT(' Total energy=',2F11.3)%
    RETURN
    SUBROUTINE ENIJ(K,L,RIJ,VIJ)D/
C
    R6=1.0/(VIJ*VIJ*VIJ)

```

```

R0=VDWR(K)+VDWR(L)
R0=R0*R0*R0*R0*R0*R0
VIJ=362.187*POLEF(K,L)*R6*((R0*0.5*R6)-1.0))
RETURN
SUBROUTINE CLEFT
EQUIVALENCE (IP1,IVEW(3)),(IP2,IVEW(4)),(IP3,IVEW(5)).
C
SCLF=SCAL(1)
CALL VOL(IQ)
IF(IQ.EQ.2)RETURN
CALL CHAMOD
NATS=0
70 WRITE(1,75)
75 FORMAT(' Input radius of spherical "probe"')
READ(1,80,ERR=70)RSPH
80 FORMAT(F9.3)
IF(RSPH.LT.0.5.OR.RSPH.GT.8.0)RSPH=2.5
WRITE(1,85)
85 FORMAT(' Increment: ', )E
READ(1,80,ERR=70)RINC
IF(RINC.LT.0.05.OR.RINC.GT.RSPH*1.8)RINC=RSPH
C
IF(NATS.EQ.0)CALL XVOL(RSPH)D/
CALL BOX2(X1,Y1,X2,Y2)
C Search through XCLFT for "holes"
X4=X2-RSPH
Y4=Y2-RSPH
Z4=Z2-RSPH
R2=RSPH**2
ICNT=0
XK=Z1
WHILE(XK.LT.Z4)DO
  XJ=Y1
  XK=XK+RINC
  WHILE(XJ.LT.Y4)DO
    XI=X1
    XJ=XJ+RINC
    WHILE(XI.LT.X4)DO
      XI=XI+RINC
      DO 200 IT=1,NATS
        DST=(XI-XCLFT(IT,IP1))*(XI-XCLFT(IT,IP1))
        * + (XJ-XCLFT(IT,IP2))*(XJ-XCLFT(IT,IP2))
        * + (XK-XCLFT(IT,IP3))*(XK-XCLFT(IT,IP3))
        IF(DST.LT.R2)GOTO 250D/
      200 CONTINUE
C Found small hole - note I,J,KD/
      ICNT=ICNT+1
      IF(ICNT.GT.400)GOTO 255/
      XIJK(1,ICNT)=XI
      XIJK(2,ICNT)=XJ
      XIJK(3,ICNT)=XK
      250 CONTINUE
    ENDWHILE
  ENDWHILE
ENDWHILE
255 WRITE(1,260)ICNT
260 FORMAT(' No. of small holes=',I6)'
C

```

CLEFT



CLEFT

```

DIFF=6.0/(Z2-Z1)
RDRW=RSPH*SCLF
DO 300 I=1,ICNT
X3=XIJK(1,I)*SCLF+128.0E
Y3=XIJK(2,I)*SCLF+128.0E
Z3=XIJK(3,I)
JCOL=9+IFIX((Z3-Z1)*DIFF)
CALL PENSEL(JCOL,0.,0)
300 CALL CIR(X3,Y3,Z3,RDRW)E
CALL CHAMOD
IF(ICNT.EQ.0)WRITE(1,320)
320 FORMAT(' No holes found')
WRITE(1,340)
340 FORMAT(' Try again with different radius? ', )
CALL ANS(YES)
IF(YES)THEN
    CALL PICCLE
    CALL DUMMY
    CALL BOX2(X1,Y1,X2,Y2)E
    GOTO 70
ENDIF
RETURN
SUBROUTINE VOL(ICUR)
EQUIVALENCE (IP1,IVEW(3)),(IP2,IVEW(4)),(IP3,IVEW(5)).
C
    CALL CURDEF(' QSH*.')
10 CALL CHAMOD
    WRITE(1,20)
20 FORMAT(' Define region of space to be examined for possible',
*' clefts.',/, ' First input x,y dimensions by defining opposite',,
*' corners of a square using the cursor.',/, 'Then input z-',
*' dimensions from the orthogonal view displayed.')i
    CALL CURSOR(ICUR,X1,Y1)E
    IF(ICUR.EQ.2)RETURN
    CALL CURSOR(ICUR,X2,Y2)E
    IF(ICUR.EQ.2)RETURN
    X1=(X1-128.0)/SCAL(1)
    X2=(X2-128.0)/SCAL(1)
    Y1=(Y1-128.0)/SCAL(1)
    Y2=(Y2-128.0)/SCAL(1)
    CALL PENSEL(14,0.,0)
    CALL BOX2(X1,Y1,X2,Y2)
    IWT=1200
    CALL SLEEP (IWT)
C
    CALL RUP
C set z-axis vertical
    DO 30 I=3,5
    IVEW(I)=IVEW(I)+1
    IF(IVEW(I).GT.3)IVEW(I)=1
30 CONTINUE
    CALL DUMMY
C
    CALL CHAMOD
    WRITE(1,40)X1,X2,Y1,Y2
40 FORMAT(/, ' X1,X2=',2F8.3, ' Y1,Y2=',2F8.3,/, 'Define z-axis (noww
*vertical) by reading in 2 points on that axis')i
    CALL CURSOR(ICUR,DUM,Z1)E

```

CLEFT

```

CALL CURSOR(ICUR,DUM,Z2)E
IF(ICUR.EQ.2)GOTO 55
IF(ICUR.EQ.4)WRITE(1,50)E
50 FORMAT(' Q=quit, S=start again, H=help')

IF(ICUR.GT.2.AND.ICUR.LT.5)GOTO 10
Z1=(Z1-128.0)/SCAL(1)
Z2=(Z2-128.0)/SCAL(1)
CALL RUB
C
C Reset IVEW to original valuesD/
55 DO 60 I=3,5
IVEW(I)=IVEW(I)-1
IF(IVEW(I).LT.1)IVEW(I)=3
60 CONTINUE
CALL DUMMY
RETURN
SUBROUTINE XVOL(RSPH)
EQUIVALENCE (IP1,IVEW(3)),(IP2,IVEW(4)),(IP3,IVEW(5)).
C
CALL FIXYZ(X1,X2,RSPH)
CALL FIXYZ(Y1,Y2,RSPH)
CALL FIXYZ(Z1,Z2,RSPH)
C Put atoms within box into XCLFT/
DO 90 I=1,NT1
IF(DUMX(IP1,I).GE.X1.AND.DUMX(IP1,I).LE.X2.AND.(
* DUMX(IP2,I).GE.Y1.AND.DUMX(IP2,I).LE.Y2.AND..
* DUMX(IP3,I).GE.Z1.AND.DUMX(IP3,I).LE.Z2)THEN
NATS=NATS+1
XCLFT(NATS,IP1)=DUMX(IP1,I)/
XCLFT(NATS,IP2)=DUMX(IP2,I)/
XCLFT(NATS,IP3)=DUMX(IP3,I)/
ICLFT(NATS)=I
ENDIF
90 CONTINUE
RETURN
SUBROUTINE FIXYZ(V1,V2,R)
IF(V1.GT.V2)THEN
STR=V1
V1=V2
V2=STR
ENDIF
IF(V2.LT.V1+R)THEN
V2=V2+R
WRITE(1,10)V1,V2
10 FORMAT('Dimensions increased, v1,v2:',2F8.3)
ENDIF
REM=(V2-V1)/R-FLOAT(IFIX((V2-V1)/R))2:
C V2=V2+(1.0-REM)*R
RETURN
SUBROUTINE CONF(RD,EN)
* ,AEND(9)
C
ENZCH=.FALSE.
CALL CHAMOD
100 WRITE(1,10)
10 FORMAT(' How many rotations (1,2 or 3)? ', )
READ(1,*,ERR=100)IROT
IF(IROT.LT.1.OR.IROT.GT.3)GOTO 100

```

CONF

```

C
  IF(IROT.NE.3)THEN
    DO 105 I=IROT+1,3
      AINIT(I)=0.
      AINC(I)=5.
      NANG(I)=1
105   AEND(I)=0.
    ENDIF
    P180=3.1415927/180.0
C
C   SET UP ROTATION ANGLES
210 IF(RD)THEN
  DO 215 I=1,IROT
    READ(2,*,ERR=9000,END=9000)II(I),IP(I),IQ(I),JJ(I)<_.
    IF(II(I).LE.N)ENZCH=.TRUE.D/_
215  CONTINUE
  ELSE
C   SET UP ROTATION ANGLES
    WRITE(1,30)
30   FORMAT(' Input atoms to define each rotation')(
    CALL TRANSF(2)
C
    DO 200 I=1,IROT
      ICNT=0
      CALL IDATOM(ICD,ICNT,4,N+1,N+NDRUG)
      IF(ICNT.NE.4)RETURN
      II(I)=ICD(1)
      IF(II(I).LE.N)ENZCH=.TRUE.D/_
      IP(I)=ICD(2)
      IQ(I)=ICD(3)
200   JJ(I)=ICD(4)
    ENDIF
    CALL RUB
C
C   FIND ATOMS FOR EACH ROTATION
    DO 230 I=1,IROT
      ID=I
      ICD(1)=II(I)
      ICD(2)=IP(I)
      ICD(3)=IQ(I)
      ICD(4)=JJ(I)
      CALL SERCH
      IF(NATOMS(I).NE.0)THEN
        CALL DIHED(II(ID),IP(ID),IQ(ID),JJ(ID),OM))
        WRITE(1,40)I,OM
40   FORMAT(' Angle for rotation',I2,' is ',F8.3)
C
110  WRITE(1,20)I
20   FORMAT(' Input start angle, final angle and increment for ',
  *   'rotation',I2)
      READ(1,*,ERR=110)AINIT(I),AEND(I),AINC(I))
      IF(ABS(AINC(I)).LT.0.01)AINC(I)=5.0I
      IF(ABS(AEND(I)).LT.0.01)AEND(I)=360.0-AINC(I)r
      AINIT(I)=AINIT(I)*P180E
      AINC(I)=AINC(I)*P180
      COST(I)=COS(AINC(I))
      SINT(I)=SIN(AINC(I))
      AEND(I)=AEND(I)*P180

```

CONF

```

C   SET ANGLE TO INITIAL VALUEE
      TAU=AINIT(I)-OM*PI80
      IF(ABS(TAU).GT.0.00000001)CALL TORROT(TAU)
      CALL DIHED(II(ID),IP(ID),IQ(ID),JJ(ID),OM)
      WRITE(1,42)OM
42   FORMAT(' New angle= ',F8.3)/
      ELSE
      WRITE(1,50)I
50   FORMAT('No atoms found for rotation',I2)
      RD=.FALSE.
      GOTO 210
      ENDIF
C 230 CONTINUE
      CALL CHAMOD
220 WRITE(1,60)
60   FORMAT(' Input criterion for a FIT (A**2): ', )r
      READ(1,70,ERR=220)FITMAXE
70   FORMAT(F8.3)
      IF(FITMAX.LT.0.01.OR.FITMAX.GT.6.)FITMAX=0.33
C
C   ROTATE AND FIND BEST FIT
      DO 300 I=1,IROT
      NANG(I)=IFIX((AEND(I)-AINIT(I))/AINC(I))

      IF(NANG(I).LE.0)NANG(I)=1
300  CONTINUE
      R2=(IROT.GT.1)
      R3=(IROT.GT.2)
      IF(.NOT.R2)NANG(2)=1
      IF(.NOT.R3)NANG(3)=1
      ICNT=0
C   CALCULATE DIST'S FOR MOLIE
      CALL DSTENZ
      IF(R2)NANG(2)=NANG(2)+1E
      IF(R3)NANG(3)=NANG(3)+1E
      CALL TRANSF(2)
C
      DO 1000 I3=1,NANG(3)
      DO 2000 I2=1,NANG(2)
      CALL GETFIT(0,I2,I3,EN,ENZCH)/
      ID=1
      CALL AMAT(AINC(1),COST(1),SINT(1))
      DO 3000 I1=1,NANG(1)
      CALL ROTCON
      CALL GETFIT(I1,I2,I3,EN,ENZCH)
3000 CONTINUE
C 3000 RESET ANGI
      TAU=-1.0*FLOAT(NANG(1))*AINC(1)%
      CT=COS(TAU)
      ST=SIN(TAU)
      CALL AMAT(TAU,CT,ST)
      CALL ROTCON
      ID=2
      IF(R2)CALL AMAT(AINC(2),COST(2),SINT(2))

      IF(R2)CALL ROTCON
2000 CONTINUE
C 2000 RESET ANG2
      TAU=-1.0*FLOAT(NANG(2))*AINC(2)%

```

```

CT=COS(TAU)
ST=SIN(TAU)
IF(R2)CALL AMAT(TAU,CT,ST)
IF(R2)CALL ROTCON
ID=3
IF(R3)CALL AMAT(AINC(3),COST(3),SINT(3))
IF(R3)CALL ROTCON
C 1000 CONTINUE
RESET ANG3
TAU=-1.0*FLOAT(NANG(3))*AINC(3)%
CT=COS(TAU)
ST=SIN(TAU)
IF(R3)CALL AMAT(TAU,CT,ST)
IF(R3)CALL ROTCON
C
IF(ICNT.EQ.0)THEN
WRITE(1,90)
90  FORMAT(' No fit found - try again? ', )
CALL ANS(YES)
IF(YES)GOTO 220
RETURN
ENDIF
C FIND & ROTATE TO BEST FIT ANGLES
FITMIN=999.
DO 400 I=1,ICNT
IF(FIT(I).GT.FITMIN)GOTO 400D/
FITMIN=FIT(I)
IMIN=I
C 400 CONTINUE
C ROTATE TO MINIMUM ANGLES
DO 410 I=1,IROT
ID=I
TAU=FLOAT(IANG(I,IMIN))*AINC(I)%
IF(ABS(TAU).GT.0.00000001)CALL TORROT(TAU)
CALL DIHED(II(ID),IP(ID),IQ(ID),JJ(ID),OM)
WRITE(1,97)ID,OM
97  FORMAT(' OM',I1,'= ',F8.3)
ANG=(TAU+AINIT(I))/PI80E
WRITE(1,80)I,ANG
80  FORMAT(' For rotation',I2,' angle= ',F8.3))
410 CONTINUE
WRITE(1,95)IMIN,FIT(IMIN)
95  FORMAT(' IMIN= ',I4,' FIT= ',F8.3) '
RETURN
9000 WRITE(1,9010)
9010 FORMAT(' ERROR/END in data file - skip minimisation? ', )>
CALL ANS(RD)
IF(.NOT.RD)GOTO 210
RETURN
SUBROUTINE GETFIT(I1,I2,I3,EN,ENZCH)ip
DATA RDEC/57.2957764/
C
C DEFINITION ID(4)=ID(1) etc.E
DO 100 I=1,3
DRG=0.
ENZ2=0.

```



```

K=I+1
DO 200 J=1,3
  DRG=(X(J, ID(I))-X(J, ID(K)))*(X(J, ID(I))-X(J, ID(K)))+DRG>>
  ENZ2=(X(J, IE(I))-X(J, IE(K)))*(X(J, IE(I))-X(J, IE(K)))<.  

