Title
Computer simulation of ligand-biomacromolecule interactions

Author
Lybrand, Terry P.

Publication Date
1984

Peer reviewed|Thesis/dissertation

Computer Simulation of Ligand-Biomacromolecule Interactions: Theory and Applications for Drug-Receptor Interactions

by

Terry P. Lybrand

B. S. in Pharmacy, University of South Carolina, 1980

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Pharmaceutical Chemistry

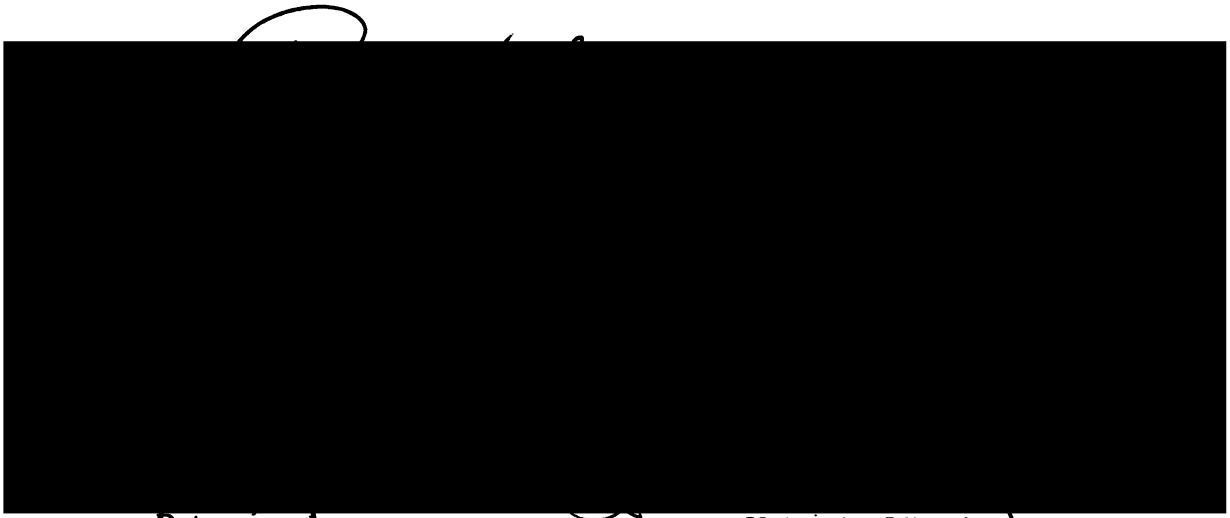
in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA

San Francisco



Date University Librarian

Degree Conferred: . . . . DEC 31 1984 . . . . . .

Abstract

Computer modeling techniques have been applied in the study of ligand–biomacromolecule interactions such as drug–receptor complexes and potential energy functions have been refined to simulate biological systems in a more realistic manner. Both receptor sites of known structure (drug–nucleic acid complexes) and receptor sites of undefined structure (e.g., opiate receptor sites) have been investigated in applications work.

Computer graphics model building studies and molecular mechanics energy minimization calculations were combined to investigate origins for experimentally observed base sequence selectivities for ethidium cation and actinomycin D. Results were in generally good agreement with experimental observations and previous theoretical calculations. Computer model structures for an $(ATGCAT)_2$:actinomycin complex were also compared with 2D–NMR NOE data. The agreement between model calculations and experimental solution structures was quite encouraging.

Modeling techniques were then extended to studies of ligand interactions with "receptor" sites of undefined structure. Electrostatic potential surface calculations were used to search for characteristics (pharmacophores) that might rationalize variations in pharmacological properties for a series of clozapine analogs (neuroleptics). Distinctive electrostatic potential surface patterns correlated nicely with the pharmacological variations between molecules. In a related study, molecular mechanics conformational analysis and electrostatic potential calculations were employed to explain differential binding behavior for a series of opioid

ligands at two distinct binding sites. Variable electrostatic potential surface patterns, together with a general lack of conformational flexibility, could be used to rationalize the differential behavior at the two binding sites. One binding site appears to require a specific conformational pharmacophore, while the other binding site seems to be extremely sensitive to electrostatic potential characteristics.

Finally, potential functions with explicit terms for nonadditive energy contributions were developed and tested for water–water and water–ion interactions. The new potential functions yielded better results for many ion–water systems than any previously reported models. They may offer the possibility of computer simulation models for ligand–biomacromolecule complexes that include solvent and counterions in a realistic manner.

Acknowledgements

This thesis is the product of a desire to apply the techniques of theoretical computational chemistry to ligand–receptor interactions. The results have been mixed, with some successes and failures. Fortunately, I have had a research advisor whose unbounded enthusiasm and energy made the failures easier to tolerate while new solutions were sought. I thank Peter Kollman for his guidance, interest in my goals and objectives, encouragement, and friendship. I cannot imagine a better advisor than you, Peter.

I also must thank Tack Kuntz for serving as my orals committee chairman and reading this dissertation. Your advice was always thoughtful, well directed, and of great benefit to me.

I am grateful to George Kenyon for serving on my dissertation committee on relatively short notice after the departure of Fred Wolff.

I would be remiss if I did not thank Fred Wolff, as he greatly influenced my decision to attend UCSF.

Thanks are also due Bob Langridge and the computer graphics laboratory staff. Without their efforts, much of this work would not have been feasible.

Special thanks go to Drs. Jim Wynn and Carl Bauguess, undergraduate professors who encouraged me to pursue science as a career and allowed me considerable freedom to explore my interests.

I cannot adequately express my thanks to my parents. Without their support and encouragement in all my endeavors, none of this would have

been possible.

This dissertation is dedicated to my wife, Becky, whose love and understanding are the most important "discoveries" I have made, or will make, in my life.

Table of Contents

Abstract	ii
Chapter 1: Introduction	1
Chapter 2: Ligand Interactions at Binding Sites of Known Structure	8
2.1 - Background	8
2.2 - Methods and Procedures	10
2.3 - Results and Discussion	13
2.4 - Conclusions	40
Chapter 3: Ligand Interactions at Binding Sites of Unknown Structure	42
3.1 - Background	42
3.2 - Methods	43
3.3 - Results and Discussion	51
3.4 - Conclusions	63
Chapter 4: Extension of Computer Models to Include Environment	65
4.1 - Background	65
4.2 - Potential Functions	66
4.3 - Parameterization and Calculations	70
4.4 - Discussion	92
4.5 - Conclusions	101
Chapter 5: Summary	102
References	105
Appendix 1: Potential Function Parameters	117
Appendix 2: NMR NOE vs. Molecular Mechanics Distances	124
Appendix 3: Source Code	137

CHAPTER 1

Introduction

Ligand–biomacromolecule interactions are of great interest and importance in biological processes. Most pharmacologically or physiologically important molecules are believed to form complexes with specific receptors as an initial step in production of pharmacological and/or physiological responses. For example, the endogenous neurotransmitter norepinephrine is thought to propagate impulses in the sympathetic autonomic nervous system by forming a complex with specific adrenergic receptors on postsynaptic neurons. Interaction of a ligand with a receptor may induce a conformational change in the receptor (generally believed to be a protein or proteinaceous complex) which, in turn, leads directly to the biological effect or initiates other molecular mechanisms that eventually produce the biological effect.

Numerous experimental techniques have been used to characterize ligand–biomacromolecule interactions such as spectroscopic methods,[1-3] thermodynamic and kinetic measurements,[4,5] and X–ray crystallographic studies.[6-8] Computer modeling techniques can provide detailed atomic resolution information about ligand–biomacromolecule interactions that complements available experimental data. Computer models may aid in rationalization of experimental results and serve as a powerful predictive tool, provided the models realistically simulate the systems of interest. The former capability alone justifies the continued refinement and application of computer modeling techniques for ligand–biomacromolecule complexes. The

latter capability has elicited great interest, especially among those engaged in drug design, and some promising results have already been obtained. For example, straightforward model building studies have led to design of new compounds (some structurally unrelated to the prototype ligands) with excellent binding affinity at target receptor sites.[9,10] Computer modeling techniques using empirical potential energy function calculations have been less successful in quantitative prediction of relative ligand affinities at a receptor site, but much useful information may still be obtained.[11] Computer models with reliable quantitative predictive capabilities would be of considerable utility in the drug design process.

This dissertation first describes applications of computer modeling techniques to ligand–biomacromolecule interactions at a receptor site of known structure. Then, extensions of these techniques to investigate ligand interactions at receptor sites of undefined composition and structure are discussed. Finally, empirical potential energy function refinements are explored which may lead to the development of more realistic computer models for ligand–biomacromolecule interactions.

Application of computer modeling techniques in the study of ligand interactions at a receptor site of known structure is more straightforward than the investigation of ligand interactions at a structurally undefined receptor site. Often, X–ray crystal structures may exist for a ligand–receptor complex and this crystal structure can be used as a guide in constructing receptor complex models for other ligands. Even if no structural data are available for ligand complexes at the receptor site, knowledge of the composition and three–dimensional structure of the receptor make it possible to propose reasonable ligand–receptor complex

models with the aid of computer graphics techniques. The ligand–receptor complex models, whether obtained directly from X–ray data or from model building studies, can then be used as a starting point for additional computer simulation such as energy minimization calculations.

As representative examples of ligand–biomacromolecule interactions at receptor sites of known structure, several drug–nucleic acid complexes were studied with molecular mechanics energy minimization. Intercalation complexes for ethidium cation with base–paired dinucleoside monophosphate and hexanucleoside pentaphosphate duplexes as well as actinomycin D (AMD) with hexanucleoside pentaphosphates were examined. The relative gas phase binding energies successfully reproduced qualitatively the binding preferences within isomeric nucleotide sequence complexes for these ligands. The calculations also suggested that sequence selectivities for ethidium were determined primarily by the relative energy cost for distortion of a given nucleic acid sequence from a B–DNA conformation to an intercalation site geometry, in good agreement with previous theoretical studies.[12-14] In contrast, the molecular mechanics calculations indicated AMD sequence preferences were governed principally by *specific* drug–nucleic acid hydrogen bonds, again in good agreement with binding models proposed on the basis of X–ray crystallographic[15,16] and spectroscopic studies.[17-23] The computer modeling studies also provided a possible rationalization for the importance of intact cyclic pentapeptide side chains in AMD. The models suggested that the cyclic pentapeptides effectively shield the specific drug–nucleic acid hydrogen bonds from solvent disruption. AMD derivatives with acyclic or modified peptide side chains have greatly diminished biological activity and nucleic acid binding affinities relative to AMD. The

models implied that acyclic or modified peptide side chains would not protect the drug–nucleic acid hydrogen bonds from solvent exposure. Finally, the computer studies indicated that AMD might exhibit some base sequence selectivity beyond intercalation site residues (i.e. AMD might display some preference for one particular tetranucleotide or hexanucleotide sequence over another). These results may be helpful in attempts to design ligands that bind specifically to a given base sequence in a DNA fragment. Such ligands could be useful experimental probes and might have therapeutic utility. For example, a ligand that selectively binds to a particular DNA fragment corresponding to a gene initiation sequence might be used to suppress production of proteins encoded by that gene.

Unlike the drug–nucleic acid intercalation complexes, most drug–receptor interactions of interest to medicinal chemists and pharmacologists involve "receptor" sites of undefined composition and three–dimensional structure. Frequently, the only information available about the nature of the ligand–receptor complex is the information contained in the ligands (agonists and antagonists) that interact with these receptors. It is assumed that a ligand possesses some set of physical features which enable it to form a complex with a receptor site. One logical approach in the study of ligand interactions at a receptor site of undefined structure involves the systematic examination of a series of ligands that interact at the receptor site. A set of physical characteristics (i.e., a pharmacophore) is sought which rationalizes pharmacological properties and binding behavior for the series of ligands at the receptor site. Two examples of possible pharmacophores are: 1) a specific spatial orientation of functional groups such as hydrogen bond donor and/or acceptor substituents and 2) a distinctive electrostatic potential pattern over the molecular surface. The

search for pharmacophoric patterns can be quite challenging, especially when the ligands in question are flexible molecules, because an extremely large number of conformations must be evaluated for each compound. If the ligands of interest are relatively rigid, the number of available conformations is limited, and each conformation can be extensively studied. Definition of a suitable pharmacophore is usually a complicated process characterized by an initial proposal for the pharmacophore based on the study of a few ligands. The initial pharmacophore hypothesis is then applied to a wider range of ligands to determine whether it satisfactorily explains binding data or pharmacological properties. If necessary, refinements are made in the initial pharmacophore proposal to account for inconsistencies with experimental data, and the refined pharmacophore is again tested against a larger data base of ligands. This procedure is continued until the refined pharmacophore successfully rationalizes the available experimental data.

Two examples of ligand interactions at unknown receptor sites are reported in this dissertation. First, clozapine analogs were studied in an attempt to rationalize variable pharmacological properties in the series of compounds. In a second study, a hypothesis was sought to explain relative binding properties for a series of opioid compounds at two different sites.

The clozapine analogs displayed two distinct pharmacological profiles.[24] Clozapine possesses neuroleptic activity comparable to that of chlorpromazine, but lacks many of the serious side effects associated with neuroleptic agents (e.g. extrapyramidal side effects). However, some clozapine analogs exhibit the traditional pharmacological profile of neuroleptic agents, with pronounced extrapyramidal side effects. All compounds were quite rigid, differing from each other only in degree and/or

position of chlorine substitution. Due to the lack of conformational freedom in these molecules, electrostatic properties were evaluated in a search for distinguishing characteristics among molecules. The electrostatic potential surface patterns allowed classification of the molecules into two categories, in good agreement with experimental classifications based on the pharmacological data. A reliable pharmacophore for definition of clozapine–like agents could be of great interest, as these compounds might be missed in typical neuroleptic screening procedures because they do not exhibit the normal pharmacological properties of a neuroleptic agent.

The opioid compounds exhibited differential binding behavior at μ opiate receptors and the λ site, a recently discovered high–affinity binding site for 4,5–epoxymorphinans.[25] Specifically, all compounds bind tightly at the μ receptor, but only the 4,5–epoxymorphinans display significant binding at the λ site. The μ site is thought to be the primary receptor responsible for mediation of opioid analgesic effects, while no function has yet been defined for the λ site. Although these ligands possessed limited structural flexibility, no conformational factors were discovered to rationalize the differential binding behavior at both sites. However, electrostatic potential surface characteristics did suggest a basis for the selective binding behavior at the λ site. Thus, it appeared that a key determinant for ligand selectivity at the λ site might involve electrostatic characteristics, whereas the μ receptor seemed rather insensitive to the electrostatic differences in these molecules.

Finally, this dissertation describes the development of models to incorporate solvent (and other environmental components such as counterions) in computer simulations in a realistic manner. The inability to properly model solvent and ion effects has been one major inadequacy of

computer models that incorporate potential energy function calculations. Thus, models which realistically simulate solvent and counterion effects should yield more reliable results in ligand–receptor complex calculations. New semi–empirical potential energy functions were derived for water–water and water–ion interactions. These new potential functions included terms for nonadditive interaction energy components such as polarization and exchange repulsion. The potential functions modeled water–water interactions in both gas and condensed phases with good success and yielded excellent results for gas phase ion hydration enthalpies as well as solvation enthalpies and structures in liquid systems.

CHAPTER 2

Ligand Interactions at Binding Sites of Known Structure

2.1 Background

One generally accepted mode for binding of planar aromatic ligands to nucleic acids is intercalation.[26,27] Ligand–nucleic acid intercalation complexes are particularly appealing problems for computer modeling techniques. The constraints imposed on nucleic acid helix duplexes by base pairing and base stacking interactions restrict the number of conformations available to the helix. Thus, reasonable intercalation complex models may be constructed for any suitable ligand using computer graphics techniques and information from detailed X–ray studies of a few well characterized intercalation structures.[28-31]

Two molecules that interact with nucleic acids via an intercalation mechanism are ethidium cation and actinomycin D (AMD). Ethidium (Figure 2.1) is a cationic dye with trypanocidal activity. It has been used for treatment of sleeping sickness and some infectious diseases in livestock[32] and is widely employed as an experimental probe in biochemical and physical studies of nucleic acids. AMD (Figure 2.2) is a chromopeptide antibiotic with utility as a chemotherapeutic agent in the treatment of Wilm's tumor[33] and gestational choriocarcinoma.[34] The biological activity of AMD is thought to arise from its ability to bind to double–stranded DNA and subsequently inhibit DNA–dependent RNA polymerase.[35] Its medicinal usefulness is restricted by a narrow therapeutic index, toxicity, and limited spectrum of

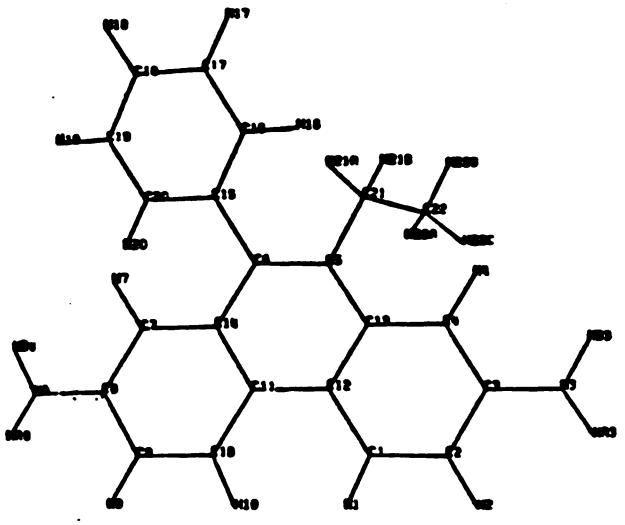


Figure 2.1: Ethidium cation

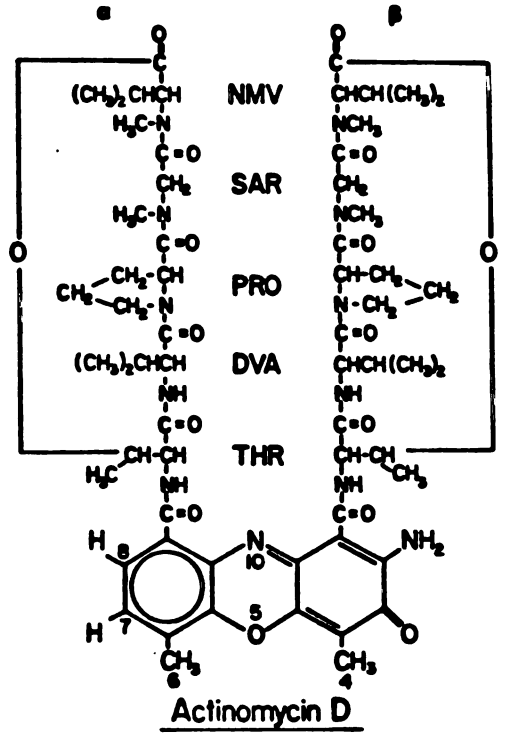


Figure 2.2: Actinomycin D (AMD)

activity.[36]

Experimental evidence exists to support an intercalation binding mode for ethidium[30,31] and AMD.[23,37] Molecular mechanics energy minimization calculations were used to determine relative sequence preferences for these ligands and examine structural details of the intercalation complexes. Complexes of ethidium with deoxydinucleoside monophosphate minihelices $d(ApT)_2$, $d(TpA)_2$, $d(A_2 \cdot T_2)$, $d(CpG)_2$, $d(GpC)_2$, and $d(G_2 \cdot C_2)$, and deoxyhexanucleoside pentaphosphate duplexes $d(CGCGCG)_2$, $d(GCGCGC)_2$, $d(G_6 \cdot C_6)$, $d(ATATAT)_2$, $d(TATATA)_2$, and $d(A_6 \cdot T_6)$ were studied. Complexes of AMD with deoxyhexanucleoside pentaphosphates $d(GCGCGC)_2$, $d(GCCGGC)_2$, $d(GCATGC)_2$, $d(GCTAGC)_2$, and $d(ATGCAT)_2$ were also investigated. The computed relative sequence preferences and structural characteristics were then compared with experimental results[15-23,30-31,37-46] and with previous theoretical studies[12-14] for these ligands.

2.2 Methods and Procedures

Molecular mechanics calculations were performed on all complexes using AMBER,[47] a package of computer programs designed to model large systems such as ligand–biomacromolecule complexes with empirical potential energy functions. The energy functions include harmonic bond stretching and bond angle bending terms, a truncated Fourier series for torsion angle terms, and standard Lennard–Jones and electrostatics terms for non–bonded interactions. The functions also include an explicit term for hydrogen bonding interactions. The form of the potential is as follows:

$$E=\sum_{bonds} K_b(R-R_e)^2 + \sum_{angles} K_\theta(\theta-\theta_e)^2 + \sum_{dihedrals} \frac{K_d}{2}\left[1+\cos(n\Phi-\gamma)\right] + \sum_{non-bonded}\left[\frac{B_{ij}}{r_{ij}^{12}} - \frac{A_{ij}}{r_{ij}^{6}} + \frac{q_i q_j}{\varepsilon_{ij} r_{ij}}\right] + \sum_{H-bonds}\left[\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}}\right] \quad (2.1)$$

In evaluating the energy of a complex, non–bonded terms were summed over all atom pairs, i and j, except those involved in 1-2 (bond) and 1-3 (bond angle) interactions. The hydrogen bond term was evaluated between atoms defined as suitable hydrogen bond donors or acceptors. Cutoffs were used to suppress evaluation of non–bonded and hydrogen bond interactions between pairs of atoms that are spatially far removed from each other and, thus, have negligible interaction energies. Values for parameters in the potential functions (K_θ,θ_e,K_b, R_e,K_d,n,γ,A_{ij},B_{ij}, C_{ij},and D_{ij}) depend on atom type and have been reported for the nucleic acids.[48,49] Additional parameters for ethidium and AMD are given in Appendix 1. The atomic charges for the intercalators are from CNDO/2 calculations. The other parameters were derived according to standard procedures.[48] Structural parameters were taken principally from standard bond lengths and angles[48] or from crystal structure data.[50] All calculations were done with a dielectric constant proportional to the magnitude of the interatomic distance. The basis for a distance-dependent dielectric constant has been described elsewhere.[11,51] Two computational models were used in these calculations. The first, a united–atom model, included only those hydrogens attached to heteroatoms in the nucleic acid helices. The second, an all–atom model, included all hydrogen atoms in the ligands and nucleic acid helices explicitly.

Initial geometries for the intercalation complexes were built with the aid of MIDAS[52] and CHEM.[53] interactive computer graphics modeling programs designed for real-time manipulation of molecular structures using an Evans

and Sutherland color PS2 picture system. Final structures after energy minimization were also analyzed using the aforementioned graphics software.

Deoxydinucleoside monophosphate minihelix intercalation complexes for ethidium were based on crystal structures for a $(CpG)_2$:ethidium complex.[30,31] All other deoxydinucleoside monophosphate intercalation complexes for ethidium were constructed by setting deoxyribose phosphate torsion angles to those values in the $(CpG)_2$ complexes. When necessary, small adjustments in the structures were made to assure that good base pair hydrogen bonds and base–ligand stacking interactions were maintained.

Intercalation geometries for base–paired deoxyhexanucleoside fragments were constructed using computer graphics model building techniques and followed a procedure similar to that employed by Alden and Arnott.[54] Starting with a hexamer fragment in a standard B-DNA geometry, the third deoxyribose unit in each strand was repuckered to a C3'-endo conformation using options in the molecular mechanics software package AMBER.[47] An intercalation site was formed between the third and fourth base pairs by altering torsional angles in that deoxydinucleoside monophosphate region of the hexamer. Finally, small adjustments were made in torsional angles throughout the hexamer fragment to assure good base pair hydrogen bonding and base stacking in the entire helix. This approach yielded an intercalation geometry with a mixed C3'-endo(3'-5')C2'-endo sugar pucker at the intercalation site and a base-base separation of about 6.8Å. The remainder of the helix had C2'-endo sugar puckers and base-base separations of about 3.4Å. The unwinding angle for the deoxyhexanucleoside duplex fragment was ~ 26°.

Ethidium and AMD molecules were incorporated in the complexes with the aid of computer graphics. The ligands were so positioned as to give good stacking interactions between chromophore and nucleic acid bases. A further effort was made to avoid any obvious bad steric contacts between bases and the cyclic pentapeptide side chains of AMD. These structures were then subjected to energy minimization. In some cases, the consequences of uniform C2'-endo sugar puckers at the intercalation site were examined as well. For these calculations, the energy-refined structures for the model-built C3'-endo(3'-5')C2'-endo mixed sugar pucker models were used as starting conformations. The C3'-endo deoxyribose units were constrained to C2'-endo conformations, the complexes were minimized with the constraints, the constraints were then removed and the complexes were relaxed completely with energy minimization. All minimizations were considered to be converged when the root mean square derivative of the energy function relative to atomic coordinate changes was 0.1 kcal/Å or less and the relative change in total energy from one cycle to the next became smaller than 1.0×10^{-5} kcal.

2.3 Results and Discussion

Ethidium complexes

Table 2.1 displays component energy terms for two low energy conformations of the intercalation complexes in each sequence at the united–atom level. The binding energy (ΔE) is the energy of the intercalation complex minus the energy of the corresponding refined nucleic acid minihelix in a B–DNA conformation.[56]

$$\Delta E = E_{complex} - (E_{B-DNA} + E_{ligand}) \qquad (2.2)$$

The drug–helix interaction energy is the total intermolecular energy between the ligand and nucleic acid helix. The helix destabilization energy is the helix *intra*– molecular energy minus the energy of a refined nucleic acid minihelix in a B–DNA conformation and represents the energy cost for creation of an intercalation site in each helix.

In the deoxydinucleoside monophosphate complexes, intrastrand hydrogen bonds between terminal hydroxy groups and phosphate oxygens of the backbone formed during minimization of some structures. As discussed previously,[56] it is desirable to compare the relative energies of these complexes without the intrastrand hydrogen bonds. Thus, the H–O5'–C5'–C4' dihedral angles were constrained at 180°, thereby preventing intrastrand hydrogen bonds in these complexes. Calculations both with and without this

TABLE 2.1*

Complex	E_T	ΔE	Drug–helix E	Destab. E
$d(CpG)_2$	-198.1(-194.7)	-5.5(-2.1)	-80.3(-76.7)	+31.4(31.5)
$d(GpC)_2$	-200.3(-196.4)	-5.2(-1.3)	-79.8(-76.0)	+30.3(31.4)
$d(G_2 \cdot C_2)$	-199.5(-195.6)	-7.9(-4.0)	-80.8(-76.9)	+29.5(29.9)
$d(TpA)_2$	-239.7(-236.2)	-9.8(-8.3)	-79.2(-75.8)	+25.8(26.1)
$d(ApT)_2$	-240.0(-236.6)	-10.6(-7.2)	-77.8(-74.1)	+23.6(23.7)
$d(A_2 \cdot T_2)$	-240.2(-236.3)	-9.4(-5.5)	-77.9(-74.7)	+24.9(26.4)

Table 2.1: Component energies of refined deoxydinucleoside monophosphate intercalation complexes at the united-atom model level for conformation A (conformation B energy components in parentheses).
* E_T : Total energy of the complex
ΔE : Binding energy
Drug–helix E : Drug–helix interaction energy
Destab. E : Helix destabilization energy

dihedral constraint indicate that there is no substantial difference in the refined structures other than the presence or absence of the intrastrand hydrogen bonds.

Upon energy refinement of the various deoxydinucleoside intercalation complexes, two distinct low energy conformations were observed. The significance of these two conformations was examined using constrained minimization. Each complex was induced to assume both conformations by constraining backbone torsional angles to the desired values, minimizing, removing the constraints, and relaxing each structure completely. For all sequence complexes, the two conformations proved to be local minimum structures. Table 2.2 lists initial and refined backbone torsion angles for the two low energy conformations of the $d(CpG)_2$:ethidium complex (note α and β angle values). Torsion angles for the two low energy conformations of all other base sequence complexes are quite similar and have not been listed here. Conformation B is a relative local minimum energy conformation corresponding to the Sobell crystal structure (initial values). Conformation A represents a new local minimum energy conformation with a much altered geometry around torsion angles α and β of strand 1. This altered conformation allows for the formation of a stronger hydrogen bond between the phosphate oxygen of strand 1 and an exocyclic amino hydrogen of ethidium (see Figure 2.3 for a comparison of the two conformations). As can be seen from Table 2.1, the predicted base sequence preferences of ethidium in either low energy conformation model do not agree well with experimental results[43-45] (the experimental binding preference is $d(CpG)_2 \geq d(G_2 \cdot C_2) > d(GpC)_2$) or previous theoretical work.[12-14] Note especially that the GpC minihelix is relatively less destabilized than its pyr–(3'-5')–pur counterpart.

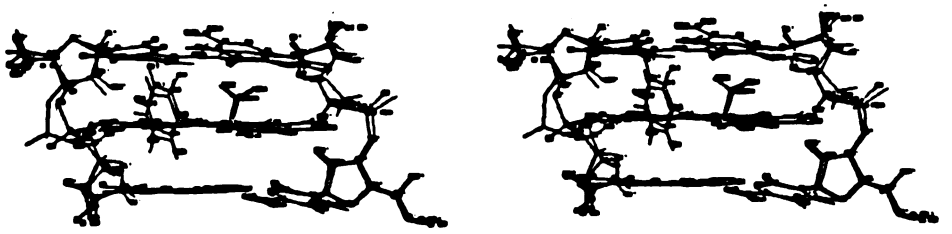


Figure 2.3: Conformation A vs B for $(CpG)_2$:Ethidium complex (Conformation A is the labeled system. Conformation B is unlabeled.)

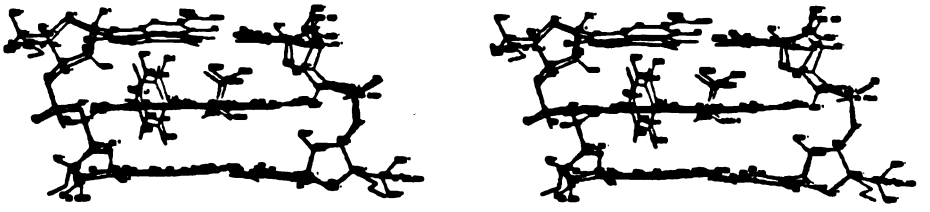


Figure 2.4: All-atom vs united-atom models for $(G_2 \cdot C_2)$:Ethidium (All-atom model is labeled.)

This is in contrast to the interpretations from previous work,[12-14] wherein the preference of ethidium cation for pyr-(3'-5')-pur sequences over pur-(3'-5')-pyr sequences was related to the greater relative destabilization of the pur-(3'-5')-pyr sequences upon opening to an intercalation geometry. Further, the homopolymer complexes appear to be unrealistically stable relative to the corresponding heteropolymers.

Careful study of the refined complexes with computer graphics suggested that explicit inclusion of hydrogens in the minihelix might alter the interactions of ethidium with the helix. Therefore, calculations were performed on all complexes using an all–atom model. These results are given in Table 2.3. As can be seen from the table, the sequence preferences are now in good agreement with experiment and previous calculations. Additionally, ethidium sequence selectivity appears to be governed primarily

TABLE 2.2

Angle	Strand 1			Strand 2		
	Initial	Refined A	Refined B	Initial	Refined A	Refined B
χ_1	210°	208°	218°	201°	210°	207°
δ_1	87°	86°	84°	84°	82°	82°
ε	226°	187°	200°	225°	203°	201°
ζ	281°	280°	276°	291°	284°	275°
α	286°	221°	316°	291°	304°	315°
β	210°	293°	195°	224°	205°	197°
γ	73°	76°	72°	55°	68°	73°
χ_2	287°	287°	266°	295°	271°	265°
δ_2	132°	120°	149°	134°	144°	150°

Table 2.2: Refined backbone torsion angles of two low energy conformations for CpG isomer deoxydinucleoside monophosphate intercalation complex (united–atom model). Refined A refers to conformation A, Refined B refers to conformation B. Angle notation is that given in Dickerson *et al.*[57] The subscripts refer to the first or second glycosidic and sugar pucker torsional angles in each strand.

by the helix destabilization energy (i.e., the energy cost for creating an intercalation site in the deoxydinucleoside monophosphate minihelix). This is also in good agreement with previous theoretical work on the origin of ethidium sequence preferences. Given these rather different results for the all–atom vs. united–atom model calculations, detailed analyses were performed to discover the basis for the discrepancies between the two models.

Extensive comparison of minimized structures for each base sequence revealed that the conformations were quite similar for all–atom and united–atom models. Root mean square deviations for atom positions in analogous all–atom and united–atom model complexes were generally about 0.2Å. Some subtle changes in ethidium orientation and position occurred, especially in the $d(G_2 \cdot C_2)$ complex, and would seem to be a result of the explicit inclusion of hydrogens at the C2' and C5' positions of deoxyribose

TABLE 2.3*

Complex	E_T	ΔE	Drug–helix E	Destab. E
$d(CpG)_2$	-208.6(-204.6)	-9.6(-5.6)	-80.3(-76.6)	+25.4(26.3)
$d(GpC)_2$	-208.4(-205.8)	-5.4(-2.8)	-78.9(-76.2)	+28.6(28.6)
$d(G_2 \cdot C_2)$	-206.8(-204.1)	-6.8(-4.1)	-78.6(-76.5)	+27.1(27.7)
$d(TpA)_2$	-166.0(-163.1)	-12.1(-9.2)	-78.7(-75.1)	+21.7(21.3)
$d(ApT)_2$	-165.2(-163.8)	-9.4(-7.8)	-78.2(-74.4)	+24.0(21.9)
$d(A_2 \cdot T_2)$	-165.5(-162.3)	-8.8(-5.2)	-77.6(-74.4)	+23.7(24.2)

Table 2.3: Component energies of refined deoxydinucleoside monophosphate intercalation complexes at the all-atom model level for conformation A (conformation B values in parentheses).
* See Table 2.1 legend for heading notation.

TABLE 2.4

Angle	Strand 1			Strand 2		
	Initial	Refined A	Refined B	Initial	Refined A	Refined B
χ_1	210°	207°	207°	201°	199°	195°
δ_1	87°	76°	75°	84°	77°	76°
ε	226°	183°	207°	225°	196°	206°
ζ	281°	269°	284°	291°	290°	276°
α	286°	222°	301°	291°	291°	309°
β	210°	290°	207°	224°	235°	202°
γ	73°	93°	68°	55°	68°	76°
χ_2	287°	278°	274°	295°	275°	270°
δ_2	132°	124°	143°	134°	149°	151°

Table 2.4: Refined backbone torsion angles of two low energy conformations for CpG isomer deoxydinucleoside monophosphate intercalation complex (all–atom model). Refined A refers to conformation A, Refined B refers to conformation B. Angle notation is that given in Dickerson *et al.*[97] The subscripts refer to the first or second glycosidic and sugar pucker torsional angles in each strand.

units around the intercalation site. Figure 2.4 shows all–atom versus united–atom model structures for the $d(G_2 \cdot C_2)$ complex. These subtle changes in the ethidium position in the all–atom model seem to account for the less favorable binding energy of the homopolymers relative to the heteropolymers, as they diminish drug–helix interactions to some extent by pushing ethidium out of the intercalation site slightly. The overall lack of conformational differences between the all–atom and united–atom models was somewhat surprising. The differences in the absolute energies of the minihelices upon conversion from a united–atom to an all–atom model arise not from major conformational differences between the two models, but rather from small differences in the parameter sets and are not physically significant. In this series of calculations, the all–atom model was more "tolerant" of small changes in bond angles (±1°) and displayed less variance in the values of dihedral angles compared to the united–atom model. The

differences in these terms accounted for most relative binding energy differences within a particular model as well as differences between models.

For the deoxyhexanucleoside pentaphosphates, two sets of calculations were performed for both all–atom and united–atom models. The model–built intercalation complexes described in the methods section were minimized and their energy components are reported in Table 2.5 for the united–atom models. A second set of calculations examined the consequences of uniform C2'–endo sugar puckers at the intercalation site in CG isomers using the procedure outlined in the methods section. Similar uniform sugar pucker calculations were not performed for the AT isomers as previous studies

TABLE 2.5*

Complex	E_T	ΔE	Drug–helix E	Destab. E
$d(ATATAT)_{2_m}$	-705.4	-38.5	-111.2	+26.5
$d(A_6 \cdot T_6)_m$	-709.5	-38.1	-110.0	+26.2
$d(TATATA)_{2_m}$	-701.6	-36.2	-107.4	+25.6
$d(GCGCGC)_{2_m}$	-591.6	-30.0	-107.5	+32.0
$d(G_6 \cdot C_6)_m$	-585.6	-37.1	-111.0	+29.1
$d(CGCGCG)_{2_m}$	-593.6	-31.6	-109.3	+30.5
$d(GCGCGC)_{2_u}$	-604.8	-43.2	-114.9	+26.2
$d(G_6 \cdot C_6)_u$	-568.2	-19.7	-101.3	+36.9
$d(CGCGCG)_{2_u}$	-598.2	-38.3	-107.3	+24.1

Table 2.5: Component energies of deoxyhexanucleoside pentaphosphate intercalation complexes at the united–atom model level (subscript m indicates mixed sugar puckers at intercalation site, subscript u indicates uniform sugar puckers).
* See Table 2.1 legend for heading notation.

suggested that they should prefer the mixed sugar pucker, even at locations other than the intercalation site.[58] The results of these calculations are also given in Table 2.5. As can be seen from the table, the relative binding energies are dependent on sugar pucker model and intercalation sequence preferences are determined by a complex set of variables including drug–helix interaction energy, helix destabilization, and conformational variations in the helix backbone. It is interesting to note that the CG heteropolymers prefer a uniform C2'-endo sugar pucker around the intercalation site whereas the homopolymer prefers a mixed C3'-endo-(3'-5')-C2'-endo sugar pucker. Table 2.6 lists backbone torsion angles after energy refinement for the *d*(*CGCGCG*)$_2$:ethidium complex at the intercalation site. The torsion angles for all other base sequence complexes are similar and have not been listed here.

