Natural Linear-Scaled Coupled-Cluster Theory with Local Transferable Triple Excitations: Applications to Peptides

Thomas F. Hughes, Norbert Flocke, and Rodney J. Bartlett*

University of Florida, Department of Chemistry, Quantum Theory Project, Gainesville, Florida 32611

Received: January 18, 2008; Revised Manuscript Received: March 27, 2008

The natural linear-scaled coupled-cluster (NLSCC) method (Flocke, N.; Bartlett, R. J. *J. Chem. Phys.* 2004, *121*, 10935) is extended to include approximate triple excitations via a coupled-cluster with single, double, and triple excitation method (CCSDT-3). The triples contribution can potentially be embedded in a larger singles and doubles region. NLSCC exploits the extensivity of the CC wave function to represent it in terms of transferable natural localized molecular orbitals (NLMOs) or functional groups thereof that are obtained from small quantum mechanical (QM) regions. Both occupied and virtual NLMOs are local because they derive from the single-particle density matrix. Noncanonical triples amplitudes are avoided by applying the unitary localization matrix to the canonical CC wave function for a QM region. A generalized NLMO code interfaced to the ACES II quantum chemistry software package provides NLMOs for the relevant number of atoms in a given functional group. Applications include linear polyglycine and the pentapeptide met-enkephalin, which was chosen as a more realistic three-dimensional system with nontrivial side chains. The results show that the triples contributions are quite large for aromatic bonds suggesting an interesting active space method for triples in which different bonds require different excitation levels. The NLSCC approach recovers a very large percentage (>99%) of the CCSD or CCSDT-3 correlation energy.

1. Introduction

Highly accurate electronic structure methods, such as coupled-cluster (CC) and configuration interaction (CI), are troubled by the task of including all possible permutations of one, two, three, etc. electrons among a large number of single-particle basis functions. The excitation operator, T, which accomplishes this is oftentimes truncated after the two-particle piece, $T = T_1 + T_2 + T_3 + \ldots \approx T_1 + T_2$, where T_N is an N-particle excitation operator, leading to singly (S) and doubly (D) excited determinants. Both the CCSD and CISD methods scale as $O(o^2v^4)$, where o and v are respectively the number of occupied and virtual orbitals, but only the exponential ansatz, unique to CC theory, additionally includes disconnected parts of triple and quadruple excitations. These disconnected triple and quadruple excitations, like T_1T_2 and T_2T_2, respectively, make truncated CC converge to the full CI (FCI) much more quickly than truncated CI, thus making CC the method of choice for high-level applications to medium-sized molecules. The remaining pieces of the triple excitation manifold are said to be connected and may be required for $\approx$1 kcal mol^{-1} accuracy in the energy and similar accuracy in other properties. The necessity of connected triples is understood from an order-by-order comparison with many-body perturbation theory (MBPT) in which it is found that the triply excited manifold is dominated by connected T_3 as opposed to T_1T_2 or T_1^3. This is in contrast to the dominant term in the quadruply excited manifold which is disconnected T_2^2 as opposed to connected T_4. These connected contributions are not obtained until explicit inclusion of T_3 as in CCSD(T), CCSDT-x (x = 1a, 1b, 2, 3), or CCSDT. These methods scale as noniterative $O(o^3v^4)$, iterative $O(o^3v^4)$, and iterative $O(o^3v^5)$, respectively, making these CC methods, without modification, prohibitive for large molecules. The exponential ansatz used in CC theory makes it a size-extensive method, which is a necessary condition for a system with a large number of electrons. For large molecules, a size-inextensive linear ansatz, for example as in truncated CI, would give a vanishing correlation energy per electron.[1,2]

Local correlation methods aim to circumvent the nonlinear scaling by exploiting an underlying simplicity in both the electronic structure and computational procedure as in the natural linear-scaled CC (NLSCC) method.[3] The NLSCC method has been shown to give very accurate correlation energies for large molecules in which the rate limiting step is the size of the quantum mechanical (QM) regions. This region can be determined self-consistently thereby making these methods systematically improvable. For this region other reduced scaling methods, which are outside of the local approximation, can offer advantages.[4–9] These methods become especially useful for three-dimensional systems where an accurate local correlation treatment can require large QM regions to represent weak interactions. In a localized basis the correlation space for a given occupied orbital does not grow with system size thereby reducing the number of free parameters and allowing unimportant interactions to be eliminated from the amplitude equations.[10] Furthermore, the method insists upon transferability of electronic structure units and as such has the potential to provide transferable electronic structure, much the way as geometric structure depends upon transferable geometric units. Therefore, high-level local correlation methods have the potential to become a standard, systematically improvable tool for calculating energies of very large molecules.

