

Structure and relative acidity for a model zinc finger

David W. Deerfield II^{a,*}, Lee G. Pedersen^b

^aPittsburgh Supercomputing Center, 4400 Fifth Ave, Pittsburgh, PA 15213, USA

^bDepartment of Chemistry, The University of North Carolina, Chapel Hill, NC, USA

Received 1 November 1996; accepted 20 March 1997

Abstract

We have performed ab initio computations examining the structure and relative acidity for a model $(\text{Zn(II)}-(\text{H}_2\text{S})_2(\text{Imidazole})_2)$ for the zinc finger binding site. All possible deprotonated forms were studied. The lowest energy complex was a mono-deprotonated form $(\text{Zn(II)}-(\text{H}_2\text{S})(\text{HS}^-)(\text{Imidazole})_2)$. In addition, we found that a triply deprotonated form $(\text{Zn(II)}-(\text{H}_2\text{S})-(\text{HS}^-)(\text{deprotonated imidazole})_2)$ rearranges to unique but not unexpected forms. © 1997 Elsevier Science B.V.

Keywords: Zinc; Imidazole; Hydrogen sulfide; Zinc finger; Quantum mechanics

1. Introduction

Many proteins, such as the transcription factors, will recognize and bind to relatively long stretches of DNA. There are a variety of proteins motifs for binding to a nucleic acid, including: zinc fingers, helix-turn-helix and homeodomains. The zinc finger is the most prevalent of the various DNA-binding motifs [1–5]. The zinc finger domain is relatively small (consensus sequence shown in Fig. 1) and will recognize only short segments of DNA [1,5]. The DNA-sequence selectivity is changed by mutating a minimal number of residues in the zinc finger domain [1,6,7]. Thus zinc finger domains represent building modules with easily modified DNA-sequence specificity. DNA-binding proteins will often contain a linear series of zinc fingers allowing for the recognition of a long DNA sequence (e.g. ZIF268 contains three [8]

and transcription factor IIIA (TFIIIA) contains nine zinc finger domains [9]). The structure and function of zinc fingers can be summarized as:

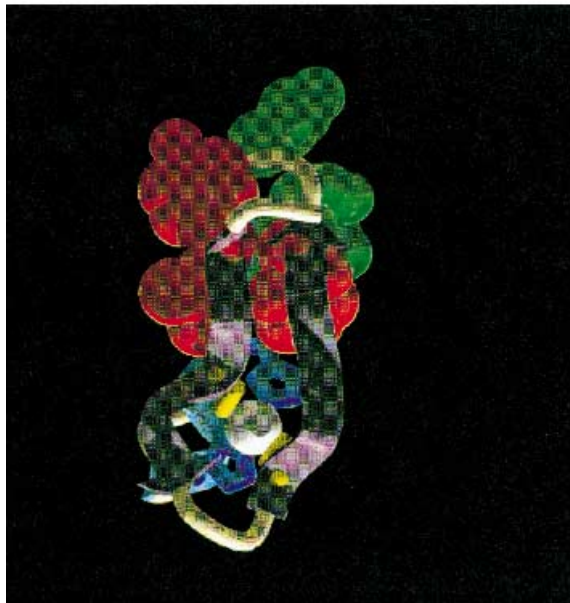
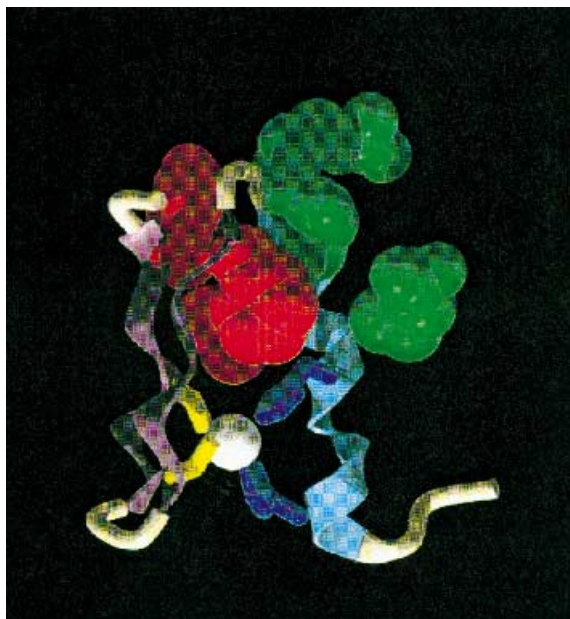
1. Each zinc finger domain interacts in the major groove with three consecutive bases from essentially one strand of duplex B-form DNA.
2. In a protein with multiple zinc fingers, the zinc finger subsites are linearly arranged.
3. The DNA is recognized via sequence specific contacts using hydrogen bonds between the protein side chains and the DNA.
4. After binding the zinc finger, the DNA is in a distorted B-form with a decreased helical twist and expanded major groove.

