INFORMATION TO USERS

This reproduction was made from a copy of a manuscript sent to us for publication and microfilming. While the most advanced technology has been used to photograph and reproduce this manuscript, the quality of the reproduction is heavily dependent upon the quality of the material submitted. Pages in any manuscript may have indistinct print. In all cases the best available copy has been filmed.

The following explanation of techniques is provided to help clarify notations which may appear on this reproduction.

1. Manuscripts may not always be complete. When it is not possible to obtain missing pages, a note appears to indicate this.

2. When copyrighted materials are removed from the manuscript, a note appears to indicate this.

3. Oversize materials (maps, drawings, and charts) are photographed by sectioning the original, beginning at the upper left hand corner and continuing from left to right in equal sections with small overlaps. Each oversize page is also filmed as one exposure and is available, for an additional charge, as a standard 35mm slide or in black and white paper format.*

4. Most photographs reproduce acceptably on positive microfilm or microfiche but lack clarity on xerographic copies made from the microfilm. For an additional charge, all photographs are available in black and white standard 35mm slide format.*

***For more information about black and white slides or enlarged paper reproductions, please contact the Dissertations Customer Services Department.**

U·M·I Dissertation Information Service

University Microfilms International
A Bell & Howell Information Company
300 N. Zeeb Road, Ann Arbor, Michigan 48106

Maye, Peter V.

MONTE CARLO COMPUTER SIMULATION STUDIES OF THE AQUEOUS HYDRATION OF THE NUCLEOTIDE BASES AND SUGARS

City University of New York PH.D. 1986

PLEASE NOTE:

In all cases this material has been filmed in the best possible way from the available copy. Problems encountered with this document have been identified here with a check mark __✓__.

1. Glossy photographs or pages _____
2. Colored illustrations, paper or print _____
3. Photographs with dark background _____
4. Illustrations are poor copy _____
5. Pages with black marks, not original copy ✓
6. Print shows through as there is text on both sides of page _____
7. Indistinct, broken or small print on several pages ✓
8. Print exceeds margin requirements _____
9. Tightly bound copy with print lost in spine _____
10. Computer printout pages with indistinct print _____
11. Page(s) _____ lacking when material received, and not available from school or author.
12. Page(s) _____ seem to be missing in numbering only as text follows.
13. Two pages numbered _____. Text follows.
14. Curling and wrinkled pages _____
15. Dissertation contains pages with print at a slant, filmed as received ✓
16. Other _____

_____

_____

MONTE CARLO COMPUTER SIMULATION STUDIES OF THE AQUEOUS HYDRATION OF THE NUCLEOTIDE BASES AND SUGARS.

by

Peter V. Maye

A dissertation submitted to the Graduate Faculty in Biochemistry in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

1986

This manuscript has been read and accepted by the Graduate Faculty in Biochemistry in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

8/15/86 — date — Chairman of Examining Committee

8/20/86 — date — Executive Officer

Supervisory Committee

The City University of New York

Abstract

MONTE CARLO COMPUTER SIMULATION STUDIES OF THE AQUEOUS HYDRATION OF THE NUCLEOTIDE BASES AND SUGARS.

by

Peter V. Maye

Advisor: Professor David L. Beveridge

Monte Carlo computer simulation studies of the hydration of the five nucleic acid bases and derivatives of ribose and deoxyribose were performed in order to obtain knowledge of the microscopic hydration of these systems. A comparison was made of biomolecular hydration descriptions obtained with the use of three available biomolecular solute-water potential functions. Such a comparison, of primary importance in the improvement of the computer simulation of biomolecular hydration, has not previously been performed. Direct comparison of simulation produced transfer energies with experimentally obtained values for nine simulations distinctly demonstrates that the Berendsen-Van Gunsteren and Kollman sets provide excellent reproduction of transfer energetics in the molecular cases examined; the Clementi potentials, however, were found to greatly overestimate solute-water interactions. Examination of the first shell energetics of thymine and uracil using BVG and Kollman potentials suggest that an average first shell

solute-water pair interaction of about one kcal/mole is reasonable for the nucleic acid bases. Molecular first shell coordination numbers are found to be essentially independent of the choice of potential functions. Functional group coordination numbers do not show definite trends with the variation of potential function and roughly correspond when rounded to the nearest integer. Finally, the simulation results provide qualitative microscopic characterization of the hydration of the five nucleic acid bases and two sugars in derivative form.

Acnowledgements

I express the deepest gratitude to my thesis advisor, Professor David L. Beveridge, for his patient intellectual (sometimes spiritual) guidance on the doctoral path. His deep concern for the scientific development of his undergraduate and doctoral students and postdoctoral researchers is striking and , without doubt, rarely paralleled.

Much gratitude must also be expressed to Dr. Mihaly Mezei, not only as the chief architect and mechanic of the Monte Carlo computer program used to perform these simulations, but also as an always helpful and expert advisor on the technique of computer simulation. The structural and energetic analysis programs used in this work were originally developed by Drs. Prem K. Mehrotra and Francis T. Marchese. The stereo views included in this dissertation were generated by a program written by Dr. Mehrotra and modified by Dr. G.Ravishankar.

The personal and intellectual companionship of fellow Ph.D. candidate Bhyravbhotla Jayaram will always be remembered and cherished.

This dissertation is dedicated to Louis and Lillian Sloane.

Table of Contents

List of Figures

List of Tables

CHAPTER I

INTRODUCTION

Hydration plays a leading role in the structural chemistry of DNA. The structural integrity of a nucleic acid duplex is strongly dependent on ambient humidity. The stacking of DNA bases, which contributes to a regular and rather rigid helical form , is encouraged by hydrophobic base-water interactions. The hydration changes occuring upon base pairing can also be a significant factor in the formation of duplex DNA in an aqueous environment. The secondary structure of DNA is a direct function of ambient humidity. Indeed, the structural regularity of polymers obtained by hydrating the DNA fibers used in early X-ray studies was critical in enabling crystallographers to obtain the sharp diffraction photographs crucial to the subsequent deduction of the three dimensional structure of DNA (1,15).

A complete understanding of the role of hydration in nucleic acid chemistry requires a precise microscopic description of polymer-water interactions in structural and energetic terms. No experimental techniques are currently available which are capable of providing a complete high resolution (on the order of atomic) description of solute-water interactions for even small solute molecules. Statistical theoretical computer simulation techniques, such as the Monte Carlo and molecular dynamics methods, provide data on the required microscopic level by generating high probability solute-

solvent configurations and computing average solution properties from a large series of such configurations. Computer simulation is a relatively new technique, however, and problems still remain. The choice of solute-water and water-water potential functions is a critical factor in the quality of a simulation.

This project focusses on the computer simulation study of the hydration of the nucleic acid constituents as a preliminary to study of polymeric DNA hydration. The nucleic acid bases and sugars provide concrete, well defined and relatively simple biomolecular systems whose structural and energetic hydration is thoroughly characterisable by Monte Carlo simulation. Further, these molecules contain several functional groups in common, facilitating an assessment of the transferability of hydration data among different solute molecules. These results are immediately applicable to the description of hydration changes upon base stacking and pairing. The results are also applicable to the construction of a description of DNA polymer hydration using an aufbau approach.

The data base of nucleic acid component hydration is obtained from Monte Carlo computer simulations of adenine, thymine, guanine, cytosine, uracil, a ribose derivative and a deoxyribose derivative using the Clementi solute-water potential functions. Additionally,

thymine and uracil were simulated using the Berendsen-VanGunsteren and Kollman solute-water potential functions in order to assess the relative characteristics of the various available potential function sets. Extensive structural and energetic analysis were performed on the simulations using the proximity criterion. This analysis provides a complete description, on the solute atomic level, of the hydration character, coordination number and solute-water/water-water energetics of the first and extended hydration shells of the nucleotide bases and sugars. With this information, one can obtain a complete theoretical description of the hydration of nucleic acid constituents, subject to the quality of the assumed potential function.

CHAPTER II

BACKGROUND

Early DNA Structural Studies and Hydration

In 1938, W.T. Astbury (2) studied a dried film of DNA using X-ray fiber photography. This study opened a new era in the biochemistry of the nucleic acids as it was the first such study addressing the 3-dimensional structure of DNA. However, the dried film produced a vague diffraction pattern capable of providing only very approximate information on DNA structure. The studies did provide an indication of 1) the structural periodicity inherent in the molecule, 2) the high length to width ratio of the molecule and 3) the high molecular weight of the molecule.

Around 1950, several researchers (among them Oster and Riley (3), Seeds and Wilkins (4), Fraser and Fraser (5)) began studying moist, hydrated DNA fibers rather than the dried films; this transition enabled crystallographers to obtain the highly detailed diffraction patterns of the macromolecule that were necessary for the ultimate solution of its three dimensional structure. The clarity, hence information content, of the diffraction pattern was dependent on the degree of hydration of the sample; this fact suggested that the structural regularity and definiteness of the macromolecule is a function of hydration.

Preliminary studies on hydrated fibers revealed that DNA fibers equilibrated with air of 50% humidity have some crystalline properties and are well oriented; consequently, Maurice Wilkins was prompted to study hydration mediated physical transformations of the DNA molecule. In 1951, Wilkins (6) described the reversible variations of fiber properties which occurred upon ambient humidity changes. The fibers were decidedly microcrystalline while maintained in air saturated with water vapor; upon reduction of the environment to 50% relative humidity, the crystallinity was markedly reduced. The fibers become increasingly more compact as their length decreased by 30% and the cross section was halved. UV and IR dichroism, as well as birefringence, were reduced. When the fibers were warmed, the resulting dehydration induced further shrinkage in length and cross section and all crystallinity vanished. Evidence thus pointed to large microscopic structural changes occuring in the DNA molecule upon variation of hydration conditions.

In 1953, Linus Pauling (7) developed a triple helical structure for DNA based on the rather poor dried film X-ray photographs of R.Corey. The phosphate groups were un-ionized retaining their hydrogens and, inside the molecule, formed a set of hydrogen bonds which held the triple helix together. The bases were turned outside the helix as a consequence of this arrangement. Pauling

arrived at the triple helical model by using, among other measurements, a DNA density value obtained from anhydrous RNA. Thus, a significant factor in Pauling's erroneous calculations and modeling resulted from his disregard for the influence of hydration on DNA structure.

The hydration of DNA then played a crucial role in the attainment, by R.Franklin and G.Gosling (8), of the first high quality DNA diffraction patterns. Diffraction patterns were obtained at 75% and 92% relative humidity. Distinctly different patterns resulted from fibers at these humidities indicative of a change in DNA structure at some point between the two humidities. The lower humidity form was referred to as the A form and the higher humidity structure was called the B form. The A form was a less crystalline fiber than the B form and more compact in size. Franklin and Gosling proposed that the structural variation was due to the effect of water on the phosphate groups of the DNA backbone. Their data implied that the phosphate groups were on the outside of the structure and that the bases were on the inside. They viewed the phosphate groups of parallel molecules as interacting via the shielding Na+ counterions. In the lower humidity A form intermolecular distances were small, while in the more extended B form intermolecular spaces were filled with water. The high quality data of Franklin and Gosling were of great importance to Watson and Crick in their solution of the three dimensional

structure of the DNA double helix.

Figure II.1 Thymine - atomic numbering system observed in text.

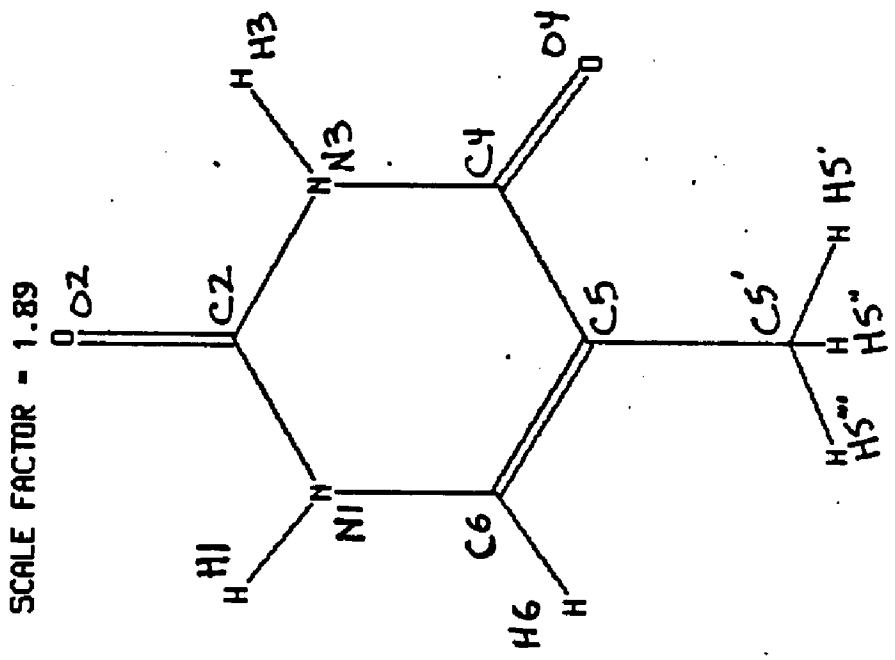


Figure II.2 Uracil - atomic numbering system observed in text.

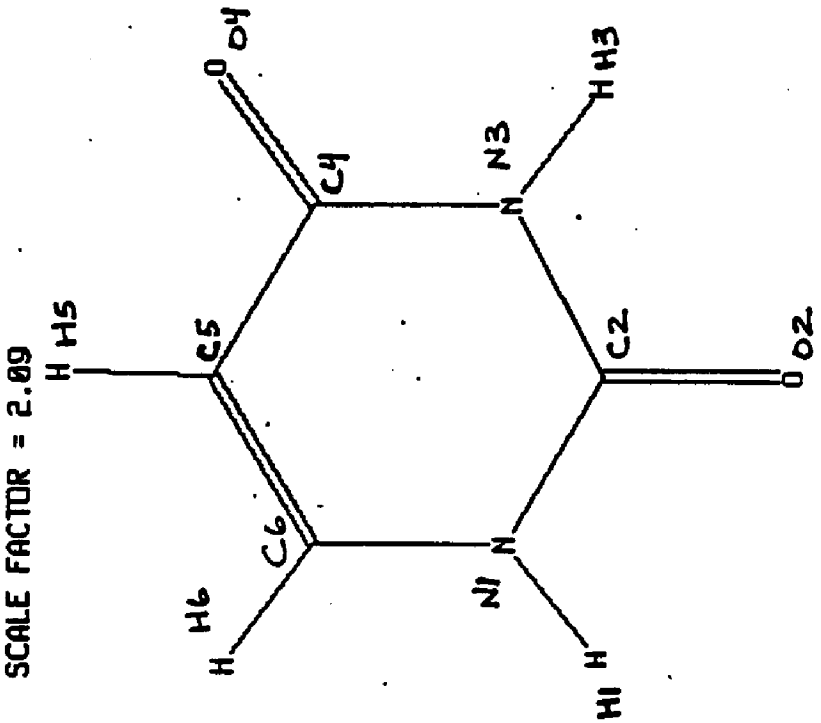
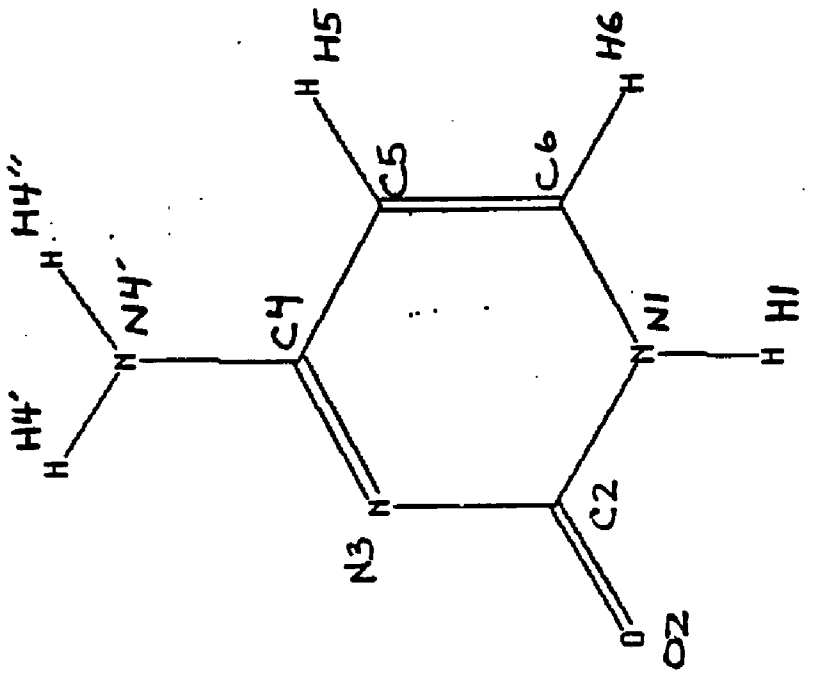
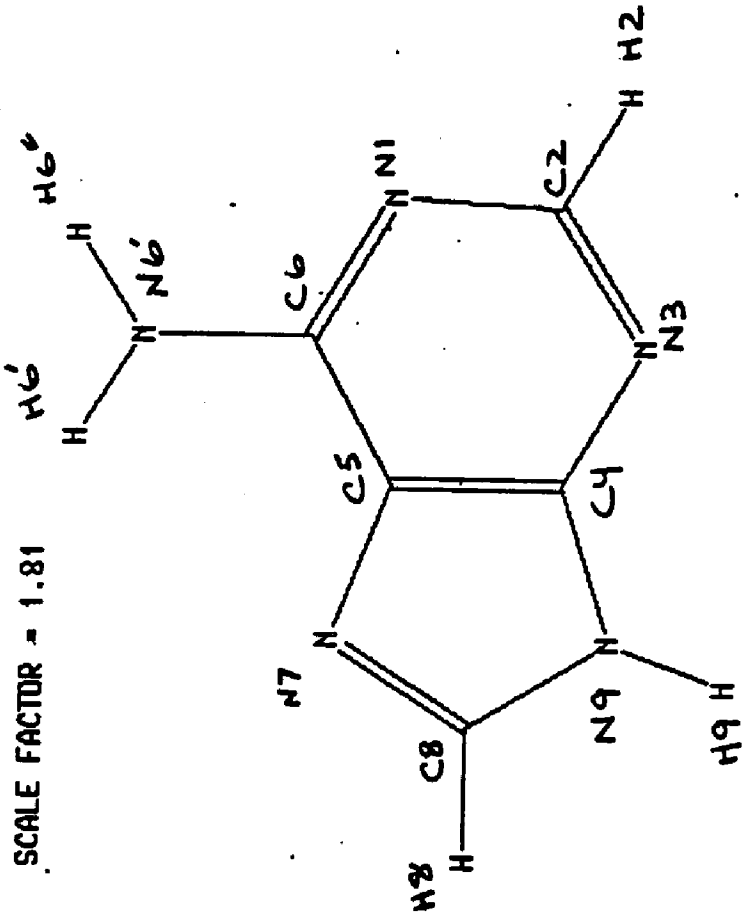


Figure II.3 Cytosine - atomic numbering system observed in text.



SCALE FACTOR = 1.89

Figure II.4 Adenine - atomic numbering system observed in text.



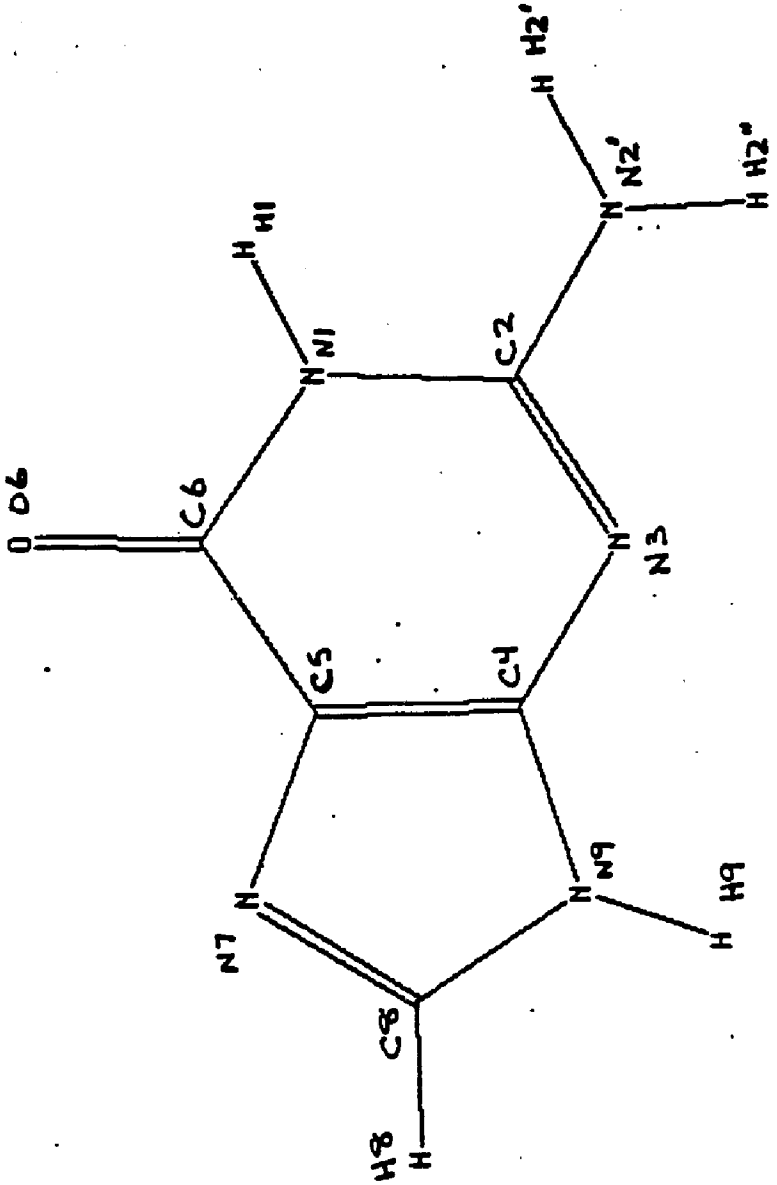


Figure II.5 Guanine - atomic numbering system observed in text.

Figure II.6 ribose derivative - atomic numbering system observed in text.

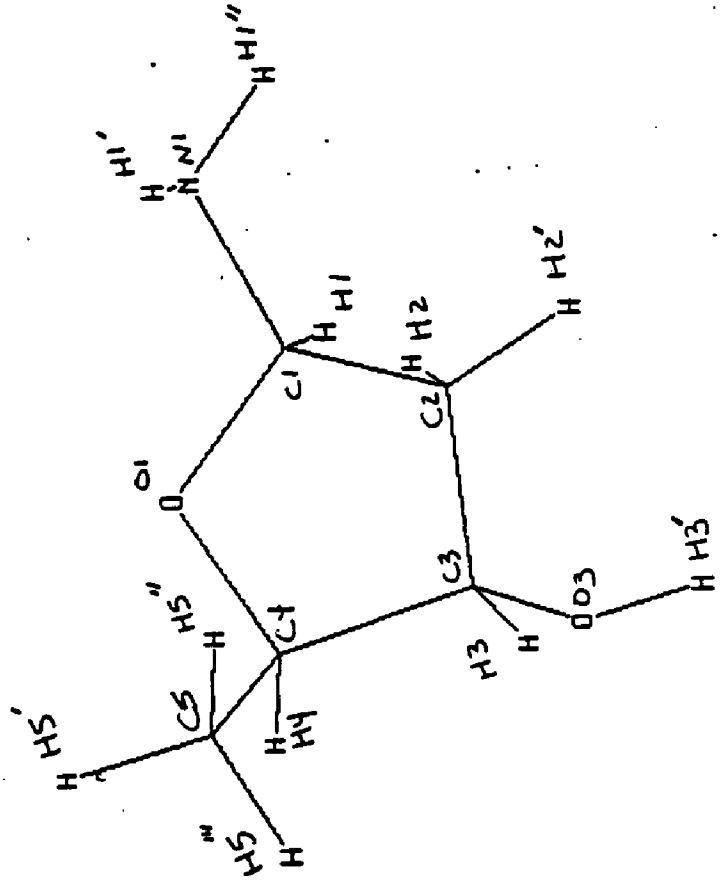
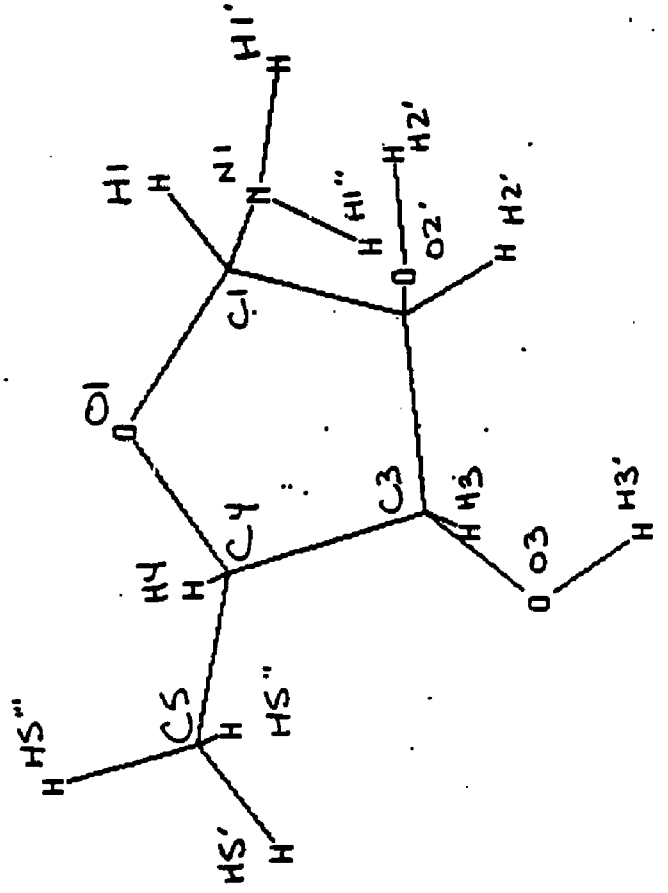


Figure II.7 deoxyribose derivative - - atomic numbering system observed in text.



Experimental Description of DNA Hydration Shells

The atomic numbering of the nucleic acid bases and sugar derivatives, discussed in this chapter and elsewhere in this dissertation, are displayed on the molecular diagrams of Figures II.1 through II.7 preceeding this section.

Diverse evidence, dating from the diffusion studies of Wang (9), has demonstrated that DNA is significantly hydrated and possesses up to two distinct layers of water, the primary and possibly secondary hydration shells. The primary layer is immediately under the influence of DNA, including both site bound waters and waters associated with site bound waters, and displays chemical properties considerably different from those of liquid (bulk) water. A secondary hydration layer, at best barely distinguishable from bulk water in its chemical properties, is also postulated by some authors.

Falk et al. (10,11), by gravimetric and spectroscopic means produced detailed data on the specific hydration of the DNA helix. Their studies showed that DNA hydration proceeds in sequential, sitewise stages correlated with increments in relative humidity of the fiber. In the region of 0% to 65% relative humidity (rh), water is adsorbed to phosphate anion sites on the DNA backbone. As the relative humidity is increased from 65% to 75%,

hydration of the phosphodiester oxygens and deoxyribose sugar ether oxygens occur. Finally, above 75% rh water fills the DNA grooves, associating with the base nitrogens and carbonyl oxygens, completing the specific hydration process. At relative humidities greater than 80% additional water merely swells the DNA fiber and gives no evidence of further specific hydration. The total number of site-bound first shell waters was estimated to be around 11 or 12.

Waters not specifically bound to DNA sites are also included in the first hydration shell. Infrared absorption signals characteristic of bulk water do not occur until 18 to 20 waters associate directly or indirectly with each nucleotide unit. Thus about 8 of 20 waters constituting the first shell of DNA are presumably associated with site-bound waters although not site-bound themselves.

Falk et al (12) later examined infrared difference spectra for DNA in the OH stretching region for levels of relative humidity ranging from 3% to 93%. This spectrum of humidities covers the entire sequence of specific hydration of the DNA helix. Their data showed that waters of the progressive hydration stages have different hydrogen bond strengths; as the relative humidity required for a particular hydration stage increases, the water-nucleotide hydrogen bond energies decrease.

The effect of lower temperatures upon the structure of the DNA hydration shell was also investigated by these authors. Strongly solute bound and ordered hydration shells should not fit easily into ordered icelike structures and should not freeze at the same temperature as bulk water. It was found that cooling a hydrated DNA film has an effect on the IR spectrum (indicative of the formation of ice I regions) only if at least 13 waters per nucleotide unit are present. Below this hydration level, no ice characteristics occur in the spectrum and thus no icelike ordering of water molecules should occur for waters directly surrounding DNA. Upon cooling of DNA films to -10 to -20 C an ice I structure forms.

Any secondary hydration shell of DNA is experimentally indistinguishable from bulk water. Falk speculated that a structure slightly different from bulk water could arise from Donnan equilibrium of counterions between the region of the charged DNA backbone and the outlying bulk solvent. Counterions concentrate close to the DNA helix and effectively empty the second shell region of ions (which could act as an ion permeable Donnan membrane) thereby concentrating waters in this region. Another mechanism which could result in a second hydration shell, suggested by Cohen and Eisenberg (13), is the specific association of counterions with the anionic oxygens of the phosphate groups. The less strongly associated waters of the first shell could then be released to the

second hydration shell.

Recent studies have focused primarily on the structure of DNA single crystals. The most important of these studies in terms of specific hydration of the DNA helix was performed by Drew and Dickerson (14), who examined the hydration of the B-DNA crystal dodecamer CGCGAATTCGCG. Waters hydrating the phosphate backbone were found to be crystallographically disordered;however, in regions near thymine methyl groups, the movement of water molecules was apparently restricted and waters were somewhat better resolved. Especially interesting was the discovery of a "spine of hydration" in the minor groove which is a regular structure of two zigzagged layers of water molecules. The spine bridges the purine nitrogen-3 and pyrimidine oxygen-2 atoms in neighboring base pairs. This bridging is especially definite in the central AATT residues of the dodecamer and would be likely to confer significant structural stability on B form helices rich in A-T regions. The 2-amino groups of the guanine bases tend to fray the hydration spine in the end regions of the dodecamer.

DNA Secondary Structural Types and Hydration

DNA secondary structure falls into three characteristic ranges, known as "families". Within each of these families (A, B or Z), related on the basis of similar structural characteristics, several "forms" are encountered. The secondary structure of DNA is influenced by several factors, especially humidity (hydration) and cation type/concentration (which is also in part a hydration effect since increasing salt concentration at a given humidity decreases water activity.

The A family (including A and A' forms) of DNA helices are encountered under ambient conditions resulting in low water activity: low humidity and moderate to high salt concentrations. RNA structures are found exclusively within this family. This family of helices is characterised by C3'-endo pucker of the deoxyribose sugars. A family structures have base pairs displaced to the periphery of the helix, resulting in an extremely deep (12 A) narrow major groove and a wide, shallow minor groove. In the A form, 11 base pairs per turn with a rise of 2.56 A per base pair yield a helix of 28.2 A per turn.

The B family (including B, a-B', b-B', C, C', D, and E forms) of helices are right handed and characterised by C2'-endo sugar pucker; this family exists under conditions of high humidity. The base pairs are centered on the helical axis resulting in major or minor grooves of approximately equal (about 8 A) depth although the minor groove is narrower. Both grooves are filled with water molecules. The minor groove is apparently lined with a regular "spine" of water molecules, lacing the groove with hydrogen bonds, which presumably contribute significantly to the stability of the high humidity form. The B family is more extended and narrower than A family helices. A B form helix contains 10 base pairs within a helix of pitch 33.8 A. A variant B family helix, the C form, exists under B-family high humidity conditions where salt concentration is raised to 3 molar or more; it is a more compact form than the B-form completing a helical turn in 9.33 base pairs.

The third family of DNA helices are the left handed Z family (including Z, Z' and Z'' forms) helices which arise from an alternating purine-pyrimidine dinucleotide repeat unit under high (> 2.5 M Na+) salt or high ethanol concentration. Six dinucleotide repeat units form a full double helical turn of 45 A. The purine residue is attached to a sugar of C3'-endo pucker with an unorthodox syn-glycosyl torsion angle while the pyrimidine residue residue is linked to a C2'-endo sugar via an anti-

glycosyl torsion angle (15).

Transitions from one family or form of DNA to another can be studied in solution via cirular dichroism, which yields a characteristic spectral shape directly related to the twist and handedness of the helix. In 1970, Tunis-Schneider and Maestre (16) examined the CD and absorbance spectra of oriented and unoriented DNA films. In their study, the relative humidity at which the fibers were maintained was varied. DNA in aqueous solution was found to provide the same spectrum as B-form DNA maintained at > 92% relative humidity. At less than 75% relative humidity a spectrum corresponding to A form DNA was obtained.

Ivanov et al. (17) studied conformational changes of double stranded DNA in solution. DNA dissolved in water, water/methanol, water/ethanol, water/isopropanol and water/dioxane were examined using circular dichroism. The effect on CD spectra of varying counterion types and concentrations was also monitored. They concluded that DNA dissolved in water or water/methanol mixtures exists as a B family structure (B-, C-, or T-DNA). The higher molecular weight alcohol systems (containing ethanol, isopropanol or dioxane) caused conformational transition of the DNA to a B form. These transitions were cooperative, sudden transformations between DNA families.

Further, Ivanov et al. found that variation in cation concentration (as well as cation type) caused a change in the winding angle between adjacent base pairs. As the cation concentration was increased, the winding angle was also increased. This resulted in a continuous, noncooperative conformational change within the B family to C or D forms. Sufficiently high concentrations of salt caused a change to the low humidity A form as expected with a large decrease in water activity.

DNA in aqueous solution, and presumably in vivo, is approximately a B form helix. The exact 10-fold repeat of the helix in the crystal structure is not found in solution; the repeat number of the classic B-form helix is apparently a functions of crystal packing (helix-helix) forces since even extremely wet crystals do not show deviation from this repeat number. Removal of the interhelical interactions upon dissolution of the polymer results in helical unwinding to a 10.3-10.6 base pairs per turn structure. This range is obtained via various means such as the wide angle X-ray studies by Wang (18), the circular dichroism studies of Chen et al. (19) and the theoretical studies of Levitt (20). Since physiological conditions, especially a high level of humidity, most closely correspond to the conditions favorable to the B-form of DNA, it is likely that the majority of in vivo DNA exists in some variation of this form.

Theoretical Studies of Nucleic Acid Component Hydration.

Port and Pullman (21), as part of a quantum mechanical study of environmental effects on molecules, studied the hydration sites of the nucleotide bases. Multipole expansions were used to represent electrostatic term only atomic orbitals in a supermolecule self-consistent field calculation. Base-water interactions were calculated by moving water molecules around the periphery of the base, keeping the water oxygen in the base plane. The water orientation was optimized at each position. A constant 2.85 Angstroms was maintained between the water oxygen and hydrogen bonding base atoms. No approach closer than the sum of atomic Van der Waals radius was permitted for two non-hydrogen bonded atoms.

Three strong interaction sites were found with adenine. The global minimum occurred where a base coplanar water bridged the ring nitrogen N3 and the NH9 imino hydrogen. An energy of -12.0 kcal/mole was found for this interaction. Another minimum (-9.9 kcal/mole) was found with a water bridge between the H6 amino hydrogen and the N7 ring nitrogen; the third interaction was a water bridging amino hydrogen H6' and the ring nitrogen N1 (-9.1 kcal/mole). The minimum interaction energies in the methine hydrogen regions were -5.1 kcal/mole (CH8) and -3.9 kcal/mole (CH2).

Three significant minimum energy interactions were found for guanine. These included an N7/O6 bridge (-11.0 kcal/mole), an NH9-N3 bridge (-10.6) and an open (unbridged) water associating with NH1. A minimum interaction energy of -5.9 kcal/mole was reported for water in the methine H8 region.

The most favorable interaction site in cytosine was an unbridged interaction with the N3 ring nitrogen at an energy of -11.1 kcal/mole. Another favorable interaction site was a water-H1 imino hydrogen association at -9.9 kcal/mole with a water hydrogen pointing at the O2 carbonyl oxygen. A minimum energy of about -4 kcal/mole was reported for the H5 methine-water interaction.

All minimum interaction sites on thymine were found to bridge imino hydrogen and carbonyl oxygen atoms. The minimum energy H1/O2 bridge resulted in a -10.1 kcal/mole interaction, an O2/H3 bridge gives -9.6 kcal/mole and an H3/O4 bridge gives -8.7 kcal/mole. No energies were reported for methine or methyl group interactions.

Three general features of base hydration emerged from their calculations. First, there was a large number of bridged interactions. Second, coplanar base-water interactions were preferred where steric clashes did not prevent such a geometry. Finally, there was a notable absence of repulsive base-water interactions (although no base-water interactions were reported in the region of

the methyl group in thymine).

Clementi and Corongiu (22) reported a theoretical study of the hydration of the nucleic acid bases and base pairs at 300 K. Monte Carlo computer simulations were performed for adenine, cytosine, guanine, thymine and uracil, each with a cluster (no periodic boundary conditions) of 40 water molecules. The number of water molecules in each case was deemed sufficient to comprise an excess of the number required to constitute a first hydration shell although proper simulation of the first shell requires the inclusion of first and second shell water-water interactions; the number of waters used cannot provide a second layer of waters.

An initial equilibration period of 500,000 moves was performed; a subsequent 400,000 to 500,000 moves produced a history file which was used for analysis. Data analysis consisted of the construction of isoenergy contour plots and probability contour plots for each simulated system.

The first shell of adenine was assigned seventeen waters on the basis of the probability contour constructed for the simulation. The NH3 imino group was assigned four waters, the N1, N3 and N7 nitrogens one water each and the amino group was assigned five water molecules. No methine group hydration data was reported. The solute binding energy was -64.09 kcal/mole, the

average water-water binding energy was -6.58 kcal/mole and the energy required to create the cluster solute cavity was 13.0 kcal/mole.

Cytosine was assigned sixteen water molecules as a first hydration shell. Three waters were assigned to the CO2 carbonyl group, six to the amino group, four to the imino group and one to the N3 ring nitrogen. No methine group hydration data was reported. The solute binding energy was -68.30 kcal/mole, the average water-water binding energy was -6.41 kcal/mole and the energy required to create a cluster solute cavity was 15.7 kcal/mole.

The guanine first shell was judged to consist of eighteen waters. The carbonyl group was assigned two waters, the amino group four waters, the NH1 imino group (no data was reported for the NH9 imino group) one water (and said to be perturbing four other waters), and the N3 and N7 ring nitrogens one water each. No methine group hydration data was reported. The solute binding energy was -154.33 kcal/mole, the average water-water binding energy was -5.69 kcal/mole and the energy require to create a cluster solute cavity was -41.5 kcal/mole. The first shell of thymine was found to contain eighteen waters. The CO2 carbonyl was assigned four waters, the CO4 carbonyl group three waters, the NH1 imino group four waters, and the NH3 imino group six waters. Neither

methyl nor methine group hydration data was reported. The solute binding energy was -36.99 kcal/mole, the average water-water binding energy was -6.44 kcal/mole and the energy required to create a cluster energy cavity was 45.8 kcal/mole.

The first shell of uracil was calculated to consist of eighteen water molecules. The CO2 carbonyl group was assigned four waters, the CO4 carbonyl group three waters, and the NH1 and NH3 imino groups three waters each. No data was presented on methine group hydration. The solute binding energy was -77.18 kcal/mole, the average water-water binding energy was -6.43 kcal/mole and the energy required to create a cluster solute cavity was 2.0 kcal/mole.

Del Bene (23), continuing a study of the molecular orbital theory of the hydrogen bond, performed ab initio self-consistent field calculations investigating the structural, energetic and electronic properties of uracil-water dimers. Three cyclic structures were found bridging 1) the NH1 imino hydrogen and CO2 carbonyl oxygen (the second most favorable interactions at -9.7 kcal/mole), 2) the CO2 carbonyl oxygen and NH3 imino hydrogen (-8.0 kcal/mole), and 3) the NH3 imino hydrogen and CO4 carbonyl oxygen (-8.9 kcal/ mole). Three favorable open (unbridged) structures were found at 1) the CO4 carbonyl oxygen (-5.4 kcal/mole), 2) the NH1

imino hydrogen (at -9.9 kcal/mole the most favorable interaction) and 3) the NH3 imino hydrogen (-8.7 kcal/mole).

Since the NH1 open structure and NH1/CO2 bridged structure were easily interconverted, they were termed "wobble" dimers. The open NH3 and bridged NH3/CO4 structures similarly constituted "wobble" dimers. The two wobble dimers were significantly more stable than the open CO4 dimer. The strongest interactions occurred through NH1, rather than NH3, due to more optimal alignments of base and water dipole moments when associations formed at NH1.

Higher order base-water structures were also investigated. Uracil-water trimer interaction energies could nearly be added up directly from the calculated dimer energies. The most stable trimer (-18.7 kcal/mole) consisted of a water bridging NH1 and CO2 and a water bridging NH3 and CO4. Tetrameric uracil-water structures showed a small cooperative effect (tetramer energies lower than the sum of dimer energies) in the hydrogen bond energies. The most stable tetramer (-24.1 kcal/mole) consisted of a NH3/CO4 bridge, an open CO4 bond and an open NH1 bond. Basically, the tetramer is a "wobble" structure while the CO4 open structure is fixed and the other two waters move freely over the other bridged and open structures described for the dimers.

Del Bene also studied (24) the properties of thymine-water complexes. Three stable thymine-water dimers structures were found. The most stable dimer was a wobble dimer consisting of an open NH1 imino hydrogen-water structure (-9.7 kcal/mole) and a bridged NH1/CO2 structure (-9.5 kcal/mole). A wobble dimer consisting of a CO2/NH3 bridged structure, an open NH3 structure and a bridged NH3/CO4 structure is next in stability at around -8.0 kcal/mole. A third structure, formed by an open water-CO4 association, on the CH5 methine side of the carbonyl group, has an energy of only -4.0 kcal/mole. No information was reported on higher order hydration structures.

Poltev, Groknlina and Malenka (25) performed studies on base-water interactions using their own solute-water potential functions. The function set was experimentally derived containing electrostatic (first power) and Lennard-Jones terms (sixth or tenth and twelve power). The tenth power Lennard-Jones term was used with atoms forming hydrogen bonds (oxygen and nitrogen atoms interacting with water atoms). Calculations were reported for the interaction energy of one water molecule with each of the nucleic acid bases. Regional minima were calculated by placing a water molecule at 10 degree angular increments about the base center and then radially and orientationally optimizing the water molecule to find the minimum interaction energy. For

adenine, the minimum was found in the N6/N7H7 region at -10.8 kcal/mole, for guanine in the O6/N7 region at -12.5 kcal/mole, for cytosine in the N3/N4H4 region at -12.0 kcal/mole, for uracil in the N1H1 region at -8.8 kcal/mole and for thymine in the N1H1/O2 region at -8.6 kcal/mole. All global minima in these calculations fall between two hydrophilic functional groups; four of five have a minimum point next to an imino functional group. The authors compare their results for base-water interactions in the metnine regions to those obtained by Clementi et al. The Poltev et al. set yield interaction energies in the range -1 to -3.5 kcal/mole here while the Clementi set yield energies up to -7 kcal/mole.

An attempt to estimate the effect of water-water interactions on base-water interaction energies was also investigated using systems of one base molecule and 2, 3, or 4 water molecules. Each of the two water systems showed several local minima. Some local minima involved water molecules interacting separately with the base; some local minima occurred where the water molecules interacted with each other at substantially favorable (-5 to -6 kcal/mole) energies and bridged two hydrophilic centers on the base. One of these bridging structures always turn out to be the global minimum (except in the case of cytosine). The global minimum with cytosine is a structure with both waters interacting with the base N3 ring nitrogen. The global minima in the bridged

structures range from -17 kcal/mole (uracil) to -20 kcal/mole (guanine). Higher order base-water structures generally resulted in minimum energy structures which mixed bridged hydrophilic centers and single interaction structures. Further information on these higher order structures was not reported; the authors considered it questionablewhether global minima had indeed been found in these cases.

Danilov, Tolokh, Poltev and Malenkov (26) performed Monte Carlo simulation studies of the hydration of uracil. Single, stacked dimer and paired dimer uracil systems were simulated. All simulations were performed at 298 K with the uracil system surrounded by a cluster of 200 water molecules. The clusters were placed inside a sphere of 22 Angstrom radius which was not completely filled by the base-water system. The stacked uracil was moved according to a Metropolis algorithm.

The simulations consisted of a 600,000 step equilibration period and a 500,000 step history production period which was used for analysis. The energetic results reported for the uracil + 200 water molecule system were: total system energy -1591.92 +/- 4.02 kcal/mole; total water-water binding energy -1530.00 +/- 4.00 kcal/mole; uracil-water binding energy -54.1 +/- .1 kcal/mole and and average of 1.75 water-water hydrogen bonds per water molecule.

The stacked base pair structure was found to be more stable than the base pair structure by -16.0 kcal/mole in the total energy, -14.0 kcal/mole in water-water energy, -7.1 kcal/mole in base-water energy and 5.1 kcal/mole in base-base energy. Five more water-water hydrogen bonds were found in the stacked uracil dimer simulation on the average. The 39 water molecules nearest to the dimers were responsible for the majority (>75%) of the energy changes. According to their simulations, the greatest stabilization of the stacked dimer structure in water arises from the decrease in water-water interaction energies around the dimer compared to the separated monomers.

Continuing the Monte Carlo studies of stacked nucleic acid bases, Danilov and Polokh (27) simulated the hydration of two stacked thymine dimers. Two cluster calculations of stacked dimers and 200 water molecules at 298 K were performed. Each system was equilibrated with 600,000 configurations before 500,000 configurations were generated for data analysis. One stacked dimer, termed the A-stack, was an antiparallel stack. The other stack of thymines, termed the O-stack, was generated from the A-stack by rotating one thymine molecule 180 degrees around the N1-H1 bond. This rotation changes the separation of the two methyl groups from 180 degrees to 60 degrees.

The O-stack was found to be the stabler dimer. This stabilization emanated completely from the water-water interactions in this system. The hydration shell of the O-stack contained six more hydrogen bonds than that of the A-stack, resulting in an enhancement of -38 kcal/mole in the water-water binding energy of the O-stack system. A small (+5.4 kcal/mole) destabilization of the O-stack base-water interaction energy accounted for nearly the remainder of the total -34 kcal/mole stabilization of the O-stack structure.

The favored approach of the methyl groups in the O-stack is likened to the association of apolar species in the hydrophobic effect. Most of the -38 kcal/mole stabilization of the O-stack structure is found in the 6.3-6.8 Angstrom layer of waters (-22.5 kcal/mole). The additional decrease in energy occurs throughout the system and suggests a rather long range cooperative structuring of the water molecules.

Theoretical Studies of Polymeric DNA Hydration

Perahia et al. (28) performed a theoretical study of the hydration of DNA using empirical potential energy functions. Water molecules were allowed to explore for potential energy minima among the hydrogen bonding sites of a B-form double stranded DNA trimer. The calculations were performed in two stages. In the first stage, waters were placed at successive hydrogen bonding sites and two of three torsional angles determining the hydrogen bonding geometry were optimized. When all possible hydrogen bonding sites were occupied, the complete hydrogen bonding geometry, including three torsional angles, two hydrogen bond angles and the hydrogen bond distance, were optimized for each water molecule.

The data produced included structural and energetic information on the hydration of the bases and sugars. Two water molecules are found to be very strongly bound. One of these waters bridged a guanine carbonyl O6 and ring nitrogen N7 with a binding energy of -19.6 kcal/mole; the other water associated only with the guanine N7 at -16.1 kcal/mole. Three water molecules were found to bind to ester oxygen on the sugar, two of which simultaneously bound to cytosine carbonyl oxygens. The bridged waters had binding energies of -11.8 and -9.4 kcal/mole; a binding energy of -11.7 kcal/mole was found

for the third water. These energies are quite negative since they are total binding energies including water-water interactions as well as solute-water energies.

The results obtained in this study were observed by the authors to correspond to the ab initio SCF calculations performed earlier by Port and Pullman on the hydration of the free bases. The favorable nature of the O6-water-N7 bridge of guanine agreed with the free base calculations. The number of favorable interactions per bases (one or two with cytosine, three or four with guanine) also appeared to correspond when allowances were made for base-pairing interactions in the polymer.

Clementi and Corongiu (29) began a series of theoretical studies of the hydration of polymeric DNA in 1979 with the publication of isopotential energy maps for the DNA double helix (GTCAGT) in the B conformation. Cylindrical contours were generated at radial distances of 19.84, 20.84 and 22.84 atomic units from the central helical axis. The contours detailed only the especially strong phosphate-water interactions along the DNA backbone. Data explicitly describing base-water or sugar-water interactions was not reported.

Next, Monte Carlo simulation of the DNA single helix in the A conformation was performed at 300 K (30). Two simulations were run including a cluster of the hexamer (GTCAGT) and an 18 water cluster and the hexamer with a

thirty water cluster. A equilibration period of 250,000 moves was followed by a history production period of 300,000 moves of Metropolis Monte Carlo (no convergence acceleration). Water molecules were placed at local energy minima determined from isoenergy contours and constrained to remain between two phosphate groups. The energy of a water molecule was determined from its interaction with the entire DNA molecule and all other waters involved in the simulation. The probability of finding water molecules, at selected gridpoints on planar surfaces perpendicular to the helical axis and coplanar with the phosphate groups of the central (C,A) residues. No explicit information on base or sugar hydration was reported. An identical study of the hydration of a B DNA single helix was performed (31). Again, no information on base or sugar hydration was reported.

Clementi and Corongiu (32) proceeded to the study of the double helical form of DNA. A dodecamer of sequence AACIGACGIAGG was simulated with 447 water molecules; this proportion of DNA to water corresponds to 95% relative humidity. Waters were confined to a cylindrical volume enclosing the DNA helix and coaxial with the helical axis. Periodic boundary conditions were not used. The Monte Carlo history used for analysis was composed of "2000 to 5000 moves per water" after equilibration. This corresponds to a simulation history of between 894,000 to 2,235,000 moves for a 447 water molecule ensemble;

unfortunately, the authors do not report the duration of the simulation any more specifically.

Graphs of the density of (water) oxygens plotted versus distance from various DNA atoms and functional groups were constructed for the simulation. Only average distributions displaying composite data obtained from all fuctional groups or atoms of a particular type were reported in the publication. Thus, the graph of the distribution of water oxygens versus distance from the ring nitrogen atoms was a composite distribution from the four N3 and N7 atoms of both adenine and guanine. Graphs of water hydrogen number versus distance from the atom or functional group were also prepared.

A water molecule was judged as bound to a particular atom if 1) the atom-water oxygen distance fell within a prescribed cutoff value and 2) if the atom-water hydrogen distance fell within a prescribed cutoff value. However, the authors did not specify how these cutoffs were determined or what the cutoff values were. The values probably are taken as the first minimum in each of the distributions mentioned above. If this is the case, then the boundary of bound waters for a specific atom is determined from the average distribution for all atoms or functional groups of that type. These average distributions are prepared for the sugar O1' atoms, the ring nitrogens (N3 and N7) of adenine and guanine, the

carbonyl groups of thymine (O2 and O4), cytosine (O2) and guanine (O6) and the amino groups of adenine, cytosine and guanine.

The total water populations and interaction energies for waters bound to each type of functional group or atom (e.g., all waters bound to the five guanine amino groups) were then calculated. These values, converted to average number of waters per site per base and the corresponding average interaction energy per water, follow. Also indicated is whether a site is partially involved in base-pairing and whether a site is in the major or minor groove. Each sugar ether oxygen has .80 waters assigned at an average energy of -1.54 kcal/mole. The adenine amino groups are assigned an average 1.11 waters at -3.63 kcal/mole each; the cytosine amino groups 1.97 waters at -3.82 kcal/mole each; the guanine amino groups .97 waters at -4.61 kcal/mole each. All amino groups have one hydrogen atom which are involved in base-pairing. The adenine and cytosine amino groups are situated in the helical major groove while the amino group of guanine is in the minor groove. The adenine N3 ring nitrogens are assigned .45 waters interacting at -8.96 kcal/mole each; the N7 atom is assigned .19 water at -21.20 kcal/mole each. The guanine N3 ring nitrogens are bound to .51 waters interacting at -9.24 kcal/mole each; the N7 atom is assigned .23 waters at -17.20 kcal/mole each. Both N3 nitrogens are involved in base pairing; the N7 ring

nitrogens occupy the major groove and the N3 atoms occupy the minor groove.

The carbonyl groups of thymine interact with .40 waters at -13.04 kcal/mole each (O2) and .92 water at -4.18 kcal/mole each (O4). The cytosine O2 carbonyl is assigned an average of .40 waters at -11.77 kcal/mole while the guanine O6 carbonyl is assigned 1.03 waters at -4.0 kcal/mole each. The thymine O4, cytosine O2 and guanine O6 atoms are involved in base-pairing. Thymine O4 and guanine O6 are in the helical major groove while thymine O2 and cytosine O2 are in the minor groove.

Some data on the hydration of the apolar sugar and base hydrogens is also presented. The average number of waters per hydrogen and average interaction energy per water for the sugar hydrogens are: H1': .65 waters/-1.90 kcal/mole; H3 and H4: 1.44 waters/-.41 kcal/mole; H2 and H2': .58 waters/-1.08 kcal/mole. The corresponding values for the base hydrogens (no further distinction is made by the authors) are: adenine: .32 waters/-6.83 kcal/mole; cytosine: .96 waters/ -2.23 kcal/mole; guanine: 1.00 waters/-4.98 kcal/mole; thymine: .50 waters/-2.26 kcal/mole.

The next study in the series (33) simulated B-DNA and 447 waters under identical conditions to the previous simulation except that 22 Na+ counterions were placed near the charged oxygen atoms of the backbone phosphate

group. The counterions were not moved in the Monte Carlo simulation.

A general increase of bound waters was noted for the polar sugar ester oxygens and base amino and carbonyl groups and ring nitrogens relative to the B-DNA simulation omitting counterions. This corresponded to the expected electrostriction of surrounding water by the charged ions. The average number of waters bound per group were: sugar ester O1': .84; -NH2 (adenine, guanine, cytosine): 1.25; ring nitrogen (adenine, guanine): .41; carbonyl (cytosine, guanine, adenine) .65. Large interaction energies for the waters were also noted; this is due to the addition of Na+-water interaction energies to the B-DNA-water interaction energies.

The next simulation experiment by Clementi and Coronqiu (33) included counterions which were moved. A ten base pair sequence of b-DNA (ACIGACGIAG) was surrounded by 20 Na+ counterions and 400 water molecules and the entire system was contained within a cylinder coaxial with the DNA helical axis. The central simulation cell was surrounded at top and bottom by identical images of itself; each time a water or counterion was moved, its corresponding top and bottom image was moved identically. Interaction energies for a water or counterion were calculated within a unit cell

centered on the water or counterion; edge effects were thus minimized. No information was presented on simulation length except that "the initial 1,000,000 configurations were disregarded".

An increase in the average number of waters bound to polar ring and sugar atoms occurs relative to the fixed counterion simulation. The average number of bound waters found per atom or group are: sugar ester O1': 1.16 waters; -NH2 group (adenine, guanine, cytosine): 1.91 waters; ring nitrogen (adenine, cytosine): 0.62 waters; carbonyl group(cytosine, guanine, thymine): 1.03 waters.

A final Monte Carlo simulation study of the hydration of B-DNA (35) examined the conterion structure around hydrated DNA at different humidities, ionic concentrations and temperatures. The simulation conditions were identical to the previously described simulation including mobile counterions. No notable new data was produced on the hydration of the bases or sugars.

CHAPTER III

THEORY AND METHODOLOGY

Statistical Thermodynamics

A liquid may be described theoretically in terms of an ensemble of N molecules, contained in a volume V, maintained at a temperature T. This so-called T,V,N ensemble forms the basis of a theoretical description of the system via the semi-classical canonical ensemble partition function Q(T,V,N):

$$Q(T,V,N) = (q^N/(8\pi^2)\Lambda^{3N} N!)\ Z(T,V,N)$$

where $$Z(T,V,N) = \int \cdots \int \exp(-E(X^N)/kT)\ dX^N$$

where q is the partition function for the internal degrees of freedom of the molecule, Λ is the one-dimensional translational partition function for each molecule and $E(X^N)$ is the energy of each configuration. The indicated integration is ideally carried out over all possible configurations of the system.

The ensemble average of any configuration dependent quantity follow from the partition function as

$$\bar{A} = \int \cdots \int A(X^N)\ P(X^N)\ dX^N$$

where $$P(X^N) = \exp(-E(X^N)/kT)/\ Z(T,V,N)$$

is the Boltzmann probability of each configuration X^N of the ensemble of molecules. The expectation value for the excess internal energy, $\bar{U}$, of the system is

$$\bar{U} = \int \cdots \int U(X^N)\ P(X^N)\ dX^N$$

and the pressure is

$$p = (kT/V)\ (N - \langle \sum_{i=1}^{N} (R_i \cdot \partial E(X^N)/\partial R_i) \rangle)$$

where the R are the coordinate vectors of each of the N molecules.

The excess constant volume heat capacity is often a quantity of interest. This quantity is related to the average fluctuation of the excess internal energy and takes the form

$$C_v = (\langle E(X^N)^2 \rangle - \langle E(X^N) \rangle^2) \ / \ kT^2 .$$

Monte Carlo Simulations and the Metropolis Method

Liquid state computer simulations proceed via numerical evaluation of the expressions for the various thermodynamic quantities. For instance, integration of the expression of the internal energy of the ensemble proceeds by averaging the values obtained for the internal energy equation over a very large number of molecular arrangements. A problem arises, however, in the direct evaluation of the integral for a completely random choice of configurations. The integrand $E(X^N) \cdot P(X^N)$ is large only in a very restricted range of configurational space and thus the "integral" converges extremely slowly. In other words, molecular configurations of significant probability (generally, lower energy) are relatively few, and with random sampling would be encountered infrequently. Such a calculation is expected to be intractable in a computer simulation.

A comprehensive treatment procedure is made possible by biasing the sampling of molecular configurations in the direction of the highest probability configurations. Sampling molecular configurations directly with a rate proportional to their Boltzmann probability achieves exactly this goal. This method, due to N.Metropolis et al., is entitled the Metropolis Monte Carlo algorithm

(MMC). It is implemented by taking the N-molecule configurations of the ensemble as the states of an irreducible Markov chain. The probability of transition between state k and state l of the chain are written

$$p = Pr(X_{t+1}^{N} = l \;/\; X_{t}^{N} = k)$$

Collection of all transition probabilities p into a matrix P forms the stochastic transition probability matrix. A stochastic walk through configurational space provides molecular configurations for the Monte Carlo integration. In the Metropolis method, the unique limiting stationary distribution p is the Boltzmann distribution:

$$\Pi = \Pi P \quad \text{where} \quad \Pi_i = P(X_i^N)$$

In an actual simulation, a molecule is selected for movement according to some sampling procedure. The molecule is then translated and rotated, generating a new molecular configuration. The quantity

$$R = \exp(-(E(X_{\beta}^{N}) - E(X_{\alpha}^{N}))/kT)$$

is then calculated. This quantity is the Boltzmann probability for the energy change. R is compared with a number randomly generated from the interval (0,1). If R is greater than the number generated, or greater than one, the move is rejected. Configurations are

accumulated with a frequency proportional to their individual probabilities $P(X^N)$: Average properties are then simply calculated as a summation over the properties of the individual configurations X^N:

$$\overline{E(X^N)} = (1/M) \sum_{T=1}^{M} E(X_T^N)$$

As the number of accepted configurations increases, the accumulated $E(X^N)$ approaches the ensemble average $\langle E(X^N) \rangle$ (36).

The Monte Carlo simulation of an aqueous solution begins by randomly placing water molecules around a solute molecule. This arbitrary configuration is extremely likely to have a very high energy (and correspondingly low Boltzmann probability), and the configurations accepted immediately following this starting configuration are likely to contribute vanishingly small amounts to the desired configurational integral. Thus, in a typical Monte Carlo simulation, the first 500,000-1,000,000 (500K-1000K) perturbations of the system are discarded. This part of the simulation is termed equilibration; during this period the configurations selected are of steadily increasing Boltzmann probability. When the energy settles down and begins to fluctuate around a steady mean value, a history of accepted configurations is collected. This proceeds for another 1000K-2000K perturbations of the system until key thermodynamic quantities (e.g., internal energy)

reach stable values. The simulation is now termed converged and the various calculated values of the system are considered to be valid. Actually, the number of configurations and amount of computer time required for simulation is appreciably greater than described above if certain convergence acceleration techniques are not implemented. These algorithms, specifically force biasing and preferential sampling, are discussed in the following section.

Convergence Acceleration

Previous work has demonstrated that biasing the one molecule perturbations constituting a Monte Carlo simulation of a liquid in the direction of the resultant of the forces exerted on the perturbed molecule appreciably accelerates simulation convergence (37). In crude Monte Carlo simulation, the probability of selecting any perturbed configuration is constant in the sampling domain. In a liquid state simulation, the sampling domain is the maximum displacement and maximum rotation allowed a molecule in a perturbation. That is:

$$\pi(R \rightarrow R') = \begin{cases} \text{constant} & \text{if } R' \in D(R) \\ 0 & \text{if } R' \notin D(R) \end{cases}$$

where $\pi(R \rightarrow R')$ is the transition probability for the displacement of one molecule from position R to position R'; where D(R) is some sampling domain. In the force bias method, $\pi(R \rightarrow R')$ is taken as:

$$\Omega \exp(\lambda(F(R)\cdot(R-R') + \tau(R)\cdot(W-W') / k\tau)) \quad \text{where } R', W' \in D(R,W)$$

$$\pi(R \rightarrow R') = \begin{cases} 0 & \text{where } R', W' \notin D(R,W) \end{cases}$$

Here, F(R) is the net force exerted on the molecule moved, $\tau(R)$ is the net torque exerted on the molecule

moved

and Q is a normalization constant depending on F(R) and I(R). The constant λ is an optimization parameter; R-R' is the displacement and W-W' is the rotation of the molecule moved. Application of the force biasing technique can reduce simulation time approximately 50%.

The method of preferential sampling makes additional convergence acceleration possible. In regular and force-biased Monte Carlo procedures, all solvent molecules are selected for displacement with equal probability. Generally, however, the properties of central interest in molecular solutions involve solvent molecules close to the solute. When preferential sampling is invoked, the probability of selection of a solvent for displacement increases near the solute. In this technique the perturbation frequency of first shell molecules is typically increased from 10% or 20% to as much as 50% of the total molecular perturbations to obtain optimal convergence.

Another useful convergence acceleration technique is the preferential sampling technique; the following formulation is due to Owicki (38). A weighting function w(R) is constructed which is some non-negative, decreasing function of R, where R is the solute-solvent distance. A probability distribution function is defined:

$$W(X_k^N) = \{W_2(X_k^N), W_3(X_k^N), \ldots, W_N(X_k^N)\}$$

as a set of probabilities for the individual solvent molecules labeled by the subscripts. The individual molecule probabilities are:

$$W_m(X_k^N) = w(R_m) \Big/ \sum_{i=2}^{N} w(R_i)$$

where the $w(R_{\ })$ are the previously described weighting functions evaluated for the distance R from the solute of each solvent m. The probability distribution $W(X_k^N)$ is sampled in order to select the solvent molecule to be moved. The preferential sampling technique has been demonstrated to significantly improve the correspondence of water-water g(R)s to experimentally obtained results. In practice, this technique is implemented in conjunction with Metropolis sampling and the force bias technique.

Analysis of Simulation Results

A Monte Carlo computer simulation of an aqueous solution consists of a series of molecular configurations each containing structural and energetic information. Analysis of the simulation proceeds via a configuration by configuration calculation of the desired quantities followed by calculation of simulation averages for these quantities and the construction of the corresponding distribution functions.

The proximity criterion (39) is central to the analysis of the simulations performed in this study. This criterion provides a description of solute hydration by assigning a water molecule to the solute atom to which it is closest. The procedure is formally equivalent to the assignment of a water molecule to a solute atom based on the inclusion of the water molecule within the limits of the Voronoi polyhedron constructed around that atom; the Voronoi polyhedron centered on an atom is simply the points in the space of the simulation cell closest to the particular atom.

The atom-water radial distribution function (g(r)) is a description of the fluctuations of water density as a function of distance from the particular atom. The total atom-water g(r) is derived by including all the water molecules in the simulation cell; the primary atom-water g(r), which is generally of much greater interest, is

calculated by including only the waters assigned to a particular atom by the proximity criterion. In general, a g(r) is calculated using the formula

$$g(R) = N(R) / \int \rho 4\pi R^2 dR = \rho'(R) / \rho$$

where N(R) is the average number of particles in a spherical shell of width dR at a radial distance R from the solute atom, ρ is the density of bulk solvent and ρ' is tne solvent density at a distance R from a solute atom (or solute center of mass).

Ihe first coordination snell of an atom is generally chosen as the distance from the atom where the first minimum in the atom-water primary g(r) is found. Ihe first shell coordination number in any configuration is just the number of water molecules contained within the atomic first shell cutoff in a configuration.

Ihe variation of the coordination number over a Monte Carlo realization is described by the quasicomponent distribution function (QCDF) of coordination number written as

$$x_c(K) = n(K) / \sum_{K=0}^{\infty} n(K)$$

Here, n is the number of simulation configurations with coordination number K and n(K) is the total number of configurations composing the simulation. Ihe numbers $x_c(K)$ are termed the quasicomponents of the coordination

number and (ideally) correspond to the mole fraction of solute molecules with the coordination number K for this atom. The average coordination number of an atom is related to the quasicomponents of coordination number by the equation

$$\bar{K} = \sum_{K=0}^{\infty} K x_c(K) = \int_0^{R_c} g(r)\, 4\pi R^2 dR$$

which also shows the relationship of the average coordination number to the primary g(R) of the atom.

The quasicomponent of atom-water pair energy in the range e to e+de is calculated from the expression

$$x_p(e) = \left\langle \sum_{i<j}^{N} (e_{ij}(X^N) - e) C_{ij}(X^N) \right\rangle / \sum_{i<j}^{N} \langle C(X^N) \rangle$$

where $e(X_{ij}^N)$ is the pair energy between water molecule i and solute molecule j and C_{ij} is a function evaluated as one if the distance between i and j is less than a cutoff distance R and zero otherwise. If R_c is the first shell cutoff of the atom to which the water molecule is assigned by proximity, then the distribution is calculated for first shell pair energies. Energetic QCDFs are continuous functions rather than discrete functions as encountered with the QCDF of coordination number. A graph of a pair energy QCDF generally contains one or more peaks reflecting one or more types of structural/energetic interactions.

The binding energy of a molecule i, in some N-molecule

configuration, is the quantity

$$B_i(X^N) = E(X_1, \ldots, X_{i-1}, X_i, X_{i+1}, \ldots, X_N) - E(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_N)$$

Then the expression

$$x_B(v) = \langle \delta(v_i(X^N) - v) C_{ij}(X^N) \rangle / \sum_{i<j}^{N} \langle C_{ij}(X^N) \rangle$$

defines the quasicomponent of binding energy for molecules with a binding energy in the range v to v+dv. The average binding energy is given by

$$\bar{v} = \int_{-\infty}^{+\infty} v x_B(v)\, dv$$

Any structural or energetic averages or QCDFs which have been calculated for atoms using the proximity criterion can be combined by simple addition to form any desired averages or QCDFs for any composite atomic structures of the molecule. Thus, distribution functions for coordination numbers and pair and binding energies are calculated for functional groups such as amino, carbonyl and methyl groups and molecular fragments such as methine purine ring fragments. Finally, averages and distribution functions for the entire molecule can be constructed providing a picture of solute hydration built from microscopic (atomic) results.

Solute-Water Potential Functions

The solute-water and water-water potential function set chosen to calculate intermolecular energies and forces in the simulation of an aqueous solution determine the quality of the simulation. Although comparisons have been made of widely used water-water potential function sets (40), no study has yet been published comparing available solute-water potential functions. Three solute-water potential function sets especially designed for biomolecular calculations include those developed by 1) Clementi and coworkers, henceforth referred to as the Clementi set, 2) Kollman and coworkers, henceforth referred to as the Kollman set and 3) Berendsen and van Gunsteren and cowokers, henceforth referred to as the BVG set. These sets have the common feature of atomic pairwise addition of interaction energies, i.e., the energy of interaction between the solute molecule and water molecule is calculated as

$$E(M,W) = \sum_{i,j} E_{ij}(M,W)$$

where $E_{ij}(M,W)$ is the interaction energy between atom i of the solute molecule M and atom j of the water molecule W. Also, each solute-water potential function set is designed to be used with a specific water-water potential function set.

The Clementi potential function set (41-43) was

calculated from SCF-LCAO-MO ab initio calculations for each of the 21 amino acids and four nucleic acid bases and one water resulting in 368 computed energies which were fit to an analytical potential of the form

$$E=\sum_{i\neq j}(-A_{ij}^{ab}/r_{ij}^{6}+B_{ij}^{ab}/r_{ij}^{12}-C_{ij}^{ab}q_iq_j/r_{ij})+E(\text{amino acid or base})+E(\text{water})$$

The i and j indexes range over the solute atoms and water atoms respectively and a refers to the electronic environment of the solute atom while b depends on whether the water atom is an oxygen or a hydrogen. A, B and C are fitting constants which are calculated for each a and b combination; E(amino acid or base) and E(water) are the total energies of the base or amino acid and water molecules. The Clementi solute-water potential function set is expressly designed for use with the MCY-CI water model and water-water potential (44) function set.

Kollman et al. (44,45) designed an analytical potential function set for the calculation of biomolecule intramolecular conformations. Their function included provisions for the calculation of nonbonded interactions including solute-water interactions. The functional form is

$$E=\sum_{i\neq j}(-A_{ij}^{ab}/r_{ij}^{6}+B_{ij}^{ab}/r_{ij}^{12}-q_iq_j/r_{ij})$$

The A and B parameters are calculated for specific solute atom-water atom interactions from

$$A_{ij}^{ab} = (A_i^a A_j^b)^{1/2} \quad ; \quad B_{ij}^{ab} = (B_i^a B_j^b)^{1/2}$$

where A and B are empirically determined. The Kollman set is used with the TIP3P water model and water-water potential function set (47).

The Berensen-VanGunsteren potential function set (48) uses the formula

$$E = \sum_{i \neq j} (-A_{ij}^{ab}/r_{ij}^6 + B_{ij}^{ab}/r_{ij}^{12} - q_i q_j / r_{ij})$$

for the solute-water oxygen energies and the equation

$$E = \sum_{i \neq j} (-q_i / r_{ij})$$

for the solute-water hydrogen energies. A method of determining solute-water energy cutoffs is also specific to this set. The solute molecule is divided into local charge groups of 2, 3 or 4 atoms each. The partial atomic charges within each charge group are chosen so that their sum is equal to zero. If the distance between the center of geometry of a charge group and the oxygen of a particular water molecule is greater than eight Angstroms, the interaction energy of the water and that fragment of the solute molecule is taken to be zero. The BVG solute-water potential function set is designed to be used with the SPC water model and water-water potential function set (49).

The calculations performed in this project all use periodic boundary conditions to minimze cell edge effects in the calculation of interaction energies. With periodic boundary conditions, the simulation cell, containing solute and solvent molecules, is completely bounded by identical cells. These cells are exact replicas of the central cell in terms of dimensions and placement of solute and solvent molecules; each time a molecule is displaced in the central cell an identical displacement occurs in all other images. Thus, a simple cubic cell is surrounded by 22 images: six at each face, eight at each edge and eight at each corner. When interaction energies are calculated for a molecule at the edge of a cell, interactions with molecules in the next cell are included; thus, no sudden truncation of energies occurs at the edge of the box. In practice, the interaction of a water molecule with the solute is calculated only within an image of the simulation cell centered on the water molecule center of mass; this is the "minimum image cutoff" and results in calculation of a water-solute pair energy including interaction with one (and only one) solute molecule. The water-water interaction energy cutoff is a spherical cutoff; if the center of masses of two water molecules do not fall within 7.75 Angstroms of each other, their interaction energy is taken as zero.

CHAPTER IV

RESULTS AND CALCULATIONS

Systems Studied.

Monte Carlo computer simulation studies were performed on aqueous solutions of the five common nucleic acid bases and derivatives of the ribose and deoxyribose sugars. These molecules constitute all nucleic acid components except the phosphate moiety. Aqueous solutions of cytosine, guanine, uracil, thymine and adenine, each with 215 waters, were simulated using the Clementi solute-water potential functions. Additionally, thymine and uracil solutions were simulated using the Kollman solute-water potential functions and the Berendsen-VanGunsteren solute-water potential functions. (C3-endo)-5-deoxy-1-C-amino-B-D-ribo-pentofuranose and (C2-endo)-2,5-dideoxy-1-C-amino-B-D-ribo-pentofuranose molecules and 202 waters were simulated using Clementi solute-water potential functions. These compounds will henceforth be referred to as ribose derivative and deoxyribose derivative, respectively.

General Simulation Characteristics

At the practical level, a simulation involves the selection of intermolecular potentials, the Monte Carlo realization and analysis of the calculation. General characteristics of each of these phases of a calculation are collected herein. Specific results will be given in the following section.

Potential Energy Contour Surfaces. Isoenergy contour surfaces for the interaction of a solute molecule and one water were calculated and drawn for each simulation using the appropriate solute-water potential function. The contours are generated by selecting a set of grid points on a molecular plane separated by .2 Angstrom intervals. A water molecule is placed at a grid point and its energy of interaction with the solute molecule is calculated as the water orientation varies. When an orientation of minimum energy is found, this energy is recorded. The calculation is repeated until all grid points have assigned energies. Lines are then drawn to connect continuous isoenergy points on the grid plane using the computer program "CONSURF".

Simulation Details. Cell dimensions are calculated from the volume of the number of water molecules used in the simulation and the partial molar volume of the solute molecule. Experimental partial molar volume values are

not available for the systems studied here. Thus, values for the partial molar volume of the nucleic acid bases and ribose and deoxyribose derivatives were obtained using the additivity rules of Kiyohara et al. (60) and Mishra et al. (61).

Geometries for all bases were obtained from the crystallographic work of Spenser. The atomic charges used in all base simulations were obtained from Clementi et al. except for the charges used for thymine and uracil in the Kollman and BVG solute-water potential function simulations; these charges were obtained with the Gaussian 80 computer program using an STO-3G basis set.

Geometries for the ribose and deoxyribose sugar derivatives were obtained from W.Olson. These geometries lacked explicit coordinates for the amino, methyl and hydroxyl hydrogen atoms; the coordinates for these atoms were determined and optimized using the MM2 molecular mechanics package available on the PROPHET computer system of Bolt, Beranek and Newman. Atomic charges for these structures were obtained by using the Gaussian 80 program with the Clementi basis set.

All simulations were carried out with a initial equilibration portion consisting of a minimum of 600,000 force biased moves, which was discarded. Translational and orientational displacements of the solute and solvent molecules were set to produce an approximately 50%

acceptance rate. An additional 600,000-2,000,000 force-biased steps were generated to produce a Monte Carlo history which was used for subsequent analysis.

Convergence of the simulation control functions, especially the mean internal energy, to stable values is the criterion for judging the length of a simulation as adequate. Figure IV.1.2 provides a typical plot of simulation control functions. The lower plots are the averaged internal energies. The curve without markings plots the calculated average internal energy as the simulation progresses. The superimposed curve marked with crosses plots the average internal energy over sequential 50,000 step blocks. Fluctuation of this quantity around the evolving mean energy is an important characteristic and reflects the sampling of a range of energies around the mean value. The simulation mean energy converges to a stable value and thus average energy values calculated from this simulation are acceptable. Another quantity which can be monitored as an index of convergence is the excess internal heat capacity. The heat capacity is plotted in the upper curve of Figure IV.1.2. This value typically requires a greater simulation length (up to 5,000,000 steps) to attain a stable value than the internal energy since it is calculated from fluctuations in the internal energy.

Thermodynamic Results. Table IV.1 contains various

Table IV.1 Calculated ensemble thermodynamic quantities for Monte Carlo simulations of aqueous nucleic acid constituents at 298 K.

MONTE CARLO MEAN ENERGY SIMULATION RESULTS ON NUCLEIC ACID CONSTITUENTS IN WATER AT 25 C

ENERGETICS OF HYDRATION OF NUCLEIC ACID CONSTITUENTS

POTENTIAL FUNCTION	THYMINE CLEMENTI	THYMINE KOLLMAN	THYMINE BVG	URACIL CLEMENTI	URACIL KOLLMAN	URACIL BVG	CYTOSINE CLEMENTI	ADENINE CLEMENTI	GUANINE CLEMENTI	RIBOSE DERIVATIVE	DEOXYRIBOSE DERIVATIVE
⟨U⟩	-1881.2	-2136.3	-2224.8	-2009.5	-2131.1	-2211.1	-1982.2	-2001.2	-1972.4	-1750.2	-1756.1
+/- 2*SD	3.8	5.5	3.9	8.1	5.6	5.4	6.2	11.9	4.6	6.9	6.4
⟨USLT'⟩	-7.4	-23.2	-28.6	-137.4	-20.9	-27.3	-108.1	-111.9	-120.0	-1.7	-23.2
+/- 2*SD	0.9	0.6	0.9	0.9	1.0	0.5	1.4	1.6	1.0	1.6	1.8
⟨UVV'⟩	-1873.8	-2113.1	-2196.2	-1872.1	-2110.2	-2184.1	-1879.1	-1889.3	-1852.4	-1748.4	-1732.9
+/- 2*SD	3.7	5.5	7.8	8.0	5.5	5.4	6.0	11.8	4.5	6.7	6.1
⟨UVV⟩	-1859.8	-2119.9	-2188.7	-1859.8	-2119.9	-2188.7	-1859.8	-1859.8	-1859.8	-1747.3	-1747.3
+/- 2*SD	6.4	6.4	6.4	6.4	6.4	6.4	6.4	6.4	6.4	6.1	6.1
⟨UREL⟩	-14.0	6.8	7.5	-12.3	9.7	4.6	-19.3	-29.5	7.4	-1.1	14.4
+/- 2*SD	7.4	8.4	10.1	10.2	8.4	8.4	8.8	13.4	7.8	9.1	8.6
⟨USLT⟩	-21.4	-16.8	-21.1	-149.7	-11.2	-22.7	-127.4	-141.4	-112.6	-2.8	-8.8
+/- 2*SD	7.5	8.4	10.1	10.2	8.5	8.4	8.9	13.5	7.9	9.2	8.8

thermodynamic quantities for the simulations of nucleic acid constituent hydration. The quantity labeled <U> is the total ensemble binding energy (containing solute-water and water-water contributions) calculated. The quantity <UWW'> is the water-water binding energy only and the quantity labeled <USLT> is the solute-water binding energy only; the sum of these two quantities is <U>. <UWW> is the total water-water binding energy obtained from a simulation of pure water using the same water-water potential function as the corresponding simulation reported here. The difference between <UWW'> and <UWW> for each simulation is the energy required to open the particular solute cavity in pure water; this energy is listed in the row labeled <UREL>. Finally, <USLT> is the vacuum to water solute transfer energy and is the sum of the solute-water binding energy <USLT> and the solute cavity energy <UREL>.

Analysis of Results. Atomic first shell quantities are computed by determination of a first shell radial cutoff for a particular atom and calculating the desired property for the water molecule within the limits of the Voronoi and first shell cutoff for that atom. A reasonably consistent manner of defining the radial cutoff of an atom is necessary for consistent property calculations. Examination of Figure IV.1 shows the g(r) functions and running coordination numbers for the N1 atom of thymine; labels for this graph are found on the

Figure IV.1 - total and primary g(r) functions and corresponding running coordination numbers for N1 nitrogen atom of thymine.

X axis - distance from Ni nitrogen atom (Angstroms).

Left Y axis - g(r).

Right Y axis - running coordination number.

Curves without crosshatches - curves for total g(r).

Curves with crosshatches - primary g(r).

Curves with peaks - total and primary g(r) functions.

Monotone increasing curves - running coordination numbers for total and primary g(r) functions.

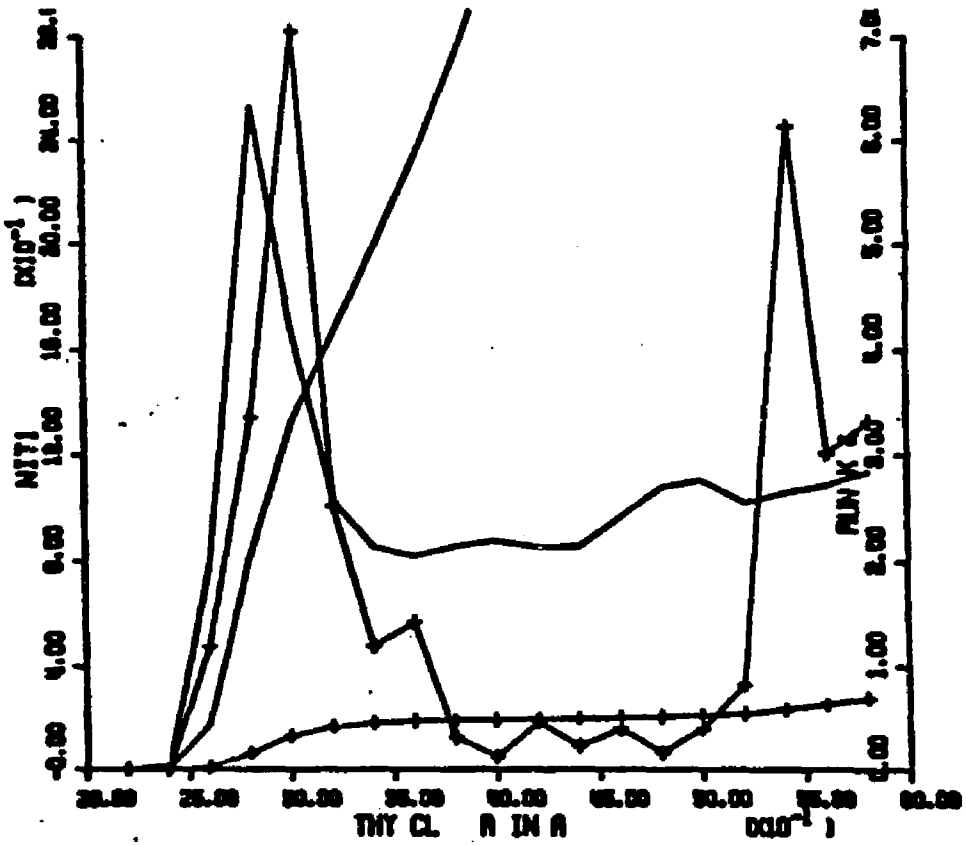


Figure IV.1 Total and primary g(r) functions and corresponding running coordination numbers for the N1 nitrogen atom of thymine - Clementi solute-water potential function simulation.

preceeding page. The curve marked with crosses displaying maxima and minima is the primary g(r) for this atom. As previously described, this function describes the fluctuation of densities of water molecules closest to a particular atom as a function of increasing distance. The radial cutoff for an atom is generally chosen as the first minimum in the primary g(r). Thus, for the thymine N1 atom, the cutoff is selected as 4.00 Anstroms.

Once the radial cutoff for an atom is defined, various quantities are calculated. The first shell coordination number of an atom is simply the number of water molecules falling inside the limits of the Voronoi polyhedron and radial cutoff of a particular atom. The distribution of coordination numbers over a Monte Carlo simulation is displayed in the bar graph of Figure IV.1.4. This graph (the quasicomponent distribution function of coordination number) reveals that in about 3% of the configurations the (NH)1 functional group has two waters in the first hydration shell, three waters in about 40% of the configurations, four waters in about 45% of the configurations, five waters in about 10% of the configurations and six waters in about 1% of the configurations. Another important quantity is the distribution of pair energies between the first shell water molecules of an atom and the solute molecule.

In the following Computational Results section, the tables and graphs for each simulation are collected and found immediately following the text section describing the particular simulation and its results. Labels for each set of graphs are described on the page preceeding each similar set of graphs, i.e., graphs of the quasicomponent distribution function of coordination number.

Computational Results

1. Thymine with Clementi Solute-Water Potentials

Computational Specifics. An aqueous solution of thymine and 215 waters was simulated using Clementi solute-water potentials and MCY-CI water-water potentials at 298 K. Face centered cubic boundary conditions with a unit cell of 14.87 Angstroms edge were used, determined by a partial molar volume of 85 cc/mole for thymine and 18.07 cc/mole for water. An initial 600,000 steps of Monte Carlo simulation were carried out and were discarded as an equilibration period. The ensemble averages and other analyses were formed over the succeeding 2000K configurations of the realization.

Potential Surface. The potential function contour surface for the interaction of thymine and one water based on Clementi solute-water potential functions is shown in Figure IV.1.1. The most favorable thymine-water interaction is found in the region of the H1 imino hydrogen at an energy of -12.25 kcal/mole. The H3 imino hydrogen is also a strong interaction site, with values of between -11.25 and -12.25 kcal/mole on its O2 side and -9.25 to -10.25 kcal/mole on its O4 side. The carbonyl group oxygens O2 and O4 have energy minima for water interactions in the range -6.25 to -7.25 and -5.25 to -6.25 kcal/mole, respectively. An interaction site near the methine H5 exhibits a binding energy of -5.25 to -6.25 kcal/mole. Interaction energies in the region of the methyl group are around 0 kcal/mole, consistent with the apolar character of this group. Hydrophobic

interactions with water are expected in this region.

Convergence and Thermodynamic Results. The control functions for the aqueous thymine simulation are given in Figure IV.1.2 and the calculated thermodynamic quantities are given in Table IV.1. The control functions show that the total binding energy reaches an apparently stable value of about -1881 kcal/mole after 1000K configurations, and continues in this region for the next 1000K. The heat capacity similarly settles to 15.71 cal/mole-K during this period. The calculated transfer energy from free space to water is found to be 21.4 kcal/mole.

Figures IV.1.4 thru IV.1.9 and IV.1.11 thru IV.1.17 contain distribution functions for the quasicomponents of coordination number and first shell pair energy for the imino, carbonyl, methyl and methine functional groups of thymine. Figures IV.1.10 and IV.1.18 contain the quasicomponents of coordination number and binding energy for the thymine molecule.

Table IV.2 contains many calculated quantities for this simulation. Entries are made for all atoms of the molecule. Composite values for each functional group and average values for each set of all functional groups of the same kind are are also tabulated. Finally, values for the entire thymine molecule are presented. The following quantities, defined in Chapter III except as

noted, are contained in the table. RFS is the radius of the first coordination sphere of an atom, VFS the the first shell volume, <K> is the number of water molecules in the first shell coordination sphere. <K/V> is the density of first shell waters with values scaled to multiples of the density of bulk water (1.00). <SLTBE> is the average solute binding energy of first shell waters or of all waters in the Voronoi polyhedron and <SLIPE> is the average first shell solute pair energy. The properties calculated for water-water interactions include <KW> the average water-water coordination number, <NNWWPE>, the average near neighbor water-water pair energy and <BEWWI> the average water-water binding energy.

Structural Results. The first shell coordination numbers for all atoms of the thymine molecule, derived from the column labeled <K> in Table IV.2, are displayed on the molecular diagram of Figure IV.1.3. The essential structural features of thymine hydration are as follows. The ring nitrogens N1 and N3 of the imino groups are sparsely hydrated due to their low solvent accessibility. However, the H1 and H3 imino hydrogens are assigned 3.18 and 3.47. waters respectively in proximity analysis. The oxygens, O2 and O4, are both significantly hydrated; O4, however, has a lower first shell coordination number (2.19 vs. 3.07) than O2. The lower first shell coordination number of O4 is paralleled to some extent by

its smaller first shell volume, 72.92 cubic Angstroms compared to 88.89 cubic Angstroms for O2, an 82% decrease in volume. The oxygen atom O2 also has a first shell density of .90, which is lower than the density of bulk water; in comparison, the O4 oxygen first shell density is 1.03. The density value for O2 suggests a hydrophobic mode of (or influence on) hydration.

The hydrogen atoms of the methyl groups are assigned first shell cutoff which are larger than those of the imino and methine hydrogens of the molecule (H5: 4.2 A; H5': 4.2 A; H5'': 4.4 A). Each hydrogen consequently has a large first shell population (H5:2.93, H5':3.31, H5'':3.19). The first shell water densities are lower than bulk water (.95, .90 and .95 respectively), also consistent with the picture of hydrophobic hydration.

A graph of the quasicomponent distribution function of coordination number for thymine is shown in Figure IV.1.10. Contributions range from 21 to 30 with the maximum contribution at 26. The average first shell coordination number of thymine in this simulation is 25.04 +/- 0.50.

Energetic Results. The first shell average solute-water pair energies assigned to each atom are displayed on the molecular diagram in Figure IV.1.11; these values are taken from the column labeled <SLIPE> in Table IV.2. The water molecules interacting most favorably with the

thymine molecule are assigned to the imino hydrogens. The H1 atom is assigned waters with average solute-water pair energies of -3.632 kcal/mole and contribute -11.560 to the total solute binding energy. Waters in the first shell of the H3 atom bind with an average pair energy of -3.473 kcal/mole and binding energy of -12.065 kcal/mole.

Weak solute-water interactions are found in the region of the methyl group. Waters in the first shells of the hydrogens of the methyl group contribute binding energies which are overall slightly positive: -.331 kcal/mole for H5, 1.447 kcal/mole for H5' and 1.008 kcal/mole for H5''. Water-water pair energies in the Voronoi polyhedra constructed around the methyl hydrogens are more negative (averaging -3.080 kcal/mole) than waters assigned to the imino or methine hydrogens and carbonyl oxygens. This pattern of energetics is chracteristic of hydrophobic hydration.

Although it seems reasonable that weak (hydrophobic) solute-water interactions would be found with waters assigned to an apolar hydrogen such as the H6 atom, relatively strong pair energies are encountered in the first shell of the H6 methine hydrogen atom. The average pair energy of -2.279 kcal/mole is about as favorable as the first shell pair energies of the carbonyl oxygens O2 (-2.555 kcal/mole) and O4 (-2.164 kcal/mole). The water-water pair energies of the water molecules in the primary

hydration region of this atom are the most positive (-2.948 kcal/mole) in the molecule.

The quasi-component distribution function for the solute-water binding energies of thymine are displayed in Figure IV.1.18. A range of values from -20.00 kcal/mole to about +5.00 kcal/mole are found with the greatest contributions in the -10.00 to -5.00 kcal/mole portion. The average solute-water binding energy is -7.4 +/- 0.9 kcal/mole.

Figure IV.1.1 - isoenergy contour surface for thymine and one water (Clementi solute-water potential functions).

X axis - X axis of plane defined by molecular ring atoms (Angstroms).

Y axis - Y axis of plane defined by molecular ring atoms (Angstroms).

Contours - isoenergy contours with one kcal/mole increments with alphabetical labels referring to contour energy values in list at right.

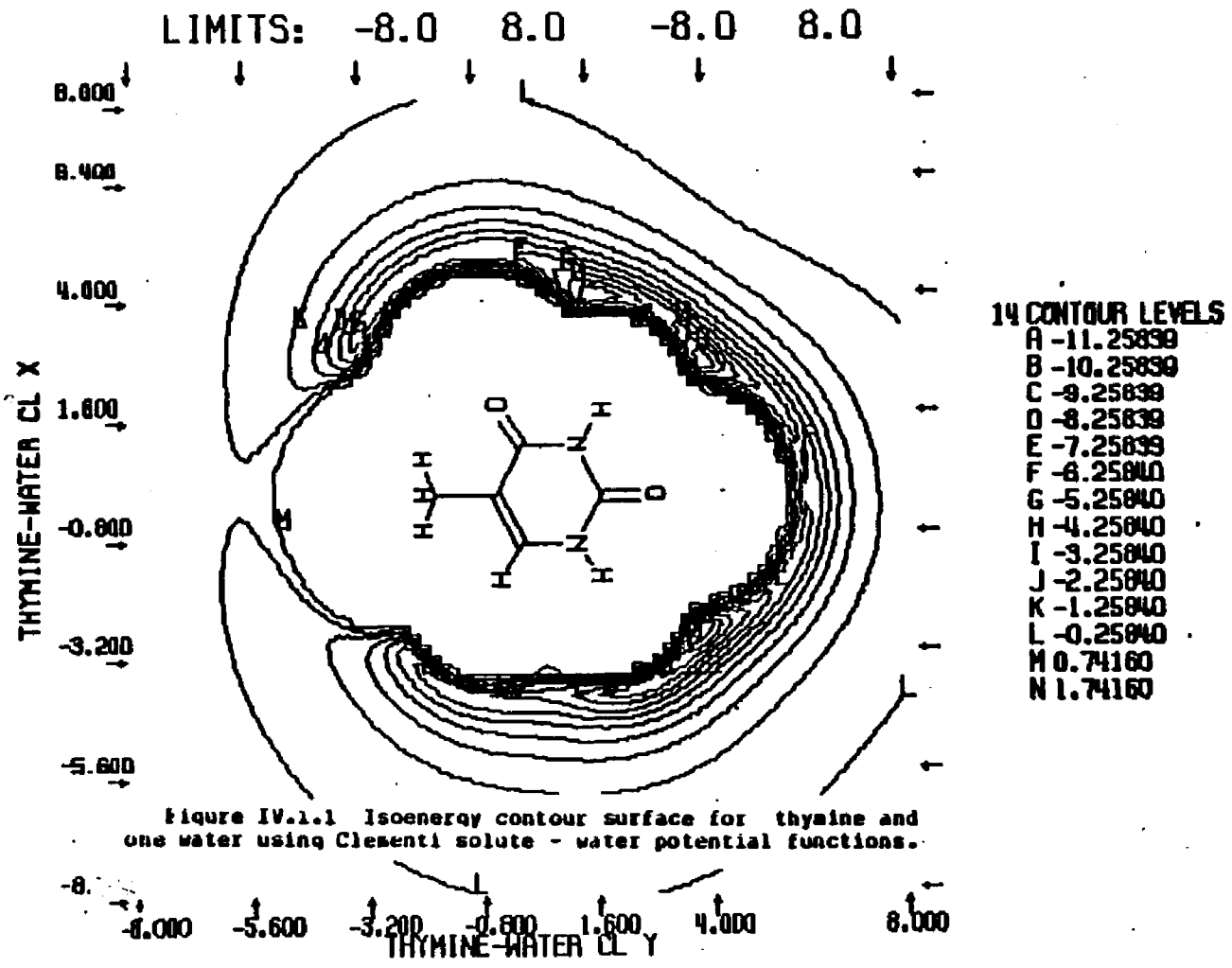


Figure IV.1.1 Isoenergy contour surface for thymine and one water using Clementi solute - water potential functions.

Figure IV.1.2 - control functions for Monte Carlo simulation of thymine and 215 waters (Clementi solute-water potential functions).

X axis - number of configurations.

Left Y axis - mean total energy (kcal/mole).

Right Y axis. - constant volume heat capacity (cal/mole-degree).

Upper curve - constant volume heat capacity.

Bottom curve without crosshatches - average total energy for entire simulation.

Bottom curve with crosshatches - average total energy for preceeding 50K configurations.

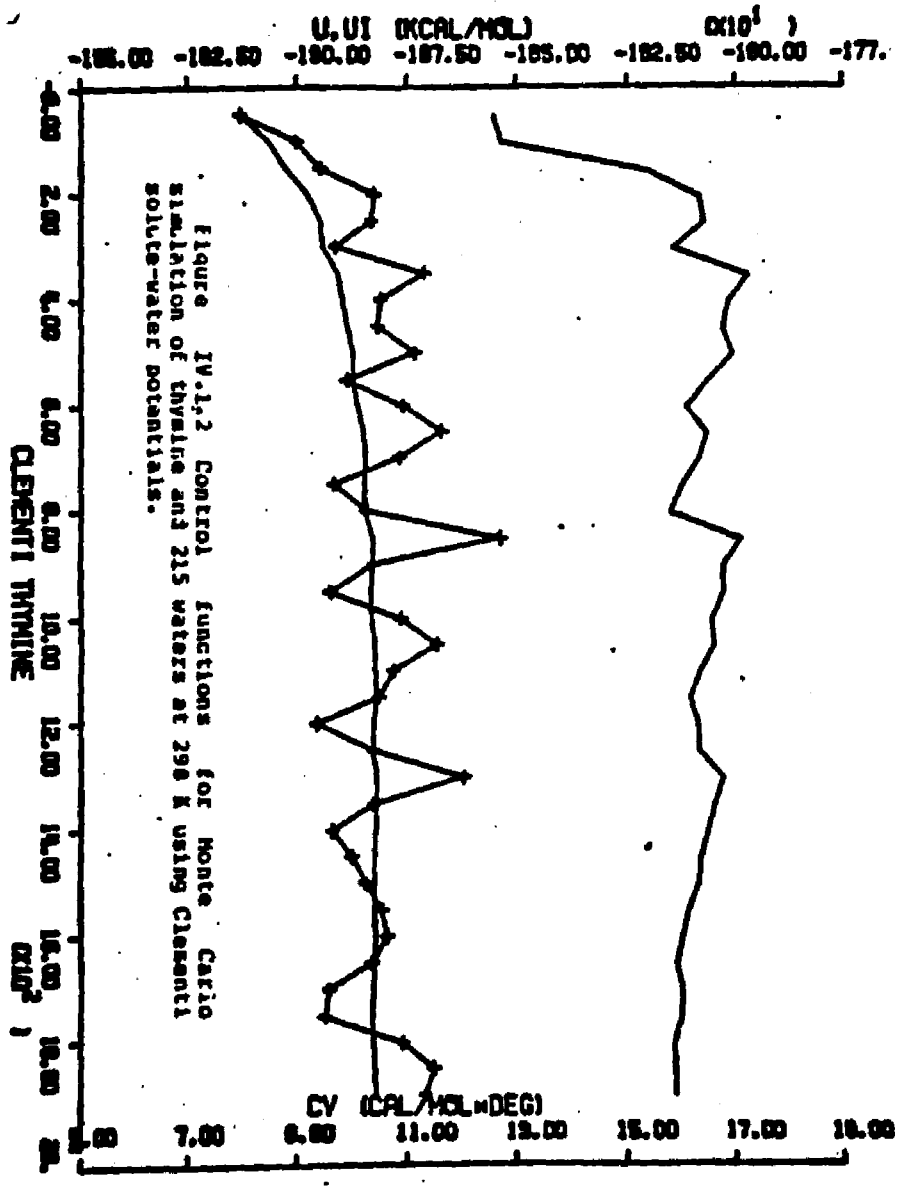


Figure IV.1.2 Control functions for Monte Carlo simulation of thymine and 215 waters at 298 K using Clementi solute-water potentials.

THYMINE IN WATER AT 298K - FORCE BIAS AND CLEMENTI POTENTIAL FUNCTIONS

LAST CONFIGURATION: 2000001

INDEX	TYPE			NFS	VFS	FIRST SHELL SOLUTE PROPERTIES <E>	<E/V>	<SLTENE>	<SLTPE>	TOTAL SLT PROPS <E>	<SLTENE>	WATER PROPERTIES NFS=3.30 RCB= 7.75 A <ND>	<RNMPE>	<RNMT>
METHYL GROUP														
1	9	6	C	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	13	3	H	4.2	92.05	2.93	0.95	0.331	0.113	20.83	5.416	4.16	-3.056	-17.447
3	14	3	H	4.2	110.38	3.31	0.98	1.447	0.437	25.98	7.588	4.18	-3.068	-17.497
4	15	3	H	4.4	100.81	3.19	0.95	1.008	0.316	18.32	4.963	4.18	-3.125	-17.629
TOTALS FOR FUNCTIONAL GROUP -CH3 :				302.45		9.43	0.93	2.785	0.295	65.13	17.959	4.17	-3.000	-17.518
STATISTICAL UNCERTAINTY (+/- 2*SD):						0.31	0.03	0.142	0.017	0.04	6.041	0.02	0.050	0.116
METHYLENE (DISUBST.)														
5	7	20	C	4.6	14.42	0.07	0.14	-0.000	-0.003	0.13	0.010	4.06	-3.233	-17.430
STATISTICAL UNCERTAINTY (+/- 2*SD):						0.03	0.05	0.000	0.002	0.00	0.072	0.37	1.163	2.575
METHYLENE (MONOSUB.)														
6	6	25	C	4.6	17.07	0.14	0.23	-0.054	-0.393	0.36	-0.010	4.22	-3.007	-17.779
7	12	16	H	3.4	60.00	2.07	1.02	-4.709	-2.279	19.03	-1.007	4.26	-2.940	-17.327
TOTALS FOR FUNCTIONAL GROUP -CH- :				78.67		2.21	0.04	-4.764	-2.160	20.19	-1.097	4.26	-2.951	-17.336
STATISTICAL UNCERTAINTY (+/- 2*SD):						0.15	0.06	0.503	0.259	0.02	0.663	0.03	0.005	0.206
AMIDE GROUP														
8	3	15	N	4.0	16.32	0.45	0.03	-0.615	-1.354	1.21	-0.513	4.10	-3.114	-10.227
9	11	16	H	3.4	71.50	3.10	1.33	-11.560	-3.632	31.02	-0.664	4.23	-3.023	-17.746
TOTALS FOR FUNCTIONAL GROUP >NH :				07.02		3.64	1.24	-12.175	-3.340	33.04	-0.177	4.23	-3.026	-17.764
10	4	15	N	3.0	12.21	0.20	0.50	-0.342	-1.002	1.50	-1.121	4.05	-3.263	-17.707
11	10	16	H	3.6	02.01	3.47	1.25	-12.065	-3.473	35.33	-0.160	4.37	-2.965	-17.719
TOTALS FOR FUNCTIONAL GROUP >NH :				95.02		3.60	1.16	-12.407	-3.374	36.03	-9.202	4.36	-2.977	-17.722
AVERAGES OVER FUNCTIONAL GRP >NH :				91.42		3.66	1.20	-12.291	-3.361	34.93	-9.229	4.29	-3.002	-17.743
STATISTICAL UNCERTAINTY (+/- 2*SD):						0.14	0.04	0.713	0.222	0.02	2.997	0.02	0.047	

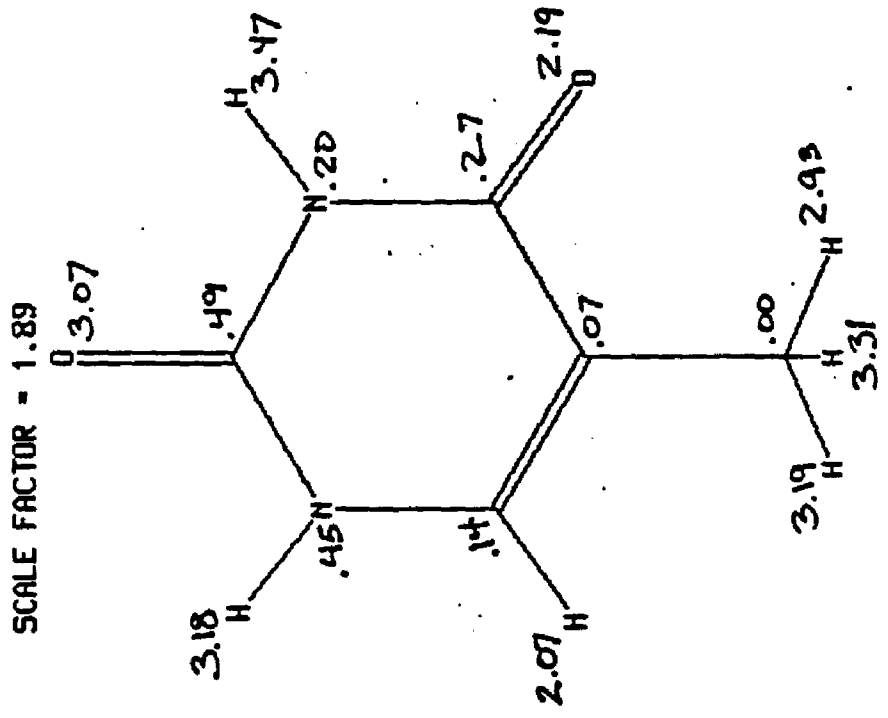
Table IV.2 Calculated structural and energetic quantities from Monte Carlo simulation of thymine and 215 waters at 298 K - Clementi solute-water potential functions.