200 * CONTINUE +ENZ2  

  DIFF=ABS(SQRT(ENZ2)-SQRT(DRG))+DIFFE(I  

100 CONTINUE  

  IF(DIFF.GT.FITMAX)RETURNE  

  IC=IC+1  

  FIT(IC)=DIFF  

  ANG(1)=(AINIT(1)+AINC(1)*I1)*RDEG'  

  ANG(2)=(AINIT(2)+AINC(2)*(I2-1))*RDEGI  

  ANG(3)=(AINIT(3)+AINC(3)*(I3-1))*RDEGI  

  IANG(1,IC)=I1  

  IANG(2,IC)=I2-1  

  IANG(3,IC)=I3-1  

  IF(EN)CALL ENERGY(EN)  

  WRITE(1,10)(ANG(I),I=1,3),DIFF,ETOT(IC)  

10 FORMAT(' For rotation: ',F6.1,':',F6.1,':',F6.1,' FIT= ',F8.3,,  

  * ' ETOT=',F8.2)  

  RETURN  

  SUBROUTINE DCKSUB(DUM,IENZ,IDRUG)  

  * XDRUG1(3),XDRUG2(3),XDRUG3(3)6.1  

  EQUIVALENCE (DL,XN(1)),(DM,XN(2)),(DN,XN(3)))  

C  

C1 PLACE XENZ1 & XDRUG1 AT ORIGIN/  

  DO 25 I=1,3  

  XENZ1(I)=DUM(I,IENZ(1))E  

  XENZ2(I)=DUM(I,IENZ(2))-XENZ1(I)&  

  XENZ3(I)=DUM(I,IENZ(3))-XENZ1(I)&  

  XDRUG1(I)=DUM(I,IDRUG(1))  

  XDRUG2(I)=DUM(I,IDRUG(2))-XDRUG1(I)DN,  

25 XDRUG3(I)=DUM(I,IDRUG(3))-XDRUG1(I)DN,  

C  

  DO 40 J=1,3  

  DO 30 I=1,NT1  

30 DUM(J,I)=DUM(J,I)-XENZ1(J)  

  DO 40 I=NT1+1,NT1+NT2  

40 DUM(J,I)=DUM(J,I)-XDRUG1(J)D/  

C  

C2 ROTATE SO THAT XENZ2, XDRUG2 & ORIGIN ARE COLINEAR  

  CALL ANG2(XENZ2,XDRUG2,TAU,COST,SINT)A  

  IF(ABS(TAU).LT.0.0015)GOTO 200  

  CALL ANORM(XENZ2,XDRUG2,XN)D/  

C TEST SIGN FROM DETERMINANT OF [XN,XENZ2,XDRUG2]  

  SGN=XENZ2(1)*((XN(2)*XDRUG2(3)-XN(3)*XDRUG2(2)))  

  SGN=SGN+XENZ2(2)*((XN(3)*XDRUG2(1)-XN(1)*XDRUG2(3)))<.  

  SGN=SGN+XENZ2(3)*((XN(1)*XDRUG2(2)-XN(2)*XDRUG2(1)))<.  

  IF(SGN.GT.0.0)THEN  

  DO 50 I=1,3  

50 XN(I)=-XN(I)  

  ENDIF  

  T=1.0-COST  

  A1=COST+(DL*DL)*T  

  A2=DL*DM*T+(DN*SINT)  

  A3=DL*DN*T-DM*SINT  

  A4=DL*DM*T-DN*SINT

```

CONF/

DOCK

```

A5=COST+(DM*DM)*T
A6=DM*DN*T+DL*SINT
A7=DN*DL*T+DM*SINT
A8=DM*DN*T-DL*SINT
A9=COST+(DN*DN)*T

C
DO 100 I=NT1+1,NT1+NT2
XX=A1*DUM(1,I)+A2*DUM(2,I)+A3*DUM(3,I)
YY=A4*DUM(1,I)+A5*DUM(2,I)+A6*DUM(3,I)
DUM(3,I)=A7*DUM(1,I)+A8*DUM(2,I)+A9*DUM(3,I))
DUM(1,I)=XX
100 DUM(2,I)=YY
XDRUG3(1)=DUM(1,IDRUG(3))
XDRUG3(2)=DUM(2,IDRUG(3))
XDRUG3(3)=DUM(3,IDRUG(3))

C
C3 ROTATE XDRUG3 ONTO PLANE CONTAINING XENZ3 & XENZ2)
200 CALL ANORM(XENZ2,XENZ3,XN1)D/
CALL ANORM(XENZ2,XDRUG3,XN2)D/
CALL ANG2(XN1,XN2,TAU,COST,SINT)&
IF(ABS(TAU).LT.0.0015)GOTO 300
DO 220 I=1,3
220 XN(I)=XENZ2(I)
CALL UNIT(XN,D)

C
C TEST SIGN FROM DETERMINANT OF [XN,XN1,XN2]
SGN=XN2(1)*((XN(2)*XN1(3)-XN(3)*XN1(2)))
SGN=SGN+XN2(2)*((XN(3)*XN1(1)-XN(1)*XN1(3)))
SGN=SGN+XN2(3)*((XN(1)*XN1(2)-XN(2)*XN1(1)))
IF(SGN.LT.0.0)THEN
DO 225 I=1,3
225 XN(I)=-XN(I)
ENDIF
T=1.0-COST
A1=COST+(DL*DL)*T
A2=DL*DM*T+(DN*SINT)
A3=DL*DN*T-DM*SINT
A4=DL*DM*T-DN*SINT
A5=COST+(DM*DM)*T
A6=DM*DN*T+DL*SINT
A7=DN*DL*T+DM*SINT
A8=DM*DN*T-DL*SINT
A9=COST+(DN*DN)*T

C
DO 240 I=NT1+1,NT1+NT2
XX=A1*DUM(1,I)+A2*DUM(2,I)+A3*DUM(3,I)
YY=A4*DUM(1,I)+A5*DUM(2,I)+A6*DUM(3,I)
DUM(3,I)=A7*DUM(1,I)+A8*DUM(2,I)+A9*DUM(3,I))
DUM(1,I)=XX
240 DUM(2,I)=YY

C
C4 SHIFT COORDS. BACK TO ORIGINAL ENZ POS'N FOR DRAWING<_.
300 DO 1000 J=1,3
DO 1000 I=1,NT1+NT2
1000 DUM(J,I)=DUM(J,I)+XENZ1(J)
RETURN
SUBROUTINE DOCK
2,NT1+NT2

```

C

```

EN=.FALSE.
EXCL=EXCL1
ISTR=IVEW(35)
4 CLOSE(2)
  CALL CHAMOD
  WRITE(1,5)
5 FORMAT(' Read data from a file? ', )S'
  CALL ANS(RD)
  IF(RD)THEN
    WRITE(1,6)
6    FORMAT(' File name: ', )
    READ(1,7)FILE
7    FORMAT(A12)
    OPEN(2,FILE=FILE)
    READ(2,*,END=4,ERR=4)(IENZ(I),IDRUG(I),I=1,3)N
  ELSE
    WRITE(1,10)
10   FORMAT(' Input 2 sets of atoms to be matched (enzyme 1st)')
    IC=0
    CALL IDATOM(IENZ,IC,3,1,NT1)
    IF(IC.NE.3)RETURN
    CALL IDATOM(IDRUG,IC,3,NT1+1,NT1+NT2)

    IF(IC.NE.6)RETURN
  ENDIF
C CHECK THAT ATOMS READ IN CORRECTLY&
DO 20 I=1,3
  IF(IENZ(I).GT.NT1.OR.IENZ(I).LT.1)THEN
    WRITE(1,11)I,IENZ(I)
    RETURN
  ENDIF
11 FORMAT(' IENZ',I1,' out of range - quit')
  IF(IDRUG(I).LT.NT1+1.OR.IDRUG(I).GT.NT1+NT2)THEN
    WRITE(1,12)I,IDRUG(I)E
    RETURN
  ENDIF
12 FORMAT(' IDRUG',I1,' out of range - quit')
20 CONTINUE
  IENZ(4)=IENZ(1)
  IDRUG(4)=IDRUG(1)
  WRITE(1,101)
101 FORMAT(' Minimise? ', )E
  CALL ANS(MINM)
  MOL1=.FALSE.
  LSTR=MOL2
  MOL2=.TRUE.
  IF(MINM)THEN
    IF(IVEW(40).NE.0)THENE
      WRITE(1,102)
102   FORMAT(' Include en. calcs in superposition? ', ).
      CALL ANS(EN)
      IF(EN)IVEW(35)=3
      EXCL1=EN
      CALL CONF(RD,EN)
    ELSE
      CALL RUB
    ENDIF
  ENDIF

```

```

        IF(RD)CLOSE(2)
        CALL DCKSUB(DUMX,IENZ,IDRUG)D/
        CALL DCKSUB(X,IENZ,IDRUG)
        CALL DUMMY
C   CHECK DIST'S
        NDIST=2
        IVEW(15)=2
        DO 400 I=1,2
            ICODEA(I)=IENZ(I+1)
400    ICODEB(I)=IDRUG(I+1)
            ICODEC(1)=0
            CALL DANG
            IF(MINM.AND.ICNT.GT.0)THEN
                IPREV=IMIN
                CALL CHAMOD
500    WRITE(1,520)FITMAX
520    FORMAT(' Current fit criterion=',F8.3,/,' input VDST '>>
        *      ,'(view all FITS < VDST): ', )'
            READ(1,540,ERR=500)VDST
540    FORMAT(F8.4)
            IF(VDST.LT.0.01)GOTO 800
            CALL RUB
C
            DO 700 J=1,ICNT
                IF(FIT(J).GT.VDST)GOTO 700D/
                DO 600 I=1,IROT
                    ID=I
                    TAU=(IANG(I,J)-IANG(I,IPREV))*AINC(I)

                    IF(ABS(TAU).GT.0.0000001)CALL TORROT(TAU))
                    ANG(I)=(TAU+AINIT(I))/PI80D/
                    CALL DIHED(II(ID),IP(ID),IQ(ID),JJ(ID),OM))
                CALL CHAMOD
                WRITE(1,*)I,OM,ANG(I)
600    CONTINUE
                IPREV=J
                CALL DCKSUB(DUMX,IENZ,IDRUG)
                CALL DCKSUB(X,IENZ,IDRUG)D/
                CALL DUMMY
                CALL DANG
            CALL CHAMOD
            WRITE(1,640)
640    FORMAT(' Return to MENU (or <return> to continue)? ', )>
            CALL ANS(YES)
            IF(YES)GOTO 800
700    CONTINUE
        ENDIF
800    EXCL1=EXCL
        MOL1=.TRUE.

        MOL2=LSTR
        IVEW(35)=ISTR
        IF(.NOT.RD)THEN
            WRITE(1,103)
103    FORMAT(' Write fit data to a file? ', )

            CALL ANS(YES)
            IF(.NOT.YES)RETURN
            WRITE(1,106)
106    FORMAT(' File name: ', )
            READ(1,107)FILE

```

```

107  FORMAT(A12)
      IF(FILE.EQ.'Q')RETURNE
      OPEN(2,FILE=FILE)
      WRITE(2,*)(IENZ(I),IDRUG(I),I=1,3) '
      IF(MINM)THEN
        DO 215 I=1,IROT
          WRITE(2,*)II(I),IP(I),IQ(I),JJ(I)'
215  CONTINUE
      ENDIF
      CLOSE(2)
      ENDIF
      RETURN
      SUBROUTINE VLJHRD(CONF)E
*      ,CONF,STAT
      DATA IBUG/11/

C
C      OPEN(12,FILE='POLEF')
C      DO 111 I=1,20
C 111 READ(12,*)(POLEF2(I,J),J=1,20)%
C      CLOSE(12)
C      SUBROUTINE FOR CALCULATING 6-12 POTENTIAL
C
C      IEX CONTAINS A LIST OF ATOMS FOR WHICH
C      SUMMATION OF 6-12 IS NOT TO TAKE PLACE
C
C      HOW MANY MOLECULES 1 OR 2
      VINTER=0.
      STAT=(IVEW(33).NE.0)
      E=SCAL(21)
      LIN=0
      IF(IVEW(33).EQ.3)LIN=1
      IF(IVEW(33).EQ.4)LIN=2
      HBD=(IVEW(34).NE.0)
      ITISHB=.FALSE.
      CUTOFF=SCAL(50)
      IF(STAT)CUTOFF=CUTOFF*2E
      IF(DBG) OPEN(IBUG,FILE='VLJ/')
      IF(DBG) WRITE(IBUG,1111) SQRT(CUTOFF)E
1111 FORMAT('/ Output from "hard" Non-bonded routine'
*      /' Note that Hydrogen bonds are not included in the total'
*      /' summation but X...H potentials (if any) are listed '>>
*      /' Cut off value=',F8.3,'A'
*      /1X,3X,2X,2X,8X,2X,'Total no. of summations',3X,'Potential')

C
      NMOL=1
      IF(NDRUG.NE.0) NMOL=2

C
C      SELECT WETHER INTRA+INTER , INTER OR INTRA
      IF(IVEW(35).EQ.2) GOTO 2000D/

C
C      LETS BEGIN WITH INTRA-MOLECULAR POTENTIAL
C
C      -----
C
      SUMTOT=0.
      SUMAT=0.
      NUMSUM=0
      NUMTOT=0
      II=1
      IF(EXCL1)II=2

```

ENERGY  
(VLJHRD)



```

DO 100 I=II,NMOL
VINTRA=0.0
VSTAT=0.0

C
C
SELECT APPROPRIATE MOLECULE PARAMETERS
IF(I.EQ.1)THEN
  NATOM=N
  N1=1
  I1=1
  I2=INFO(1)
ELSE
  NATOM=NDRUG+N
  N1=N+1
  I1=INFO(1)+1
  I2=INFO(1)+INFO(2)
ENDIF

C
C
NOW BEGIN SUMMATION
DO 90 J=N1,NATOM
IF(STAT)VST=0.0
DO 80 K=J,NATOM
IF(J.EQ.K) GO TO 80
C
HYD BONDING
IF(HBD.AND.INFO(I).NE.0)THEND/
  ITISHB=.FALSE.
  DO 81 IHB=I1,I2
    IF(IHX(IHB).EQ.J.AND.IHA(IHB).EQ.K) ITISHB=.TRUE.<.
    IF(IHA(IHB).EQ.J.AND.IHX(IHB).EQ.K) ITISHB=.TRUE.<.
    IF(ITISHB.AND..NOT.DBG.AND..NOT.STAT) GO TO 80E
    IF(ITISHB) GO TO 61
81 CONTINUE
  ENDIF

C
JFIX=J
IF(EXCL1)JFIX=J-N
DO 60 L=1,M(JFIX)
IF(IEX(JFIX,L).EQ.K) GO TO 80/
C 60 CONTINUE
C
IT'S ALL COOL TO SUM ATOM J & K%
61 CONTINUE
RJK=0.0
DO 45 L=1,3
  RJK=RJK+((X(L,J)-X(L,K))*(X(L,J)-X(L,K)))
C 45 CONTINUE
IF(RJK.GT.CUTOFF) GOTO 80
IF(RJK.LT.0.000001)THENE
  VINTRA=9999999.9
  WRITE(1,156)J,K
156 FORMAT(' 2 ATOMS COINCIDENT IN VLJHRD.INTRA:',2I5)<.
  GOTO 95
ENDIF
R6=RJK*RJK*RJK
C
ACTUAL DEF OF VMINIMUM=VM/(2*RO)&
L1=ITYPE(J)
L2=ITYPE(K)
RO=VDWR(L1)+VDWR(L2)

```