Zinc fingers, as the name implies, are Zn(II) ion binding domains. In the absence of Zn(II) ions, the protein is largely unfolded [10]. Addition of Zn(II) ions results in the protein folding to a relatively compact domain [1,5]. In the original zinc fingers, the Zn(II) ion is tetrahedrally bound to four ligands

* Corresponding author. Tel.: 001 412 268 4960; fax: 001 412 268 5832; e-mail: deerfiel@psc.edu



Fig. 1. The sequence profile for a zinc finger. The Zn(II) ion is bound by the side chains from two cysteines (yellow) and two histidines (blue). The hydrophobic core (red) is conserved and required for the proper folding of the zinc finger domain. DNA sequence specificity resides in Residues 13, 16 and 19. In addition, the side chain from Residue 15 has hydrogen bond interactions with the DNA. The secondary structure is given below the sequence and is from ZIF268 (PDB entry = 1zaa). (The coloring scheme used for this figure is the same as used in Fig. 2.)



from the protein [2]: the side chains from two cysteines and two histidines (Fig. 2). The conserved hydrophobic residues (Red in Figs. 1 and 2) form a core of the protein that is essential for proper folding [10]. Residues 13, 16 and 19 are responsible for DNA recognition (Figs. 1 and 2). The zinc finger domain contains an unusually large number of secondary structure elements, thus many positions are intolerant to proline substitution [1].

At pH 5.5 and below, but in the presence of Zn(II) ions, the zinc finger is in a slow equilibrium between the folded form and an unfolded form [9,11,12]. Recent pH titration studies have shown that the unfolding is a two proton process that ultimately results in the protonation of both imidazole rings. The NMR studies on a consensus sequence zinc finger [11] reveal that there is an intermediate structure in which one imidazole is protonated and no longer complexed with the Zn(II) ion leaving only three protein-supplied ligands about the Zn(II) ion. The coordination number of the resulting Zn(II) ion was not reported.

We have a longstanding interest in models for metal ion binding sites [13–19] and now turn to models for the zinc finger metal ion site. Previous researchers performing computational studies on similar systems

Fig. 2. Two molecular graphics representations from different camera positions of the first zinc finger from ZIF268 (Brookhaven PDB entry 1zaa). A spline is drawn through the C(alpha) carbon coordinates with the sheets shown as arrows (violet), the helix as a ribbon (sky blue) and the random coil as a tube (wheat). In the top figure, the N-terminus is to the upper left and the C-terminus is to the lower right. The white sphere is the Zn(II) ion (lower center), the bonds are shown to the two cysteines (yellow) and histidines (blue) that are interacting with the Zn(II) ion. The side chains for the hydrophobic residues that are essential for the proper folding of the domain are shown as the red spheres and the side chains for the DNA recognition site are shown as the green spheres. The models were created using the molecular graphics program GRAPHX and ray-traced using ART.

