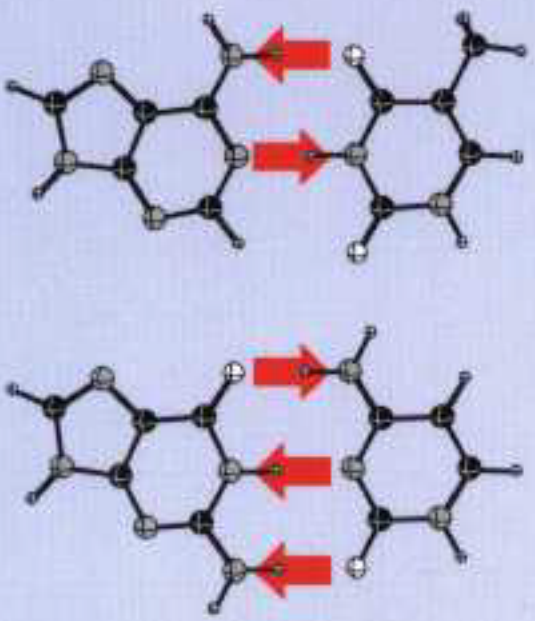
Structure and Bonding of DNA

Development and Application of Parallel and Order-N DFT Methods



Célia Fonseca Guerra

Structure and Bonding of DNA

Development and Application of Parallel and Order-N DFT Methods

Célia Fonseca Guerra

VRIJE UNIVERSITEIT

Structure and Bonding of DNA

Development and Application of Parallel and Order-N DFT Methods

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad van doctor aan
de Vrije Universiteit te Amsterdam,
op gezag van de rector magnificus
prof.dr. T. Sminia,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de faculteit der exacte wetenschappen / scheikunde
op dinsdag 22 februari 2000 om 13.45 uur
in het hoofdgebouw van de universiteit,
De Boelelaan 1105

door

Célia Fonseca Guerra

geboren te Amsterdam

Promotoren:	prof.dr. J.G. Snijders
	prof.dr. E.J. Baerends
Copromotor:	dr. F.M. Bickelhaupt

para meus pais

Contents

Chapter 1

General Introduction

The objective of the work presented in this thesis has been to make a contribution to the development of quantum biology by carrying out the first density functional theoretical (DFT) investigation on larger segments of deoxyribonucleic acid (DNA). The challenges associated with this objective are twofold. In the first place, we wish to describe the structure and energetics of the DNA segments accurately and, in particular, we try to achieve a better understanding of the nature and behavior of this complex molecule of heredity on the basis of its electronic structure. The computational effort connected with first-principles quantum chemical studies on these biochemical systems is enormeous and, until recently, calculations on these systems have been out of reach. Thus, finding and implementing speed-up techniques that make our model systems computationally accessible constitutes the other challenge of this work that, in fact, had to be tackled first. This chapter introduces a number of basic concepts concerning the speed-up techniques of parallelization and linearization treated in chapters 2 and 3. It also summarizes the chemical questions concerning DNA that will be addressed in chapters 4 - 6.

1.1 Speed-up Techniques

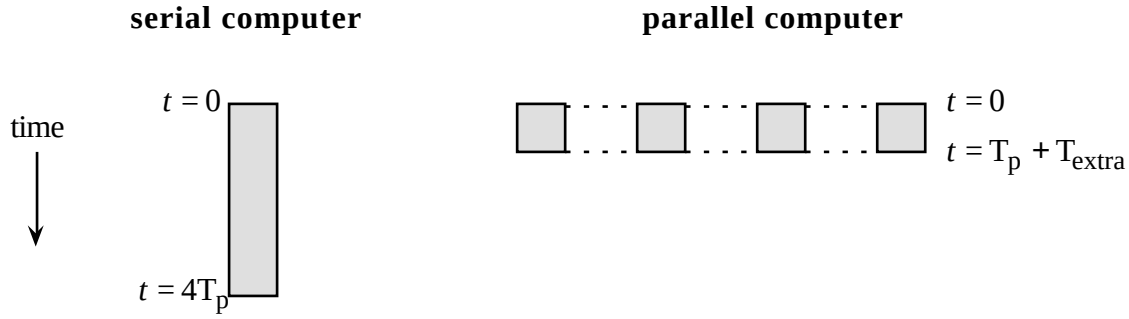
Over the years, the Amsterdam Density Functional (ADF) program[1] has proven to be a valuable and versatile quantum chemical tool for investigating problems in organic, organometallic and inorganic chemistry[2] both accurately and insightfully on the basis of density functional theory (DFT).[3] Relatively large model (reaction) systems, typically containing up to 30 atoms, can be studied routinely. Yet, this is small in view of the computational challenge provided already by small model systems for biological molecules, such as the DNA segments under consideration in this thesis. Not only do they contain much more atoms, namely up to 122, but they also lack symmetry, a property that otherwise would have reduced the computational effort. In addition, the weak hydrogen bonds that hold the DNA base pairs together are associated with very shallow potential energy surfaces. This further increases the computational cost as it imposes high demands on the numerical accuracy of a calculation (i.e. precision of numerical integration, SCF and geometry convergence criteria). Therefore, until recently, quantum chemical studies including geometry optimizations and vibrational analyses of such biological molecules were out of reach. The purpose of developing and implementing the speed-up techniques described in this thesis has been to eliminate these limitations in so far that state-of-the-art DFT computations on DNA segments (and other comparably large molecular systems) have become feasible.

Parallelization

The first speed-up technique used to reduce the computational time is the parallelization of the code. In this approach, the computational work is distributed equally over the processors or nodes of the parallel machine (see Scheme 1).

In the ideal case, the computational time on a parallel computer is just equal to the time of the computation on a serial machine divided by the number of nodes of the parallel machine. In Scheme 1, however, the elapsed time is set equal to $T_p + T_{extra}$. This extra time comes from the start-up time and the communication between the nodes during the calculations (represented with dashed lines in Scheme 1). All nodes must become aware of each other and know their part of the work and, at the end of the task, the outcome has to be sent to one of the nodes that gathers all data. Another aspect that can enlarge this extra time (not represented in Scheme 1) is an unequal

partioning of the work, the so-called load imbalance. Therefore, to obtain a perfectly parallelized program, the communciation and load imbalances have to be kept to a minimum. In chapter 2 the parallelization of the ADF program is described in detail and speed-ups obtained on different parallel architectures are given.



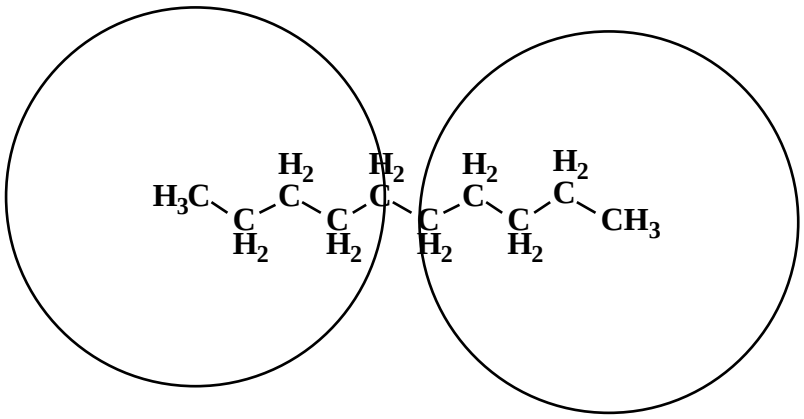
Scheme 1

Linearization

The second technique applied to reduce the computational cost is the so-called linearization of the code or order-N method. The quantum chemcial code is adapted in such a way that the time spent in a calculation scales linear with the number of atoms (N).

The decrease of computational cost is accomplished by removing from the calculation interactions between atoms that are at such a large distance from each other that they do not "feel" each others presence. A radius is defined for each atom such that, only when the radii of two atoms overlap, the matrix elements between them are calculated, otherwise they are skipped (see Scheme 2). This technique prevents the program from calculating an enormous amount of zeros.

The basic problem encountered in this linearization process is the definition of this radius. We need a mathematical expression that enables the program to determine for each atom a proper radius, which must remain large enough to conserve a certain accuracy of the calculation. Chapter 3 describes this linearization technique and shows the scaling that has been achieved for the various subroutines of the program.

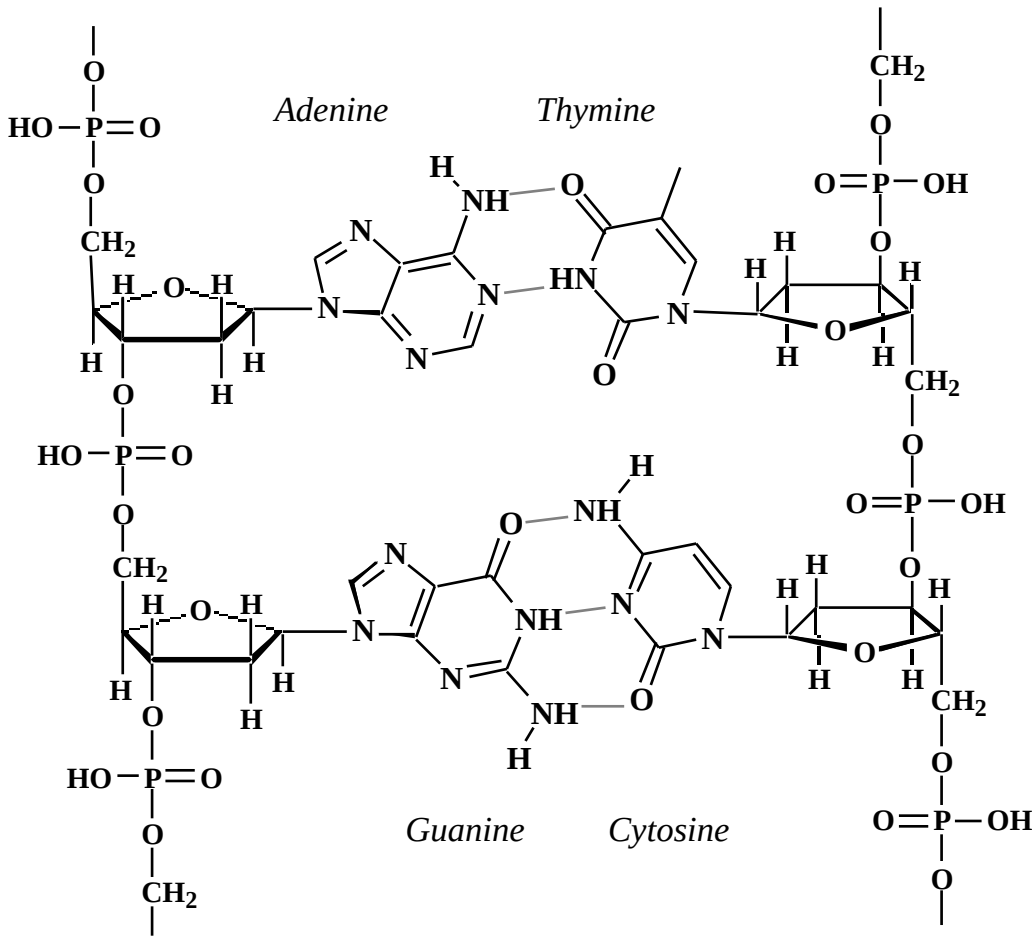


Scheme 2

1.2 DNA: Molecule of Heredity

Our genetic information is held in deoxyribonucleic acid or DNA molecules. These are very long, linear polymers made up of a large number of deoxyribonucleotides, each composed of one of four nitrogenous bases, a sugar (deoxyribose) and a phosphate group (see Scheme 3).[4] Two of the bases, adenine (A) and guanine (G), derive from purine, the other two, thymine (T) and cytosine (C), from pyrimidine. In 1953, James Watson and Francis Crick proposed in their famous article in Nature[5] the three-dimensional structure of DNA, which they had deduced from X-ray diffraction photographs. This structure consists of two right-handed helical chains of polynucleotides coiled around a common axis (see Scheme 3). The purine and pyrimidine bases are on the inside of the helix, whereas the phosphate and sugar units are on the outside. The planes of the bases are perpendicular to the axis of the helix and the hydrogen bonds between the bases hold the two chains together. The pairing occurs selectively between adenine and thymine and between guanine and cytosine. Furthermore, there is no restriction to the sequence of bases along a chain of nucleotides and it is this sequence of bases that carries the genetic information.

It was, however, only in the mid-seventies that crystallographic structures[4b,6] appeared on small segments of ribonucleic acid (RNA), which for the first time showed the Watson-Crick base pairing for AU and GC dimers. Nowadays, these data are supplemented by high resolution crystal structures[7] of much longer DNA oligomers. All this experimental work confirms the idea of Watson and Crick and contributes to our knowledge of the structure of DNA.



Scheme 3

Understanding the Structure and Nature of DNA

But *why* does this structure of DNA arise and, in particular, what causes selective molecular recognition between A and T, and G and C in the Watson-Crick base pairs? Is it the hydrogen bonds holding together the DNA base pairs that causes this selectivity and intrinsic stability of the genetic code? And what is really the nature of these hydrogen bonds, i.e., are they electrostatic phenomena reinforeced by resonance assistance of the π electrons as suggested by Gilli,[8] and how are they influenced by the environment in the crystal (or under physiological conditions)? These are basic questions concerning our conception of DNA and the genetic code and they lend themselves to be tackled quantum chemically.

In the past decade, *ab initio* and DFT quantum chemical studies[9] have appeared on the geometry and energy of simple models for AT and GC pairs. They were confined, however, to model base pairs in which the effect of the sugar-phosphate backbone was simulated by a methyl group at each base. Watson-Crick pairs of nucleosides or nucleotides, let alone oligomers of DNA base pairs, were beyond reach. Moreover, there is very little known about the nature of the hydrogen bonds in DNA base pairs. The current conception is still that of a predominantly electrostatic interaction. Based on a statistical evaluation of numerous X-ray cristallographic structures, Gilli et al.[8] have suggested that the electrostatic forces of hydrogen bonds between monomers with conjugated π-ellectrons systems (e.g., the bases in DNA base pairs) can be reinforced by resonance-assistance of the π-electrons, the so-called resonance-assisted hydrogen bonding (RAHB). So far, this hypothesis has not been verified by detailed quantum chemical analyses. Through an analysis of the electronic structure in the context of the Kohn-Sham molecular orbital (KS-MO) model and a corresponding quantative decomposition of the bond energy into the electrostatic interaction, the repulsive orbital interactions (Pauli repulsion) and the bonding orbital interaction (charge transfer and polarization), we tackle the open questions concerning DNA and the Watson-Crick base pairing mentioned in this introduction plus many other issues (e.g., the hypothesis of C–H•••O[10] hydrogen bonding in the AT pair). Chapter 4 briefly summarizes our key findings and how they lead to the unraveling of a hitherto unresolved discrepancy between experimental[4b,6] (X-ray crystallographic) and theoretical[8d,e,i-l] (both *ab initio* and hybrid DFT) structures of Watson-Crick base pairs. Thereafter, chapter 5 elaborates on the nature of the Watson-Crick base pairs. Chapter 6 presents the results of an extensive study on the effect of the incorporation of the sugar and phosphate group of the DNA backbone and the influence of the environment (water molecules and counterions). The thesis ends with a summary of our findings and an evaluation of the problems that remain to be solved before we can further proceed on the road toward quantum biology.

References

[1] a) C. Fonseca Guerra, O. Visser, J. G. Snijders, G. te Velde, E. J. Baerends, in *Methods and Techniques for Computational Chemistry*, (Eds.: E. Clementi, G. Corongiu), STEF, Cagliari **1995**, p. 305-395

b) E. J. Baerends, D. E. Ellis, P. Ros, Chem. Phys. 1973, 2, 41;

c) E. J. Baerends, P. Ros, *Chem. Phys.* **1975**, *8*, 412

d) E. J. Baerends, P. Ros, *Int. J. Quantum. Chem. Symp.* **1978**, *12*, 169

e) W. Ravenek, in *Algorithms and Applications on Vector and Parallel Computers*, (Eds.: H. H. J. Riele, T. J. Dekker, H. A. van de Vorst), Elsevier, Amsterdam, **1987**

f) P. M. Boerrigter, G. te Velde, E. J. Baerends, *Int. J. Quantum Chem.* **1988**, *33*, 87

g) G. te Velde, E. J. Baerends, *J. Comp. Phys.* **1992**, *99*, 84;

h) J. G. Snijders, E. J. Baerends, P. Vernooijs, *At. Nucl. Data Tables* **1982**, *26*, 483

i) J. Krijn, E. J. Baerends, *Fit-Functions in the HFS-Method; Internal Report (in Dutch)*, Vrije Universiteit, Amsterdam, **1984**

j) L. Versluis, T. Ziegler, *J. Chem. Phys.* **1988**, *88*, 322

k) L. Fan, L. Versluis, T. Ziegler, E. J. Baerends, W. Ravenek, *Int. J. Quantum. Chem., Quantum. Chem. Symp.* **1988**, *S22*, 173

l) J. C. Slater, Quantum Theory of Molecules and Solids, Vol. 4, McGraw-Hill, New York, 1974

m) L. Fan, T. Ziegler, *J. Chem. Phys.* **1991**, *94*, 6057

[2] For applications see, for example: a) T. Ziegler, *Can. J. Chem.* **1995**, *73*, 743

b) T. Ziegler, *Chem. Rev.* **1991**, *91*, 651

c) F. M. Bickelhaupt, *J. Comput. Chem.* **1999**, *20*, 114 and references cited therein.

[3] a) P. Hohenberg, W. Kohn, Phys. Rev. **1964**, *136B*, 864

b) W. Kohn, L.J. Sham, Phys. Rev. **1965**, *140A*, 1133

c) A. D. Becke, *J. Chem. Phys.* **1986**, *84*, 4524

d) A. Becke, *Phys. Rev. A* **1988**, *38*, 3098

e) S. H. Vosko, L. Wilk, M. Nusair, *Can. J. Phys.* **1980**, *58*, 1200

f) J. P. Perdew, *Phys. Rev. B* **1986**, *33*, 8822; Erratum: *Phys. Rev. B* **1986**, *34*, 7406)

[4] a) L. Stryer, *Biochemistry*, W.H. Freeman and Company, New York, **1988**

b) W. Saenger, *Principles of Nucleic Acid Structure*, Springer-Verlag, New York, Berlin, Heidelberg, Tokyo, **1984**

[5] J. D. Watson, F. H. C. Crick, *Nature* **1953**, *171*, 737

[6] a) N. C. Seeman, J. M. Rosenberg, F. L. Suddath, J. J. P. Kim, A. Rich, *J. Mol. Biol.* **1976**, *104*, 109

b) J. M. Rosenberg, N. C. Seeman, R. O. Day, A. Rich, *J. Mol. Biol.* **1976**, *104*, 145

[7] a) X.Q. Shui, L. Mc-Fail-Isom, G.G. Hu, L.D. Williams, *Biochemistry* **1998**, *37*, 8341

b) V. Tereshko, G. Minasov, M. Egli, *J. Am. Chem. Soc.* **1999**, *121*, 470

[8] a) G. Gilli, F. Bellucci, V. Ferretti, V. Bertolasi, *J. Am. Chem. Soc.* **1989**, *111*, 1023

b) P. Gilli, V. Ferretti, V. Bertolasi, G. Gilli, in *Advances in Molecular Structure Research, Vol. 2*, (Eds.: M. Hargittai, I. Hargittai), JAI Press, Greenwich, CT, **1996**, p. 67-102

[9] a) J. P. Lewis, O. F. Sankey, *Biophys. J.* **1995**, *69*, 1068

b) Y. S. Kong, M. S. Jhon, P. O. Löwdin, *Int. J. Quantum. Chem., Symp. QB* **1987**, *14*, 189

c) C. Nagata, M. Aida, *J. Molec. Struct.* **1988**, *179*, 451

d) I. R. Gould, P. A. Kollman, *J. Am. Chem. Soc.* **1994**, *116*, 2493

e) J. Sponer, J. Leszczynski, P. Hobza, J. Phys. Chem. 1996, 100, 1965

f) J. Sponer, J. Leszczynski, P. Hobza, *J. Biomol. Struct. Dyn.* **1996**, *14*, 117

g) J. Sponer, P. Hobza, J. Leszczynski, in *Computational Chemistry. Reviews of Current Trends*, (Ed.: J. Leszczynski), World Scientific Publisher, Singapore, **1996**, p. 185-218

h) M. Hutter, T. Clark, *J. Am. Chem. Soc.* **1996**, *118*, 7574

i) K. Brameld, S. Dasgupta, W. A. Goddard III, *J. Phys. Chem. B* **1997**, *101*, 4851

j) M. Meyer, J. Sühnel, *J. Biomol. Struct. Dyn.* **1997**, *15*, 619

k) R. Santamaria, A. Vázquez, *J. Comp. Chem.* **1994**, *15*, 981

l) J. Bertran, A. Oliva, L. Rodríguez-Santiago, M. Sodupe, *J. Am. Chem. Soc.* **1998**, *120*, 8159

[10] G. A. Leonard, K. McAuley-Hecht, T. Brown, W. N. Hunter, *Acta Cryst.* **1995**, *D51*, 136

Chapter 2

Parallelization of the Amsterdam Density Functional Program

The Amsterdam Density Functional (ADF) Program has been parallelized using the Single Program Multiple Data (SPMD) model. The subroutines dealing with the numerical integration or loops over the pairs of atoms appear in the serial calculations as the most time-consuming. Therefore, the integration points and the pairs of atoms are distributed in the beginning of the calculation over the nodes of the parallel machine in such a way that the load on the nodes is perfectly balanced. This static load balancing is used to minimize the communication between the nodes of the parallel machine as much as possible. The performance of the parallelized code has been tested on different platforms (distributed memory and shared memory parallel machines) for different types of calculations.

2.1 Introduction

This chapter concerns the parallelization of the code of the ADF program.[1] In section 2.2, the basic elements of the ADF methodology are discussed, which are used for the parallelization of the program. (For more information about the features of ADF see ref. [1a]). In section 2.3 an overview of the entire code is given and in section 2.4 some serial timing examples are presented. Next, we discuss considerations concerning the parallelization strategy (section 2.5) and the actual parallelization of the code (section 2.6). The performance of the parallelized code on different parallel platforms is given in section 2.7.

2.2 Methodological Details of ADF

In this section, we discuss the numerical integration and the fit of the density in ADF which are important features of the program for the process of the parallelization.

2.2.1 Solving the Kohn-Sham Equation Using a Basis Set Expansion

Basis Set

The Kohn-Sham equation[2]

$$h_{\mathrm{KS}}\phi_i(\mathbf{r}) = \left[-\tfrac{1}{2}\nabla^2 + V_{\mathrm{N}}(\mathbf{r}) + V_{\mathrm{C}}(\mathbf{r}) + V_{\mathrm{XC}}(\mathbf{r})\right]\phi_i(\mathbf{r}) = \varepsilon_i\phi_i(\mathbf{r}) \qquad (2.2.1)$$

where V_{N} is the nuclear potential, V_{XC} the exchange-correlation potential and the Coulomb potential V_{C} is given by

$$V_{\mathrm{C}}(\mathbf{r}_1) = \int \frac{\rho(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|} d\mathbf{r}_2 \qquad (2.2.2)$$

may be solved by expanding the solutions $\{\phi_i(\mathbf{r}_1), i = 1,\mathrm{n}\}$ in a set of basis functions $\{\chi_\mu(\mathbf{r}_1), \mu = 1,\mathrm{m}\}$, for which in ADF Slater type orbitals ($\chi_\mu(\mathbf{r}) = r^{n-1}e^{-\alpha r}Y_{lm}(\vartheta,\varphi)$) are used

$$\phi_i(\mathbf{r}_1) = \sum_{\mu=1}^{\mathrm{m}} C_{i\mu}\chi_\mu(\mathbf{r}_1) \qquad (2.2.3)$$

With this expansion equation (2.2.1) is transformed into a secular equation from which the eigenvalues and eigenfunctions are obtained:

$$\sum_{\mu=1}^{\mu=m} \left[F_{\nu\mu} - \varepsilon_i S_{\nu\mu} \right] = 0, \quad \nu = 1,m \tag{2.2.4}$$

with $F_{\nu\mu} = \int \chi_\nu(\mathbf{r}_1) h_{KS} \chi_\mu(\mathbf{r}_1) d\mathbf{r}_1$ (2.2.5)

and $S_{\nu\mu} = \int \chi_\nu(\mathbf{r}_1) \chi_\mu(\mathbf{r}_1) d\mathbf{r}_1$ (2.2.6)

Numerical Integration

The matrix elements $F_{\nu\mu}$ of the Fock matrix and other matrices are obtained by numerical integration[1g,h]

$$F_{\nu\mu} = \sum_k w_k \chi_\nu(\mathbf{r}_k) h_{KS} \chi_\mu(\mathbf{r}_k) \tag{2.2.7}$$

where w_k is a weight factor. In ADF the whole molecular volume is spatially partitioned into Voronoi cells.[1h] Each Vornoi cell is divided into an atomic sphere around the atomic nucleus and polyhedra.

Numerical integration requires the evaluation at all grid points of the values of all basis functions and of their second derivatives to obtain $\nabla^2\chi$. This also allows the calculation of the density (and derivatives of the density if required) to evaluate $V_{XC}[\rho](\mathbf{r}_k)$. The evaluation of the nuclear Coulomb potential at each point is trivial, but this is not the case for the Coulomb potential of the electronic charge distribution, which we will discuss next.

Coulomb Potential

The evaluation of the Coulomb potential at grid points will be explained in detail because of its importance for the parallelization. Analytical evaluation of the Coulomb matrix elements would require the calculation of two-electron integrals, with the characteristic n^4 problem (n is the number of basis functions). Numerical evaluation of the Coulomb matrix elements requires the evaluation of the Coulomb potential in the grid points, which leads to a large number (n^2p, p is the

number of grid points) of nuclear attraction type of integrals:

$$V_C(\mathbf{r}_k) = \int \frac{\rho(\mathbf{r})}{|\mathbf{r}-\mathbf{r}_k|} d\mathbf{r} = \sum_{\mu\nu} P_{\mu\nu} \int \frac{\chi_\mu(\mathbf{r})\chi_\nu(\mathbf{r})}{|\mathbf{r}-\mathbf{r}_k|} d\mathbf{r} \tag{2.2.8}$$

It is possible to alleviate either the n^4 or the n^2p problem by approximating the exact density with an expansion in one-centre fit functions[1b]

$$\rho(\mathbf{r}) = \sum_{\mu,\nu} P_{\mu\nu}\chi_\mu(\mathbf{r})\chi_\nu(\mathbf{r}) \approx \sum_i a_i f_i(\mathbf{r}) \tag{2.2.9}$$

The expression of the Coulomb potential in the integration point $\mathbf{r}_k$ now becomes

$$V_C(\mathbf{r}_k) = \sum_i a_i \int \frac{f_i(\mathbf{r})}{|\mathbf{r}-\mathbf{r}_k|} d\mathbf{r} \tag{2.2.10}$$

To determine the fit coefficients the density is split in a sum of densities of atom pairs

$$\rho = \sum_{A,B} \rho_{AB} \tag{2.2.11}$$

where $$\rho_{AB} = \sum_{\mu}^{A} \sum_{\nu}^{B} P_{\mu\nu}\chi_\mu \chi_\nu \tag{2.2.12}$$

and each ρ_{AB} is then approximated by an expansion of fit functions on atom A and fit functions on atom B

$$\tilde{\rho}_{AB} = \sum_i a_i f_i^A + \sum_j a_j f_j^B \tag{2.2.13}$$

Minimizing the difference between the exact and the approximated density $\int |\rho_{AB} - \tilde{\rho}_{AB}|^2 d\tau$ under the constraint that the number of electrons remains the same $\int \rho_{AB} d\tau = \int \tilde{\rho}_{AB} d\tau = N_{AB}$, leads for each atom or pair of atoms to the following set of linear equations:

$$\mathbf{Sa} = \mathbf{t} + \lambda\mathbf{n} \tag{2.2.14}$$

where

$$S_{ij} = \int f_i^A f_j^B d\tau \tag{2.2.15}$$

$$n_i = \int f_i^{A/B} d\tau \tag{2.2.16}$$

$$t_i = \sum_{\mu}^{A,\nu} P_{\mu\nu}^{AB} T_{\mu\nu i} \tag{2.2.17}$$

$$T_{\mu\nu i} = \int \chi_\mu^A \chi_\nu^B f_i^{A/B} d\tau \tag{2.2.18}$$

For the Lagrange multiplier we find from the charge preservation condition

$$\lambda = \frac{N_{AB} - \mathbf{n}^\dagger \mathbf{S}^{-1} \mathbf{t}}{\mathbf{n}^\dagger \mathbf{S}^{-1} \mathbf{n}} \tag{2.2.19}$$

As will be mentioned in the next section ADF makes use of symmetry, which for the fitting of the density implies that only the fit coefficients of the symmetry unique atoms and atom pairs are determined.

Exchange Correlation Potential

To be able to calculate the self-consistent solutions of the Kohn-Sham equation, the potential V_{XC} is derived from an approximate expression, $\tilde{E}_{\mathrm{XC}}$, for the exact exchange-correlation energy, E_{XC}. Various approximations have been implemented in ADF. The exchange can be approximated by the local-density approximation (Slaters Xα expression),[1m] with or without the non-local corrections due to Becke.[3a] The correlation can be treated in the local-density approximation using the Vosko-Wilk-Nusair (VWN) parametrization,[3b] also with or without non-local corrections proposed by Perdew.[3c] Also the nonlocal correlation corrections of Perdew and Wang[3d] have been implemented.

2.3 Code Structure

For a single point calculation (calculation at a fixed geometry) the ADF program consists of different parts as we can see from Figure 2.1 where the flow of the program is represented. In the first part, the input is read and then the numerical data to be used in the SCF part is prepared. The next part contains the SCF and in the last part the population and energy analysis is done.

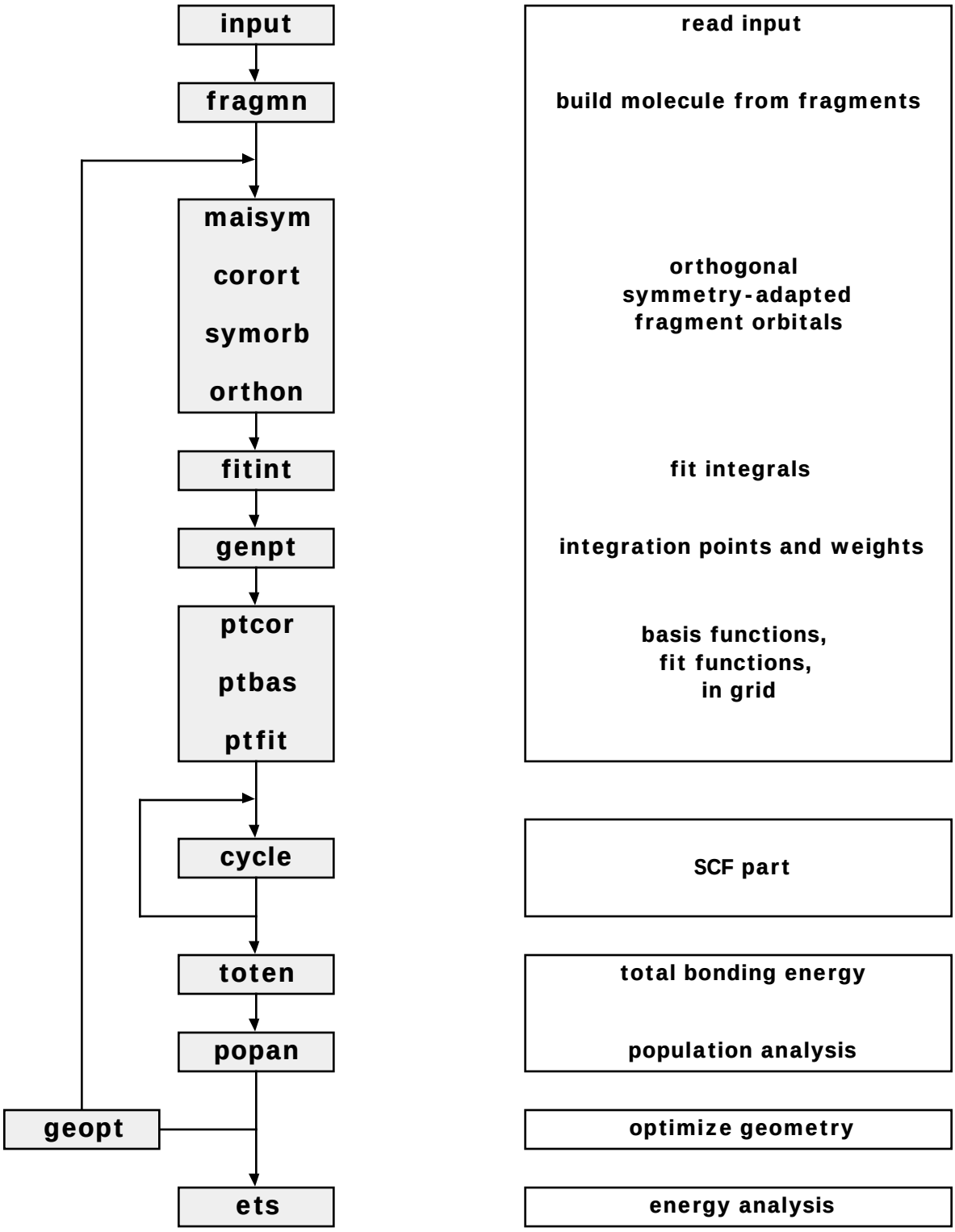


Figure 2.1. Flow of Amsterdam Density Functional (ADF) program.

If we look into more detail, we see that several setups are done in the second part before the SCF. After the input has been read and the molecule has been built from the (atomic or molecular) fragments, the fragment orbitals are orthogonalized to the core (CORORT), symmetrized (SYMORB) with symmetry information from MAISYM and then mutually orthogonalized (ORTHON) to construct the orthogonal symmetry-adapted fragment orbitals.

For the density fit the fit integrals (2.2.15), (2.2.16) and (2.2.18) are precomputed by the subroutine FITINT and written to file. Then the program generates the integration points and the weights of the integration points. The last three routines of this part calculate the value of the core density, the core Coulomb potential, the basis functions, fit functions and the fit Coulomb potential in the integration points and write results to file.

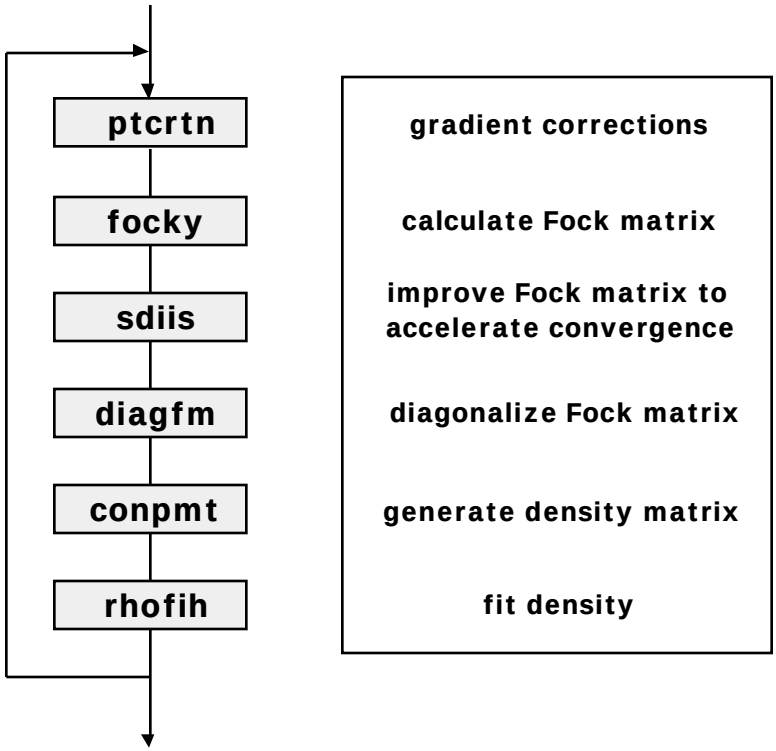


Figure 2.2. The SCF part of ADF program, handled by the routine CYCLE

To be able to do calculations on large molecules a direct SCF option has been included in ADF. Instead of reading the values of the functions in the integration points from file, they are

recalculated each iteration during the SCF. This saves a lot of disk space, but it also costs much more CPU time. All calculations described in this chapter use the direct SCF only for the fit functions. The values of the basis functions in the integration points are calculated once and written to file.

The SCF part is handled by CYCLE. It consists of six major subroutines (see Figure 2.2). In FOCKY the Fock matrix is calculated on the orthogonal symmetry-adapted basis set. To accelerate the SCF convergence the ADF program uses DIIS (Direct Inversion of Iterative Subspace).[4] The essence of the DIIS algorithm is that the results of the previous cycles are used to make a better guess for the new Fock matrix. The routine DIAGFM diagonalizes this Fock matrix. Finally, the subroutine CONPMT generates the density matrix on the primitive basis set from the calculated eigenvectors to enable RHOFIH to determine the fit coefficients for the new density. RHOFIH reads for each atom pair from file the integrals (2.2.18) and calculates from equation (2.2.14) the coefficients for that atom pair. This cycle from FOCKY to RHOFIH is repeated until convergence has been reached.

The density gradient corrections (GGA's) depend at a certain point on the first and second derivative of the density at that same point. So, to evaluate the Fock matrix the value of the GGA potential in the integration points is needed. Therefore, the derivatives of the fit functions in the integration points are calculated and knowledge of the fit coefficients makes it then possible to calculate the numerical value of these corrections in all integration points. The correction to the XC-potential is calculated by the routine PTCRTN, which is called before FOCKY in CYCLE. The value of the derivatives of the fit functions in the integration points is recalculated each iteration and the corrections in each integration point are written to file.

After convergence is reached, the population analysis and decomposition of the energy is done. TOTEN computes by numerical integration energy terms related to the following densities: the sum of fragment densities, the density ρ^0 of orthogonalized fragments (see ref [1a]), and the SCF density. Then POPAN performs a Mulliken population analysis on the basis of the primitive STO's (if required for individual orbitals) and calculates the atom-atom population matrix and the charge of the atoms. The last routine ETS calculates the interaction energy and performs the population analysis in terms of symmetrized fragment orbitals.

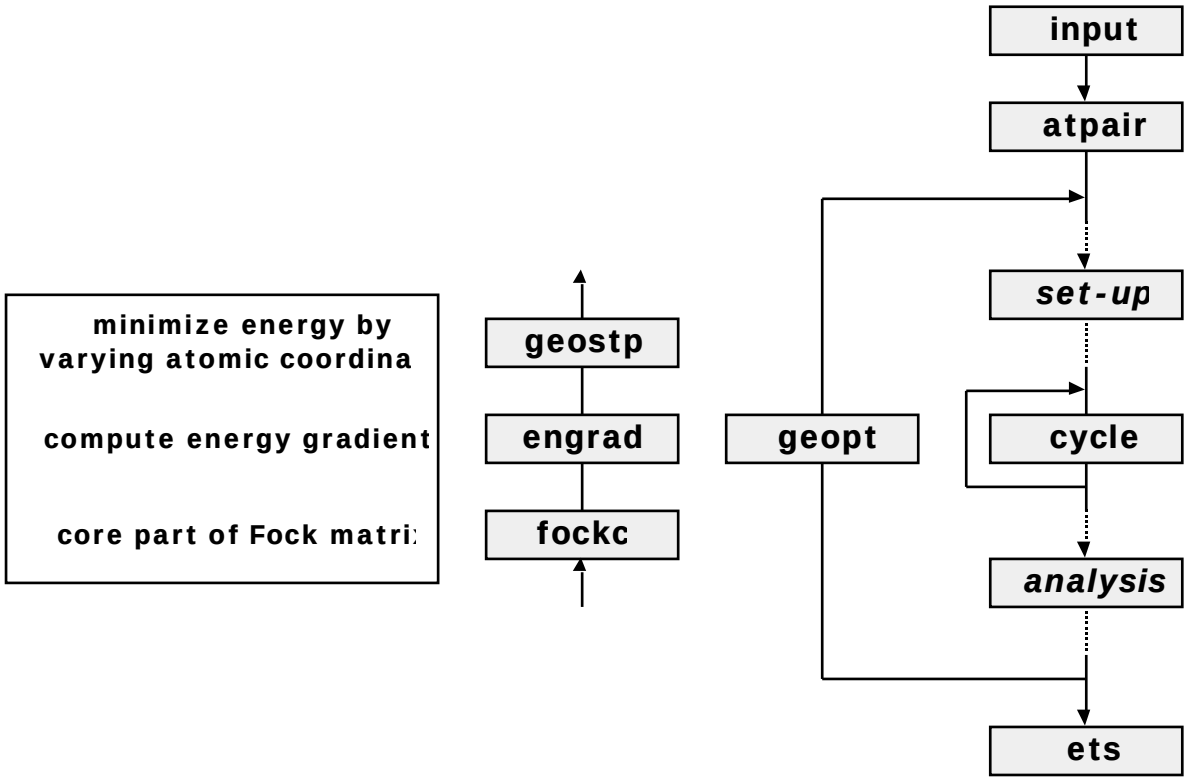
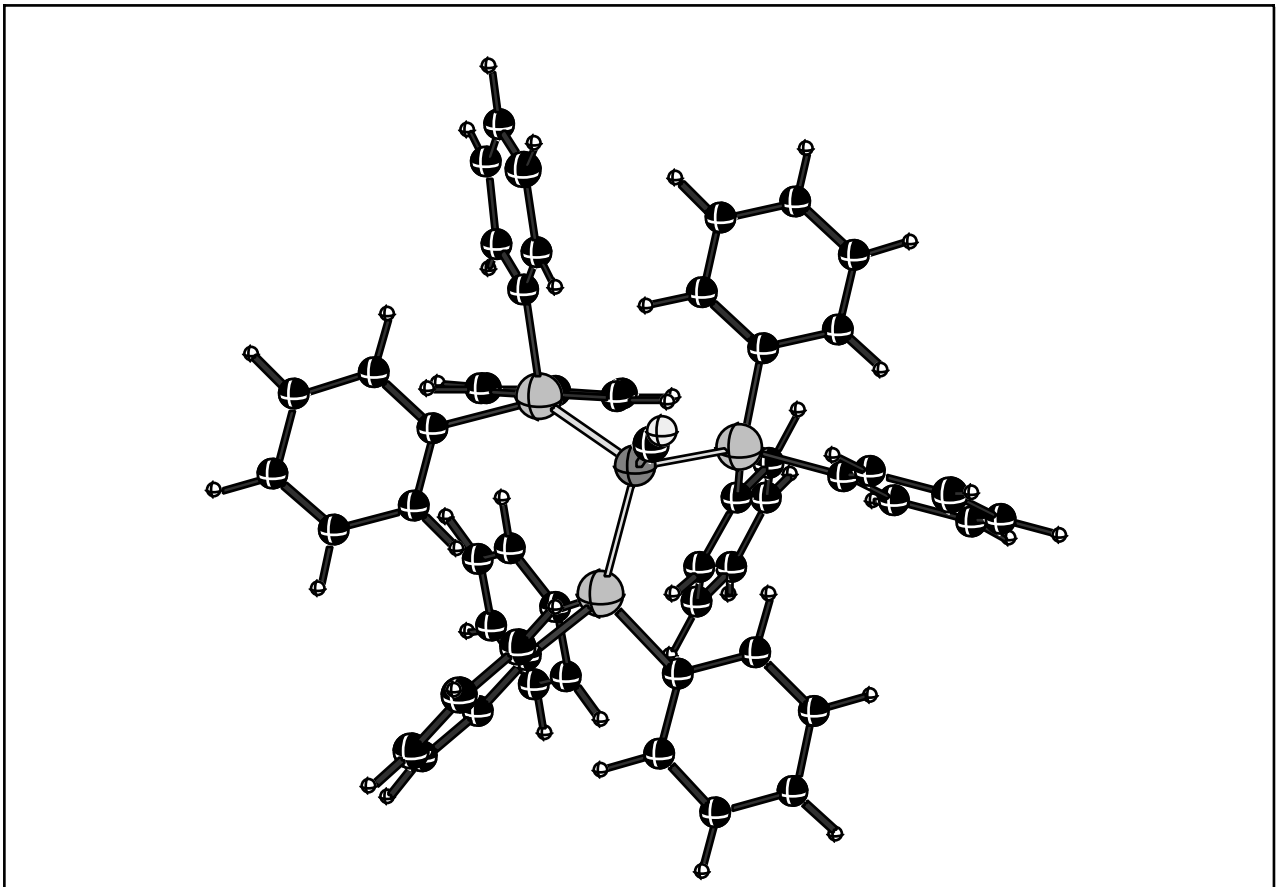


Figure 2.3. Flow of geometry optimization in ADF program.

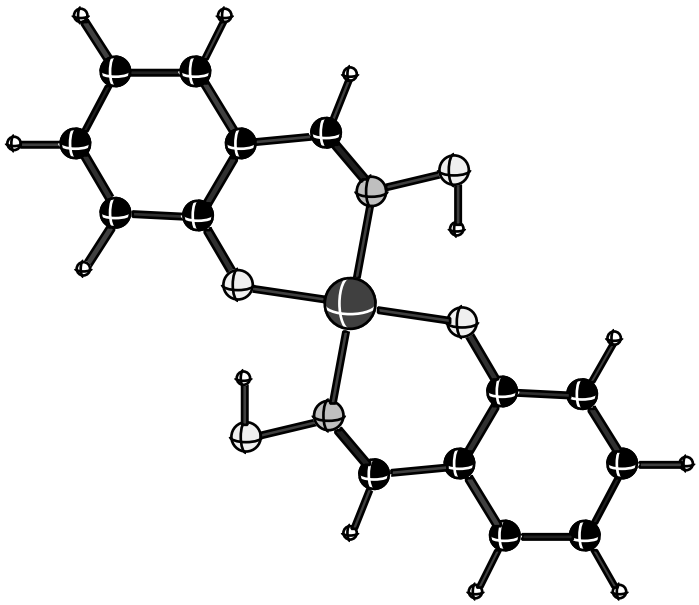
In Figure 2.3 (see also Figure 2.1) the flow of the program for a geometry optimization[1k,p] is shown. After reading the input the routine ATPAIR is called once to calculate the atom-pair electrostatic interactions for geometry updates. Then the geometry cycles are started.

The routine that handles the geometry optimization is GEOPT. After the population analysis has been done ADF calls GEOPT to generate the new atomic coordinates. Then ADF performs the whole setup from MAISYM to PTFIT again, it calculates the new converged energy and performs the population analysis, and as long as the geometry has not converged or the maximum number of geometry cycles has not been reached, GEOPT generates new atomic coordinates.

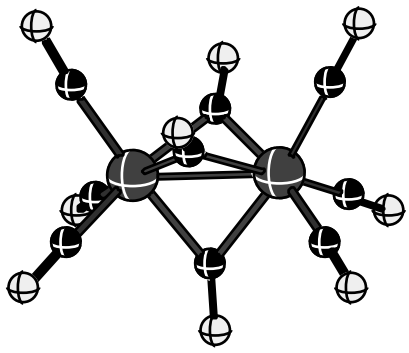
The routine GEOPT consists essentially of three routines, FOCKC, ENGRAD and GEOSTP as can be seen in Figure 2.3. FOCKC calculates the core part of the Fock matrix and overlap matrix, ENGRAD calculates the energy gradients and GEOSTP does a quasi-newtonminimization of the energy by varying the atomic coordinates. See for details section 2.2.5 in ref [1a].



a tris-(triphenylphosphine)-platinum(0)-carbonyl



b bis-(salicylatdoximato)-copper(II)



c diironnonacarbonyl

Figure 2.4. Molecules used for benchmark: $Pt(P(Ph)_3)_3CO$, $Cu(C_7H_6O_2N)_2$ and $Fe_2(CO)_9$.

2.4 Performance Analysis

In this section, we study the serial timing results for $Fe_2(CO)_9$, $Cu(C_7H_6O_2N)_2$ and $Pt(P(Ph)_3)_3CO$ (see Figure 2.4) on an IBM SP1. These molecules have been used for the benchmarking of the parallel machines. The serial timing results for $Pt(P(Ph)_3)_3CO$ were estimated from a two-node run, because we are not able to run this molecule on a single node of the SP1 due to lack of disk space. (A node is defined as a single processor.)

In the Tables 2.1 to 2.2 the timing results for a single point calculation with and without the gradient corrections to the density functionals are shown and the geometry optimization of $Fe_2(CO)_9$ with gradient corrections. For all time-consuming routines mentioned in the previous section the percentage of the overall CPU time is shown.

Table 2.1. Serial timing results for single point calculation

Molecule	$Fe_2(CO)_9$	$Cu(C_7H_6O_2N)_2$	$Pt(P(Ph)_3)_3CO$
ADF total CPU time	4.67 min	55.5 min	32.3 hours
ORTHON	.5%	.2%	.3%
FITINT**	13.8%	6.5%	3.4%
GENPT*	2.8%	.9%	.2%
PTCOR*	.5%	.3%	.1%
PTBAS*	6.5%	9.4%	20.6%
FOCKY*	50.3%	71.7%	66.0%
SDIIS	.7%	.4%	.3%
DIAGFM	.4%	.2%	.2%
CONPMT**	.3%	< .1%	< .1%
RHOFIH**	2.0%	.7%	.5%
TOTEN*	8.3%	8.2%	7.8%
ESTAT**	8.9%	0.9%	.2%

*Routine dealing with numerical integration

**Routine dealing with loops over atom pairs

We have used a double-ζ STO basis for all atoms. For carbon, oxygen and nitrogen the 1s orbitals have been frozen. For copper and phosphorus the core has been frozen up to 2p, for iron up to 3p, and for platinum up to 5p. The symmetries used for the different molecules are D_{3h} for $Fe_2(CO)_9$, C_i for $Cu(C_7H_6O_2N)_2$ and C_1 for $Pt(P(Ph)_3)_3CO$.

From Table 2.1 we see that FOCKY is the most expensive routine in all cases. In Figure 2.5 the serial timing results are graphically represented. We grouped together all routines dealing with numerical integration and all routines dealing with atom pairs. Evidently, the largest amount of time is spent in the routines for the numerical integration. Especially for large molecules with low symmetry the numerical integration is important. Next are the routines dealing with the atom pairs. If we parallelize all these routines, about 95% of the program is running in parallel.

When gradient corrections are added, we see from Table 2.2 that the numerical integration takes relatively even more time.

The derivatives of the energy with respect to the nuclear coordinates are also evaluated using numerical integration. Thus, for geometry optimizations the numerical integration code is even more important (see Table 2.3).

Table 2.2. Serial timing results for single point calculation with gradient corrections

Molecule	$Fe_2(CO)_9$	$Cu(C_7H_6O_2N)_2$
ADF total CPU time	13.8 min	2.76 hours
PTCRTN	58.6%	59.4%

Table 2.3. Serial timing results for geometry optimization with gradient corrections

Molecule	$Fe_2(CO)_9$
ADF total CPU time	22.1 hours
ATPAIR	.1%
GEOSTP	<< .1%
FOCKC	1.7%
ENGRAD	12.4%

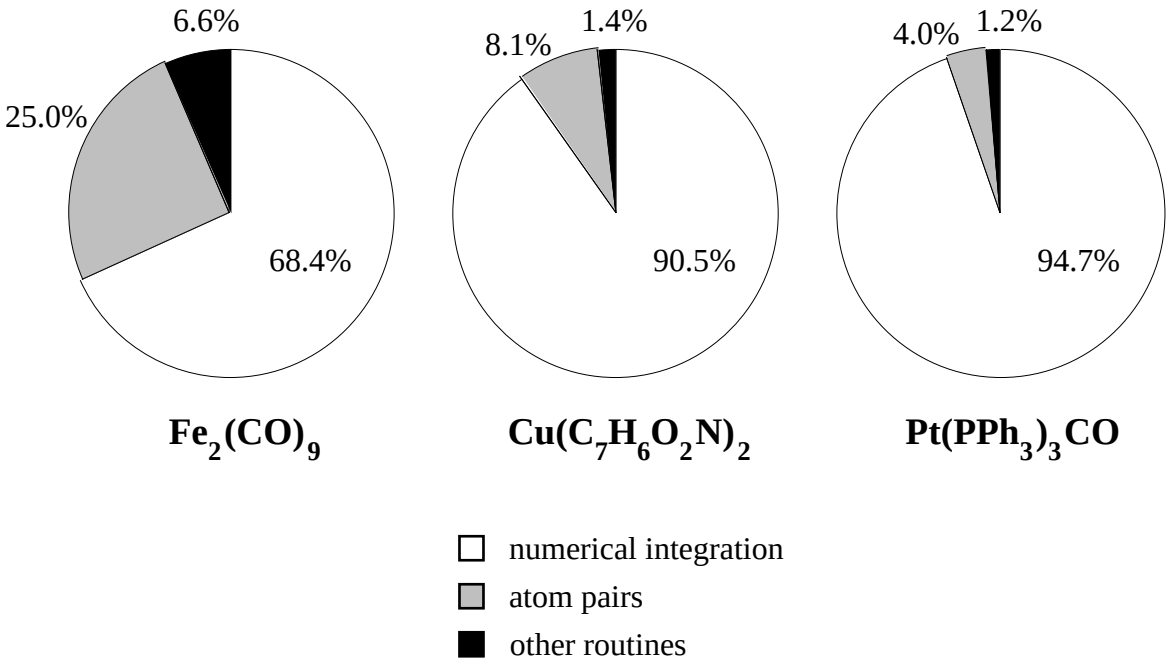


Figure 2.5. Diagram of the serial timing results for benchmark molecules.

2.5 Parallelization Model

2.5.1 Single Program Multiple Data Model

The parallelization paradigm that we use within ADF is the single program multiple data (SPMD) model. This model is shown in Figure 2.6. First ADF is started on one of the nodes of the parallel machine. This process is called the parent. Then the parent creates the child processes (normally on different CPU's), reads the input and broadcasts the input to the children. After this, we have a copy of ADF running on all the nodes with exactly the same data. In the serial parts of the program all these copies perform the same calculation, duplicating each others work. In the parallel parts each copy handles a part of the problem and the results are combined. After all copies have finished their calculation, the child processes are stopped.

One of the advantages of the SPMD approach is that we can reduce the serial and increase the parallel part of the code step by step during the parallelization process. As a result we have always a running program. This not only makes debugging much easier, but also the determination of the most time-consuming serial parts that are left over.

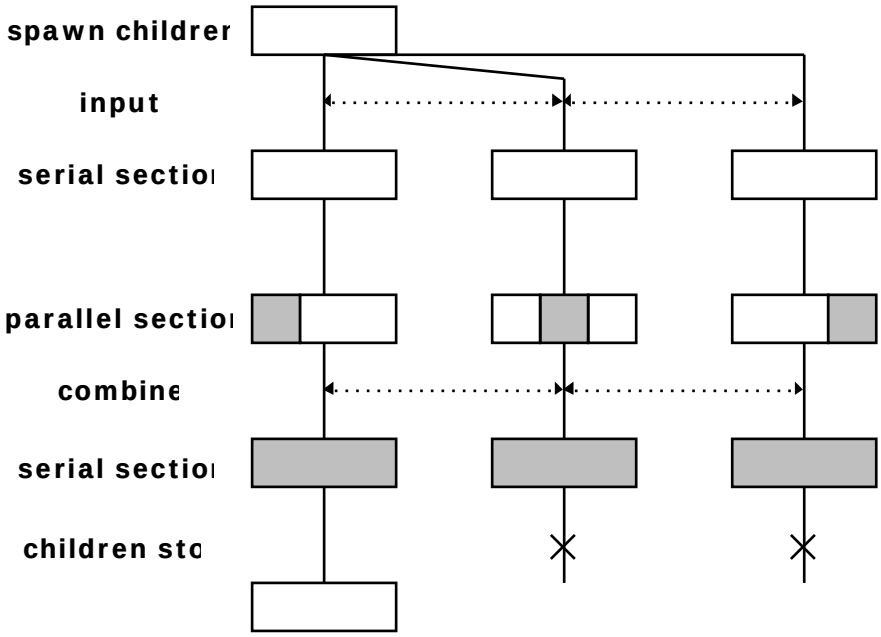


Figure 2.6. Single Program Multiple Data (SPMD) model.

Another advantage is that the structure of the parallel program is virtually identical to the structure of the serial program. This makes it much easier to maintain and extend the functionality of both versions simultaneously. In fact no separate serial source code exists.

Duplicating the work in the serial sections (leading to data replication) reduces the amount of communication: the results of the serial sections need not to be sent to other nodes. If we would avoid the serial section, the other nodes would have to wait anyway. So, there is no negative effect on the total elapsed time.

2.5.2 Load Balancing

For the load balancing there are two possibilities: the static and the dynamic load balancing. We expected that communication between the nodes would be expensive and might become a bottleneck on large parallel machines. So, we wanted a balancing that would lead to as little communication as possible to get a scalable parallel program. This requirement of small communication time is satisfied by coarse grain static load balancing. The data is kept local and only a very small amount of data is combined over the nodes.

Although dynamic load balancing has the advantage that the parallel program can perform reasonably on loaded machines because of its ability to adapt to external circumstances, we expected that the parallelization of ADF would benefit more from a static load balancing. The dynamic load balancing did not satisfy our demand of small communication time because the repeated distribution of the data requires much more communication. In the case of dynamic load balancing each time a task has finished its work, there has to be communication to assign more work to that task, while in the case of the static load balancing this has to be done only once.

The static load balancing in combination with the SPMD model allows us to generate most data locally and keep it on file or in memory on the same node throughout the execution of the program. The partioning of data files over the nodes gives an enormous reduction of disk requirements per node. The use of distributed matrices lowers to some extent the memory requirements. The resulting files are N times smaller on an N-processor machine and the size of some matrices is compressed. If the hardware allows it, I/O can be done in parallel which gives a decrease of I/O elapsed time. So, in our case parallel machines which have the capability to perform I/O in parallel are preferred.

2.5.3 Communication

For the communication between the nodes we have used for portability reasons the public domain software library PVM,[5] which has been designed to treat a collection of possibly heterogeneous computers as one computer, the so-called parallel virtual machine.

The PVM system consists of two parts. The first part is a daemon that runs on all computers forming the virtual machine. The second part provides a library of PVM interface routines for spawning processes and message passing. The processes that are spawned by PVM, such as the copies of ADF on the different nodes, are given a task identifier. In ADF the task identifier is held in MYTID. The routines based on PVM that are used for combining data, like PPCBNR, are discussed in section 2.6.5.

For the SP2 we have used the vendor implementation PVMe. This enables us to make use of the high speed communication hardware of the SP2. A disadvantage of PVMe is that you are not allowed to use more than 128 nodes.

2.6 Parallelization of ADF

2.6.1 Numerical Integration

In the ADF program most of the integrals are calculated using numerical integration. The serial timing results in section 2.4 showed us that the numerical integration is also the most time-consuming part of the ADF program. In this section, we discuss the parallelization of the numerical integration first and then the parallelization of the generation of the integration points.

Use of Numerical Integration

The numerical integration is parallelized by distributing the integration points over the nodes of the parallel machine. The integration points are collected in NBLOCK blocks, each with LBLOCK integration points. These blocks are distributed over the nodes. This distribution is determined by the subroutine PPIBLK. The function IB2TID returns for each block the task identifier of the task it belongs to. All routines dealing with integration points (FOCKY, PTCRTN etc.) have been parallelized. The balancing of the numerical integration is close to perfect since the amount of work required for each integration point does not depend on the integration point considered.

```
do iblock = 1, nblock
      if (ib2tid(iblock) = mytid) then
         read r_k
         for all i
            do k = 1, lblock
               calculate χ_i(r_k)
            continue
            write χ_i(r_k) to file
      ifend
continue
```

Figure 2.7. Pseudo code for calculation of basis functions in the integration points.

In Figure 2.7 pseudo code for the parallelization of the calculation of the basis functions in the integration points is shown. The loop over the blocks of integration points is the same as in the serial code. If the block is one of the blocks to be handled by this task, the values of all the basis functions in the integration points of that block are calculated. The variable MYTID gives the task identifier of this ADF task. After finishing the loop, each node holds on a local file the value of the basis functions in the integration points belonging to that node.