```

      RO=RO*RO*RO*RO*RO*RO
C      VM=362.187*POL(L1)*POL(L2)/(POLEF(L1)+POLEF(L2))I
      VM=362.187*POLEF2(L1,L2)E
      Z=1.0-(RO/(2.0*R6))
      V=-1.0*VM*Z/R6
C
      IF(STAT)THEN
        RJK=SQRT(RJK)
        IF(LIN.EQ.2)THEN
          E=SCAL(39)*RJK
        ELSEIF(LIN.NE.0)THEN
          IF(RJK.GT.7.0)E=4.0E
          IF(RJK.LT.7.0)E=0.75*RJK-1.25'
          IF(RJK.LT.3.0)E=1.0E
        ENDIF
        VST=VST+332.0*PC(J)*PC(K)/(RJK*E)L1)
      ENDIF
C
C      IF THIS IS A HB THEN RECORD AND GET THE HELL OUTI
      IF(HBD.AND.ITISHB)THEN
        IF(DBG)WRITE(BUG,1118) IAT(J),IAT(K),V
1118      FORMAT(1X,7X,A4,'...',2X,A4,14X,F10.1)
      ELSE
        VINTRA=VINTRA+V
        IF(DBG)THEN
          SUMAT=SUMAT+V
          NUMSUM=NUMSUM+1
        ENDIF
      ENDIF
C
C      80 CONTINUE
C
      IF(STAT)VSTAT=VSTAT+VSTE
      IF(DBG)THEN
        IF(NUMSUM.GT.0)WRITE(BUG,1112) J,IAT(J),RES(J),MRES(J),
*      NUMSUM,SUMAT,VST
1112      FORMAT(1X,I4,' ] ',2X,A4,2X,A3,I3,5X,I4,7X,2F11.2)<_
        SUMTOT=SUMTOT+SUMAT
        NUMTOT=NUMTOT+NUMSUM
        SUMAT=0.0
        NUMSUM=0
      ENDIF
C      90 CONTINUE
C
      95 IF(DBG)WRITE(BUG,1114) NUMTOT,SUMTOT,
1114      FORMAT(2X,'Total: Summations & Intra Pot = ',I6,3X,F10.1/)
C
      IF(.NOT.CONF)WRITE(1,1200)I,VINTRA
      IF(.NOT.CONF.AND.STAT)WRITE(1,1201)I,VSTAT
1200      FORMAT(' For MOL',I1,' VINTRA=',F10.3)
1201      FORMAT(' For MOL',I1,' VSTAT=',F10.3)
      SCAL(26)=SCAL(26)+VINTRA
      SCAL(24)=SCAL(24)+VSTAT
C      100 CONTINUE
C
C      DID WE WANT INTER MOLECULAR POTENTIAL AS WELL
      IF(IVEW(35).EQ.0.OR.IVEW(35).EQ.3) GO TO 2000
C

```

ENERGY

```

C      INTER-MOLECULAR POTENTIAL
C      -----
C
2000  SUMAT=0.0
      SUMTOT=0.
      VSTAT=0.0
      NIMSUM=0
      IF(DBG) WRITE(IBUG,1116)E
1116  FORMAT('/ Inter-Molecular Potentials [Rij < 12A] ')<.
C
      DO 130 J=1,N
      IF(.NOT.ACTSIT(J))GOTO 130
      IF(STAT)VST=0.0
      DO 120 K=N+1,N+NDRUG
C
C      HBONDS ?
      IF(HBD.AND.INFO(3).NE.0)THEND/
      ITISHB=.FALSE.
      DO 101 IHB=1,INFO(3)
      IF(IHC(IHB).EQ.J.AND.IHY(IHB).EQ.K)THEN
          ITISHB=.TRUE.
          IF(.NOT.DBG.AND..NOT.STAT)GOTO 120
          GOTO 105
101  ENDIF
      CONTINUE
      ENDIF
C
C
105  RJK=0.0
      DO 110 L=1,3
      RJK=RJK+((X(L,J)-X(L,K))*(X(L,J)-X(L,K)))
110  CONTINUE
C
      IF(RJK.GT.CUTOFF) GOTO 120
      IF(RJK.LT.0.000001)THENE
          VINTER=9999999.9
          WRITE(1,112)J,K
112  FORMAT(' 2 ATOMS COINCIDENT IN HRD.INTER:',2I5)'
          GOTO 135
      ENDIF
      R6=1/(RJK*RJK*RJK)
      L1=ITYPE(J)
      L2=ITYPE(K)
      R0=VDWR(L1)+VDWR(L2)
      R0=R0*R0*R0*R0*R0*R0
C      VM=362.187*POL(L1)*POL(L2)/(POLEF(L1)+POLEF(L2))'
      VM=362.187*POLEF2(L1,L2)E
      V=VM*R6*((R0*.5*R6)-1.)E
C
      IF(STAT)THEN
          RJK=SQRT(RJK)
          IF(LIN.EQ.2)THEN
              E=SCAL(39)*RJK
          ELSEIF(LIN.NE.0)THEN
              IF(RJK.GT.7.0)E=4.0E
              IF(RJK.LT.7.0)E=0.75*RJK-1.25'
              IF(RJK.LT.3.0)E=1.0E

```

ENERGY

```

      ENDIF
      VST=VST+332.0*PC(J)*PC(K)/(RJK*E)L1)
    ENDIF
  C   HBOND
      IF(HBD.AND.ITISHB)THEN
        IF(DBG)WRITE(IBUG,1118) IAT(J),IAT(K),V
      ELSE
        VINTER=VINTER+V
        IF(DBG)THEN
          SUMAT=SUMAT+V
          NUMSUM=NUMSUM+1
        ENDIF
      ENDIF
  C
  C 120 CONTINUE
      IF(STAT)VSTAT=VSTAT+VSTE
      IF(DBG)THEN
        IF(NUMSUM.GT.0)WRITE(IBUG,1113) J,IAT(J),RES(J),MRES(J)>
        *      ,NUMSUM,SUMAT,VST
1113  FORMAT(1X,I4,' ] ',A4,2X,A3,I3,4X,I5,7X,2F12.2),
        SUMTOT=SUMTOT+SUMAT
        NUMTOT=NUMTOT+NUMSUM
        SUMAT=0.0
        NUMSUM=0
      ENDIF
  C 130 CONTINUE
135  IF(DBG) WRITE(IBUG,1117) NUMTOT,SUMTOT,VSTAT
1117 FORMAT(2X,'Total: Summations & Inter Pot = ',I6,3X,2F12.2)
      IF(.NOT.CONF)CALL ENOUT(VINTER,'INTER=')

      IF(.NOT.CONF.AND.STAT)CALL ENOUT(VSTAT,'STAT2=')X
      SCAL(26)=SCAL(26)+VINTERE
      SCAL(24)=SCAL(24)+VSTATE
  C
200  IF(DBG) CLOSE(IBUG)
      RETURN
      SUBROUTINE FRAG(IGTYP,AUTO,HYD,IATM)AT
      INTEGER F1,F2
      DATA NUMATM/2,3,3,4,3,3/E
  C
      IGTYP=IGTYP
      ICOD(1)=IATM
      IF(AUTO) GOTO 131
      KILLF=0
      LIM=2
      IF(IVEW(44).EQ.3.OR.IVEW(44).EQ.4)LIM=1
      IF(IVEW(44).GT.4.AND.IVEW(44).LT.7)WRITE(1,13)
13  FORMAT(' Input 2 atoms to define fragment -''
      *,' atom 1 to atom 2 inclusive will be deleted')X
      ICNT=0
      CALL IDATOM(ICOD,ICNT,LIM,-1,NT1+NT2)
      IF(ICNT.LT.LIM)RETURN
      IATM=ICOD(1)
      IGTYP=ICOD(1)
  C
      GO TO (110,120,130,140,150,160,180) ,IVEW(44))
  C

```

FRAG

(EDIT MOL)

FRAG

(EDITMOL)

```

C      Break bond
110  L=ICOD(1)
      M=ICOD(2)
      DO 200 I=1,NBONDS+MBONDSE
      J=IA(I)
      K=IB(I)
      IF(J.EQ.L.AND.K.EQ.M) GO TO 210%
      IF(J.EQ.M.AND.K.EQ.L) GO TO 210%
200  CONTINUE
      WRITE(I,202)L,M
202  FORMAT(' No bond between atoms ',2I6)I
      RETURN

C
210  CONTINUE
CTRN  CALL TRANSF(2)
      CALL PENSEL(0,0.,0)
      CALL LINE(L,M)
      DO 220 J=1,NBONDS+MBONDSE
      IA(J)=IA(J+1)
      IB(J)=IB(J+1)
220  CONTINUE
      IF(I.GT.NBONDS) MBONDS=MBONDS-1%
      IF(I.LE.NBONDS) NBONDS=NBONDS-1%
      DO 241 I1=1,2
      DO 250 I=1,NCON(ICOD(I1))
      IF(ICON(I,ICOD(I1)).EQ.ICOD(3-I1))THEN
        NCON(ICOD(I1))=NCON(ICOD(I1))-1'
        DO 262 J=1,NCON(ICOD(I1))D/
262  ICON(J,ICOD(I1))=ICON(J+1,ICOD(I1))N
        GOTO 241
      ENDIF
250  CONTINUE
241  GO TO 3000

C
C      Make bond
120  I=NBONDS+MBONDS
      K=I
      IFROM=1
      IF(ICOD(1).GT.NT1) IFROM=NBONDS+1'
      DO 230 J=IFROM,I
      IF(K.EQ.0) GO TO 230
      IA(K+1)=IA(K)
      IB(K+1)=IB(K)

      K=K-1
230  CONTINUE
      IA(IFROM)=ICOD(1)
      IB(IFROM)=ICOD(2)
      IF(IFROM.EQ.1) NBONDS=NBONDS+1
      IF(IFROM.GT.1) MBONDS=MBONDS+1
      DO 125 I=1,2
      IF(NCON(ICOD(I)).GT.3)GOTO 126
      NCON(ICOD(I))=NCON(ICOD(I))+1/
125  ICON(NCON(ICOD(I)),ICOD(I))=ICOD(3-I)N
CTRN  CALL TRANSF(2)
      CALL PENSEL(4,0.,0)
      CALL LINE(ICOD(1),ICOD(2))
      GO TO 3000

```



FRAG

```

126 WRITE(1,127)IAT(ICOD(I)),ICOD(I)&
127 FORMAT(' Too many bonds for atom ',A4,I4)
      GOTO 3000
C
CTRNC 129 CALL TRANSF(2)
      129 CALL HELP
          CALL CHAMOD
C      ADD ATOMS IN HYDRO FORME
130 CONTINUE
    WRITE(1,1)
      1 FORMAT(/' ITYPE [0 for HELP]: ', )
      READ(1,2,ERR=129) IGTYP
      2 FORMAT(I5)
      IF(IGTYP.LE.0.OR.IGTYP.GT.6)GOTO 129
CTRNC  CALL TRANSF(2)
      CALL RUB
C
131 JOIN2=1
    IF(IGTYP.EQ.3.OR.IGTYP.EQ.6) JOIN2=2
    LIMID=NUMATM(IGTYP)
C
      ID=1
      IJOIN=1
      IOLD=0
239 CONTINUE
    DO 240 I=1,NCON(IATM)
      K=ICON(I,ICOD(IJOIN))
      IF(K.EQ.IOLD) GO TO 240E
      IF(ID.EQ.LIMID) GO TO 260
      ID=ID+1
      ICOD(ID)=K
      IJOIN=JOIN2
      IF(IJOIN.EQ.2) IOLD=ICOD(1)D/
      IF(IJOIN.EQ.2) GO TO 239E
240 CONTINUE
    IF(ID.EQ.LIMID) GOTO 260E
    I=ICOD(1)
    WRITE(1,245) IAT(I),I,IGTYP,ITYPE(I),HYD
245 FORMAT(' NO MATCH ATOM ',A4,' NO.',I4,
* ' ', IGTYP=' ,I2,' ITYPE=' ,I3,' hydrogen ',A4)X
    RETURN
260 IF(.NOT.AUTO.AND.DRAW)CALL RUB
    CALL HYDRO(AUTO,HYD,IATM,LIMID)%
    IF(AUTO)RETURN
    GO TO 3000
C
C      DELETE ATOM
140 CALL RUB
    IVEW(29)=0
    IF(ICOD(1).GT.NT1)IVEW(29)=2D/
    CALL DELATM(ICOD(1))
    CALL BONDS
    GO TO 3000
C
150 CONTINUE
    IVEW(29)=0
    IF(ICOD(1).GT.NT1)IVEW(29)=2D/
    ICD(1)=ICOD(1)

```

FRAG

```

      ICD(2)=ICOD(1)
      ICD(3)=ICOD(2)
      IDKIL=7
      NATOMS(IDKIL)=0
      IEORD=-1
      CALL SERCH
      IF(NATOMS(IDKIL).EQ.0)THEN
        WRITE(1,155)
155   FORMAT(' No atoms found - quit!')hyd
        RETURN
      ENDIF
      CALL RUB
      DO 165 KILLF=1,NATOMS(IDKIL)D/
        ICD(1)=JCON(IDKIL,KILLF)
      DO 166 J=KILLF,NATOMS(IDKIL)D/
        JCK=JCON(IDKIL,J)
        IF(JCK.LE.ICOD(1)) GO TO 166D/
        JCON(IDKIL,J)=JCK-1
166   CONTINUE
      CALL DELATM(ICOD(1))
C 165 CONTINUE
      CALL BONDS
      GOTO 3000
C
C 160 CONTINUE
C Delete specific fragment
      F1=ICOD(1)
      F2=ICOD(2)
      IF(F1.LE.NT1)THEN
        N1=1
        N2=NT1
        IVEW(29)=0
      ELSE
        N1=NT1+1
        N2=NT1+NT2
        IVEW(29)=2
      ENDIF
      IKILL(1)=F1
      ICBR(1)=1
      ICNT=1
      GOTF2=.FALSE.
      ISTOP=-99
      ICUR=F1
      DO 10 I=1,100
10   IBR(I)=1
      DO 15 I=N1,N2
15   TERM(I)=(NCON(I).EQ.1)
C
1000 CALL BONDED(IC,ICUR,GOTF2)
      IF(GOTF2)GOTO 2000
C   If don't find suitable atom(s) go back to last "branch" and find a suitable
C   route
      WRITE(1,1001)IC,ICUR
1001 FORMAT(' IC,ICUR:',2I6)E
      IF(IC.LT.IBR(ICNT))THENE
        ICNT=ICNT-1
        IF(ICNT.LE.ISTOP)THENE

```

FRAG

```

        IOLD=ISTOP-1
        CALL NWPATH(IOLD,OK)E
        IF(OK)GOTO 1000
        CALL CHAMOD
        WRITE(1,2710)
        RETURN
    ENDIF
DO 20 I=ICNT,1,-1
C   If atom contains a branch that hasn't been tried then:<
    IF(IBR(I).LT.ICBR(I))THEND/
        ICNT=I+1
        IBR(I)=IBR(I)+1
        ICUR=IBD(IBR(I),I)
        IKILL(ICNT)=ICUR
        GOTO 1000
    ELSE
        IBR(I)=1
20  CONTINUE
    WRITE(1,21)F2
21  FORMAT(' Searched all branches and can't find atom',I5)
    WRITE(1,22)(IKILL(I),I=1,44)
22  FORMAT(' IKILL:',18I4)E
    RETURN
ELSE
C   Go on to next atom of branch
    ICBR(ICNT)=IC
    IKILL(ICNT+1)=IBD(IBR(ICNT),ICNT)d c
    ICNT=ICNT+1
    ICUR=IKILL(ICNT)
    GOTO 1000
ENDIF
C
2000 CONTINUE
    IOLD=ICNT
    ICNT=ICNT+1
    IKILL(ICNT)=F2
    CALL ADTERM
    CALL DRFRAG
    CALL CHAMOD
    WRITE(1,2010)
2010 FORMAT(' Search for a different route between at1 & at2?',/,
    * ' (otherwise these atoms will be deleted): ', )
    CALL ANS(YES)
    IF(YES)THEN
CTRN        CALL TRANSF(2)
            CALL RUB
            CALL DUMMY
            CALL NWPATH(IOLD,OK)
            GOTF2=.FALSE.
            IF(OK)GOTO 1000
            CALL CHAMOD
            WRITE(1,2710)
2710  FORMAT(' Can't find another path - quit!'))
            RETURN
    ENDIF
CTRN        CALL TRANSF(2)
C   Sort IKILL into ascending order/

```

~~SURF~~  
FRAS

