4.1 Introduction

The Hessian matrix, comprising of second order partial derivatives of molecular energy with respect to nuclear coordinates, is one of the most commonly used entity for characterizing the potential energy surfaces (PES) of chemical systems and reactions. Diagonalization of the mass-weighted Hessian provides harmonic normal modes and corresponding eigenvalues produce the vibrational frequencies. Absence of imaginary frequencies confirms a minimum on the PES of a system. While occurrence of just one imaginary frequency generally indicates a transition state in which that imaginary frequency corresponds to the newly establishing bond during product formation. Thus, computation of Hessian matrix is useful as a diagnostic tool for the nature of a stationary point as well as for obtaining the vibrational frequencies. Further, many algorithms such as Davidon-Fletcher-Powell (DFP) algorithm and Broyden-Fletcher-Goldfarb-Shanno (BFGS) method, involve Hessian matrix for speeding up the process of geometry optimization.

There are two ways in which the Hessian calculation is carried out: numerical one using a finite difference of analytic gradients and fully analytic. The computational power and memory requirements are rather different for these two methods. When the Hessian is computed numerically, at least $3M+1$ energy and gradient evaluations are necessary for a system containing $M$ atoms. The analytic approach is usually preferable over the numerical one as it not only produces more accurate vibrational frequencies but also saves computational time considerably for large systems. However, in either case the computation of Hessian matrix at \textit{ab initio} level of theory is an expensive proposition in terms of CPU and memory requirements.
To surmount this scalability issue, computational chemists have developed a variety of means for fast evaluation of Hessian which are based on either parallelization of the algorithm or obtaining an approximate Hessian. For instance, a recent paper by Alexeev et al.\textsuperscript{6} discusses a new parallel algorithm for analytic Hessian calculation and scales well over a large number of processors for systems such as luciferin and capreomycin. Using an advanced HP IA64 based supercomputer interconnected by a very high speed network, the largest of the test cases is capreomycin. This calculation involves 778 basis functions and took about 75 minutes on 256 processors for the Hessian and frequency calculation at the HF/6-31G* level of theory. Izmaylov and Scuseria\textsuperscript{7} also recently proposed a method for efficient evaluation of analytical vibrational frequencies within HF and DFT frameworks for solids and periodic systems under periodic boundary conditions. They applied this method for evaluating the vibrational frequencies of various periodic non-conducting systems such as boron nitride sheet and bulk diamond, mostly using 6-31G** basis set. Employing efficient numerical quadrature, integral pre-screening based on rigorous estimates and exploitation of point group symmetry of the molecule, Deglmann and co-workers evaluated\textsuperscript{8} vibrational frequencies for large linear molecules and atomic sheets comprising up to 100 heavy atoms on a personal computer (PC) at DFT level of theory. Besley and Metcalf, in a recent interesting study,\textsuperscript{9} demonstrated a computation of frequencies from partial Hessian for amide I band of a few polypeptides and proteins at the HF and DFT levels of theory. Their method includes the effects of only carbon, oxygen and nitrogen present on the backbone chain of amino acids. This approximation basically explores the nature of the amide I band, which is largely localized on carbonyl groups of the backbone amide residues. For a set of model polypeptides at HF/STO-3G level of theory, they observed a mean absolute error of the order of 15 cm\textsuperscript{-1}. For the model test cases using their method, a time of 10\%–15\% of the full
Hessian calculation is observed which, certainly is a large gain in terms of computational requirements, albeit with lesser gains in terms of memory requirements. This method is integrated with the Q-Chem package. Neugebauer et al.\textsuperscript{10} developed a method for approximate calculations of IR and Raman frequencies and applied to buckminster fullerene at various basis sets. Their study is probably the first computational study on IR and Raman spectra of C\textsubscript{60} molecule using numerical Hessian derived from analytical gradients.

A few of the largest frequency calculations performed at \textit{ab initio} level of theory include a study by Pathak, Rastogi and another report by Requena \textit{et al.}. Pathak and Rastogi reported\textsuperscript{11} computation of IR frequencies for a variety of polycyclic aromatic hydrocarbons (PAHs) at the B3LYP/4-31G level of theory and compared their results with the corresponding experimental ones. All of their calculations are performed on high performance machines using standard GAMESS package.\textsuperscript{12} One of the largest calculations that they performed is on PAH C\textsubscript{96}H\textsubscript{24} at B3LYP/4-31G level of theory with 912 basis functions. Although the number of basis functions employed is indeed large, the authors’ use symmetry, prevalent in the case of PAHs, which considerably reduces the required computational cost. Requena \textit{et al.} reported\textsuperscript{13} frequency calculation of \textbeta-carotene and its derivatives, viz. capsanthin and capsorubin, at B3LYP/6-311G** level of theory, involving a maximum of 1096 basis functions.