The elements of the Fock matrix are calculated using numerical integration. In Figure 2.8 we present pseudo code for the parallelization of this routine. The loop is again the same as in the serial code. First, all matrix elements are initialized. Next, the loop over the blocks of integration points is started. Again if the block is one of the blocks to be handled by this task, the partial results are calculated for all the matrix elements. All data required to evaluate the Fock matrix is locally available. After finishing the loop the incomplete Fock matrices are summed by the subroutine PPCBNR.

This approach of parallelization of the numerical integration, where the data per point is held locally and partial results are combined, minimizes the rather time-consuming communication between the tasks.

```
F_ij = 0
do iblock = 1, nblock
      if (ib2tid(iblock) = mytid) then
         read r_k , χ_i(r_k) etc.
         for all i, j
            do k = 1, lblock
               F_ij = F_ij + w_k χ_i(r_k) F χ_j(r_k)
            continue
      ifend
continue
ppcbnr (F, fsize)
```

Figure 2.8. Pseudo code for the calculation of Fock matrix with numerical integration.

Generation of the Integration Points

After having parallelized all routines involving numerical integration, we saw in our timing results that the generation of the integration points (GENPT) takes a significant amount of time. The standard integration scheme of ADF generates the integration points for three different parts of the three-dimensional space; the atomic spheres, the atomic polyhedra, and the layers around the molecule.

Most of the time required by GENPT is consumed by the routines which actually generate the points. The routines generating the points around the atoms (in the atomic spheres and in the atomic polyhedra) have been parallelized by distributing the atoms over the tasks. The routine which generates the integration points in the outer region has been parallelized by distributing the different layers of the outer region. The result is that each task generates only part of the points. For large molecules this load balancing turns out to be satisfactory.

After the generation of the integration points all the nodes have their own set of integration points, possibly not very well balanced. To acquire a good load balancing for the numerical integration some of the points need to be moved to other nodes, so that all the nodes have exactly the same number of integration points. First the routine PPIBLK is called to determine the total number of integration points and to distribute the blocks of integration points equally over the nodes, so that routines dealing with the integration points have a good load balance. Then the routine PPRDSU is called to determine which node has integration points in excess and which node needs additional integration points and also to determine the redistribution of the points such that the actual communication is minimal. After this setup phase, the routine WRINPT loops over all blocks of integration points. If a block of integration points needs to be handled on this node, WRINPT tries to get the points from the list of points generated on this node, but when there are no more points available WRINPT uses PPGATH to get the required points from another node. The complete blocks of points on this node are written to file. After the loop over the blocks has been finished the remaining integration points are made available to the other nodes by calling the routine PPSCAT. At the end of WRINPT, each node holds only those blocks of integration points which it will handle during the rest of the run.

Geometry Optimization

The geometry optimization consists essentially of three routines: FOCKC, ENGRAD and GEOSTP. As mentioned in section 2.4 the execution time of GEOSTP is negligible, so we have not parallelized this routine. The parallelization of the other routines was rather straightforward: we could apply the same techniques as we used for the numerical integration.

The routine FOCKC, which calculates the core part of the Fock and overlap matrices, has the same structure as FOCKY which calculates the valence Fock matrix. It loops over all blocks of integration points, reads relevant data per point (such as core density, core potential), and adds everything together to get the Fock matrix. So, the loop over the blocks of integration points in FOCKC has been parallelized in the same way as FOCKY.

ENGRAD calculates the gradient of the energy with respect to the nuclear coordinates. It consists of a big loop over all blocks of integration points. For all integration points in that block the relevant contributions to the integrals making up the gradients are calculated and summed up. This routine has also been parallelized by distributing the blocks of integration points and combining the results at the end.

The routine ATPAIR that is called once for the geometry optimization, has also been parallelized. It calculates atom-atom electrostatic interaction energies that are used by the geometry optimization algorithm. ATPAIR consists of a loop over all atom pairs. For each atom pair it calls PAIRPT to calculate the electrostatic interaction. PAIRPT loops over a number of atom-atom distances, and for each of these distances the required data is calculated.

PAIRPT has been parallelized by distributing the different distances over all the nodes. The alternative is to distribute the atom pairs, but then it is more difficult to obtain a good load balance. However, since the number of distances is limited (42 in the current implementation) this routine scales only up to 42 nodes. If the time required by this routine would become excessive, we might parallelize it further by distributing both atom pairs and atom-atom distances.

After each node has calculated the data for their assigned distances, the results are distributed over all nodes using PPGDV *(vide infra)*.

2.6.2 Atom pairs

In the ADF program we distinguish two kinds of atom pairs: the symmetry-unique atom pairs and all atom pairs. For the density fit only the symmetry-unique atom pairs are used, but in other parts of the program the data of all atom pairs are needed. In the next part, we discuss first the distribution of the symmetry-unique pairs for the density fit and then of all atom pairs.

Symmetry-unique Atom pairs

The distribution of the atom pairs is much more involved than the distribution of the integration points. The symmetry-unique atom pairs are used by FITINT, RHOFIH and CONPMT. For these atom pairs CONPMT calculates the density matrix on the primitive basis set, and FITINT and RHOFIH perform the density fit. The problem is the determination of the cost of an atom pair. We have on one side the routines FITINT and RHOFIH, and on the other side routine CONPMT. Because FITINT takes much more of the overall execution time than RHOFIH or the other routines, we have determined experimentally the cost of an atom pair in FITINT. From our timing information we obtained that the cost of an atom pair was approximately proportional to the number of integrals $T_{\mu\nu i}$ (2.2.18) used to calculate the vector **t** (2.2.17) in RHOFIH.

The routine PPPAIR, which does the load balancing for the atom pairs, distributes the atom pairs by first sorting the atom pairs according to decreasing number of integrals $T_{\mu\nu i}$ and then assigning the next atom pair to the node with the lowest cumulative weight. The function IP2TID returns information about this distribution.

The load balancing achieved by this distribution is close to perfect for FITINT when there are enough atom pairs to be distributed. It also works very well for RHOFIH, because in RHOFIH the most time-consuming part is the calculation of the vector **t** (2.2.17) and this is of course proportional to the number of integrals $T_{\mu\nu i}$.

In Figure 2.9 pseudo code for FITINT and RHOFIH is presented. Here, we see again the advantage of the static load balancing, as we saw in the case of the numerical integration. The routine that does the preparation, in this case FITINT, writes the data to a local file on the node, and the routine which actually does the calculation, in this case RHOFIH, reads the data from the local file and calculates the fit coefficients. Only a relatively small amount of data, the fit coefficients, has to be combined at the end of RHOFIH.

The routine CONPMT uses the same distribution of the atom pairs as RHOFIH and FITINT. We can see from equation (2.2.17) that only the elements of the density matrix that belong to the atom pairs on a node are needed on that particular node. Also the density of the symmetry equivalent atom pairs is not fitted by RHOFIH, so the elements of the density matrix of those atom pairs are not needed and therefore not calculated.

```
do ipair = 1, npair
      if (ip2tid(1,ipair) = mytid) then
         Calculate fit integrals for ipair
         Write to file
      ifend
continue
```

```
do ipair = 1, npair
      if (ip2tid(1,ipair) = mytid) then
         Read from file
         Fit density for ipair
      ifend
continue
ppcbnr (fit coefs)
```

Figure 2.9. Pseudo code of FITINT and RHOFIH: routines for the fitting of the density.

```
do ipair = 1, npair
      if (ip2tid(1,ipair) = mytid) then
         for all μν of ipair
            Calculate P_μν
      ifend
continue
```

Figure 2.10. Pseudo code for the construction of the density matrix in CONPMT.

In Figure 2.10 pseudo code for the calculation of the density matrix is represented. Again we benefit from the choice of the static load balancing. The elements of the density matrix that are calculated on a particular node by CONPMT are also the elements of the density matrix that are needed in RHOFIH to fit the density as we can see from equation (2.2.17). No communication is needed in this case. This static distribution gives us also another advantage. Since only part of the density matrix is calculated, we need much less memory to store it. So, we achieve a compression of the needed memory that scales with the number of processors.

In the routine CONPMT the cost of an atom pair is proportional to the square of the number of primitive basis functions of that atom pair. It is not obvious that the chosen balancing, which is based on the balancing for FITINT, would be proper for this routine, but our timing results show that the CPU times are well balanced. The approximation would be justified if the number of fit functions would be equal for all atom pairs. Deviations are not important since CONPMT does not take much time.

All Atom pairs

In the ADF program there are some routines with loops over the atom pairs that do not use symmetry. For instance CLSMAT, which calculates the overlap matrix on the primitive basis, and CNTPMT which calculates the total density matrix on the primitive basis. For these routines the cost of an atom pair is proportional to the number of elements of the overlap or density matrix. The distribution of the atom pairs is done by PPPAIR by first sorting the atom pairs according to decreasing number of elements of the density matrix and then assigning the next atom pair to the node with the lowest cumulative weight. The function IP2TID returns information about this distribution.

The loop for constructing the density matrix in the routine CNTPMT looks almost the same as the loop in CONPMT for constructing the partial density matrix. The only difference in that loop is that IP2TID (1,IPAIR) is replaced by IP2TID (2,IPAIR) to handle all atom pairs and the density matrix of the different nodes is combined at the end of the loop.

2.6.3 PP Library

We implemented a high level parallel library to hide the low level message passing code from the application programmer. This library makes it possible to use several message passing interfaces without making changes to other parts of ADF. Currently a PVM, PVMe and MPI version of the library is available.

Besides the routines that combine incomplete or distributed matrices, the PP Library also contains routines that find the maximum or minimum value of a variable over the nodes. Other utilities are for instance the barrier routine PPBARR, which synchronizes all the nodes by having all children send a message to the parent and wait for the parent to reply. It is mainly used to measure the pure elapsed time of the individual routines without time spent in the other routines.

The library also contains a routine PPINIT which initializes the parallel version of ADF. This routine is called by the parent and the children. The parent reads the first line of input, and based upon that line generates a number of children. After creating the children, PPINIT waits in a barrier so after the call to PPINIT one can assume that the children are alive and well. All tasks will be added to the PVM group 'ADF'. Furthermore, the library contains high-level communication routines for combining incomplete or distributed matrices. For these routines a number of algorithms are available. At installation time the user can decide the algorithm to use. We have defined the incomplete matrices as those that have the same size as in a serial run, but the matrices on the different nodes contain only a part of the total result. So, the local matrices have to be summed over the nodes to get the final matrix. The distributed matrices on the other hand have their elements distributed over the nodes. Their size per node will be smaller than in a serial run. To get the total matrix the distributed elements have to be gathered.

For the calculations on one of the platforms (SP2) we used the IBM implementation of PVM (PVMe). It enables us to utilize the high speed communication hardware of the SP2. However, the calls in PPINIT had to be adapted to make use of PVMe due to incomplete compatibility with the public domain version of PVM.

Combining Incomplete Matrices

For the routine that handles the combining of the incomplete matrices, PPCBNR, three different algorithms[6] have been implemented.

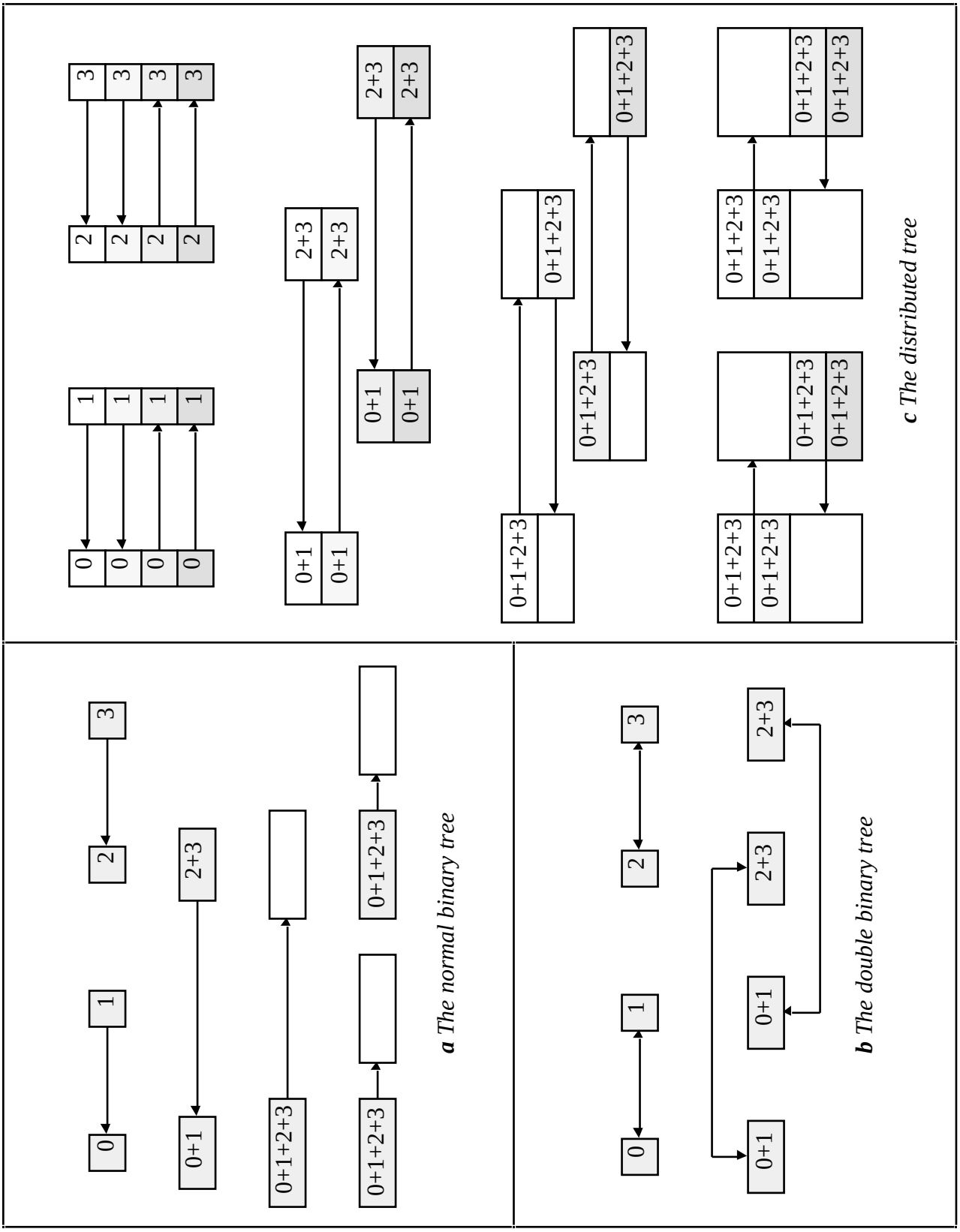


Figure 2.11. Algorithms for combining incomplete matrices residing on all nodes. The different shades of grey in the distributed tree algorithm represent different parts of the matrix on the nodes.

The first algorithm is the well-known binary tree algorithm, see Figure 2.11a. In this algorithm each even node gets the incomplete matrix from the uneven node on the right and adds these two incomplete matrices. In the next step the uneven nodes are left out and the process is repeated for the remaining nodes. This continues until the total matrix is on the parent. Then the total matrix is broadcast to the children by walking up the binary tree.

In Figure 2.11b the "double binary tree" algorithm is presented. In the first step each pair of nodes next to each other exchange their incomplete matrix and sum these up. In the next step this is done again but now between the pair of nodes that are one node further away from each other. This exchanging and summing is continued until all the nodes contain the total matrix.

In double binary tree algorithm all the nodes send and receive at the same time and they also sum the incomplete matrices at the same time. The advantage of this algorithm is that less communication is required. In the normal binary tree the number of communication steps is equal to 2logN (N = number of nodes). But with the double binary tree the number of communication steps is only equal logN.

The last algorithm, the "distributed tree" algorithm is shown in Figure 2.11c. This algorithm does the summation of the matrix elements in parallel. Each pair of nodes next to each other exchange half of their incomplete matrix and sum this up. In the next step the pair of nodes that are one node further away from each other repeat this process. The half of the incomplete matrix that they contain, is divided in two parts. One quarter of the incomplete matrix is send to the other node and the quarter that remains on the node, is summed up with the part received from the other node. The process is repeated until each node contains a part of the total matrix. So, the total matrix is now distributed over the nodes. Each part of the total matrix on a node is equal to the number of elements of the total matrix divided by the number of nodes. To get the total matrix on all the nodes, each node exchanges its part with the node that has been the last one with which it has communicated. This process is repeated until finally half of the total matrix is on each node and is exchanged between the two neighbor nodes.

This algorithm is particular suitable when the summation of the matrix becomes expensive, but the communication is relatively cheap. In the double tree algorithm each node did the same summation but in the distributed tree algorithm the summation is done in parallel. The number of communication steps is equal to 2logN.

For simplicity we have assumed that the number of nodes equals a power of 2. For the binary tree this is not a problem. The other two algorithms are only used for the largest power of two partition of the nodes. The data from the nodes in excess are explicitly handled before and after the combine requiring two additional steps.

To choose the algorithm to be used on the parallel platform, we have measured the three different algorithms on the machines that are used. The platforms used are a workstation cluster of 6 IBM RS6000/250 workstations connected by ethernet, an 8-node IBM SP1 with FDDI for communication and the IBM SP2 with the high performance switch connecting the processors.

The three algorithms were used to combine a matrix of 1000 elements and a matrix of 100000 elements. In Figure 2.12, 2.13 and 2.14 the results for the combining of the largest matrix are given for the workstation cluster, the SP1 and the SP2. It shows that the fastest algorithm on de cluster is the binary tree. For the SP1 we also choose the binary tree. On the SP2, however, the distributed tree is much faster than the other two so PPCBNR uses the distributed tree algorithm on the SP2. The results for the smaller matrix give the same conclusions.

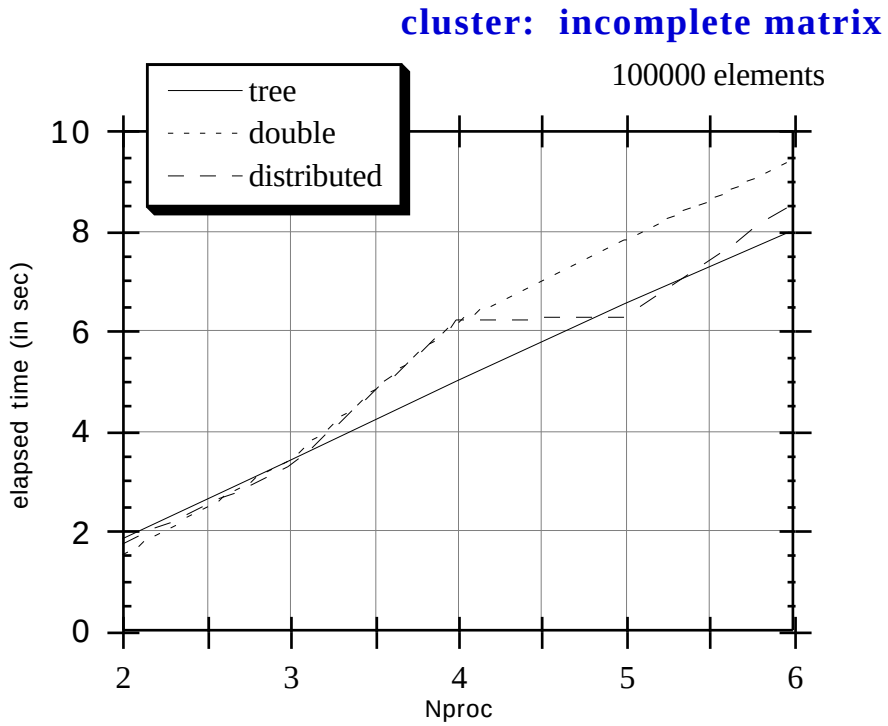


Figure 2.12. Comparison of algorithms for combining on a workstation cluster. (Nproc = number of nodes/processors of the parallel machine)

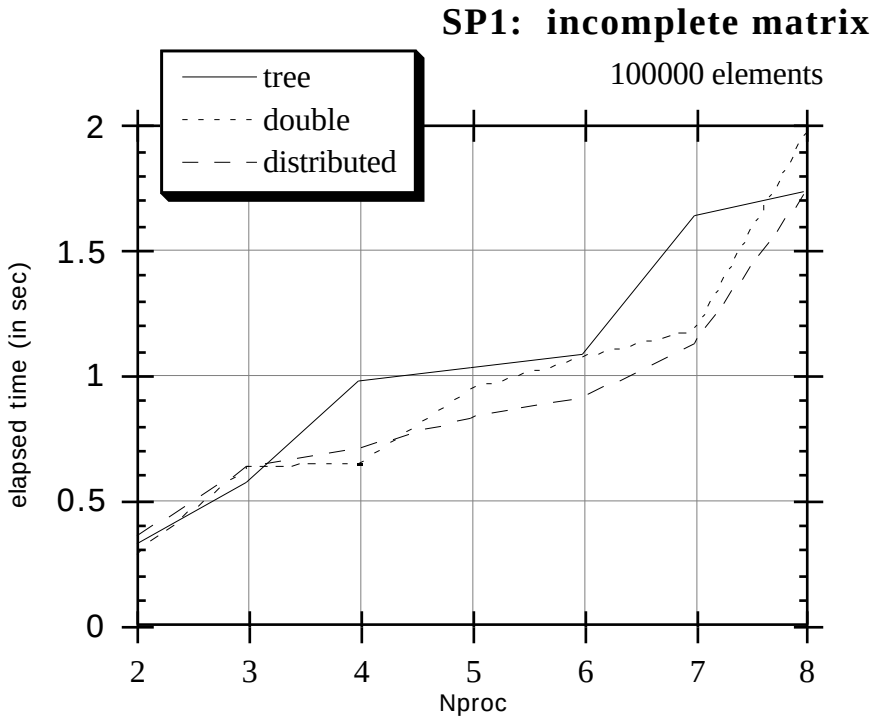


Figure 2.13. Comparison of algorithms for combining on an 8-node IBM SP1.

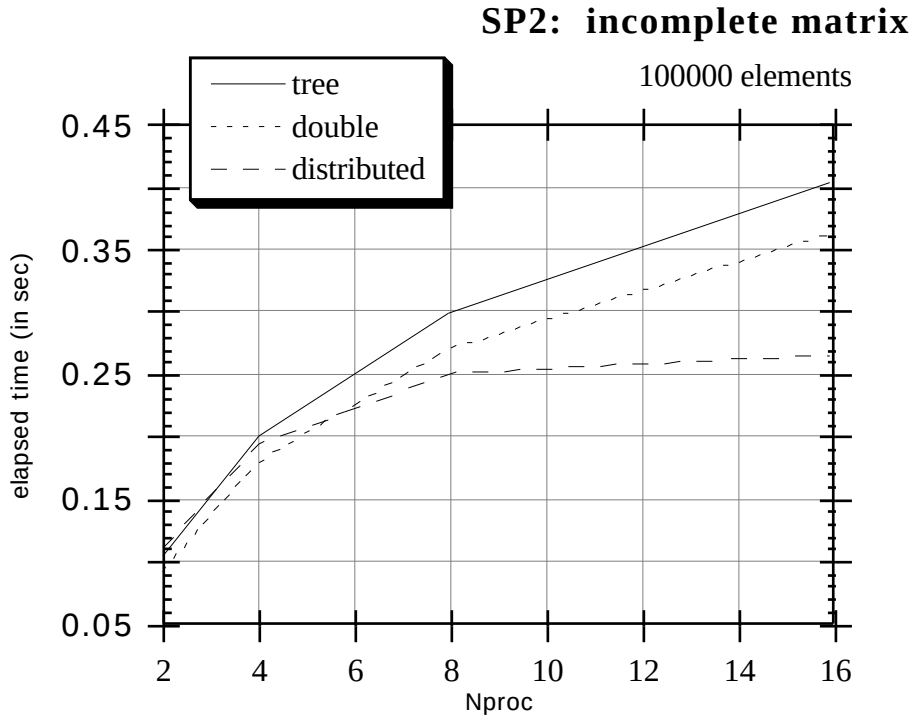


Figure 2.14. Comparison of algorithms for combining on a 16-node IBM SP2.

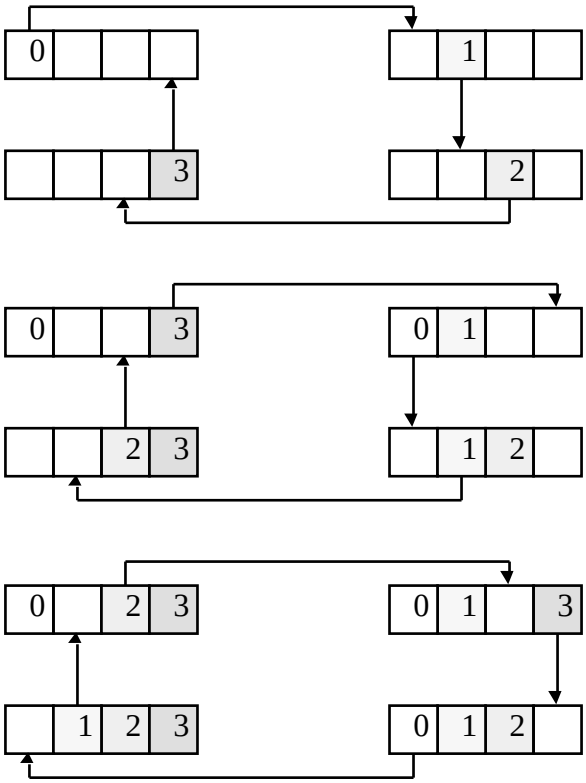


Figure 2.15. The ring algorithm for gathering distributed matrices.

Gathering Distributed Matrices

The combining of the distributed matrices is done by the routine PPGDV. It has the choice between three different algorithms. The first algorithm, the "wild" algorithm is rather trivial. Each node just sends its part of the matrix to the other nodes. The second algorithm is the so-called ring algorithm. In Figure 2.15 this algorithm is shown. The nodes of the parallel machine are seen as a ring. In each step a node sends to its right node the part of the matrix that the receiving node does not contain and receives a new part of the matrix from its left node. This process is repeated until the whole matrix is on all the nodes. The last algorithm is the binary tree algorithm, see Figure 2.11a. The difference with binary tree of the incomplete matrices, is that no summation is needed and the indices have to be sent.

We have measured on the different platforms the time for gathering distributed matrices of 1000 and 100000 elements. The results for gathering the largest matrix are shown in Figures 2.16, 2.17 and 2.18. From these Figures we can immediately conclude that the ring algorithm is the fastest on all platforms. The same result was obtained for smaller matrices.

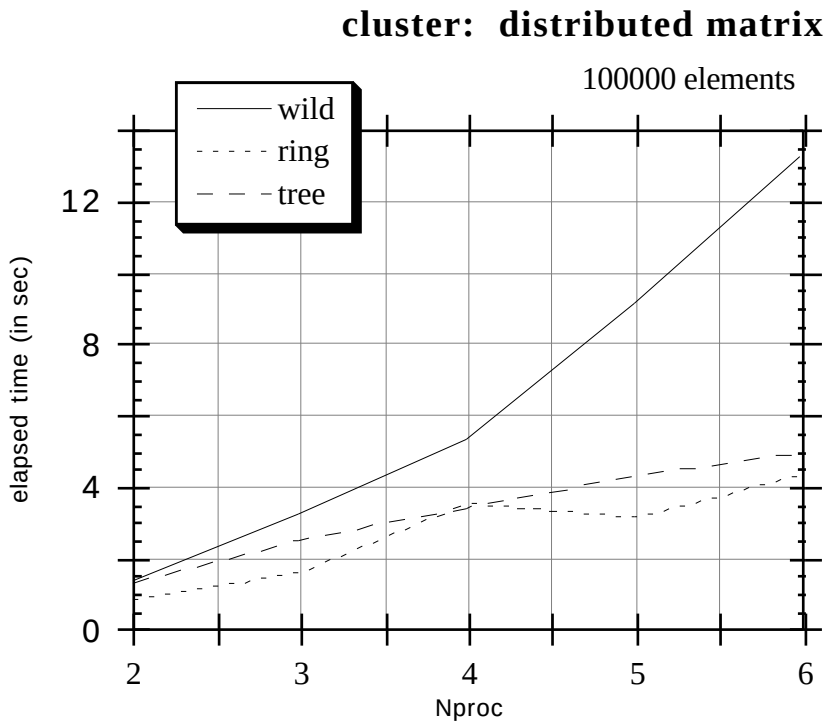


Figure 2.16. Comparison of algorithms for gathering on a workstation cluster.

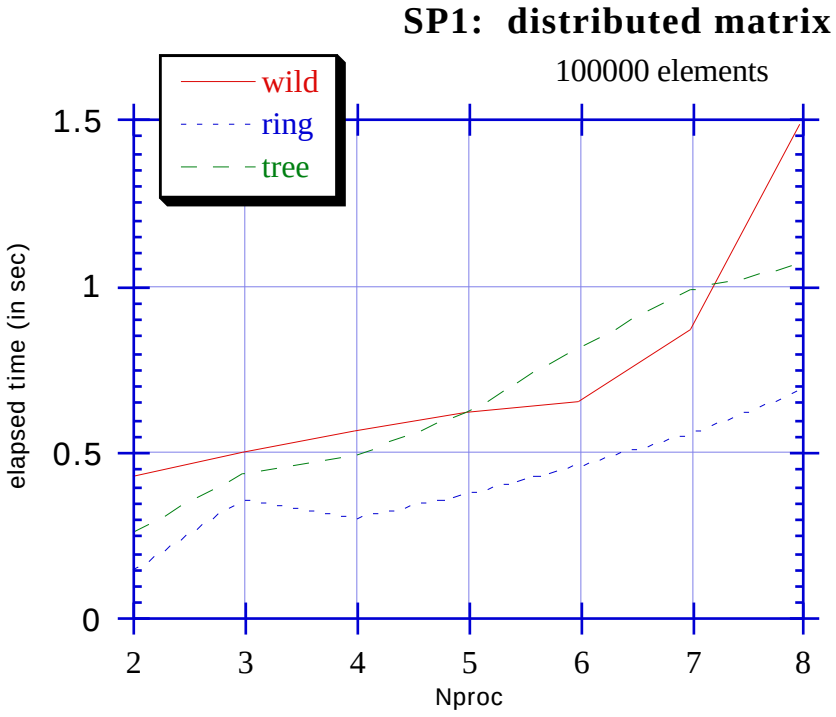


Figure 2.17. Comparison of algorithms for gathering on an 8-node IBM SP1.

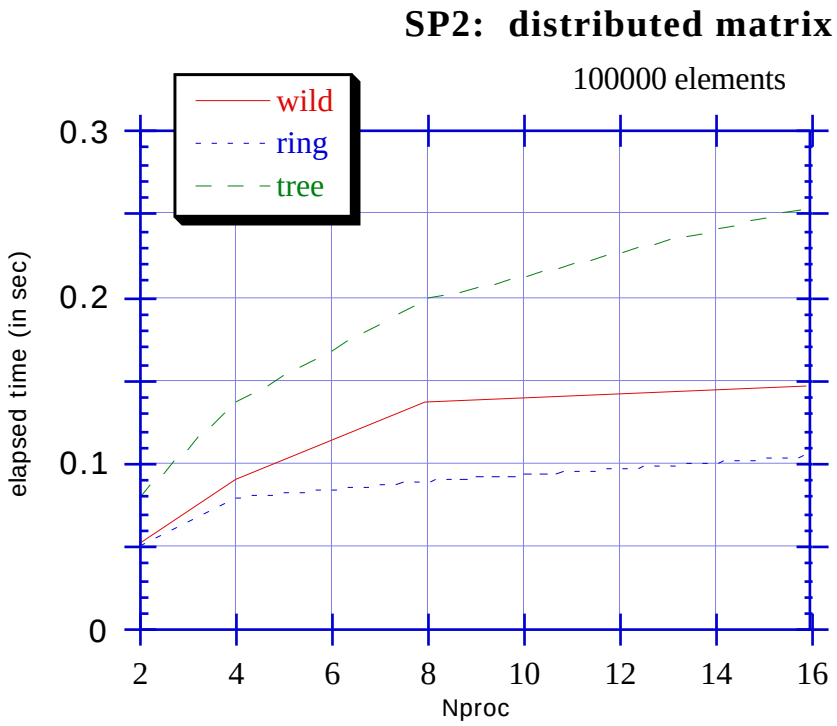


Figure 2.18. Comparison of algorithms for gathering on a 16-node IBM SP2.

2.7 Parallel Performance

In this section, we describe first the molecules and the parallel platforms we have used for the benchmarking of the parallel ADF code. Then the timing definitions used are explained. After this, the results per molecule on the different parallel platforms are given.

Benchmark Molecules

To benchmark the parallel ADF code we have chosen molecules that will illustrate the different aspects of ADF: a single-point calculation for a large molecule ($Pt(P(Ph)_3)_3CO$), a single-point calculation with gradient corrections for a medium-sized molecule ($Cu(C_7H_6O_2N)_2$), a geometry optimization with non-local corrections was done for a medium-sized molecule ($Fe_2(CO)_9$), and finally a single point calculation including gradient corrections for $Pt(P(Ph)_3)_3CO$. The basis set and frozen cores used, are the same as in section 2.4. All benchmark runs were performed using the disk-based algorithm.

Benchmark Platforms

The calculations have been done on different hardware platforms: a workstation cluster of 6 IBM RS6000/250 workstations connected by ethernet, an 8-node IBM SP1 using an FDDI network for communication, a 28-node IBM SP2 at CSSR in Italy (indicated in graphs as SP2) and a 512-node IBM SP2 at Cornell, U.S.A. (indicated in graphs as SP2*). Both SP2 platforms use a high-performance switch for communication. Beside these distributed memory machines we have also used a platform consisting of 4 SGI Power Challenge machines (EPCA, indicated in Figures and Tables as SGI*). For the communication between the machines either HIPPI or ethernet was used, while between the CPU's in the same machine messages are passed through shared memory. SGI (Boston, U.S.A.) performed some of the benchmark calculations on their SGI Power Challenge platform (indicated in the graphs as SGI) after some minor revisions in the ring algorithm (see section 2.6.5).

We also tried to use the Parsytec PowerPC601. However, on this machine the hardware I/O configuration leads to an extremely low I/O performance making the performance of ADF unacceptable (10 to 20 times more elapsed time needed than on other machines with similar processors). Running a large calculation on this machine is impossible.

2.7.1 Timing Definitions

For the presentation of our timing results we have used the following definition for the speed-up: $S = T_S / T_P$, where T_S and T_P are the elapsed times of the whole program executed on a single node respectively a number of nodes. In the elapsed time the startup time of all children is also included, because that is the real time that a user has to wait for his/her job to be finished.

To find out if communication, load imbalances, I/O or page faults influence the measured speed-up we compare our speed-up results with the speed-ups from Amdahl's law.[7]

$$T_p = t_s + \frac{t_p}{n} \qquad (8.1.1)$$

where t_s and t_p are the elapsed time of the serial respectively parallel part of the program executed on a single processor. There are several reasons for an increase of the deviation between the curve of the measured speed-up and the speed-up obtained from Amdahl's law. It can be

caused by an increase of the communication time between the nodes, by load imbalances or by shared disk access. To obtain the times t_s and t_p the program prints the percentage of time that it runs in the parallel parts, when executed on a single node. When we could not run on a single node, this percentage was obtained from a calculation on a small number of nodes, assuming Amdahl's law to be exact for that number of nodes. Furthermore, for the timing of some individual routines we have used PPBARR to synchronize all tasks before starting timers for that routine. This prevents imbalances from previous routines to be measured as part of the routine measured. The speed-up shown for the individual routines have all been measured on the 8-node IBM SP1 using FDDI.

In some of the graphs where two speed-ups are shown, one of the speed-ups is raised by one otherwise only one line would be visible. In the graphs with three speed-ups one of them is raised by two.

2.7.2 Single Point: $Pt(P(Ph)_3)_3CO$

In this section the results for the parallelization of a single point calculation are shown. The molecule considered here is $Pt(P(Ph)_3)_3CO$. As mentioned before this molecule cannot be run on a single node, therefore the speed-up at 2 nodes is set to 2. On the 512-node IBM SP2 at Cornell it was only possible to run it on 4 nodes, so the speed-up is set to 4 at 4 nodes.

Figure 2.20 shows the routine FOCKY that calculates the Fock matrix by numerical integration. As expected, this routine scales exactly with the number of nodes because of the perfect load balancing that is achieved with the numerical integration. This can be seen in Figure 2.19 where the CPU times for the fastest and slowest node are shown.

Table 2.4. Elapsed time of routines on 8 nodes of the IBM SP1

Routine	Speedup	Elapsed time (in sec)
FOCKY	7.9	9918
FITINT	8.0	555
CONPMT	4.8	8
RHOFIH	8.7	494

The speed-ups for the routines that are concerned with the density fit are shown in Figure 2.21 (see also Table 2.4). We see that the routine FITINT which imposes the load balancing for the atom pairs, scales exactly with the number of processors. This is due to the fact that for this large molecule without symmetry there are enough atom pairs to distribute over the nodes and that FITINT does not communicate. Therefore, none of the nodes will be waiting for the others to finish. The routine RHOFIH scales satisfactorily, but CONPMT does not scale so well. The speed-up of the routine CONPMT is not so good due to 10% of serial time left for the 2-node run. However, this is not so important because it is less than 1‰ of the overall elapsed time. The graph shows furthermore that the speed-up of RHOFIH increases dramatically at 7 nodes to 8.6 and that of CONPMT peaks at 7 nodes. This behavior is caused by an enormous decrease of the number of page faults when going form 6 to 7 nodes.

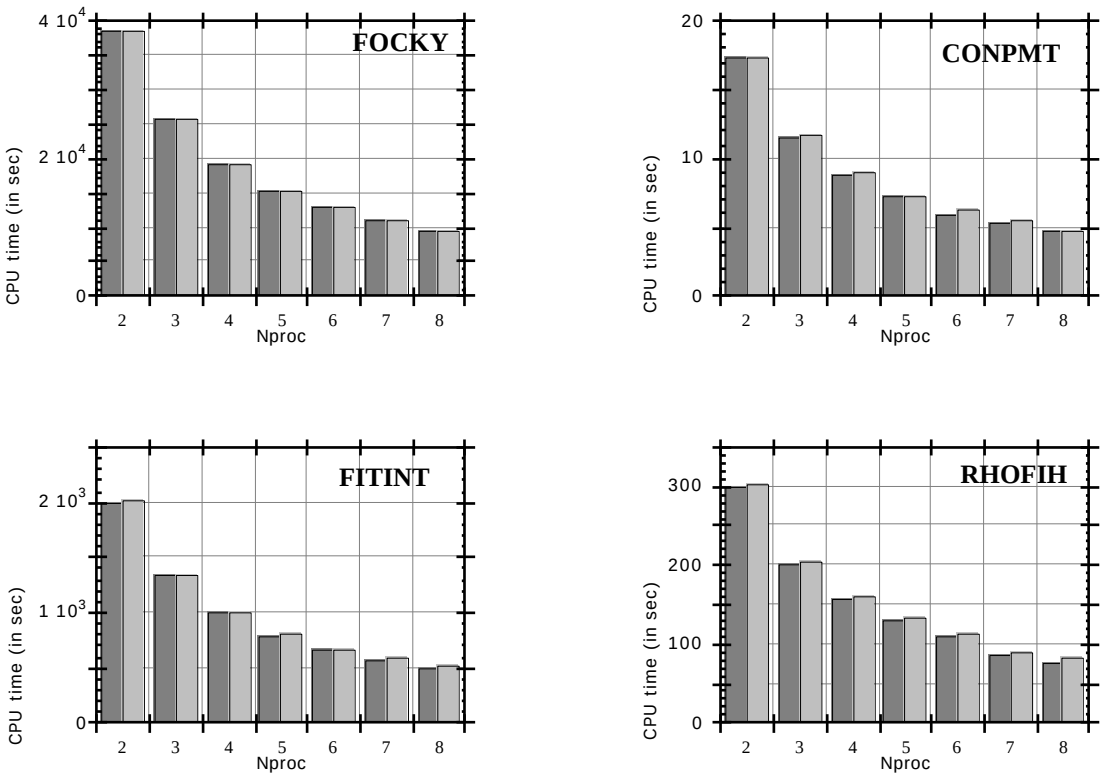


Figure 2.19. CPU times of the least and most loaded nodes for the routines FOCKY, FITINT, RHOFIH and CONPMT.

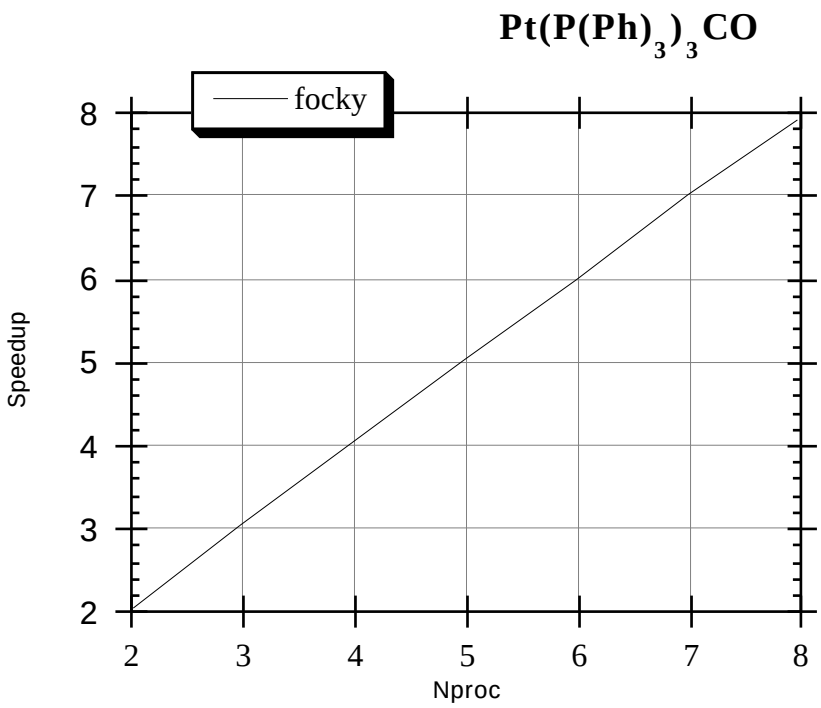


Figure 2.20. Speedup of the routine FOCKY on the IBM SP1.

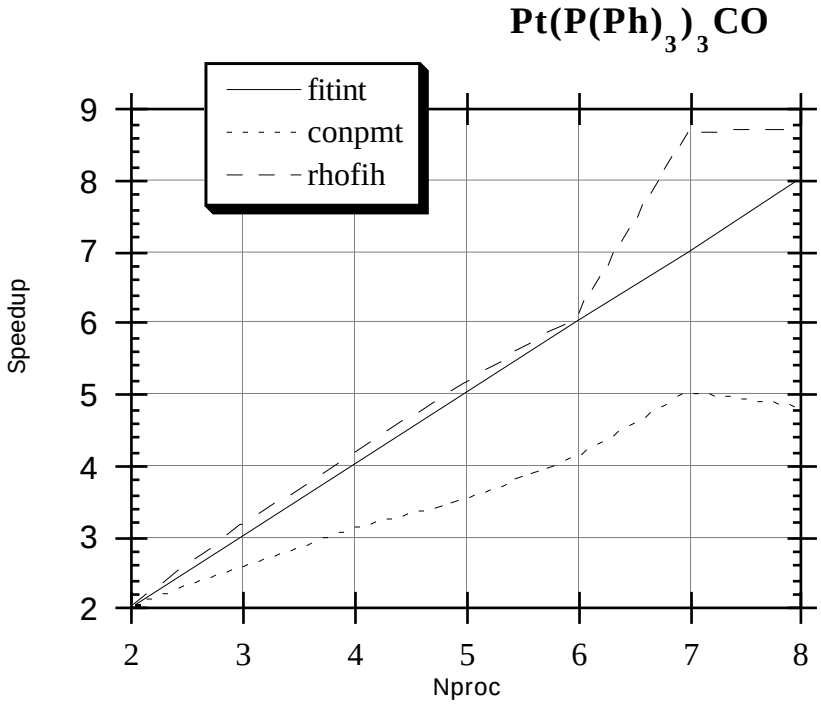


Figure 2.21. Speedup of the routines FITINT, CONPMT and RHOFIH on the IBM SP1.

To examine the load balancing for the routines that deal with atom pairs the CPU times for the fastest and slowest node are shown in Figure 2.19. We can see that the chosen load balancing works very well for FITINT and RHOFIH but also for CONPMT it seems to be nearly perfect.

Results on the Benchmark Platforms

The last three Figures of this section show the speed-ups for the whole program on different platforms and the speed-up predicted by Amdahl's law (see also Table 2.5). On a single node 99.4% of the elapsed time is spent in the parallel parts. We see that the speed-up of the whole program for the cluster and the IBM SP1 does not deviate from the speed-up predicted by Amdahl's law (Figure 2.22). For the number of nodes used there is no real communication problem. The ethernet connection that is used for the cluster is quite acceptable for 6 workstations.

The results on the 512-node IBM SP2 are shown in Figure 2.23. Comparing the SP2 and Amdahl we see a deviation starting at 64 nodes. From our timing results we found that this was due to the routine PPINIT, which starts all children on the different nodes. As Amdahl does not take the starting time into account we have to compare the speed-up of the program without PPINIT with the theoretical speed-up to investigate the parallel performance of the program. The speed-up without PPINIT indeed follows Amdahl closely with only an (unexplained) deviation after 120 nodes. We see no load balancing problems nor communication problems. However, for the end-users of the program the "true" speed-up is the only interesting information.

In Figure 2.24 the scaling of the program on the distributed memory machine, the IBM SP2 (Cornell), is compared with the scaling on a shared memory machine, the SGI Power Challenge (Boston). One of the SGI results shows the speed-up using only one machine, and the other two show the results using two or three machines connected by HIPPI. The results on one machine have only been obtained up to 8 CPU's. It is almost not visible because the graph coincides with the other two SGI graphs. All three SGI Power Challenge results show that the scaling of the execution time with the number of processors is not so good compared to the scaling of the SP2. The reason for this scaling behavior is probably the shared access of the CPU's to the disks.

A CPU of an SGI Power Challenge is roughly twice as fast as a CPU in one node of the IBM SP2. In Table 2.5 we can see that the run on 64 nodes of the IBM SP2 takes almost the same amount of time as the run on 2 SGI Power Challenge machines with each 16 CPU's.

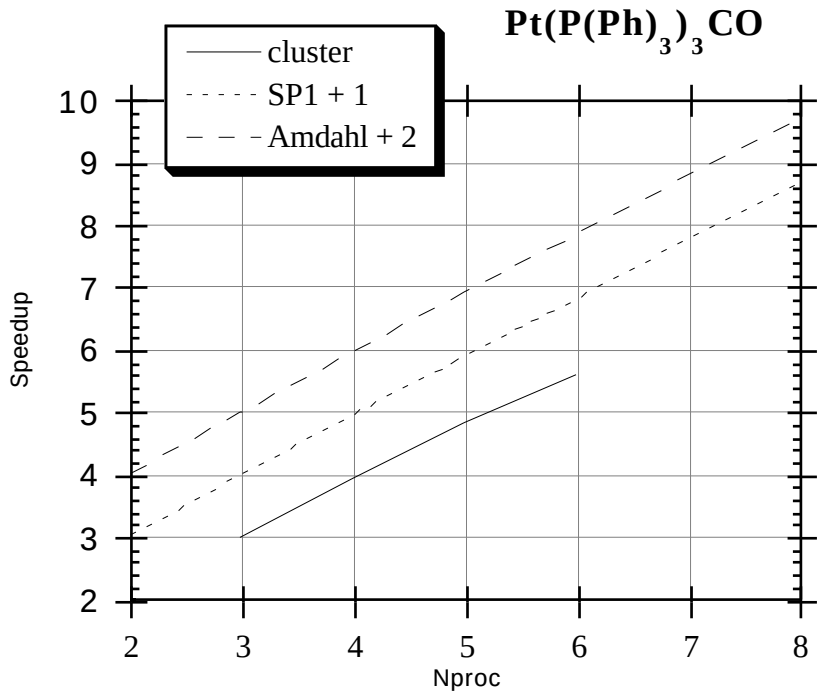


Figure 2.22. The speed-up of ADF on the cluster and the SP1 (raised by one) and the speed-up by Amdahl's law (raised by two).

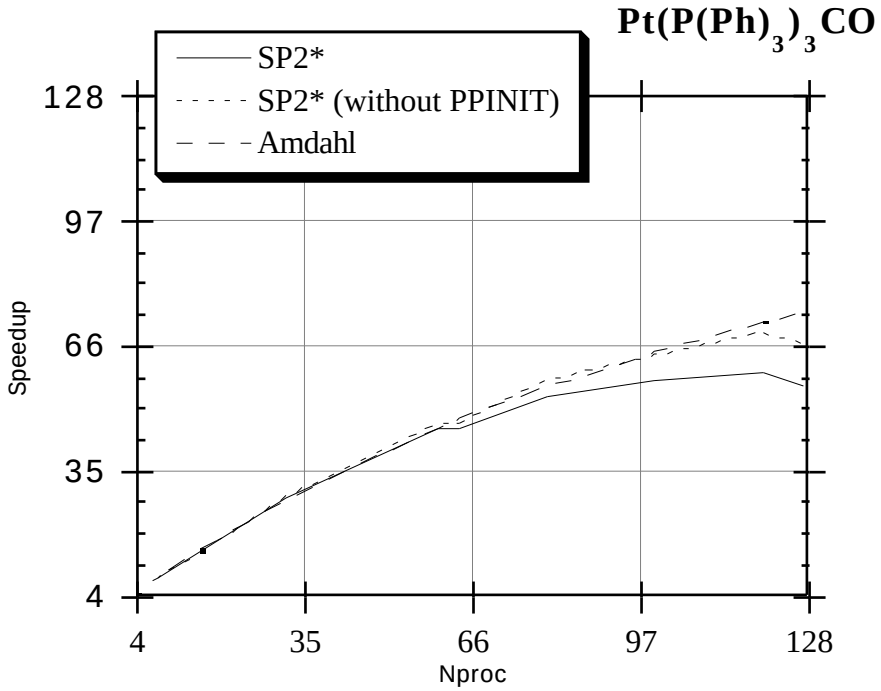


Figure 2.23. The speed-up of ADF on the IBM SP2 at Cornell and the speed-up by Amdahl's law.

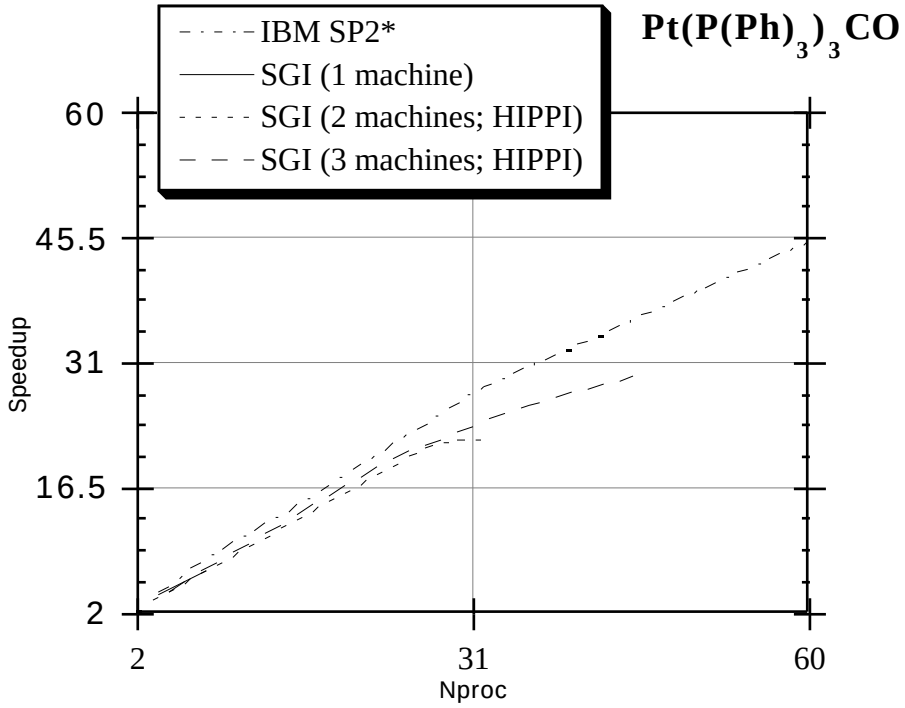


Figure 2.24. The speed-up of ADF on the SGI Power Challenge (Boston) and the IBM SP2 (Cornell).

Table 2.5. Elapsed time of ADF program for $Pt(P(Ph)_3)_3CO$

Parallel platform	Network	Nproc	Speedup	Elapsed time (in sec)
cluster of IBM RS6000/250	ethernet	6	5.6	48891
IBM SP1	FDDI	8	7.6	16545
IBM SP2 (Cornell)	switch	8	8.0	12310
		32	27.9	3546
		64	44.8	2209
SGI Power Challenge (1 machine)		8	6.5	7677
SGI Power Challenge (2 machines)	HIPPI	(16+16)	21.8	2224
SGI Power Challenge (3 machines)	HIPPI	(8+8+8)	19.5	2486

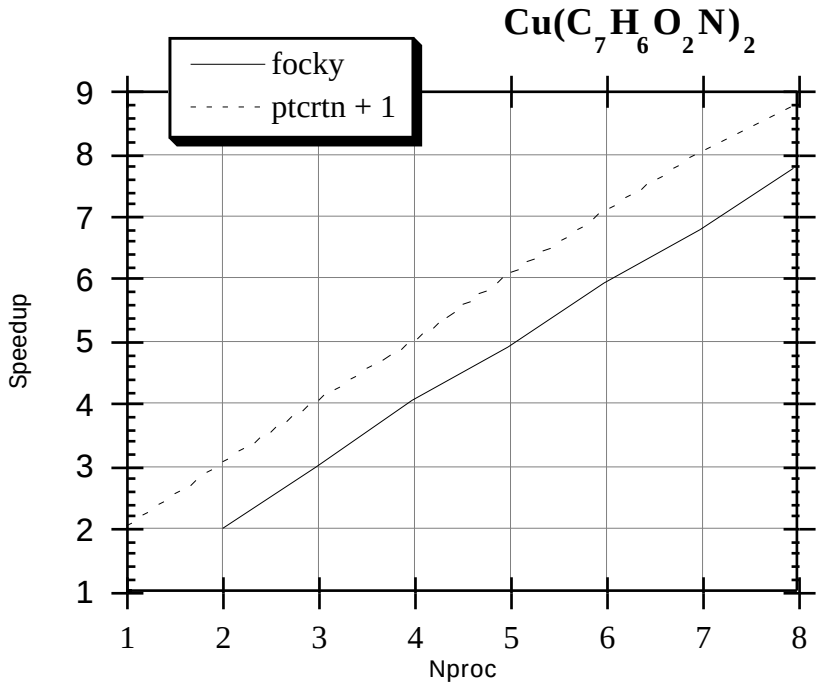


Figure 2.25. Speedup of the routines FOCKY and PTCRTN (raised by one) on the IBM SP1.

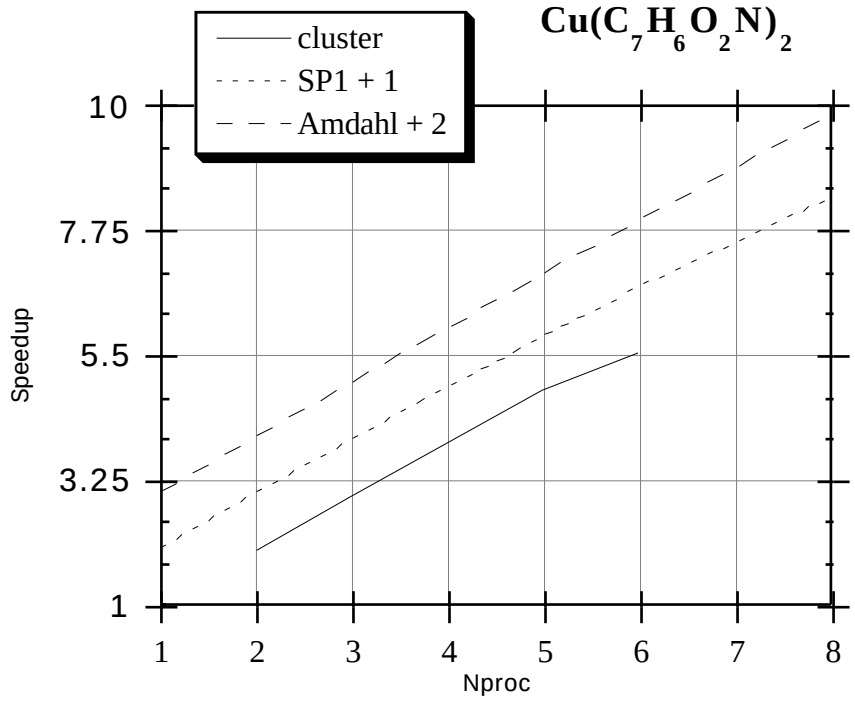


Figure 2.26. The speed-up of ADF on the cluster and the SP1 (raised by one) and the speed-up by Amdahl's law (raised by two).

2.7.3 Gradient Corrections: $Cu(C_7H_6O_2N)_2$

The results for the single point calculation with gradient corrections during the SCF do not differ much from the previous results. The overall elapsed time is again dominated by the numerical integration. Figure 2.25 shows the speed-up for the routines FOCKY and PTCRTN. The last routine calculates the gradient corrections to the exchange-correlation potential in the integration points. The difference between the two routines is that FOCKY has to combine the Fock matrix at the end while PTCRTN does not communicate at all. As expected both routines scale exactly with the number of nodes (see Table 2.6). Further we see that for the number of nodes used, the communication in FOCKY is not visible yet.

Table 2.6. Elapsed time of routines on 8 nodes of the IBM SP1

Routine	Speedup	Elapsed time (in sec)
FOCKY	7.8	314
PTCRTN	7.8	773

Results on the Benchmark Platforms

In Figure 2.26 the speed-up of the whole program on workstation cluster and the IBM SP1 is shown (see also Table 2.7). The parallel part of the program is in this case for a single node run equal to 99.6% of the elapsed time. The speed-up for the workstation cluster is good, 5.5 for the 6-node run. For both machines we see that the speed-up deviates somewhat from Amdahl's law.

In Figure 2.27 the results on the 512-node IBM SP2 are shown. The program scales satisfactorily until 32 processors. Increasing the number of processors by a factor of two after 32 has almost no effect on the speed-up. It even falls down going from 64 to 128 nodes. This poor scaling behavior at the larger number of nodes is caused by an unfavorable ratio of computation and communication. The computation time becomes relatively too small for the communication needed and therefore some parallel parts of the program become more expensive than in the serial run.

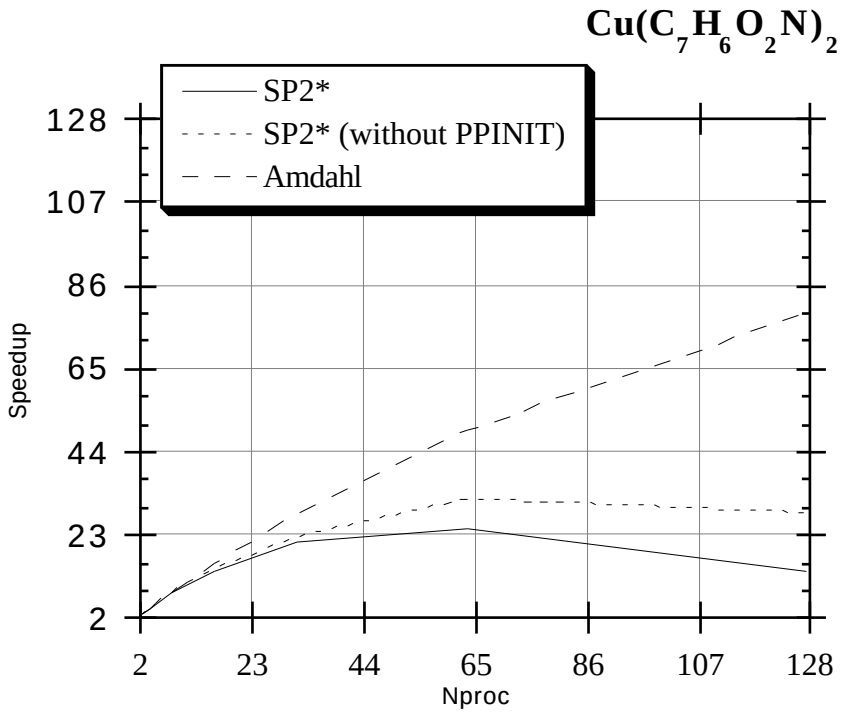


Figure 2.27. The speed-up of ADF on the IBM SP2 at Cornell and the speed-up by Amdahl's law.