```
IFAIL=0
CALL M01AQF(IKILL,1,ICNT,IFAIL)
IF(IFAIL.NE.0)WRITE(1,2020)IFAIL
2020 FORMAT(' ERROR in M01AQF, IFAIL=',I5)
CALL RUB
DO 2500 I=1,ICNT
  IDEL=IKILL(I)-I+1
  CALL DELATM(IDEL)
2500 CONTINUE
  CALL BONDS
  GOTO 3000

C
C  Split molecule
180 ICD(1)=ICOD(1)
  ICD(2)=ICOD(1)
  ICD(3)=ICOD(2)
  ICD(4)=ICOD(2)
  IDKIL=9
  DO 185 I=1,NT1+NT2
185 SORT(I)=.FALSE.
  CALL SERCH
  NT2=NATOMS(IDKIL)
  NT1=NT1-NT2
  N1=1
  N2=NT1+NT2
  CALL TRMEND(.TRUE.)
  IVEW(29)=1
  CALL BONDS
3000 CONTINUE

C
  IF(ICOD(1).LE.NT1)THEN
C  Redefine Metal ion and H2O pos'ns
  IVEW(48)=0
  IVEW(49)=0
  DO 3005 I=1,NT1
  IF(ITYPE(I).GT.17.AND.IVEW(49).LT.1)IVEW(49)=I
  IF(RES(I).EQ.'HOH')GOTO 3006
3005 CONTINUE
3006 IF(RES(I).EQ.'HOH')IVEW(48)=I
  ENDIF
  IF(IVEW(40).EQ.0)RETURN
  CALL CHAMOD
  WRITE(1,3040)
3040 FORMAT(' Change energy parameters? ', )
  CALL ANS(YES)
  IF(YES)CALL SETUPE
  RETURN
  SUBROUTINE BONDED(IC,ICUR,GOTF2)
  INTEGER F1,F2

C
  I2=2
  IF(ICUR.EQ.F1)I2=3
  IC=0
  DO 100 I=1,NCON(ICUR)
  K=ICON(I,ICUR)

C  Make sure that not going round in circles
  DO 50 L=1,ICNT-1
  IF(K.EQ.IKILL(L))GOTO 100
50 CONTINUE
  IF(.NOT.TERM(K))THEN
    IC=IC+1
    IBD(IC,ICNT)=K
  ENDIF
```

FRAG

```

        IF(K.EQ.F2)GOTF2=.TRUE.
        IF(GOTF2.OR.IC.GT.I2)RETURN
100 CONTINUE
    RETURN
    SUBROUTINE DRFRAG
C   Draw chosen fragment for deletion
C
        CALL PENSEL(4,0.,0)
        DO 200 I=1,ICNT
            K=IKILL(I)
            DO 100 J=1,NCON(K)
                L=ICON(J,K)
                IF(L.GT.K)CALL LINE(L,K)
100 CONTINUE
200 CONTINUE
    RETURN
    SUBROUTINE DELATM(I)
C
C   REMOVE ATOMS
    DO 300 J=I,NT1+NT2
        K=J+1
        DO 290 L=1,3
            DUMX(L,J)=DUMX(L,K)
290  X(L,J)=X(L,K)
            IAT(J)=IAT(K)
            ITYPE(J)=ITYPE(K)
            PC(J)=PC(K)
            SPE(J)=SPE(K)
300 CONTINUE
        IF(I.GT.NT1) NT2=NT2-1
        IF(I.LE.NT1) NT1=NT1-1
C
C   REDO LABEL'S AND DIST LIST
    NLAB=0
    NDIST=0
C
    RETURN
    SUBROUTINE ADTERM
    INTEGER F1,F2
C
        ICNT2=ICNT
        DO 200 I=1,ICNT2
            ICUR=IKILL(I)
            DO 100 L=1,NCON(ICUR)
                K=ICON(L,ICUR)
                IF(TERM(K).AND.K.NE.F1.AND.K.NE.F2)THEN
                    ICNT=ICNT+1
                    IKILL(ICNT)=K
                ENDIF
100 CONTINUE
200 CONTINUE
        WRITE(1,1)ICNT2,ICNT
1  FORMAT(' OLD NO.: ',I5,' NEW: ',I5)
    RETURN
    SUBROUTINE NWPATH(IOLD,OK)
        ICNT=IOLD
        OK=.FALSE.
        DO 100 I=ICNT,1,-1
            IF(IBR(I).LT.ICBR(I))THEN
C   ISTOP used to stop fragment being 'reversed' beyond ISTOP
                ISTOP=I
                ICNT=I+1
                IBR(I)=IBR(I)+1

```



FRAG

```

    ICUR=IBD(IBR(I),I)
    IKILL(ICNT)=ICUR
    OK=.TRUE.
    RETURN
ENDIF
100 CONTINUE
RETURN
SUBROUTINE CONCAT
C
    IF(NT2.LT.1)RETURN
    NT1=NT1+NT2
    NT2=0
    NBONDS=NBONDS+MBONDS
    MBONDS=0
    MOL2=.FALSE.
    MOL1=.TRUE.
    CALL STAR(1,MOL1)
    CALL STAR(2,MOL2)
    CALL CHAMOD
    WRITE(1,10)
10 FORMAT(' MOL1 & MOL2 atoms joined.',/, 'Re-calculate bonding'
    * , 'matrices? ', )
    IVEW(29)=1
    CALL ANS(YES)
    IF(YES)CALL BONDS
    RETURN
SUBROUTINE INVERT
EQUIVALENCE (INVN(1),I4),(INVN(2),I3),(INVN(3),I2),(INVN(4),I1)
C
C1 Input atom to be inverted
    IC=0
    CALL IDATOM(ICD,IC,1,1,NT1+NT2)
    IF(IC.LT.1)RETURN
    INV=ICD(1)
    IF(NCON(INV).LT.4)THEN
        WRITE(1,10)NCON(INV),INV
10    FORMAT(' Only ',I1,' atoms bonded to atom ',I4,' - quit')
        RETURN
    ELSE
        DO 100 I=1,4
100    INVN(I)=ICON(I,INV)
    ENDIF
C
    ID=6
    CALL DIHED(I1,INV,I3,I4,OM1)
    CALL DIHED(I2,INV,I3,I4,OM2)
    OM1=OM1*3.1415927/180.0
    OM2=OM2*3.1415927/180.0
    ICODES(1)=INV
    ICODES(2)=INV
    ICODES(3)=I1
    CALL SERCH
    ANG=OM2-OM1
    CALL RUB
    IP(ID)=I3
    IQ(ID)=INV
    CALL TORROT(ANG)
C
    ID=7
    ICODES(3)=I2
    CALL SERCH
    ANG=OM1-OM2
    IP(ID)=I3

```

MND0IN

```

      IQ(ID)=INV
      CALL TORROT(ANG)
      RETURN
      SUBROUTINE MND0IN
      DATA ITYP/1,6*6,4*7,3*8,2*16,15,3*99/
C
      CALL CHAMOD
      WRITE(1,10)
10  FORMAT(' This routine converts cartesian coordinates into '
      *, 'internal coordinates',/, ' ready for MND0 input on the Cray')
      N1=NT1+1
      N2=NT1+NT2
      NMOL=2
      IF(NT1.GT.50)THEN
        IF(NT2.LT.4)THEN
          WRITE(1,20)NT1
20    FORMAT(1X,I5,' is too many atoms for MND0 input')
          RETURN
        ELSEIF(NT2.GT.50)THEN
          WRITE(1,20)NT2
          RETURN
        ENDIF
      ELSE
        IF(NT2.LT.4)THEN
          N1=1
          N2=NT1
          NMOL=1
        ELSE
          WRITE(1,30)
30    FORMAT(' Default is to output MOL2 coord's, do you wish'
      *, ' to change this? ', )
          CALL ANS(YES)
          IF(YES)THEN
            N1=1
            N2=NT1
            NMOL=1
          ENDIF
        ENDIF
      ENDIF
      NOHYD=.TRUE.
      DO 11 I=N1,N2
        IF(ITYPE(I).EQ.1)NOHYD=.FALSE.
        IF(ITYPE(I).GT.17)WRITE(1,12)I,ITYPE
12  FORMAT(' WARNING - atom ',I4,' type ',I2,' not parameterised '
      *, 'and will be treated',/, ' as a dummy atom.')
11  CONTINUE
        IF(NOHYD)WRITE(1,35)NMOL
35  FORMAT(15X,' ***** WARNING *****',//,
      *, 'No hydrogens present in MOL',I1)
C
      IF(N1.EQ.1)THEN
        IVEW(29)=1
      ELSE
        IVEW(29)=2
      ENDIF
C
C  INITIALISE AND TREAT FIRST 3 ATOMS SEPARATELY
      NATS=N2-N1+1
      I1=1
      DO 40 I=1,3
        R(I)=0.
        BA(I)=0.
        TOR(I)=0.

```

```

      DO 40 J=1,3
40  IJKS(J,I)=0
      NATM(1)=N1
C    Try initially to obtain connectivity w/o re-ordering atoms
      FAIL=.FALSE.
      ORDER=.FALSE.
60  IF(ORDER)THEN
C    Re-order atoms so that TERM ats are at end
      CALL TRMEND(.FALSE.)
      CALL NXTAT(2,FAIL)
      IF(FAIL)WRITE(1,401)
      CALL NXTAT(3,FAIL)
      IF(FAIL)WRITE(1,401)
      ELSE
        DO 80 I=1,NATS
80    NATM(I)=N1+I-1
      ENDIF
      IATM(1)=ITYP(ITYPE(N1))
      IATM(2)=ITYP(ITYPE(NATM(2)))
      IATM(3)=ITYP(ITYPE(NATM(3)))
      CALL DIST(NATM(1),NATM(2),R(2))
      CALL DIST(NATM(2),NATM(3),R(3))
      CALL ANGLE(NATM(1),NATM(2),NATM(3),BA(3))
C
      DO 100 N=4,NATS
      IF(ORDER)CALL NXTAT(N,FAIL)
      IF(FAIL)THEN
        WRITE(1,401)N
        GOTO 120
      ENDIF
      IATM(N)=ITYP(ITYPE(NATM(N)))
      CALL ZMAT(N,IFAIL)
      IF(IFAIL.NE.0)THEN
        IF(.NOT.ORDER)THEN
          ORDER=.TRUE.
          WRITE(1,403)
403    FORMAT(' Atoms re-ordered to obtain connectivity')
          GOTO 60
        ENDIF
        WRITE(1,402)IFAIL,N
      ELSE
        CALL DIST(NATM(N),IJKS(1,N),R(N))
        CALL ANGLE(NATM(N),IJKS(1,N),IJKS(2,N),BA(N))
        CALL DIHED(NATM(N),IJKS(1,N),IJKS(2,N),IJKS(3,N),TOR(N))
C    Invert torsion angle because of different MNDO default
        TOR(N)=-1.0*TOR(N)
      ENDIF
100  CONTINUE
C
120  CONTINUE
      IF(ORDER)THEN
C    Re-order IJKS into same order as NATM
        DO 300 I=1,3
        DO 300 N=4,NATS
        DO 400 M=1,NATS
          IF(IJKS(I,N).EQ.NATM(M))THEN
            IJKS(I,N)=M
            GOTO 300
          ENDIF
400    CONTINUE
300    CONTINUE
        ELSEIF(N1.GT.1)THEN
          DO 450 I=1,3

```

MNDOIN

```

      DO 450 J=4,NATS
450    IJKS(I,J)=IJKS(I,J)-N1+1
      ENDIF
C
      DO 500 N=1,NATS
      WRITE(1,510)NATM(N),IATM(N),R(N),I1,BA(N),I1,TOR(N),I1,(IJKS(
*      I,N),I=1,3)
510    FORMAT(I4,3(I4,F10.3),4I4)
500    CONTINUE
      NORB=0
      DO 512 I=N1,N2
      IF(ITYPE(I).NE.1)NORB=NORB+1
512    CONTINUE
      NORB=NORB*3+NATS
515    WRITE(1,511)NATS,NORB
511    FORMAT(' No. of atoms=',I4,' no. of orbitals=',I4,
*    '/', ' Molecular charge: ', )
      READ(1,516,ERR=515)ICHG
516    FORMAT(I2)
      IF(ICHG.LT.-9.OR.ICHG.GT.9)ICHG=0
      WRITE(1,521)
521    FORMAT(' CNDO? (default MNDO) ', )
      METH='MNDO '
      READ(1,522)MT
522    FORMAT(A1)
      IF(MT.EQ.'C'.OR.MT.EQ.'c')METH='CNDO2'
514    WRITE(1,517)
517    FORMAT(' Convergence criterion for Geom. Opt. (default 0.02):', )
      READ(1,518,ERR=514)CONV
518    FORMAT(F10.6)
      IF(CONV.LT.0.000001)CONV=0.02
      WRITE(1,520)
520    FORMAT(' Name of output file: ', )
      READ(1,530)FILE
530    FORMAT(A12)
      IF(FILE.EQ.'          ')FILE='MNDO.DAT '
      OPEN(10,FILE=FILE)
      ITIM=NATS*NATS*0.035+2
      IF(ITIM.LT.10)ITIM=10
      WRITE(10,540)ITIM,METH,CONV,ICHG,(TITLE(NMOL,I),I=1,8)
540    FORMAT(' JOB,US=GPAP571,JN=PRMNDO,T=',I2,',MFL=150000,SD.',/,
*    'ACCESS,DN= BLD,PDN=MNDBIN,ID=GPAP571.',/,
*    'LDR,SET=ZERO.',/,
*    'EXIT,U.',/, 'DISPOSE,DN=TAPE11,DC=ST,',
*    'TEXT=' 'UNIT=DISC,DSN=GPAP571.TPRM.C,' ' ',/,
*    'DISP=(NEW,CATLG,DELETE),DCB=DCB.FB80' ' ',/,
*    '/EOF',/,A5,/, ' 1',28X,F10.6,'coords from Prime',/,I2,30X,8A4)
      DO 600 N=1,NATS
      WRITE(10,550)IATM(N),R(N),I1,BA(N),I1,TOR(N),I1,(IJKS(
*      I,N),I=1,3),N
550    FORMAT(3(I4,F10.3),4I4,I6)
600    CONTINUE
      WRITE(10,610)
610    FORMAT('0 0 0 0 0 0 0 0 0 0',/, '99',/, '/EOF')
      CLOSE(10)
      RETURN
401    FORMAT(' Failed to find atom ',I4)
402    FORMAT(' Failed to find pointer ',I1,' for atom ',I4)
      SUBROUTINE NXTAT(N,FAIL)
C    Search for the next lowest new atom bound to atom (N-1).
C    If no new atom is found repeat for atom (N-2), (N-3) etc.
C
      M=N-1

```

MNDOWN