THYMINE IN WATER AT 298K - FORCE BIAS AND CLEMENTI POTENTIAL FUNCTIONS

CARBONYL GROUP													
12	1	27	O	3.6	88.89	3.87	1.83	-7.852	-2.555	34.86	-5.289	4.23	-3.818	-17.279
13	5	26	C	4.2	19.64	0.49	0.75	-0.611	-1.246	1.41	-0.495	4.16	-3.145	-17.521
TOTALS FOR FUNCTIONAL GROUP C=O :					108.52	3.56	0.98	-8.463	-2.375	35.47	-5.784	4.23	-3.816	-17.289
14	2	27	O	3.6	72.92	2.19	0.98	-4.749	-2.164	23.67	-0.857	4.26	-2.968	-17.421
15	8	26	C	4.8	17.53	0.27	0.46	-0.020	-0.073	0.54	0.015	3.93	-3.283	-17.447
TOTALS FOR FUNCTIONAL GROUP C=O :					90.45	2.47	0.82	-4.769	-1.934	24.21	-0.843	4.25	-2.975	-17.422
AVERAGES OVER FUNCTIONAL GRP C=O :					99.49	3.01	0.91	-6.616	-2.155	29.84	-2.873	4.24	-2.995	-17.355
STATISTICAL UNCERTAINTY (+/- 2*SD):						0.12	0.04	0.423	0.156	0.82	1.010	0.02	0.050	0.120
MOLECULAR SUM/AVERAGE:					777.4	25.04	0.96	-39.792	-1.589	215.00	-7.333	4.24	-3.019	-17.524
STATISTICAL UNCERTAINTY (+/- 2*SD):						0.50	0.02	1.240	0.057	0.07	1.350	0.01	0.037	0.064

Table IV.2 Calculated structural and energetic quantities from Monte Carlo simulation of thymine and 215 waters at 298 K - Clementi solute-water potential functions.

Figure IV.1.3 First shell coordination numbers for the atoms of thymine - Clementi solute-water potential function simulation.



Figures IV.1.4 through IV.1.10 - calculated quasicomponent distribution function of first shell coordination number for thymine and functional groups (Clementi potentials).

X axis - coordination number.

Y axis - quasicomponent of coordination number.

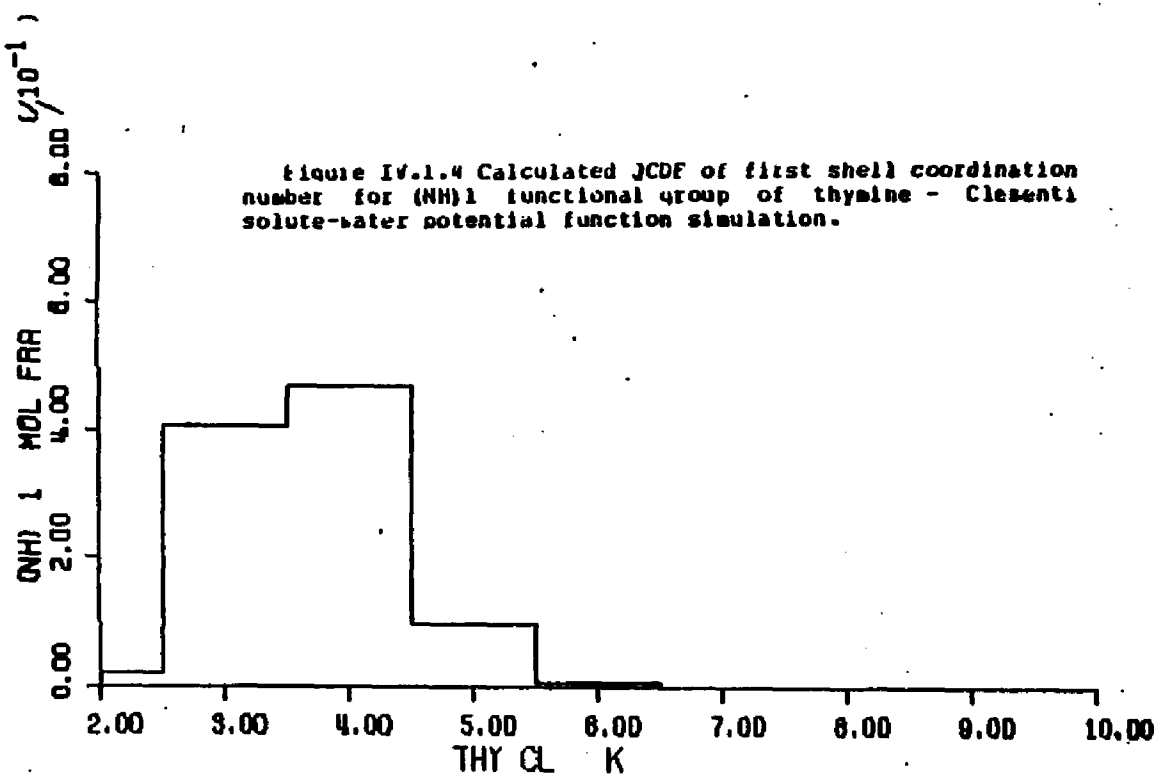


Figure IV.1.4 Calculated JCDF of first shell coordination number for (NH)1 functional group of thymine - Clementi solute-water potential function simulation.

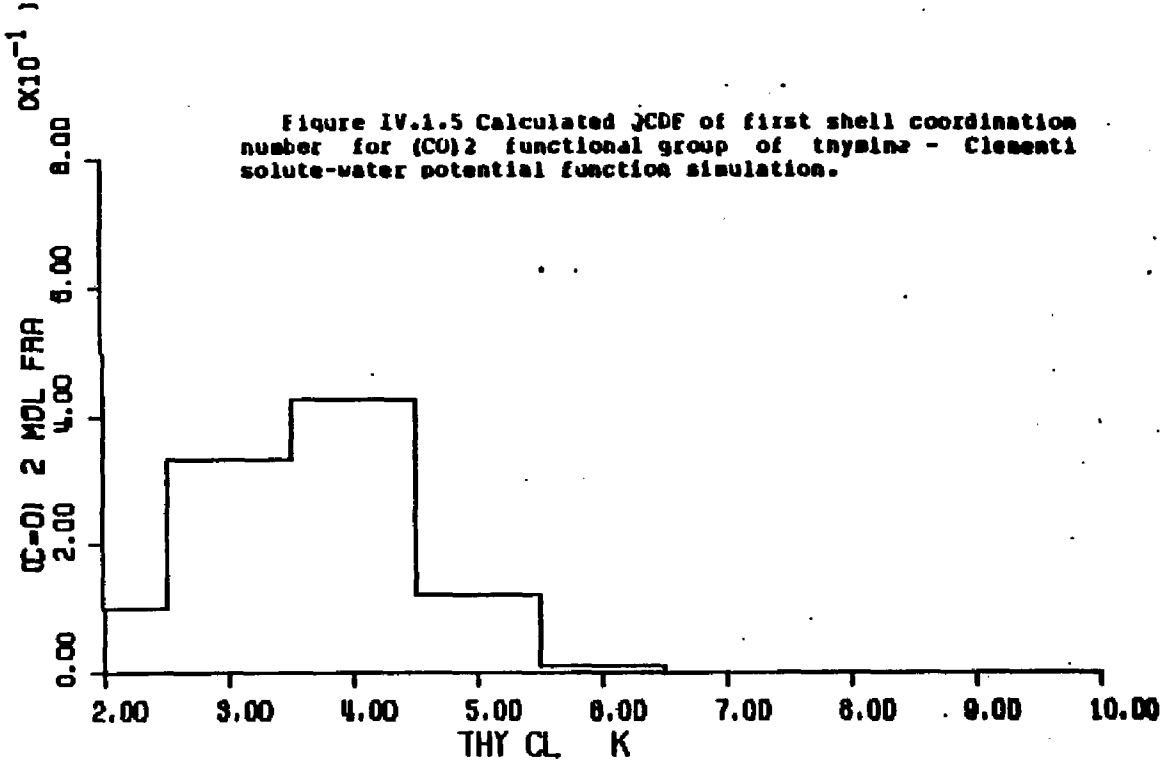


Figure IV.1.5 Calculated JCDF of first shell coordination number for (CO)2 functional group of thymine - Clementi solute-water potential function simulation.

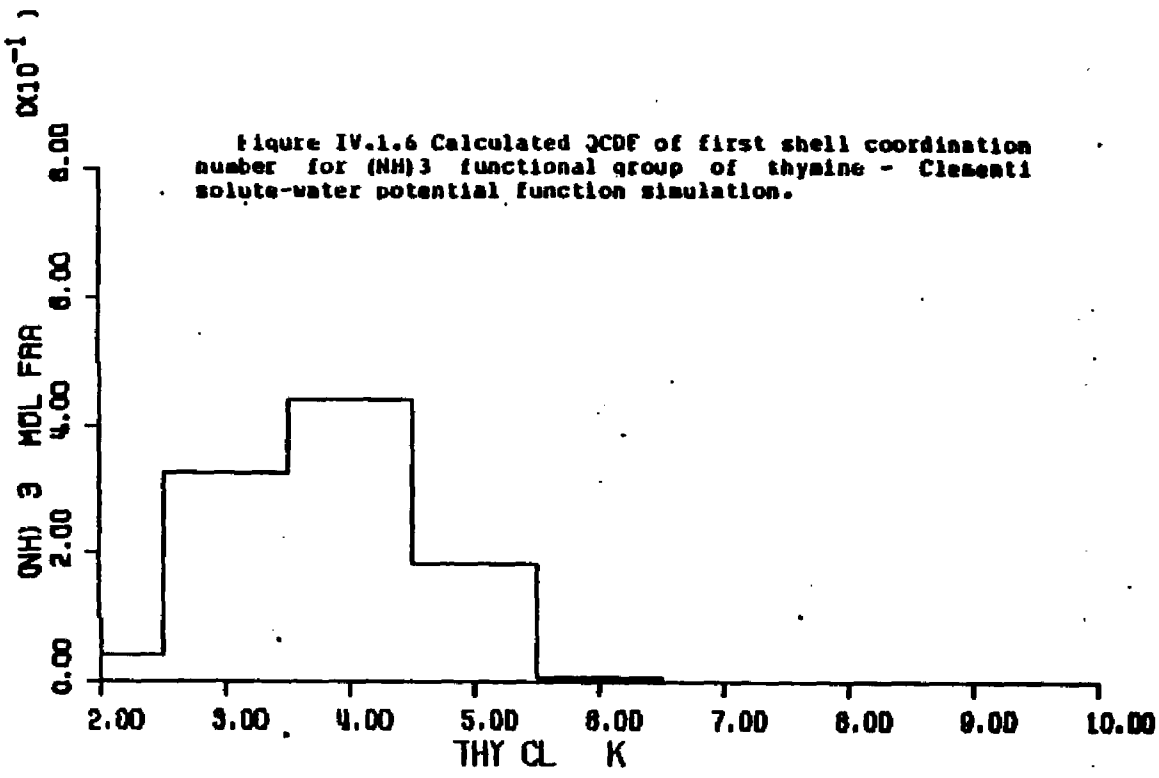


Figure IV.1.6 Calculated RDF of first shell coordination number for $(NH)_3$ functional group of thymine - Clementi solute-water potential function simulation.

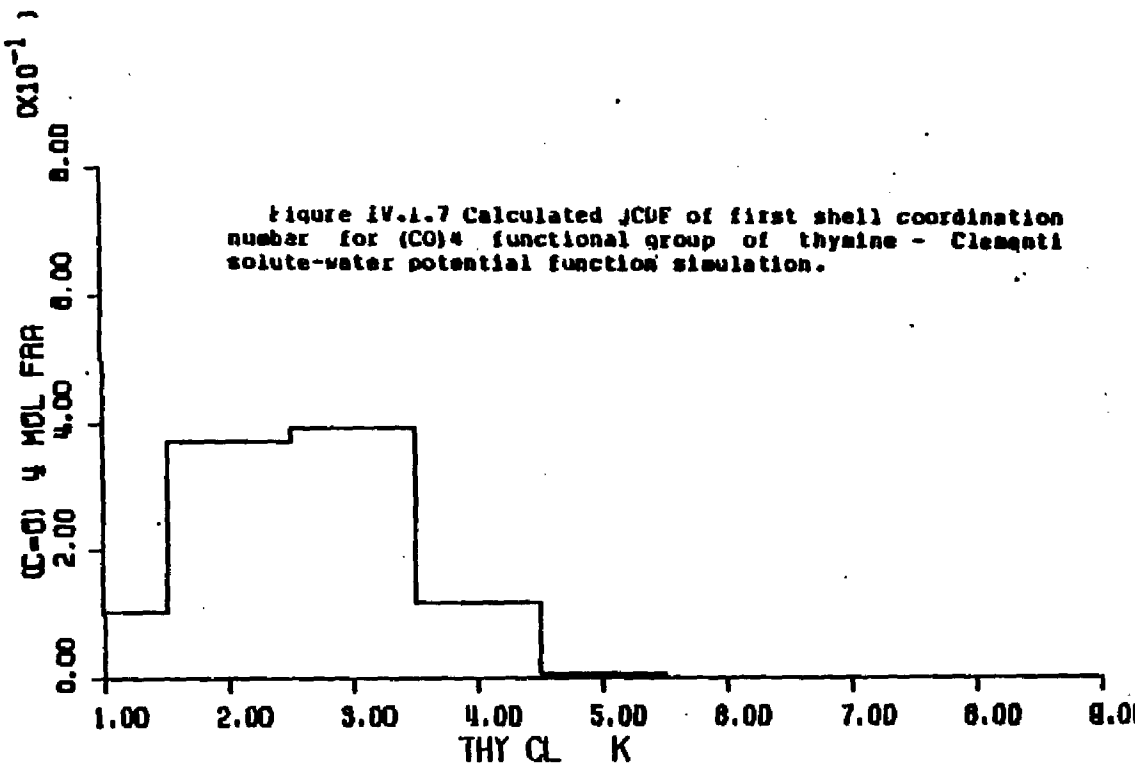


Figure IV.1.7 Calculated g_{CDF} of first shell coordination number for $(CO)_4$ functional group of thymine - Clementi solute-water potential function simulation.

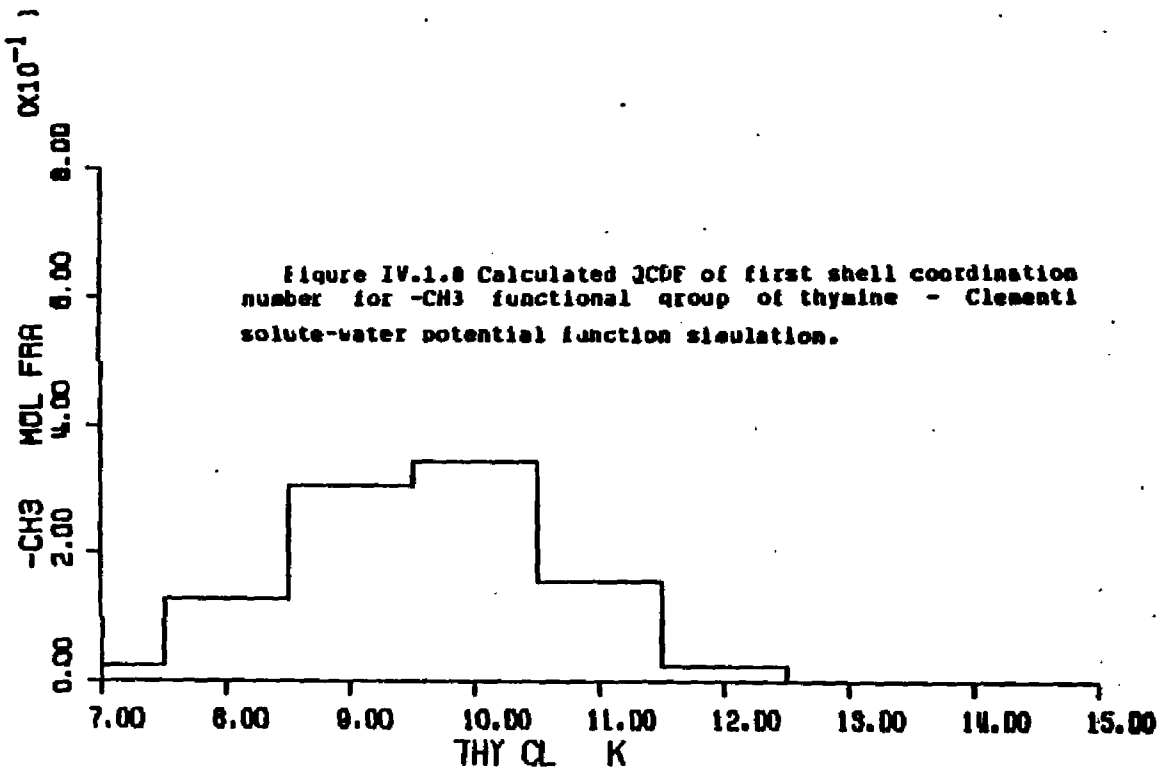


Figure IV.1.8 Calculated JCDF of first shell coordination number for -CH3 functional group of thymine - Clementi solute-water potential function simulation.

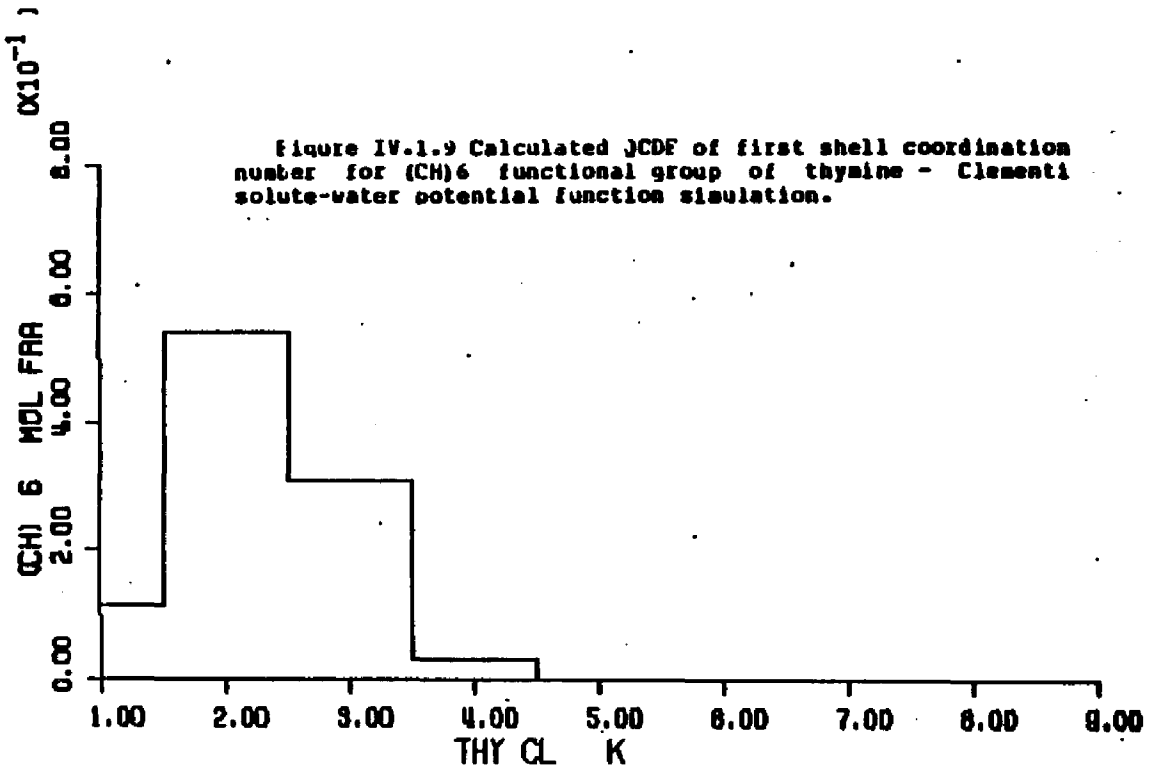


Figure IV.1.9 Calculated JCDF of first shell coordination number for (CH)6 functional group of thymine - Clementi solute-water potential function simulation.

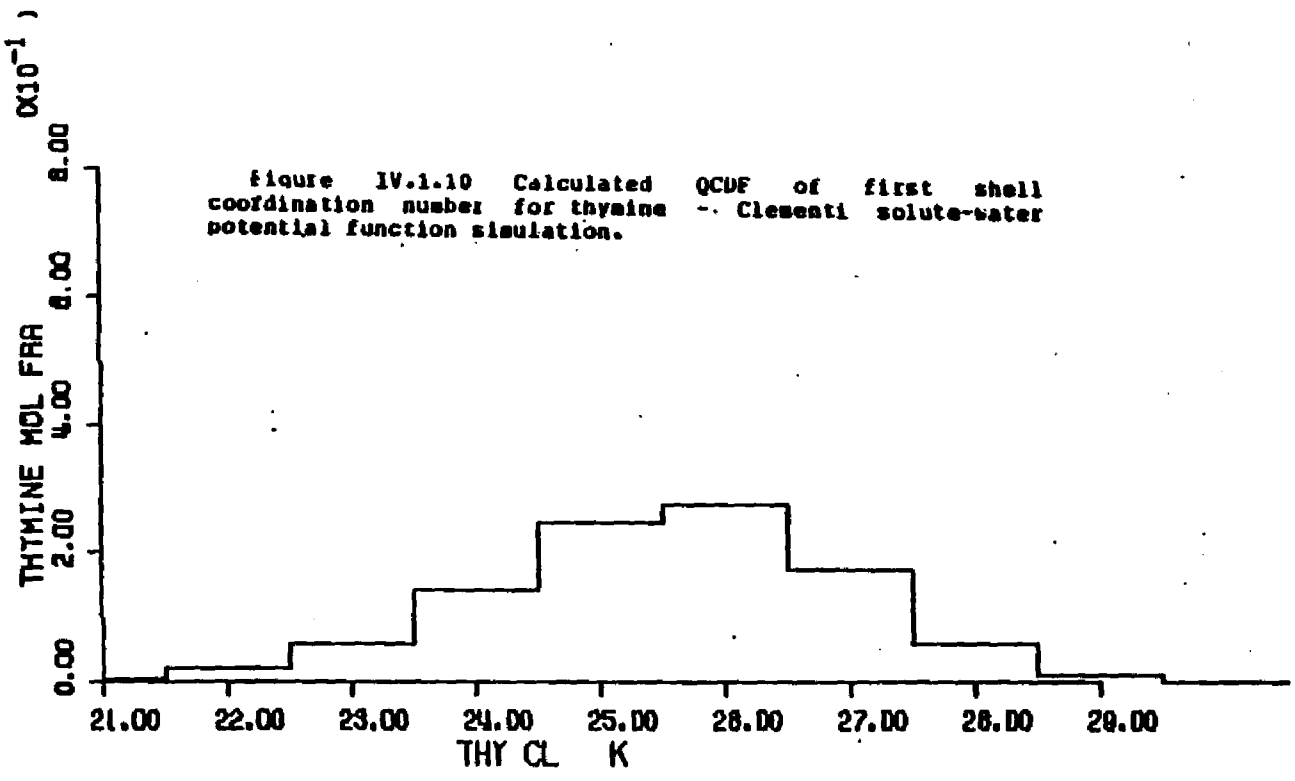


Figure IV.1.10 Calculated QCUF of first shell coordination number for thymine -- Clementi solute-water potential function simulation.

Figure IV.1.11 Average first shell solute-water pair energies of waters assigned to the atoms of thymine - Clementi solute-water potential function simulation.

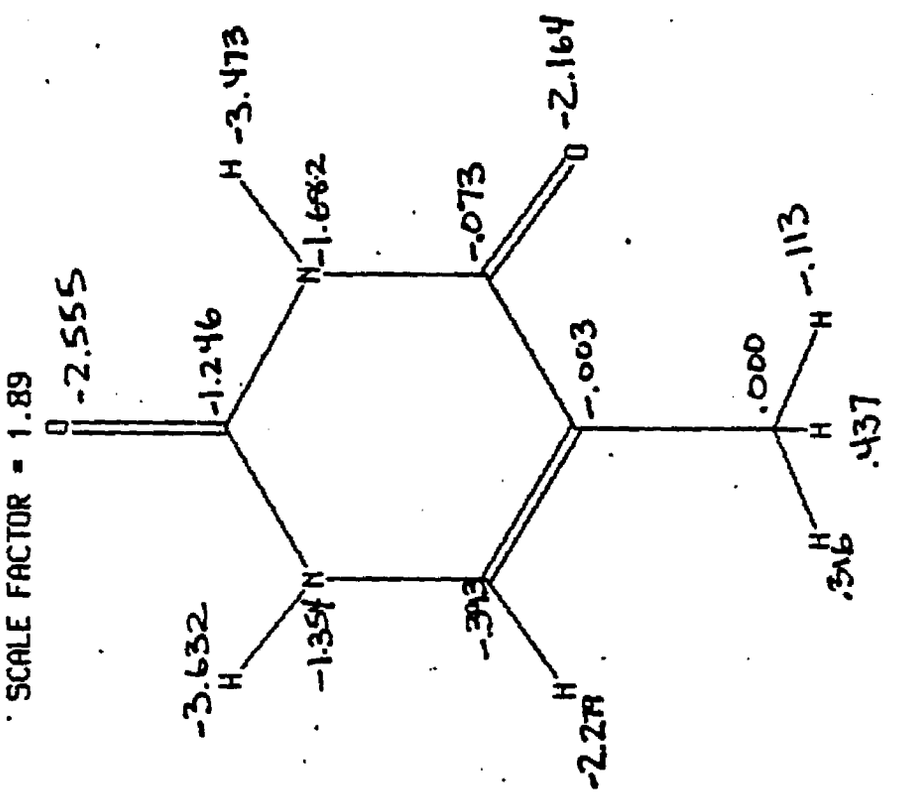


Figure IV.1.12 through IV.1.17 - calculated quasicomponent distribution function of average first shell pair energy for functional groups of thymine (Clementi potentials).

X axis - pair energy (kcal/mole).

Left Y axis - quasicomponent of pair energy.

Right Y axis - running coordination number.

Figure IV.1.12 Calculated QCDF of solute-water pair energies of waters of the (NH)1 functional group of tnymine - Clementi solute-water potential function simulation.

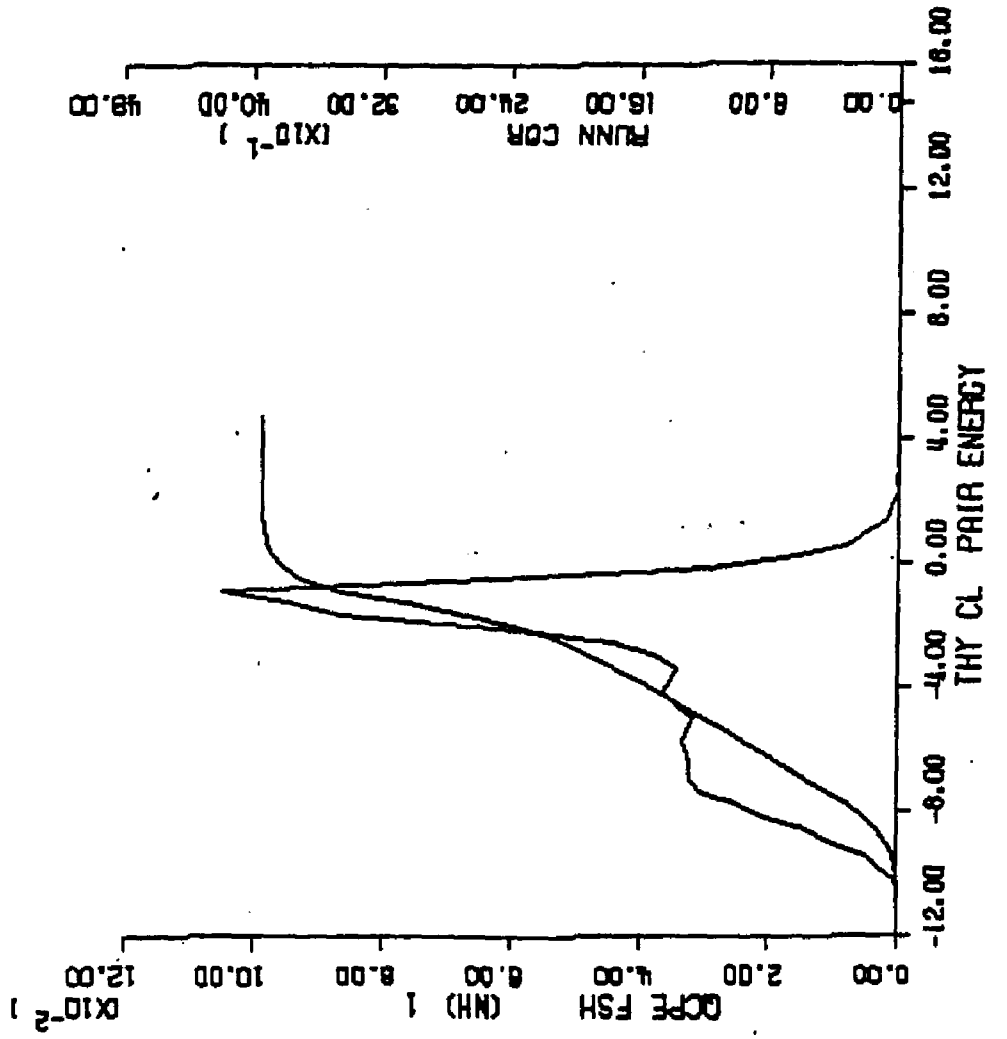


Figure IV.1.13 Calculated QCDF of solute-water pair energies of waters of the (CO)2 functional group of thymine - Clementi solute-water potential function simulation.

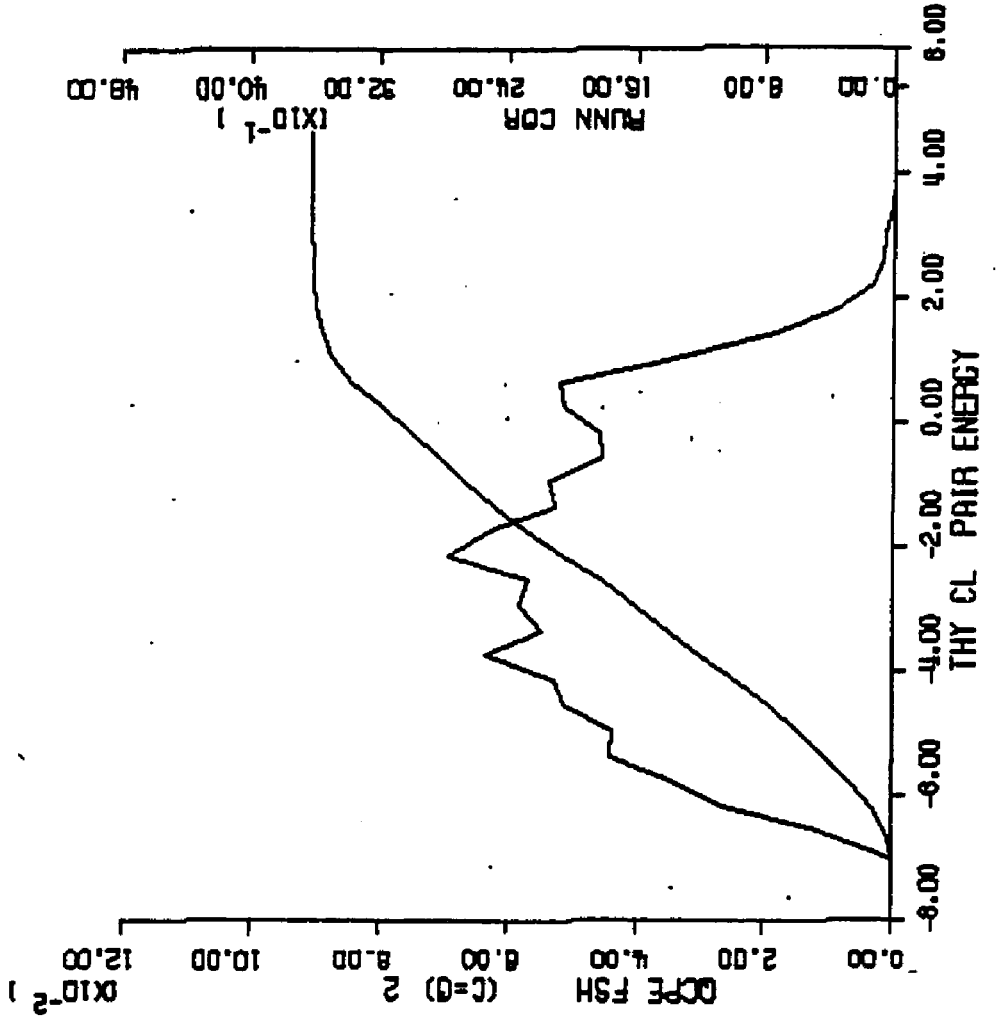


Figure IV-1.19 Calculated vCDF of solute-water pair energies of waters of the (NH)3 functional group of thymine - Clementi solute-water potential function simulation.

Figure IV.1.15 Calculated CDF of solute-water pair energies of waters of the (CO)4 functional group of thymine - Clementi solute-water potential function simulation.

Figure IV.1.1b Calculated QCDF of solute-water pair energies of waters of the -CH3 functional group of thymine - Clementi solute-water potential function simulation.

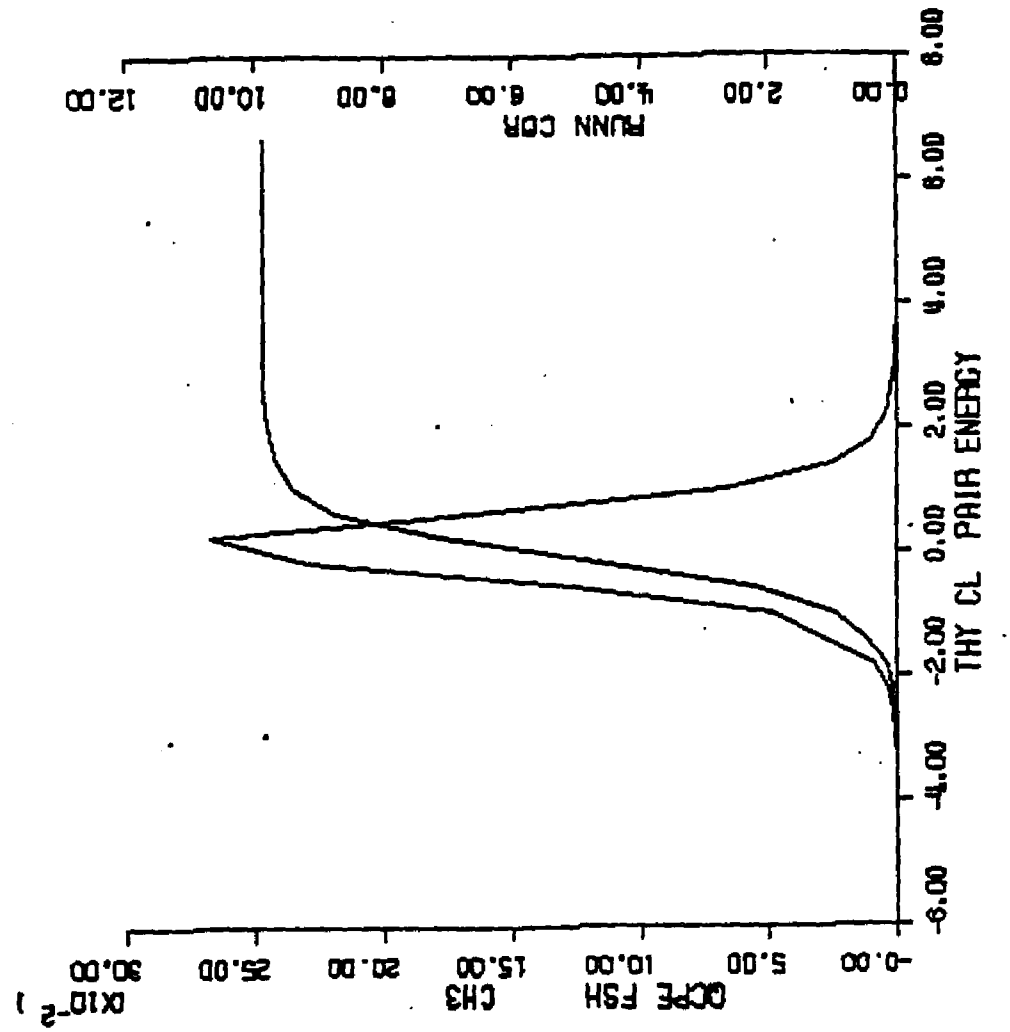


Figure IV.1.17 Calculated QCDF of solute-water pair energies of waters of the (CH)6 functional group of thymine - Clementi solute-water potential function simulation.

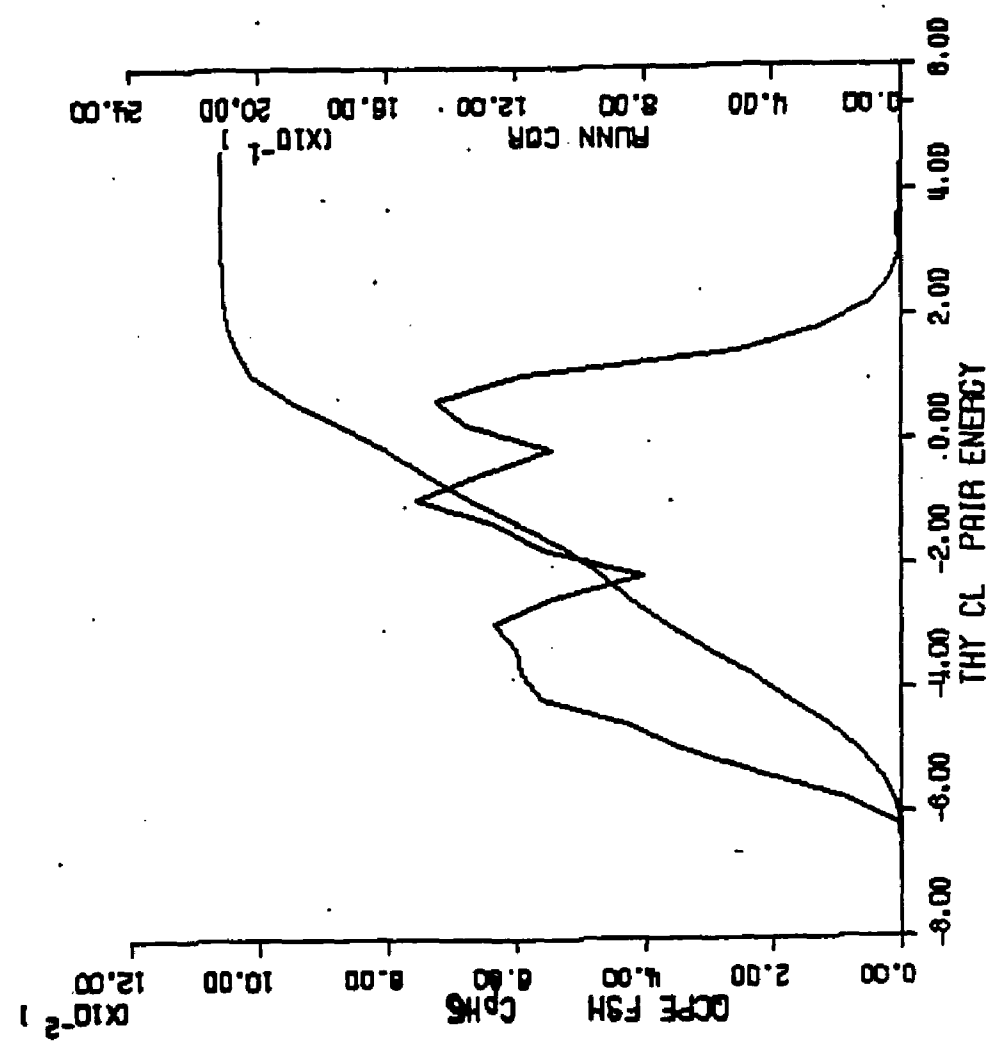
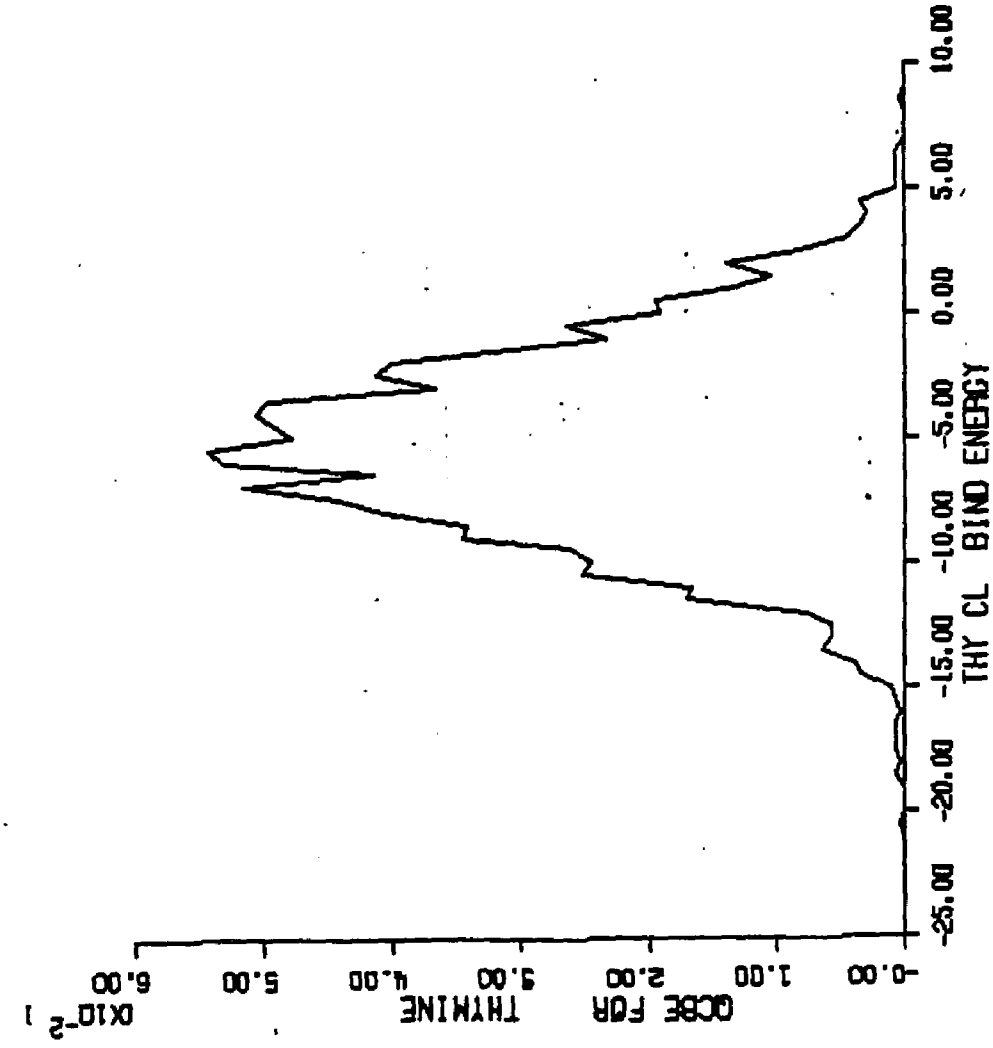


Figure IV.1.18 - calculated quasicomponent distribution function of binding energy for thymine (Clementi potentials).

X axis - binding energy (kcal/mole).

Y axis - quasicomponent of binding energy.

Figure IV.1.18 Calculated QCDF of total binding energy for thymine - Clementi solute-water potential function simulation.



2. Thymine With Kollman Solute-Water Potentials

Computational Specifics. An aqueous solution of thymine and 215 waters was simulated using the Kollman solute-water potentials and the TIP3P water-water potentials at 298 K. Simulation conditions were identical to the thymine simulation described previously using Clementi solute-water potentials. Face centered cubic boundary conditions were used with a unit cell edge of 14.87 Angstroms. The simulation consisted of an equilibration period of 600,000 moves, which was discarded, and a production period of 2,000,000 moves. All analyses and ensemble averages were formed over the last 2000K configurations.

Potential Surface. The potential function surface for the interaction of thymine and one water, using Kollman solute-water potential functions for the calculation, is shown in Figure IV.2.1. The lowest energy solute-water interaction occurs in the region of the H1 imino hydrogen atom at an energy of -6.76 kcal/mole. Nearly as favorable an interaction is found at the H3 hydrogen site with energies between -4.76 and -5.76 kcal/mole. The carbonyl O2 and O4 oxygens bind water somewhat more weakly with the range of interactions for waters in the region of both atoms falling between -2.76 and -3.76 kcal/mole. The apolar methine H6 atom interacts even more weakly with water, with interactions ranging from -1.76 to -2.76 kcal/mole. Finally, thymine-water interactions in the vicinity of the methyl group are

almost negligible with energies around 0.0 kcal/mole.

Convergence and Thermodynamic Results. The control functions for the Kollman potential aqueous thymine simulation are shown in Figure IV.2.2; the calculated thermodynamic quantities are given in Table IV.1. The control function plots reveal that the total binding energy reaches an apparently stable value around 800K configurations and after the next 1200K configurations this value is found to be -2136.3 +/- 5.5 kcal/mole. The heat capacity attains a final value of 18.1 cal/mole-K after the 2000K configurations. The vacuum to water transfer energy is calculated to be -16.8 kcal/mole. Figures IV.2.4 thru IV.2.9 and IV.2.12 thru IV.2.17 contain distribution functions for the quasicomponents of coordination number and first shell pair energy for the imino, carbonyl, methyl and methine functional groups of thymine. Figures IV.2.10 and IV.2.18 contain the quasicomponents of coordination number and binding energy for the thymine molecule.

Structural Results. The atomic first shell coordination numbers, derived from the column labeled <K> in Table IV.3, are displayed on the molecular diagram of Figure IV.2.3. The H1 and H3 imino hydrogens are fully hydrated with 2.67 and 2.91 waters, respectively, in each first shell. The first shell water densities of the polar atoms is about that of bulk water (H1: .97; H3:

1.03). The carbonyl oxygens O2 and O4 are also significantly hydrated (3.05 and 3.18 first shell waters respectively) with first shell water densities about equal to bulk water (.99 and .96 respectively). Thus, none of the polar ring substituents show the increased first shell concentration expected of polar atoms.

The methyl group is represented as a united atom in the Kollman potential function set; in other words the hydrogens are not explicitly represented but are effectively included in the potential which is centered on the methyl carbon atom. Thus, the first coordination shell volume of the Kollman methyl group (197.95 cubic Angstroms) is smaller than that of the Clementi methyl group (300.00 cubic Angstroms) The coordination number of the Kollman methyl group is 6.20 (vs. 9.43 for the Clementi methyl group). The extended first shell radius (4.6 Angstroms) and low first shell water density (.94) is characteristic of the hydration of apolar atoms.

Another hydrophobic hydration pattern is seen with the C6H6 methine united atom. The extensive first shell of radius 4.8 Anstroms contains 3.00 water molecules. The first shell water density of .92 is the lowest in the molecule (with the exception of substituted ring atoms with much reduced solvent accessibilty).

The QCDF of coordination number for thymine is displayed in Figure IV.2.10. The contributions fall in

the range 19 to 29; the most common coordination number is 25. The average first shell coordination number is 22.97 +/- 0.6.

Energetic Results. First shell solute pair energies for the atoms of thymine are shown with the molecular structure in Figure IV.2.11. The waters assigned to the imino hydrogens interact with thymine at lower energies than any other waters. The H1 hydrogen has first shell waters with average pair energies of -1.268 kcal/mole while the waters of hydrogen H3 each interact with an average -1.073 kcal/mole. The total binding energy of the H1 and H3 waters is -6.504 kcal/mole, about 40% of the total thymine first shell binding energy of -16.943 kcal/mole. The pair energies for the carbonyl oxygen waters are somewhat less strong (O2: -0.766 kcal/mole and O4: -0.766 kcal/mole).

The waters of the methyl group have the least favorable average pair energies of the molecule: -.394 kcal/mole. Overall, the methyl group waters contribute a solute binding energy of -2.443 kcal/mole which is comparable to the oxygen 2 (-2.894 kcal/mole) and oxygen 4 (-2.600 kcal/mole) contributions. This is unusual considering the apolar nature of the methyl group and the polar nature of the oxygen atom. The average water-water pair energies in this region are -3.189 kcal/mole, only slightly more enhanced than those of the imino hydrogens

(-3.177 and -3.153 kcal/mole), carbonyl O4 (-3.137 kcal/mole) and less favorable than those of carbonyl O2 (-3.211 kcal/mole). The water-water pair energies in the apolar methyl region would generally be expected to be less favorable than those in the polar oxygen region. The waters of the methine group interact with an average pair energy of -.591. The water-water pair energies of -3.268 kcal/mole are enhanced considerably relative to other waters in the simulation.

The QCDF of solute-water binding energies for all 215 waters is displayed in Figure IV.2.18. Energy values ranging from -33 to -14 kcal/mole are encountered; the most significant contributions range from -27 to -21 kcal/mole. The average solute binding energy is -22.97 +/- 0.6 kcal/mole.

Figure IV.2.1 - isoenergy contour surface for thymine and one water (Kollman solute-water potential functions).

X axis - X axis of plane defined by molecular ring atoms (Angstroms).

Y axis - Y axis of plane defined by molecular ring atoms (Angstroms).

Contours - isoenergy contours with one kcal/mole increments with alphabetical labels referring to contour energy values in list at right.

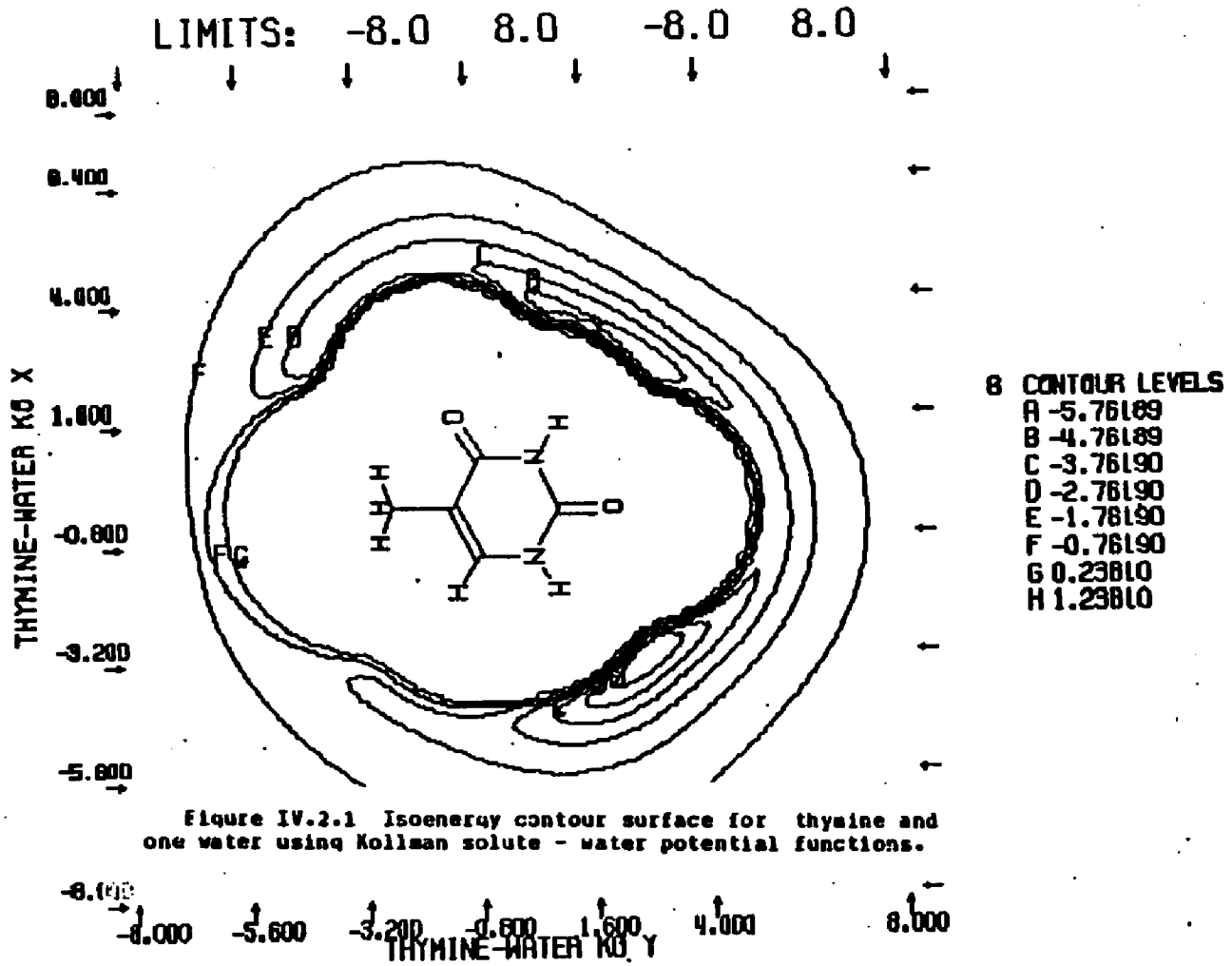


Figure IV.2.1 Isoenergy contour surface for thymine and one water using Kollman solute - water potential functions.

Figure IV.2.2 - control functions for Monte Carlo simulation of thymine and 215 waters (Kollman solute-water potential functions).

X axis - number of configurations.

Left Y axis - mean total energy (kcal/mole).

Right Y axis. - constant volume heat capacity (cal/mole-degree).

Upper curve - constant volume heat capacity.

Bottom curve without crosshatches - average total energy for entire simulation.

bottom curve with crosshatches - average total energy for preceeding 50K configurations.

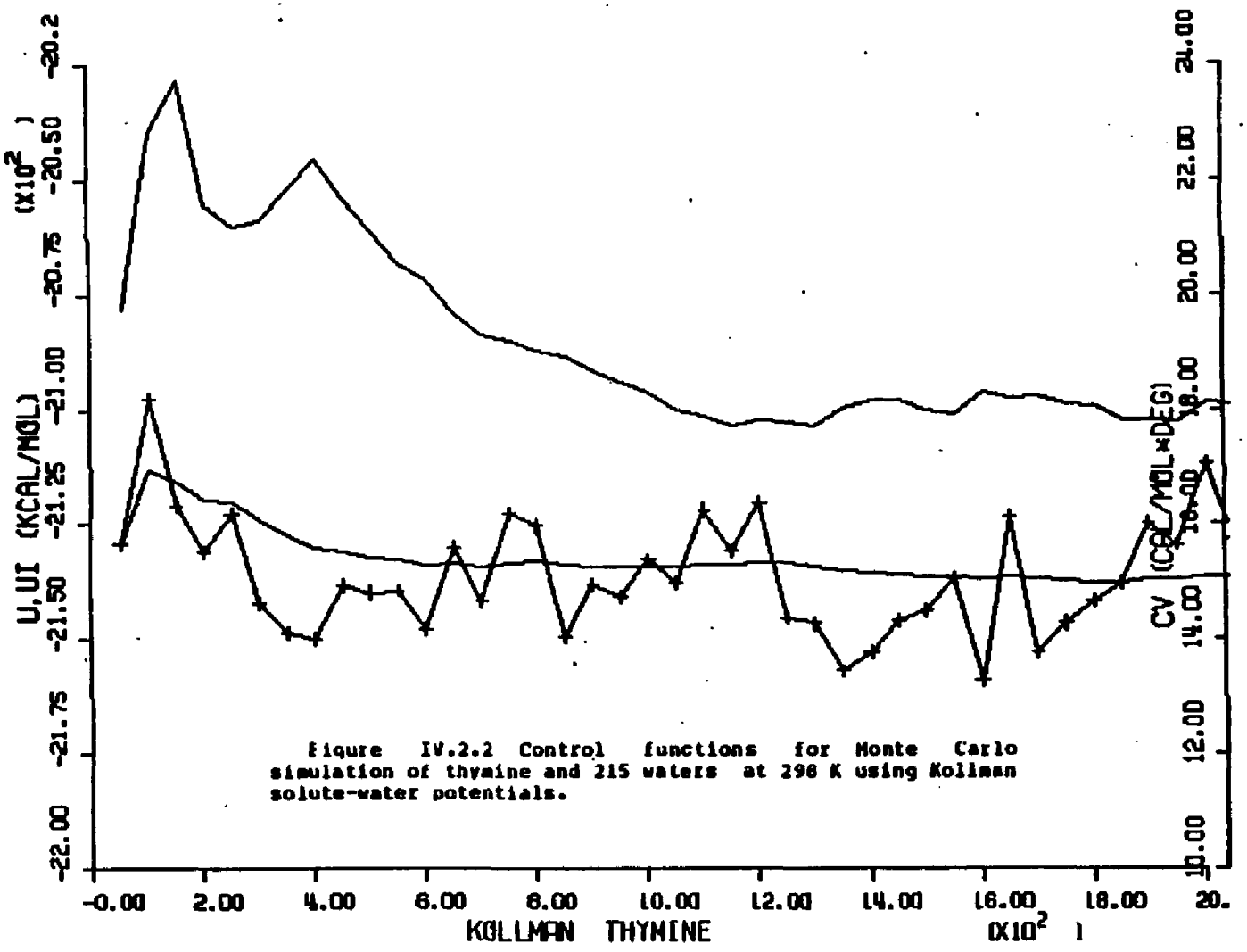


Figure IV.2.2 Control functions for Monte Carlo simulation of thymine and 215 waters at 298 K using Kollman solute-water potentials.

THYMINE IN WATER AT 298K - FORCE BIAS AND KOLLMAN POTENTIAL FUNCTIONS

LAST CONFIGURATION: 2000001						FIRST SHELL SOLUTE PROPERTIES					TOTAL SLT PROPS		WATER PROPERTIES RFSW=3.30 RCB= 7.75 A		
AT NO	INDEX	TYPE			RFS	VPS	<K>	<K/V>	<ELTBE>	<ELTPE>	<K>	<ELTBE>	<NW>	<EBWPE>	<EBWT>
RING CARBON															
1	9	15	C	C5	4.2	22.25	0.47	0.64	-0.253	-0.535	1.769	-0.380	4.34	-3.242	-19.827
STATISTICAL UNCERTAINTY (+/- 2*SD)							0.05	0.11	0.000	0.004	0.002	0.072	0.37	1.163	2.575
METHYLENE GROUP															
2	11	14	C	C6H6	4.8	97.48	3.00	0.92	-1.774	-0.591	15.835	-2.237	4.32	-3.268	-19.643
STATISTICAL UNCERTAINTY (+/- 2*SD)							0.28	0.09	0.418	0.114	0.012	0.288	0.06	0.091	0.286
METHYL GROUP															
3	10	4	C	CH3 5*	4.6	197.95	6.20	0.94	-2.443	-0.394	45.771	-3.288	4.42	-3.189	-19.748
STATISTICAL UNCERTAINTY (+/- 2*SD)							0.34	0.05	0.338	0.045	0.020	0.100	0.03	0.052	0.122
IMINO GROUPS															
4	1	30	N	N1	4.0	16.31	0.33	0.61	-0.260	-0.709	1.331	-0.330	4.40	-3.203	-19.775
5	02	19	H	H1	3.4	77.05	2.67	1.04	-03.302	-1.260	39.521	-4.013	4.39	-3.177	-19.661
TOTALS FOR FUNCTIONAL GROUP						93.37	3.00	0.96	-03.642	-1.215	40.853	-5.143	4.39	-3.178	-19.665
06	5	30	N	N3	4.2	17.30	0.29	0.50	-0.213	-0.739	1.250	-0.247	4.45	-3.107	-19.513
07	06	19	H	H3	3.6	82.74	2.91	1.05	-03.122	-1.073	34.606	-3.904	4.49	-3.153	-19.090
TOTALS FOR FUNCTIONAL GROUP						100.04	3.20	0.96	-03.336	-1.043	35.064	-4.231	4.49	-3.151	-19.076
AVERAGES OVER FUNCTIONAL GROUPS:—						96.70	3.10	0.96	-03.489	-1.129	38.358	-4.607	4.44	-3.164	-19.771
STATISTICAL UNCERTAINTY (+/- 2*SD)							0.13	0.04	0.373	0.099	0.013	0.190	0.03	0.040	0.094

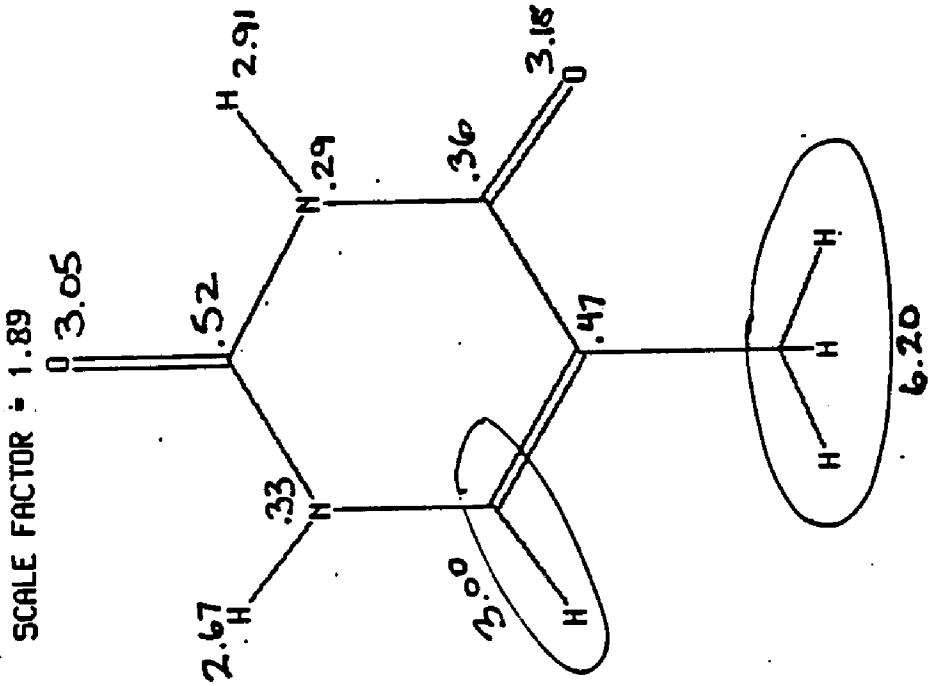
Table IV.3 Calculated structural and energetic quantities from Monte Carlo simulation of thymine and 215 waters at 298 K - Kollman solute-water potential functions.

THYMINE IN WATER AT 298K - FORCE BIAS AND KOLLMAN POTENTIAL FUNCTIONS

CARBONYL GROUPS													
08	3	01	C C2	4.2	19.68	0.52	0.78	-0.494	-0.959	01.678	-0.565	4.41	-3.149	-19.689
09	4	34	O 02	3.6	80.93	3.05	1.03	-2.401	-0.787	34.431	-3.671	4.35	-3.211	-19.701
TOTALS FOR FUNCTIONAL GROUP					100.61	3.56	0.90	-2.894	-0.812	36.101	-4.236	4.35	-3.200	-19.701
10	7	01	C C4	4.2	19.28	0.36	0.55	-0.164	-0.459	01.433	-0.101	4.39	-3.134	-19.516
11	0	34	O O4	3.0	100.31	3.10	0.95	-2.436	-0.766	37.375	-3.467	4.40	-3.137	-19.024
TOTALS FOR FUNCTIONAL GROUP					119.50	3.54	0.00	-2.000	-0.735	30.000	-3.040	4.40	-3.137	-19.013
AVERAGES OVER FUNCTIONAL GROUPS:--					114.10	3.55	0.93	-2.747	-0.774	37.454	-3.942	4.42	-4.172	-19.747
STATISTICAL UNCERTAINTY (+/- 2*SD)						0.15	0.04	0.297	0.069	0.013	0.100	0.03	0.040	0.095
THYMINE														
MOLECULAR SUM/AVERAGE:					739.20	22.97	0.93	-16.943	-0.730	215.000	-23.004	4.42	-3.100	-19.740
STATISTICAL UNCERTAINTY (+/- 2*SD)						0.50	0.02	1.002	0.039	0.044	0.501	0.02	0.024	0.050

Table IV.3 Calculated structural and energetic quantities from Monte Carlo simulation of thymine and 215 waters at 298 K - Kollman solute-water potential functions.

Figure IV.2.3 First shell coordination numbers for the atoms of thymine - Kollman solute-water potential function simulation.



Figures IV.2.4 through IV.2.10 - calculated quasicomponent distribution function of first shell coordination number for thymine and functional groups (Kollman potentials).

X axis - coordination number.

Y axis - quasicomponent of coordination number.

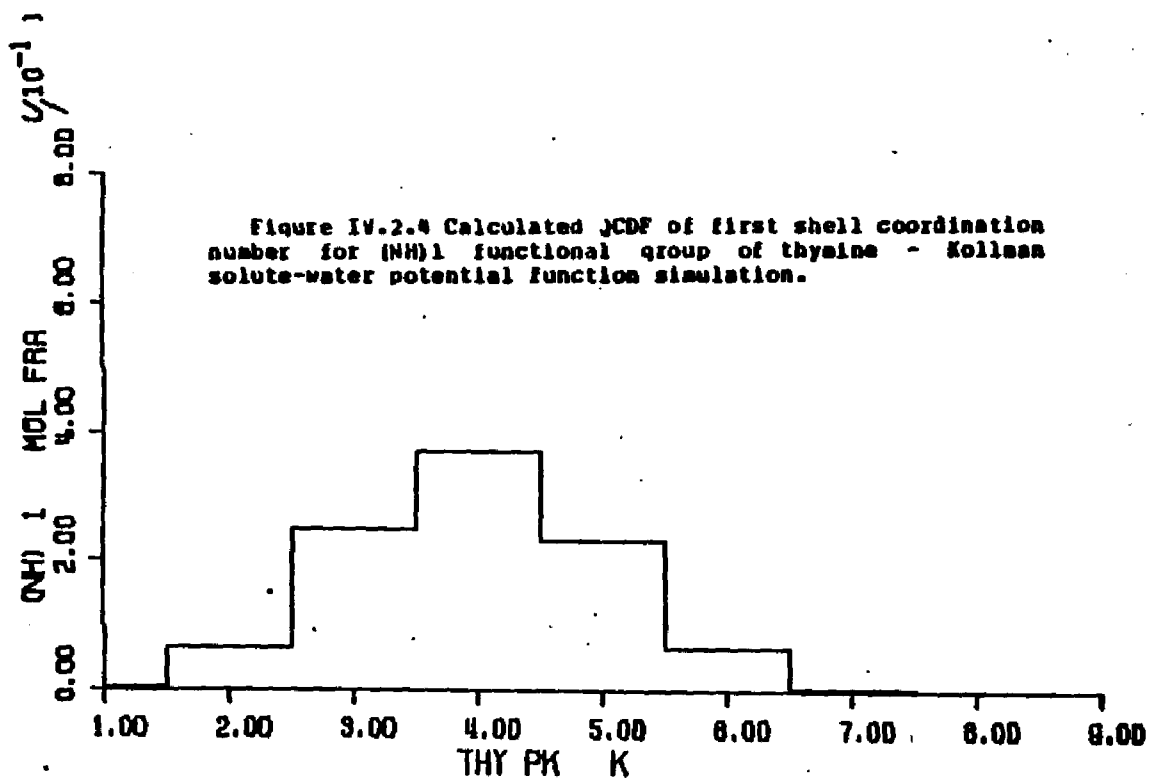


Figure IV.2.4 Calculated JCDF of first shell coordination number for (NH)1 functional group of thymine - Kollman solute-water potential function simulation.

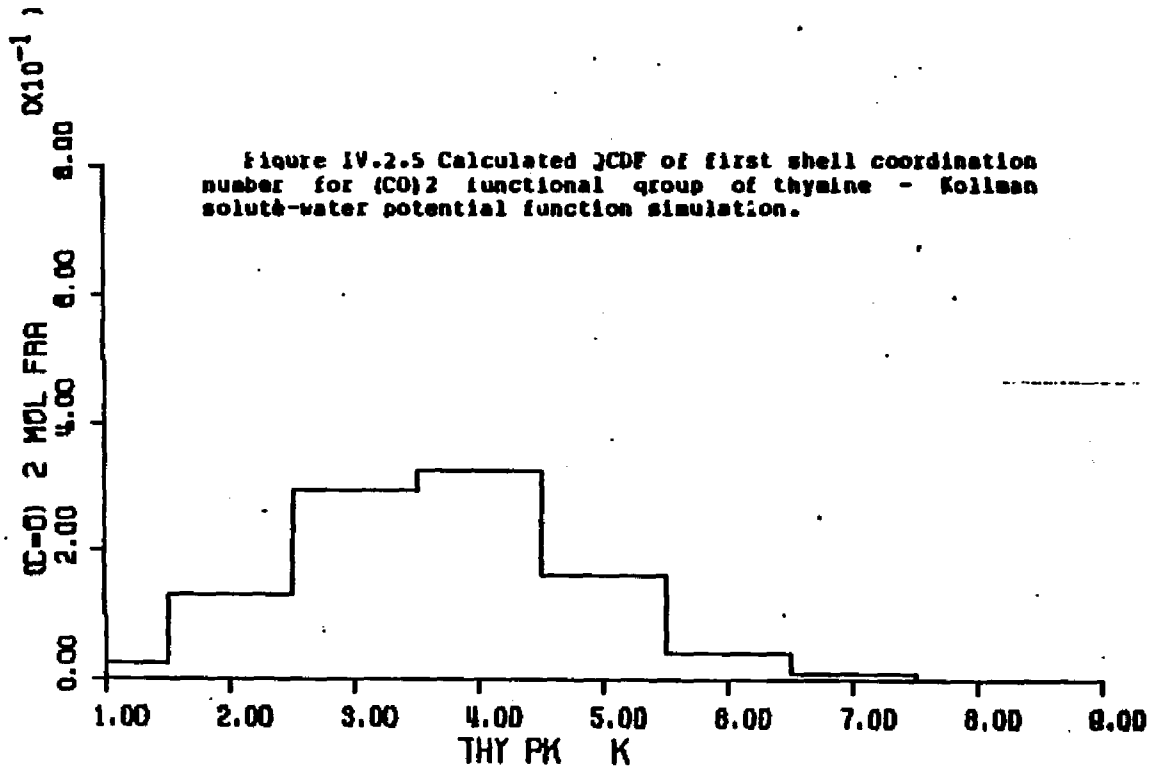


Figure IV.2.5 Calculated)CDF of first shell coordination number for (CO)2 functional group of thymine - Kollman solute-water potential function simulation.

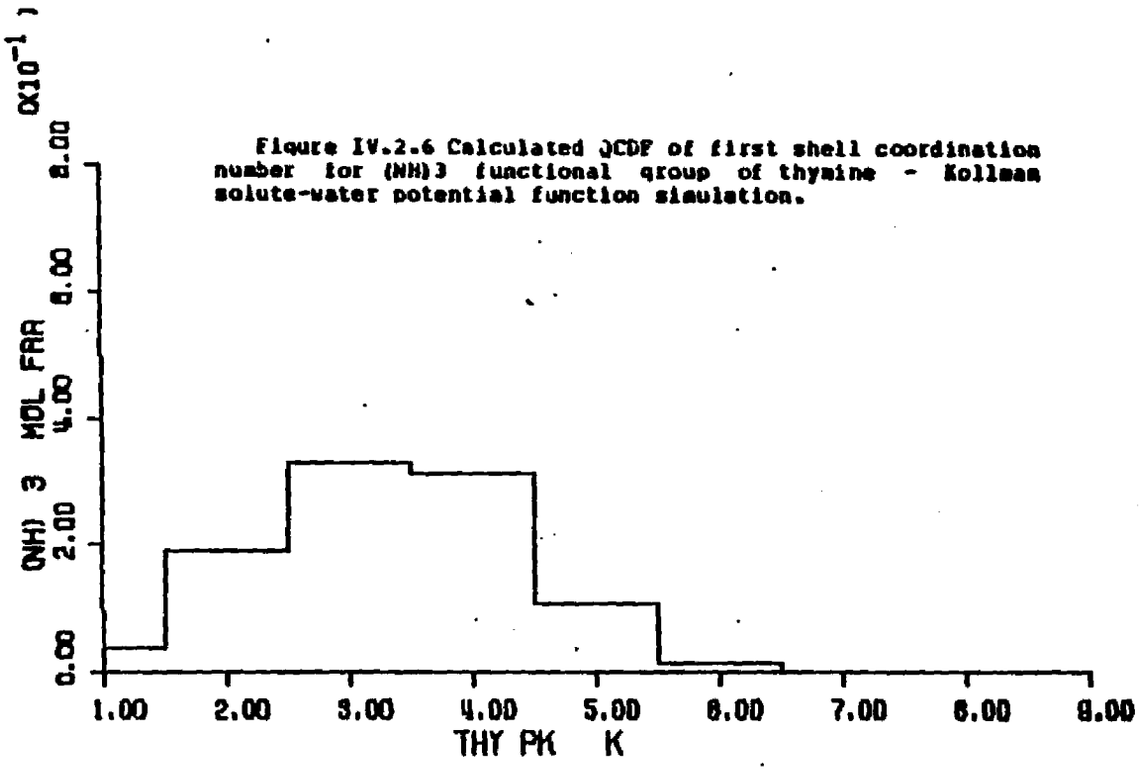


Figure IV.2.6 Calculated QCDF of first shell coordination number for (NH)3 functional group of thymine - Kollman solute-water potential function simulation.

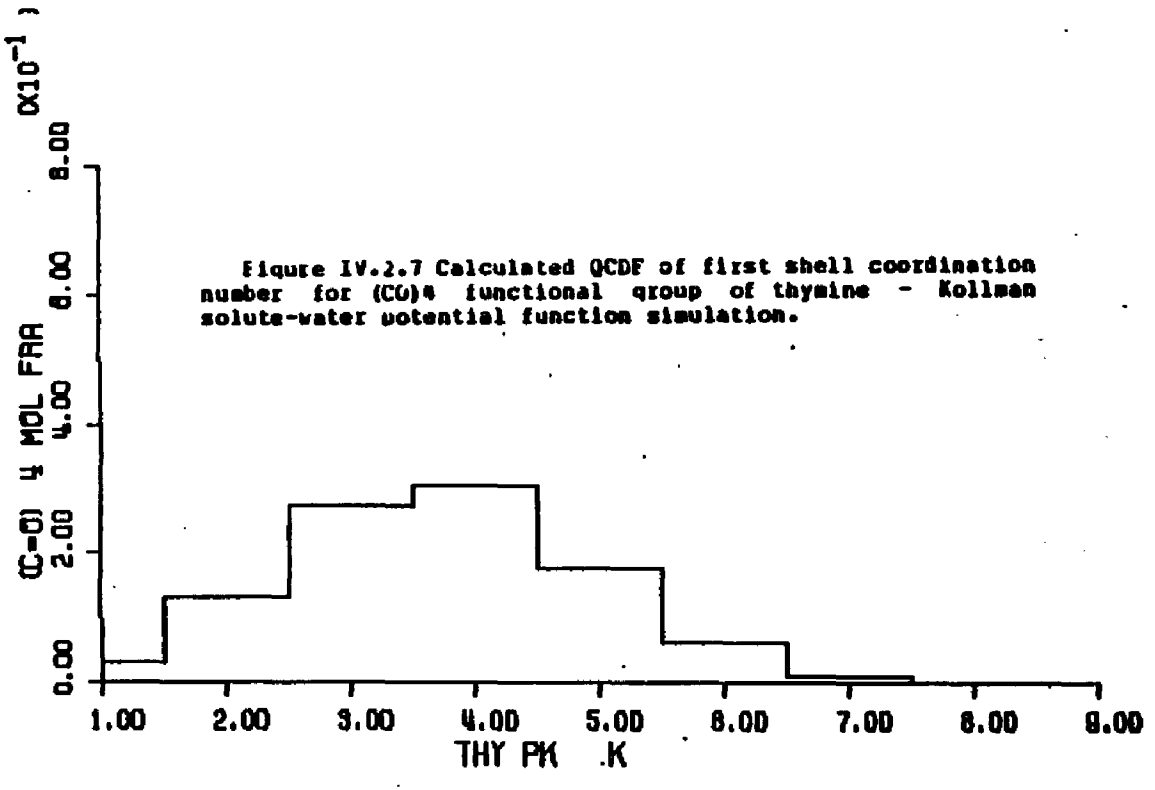


Figure IV.2.7 Calculated QCDF of first shell coordination number for $(CO)_4$ functional group of thymine - Kollman solute-water potential function simulation.

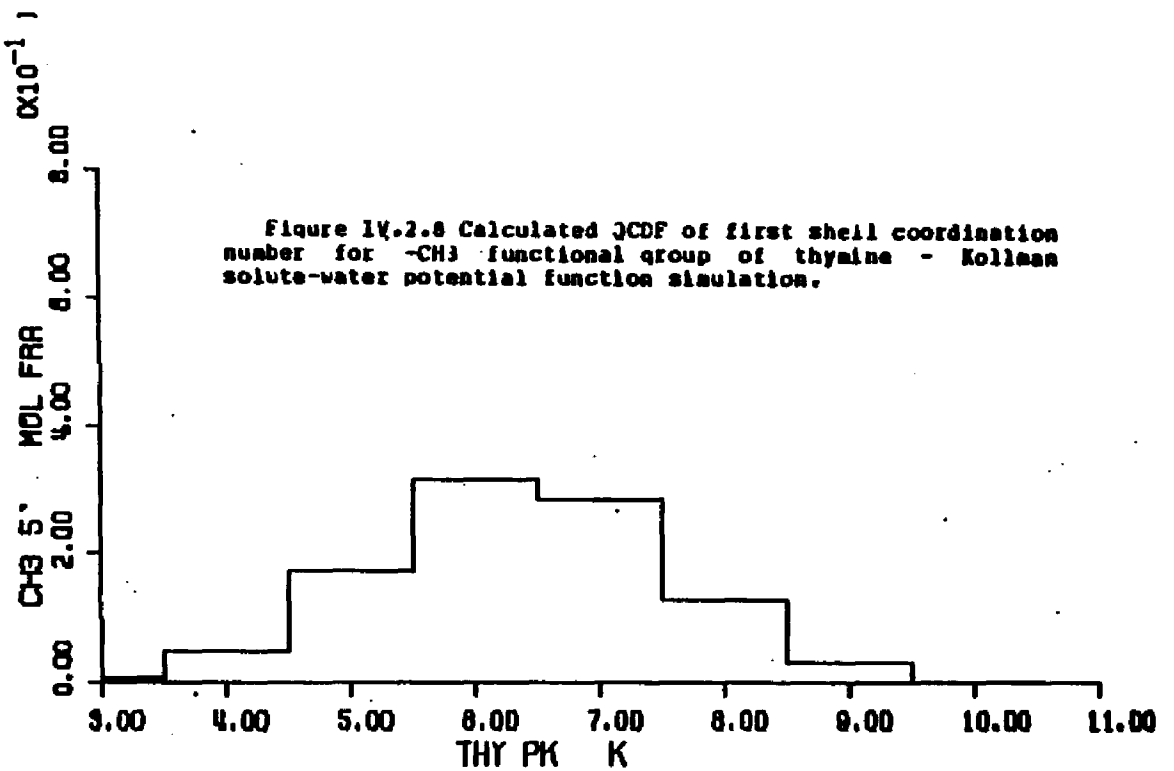


Figure IV.2.8 Calculated ICDF of first shell coordination number for -CH3 functional group of thymine - Kollman solute-water potential function simulation.

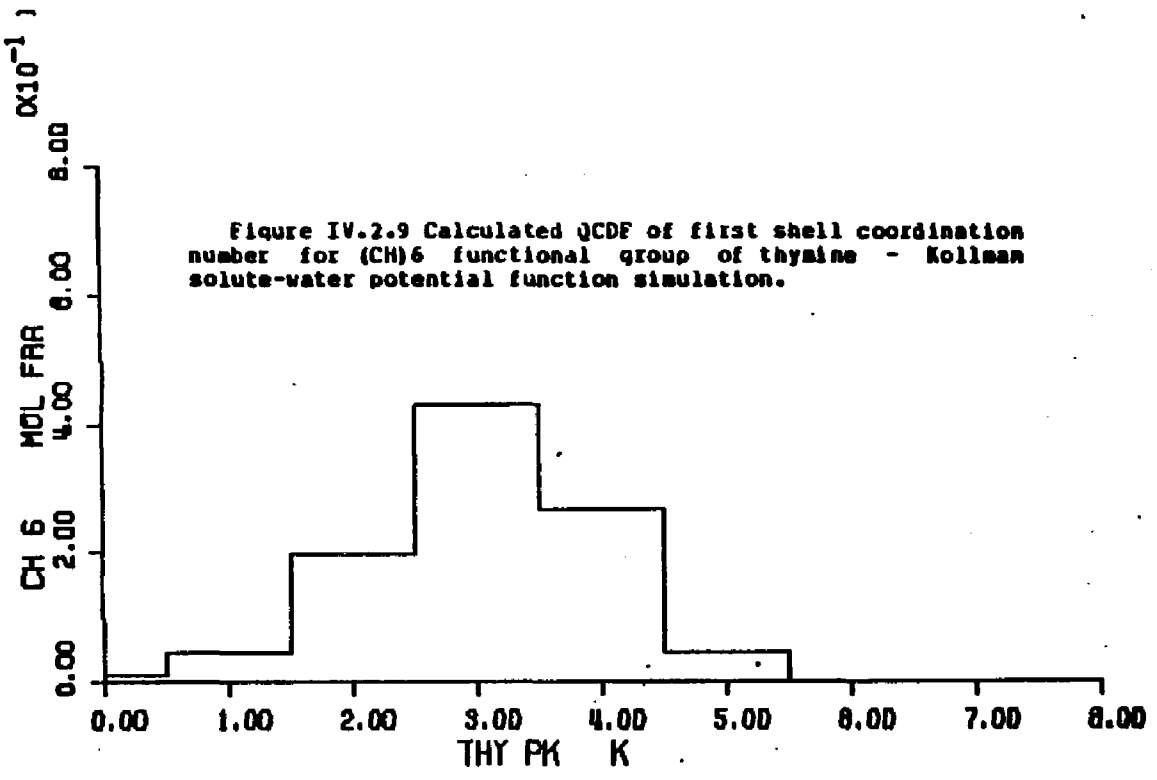


Figure IV.2.9 Calculated QCDF of first shell coordination number for (CH)6 functional group of thymine - Kollman solute-water potential function simulation.

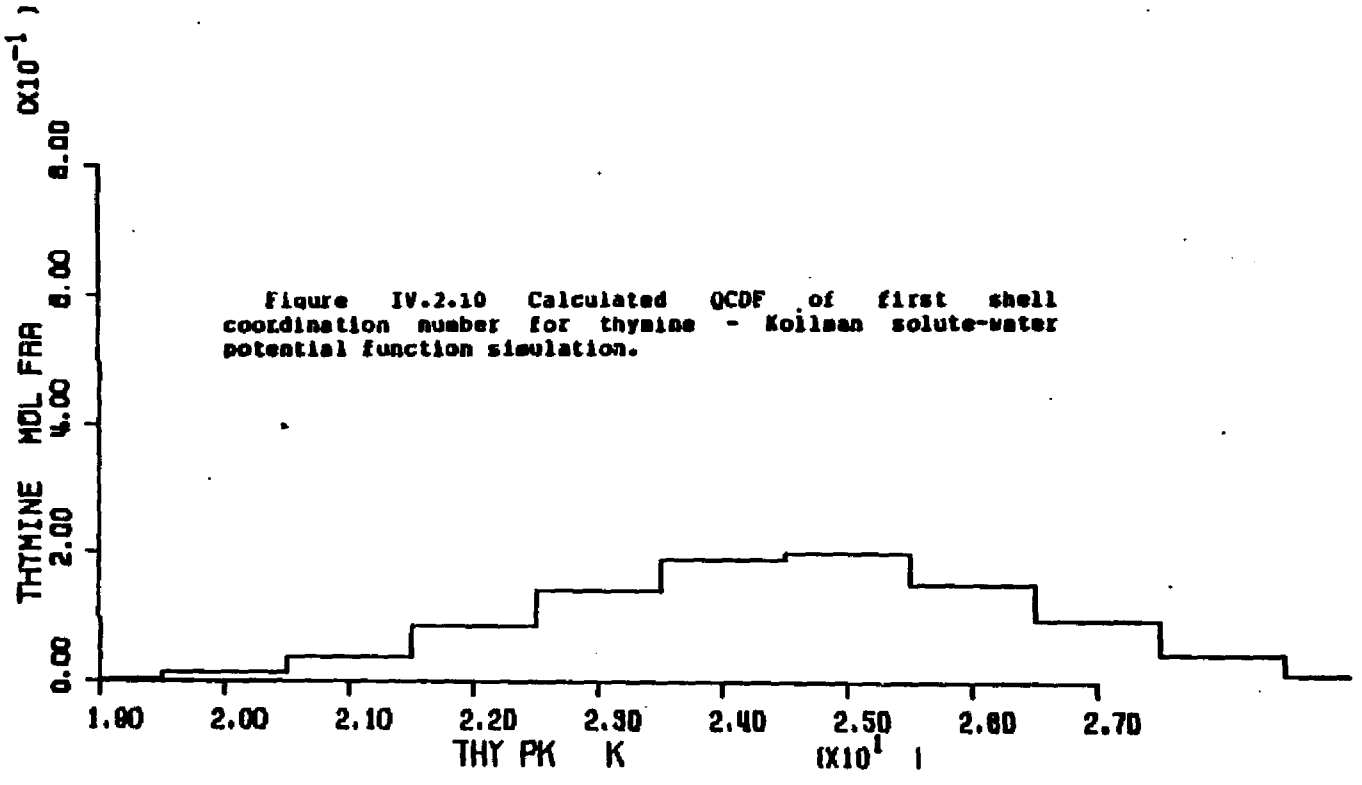


Figure IV.2.10 Calculated QCDF of first shell coordination number for thymine - Kollman solute-water potential function simulation.

Figure IV.2.11 Average first shell solute-water pair energies of waters assigned to the atoms of thymine - Kollman solute-water potential function simulation.

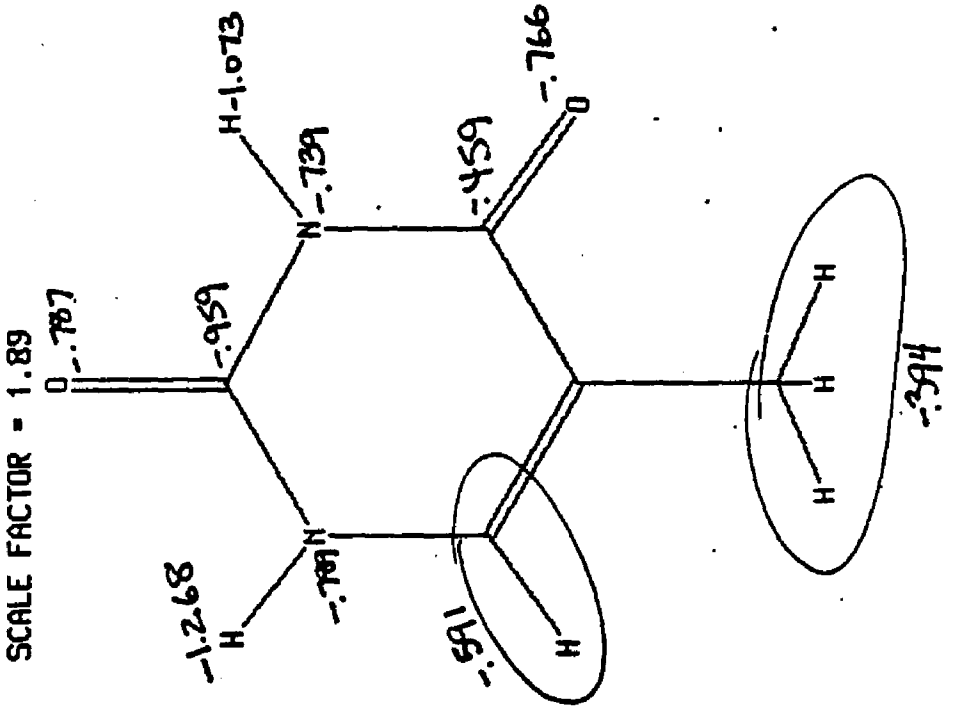


Figure IV.2.12 through IV.2.17 - calculated quasicomponent distribution function of average first shell pair energy for functional groups of thymine (Kollman potentials).

X axis - pair energy (kcal/mole).

Left Y axis - quasicomponent of pair energy.

Right Y axis - running coordination number.

Figure IV.2.12 Calculated QCDF of solute-water pair energies of waters of the (NH)1 functional group of thymine - Kollman solute-water potential function simulation.

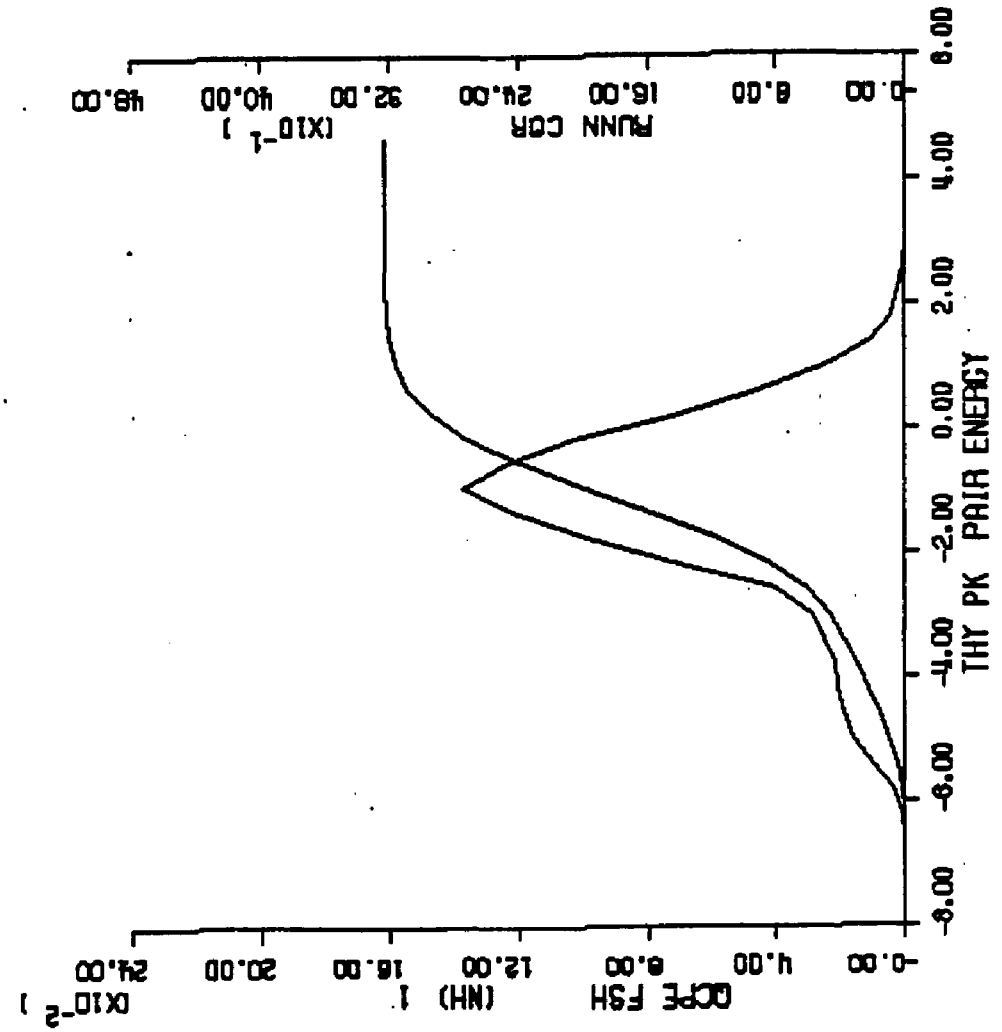


Figure IV.2.13 Calculated QCDF of solute-water pair energies of waters of the (CO)2 functional group of thymine - Kollman solute-water potential function simulation.

Figure IV.2.1a Calculated QCOF of solute-water pair energies of waters of the (NH)3 functional group of tyrosine - Kollman solute-water potential function simulation.

Figure IV.2.15 Calculated QCDF of solute-water pair energies of waters of the (CO)* functional group of thymine - Kollman solute-water potential function simulation.

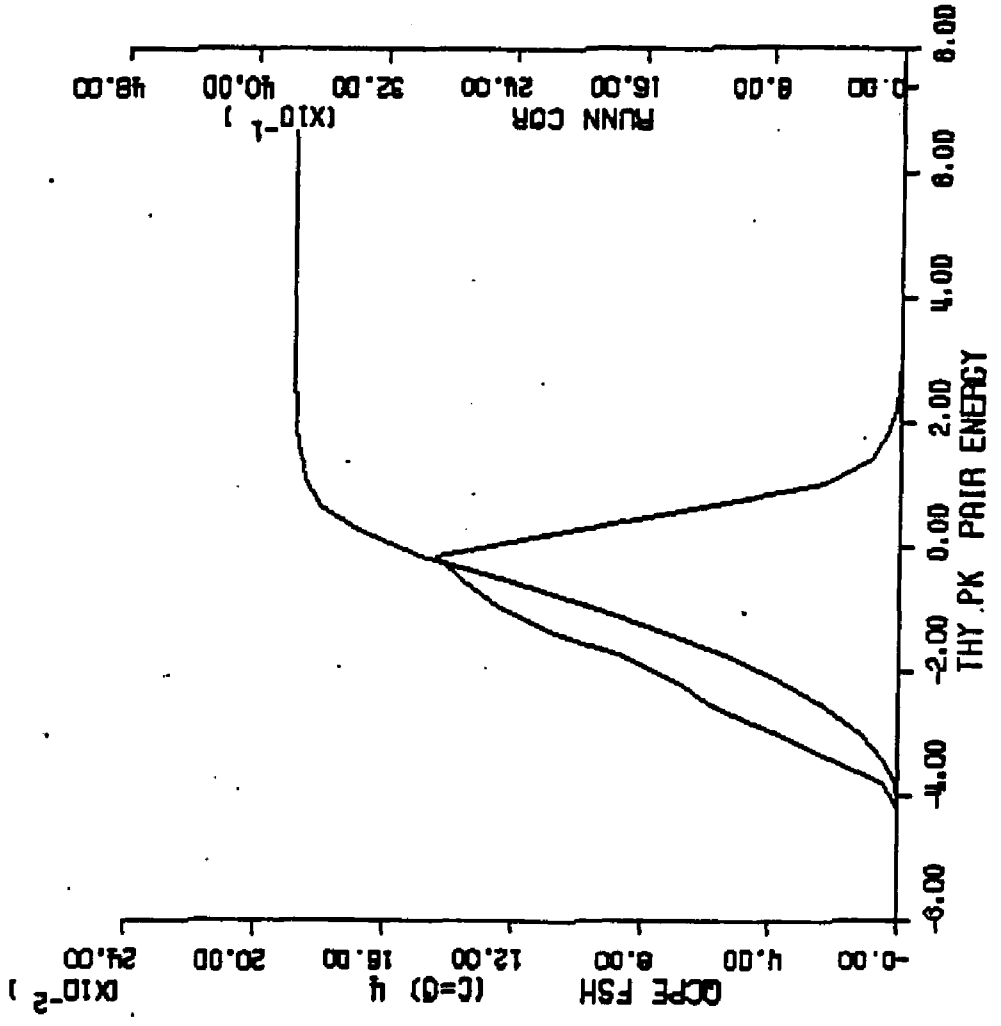


Figure IV.2.16 Calculated CCDF of solute-water pair energies of waters of the -CH3 functional group of thymine - Kollman solute-water potential function simulation.

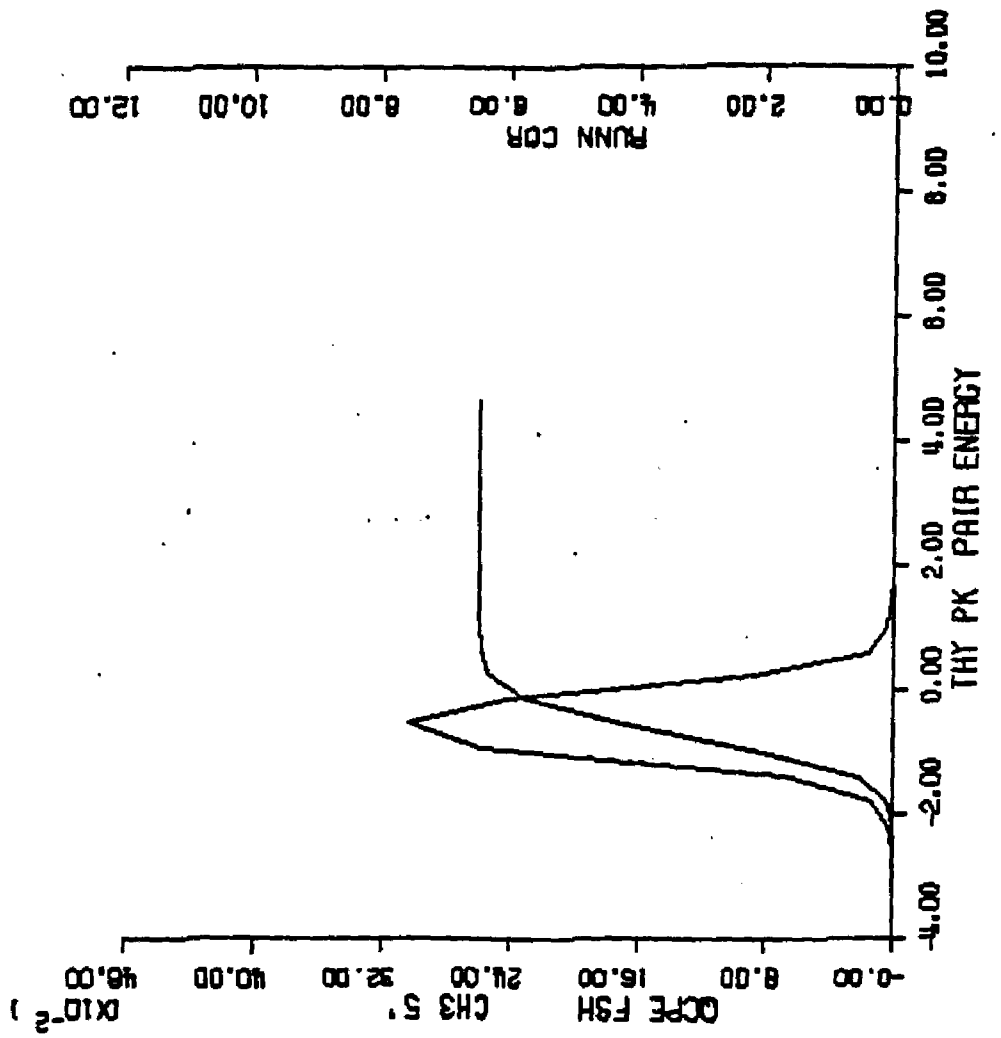


Figure IV.2.17 Calculated QCDF of solute-water pair energies of waters of the (CH)6 functional group of thymine - Kollman solute-water potential function simulation.

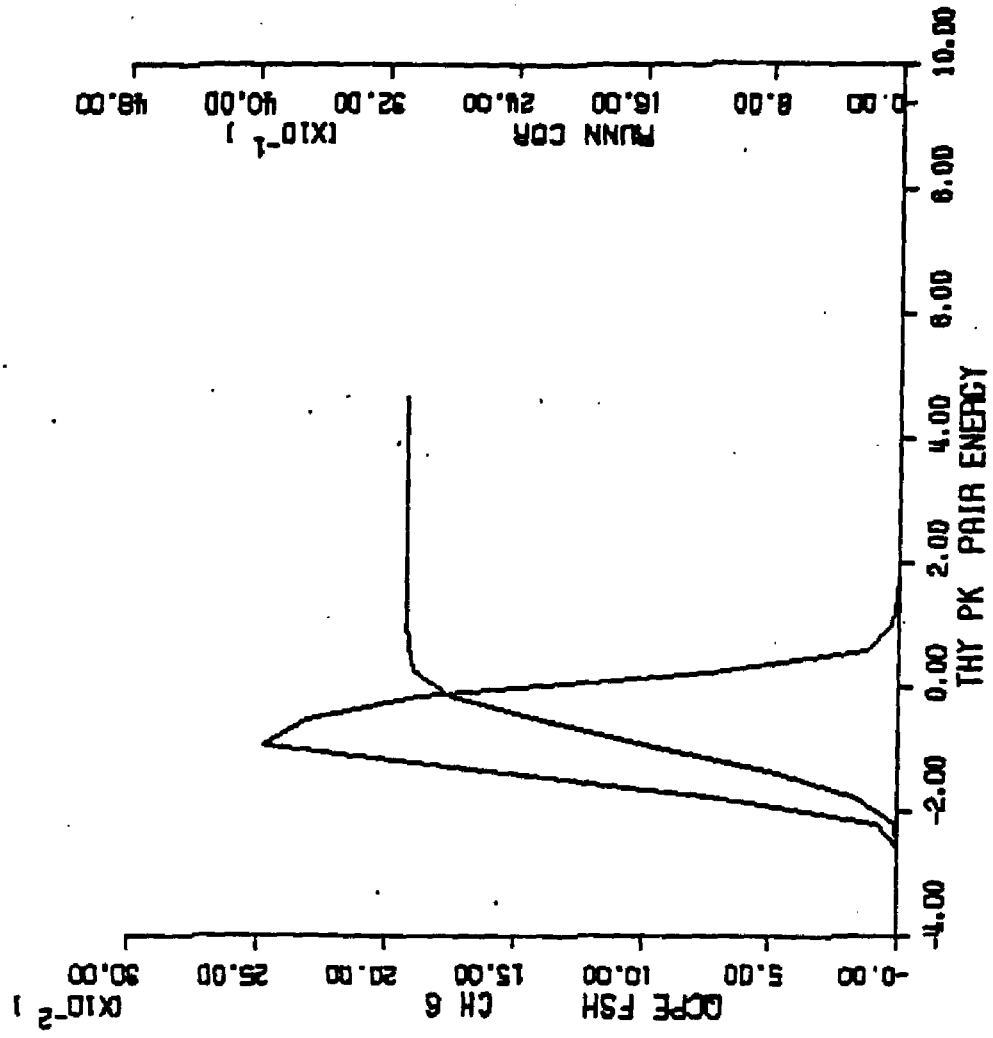
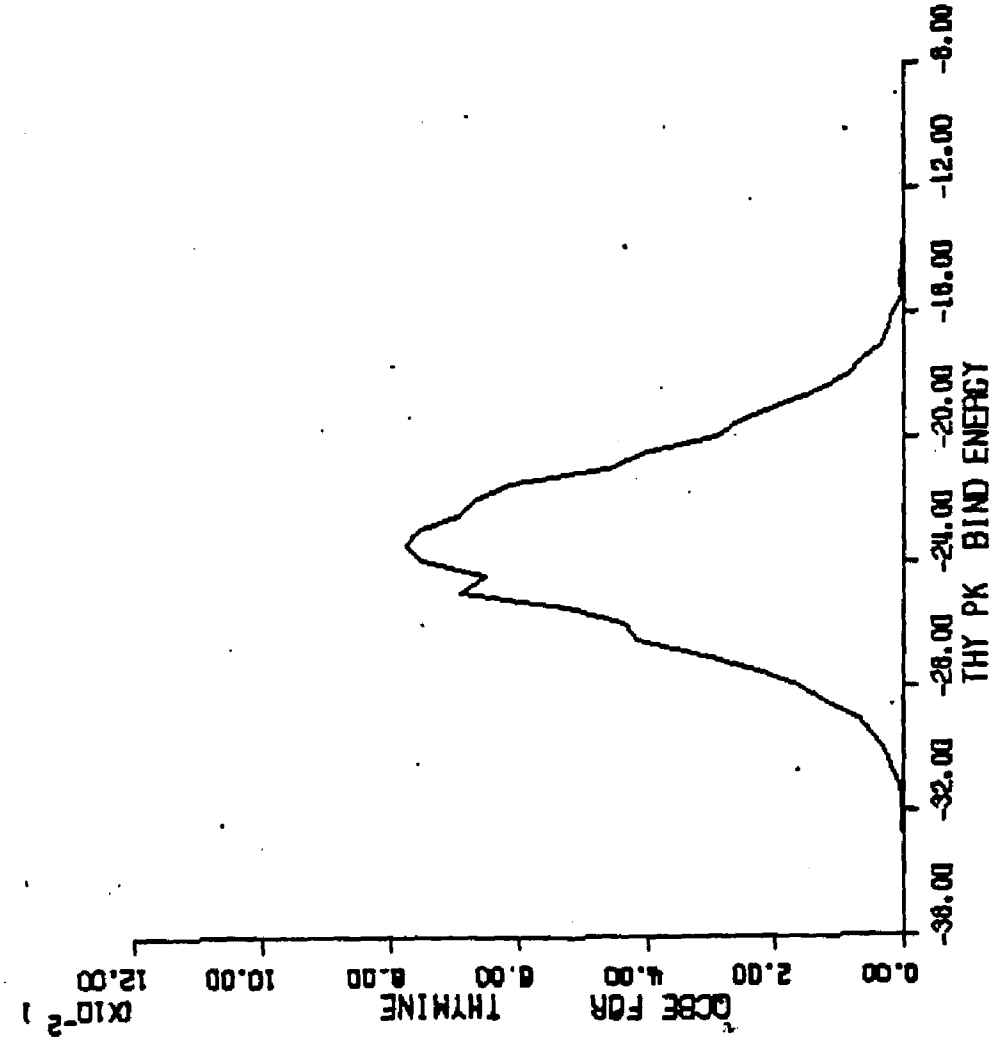


Figure IV.2.18 - calculated quasicomponent distribution function of binding energy for thymine (Kollman potentials).