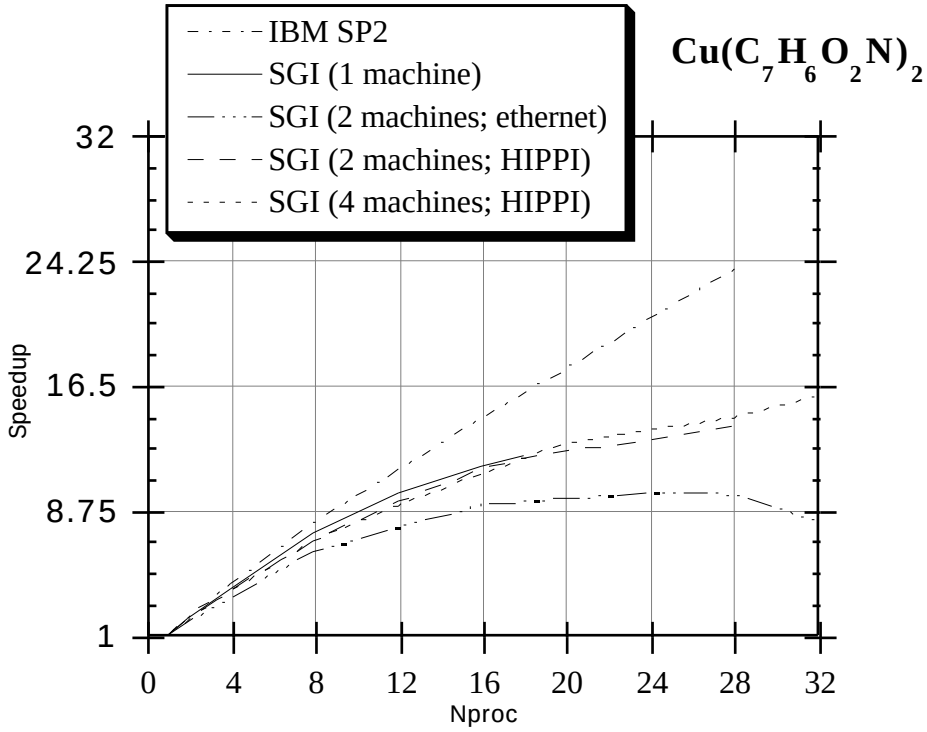


Figure 2.28. The speed-up of ADF on the SGI Power Challenge (Boston) and the IBM SP2 (CSSR, Italy).

For this molecule the results of the SGI Power Challenge (Boston) were also compared with the IBM SP2 (CSSR, Italy) results. From Figure 2.28 we can conclude again that the SGI does not scale as well as the IBM SP2. This can probably be ascribed to the difference in I/O performance between the machines. The graph (Figure 2.28) for the ethernet connection clearly shows that ethernet is too slow. We can see also that it does not make much difference if we use the HIPPI connection or shared memory.

Table 2.7. Elapsed time of ADF program for $Cu(C_7H_6O_2N)_2$

Parallel platform	Network	Nproc	Speedup	Elapsed time (in sec)
cluster of IBM RS6000/250	ethernet	6	5.5	3786
IBM SP1	FDDI	8	7.3	1404
IBM SP2 (at CSSR)	switch	8	7.9	922
		16	14.4	503
IBM SP2 (at Cornell)	switch	8	7.7	912
SGI Power Challenge (1 machine)		8	7.2	536
SGI Power Challenge (2 machines)	ethernet	(4+4)	6.2	634
SGI Power Challenge (2 machines)	HIPPI	(4+4)	6.8	572
SGI Power Challenge (4 machines)	HIPPI	(2+2+2+2)	6.8	574

2.7.4 Geometry Optimization: $Fe_2(CO)_9$

In this section we discuss the speed-up for geometry optimizations with gradient corrections. The molecule considered is $Fe_2(CO)_9$. In Figure 2.29 the speed-ups of the two important routines for the geometry optimization (FOCKC and ENGRAD) are given. They scale both exactly with the number of processors (see Table 2.8). The load balancing for the numerical integration is perfect and there is no real communication problem.

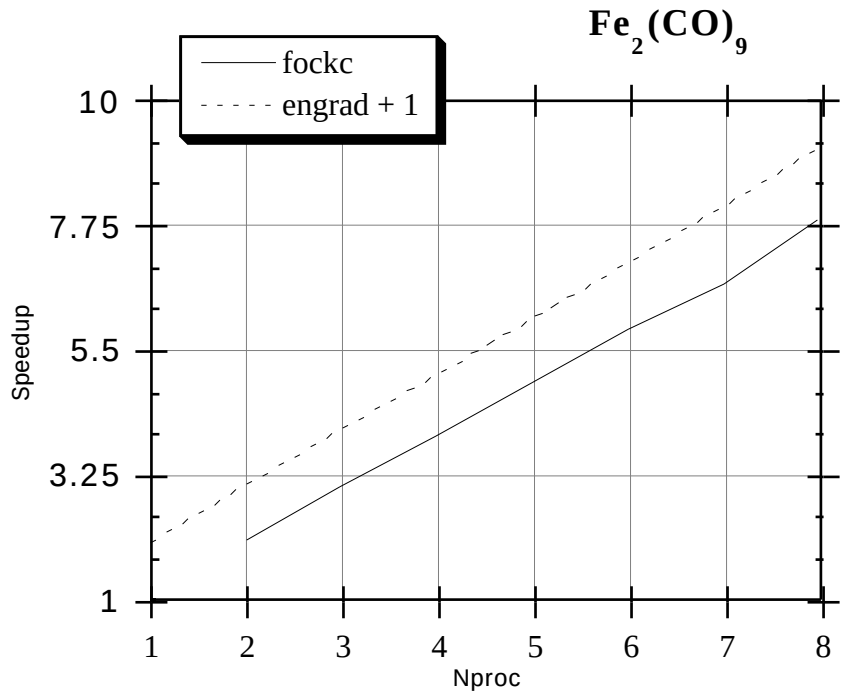


Figure 2.29. Speedup of the routines FOCKC and ENGRAD (raised by one) on the IBM SP1.

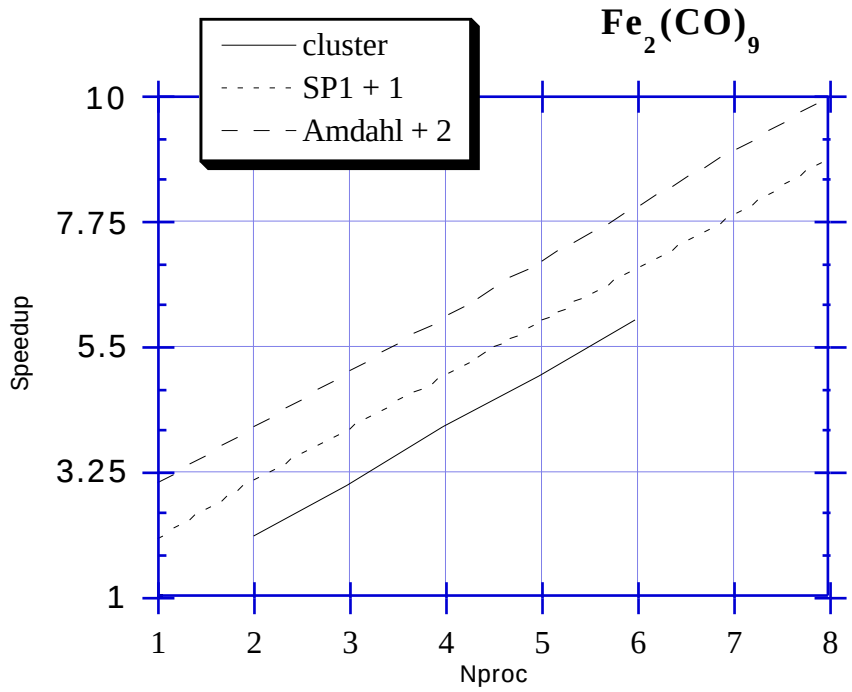


Figure 2.30. The speed-up of ADF on the cluster and the SP1 (raised by one) and the speed-up by Amdahl's law (raised by two).

Table 2.8. Elapsed time of routines on 8 nodes of the IBM SP1

Routine	Speedup	Elapsed time (in sec)
FOCKC	7.9	180
ENGRAD	8.1	1241

Results on the Benchmark Platforms

The speed-up of the whole program is shown in Figure 2.30 for the workstation cluster and the IBM SP1 (see also Table 2.9). The parallel part is for a single node run equal to 99.9% of the elapsed time. The measured speed-up on the cluster and on the SP1 follow the speed-up of Amdahl's law extremely well, so their is no load imbalance and no communication problem at all on both platforms. Thus with a small workstation cluster we are able to reduce an optimization of one week to one day.

$Fe_2(CO)_9$ is a rather small molecule compared to the two previous molecules, and also a molecule with a much higher symmetry. There could be a problem with the load balancing of the routines dealing with the density fit. In fact a very small load imbalance starts to show up for FITINT at the 8 node run. This is caused by the iron-iron pair, which has much more integrals than the other atom pairs. The imbalance for RHOFIH and CONPMT is almost negligible. These three routines only take .7% of the overall elapsed time. So, this small load imbalance will not easily cause a decrease of the speed-up.

In Figure 2.31 the speed-up is shown for both SP2 machines. As in the case of the previous molecule the speed-up starts to deviate from Amdahl's law for larger number of nodes. This is also caused by communication. Furthermore, we can see that the 28-node SP2 in Italy has a better scaling behavior than the 28-node partition of the SP2 at Cornell (SP2*). The difference in the scaling behavior might be caused by the different file sharing systems of the machines. On the 28-node SP2 in Italy the files were NFS mounted and on the Cornell machine they were AFS mounted. The results of this calculation on the SGI Power Challenge configuration, EPCA, are shown in Figure 2.32. We see again a better scaling behavior on the IBM SP2. The difference between speed-ups using one machine or two machines is caused by the ethernet connecting the two machines. This communication network is too slow for the number of processors used.

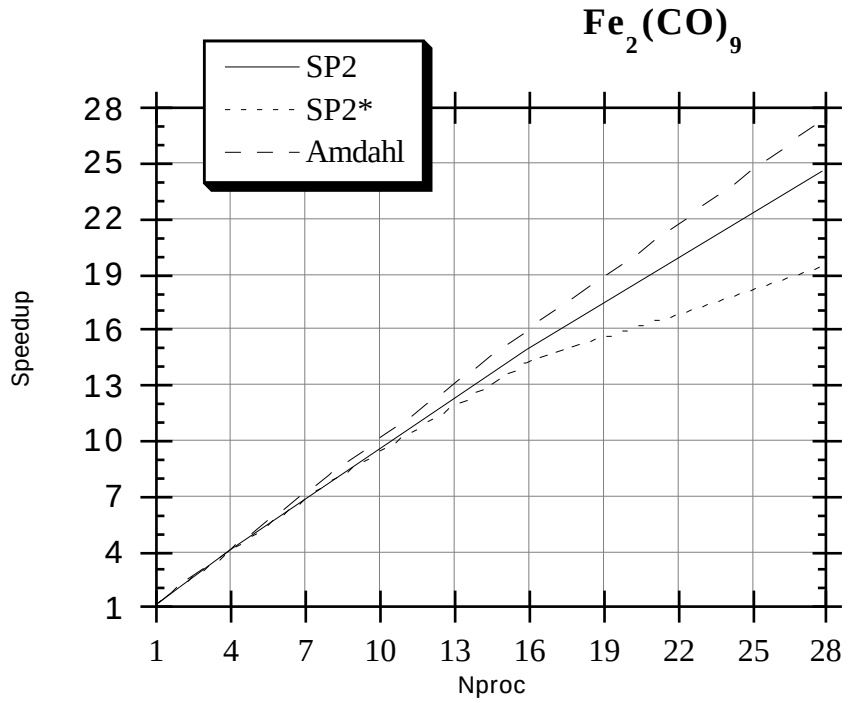


Figure 2.31. The speed-up of ADF on the SP2 at CSSR in Italy and the SP2* at Cornell (U.S.A.) and the speed-up by Amdahl's law.

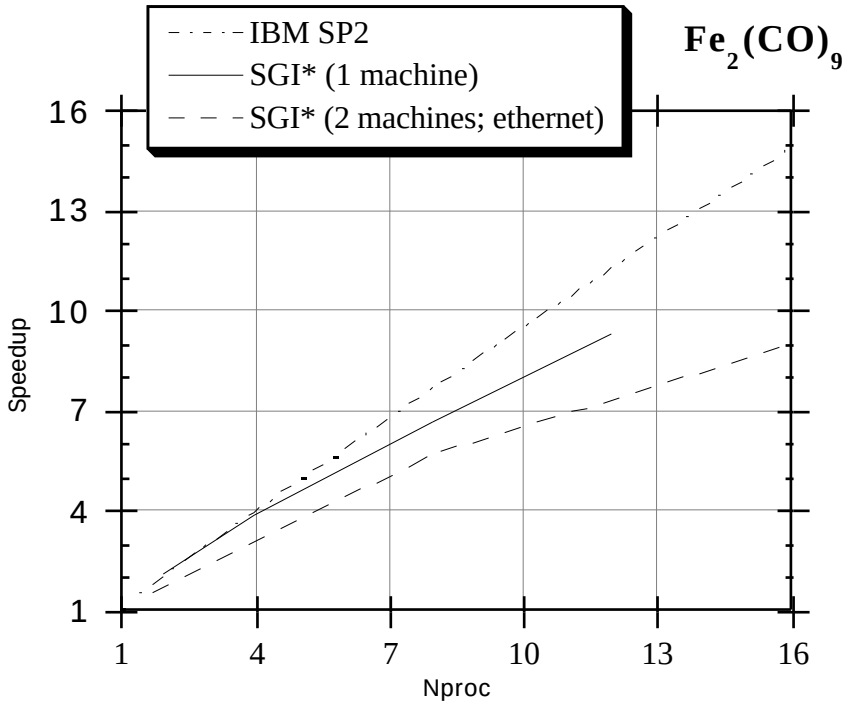


Figure 2.32. The speed-up of ADF on the SGI Power Challenge (EPCA) and the IBM SP2 (CSSR).

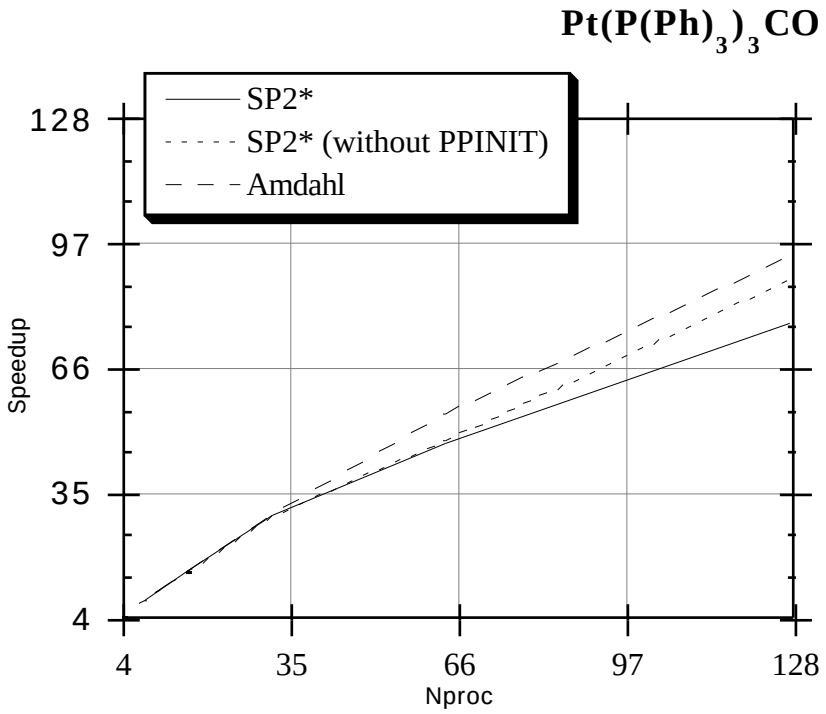


Figure 2.33. The speed-up of ADF on the IBM SP2 at Cornell and by Amdahl's law.

Table 2.9. Elapsed time of ADF program for $Fe_2(CO)_9$

Parallel platform	Network	Nproc	Speedup	Elapsed time (in sec)
cluster of IBM RS6000/250	ethernet	6	5.9	29227
IBM SP1	FDDI	8	7.8	10483
IBM SP2 (at CSSR)	switch	8	7.6	7368
		16	14.8	3798
IBM SP2 (at Cornell)	switch	8	7.6	7225
SGI* Power Challenge (1 machine)		8	6.6	4946
SGI* Power Challenge (2 machines)	ethernet	(4+4)	5.6	5777

2.7.5 Gradient Corrections: $Pt(P(Ph)_3)_3CO$

The last calculation is an extremely large one: a single-point calculation with gradient corrections for $Pt(P(Ph)_3)_3CO$. It takes almost 13 hours on 4 nodes of the IBM SP2 and 40 minutes on 128 nodes. Figure 2.33 shows that the program scales very well up to 128 nodes. Theoretically this calculation would take more than two days on a single node. It is a great achievement that the result can now be obtained in 40 minutes.

Table 2.9. Elapsed time of ADF program for $Pt(P(Ph)_3)_3CO$ (with gradient corrections)

Parallel platform	Network	Nproc	Speedup	Elapsed time (in sec)
SP2 (at Cornell)	switch	128	77	2390
SP2 without PPINIT (at Cornell)	switch	128	88	2091

2.8 Conclusions

From the Figures that show the speed-ups of the individual routines, we see that the elapsed time of the routines scales reasonably with the number of nodes. The routine, CONPMT, that does not scale so well with the number of nodes, is also much less time-consuming.

We have chosen in our parallelization strategy to keep most data on local disks. The disk space needed by the program scales inversely with the number of nodes. From Figure 2.21, where the speed-up of RHOFIH is shown, we saw that the partioning of the file with fit integrals over the nodes decreased the number of page faults dramatically.

Most of the matrices have not yet been distributed over the nodes. Therefore, the memory size needed will not scale with the number of nodes. Matrices, such as the Fock matrix, that are needed completely on all the nodes, can be distributed, but this will lead to an enormous increase of communication time.

The speed-ups of the two most intensive computational kernels, FOCKY and PTCRTN, show

that we can achieve a perfect scaling. It can be very misleading to present the scaling of the most time-consuming routines only. Therefore, we show normally the scaling behavior of the total ADF program. The total elapsed time includes the starting of ADF on the other nodes of the parallel machine.

The speed-ups of the overall program on the workstation cluster for the three different molecules, show that a cluster of 6 (or 8) workstations connected by ethernet can be used adequately. Only when more workstations are added to the cluster the communication might become a bottleneck.

On the 8-node IBM SP1 and the 28-node IBM SP2 we see that the speed-up almost follows Amdahl's law. The small deviation can be caused by load imbalance or communication. It is known that there can be a small load imbalance for the routines that deal with the density fit. However, these routines take a very small amount of time, and small deviations in the load balancing will hardly influence the speed-up. From all the speed-ups of the calculations shown here, we can conclude that our parallelization strategy turned out to be a good choice.

Comparing the SGI Power Challenge with the IBM SP2 we saw that the SGI machine does not scale as well as the SP2. This is probably due to shared access of the CPU's to the disks. The Parsytec PowerPC601 was unusable because of the bad I/O performance. For a good scaling of the disk-based calculations with the ADF program it is required that each node can do its own I/O.

The results on the 512-node IBM SP2 show that the number of nodes claimed should be conformable with the size of the calculation to use the ADF program effectively on a large parallel platform. So medium sized calculations should be run on moderate number of nodes.

The results obtained for the calculation of $Pt(P(Ph)_3)_3CO$ with gradient corrections demonstrate that the ADF program is able to use 128 nodes effectively for a system of this size. The speed-up of almost 80 for the whole program on 128 nodes shows that parallel computing can be a significant advance over serial computing.

With the present degree of parallelization almost 2 orders of magnitude can be gained by using a parallel implementation. This opens the perspective of routinely performing calculations on molecules of a size that have until now been unachievable.

References

[1] a) C. Fonseca Guerra, O. Visser, J. G. Snijders, G. te Velde, E. J. Baerends, in *Methods and Techniques for Computational Chemistry*, (Eds.: E. Clementi, G. Corongiu), STEF, Cagliari **1995**, p. 305-395

b) E. J. Baerends, D. E. Ellis, P. Ros, Chem. Phys. **1973**, 2, 41

c) E. J. Baerends, P. Ros, *Chem. Phys.* **1975**, *8*, 412

d) E. J. Baerends, P. Ros, *Int. J. Quantum. Chem. Symp.* **1978**, *12*, 169

e) W. Ravenek, in *Algorithms and Applications on Vector and Parallel Computers*, (Eds.: H. H. J. Riele, T. J. Dekker, H. A. van de Vorst), Elsevier, Amsterdam, **1987**

g) P. M. Boerrigter, G. te Velde, E. J. Baerends, *Int. J. Quantum Chem.* **1988**, *33*, 87

h) G. te Velde, E. J. Baerends, *J. Comp. Phys.* **1992**, *99*, 84;

i) J. G. Snijders, E. J. Baerends, P. Vernooijs, *At. Nucl. Data Tables* **1982**, *26*, 483

j) J. Krijn, E. J. Baerends, *Fit-Functions in the HFS-Method; Internal Report (in Dutch)*, Vrije Universiteit, Amsterdam, **1984**

k) L. Versluis, T. Ziegler, *J. Chem. Phys.* **1988**, *88*, 322

l) L. Fan, L. Versluis, T. Ziegler, E. J. Baerends, W. Ravenek, *Int. J. Quantum. Chem., Quantum. Chem. Symp.* **1988**, *S22*, 173

m) J. C. Slater, Quantum Theory of Molecules and Solids, Vol. 4, McGraw-Hill, New York, 1974

n) L. Fan, T. Ziegler, *J. Chem. Phys.* **1991**, *94*, 6057

o) T. Ziegler, *Chem. Rev.* **1991**, *91*, 651

p) L. Fan, T. Ziegler, J. Chem. Phys. **1991**, *95*, 7401

[2] a) P. Hohenberg, W. Kohn, Phys. Rev. **1964**, *136B*, 864

b) W. Kohn, L.J. Sham, Phys. Rev. **1965**, *140A*, 1133

[3] a) A. Becke, *Phys. Rev. A* **1988**, *38*, 3098

b) S. H. Vosko, L. Wilk, M. Nusair, *Can. J. Phys.* **1980**, *58*, 1200

c) J. P. Perdew, *Phys. Rev. B* **1986**, *33*, 8822; Erratum: *Phys. Rev. B* **1986**, *34*, 7406)

d) Y. Wang, J.P. Perdew, Phys.Rev. B **1991**, *44*, 13298

[4] a) P. Pulay, Chem. Phys. Lett. **1980**, *73*, 393

b) P. Pulay, J. Comp. Chem. **1982**, *3*, 556

c) T.P. Hamilton, P. Pulay, J. Chem. Phys. Lett. **1986**, *84*, 5728

[5] V. Sunderam, *Concurrency: Practice Experience* **2**(4), 1990.

PVM is available via electronic mail from netlib@ornl.gov.

[6] a) G. Fox, M. Johnson, G. Lyzenga, S. Otto, J. Salmon, D. Walker, *Solving Problems on Concurrent Processors*, volume I, Prentice Hall, **1988**

b) R.A. van de Geijn, LAPACK Working Note 29, University of Tennessee

[7] G.M. Amdahl, *Validity of the single processor approach to achieving large scale computing capabilities.* In AFIPS Conference Proceedings **1967**, p. 483-485

Chapter 3

Towards an Order-N DFT Method

One of the most important steps in a Kohn-Sham type DFT calculation is the construction of the matrix of the Kohn-Sham operator (the "Fock" matrix). It is desirable to develop an algorithm for this step that scales linearly with system size. We discuss attempts to achieve linear scaling for the calculation of the matrix elements of the exchange-correlation and Coulomb potentials within a particular implementation (the ADF code) of the KS method. In the ADF scheme the matrix elements are completely determined by a 3D numerical integration, the value of the potentials in each grid point being determined with the help of an auxiliary function representation of the electronic density. Nearly linear scaling for building the total Fock matrix is demonstrated for systems of intermediate size (in the order of 1000 atoms). For larger systems further development will be desirable for the treatment of the Coulomb potential.

3.1 Introduction

There is currently much interest in the development of $O(N)$ methods (linear scaling of computation time with system size) for both Hartree-Fock (HF) and Kohn-Sham (KS) calculations. In contrast to most other current schemes for KS calculations, the Amsterdam Density Functional code[1] (ADF) uses 3D numerical integration for the evaluation of the "Fock" matrix elements throughout (in spite of some historical and logical inconsistency, we denote the matrix of the KS one-electron operator simply as Fock matrix). The use of 3D numerical integration also for the elements of the electronic Coulomb potential is particular to the ADF code. It is the purpose of this contribution to discuss the way in which linear scaling can be achieved in setting up the Fock matrix in this method. We will first, in this introduction, briefly consider the scaling of various steps in the ADF method, comparing to alternative DFT methods. In order to achieve linear scaling we need to define a finite spatial region over which a basis function can be considered to be nonzero. A definition of the spatial cut-off for the basis functions that takes into account the effect of the cut-off on the matrix elements to be constructed, will be discussed in section 3.2. Section 3.3 treats the linearization of the density fitting, section 3.4 the linearization of the evaluation of the exchange-correlation and Coulomb potentials in the grid points, section 3.5 the building of the Fock matrix and sections 3.6 and 3.7 offer results and conclusions respectively.

The matrix elements of the exchange-correlation (XC) potential and the Coulomb potential are both evaluated by numerical integration, using P integration points.

$$G_{\mu\nu} = \left\langle \chi_\mu \middle| \hat{V}_{\mathrm{XC}} + \hat{V}_{\mathrm{Coul}} \middle| \chi_\nu \right\rangle \approx \sum_{k=1}^{P} \chi_\mu^*(\mathbf{r}_k)\left(V_{\mathrm{XC}}(\mathbf{r}_k) + V_{\mathrm{Coul}}(\mathbf{r}_k)\right)\chi_\nu(\mathbf{r}_k)w_k \qquad (3.1.1)$$

Using 3D numerical integration in this way has the advantage that it is possible to use almost any type of basis function, although there may be some restrictions arising from the required evaluation of $V_{\mathrm{Coul}}(\mathbf{r}_k)$ at the grid points (the Poisson solver may or may not use the basis functions). STO's have been chosen as a convenient and physically well motivated basis. Apart from enabling the one-electron wavefunctions and the total density to have the correct behavior in

the tail and at the nucleus, they are also efficient since their number can typically be a factor three lower than the number of Gaussians. Another choice that has been made are numerical atomic orbitals (NAO's, also called single-site-orbitals, SSO's).[2-6] With N basis functions, there is an NP step in the calculation for the evaluation of the basis functions at all points, assuming for the time being that no cut-offs are applied. The value of the XC potential at each point simply follows, with the current approximate functionals, from the density and its gradients at that point. The density can be calculated either through the density matrix, yielding a N^2P step, or in terms of the occupied orbitals, yielding a N_eNP step.[7] Here N_e is the number of electrons. Of course not every size parameter is equal, in particular one usually has for large molecules

$$P \gg N' > N > N_e \tag{3.1.2}$$

where N' is the number of auxiliary functions to be discussed below. However, each of these parameters scales linearly with system size, so the sum of their exponents can be used roughly as the scaling order and the density evaluation appears to be an $O(N^3)$ step. In fact, however, a set of atom-centred auxiliary functions (density fitting basis) was introduced,[1] which allowed to write the density as a linear expansion in N' functions,

$$\rho(\mathbf{r}) \rightarrow \tilde{\rho}(\mathbf{r}) = \sum_A \sum_i a_i f_i(\mathbf{r}) \tag{3.1.3}$$

making evaluation of the density and of V_{XC} an $N'P$, i.e. $O(N^2)$ step. The determination of the fit coefficients a_i was also implemented as an $O(N^2)$ step (for more about the fitting procedure, see section 3.3). The Coulomb potential at each point is calculated from the functions f^c, obtained as Coulomb integrals from the fitfunctions f,

$$f_i^c(\mathbf{r}_k) = \int \frac{f_i(\mathbf{r})}{|\mathbf{r} - \mathbf{r}_k|} d\mathbf{r} \tag{3.1.4}$$

$$V_{\mathrm{Coul}}(\mathbf{r}_k) \rightarrow \tilde{V}(\mathbf{r}_k) = \int \frac{\tilde{\rho}(\mathbf{r})}{|\mathbf{r} - \mathbf{r}_k|} d\mathbf{r} = \sum_A \sum_i a_i f_i^c(\mathbf{r}_k) \tag{3.1.5}$$

This makes the evaluation of the Coulomb potential at the grid points also an $N'P$ ($O(N^2)$) step, though with a much larger prefactor than the density and V_{XC} evaluation. Finally, setting up the G matrix according to eq. (3.1.1), when the values of the potentials at all points are available, as well

as the values of the basis functions at all points, is a N^2P step, i.e. $O(N^3)$, as is the subsequent diagonalization. The scalings of the computation steps in the original ADF implementation[1] are summarized below:

fit coeffecients determination:	$O(N^2)$
potentials in grid points:	$aN'P+bN'P$
functions in grid points:	NP
Fock matrix set-up:	N^2P
diagonalization:	$O(N^3)$

For many systems, in particular systems incorporating transition metal atoms (transition-metal complexes or clusters with a limited number of metal atoms) which were the prime application targets, the dominant steps were the fit coefficient determination and the Fock matrix set-up, so the scaling was typically between N^2 and N^3. This constituted a significant improvement over the then N^4 scaling of *ab initio* HF calculations.

We will address in this paper the reduction of the various steps in the Fock-matrix set-up to linear scaling. It is to be noted that currently a variety of methods is being explored to achieve linear scaling for the matrix elements of the Coulomb potential.[8-21] There are other steps in the calculation which we do not discuss in this paper, but which have already received considerable attention elsewhere, such as the scaling of the point generation itself[22] and, of course, the diagonalization, see ref..[23-32]

3.2 Definition of Cut-offs for Matrix Elements

It is natural to define for each function a radius outside which it is negligible, and to determine the cut-off radius of each atom as the maximum of the radii of its functions. In the literature[7,22] the function radius λ_μ has usually been determined from the function values by the condition $|\chi_\mu(r)|<\tau$ for $|r|>\lambda_\mu$. This condition does not take into account that cutting at a given function value

is not quite the same for weakly decaying functions (small exponent) as for sharply decaying functions. A (much) larger contribution to for instance the normalization integral would be neglected in this way for a (very) weakly decaying function. In order to have approximately the same relative error in matrix elements of both diffuse and contracted functions we determine the function radius according to a somewhat different algorithm, without of course any change in the underlying idea.

The relative weight of the radial part of a basis or fit function beyond a certain cut-off radius r_0 is calculated as the ratio between the integral of the tail of the function beyond the cut-off point r_0 and the total radial integral,

$$w(\chi;r_0) = \frac{\int_{r_0}^{\infty} r^2\chi(r)dr}{\int_{0}^{\infty} r^2\chi(r)dr} = \frac{\int_{r_0}^{\infty} r^{n+1}e^{-\alpha r}dr}{(n+1)!/\alpha^{n+2}} \qquad (3.2.1)$$

Here n is the main quantum number, i.e. the radial part of χ is $r^{n-1}e^{-\alpha r}$. Since we wish to have a very efficient algorithm, in view of the frequent tests on these weights, we avoid the time-consuming integration over the partial region (r_0,∞) in the numerator by using a simple exponential function to approximate the tail of the function $r^2\chi$,

$$g(r) = Ce^{-\beta r} \qquad (3.2.2)$$

The parameters β and C are determined by

$$g(r_0) = r_0^2\chi(r_0) \qquad \text{and} \qquad g'(r_0) = \frac{d}{dr}r^2\chi\Big|_{r=r_0} \qquad (3.2.3)$$

At large cut-off points the function g simulates the tail of $r^2\chi$ almost perfectly. Equations (3.1.1) and (3.1.3) lead to the following expression of the relative weight of the tail

$$w(\chi;r_0) = \frac{\alpha^{n+2}r_0^{n+1}e^{-\alpha r_0}}{(n+1)!\left(\alpha-(n+1)/r_0\right)} \qquad (3.2.4)$$

If we would have taken in eq. (3.2.1) the normalization integral rather than the direct radial

integral over (the radial part of) the function, approximately the square of the present expression would have resulted (apart from a factor of the order of magnitude 1) for the contribution of the region beyond $r=r_0$ to the normalization integral. Therefore the typical choice for the weight of 0.1% ($w=10^{-3}$) implies neglecting a contribution 10^{-6} to the normalization integral. The weight of the neglected tails, from which the cut-off radii r_0 (denoted λ_μ) are determined, is of course chosen dependent on the desired maximum "cut-off error". The latter has to be in keeping with the precision of the numerical integration. An atomic radius can be determined as $\lambda_A = \max_{\mu\ A}\left(\lambda_\mu\right)$. All matrix elements of a pair AB can be excluded if $R_{AB} > \lambda_A + \lambda_B$. For a given atom A we have a neighborhood of atoms B which are such (radius λ_B larger than $R_{AB} - \lambda_A$) that the pairs AB have to be treated. This will always be a small number, that will not, for large molecules, increase with system size. Parts of the code with an outer and inner loop over the atoms ($O(N^2)$) transform into a single loop over the atoms with, for each atom, a small inner loop over the neighboring atoms ($O(N)$). As an additional refinement, for an atom pair AB that has to be considered, the function pairs may be checked and be excluded if $R_{AB} > \lambda_\mu + \lambda_\nu$. This will not change the $O(N)$ scaling but will improve the prefactor.

For matrix elements that according to the chosen criterion have to be evaluated, one can still restrict the number of sample points in the numerical integration by using essentially the same criterion to neglect distant points, see section 3.4.

3.3 Linearization of the Density Fitting

The linearization of the density fitting is trivial once an important point is recognized concerning the implementation of the density fitting in the ADF program. To make this clear we briefly review the density fitting as introduced in ref. [1].

The fit coefficients for a density $\rho(r)$, $\rho(\boldsymbol{r}) \rightarrow \tilde{\rho}(\boldsymbol{r}) = \sum a_i f(\boldsymbol{r})$, are determined from a least squares fitting procedure, minimizing the deviation D between true and fitted density subject to the condition that charge is conserved

$$D = \int \left(\rho - \tilde{\rho}\right)^2 d\tau; \quad \int \tilde{\rho} d\tau = \int \rho d\tau = N \tag{3.3.1}$$

Introducing the Lagrangian multiplier 2λ one obtains the equation

$$\boldsymbol{S}\boldsymbol{a} = \boldsymbol{t} + \lambda\boldsymbol{n} \tag{3.3.2}$$

where a is a column vector with the fit coefficients, S is the matrix of fit function overlap integrals, n is a column with the integrals $n_i = \int f_i d\boldsymbol{r}$ and t is a column vector collecting the overlaps $t_i = \int \rho(\boldsymbol{r})\, f_i(\boldsymbol{r}) d\boldsymbol{r}$. Solving for the Lagrange multiplier gives explicitly

$$\begin{aligned} \boldsymbol{a} &= \boldsymbol{S}^{-1}\boldsymbol{t} + \lambda\boldsymbol{S}^{-1}\boldsymbol{n} \\ &= \boldsymbol{S}^{-1}\boldsymbol{t} + \frac{N - \boldsymbol{n}^{\dagger}\boldsymbol{S}^{-1}\boldsymbol{t}}{\boldsymbol{n}^{\dagger}\boldsymbol{S}^{-1}\boldsymbol{n}}\,\boldsymbol{S}^{-1}\boldsymbol{n} \end{aligned} \tag{3.3.3}$$

Dunlap et al.[33] have changed the deviation to be minimized from the least squares deviation in the density to the least squares deviation in the Coulomb self repulsion of $\rho - \tilde{\rho}$. This allows the electronic Coulomb energy, evaluated from the fitted density, to approach the true Coulomb energy from above ("variational fitting for the Coulomb term in the energy").

Carrying out the fitting process as sketched here for the whole density at once will be an $O(N'^3)$ process in view of the linear system of equations (3.3.2) to be solved. Considering possibilities to make the fitting scale more favourably, we first note that apart from the charge conservation condition showing up in the second (Lagrange multiplier) term of the equation for a, the determination of the fit coefficients is a linear process in the fitted object, the density $\rho(r)$: it is possible to write ρ as a sum of parts ($\rho=\sum_i \rho^i$ say) and determine the coefficients as the sum of the coefficients arising from fitting the parts separately, *provided all functions are used for each term* ρ^i.

It is natural to express the charge density $\rho(r)$ as a sum of one- and two-center charge distributions ρ^{AB}:

$$\rho = \sum_{\mu,\nu} P_{\mu\nu}\chi_\mu\chi_\nu = \sum_{A\le B}\sum_{\mu\,A}\sum_{\nu\,B} P_{\mu\nu}\chi_\mu\chi_\nu = \sum_{A\le B}\rho^{AB} \tag{3.3.4}$$

The use of all fit functions for the fit of each ρ^{AB} term implies that apart from functions at A and B, also functions at neighboring centers, C say, would have to be used. The ADF implementation[1] however, carries out the fitting for pair densities ρ^{AB} separately and restricts the fitting of ρ^{AB} to fit functions on A and B only, making the fitting an $O(N^2)$ process. This policy does not seem to have been followed by either Sambe and Felton[34] or Dunlap et al.[33] or

any of the more recent implementations of the use of auxiliary functions, except for the recent work of Gallant and St-Amant,[35] who also partition the total density into a sum of subsystem densities. We briefly review the arguments in favour of atomic pair fitting:

1. The process becomes $O(N^2)$. Moreover, the fitting of each pair density ρ^{AB} with functions on A and B only, implies that the dimension of each linear system (3.3) remains modest and is independent of system size.

2. The possible benefit in fitting ρ^{AB} that could be derived from functions on other atoms is heavily system dependent. Since the presence of atoms C close to ρ^{AB} is accidental, the quality of the fit could vary from one system to another depending on the accidental surrounding of a pair AB by other atoms. A reliable fit basis will have to be able to fit a density ρ^{AB} accurately, with approximately the same precision, under all circumstances, also in cases where other atoms are remote or not present at all, for instance in the diatomic molecule AB.

3. It is not possible to take the accidental presence of atoms C into account when systematically developing sets of fit functions. For fit functions on A this is possible to a certain extent for directly bonded atoms B, since the variation in the number and type of chemical bonds formed by an atom A is limited (nearest neighbor distances for instance fall in a certain range). Fit function sets for an atom A can therefore be generated that can cope with the presence of any "normal" two-center bond with an atom B. The presence of "third parties" C is not certain and can therefore not be exploited in a generation scheme. It is possible that the fit sets generated for reliable pair fits are actually larger than would be necessary in some particular case - the ADF fit sets are indeed relatively large - but this is not a problem since the linear system (3.3) always remains small.

4. The pair fitting scheme can be trivially scaled down to $O(N)$, see below.

Given a pair-wise fitting scheme, linear scaling can simply be achieved by including only atom pairs for which the sum of the radii, determined as described in section 2, is larger than the distance between the atoms. Furthermore, a smaller prefactor can be obtained by eliminating the pairs of basis functions for which the sum of the radii is smaller than the distance between their atoms. We note that the pair-wise fitting has also proven valuable during the parallelization of the ADF program. As long as the number of nodes is smaller than the number of atom pairs, distribution of the one- and two-centre densities over the nodes of a parallel machine, with proper

account of their computational "weights" to achieve load balancing, yields a perfectly scaling parallel implementation.[36] In the present case of course only the pairs that are not excluded are to be distributed.

3.4 Linearization of Manipulations Involving Grid Points: Coulomb and XC Potential Evaluation

The calculation of the XC potential will already be linear when we apply cut-offs to the evaluation of the values of fit functions at the grid points (and derivatives) to obtain the density and XC potential. Since it is the value of a fit function at the grid point that enters the density evaluation, cut-offs can be based on a threshold for the function value: $f_i(r)$ is considered negligible outside a radius r_0 at which the radial part of f_i is equal to a given small threshold, $f_i(\mathbf{r}_0)/\, Z_{lm}(\hat{\mathbf{r}}_0) = \tau$ (Z_{lm} is a real spherical harmonic). We use the weight criterion (3.1.4), since the density is actually used for the XC potential, which enters the spatial integrals making up the Fock matrix elements. The XC potential evaluation is much cheaper than the evaluation of the Coulomb potential, but in our scheme the calculation of the two potentials is closely related. For the distance cut-offs in the Coulomb potential we will follow the same strategy as implemented in the ADF-BAND code for infinite periodic systems.[5] The Coulomb potential in a certain integration point can be written as a sum over the atoms:

$$V_C(\mathbf{r}_k) = \sum_A V_C^A(\mathbf{r}_k) \quad \text{with} \quad V_C^A(\mathbf{r}_k) = \sum_{i \in A} a_i \int \frac{f_i(\mathbf{r}')}{|\mathbf{r}_k - \mathbf{r}'|} d\mathbf{r}' \tag{3.4.1}$$

Applying the expansion of $|\mathbf{r} - \mathbf{r}'|$ in spherical harmonics and using the exponential form of the STO fit functions gives us the following spherical harmonics expansion for the Coulomb potential from each atom:

$$V_C^A(\mathbf{r}_k) \equiv \sum_l \sum_{m=-l}^{m=l} \frac{4\pi}{2l+1} Z_{lm}(\hat{R}_{kA}) I_{lm}^A(R_{kA}) \tag{3.4.2}$$

with $I_{lm}^A(R_{kA}) = \sum_{i \in A} \delta(l, l_i)\delta(m, m_i) a_i I(n_i, l_i, \alpha_i; R_{kA})$

Here R_{kA} is the distance vector from nucleus A to the point r_k, and I_{lm}^{A} is the radial part of the (lm) term in the spherical harmonics expansion. The function I is obtained from incomplete Gamma functions. Going back to the algorithms used to evaluate incomplete Gamma functions, it is apparent that this function can be written as the sum of a multipolar and an exponentially decaying part:

$$I(n_i, l_i, \alpha_i; R_{kA}) = \frac{1}{(R_{kA})^{l_i+1}} \frac{(n_i + l_i + 1)!}{(\alpha_i)^{n_i+l_i+2}} + e^{-\alpha_i R_{kA}} J(n_i, l_i, \alpha_i; R_{kA}) \qquad (3.4.3)$$

The function J consists of a power series in R_{kA}, with n_i as highest power. This is just one factor R_{kA} more than in the value of the STO fit function itself, which of course has a simple $r^{n_i-1}\ e^{-\alpha_i r}$ radial behavior. For the sake of clarity one cut-off radius for both the exponential part of the Coulomb potential and the function value for the density evaluation, is used, where the $r^{n_i}\ e^{-\alpha_i r}$ radial behaviour of the former is to be taken into account when determining the cut-off threshold.

The multipolar part of the Coulomb potential is long-ranged, which hampers achieving linear scaling. It is advantageous to first compute the "strength" M_{lm}^{A} of an atomic multipolar term of nucleus A as

$$M_{lm}^{A} = \sum_{i \in A} \delta(l, l_i)\delta(m, m_i) a_i \frac{(n_i + l + 1)!}{(\alpha_i)^{n_i+l+2}} \qquad (3.4.4)$$

so that the multipolar term becomes

$$Q_{lm}(R_{kA}) = \frac{M_{lm}^{A}}{(R_{kA})^{l+1}} \qquad (3.4.5)$$

The multipolar strength can be calculated once for all atoms (at a certain SCF cycle), and then the possibility to neglect higher order multipoles can, for a particular block of grid points, be based on the strength M_{lm}^{A} as well as the distances R_{kA}. For the size of systems we have been considering, it has never been possible to neglect also the l=0 terms, not even from the most distant atoms, so there has always remained an $O(N^2)$ term in the Coulomb potential evaluation. It has been possible to neglect higher multipole terms, in particular above l=1. We will investigate below to what extent the $O(N^2)$ scaling of the l=0 potential terms, which only require a small fraction of the total time in the calculations without cut-offs, affects the overall scaling in the systems considered (in the order

of a several hundreds of atoms). Obviously, the structure of the whole scheme is suitable for fast multipole techniques,[8-10,20] as are being applied now in Gaussian based codes.[5,8-12,14] Nevertheless, it appears that such techniques would only pay off, and are actually only needed in the present scheme, when the size of systems becomes one order of magnitude larger again.

3.5 Linearization of Manipulations Involving Grid Points: Function Evaluation and Fock Matrix Set-up.

Pérez-Jordá and Yang[7] have developed an $O(N)$ method for the evaluation of the density at the grid points, and of matrices such as the overlap matrix, based on the construction of the sets $S(r_k)$ of basis functions that have nonnegligible values at a given grid point r_k (we ignore here the partition function p_α that also plays a role in the development of ref.):[7]

$$S(\mathbf{r}_k): \left\{ \chi_\mu \,\middle|\, |\chi_\mu(\mathbf{r}_k)| > \tau \right\} \tag{3.5.1}$$

They determine the set of relevant basis functions for each grid point with an efficient tree algorithm. However, ADF uses (as most DFT implementations, see TURBOMOLE[37] and GAUSSIAN94)[38] for efficiency reasons blocks of points, so that vectorization in inner loops over grid points can be used. Moreover, as explained in section 3.2, we do not base the cut-off decisions on a radius for each basis function that is determined by a fixed threshold for the function value. We wish to base the cut-off decisions on an estimate of the percentage of the matrix elements that will be neglected. Therefore the set of functions that has to be taken into account for a given block of points B is based on whether or not all the points of the block are located in the region of space beyond the radius r_0 which according to section 2 contributes less than a given percentage (w of eq. 3.2.4) to the radial integral over the whole space. We define the tail function $T(\chi_\mu;r_k)$ for function χ_μ with respect to a point r_k as the weight in the sense of section 2 of the radial integral of χ_μ beyond a radius which is the distance from the center of the function to the point r_k, or

$$T(\chi_\mu^A;\boldsymbol{r}_k) = w(\chi_\mu; |\boldsymbol{r}_k - \boldsymbol{R}_A|) \tag{3.5.2}$$

In case a block of points B is specified ($T(\chi_\mu;B)$), the radius is to be taken as the distance from the center of the function to the nearest point of the block. The use of the tail function allows a more even-handed treatment of diffuse and contracted functions than the uniform cutting of basis functions at a given value. We now define the set of functions S(B) that belong to a block of points B as follows:

χ_μ (on atom A) belongs to S(B) if there exists a χ_ν (on atom C) such that both the following conditions are satisfied:

$$\text{a) } R_{AC} < \lambda_\mu + \lambda_\nu \text{ ;} \quad \text{b) } T(\chi_\mu;B)*T(\chi_\nu;B) > \text{threshold} \tag{3.5.3}$$

Atoms A and C are chosen from the set of atoms that possess functions with a tail $T(\chi_\mu;B)$ over the block larger than a given threshold. The condition 3.5.3a then means that the function χ_μ is sufficiently close to some other function χ_ν to have a nonnegligible matrix element with it. The second condition checks if the contribution of the block B to the numerical integration that makes up the $\mu\nu$ matrix element would indeed be nonnegligible.

Stratman et al.[22] have pointed out that the points in the block should be localized in space in order to keep the set of relevant basis functions limited. As a matter of fact, the Voronoy polyhedron scheme of point generation in ADF already generates points in spatial cells (small spheres around an atom, or wedges of the atom's Voronoy cell).[39,40]

As pointed out in refs.[7,22] the creation of the sets S(B) of basis (and fit) functions belonging to block B with just condition (5.1) is an $O(N^2)$ step, the number of checks required being in principle (number of points)*(number of functions). In our case, where pair tests are being done for each block, cf. (5.3), it is in principle an $O(N^3)$ step. However, with present target sizes of molecules of up to 1000 atoms, it is still very cheap. It is organized efficiently in a tree structure by first, for a given block of points, excluding (most) atoms on account of their distance from the block, and considering only pairs $\mu\nu$ belonging to the (small) set of atoms "belonging" to the block.

Once the sets of functions S(B) belonging to the blocks B have been determined, the calculation of the function values is evidently $O(N)$.

The calculation of matrices of operators like the identity (overlap matrix) or $V_N(r)$, $V_{Coul}(r)$,

V_{XC}(r) now also is automatically linear, if the values of the potentials in the points are known (see previous section), since for each block of points B a limited number of basis functions has to be processed, which is independent of system size. Let us denote the total operator value with $\Omega(r)$ (operators like the kinetic energy $\hat{T} = \sum_i -\frac{1}{2}\ ^2(i)$ can be treated in a completely analogous way), then for the matrix of the operator $\hat{\Omega}$, and also for the density if it is calculated from the density matrix, we have the following double loop over the basis functions:

{*Construct contributionof block B to matrixelements; construct valuesof densityin pointsof B*}
For μ $S(B)$
for all r_k B: $\xi_\mu(r_k) = \Omega(r_k)\chi_\mu(r_k)$

For ν $S(B)$
if $R_{AC} < \lambda_\mu + \lambda_\nu$ *then*
if $T(\chi_\mu; B)$ $T(\chi_\nu; B) >$ *threshold then*
{*add contributionof B to* $\Omega_{\mu\nu}$}
$\Omega_{\mu\nu} = \Omega_{\mu\nu} + \sum_{r_k \ B} \chi_\nu(r_k)\ \xi_\mu(r_k)$ (*operatormatrix*)
{*calculate densityin pointsof B*}
for all r_k B: $\rho(r_k) = \rho(r_k) + P_{\mu\nu}\chi_\mu(r_k)\chi_\nu(r_k)$ (*density*)
End ν
End μ

We have included here the density evaluation from the density matrix to show that it follows essentially the same route, although in our implementation the density is directly evaluated from the fit functions during the SCF. Note that the use of the tail function $T(\chi_\mu; B)$ does not change the scaling, which would also become linear if only tests on a cut-off radius would be performed. However, it allows an additional refinement by which the contracted or diffuse nature of basis functions, which is important for the contribution of block B to the matrix elements, is properly accounted for.

3.6 Results and Discussion

This section contains an overview of the effect of the implementation of cut-offs in various stages of the calculation of the (electronic potentials part of the) Fock matrix, including graphs with timings for the separate routines of the original code of ADF and the adapted code including the cut-offs. As test case we choose alkanes with the following geometry: the carbon atoms lie in one plane and form a fully extended zig-zag chain, the hydrogen atoms are below and above the plane. As second test case we took hypothetical 2-dimensional systems, namely hexagonal graphitic sheets, which are terminated by nitrogens. Figure 3.1 shows the largest one. The fitting of our timing curves to a power function is done with the Levenberg-Marquardtalgorithm.[41] For the investigation of the computational efficiency in a three-dimensional system we have chosen the $Pt(P(Ph)_3)_3CO$ complex.

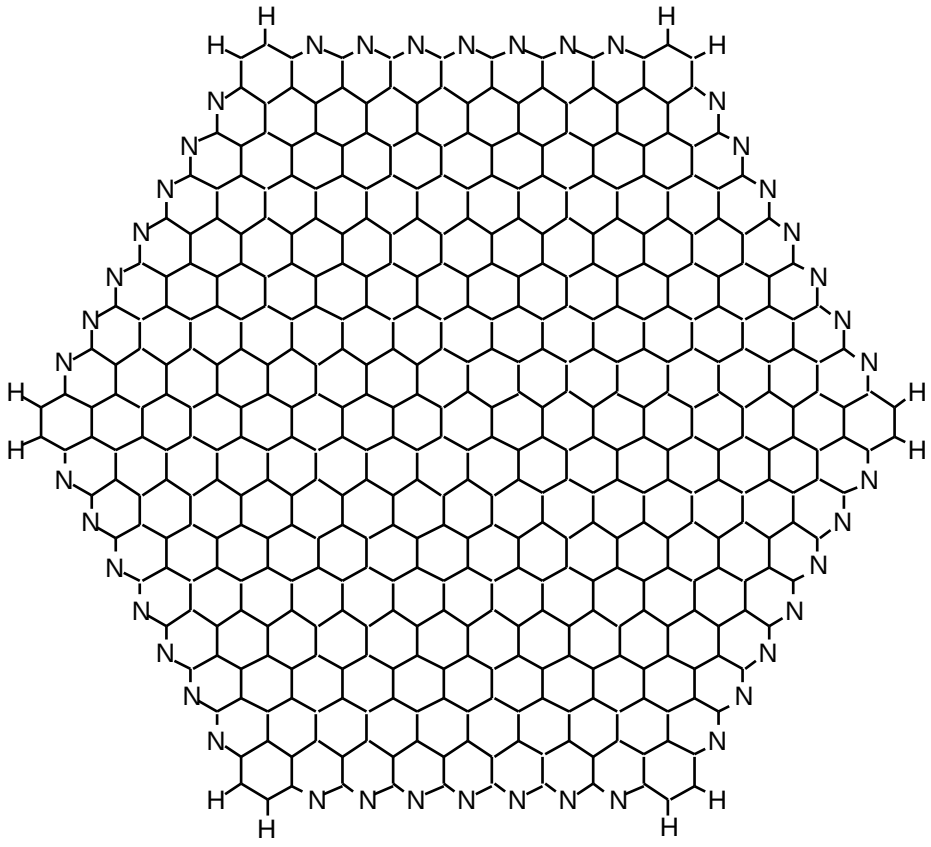


Figure 3.1. The largest planar aromatic system used for the timing results.

3.6.1 Density Fit

Figures 3.2 and 3.3 display the scaling behavior of the two routines involved in the fitting of the density. The first routine (FITINT) calculates prior to the SCF cycles for each pair of atoms the elements of the matrices *S* and *n*, and the individual integrals $t_{\mu\nu,i} = \left\langle \chi_\mu^A \chi_\nu^B f_i^{AorB} \right\rangle$, which are contracted on each cycle with the density matrix elements $P_{\mu\nu}$ to generate the t vector (cf. eqs. 3.3.2, 3.3.3), and writes these integrals to file (timing in Figure 3.2). The second routine (RHOFIH) reads on every SCF cycle the matrix elements from file and performs the fitting of the present density by solving equation (3.3.3) for each pair of atoms (timing in Figure 3.3).

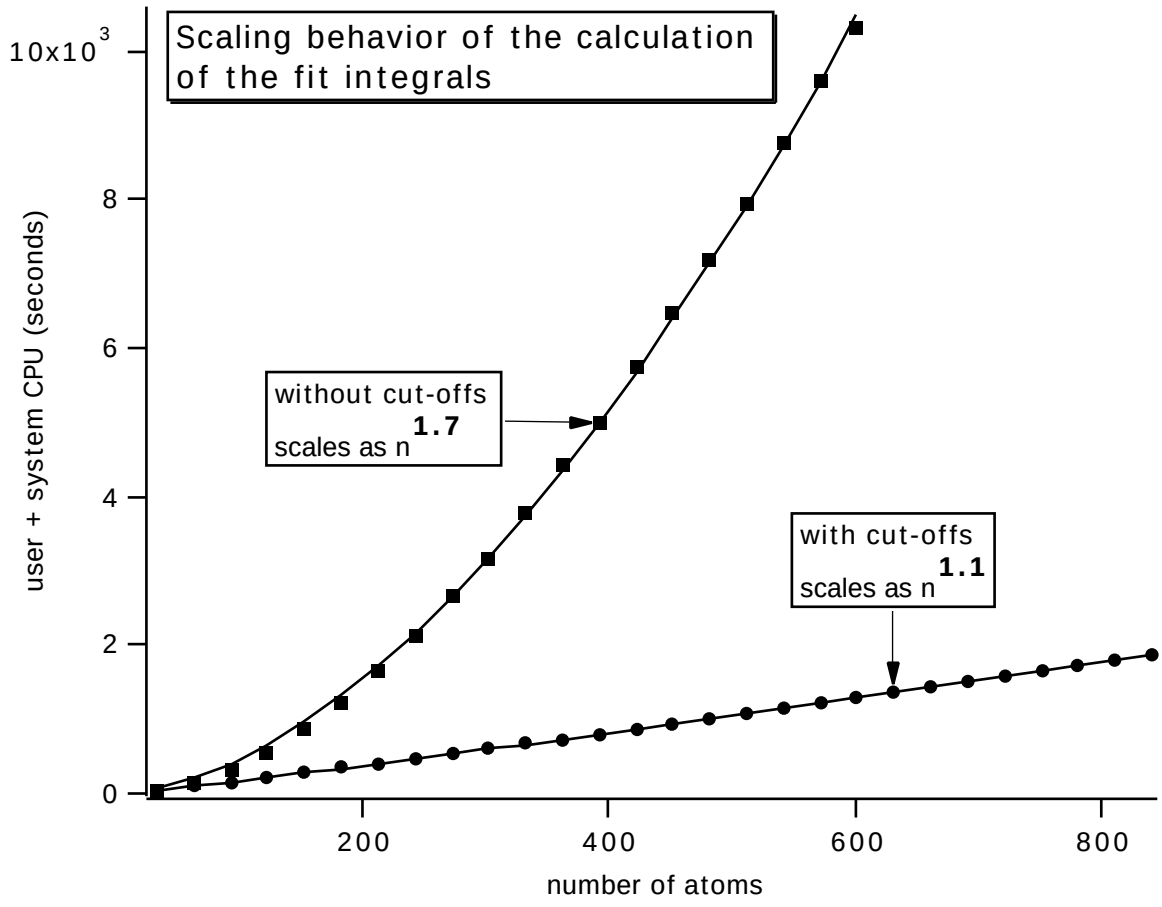


Figure 3.2. The scaling behavior of the routine (FITINT) that calculates and writes the fit integrals to file, is shown with and without the use of cut-offs, for a series of alkane chains with the indicated total numbers of atoms.

Both Figures show the scaling with and without the application of cut-offs to the basis functions. The use of cut-offs implies that we are only fitting atom-pair densities of "neighbouring" atoms. The calculation of the fit integrals, which is dominated by the integrals $t_{\mu\nu,i} = \left\langle \chi_\mu^A \chi_\nu^B f_i^{AorB} \right\rangle$, meets our expectations in that Figure 3.2 shows a decrease of an almost quadratic scaling to an almost linear scaling. The small deviations from quadratic and linear scaling respectively are due to special circumstances. The lower scaling than 2.0 in the old code arises from atom pairs with large distance being already cheaper with the current integral routines. The small deviation from linear scaling of the calculation with cut-offs may be due to IO effects perturbing the timing, since we could determine this to be the origin also of the scaling of the routine RHOFIH (Figure 3.3) being not as predicted.