Though parallelization of conventional Hessian calculation procedure does make it easier to handle larger molecules, it does not solve the core issue of non-linear scaling. Therefore, there is again restriction on the size of molecule that can be handled with these highly parallel codes. However, the scalability of \textit{ab initio} methods is addressed by divide-and-conquer (DC) methods such as molecular tailoring approach. In the present Chapter, an overview of applicability of MTA to
calculation of Hessian matrix and vibrational spectrum is presented. The implementation of the methodology for MTA-based Hessian and frequency calculation is discussed in Section 1.5. The MTA-based approximate Hessian matrix is assembled from those of fragments, according to Eq. 1.82 at the required level of theory. Further Eq. 1.71 to 1.73 are employed to obtain vibrational frequencies from the Hessian matrix. This method is tested and benchmarked for a structurally diverse, spatially large molecules and molecular clusters. The advantage factors and accuracy achieved by MTA with reference to the actual calculations (whenever possible) is judged (Section 4.2) before applying the method to systems of chemical interest (Sections 4.3 and 4.4).

4.2 Benchmarks

The method of MTA-based Hessian and frequency calculation is put to test on a number of molecular systems employing the HF or DFT level of theory with a variety of basis sets. For the benchmarking purpose, MTA-enabled GAMESS is employed. The spatially extended molecules used as test cases (for final optimized structures, see to Ref. 14) are chemically and biologically diverse in nature and the calculations are performed without enforcing symmetry. For all the test cases, a fragmentation scheme with a minimum R-goodness value, Rg, (refer to Subsection 1.5.2) of at least 3 Å is employed to ensure the reliability of the calculations performed.

The smallest of the test systems viz., the cluster of 37 water molecules, α-tocopherol, and folic acid have substantial chemical or biological significance. A fat-soluble antioxidant, α-tocopherol is the form of vitamin E that is preferentially absorbed in humans. Folic acid is a form of water soluble vitamin B. All these systems involve less than 600 basis functions at the specified level of theory (cf. Table 4.1), making it possible to run the actual frequency calculation using a
standard quantum chemical package, GAMESS. Even in such cases, MTA offers a notable advantage in terms of computational resources as well as time required for the calculation. For instance, in the case of α-tocopherol, the total time taken for this calculation is about 11 hours with MTA-enabled GAMESS as against the staggering 50 hours taken for the actual job on the same hardware (cluster of three Pentium Dual Core processors at 2.8 GHz with 2 GB of RAM and 80 GB of disk each). It is also worth noting that the maximum memory required per node for this job is about 600 MB for MTA while that for the actual job is 1.8 GB. Except for α-tocopherol, all other calculations reported in Table 4.1 are performed on a cluster of four Intel Core 2 Quad processors at 2.4 GHz with 4 GB RAM and 250 GB disk each. MTA-based Hessian computation for (H$_2$O)$_{37}$ achieved advantage factor of 6.9 taking ~3 hours to complete while the corresponding actual one took ~20 hours. It is expected that the advantage factor $T_r$ would enhance as the basis functions involved in calculation increase. This assumption is further cemented when $T_r$ value of 1.7 increases to 4.5 for increase in number of basis functions from 324 to 575 in the case of folic acid at B3LYP level of theory (cf. Table 4.1).

In terms of the accuracy of the Hessian matrix computed using MTA as compared to the actual one, the maximum error is $1.8 \times 10^{-4}$ a.u. and root mean square deviation (RMSD) is $1.0 \times 10^{-5}$ a.u. for α-tocopherol (see Table 4.1). Consistency in errors in case of folic acid when different basis sets are employed again assures the basis set independence of the MTA method as discussed in Subsection 2.2.3. For all the test cases examined, the maximum error in Hessian elements is of the order of $10^{-3}$ a.u. with RMSD of $10^{-4}$ to $10^{-5}$ a.u. It is also evident from Table 4.1 that the maximum error in vibrational frequencies is less than 0.5 cm$^{-1}$ with RMSD of the order of $10^{-2}$ cm$^{-1}$. Hence, MTA-based Hessian matrix and corresponding vibrational frequencies are, in general, quite reliable.
### Table 4.1

Level of theory, basis set, number of atoms ($N_{at}$) and basis functions ($N_{BF}$) employed for each test system. The maximum and RMS errors, $\varepsilon_{\text{MAX}}$ and $\varepsilon_{\text{RMSD}}$, in Hessian matrix (in a.u.) and vibrational frequencies† (in cm$^{-1}$) are listed. $T_{\text{MTA}}$ (in min) is time taken by the MTA calculation and $T_r$, the advantage factor over the actual calculation ($T_r = T_{\text{actual}}/T_{\text{MTA}}$).

<table>
<thead>
<tr>
<th>System/Level of theory</th>
<th>$N_{at}/N_{BF}$</th>
<th>Hessian Matrix</th>
<th>Frequency</th>
<th>$T_{\text{MTA}}$</th>
<th>$T_r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>($\text{H}<em>2\text{O})</em>{37}$ HF/6-31G</td>
<td>111/481</td>
<td>2.91 0.07 0.48 (3188.9) 0.028</td>
<td>172 6.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha$-tocopherol HF/6-31G*</td>
<td>81/565</td>
<td>0.18 0.01 0.01 (243.6) 0.009</td>
<td>666 4.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Folic acid (a) B3LYP/6-31G</td>
<td>51/326</td>
<td>3.24 0.18 0.47 (1081.3) 0.027</td>
<td>270 1.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Folic acid (b) B3LYP/6-31G**</td>
<td>51/575</td>
<td>3.24 0.18 0.32 (1508.9) 0.036</td>
<td>960 4.5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

† Comparisons only for frequencies greater than 50 cm$^{-1}$ are reported. The errors in frequencies below 50 cm$^{-1}$ are also small; however, the percentage error gets inflated.