```

1 NM1=NATM(M)
  FAIL=.FALSE.
  DO 100 I=1,NCON(NM1)
    K=ICON(I,NM1)
C   check that K not already recorded
      DO 40 L=1,N-1
        IF(K.EQ.NATM(L))GOTO 100
40    CONTINUE
      NATM(N)=K
      RETURN
100 CONTINUE
C
  M=M-1
  FAIL=(M.LT.1)
  IF(FAIL)RETURN
  GOTO 1
  SUBROUTINE ZMAT(N,IFAIL)
C   Find I,J & K for the MNDO Z-matrix.
C   I is the lowest atom bound to atom N.
C   J is the lowest atom bound to atom I.
C   K is the lowest atom bound to atom J.
C
  IFAIL=0
  Istor(1)=NATM(N)
C
  DO 1000 IJK=1,3
    IF(IJK.EQ.1)ICUR=NATM(N)
    IF(IJK.GT.1)ICUR=IJKS(1,N)
C
    1 DO 100 I=1,NCON(ICUR)
      K=ICON(I,ICUR)
C   J & K must already be in NATM.
      KOK=.FALSE.
      DO 20 L=1,N
        IF(K.EQ.NATM(L))KOK=.TRUE.
20    CONTINUE
      IF(KOK)THEN
C   check that K not already included
        DO 40 L=1,IJK
          IF(K.EQ.Istor(L))GOTO 100
40      CONTINUE
          Istor(IJK+1)=K
          IJKS(IJK,N)=K
          GOTO 1000
        ENDIF
100 CONTINUE
C
    IF(IJK.LT.3.OR.ICUR.EQ.IJKS(2,N))THEN
      IFAIL=IJK
      RETURN
    ELSE
      ICUR=IJKS(2,N)
      GOTO 1
    ENDIF
1000 CONTINUE
C
  RETURN
SUBROUTINE POSN(IGEN,IUPTO,HYD,ICOD)
C
  I1=IUPTO+IGEN+1
  IF(IUPTO.GT.NT1)THEN
    NT2=NT2+IGEN
    I2=NT1+NT2

```

POSN



POSN

```

        LIM=I2+1
        IVEW(29)=2
    ELSE
        NT1=NT1+IGEN
        I2=NT1
        LIM=NT1+1
        IVEW(29)=0
    ENDIF
C
C
        I3=I2
        DO 10 II=I1,I3
        L=I2-IGEN
        DO 20 J=1,3
        X(J,I2)=X(J,L)
        DUMX(J,I2)=DUMX(J,L)
20    CONTINUE
        IAT(I2)=IAT(L)
        ITYPE(I2)=ITYPE(L)
        PC(I2)=PC(L)
        MRES(I2)=MRES(L)
        RES(I2)=RES(L)
        SPE(I2)=SPE(L)
        LAB(I2)=LAB(L)
        I2=I2-1
10    CONTINUE
C
        II=0
        DO 30 I=IUPTO+1,IUPTO+IGEN
        II=II+1
        IAT(I)=HYD(I-IUPTO)
        RES(I)=RES(IUPTO)
        SPE(I)=SPE(IUPTO)
        LAB(I)=.FALSE.
        MRES(I)=MRES(IUPTO)
        ITYPE(I)=ICOD(I-IUPTO)
        DO 30 J=1,3
        X(J,I)=XHYD(J,II)
        DUMX(J,I)=XHYDX(J,II)
30    CONTINUE
C
        KRES=MRES(IUPTO)+1
        DO 40 I=KRES,NRES
40    IRES(I)=IRES(I)+IGEN
C
        CALL BONDS
C
        WRITE(1,50) IGEN,IAT(IUPTO),IUPTO,HYD(1),RES(IUPTO),MRES(IUPTO)
50    FORMAT(1X,I4,' atoms added to ',A4,' atom',I5,
        *      ' atom type: ',A4,' residue: ',A3,I4)
C
        IUPTO=IUPTO+IGEN
C
        RETURN
        SUBROUTINE ADDH
C
        IVEW(44)=3
        L=.TRUE.
        ICNT=0
C
        ITYP=3
        CALL FRAG(ITYP,L,'HN ',1)
        DO 100 I=2,NT1
        ITYP=1

```

ADDH

ADDH  
(Proteins)

```

IF(ITYPE(I).EQ.13) CALL FRAG(ITYP,L,'HO1 ',I)
ITYP=2
CRES=RES(I)
IF(IAT(I).EQ.'N '.AND.CRES.NE.'PRO')
* CALL FRAG(ITYP,L,'HN1 ',I)
IF(IAT(I).EQ.'NE '.OR.IAT(I).EQ.'NE1 ')CALL FRAG(ITYP,L,
* 'H1N ',I)
IF(IAT(I).EQ.'NE2 '.AND.CRES.EQ.'HIS')CALL FRAG(ITYP,L,'H1NH'
* ,I)
ITYP=3
IF(IAT(I).EQ.'ND2 '.OR.IAT(I).EQ.'NH1 '.OR.
* IAT(I).EQ.'NH2 ') CALL FRAG(ITYP,L,'H2N ',I)
IF(IAT(I).EQ.'NE2 '.AND.CRES.EQ.'GLN')CALL FRAG(ITYP,L,'H2NG'
* ,I)
ITYP=6
IF(IAT(I).EQ.'NZ ') CALL FRAG(ITYP,L,'H6N ',I)
100 CONTINUE
C
WRITE(1,110)ICNT
110 FORMAT(1X,I4,' additions - now repositioning arrays')
NOLD=NT1
CALL HPOSN(1,NOLD)
WRITE(1,150) NOLD,NT1
150 FORMAT(' Old number of atoms=',I5,' New=',I5,/,
* ' Now writing to file H.OUT')
OPEN(10,FILE='H.OUT')
WRITE(10,240)(TITLE(1,I),I=1,20)
240 FORMAT(20A4)
DO 200 I=1,NT1
WRITE(10,250) IAT(I),RES(I),MRES(I),
* (X(M,I),M=1,3),PC(I),ITYPE(I),SPE(I)
250 FORMAT(A4,A3,I4,4F8.3,I4,1X,L1)
200 CONTINUE
CLOSE(10)
RETURN
SUBROUTINE HPOSN(N1,N2)
C
C ICNT now used for the total no. of atoms added
ICNT=1
ICUR=N1-1
C
DO 100 I=N1,N2
ICUR=ICUR+1
DO 10 K=1,3
10 XX(K,ICUR)=X(K,I)
SSPE(ICUR)=SPE(I)
RRES(ICUR)=RES(I)
MMRES(ICUR)=MRES(I)
IIAT(ICUR)=IAT(I)
IITYP(ICUR)=ITYPE(I)
C
IF(I.EQ.NATCHH(ICNT))THEN
DO 20 J=1,IGEN(ICNT)
DO 30 K=1,3
30 XX(K,ICUR+J)=HX(K,J,ICNT)
SSPE(ICUR+J)=SPE(I)
RRES(ICUR+J)=RES(I)
MMRES(ICUR+J)=MRES(I)
IIAT(ICUR+J)=HNAM(ICNT)
IITYP(ICUR+J)=1
20 CONTINUE
ICUR=ICUR+IGEN(ICNT)
ICNT=ICNT+1

```

```

ENDIF
100 CONTINUE
ICNT=ICNT-1
C
IF(NT2.NE.0.AND.N1.LT.NT1)THEN
  IDIF=ICUR-NT1
  DO 120 I=NT1+NT2,NT1+1,-1
    J=I+IDIF
    SPE(J)=SPE(I)
    STUPID=RES(I)
    RES(J)=STUPID
    ITYPE(J)=ITYPE(J)
    MRES(J)=MRES(I)
    IAT(J)=IAT(I)
    DO 120 J=1,3
120   X(K,J)=X(K,I)
  ENDIF
  IF(N1.LT.NT1)THEN
    NT1=ICUR
  ELSE
    NT2=ICUR-NT1
  ENDIF
  DO 200 I=1,NT1
    DO 210 K=1,3
210   X(K,I)=XX(K,I)
    SPE(I)=SSPE(I)
    RES(I)=RRES(I)
    MRES(I)=MMRES(I)
    IAT(I)=IIAT(I)
    ITYPE(I)=IITYP(I)
200 CONTINUE
C ADJUST RES END POINTS (IRES)
  IRES(1)=1
  JCNT=1
  DO 300 I=2,NT1
    IF(MRES(I).NE.MRES(I-1))THEN
      JCNT=JCNT+1
      IRES(JCNT)=I
    ENDIF
300 CONTINUE
  IRES(JCNT+1)=NT1
  IF(JCNT.NE.NRES)WRITE(1,310)JCNT,NRES
310 FORMAT( 'New NRES:',I4,' old:',I4)
  RETURN
  SUBROUTINE SMLADH(M1,M2)
  DATA KCON/1,3,4*4,3,2,3,4,99,1,2,1,2,1,4*99/
C
  T=.TRUE.
  N1=M1
  N2=M2
  NMOL=1
  IF(N1.GT.NT1)NMOL=2
C ICNT= no. of atoms to which H(s) added
  ICNT=0
C
  DO 1000 I=N1,N2
    JCON=KCON(ITYPE(I))
    IF(JCON.EQ.1)GOTO 1000
    IF(JCON.NE.99)THEN
      LCON=NCON(I)
      IF(LCON.GE.JCON)GOTO 1000
    IF
6 Have found atom suitable for H-addn, find ITYP and add H(s)

```

```

IF(LCON.EQ.1.AND.JCON.EQ.2)THEN
  ITYP=1
  CALL FRAG(ITYP,T,'HA ',I)
ELSEIF(LCON.EQ.2.AND.JCON.EQ.3)THEN
  ITYP=2
  CALL FRAG(ITYP,T,'HA ',I)
ELSEIF(LCON.EQ.1.AND.JCON.EQ.3)THEN
  ITYP=3
  CALL FRAG(ITYP,T,'HA ',I)
ELSEIF(LCON.EQ.3.AND.JCON.EQ.4)THEN
  ITYP=4
  CALL FRAG(ITYP,T,'HA ',I)
ELSEIF(LCON.EQ.2.AND.JCON.EQ.4)THEN
  ITYP=5
  CALL FRAG(ITYP,T,'HA ',I)
ELSEIF(LCON.EQ.1.AND.JCON.EQ.4)THEN
  ITYP=6
  CALL FRAG(ITYP,T,'HA ',I)
ENDIF
ELSE
  WRITE(1,120)IAT(I),I,ITYPE(I)
120  FORMAT(' Not programmed for atom ',A4,' no.',I5,' type',
  *  I3,/, ' this atom must be treated separately.')
ENDIF
1000 CONTINUE
C
IF(ICNT.NE.0)THEN
  WRITE(1,110)ICNT
110  FORMAT(1X,I4,' additions - now repositioning arrays')
  NOLD=NT1
  IF(NMOL.EQ.2)NOLD=NT2
  CALL HPOSN(N1,N2)
  IF(NMOL.EQ.1)THEN
    N3=1
    N4=NT1
  ELSE
    N4=NT1+NT2
    N3=NT1+1
  ENDIF
  WRITE(1,150) NOLD,N4
150  FORMAT(' Old number of atoms=',I5,' New=',I5,/,
  *  ' Now writing to file H.OUT')
  OPEN(10,FILE='H.OUT')
  WRITE(10,240)(TITLE(NMOL,I),I=1,20)
240  FORMAT(20A4)
  DO 200 I=N3,N4
    WRITE(10,250) IAT(I),RES(I),MRES(I),
    *  (X(M,I),M=1,3),PC(I),ITYPE(I),SPE(I)
250  FORMAT(A4,A3,I4,4F8.3,I4,1X,L1)
200  CONTINUE
    CLOSE(10)
    IF(NMOL.EQ.2)JUST2=T
    CALL BONDS
    JUST2=.FALSE.
    DO 2000 I=N1,NT1+NT2
      DUMX(1,I)=X(1,I)-XMN(1)
      DUMX(2,I)=X(2,I)-XMN(2)
      DUMX(3,I)=X(3,I)-XMN(3)
      XX=A(1,1)*DUMX(1,I)+A(1,2)*DUMX(2,I)+A(1,3)*DUMX(3,I)
      YY=A(2,1)*DUMX(1,I)+A(2,2)*DUMX(2,I)+A(2,3)*DUMX(3,I)
      ZZ=A(3,1)*DUMX(1,I)+A(3,2)*DUMX(2,I)+A(3,3)*DUMX(3,I)
      DUMX(1,I)=XX+F(1)
      DUMX(2,I)=YY+F(2)

```

READ AT

```

      DUMX(3,I)=ZZ+F(3)
2000  CONTINUE
      ELSE
        WRITE(1,270)
270   FORMAT(' ***** No atoms added *****')
      ENDIF
      RETURN
      SUBROUTINE READAT(BOND)
1000  WRITE(1,2)
      2  FORMAT(/,' Menu',T40,' MOL2 input',//,' SEe data',
        * T40,' Add hydrogen atoms',//,' FILE input',T40,' CHarges input'
        * ,//,' NOn-standard input file format',T40,
        * ' PRe-processed input file (binary)',//,
        * ' EXclude hydrogens on input',T40,' CLear data and re-input'
        * ,//,' RETurn to operating system',T41,' INput rot/trans matrices'
        * ,//,' STop',T40,' GO - to next step (bonding or SETUPE)',
        */
        *,'_____')
9600  WRITE(1,1) NMOL
      1  FORMAT(/' Data Input for Molecule ',I1,
        READ(1,3,ERR=9600) JFUNCT
      3  FORMAT(A2)
        JST=.FALSE.
        DO 10 I=1,15
          IF(JFUNCT.EQ.IFUNCT(I)) GO TO 20
10     CONTINUE
        WRITE(1,31)
      31  FORMAT(' Input 1st 2 letters of function name (capitals!)')
        GO TO 9600
C
      20  GO TO (1000,1500,2000,3000,3500,4000,5000,6000,6500,7000,
        *8000,8500,9000,9500,1000) I
C
C   Input atomic charges
1400  WRITE(1,1410)FILE
1410  FORMAT(' Data error in file ',A12)
        BACKSPACE IO2
        READ(IO2,4)COORD
        WRITE(1,1420)COORD
1420  FORMAT(' Data: ',18A4)
        CLOSE(IO2)
1500  IF(NT1.EQ.0)GOTO 9600
        WRITE(1,103)
        READ(1,102)FILE
        IF(FILE.EQ.SP12)GOTO 1500
        IF(FILE.EQ.'Q')GOTO 9600
        OPEN(IO2,FILE=FILE,STATUS='OLD',ERR=1500)
        IF(NMOL.EQ.1)THEN
          N1=1
          N2=NT1
        ELSE
          N1=NT1+1
          N2=NT1+NT2
        ENDIF
        WARN=.FALSE.
        DO 1600 I=N1,N2
          READ(IO2,*,END=1700,ERR=1400)PC(I)
          IF(PC(I).GT.0.1.AND.ITYPE(I).GT.6.AND.ITYPE(I).LT.18)WARN=.TRUE.
1600  CONTINUE
          IF(WARN)WRITE(1,1520)NMOL
1520  FORMAT(' COMMENT: at least 1 electronegative atom in MOL',I1,
        *' has been assigned a positive charge')
        CLOSE(IO2)

```



REMARK