TABLE 2.6

Angle	Strand 1			Strand 2		
	B–DNA	A	B	B–DNA	A	B
χ_3	63°	34°	69°	63°	28°	78°
δ_3	140°	84°	147°	140°	83°	144°
ε	185°	188°	193°	185°	184°	188°
ζ	252°	280°	260°	252°	270°	191°
α	286°	197°	175°	286°	202°	180°
β	178°	186°	168°	178°	189°	193°
γ	59°	182°	181°	59°	182°	181°
χ_4	68°	68°	69°	66°	69°	
δ_4	144°	157°	140°	144°	159°	158°

Table 2.6: Refined backbone torsion angles around the intercalation site in (*CGCGCG*)$_2$:Ethidium complexes with A) mixed C3'–endo(3'–5')–C2'–endo sugar puckers and B) uniform C2'–endo sugar puckers at the intercalation site. Torsion angles for a B–DNA helix duplex[56] are given for comparison. Angle notation is that given in Dickerson *et al.*[57]

As with the deoxydinucleoside monophosphate intercalation complexes, calculations were performed on the deoxyhexanucleoside complexes using an all–atom model. The results of these calculations are given in Table 2.7 for mixed and uniform sugar pucker conformations. The backbone torsion angles vary little (±2° for most angles) from the values reported for the united–atom models in Table 2.6. Figure 2.5 displays the mixed versus uniform sugar pucker models for the $d(CGCGCG)_2$:ethidium complexes. Like the deoxydinucleoside complexes, use of the all–atom model in the deoxyhexanucleoside intercalation complexes alters the results. In the CG isomer complexes, the pyr-(3'-5')-pur sequence is preferred with the homopolymer next and the pur-(3'-5')-pyr sequence least favorable as was

TABLE 2.7*

Complex	E_T	ΔE	Drug–helix E	Destab. E
$d(ATATAT)_{2_m}$	-423.4	-42.0	-111.7	+23.7
$d(A_6 \cdot T_6)_m$	-431.4	-40.8	-109.7	+24.1
$d(TATATA)_{2_m}$	-422.8	-42.0	-106.4	+20.5
$d(GCGCGC)_{2_m}$	-561.1	-32.1	-109.8	+32.2
$d(G_6 \cdot C_6)_m$	-557.7	-35.1	-109.9	+30.3
$d(CGCGCG)_{2_m}$	-563.9	-36.6	-108.6	+27.2
$d(GCGCGC)_{2_u}$	-559.8	-34.4	-114.8	+35.1
$d(G_6 \cdot C_6)_u$	-535.1	-32.7	-104.4	+27.0
$d(CGCGCG)_{2_u}$	-560.2	-36.2	-106.7	+26.2

Table 2.7: Component energies of deoxyhexanucleoside pentaphosphate intercalation complexes at the all–atom level (subscript m indicates mixed sugar puckers at intercalation site, subscript u indicates uniform sugar puckers).
* See Table 2.1 legend for heading notation.

the case in the deoxydinucleoside complexes. In the AT isomer complexes, pyr-(3'-5')-pur and pur-(3'-5')-pyr sequences bind ethidium cation equally well with the homopolymer complex being less favorable. The discrepancy in ethidium binding behavior in AT isomers between deoxydinucleoside and deoxyhexanucleoside complexes may be explained at least in part by the tendency of AT hexamers to undergo more extensive sugar repuckering during energy refinement, thus resulting in more numerous relative local minima. This behavior is less common in the CG isomers which present a consistent picture between deoxydinucleoside and deoxyhexanucleoside models. Further, the primary determinant of sequence preference appears to be the helix destabilization energy as was the case with the deoxydinucleoside intercalation complexes.

Discussion

The ethidium calculations reported here raise several interesting points. First, the deoxydinucleoside complex calculations at the all–atom model level give results in excellent agreement with both experiment and other theoretical studies, whereas the energies calculated at the united–atom model level are somewhat contradictory to the all–atom results. This trend is also observed at the deoxyhexanucleoside level. The results suggest that helix destabilization (i.e. the energy required to form an intercalation site in a helix fragment of some given sequence) is a principle determinant in governing base sequence preference for ethidium cation intercalation. The more sophisticated all–atom models suggest that intercalation in pyr-(3'-5')-pur sequences is preferred over pur-(3'-5')-pyr sequences in the deoxydinucleoside complexes. The all–atom models also predict that the CG homopolymer forms a strong complex with ethidium, in agreement with the experimental work of Kastrup *et al.*[46] Two local minimum energy

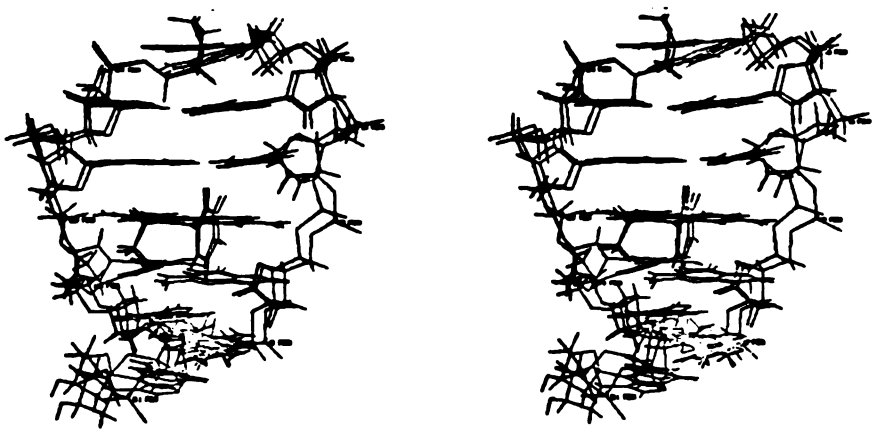


Figure 2.5: $(CGCGCG)_2$**:Ethidium complex in mixed and uniform sugar pucker models (mixed sugar pucker model is labeled at phosphate atoms.)**

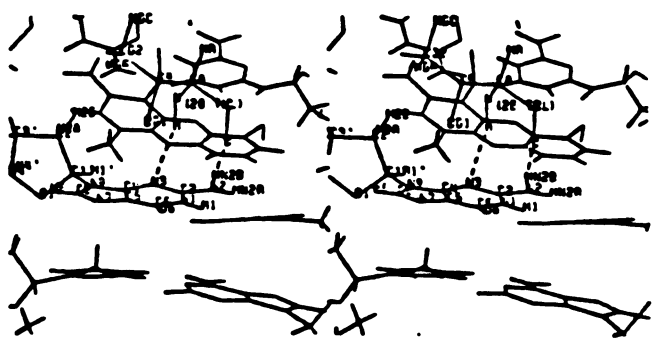


Figure 2.6: AMD–DNA hydrogen bond network. The cyclic pentapeptide residues have been clipped away to reveal the threonine–guanine interactions in detail.

conformations are observed in the deoxydinucleoside models. One, conformation B, corresponds to the conformation represented by the Sobell crystal structure model.[30,31] The other conformation results when the phosphodiester backbone distorts to form a stronger hydrogen bond between a phosphate oxygen and an amino hydrogen of the ethidium chromophore. It is quite possible that this altered conformation is the result of *in vacuo* calculations. The inclusion of explicit solvent and counterions in the calculation might decrease the probability that this conformation is observed in minimizations, as the phosphate oxygens would interact strongly with the environment. It should be emphasized that the computed ethidium sequence preferences in the deoxydinucleoside complexes are independent of conformation (A or B).

The relative sequence preference in the CG deoxyhexanucleoside complexes follows the pattern established in the CG deoxydinucleosides. The relative sequence preference in the AT deoxyhexanucleosides differs from the deoxydinucleoside results in that the pyr-(3'-5')-pur and pur-(3'-5')-pyr binding affinities are isoenergetic. These results are still consistent with the experimental observations of Bresloff and Crothers,[46] which suggest only that the heteropolymer intercalation site is more favorable than the homopolymer site in DNA fragments. Based on the ethidium binding constants to AT heteropolymer and homopolymer fragments,[46] intercalation in the heteropolymer sequence should be favored by about 1.95 kilocalories per mole. The computed value of 1.2 kilocalories per mole is in reasonable agreement with the experimental value. Analogous experimental values are not available for GC heteropolymer and homopolymer fragments, but binding to IC heteropolymer is favored by 1.89 kilocalories per mole over the IC

homopolymer (1.57 kilocalories per mole for a second weaker binding site of IC heteropolymer).[46] The calculations suggest that binding to GC heteropolymer should be favored by about 1.2 kilocalories per mole over GC homopolymer, which seems reasonable given the experimental values for the IC heteropolymer and homopolymer fragments.

A second interesting result of these calculations is the ability of CG deoxyhexanucleoside intercalation complexes to form energetically feasible, uniform C2'-endo sugar pucker geometries (with the exception of the CG homopolymer in the united–atom model calculations). That uniform C2'-endo sugar pucker geometry intercalation sites prove energetically feasible should not be a surprise in view of previous model–building studies by Alden and Arnott,[54] which suggested such intercalation site geometries were possible in deoxyhexanucleoside fragments.

Calculations using the united–atom models did not reproduce experimental results well whereas the explicit all–atom model results were in good agreement with experiment and previous theoretical work. However, the refined structures from united–atom and all–atom models have quite similar conformations as has been noted. Detailed analyses revealed that the differences in the two models arise from subtle differences in the potential function parameters for united– versus all–atom models. It seems that the more sophisticated all–atom model may be necessary to compute reliably relative binding energies when the ligand interacts with the biomacromolecule via nonspecific interactions. Ligands that interact with biomacromolecules via specific group–group interactions such as hydrogen bond and/or charge–charge interactions can probably be modeled with fewer problems than those encountered with ethidium because strong

specific interactions tend to overshadow weaker nonspecific interactions such as van der Waals forces. This belief is based on results obtained in calculations on deoxyhexanucleoside pentaphosphate:actinomycin–D complexes discussed below.

Actinomycin D complexes

Results for united–atom and all–atom calculations on AMD deoxyhexanucleoside pentaphosphate complexes are reported in Table 2.8 and indicate that, while the two models are not identical, the qualitative features are similar. Therefore, the following discussion will focus on the all–atom model results, keeping in mind that the united–atom model results lead to the same general conclusions.

For the AMD complexes, the drug-helix interaction energy term is decomposed into 'drug-site' and 'drug-other' components. The 'drug-site' term represents the intermolecular interaction energy between AMD and the deoxydinucleoside monophosphate fragment which forms the intercalation site. The 'drug-other' term represents the intermolecular interaction energy between AMD and the deoxydinucleoside diphosphate fragments on each side of the intercalation site. All other terms are analogous to those in Tables 2.5 and 2.7 for ethidium complexes.

Focusing attention first on the $d(GCXYGC)_2$:AMD complexes, Table 2.8 indicates that AMD strongly prefers to bind to sequences with $XY=GC$. The origin of this preference appears to depend primarily on 'drug-site' interaction energies, with helix destabilization energies modulating the overall binding energy preferences in some cases.

TABLE 2.8*

Complex	E_T	ΔE	Drug–site	Drug–other	Destab. E
$d(ATGCAT)_{2_{mix}}$	-572.3 (-720.8)	-76.2 (-78.1)	-102.9 (-103.0)	-38.9 (-36.4)	+48.8 (+45.5)
$d(ATGCAT)_{2_{uni}}$	-582.0 (-725.6)	-85.9 (-82.9)	-109.2 (-108.3)	-39.1 (-37.0)	+49.1 (+49.4)
$d(GCGCGC)^{**}_{2_{mix}}$	-660.1 (-649.5)	-74.1 (-80.4)	-101.9 (-105.1)	-35.5 (-33.1)	+47.3 (+42.9)
$d(GCGCGC)^{**}_{2_{uni}}$	-659.3 (-649.5)	-73.3 (-80.4)	-109.6 (-105.1)	-35.3 (-33.1)	+57.4 (+42.9)
$d(GCCGGC)_{2_{mix}}$	-650.7 (-638.6)	-62.9 (-75.5)	-94.8 (-92.3)	-36.0 (-33.9)	+52.9 (+42.3)
$d(GCATGC)_{2_{mix}}$	-602.7 (-675.3)	-61.4 (-72.2)	-89.3 (-87.9)	-36.0 (-33.6)	+47.7 (+38.6)
$d(GCTAGC)_{2_{mix}}$	-604.0 (-671.5)	-64.5 (-69.7)	-87.0 (-83.7)	-35.9 (-33.9)	+43.3 (+37.8)

Table 2.8: Component energies of refined deoxyhexanucleoside pentaphosphate intercalation complexes. Subscripts mix and uni refer to mixed and uniform sugar pucker intercalation models, respectively. Values in parentheses are for united–atom model calculations.
* E_T : Total energy of the complex ΔE : Binding energy
Drug–site : The intermolecular interaction energy between AMD and the deoxydinucleoside monophosphate portion of the helix which comprises the intercalation site.
Drug–other : The intermolecular interaction energy between AMD and the deoxydinucleoside diphosphate fragments on each side of the intercalation site.
Destab. E : The helix destabilization energy
** The united–atom model for $(GCGCGC)_2$ has a mixed C3'-endo(3'-5')- C2'-endo sugar pucker in one strand and a uniform C2'-endo sugar pucker in the other strand at the intercalation site.

Visual analysis of these complexes with computer graphics rationalizes the numerical results. A G-(3'-5')-C sequence at the intercalation site allows the formation of good hydrogen bonds between the N3 ring nitrogen and a 2-amino proton of guanine with a threonine amide hydrogen and carbonyl oxygen, respectively. These hydrogen bonds are similar to those observed in

the Sobell[15] and Berman[16] crystal structures and proposed in the earlier Sobell model[59] (see Figure 2.6 for diagram detailing these interactions). In the refined model structures, threonine–guanine hydrogen bonds are observed for each cyclic pentapeptide interaction with a strand of the nucleic acid helix. In contrast, a revised model by Sobell[60] predicts only one set of threonine–guanine hydrogen bonds. All other intercalation site base sequence possibilities either diminish or abolish this network of hydrogen bonds between drug and nucleic acid. For example, an A-(3'-5')-T sequence at the intercalation site preserves the purine N3- threonine NH hydrogen bond but adenine possesses no amino group in the 2 position so the purine 2-amino-threonine carbonyl hydrogen bond is lost. A T-(3'-5')-A sequence allows no hydrogen bonds to form between AMD and the edges of the bases at the intercalation site. A C-(3'-5')-G sequence at the intercalation site enables the formation of an altered network of hydrogen bonds. The threonine NH and carbonyl groups hydrogen bond to the cytosine O2 oxygen and the 2-amino group of the base-paired guanine, respectively. However, these hydrogen bonds are not so strong as those formed in the GC sequence. The hydrogen bond distances are notably longer and the hydrogen bond geometries are more distorted. There is probably also additional strain introduced in the complex to allow this alternate hydrogen bond network to form since the geometry is not ideal for such a network.

The numerical results (i.e., the 'drug-other' interaction energy) suggest that AMD interactions with those portions of the helix not comprising the intercalation site are constant throughout the series $d(GCXYGC)_2$:AMD. Computer graphics comparison of these various complexes reveals that all sequences adopt quite similar helical conformations except in the immediate

vicinity of the intercalation site and that AMD displays a consistent conformation in all complexes. Given these observations, it is not surprising that the 'drug-other' term varies little in the $d(GCXYGC)_2$:AMD series.

The $d(GCGCGC)_2$:AMD complex exhibits a much stronger 'drug-site' interaction energy for the uniform sugar pucker model versus the mixed sugar pucker model. Table 2.9 reports backbone torsion angles at the intercalation site for the AMD complexes. This stronger interaction appears to arise from consistently stronger hydrogen bonds (guanine-threonine and chromophore amino-DNA phosphate) in this conformation. Each hydrogen bond in the uniform sugar pucker model is somewhat shorter with a better hydrogen bond geometry than its counterpart in the mixed sugar pucker model. However, the helix destabilization energy strongly favors the formation of a mixed sugar pucker intercalation geometry over the uniform sugar pucker site geometry. Thus, the two conformations have similar binding energies.

The $d(ATGCAT)_2$:AMD complex also exhibits a much stronger 'drug-site' interaction energy for the uniform sugar pucker model versus the mixed sugar pucker model, for the same reasons as the $d(GCGCGC)_2$:AMD complex. However, in this complex there is little difference in the helix destabilization energy of the uniform and mixed sugar pucker models, so AMD binding to the uniform sugar pucker model is preferred by nearly 10 kilocalories per mole over the mixed sugar pucker model. From Table 2.8, it may be noted that 'drug-site' interaction energy terms are nearly identical for analogous sugar pucker models in $d(ATGCAT)_2$:AMD and $d(GCGCGC)_2$:AMD complexes but the 'drug-other' interaction energy term is 3.4-3.8 kilocalories per mole more favorable for the $d(ATGCAT)_2$:AMD complexes. Careful study of these

TABLE 2.9

Angle	Strand 1			Strand 2		
	B–DNA	A	B	B–DNA	A	B
χ_3	65°	51°	100°	65°	44°	98°
δ_3	140°	90°	139°	140°	82°	160°
ε	185°	188°	195°	185°	185°	194°
ζ	240°	239°	179°	240°	266°	265°
α	289°	195°	179°	289°	198°	174°
β	178°	201°	205°	178°	202°	177°
γ	61°	176°	171°	62°	170°	176°
χ_4	66°	74°	76°	66°	70°	54°
δ_4	148°	155°	158°	148°	156°	116°

Table 2.9: Refined backbone torsion angles around the intercalation site in $(GCGCGC)_2$:AMD complexes with A) mixed C3'–endo(3'–5')–C2'–endo sugar puckers and B) uniform C2'–endo sugar puckers at the intercalation site for the all–atom models. Torsion angles for a B–DNA helix duplex[55] are included for reference. Angle notation is that given in Dickerson *et al.*[57]

complexes using computer graphics indicates that the N-methyl groups of N-methyl valine residues form unfavorable steric contacts with the 2-amino group of the terminal guanine residues in $d(GCGCGC)_2$:AMD complexes. This repulsive steric interaction is not present in the $d(ATGCAT)_2$:AMD complexes. The absence of this steric repulsion seems to account for the improved drug-helix interaction in the $d(ATGCAT)_2$:AMD complexes.

As mentioned above, the AMD conformation does not vary in different complex models. This AMD conformation displays no significant deviations from the crystal structure conformation.[15] For example, the intra–annular hydrogen bonds between valine residues in each cyclic pentapeptide are maintained in the drug–DNA complexes.

The energy-refined $d(ATGCAT)_2$:AMD structures have been compared with results from 2-D NMR NOE experiments.[23] The NOE distances were

assigned by using a simple formula to relate an integrated proton-proton peak intensity to the peak intensity of a well defined reference interaction:

$$\frac{r_{ij}}{r_{kl}} = \left[\frac{\eta_{kl}}{\eta_{ij}}\right]^{\frac{1}{6}} \qquad (2.3)$$

where r_{kl} is the distance of the known reference interaction and η_{ij} and η_{kl} are the integrated peak intensities for the unknown and reference distances, respectively. The cytosine base H5-H6 peak, which corresponds to an interatomic distance of 2.46Å, was chosen as the reference interaction. The experimental distances are typically accurate to within ± 0.5Å. Appendix 2 contains a comparison of representative NMR and molecular mechanics model distances.

The overall agreement between molecular mechanics model structures and NMR results is quite respectable. Nearly three hundred distances from NOE experiments were compared with the corresponding model structure distances. The intramolecular AMD distances from the molecular mechanical model are in good agreement with the NOE distances and most AMD–DNA intermolecular distances also show good agreement between model structures and experimental results. Less than ten percent of the compared distances (twenty six total distances) display large discrepancies (≥1.5Å) between experimental and computed results. These twenty six problem comparisons involve mainly internal helix distances such as base proton-deoxyribose proton and base proton-base proton interactions encompassing terminal residues. Only two of the serious distance discrepancies involve AMD–helix distances (the H1' proton on the second thymidine residue in each strand with the N-methyl valine methyl protons in AMD) and only one serious discrepancy involves an AMD internal distance (a proline alpha

proton with a D-valine beta proton).

Discussion

The results outlined here are in good agreement with experiment and model-building studies. The structures clearly display the hydrogen bond network of the earlier Sobell model.[69] The computed gas phase energy results suggest that AMD should strongly prefer to intercalate on the 3' side of guanine residues as has been shown by experimental work and rationalizes this preference on the basis of specific strong hydrogen bonds formed when AMD intercalates at this position. The model structures also appear to agree well with the conformations predicted by 2-D NMR experiments. The all–atom and united–atom models give comparable results for AMD interaction with nucleic acids. As discussed above, specific interactions between ligand and macromolecule such as hydrogen bonds and charge interactions should generally overwhelm much weaker nonspecific van der Waals interactions involving CH, CH_2, and CH_3 groups, where the two models differ. Since hydrogen bonding and charge–charge interactions are handled identically by both all–atom and united–atom models, it is not surprising that the AMD results are comparable for both models.

Although the calculations are gas phase computations and overestimate the magnitude of the ΔE values, some results may be cautiously extrapolated to solution phase data. If differential desolvation effects for complex formation in varying base sequences are not substantial, computed relative binding energies, $\Delta\Delta E$ ($\Delta\Delta E = \Delta E_{complex\ A} - \Delta E_{complex\ B}$) can be related to relative free energies $\Delta\Delta G$ for solution phase systems ($\Delta\Delta G \sim \Delta\Delta E$). For example, it seems reasonable to assume that the desolvation energy for the $d(GCGCGC)_2$ and $d(GCCGGC)_2$ helices might not be radically different. Therefore, one can

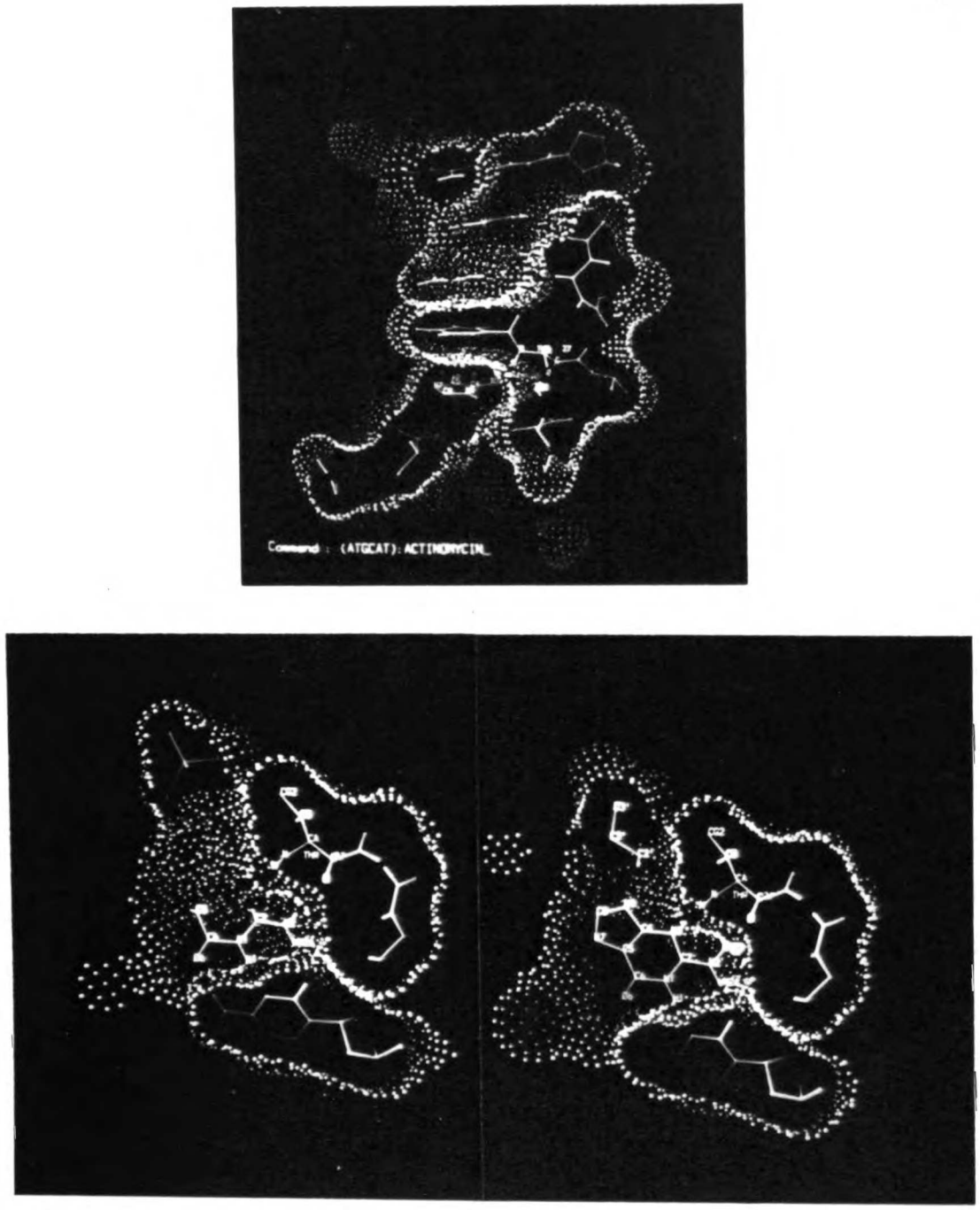


Figure 2.7: Top: Side view of $(ATGCAT)_2$:AMD complex.

Bottom: Detail of threonine–guanine interactions. DNA helix is blue and AMD is yellow. The threonine residues are green and the guanine residues are blue.

calculate with some confidence that AMD would indeed prefer to bind to the $d(GCGCGC)_2$ sequence versus the $d(GCCGGC)_2$ sequence in solution. Experimental results for AMD complex formation with deoxydinucleotides support this prediction, as Krugh found that AMD prefers to interact with (*dpGpC*) over (*dpCpG*) by approximately 5.5 kilocalories per mole.[20] This relative free energy difference cannot be directly related to the computed relative binding energies, however. The Krugh experiments reflect the interaction of AMD with two (*dpGpC*) deoxydinucleotides to form a minihelix intercalation complex, whereas AMD interacts with only one (*dpCpG*) deoxydinucleotide and no minihelix intercalation complex is formed. Nonetheless, the Krugh data suggest a strong preference for AMD to interact with the (*dpGpC*) sequence relative to the (*dpCpG*) sequence. The computed binding preference of AMD for $d(GCGCGC)_2$ over $d(GCCGGC)_2$ of 11.0 kilocalories per mole is probably somewhat excessive (this value corresponds to a ratio of binding constants for $d(GCGCGC)_2$ versus $d(GCCGGC)_2$ of 1×10^8). If this is an overestimation of the relative stabilities, it most likely is due to lack of solvent molecules in the calculations. In aqueous solution, both AMD and the DNA hexamer fragment would have water molecules forming hydrogen bonds with threonine and guanine residues. Upon complexation, the waters would be liberated and the threonine–guanine hydrogen bond network would form. Little is gained energetically from the formation of specific hydrogen bonds between threonine and guanine residues as water molecules effectively fulfill the role of hydrogen bond donors and acceptors in the uncomplexed state. However, much is lost if these interactions are not present. (Recall that threonine–guanine hydrogen bonds are the basis for AMD sequence selectivity.) When the AMD-DNA complex is formed, the large numbers of water molecules liberated from the surface of AMD and the minor

groove of the DNA fragment make complex formation highly favored on an entropic basis. Experimental studies have indeed shown that AMD complexation with DNA is an entropically driven process with little enthalpic contribution ($\Delta H \sim 0$, ΔS = 31eu/mole).[61] In gas phase calculations, the enthalpic contribution for threonine–guanine hydrogen bond formation is overestimated as the uncomplexed reference state has no water molecules to fulfill hydrogen bonding needs. Therefore, the relative preference of AMD for $d(GCGCGC)_2$ vs. $d(GCCGGC)_2$ is probably overestimated.

There is no reason to assume *a priori* that the desolvation energy for the $d(GCGCGC)_2$ and $d(ATGCAT)_2$ helices would be similar. Thus, no statements can be made about the relative preference of AMD for these two sequences in solution, based on gas phase calculations. To address this preference, calculations must be performed including solvent molecules and counterions explicitly.

Even though qualitative relative binding preferences cannot be addressed due to differential desolvation effects in most complexes, statements can be made about the intrinsic AMD-helix interactions observed in the calculations. For example, it has been proposed in the past that the pentapeptide side chains might form specific interactions with the nucleic acid helix. The model structures reveal no such interactions except the threonine-guanine hydrogen bonds and the unfavorable steric interaction between N–methylvaline and 5'-terminal guanine residues in the $d(GCXYGC)_2$:AMD complexes. This is not to imply that appropriately modified cyclic pentapeptides could not be found which do exhibit specific interactions. An extensive computer graphics study of these complexes suggests that it might be possible to make conservative modifications in the

amino acid substituents and maintain strong binding while introducing some degree of selectivity for base pairs adjacent to the intercalation site via specific pentapeptide–nucleic acid interactions (i.e., develop AMD analogs selective for given tetramer $d(-XGCY-)_2$ or hexamer $d(-XYGCX'Y'-)_2$ base sequences).

Any modifications made in an effort to introduce some degree of base sequence selectivity should probably be conservative changes. Even though no specific pentapeptide-nucleic acid interactions are observed (except those involving threonine), it is well documented that acyclic pentapeptides or AMD derivatives without the full compliment of five amino acids do not bind well to DNA and have little biological activity.[62,63] The importance of these cyclic pentapeptides may be to serve as "flaps" which effectively shield the guanine-threonine hydrogen bond network from solvent exposure. Solvent-accessible molecular surfaces were computed for several complex structures using an algorithm developed by Connolly.[64] These surfaces demonstrate that the cyclic pentapeptide "flaps" do indeed prohibit facile access of water to the hydrogen bond network. Using a binding model first proposed by Müller and Crothers to explain the complex kinetic behavior of AMD-nucleic acid interactions,[17] the following steps can be postulated. Initial intercalation of the chromophore on the 3' side of a guanine residue in a breathing DNA molecule occurs, followed by small conformational changes in the pentapeptide rings which allow them to fit snugly in the minor groove, forming strong but nonspecific hydrophobic interactions with the DNA helix and protecting the guanine-threonine hydrogen bonds from disruption by solvent. This series of steps may account for the tight binding and slow dissociation of AMD-nucleic acid complexes. However, it does not seem to

explain the complicated kinetic behavior observed in AMD complexation with nucleic acids. The proposal of Müller and Crothers suggests that large conformational changes occur in the cyclic pentapeptides upon binding to DNA and that this behavior rationalizes the unusual kinetics. The molecular mechanics models suggest that little change occurs in pentapeptide conformations when AMD complexes with DNA, in accord with recent experimental results.[23] The root mean square deviation in pentapeptide conformation between complexed and free AMD molecules is ~ 0.2Å/atom for nonhydrogen atoms (the RMS deviation is ~ 0.3Å/atom if hydrogen atoms are included in the statistics) for the molecular mechanics models.

The agreement between experimental NMR NOE distances and distances computed from the energy refined model structures is encouragingly good and lends support to the validity of molecular mechanics as a tool for examination of structural aspects of biomacromolecular interactions. The best agreement relates to AMD internal conformation and DNA-AMD contact distances. These portions of the structure are relatively rigid due to the tight complex formed between AMD and the nucleic acid helix. As mentioned previously, only twenty six distances out of nearly three hundred compared between NOE data and molecular mechanics models differ by 1.5Å or more. In all these cases, the distance predicted by molecular mechanics is longer than that determined from NOE experiments. Six of these distances involve base and/or deoxyribose protons of terminal nucleoside residues. Detailed computer graphics analysis of the molecular mechanics structures suggests that the distances from NOE data might be shorter because the complex in solution may have exhibited some fraying at the terminal residues. If these terminal nucleoside residues did occasionally break their Watson–Crick base pair conformations, the distances in question could have been shortened due

to closer base-base interactions resulting from the enhanced conformational flexibility. Thirteen of the distances exhibiting poor agreement between experiment and molecular mechanical models involve deoxyribose proton interactions with other deoxyribose and/or base protons of the same residue or AMD protons. Sugar repuckering is one possible explanation for the shorter experimental distances observed for these interactions. However, several of this group of thirteen problem distances probably would not exhibit good agreement with the computational models even if considerable sugar repuckering were taken into account. At present, molecular dynamics calculations on the $d(ATGCAT)_2$:AMD complex are being used to evaluate the effects of molecular motions on the experiment-model agreement. The preliminary results after only ~28 picoseconds of simulation suggest that dynamic motion of the complex is likely to account for most discrepancies less than 1.5Å and may greatly improve experiment-computational model agreement for the twenty six problem distances. However, there are still some discrepancies that are not easily explained by any type of realistic molecular motions. Several of this group of twenty six distances from NOE data are impossibly short due to constraints imposed by covalent bond topology. At least four distances which involve interactions with freely rotating methyl groups seem unrealistically short based on computer modeling results and cannot be rationalized by base sliding, sugar repuckering, or other dynamic motion. Considering all these facts, an additional source of disagreement could relate to the simple expression used to compute interatomic distances from the NMR peak intensities. Some of these experimental distances may be unrealistically short due to complicated and more efficient relaxation mechanisms than have been assumed in the computation of distances from NOE data.

2.4 Conclusions

The good agreement with experimental sequence selectivities for ethidium and AMD and 2–D NMR results for the $d(ATGCAT)_2$:AMD complex illustrate the capabilities of molecular mechanics computer modeling studies for ligand–biomacromolecule complexes. These models have been used to suggest hypotheses for several experimental observations (e.g., the origin of ethidium sequence selectivity, the role of intact cyclic pentapeptides in AMD–DNA complexes). The most useful current role of computer modeling techniques is perhaps their application in analysis of experimental results and evaluation of possible explanations for these experimental results. Computer modeling techniques are not yet sufficiently refined to allow quantitative predictions of relative binding affinities for several ligands at a receptor site (or one ligand at closely related receptor sites). One major deficiency for these potential energy calculations at present is the inability to properly include solvent (and counterions) in the models. Explicit inclusion of the solvent environment in these calculations will allow direct comparison of relative binding energies for $d(ATGCAT)_2$:AMD and $d(GCGCGC)_2$:AMD complexes, for example.

One further point is illustrated by the examples in this chapter. *Nonspecific* ligand–biomacromolecule interactions (e.g., van der Waals interactions) may be intrinsically more difficult to model with confidence than specific ligand–biomacromolecule interactions (e.g., hydrogen bonds, charge–charge interactions). Results for ethidium–nucleic acid complexes (predominantly nonspecific interactions) were model–dependent (united–atom vs all–atom models gave somewhat different results), while results for AMD–nucleic acid interactions (predominantly specific

interactions) were model–independent. This further emphasizes the current inability of potential function models to predict free energies of binding for ligand–biomacromolecule interactions accurately.

CHAPTER 3

Ligand Interactions at Binding Sites of Unknown Structure

3.1 Background

As stated previously, most drug–receptor interactions of interest to medicinal chemists and pharmacologists involve undefined receptors. Although the composition and three–dimensional structure for many important receptor sites are unknown, extensive data may be available on pharmacological properties and binding affinities for ligands that interact with the receptors. Two examples of undefined receptors with extensively studied ligands are CNS target receptors for psychoactive compounds and opioid receptors.

Many psychoactive compounds, such as neuroleptic (antipsychotic) agents, form complexes with undefined CNS receptors. For example, clozapine (Figure 3.1) is an unusual neuroleptic agent with potent antipsychotic activity[65] but marked anticholinergic properties and little propensity to induce extrapyramidal side effects such as parkinsonian syndrome, akathisia, and tardive dyskinesia.[24] (Most commonly used neuroleptic agents exhibit extrapyramidal side effects as a major undesired property.) Unfortunately, clozapine produced agranulocytosis as a serious toxic side effect in clinical use.[66] Thus, there has been considerable interest in the development of compounds that retain clozapine's desirable pharmacological properties without the hematological toxicity. Several clozapine analogs have been tested; however, one analog, HUF–2046,

possesses a quite different pharmacological profile (e.g., traditional neuroleptic pharmacological properties including serious extrapyramidal side effects) and the dichloroclozapine analog displays pharmacological properties common to both clozapine and HUF–2046. These compounds differ from clozapine only in degree and/or position of chlorine substitution. Therefore, computer modeling techniques were used in an attempt to rationalize the variable pharmacological behavior in this set of structurally similar compounds.

In a related study, computer modeling techniques were employed to explain differential binding behavior for a series of opioid ligands (Figure 3.2a-e) at the μ receptor site and the λ site, a recently discovered opioid binding site.[25] The λ site appears to be extremely selective for 4,5–epoxymorphinans while the μ site does not discriminate between 4,5–epoxymorphinans (e.g., oxymorphone, Figure 3.2a) and morphinans (e.g., levorphanol, Figure 3.2e). Pharmacophoric patterns were sought to explain the selective opioid binding behavior at the λ site relative to the μ receptor for these opioid ligands.

3.2 Methods

The approach used in these studies involves an integrated application of conformational analysis using molecular mechanics minimization (for those ligands with internal flexibility), computation of electrostatic potential surfaces, and interactive computer graphics analysis of three–dimensional structural features for the ligands. Conformational analysis has been used frequently to search for structural pharmacophores in drug molecules.[67-70] Electrostatic potential calculations have also been heavily utilized in the

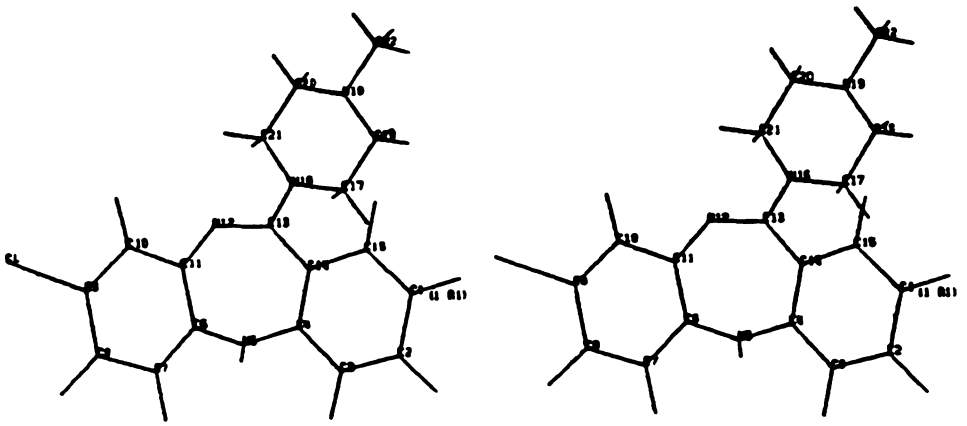


Figure 3.1: Clozapine. Clozapine analogs differ from clozapine as follows:

HUF–2046 – Chlorine substituent on C1 carbon rather than C9 carbon.

des– chloroclozapine – No chlorine substituents in the molecule.

Dichloroclozapine – Chlorine substituents at both C1 and C9 carbons.