All the vibrational frequencies reported in the present work are unscaled values obtained from the computation. Important vibrational frequencies for $\alpha$-tocopherol are: O-H stretching at 3604.3 cm$^{-1}$ (3604.4 cm$^{-1}$ for MTA) and C-O stretching at 1111.8 cm$^{-1}$ (1111.7 cm$^{-1}$ for MTA). For ($\text{H}_2\text{O})_{37}$, a wide range of O-H stretching frequencies from 3173 to 4106 cm$^{-1}$ is obtained due to the extensive hydrogen bonding between the water molecules. The important frequencies for folic acid are the N-H stretches at 3766.3 and 3621.6 cm$^{-1}$, O-H stretches in the range of 3715–3762 cm$^{-1}$, three C=O stretches at 1750.8, 1835.3, and 1847.6 cm$^{-1}$, while C-N stretching frequency is observed at 1677.5 cm$^{-1}$. For all these cases, the difference between MTA and actual frequencies is seen to be within 1 cm$^{-1}$. In general, when the complete range of frequencies is compared, the maximum deviation is less than 0.5% (cf. Table 4.1).
Next in line are a few larger test cases for which the actual Hessian and frequency calculations cannot be performed on the available hardware. Out of these cases, cholesterol, capreomycin and \( (H_3BO_3)_{40} \) (nanotube of 40 orthoboric acid molecules, refer to Section 3.2 for its structure) are treated at the HF level, while \( \alpha\)-cyclodextrin is treated at the DFT level of theory. MTA-optimized geometry of cholesterol is subjected to MTA-based vibrational frequency analysis. Chemically, cholesterol is a combination of steroid and alcohol and plays an important role in the cell membrane of animal tissues. Cholesterol is subjected to MTA-based vibrational frequency calculation at HF/6-31G* level of theory and includes 74 atoms and 650 basis functions. O-H stretch frequency for this molecule is found at 4179.0 cm\(^{-1}\). Capreomycin (91 atoms), a peptide antibiotic used in combination with other antibiotics for treating tuberculosis, is subjected to geometry optimization using MTA at HF/6-31+G** level of theory and involves 1113 basis functions. The characteristic frequencies of C=O and N-H in peptide linkages of capreomycin are obtained in ranges of 1885.6–1950.1 and 3782.2–3927.4 cm\(^{-1}\), respectively, while the O-H stretch is obtained at 4197.4 cm\(^{-1}\). The frequency analysis of both of these cases confirmed the local minimum structure with all frequencies being real.

MTA-optimized geometry of \( (H_3BO_3)_{40} \) at HF/6-31++G** level of theory (as discussed in Section 3.2) is subjected to MTA-based frequency analysis. This is carried out using two different fragmentation schemes to appraise the internal consistency of the MTA results. These schemes have R-goodness values of 3 and 4.1 Å, respectively. The results (in terms of the Hessian and IR frequencies) obtained from both the schemes are found to be in excellent agreement (cf. Table 4.2) with each other. The frequencies from both the schemes are real and match to within 0.5 cm\(^{-1}\). The time taken by scheme 1 is 25 hours on five Pentium Dual Core processors at 2.8 GHz with 2 GB of RAM each, while that for scheme 2 is 12 hours
on thirteen nodes of the same configuration. It is worth noting here that the number of basis functions involved in these calculations is 3760, for which actual calculation cannot be performed at all, on any commonly available hardware. For this test system, frequencies within range of 3970–4190 cm\(^{-1}\) correspond to O-H stretches, while those in the range of 1120–1600 cm\(^{-1}\) are due to B-O stretches.

**Table 4.2.** Level of theory, basis set, number of atoms (\(N_{\text{at}}\)) and basis functions (\(N_{\text{BF}}\)) employed for each test system. The maximum and RMS deviations among two and/or three fragmentation schemes of MTA, \(\varepsilon_{\text{MAX}}\) and \(\varepsilon_{\text{RMSD}}\) in Hessian matrix (in a.u.) and vibrational frequencies\(^{\dagger}\) (in cm\(^{-1}\)) are reported.