X axis - binding energy (kcal/mole).

Y axis - quasicomponent of binding energy.

Figure IV.2.18 Calculated GCDF of total binding energy for thymine - Kollman solute-water potential function simulation.



3. Thymine With BVG Solute-Water Potentials

Computational Specifics. An aqueous solution of thymine and 215 waters was simulated using the BVG solute-water potentials and the SPC water-water potentials at 298 K. Simulation conditions were identical to the thymine simulation described previously using Clementi and Kollman solute-water potentials. Face centered cubic boundary conditions were used with a unit cell edge of 14.87 Angstroms. The simulation consisted of an equilibration period of 600,000 moves, which was discarded, and a production period of 2,000,000 moves. All analyses and ensemble averages were formed over the last 2000K configurations.

Potential Surface. The potential energy surface for the interaction of thymine and one water, using BVG solute-water potential functions for the calculation, is shown in Figure IV.3.1. The vicinity of the H3 imino hydrogen contains the most favorable solute-water interaction at an energy of -8.10 kcal/mole. Strong though less favorable interactions are found in the region of the H1 hydrogen at energies between -6.10 and -7.10 kcal/mole. The polar carbonyl oxygens interact more weakly with water. The O4 oxygen binds with energies between -3.1 and -4.1 kcal/mole and the O2 oxygen bind between -3.1 and -4.1 kcal/mole where it is flanked by the imino hydrogens and between -2.1 and -3.1 kcal/mole colinear with the C=O bond. The apolar methine region solute-water interactions range from -1.10 to

-2.10 kcal/mole. Interaction energies in the region of the methyl group range from -0.1 to -2.1 kcal/mole with the lower energies on the O4 side of the methyl group.

Convergence and Thermodynamic Results. The control functions for the BVG potential aqueous thymine simulation are shown in Figure IV.3.2; the calculated thermodynamic quantities are given in Table IV.1. The control function plots reveal that the total binding energy is rather stable for the last 1000K or so its value after 2000K is -2224.8 +/- 3.9 kcal/mole. The heat capacity reaches the value 15.4 cal/mole-K after 2000K configurations. The vacuum to water transfer energy is calculated to be -21.1 kcal/mole. Figures IV.3.4 thru IV.3.9 and IV.3.12 thru IV.3.17 contain distribution functions for the quasicomponents of coordination number and first shell pair energy for the imino, carbonyl, methyl and methine functional groups of thymine. Figures IV.3.10 and IV.3.18 contain graphs of the quasicomponents of coordination number and binding energy for the thymine molecule.

Structural Results. The first shell coordination numbers for the atoms of thymine are displayed on the molecular diagram of Figure IV.3.3. The imino hydrogens both have significant first shell populations. The H1 hydrogen contains 2.51 waters in a 3.4 Angstrom shell and the H3 hydrogen contains 3.24 waters in a 3.8 Angstrom

shell. Neither shell is packed at a high density; solvent densities are around bulk water values in both cases (H1: .97 x bulk water;H3: 1.03). This results suggests that little hydrophilic water concentration is occurring in these solvation shells. Similarly, the carbonyl oxygens have significant first shell populations (O2: 3.79 waters; O4: 2.83) with low solvent densities (O2: .99; O4: .96).

The apolar methyl group united atom, CH3 5', has a large first shell radial cutoff (4.6 Angstroms). A first shell population of 6.47 waters is calculated; the combination of extended first shell radius, large coordination number and low solvent density (.97) is generally seen with apolar solute-water interactions. The methine united atom, CH6, is assigned 1.99 waters in a 4.0 Angstrom first shell. The first shell solvent density is .88; thus, the methine united atom shows the characteristic hydration of apolar moieties.

The quasicomponent distribution function of coordination number for thymine is found in Figure IV.3.10. The coordination numbers range from 18 to 28 and the greatest contributions arises from coordination numbers of 23 and 24. The average coordination number is 22.84 +/- .85.

Energetic Results. The first shell solute pair energies for the waters assigned to the atoms of thymine

are displayed on the molecular diagram of Figure IV.3.11. The waters of the H3 imino hydrogen have the lowest average pair energies (-1.743 kcal/mole) and make the greatest contribution to the total solute binding energy (-5.647 kcal/mole). The water-water pair energies in the extended coordination shell of this atom are the most destabilized overall (-3.294 kcal/mole); this result is consistent with the solute-water pair energies found here. The waters of the H1 imino hydrogen show less favorable solute-water pair energies (average -.876 kcal/mole) presumably due to the influence of the neighboring apolar methine group.

The waters of the carbonyl oxygen O2 and O4 are rather weakly bound (average pair energies -.949 and -.947 kcal/mole,respectively). The average pair energies of the ring carbonyl carbon waters are actually greater than those of the carbonyl oxygen waters (-1.210 > -.949 kcal/mole for C=O 2 and -1.129 for C=O 4); the same unexpected ordering is also seen with N1 and H1.

The waters assigned to the CH3 5' methyl group interact the least favorably of all at an average -.590 kcal per solute-water interaction. The extended hydration shell of this ring substituent has the enhanced water-water pair energies (-3.425 kcal/mole vs. -3.390 kcal/mole average for this simulation) expected with apolar hydration.

The quasicomponent distribution function of solute-water binding energies for thymine is displayed in Figure IV.3.18. Energy values range from -39 to -17 kcal/mole; the most significant contributions range from -32 to -25 kcal/mole. The average solute binding energy is -28.6 +/- .9 kcal/mole.

Figure IV.3.1 - isoenergy contour surface for thymine and one water (BVG solute-water potential functions).

X axis - X axis of plane defined by molecular ring atoms (Angstroms).

Y axis - Y axis of plane defined by molecular ring atoms (Angstroms).

Contours - isoenergy contours with one kcal/mole increments with alphabetical labels referring to contour energy values in list at right.

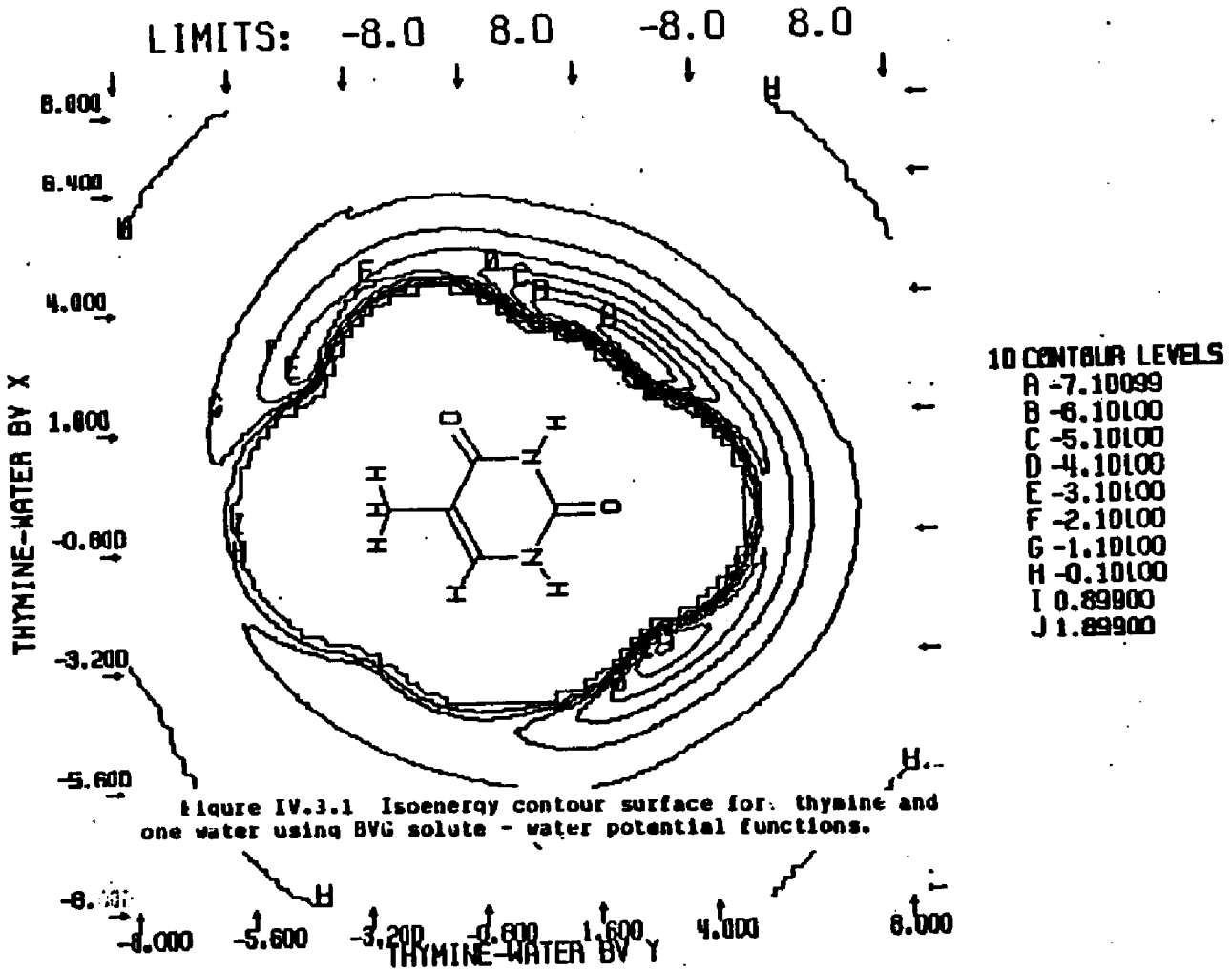


Figure IV.3.1 Isoenergy contour surface for thymine and one water using BVG solute - water potential functions.

Figure IV.3.2 - control functions for Monte Carlo simulation of thymine and 215 waters (BVG solute-water potential functions).

X axis - number of configurations.

Left Y axis - mean total energy (kcal/mole).

Right Y axis. - constant volume heat capacity (cal/mole-degree).

Upper curve - constant volume heat capacity.

Bottom curve without crosshatches - average total energy for entire simulation.

Bottom curve with crosshatches - average total energy for preceeding 50K configurations.

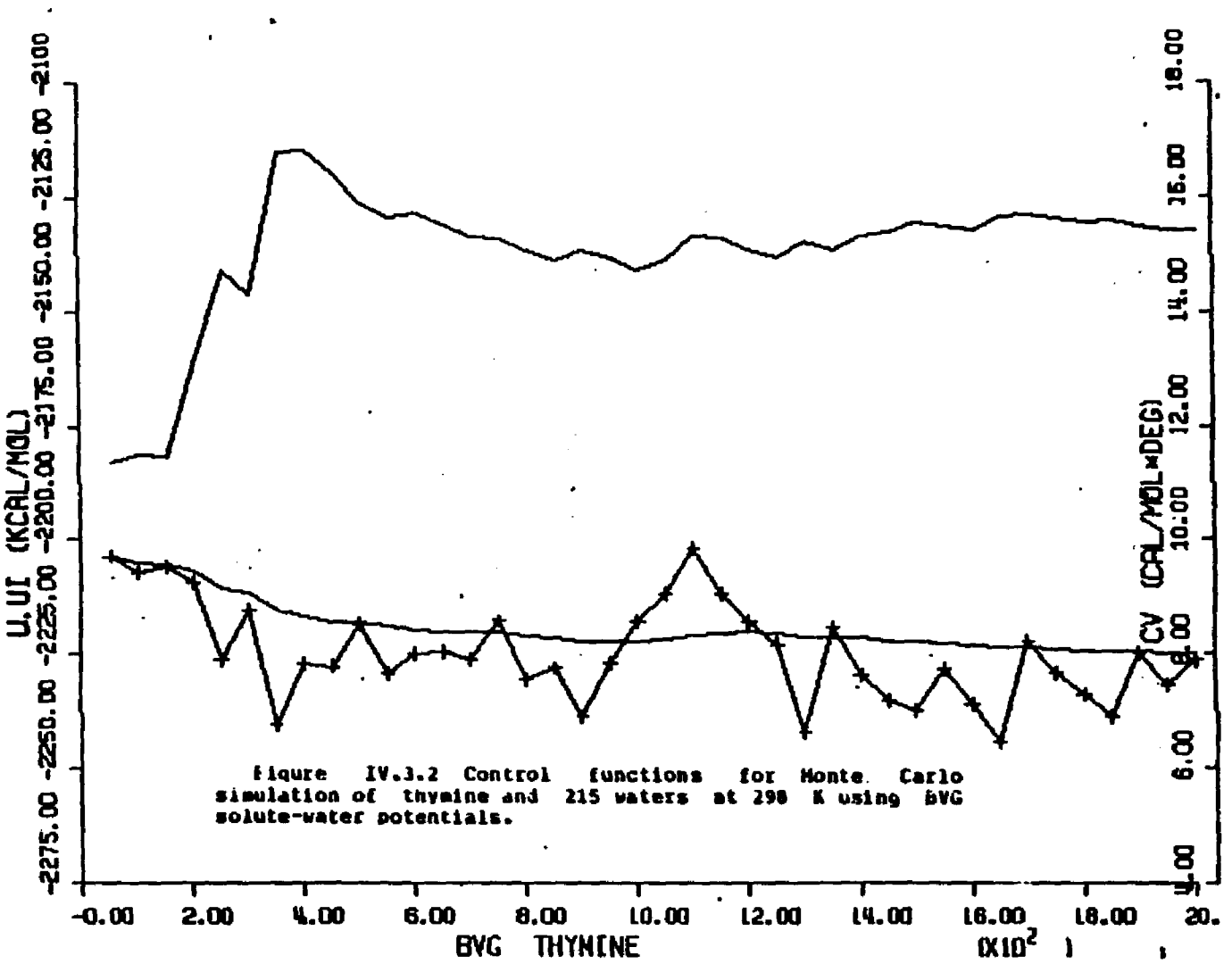


Figure IV.3.2 Control functions for Monte Carlo simulation of thymine and 215 waters at 298 K using BVG solute-water potentials.

THYMINE IN WATER AT 298K - FORCE BIAS AND BVG POTENTIAL FUNCTIONS

LAST CONFIGURATION: 2000002 FIRST SHELL SOLUTE PROPERTIES TOTAL SLT PROPS WATER PROPERTIES RMIN=3.30 RCN= 7.75 A

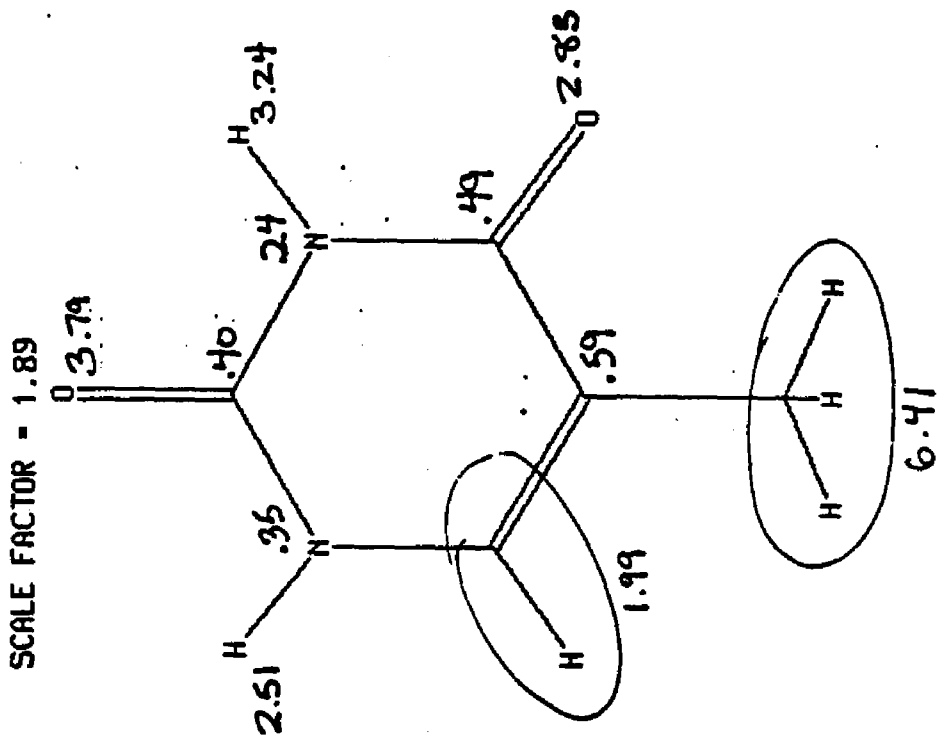
INDEX	TYPE				RPS	VPS	<E>	<E/V>	<SLTBE>	<SLTPE>	<E>	<SLTBE>	<RD>	<RMAPE>	<RMAT>
METHYLENE (DISUBST.)															
1	18	18	C	5	4.2	22.25	0.59	0.79	-0.656	-1.116	1.82	-0.745	4.18	-3.477	-20.121
STATISTICAL UNCERTAINTY (+/- 2*SD):							0.14	0.18	0.243	0.469	0.81	0.386	0.25	0.161	0.645
METHINE GROUP															
2	3	17	CH6		4.8	67.29	1.99	0.88	-1.676	-0.843	16.15	-2.234	4.31	-3.447	-20.495
STATISTICAL UNCERTAINTY (+/- 2*SD):							0.25	0.11	0.338	0.193	0.82	0.388	0.89	0.054	0.228
METHYL GROUP															
3	11	15	CH3	5'	4.6	197.96	6.41	0.97	-3.783	-0.598	45.97	-4.642	4.36	-3.425	-20.578
STATISTICAL UNCERTAINTY (+/- 2*SD):							0.45	0.07	0.425	0.075	0.83	0.379	0.85	0.032	0.131
AMIDE GROUP															
4	1	31	N	1	4.4	18.82	0.35	0.58	-0.354	-1.012	1.34	-0.436	4.34	-3.461	-20.687
5	2	19	H	1	3.4	77.85	2.51	0.97	-2.198	-0.876	38.41	-3.813	4.34	-3.434	-20.510
TOTALS FOR FUNCTIONAL GROUP >NH :						95.87	2.86	0.90	-2.552	-0.893	39.75	-4.249	4.34	-3.435	-20.513
6	6	31	N	3	4.8	16.47	0.24	0.43	-0.238	-0.995	1.26	-0.322	4.35	-3.383	-20.315
7	7	19	H	3	3.8	94.89	3.24	1.03	-5.647	-1.743	35.37	-6.653	4.51	-3.294	-20.628
TOTALS FOR FUNCTIONAL GROUP >NH :						110.56	3.48	0.94	-5.886	-1.691	36.63	-6.975	4.51	-3.295	-20.610
AVERAGES OVER FUNCTIONAL GRP >NH :						102.82	3.17	0.92	-4.219	-1.292	38.19	-5.612	4.42	-3.365	-20.561
STATISTICAL UNCERTAINTY (+/- 2*SD):							0.22	0.06	0.477	0.165	0.82	0.356	0.84	0.024	0.102

Table IV.4 Calculated structural and energetic quantities from Monte Carlo simulation of thymine and 215 waters at 298 K - B VG solute-water potential functions.

Table IV.4 Calculated structural and energetic quantities from Monte Carlo simulation of thymine and 215 waters at 298 K - BVG solute-water potential functions.

CARBONYL GROUP													
8	4	18	C	2	4.2	19.68	0.48	0.61	-0.486	-1.218	1.62	-0.584	4.36	-3.442	-20.716
9	5	3	O	2	4.8	114.73	3.79	0.99	-3.598	-0.949	34.69	-4.454	4.39	-3.418	-20.589
TOTALS FOR FUNCTIONAL GROUP C=O :						134.41	4.19	0.93	-4.084	-0.974	36.31	-5.038	4.39	-3.419	-20.594
10	8	18	C	4	4.8	10.34	0.49	0.81	-0.558	-1.129	1.62	-0.668	4.21	-3.362	-19.942
11	9	3	O	4	3.6	88.21	2.83	0.96	-2.676	-0.947	36.76	-3.723	4.42	-3.341	-20.388
TOTALS FOR FUNCTIONAL GROUP C=O :						106.55	3.32	0.93	-3.234	-0.974	38.37	-4.383	4.41	-3.342	-20.361
AVERAGES OVER FUNCTIONAL GRP C=O :						120.48	3.76	0.93	-3.659	-0.974	37.34	-4.711	4.40	-3.380	-20.478
STATISTICAL UNCERTAINTY (+/- 2*SD):							0.24	0.06	0.388	0.114	0.82	0.382	0.04	0.024	0.182
MOLECULAR SUM/AVERAGE:						734.1	22.84	0.93	-21.871	-0.958	215.80	-28.267	4.39	-3.398	-20.525
STATISTICAL UNCERTAINTY (+/- 2*SD):							0.85	0.03	1.382	0.865	0.86	1.868	0.02	0.014	0.861

Figure IV.3.3 First shell coordination numbers for the atoms of thymine - BVG solute-water potential function simulation.



Figures IV.3.4 through IV.3.10 - calculated quasicomponent distribution function of first shell coordination number for thymine and functional groups (BVG potentials).

X axis - coordination number.

Y axis - quasicomponent of coordination number.

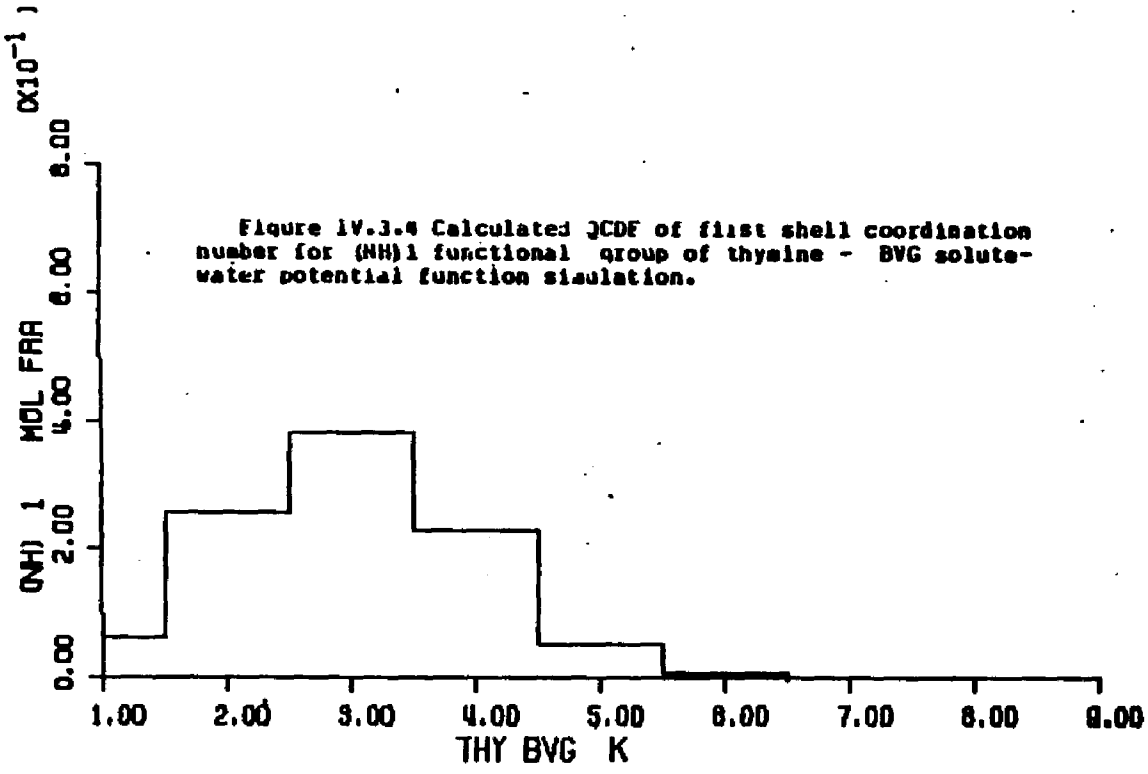


Figure IV.3.4 Calculated JCDF of first shell coordination number for (NH)1 functional group of thymine - BVG solute-water potential function simulation.

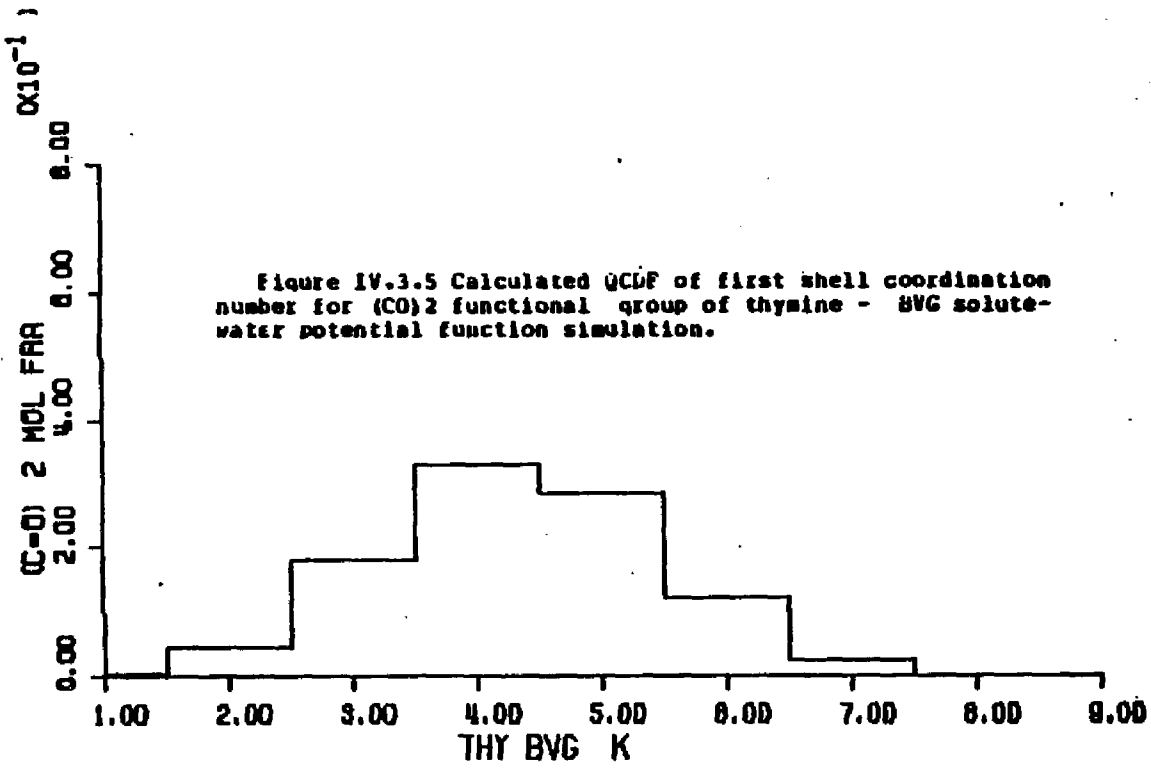


Figure IV.3.5 Calculated QCDF of first shell coordination number for (CO)2 functional group of thymine - BVG solute-water potential function simulation.

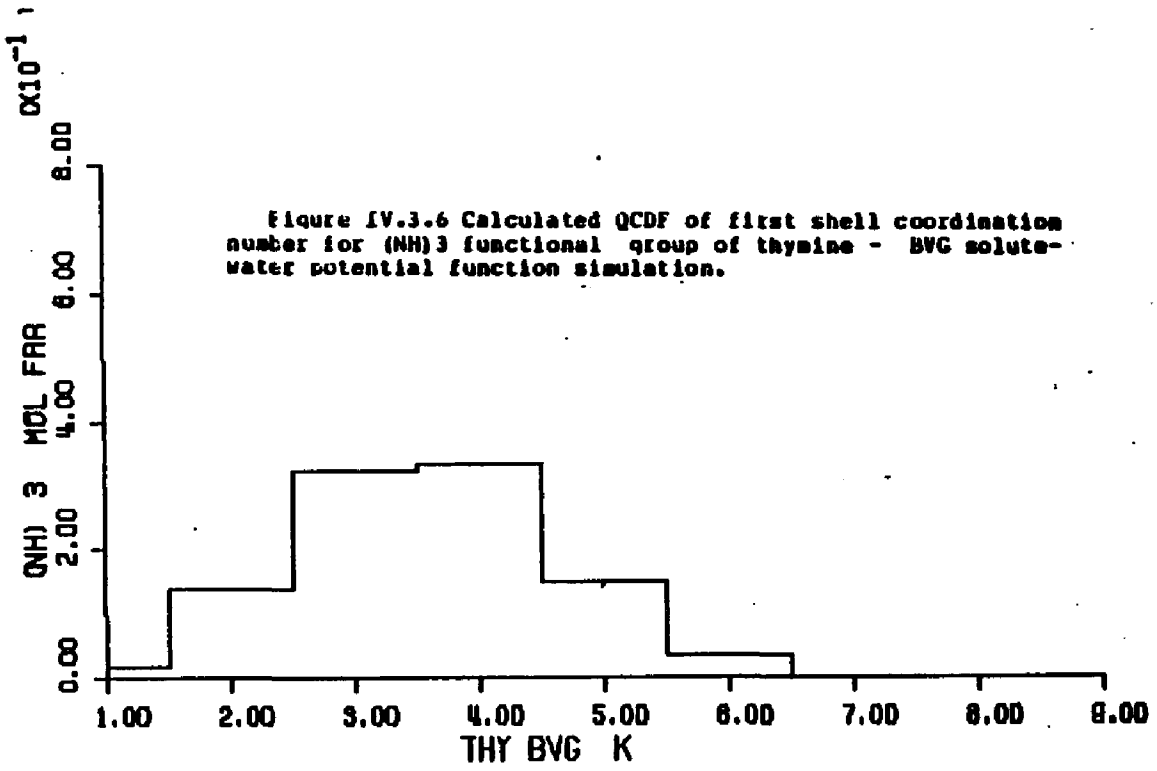


Figure IV.3.6 Calculated QCDF of first shell coordination number for (NH)3 functional group of thymine - BVG solute-water potential function simulation.

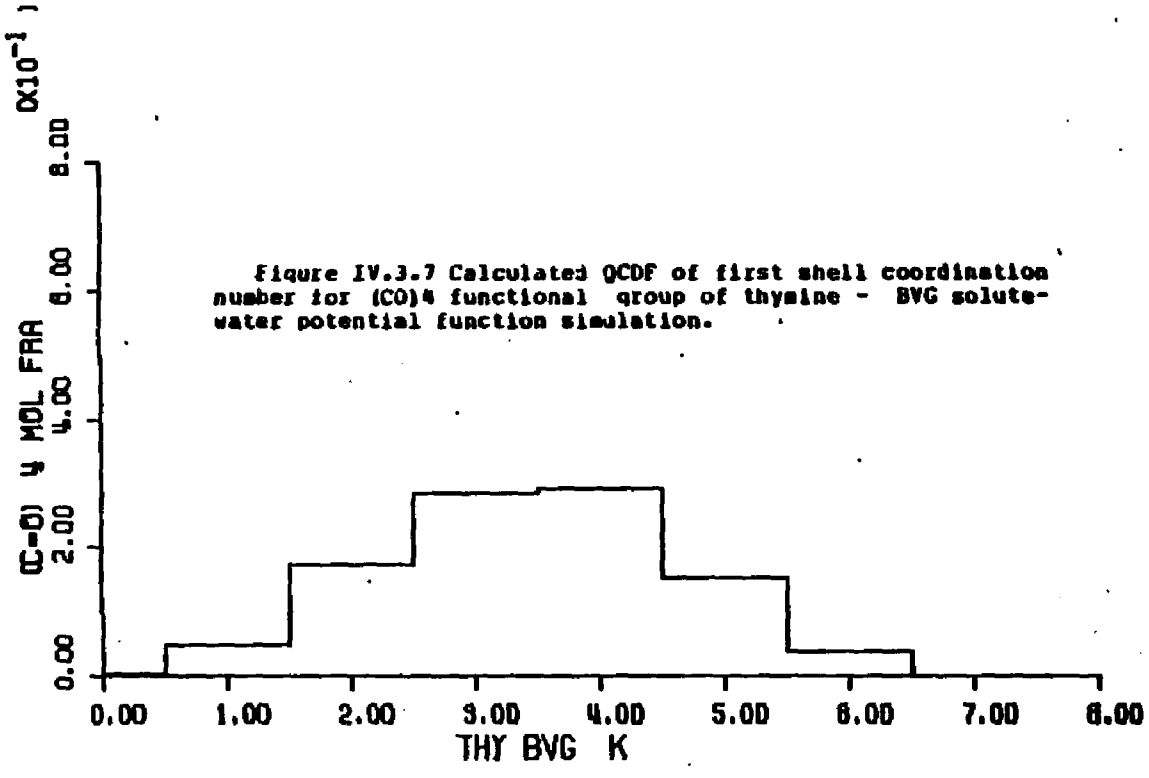


Figure IV.3.7 Calculated QCDF of first shell coordination number for (CO)4 functional group of thymine - BVG solute-water potential function simulation.

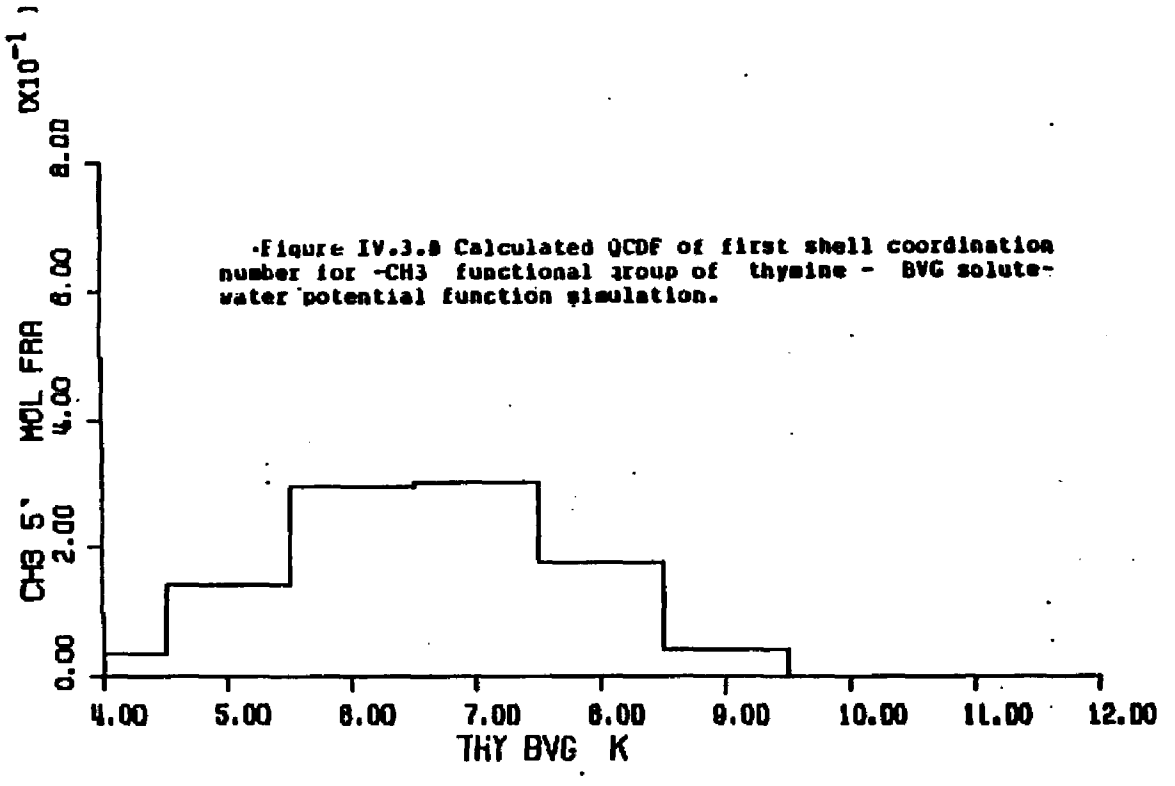


Figure IV.3.8 Calculated QCDF of first shell coordination number for -CH3 functional group of thymine - BVG solute-water potential function simulation.

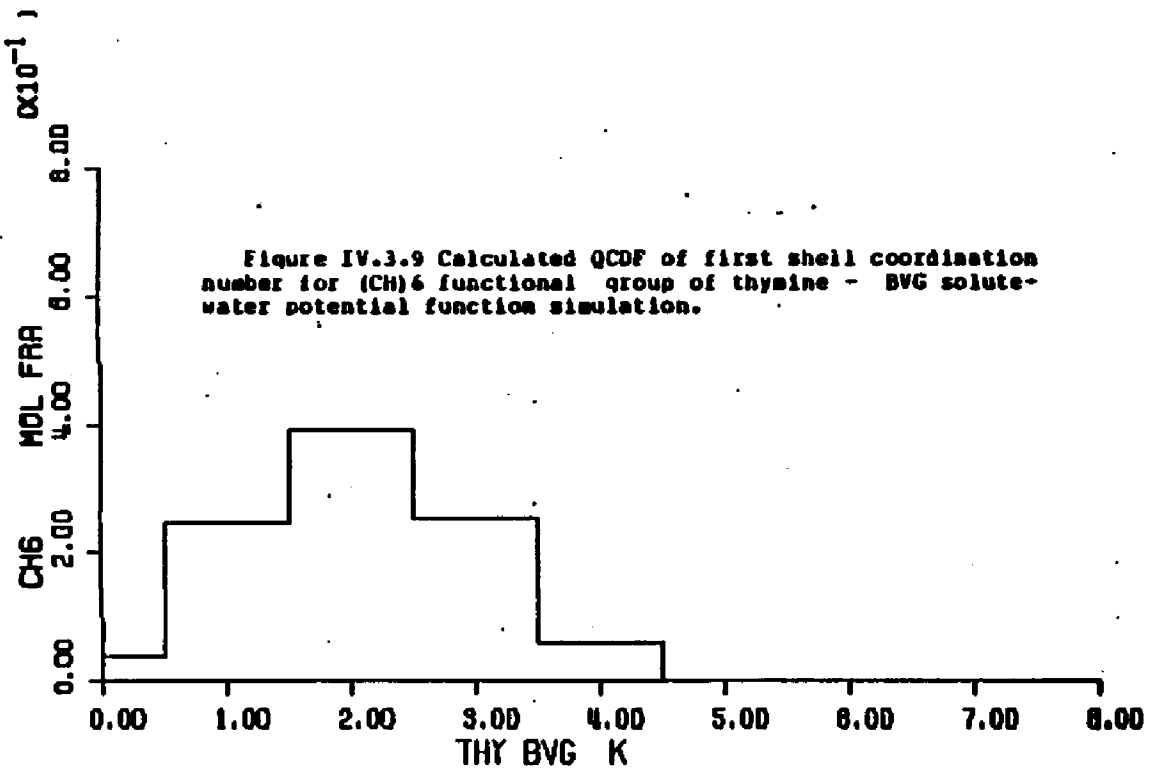


Figure IV.3.9 Calculated QCDF of first shell coordination number for (CH)6 functional group of thymine - BVG solute-water potential function simulation.

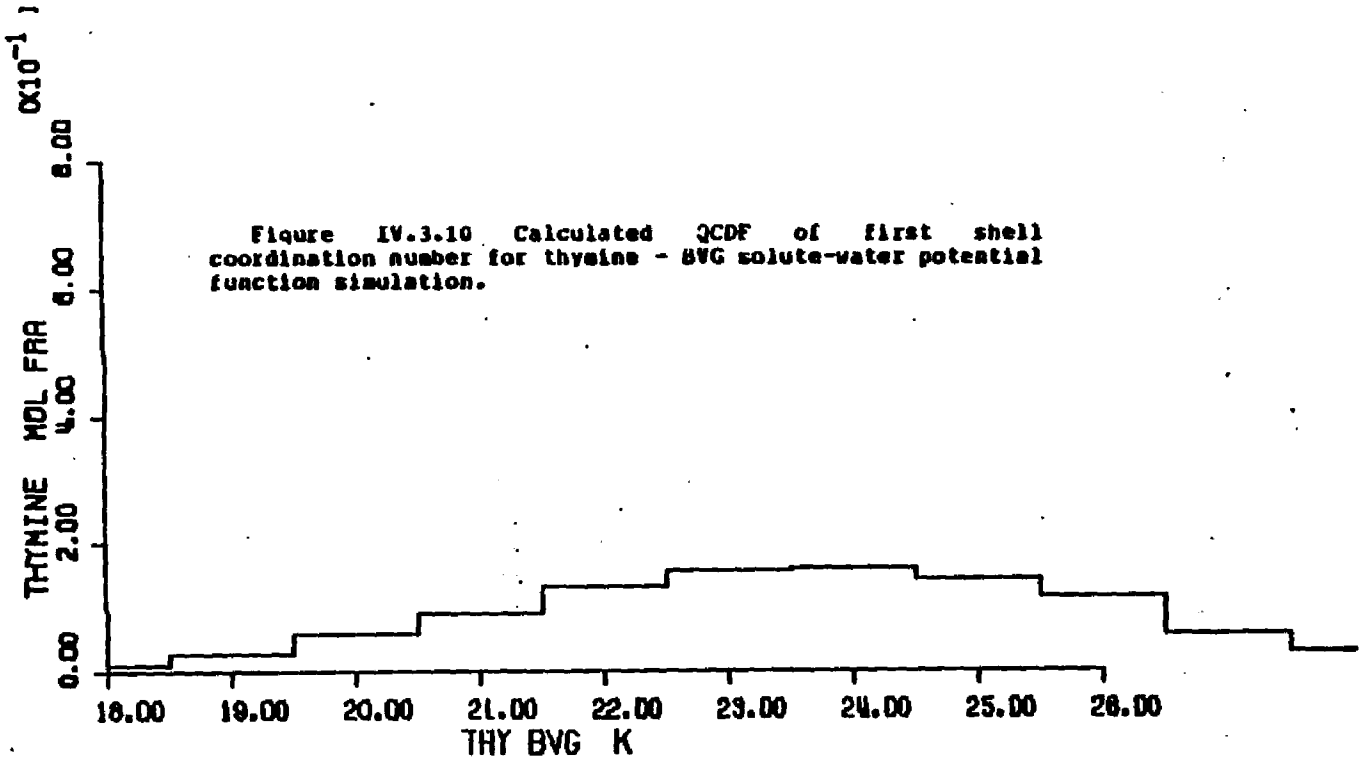


Figure IV.3.10 Calculated QCDF of first shell coordination number for thymine - BVG solute-water potential function simulation.

Figure IV.3.11 Average first shell solute-water pair energies of waters assigned to the atoms of thymine - BVG solute-water potential function simulation.

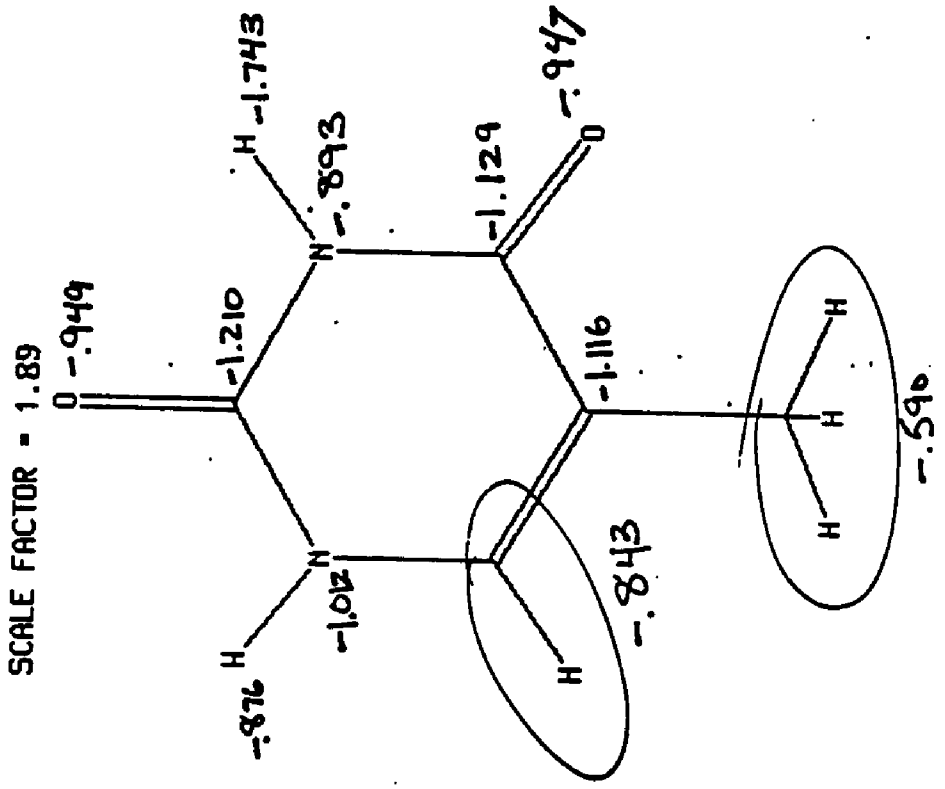


Figure IV.3.12 through IV.3.17 - calculated quasicomponent distribution function of average first shell pair energy for functional groups of thymine (BVG potentials).

X axis - pair energy (kcal/mole).

Left Y axis - quasicomponent of pair energy.

Right Y axis - running coordination number.

Figure IV.3.12 Calculated QCDF of solute-water pair energies of waters of the (NH)1 functional group of thymine - BVC solute-water potential function simulation.

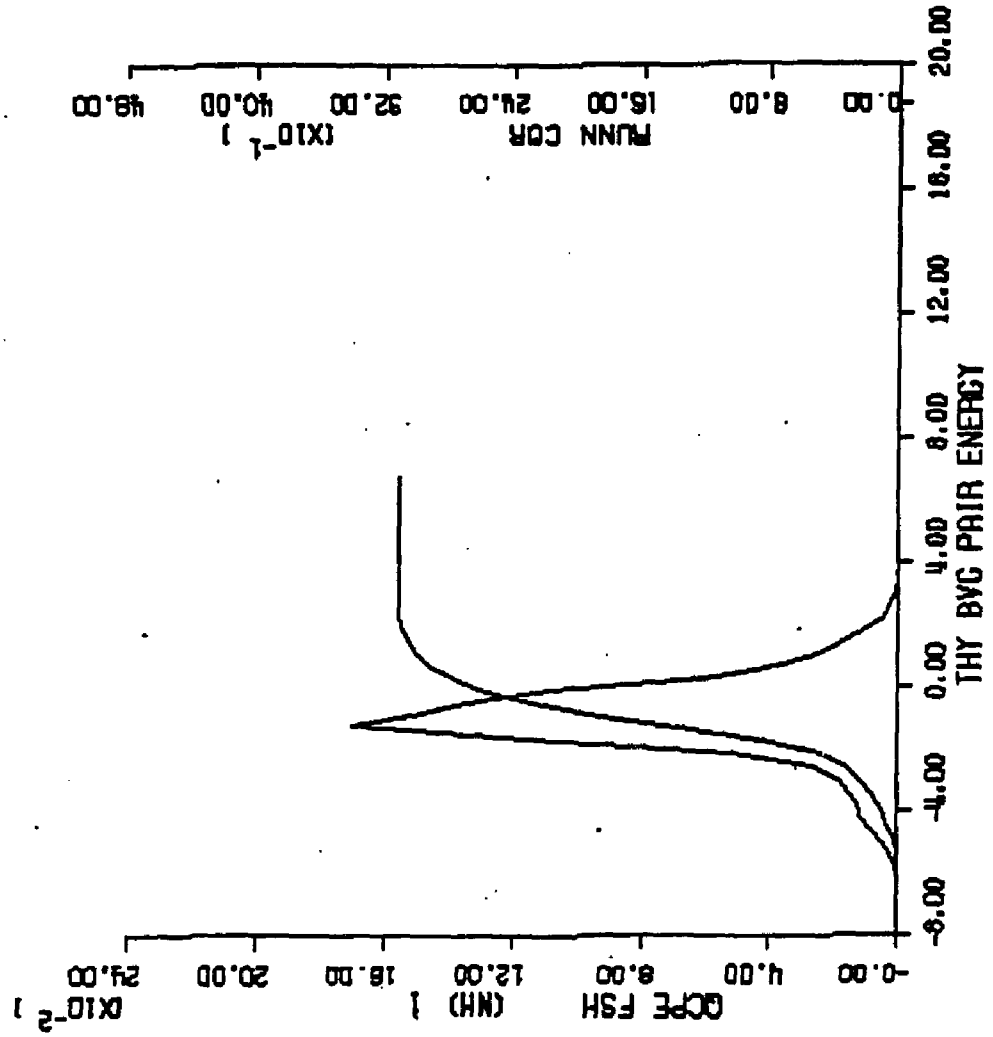


Figure IV.3.13 Calculated QCDF of solute-water pair energies of waters of the $(CO)_2$ functional group of thymine - SPC solute-water potential function simulation.

Figure IV.3.14 Calculated QCDF of solute-water pair energies of waters of the (NH)3 functional group of thymine - BVG solute-water potential function simulation.

Figure IV.3.15 Calculated QCDF of solute-water pair energies of waters of the (CO)4 functional group of thymine - BVG solute-water potential function simulation.

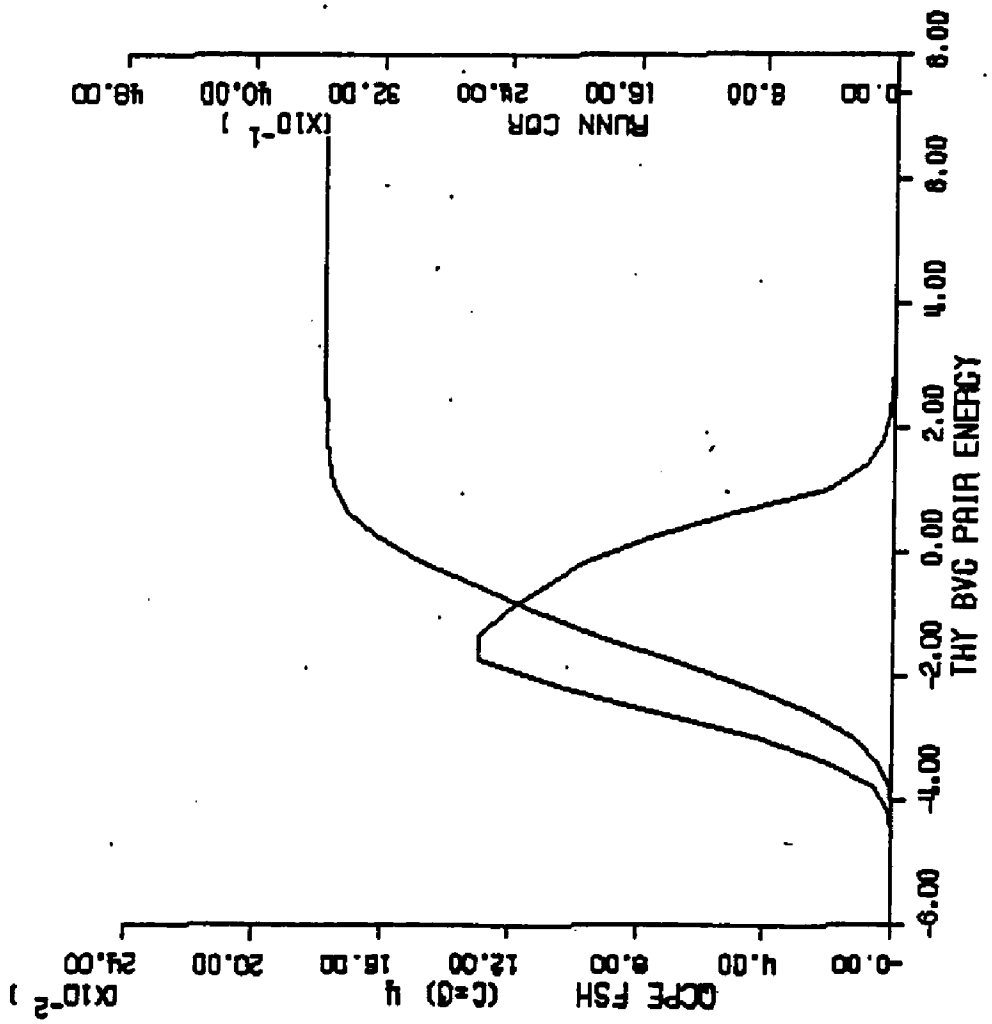


Figure IV.3.16 Calculated QCDF of solute-water pair energies of waters of the -CH3 functional group of thymine - bVG solute-water potential function simulation.

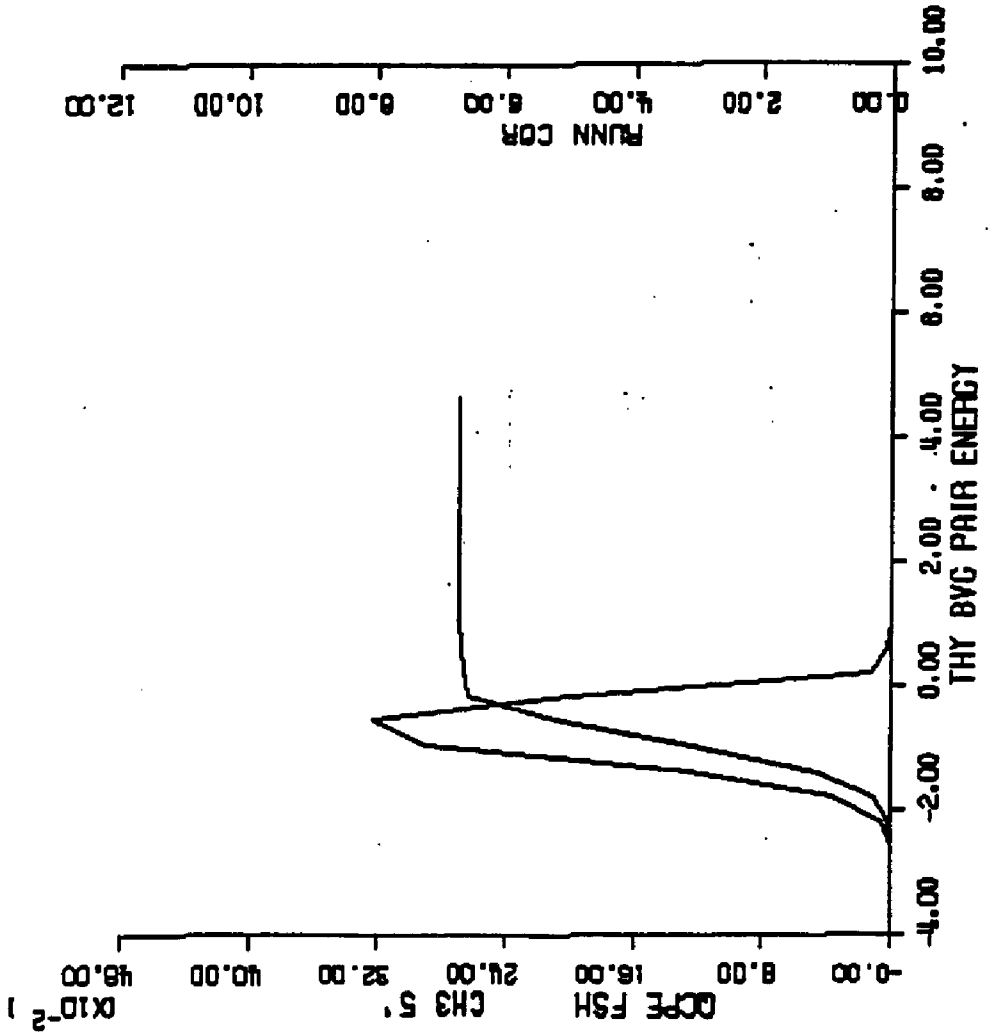


Figure IV.3.17 Calculated GCDF of solute-water pair energies of waters of the (CH)6 functional group of thymine - dV3 solute-water potential function simulation.

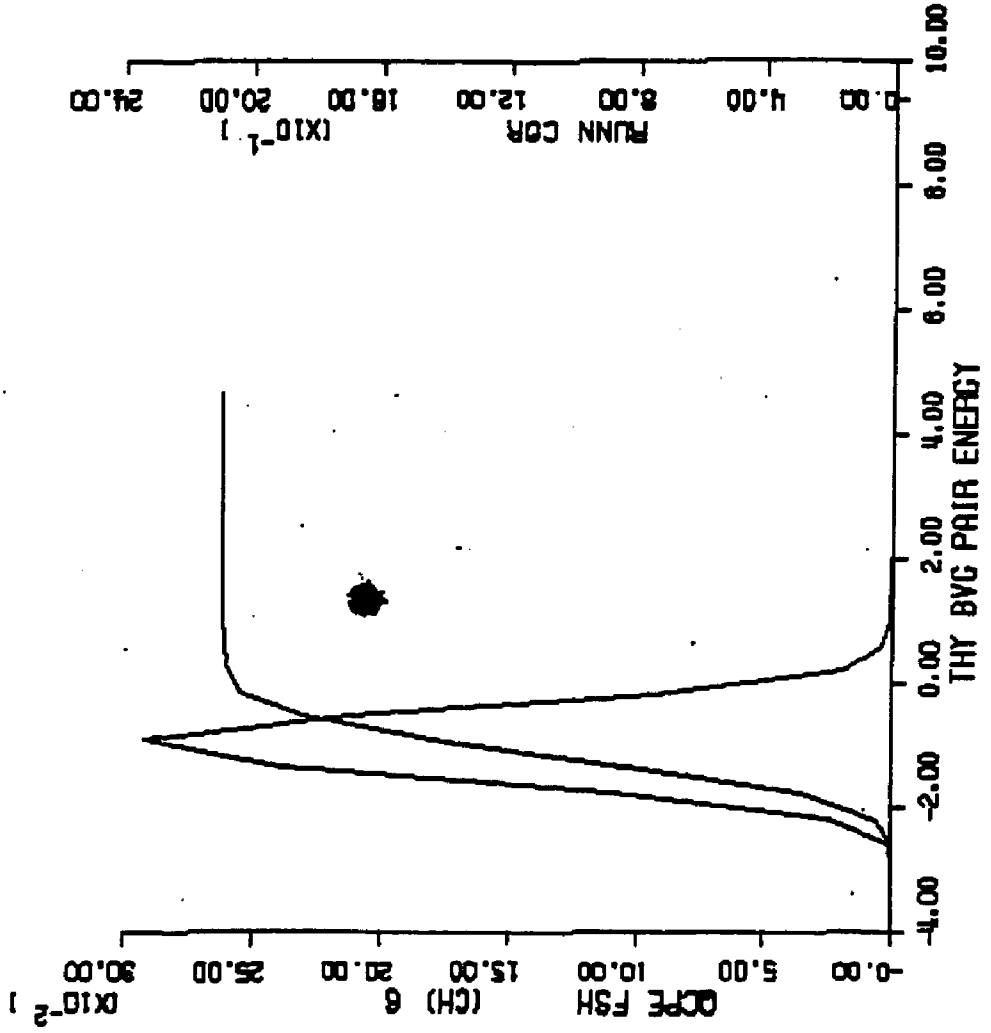
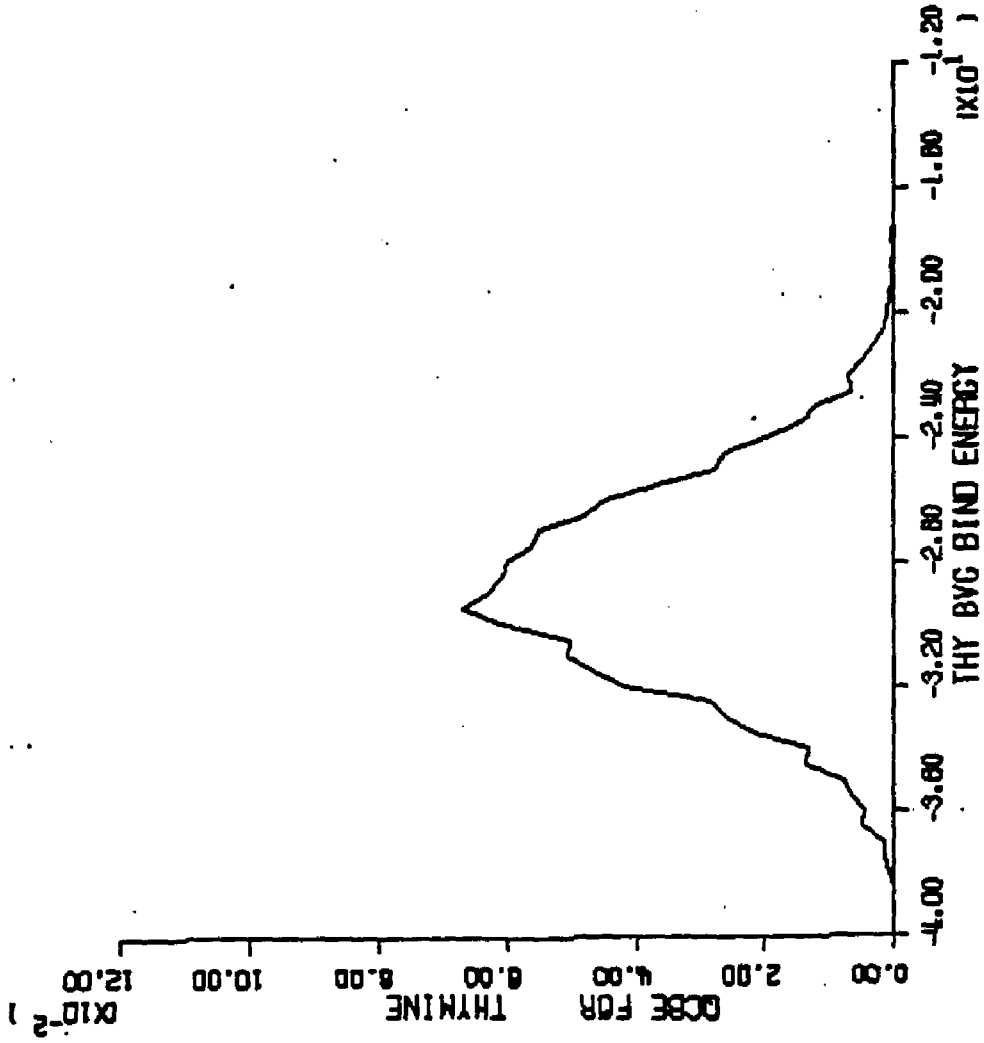


Figure IV.3.18 - calculated quasicomponent distribution function of binding energy for thymine (BVG potentials).

X axis - binding energy (kcal/mole).

Y axis - quasicomponent of binding energy.

Figure IV.3.18 Calculated QCDF of total binding energy for thymine - bVG solute-water potential function simulation.



4. Uracil with Clementi Solute-Water Potentials

Computational Specifics. An aqueous solution of uracil and 215 waters was simulated using Clementi solute-water potentials and MCY-CI water-water potentials at 298 K. Face centered cubic boundary conditions with a unit cell of 14.87 Angstroms edge were used, determined by a partial molar volume of 75 cc/mole for uracil and 18.07 cc/mole for water. An initial 600,000 steps of Monte Carlo simulation were carried out and were discarded as an equilibration period. The ensemble averages and other analyses were formed over the succeeding 2000K configurations of the realization.

Potential Surface. The potential energy surface for uracil and one water, using Clementi solute-water potentials for calculation, is shown in Figure IV.4.1. The peak between the H1 hydrogen and O2 oxygen is the lowest energy point on the surface with a minimum interaction of -13.40 kcal/mole. Energies become as high as -6.40 kcal/mole as the region near the H1 hydrogen is traversed towards the H6 methine hydrogen. The region of the H3 imino hydrogen has minimum solutewater interactions of -11.40 to -12.40 kcal/mole on the O2 side. A significant minimum also occurs on the O4 side of this hydrogen between -10.4 and -11.4 kcal/mole. The interactions in the regions immediately surrounding the carbonyl oxygens (coaxial to the C=O bonds) are less favorable, ranging from -6.4 to -7.4 kcal/mole for O4 and from -7.4 to -8.4 kcal/mole for O2. The region of the

methine hydrogens contain interactions as favorable (-6.4 to -8.4 kcal/mole) as the carbonyl oxygen regions.

Convergence and Thermodynamic Results. The control functions for the Clementi potential uracil simulation are shown in Figure IV.4.2. Calculated thermodynamic quantities are given in Table IV.1. The total binding energy remains within 10 kcal/mole for the final 1000K of the simulation attaining a final value of -2009.5 +/- 8.1 kcal/mole. The heat capacity reaches a final value of 18.1 cal/mole-K. The vacuum to water transfer energy is -149.7 kcal/mole.

Figures IV.4.4 thru IV.4.9 and IV.4.12 and IV.4.17 contain distribution functions for the quasicomponents of coordination number and average first shell pair energy for the imino, carbonyl and methine functional groups of uracil.

Structural Results. The atomic first shell coordination numbers, derived from the column labeled <K> in Table IV.5 are displayed on a molecular diagram in Figure IV.4.3. Both imino hydrogens are assigned the same first shell radii of 3.4 Angstroms. First shell population densities for these atoms are considerably higher than bulk water (1.32 for H1 and 1.42 for H3); accordingly, the first shell populations are high (H1: 3.16, H3: 3.43). The polar oxygens (O2 and O4) also have significant first shell populations of 2.31 and 3.07

waters respectively. The first shell densities, however, are significantly lower than bulk water at values of .89 for O2 and .90 for O4.

Although the methine hydrogens are considered apolar, the first shell population densities are both higher than bulk water. The H5 density is 1.05; the H6 density is considerably higher than bulk solvent at 1.27.

The QCDF of coordination number for uracil and Clementi solute-water potentials is displayed in Figure IV.4.10. The contributions fall in the range 18 to 25; the most common coordination number is 21. The average first shell coordination number is 19.76 +/- .31.

Energetic Results. First shell solute-water pair energies for the atoms of uracil are shown in Figure IV.4.11. The waters of the imino hydrogens have the most favorable pair energies of all first shell waters. The H1 hydrogen waters have the lowest solute binding energies (<SLIPE> = -5.524 kcal/mole) while the H3 waters contribute the lowest total solute binding energy to the molecule (<SLIBE> = -17.511 kcal/mole, <SLIPE> = -5.106 kcal/mole). Together the imino hydrogens contribute about 50% (-34.984 kcal/mole) of the total solute binding energy (-72.453 kcal/mole). The polar carbonyl oxygens are assigned waters with average pair energies (-3.623 kcal/mole for O2 and -2.458 kcal/mole for O4) approximately the same as those for the apolar methine

hydrogens (-2.493 kcal/mole for H5 and -3.473 kcal/mole). Waters assigned to the methine hydrogens have the most favorable water-water pair energies (-3.113 kcal/mole for H5 waters and -3.064 kcal/mole for H6 waters); all waters assigned to other ring substituents have average water-water pair energies -3.000 kcal/mole. This result is reasonable considering the apolar character of the methine hydrogens.

The QCDF of solute binding energies for uracil and Clementi potentials is displayed in Figure IV.4.18. Energy values ranging from around -153 to -123 kcal/mole are seen; the most significant contributions are from about -142 to -135 kcal/mole The average solute binding energy is -137.4 +/- 1.0 kcal/mole.

Figure IV.4.1 - isoenergy contour surface for uracil and one water (Clementi solute-water potential functions).

X axis - X axis of plane defined by molecular ring atoms (Angstroms).

Y axis - Y axis of plane defined by molecular ring atoms (Angstroms).

Contours - isoenergy contours with one kcal/mole increments with alphabetical labels referring to contour energy values in list at right.

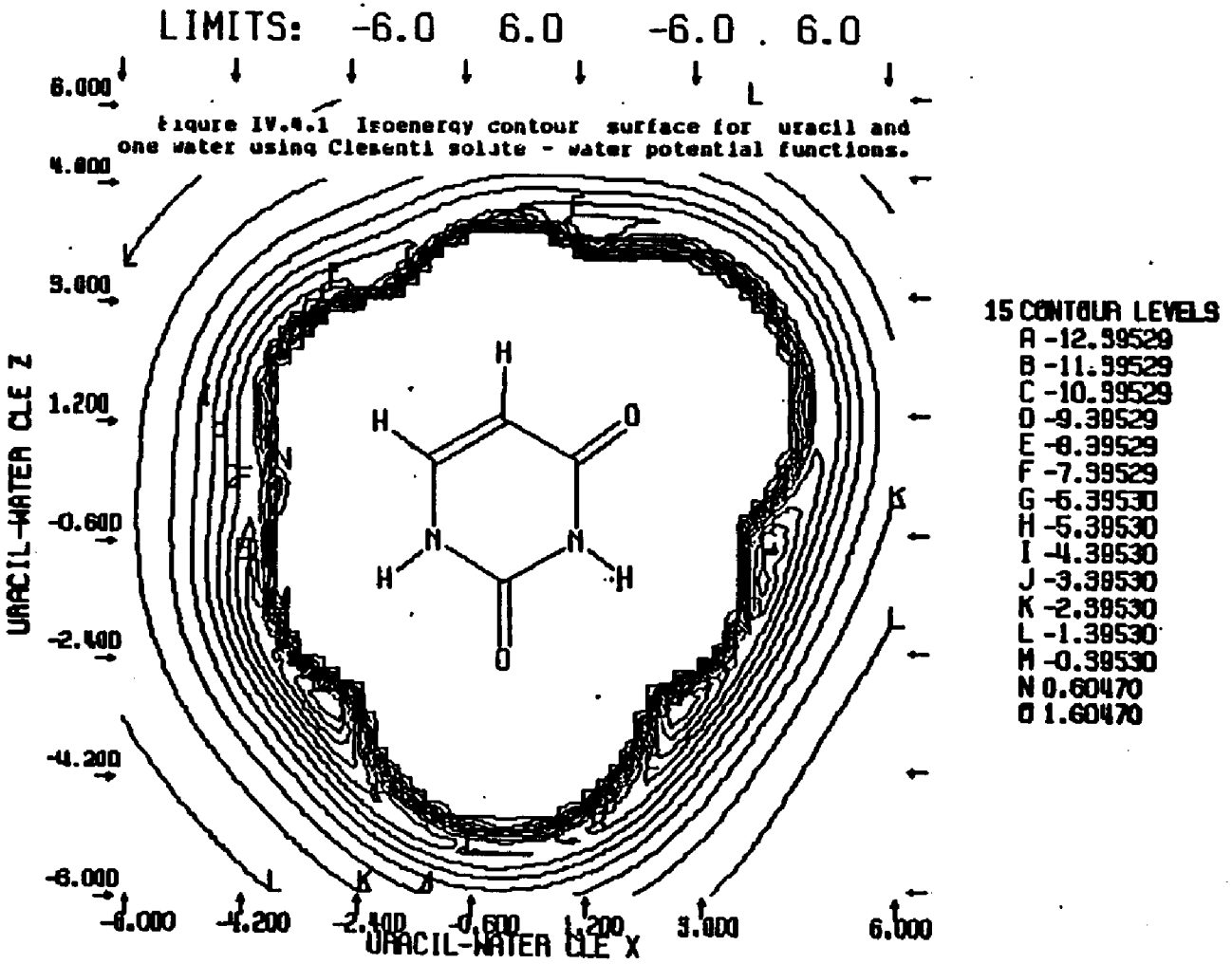


Figure IV.4.1 Isoenergy contour surface for uracil and one water using Clementi solute - water potential functions.

Figure IV.4.2 - control functions for Monte Carlo simulation of uracil and 215 waters (Clementi solute-water potential functions).

X axis - number of configurations.

Left Y axis - mean total energy (kcal/mole).

Right Y axis. - constant volume heat capacity (cal/mole-degree).

Upper curve - constant volume heat capacity.

Bottom curve without crosshatches - average total energy for entire simulation.

Bottom curve with crosshatches - average total energy for preceeding 50K configurations.

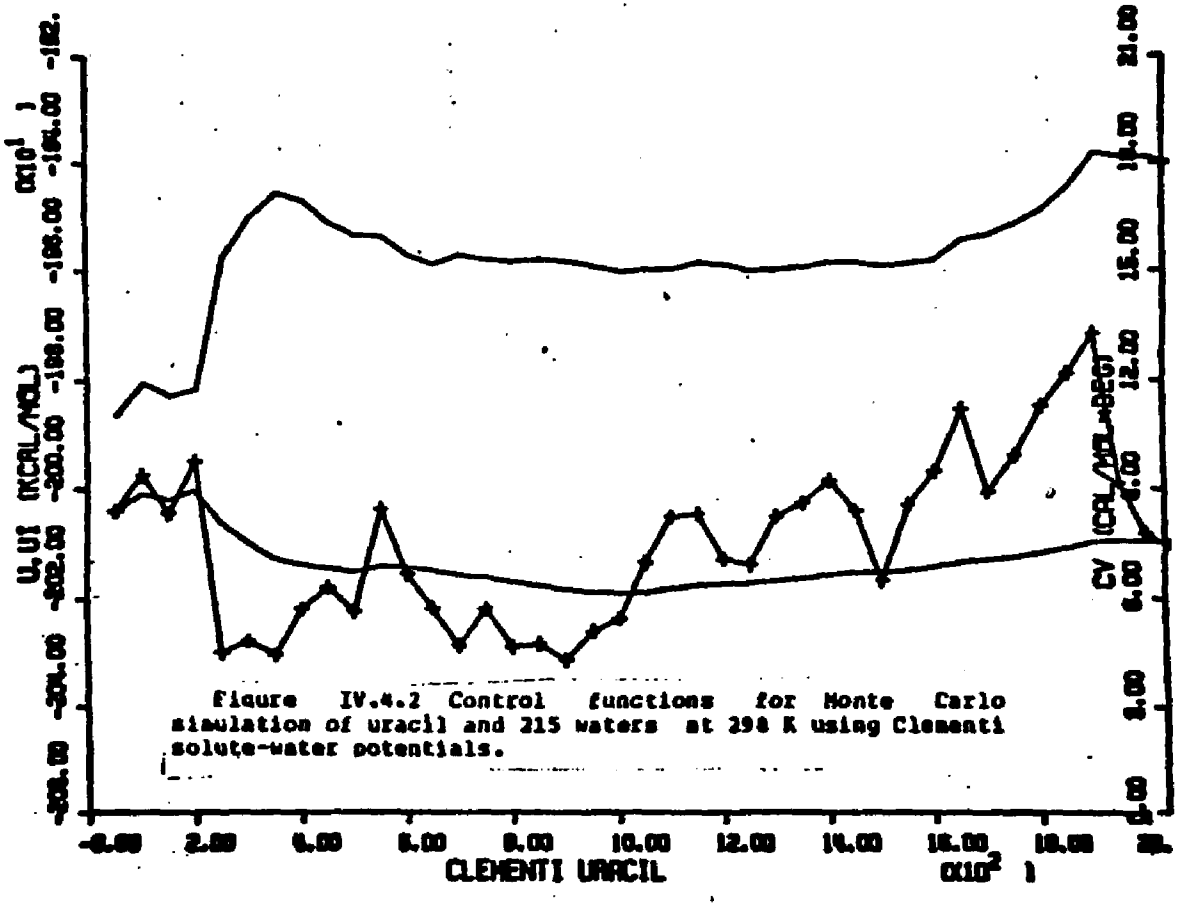


Figure IV.4.2 Control functions for Monte Carlo simulation of uracil and 215 waters at 298 K using Clementi solute-water potentials.

URACIL IN WATER AT 298K - FORCE BIAS AND CLEMENTI POTENTIAL FUNCTIONS

LAST CONFIGURATION: 2000001					FIRST SHELL SOLUTE PROPERTIES				TOTAL SLT PROPS		WATER PROPERTIES RFSH=3.30 RCS= 7.75 A		
AT NO	INDEX	TYPE		RFS	VFS	<K>	<K/V>	<SLTNE>	<SLTPE>	<K>	<SLTNE>	<ND>	<NNNPE>	<NNWT>
METHYLENE GROUPS														
1	5	25	C C5	4.2	17.88	0.24	0.41	-0.011	-0.045	1.107	-0.328	4.18	-2.976	-17.524
2	10	16	H H5	3.6	81.10	2.84	1.05	-7.868	-2.493	32.252	-16.862	4.14	-3.113	-17.463
TOTALS FOR FUNCTIONAL GROUP					98.98	3.08	.93	-7.879	-2.300	33.359	-16.389	4.14	-3.109	-17.465
3	6	25	C C6	3.6	14.86	0.24	0.48	-0.207	-0.066	1.241	-0.644	4.20	-2.994	-17.083
4	11	16	H H6	3.6	55.67	2.37	1.27	-0.232	-3.473	32.471	-16.990	4.17	-3.064	-17.568
TOTALS FOR FUNCTIONAL GROUP					70.53	2.61	1.11	-0.439	-3.235	33.712	-17.633	4.17	-3.061	-17.569
AVERAGES OVER FUNCTIONAL GROUPS:—					84.75	2.84	1.00	-7.759	-2.767	33.536	-17.011	4.16	-3.005	-17.517
STATISTICAL UNCERTAINTY (+/- 2*SD)						0.09	0.03	0.388	0.155	0.020	0.270	0.02	0.032	0.150
IMINO GROUPS														
5	6	15	N N3	3.8	15.58	0.91	1.75	-3.023	-3.323	1.003	-3.432	4.78	-3.177	-20.165
6	00	16	H H3	3.4	72.33	3.43	1.42	-17.511	-5.106	34.500	-28.670	4.27	-2.960	-17.404
TOTALS FOR FUNCTIONAL GROUP					87.91	4.34	1.40	-20.514	-4.732	36.303	-32.111	4.29	-2.970	-17.606
7	7	15	N N1	3.6	14.60	0.66	1.35	-1.655	-2.515	1.567	-2.136	4.12	-3.142	-17.009
8	12	16	H H1	3.4	71.62	3.16	1.32	-17.473	-5.524	31.996	-27.339	4.23	-2.950	-17.249
TOTALS FOR FUNCTIONAL GROUP					86.22	3.82	1.33	-19.127	-5.006	33.563	-29.475	4.23	-2.959	-17.275
AVERAGES OVER FUNCTIONAL GROUPS:—					87.06	4.00	1.40	-19.831	-4.869	34.973	-30.793	4.26	-2.969	-17.440
STATISTICAL UNCERTAINTY (+/- 2*SD)						0.13	0.05	0.770	0.267	0.021	0.493	0.02	0.030	0.146

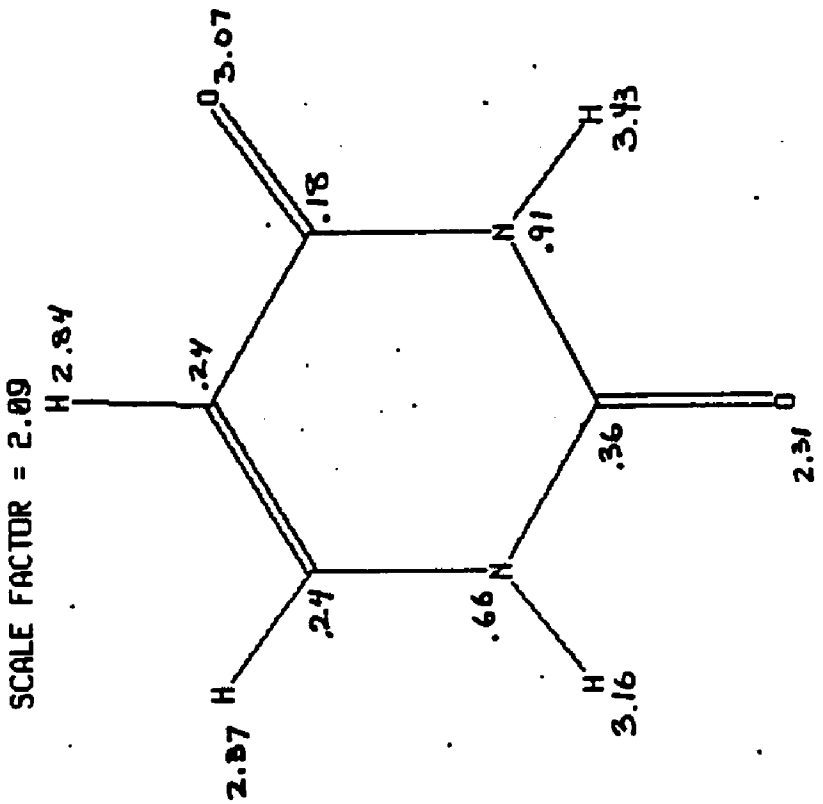
Table IV.5 Calculated structural and energetic quantities from Monte Carlo simulation of uracil and 215 waters at 298 K - Clementi solute-water potential functions.

URACIL IN WATER AT 298K - FORCE BIAS AND CLEMENTI POTENTIAL FUNCTIONS

CARBONYL GROUPS													
9	1	27	O O2	3.4	77.59	2.31	0.89	-0.383	-3.623	38.736	-21.896	4.23	-2.986	-17.493
10	2	26	C C2	4.0	18.77	0.36	0.57	-0.966	-2.693	1.256	-1.336	4.26	-3.128	-17.981
TOTALS FOR FUNCTIONAL GROUP					96.36	2.67	0.83	-9.349	-3.498	39.992	-22.432	4.23	-2.998	-17.588
11	4	26	C C4	3.4	15.41	0.18	0.34	-0.389	-2.196	1.337	-0.947	4.25	-2.982	-17.558
12	9	27	O O4	3.8	102.25	3.07	0.90	-7.536	-2.458	36.654	-18.390	4.26	-2.998	-17.588
TOTALS FOR FUNCTIONAL GROUP					117.66	3.24	0.82	-7.925	-2.443	37.991	-19.338	4.26	-2.998	-17.518
AVERAGES OVER FUNCTIONAL GROUPS:—					107.01	2.96	0.83	-0.637	-2.971	38.991	-20.885	4.25	-2.994	-17.589
STATISTICAL UNCERTAINTY (+/- 2*SD)						0.09	0.03	0.317	0.154	0.022	0.317	0.02	0.028	0.139
URACIL														
MOLECULAR SUM/AVERAGE:					557.66	19.76	1.06	-72.453	-3.666	215.000	-137.783	4.22	-3.814	-17.491
STATISTICAL UNCERTAINTY (+/- 2*SD)						0.31	0.02	1.465	0.090	0.033	1.851	0.01	0.015	0.060

Table IV.5 Calculated structural and energetic quantities from Monte Carlo simulation of uracil and 215 waters at 298 K - Clementi solute-water potential functions.

Figure IV.4.3 First shell coordination numbers for the atoms of uracil - Clementi solute-water potential function simulation.



Figures IV.4.4 through IV.4.10 - calculated quasicomponent distribution function of first shell coordination number for uracil and functional groups (Clementi potentials).

X axis - coordination number.

Y axis - quasicomponent of coordination number.

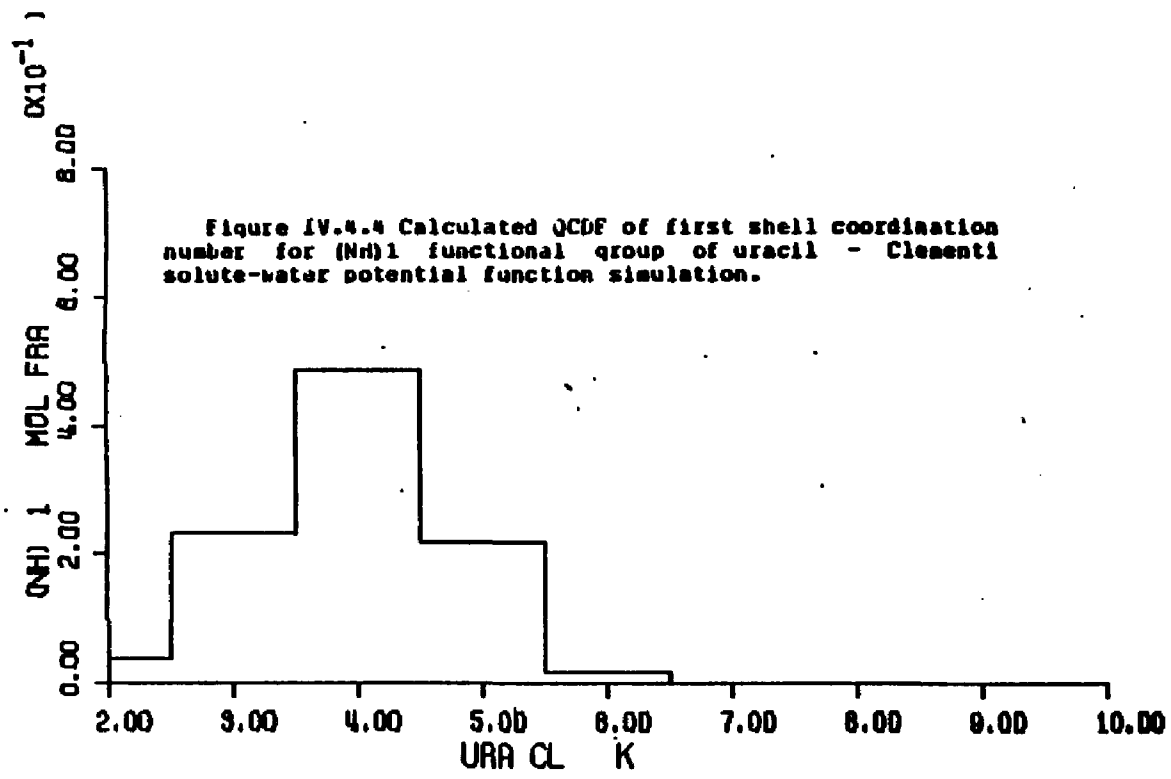


Figure IV.4.4 Calculated QCDF of first shell coordination number for (NH)1 functional group of uracil - Clementi solute-water potential function simulation.

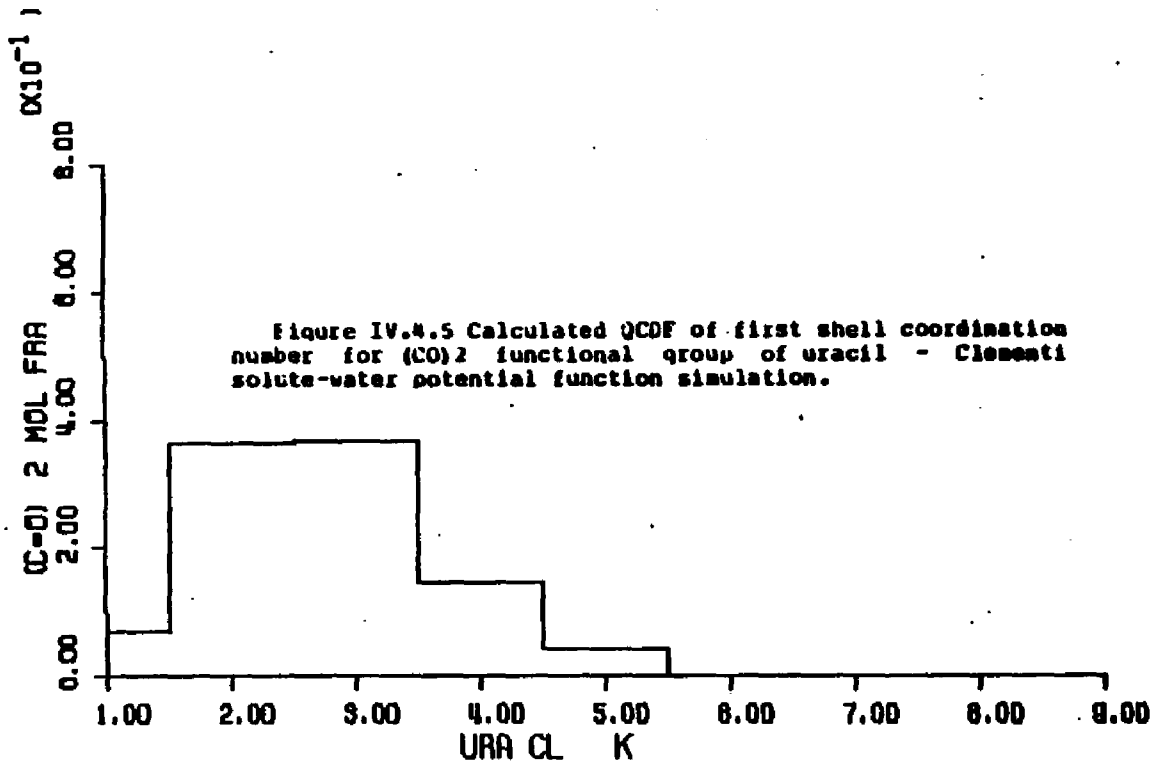


Figure IV.4.5 Calculated QCDF of first shell coordination number for (CO)2 functional group of uracil - Clementi solute-water potential function simulation.

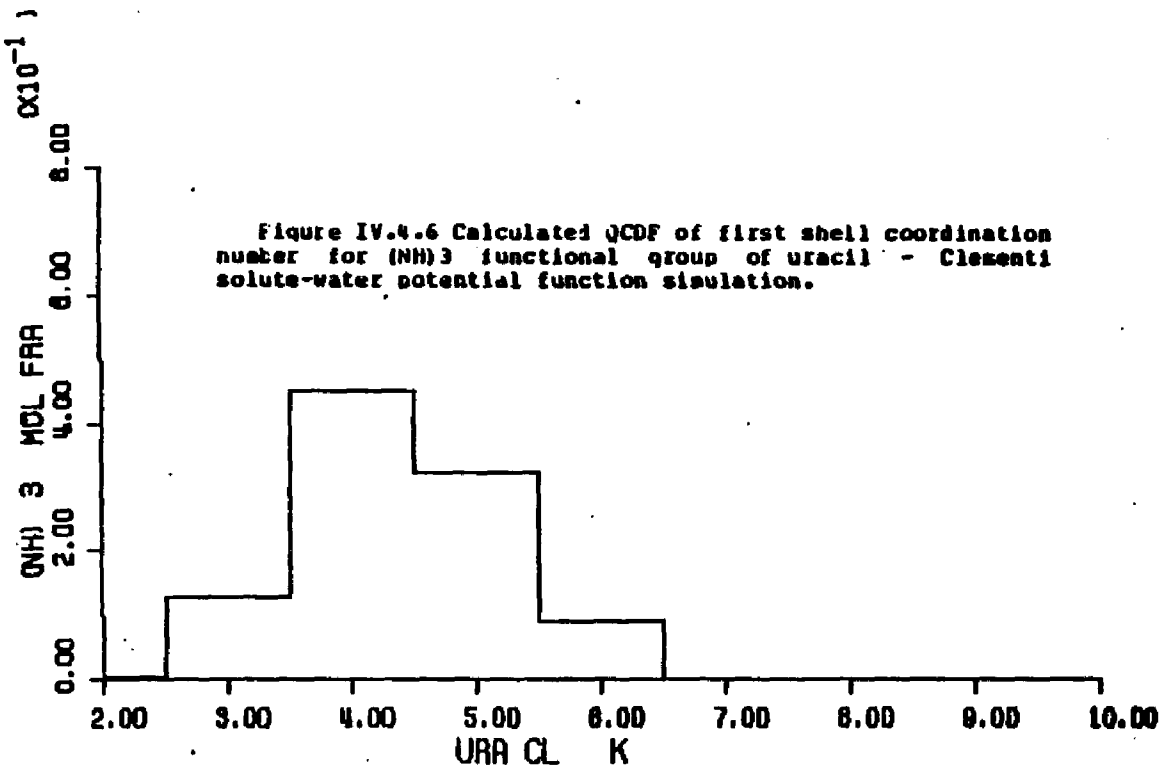


Figure IV.4.6 Calculated QCDF of first shell coordination number for (NH)3 functional group of uracil - Clementi solute-water potential function simulation.

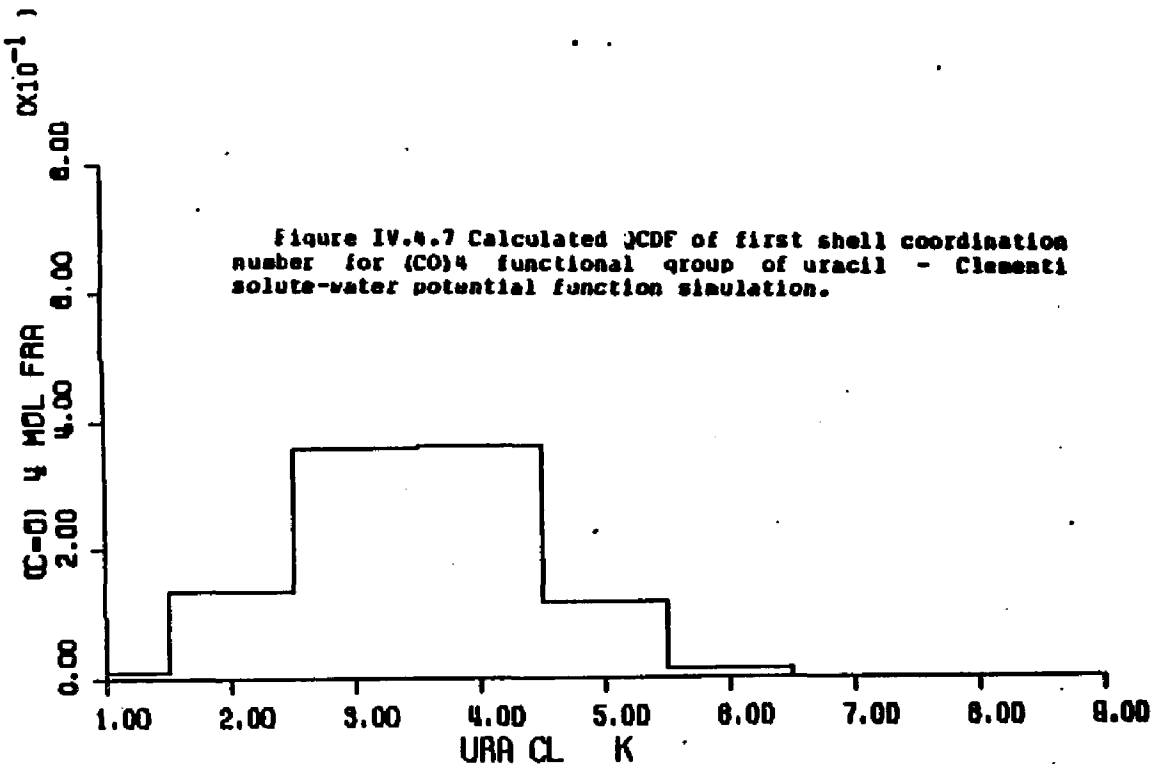


Figure IV.4.7 Calculated CDF of first shell coordination number for (CO)4 functional group of uracil - Clementi solute-water potential function simulation.

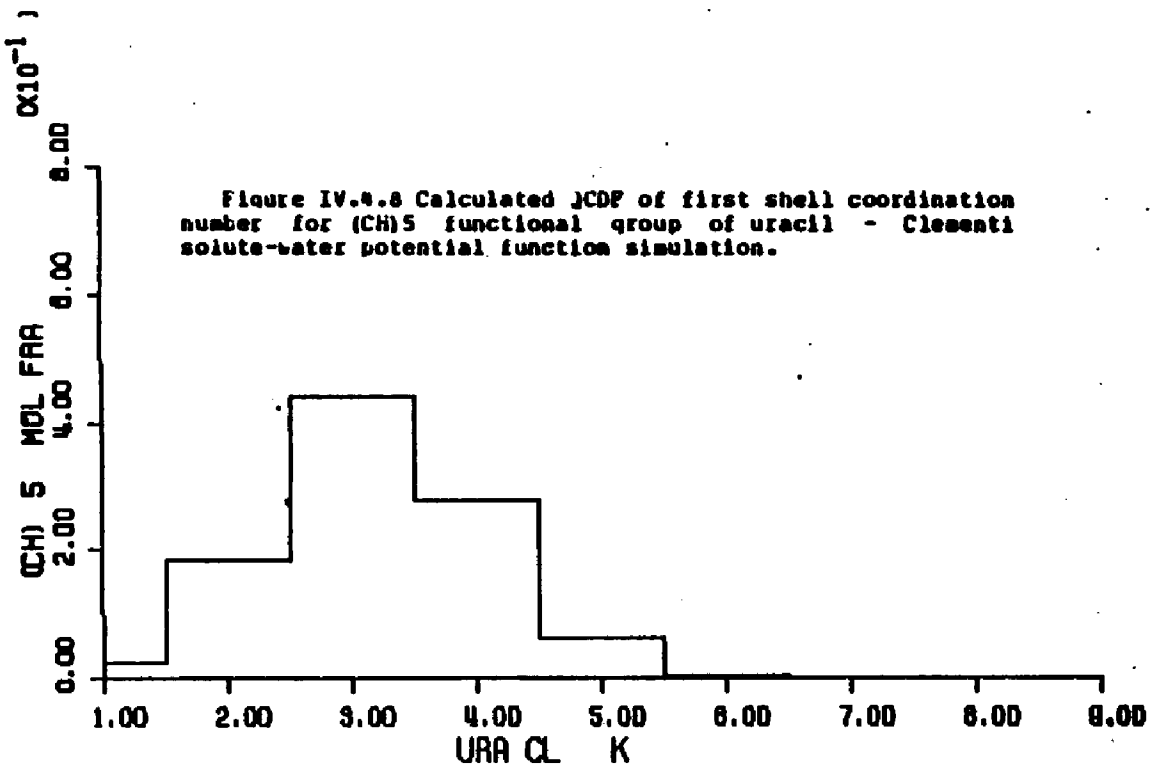


Figure IV.4.8 Calculated JCDF of first shell coordination number for (CH)5 functional qroup of uracil - Clementi solute-water potential function simulation.

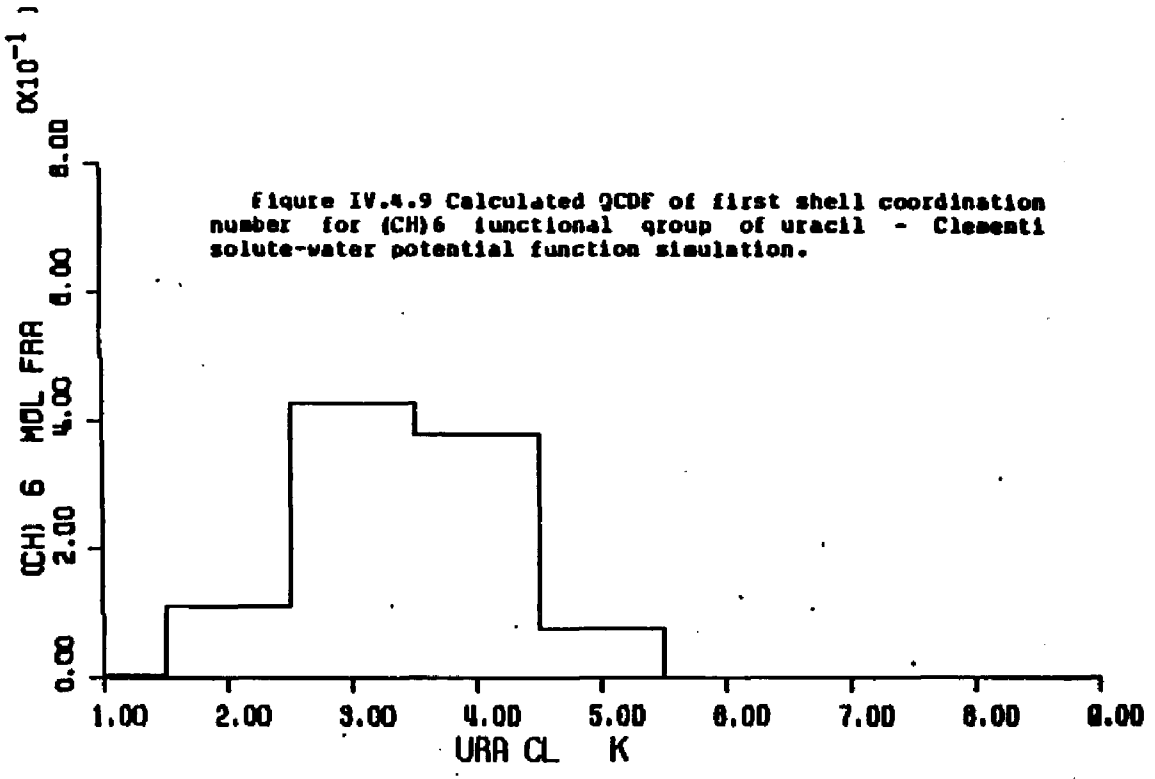


Figure IV.4.9 Calculated QCDF of first shell coordination number for (CH)6 functional group of uracil - Clementi solute-water potential function simulation.