Natural localized molecular orbitals (NLMOs)[11,12] are used to introduce a scale that offers certain advantages to local correlation methods and NLSCC in particular. One benefit of using NLMOs is that the virtual space is composed of localized orthogonal orbitals that are obtained with little convergence difficulties even for large, diffuse basis sets. In the NLMO search

* To whom correspondence should be addressed. E-mail: bartlett@qtp.ufl.edu.

10.1021/jp800516q CCC: $40.75
Published on Web 06/11/2008

for n-centered orbitals, where n is typically small $n < N$, with N the number of centers in the molecule, only small diagonalizations of the single-particle density matrix are required ensuring that the NLMO determination is quite fast.[12] In contrast to other localization methods which use an initial set of orbitals, the NLMO procedure is based on a single-particle density matrix, $\gamma = \mathbf{C}^{\dagger}\mathbf{S}^{\dagger}\mathbf{PSC}$ with **C**, **S**, and **P** being the orbitals, overlap, and single-particle density matrix in atomic orbitals, which can be correlated or uncorrelated. For example, localized correlated orbitals may have certain properties that are more desirable in treating QM regions that would otherwise have been quite large. Restricted as well as unrestricted NLMOs can be obtained from the corresponding restricted Hartree−Fock (RHF) or unrestricted (UHF) density matrices. The NLMOs derive from non-HF natural bond orbitals (NBOs) resulting in the formation of bond/antibond pairs. As previously mentioned, in the ACES II NLMO program the bonds can be n-centered with $1 \leq n \leq N$ where N is the number of centers in the molecule and are classified as occupied cores ($n = 1$), lone pairs ($n = 1$), bonds ($n \geq 2$), virtual antibonds ($n \geq 2$) and Rydbergs ($n = 1$). The $n > 2$-centered bonds are useful for describing delocalized moieties, for example, aromatic molecules, without resorting to more artificial Kekulé structures.

A decomposition of the electronic structure into NLMO and/or functional group (groups of $n > 2$ NLMOs) contributions is possible due to the high transferability of the correlated amplitudes expressed in these localized orbitals. Another useful quality with regard to defining functional groups is that the NLMO procedure maintains σ/π separation. By transferability we mean that the orbitals are roughly independent of their environment. The remarkable transferability of localized chemical bonds and/or functional groups, and their associated properties among the ground states of molecules is one of the most fundamentally important and useful concepts in chemistry. Exceptions to this rule include certain pathological cases, for example, D_{3d} cyclohexane cation, where the unpaired hole remains delocalized, despite the fact that all the other electrons in the molecule are localized. For example, the concept of transferable groups in conjugated systems was recently investigated for substituted polyenes up to second-order many-body perturbation theory (MBPT2) where it was found that effects due to substitution propagate to three or four methylene group distances.[13] Not surprisingly, the majority of applications of local correlation methods have been on saturated molecules for which there is a large homolumo gap and therefore a more local electronic structure.

Correlated methods such as MBPT2 and CC are a minimal requirement for predicting effects such as dispersion interactions, which can play important collective roles in large, albeit conventionally unattainable, systems, such as peptides. There are local CC methods that have been designed for single and double excitations;[3,10,14–27] however, less work has been done on extending the local CC methods to higher excitations, for example, triple excitations.[28–33] Disconnected T_1T_2 or even T_1^3 obtained with CCSD accounts for some of the triples contribution to the correlation energy, but typically the connected T_3 contributions make significant contributions. It is expected that the triples amplitudes are longer-ranged than the doubles amplitudes because only two of the three occupied-virtual pairs need to be spatially close to make a significant contribution. This could result in significant modification of local correlation methods that were intended for CCSD. Additionally, as the excitation level increases the nonlinearity of the CC equations leads to complicated terms in which there are one, two, three, etc. excitation operators making it difficult to determine if the overall contribution is even local.

As an example, consider the local CCSD methods that have been extended to include connected triple excitations for both noniterative CCSD(T)[31,33] and iterative CCSDT-1b.[32] The easiest way to include connected triples contributions is via perturbation theory giving the well-known CCSD(T) method[34] and its non-HF generalization.[35] Perturbative triples corrections like CCSD(T) in a localized basis require both an iterative solution and storage of the triples amplitudes,[35] which can destroy performance due to large memory requirements. It is important to have approximations that eliminate the need to store the triples amplitudes. Note that for local perturbative triples approximations, like local CCSD(T), an approximate triples energy contribution is added to an already approximate singles and doubles energy. This is in contrast with approximate iterative triples methods in which the triples are coupled to the single and double equations, and vice versa. Due to the fact that CCSDT-3 contains many more nonlinear triples contributions than CCSD(T) or CCSDT-1x (x = a, b) the triples excitation regions may need to be larger than the double excitation regions. Current results indicate that the T_3 contribution to the CCSDT-3 wave function is only slightly longer-ranged than the doubles and can thus be calculated using local correlation algorithms similar to those used previously for CCSD. This paper considers the effect of triple excitations at the CCSDT-3 level with the NLSCC method.

