

Konrad-Zuse-Zentrum für Informationstechnik Berlin
Heilbronner Str. 10, W-1000 Berlin 31



Klaus Schöffel

Computational Chemistry
Software at ZIB

Herausgegeben vom
Konrad-Zuse-Zentrum für Informationstechnik Berlin
Heilbronner Str. 10
1000 Berlin 31
Verantwortlich: Dr. Klaus André
Umschlagsatz und Druck: Rabe KG Buch-und Offsetdruck Berlin

ISSN 0933-789X

Abstract

The following report intends to provide a survey over the computational chemistry molecular structure software installed on the supercomputers CRAY X-MP/216 and CRAY Y-MP2E/164 at ZIB. It shows what kind of problems can be tackled with the existing chemistry software, which covers a wide range of *ab initio*, semiempirical, molecular mechanics, and dynamics applications.

Preface

This report is the second issue of *Computational Chemistry at ZIB* within one year, which reflects the tremendous progress in Computational Chemistry software development.

During the last year we were able to enlarge our software substantially. With the introduction of *UniChem* a new level of chemistry software is reached. This integrated software package is a general toolkit for electronic structure investigations which gives an easy access to sophisticated quantum chemical methods for non-experts.

Through this toolkit the way a computational chemist works changed. With the use of interactive graphic the chemist can now concentrate himself doing his scientific work and there is no need to spend hours preparing input and studying thousands of output lines. It should be mentioned that a second toolkit *SPARTAN* is now commercially available.

A very promising and exciting development of general computational chemistry software toolkits has started which cover a wide range of computational techniques, integrated in an unique graphical user interface.

Contents

1	Introduction	1
2	Toolkits	3
2.1	UniChem	3
2.2	SPARTAN	4
3	ab initio Programs	7
3.1	CADPAC	7
3.2	GAMESS-USA	8
3.2.1	MOLPLT	9
3.2.2	MEPMAP	10
3.2.3	DENDIF	10
3.2.4	PLTORB	10
3.3	GAMESS-UK	11
3.4	GAUSSIAN90	12
3.4.1	The capabilities of the GAUSSIAN90 system	13
3.4.2	Improvements over GAUSSIAN88	14
3.5	GRADSCF	14
3.6	RPAC	15
3.6.1	PSI88A	16
4	Density-Functional Programs	17
4.1	DGauss	17
5	Semiempirical Molecular Orbital Programs	20
5.1	MNDO90	20
5.2	MOPAC	20
5.2.1	XDRAW2	22
6	Molecular Modelling Programs	23
6.1	AMBER	23
6.1.1	PREP Module	23
6.1.2	LINK Module	23
6.1.3	EDIT Module	24
6.1.4	NUCGEN Module	24

6.1.5	PARM Module	24
6.1.6	BORN Module	25
6.1.7	NEWTON Module	25
6.1.8	QUEST Module	25
6.1.9	GIBBS Module	26
6.1.10	ANALYS Module	26
6.1.11	MDANAL Module	26
6.2	GROMOS	26
6.2.1	Programs that build a molecular topology	27
6.2.2	Transformation Programs	28
6.2.3	Programs that generate atom coordinates	28
6.2.4	Programs that perform simulations	29
6.2.5	Programs that analyse configurations or sequences of configurations	29
6.2.6	Merge Programs	30
6.2.7	Programs for MD X-Ray refinement	30
6.3	Comparison between GROMOS and AMBER	30
6.3.1	Application	30
6.3.2	Structure of the Programs	31
6.3.3	Examples	32
6.4	Prophet	32
7	Plans	34
7.1	ab initio Programs	34
7.1.1	MOLCAS	34
7.1.2	GAUSSIAN 92	35
7.2	Semiempirical Molecular Orbital Programs	35
7.2.1	Argus	35
7.3	Density Functional Programs	36
7.3.1	Dmol	36
7.4	Molecular Modelling / Force Field Programs	36
7.4.1	X-Plor	36

Chapter 1

Introduction

Chemistry was a long time seen as a real experimental science. But this traditional picture is no more valid. During the last 20 years the computer has received more and more attention for chemistry. This is mainly due to increasing hardware performance and development of sophisticated program packages. These new programs became important research tools in chemistry. The sometimes huge output data can be transformed into the chemists language by various analysis and visualization tools.

It is nowadays possible to simulate chemical and physical processes on the atomic and molecular level by using fast computers. This offers the opportunity to avoid expensive experimental investigations. On the other hand, computer simulation gives the possibility to study chemical species which are inaccessible to experiments.

Calculations in *Computational Chemistry* are in many cases numerically highly intensive. The use of sophisticated hard- and software is therefore necessary. The development of vector computers had a great impact on this field, because many problems in *Computational Chemistry* can be reduced to simple matrix multiplications which can easily be vectorized and run at high Mflops rates. The today available parallel computers with vector hardware open new perspectives. These machines have the potential to perform expensive numerical calculations, particularly simulations, in some cases more effectively than conventional supercomputers. But nevertheless *Computational Chemistry* is still a challenge for computer vendors and software developing scientists.

Here at ZIB the chemistry community is using our high performance supercomputers CRAY X-MP/216 and CRAY Y-MP2E/164 to a large extend for a broad spectrum of chemical projects:

- Molecular orbital calculations for large organic systems
- Quantum mechanical electronic structure calculations for molecules and clusters
- Quantum mechanical calculations of photochemical and photophysical processes
- Molecular dynamics of clusters
- Quantum mechanical calculations of unusual gas phase molecules
- Quantum theoretical calculations on ions (PTICR)

- X-Ray refinements for transition metal complexes
- Determination of the second virial coefficient of molecule models in the continuum
- Determination of the second osmotic virial coefficient of polymers
- *ab initio* treatment of isoelectronic molecules
- Molecular dynamics calculations for macromolecules
- Computer simulation of the molecule movement in liquid ethen

It is clear that these research projects require sophisticated soft- and hardware. As a kind of user support at ZIB we felt it important to provide a set of standard up to date computational chemistry programs which utilize to a large extend the given vector facilities and contain the currently best algorithms.

The X Window system is getting more and more important for distributed graphic applications. We support X Window graphical output for some of the programs. This allows to display 3-dimensional structures, orbital plots or electrostatic potential maps in an easy way on workstations within the network. Extensive distributed graphic applications are used in the toolkits, like *UniChem* or *SPARTAN*. Here, the compute intensive part of the investigation is running on a CRAY and the course of the calculation can be followed simultaneously on the users remote workstation, where the final results may be visualized as well.

This report intends to provide a comprehensive information about the computational chemistry software installed at the CRAY computers at ZIB and it should show what kind of problems can be tackled with the existing chemistry software. For detailed information, the user is referred to the documentation-information system on the unix front-end *ufer.zib-berlin.de* and to the manuals which are available at ZIB on request.

Chapter 2

Toolkits

2.1 UniChem

UniChem is an advanced chemistry toolkit for molecular electronic structure investigations.

UniChem couples the power of the CRAY Y-MP2E supercomputer with the graphical display capability of Silicon Graphics (or IBM RS6000) workstations.

The workstation occupies one side of the *UniChem* window. It enables the user to:

- build molecular structures, either with single atoms or by using a molecular fragment library
- edit structures
- set up the input to the calculation by using a graphical point- and click display
- launch the calculation on the CRAY computer by using simple menus.
- monitor the progress of the calculation running on the CRAY
- analyse graphically the computed data of the calculation.

The quantum chemical programs that reside on the CRAY computer occupy the other side of the *UniChem* window. Each application solves the non-relativistic Schrödinger equation, but with different simplification. The codes work under the Born-Oppenheimer approximation, which allows the separation of nuclear and electronic motion.

The Quantum Chemistry Programs available under *UniChem* are:

- MNDO90, a semiempirical program
- DGauss, a density functional program
- CADPAC, an ab initio Hartree-Fock program

The *UniChem* environment not only provides the user with a common interface for MNDO90, DGauss, and CADPAC, but also offers many of the same types of calculations within each method. All of the chemistry codes can perform geometry optimizations (minima and transition states) and calculate second derivatives and dipole moment derivatives. The user can take geometry and second derivatives from any level of theory and use them as input to another level. *UniChem* also provides common analysis and display tools such as visualization of electrostatic potential, molecular orbitals, total electron density, electronic populations, permanent moments, polarizabilities and vibrational spectra.