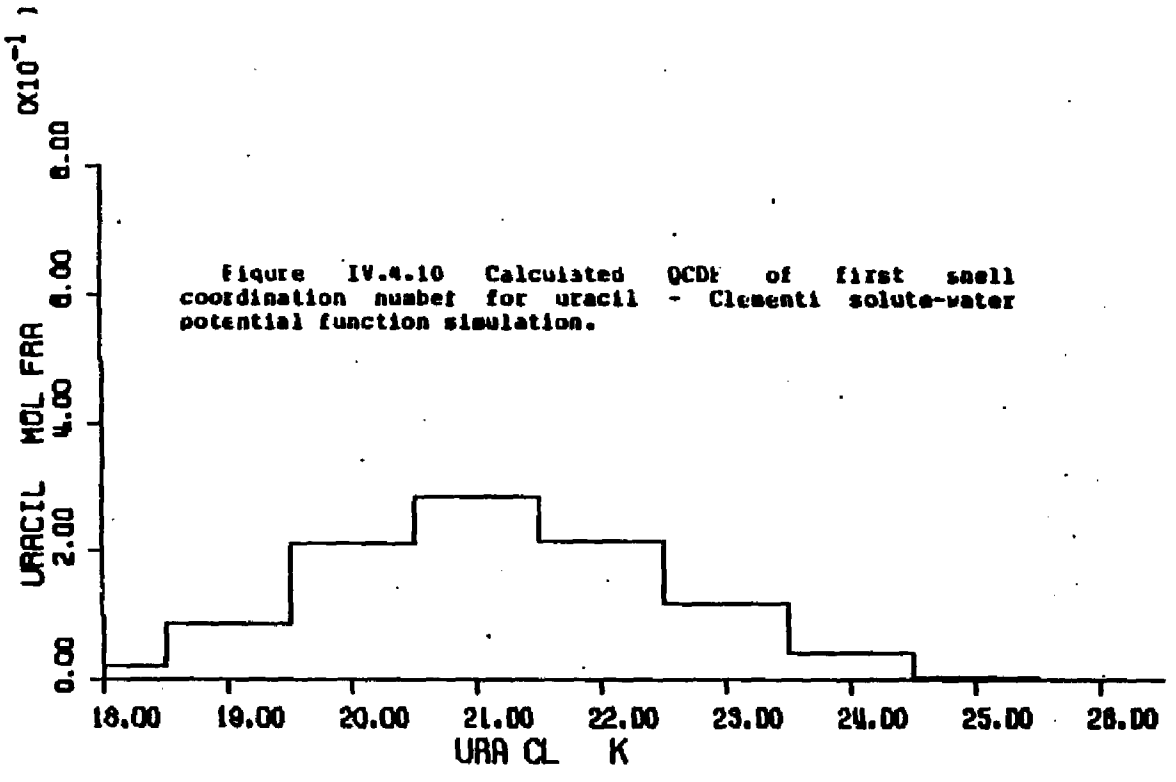


Figure IV.4.10 Calculated QCDF of first shell coordination number for uracil - Clementi solute-water potential function simulation.

Figure IV.4.11 Average first shell solute-water pair energies of waters assigned to the atoms of uracil - Clementi solute-water potential function simulation.

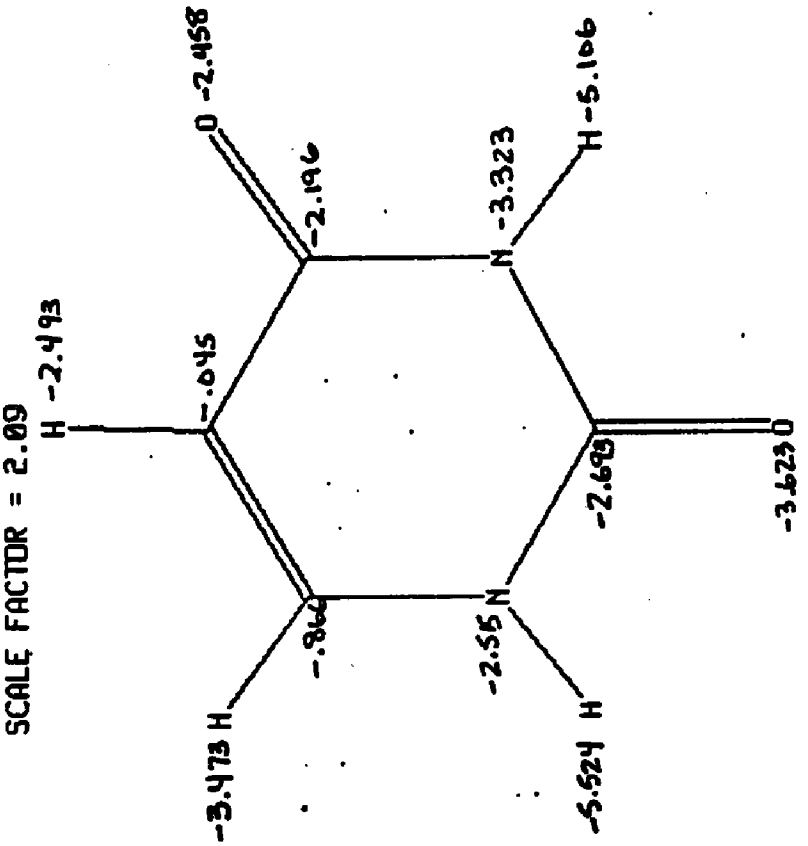


Figure IV.4.12 through IV.4.17 - calculated quasicomponent distribution function of average first shell pair energy for functional groups of uracil (Clementi potentials).

X axis - pair energy (kcal/mole).

Left Y axis - quasicomponent of pair energy.

Right Y axis - running coordination number.

Figure IV.4.12 Calculated QCDF of solute-water pair energies of waters of the (NH)1 functional group of uracil - Clementi solute-water potential function simulation.

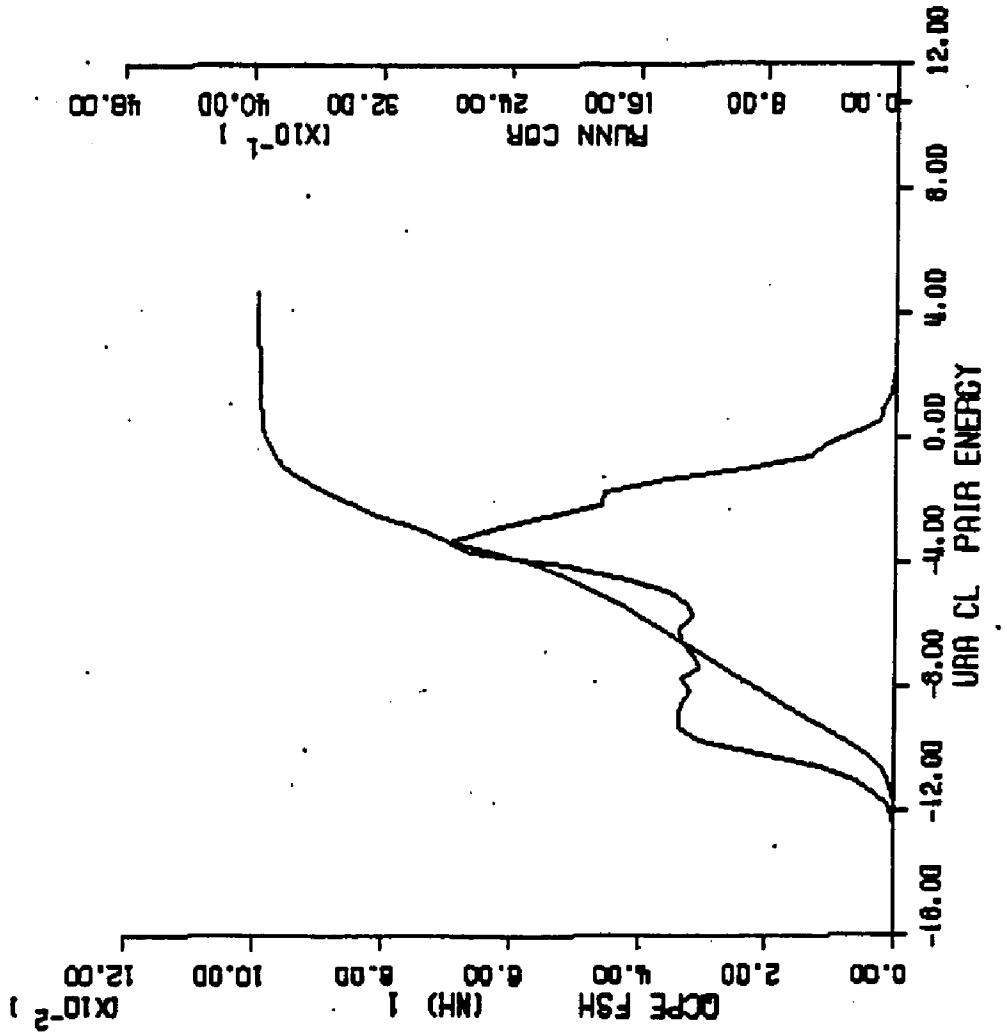


Figure IV.4.13 Calculated QCDF of solute-water pair energies of waters of the (CO)2 functional group of uracil - Clementi solute-water potential function simulation.

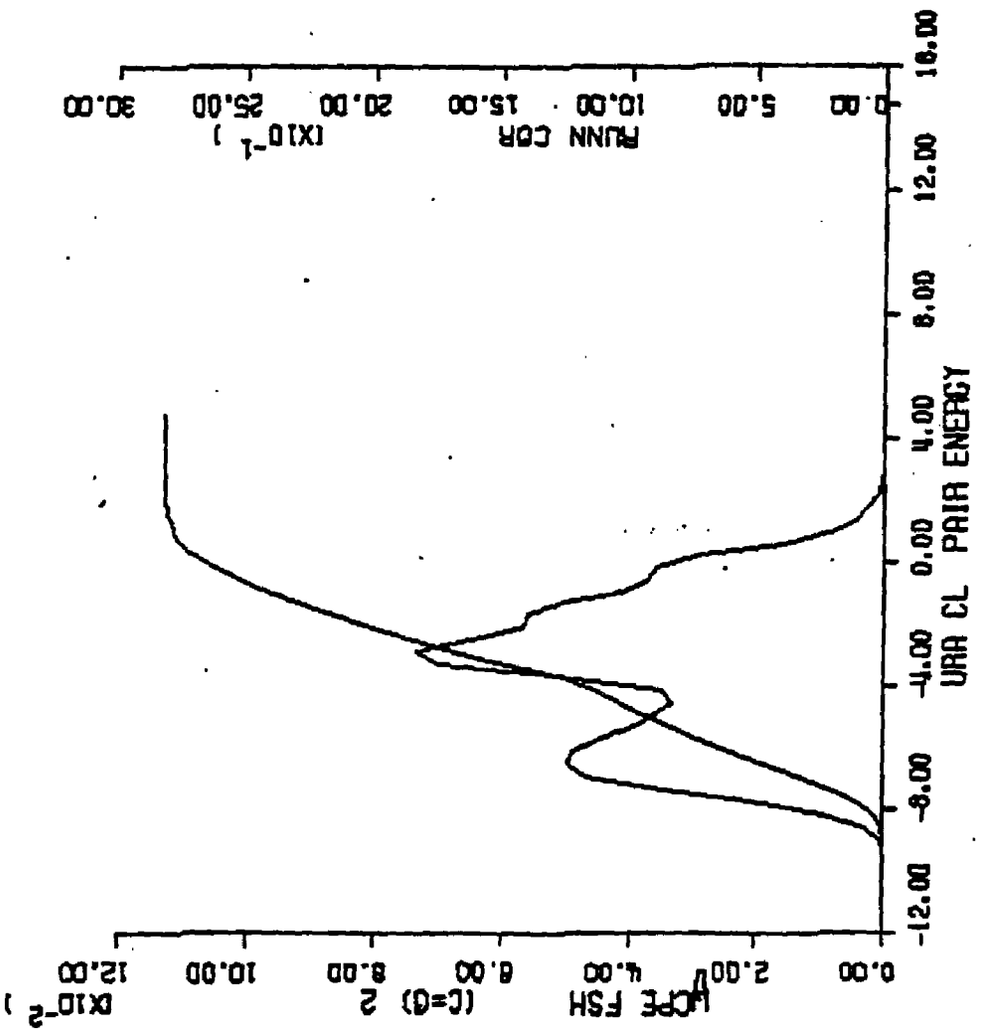


Figure IV.4.14 Calculated qCDF of solute-water pair energies of waters of the (NH)3 functional group of uracil - Clementi solute-water potential function simulation.

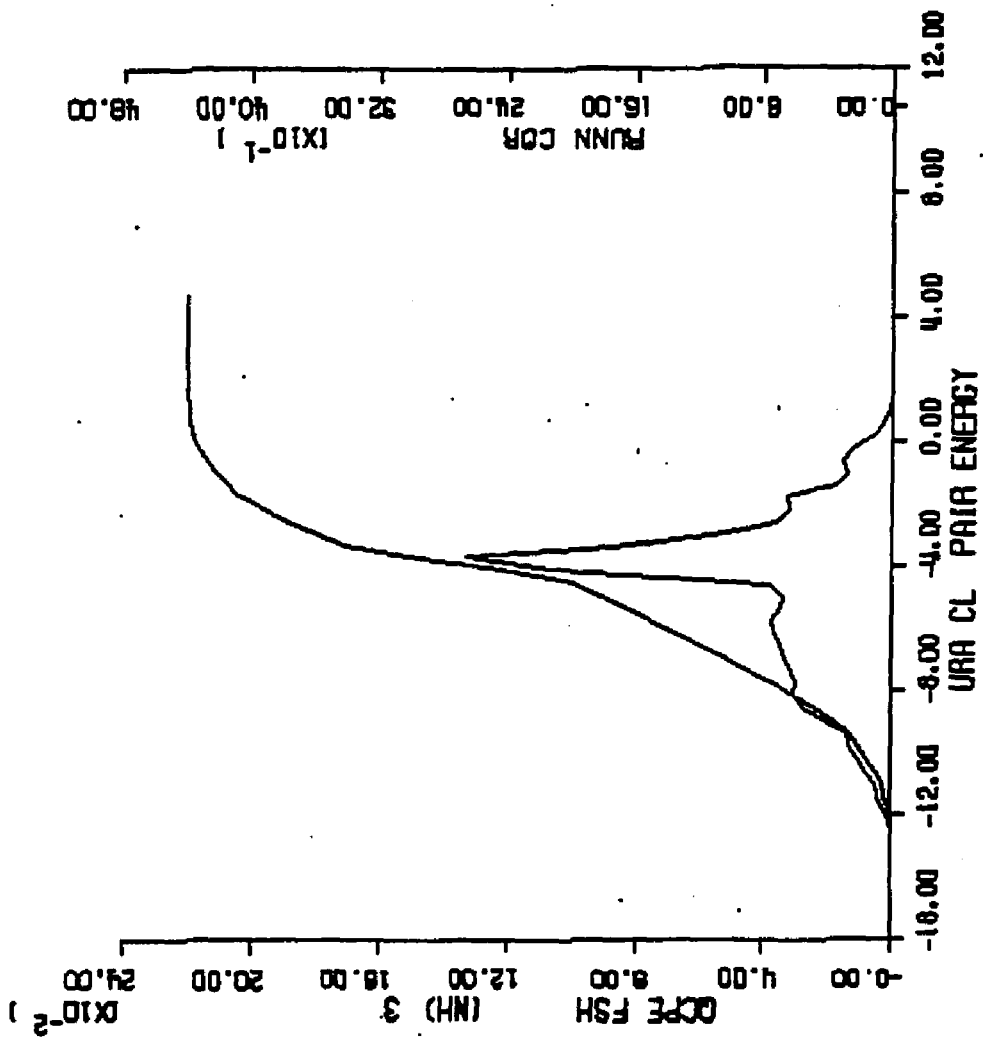


Figure IV.4.15 Calculated QCDF of solute-water pair energies of waters of the (CO)4 functional group of uracil - Clementi solute-water potential function simulation.

Figure IV.4.16 Calculated QCDF of solute-water pair energies of waters of the $(CH_3)_5$ functional group of uracil - Clementi solute-water potential function simulation.

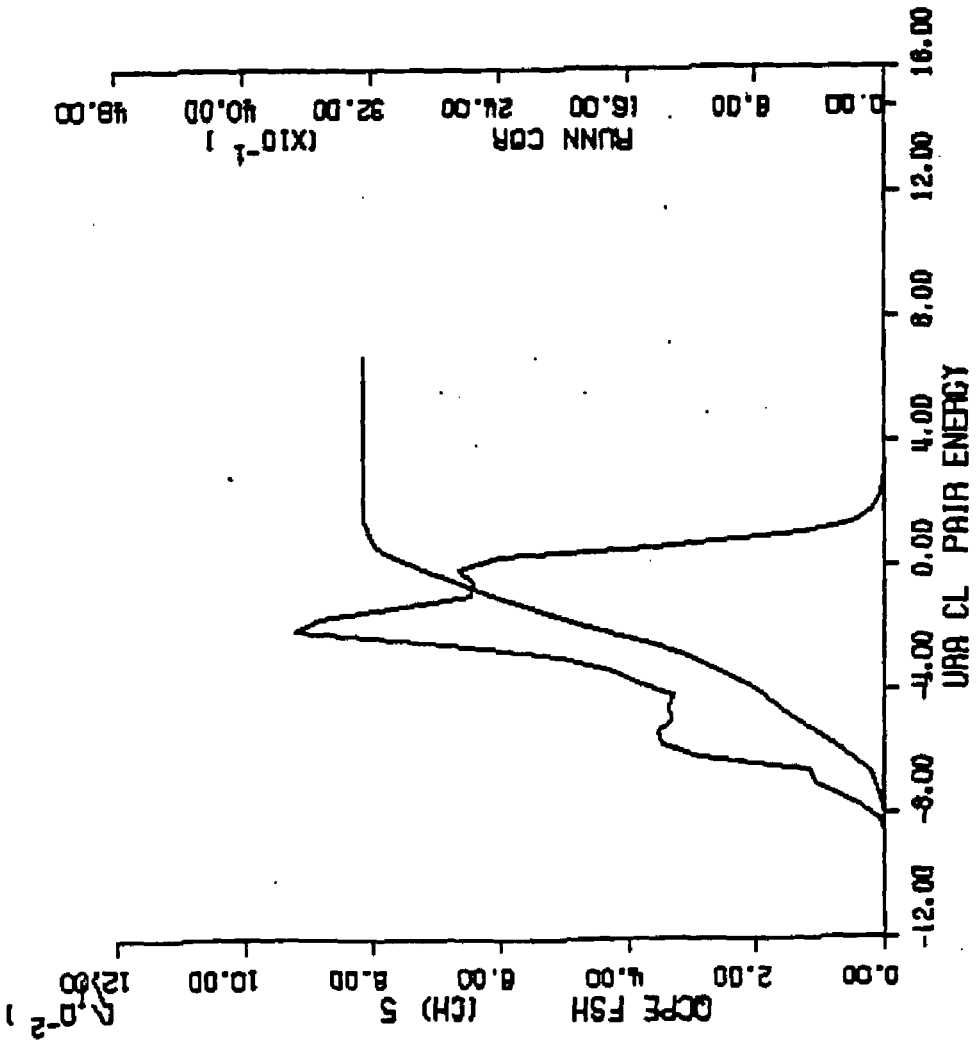


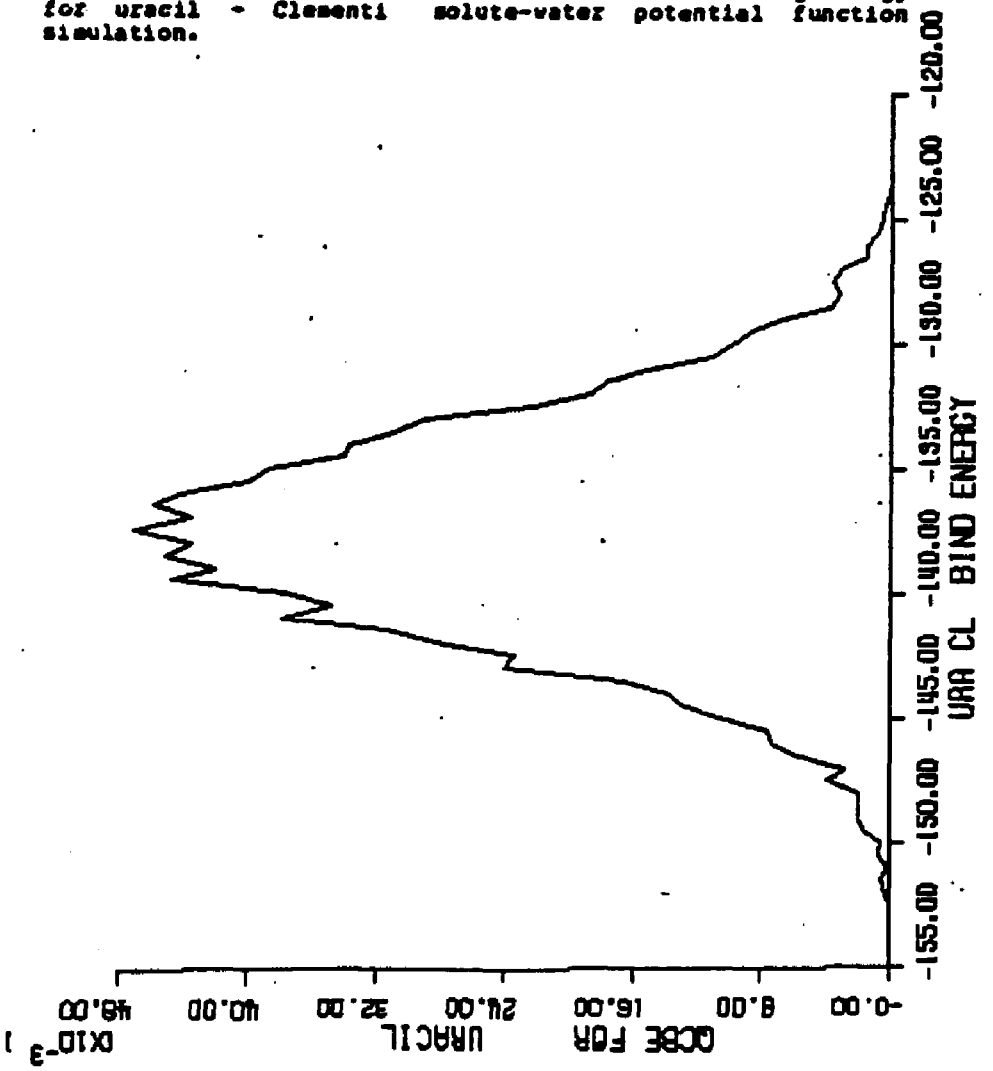
Figure IV-4.17 Calculated QCSF of solute-water pair energies of the (H$_2$O) functional group of uracil - Clementi solute-water potential function simulation.

Figure IV.4.18 - calculated quasicomponent distribution function of binding energy for uracil (Clementi potentials).

X axis - binding energy (kcal/mole).

Y axis - quasicomponent of binding energy.

Figure IV.4.18 Calculated QCDF of total binding energy for uracil - Clementi solute-water potential function simulation.



5. Uracil with Kollman Solute-Water Potentials

Computational Specifics. A solution of uracil and 215 waters was simulated using Kollman solute-water potentials and TIP3P water-water potentials at 298 K. Face centered cubic boundary conditions were used with a unit cell edge of 14.87 Angstroms. The dimensions were determined using a partial molar volume of 75 cc/mole for uracil and 18.07 cc/mole for water. The initial 600,000 steps of Monte Carlo simulation were discarded as an equilibration period. Ensemble averages and other analyses were calculated using the succeeding 2000K configurations of the simulation.

Potential Surface. The potential energy surface for the interaction of uracil and one water molecule using Kollman solute-water potential functions is shown in Figure IV.5.1 The lowest energy solute-water interaction is found in the area between the O2 carbonyl oxygen and the H1 imino hydrogen atoms at an energy of -6.98 kcal/mole. The region of the H3 hydrogen similarly contains quite favorable solute-water interaction energies between -4.98 and -5.98 kcal/mole. The regions surrounding both the carbonyl O2 and the carbonyl O4 atoms contain interaction energies between -2.98 and -3.98 kcal/mole. Finally, the apolar methylene region minimum solute-water energies fall between -.98 and -2.98 kcal/mole.

Convergence and Thermodynamic Results. Control

functions for this simulation are found in Figure IV.5.2. Examination of the mean energy curve reveals that the total binding energy remains within about five kcal/mole for the last 1000K of the simulation reaching a final value of -2131.1 +/- 5.6 kcal/mole. The vacuum to water solute transfer energy is -11.2 kcal/mole. The heat capacity reaches a final value of 18.4 cal/mole-K.

Figures IV.5.4 thru IV.5.9 and IV.5.12 thru IV.5.17 contain distribution function plots for the quasicomponents of coordination number and average first shell pair energies for the imino, carbonyl and methine functional groups.

Structural Results. The first shell coordination numbers for uracil in this simulation are listed in the column labeled <K> in Table IV.6 and are displayed on the molecular diagram in Figure IV.5.3. The ring nitrogens of uracil are unhydrated due to their low solvent accessibility. The imino hydrogens, H1 and H3, are fully hydrated. The H1 hydrogen has the highest coordination number in the molecule and contains 3.32 waters in a first shell of radius 3.8 Angstroms. The first shell water density is .98 of bulk water; although this density is lower than bulk water, it is the highest first shell water density found in the molecule. The H3 hydrogen contains about one less (2.57) water in the first coordination shell. The low water density of .93 is

nonetheless the second highest in the molecular coordination shell. The carbonyl oxygens (O2 and O4) also turn out to be significantly hydrated by 2.22 waters (O2) and 3.09 waters (O4).

The united atoms (C5H5 and C6H6) have the extended first shell characteristic of apolar atoms (both radial cutoffs are 4.6 Angstroms). On the average, these ring fragments have 3.08 (C5H5) and 3.29 (C6H6) waters in their coordination shells. The coordination shell water densities for C5H5 and C6H6 are .91 and .92 respectively, demonstrating that, in this simulation, there is only a small difference between the first shell structural hydration character of polar and apolar ring substituents.

The QCDF of coordination number for uracil in this simulation is displayed in Figure IV.5.10. Contributions range from 14 to 24; the most common values are 19 and 20 The average first shell coordination number is 18.76 +/- .70.

Energetic Results. The first shell average pair energies collected for the waters of the first hydration shell of each atom of uracil are displayed on the molecular diagram of Figure IV.5.11. The values in this table are taken from the column labeled <SLTPE> in Table IV.6. The H3 imino hydrogen is assigned relatively strongly bound waters at average pair energies of -1.299

kcal/mole resulting in a total solute binding energy contribution of -3.341 kcal/mole. The waters assigned to the H1 hydrogen are bound less strongly at an average of -.664 kcal/mole. The apolar methine united atoms posess the waters with the most positive average pair energies (-.394 kcal/mole for C5H5 and -.515 kcal/mole for C6H6).

The quasicomponent distribution function of solute binding energy for uracil is shown in Figure IV.5.18. The values range from -30 to -10 kcal/mole with the most significant contributions occurring in the -23 to -17 kcal/mole range. The average solute binding energy is -20.9 +/- 1.0 kcal/mole. The average water-water pair energies in the C6H6 region are the lowest of any region of the molecule at -3.204 kcal/mole.

Figure IV.5.1 - isoenergy contour surface for uracil and one water (Kollman solute-water potential functions).

X axis - X axis of plane defined by molecular ring atoms (Angstroms).

Y axis - Y axis of plane defined by molecular ring atoms (Angstroms).

Contours - isoenergy contours with one kcal/mole increments with alphabetical labels referring to contour energy values in list at right.

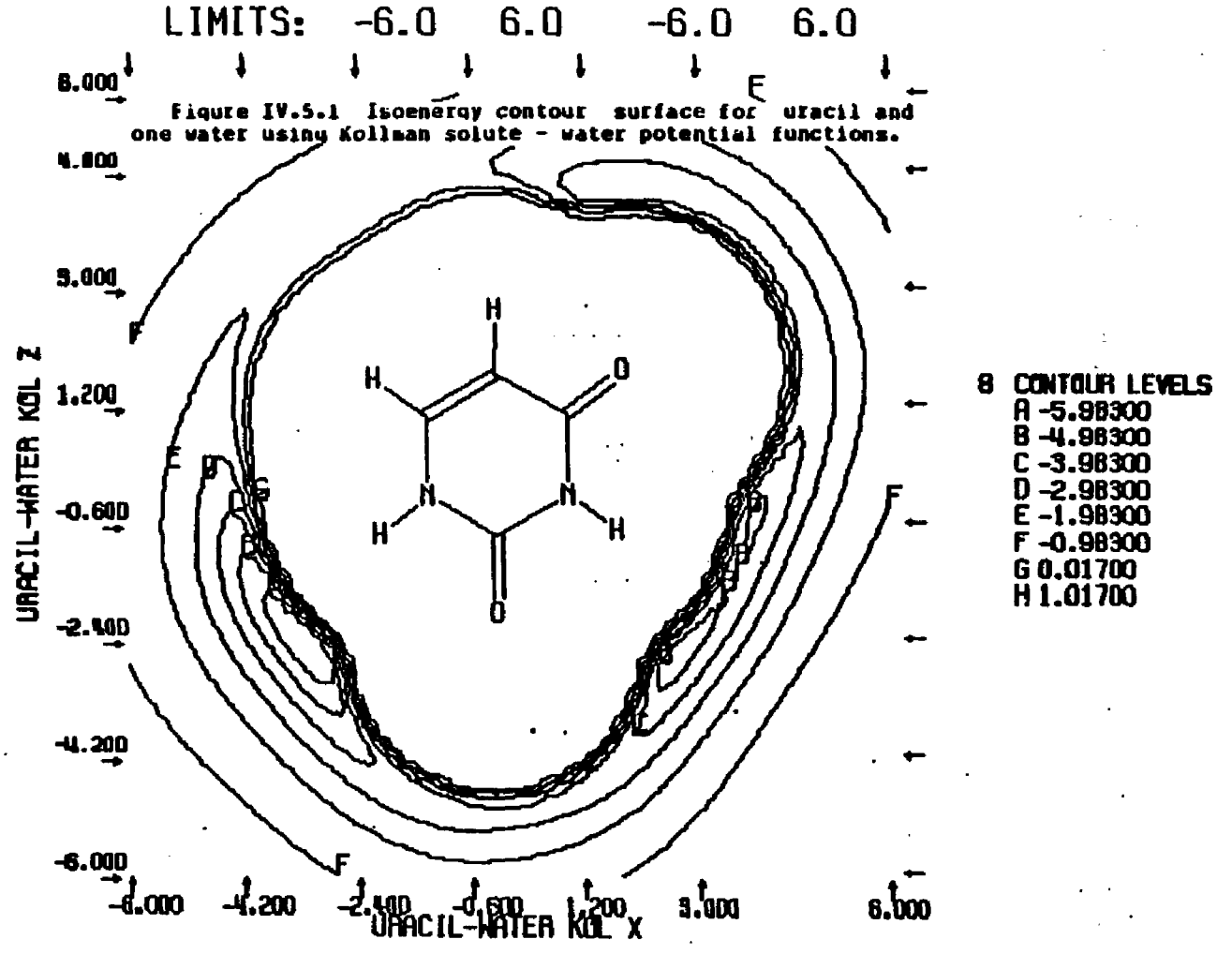


Figure IV.5.1 Isoenergy contour surface for uracil and one water using Kollman solute - water potential functions.

Figure IV.5.2 - control functions for Monte Carlo simulation of uracil and 215 waters (Kollman solute-water potential functions).

X axis - number of configurations.

Left Y axis - mean total energy (Kcal/mole).

Right Y axis. - constant volume heat capacity (cal/mole-degree).

Upper curve - constant volume heat capacity.

Bottom curve without crosshatches - average total energy for entire simulation.

Bottom curve with crosshatches - average total energy for preceeding 50K configurations.

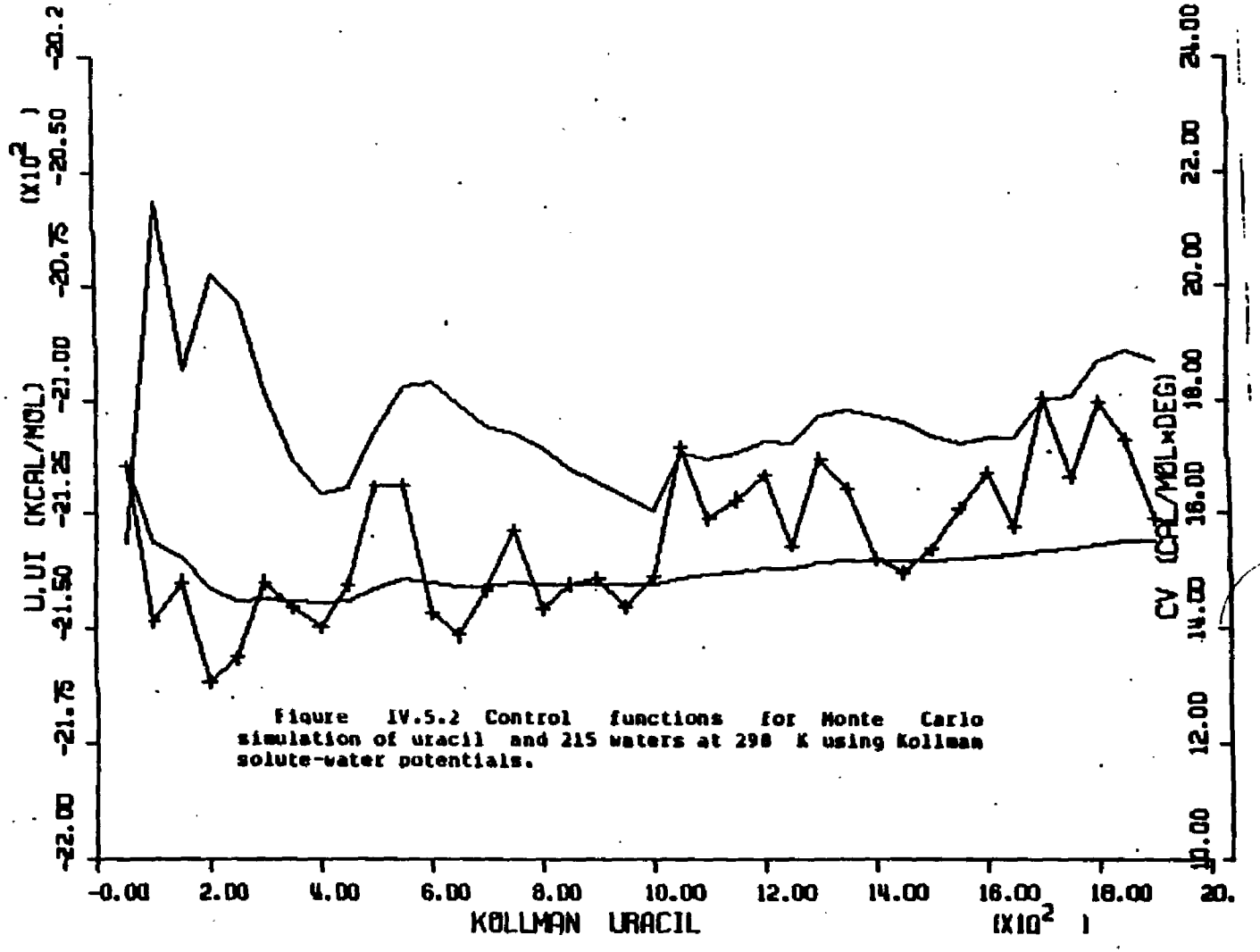


Figure IV.5.2 Control functions for Monte Carlo simulation of uracil and 215 waters at 298 K using Kollman solute-water potentials.

URACIL IN WATER AT 298K - FORCE BIAS AND KOLLMAN POTENTIAL FUNCTIONS

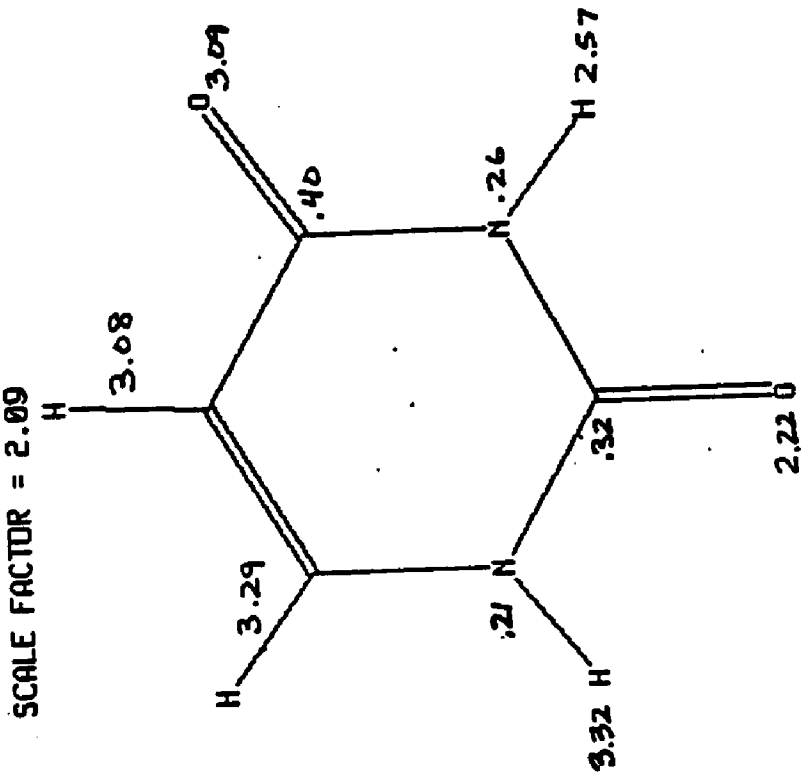
LAST CONFIGURATION: 2000001					FIRST SHELL SOLUTE PROPERTIES				TOTAL SLT PROPS		WATER PROPERTIES RPS=3.38 RCS= 7.75 A		
AT NO	INDEX	TYPE		NPS	VPS	<E>	<E/V>	<ELTSE>	<ELTPE>	<E>	<ELTSE>	<RO>	<ORNAPE>	<ORNWT>
METHYLENE GROUPS														
1	89	18	C5H5	4.6	181.15	3.88	0.91	-1.214	-0.394	23.616	-1.683	4.42	-3.188	-19.761
2	18	18	C6H6	4.6	187.34	3.29	0.92	-81.694	-0.515	26.188	-1.961	4.44	-3.284	-19.896
AVERAGES OVER FUNCTIONAL GROUPS:—					184.25	3.18	0.91	-81.454	-0.455	24.862	-1.822	4.43	-3.196	-19.829
STATISTICAL UNCERTAINTY (+/- 2*SD)						0.25	0.07	0.186	0.057	0.017	0.182	0.02	0.028	0.141
IMINO GROUPS														
3	1	38	N N1	3.6	14.64	0.21	0.43	-0.188	-0.086	1.242	-0.324	4.24	-3.274	-19.358
4	82	19	H H1	3.8	101.72	3.32	0.98	-82.289	-0.664	39.488	-2.735	4.48	-3.196	-19.762
TOTALS FOR FUNCTIONAL GROUP					116.36	3.54	0.91	-82.397	-0.678	40.738	-3.059	4.39	-3.199	-19.749
85	5	38	N N3	3.8	15.57	0.26	0.51	-0.287	-0.783	1.254	-0.316	4.24	-3.298	-19.318
86	86	19	H H3	3.6	82.77	2.57	0.93	-83.341	-1.299	35.711	-4.521	4.42	-3.154	-19.538
TOTALS FOR FUNCTIONAL GROUP					898.34	2.84	0.86	-83.548	-1.251	36.964	-4.837	4.41	-3.159	-19.538
AVERAGES OVER FUNCTIONAL GROUPS:—					187.35	3.19	0.89	-82.973	-0.964	38.847	-3.948	4.48	-3.179	-19.648
STATISTICAL UNCERTAINTY (+/- 2*SD)						0.28	0.05	0.381	0.121	0.022	0.315	0.02	0.023	0.112

Table IV.6 Calculated structural and energetic quantities from Monte Carlo simulation of uracil and 215 waters at 298 K - Kollman solute-water potential functions.

Table IV.6 Calculated structural and energetic quantities from Monte Carlo simulation of uracil and 215 waters at 296 K - Kollman solute-water potential functions.

CARBONYL GROUPS													
87	3	81	C	C2	4.2	19.75	8.32	8.48	-8.295	-8.925	81.373	-8.378	4.28	-3.217	-19.167
88	4	34	0	82	3.4	77.67	2.22	8.86	-2.822	-8.918	37.888	-3.178	4.45	-3.188	-19.557
TOTALS FOR FUNCTIONAL GROUP						897.42	2.54	8.78	-2.316	-8.912	38.381	-3.548	4.45	-3.188	-19.557
89	7	81	C	C4	4.2	19.21	8.48	8.62	-8.359	-8.987	81.518	-8.444	4.31	-3.266	-19.362
18	8	34	0	04	3.6	899.86	3.89	8.92	-2.731	-8.884	47.682	-4.534	4.47	-3.161	-19.891
TOTALS FOR FUNCTIONAL GROUP						119.87	3.48	8.87	-3.898	-8.887	49.288	-4.978	4.46	-3.164	-19.874
AVERAGES OVER FUNCTIONAL GROUPS:						188.25	3.81	8.83	-2.783	-8.988	43.791	-4.259	4.46	-3.134	-19.789
STATISTICAL UNCERTAINTY (+/- 2*SD)							8.18	8.85	8.356	8.116	8.823	8.328	8.82	8.821	8.186
URACIL															
MOLECULAR SUM/AVERAGE:						639.78	18.76	8.88	-14.258	-8.768	215.888	-28.859	4.43	-3.166	-19.722
STATISTICAL UNCERTAINTY (+/- 2*SD)							8.78	8.83	1.864	8.856	8.872	8.961	8.81	8.814	8.868

Figure IV.5.3 First shell coordination numbers for the atoms of uracil - Kollman solute-water potential function simulation.



Figures IV.5.4 through IV.5.10 - calculated quasicomponent distribution function of first shell coordination number for uracil and functional groups (Kollman potentials).

X axis - coordination number.

Y axis - quasicomponent of coordination number.

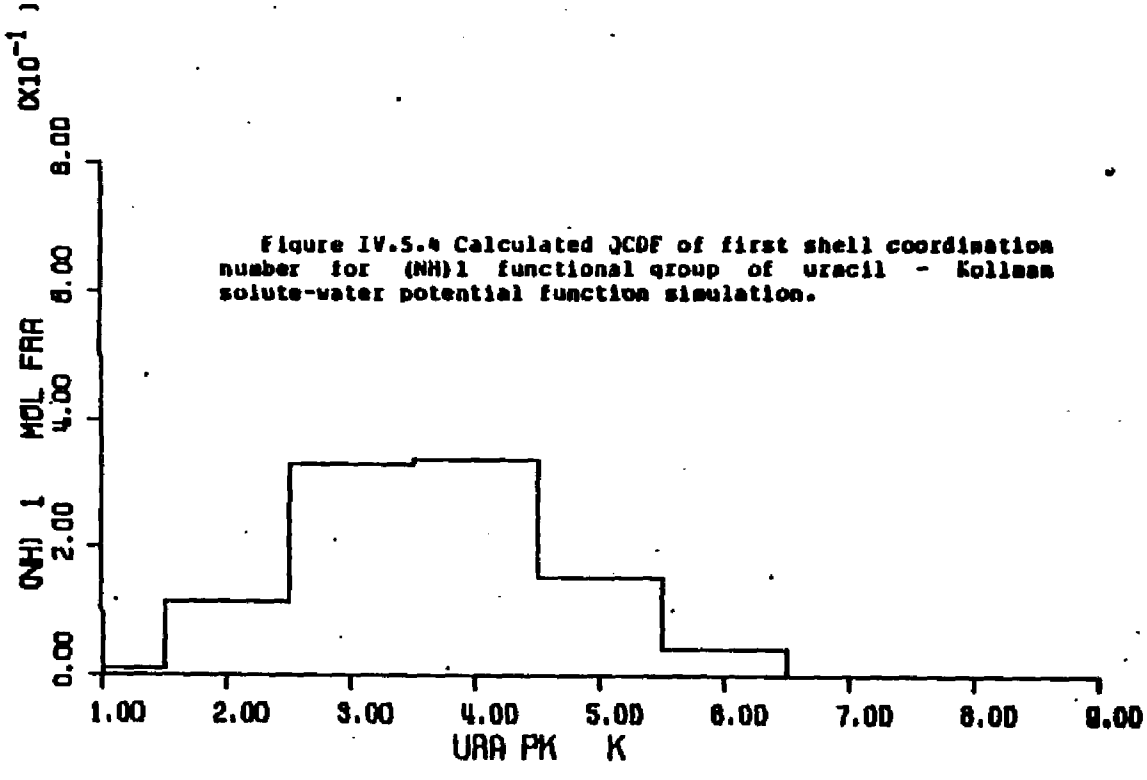


Figure IV.5.4 Calculated JCDF of first shell coordination number for (NH)1 functional group of uracil - Kollman solute-water potential function simulation.

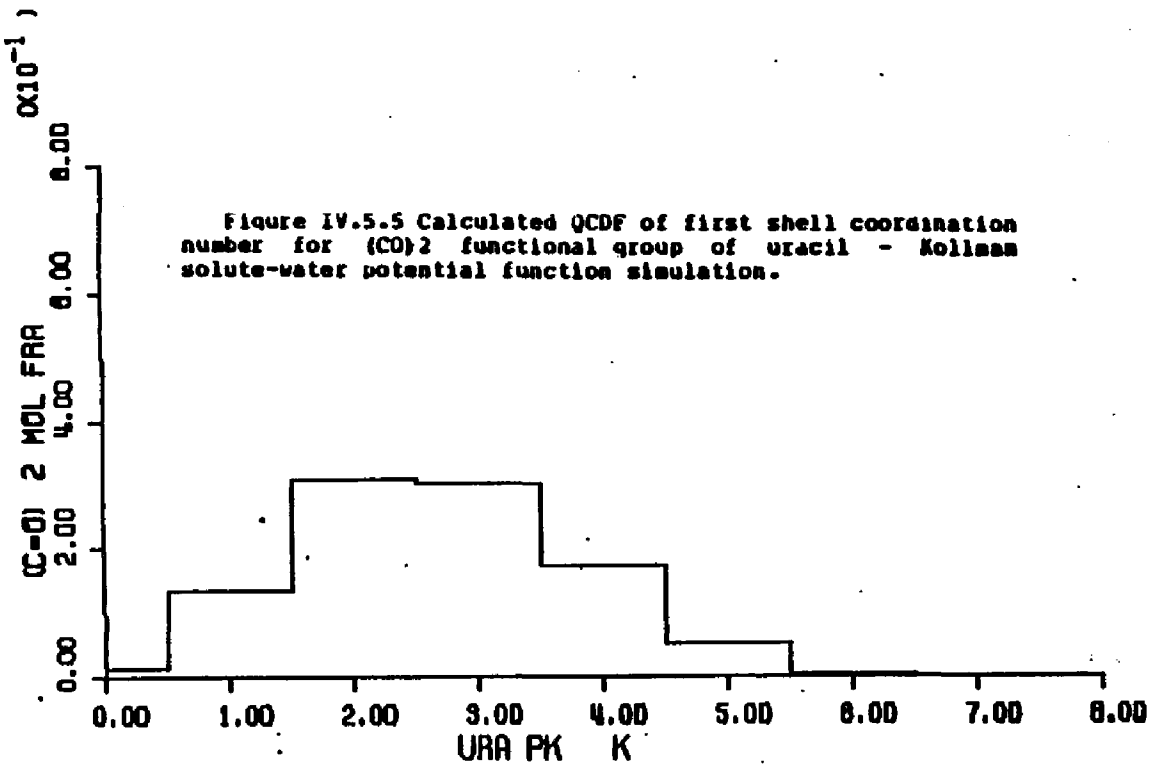


Figure IV.5.5 Calculated QCDF of first shell coordination number for (CO)2 functional group of uracil - Kollman solute-water potential function simulation.

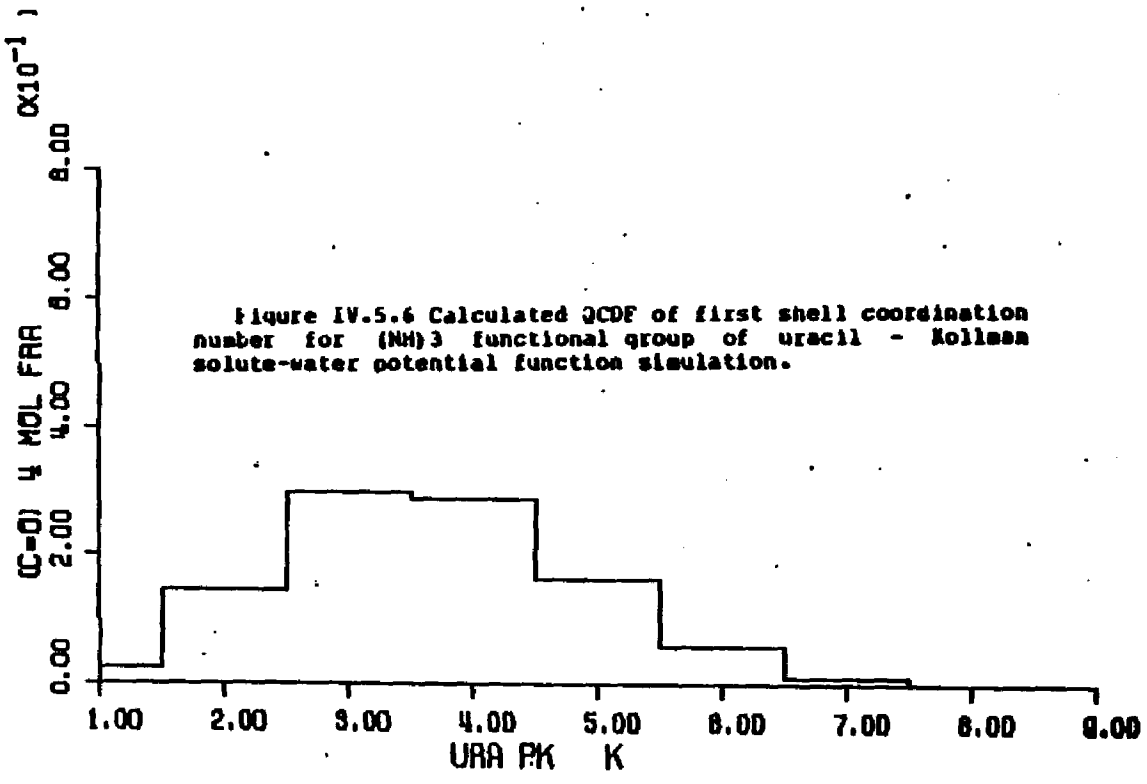


Figure IV.5.6 Calculated QCDF of first shell coordination number for (NH)3 functional group of uracil - Kollman solute-water potential function simulation.

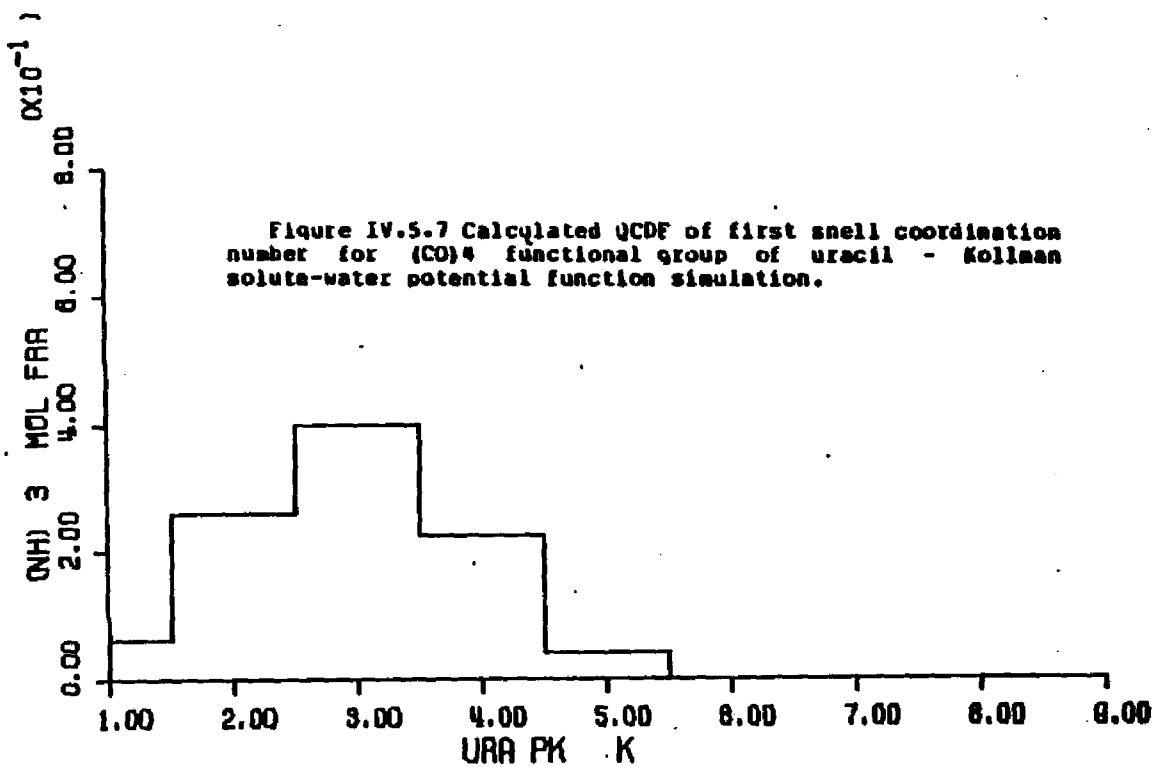


Figure IV.5.7 Calculated QCDF of first shell coordination number for (CO)4 functional group of uracil - Kollman solute-water potential function simulation.

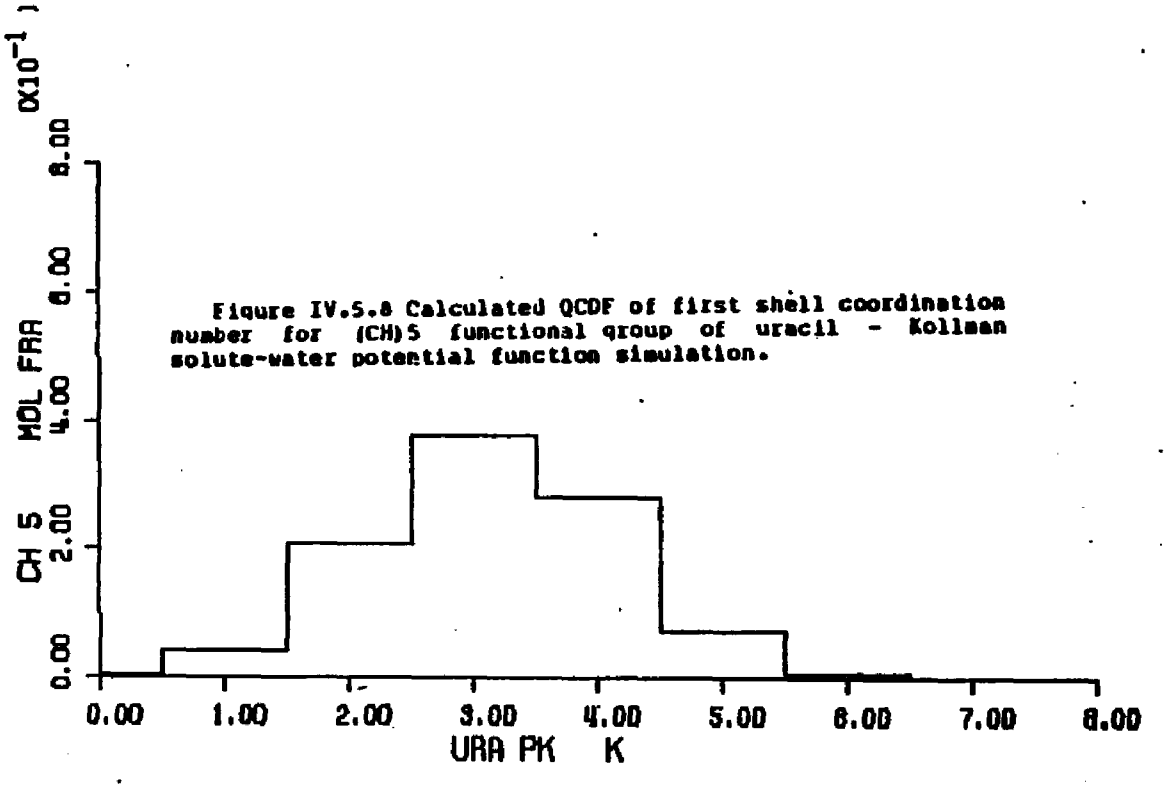


Figure IV.5.8 Calculated QCDF of first shell coordination number for (CH)5 functional qroup of uracil - Kollman solute-water potential function simulation.

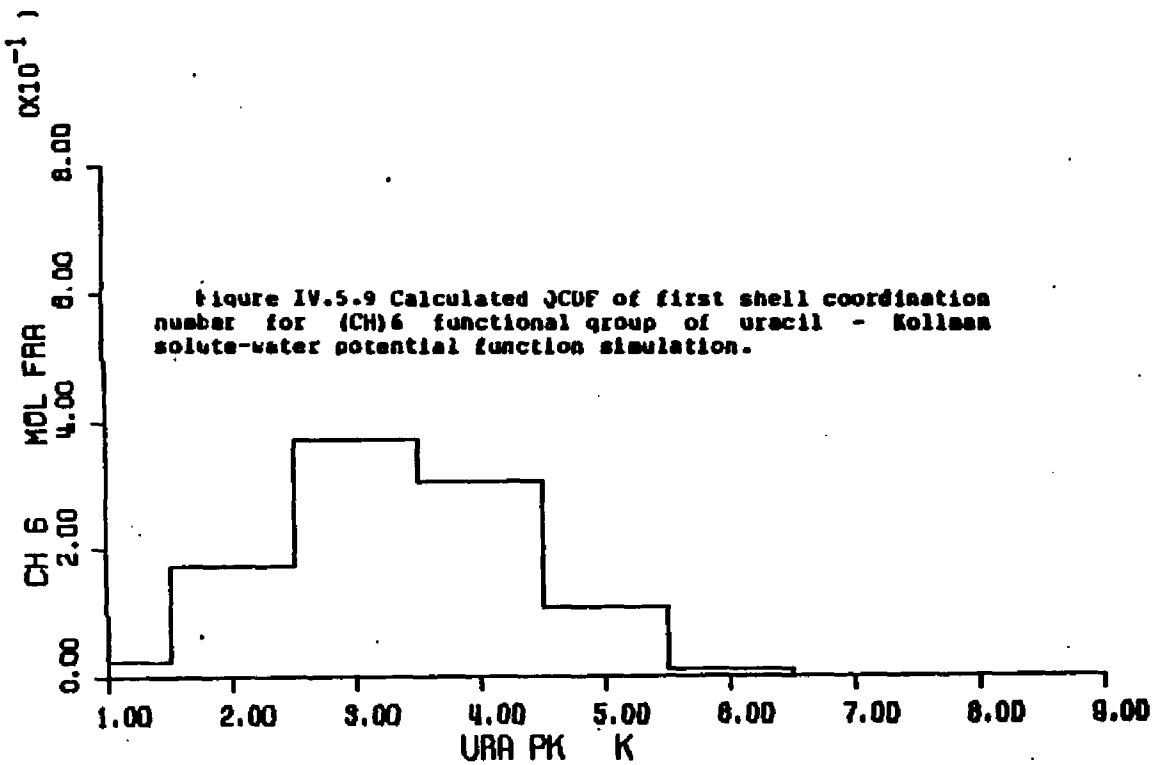


Figure IV.5.9 Calculated QCUF of first shell coordination number for (CH)6 functional qroup of uracil - Kollman solute-water potential function simulation.

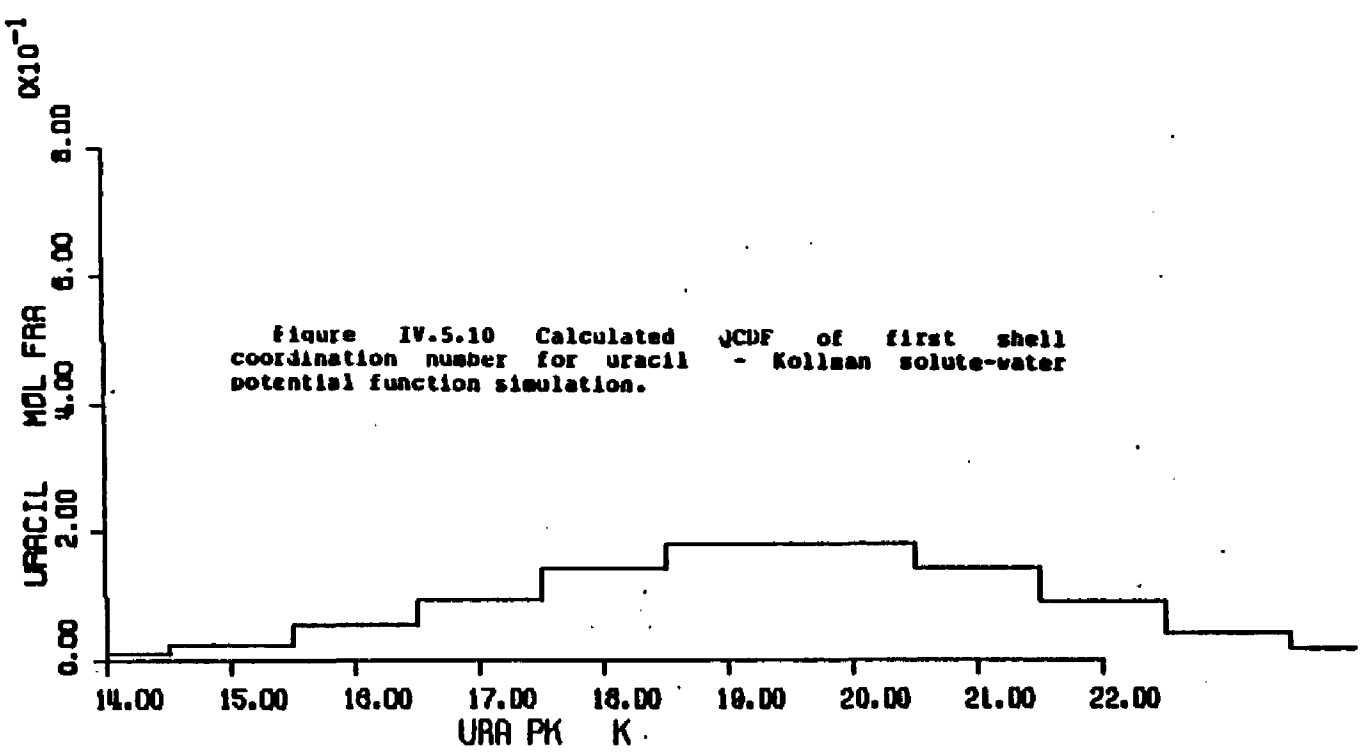


Figure IV.5.10 Calculated JCDF of first shell coordination number for uracil - Kollman solute-water potential function simulation.

Figure IV.5.11 Average first shell solute-water pair energies of waters assigned to the atoms of uracil - Kollman solute-water potential function simulation.

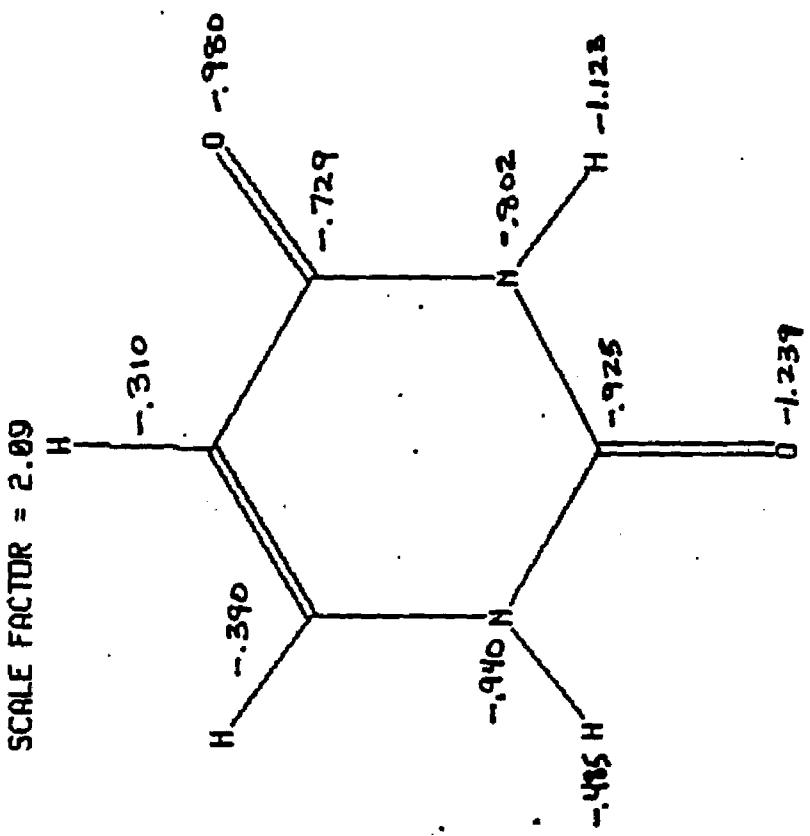


Figure IV.5.12 through IV.5.17 - calculated quasicomponent distribution function of average first shell pair energy for functional groups of uracil (Kollman potentials).

X axis - pair energy (kcal/mole).

Left Y axis - quasicomponent of pair energy.

Right Y axis - running coordination number.

Figure IV.5.12 Calculated QCDF of solute-water pair energies of waters of the (NH)1 functional group of uracil - Pollman solute-water potential function simulation.

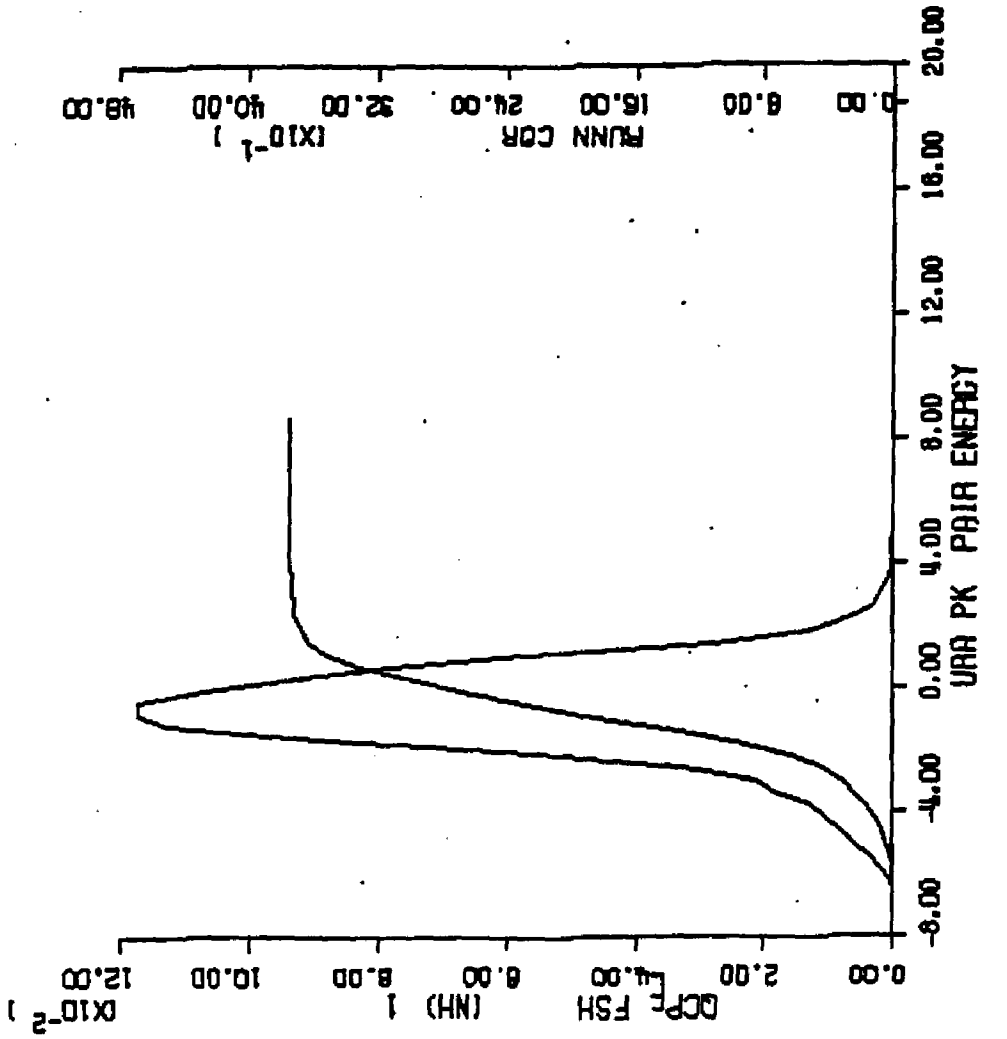


Figure IV.5.13 Calculated QCDF of solute-water pair energies of waters of the (CO)2 functional group of uracil - Kollman solute-water potential function simulation.

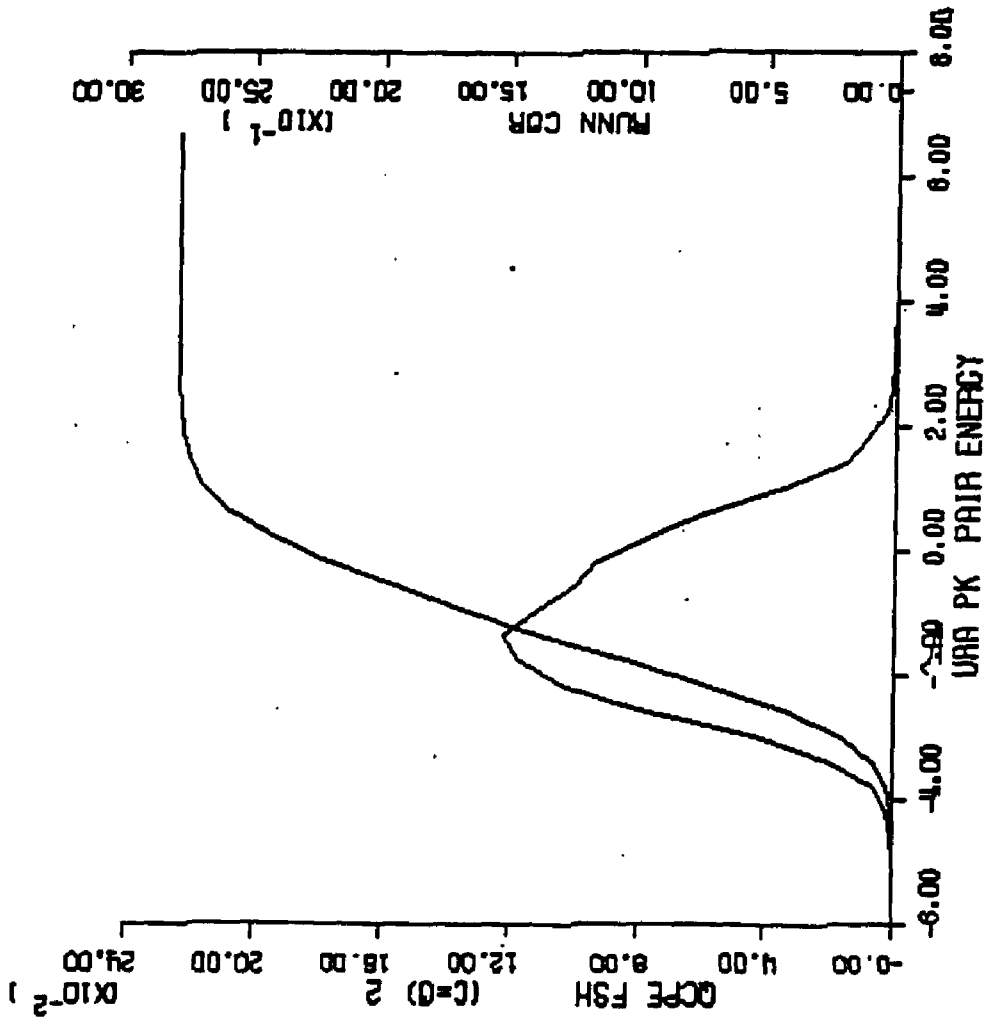


Figure IV.5.14 Calculated ΔCDF of solute-water pair energies of waters of the (NH)3 functional group of uracil - Kollman solute-water potential function simulation.

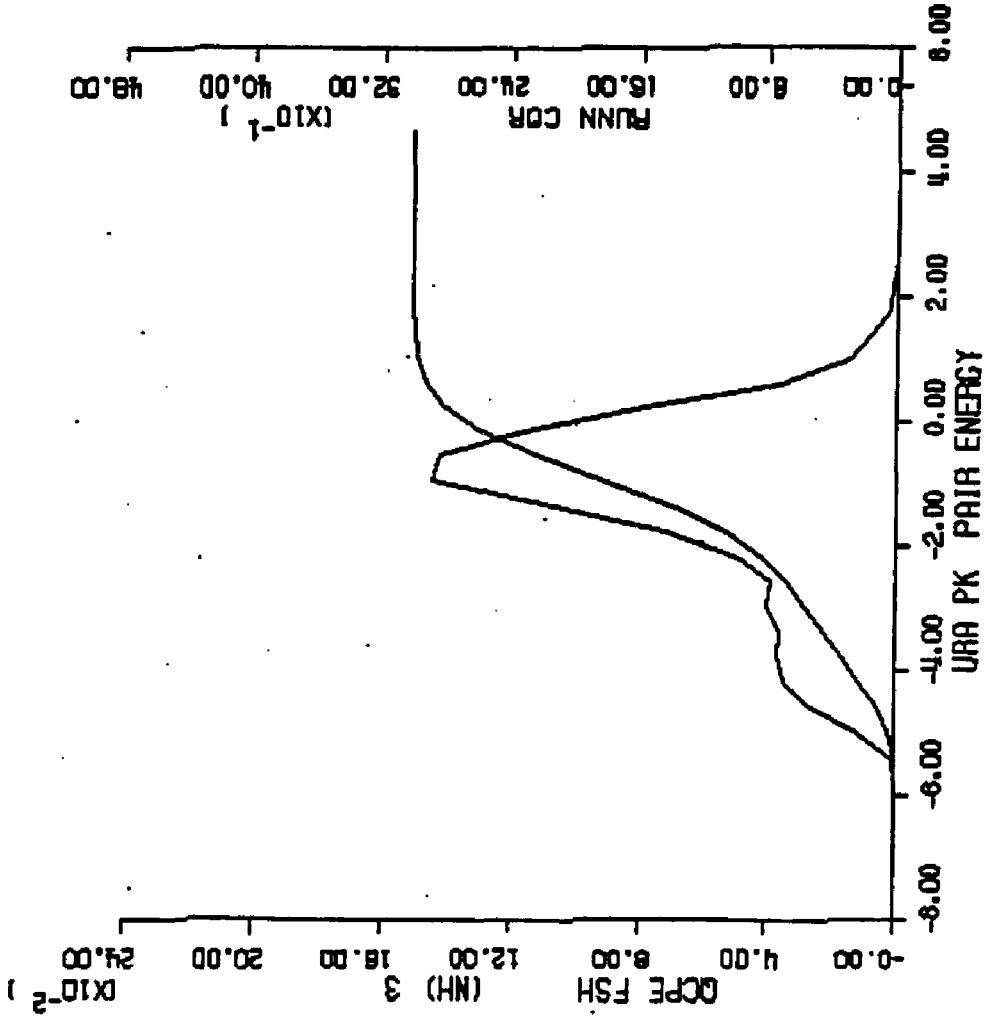


Figure IV.5.15 Calculated QCDF of solute-water pair energies of waters of the (CO)$_4$ functional group of uracil - Kollman solute-water potential function simulation.

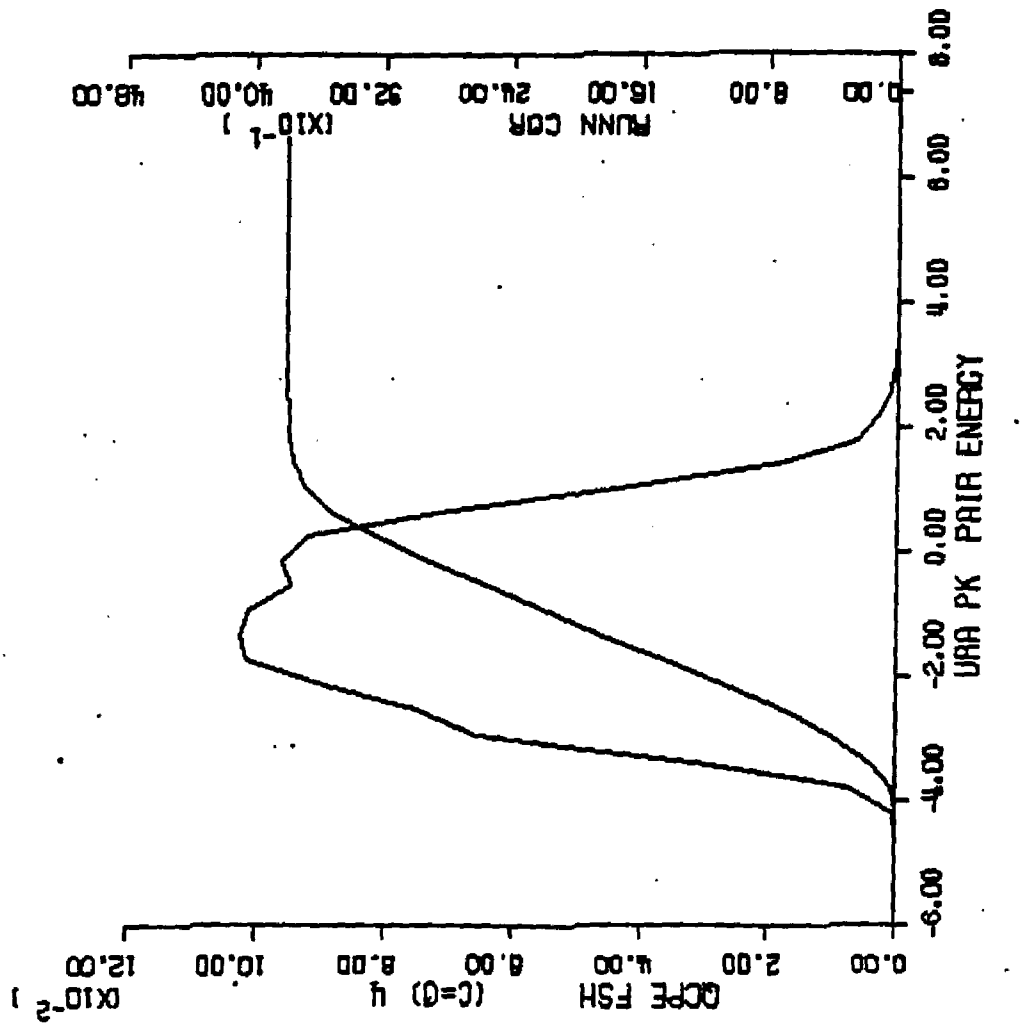


Figure IV.5.16 Calculated QCDF of solute-water pair energies of waters of the (Cd)5 functional group of uracil - Kollman solute-water potential function simulation.

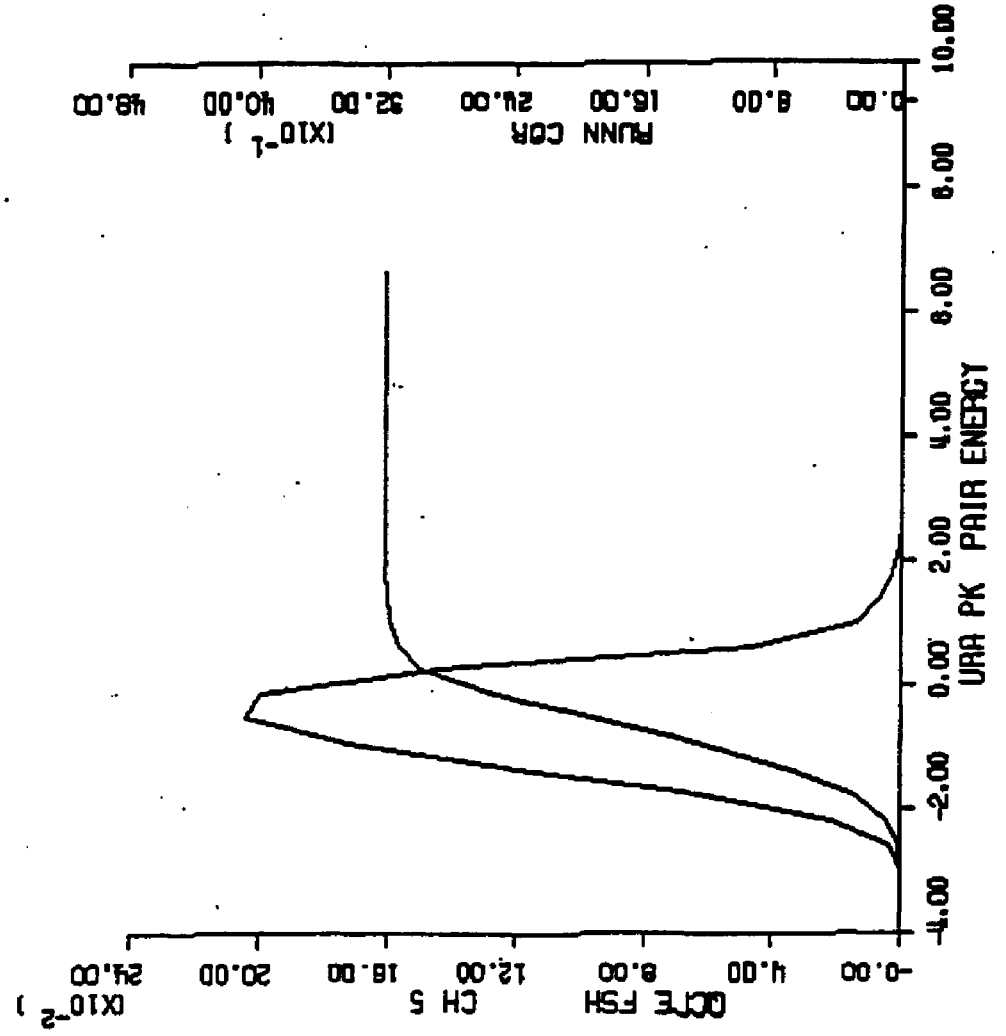


Figure IV.5.17 Calculated QCDF of solute-water pair energies of waters of the (CH)6 functional group of uracil - Kollman solute-water potential function simulation.

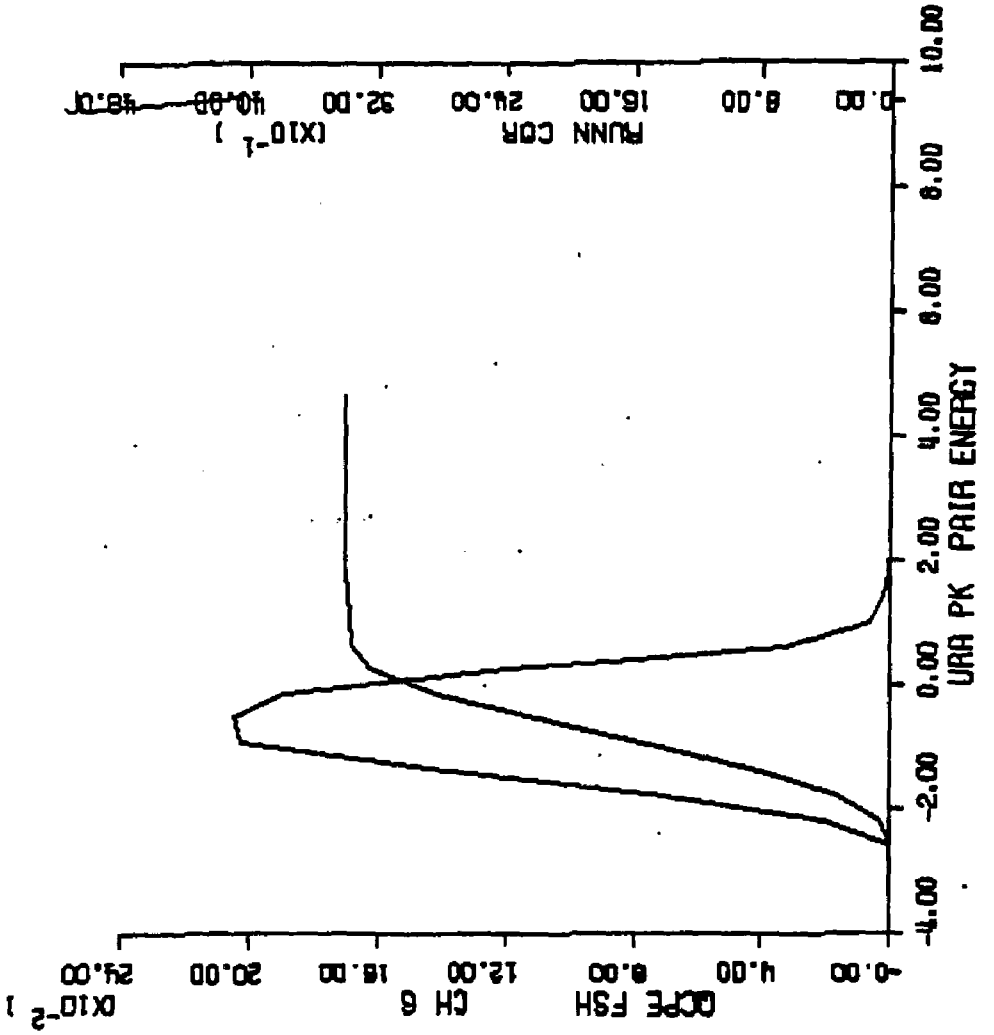
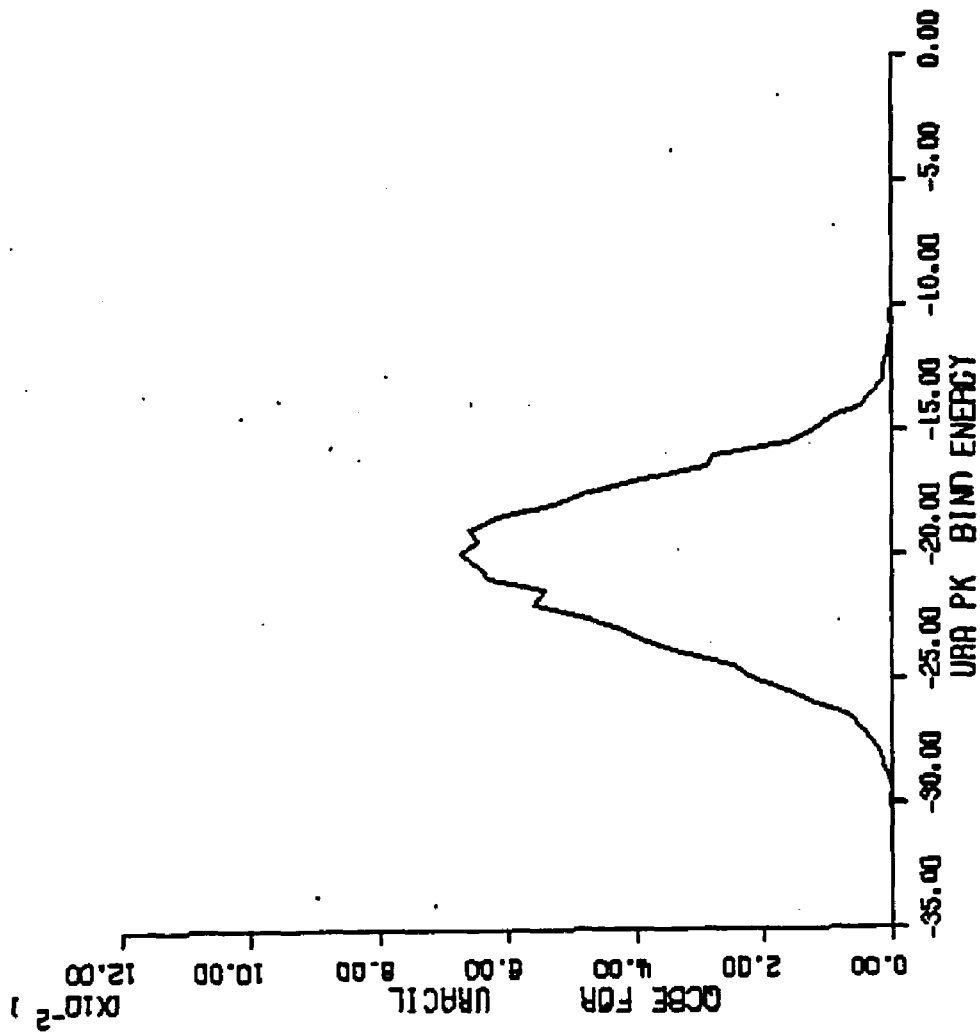


Figure IV.5.18 - calculated quasicomponent distribution function of binding energy for uracil (Kollman potentials).

X axis - binding energy (kcal/mole).

Y axis - quasicomponent of binding energy.

Figure IV.5.18 Calculated GCDF of total binding energy for uracil - Kollman solute-water potential function simulation.



6. Uracil with BVG Solute-Water Potentials

Computational Specifics. An aqueous solution of uracil and 215 waters was simulated using the Berendsen-VanGunsteren solute-water potential functions and SPC water-water potentials functions at 296 K. Face centered cubic boundary conditions were used; the unit cell edge was 14.87 cubic Angstroms. The cell edge was calculated using a partial molar volume of 18.07 cc/mole for water and an estimated 65 cc/mole for the uracil molecule. The simulation was performed with an initial 600,000 move equilibration segment which was discarded. A subsequent 2,000,000 force-biased moves were saved and used to compute all ensemble averages and perform analyses.

Potential Surface. The potential energy surface for the interaction of uracil and one water molecule using BVG solute-water potentials is shown in Figure IV.6.1. The minimum interaction energy occurs near the H3 imino hydrogen at -8.39 kcal/mole. Another minimum is found between the H1 imino hydrogen and O2 carbonyl oxygen where the energy ranges between -4.39 and -5.39 kcal/mole. Interaction energies fall between -2.39 and -3.39 kcal/mole in the vicinity of the carbonyl oxygens, directly in line with the carbonyl double bonds. Finally, the H5 and H6 methine hydrogens each have nearby minima with energies in the range -3.39 to -4.39 kcal/mole. It is striking that the waters near methine hydrogens interact at energies as favorable as waters in the O2, O4 and H1 regions.

Convergence and Thermodynamic Results. Control functions for the simulation of uracil and 215 waters are shown in Figure IV.6.2. Thermodynamic quantities calculated for this simulation are listed in Table IV.1. The mean energy remains essentially stable for the last 600K moves of the simulation around -2210 kcal/mole and after 2000K moves has apparently converged to a final value of -2211.1 +/- 5.4 kcal/mole. The heat capacity reaches the value 18.5 cal/mole-K and a vacuum to water transfer energy for uracil of -22.7 kcal/mole is calculated. Graphs of the quasicomponents of coordination number and first shell pair energy for the imino, carbonyl and methine functional groups of uracil are displayed in Figures IV.6.4 through IV.6.8 and IV.6.12 thru IV.6.17. Figures IV.6.9 and IV.6.18 contain graphs of the quasicomponents of molecular coordination number and binding energy for this simulation.

Structural Results. The molecular diagram of Figure IV.6.3 displays the first shell coordinates numbers found for uracil in this simulation; these numbers are obtained from the <K> column of Table IV.7. The imino hydrogens are seen to have well populated first shells. The H1 hydrogen has 3.51 waters within a 3.8 A cutoff. The first shell of the H3 atom is somewhat less populated with 2.2 waters in a 3.2 A shell. The polar carbonyl oxygens have similar first shell cutoffs (O2: 3.6 A;O4: 3.8 A) but differ by nearly one and one half waters in

first shell population (O2: 2.93; O4: 4.34). This difference is paralleled by a roughly proportional difference in first shell volumes (O2: 89.09 cubic A; O4: 114.53).

The methine (CH)5 and (CH)6 united atom ring fragments have similar structural hydration characteristics. Both first hydration shells have the same extended radial cutoff value (4.6 Angstroms). In addition both have the high first shell populations ((CH)5: 3.13; (CH)6: 3.47) and low first shell densities ((CH)5: 0.97;(CH)6: 0.93) often found in the hydration of apolar hydrogens.

The quasicomponent distribution function of coordination number for uracil in this simulation is found in Figure IV.6.10. The coordination numbers range from 16 to 26 and the greatest contributions are from coordination numbers of 21 and 22. The average molecular coordination number is 21.00 +/- 0.54.

Energetic Results. First shell pair energies for waters assigned to the atoms of the uracil in this simulation are displayed in the molecular diagram of Figure IV.6.11; these values are obtained from the first shell <SLIPE> column of Table IV.7. The waters of the H3 imino hydrogen are the most strongly bound to the solute at an average -2.432 kcal/mole. These waters contribute the most to the total solute binding energy (-5.348 kcal/mole). The waters of the H1 imino hydrogen are

considerably less favorably bound at an average -.770 kcal/mole each.

The waters of the methine ring fragments are moderately strongly bound compared with other waters. For instance, the waters of the O4 carbonyl oxygen bind at an average -.563 kcal/mole whereas the waters of (CH)5 bind at -.683 kcal/mole each and the waters of (CH)6 bind at -.781 kcal/mole each.

The quasicomponent distribution function of first shell pair energy for uracil in this simulation is shown in Figure IV.6.18. Energies range from -38 to -17 kcal/mole with significant contributions in the range -31 to -25 kcal/mole. The average solute binding energy is -27.0 +/- 0.5 kcal/mole.

Figure IV.6.1 - isoenergy contour surface for uracil and one water (BVG solute-water potential functions).

X axis - X axis of plane defined by molecular ring atoms (Angstroms).

Y axis - Y axis of plane defined by molecular ring atoms (Angstroms).

Contours - isoenergy contours with one kcal/mole increments with alphabetical labels referring to contour energy values in list at right.

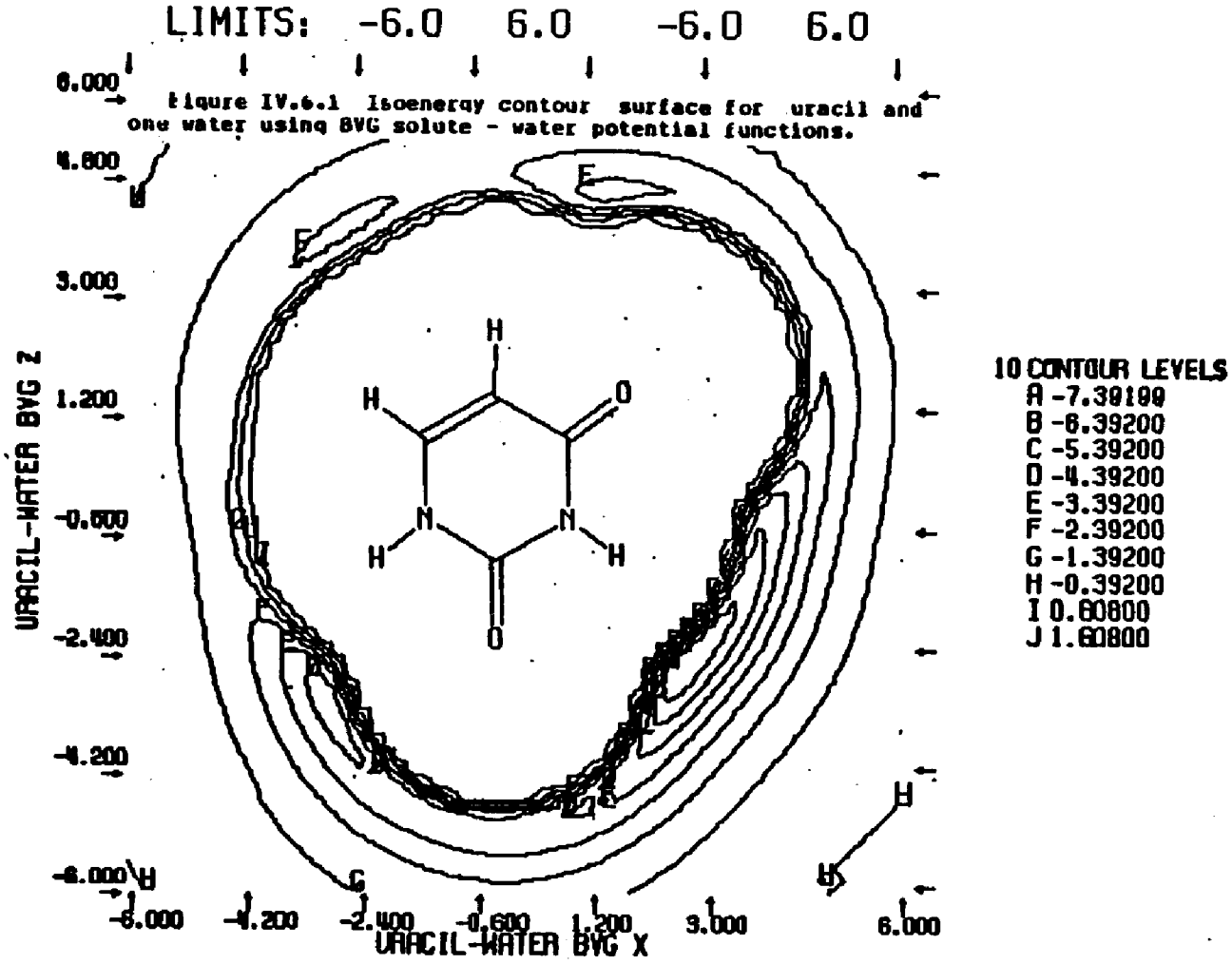


Figure IV.6.1 Isoenergy contour surface for uracil and one water using BVG solute - water potential functions.

Figure IV.6.2 - control functions for Monte Carlo simulation of uracil and 215 waters (BVG solute-water potential functions).

X axis - number of configurations.

Left Y axis - mean total energy (kcal/mole).

Right Y axis. - constant volume heat capacity (cal/mole-degree).

Upper curve - constant volume heat capacity.

Bottom curve without crosshatches - average total energy for entire simulation.

Bottom curve with crosshatches - average total energy for preceeding 50K configurations.

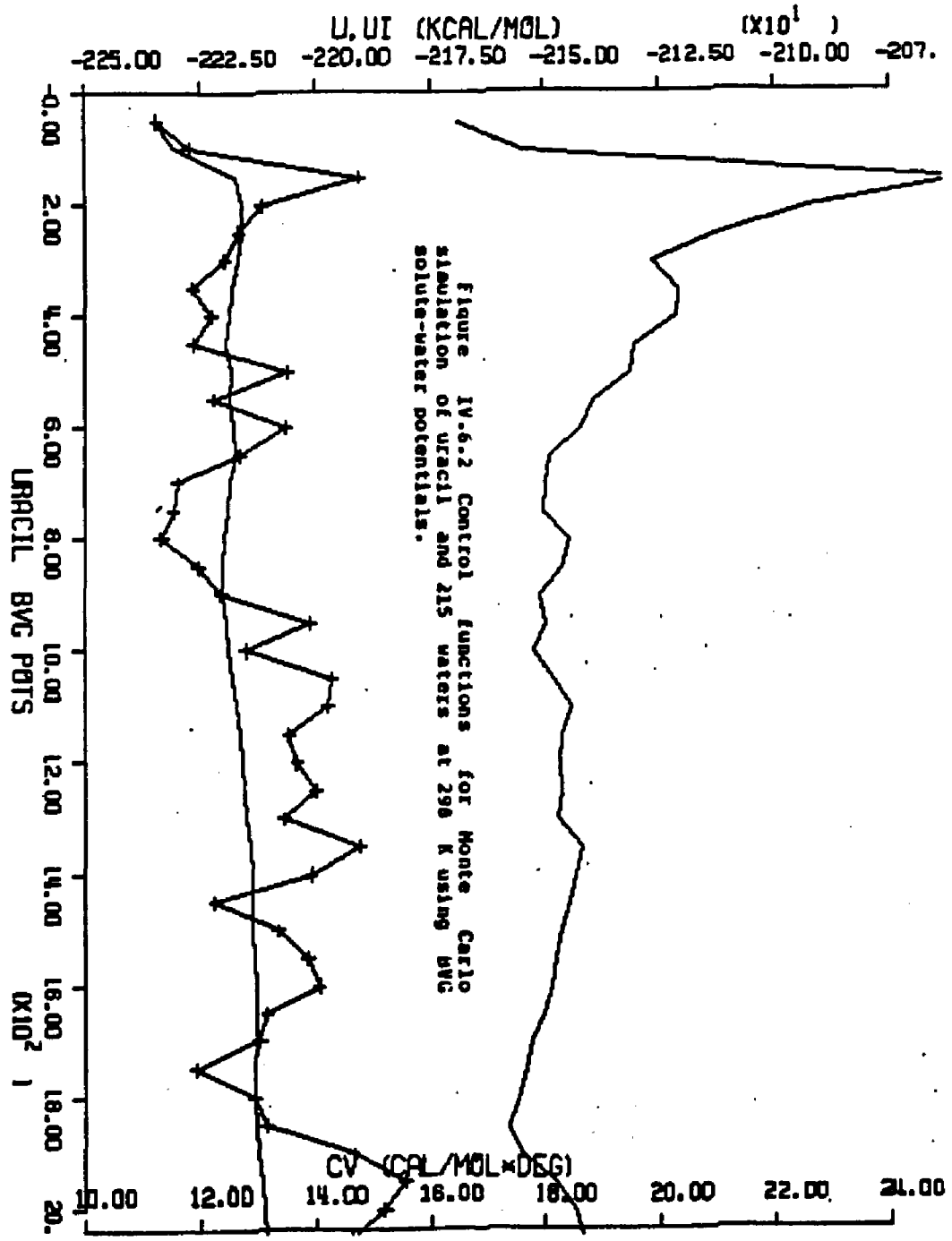


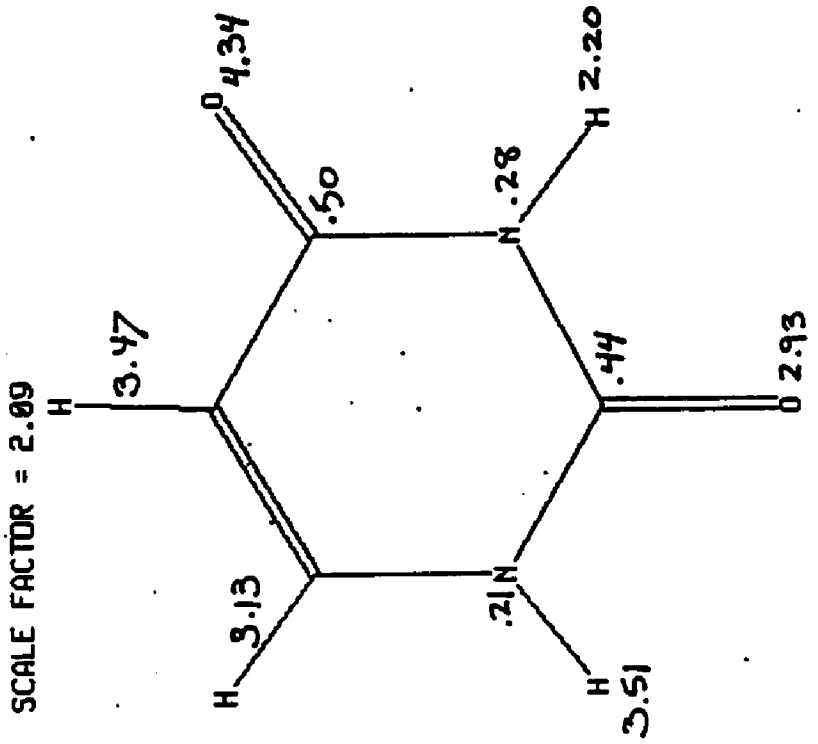
Figure IV.6.2 Control functions for Monte Carlo simulation of uracil and 215 waters at 298 K using BWC solute-water potentials.