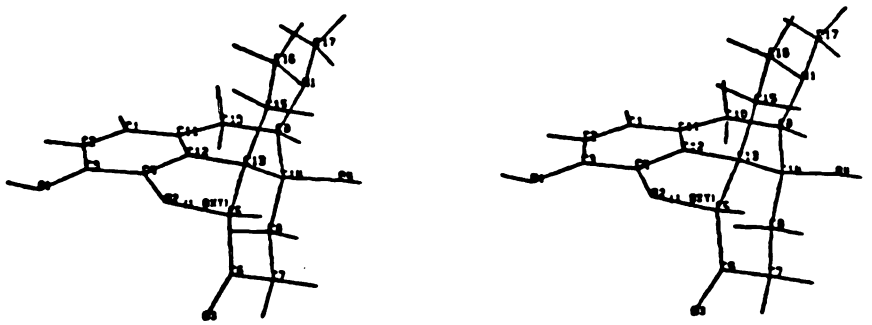


Figure 3.2a: Oxymorphone

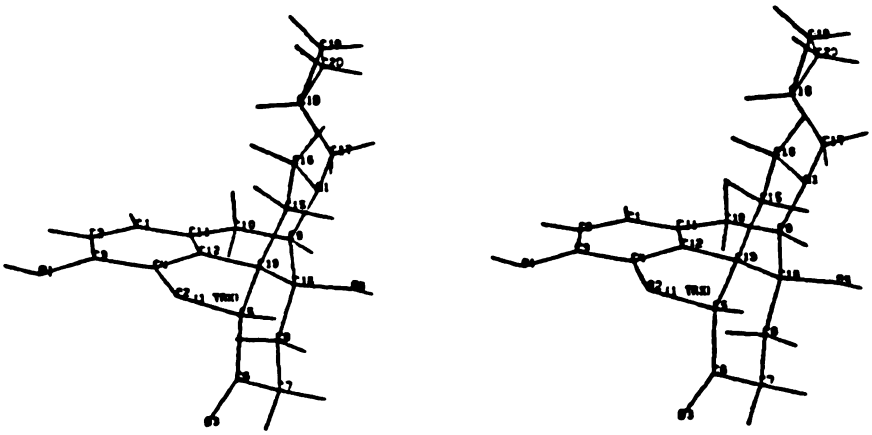


Figure 3.2b: Naltrexone

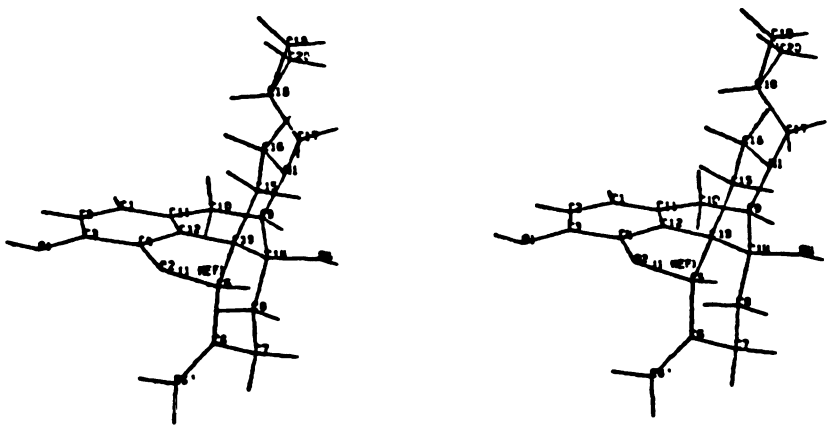


Figure 3.2c: Nalmefene

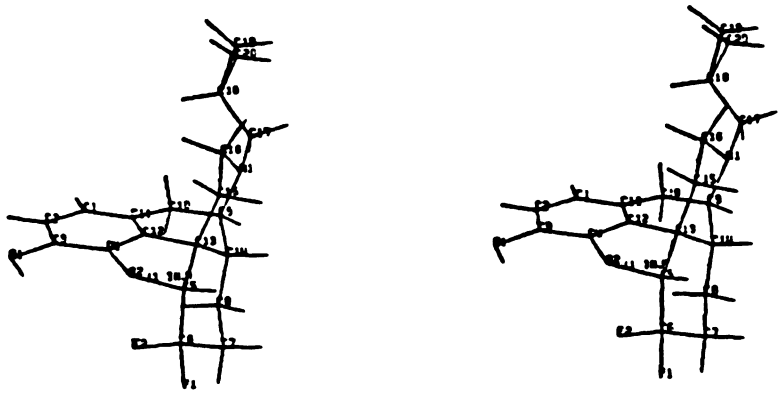


Figure 3.2d: INJ–6471

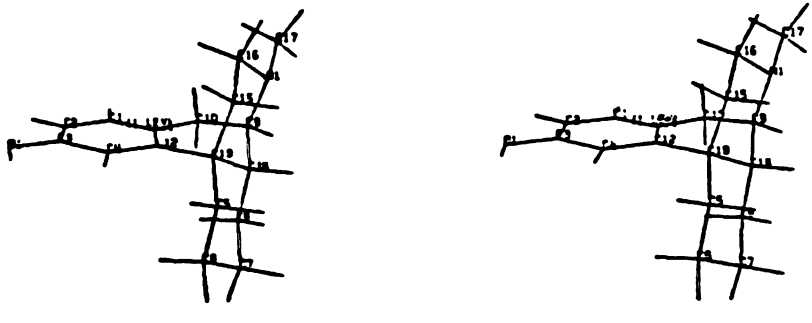


Figure 3.2e: Levorphanol

search for pharmacophoric patterns.[71,72] Both techniques have provided encouraging results, when applied judiciously.

All energy minimizations were performed using a Newton–Raphson algorithm in the molecular mechanics package AMBER.[47] The potential function parameters used in these calculations were taken directly from the program MM2 developed by Allinger.[73] The potentials used in this work differed from the standard MM2 potentials in that atomic partial charges were employed rather than bond dipoles to evaluate electrostatic interactions, out–of–plane and stretch–bend components were not included in the bond angle term, the van der Waals term employed a 6–12 function rather than an exp–6 function, and hydrogen van der Waals parameters were centered on the nuclei rather than shifted along the carbon–hydrogen bond vector as is done in the MM2 program. The expression for the potential energy is as follows:

$$E = \sum_{bonds} K_b (R - R_e)^2 + \sum_{angles} K_\theta (\Theta - \Theta_e)^2 + \sum_{dihedrals} \frac{K_d}{2} \left[1 + \cos(n\Phi - \gamma)\right] + \sum_{non-bonded} \left[\frac{\sqrt{B_i B_j}}{r_{ij}^{12}} - \frac{\sqrt{A_i A_j}}{r_{ij}^{6}} + \frac{q_i q_j}{\varepsilon_{ij} r_{ij}} \right] \quad (3.1)$$

Most details of the energy minimization calculations were identical to the procedures outlined in chapter 2. Non–bonded interactions were evaluated for all atom pairs not involved in bond or bond angle terms. The partial charges for the various molecules studied were determined from CNDO/2 calculations and a dielectric constant ε equal to the magnitude of the interatomic distance r_{ij} was used.

All structural models for clozapine and opioid ligands were based on coordinates from X–ray crystallographic studies when available.[74-79] The opioid compounds with no available X–ray coordinate data were model–built

using crystal coordinates from closely related compounds for analogous sections of the molecules and standard bond lengths and angles[48] for unique sections. All opioid structures were then energy refined to remove any existing strain in the crystal structure or model–built compounds. Analogous calculations were not performed for the rigid clozapine analogs. Only the piperazinyl substituent of the clozapine analogs possesses conformational flexibility, and there was no reason to suspect that its conformation should vary between different analogs. Minimizations were considered converged when the root mean square derivative of the energy function with respect to the atomic coordinate changes was ~ 1.0×10^{-4} kcal/mole Å or less.

Three–dimensional structural features for the various molecules were analyzed using MIDAS,[52] and solvent accessible molecular surfaces were computed using the algorithm developed by Connolly.[64] Several procedures were used to calculate molecular electrostatic potential surfaces for the clozapine analogs. In the most primitive approach, Mulliken populations determined within the CNDO/2 formalism were used to assign partial charges to each atom in the molecule of interest. Next, a solvent–accessible molecular surface for the molecule was calculated. The electrostatic potential at each surface point due to the partial charges centered on the atoms of the molecule was then computed. Finally, the computed molecular electrostatic potential surfaces were displayed with MIDAS. A special coloring scheme, calculated and scaled by features within MIDAS, was used to achieve optimal color differentiation corresponding to the electrostatic potential gradient over the molecular surface. In this coloring scheme, those surface points associated with the regions of strongest negative potential were colored green. The surface points associated with the regions of

strongest positive potential were colored blue. Scaling algorithms within MIDAS then colored all points of intermediate electrostatic potential accordingly. For example, points associated with regions of approximately neutral potential were colored blue. A second approach in computing molecular electrostatic potential surfaces was quite similar to the first, differing only in that Mulliken populations were calculated at the *ab initio* level using an STO–3G basis set.

A third approach to computing molecular electrostatic potential surfaces followed a procedure used previously by Singh and Kollman.[80] First, *ab initio* wavefunctions for the molecules were calculated using an STO–3G basis set. Then, the electrostatic potential at each point of four contours of Connolly solvent–accessible surfaces (contours at 1.4Å, 1.6Å, 1.8Å, and 2.0Å beyond the molecular surface) were calculated using the STO–3G wavefunctions. Next, the quantum–mechanically calculated molecular electrostatic potential surfaces were used in a non–linear least squares fitting algorithm to obtain an analytical atom–centered point charge model which reasonably reproduced the quantum mechanical electrostatic potential over the four surface contours. The point charge models produced by this method were then used analogously to the Mulliken population charges in the two previous procedures to compute the molecular electrostatic potential surfaces for display by MIDAS.

In some cases, namely the *ab initio* computations, calculations could not be conveniently performed on the entire molecule due to the large number of basis functions. Therefore, molecules were divided into fragments (dibenzodiazepine and methyl piperazine derivatives for the clozapine analogs) with common overlap regions, calculations were done on the

fragments, and the final point charge models for the molecules were derived by piecing together the fragments with adjustment of the partial charges in the common overlap regions to conserve the molecular charge. This approach was also used to derive point charge models for isomeric clozapine analogs so as to avoid the need for a full series of *ab initio* calculations of wave functions and quantum mechanical electrostatic potentials for each isomer. The partial charge adjustment in overlap regions necessary for maintenance of molecular charge proved to be quite small (never more than ± 0.05 electrons on each atom), suggesting that this approach for development of partial charge models should not differ substantially from full *ab initio* calculations on each molecule.

Point charge models for protonated species of the clozapine analogs and all opioid ligands were taken from CNDO/2 Mulliken populations only. This simplified approach was taken because initial evaluation of electrostatic potential surfaces for the unprotonated clozapine analogs indicated that all three computational models gave the same qualitative results. These results are discussed in more detail below. Within a given point charge model, the molecular electrostatic potentials were calculated both at the solvent accessible molecular surface and 1.4Å above the molecular surface along a normal vector of each surface point. The electrostatic potential beyond the molecular surface of a ligand is of interest as it more realistically represents the electrostatic potential that a receptor might experience as the ligand approaches.

3.3 Results and Discussion

Clozapine analogs

The resultant electrostatic potential surfaces for clozapine, HUF–2046, and the dichloro–clozapine analog are presented in Figure 3.3. As can be seen from these color stereoviews, the electrostatic potential surface characteristics for clozapine and HUF–2046 are quite different. The electrostatic potential surface for the dichloroclozapine analog exhibits characteristics of both the clozapine and HUF–2046 electrostatic potential surfaces. The electrostatic potential surface for the *des*– chloroclozapine analog is qualitatively similar to the clozapine surface and has not been displayed here.

Table 3.1 lists pharmacological properties for the four analogs. The resultant molecular electrostatic potential surfaces can be correlated quite nicely with the pharmacological variance between molecules. Clozapine and the *des*– chloro compound both exhibit similar pharmacological profiles, each being devoid of extrapyramidal side effects and each exhibiting appreciable anticholinergic activity.[24] These two molecules also displayed the same qualitative molecular electrostatic potential surface characteristics. For example, the electrostatic potential surface gradient varies smoothly from positive potential around C1 to relative neutrality in the central ring to relatively negative potential around C9. In contrast, HUF-2046, which displays a quite different pharmacological profile (extrapyramidal side effects comparable to chlorpromazine and little perceivable anticholinergic activity[24]), exhibits an electrostatic potential surface gradient which ranges from slightly negative potential around C9 to strongly positive potential in the central ring to strongly negative potential around C1. Finally, the

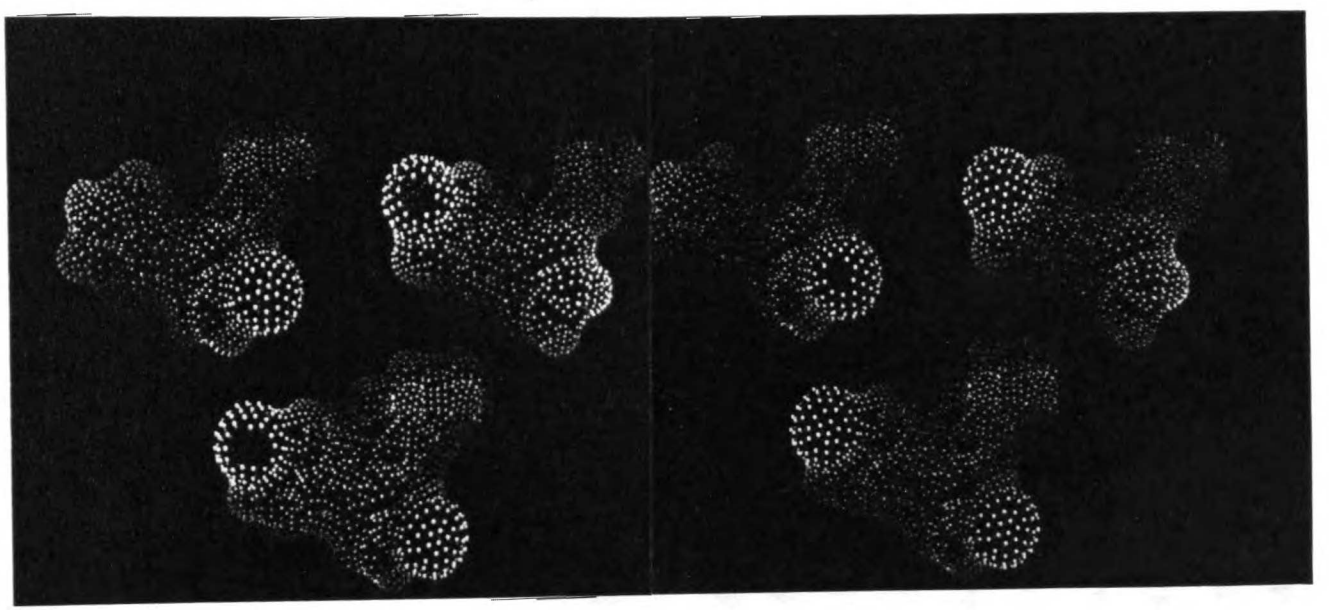


Figure 3.3: Electrostatic potential surfaces for clozapine (upper right), HUF-2046 (upper left), and dichloroclozapine (bottom).

dichloro compound has pharmacological properties common to both clozapine and HUF-2046 (extrapyramidal side effects and anticholinergic activity[24]). Likewise, characteristics of the electrostatic potential surfaces of both clozapine and HUF-2046 are visible in the potential surface of the dichloro compound. It exhibits strongly negative potential around C1 and C9 regions with strongly positive potential in the central ring region.

The results proved to be relatively insensitive to the method used for calculation of the point charges. The most sophisticated procedure, *ab initio* least squares fit for partial charge determination, led to the most marked gradient over the potential from negative to positive regions. However, since all partial charge models were colored with the scaling algorithm discussed above, the method of charge determination made little difference when the electrostatic potentials were displayed. The qualitative features of the electrostatic potential for each molecule were the same, regardless of

TABLE 3.1

Compound	EPSE	Antichol.
Clozapine	-	++
des–chloroclozapine	-	+
HUF–2046	++	-
Dichloroclozapine	+	+

Table 3.1: Relative pharmacological properties for clozapine analogs.[24] The electrostatic potential in these molecules generally ranges from -12.0 kilocalories/mole to +12 kilocalories/mole.
EPSE : Extrapyramidal side effects
Antichol. : Anticholinergic properties
+ : Observation of an effect in the listed compound. Number of + signs indicates relative potency of the compound in producing this effect.
- : Property is absent or negilible in this compound

computational technique. Of course, comparisons should only be made amongst models computed by the same procedure. For example, a CNDO/2 potential surface cannot be directly compared to an *ab initio* surface as the limits for positive and negative potential values are quite different for these two computational methods, and it is the boundary potential values which control the scaled coloring algorithm in MIDAS.

The electrostatic potential surfaces for the protonated species were less helpful in distinguishing differences between these molecules. The full positive charge was the predominant feature for all four molecules and masked most of the other features of these potential surfaces. However, the electrostatic potential for the protonated species is not necessarily pertinent to the ligand–receptor interaction process. Protonated molecules (and other cationic or anionic molecules) do not exist as isolated charges and counterions must be considered in modeling such species in solution. Calculations by Weinstein *et al*, on protonated molecules with associated counterions indicate that the electrostatic potentials for the ion pair neutralized molecule and corresponding unionized molecule are nearly identical.[71]

Finally, the electrostatic potentials computed 1.4Å above the molecular surface displayed patterns and properties similar to those potentials computed on the molecular surface. Thus, analysis of these surfaces for the four molecules led to the same observations outlined above for potentials on the molecular surface. Although the potentials on the molecular surface and 1.4Å above the molecular surface were qualitatively similar for the clozapine analogs, there is no reason to assume that this observation will be true for all molecules studied with electrostatic potential surface calculations.

Discussion

The characteristics of the computed electrostatic potential surfaces for the four clozapine analogs define two electrostatic potential surface patterns (or "pharmacophores"), which can be related to clozapine–like or nonclozapine–like (i.e., traditional) pharmacological profiles. Both clozapine and the des–chloro analog fit the first pattern and HUF–2046 fits the second pattern. The electrostatic potential surface for the dichloro analog exhibits properties of both patterns, defining it as a compound with properties of both classes (i.e., traditional and nontraditional or clozapine–like pharmacological properties). The assignments based on the electrostatic potential surface patterns correlate well with the pharmacological data for these compounds.[24] Several other clozapine analogs exist which differ from those examined in this work in that they contain heteroatom substitutions in the three–member ring system. The two computed electrostatic pharmacophoric patterns defined in this work can probably be used to characterize these compounds properly as either traditional or nontraditional neuroleptic agents. The heteroatom substitutions are unlikely to alter the overall electrostatic potential characteristics dramatically from those of the four compounds discussed here. The significance of these two electrostatic potential patterns is not clear. The two patterns may imply that two distinct receptor types or subtypes are involved. Alternately, they may suggest two different binding modes at one distinct receptor site. Questions such as these could possibly be addressed by performing competition binding experiments.

The same *qualitative* results were obtained regardless of the procedure used to develop the point charge model, which suggests that the results are

not crucially dependent on the computational methods. The only difference between the CNDO/2 model electrostatic potential surfaces and those obtained by the *ab initio* least squares fit method is the magnitude of the gradient (degree of sharpness) between the regions of greatest positive and negative potential in the molecules. Since a scaled coloring scheme has been used to display the molecular surfaces, the same picture is observed whether CNDO/2 Mulliken population charges or *ab initio* least squares fit charges are used. The protonated species were somewhat less informative as the formal positive charge tended to overwhelm most other aspects of the electrostatic potential surface. However, as pointed out above, the fully charged, unneutralized species is probably not a realistic model for a molecule in a biological system.

Similar qualitative patterns were also obtained when the electrostatic potential was calculated on the molecular surface or 1.4Å above the molecular surface. As mentioned previously, the potential above the molecular surface represents the electrostatic potential the receptor site would experience as the ligand molecule approached and thus should play a key role in recognition and spatial orientation of the ligand by the receptor site. The potential on the molecular surface represents the electrostatic potential the receptor site experiences after a tightly bound complex has been formed. The degree of complementarity between the electrostatic potentials of the ligand and receptor site determine to a large extent the strength of the ligand–receptor complex. The results from this study suggest that the approach may be useful for correlation of electronic/electrostatic characteristics with pharmacological properties, especially when applied to a series of conformationally rigid molecules such as the clozapine analogs. The lack of conformational freedom in a molecule greatly simplifies the search

for tentative pharmacophores.

Opioid ligands

Table 3.2 lists a number of opioid ligands, along with their observed K_D values at the λ site and K_I (ED_{50}) values at the μ site. Several general qualitative assessments can be made simply by observing substitution patterns that radically alter λ site binding affinity. For example, it appears that modification of the phenolic hydroxyl group to form an ether (morphine → codeine) destroys all affinity for the λ site. This suggests the phenolic OH group may bind in a region of extreme steric sensitivity (i.e. a region which is unable to accommodate groups bulkier than –OH). It is also possible that the phenolic –OH group is a hydrogen bond donor at the binding site; the $-OCH_3$ derivative (codeine) is unable to function as a hydrogen bond donor. Acetylation of the hydroxyl group at the C6 position in morphine also abolishes affinity for the λ site. This result suggests the λ site may have an additional region of steric intolerance.

Closer inspection of Table 3.2 indicates that 4,5–epoxymorphinans (e.g., oxymorphone) tend to bind much more tightly at the λ site than derivatives lacking the epoxy bridge (e.g., levorphanol). This trend is not observed for the μ receptor. Energy minimization calculations were performed to examine the conformational effects produced by either the presence or absence of the epoxy bridge. Although some differences exist in conformation between the two families, the overall deviations are not striking. Computer graphics analysis of solvent accessible surfaces for the various molecules indicates that the gross steric consequences of the differing conformations are small. One notable difference that was initially intriguing concerned the orientation of potential hydrogen bond acceptors in

TABLE 3.2

Compound	λ	μ
Oxymorphone	2.0	1.0[81]
Naltrexone	8.0	0.5[82]
Morphine	4.4×10^1	3.5[82]
INJ 6471-3	2.2×10^1	NA
Nalmefene	1.0×10^2	NA
S-20682	3.3×10^3	NA
Levorphanol	$>1.0 \times 10^4$	0.7[82]
Levallorphan	$\sim 1.0 \times 10^4$	0.7[82]
Codeine	$>1.0 \times 10^4$	2.0×10^4
6—acetylmorphine	$>1.0 \times 10^4$	NA

Table 3.2: K_D values (nM) for some ligands at the λ binding site and K_I (ED_{50} for reference 81 entries) values (nM) for the ligands at the μ receptor. K_I and ED_{50} values may differ from the corresponding K_D values for these ligands at the μ receptor by a factor of 2-3.
NA : Not available

the C6 position. When 4,5—epoxymorphinans and morphinans were displayed simultaneously, with nitrogen atoms and phenyl rings precisely superimposed for all molecules, different spatial orientations of the oxo— or hydroxy groups in the C6 position were observed for 4,5—epoxymorphinans versus morphinans. The spatial orientation of potential hydrogen bond acceptors appears to be unimportant, however. Compounds with mono— or difluoro— substitution (e.g. INJ--6471) at the C6 position display high affinity for the λ site, but fluorine substituents are not good hydrogen bond

acceptors.

Since no obvious conformational properties were discovered to rationalize the binding data for most of these ligands, electrostatic potential surfaces were calculated for a representative subset of compounds. The electrostatic potentials were computed as outlined in the methods section using partial charge models taken from CNDO/2 Mulliken populations. The resultant molecular electrostatic potential surfaces were displayed with MIDAS using the coloring scheme described above.

Distinctive differences in electrostatic potential characteristics of high– and low–affinity λ site ligands were clearly evident. Electrostatic potential surfaces for three ligands are shown in Figure 3.5. High–affinity ligands, such as oxymorphone, displayed a strongly negative potential from the phenolic oxygen through the epoxy ether bridge to the C6 substituent (carbonyl oxygen in oxymorphone). Ligands with weaker λ site binding affinity (e.g., nalmefene) displayed discontinuous or weakly negative potential across the same region, and compounds with no detectable binding affinity (e.g., levorphanol) exhibited relatively neutral or weakly positive potential in this region. Most of the low–affinity compounds lack electronegative substituents at the C6 position and/or the 4,5–epoxy bridge, so it is not surprising that they do not exhibit continuous strong negative potential across this region. Other regions of these molecules displayed no perceivable consistent differences in electrostatic potential characteristics between high– and low–affinity λ ligands.

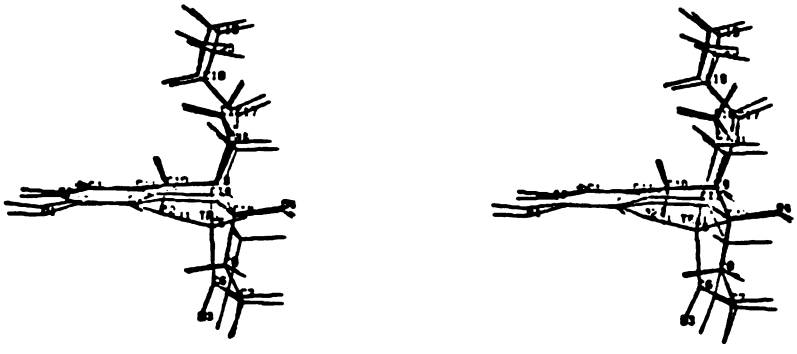


Figure 3.4: Conformational comparison for a 4,5–epoxymorphinan (naltrexone, labeled molecule) vs a morphinan (S–20682, unlabeled molecule). These two molecules are structurally identical except that the morphinan lacks the 4,5 epoxy bridge.

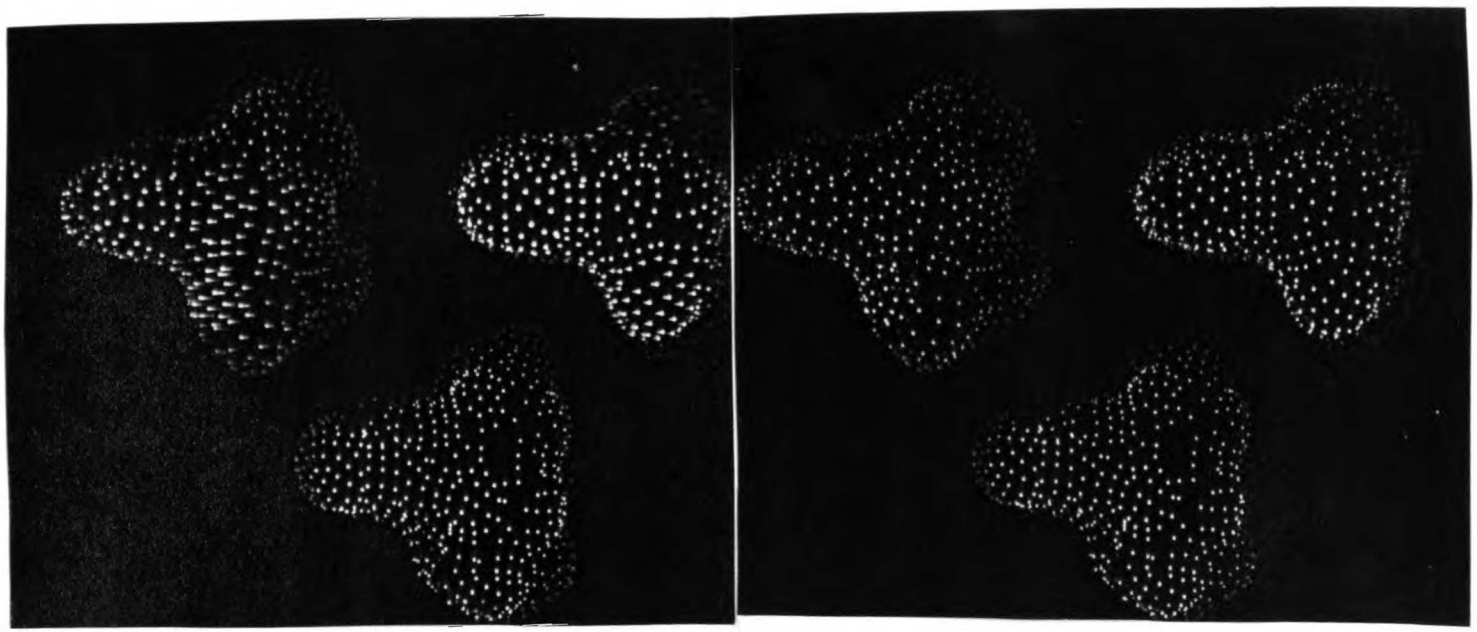


Figure 3.5: Electrostatic potential surfaces for oxymorphone (upper right), nalmefene (upper left), and levorphanol (bottom). The electrostatic potential ranges for these molecules are: oxymorphone, -16 kcal/mole → +14 kcal/mole; nalmefene, -12 kcal/mole → +14 kcal/mole; levorphanol, -5 kcal/mole → +6 kcal/mole.

Discussion

The molecular electrostatic potential surface characteristics allow for classification of these compounds as high– or low–affinity λ site ligands in good agreement with experimental binding data. The strongly negative electrostatic potential across the front edge of the surface is characteristic of all high–affinity ligands and is either considerably weakened or absent in the low–affinity compounds. The differences in electrostatic potential surface characteristics appear to be completely unrelated to μ receptor binding affinity. The models suggest that the λ site may differ from the μ receptor site in that the λ site is particularly sensitive to the gross electrostatic potential surface characteristics of a ligand relative to the μ receptor. There is also some suggestion that the λ site may impose greater steric constraints on potential ligands (e.g., intolerance to O3 substitution). However, ligands with greater substituent variation would have to be studied to address this possible difference in the two binding sites.

It may be feasible to correlate relative binding affinity for ligands at the λ site with quantitative characteristics of their electrostatic potentials. Such correlations have been made in previous work, although the method for electrostatic potential calculation may be more crucial in studies of this type.[71] Specifically, it may be necessary to use electrostatic potentials computed directly from semi–empirical or *ab initio* wavefunctions, rather than from analytical point charge models, if quantitative comparisons between molecules are made.

No conformational characteristics were discovered which allow discrimination between high– and low–affinity λ site ligands. The fact that no significant conformational differences exist between these ligands can be

used to rationalize the uniformly strong binding of these ligands at the μ site. For example, the results of the conformational calculations for these ligands are consistent with a current pharmacophore model for the μ receptor site.[69,83] This pharmacophore model considers only conformational characteristics to explain μ receptor binding for many agonists and antagonists. It suggests that the presence and relative spatial orientation of the nitrogen lone pair electrons and a phenyl ring are two key determinants of μ receptor affinity.

3.4 Conclusions

These results for clozapine analogs and opioid ligands, along with previous work by other groups, illustrate the potential utility of computer modeling techniques in the study of drug–receptor interactions at binding sites of unknown structure and composition. A key feature in this work has been the integration of conformational analysis using molecular mechanics (for flexible or semi–flexible molecules) with electrostatic potential surface calculations and interactive color computer graphics studies. The absence of any one component may greatly diminish the chances for success in a study of this nature. Such studies are more likely to yield useful information when the compounds in question are rigid or semi–rigid molecules. The procedures outlined here will probably have far less chance for success if the ligands of interest are conformationally "floppy" molecules, with many degrees of internal freedom and numerous local minimal energy conformations.

Establishment of reliable pharmacophore models for drug–receptor interactions may be of great benefit in the drug design process. Clozapine is

a prime example. This compound was originally developed as a potential anti–anxiety agent and neuroleptic activity was not considered, because clozapine does not exhibit the normal pharmacological profile expected of antipsychotic agents in *in vitro* and *in vivo* animal screening procedures. Its neuroleptic properties were not discovered until the drug was utilized in clinical trials. Future screening for neuroleptics with clozapine–like properties could be facilitated if proposed pharmacophores, such as electrostatic potential surface characteristics, prove to have reliable predictive capabilities. Similarly, reliable pharmacophores defining features necessary for ligand interaction at specific opioid receptor subtypes might aid in development of useful drugs. For example, there is currently much interest in compounds with selectivity for the κ opioid receptor, as these agents might produce analgesia with far less dependence liability than most opioid analgesic agents now available.[84]

CHAPTER 4

Extension of Computer Models to Include Environment

4.1 Background

Many computer simulation models for ligand–biomacromolecule complexes are characterized by several inadequacies. Inability realistically to incorporate environmental components, such as solvent and counterions, has been perhaps the most serious deficiency. Numerous potential energy functions for water have been introduced over the past few years; examples include some based on quantum mechanical calculations,[85-88] others parameterized on a strictly empirical basis.[89-92] With a few exceptions,[93-95] all these potentials are pair-wise additive. They attempt to describe the behavior and properties of bulk phase (solid and liquid) water as a sum of pair interactions. This is a simplification adopted primarily to facilitate computational efficiency in large calculations such as Monte Carlo or molecular dynamics simulations of pure water and water–solute systems. It is widely recognized that many body or non–pairwise additive interactions make an appreciable contribution to the potential energy of condensed phase water systems.[94,96-98] The empirically derived potential functions attempt to account for this non–additivity in some average manner; however, it has been shown that most of these pair potentials do not realistically reproduce both gas and condensed phase water properties.[99,100] Therefore, the performance of potential energy functions which include terms for many–body effects in pure water and water–solute systems have

been evaluated.

4.2 Potential Functions

First, a potential function was developed for water–water interactions which includes a nonadditive term in addition to traditional pairwise additive electrostatic and van der Waals components. Then, in an extension of this step, water–solute potential functions were derived to model hydration and solvation behavior, again including terms for nonadditive contributions.

The work first focused on the water potential, as there seemed to be little reason to pursue the development of nonadditive water–solute potentials if water–water interactions could not be satisfactorily modeled with a nonadditive function. The nonadditive energy term for water–water interactions is a self–consistent field polarization calculation based on equations from classical electrostatics. For the pairwise additive portion of the potential functions, several analytical forms were evaluated. Finally, the RWK2 potential developed by Reimers, Watts, and Klein[101] was adopted and modified to include an SCF polarization term. This approach was taken after disappointing results with other potential function forms, which are discussed more fully below.

The potential function for water interactions includes no intramolecular degrees of freedom (i.e., the water molecules are fixed at equilibrium internal geometries). The intermolecular potential can be decomposed into additive and nonadditive terms which are evaluated separately in calculations. The additive component adopts the analytical form of the RWK2 potential function although the parameters are not identical to those derived by Reimers and coworkers[101]

$$E_{pair} = \sum_{i,j} \frac{q_i q_j}{r_{ij}} + \sum_{O,O} A_{OO} \exp\left\{-\beta_{OO} r_{OO}\right\} + \sum_{H,H} A_{HH} \exp\left\{-\beta_{HH} r_{HH}\right\} +$$
$$\sum_{O,H} A_{OH} \exp\left\{-\alpha_{OH}\left(r_{OH} - r_{min}\right)\right\}\left[\exp\left\{-\alpha_{OH}\left(r_{OH} - r_{min}\right)\right\} - 2\right] \qquad (4.1)$$
$$+ \sum_{O,O} -F\left[C_6\left(\frac{g_6}{R_{sc}}\right)^6 + C_8\left(\frac{g_8}{R_{sc}}\right)^8 + 1.5C_{10}\left(\frac{g_{10}}{R_{sc}}\right)^{10}\right]$$

The first term in equation (1) represents the Coulombic or electrostatic interactions. The second and third terms are exponential repulsions for oxygen-oxygen and hydrogen-hydrogen interactions, respectively. The fourth term is a Morse function for oxygen-hydrogen interactions. The final term in equation (1) is a molecular dispersion term developed by Scoles and coworkers.[102] The coefficients for water–water interactions are taken from work by Meath *et al.*[103] (see reference 101 also) The components of the dispersion term are:

$$g_n = 1 - \exp\left[-\frac{2.1}{n} R_{su} - \frac{0.109}{\sqrt{n}} R_{su}^2\right] \qquad (4.1a)$$

$$R_{sc} = R_{OO}\left[\frac{I_{H_2O}}{I_H}\right]^{\frac{2}{3}} = 0.94834673 \times R_{OO}\,,\ I = \textit{ionization energy} \qquad (4.1b)$$

$$F = 1 - R_{su}^{2.326}\exp\left[-R_{su}\right]\ ;\ R_{su} = \frac{R_{sc}}{0.529177} \qquad (4.1c)$$

where R_{OO} is the intermolecular oxygen–oxygen distance. The nonadditive component is an SCF polarization energy calculation using standard formulas of classical electrostatics. The electric field $\vec{E}$ at a point j is the negative gradient of the potential at that point:

$$\vec{E}_j = -\nabla\varphi_j;\ \varphi_j = \sum_{i;i\neq j} \frac{q_i}{r_{ij}} + \sum_{i;i\neq j} \frac{\vec{\mu}_i \vec{r}_{ij}}{r_{ij}^3} + \textit{quadrupole} + \cdots \qquad (4.2)$$

where q_i is the charge on atom i and $\vec{\mu}_i$ is the induced dipole on atom i. In evaluating the electric field, the quadrupole, octapole, and higher terms are

ignored. The contributions from these higher order terms are much smaller than the charge and dipole terms and the additional computational expense for inclusion of these terms is impractical. Therefore, the expression for the electric field becomes

$$\vec{E}_j = \sum_{i \neq j} \frac{q_i \vec{r}_{ij}}{r_{ij}^3} + \sum_{i \neq j} \frac{\vec{\mu}_i \vec{r}_{ij}}{r_{ij}^5} \vec{r}_{ij} \tag{4.3}$$

The water molecule consists of a series of fixed point charges and polarizable centers (Figure 4.1). The point charges reproduce the permanent dipole and quadrupole moments of the water monomer.[104] The point polarizabilities for the water monomer are taken from Applequist.[105] The basis for this choice of point polarizabilities will be discussed in detail below. Each polarizable center is capable of acquiring an induced dipole according to the formula:

$$\vec{\mu}_j = \alpha_j \vec{E}_j \tag{4.4}$$

where $\vec{E}_j$ is determined from Equation (4.3). In the initial step of the SCF polarization energy calculation, all induced dipoles, $\vec{\mu}_j$, are set to zero. The field due to fixed point charges is evaluated and the dipoles induced by this point charge field are computed in step two. Then, the electric field, $\vec{E}$, is recalculated including fixed point charges and induced dipoles determined in step two. From this new value for the electric field, new induced dipoles are calculated and substituted back into equation (4.3). Equations (4.3) and (4.4) are solved in this iterative procedure until the variation in induced dipoles from one cycle to the next becomes smaller than some predefined convergence value. The polarization energy is then

$$E_{pol} = -\frac{1}{2} \sum_j \alpha_j \left(\vec{E}_j \cdot \vec{E}_j \right) \tag{4.5}$$

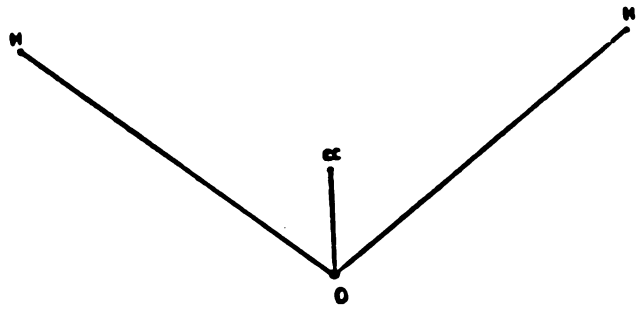


Figure 4.1: Water monomer. $R_{O-H} = 0.9572\text{Å}$, $R_{CC} = 0.2600\text{Å}$, $\theta_{H-O-H} = 104.52°$.

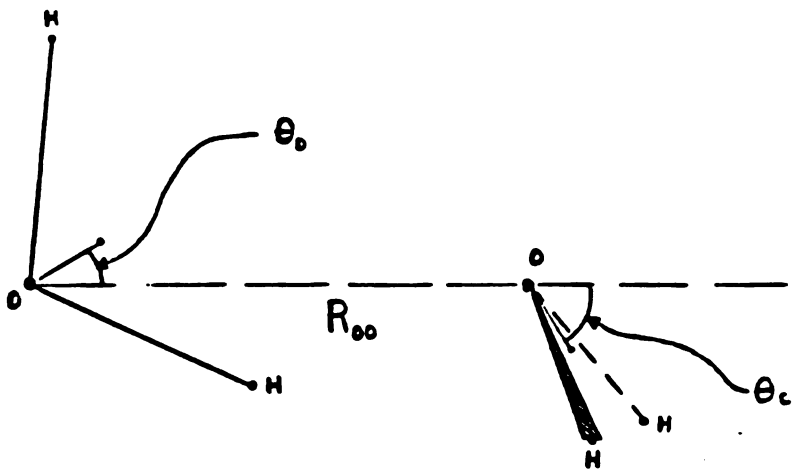


Figure 4.2: Water dimer geometry.