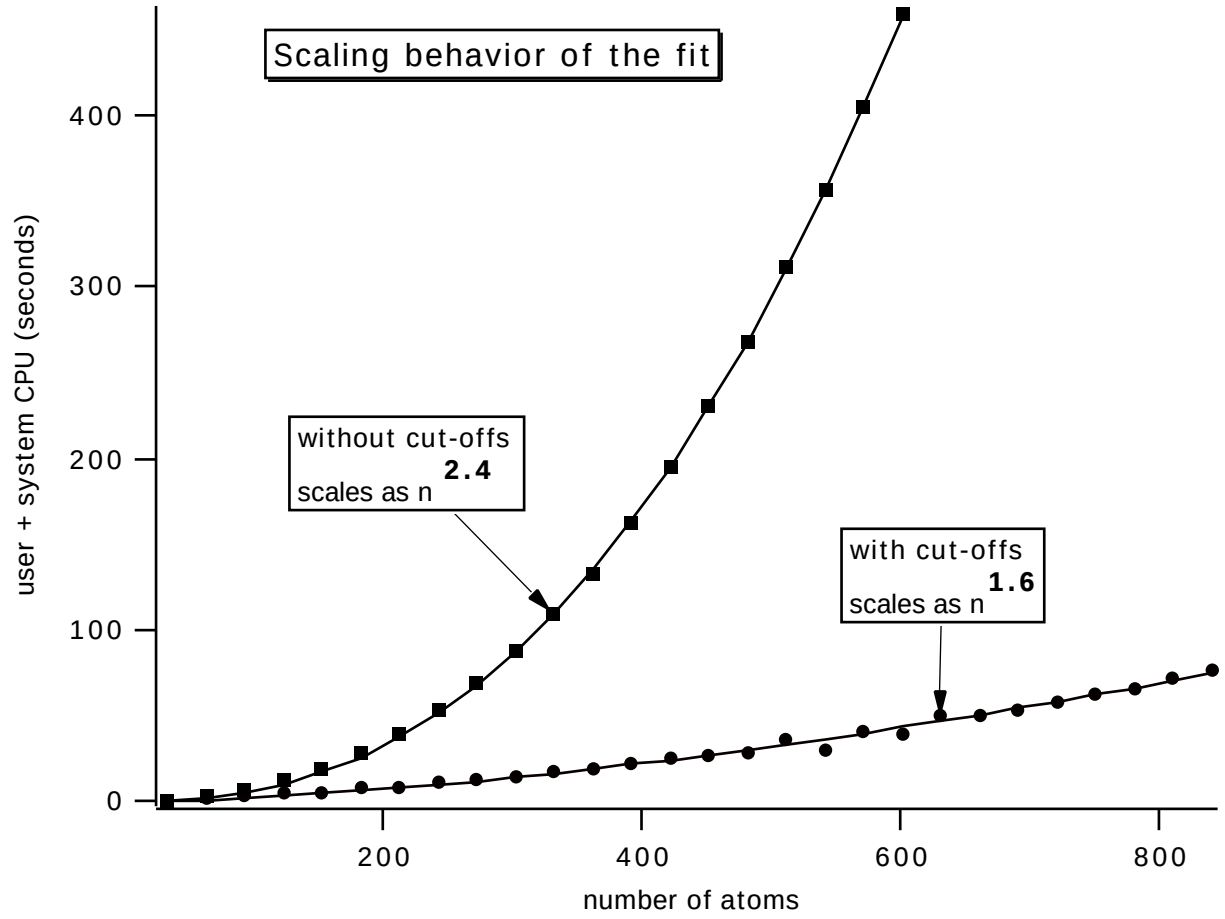


Figure 3.3. For the alkane chains, the scaling behaviour is shown of the routine (RHOFIH) that reads the fit integrals from file and solves the least squares equation for the fit coefficients.

The scaling of RHOFIH without cut-offs (3.2.4) as well as with the cut-offs (3.1.6) is higher than predicted. Analysis of this routine showed that the discrepancy between the actual scaling and the predicted one is caused by an I/O bottleneck. The routine is not computationally intensive but spends the largest part of its time in I/O manipulations, viz. the reading of the fit integrals (compare the absolute times in Figures 3.2 and 3.3). For the smaller molecules the file with the fit integrals will fit into memory. However for the larger molecules this becomes impossible, as the size of this file scales linearly with the number of atoms. This causes at certain molecular sizes a tremendous growth of paging which perturbs the timings. The same will be happening in the FITINT routine (Figure 3.2) and may become visible for the smaller computation time in the calculation with cut-offs.

3.6.2 The Coulomb Potential

In Figure 3.4 the scaling of the calculation of the Coulomb and the XC potential are shown with and without cut-offs. The reduction in the scaling power is rather small, although even this already implies a reduction in computation time with a factor of 10 for the 600-atom system. We examine these timings more closely to understand the small decrease in the scaling power. In Figure 3.5 the different components of the calculation of the Coulomb and XC potential are presented.

First, the density in the grid points $\rho(r_k)$ is evaluated as written in equation (3.1.3). Due to the cut-offs for each grid point only fit functions in its neighbourhood are used. This results in a perfectly linear scaling of this step. The evaluation of the XC potential is also linear, because it only manipulates the density in the grid points. With $V_{\text{Coulomb}}^{\text{sr}}$ we denote the calculation of the short-range (exponential) part of the Coulomb potential, equation (3.4.3). For this component of the Coulomb potential the same cut-off radii are applied as for the density. Therefore, also this component of the potential scales linearly.

For the evaluation of the multipolar parts $Q_{lm}^{A}(R_{kA}) = M_{lm}^{A} / (R_{kA})^{l+1}$ of the Coulomb potential, $1/R_{kA}$ has to be calculated for each grid point r_k with respect to all atoms in the system. A few computational operations are needed to determine $1/R_{kA}$: first, $(R_{kA})^2$ from ${x_{kA}}^2 + {y_{kA}}^2 + {z_{kA}}^2$, then R_{kA} from $\sqrt{(R_{kA})^2}$ and finally $1/R_{kA}$. We show explicitly just how much time the simple

evaluation of the $1/R_{kA}$ values takes by adding the computation time to that of V_{XC} and $V_{Coulomb}^{sr}$. This step of course scales quadratically, and since it takes an amount of time that is comparable to the previous steps, its addition changes the scaling from linear to ca. 1.5 for systems up to ca. 1000 atoms. The relatively large computation time of this simple step can be attributed to the evaluation of the square root being extremely expensive. The last curve shows the scaling for the complete evaluation of the Coulomb and the XC potential. Comparing the last two curves we see that once the $1/R_{kA}$ are known, only little time is needed to evaluate the full multipolar contributions of equation (3.4.5).

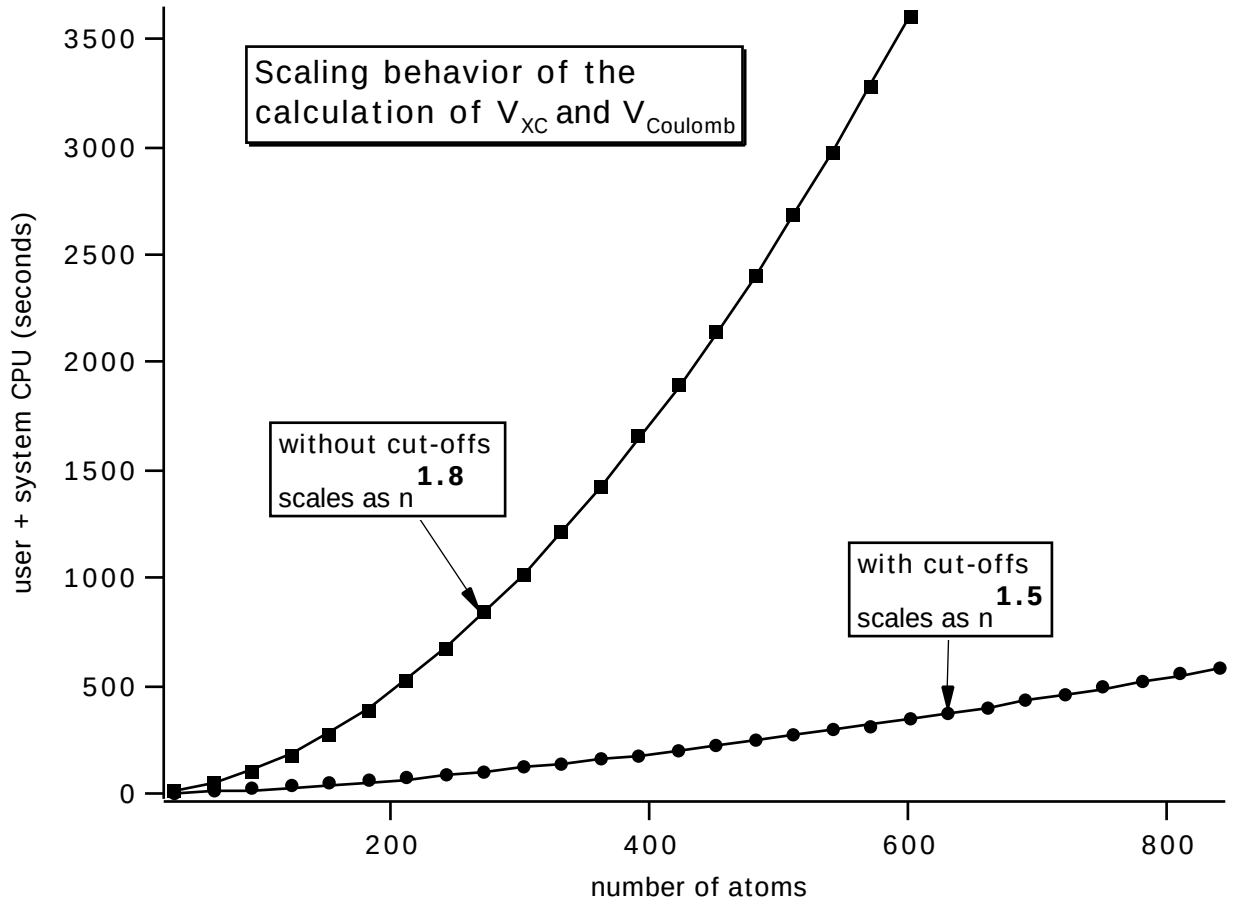


Figure 3.4. For the alkane chains, the scaling behaviour is shown of the routine that calculates the Coulomb and the XC potential in the grid points.

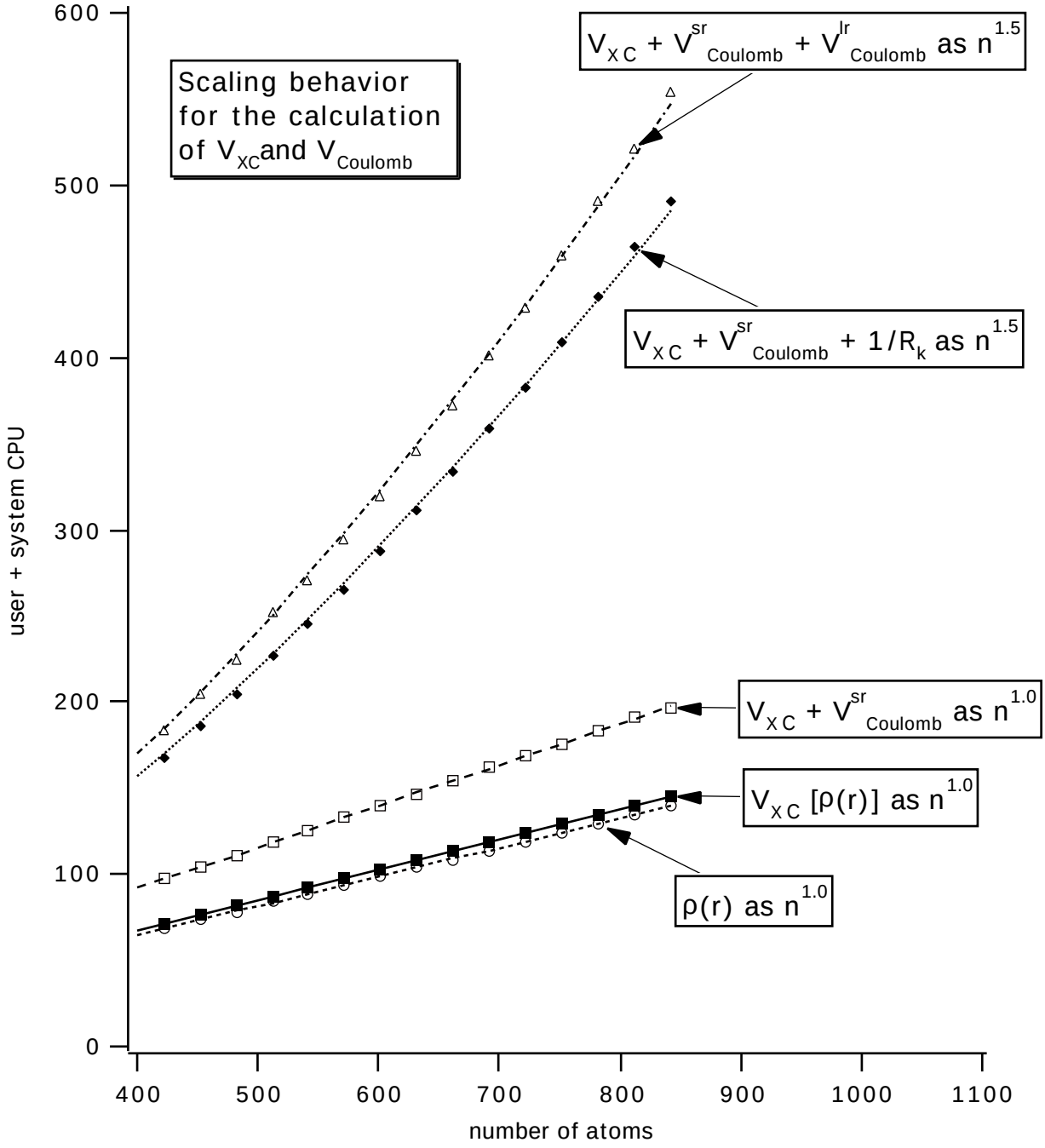


Figure 3.5. Timing and scaling behavior for the various steps in the calculation of the Coulomb and the XC potentials in the grid points when cut-offs are used (alkane chains). $V^{sr}_{Coulomb}$ is the short range (exponential) part of the Coulomb potential, $V^{lr}_{Coulomb}$ is the long range (multipolar) part.

Since the evaluation of the long-range part of the Coulomb potential is a quadratically scaling step in our present implementation, its computation time would ultimately start to dominate. For the size of systems we are interested in, the evaluation of the long-range part, although algorithmically very simple, already takes an amount of time that is comparable to (even already somewhat larger than) the sum of the exchange-correlation potential and the short-range part of the Coulomb potential. Since several of the other computation steps can be scaled down very successfully (see the comparison for total timings below), the computation time for the long-range Coulomb potential changes from a modest fraction of the time without cut-offs to a significant fraction after application of cut-offs. The computational expense is still acceptable, but it is one of the prime targets for future efforts towards computational speed-up.

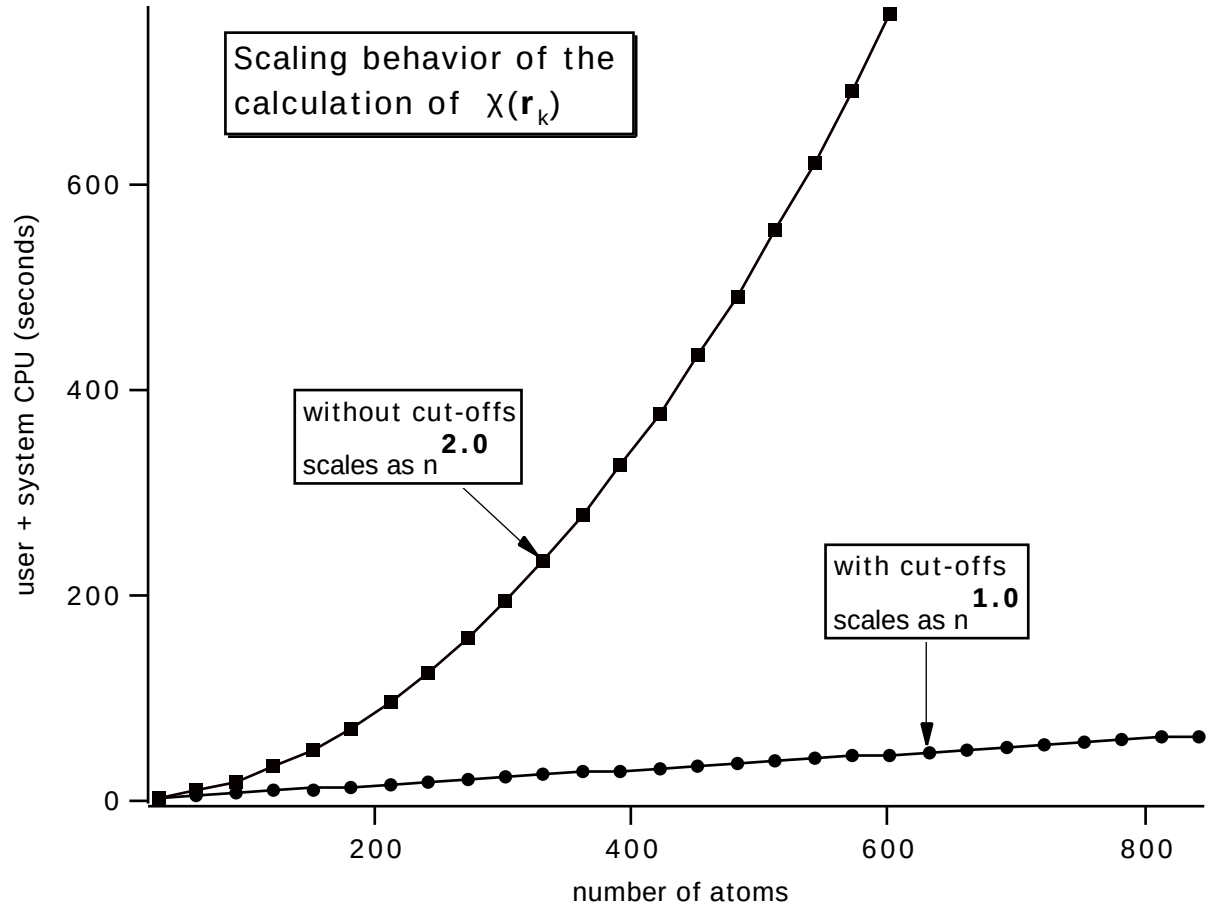


Figure 3.6. The scaling behaviour (alkane chains) of the routine that calculates the values of the basis functions in the grid points, is shown with and without the use of cut-offs.

3.6.3 The Fock Matrix

Besides the potential in the integration points we need the values of the basis functions in the integration points to set up the Fock matrix. Figure 3.6 shows the two curves, with and without cut-offs, for the evaluation of $\chi(r_k)$. The original code of the routine scales quadratically due to the loop over the functions and the loop over the integration points. The adapted code scales perfectly linearly as expected.

Having all ingredients available (function and potential values in the grid points) the final set-up of the Fock matrix is shown in Figure 3.7. A double loop over the atoms and a loop over the integration points in the original code give a cubic scaling. The adapted code shows exactly linear scaling with system size. Dramatic improvement of the computational efficiency results.

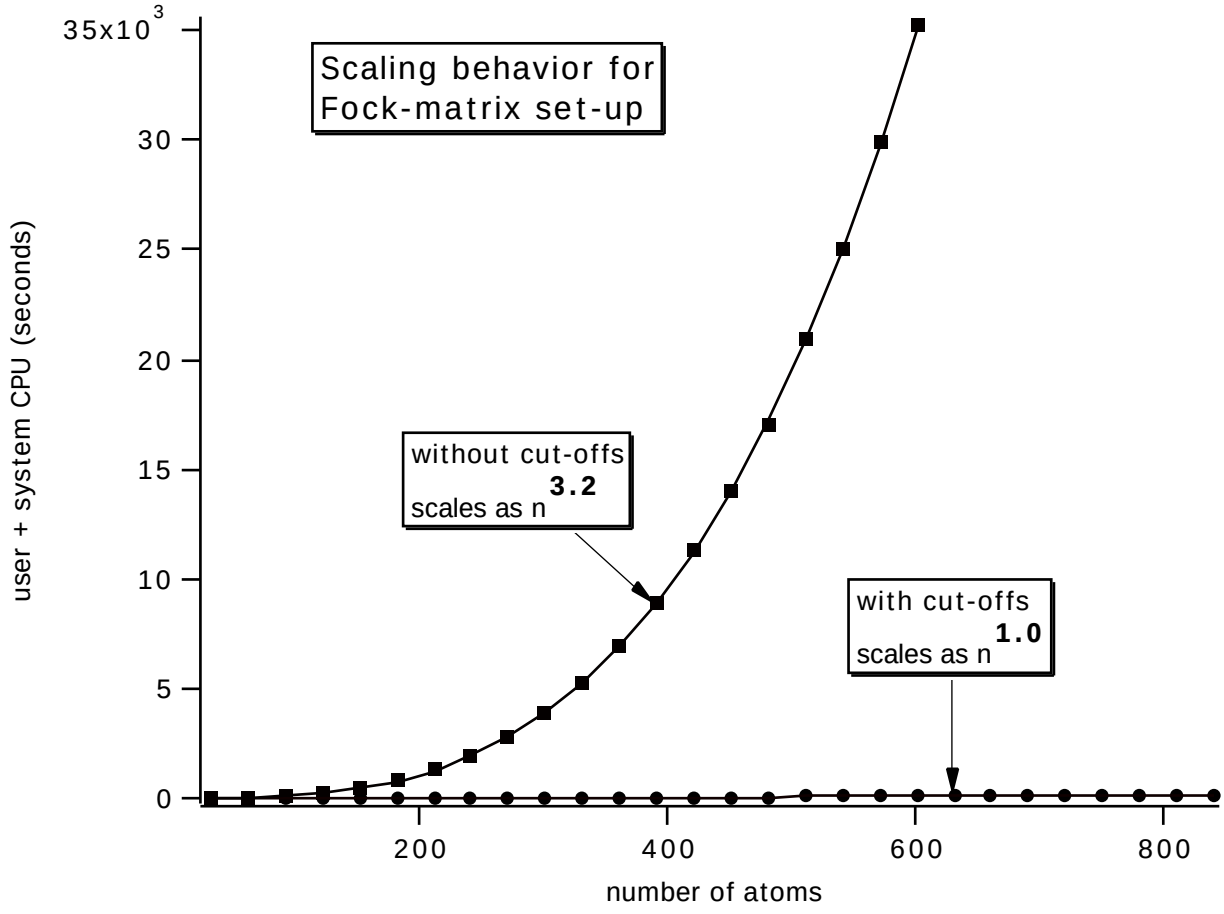


Figure 3.7. The scaling behaviour (alkane chains) for the final set-up of the Fock matrix, when potentials and basis function values are known, with and without the use of cut-offs.

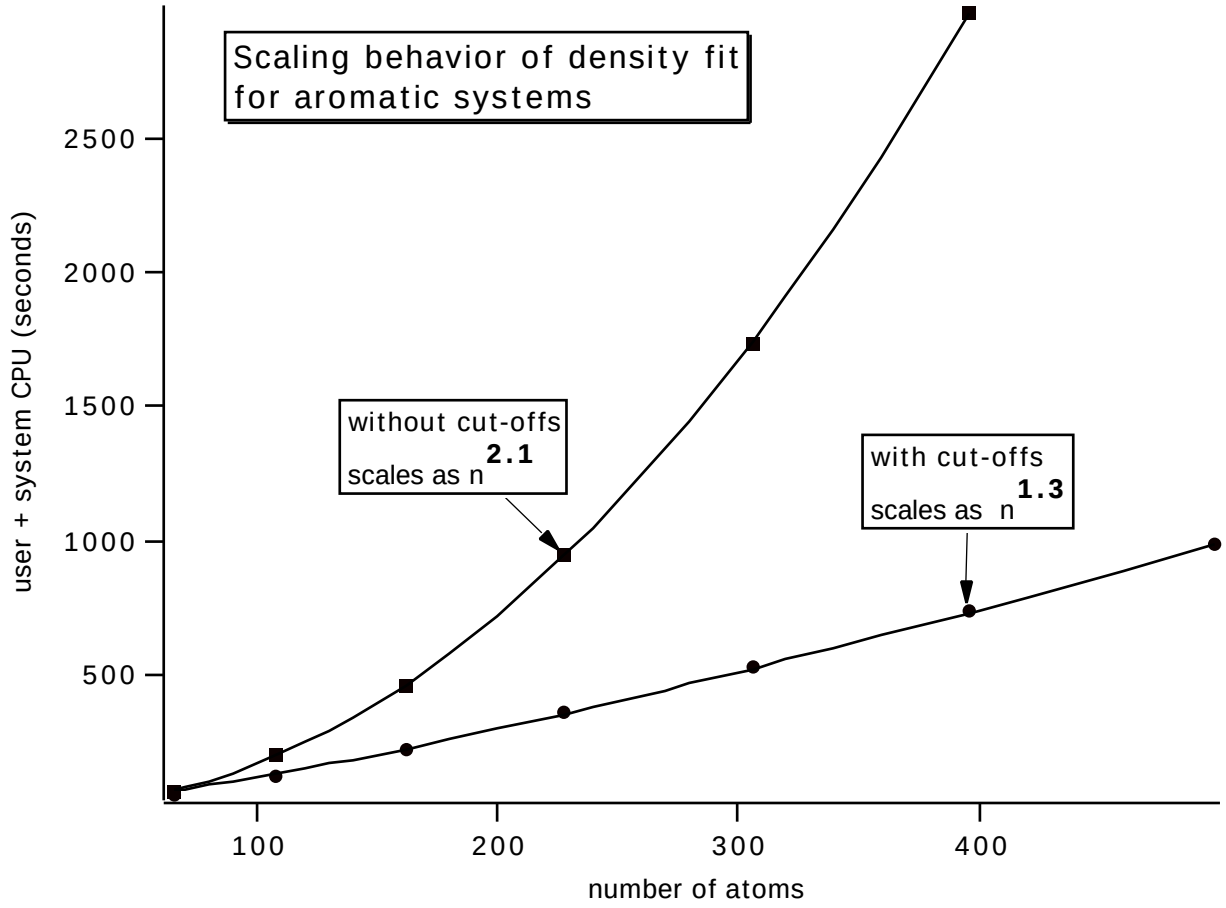


Figure 3.8. The combined scaling behaviour in the planar aromatic systems of the two routines (FITINT and RHOFIH) that are involved in the fitting of the density, is shown with and without the use of cut-offs (one call of each routine).

3.6.4 Aromatic Systems

The applicability of our method on 2-dimensional systems is demonstrated with the results of the aromatic systems. We do not show the individual routines but the timing is shown for the sum of the two routines involved in the density fit (Figure 3.8) and the calculation of the total Fock matrix (Figure 3.9), including the calculation of function values, of the potentials, and of the matrix elements of the Fock matrix with both the kinetic energy and the electronic potentials as operators. In Figure 3.8 the scaling behavior with cut-offs is dominated by the evaluation of the fit integrals. The aromatic systems are not yet large enough to enter the size range where the cut-offs

cause the number of pairs to be handled to increase linearly (note that in one dimension even the largest aromatic system has the size of only a C_{35} alkane chain). The number of pairs still increases like ca. 1.3, explaining the observed scaling. The poorer scaling of RHOFIH (1.7) is not visible due to the the computation time being more than an order of magnitude less than that of FITINT. In Figure 3.9 the scaling achieved in the calculation of the Fock matrix with cut-offs applied is an average over the various components of the calculation. The quadratic scaling of the long-range Coulomb part - still relatively cheap at this system size - is not yet important, the 1.3 scaling has its origin in the increase of the number of pairs, as in the case of the fit integrals.

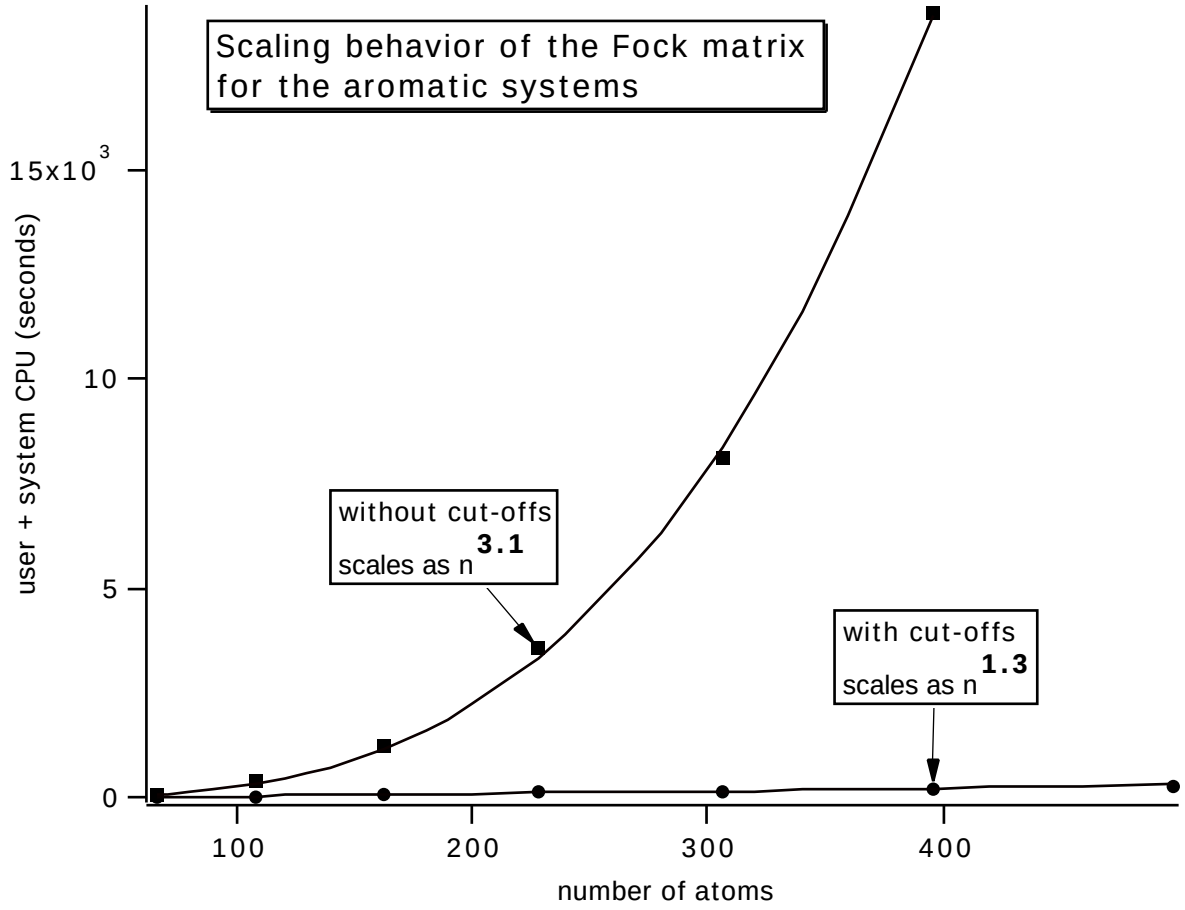


Figure 3.9. The combined scaling behaviour of all routines that are used to set-up the Fock matrix: the calculation of the Coulomb and the XC potential in the grid points, the evaluation of the basis functions in the grid points, the initialization of the Fock matrix with the kinetic energy part, and the final set-up of the Fock matrix. The timings are shown with and without the use of cut-offs.

3.6.5 Relative Error

Our objective throughout this project has been to improve the scaling with system size, while retaining sufficient precision in the results. It is natural to gauge the required precision against the precision to which the numerical integration is typically set. A key quantity is the bond energy, i.e. the energy of the molecule minus the sum of the energies of the atoms. With increasing system size, the *relative* error in this quantity is required to stay constant (approximately), implying a constant absolute error in the bonding energy *per bond*, which of course also stays roughly constant irrespective of system size. The numerical integration has the required property of yielding a constant relative error independent of system size, its magnitude being determined of course by the setting of the accuracy parameter for the numerical integration. The cut-off parameters should be set in such a way that if the numerical integration affords a bonding energy with a relative error of only 0.1% say, also the calculation done with the cut-off radii should return bonding energies with a relative error not larger than 0.1%.

The setting of the cut-off parameters has been tested for the following molecules as test cases: $C_{50}H_{102}$, $C_{100}H_{202}$ and $Pt(P(Ph)_3)_3CO$. Each cut-off parameter was investigated separately to prevent spurious results from accidental cancellation of errors and also with a high numerical precision (relative error less than 10^{-4} %) to avoid numerical noise. Our goal has been to get for each cut-off parameter a setting, that gives a relative error in the bonding energy smaller than 0.1% and at the same time stability in the computational process in the sense that not just the final bonding energy is stable to the required precision but also the SCF cycling proceeds in the same way, so that also the same number of SCF cycles is needed to reach convergence. The latter requirement resulted for some of the cut-offs in values that lead to a relative error in the bonding energy much less than 0.1%.

The cut-off used in the fitting of the density is determined by the radius of a basis function as calculated from the weight parameter w for the neglect of the tail of the basis functions given in eq. (3.2.4). The radii are used to decide which pairs of basis functions have negligible overlap. Neglect of basis function pairs $\chi_\mu\chi_\nu$ in the fitting of the density leads to loss of charge (the contribution $P_{\mu\nu}S_{\mu\nu}$ is neglected) and to a relative error in the bonding energy. A good choice for the cut-off criterion w turned out to be 0.05%. This gives for the carbon atom a radius of 5.8 Å. The loss of charge is almost negligible: less than 10^{-7}%. Although this is a small charge deficit, it

turned out to be important to rescale the fit coefficients so the exact charge is retained. This rescaling of the number of electrons in the fit density proved to be particularly beneficial for stability in the (number of) SCF cycles. The relative error in the bonding energy due to this cut-off in the density fit is very small: less than 10^{-6}%.

The calculation of the Coulomb and XC potential in the grid points uses two cut-off criteria. One is for the cut-off of the tails of the fit functions, which is used for both the density evaluation and the short-range Coulomb potential. The weight function w of eq. (3.2.4) is used as criterion to determine the function radius. The other cut-off is for the multipolar part of the Coulomb potential (equation (3.4.5)). The influence of the two cut-off parameters on the relative error of the bonding energy and the convergence of the SCF was investigated independently. The cut-off criterion of the fit functions was set to 0.005%, which implies a maximum radius of carbon of 7.4 Å. The relative error in the bond energy was for $Pt(P(Ph)_3)_3CO$ about 0.01%, but for the alkanes it was much smaller: about 10^{-5}%. The cut-off for the multipolar part was set to 10^{-5} a.u., i.e. a multipole term $M_{lm}^{A}/(R_{kA})^{l+1}$ is evaluated in all the points of a block B if $M_{lm}^{A}/(R_{kA})^{l+1}$ in the point r_k of the block nearest to atom A is larger than 10^{-5} a.u.. In this case the error in the energy was less than 10^{-5}%.

For the building of the Fock-matrix two cut-off criteria are used: a) the threshold that determines if two basis functions have a sufficiently large overlap and therefore Fock matrix element, analogous to the (independent) threshold used to decide which basis function pairs can be neglected in the density fit; b) the tail-function indicating the weight of a basis function over a block of points, which determines if the points of a block have to be used for the numerical integration of a Fock matrix element, eq. (3.5.3). We will refer to the former cut-off as the overlap threshold. Again the two cut-off criteria are investigated separately. We investigated the error for the building of the Fock matrix by comparing the matrix of the unit operator, calculated numerically with either one of the cut-offs applied, to the analytically calculated "exact" overlap matrix. The radii of the basis functions for the overlap threshold were determined with w (eq. 3.2.4) set to 0.05%. The maximum radius of the carbon atom in the building of the Fock-matrix then equals 7.4 Å. The maximum absolute error in any overlap matrix element is less than 10^{-4} for this cut-off. To determine if a block B has to be used for a $\langle \chi_\mu | \hat{h}_s | \chi_\nu \rangle$ matrix element of the Fock matrix, the threshold of eq. (3.5.3b) was set to 10^{-4}, which resulted in a maximum absolute error

of also 10^{-4}.

We have finally ascertained that with the above mentioned setting of the cut-off criteria the relative error in the bond energy does not increase with the system size. Figure 3.10 displays the relative error of the bond energy with a precision of 10^{-6} (10^{-4}%) in the numerical integrations of the Fock matrix elements, as well as with a numerical precision of 10^{-3}. A numerical precision of 10^{-6} assures that we do not have errors due to numerical integration and cut-offs that accidentally cancel each other. The results with numerical precision of 10^{-3} are displayed because this precision was used for the timings. We see that the cut-off criteria satisfy our requirement of a relative error less than 0.1% and of a stable relative error, independent of system size.

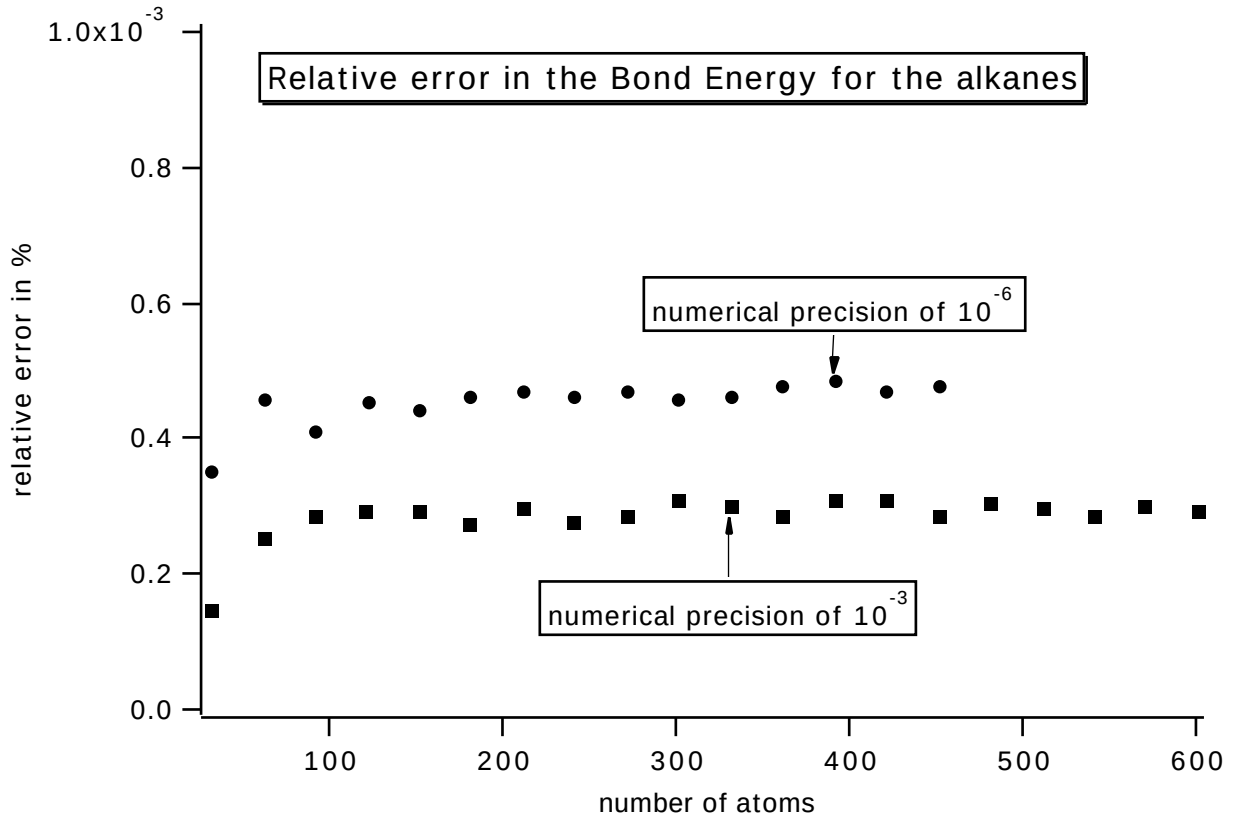


Figure 3.10. The relative error in the bond energy (total energy minus sum of atomic energies) for the alkanes. The cases with a numerical integration precision of 10^{-3} and 10^{-6} are shown.

Table 3.1. Timing results (times in seconds) for $Pt(P(Ph)_3)_3CO$

	old	new	speedup
calculation of fit integrals	2892.1	2582.7	1.1
calculation of kinetic energy matrix	4089.5	759.9	5.4
calculation of V_{Coulomb} and V_{XC}	926.1	266.3	3.5
calculation of $\chi(r_k)$	162.3	90.5	1.8
set-up of Fock matrix	2292.4	416.1	5.5
determination of fit coefficients	144.7	95.5	1.5
diagonalization	0.8		

Table 3.2. Timing results (times in seconds) for $C_{35}H_{72}$

	old	new	speedup
calculation of fit integrals	420.5	179.6	2.3
calculation of kinetic energy matrix	343.2	28.7	12.0
calculation of V_{Coulomb} and V_{XC}	138.8	31.3	4.4
calculation of $\chi(r_k)$	24.7	7.3	3.4
set-up of Fock matrix	170.4	11.5	14.8
determination of fit coefficients	9.4	3.6	2.6
diagonalization	8.0		

Table 3.3. Timing results (times in seconds) for $C_{200}H_{402}$

	old	new	speedup
calculation of fit integrals	10342.2	1280.8	8.1
calculation of kinetic energy matrix	62777.0	205.0	306.2
calculation of V_{Coulomb} and V_{XC}	3602.8	352.9	10.2
calculation of $\chi(r_k)$	765.5	44.5	17.2
set-up of Fock matrix	35277.0	74.8	471.6
determination of fit coefficients	459.0	39.6	11.6
diagonalization	53.5		

3.7 Summary

We present in Tables 1-3 some overall timings for the old and new codes. As a genuinely three-dimensional system $Pt(P(Ph)_3)_3CO$ is included, where three $P(Ph)_3$ ligands are directly coordinated to the Pt atom, as is the CO ligand. The number of atoms (105) is equal to that in the chain-like alkane $C_{35}H_{72}$. The 602 atom alkane $C_{200}H_{402}$ is more typical for the size of system for which the present development is intended. The numbers of STO basis functions are 685, 459 and 2604 for these three systems respectively.

In the old situation, the calculation of the kinetic energy matrix, which is presented separately since it is carried out once, prior to the SCF cycles, and the setting up of the Fock matrix of the electronic potentials, are the most time-consuming parts of the calculation for $Pt(P(Ph)_3)_3CO$ and in particular for $C_{200}H_{402}$. They are also the steps for which (by far) the largest speed-up factors are achieved. Although the total speed-up is already quite significant for the 105-atom systems, the real benefit of the better scaling is of course most apparent in the largest system. In all systems the relative weights of computational burden shift from the matrix evaluation to the calculation of the fit integrals. This is notably the case for the Pt complex, which is related to the large number of basis functions on the Pt atom. It should be realized however, that the calculation of the fit integrals is only executed once, as is the calculation of the kinetic energy matrix, whereas the other steps are repeated on each cycle (in the so-called direct-SCF mode). Incidentally we note that the diagonalization times are still fairly modest and the diagonalization is, in spite of its N^3 scaling, not yet a bottleneck.

We conclude that significant advances have been made towards an $O(N)$ DFT method. The calculation of the fit integrals is still relatively time-consuming but has scaled down to linear scaling. The calculation of the Coulomb and XC potentials is, in the alkane chains, 4-6 times less expensive, but since its scaling has not been reduced to linear, it will eventually dominate and should be the next target for algorithmic development.

References

[1] E. J. Baerends, D. E. Ellis, P. Ros, *Chem. Phys.* **1973**, *2*, 41

[2] D. E. Ellis, H. Adachi, F. W. Averill, *Surf. Sci.* **1976**, *58*, 497

[3] A. Rosén, D. E.Ellis, H. Adachi, F. W. Averill, *J. Chem. Phys.* **1976**, *65*, 3629

[4] B.Delley, D. E. Ellis, *J. Chem. Phys.* **1982**, *76*, 1949

[5] G. te Velde, E. J. Baerends, *Physical Review B* **1991**, *44*, 7888

[6] B. Delley, *J. Chem. Phys.* **1990**, *92*, 508

[7] J. M. Pérez-Jordá, W. Yang, *Chem. Phys. Lett* **1995**, *241*, 469

[8] V. Rohklin, *J. Comput. Phys.* **1985**, *60*, 187

[9] L.Greengard, V. Rokhlin, *J. Comput. Phys.* **1987**, *73*, 325

[10] L. Greengard, V. Rohklin, *Comm. Pure Appl. Math.* **1991**, *44*, 419

[11] C. A.White, B. G. Johnson, P. M. W. Gill, M. Head-Gordon, *Chem. Phys. Lett* **1994**, *230*, 8

[12] M. C. Strain, G. E. Scuseria, M. J. Frisch, *Science* **1996**, 51

[13] K. N. Kudin, G. E. Scuseria, *Chem. Phys. Lett* **1998**, *283*, 61

[14] R. Kutteh, E. Aprà, J. Nichols, *Chem. Phys. Lett* **1995**, *238*, 173

[15] R. D. Adamson, J. P. Dombroski, P. M. W. Gill, *Chem. Phys. Lett* **1996**, *254*, 329

[16] J. P. Dombroski, S. W. Taylor, P. M. W. Gill, *J. Phys. Chem.* **1996**, *100*, 6272

[17] P. M. W. Gill, *Chem. Phys. Lett* **1997**, *270*, 193

[18] S. Goedecker, O. V. Ivanov, *Sol. Stat. Comm.* **1998**, *105*, 665

[19] J. M. Pérez-Jordá, W. Yang, *J. Chem. Phys.* **1997**, *107*, 1218

[20] M. Challacombe, E. Scghwegler, J. Almlöf, *J. Chem. Phys.* **1996**, *104*, 4685

[21] M. Challacombe, E. Schwegler, *J. Chem. Phys.* **1997**, *106*, 5526

[22] R. E. Stratman, G. E. Scuseria, M. J. Frisch, *Chem. Phys. Lett* **1996**, *257*, 213

[23] X. P. Li, R. W. Nunes, D. Vanderbilt, *Phys. Rev. B* **1993**, *47*, 1089

[24] G. Galli, M. Parrinello, *Phys. Rev. Lett.* **1992**, *69*, 3547

[25] E. Hernández, M. J. Gillan, *Phys. Rev. B* **1995**, *51*, 10157

[26] E. Hernández, M. J. Gillan, C. M. Goringe, *Phys. Rev. B* **1996**, *53*, 7147

[27] P. Ordejón, D. A. Drabold, R. M. Martin, M. P. Grumbach, *Phys. Rev. B* **1995**, *51*, 1456

[28] D. Sánchez-Portal, P. Ordejón, E. Artacho, J. M. Soler, *Int. J. Quantum Chem.* **1997**, *65*, 453

[29] J. P. Stewart, *Int. J. Quantum Chem.* **1996**, *58*, 133

[30] C. H. Xu, G. E. Scuseria, *Chem. Phys. Lett* **1996**, *262*, 219

[31] J. M. Millam, G. E. Scuseria, *J. Chem. Phys.* **1997**, *106*, 5569

[32] A. D. Daniels, J. M. Millam, G. E. Scuseria, *J. Chem. Phys.* **1997**, *107*, 425

[33] B. I. Dunlap, J. W. D. Connolly, J. R. Sabin, *J. Chem. Phys.* **1979**, *71*, 3396

[34] H. Sambe, R. H. Felton, *J. Chem. Phys.* **1975**, *62*, 1122

[35] R. T. Gallant, A. St-Amant, *Chem. Phys. Lett* **1996**, *256*, 569

[36] C. Fonseca Guerra, O. Visser, J. G. Snijders, G. te Velde, E. J. Baerends, in *Methods and Techniques for Computational Chemistry*, (Eds.: E. Clementi, G. Corongiu), STEF, Cagliari **1995**, p. 305-395

[37] O. Treutler, R. Ahlrichs, *J. Chem. Phys.* **1995**, *102*, 346

[38] M. J. Frisch, G. W. Trucks, H. B. Schlegel, P. M. W. Gill, B. G. Johnson, M. A. Robb, J. R. Cheeseman, T. Keith, G. A. Petersson, J. A. Montgomery, K. Raghavachari, M. A. Al-Laham, V. G. Zakrzewski, J. V. Ortiz, J. B. Foresman, J. Cioslowski, B. B. Stefanov, A. Nanayakkara, M. Challacombe, C. Y. Peng, P. Y. Ayala, W. Chen, M. W. Wong, J. L. Andres, E. S. Replogle, G. Gomperts, R. L. Martin, D. J. Fox, J. S. Binkley, D. J. Defrees, J. Baker, J. P. Stewart, M. Head-Gordon, C. Gonzalez, J. A. Pople, *GAUSSIAN 94*; Gaussian Inc.: Pittsburg PA, **1995**

[39] P. M. Boerrigter, G. te Velde, E. J. Baerends, *Int. J. Quantum Chem.* **1988**, *33*, 87.

[40] G. te Velde, E. J. Baerends, *J. Comput. Phys.* **1992**, *99*, 84

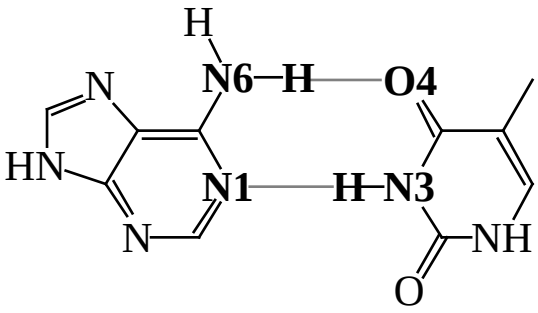
[41] W. H. Press, B. P. Flannery, S. A. Teukolsky, W. T. Vetterling, *Numerical Recipes*; Cambridge University Press: Cambridge, **1986**

Chapter 4

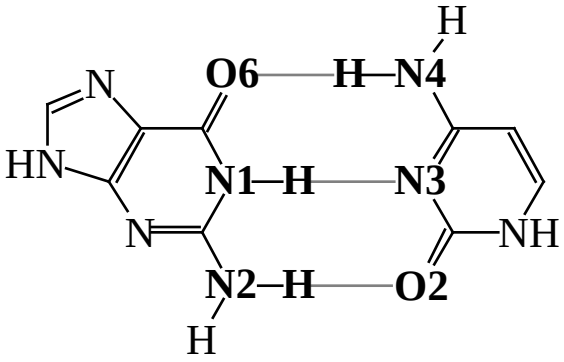
Charge Transfer and Environment Effects Responsible for Characteristics of DNA Base Pairing

A hitherto unresolved discrepancy between theory and experiment is unraveled. Charge transfer and the influence of the environment in the crystal are vital for understanding the nature and for reproducing the structure of hydrogen bonds in DNA base pairs. The introduction of water molecules and a sodium counter ion into the theoretical model deforms the geometry of AT and GC in such a way that excellent agreement with the experimental structures is obtained.

Although it is one of the weakest chemical interactions, the hydrogen bond plays a key role in the chemistry of life, being involved, amongst others, in various types of self-organization and molecular recognition. A point in case is the hydrogen bonds in Watson-Crick base pairs, i.e. adenine-thymine (AT) and guanine-cytosine (GC), that hold together the two helical chains of nucleotides in DNA and form the basis of the genetic code. These hydrogen bonds are commonly believed to be predominantly electrostatic phenomena that, as suggested by Gilli et al., are substantially reinforced by resonance in the π-electron system which makes the proton-acceptor atom more negative and the proton-donor atom more positive, the so-called resonance-assisted hydrogen bonding (RAHB). [1]



AT (**1**)



GC (**2**)

In the present communication, we provide evidence from quantum chemical analyses that challenges this picture and emphasizes the importance of the charge-transfer nature of and environment effects on the hydrogen bonds in DNA base pairs. This has led us to the solution of a hitherto unresolved and significant discrepancy between experimental [2] and theoretical [3] values for distances between the proton-donor and proton-acceptor atoms in AT and GC base pairs. Our evidence is based on a thorough nonlocal density functional theoretical (DFT) investigation with the ADF program (at BP86/TZ2P) of various AT and GC model systems.[4,5]

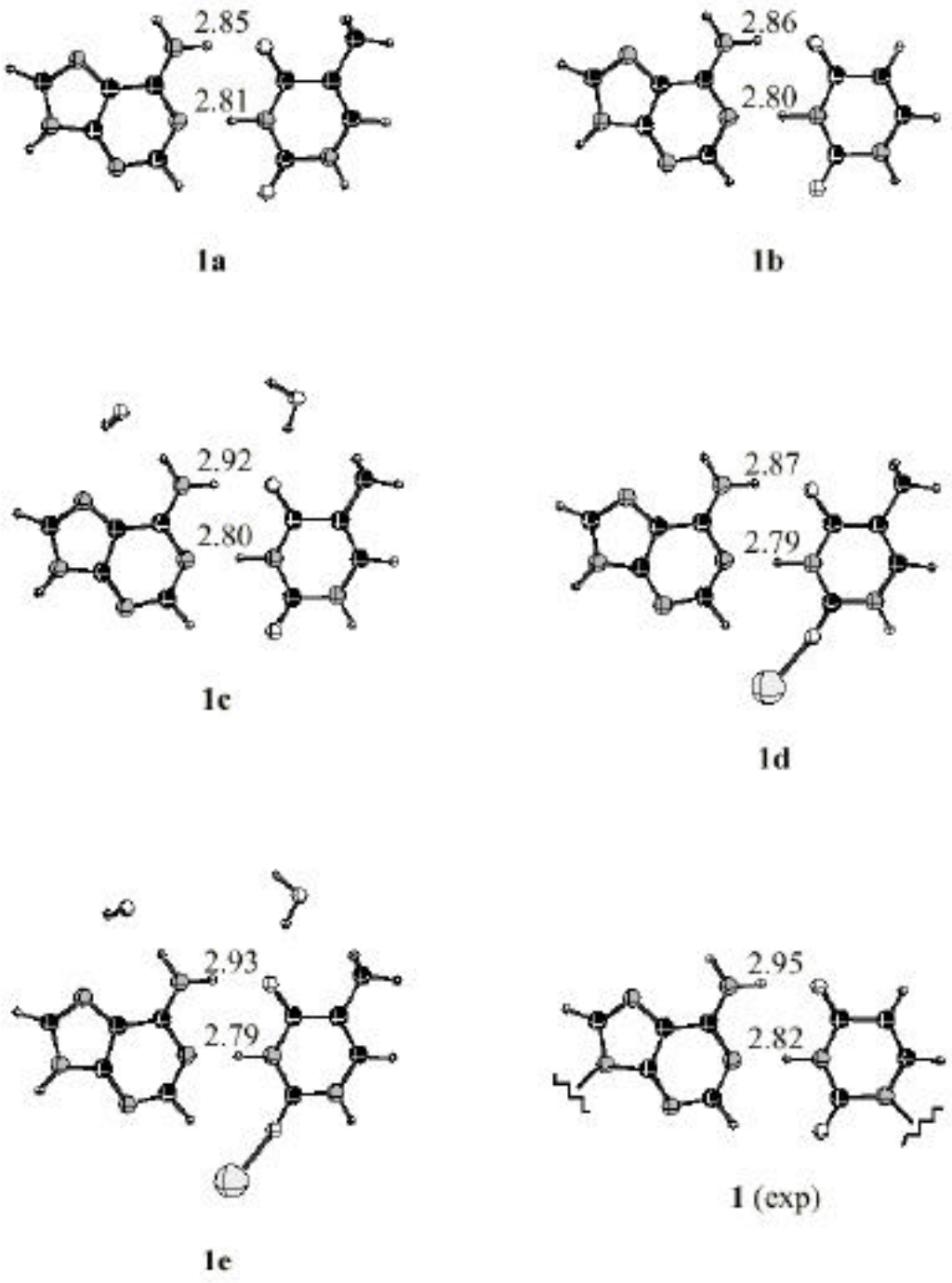


Figure 4.1. N6-O4 and N1-N3 distances in adenine-thymine (**1a**), adenine-uracil (**1b**) and various AT model systems containing water molecules and/or a sodium ion (**1c-e**) from BP86/TZ2P computations and in the crystal of sodium adenylyl-3',5'-uridine hexahydrate from X-ray diffraction (**1**). [2b]

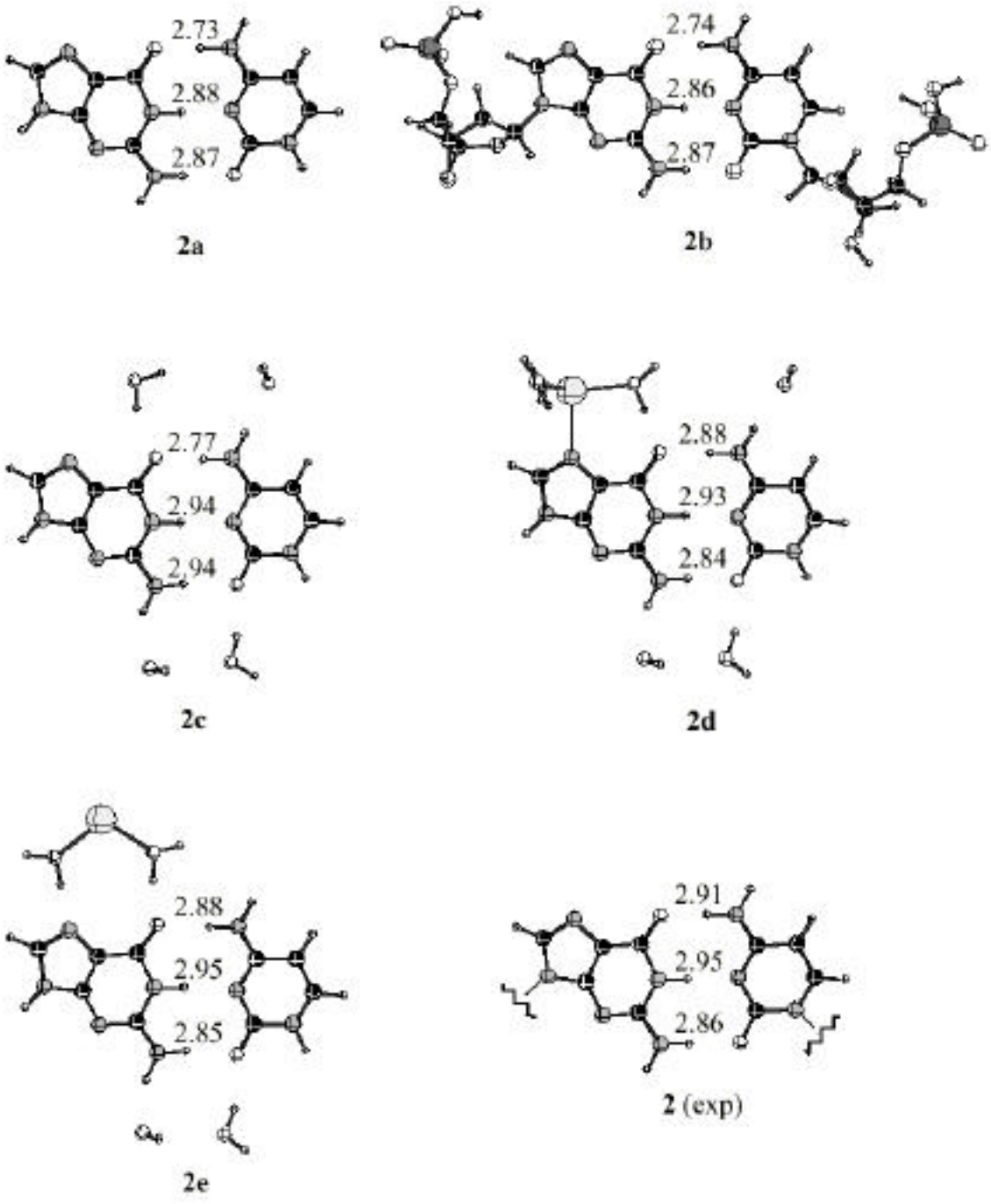


Figure 4.2. O6-N4, N1-N3 and N2-O2 distances in guanine-cytosine (**2a**), GC with inclusion of the backbone (**2b**) and various GC model systems containing water molecules and/or a sodium ion (**2c-e**) from BP86/TZ2P (**2a,2c-e**) and BP86/DZP (**2b**) computations and in the crystal of sodium guanylyl-3',5'-cytidine nonahydrate from X-ray diffraction (**2**).[2c]

Whereas our base-pairing enthalpies (298 K, BSSE corrected) of –11.8 and –23.8 kcal mol^{-1} for AT and GC are in excellent agreement with gas-phase experimental values (–12.1 and –21.0 kcalmol^{-1}),[6] we arrive still at the same striking discrepancies with experimental (X-ray crystal) structures[2] that were encountered before in conventional *ab initio* (HF) and hybrid DFT (B3LYP) studies.[3] As shown in Figure 4.1, we find N6-O4 and N1-N3 hydrogen-bond distances in AT of 2.85 and 2.81 Å (**1a**) that are essentially equal to those in AU (**1b**). These values have to be compared with 2.95 and 2.82 Å from experiment (**1**). Even more eye-catching, as can be seen in Figure 4.2, is the situation for the three hydrogen bonds in GC, i.e. O6–N4, N1-N3 and N2-O2, for which we find a bond length pattern that is short-long-long (2.73, 2.88 and 2.87 Å, **2a**) at significant variance with the experimental values which are long-long-short (2.91, 2.95 and 2.86 Å, **2**). We have verified that these inconsistencies are not induced by our neglecting the glycosidic N–C bond. Methylation of the bases at N9 (adenine, guanine) or N1 (thymine, cytosine), for example, which is a way to mimic the glycosidic N–C bond, has basically no effect on the hydrogen bonds in AT and GC base pairs: hydrogen bond energies (zero K, no BSSE correction) differ by 0.0 and 0.3 kcal mol^{-1}, respectively, and the largest change in hydrogen bond distances amounts to 0.01 Å at BP86/TZ2P (not shown in Figure). Likewise, the hydrogen bond distances of the GC pair consisting of *nucleotides* (**2b**, Figure 4.2) differ only slightly from those in the plain GC pair (**2a**), i.e. by 0.02 Å or less at BP86/DZP.