URACIL IN WATER AT 298K - FORCE BIAS AND BVG POTENTIAL FUNCTIONS

LAST CONFIGURATION: 2000001						FIRST SHELL SOLUTE PROPERTIES					TOTAL SLT PROPS		WATER PROPERTIES RFSW=3.38 RCB= 7.75 A		
INDEX	TYP				RFS	VFS	<K>	<K/V>	<SLTDE>	<SLTPE>	<K>	<SLTDE>	<NN>	<NNMPE>	<EBMT>
1	3	17	CH	6	4.6	187.34	3.47	0.97	-2.709	-0.781	26.39	-3.538	4.36	-3.343	-20.268
2	10	17	CH	5	4.6	181.15	3.13	0.93	-2.141	-0.683	23.26	-2.718	4.40	-3.365	-20.471
AVERAGES OVER FUNCTIONAL GRP =C= :						184.25	3.30	0.95	-2.425	-0.732	24.83	-3.128	4.38	-3.354	-20.369
STATISTICAL UNCERTAINTY (+/- 2*SD):							0.15	0.04	0.158	0.063	0.01	0.156	0.02	0.049	0.163
AMIDE GROUP															
3	1	31	N	1	3.6	14.64	0.21	0.43	-0.189	-0.898	1.26	-0.317	4.29	-3.411	-20.351
4	2	19	H	1	3.8	101.72	3.51	1.03	-2.706	-0.770	39.65	-3.797	4.37	-3.359	-20.354
TOTALS FOR FUNCTIONAL GROUP >NH :						116.36	3.72	0.96	-2.895	-0.777	40.91	-4.115	4.37	-3.360	-20.354
5	6	31	N	3	3.8	15.57	0.28	0.53	-0.298	-1.882	1.32	-0.386	4.27	-3.489	-20.254
6	7	19	H	3	3.2	62.71	2.20	1.05	-5.348	-2.432	36.21	-7.153	4.39	-3.367	-20.475
TOTALS FOR FUNCTIONAL GROUP >NH :						78.28	2.47	0.95	-5.647	-2.282	37.53	-7.539	4.39	-3.368	-20.467
AVERAGES OVER FUNCTIONAL GRP >NH :						97.32	3.10	0.95	-4.271	-1.530	39.22	-5.827	4.38	-3.364	-20.411
STATISTICAL UNCERTAINTY (+/- 2*SD):							0.15	0.05	0.287	0.135	0.02	0.232	0.02	0.039	0.130
CARBONYL GROUP															
7	4	18	C	2	4.2	19.75	0.44	0.66	-0.335	-0.769	1.57	-0.388	4.25	-3.425	-20.332
8	5	3	O	2	3.6	89.09	2.93	0.98	-2.931	-1.001	36.57	-4.590	4.38	-3.400	-20.284
TOTALS FOR FUNCTIONAL GROUP C=O :						108.85	3.36	0.92	-3.266	-0.971	38.14	-4.978	4.38	-3.401	-20.286
9	8	18	C	4	4.2	19.21	0.50	0.78	-0.643	-1.279	1.51	-0.695	4.22	-3.394	-19.915
10	9	3	O	4	3.8	114.53	4.34	1.13	-2.443	-0.563	47.25	-3.798	4.43	-3.337	-20.571
TOTALS FOR FUNCTIONAL GROUP C=O :						133.74	4.84	1.08	-3.085	-0.637	48.76	-4.493	4.42	-3.339	-20.551
AVERAGES OVER FUNCTIONAL GRP C=O :						121.29	4.10	1.01	-3.176	-0.804	43.45	-4.736	4.36	-3.370	-20.419
STATISTICAL UNCERTAINTY (+/- 2*SD):							0.17	0.04	0.185	0.062	0.02	0.179	0.02	0.037	0.124
MOLECULAR SUM/AVERAGE:						645.7	21.00	0.97	-19.742	-0.948	215.00	-27.365	4.37	-3.362	-20.400
STATISTICAL UNCERTAINTY (+/- 2*SD):							0.54	0.03	0.720	0.045	0.05	0.658	0.01	0.024	0.079

Table IV.7 Calculated structural and energetic quantities from Monte Carlo simulation of uracil and 215 waters at 298 K - BVG solute-water potential functions.

Figure IV.6.3 First shell coordination numbers for the atoms of uracil - BVG solute-water potential function simulation.



Figures IV.6.4 through IV.6.10 - calculated quasicomponent distribution function of first shell coordination number for uracil and functional groups (BVG potentials).

X axis - coordination number.

Y axis - quasicomponent of coordination number.

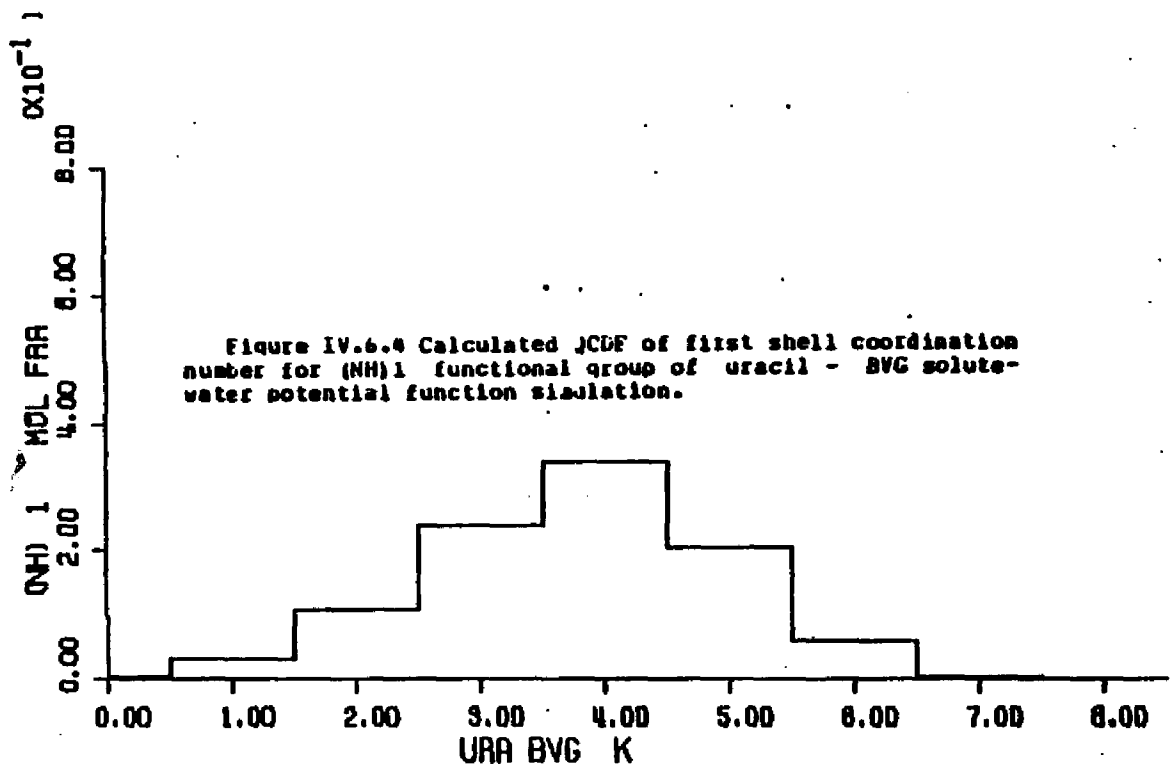


Figure IV.6.4 Calculated JCDF of first shell coordination number for (NH)1 functional group of uracil - BVG solute-water potential function simulation.

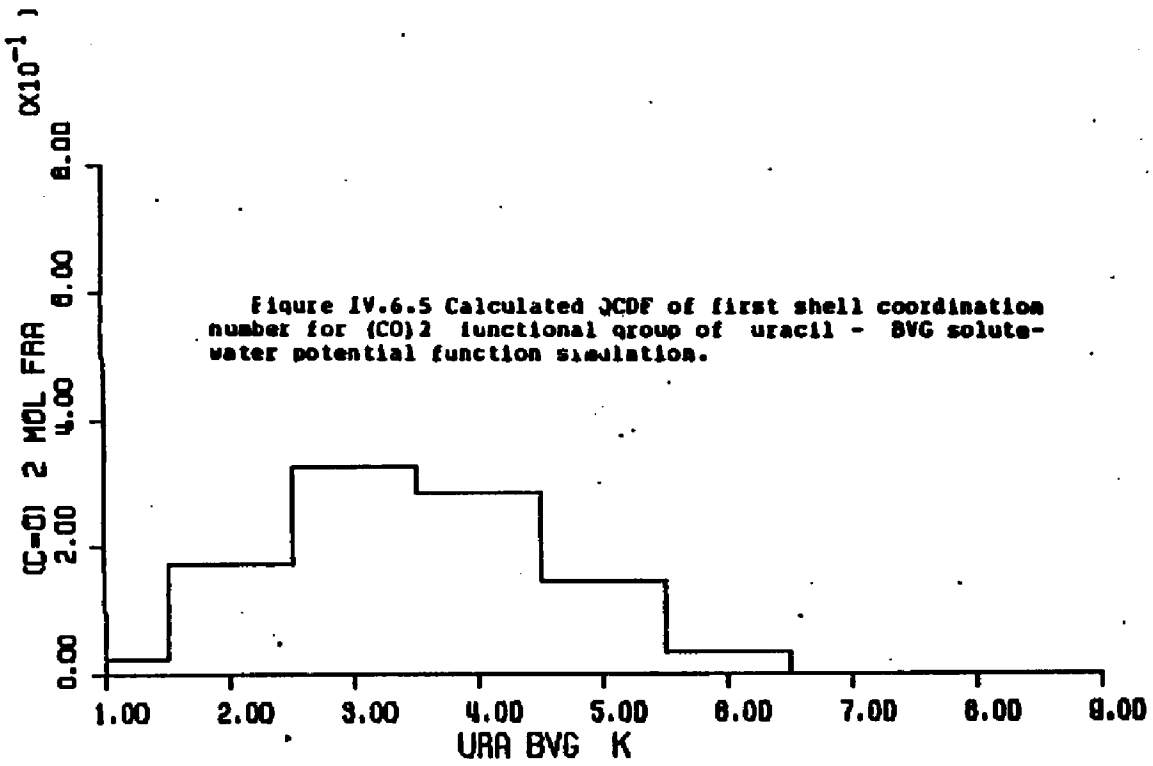


Figure IV.6.5 Calculated JCDF of first shell coordination number for (CO)2 functional group of uracil - BVG solute-water potential function simulation.

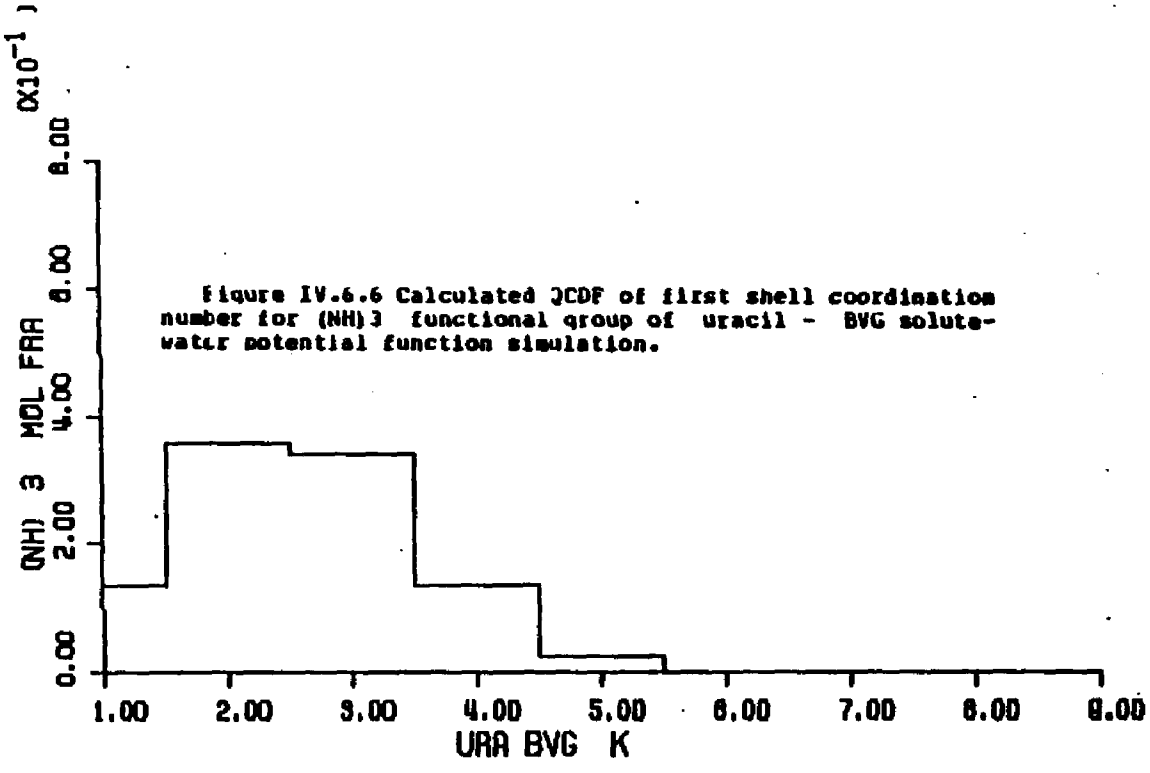


Figure IV.6.6 Calculated JCDF of first shell coordination number for (NH)3 functional group of uracil - BVG solute-water potential function simulation.

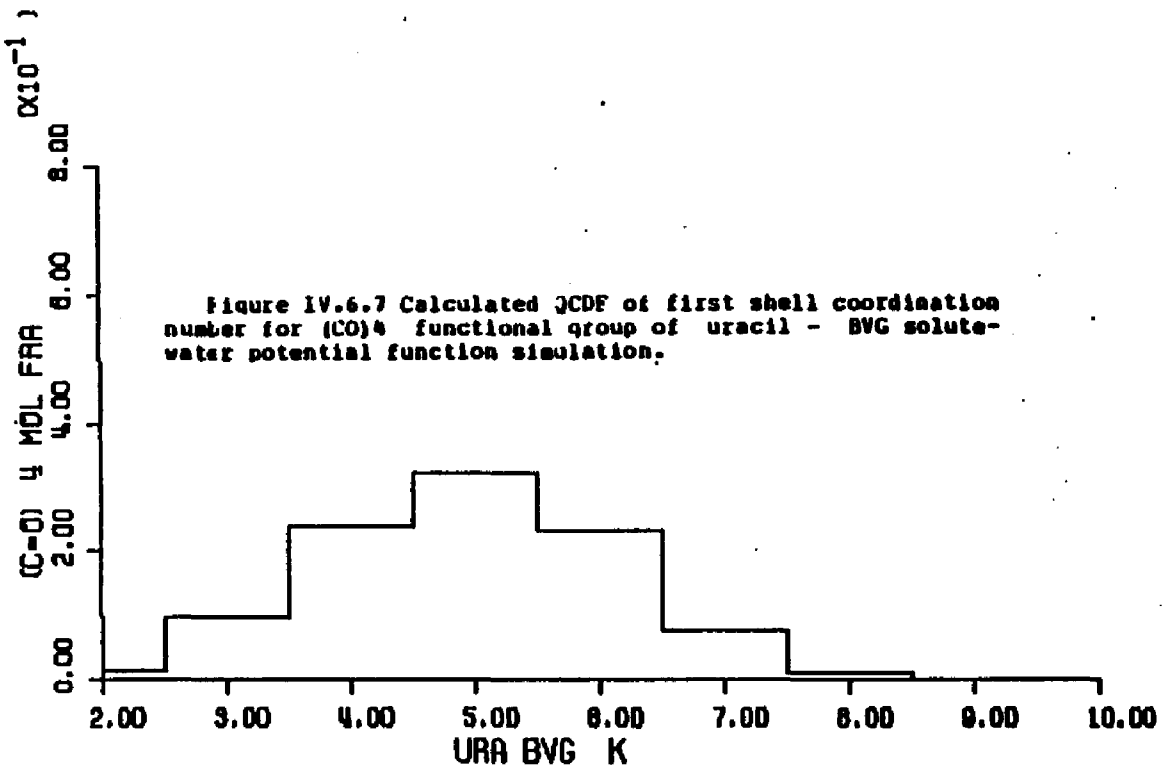


Figure IV.6.7 Calculated QCDF of first shell coordination number for $(CO)_4$ functional group of uracil - BVG solute-water potential function simulation.

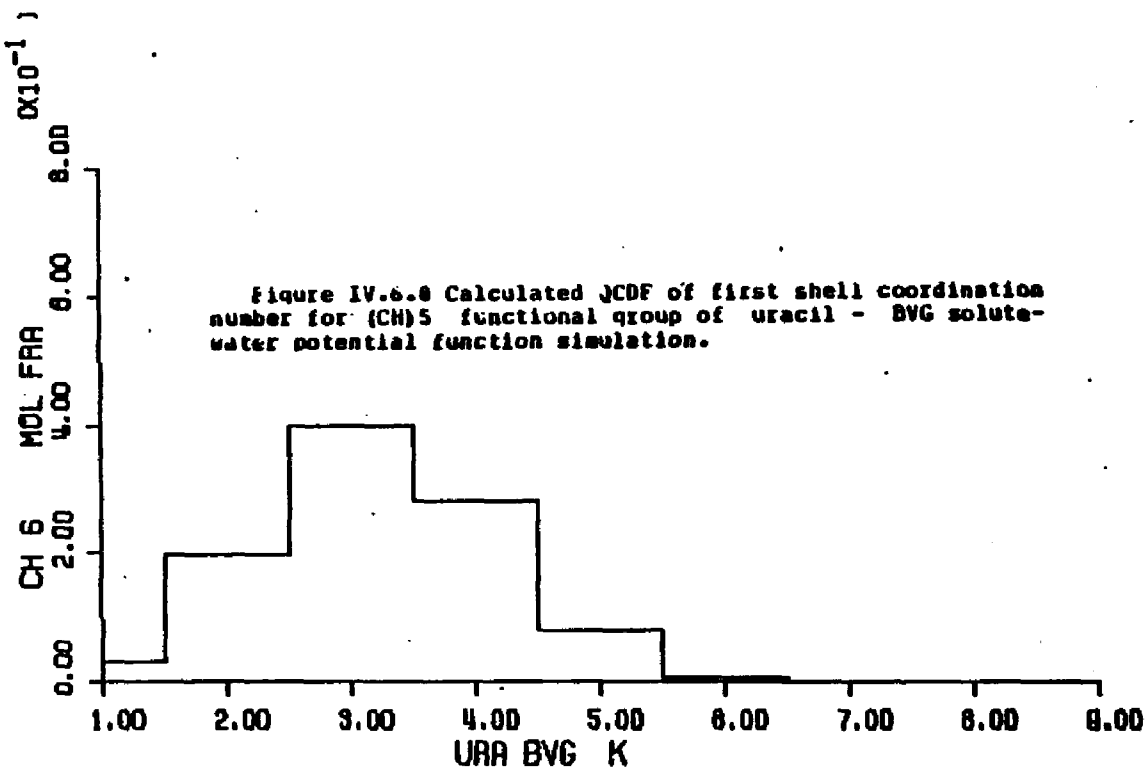


Figure IV.6.8 Calculated JCDF of first shell coordination number for (CH)5 functional group of uracil - BVG solute-water potential function simulation.

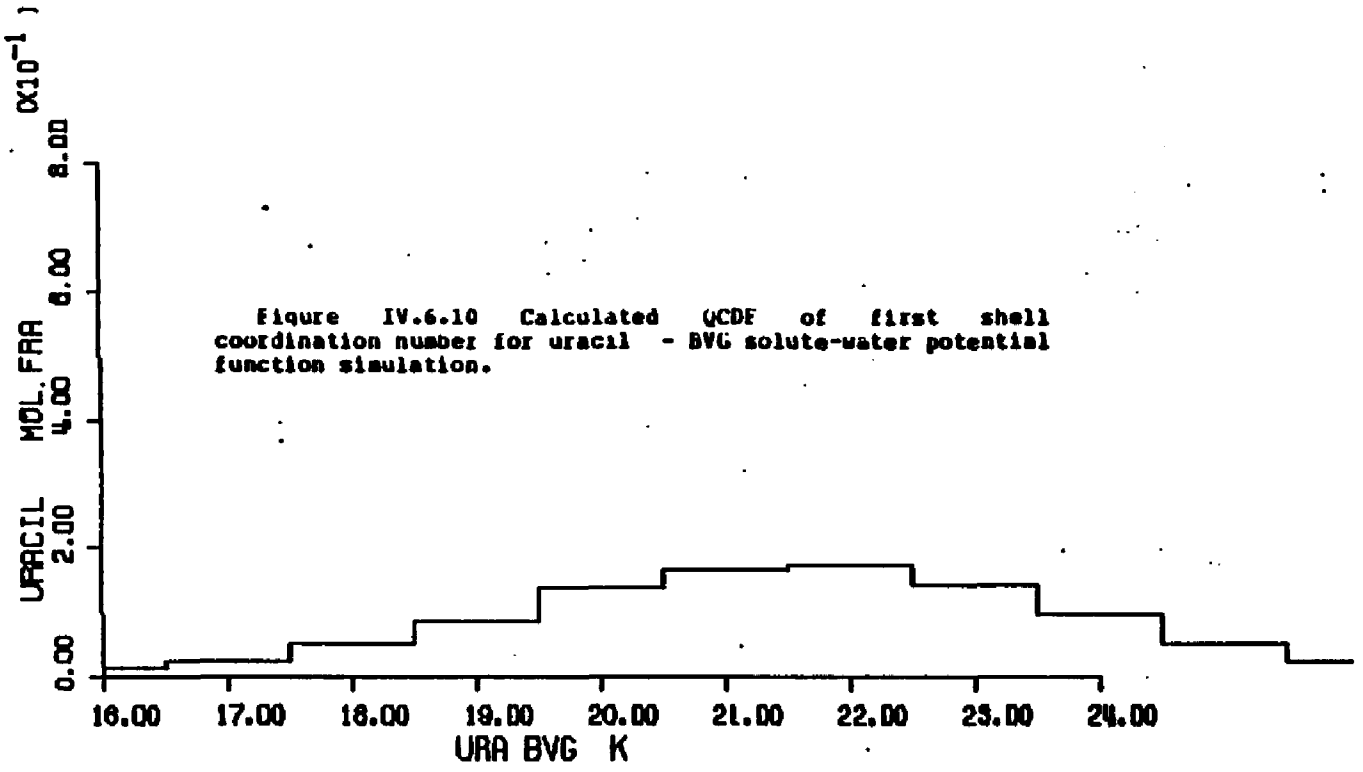


Figure IV.6.10 Calculated QCDF of first shell coordination number for uracil - BVG solute-water potential function simulation.

Figure IV.6.11 Average first shell solute-water pair energies of waters assigned to the atoms of uracil - BVG solute-water potential function simulation.

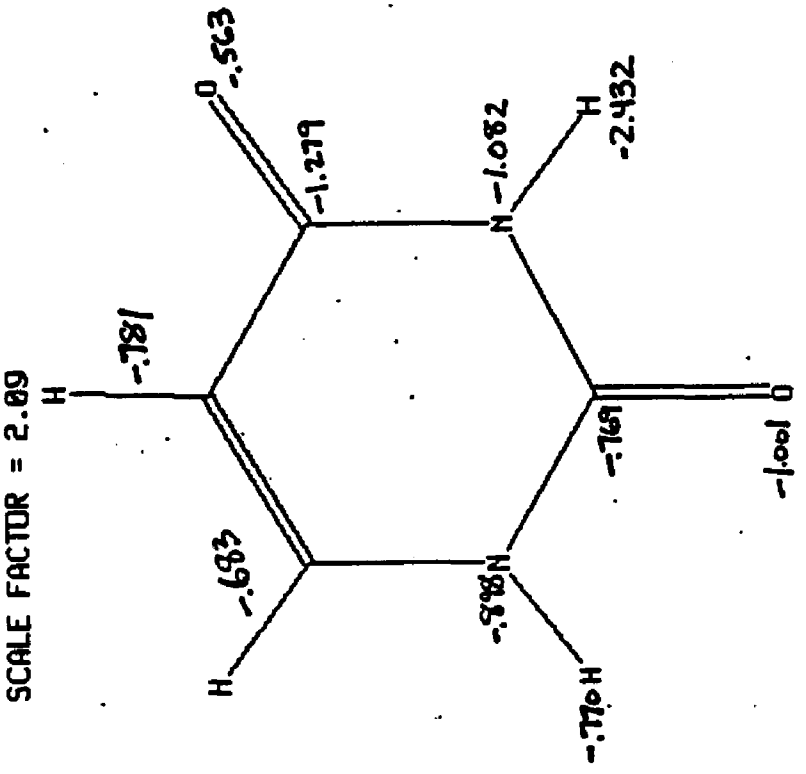


Figure IV.6.12 through IV.6.17 - calculated quasicomponent distribution function of average first shell pair energy for functional groups of uracil (BVS potentials).

X axis - pair energy (kcal/mole).

Left Y axis - quasicomponent of pair energy.

Right Y axis - running coordination number.

Figure IV.6.12 Calculated QCDF of solute-water pair energies of waters of the (NH)1 functional group of uracil - 6VG solute-water potential function simulation.

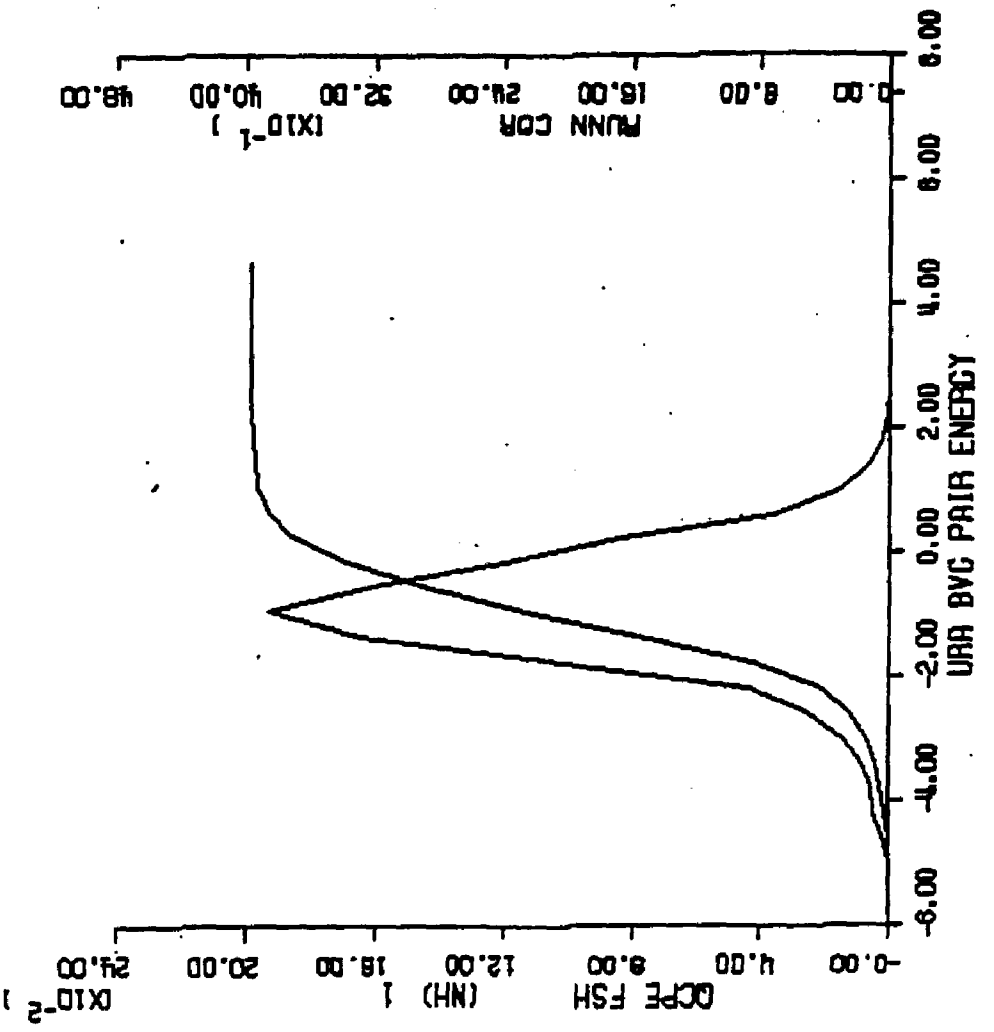


Figure IV.6.13 Calculated QCDF of solute-water pair energies of waters of the (C=O)2 functional group of uracil - BVG solute-water potential function simulation.

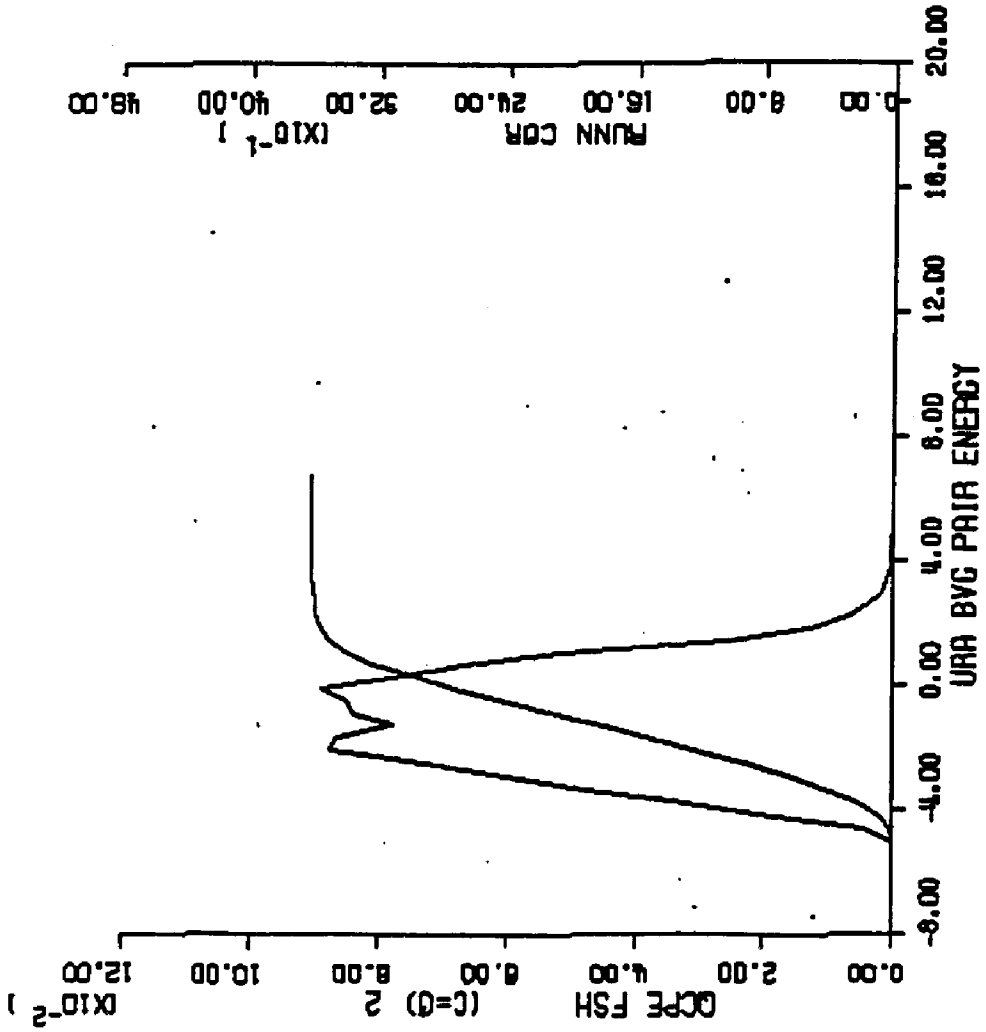


Figure IV.-14 Calculated QCDF of solute-water pair energies of the (NH)3 functional group of uracil - aqueous solute-water potential function simulation.

Figure IV.6.15 Calculated QCOF of solute-water pair energies of waters of the (C)4 functional group of uracil - BVG solute-water potential function simulation.

Figure IV.6.16 Calculated QCDF of solute-water pair energies of waters of the (CH)5 functional group of uracil - BVG solute-water potential function simulation.

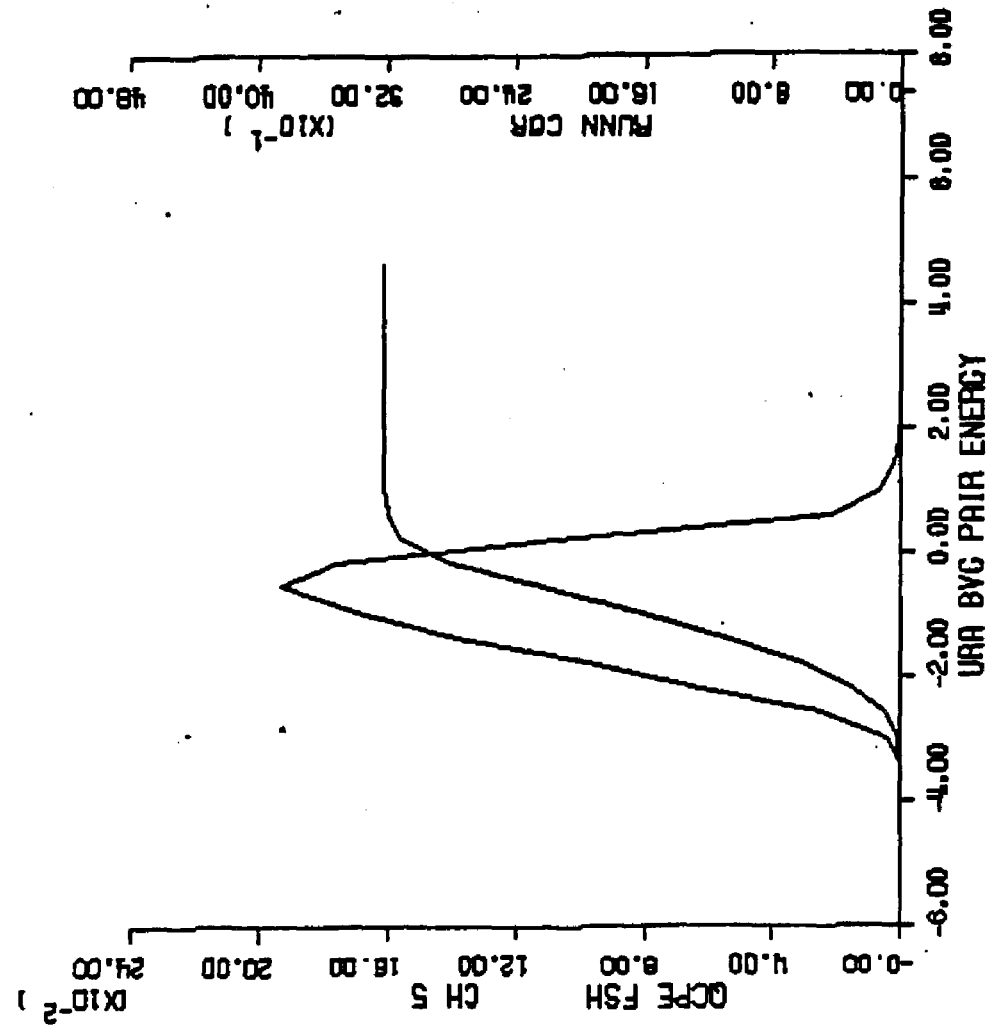


Figure IV.e.17 Calculated QCDF of solute-water pair energies of waters of the (CH)6 functional group of uracil - BVG solute-water potential function simulation.

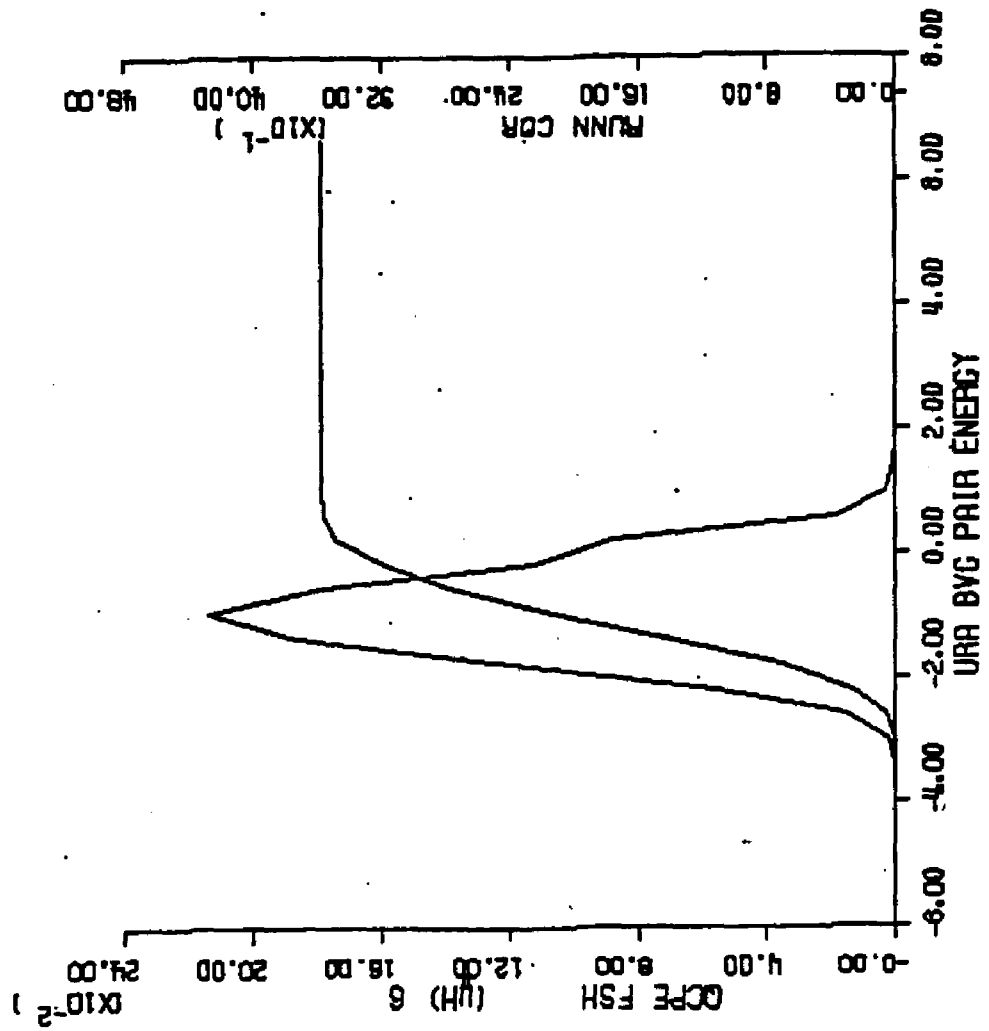
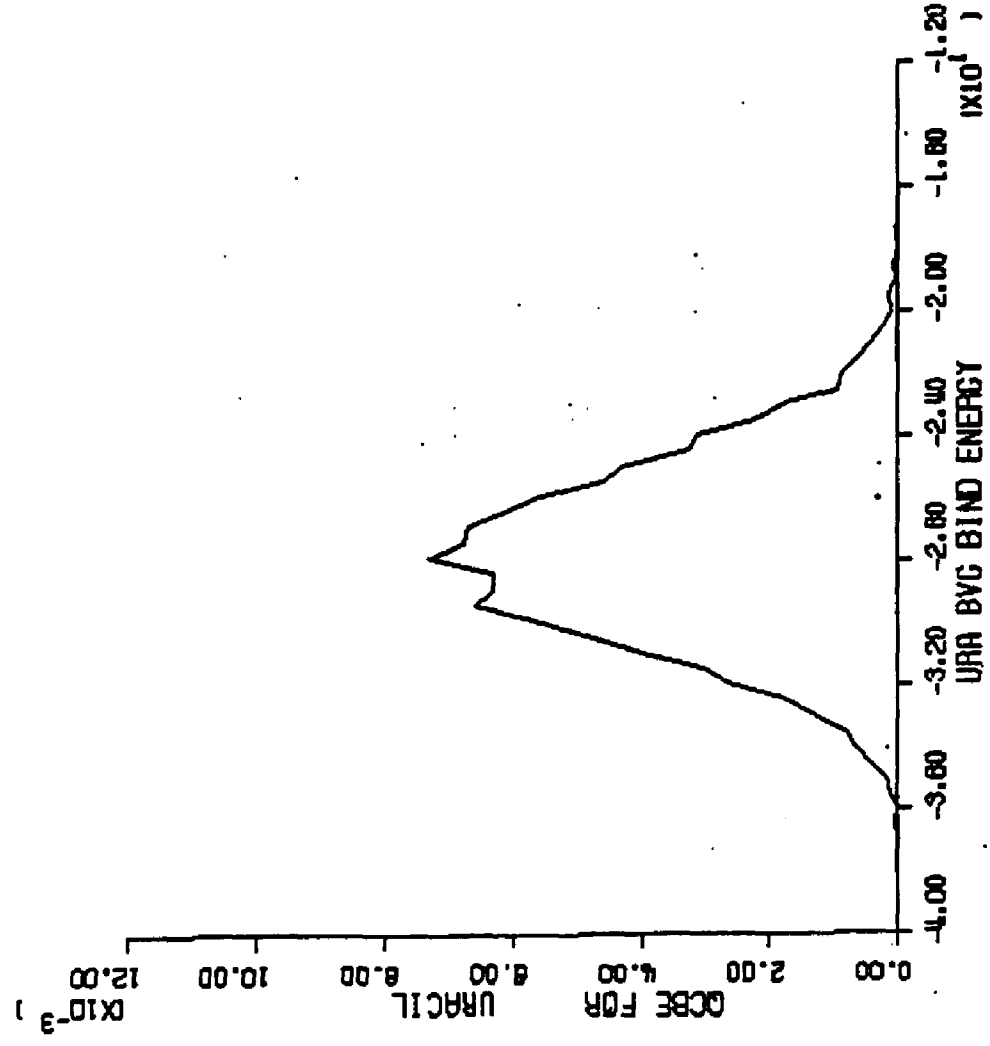


Figure IV.6.16 - calculated quasicomponent distribution function of binding energy for uracil (BVG potentials).

X axis - binding energy (kcal/mole).

Y axis - quasicomponent of binding energy.

Figure IV.6.18 Calculated QCDF of total binding energy for uracil - SVG solute-water potential function simulation.



7. Cytosine with Clementi Solute-Water Potentials

Computational Specifics. An aqueous solution of cytosine and 215 waters was simulated using Clementi solute-water potential functions and MCY-CI potential functions at 298 K. Simple cubic boundary conditions with a unit cell edge length of 18.73 Angstroms, calculated from a partial molar volume of 65 cc/mole for cytosine and 18.07 cc/mole for water, were used. An initial equilibration period of 600,000 moves was discarded; a subsequent 2,000,000 step production period was used in the formation of ensemble averages and for all subsequent analysis.

Potential Surface. The contour surface for the potential energy of the interaction of cytosine and one water molecule is shown in Figure IV.7.1. The global minimum is found next to the imino H1 hydrogen at an energy of -12.92 kcal/mole. The next deepest minima are proximal to the amino hydrogens, H4' and H4'', at energies between -8.92 and -9.92 kcal/mole. The O2 carbonyl oxygen is surrounded by a region containing minimum interactions between -7.92 and -8.92 kcal/mole

The area near the methine hydrogen H6 is also under the influence of the nearby imino hydrogen and contains rather favorable interactions in the range -7.92 and -8.92 kcal/mole. The region of the other methine hydrogen, H5, does not contain such favorable interactions; energies here range between -5.92 and -6.92

kcal/mole. Finally, the region adjacent to the bare ring nitrogen, N3, contains the least favorable interactions of all; energies are found from -4.92 to -6.92 kcal/mole.

Convergence and Thermodynamic Results. The control functions for the aqueous cytosine simulation are given in Figure IV.7.1 and the calculated thermodynamic quantities are given in Table IV.1. Examination of the control functions reveal that the mean total binding energy is restricted to within about 3 kcal/mole of -1980 kcal/mole for the final 500 K moves of the simulation and attains a final value of -1982.2 +/- 6.2 kcal/mole after 2000 K. A final heat capacity value of 19.1 cal/mole-K is obtained. The calculated vacuum to water transfer energy is -127.4 kcal/mole.

Figures IV.7.4 thru IV.7.9 and IV.7.12 thru IV.7.17 contain distribution functions for the quasicomponents of coordination number and average first shell pair energies for the imino, amino, carbonyl and methine functional groups and bare ring nitrogen of cytosine. Figure IV.7.4 and IV.7.18 contain the quasicomponent distribution function of first shell coordination number and binding energy for cytosine.

Structural Results. The coordination numbers of the atoms of cytosine are displayed on the molecular diagram of Figure IV.7.3. The imino hydrogen, H1, significantly hydrated as usual, assigned 2.76 waters in a 3.2 Angstrom

shell. The solvent density is very high (1.37 x bulk water) reflecting strong solute-water associations and concentrated first shells. The amino hydrogens have tight first shells (2.4 and 2.2 Angstroms for H4' and H4'' respectively) each containing about one water (H4': 1.07 waters; H4'': .83). The carbonyl O2 oxygen and the N3 ring nitrogen are both significantly hydrated. The carbonyl oxygen contains 3.48 waters in a somewhat diffuse first shell (of density .92 x bulk water) more typical of apolar hydration than of polar hydration. The first shell population of the bare ring nitrogen N3 is 2.07. This atom has a larger degree of access to solvent than other ring atoms and consequently makes the largest contribution to ring constituent / pi cloud hydration in cytosine.

Finally, the methine H5 and H6 hydrogens are both fully hydrated (2.46 and 3.77 waters respectively). The first shell solvent densities are lower than (H5: .90) or equal to (H6: 1.01) water bulk density; these results are characteristic of apolar hydration.

The quasicomponent distribution function of coordination number for cytosine is shown in Figure IV.7.10. The values range from 16 to 23; the maximumm contribution is from a coordination number of 19. The average coordination number for cytosine is 18.83 +/- .31.

Energetic Results. The average first shell solute-water pair energies of the waters assigned to the atoms of cytosine are displayed on the molecular diagram of Figure IV.7.11. The waters of the H1 imino hydrogen are the most strongly bound of the cytosine hydration shell at an average of -6.215 kcal/mole. The solute binding energy contribution of these waters is also the greatest overall at -17.170 kcal/mole. Average pair energies for the waters of the imino hydrogens are nearly as favorable (H4'= -6.055 kcal/mole; H4''= -5.688 kcal/mole).

Rather low average pair energies are seen with the waters assigned to the methine hydrogens (H5= -1.619 kcal/mole; H6= -2.417 kcal/mole). Thus, the relatively weak solute-water pair energies indicative of apolar hydration are found here.

The quasicomponent distribution function of solute-water binding energy for cytosine is displayed in Figure IV.7.18. The values range from -126 to -93 kcal/mole; the most significant contributions arise between -116 and -105 kcal/mole. The average solute binding energy for cytosine is -108.1 +/- 1.4 kcal/mole.

Figure IV.7.1 - isoenergy contour surface for cytosine and one water (Clementi solute-water potential functions).

X axis - X axis of plane defined by molecular ring atoms (Angstroms).

Y axis - Y axis of plane defined by molecular ring atoms (Angstroms).

Contours - isoenergy contours with one kcal/mole increments with alphabetical labels referring to contour energy values in list at right.

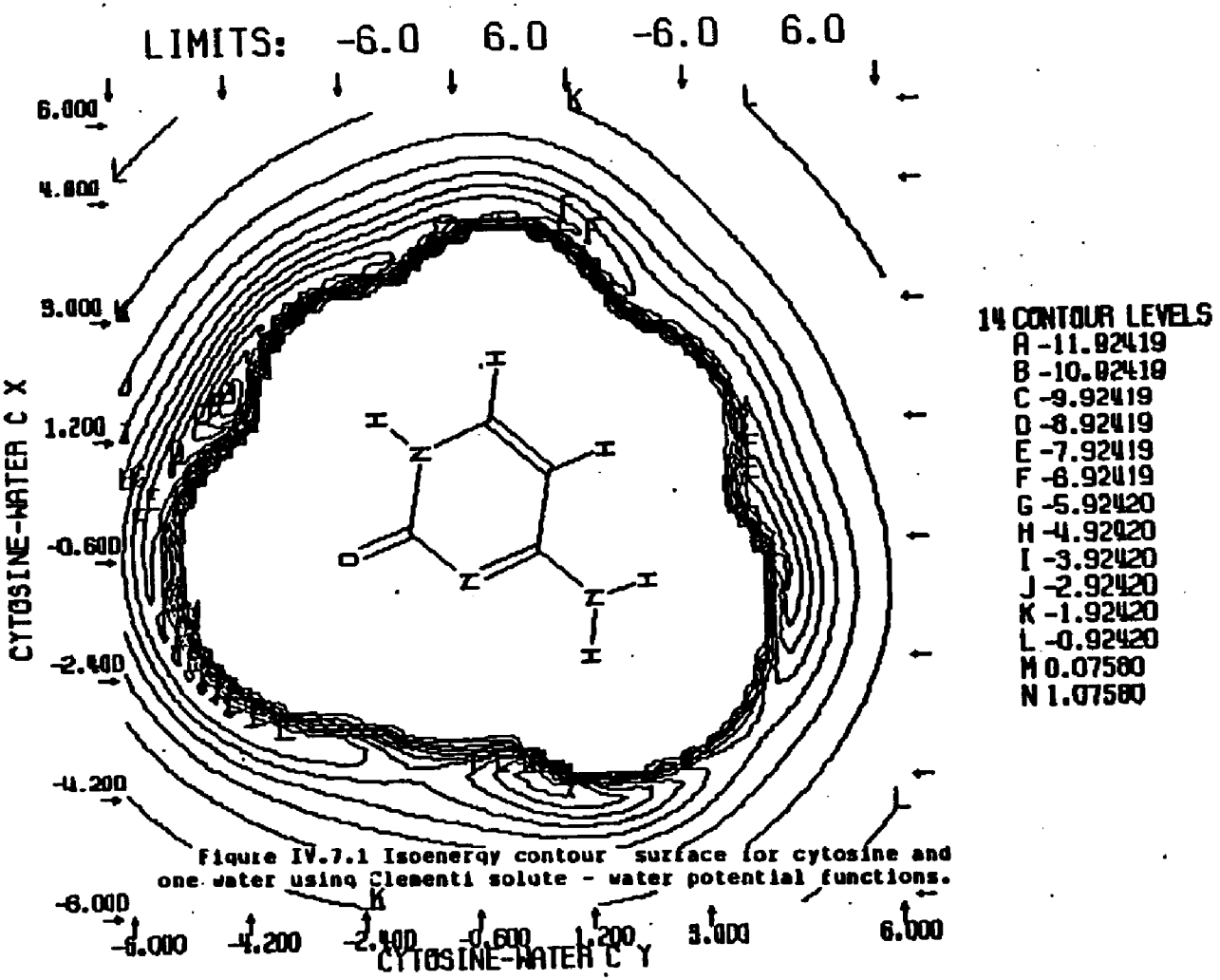


Figure IV.7.1 Isoenergy contour surface for cytosine and one water using Clementi solute - water potential functions.

Figure IV.7.2 - control functions for Monte Carlo simulation of cytosine and 215 waters (Clementi solute-water potential functions).

X axis - number of configurations.

Left Y axis - mean total energy (kcal/mole).

Right Y axis. - constant volume heat capacity (cal/mole-degree).

Upper curve - constant volume heat capacity.

Bottom curve without crosshatches - average total energy for entire simulation.

Bottom curve with crosshatches - average total energy for preceeding 50K configurations.

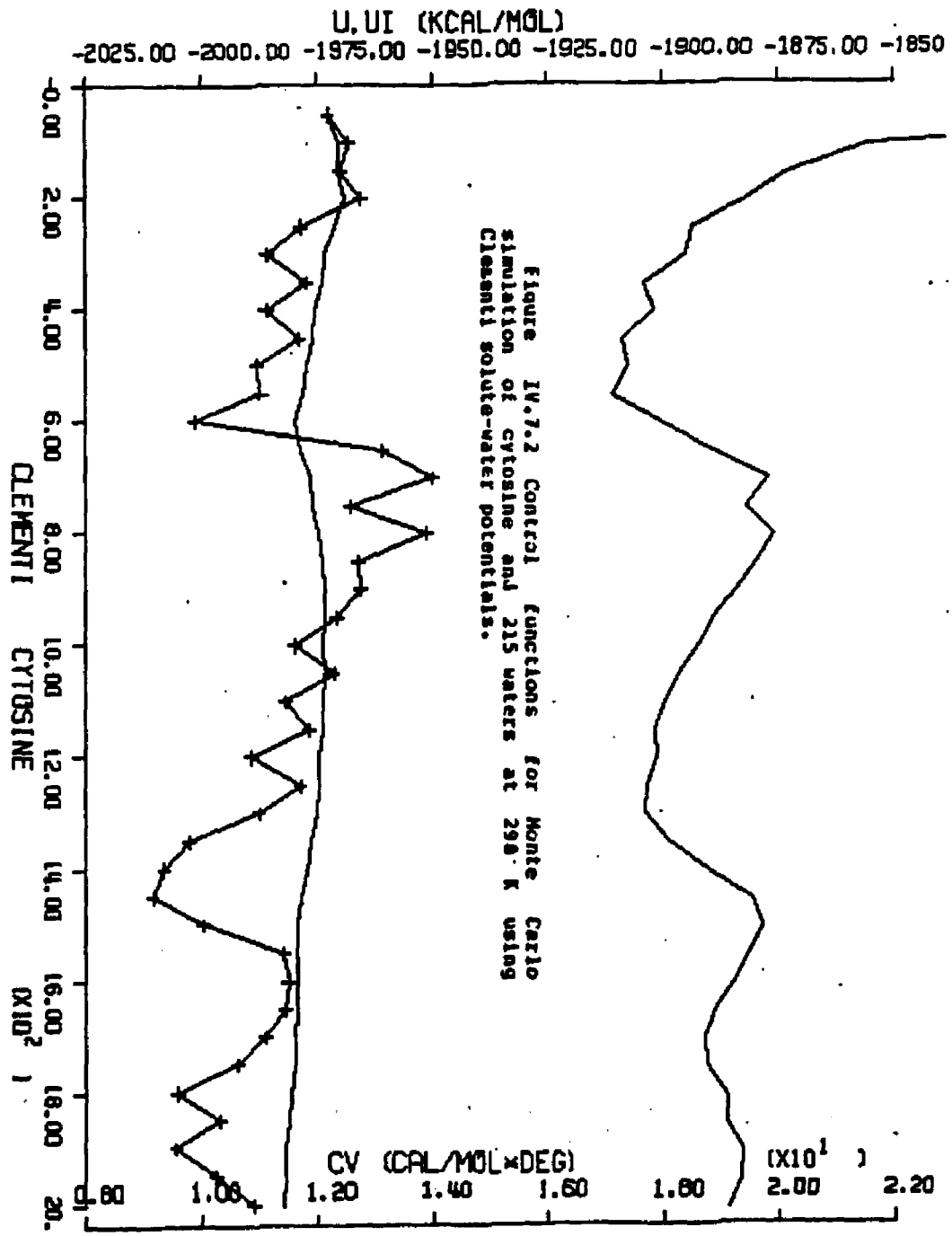


Figure IV.7.2 Control functions for Monte Carlo simulation of cytosine and 215 waters at 298.1 K using Clementi solute-water potentials.

CYTOSINE IN WATER AT 298K - FORCE BIAS AND CLEMENTI POTENTIAL FUNCTIONS

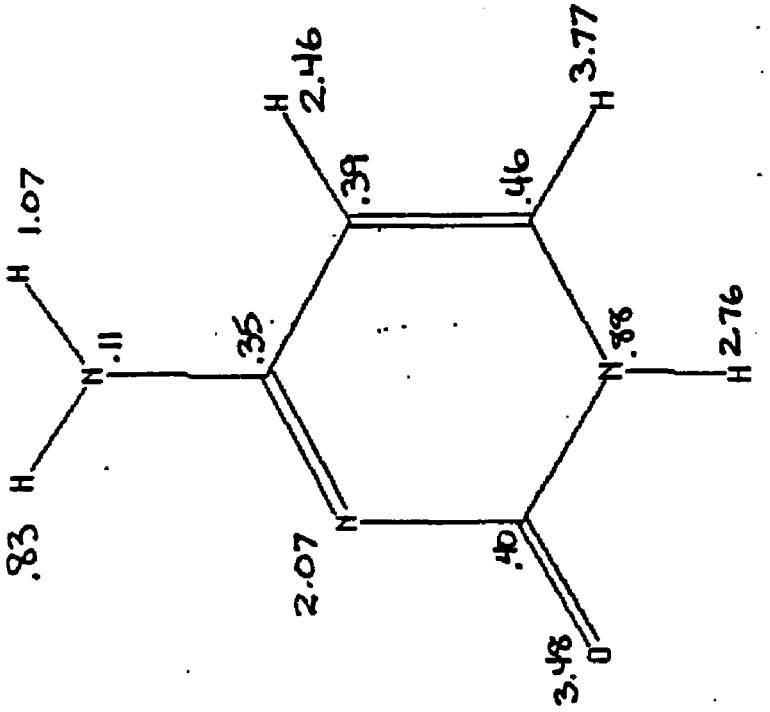
LAST CONFIGURATION: 2000001					FIRST SHELL SOLUTE PROPERTIES				TOTAL SLT PROPS		WATER PROPERTIES RFSW=3.30 RCB= 7.75 A		
INDEX	TYPE				NFS	VFS	<E>	<E/V>	<SLTSE>	<SLTPE>	<E>	<SLTSE>	<DD>	<SSSHPE>	<SSMAT>
METHYLENE (DISUBST.)															
1	6	24	C4		4.8	18.64	0.35	0.56	-0.458	-1.388	1.29	-0.946	4.69	-2.783	-18.618
STATISTICAL UNCERTAINTY (+/- 2*SD):							0.04	0.07	0.082	0.244	0.08	0.158	0.17	0.228	0.871
METHYLENE (MONOSUB.)															
2	7	25	C5		4.2	17.72	0.39	0.66	-0.882	-0.211	1.11	-0.261	4.65	-2.944	-19.060
3	11	16	H5		3.8	81.86	2.46	0.98	-3.986	-1.619	19.65	-7.937	4.24	-3.089	-17.548
TOTALS FOR FUNCTIONAL GROUP				=CH- :		99.58	2.85	0.86	-4.868	-1.427	20.75	-8.198	4.26	-3.085	-17.652
4	8	25	C6		3.8	15.79	0.24	0.46	-0.179	-0.732	0.94	-0.328	4.98	-2.999	-21.034
5	12	16	H6		4.8	111.21	3.77	1.01	-9.078	-2.487	36.59	-16.258	4.22	-3.081	-17.466
TOTALS FOR FUNCTIONAL GROUP				=CH- :		127.00	4.02	0.95	-9.257	-2.385	37.53	-16.585	4.23	-3.081	-17.542
AVERAGES OVER FUNCTIONAL GRP				=CH- :		113.29	3.43	0.91	-6.663	-1.866	29.14	-12.392	4.25	-3.083	-17.597
STATISTICAL UNCERTAINTY (+/- 2*SD):							0.09	0.02	0.274	0.079	0.01	0.389	0.02	0.036	0.122
AMIDE GROUP															
6	2	15	N1		3.4	13.46	0.88	1.96	-2.623	-2.971	1.67	-3.197	4.29	-3.085	-18.823
7	13	16	H1		3.2	68.32	2.76	1.37	-17.178	-6.215	31.38	-25.529	4.24	-2.978	-17.614
TOTALS FOR FUNCTIONAL GROUP				>NH :		73.77	3.65	1.48	-19.793	-5.429	32.97	-28.726	4.25	-2.976	-17.674
STATISTICAL UNCERTAINTY (+/- 2*SD):							0.14	0.05	1.118	0.314	0.01	0.951	0.03	0.048	0.164
AMINE GROUP															
8	4	11	N4'		3.8	12.38	0.11	0.27	-0.893	-0.834	0.85	-0.451	4.47	-2.881	-17.813
9	9	1	H4''		2.4	32.88	1.07	1.80	-6.072	-5.688	37.77	-15.830	4.25	-3.080	-17.473
10	10	1	H4'		2.2	23.90	0.83	1.84	-5.050	-6.055	30.38	-11.565	4.19	-3.051	-17.486
TOTALS FOR FUNCTIONAL GROUP				-NH2 :		68.28	2.01	0.88	-11.216	-5.578	68.91	-27.846	4.22	-3.020	-17.483
STATISTICAL UNCERTAINTY (+/- 2*SD):							0.10	0.04	0.052	0.434	0.02	0.630	0.02	0.034	0.112

Table IV.8 Calculated structural and energetic quantities from Monte Carlo simulation of cytosine and 215 waters at 298 K - Clementi solute-water potential functions.

SUBSTITUTED IMINO GR												
11	3	12	N 3	4.8	58.98	2.87	1.22	-4.582	-2.178	5.93	-6.314	4.25	-2.979	-17.891
STATISTICAL UNCERTAINTY (+/- 2*SD):						0.10	0.06	0.337	0.167	0.01	0.493	0.07	0.113	0.373
CARBONYL GROUP														
12	1	27	O 2	3.8	112.58	3.48	0.92	-10.374	-2.982	46.38	-19.983	4.18	-3.029	-17.314
13	5	26	C 2	3.8	16.77	0.40	0.71	-0.959	-2.393	1.24	-1.235	4.28	-3.028	-17.956
TOTALS FOR FUNCTIONAL GROUP C=O :					129.35	3.88	0.90	-11.333	-2.921	47.62	-21.218	4.18	-3.029	-17.330
STATISTICAL UNCERTAINTY (+/- 2*SD):						0.14	0.03	0.628	0.164	0.01	0.584	0.02	0.041	0.133
MOLECULAR SUM/AVERAGE:					567.6	18.83	0.99	-60.619	-3.228	215.00	-109.033	4.23	-3.007	-17.500
STATISTICAL UNCERTAINTY (+/- 2*SD):						0.31	0.02	1.507	0.082	0.03	1.424	0.01	0.019	0.063

Table IV.8 Calculated structural and energetic quantities from Monte Carlo simulation of cytosine and 215 waters at 298 K - Clementi solute-water potential functions.

Figure IV.7.3 First shell coordination numbers for the atoms of cytosine - Clementi solute-water potential functions.



SCALE FACTOR = 1.89

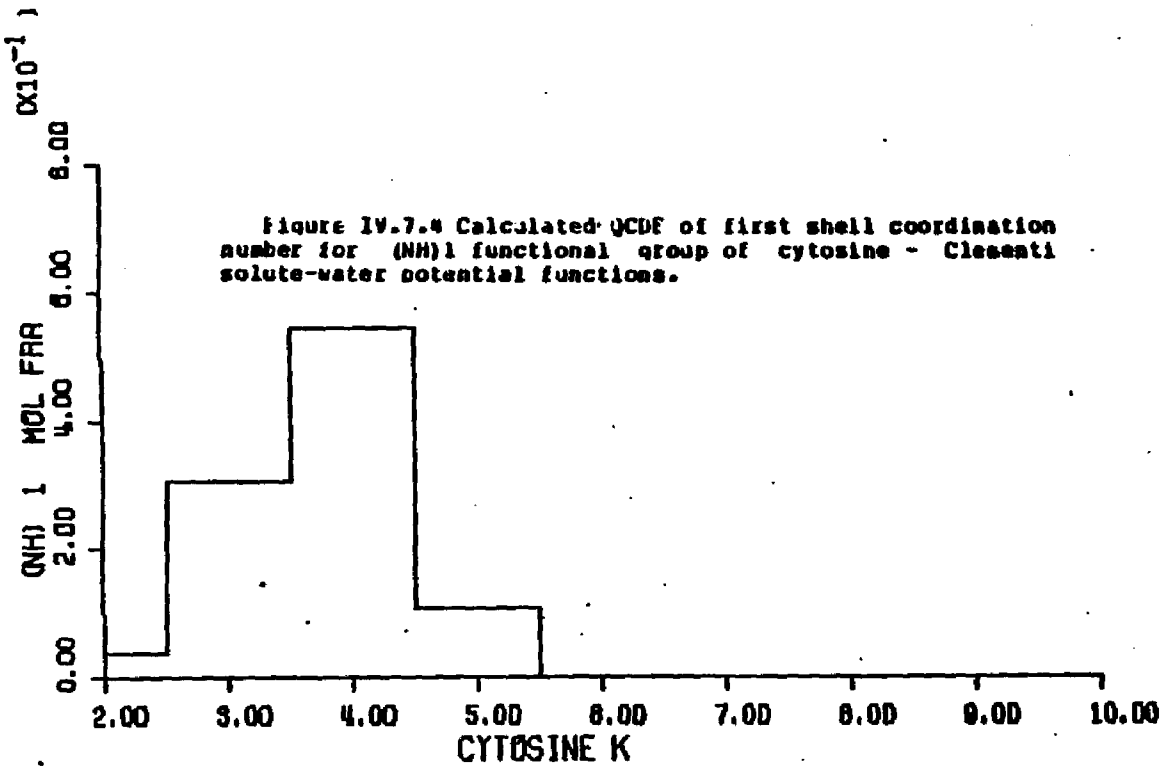


Figure IV.7.4 Calculated QCDF of first shell coordination number for (NH)1 functional group of cytosine - Clementi solute-water potential functions.

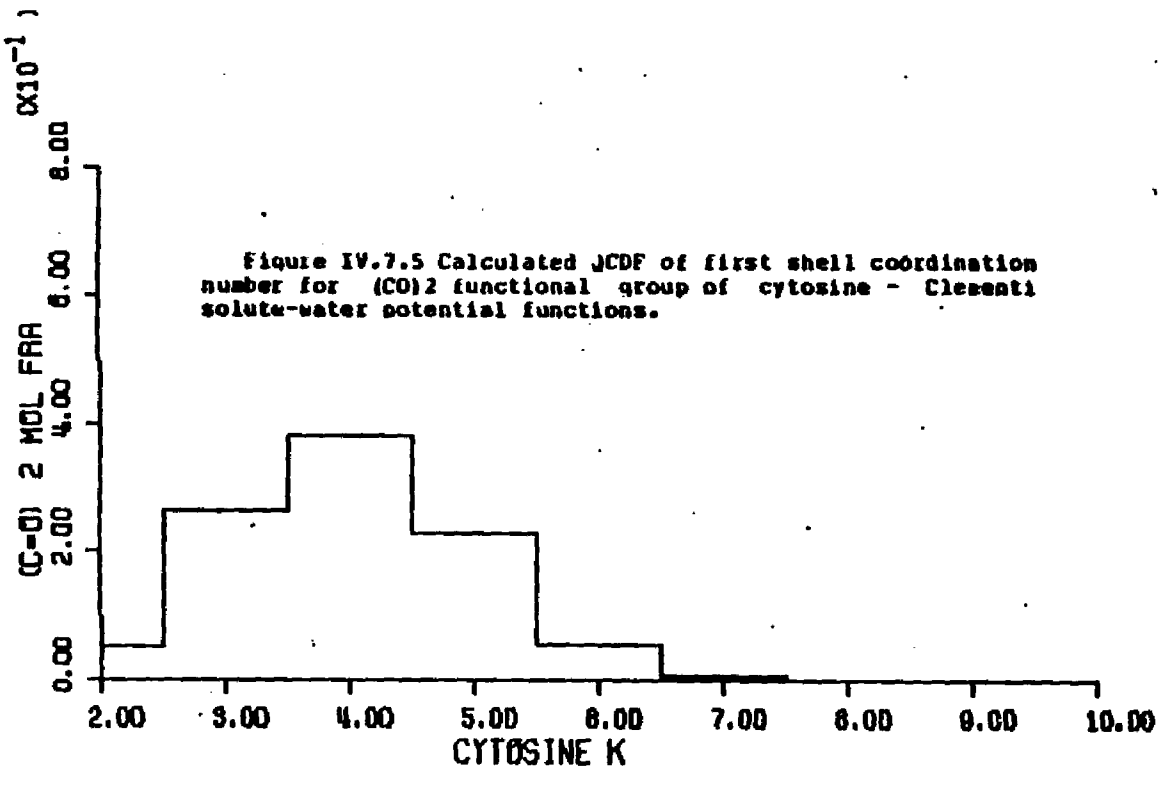


Figure IV.7.5 Calculated JCDF of first shell coordination number for (CO)2 functional group of cytosine - Clementi solute-water potential functions.

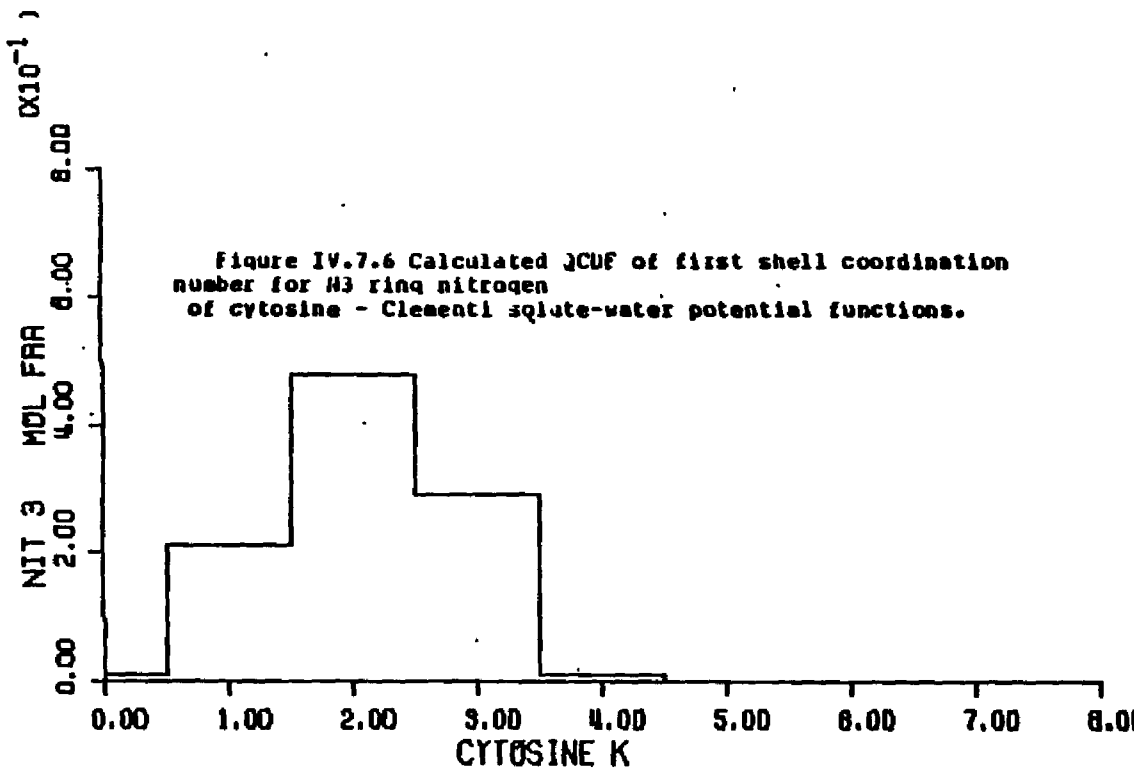


Figure IV.7.6 Calculated JCUF of first shell coordination number for N3 ring nitrogen of cytosine - Clementi solute-water potential functions.

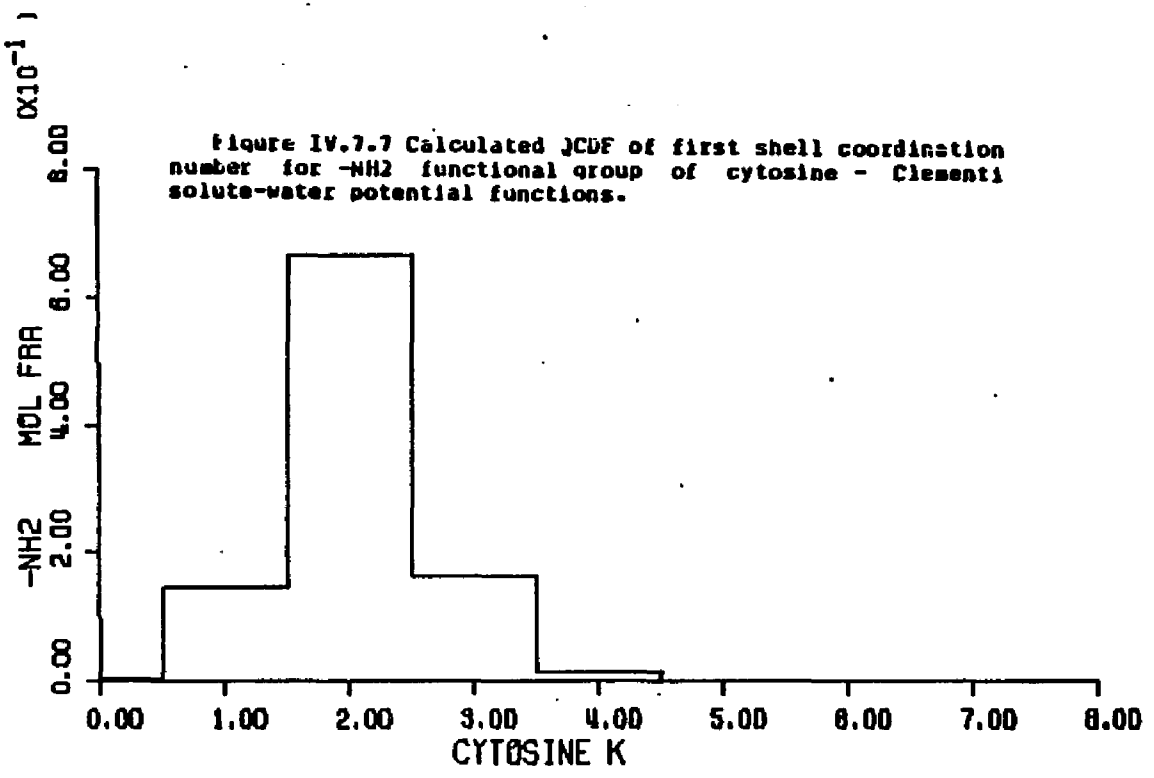


Figure IV.7.7 Calculated JCDF of first shell coordination number for -NH2 functional group of cytosine - Clementi solute-water potential functions.

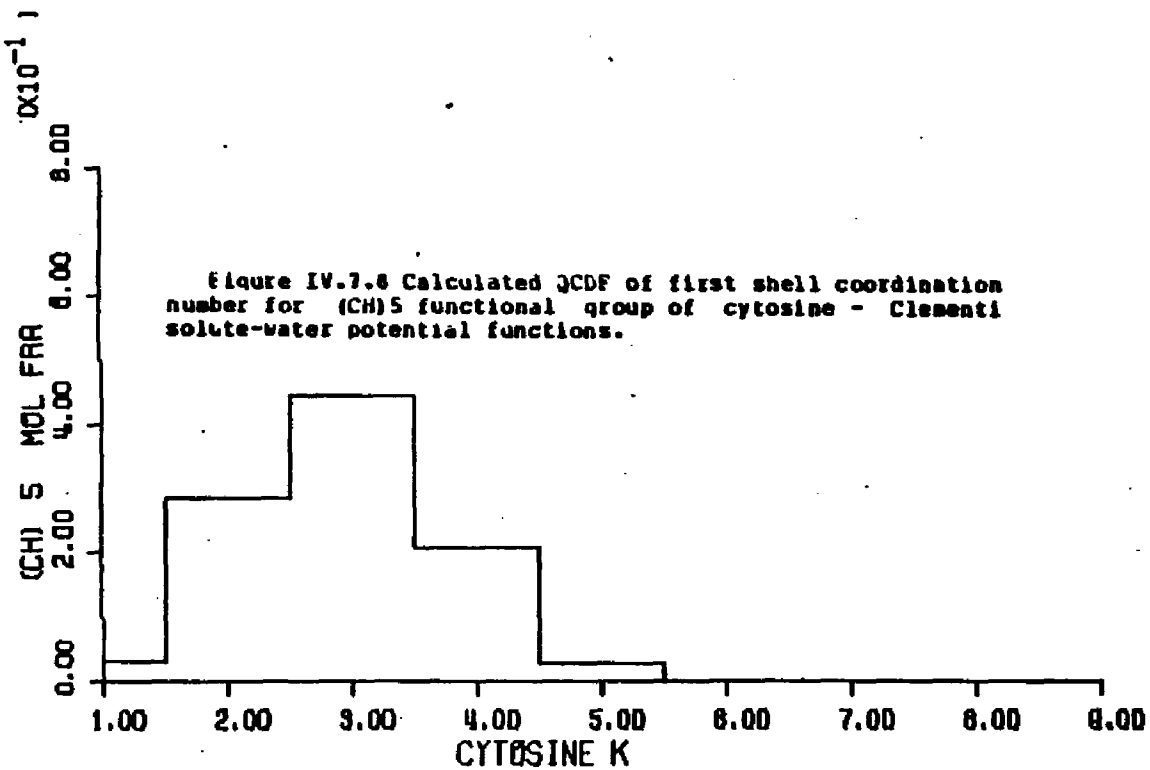


Figure IV.7.6 Calculated RDF of first shell coordination number for (CH)5 functional group of cytosine - Clementi solute-water potential functions.

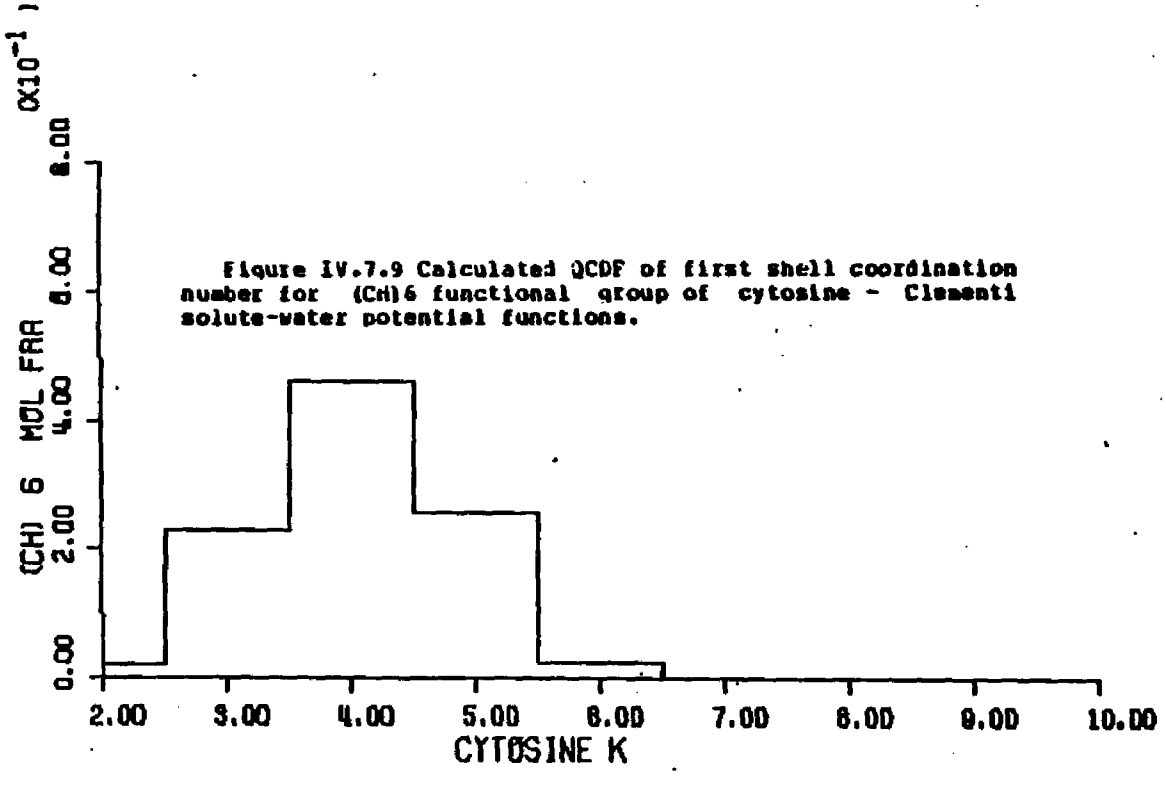


Figure IV.7.9 Calculated QCDF of first shell coordination number for (CH)6 functional group of cytosine - Clementi solute-water potential functions.

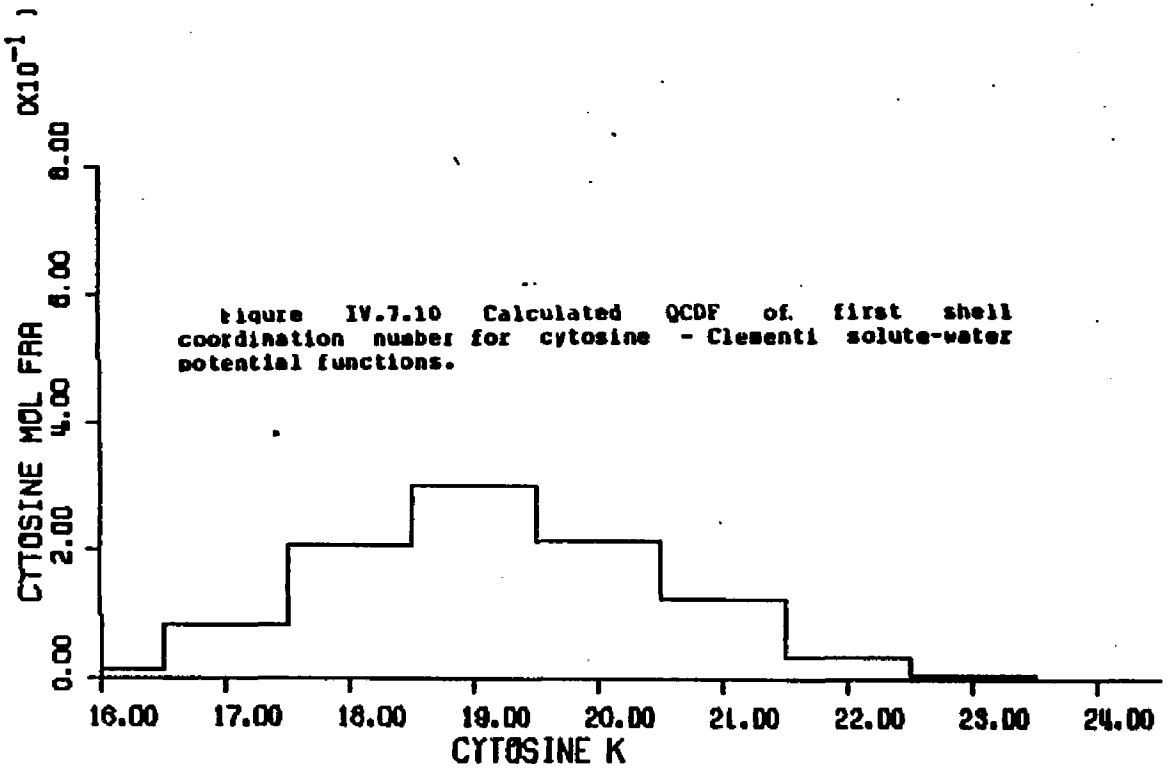


Figure IV.7.10 Calculated QCDF of first shell coordination number for cytosine - Clementi solute-water potential functions.

Figure IV.7.11 Average first shell solute-water pair energies of waters assigned to the atoms of cytosine - Clementi solute-water potential function simulation.

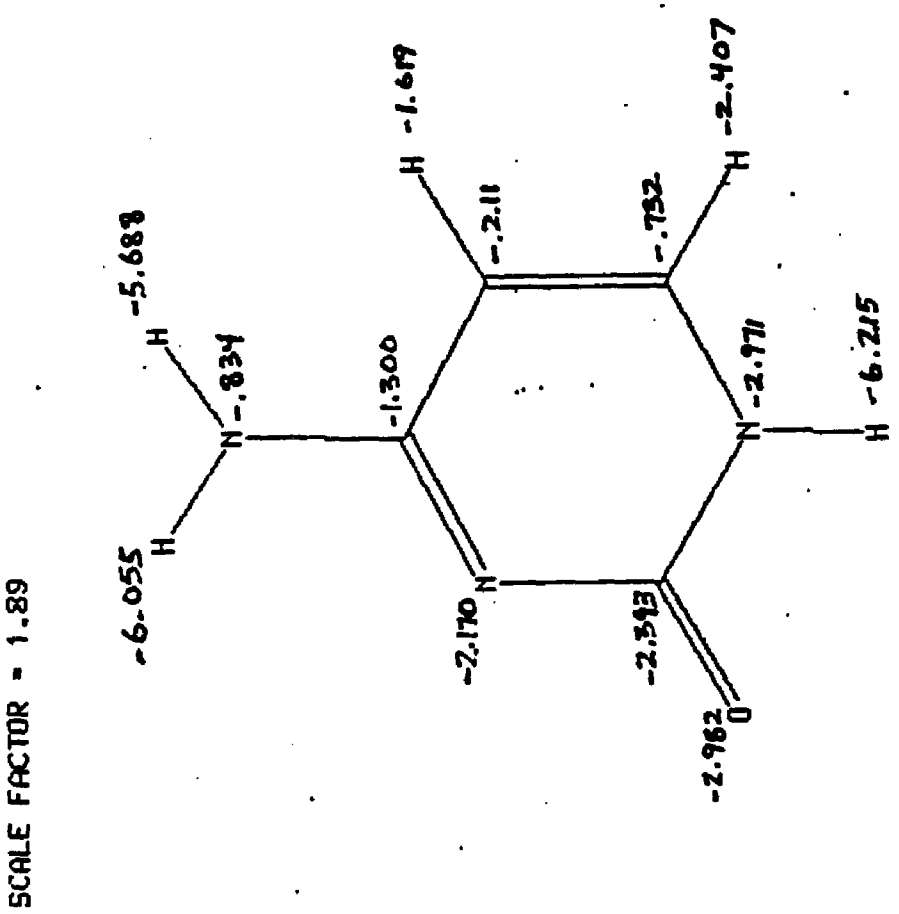


Figure IV.7.12 Calculated QCDF of solute-water pair energies of waters of the (NH)1 functional group of cytosine - Clementi solute-water potential functions.

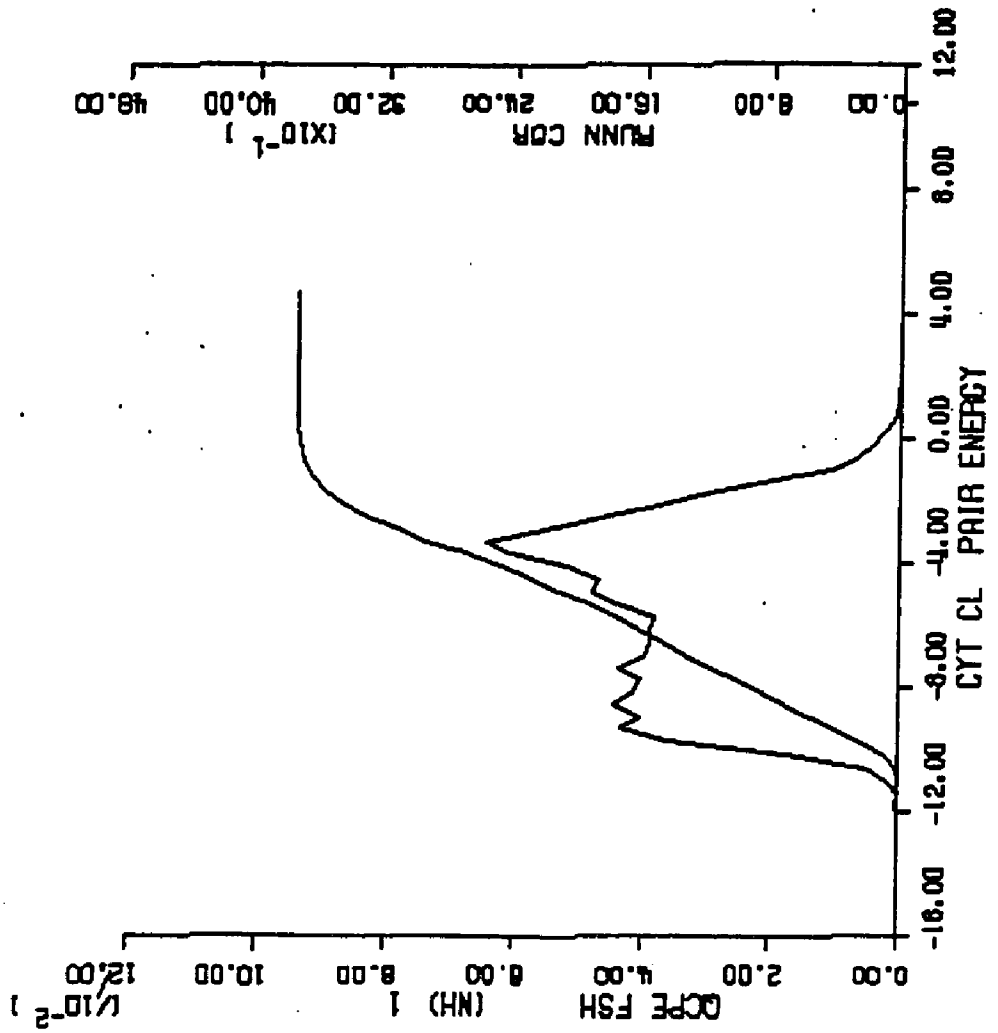


Figure IV-7.13 Calculated QCPE of solute-water pair energies of waters of the (C)2 functional group of cytosine - Clementi solute-water potential functions.

Figure IV.7.14 Calculated CDF of solute-water pair energies of waters of the N3 ring nitrogen of cytosine - Clementi solute-water potential functions.

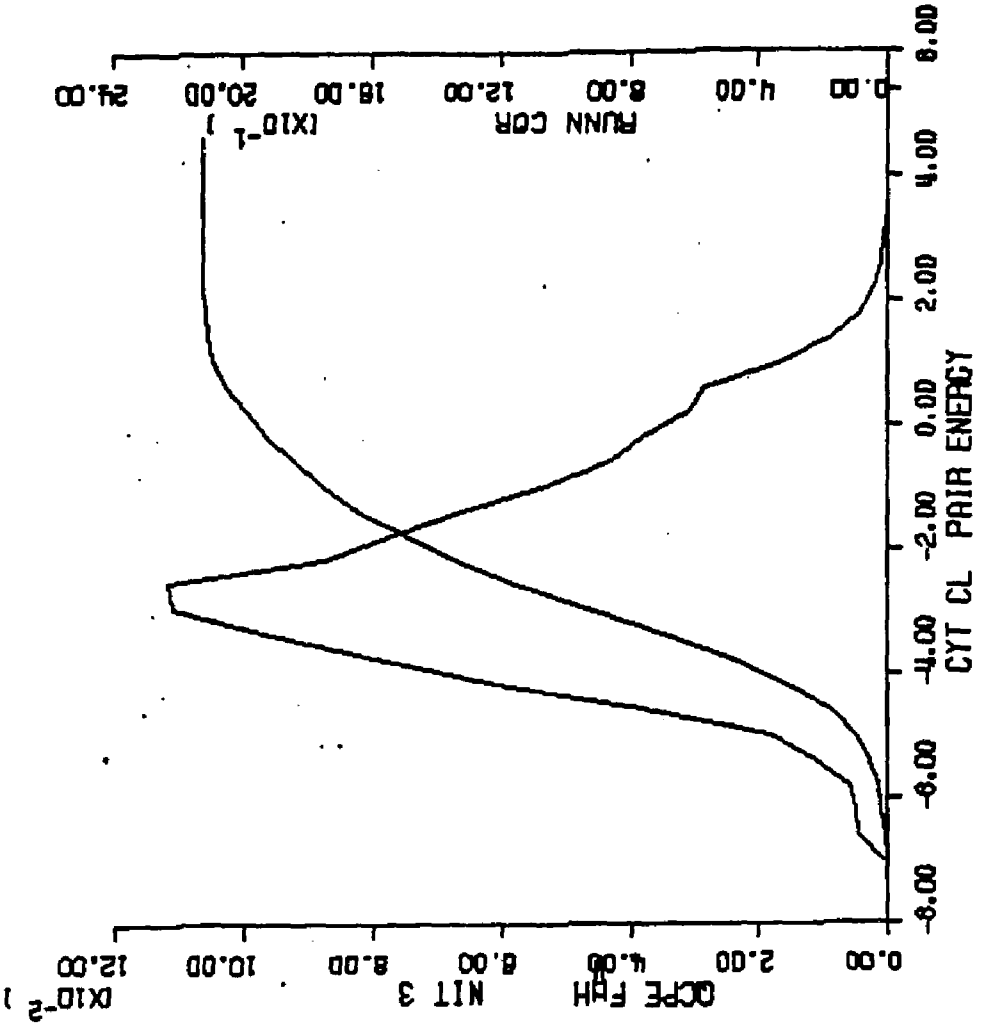


Figure IV.1.15 Calculated VCOF of solute-water pair energies of water molecules of the -NH2 functional group of cytosine - Chemical solute-water potential functions.

Figure IV.7.16 Calculated QCDF of solute-water pair energies of waters of the (CH)5 functional group of cytosine - Clementi solute-water potential functions.

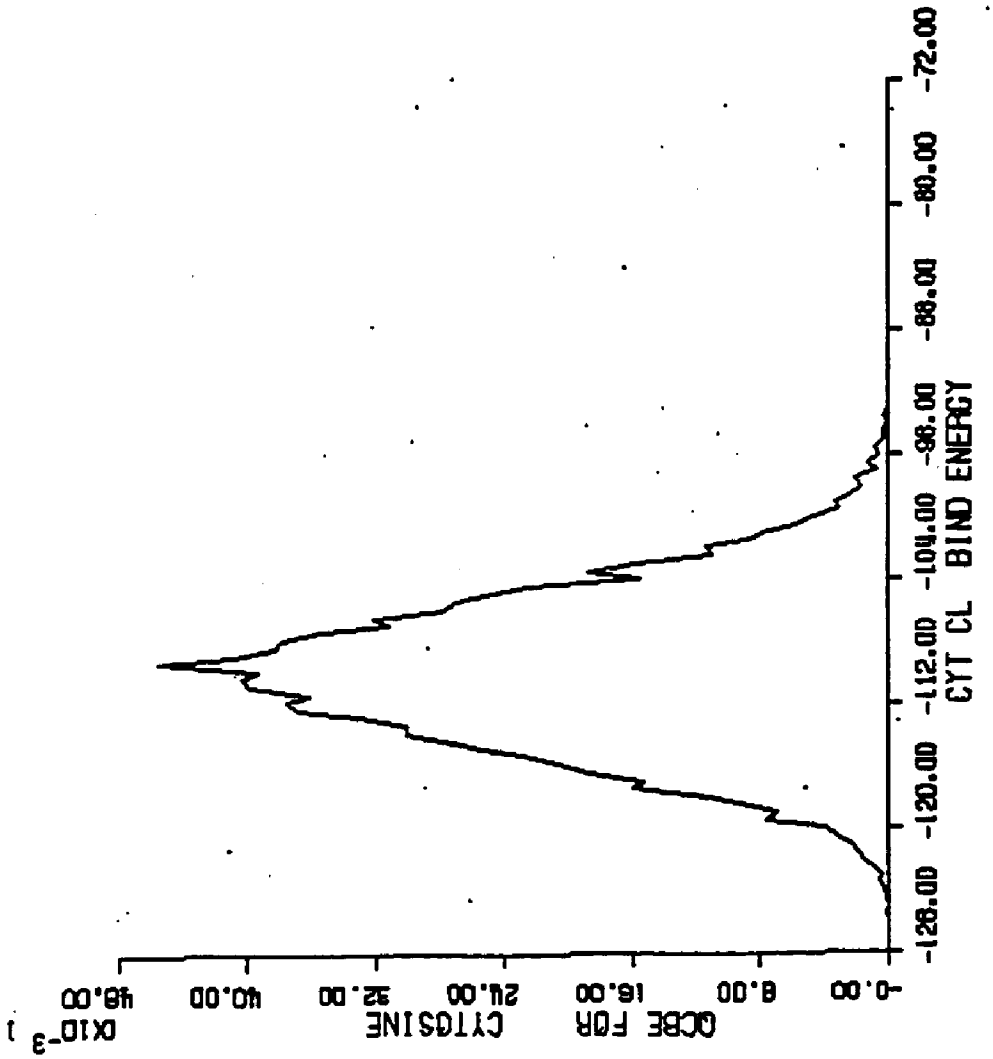
Figure IV.7.17 Calculated QCDF of solute-water pair energies of waters of the (CH)6 functional group of cytosine - Clementi solute-water potential functions.

Figure IV.7.18 - calculated quasicomponent distribution function of binding energy for cytosine (Clementi potentials).

X axis - binding energy (Kcal/mole).

Y axis - quasicomponent of binding energy.

Figure IV.7.18 Calculated QCDF of total binding energy for cytosine - Clementi solute-water potential functions.



8. Adenine with Clementi Solute-Water Potentials

Computational Specifics. A solution of adenine and 215 waters was simulated using Clementi solute-water potential functions and MCY-CI water-water potential functions at 298 K. Simple cubic boundary conditions were used with a unit cell edge length of 18.75 Angstroms. These dimensions were determined using a partial molar volume of 85 cc/mole for adenine and 18.07 cc/mole for water. The initial 600,000 steps of Monte Carlo simulation were discarded as an equilibration period. Ensemble averages and other analyses were performed using the succeeding 2000K configurations of the simulation.

Potential Surface. The potential surface for the interaction of adenine and one water molecule is shown in Figure IV.8.1. The global minimum is found near the H9 imino nitrogen at -10.24 kcal/mole. Another deep minimum occurs near the H2'' amino hydrogen at an energy somewhat lower than -8.24 kcal/mole; the other amino nitrogen, H2', has a shallower minimum lying between -6.24 and -8.24 kcal/mole. Potential energies in the region of the bare ring nitrogens, N1, N3 and N7 are only moderately favorable, falling between -2.24 and -4.24 kcal/mole in all three cases. Interactions at the supposedly apolar H2 and H8 methine positions are more favorable. The energies in the H2 region range between -4.24 and -6.24 kcal/mole; near the H8 hydrogen, a small region is found in the even more favorable -6.24 to -8.24 kcal/mole

range.

Convergence and Thermodynamic Results. The control functions for the adenine simulation are given in Figure IV.8.2. and the calculated thermodynamic quantities are given in Table IV.1. The total binding energy remains within about 5 kcal/mole of -2010 kcal/mole for the last 500K of the simulation. The mean total binding energy after 2000 K moves is -2001.2 +/- 11.9 kcal/mole; the heat capacity reaches a final value of 25.5 cal/mole-K. The vacuum to water transfer energy is calculated to be -141.4 kcal/mole.

Figures IV.8.4 through IV.8.10 and IV.8.13 thru IV.8.19 contain distribution functions for the quasicomponents of coordination number and first shell pair energy for the amino, imino and methine functional groups and base ring nitrogen atoms of adenine. Figures IV.8.11 and IV.8.20 contain the quasicomponents of coordination number and binding energy for the adenine molecule.

Structural Results. The atomic first shell coordination numbers for adenine are displayed on the molecular diagram of Figure IV.8.3. The imino H9 atom has a fully populated first shell (3.61 waters) limited by a 3.4 Angstrom cutoff and is thus similar to the imino hydrogens previously discussed. The first shell solvent density (1.34 times bulk water) is the highest in the

adenine coordination shell and clearly reflects the hydrophilic hydration of this atom. The amino hydrogens, H6' and H6'' each have very small first shell cutoffs of 2.2 Angstroms. The first shell of H6' contains .81 waters while the first shell of H6'' contains .90 waters.

The base ring nitrogens have significant first shell populations. The atom N1 is assigned 2.21 waters, N3 is assigned 1.14 waters and N7 is assigned 1.91 waters. These relatively apolar ring atoms have both in plane and out of plane access to solvent but do not have the degree of solvent exposure of a ring substituent.

The methine hydrogens, H2 and H8, are assigned identical 3.6 Angstrom cutoffs resulting in first shell populations of 3.81 and 3.35 waters respectively. The first shell solvent densities are somewhat above bulk water (1.12 and 1.03 times bulk water density, respectively).

The quasicomponent distribution function of coordination number for adenine is displayed in Figure IV.8.11. The contributions fall in the range 17 to 24; the most common coordination number is 21. The average first shell coordination number is 20.40 +/- .32.

Energetic Results. Average first shell pair energies for the waters assigned to each of the atoms of adenine are shown on the molecular diagram of Figure IV.8.12.

The waters of the amino hydrogens H6' and H6'' interact most favorably with adenine (H6'':<SLIPE>=-5.662 kcal/mole; H6':<SLIPE>=-5.026 kcal/mole). The waters assigned to the imino H9 atom associate nearly as favorably with the solute at an average -4.792 kcal/mole each. The large first shell population of H9, however, results in the greatest total solute binding energy contribution of all waters in the adenine coordination shell (<SLIBE>=-17.298 kcal/mole). The waters of the methine hydrogens, H2 and H8, interact relatively weakly with the solute (H2:<SLIPE>=-1.304 kcal/mole;H8:<SLIPE>=-1.996 kcal/mole). These low interaction energies are expected for apolar atoms. The average water-water pair energies, however, are destabilized in this region relative to water-water pair energies in other parts of the extended adenine hydration shell; this result is not expected for apolar atoms. Average solute-water pair energies for the waters of the apolar bare ring nitrogens, N1, N3 and N9, are comparable to those of the methine waters (-1.855, -1.817 and -1.300 kcal/mole respectively).

The quasicomponent distribution function of solute-waters binding energy for adenine is displayed in Figure IV.8.20. Energy values ranging from -125 to -110 kcal/mole are encountered; the most significant contributions are in the -118 to -108 kcal/mole range. The average solute binding energy is -116.9 +/- 1.6

kcal/mole.

Figure IV.8.1 - isoenergy contour surface for adenine and one water (Clementi solute-water potential functions).

X axis - X axis of plane defined by molecular ring atoms (Angstroms).

Y axis - Y axis of plane defined by molecular ring atoms (Angstroms).

Contours - isoenergy contours with one kcal/mole increments with alphabetical labels referring to contour energy values in list at right.