2.2 SPARTAN

SPARTAN has been written for efficiency both on the current generation of vector and parallel architecture computers, as well as on modern graphic workstations, and has been designed to take advantage of large real memories if available. *SPARTAN* is intended not only to provide a convenient environment in which to carry out individual semi-empirical and *ab initio* Hartree Fock and correlated level calculations of relative energies, geometrical structures and vibrational frequencies on diverse molecular systems, but also to facilitate processing of large numbers of closely related calculations.

Most important, *SPARTAN* is intended to be utilized by chemists, preliminary experimental chemists. With this in mind, particular attention has been paid to the interface linking *SPARTAN* to the user. Modern computer graphics techniques have been extensively employed, not only to greatly reduce the drudgery and possibility of error associated with the construction of program input, but also to guide the interpretation of program output.

The program architecture of *SPARTAN* is not only appropriate for current generation of scalar and vector computers, but also one which will easily migrate onto future parallel and possibly massive parallel platforms.

SPARTAN provides a graphical user interface which allows not only convenient access to the full range of program capabilities, but also to permit visualization of the results of the quantum mechanical calculations.

While *SPARTAN*'s *ab initio* and semi-empirical modules maintain essential input compatibility with other electronic structure codes, particular, the widely used GAUSSIAN series of *ab initio* programs and the AMPAC and MOPAC semi-empirical codes. In particular, with regard to their ability to accept molecular geometry in terms of a Z-matrix, as well as providing a wide range of text output, they integrate tightly into an overall graphical environment.

The *SPARTAN* quantum chemistry environment presently comprises four independent program modules:

- **INTERFACE:** serves the following functions
 - input preparation designating the quantum chemical calculation to be performed by the AB INITIO and/or SEMI EMPIRICAL modules
 - the presentation of input designating graphical surfaces to be constructed by the GRAPHICS module for later display
 - the display of textual output resulting from the calculation
 - the display and manipulation of structures and graphical surfaces resulting from the calculation
 - construction of molecules from fragments
 - minimization of the resulting geometries using molecular mechanics
 - displaying and manipulation of molecular structures in terms of:
 - * Wire Frames
 - * Tubes
 - * Balls and Spokes
 - * Space filling models
 - dialogues for program access
 - display output of graphics module in terms of
 - * Dots
 - * Meshes
 - * Translucent or opaque solids
 - Model builder and associated molecular mechanics force fields for inorganic and organometallic molecules, as well as reaction states
 - Tools allowing equilibrium and transition structures to be optimized subject to geometrical constraints
 - Tools to make the comparison of graphical surfaces for two or more molecules easier
 - Display of multiple contour maps and multiple surfaces
 - Tools which make conformational searches much easier
 - Animation of normal mode vibrations
- **AB INITIO:** module is limited to the following functions:
 - the calculation of the energy and wavefunction corresponding to a single geometry.
 - the calculation of equilibrium or transition state geometries
 - the calculation of normal-mode vibrational frequencies
 - HF and MP2 models for closed (RHF) and open shell (UHF) system
 - Analytical HF second derivatives
 - Direct RHF and UHF treatment for very big systems

- Direct MP2 energies and first derivatives
- Internal basis set library
- Mulliken and NBO analysis
- **SEMI EMPIRICAL:** module is limited to three functions:
 - the calculation of the energy and wavefunction corresponding to a single geometry.
 - the calculation of equilibrium or transition state geometries
 - the calculation of normal-mode vibrational frequencies
 - MNDO, AM1, and PM3 hamiltonians
 - Calculation of electronic spectra using CI techniques
 - Limited to 200 atoms
- **GRAPICS:** module is responsible for the actual calculation of surfaces with *ab initio* or semi-empirical wavefunctions.
 - Preparation of electron and spin densities and molecular electrostatic potentials surfaces, as well as surfaces of molecular orbitals themselves.
 - Encoding properties onto those surfaces
 - Steric surface (region of the total density surface which are associated to a reagent of given size)
 - Construction of volumes (collection of data in 3 directions)
 - First derivatives of existing surfaces and properties

Chapter 3

ab initio Programs

3.1 CADPAC

Within the Hartree-Fock theory, the electronic energy equation is exact, but the wavefunctions used to solve it are not. These wavefunctions are a Slater determinant of linear combinations of one-electron hydrogen-like atomic orbitals represented by Cartesian Gaussians. Solving the Hartree-Fock equations yields the energy and wavefunction - the molecular orbitals. But because HF theory does not treat electron correlation correctly, it yields molecular orbitals that are less delocalized than those from density functional methods. However, HF theory can be extended easily to obtain accurate electron-correlation treatments. With CADPAC, users can choose from a wide range of correlated methods. These include Møller-Plesset perturbation theory, coupled cluster doubles, and configuration interaction with single and double excitations from the reference state. Of the three techniques present in *UniChem*, these correlated methods offer the greatest accuracy, but at a cost. The computational demands of many of these correlated methods scale as N^5 or even greater, where N is the number of basis functions. Such a scaling law prevents correlated methods from being applied to large systems. In addition to its collection of correlated methods CADPAC offers a rich set of standard HF-based properties. These range from permanent moments to dipole-dipole polarizability to IR and Raman intensities, and harmonic frequencies. For moderately sized systems, these HF methods scale as N^4 , where N is the Number basis functions, but reduce to $N^{2.5}$ or N^3 for larger systems.

Summary of the Capabilities of the package:

- evaluation of 1 and 2-electron integrals over contracted Cartesian gaussian basis functions of type s, p, d, or f.
- SCF calculation for closed shell, open shell, UHF and generalized open shell techniques
- calculation of 1-electron properties for these types of wavefunctions, including a distributed multipole analysis
- calculation of the gradients of the SCF energy

- use of gradients for automatic geometry optimization, and for the calculation of force constants by numerical differentiation
- transformation of the integrals from atomic orbital to the molecular orbital basis
- Møller-Plesset perturbation theory calculations to fourth order in the energy with and without contribution from triple excitations and second order in the 1-electron properties
- Coupled Hartree-Fock calculations of polarizabilities, including frequency dependence, and magnetizabilities
- Coupled Hartree-Fock calculations of the perturbation due to nuclear displacement
- calculation of the dipole and quadrupole moment derivatives
- calculation of the second derivatives of the energy by analytic methods
- analytic calculations of the polarizability derivatives
- calculation of infrared and Raman intensities, and the study of vibrational circular dichroism
- calculation of MP2 gradients, dipole moment derivatives, polarizabilities and force constants using analytic algorithms
- spin-projected UHF MP2 energies
- configuration Interaction Calculation using a wavefunction consisting of single and double excitations from the SCF wavefunction
- calculation of wavefunction within the Coupled Electron Pair Approximation (CEPA)
- calculation of wavefunction within the Coupled Pair Functional Approximation (ACPF)
- use the Coupled Cluster approach with double excitations from the SCF wavefunction

CADPAC is part of *UniChem*.

3.2 GAMESS-USA

GAMESS-USA (Generalized Atomical and Molecular Electronic Structure System, developed in USA) is an molecular orbital program developed through the National Resource of Computational Chemistry by M. Dupuis, D. Spangler and J. Wendoloski. The program is available for a wide range of computers: Alliant, Apollo, Ardent, Celerity, Cray X/Y-MP (Unicos), Convex, DecStation, FPS500, IBM RS/6000, MIPS, SGI and SUN. It is capable of performing a wide range of quantum chemical computations such as:

- Calculates RHF- SCF molecular wavefunctions.
- Calculates UHF- SCF molecular wavefunctions.
- Calculates ROHF- SCF molecular wavefunctions.
- Calculates GVB- SCF molecular wavefunctions.
- Calculates MCSCF- wavefunctions.
- Calculates CI- wavefunctions using the unitary group method.
- Calculates analytic energy gradients for all these wavefunctions except CI.
- Optimizes molecular geometries using an energy gradient in terms of Cartesian or internal coordinates.
- Searches for potential energy surface saddle points.
- Traces the intrinsic reaction path from a saddle point to reactants or products.
- Computes normal modes, vibrational frequencies and IR intensities.
- Computes radiative transition probabilities.
- Obtains Boys localized orbitals.
- Calculates the following molecular properties:
 - dipole, quadrupole, and octupole moments
 - electrostatic potential
 - electric field and electric field gradients
 - electron density and spin density
 - Mulliken and Lowdin population analysis
 - virial theorem and energy components.