To trace the origin of the discrepancy between quantum chemical and experimental structures, we have analyzed the A–T and G–C interactions of C_s-symmetric base pairs (whose hydrogen bonds differ by less than 0.005 Å and 0.1 kcalmol^{-1} from those of the C1 symmetric **1a** and **2a**) in the conceptual framework provided by the Kohn-Sham molecular orbital (KS-MO) model through a decomposition of the actual interaction energy (ΔE_{int}) into the classical electrostatic interaction (ΔV_{elst}), the attractive orbital interactions comprising charge transfer and polarization (ΔE_{oi}) and the Pauli repulsive orbital interactions between closed shells (ΔE_{Pauli})[5] It appears that in both DNA base pairs, AT and GC, the bonding orbital interactions associated with hydrogen bonding are of comparable magnitude as the electrostatic attraction (for AT, ΔE_{oi} and ΔV_{elst} are -22.4 and -32.1 kcal mol^{-1}, and for GC, -34.1 and -48.6 kcal mol^{-1}). A more detailed examination of ΔE_{oi} and the associated changes in the wavefunction (i.e. orbital mixings) shows that the hydrogen-bonding orbital interactions are predominantly provided by charge-transfer interactions

in the σ-electron system between a lone pair on nitrogen or oxygen on one base and the N–H σ* acceptor orbitals of the other base. Figure 4.3 shows the relevant frontier-orbital interactions for AT that emerge from our Kohn-Sham MO analysis (a similar diagram can be drawn for GC). Indeed, we do find orbital interactions in the π-electron system that are reminiscent of the RAHB model: by polarization of the π-charge distribution *within* a DNA base, they compensate the build-up of charge caused by charge-transfer hydrogen bonding in the σ-electron system. The corresponding π interaction amounts to –1.7 and –4.8 kcal mol^{-1} for AT and GC, that is, only 3 and 6 % of the total attractive interactions $\Delta E_{oi} + \Delta V_{elst}$. In that respect, π assistance is of minor importance. On the other hand, the π interactions contribute 14 - 20 % of the net bond enthalpies and, on the very shallow potential energy surface, they are able to bring about a shortening of the hydrogen bonds by 0.1 Å. In this sense, one may speak of a certain π assistance, in spite of this term being rather small compared to the charge-transfer and electrostatic interactions.

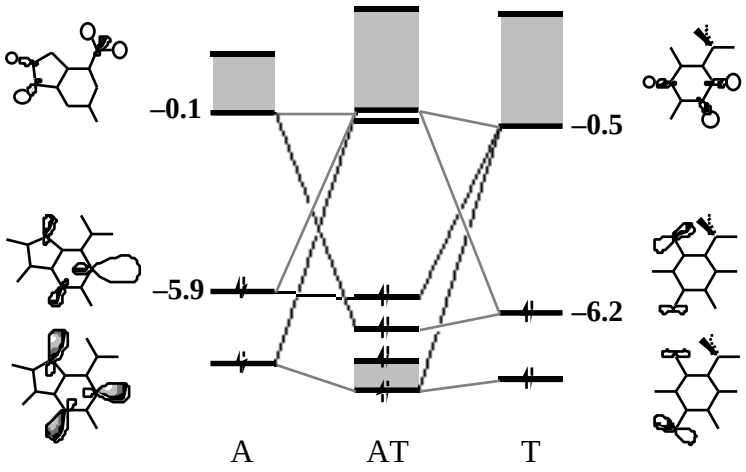


Figure 4.3. Frontier orbital interactions between adenine and thymine in AT (**1a**) from BP86/TZ2P Kohn-Sham DFT analyses (base HOMO and LUMO energies in eV). The group of lowest unoccupied orbitals involved is represented by a block.

Additional support for charge transfer comes from an analysis of the deformation density $\rho_{pair}(\mathbf{r}) - \rho_{base1}(\mathbf{r}) - \rho_{base2}(\mathbf{r})$, i.e. the redistribution of charge density that is caused by the formation of a DNA base pair from its constituting bases. This can be quantified using an extension of the Voronoi deformation density (VDD) method,[7] in which the change in atomic charges associated with base pairing, ΔQ_A, is defined, and related to the deformation density, by Equation 1:

$$\Delta Q_A^{VDD} = - \int_{\substack{\text{Voronoi cell} \\ \text{of A in pair}}} \left[\rho_{pair}(\mathbf{r}) - \rho_{base1}(\mathbf{r}) - \rho_{base2}(\mathbf{r})\right] d\mathbf{r} \qquad (4.1)$$

The interpretation of VDD atomic charges is rather straightforward. Instead of measuring the amount of charge associated with a particular atom A, they directly monitor how much charge flows, due to chemical interactions between DNA bases, out of ($\Delta Q_A > 0$) or into ($\Delta Q_A < 0$) the Voronoi cell of atom A, that is, the region of space that is closer to nucleus A than to any other nucleus. We have verified that all VDD values are stable with respect to variations of the basis by computing the VDD charges at BP86 with three different Slater-type basis sets, i.e. DZ (unpolarized double-ζ), DZP (singly polarized double-ζ), and TZ2P (doubly polarized triple-ζ): the maximum deviation is 0.01 electron. This affects neither the physical picture nor our conclusions. Full details of our analyses are presented in the next chapter.

In AT, the donor-acceptor interaction associated with hydrogen bond N1•••H-N3 (involving two donating adenine orbitals with N1 lone-pair character) leads to a net transfer of 0.05 electrons from A to T (computed with σ virtuals on T only and all other virtuals removed). This is counter-acted by a transfer of 0.04 electrons back from T to A due to hydrogen bond N6-H•••O4 (involving one donating thymine orbital with O4 lone-pair character), leading to a slight build-up of negative charge on thymine. Likewise, for the GC pair, we find a build-up of negative charge on guanine that stems from a transfer of 0.05 electrons from G (with one lone-pair donating atom, O6) to C via hydrogen bond O6•••H-N4 that is outweighed by the transfer of 0.07 electrons back from C (with two donating atoms, N3 and O2) to G via the two hydrogen bonds N1-H•••N3 and N2-H•••O2.

The simultaneous occurrence of hydrogen bonds that donate charge in opposite directions reduces the net build-up of charge on the DNA bases, and one might thus expect that these

hydrogen bonds reinforce each other. However, the corresponding analysis of the bond energy shows that there is no such synergism. The deformation density (not shown here) reveals that this is so because each hydrogen bond creates its own local charge separation which is not affected by the occurrence of another hydrogen bond.

It appears from a breakdown of the VDD charges into contributions from the σ- end π-electron system ($\Delta Q_A = \Delta Q_A^\sigma + \Delta Q_A^\pi$) that virtually all charge transfer occurs in σ symmetry. The ΔQ_A^π values reveal that the π system of each individual DNA base polarizes in such a way that the accumulation of positive or negative charge around the donating or accepting atoms, caused by the charge transfer in the σ-electron system, is counteracted and partly relieved. However, we do not find any synergism between ΔE_σ and ΔE_π, i.e. ΔE_σ is not increased by the occurrence of the π-electron polarization.

The extremely shallow potential energy surface that we find for the hydrogen bonds in DNA base pairs makes it plausible that, in the crystal or under physiological conditions, the structure of AT (1) and GC (2) is significantly influenced by interactions with the environment, such as, hydration, coordination of alkali metal ions (Na+), and hydrogen bonding to hydroxyl groups of the sugar. Therefore, we have tried to simulate the major environment effects that occur in the crystals used in the experimental X-ray structure determinations of AT and GC, i.e. the crystal of sodium adenylyl-3',5'-uridine (ApU) hexahydrate (**1**)[2b] and that of sodium guanylyl-3',5'-cytidine (GpC) nonahydrate (**2**).[2c] And indeed, as can be seen from Figures 4.1 and 4.2, the addition of water molecules (simulating a part of the hydration and hydrogen bonding with ribose OH groups) and the introduction of the sodium counter ions along **1a-e** and **2a-e** deforms the geometry of AT and GC in such a way that excellent agreement between our DFT (e.g. **1e**: N6-O4 and N1-N3 are 2.93 and 2.79 Å; **2e**: O6-N4, N1-N3 and N2-O2 are 2.88, 2.95 and 2.85 Å) and the experimental structures is obtained. Our results show that present-day approximate DFT not only provides a highly efficient but, if appropriate model systems are chosen, also a suitable and accurate alley towards describing and understanding hydrogen bonding in DNA base pairs.

References

[1] a) G. Gilli, F. Bellucci, V. Ferretti, V. Bertolasi, *J. Am. Chem. Soc.* **1989**, *111*, 1023

b) H. Umeyama, K. Morokuma, *J. Am. Chem. Soc.* **1977**, *99*, 131

c) G. A. Jeffrey, W. Saenger, *Hydrogen Bonding in Biological Structures*, Springer-Verlag, Berlin, New York, Heidelberg, **1991**, p. 37

d) G. A. Jeffrey, *An Introduction to Hydrogen bonding*, Oxford University Press, New York, Oxford **1997**, p. 103

[2] a) W. Saenger, *Principles of Nucleic Acid Structure*, Springer-Verlag, New York, **1984**;

b) N. C. Seeman, J. M. Rosenberg, F. L. Suddath, J. J. P. Kim, A. Rich, *J. Mol. Biol.* **1976**, *104*, 109

c) J. M. Rosenberg, N. C. Seeman, R. O. Day, A. Rich, *J. Mol. Biol.* **1976**, *104*, 145

[3] a) J. Sponer, J. Leszczynski, P. Hobza, *J. Phys. Chem.* **1996**, *100*, 1965

b) K. Brameld, S. Dasgupta, W. A. Goddard III, *J. Phys. Chem. B* **1997**, *101*, 4851

c) J. Bertran, A. Oliva, L. Rodríguez-Santiago, M. Sodupe, *J. Am. Chem. Soc.* **1998**, *120*, 8159

[4] a) C. Fonseca Guerra, O. Visser, J. G. Snijders, G. te Velde, E. J. Baerends, in *Methods and Techniques for Computational Chemistry*, (Eds.: E. Clementi, G. Corongiu), STEF, Cagliari, **1995**, p. 305

b) A. Becke, *Phys. Rev. A* **1988**, *38*, 3098

c) J. P. Perdew, *Phys. Rev. B* **1986**, *33*, 8822 (Erratum: *ibid.* **1986**, *34* , 7406).

[5] a) F. M. Bickelhaupt, N. M. M. Nibbering, E. M. van Wezenbeek, E. J. Baerends, *J. Phys. Chem.* **1992**, *96*, 4864

b) T. Ziegler, A. Rauk, *Theor. Chim. Acta* **1977**, *46*, 1

[6] Mass spectrometric data from I. K. Yanson, A. B. Teplitsky, L. F. Sukhodub, *Biopolymers* **1979**, *18*, 1149 with corrections proposed in Ref. 3b.

[7] F. M. Bickelhaupt, N. J. R. van Eikema Hommes, C. Fonseca Guerra, E. J. Baerends, *Organometallics* **1996**, *15*, 2923

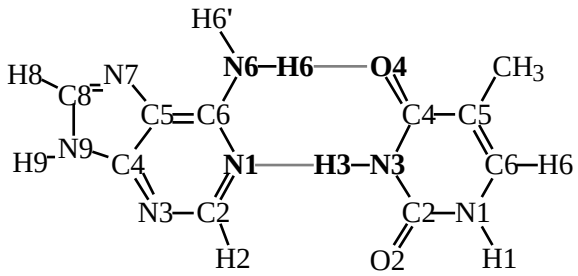
Chapter 5

The Nature of the Hydrogen Bond in DNA Base Pairs: the Role of Charge Transfer and Resonance Assistance

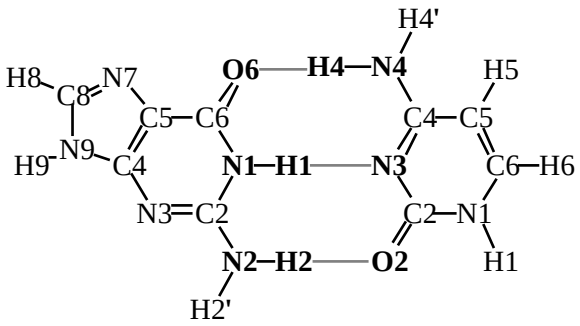
The view that the hydrogen bonds in the Watson-Crick adenine–thymine (AT) and guanine–cytosine (GC) base pairs are in essence electrostatic interactions with substantial resonance assistance from the π electrons is questioned. Our investigation is based on a state-of-the-art density functional theoretical (DFT) approach (BP86/TZ2P) which has been shown to properly reproduce experimental data. Through a quantitative decomposition of the hydrogen bond energy into its various physical terms, we show that, at variance with widespread belief, donor–acceptor orbital interactions (i.e. charge transfer) in σ symmetry between N or O lone pairs of one base and N–H σ* acceptor orbitals on the other base do provide a substantial bonding contribution which is, in fact, of the same order of magnitude as the electrostatic interaction term. The overall orbital interactions are reinforced by a small π component, stemming from polarization in the π-electron system of the individual bases. This π component is, however, one order of magnitude smaller than the σ term. Furthermore, we have investigated the synergism in a base pair between charge-transfer from one base to the other through one hydrogen bond and in the opposite direction through another hydrogen bond, as well as the cooperative effect between the donor–acceptor interactions in the σ- and polarization in the π-electron system. The possibility of C–H•••O hydrogen bonding in AT is also examined. In the course of these analyses, we introduce an extension of the Voronoi deformation density (VDD) method which monitors the redistribution of the σ- and π-electron densities individually out of ($\Delta Q > 0$) or into ($\Delta Q < 0$) the Voronoi cell of an atom upon formation of the base pair from the separate bases.

5.1 Introduction

Although it is the weakest chemical interaction, the hydrogen bond plays a key role in the chemistry of life.[1] Apart from providing water with physical properties that make it the ideal medium for many processes of life to take place in, it is responsible for various types of self-organization and molecular recognition, such as the folding of proteins. As proposed already in 1953 by Watson and Crick,[1c] hydrogen bonds are also essential to the working of the genetic code contained in DNA.[1] The latter consists of two helical chains of nucleotides which are held together by the hydrogen bonds that arise between a purine- and a pyrimidine-derived nucleic base. In particular, this base pairing occurs specifically between adenine (A, a purine) and thymine (T, a pyrimidine), and between guanine (G, a purine) and cytosine (C, a pyrimidine), giving rise to the so-called Watson-Crick AT and GC pairs (see Scheme 1).



Adenine–Thymine



Guanine–Cytosine

Scheme 1

In the past decade, *ab initio* and DFT quantum chemical studies[2] have appeared on the geometry, energy and other aspects of the hydrogen bonds that hold together AT and GC pairs. The adequacy of DFT for hydrogen bonded systems has received much attention lately.[3] It is known from the investigations of Sim et al.[3a] on the water dimer and the formamide-water complex that DFT with nonlocal gradient corrections is capable of describing hydrogen-bonded

systems reasonably well. They found that the DFT results are of comparable quality as those from correlated *ab initio* methods. Others[2e,j-l] have shown that this holds true for the strength of hydrogen bonds in DNA base pairs, too, while for the corresponding structures minor but significant deviations from experimental values were obtained with both DFT and *ab initio* methods. Very recently, we[4,5] have shown that these structural deviations are due to intermolecular interactions of the base pairs with the environment in the crystal. These discrepancies can be resolved if the most important environment effects are incorporated into the model system, yielding DFT structures for DNA base pairs in excellent agreement with experiment.[4,5]

For a true comprehension of the structure, properties and behavior of DNA base pairs, a sound understanding of the hydrogen bonds involved is indispensable. Yet, this nature is not at all clear. The importance, for example, of covalence in these hydrogen bonds, i.e. the magnitude of donor–acceptor orbital interactions, is still unknown. Based on the work of Umeyama and Morokuma[6] on dimers and codimers of HF, H_2O, NH_3 or CH_4, weak and medium range hydrogen bonds are generally believed to be predominantly electrostatic in nature. On the other hand, Gilli et al.[7] suggested that the relatively strong hydrogen bonds in DNA base pairs cannot be understood on the basis of electrostatic interactions only. In their work on β-diketone enols,[7a,c] they ascribed the strong intra- and intermolecular hydrogen bonds found in the corresponding monomers and dimers to a phenomenon, first appreciated by Huggins,[8] that they designated resonance-assisted hydrogen bonding (RAHB): resonance in the π-system assists the hydrogen bond by making the proton-acceptor more negative and the proton-donor more positive. Because of the close similarity between the hydrogen bonding patterns in β-diketone enols (monomers and dimers) and those in DNA base pairs – both involve hydrogen bonds between proton-acceptor and proton-donor atoms that are connected through a conjugated π-system – they suggested that "nature itself may have taken advantage of the greater energy of RAHB to keep control of molecular associations whose stability is essential for life".

In this work, we try to clarify the nature of the hydrogen bonds in the Watson-Crick DNA base pairs using nonlocal density functional theory (DFT). In the conceptual framework provided by Kohn-Sham molecular orbital (KS-MO) theory,[9] we investigate the hydrogen bonding mechanism through an analysis of the electronic structure and a quantitative decomposition of the

bond energy into the electrostatic interaction, the repulsive orbital interactions (Pauli repulsion) and the bonding orbital interactions (charge transfer and polarization). This enables us to address a number of fundamental questions. How important are electrostatics and charge transfer really? And is there a synergism between charge transfer from one base to the other through one hydrogen bond, and in the opposite direction through another hydrogen bond? In other words, does the overall hydrogen bond strength benefit from this mechanism that reduces the net build-up of charge on a base caused by the individual hydrogen bonds? Furthermore, we try to find evidence for the resonance-assisted hydrogen bonding proposed by Gilli et al.[7a] and we test the hypothesis[10] of C–H•••O hydrogen bonding in the AT base pair.

Complementary to the analysis of the orbital electronic structure, we have also studied the electronic density of the DNA bases and, in particular, how this is affected by the formation of the hydrogen bonds in the base pairs. For this purpose, we have developed two extensions to the Voronoi Deformation Density (VDD) method:[11] (i) a scheme for computing changes in the atomic charges of a polyatomic fragment due to the chemical interaction with another fragment, and (ii) a partitioning of these changes in atomic charges into the contributions from different irreducible representations. These new features in VDD enable us to compute the change in σ and π density in the Voronoi cell of a particular atom due to the DNA base-pairing interaction.

5.2 Theoretical Methods

5.2.1 General Procedure

All calculations were performed using the Amsterdam Density Functional (ADF) program[12] developed by Baerends et al.,[12a-d] vectorized by Ravenek[12e] and, parallelized[12a] as well as linearized[12f] by Fonseca Guerra et al.. The numerical integration was performed using the procedure developed by te Velde et al..[12g,h] The MOs were expanded in a large uncontracted set of Slater type orbitals (STOs) containing diffuse functions: TZ2P (no Gaussian functions are involved).[12i] The basis set is of triple-ζ quality for all atoms and has been augmented with two sets of polarization functions, i.e. 3d and 4f on C, N, O, and 2p and 3d on H. The 1s core shell of carbon, nitrogen and oxygen were treated by the frozen-core approximation.[12b] An auxiliary set

of s, p, d, f and g STOs was used to fit the molecular density and to represent the Coulomb and exchange potentials accurately in each self-consistent field cycle.[12j]

Geometries and energies were calculated using nonlocal density-functionals (NL). Equilibrium structures were optimized using analytical gradient techniques.[12k] Frequencies[12l] were calculated by numerical differentiation of the analytical energy gradients using the nonlocal density functionals .

Exchange is described by Slater's Xα potential[12m] with corrections due to Becke[12n,o] added self-consistently and correlation is treated in the Vosko-Wilk-Nusair (VWN) parametrization[12p] with nonlocal corrections due to Perdew[12q] added, again, self-consistently (BP86).[12r]

Bond enthalpies at 298.15 K and 1 atmosphere (ΔH_{298}) were calculated from 0 K electronic bond energies (ΔE) according to equation (5.2.1), assuming an ideal gas.[13]

$$\Delta H_{298} = \Delta E + \Delta E_{\text{trans},298} + \Delta E_{\text{rot},298} + \Delta E_{\text{vib},0} + \Delta(\Delta E_{\text{vib}})_{298} + \Delta(pV) \qquad (5.2.1)$$

Here, $\Delta E_{\text{trans},298}$, $\Delta E_{\text{rot},298}$ and $\Delta E_{\text{vib},0}$ are the differences between products and reactants in translational, rotational and zero point vibrational energy, respectively; $\Delta(\Delta E_{\text{vib}})_{298}$ is the change in the vibrational energy difference as one goes from 0 to 298.15 K. The vibrational energy corrections are based on our frequency calculations. The molar work term $\Delta(pV)$ is $(\Delta n)RT$; $\Delta n = -1$ for two fragments combining to one molecule. Thermal corrections for the electronic energy are neglected. The basis set superposition error (BSSE), associated with the hydrogen bond energy, has been computed via the counterpoise method,[14] using the individual bases as fragments.

5.2.2 Bonding Energy Analysis

The bonding in the AT and GC systems was analyzed using the extended transition state (ETS) method developed by Ziegler and Rauk.[15] The overall bond energy ΔE is made up of two major components (eq 5.2.2).

$$\Delta E = \Delta E_{\text{prep}} + \Delta E_{\text{int}} \qquad (5.2.2)$$

In this formula the preparation energy ΔE_{prep} is the amount of energy required to deform the

separate bases from their equilibrium structure to the geometry that they acquire in the base pair. The interaction energy ΔE_{int} corresponds to the actual energy change when the prepared bases are combined to form the base pair. The interaction energy is further split up into three physically meaningful terms (eq 5.2.3):

$$\Delta E_{int} = \Delta V_{elstat} + \Delta E_{Pauli} + \Delta E_{oi} \qquad (5.2.3)$$

The term ΔV_{elstat} corresponds to the classical electrostatic interaction between the unperturbed charge distributions of the prepared (i.e. deformed) bases and is usually attractive. The Pauli-repulsion ΔE_{Pauli} comprises the destabilizing interactions between occupied orbitals and is responsible for the steric repulsion. The orbital interaction ΔE_{oi} accounts for charge transfer (interaction between occupied orbitals on one moiety with unoccupied orbitals of the other, including the HOMO-LUMO interactions) and polarization (empty/occupied orbital mixing on one fragment). It can be decomposed into the contributions from each irreducible representation Γ of the interacting system (eq 5.2.4).[15] In systems with a clear σ, π separation (like our DNA base pairs), this symmetry partitioning proves to be most informative.

$$\Delta E_{oi} = \sum_{\Gamma} \Delta E_{\Gamma} \qquad (5.2.4)$$

5.3 Results and Discussion

5.3.1 Geometry and Hydrogen Bond Strength

The results of our BP86/TZ2P study on the formation of the adenine-thymine and guanine-cytosine complexes are summarized and compared with literature in Tables 5.1 (energies), 5.2 and 5.3 (geometries). Scheme 1 defines the proton donor–acceptor distances used throughout this work. The structures calculated in C_1 point group symmetry, without any symmetry constraints, were confirmed to be energy minima through a vibrational analysis that revealed zero imaginary frequencies. The choice for the BP86 density functional[12n-q] is based on our investigation[5] on the performance of various nonlocal density functionals for these systems which showed that BP86 agrees slightly better with experiment than PW91[16] and BLYP.[12o,17]

Table 5.1. Hydrogen bond energies (in kcal/mol) of AT and GC.[a]

Base pair	ΔE	ΔE_{BSSE}	ΔH_{298}	ΔH_{exp}
AT [b]	-13.0	-12.3	-11.8	-12.1[d]
AT [c]	-13.0	-12.3		
GC [b]	-26.1	-25.2	-23.8	-21.0[d]
GC [c]	-26.1	-25.2		

[a] BP86/TZ2P. ΔE and ΔE_{BSSE} are the zero K bond energy without and with correction for the BSSE, respectively. ΔH_{298} is the 298 bond enthalpy.

[b] Full optimization of base pair and separate bases.

[c] Base pair optimized in C_s symmetry; full optimization of separate bases.

[d] ΔH_{exp}, experimental ΔH from mass spectrometry data[18] with corrections for AT according to Brameld et al.[2i]

Table 5.2. Distances (in Å) between proton-donor and acceptor atoms of AT.[a]

Level of theory	N6(H)•••O4	N1•••(H)N3
BP86/TZ2P [b]	2.85	2.81
BP86/TZ2P [c]	2.85	2.81
HF/6-31G** [c,d]	3.09	2.99
HF/cc-pVTZ(-f) [b,e]	3.06	2.92
B3LYP/6-31G** [c,f]	2.94	2.84

[a] BP86/TZ2P. See Scheme 1.

[b] Full optimization in C_1 symmetry.

[c] Optimized in C_S symmetry.

[d] Sponer et al.[2e]

[e] Brameld et al.[2i]

[f] Bertran et al.[2l]

Table 5.3. Distances (in Å) between proton-donor and acceptor atoms of GC.[a]

Level of theory	N2(H)•••O2	N1(H)•••N3	O6•••(H)N4
BP86/TZ2P [b]	2.87	2.88	2.73
BP86/TZ2P [c]	2.87	2.88	2.73
HF/6-31G** [c,d]	3.02	3.04	2.92
HF/cc-pVTZ(-f) [b,e]	2.92	2.95	2.83
B3LYP/6-31G** [c,f]	2.92	2.93	2.79

[a] BP86/TZ2P. See Scheme 1.
[b] Full optimization in C_1 symmetry.
[c] Optimized in C_S symmetry.
[d] Sponer et al.[2e]
[e] Brameld et al.[2i]
[f] Bertran et al.[2l]

The computed BP86/TZ2P bond enthalpies for the AT and GC pairs of –11.8 and –23.8 kcal/mol agree well with the experimental results of –12.1 and –21.0 kcal/mol,[18] deviating by as little as +0.3 and –2.8 kcal/mol, respectively (see Table 5.1). The basis set superposition error (BSSE) of some 0.7 kcal/mol is quite small. An important point is that there is essentially no difference between both geometries and bond energies associated with DNA bases and base pairs optimized in C_S symmetry, and those obtained in C_1 symmetry, i.e. without any symmetry restrictions. The various hydrogen bond lengths in AT and GC, i.e. the distances between the proton-donor and proton-acceptor atoms, differ by less than 0.01 Å (see Tables 5.2 and 5.3). Likewise, the formation of C_S-symmetric base pairs (again from fully optimized bases) yields bond energies ΔE that differ by less than 0.1 kcal/mol from those for the same process without symmetry constraint (see Table 5.1). As a consequence, we may analyse the A–T and G–C bonding mechanisms in C_S symmetry, enabling us to decompose the orbital interactions into a σ- and a π component (eq 5.2.4).

As mentioned in our communication[4] and further investigated in Ref. [5], gas-phase theoretical geometries can not be directly compared with experimental X-ray crystal structures[1d,19] that are subjected to and influenced by packing forces as well as intermolecular

interactions. Therefore, in the present study, we restrict ourselves to a brief comparison between our results and those from a few other theoretical studies (for an exhaustive comparison with other theoretical[2d,e,i,k,l] and experimental[1d,19] studies, we refer to Ref. [5]). The Hartree-Fock approach (HF/6-31G**)[9,13] yields distances that are up to 0.2 Å longer than our BP86/TZ2P values. The agreement between our distances and those obtained by Bertran et al.[2l] at B3LYP/6-31G** is better, the latter being only up to 0.1 Å longer than ours. The remaining variance is probably not only due to the different functionals but also the different basis sets as well as technical differences between the programs used.

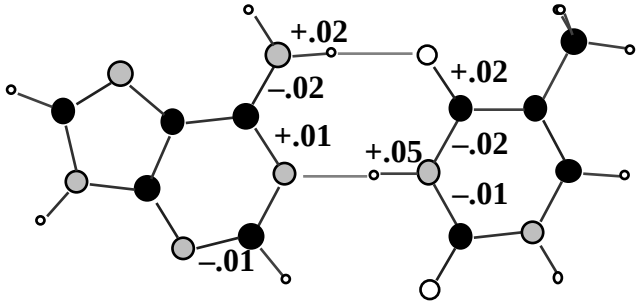


Figure 5.1. Deformation (in Å) of the individual bases caused by hydrogen bonding in the base pairs, from BP86/TZ2P optimizations without any symmetry constraint (only changes in bond length ≥ 0.01Å are given).

The deformation of the bases (i.e. changes in bond lengths larger than 0.1 Å) caused by the formation of the hydrogen bonds is shown in Figure 5.1. All the N–H bonds that participate in hydrogen bonding expand by 0.02 - 0.05 Å. The largest elongations are found for the N3–H3 of thymine (+0.05 Å) and the N4–H4 of cytosine (+0.04 Å). The C=O distances of oxygen atoms involved in hydrogen bonding increase by some 0.02 Å. Furthermore, we see that G–C base pairing leads to somewhat stronger distortions of the corresponding bases than A–T base pairing. In the next section, we will explain how charge-transfer interactions in the σ-system and polarization in the π-system are responsible for these deformations.

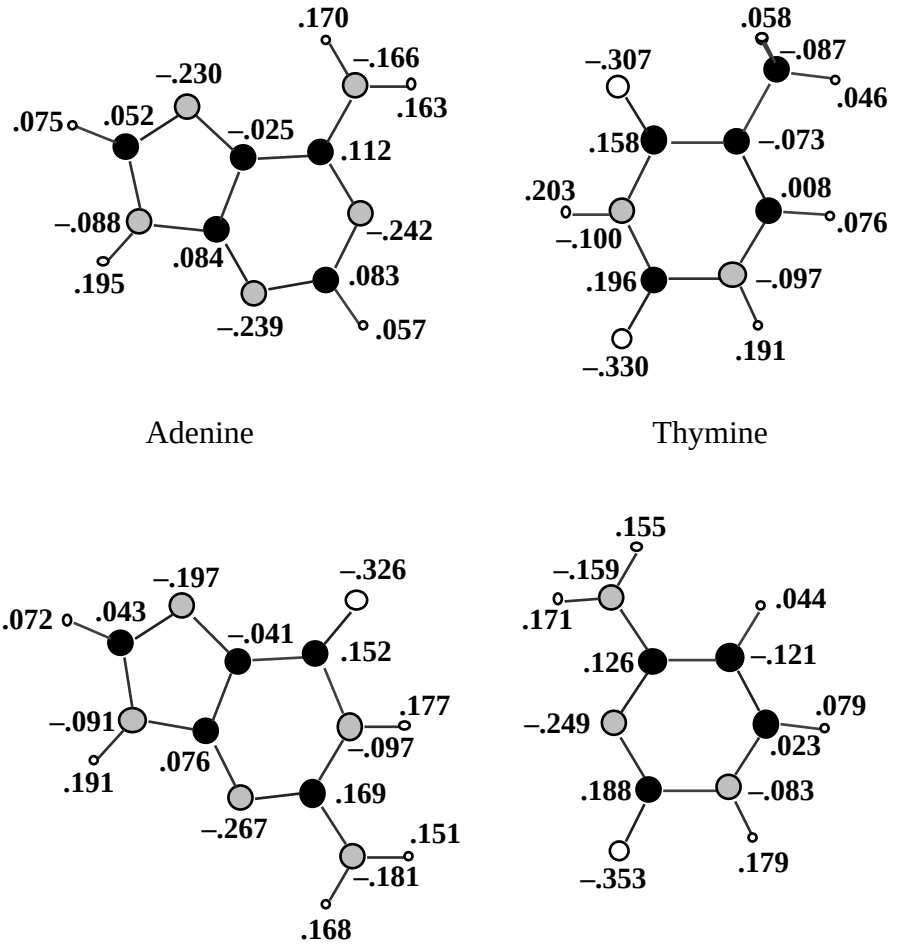


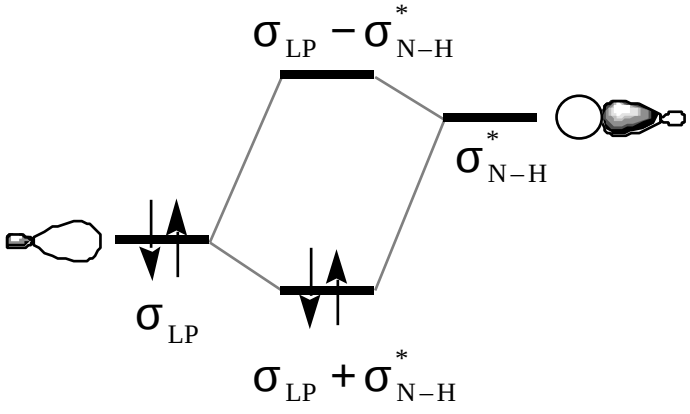
Figure 5.2. VDD atomic charges (in electrons) of the *isolated* bases adenine, thymine, guanine and cytosine obtained at BP86/TZ2P (see Scheme 1).

5.3.2 Nature of the Hydrogen Bond

Electronic Structure of DNA Bases

In order to form stable base pairs, DNA bases must be structurally and electronically complementary. The role of structural complementarity has been discussed very recently by Kool and others.[20] Here, we focus on the electronic structure of the four DNA bases and their capability to form stable A–T and G–C hydrogen bonds. First, we examine if the bases do possess the right charge distribution for achieving a favorable electrostatic interaction in the Watson-Crick base pairs. This turns out to be the case, as can be seen from Figure 5.2, which displays the VDD atomic charges[11] (see also section 5.3.3) for the separate, noninteracting bases: all proton-acceptor atoms have a negative charge whereas the corresponding protons they face are all positively charged.

Next, we consider the possibility of charge-transfer interactions in the σ-electron system. Scheme 2 displays the basic features in the electronic structures that are required in order for these donor–acceptor orbital interactions to occur: a lone pair on a nitrogen or oxygen atom of one base pointing toward (and donating charge into) the unoccupied σ^* orbital an N-H group of the other base; this leads to the formation of a weak $\sigma_{LP} + \sigma^*_{N-H}$ bond.



Scheme 2

Of course, the electronic structure and bonding mechanism in DNA base pairs, with two or three hydrogen-bonding contacts occurring simultaneously, are somewhat more complicated. Not only

the HOMOs and LUMOs of the σ-electron system but also some of the other high-energy occupied and low-energy unoccupied orbitals of the bases are involved in frontier-orbital interactions. However, the basic bonding pattern should still be that of Scheme 2: the occupied orbitals at high energy must have lone-pair character on the charge-donating nitrogen or oxygen atoms and the unoccupied orbitals at low energy must be σ^* antibonding on the charge-accepting N–H group (*vide infra*). Indeed, as can be seen from the contour plots of the DNA-base frontier orbitals in Figures 5.3 and 5.4 – and anticipating the outcome of our orbital-interaction analyses – this turns out to be the case.

We begin with the bases of the AT pair (see Figure 5.3). Adenine has two occupied orbitals, the σ_{HOMO-1} and the σ_{HOMO}, that have lone-pair-like lobes on the nitrogen atoms N1, N3 and N7 (see also Scheme 1). Through their lobe on N1, they can overlap with and donate charge into the lowest unoccupied orbitals of thymine which all have N3–H3 σ^* character (they have also σ^* character on C–H and other N–H groups of thymine but this is of no direct importance for A–T bonding); one of these thymine acceptor orbitals, the σ_{LUMO+1}, is shown in Figure 5.3. Likewise, the σ_{HOMO-1} and σ_{HOMO} of thymine are essentially lone pairs on the oxygen atoms O2 and O4. With their lobe on O4, they can overlap and interact with the complementary N6–H6 σ^* antibonding virtuals on adenine, e.g. the σ_{LUMO} (Figure 5.3).

The situation for the bases of the GC pair is very similar (see Figure 5.4). The σ_{HOMO} of guanine is basically a lone pair on O6 that points toward and can donate charge into the lowest unoccupied orbitals of cytosine that have N4–H4 σ^* antibonding character, e.g. the σ_{LUMO} (Figure 5.4). The σ_{HOMO-1}, and σ_{HOMO} of cytosine are a lone pair on N3 and O2, respectively. They can overlap and interact with the lowest unoccupied orbitals on guanine with N1–H1 and N2–H2 σ^* antibonding character. Interestingly, the σ_{LUMO}, the σ_{LUMO+2} (not shown in Figure) and the σ_{LUMO+3} of guanine can be conceived as the totally bonding (plus-plus-plus), the nonbonding (plus-null-minus) and the antibonding (plus-minus-plus) combinations, respectively, of the three N–H σ^* orbitals corresponding to the N2–H2', N2–H2 and N1–H1 groups.

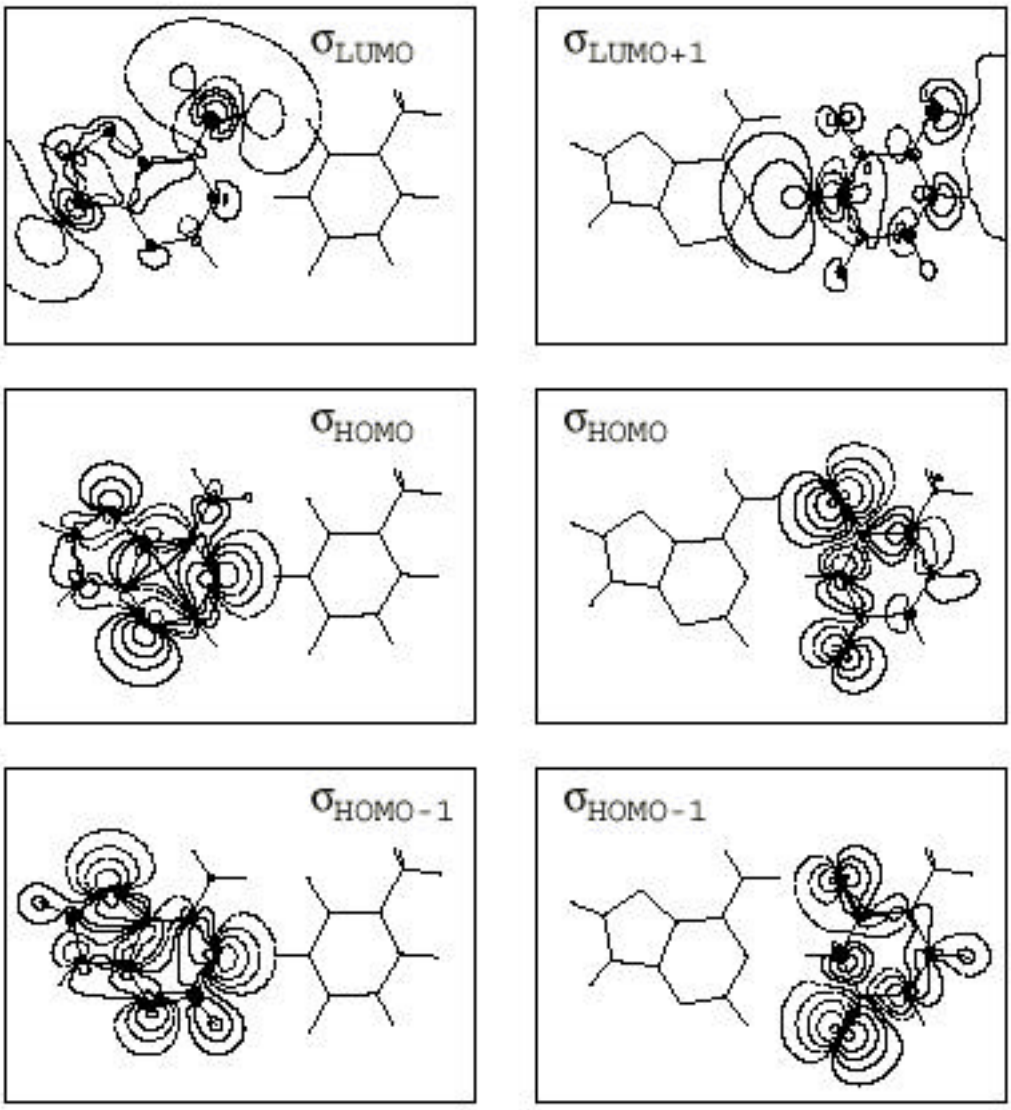


Figure 5.3. Contour plots of the σ_{HOMO-1}, σ_{HOMO} and σ_{LUMO} of adenine and the σ_{HOMO-1}, σ_{HOMO} and σ_{LUMO+1} of thymine obtained at BP86/TZ2P (Scan values: ±0.5, ±0.2, ±0.1, ±0.05, ±0.02. Solid and dashed contours refer to positive and negative values, respectively). For each fragment molecular orbital (FMO), both its own base and the other base in the Watson-Crick pair are shown as wire frames.

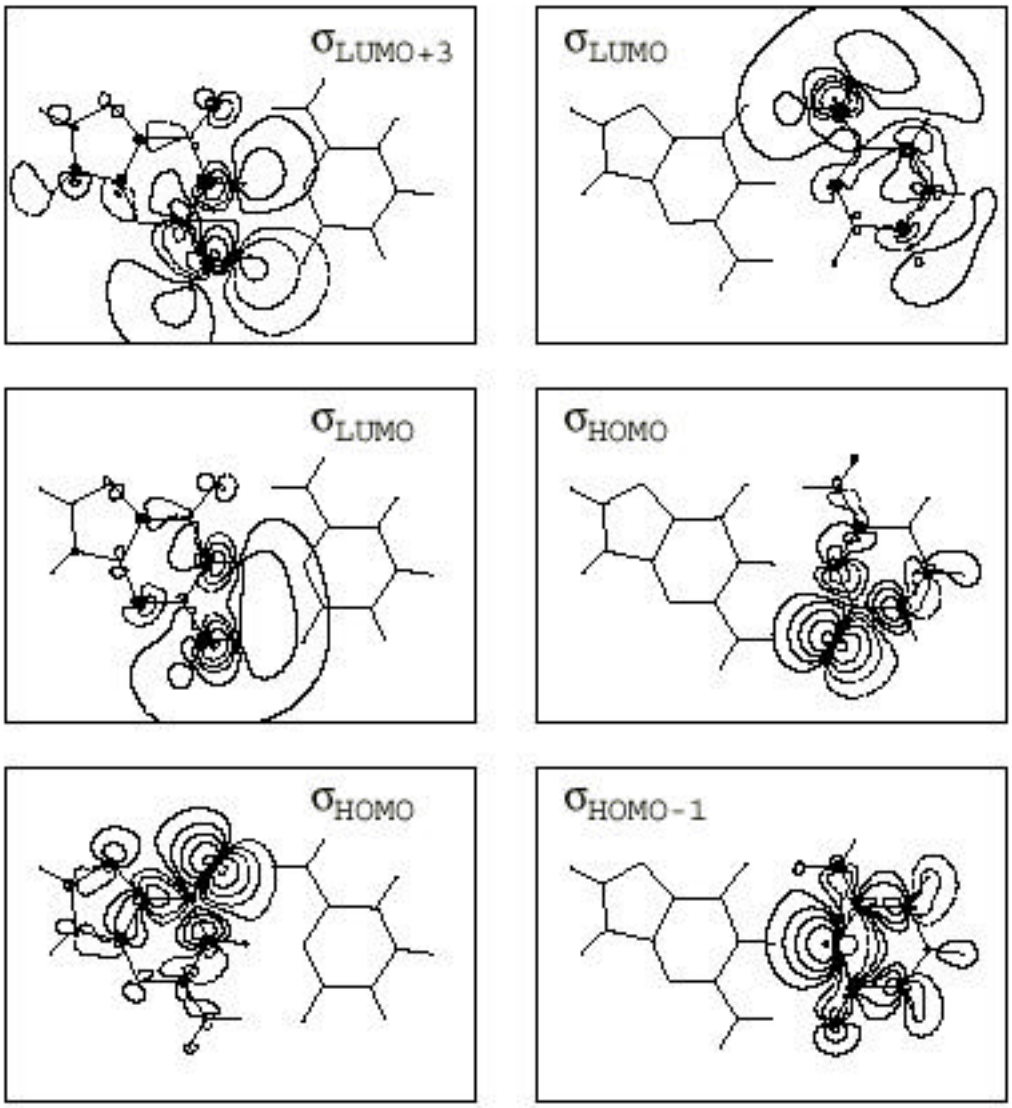


Figure 5.4. Contour plots of the σ_{HOMO}, σ_{LUMO} and σ_{LUMO+3} of guanine and σ_{HOMO-1}, σ_{HOMO} and σ_{LUMO} cytosine obtained at BP86/TZ2P (see also legend to Figure 5.3).

A–T Orbital Interactions

Now, let us analyse how the frontier orbitals of the bases really interact in the Watson-Crick base pairs. Figures 5.5 and 5.6 show schematically the resulting MO diagrams for the σ-electron systems; relevant overlaps between occupied and virtual frontier orbitals are given in Table 5.4. The Kohn-Sham MO analyses of the A–T and G–C base-pairing interactions do indeed yield the bonding mechanism that we expected on the basis of the above qualitative considerations on the character and shape of the DNA-base orbitals. The picture is only complemented by a few repulsive four-electron orbital interactions that we did not consider above.

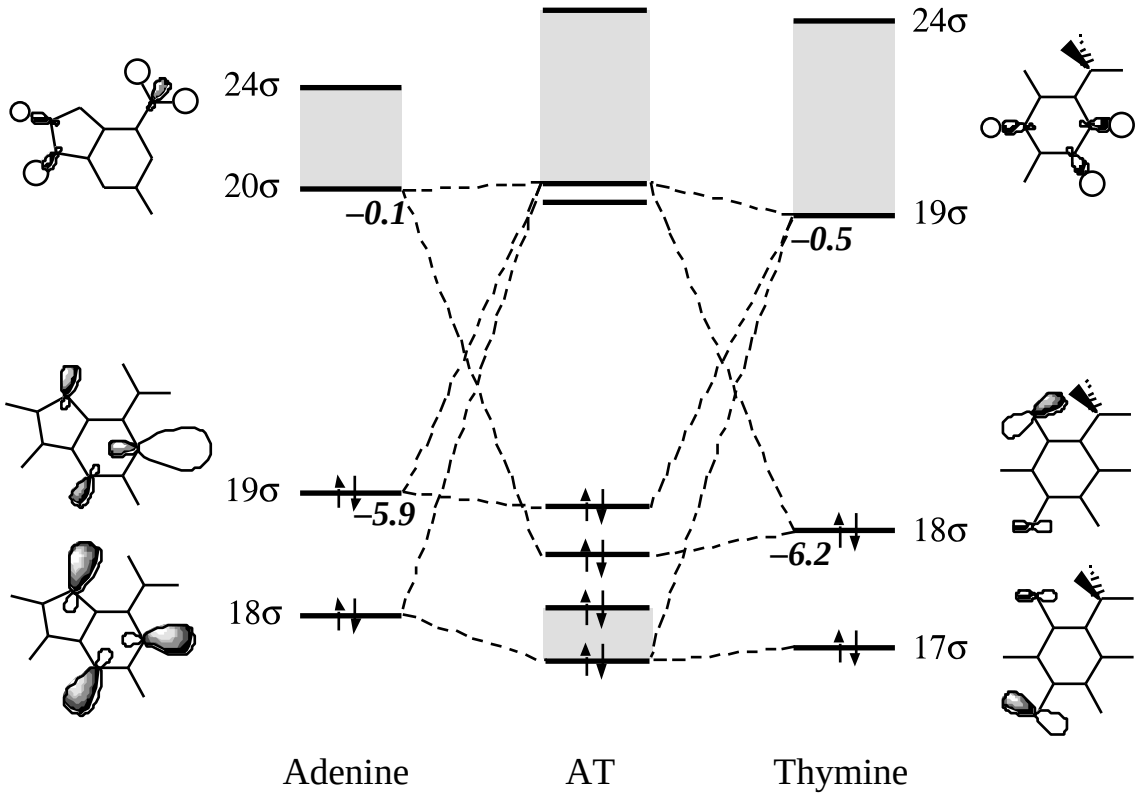


Figure 5.5. Diagram for the donor–acceptor interactions in the N6(H)•••O4 and N1•••(H)N3 hydrogen bonds between adenine and thymine with σ_{HOMO} and σ_{LUMO} energies in eV, obtained at BP86/TZ2P (the lowest unoccupied orbitals that participate in these interactions are represented by a block).

For AT, we find charge-transfer hydrogen bonding from A to T, through N1•••H3–N3, and the other way around from T to A, through N6–H6•••O4. The N1•••H3–N3 bond arises from the donor–acceptor interaction between the two σ_{HOMO-1} and σ_{HOMO} nitrogen lone-pair orbitals of adenine (18σ and 19σ in Figure 5.5) and the lowest unoccupied N3–H3 σ^* orbitals of thymine (19σ through 24σ, represented as a block in Figure 5.5). The N6–H6•••O4 bond, donating charge in the opposite direction, is provided by the interaction between the σ_{HOMO} oxygen lone-pair orbital of thymine (i.e. 18σ) and the lowest unoccupied N6–H6 σ^* orbitals of adenine (i.e. 20σ through 24σ, represented as a block in Figure 5.5). In addition, there is a repulsive orbital interaction between the σ_{HOMO-1} of adenine and the σ_{HOMO-1} of thymine, with a mutual orbital overlap of 0.19, which splits the A–T bonding combination of the adenine 18σ with the thymine 19σ through 24σ into the σ_{HOMO-3} and σ_{HOMO-2} of the AT base pair; this "split orbital level" is represented as a block in the MO diagram. It is the donation of charge into the N–H antibonding σ^* orbitals of adenine and thymine that is responsible for the slight elongation observed for the N–H bonds involved in hydrogen bonding (see Figure 5.1). The adenine σ_{LUMO}, σ_{LUMO+2} and σ_{LUMO+3}, for example, acquire populations of 0.05, 0.03 and 0.03 electrons, respectively (not shown in Table). The thymine σ_{LUMO} through σ_{LUMO+3} and σ_{LUMO+5} each gain 0.02 electrons. Also other deformations that occur upon base pairing are caused by these charge-transfer interactions in the σ-electron system but also by polarization (i.e. occupied–empty mixing of orbitals on the same base) in the π-electron system (*vide infra*).

As follows from the total VDD charges of the individual DNA bases in the Watson-Crick base pairs in Table 5.5, the charge-transfer from A to T associated with the N1•••H3–N3 bond is stronger than that back from T to A through the N6–H6•••O4 bond. This leads to an accumulation of negative charge of –0.03 electrons on thymine. Two factors are responsible for this build-up of charge. In the first place, the N1•••H3–N3 bond comprises two donor orbitals on adenine for charge transfer into virtuals of thymine, whereas only one donor orbital on thymine is involved in the N6–H6•••O4 bond for charge transfer back into virtuals on adenine. Secondly, the overlaps between the donor orbitals of adenine (18σ and 19σ) and the lowest unoccupied acceptor orbitals of thymine (19σ through 24σ) are with values of 0.06 - 0.19 significantly larger than those between the donor orbital of thymine and the acceptor orbitals of adenine that amount to 0.03 - 0.09 (see Table 5.4).

Table 5.4. Overlaps between σ frontier orbitals of DNA bases in AT and GC.[a]

$\langle\sigma_A^{occ}\vert\sigma_T^{virt}\rangle$	$\vert 19\sigma_T\rangle$	$\vert 20\sigma_T\rangle$	$\vert 21\sigma_T\rangle$	$\vert 22\sigma_T\rangle$	$\vert 24\sigma_T\rangle$
$\langle 18\sigma_A\vert$	.06	.16	.12	.08	.06
$\langle 19\sigma_A\vert$	.08	.19	.14	.09	.06

$\langle\sigma_T^{occ}\vert\sigma_A^{virt}\rangle$	$\vert 20\sigma_A\rangle$	$\vert 21\sigma_A\rangle$	$\vert 22\sigma_A\rangle$	$\vert 23\sigma_A\rangle$	$\vert 24\sigma_A\rangle$
$\langle 18\sigma_T\vert$	.08	.07	.04	.03	.09

$\langle\sigma_G^{occ}\vert\sigma_C^{virt}\rangle$	$\vert 17\sigma_C\rangle$	$\vert 18\sigma_C\rangle$	$\vert 19\sigma_C\rangle$	$\vert 20\sigma_C\rangle$	$\vert 21\sigma_C\rangle$
$\langle 20\sigma_G\vert$	.003	.002	.005	.003	.005
$\langle 21\sigma_G\vert$	.08	.06	.09	.08	.04

$\langle\sigma_C^{occ}\vert\sigma_G^{virt}\rangle$	$\vert 22\sigma_G\rangle$	$\vert 24\sigma_G\rangle$	$\vert 25\sigma_G\rangle$	$\vert 27\sigma_G\rangle$
$\langle 15\sigma_C\vert$	.22	.24	.16	.04
$\langle 16\sigma_C\vert$	.15	.03	.09	.02

[a] BP86/TZ2P.

Note that the π-electron density does not contribute to the net A–T charge transfer ($\Delta Q_{total}^{\pi} = 0$, Table 5.5) which is thus entirely due to the σ-orbital interactions. The absence of A–T charge transfer in the π-electron system is due to the extremely small π-orbital overlaps (in the order of 10^{-3}), which are one to two orders of magnitude smaller than those occurring between σ-orbitals. There is however occupied-virtual mixing within the π-system of each individual base. This is ascribed mainly to the electrostatic potential that one base experiences from the other base. This π-polarization is responsible for a sizeable charge reorganization as discussed in the section 5.3.4.

Table 5.5. Total charge transfer (in electrons) between the individual DNA bases in Watson-Crick base pairs calculated with the extensions of the VDD method.[a]

	Adenine	Thymine	Guanine	Cytosine
$\Delta Q_{total}^{\sigma}$	.03	–.03	–.03	.03
ΔQ_{total}^{π}	.00	.00	.00	.00
ΔQ_{total}	.03	–.03	–.03	.03
σ *virtuals on T and C only, no* π *virtuals at all* [b]				
$\Delta Q_{total}^{\sigma}$	.05	–.05	.05	–.05
σ *virtuals on A and G only, no* π *virtuals at all* [c]				
$\Delta Q_{total}^{\sigma}$	–.04	.04	–.07	.07

[a] BP86/TZ2P.

[b] Only charge transfer from A to T and from G to C possible.

[c] Only charge transfer from T to A and from C to G possible.

We have also tried to infer the amount of charge transfer associated with the individual N1•••H3–N3 and N6–H6•••O4 hydrogen bonds by removing either the σ virtuals from thymine (switching off N1•••H3–N3) or from adenine (switching off N6–H6•••O4) while at the same time all π virtuals are removed from both DNA bases (switching off polarization of the π electrons; see also section 5.3.5). The results (entries four and five in Table 5.5) confirm that more charge is transferred from A to T through N1•••H3–N3 (0.05 electrons) than back from T to A via N6–H6•••O4 (0.04 electrons). Note, however, that the difference between the amount of charge transferred in opposite directions through either of the two hydrogen bonds is somewhat smaller without (0.01 electrons, i.e. the *difference* between entries four and five in Table 5.5) than with all other interfering orbital interactions (0.03 electrons, see entry one of Table 5.5).

C–H•••O Hydrogen Bonding in AT ?

Leonard et. al.[10a] suggested that there is also a hydrogen bond between the C2–H2 bond of adenine and the oxygen atom O2 of thymine that would contribute to the stability of the AT pair.

However, our analyses show that this is not the case. In the first place, already the distance between this C–H bond and O atom is to large to be indicative for a hydrogen bonding interaction (C2–O2 = 3.63 Å and H2–O2 = 2.81 Å). But more importantly, we do not find any donor–acceptor orbital interaction corresponding with a C2–H2•••O2 bond. Accordingly, neither the appropriate donor orbital of thymine (the O2 lone-pair orbital of thymine, i.e. σ_{HOMO-1} or 17σ in Figure 5.5) is depopulated nor does the C2–H2 antibonding acceptor orbital of adenine (i.e. the σ_{LUMO+2}; not shown in Figure) acquire any population. In line with this, the C2–H2 bond distance does not expand but remains unchanged. To get a more quantitative idea of the strength of the C2–H2•••O2 interaction, we have analyzed this bond separately from the other bonds, by rotating thymine 180° around an axis through its O2 atom and parallel to the N1–N3 bond (this yields a structure in which both N6–H6•••O4 and N1•••H3–N3 bonds are broken whereas the C2–H2•••O2 moiety is preserved). What we get is a weakly repulsive net interaction energy of only 1.6 kcal/mol, which arises from +1.0 kcal/mol electrostatic repulsion, +1.2 kcal/mol Pauli repulsion and –0.6 kcal/mol bonding orbital interaction. Thus, we must reject the hypothesis of a stabilizing C–H•••O hydrogen bond in AT. This supports Shishkin et al.[10b] who have ruled out C–H•••O hydrogen bonding in AT on the basis of a computed (HF/6-31G*) increase of the C–H stretching frequency of adenine in the base pair.

G–C Orbital Interactions

The MO diagram for GC looks somewhat more complicated than that for AT. This is not the result of a more complicated bonding mechanism but follows simply from the fact that there are now three instead of only two hydrogen bonds. We find for GC one charge-transfer interactions from G to C, via O6•••H4–N4, and two back from C to G, via N1–H1•••N3 and N2–H2•••O2 (see Scheme 1). The O6•••H4–N4 bond is provided by a donor–acceptor interaction between the σ_{HOMO} of guanine, an oxygen O6 lone-pair orbital (21σ in Figure 5.6) and the lowest unoccupied N4–H4 antibonding acceptor orbitals on the amino group of cytosine (17σ through 21σ, represented as a block in Figure 5.6). The resulting bonding combination is split into two levels (i.e. the σ_{HOMO} and σ_{HOMO-1} of the GC pair) due to the admixing of the guanine σ_{HOMO-1} (20σ in Figure 5.6) which does however not contribute to the donor–acceptor interaction.