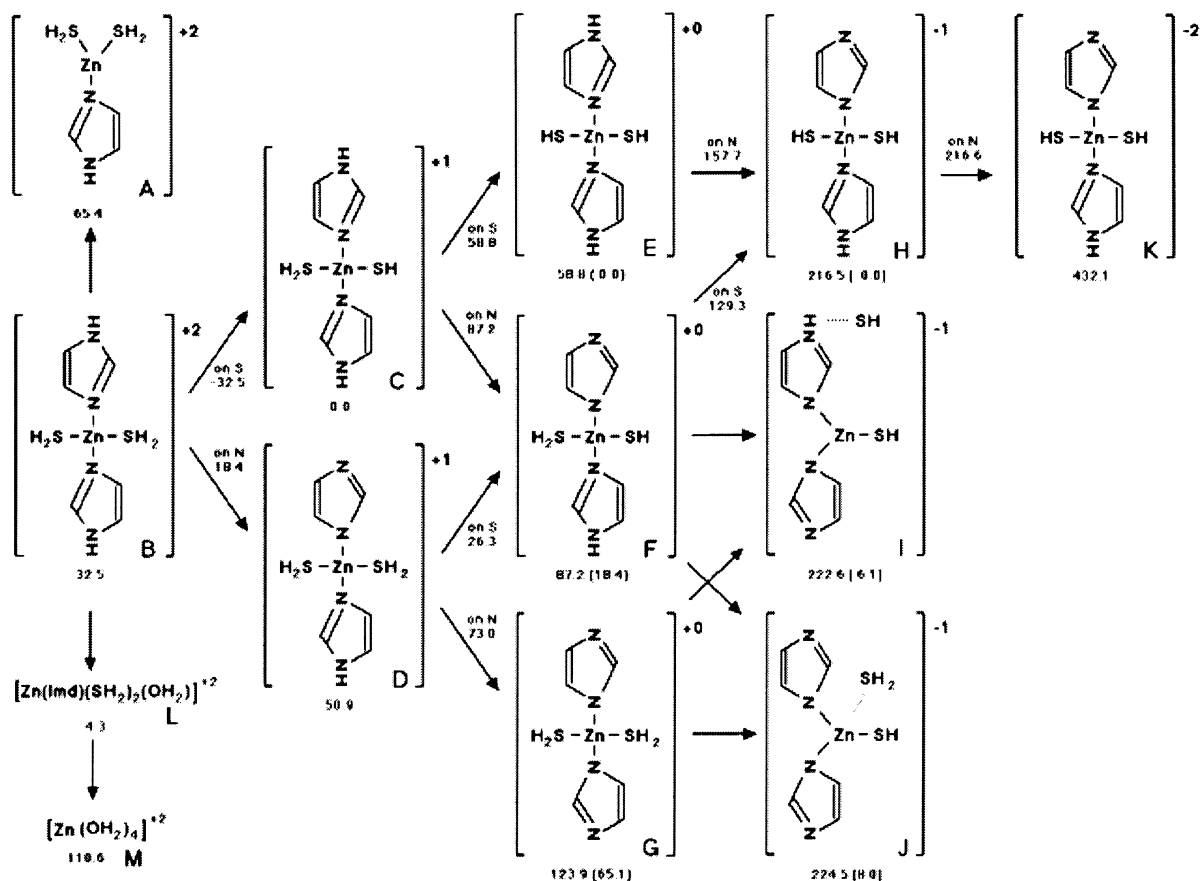


Fig. 3. Graphical representation of the results from this study. The numbers are the relative energies of the various complexes, and the numbers in parentheses are the relative energies for the complexes with the same deprotonation state. All energies are corrected for zero point energy and to 298°.

(e.g. alcohol dehydrogenase) have chosen to doubly deprotonate (on sulfur) the complex to provide an overall neutral Zn(II) ion site [20–23]. We have interest in examining whether this deprotonation model is appropriate for zinc fingers.

2. Computational methodology

Ab initio quantum mechanical calculations were performed using the GAUSSIAN94 program [24] on the Cray C916/16512 at the Pittsburgh Supercomputing Center and the T90 at the North Carolina Supercomputing Center. Optimizations employed the eigenvalue-following methods [25,26] with the initial force constants calculated analytically. The internal 3-21G* basis set was employed. All resulting

energy minimized structures were shown to be energy minima by confirming that all harmonic frequencies were real. The relative energies in the tables are reported in kcal mol⁻¹ as both the computed difference in energy (*comp*) and corrected for zero point energy and to 298° (*corr*) using the thermochemistry code in the GAUSSIAN94 program [24]. The energies reported in the figures are corrected. In addition, single-point computations were performed using a 6-311G* basis set at the 3-21G* optimized geometry.