2. Theoretical Method

As previously mentioned, the CCSD(T) method provides the easiest connected triples contribution; however, as in the context of second-order with noncanonical HF orbitals, the Fock matrix is block-diagonal, meaning that perturbative methods like CCSD(T) must be solved iteratively, thereby requiring storage of the corresponding triples amplitudes.[35] One disadvantage of localized orbitals is that they are not eigenfunctions of any common energy operators,

$$h|p\rangle = \sum_q \varepsilon_{pq}|q\rangle \tag{1}$$

where h is some single-particle energy operator, p and q are localized occupied or virtual orbitals, and ε is an energy matrix. The orbitals might diagonalize some other energy operator,

$$h^{\text{eff}}|p\rangle = \varepsilon_p^{\text{eff}}|p\rangle \tag{2}$$

or they might diagonalize some other operator, for example the NLMOs block-diagonalize the density matrix. This results in perturbative methods needing to be solved iteratively. This iterative solution negates one of the most useful properties of perturbative methods in that they are noniterative for canonical orbitals. If the equations have to be iterated, it is worthwhile to include more terms per iteration amounting to a CC calculation. Because the overhead is in the storage of triples as opposed to the inclusion of more terms, we choose to work with CCSDT-3, which is as close to CCSDT as possible while still being $O(N^7)$. Additionally, for CCSDT-3 we recognize that with canonical or semicanonical orbitals the triples amplitudes need not be stored in contrast to CCSDT.

The NLSCCSDT-3 method follows straightforwardly from the NLSCCSD method.[3] The ground state NLSCCSDT-3 wave function is rewritten from the CCSDT-3 wave function without loss of generality as

$$|\Psi\rangle = e^{T_1+T_2+T_3}|0\rangle = \prod_i e^{T_1^i+T_2^i+T_3^i}|0\rangle \quad (3)$$

where $|0\rangle$ is the single-determinant reference function. The operator T_3^i for an excitation specific to orbital i is given as

$$T_3 = \sum_i T_3^i = \sum_i \left(\frac{1}{36}\sum_{jkabc} t_{ijk}^{abc} a^\dagger b^\dagger c^\dagger kji\right) \quad (4)$$

with $i, j, k, \ldots \in \{occ\}$ and $a, b, c, \ldots \in \{vir\}$. The assumption is that the operators T^i in the NLMO basis only need to be locally correlated due to the at worst R^{-3} decay of the "dipole−dipole" cluster amplitudes. We approximate T_1^i and T_2^i by only summing over a third QM region, QM3, and likewise for T_3^i with respect to a second QM region, *QM*2, both composed of orbitals that are spatially close and closer to i, respectively. This is shown below for triple excitations,

$$\sum_{jkabc} = \sum_{jkabc\in QM2} + \sum_{P(jkabc)\notin QM2} \approx \sum_{jkabc\in QM2} \quad (5)$$

with $P(jkabc)$ a permutation operator which introduces local orbitals outside of QM2 . Orbital i is defined to be in the first QM region, QM1, with QM1 ⊂ QM2 ⊂ QM3 giving T_3^i as

$$T_3^i \approx \frac{1}{36}\sum_{jkabc\in QM2} t_{ijk}^{abc} a^\dagger b^\dagger c^\dagger kji \qquad i \in \text{QM1},\ j,\ k, a,\ b, c \in \text{QM2} \quad (6)$$

and T_2^i as

$$T_2^i \approx \frac{1}{4}\sum_{jab\in QM3} t_{ij}^{ab} a^\dagger b^\dagger ji \qquad i \in \text{QM1},\ j, a, b \in \text{QM3} \quad (7)$$

The single excitation operator, T_1^i, is given as

$$T_1^i \approx \sum_{a\in QM3} t_i^a a^\dagger i \qquad i \in \text{QM1}, a \in \text{QM3} \quad (8)$$