3.2.1 MOLPLT

This program draws ball-and-stick molecular figures, automatically centring the molecule. Optionally, the program will draw a normal mode displacement at each atom. Once drawn on a graphics terminal, almost everything about the picture can be changed interactively. The changes include rotation of the molecule, colour and size changes, and rescaling of the normal mode. Once the molecule looks right, the current variable values can be written back to disk, so that a hardcopy run may be made.

3.2.2 MEPMAP

This program draws a contour map of a molecular electrostatic potential grid. The grid is produced by the GAMESS-USA program before MEPMAP is run. The MEP plot shows the energy of attraction or repulsion for a test positive charge, $+q$. The units of the contour lines are kcal/mol $\cdot e$, where e is the electron charge. Therefore, negative contours show where electrophilic reagents will prefer to attack a molecule, while positive contours show the preferred sites of nucleophilic attack. Note that the MEP map is computed by GAMESS-USA in the absence of the test charge, and that the MEP seen by a full $+1$ charge will actually differ substantially from that seen by a $+0.001$ test charge.

3.2.3 DENDIF

This program will draw a contour plot of the total electron density of a molecule. It also will plot a density difference map. The density difference map might be a whole molecule's density, minus the density of the individual atoms. Note that the fragments could be molecular subunits larger than atoms.

The necessary information to run this program is a set of molecular orbital grids of the natural orbitals of the whole molecule, produced by program PLTORB. If a density difference map is desired, sets of orbital amplitude grids must be provided for each fragment's natural orbitals.

3.2.4 PLTORB

This program draws contour plots of molecular orbitals expanded in a basis of contracted Cartesian Gaussian functions. The basis may contain s , p , and d functions. An arbitrary orientation of the plotting plane can be specified. The program contains an algorithm which attempts to draw the orbital with a reasonable phase, but this phase choice can be overridden if desired.

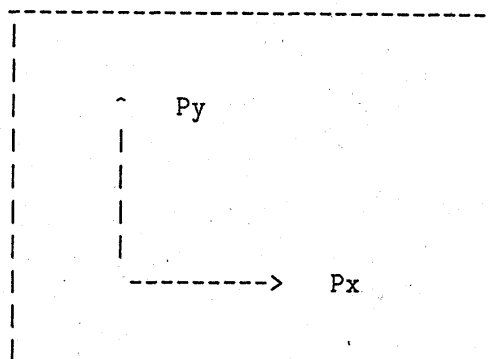
The Cartesian coordinates of the nuclei are referred to as the fundamental, or F axes. The plot is drawn in the plotting, or P axes, specifically, the P_x and P_y plane is plotted, in this fashion:


```

1  1
a  a
b  b
e  e
1  1

1  2

```



The X windows implementation rotates the entire picture, including axes, 90 degrees clockwise from this picture.

In the most common situation, the plane you want to plot is one of the axis planes of the original fundamental coordinate system. In this case you can ask for the plane by name, for example, entering ZX makes the Px direction point along Fz, and the Py direction is the Fx direction. You can choose an arbitrary orientation of the Px,Py plane by inputting a translation from the F coordinate origin to the P coordinate origin, along with a 3x3 orthogonal transformation matrix to rotate the F axes into the P axes.

If all the orbitals which are occupied are plotted, the resulting information can be used to draw total density or density difference maps (see program DENDIF).

3.3 GAMESS-UK

GAMESS-UK (Generalized Atomical and Molecular Electronic Structure System, developed in UK) general purpose *ab initio* molecular electronic structure program for performing SCF- and MCSCF-gradient calculations, together with a variety of techniques for post Hartree Fock calculations. On-going development of the code is carried out at Daresbury, with the program currently available on a wide range of machines, including the CRAY X-MP, Amdahl VP, FPS-264, Convex C2, Cyber-205, Alliant FX8, Apollo DN1000, SGI Power Series, VAX and SUN.

The program utilises the Rys Polynomial or Rotation techniques to evaluate repulsion integrals over s,p,d and f type Cartesian Gaussian orbitals. Open- and closed-shell SCF treatments are available within both the RHF and UHF framework, with convergence controls provided through a hybrid scheme of level shifters and the DIIS method. In addition generalised valence bond, complete active space SCF, and more general MCSCF calculations may be performed. The analytic energy gradient is available for each class of wavefunction above. Gradients for s and p Gaussians are evaluated using the algorithm due to Schlegel, while gradients involving d and f Gaussians utilise the Rys Polynomial Method. Geometry optimization is performed using a quasi-Newton rank-2 update method, while transition state location is available through either a synchronous transit trust region or 'hill-walking' method. Force constants may be evaluated by numerical differentiation or analytical by using a Coupled Hartree-Fock scheme.

Møller Plesset perturbation calculation may be carried out up to fourth order. 1st and 2nd derivatives are available analytically. Direct MP2 calculations are possible for large systems.

Ab initio core potentials are provided in both semi-local and non-local formalism for valence-only molecular orbital treatments.

Conventional CI (using the table-driven selection algorithms within the framework of MR-DCI calculations, Direct-CI and Full-CI correlation treatments are available. The direct calculation of molecular valence ionization energies may be performed through Green's function techniques, using either the outer-valence Green's function (OVGF) or the two-particle-hole Tamm-Dancoff method (2ph-TDA).

A variety of wavefunction analysis methods are provided, including population analysis and distributed multipole analysis, localised orbitals, graphical analysis and calculation of 1-electron properties.

With conventional processing of two-electron integrals, GAMESS-UK is limited to a maximum of 255 basis functions. Direct-SCF calculations lift this restriction however - the program structure is open-ended in this mode, so that calculations up to 1000 basis functions have been performed.

In addition to the functionality outlined above, the program now includes a utilities module concerned with house-keeping, library file creation, file manipulation etc.

Areas of application for GAMESS-UK include:

- Theoretical studies of reaction surfaces. SCF and MCSCF calculations of equilibrium and transition state geometries, and the evaluation of force constants and vibrational frequencies at these stationary points.
- Energy assignments in the entire field of electronic spectroscopy. Calculation of spectroscopic properties of highly excited molecular and ionic states. Evaluation of transition moments.
- Graphical and numerical analysis of Hartree Fock and post-Hartree Fock wavefunctions.
- Generation of zero-order wavefunctions required in the more extensive treatments of dynamical correlation energy. Applicability to general systems with many electrons in this treatment of dynamical correlation energy. Benchmark treatments of correlation energy using Full-CI calculations.
- Theoretical mechanistic studies of chemisorption in heterogeneous catalysis.
- Treatment of bio-organic and related molecules through Direct-SCF calculations and electrostatic potential analysis.

3.4 GAUSSIAN90

GAUSSIAN90 is a connected system of programs for performing semiempirical and *ab initio* molecular orbital calculations. It has been designed with the needs of the user in mind. Thus, all of the standard input is free-format and mnemonic.

GAUSSIAN90 provides an analysis of molecular potential energy surfaces to determine molecular structures and spectral properties of stable and unstable species. It provides a wide number of *ab initio* models, and some semiempirical models, using a user friendly interface.

3.4.1 The capabilities of the GAUSSIAN90 system

- Calculation of one- and two-electron integrals over s, p, d and f contracted Gaussian functions. The basis functions can either be cartesian gaussians or pure angular momentum functions, and a variety of basis sets are stored in the program and can be requested by name. Integrals may be stored externally or recomputed as needed via the Direct SCF procedures.
- Self-consistent field calculations for restricted closed-shell (RHF), unrestricted open-shell (UHF), and open-shell restricted (ROHF) Hartree-Fock wavefunctions as multiconfigurational wavefunctions that fall within the Generalized Valence Bond- Perfect Pairing (GVB-PP) formalism.
- Evaluation of various one-electron properties of the Hartree-Fock, MP2 and CI energies and wavefunctions, including Mulliken population analysis, multipole moments and electrostatic fields.
- Semi-empirical calculations using the CNDO, INDO, MINDO/3, MNDO and AM1 model Hamiltonians.
- Automated geometry optimization to either minima or saddle points, numerical differentiation to produce force constants, polarizabilities, dipole derivatives, and reaction path following.
- Correlation energy calculations using Møller-Plesset perturbation theory carried out to second (MP2), third (MP3) or fourth (MP4) order.
- Correlation energy calculations using configuration interaction (CI) using either all double excitations (CID) or all single and double excitations (CISD).
- Correlation energy calculations using coupled cluster theory with double substitutions (QCISD).
- Analytic computation of the nuclear coordinate gradient of the RHF, UHF, ROHF, GVB-PP, MP2, RCID and RCISD.
- Computation of force constants (nuclear coordinate second derivatives), polarizabilities, hyperpolarizabilities, dipole derivatives and polarizability derivatives analytically for RHF and UHF energies, and numerically for RHF, UHF, MP2, CID and CISD energies.
- Harmonic vibrational analysis. Determination of intensities for vibrational transitions at the HF, MP2, and CI levels.
- Testing the SCF wave functions for stability under release of constraints.