3. Results and discussion

The results for this study are summarized in Table 1 and Fig. 3. The lowest energy complex (C) is deprotonated on only one hydrogen sulfide. That is, the

Table 1
Geometries and relative energies of the model zinc finger complexes studied. All structures were shown to be minima by confirming that all harmonic frequencies were real. These results are presented graphically in Fig. 3

Complex ^a	Relative energy ^b		Ionic distances				Interatomic angles			
	Comp	Corr	6-311G** 3-21G ^e				N1–Zn–N2			
			Zn–N1	Zn–N2	Zn–S1	Zn–S2	N1–Zn–N1	N1–Zn–S1	N1–Zn–S2	N1–Zn–S2
A Zn(lmd)(SH ₂)(SH ₂) ^c + ImdH	68.2	65.4	82.4	–	2.455	2.452	–	124.0	122.3	
B Zn(lmd)(lmd)(SH ₂)(SH ₂) ^d	34.2	32.5	20.3	1.918	1.922	2.601	120.1	110.1	111.0	
C Zn(lmd)(lmd)(SH ₂)(SH [–])	0.0	0.0	0.0	1.952	1.952	2.300	119.4	102.2	114.3	
D Zn(lmd)(lmd)(Im [–])(SH ₂)(SH ₂)	53.6	50.9	48.2	1.926	1.854	2.610	121.6	111.5	105.8	
E Zn(lmd)(lmd)(SH [–])(SH [–])	57.6	58.8	69.1	2.001	1.996	2.367	113.7	100.8	108.0	
F Zn(lmd)(Im [–])(SH ₂)(SH [–])	88.3	87.2	94.6	1.971	1.884	2.324	116.6	102.2	106.7	
G Zn(Im [–])(Im [–])(SH ₂)(SH ₂)	127.6	123.9	128.6	1.861	1.861	2.681	136.6	105.3	103.4	
H Zn(lmd)(Im [–])(SH [–])(SH [–])	216.4	216.5	233.9	2.027	1.931	2.455	111.3	99.3	98.0	
I Zn(Im [–])(lmd–SH)(SH [–]) ^c	223.1	222.6	237.9	1.885	1.926	2.289	117.3	135.2	126.0	
J Zn(Im [–])(Im [–])(SH [–])*SH [–] ‡	226.8	224.5	239.5	1.868	1.875	2.349	122.2	123.5	138.0	
K Zn(Im [–])(Im [–])(SH [–])(SH [–])	433.1	432.1	455.7	1.967	1.963	2.474	109.2	112.7	105.9	
L Zn(lmd)(SH ₂) ₂ (OH ₂) ^e + ImdH	4.5	4.3	13.7							
M Zn(OH ₂) ₂ + 2SH ₂ + 2lmd	110.4	110.6	57.1							

^a lmd is imidazole, Im[–] is the N3 deprotonated imidazole, ImdH is imidazolium, SH₂ is hydrogen sulfide, SH[–] is deprotonated hydrogen sulfide and OH₂ is water. In one case (complex I), the hydrogen sulfide rearranged from a contact ion pair with the Zn(II) ion to a position that is best described as the H₂S transferring a proton to a formerly deprotonated imidazole with the sulfur hydrogen bonding to the transferred proton (lmd–SH).

^b The relative energy given in kcal mol^{–1} and is computed for the transfer of a proton from the initial complex to a water. The lowest energy complex (C) has been assigned a value of 0.0. Zn(lmd)(lmd)(SH₂)(SH₂) + H₂O → Zn(lmd)(lmd)(SH₂)(SH[–]) + (H₃O⁺), ΔE = –34.2 kcal mol^{–1} (at 3-21G^e).

The *corr* relative energies are corrected for zero point energy and to 298° using the thermochemistry capability in the GAUSSIAN92 and GAUSSIAN94 software package. The 6-311G**//3-21G^e relative energies are computed difference in energy by calculating the HF/6-311G* single-point energy at the HF/3-21G* optimized geometry.

^c The Zn(II) ion is three-coordinate for three of the complexes, with the ligands in a trigonal planar arrangement. (Complex A deviation from the plane [Zn,N,N,S] = 0.003 Å; complex I deviation from the plane [Zn,N,N,S] = 0.000 Å; complex J deviation from the plane [Zn,N,N,S] = 0.004 Å.)

^d Computed energy of the fully optimized Zn(lmd)(lmd)(SH₂)(SH₂) complex is –3009.606514 (at 3-21G^e) a.u.