Note that the QM1 regions are disjoint from one another as opposed to the QM2 and QM3 regions. This reduces the number of singles and doubles parameters and significantly reduces the number of triples parameters leading to linear or even sublinear scaling with respect to the number of basis functions. By adopting a hybrid approach to NLSCC where NLSCCSDT-3 is performed in a small region, QM2, and NLSCCSD is performed in a larger region, QM3, it is possible to take advantage of simplifications that manifest in other hybrid methods that mix small CCSD regions with large MBPT2 regions. These latter hybrid schemes are based on the assumption that at large enough distances the localized electron pairs are only weakly interacting and are therefore well approximated by MBPT2, which does not allow these pairs to influence one another to the extent that they do within CC theory. A hybrid approach to treating triples within the NLSCC framework, as opposed to a full NLSCCSDT-3 treatment, should benefit from similar simplifications where the focus is on a small triples region in which the triples influence one another more strongly than those which are scattered over a larger region. For those systems that are well described by single reference, as opposed to a multireference system, it is expected that the hybrid NLSCC methods for triples would compare well with a full NLSCCSDT-3 method.

Amplitudes in QM2−QM1 have an indirect influence on those amplitudes T^i. As an example, consider a QM2 region with three occupied orbitals, i, j, and k and an arbitrary number of virtuals and let $i \in$ QM1 and $j, k \in$ QM2−QM1. The T_2 equation will depend on t_{ij}, t_{ik}, and t_{jk}, where we have dropped virtual orbital labels for convenience. It is tempting to think about dropping the t_{jk} term and representing the NLSCC wave function entirely in terms of QM1 → QM1 and QM1 → QM2 excitations; however, the QM2 → QM2 excitations have an important, albeit indirect, effect on t_{ij} and t_{ik}. Obviously, those outside of the largest QM region defined do not make a contribution and thus are neglected within NLSCC. This choice of QM regions should be done self-consistently to ensure that we are including any important longer-range physics and an automated QM determination is currently underway. We should also choose the regions to minimize tail contributions. In our current implementation we choose these regions using Molden[36] based on results from calculations performed on substituted alkanes of varying lengths to be discussed in the Results and Discussion.

In the limit of perfect localization where $|0\rangle = |i\rangle|j\rangle|k\rangle\ldots$, the NLSCCSDT-3 wave function becomes a product, by virtue of its size-extensive exponential ansatz,

$$|\Psi\rangle = |\Psi_i\rangle|\Psi_j\rangle|\Psi_k\rangle\ldots \quad (9)$$

Due to the orthogonality tails in the orthonormal NLMO basis there will be higher order terms shown in the following many-body expansion,

$$|\Psi\rangle = |\Psi_i\rangle|\Psi_j\rangle|\Psi_k\rangle\ldots + |\Psi_{ij}\rangle|\Psi_k\rangle\ldots + |\Psi_{ik}\rangle|\Psi_j\rangle\ldots + |\Psi_{jk}\rangle|\Psi_i\rangle\ldots + |\Psi_{ijk}\rangle \quad \ldots (10)$$

To expedite the evaluation of triples within NLSCCSDT-3, it is useful to first consider the dominant linear term in the T_3^i equation,

$$\langle {}_{ijk}^{abc}|(W_N T_2)_c|0\rangle \quad (11)$$

with

$$H_N = \sum_{pq} F_{Npq}\{p^\dagger q\} + \frac{1}{4}\sum_{pqrs} W_{Npqrs}\{p^\dagger q^\dagger sr\} \quad (12)$$

and

$$H_N = H - \langle 0|H|0\rangle \quad (13)$$

which is the cornerstone of CCSD(T) and CCSDT-1x (x = a, b) methods. Within the context of NLSCCSDT-3, nonlinear contributions to T_3, for example, $\langle {}_{ijk}^{abc}|(W_N T_2^2)_c|0\rangle$, allow the localized regions to indirectly influence one another, for example, by certain strong inter-regional excitations, without destroying the locality or transferability of the individual regions. These types of terms are incomplete in linear versions such as CCSDT-1x (x = a, b) or CCSD(T). Nonlinear terms involving T_3 are limited to the doubles equation $\langle {}_{ij}^{ab}|(W_N T_1 T_3)_c|0\rangle$ within the NLSCCSDT-3 approximation, meaning that all other terms involving T_3 are linear. Equations 14, 15, and 16 show the coupled nonlinear T_1, T_2, and T_3 CCSDT equations, where in the latter those beneficial terms that were previously mentioned, namely, those of CCSDT-3, are underbraced. The CCSDT-3 T_1 and T_2 equations are the same as they are for CCSDT. Prefactors have been dropped for simplicity.

$$\langle {}_i^a|(H_N(1 + T_2 + T_1 + T_1T_2 + T_1^2 + T_1^3 + T_3))_c|0\rangle = 0 \quad (14)$$

$$\langle {}_{ij}^{ab}|(H_N(1 + T_2 + T_2^2 + T_1 + T_1T_2 + T_1^2 + T_1^2T_2 + T_1^3 + T_1^4 + T_3 + T_1T_3))_c|0\rangle = 0 \quad (15)$$