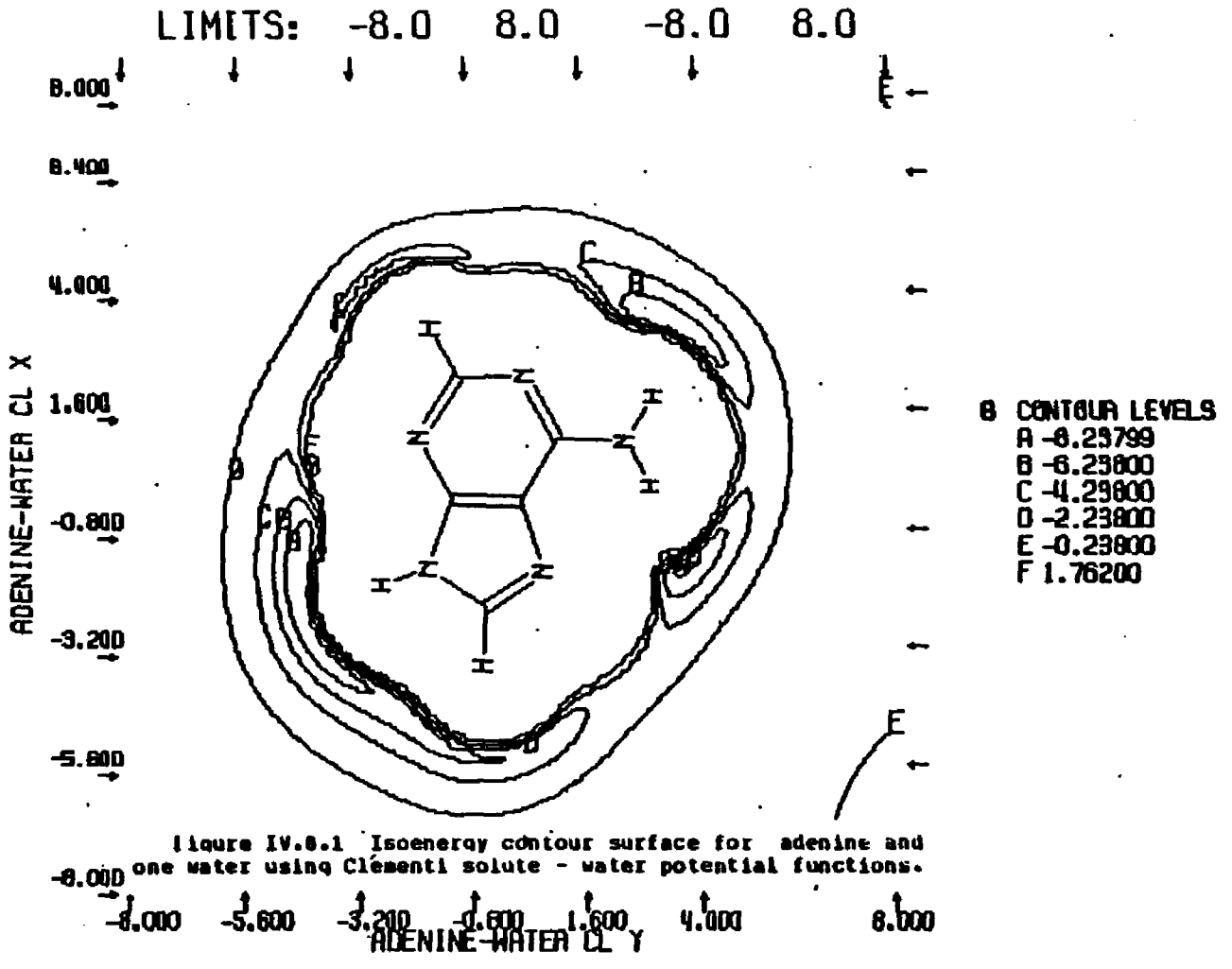


Figure IV.8.1 Isoenergy contour surface for adenine and one water using Clementi solute - water potential functions.

Figure IV.8.2 - control functions for Monte Carlo simulation of adenine and 215 waters (Clementi solute-water potential functions).

X axis - number of configurations.

Left Y axis - mean total energy (Kcal/mole).

Right Y axis. - constant volume heat capacity (cal/mole-degree).

Upper curve - constant volume heat capacity.

Bottom curve without crosshatches - average total energy for entire simulation.

Bottom curve with crosshatches - average total energy for preceeding 50K configurations.

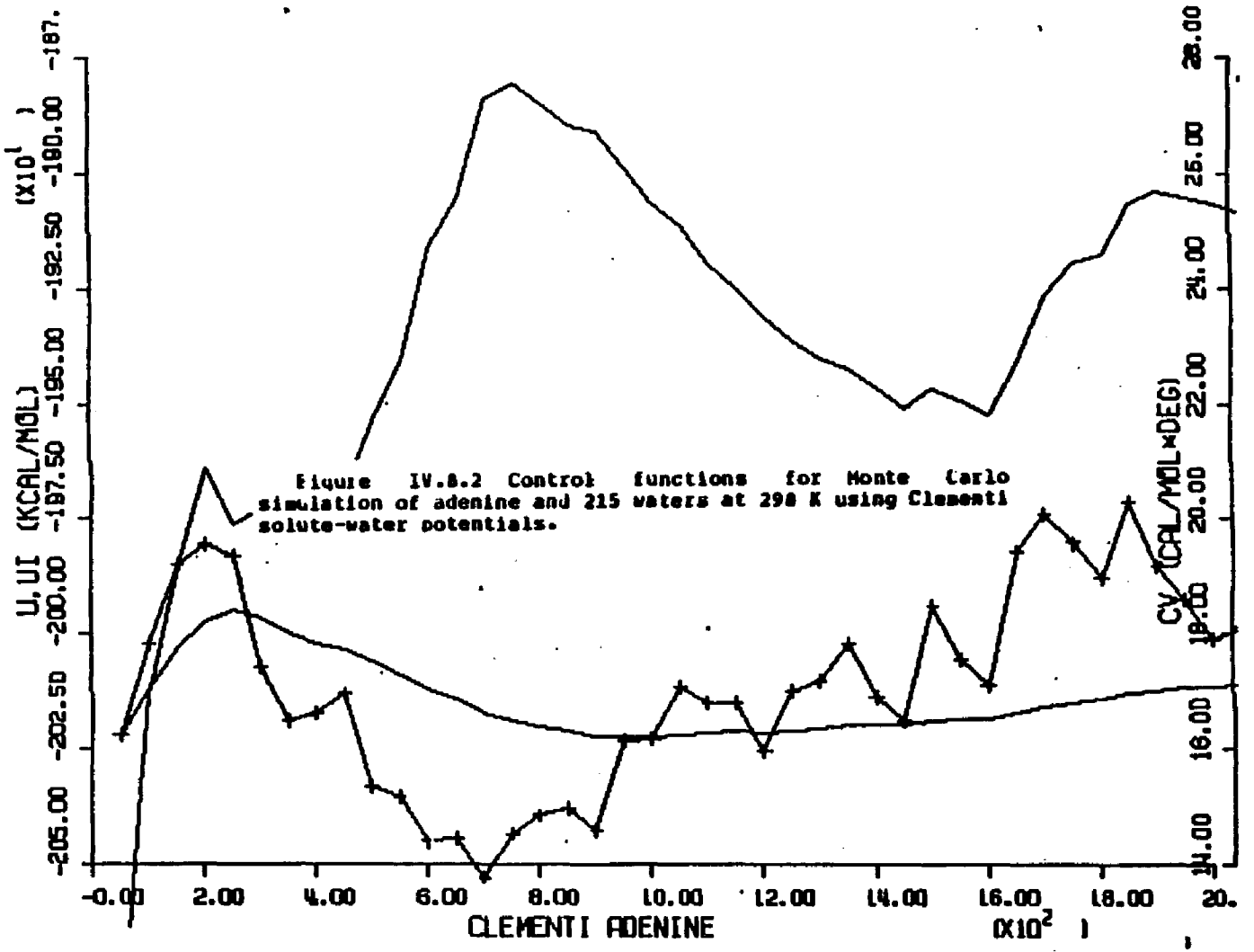
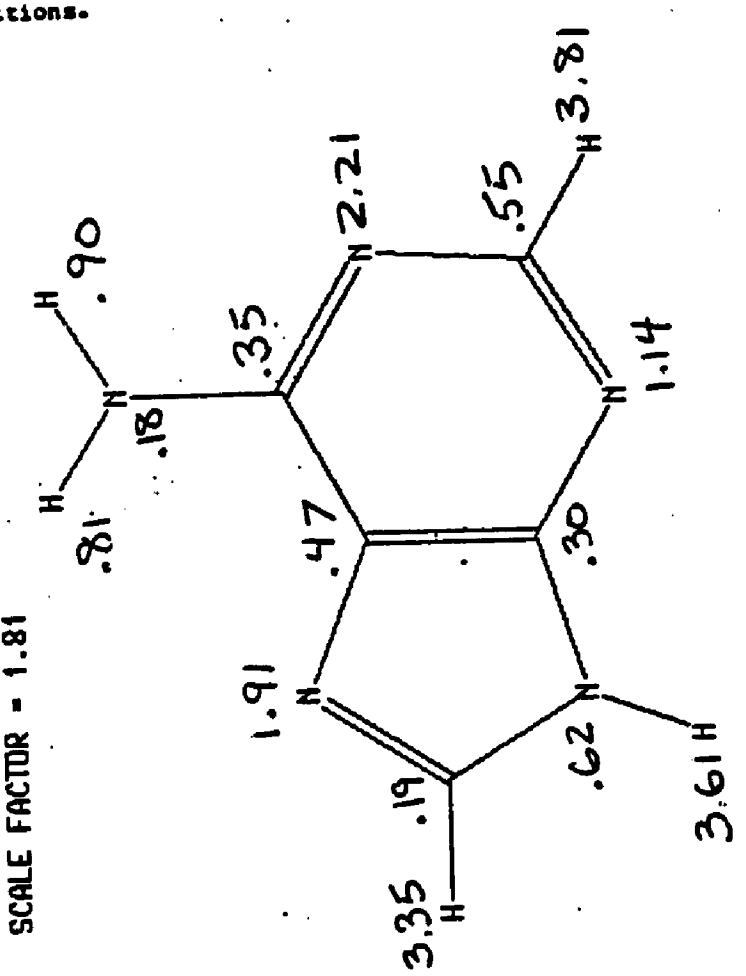


Figure IV.8.2 Control functions for Monte Carlo simulation of adenine and 215 waters at 298 K using Clementi solute-water potentials.

Figure IV.a.3 First shell coordination numbers for the atoms of adenine - Clementi solute-water potential functions.



ADENINE IN WATER AT 298K - FORCE BIAS AND CLEMENTI POTS.

LAST CONFIGURATION: 2000001					FIRST SHELL SOLUTE PROPERTIES					TOTAL SLT PROPS		WATER PROPERTIES RFSH=3.30 RCB= 7.75 A		
INDEX TYPE					NFS	VFS	<E>	<E/V>	<ELTSE>	<ELTPE>	<E>	<ELTSE>	<RH>	<RHAMPS>	<RHAMT>
METHYLENE (DISUBST.)															
1	6	24	C		4.8	22.34	0.35	0.47	-0.483	-1.142	1.31	-0.628	4.15	-3.151	-18.187
2	7	19	C		4.2	20.26	0.47	0.69	-0.974	-2.083	1.47	-1.243	4.07	-3.069	-17.518
3	8	20	C		3.8	18.23	0.38	0.58	-0.948	-3.091	1.21	-1.233	4.12	-3.064	-17.365
AVERAGES OVER FUNCTIONAL GRP =C< :						20.28	0.38	0.55	-0.772	-2.105	1.33	-1.035	4.12	-3.095	-17.687
STATISTICAL UNCERTAINTY (+/- 2*SD):							0.03	0.04	0.098	0.221	0.0	0.089	0.06	0.168	0.000
METHYLENE (MONOSUB.)															
4	9	25	C		5.0	28.33	0.55	0.81	-0.974	-1.772	1.38	-1.283	4.21	-3.197	-17.869
5	11	16	H		3.6	101.34	3.81	1.12	-4.966	-1.304	44.59	-16.286	4.29	-3.028	-17.950
TOTALS FOR FUNCTIONAL GROUP =CH- :						121.67	4.36	1.07	-5.940	-1.363	45.89	-17.489	4.29	-3.033	-17.948
6	10	25	C		3.8	15.17	0.19	0.37	-0.568	-2.701	1.28	-0.888	4.09	-3.123	-17.314
7	13	16	H		3.6	97.75	3.35	1.03	-6.694	-1.996	41.51	-16.931	4.31	-2.966	-17.687
TOTALS FOR FUNCTIONAL GROUP =CH- :						112.91	3.54	0.94	-7.201	-2.033	42.71	-17.820	4.30	-2.970	-17.677
AVERAGES OVER FUNCTIONAL GRP =CH- :						117.29	3.95	1.01	-6.571	-1.698	44.30	-17.654	4.30	-3.001	-17.812
STATISTICAL UNCERTAINTY (+/- 2*SD):							0.18	0.03	0.314	0.067	0.0	0.323	0.01	0.035	0.171
AMIDE GROUP															
8	4	15	N		4.0	15.57	0.62	1.19	-2.312	-3.723	1.46	-2.567	4.11	-3.053	-17.886
9	12	16	H		3.4	78.97	3.61	1.37	-17.298	-4.792	38.73	-28.638	4.25	-3.047	-17.773
TOTALS FOR FUNCTIONAL GROUP >NH :						94.54	4.23	1.34	-19.610	-4.635	40.19	-31.205	4.24	-3.047	-17.745

Table IV.9 Calculated structural and energetic quantities from Monte Carlo simulation of adenine and 215 waters at 298 K - Clementi solute-water potential functions.

ADENINE IN WATER AT 298K - FORCE BIAS AND CLEMENTI POTENTIAL FUNCTIONS

AMINE GROUP													
10	5	11	N	4.4	14.38	0.18	0.37	-0.871	-0.399	0.73	-0.244	4.30	-3.015	-17.984
11	14	1	H	2.2	24.45	0.81	0.99	-4.884	-5.026	24.84	-11.841	4.17	-3.051	-17.529
12	15	1	H	2.2	26.15	0.90	1.03	-5.078	-5.662	35.24	-16.000	4.22	-3.034	-17.760
TOTALS FOR FUNCTIONAL GROUP -NH2 :					64.98	1.89	0.87	-9.233	-4.891	60.81	-28.085	4.28	-3.040	-17.669
STATISTICAL UNCERTAINTY (+/- 2*SD):						0.10	0.05	0.902	0.396	0.0	0.623	0.02	0.043	0.286
SUBSTITUTED IMINO GR														
13	1	12	N	4.6	65.02	2.21	1.02	-4.102	-1.855	7.16	-5.685	4.22	-3.089	-17.673
14	2	12	N	3.6	49.35	1.14	0.69	-2.064	-1.017	9.04	-5.165	4.27	-3.100	-17.825
15	3	12	N	4.0	52.00	1.91	1.00	-2.485	-1.300	6.01	-3.090	4.15	-3.066	-17.637
AVERAGES OVER FUNCTIONAL GRP -N- :					55.72	1.75	0.94	-2.884	-1.657	7.40	-4.914	4.21	-3.088	-17.712
STATISTICAL UNCERTAINTY (+/- 2*SD):						0.05	0.03	0.169	0.080	0.0	0.179	0.03	0.071	0.339
MOLECULAR SUM/AVERAGE:					622.0	20.40	0.98	-52.953	-2.596	215.00	-112.443	4.25	-3.032	-17.750
STATISTICAL UNCERTAINTY (+/- 2*SD):						0.32	0.02	1.574	0.064	0.0	1.319	0.01	0.022	0.109

Table IV.9 Calculated structural and energetic quantities from Monte Carlo simulation of adenine and 215 waters at 298 K - Clementi solute-water potential functions.

Figures IV.8.4 through IV.8.11 - calculated quasicomponent distribution function of first shell coordination number for adenine and functional groups (Clementi potentials).

X axis - coordination number.

Y axis - quasicomponent of coordination number.

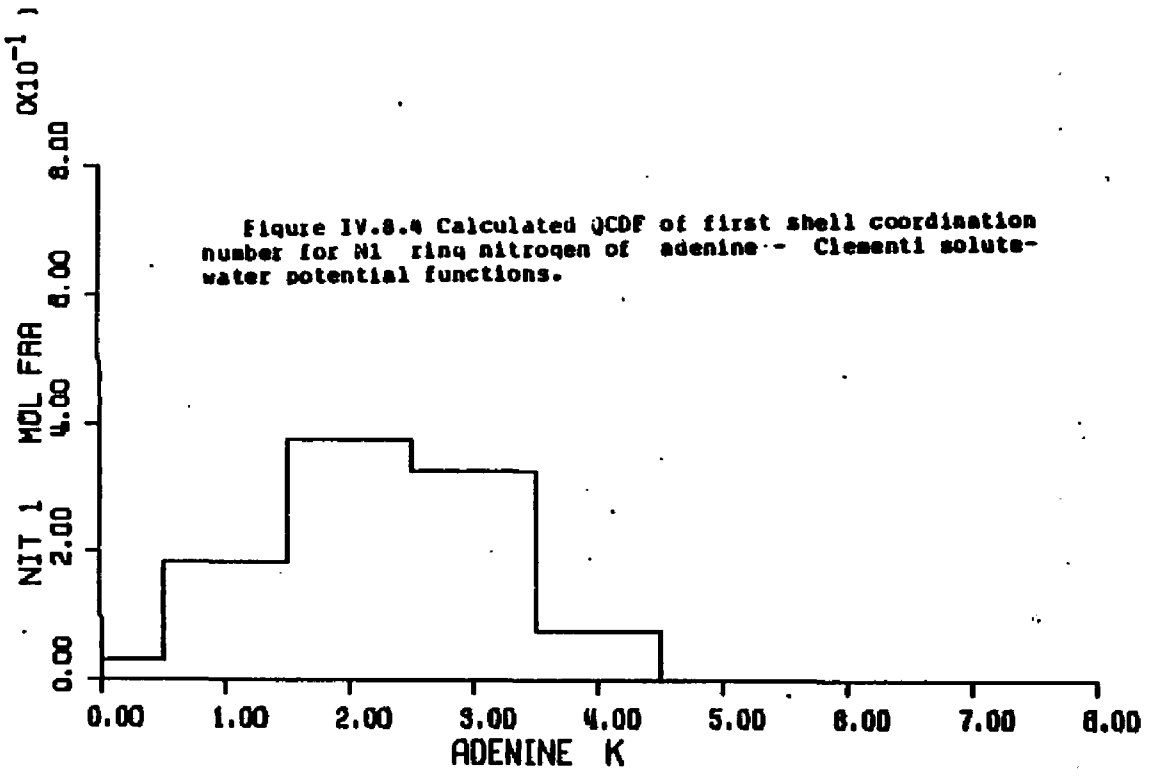


Figure IV.8.4 Calculated QCDF of first shell coordination number for Ni ring nitrogen of adenine -- Clementi solute-water potential functions.

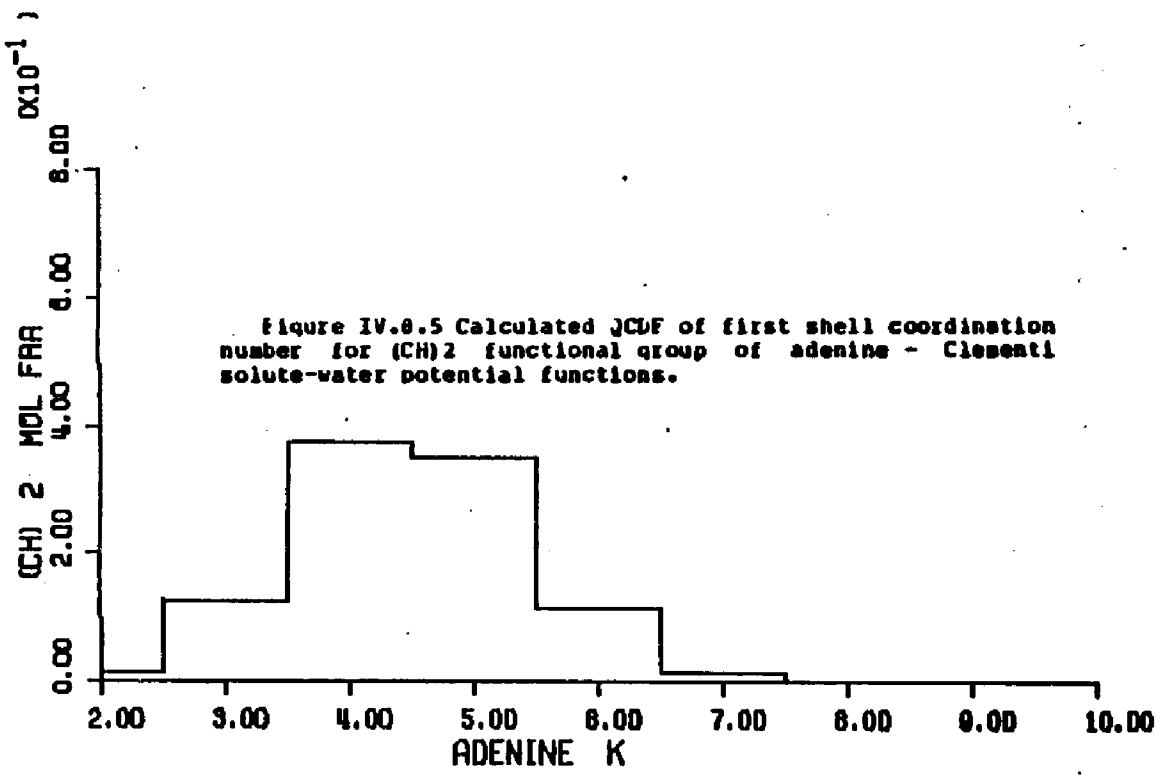


Figure IV.8.5 Calculated JCDF of first shell coordination number for (CH)2 functional group of adenine - Clementi solute-water potential functions.

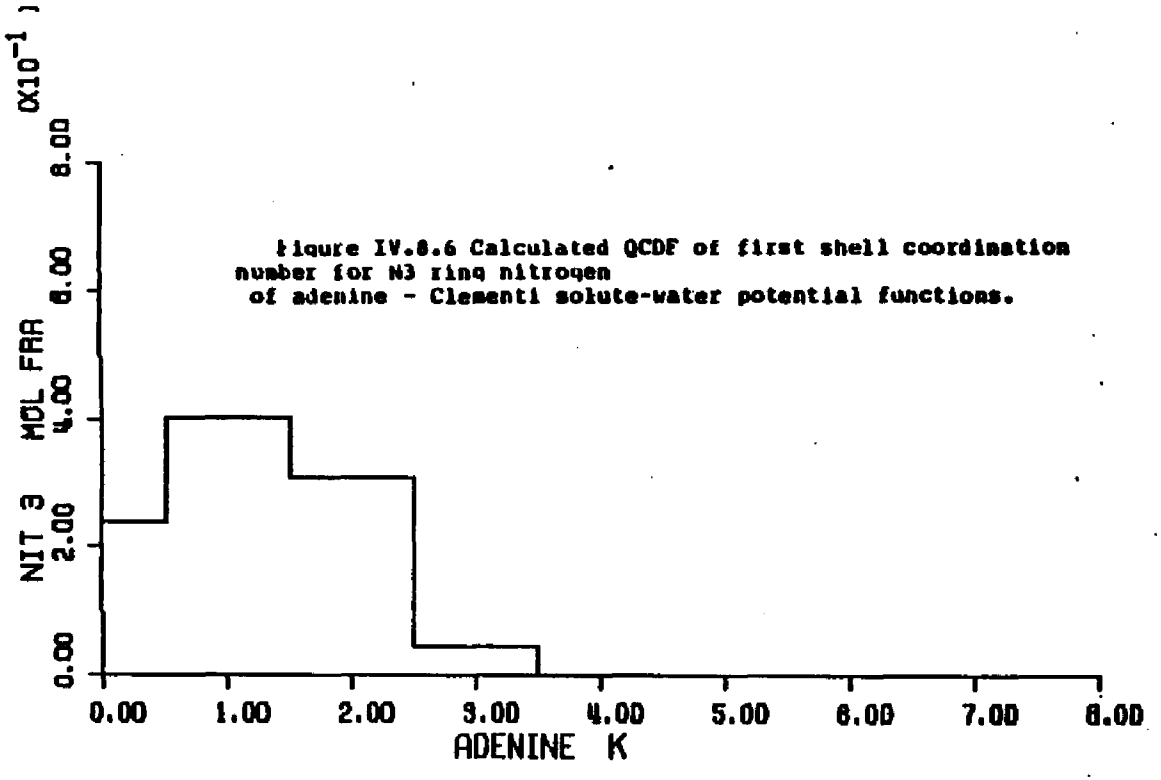


Figure IV.8.6 Calculated QCDF of first shell coordination number for N3 ring nitrogen of adenine - Clementi solute-water potential functions.

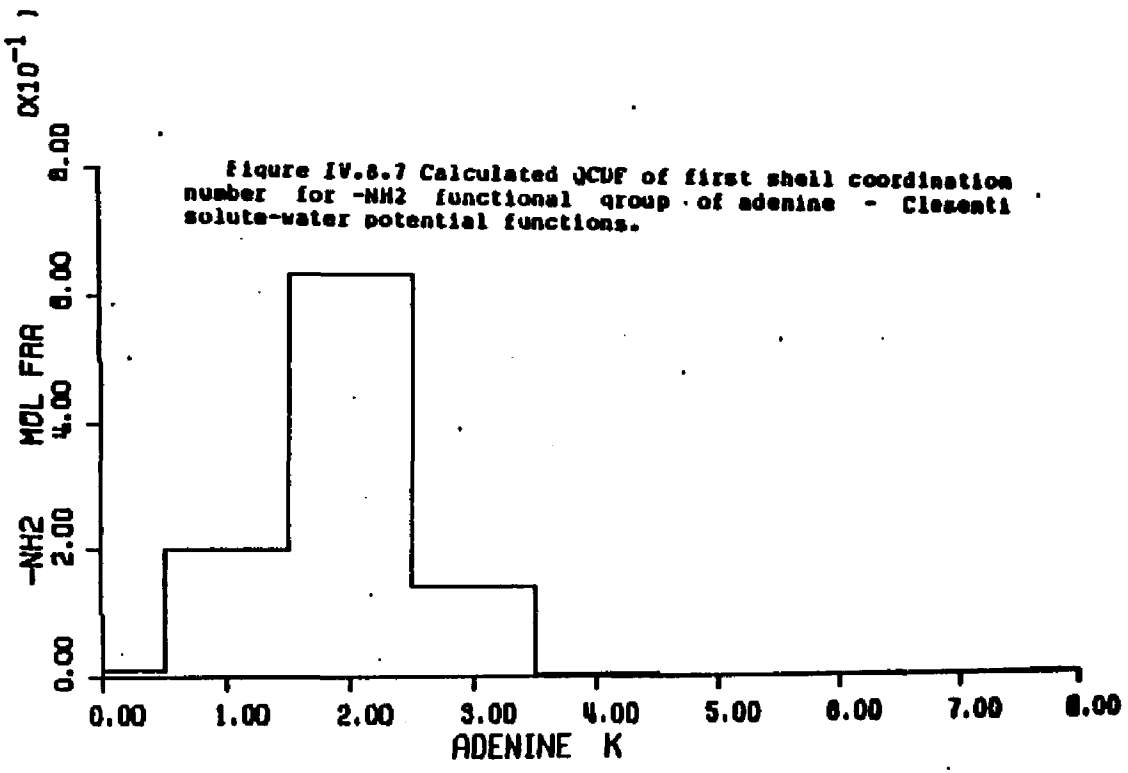


Figure IV.8.7 Calculated QCUF of first shell coordination number for -NH2 functional group of adenine - Clementi solute-water potential functions.

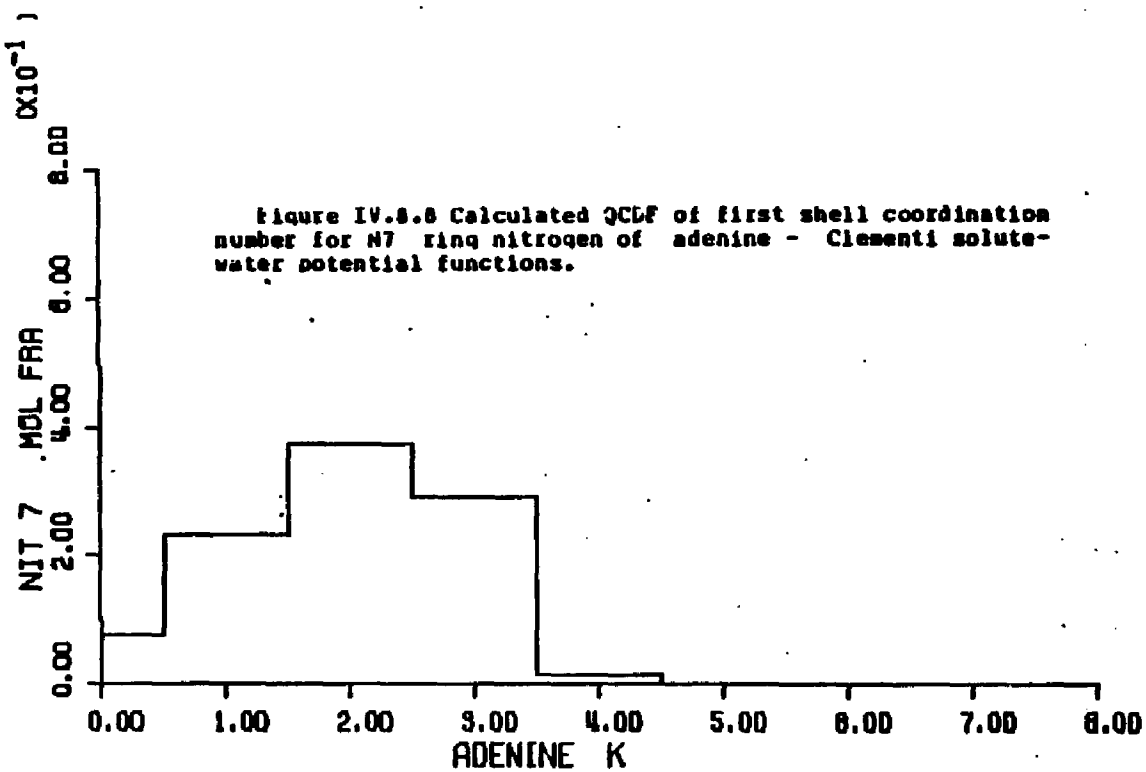


Figure IV.8.8 Calculated QCDF of first shell coordination number for N7 ring nitrogen of adenine - Clementi solute-water potential functions.

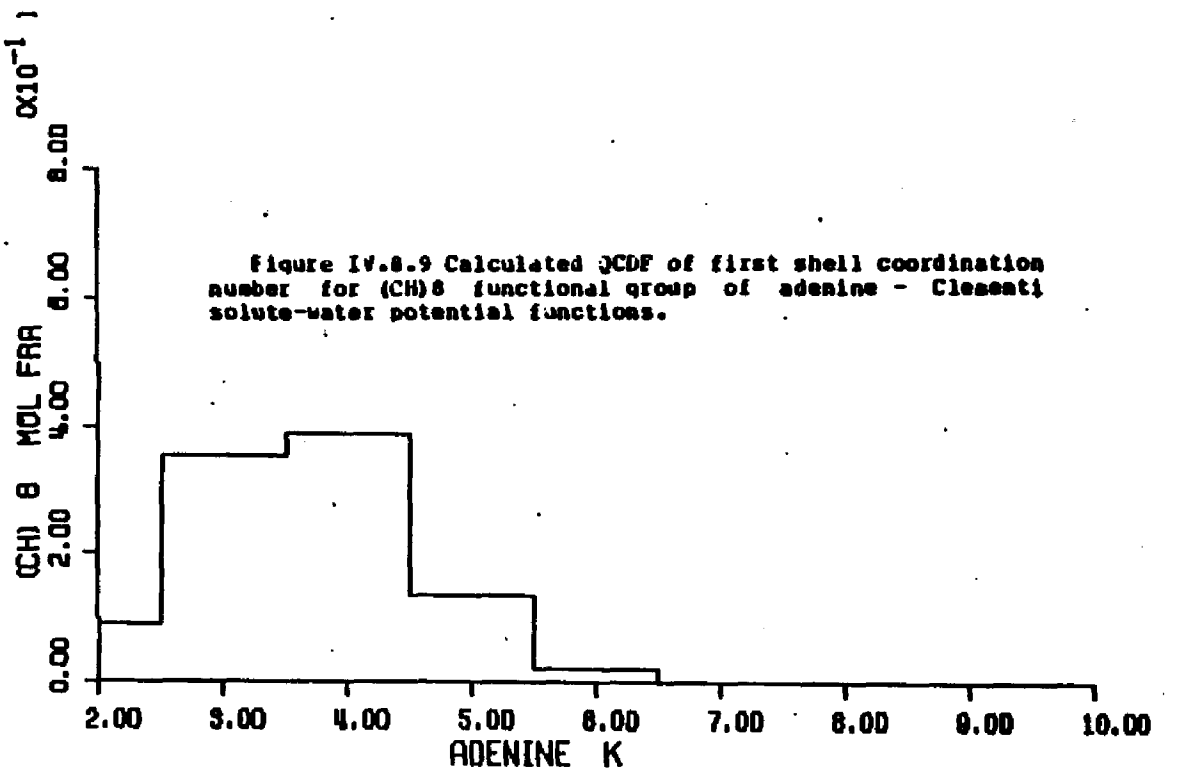


Figure IV.8.9 Calculated QCDF of first shell coordination number for (CH)8 functional group of adenine - Clementi solute-water potential functions.

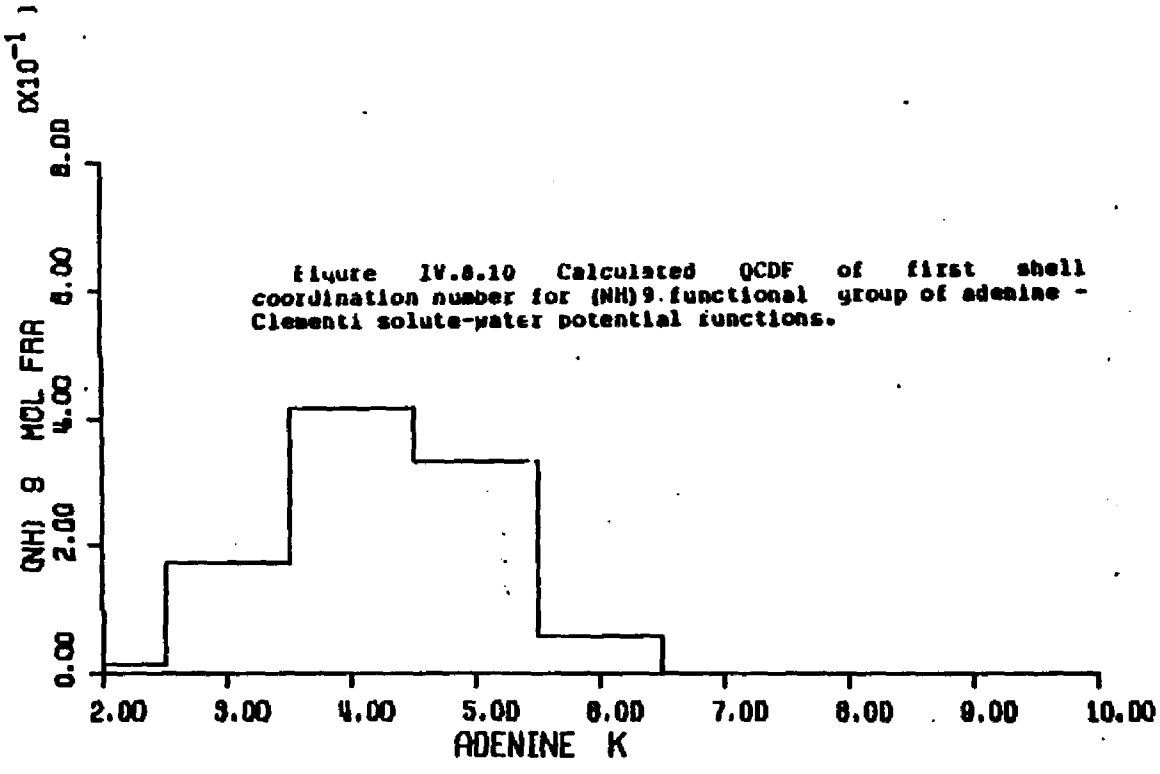


Figure IV.8.10 Calculated QCDF of first shell coordination number for (NH)9 functional group of adenine - Clementi solute-water potential functions.

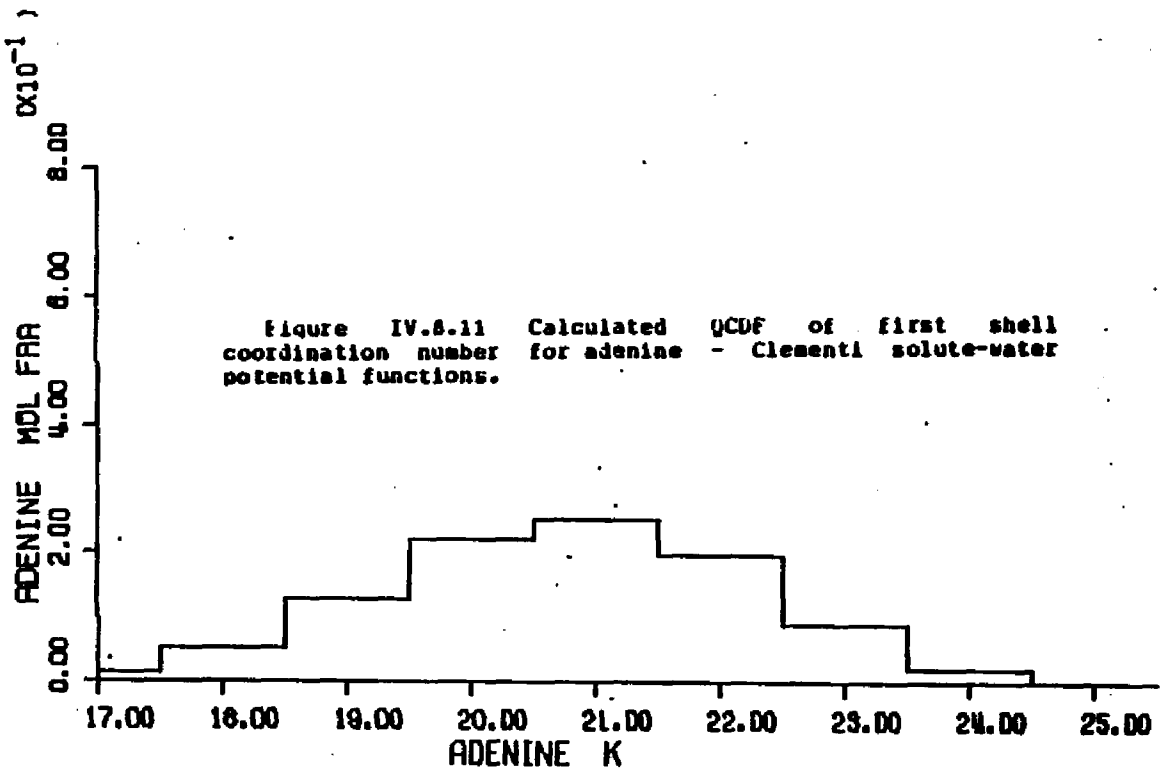


Figure IV.8.11 Calculated QCDF of first shell coordination number for adenine - Clementi solute-water potential functions.

Figure IV.8.12 Average first shell solute-water pair energies of waters assigned to the atoms of adenine - C.Gsenti solute-water potential functions.

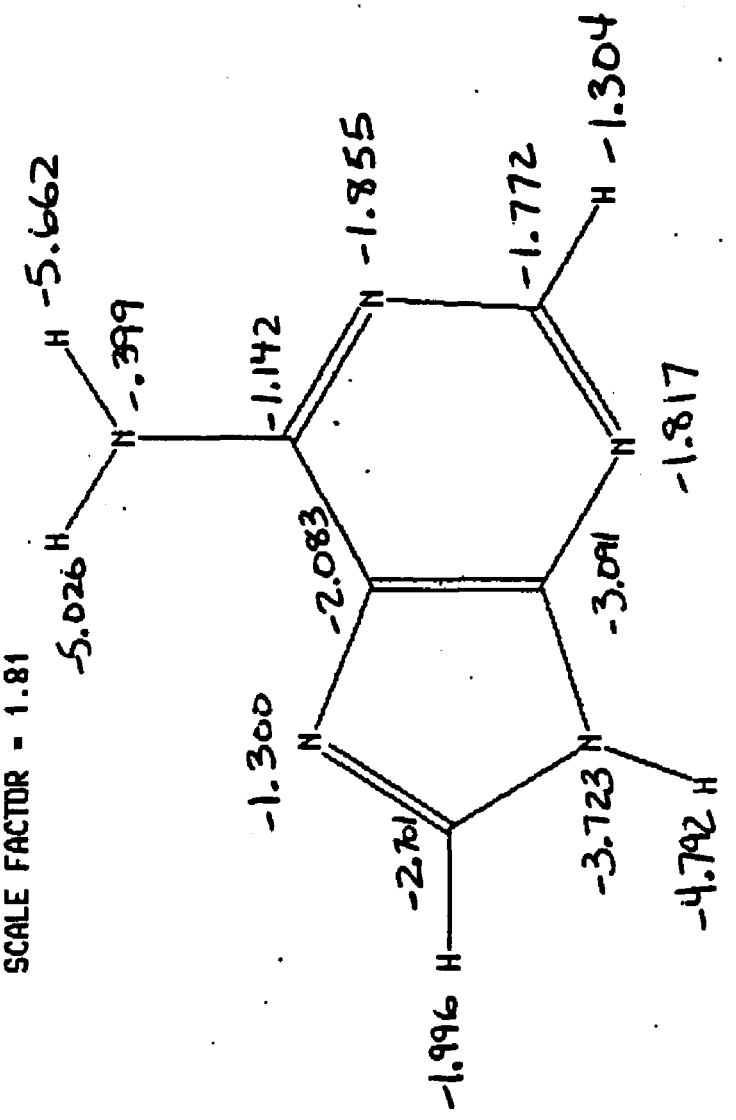


Figure IV.8.13 through IV.8.19 - calculated quasicomponent distribution function of average first shell pair energy for functional groups of adenine (Clementi potentials).

X axis - pair energy (kcal/mole).

Left Y axis - quasicomponent of pair energy.

Right Y axis - running coordination number.

Figure IV.6.13 Calculated ᵥCDF of solute-water pair energies of waters of the N1 ring nitrogen of adenine - Clementi solute-water potential functions.

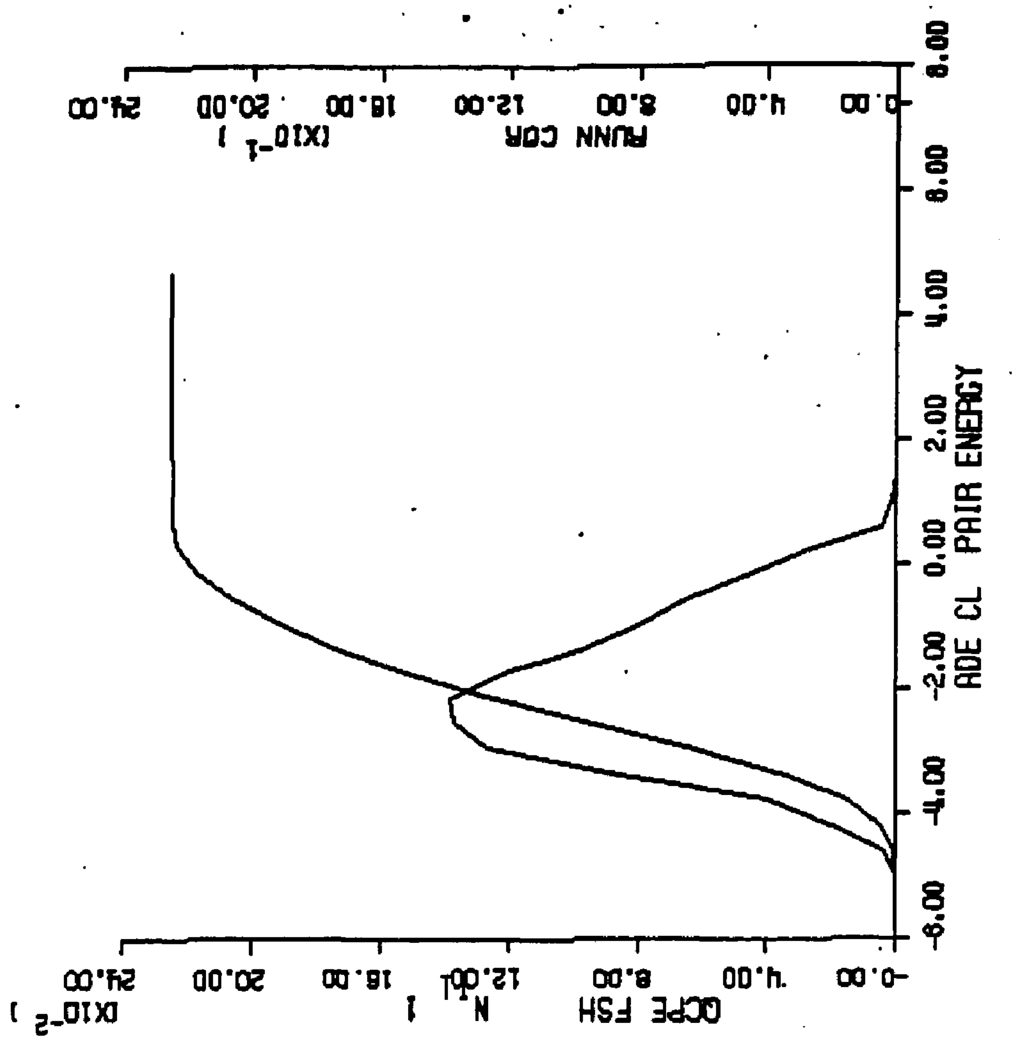


Figure IV.B.14 Calculated QCDF of solute-water pair energies of waters of the (CH)2 functional group of adenine - Clementi solute-water potential functions.

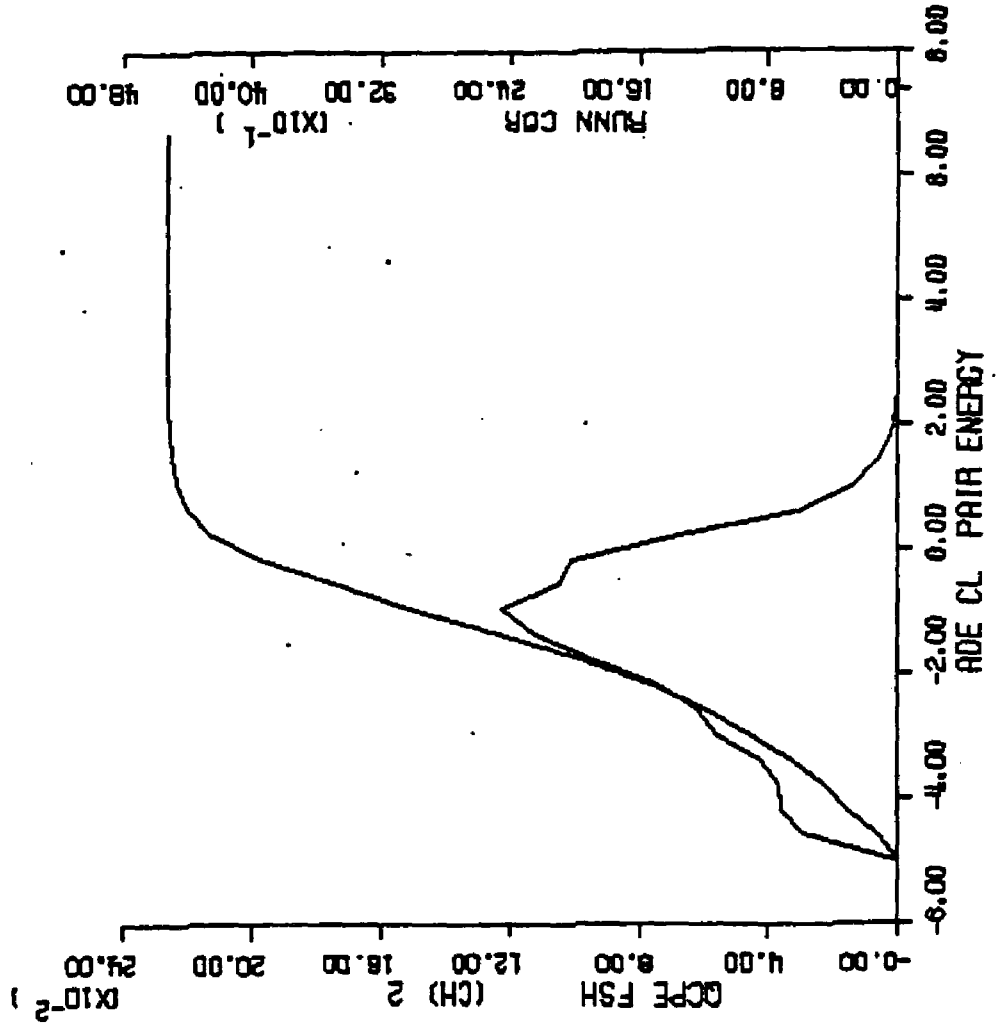


Figure IV-8.15 Calculated vCDF of solute-water pair energies of waters of the N1 and nitrogen of adenine - Clementi solute-water potential functions.

Figure IV.8.16 Calculated QCDF of solute-water pair energies of waters of the -NH2 functional group of adenine - Clementi solute-water potential functions.

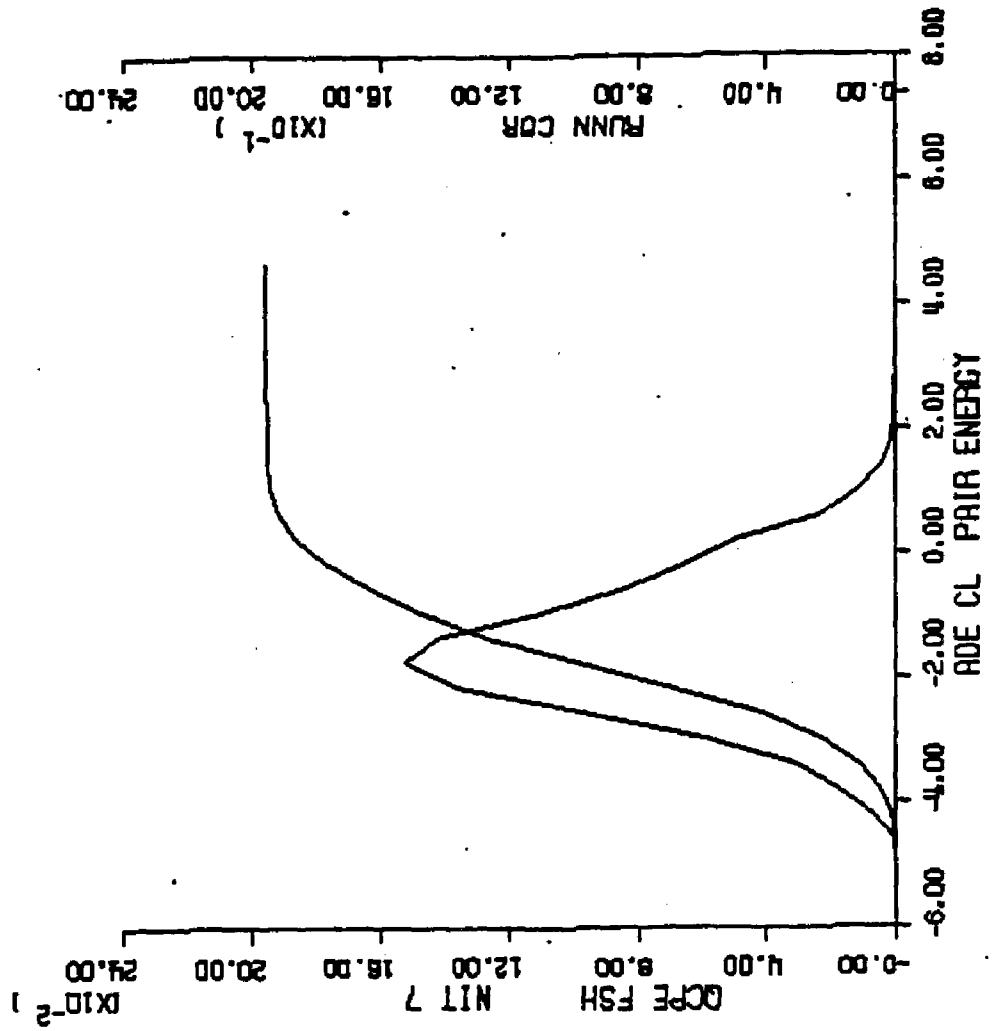


Figure IV.8.17 Calculated QCDF of solute-water pair energies of waters of the N7 ring nitrogen of adenine - Clementi solute-water potential functions.

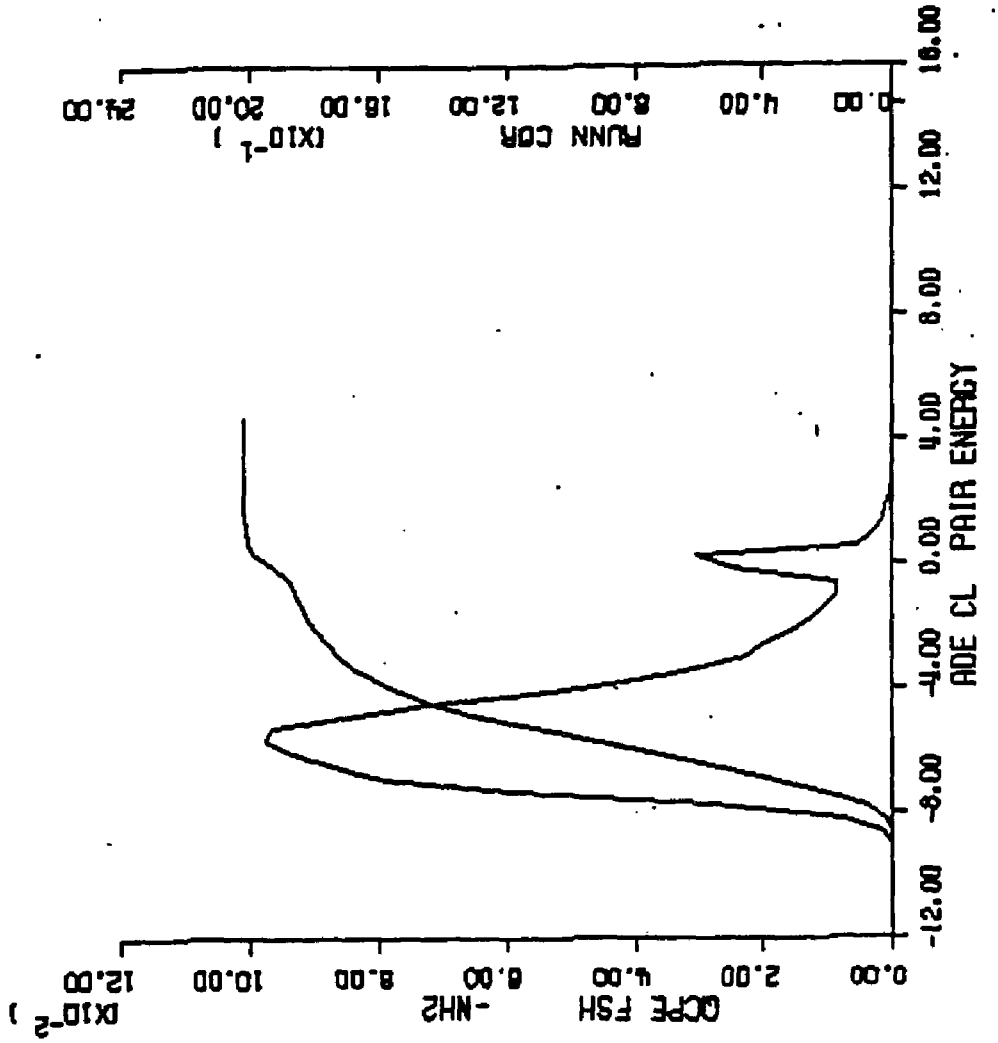


Figure IV.8.18 Calculated QCDF of solute-water pair energies of waters of the (CH)8 functional group of adenine - Clementi solute-water potential functions.

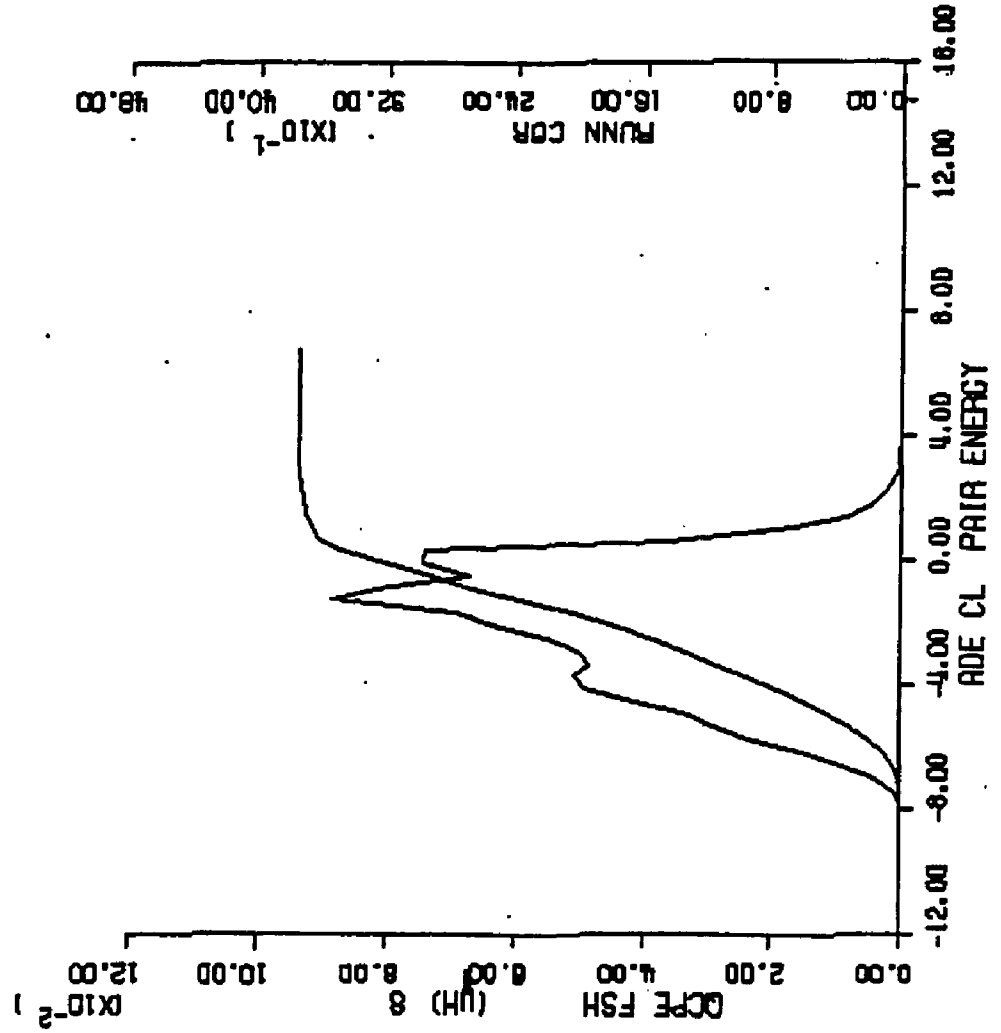


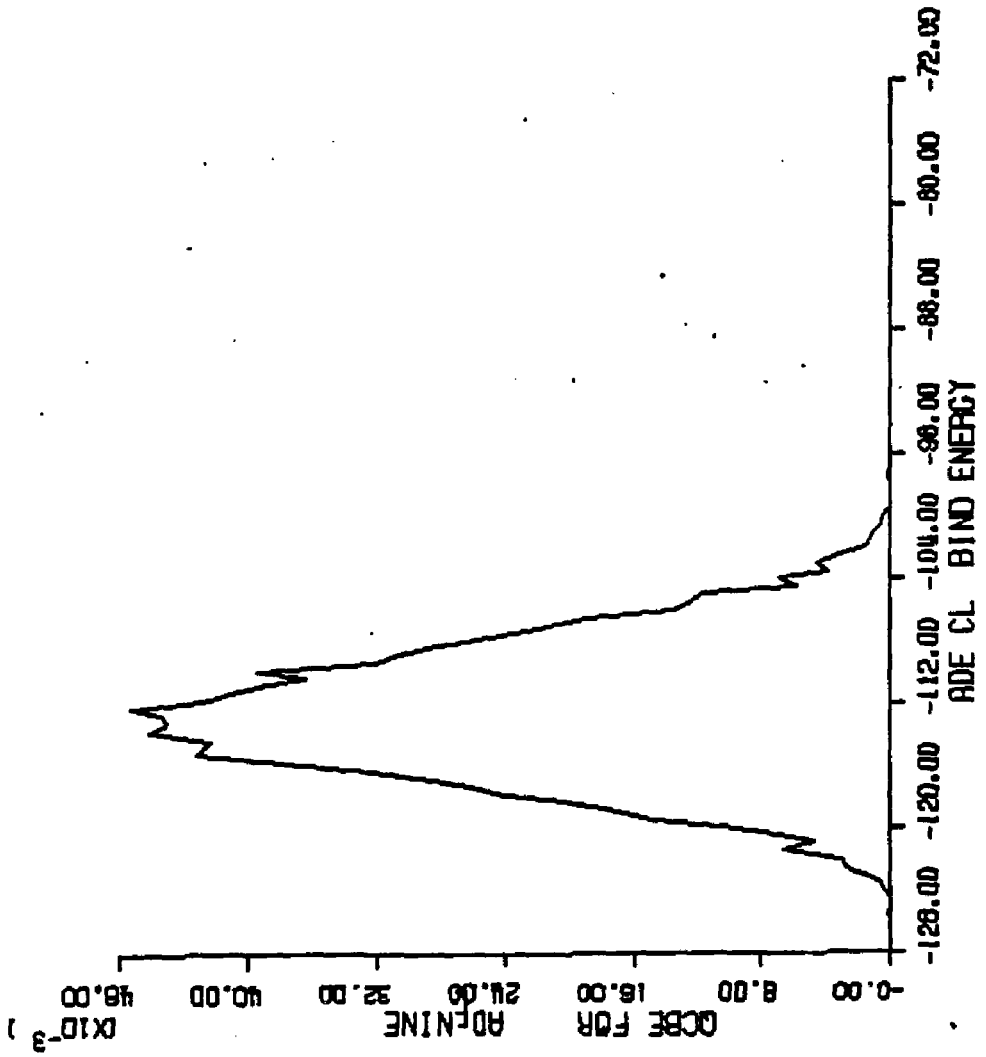
Figure IV.8.19 Calculated QCDF of solute-water pair energies of states of the (NH) 9 functional group of adenine - Classical solute-water potential functions.

Figure IV.8.20 - calculated quasicomponent distribution function of binding energy for adenine (Clementi potentials).

X axis - binding energy (Kcal/mole).

Y axis - quasicomponent of binding energy.

Figure IV.8.20 Calculated QCDF of total binding energy for adenine - Clementi solute-water potential functions.



9. Guanine with Clementi Solute-Water Potentials

Computational Specifics. An aqueous solution of guanine and 215 waters was simulated using Clementi solute-water potentials and the MCY-CI water-water potential function set at 298 K. A unit cell of 18.75 Angstroms was used with simple cubic periodic boundary conditions. An initial equilibration period of 600,000 moves was discarded; a subsequent 2,000,000 move production period was used in the formation of ensemble averages and as data for all analysis.

Potential Surface. The isoenergy contour surfaces for the interaction of guanine and one water molecule, using Clementi solute-water potential functions for their calculation, is shown in Figure IV.9.1. The global minimum lies between the H1 imino hydrogen and the H2 amino hydrogen at -13.66 kcal/mole. Another deep well is found on the carbonyl side of this imino hydrogen between -11.66 and -13.66 kcal/mole. The next most favorable interaction region occurs on the side of the H9 imino hydrogen towards C4 and N3 at about -9.66 kcal/mole. The region near the second amino hydrogen, H2', has solute-water interactions falling between -7.66 and -9.66 kcal/mole.

Interaction energies in the region surrounding the O6 carbonyl oxygen range from -5.66 to -7.66 kcal/mole. The areas surrounding the bare ring nitrogens, N3 and N7, have guanine -water interactions falling between -3.66

and -7.66 kcal/mole. Finally, the region of the methine H8 hydrogen shows energies ranging from -3.66 to -5.66 kcal/mole (on the N7 side) to between -5.66 and -7.66 kcal/mole (on the N9H9 side).

Convergence and Thermodynamic Results. The control function for this simulation are shown in Figure IV.9.2; the calculated thermodynamic quantities are listed in Table IV.1. The mean energy remains around -1970 kcal/mole for the last 800K of the simulation and attains a final value of -1972.4 +/- 4.6 kcal/mole. The vacuum to water transfer energy is calculated to be -112.6 kcal/mole. The final value of the heat capacity is 16.9 kcal/mole.

Figures IV.9.4 thru IV.9.10 and IV.9.13 thru IV.9.19 contain distribution functions for the quasicomponent of coordination number for the imino, amino, carbonyl, methine and bare ring nitrogen functional groups of guanine. Figures IV.9.11 and IV.9.20 contain the quasicomponents of coordination number and binding energy for the guanine molecule.

Structural Results. The atomic first shell coordination numbers, derived from the column labeled <K> of Table 10, are displayed on the molecular diagram of Figure IV.10.3. The imino nitrogens, N1 and N3, and the N2 amino nitrogen, are sparsely hydrated due to low solvent accessibility. The imino hydrogen, H1 and H9,

are both significantly hydrated (H1: 2.77 waters; H9: 3.63 waters). The lower first shell population of the H1 hydrogen is roughly proportional to its smaller first shell volume (51.57 cubic Angstroms vs. 78.78 cubic Angstroms for H9). The amino hydrogens have first shell populations of about one water each (H2: .86 waters; H2': .97 waters).

First shell populations of the base ring nitrogens N3 and N7 are significant as a result of the appreciable exposure of these unsubstituted atoms to in plane solvation. The N7 population (2.40 waters) is greater than the N3 population (1.64) despite the more extended first shell radius of N3 (4.8 Angstroms vs. 4.0 Angstroms for N7). The greater first shell population for N7 is proportional, however, to its greater first shell volume (68.36 vs. 59.57) resulting from less crowding by peripheral ring substituents at the N7 position of the five membered imidazole ring.

The O6 carbonyl oxygen contains a first shell population typical for this functional group (2.53 waters) at a density somewhat lower than bulk water (.93). The methine hydrogen, H8, is hydrated by 3.22 waters at a higher than bulk water density (1.14); this is a high density for an apolar hydrogen.

The quasicomponent distribution of coordination number for guanine is found in Figure IV.9.11. The coordination

numbers range from 18 to 25 and the greatest contribution arises from a coordination number of 22. The average molecular coordination number is 21.50 +/- 0.41.

Energetic Results. First shell solute-water pair energies for waters assigned to each atom of guanine are shown on the molecular diagram of Figure IV.9.12, derived from the column labeled <SLTPE> of Table IV.10. The waters assigned to the H2'' and H2' amino hydrogens have the lowest pair energies in the molecular hydration shell at -8.532 and -7.021 kcal/mole respectively. The waters of the imino hydrogen H1, however, provide the greatest overall contribution to the solute binding energy (-15.534 kcal/mole; <SLTPE>=-4.276 kcal/mole). The waters of H9 also contribute considerably (-13.122 kcal/mole; <SLTPE>= -4.403 kcal/mole). Together, the waters of H1 and H9 provide about one quarter of the total solute binding energy of -120 kcal/mole. Water-water interactions in the extended hydration shell of the hydrogen H1 are the most destabilized (-2.948 kcal/mole) of all the waters (average water-water pair energy= -3.005 kcal/mole). Again, this result is expected with the hydration of solvent breaking, polar moieties. The first shell waters of the other polar ring substituent, the carbonyl oxygen O2, are also moderately favorably bound (<SLTPE>= -3.251 kcal/mole).

The waters of the less polar N3 and N7 ring

constituents are relatively weakly bound. The average solute-water pair energy for a N3 water is -2.075 kcal/mole; a N7 water is bound with an average -1.504 kcal/mole. On the other hand, the waters of the presumably apolar methine H8 atom display more favorable average pair interactions (-2.140 kcal/mole). Waters in the extended hydration shell of this atom are destabilized (-2.982 kcal/mole) relative to bulk water; this result is expected from polar substituents.

The quasicomponent distribution function for the solute binding energy of guanine is displayed in Figure IV.9.20. Energy values range from -137 to -104 kcal/mole. The most significant contribution arise in the range -124 to -116 kcal/mole. The average solute binding energy is -120.0 +/- 1.0 kcal/mole.

Figure IV.9.1 - isoenergy contour surface for guanine and one water (Clementi solute-water potential functions).

X axis - X axis of plane defined by molecular ring atoms (Angstroms).

Y axis - Y axis of plane defined by molecular ring atoms (Angstroms).

Contours - isoenergy contours with one kcal/mole increments with alphabetical labels referring to contour energy values in list at right.

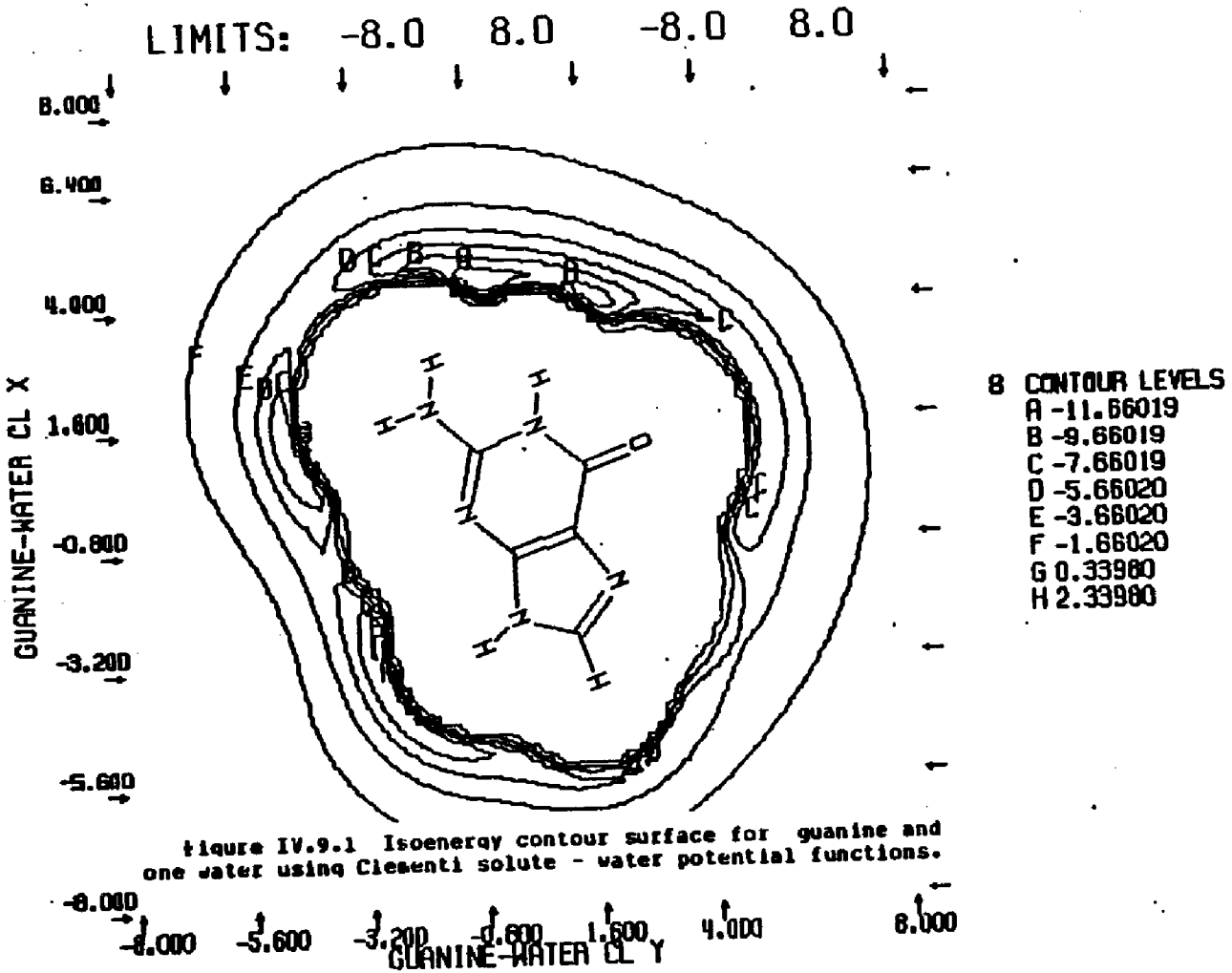


Figure IV.9.1 Isoenergy contour surface for guanine and one water using Clementi solute - water potential functions.

Figure IV.9.2 - control functions for Monte Carlo simulation of guanine and 215 waters (Clementi solute-water potential functions).

X axis - number of configurations.

Left Y axis - mean total energy (Kcal/mole).

Right Y axis. - constant volume heat capacity (cal/mole-degree).

Upper curve - constant volume heat capacity.

Bottom curve without crosshatches - average total energy for entire simulation.

Bottom curve with crosshatches - average total energy for preceeding 50K configurations.

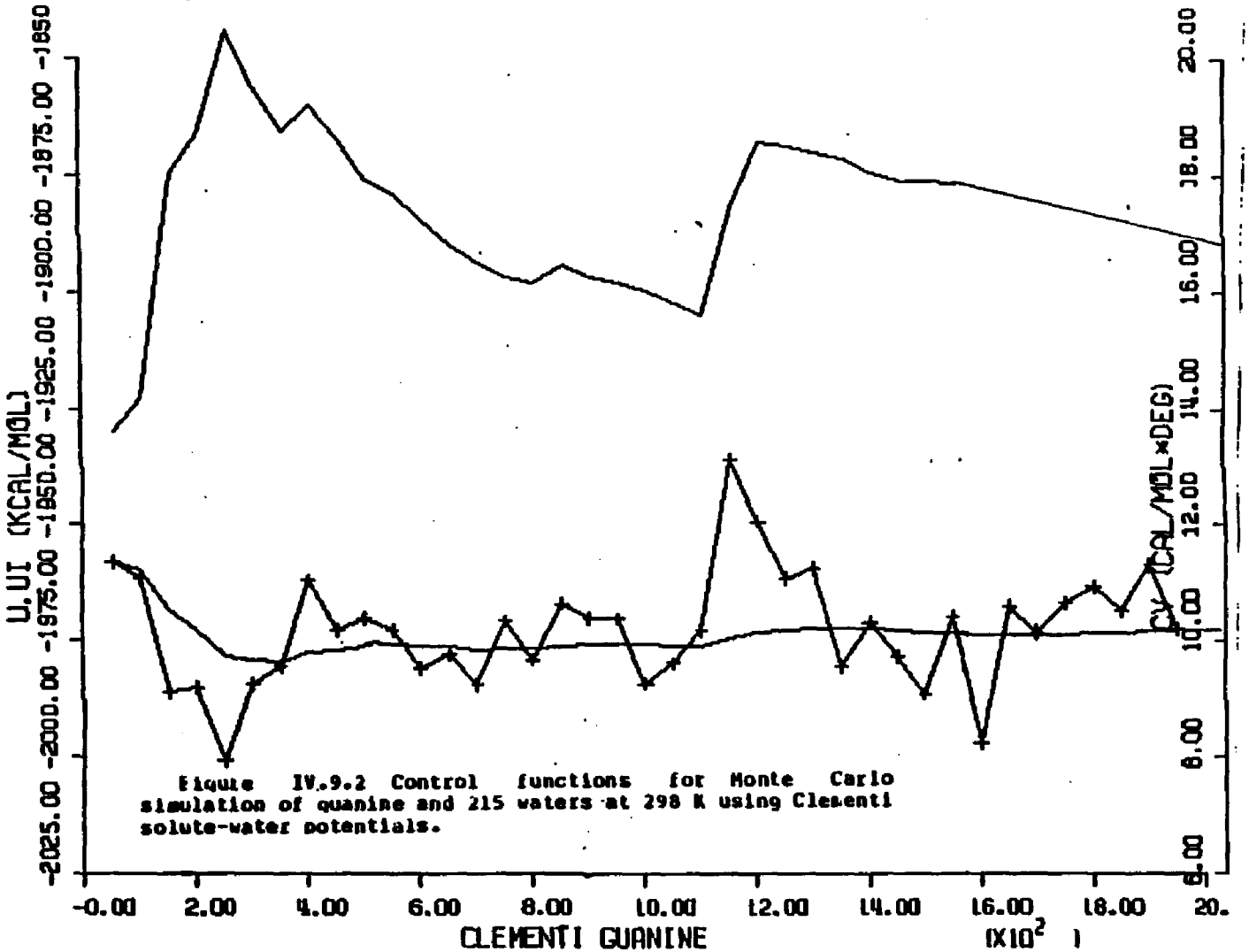


Figure IV.9.2 Control functions for Monte Carlo simulation of guanine and 215 waters at 298 K using Clementi solute-water potentials.

Table IV.10 Calculated structural and energetic quantities from Monte Carlo simulation of guanine and 215 waters at 298 K - Clementi solute-water potential functions.

GUANINE IN WATER AT 298K - FORCE BIAS AND CLEMENTI POTENTIAL FUNCTIONS

LAST CONFIGURATION: 1997898					FIRST SHELL SOLUTE PROPERTIES					TOTAL SLT PROPS		WATER PROPERTIES RFSW=3.38 RCE= 7.75 A		
INDEX	TYPE				NFS	VFS	<K>	<K/V>	<SLTBE>	<SLTPE>	<K>	<SLTBE>	<KW>	<SWWPE>	<SEWMT>
METHYLENE (DISUBST.)															
1	7	24	C	2	5.8	22.93	0.73	0.95	-1.695	-2.329	1.74	-1.948	4.04	-3.252	-17.378
2	8	28.	C	4	4.4	21.08	0.32	0.46	-0.792	-2.447	1.41	-1.083	4.16	-3.123	-17.115
3	9	19	C	5	4.2	20.71	0.38	0.55	-0.279	-0.733	1.58	-0.489	4.11	-3.123	-17.148
AVERAGES OVER FUNCTIONAL GRP =C< :						21.57	0.48	0.66	-0.922	-1.836	1.55	-1.173	4.10	-3.166	-17.208
STATISTICAL UNCERTAINTY (+/- 2*SD):							0.04	0.05	0.065	0.128	0.0	0.067	0.06	0.135	0.352
METHYLENE (MONOSUB.)															
4	11	25	C	8	3.8	15.36	0.29	0.57	-0.591	-2.020	1.32	-0.866	4.19	-2.989	-17.004
5	12	16	H	8	3.4	84.75	3.22	1.14	-6.903	-2.140	37.45	-13.151	4.21	-2.982	-16.930
TOTALS FOR FUNCTIONAL GROUP =CH- :						100.11	3.52	1.85	-7.494	-2.130	38.78	-14.018	4.21	-2.982	-16.932
STATISTICAL UNCERTAINTY (+/- 2*SD):							0.17	0.05	0.335	0.009	0.0	0.279	0.02	0.044	0.120
AMIDE GROUP															
6	2	15	N	1	3.6	14.32	0.37	0.78	-0.939	-2.519	1.37	-1.200	4.17	-3.248	-18.019
7	16	16	H	1	3.2	51.57	2.77	1.68	-12.182	-4.403	21.12	-17.468	4.32	-2.948	-17.165
TOTALS FOR FUNCTIONAL GROUP >NH :						65.89	3.14	1.43	-13.122	-4.179	22.49	-18.668	4.31	-2.966	-17.217
8	5	15	N	9	3.8	14.88	0.61	1.24	-1.931	-3.151	1.71	-2.211	4.17	-3.032	-17.126
9	13	16	H	9	3.4	78.78	3.63	1.38	-15.534	-4.276	31.44	-21.618	4.28	-2.979	-17.147
TOTALS FOR FUNCTIONAL GROUP >NH :						93.58	4.25	1.36	-17.465	-4.114	33.15	-23.828	4.27	-2.982	-17.146
AVERAGES OVER FUNCTIONAL GRP >NH :						79.74	3.69	1.39	-15.294	-4.146	27.82	-21.244	4.29	-2.974	-17.182
STATISTICAL UNCERTAINTY (+/- 2*SD):							0.12	0.05	0.471	0.119	0.0	0.353	0.02	0.037	0.102
AMINE GROUP															
10	6	11	N	2'	3.8	12.27	0.27	0.66	-0.534	-1.983	1.10	-0.793	4.19	-3.074	-17.135
11	14	1	H	2"	2.2	26.08	0.97	1.11	-6.814	-7.021	30.94	-16.741	4.10	-2.992	-16.926
12	15	1	H	2'	2.2	23.75	0.86	1.08	-7.334	-8.532	25.84	-15.718	4.40	-2.983	-17.475
TOTALS FOR FUNCTIONAL GROUP -NH2 :						62.10	2.10	1.01	-14.682	-6.994	57.88	-33.251	4.28	-2.990	-17.170
STATISTICAL UNCERTAINTY (+/- 2*SD):							0.13	0.06	0.849	0.378	0.0	0.541	0.02	0.036	0.099

Table IV.10 Calculated structural and energetic quantities from Monte Carlo simulation of guanine and 215 waters at 298 K - Clementi solute-water potential functions.

GUANINE IN WATER AT 298K - FORCE BIAS AND CLEMENTI POTENTIAL FUNCTIONS

SUBSTITUTED IMINO GR													
13	3	12	N 3	4.0	59.57	1.64	0.82	-3.402	-2.075	5.03	-4.311	4.10	-3.160	-17.545
14	4	12	N 7	4.0	60.36	2.40	1.05	-3.616	-1.504	14.64	-5.634	4.14	-3.049	-17.047
AVERAGES OVER FUNCTIONAL GRP -N- :					63.97	2.02	0.95	-3.509	-1.789	9.84	-4.973	4.16	-3.104	-17.296
STATISTICAL UNCERTAINTY (+/- 2*SD):						0.09	0.04	0.146	0.070	0.0	0.139	0.03	0.064	0.172
CARBONYL GROUP														
15	1	27	O 6	3.4	-81.61	2.53	0.93	-0.237	-3.251	36.76	-16.102	4.15	-3.033	-16.986
16	10	26	C 6	4.2	19.23	0.48	0.75	-0.449	-0.927	1.63	-0.650	4.09	-3.207	-17.567
TOTALS FOR FUNCTIONAL GROUP C=O :					100.83	3.02	0.90	-0.686	-2.078	38.38	-16.752	4.15	-3.040	-17.011
STATISTICAL UNCERTAINTY (+/- 2*SD):						0.15	0.05	0.419	0.130	0.0	0.335	0.02	0.045	0.121
MOLECULAR SUM/AVERAGE:					615.2	21.50	1.05	-71.233	-3.314	215.08	-119.974	4.23	-3.005	-17.101
STATISTICAL UNCERTAINTY (+/- 2*SD):						0.41	0.02	1.287	0.056	0.0	1.013	0.01	0.019	0.051

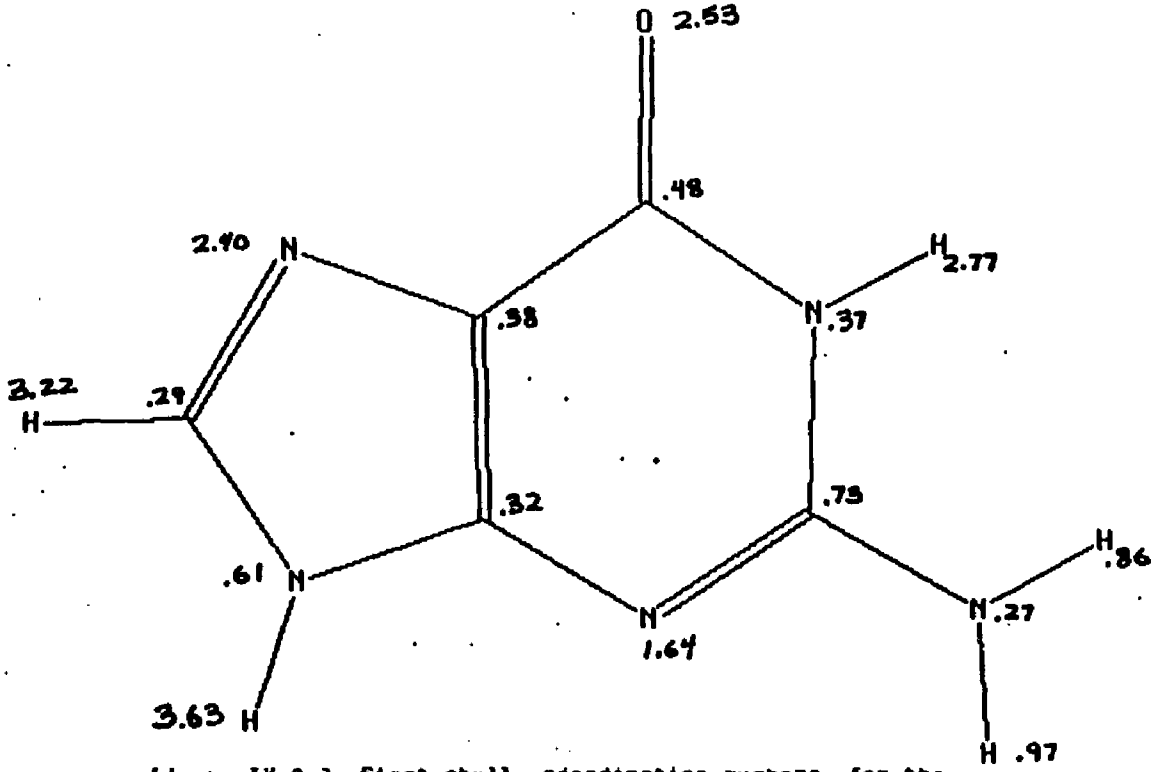


Figure IV.9.3 First shell coordination numbers for the atoms of guanine - Clementi solute-water potential functions.

Figures IV.9.4 through IV.9.11 - calculated quasicomponent distribution function of first shell coordination number for guanine and functional groups (Clementi potentials).

X axis - coordination number.

Y axis - quasicomponent of coordination number.

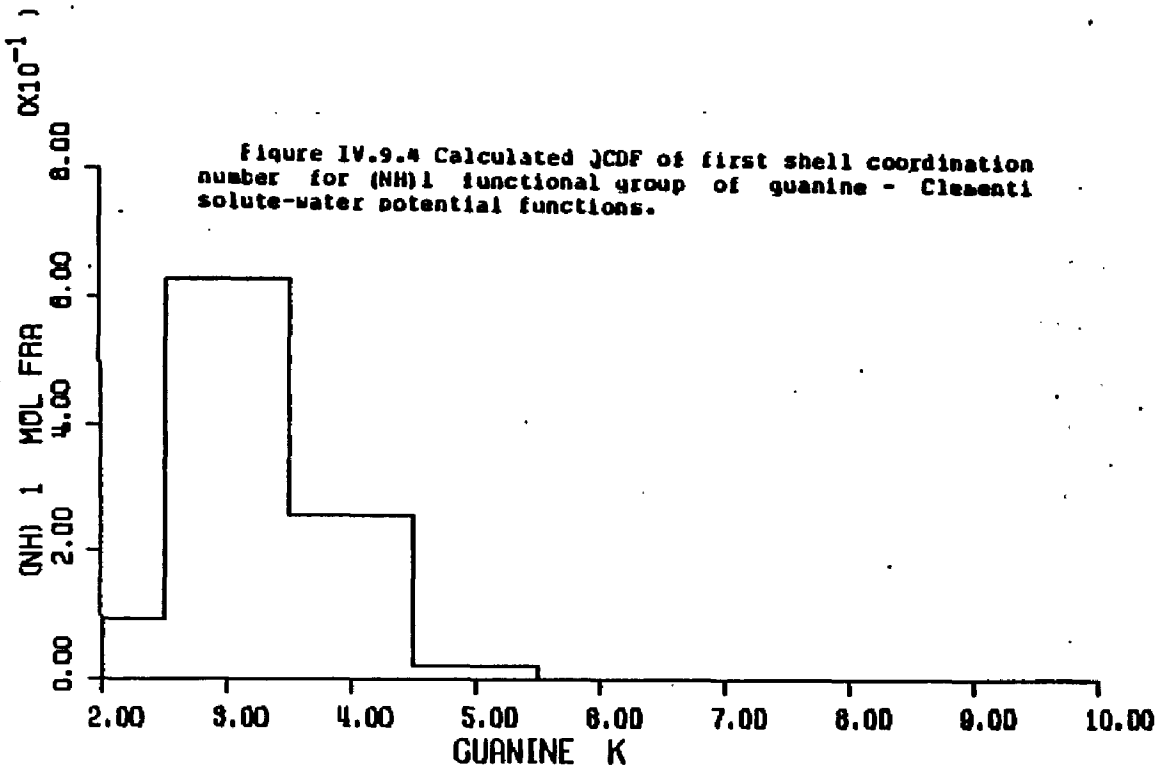


Figure IV.9.4 Calculated JCDF of first shell coordination number for (NH)1 functional group of guanine - Clementi solute-water potential functions.

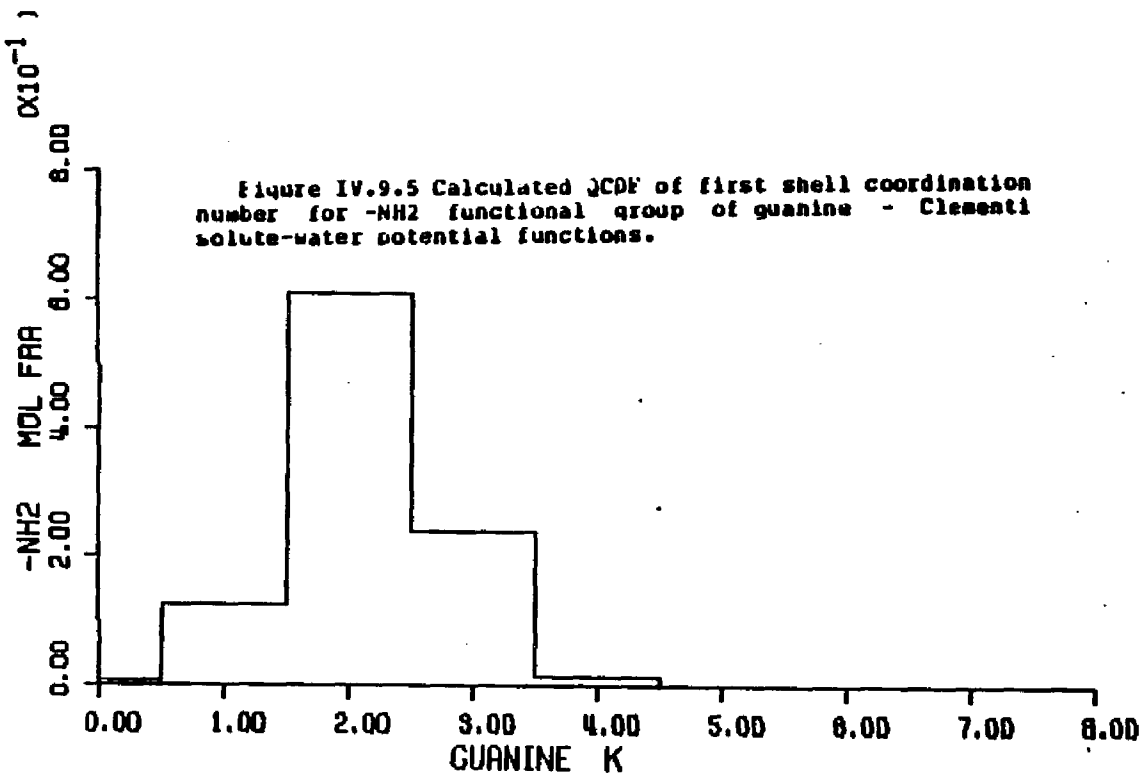


Figure IV.9.5 Calculated QCDF of first shell coordination number for -NH2 functional group of guanine - Clementi solute-water potential functions.

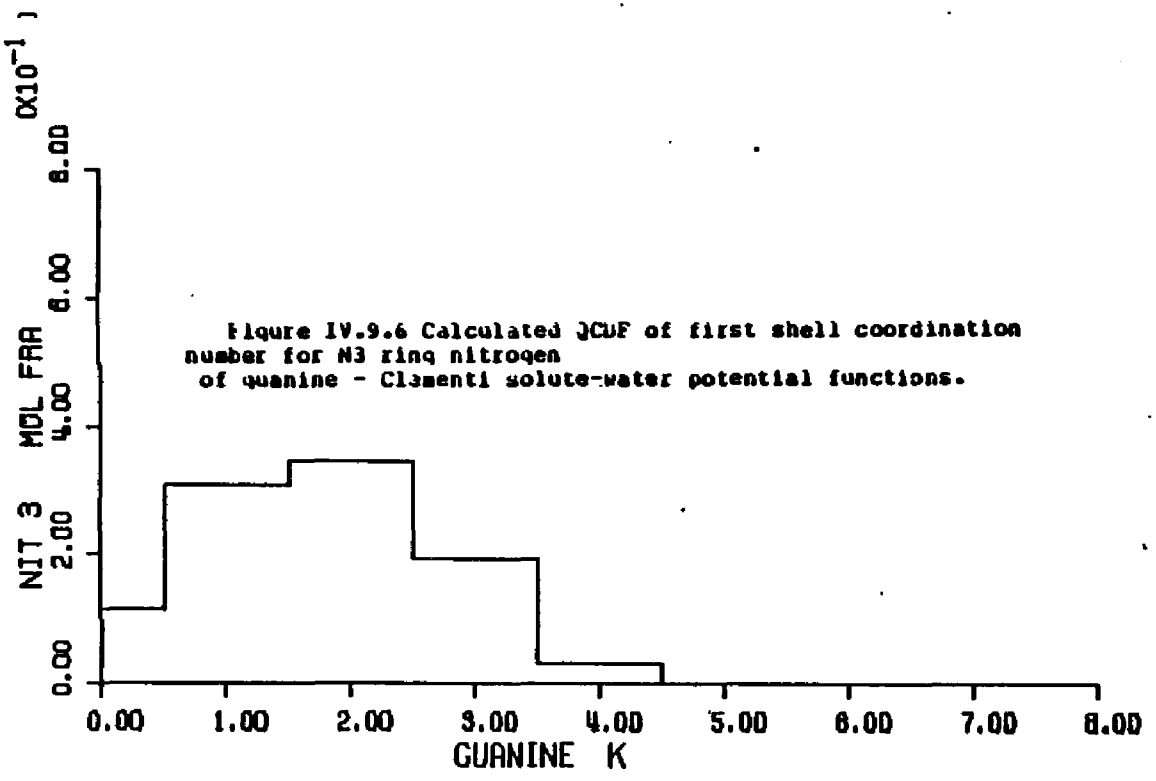


Figure IV.9.6 Calculated JCUF of first shell coordination number for N3 ring nitrogen of guanine - Clementi solute-water potential functions.

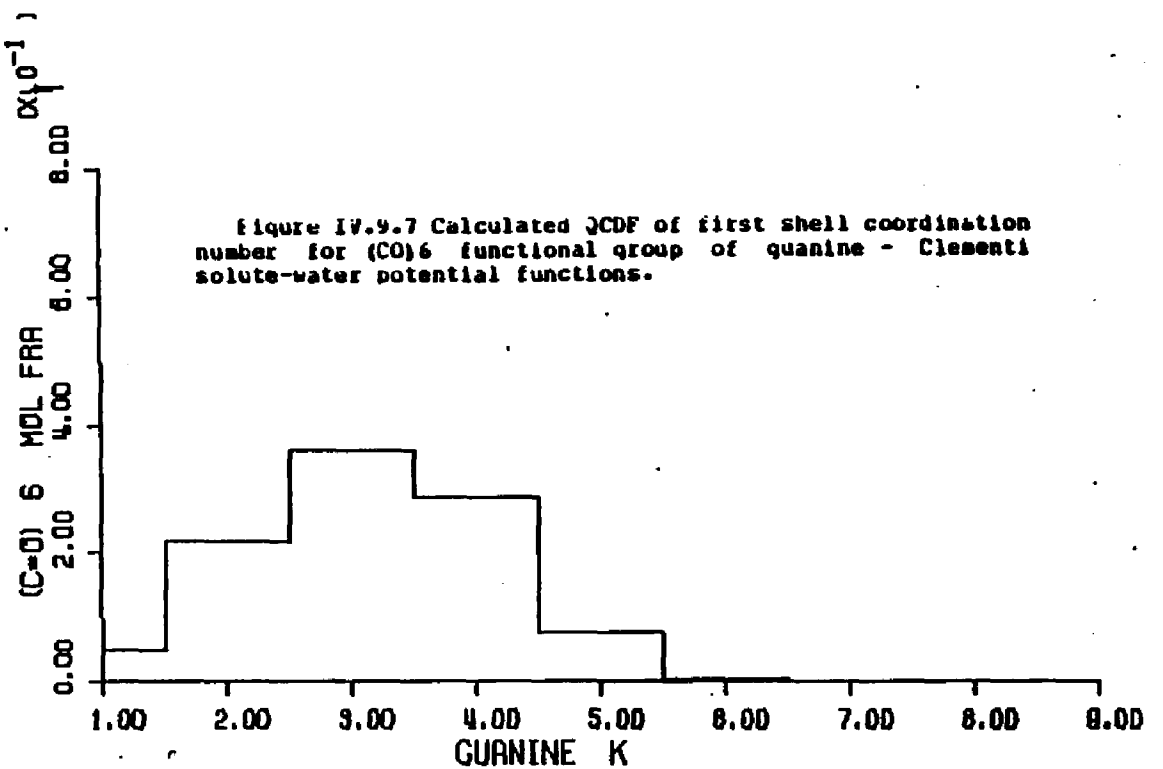


Figure IV.9.7 Calculated QCDF of first shell coordination number for (CO)6 functional group of guanine - Clementi solute-water potential functions.

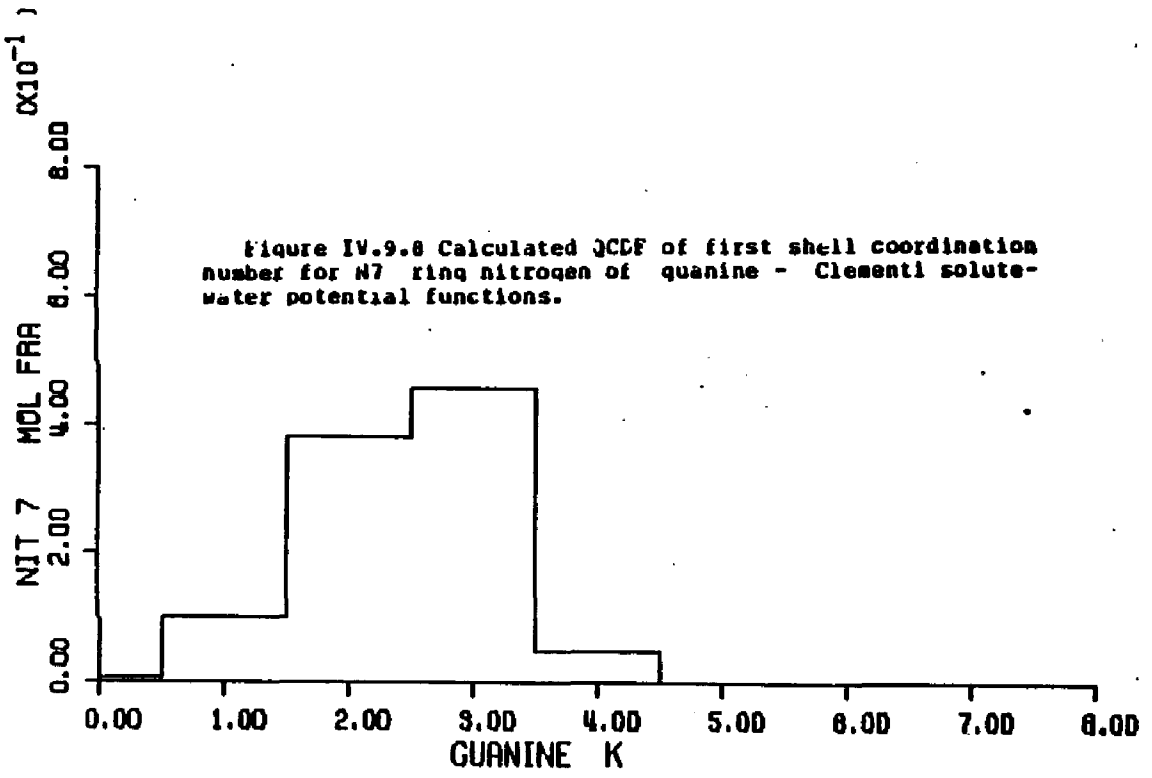


Figure IV.9.8 Calculated JCDF of first shell coordination number for N7 ring nitrogen of guanine - Clementi solute-water potential functions.

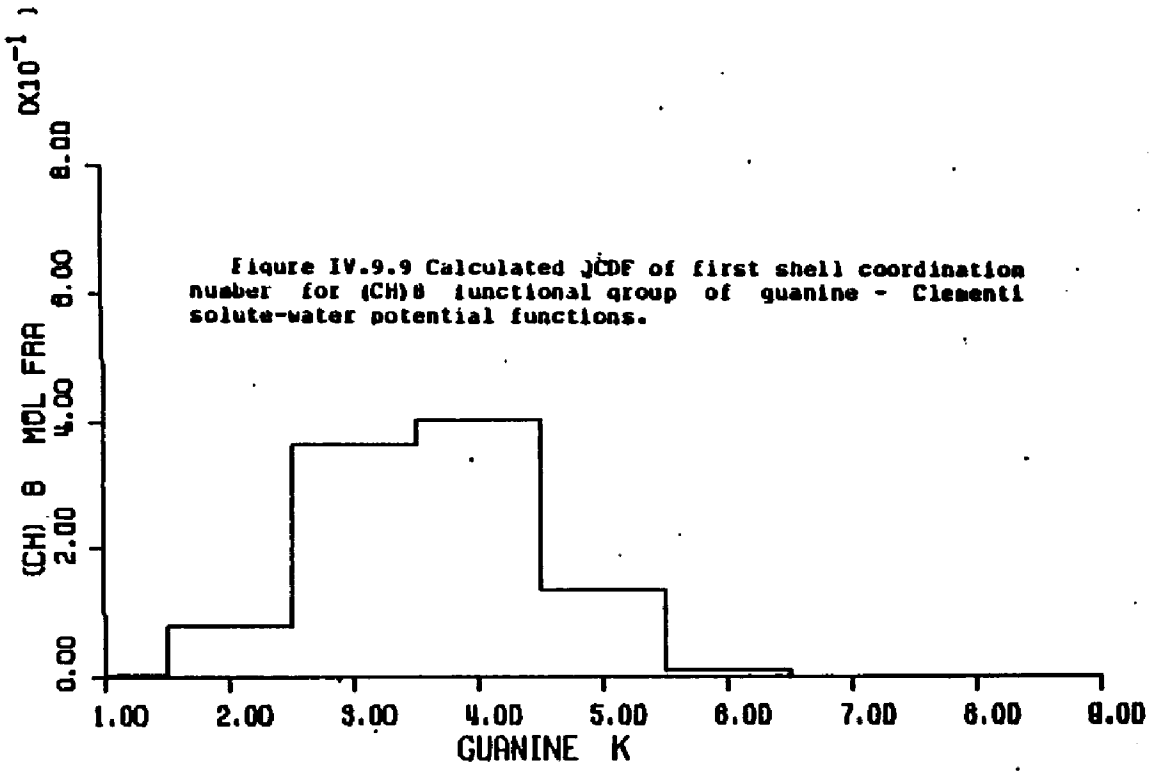


Figure IV.9.9 Calculated CDF of first shell coordination number for (CH)8 functional group of guanine - Clementi solute-water potential functions.