^e The geometry of this complex will be reported in a future publication [19].

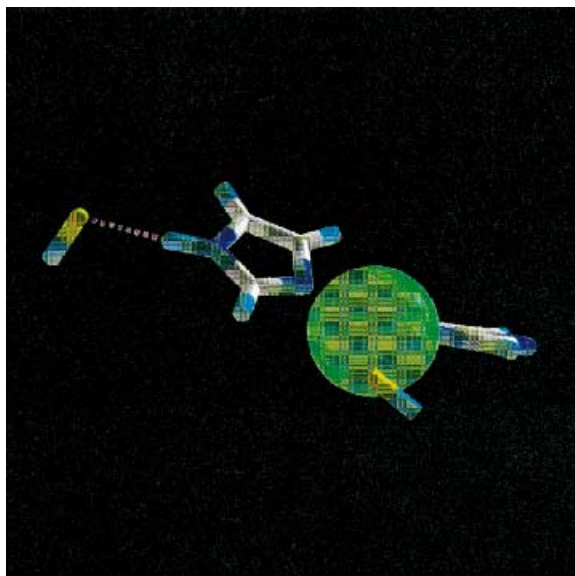


Fig. 4. Molecular graphics representation of complex I in Table 1.

neutral complex (doubly deprotonated) is not the lowest energy. The lowest energy complex (C) is substantially more stable (by $72.4 \text{ kcal mol}^{-1}$) than tetraaquozinc ($\text{Zn(II)(H}_2\text{O)}_4$). It is interesting that the protonation and deprotonation energies for the lowest energy complex (C) are very similar. Triple deprotonation (both imidazoles and one H_2S) of the parent complex resulted in rearrangement with the H_2S no longer in a contact ion pair with the Zn(II) ion (Fig. 4 and Fig. 5).

The Zn–S distance is more susceptible to deprotonation than the Zn–N(Imidazole) distance ($\Delta(\text{Zn–S}) = 0.254 \text{ \AA}$, 10.6%; $\Delta(\text{Zn–N}) = 0.057 \text{ \AA}$, 3.0%). The N–Zn–S angle was sensitive to the protonation state of the sulfur. This is clearly seen in the lowest energy complex (C), for which the N–Zn– SH_2 angle is 102.2° and the Imd–Zn–SH^- is 114.3° . However, the N–Zn–N(Imd or Im^-) angle showed only slight sensitivity to the protonation state. Furthermore, deprotonation on sulfur is preferred by an average of $38.6 \text{ kcal mol}^{-1}$ as compared to deprotonation of an imidazole.

Several comments can be made about the pH titration of zinc fingers. During the protonation and hydrolysis of zinc fingers, this study predicts that the lowest energy complex (C) is initially protonated to give the parent complex (B). Further protonation on imidazole

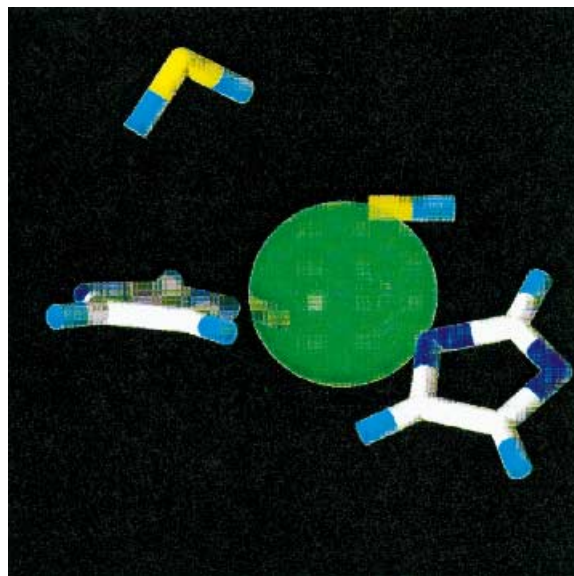


Fig. 5. Molecular graphics representation of Complex J in Table 1.

results in a partially aquated form (L) and then finally the fully aquated Zn(II) ion (M). A species has been observed in NMR studies that is consistent with either complex A or the partially aquated form (L). These computations strongly suggest that the partially aquated form (L) is the observed species and a reasonable intermediate at low pH in the hydrolysis of zinc fingers.