It can be seen that CCSDT-3 is as close to CCSDT as possible, but still O(N^7) and does not require storing the T_3 amplitudes

$$\langle {}^{abc}_{ijk}|(H_N(\underbrace{T_2}+\underbrace{T_3}+\underbrace{T_2^2}+\underbrace{T_1T_2}+T_2T_3+T_1T_3+ \underbrace{T_1^2T_2}+\underbrace{T_1T_2^2}+T_1^2T_3+\underbrace{T_1^3T_2})_c|0\rangle = 0 \tag{16}$$

in a canonical or semicanonical basis because in those bases the T_3-into-T_3 contributions vanish. For example, the only T_3-into-T_3 contribution in CCSDT-3 is

$$\langle {}^{abc}_{ijk}|(F_NT_3)_c|0\rangle \tag{17}$$

which for canonical HF, or semicanonical orbitals, $F_{pq} = \varepsilon_p\delta_{pq}$ becomes diagonal, establishing that the set of triples need not be stored. The CCSDT-3 method is correct through fourth order in the energy and second-order in the wave function.

In the case of NLMOs there are off-diagonal occupied−occupied and virtual−virtual terms in F_N that require the amplitudes be stored in the conventional implementation. The semicanonical transformation that eliminates these terms in the case of a non-HF reference function,[35] thereby preventing storage, cannot be used here because it would simply return the delocalized canonical HF orbitals. In NLSCCSDT-3 we avoid storing triples amplitudes in the localized basis by first solving the CC equations for our small QM regions in the canonical basis and then subsequently unitarily transforming amplitudes and integrals into NLMOs, shown below using the summation convention.

$$t^{a'b'\ldots}_{i'j'\ldots} = U_{aa'}U_{bb'}U_{ii'}U_{jj'}\ldots t^{ab\ldots}_{ij\ldots} \tag{18}$$

$$w_{p'q'r's'} = U_{rr'}U_{ss'}U_{pp'}U_{qq'}w_{pqrs} \tag{19}$$

That is we apply the localization matrix to the canonical CC wave function from a QM region instead of to the reference function. This a posteriori transformation does not affect our choice of QM regions because the transferability of the NLMOs necessitates that they can be determined from either a global or fragmented reference function. In eqs 18 and 19 U is the localization transformation obtained in the NLMO procedure from $C' = CU$ where C' are the NLMOs and C are the canonical HF orbitals. The distinction here in comparison to many other local CC methods is that by virtue of the fact that CC is an infinite-order method, in comparison to noniterative perturbative methods, we are free to apply the localization to the final correlated wave function instead of to the reference function. Therefore, within NLSCCSDT-3 we do not have to store the triples amplitudes.

The total correlation energy, $E = \langle 0|(H_Ne^T)_c|0\rangle = \langle 0|\bar{H}|0\rangle$,

$$E = \sum_{ia} f_i^a t_i^a + \frac{1}{4}\sum_{ijab} w_{ij}^{ab}(t_{ij}^{ab} + t_i^a t_j^b - t_i^b t_j^a) \tag{20}$$

is then a sum of orbital correlation energies,

$$E = \sum_i E_i \tag{21}$$

Using our truncation of excitation operators, we have $E_i = E_i(\mathrm{QM1}) + E_i(\notin \mathrm{QM1})$ and so we choose i to minimize E_i ($\notin$ QM1) obtaining $E_i \approx E_i(\mathrm{QM1})$ to a good approximation. This obviously becomes exact in the limit of QM2 being the entire molecule. If we designate E^{T} as the total CCSDT-3 correlation energy, we have from eq 21 that

$$E^{\mathrm{T}} = \sum_i E_i^{\mathrm{T}} \tag{22}$$

in terms of triply correlated NLMO and/or functional group energies. We consider the following heirarchy in terms of the QM regions within the language of NLSCC, QM1 ⊂ QM2 ⊂ QM3, where ideally we would have triples contributions to QM1 from as large a region as possible given by QM3. If we consider rewriting E_i^{T} in terms of QM2 and QM3 contributions as

$$E_i^{\mathrm{T}} = E_i^{\mathrm{T}}(\mathrm{QM3}_i) + E_i^{\mathrm{T}}(\mathrm{QM2}_i) \tag{23}$$

an interesting hybrid approach can be developed by assuming that $E_i^{\mathrm{T}}(\mathrm{QM3}_i) \approx E_i^{\mathrm{D}}(\mathrm{QM3}_i)$, thereby avoiding longer-range contributions to the triples equation resulting in