3.4.2 Improvements over GAUSSIAN88

- Configuration Interaction with single excitations for energies and gradients. This allows excited states structures, adiabatic excitation energies, 0-0 transitions to be predicted and related properties such as oscillator strengths and electron densities to be computed.
- Ground state vibrational frequencies and excited state properties are now possible on the context of a direct SCF calculation.
- Semi-direct MP2 method optimizes the calculation procedure for the particular disk and memory configuration of the computer system.
- Incore methods for SCF energies, gradients and second derivatives and MP2 energies and gradients are now available.
- Gradients as well as energies can be calculated at the Quadratic CI level. Associated properties, including polarizabilities, hyperpolarizabilities and electron densities, are also computed using the QCI level of correlation. The Complete Active Space SCF (CASSCF) method makes it possible to study systems with unusual electronic structures.

3.5 GRADSCF

GRADSCF (GRADient-SCF-program) is able to predict equilibrium structures, locates saddle points on many dimensional potential surfaces, and calculates harmonic force constants and vibrational frequencies. The most prominent feature of GRADSCF is its ability to calculate analytically the gradient of the potential energy with respect to the nuclear coordinates for several types of SCF or limited MCSCF wavefunctions. It is available only for CRAY X/Y-MP computers.

The program system is capable of the following functions:

- Evaluate an SCF wavefunction and energy at a specified geometry. The results of such a calculation can be further used to evaluate some one-electron properties.
- Locate an optimum geometry for a molecular system which lies at an energy minimum, subject only to the implied symmetry constraints present in the initial structure.
- Evaluate the matrix of second derivatives, from which the force constants and vibrational frequencies can be derived. This can be done either analytically or by finite difference techniques. The analytical calculation is especially efficient. If done analytically, the atomic polar tensors, from which the analytic infrared intensities are obtained, can be evaluated.
- Locate stationary points which are not necessarily minima with respect to the energy, and thereby find saddle points, or transition states, in chemical reactions.

- Correlation energy corrections for closed-shell SCF wavefunctions can be included at the MP2 level of theory.
- Properties calculated from the wavefunctions include a Mulliken population analysis, dipole and quadrupole moments, and expectation values. For closed shell SCF wavefunctions, the program is also capable of evaluating analytically the polarizability tensor.

3.6 RPAC

The program RPAC (Random Phase Approximation) is designed to calculate a variety of molecular linear response properties. It is available for CRAY X-MP, Convex and IBM (MVS) computers. The area of applications includes:

- electronic excitation energies
- oscillator and rotatory strengths
- nuclear magnetic shieldings for diamagnetic polyatomic molecules
- static and dynamic polarizabilities
- analyse of the result into molecular structure features

The excitation properties can be calculated for both singlet and triplet transitions using one of the following particle-hole schemes:

- the random phase approximation (RPA)
- diagonalisation of $A+B$ or $A-B$ matrices
- monoexcited CI (TDA)
- single excitation approximation (STA)

The RPA provides the correct linear response of a closed shell Hartree-Fock ground state to an external perturbation. It is particularly suited for spin-singlet properties. The magnetic shieldings can be calculated by the LORG (localized orbital/local origin) variant of the coupled Hartree-Fock theory (CHF) or by common-origin CHF theory, both of which use the inverse of $A-B$ in the calculation of the paramagnetic contribution. Polarizabilities may be calculated using RPA, TDA, or STA methods.

The LORG method and some of the methods for the analysis of the computed results require the use of localized molecular orbitals. Therefore the program includes routines for Forster-Boys localization of occupied orbitals and intensity localization of virtual orbitals.

The analysis of the excitations and the optical properties includes the calculation of gross transition and rearrangement charges and of effective bond intensity contributions mimicking the terms in Kirkwood's theory of optical activity. Alternatively the excitations

can be analyzed by plotting relevant molecular orbitals or various densities, namely ground and excited state densities or rearrangement and transition densities. These plots are generated as crude printer plots or as files for off-line plots (see. PSI88A).

The nuclear magnetic shieldings and polarizabilities can be analyzed into bond and bond-bond coupling contributions. For the shielding tensors the options allow analysis into contributions along bond directions. Shielding vectors may be plotted.

The program includes iterative methods for solving the RPA or CHF equations that allow the use of the full set of particle-hole configurations supported by the orbital basis. Limits in the current version of the program are:

- No. of atoms: 50
- No. of atomic orbitals: 255 (s,p, and only d)
- Connected to GAMESS-USA, interface to GAMESS-UK and GAUSSIAN 88 in progress

3.6.1 PSI88A

This program accepts a set of nuclear shielding tensors and determines contours of the shielding (tb) vector in the planes which are specified. Closed contours at the specified level(s) are drawn through the shielding vector arrays and the points (in angstroms) for each contour are stored as output. The total number of points, the number of curves along with the number of points in each curve and an indicator for whether the contour level is positive or negative are stored in the output disk file. For proper hidden line elimination in the 3-D drawing program, it is essential that all contours which are constructed be closed. If this condition is not met, an error message is printed. The current program version generates output in the HPGL-plotter language.

Chapter 4

Density-Functional Programs

4.1 DGAUSS

DGAUSS is a density functional program. In the density functional theory, the energy equation is approximate due to the presence of an exchange-correlation potential. This exchange-correlation potential is a quantum effect and describes the instantaneous electron-electron repulsion; it differs from the Coulomb, or electrostatic, repulsion. Density functional programs have the electron density as their conceptual centrepiece. DGAUSS solves for the electron density using a local density approximation that sometimes overemphasizes electron correlation, yielding an electronic structure that has the electron density too delocalized over the molecular framework. By applying a nonlocal correction, DGAUSS can compensate for this error, leading to results typical for correlated Hartree-Fock methods. Systems with many electrons that are known to be sensitive to electron correlation, such as organo metallic compounds and transition metal complexes, can be handled. Density functional methods scale as N^3 , where N is the number of basis functions. Molecules with over 150 atoms are manageable with DGAUSS.

DGAUSS key features:

- DGAUSS uses orbital basis sets which are constructed from linear combinations of Gaussian-type functions. In fact, these basis sets can be identical to those used in Hartree Fock calculations.
- DGAUSS uses a self-consistent-field (SCF) procedure.
- DGAUSS calculates the total energy as a function of molecular geometry.
- DGAUSS can calculate energy gradients, that is, the forces on each atom within a certain molecular geometry. These energy gradients are used to find stable molecular structures and to search for transition states.
- DGAUSS can calculate vibrational frequencies and force constants using second derivatives obtained as difference of gradients at small nuclear displacements. DGAUSS calculates molecular orbitals (MOs) corresponding to orbital energies.

These MOs can be used to derive molecular properties such as charges, multipole moments, and electrostatic potentials.

- DGauss can be applied in a spin-restricted and spin-unrestricted form. The spin density can be calculated.
- In DGauss, the electron density is expanded in Gaussian-type functions, referred to as "auxiliary basis set".
- DGauss includes, in an approximative way, electron exchange and correlation effects. The exchange-correlation contributions are evaluated on a finite grid and then fitted to another auxiliary basis set, which is closely related to one used to represent the electron density.
- DGauss calculates one-, two-, and three centre integrals over Gaussian-type functions, but no four-centre integrals are needed.

Most integrals are calculated analytically, but certain integrals involving exchange-correlation contributions are evaluated by a numerical integration technique using a finite grid. This introduces a numerical noise which can be controlled by the grid resolution. Computations using finer grids are significantly more compute intensive.