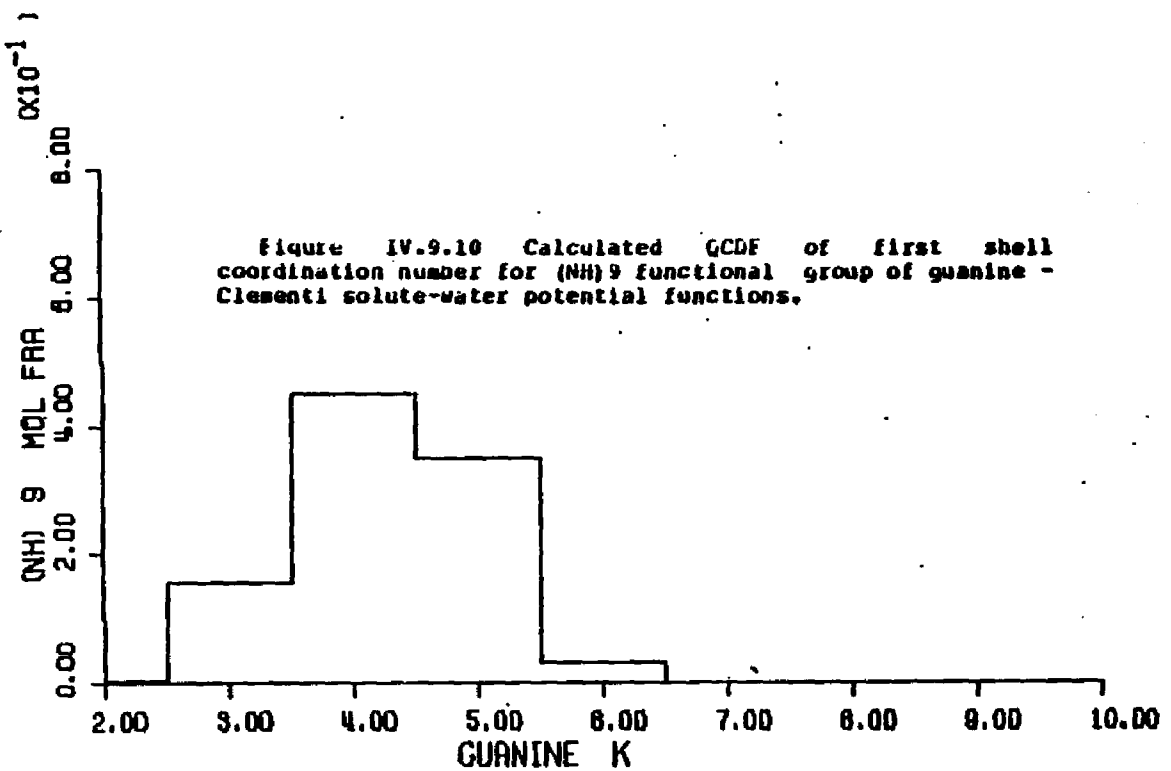


Figure IV.9.10 Calculated GCDF of first shell coordination number for (NH)9 functional group of guanine - Clementi solute-water potential functions.

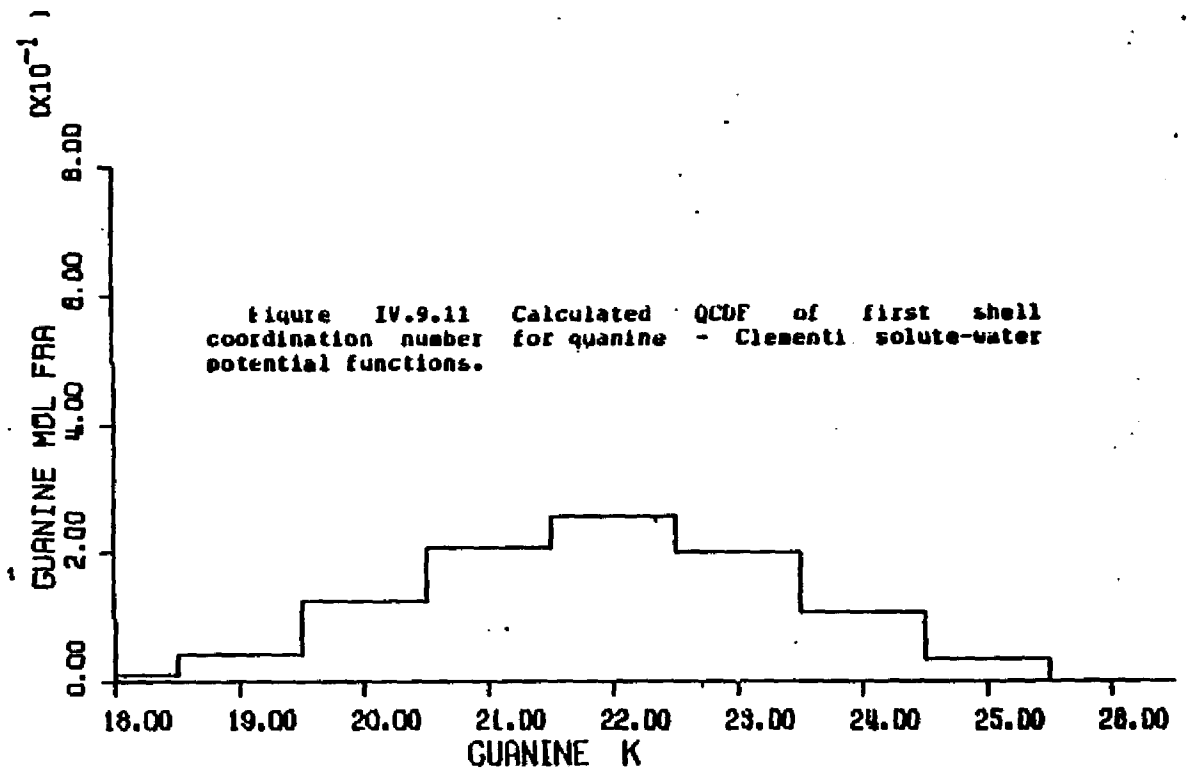


Figure IV.9.11 Calculated QCDF of first shell coordination number for guanine - Clementi solute-water potential functions.

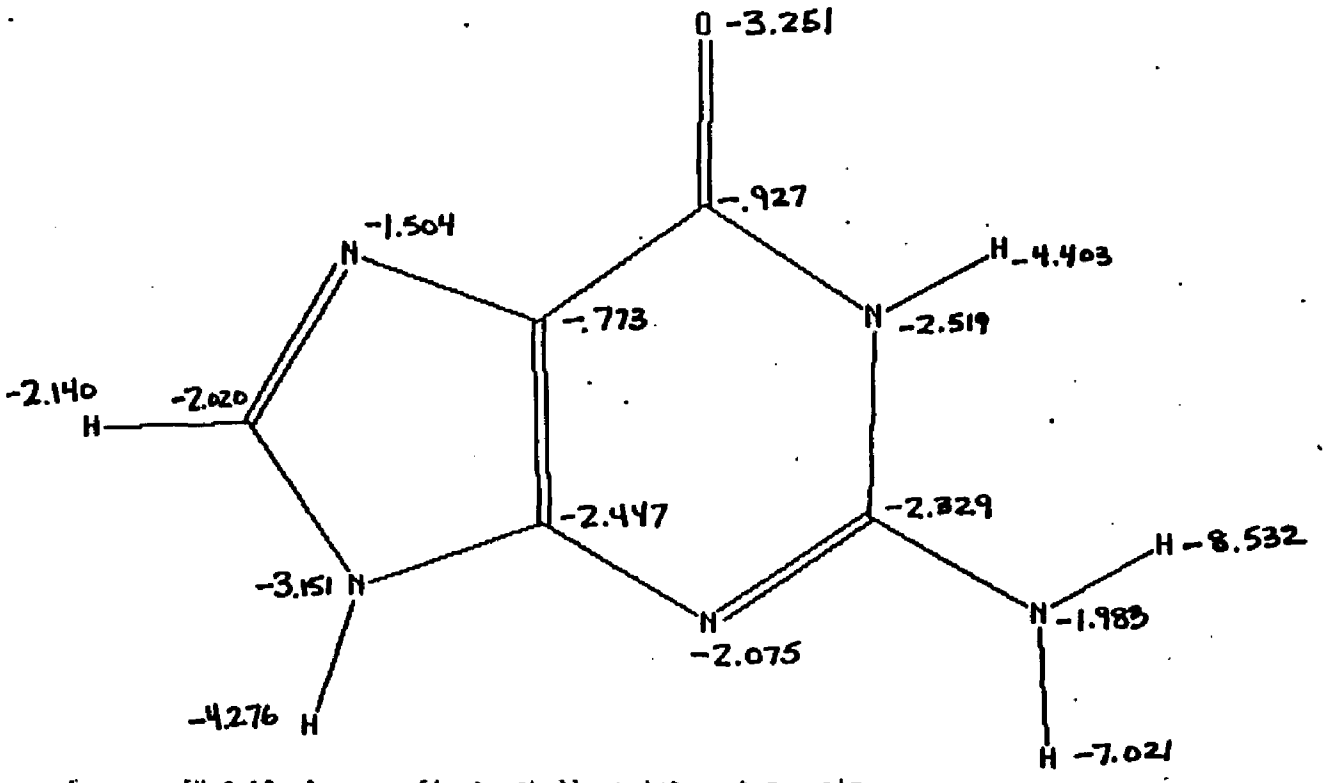


Figure IV.9.12 Average first shell solute-water pair energies of waters assigned to the atoms of guanine - Clementi solute-water potential functions.

Figure IV.9.13 through IV.9.19 - calculated quasicomponent distribution function of average first shell pair energy for functional groups of guanine (Clementi potentials).

X axis - pair energy (kcal/mole).

Left Y axis - quasicomponent of pair energy.

Right Y axis - running coordination number.

Figure IV.6.13 Calculated QCDF of solute-water pair energies of waters of the (NH)1 functional group of guanine - Clementi solute-water potential functions.

Figure IV.9.14 Calculated VCOE of solute-water pair energies of waters of the -NH2 functional group of guanine - Clementi solute-water potential functions.

Figure IV.9.15 Calculated GCDF of solute-water pair energies of waters of the N3 ring nitrogen of guanine - Clementi solute-water potential functions.

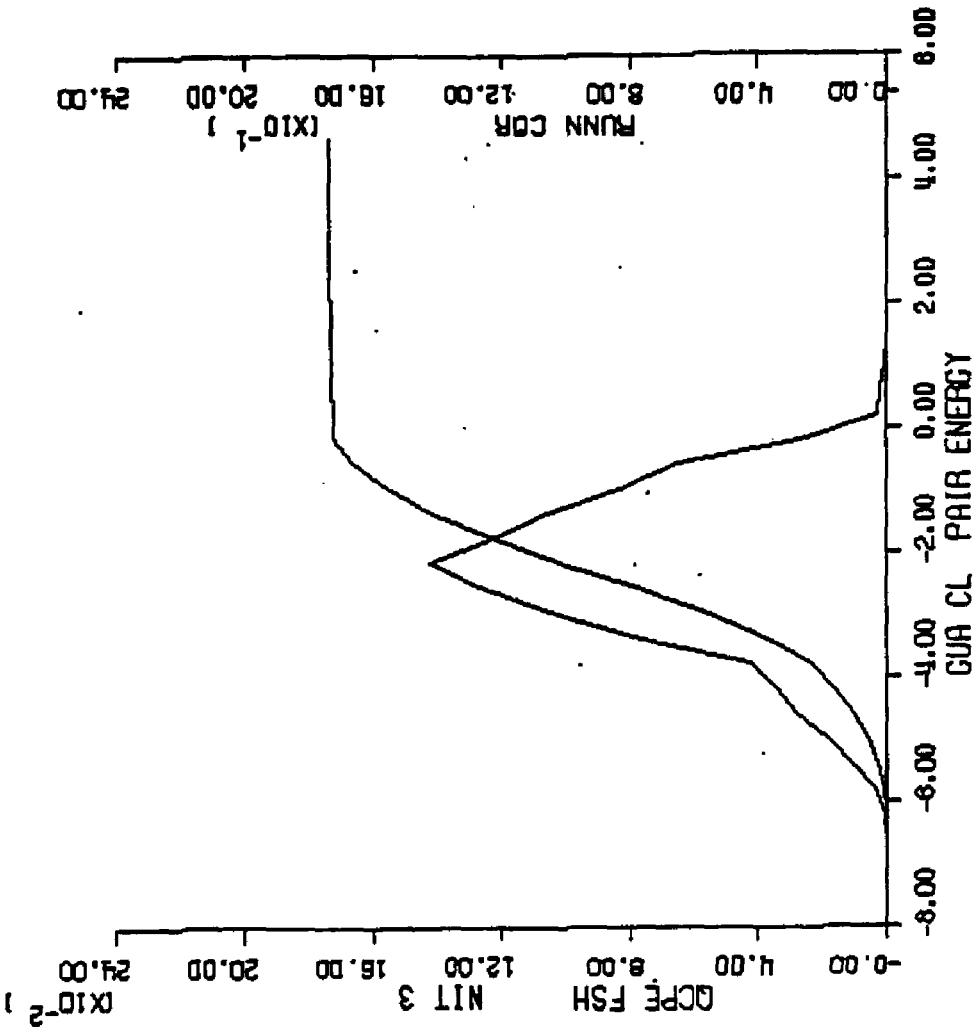


Figure IV.9.16 Calculated QCDF of solute-water pair energies of waters of the (CO)6 functional group of guanine - Clementi solute-water potential functions.

Figure IV.9.17 Calculated QCDF of solute-water pair energies of waters of the N7 ring nitrogen of guanine - Clementi solute-water potential functions.

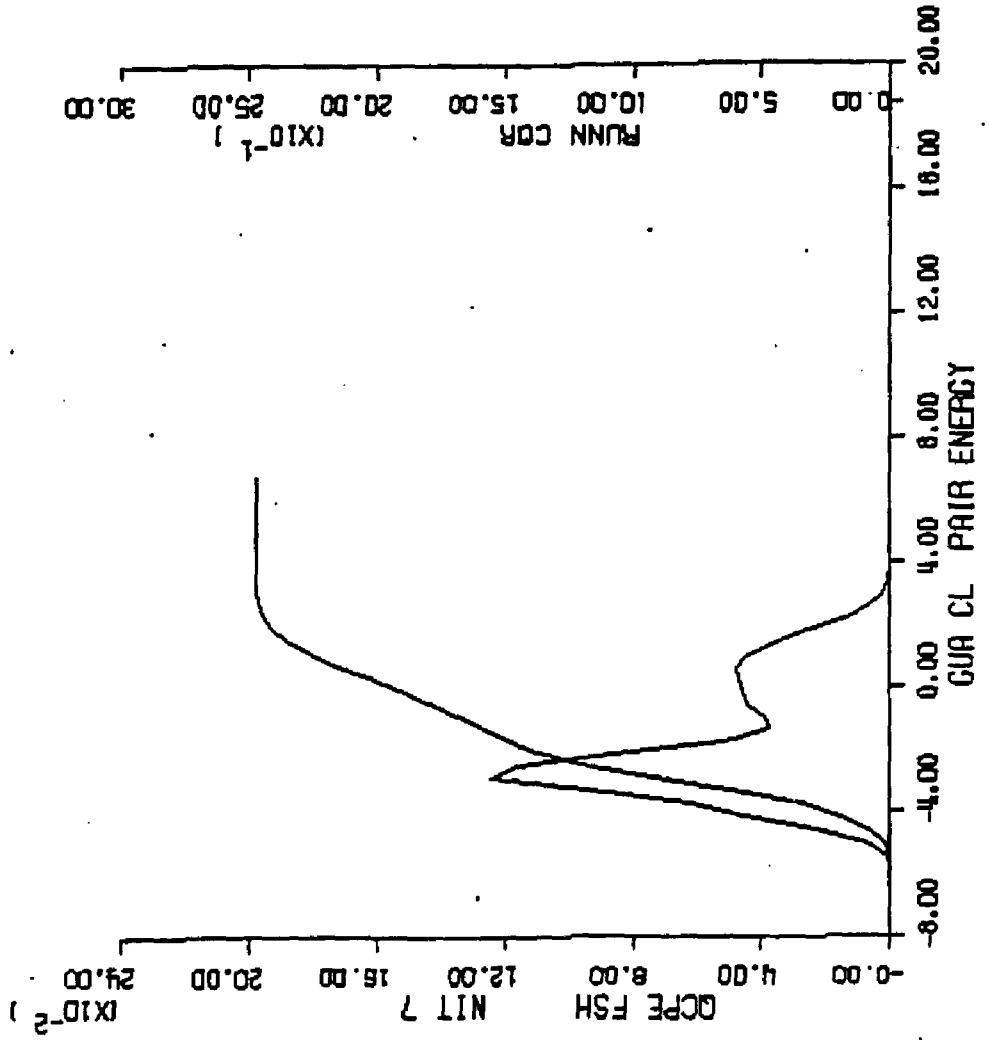


Figure IV.5.18 Calculated QCOF of solute-water pair energies of waters of the (CH)$_8$ functional group of guanine - Clementi solute-water potential functions.

Figure IV.9.19 Calculated QCDF of solute-water pair energies of waters of the (NH)9 functional group of guanine - Clementi solute-water potential functions.

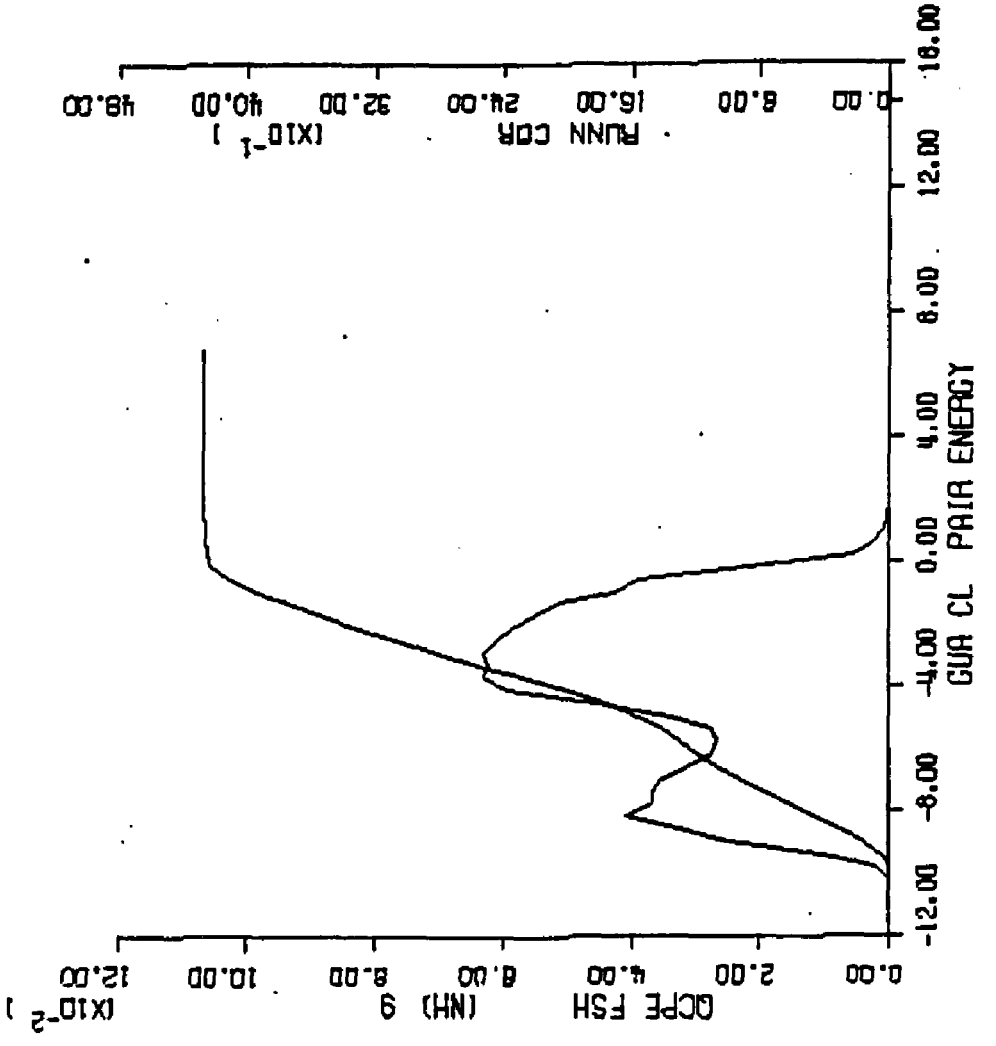
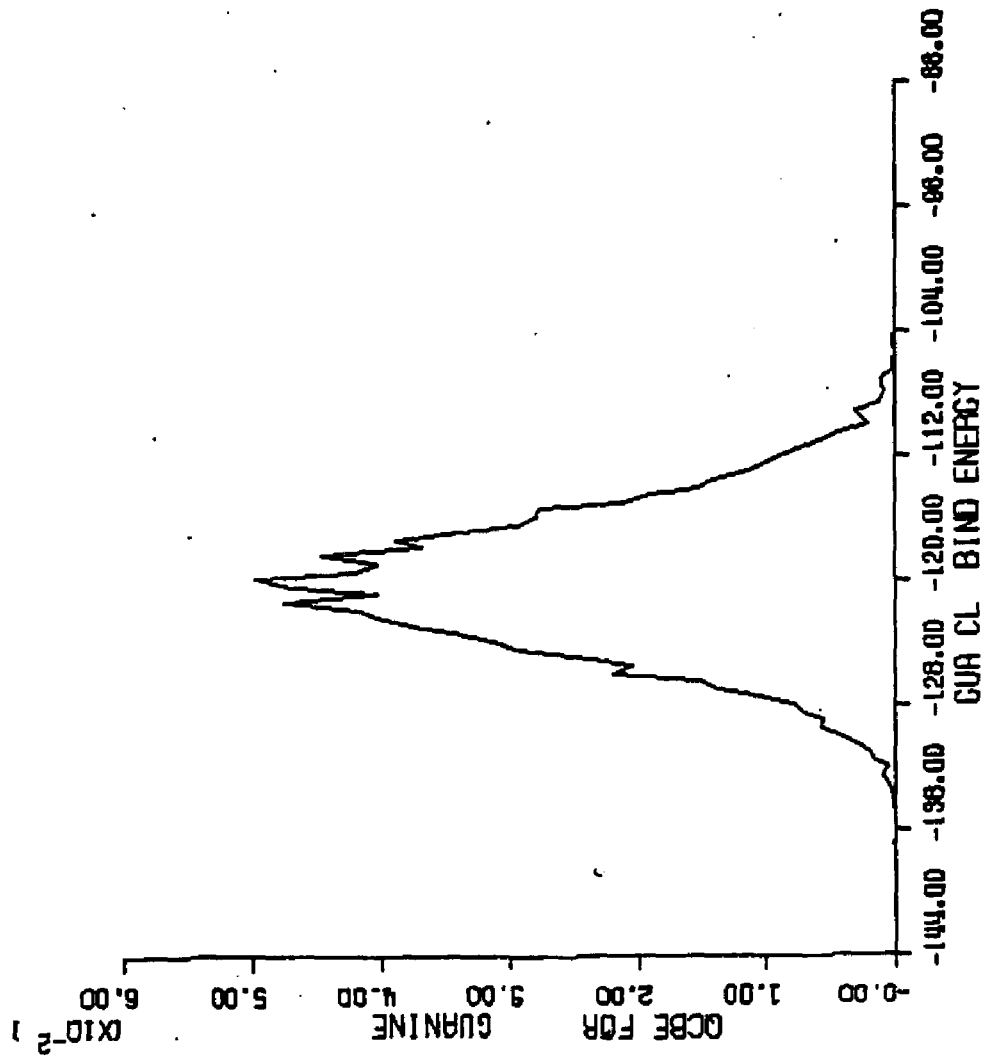


Figure IV.9.20 - calculated quasicomponent distribution function of binding energy for guanine (Clementi potentials).

X axis - binding energy (kcal/mole).

Y axis - quasicomponent of binding energy.

Figure IV.9.20 Calculated GCDF of total binding energy for guanine - Clementi solute-water potential functions.



10. (C3-Endo)-5-Deoxy-1-C-Amino-B-D-Ribo-Pentofuranose

Clementi Solute-Water Potentials

Computational Specifics. An aqueous solution of a ribose derivative and 202 waters was simulated using Clementi solute-water potential functions and MCY-CI water-water potential functions at 298 K. The sugar ring was puckered in the C3'-endo form. Face centered cubic boundary conditions with a unit cell edge of 14.64 Angstroms were used. The cell edge was obtained from partial molar volumes of 110 cc/mole for the ribose derivative and 18.07 cc/mole for the water molecules. The simulation was performed with an initial equilibration period of 500,000 non-force biased and 500,000 force biased moves which were discarded and a subsequent 2,000,000 force bias moves which was used to form all analyses and ensemble averages.

Potential Surface. The potential function surface for the the interaction of the ribose derivative and one water molecule using Clementi solute-water potentials is shown in Figure IV.10.1. The plane chosen for the calculation of contours is the best fit to the molecular ring The lowest solute-water interaction energy occurs at the H4' hydroxyl hydrogen atom at an energy of -7.87 kcal/mole. The hydrogen atom is slightly rotated above the countour plane. The oxygen atom lying essentially in the diagram plane is seen to have a surrounding region with solute-water interactions of from -3.87 to -5.67 kcal/mole. The broad area next to the ester oxygen also contains very favorable solute-water interactions;

energies here fall between -5.87 and -7.87 kcal/mole. Another strong interaction region is seen where one amino hydrogen lies in the contour plane. Two small contours are found next to this atom representing solute-water interactions of -5.87 to -7.87 kcal/mole. Finally, an apolar interaction region occurs peripheral to the substituent methy group. Interaction energies are positive here and are not more favorable than 0.12 kcal/mole.

Convergence and Thermodynamic Results. Control functions for the simulation of the aqueous ribose derivative are shown in Figure IV.10.2. The thermodynamic quantities calculated for the simulation are listed in Table IV.1. The mean total energy stabilizes around -1750 kcal/mole for the last 800K moves of the simulation and apparently converges to a value of -1750.2 +/- 6.9 kcal/mole after 2000K moves; at this point the heat capacity has the value 18.2 cal/mole-K. The free space to water transfer energy is -2.8 kcal/mole. Graphs of the quasicomponents of coordination number and first shell pair energies are displayed in Figures IV.10.4 thru IV.10.12 and IV.10.15 and IV.10.23 for the ester oxygen and methine, methyl, amino and hydroxyl functional groups of the ribose derivative. Figures IV.10.13 and IV.10.24 contain graphs of the quasicomponents of molecular coordination number and binding energy for this simulation.

Structural Results. The molecular diagram of Figure IV.10.3 displays the first shell coordination numbers found for the atoms of the ribose derivative. These numbers are obtained from the column labeled <K> in Table IV.11. The structural hydration of the ether and hydroxyl oxygens and hydroxyl hydrogens of the molecule is rather uniform. The ether oxygen (O1) of the sugar ring is populated by .77 waters whereas the O2' and O3' oxygen atoms of the hydroxyl groups contain .83 and 1.22 waters within their first shells, respectively. First shell cutoffs (corresponding first shell volumes) are essentially identical (O1: 3.6 A; O2': 3.4 A; O3': 3.6 A). The hydroxyl group hydrogens (H2' and H3') each have small first shell cutoffs of 2.2 Angstroms. These atoms also have virtually identical coordination numbers of one water (H2': 1.00; H3': 1.01).

The amino group hydrogens, H1' and H1'', differ markedly in their coordination numbers (H1': .91; H1'': 2.90). This variation presumably arises from the high degree of molecular asymmetry and consequent asymmetry of the hydration shell.

The hydrogens H1, H2, H3, and H4 are the methine hydrogens of the molecule. The coordination numbers of these atoms vary widely but are seen to be roughly proportional to the respsective first shell volumes (H1: 2.32 waters, 69.16 cubic A; H2: 1.08, 45.01; H3: .78,

37.47; H4, 1.30, 55.14). The methyl group similarly shows variations in the first shell populations of its hydrogens (H5': 1.73, H5'': 1.95, H5''' 3.14).

The QCDF of coordination number for the ribose derivative is found in Figure IV.10.13. The coordination numbers range from 17 to 27 and the greatest contributions are from coordiation numbers of 21 and 22. The average molecular coordination number is 21.30 +/- 0.88.

Energetic Results. First shell pair energies for waters assigned to the atoms of the ribose derivative are displayed in the molecular diagram of Figure IV.10.14. These values are also listed in the column labeled <K> in Table IV.11. The waters of the three oxygen atoms of the molecule are found to bind to the solute at similar energies. The waters of the O1 ether oxygen bind at -3.644 kcal/mole while those of the hydroxyl oxygens, O2' and O3', have average pair energies of -3.536 and -2.572 kcal/mole respectively.

The waters of the hydroxyl hydrogen H2' are bound to the solute more favorably than any other waters of the system at an average energy of -6.967 kcal/mole each. Consequently, these waters contribute the greatest solute-water binding energies to the system (-7.022 kcal/mole). The H3' hydroxyl waters are less strongly bound at an average -4.163 kcal/mole.

The waters of the amino hydrogens, H1' and H1'', have quite different pair energy values of -4.355 and -1.406 kcal/mole. The weaker average pair energy of the H1'' waters correlates with the increased number of waters in the coordination shell of the atom; these additional waters interact less favorably with the solute.

The solute-water pair interactions of the waters of the apolar hydrogens are very weak. The average first shell pair energies of the methine hydrogens range from -.502 (H1) to .004 (H4) kcal/mole. The corresponding energy values for the methyl hydrogens are equally unfavorable (H5': .12 kcal/mole,; H5'': -.907 kcal/mole; H5''': .064 kcal/mole).

The graph of the quasicomponents of solute binding energy for the ribose derivative simulation is Figure IV.10.24. Energies range from -15 to 14 kcal/mole with significant contributions in the range -6 to 2 kcal/mole. The average solute binding energy is -1.7 +/- 1.6 kcal/mole.

Figure IV.10.1 - isoenergy contour surface for ribose derivative and one water (Clementi solute-water potential functions).

X axis - X axis of plane defined by molecular ring atoms (Angstroms).

Y axis - Y axis of plane defined by molecular ring atoms (Angstroms).

Contours - isoenergy contours with one kcal/mole increments with alphabetical labels referring to contour energy values in list at right.

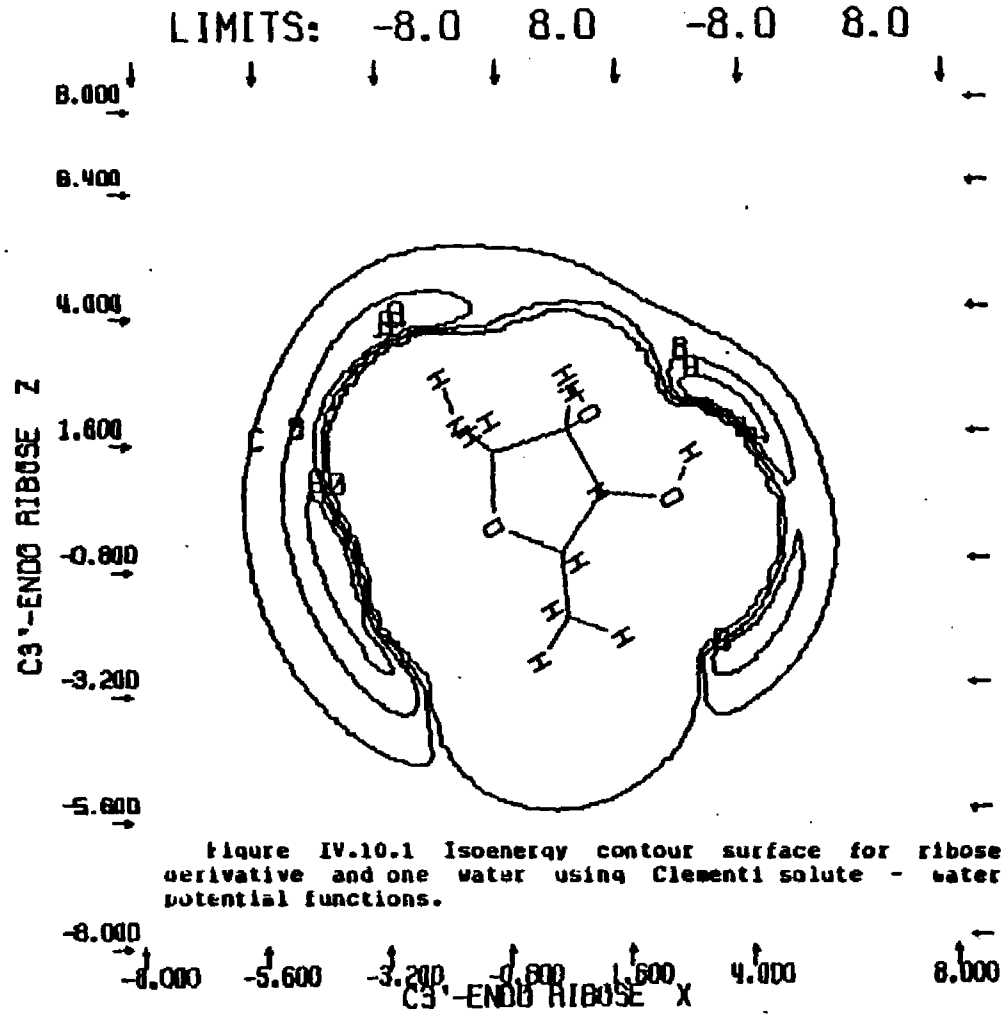


Figure IV.10.1 Isoenergy contour surface for ribose derivative and one water using Clementi solute - water potential functions.

Figure IV.10.2 - control functions for Monte Carlo simulation of ribose derivative and 202 waters (Clementi solute-water potential functions).

X axis - number of configurations.

Left Y axis - mean total energy (Kcal/mole).

Right Y axis. - constant volume heat capacity (cal/mole-degree).

Upper curve - constant volume heat capacity.

Bottom curve without crosshatches - average total energy for entire simulation.

bottom curve with crosshatches - average total energy for preceeding 50K configurations.

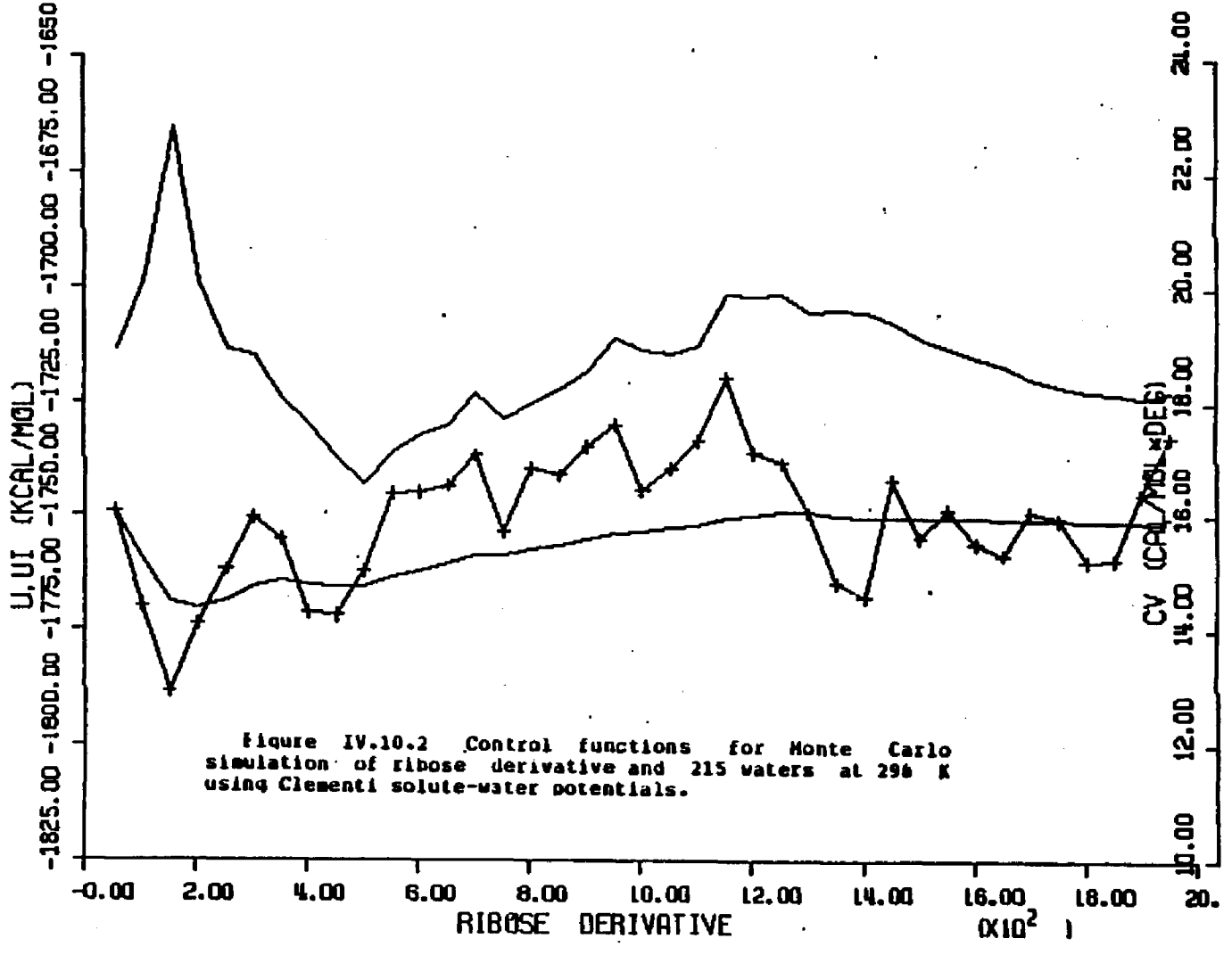


Figure IV.10.2 Control functions for Monte Carlo simulation of ribose derivative and 215 waters at 298 K using Clementi solute-water potentials.

Table IV.11 Calculated structural and energetic quantities from Monte Carlo simulation of C1'a-ino-C4'-methyl-ribose derivative and 202 waters at 298 K - Clementi solute-water potential functions.

RIBOSE DERIVATIVE IN WATER AT 298K - FORCE BIAS AND CLEMENTI POTENTIAL FUNCTIONS

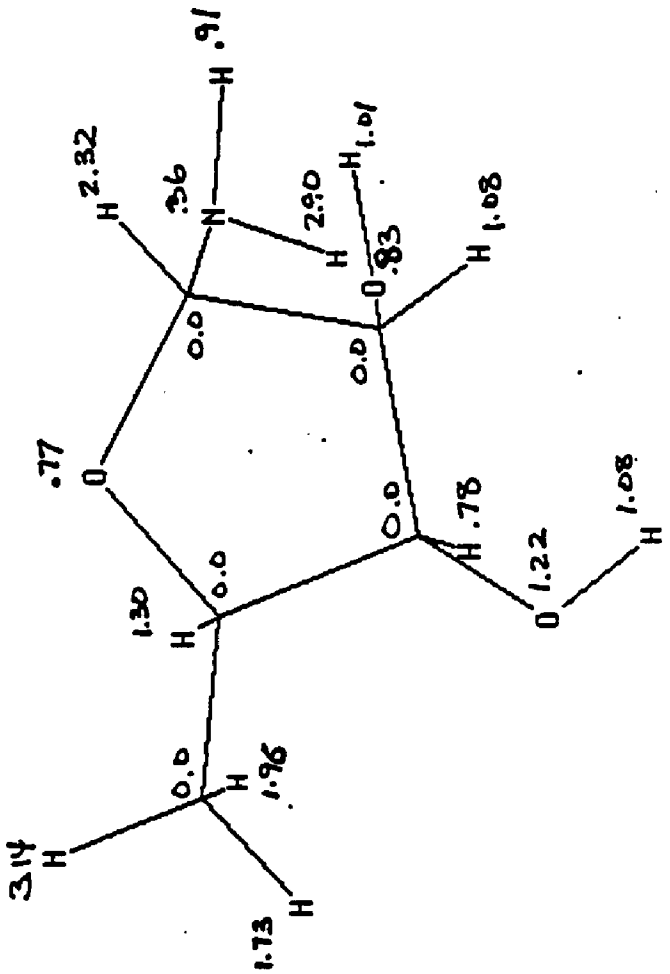
LAST CONFIGURATION: 2000001					FIRST SHELL SOLUTE PROPERTIES					TOTAL SLT PROPS		WATER PROPERTIES RFSH=3.30 RCH= 7.75 A		
INDEX	TYPE				RFS	VFS	<K>	<K/V>	<SLTHE>	<SLTPE>	<K>	<SLTHE>	<HH>	<HWSPE>	<HWWT>
METHYNE GROUP															
1	2	8	C		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	6	59	H		4.2	69.16	2.32	1.00	-1.166	-0.502	14.92	0.360	4.14	-3.006	-17.338
TOTALS FOR FUNCTIONAL GROUP				>CH- :	69.16	2.32	1.00	-1.166	-0.502	14.92	0.360	4.14	-3.006	-17.338	
3	3	17	C		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	8	59	H		4.0	45.00	1.12	0.74	0.392	0.351	7.05	1.939	4.13	-3.106	-17.275
TOTALS FOR FUNCTIONAL GROUP				>CH- :	45.00	1.12	0.74	0.392	0.351	7.05	1.939	4.13	-3.106	-17.275	
5	4	17	C		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	10	59	H		3.0	37.49	0.75	0.60	-0.257	-0.341	4.94	0.340	4.11	-3.094	-17.092
TOTALS FOR FUNCTIONAL GROUP				>CH- :	37.49	0.75	0.60	-0.257	-0.341	4.94	0.340	4.11	-3.094	-17.092	
7	5	17	C		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8	12	59	H		3.0	55.13	1.37	0.74	0.029	0.021	13.06	2.016	4.10	-3.105	-17.323
TOTALS FOR FUNCTIONAL GROUP				>CH- :	55.13	1.37	0.74	0.029	0.021	13.06	2.016	4.10	-3.105	-17.323	
AVERAGES OVER FUNCTIONAL GRP				>CH- :	51.69	1.39	0.80	-0.250	-0.110	10.20	1.166	4.12	-3.110	-17.257	
STATISTICAL UNCERTAINTY (+/- 2*SD):						0.09	0.05	0.030	0.021	0.01	1.001	0.03	0.005	0.210	
METHYL GROUP															
9	13	6	C		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10	18	3	H		3.4	57.61	1.75	0.91	0.046	0.026	16.93	2.921	4.12	-2.995	-17.130
11	19	3	H		3.6	60.64	1.95	0.96	-1.002	-0.925	15.11	0.010	4.11	-3.029	-17.021
12	20	3	H		4.0	95.04	3.12	0.90	0.256	0.002	23.01	3.571	4.20	-3.090	-17.610
TOTALS FOR FUNCTIONAL GROUP				-CH3 :	213.29	6.02	0.96	-1.501	-0.220	55.04	6.510	4.15	-3.040	-17.307	
STATISTICAL UNCERTAINTY (+/- 2*SD):						0.41	0.06	0.102	0.035	0.03	0.595	0.03	0.071	0.100	
AMINE GROUP															
13	7	11	N		3.4	17.59	0.37	0.62	-1.503	-4.090	3.60	-1.163	4.22	-2.966	-16.604
14	14	1	H		2.2	20.30	0.91	1.33	-3.936	-4.334	15.70	-2.002	4.11	-3.007	-17.064
15	15	1	H		3.6	75.53	2.90	1.15	-4.090	-1.411	22.66	-0.000	4.17	-2.973	-17.003
TOTALS FOR FUNCTIONAL GROUP				-NH2 :	113.51	4.17	1.10	-9.529	-2.202	41.95	-4.063	4.15	-3.015	-16.990	
STATISTICAL UNCERTAINTY (+/- 2*SD):						0.32	0.00	1.314	0.469	0.03	7.400	0.03	0.001	0.204	

RIBOSE DERIVATIVE IN WATER AT 298K - FORCE BIAS AND CLEMENTI POTENTIAL FUNCTIONS

LAST CONFIGURATION: 2000001				FIRST SHELL SOLUTE PROPERTIES					TOTAL SLT PROPS		WATER PROPERTIES RFSW=3.38 RCE= 7.75 A		
ESTER OXYGEN													
16	1	68	O	3.6	33.66	0.76	0.67	-2.713	-3.591	4.35	-2.521	4.48	-3.037	-18.838
STATISTICAL UNCERTAINTY (+/- 2*SD):						0.14	0.12	0.879	1.733	0.81	11.922	0.12	0.253	0.671
HYDROXYL GROUP														
17	9	9	O	3.4	26.44	0.82	0.92	-2.893	-3.544	3.61	-2.673	3.96	-2.984	-16.213
18	16	57	H	2.2	23.54	1.01	1.28	-7.844	-6.982	25.58	-2.483	4.07	-3.058	-16.843
TOTALS FOR FUNCTIONAL GROUP -OH :					49.97	1.83	1.09	-9.938	-5.444	29.19	-5.076	4.06	-3.049	-16.765
19	11	9	O	3.6	33.87	1.22	1.11	-3.142	-2.572	5.11	-2.838	4.04	-2.978	-16.419
20	17	57	H	2.2	23.54	1.00	1.27	-4.136	-4.149	24.77	2.528	4.13	-3.079	-17.294
TOTALS FOR FUNCTIONAL GROUP -OH :					56.61	2.22	1.17	-7.278	-3.288	29.88	-0.311	4.12	-3.062	-17.143
AVERAGES OVER FUNCTIONAL GRP -OH :					53.29	2.02	1.13	-8.688	-4.362	29.54	-2.693	4.09	-3.056	-16.954
STATISTICAL UNCERTAINTY (+/- 2*SD):						0.16	0.09	1.286	0.918	0.82	3.457	0.03	0.069	0.171
MOLECULAR SUM/AVERAGE:					673.8	21.36	0.95	-31.968	-1.496	282.88	-1.598	4.13	-3.058	-17.151
STATISTICAL UNCERTAINTY (+/- 2*SD):						0.73	0.03	1.948	0.136	0.86	1.189	0.02	0.037	0.094

Table IV.11 Calculated structural and energetic quantities from Monte Carlo simulation of C1'amino-C4'-methyl-ribose derivative and 202 waters at 298 K - Clementi solute-water potential functions.

Figure IV.10.3 First shell coordination numbers for the atoms of ribose derivative - Clementi solute-water potential functions.



Figures IV.10.4 through IV.10.13 - calculated quasicomponent distribution function of first shell coordination number for ribose derivative and functional groups (Clementi potentials).

X axis - coordination number.

Y axis - quasicomponent of coordination number.

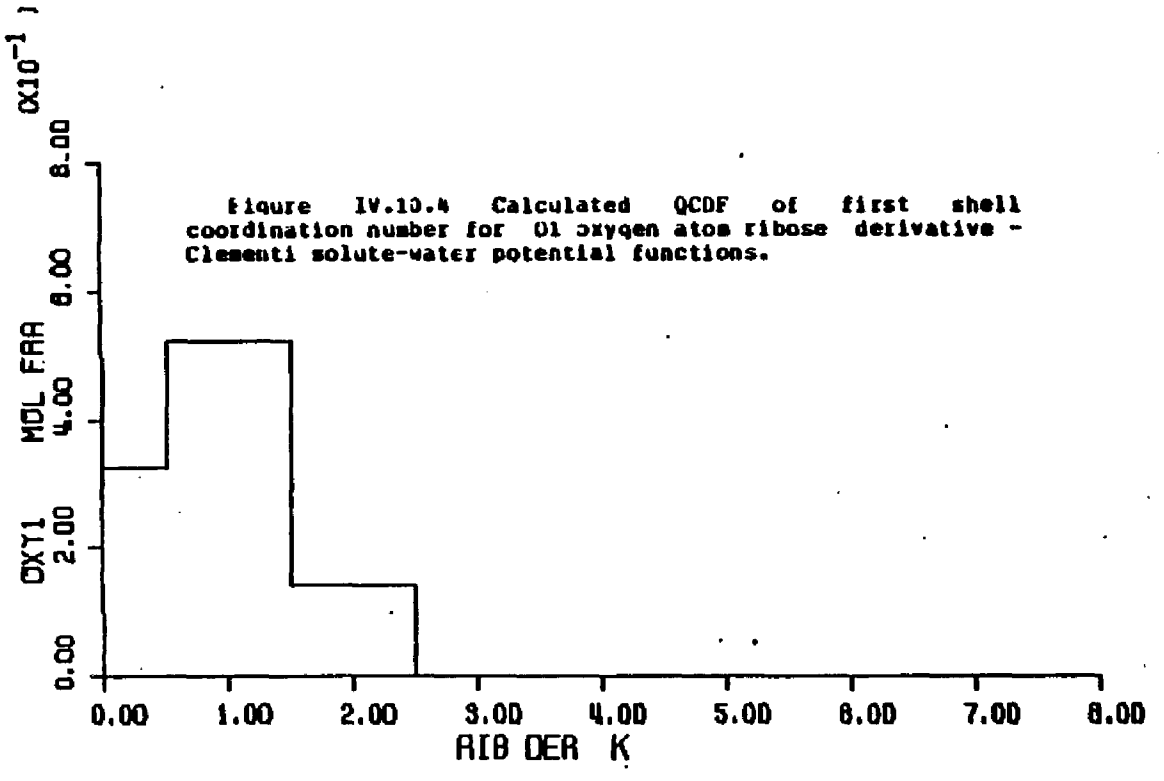


Figure IV.10.4 Calculated QCDF of first shell coordination number for O1 oxygen atom ribose derivative - Clementi solute-water potential functions.

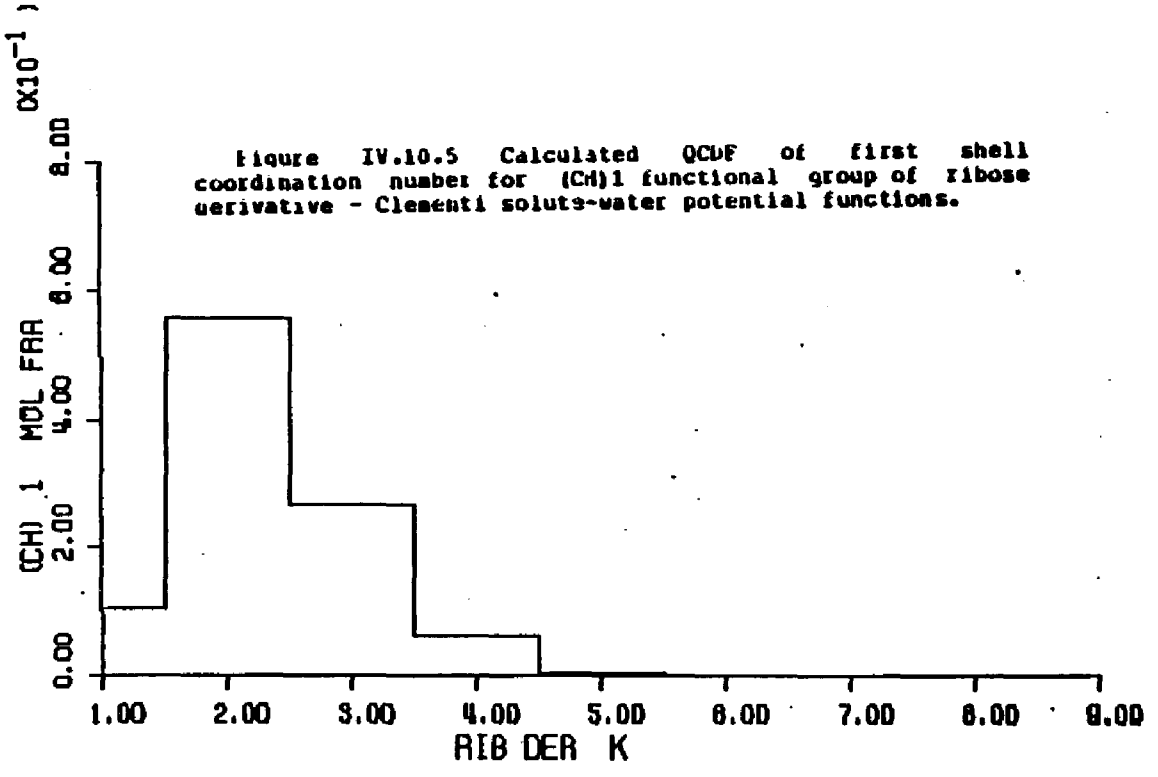


Figure IV.10.5 Calculated QCDF of first shell coordination number for (CH)1 functional group of ribose derivative - Clementi solute-water potential functions.

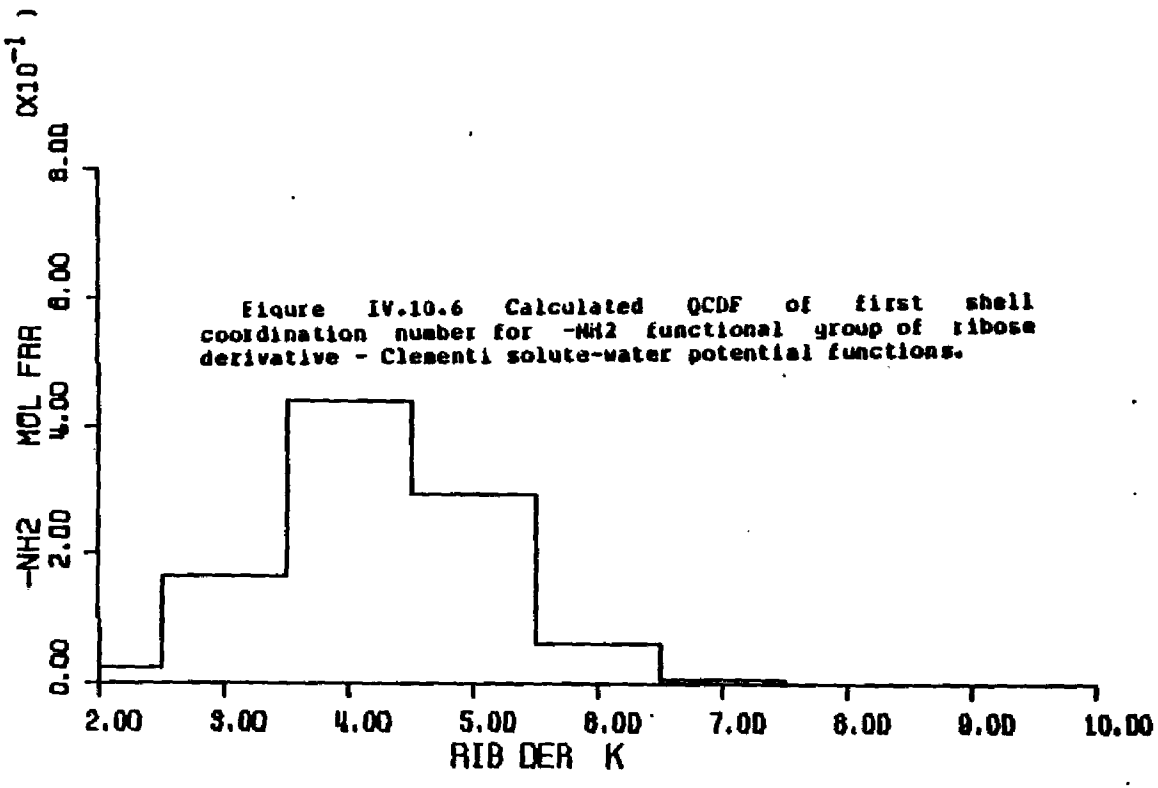


Figure IV.10.6 Calculated QCDF of first shell coordination number for -NH2 functional group of ribose derivative - Clementi solute-water potential functions.

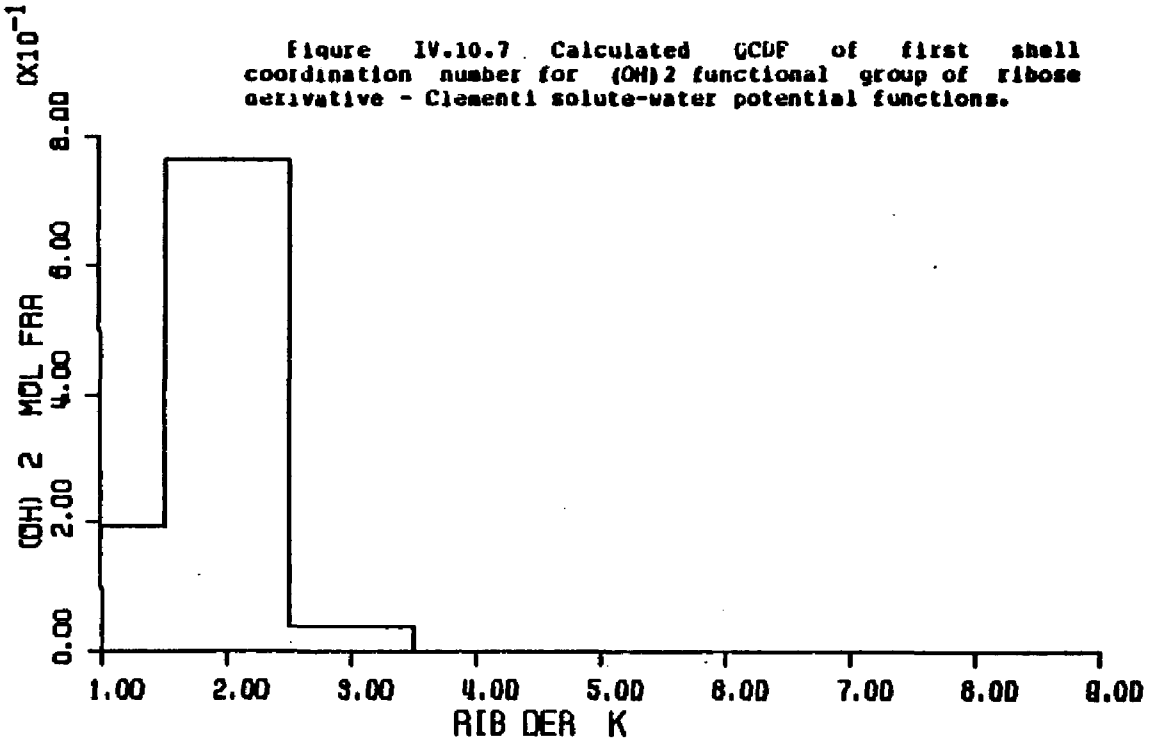


Figure IV.10.7 Calculated GCDF of first shell coordination number for (OH)2 functional group of ribose derivative - Clementi solute-water potential functions.

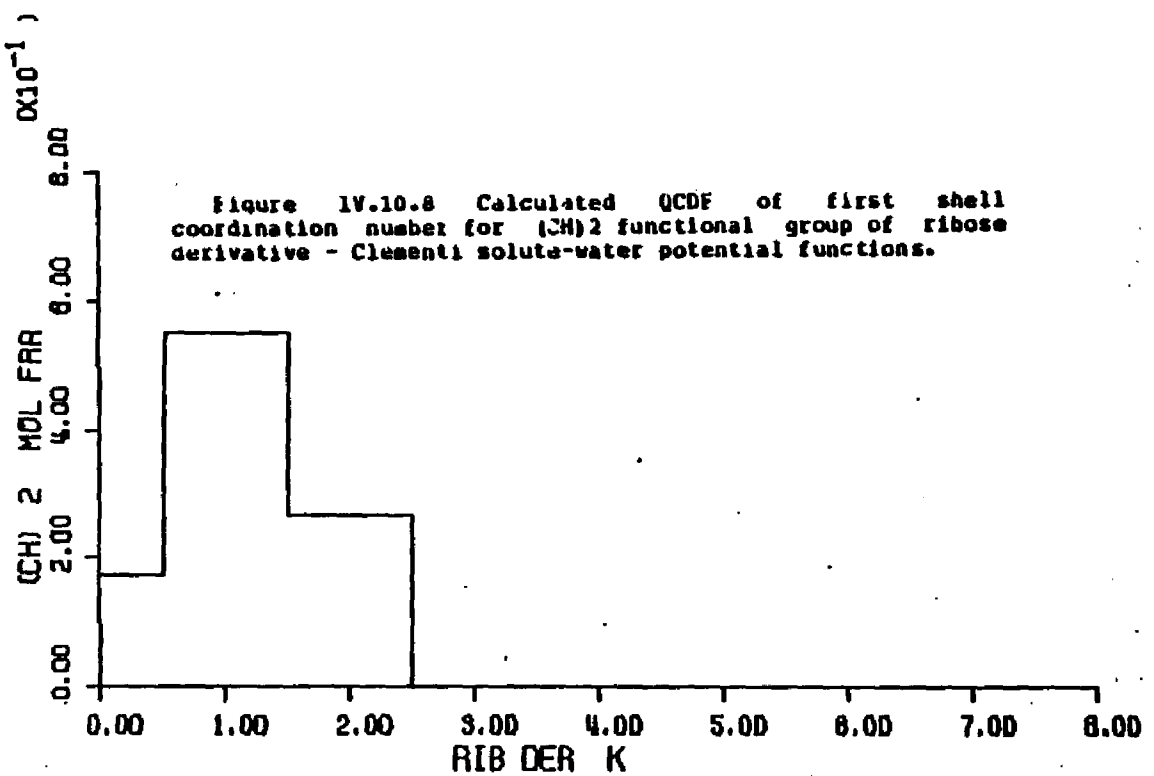


Figure IV.10.8 Calculated QCDF of first shell coordination number for (CH)2 functional group of ribose derivative - Clementi solute-water potential functions.

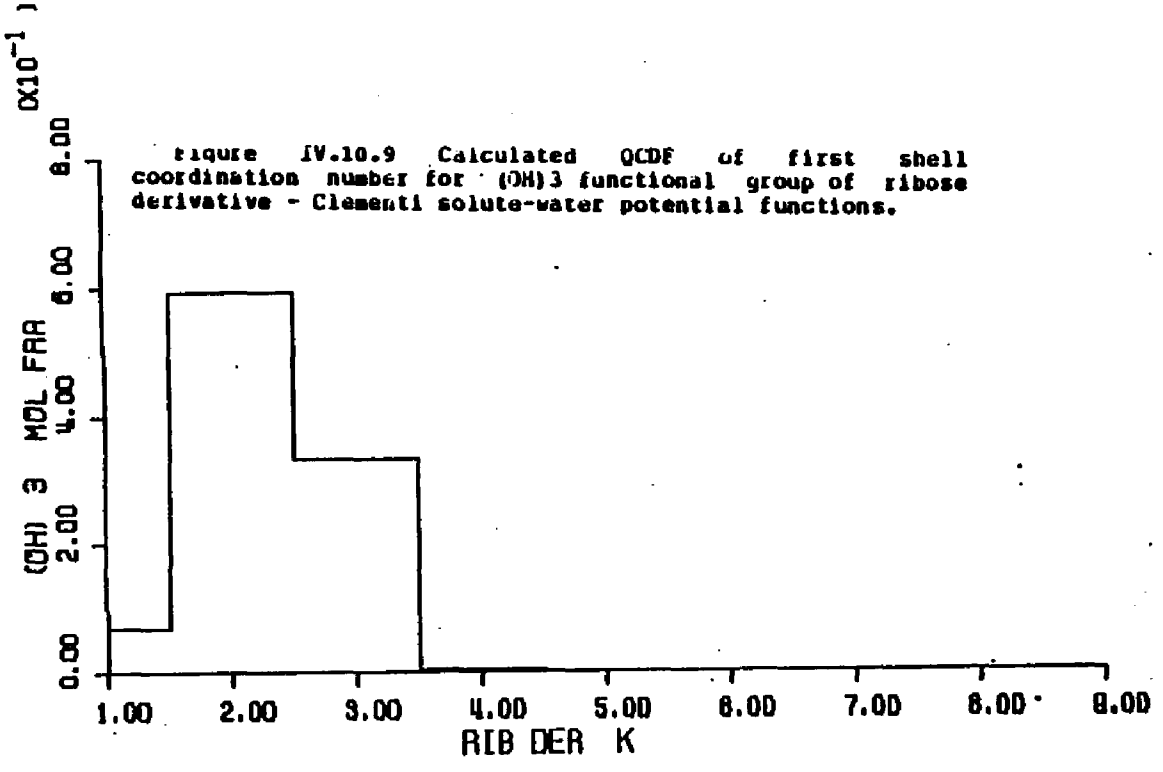


Figure IV.10.9 Calculated QCDF of first shell coordination number for (OH)3 functional group of ribose derivative - Clementi solute-water potential functions.

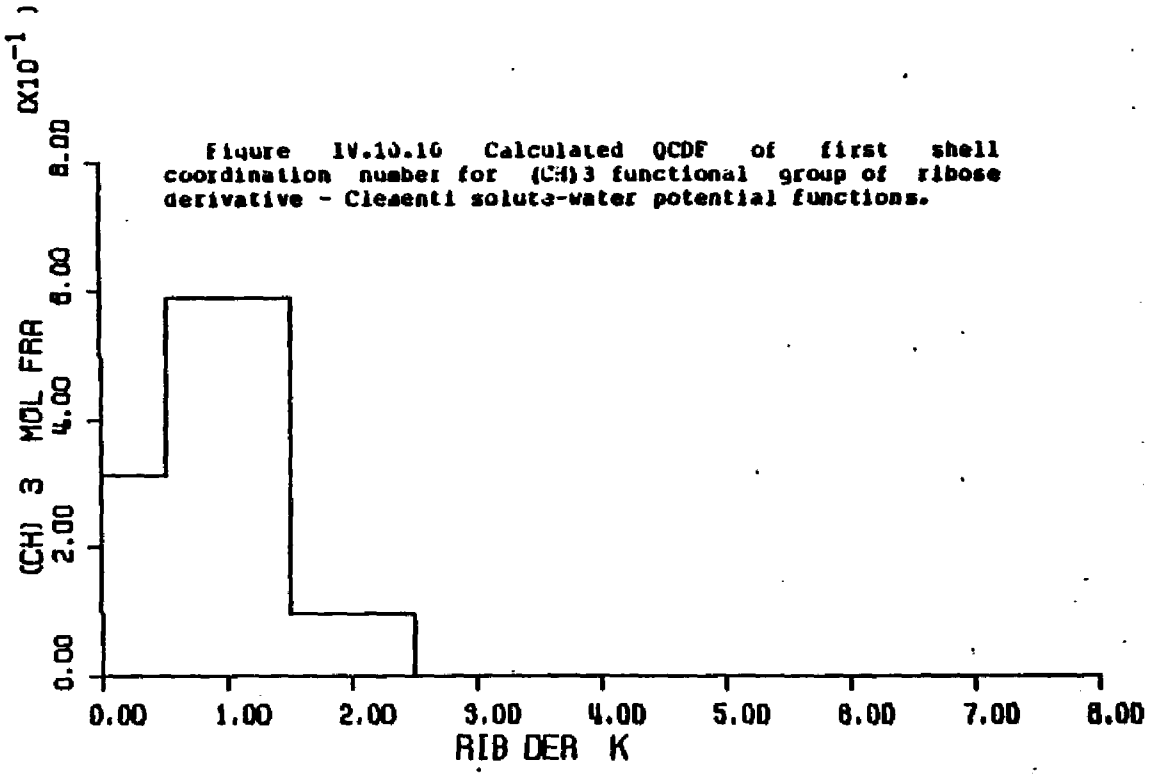


Figure IV.10.10 Calculated QCDF of first shell coordination number for (CH)3 functional group of ribose derivative - Clementi solute-water potential functions.

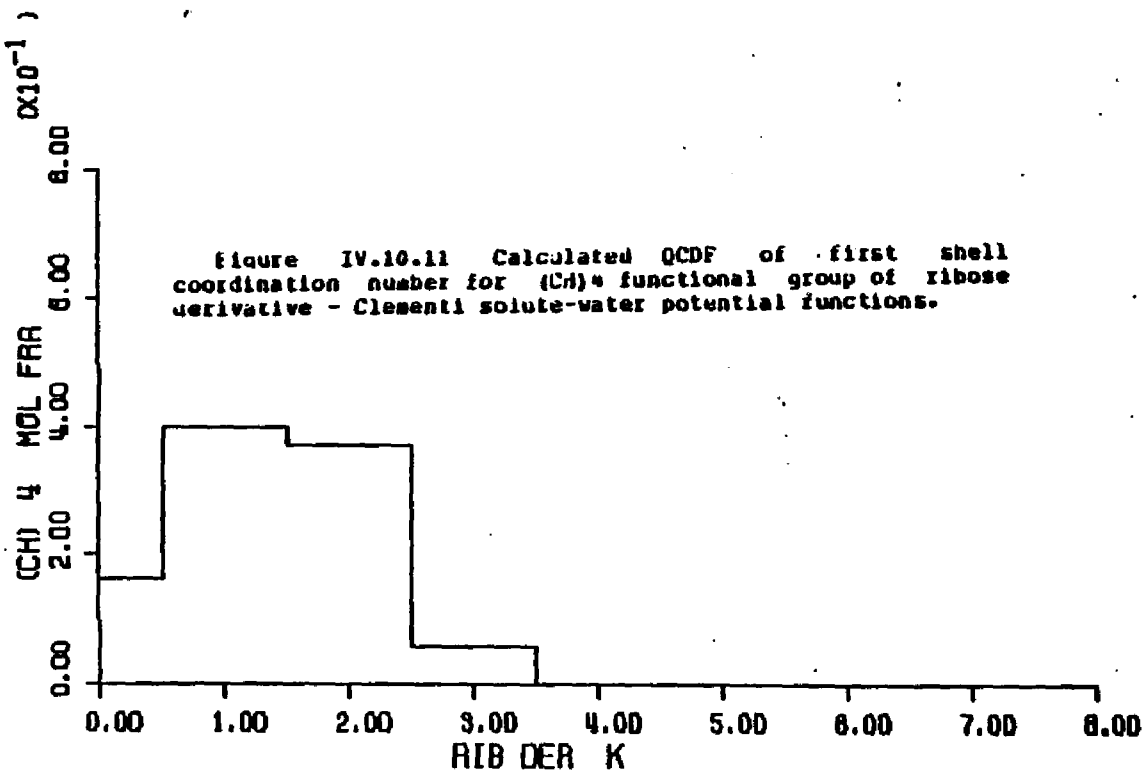


figure IV.10.11 Calculated QCDF of first shell coordination number for (CH)4 functional group of ribose derivative - Clementi solute-water potential functions.

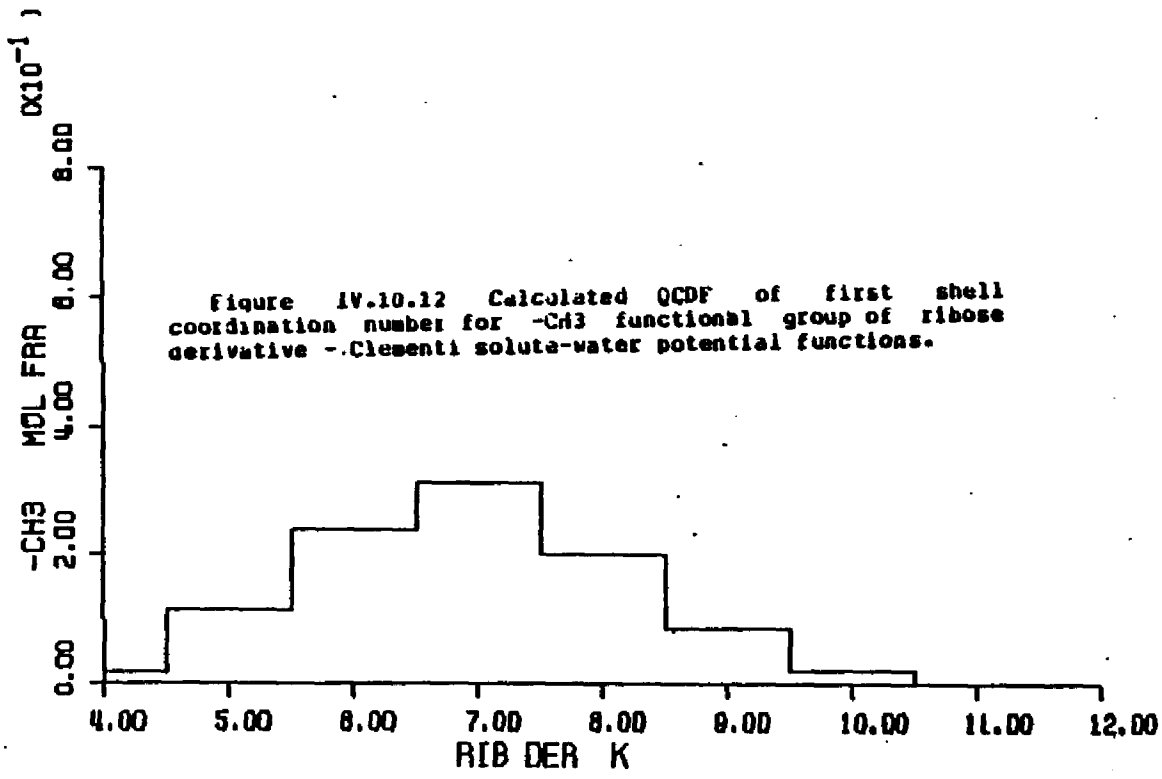


Figure IV.10.12 Calculated QCDF of first shell coordination number for -Cd3 functional group of ribose derivative -.Clementi solute-water potential functions.

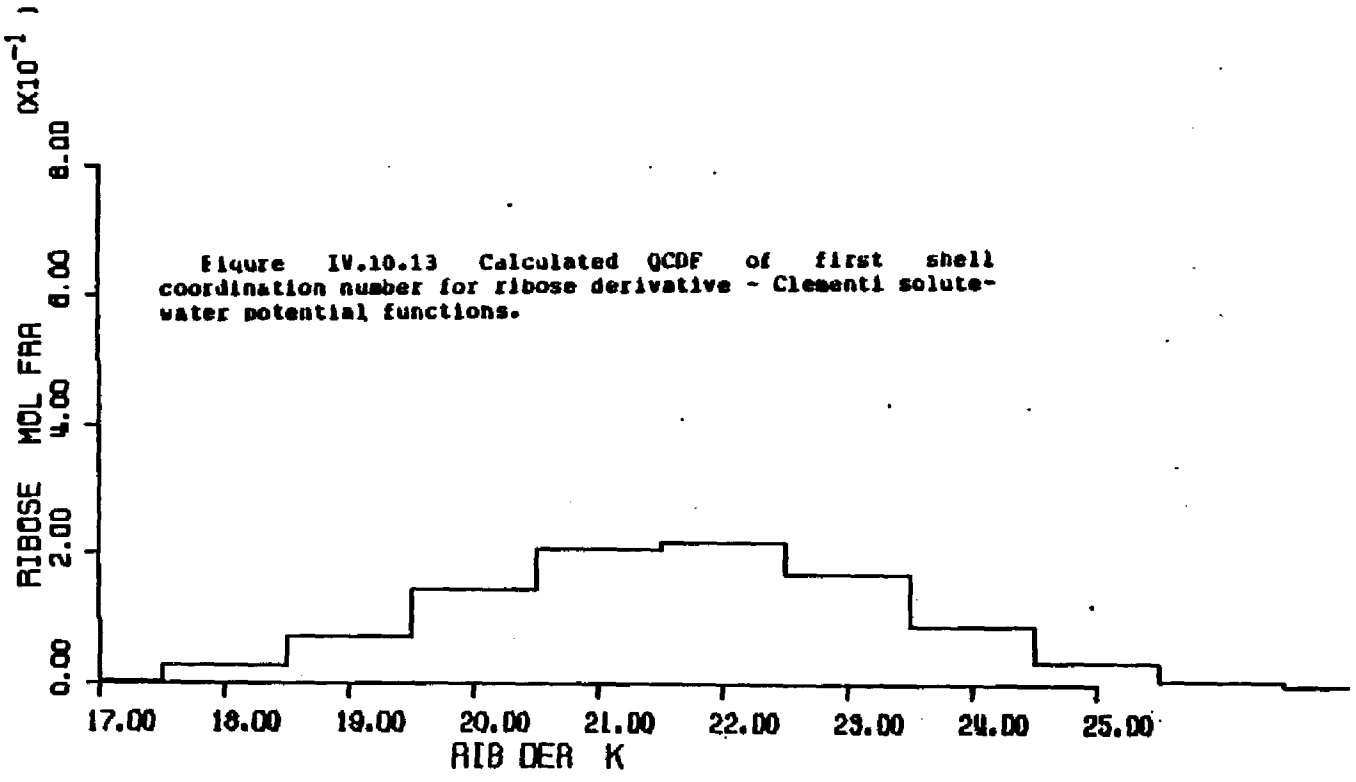


Figure IV.10.13 Calculated QCDF of first shell coordination number for ribose derivative - Clementi solute-water potential functions.

Figure IV.10.14 Average first shell solute-water pair energies of waters assigned to the atoms of ribose derivative - Clementi solute-water potential functions.

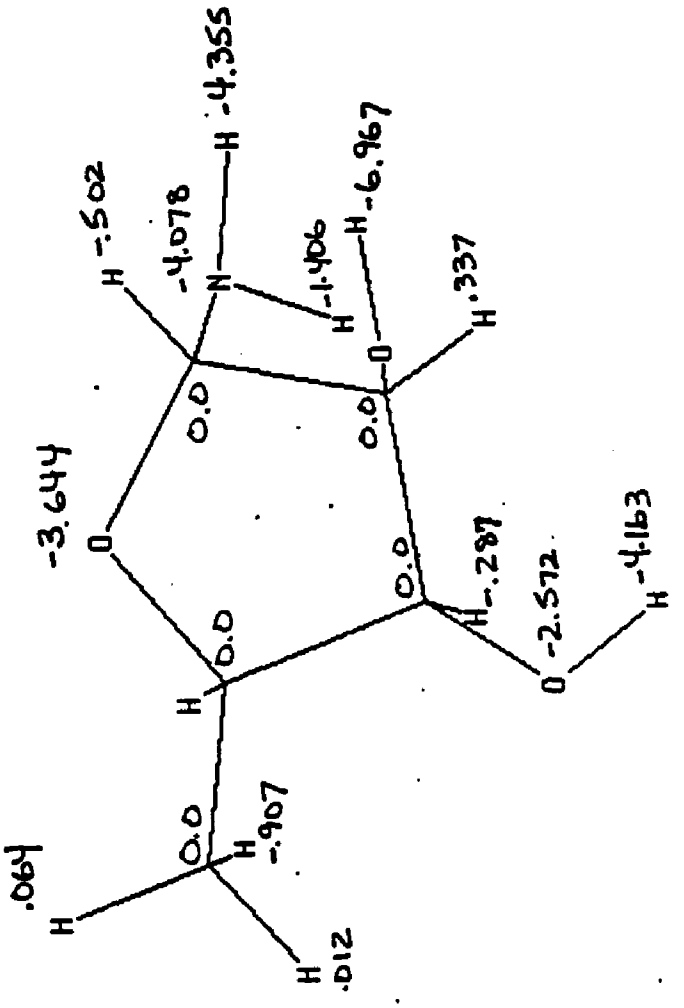


Figure IV.10.15 through IV.10.23 - calculated quasicomponent distribution function of average first shell pair energy for functional groups of ribose derivative (Clementi potentials).

X axis - pair energy (kcal/mole).

Left Y axis - quasicomponent of pair energy.

Right Y axis - running coordination number.

Figure IV.10.15 Calculated HOMO of solute-water pair derivative - Clementi solute-water potential functions.
energies of waters of the 11 oxygen atom group of ribose

Figure IV.10.17 Calculated QCDF of solute-water pair energies of waters of the -NH2 functional group of ribose derivative - Clementi solute-water potential functions.

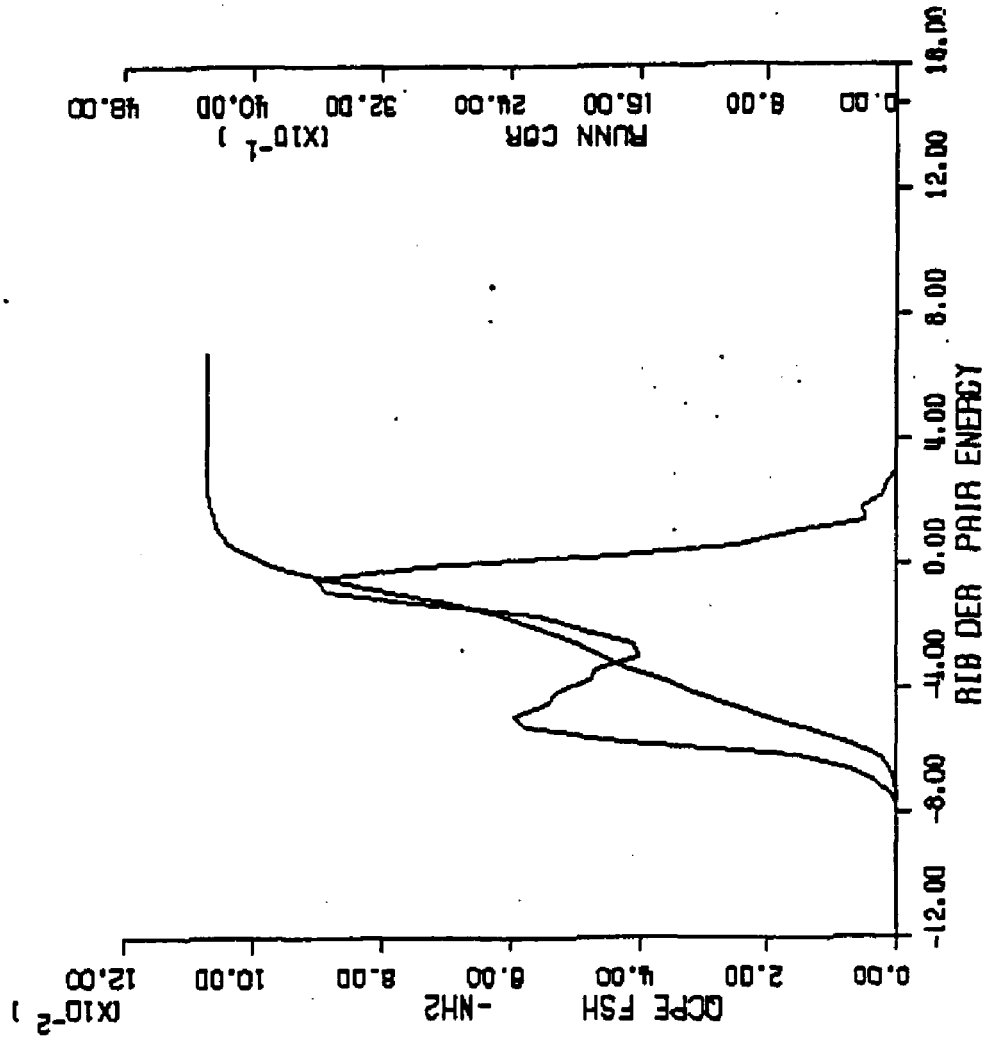


Figure IV.10.22 Calculated QCOF of solute-water pair energies of waters of the (CH)$_4$ functional group of ribose derivative - Clementi solute-water potential functions.

Figure IV.10.16 Calculated QCDF of solute-water pair energies of waters of the (CH)1 functional group of ribose derivative - Clementi solute-water potential functions.

Figure IV.10.18 Calculated QCDF of solute-water pair energies of waters of the (OH)2 functional group of ribose derivative - Clementi solute-water potential functions.

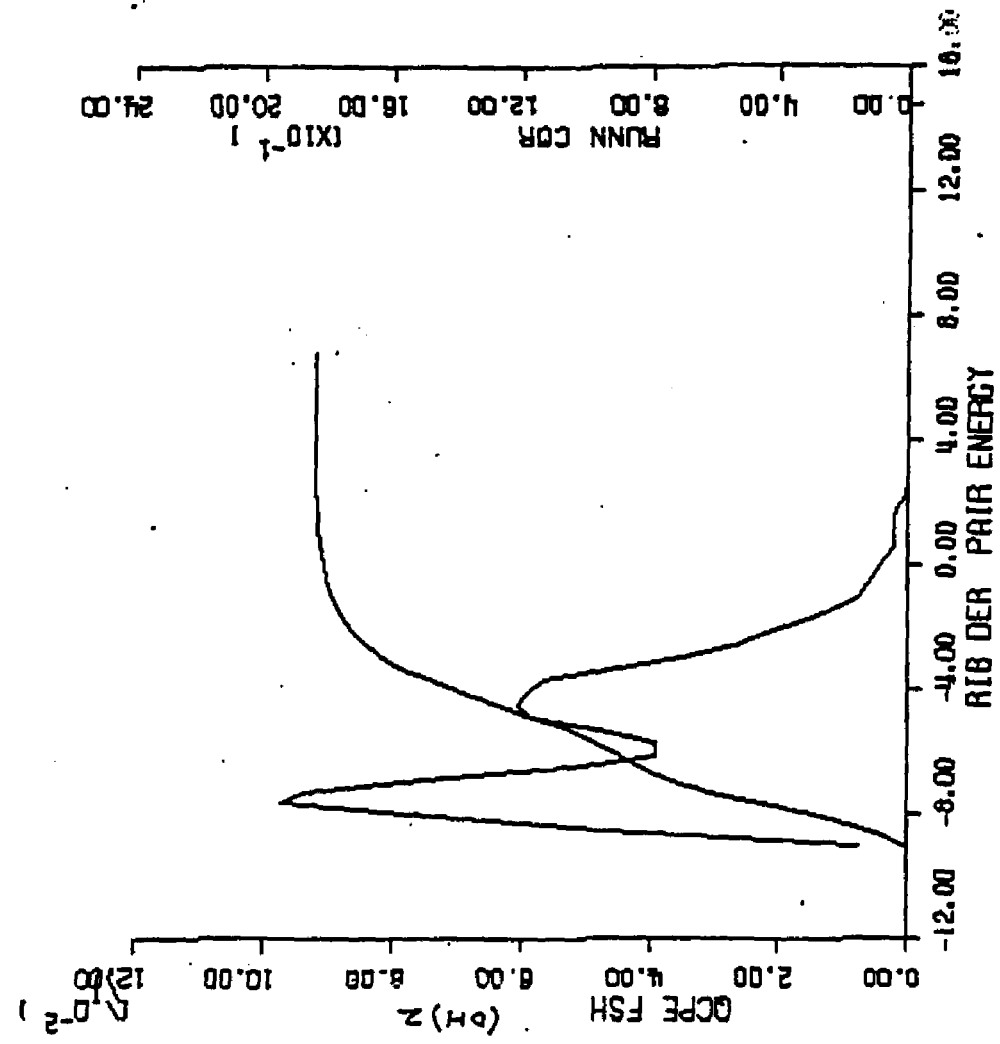


Figure IV.10.19 Calculated QCOF of solute-water pair energies of waters of the (CH)2 functional group of ribose derivative - Clementi solute-water potential functions.

Figure IV.10.20 Calculated QCDF of solute-water pair energies of waters of the (OH)3 functional group of ribose derivative - Clementi solute-water potential functions.

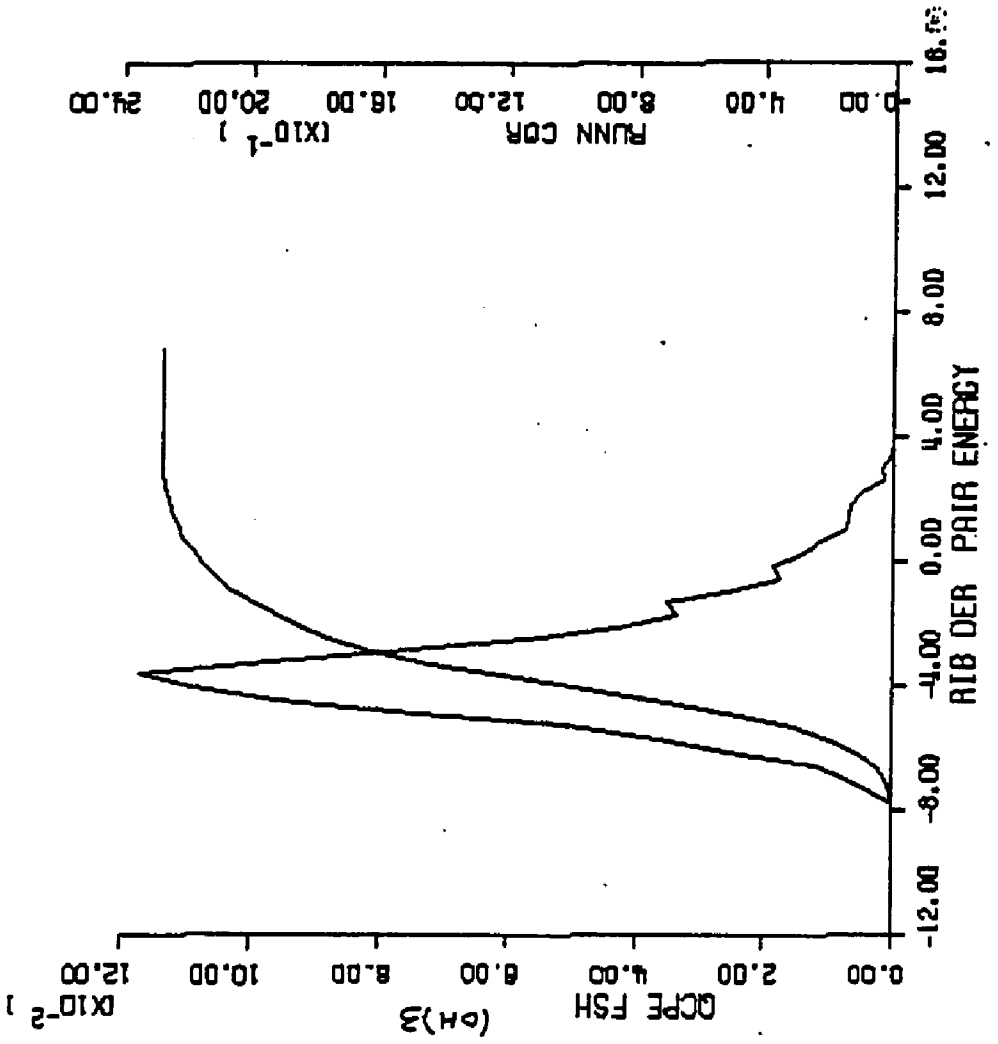


Figure IV.10.21 Calculated QCDF of solute-water pair energies of waters of the $(CH)_3$ functional group of ribose derivative - Clementi solute-water potential functions.

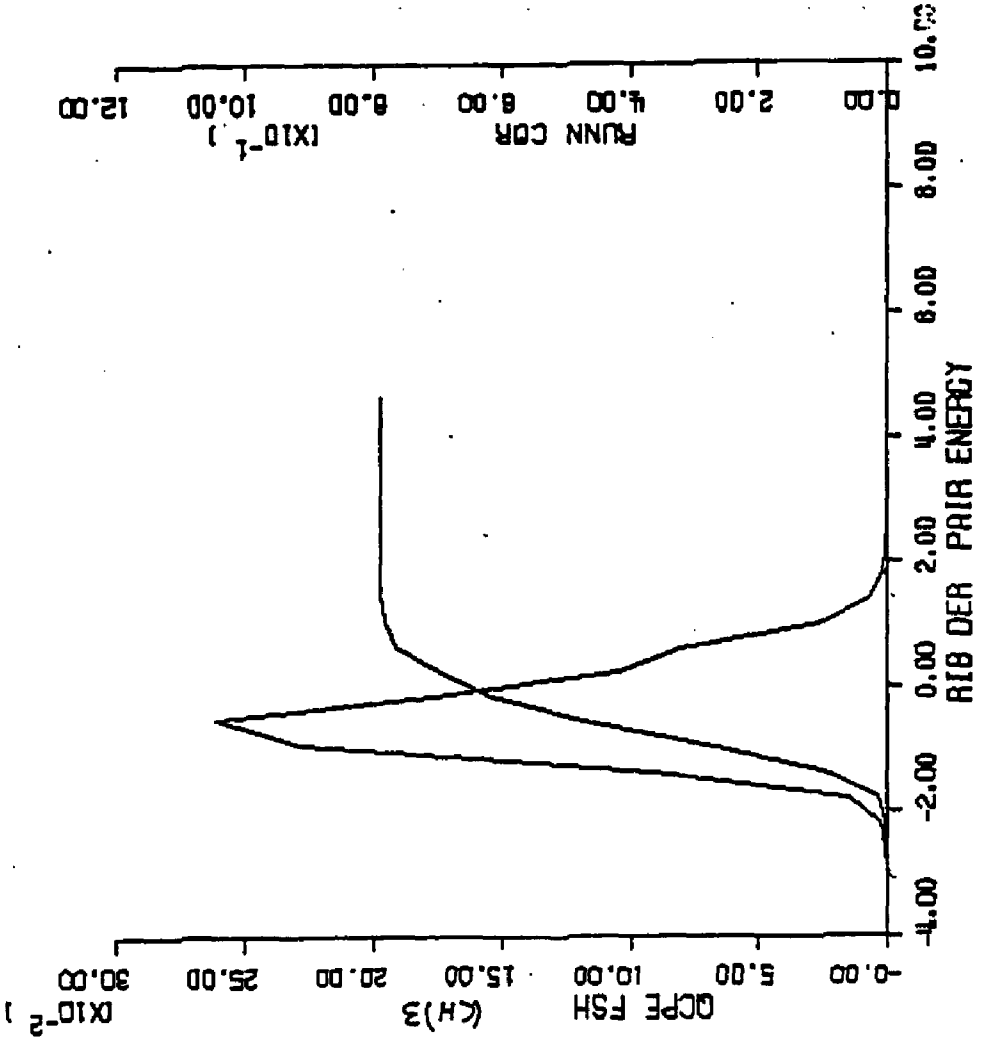


Figure IV.10.23 Calculated QCDF of solute-water pair energies of waters of the -CH3 functional group of ribose derivative - Clementi solute-water potential functions.

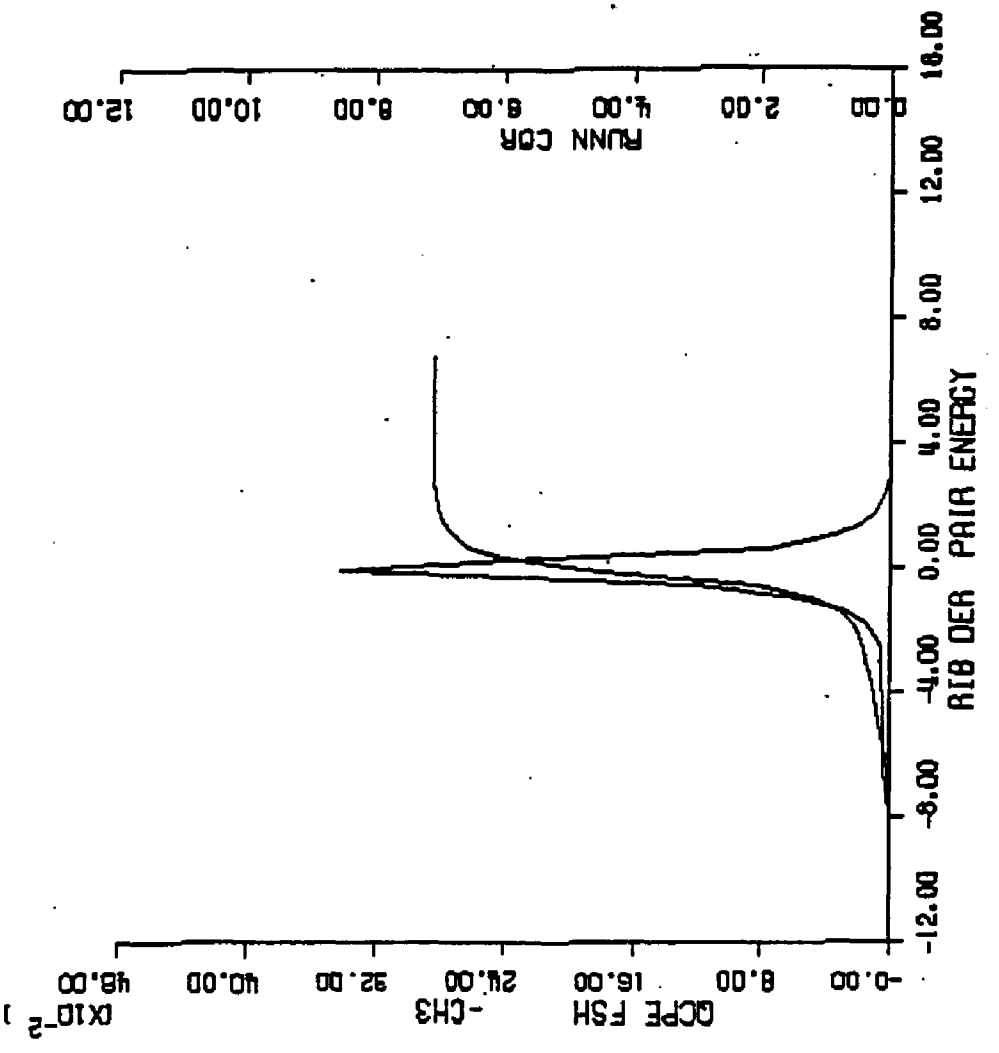
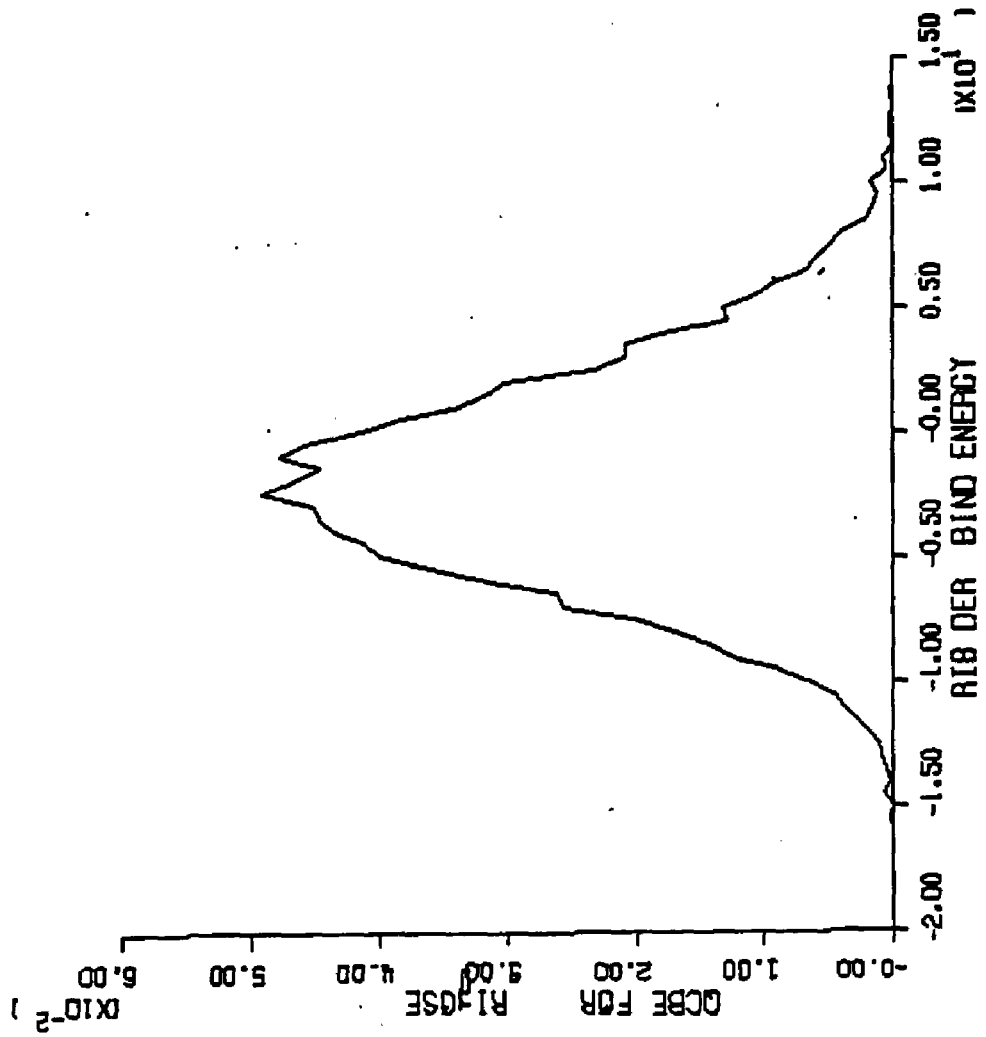


Figure IV.10.24 - calculated quasicomponent distribution function of binding energy for ribose derivative (Clementi potentials).

X axis - binding energy (kcal/mole).

Y axis - quasicomponent of binding energy.

Figure IV.10.24 Calculated QCDF of total binding energy for ribose derivative - Clementi solute-water potential functions.



11. (C2-Endo)-2,5-Dideoxy-1-C-Amino-B-D-Ribo-Pentofuranose with Clementi Solute-Water Potentials

Computational Specifics. An aqueous solution of a deoxyribose derivative and 202 waters was simulated using Clementi solute-water potentials and MCY-CI water-water potentials at 298 K. The sugar ring was puckered in the C2'-endo form. Face centered cubic boundary conditions with a unit cell edge length of 14.63 Angstroms were used. The unit cell edge was obtained from partial molar volumes of 105 cc/mole for the deoxyribose derivative and 18.07 cc/mole for water. The simulation was performed with an initial equilibration period of 500,000 non-force biased and 500,000 force biased moves which were discarded, and a subsequent 2,000,000 moves which were used to perform all analyses and calculate ensemble averages.

Potential Surface. The potential energy contour surface for the interaction of the deoxyribose derivative and one water is shown in Figure IV.11.1. The contour plane chosen is the best fit to the plane of the ring atoms of the deoxyribose derivative. The minimum is found in the region surrounding the ester oxygen where the lowest energy is -7.99 kcal/mole. The region surrounding the molecular plane, excepting the ether oxygen, from the ether oxygen side of the methyl group to the amino group generally shows interactions in the range -1.99 kcal/mole to -3.99 kcal/mole. The area surrounding the CH2 2 methylene group of the molecular ring also contains interactions between -1.99 and -3.99 kcal/mole.

The region peripheral to the 5- methyl group shows the least favorable interactions overall; here, energies between .01 and -1.99 kcal/mole are encountered.

Convergence and Thermodynamic Results. The control functions for the deoxyribose derivative simulation are found in Figure IV.. The mean internal energy stabilizes at about -1755 kcal/mole for the value of -1756.1 +/- 6.4 kcal/mole after 2000K moves. The heat capacity attains the final value 18.8 cal/mole-K. The calculated vacuum to water transfer energy for the deoxyribose derivative is -8.8 kcal/mole.

Figures IV.11.4 thru IV.11.11 and IV.11.14 thru IV.11.21 contains graphs of the quasicomponent distribution functions of coordination number and binding energy for the amino, hydroxyl, methylene, methine, and methyl functional groups and ether oxygen atom of the deoxyribose derivative. Figures IV.11.12 and IV.11.22 contain graphs of the quasicomponent distribution functions of first shell coordination number and binding energy for the deoxyribose derivative. Table IV.12 contains the complete structural and energetic data for the deoxyribose derivative and its atoms and functional groups.

Structural Results. First shell coordination numbers for the atoms of the deoxyribose derivative are displayed on the molecular diagram of Figure IV.11.3. The amino

hydrogen atoms, H1' and H1'', again have significantly different radial cutoff values (3.2 and 2.2 Angstroms, respectively) and correspondingly different first shell populations (2.11 and .70, respectively).

The O1 ester oxygen, O3' hydroxyl oxygen and H3' hydroxyl hydrogen each have about one coordinated water molecule (1.02, .85, and 1.02, respectively), as seen with the ribose derivative. The first shell solvent density of the H4' hydrogen is especially high (1.34) signifying a highly localized water molecule.

The methine and methylene hydrogens have relatively extended first shell cutoffs (H1: 3.6 Angstroms; H2': 4.4 A; H3: 3.6 A; H4: 4.4 A) and first shell solvent densities less than or about equal to that of bulk water (H2: 1.07; H3': .95; H4: .85; H5: .89). These results are typical of apolar hydration. The methylene hydrogen H2 has zero coordination number; due to molecular geometry, no waters are assigned to this hydrogen by proximity. The methyl group cutoffs are similarly extended (H5': 3.8A; H5'': 3.8 A; H5''': 4.2 A) with low first shell solvent densities (H5' : .95; H5'': .95; H5''': .87).