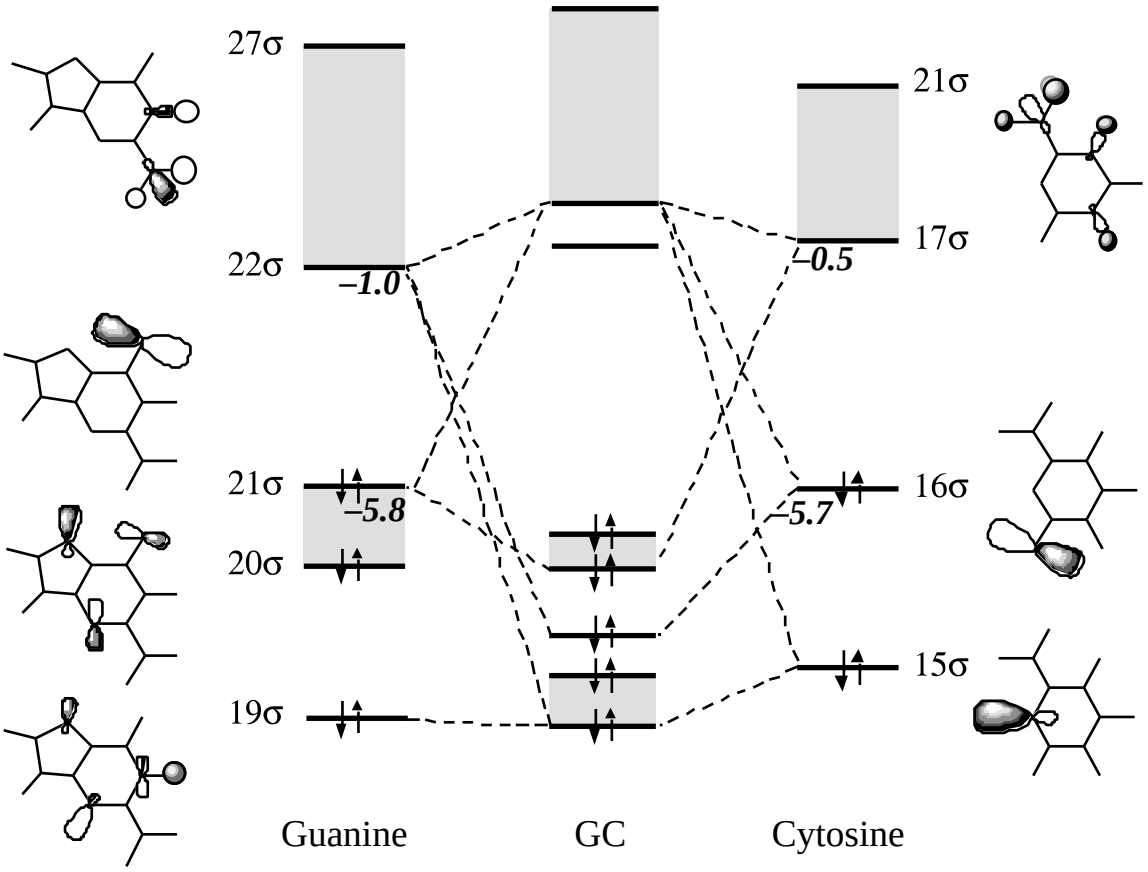


Figure 5.6. Diagram for the donor–acceptor interactions in the O6•••(H)N4, N1(H)•••N3 and N2(H)•••O2 hydrogen bonds between guanine and cytosine with σ_{HOMO} and σ_{LUMO} energies in eV, obtained at BP86/TZ2P (the lowest unoccupied orbitals that participate in these interactions are represented by a block).

The two N1–H1•••N3 and N2–H2•••O2 bonds are provided by the donor–acceptor interactions of the cytosine lone-pair orbitals on oxygen O2 (the σ_{HOMO-1}, i.e. 15σ) and nitrogen N3 (the σ_{HOMO}, i.e. 16σ), respectively, with the lowest unoccupied acceptor orbital of guanine (22σ through 27σ, represented as a block), which are N1–H1 and N2–H2 antibonding (see Figure 5.6). The bonding combination between cytosine σ_{HOMO-1} and guanine virtuals is split into two levels (i.e. σ_{HOMO-4} and σ_{HOMO-3} of the GC pair, indicated as a block in the MO diagram) due to an additional four-electron repulsion that the σ_{HOMO-1} of cytosine (i.e. the 15σ) experiences with the σ_{HOMO-3} of guanine (i.e. the 19σ). The slight elongation of the N–H bonds that

participate in hydrogen bonding (see Figure 5.1) is caused by the donation of charge into the corresponding N–H antibonding σ^* orbitals of guanine and cytosine (e.g. 0.05 and 0.02 electrons, respectively, in the corresponding σ_{LUMO}'s; not given in Table).

The fact that there are two hydrogen bonds donating charge from C to G and only one donating charge from G to C leads to a net accumulation of negative charge on guanine (–0.03 electrons, Table 5.5). Using the same procedure as for AT (*vide supra*), the amount of charge-transfer from G to C associated with the individual O6•••H4–N4 bond is estimated to be 0.05 electrons (entry four in Table 5.5) which is indeed exceeded by the transfer of 0.07 electrons back from C to G caused by the N1–H1•••N3 and N2–H2•••O2 bonds together (entry five in Table 5.5; see also section 5.3.5).

Note that, as for AT, due to very small overlaps (in the order of 10^{-3}), the π-orbital interactions do not contribute to the net G–C charge transfer ($\Delta Q^{\pi}_{total} = 0$ and $\Delta Q_{total} = \Delta Q^{\sigma}_{total}$, see Table 5.5). But, again as for AT, the π-electron systems of guanine and cytosine are significantly polarized (mainly due to the electrostatic potential that the bases experience from each other) leading to a sizeable charge reorganization within each base (see section 5.3.4).

Quantitative Decomposition of the Hydrogen Bond Energy

Now that we know that the DNA bases have suitable charge distributions for electrostatically attracting each other and after having established the occurrence of σ charge transfer and π polarization (see also section 5.3.4), we want to quantitatively assess the importance of the various components of the A–T and G–C base-pairing energy. Thus, we have carried out a bond energy decomposition for the Watson-Crick base pairs for two geometries (see Table 5.6): (i) the equilibrium geometry (AT and GC), and (ii) a geometry derived from the former by freezing the structures of the individual bases and pulling them 0.1 Å apart along an axis parallel to the hydrogen bonds ($AT_{0.1Å}$ and $GC_{0.1Å}$). The latter corresponds to the slightly longer hydrogen bonds observed experimentally in X-ray crystal structure determinations,[1d,19] and its analysis serves to get an idea if the nature of the hydrogen bonds is affected by structural perturbations that may occur in crystals (or under physiological conditions). The orbital interaction is divided into a σ-component and a π-component. ΔE_{σ} consists mainly of the electron donor–acceptor interactions mentioned above. The π-component accounts basically for the polarization in the π-

system (*vide supra*) which turns out to partly compensate the local build-up of charge caused by the charge-transfer interactions in the σ-system (see section 5.3.4).

Table 5.6. Bond energy decomposition for the Watson-Crick Base Pairs (in kcal/mol) in the optimized geometry (AT and GC) and with the base–base distance elongated by 0.1 Å ($AT_{0.1Å}$ and $GC_{0.1Å}$).[a]

	AT	$AT_{0.1Å}$	GC	$GC_{0.1Å}$
Orbital Interaction Decomposition				
ΔE_σ	–20.7	–15.9	–29.3	–22.8
ΔE_π	–1.7	–1.3	–4.8	–3.9
ΔE_{oi}	–22.4	–17.2	–34.1	–26.7
Bond Energy Decomposition				
ΔE_{Pauli}	39.2	28.6	52.1	37.5
ΔV_{elstat}	–32.1	–26.5	–48.6	–41.0
$\Delta E_{Pauli} + \Delta V_{elstat}$	7.1	2.1	3.5	–3.5
ΔE_{oi}	–22.4	–17.2	–34.1	–26.7
ΔE_{int}	–15.3	–15.1	–30.6	–30.2
ΔE_{prep}	2.3		4.1	
ΔE	–13.0		–26.5	

[a] BP86/TZ2P. Bond energies with respect to bases optimized in C_s symmetry.

The striking result of our analysis is that charge-transfer orbital interactions are not at all a negligible or minor component in the hydrogen bond energy of Watson-Crick base pairs (see Table 5.6). Instead, what we find is that charge transfer is of the same order of magnitude as the electrostatic interaction! For AT, ΔE_{oi} is –22.4 kcal/mol and ΔV_{elstat} is –32.1 kcal/mol, and for GC, ΔE_{oi} is –34.1 kcal/mol and ΔV_{elstat} is –48.6 kcal/mol. Interestingly, we see that the electrostatic interaction alone is not capable of providing a net bonding interaction; it can only

partly compensate the Pauli-repulsive orbital interactions ΔE_{Pauli}. Without the bonding orbital interactions, the net interaction energies of AT and GC at their equilibrium structures would be repulsive by 7.1 and 3.5 kcal/mol, respectively (Table 5.6). This parallels the finding of Reed and Weinhold[21] that the water dimer at equilibrium distance would be repulsive without the charge-transfer interactions.

Thus, our analyses disprove the established conception that hydrogen bonding in DNA base pairs is a predominantly electrostatic phenomenon. Almost all arguments we found in the literature in favor of the electrostatic model were eventually based on the work of Umeyama and Morokuma[6] on the hydrogen bond in water dimers and other neutral hydrogen-bound complexes (see introduction). But in fact, the analyses of Umeyama and Morokuma do reveal a significant charge-transfer component. They[6] found that for the water dimer, for example, the totalattractive interaction is provided for 72% by electrostatic interaction, for 21% by charge transfer and for 6% by polarization. We feel that the conclusions of Umeyama and Morokuma are not well represented if this charge-transfer component they found is completely ignored.

In the present work, for both Watson-Crick pairs, i.e. AT and GC in their equilibrium geometry, we find that ΔE_{oi} provides even 41% of all attractive interactions, while electrostatic forces contribute 59% (Table 5.6). The ΔE_{oi} can be further split into 38% ΔE_{σ} and 3% ΔE_{π} for AT, and 35% ΔE_{σ} and 6% ΔE_{π} for GC. In the complexes with the 0.1 Å elongated hydrogen bonds, i.e. $AT_{0.1Å}$ and $GC_{0.1Å}$, ΔE_{oi} provides still 39% of all attractive interactions (Table 5.6). We conclude that, at variance with current belief, charge transfer plays a vital role in the hydrogen bonds of DNA base pairs.

We were also interested into how the bonding mechanism is affected by more severe changes in the geometry, for example, if the A–T or G–C bond is still further elongated in the way described above for $AT_{0.1Å}$ and $GC_{0.1Å}$ (see also Table 5.6). Thus, we have analyzed the A–T and G–C bond energy as a function of the base–base distance *r*; the results are shown in Figures 5.7 and 5.8, respectively. Around the equilibrium distance, ΔE_{oi} and ΔV_{elstat} are of the same order of magnitude as discussed above. But at larger hydrogen-bond distances, solely ΔV_{elstat} survives as the only significant term causing attraction. The reason why ΔE_{oi} disappears faster with increasing base–base distance r is that the overlap, necessary for donor–acceptor interactions to occur, vanishes exponentially whereas ΔV_{elstat} decays more slowly as $1/r^3$.[13]

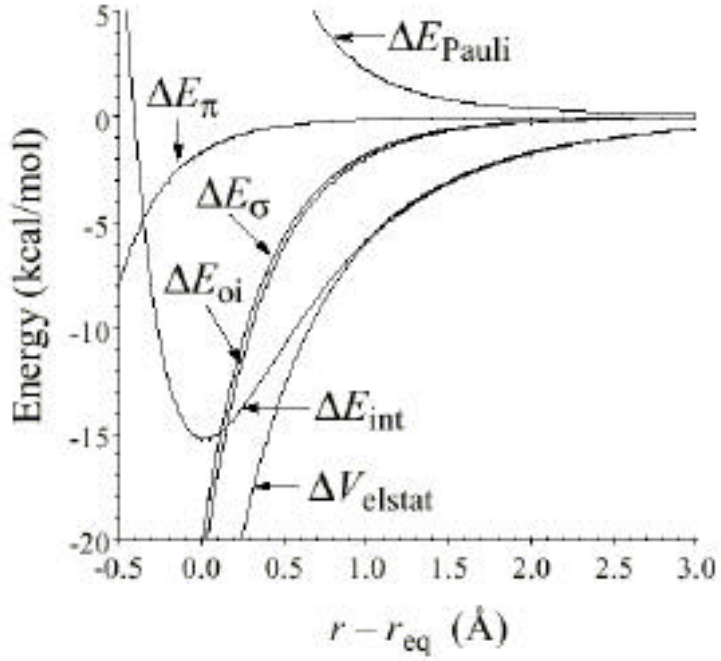


Figure 5.7. Bond energy decomposition (at BP86/TZ2P) as function of the adenine-thymine distance ($r - r_{\text{eq}} = 0.0$ corresponds to the equilibrium distance).

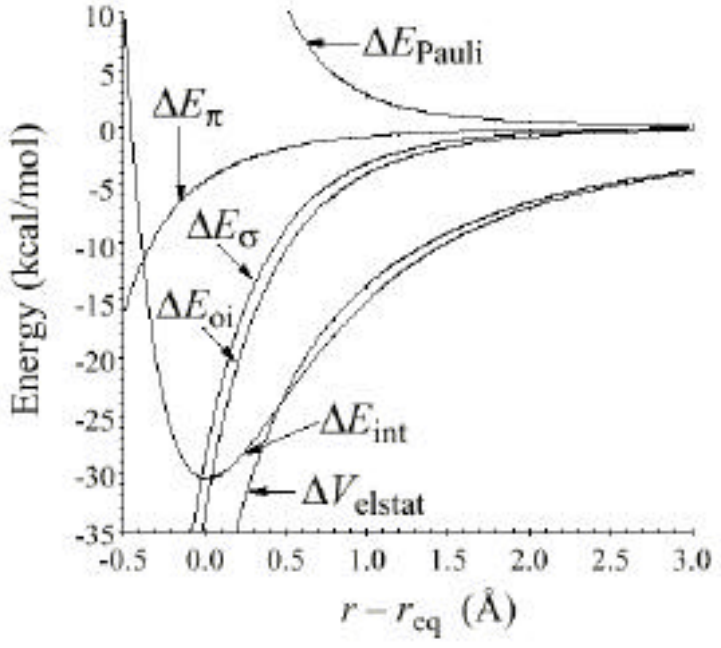


Figure 5.8. Bond energy decomposition (at BP86/TZ2P) as function of the guanine-cytosine distance ($r - r_{\text{eq}} = 0.0$ corresponds to the equilibrium distance).

5.3.3 Extension of VDD Method for Analysing Charge Distribution

The base-pairing interactions, in particular σ charge transfer and π polarization, discussed in the previous section modify the charge distribution around the nuclei. We have analyzed this reorganization of the charge distribution using the Voronoi deformation density (VDD) method, introduced in Ref. [11a]. The VVD charge Q_A^{VDD} of an atom A monitors how much electronic charge moves into ($Q_A^{\mathrm{VDD}} < 0$) or out of ($Q_A^{\mathrm{VDD}} > 0$) a region of space around nucleus A that is closer to this than to any other nucleus. This particular compartment of space is the Voronoi cell of atom A,[12h] and it is bounded by the bond midplanes on and perpendicular to all bond axes between nucleus A and its neighbouring nuclei (cf. the Wigner-Seitz cells in crystals). The VVD charge Q_A^{VDD} is computed as the (numerical) integral of the deformation density $\Delta\rho(\mathbf{r}) = \rho(\mathbf{r}) - \sum_B \rho_B(\mathbf{r})$ in the volume of the corresponding Voronoi cell (eq 5.3.1).

$$Q_A^{\mathrm{VDD}} = -\int_{\substack{\text{Voronoi}\\ \text{cell of A}}} \left(\rho(\mathbf{r}) - \sum_B \rho_B(\mathbf{r})\right) d\mathbf{r} \qquad (5.3.1)$$

Here, $\rho(\mathbf{r})$ is the electron density of the molecule and $\sum_B \rho_B(\mathbf{r})$ the superposition of atomic densities $\sum_B \rho_B$ of a fictitious promolecule without chemical interactions that is associated with the situation in which all atoms are neutral. As has been shown before, the VDD method yields chemically meaningful atomic charges that display hardly any basis set dependence.[11] Note, however, that the value of Q_A^{VDD} does depend on both the chosen reference density (i.e. the promolecule) and the shape of the Voronoi cell.

Front Atom Problem and its Solution: An Extension of the VDD Method

For the DNA base pairs, we want to know the charge rearrangement associated with the base-pairing interaction, in particular that on the front atoms on each base, i.e. the atoms pointing toward the other base. It may seem to be a plausible approach to simply compute for each atom A the difference between the atomic charge in the base pair, $Q_{A,\,\mathrm{pair}}^{\mathrm{VDD}}$, and that in the separate base, $Q_{A,\,\mathrm{base}}^{\mathrm{VDD}}$ (eq 5.3.2):

$$\Delta Q_A^{\mathrm{VDD}} = Q_{A,\,\mathrm{pair}}^{\mathrm{VDD}} - Q_{A,\,\mathrm{base}}^{\mathrm{VDD}}$$

$$= -\Big[\int_{\substack{\text{Voronoi cell}\\ \text{of A in pair}}} \Big(\rho_{\mathrm{pair}}(\mathbf{r}) - \sum\nolimits_B \rho_B(\mathbf{r})\Big) d\mathbf{r} \;-\; \int_{\substack{\text{Voronoi cell}\\ \text{of A in base}}} \Big(\rho_{\mathrm{base}}(\mathbf{r}) - \sum\nolimits_B \rho_B(\mathbf{r})\Big) d\mathbf{r} \Big] \quad (5.3.2)$$

However, the effect of A–T and G–C hydrogen bonding on the atomic charges is about an order of magnitude smaller than the charge rearrangements due to the primary process of strong chemical bond formation within the individual bases. In that case, $\Delta Q_A^{\mathrm{VDD}}$ as defined in eq 5.3.2 is not a reliable indicator of the charge flow due to hydrogen bonding, at least not for the front atoms that form the bonds with the opposite base. Note that $Q_{A,\,\mathrm{pair}}^{\mathrm{VDD}}$ and $Q_{A,\,\mathrm{base}}^{\mathrm{VDD}}$ differ in two respects: (i) the different molecular densities ρ_{pair} and ρ_{base}, and (ii) the altered Voronoi cell. For the front atoms, the latter effect is important since in a free base the Voronoi cell of such an atom will extend to infinity in the direction where the second base will be located. In the pair, of course, the Voronoi cell of the front atom will have as one of its faces the bond midplane perpendicular to the bond to the other base and cutting that bond in half. This drastic change of the shape of the Voronoi cell has as much effect on the VDD charge as the subtle change of the density from ρ_{base} to ρ_{pair}, rendering eq 5.3.2 useless. We wish to emphasize that other methods for the calculation of atomic charges (Mulliken, Hirshfeld, Bader), where the presence of the new neighbour atom in the other base directly affects the atomic charge evaluation on the front atom, are in principle subject to the same kind of problem when the small change in atomic charge due to hydrogen bonding is calculated as the difference of the "large" charges in the pair and the base.

The VDD charge analysis offers a natural solution to this problem. The relevant density difference, caused by the hydrogen bonding between the bases, is the difference between the SCF density of the pair as final density and the superposition of the densities of the bases as initial density. Integration of this deformation density, which is plotted in Figure 5.11 (*vide infra*), over the Voronoi cells of the atoms *in the pair* will reflect the charge flow due to the hydrogen bonding interaction (eq 5.3.3).

$$\Delta Q_A^{\mathrm{VDD}} = -\int_{\substack{\text{Voronoi cell}\\ \text{of A in pair}}} \Big[\rho_{\mathrm{pair}}(\mathbf{r}) - \rho_{\mathrm{base1}}(\mathbf{r}) - \rho_{\mathrm{base2}}(\mathbf{r})\Big] d\mathbf{r} \quad (5.3.3)$$

The calculation of a small difference of two large numbers that are not completely comparable, as

in eq 5.3.2, is now avoided. Only one Voronoi cell is used, the one in the pair, which eliminates the problem identified above. This method for "measuring" the charge rearrangement due to the weak hydrogen bonding is of course in the spirit of the VDD calculation of atomic charges resulting from chemical bond formation as in eq 5.3.1 since it integrates the relevant density difference over an appropriate atomic part of space.

Decomposition of VDD Charges into σ and π Components

To analyze the charge rearrangement caused by charge transfer in the σ-system and that caused by polarization in the π-system separately, we introduce a further extension of the VDD method: ΔQ_A that properly accounts for the effect of base pairing according to eq 5.3.3 is decomposed into the contributions of the σ- and π-deformation densities ΔQ_A^{σ} and ΔQ_A^{π} (eq 5.3.4):

$$\Delta Q_A^{\Gamma} = -\int_{\substack{\text{Voronoi cell}\\ \text{of A in pair}}} \left[\rho_{\text{pair}}^{\Gamma}(\mathbf{r}) - \rho_{\text{base1}}^{\Gamma}(\mathbf{r}) - \rho_{\text{base2}}^{\Gamma}(\mathbf{r})\right] d\mathbf{r} \qquad (5.3.4)$$

The density ρ^{Γ} is obtained as the sum of orbital densities of the occupied molecular orbitals belonging to the irreducible representation Γ (eq 5.3.5):

$$\rho^{\Gamma} = \sum_{i}^{\text{occ}} \left|\psi_i^{\Gamma}\right|^2 \qquad (5.3.5)$$

5.3.4 Charge Redistribution due to Hydrogen Bonding

The changes in atomic charge ΔQ_A caused by hydrogen bonding in AT and GC (eq 5.3.3) are collected in Figures 5.9 and 5.10, respectively. An unexpected pattern emerges for the ΔQ_A's of the atoms directly involved in hydrogen bonds. Instead of losing density as one would at first expect on the basis of the orbital interactions (see Scheme 2), the electron-donor atoms (oxygen and nitrogen) *gain* density and become more negative! For AT, we find that adenine N1 and thymine O4 gain negative charges of –0.031 and –0.037 electrons, respectively (Figure 5.9). Likewise, in GC, the negative charge on guanine O6 increases by –0.049 electrons, and the electron-donor atoms in cytosine, O2 and N3, gain negative charges of –0.030 and –0.037 electrons, respectively (Figure 5.10). Surprizing is also that the electronic density at the hydrogen

atom of the electron-accepting N–H group decreases upon formation of the complex, yielding ΔQ_A values ranging from +0.035 to +0.048 electrons (Figures 5.9 and 5.10). An increase of electron density would have been expected due to the charge-transfer interactions (see Scheme 2). Furthermore, we only find a moderate accumulation of negative charge on the nitrogen atoms of the electron-accepting N–H groups (Figures 5.9 and 5.10).

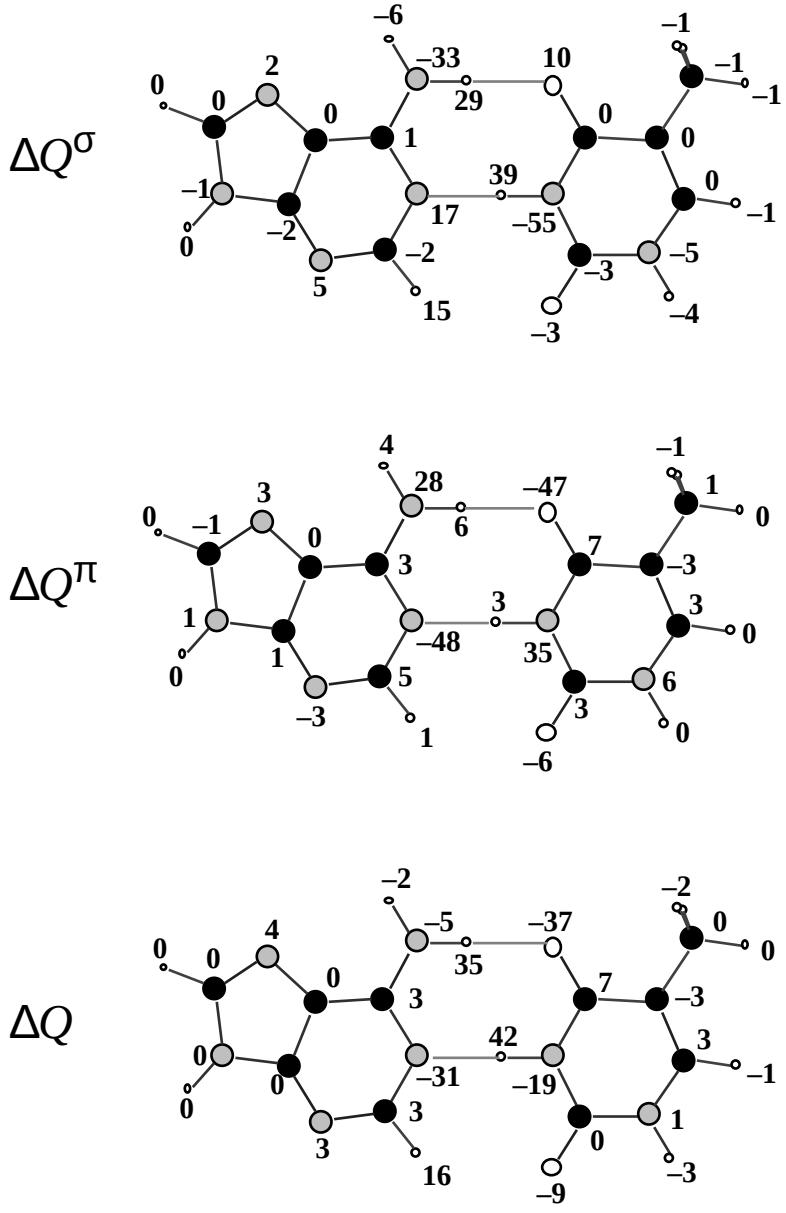


Figure 5.9. Changes in σ, π and total VDD atomic charges (in mili-electrons) on forming the N6(H)•••O4 and N1•••(H)N3 hydrogen bonds between adenine and thymine in AT (see Scheme 1) calculated at BP86/TZ2P.

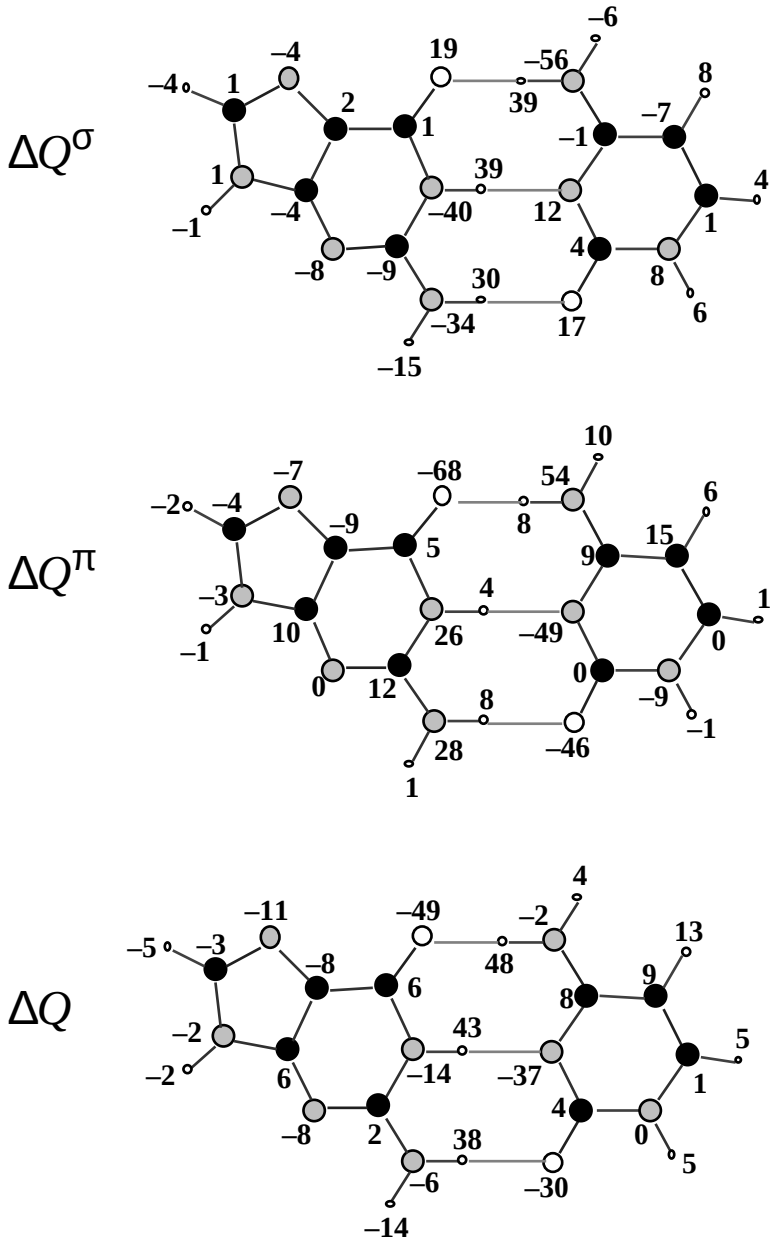


Figure 5.10. Changes in σ, π and total VDD atomic charges (in mili-electrons) on forming the O6•••(H)N4, N1(H)•••N3 and N2(H)•••O2 hydrogen bonds between guanine and cytosine in GC (see Scheme 1) calculated at BP86/TZ2P.

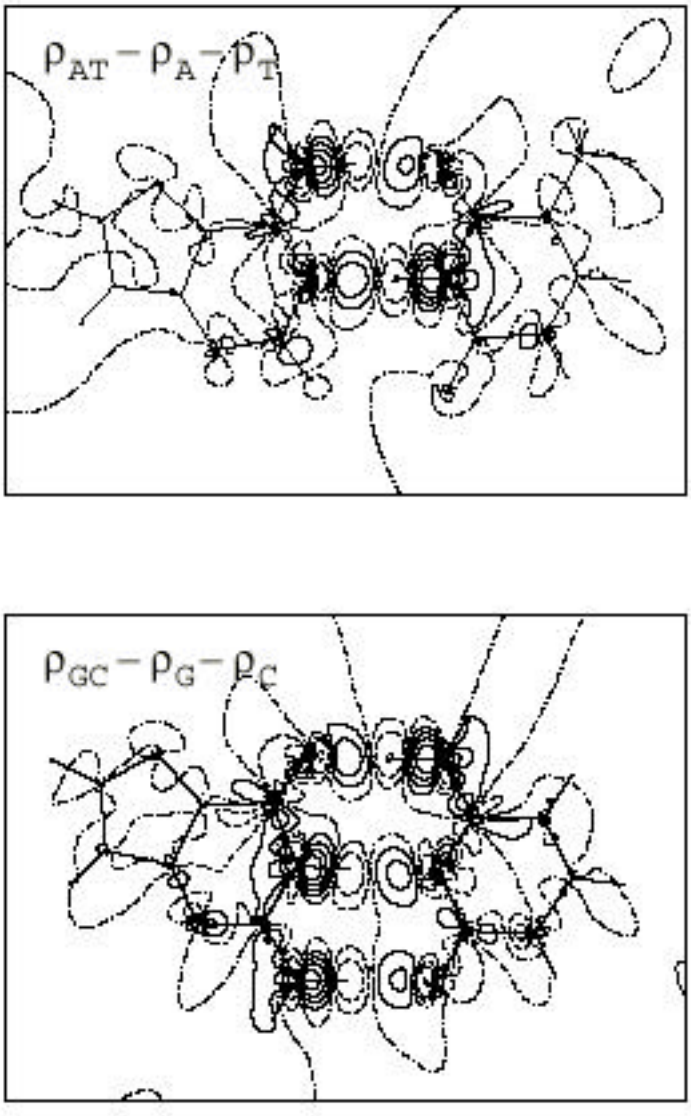


Figure 5.11. Contour plots for AT and GC of the difference between the density of the base pair and the superposition of densities of the individual bases calculated at BP86/TZ2P (Scan values: ±0.05, ±0.02, ±0.01, ±0.005, ±0.002, 0. Solid, dashed and dash-dotted contours indicate positive, negative and zero values, respectively).

How do these ΔQ_A values arise or, in other words, what is the physics behind these numbers? We have tried to find out by decomposing ΔQ_A into its σ and π components ΔQ_A^{σ} and ΔQ_A^{π} (eq 5.3.4) which are also shown in Figures 5.9 and 5.10. The ΔQ_A^{σ} values reveal a clear charge-transfer picture for AT and GC: negative charge is lost on the electron-donor atoms whereas there is a significant accumulation of negative charge on the nitrogen atoms of the electron-accepting N–H bonds. It is the reorganization of charge stemming from π polarization, as reflected by the ΔQ_A^{π} values, that causes the counterintuitive pattern of the overall charge rearrangement monitored by ΔQ_A. Note that ΔQ_A^{σ} and ΔQ_A^{π} are of the same order of magnitude whereas ΔE_{π} is an order of magnitude smaller than ΔE_{σ} (see section 5.3.2). The π-electron density of the bases is polarized in such a way that the build-up of charge arising from charge-transfer interactions in the σ-system is counteracted and compensated: the electron-donor atoms gain π density and the nitrogen atoms of the electron-accepting N–H bonds loose π-density (compare ΔQ_A^{σ} and ΔQ_A^{π} in Figures 5.9 and 5.10). This suggests that there may be some kind of cooperativity between the σ charge transfer and π polarization which is reminiscent of the resonance-assistance proposed by Gilli et al..[7a] In the following section, we examine if such a synergism between ΔE_{σ} and ΔE_{π} interactions really exists.

But first we want to resolve the still open question why hydrogen bonding makes the H atoms involved more positive (Figures 5.9 and 5.10). This turns out to be a subtle mechanism. To get an idea how the positive ΔQ_A charges of these H atoms arise, we have plotted the corresponding deformation densities for the formation of AT and GC from their separate bases (i.e. the density of the base pair minus the superimposed densities of the bases) in Figure 5.11. These deformation-density plots nicely show the depletion of charge around the hydrogen bonding H atoms that the VDD charges had already detected. A more detailed examination reveals that an important portion of this charge depletion stems from the Pauli repulsion (i.e. ΔE_{Pauli}) between the occupied orbitals of the two bases, in particular the strongly overlapping O or N lone pairs of one base and the *occupied* N–H σ bonding orbitals of the other base. But also the bonding orbital interactions (i.e. ΔE_{oi}) contribute to this feature in the deformation density. Morokuma and coworkers have shown,[6b,c] that charge depletion around the hydrogen bonding H atom in, for example, the water dimer is due to a large extent to *polarization* in the σ-electron system even though this term does not contribute much to the interaction energy. A further mechanism that may contribute to the

depletion of charge around these H atoms is that the lone pairs that donate charge, penetrate deeply into the space around the hydrogen nucleus of the partner N–H bond, i.e. the Voronoi cell of that hydrogen atom. Consequently, as the lone pair gets depopulated during charge transfer, it causes a depletion of charge not only on the donor atom but also in the Voronoi cell of the "accepting" hydrogen atom. Meanwhile, the N–H acceptor orbitals have a compact high amplitude character around the nitrogen atom whereas they are more extended and diffuse on hydrogen (see Figures 5.3 and 5.4). This makes that the electronic charge accepted during charge transfer appears in a region closely around the nucleus of nitrogen and more distant from that of hydrogen. Thus, we find that thanks to the two extensions presented here the VDD method has become a valuable tool for monitoring and analyzing even very subtle charge rearrangements.

5.3.5 Synergism in Hydrogen Bonding

At this point, we are left with three questions concerning DNA base pairing: (i) do the hydrogen bonds that donate charge in opposite directions reinforce each other by reducing the net build-up of charge on each base? (ii) is there a cooperative effect or resonance-assistance by the π-electron system as suggested by Gilli et al.?[7a] and (iii) how important is π-polarization for the hydrogen bonding structure (i.e. bond distances)? To answer these questions, we have carried out further detailed analyses of the base-pairing energies, in which individual types of orbital interactions are considered while others are switched off by removing the appropriate σ or π virtuals from the respective DNA bases. The results are collected in Tables 5.7 and 5.8. Our notation is exemplified for the AT pair: A(σ,π)T(σ,π) corresponds to a regular computation on AT in which all σ and π virtuals are included; A(σ,–)T(σ,–), for example, indicates that all σ virtuals are available on A and T whereas the π virtuals have been removed from both bases.

Synergism Between Individual Hydrogen Bonds in DNA Base Pairs ?

The synergism within the σ-system between charge transfer from one base to the other through one hydrogen bond and back through the other hydrogen bond (AT) or bonds (GC) is obtained as the difference between ΔE_σ in entry IIIa and entry IIa+b in Table 5.7 or 5.8. In IIIa, charge-transfer interactions in both directions occurs simultaneously, whereas IIa+b gives the sum of the

situations with charge-transfer interaction forth only and back only; π-polarization is completely switched off. The anticipated synergic effect does not occur: we find that ΔE_σ(IIIa) – ΔE_σ(IIa+b) is close to zero with values of +0.8 and +1.1 kcal/mol for AT and GC (see Tables 5.7 and 5.8). This suggests that the hydrogen bonds donating charge in opposite directions operate independently. This is nicely confirmed by comparing the regular deformation density (e.g. $\Delta\rho_{AT} = \rho_{AT} - \rho_A - \rho_T$ see Figure 5.11) with the deformation densities belonging to IIa and IIb (i.e. $\Delta\rho_{A(\sigma,-)T(-,-)} = \rho_{A(\sigma,-)T(-,-)} - \rho_A - \rho_T$ and $\Delta\rho_{A(-,-)T(\sigma,-)} = \rho_{A(-,-)T(\sigma,-)} - \rho_A - \rho_T$, not shown in Figure). This comparison shows that the charge-transfer processes that donate charge in opposite directions do not affect each others locally induced (and conversely oriented) charge separations, while their simultaneous occurrence still does reduce the *net* build-up of charge. The fact that ΔE_σ(III) – ΔE_σ(IIa+b) is even slightly destabilizing can be ascribed to the repulsion accompanying the simultaneous occurrence in III but not in IIa or IIb of an accumulation of density both at the donor and acceptor atoms next to each other on the same base (see Figure 5.11).

Table 5.7. Analysis of the synergy between σ- and π-orbital interactions in A–T (in kcal/mol).[a]

		virtuals available[b]	ΔE_σ	ΔE_π	ΔE_{oi}
I		A(σ,π)T(σ,π)	–20.7	–1.7	–22.4
II	a	A(–,–)T(σ,–)	–12.9		
	b	A(σ,–)T(–,–)	–8.3		
	a+b		–21.2		
	c	A(–,–)T(–,π)		–0.7	
	d	A(–,π)T(–,–)		–0.7	
	c+d			–1.4	
III	a	A(σ,–)T(σ,–)	–20.4		–20.4
	b	A(–,π)T(–,π)		–1.3	–1.3
	a+b				–21.7

[a] BP86/TZ2P.

[b] A(σ,–)T(σ,–) for example indicates: σ virtuals available on and π virtuals removed from both A and T.

Table 5.8. Analysis of the synergy between σ- and π-orbital interactions in G–C (in kcal/mol).[a]

		virtuals available[b]	ΔE_σ	ΔE_π	ΔE_{oi}
I		G(σ,π)C(σ,π)	–29.3	–4.8	–34.1
II	a	G(–,–)C(σ,–)	–13.6		
	b	G(σ,–)C(–,–)	–16.4		
	a+b		–30.0		
	c	G(–,–)C(–,π)		–2.0	
	d	G(–,π)C(–,–)		–1.6	
	c+d			–3.6	
III	a	G(σ,–)C(σ,–)	–28.9		
	b	G(–,π)C(–,π)		–3.8	–3.8
	a+b				–32.7

[a] BP86/TZ2P

[b] G(σ,–)C(σ,–) for example indicates: σ virtuals available on and π virtuals removed from both G and C.

In the same manner, we can compute the synergism between the π-polarizations occurring on each of the bases as the difference between ΔE_π in entry IIIb and entry IIc+d in Table 5.7 or 5.8. In IIIb, π-polarization occurs on both bases simultaneously, whereas IIc+d gives the sum of the situations with π-polarization on one base only and on the other base only; charge transfer in the σ-electron system is completely switched off. Again, there is no synergic effect with ΔE_π(IIIb) – ΔE_π(IIc+d) being virtually zero (0.1 and 0.2 kcal/mol for AT and GC). Thus, the π-polarizations occurring in each individual base of a base pair are independent.

Synergism Between σ Charge Transfer and π Polarization ?

The synergism between charge-transfer in the σ-electron system (ΔE_σ) and polarization in the π-electron system (ΔE_π) can be computed as the difference between ΔE_{oi} in entry I and entry IIIa+b in Tables 5.7 or 5.8. In I, all σ charge-transfer and π-polarization interactions occur simultaneously, whereas IIIa+b gives the sum of the situations in which there is σ charge-transfer

interaction only and π-polarization only. We find very small synergic effects ΔE_{oi}(I) – ΔE_{oi}(IIIa+b) of –0.7 and –1.4 kcal/mol for AT and GC. The overall synergic effect is composed of a synergic stabilization in the σ charge-transfer interaction ΔE_{σ}(I) – ΔE_{σ}(IIIa) of –0.3 and –0.4 kcal/mol, and a synergic stabilization in the π polarization ΔE_{π}(I) – ΔE_{π}(IIIa) of –0.4 and –1.0 kcal/mol for AT and GC, respectively.

We conclude that the π electrons give almost no assistance to the donor–acceptor interactions in the hydrogen bonds in the sense of a synergism. Energetically, the main assistance caused by the π electrons is simply the small although not negligible term ΔE_{π} which contributes –1.7 and –4.8 kcal/mol to the net hydrogen bond energy (see Tables 5.3, 5.4, 5.7 and 5.8; see also section 5.3.2).

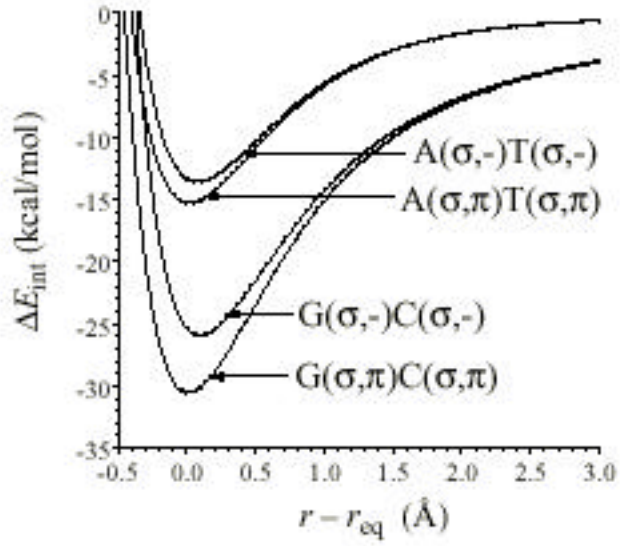


Figure 5.12. Interaction energy of AT and GC with and without π-virtuals as function of the base-base distance calculated at BP86/TZ2P ($r - r_{eq} = 0.0$ corresponds to the equilibrium distance).

But how important is this ΔE_π term for the *structure*, that is, the hydrogen bond distances of the DNA base pairs? We can determine this influence, by computing the bond energy with π polarization switched on and off (i.e. with or without the π virtuals of the bases) as a function of the base–base distance (we follow the procedure for varying the bond length described before in section 5.3.2). The resulting bond energy curves are shown in Figure 5.12. The comparison between the curves of $A(\sigma,\pi)T(\sigma,\pi)$ and $G(\sigma,\pi)C(\sigma,\pi)$ (i.e. π polarization switched on) and those of $A(\sigma,-)T(\sigma,-)$ and $G(\sigma,-)C(\sigma,-)$ (i.e. π polarization off) shows that without π polarization the equilibrium hydrogen bond distances expand for both base pairs by some 0.1 Å. This would yield hydrogen bond lengths for AT of 2.95 and 2.91 Å and for GC of 2.97, 2.98 and 2.83 Å. One might conceive the extra bond shortening caused by π polarization as some kind of resonance-assistance. However, we stress again that ΔE_π is only a minor bonding component and that there is no resonance-assistance in the sense of a synergism between σ-charge transfer and π polarization.

5.4 Conclusions

The hydrogen bond in DNA base pairs is, at variance with widespread belief, not a pure or essentially electrostatic phenomenon. Instead, as follows from our BP86/TZ2P investigation, it has a substantial charge-transfer character caused by donor–acceptor orbital interactions (between O or N lone pairs and N–H σ* acceptor orbitals) that are of the same order of magnitude as the electrostatic term. Polarization in the π-electron system provides an additional stabilizing term. This is, however, one order of magnitude smaller than the σ donor–acceptor interactions. It still has the effect of reducing the base–base bond distance by 0.1 Å. A more detailed bond analysis shows that no substantial synergism occurs between the individual hydrogen bonds in the base pairs nor between σ orbital interactions and π polarization. And there is no C–H•••O hydrogen bond in AT. The occurrence of charge transfer and polarization in the σ- and π-electron system, respectively, is confirmed by our complementary analysis of the electron density distribution with the extensions of the VDD method that we have introduced in the present work.

It is evident that many other factors are of great importance for the working of the molecular genetic machinery (e.g., structural complementarity of bases, hydrophobic interactions and other

medium effects, interaction with enzymes and other proteins, etc.).[1d,22] However, regarding the intrinsic cohesion of DNA, we may conclude that it is the chemical charge-transfer nature of the hydrogen bond in Watson-Crick base pairs, rather than resonance-assistance by the π-electron system, that together with the classical electrostatic interaction is vital to the behavior and the stability and, thus, the evolution of nature's genetic code.

References

[1] a) G. A. Jeffrey, W. Saenger, *Hydrogen Bonding in Biological Structures*, Springer-Verlag, Berlin, New York, Heidelberg, **1991**

b) G. A. Jeffrey, *An Introduction to Hydrogen bonding*, Oxford University Press, New York, Oxford **1997**, Chapter 10

c) J. D. Watson, F. H. C. Crick, *Nature* **1953**, *171*, 737

d) W. Saenger, *Principles of Nucleic Acid Structure*, Springer-Verlag, New York, Berlin, Heidelberg, Tokyo, **1984**

[2] a) J. P. Lewis, O. F. Sankey, *Biophys. J.* **1995**, *69*, 1068

b) Y. S. Kong, M. S. Jhon, P. O. Löwdin, *Int. J. Quantum. Chem., Symp. QB* **1987**, *14*, 189

c) C. Nagata, M. Aida, *J. Molec. Struct.* **1988**, *179*, 451

d) I. R. Gould, P. A. Kollman, *J. Am. Chem. Soc.* **1994**, *116*, 2493

e) J. Sponer, J. Leszczynski, P. Hobza, *J. Phys. Chem.* **1996**, *100*, 1965

f) J. Sponer, J. Leszczynski, P. Hobza, *J. Biomol. Struct. Dyn.* **1996**, *14*, 117

g) J. Sponer, P. Hobza, J. Leszczynski, in *Computational Chemistry. Reviews of Current Trends*, (Ed.: J. Leszczynski), World Scientific Publisher, Singapore, **1996**, p. 185-218

h) M. Hutter, T. Clark, *J. Am. Chem. Soc.* **1996**, *118*, 7574

i) K. Brameld, S. Dasgupta, W. A. Goddard III, *J. Phys. Chem. B* **1997**, *101*, 4851

j) M. Meyer, J. Sühnel, *J. Biomol. Struct. Dyn.* **1997**, *15*, 619

k) R. Santamaria, A. Vázquez, *J. Comp. Chem.* **1994**, *15*, 981

l) J. Bertran, A. Oliva, L. Rodríguez-Santiago, M. Sodupe, *J. Am. Chem. Soc.* **1998**, *120*, 8159

[3] a) F. Sim, A. St-Amant, I. Papai, D. R. Salahub, *J. Am. Chem. Soc.* **1992**, *114*, 4391

b) H. Guo, S. Sirois, E. I. Proynov, D. R. Salahub, in *Theoretical Treatment of Hydrogen Bonding*, (Ed.: D. Hadzi), Wiley, New York, **1997**

c) S. Sirois, E. I. Proynov, D. T. Nguyen, D. R. Salahub, *J. Chem. Phys.* **1997**, *107*, 6770

d) P. R. Rablen, J. W. Lockman, W. L. Jorgensen, *J. Phys. Chem.* **1998**, *102*, 3782

e) K. Kim, K. D. Jordan, *J. Phys. Chem.* **1994**, *98*, 10089

f) J. J. Novoa, C. Sosa, *J. Phys. Chem.* **1995**, *99*, 15837

g) Z. Latajka, Y. Bouteiller, *J. Chem. Phys.* **1994**, *101*, 9793

h) J. E. Del Bene, W. B. Person, K. Szczepaniak, *J. Phys. Chem.* **1995**, *99*, 10705

i) J. Florian, B. G. Johnson, *J. Phys. Chem.* **1995**, *99*, 5899

j) J. E. Combariza, N. R. Kestner, *J. Phys. Chem.* **1995**, *99*, 2717

k) B. Civalleri, E. Garrone, P. Ugliengo, *J. Molec. Struct.* **1997**, *419*, 227

l) M. Lozynski, D. Rusinska-Roszak, H.-G. Mack, *J. Phys. Chem.* **1998**, *102*, 2899

m) A. K. Chandra, M. Nguyen, *Chem. Phys.* **1998**, *232*, 299

n) B. Paizs, S. Suhai, *J. Comp. Chem.* **1998**, *19*, 575

o) M. A. McAllister, *J. Molec. Struct.* **1998**, *427*, 39

p) Y. P. Pan, M. A. McAllister, *J. Molec. Struct.* **1998**, *427*, 221

q) L. Gonzalez, O. Mo, M. Yanez, *J. Comp. Chem.* **1997**, *18*, 1124

[4] C. Fonseca Guerra, F. M. Bickelhaupt, *Angew. Chem.* **1999**, *111*, in press

[5] C. Fonseca Guerra, F. M. Bickelhaupt, J. G. Snijders, E. J. Baerends, submitted

[6] a) H. Umeyama, K. Morokuma, *J. Am. Chem. Soc.* **1977**, *99*, 1316

b) S. Yamabe, K. Morokuma, *J. Am. Chem. Soc.* **1975**, *97*, 4458

c) K. Morokuma, *Acc. Chem. Res.* **1977**, *10*, 294

[7] a) G. Gilli, F. Bellucci, V. Ferretti, V. Bertolasi, *J. Am. Chem. Soc.* **1989**, *111*, 1023

b) P. Gilli, V. Ferretti, V. Bertolasi, G. Gilli, in *Advances in Molecular Structure Research, Vol. 2*, (Eds.: M. Hargittai, I. Hargittai), JAI Press, Greenwich, CT, **1996**, p. 67-102

c) G. Gilli, V. Bertolasi, V. Ferretti, P. Gilli, *Acta Cryst.* **1993**, *B49*, 564

d) V. Bertolasi, P. Gilli, V. Ferretti, G. Gilli, *Acta Cryst.* **1995**, *B51*, 1004

[8] M. L. Huggins, *Angew. Chem.* **1971**, *83*, 163; *Angew. Chem. Int. Ed. Engl.* **1971**, *10*, 147

[9] a) F. M. Bickelhaupt, E. J. Baerends, *Rev. Comput. Chem.*, submitted

b) R. Hoffmann, *Angew. Chem.* **1982**, *94*, 725; *Angew. Chem. Int. Ed. Engl.* **1982**, *21*, 711

c) T. A. Albright, J. K. Burdett, M. Whangbo, *Orbital Interactions in Chemistry*, Wiley, New York, **1985**

d) B. Gimarc, *The Qualitative Molecular Orbital Approach*, Academic Press, New York, **1979**

[10] a) C–H•••O hydrogen bonding in AT has been suggested by: G. A. Leonard, K. McAuley-Hecht, T. Brown, W. N. Hunter, *Acta Cryst.* **1995**, *D51*, 136

b) It has been ruled out for AT by: O. V. Shishkin, J. Sponer, P. Hobza, *J. Molec. Struct.* **1999**, *477*, 15

For studies that do find C–H•••O hydrogen bonding in other systems, see, for example: c) K. N. Houk, S. Menzer, S. P. Newton, F. M. Raymo, J. F. Stoddart, D. J. Williams, *J. Am. Chem. Soc.* **1999**, *121*, 1479

d) P. Seiler, G. R. Weisman, E. D. Glendening, F. Weinhold, V. B. Johnson, J. D. Dunitz, *Angew. Chem.* **1987**, *99*, 1216; *Angew. Chem. Int. Ed. Engl.* **1987**, *26*, 1175

[11] a) F. M. Bickelhaupt, N. J. R. van Eikema Hommes, C. Fonseca Guerra, E. J. Baerends,

Organometallics **1996**, *15*, 2923

b) F. M. Bickelhaupt, C. Fonseca Guerra, J.-W. Handgraaf, E. J. Baerends, to be submitted

[12] a) C. Fonseca Guerra, O. Visser, J. G. Snijders, G. te Velde, E. J. Baerends, in *Methods and Techniques for Computational Chemistry*, (Eds.: E. Clementi, G. Corongiu), STEF, Cagliari **1995**, p. 305-395

b) E. J. Baerends, D. E. Ellis, P. Ros, *Chem. Phys.* **1973**, *2*, 41

c) E. J. Baerends, P. Ros, *Chem. Phys.* **1975**, *8*, 412

d) E. J. Baerends, P. Ros, *Int. J. Quantum. Chem. Symp.* **1978**, *12*, 169

e) W. Ravenek, in *Algorithms and Applications on Vector and Parallel Computers*, (Eds.: H. H. J. Riele, T. J. Dekker, H. A. van de Vorst), Elsevier, Amsterdam, **1987**

f) C. Fonseca Guerra, J. G. Snijders, G. te Velde, E. J. Baerends, *Theor. Chem. Acc.* **1998**, *99*, 391

g) P. M. Boerrigter, G. te Velde, E. J. Baerends, *Int. J. Quantum Chem.* **1988**, *33*, 87

h) G. te Velde, E. J. Baerends, *J. Comp. Phys.* **1992**, *99*, 84

i) J. G. Snijders, E. J. Baerends, P. Vernooijs, *At. Nucl. Data Tables* **1982**, *26*, 483

j) J. Krijn, E. J. Baerends, *Fit-Functions in the HFS-Method; Internal Report (in Dutch)*, Vrije Universiteit, Amsterdam, **1984**

k) L. Versluis, T. Ziegler, *J. Chem. Phys.* **1988**, *88*, 322

l) L. Fan, L. Versluis, T. Ziegler, E. J. Baerends, W. Ravenek, *Int. J. Quantum. Chem., Quantum. Chem. Symp.* **1988**, *S22*, 173

m) J. C. Slater, *Quantum Theory of Molecules and Solids, Vol. 4*, McGraw-Hill, New York, **1974**

n) A. D. Becke, *J. Chem. Phys.* **1986**, *84*, 4524

o) A. Becke, *Phys. Rev. A* **1988**, *38*, 3098

p) S. H. Vosko, L. Wilk, M. Nusair, *Can. J. Phys.* **1980**, *58*, 1200

q) J. P. Perdew, *Phys. Rev. B* **1986**, *33*, 8822; Erratum: *Phys. Rev. B* **1986**, *34*, 7406);

r) L. Fan, T. Ziegler, *J. Chem. Phys.* **1991**, *94*, 6057

s) T. Ziegler, *Chem. Rev.* **1991**, *91*, 651

[13] P. W. Atkins, *Physical Chemistry*, Oxford University Press, Oxford, **1982**

[14] S. F. Boys, F. Bernardi, *Mol. Phys.* **1970**, *19*, 553

[15] a) F. M. Bickelhaupt, N. M. M. Nibbering, E. M. van Wezenbeek, E. J. Baerends, *J. Phys. Chem.* **1992**, *96*, 4864

b) T. Ziegler, A. Rauk, *Inorg. Chem.* **1979**, *18*, 1755

c) T. Ziegler, A. Rauk, *Inorg. Chem.* **1979**, *18*, 1558

d) T. Ziegler, A. Rauk, *Theor. Chim. Acta* **1977**, *46*, 1

[16] a) J. P. Perdew, in Electronic Structure of Solids, (Eds.: P. Ziesche, H. Eschrig), Akademie Verlag, Berlin, 1991, p. 11-20

b) J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, C. Fiolhais, Phys. Rev. B 1992, 46, 6671

[17] a) C. Lee, W. Yang, R. G. Parr, *Phys. Rev. B* **1988**, *37*, 785

b) B. G. Johnson, P. M. W. Gill, J. A. Pople, *J. Chem. Phys.* **1992**, *98*, 5612

[18] I. K. Yanson, A. B. Teplitsky, L. F. Sukhodub, *Biopolymers* **1979**, *18*, 1149

[19] a) N. C. Seeman, J. M. Rosenberg, F. L. Suddath, J. J. P. Kim, A. Rich, *J. Mol. Biol.* **1976**, *104*, 109-144

b) J. M. Rosenberg, N. C. Seeman, R. O. Day, A. Rich, *J. Mol. Biol.* **1976**, *104*, 145

[20] The role of structural complementarity for Watson-Crick base pairing is discussed, for example, in: a) S. Moran, R. X.-F. Ren, S. Rumney, E. T. Kool, *J. Am. Chem. Soc.* **1997**, *119*, 2056

b) U. Diederichsen, *Angew. Chem.* **1998**, *110*, 2056; Angew. Chem. Int. Ed. 1998, 37, 1655

c) See, however, also: T. A. Evans, K. R. Seddon, *Chem. Commun.* **1997**, 2023

d) For a very recent discussion of the role of minor groove interactions between DNA and polymerase for replication, see: J. C. Morales, E. T. Kool, *J. Am. Chem. Soc.* **1999**, *121*, 2323

[21] a) A. E. Reed, F. Weinhold, *J. Chem. Phys.* **1983**, *78*, 4066

See also: b) A. E. Reed, F. Weinhold, L. A. Curtiss, D. J. Pochatko, *J. Chem. Phys.* **1986**, *84*, 5687

c) L. A. Curtiss, D. J. Pochatko, A. E. Reed, F. Weinhold, *J. Chem. Phys.* **1985**, *82*, 2679

[22] a) L. Stryer, *Biochemistry*, W.H. Freeman and Company, New York, **1988**

b) M. W. Strickberger, *Evolution*, Jones and Bartlett Publishers, Boston, **1996**, Chapters 7 and 8

c) A. Kornberg, T. A. Baker, *DNA Replication*, Freeman and Company, New York, **1992**

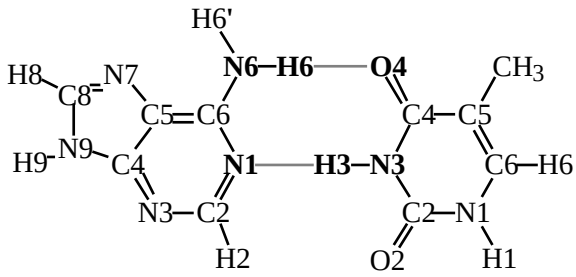
Chapter 6

Hydrogen Bonding in DNA Base Pairs: Reconciliation of Theory and Experiment

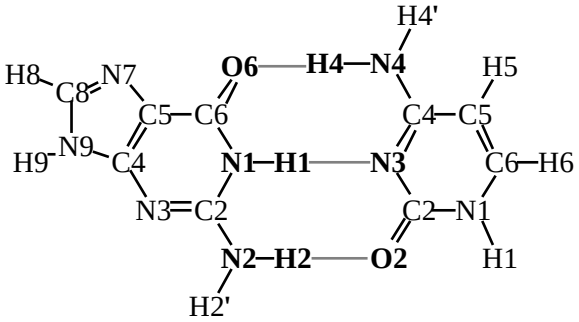
Up till now, there has been a significant disagreement between theory and experiment regarding hydrogen bond lengths in Watson-Crick base pairs. To investigate the possible sources of this discrepancy, we have studied numerous model systems for adenine–thymine (AT) and guanine–cytosine (GC) base pairs at various levels (i.e., BP86, PW91 and BLYP) of nonlocal density functional theory (DFT) in combination with different Slater-type orbital (STO) basis sets. Best agreement with available gas-phase experimental A–T and G–C bond enthalpies (–12.1 and –21.0 kcal/mol) is obtained at the BP86/TZ2P level, which (for 298 K) yields –11.8 and –23.8 kcal/mol. However, the computed hydrogen bond lengths show again the notorious discrepancy with experimental values. The origin of this discrepancy is not the use of the plain nucleic bases as models for nucleotides: the disagreement with experiment remains no matter if we use hydrogen, methyl, deoxyribose or 5'-deoxyribose monophosphate as the substituents at N9 and N1 of the purine and pyrimidine bases, respectively. Even the BP86/DZP geometry of the Watson-Crick-type dimer of deoxyadenylyl-3',5'-deoxyuridine including one Na^+ ion (with 123 atoms our largest model for sodium adenylyl-3',5'-uridine hexahydrate, the crystal of which had been studied experimentally with the use of X-ray diffraction) still shows this disagreement with experiment. The source of the divergence turns out to be the molecular environment (water, sugar hydroxyl groups, counterions) of the base pairs in the crystals studied experimentally. This has been missing, so far, in all theoretical models. After we had incorporated the major elements of this environment in our model systems, excellent agreement between our BP86/TZ2P geometries and the X-ray crystal structures was achieved.