Water–solute potential functions also include additive and nonadditive components. For these potentials, a simple functional form is employed for pairwise additive terms

$$E_{pair} = \sum_{i,j} \frac{q_i q_j}{r_{ij}} + \sum_{i,j} \left[\frac{\sqrt{A_i A_j}}{r_{ij}^{12}} - \frac{\sqrt{C_i C_j}}{r_{ij}^{6}} \right] \tag{4.6}$$

The nonadditive component again consists of an SCF polarization energy calculation. For some solutes, namely simple ionic species like Na^+ or Cl^-, a nonadditive exchange repulsion (EX) component has also been included in the many body potential function term. The exchange repulsion is evaluated over ion–water trimers (Ion-$(H_2O)_2$) and has the following form

$$E_{three-body} = A\left\{\exp\left[-\alpha r_{12}\right]\exp\left[-\alpha r_{13}\right]\exp\left[-\beta r_{23}\right]\right\} \tag{4.7}$$

where r_{12} and r_{13} are ion–oxygen distances for the ion–water trimer and r_{23} is the oxygen–oxygen distance for the two water molecules involved in the ion–water trimer. Thus, the total energy for an ion–water system is

$$E = \left[(4.6) + (4.5) + (4.7)\right]_{ion-water} + \left[(4.1) + (4.5)\right]_{water-water} \tag{4.8}$$

4.3 Parameterization and Calculations

Parameters were first derived for the water potential. Criteria for the water potential included that it satisfactorily predict lattice energies and densities for ice Ih and ice VII,[90,101,106,107] as well as reproduce second virial coefficient values for steam[108,109] and energetic and structural data for water dimer.[101,110] The RWK2 potential was previously parameterized so as to give good results for these various test cases. Thus, the analytical form of the RWK2 function was chosen after unsuccessful attempts to reproduce the test cases with other functional forms. These test conditions represent both gas

(minimal nonadditivity) and condensed phase (appreciable nonadditivity) properties of water. They can also be evaluated relatively rapidly, allowing parameter set evaluation and adjustment in a reasonable time period. In the first step of parameter development, values were chosen to reproduce water dimer results accurately. These parameters were then used to evaluate lattice energies and optimal densities for ice Ih and ice VII. Finally, second virial coefficients were computed at several temperatures.

Water dimer energies and geometries for various choices of parameters were computed with a program developed to perform rigid body minimization using a modified Newton–Raphson algorithm (the program is discussed in more detail in Appendix 3). This step was quite rapid and served as an excellent initial screening mechanism, allowing rapid elimination of unsuitable parameter choices (See Figure 4.2 for dimer geometry).

Lattice energies and densities were also computed using an option of the rigid body minimization program. Rather than performing full minimization of the lattice structures, the lattices were uniformly expanded or contracted in a stepwise procedure. The lattice energy was computed at each density until the optimal lattice energy and density were discovered. The lattice structures for ice Ih and ice VII were based on the coordinates derived by Cota and Hoover.[111] The ice Ih lattice contains 96 water molecules in a "unit cell" and the ice VII lattice has 144 water molecules in the "unit cell". The lattice energy was computed as the single molecule interaction energy with all neighboring molecules within a 6-7 Å radius (molecule based cutoff), averaged over all molecules in the unit cell. Tests of several different cutoff distances suggested that a 7Å cutoff radius was reasonable for both pair–additive and nonadditive terms in the potential function. Extension to

longer cutoff distances lowered the computed lattice energies negligibly but markedly increased the computation time, especially for the SCF polarization term. Periodic boundary conditions were used to generate neighbors around the unit cell, thus preventing surface or edge effects which might give misleading results for computed lattice energies.

Computation of the second virial coefficients was the final and most time–consuming step in parameter evaluation. By fixing the position of one water molecule and considering the molecular symmetry, the second virial coefficient can be written as the integral

$$B(T) = -\frac{N}{2\pi^2}\int_0^{\infty} r^2 dr \int_0^{\frac{\pi}{2}} \sin\Theta d\Theta \int_0^{\pi} d\varphi \int_0^{2\pi} d\alpha \int_0^{\pi} \sin\beta d\beta \int_0^{\pi} d\gamma \times f\left[r,\Theta,\varphi,\alpha,\beta,\gamma\right] \tag{4.9}$$

where N is Avogadro's number and $f(r,\Theta,\varphi,\alpha,\beta,\gamma)$ is the Mayer f-function

$$f(r,\Theta,\varphi,\alpha,\beta,\gamma)=\exp\left[-\frac{E(r,\Theta,\varphi,\alpha,\beta,\gamma)}{kT}\right] - 1 \tag{4.10}$$

with $E(r,\Theta,\varphi,\alpha,\beta,\gamma)$ the potential energy at the given coordinates, k Boltzmann's constant, and T the temperature in degrees Kelvin. Molecule one is fixed with the oxygen atom at the origin and its symmetry axis in the X–direction. The oxygen atom of molecule two has spherical polar coordinates (r,Θ,φ) and the molecule is oriented by Euler rotation angles (α,β,γ) with respect to the reference axes (principal axes of molecule one).

The integration was performed in a procedure similar to that outlined by Reimers.[101] A non–product numerical integration formula was used which requires $n+2$ function evaluations in an n-dimensional unit cube.[112,113] Based on test cases, this method offers at least an order of magnitude

improvement in efficiency over product Gaussian quadrature or product Simpson's rule schemes for evaluating this integral. The r interval was truncated at 12Å, with the region less than 2.5Å approximated by a hard sphere potential. The region from 2.5Å → 12Å was partitioned into twelve uneven intervals and each interval was evaluated using a five point Gaussian quadrature. Contributions for distances greater than 12Å are quite small and were neglected. A hard sphere potential was used to evaluate the integral inside 2.5Å as some configurations generated by the integration algorithm lead to divergence of the SCF polarization calculation because polarizable centers are positioned unrealistically close to each other. Such points would not be observed in normal minimization or molecular dynamics calculations as extremely large energy barriers must be overcome to reach these divergent energy points. Each angle variable range was divided into five intervals of equal size and the five–dimensional angle integral was evaluated using the non–product formula after transformation to the unit hypercube.

These three test procedures were used for the parameter evaluation process. Initially, attempts were made to fit standard analytical expressions, such as 6–12 or exp–6 plus electrostatics functions including the SCF polarization term, to gas phase and ice lattice data. Reproducing ice lattice energies and densities proved to be difficult using this approach. It was also impossible to fit the quantum mechanical points of Clementi[87] to these functional forms and satisfactorily to reproduce second virial coefficient data. This inability to develop a suitable dimer potential using the Clementi CI points has been observed before. It has been suggested that the 64 dimer points in the Clementi configuration interaction calculations are insufficient to determine the water dimer energy hypersurface adequately.[114] After

considerable effort with these more traditional functional forms, the analytical form of the RWK2 pair potential was adopted. This potential is one of the more successful water potentials developed to date as it reproduces many gas and condensed phase properties reasonably well.[101] The parameters were readjusted and the SCF polarization term was added to this modified RWK2 function. The partial charges were fixed at the values derived by Reimers, as these values properly reproduce the permanent dipole and quadrupole moments of the water monomer. The importance of a model with the appropriate quadrupole behavior for liquid water simulations has been observed previously.[116] The molecular dispersion term was also left unaltered. This term was carefully parameterized and was intended to be a nonadjustable correction term for dispersion effects in water–water interactions. That left the exponential repulsion, Morse hydrogen bonding, and polarization terms as adjustable components in the potential function development process. Only the exponential prefactors A_{OO}, A_{HH}, and A_{OH} of the exponential repulsion and Morse terms were modified in the "fitting" procedure. After initial trial modifications, it was observed that alteration of the hydrogen repulsion term (i.e., A_{HH}) had relatively little effect on the potential function behavior. Thus, the "adjustable" parameters in the development of a nonadditive water potential from the RWK2 function included only the A_{OO}, A_{OH}, and point polarizability terms.

Initially, an isotropic molecular point polarizability centered on the negative charge center (the approximate center of mass of the molecule) was assigned to each monomer. The molecular polarizability was equal to the experimental molecular polarizability[116] (1.48Å^3) and the exponential repulsions were concurrently stiffened to counter the additional attraction

from the polarization. This initial guess allowed for good dimer energies and geometries, but invariably gave ice lattice energies that were much too attractive. An analysis of the polarization calculation results indicated that the induced dipoles were too large (≥ 1.5D) with this model, thus leading to polarization energies which were excessive. Given this result, the single molecular polarizability was abandoned in favor of atom–centered polarizable centers.

First, the atomic polarizabilities developed by Thole[117] were tested. These values were derived so as to reproduce accurately the experimental molecular polarizability (see reference 117 for details). Using the Thole polarizability values, better results were obtained but the induced dipoles were still rather large. Because the polarization energy in the ice lattice structures was too attractive due to the large induced dipoles, it was difficult to develop a set of parameters which simultaneously predicted both the lattice energies and densities for the two forms of ice. Next, the atomic polarizabilities for water developed by Applequist[105] were examined. The Applequist values were less appealing than the Thole atomic polarizabilities in one regard as they do not reproduce the experimental molecular polarizability well (see reference 105). However, in the lattice calculations the Applequist polarizabilities led to induced dipoles on water molecules in good agreement with those estimated by Coulson and Eisenberg[118] ($\vec{\mu}_{induced}$ = 0.65-0.75D for net water dipoles of 2.5-2.6D) for water in an ice lattice. Additionally, Applequist has derived atomic polarizabilities for many other atom types which may be helpful in extension of the approach to water–solute interactions. Using the Applequist atomic polarizabilities on oxygen and hydrogen nuclei (no polarizability on the negative charge

center), it was possible to modify the oxygen–oxygen repulsion and Morse term parameters of Reimers,[101] yielding a water potential that gave good results for the three test criteria. The parameters for the function are given in Appendix 1. Tables 4.1-3 contain results obtained using this potential function for water dimer minimizations, ice lattice computations, and second virial coefficient calculations, respectively.

After defining parameters for the water–water interaction potential, development of water–ion potential parameters proved relatively easy. For

TABLE 4.1

	Pot. Func.	Expt.
E(kcal)	-5.7	-5.4 ± 0.2
$R_{O\text{-}O}$(Å)	2.85	2.98
θ_C	52.74°	58.50°
θ_D	55.00°	50.20°

Table 4.1: Water dimer results (after minimization). The second column displays results from the many body potential function. The third column displays experimental data. See Figure 4.2 for the dimer geometry and angle notation.

TABLE 4.2

	Icelh		IceVII	
	ρ(kg/m^3)	E(kcal/mol)	ρ(kg/m^3)	E(kcal/mol)
Pot. Func.	941	-14.36	1415	-12.87
Expt.	936	-14.08	1430	-13.00

Table 4.2: Optimal calculated ice lattice energies and densities versus experimental values.

TABLE 4.3

Temperature(K)	423	573	873	1173
Expt.	-330.0	-117.5	-35.2	-11.6
This work	-358.0	-126.5	-33.0	-8.1
RWK2[101]	-371.5	-128.6	-33.4	-8.3
CI[87,88]	-481.4	-145.9	-27.6	+0.2
HF[85,88]	-241.0	-80.4	-8.4	+10.4
TIP4P[91]	-541.3	-168.6	-41.4	-11.1

Table 4.3: Second virial coefficients(cm^3/mole) for various potential functions and experimental results (all functions evaluated using the procedure outlined in section 4.3)

these potentials, many parameters were already fixed (e.g., partial charges and atomic polarizabilities for water molecules and charges for ions). This left only ionic polarizabilities and parameters for the 6-12 van der Waals term. Many ionic polarizability values are available in the literature.[119-122] Evaluation of several sets of values suggested little variability in results with different polarizability sets. Finally, the free ion polarizabilities developed by Sangster[121] were chosen as these values were consistently derived and represented a set with proven utility in lattice calculations. Next, the 6–12 term parameters were adjusted so as to reproduce experimental and/or accurate *ab initio* quantum mechanical interaction energies and geometries for ion-$(H_2O)_1$ complexes.[123-126] The parameters for the ion–water potential are given in Appendix 1. Then, the energy of ion-$(H_2O)_2$ clusters was evaluated at various geometries using both *ab initio* techniques and the potential functions. For the *ab initio* calculations, a 4–31G basis set was used for water

molecules and Cl^-, with Roos–Siegbahn (10s/4p → 6s/4p) basis sets[127] for Na^+ and Mg^{++}, and a (12s/6p → 8s/4p) basis set[128] for K^+. The energy nonadditivity for potential function and *ab initio* calculations

$$E_{nonadditive} = E_{cluster} - \left\{ E_{ion-water\ 1} + E_{ion-water\ 2} \right\} \qquad (4.11)$$

was computed to evaluate the potential function performance. Although these basis sets overestimate two body (ion–$(H_2O)_1$) interaction energies, the three body nonadditivities are likely to exhibit less basis set dependence.[129] These calculations revealed that the SCF polarization did not adequately model the nonadditivity for ion-water geometries where both waters were in the first coordination sphere of the ion (see Figure 4.3a) As discussed previously,[129] this is due to a larger contribution of exchange repulsion and/or charge transfer to the nonadditivity for these geometries. For longer ion–water intermolecular separations (e.g. one water in the first coordination sphere and the second water ≥ ~ 4Å from the ion, see Figure 4.3b), the SCF polarization term alone adequately reproduces the nonadditivity of these systems. The relative differences in nonadditive three–body energies were fit to the analytical expressions in equation (4.7)

$$E_{three-body} = E_{quan.\ mech.\ nonadditivity} - E_{pot.\ func.\ nonadditivity} \qquad (4.12)$$

using a non–linear least squares fitting algorithm.[130] In the fitting procedure, all ion three body functions were forced to decay smoothly to zero by 4.0-4.5Å in order to obtain functions which were well–behaved in minimization calculations. In earlier least squares fits, this constraint was not imposed. Instead, a cutoff distance was used to suppress evaluation of long range three body interactions that made negligible contributions. This approach frequently led to entrapment problems during minimizations, possibly due to

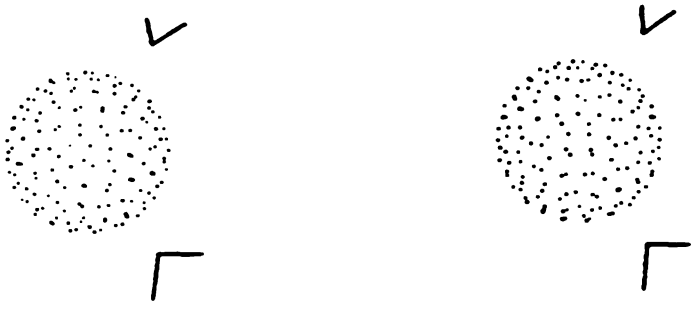


Figure 4.3a: Ion–$(H_2O)_8$. Both waters in first coordination sphere.

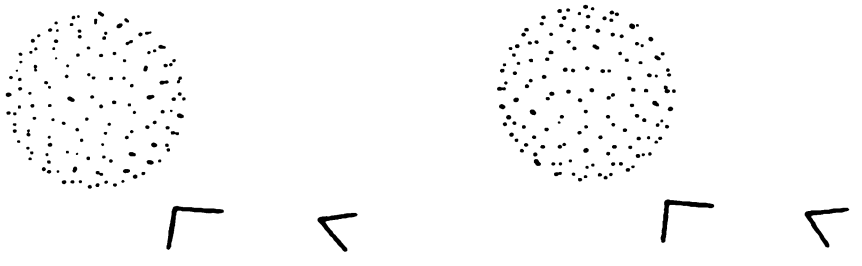


Figure 4.3b: Ion–$(H_2O)_8$. Only one water in first coordination sphere.

discontinuities in the derivatives produced by the rigid cutoff. Inclusion of these long range three body interactions also caused problems. In a solvated ion system, the ion literally has thousands of three body interactions if long range interactions are included. Even though the fitted three body terms are quite small beyond 4Å, they do not decay to zero until 6–7Å. With thousands of three body interactions, these negligible long range three body terms sum to unreasonably large repulsive contributions. This behavior arises because the analytic expression for the three body term cannot be calibrated to reproduce exactly the quantum mechanical nonadditivities (i.e. the analytic three body term does not decay to zero as rapidly as the quantum mechanical calculations suggest it should). Requiring that the functions smoothly decay to zero gave the desired functional behavior for minimization calculations, avoided the cumulative effect of thousands of very small long range three body repulsions, and only slightly reduced the quality of fit for the three body term. Typically, nonadditivities for twelve to fourteen different ion–$(H_2O)_2$ conformations of varying ion–water distances and spatial orientations were computed with both the potential functions and *ab initio* techniques. Applying the constraint that the analytic functions decay to zero by 4–4.5Å as discussed above, these conformations and corresponding energy nonadditivities were used to calibrate the three body term. The root mean square error for the fit of the analytic three body term to the quantum mechanically computed nonadditivities is on the order of 0.5 kcal/mole. This fit generally slightly underestimates the three body repulsion. Four body effects for ion–$(H_2O)_3$ systems computed at the *ab initio* level reveal this term to be small relative to the three body nonadditivity and of opposite sign. Thus, the tendency to underestimate three body corrections with the analytic function results in a fortuitous,

although inadvertent, correction for four body terms. For Na^+, K^+, and Cl^-, the three body nonadditivity is a smooth function which becomes increasingly repulsive as the the ion–water distance becomes shorter. The parameters of the three body term for these ions are given in Appendix 1.

To determine the quality of the monovalent ion–water potentials, the rigid body minimization program was used to refine ion–water clusters (ion-$(H_2O)_n$). The results were compared with the experimental data of Kebarle[123,124] and potential function calculations by other workers.[121,122] Multiple initial configurations were used for each ion–water cluster minimization and the calculations were considered converged when the root mean square deviation of the energy function with respect to atomic coordinate changes was 1.0×10^{-4} kcal/Å or less. The computed ΔE values for formation of the various complexes along with experimental data are given in Table 4.4 for Na^+, K^+, and Cl^-, and energy components for some of the complexes are displayed in Table 4.5. The results of the potential function calculations show quite good agreement with experimental hydration enthalpies for the ions, whereas results from the other potentials are less satisfactory. The computed ΔE values are not exactly equal to ΔH values but are related by the equation

$$\Delta H = \Delta E + p\Delta V \qquad (4.13)$$

under constant pressure conditions. In condensed phase systems, ΔH and ΔE values are compared directly because the volume change (ΔV) in reactions is generally assumed to be negligible.[133] It is not clear that this would be a valid assumption for gas phase processes such as ion hydration reactions. A correction factor

$$p\Delta V = \Delta nRT\left[1 + \frac{B(T)}{V} + \frac{C(T)}{V^2} + \cdots\right] \qquad (4.14)$$

can be described where $B(T)$ is the second virial coefficient and $C(T)$ is the third virial coefficient. Since second virial coefficients have been computed for the water potential, these values were used to estimate the contribution from the second term in equation (4.14). At 298K, this term should be small and of opposite sign compared to the first order term (ΔnRT). The assumption is made that the higher order terms (e.g., third virial coefficient) will make even smaller contributions to the correction factor than the second virial coefficient term. As ΔE values for ion–$(H_2O)_1$ complex formation were calibrated to reproduce experimental ΔH values, no correction is needed here. Using equation (4.14) to estimate corrections for ΔH values, the computed ΔE values are modified by only 2–3 kilocalories per mole for the larger clusters (n=5,6) and correspondingly less for smaller clusters. Even with the correction factor, the computed $\Delta E(\Delta H)$ values usually fall within the range of estimated experimental error (± 1–3 kilocalories per mole) for the ion hydration enthalpies.[134]

Analysis of the refined structures reveals some interesting features for the cation clusters in particular. Both sodium and potassium form triangular complexes with three waters, with the cation sitting in the center of an equilateral triangle formed by the waters. With four waters, the cations sit in the center of a regular tetrahedron, with a water molecule forming each corner of the tetrahedron. For five water complexes, the lowest energy configurations consist of four waters forming a tetrahedral cage around the cation, with the fifth water hydrogen bonding to this tetrahedral "inner sphere". For the sodium complex, the fifth "outer sphere" water bridges between two "inner sphere" waters, linking two corners of the tetrahedron.

TABLE 4.4

n	Na^+	K^+	Cl^-
1	-23.9(-24.0)	-17.8(-17.9)	-13.0(-13.1)
2	-44.7(-43.8)	-33.5(-34.0)	-26.1(-25.8)
3	-61.3(-59.6)	-47.1(-47.2)	-37.8(-37.5)
4	-74.1(-73.4)	-58.8(-59.0)	-49.6(-48.6)
5	-86.2(-85.7)	-69.2(-69.7)	
6	-98.0(-96.4)	-79.3(-79.7)	

Table 4.4: ΔE (kcal/mole) of hydration for ion-$(H_2O)_n$ complexes after minimization (n = number of water molecules in the complex). Experimental hydration enthalpies are in parentheses.[31,32] Results using the Jorgensen potentials for the sodium ion hydration are: -24.1(n = 1), -44.1(n = 2), -63.5(n = 3), -80.7(n = 4), -94.4(n = 5), and -106.4(n = 6). The hexahydrate complex with the Jorgensen potential favors an octahedral water configuration around the cation. Results from *ab initio* calculations are 1) sodium hexahydrate complex in 4 + 2 configuration (ΔE = -141.6 kcal; E_T = -617.271517a.u.) and octahedral configuration (ΔE = -129.5 kcal); 2) potassium hexahydrate in a 4 + 2 configuration (ΔE = -107.1 kcal; E_T = -1054.423830.u.) and octahedral configuration (ΔE = -100.1 kcal); 3) chloride tetrahydrate complex in clustered configuration (ΔE = -75.4 kcal; E_T = -762.775958a.u.) and tetrahedral configuration (ΔE = -60.4 kcal).

For the potassium complex, the fifth "outer sphere" water forms a linear hydrogen bond with one of the "inner sphere" waters. The sodium– and potassium–hexahydrate complexes are logical extensions of the pentahydrate complexes. Sodium displays its tetrahedral "inner sphere" with two "outer sphere" molecules bridging the corners of the tetrahedron (Figure 4.4a). The potassium complex has two "outer sphere" water molecules forming linear hydrogen bonds with two "inner sphere" water molecules (Figure 4.4b). It appears that the larger size of the potassium–four water tetrahedron precludes energetically feasible structures where the "outer sphere" molecules bridge two "inner sphere"

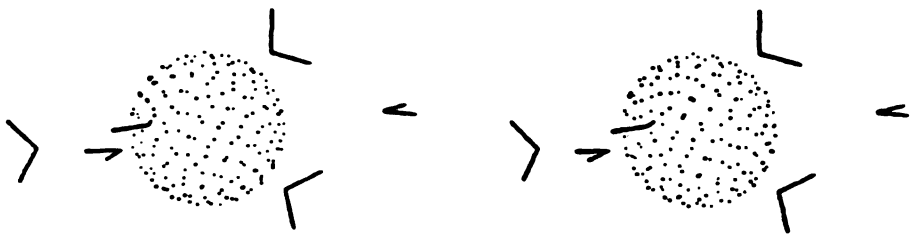


Figure 4.4a: $Na^+-(H_2O)_6$

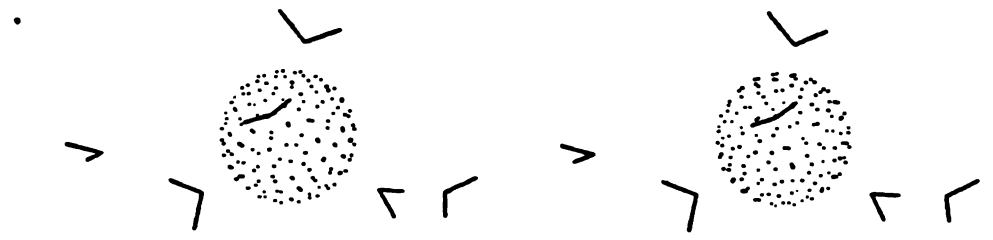


Figure 4.4b: $K^+-(H_2O)_6$

molecules. For sodium, the analogous octahedral hexahydration complex is ~ 10 kilocalories per mole higher in energy than the 4 + 2 inner sphere/outer sphere hydration complex. For potassium, the best octahedral hexahydration complex is ~ 4 kilocalories per mole higher in energy than the 4 + 2 complex. Several *ab initio* quantum mechanical calculations were performed on the sodium–hexahydrate and potassium–hexahydrate complexes in both 4 + 2 "inner sphere–outer sphere" and octahedral configurations using the basis sets described above. For sodium, the quantum mechanical results find the 4 + 2 structure to be more stable than the octahedral conformation by about 11.5 kilocalories per mole, in good agreement with the relative stabilities predicted by the potential functions. The quantum mechanical calculations for potassium find the 4 + 2 structure to be more stable by 7.0 kilocalories per mole (the quantum mechanically computed ΔE values for these complexes are given in the legend of Table 4.4). The basis sets used overestimate the dipole moment of water and, as a result, overestimate the stability of $X^+-(H_2O)_1$ interactions. However, these basis sets would also be expected to overestimate water–water repulsions in the hexahydrate complexes. Thus, there may be an approximate cancellation of the errors which arise from basis set inadequacies. The polarization function calculations of Perez *et al.*[132] for Na^+–water clusters also exhibit non–octahedral geometries for $n = 6$, although their structures are quite different from those described here. By contrast, all the pair potentials appear to predict highly symmetric structures for these test cases.

In chloride hydration complex structures, the water molecules form linear hydrogen bonds to the anion. These complexes differ from the cation hydration complexes in one significant aspect: the water–water interactions

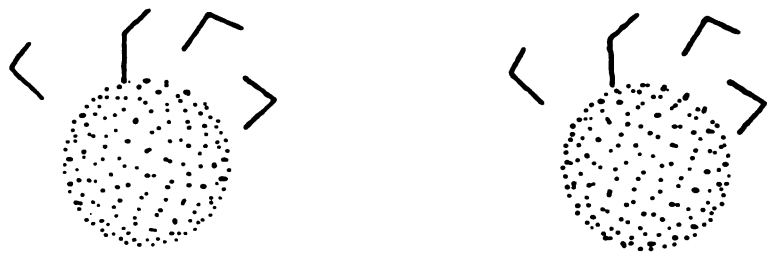


Figure 4.4c: $Cl^{-}-(H_2O)_4$

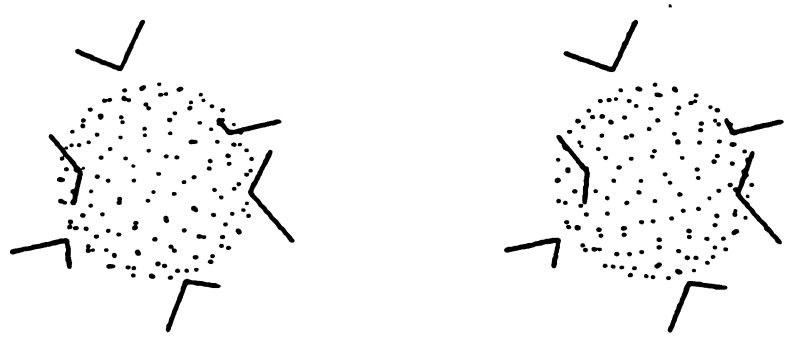


Figure 4.4d: $Mg^{+}-(H_2O)_6$

in the chloride hydration complexes are attractive. The waters form linear hydrogen bonds to the anion and orient themselves so as to form weak hydrogen bonds with each other (see Figure 4.4c for the Cl^-–$(H_2O)_4$ complex). These attractive water–water interactions thus supplement the favorable ion–water interactions formed in the hydration complexes. In the cation hydration complexes, the water–water interactions are repulsive and counterbalance the attractive ion–water interactions in complex formation. The net results of these differing behaviors can be seen in Table 4.4. The Cl^-–$(H_2O)_4$ complex energy is very nearly four times the Cl^-–$(H_2O)_1$ complex energy. By contrast, the Na^+–$(H_2O)_4$ complex energy is significantly less than four times the Na^+–$(H_2O)_1$ complex energy. The K^+ complexes behave as the sodium complexes, although the effect is less pronounced because the water–water distances are much larger in the potassium complexes, thus diminishing the water–water repulsions. The potential functions predict the chloride–tetrahydrate cluster to be more stable than a tetrahedral hydration complex by slightly more then 10 kilocalories per mole; *ab initio* calculations using the basis sets described above find the clustered structure is more stable than the tetrahedral configuration by nearly 15 kilocalories per mole (quantum mechanical results listed in the legend of Table 4.4).

For Mg^{++}, the three body correction (Equation 4.12) is attractive at short distances (≤ 2Å) and somewhat repulsive at intermediate distances (2-3Å) before decaying to zero by 4Å. The simple exponential repulsion term used in three body corrections for the monovalent cations along with the SCF polarization term cannot properly model the observed three body behavior for Mg^{++}. This result is perhaps reflective of observations by Corongiu and Clementi that divalent cations interacting with water behave differently than

TABLE 4.5

COMPLEX	E_{TOT}[a]	E_{PAIR}[b]	E_{POL}[c]	$E_{THREE\ BODY}$[d]
Na^+-$(H_2O)_1$[e]	-23.9	-19.6	-4.3	0.0
Na^+-$(H_2O)_4$[f]	-74.1	-73.1	-7.5	+6.5
Na^+-$(H_2O)_6$[g]	-98.0	-96.8	-10.5	+9.3
Na^+-$(H_2O)_{6OCT}$[h]	-88.0	-99.7	-5.3	+17.0
Mg^{++}-$(H_2O)_1$[i]	-78.5	-46.4	-32.1	0.0
Mg^{++}-$(H_2O)_{6OCT}$[j]	-310.9	-254.5	-56.4	0.0
Cl^--$(H_2O)_1$[k]	-13.0	-10.6	-2.4	0.0
Cl^--$(H_2O)_4$[l]	-49.5	-48.5	-7.0	+5.9

Table 4.5: Energy components (kcal/mole) for some ion-$(H_2O)_n$ complexes after minimization.
a) Total energy of complex formation.
b) Portion of total complex formation energy from pairwise–additive terms.
c) Portion of total complex formation energy from polarization term.
d) Portion of total complex formation energy from three body exchange repulsion term.
e) Optimized Na^+–$(H_2O)_1$ structure, C_{2v} symmetry, R(Na^+–O) = 2.21Å
f) Optimized Na^+–$(H_2O)_4$ structure, tetrahedral cage of water molecules around ion, R(Na^+–O) = 2.33Å
g) Optimized Na^+–$(H_2O)_6$ structure, 4 + 2 symmetry (see Figure 4.4a).
h) Optimized Na^+–$(H_2O)_6$ structure, octahedral water coordination around ion, R(Na^+–O) = 2.42Å
i) Optimized Mg^{++}–$(H_2O)_1$ structure, C_{2v} symmetry, R(Mg^{++}–O) = 1.86Å
j) Optimized Mg^{++}–$(H_2O)_6$ structure, octahedral coordination (see Figure 4.4d) R(Mg^{++}–O) = 2.10Å
k) Optimized Cl^-–$(H_2O)_1$ structure, linear water hydrogen bond with anion, R(Cl^-–O) = 3.33Å
l) Optimized Cl^-–$(H_2O)_4$ structure, linear hydrogen bonds to anion (see Figure 4.4c), R(Cl^-–O) = 3.50Å

monovalent cations.[135] For example, charge transfer is insignificant for monovalent cations but appreciable for divalent cations.[135] The importance

of charge transfer in divalent cation systems may partially account for the unusual behavior observed in the magnesium complexes. Since the region around the optimal ion–water interaction distances [1.86Å for Mg^{++}–$(H_2O)_1$ and ~ 2.1Å for Mg^{++}–$(H_2O)_6$ clusters] requires generally small three body corrections and because the simple analytical expression (Equation 4.7) cannot properly model the three body corrections, the three body exchange repulsion term for magnesium has been omitted. To test the effect of this omission, *ab initio* calculations were performed on two relative minimum energy structures for the Mg^{++}–$(H_2O)_6$ complex using the basis sets described above. One structure corresponded to an octahedral hydration complex and the other structure represented the lowest energy 4 + 2 "inner sphere–outer sphere" geometry found for the hydration complex. The potential functions predicted the octahedral complex to be more stable by ~ 9.0 kilocalories per mole. The *ab initio* calculations indicated that the octahedral complex was more stable by 15.0-16.0 kilocalories per mole. Thus, the omission of the three body term does not appear to have flawed seriously the qualitative behavior of the potential function. It is not clear why explicit three body exchange repulsion terms are required to reproduce the quantum mechanical nonadditivities for Na^+, K^+, and Cl^-, but are not essential to reproduce quantum mechanical nonadditivities for Mg^{++}. The computed quantum mechanical nonadditivities are, as expected, much larger for Mg^{++} than for monovalent ions. However, the nonadditivity from the SCF polarization term alone in the molecular mechanical model seems to mimic the total quantum mechanical nonadditivity reasonably for Mg^{++}, although it tends to underestimate the repulsive nonadditivity at short Mg^{++}–water distances (≤ 1.85Å) and slightly overestimate it at intermediate Mg^{++}–water distances (2.0–3.0Å). It is also quite likely that the omission of

the exchange repulsion term for Mg^{++} systems does not lead to serious errors because the absolute value of the three body correction in these systems is a much smaller fraction of the total complex energy as compared to Na^{+} or K^{+} systems.

Given the encouraging results with the gas phase ion hydration enthalpies, solvation enthalpies and coordination numbers were estimated for the ions. These calculations were especially important for magnesium, as this was the only direct comparison with experimental data for this ion. To estimate these quantities, twenty to forty cycles of minimization were performed on a cluster of ~ 25 water molecules (an adequate number of water molecules to form two complete solvation shells) surrounding the ion using the rigid body minimization program. The cluster molecules were arbitrarily distributed around the ions with no set number of water molecules in the first solvation shell. The minimization with the many body potential functions was used to define the nature of the first and second solvation shells (i.e., ion–water distances and coordination numbers). Then, each ion–water cluster was immersed in a water bath of ~ 200 molecules taken from a Jorgensen Monte Carlo simulation.[91,136] The Jorgensen TIP3P potential[92] (two body potential) was used to relax the water bath environment around the ion, keeping the first and second solvation shells rigidly fixed at the configurations created with the many body functions. Several hundred cycles of minimization were applied to "relax" the bulk water environment around the ion and inner solvation shell waters. After allowing the bulk water environment to relax around the inner core, solvation enthalpies were estimated ($\Delta H \simeq \Delta E$ for condensed phase systems) using the many body potential functions. The solvation enthalpies were computed as

the interaction energy of the ion plus its first solvation shell waters with each other and the remainder of the system, minus the interaction energy of an analogous number of waters with the environment in a bulk water calculation. For example, the sodium solvation system displayed a clearly defined first solvation shell of six waters; therefore, the computed solvation energy for sodium is the interaction energy of the cation and the six waters in the first solvation shell with the environment (~ 220 water molecules) minus the interaction energy of six water molecules with the environment (~ 220 water molecules) in a bulk water system.

The results for the ion solvation enthalpy and coordination number estimations are given in Table 4.6 (representative ion–water structures are shown in Figures 4.5-8). The computed values are in good agreement with experiment.[137-139] The structures are also in reasonable agreement with experiment[140] and previous computer simulations[141] based on $R_{ion\text{-}O}$ and $R_{ion\text{-}H}$ distances when one considers that only limited minimization was performed on these systems. Coordination numbers were determined from $R_{ion\text{-}O}$ distance distributions and computer graphics analysis of the systems. Na^+ and Mg^{++} both display six water first solvation shells, although sodium cation holds its first solvation shell waters less tightly than magnesium. The results for K^+ are less well defined. The first solvation shell for K^+ ranges from four to six waters in the calculations, although analyses of these systems with computer graphics indicate that only four water molecules usually form good dipole–ion interactions. The other water molecules are members of hydrogen bond networks involving primarily second shell molecules that seem to have gotten "trapped" close to the cation. These water molecules moved slowly away from the cation as the minimization progressed, but

never moved beyond the range of apparent first solvation shell waters during the limited minimization, based on ion–oxygen distances. Chloride has seven waters forming good linear hydrogen bonds with the anion, but the distinction between the first and second solvation shell waters is much less clear for Cl^- versus the cations when only ion–oxygen or ion–hydrogen distances are used to define shells. This observation is also consistent with previous computer simulation results.[141] Given the results for gas phase hydration of sodium which predicted a 4 + 2 conformation for the hexahydrate structure, an attempt was made to evaluate the solvation enthalpy for a sodium cation exhibiting a coordination number of four. This was done by constraining four water molecules in a tetrahedral configuration around the cation and proceeding with the steps outlined above for estimation of solvation enthalpies. This evaluation gave an estimated solvation enthalpy of -75 to -80 kilocalories per mole, suggesting that sodium prefers a coordination number of six in solution. It appears that the first shell water interactions with the second shell waters are much poorer for the $N_C = 4$ configuration compared to the $N_C = 6$ configuration, thus leading to a notably less favorable estimated solvation enthalpy.