Although DGauss calculates MOs, density functional theory does not require the concept of a many-electron wave function. In density functional theory, the total electron density and not the total many-electron wave function plays the central role of the theory.

DGauss may be applied to a wide range of compounds including:

- Metal clusters
- Biological compounds
- Organometallics
- Organic Compounds

with the following main applications:

- Electronic energy
- Equilibrium geometry
- Harmonic vibrational frequencies
- Molecular orbital coefficients
- Orbital eigenvalues and occupations
- Atomic population analyses
- Dipole moments
- Molecular polarizability

- Data for subsequent visualization: spin density, electrostatic potential, total density, molecular orbitals

DGauss is a part of *UniChem*.

Chapter 5

Semiempirical Molecular Orbital Programs

5.1 MNDO90

MNDO90 is a semiempirical quantum chemistry program which allows calculations for MNDO, AM1, PM3, MINDO/3, CNDO/2m, MNDO/H and MND0C. Available SCF treatments include closed shell RHF, open shell UHF, and open shell half electron RHF, while electron correlation can be treated by second order perturbation theory (e.g. MP2, BWEN) or by minimal configuration interaction. MNDO90 may be used to calculate molecular structures, vibrational frequencies, charge distributions, and other molecular properties. Geometry optimizations for minima and transition states can be either complete or constrained, and can use cartesian or internal coordinates. Force constant calculations yield harmonic vibrational frequencies, infrared intensities, and thermodynamic properties. There are two input modes available. Keyword controlled input is designed to be compatible to MOPAC(5.0) allowing the use of identical input files. Some options of the program are not accessible via keywords, but only in the numeric input mode. The code is highly vectorized. MNDO90 is a part of *UniChem*.

5.2 MOPAC

MOPAC V6.0 is a general-purpose semi-empirical Molecular Orbital PACkage for the study of chemical reactions. The program is available for nearly all known computers, ranging from small IBM PC to large mainframes like the CRAY Y-MP.

The semi-empirical Hamiltonians MNDO, MINDO, and AM1 are implemented, and calculations of vibrational spectra, thermodynamic quantities, isotopic substitution effects and force constants for molecules, radicals, ions, and polymers are combined in a fully integrated package. Within the electronic part of the calculation, eigenvectors and localized orbitals, chemical bond indices, charges, molecular orbitals etc. are calculable. For studying chemical reactions, a transition-state location routine and two transition-state optimizing routines are available. For potential users to get the most out of the program they must understand how the program works, how to enter data,

how to interpret the results, and what to do when things go wrong.

While MOPAC calls upon many concepts in quantum theory and thermodynamics and uses some fairly advanced mathematics, the user who is not familiar with these specialized topics should not feel excluded from using it. On the contrary, MOPAC is written with the non-theoretician in mind. To this end the data are kept as simple as possible; this means that users can give their attention to the chemistry involved, and not concern themselves with quantum and thermodynamic exotica.

Areas of applications:

- MNDO, MINDO/3, and AM1 Hamiltonians.
- RHF and UHF methods.
- Extensive Configuration Interaction
 - 100 configurations
 - Singlets, Doublets, Triplets, Quartets, Quintets, and Sextets
 - Excited states
 - Geometry optimizations on specified states
- Single SCF calculation
- Geometry optimization
- Gradient minimization
- Transition state location
- Reaction path coordinate calculation
- Force constant calculation
- Normal coordinate analysis
- Transition dipole calculation
- Thermodynamic properties calculation
- Localized orbitals
- Covalent bond orders
- Bond analysis into σ and Π
- One dimensional polymer calculation
- Dynamic Reaction Coordinate calculation
- Intrinsic Reaction Coordinate calculation

5.2.1 XDRAW2

XDRAW2 is a powerful editing program specifically written to interface with MOPAC. Among the various facilities it offers are:

- The on-line editing and analysis of a data file, starting from scratch or from an existing data file, an archive file, or from a results file.
- The option of continuous graphical representation of the system being studied. Several types of terminals are supported, including DIGITAL, HEATHKIT, TEKTRONIX, X-Window and TERA-K terminals.
- The drawing of electron density contour maps generated by DENSITY on graphical devices.
- The drawing of solid-state band structures generated by MOSOL.
- The sketching of molecular vibrations, generated by a normal coordinate analysis.

Chapter 6

Molecular Modelling Programs

6.1 AMBER

AMBER V3.0A is an acronym for Assisted Model Building with Energy Refinement. AMBER is a set of programs called modules to do different tasks and the function of different modules are documented separately. It is available for Convex, CRAY X/Y-MP, SGI, Stellar and SUN.

AMBER is a general purpose molecular mechanics and dynamics program designed for the refinement of macro molecular conformations using the analytical potential energy function. The empirical parameters for the potential function are derived from both experimental results on smaller molecules of diverse chemical structure and ab initio quantum mechanical results on smaller molecules. The transferability of these parameters to related system has been the success of the use of molecular mechanics approach to study the conformations of macro molecules such as DNA and protein which would otherwise be an impossible task.

Molecular mechanics method is becoming increasingly important in the study of the binding of small molecules to macro molecules of biological interest. Their binding affinity could be satisfactorily determined by the existing potential energy function.

6.1.1 PREP Module

The purpose of this module is to generate the necessary connectivity and internal parameter list for a given residue or set of residues for linking this residue with others to form a macromolecule. The geometry of the residue is to be described by a tree structure for the use in the next module (LINK) to successfully connect these residues into a super molecule. The standard residues for protein and nucleic acids are kept as a data base and can be input to this module to generate the data base for the link module

6.1.2 LINK Module

The purpose of this module is to link different residues to form a macromolecule. It reads the geometries of a given set of residues either from the data base or from specified files and links them together in a specific manner to form the macro- molecules. It adopts the

tree convention for the residue connectivity and always connects the first main type atom of the current residue to the last main type atom of the previous residue. In addition to this standard linking process, it can cross link specified atoms in a molecule or different molecules to form a covalent bond. The macro-molecule can be specified as a single molecule or set of molecules. For example, double stranded DNA can be defined as two molecules and each strand constitutes a molecule (different molecules are not linked by a covalent bond unless explicitly specified by the cross linking information).

The residues which constitute nucleic acids and proteins are put in a data base and the program will fetch them whenever they are needed.

It is important to note that this module will generate the connectivity and the internal parameter lists such as bond, bond angle, dihedral angle, and excluded atoms pointers correctly. However, the coordinates are not meaningful except for smaller molecules. For example, the coordinates generated for double helical DNA will be completely meaningless, since it does not use any of the topological property of the actual double helix. The same is true for proteins. Hence, the user needs to read the "correct" coordinates at the next stage (the EDIT module).

6.1.3 EDIT Module

The main purpose of this module is to read the initial cartesian coordinates of the macromolecules preferably in pdb format for the energy refinement. This is necessary because the coordinates passed from the LINK module would not be realistic except for smaller molecules. This module reads those coordinates obtained either from x-ray crystallography or from user model building. The coordinates are assigned uniquely for each atom based on its graph name and the residue to which it belongs. In addition, some of the undefined atom coordinates such as beginning hydrogens are generated from connectivity information.

In addition, this module performs some other options to manipulate the coordinates and can also put counter-ions and solvent molecules around the solute molecule if necessary.

6.1.4 NUCGEN Module

The purpose of this module is to generate the cartesian coordinates for either double helical DNA or RNA with a given type of conformation. The helical topology of the double helix is stored for individual types in terms of cylindrical coordinates and the program loads the required topology and applies the two-fold symmetry with necessary helical repeat and height values.

The cartesian coordinates are output in the pdb format suitable for the minimisation. The residue informations are read as in the LINK module either for DNA or RNA.

6.1.5 PARM Module

The purpose of this module is to allocate the force field parameters for all the constants in the potential energy function using the lists passed from the edit module. The

constants such as equilibrium bondlengths, bondangles, dihedral angles, and their force constants and the non-bonded 6-12 parameters and H-bond 10-12 parameters are read in from the user source and allocated to all the quantities occurring in the system the respective values. The parameters are allocated based on atom symbols and all the quantities which are equivalent are assigned the same value. Once the assignment of the parameters is completed the information needed for the next modules is written out. This can subsequently be used by different modules, namely, BORN, NEWTON, NMODE, QUEST, GIBBS, ANALYS and MDANAL to perform different tasks.