<table>
<thead>
<tr>
<th>System/Level of theory</th>
<th>Nat/(N_{\text{BF}})</th>
<th>Hessian Matrix</th>
<th>Frequency (in cm(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\beta)-carotene(^\dagger) (a) HF/6-31G*</td>
<td>96/712</td>
<td>(\varepsilon_{\text{MAX}} \times 10^3)</td>
<td>(\varepsilon_{\text{RMSD}})</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.94</td>
<td>0.00</td>
</tr>
<tr>
<td>((\text{H}_3\text{BO}<em>3)</em>{40})(^\dagger) HF/6-31G++G**</td>
<td>280/3670</td>
<td>0.24</td>
<td>0.01</td>
</tr>
<tr>
<td>(\alpha)-cyclodextrin(^\dagger) B3LYP/6-31G*</td>
<td>126/1110</td>
<td>2.65</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.54</td>
<td>0.06</td>
</tr>
<tr>
<td>(\beta)-carotene(^\dagger) (b) B3LYP/6-31G*</td>
<td>96/712</td>
<td>0.51</td>
<td>0.00</td>
</tr>
</tbody>
</table>

* Comparisons only for frequencies greater than 50 cm\(^{-1}\) are reported. The errors in frequencies below 50 cm\(^{-1}\) are also small; however, the percentage error gets inflated.

\(^\dagger\) For \(\beta\)-carotene and \((\text{H}_3\text{BO}_3)_{40}\), a comparison of schemes 1 and 2 is reported.

\(^\ddagger\) For \(\alpha\)-cyclodextrin, comparisons of schemes 1 and 3 (first row) and schemes 2 and 3 (second row) are reported, scheme 3 being the best one.

A cyclic oligosaccharide composed of six glucose units, viz. \(\alpha\)-cyclodextrin, is chosen as another test case at B3LYP/6-31G* level. For MTA-based frequency analysis, \(\alpha\)-cyclodextrin is fragmented with three different schemes. Of these schemes, scheme 3 is the best one with a minimum R-goodness of 3.6 Å as both the other schemes have Rg value of 3.3 Å. Comparisons of the results obtained from schemes 1 and 2 with those from scheme 3 are displayed in Table 4.2, which
evidently bring out self-consistency between the results obtained by the MTA method. The O-H stretching frequencies are obtained in the range of 3562–3745 cm\(^{-1}\), while C-O stretching frequencies are in the range of 1108–1111 cm\(^{-1}\) from all the three schemes.

Within DFT, the Hessian evaluation is generally done numerically. Hence, even if a less accurate two-point formula is used, it is necessary to evaluate energy and gradients at 3\(N_a\)+1 steps, where \(N_a\) is the number of atoms in the system. Therefore, the conventional computation of Hessian at the DFT level for a large molecular system consisting of more than 100 atoms is more demanding in terms of hardware as well as computational time. Within MTA, the calculations are performed on individual fragments which, being much smaller, the cumulative effort required for energy-gradient evaluation at each step for all fragments is much less than that for the whole molecule.

In summary, the test cases presented in this work clearly show that MTA indeed enables calculations of the Hessian matrix and vibrational frequencies that are otherwise not possible on contemporary hardware without significant loss of accuracy. In the next Section, this method of evaluating Hessian matrix and vibrational spectra using MTA-enabled GAMESS is applied to various conformers of novel hosts cucurbit[n]urils. Although the current study is mainly restricted to the HF and DFT levels of theory, with view of versatile nature of MTA, it is applicable to higher correlated methods such as MP2 as well with equal ease. This is demonstrated in the Section 4.4 for the case of acetylene clusters.
4.3 Simulation of Vibrational Spectra of Novel Hosts Cucurbit[n]urils

Cucurbit[n]uril ([C₆H₆N₄O₂]ₙ, n = 5 to 10 and denoted as CB[n]) constitute a family of macrocyclic cavitands. CB[n] homologues can be obtained in different proportions from condensation of glycourea and formaldehyde under acidic conditions by varying the external stimuli such as temperature and the amount as well as type of the acid used. In these novel hosts, the guest molecules can be held via ion-dipole interactions as well as hydrogen bonding. A large affinity of CB[n] hosts towards an electrophile or metal ion and selective binding are explored in separation technology, supramolecular chemistry, catalysis, molecular recognition and nanotechnology. The enantiomers of CB[n] with the methine protons from glycouril unit(s) (cf. Figure 4.1) directing within the cavity, generally known as inverted cucurbit[n]urils (iCB[n], where x is a number of inverted units), have also been synthesized. CB[n] homologues yield a wide range for the cavity size, while different shapes of the cavity can be achieved by inverting the glycouril unit(s). It has also been pointed out that compared to CB[n], varying shape of iCB[n] cavities conduce selective recognition. In order to understand the interactions of CB[n] with different guests, vibrational spectroscopy may be employed as one of the tools. However, the vibrational spectra of CB[n] macrocycles are not reported in the literature as yet. Therefore, analysis of normal vibrations in these hosts may prove useful as an initial step towards studying their interactions with different guest molecules. Besides, the change in the vibrational spectra can also be studied to differentiate between CB[n] and its iCB[n] derivatives. The assignment of peaks in vibrational spectroscopy for larger molecules is only possible by means of a theoretical calculation. These calculations for CB[n] turn out to be computationally demanding even at HF level of theory, owing to larger size of these hosts. In view of this, MTA-based geometry optimization followed by Hessian and vibrational spectra calculation via MTA-enabled GAMESS is performed.
on CB[n] (n = 5 to 8) and their conformers with one inverted glycouril at HF/6-31G* level of theory. The complete spectra of all the conformers considered in this study are given in the appendix (Appendix 4-A) of this Chapter.