4.4 Discussion

These many body potential functions have produced results in excellent agreement with a number of experiments. While various pair potentials have yielded good results for some of these test cases, there appear to be no pair potentials which can produce good agreement with experiment for this wide range of environments. For example, the Jorgensen pair potentials give good results for liquid water and reasonable results for sodium and chloride ion

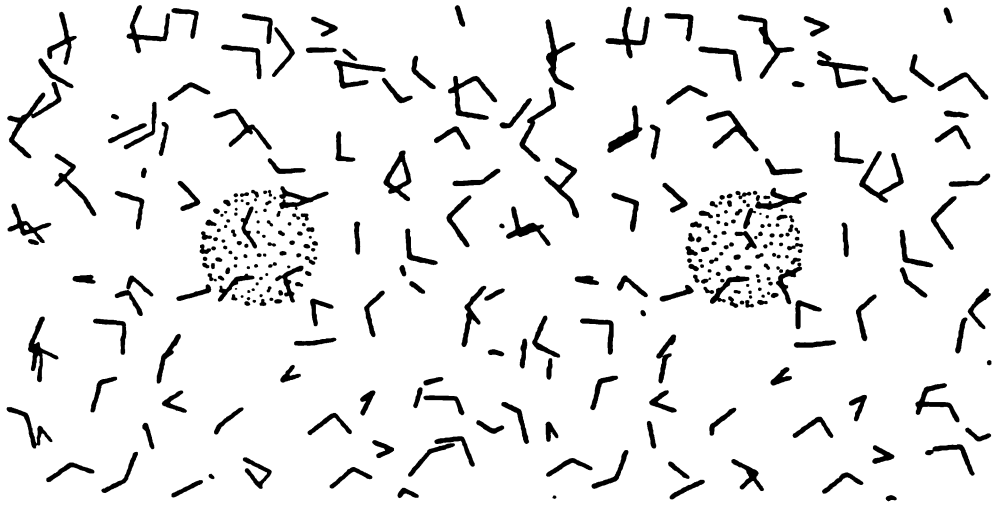


Figure 4.5: Solvated sodium cation

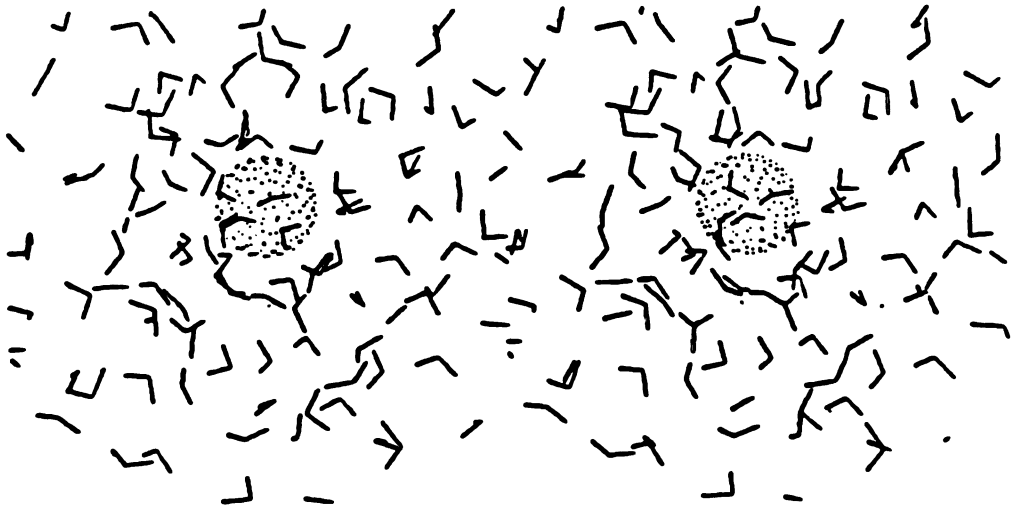


Figure 4.6: Solvated potassium cation

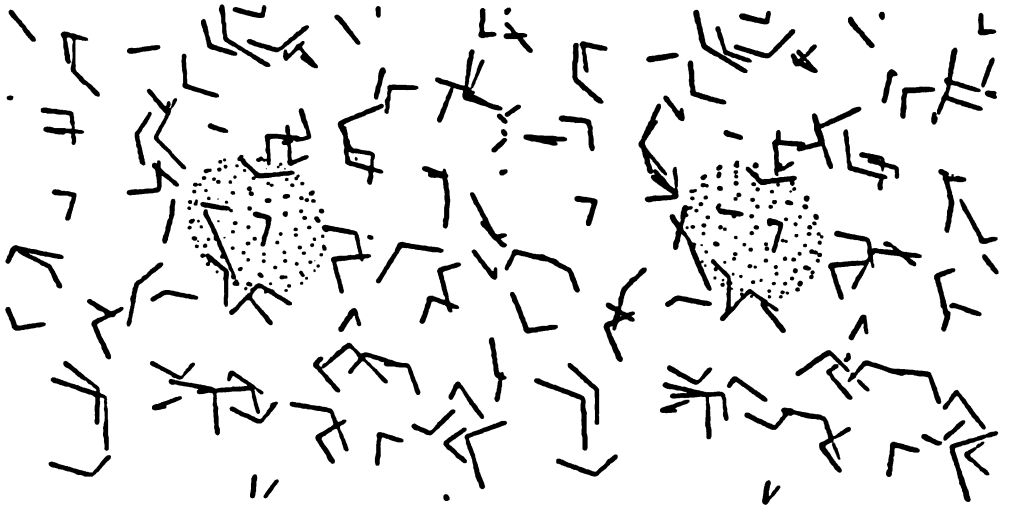


Figure 4.7: Solvated magnesium cation

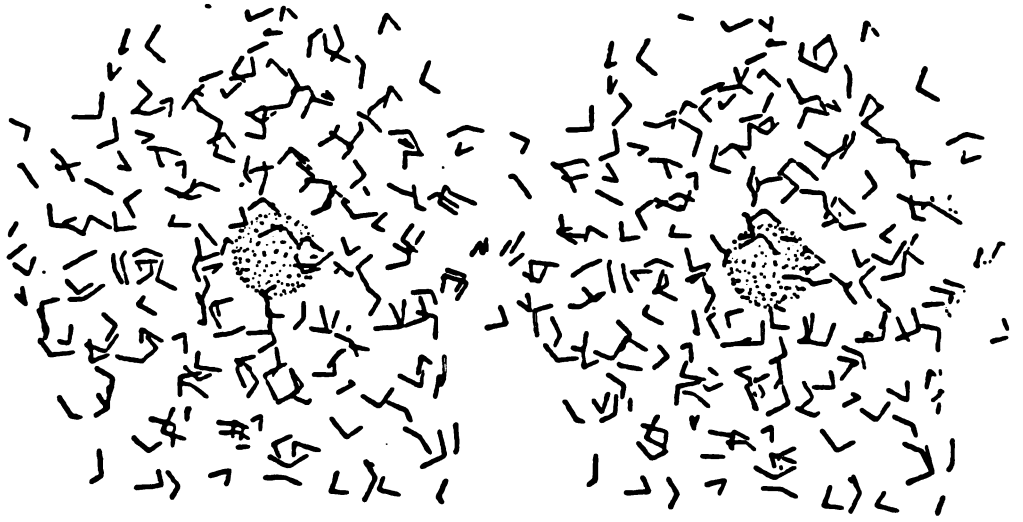


Figure 4.8: Solvated chloride anion

TABLE 4.6

Ion	$-\Delta E_{solvation}$[a](kcal/mole)	$R_{O\text{-}Ion}$(Å)	N_C
Na^+	95.0(102-106)	2.4(2.4)	6(4,6)
K^+	82.0[b](83-86)	2.7(2.7-2.9)	4-5(4)
Mg^{++}	450.0(477)	2.1(2.1)	6(6)
Cl^-	85.0(81-83)	3.50(3.1-3.4)	7(7)

Table 4.6: Solvation energies, ion–oxygen distances for first coordination shell waters, and coordination number for various ions tested. Experimental solvation enthalpies,[137-139] ion–oxygen distances,[140] and coordination numbers[140] are in parentheses.
a) Solvation energies were estimated by computing the total interaction energy of the ion and first coordination shell waters with each other and the environment (~ 220 water molecules), corrected for a corresponding number of water molecules in a pure water energy refinement.
b) There is some ambiguity in the coordination number for potassium cation in these calculations. Four water molecules are clearly coordinated with the ion but a fifth water molecule also lies within the first coordination shell sphere, based on ion–oxygen distance distributions (see Figure 4.6). Because of this ambiguity, the estimated solvation energy for potassium has error limits of ± 5.0 kcal/mole.

solvation studies but do not reliably reproduce the gas phase ion hydration enthalpies, based on tests of these functions (see legend in Table 4.4).

The results obtained from the gas phase ion hydration calculations are especially gratifying. The many body potentials predict hydration enthalpies that are usually within one or two kilocalories/mole of the experimental value. No previous potential functions, either pair additive or many body functions, have reproduced experimental numbers so accurately. These calculations have also suggested rather unusual conformations for some of the water–cation clusters. The 4 + 2 "inner sphere–outer sphere" sodium and potassium hexahydrate geometries were unsuspected but perhaps not a complete surprise in light of some diffraction studies of solvated sodium and

potassium systems, which suggested coordination numbers of four for these cations.[140]

The energies for some ion hydration complexes are broken down into additive and nonadditive components in Table 4.5. This energy decomposition emphasizes the importance of the nonadditive components in successfully predicting complex energies. All effective pair potentials that reproduce single water interactions with the ions invariably overestimate the stability of the larger ion–water clusters. The nonadditive components (especially polarization) allow the van der Waals parameters to be so adjusted as to yield values in good agreement with experiment for ion–single water and ion–many water clusters as indicated in Table 4.4. The necessity of both polarization and three body exchange repulsion terms can also be seen in Table 4.5. For some complexes, the two terms effectively cancel each other (e.g. $Na^+-(H_2O)_6$ in the 4 + 2 geometry). However, correct prediction of the relative stability of $Na^+-(H_2O)_6$ complexes in 4 + 2 and octahedral configurations would not be possible without both three body and polarization terms. The pair additive component actually favors the octahedral geometry; even with polarization the octahedral geometry is only slightly less favorable than the 4 + 2 configuration. Inclusion of both polarization and three body corrections is necessary to properly reproduce the quantum mechanical relative stabilities for these complexes.

The unusual behavior of the three body correction for Mg^{++} systems made it impossible to apply the simple analytical three body function used for the monovalent species. Fortunately, the three body corrections appear to be relatively less important for Mg^{++}, perhaps due to the increased importance of attractive charge transfer effects in divalent cations which

may counterbalance exchange repulsion effects. It would be possible to calibrate a three body term for Mg^{++} with the procedure used for the monovalent ions. However, the analytical expression must be more complicated than Equation 4.7 to model the unusual behavior observed.

The apparent success of the many body potential functions in reproducing solvation enthalpies should not be overstated. The method for determining the solvation enthalpies is crude at best; Monte Carlo simulations on these systems are necessary to evaluate more rigorously the performance of the potential functions. The results from the crude computation of solvation enthalpies does suggest that full simulations will likewise give values in good agreement with experimental data. Boltzmann averaging effects would tend to increase the calculated energies above those of the lowest energy configuration, but as it is quite unlikely that the lowest energy configurations for these solvated ion systems have been located with such limited minimization, it seems unlikely that detailed simulations would yield solvation enthalpies less attractive than those estimated. If anything, detailed simulations may further improve the agreement between potential function results and experimental values. Monte Carlo simulations should also improve the structural results for these systems. The limited minimizations are definitely not adequate to predict confidently features of radial distribution functions, for example.

The advantages of the many body potential functions are indicated by some of the results; namely, the excellent agreement between the computed energies and the experimental data. It appears that these potential functions will allow calculation of both gas phase ion hydration enthalpies and solvation enthalpies. As has been emphasized already, no pair potential which

produces reasonable results for the solvation enthalpies and ion solution structures is also capable of reproducing the gas phase data for ion–water clusters. Little or no work has been done to date on potential functions which reliably model divalent cation–water interactions; explicit inclusion of many body terms for these systems may be especially important if one wishes to reproduce experimental solvation enthalpies, for example. These calculations have indicated that magnesium does induce much larger dipoles in surrounding water molecules than do the monovalent cations and anion studied, thus leading to much more attractive polarization energies for the solvated magnesium system. Yet the van der Waals interactions are not radically different for species like magnesium and sodium cations. It would prove quite challenging to model successfully the relative behavior of magnesium versus sodium cations with effective pair potentials, particularly if good quantitative agreement with experimental quantities like gas phase hydration enthalpies is desired.

It should be emphasized that this many body potential has been derived from an effective two body potential. The original RWK2 function modeled condensed phase water systems with some success because nonadditive effects were incorporated in the potential in some average manner. The parameters for the RWK2 function were so derived as to reproduce both water vapor and ice lattice data; this procedure automatically includes nonadditive effects to reproduce the condensed phase data. The evidence for effective nonadditivity in the RWK2 function is the minimized water dimer obtained with this potential. The energy is in respectable agreement with experiment and the best quantum mechanical calculations (-5.9 kcal/mole vs. -5.4 kcal/mole from experimental work) but the O–O dimer distance is much too short (2.74Å vs. 2.98Å from experimental studies).[101] Some effective

nonadditivity has been removed from the RWK2 function during parameterization of the many body potential, but some intrinsic effective pair potential character remains (e.g., the many body potential predicts a dimer energy of -5.7 kcal/mole and and O–O dimer distance of 2.85Å). It is perhaps not unreasonable to retain some effective nonadditivity in the pair additive terms. The SCF polarization term models the attraction due to induced dipoles, possibly the most important nonadditive energy contribution in water. However, other nonadditive effects are known from quantum mechanical studies, such as charge transfer and exchange repulsion interactions. The functional form of the Morse term should approximately model charge transfer effects,[129] so the retention of this term in the water potential may actually model these nonpolarization nonadditivities. Given the results discussed here, it should be possible to add other nonadditive terms for water–water interactions (analogous to the three body exchange repulsion in ion–water interactions, for example) and develop a potential function that is indeed a true two body potential for water dimer (a potential that would predict the proper water dimer energy and geometry), with the proper nonadditive behavior in condensed phase systems.

There are many applications where these potential functions may be utilized, but one area for which they may be particularly useful is ionophore–ion interactions. In these systems, charge distribution in the ionophore is significantly altered by the presence of an ion. In previous calculations, it has been necessary to increase substantially partial charges from those values which reproduce the dipole moment in 18–crown–6 in order to reproduce experimental cation binding data.[142] Preliminary tests

with these many body functions suggest that fixed partial charges, which correctly reproduce ionophore properties such as dipole moments, can be used successfully to calculate cation complexation enthalpies.[143] The SCF polarization term leads to induced dipoles when an ion is complexed; in the old two body potential model,[142] the increased partial charges were necessary to model polarization (induced dipole) effects. Thus, these many body potential functions may make it possible to model complicated systems like the ionophore–cation complexes with a consistent formalism (i.e., fixed partial charges on each atom for *all* states).

The principle disadvantage at present is the much greater computer time requirements for calculations with many body functions. Although the three body terms for ion–water interactions add little computational overhead, the SCF polarization calculations can be quite time consuming in large systems. Additionally, analytical gradients for the SCF polarization term are not straightforward and have not yet been derived. Therefore, numerical gradient techniques have been used for all calculations discussed here. This is an even greater hindrance to calculations on large systems. Development of analytical gradients for the SCF polarization term will allow application of these many body potential functions to molecular dynamics simulations and minimization problems involving much larger systems. It now appears that the SCF polarization calculation may prove amenable to efficient adaptation to array processors and supercomputers, thus greatly enhancing computation speed and making extensive calculations with these many body potential functions feasible for larger systems.

4.5 Conclusions

The results obtained in these initial tests suggest that the many body potential functions may more realistically model water–water and water–solute (specifically, water–ion) interactions than other presently available potential functions. The potentials must still be tested in Monte Carlo or molecular dynamics liquid state simulations before it is certain that the apparent good reproduction of solvation enthalpies and ion–water structure is not spurious. The potential function performance in gas phase ion–water cluster calculations is clearly better than any existing pair potentials.

These potential functions are novel in that they include not only a nonadditive polarization term but also an exchange repulsion term derived directly from quantum mechanical calculations. A further refinement of these potential functions is the inclusion of intramolecular degrees of freedom (e.g., bond stretching and angle bending) for the water molecules to correct for the inadequacies of rigid water molecule models.[101]

Thus, it seems that potential functions which include many body effects explicitly may offer computer simulation results improved over those obtained with effective pair potentials, particularly for charged systems like solvated ion complexes. Considering the importance of solvent and counterions in biological systems, improved water and ion potential functions may lead to much more realistic and reliable computer simulation models for ligand–biomacromolecule complexes.

CHAPTER 5

SUMMARY

Several examples of computer modeling applications in ligand–biomacromolecule interactions have been presented. These examples hopefully illustrate the utility of computer models, when used judiciously and with reasonable expectations, in the study of complex biomolecular interactions.

The nucleic acid intercalation complex models for ethidium and actinomycin D produced results in generally good agreement with experimental studies. Conflicting base sequence preferences for ethidium intercalation obtained with all–atom and united–atom potential function models illustrate the need for caution when analyzing these results. As discussed in chapter 2, molecular interactions dominated by nonspecific forces (e.g. van der Waals forces) can be particularly difficult to model. In contrast, the results for actinomycin D–nucleic acid complexes were consistent, regardless of model type used. The lack of explicit inclusion of solvent in these calculations further prohibits the extraction of quantitative data from these studies. However, these models have provided interesting explanations for experimentally observed phenomena. For example, cyclic pentapeptide side chains are necessary for formation of strong DNA–actinomycin complexes. The computer models suggest these cyclic pentapeptides form "flaps" which shield the guanine–threonine hydrogen bond network from solvent disruption, thus offering a possible explanation for the strong binding and unusually slow dissociation kinetics of

actinomycin–DNA complexes. The models also suggest that acyclic pentapeptide side chains or shortened peptide side chains (i.e. di– or tripeptides) cannot form an effective shield for the guanine–threonine hydrogen bonds, thus rationalizing the greatly diminished binding affinity and biological activity of these modified actinomycin molecules.

Pharmacophores derived using computer modeling techniques were able to rationalize variable pharmacological properties for several clozapine analogs and differential binding behavior for a series of opioid ligands at the μ receptor and the λ site. These results illustrate the potential utility of computer modeling techniques in the study of ligand interactions at unknown or poorly defined receptor sites. The search for pharmacophoric patterns is an especially challenging task for computer modeling techniques; proposed pharmacophores must constantly be critically evaluated against biological data and refined until they explain that data. Pharmacophores that reliably predict biological activity for a wide range of ligands may be powerful tools in the design of new drugs. The pharmacophore may suggest new compounds that retain the basic features necessary for productive interaction with the receptor, but with modifications that improve absorption and distribution, for instance. Pharmacophoric patterns may also provide insight into the nature of unknown receptor sites. Since the pharmacophore defines those physical characteristics in the ligands (three–dimensional shapes and volumes, spatial orientation of hydrogen bond donor/acceptor groups, electrostatic potential properties) important for formation of a complex at the receptor site, it indirectly provides information about the receptor geometry and composition.

The many body potential functions derived for water–water and water–ion interactions have produced results in excellent agreement with experiment. These potential functions appear to be the first potentials to predict successfully gas phase ion hydration enthalpies. Estimated ion solvation enthalpies and ion–water structure for liquid phase calculations also appear to be in good agreement with experimental results. The reliability of these nonadditive potential functions must still be evaluated in full scale liquid state computer simulations, but the preliminary results are encouraging. These many body potentials may lead to improved computer models for ligand–biomacromolecule complexes by realistically incorporating environmental effects of solvent and counterions.

Computer models and simulation for ligand–biomacromolecule interactions are still relatively new techniques. Only within the past decade or so has much of the requisite hardware (interactive color graphics devices like the Evans and Sutherland PS2, superminicomputers such as the VAX 11/780) necessary for these studies become readily available to most scientists. Computer modeling techniques will probably prove to be of ever–increasing reliability and importance in the study of ligand–biomacromolecule interactions. Improvements in computer technology continue to enhance performance, making larger, more sophisticated simulations feasible. The mathematical models (e.g. potential functions) and algorithms are likewise undergoing continual refinement. Computer modeling techniques, used in concert with experimental studies, should provide much useful information about ligand–biomacromolecule complexes in the future.

REFERENCES

1. A. S. Mildvan, *Phil. Trans. R. Soc. Lond.* **B293**, 65, (1981)

2. D. D. Saperstein, A. J. Rein, M. Poe, and M. F. Leahy, *J. Am. Chem. Soc.* **100**, 4296, (1978)

3. B. J. Kimber, D. V. Griffiths, B. Birdsall, R. W. King, P. Scudder, J. Feeney, G. C. K. Roberts, and A. S. V. Burger *Biochemistry* **16**, 3492, (1977)

4. R. W. Oberfelder, L. L.–Y. Lee, and J. C. Lee, *Biochemistry* **23**, 3813, 3822, (1984)

5. P. F. Cook, G. L. Kenyon, and W. W. Cleland *Biochemistry* **20**, 1204, (1981)

6. D. J. Baker, C. R. Beddell, J. N. Champness, P. J. Goodford, F. E. A. Norrington, D. R. Smith, and D. K. Stammers, *FEBS Lett.* **126**, 49, (1981)

7. G. J. Quigley, A. H. J. Wang, G. Ughetto, G. V. D. Marel, J. H. Van Boom, and A. Rich *Proc. Natl. Acad. Sci. USA* **77**, 7204, (1980)

8. D. A. Matthews, R. A. Alden, J. T. Bolin, D. J. Filman, S. T. Freer, R. Hamlin, W. G. J. Hol, R. L. Kislink, E. J. Pastore, L. T. Plante, N. Xuong, and J. Kraut *J. Biol. Chem.* **253**, 6946, (1978)

9. C. R. Beddell, P. J. Goodford, F. E. A. Norrington, S. Wilkinson, and R. Wootton *Br. J. Pharmacol.* **57**, 201, (1976)

10. L. F. Kuyper, B. Roth, D. P. Baccanari, R. Ferone, C. R. Beddell, J. N. Champness, D. K. Stammers, J. G. Dann, F. E. A. Norrington, D. J. Baker, and

P. J. Goodford *J. Med. Chem.* **25**, 1120, (1982)

11. J. M. Blaney, P. K. Weiner, A. Dearing, P. A. Kollman, E. C. Jorgensen, S. J. Oatley, J. M. Burridge, and C. C. F. Blake *J. Am. Chem. Soc.* **104**, 6424, (1982)

12. G. R. Pack and G. H. Loew, *Biochim Biophys Acta* **519**, 163, (1978)

13. R. L. Ornstein and R. Rein, *Biopolymers* **18**, 2821, (1979)

14. K. J. Miller, R. Brodzinsky, and S. Hall, *Biopolymers* **19**, 2091, (1980)

15. H. M. Sobell, S. C. Jain, and T. D. Sakore, *Nature (London)* **231**, 200, (1971)

16. F. Takusagawa, M. Dabrow, S. Neidle, and H. Berman, *Nature (London)* **296**, 466, (1982)

17. W. Müller, and D. M. Crothers, *J Mol Biol* **35**, 251, (1968)

18. D. Patel, *Biochemistry* **13**, 2388, (1974)

19. *ibid*, 2396

20. T. R. Krugh, *Proc Nat Acad Sci USA* **69**, 1911, (1972)

21. C. G. Reinhardt, and T. R. Krugh, *Biochemistry* **16**, 2890, (1977)

22. S. K. Sengupta *et al*, *J Med Chem* **22**, 797, (1979)

23. S. C. Brown, K. Mullis, C. Levenson, and R. H. Shafer, *Biochemistry* **23**, 403, (1984)

24. H. R. Bürki, A. C. Sayers, W. Ruch, and H. Asper, *Arzneimittel-Forschung* **27**, 1561, (1977)

25. J. Grevel and W. Sadeé, *Science* **221**, 1198, (1983)

26. L. S. Lerman, *J Mol Biol.* **3**, 18, (1961)

27. V. Bloomfield, D. Crothers, and I. Tinoco, Jr., in *Physical Chemistry of Nucleic Acids*, (Harper and Row, New York, 1974) 456

28. T. D. Sakore, S. C. Jain, C. C. Tsai, and H. M. Sobell *Proc. Natl. Acad. Sci. USA* **74**, 188, (1977)

29. H.–S. Sheih, H. M. Berman, M. Dabrow, and S. Neidle *Nucleic Acids Research* **8**, 35, (1980)

30. S. C. Jain and H. M. Sobell, *J Biomol Struct Dyn* **1**, 1161, (1984)

31. *ibid*, 1179

32. B. A. Newton, B. A., in *Advances in Chemotherapy*, Goldin, A. and Hawking, F. Eds., Vol. I, (Academic Press, New York, 1964) 35

33. S. Farber *J Am Med Assoc* **198**, 826, (1966)

34. J. L. Lewis, Jr., *Cancer* **30**, 1517, (1972)

35. I. H. Goldburg, M. Rabinowitz, and E. Reich, *Proc Natl Acad Sci USA* **48**, 2094, (1962)

36. S. Perry, *Cancer Chemother Rep* **58**, 117, (1974)

37. F. Takusagawa, B. M. Goldstein, S. Youngster, R. A. Jones, and H. M. Berman, *J Biol Chem* **259**, 4714, (1984)

38. D. J. Patel, *Biopolymers*, **16**, 2739, (1977)

39. D. J. Patel and L. L. Canuel, *Proc Natl Acad Sci U.S.A.* **74**, 2624, (1977)

40. D. J. Patel and L. L. Canuel, *Proc Natl Acad Sci U.S.A.* **73**, 3343, (1976)

41. D. J. Patel and L. L. Canuel, *Eur J Biochem* **96**, 267, (1979)

42. B. C. Baguley and E.–M. Falkenhaug, *Nucleic Acids Research* **5**, 161, (1978)

43. C. G. Reinhardt and T. R. Krugh, *Biochemistry* **17**, 4845, (1978)

44. T. R. Krugh and C. G. Reinhardt, *J Mol Biol* **97**, 133, (1975)

45. R. V. Kastrup, M. A. Young, and T. R. Krugh, *Biochemistry* **17**, 4855, (1978)

46. J. L. Bresloff and D. M. Crothers, *Biochemistry* **20**, 3547, (1981)

47. P. Weiner, and Kollman, P. A. *J Comput Chem* **2**, 287, (1981)

48. S. J. Weiner, P. A. Kollman, D. Case, U. C. Singh, C. Ghio, G. Alagona, and P. Weiner, *J Am Chem Soc* **106**, 765, (1984)

49. S. J. Weiner, P. A. Kollman, D. Case, and D. Nguyen, *J Comput Chem* (submitted for publication)

50. E. Subramanian, J. Trotter, and C. E. Bugg, *J Cryst Mol Struct* **1**, 3, (1971)

51. A. Hopfinger, *Conformational Properties of Macromolecules* (Academic Press, New York, 1973)

52. UCSF MIDAS, Molecular Interactive Display and Simulation, Laurie Gallo, Conrad Huang, and Thomas Ferrin, University of California, San Francisco (1983)

53. Written by A. Dearing at UCSF, 1980

54. C. J. Alden and S. Arnott, *Nucleic Acids Research* **2**, 1701, (1975)

55. P. A. Kollman, (unpublished results)

56. T. P. Lybrand, A. Dearing, P. Weiner, and P. A. Kollman, *Nucleic Acids Research* **9**, 6995, (1981)

57. R. E. Dickerson, *et al*, *Science* **216**, 475, (1982)

58. S. Rao and P. A. Kollman, "On the Role of Mixed and Uniform Sugar Puckers in DNA Double Helices", *J Am Chem Soc* (in press)

59. H. M. Sobell, and S. C. Jain, *J Mol Biol* **68**, 21, (1972)

60. H. M. Sobell, C. Tsai, S. C. Jain, and S. G. Gilbert, *J Mol Biol* **114**, 333, (1977)

61. M. Gellert, C. E. Smith, D. Neville, and G. Felsenfeld, *J Mol Biol* **11**, 445, (1965)

62. J. Meienhofer and E. Atherton, in *Structure Activity Relationships among the Semisynthetic Antibiotics* D. Perlman, Ed. (Academic Press, New York, 1977) 472

63. A. K. Azad Chowdhury, J. R. Brown, and R. B. Longmore, *J Med Chem* **21**, 607, (1978)

64. M. Connolly, *QCPE Bull* **1**, 75, (1981)

65. J. Schmutz, and C. W. Picard, *Handb. Exp. Pharmacol.* **55**, 3, (1980)

66. J. Idänpään-Heikkilä, E. Alkava, N. Olkinuora, and I. P. Palva, *Eur. J. Clin. Pharmacol.* 11, 193, (1977)

67. M. Froimowitz, *J. Med. Chem.* **25**, 1127, (1982)

68. W. E. Klunk, B. L. Kalman, J. A. Ferrendelli, and D. F. Covey, *Mol Pharmacol* **23**, 511, (1983)

69. C. Humblet, and G. R. Marshall, *Drug Development Res* 1, 409, (1981)

70. S. K. Burt G. H. Loew, and G. M. Hashimoto, *Ann N Y Acad Sci* **367**, 219, (1981)

71. H. Weinstein, R. Osman, J. P. Green, and S. Topiol, *Chem. Appl. At. Mol. Electrost. Potentials, Proc. Symp. Role Electrost. Potentials Chem.* P. Politzer and D. G. Truhlar, eds. (Plenum Press, New York, 1981) 309

72. J. J. Kaufman, P. C. Hariharan, H. E. Popkie, and C. Petrongolo, *Ann. N. Y. Acad. Sci.* **367**, 452, (1981)

73. N. L. Allinger and Y. H. Yun, Operating instructions for MM2 and MM2P programs (1980)

74. T. J. Petcher, and H.–P. Weber, *J Chem Soc, Perkin Trans II*, 12 1415, (1976)

75. I. L. Karle, *Acta Cryst* **B30**, 1682, (1974)

76. R. J. Sime, M. Dobler, and R. L. Sime, *Acta Cryst* **B32**, 2937, (1976)

77. D. Canfield, J. Barrick, and B. C. Giessen, *Acta Cryst* **B35**, 2806, (1979)

78. T. G. Cochran and J. E. Abola, *Acta Cryst* **B31**, 919, (1975)

79. I. L. Karle, R. D. Gilardi, A. V. Fratini, and J. Karle, *Acta Cryst* **B25**, 1469, (1969)

80. U. C. Singh and P. A. Kollman, *J. Comput. Chem.* **5**, 129, (1984)

81. C. B. Pert, G. Pasternak, and S. H. Snyder, *Science* **182**, 1359, (1973)

82. P. L. Wood, S. E. Charleson, D. Lane, and R. L. Hudgin, *Neuropharmacology* **20**, 1215, (1981)

83. D. S. Fries and P. S. Portoghese, *J Med Chem* **19**, 1155, (1976)

84. J. M. Blaney, (personal communication)

85. H. Popkie, H. Kistenmacher, and E. Clementi, *J Chem Phys* **59**, 1325 (1973)

86. H. Kistenmacher, G. C. Lie, H. Popkie, and E. Clementi, *J Chem Phys* **61**, 546 (1974)

87. O. Matsuoka, E. Clementi, and M. Yoshimine, *J Chem Phys* **64**, 1351 (1976)

88. G. C. Lie, E. Clementi, and M. Yoshimine, *J Chem Phys* **64**, 2314 (1976)

89. F. H. Stillinger and A. Rahman, *J Chem Phys* **60**, 1545 (1974)

90. L. L. Shipman and H. A. Scheraga, *J Phys Chem* **78**, 909 (1974)

91. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, *J Chem Phys* **79**, 926 (1983)

92. W. L. Jorgensen, *J Am Chem Soc* **103**, 335 (1981)

93. F. H. Stillinger and C. W. David, *J Chem Phys* **69**, 1473 (1978)

94. F. H. Stillinger, *Science* **209**, 451 (1980)

95. P. Barnes, J. L. Finney, J. D. Nicholas, and J. E. Quinn, *Nature* **282**, 459 (1979)

96. R. O. Watts and I. J. McGee, *Liquid State Chemical Physics* (Wiley-Interscience, New York, 1976)

97. J. L. Finney, *Farad Trans Chem Soc* **66**, 80, 86 (1978)

98. G. D. Zeiss, W. J. Meath, J. C. F. McDonald, and D. J. Dawson, *Mol Phys* **39**, 1055 (1980)

99. I. R. McDonald and M. L. Klein, *Farad Disc Chem Soc* **66**, 48 (1978)

100. M. D. Morse and S. A. Rice, *J Chem Phys* **76**, 650 (1982)

101. J. R. Reimers, R. O. Watts, and M. L. Klein, *Chem Phys* **64**, 95 (1982)

102. C. Douketis, G. Scoles, S. Marchetti, M. Zen, and A. J. Thakker, *J Chem Phys* **76**, 3057 (1982) and references cited therein

103. D. J. Margoliash, T. R. Proctor, G. D. Zeiss, and W. J. Meath, *Mol Phys* **35**, 747 (1978)

104. D. Eisenberg and W. Kauzmann, *The Structure and Properties of Water* (Oxford University Press, London, 1969)

105. J. Applequist, J. R. Carl, and K-K. Fung, *J Am Chem Soc* **94**, 2952 (1972)

106. E. Whalley, in *Physics and Chemistry of Ice* E. Whalley, S. J. Jones, and L. W. Gold, eds. (Royal Society Canada, Ottawa, 1973)

107. B. Olinger and P. M. Halleck, *J Chem Phys* **62**, 94 (1975)

108. G. S. Kell, G. E. McLaurin, and E. Whalley, *J Chem Phys* **48**, 3805 (1968)

109. M. P. Vukalovich, M. S. Traskhtengerts, and G. A. Spiridonov, *Teploenergetika* **14**, 65 (1967)

110. J. A. Odutola and T. R. Dyke, *J Chem Phys* **72**, 5062 (1980) and references cited therein

111. E. Cota and W. G. Hoover, *J Chem Phys* **67**, 3839 (1977)

112. A. H. Stroud, *Approximate Calculation of Multiple Integrals* (Prentice-Hall, 1971)

113. D. J. Evans, in *Computational Methods in Mathematical Physics*, edited by R. S. Anderssen and R. O. Watts (Queensland, University Press, 1974)

114. J. L. Finney, (personal communication)

115. I. R. McDonald and M. L. Klein, *J Chem Phys* **68**, 4875 (1978) and references cited therein

116. P. W. Atkins, in *Physical Chemistry* (W. H. Freeman and Co., San Francisco, 1982) 773

117. B. T. Thole, *Chem Phys* **59**, 341 (1981)

118. C. A. Coulson and D. Eisenberg, *Proc R Soc* **A291**, 445, 454 (1966)

119. L. Pauling, *Proc R Soc Lond, Sec A* **114**, 181 (1927)

120. K. Fajans and N. Bauer, *J Am Chem Soc* **64**, 3023 (1942)

121. M. J. L. Sangster and R. M. Atwood, *J Phys, C; Solid State Physics* **11**, 1541 (1978)

122. H. Coker, *J Phys Chem* **80**, 2078, 2084 (1976)

123. I. Dzidic and P. Kebarle, *J Phys Chem* **74**, 1466 (1970)

124. M. Arshadi, R. Yamdagni, and P. Kebarle, *J Phys Chem* **74**, 1475 (1970)

125. H. Kistenmacher, H. Popkie, and E. Clementi, *J Chem Phys* **59**, 5842 (1973)

126. P. A. Kollman and I. D. Kuntz, *J Am Chem Soc* **94**, 9236 (1972)

127. B. Roos and P. Siegbahn, *Theoret Chim Acta* **17**, 209 (1970)

128. B. Roos, A. Veillard, and G. Vinot, *Theoret Chim Acta* **20**, 1 (1971)

129. P. A. Kollman, *J Am Chem Soc* **99**, 4875 (1979)

130. International Mathematical and Statistical Libraries, Inc., Houston, Texas

131. H. Kistenmacher, H. Popkie, and E. Clementi, *J Chem Phys* **61**, 799 (1974)

132. P. Perez, W. K. Lee, and E. W. Prohofsky, *J Chem Phys* **79**, 388 (1983)

133. P. Atkins, in *Physical Chemistry* (W. H. Freeman and Co., San Francisco, 1982) 111

134. H. Kistenmacher, H. Popkie, and E. Clementi, *J Chem Phys* **59**, 5842 (1973)

135. G. Corongiu and E. Clementi, *J Chem Phys* **69**, 4885 (1978)

136. Coordinates from a Monte Carlo simulation provided by Prof. W. L. Jorgensen

137. H. L. Friedman and C. V. Krishnan, in *Water, A Comprehensive Treatise* edited by F. Franks (Plenum Press, New York, 1972) Vol. 3, Ch. 1

138. J. O'M. Bockris and A. K. N. Reddy, *Modern Electrochemistry* (Plenum Press, New York, 1970) Vol. 1, Ch. 2

139. J. E. Desnoyers and C. Jolicoeur, in *Modern Aspects of Electrochemistry* edited by J. O'M. Bockris and B. E. Conway (Plenum Press, New York, 1969) Vol. 5, Ch. 1

140. J. E. Enderby and G. W. Neilson, in *Water, A Comprehensive Treatise* edited by F. Franks (Plenum Press, New York, 1979) Vol. 6, Ch. 1

141. J. Chandrasekhar, D. C. Spellmeyer, and W. L. Jorgensen, *J Am Chem Soc* **106**, 903 (1984) and references cited therein

142. G. Wipff, P. Weiner, and P. Kollman, *J Am Chem Soc* **104**, 3249 (1982)

143. G. Wipff, T. Lybrand, and P. Kollman, (in preparation)

Appendices

Appendix 1: Parameters

These tables contain potential function paramenters for ethidium cation and actinomycin D not given in references 48 and 49. Table A1.1 contains bond, bond angle and torsion angle parameters. Tables A1.2–8 contain partial charges for these molecules.

Tables A1.9–11 contain many body potential function parameters for water and ion work discussed in Chapter 4.

TABLE A1.1

Bond	k_b (kcal/mole Å^2)	r_b (Å)
CA-CA	469.0	1.40
CA-CC	317.0	1.49
CA-HC	331.0	1.08
CB-CC	469.0	1.42
CC-N*	514.0	1.34
CT-N*	337.0	1.50
CX-CA	416.0	1.44
CX-O	570.0	1.23

Angle	k_θ (kcal/mole radians^2)	θ_o (degrees)
CT-CT-HC	35.0	109.50
N*-CT-HC	35.0	109.50
CA-CA-CB	70.0	120.00
CA-CA-CC	70.0	120.00
CA-CA-CX	70.0	120.00
CB-CA-CX	70.0	120.00
CT-CA-CX	70.0	120.00
CA-CA-N2	70.0	120.00
CX-CA-N2	70.0	120.00
CA-CB-CC	70.0	120.00
CA-CC-CB	70.0	118.00
CA-CB-N*	70.0	120.00
CA-CC-N*	70.0	121.00
CB-CB-CB	70.0	119.50
CB-CB-CC	70.0	119.00
CB-CC-N*	70.0	121.00
CA-CX-CA	70.0	120.00
CA-CX-O	70.0	120.00
CB-N*-CC	70.0	122.00
CB-N*-CT	70.0	119.00
CC-N*-CT	70.0	119.00
CA-CA-CA	70.0	120.00
CA-CA-HC	40.0	120.00
CB-CA-HC	40.0	120.00

Dihedral	k_d (kcal/mole)	γ (degrees)	n
X-CA-CA-X	11.0	180.00	4
X-CA-CX-X	7.0	180.00	4
X-CB-CC-X	9.2	180.00	4
X-CC-N*-X	10.0	180.00	4
X-CT-N*-X	0.0	0.00	6

Table A1.1: Additional bond, bond angle, and torsion angle parameters not given in references 48 and 49.

TABLE A1.2

ATOM	TYPE	CHARGE	ATOM	TYPE	CHARGE	ATOM	TYPE	CHARGE
N8	N2	-.2269	N3	N2	-.2232	H17	HC	.0205
HA8	H2	.1281	HA3	H2	.1330	C18	CA	.0151
HB8	H2	.1244	HB3	H2	.1262	H18	HC	.0204
C8	CA	.1757	C4	CA	-.1219	C19	CA	.0183
C9	CA	-.0095	H4	HC	.0163	H19	HC	.0208
H9	HC	.0258	C13	CB	.1277	C20	CA	.0052
C10	CA	-.0017	N5	N*	-.0275	H20	HC	.0068
H10	HC	.0167	C6	CC	.1985	C21	CT	.0951
C11	CB	.0605	C14	CB	-.0077	H21A	HC	.0264
C12	CB	-.0296	C7	CA	-.0676	H21B	HC	.0183
C1	CA	.0391	H7	HC	.0109	C22	CT	-.0325
H1	HC	.0151	C15	CA	.0039	H22A	HC	.0288
C2	CA	-.0581	C16	CA	.0044	H22B	HC	.0418
H2	HC	.0252	H16	HC	.0051	H22C	HC	.0243
C3	CA	.2097	C17	CA	.0181			

Table A1.2: Atomic charges and atom types for ethidium cation.

TABLE A1.3

ATOM	TYPE	CHARGE	ATOM	TYPE	CHARGE	ATOM	TYPE	CHARGE
C2	CA	.1477	C12	CB	.1410	H8	HC	-.0027
N2	N2	-.2070	O5	OS	-.2089	C9	CA	-.0528
H2A	H2	.1274	C13	CB	.1110	CO9	C	.3713
H2B	H2	.1278	C6	CA	.0164	O9	O	-.3654
C3	CX	.2199	C6'	CT	-.0127	C14	CB	.0829
O3	O	-.2628	H6A	HC	.0119	N10	NC	-.1761
C4	CA	-.0719	H6B	HC	.0012	C11	CB	.1192
C4'	CT	.0038	H6C	HC	.0127	C1	CA	-.1569
H4A	HC	.0086	C7	CA	-.0018	CO1	C	.3800
H4B	HC	.0036	H7	HC	.0018	O1	O	-.3889
H4C	HC	.0074	C8	CA	.0106			

Table A1.3: Atomic charges and atom types for actinomycin D chromophore.