6.1 Introduction

Hydrogen bonds are important in many fields of biological chemistry. They play, for instance, a key role in the working of the genetic code.[1] Already in 1953, Watson and Crick[1d] proposed a structure for DNA in which two helical chains of nucleotides are held together by the hydrogen bonds that occur in a selective fashion between a purine and a pyrimidine nucleic base giving rise to the Watson-Crick pairs adenine–thymine (AT) and guanine–cytosine (GC), see Scheme 1.



Adenine–Thymine



Guanine–Cytosine

Scheme 1

Until recently, hydrogen bonds were conceived as predominantly electrostatic phenomena that in the case of DNA base pairs are reinforced by polarization of the π-electron system (Resonance Assisted Hydrogen Bonding, RAHB).[2] Very recently,[3,4] we have shown through detailed analyses of the bonding mechanism that donor–acceptor orbital interactions between the DNA bases in the Watson-Crick pairs are of comparable strength as electrostatic interactions. The donor–acceptor or charge-transfer term is provided by the interactions of lone-pair orbitals on O or N of one base with N–H σ* acceptor orbitals of the other base. This picture complements and is in perfect agreement with experimental evidence[5] for a partial covalent character of hydrogen bonds obtained, lately, by different groups through X-ray diffraction studies on ice[5a] and NMR investigations on hydrogen bonds in RNA[5b] and in proteins.[5c,d]

In the present paper, we wish to address a different point. Whereas both density functional and traditional *ab initio* methods satisfactorily reproduce experimental A–T and G–C hydrogen bond enthalpies,[6] there has been a significant discrepancy between theory[7] and experiment[1c,8] regarding hydrogen bond lengths in the Watson-Crick base pairs. Our purpose is to find and understand the source of this disagreement using modern density functional theory (DFT) and, in this way, to arrive at a suitable quantum chemical approach for biochemical molecules that involve hydrogen bonds.

First, an extensive comparison is done between the performances of a number of density functionals (BP86, PW91, BLYP) in combination with different Slater-type orbital (STO) basis sets. The suitability of DFT for hydrogen-bonded systems has been the subject of many investigations.[9] In a study on the water dimer and the formamide-water complex, for example, Sim *et al.*[9a] found that nonlocal DFT performs satisfactorily, yielding results that compare well with those from correlated *ab initio* methods. At this point, however, we anticipate that whereas our highest-level base-pairing enthalpies are in excellent agreement with gas-phase experimental values, we still arrive at the notorious discrepancies with experimental (X-ray crystal) structures that were encountered before.

In a preliminary communication,[4] we have briefly reported how this, and the fact that Watson-Crick base pairing is associated with very shallow potential energy surfaces, has led us to study the possible effects of using better models for the glycosidic N–C bond as well as the influence of the molecular environment that the bases experience in the crystals studied experimentally.[1c,8] Here, we present a full account of our investigations which have been extended, meanwhile, to larger DNA segments. In particular, we present the results of the first high-level DFT study of the Watson-Crick-type dimer of adenylyl-3',5'-uridine, *i.e.*, $(ApU)_2$, including a full geometry optimization and bond analysis. Furthermore, the structure of and bonding in Watson-Crick pairs of, amongst others, methylated bases, nucleosides and nucleotides are examined.

It is important to note that the experimental AT (or AU) and GC structures were obtained from X-ray diffraction at crystals of sodium adenylyl-3',5'-uridine hexahydrate and sodium guanylyl-3',5'-cytidine nonahydrate.[1c,8] In these crystals, the functional groups of the DNA bases that are involved in Watson-Crick hydrogen bonding can also enter (hydrogen bonding) interactions with water molecules, hydroxyl groups of sugar residues, and Na^+ counterions. We simulate this

environment by incorporating up to six water molecules (modelling both crystal water and sugar OH groups) and one Na^+ ion into our model systems. Also the effect was studied of a sodium ion on the hydrogen bonds in the adenylyl-3',5'-uridine pair, $(ApU)_2$. We are interested in both the structural and energetic consequences and we try to rationalize them in terms of the Kohn-Sham MO model and by analyzing the electron density redistribution associated with particular chemical interactions.

6.2 Theoretical Methods

6.2.1 General Procedure

All calculations were performed using the Amsterdam Density Functional (ADF) program[10] developed by Baerends *et al.*[10a-d] and parallelized[10a] as well as linearized[10e] by Fonseca Guerra *et al.*. The numerical integration was performed using the procedure developed by Boerrigter, te Velde, and Baerends.[10f,g] The MOs were expanded in a large uncontracted set of Slater type orbitals (STOs) containing diffuse functions: DZP and TZ2P.[10h] The DZP basis set is of double ζ quality for all atoms and has been augmented with one set of polarization functions: 3d on C, N, O; and 2p on H. The TZ2P basis set is of triple-ζ quality for all atoms and has been augmented with two sets of polarization functions: 3d and 4f on C, N, O, P; and 2p and 3d on H. The 1s core shell of carbon, nitrogen, oxygen and the 1s 2s 2p core shells of phosphorus were treated by the frozen-core (FC) approximation.[10b] An auxiliary set of s, p, d, f and g STOs was used to fit the molecular density and to represent the Coulomb and exchange potentials accurately in each SCF cycle.[10i]

Energies and geometries were calculated at four different levels of theory: (i) the local density approximation (LDA), where exchange is described by Slater's Xα potential[10j] with $\alpha = \frac{2}{3}$ and correlation is treated in the Vosko-Wilk-Nusair (VWN) parametrization:[10k] (ii) LDA with nonlocal corrections to exchange due to Becke[10l,m] and correlation due to Perdew[10n] added self-consistently[10o] (BP86); (iii) LDA with nonlocal corrections to exchange and correlation due to Perdew and Wang[10p,q] also added self-consistently (PW91); (iv) LDA with nonlocal corrections to exchange due to Becke[10m] and correlation due to Lee-Yang-Parr[10r,s] added, again, self-consistently (BLYP).

Geometries were optimized using analytical gradient techniques.[10t] Frequencies[10u] were calculated by numerical differentiation of the analytical energy gradients and using the nonlocal density functionals. The basis set superposition error (BSSE), associated with the hydrogen bond energy, has been computed via the counterpoise method,[10v] using the individual bases as fragments.

Bond enthalpies at 298.15 K and 1 atmosphere (ΔH_{298}) were calculated from 0 K electronic bond energies (ΔE) according to equation 6.2.1, assuming an ideal gas.[11]

$$\Delta H_{298} = \Delta E + \Delta E_{\text{trans},298} + \Delta E_{\text{rot},298} + \Delta E_{\text{vib},0} + \Delta(\Delta E_{\text{vib}})_{298} + \Delta(pV) \qquad (6.2.1)$$

Here, $\Delta E_{\text{trans},298}$, $\Delta E_{\text{rot},298}$ and $\Delta E_{\text{vib},0}$ are the differences between products and reactants in translational, rotational and zero point vibrational energy, respectively; $\Delta(\Delta E_{\text{vib}})_{298}$ is the change in the vibrational energy difference as one goes from 0 to 298.15 K. The vibrational energy corrections are based on our frequency calculations. The molar work term $\Delta(pV)$ is $(\Delta n)RT$; $\Delta n = -1$ for two fragments combining to one molecule. Thermal corrections for the electronic energy are neglected.

6.2.2 Bond Analysis

The bonding in the various AT and GC model systems was analyzed in the conceptual framework provided by the Kohn-Sham molecular orbital (KS-MO) model[12] using the extended transition state (ETS) method developed by Ziegler and Rauk to decompose the bond energy.[13] The overall bond energy ΔE is made up of two major components (eq 6.2.2).

$$\Delta E = \Delta E_{\text{prep}} + \Delta E_{\text{int}} \qquad (6.2.2)$$

The preparation energy ΔE_{prep} is the amount of energy required to deform the separate molecular fragments (*e.g.*, nucleic bases) from their equilibrium structure to the geometry they acquire in the composite system (*e.g.*, the base pair). The interaction energy ΔE_{int} corresponds to the actual energy change when the prepared fragments are combined to form the composite system. It is further split up into three physically meaningful terms (eq 6.2.3):

$$\Delta E_{\text{int}} = \Delta V_{\text{elstat}} + \Delta E_{\text{Pauli}} + \Delta E_{\text{oi}} \qquad (6.2.3)$$

The term ΔV_{elstat} corresponds to the classical electrostatic interaction between the unperturbed charge distributions of the prepared fragments and is usually attractive. The Pauli-repulsion ΔE_{Pauli} comprises the destabilizing interactions between occupied orbitals and is responsible for the steric repulsion. The orbital interaction ΔE_{oi} accounts for charge transfer (interaction between occupied orbitals on one moiety with unoccupied orbitals of the other, including the HOMO-LUMO interactions) and polarization (empty/occupied orbital mixing on one fragment due to the presence of another fragment). It can be decomposed into the contributions from each irreducible representation Γ of the interacting system (eq. 6.2.4).[13] In systems with a clear σ, π separation (*e.g.*, flat DNA base pairs) this symmetry partitioning proves to be most informative.

$$\Delta E_{\text{oi}} = \sum_{\Gamma} \Delta E_{\Gamma} \tag{6.2.4}$$

6.2.3 Analysis of the Charge Distribution

The electron density distribution is analyzed using the Voronoi deformation density (VDD) method introduced in Ref. [14] and further developed in Ref. [3] to enable a correct treatment of even the subtle changes in atomic charges ΔQ_A^{VDD} caused by weak chemical interactions (such as hydrogen bonds) between molecular fragments as well as a decomposition into the contributions from the σ- and π-electron systems. VDD atomic charges ΔQ_A^{VDD} are defined and related to the deformation density $\rho_{\text{total system}}(\mathbf{r}) - \rho_{\text{subsystem1}}(\mathbf{r}) - \rho_{\text{subsystem2}}(\mathbf{r})$ by equation 6.2.5.[3]

$$\Delta Q_A^{\text{VDD}} = - \int_{\substack{\text{Voronoi cell of A} \\ \text{in total system}}} \Big(\rho_{\text{total system}}(\mathbf{r}) - \rho_{\text{subsystem1}}(\mathbf{r}) - \rho_{\text{subsystem2}}(\mathbf{r}) \Big) d\mathbf{r} \tag{6.2.5}$$

The interpretation of VDD atomic charges is rather straightforward. Instead of measuring the amount of charge associated with a particular atom A, they directly monitor how much charge flows, due to chemical interactions, out of ($\Delta Q_A^{\text{VDD}} > 0$) or into ($\Delta Q_A^{\text{VDD}} < 0$) the Voronoi cell of atom A, that is, the region of space that is closer to nucleus A than to any other nucleus. The Voronoi cell of atom A is bounded by the bond midplanes on and perpendicular to all bond axes between nucleus A and its neighbouring nuclei (cf. the Wigner-Seitz cells in crystals).[10g,15]

Table 6.1. Hydrogen-bond strength (kcal/mol) in AT (plain nucleic bases, unless stated otherwise).

	ΔE^{AT}	ΔE^{AT} (BSSE)	ΔH^{AT}_{298}
Experiment[a]			–12.1
DFT with STO basis (this work)			
BP86/TZ2P (C_1)[b]	–13.0	–12.3	–11.8
BP86/TZ2P (pair: C_s, bases: C_1)[c]	–13.0	–12.3	
BP86/TZ2P (C_s)[d]	–13.0	–12.3	
BP86/DZP (C_1)[b]	–15.8	–12.7	
PW91/TZ2P (C_1)[b]	–15.2	–14.5	–14.0[j]
BLYP/TZ2P (C_1)[b]	–14.5	–13.7	–13.2[j]
DFT with GTO basis (others)			
BP86/DZVP (C_1)[b,e]		–13.9	
B3LYP/6-31G** (C_1)[b,f[		–12.3	–10.9
Ab Initio with GTO basis (others)			
MP2/DZP//HF/6-31G* (C_1)[b,g]		–14.7	–11.9
MP2/6-31G*(0.25)//HF/6-31G** (C_s)[d,h]		–11.8	–9.5
LMP2/cc-pVTZ(-f)//HF/cc-pVTZ(-f) (C_1)[i]		–10.8	–10.2

[a] ΔH_{exp} from mass spectrometry data of Yanson et al.[6] for AT with 9-methyladenine and 1-methylthymine (**1c**) with corrections according to Brameld *et al.*[7e]

[b] Full optimization in C_1 symmetry of base pair and bases.

[c] Base pair optimized in C_S (**1a**) and bases in C_1 symmetry.

[d] Optimization in C_s symmetry of base pair (**1a**) and bases.

[e] Santamaria *et al.*[7a]

[f] Bertran *et al.*[7b]

[g] Gould *et al.*;[7c] AT with 9-methyladenine and 1-methylthymine (**1c**).

[h] Sponer *et al.*[7d,f]

[i] Brameld *et al.*;[7e] AT with 9-hydroxymethyladenine and 1-hydroxymethylthymine. Base pair optimized in C_1 and bases in C_s symmetry.

[j] ΔH^{AT}_{298} was obtained with thermal energy corrections from BP86/TZ2P (C_1)

Table 6.2. Hydrogen-bond strength (kcal/mol) in GC (plain nucleic bases, unless stated otherwise).

	ΔE^{GC}	ΔE^{GC} (BSSE)	ΔH_{298}^{GC}
Experiment[a]			–21.0
DFT with STO basis (this work)			
BP86/TZ2P (C_1)[b]	–26.1	–25.2	–23.8
BP86/TZ2P (pair: C_s, bases: C_1)[c]	–26.1	–25.2	
BP86/TZ2P (C_s)[d]	–26.5	–25.7	
BP86/DZP (C_1)[b]	–28.8	–25.1	
PW91/TZ2P (C_1)[b]	–28.5	–27.7	–26.3[j]
BLYP/TZ2P (C_1)[b]	–28.3	–27.4	–26.0[j]
DFT with GTO basis (others)			
BP86/DZVP (C_1)[b,e]		–27.7	
B3LYP/6-31G** (C_1)[b,f]		–25.5	–24.0
Ab Initio with GTO basis (others)			
MP2/DZP//HF/6-31G* (C_1)[b,g]		–28.0	–25.4
MP2/6-31G*(0.25)//HF/6-31G** (C_s)[d,h]		–23.8	–20.8
LMP2/cc-pVTZ(-f)//HF/cc-pVTZ(-f) (C_1)[i]		–22.4	–21.2

[a] ΔH_{exp} from mass spectrometry data of Yanson et al.[6] for GC with 9-methylguanine and 1-methylcytosine (**2b**).

[b] Full optimization in C_1 symmetry of base pair and bases.

[c] Base pair optimized in C_s (**2a**) and bases in C_1 symmetry.

[d] Optimization in C_s symmetry of base pair (**2a**) and bases.

[e] Santamaria *et al.*[7a]

[f] Bertran *et al.*[7b]

[g] Gould *et al.*;[7c] GC with 9-methylguanine and 1-methylcytosine (**2b**).

[h] Sponer *et al.*[7d,f]

[i] Brameld *et al.*;[7e] GC with 9-hydroxymethylguanine and 1-hydroxymethylcytosine. Base pair optimized in C_1 and bases in C_s symmetry.

[j] ΔH_{298}^{GC} was obtained with thermal energy corrections from BP86/TZ2P (C_1).

6.3 Watson-Crick Pairs of Plain Nucleic Bases

6.3.1 Hydrogen Bond Strength

To examine the performance of the different density functionals and STO basis sets, we have studied the formation of the plain adenine–thymine and guanine–cytosine complexes (see Scheme 1) at five different levels of theory: BP86/TZ2P, BLYP/TZ2P, PW91/TZ2P, BP86/DZP and also LDA/TZ2P. In Tables 6.1-6.4, the results of our calculations are summarized and compared with those from other theoretical[7] and experimental[1c,6,8] studies. The LDA functional leads to overbinding (i.e., hydrogen bonds are too short and too strong at LDA/TZ2P; data not shown here), in line with general experience, and will not be further discussed.

Tables 6.1 and 6.2 show that BP86/TZ2P provides A–T and G–C bond enthalpies (–11.8 and –23.8 kcal/mol) that agree within 0.3 and 2.8 kcal/mol, respectively, with the corresponding gas-phase experimental[6] values (–12.1 and –21.0 kcal/mol). The PW91/TZ2P and the BLYP/TZ2P levels yield bond enthalpies that deviate somewhat more from experiment: they are too strongly binding by 1-2 kcal/mol for AT and by some 5 kcal/mol for GC. Also the DFT and *ab initio* results of others agree reasonably well with experiment with (Tables 6.1 and 6.2).

Whereas the basis set superposition error (BSSE) is less than a kcal/mol in the case of the TZ2P basis, it rises to 3.7 kcal/mol if the smaller DZP basis is used [compare BP86/TZ2P (C_1) and BP86/DZP (C_1) entries in Tables 6.1 and 6.2]. Note, however, that the BSSE corrected BP86/DZP values agree within 0.4 (AT) or 0.1 kcal/mol (GC). This suggests that we can use the DZP basis as a suitable and efficient alternative to the much larger TZ2P basis for studying very large systems involving hydrogen bonds, provided that energies are corrected for the BSSE.

We have also studied the effect of symmetry constraints on the base-pairing energies by examining, at BP86/TZ2P, three different situations (see Tables 6.1 and 6.2): (i) both the base pair and the separate bases are fully optimized in C_1 symmetry; (ii) a C_s symmetry constraint applies to the base pair but the bases are still fully optimized in C_1 symmetry; and (iii) a C_s symmetry restriction applies to both base pair and bases. There appears to be virtually no difference between base-pairing energies computed according to (i) and (ii) but those of (iii) deviate slightly (by half a kcal/mol) in the case of GC. This is not difficult to understand: whereas the Watson-Crick base pairs and also, although to a slightly lesser extent, the bases adenine,

thymine and cytosine are nearly planar, in guanine the pyramidalization of the N2 amino group is quite pronounced (see section 6.3.2). Thus, reliable bond energies can be obtained efficiently from C_s-optimized Watson-Crick pairs (provided the separate bases are fully optimized). These results also show that the A–T and G–C bond analyses can be carried out in C_s symmetry, which allows for a decomposition of orbital interactions into σ and π contributions (eq 6.2.4).

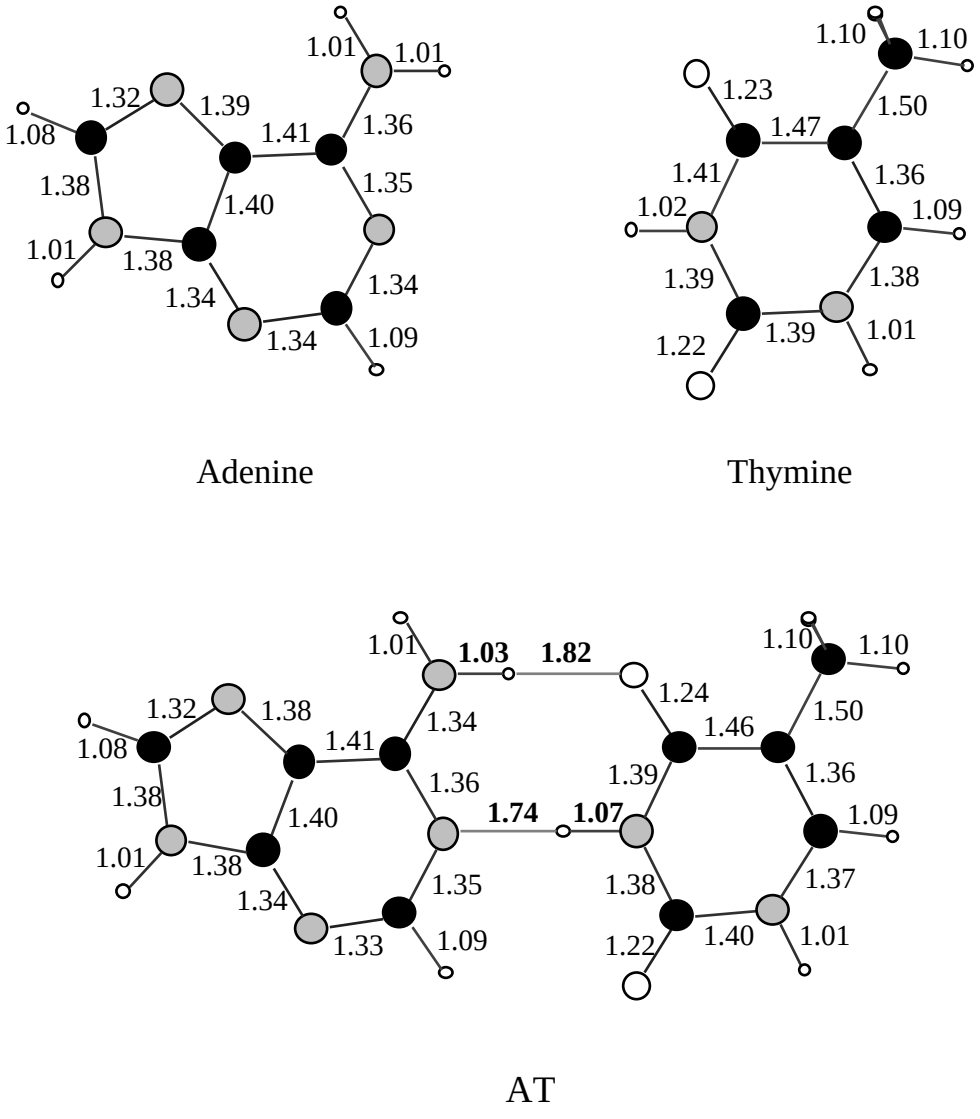


Figure 6.1. Bond distances (Å) from BP86/TZ2P calculations without any symmetry constraint for adenine, thymine and the Watson-Crick pair AT (see Scheme 1).

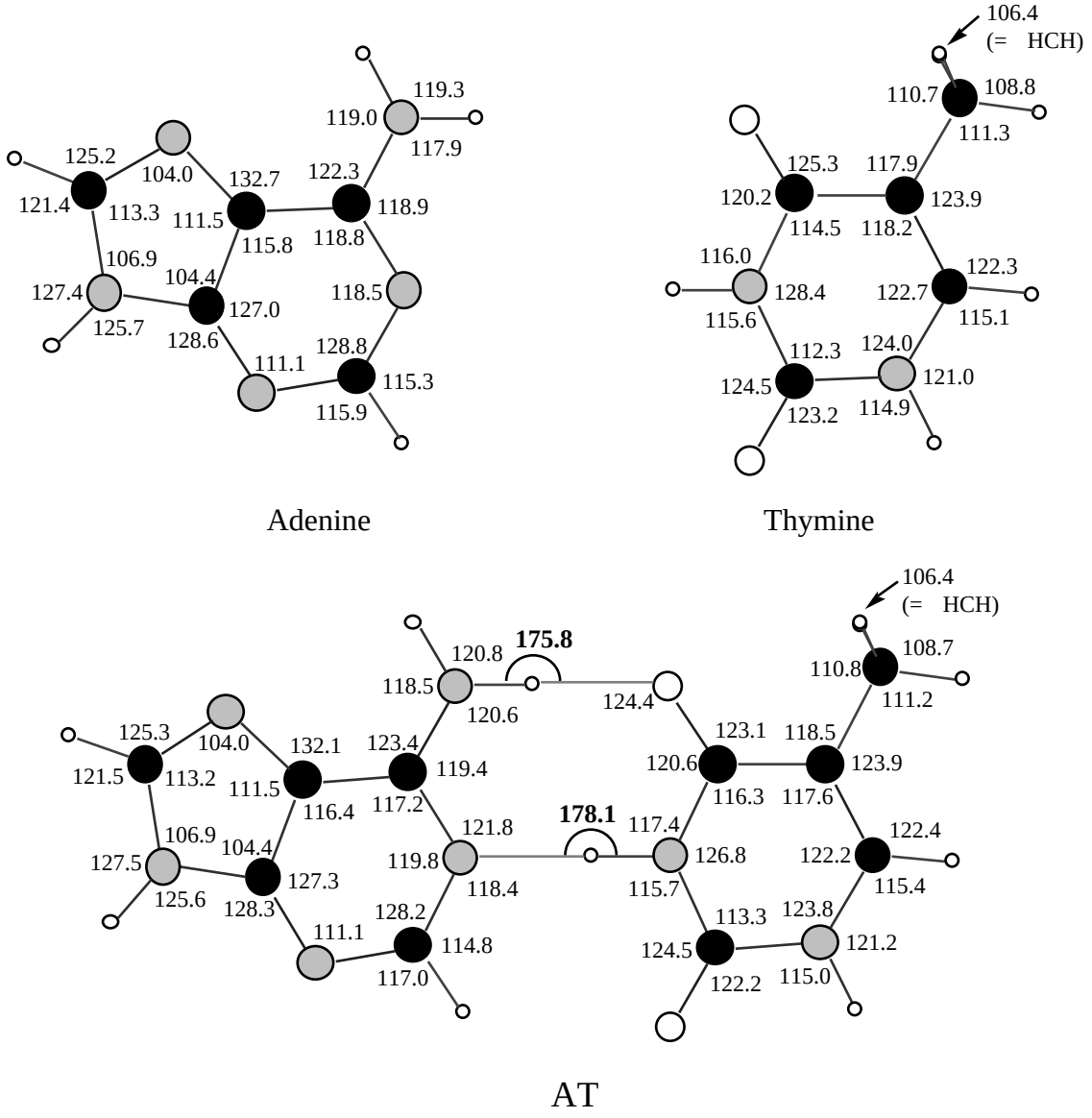


Figure 6.2. Bond angles (in degrees) from BP86/TZ2P calculations without symmetry constraints for adenine, thymine and the Watson-Crick pair AT (see Scheme 1).

6.3.2 Structure of Bases and Watson-Crick Pairs

The BP86/TZ2P geometries of Watson-Crick pairs and separate bases are shown in Figures 6.1-6.4. All structures have been verified to be energy minima through vibrational analyses (no imaginary frequencies). As mentioned above, the amino groups of adenine (H6'N6C6C5 and H6N6C6N1 are 11.9° and 11.4°) and cytosine (H4'N4C4C5 and H4N4C4N3 are 15.1° and

10.6°) are only slightly pyramidal whereas that of guanine (H2'N2C2N3 and H2N2C2N1 are 13.2° and 33.1°!) is more strongly pyramidalized in line with previous results[16] (we give absolute values of dihedral angles). The base pairs, in which the amino groups adopt a planar conformation, deviate only slightly from C_s symmetry. Furthermore, the hydrogen bonds in AT (N6H6O4 and N1H3N3 are 175.8° and 178.1°) and GC (O6H4N4, N1H1N3 and N2H2O2 are 178.7°, 178.2° and 179.4°) are essentially linear (see Figures 6.2 and 6.4).

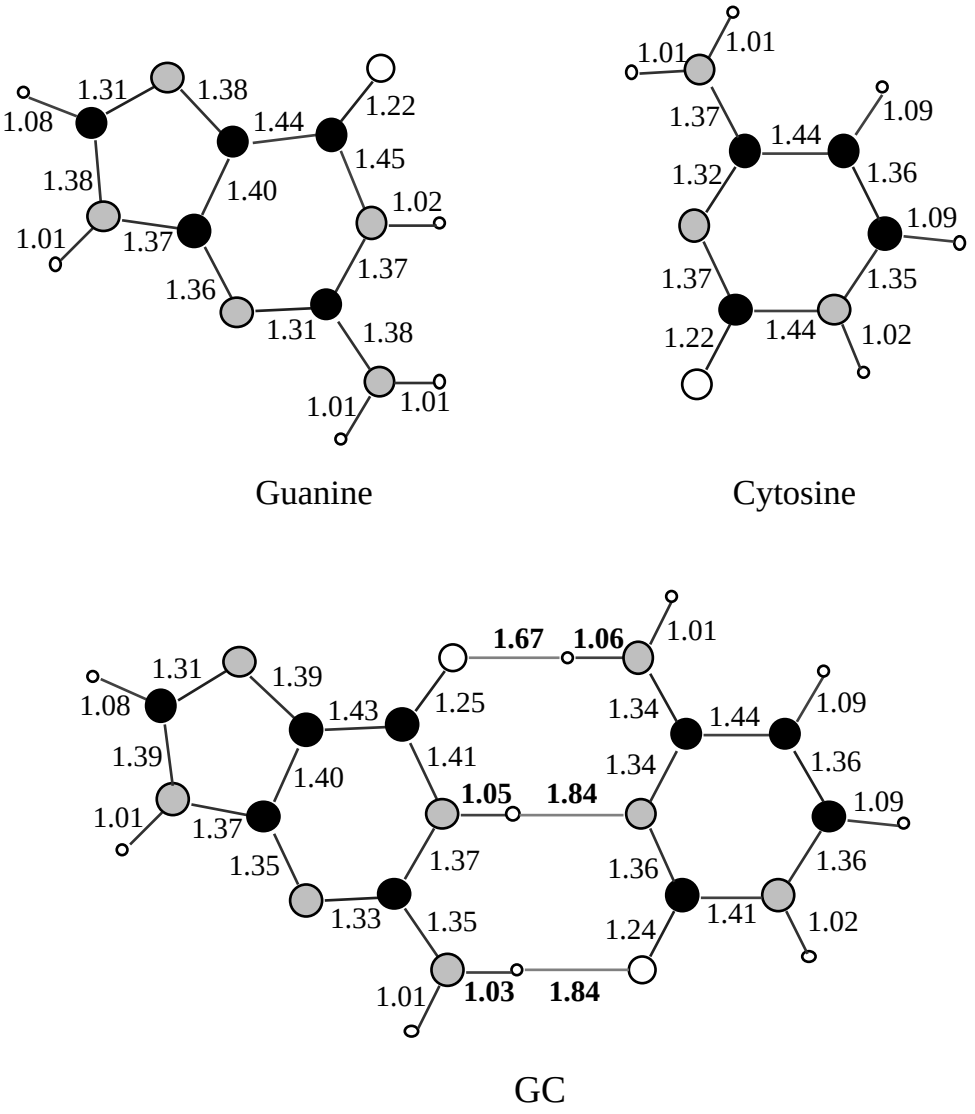


Figure 6.3. Bond distances (Å) from BP86/TZ2P calculations without any symmetry constraint for guanine, cytosine and the Watson-Crick pair GC (see Scheme 1).

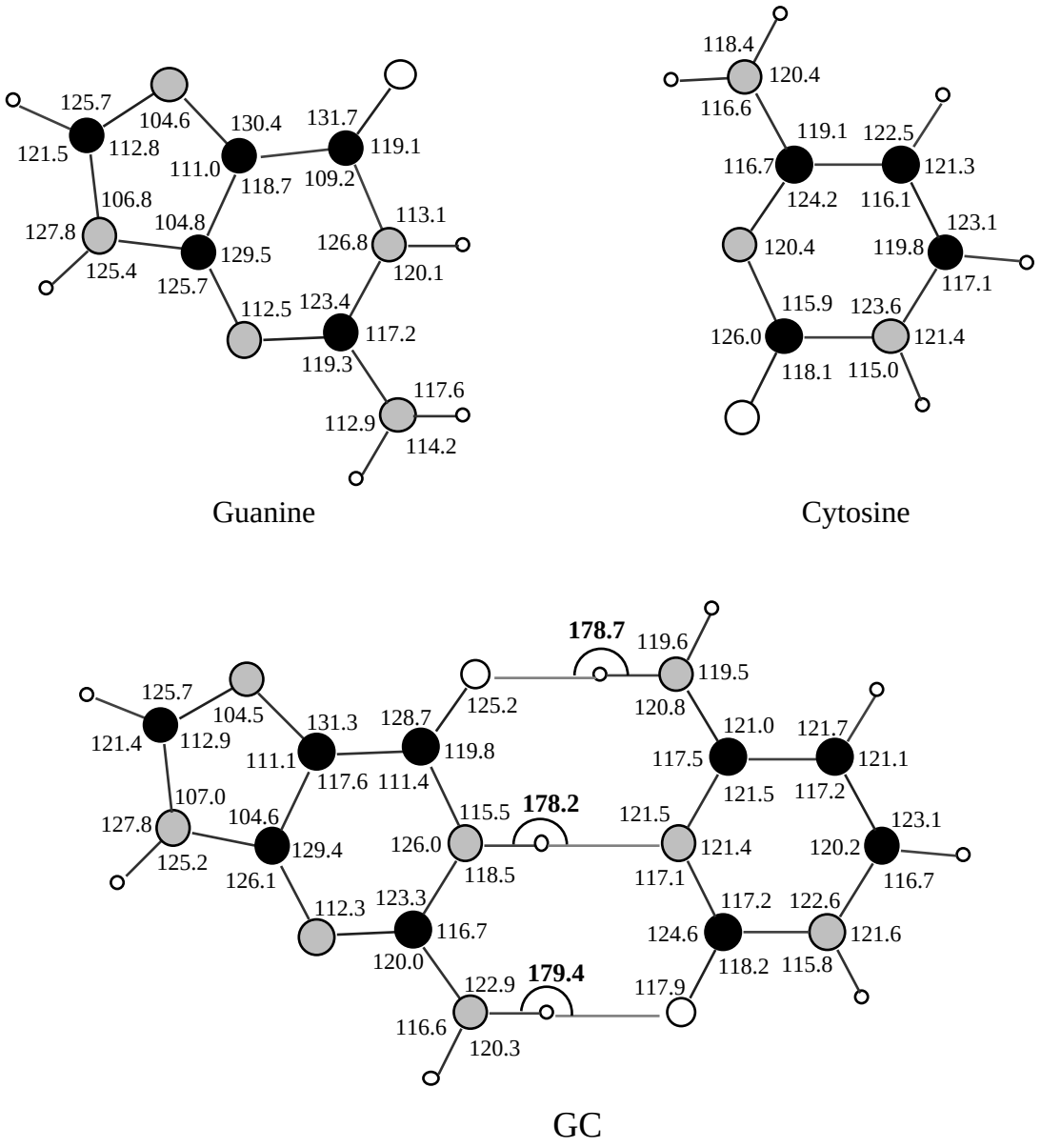


Figure 6.4. Bond angles (in degrees) from BP86/TZ2P calculations without symmetry constraints for guanine, cytosine and the Watson-Crick pair GC (see Scheme 1).

Regarding the Watson-Crick hydrogen bond distances, we arrive at striking discrepancies with experimental structures (see Tables 6.3 and 6.4). At BP86/TZ2P, we find N6-O4 and N1-N3 hydrogen-bond distances in AT of 2.85 and 2.81 Å; this result is not influenced by applying a C_s symmetry constraint (Table 6.3). Only slightly shorter N6-O4 and N1-N3 distances are obtained with the smaller DZP basis (i.e. BP86/DZP) and with the other nonlocal functional (PW91/TZ2P

and BLYP/TZ2P). These values have to be compared with 2.95 and 2.82 Å from experiment.[8a] Even more eye-catching, as can be seen in Table 6.4, is the situation for the three hydrogen bonds in GC, i.e. O6–N4, N1-N3 and N2-O2, for which we find a bond length pattern that is short-long-long, i.e., 2.73, 2.88 and 2.87 Å at BP86/TZ2P, at significant variance with the experimental values[8b] which are long-long-short (2.91, 2.95 and 2.86 Å). Again, the smaller DZP basis (i.e. BP86/DZP) and the other nonlocal functionals (PW91/TZ2P and BLYP/TZ2P) perform comparably well, yielding hydrogen bonds that are only slightly (i.e. 0.01-0.03 Å) shorter.

This disagreement between theoretical and experimental[1c,8] Watson-Crick hydrogen bond length is not new. It has been encountered before in several DFT and *ab initio* studies (see Tables 6.3 and 6.4).[7] For example, Hartree-Fock (HF)[7c-e] gives hydrogen bonds that are up to 0.2 Å longer than both our computed and the experimental[8] values and in case of GC the wrong bond length pattern of long-long-short is found (Table 6.4). Furthermore, whereas the DFT results of Bertran et al.[7b] and Santamaria et al.[7a] for AT are in good agreement with experimental structures,[8a] their geometries for GC differ again significantly from experimental ones.[8b] Difference between our and the other DFT geometries (the latter are up to 0.1 Å longer) can be ascribed, amongst others, to the use of different basis sets: STOs in our calculations and GTOs in those of Bertran et al.[7b] and Santamaria et al..[7a]

It is important to realize that the experimental structures stem from X-ray diffraction measurements on crystals of sodium adenylyl-3',5'-uridine hexahydrate (**1**)[8a] for AT (or AU) and sodium guanylyl-3',5'-cytidine nonahydrate (**2**)[8b] for GC. The base pairs in these crystals differ from the theoretical model systems studied so far, in two important fashions: (i) they are part of a small double helix consisting of two base pairs in which bases along a strand are connected via a ribose–phosphate–ribose backbone, and (ii) they experience interactions with the environment in the crystal, in particular water molecules, ribose OH groups and counterions. In view of the very shallow potential energy surfaces that we find for Watson-Crick base pairing, it seems plausible that the effects of the backbone and the molecular environment in the crystal could cause the discrepancy with more simplistic AT and GC models. This has led us to study the effect of the backbone (section 6.4) and the molecular environment (section 6.5) at the BP86/TZ2P level which yields our best hydrogen bond enthalpies (see section 6.3.1).

Table 6.3. Hydrogen bond distances (Å) in AT (plain nucleic bases, unless stated otherwise).

	N6–O4	N1–N3
Experiment[a]		
A2U1	2.95	2.82
A1U2	2.93	2.85
DFT with STO basis (this work)		
BP86/TZ2P (C_1)[b]	2.85	2.81
BP86/TZ2P (C_s)[c]	2.85	2.81
BP86/DZP (C_1)[d]	2.84	2.79
PW91/TZ2P (C_1)[b]	2.85	2.79
BLYP/TZ2P (C_1)[b]	2.84	2.78
DFT with GTO basis (others)		
BP86/DZVP (C_1)[b.d]	2.95	2.87
B3LYP/6-31G** (C_1)[b,e]	2.94	2.84
Ab Initio with GTO basis (others)		
HF/6-31G* (C_1)[b.f	3.08	3.01
HF/6-31G** (C_s)[c,g]	3.09	2.99
HF/cc-pVTZ(-f) (C_1)[b,h]	3.06	2.92

[a] X-ray crystallographic measurements by Seeman et al.[8a] on sodium adenylyl-3',5'-uridine hexahydrate (**1**) containing the Watson-Crick-type dimer $(ApU)_2$. There are two values for each hydrogen bond length because the two AU pairs (A2U1 and A1U2) have different environments (see Scheme 2).

[b] Full optimization in C_1 symmetry.

[c] Optmization in C_s symmetry (**1a**).

[d] Santamaria *et al.*[7a]

[e] Bertran *et al.*[7b]

[f] Gould *et al.*;[7c] AT with 9-methyladenine and 1-methylthymine (**1c**).

[g] Sponer *et al.*[7d]

[h] Brameld *et al.*;[7e] AT with 9-hydroxymethyladenine and 1-hydroxymethylthymine.

Table 6.4. Hydrogen bond distances (Å) in GC (plain nucleic bases, unless stated otherwise).

	O6–N4	N1–N3	N2–O2
Experiment			
GpC [a]	2.91	2.95	2.86
DFT with STO basis (this work)			
BP86/TZ2P (C_1)[b]	2.73	2.88	2.87
BP86/TZ2P (C_s)[c]	2.73	2.88	2.87
BP86/DZP (C_1)[b]	2.71	2.87	2.87
PW91/TZ2P (C_1)[b]	2.72	2.88	2.87
BLYP/TZ2P (C_1)[b]	2.71	2.86	2.84
DFT with GTO basis (others)			
BP86/DZVP (C_1)[b,d]	2.78	2.93	2.93
B3LYP/6-31G** (C_1)[b,e]	2.79	2.93	2.92
Ab Initio with GTO basis (others)			
HF/6-31G* (C_1)[b.f]	2.93	3.05	3.01
HF/6-31G** (C_s)[c,g]	2.92	3.04	3.02
HF/cc-pVTZ(-f) (C_1)[b,h]	2.83	2.95	2.92

[a] X-ray crystallographic measurements by Rosenberg et al.[8b] on sodium guanylyl-3',5'-cytidine nonahydrate (**2**) containing the Watson-Crick-type dimer $(GpC)_2$.

[b] Full optimization in C_1 symmetry.

[c] Optmization in C_s symmetry (**2a**).

[d] Santamaria *et al.*[7a]

[e] Bertran *et al*.[7b]

[f] Gould *et al.*;[7c] GC with 9-methylguanine and 1-methylcytosine (**2b**).

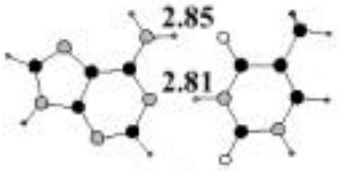
[g] Sponer *et al*.[7d]

[h] Brameld *et al.*;[7e] GC with 9-hydroxymethylguanine and 1-hydroxymethylcytosine.

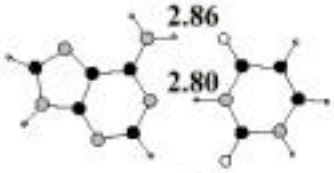
6.4 The Effect of the Backbone

The effect of the backbone on Watson-Crick base pairing is studied by stepwise going from Watson-Crick pairs of plain nucleic bases via nucleotides to strands consisting of two nucleotides (**1a**-**1e**, **1k**, **2a**-**2d**). The results are summarized in Figures 6.5-6.7 (geometries) and Tables 6.5 and 6.6 (Watson-Crick-pairing energies). In the first place, comparison of plain AT (**1a**) and AU (**1b**) shows that methylation at C5 of uracil has basically no influence on Watson-Crick pairing, i.e., hydrogen-bond distances and energies ΔE_{int} (eqs 6.2.4 and 6.2.5) differ by only 0.01 Å and 0.2 kcal/mol, respectively. The same holds for methylation at the positions where the glucosidic N–C bond occurs in nucleosides (N9 in A en G, N1 in T and C): with bases methylated in this way, AT (**1c**) and GC(**2b**) hydrogen-bond distances and energies differ only by up to 0.01 Å and 0.3 kcal/mol, respectively, from **1a** and **2a**.

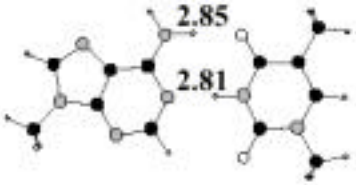
Similarly, only very small effects occur on substituting hydrogen by 2'-deoxyribose (**1d** and **2c**) or neutral 2'-deoxyribose-5'-phosphate residues (**1e** and **2d**) at the N9 and N1 atoms of the purine and pyrimidine bases, respectively. For AT, the hydrogen-bond energy decreases by only 0.3 kcal/mol on going from the Watson-Crick pair of plain nucleic bases (**1a**, –13.0 kcal/mol) to those of either nucleosides or nucleotides (**1d** and **1e**, both –12.7 kcal/mol; Table 6.5). At the same time, the N6–O4 and N1–N3 hydrogen-bond distances go from 2.85 and 2.81 Å (**1a**) to 2.87 and 2.77 Å in **1d** and to 2.83 and 2.76 Å in **1e** (Figure 6.5). This does not resolve the discrepancy with the experimental values of 2.95 and 2.82 Å (**1**). And again, for GC, the hydrogen-bond energy changes only slightly as we go from the Watson-Crick pair of plain nucleic bases (**2a**, –26.1 kcal/mol) to those of either nucleosides or nucleotides (**2c** and **2d**, both –25.3 kcal/mol; Table 6.6). The O6–N4, N1–N3 and N2–O2 hydrogen-bond distances change only by up to 0.03 Å along **2a**, **2c** and **2d** (Figure 6.6). Thus, we still have the erroneous bond-length pattern of short-long-long at variance with the experimental order of long-long-short (**2**).



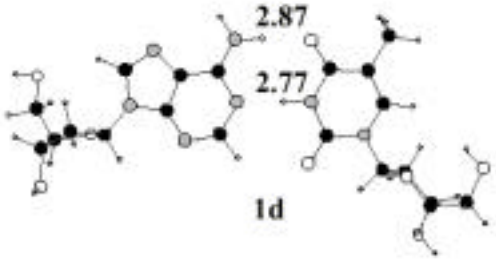
1a



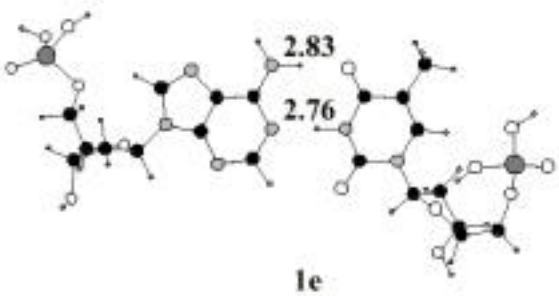
1b



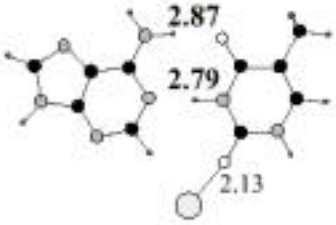
1c



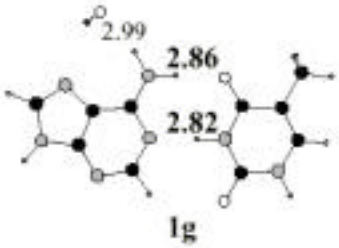
1d



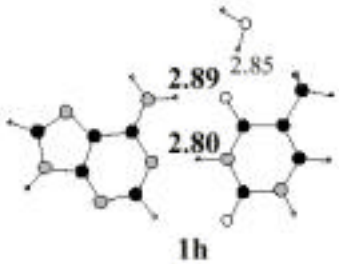
1e



1f



1g



1h

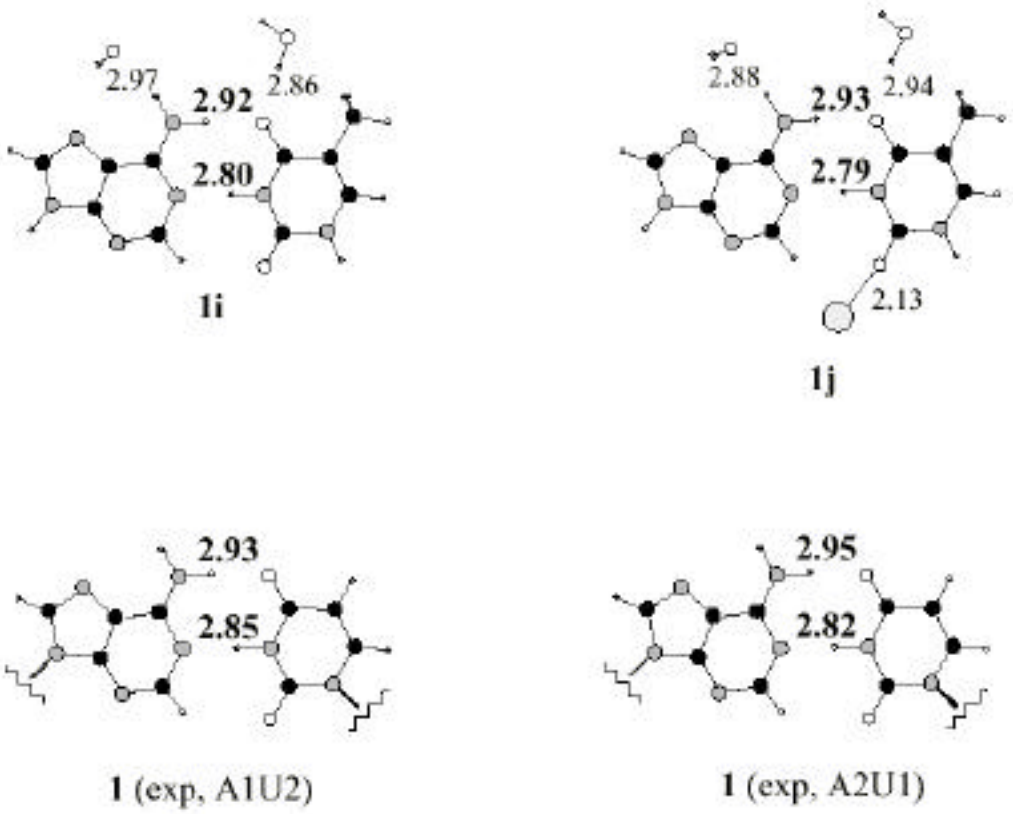
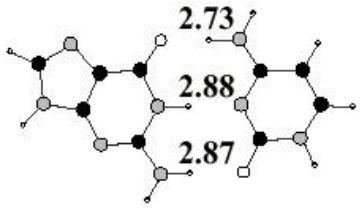
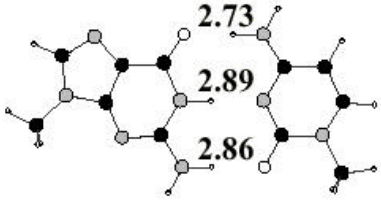


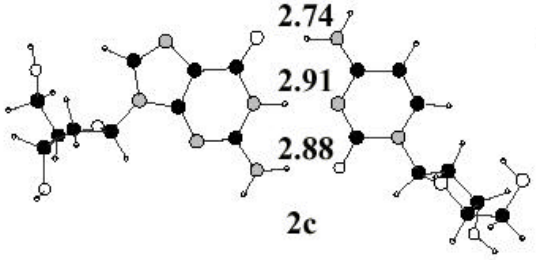
Figure 6.5. N6–O4 and N1–N3 distances in AT (**1a**), AU (**1b**), methylated AT (**1c**), AT with deoxyribose residues (**1d**), AT with deoxyribose 5'-phosphate residues (**1e**) and various AT crystal model systems (**1f-1j**) from BP86/TZ2P (**1a-1d**, **1f-1j**) and BP86/DZP (**1e**) computations, and from the X-ray crystal structure of sodium adenylyl-3',5'-uridine hexahydrate (**1**).[8a] Geometries of **1d-1e** were optimized without any symmetry constraint whereas for the other systems (**1a-1c** and **1f-1i**) C_S symmetry has been used. We also show the distances between the oxygen of water and the proton-donor or proton-acceptor atom of the bases, and those between Na^+ and O2 of thymine. Note that there are two experimental values for both N6–O4 and N1–N3 because the two AU pairs in the crystal of **1** (see Scheme 2) experience different environments.



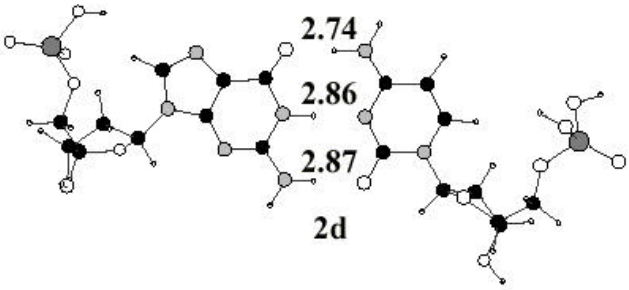
2a



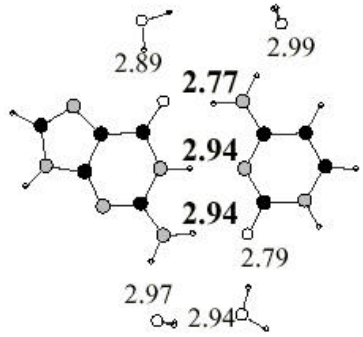
2b



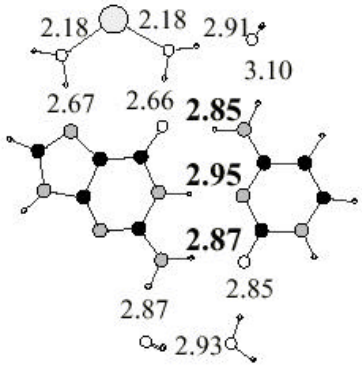
2c



2d



2e



2f

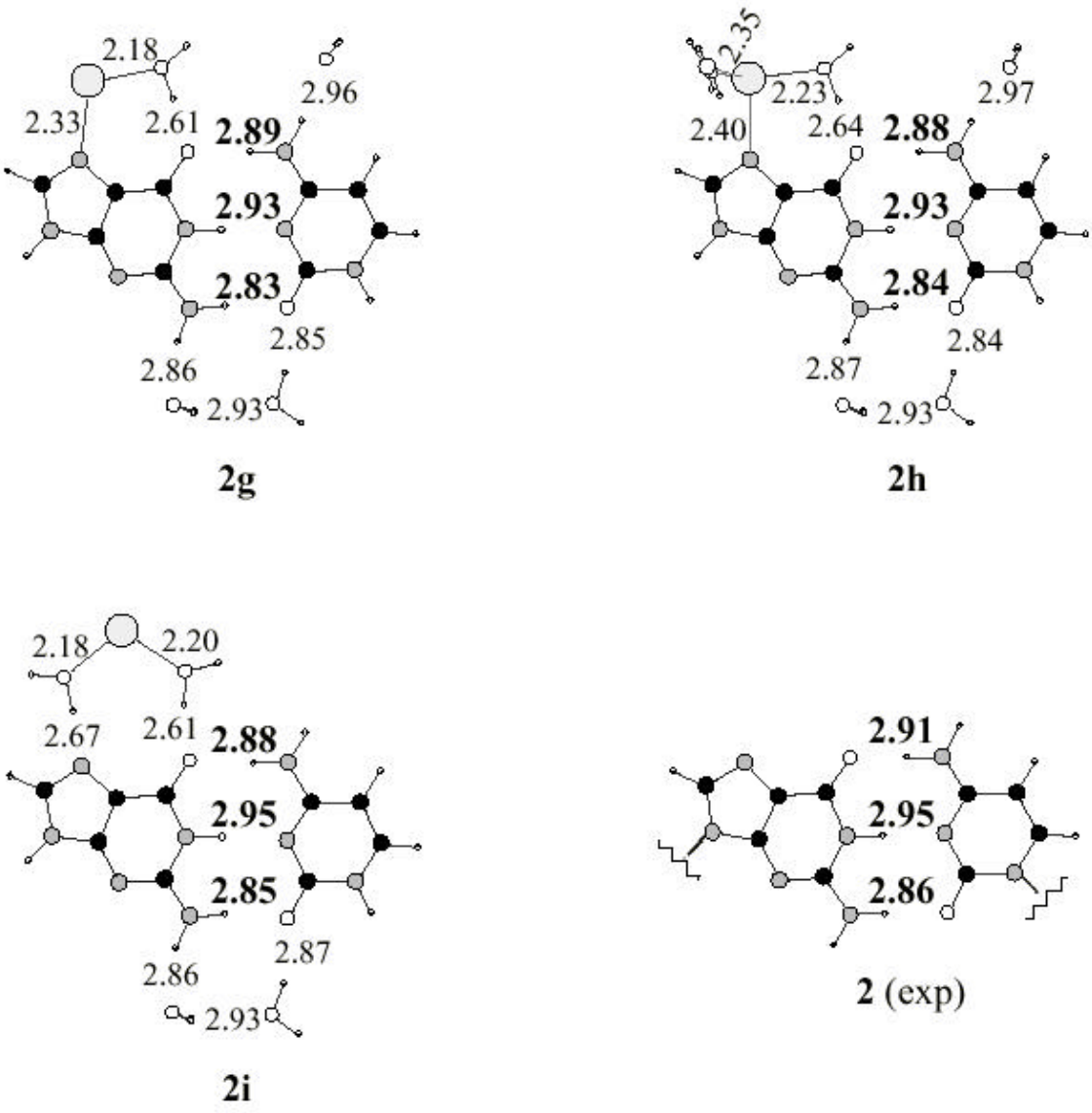


Figure 6.6. O6–N4, N1–N3 and N2–O2 distances in GC (**2a**), methylated GC (**2b**), GC with deoxyribose residues (**2c**), GC with deoxyribose 5'-phosphate residues (**2d**) and various GC crystal model systems (**2e**-**2i**) from BP86/TZ2P (**2a**-**2c**, **2e**-**2i**) and BP86/DZP (**2d**) computations, and from the X-ray crystal structure of sodium guanylyl-3',5'-cytidine nonahydrate (**2**).[8b] Geometries of **2c**-**2d** were optimized without any symmetry constraint whereas for the other systems (**2a**-**2b** and **2e**-**2i**) C_s symmetry has been used. We also show the distances between the oxygen of water and the proton-donor or proton-acceptor atom of the bases, those between Na^+ and lone-pair donating atoms of guanine or water molecules, and those between oxygen atoms of water molecules.

Table 6.5. Analysis of the A-T interaction (kcal/mol) in **1a-1k** (with environment effects in **1f-1j**).[a]

	1a	**1b**	**1c**	**1d**	**1e**[b]	**1f**	**1g**	**1h**	**1i**	**1j**	**1k**[b]
ΔE_{Pauli}	38.7	38.7	38.3	41.1	45.0	39.9	37.5	37.5	37.0	38.5	89.7
ΔV_{elstat}	–31.8	–32.0	–31.7	–32.9	–35.1	–33.4	–30.7	–30.7	–30.2	–33.2	–69.1
$\Delta E_{\text{Pauli}}+\Delta V_{\text{elstat}}$	6.9	6.7	6.6	8.2	9.9	6.5	6.8	6.9	6.8	5.3	20.6
ΔE_{σ}	–20.4	–20.5	–20.1			–24.3	–19.5	–19.8	–19.2	–23.2	
ΔE_{π}	–1.7	–1.7	–1.7			–4.2	–1.6	–1.6	–1.5	–4.0	
ΔE_{oi}	–22.1	–22.2	–21.8	–23.3	–25.9	–28.5	–21.1	–21.4	–20.7	–27.2	–50.7
ΔE_{int}	–15.2	–15.5	–15.2	–15.1	–16.0	–22.0	–14.3	–14.5	–13.9	–21.9	–30.1
ΔE_{prep}	2.2	2.3	2.1	2.4	3.3						9.2
ΔE	–13.0	–13.2	–13.1	–12.7	–12.7						–20.9

[a] BP86/TZ2P. See Figures 5 and 7. All bond energies relative to bases fully optimized in C_1 symmetry.