Figure 4.1. Glycouril unit with numbering used in the text.

The overall pattern of the IR spectra for CB[n], n = 5 to 8, and their conformers with one inverted glycouril is very similar. The prominent peaks in all these spectra are seen in the ranges 500-1200 cm\(^{-1}\) (region of bending vibrations), 1200-1700 cm\(^{-1}\) (C-C and C-N stretching region), 1800-2200 cm\(^{-1}\) (C=O stretching region) and 3200-3400 cm\(^{-1}\) (C-H stretching region). In general, as n increases from 5 to 8, the spectral patterns in all the above-mentioned regions remain very similar with the gain in intensities of the respective peaks (refer to Appendix 4-A). When one glycouril is inverted in \(i\)CB[n], n = 6 to 8, the spectral pattern changes as compared to the corresponding CB[n] due to splitting of peaks. For instance, in region of C-H stretches i.e. 3200 to 3400 cm\(^{-1}\), only one intense peak appears in spectrum of CB[6] which is split to many smaller peaks in case of \(i\)CB[6]. In all the spectra, the most intense peak is that corresponding to C=O stretch. The changes in this peak for various conformers of CB[n] and \(i\)CB[n] can be utilized for identification of the conformers. A detailed discussion regarding shifts and/or splitting of C=O stretching peaks in these conformers follows.
Figure 4.2. The peaks corresponding to C=O stretches for CB[n], n = 5 to 8 and iCB[n], n = 6 to 8 in vibrational spectra evaluated using MTA at HF/6-31G* level of theory.
The C=O stretching peaks in the vibrational spectra of CB[n] and their inverted conformers is displayed in Figure 4.2. The most striking observation in the vibrational spectra of CB[n], n = 5 to 8 (Figure 4.2), is that there exists a highly intense peak in range 1800 to 2200 cm\(^{-1}\) which is due to the C=O stretching vibrations. This peak is assigned to asymmetric stretching of the two C=O groups (viz. C2=O1 and C2’=O1’ as per the numbering shown in Figure 4.1) within a glycouril unit. In other words, the asymmetric stretch here indicates that when a C2=O1 bond expands, C2’=O1’ bond contracts within each glycouril unit. The larger intensities of the C=O bands are attributed to the simultaneous vibration of the large number of C=O groups present within these systems. Thus, it is but natural that as the number of C=O bonds increases in CB[n] from n = 5 to 8, the peak becomes more intense. Specifically, the intensity of C=O stretch in CB[5] is about 90 units (D\(^2\).amu\(^{-1}\).Å\(^{-2}\)) which increases to ~110 and ~136 units for CB[6] and CB[7] respectively, finally reaching almost double value of ~160 units in case of CB[8]. It is also clearly seen from Figure 4.2 that there is downshift in this C=O stretching peak in spectra of CB[n], as n goes from 5 to 8. For instance, the most intense peak in CB[5] appears at 2042.0 cm\(^{-1}\) while that for CB[8] downshifts to 2016.9 cm\(^{-1}\) through 2029.8 and 2021.4 cm\(^{-1}\) for CB[6] and CB[7] respectively. This implies that on going from CB[5] to CB[8], the C=O bonds become weaker. In the region of C=O stretching frequency, there also exists a small peak in all the spectra for CB[n], n = 5 to 8 and is nearly at constant position in range 2021-2023 cm\(^{-1}\). This peak is assigned to C=O stretches symmetric with respect to the two C=O groups (C2=O1 and C2’=O1’) within glycouril units vibrating in phase and hence, expanding or compressing CB[n] along the symmetry axis during the vibration. A recent paper by Gobre et al. reports\(^{26}\) vibrational spectra for CB[n], n = 5 to 12 at B3LYP/6-31G* level of theory. The changes in spectral patterns for CB[n], n = 5 to 8 obtained by MTA are in agreement with those discussed\(^{26}\) by Gobre et al.
When one of the glycouril unit is inverted, the above mentioned two C=O stretching peaks split into 6 peaks for $i\text{CB}[n]$, $n = 6$ to 8. The C=O stretching frequencies for the inverted glycouril are shifted downwards as compared to the other glycourils implying weaker of C=O bonds in the inverted glycouril. More specifically, these peaks appear at 1965.9 and 1978.6 cm$^{-1}$ for asymmetric and symmetric stretching, respectively, of the two C=O groups in inverted glycouril unit of $i\text{CB}[6]$. The corresponding peaks in $i\text{CB}[7]$ and $i\text{CB}[8]$ are slightly downshifted and appear at 1957.4 and 1971.9 cm$^{-1}$ (for $i\text{CB}[7]$) and 1957.6 and 1973.6 cm$^{-1}$ (for $i\text{CB}[8]$) with the intensities of the order of 10 units.