TABLE A1.4

ATOM	TYPE	CHARGE	ATOM	TYPE	CHARGE	ATOM	TYPE	CHARGE
N	N	-.1521	HB	HC	-.0152	OG1	OS	-.2478
H	H	.1500	CG2	CT	-.0401	C	C	.3454
CA	CT	.0122	HGA	HC	.0163	O	O	-.3798
HA	HC	-.0028	HGB	HC	.0150			
CB	CT	.1781	HGC	HC	.0135			

Table A1.4: Atomic charges and atom types for threonine.

TABLE A1.5

ATOM	TYPE	CHARGE	ATOM	TYPE	CHARGE	ATOM	TYPE	CHARGE
N	N	-.1981	CG1	CT	-.0144	H2B	HC	-.0011
H	H	.1136	H1A	HC	.0063	H2C	HC	-.0055
CA	CT	.0681	H1B	HC	.0020	C	C	.3297
HA	HC	.0078	H1C	HC	.0103	O	O	-.3649
CB	CT	.0417	CG2	CT	-.0157			
HB	HC	-.0038	H2A	HC	.0254			

Table A1.5: Atomic charges and atom types for D-valine.

TABLE A1.6

ATOM	TYPE	CHARGE	ATOM	TYPE	CHARGE	ATOM	TYPE	CHARGE
N	N	-.1581	HG1	HC	.0030	CA	CT	.0470
CD	CT	.1202	HG2	HC	-.0022	HA	HC	.0128
HD1	HC	-.0076	CB	CT	.0229	C	C	.3324
HD2	HC	-.0093	HB1	HC	-.0075	O	O	-.3663
CG	CT	.0058	HB2	HC	.0041			

Table A1.6: Atomic charges and atom types for proline.

TABLE A1.7

ATOM	TYPE	CHARGE	ATOM	TYPE	CHARGE	ATOM	TYPE	CHARGE
N	N	-.1483	HNC	HC	-.0005	C	C	.3421
CN	CT	.0943	CA	CT	.0348	O	O	-.3429
HNA	HC	.0083	HA1	HC	.0355			
HNB	HC	-.0062	HA2	HC	.0010			

Table A1.7: Atomic charges and atom types for sarcosine.

TABLE A1.8

ATOM	TYPE	CHARGE	ATOM	TYPE	CHARGE	ATOM	TYPE	CHARGE
N	N	-.1621	CB	CT	.0329	H2A	HC	-.0028
CN	CT	.0914	HB	HC	.0163	H2B	HC	.0062
HNA	HC	-.0104	CG1	CT	-.0117	H2C	HC	.0017
HNB	HC	.0004	H1A	HC	-.0016	C	C	.3649
HNC	HC	-.0058	H1B	HC	.0038	O	O	-.3151
CA	CT	.0773	H1C	HC	.0105			
HA	HC	.0016	CG2	CT	-.0125			

Table A1.8: Atomic charges and atom types for N-methyl valine.

TABLE A1.9

A_{OO}	4.05000 × 10^6 kcal/mole
A_{HH}	6.31918 × 10^2 kcal/mole
A_{OH}	1.30000 kcal/mole
β_{OO}	4.97019Å^{-1}
β_{HH}	3.28059Å^{-1}
β_{OH}	7.36154Å^{-1}
r_{min}	1.63781Å
C_6	6.25450 × 10^2 kcal/mole Å^6
C_8	3.39000 × 10^3 kcal/mole Å^8
C_{10}	2.12000 × 10^4 kcal/mole Å^{10}
q_H	0.60000e
q_O	0.00000e
q_{CC}	-1.20000e

Table A1.9: Water–water interaction potential parameters.

TABLE A1.10

Atom type	r^*(Å)	ε(kcal)	α(Å^3)
O	1.77	0.15	0.465
H	1.00	0.02	0.135
Mg^{++}	1.16	0.05	0.120
Na^+	1.28	0.10	0.240
K^+	1.70	0.10	1.060
Cl^-	2.25	0.20	3.240

Table A1.10: Van der Waals parameters and polarizabilities for atoms. The relationship between r^*, ε values and A, C values in equation (4.6) is:

$$C_i = 2\varepsilon_i \left[r_i^* + r_i^*\right]^6$$
$$A_i = \varepsilon_i \left[r_i^* + r_i^*\right]^{12}$$

TABLE A1.11

Ion	A(kcal)	α(Å^{-1})	β(Å^{-1})
Na^+	1430.86	0.787	0.927
K^+	1430.86	0.787	0.927
Cl^-	41227.0	1.309	0.436

Table A1.11: Parameters for three–body terms for ion-$(H_2O)_2$ interactions.

Appendix 2: NMR NOE vs. Molecular Mechanics Model Distances

TABLE A2.1

proton	cross peak	Calculated Distance[a]	Model Distance[b]
$A^8_{4,9}$	C^8_9	3.6	5.6
	$T^8_{5,9}$	3.5	5.9
	$A^{1'}_{4,9}$	2.7	3.9
	$C^{1'}_9$	3.5	3.4
	$A^{3'}_4$	3.0	4.4
	$T^{3'}_5$	3.7	9.7
	$A^{4'}_4$	3.6	5.0
	$A^{5',5''}_4$	3.3	4.4
	$A^{2',2''}_4$	2.4,2.8	2.3,3.7
	$C^{2''}_9$	3.1	2.4
	$T^5_{5,9}$	2.9	3.8
	$C^{2'}_9$	3.6	4.1

$A^{6}_{t,b}$	$T^{6}_{t,b}$	3.7	6.0
	$A^{1'}_{t,b}$	2.7	3.9
	$C^{1'}_{b}$	4.0	3.3
	$A^{3'}_{t,b}$	3.0	4.5
	$A^{4'}_{t,b}$	3.5	5.0
	$A^{2'}_{t,b}$	2.4	2.3
	$A^{2''}_{t,b}$	2.5	3.7
	$C^{2'}_{b}$	3.1	2.3
	$T^{5}_{t,b}$	2.8	4.0
	$C^{2''}_{b}$	3.1	4.0
$A^{6}_{t,b}$	$T^{6}_{t,b}$	4.3	4.4
	$A^{1'}_{t,b}$	2.6	3.9
	$A^{3'}_{t,b}$	3.5	4.4
	$A^{4'}_{t,b}$	3.8	4.8
	$A^{5',5''}_{t,b}$	3.1	4.1
	$A^{2'}_{t,b}$	2.9	3.8
	$A^{2''}_{t,b}$	2.6	2.4
	$T^{5}_{t,b}$	3.0	3.1
$A^{2}_{t,q}$	$A^{1'}_{t,q}$	3.4	4.6
	$A^{1'}_{t,b}$	3.8	5.0
	$Sar\,\alpha_1$	3.1	3.6
	$Sar\,\alpha_2$	3.0	3.5
	$Nmv\ N-CH_3$	2.4	2.6

$A^2_{t.b}$	$A^{1'}_{t.b}$	3.4	4.6
	$Sar\,\alpha_1$	3.3	4.0
	$Sar\,\alpha_2$	3.2	3.7
	$Nmv\ N-CH_3$	2.6	2.6
$A^2_{t.q}$	$A^{1'}_{t.b}$	3.9	7.5
	$T^{2',2''}_t$	3.6	6.0
	$T^5_{t.q}$	3.8	7.0
G^8_q	$T^6_{t.q}$	3.4	4.2
	$G^{1'}_q$	2.9	3.9
	$T^{2'}_{t.q}$	3.1	2.5
	$G^{4'}_q$	3.3	4.8
	$G^{5',5''}_q$	3.8	4.9,5.3
	$G^{2'}_q$	2.1	2.8
	$4-CH_3$	3.3	3.7
	$T^{2''}_{t.q}$	2.8	2.4

	$T_{t,b}^{6}$	3.0	4.9
	$H7$	3.0	3.6
	C_b^5	3.7	6.0
	$C_b^{1'}$	3.7	8.8
	$G_b^{1'}$	3.0	3.9
	$T_{t,b}^{1'}$	3.6	4.4
	$G_b^{3'}$	3.2	2.5
G_b^8	$T_{t,b}^{3'}$	3.2	4.3
	$G_b^{4'}$	3.5	4.5
	$G_b^{5',5''}$	3.2	4.3,4.6
	$G_b^{2'}$	2.6	4.4
	$G_b^{2''}$	2.3	3.4
	$6-CH_3$	3.0	4.1
	$T_{t,b}^{2'}$	2.9	2.6
	$T_{t,b}^{2''}$	2.8	2.4
	$H7$	3.3	3.5
	C_b^5	2.4	2.4
	$C_b^{1'}$	2.4	3.8
	$C_b^{3'}$	2.8	4.3
	$C_b^{4'}$	3.6	4.6
C_b^8	$C_b^{5'}$	3.3	4.2
	$C_b^{5''}$	3.4	2.5
	$C_b^{2'}$	2.5	2.2
	$C_b^{2''}$	2.6	3.6
	$A_{t,b}^{8}$	4.5	5.6

C_q^6	$C_q^{5'}$	2.4	2.4
	$C_q^{1'}$	3.2	3.7
	$C_q^{3'}$	3.0	4.2
	$C_q^{5'}$	3.1	2.6
	$C_q^{5''}$	3.1	4.3
	$C_q^{2'}$	2.3	2.1
	$C_q^{2''}$	2.5	3.5
$T_{t,b}^6$	$A_{t,b}^{1'}$	2.8	3.1
	$T_{t,b}^{3'}$	2.8	4.2
	$T_{t,b}^{4'}$	2.7	4.6
	$T_{t,b}^{4'}$	2.7	4.1
	$T_{t,b}^{5''}$	2.7	4.4
	$T_{t,b}^{2'}$	2.2	2.3
	$T_{t,b}^{2''}$	2.2	3.7
	$A_{t,b}^{2'}$	3.0	4.4
	$A_{t,b}^{2''}$	2.7	2.6
	T_t^5	2.4	2.3
$T_{t,q}^6$	$A_{t,q}^{1'}$	3.2	3.2
	$T_{t,q}^{1'}$	3.1	3.7
	$T_{t,q}^{3'}$	2.9	4.2
	$T_{t,q}^{5'}$	3.1	4.1
	$T_{t,q}^{5''}$	3.1	4.4
	$T_{t,q}^{2'}$	2.4	2.2
	$T_{t,q}^{2''}$	2.4	3.7
	$T_{t,q}^5$	2.6	2.4
	$A_{t,q}^{2'}$	2.7	2.5
	$A_{t,q}^{2''}$	2.7	4.3

$T^6_{i,b}$	$A^6_{i,b}$	3.9	4.4
	G^6_b	≥3.9	4.9
	$A^{1'}_{i,b}$	3.6	3.8
	$T^{1'}_{i,b}$	3.1	3.7
	$T^{3'}_{i,b}$	2.7	3.7
	$T^{2'}_{i,b}$	2.4	2.7
	$T^{2''}_{i,b}$	2.7	4.1
	$T^5_{i,b}$	2.5	2.4
	$A^{2'}_{i,b}$	2.8	2.4
	$A^{2''}_{i,b}$	3.0	2.7
$H8$	G^8_b	3.2	4.9
	C^6_b	3.6	3.5
	$H7$	2.5	2.5
	C^5_b	3.0	4.3
	$C^{1'}_b$	3.0	4.1
	$G^{1'}_b$	3.0	4.2
	$G^{2'}_b$	2.7	2.1
	$G^{2''}_b$	2.8	3.3
	$6-CH_3$	3.2	4.7
	$Thr\,\gamma$	2.6	2.4
$T^6_{i,q}$	$A^6_{i,q}$	4.2	4.5
	G^6_q	3.7	4.2
	$A^{1'}_{i,q}$	3.4	3.9
	$T^{1'}_{i,q}$	3.0	3.7
	$T^{3'}_{i,q}$	2.7	3.9
	$T^{4'}_{i,q}$	3.0	4.6
	$T^{2'}_{i,q}$	2.4	2.5
	$T^{2''}_{i,q}$	2.8	3.9
	$T^5_{i,q}$	2.4	2.4
	$A^{2'}_{i,q}$	2.9	2.4
	$A^{2''}_{i,q}$	3.1	3.3

	G_b^5	3.2	3.6
	C_b^6	3.1	3.5
	$H8$	2.5	2.5
	C_b^5	2.8	2.9
$H7$	$C_b^{1'}$	2.8	5.1
	$G_b^{1'}$	3.5	5.0
	$G_b^{2'}$	3.0	2.7
	$G_b^{2''}$	3.2	4.5
	$6-CH_3$	2.4	2.4
	$A_{t,b}^6$	2.6	3.9
	$T_{t,q}^6$	3.7	11.2
	$T_{t,b}^6$	3.1	3.8
	$A_{t,b}^{3'}$	3.3	3.8
	$A_{t,b}^{4'}$	3.4	3.3
	$T_{t,b}^{3'}$	4.0	5.1
$A_{t,b}^{1'}$	$T_{t,b}^{4'}$	3.7	3.9
	$T_{t,b}^{5'}$	3.7	4.9
	$T_{t,b}^{5''}$	3.7	4.7
	$T_{t,b}^{2'}$	2.6	3.1
	$T_{t,b}^{2''}$	2.6	2.4
	$T_{t,b}^5$	3.5	4.7
	C_q^6	2.5	2.4
	$C_q^{3'}$	4.0	6.6
	$C_q^{2'}$	3.4	4.4
C_q^5	$C_q^{2''}$	3.2	5.4
	$C_q^{5'}$	4.2	6.5
	$C_q^{5''}$	4.2	4.7

$G_q^{1'}$	G_q^8	3.0	3.9
	$G_q^{2'}$	2.4	3.0
	$G_q^{2''}$	2.4	2.2
	$G_q^{4'}$	2.8	2.7
	$Thr\,\gamma$	2.4	3.7
	$Nmv\ N-CH_3$	2.9	2.6
$A_{t,q}^{1'}$	$A_{t,q}^8$	2.7	3.9
	$A_{t,q}^2$	3.2	4.6
	$T_{t,q}^8$	2.9	3.2
	$A_{t,q}^{2'}$	2.1	3.0
	$A_{t,q}^{2''}$	2.1	2.4
	$A_{t,q}^{3'}$	3.3	3.8
	$A_{t,q}^{4'}$	2.8	3.4
	$A_{t,q}^{5'}$	2.6	4.9
	$A_{t,q}^{5''}$	2.6	4.7
$G_b^{1'}$	G_b^8	3.6	3.9
	$T_{t,b}^8$	2.8	7.7
	$H8$	2.8	4.2
	$H7$	3.6	5.1
	$G_b^{2'}$	2.6	2.3
	$G_b^{2''}$	2.6	2.8
	$G_b^{3'}$	4.0	3.8
	$G_b^{4'}$	3.7	3.1
	$G_b^{5'}$	3.1	5.1
	$G_b^{5''}$	3.1	4.6
	$Thr\,\gamma$	2.4	3.7
	$Nmv\ N-CH_3$	2.8	2.5

$T_{i,q}^{1'}$	$T_{i,q}^{6}$	3.0	3.7
	$T_{i,q}^{3'}$	2.8	3.8
	$T_{i,q}^{4'}$	2.7	2.4
	$T_{i,q}^{2'}$	2.5	3.0
	$T_{i,q}^{2''}$	2.5	2.4
	$Nmv\ N-CH_3$	2.7	5.0
$T_{i,b}^{1'}$	$T_{i,b}^{6}$	3.0	3.7
	$T_{i,b}^{3'}$	2.9	3.8
	$T_{i,b}^{4'}$	2.7	2.4
	$T_{i,b}^{2'}$	2.4	3.0
	$T_{i,b}^{2''}$	2.6	2.3
	$Nmv\ N-CH_3$	3.1	4.9
$C_q^{1'}$	C_q^{6}	3.0	3.8
	$C_q^{2'}$	2.6	3.1
	$C_q^{2''}$	2.6	2.4
	$C_q^{3'}$	2.9	3.8
	$C_q^{4'}$	2.9	3.4
	$C_q^{5'}$	3.2	4.9
	$C_q^{5''}$	3.2	4.1
	$Pro\,\beta_1$	3.0	2.7
	$Pro\,\beta_2$	2.8	2.5
	$Pro\,\gamma$	2.8	2.6

$A_{1,q}^{3'}$	$A_{1,q}^{8}$	3.2	4.4
	$A_{1,q}^{1'}$	3.0	3.8
	$A_{1,q}^{2'}$	2.2	2.4
	$A_{1,q}^{2''}$	2.2	2.7
	$A_{1,q}^{4'}$	2.6	2.7
	$A_{1,q}^{5'}$	2.3	2.4
	$A_{1,q}^{5''}$	2.3	3.6
$A_{1,b}^{3'}$	$A_{1,b}^{2'}$	2.4	2.4
	$A_{1,b}^{2''}$	2.4	2.7
	$A_{1,b}^{4'}$	2.6	2.7
	$A_{1,b}^{5'}$	2.3	2.3
	$A_{1,b}^{5''}$	2.3	3.6
$A_{t,q}^{3'}$	$A_{t,q}^{8}$	4.1	4.4
	$A_{t,q}^{1'}$	3.2	3.8
	$A_{t,q}^{2'}$	2.7	2.4
	$A_{t,q}^{2''}$	2.5	2.7
	$A_{t,q}^{4'}$	2.7	2.8
	$A_{t,q}^{5'}$	2.5	2.5
	$A_{t,q}^{5''}$	2.5	3.7

$T^{3'}_{t,a}$	$T^{0}_{t,a}$	2.8	3.9
	$T^{1'}_{t,a}$	4.0	3.8
	$T^{2'}_{t,a}$	3.0	2.3
	$T^{2''}_{t,a}$	2.7	2.9
	$T^{4'}_{t,a}$	2.5	3.0
$T^{3'}_{t,b}$	$T^{0}_{t,b}$	2.8	3.7
	$T^{1'}_{t,b}$	3.5	3.8
	$T^{2'}_{t,b}$	2.8	2.3
	$T^{2''}_{t,b}$	2.6	2.9
	$T^{4'}_{t,b}$	3.0	3.0
	$T^{5'}_{t,b}$	3.3	3.7
	$T^{5''}_{t,b}$	3.3	2.8
$A^{3'}_{t,a}$	$A^{0}_{t,a}$	4.1	4.4
	$A^{1'}_{t,a}$	3.2	3.8
	$A^{2'}_{t,a}$	2.7	2.4
	$A^{2''}_{t,a}$	2.5	2.7
	$A^{4'}_{t,a}$	2.7	2.8
	$A^{5'}_{t,a}$	2.5	2.5
	$A^{5''}_{t,a}$	2.5	3.7

$T^{3'}_{t,g}$	$T^{6}_{t,g}$	3.0	4.2
	$T^{1'}_{t,g}$	3.1	3.8
	$T^{2'}_{t,g}$	2.4	2.4
	$T^{2''}_{t,g}$	2.4	2.7
	$T^{4'}_{t,g}$	3.2	2.8
	$T^{5'}_{t,g}$	2.6	3.7
	$T^{5''}_{t,g}$	2.6	2.4
Thr α	*Thr* β	2.6	2.5
	Pro δ_1	3.2	2.2
	Thr γ	2.7	2.6
Thr β_1	*Thr* α	2.5	2.5
	Thr γ	2.3	2.4
	Pro δ	2.6	2.8
	Dua γ_1	2.4	2.3
Thr β_2	*Thr* α	2.4	2.4
	Thr γ	2.3	2.4
	Dua γ_1	2.5	2.7

Pro α	*Sar* α_1	2.5	2.1
	Sar α_2	2.6	3.6
	Dva α	2.4	2.3
	Pro β_1	2.3	2.3
	Dva β	3.2	4.9
	Pro β_2	2.9	2.7
	Pro γ	2.9	3.8
	Dva γ	3.7	4.7
Sar α_1	$A^2_{4',q}$	3.3	3.6
	Pro α	2.6	2.1
	Dva α	2.8	3.5
	NmvN$-CH_3$	2.4	2.2

Table A2.1: Comparison of 2D NMR NOE distances with moecular mechanics model distances (all distances in angstroms). Nomenclature is:

$A^2_{4',q}$ refers to the H2 proton of the inside adenine residue in the hexanuceoide strand on the quinoidal side of the AMD molecule.

$T^{1'}_{5',b}$ refers to the H1' proton of the terminal thymine residue in the hexanucleoside strand on the benzenoid side of the AMD molecule.

H7 refers to the H7 hydrogen on the chromophore.

Thr β_2 refers to the second beta proton of a threonine residue.

a - NMR NOE distances calculated from relative integrated peak intensities.

b - Interatomic distances from the molecular mechanics model structure. These values are for the mixed sugar pucker model. The uniform sugar model distances are quite similar.