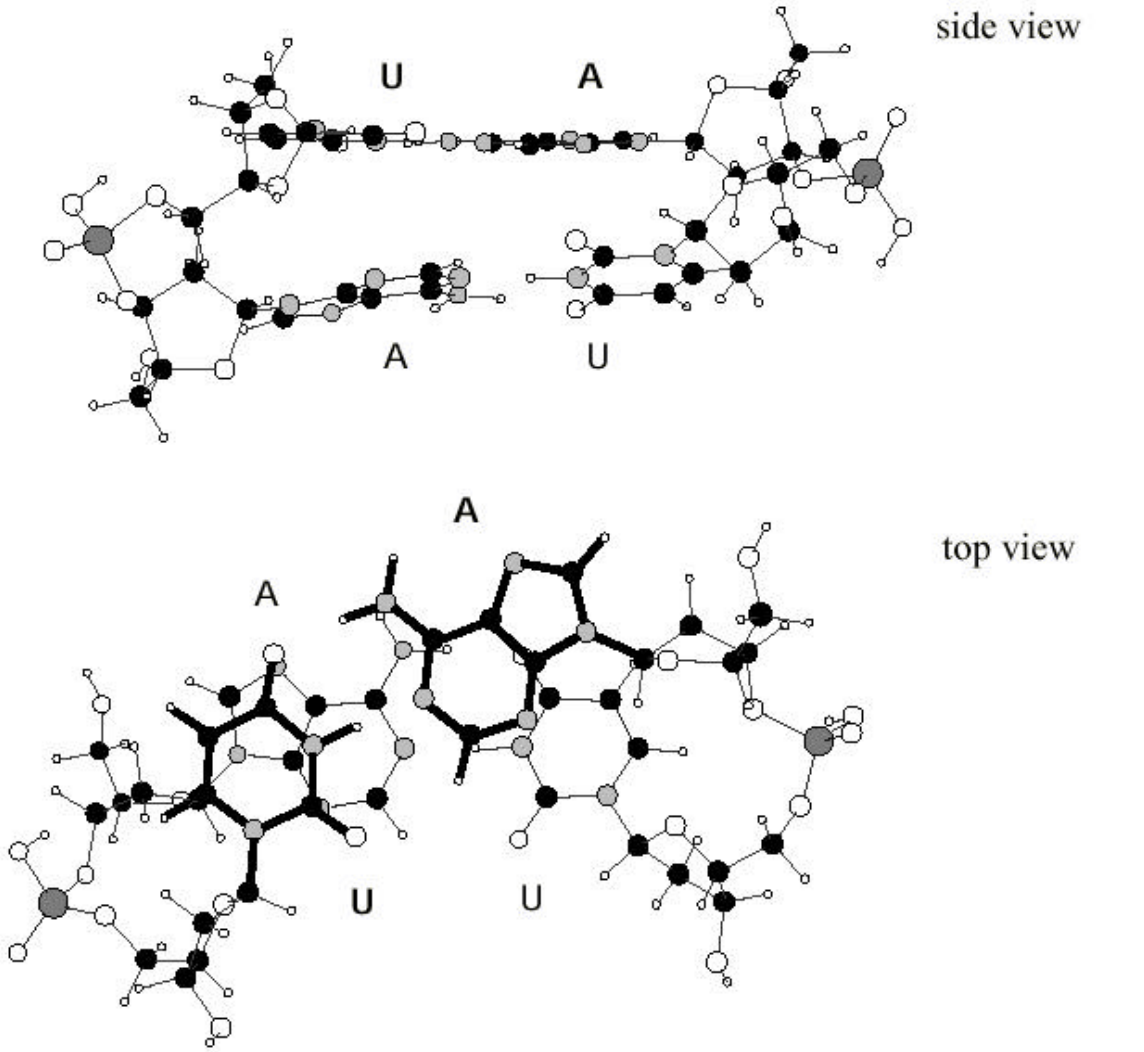
[b] BP86/TZ2P//BP86/DZP

Table 6.6. Analysis of the G-C interaction (kcal/mol) in **2a**-**2d** (with environment effects in **2e**-**2i**).[a]

	2a	**2b**	**2c**	**2d**[b]	**2e**	**2f**	**2g**	**2h**	**2i**
ΔE_{Pauli}	51.9	51.1	48.6	53.7	47.7	51.9	44.7	45.0	43.2
ΔV_{elstat}	–48.5	–47.8	–46.0	–48.6	–50.7	–56.0	–51.6	–51.4	–46.5
$\Delta E_{\text{Pauli}}+\Delta V_{\text{elstat}}$	3.4	3.3	2.6	5.1	–3.0	–4.1	–6.9	–6.4	–3.3
ΔE_{σ}	–29.2	–28.9			–25.1	–28.1	–24.5	–24.5	–23.5
ΔE_{π}	–4.8	–4.6			–4.8	–4.1	–4.6	–4.6	–4.2
ΔE_{oi}	–34.0	–33.5	–32.1	–35.2	–29.9	–32.2	–29.1	–29.1	–27.7
ΔE_{int}	–30.6	30.2	–29.5	–30.1	–32.9	–36.3	–36.0	–35.5	–31.0
ΔE_{prep}	4.5	4.4	4.2	4.8					
ΔE	–26.1	–25.8	–25.3	–25.3					

[a] BP86/TZ2P. See Figure 6. All bond energies relative to bases fully optimized in C_1 symmetry.

[b] BP86/TZ2P//BP86/DZP.



1k

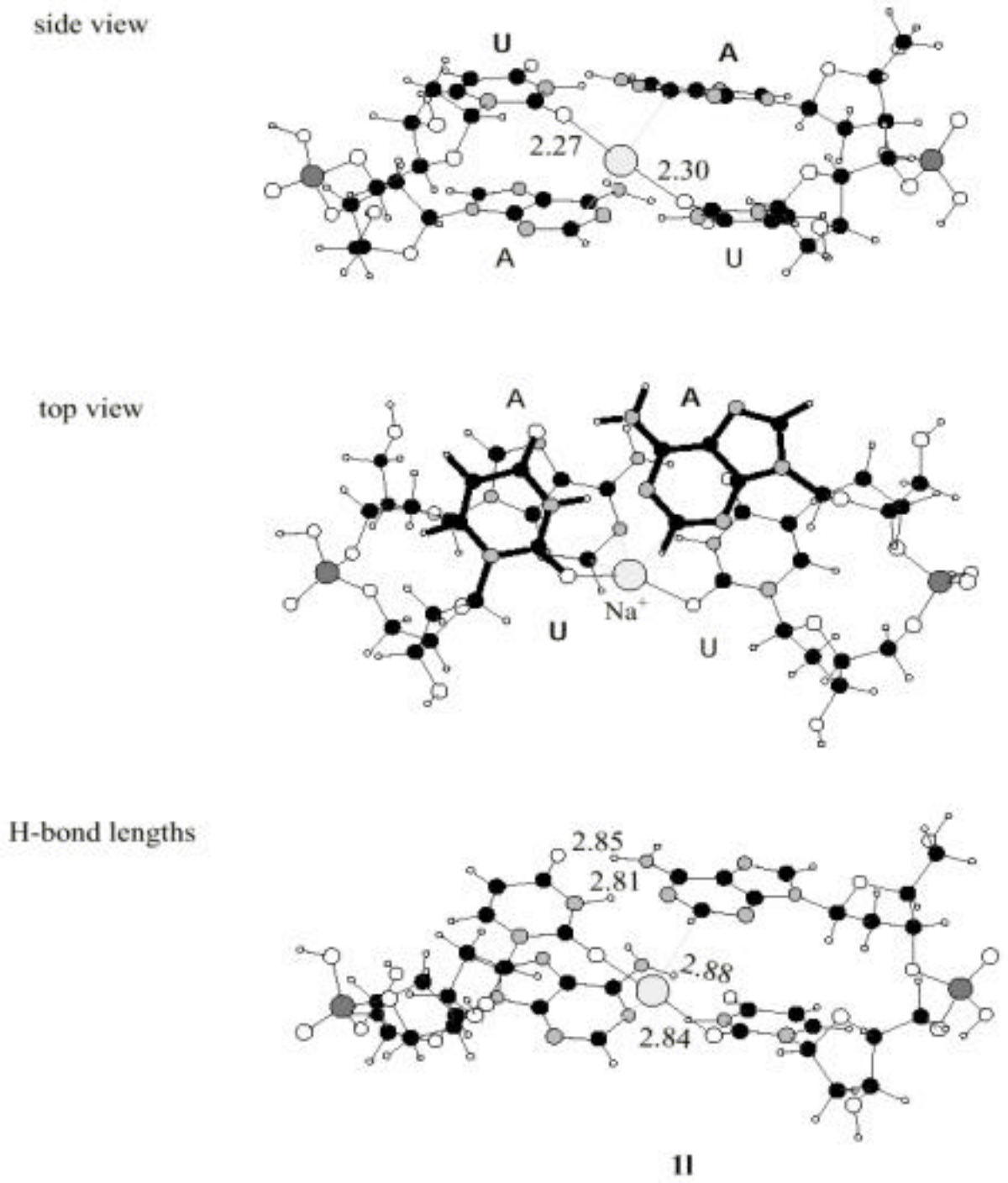
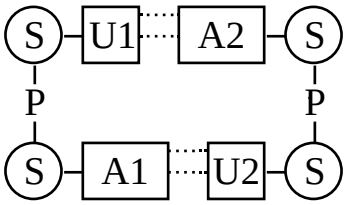


Figure 6.7. Different perspectives of the BP86/DZP structure of the Watson-Crick-type dimer of deoxyadenylyl-3',5'-deoxyuridine, $(dApdU)_2$ (see also Scheme 2), with Na^+ ion (**1l**) and without Na^+ ion (**1k**), both optimized in C_1 symmetry without any symmetry constraint. The illustration shows N6–O4 and N1–N3 distances in AU pairs and the distances between the O2 atoms of each uracil base and the Na^+ ion.

We went even one step further by studying the Watson-Crick complex of a strand of two nucleotides, namely that of deoxyadenylyl-3',5'-deoxyuridine, i.e., $(dApdU)_2$ (**1k**). This is a model for the corresponding adenylyl-3',5'-uridine complex $(ApU)_2$ in the crystal (**1**) studied by Seeman et al.[8a] (we have only removed the 2'-OH groups of ribose to somewhat reduce the immense computational cost). The structure of both our model $(dApdU)_2$ (**1k**) and the $(ApU)_2$ complex (**1**) is illustrated by Scheme 2.



Scheme 2

The BP86/DZP geometry of **1k** is shown from different perspectives in Figure 6.7, left. As can be seen, the AU hydrogen-bond distances in **1k** (Figure 6.7) differ only slightly, i.e., at most by 0.03 Å from those of plain AT (**1a**) also obtained at BP86/DZP (Table 6.3). The Watson-Crick-pairing energy ΔE of **1k** equals –20.9 kcal/mol at BP86/TZ2P// BP86/DZP (Table 6.5). Note that, although **1k** involves two AU pairs, this is significantly *less* than twice the pairing energy ΔE of AT (**1a**) or AU (**1b**). This can be ascribed to the strain in the backbone, which shows up in the much higher preparation energy ΔE_{prep} of 9.2 kcal/mol, and not to the actual interaction energy ΔE_{int} of –30.2 kcal/mol between the strands which, in fact, *is twice as strong* as that of a single base pair (Table 6.5).

In conclusion, the backbone has only a marginal influence on Watson-Crick hydrogen bonds and is thus not the source for the disagreement between theoretical and experimental structures mentioned above. But we can make use of this finding for reducing the computational cost in our further investigations on the effect of the crystal environment that the bases experience in **1** and **2** by leaving out the backbone.

6.5 The Effect of the Crystal Environment

6.5.1 Environment Effects on Watson-Crick Structures

As will appear in the following, reconciliation of theory and experiment regarding AT and GC structures is achieved if one incorporates the effects of the molecular environment on the Watson-Crick pairs in the crystals of sodium adenylyl-3',5'-uridine hexahydrate (**1**) and sodium guanylyl-3',5'-cytidine nonahydrate (**2**) into the theoretical model systems. We begin with AT or AU. In **1**,[8a] the amino-group of adenine and the O4 atom of uridine interact with two water molecules (A1U2, see Scheme 2) or with two 3'-ribose-OH groups of another $(ApU)_2$ complex (A2U1). We have modelled these interactions at BP86/TZ2P by introducing successively two water molecules at the corresponding positions in AT (compare **1a** and **1g-1i** in Figure 6.5). The N1–N3 bond is not much affected but the N6–O4 expands, only slightly for one H_2O (**1g** and **1h**) but significantly for two water molecules (**1i**). This leads to hydrogen bond lengths in close agreement with experiment (**1i**: N6–O4 and N1–N3 are 2.92 and 2.80 Å).

The effect of sodium counter ions is modest. In the crystal (**1**), one of the sodium ions bridges the O2 atoms of the two uracil bases (see Scheme 2 and Figure 6.7). First, we have modelled this by adding a Na^+ ion to plain (**1a**) and dihydrated AT (**1i**): the changes in hydrogen bond length in the resulting systems **1f** and **1j**, respectively, are marginal, i.e. 0.01-0.02 Å (Figure 6.5). The bridging Na^+ ion in the real crystal (**1**) may have a somewhat more pronounced effect because, there, it binds simultaneously to two uridine bases that are part of opposite strands of the double-helical segment and involved in different AU pairs (Figure 6.7). We have studied this at BP86/DZP by adding a Na+ ion to (dApdU)2 (**1l**, see Figure 6.7). Indeed, with hydrogen-bond elongations of 0.02-0.07 Å on going from **1k** to **1l**, the effect is a bit more pronounced than in the case of the flat model systems (**1a**, **1f** and **1j**). But eventually, the hydrogen bond lengths in **1l** still deviate significantly from the experimental values for **1** (compare Figures 6.5 and 6.7). We note that at variance with the situation in our model **1l**, the sodium ion does not enter into the space between the layers of the two AU pairs. Instead, it remains in the minor groove where it can bind also to water molecules.

Next, we consider the environment effects on the structure of GC pairs in **2**.[8b] Here, the N7 and O6 and N2 positions of guanine are involved in hydrogen bonds with water molecules that, in

case of N7 and O6, coordinate to a Na^+ ion. The O2 position of cytosine also forms a hydrogen bond with a water molecule whereas N4 hydrogen binds to a 3'-ribose-OH group of a neighboring $(GpC)_2$ complex. We have modelled the interactions of GC with its environment at BP86/TZ2P by introducing up to six water molecules and one sodium cation (compare **2a** and **2e**-**2i** in Figure 6.6). This time, the sodium ion appears to be crucial. Introducing four water molecules (one at each position, O6 and N2 of guanine, and N4 and O2 of cytosine) leads to a significant elongation of the O6–N4, N1–N3 and N2–O2 hydrogen bonds which are now 2.77, 2.94 and 2.94 Å (**2e**) but we still have the wrong bond length pattern short-long-long (**2e**) instead of long-long-short in the crystal (**2**, see Figure 6.6). The situation improves significantly if, in addition, a sodium cation is introduced. This has been done in **2f**-**2i**: all these model systems (but **2f**) show the correct hydrogen-bond length pattern (long-long-short) with O6–N4, N1–N3 and N2–O2 distances that, especially for **2i**, agree excellently (i.e., within 0.01-0.03 Å) with the X-ray data (**2**: 2.91, 2.95 and 2.86 Å).

6.5.2 Environment Effects on Watson-Crick Bond Strength

To analyse how the Watson-Crick interaction energy ΔE_{int} (eqs 6.2.4 and 6.2.5) is affected by the environment, we divide our model systems (**1f**-**1j** and **2e**-**2i**) into two subsystems, each of which consists of one of the bases plus the environment molecules that are closest to that base (see Figures 6.5 and 6.6). For example, the sodium ion in **1f** and **1j** belongs to thymine. First, we examine the interaction in AT systems (**1a**, **1f**-**1j**, Table 6.5). The introduction of water molecules has little effect. In **1g**-**1i**, the Watson-Crick interaction energy ΔE_{int} decreases only slightly (at most by 1.6 kcal/mol) with respect to **1a**. The presence of the Na^+ ion in **1f** and **1j** causes stronger orbital interactions as a result of which the hydrogen-bond interaction energy ΔE_{int} increases by some 7 kcal/mol. In the case of GC (**2a**, **2e**-**2i**, Table 6.6), hydration and the introduction of a sodium cation leads in all cases to a moderate increase of 0.4-5.7 kcal/mol of the hydrogen-bond interaction ΔE_{int} with respect to **2a**. This is caused by a slight increase of the electrostatic attraction (and a reduction of Pauli repulsion).

We conclude that hydration and counterions combined have a clearly visible effect on the hydrogen-bond structure and strength of Watson-Crick pairs. Although we are with our model systems of course still far removed from the real crystal, we have been able to incorporate the

most important interactions with the crystal environment, and this has brought theoretical and experimental structures into agreement. It is interesting to note that the species we have studied may also be conceived as microsolvated base pairs and, in this respect, they are also simple models for Watson-Crick systems that are exposed to hydration and ions under physiological conditions.

Table 6.7. Analysis of the interaction energy (kcal/mol) and charge transfer (electrons) between AT and the environment in **1f**-**1j**.[a]

	1f	**1g**	**1h**	**1i**	**1j**
AT with	Na^+	H_2O on A	H_2O on T	2 H_2O	2 H_2O + Na^+
Interaction Energy Decomposition[b]					
ΔE_{Pauli}	12.1	13.7	11.0	24.7	35.6
ΔV_{elstat}	–32.9	–13.3	–10.1	–23.5	–54.5
$\Delta E_{Pauli}+\Delta V_{elstat}$	–20.8	.4	.9	1.2	–18.9
ΔE_{σ}	–10.4	–5.6	–5.1	–10.7	–20.9
ΔE_{π}	–8.2	–.5	–.6	–.9	–8.9
ΔE_{oi}	–18.6	–6.1	–5.7	–11.6	–29.8
ΔE_{int}	–39.4	–5.7	–4.8	–10.4	–48.7
VDD Charge of AT[c]					
ΔQ_{AT}	+0.04	–0.02	+0.03	+0.01	+0.01

[a] BP86/TZ2P. See Figure 6.5.

[b] See section 6.2.2.

[c] See section 6.2.3.

Table 6.8. Analysis of the interaction energy (kcal/mol) and the charge transfer (electrons) between GC and the environment in **2e**-**2i**.[a]

	2e	**2f**	**2g**	**2h**	**2i**
GC with	4 H_2O	5 H_2O + Na^+	4 H_2O + Na^+	6 H_2O + Na^+	4 H_2O + Na^+
Interaction Energy Decomposition[b]					
ΔE_{Pauli}	32.8	65.9	61.2	57.6	65.0
ΔV_{elstat}	–32.2	–68.0	–84.2	–75.1	–67.9
$\Delta E_{Pauli}+\Delta V_{elstat}$	.6	–2.1	–23.0	–17.5	–2.9
ΔE_{σ}	–17.1	–42.0	–35.8	–31.9	–43.0
ΔE_{π}	–1.5	–6.1	–8.2	–6.2	–6.9
ΔE_{oi}	–18.6	–48.1	–44.0	–38.1	–49.9
ΔE_{int}	–18.0	–50.2	–67.0	–55.6	–52.8
VDD Charge of GC[c]					
ΔQ_{GC}	+0.02	+0.15	+0.08	+0.07	+0.18

[a] BP86/TZ2P. See Figure 6.6.

[b] See section 6.2.2.

[c] See section 6.2.3.

6.5.3 Analysis of Interaction with Environment

Finally, we take a short look at the interaction between the Watson-Crick pairs and the environment. We want to know the strength and the nature of these intermolecular forces. For that purpose, we divide our AT and GC models **1f**-**1j** and **2e**-**2i** again into two subsystems using, however, a partitioning that differs from the above one. This time, the first subsystem is the plain base pair and the second subsystem consists of the surrounding water molecules and/or sodium cation. The analyses of both the interaction (section 6.2.2) and the associated charge transfer (section 6.2.3) between these fragments in the various AT and GC systems are summarized in Tables 6.7 and 6.8.

If the environment consists of only water molecules that hydrogen bind to the base pairs (as in **1g**-**1i** and **2e**), the interaction ΔE_{int} is between –4.8 through –18.0 kcal/mol which comes down to roughly –5 kcal/mol per H_2O. Orbital interactions are relatively important here. Although they are about half as strong as the electrostatic interaction, they still are crucial for achieving net binding. As revealed by the VDD analyses, the interaction between the base pairs and the water molecules is accompanied by charge transfer from or to the environment: in **1g** (ΔQ_{AT} = –0.02 e) and **1h** (ΔQ_{AT} = +0.03 e) the AT pair accepts and donates electrons, respectively, from the water molecule (the VDD charge of a base pair, ΔQ_{AT} or ΔQ_{GC}, is computed as the sum of the VDD charges of all atoms that belong to that pair). In **1i** and **2e**, donation of electrons to and acceptation of electrons from the water molecules of the environment occur simultaneously and almost cancel (in **1i**: ΔQ_{AT} = +0.01 e, and in **2e**: ΔQ_{GC} = +0.02 e).

Introducing a sodium cation into the environment (as in **1f**, **1j** and **2f**-**2i**) increases all components of the interaction energy (Tables 6.7 and 6.8). If Na^+ interacts directly with the base pair (**1f**, **1j**, **2g** and **2h**), the electrostatic interaction gains in importance in the sense that it can overcome the Pauli repulsion and provide net bonding on its own. This is still substantially reinforced by sizeable orbital interactions ΔE_{oi} (roughly half as strong as ΔV_{elstat}) that lead to a charge transfer of up to 0.18 electrons in the case of **2i** from base pair to environment. On the other hand, when the sodium cation is separated from the base pair by a shell of water molecules (as in **2f** and **2i**), the electrostatic interaction ΔV_{elstat} becomes smaller with respect to the other components of the interaction and merely compensates the Pauli repulsion. In these cases, both ΔV_{elstat} and ΔE_{oi} are needed to achieve substantial net bonding between base pair and environment.

6.6 Conclusions

We have unravelled a hitherto unresolved discrepancy between theoretical and experimental hydrogen bond lengths in Watson-Crick base pairs. The disagreement was caused by a deficiency in the model systems used so far in theoretical computations, namely, the absence of the molecular environment (i.e. water, sugar OH groups, counterions) that the base pairs experience in the

crystals studied experimentally. If we incorporate the major elements of this environment into our model, simulating them by up to six water molecules and one Na^+ ion, we achieve excellent agreement with experiment at BP86/TZ2P.

On the other hand, whether plain nucleic bases or more realistic models for the nucleotides are used is much less important. Neither hydrogen bond lengths nor strengths are significantly affected if we use methyl, ribose, or 5'-ribose monophosphate instead of hydrogen as the substituents at N9 and N1 of the purine and pyrimidine bases, respectively. Even the Watson-Crick-type dimer of deoxyadenylyl-3',5'-deoxyuridine [$(dApdU)_2$, a model for a DNA segment of two base pairs] yields hydrogen bond lengths that differ only slightly from those in a plain AT or AU base pair.

Furthermore, we find that the BP86 functional yields A–T and G–C bond enthalpies in excellent agreement with experiment, especially in combination with the TZ2P basis. But also the smaller and, thus, more economic DZP basis leads to satisfactory results, provided that bond energies are corrected for the basis set superposition error. On the other hand, PW91 and BLYP functionals furnish hydrogen bonds that are up to 0.03 Å shorter and up to 2.5 kcal/mol more binding than those obtained with BP86.

Finally, our finding that present-day density functional theory is very well able to adequately describe biologically relevant molecules involving hydrogen bonds has an important consequence for future quantum biochemical studies. It is a justification for tackling this type of computationally extremely demanding problems with DFT as an efficient alternative to traditional (*i.e.*, Hartree-Fock-based) *ab initio* methods.

References

[1] a) G.A. Jeffrey, W. Saenger, *Hydrogen Bonding in Biological Structures*; Springer-Verlag: Berlin, New York, Heidelberg, 1991.

b) G.A. Jeffrey, *An Introduction to Hydrogen bonding*; Oxford University Press: New York, Oxford, 1997, Chapter 10.

c) W. Saenger, *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, Berlin, Heidelberg, Tokyo, 1984.

d) J.D. Watson, F.H.C. Crick, *Nature* **1953**, *171*, 737.

[2] a) P. Gilli, V. Ferretti, V. Bertolasi, G. Gilli, In: Advances in Molecular Structure Research; Hargittai, M. and Hargittai, I., Eds.; JAI Press: Greenwich, CT, 1996; Vol. 2; p. 67-102.

b) G. Gilli, F. Bellucci, V. Ferretti, V. Bertolasi, *J. Am. Chem. Soc.* **1989**, *111*, 1023.

c) G. Gilli, V. Bertolasi, V. Ferretti, P. Gilli, *Acta Cryst.* **1993**, *B49*, 564.

[3] C. Fonseca Guerra, F.M. Bickelhaupt, J.G. Snijders, E.J. Baerends, *Chem. Eur. J.* **1999**, *5*, accepted.

[4] C. Fonseca Guerra, F.M. Bickelhaupt, *Angew. Chem.* **1999**, *111*, in press.

[5] a) E.D. Isaacs, A. Shukla, P.M. Platzman, D.R. Hamann, B. Barbiellini, C.A. Tulk, C. A. *Phys. Rev. Lett.* **1999**, *82*, 600.

b) A. J. Dingley, S. Grzesiek, *J. Am. Chem. Soc.* **1998**, *120*, 8293.

c) F. Cordier, S. Grzesiek, *J. Am. Chem. Soc.* **1999**, *121*, 1601.

d) G. Cornilescu, J.-S. Hu, A. Bax, *J. Am. Chem. Soc.* **1999**, *121*, 2949.

[6] I. K. Yanson, A. B. Teplitsky, L. F. Sukhodub, *Biopolymers* **1979**, *18*, 1149.

[7] a) R. Santamaria, A. Vázquez, *J. Comp. Chem.* **1994**, *15*, 981.

b) J. Bertran, A. Oliva, L. Rodríguez-Santiago, M. Sodupe, *J. Am. Chem. Soc.* **1998**, *120*, 8159.

c) I.R. Gould, P.A. Kollman, *J. Am. Chem. Soc.* **1994**, *116*, 2493.

d) J. Sponer, J. Leszczynski, P. Hobza, *J. Phys. Chem.* **1996**, *100*, 1965.

e) K. Brameld, S. Dasgupta, W.A. Goddard III, *J. Phys. Chem. B* **1997**, *101*, 4851.

f) P. Hobza, J. Sponer, *Chem. Phys. Lett.* **1996**, *261*, 379

[8] a) N. C. Seeman, J. M. Rosenberg, F.L. Suddath, J.J.P. Kim, A. Rich, *J. Mol. Biol.* **1976**, *104*, 109.

b) J.M. Rosenberg, N.C. Seeman, R.O. Day, A. Rich, *J. Mol. Biol.* **1976**, *104*, 145.

[9] a) F. Sim, A. St-Amant, I. Papai, D.R. Salahub, *J. Am. Chem. Soc.* **1992**, *114*, 4391.

b) H. Guo, S. Sirois, E.I. Proynov, D.R. Salahub, In: Theoretical Treatment of Hydrogen

Bonding; Hadzi, D., Ed.; Wiley: New York, 1997.

c) S. Sirois, E.I. Proynov, D.T. Nguyen, D.R. Salahub, *J. Chem. Phys.* **1997**, *107*, 6770.

d) K. Kim, K.D. Jordan, *J. Phys. Chem.* **1994**, *98*, 10089.

e) J.J. Novoa, C. Sosa, *J. Phys. Chem.* **1995**, *99*, 15837.

f) Z. Latajka, Y. Bouteiller, *J. Chem. Phys.* **1994**, *101*, 9793.

g) J.E. Del Bene, W.B. Person, K. Szczepaniak, *J. Phys. Chem.* **1995**, *99*, 10705.

h) J. Florian, B.G. Johnson, *J. Phys. Chem.* **1995**, *99*, 5899.

i) J.E. Combariza, N.R. Kestner, *J. Phys. Chem.* **1995**, *99*, 2717.

j) B. Civalleri, E. Garrone, P. Ugliengo, *J. Molec. Struct.* **1997**, *419*, 227.

k) M. Lozynski, D. Rusinska-Roszak, H.-G.Mack, *J. Phys. Chem.* **1998**, *102*, 2899.

l) A.K. Chandra, M. Nguyen, *Chem. Phys.* **1998**, *232*, 299.

m) B. Paizs, S. Suhai, *J. Comp. Chem.* **1998**, *19*, 575.

n) M. A. McAllister, *J. Molec. Struct.* **1998**, *427*, 39.

o) Y. P. Pan, M.A. McAllister, *J. Molec. Struct.* **1998**, *427*, 221.

p) L. Gonzalez, O. Mo, M. Yanez, *J. Comp. Chem.* **1997**, *18*, 1124.

q) P.R. Rablen, J.W. Lockman, W.L. Jorgensen, *J. Phys. Chem.* **1998**, *102*, 3782.

[10] a) C. Fonseca Guerra, O. Visser, J.G. Snijders, G. te Velde, E.J. Baerends, In: Methods and Techniques for Computational Chemistry; Clementi, E. and Corongiu, G., Eds.; STEF: Cagliari, 1995; p. 305-395.

b) E.J. Baerends, D.E. Ellis, P. Ros, *Chem. Phys.* **1973**, *2*, 41.

c) E.J. Baerends, P. Ros, *Chem. Phys.* **1975**, *8*, 412.

d) E.J. Baerends, P. Ros, *Int. J. Quantum Chem., Quantum Chem. Symp.* **1978**, *S12*, 169.

e) C. Fonseca Guerra, J.G. Snijders, G te Velde, E.J. Baerends, *Theor. Chem. Acc.* **1998**, *99*, 391.

f) P.M. Boerrigter, G. te Velde, E.J. Baerends, *Int. J. Quantum Chem.* **1988**, *33*, 87.

g) G. te Velde, E.J. Baerends, *J. Comp. Phys.* **1992**, *99*, 84.

h) J.G. Snijders, E.J. Baerends, P. Vernooijs, *At. Nucl. Data Tables* **1982**, *26*, 483.

i) J. Krijn, E.J. Baerends, *Fit-Functions in the HFS-Method; Internal Report* (in Dutch); Vrije Universiteit: Amsterdam, 1984.

j) J.C. Slater, *Quantum Theory of Molecules and Solids Vol. 4*; McGraw-Hill: New York, 1974.

k) S.H. Vosko, L. Wilk, M. Nusair, *Can. J. Phys.* **1980**, *58*, 1200.

l) A.D. Becke, *J. Chem. Phys.* **1986**, *84*, 4524.

m) A. Becke, *Phys. Rev. A* **1988**, *38*, 3098.

n) J.P. Perdew, *Phys. Rev. B* **1986**, *33*, 8822 (Erratum: *Phys. Rev. B* **1986**, *34*, 7406).

o) L. Fan, T. Ziegler, *J. Chem. Phys.* **1991**, *94*, 6057.

p) J.P. Perdew, In: Electronic Structure of Solids; Ziesche, P. and Eschrig, H., Eds.; Akademie Verlag: Berlin, 1991; p. 11-20.

q) J.P. Perdew, J.A. Chevary, S.H. Vosko, K.A. Jackson, M.R. Pederson, D.J. Singh, C. Fiolhais, *Phys. Rev. B* **1992**, *46*, 6671.

r) C. Lee, W. Yang, R.G. Parr, *Phys. Rev. B* **1988**, *37*, 785.

s) B.G. Johnson, P. M. W. Gill, J.A. Pople, *J. Chem. Phys.* **1992**, *98*, 5612.

t) L. Versluis, T. Ziegler, *J. Chem. Phys.* **1988**, *88*, 322.

u) L. Fan, L. Versluis, T. Ziegler, E.J. Baerends, W. Ravenek, *Int. J. Quantum. Chem., Quantum. Chem. Symp.* **1988**, *S22*, 173.

v) S.F. Boys, F. Bernardi, *Mol. Phys.* **1970**, *19*, 553.

[11] P. W. Atkins, *Physical Chemistry*; Oxford University Press: Oxford, 1982.

[12] F.M. Bickelhaupt, E.J. Baerends, In: Rev. Comput. Chem.; Lipkowitz, K. B. and Boyd, D. B., Eds.; Wiley-VCH: New York, scheduled for Vol. 15.

[13] a) F.M. Bickelhaupt, N. M. M. Nibbering, E.M. van Wezenbeek, E.J. Baerends, *J. Phys. Chem.* **1992**, *96*, 4864.

b) T. Ziegler, A. Rauk, *Inorg. Chem.* **1979**, *18*, 1755.

c) T. Ziegler, A. Rauk, *Inorg. Chem.* **1979**, *18*, 1558.

d) T. Ziegler, A. Rauk, *Theor. Chim. Acta* **1977**, *46*, 1.

[14] F.M. Bickelhaupt, N. J. R. van Eikema Hommes, C. Fonseca Guerra, E.J. Baerends, *Organometallics* **1996**, *15*, 2923.

[15] C. Kittel, *Introduction to Solid State Physics*; Wiley: New York, 1986.

[16] a) J. Sponer, P. Hobza, J. Leszczynski, In: Computational Chemistry. Reviews of Current Trends; Leszczynski, J., Ed.; World Scientific Publisher: Singapore, 1996; p. 185-218.

b) J. Sponer, P. Hobza, *Int. J. Quantum Chem.* **1996**, *57*, 959.

c) E.L. Stewart, C.K. Foley, N.L. Allinger, J.P. Bowen, *J. Am. Chem. Soc.* **1994**, *116*, 7282.

d) J. Sponer, P. Hobza, *J. Phys. Chem.* **1994**, *98*, 3161.

Summary

In this thesis, the results are presented of a density functional theoretical (DFT) investigation on the structure and nature of the deoxyribonucleic acid (DNA) molecule, the carrier of the hereditary information. The purpose is to provide a foundation for an accurate, quantitative description of the structure and energetics of DNA, the influence of the molecular environment (such as, *e.g.*, water molecules and counterions) and, in particular, a better understanding of the nature and behavior of this molecule that is of crucial importance for the existence of all life (see general introduction, chapter 1). However, the quantum chemical calculations that need to be done for achieving this goal are associated with a computational cost that, until recently, has been out of reach, even with the use of supercomputers. This computational problem had to be tackled in the first place before the actual studies on DNA could be started. For this aim, the Amsterdam Density Functional (ADF) program was further developed by implementing two speed-up techniques: parallelization and linearization of the code.

In chapter 2, the parallelization of the ADF program is described. Parallelization of a program means that it is redesigned in such a way that the computational job is executed not only by one but, simultaneously, by several processors (the "nodes" of the parallel computer). In this case, the Single Program Multiple Data (SPMD) model is used, which implies that all nodes execute exactly the same program, each one however processing a different part of the data.

From serial calculations it appeared that the subroutines, in which numerical integration or loops over the atom pairs are carried out, consume most of the computational time. It was therefore obvious to distribute the integration points and atom pairs over the nodes. It appears to be very important that the data are distributed equally to achieve that each node has the same load. Furthermore, a static load balancing was chosen, that is, the data are distributed over the nodes only once, in the beginning of the program. In this way, the communication between the nodes of the parallel machine is kept to a minimum which is beneficial for the efficiency.

From the Figures, that show the speed-ups of the individual subroutines, we can conclude that indeed the computational time of the modified routines scales well with the number of nodes: the CPU time is halved when the number of nodes is doubled. Overall, the ADF program scales well up to about 32 nodes for medium-sized calculations and for large calculations even up to 128 nodes. This means that the parallelization opens the perspective of routinely performing quantum chemical investigations on molecules that are more than one order of magnitude larger than has been feasible before.

The principle of the other speed-up method, the linearization of the code discussed in chapter 3, is based on the fact that, simply speaking, atoms far away from each other do not "feel" the presence of the other and, thus, it is not necessary to calculate explicitly the very small interaction. By neglecting the latter, one must compute the interactions of each atom with its nearest neighbors only. In case of large molecules, this can lead to a situation in which the computational cost does no longer rise cubically, as usual, but linearly with the size of the system. If this goal is really achieved, one speaks of linear scaling.

The study presented in chapter 3 mainly deals with the construction of the matrix of the Kohn-Sham operator which, by analogy with the Hartree-Fock method, is also called the Fock matrix. This is the most expensive part in a Kohn-Sham DFT calculation. The matrix elements of this operator, of which the exchange-correlation and Coulomb potential are part, are calculated in ADF through numerical integration. The values of these potentials in the integration points are obtained with the help of so-called fit functions, which are used for the description of the electronic density. Now, one tries to achieve linearization of the code that calculates these potentials in the integration points by cutting off the fit functions. As soon as an integration point lies outside the so-called cut-off radius of the fit function, it is skipped in the calculation of that fit function.

Although it may seem plausible to define the cut-off radius of a fit or basis function as the radius where the function value equals a certain threshold, the function is instead cut off at that radius at which *the relative weight of the tail* of the function (*i.e.*, the ratio between the radial integral of this tail beyond the cut-off radius and that of the total function) matches a certain threshold. The thresholds are set beforehand and determine the cut-off radii.

The calculation of the Fock-matrix elements is further accelerated by taking, just as in the case of the fit functions, only those points for the evaluation of the basis functions in the integration points that are located within the sphere around the nucleus defined by the cut-off radius. Furthermore, the Fock matrix elements between two basis functions are calculated only when the distance between the nuclei is smaller than the sum of the corresponding cut-off radii.

A similar approach is used for setting up the fitted density. The fitted density, which, as mentioned before, is used for calculating efficiently the values of the potentials in the integration points, is determined by first writing the exact density as a sum over densities of all possible pairs of atoms, $\rho_{exact}(\mathbf{r}) = \Sigma_{A,B}\rho_{AB}$ with $\rho_{AB} = \Sigma_{i\ A,j\ B} P_{ij}\chi_i^A\chi_j^B$, and then approximating each ρ_{AB} by an expansion of fit functions on atom A and B, $\rho_{AB} = \Sigma_{i\ A} a_i f_i^A + \Sigma_{j\ B} a_j f_j^B$. This procedure is linearized by excluding pairs of atoms that are further away from each other than the sum of the corresponding atomic cut-off radii (the atomic cut-off radius is the largest cut-off radius in the set of fit functions on that atom).

The results presented in chapter 3 show that for 1-dimensional systems, such as the zigzag chains of n-alkanes, perfect linear scaling can be obtained. Even the computational cost for setting up the Fock matrix, previously scaling cubically, drops to linear scaling. The 2-dimensional systems (which derive from aromatic polycyclic hydrocarbons) do not yet have the critical size necessary for *linear* scaling. Yet, spectacular speed-ups are obtained also in this case. The scaling of the Fock-matrix setup goes from cubic (N^3) to almost linear ($N^{1.3}$). We can therefore conclude that the linearization has led to a further considerable increase of the efficiency of ADF. Together with the parallelization of the code, this paves the way to the research described in the next three chapters.

Chapter 4 is the introduction to the second part of this thesis, which is devoted to DNA. It gives a brief overview of the key findings. In the first place, the bonding analyses of the Watson-Crick base pairs adenine–thymine (AT) and guanine–cytosine (GC) show that the present conception of hydrogen bonds in DNA need to be substantially adjusted: they are not plain electrostatic phenomena reinforced, as suggested by Gilli, through resonance in the π-electronic system (the so-called Resonance Assisted Hydrogen Bonding or RAHB). Instead, charge transfer appears to contribute considerably to the strength of these hydrogen bonds. To understand and, in

particular, to reproduce correctly the experimental structures, the interaction of the DNA base pairs with the molecular environment in the crystal (or under physiological conditions) turns out to be of crucial importance. This insight has led to the solution of a hitherto unresolved discrepancy between experimental (X-ray) and theoretical (*ab initio* and DFT) structures of AT (or AU) and GC. When the most important hydrogen bond interactions between the base pair and its environment are taken into account, the available nonlocal density functionals yield results that agree excellently with the experiment. This finding has an enormous scope. It shows that the presently available density functionals are, in principle, capable of adequately describing biological molecules containing hydrogen bonds. This justifies, for future work on such molecules, the use of DFT as an efficient alternative to the much more expensive traditional (i.e., Hartree-Fock-based) *ab initio* methods.

Chapter 5 elaborates on the nature of the hydrogen bonds in DNA base pairs. From detailed analyses at the BP86/TZ2P level of DFT, it appears that the charge-transfer interaction between the DNA bases in the Watson-Crick pairs is caused by donor–acceptor orbital interactions between the lone pair on the oxygen or nitrogen of one base and the N–H antibonding σ acceptor orbitals of the opposite base. Polarization effects in the π-electron system that are reminiscent of Gilli's RAHB model could indeed be revealed. Energetically, however, they are of minor importance because they are one order of magnitude smaller than the σ-interactions. Furthermore, there appears to be neither any synergism between the charge transfer from one base to the other through one hydrogen bond and back through the other hydrogen bond, nor between the σ- and the π-components of the interaction between the bases. In this respect, there is no "resonance assistance" as meant in the RAHB. On the other hand, on the extremely flat potential surface, the π-interactions are still capable of shortening the hydrogen bonds between the bases by 0.1 Å. In this sense, one may speak of a certain π assistance. Another point of discussion in the literature concerns the existence of the C–H•••O hydrogen bond in AT. We show that this does not exist.

In order to analyze the extremely subtle charge-transfer interactions in Watson-Crick pairs the Voronoi Deformation Density (VDD) method was further developed, to make it possible to correctly monitor very small charge reorganizations caused by weak chemical interactions (such as the hydrogen bonds investigated here) and, in addition, to decompose them into the contributions

of the σ- and π-electron system. The VDD analyses confirm the existence of charge transfer in the σ- and polarization in the π-electron system.

Chapter 6 describes our search for the cause of the hitherto existing discrepancy between theory (Hartree-Fock or DFT) and experiment (X-ray diffraction) regarding hydrogen bond lengths in Watson-Crick base pairs. In the first place, we have investigated which of the nonlocal density functionals available in ADF (BP86, PW91 and BLYP) is best capable of reproducing the experimental hydrogen bond enthalpies. Computations with a very large doubly polarized STO basis of triple-ζ quality pointed to BP86 as the most appropriate functional: the theoretical A–T and G–C bond enthalpies of –11.8 kcal/mol and –23.8 kcal/mol, calculated (for 298 K) at BP86/TZ2P, are in excellent agreement with the experimental values of –12.1 and –21.0 kcal/mol, respectively. The optimized geometries appear to be much less dependent on the nonlocal density functional used. Computations with the BP86 functional furthermore showed that the calculated geometries and bond enthalpies hardly change if a singly-polarized double-ζ basis (DZP) instead of the TZ2P basis is used, provided the bond enthalpies are corrected for the Basis Set Superposition Error (BSSE). Nevertheless, to be safe, all calculations described in chapters 4 to 6 were performed at the BP86/TZ2P level, unless stated otherwise.

After having established an adequate theoretical level, the model systems were examined in order to unravel, once and for all, the discrepancy between theory and experiment mentioned above. So far, only the "plain" base pairs adenine–thymine (or adenine–uracil) and guanine–cytosine had served as model systems. To investigate if modelling the glycosidic N–C bond between base and sugar with an N–H bond could have led to the established disagreement, we have investigated the Watson-Crick pairs of the corresponding methylated bases as well as the Watson-Crick pairs of the corresponding nucleosides and even nucleotides, that is, bases that are substituted with a sugar or a sugar-phosphate group, respectively, at nitrogen atoms concerned (i.e., N1 for T and C, N9 for A and G). This appeared to have no influence on the calculated hydrogen bond lengths: the discrepancy between theory and experiment remained. However, as soon as we had incorporated the most important elements of the molecular environment that the base pairs experience in the crystals studied experimentally, computed values were obtained in excellent agreement with experiment. As already concluded above, this finding has an enormous

scope because it shows that the already existing density functionals are capable of adequately describing biological molecules containing hydrogen bonds. One may hope that, besides the theoretical-biochemical insights acquired and the acceleration of the software, this means a step forward in the development of quantum biology.

Samenvatting

In dit proefschrift worden de resultaten gepresenteerd van dichtheidsfunctionaaltheoretisch (DFT) onderzoek naar de structuur en aard van het deoxyribonucleïnezuur- ofwel DNA-molecuul, de drager van de erfelijke informatie. Doel van dit onderzoek is het leggen van een basis voor een nauwkeurige, kwantitatieve beschrijving van de structuur en energetica van DNA, de invloed van de moleculaire omgeving (zoals bijv. water moleculen en tegenionen) en met name het verkrijgen van een diepergaand begrip van de geaardheid en het gedrag van dit voor het bestaan van ieder leven cruciale molecuul (zie algemene inleiding, hoofdstuk 1). De voor het bereiken van dit doel benodigde quantumchemische berekeningen zijn echter extreem rekenintensief en waren tot voor kort, zelfs onder gebruikmaking van supercomputers niet haalbaar. Daarom moest in de eerste plaats dit rekenkundige probleem worden aangepakt, voordat met het onderzoek aan DNA kon worden begonnen. Hiertoe werd het Amsterdam-Dichtheids-Functionaal- (ADF) programma verder ontwikkeld door de implementatie van twee versnellingstechnieken: parallelisatie en linearisatie van de code.

In hoofdstuk 2 wordt de parallelisatie van het ADF-programma beschreven. Parallellisatie van een programmacode houdt in, dat deze zodanig wordt opgezet, dat het te verwerken rekenkundige probleem niet alleen door één maar door meerdere processoren (de "knopen" van de parallelle computer) tegelijk wordt verwerkt. In dit geval wordt gebruik gemaakt van het *Single Program Multiple Data* - (SPMD) model, hetgeen betekent, dat alle knopen exact hetzelfde programma uitvoeren, maar verschillende delen van de data verwerken.

Uit seriële berekeningen bleek, dat de subroutines, waarin numerieke integratie plaatsvindt of loops over de paren van atomen worden genomen, de meeste rekentijd verbruiken. Het lag dus voor de hand om de integratie punten en de paren van atomen over de verschillende knopen te verdelen. Het blijkt zeer belangrijk te zijn dat de gegevens gelijkmatig verdeeld worden om zodoende iedere knoop dezelfde belasting te geven. Verder is er gekozen voor een statische

belasting van de knopen, d.w.z de gegevens worden éénmalig (aan het begin van het programma) over de knopen verdeeld. Hierdoor blijft de communicatie tussen de knopen van de parallele machine zo klein mogelijk, hetgeen de efficiëntie ten goede komt.

Uit de figuren, die de versnellingsfactoren voor de individuele subroutines tonen, kunnen we opmaken dat de rekentijd van de aangepakte routines inderdaad goed schaalt met het aantal knopen: de rekentijd halveert wanneer het aantal knopen verdubbelt. Het ADF-programma in zijn geheel schaalt goed tot ongeveer 32 knopen voor middelgrote berekeningen en voor grote berekeningen zelfs tot 128 knopen. Dit betekent dat de parallelisatie het perspectief opent op routinematig quantumchemisch onderzoek aan moleculen, die ruim een orde groter zijn dan wat voorheen haalbaar was.

Het principe van de andere versnellingsmethode, de in hoofdstuk 3 besproken linearisatie van de code, berust eenvoudig gezegd op het feit, dat ver van elkaar gelegen atomen elkaar niet "voelen" en het dus niet nodig is hun zeer geringe wisselwerking expliciet te berekenen. Door deze te verwaarlozen, hoeft men slechts de wisselwerking van ieder atoom met zijn naaste buren te berekenen. Bij grote moleculen leidt dit ertoe, dat de rekenkosten niet meer zoals gewoonlijk kubisch maar lineair met de grootte van het systeem stijgen. Als dit doel daadwerkelijk bereikt wordt, spreekt men van lineaire schaling.

De in hoofdstuk 3 gepresenteerde studie houdt zich vooral bezig met de constructie van de matrix van de Kohn-Sham-operator, die in analogie met de Hartree-Fock-methode ook wel Fock-matrix wordt genoemd. Dit is de duurste stap in een Kohn-Sham-DFT-berekening. De matrixelementen van deze operator, waar de exchange-correlatie- en de Coulomb-potentiaal deel van uit maken, worden in ADF d.m.v. numerieke integratie berekend. De waarden van deze potentialen in de integratiepunten worden verkregen uit de zogenaamde fitfuncties, die voor de beschrijving van de electronische dichtheid worden gebruikt. Men tracht nu de linearisatie van de code, die deze potentialen in de integratiepunten berekent, te bereiken door de fitfuncties af te kappen. Zodra een integratiepunt buiten de zogenaamde afkapstraal van de fitfunctie valt, wordt dit integratiepunt bij de berekening van deze fitfunctie overgeslagen.

Hoewel het misschien voor de hand lijkt te liggen om de afkapstraal voor een fit- of basisfunctie te definiëren als de straal, waarbij de functiewaarde een zekere drempelwaarde

bereikt, wordt in plaats daarvan de functie bij die straal afgekapt, waarbij het *relatieve gewicht van de staart* van de functie (d.w.z. de verhouding tussen de radiële integraal van deze staart voorbij de afkapstraal en die van de totale functie) een zekere drempelwaarde evenaart. De drempelwaarden worden vooraf vastgesteld en bepalen de afkapstralen.

De berekening van de Fock-matrixelementen wordt verder versneld door voor de evaluatie van de basisfuncties in de integratiepunten net als bij de fitfuncties alleen die punten mee te nemen die binnen de door de afkapstraal gedefiniëerde bol om de kern vallen. Tevens geldt, dat Fock-matrixelementen van twee basisfuncties alleen worden uitgerekend, indien de afstand tussen de kernen kleiner is dan de som van de corresponderende afkapstralen.

Bij het opzetten van de fitdichtheid wordt eenzelfde aanpak gebruikt. De fitdichtheid, die zoals eerder genoemd, gebruikt wordt voor het efficiënt berekenenen van de waarden van de potentialen in de integratiepunten, wordt bepaald door de exacte dichtheid te schrijven als een som van dichtheden van alle mogelijke atoomparen, $\rho_{exact}(\mathbf{r}) = \Sigma_{A,B}\rho_{AB}$ met $\rho_{AB} = \Sigma_{i\ A,j\ B} P_{ij}\chi_i^A\chi_j^B$, en vervolgens iedere ρ_{AB} te benaderen door een expansie van fitfuncties op atoom A en atoom B, $\rho_{AB} = \Sigma_{i\ A} a_i f_i^A + \Sigma_{j\ B} a_j f_j^B$. Deze procedure wordt nu gelineariseerd door paren van atomen uit te sluiten, die verder uiteen liggen dan de som van de corresponderende atomaire afkapstralen (de atomaire afkapstraal correspondeert met de grootste afkapstraal in de set van fitfuncties op het atoom).

Uit de in hoofdstuk 3 getoonde resultaten blijkt, dat voor 1-dimensionale systemen zoals de zigzagketens van n-alkanen perfecte lineaire schaling verkregen kan worden. Zelfs de voorheen kubisch schalende rekenkosten voor het opzetten van de Fock-matrix zakken in naar lineaire schaling. De 2-dimensionale systemen (die zijn afgeleid van aromatische polycyclische koolwaterstoffen) blijken nog niet de voor *lineaire* schaling benodigde kritische grootte te bezitten. Desondanks worden ook hier spectaculaire versnellingen behaald. De schaling voor het opzetten van de Fock-matrix loopt terug van kubisch (N^3) naar bijna lineair ($N^{1.3}$). Er kan dan ook geconcludeerd worden dat de linearisatie tot een verdere aanzienlijke verhoging van de efficientie van ADF heeft geleid. Samen met de parallelisatie van de code effent dit de weg naar het in de volgende drie hoofdstukken beschreven onderzoek aan DNA.

Hoofdstuk 4 vormt de inleiding tot dit tweede, aan DNA gewijde deel van het proefschrift en

geeft een beknopt overzicht van de belangrijkste bevindingen. Ten eerste tonen de bindingsanalyses van de Watson-Crick-basenparen adenine-thymine (AT) en guanine-cytosine (GC) dat de gangbare opvatting over de waterstofbruggen in DNA substantiëel bijgesteld moet worden: deze berusten níet op in essentie electrostatische wisselwerkingen, die zoals gepostuleerd door Gilli versterkt worden door resonantie in het π-electronensysteem (de zogenaamde *Resonance Assisted Hydrogen Bonding* of RAHB). Ladingsoverdracht blijkt juist een aanzienlijke bijdrage aan de bindingssterkte van deze waterstofbruggen te geven. Voor het begrijpen en met name het correct reproduceren van de experimenteel bepaalde structuren blijkt tevens de wisselwerking van de DNA-basenparen met de moleculaire omgeving in het kristal (of onder fysiologische omstandigheden) van cruciaal belang te zijn. Dit inzicht heeft geleid tot de oplossing van een daarvoor onbegrepen discrepantie tussen experimenteel (Röntgen) en theoretisch (*ab initio* en DFT) bepaalde structuren van AT (of AU) en GC. Wanneer de belangrijkste waterstofbrug-wisselwerkingen tussen het basenpaar en zijn omgeving worden meegenomen, dan blijken de beschikbare nietlokale dichtheidsfunctionalen resultaten op te leveren die uitstekend met het experiment overeenstemmen. Deze vinding heeft een enorme draagwijdte. Zij toont aan, dat de tegenwoordig beschikbare dichtheidsfunctionalen principiëel in staat zijn waterstofbruggen bevattende biologische moleculen adequaat te beschrijven. Dit rechtvaardigt het gebruik van DFT als efficiënt alternatief voor de veel duurdere traditionele, d.w.z. op Hartree-Fock-theorie gebaseerde *ab initio* methoden voor toekomstig werk aan soortgelijke moleculen.

Hoofdstuk 5 gaat dieper in op de geaardheid van de waterstofbruggen in DNA-basenparen. Uit gedetailleerde analyses op het BP86/TZ2P-niveau van DFT blijkt dat de ladingsoverdrachtwisselwerking tussen de DNA-basen in de Watson-Crick-paren verzorgd wordt door donor/acceptor-orbitaalinteracties tussen het vrije electronenpaar op zuurstof of stikstof van de ene base en de N–H-antibindende σ -acceptororbitalen van de tegenoverliggende base. Verder konden daadwerkelijk polarisatie-effecten in het π-electronensysteem vastgesteld worden, die aan het door Gilli gepostuleerde RAHB-model doen denken. Energetisch gezien spelen deze echter geen rol, daar zij ongeveer één orde kleiner zijn dan de σ-interacties. Er blijkt ook geen sprake te zijn van synergie tussen de ladingsoverdracht van één base naar de andere via één van de waterstofbruggen en terug via een andere waterstofbrug en evenmin tussen de σ- en de π-componenten van de wisselwerking

tussen de basen. In die zin is er dus geen sprake van de in het RAHB-model bedoelde "resonantie-ondersteuning". Anderzijds zijn de π-interacties op het uiterst vlakke potentiaaloppervlak toch in staat om de lengte van de waterstofbruggen tussen de basen met 0.1 Å te verkorten. Zo gezien kan men wel van enige π-ondersteuning spreken. Een ander discussie punt in de literatuur betreft het wel of niet bestaan van een C–H•••O-waterstofbrug in AT. Ons onderzoek toont aan dat deze er niet is.

Ten behoeve van de analyse van de zeer subtiele ladingsoverdracht-interacties in de Watson-Crick-paren werd de Voronoi-Deformatie-Dichtheid- (VDD) methode verder ontwikkeld, zodat deze nu in staat is zeer kleine ladinsgverschuivingen veroorzaakt door zwakke chemische interacties (zoals de hier onderzochte waterstofbruggen) op correcte wijze te registreren en deze bovendien te verdelen in de bijdragen van het σ- en π-electronensysteem. De VDD-analyses bevestigen het voorkomen van ladingsoverdracht in het σ- en polarisatie in het π-systeem.

Hoofdstuk 6 beschrijft de zoektocht naar de oorzaak van de tot dan toe bestaande discrepantie tussen de theoretische (met Hartree-Fock of DFT berekende) en de experimenteel (uit Röntgen-kristalstructuren) bepaalde waterstofbruglengtes voor Watson-Crick-basenparen. In eerste instantie werd uitgezocht, welke van de in ADF beschikbare, niet-locale dichtheidsfunctionalen (BP86, PW91 en BLYP) de experimentele waterstofbrug-bindingsenthalpieën het beste reproduceert. Berekeningen met een zeer grote tweevoudig gepolariseerde STO-basis van tripel-ζ-kwaliteit (TZ2P) wezen op BP86 als de meest geschikte functionaal: de op het BP86/TZ2P-niveau berekende A–T- en G–C-bindingsenthalpieën (voor 298 K) van –11.8 kcal/mol en –23.8 kcal/mol komen uitstekend overeen met de experimentele waarden van –12.1 respectievelijk –21.0 kcal/mol. De geoptimaliseerde geometrieën blijken veel minder van de gebruikte nietlokale dichtheidsfunctionaal af te hangen. Verder bleek uit berekeningen met de BP86-functionaal, dat de berekende geometrieën en bindingsenthalpiën nauwelijks veranderen, indien een enkelvoudig gepolariseerde dubbel-ζ-basis (DZP) i.p.v. de TZ2P basis wordt gebruikt; bindingsenthalpiën moeten dan in het geval van de DZP basis wel voor de *Basis Set Superposition Error* (BSSE) gecorrigeerd worden. Zekerheidshalve werden toch alle verdere in de hoofdstukken 4 t/m 6 beschreven berekeningen op het BP86/TZ2P-niveau uitgevoerd, tenzij anders aangegeven.

Na het vaststellen van een adequaat rekenniveau, werden de modelsystemen geïnspecteerd om

de reeds hiervoor genoemde discrepantie tussen theorie en experiment eens en voor al op te lossen. De tot dan toe onderzochte modelsystemen waren de "gewone" basenparen adenine–thymine (of adenine–uracil) en guanine–cytosine. Om na te gaan of het modelleren van de glycosidische N–C-binding tussen base en suiker door een N–H binding tot de vastgestelde afwijking geleid zou kunnen hebben, werden de Watson-Crick-paren van de corresponderende gemethyleerde basen onderzocht alsmede de Watson-Crick-paren van basen, die op het betreffende stikstofatoom (N1 bij T en C, N9 bij A en G) een suiker- of zelfs een suikerfosfaatgroep hebben. Dit bleek geen invloed op de berekende waterstofbruglengtes te hebben; de discrepantie tussen theorie en experiment bleef bestaan. Wanneer echter de belangrijkste elementen van de moleculaire omgeving van de basenparen in het experimenteel onderzochte kristal in het modelsysteem worden opgenomen, verkrijgt men berekende waarden, die uitstekend met de experimentele overeenkomen. Zoals reeds hiervoor werd geconcludeerd, heeft deze vinding een enorme draagwijdte, omdat hierdoor wordt aangetoond, dat de reeds tegenwoordig beschikbare dichtheidsfunctionalen in staat zijn waterstofbruggen bevattende biologische moleculen adequaat te beschrijven. Men mag hopen dat dit, naast de hier gewonnen theoretisch-biochemische inzichten en de versnelling van de software, een extra stap voorwaarts betekent in de ontwikkeling van de quantumbiologie.

Dankwoord

Mijn proefschrift wil ik afsluiten met het bedanken van een aantal mensen, die direct of indirect bijgedragen hebben aan het welslagen van dit promotie-onderzoek.

Ik begin bij mijn co-promotor en tevens echtgenoot, Matthias Bickelhaupt, die enorm heeft bijgedragen aan de totstandkoming van dit proefschrift. In zijn hoedanigheid van copromotor bedank ik hem voor de begeleiding van het DNA onderzoek. Discussies over het onderzoek heb ik altijd als zeer inspirerend ervaren. Ik hoop dan ook dat onze samenwerking in de toekomst voortgezet kan worden. In zijn hoedanigheid van echtgenoot bedank ik hem voor zijn steun en aanmoediging om door te zetten. Hoewel dat niet altijd even gemakkelijk ging vanwege een grote geografische afstand en een behoorlijk aantal tijdzones tussen ons, was het altijd mogelijk om via talken, e-mail of telefoon de hindernissen, die zich zoal bij het promotie-onderzoek voordoen, te bespreken.

Mijn promotoren Evert Jan Baerends en Jaap Snijders bedank ik voor hun begeleiding bij de parallellisatie en linearisatie van het ADF-programma, maar ook voor de vrijheid die zij mij gegeven hebben bij het vormgeven van mijn onderzoek. Hun vertrouwen in mij en mijn streven om de ontwikkeling van de programmatuur met een chemische probleemstelling te verbinden heeft tot een goede afloop van dit promotie-onderzoek geleid.

Pieter Vernooys ben ik veel dank verschuldigd voor het oplossen van de vele computer-problemen die zich bij dit werk voorgedaan hebben. Ook zijn collegialiteit heb ik als zeer waardevol ervaren.

Olivier Visser bedank ik voor de prettige samenwerking tijdens het parallelisatie project.

Mijn kamergenoten Vincent Osinga en Pier Philipsen bedank ik voor de gezellige sfeer in R129 en hun collegialiteit.

Verder bedank ik de overige (ex-)TC leden (Remco Bouten, Roxanne Bruinsma, Francesco Buda, Andreas Ehlers, Bernd Ensing, Ad van Eulem, Stan van Gisbergen, Jonathan Go, Oleg Gritsenko, Jeroen Groeneveld, Myrta Grüning, Jan-Willem Handgraaf, Sophie de Jong, Eric

Kirchner, Geert-Jan Kroes, Robert van Leeuwen, Erik van Lenthe, Evert Jan Meyer, Roar Olsen, Chris Oostenbrink, Walter Ravenek, Angela Rosa, Pieter Schipper, Antoinette Sisto, Bert te Velde, Luuk Visscher, Margot Vlot en Gijsbert Wiesenekker) voor de goede sfeer bij theoretische chemie.

I thank Professor Gernot Frenking, Dr. Miquel Solà and Professor Tom Ziegler for giving me the opportunity to present my work in their group.

I also thank the members of the reading committee, consisting of Prof. Dr. C. Altona, Prof. Dr. H. Bal, Prof. Dr. J. Betran, Prof. Dr. H.A. Raué and Dr. L. Visscher, for critically reading the manuscript of this thesis.

Mijn schoonouders en mijn schoonzus Christina bedank ik voor hun belangstelling voor mijn werk.

Tenslotte gaat er heel veel dank uit naar mijn ouders en mijn broer Fred voor hun interesse in en steun bij mijn werk. Hun luisterend oor en goede raad is voor mij zeer waardevol geweest en heeft het verloop van dit promotie-onderzoek positief beïnvloed .