The most intense peaks in this region are assigned to asymmetric C=O stretch (asymmetric within a glycouril) in rest of the glycouril units. In case of $i\text{CB}[7]$ and $i\text{CB}[8]$, these peaks appear at 2015.0 and 2018.0 cm$^{-1}$ respectively with corresponding intensities being ~120 and 135 units. In $i\text{CB}[6]$, this peak is at 2025.7 cm$^{-1}$ with lesser intensity of ~70 units as compared to $i\text{CB}[7]$ and $i\text{CB}[8]$. The asymmetric C=O stretch within a glycouril in those two adjacent to the inverted one show a peak at 2022.5 cm$^{-1}$ with intensity ~40 units in $i\text{CB}[6]$. However, in $i\text{CB}[7]$, $i\text{CB}[8]$, these peaks due to the glycourils adjacent to the inverted one are 2014 and 2001 cm$^{-1}$ respectively with smaller intensities of the order of 5 units.

Though the changes in only C=O stretching peaks for differentiating the conformers is emphasized in this thesis, it must be noted that the changes and splitting in the other peaks as $n$ increases or one glycouril is inverted, are also important for identification of the conformer.

All the calculations in this work are performed on a cluster of eight Core 2 Quad processors at 2.4 GHz with 4 GB RAM and 250 GB disk each. The largest of these calculations involves 1536 basis functions and took about 60 hours on the above-mentioned hardware. As learned from the benchmarked presented in
previous Section, the actual calculations for these systems are expected to take substantial amount computational time. Also, the benchmarks in Section 4.3 assure the reliability of MTA-based calculations. Therefore, actual calculations are not performed for any of these systems. The main aim of this investigation is to demonstrate the applicability of MTA and it can be concluded from this study that MTA substantially reduces the computational cost as well as time of Hessian matrix and vibrational spectra calculations of large molecular systems.

4.4 Exploring Local Minimum Energy Structures for Acetylene Clusters

The significance of studying molecular clusters bound by $\pi\cdots\pi$ and $\text{CH} \cdots \pi$ interactions has earlier been discussed in Section 3.3. The investigations performed on benzene and acetylene clusters employing better correlated methods as prototypical examples of such weakly bound clusters are also presented there. In Subsection 3.3.3, Plug-In with Gaussian 09$^{27}$ and CFOUR$^{28}$ are used for geometry optimization of acetylene clusters. However, to check whether the optimized geometry is a local minimum, a Hessian and frequency calculation at that geometry is a necessary step.

In the present Section, an attempt to scan local minima of higher clusters of acetylene ($\text{AC}$) viz. octamers and decamers, at MP2 level of theory is made. The initial geometries of ($\text{AC})_8$ and ($\text{AC})_{10}$ are built with the aid of cluster building algorithm$^{29}$ taking the hexamers (reported in Section 3.3.3) as starting structures. The cluster building algorithm and the basic ideas of previously proposed Electrostatic Potential for Intermolecular Complexation (EPIC) model$^{30,31}$ are briefly outlined and illustrated in Section 3.3.2. Through the cluster building algorithm one can add monomer, dimer as well as trimer to the parent cluster system.
To begin with, various \((\text{AC})_8\) structures are built from the three energetically favourable \((\text{AC})_6\) structures reported in Subsection 3.3.3 by adding either 2 acetylene monomers or a T-shaped acetylene dimer. MESP calculated at MP2/3-21G level of theory is employed for building these \((\text{AC})_8\) structures. It has been known that at a fixed geometry, change in basis set does not affect the essential topographical features of MESP, which is why a small basis set viz. 3-21G is chosen for MESP calculations. This exercise resulted into 40 different structures of octamers, which are subjected to geometry optimization MP2/3-21G level of theory. This small basis set is chosen in order to scan the best candidates for energetically most favourable assemblies. Out of these 40 structures, best 15 are further optimized at MP2/aug-cc-pVDZ (aug-cc-pVDZ named as DZ henceforth) employing the tool of MTA. Although geometry optimization of \((\text{AC})_8\) is possible with conventional methods at MP2/DZ level of theory, MTA-Plug-In with Gaussian 09 back-end is used to accelerate this process, resulting into a substantial reduction in computational resources and time. These are further subjected to a single point energy calculation at MP2/TZ (TZ stands for aug-cc-pVTZ). The five most stable \((\text{AC})_8\) are subjected to MTA-based Hessian calculation MP2/DZ level to confirm their local minimum nature.

For building up decamers, use is made of first three lowest energy octameric structures which have been described above. MESP of these best three \((\text{AC})_8\) is calculated at MP2/3-21G level of theory and fed to cluster building algorithm to generate several good guess geometries for \((\text{AC})_{10}\). Following this, these decamer geometries are scanned at MP2/3-21G level, out of which, the best 10 lowest energy decamers are chosen for MTA-based optimization at MP2/DZ level of theory and single point energy calculation at MP2/TZ for the final optimized geometry. Here also, the MTA-based Hessian calculations are performed on the 5 most stable
structures at MP2/DZ level. The stabilization energies at MP2/CBS limit for all these octamers and decamers are evaluated as per Eq. 3.1 and 3.4.