```

      GOTO 9600
C
1700 WRITE(1,1710)N2-N1+1
1710 FORMAT(' File too short - ',I4,' atoms required')
      CLOSE(IO2)
      GOTO 9600
C
C      INPUT FOR SECOND MOLECULE
2000 IF(NT1.EQ.0)THEN
      WRITE(1,2010)
2010  FORMAT(' No MOL1 data read in yet!')
      GOTO 9600
      ENDIF
      IF(MOL2)N=NT1
      NMOL=2
      MOL2=.TRUE.
      IF(EMP)WRITE(1,2020)
2020 FORMAT(' Reverted back to standard file format')
      EMP=.FALSE.
      NSTD=.FALSE.
      GO TO 9600
C
C      SEE INPUTED DATA
C      ADD HYDROGENS
3500 IF(NT1.EQ.0) GOTO 9600
      IF(DONE)WRITE(1,3510)
3510 FORMAT(' Hydrogens cannot be added to a preprocessed file')
      IF(DONE)GOTO 9600
      WRITE(1,26)
26  FORMAT(' This needs to be set up beforehand - do you wish to',
*      ' continue? ', )
      CALL ANS(YES)
      IF(.NOT.YES)GOTO 9600
      IF(ITYPE(1).NE.9)WRITE(1,261)
261 FORMAT(' This only works for proteins with standard residues!')
      IF(ITYPE(1).NE.9)GOTO 9600
      CALL BONDS
      CALL ADDH
      WRITE(1,3520)
3520 FORMAT(' Need to re-calculate bonding matrices')
      GO TO 9600
C
C      NO HYDS
6500 NOH=.NOT.NOH
      WRITE(1,6510)
6510 FORMAT(' Hydrogen atoms excluded on input')
      GO TO 9600
C
C *** INPUT FROM DATA FILE
3990 CLOSE(IO2)
4000 IF(DONE) GOTO 9600
      WRITE(1,103)
103  FORMAT(' File name:', )
      READ(1,102)FILE
      IF(FILE.EQ.SP12) GOTO 4000
      IF(FILE.EQ.'C')GOTO 9600
102  FORMAT(A12)
      OPEN(IO2,FILE=FILE,STATUS='OLD',ERR=4000)
C
      READ(IO2,4,END=3990)(TITLE(NMOL,I),I=1,20)
4  FORMAT(20A4)
      WRITE(1,6)(TITLE(NMOL,I),I=1,20)
6  FORMAT(' Title of your File is: ',/,1X,20A4)

```

# READAT

```

C      INIT=N+1
      INRES=NRES
      IRES(1)=1
      DO 100 I=1,3410
      N=N+1
30 IF(.NOT.EMP)READ(IO2,71,END=110,ERR=4900) IAT(N),RES(N),
*      MRES(N),(X(J,N),J=1,3),PC(N),ITYPE(N),SPE(N)
C      *      MRES(N),ITYPE(N),(X(J,N),J=1,3),PC(N),SPE(N)
      IF(EMP)THEN
        IF(NSTD)THEN
          READ(IO2,4110,END=110)IAT(N),(COORD(J),J=1,18)
          WRITE(13,4110)COORD
          BACKSPACE 13
          READ(13,*,END=110,ERR=4900)(X(J,N),J=1,3)
4110      FORMAT(19A4)
        ELSEIF(BROK)THEN
          READ(IO2,7,END=110,ERR=4900)IAT(N),RES(N),MRES(N),(X(J,N),
*          J=1,3)
          ELSE
            READ(IO2,7,END=110,ERR=4900)IAT(N),RES(N),MRES(N),(X(J,N),
*            J=1,3),PC(N),ITYPE(N)
            GOTO 1112
7          FORMAT(13X,A4,A3,2X,I4,4X,4F8.3,6X,I4)
        ENDIF
        KAT=IAT(N)
        IF(JAT(1).EQ.'H')ITYPE(N)=1
        IF(JAT(1).EQ.'C')ITYPE(N)=3
        IF(JAT(1).EQ.'N')ITYPE(N)=9
        IF(JAT(1).EQ.'O')ITYPE(N)=13
        IF(JAT(1).EQ.'S')ITYPE(N)=15
        IF(JAT(1).EQ.'P')ITYPE(N)=17
        IF(JAT(1).NE.'H'.AND.JAT(1).NE.'C'.AND.JAT(1).NE.'N'.AND.
*        JAT(1).NE.'O'.AND.JAT(1).NE.'S'.AND.JAT(1).NE.'P')THEN
          ITYPE(N)=20
          WRITE(1,1111)IAT(N),ITYPE(N)
1111      FORMAT(' Not set up for atom ',A4,' ITYPE set to ',I2)
        ENDIF
      ENDIF
1112 IF(NOH.AND.ITYPE(N).EQ.1) GO TO 30
      IF(MRES(N).GT.-NRES2)NRES2=-MRES(N)
      IF(N.LT.2)GOTO 100
      IF(MRES(N).NE.MRES(N-1))THEN
        NRES=NRES+1
        IRES(NRES)=N
C      MRES(N)=NRES
      ENDIF
100 CONTINUE
71 FORMAT(A4,A3,I4,4F8.3,I4,1X,L1)
      WRITE(1,101) 3410
101 FORMAT(' MAX NO. OF ',I4,' ATOMS READ *****')
110 CLOSE(IO2)
      IF(NSTD)CLOSE(13)
C      FIX END PT. FOR FINAL RESIDUE
      N=N-1
      IRES(NRES+1)=N
      NATS=N-INIT+1
      JRES=NRES-INRES
      WRITE(1,8) NATS,JRES
8 FORMAT(' No. of Atoms Read= ',I5,' No. of residues=',I5)
      IF(NMOL.EQ.1)NT1=N
      IF(MOL2) NT2=N-NT1
      GO TO 9600

```

```

C
4900 WRITE(1,4320)
4320 FORMAT(' Data error - is file format correct?')
      BACKSPACE IO2
      READ(IO2,4)COORD
      WRITE(1,4330)COORD
4330 FORMAT(' Data: ',18A4)
      CLOSE(IO2)
      IF(NSTD)CLOSE(13)
      N=NT1
      GOTO 9600

C
C *** NON STD. FILE FORMAT
5000 EMP=.NOT.EMP
      IF(EMP)THEN
        WRITE(1,5010)
5010  FORMAT(' Is file format Camb? ', )
        CALL ANS(BROK)
        IF(BROK)GOTO 9600
        IF(NSTD)CLOSE(13)
        CALL ANS(NSTD)
        IF(NSTD)THEN
          WRITE(1,5020)
5020  FORMAT(' Only IAT and coords are read in, all else =0')
          OPEN(13,STATUS='SCRATCH')
        ENDIF
      ENDIF
      GO TO 9600

C  Input rot/trans matrices from data file
8500 WRITE(1,103)
      READ(1,102)FILE
      IF(FILE.EQ.SP12)GOTO 8500
      IF(FILE.EQ.'Q')GOTO 9600
      OPEN(IO2,FILE=FILE,STATUS='OLD',ERR=8500)
      DO 8600 I=1,3
8600  READ(IO2,*,ERR=8500,END=8500)(A(I,J),J=1,3)
      READ(IO2,*,ERR=8500,END=8500)F,SCAL(1)
      CLOSE(IO2)
      NOSCL=.TRUE.
      N3=1
      GOTO 9600

C
C *** GO
9000 IF(NT1.LT.1.OR.(MOL2.AND.NT2.LT.1))THEN
      WRITE(1,9010)NMOL
9010  FORMAT(' No MOL',I1,' data read in yet!')
      GOTO 9600
    ENDIF
    IF(JUST2.OR.NOSCL)THEN
C  Set scale using original atoml pos'n & stored rotation matrices
      DO 9200 I=N3,NT1+NT2
        DUMX(1,I)=X(1,I)-XMN(1)
        DUMX(2,I)=X(2,I)-XMN(2)
        DUMX(3,I)=X(3,I)-XMN(3)
        XX=A(1,1)*DUMX(1,I)+A(1,2)*DUMX(2,I)+A(1,3)*DUMX(3,I)
        YY=A(2,1)*DUMX(1,I)+A(2,2)*DUMX(2,I)+A(2,3)*DUMX(3,I)
        ZZ=A(3,1)*DUMX(1,I)+A(3,2)*DUMX(2,I)+A(3,3)*DUMX(3,I)
        DUMX(1,I)=XX
        DUMX(2,I)=YY
        DUMX(3,I)=ZZ
      END DO
    END IF
  
```

C SET NO. OF 1st H2O & POSITION OF CALCIUM  
 DO 9005 I=1,NT1  
 IF(ITYPE(I).GT.17.AND.IVEW(49).LT.1)IVEW(49)=I  
 IF(RES(I).EQ.'HOH')GOTO 9006  
 9005 CONTINUE  
 9006 IF(RES(I).EQ.'HCH')IVEW(48)=I  
 C  
 IF(NSTD.OP.BPOK)RETURN  
 IF(EMP)THEN  
 DO 9020 J=1,NT1-NT2  
 IF(ITYPE(I).GT.20)STOP 'ERROR IN ITYPE CONVERSION (>20)'  
 ITYPE(I)=IEMP(ITYPE(I))  
 9020 CONTINUE  
 WRITE(1,9030)  
 9030 FORMAT(' EMP atom types converted \*\*\*\*\*')  
 ENDIF  
 ENDIF  
 RETURN  
 C  
 9500 PAUSE 'Note: do not use ED or SLIST etc.'  
 GOTO 9600  
 SUBROUTINE ROCK  
 \* /CRY/SPE(3410),ACTSIT(3410),GRAY  
 \* /VIEW/IVEW(50),DBC  
 \* /DUM/DUMY(3,3410)  
 C  
 CALL CHAMOP  
 10 WRITE(1,1)  
 1 FORMAT(' Delay time (millisecs): ', )  
 READ(1,2,FPP=10)IWT  
 2 FORMAT(14)  
 IF(IWT.LT.1)IWT=300  
 IF(IVEW(1).FO.8)IVEW(1)=7  
 CALL PICCIE  
 IF(.NOT.GRAY)CALL DRUMCL  
 IVEW(16)=IVEW(4)  
 IVEW(22)=1  
 IF(GRAY)CALL GREY  
 XR=5.  
 CALL ROT(1,XR,.FALSE.)  
 IF(GRAY)THEN  
 IVEW(22)=2  
 CALL GREY  
 ELSE  
 CALL DRUMCL  
 ENDIF  
 IVEW(22)=0  
 XR=360.0-XR  
 CALL ROT(1,XR,.FALSE.)  
 C  
 20 DO 1000 I=1,20  
 IF(GRAY)THEN  
 CALL ONOFF(1,.FALSE.)  
 CALL ONOFF(4,.TRUE.)  
 CALL SLEEP (IWT)  
 CALL ONOFF(4,.FALSE.)  
 CALL ONOFF(1,.TRUE.)  
 ELSE  
 DO 100 J=1,7  
 100 CALL PENDEF(J,0.,IVEW(23))  
 CALL DEFCOL(8,13,15,15)  
 CALL PENDEF(9,0.,0)  
 CALL PENDEF(10,0.,15)

ROCK

```

CALL PENDEF(11,0.,3857)
CALL DEFCOL(12,13,13,0)
CALL DEFCOL(13,13,2,13)
CALL DEFCOL(14,11,11,4)
CALL SLEEP (IWT)
DO 200 J=8,14
200 CALL PENDEF(J,0.,JVEW(23))
CALL DEFCOL(1,13,15,15)
CALL PENDEF(2,0.,0)
CALL PENDEF(3,0.,15)
CALL PENDEF(4,0.,3857)
CALL DEFCOL(5,13,13,0)
CALL DEFCOL(6,13,2,13)
CALL DEFCOL(7,11,11,4)
ENDIF
CALL SLEEP (IWT)
1000 CONTINUE
CALL CHAMOD
WRITE(1,30)
30 FORMAT(' Repeat? ', )
CALL ANS(YES)
IF(.NOT.YES)RETURN
25 WRITE(1,1)
READ(1,2,ERR=25)IWT
IF(IWT.LT.1)IWT=300
GOTO 20
SUBROUTINE ONOFF(I1,OFF)
C
I2=I1+7
C
DO 20 J=1,3
IF(OFF)THEN
CALL PENDEF(I+I1,0.,0)
IF(MOI2)CALL PENDEF(I+I2,0.,0)
ELSE
I3=(4*I)+3
CALL DEFCOL(I+I1,I3,I3,0)
IF(MOI2)CALL DEFCOL(I+I2,0,I3,I3)
ENDIF
20 CONTINUE
C
I3=1
IF(I1.EQ.4)I3=8
IF(OFF)CALL PENDEF(I3,0.,0)
IF(.NOT.OFF)CALL DEFCOL(I3,12,0,12)
RETURN
SUBROUTINE SFRCH
INTEGER COUNT
C
DO 25 I=1,135
25 JCON(ID,I)=0
C
C
SET VARIABLES
IF(ICD(1).EQ.ICD(2))THEN
JCON(ID,1)=ICD(1)
JCON(ID,2)=ICD(3)
SORT(ICD(3))=.TRUE.
COUNT=2
NATOM=2
I1=1
ELSE
JCON(ID,1)=ICD(1)
JCON(ID,2)=ICD(2)

```



ROCK

```

      JCON(ID,3)=ICD(3)
      JCON(ID,4)=ICD(4)
      SORT(ICD(3))=.TRUE.
      SORT(ICD(4))=.TRUE.
      NATOM=4
      COUNT=3
      I1=2
      ENDIF
      MOL2=(ICD(1).CT.NT1)
C
31 IIA=JCON(ID,COUNT)
   DO 20 I=1,NCON(IIA)
      IIB=ICON(I,IIA)
C
      IF(COUNT.CT.135) STOP 'COUNT .CT. 135 IN SEARCH'
C
      DO NOT INCLUDE PPREVIOUS ATOMS IN ICON
      DO 20 I=1,NATOM
      IF(K.FO.COUNT)GOTO 20
      IF(IIB.FO.JCON(ID,K)) GO TO 30
20 CONTINUE
C
      NATOM=NATOM+1
      JCON(ID,NATOM)=IIB
      SORT(IIB)=.TRUE.
30 CONTINUE
      COUNT=COUNT+1
      IF(COUNT.CT.NATOM)GOTO 21
      GO TO 31
C
21 NATOMS(ID)=NATOM-I1
C
   Extract first I1 atoms
      DO 50 I=1,NATOMS(ID)
      IF(JCON(ID,I).FO.0)WRITE(1,2)ID,I
2  FORMAT(' NOTE:  JCON(',I1,',',',I3,')=0')
50 JCON(ID,I)=JCON(ID,I+I1)
      RETURN
      SUBROUTINE STEREO
C
      ZF=0.4
      IF(IVEW(1).CT.8)ZF=.7
C
      CALL ZOOM(ZF,.FALSE.)
      SCAL(1)=SCAL(1)*ZF
      CALL CHAMOD
      WRITE(1,5078)
5078 FORMAT(' Input viewing dist:', )
      READ(1,5079)Z2
5079 FORMAT(F10.3)
      IF(Z2.LT.100..OR.Z2.CT.2000.)Z2=500.
C
      PROJECT VIEW AT ANGLE .0524 RADS (6 DEGREES) L
      X2=-(Z2*.0524)
      CALL PICCLE
C
      SHIFT TO CENTRE OF SCREEN
      CALL SHIFT2(128.,128.)
      CALL PROJ3(X2,0.,Z2)
C
      SHIFT BACK
      CALL SHIFT2(-192.,-128.)
      IVEW(9)=2
C
      CALL TRABFC
      CALL DUMMY
C
      CALL TRAEND
C
      CALL TRABFC
C
      REPEAT ON RHS
      CALL SHIFT2(192.,128.)

```

STEREO