6.1.6 BORN Module

The purpose of this module is to minimize the energy of a system in cartesian space using either steepest descent and conjugate gradient minimizer or conjugate minimizer alone. The module has been fully vectorized for CRAY.

There are provisions for putting the position constraints for groups of atoms with different harmonic force constraints for each group and for the selective minimisation of certain part of the system keeping the rest of the system as frozen environment. The forces between the unfrozen and frozen part are calculated while that between the frozen part alone is ignored.

6.1.7 NEWTON Module

The purpose of this module is to perform molecular dynamics. The driver for the molecular dynamics is taken from the program written by Van Gunsteren, Netherland. The module has been fully vectorized for CRAY.

6.1.8 QUEST Module

This module is a hybrid of ab initio quantum mechanical program GAUSSIAN80 (UCSF) and the empirical molecular mechanical program AMBER(UCSF) for simulating the molecular interactions and reactions in solution and enzymes. The atoms in the system are divided into two sets, namely, the G80 atoms and the AMBER atoms. The G80 atoms are treated quantum mechanically while the AMBER atoms are treated by molecular mechanics method. The interaction between the two systems is treated either purely by the molecular mechanics method or by a quasi quantum mechanical way, meaning that AMBER atoms are assumed as point charges and the charges are entering into the one-electron Hamiltonian of the G80 atoms and the repulsive interaction between the G80 and the AMBER atoms are determined by a 6-12 potential. In addition, the geometry optimization of smaller molecules by ab initio method can be routed through AMBER to this module so that one does not need to give a guess for the Hessian. There is an option to calculate the Hessian by molecular mechanics method which is found to be quite accurate for the initial guess of the second derivative so that the geometry optimization would converge rather rapidly.

6.1.9 GIBBS Module

The purpose of this module is to perform Free Energy Perturbation calculation using molecular dynamics approach. There are two methods available for evaluating the Free Energy change for a given system. One is the Monte Carlo type Mutation, which means that the Free Energy is calculated at discrete values of the mutation parameter Lambda. The number of steps for taking the average and the width of the window can be changed at each run. The program automatically tabulates the accumulated free energy and other statistics for each run. The other is the Molecular Dynamics type Mutation, which means that the Mutation parameter Lambda is a function of time and the Free Energy is accumulated over a period of time by defining the conversion time and the number of MD steps.

6.1.10 ANALYS Module

The purpose is to do different types of analysis to get the meaningful results out of the energy refined structure. There are several types of options available to perform the analysis.

The important part of the program is to decompose the energy among different groups of atoms in order to get the interaction energies between different parts of the system. It can also decompose the gradients into individual groups so that one can understand the direction of forces in different groups of atoms. The program puts those atoms which are not defined as groups explicitly into a separate group.

6.1.11 MDANAL Module

The purpose of this module is to do various analysis on the molecular dynamics trajectory of a system. There are several options available for doing different types of analysis such as internal coordinates averaging, cartesian coordinate averaging, and special DNA helical parameters analysis. It can also make the output for dynamics trajectory and normal mode movies in the Evans and Sutherland PS2 picture system.

6.2 GROMOS

GROMOS (GRONingen MOLEcular Simulation package) is a general-purpose molecular dynamics computer simulation package for the study of biomolecular systems. The program is available for: CDC, IBM, DEC, VAX, Amdahl, Cyber-205, CRAY X/Y-MP, Fujitsu, Convex and FPS. Its purpose is threefold:

- Simulation of arbitrary molecules in solution or crystalline state by the method of molecular dynamics (MD) or stochastic dynamics (SD).
- Energy minimisation of arbitrary molecules.
- Analysis of conformations obtained by experiment or by computer simulation.

The simulation package contains the GROMOS force field (proteins, nucleotides, sugars, etc.).

Applications of GROMOS are:

- Prediction of the dependence of a molecular conformation on the type of environment (water, apolar solvent, crystal, etc.)
- Calculation of relative binding constants by evaluating free energy differences between various molecular complexes
- Prediction of the energetic and structural changes caused by modifications of amino acids in enzymes or base pairs in DNA
- Derivation of (3D) molecular structure on the basis of 2D NMR data by using restrained MD techniques
- Dynamic modelling of molecular complexes by searching the configuration space using MD
- Prediction of properties of materials under extreme conditions of temperature and pressure, which may be experimentally inaccessible

6.2.1 Programs that build a molecular topology

- PRORMT: The program generates a molecular topology by plain reading; useful for small molecules, when a (formatted) molecular topology has been edited (changed by hand) or when special interaction parameters are to be used
- PROGMT: The program generates a molecular topology file containing the information that is required for a MD or SD simulation or energy minimization of a (set of) molecule(s). Given the amino acid sequence, the disulphide bridges and the terminal groups, it forms the peptide links and disulphide bridges between the residues, makes lists of bonds, bond-angles, improper dihedrals, dihedrals, and lists of atoms (neighbours) to be excluded from non-bonded interactions. Given the nucleotide sequence, the same can be done for nucleic acids. Given the glucose sequence, the same can be done for saccharides. Besides, a solvent topology file can be generated. The program can handle:
 - all kinds of amino acid residues, provided they are only linked together by peptide links or cys-cys bridges, two separate residue-chains can be generated, all kinds of nucleotides, provided they are only linked together by phospho-ester links, two separate nucleotide chains can be generated, all kinds of sugars, provided they are only linked together by c4-o4-c1 links,
 - an arbitrary number of atoms to be added to the two ends of the chains; in addition, the name, type and charge of the atoms making up the ends of the chains can be changed,

- a heme-group, linked to a histidine residue as well as not linked to the residue-chain,
 - any set of molecules, provided their topologies are given (like residues) and no links to the residue-, nucleotide- or glucose-chains are to be made,
 - an arbitrary number of atom types, bare as well as extended atoms,
 - non-bonded interactions requiring the exclusion of neighbours up to the second or up to the third neighbour inclusive.
- PROMMT: The program merges two molecular topology files
 - PROCMT: The program allows for changes in a molecular topology file by producing a formatted molecular topology, or removes specific parts (atoms) from a molecular topology file, according to a distance criterion, or an atom pointer list.

6.2.2 Transformation Programs

Programs that transform a given atom coordinate sequence to the atom sequence in the molecular topology, and to the GROMOS coordinate format:

- PROBRK: The program reads Brookhaven tape format.
- PROCS1: The program changes atom sequence and format.
- PROCS2: The program changes atom sequence and format, transforms from oblique to cartesian coordinates.

6.2.3 Programs that generate atom coordinates

- PROGCA: The program generates cartesian coordinates from internal coordinates (bond-lengths, -angles and dihedrals).
- PROSSC: The program substitutes amino acid side-chain coordinates.
- PROGCH: The program generates hydrogen atom coordinates (polar hydrogens).
- PROGHW: item, but for waters
- PROCRY : The program performs crystal symmetry transformations.
- PROBOX: The program puts solute molecule(s) in box with solvent molecules.
- PROION: The program substitutes charged ions for solvent molecules.

6.2.4 Programs that perform simulations

- PROEM: The program performs energy minimization (steepest descents, conj. grad.).
- PROMD: The program performs molecular dynamics (classical, constant temperature or constant pressure).
- PROSD: stochastic dynamics

6.2.5 Programs that analyse configurations or sequences of configurations

- PROAVX: The program calculates the means and root mean square fluctuations of the atom coordinates from a sequence of given protein configurations. It also gives the RMS fluctuations in the atomic principal axes systems. Besides, it yields the third and fourth moments of the distributions of atom coordinates, and the angles between the principal axes and the bondvectors to neighbour atoms.
- PROAVQ: The program calculates for quantities q the means, the root mean square fluctuations, the third and fourth moments, the minima and maxima from a sequence of given protein configurations. q may denote bond-lengths, bond-angles or (improper) dihedrals. The corresponding energies are also averaged. When dihedrals are considered, the transitions between the potential minima are monitored.
- PROAJC: The program averages nmr j-coupling constants.
- PRODR: The program averages atom-atom distance restraints.
- PROAVN: The program averages number of neighbour atoms.
- PROAHB: The program averages hydrogen bonds.
- PROMHB: The program monitors hydrogen bonds (lifetimes).
- PROCOC: The program averages solute or solvent site occupancies.
- PROAVS: The program averages solvent atom positions.
- PROCOX: The program compares two solute configurations: atom positions.
- PROCOQ: The program compares two solute configurations: internal coord.
- PROCOB: The program compares two solute configurations: isotr. b-factors.
- PROCAB: The program compares two solute configurations: aniso. b-factors.
- PROCOD: The program compares two solute configurations: distances.
- PRONBL: The program generates a neighbour list.
- PROCHB: The program generates a hydrogen bond list.