$$E_i^{\mathrm{T}} \approx E_i^{\mathrm{D}}(\mathrm{QM3}_i) + E_i^{\mathrm{T}}(\mathrm{QM2}_i) \tag{24}$$

where E_i^{D} is a doubly correlated bond energy. Hybrid approaches are labeled as NLS(CCSDT-3/CCSD). This approximation is justified in the case of NLS(CCSD/MBPT2) by considering that the effects of coupled excitations that CC offers in comparison to MBPT2 would be less important at large distances. At large distances the localized electron pairs are only weakly interacting and so a theory in which the pairs are not allowed to influence one another, such as MBPT2 in contrast to CC, is appropriate.

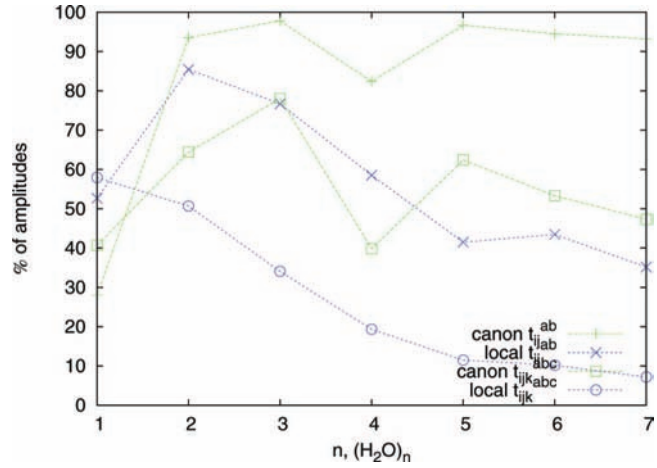

Figure 1. The percentage of canonical (canon) and NLMO (local) based CCSDT-3 $|T_2|$ and $|T_3|$ amplitudes, which are above the 5.0×10^{-8} threshold as a function of C_1 water clusters of varying size.

3. Results and Discussion

All calculations were performed using a modified version of the serial ACES II,[37] quantum chemistry software package. With the exception of those for met-enkephalin, results were obtained using a 375 MHz power3 processor using a maximum of 4 of the available 8 GB of shared memory over four processors. The calculations on met-enkephalin were performed on a 8 dual-core ia64 processor SGI Altix machine with 256 GB of shared memory using multithreaded libraries. Such calculations made use of some recent advances in the ACES II code made to take advantage of large shared memory machines. These advances included updating some old memory limitations to allow for large reference functions (739 basis functions for the case of cc-pVDZ met-enkephalin) and in-core contractions in the context of many-body methods.

Geometries for polyglycine were taken from ref 10. The geometry for met-enkephalin was taken from the Protein Data Bank (1PLW.pdb, model 1).[38] Geometries of test systems are available upon request. Given that the NLSCC method is implemented as a pilot code in ACES II, we use a manual determination of the QM regions using Molden.[36]

We first consider a quasi-linear translationally periodic system because for these types of systems the arguments in favor of

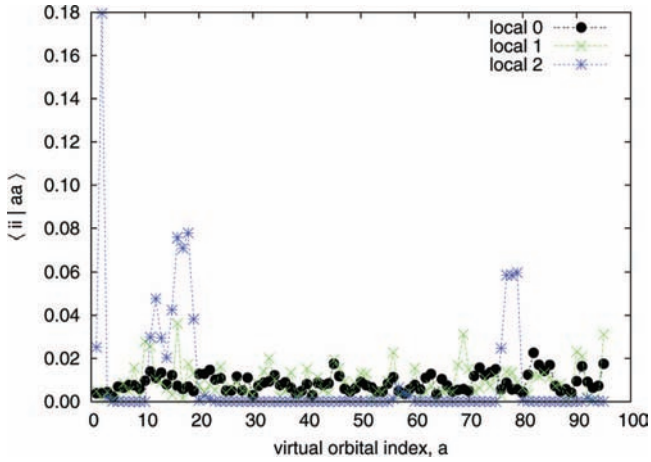

Figure 2. Profile of the absolute value of diagonal two-electron integrals, $\langle iiaa \rangle$, with respect to different virtual bases from a C_1 $(H_2O)_5$ calculation. The key is understood as follows with the number giving the degree of localization: local 0 (canonical occupied and canonical virtual), local 1 (localized occupied and canonical virtual), and local 2 (localized occupied and localized virtual).