Appendix 3: Source code

This appendix contains source code for the programs RIDMIN.FOR (rigid body minimization) and NPQUAD.FOR (nonproduct N–dimensional numerical integration for second virial coefficient calculations). Documentation and operating instructions are given in comment cards within the source code. Both programs are set up to run under the VAX/VMS 3.6 operating system.

```
C     RIDMIN.FOR
C
C     A PROGRAM FOR MINIMIZATION OF INTERACTION ENERGY OF
C     RIGID BODIES. WRITTEN OCTOBER, 1982 BY T. LYBRAND
C     USING A MINIMIZER BASED ON QUASI-NEWTON METHODS OF
C     FLETCHER, et. al.
C
      IMPLICIT REAL*8 (A-H,O-Z)
      COMMON
     ./PARMS/ RE(2Ø),EPS(2Ø),Q(2Ø),POL(2Ø)
     ./COORD/ X(15Ø,1Ø),Y(15Ø,1Ø),Z(15Ø,1Ø),ITYPE(15Ø,1Ø)
     ./BOOK/ NMOL,NAT(15Ø),RMS,ETB
     ./POLARZ/ ISCFP,MAXIT,NPOL,TOL,RCUT
     ./FLAGS/ IPRDIP,ISTAT
     ./ROT/ CONV,XI(15Ø,1Ø),YI(15Ø,1Ø),ZI(15Ø,1Ø)
C
      INTEGER NMOL,NVAR,NSIG,MAXFN,MAXIT,IOPT,IER,ISCF,IPRDIP
      INTEGER MAXITN,ISCFP,NPOL,IPRDER,IPRCOR,IPRPDB
      DIMENSION V(9ØØ),W(27ØØ),G(9ØØ),H(4Ø545Ø)
      DIMENSION RI(1Ø,1Ø)
      DIMENSION TITLE(2Ø)
      DIMENSION ALAB(15Ø,1Ø),MLAB(15Ø)
      CHARACTER*4 ALAB
      CHARACTER*3 MLAB
C     ********************************************************************************
C     ********************************************************************************
C                              INPUT DOCUMENTATION
C
C
C     NMOL = TOTAL NUMBER OF MOLECULES IN SYSTEM (DETERMINED
C            BY SUBROUTINE READIN AUTOMATICALLY).
C     NAT(I) = NUMBER OF ATOMS IN MOLECULE I (ALSO DETERMINED
C              BY SUBROUTINE READIN).
C     NVAR = NUMBER OF VARIABLES TO BE MINIMIZED (DETERMINED BY PROGRAM).
C
C     ********************************************************************************
C              USER SUPPLIED OPTIONS AND PARAMTERS (INPUT UNIT 5)
C
C     CARD 1    -      TITLE CARD
C
C                      TITLE (2ØA4)
C
C     ---------------------------------------------------------------------------
C     CARD 2    -      MINIMIZER CONTROL FLAGS
C
C                      NSIG,MAXFN,MAXITN,IOPT,ISTAT,IRST (6I5)
C
C     NSIG = CONVERGENCE CRITERION (NUMBER OF DIGITS OF ACCURACY FOR MINIMIZER).
C     MAXFN = MAXIMUM NUMBER OF ALLOWED FUNCTION EVALUATIONS (INCLUDING
C             FUNCTION EVALUATIONS NECESSARY FOR DETERMINING GRADIENTS).
C     MAXITN = MAXIMUM NUMBER OF MINIMIZATION CYCLES.
C     IOPT = MINIMIZATION OPTION FLAG
C            IOPT = Ø : MINM INITIALIZES THE HESSIAN TO IDENTITY MATRIX.
C            IOPT = 1 : USER INITIALIZES HESSIAN TO BE A
C                       POSITIVE DEFINITE MATRIX (NOT ALLOWED).
C            IOPT = 2 : MINM COMPUTES DIAGONAL VALUES OF THE
C                       HESSIAN AND SETS H TO A DIAGONAL MATRIX
C                       CONTAINING THESE VALUES.
C            IOPT = 3 : MINM ESTIMATES THE HESSIAN MATRIX.
C     ISTAT  = 1 : DO SINGLE POINT ENERGY EVALUATION WITH NO MINIMIZATION.
C            = Ø : DO FULL MINIMIZATION.
C     IRST   = 1 : RESTART CALCULATION USING INFORMATION IN THE RESTART FILE.
C            = Ø : NORMAL FRESH START. NO RESTART FILE IS READ. NOTE THAT
```

```
C                RIDMIN AUTOMATICALLY CREATES A RESTART FILE EACH TIME YOU
C                RUN THE PROGRAM. THE FILE IS NAMED RSTMIN.DAT (UNIT 8) AND CONTAINS
C                ALL INFORMATION NECESSARY FOR THE PROGRAM TO RESUME A
C                CALCULATION FROM THE LAST TERMINATION (I.E. COMPUTER CRASH, ETC.)
C                THE RESTART FILE FOR RESUMING A CALCULATION SHOULD BE NAMED
C                RST.DAT . ALTERNATELY, USE ASSIGN STATEMENTS TO ASSIGN YOUR
C                RESTART FILE TO RST.DAT (UNIT 7).
C
C     ------------------------------------------------------------------------
C     CARD 3   -      POLARIZATION CALCULATION CONTROL FLAGS
C
C                     ISCF,MAXIT,NPOL,TOL,RCUT (3I5,2F10.6)
C
C     ISCFP = 1 : DO SCF POLARIZATION CALCULATION.
C           = 0 : NO SCF CALCULATION DONE.
C     MAXIT = MAXIMUM NUMBER OF ALLOWED ITERATIONS IN SCF CALCULATION.
C     NPOL = SCF POLARIZATION CALCULATION IS DONE EVERY NPOL CYCLES OF
C            MINIMIZATION.
C     TOL = CONVERGENCE CRITERION FOR SCF POLARIZATION CALCULATION (EQUALS
C           THE MAXIMUM ALLOWED RMS DIFFERENCE BETWEEN INDUCED DIPOLES OF
C           ITERATION I AND ITERATION I-1).
C     RCUT = CUTOFF DISTANCE FOR SUPPRESSING POLARIZATION CALCULATION. ALL
C            ATOM PAIRS SHORTER THAN RCUT DISTANCE APART ARE NOT ALLOWED TO
C            POLARIZE EACH OTHER. RCUT SHOULD BE SET TO A VALUE LONGER THAN
C            THE BOND LENGTHS AND 1-3 DISTANCES IN YOUR SYSTEM.
C
C     ------------------------------------------------------------------------
C     CARD 4   -      PRINTOUT CONTROL FLAGS
C
C                     IPRDIP,IPRDER,IPRCOR,IPRPDB,IPRSTS,IPRDIS (6I5)
C
C     IPRDIP = 1 : PRINT OUT INDUCED DIPOLES EACH TIME INDIP IS CALLED.
C            = 0 : NO INDUCED DIPOLE PRINTOUT.
C     IPRDER = 1 : PRINT OUT VARIABLES AND DERIVATIVES EACH CYCLE.
C            = 0 : NO DERIVATIVE PRINTOUT
C     IPRCOR = 1 : PRINT OUT FINAL COORDINATES IN FORMATTED OUTPUT FILE (UNIT 6)
C            = 0 : NO FINAL COORDINATE PRINTOUT IN FORMATTED OUTPUT FILE.
C     IPRPDB = 1 : PRINT OUT FINAL COORDINATES IN PROTEIN DATA BANK FILE(UNIT 7)
C            = 0 : NO PROTEIN DATA BANK FILE GENERATED.
C     IPRSTS = 1 : PRINT OUT STATISTICS ON ATOM MOVEMENTS DURING MINIMIZATION
C            = 0 : NO BAR GRAPH PRINTOUT OF ATOM MOVEMENTS.
C     IPRDIS = 1 : PRINT OUT INTRA- AND INTERMOLECULAR ATOM-ATOM DISTANCES.
C            = 0 : NO ATOM-ATOM DISTANCES PRINTED.
C     ************************************************************************
C     NOTE ON PRINTOUT OPTIONS - A POSITIVE PRINTOUT FLAG (I.E. IPR*** = 1)
C     WILL GENERATE LARGE FORMATTED OUTPUT FILES. IT IS SUGGESTED THAT THESE
C     OPTIONS BE USED PRIMARILY FOR DIAGNOSTIC PURPOSES AND THAT THEY BE
C     SWITCHED OFF WHEN FULL MINIMIZATION IS DONE. THE PROTEIN DATA BANK FILE
C     SHOULD BE USED (I.E. IPRPDB = 1) IF FINAL COORDINATES ARE TO BE SAVED.
C     THE FORMAT OF THE PROTEIN DATA BANK FILE GENERATED BY THIS FLAG MAKES IT A
C     SUITABLE COORDINATE INPUT FILE IF ONE WISHES TO RESTART THE MINIMIZATION.
C     ************************************************************************
C
C     ------------------------------------------------------------------------
C     CARDS 5-24   -      POTENTIAL FUNCTION PARAMETERS
C
C                         I,RE,EPS,Q,POL (I5,4F10.6)
C
C     I : ATOM TYPE NUMBER
C     RE : ATOMIC RADIUS OF ATOM TYPE I.
C     EPS : VAN DER WAAL'S FUNCTION WELL DEPTH OF ATOM TYPE I.
C     Q : PARTIAL CHARGE OF ATOM TYPE I.
C     POL : ATOMIC POLARIZABILITY OF ATOM TYPE I.
C
```

```
C     CARDS 5-24 CORRESPOND TO THE ENERGY FUNCTION PARAMETERS WHICH ARE ASSIGNED
C     ACCORDING TO ATOM TYPE. THERE IS A MAXIMUM OF 20 ATOM TYPES ALLOWED IN
C     THE CURRENT VERSION OF THE PROGRAM.
C     ************************************************************************
C                          COORINATE INPUT (INPUT UNIT 2)
C
C               SYM,ALAB,MLAB,X,Y,Z,ITYPE (A4,8X,A4,1X,A3,10X,3F8.3,I2)
C
C     SYM : HEADER. IF SYM = 'ATOM', READ REST OF LINE.
C                   IF SYM = 'TER', START INPUT FOR NEW MOLECULE ON NEXT LINE.
C                   IF SYM = 'END', SIGNALS END OF COORDINATE INPUT.
C     ALAB : ATOM LABEL
C     MLAB : MOLECULE LABEL
C     X : X COORDINATE OF ATOM
C     Y : Y COORDINATE OF ATOM
C     Z : Z COORDINATE OF ATOM
C     ITYPE : ATOM TYPE
C
C     ************************************************************************
C     ************************************************************************
      CONV=(2.D0*DASIN(1.D0))/180.D0
C     Input file with coordinates
      open(unit=2,file='in1',status='old',readonly)
      rewind 2
C
C     Input file with energy potential
C     parameters for EPARM
      open(unit=5,file='in2',status='old',readonly)
      rewind 5
C
C     Output file (formatted)
      open(unit=6,file='iout1',status='new',shared,blocksize=132,
     .buffercount=1)
      rewind 6
C
C     INPUT STATEMENTS
C
      READ(5,9)(TITLE(I),I=1,20)
    9 FORMAT(20A4)
      READ(5,10)NSIG,MAXFN,MAXITN,IOPT,ISTAT,IRST
   10 FORMAT(6I5)
C
C     OPEN RESTART FILE IF CALCULATION IS BEING RESUMED
C
      IF(IRST.EQ.1) THEN
        OPEN(UNIT=7,FILE='RST',STATUS='OLD',FORM='UNFORMATTED')
        REWIND 7
      END IF
C
      READ(5,11)ISCFP,MAXIT,NPOL,TOL,RCUT
   11 FORMAT(3I5,2F10.4)
      READ(5,12)IPRDIP,IPRDER,IPRCOR,IPRPDB,IPRSTS,IPRDIS
   12 FORMAT(6I5)
      NN=0
   13 READ(5,14,END=15)(I,RE(I),EPS(I),Q(I),POL(I))
   14 FORMAT(I5,4F10.4)
      NN=NN+1
      GO TO 13
   15 CONTINUE
C
C     COORDINATE INPUT
C
      CALL READIN(ALAB,MLAB)
C
```

```
C      INITIALIZE EULER ANGLE VARIABLES.
C
       IF(IRST.NE.1) THEN
       NVAR=(NMOL-1)*6
       DO 1 I=1,NVAR
     1 V(I)=0.D0
       END IF
C
C      OUTPUT STATEMENTS
C
       WRITE(6,10000)(TITLE(I),I=1,15)
10000 FORMAT(1H ,10X,20A4)
       WRITE(6,9997)NN
 9997 FORMAT(//,1H ,20X,I2,1X,'ATOM TYPE PARAMETERS READ IN',/,
      .5X,'ATOM TYPE',12X,'RE',13X,'EPS',12X,'Q',13X,'POL',/)
       WRITE(6,9998)(I,RE(I),EPS(I),Q(I),POL(I),I=1,NN)
 9998 FORMAT(5X,I5,10X,F10.4,5X,F10.4,5X,F10.4,5X,F10.4)
C
C      BEGIN MINIMIZATION.
C
       CALL SECOND(T1)
       IF(ISTAT.EQ.0) THEN
C
C      OUTPUT FILE WITH RESTART INFORMATION (UNFORMATTED)
C
          OPEN(UNIT=8,FILE='RSTMIN',STATUS='NEW',FORM='UNFORMATTED')
          REWIND 8
C
          CALL MINM(NVAR,NSIG,MAXFN,IOPT,V,H,G,F,W,IER,IPRDER,MAXITN,
      .ITN,IRST)
          CALL MOVE(V)
       ELSE IF(ISTAT.EQ.1) THEN
          WRITE(6,150)
  150     FORMAT(/,1H ,5X,'********** SINGLE POINT ENERGY '
     1 'EVALUATION ******',/)
          CALL EPARM(V,F,0,1)
       END IF
       IF(IPRSTS.EQ.1) CALL STAT(ALAB,MLAB)
       IF(IPRDIS.EQ.1) THEN
C
C      INTRAMOLECULAR DISTANCES
C
       DO 25 I=1,NMOL
       NATI=NAT(I)
       DO 20 J=1,NATI
       DO 20 K=1,NATI
       XD=X(I,J)-X(I,K)
       YD=Y(I,J)-Y(I,K)
       ZD=Z(I,J)-Z(I,K)
   20 RI(J,K)=DSQRT(XD*XD+YD*YD+ZD*ZD)
       WRITE(6,100)I,MLAB(I)
  100 FORMAT(//,1H ,5X,'INTRAMOLECULAR DISTANCES FOR MOLECULE',
     1I5,2X,A3,/)
       DO 30 M=1,NATI,10
       N=M+9
       IF(N.LE.NATI) GO TO 32
   31 N=NATI
   32 WRITE(6,101)(ALAB(I,II),II=M,N)
  101 FORMAT(5X,10(3X,A4))
       DO 30 III=1,NATI
       WRITE(6,102)ALAB(I,III),(RI(III,JJJ),JJJ=M,N)
  102 FORMAT(1X,A4,10(F7.3))
   30 CONTINUE
   25 CONTINUE
```

```
      WRITE(6,110)
  110 FORMAT(//,1H ,10X,'INTERMOLECULAR DISTANCES',/)
      DO 35 I=1,NMOL
      NATI=NAT(I)
      DO 35 K=1,NATI
      II=I+1
      DO 35 J=II,NMOL
      NATJ=NAT(J)
      DO 35 L=1,NATJ
      XD=X(I,K)-X(J,L)
      YD=Y(I,K)-Y(J,L)
      ZD=Z(I,K)-Z(J,L)
      DIS=DSQRT(XD*XD+YD*YD+ZD*ZD)
      WRITE(6,120)I,ALAB(I,K),J,ALAB(J,L),DIS
  120 FORMAT(1X,'MOLECULE ',I3,2X,'ATOM ',A4,3X,'MOLECULE ',
     .I3,2X,'ATOM ',A4,' DISTANCE IS ',F6.2,' ANGSTROMS',/)
   35 CONTINUE
      END IF
      CALL SECOND(T2)
      TIME = T2-T1
C
C     PROTEIN DATA BANK FILE WITH FINAL COORDINATES.
C
      IF(IPRPDB.EQ.1) THEN
         open(unit=7,file='fout2',status='new')
         rewind 7
         CALL OUTPUT(ALAB,NLAB)
      END IF
C
C
C     FORMATTED OUTPUT FILE (SEE ASSORTED PRINT FLAGS ABOVE)
C
      IF(ISCFP.EQ.1) WRITE(6,9999)NPOL
 9999 FORMAT(/,1H ,5X,'SCF POLARIZATION CALCULATION EVERY',I5,
     .   ' CYCLES',/)
      WRITE(6,10001)NMOL
10001 FORMAT(///,1H ,5X,'NUMBER OF MOLECULES IS',I5,/)
      IF(IPRCOR.EQ.1) THEN
         DO 3 I=1,NMOL
         NATI=NAT(I)
         WRITE(6,10002)I,NATI
10002    FORMAT(1H ,5X,'NUMBER OF ATOMS IN MOLECULE',I5,1X,
     .                'IS',I5,/)
         WRITE(6,10003)I
10003    FORMAT(1H ,5X,'FINAL COORDINATES OF MOLECULE',I5,1X,
     .      'ARE',/)
         WRITE(6,10004)
10004    FORMAT(1H ,11X,'X',14X,'Y',14X,'Z',14X,'Q',5X,'TYPE',/)
         DO 2 K=1,NATI
         INDEX=ITYPE(I,K)
         QIK=Q(INDEX)
         WRITE(6,10005)K,X(I,K),Y(I,K),Z(I,K),QIK,ITYPE(I,K)
10005    FORMAT(1H ,I3,'.',2X,F10.6,5X,F10.6,5X,F10.6,5X,F10.6,2X,I2)
    2    CONTINUE
         WRITE(6,10006)
10006    FORMAT(//)
    3    CONTINUE
      END IF
      IF(IER.NE.0) WRITE(6,10007)
10007 FORMAT(/,1H ,5X,'****************** PREMATURE TERMINATION!
     .******************',/)
      IF(IER.EQ.129) WRITE(6,10008)
10008 FORMAT(/,1H ,5X,'************ HESSIAN IS NOT POSITIVE
     . DEFINITE! ************',/)
```

```
      IF(IER.EQ.13Ø) WRITE(6,1ØØØ9)
1ØØØ9 FORMAT(/,1H ,5X,'********** ROUNDING ERRORS HAVE BECOME
     . EXCESSIVE! **********',/)
      IF(IER.EQ.131) WRITE(6,1ØØ1Ø)
1ØØ1Ø FORMAT(/,1H ,5X,'**** MAXIMUM NUMBER OF FUNCTION EVALUATIONS
     .EXCEEDED! ****',/)
      IF(IER.EQ.132) WRITE(6,1ØØ11)
1ØØ11 FORMAT(/,1H ,5X,'**** MAXIMUM NUMBER OF MINIMIZATION CYCLES
     .EXCEEDED! *****',/)
      WRITE(6,1ØØ12)
1ØØ12 FORMAT(///,1H ,2ØX,'FINAL RESULTS ',/)
      WRITE(6,1ØØ13)
1ØØ13 FORMAT(3X,'CYCLE',7X,'TOTAL',7X,'ELECTROSTATIC',6X,'ATOT',7X,
     .'BTOT',6X,'POLARIZATION',5X,'DISPERSION',5X,'THREE-BODY',/)
      CALL EPARM(V,E,ITN,1)
      WRITE(6,1ØØ14)(W(I),I=1,2)
1ØØ14 FORMAT(/,1H ,6X,'RMS DERIVATIVE =',D2Ø.8,/,1H ,
     .       5X,'NUMBER OF FUNCTION EVALUATIONS =',F7.Ø)
      WRITE(6,1ØØ15)TIME
1ØØ15 FORMAT(1H ,5X,'TOTAL TIME =',F8.3,' SECONDS',//)
C
      STOP
      END
      SUBROUTINE EPARM(V,E,ITN,IPRCOM)
C
C     THIS SUBROUTINE CALCULATES THE INTERACTION ENERGY OF THE
C     SYSTEM.
C
      IMPLICIT REAL*8(A-H,O-Z)
      COMMON
     ./PARMS/ RE(2Ø),EPS(2Ø),Q(2Ø),POL(2Ø)
     ./COORD/ X(15Ø,1Ø),Y(15Ø,1Ø),Z(15Ø,1Ø),ITYPE(15Ø,1Ø),
     ./BOOK/ NMOL,NAT(15Ø),RMS,ETB
     ./POLARZ/ ISCFP,MAXIT,NPOL,TOL,RCUT
     ./FLAGS/ IPRDIP,ISTAT
     ./ROT/ CONV,XI(15Ø,1Ø),YI(15Ø,1Ø),ZI(15Ø,1Ø)
C
      DIMENSION V(9ØØ)
C
      IF(ISTAT.EQ.Ø) CALL MOVE(V)
C
C     DO ADDITIVE ENERGY TERMS FIRST OVER ALL ATOM PAIRS (INTERMOLECULAR
C     INTERACTIONS ONLY)
C        E = BTOT/R**12 - ATOT/R**6 + QTOT/R
C
      QTOT=Ø.DØ
      ATOT=Ø.DØ
      BTOT=Ø.DØ
      PTOT=Ø.DØ
      D2=Ø.DØ
      ETB=Ø.DØ
      I=1
   1Ø CONTINUE
      NATI=NAT(I)
      DO 2Ø K=1,NATI
      INDEX=ITYPE(I,K)
      QA=Q(INDEX)
      REA=RE(INDEX)
      EPSA=EPS(INDEX)
C
C     ASSIGN PARAMETERS BY ATOM TYPE
C
C     SECOND MOLECULE
      JJ=I+1
```

```
      DO 3Ø J=JJ,NMOL
      NATJ=NAT(J)
      DO 4Ø L=1,NATJ
      JNDEX=ITYPE(J,L)
      QB=Q(JNDEX)
      REB=RE(JNDEX)
      EPSB=EPS(JNDEX)
C
C     ATOM-ATOM DISTANCE
C
      XD=X(I,K)-X(J,L)
      YD=Y(I,K)-Y(J,L)
      ZD=Z(I,K)-Z(J,L)
      DENOM=DSQRT(XD*XD+YD*YD+ZD*ZD)
      DENOM6=DENOM*DENOM*DENOM*DENOM*DENOM*DENOM
      DEN12=DENOM6*DENOM6
C
C     ELECTROSTATIC INTERACTIONS
C
      QTOT = QTOT + (331.62DØ*QA*QB)/DENOM
C
C     ************************************************************
C     RWK2 POTENTIAL FUNCTION FOR WATER-WATER INTERACTIONS
C
C     DISPERSION AND O-O REPULSION
C
      IF(INDEX.EQ.1.AND.JNDEX.EQ.1) THEN
        AOO=4Ø5ØØØØ.DØ
        BOO=-4.97Ø19DØ
        RSC=DENOM*Ø.94834673DØ
        RAU=RSC/Ø.52917715DØ
        RSQ=RAU*RAU
        F=1.DØ-((RAU**2.326DØ)*DEXP(-RAU))
        G6=1.DØ-DEXP(-Ø.35ØØØDØ*RAU-Ø.Ø44499DØ*RSQ)
        G8=1.DØ-DEXP(-Ø.2625ØDØ*RAU-Ø.Ø38537DØ*RSQ)
        G1Ø=1.DØ-DEXP(-Ø.21ØØØDØ*RAU-Ø.Ø34469DØ*RSQ)
        D2=D2-(F*(625.45DØ*(G6/RSC)**6 + 339Ø.DØ*(G8/RSC)**8
     1  + 31ØØØ.DØ*(G1Ø/RSC)**1Ø))
        BTOT=BTOT+AOO*DEXP(BOO*DENOM)
C
C     H-H REPULSION
C
      ELSE IF(INDEX.EQ.2.AND.JNDEX.EQ.2) THEN
        AHH=631.918DØ
        BHH=-3.28Ø59DØ
        BTOT=BTOT+AHH*DEXP(BHH*DENOM)
C
C     MORSE POTENTIAL FOR O-H INTERACTIONS
C
      ELSE IF(INDEX.EQ.1.AND.JNDEX.EQ.2.OR.INDEX.EQ.2.AND.JNDEX.
     1EQ.1) THEN
        AOH=1.3DØ
        BOH=-7.36154DØ
        RMIN=1.6378ØØDØ
        BX=DEXP(BOH*(DENOM-RMIN))
        ATOT=ATOT+AOH*BX*(BX-2.DØ)
      ELSE IF(INDEX.EQ.3.AND.JNDEX.NE.4) THEN
        GO TO 4Ø
      ELSE IF(INDEX.NE.4.AND.JNDEX.EQ.3) THEN
        GO TO 4Ø
      ELSE
        EPSX = DSQRT(EPSA*EPSB)
        RMIN = REA+REB
        AC = 2.DØ*EPSX*(RMIN**6)
```

```
        BC = EPSX*(RMIN**12)
        ATOT = ATOT - AC/DENOM6
        BTOT = BTOT + BC/DEN12
      END IF
C     ************************************************************
   40 CONTINUE
   30 CONTINUE
      IF(INDEX.GE.4) CALL THREEB(I,K,INDEX)
   20 CONTINUE
      I=I+1
      IF(I.LT.NMOL) GO TO 10
      ETBSAVE=ETB
C
C     DO POLARIZATION CALCULATION
C
      CALL POLAR(PTOT,ITN)
      VDW=ATOT+BTOT
      E=QTOT+ATOT+BTOT+PTOT+D2+ETBSAVE
      IF(IPRCOM.EQ.1) THEN
         write(6,102)itn,e,qtot,atot,btot,ptot,D2,etbsave
  102    format(1h ,I5,7(6x,f8.3),/)
      END IF
      RETURN
      END
      SUBROUTINE POLAR(PTOT,ITN)
C     THIS SUBROUTINE CALCULATES THE POLARIZATION ENERGY
C     OT THE SYSTEM, WITH THE OPTION OF AN SCF CALCULATION
C     INCLUDING PARTIAL CHARGES AND INDUCED DIPOLES IN THE
C     ELECTROSTATIC FIELD.
C
      IMPLICIT REAL*8(A-H,O-Z)
      COMMON
     ./PARMS/ RE(20),EPS(20),Q(20),POL(20)
     ./COORD/ X(150,10),Y(150,10),Z(150,10),ITYPE(150,10)
     ./BOOK/ NMOL,NAT(150),RMS,ETB
     ./POLARZ/ ISCFP,MAXIT,NPOL,TOL,RCUT
     ./FLAGS/ IPRDIP,ISTAT
     ./ROT/ CONV,XI(150,10),YI(150,10),ZI(150,10)
C
      DIMENSION ECX(150,10),ECY(150,10),ECZ(150,10)
      DIMENSION EDX(150,10),EDY(150,10),EDZ(150,10)
      DIMENSION ETX(150,10),ETY(150,10),ETZ(150,10)
      DIMENSION UX(150,10),UY(150,10),UZ(150,10)
C
      ZERO=0.D0
      PTOT1=0.D0
      PTOT2=0.D0
      PTOT3=0.D0
C
C     CALCULATE ELECTROSTATIC FIELD.
C
      CALL CFIELD(ECX,ECY,ECZ)
      CALL INDIP(ECX,ECY,ECZ,UX,UY,UZ,0,ISCF)
C
C     SCF LOOP (IF ISCF=1).
C
      NP=MOD(ITN,NPOL)
      IF(NP.EQ.0.AND.ISCFP.EQ.1.AND.ITN.GT.1) THEN
         ISCF=1
      ELSE IF(ISCFP.EQ.1.AND.ITN.EQ.1) THEN
         ISCF=1
      ELSE IF(NP.NE.0.OR.ISCFP.NE.1) THEN
         ISCF=0
      ELSE IF(ISTAT.EQ.1) THEN
```

```
         ISCF=ISCFP
      END IF
      ITER=0
      IF(ISCF.NE.1) GO TO 20
   10 CONTINUE
      ITER=ITER+1
      IF(ITER.GT.MAXIT) WRITE(6,2)rms
    2 FORMAT(1H ,5X,'NUMBER OF ALLOWED ITERATIONS IN SCF
     . CALCULATION EXCEEDED',/,6X,'AT SCF TERMINATION RMS =',E20.7,/)
      IF(ITER.GT.MAXIT) GO TO 40
   20 CALL DFIELD(EDX,EDY,EDZ,UX,UY,UZ)
      DO 30 I=1,NMOL
      NATI=NAT(I)
      DO 30 K=1,NATI
      ETX(I,K)=ECX(I,K)+EDX(I,K)
      ETY(I,K)=ECY(I,K)+EDY(I,K)
   30 ETZ(I,K)=ECZ(I,K)+EDZ(I,K)
      CALL INDIP(ETX,ETY,ETZ,UX,UY,UZ,ITER,ISCF)
      IF(ISCF.NE.1) GO TO 40
      IF(RMS.GT.TOL) GO TO 10
C
C     END OF SCF LOOP
C
   40 CONTINUE
C
C     CALCULATE POLARIZATION ENERGY DUE TO FINAL FIELD
C
      DO 50 I=1,NMOL
      NATI=NAT(I)
      DO 50 K=1,NATI
      NDEX=ITYPE(I,K)
      ALP=POL(NDEX)
      IF(ALP.EQ.ZERO) GO TO 50
      EX = ETX(I,K)
      EY = ETY(I,K)
      EZ = ETZ(I,K)
      E2 = (EX*EX)+(EY*EY)+(EZ*EZ)
      PTOT3 = PTOT3 + (331.62D0*(0.5D0*ALP*(E2)))
   50 CONTINUE
      PTOT = -PTOT3
      RETURN
      END
      SUBROUTINE CFIELD(ECX,ECY,ECZ)
C     THIS SUBROUTINE CALCULATES THE ELECTROSTATIC FIELD
C     DUE TO THE PARTIAL CHARGES.
C
      IMPLICIT REAL*8(A-H,O-Z)
      COMMON
     ./PARMS/ RE(20),EPS(20),Q(20),POL(20)
     ./COORD/ X(150,10),Y(150,10),Z(150,10),ITYPE(150,10)
     ./BOOK/ NMOL,NAT(150),RMS,ETB
     ./POLARZ/ ISCFP,MAXIT,NPOL,TOL,RCUT
     ./FLAGS/ IPRDIP,ISTAT
     ./ROT/ CONV,XI(150,10),YI(150,10),ZI(150,10)
C
      DIMENSION ECX(150,10),ECY(150,10),ECZ(150,10)
C
      ZERO=0.D0
      DO 10 I=1,NMOL
      NATI=NAT(I)
      DO 10 K=1,NATI
      ECX(I,K)=0.D0
      ECY(I,K)=0.D0
      ECZ(I,K)=0.D0
```

```
      INDEX=ITYPE(I,K)
      IF(POL(INDEX).EQ.ZERO) GO TO 10
      DO 10 J=1,NMOL
      IF(J.EQ.I) GO TO 20
      NATJ=NAT(J)
      DO 10 L=1,NATJ
      JNDEX=ITYPE(J,L)
      QQ=Q(JNDEX)
      XX=X(I,K)-X(J,L)
      YY=Y(I,K)-Y(J,L)
      ZZ=Z(I,K)-Z(J,L)
      DENOM=DSQRT(XX*XX+YY*YY+ZZ*ZZ)
      DENOM3=DENOM*DENOM*DENOM
      ECX(I,K) = ECX(I,K) + ((XX*QQ)/DENOM3)
      ECY(I,K) = ECY(I,K) + ((YY*QQ)/DENOM3)
      ECZ(I,K) = ECZ(I,K) + ((ZZ*QQ)/DENOM3)
   20 CONTINUE
   10 CONTINUE
      RETURN
      END
      SUBROUTINE DFIELD(EDX,EDY,EDZ,UX,UY,UZ)
C     THIS SUBROUTINE CALCULATES THE PORTION OF THE
C     ELECTROSTATIC FIELD DUE TO INDUCED DIPOLES.
C
      IMPLICIT REAL*8(A-H,O-Z)
      COMMON
     ./PARMS/ RE(20),EPS(20),Q(20),POL(20)
     ./COORD/ X(150,10),Y(150,10),Z(150,10),ITYPE(150,10),
     ./BOOK/ NMOL,NAT(150),RMS,ETB
     ./POLARZ/ ISCFP,MAXIT,NPOL,TOL,RCUT
     ./FLAGS/ IPRDIP,ISTAT
     ./ROT/ CONV,XI(150,10),YI(150,10),ZI(150,10)
C
      DIMENSION EDX(150,10),EDY(150,10),EDZ(150,10)
      DIMENSION UX(150,10),UY(150,10),UZ(150,10)
C
      ZERO=0.D0
      DO 10 I=1,NMOL
      NATI=NAT(I)
      DO 10 K=1,NATI
      EDX(I,K)=0.D0
      EDY(I,K)=0.D0
      EDZ(I,K)=0.D0
      INDEX=ITYPE(I,K)
      IF(POL(INDEX).EQ.ZERO) GO TO 10
      DO 10 J=1,NMOL
      NATJ=NAT(J)
      DO 10 L=1,NATJ
C
C     DO NOT ALLOW ATOMS TO POLARIZE THEMSELVES.
C
      IF(J.EQ.I.AND.L.EQ.K) GO TO 20
      JNDEX=ITYPE(J,L)
      XX=X(I,K)-X(J,L)
      YY=Y(I,K)-Y(J,L)
      ZZ=Z(I,K)-Z(J,L)
      R=DSQRT(XX*XX+YY*YY+ZZ*ZZ)
      R2=R*R
      DENOM5=R2*R2*R
      IF(R.LT.RCUT.AND.I.EQ.J) THEN
      GO TO 20
      ELSE IF(R.GE.RCUT.OR.I.NE.J) THEN
      EDX(I,K) = EDX(I,K) + (((3.D0*XX*XX)-R2)*UX(J,L)
     .                    + (3.D0*XX*YY)*UY(J,L)
```

```
     .                       + (3.DØ*XX*ZZ)*UZ(J,L))/DENOM5
      EDY(I,K) = EDY(I,K) + ((3.DØ*XX*YY)*UX(J,L)
     .                       + ((3.DØ*YY*YY)-R2)*UY(J,L)
     .                       + (3.DØ*YY*ZZ)*UZ(J,L))/DENOM5
      EDZ(I,K) = EDZ(I,K) + ((3.DØ*XX*ZZ)*UX(J,L)
     .                       + (3.DØ*YY*ZZ)*UY(J,L)
     .                       + ((3.DØ*ZZ*ZZ)-R2)*UZ(J,L))/DENOM5
      END IF
   2Ø CONTINUE
   1Ø CONTINUE
      RETURN
      END
      SUBROUTINE INDIP(EX,EY,EZ,UX,UY,UZ,ITER,ISCF)
C     THIS SUBROUTINE CALCULATES DIPOLES INDUCED BY THE
C     ELECTROSTATIC FIELD.
C
      IMPLICIT REAL*8(A-H,O-Z)
      COMMON
     ./PARMS/ RE(2Ø),EPS(2Ø),Q(2Ø),POL(2Ø)
     ./COORD/ X(15Ø,1Ø),Y(15Ø,1Ø),Z(15Ø,1Ø),ITYPE(15Ø,1Ø)
     ./BOOK/ NMOL,NAT(15Ø),RMS,ETB
     ./POLARZ/ ISCFP,MAXIT,NPOL,TOL,RCUT
     ./FLAGS/ IPRDIP,ISTAT
     ./ROT/ CONV,XI(15Ø,1Ø),YI(15Ø,1Ø),ZI(15Ø,1Ø)
C
      DIMENSION EX(15Ø,1Ø),EY(15Ø,1Ø),EZ(15Ø,1Ø)
      DIMENSION UX(15Ø,1Ø),UY(15Ø,1Ø),UZ(15Ø,1Ø)
      DIMENSION UIN(15Ø,1Ø),UINOLD(15Ø,1Ø)
C
      ZERO=Ø.DØ
      IATOT=Ø
      RMSU=Ø.DØ
      DO 1Ø I=1,NMOL
      NATI=NAT(I)
      IATOT=IATOT+NATI
      UXX=ZERO
      UYY=ZERO
      UZZ=ZERO
      DO 15 K=1,NATI
      IF(ITER.EQ.1) UIN(I,K)=Ø.DØ
      UINOLD(I,K)=UIN(I,K)
      NDEX=ITYPE(I,K)
      ALP=POL(NDEX)
      IF(ALP.EQ.ZERO) GO TO 15
      UX(I,K) = ALP*EX(I,K)
      UY(I,K) = ALP*EY(I,K)
      UZ(I,K) = ALP*EZ(I,K)
      UIN(I,K)=DSQRT(UX(I,K)**2 + UY(I,K)**2 + UZ(I,K)**2)*4.8DØ
C
C     COMPARE OLD AND NEW INDUCED DIPOLES.
C
      DIFF=UIN(I,K)-UINOLD(I,K)
      RMSU=RMSU+(DIFF*DIFF)
      UXX = UXX + UX(I,K)
      UYY = UYY + UY(I,K)
      UZZ = UZZ + UZ(I,K)
   15 CONTINUE
C
C     PRINTOUT FOR INDUCED DIPOLES
C
      U = DSQRT(UXX*UXX + UYY*UYY + UZZ*UZZ)*4.8DØ
      IF(IPRDIP.EQ.1) THEN
         WRITE(6,1ØØ)I,U
  1ØØ    FORMAT(1h ,5x,'INDUCED DIPOLE ON MOLECULE
```

```
     .    ',14,1x,'IS',f10.2,1x,'DEBYE',/)
          END IF
       10 CONTINUE
          IF(ISCF.NE.1) GO TO 20
          ATOMN=FLOAT(IATOT)
          RMS=DSQRT(RMSU/ATOMN)
       20 CONTINUE
          RETURN
          END
          SUBROUTINE MOVE(V)
C         THIS SUBROUTINE TRANSLATES AND ROTATES MOLECULES
C         TWO THRU NMOL AS DIRECTED BY THE MINIMIZER.
C
          IMPLICIT REAL*8 (A-H,O-Z)
          COMMON
         ./PARMS/ RE(20),EPS(20),Q(20),POL(20)
         ./COORD/ X(150,10),Y(150,10),Z(150,10),ITYPE(150,10)
         ./BOOK/ NMOL,NAT(150),RMS,ETB
         ./POLARZ/ ISCFP,MAXIT,NPOL,TOL,RCUT
         ./FLAGS/ IPRDIP,ISTAT
         ./ROT/ CONV,XI(150,10),YI(150,10),ZI(150,10)
C
          DIMENSION V(900),VE(150,6)
C
C         PUT EULER VARIABLES IN A 2-DIMENSIONAL ARRAY INDEXED
C         TO CORRESPOND TO THE MOLECULE COORDINATE INDICES.
C
          NVAR=(NMOL-1)*6
          J=1
          K=0
          DO 1 I=1,NVAR
          K=K+1
          VE(J,K)=V(I)
          NR=MOD(I,6)
          IF(NR.EQ.0) J=J+1
          IF(NR.EQ.6) K=0
        1 CONTINUE
C
C         TRANSLATION AND ROTATION OF MOLECULE I. NOTE THAT
C         THE INDEX I STARTS AT 2 AS MOLECULE ONE IS HELD
C         FIXED AT ITS INITIAL POSITION FOR GREATER COMPUTATIONAL
C         EFFICIENCY.
C
          DO 10 I=2,NMOL
          J=I-1
          REPHI=VE(J,1)*0.1D0
          RETHET=VE(J,2)*0.1D0
          REPSI=VE(J,3)*0.1D0
          RS=VE(J,4)*0.1D0
          RTHETA=VE(J,5)*0.1D0
          RPHI=VE(J,6)*0.1D0
          XO=RS*DSIN(RTHETA)*DCOS(RPHI)
          YO=RS*DSIN(RTHETA)*DSIN(RPHI)
          ZO=RS*DCOS(RTHETA)
C
C         THE ROTATION MATRIX ELEMENTS FOR MOVEMENT OF MOLECULES
C         IN EULER SPACE ARE BASED ON THE CONVENTION OF GOLDSTEIN
C         IN "CLASSICAL MECHANICS", (1965), P.109 .
C
          A11=DCOS(REPSI)*DCOS(REPHI)-DCOS(RETHET)*DSIN(REPHI)*DSIN(REPSI)
          A12=DCOS(REPSI)*DSIN(REPHI)+DCOS(RETHET)*DCOS(REPHI)*DSIN(REPSI)
          A13=DSIN(REPSI)*DSIN(RETHET)
          A21=-DSIN(REPSI)*DCOS(REPHI)-DCOS(RETHET)*DSIN(REPHI)*DCOS(REPSI)
          A22=-DSIN(REPSI)*DSIN(REPHI)+DCOS(RETHET)*DCOS(REPHI)*DCOS(REPSI)
```

```
      A23=DCOS(REPSI)*DSIN(RETHET)
      A31=DSIN(RETHET)*DSIN(REPHI)
      A32=-DSIN(RETHET)*DCOS(REPHI)
      A33=DCOS(RETHET)
      NATI=NAT(I)
      DO 20 K=1,NATI
      X(I,K)=XI(I,K)*A11+YI(I,K)*A12+ZI(I,K)*A13+XO
      Y(I,K)=XI(I,K)*A21+YI(I,K)*A22+ZI(I,K)*A23+YO
   20 Z(I,K)=XI(I,K)*A31+YI(I,K)*A32+ZI(I,K)*A33+ZO
   10 CONTINUE
      RETURN
      END
      SUBROUTINE MINM(N,NSIG,MAXFN,IOPT,X,H,G,F,W,IER,IPRDER,MAXITN,
     .ITN,IRST)
C
      INTEGER             N,NSIG,MAXFN,IOPT,IER,MAXITN,IPRDER
      DOUBLE PRECISION    X(900),G(900),H(1),F,W(1)
      INTEGER             IG,IGG,IS,IDIFF,IR,IJ,I,J,NM1,JJ,JP1,L,KJ,K,
     .                    IFN,LINK,ITN,II,IM1,JNT,NP1,JB,NJ
      DOUBLE PRECISION    REPS,AX,ZERO,ONE,HALF,SEVEN,FIVE,TWELVE,TEN,HH,
     .                    EPS,HJJ,V,DF,RELX,GS0,DIFF,AEPS,ALPHA,FF,TOT,
     .                    F1,F2,Z,GYS,DGS,SIG,ZZ,GNRM,P1,HHH,GHH,H2,F11,
     .                    F12,F21,F22,HMAX,HMIN,E,RMS
      DATA                REPS/1.0D-20/,AX/0.1D0/
      DATA                ZERO/0.0D0/,ONE/1.0D0/,HALF/0.5D0/,
     .                    SEVEN/7.0D0/,FIVE/5.0D0/,TWELVE/12.0D0/,
     .                    TEN/10.0D0/,P1/0.1D0/
C
C     INPUT FOR RESTART INFORMATION IF IRST = 1
C
      IF(IRST.EQ.1) THEN
         REWIND 7
         READ(7)N,NSIG,IOPT,IPRDER
         READ(7)F,(X(I),I=1,N),(G(I),I=1,N),(W(I),I=1,3*N)
         NN2=N*(N+1)/2
         READ(7)(W(I),I=1,NN2)
         READ(7)IG,IGG,IS,IDIFF,IR,IJ,I,J,NM1,JJ,JP1,L,KJ,K,
     .          IFN,LINK,ITN,II,IM1,JNT,NP1,JB,NJ
         READ(7)HH,EPS,HJJ,V,DF,RELX,GSO,DIFF,AEPS,ALPHA,FF,
     .          TOT,F1,F2,Z,GYS,DGS,SIG,ZZ,GNRM,HHH,GHH,H2,F11,
     .          F12,F21,F22,HMAX,HMIN,E,RMS
         GO TO 105
      END IF
C
C   INITIALIZATION OF CONTROL FLAGS AND VARIABLES
C
      IER = 0
      HH = DSQRT(REPS)
      H2 = DSQRT(HH)
      EPS = TEN**(-NSIG)
      IG = N
      IGG = N+N
      IS = IGG
      IDIFF = 1
      IR = N
      W(1) = -ONE
      W(2) = ZERO
      W(3) = ZERO
C
C     EVALUATE FUNCTION AT STARTING POINT
C
      DO 5 I=1,N
      G(I) = X(I)
    5 CONTINUE
```

```
      CALL EPARM(G,F,ITN,0)
      IFN = 1
      IF (IOPT.EQ.1) GO TO 50
C
C     SET OFF-DIAGONAL ELEMENTS OF H TO 0.0
C
      IF (N.EQ.1) GO TO 20
      IJ = 2
      DO 15 I=2,N
         DO 10 J=2,I
            H(IJ) = ZERO
            IJ = IJ+1
   10    CONTINUE
         IJ = IJ+1
   15 CONTINUE
   20 IF (IOPT.NE.0) GO TO 30
C
C     SET DIAGONAL ELEMENTS OF H TO ONE
C
      IJ = 0
      DO 25 I=1,N
         IJ = IJ+I
         H(IJ) = ONE
   25 CONTINUE
      GO TO 95
C
C     GET DIAGONAL ELEMENTS OF HESSIAN
C
   30 IM1 = 1
      NM1 = 1
      NP1 = N+1
      DO 35 I=2,NP1
         HHH = H2*DMAX1(DABS(X(IM1)),AX)
         G(IM1) = X(IM1)+HHH
         CALL EPARM(G,F2,ITN,0)
         G(IM1) = X(IM1)-HHH
         CALL EPARM(G,FF,ITN,0)
         H(NM1) = (FF-F+F2-F)/(HHH*HHH)
         G(IM1) = X(IM1)
         IM1 = I
         NM1 = I+NM1
   35 CONTINUE
      IFN = IFN+N+N
      IF (IOPT.NE.3 .OR. N.EQ.1) GO TO 50
C
C     GET THE REST OF THE HESSIAN
C
      JJ = 1
      II = 2
      DO 45 I=2,N
         GHH = H2*DMAX1(DABS(X(I)),AX)
         DO 40 J=1,JJ
            HHH = H2*DMAX1(DABS(X(J)),AX)
            G(I) = X(I)+GHH
            G(J) = X(J)+HHH
            CALL EPARM(G,F22,ITN,0)
            G(I) = X(I)-GHH
            CALL EPARM(G,F12,ITN,0)
            G(J) = X(J)-HHH
            CALL EPARM(G,F11,ITN,0)
            G(I) = X(I)+GHH
            CALL EPARM(G,F21,ITN,0)
            H(II) = (F22-F21-F12+F11)/(4.D0*HHH*GHH)
            G(J) = X(J)
```

```
            II = II+1
   4Ø    CONTINUE
         G(I) = X(I)
         JJ = JJ+1
         II = II+1
   45 CONTINUE
      IFN = IFN+((N*N-N)*2)
C
C     ADD MULTIPLE OF IDENTITY TO MAKE DIAGONAL ELEMENTS POSITIVE
C
   5Ø HMIN = H(1)
      HMAX = H(1)
      NM1 = 1
      DO 55 I=1,N
         HMIN = DMIN1(HMIN,H(NM1))
         HMAX = DMAX1(HMAX,H(NM1))
         NM1 = NM1+I+1
   55 CONTINUE
      HMIN = DMAX1(Ø.Ø1DØ*(DABS(HMAX)+DABS(HMIN))-HMIN,Ø.ØDØ)
      NM1 = 1
      DO 6Ø I=1,N
         H(NM1) = H(NM1)+HMIN
         NM1 = NM1+I+1
   6Ø CONTINUE
C
C     FACTOR H TO L*D*L-TRANSPOSE
C
      IR = N
      IF (N.GT.1) GO TO 65
      IF (H(1).GT.ZERO) GO TO 95
      H(1) = ZERO
      IR = Ø
      GO TO 9Ø
   65 NM1 = N-1
      JJ = Ø
      DO 85 J=1,N
         JP1 = J+1
         JJ = JJ+J
         HJJ = H(JJ)
         IF (HJJ.GT.ZERO) GO TO 7Ø
         H(JJ) = ZERO
         IR = IR-1
         GO TO 85
   7Ø    IF (J.EQ.N) GO TO 85
         IJ = JJ
         L = Ø
         DO 8Ø I=JP1,N
            L = L+1
            IJ = IJ+I-1
            V = H(IJ)/HJJ
            KJ = IJ
            DO 75 K=I,N
               H(KJ+L) = H(KJ+L)-H(KJ)*V
               KJ = KJ+K
   75       CONTINUE
            H(IJ) = V
   8Ø    CONTINUE
   85 CONTINUE
   9Ø IF (IR.EQ.N) GO TO 95
      IER = 129
      GO TO 9ØØØ
   95 ITN = Ø
      DF = -ONE
C
```

```
C     EVALUATE GRADIENT W(IG+I),I=1,...,N
C
  100 LINK = 1
      GO TO 200
  105 CONTINUE
C
C     BEGIN ITERATION LOOP
C
C     WRITE RESTART INFORMATION USING VALUES FOR CYCLE JUST COMPLETED
C
      REWIND 8
      WRITE(8)N,NSIG,IOPT,IPRDER
      WRITE(8)F,(X(I),I=1,N),(G(I),I=1,N),(W(I),I=1,3*N)
      NN2=N*(N+1)/2
      WRITE(8)(H(I),I=1,NN2)
      WRITE(8)IG,IGG,IS,IDIFF,IR,IJ,I,J,NM1,JJ,JP1,L,KJ,K,
     .        IFN,LINK,ITN,II,IM1,JNT,NP1,JB,NJ
      WRITE(8)HH,EPS,HJJ,V,DF,RELX,GSO,DIFF,AEPS,ALPHA,FF,
     .        TOT,F1,F2,Z,GYS,DGS,SIG,ZZ,GNRM,HHH,GHH,H2,F11,
     .        F12,F21,F22,HMAX,HMIN,E,RMS
      ENDFILE 8
C
      IF (IFN.GE.MAXFN) GO TO 240
      ITN = ITN+1
      IF (ITN.GT.MAXITN) GO TO 241
      IF(ITN.EQ.1) THEN
         WRITE(6,106)
  106    FORMAT(/,1H ,20X,'ENERGY TERMS (KILOCALORIES)',//)
         WRITE(6,107)
  107 FORMAT(3X,'CYCLE',7X,'TOTAL',7X,'ELECTROSTATIC',6X,'ATOT',7X,
     .'BTOT',6X,'POLARIZATION',5X,'DISPERSION',5X,'THREE-BODY',
     .5X,'RMS DERIV',/)
      END IF
      CALL EPARM(X,E,ITN,1)
C
      DO 110 I=1,N
         W(I) = -W(IG+I)
  110 CONTINUE
C
C     DETERMINE SEARCH DIRECTION W BY SOLVING H*W = -G WHERE
C     H = L*D*L - TRANSPOSE
C
      IF (IR.LT.N) GO TO 140
C
C     N .EQ. 1
C
      G(1) = W(1)
      IF (N.GT.1) GO TO 115
      W(1) = W(1)/H(1)
      GO TO 140
C
C     N .GT. 1
C
  115 II = 1
C
C     SOLVE L*W = -G
C
      DO 125 I=2,N
         IJ = II
         II = II+I
         V = W(I)
         IM1 = I-1
         DO 120 J=1,IM1
            IJ = IJ+1
```

```
            V = V-H(IJ)*W(J)
  120     CONTINUE
          G(I) = V
          W(I) = V
  125 CONTINUE
C
C     SOLVE (D*LT)*Z = W WHERE LT = L-TRANSPOSE
C
      W(N) = W(N)/H(II)
      JJ = II
      NM1 = N-1
      DO 135 NJ=1,NM1
C
C     J = N-1,N-2,...,1
C
         J = N-NJ
         JP1 = J+1
         JJ = JJ-JP1
         V = W(J)/H(JJ)
         IJ = JJ
         DO 130 I=JP1,N
            IJ = IJ+I-1
            V = V-H(IJ)*W(I)
  130    CONTINUE
         W(J) = V
  135 CONTINUE
C
C     DETERMINE STEP LENGTH ALPHA
C
  140 RELX = ZERO
      GS0 = ZERO
      DO 145 I=1,N
         W(IS+I) = W(I)
         DIFF = DABS(W(I))/DMAX1(DABS(X(I)),AX)
         RELX = DMAX1(RELX,DIFF)
         GS0 = GS0+W(IG+I)*W(I)
  145 CONTINUE
      IF (RELX.EQ.ZERO) GO TO 245
      AEPS = EPS/RELX
      IER = 130
      IF (GS0.GE.ZERO) GO TO 245
      IF (DF.EQ.ZERO) GO TO 245
      IER = 0
      ALPHA = (-DF-DF)/GS0
      IF (ALPHA.LE.ZERO) ALPHA = ONE
      ALPHA = DMIN1(ALPHA,ONE)
      IF (IDIFF.EQ.2) ALPHA = DMAX1(P1,ALPHA)
      FF = F
      TOT = ZERO
      JNT = 0
C
C     SEARCH ALONG  X+ALPHA*W
C
  150 IF (IFN.GE.MAXFN) GO TO 240
      DO 155 I=1,N
         W(I) = X(I)+ALPHA*W(IS+I)
  155 CONTINUE
      CALL EPARN(W,F1,ITN,0)
      IFN = IFN+1
      IF (F1.GE.F) GO TO 180
      F2 = F
      TOT = TOT+ALPHA
  160 IER = 0
      F = F1
```

```
      DO 165 I=1,N
         X(I) = W(I)
  165 CONTINUE
      IF (JNT-1) 170, 200, 205
  170 IF (IFN.GE.MAXFN) GO TO 240
      DO 175 I=1,N
         W(I) = X(I)+ALPHA*W(IS+I)
  175 CONTINUE
      CALL EPARM(W,F1,ITN,0)
      IFN = IFN+1
      IF (F1.GE.F) GO TO 205
      IF (F1+F2.GE.F+F .AND. SEVEN*F1+FIVE*F2.GT.TWELVE*F) JNT = 2
      TOT = TOT+ALPHA
      ALPHA = ALPHA+ALPHA
      GO TO 160
  180 CONTINUE
      IF (F.EQ.FF .AND. IDIFF.EQ.2 .AND. RELX.GT.EPS) IER = 130
      IF (ALPHA.LT.AEPS) GO TO 245
      IF (IFN.GE.MAXFN) GO TO 240
      ALPHA = HALF*ALPHA
      DO 185 I=1,N
         W(I) = X(I)+ALPHA*W(IS+I)
  185 CONTINUE
      CALL EPARM(W,F2,ITN,0)
      IFN = IFN+1
      IF (F2.GE.F) GO TO 195
      TOT = TOT+ALPHA
      IER = 0
      F = F2
      DO 190 I=1,N
         X(I) = W(I)
  190 CONTINUE
      GO TO 200
  195 Z = P1
      IF (F1+F.GT.F2+F2) Z = ONE+HALF*(F-F1)/(F+F1-F2-F2)
      Z = DMAX1(P1,Z)
      ALPHA = Z*ALPHA
      JNT = 1
      GO TO 150
  200 IF (TOT.LT.AEPS) GO TO 245
  205 ALPHA = TOT
C
C     SAVE OLD GRADIENT
C
      DO 210 I=1,N
         W(I) = W(IG+I)
  210 CONTINUE
C
C     EVALUATE GRADIENT W(IG+I), I=1,...,N
C
      LINK = 2
      GO TO 280
  215 IF (IFN.GE.MAXFN) GO TO 240
      GYS = ZERO
      DO 220 I=1,N
         GYS = GYS+W(IG+I)*W(IS+I)
         W(IGG+I) = W(I)
  220 CONTINUE
      DF = FF-F
      DGS = GYS-GS0
      IF (DGS.LE.ZERO) GO TO 105
      IF (DGS+ALPHA*GS0.GT.ZERO) GO TO 230
C
C     UPDATE HESSIAN H USING COMPLEMENTARY DFP FORMULA
```

```
C
      SIG = ONE/GSØ
      IR = -IR
      CALL ZXMJN(H,N,W,SIG,G,IR,Ø,ZERO)
      DO 225 I=1,N
         G(I) = W(IG+I)-W(IGG+I)
  225 CONTINUE
      SIG = ONE/(ALPHA*DGS)
      IR = -IR
      CALL ZXMJN(H,N,G,SIG,W,IR,Ø,ZERO)
      GO TO 1Ø5
C
C     UPDATE HESSIAN USING DFP FORMULA
C
  23Ø ZZ = ALPHA/(DGS-ALPHA*GSØ)
      SIG = -ZZ
      CALL ZXMJN(H,N,W,SIG,G,IR,Ø,REPS)
      Z = DGS*ZZ-ONE
      DO 235 I=1,N
         G(I) = W(IG+I)+Z*W(IGG+I)
  235 CONTINUE
      SIG = ONE/(ZZ*DGS*DGS)
      CALL ZXMJN(H,N,G,SIG,W,IR,Ø,ZERO)
      GO TO 1Ø5
C
C     MAXIMUM NUMBER OF MINIMIZATION CYCLES OR FUNCTION
C     EVALUATIONS EXCEEDED.
C
  24Ø IER = 131
      GO TO 25Ø
  241 IER = 132
      GO TO 25Ø
C
  245 IF (IDIFF.EQ.2) GO TO 25Ø
C
C     CHANGE TO CENTRAL DIFFERENCES
C
      IDIFF = 2
      GO TO 1ØØ
  25Ø IF (IER.NE.Ø) GO TO 255
      IF (RELX.LE.EPS) GO TO 255
      GO TO 1ØØ
C
C     MOVE GRADIENT TO G AND RETURN
C
  255 GNRM = ZERO
      DO 26Ø I=1,N
         G(I) = W(IG+I)
         GNRM = GNRM+G(I)*G(I)
  26Ø CONTINUE
      RMS = DSQRT(GNRM/DFLOAT(N))
      GNRM = DSQRT(GNRM)
      W(1) = RMS
      W(2) = IFN
      W(3) = -DLOG1Ø(DMAX1(REPS,RELX))
C
C     COMPUTE H = L*D*L-TRANSPOSE
C
      IF (N.EQ.1) GO TO 9ØØØ
      NP1 = N+1
      NM1 = N-1
      JJ = (N*(NP1))/2
      DO 275 JB=1,NM1
         JP1 = NP1-JB
```

```
         JJ = JJ-JP1
         HJJ = H(JJ)
         IJ = JJ
         L = 0
         DO 270 I=JP1,N
            L = L+1
            IJ = IJ+I-1
            V = H(IJ)*HJJ
            KJ = IJ
            DO 265 K=I,N
               H(KJ+L) = H(KJ+L)+H(KJ)*V
               KJ = KJ+K
  265       CONTINUE
            H(IJ) = V
  270    CONTINUE
         HJJ = H(JJ)
  275 CONTINUE
      GO TO 9000
C
C     EVALUATE GRADIENT
C
  280 IF (IDIFF.EQ.2) GO TO 290
C
C     FORWARD DIFFERENCES    GRADIENT = W(IG+I), I=1,...,N
C
      DO 285 I=1,N
         Z = HH*DMAX1(DABS(X(I)),AX)
         ZZ = X(I)
         X(I) = ZZ+Z
         CALL EPARM(X,F1,ITN,0)
         W(IG+I) = (F1-F)/Z
         X(I) = ZZ
  285 CONTINUE
      IF(IPRDER.EQ.1) WRITE(6,99997)
      IF(IPRDER.EQ.1) WRITE(6,99998)(I,X(I),I,W(IG+I),I=1,N)
      IF(IPRDER.EQ.1) WRITE(6,99999)
      IFN = IFN+N
      GO TO (105, 215), LINK
C
C     CENTRAL DIFFERENCES     GRADIENT = W(IG+I), I=1,...,N
C
  290 DO 295 I=1,N
         Z = HH*DMAX1(DABS(X(I)),AX)
         ZZ = X(I)
         X(I) = ZZ+Z
         CALL EPARM(X,F1,ITN,0)
         X(I) = ZZ-Z
         CALL EPARM(X,F2,ITN,0)
         W(IG+I) = (F1-F2)/(Z+Z)
         X(I) = ZZ
  295 CONTINUE
      IF(IPRDER.EQ.1) WRITE(6,99997)
      IF(IPRDER.EQ.1) WRITE(6,99998)(I,X(I),I,W(IG+I),I=1,N)
      IF(IPRDER.EQ.1) WRITE(6,99999)
      IFN = IFN+N+N
      GO TO (105, 215), LINK
99997 FORMAT(/,2(6X,'VARIABLE',15X,'GRADIENT'))
99998 FORMAT(2(5X,'X(',I3,')=',F6.1,5X,'G(',I3,')=',F10.6))
99999 FORMAT(//)
 9000 CONTINUE
 9005 RETURN
      END
      SUBROUTINE ZXMJN  (A,N,Z,SIG,W,IR,MK,EPS)
C
```

```
      INTEGER             N,IR,MK
      DOUBLE PRECISION    A(1),Z(9ØØ),SIG,W(9ØØ),EPS
      INTEGER             J,JJ,IJ,JP1,I,II,MM
      DOUBLE PRECISION    ZERO,ONE,FOUR,TI,V,TIM,AL,R,B,GM,Y
      DATA                ZERO/Ø.ØDØ/,ONE/1.ØDØ/,FOUR/4.ØDØ/
C
C     UPDATE FACTORS GIVEN IN A SIG*Z*Z-TRANSPOSE IS ADDED
C
      IF (N.GT.1) GO TO 5
C
C     N .EQ. 1
C
      A(1) = A(1)+SIG*Z(1)*Z(1)
      IR = 1
      IF (A(1).GT.ZERO) GO TO 9ØØ5
      A(1) = ZERO
      IR = Ø
      GO TO 9ØØ5
C
C     N .GT. 1
C
    5 IF (SIG.GT.ZERO) GO TO 65
      IF (SIG.EQ.ZERO.OR.IR.EQ.Ø) GO TO 9ØØ5
      TI = ONE/SIG
      JJ = Ø
      IF (MK.EQ.Ø) GO TO 15
C
C     L*W = Z ON INPUT
C
      DO 1Ø J=1,N
         JJ = JJ+J
         IF (A(JJ).NE.ZERO) TI = TI+(W(J)*W(J))/A(JJ)
   1Ø CONTINUE
      GO TO 4Ø
C
C     SOLVE L*W = Z
C
   15 DO 2Ø J=1,N
         W(J) = Z(J)
   2Ø CONTINUE
      DO 35 J=1,N
         JJ = JJ+J
         V = W(J)
         IF (A(JJ).GT.ZERO) GO TO 25
         W(J) = ZERO
         GO TO 35
   25    TI = TI+(V*V)/A(JJ)
         IF (J.EQ.N) GO TO 35
         IJ = JJ
         JP1 = J+1
         DO 3Ø I=JP1,N
            IJ = IJ+I-1
            W(I) = W(I)-V*A(IJ)
   3Ø    CONTINUE
   35 CONTINUE
C
C     SET TI, TIM AND W
C
   4Ø IF (IR.LE.Ø) GO TO 45
      IF (TI.GT.ZERO) GO TO 5Ø
      IF (MK-1) 65,65,55
   45 TI = ZERO
      IR = -IR-1
      GO TO 55
```

```
   50 TI = EPS/SIG
      IF (EPS.EQ.ZERO) IR = IR-1
   55 TIM = TI
      II = JJ
      I = N
      DO 60 J=1,N
         IF (A(II).NE.ZERO) TIM = TI-(W(I)*W(I))/A(II)
         W(I) = TI
         TI = TIM
         II = II-I
         I = I-1
   60 CONTINUE
      NM = 1
      GO TO 70
   65 NM = 0
      TIM = ONE/SIG
   70 JJ = 0
C
C     UPDATE A
C
      DO 110 J=1,N
         JJ = JJ+J
         IJ = JJ
         JP1 = J+1
C
C     UPDATE A(J,J)
C
         V = Z(J)
         IF (A(JJ).GT.ZERO) GO TO 85
C
C     A(J,J) .EQ. ZERO
C
         IF (IR.GT.0.OR.SIG.LT.ZERO.OR.V.EQ.ZERO) GO TO 80
         IR = 1-IR
         A(JJ) = (V*V)/TIM
         IF (J.EQ.N) GO TO 90005
         DO 75 I=JP1,N
            IJ = IJ+I-1
            A(IJ) = Z(I)/V
   75    CONTINUE
         GO TO 90005
   80    TI = TIM
         GO TO 110
C
C     A(J,J) .GT. ZERO
C
   85    AL = V/A(JJ)
         TI = W(J)
         IF (NM.EQ.0) TI = TIM+V*AL
         R = TI/TIM
         A(JJ) = R*A(JJ)
         IF (R.EQ.ZERO) GO TO 115
         IF (J.EQ.N) GO TO 115
C
C     UPDATE REMAINDER OF COLUMN J
C
         B = AL/TI
         IF (R.GT.FOUR) GO TO 95
         DO 90 I=JP1,N
            IJ = IJ+I-1
            Z(I) = Z(I)-V*A(IJ)
            A(IJ) = A(IJ)+B*Z(I)
   90    CONTINUE
         GO TO 105
```

```
   95    GM = TIM/TI
         DO 1ØØ I=JP1,N
            IJ = IJ+I-1
            Y = A(IJ)
            A(IJ) = B*Z(I)+Y*GM
            Z(I) = Z(I)-V*Y
  1ØØ    CONTINUE
  1Ø5    TIM = TI
  11Ø CONTINUE
  115 IF (IR.LT.Ø) IR = -IR
 9ØØ5 CONTINUE
      RETURN
      END
      SUBROUTINE READIN(ALAB,MLAB)
C     SUBROUTINE FOR READING IN COORDINATE DATA. THE SUBROUTINE
C     READS IN THE COORDINATES, COUNTS THE TOTAL NUMBER OF
C     MOLECULES AND THE NUMBER OF ATOMS IN EACH MOLECULE, AND
C     STORES THE COORDINATES IN A 2-DIMENSIONAL ARRAY. THE
C     FIRST INDEX CORRESPONDS TO MOLECULE NUMBER AND THE
C     SECOND INDEX CORRESPONDS TO ATOM NUMBER.
C
      IMPLICIT REAL*8(A-H,O-Z)
      COMMON
     ./PARMS/ RE(2Ø),EPS(2Ø),Q(2Ø),POL(2Ø)
     ./COORD/ X(15Ø,1Ø),Y(15Ø,1Ø),Z(15Ø,1Ø),ITYPE(15Ø,1Ø)
     ./BOOK/ NMOL,NAT(15Ø),RMS,ETB
     ./POLARZ/ ISCFP,MAXIT,NPOL,TOL,RCUT
     ./FLAGS/ IPRDIP,ISTAT
     ./ROT/ CONV,XI(15Ø,1Ø),YI(15Ø,1Ø),ZI(15Ø,1Ø)
      DIMENSION ALAB(15Ø,1Ø),MLAB(15Ø)
      CHARACTER*4 ALAB
      CHARACTER*3 MLAB
      CHARACTER*4 AAA
      CHARACTER*3 BBB
C
C     INITIATE COUNTERS.
C
      I=Ø
   1Ø CONTINUE
      I=I+1
      K=Ø
   2Ø CONTINUE
      K=K+1
C
C     READ DATA ONE LINE AT A TIME.
C
      READ(2,1)SYM,AAA,BBB,XX,YY,ZZ,IAC
    1 FORMAT(A4,8X,A4,1X,A3,1ØX,3FØ.3,I2)
      IF(SYM.EQ.'TER ') NAT(I)=K-1
      IF(SYM.EQ.'TER ') GO TO 1Ø
      IF(SYM.EQ.'END ') NMOL=I-1
      IF(SYM.EQ.'END ') GO TO 5Ø
C
C     STORE DATA IN ARRAY.
C
      ALAB(I,K)=AAA
      MLAB(I)=BBB
      X(I,K)=XX
      XI(I,K)=XX
      Y(I,K)=YY
      YI(I,K)=YY
      Z(I,K)=ZZ
      ZI(I,K)=ZZ
      ITYPE(I,K)=IAC
```

```
      GO TO 20
   50 CONTINUE
      RETURN
      END
      SUBROUTINE OUTPUT(ALAB,MLAB)
C     THIS SUBROUTINE WRITES A FILE CONTAINING THE FINAL
C     COORDINATES IN PROTEIN DATA BANK FORMAT (IF IPRPDB = 1)
C
      IMPLICIT REAL*8(A-H,O-Z)
      COMMON
     ./PARMS/ RE(20),EPS(20),Q(20),POL(20)
     ./COORD/ X(150,10),Y(150,10),Z(150,10),ITYPE(150,10)
     ./BOOK/ NMOL,NAT(150),RMS,ETB
     ./POLARZ/ ISCFP,MAXIT,NPOL,TOL,RCUT
     ./FLAGS/ IPRDIP,ISTAT
     ./ROT/ CONV,XI(150,10),YI(150,10),ZI(150,10)
      DIMENSION ALAB(150,10),MLAB(150)
      CHARACTER*4 ALAB
      CHARACTER*3 MLAB
C
      I=0
   10 CONTINUE
      J=0
      I=I+1
      IF(I.GT.NMOL) THEN
         WRITE(7,50)
         RETURN
      END IF
      NATI=NAT(I)
   20 J=J+1
      IF(J.GT.NATI) THEN
         WRITE(7,40)
         GO TO 10
      END IF
      WRITE(7,30,ERR=99)ALAB(I,J),MLAB(I),I,X(I,J),Y(I,J),
     .Z(I,J),ITYPE(I,J)
      GO TO 20
C
   30 FORMAT('ATOM',8X,A4,1X,A3,1X,I5,4X,3F8.3,I2)
   40 FORMAT('TER')
   50 FORMAT('END')
   99 WRITE(6,100)
  100 FORMAT(///,1H ,'****** PROBLEMS WRITING PDB FILE! ******',///)
      RETURN
      END
      SUBROUTINE STAT(ALAB,MLAB)
C     PERFORM STATISTICS ON ATOM AND MOLECULE MOVEMENTS
C     OF SYSTEM DURING MINIMIZATION.
C
      IMPLICIT REAL*8(A-H,O-Z)
      COMMON
     ./PARMS/ RE(20),EPS(20),Q(20),POL(20)
     ./COORD/ X(150,10),Y(150,10),Z(150,10),ITYPE(150,10)
     ./BOOK/ NMOL,NAT(150),RMS,ETB
     ./POLARZ/ ISCFP,MAXIT,NPOL,TOL,RCUT
     ./FLAGS/ IPRDIP,ISTAT
     ./ROT/ CONV,XI(150,10),YI(150,10),ZI(150,10)
      DIMENSION ALAB(150,10),MLAB(150)
      DIMENSION HISTO(136)
      CHARACTER*4 ALAB
      CHARACTER*3 MLAB
      CHARACTER*1 HISTO
      CHARACTER*1 STAR
      DATA NLIM /90/
```

```
C
      STAR='*'
      RRMS=0.D0
      AV=0.D0
C
C     LOAD HISTOGRAM ARRAY.
C
      DO 1 K=1,90
    1 HISTO(K)=STAR
C
C     DO STATISTICS OVER ATOMS OF INDIVIDUAL MOLECULES FIRST,
C     THEN STATISTICS FOR ENTIRE SYSTEM.
C
      DO 10 I=1,NMOL
      NATI=NAT(I)
      NATOM=NATOM+NATI
      WRITE(6,100)I,MLAB(I)
  100 FORMAT(/,1H ,15X,'MOLECULE ',I3,2X,A3,/)
      DO 20 J=1,NATI
      RMSM=0.D0
      AVM=0.D0
      DX=X(I,J)-XI(I,J)
      DY=Y(I,J)-YI(I,J)
      DZ=Z(I,J)-ZI(I,J)
      RMSM2=DX*DX+DY*DY+DZ*DZ
      R=DSQRT(RMSM2)
      IR=IDINT(R*10.D0)
      IF(IR.LT.1) IR=1
      IF(IR.GT.NLIM) IR=NLIM
      RMSM=RMSM+RMSM2
      AVM=AVM+R
      WRITE(6,101)J,ALAB(I,J),R,(HISTO(IJK),IJK=1,IR)
  101 FORMAT(1H ,5X,I5,'.',1X,A4,2X,F12.4,2X,100A1)
      RRMS=RRMS+RMSM
      AV=AV+AVM
   20 CONTINUE
      RNATI=DFLOAT(NATI)
      RMSM=DSQRT(RMSM/RNATI)
      WRITE(6,102)I,RMSM
  102 FORMAT(/1H ,5X,'ROOT MEAN SQUARE DEVIATION FOR
     1 MOLECULE',I5,' = ',F10.6,/)
   10 CONTINUE
      RNAT=DFLOAT(NATOM)
      AV=AV/RNAT
      SD=DSQRT(DABS((RRMS/RNAT)-AV*AV))
      RRMS=DSQRT(RRMS/RNAT)
      WRITE(6,103)NMOL,NATOM
  103 FORMAT(1H ,5X,'SYSTEM CONTAINS ',I3,' MOLECULES AND'
     1,I5,' TOTAL ATOMS',/)
      WRITE(6,104)RRMS,AV,SD
  104 FORMAT(1H ,5X,'ROOT MEAN SQUARE DEVIATION FOR SYSTEM = ',
     1F10.6,' ANGSTROMS',/,6X,'MEAN DISPLACEMENT = ',F10.6,
     1' ANGSTROMS',/,6X,'STANDARD DEVIATION = ',F10.6,
     1' ANGSTROMS',///)
      RETURN
      END
```

```
C     NPQUAD.FOR
C
C     THIS ALGORITHM NUMERICALLY EVALUATES A MULTIDIMENSIONAL
C     INTEGRAL USING A NON-PRODUCT QUADRATURE FORMULA WHICH
C     REQUIRES n + 2 FUNCTION EVALUATIONS IN AN N-DIMENSIONAL
C     SPACE. THE FORMULA IS A QUADRATURE FORMULA OF DEGREE TWO;
C     FOR ACCURATE EVALUATION, EACH DIMENSION OF THE INTEGRAL
C     SHOULD BE SUBDIVIDED INTO APPORPRIATELY SMALL SEGMENTS
C     SO AS TO BE REASONABLY APPROXIMATED BY THE QUADRATURE
C     (I.E. THIS FORMULA WILL EVAUATE ANY POYNOMIAL OF DEGREE
C     LESS THAN TWO EXACTLY). AFTER DIVIDION INTO SUBSEGMENTS,
C     EACH SUBSEGMENT MUST BE TRANSFORMED TO THE N-DIMENSIONAL
C     UNIT HYPERCUBE. THE FORMULA EFFICIENTLY EVALUATES A
C     MULTIDIMENSIONAL INTEGRAL BY COMPUTING THE INTEGRAL VALUE
C     AT THE MINIMAL NUMBER OF POINTS NECESSARY TO EFFECTIVELY
C     SAMPLE THE N\-DIMENSIONAL SPACE.
C
C     THE CURRENT PROGRAM IS SET UP TO COMPUTE THE SECOND
C     VIRIAL COEFFICIENT FOR WATER USING THE CLEMENTI CI
C     POTENTIAL. THE ONLY INPUT PARAMETER FOR THE PROGRAM
C     IS THE TEMPERATURE AT WHICH THE SECOND VIRIAL COEFFICIENT
C     IS TO BE EVALUATED. THE PROGRAM IS MODULAR, AND ANY
C     APPROPRIATE POTENTIAL FUNCTION CAN BE SUBSTITUTED.
C
C
C     NON-PRODUCT N-DIMENSIONAL QUADRATURE INTEGRATION
C
      IMPLICIT DOUBLE PRECISION (A-H,O-Z)
      DIMENSION G(5),H(5),V(7,5)
C
      OPEN(UNIT=5,FILE='TC',STATUS='OLD')
      REWIND 5
      OPEN(UNIT=6,FILE='VIRIAL',STATUS='NEW',SHARED)
      REWIND 6
      OPEN(UNIT=8,FILE='RST',STATUS='NEW',FORM='UNFORMATTED')
      REWIND 8
C
C     INPUT CARD
C     TEMP: TEMPERATURE FOR SECOND VIRIAL COEFFICIENT CALCULATION
C     IRST: RESTART PARAMETER.
C           IRST = 0 : NORMAL FRESH START
C           IRST = 1 : RESTART THE CALCULATION USING INFORMATION
C                      CONTAINED IN THE RESTART FILE RSTVIR.DAT. A RESTART
C                      FILE IS GENERATED BY THE PROGRAM AUTOMATICALLY
C                      EACH TIME A JOB IS INITIATED.
C
      READ(5,1)TEMP,IRST
    1 FORMAT(D20.8,I2)
      WRITE(6,2)TEMP
    2 FORMAT(1H ,'SECOND VIRIAL COEFFICIENT CALCULATION AT ',
     1F8.2,2X,'DEGREES CELSIUS',///)
C
C     SET CONSTANTS
C
      DENOM=TEMP*1.987190D-3
      PI=2.D0*DASIN(1.D0)
      PI5=PI/5.D0
      PI53=PI5*3.D0
      PI55=PI5*5.D0
      PI57=PI5*7.D0
      PI59=PI5*9.D0
      PI10=PI/10.D0
      PI103=PI10*3.D0
```

```
      PI105=PI10*5.D0
      PI107=PI10*7.D0
      PI109=PI10*9.D0
      PI20=PI/20.D0
      PI203=PI20*3.D0
      PI205=PI20*5.D0
      PI207=PI20*7.D0
      PI209=PI20*9.D0
      WT=32.D0/7.D0
      G(1)=0.9061798D0
      G(2)=-0.9061798D0
      G(3)=0.5384693D0
      G(4)=-0.5384693D0
      G(5)=0.0000000D0
      H(1)=0.2369269D0
      H(2)=0.2369269D0
      H(3)=0.4786287D0
      H(4)=0.4786287D0
      H(5)=0.5688889D0
      V(1,1)=0.4409586D0
      V(1,2)=0.7637626D0
      V(1,3)=-V(1,1)
      V(1,4)=V(1,2)
      V(1,5)=-0.8819171D0
      V(2,1)=-V(1,1)
      V(2,2)=V(1,2)
      V(2,3)=-V(1,1)
      V(2,4)=-V(1,2)
      V(2,5)=-V(1,5)
      V(3,1)=V(1,5)
      V(3,2)=0.0000000D0
      V(3,3)=-V(1,5)
      V(3,4)=V(3,2)
      V(3,5)=-V(1,5)
      V(4,1)=-V(1,1)
      V(4,2)=-V(1,2)
      V(4,3)=V(1,1)
      V(4,4)=V(1,2)
      V(4,5)=-V(1,5)
      V(5,1)=-0.6236096D0
      V(5,2)=-V(5,1)
      V(5,3)=V(5,1)
      V(5,4)=-V(5,1)
      V(5,5)=V(5,1)
      V(6,1)=-V(1,5)
      V(6,2)=V(3,2)
      V(6,3)=-V(1,5)
      V(6,4)=V(3,2)
      V(6,5)=-V(1,5)
      V(7,1)=V(3,2)
      V(7,2)=V(3,2)
      V(7,3)=V(3,2)
      V(7,4)=V(3,2)
      V(7,5)=V(3,2)
C
C     OUTER LOOP: 5 PT. GAUSSIAN QUADRATURE FOR R DIMENSION
C
      TOT=0.D0
      IRINIT=1
      IF(IRST.EQ.1) THEN
        OPEN(UNIT=7,FILE='RSTVIR',STATUS='OLD',FORM='UNFORMATTED')
        REWIND 7
        READ(7)INIT,RTOT
        IRINIT=INIT+1
```