**Table 4.3.** MTA-based stabilization energies $\Delta E$ (in kcal/mol) of acetylene octamers $(AC)_8$ and decamers $(AC)_{10}$ for MP2/XZ ($X = D, T$) and at CBS limit.

<table>
<thead>
<tr>
<th>System</th>
<th>$\Delta E_{DZ}$</th>
<th>$\Delta E_{TZ}$</th>
<th>$\Delta E_{CBS}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oct-01</td>
<td>-32.3</td>
<td>-27.1</td>
<td>-24.9</td>
</tr>
<tr>
<td>Oct-02</td>
<td>-32.3</td>
<td>-26.9</td>
<td>-24.6</td>
</tr>
<tr>
<td>Oct-03</td>
<td>-32.0</td>
<td>-26.6</td>
<td>-24.3</td>
</tr>
<tr>
<td>Oct-04</td>
<td>-31.7</td>
<td>-26.4</td>
<td>-24.2</td>
</tr>
<tr>
<td>Oct-05</td>
<td>-31.7</td>
<td>-26.5</td>
<td>-24.3</td>
</tr>
<tr>
<td>Dec-01</td>
<td>-44.7</td>
<td>-37.3</td>
<td>-34.2</td>
</tr>
<tr>
<td>Dec-02</td>
<td>-43.5</td>
<td>-36.2</td>
<td>-33.1</td>
</tr>
<tr>
<td>Dec-03</td>
<td>-42.8</td>
<td>-35.7</td>
<td>-32.7</td>
</tr>
<tr>
<td>Dec-04</td>
<td>-42.4</td>
<td>-35.2</td>
<td>-32.2</td>
</tr>
<tr>
<td>Dec-05</td>
<td>-42.3</td>
<td>-35.1</td>
<td>-32.1</td>
</tr>
</tbody>
</table>

The stabilization energies for the lowest 5 $(AC)_8$ and $(AC)_{10}$ are reported in Table 4.3, along with the corresponding MP2/TZ MTA-single point energies. These stabilization energies are very close, with all energies lying within 0.6 kcal/mol (cf. Table 4.3), of each other. Structures of the best three $(AC)_8$ and $(AC)_{10}$ are displayed in Figure 4.3. It should be noted that all best five structures of $(AC)_8$ and $(AC)_{10}$ listed in Table 4.3 are found to be non-planar. In the lowest energy octamer, i.e. Oct-01 (cf. Figure 4.3) all the interactions are T-shaped and maximum in number among all the other octameric structures scanned. The second best structure, Oct-02 exhibits lesser number of T-shaped interactions than in Oct-01. It also exhibits a few $\text{CH} \cdots \pi$ stacking interactions. This shows that the T-shaped interactions play an important role in stabilization of geometry of an acetylene cluster. The most stable decamer, Dec-01 has maximum number of T-shaped interactions, although there are some stacking interactions present as well. All the decameric structures
in Figure 4.3 show a kind of zig-zag arrangement of monomers. In the most stable decamer, the monomers are arranged in highly ordered manner while the disorder goes on increasing as the energy of the cluster increases.

Figure 4.3. Best three MTA-optimized structures of $(AC)_8$ and $(AC)_{10}$ at MP2/DZ level of theory.

The efficacy of MTA in calculating highly accurate Hessian and frequencies, with large computational time advantage, has been already demonstrated in Section 4.2 for structurally diverse, spatially extended molecules. However, to assure the similar accuracy as for the test cases presented in Section 4.2, a test calculation for Hessian matrix and vibrational spectrum is performed for Oct-01 at MP2/3-21G level of theory, for which conventional calculation is readily possible. The earlier studies on MTA have revealed that the accuracy of MTA is independent
of the basis set and level of theory used (also refer to Subsection 2.2.3). Therefore, the accuracy of MTA for 3-21G and DZ basis sets is expected to be similar. The MTA-estimated Hessian is computed based on only tetramers as main fragments and their overlaps. On comparing the MTA-based Hessian matrix with the corresponding actual one, the maximum error in the Hessian element is seen to be $4.4 \times 10^{-4}$ a.u. for the actual value of 0.0423 a.u., with $2.4 \times 10^{-5}$ a.u. as the RMS error. On similar lines, when MTA-based dipole derivative tensors (calculated as per Eq. 1.83) are compared, the maximum error is 0.0075 (corresponding actual value being 0.2365 a.u.) while the RMS error is 0.0020 a.u.

**Figure 4.4.** A comparison of the vibrational spectrum of Oct-01 obtained by MTA with that obtained by actual calculation at MP2/3-21G level of theory.
A comparison of MTA-based vibrational spectrum for Oct-01 with that obtained by conventional (actual) calculation on the entire cluster is shown in Figure 4.4. MTA-based vibrational frequencies indeed turn out to be real and match well with their counterparts in actual calculation, while intensities show a reasonably good agreement. In both of these spectra, the vibrational frequencies appear in ranges of 700 to 850 cm$^{-1}$ and 3400 to 3500 cm$^{-1}$. The most intense peaks of C=C stretching are around 780 to 800 cm$^{-1}$ in both MTA-based as well as actual spectra. However, intensity of one of these peaks is slightly overestimated by MTA. The intensities of C-H stretch modes in 3400 to 3500 cm$^{-1}$ region are quite accurately mimicked by MTA as in actual spectra.