- PROCPS: The program determines solute-solvent nearest neighbours.
- PROCOS: The program compares two solvent configurations: atom positions.
- PROTCF: The program calculates various time correlation functions.

6.2.6 Merge Programs

Programs that merge coordinate files or transform atom coordinates to a special format (interfacing)

- PROTCF : The program calculates various time correlation functions.
- PROMCF: The program merges or reduces coordinate files.
- PROPSF: The program converts to oblique contra-variant fractional coord.
- PROPDF: The program converts to master coordinate file format (display).

6.2.7 Programs for MD X-Ray refinement

Programs that form the molecular dynamics crystallographic (x-ray) refinement (mdxref) supplement to GROMOS87 have not yet completely been adapted to the GROMOS style and are kept in a separate directory:

- PROISF: The program initiates an observed structure factor file in the so called master data file motif format
- PROESF: The program expands the observed structure factors mdf-format to space group p1
- PROEMX: The program performs energy minimization with forces derived from the difference between observed and calculated structure factors
- PROMDX: The program performs molecular dynamics with forces derived from the difference between observed and calculated structure factors

6.3 Comparison between GROMOS and AMBER

6.3.1 Application

Both programs are used for the calculation of structures of biological macromolecules as proteins or nucleic acids. In the calculations, the conformational space, which is accessible to a given molecule, is searched for the global minimum in the energy. In the programs two methods are available for this search, energy minimization (EM) and molecular dynamics (MD). As there are many local minima in the multidimensional conformational space, the MD serves better than the EM as the latter is often trapped by these local minima. Additionally, the MD can include effects of temperature and flexibility of the molecule.

The calculations can be performed both in vacuum and in solvent. In GROMOS, there is the possibility to include a solvent implicit as a source of friction, so that the Langevin-equations of motion are used instead those of Newton. In both programs external conformational data can be introduced as additional constraints for distances and angles. In this way experimental results, e. g. from NMR-spectroscopy, can be used. GROMOS is in this point more flexible, as it provides a better potential-function for the restraints.

6.3.2 Structure of the Programs

Both programs show a similar structure, each being divided in three major building blocks: preparation, simulation and analysis. In the preparation, the topology and the coordinates of the macromolecules are built up. As proteins and nucleic acids are linear polymers of standard building blocks, both programs provide these blocks. So the topology of the whole molecule is obtained by only giving the sequence of the standard residues. Coordinates can then be read in for parts or the whole molecule. The molecule can then be dissolved and counterions can be added. In AMBER, there are four programs for this purpose: PREP to construct building blocks which are not yet in the database; LINK to generate the topology from the sequence; EDIT to read coordinates and solvate; PARM to read all force field parameters which are needed for the simulation. In GROMOS, there are fourteen programs, which perform the same work. There are only little differences in the functions of the preparation modules between AMBER and GROMOS. In AMBER these functions are just more concentrated than in GROMOS.

In the simulation, the energy of the molecule can be minimized, or a trajectory in phase space can be calculated. Each function is implemented in an own program, MIN/MOLDYN in AMBER and PROEM/PROMD in GROMOS. In addition AMBER provides the calculation of normal modes with NMODE and GROMOS provides stochastic dynamics with PROSD.

In the analysis, the generated coordinate sets and trajectories can be used to derive several additional data to simplify the interpretation of the calculation. So internal coordinates can be calculated from cartesian. Other geometric data as distance matrices can be calculated, too. The total energy can be decomposed in the individual contributions of free selectable groups. Parts of the MD-trajectory can be averaged to obtain the mean coordinates and energies. To observe time dependent fluctuations, coordinates and energies can be printed via the simulation time. Additionally, auto- and cross correlations can be calculated. Besides these functions, which are common to both programs, GROMOS provides the calculation of coupling constants, temperature factors, site occupancies and comparisons of solute positions. In AMBER, there are only two programs for analysis, ANAL for EM-calculations and MDANAL for MD-calculations. In GROMOS there are nineteen programs to analyse EM- and MD-calculations.

In both programs, the communication between the single modules is done by reading and writing files, which contain the necessary data. In the preparation, the topology and the starting coordinates are written to separate files. These are read by the simulation modules, which in turn produce new coordinate files. The MD modules produce a whole trajectory, so that a number of coordinate sets is written to a file. Additionally, energies

and velocities can be written in this manner. In the analysis, these files are read, and new ones are produced, which include the final data.

6.3.3 Examples

To cover the most applications in the test, a DNA-doublestrand of eight nucleotides and a linear decapeptid were chosen. The starting coordinates were obtained from the protein data bank at the Brookhaven laboratory. As these do not include polar hydrogen atoms, which are needed for the united-atom calculations, these had to be generated by the GROMOS module PROGCH. In AMBER they are generated automatically from the internal coordinates. From each molecule EM and MD calculations were performed in vacuum and in water. At the moment, the calculations of the DNA with AMBER are still in progress, so that only the results with the peptide will be presented. In GROMOS the solvation was performed with the PROBOX module, in AMBER this was done in the EDIT module. In both programs the same parameters for the solvation were used, so that the molecules were immersed in the same number of solvent molecules. The dna was dissolved in 1675 and the peptide in 524 molecules of water. The calculation times for comparison were chosen for 500 steps of calculation, which corresponds to one picosecond (ps) of MD simulation.

	DNA (GROMOS)	Peptide (GROMOS)	Peptide (AMBER)
EM/vac	243 s	40 s	9 s
EM/solv	1010 s	384 s	250 s
MD/vac	1675 s	325 s	10 s
MD/solv		456 s	200 s

As a picosecond is 10^{-12} second, the MD simulation with both programs is roughly 10^{13-15} times slower than nature. A comparison for GROMOS to the results from other vector computers shows for similar molecules (peptide in solution) nearly the same results. On a CONVEX C2 450 s and on a CRAY-2 only 250 s are needed for 500 steps in GROMOS. A comparison of the conformations and energies at the end of the calculations shows only slight differences between the programs. The remarkable difference is the much better performance of AMBER. On the other hand, the simulation modules of AMBER need 1.7 MB of memory while those from GROMOS need only 1 MB.

6.4 Prophet

The PROPHET environment includes tools for data management, statistical analysis, graphical displays, modelling and simulation, molecular structure manipulation and sequence handling. All tools are integrated by a consistent user interface. In addition, PROPHET is an extensible system that includes a high-level programming language that can be used to customize the scientists work environment.

- **Structural Biochemistry**
 - Tools for building molecules and multimolecules
 - Tools for displaying molecules and multimolecules
 - Tools for manipulating molecules
 - Tools for computing and analysing molecular structures
 - Tools for quantitative structure activity relationship studies
 - Tools for database accessing
- **Modelling**
 - Tools for regression analysis
 - Tools for compartmental analysis
 - Tools: binkin2, dragmodel, fitcomp, fitria, fitsat, ligandplot, probit
- **Statistical Computation**
 - Descriptive Statistic
 - Statistical Testing
 - Curve Fitting
 - Interface to SAS

Chapter 7

Plans

The following program packages are candidates for future installations. If there are other programs which seem to be more likely for installation, please do not hesitate to contact us.

7.1 ab initio Programs

7.1.1 MOLCAS

MOLCAS is a package consisting of a set of computer programs for ab initio quantum chemical computations, coupled to a unit to perform a wide variety of tasks. These programs are assembled by the Quantum Chemistry group at Lund University, Sweden.

MOLCAS has been designed for accurate quantum chemical calculations on ground and excited states of molecular systems, with special emphasis on multiconfigurational wavefunctions.

Highlights of MOLCAS:

- Driver routine for ease of use
- On-line help
- Integral generators
- Closed shell SCF wavefunction generator
- Complete and Restricted Active Space SCF wavefunction generator
- Multi Reference SDCl wavefunction generator
- Average Coupled Pair Functional wavefunction generator
- Coupled Pair (or Modified Coupled Pair) Functional wavefunction generator
- Many Body Perturbation Theory, up to full 4'th order
- One electron properties from any wavefunction generator

- ANO basis set library

The program will be released in the second part of 92.