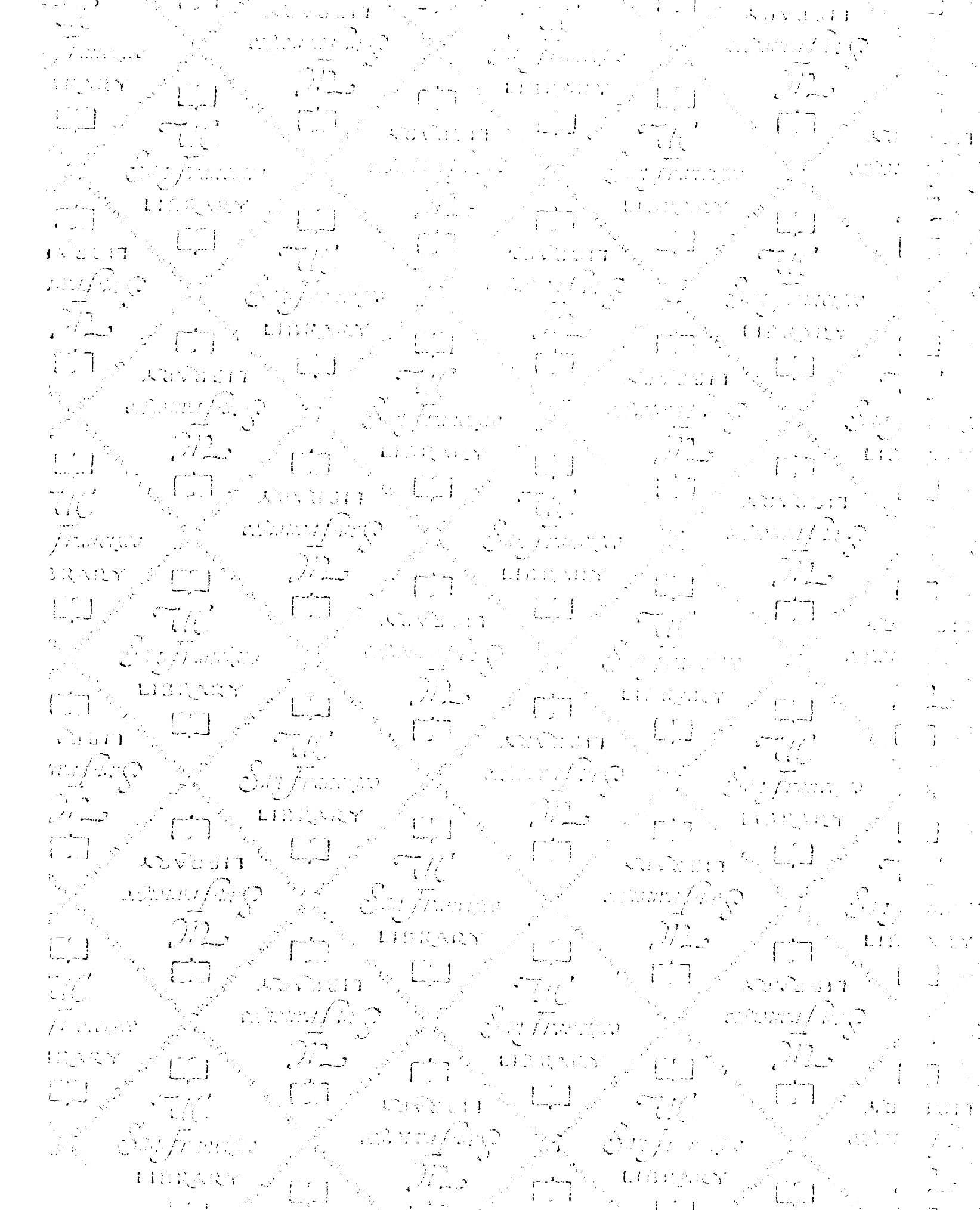
```
        TOT=RTOT
        IF(IRINIT.GT.12) GO TO 2ØØ
      END IF
      DO 1ØØ IR=IRINIT,12
      AG1=Ø.DØ
      DO 2Ø I=1,5
      IF(IR.EQ.1) THEN
         XR=Ø.2DØ*G(I)+2.6DØ
         GFAC=Ø.2DØ
      ELSE IF(IR.EQ.2) THEN
         XR=Ø.2DØ*G(I)+3.ØDØ
         GFAC=Ø.2DØ
      ELSE IF(IR.EQ.3) THEN
         XR=Ø.3DØ*G(I)+3.5DØ
         GFAC=Ø.3DØ
      ELSE IF(IR.EQ.4) THEN
         XR=Ø.3DØ*G(I)+4.1DØ
         GFAC=Ø.3DØ
      ELSE IF(IR.EQ.5) THEN
         XR=Ø.3DØ*G(I)+4.7DØ
         GFAC=Ø.3DØ
      ELSE IF(IR.EQ.6) THEN
         XR=Ø.4DØ*G(I)+5.4DØ
         GFAC=Ø.4DØ
      ELSE IF(IR.EQ.7) THEN
         XR=Ø.4DØ*G(I)+6.2DØ
         GFAC=Ø.4DØ
      ELSE IF(IR.EQ.8) THEN
         XR=Ø.4DØ*G(I)+7.ØDØ
         GFAC=Ø.4DØ
      ELSE IF(IR.EQ.9) THEN
         XR=Ø.5DØ*G(I)+7.9DØ
         GFAC=Ø.5DØ
      ELSE IF(IR.EQ.1Ø) THEN
         XR=Ø.5DØ*G(I)+8.9DØ
         GFAC=Ø.5DØ
      ELSE IF(IR.EQ.11) THEN
         XR=Ø.6DØ*G(I)+1Ø.DØ
         GFAC=Ø.6DØ
      ELSE IF(IR.EQ.12) THEN
         XR=Ø.7DØ*G(I)+11.3DØ
         GFAC=Ø.7DØ
      END IF
      HR=H(I)
C
C     INNER LOOPS FOR NON-PRODUCT QUADRATURE INTEGRATION OF
C     ANGULAR SPACE.
C
      A1=Ø.DØ
      DO 1Ø IE=1,5
      DO 1Ø ID=1,5
      DO 1Ø IC=1,5
      DO 1Ø IB=1,5
      DO 1Ø IA=1,5
      ARG=Ø.DØ
      DO 5 N=1,7
      V1=V(N,1)
      V2=V(N,2)
      V3=V(N,3)
      V4=V(N,4)
      V5=V(N,5)
C
C     PARTITION 5-DIMENSIONAL ANGLE SPACE INTO 3125 HYPERCUBES
C
```

```
      IF(IE.EQ.1) THEN
         XE=PI1Ø*V1+PI1Ø
      ELSE IF(IE.EQ.2) THEN
         XE=PI1Ø*V1+PI1Ø3
      ELSE IF(IE.EQ.3) THEN
         XE=PI1Ø*V1+PI1Ø5
      ELSE IF(IE.EQ.4) THEN
         XE=PI1Ø*V1+PI1Ø7
      ELSE IF(IE.EQ.5) THEN
         XE=PI1Ø*V1+PI1Ø9
      END IF
C
      IF(ID.EQ.1) THEN
         XD=PI1Ø*V2+PI1Ø
      ELSE IF(ID.EQ.2) THEN
         XD=PI1Ø*V2+PI1Ø3
      ELSE IF(ID.EQ.3) THEN
         XD=PI1Ø*V2+PI1Ø5
      ELSE IF(ID.EQ.4) THEN
         XD=PI1Ø*V2+PI1Ø7
      ELSE IF(ID.EQ.5) THEN
         XD=PI1Ø*V2+PI1Ø9
      END IF
C
      IF(IC.EQ.1) THEN
         XC=PI5*V3+PI5
      ELSE IF(IC.EQ.2) THEN
         XC=PI5*V3+PI53
      ELSE IF(IC.EQ.3) THEN
         XC=PI5*V3+PI55
      ELSE IF(IC.EQ.4) THEN
         XC=PI5*V3+PI57
      ELSE IF(IC.EQ.5) THEN
         XC=PI5*V3+PI59
      END IF
C
      IF(IB.EQ.1) THEN
         XB=PI2Ø*V4+PI2Ø
      ELSE IF(IB.EQ.2) THEN
         XB=PI2Ø*V4+PI2Ø3
      ELSE IF(IB.EQ.3) THEN
         XB=PI2Ø*V4+PI2Ø5
      ELSE IF(IB.EQ.4) THEN
         XB=PI2Ø*V4+PI2Ø7
      ELSE IF(IB.EQ.5) THEN
         XB=PI2Ø*V4+PI2Ø9
      END IF
C
      IF(IA.EQ.1) THEN
         XA=PI1Ø*V5+PI1Ø
      ELSE IF(IA.EQ.2) THEN
         XA=PI1Ø*V5+PI1Ø3
      ELSE IF(IA.EQ.3) THEN
         XA=PI1Ø*V5+PI1Ø5
      ELSE IF(IA.EQ.4) THEN
         XA=PI1Ø*V5+PI1Ø7
      ELSE IF(IA.EQ.5) THEN
         XA=PI1Ø*V5+PI1Ø9
      END IF
C
C     FUNCTION FOR INTEGRATION VIA NON-PRODUCT QUADRATURE FORMULA
C
C     XR : R
C     XA : BETA
```

```
C      XB : THETA
C      XC : ALPHA
C      XD : PHI
C      XE : GAMMA
C
       CALL FUNC(XR,XB,XD,XC,XA,XE,ENER)
C
       FMAYER=(DEXP(-ENER/DENOM))-1.DØ
C
       AINT=(DSIN(XB)*DSIN(XA)*XR*XR)*FMAYER
       ARG=ARG+WT*AINT
     5 CONTINUE
       A1=A1+ARG
    1Ø CONTINUE
       AG1=AG1+A1*HR
    2Ø CONTINUE
       VAL=(AG1*PI*PI*PI*PI*PI*GFAC)/1.D6
       WRITE(6,1ØØØ)VAL
 1ØØØ FORMAT(1H ,5X,'INT = ',D2Ø.8)
       TOT=TOT+VAL
C
C      WRITE INTERMEDIATE RESULTS TO RESTART FILE
C
       REWIND 8
       WRITE(8)IR,TOT
       ENDFILE 8
   1ØØ CONTINUE
   2ØØ HSPH=(2.DØ*PI*2.4DØ*2.4DØ*2.4DØ/3.DØ)*(6.Ø22Ø5D23/1.D24)
       VIR=(TOT*6.Ø22Ø5D23/(-1.ØD24*2.DØ*PI*PI))+HSPH
       WRITE(6,2ØØØ)VIR
 2ØØØ FORMAT(///,1H ,'2ND VIRIAL COEFFICIENT = ',D2Ø.8,2X,
      1'CM**3/MOLE')
       STOP
       END
       SUBROUTINE FUNC(R,THETA,PHI,ALPHA,BETA,GAMMA,E)
       IMPLICIT DOUBLE PRECISION (A-H,O-Z)
C
C      CLEMENTI CI POTENTIAL
C
       COMMON /COORDS/ XB1,YB1,ZB1,XB2,YB2,ZB2,XB3,YB3,ZB3,XB4,YB4,ZB4
C
       XA1=Ø.ØDØ
       YA1=Ø.ØDØ
       ZA1=Ø.ØDØ
       XA2=Ø.5ØGØ144DØ
       YA2=Ø.75Ø84Ø1DØ
       ZA2=Ø.ØDØ
       XA3=Ø.5Ø6Ø144DØ
       YA3=-Ø.75Ø84Ø1DØ
       ZA3=Ø.ØDØ
       XA4=Ø.2677DØ
       YA4=Ø.ØDØ
       ZA4=Ø.ØDØ
       XB1=XA1
       YB1=YA1
       ZB1=ZA1
       XB2=XA2
       YB2=YA2
       ZB2=ZA2
       XB3=XA3
       YB3=YA3
       ZB3=ZA3
       XB4=XA4
       YB4=YA4
```

```
      ZB4=ZA4
      AOO=1088213.ØDØ
      AOH=1455.427DØ
      AHH=666.3373DØ
      AHO=273.5954DØ
      BOO=-5.152712DØ
      BOH=-2.961895DØ
      BHH=-2.760844DØ
      BHO=-2.233264DØ
      Q2=Ø.717516DØ*Ø.717516DØ
C
      CALL MOVE(ALPHA,BETA,GAMMA,R,THETA,PHI)
C
      R44=DSQRT((XA4-XB4)**2+(YA4-YB4)**2+(ZA4-ZB4)**2)
      R42=DSQRT((XA4-XB2)**2+(YA4-YB2)**2+(ZA4-ZB2)**2)
      R43=DSQRT((XA4-XB3)**2+(YA4-YB3)**2+(ZA4-ZB3)**2)
      R21=DSQRT((XA2-XB1)**2+(YA2-YB1)**2+(ZA2-ZB1)**2)
      R22=DSQRT((XA2-XB2)**2+(YA2-YB2)**2+(ZA2-ZB2)**2)
      R23=DSQRT((XA2-XB3)**2+(YA2-YB3)**2+(ZA2-ZB3)**2)
      R24=DSQRT((XA2-XB4)**2+(YA2-YB4)**2+(ZA2-ZB4)**2)
      R31=DSQRT((XA3-XB1)**2+(YA3-YB1)**2+(ZA3-ZB1)**2)
      R32=DSQRT((XA3-XB2)**2+(YA3-YB2)**2+(ZA3-ZB2)**2)
      R33=DSQRT((XA3-XB3)**2+(YA3-YB3)**2+(ZA3-ZB3)**2)
      R34=DSQRT((XA3-XB4)**2+(YA3-YB4)**2+(ZA3-ZB4)**2)
      R11=DSQRT((XA1-XB1)**2+(YA1-YB1)**2+(ZA1-ZB1)**2)
      R12=DSQRT((XA1-XB2)**2+(YA1-YB2)**2+(ZA1-ZB2)**2)
      R13=DSQRT((XA1-XB3)**2+(YA1-YB3)**2+(ZA1-ZB3)**2)
C
      QTOT=(4.DØ*Q2/R44)+(-2.DØ*Q2/R42)+(-2.DØ*Q2/R43)+
     I     (-2.DØ*Q2/R24)+(Q2/R22)+(Q2/R23)+(-2.DØ*Q2/R34)+
     I     (Q2/R32)+(Q2/R33)
      BTOT=AOO*DEXP(R11*BOO)+AOH*DEXP(R12*BOH)+AOH*DEXP(R13*BOH)
     I    +AOH*DEXP(R21*BOH)+AHH*DEXP(R22*BHH)+AHH*DEXP(R23*BHH)
     I    +AOH*DEXP(R31*BOH)+AHH*DEXP(R32*BHH)+AHH*DEXP(R33*BHH)
      ATOT=AHO*DEXP(R12*BHO)+AHO*DEXP(R13*BHO)+AHO*DEXP(R21*BHO)
     I    +AHO*DEXP(R31*BHO)
C
      QTOT=331.62DØ*QTOT
      E=QTOT+BTOT-ATOT
      RETURN
      END
      SUBROUTINE MOVE(EPHI,ETHET,EPSI,R,THETA,PHI)
C
C     TRANSFORM COORDINATES FROM EULER SPACE TO CARTESIAN SPACE
C
      IMPLICIT DOUBLE PRECISION (A-H,O-Z)
      COMMON /COORDS/ XB1,YB1,ZB1,XB2,YB2,ZB2,XB3,YB3,ZB3,XB4,YB4,ZB4
C
      XO=R*DSIN(THETA)*DCOS(PHI)
      YO=R*DSIN(THETA)*DSIN(PHI)
      ZO=R*DCOS(THETA)
C
      A11=DCOS(EPSI)*DCOS(EPHI)-DCOS(ETHET)*DSIN(EPHI)*DSIN(EPSI)
      A12=DCOS(EPSI)*DSIN(EPHI)+DCOS(ETHET)*DCOS(EPHI)*DSIN(EPSI)
      A13=DSIN(EPSI)*DSIN(ETHET)
      A21=-DSIN(EPSI)*DCOS(EPHI)-DCOS(ETHET)*DSIN(EPHI)*DCOS(EPSI)
      A22=-DSIN(EPSI)*DSIN(EPHI)+DCOS(ETHET)*DCOS(EPHI)*DCOS(EPSI)
      A23=DCOS(EPSI)*DSIN(ETHET)
      A31=DSIN(ETHET)*DSIN(EPHI)
      A32=-DSIN(ETHET)*DCOS(EPHI)
      A33=DCOS(ETHET)
C
      WX1=XB1*A11+YB1*A12+ZB1*A13+XO
      WY1=XB1*A21+YB1*A22+ZB1*A23+YO
```

```
WZ1=XB1*A31+YB1*A32+ZB1*A33+ZO
WX2=XB2*A11+YB2*A12+ZB2*A13+XO
WY2=XB2*A21+YB2*A22+ZB2*A23+YO
WZ2=XB2*A31+YB2*A32+ZB2*A33+ZO
WX3=XB3*A11+YB3*A12+ZB3*A13+XO
WY3=XB3*A21+YB3*A22+ZB3*A23+YO
WZ3=XB3*A31+YB3*A32+ZB3*A33+ZO
WX4=XB4*A11+YB4*A12+ZB4*A13+XO
WY4=XB4*A21+YB4*A22+ZB4*A23+YO
WZ4=XB4*A31+YB4*A32+ZB4*A33+ZO
XB1=WX1
YB1=WY1
ZB1=WZ1
XB2=WX2
YB2=WY2
ZB2=WZ2
XB3=WX3
YB3=WY3
ZB3=WZ3
XB4=WX4
YB4=WY4
ZB4=WZ4
RETURN
END
```


SAN FRANCISCO
PUBLIC
LIBRARY