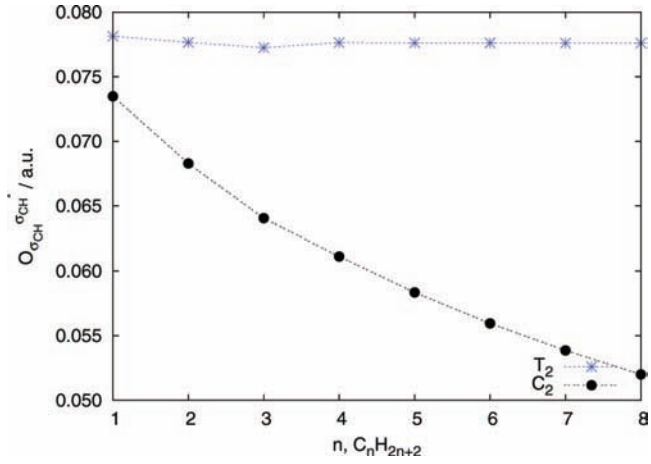

Figure 3. Diagonal $|T_2|$ (CCSD) and $|C_2|$ (CISD) amplitudes for a σ_{CH}-bond belonging to a methyl group as a function of alkane length. The plot demonstrates that for the size-extensive CCSD wave function the amplitudes are transferable. This is in contrast to the CISD wave function which is size-inextensive.

locality are most simply understood. Polyglycine is meant to represent the simplest peptide but still containing a variety of functional groups. For example, it has a three-center peptide-bond. Met-enkephalin is meant to represent a more realistic application because it is three-dimensional and has nontrivial side chains. It is a conformationally flexible neuropeptide with a high affinity for the opiate receptor[38] for which there has been recent interest from the theoretical chemistry community.[39–42] For polyglycine we limit ourselves to two-center bonds, although we certainly could have used three-center bonds, whereas for met-enkephalin we use up to six-centered bonds. We use a cc-pVDZ basis unless specified otherwise; however, for more practical applications a triple-ζ or better basis should be used.

Three of the figures in this article establish transferability by plotting a quantity for a functional group, for example a methyl group or a σ -bond, belonging to one side of a quasi-linear molecule as a function of its distance from a perturbation on the opposite side of the molecule. This distance is taken to be the number of methylene groups between the two. Perturbations are taken as simple functional groups, R = $-CH_3$, $-OH$, $-NH_2$, $-F$, $-H$.

3.1. Motivation for NLMOs. It is demonstrated in Figure 1 for C_1 water clusters that the NLMOs can localize the triples contribution to the CCSDT-3 wave function. These water clusters were chosen because they are examples of small systems which have C_1 symmetry; however, other C_1 molecules such as distorted alkanes, etc. could be used. It can be shown that the number of negligible amplitudes resulting from spatial locality is greater than the number due to symmetry. For a large symmetric molecule it is necessary that the molecule be sufficiently large so as to pass the crossover point for which spatial locality, as opposed to symmetry, dominates the number of negligible amplitudes. In the canonical basis ≈90% of the $|T_2|$ amplitudes on average are above the 5.0×10^{-8} threshold whereas for the $|T_3|$ amplitudes this number reduces to ≈50%. These averages ignore the contribution from a single water molecule which because of the C_{2v} symmetry gives a crossover in the number of amplitudes above the threshold for canonical versus local bases. It is not surprising that the size of $|T_3|$ is smaller than $|T_2|$ because these clusters have relatively simple, nondegenerate ground states thereby lacking nondynamical correlation effects usually associated with higher excitation operators. In the localized NLMO basis both $|T_2|$ and $|T_3|$ are sparse, giving, for example, only ≈35% and ≈10% amplitudes above threshold, respectively, for the water heptamer. The decay of localized amplitudes is monotonic for $|T_3|$ and nearly monotonic for $|T_2|$; however, due to the lack of symmetry in