```

      X2=(Z2*.0524)
      CALL PROJ3(X2,0.,Z2)
      CALL SHIFT2(-64.,-128.)
      CALL DUMMY
C     CALL TRAEND
      CALL SHIFT2(-64.,0.)
      SCAL(1)=SCAL(1)/ZF
C     CALL ZOOM(ZF,.FALSE.)
      RETURN
      SUBROUTINE SURF(SRF)
      EQUIVALENCE (IP1,IVEW(3)),(IP2,IVEW(4)),(IP3,IVEW(5))
C
      SCALF=SCAL(1)
      IF(SRF)THEN
        WRITE(1,1)
1       FORMAT(' Draw existing SURF? ', )
        CALL ANS(YES)
        IF(YES)GOTO 450
      ENDIF
C   Define region of MOLL to be represented
      DIM=1.8
      CALL VOL(10)
      IF(10.EQ.2)RETURN
      CALL XVOL(DIM)
      IF(NATS.LT.1)THEN
        WRITE(1,6)
6       FORMAT(' No atoms in box - quit!')
        RETURN
      ENDIF
      CALL CHAMCD
      WRITE(1,5)NATS
5      FORMAT(' No. of atoms in box=',I4)
C   Derive index for XCLFT in decending Z order
      DO 200 I=1,NATS
200    ZED(I)=XCLFT(I,IP3)
        IFAIL=0
        CALL MOIABF(ZED,I,NATS,IND,IW,IFAIL)
        IF(IFAIL.NE.0)WRITE(1,40)IFAIL
40     FORMAT(' ERROR DETECTED IN MOIABF: IFAIL=',I4)
C
C   Define points constituting UPPER surface
10    WRITE(1,20)
20    FORMAT(' Point separation (Angstroms): ', )
      READ(1,30,ERR=10)PIX
30    FORMAT(F8.5)
      IF(PIX.LT.0.08.OR.PIX.GT.0.8)PIX=0.35
C     PIX=PIX*SCAL(1)
C     IF(PIX.LT.1.1)PIX=1.1
      PIX2=(PIX*0.5)**2
      PIX3=(PIX-(1.0/SCALF))**2
      ICD=0
C
      DO 1000 I=1,NATS
        WRITE(1,*)ZED(I),IND(I)
        II=ICLFT(IND(I))
        XO=XCLFT(II,IP1)
        YO=XCLFT(II,IP2)
        ZO=XCLFT(II,IP3)
C
C   Check that not "underneath" another atom
      DO 100 J=1,I-1
        JJ=ICLFT(IND(J))
        DST=(XO-XCLFT(JJ,IP1))*(XO-XCLFT(JJ,IP1))

```

SURF

```

      *  +(Y0-XCLFT(JJ,IP2))*(Y0-XCLFT(JJ,IP2))
C      *  +(Z0-XCLFT(JJ,IP3))*(Z0-XCLFT(JJ,IP3))
      IF(DST.LT.PIX2)GOTO 1000
100 CONTINUE
C
C   Determine hydrophobicity of atom
      CFC=ABS(PC(II))
      IPH=10
      IF(CFC.GE.0.1.AND.CFC.LE.0.2)IPH=13
      IF(CFC.GT.0.2)IPH=7
      WRITE(1,101)II,IPH
101  FORMAT(' II=',I5,'    IPH=',I3)
C
      R=RVDW(ITYPE(II))
      AINC=ASIN(0.5*PIX/R)*2.0
      PIX1=1.0/(R*AINC)
      NPTS=IFIX(0.5*3.1415927/AINC)
      A1=-AINC
C
      DO 400 J=1,NPTS
        A1=A1+AINC
        CA1=COS(A1)
        SA1=SIN(A1)
        A2=-AINC
        MPTS=IFIX(2.0*3.1415927*R*CA1*PIX1)
        DO 300 K=1,MPTS
          A2=A2+AINC
          CA2=COS(A2)
          SA2=SIN(A2)
          XI=R*CA2*CA1+X0
          YI=R*SA2*CA1+Y0
          ZI=R*SA1+Z0
C
C   Test for closeness of grid point to neighbouring atoms
C   DO 60 L=1,NCON(II)
      M=ICON(L,II)
      DST=(XI-DUMX(IP1,M))*(XI-DUMX(IP1,M))
      *  +(YI-DUMX(IP2,M))*(YI-DUMX(IP2,M))
      *  +(ZI-DUMX(IP3,M))*(ZI-DUMX(IP3,M))
      IF(DST.LT.RVDW(ITYPE(M))*RVDW(ITYPE(M)))GOTO 300
60 CONTINUE
C
C   Test for overlap with other grid points
      I2=ICD-70
      IF(I2.LT.1)I2=1
      DO 80 L=I2,ICD
        DST=(XI-GRID(IP1,L))*(XI-GRID(IP1,L))
        *  +(YI-GRID(IP2,L))*(YI-GRID(IP2,L))
        IF(DST.LT.PIX3)GOTO 300
80 CONTINUE
C
C   All OK so store grid point
      ICD=ICD+1
      GRID(IP1,ICD)=XI
      GRID(IP2,ICD)=YI
      GRID(IP3,ICD)=ZI
      ICOL(ICD)=IPH
300 CONTINUE
400 CONTINUE
1000 CONTINUE
      SRF=.TRUE.
C
C   Display surface

```

SURF

```
450 WRITE(1,500)IGD
500 FORMAT(' No. of grid points=',I5)
    CALL SRFCOL
    ZMIN=9999.
    ZMAX=-9999.
    DO 600 I=1,IGD
    IF(GRID(IP3,I).GT.ZMAX)ZMAX=GRID(IP3,I)
    IF(GRID(IP3,I).LT.ZMIN)ZMIN=GRID(IP3,I)
600 CONTINUE
C
    DIFF=2.0/(ZMAX-ZMIN)
    DO 2000 I=1,IGD
    ZI=GRID(IP3,I)
    JCOL=ICOL(I)+DIFF*(ZI-ZMIN)
    CALL PENSEL(JCOL,0.0,0)
    XI=GRID(IP1,I)*SCALF+128.0
    YI=GRID(IP2,I)*SCALF+128.0
    CALL MOVTO3(XI,YI,ZI)
    CALL DOT(1.0)
2000 CONTINUE
```

Appendix A7. Routines in MOLEFC5 which required major modification for efficient use in IMDAC.

1) BONDS.

For PA2 (957 atoms - excluding hydrogens) the original bonding routine required ca 10 minutes on the Prime simply to calculate the bonding matrices - this was clearly unacceptable! As the maximum length of any standard residue in a protein is 17 atoms, the bonding routine (BONDS) could be altered so that a search through  $(n \times n)/2$  atoms to find all the bonds could be replaced by a search through only  $(n \times 17)$  atoms ( $n$  is the number of atoms in the protein). For small molecules, which do not have a regular sequence of atoms, a search through  $n \times n$  atoms is performed. (Note that the limited bonding search for proteins necessitates the manual input of S-S bonding data - if required.)

Another necessary modification to BONDS was the addition of atom-specific bonding criteria because, for example, N-H bonds are clearly incompatible in length with eg S-C or S-S bonds. (Bonding radii were taken from the Cambridge Crystal Data File, with some small empirical adjustments to allow for observed differences between bond-lengths in the crystal and in the gas-phase).

2) Energy calculations - VLJHPD, VSTAT, VHBOND, FINDHB, and MINIM.

Major changes were required in the non-bonded energy routines (VLJHPD, VSTAT and VHBOND) due to the severe restriction in the original programme of having to read-in exclusion ( $IEX_{i,j}$  and  $N_i$ ) and H-bond terms from a data file, for every 50 atoms. (This very slow process had been required because of the small memory of the PDP11 computer for which the programmes were originally written.) As the Prime has 3 MB of real memory all the file input/output (I/O) code could be replaced with code for larger IEX, N and H-bonding arrays. The routine FINDHB (which sets up the arrays of atoms which could be



involved in H-bonds) had to be substantially re-written for this change - mainly due to the rather untidy nature of the original code. Other features we have added to speed up the energy calculations are:

- (i) An optional cutoff - assumes that the energy of a pair of atoms separated by more than CUTOFF Å (default 12Å) will be negligible.
- (ii) The storing of all constant terms in one array (POLFF). This reduces the number of mathematical operations required.
- (iii) The adding of logical variables - much faster than the original use of integer variables.
- (iv) The exclusion of non-bonded intramolecular energy for calculations on large proteins. This very slow process (requiring  $(n \times n)/2$  calculations) is not needed since the structure of the protein will not change.
- (v) The routines for calculating 6-12 (VLJHRD) and electrostatic potentials (VSTAT) have been combined, almost halving the computer time (and code!) required, while maintaining all the original features.
- (vi) Addition of an option for the isolation of the active site residues of protein (SITE - 7.2.2.5).

The gain execution in speed obtained with all these changes (excluding replacement of file I/O code which could not be measured) is roughly 200 fold for PA2 (from ca 20 minutes to ca 5 seconds) and greater for larger proteins such as TLN. All changes were thoroughly tested by comparing results from the old code with results from the new. In all cases results were reproduced to within 2 decimal places, except for cutoffs of  $< \text{ca } 7\text{\AA}$ .

### 3) Change in the way the bonding matrices are stored

A rapid means of obtaining molecular connectivities was required for some of the new IMDAC subroutines to execute with reasonable speed - in particular the routines SURF and ADDH (see A6.5). This necessitated a change in the way bonds are stored, since the original

method of storing the atom number for each bonded pair of atoms in the arrays IA and IB gave no measure of the number of atoms connected to each atom, except by going through a somewhat lengthy procedure. We now therefore store the molecular connectivity in the arrays NCON and ICON, where NCON contains the number of atoms bonded to each atom, and ICON contains the numbers of the bonded atoms. Setting up the arrays NCON and ICON slows down the bonding routine by only an insignificant amount, but greatly speeds up and simplifies all the routines for which molecular connectivity is required (SURF, FRAG, SERCH, MNDOIN, MM2IN, EXCLUD, DANG and others).

This change in the way bonds are stored meant that a new means of drawing the molecules had to be devised. The old code for drawing from atom  $IA_i$  to  $IB_i$  was replaced by:

```
DO 100 I= atom1, atom2
```

```
DO 100 J=1,NCON(I)
```

```
L=ICON (J,I)
```

```
IF(L.CT.I) draw a line from atom L to atom I.
```

C The IF statement ensures that each line is only drawn once.