The graph of the quasicomponent distribution function of first shell coordination number for the deoxyribose derivative is shown in Figure IV.11.12. Values range from 18 to 28; thelargest contributions arise from

coordination numbers of 22 and 23. The average first shell coordination number for C2-amino-C5-methyl-deoxyribose is 22.21 +/- .67.

Energetic Results. The average first shell solute-water pair energies for the waters assigned to the atoms of the deoxyribose derivative are displayed on the molecular diagram of Figure IV.11.13. The most favorable average solute-water pair energy is found with the waters of the H3' hydroxyl hydrogen atom at -7.271 kcal/mole; the solute binding energy of -7.419 kcal/mole is the single greatest contribution to the total solute binding energy in the molecular hydration shell (about 1/3 of a total 23.2 kcal /mole). The next most favorable interaction arises from the waters of the ester oxygen (-2.972 kcal/mole) followed by the waters of the amino hydrogens (H1'': -2.915 kcal/mole; H1': -2.593 kcal/mole).

The energetics of the waters of the apolar methylene and methine hydrogens are all fairly similar. The average pair energies range from about -1 kcal/mole (H1: -.924 kcal/mole) to slightly positive (H3: .279 kcal/mole). The waters assigned to the methyl hydrogens are also not especially favorably bound (H5'<SLIPE>= .099 kcal/mole; H5'' :-.476 kcal/mole; H5''' : -.013 kcal/mole).

The graph of the quasicomponent distribution function of solut binding energy is shown in Figure IV.11.22. Values ranging from -37 to -8 kcal/mole are seen; the major contributions arise in the range -29 to -18 kcal/mole. The calculated average solute binding energy for the deoxyribose derivative is -23.2 +/- 1.8 kcal/mole.

Figure IV.11.1 - isoenergy contour surface for deoxyribose derivative and one water (Clementi solute-water potential functions).

X axis - X axis of plane defined by molecular ring atoms (Angstroms).

Y axis - Y axis of plane defined by molecular ring atoms (Angstroms).

Contours - isoenergy contours with one kcal/mole increments with alphabetical labels referring to contour energy values in list at right.

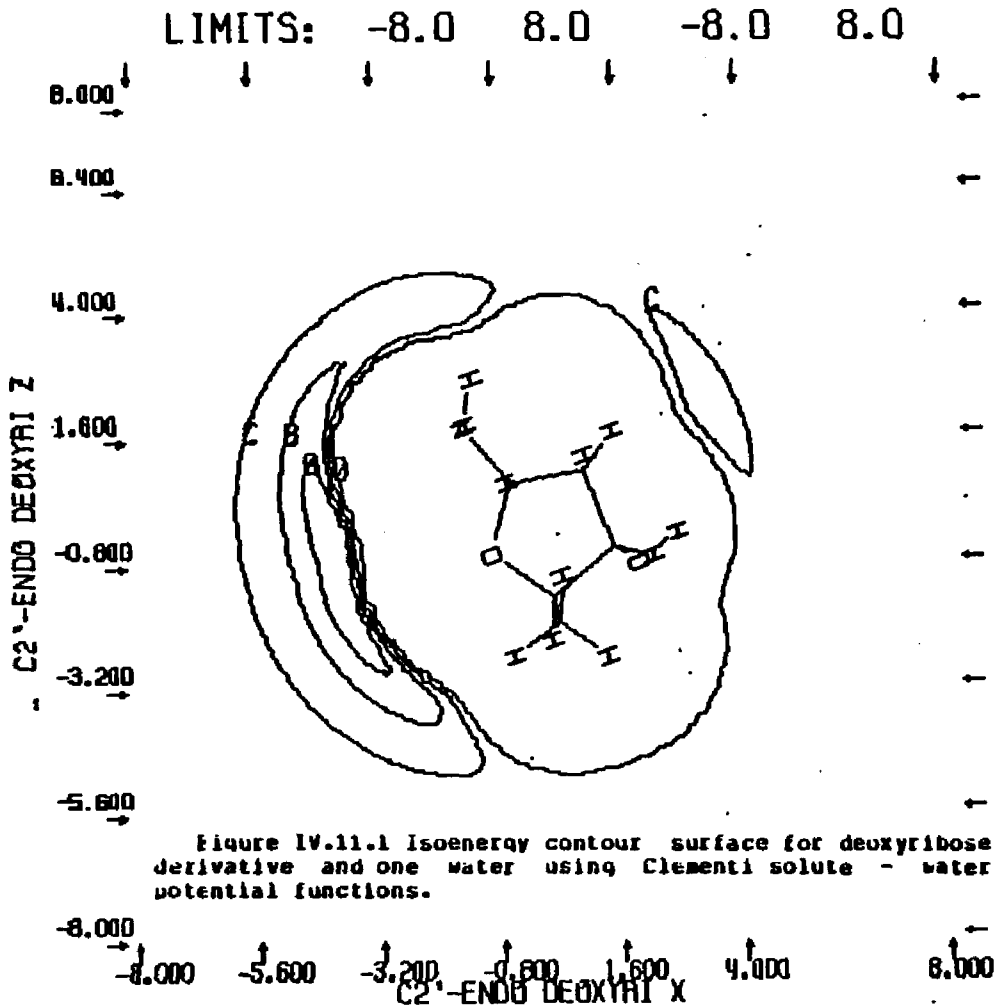


Figure IV.11.1 Isoenergy contour surface for deoxyribose derivative and one water using Clementi solute - water potential functions.

Figure IV.11.2 - control functions for Monte Carlo simulation of deoxyribose derivative and 202 waters (Clementi solute-water potential functions).

X axis - number of configurations.

Left Y axis - mean total energy (kcal/mole).

Right Y axis. - constant volume heat capacity (cal/mole-degree).

Upper curve - constant volume heat capacity.

Bottom curve without crosshatches - average total energy for entire simulation.

Bottom curve with crosshatches - average total energy for preceeding 50K configurations.

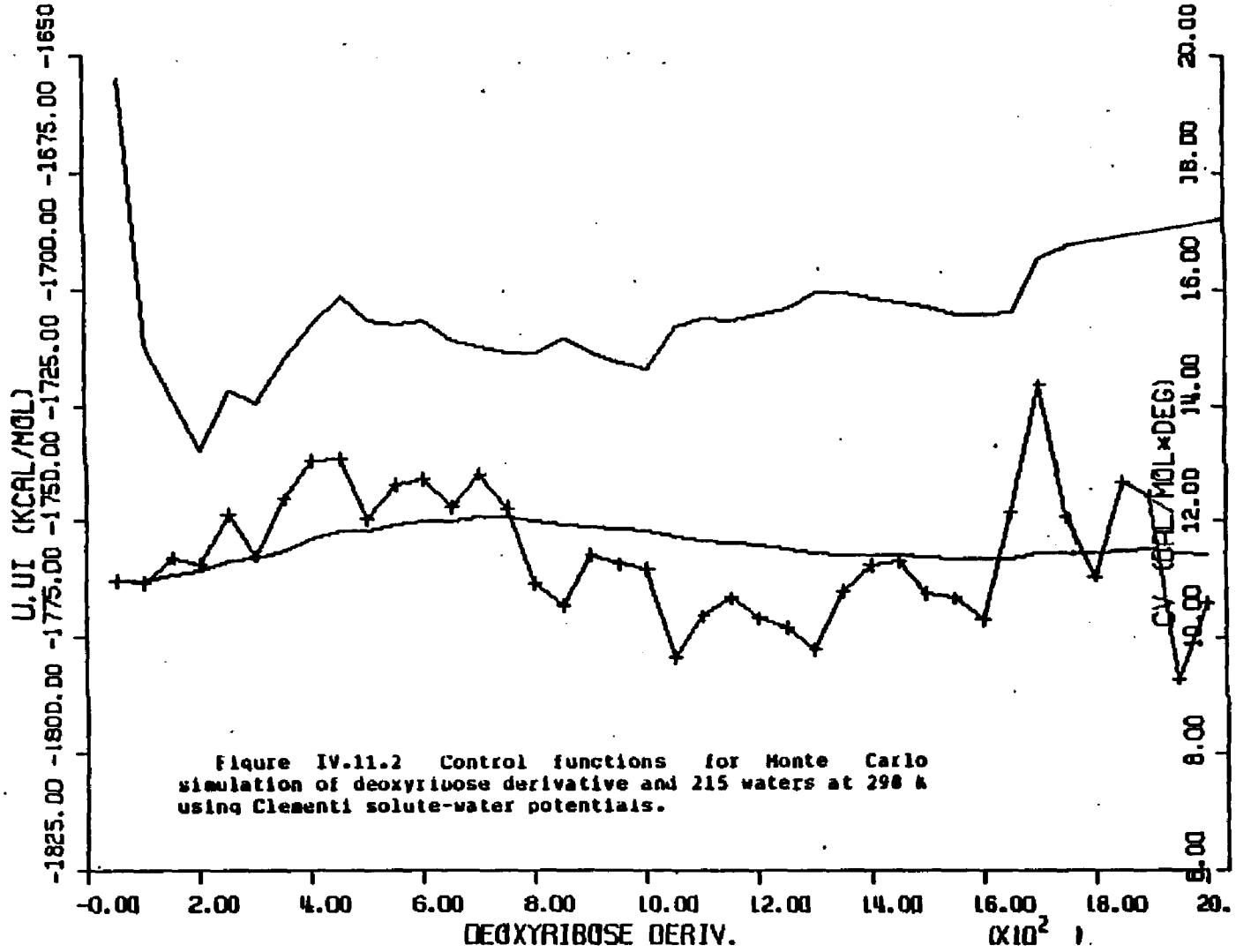


Figure IV.11.2 Control functions for Monte Carlo simulation of deoxyribose derivative and 215 waters at 298 K using Clementi solute-water potentials.

DEOXYRIBOSE DERIVATIVE IN WATER AT 298K - FORCE BIAS AND CLEMENTI POTS. POTENTIAL FUNCTIONS

LAST CONFIGURATION: 2000002					FIRST SHELL SOLUTE PROPERTIES				TOTAL SLT PROPS		WATER PROPERTIES RFSW=3.30 RCE= 7.75 A		
INDEX	TYPE			RFS	VFS	<K>	<K/V>	<SLTBE>	<SLTPE>	<K>	<SLTBE>	<KW>	<SWWPE>	<BEWT>
METHYLENE (MONOSUB.)														
1	2	8	C	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	6	59	H	3.6	55.07	1.96	1.07	-1.815	-0.924	15.76	-2.228	4.13	-3.074	-17.072
TOTALS FOR FUNCTIONAL GROUP >CH- :					55.07	1.96	1.07	-1.815	-0.924	15.76	-2.228	4.13	-3.074	-17.072
STATISTICAL UNCERTAINTY (+/- 2*SD):						0.20	0.11	0.304	0.205	0.02	0.624	0.07	0.048	0.216
7	5	17	C	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8	12	59	H	4.4	84.20	2.50	0.89	0.238	0.095	16.71	0.145	4.35	-3.014	-17.674
TOTALS FOR FUNCTIONAL GROUP >CH- :					84.20	2.50	0.89	0.238	0.095	16.71	0.145	4.35	-3.014	-17.674
STATISTICAL UNCERTAINTY (+/- 2*SD):						0.22	0.08	0.035	0.019	0.02	0.039	0.07	0.045	0.218
17	4	17	C	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
14	18	59	H	3.6	44.49	1.26	0.85	0.352	0.279	8.65	0.935	4.16	-2.944	-16.949
TOTALS FOR FUNCTIONAL GROUP >CH- :					44.49	1.26	0.85	0.352	0.279	8.65	0.935	4.16	-2.944	-16.949
STATISTICAL UNCERTAINTY (+/- 2*SD):						0.08	0.08	0.037	0.050	0.01	0.003	0.07	0.046	0.213
METHYLENE GROUP														
9	3	17	C	3.4	11.10	0.07	0.19	0.072	1.016	0.15	0.065	4.18	-2.863	-15.865
18	8	59	H	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
19	9	59	H	4.4	92.55	2.92	0.95	-0.811	-0.277	16.34	-0.027	4.20	-3.032	-17.079
TOTALS FOR FUNCTIONAL GROUP -CH2 :					103.65	2.99	0.89	-0.739	-0.247	16.49	-00.038	4.20	-3.014	-16.949
STATISTICAL UNCERTAINTY (+/- 2*SD):						0.04	0.10	0.063	1.191	0.00	0.107	0.72	0.458	2.000

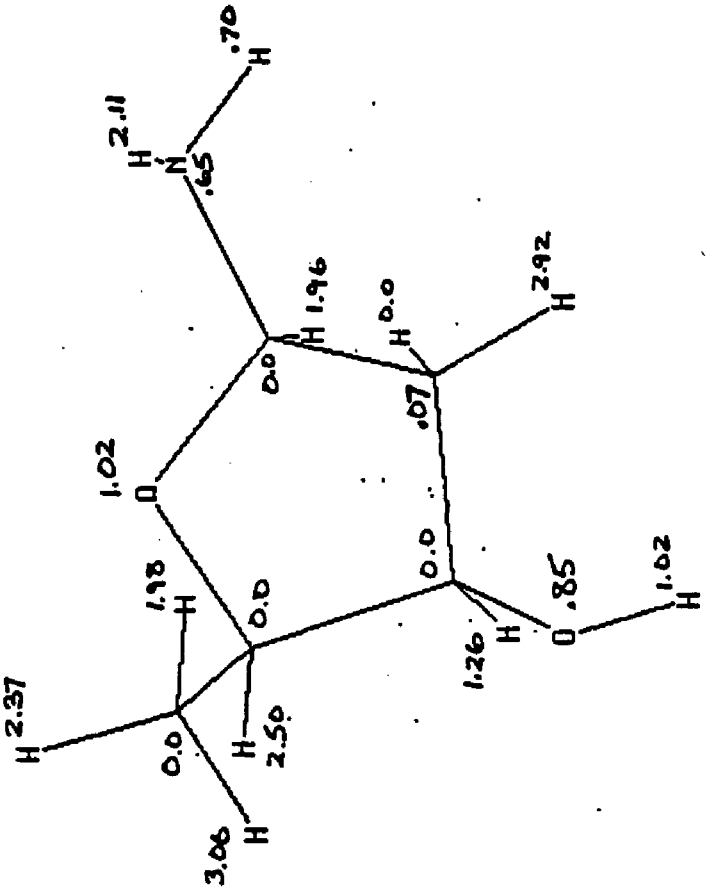
Table IV.12 Calculated structural and energetic quantities from Monte Carlo simulation of C1'amino-C4'-methyl-deoxyribose derivative and 202 waters at 298 K - Clementi solute-water potential functions.

DEOXYRIBOSE DERIVATIVE IN WATER AT 298K - FORCE BIAS AND CLEMENTI POTENTIAL FUNCTIONS

LAST CONFIGURATION: 2000002					FIRST SHELL SOLUTE PROPERTIES					TOTAL SLT PROPS		WATER PROPERTIES RFSW=3.38 RCB= 7.75 A		
METHYL GROUP														
3	13	6	C	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	17	3	H	3.8	74.63	2.37	0.95	0.234	0.099	18.93	0.656	4.15	-3.092	-17.158
5	18	3	H	3.8	62.36	1.98	0.95	-0.942	-0.476	14.18	-0.828	4.17	-3.027	-17.083
6	19	3	H	4.2	105.36	3.06	0.87	-0.848	-0.013	22.94	0.583	4.18	-3.044	-16.961
TOTALS FOR FUNCTIONAL GROUP -CH3 :					242.35	7.41	0.91	-0.747	-0.101	55.97	0.411	4.17	-3.056	-17.056
STATISTICAL UNCERTAINTY (+/- 2*SD):						0.39	0.05	0.064	0.012	0.04	0.061	0.04	0.025	0.115
AMINE GROUP														
10	7	11	N	3.4	17.72	0.38	0.65	-1.906	-4.968	4.42	-2.156	4.05	-3.088	-16.442
11	14	1	H	3.2	54.19	2.11	1.16	-5.464	-2.593	20.98	-6.196	4.19	-3.021	-16.979
12	15	1	H	2.2	23.89	0.70	0.88	-2.040	-2.915	23.08	-4.065	4.15	-3.090	-17.013
TOTALS FOR FUNCTIONAL GROUP -NH2 :					95.81	3.19	1.00	-9.410	-2.949	48.48	-12.418	4.16	-3.060	-16.946
STATISTICAL UNCERTAINTY (+/- 2*SD):						0.25	0.08	1.237	0.514	0.03	1.983	0.04	0.027	0.122
ESTER OXYGEN														
13	1	68	O	4.0	38.61	1.02	0.79	-3.035	-2.972	5.52	-2.827	4.10	-3.000	-16.682
STATISTICAL UNCERTAINTY (+/- 2*SD):						0.14	0.11	0.705	0.916	0.01	1.337	0.12	0.078	0.357
HYDROXYL GROUP														
15	11	9	O	3.4	38.59	0.85	0.83	-1.434	-1.693	7.37	-1.181	4.05	-3.030	-16.834
16	16	57	H	2.2	22.82	1.02	1.34	-7.419	-7.271	27.05	-6.051	4.15	-3.025	-16.743
AVERAGES OVER FUNCTIONAL GRP -OH :					53.41	1.87	1.07	-8.853	-4.434	34.42	-7.232	4.09	-3.028	-16.795
STATISTICAL UNCERTAINTY (+/- 2*SD):						0.08	0.08	0.376	0.589	0.01	0.356	0.04	0.028	0.129
MOLECULAR SUM/AVERAGE:					717.6	22.21	0.93	-24.010	-1.881	202.00	-23.176	4.17	-3.041	-17.016
STATISTICAL UNCERTAINTY (+/- 2*SD):						0.67	0.03	1.196	0.071	0.07	1.013	0.02	0.013	0.060

Table IV.12 Calculated structural and energetic quantities from Monte Carlo simulation of C1'amino-C4'-methyl-deoxyribose derivative and 202 waters at 298 K - Clementi solute-water potential functions.

Figure IV.11.3 First shell coordination numbers for the atoms of deoxyribose derivative -Clementi solute-water potential functions.



Figures IV.11.4 through IV.11.12 - calculated quasicomponent distribution function of first shell coordination number for deoxyribose derivative and functional groups (Clementi potentials).

X axis - coordination number.

Y axis - quasicomponent of coordination number.

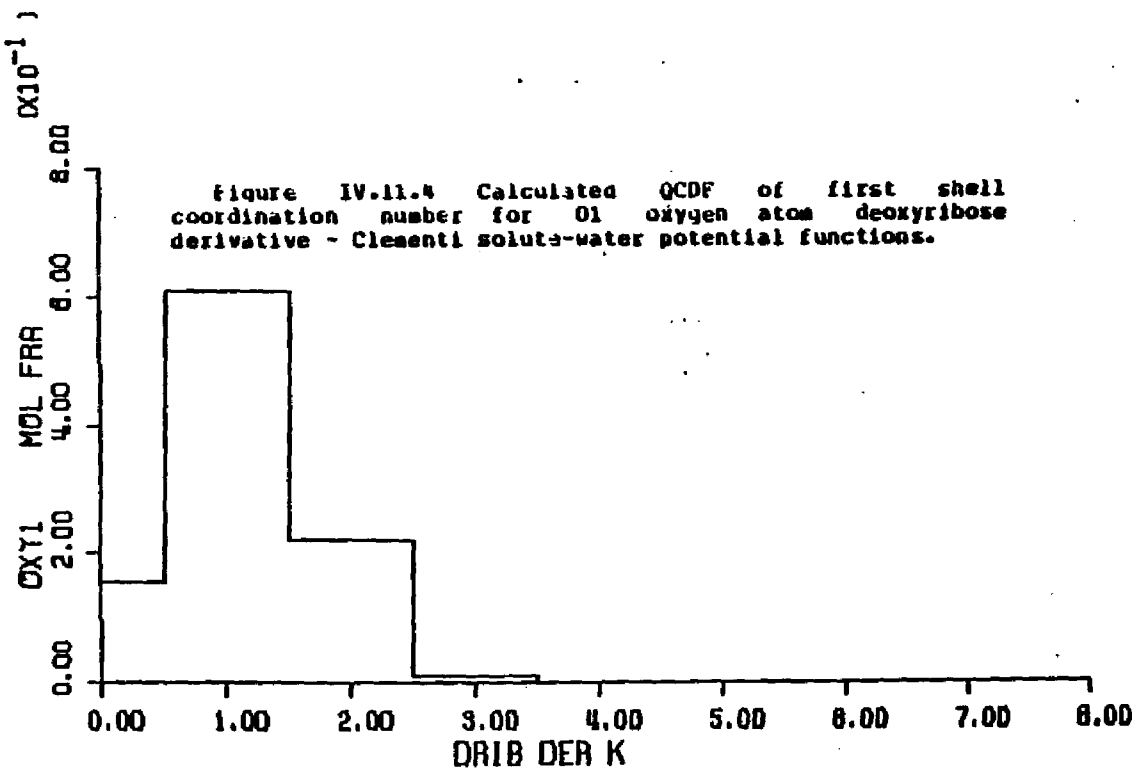


Figure IV.11.4 Calculated QCDF of first shell coordination number for O1 oxygen atom deoxyribose derivative - Clementi solute-water potential functions.

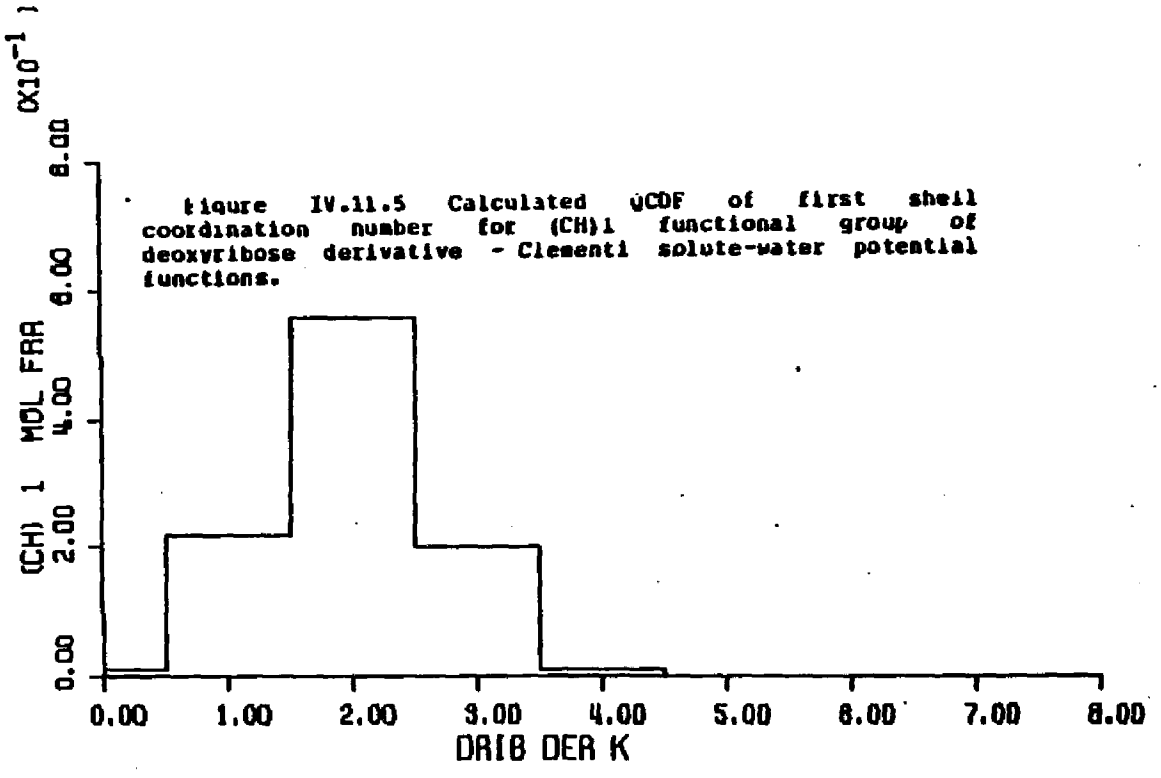


Figure IV.11.5 Calculated QCDF of first shell coordination number for (CH)1 functional group of deoxyribose derivative - Clementi solute-water potential functions.

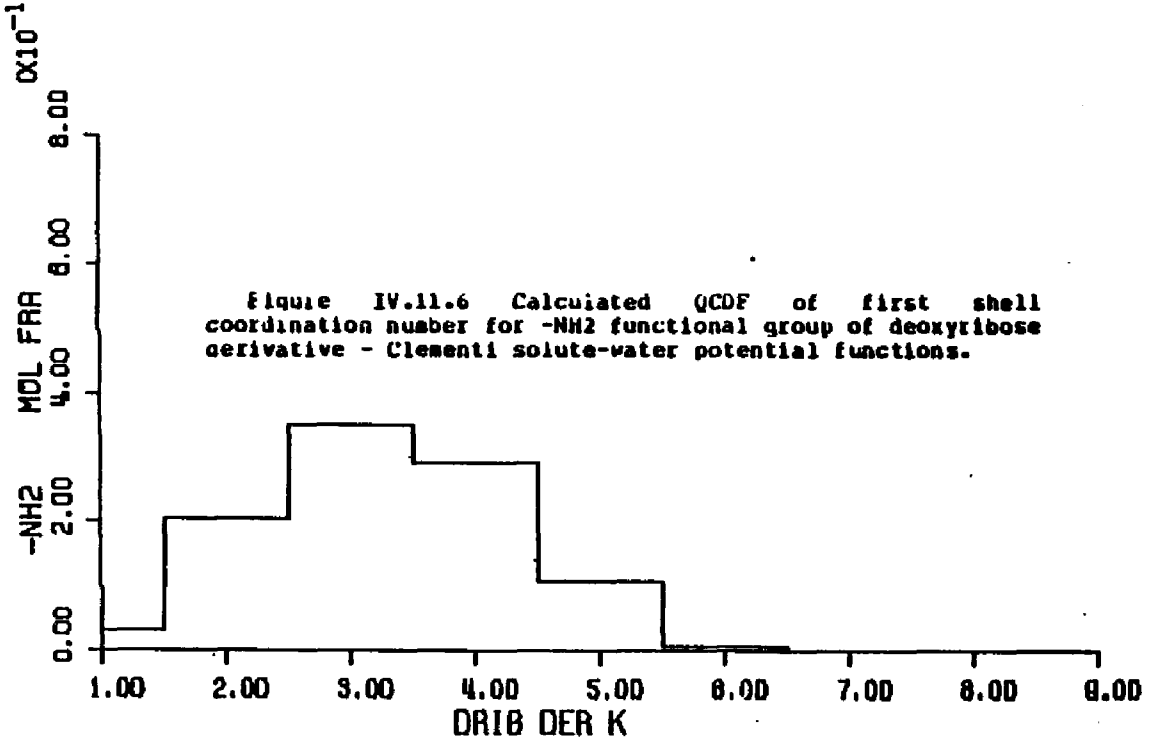


Figure IV.11.6 Calculated QCDF of first shell coordination number for -NH2 functional group of deoxyribose derivative - Clementi solute-water potential functions.

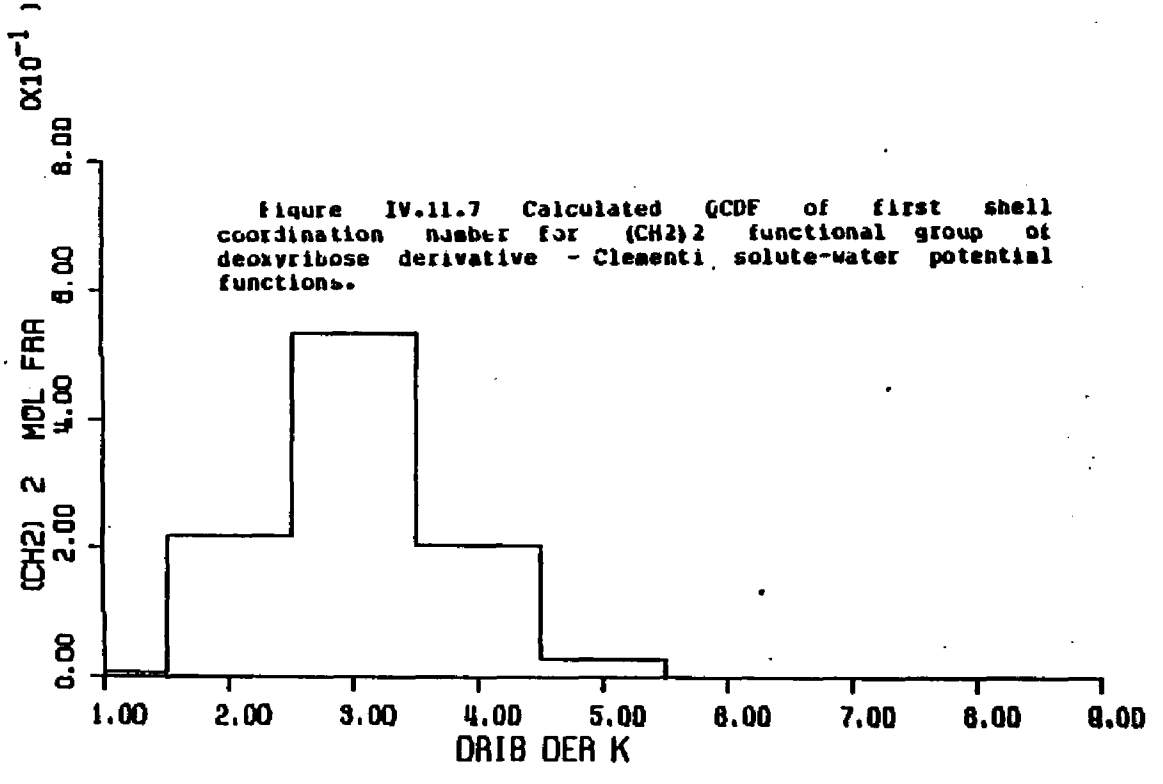


Figure IV.11.7 Calculated QCDF of first shell coordination number for (CH2)2 functional group of deoxyribose derivative - Clementi solute-water potential functions.

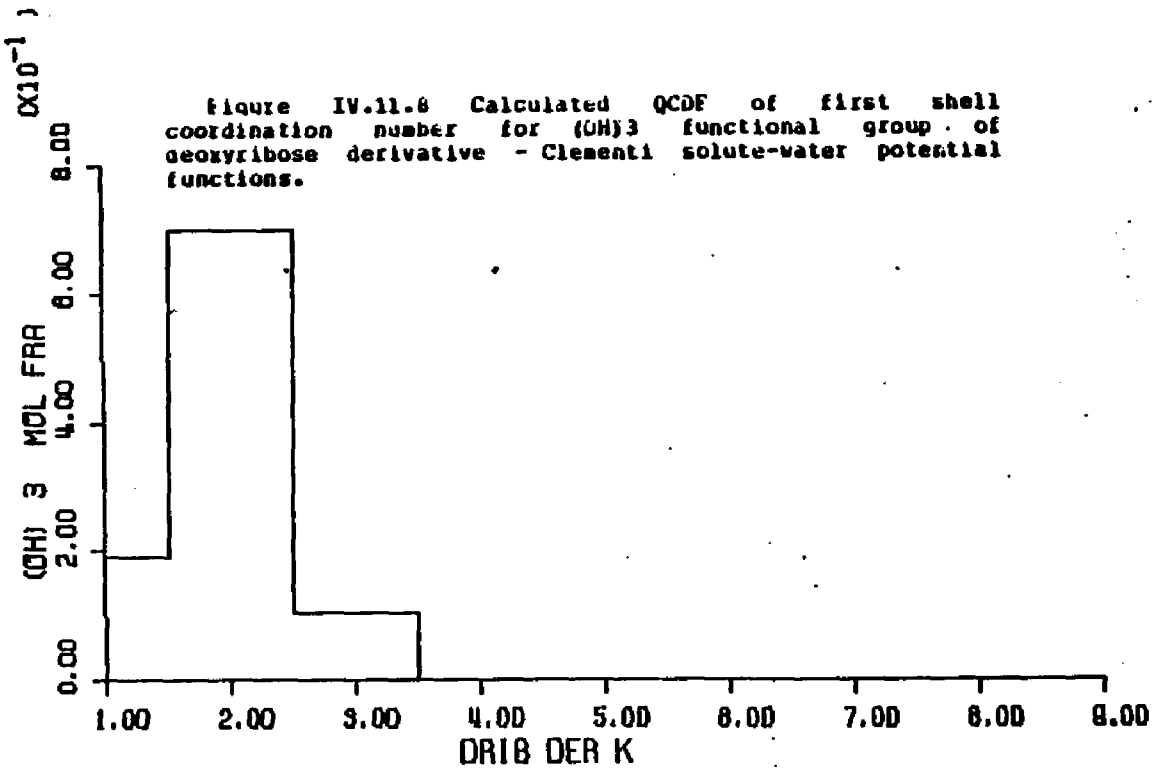


Figure IV.11.6 Calculated QCDF of first shell coordination number for (OH)3 functional group of deoxyribose derivative - Clementi solute-water potential functions.

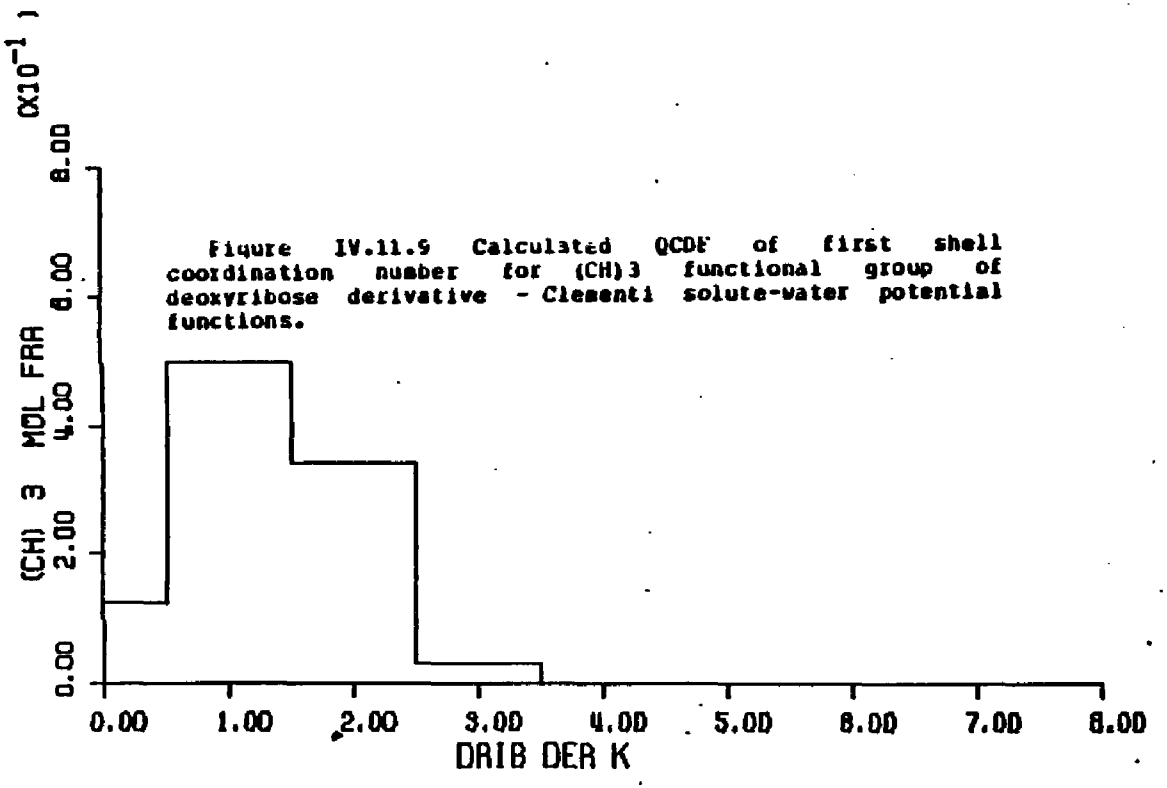


Figure IV.11.9 Calculated QCDF of first shell coordination number for (CH)3 functional group of deoxyribose derivative - Clementi solute-water potential functions.

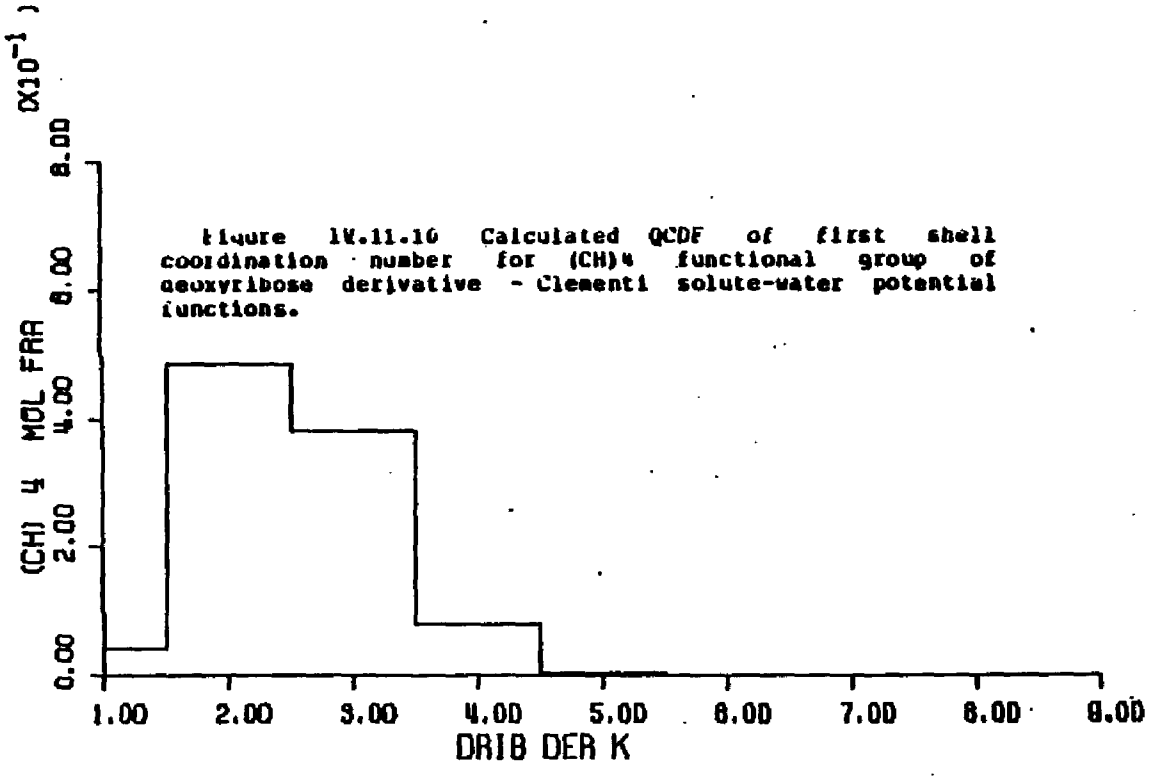


Figure IV.11.10 Calculated QCDF of first shell coordination number for $(CH)_4$ functional group of deoxyribose derivative - Clementi solute-water potential functions.

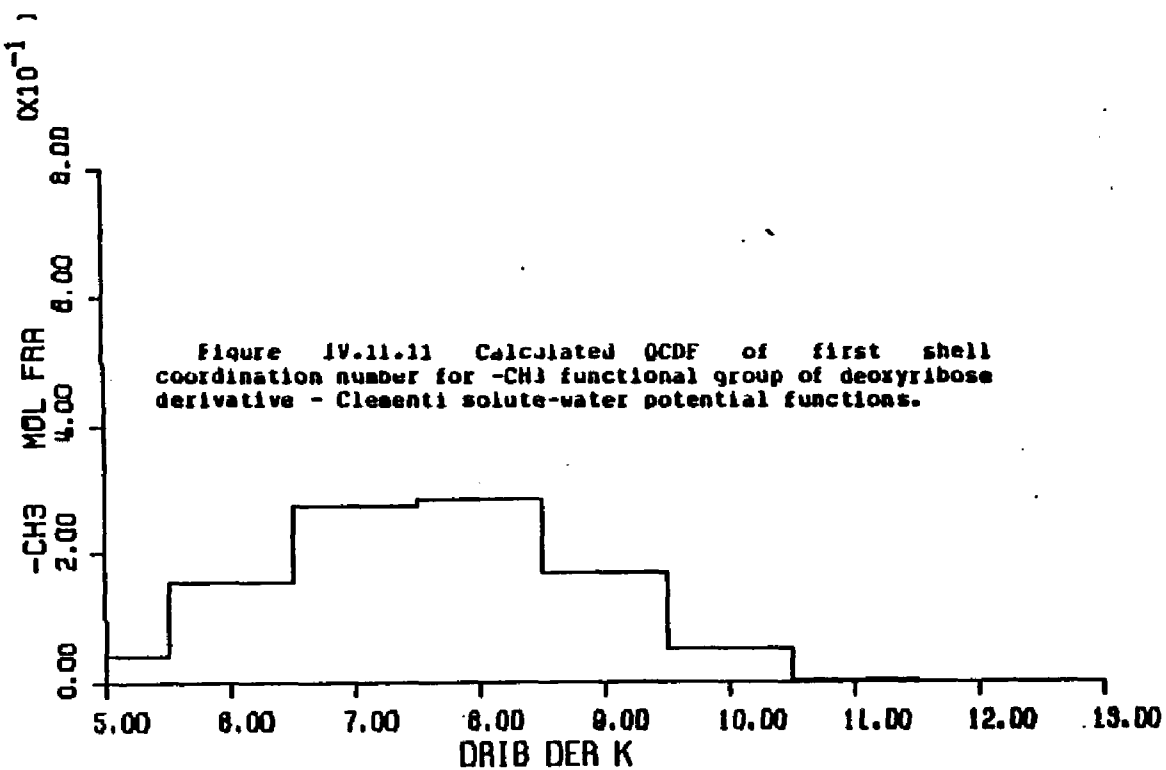


Figure IV.11.11 Calculated QCDF of first shell coordination number for -CH3 functional group of deoxyribose derivative - Clementi solute-water potential functions.

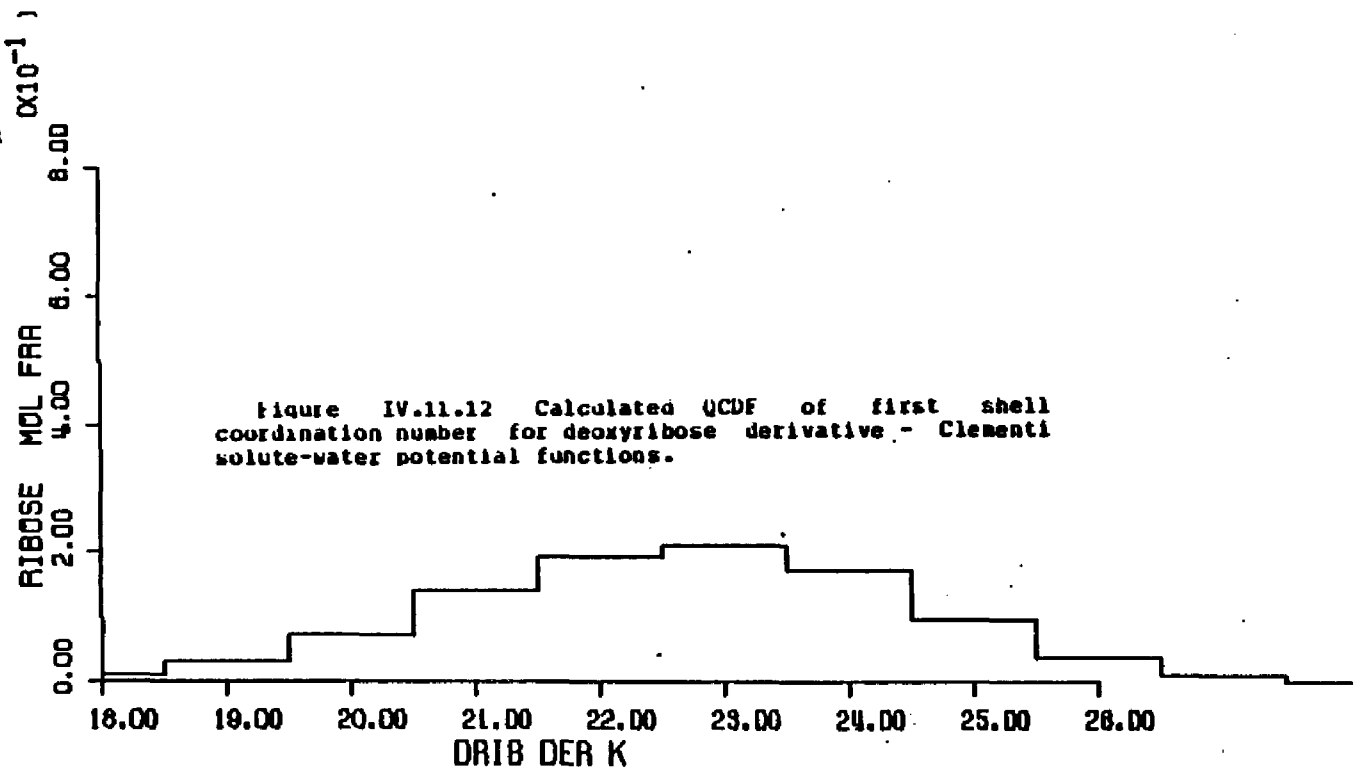


Figure IV.11.12 Calculated QCDF of first shell coordination number for deoxyribose derivative - Clementi solute-water potential functions.

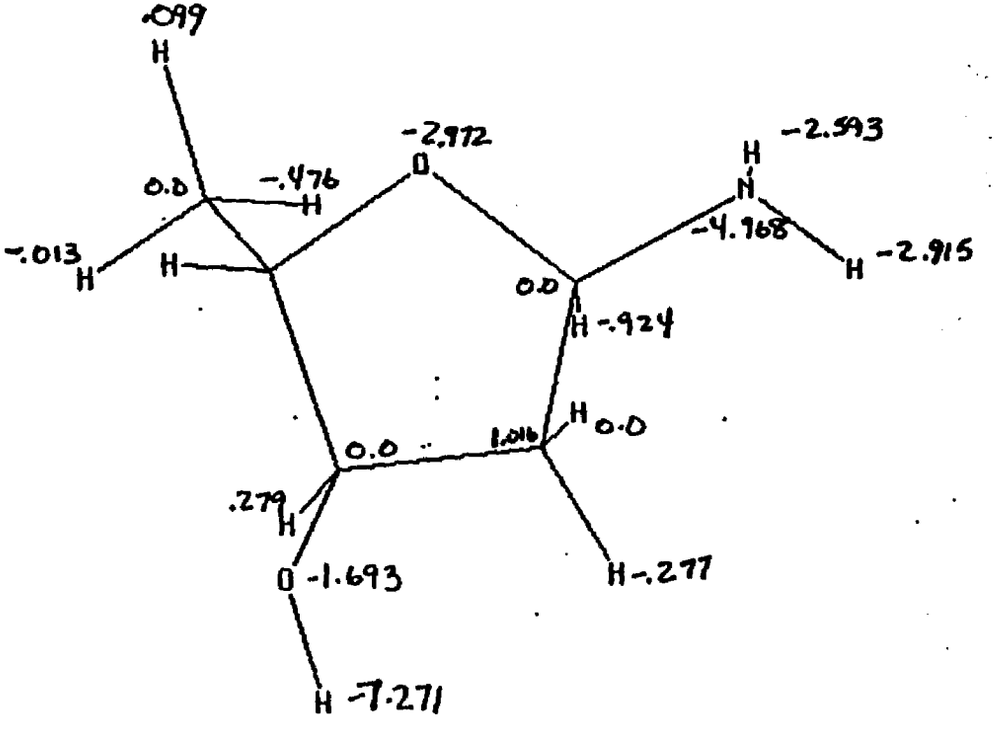


Figure IV.11.13 Average first shell solute-water pair energies of waters assigned to the atoms of deoxyribose derivative - Clementi solute-water potential functions.

Figure IV.11.14 through IV.11.21 - calculated quasicomponent distribution function of average first shell pair energy for functional groups of deoxyribose derivative (Clementi potentials).

X axis - pair energy (kcal/mole).

Left Y axis - quasicomponent of pair energy.

Right Y axis - running coordination number.

Figure IV.II.14 Calculated QCOF of solute-water pair energies of -ates of the O1 oxygen atom group of deoxyribose derivative - Clementi solute-water potential functions.

Figure IV.11.15 Calculated QCUF of solute-water pair energies of waters of the (CH)1 functional group of deoxyribose derivative - Clementi solute-water potential functions.

Figure IV.11-16 Calculated DCDF of solute-water pair potential energy derivative - Classical solute-water functions.

Figure IV.11.17 Calculated QCPE of solute-water pair energies of water of the (CH2)2 functional group of deoxyribose derivative - Clementi solute-water potential functions.

Figure IV.11.18 Calculated QCDF of solute-water pair energies of waters of the (OH)3 functional group of deoxyribose derivative - Clementi solute-water potential functions.

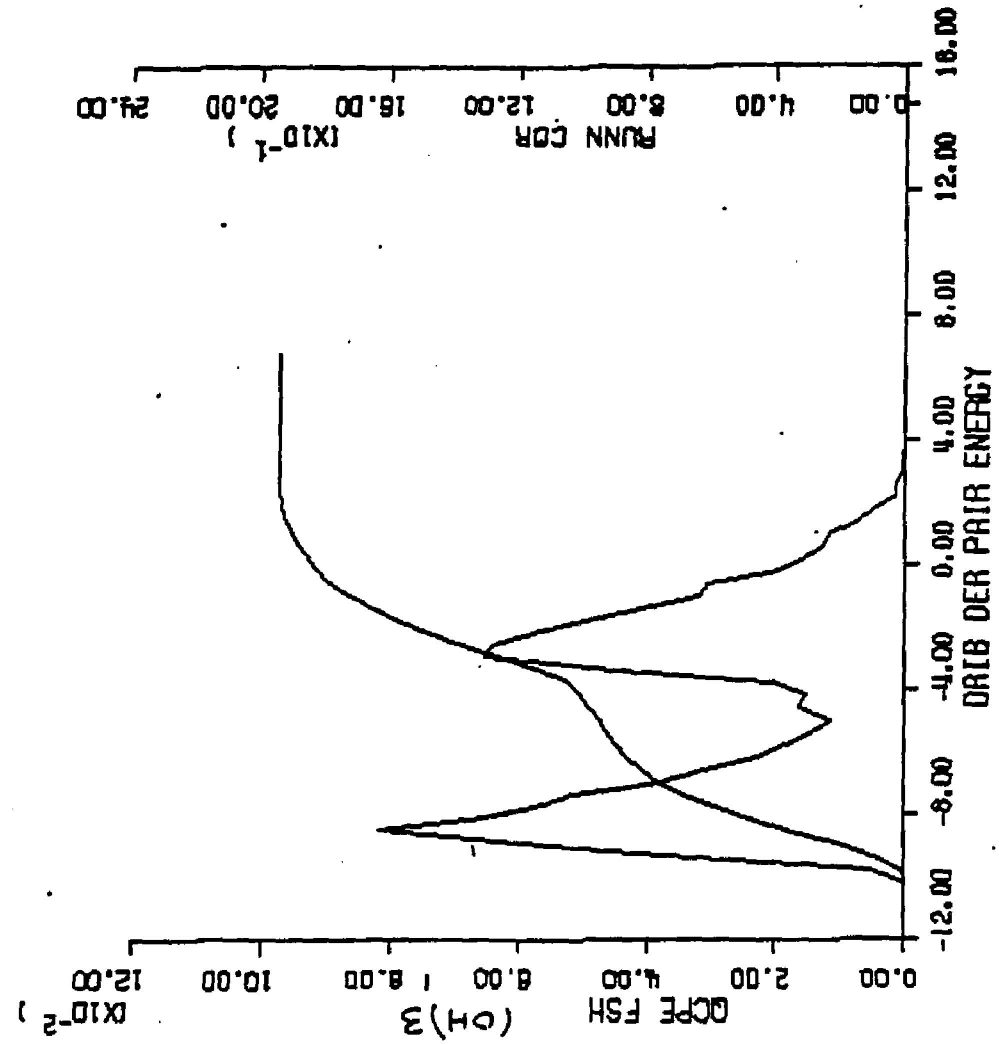


Figure IV.11.19 Calculated QCDF of solute-water pair energies of waters of the (CH)3 functional group of deoxyribose derivative - Clementi solute-water potential functions.

Figure IV.11.20 Calculated QCOF of solute-water pair potential for the (CH)$_4$ functional group of methane derivatives - Classical solute-water pair functions.

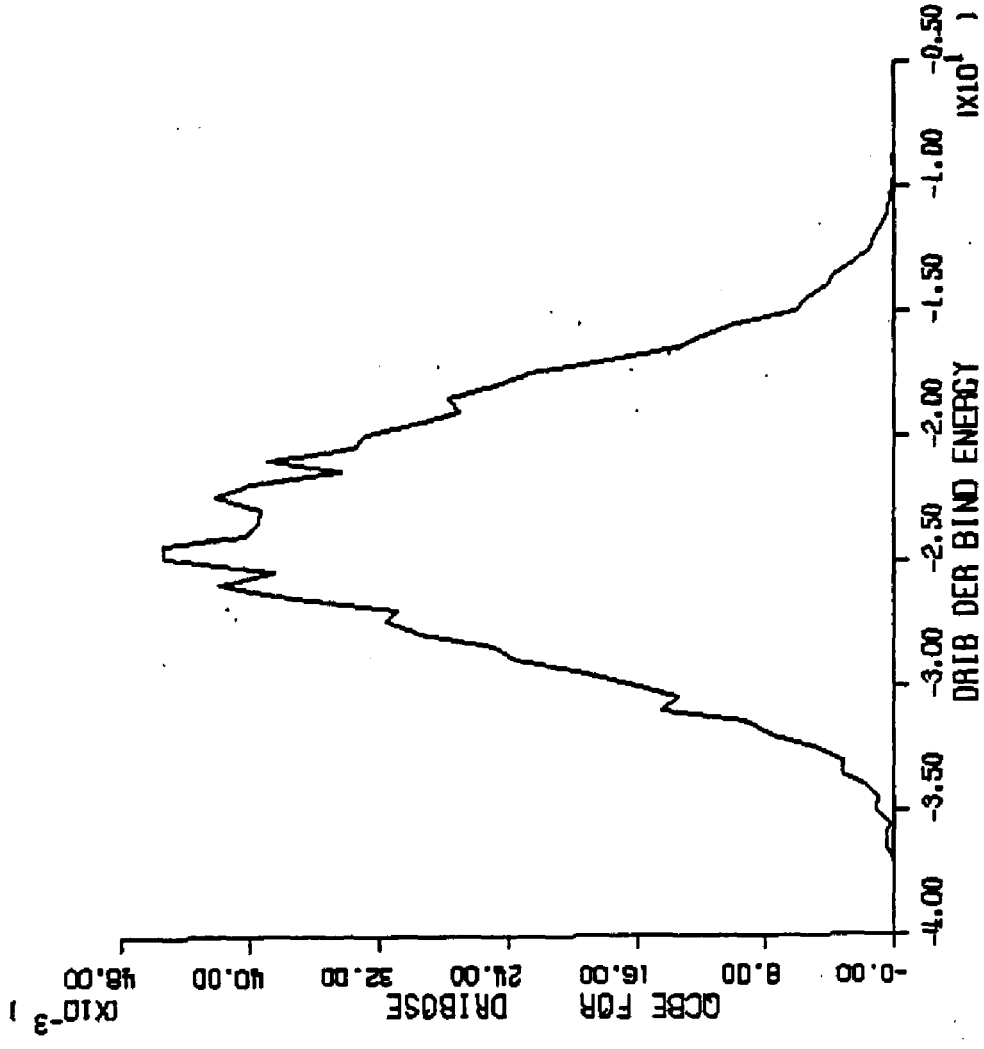
Figure IV-11.21 Calculated QCPE of solute-water pair potential as a function of the -CH3 group of deoxyribose derivative - solute-water clusters.

Figure IV.11.22 - calculated quasicomponent distribution function of binding energy for deoxyribose derivative (Clementi potentials).

X axis - binding energy (kcal/mole).

Y axis - quasicomponent of binding energy.

Figure IV.11.22 Calculated QCDF of total binding energy for deoxyribose derivative - Clementi solute-water potential functions.



CHAPTER V

SYNTHESIS AND DISCUSSION OF RESULTS

Aqueous Hydration of Thymine

The essential structural feature of the aqueous hydration of thymine emerging from the Clementi solute-water potential function simulation is a first hydration shell consisting of 25 waters. Twenty three of these waters constitute the exocyclic ring substituent hydration. The other two waters constitute the above and below plane hydration of the ring atoms. A representative stereo picture of the first hydration shell of thymine is shown in Figure V.1

The hydration of the exocyclic ring substituents of thymine has varied character. The only atoms of thymine with the structural and energetic features of hydrophilic hydration are the imino hydrogens. These atoms have large first shell populations with average solute-water pair energies of -3.552 kcal/mole. When this quantity is compared to the average water-water pair energy of -3.010 kcal/mole for MCY-CI bulk water, it is apparent that the hydration of these atoms is solvent structure breaking (relative to bulk solvent). The carbonyl, methine and methyl functional groups all are assigned first shell waters with first shell solute-water pair energies more positive than the average water-water pair energy reflecting a prevalence of water-water interactions in these regions.

The vacuum to water transfer energy of thymine was calculated to be -21.4 +/- 7.5 kcal/mole. The typical solute-water interaction in the first hydration shell of thymine (-1.589 kcal/mole) is considerably less than the average bulk water-water pair energy. The solute-water interactions of the ring substituents predominate not only in sheer number (23 vs. 2) but also in average solute-water pair energy and binding energy (only -.793 of the solute binding energy of -39.8 kcal/mole arises from ring atom hydration).

An interesting result is the unfavorable solute-water interactions occurring in the extended (beyond first shell) hydration layers of thymine. The binding energy of these waters (+32.459 kcal/mole) is nearly as unfavorable as the the first shell interactions are favorable (-39.792 kcal/mole). About half of this positive interaction energy arises in the extended methyl group shell suggesting that the overall unfavorable solute-water interactions in the extended molecular shell region is the result of a long range hydrophobic effect due to the methyl group.

A comparison of results for the hydration of thymine obtained from the Clementi, Kollman and BVG solute-water potential function simulations follows. The Clementi thymine first hydration shell contains 25.04 waters, the Kollman thymine first shell contains 22.94 waters and the

BVG thymine first shell contains 22.84 waters. The explicit hydrogen atoms of the Clementi thymine molecule have much larger first shell volumes and consequently a significantly greater first shell population than the united atom Kollman and BVG methyl groups. When an adjustment in the first shell coordination number of the Clementi methyl group is made proportional to the ratio of the volume of the BVG or Kollman first shells (these are the same) to the Clementi methyl first shell volume, the coordination number of Clementi thymine is 21.78. Thus no appreciable difference is found in first shell hydration structure with the variety of solute-water potential functions.

The average first shell solute-water pair energies for Kollman thymine (-.738 kcal/mole) and BVG thymine (-.958 kcal/mole) are more positive than those of Clementi thymine (-1.589 kcal/mole). Also, there are no atomic first shell solute-water pair energies greater than the average water-water pair energies in either the Kollman thymine or BVG thymine simulations. Thus, thymine hydration is generally more typical of apolar hydration when a simulation is performed with either the Kollman or BVG solute-water potential functions.

Values for the solid to water transfer energy (heat of solution) for thymine are calculated for each simulation by adding the vacuum to water transfer energy (<USLT> in

Table IV.1) to the experimentally determined sublimation energy of thymine (+29.7 kcal/mole) (50). The values thus obtained are: Clementi thymine: +8.3 kcal/mole, Kollman thymine: +12.9 kcal/mole and BVG thymine: +8.6 kcal/mole. When these values are compared to experimentally determined values for the heat of solution of thymine ranging from +5.3 to +5.8 kcal/mole (50-52), it is seen that the Clementi solute-water potential function simulation produces a value very close to experiment followed closely by the value obtained from the BVG thymine simulation.

SCALE FACTOR = .48

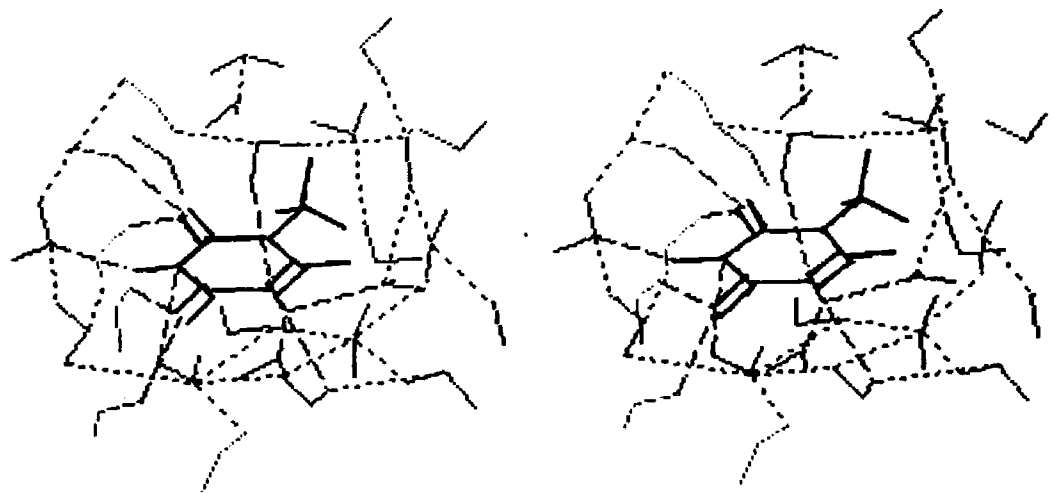


Figure V.1 Stereo view of a representative first hydration shell configuration of thymine - Clementi solute-water potential functions.

Aqueous Hydration of Uracil

The basic structural feature of the aqueous hydration of uracil obtained from the Clementi solute-water potential function simulation is a first hydration shell consisting of 19 to 20 water molecules. Seventeen of these waters constitute the first shell hydration of the exocyclic ring substituents. The other two or three of these waters hydrate the ring atoms of uracil. A stereo picture of a representative first shell hydration structure of uracil from this simulation is shown in Figure V.2.

The imino hydrogens both show definite characteristics of hydrophilic hydration with fully populated, high solvent density (average 1.37 x bulk water) first shells with strong solute-water interactions (average -5.309 kcal/mole) The O2 carbonyl oxygen and H5 methine hydrogen first shell waters both have solute-water pair energies greater than bulk water (-3.010 kcal/mole). These pair energies are enhanced by favorable interactions with the nearby imino hydrogens.

The vacuum to water transfer energy of uracil was found to be -149.7 +/- 10.2 kcal/mole. The average first shell solute-water pair energy (-3.666 kcal/mole) is somewhat greater than bulk MCY-CI water. The waters of the exocyclic ring substituents provide a considerable

majority (-66.202 kcal/mole of -72.453 kcal/mole) of the first shell solute-water binding energy. Hydration of the ring atoms provides only -6.251 kcal/mole to this quantity.

Comparison of the structural hydration results for the Clementi, Kollman and BVG solute-water potential function uracil-water simulations again shows no trend in the first shell molecular coordination numbers. Clementi uracil has a first shell consisting of 19.76 waters, Kollman uracil has a first shell consisting of 18.76 waters and BVG uracil has a first shell consisting of 21.00 waters. The Clementi uracil first shell is denser (1.06) than the Kollman (.97) or BVG (.88) uracil first shells.

The denser Clementi first shell reflects the tighter clustering of water molecules in the relatively strongly bound first hydration shell of the Clementi uracil molecule. The order of first shell solute-water pair energies is: Clementi uracil (-3.666 kcal/mole) > BVG uracil (-.940 kcal/mole) > Kollman uracil (-.760 kcal/mole). The Clementi solute-water interactions are about four times more favorable than the Kollman or BVG energies; none of the first shell waters show hydrophilic hydration energetics in the Kollman and BVG simulations.

Solid to water transfer energies (heats of solution) are calculated for uracil using the value <USLT> obtained

from Table IV.1 for each simulation and and experimentally determined value for the heat of sublimation of uracil (+28.8 kcal/mole) (50). The following values are obtained: Clementi uracil: -120.9 kcal/mole, Kollman uracil: +17.6 kcal/mole and BVG uracil: +6.1 kcal/mole. Experimentally determined heats of solution for uracil range from +6.4 to +7.1 kcal/mole (50-52,55); in this case, the BVG uracil-water simulation produces a transfer energy comparing very well with experimental data. The Clementi solute-water potential function simulation considerably overestimates the heat of solution.

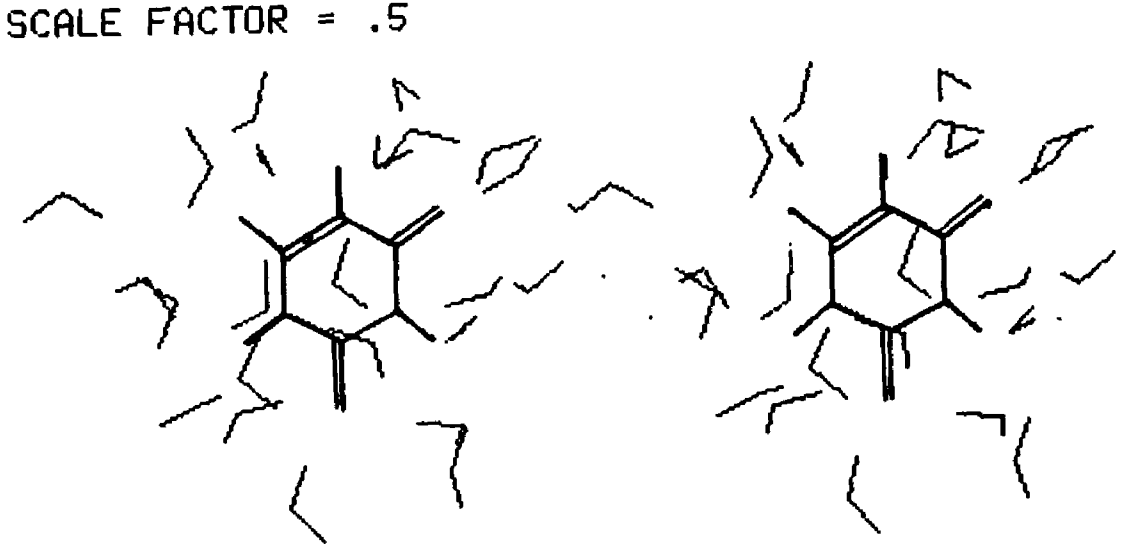


Figure V.2 Stereo view of a representative first hydration shell configuration of uracil - Clementi solute-water potential functions.

Aqueous Hydration of Cytosine

The first hydration shell of cytosine was found to consist of nineteen water molecules. Sixteen of these water molecules constitute the first shell hydration of the exocyclic ring substituents of cytosine. If the out of plane hydration of the bare ring nitrogen is assigned .59 waters (this number is the average first shell coordination number of the base imino nitrogens from the Clementi solute-water potential function base-water calculations), then the ring atoms are assigned about three water molecules. A stereo view of a representative first shell hydration structure of cytosine from this simulation is shown in Figure V.3.

The waters of the imino and amino hydrogen atoms show definite characteristics of hydrophilic interaction with the solute molecule with average solute-water pair energies of -6.215 kcal/mole (imino hydrogen) and -5.854 kcal/mole (amino hydrogens). All other first shell waters have average solute-water pair energies less favorable than the average bulk water water-water interaction. Typical apolar solute-water interactions are seen with the waters of all ring carbon atoms and the methine H5 hydrogen atom. The H6 methine hydrogen atom waters interact with the solute somewhat more favorably than the H5 waters because of the strong potential of the

neighboring H1 imino hydrogen.

The average first shell solute-water pair energy is -3.220 kcal/mole which is comparable to the average bulk water water-water interaction of -3.010 kcal/mole. The total solute-water binding energy of the exocyclic ring substituent atoms (-54.957 kcal/mol) dominate the molecular first shell binding energy of -60.619 kcal/mole. The remaining tenth (-5.666 kcal/mole) of the first shell solute-water binding energy is contributed by the ring atom waters.

The vacuum to water transfer energy for cytosine was calculated to be -127.4 +/- 8.9 kcal/mole. The solid to water transfer energy for cytosine can be calculated for this simulation if the sublimation energy of cytosine is estimated. A value of +30 kcal/mole for this quantity is reasonable since known experimental values for uracil, thymine, adenine and various methylated bases range from +25 to +35 kcal/mole. Using this estimated value, the solid to water transfer energy of cytosine is calculated to be -97.4 kcal/mole. Comparison with the experimental value of +7.7 kcal/mole (50) demonstrates that the Clementi solute-water potentials overestimate the solute-water interactions.

SCALE FACTOR = .5

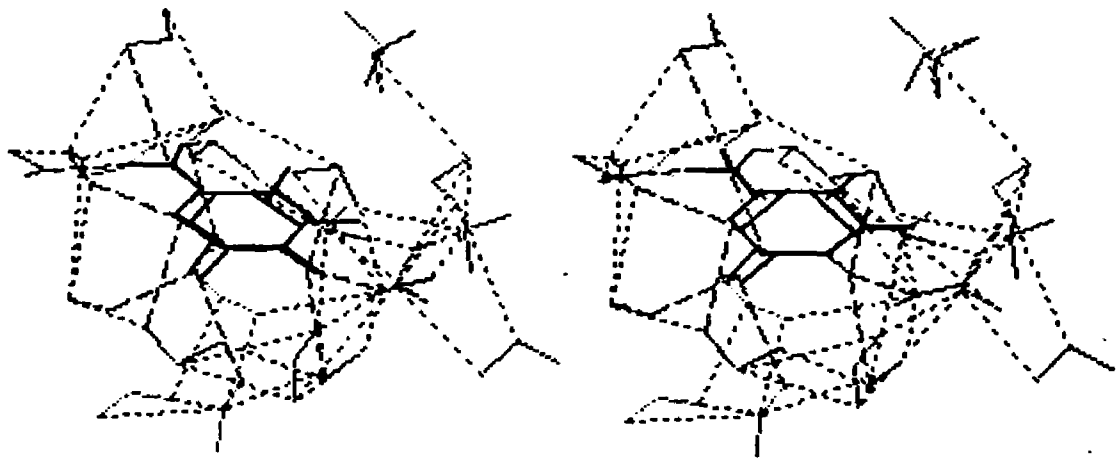


Figure V.3 Stereo view of a representative first hydration shell configuration of cytosine- Clementi solute-water potential functions.

Aqueous Hydration of Adenine

The aqueous hydration of adenine is found to have the basic structural feature of a first hydration shell consisting of twenty or twenty one water molecules. Twelve or thirteen of these water molecules are found in the first hydration shell of the exocyclic ring substituents. The ring atoms are assigned four or five waters using the same approximated value for the out of plane bare ring nitrogen described for cytosine. A stereo view of a representative first shell hydration structure of adenine from this simulation is shown in Figure V.4.

The waters of the imino and amino hydrogen atoms of adenine again show decidedly hydrophilic character in their interactions with the solute. The average solute-water interactions for the imino hydrogen waters are -4.792 kcal/mole and -5.357 kcal/mole. All other first shell waters have average solute-water pair energies less than the bulk water water-water interaction with the exception of the N9 imino nitrogen waters.

The average first shell solute-water pair energy (-2.596 kcal/mole) is less favorable than the average bulk water water-water interaction of -3.010 kcal/mole due to the weak solute-water interactions of all ring atoms (including the bare ring nitrogens) and the two

methine ring fragments. The exocyclic ring substituent atoms contribute a substantial majority (-43.838 kcal/mole) of the first shell solute-water binding energy of -52.953 kcal/mole. The remaining sixth of the first shell binding energy is obtained via the waters of the ring atoms from out of plane interactions (-9.115 kcal/mole). The vacuum to water transfer energy of adenine is found to be -141.4 +/- 13.5 kcal/mole. The experimentally determined sublimation energy of adenine is +30.4 kcal/mole (53). When this value is added to the vacuum to water transfer energy obtained from this simulation, a value of -111.0 kcal/mole is found for the solid to water transfer energy. Experimentally obtained value for this quantity range from +7.1 to +8.0 kcal/mole (52-56). The Clementi solute-water potential functions again overestimate solute-water interactions.

SCALE FACTOR = .46

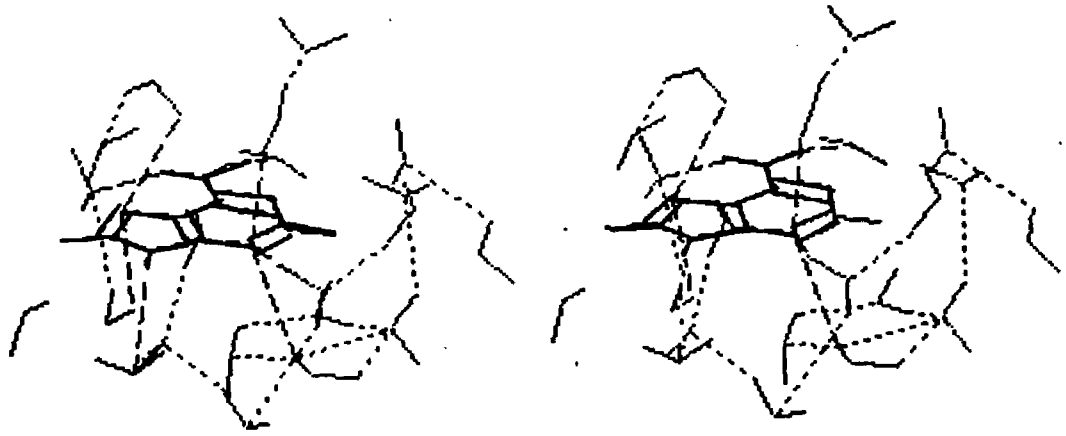


Figure V.4 Stereo view of a representative first hydration shell configuration of adenine - Clementi solute-water potential functions.

Aqueous Hydration of Guanine

The basic structural feature of the aqueous hydration of guanine is a first hydration shell consisting of twenty one or twenty two water molecules. Seventeen of these waters constitute the first hydration shell of the exocyclic ring substituent atoms. The other four or five waters are assigned to the ring atoms using the same approximation to the out of plane hydration of the bare ring nitrogen atoms described for cytosine. A stereo view of a representative first hydration shell structure of guanine from this simulation is shown in Figure V.5.

The waters of the amino hydrogen atoms show definite characteristics of hydrophilic interaction with the solute molecule with solute-water pair energies of -4.331 kcal/mole (imino hydrogens) and -7.731 kcal/mole (amino hydrogens). All other first shell waters have average solute-water interactions less than the average bulk water water-water interaction energy except for the imino nitrogen N9 and carbonyl oxygen O6 waters.

The average first shell solute-water pair energy is -3.314 kcal/mole which is slightly more favorable than the average bulk water interaction. Most (-61.912 kcal/mole) of the first shell solute-water binding energy of -71.233 kcal/mole arises in the first hydration shells of the exocyclic ring substituent atoms. About one-

seventh (-9.321 kcal/mole) of this energy comes from the out of plane interactions of the waters of the ring atoms.

The vacuum to water transfer energy for guanine is -112.6 +/- 7.9 kcal/mole. If the sublimation energy of guanine is estimated to be +30 kcal/mole as in the case of cytosine and added to the vacuum to water transfer energy obtained in the simulation, a solid to water transfer energy of -82.6 kcal/mole is obtained. The experimentally obtained value for this quantity is +11.8 kcal/mole (54); the Clementi solute-water potential functions again overestimate solute-water interaction energies.

SCALE FACTOR = .44

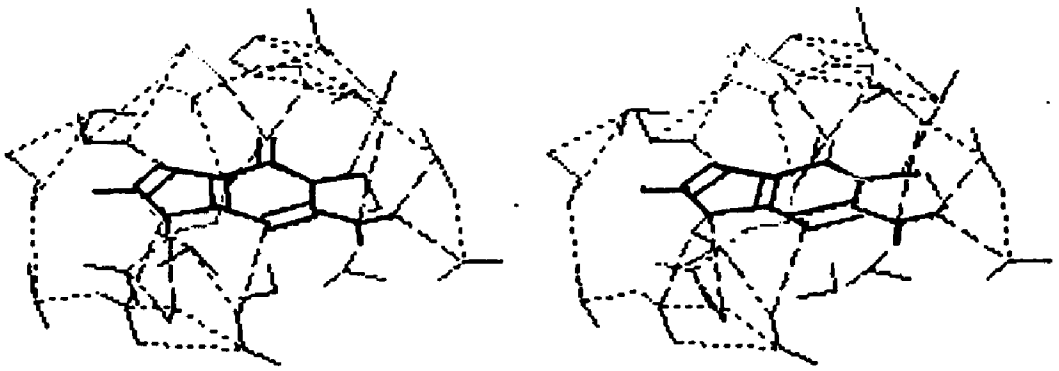


Figure V.5 Stereo view of a representative first hydration shell configuration of guanine - Clementi solute-water potential functions.

Aqueous Hydration of (C3-Endo)-5-Deoxy-1-C-Amino-B-D-Ribo-Pentofuranose

The basic structural feature of the hydration of the ribose derivative is a first shell consisting of twenty one or twenty two waters molecules. The majority (twelve or thirteen) of the first shell waters are assigned to the apolar constituents of the molecule (methine and methyl groups). The remainder of the first shell (nine waters) hydrate the polar constituents (ether oxygen, hydroxyl and amino groups) of the ribose derivative. A stereo view of a representative first shell configuration for the ribose derivative obtained from this simulation is shown in Figure V.6.

The hydrophilic hydration of the molecule is constituted by the waters of the hydroxyl hydrogen atoms (average -5.562 kcal/mole) and the H2' amino hydrogen atom (-4.331 kcal/mole). When a water molecule is found within the first shell of the ether oxygen (first shell coordination number only .76) the interaction is also quite favorable (-3.591 kcal/mole). Clearly hydrophobic interactions occur with the waters of the ring methylene hydrogens with extended cutoffs (average 4.0 Angstroms), low first shell densities (.80 x bulk water) and very weak average solute-water pair interactions (-.118 kcal/mole). The waters of the methyl group show similar

data characteristic of hydrophobic solute-water interactions.

The vacuum to water transfer energy of the ribose derivative is calculated to be -2.8 +/- 9.2 kcal/mole. The average solute-water pair energy in the first hydration shell is -1.496 kcal/mole; the first shell of this molecule thus appears to be dominated by water-water interactions. The most important solute-water associations occur with the four waters hydrating the two hydroxyl groups. These contribute the majority (-17.216 kcal/mole) of the total first shell binding energy (-31.960 kcal/mole) of this molecule. Nearly the remainder (-9.529 kcal/mole) is provided by the waters of the amino group.

SCALE FACTOR = .46

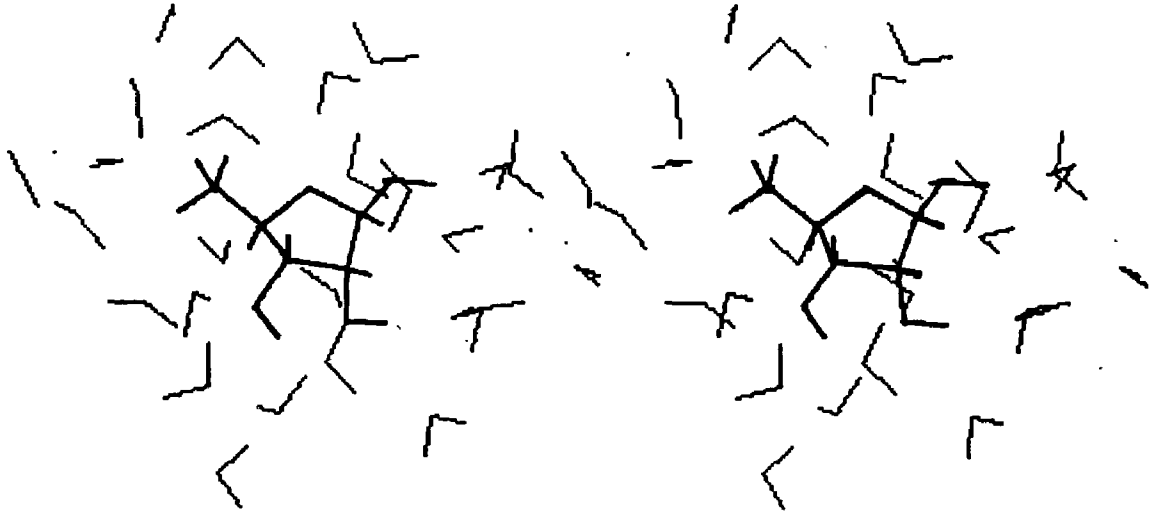


Figure V.6 Stereo view of a representative first hydration shell configuration of ribose derivative - Clementi solute-water potential functions.

Aqueous Hydration of (C2-Endo)-2,5-Dideoxy-1-C-Amino-B-D-Ribo-Pentofuranose

The basic structural feature obtained from the deoxyribose derivative simulation is a first hydration shell containing twenty two waters. The bulk of the waters (sixteen) are assigned to the apolar methylene and methyl constituents of the deoxyribose derivative. The remainder (seven) of the water molecules hydrate the polar ether oxygen, hydroxyl and amino molecular constituents. A stereo view of a representative configuration of the first hydration shell obtained from this simulation is found in Figure V.7.

A definite hydrophilic interaction is found with the water of the hydroxyl hydrogen with an average solute interaction of -7.271 kcal/mole. Other significant, moderate solute-water interactions occur with the waters of the ester oxygen and amino hydrogens. Hydrophobic solute-water interactions are found with the waters of the methylene ring fragments with the typical extended first shell cutoffs, low first shell densities and very weak average solute pair interactions of -.236 kcal/mole. the waters assigned to the methyl group also display hydrophobic interactions with average solute pair energies of -.107 kcal/mole.

The vacuum to water transfer energy is found to be -8.8 +/- 8.8 kcal/mole. The average first shell solute-water interactions is -1.081 kcal/mole demonstrating that water-water interactions predominate in the first shell. The preponderance of the first shell solute-water binding energy arises with the three amino waters (-9.410 kcal/mole) and the two waters of the hydroxyl group (-8.853 kcal/mole).

The results for each simulation were accompanied by plots of the simulation control functions, including the running mean total energy. Attainment of a stable mean total energy over the last 500K to 1000K configurations of a simulation is one index of simulation reliability. The stability of this quantity varied from a case where virtually no variation was seen in the total mean energy over the last 1000K of the simulation (thymine, Clementi potentials) to a case where the mean total energy drifted somewhat less than 10 kcal/mole (less than .05 kcal/mole/particle) over the last 1000K configurations of the simulation (uracil, Clementi potentials).

Each simulation was run for a period sufficient for one large scale fluctuation to appear in the 50K block averages of the total energy calculated from the simulation. These periodic large scale fluctuations, also known as grand cycles, have been found to occur about every 2000K configurations in extended Monte Carlo

liquid state simulations. Only very small changes in mean total energy occurs (about .03 kcal/mole/particle) when simulation is performed for more than one grand cycle (59); considering the extremely large amounts of computer time required for a Monte Carlo simulation, it is felt that attainment of one full grand cycle is sufficient to reasonably insure the reliability of a simulation despite a small amount of drift in the mean total energy. For this reason, it is felt that the calculations reported here are quite reliable.

SCALE FACTOR = .47

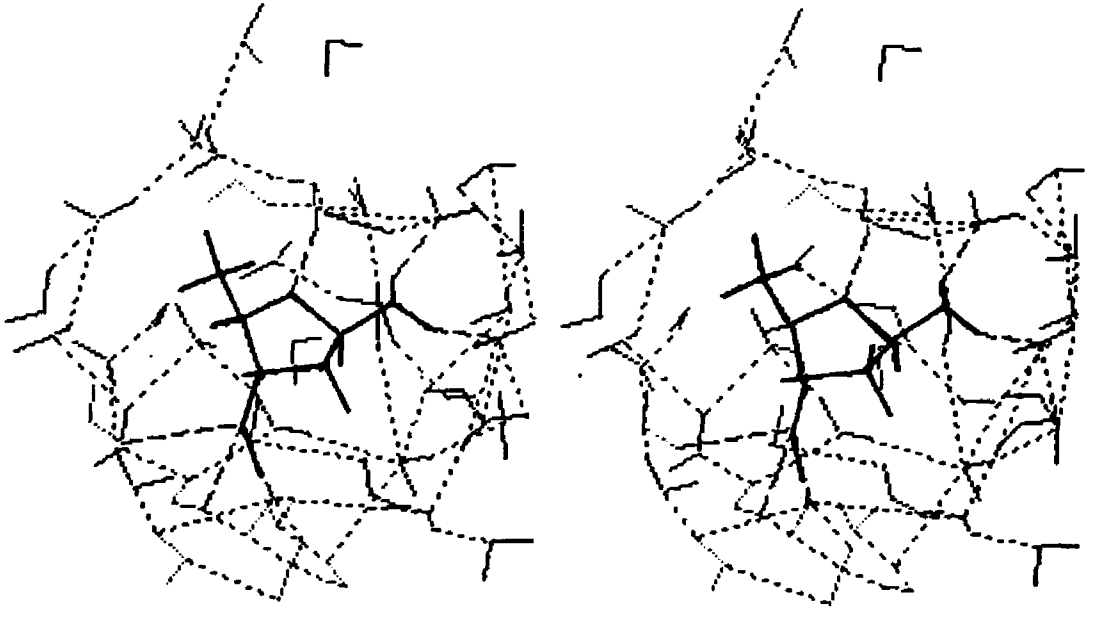


Figure V.7 Stereo view of a representative first hydration shell configuration of deoxyribose derivative - Clementi solute-water potential functions.

CHAPTER VI

SUMMARY AND CONCLUSIONS

Monte Carlo computer simulation studies of the hydration of the five nucleic acid bases and derivatives of ribose and deoxyribose were performed to obtain knowledge of the microscopic hydration of these systems. This chapter is a concise summary of the results and a survey of possible extension of this line of research.

The simulation of uracil and thymine hydration were performed with each of three different solute-water potential functions in order to assess the quality of these potentials and determine the effects of variation of this methodological variable on simulation results. Comparison of simulation transfer energies with experimentally obtained transfer energies permits determination of the quality of simulation energetics. The simulation transfer energies calculated for thymine are comparable to experimental solid to water transfer energies (average +5.5 kcal/mole) with all three solute-water potential functions (Clementi: +8.3 kcal/mole, Kollman: +17.9 kcal/mole, BVG: +8.6 kcal/mole). On the other hand, transfer energetics calculated for uracil differ greatly between the Clementi simulation and the Kollman and BVG simulations. The calculated solid to water transfer energies using Kollman potentials (+17.6 kcal/mole) and BVG potentials (+6.1 kcal/mole) both agree well with experiment (average +6.8 kcal/mole). The corresponding transfer energy calculated using the Clementi potentials (-120.9 kcal/mole) is over 120

kcal/mole more negative than the experimental value. This trend of transfer energy overestimation is continued in the other base simulations using Clementi potentials. The values calculated for cytosine, guanine and adenine are all over 100 kcal/mole more negative than experimental results. The conclusion is that the BVG and Kollman potentials calculate reasonable transfer energies in the cases investigated here. The Clementi set, however, consistently overestimates solute-water interaction energies. The best possible estimate of average first shell pair energies for the bases are on the order of those calculated for uracil and thymine using BVG and Kollman potentials, i.e., about -1.0 kcal/mole.

The large negative errors found for the solute transfer energies of adenine, guanine, cytosine and uracil with the Clementi potentials can probably be attributed to the use of very small (subminimal) basis sets in the ab initio calculations used to determine the base-water interaction energies upon which these potentials are based. The intermolecular interaction energies are calculated by subtracting the sum of the energies of the isolated molecules from the energy of the interacting molecules. The energy of the interacting molecules is lowered relative to the sum of the isolated molecules via the formation of new, lower energy, orbitals. The calculated energies of the isolated

molecules, however, become increasingly positive (and less realistic) with smaller basis sets. Thus, the increased difference between the energies calculated for the interacting molecules and the isolated molecules with small basis sets results in the calculation of unrealistically negative interaction energies. This error is known as basis set superposition error and can be corrected via the counterpoise correction method (56) or avoided using large (extended) basis sets. Thus, the use of larger basis sets or correction methods is indicated in the development of future biomolecular solute-water potentials via ab initio methods.

Although the Clementi potential solute transfer energies are considerably overestimated for all bases except thymine, relative energetics appear reasonable. The observed order of the strength of solute-water interactions for waters assigned to the various base functional groups (imino and amino > carbonyl > methine > methyl) is a trend seen in the five Clementi base-water simulations as well as the six uracil and thymine potential function comparisons; this trend corresponds to the decreasing polarity of the functional groups.

Another question addressed was the effect of potential function choice on solute coordination number. First shell molecular coordination numbers for uracil and thymine appeared essentially independent of energetic

trends. Average coordination numbers for the different functional groups of uracil and thymine with each potential set are calculated in order to assess functional group coordination number transferability. The averages for the imino groups are: Clementi potentials: 3.87, Kollman potentials: 3.10 and BVG potentials: 3.18. The carbonyl groups show exactly the opposite trend. The average coordination numbers are : Clementi potentials: 2.98, Kollman potentials: 3.24 and BVG potentials: 3.92.

Again, no trend is seen in the methine group average coordination numbers: Clementi potentials: 2.51, Kollman potentials: 3.12 and BVG potentials : 2.86. Finally, when the coordination numbers of the BVG and Kollman methyl groups are adjusted to compensate for the reduced first shell volume of the BVG and Kollman united atom methyl groups, the following closely corresponding numbers are obtained: Clementi potentials: 9.43, Kollman potentials: 9.47 and BVG potentials: 9.79. The range of values for these coordination number sets is quite tight; when the averages are rounded to the nearest integer in each case the following numbers are found: imino groups: 3 or 4 waters; carbonyl groups: 3 or 4 waters; methine groups: 3 waters; methyl groups 9 or 10 waters. The coordination numbers are seen to be roughly comparable from potential to potential. There is also no apparent correlation between pair energy strength and coordination

number; no means of determining which coordination numbers are the most accurate are available.

Some assessment of the changes in coordination number upon base dimerization can be made using these results. Using the coordination numbers calculated here, the formation of a base pair between adenine and thymine or uracil would displace about seven first shell waters (interacting at about seven kcal/mole) and formation of a cytosine-guanine base pair would displace about nine first shell waters (interacting at about nine kcal/mole). Construction of a base stacked dimer, on the other hand, would displace only two to four waters per dimer. Base pairing would require displacement of a greater number of first shell waters than base stacking and thus remove more favorable solute-water interactions. Solute-water interactions thus may contribute to the well known preference of bases to form stacked dimers in aqueous solutions, even though the dominant stabilizing force is expected to be hydropobic. Other factors would, of course, be important such as base-base interaction energies and changes in water-water energies and configurational entropies upon dimerization.

The accuracy of solute-water energetics for the pentose systems studied cannot be judged since experimental transfer energies for these molecules are not available. However, the order of solute-water

interaction energies for the waters of the various functional groups reasonably correlates with functional group polarity (hydroxyl > ether > methylene).

Coordination numbers for each of the polar groups moieties (hydroxyl groups and ether oxygens) are all fairly close while the coordination numbers for the apolar methylene groups vary more widely. The ether oxygen is essentially assigned one water molecule in each sugar derivative. Each hydroxyl group is assigned two waters: one to the oxygen atom and one to the hydrogen atom. Considerable variation is seen among the coordination numbers for the methylene groups; this result is largely due to the variation in first shell volumes due to the geometry of the puckered sugar rings.

The generalization of functional group coordination numbers among different molecular systems is an interesting question. Average coordination numbers are calculated for the various functional groups of the Clementi potential base and sugar simulations These are: imino: 3.84, carbonyl: 3.14, methine: 3.27, methyl: 7.89, amino: 2.00, bare ring nitrogen: 1.90, hydroxyl: 1.97, ether oxygen: .89 and methylene: 1.78. An assessment of coordinate number transferability among these simulations can be made by examination of the range of coordination numbers for each functional group when each coordination number is rounded to the nearest

integer. The results are: imino: 3 or 4 waters, carbonyl: 3 or 4 waters, methine: 3 or 4 waters, methyl: 7 to 9 waters, amino: 2 waters, bare ring nitrogens:1 or 2 waters, hydroxyl: 2 waters, ether oxygen: 2 waters and methylene : 1 to 3 waters. The same rough transferability of coordination number noted in the potential function comparison holds for the functional groups of the seven Clementi potential simulations.

The previous Monte Carlo study of the hydration of the five nucleic acid bases by Clementi et al. was described in the Background section of this dissertation. Each of these bases was simulated with a cluster of only 40 water molecules. B.Jayaram, in this laboratory, demonstrated that waters surrounding the apolar atoms of dimethyl phosphate anion rapidly diffuse away from the solute. Thus, the first shells of the methyl group, methine groups and ring atoms of the Clementi et al. simulations can be considered incomplete; no data is reported for apolar hydration and the first hydration shells are lower in population than the corresponding base shell reported here. Since only 40 waters are included in the Clementi et al. base simulations, two complete hydration shells cannot surround the solute. The properties of the first hydration shell are not likely to be accurate unless the water-water influences of a second shell are present. Further, the lack of periodic boundary conditions logically results in edge effects in the energies

calculated for the scant number of waters included. It has been shown that the ensemble mean energy obtained in Monte Carlo simulations of aqueous systems using force bias converges in about half the number of moves required for the same system using the straight Metropolis algorithm. Thus, the Clementi et al. base simulations used the equivalent of only 250,000 force biased equilibration steps and 250,000 force biased history production steps. Examination of any of the control functions for the 2,000,000 step force biased histories compiled here reveals that virtually none of the mean energies have settled to their final values after 500,000 moves much less 250,000 moves.

The simulations performed here differ from the Clementi et al. calculations in several ways. These include: 1) the use of periodic boundary conditions, 2) the use of sufficient waters in the simulation to constitute more than two complete hydration layers, 3) the use of sufficient simulation length to reasonably assure convergence and 4) the use of convergence acceleration techniques such as force bias and preferential sampling. Further, the simulations are structurally and energetically analyzed using well defined, consistent rules based on the proximity criterion. These differences in methodology produce the differences in structural and energetic results between the base simulations reported here and the Clementi et

al. base simulations. Because of the differences in methodology and analysis, it is felt that the simulations reported here contribute the most reliable information to date on nucleic acid base hydration.

These difficulties are presumably compounded in the complex helical DNA systems studied by Clementi where a large molecule (24 nucleotide units) is simulated with an inadequate number of waters (447) with inadequate run lengths ("between 894,000 and 2,235,000 moves") without periodic boundary conditions or convergence acceleration.

Several conclusions may be drawn from this study. A comparison was made of biomolecular hydration descriptions obtained with the use of three available biomolecular solute-water potential functions. Such a comparison, of primary importance in the improvement of the computer simulation of biomolecular hydration, has not previously been performed. Direct comparison of simulation produced transfer energies with experimentally obtained values for nine simulations distinctly demonstrates that the Berendsen-Van Gunsteren and Kollman sets provide excellent reproduction of transfer energetics in the molecular cases examined; the Clementi potentials, however, were found to greatly overestimate solute-water interactions. Examination of the first shell energetics of thymine and uracil using BVG and Kollman potentials suggest that an average first shell

solute-water pair interaction of about one kcal/mole is reasonable for the nucleic acid bases. Molecular first shell coordination numbers are found to be essentially independent of the choice of potential functions. Functional group coordination numbers do not show definite trends with the variation of potential function and are roughly transferable when rounded to the nearest integer. Finally, the simulation results provide qualitative microscopic characterization of the hydration of the five nucleic acid bases and two sugars in derivative form. The simulations were performed using methodology (periodic boundary conditions, adequate numbers of waters, convergence acceleration and sufficient run length for the number of molecules in the simulation) which I believe to supercede the previous simulation work of Clementi.

A very interesting focus for future simulation study is the effect of aqueous hydration on the base stacking interaction which is perhaps the primary contributor to helical stability. The identification of solute-water potential functions providing high quality solute-water energetics was the necessary prerequisite for this type of study. Simulation of aqueous stacked dimers would proceed using a potential of mean force Monte Carlo algorithm which calculates the energies of solvent influenced solute-solute interactions. The result is the identification of the most favorable solute-solute

configurations. Study of stacked base dimers moved on a vertical coordinate would determine the most favorable solute-solute separation while simulation of parallel displacement of one base in the dimer at a constant vertical distance would calculate the most favorable dimer overlap. Subsequent higher order systems for simulation would be vertically stacked base pairs where one pair is moved to simulate the relative displacement in base pairs in helical winding and unwinding. Sequence dependent effects would be an interesting facet of such a study. Finally, a study where one base in a stacked dimer is slid parallel to the base planes would simulate solvent effects on helical opening and unopening.

REFERENCES

1. H.DeVoe and I.Tinoco, J.Mol.Bio., 4, 500 (1962).

2. W.T.Astbury and F.O.Bell, Nature, 141, 747 (1938).

3. D.P.Riley and G.Oster, Biochim.Biophys.Acta, 7, 526 (1951).

4. W.E.Seeds and M.H.F.Wilkins, Farad.Soc.Discuss., Cambridge (1950).

5. M.J.Fraser and R.D.B.Fraser, Nature, 167, 761 (1951).

6. M.H.F.Wilkins, R.G.Gosling and W.E.Seeds, Nature, 167, 759 (1951).

7. L.Pauling and R.B.Corey, Proc.Nat.Acad.Sci., USA, 29, 84 (1953).

8. R.E.Franklin and R.G.Gosling, Acta Cryst., 6, 673 (1953).

9. J.H.Wang, J.Amer.Chem.Soc., 77, 258 (1955).

10. M.Falk, K.A.Hartman and R.C.Lord, J.Amer.Chem.Soc., 84, 3843 (1962).

11. M.Falk, K.A.Hartman and R.C.Lord, J.Amer.Chem.Soc., 85, 387 (1963).

12. M.Falk, G.Poole and C.G.Goymour, Can.J.Chem., 48, 1536 (1970).

13. G.Cohen and H.Eisenberg, Biopolymers, 6, 1077 (1968).

14. H.R.Drew and R.E.Dickerson, J.Mol.Bio., 151, 535 (1981).

15. W.Saenger, "Principles of Nucleic Acid Structure", SpringerVerlag, New York, 220-241 (1984).

16. M.J.B.Tunis-Schneider and M.F.Maestre, J.Mol.Bio., 52, 521 (1970).

17. V.I.Ivanov, L.E.Minchenkova, A.K.Schyolkina and A.I.Poletayev, Biopolymers, 12, 89 (1973).

18. J.C.Wang, Proc.Nat.Acad.Sci, USA, 76, 200 (1979).

19. A.Cnan, R.Kilkuskie and S.Hanlon, Biochemistry, 18, 84 (1979).

20. M.Levitt, Proc.Nat.Acad.Sci., USA, 75, 640 (1978).

21. G.N.J.Port and A.Pullman, FEBS Letters, 31, 70 (1973).

22. E. Clementi and G. Corongiu, J.Chem.Phys., 72, 3979 (1980).

23. J.E.Del Bene, J. Computational Chem., 2, 188 (1981).

24. J.E.Del Bene, J.Chem.Phys., 76,1058 (1982).

25. V.I.Poltev, I.I.Grokhlina and G.G.Malenkov, J.Biomol. Struct.Dyn., 2, 413 (1984).

26. V.I Danilov, I.S. Tolokh, V.I. Poltev and G.G.Malenkov, FEBS Letters, 167, 245 (1984).

27. V.I.Danilov and I.S.Tolokh, FEBS Letters, 173, 347 (1984).

28. D.Perahia,M.S.Jhon and B.Pullman, Biochim.Biophys.Acta, 474, 349 (1977).

29. E.Clementi and G.Corongiu, Chem.Phys.Letters, 60, 175, (1979).

30. E.Clementi and G.Corongiu, Biopolymers, 18, 2431 (1979).

31. E.Clementi and G.Corongiu, Int.Journal of Quantum Chem., 16, 897 (1979).

32. E.Clementi and G.Corongiu, Annals New York Academy of Sciences, 83 (1981).

33. G.Corongiu and E.Clementi, Biopolymers, 20, 2427 (1981).

34. E.Clementi and G.Corongiu, Biopolymers, 21, 703 (1982).

35. E.Clementi and G.Corongiu, Biopolymers, Int.Journal of Quantum Chem., 22, 595 (1982).

36. D.L.Beveridge, M.Mezei, P.K.Mehrotra, F.T.Marchese, G.Ravi-Shankar, I.Vasu and S.Swaminathan in "Molecular Based Study of Fluids", eds. J.M.Haile and G.A.Mansoori, ACS, Washington, D.C. (1983).

37. M.Rao, C.Pangali and B.J.Berne, Mol.Phys. 37, 1773 (1979).

38. J.C.Owicki and H.A.Scheraga, Chem.Phys.Lett., 47, 600 (1977).

39. S.Swaminathan and D.L.Beveridge, J.Amer.Chem.Soc., 99, 8392 (1977).

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

41. E.Clementi, F.Cavallone and R.Scordamaglia, J.Amer.Chem.Soc., 99, 5531 (1977).

42. R.Scordamaglia, F.Cavallone and E.Clementi, J.Amer.Chem.Soc., 99, 5545 (1977).

43. G.Bolis and E.Clementi, J.Amer.Chem.Soc., 99, 5530 (1977).

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

45. J.Blaney, P.Weiner, A.Dearing, P.Kollman, E.Jorgensen, S.Oatley, J.Burridge and C.Blake, J.Amer.Chem.Soc., 104, 6424 (1982).

46. G.Wipff, A.Dearing, P.Weiner, J.Blaney and P.Kollman, J.Amer.Chem.Soc., 105, 997 (1983).

47. W.C.Jorgensen, J.Chem.Phys., 77, 4156 (1962).

48. H.J.C.Berensen, Private Communication.

49. H.J.C.Berendsen, J.P.M.Postma, W.F.vanGunsteren and J.Hermans, in "Intermolecular Forces", ed. B.Pullman, Reidel, Dordrecht, 331 (1981).

50. A.B.Teplitsky, I.K.Yanson, O.T.Glukhova, A.Zielenkiewicz, W.Zielenkiewicz and K.C.Wierzchowski, Biophys.Chem., 11, 17 (1980).

51. J.K.Ahmed, G.A.W.Derwish and F.I. Kanbour, J.Sol.Chem., 10, 343 (1981).

52.J.N.Spenser and I.A.Judge, J.Sol.Chem., 12, 847 (1983).

53. A.Zielenkiewisz, W.Zielenkiewicz, L.F. Sukhodub, O.T.Glukhova, A.B.Teplitsky and K.C.Wierzchowski, J.Sol.Chem., 15, 757 (1984).

54. H.DeVoe and S.P.Wasik, J.Sol.Chem., 13, 51 (1984).

55. R.L.Scruggs, E.K.Achter and P.D.Ross, Biopolymers, 11, 1961 (1972).

56. F.I.Kanbour, J.K.Ahmed and G.A.W.Derwish, J.Sol.Chem., 12, 763 (1983).

57. B.Johnson, G.Karlstrom and S.Romano, J.Chem.Phys., 74, 2896 (1970).

58. S.F.Boys and F.Bernardi, Mol.Phys., 19, 553 (1970).

59. M.Mezei, S.Swaminathan and D.L.Beveridge, J.Chem.Phys., 71, 3366 (1979).