7.1.2 GAUSSIAN 92

No information is available. GAUSSIAN92 will be shown on the ACS meeting in Summer 92.

7.2 Semiempirical Molecular Orbital Programs

7.2.1 Argus

Argus is an experimental code developed for designing new model Hamiltonians and doing UV/VIS spectra of large systems. While it is well designed and structured, it does not deliver an exhaustive set of options like MOPAC or GAUSSIAN.

Development thus far has been almost totally driven by personal research interests. Some key features are missing. However, there will be ongoing development of Argus at Pacific North-West Laboratory.

- Self Consistent Field (SCF):
 - Restricted Hartree-Fock (RHF)
 - Unrestricted Hartree Fock (UHF)
 - Complex RHF (CRHF)
 - Complex UHF (CUHF).
 - Self Consistent Reaction Field (SCRf) for simulating solvent effects (SCRf for closed shell systems SCF/CI only).
- Configuration Interaction (CI)
- Random Phase Approximation (RPA)
- limited to closed shell systems; Single excited configurations; Singlet states only
- geometry optimization INDO1 systems with S, P atomic orbitals.
- properties calculations
 - dipole moments
 - oscillator strengths (dipole length and dipole velocity)
 - rotary strengths
 - ZDO and Mulliken atomic charges over atoms or fragments for ground and excited states.
- Hartree-Fock Classical Molecular Dynamics. INDO1 systems with S, P atomic orbitals only.

- Hamiltonians
 - INDO1 (spectroscopic and non-spectroscopic)
 - NDDO1
 - EHT
 - ab initio (STO-NG contracted basis sets)

7.3 Density Functional Programs

7.3.1 Dmol

Dmol is a first-principles (*ab initio*) quantum chemistry software package that performs accurate theoretical calculations on a wide range of compounds, including metal clusters, biological compounds, organometallics, and organic compounds. DMol calculates variational self-consistent solutions to the local density functional (LDF) equations, expressed in a numerical atomic orbital basis. The solutions to these equations provide the molecular wavefunctions and electron densities which may subsequently be used to evaluate the energetics, and electronic and magnetic properties of the systems. In addition, evaluation of energy gradients provides a convenient method for determining the equilibrium geometries of these systems. These results provide a reliable, predictive method for theoretically exploring the properties of unknown compounds, as well as a method for explaining, on a microscopic scale, the known properties of existing compounds. The relatively low computational requirements of the method allow the study of larger molecular systems than would be possible with alternative *ab initio* methods.

7.4 Molecular Modelling / Force Field Programs

7.4.1 X-Plor

X-PLOR is a macromolecular modelling program that refines crystal structure data more efficiently than traditional methods. It combines x-ray diffraction data with molecular dynamics to produce structures that are both energy minimized and well fitted to the x-ray data. The use of high temperature molecular dynamics allows the refinement process to escape from local minima, thus decreasing the number of times human intervention is necessary. The energy model used for X-PLOR's molecular dynamics is similar to that of CHARMM. X-PLOR has been implemented with particular attention to efficiency.

Veröffentlichungen des Konrad-Zuse-Zentrum für Informationstechnik Berlin
Technical Reports

März 1992

- TR 91- 1. F. Bornemann; B. Erdmann; R. Roitzsch. *KASKADE - Numerical Experiments.*
- TR 91- 2. J. Lügger; W. Dalitz. *Verteilung mathematischer Software mittels elektronischer Netze: Die elektronische Softwarebibliothek eLib.*
- TR 91- 3. S. W. C. Noelle. *On the Limits of Operator Splitting: Numerical Experiments for the Complex Burgers Equation.*
- TR 91- 4. J. Lang. *An Adaptive Finite Element Method for Convection-Diffusion Problems by Interpolation Techniques.*
- TR 91- 5. J. Gottschewski. *Supercomputing During the German Reunification.*
- TR 91- 6. K. Schöffel. *Computational Chemistry Software for CRAY X-MP/24 at Konrad-Zuse-Zentrum für Informationstechnik Berlin.*
- TR 91- 7. F. A. Bornemann. *An Adaptive Multilevel Approach to Parabolic Equations in Two Space Dimensions.*
- TR 91- 8. H. Gajewski; P. Deuffhard; P. A. Markowich (eds.). *Tagung NUMSIM '91 _5.-8. Mai 1991_ Collected Abstracts and Papers .*
- TR 91- 9. P. Deuffhard; U. Nowak; U. Pöhle; B. Ch. Schmidt; J. Weyer. *Die Ausbreitung von HIV/AIDS in Ballungsgebieten.*
- TR 91-10. U. Nowak; L. Weimann. *A Family of Newton Codes for Systems of Highly Nonlinear Equations.*
- TR 92- 1. K. Schöffel. *Ab initio Quantum Chemical Calculations with GAMESS-UK and GAUSSIAN90 Program Packages - A Comparison - .*
- TR 92- 2. K. Schöffel. *Computational Chemistry Software at ZIB.*

Veröffentlichungen des Konrad-Zuse-Zentrum für Informationstechnik Berlin
Preprints

März 1992

- SC 91- 1. F. A. Bornemann. *An Adaptive Multilevel Approach to Parabolic Equations III.*
- SC 91- 2. R. Kornhuber; R. Roitzsch. *Self Adaptive Computation of the Breakdown Voltage of Planar pn-Junctions with Multistep Field Plates.*
- SC 91- 3. A. Griewank. *Sequential Evaluation of Adjoint and Higher Derivative Vectors by Overloading and Reverse Accumulation.*
- SC 91- 4. P. Deuffhard; F. Potra. *A Refined Gauss-Newton-Mysovskii Theorem.*
- SC 91- 5. B. Fiedler; J. Scheurle. *Discretization of Homoclinic Orbits, Rapid Forcing and "Invisible" Chaos.*
- SC 91- 6. R. H. W. Hoppe; R. Kornhuber. *Multilevel Preconditioned CG-Iterations for Variational Inequalities.*
- SC 91- 7. J. Lang; A. Walter. *An Adaptive Discontinuous Finite Element Method for the Transport Equation.*
- SC 91- 8. K. Gatermann; A. Hohmann. *Hexagonal Lattice Dome - Illustration of a Nontrivial Bifurcation Problem.*
- SC 91- 9. F. A. Bornemann. *A Sharpened Condition Number Estimate for the BPX Preconditioner of Elliptic Finite Element Problems on Highly Nonuniform Triangulations.*
- SC 91-10. G. M. Ziegler. *Higher Bruhat Orders and Cyclic Hyperplane Arrangements.*
- SC 91-11. B. Sturmfels; G. M. Ziegler. *Extension Spaces of Oriented Matroids.*
- SC 91-12. F. Schmidt. *An Adaptive Approach to the Numerical Solution of Fresnel's Wave Equation.*
- SC 91-13. R. Schöpfung; P. Deuffhard. *OCCAL: A mixed symbolic-numeric Optimal Control Calculator.*
- SC 91-14. G. M. Ziegler. *On the Difference Between Real and Complex Arrangements.*
- SC 91-15. G. M. Ziegler; R. T. Zivaljevic. *Homotopy Types of Subspace Arrangements via Diagrams of Spaces.*
- SC 91-16. R. H. W. Hoppe; R. Kornhuber. *Adaptive Multilevel - Methods for Obstacle Problems.*
- SC 91-17. M. Wulkow. *Adaptive Treatment of Polyreactions in Weighted Sequence Spaces.*
- SC 91-18. J. Ackermann; M. Wulkow. *The Treatment of Macromolecular Processes with Chain-Length-Dependent Reaction Coefficients - An Example from Soot Formation.*
- SC 91-19. C. D. Godsil; M. Grötschel; D. J. A. Welsh. *Combinatorics in Statistical Physics.*
- SC 91-20. A. Hohmann. *An Adaptive Continuation Method for Implicitly Defined Surfaces.*
- 92- 1. F. Bornemann; H. Yserentant. *A Basic Norm Equivalence for the Theory of Multilevel Methods.*
- 92- 2. J. Ackermann; K. Helfrich. *Radius of Convergence of the $1/Z$ -Expansion for Diatomic Molecules: The Ground State of the Isoelectronic H_2 Sequence.*