```
100 CONTINUE
```

This code is as fast as the original code, but has the advantage that drawing is performed from atom-to-atom, instead of bond-to-bond, which makes the drawing of parts of a molecule (eg individual residues) much easier.

4) IDATOM - this routine for the input of atoms via a cursor on the graphics screen was essentially completely re-written and is therefore listed in appendix A6.5. The main change was that if the lower number for the range of atoms to be searched for matching the (input) atom on the screen is negative, then the first atom input defines the 'working' molecule and subsequent input atoms are limited to that molecule only. This is very useful for work with proteins or with two

close molecules so that unwanted atoms from the wrong molecule are avoided.

All of the original MOLFCS routines have been modified in some way to incorporate the extra options we have added and to clarify the logical structure of the routines. (Most of the routines have been essentially replaced.)

## References.

- R.J.Abraham and E.Bretschneider, in 'Internal rotation in molecules', (W.J.Orville-Thomas, Ed), Wiley-Interscience, (1974) 481-584.
- R.J.Abraham and G.Gatti, J. Chem. Soc. (B), (1969)961-968.
- G.Alagona, A.Pullman, E.Scrocco and J.Thomasi, Int. J. Peptide Protein Res., **5** (1973) 251.
- N.Allinger, Adv. Phys. Org. Chem., **13** (1976) 1-82.
- N.L.Allinger, J. Am. Chem. Soc., **99** (1977) 8127.
- N.Allinger, QCPE Bulletin, **3.2** (1983) 32-33.
- J.Altmann, minutes of Computational Chemistry Working Party (ULCC), Oct. 1984.
- P.R.Andrews, J.M.Carson, A.Caseli, M.J.Spark and R.Woods, J. Med. Chem., **28** (1985) 393-399.
- P.R.Andrews and G.A.R.Johnston, Nature New Biol., **27** (1973) 29-30.
- P.R.Andrews and G.A.R.Johnston, Biochem. Pharmacol., **28** (1979a) 2697-2702.
- P.R.Andrews and G.A.R.Johnston, J. Theoret. Biol., **79** (1979b) 263-273.
- D.R.Armstrong, R.J.Breckenridge and C.J.Suckling, J. Theoret. Biol., **97** (1982) 267-276.
- J.Arnt and P.Krogsgaard-Larsen, Brain Res., **177** (1979) 395.
- P.M.Beart, D.R.Curtis and G.A.R.Johnston, Nature New Biol.,
- F.C.Bernstein et al, J. Mol. Biol., **112** (1977) 535-542. **234** (1971) 80-81.
- F.Billes, J. Mol. Struct., **142** (1986) 343-346.
- R.C.Bingham, M.J.S.Dewar and D.H.Lo, J. Am. Chem. Soc., **97** (1975) 1285.
- G.Binsch and H.Kessler, Ang. Chem. (Int. Edn.), **19** (1980) 411-494.
- T.Blair and G.A.Webb, J. Med. Chem., **20** (1977) 1206.
- B.W.Borthwick, PhD Thesis, The City University, London (1977).
- B.W.Borthwick and E.G.Steward, J. Molec. Struct., **7** (1977) 41-48.
- R.Brakaspathy and S.Singh, Proc. Indian Acad. Sci., Chem. Sci., **96** (1986) 285-290.
- J.L.Burch, K.S.Raghuveer and R.E.Christofferson, in "Environmental effects on molecular structure and properties", B.Pullman (Ed.), The Jerusalem symposia on quantum chemistry and biochemistry, volume 8, Reidel, Dordrecht, Holland, 1976.

- U.Burkert and N.L.Allinger, in "Molecular mechanics", Am. Chem. Soc. Monographs, Washington, DC, 1982.
- K.Y.Burstein and A.N.Isaev, *Theor. Chim. Acta*, **64** (1984) 397-401.
- B.Busetta, I.J.Tickle and T.L.Blundell, *J. Appl. Cryst.*, **16** (1983) 432-437.
- B.L.Bush and T.A.Halgren, Fifth International Meeting of the Molecular Graphics Society, Cap d'Agde, France, April 1986.
- L.R.Chang and E.A.Barnard, *J. Neurochem.*, **39** (1982)1507-1518.
- C.Chotia, A.M.Lesk, M.Levitt, A.G.Amit, R.A.Mariuzza, S.E.V.Phillips and R.J.Poljak, *Science*, **233** (1986) 755-758.
- G.R.Clarke, PhD Thesis, The City University, London (1976).
- G.R.Clarke, personal communication (1981).
- J.B.Collins, P.V.R.Scheyer, J.S.Binkley and J.A.Pople, *J. Chem. Phys.*, **64** (1976) 5142.
- M.L.Connolly, *Science*, **221** (1983) 709-714.
- M.L.Connolly, *J. Mol. Graph.*, **3** (1985) 19-24.
- E.Costa and A.Guidotti, *Bioch. Pharmacol.*, **34** (1985) 3399-3403.
- R.W.Counts, *QCPE Bulletin*, **5** (1985) 113-114.
- D.R.Curtis, A.W.Duggan, D.Felix and G.A.R.Johnston, *Nature*, **226** (1970) 1222-1224.
- E.K.Davies, *Comput.-Aided Mol. Des.*, Proc. 2 day conf., 1984, (Pub. 1985), Oyez Sci. Tech. Serv., London, ppl47-155.
- F.V.DeFeudis, *Trends Pharm. Sci.*, **2** (1981) vi-ix.
- M.J.S.Dewar, *J. Mol. Struct.*, **100** (1983) 41-50.
- M.J.S.Dewar and W.Thiel, *J. Am. Chem. Soc.*, **99** (1977) 4899-4907 and 4907-4917.
- R.Diamond, A.Wynn, K.Thomsen and J.Turner, in "Computational Crystallography", D.Sayre Ed., Oxford Univ. Press, 1982.
- B.W.Dijkstra, J.Drenth and K.H.Kalk, *Nature*, **289** (1981) 604-606.
- S.Diner, J.P.Malrieu and P.Claverie, *Theor. Chim. Acta*, **15** (1969) 100.
- P.A.Dobosh and N.S.Oslund, *QCPE programme* 281 (1975).
- M.Dupuis, D.Spangler and J.Wendoloski, *NRCC Software Catalogue*, **1**, QG01 (GAMESS) 1980.
- J.T.Edward, P.G.Farrell and J.L.Job, *J. Phys. Chem.*, **77** (1973) 2191-2195.



- V.Elango, A.J.Freyer, G.Blasko and M.Shamma, *J.Nat.Prod.*, **45** (1982) 517-522.
- J.W.Emsley, J.Feeney and L.H.Sutcliffe, *High resolution NMR spectroscopy*, Pergamon Press (1965) 595-604.
- S.J.Enna, J.F.Collins and S.H.Snyder, *Brain Res.*, **124** (1977) 185-190.
- E.Fos, J.Vilarrasa and J.Fernandez, *J. Org. Chem.*, **50** (1985) 4894-4899.
- M.C.Fournie-Zaluski, E.Lucas, G.Waksman and B.P.Roques, *Eur. J. Biochem.*, **139** (1984) 267-274.
- M.C.Fournie-Zaluski et al, *J. Med. Chem.*, **28** (1985) 1158-1169.
- M.C.Fournie-Zaluski, E.Lucas-Soroca and B.P.Roques, *J. Med. Chem.*, **29** (1986) 751-757.
- D.Friesen and K.Hedberg, *J. Am. Chem. Soc.*, **102** (1980) 3987-3994.
- I.J.Frigerio, I.D.Rae and M.G.Wong, *Aust. J. Chem.*, **35** (1982) 1604-1614.
- D.W.Gallager, J.W.Thomas and J.F.Tallman, *Bioch. Pharmacol.*, **27** (1978) 2745-2749.
- A.Galli, L.Zilletti, M.Scotton, G.Adembri and A.Giotti, *Pharmacol. Res. Comm.*, **3** (1980) 267-272.
- R.D.Gilardi, *Nature New Biol.*, **245** (1973) 86-88.
- E.W.Gill, *Proc. Roy. Soc. B*, **150** (1959) 381-402.
- E.W.Gill, in "Progress in medicinal chemistry" Vol.14, G.P.Ellis and G.B.West (Eds), Butterworths (1965) pp39-85.
- A.Colebiewski and A.Parcezewski, *Chem. Rev.*, **74** (1974) 519.
- C.Corinsky and D.S.Moss, *J. Cryst. Mol. Struct.*, **3**(1973) 299-307.
- A.R.Gregory and M.N.Paddon-Row, *J. Am. Chem. Soc.*, **98** (1976) 7521-7523.
- A.R.Gregory and M.Przybylska, *J. Am. Chem. Soc.*, **100** (1978) 943-949.
- A.Guidotti, *Adv. Bioch. Psychopharmacol.*, **21** (1980) 271
- A.Guidotti et al, *Nature*, **275** (1978) 553
- H.Gunther, 'NMR spectroscopy', John Wiley and Sons, 1980, p33.
- C.W.Haigh and R.B.Mallion, *Mol. Phys.*, **22** (1971) 955-970.
- C.W.Haigh and R.B.Mallion, *Prog. NMR Spectroscopy*, **13** (1980) 303-344.
- I.Haneef, personal communication (1985).
- M.G.Hicks and W.Thiel, *J. Comput. Chem.*, **7** (1986) 213-218.

- D.R.Hill and N.G.Bowery, *Nature*, **290** (1981) 149-152.
- T.Hirano and T.Miyajima, *J. Mol. Struct.*, **126** (1985) 141-144.
- P.B.Hitchcock, R.Mason, K.M.Thomas and G.G.Shipley, *Proc. Nat. Acad. Sci. USA*, **71** (1974) 3036-3040.
- A.J.Hopfinger, *J. Med. Chem.*, **28** (1985) 229-237.
- W.Hunkeler, H.Mohler, L.Pieri, E.P.Bonetti, R.Cumin, R.Schaeffner and W.Haefly, *Nature*, **290** (1981) 514-516.
- S.Islam, personal communication (1984).
- L.M.Jackman and S.Sternhell, 'Applications of NMR Spectroscopy in Organic Chemistry', Pergamon Press (1969).
- G.A.R.Johnston, in 'GABA in central nervous system action', (E.Roberts, T.N.Chase and D.B.Towers, Eds), Raven Press, New York, (1975) pp395-411.
- G.A.R.Johnston, P.M.Beart, D.R.Curtis, C.J.A.Game, R.M.McCulloch and R.M.Maclachlan, *Nature New Biol.*, **240** (1972) 219-220.
- G.A.R.Johnston and M.Willow, *Trends Pharm. Sci.*, **3** (1982) 328-330.
- G.A.R.Johnston and R.D.Allen, *Neuropharmacology*, **23** (1984) 831-832.
- F.Jordan, S.Nishikawa and P.Hemmes, *J. Am. Chem. Soc.*, **102** (1980) 3913-3917.
- J.Kardos, G.Blasko, P.Kerekes, I.Kovaks and M.Simonyi, *Biochem. Pharmacol.*, **33** (1984) 3537-3545.
- M.Karplus, *J. Chem. Phys.*, **30**(1959) 11-15.
- A.R.Katritzky, R.C.Patel and F.G.Riddell, *Angew. Chem. Int. Ed.*, **20** (1981) 521-529
- L.B.Kier and J.M.George, *Experientia*, **29** (1973) 501-502.
- L.B.Kier and E.B.Trutt, *Experientia*, **26** (1970) 988.
- H-J.Kohler, *Z. Chem.*, **11** (1971) 467.
- W.Klyne and V.Prelog, *Experientia*, **16** (1960) 521.
- P.Krogsgaard-Larsen, *J. Med. Chem.*, **24** (1981) 1377-1383
- P.Krogsgaard-Larsen and A.V.Christensen, *Ann. Rep. Med. Chem.*, **15** (1980) 41-50.
- P.Krogsgaard-Larsen, T.Honore and K.Thyssen, in 'GABA Neurotransmitters', (P.Krogsgaard-Larsen, J.Scheel-Kruger and H.Kofod eds), Munksgaard, Copenhagen, (1978) 201-216.
- P.Krogsgaard-Larsen, P.Jacobsen and E.Falch, in 'The GABA receptors' (Ed Enna, S.J.) 149-176 (The Humana Press, 1983).
- P.Krogsgaard-Larsen and E.O.Nielsen, *Acta Neur. Scand.*, **69** (1984) 317-320

- J.Kroon and J.A.Kanters, *J. Mol. Struct.*, **24** (1975) 109-129.
- J.Koller, V.Harb, M.Hodoscek and D.Hadzi, *Theochem.*, **23** (1985) 343-350.
- J.B.Lambert, R.J.Nienhuis and J.W.Keepers, *Ang. Chem. (Int. Edn.)*, **20** (1981) 487-500.
- G.Lambrecht and E.Mutschler, in "Proceedings of the seventh Jerusalem symposium on quantum and molecular pharmacology", B.Pullman and E.D.Bergmann (Eds), Reidal, Boston, USA, (1974).
- K.G.Lloyd, S.Dreksler and E.D.Bird, *Life Sciences*, **21** (1977) 747-754.
- G.Maksay and M.K.Ticku, *Neurosci. Lett.*, **51** (1984)219-224
- R.B.Mallion, *J. Chem. Soc. B*, (1971) 681-686.
- I.L.Martin, *Trends Pharm. Sci.*, **5** (1984)343-347
- I.L.Martin and J.M.Candy, *Neuropharmacol.*, **17** (1978) 993-998
- Y.Martin, *J. Med. Chem.*, **24** (1981) 229-237.
- H.Mohler and T.Okada, *Science*, **198** (1977a) 849
- H.Mohler and T.Okada, *Nature*, **267** (1977b) 65-67.
- H.Mohler and T.Okada, *Mol. Pharmacol.*, **14** (1978) 256-265.
- A.J.Morffew, *J.Molec. Graphics*, **2** (1984) 124-128.
- A.F.Monzingo and B.W.Matthews, *Biochemistry*, **21** (1982) 3390-3394.
- A.F.Monzingo and B.W.Matthews, *Biochemistry*, **23** (1984) 5724-5729.
- D.S.Moss and F.E.Watson, *Acta Cryst.* **C40** (1984)1960-1963
- J.Murto, M.Rasanen, A.Aspiala and T.Lotta, *Theochem*, **17** (1984) 99-112.
- J.D.Neece, PhD thesis, Univ. of Arkansas, 1980.
- S.H.Nicholson, C.J.Suckling and L.L.Iversen, *J. Neurochem.*, **32** (1979) 249-252.
- P.van Nuffel, L.van den Enden, C.van Alsenoy and H.J.Geise, *J. Molec. Struct.*, **116** (1984) 99-118.
- R.W.Olsen, *J. Neurochem.*, **37** (1981) 1-13.
- L.Onsager, *J. Am. Chem. Soc.*, **58** (1936) 1486.
- J.A.Osborne, *Phys. Rev.*, **62** (1945) 351.
- M.N.Palfreyman, personal communications (1985 and 1986).
- P.J.D.Park, R.A.Rethvick and B.H.Thomas, in "Internal Rotation in molecules", (W.J.Orville-Thomas, Ed), Wiley-Interscience, (1974) 57-113.
- W.J.E.Parr and T.Schaefer, *Acc. Chem. Res.*, **13** (1980) 400-406
- S.J.Perkins and K.Wuthrich, *Biochimica Biophysica Acta*, **576** (1979) 409-423.

- G.W.Pooler and E.G.Steward, *J. Molec. Struct.*, 140 (1986) 131-139.
- G.W.Pooler and E.G.Steward, *J. Molec. Struct.*, in press.
- J.A.Pople and D.L.Beveridge, 'Approximate molecular orbital theory', McGraw Hill, 1970.
- J.A.Pople, D.L.Beveridge and P.A.Dobosh, *J. Chem. Phys.*, 47 (1967) 2026.
- J.A.Pople and G.A.Segal, *J. Chem. Phys.*, 44 (1966) 1420.
- G.N.J.Port and A.Pullman, *Theoret. Chim. Acta*, 31 (1973) 231-237.
- G.N.J.Port and A.Pullman, *Int. J. Quant. Chem.*, 74 (1974) 21.
- R.Potenzzone, Jr., E.Cavicchi, H.J.R.Weintraub and A.J.Hopfinger, *Computers and Chemistry*, 1 (1977) 187-194.
- C.J.Pouchert and J.R.Campbell (Eds), *The Aldrich Library of NMR Spectra*, Aldrich chemical company, 1974.
- M.A.Prazeres, PhD Thesis, The City University, London (1982)
- S.Profeta, Jr. and N.L.Allinger, *J. Am. Chem. Soc.*, 107 (1985) 1907-1918.
- B.Pullman, in "Proceedings of the seventh Jerusalem symposium on quantum and molecular pharmacology", B.Pullman and E.D.Bergmann (Eds), Reidal, Boston, USA, (1974) pp10-36.
- A.Pullman and B.Pullman, *Quart. Rev. Biophys.*, 7 (1975) 505-566.
- L.Radom and N.V.Riggs, *Aust. J. Chem.*, 33 (1980) 249-255.
- L.Radom, J.Baker, P.M.W.Gill, R.H.Nobes and N.V.Riggs, *J. Mol. Struct.*, 126 (1985) 271-290.
- W.G.Richards, *J. Med. Chem.*, 19 (1976) 1250-1252.
- B.P.Roques, *J. Pharmacol.*, (Paris) 16(suppl) (1985) 5-31.
- S.Safe and R.Y.Moir, *Can. J. Chem.*, 42 (1964) 160-162.
- J.Sandstrom, *Dynamic NMR spectroscopy*, Academic Press, 1982.
- R.Schulz, A.Schweig and W.Zittlau, *Theochem.*, 22 (1985) 115-120.
- M.Shamma and V.St.Georgiev, *Tet. Lett.*, 27 (1974) 2339-2342.
- E.Sigel, F.A.Stephenson, C.Mamalaki and E.A.Barnard, *J. Recept. Res.*, 4 (1984) 175-188
- M.A.Simmonds, *Nature*, 284 (1980) 558-560
- M.A.Simmonds, in 'Actions and interactions of GABA and benzodiazepines', (Ed. N.G.Bowery), (Raven Press, New York 1984) 27-41
- M.Snarey, personal communication (1982).
- R.Squires and C.Braestrup, *Nature*, 266 (1977) 732
- R.A.Spragg, *J. Chem. Soc. (B)*, (1968)1128

- E.G.Steward, P.W.Borthwick, G.R.Clarke and D.Warner, *Nature*, **256** (1975) 600-602
- E.G.Steward, R.Player, J.P.Quilliam, D.A.Brown and M.J.Pringle, *Nature*, **233** (1971) 87-88.
- E.G.Steward, R.B.Player and D.Warner, *Acta Cryst.*, **B29** (1973) 2038-2040.
- J.J.P.Stewart, *QCPE Bulletin*, **5** (1985) 126-130 (and the preceding paper).
- R.D.Stowlow, P.W.Samuel and T.W.Giants, *J. Am. Chem. Soc.*, **103** (1981) 197-198.
- J.F.Tallman, J.W.Thomas and D.W.Gallager, *Nature*, **274** (1978) 383-385.
- K.Tanaka, H.Akutsu, Y.Ozaki, Y.Kyogoku and K.Tomita, *Bull. Chem. Soc. Japan*, **51** (1978) 2654-2658.
- W.Thiel, *J. Am. Chem. Soc.*, **103** (1981) 1413, 1420.
- G.Toffano, A.Guidotti and E.Costa, *Proc. Nat. Acad. Sci. USA*, **75** (1978) 4024-4028.
- W.F.van Gunsteren, Fifth International Meeting of the Molecular Graphics Society, Cap d'Agde, France, April 1986.
- H.M.Verheij et al, *Biochemistry*, **19** (1980) 743-750.
- D.E.Walters and A.J.Hopfinger, in *QSAR in design of bioactive compounds, first telesymposium on medicinal chemistry*, 1984. pp279-286.
- D.Warner and E.G.Steward, *J. Mol. Struct.*, **25** (1975) 403-411.
- H.J.R.Weintraub and A.J.Hopfinger, *Int. J. Quant. Chem.: Quantum Biology Symp. no. 2*, (1975) 203-208.
- T.Weller, D.Klopper and H.J.Kohler, *Chem. Phys. Lett.*, **36** (1975) 475-477.
- J.W.Wilson III, C.L.Zircle, E.L.Anderson, J.J.Stehle and G.E.Ullyot, *J. Org. Chem.*, **16** (1951) 792-799.
- M.Withnall, personal communication (1984).