The observed rearrangement of the triply deprotonated complexes was not unexpected since we have observed this phenomenon previously in similar model systems (e.g. $\text{Zn(II)(SH}^-\text{)}_2(\text{OH}^-)(\text{H}_2\text{S})$) [19]. Due to the relatively high energy of these complexes, these rearranged forms are probably only computational curiosities.

Acknowledgements

We would like to thank the Pittsburgh Supercomputing Center for time on the Cray C916/16512 and the North Carolina Supercomputing Center for time on the Cray T90. This work was supported in part by a grant from the Pittsburgh Supercomputing Center through the NIH-National Center for Research Resources cooperative agreement P41 RR06009 and through individual NIH grants HL-27995 and

HL-26309 (L.G.P). L.G.P. especially thanks NIEHS for providing summer support during the course of this work.

References

- [1] J.M. Berg, *Acc. Chem. Res.* 28 (1995) 14.
- [2] M. Schmiedeskamp, R.E. Klevit, *Curr. Opin. Struct. Biol.* 4 (1994) 28.
- [3] R. Kaptein, *Curr. Opin. Struct. Biol.* 2 (1992) 109.
- [4] C.O. Pabo, R.T. Sauer, *Annu. Rev. Biochem.* 61 (1992) 1053.
- [5] D. Rhodes, A. Klug, *Sci. Am.*, (1993) 56.
- [6] C.A. Kim, J.M. Berg, *Nature Struct. Biol.* 11 (1996) 940.
- [7] J.R. Desjarlais, J.M. Berg, *Proc. Natl. Acad. Sci. U.S.A.* 90 (1993) 2256.
- [8] N.P. Pavletich, C.O. Pabo, *Science* 252 (1991) 809.
- [9] J.J. Hayes, T.D. Tullius, *J. Mol. Biol.* 227 (1992) 407.
- [10] S.F. Michael, V.J. Kilfoil, M.H. Schmidt, B.T. Amann, J.M. Berg, *Proc. Natl. Acad. Sci. U.S.A.* 89 (1992) 4796.
- [11] B.A. Krizek, B.T. Amann, V.J. Kilfoil, D.L. Merkle, J.M. Berg, *J. Am. Chem. Soc.* 113 (1991) 4518.
- [12] G. Párraga, S. Horvath, L. Hood, E.T. Young, R.E. Klevit, *Proc. Natl. Acad. Sci. U.S.A.* 87 (1990) 137.
- [13] D.W. Deerfield II, R.G. Hiskey, H.B. Nicholas, L.G. Pedersen, *Proteins* 6 (1989) 168.
- [14] D.W. Deerfield II, R.G. Hiskey, M.A. Lapadat, L. Spemulli, L.G. Pedersen, *J. Biomol. Struct. Dyn.* 6 (1989) 1077.
- [15] D.W. Deerfield II, R.G. Hiskey, H.B. Nicholas, L.G. Pedersen, *Proteins* 6 (1989) 168.
- [16] D.W. Deerfield II, D.J. Fox, M. Head-Gordon, R.G. Hiskey, L.G. Pedersen, *J. Am. Chem. Soc.* 113 (1991) 1892.
- [17] D.W. Deerfield II, D.J. Fox, M. Head-Gordon, R.G. Hiskey, L.G. Pedersen, *Proteins* 21 (1995) 244.
- [18] D.W. Deerfield II, L.G. Pedersen, in preparation.
- [19] D.W. Deerfield II, C.W. Carter, L.G. Pedersen, in preparation.
- [20] O. Tapia, R. Cardenas, J. Andres, J. Krechl, M. Campillo, F. Colonna-Cesari, *Int. J. Quantum Chem.* 39 (1991) 767.
- [21] U. Ryde, *Int. J. Quantum Chem.* 52 (1994) 1229.
- [22] U. Ryde, *Protein Sci.* 4 (1995) 1124.
- [23] U. Ryde, *Proteins* 21 (1995) 40.
- [24] 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, R. 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., Pittsburgh, PA, 1995.
- [25] J. Baker, *J. Comput. Chem.* 7 (1986) 385.
- [26] J. Baker, *J. Comput. Chem.* 8 (1987) 563.