**Figure 4.5.** MTA-based vibrational spectra for two best acetylene octamers and decamers at MP2/DZ level of theory.
With this assurance of accuracy, MTA-based Hessians for best five acetylene octamers are evaluated at MP2/DZ level of theory. All the frequencies for these five octamers are found to be real confirming the local minimum nature of these structures. The vibrational spectra for Oct-01 and Oct-02 calculated by employing MTA are shown in Figure 4.5. The overall nature of the spectrum and the regions of vibrational frequencies (viz. 700-800 cm\(^{-1}\) and 3350-3450 cm\(^{-1}\)) are similar for both the octamers. However, in case of Oct-02, some of the C-H stretches are slightly downshifted and hence, more widespread in the region of 3350-3450 cm\(^{-1}\).

As in the case of octamers, the MTA-based Hessian calculations are carried out on five best decamers at MP2/DZ level of theory. All the frequencies for each decamer are also found to be real, confirming the nature of the structures as stationary points on the PES. Yet again, the main features of Dec-01 and Dec-02 are similar, and the spectral bands in Dec-02 are more widespread (cf. Figure 4.5). Also, C=\(\text{C}\) stretches in spectra of Dec-01 appear more intense and close to each other, while those for Dec-02 are farther and less intense as compared to those of Dec-01.

### 4.5 Concluding Remarks

This Chapter deals with benchmarking and application of linear scaling method MTA for calculating Hessian matrix and vibrational spectra of large molecules and clusters at \textit{ab initio} level of theory. The initial testing of MTA-based Hessian matrix with reference to the corresponding actual one shows that the former one is highly accurate with maximum and RMS errors of the order of $10^{-3}$ and $10^{-4}$ a.u. respectively. Consequently, the vibrational frequencies obtained from this Hessian are also accurate enough with the errors less than 1 cm\(^{-1}\). It must be noted that the frequencies reported in this Chapter are not scaled, however, for a comparison with experimental spectra, whenever available, an appropriate scaling
factor could be employed. Thus, in general, it can be concluded that MTA is successful in estimating the Hessian and vibrational frequencies of large molecular systems. Besides the accuracy, MTA also shows substantial time advantage and reduction in memory and CPU requirements over the actual calculation for the test cases examined. The Hessian and frequency calculation, without any symmetry constraints, for a molecule such as \((\text{H}_3\text{BO}_3)_{40}\) with 3670 basis functions, is one of the largest reported in the literature.

After assuring the reliability of MTA for computing vibrational spectra at HF and DFT levels of theory, the technique is applied further to novel hosts cucurbit\([n]\)uril with \(n = 5\) to 8 and their conformers with one glycouril unit inverted. The vibrational spectra of these molecules may be employed to differentiate between the conformers. This is illustrated by the change in most intense peaks due to C=O stretching vibrations in all of these spectra. As \(n\) increases from 5 to 8, the most intense peak in spectra of CB[5] at 2042.0 cm\(^{-1}\) downshifts to 2016.9 cm\(^{-1}\) in spectra of CB[8] through 2029.8 and 2021.4 cm\(^{-1}\) for CB[6] and CB[7] respectively. Also, these peaks gain intensities from CB[5] to CB[8]. Further, when the spectrum of CB\([n]\) is compared to that of \(i\)CB\([n]\), these two peaks due to C=O stretching vibrations are split into a pattern of about 6 peaks. Among the \(i\)CB\([n]\), the systems can be identified on the basis of positions and intensities of these 6 peaks due to C=O stretches. Though the entire vibrational spectra of these systems are not compared, this investigation emphasizes on the usefulness of MTA for calculations of vibrational spectra of large molecules.

Again exploiting the versatile nature of MTA, the method is applied for confirming minimum nature of acetylene octamers and decamers at MP2/aug-cc-pVDZ level of theory via Plug-In utility. The best candidates for \((\text{AC})_8\) and \((\text{AC})_{10}\) are subjected to geometry optimization via MTA-Plug-In with back-end of G09 at the
aforesaid level of theory. The final optimized geometries of these systems are further fed to MTA-based Hessian calculations at the same level of theory. All the frequencies for all of these clusters are found to be real, confirming the local minimum nature on the PES. This also reaffirms the trustworthiness of MTA-based geometry optimization. It may be of interest to a chemist that the vibrational spectra of these clusters are quite different from each other and may be useful in differentiating these clusters if they could be experimentally generated.

Overall, the cases presented in this Chapter demonstrate the efficiency of MTA in handling the expensive calculations of geometry optimization and Hessian matrix of large clusters at HF, DFT and MP2 levels of theory, without sacrificing the accuracy.
References


Appendix 4-A

MTA-based vibrational spectra for CB[n], n = 5 to 8 and iCB[n], n = 6 to 8, calculated at HF/6-31G* level of theory.