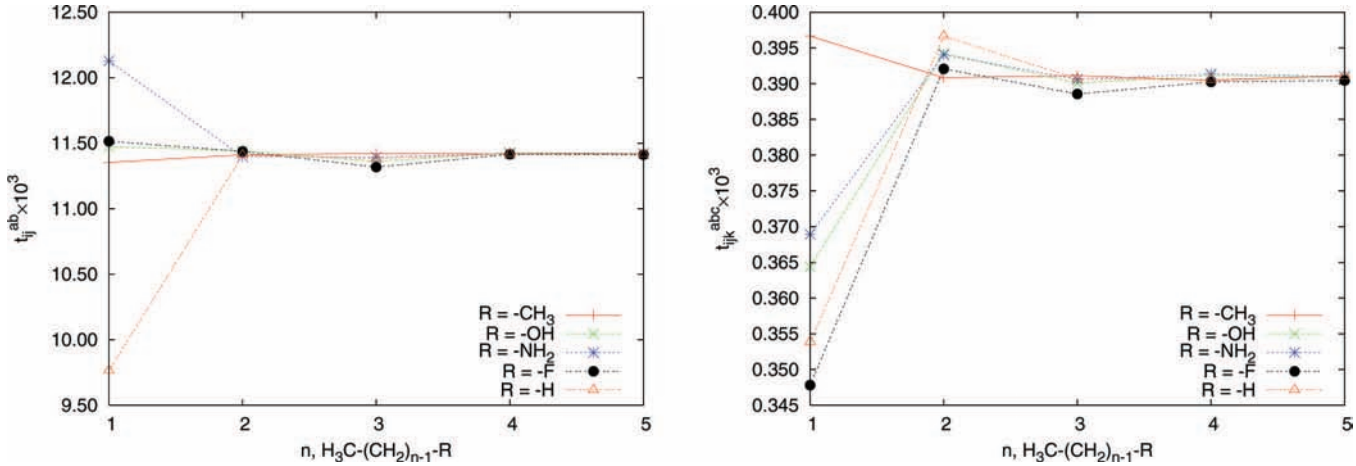

Figure 4. Transferability of $|T_2|$ and $|T_3|$ from CCSDT-3 in the NLMO basis for substituted alkanes. The amplitudes are confined to the methyl cap on the opposite side of the substituent and the *x*-axis represents increasing distance between the two in terms of the number of methylene groups.

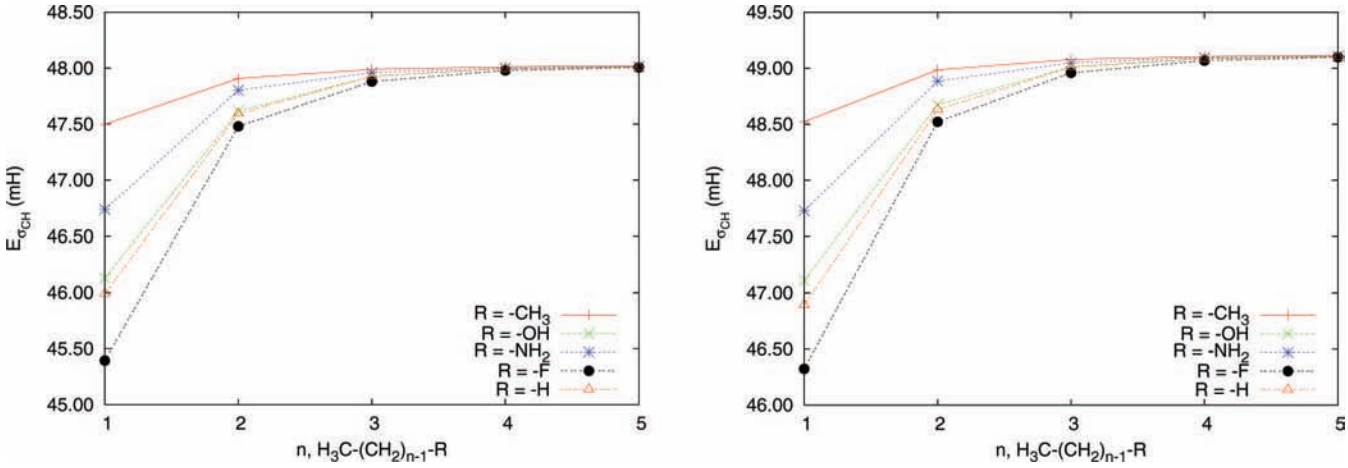

Figure 5. Transferability of CCSD and CCSDT-3 absolute value bond energies in the NLMO basis for substituted alkanes. The σ_{CH}-bond is on the opposite side of the substituent, and the x-axis represents increasing distance between the two in terms of the number of methylene groups.

the clusters it is not necessarily expected that the behavior be monotonic. By virtue of the localization scale, we can get a handle on the dominant amplitudes.

Figure 2 examines the locality of the virtual NLMOs, specifically the absolute value of diagonal two-electron integrals, $\langle iilaa \rangle$, for a $(H_2O)_5$ calculation. The profile of the integrals with respect to different virtual bases is shown where the key is understood as follows with the number giving the degree of localization: local 0 (canonical occupied and canonical virtual), local 1 (localized occupied and canonical virtual), and local 2 (localized occupied and localized virtual). For the localized occupied calculations $i = \sigma_{OH}$ and for the canonical occupied calculations i is delocalized. The local 0 curve shows that all the virtual orbitals give significant interactions with the chosen occupied. The local 1 curve shows that localizing the occupied orbitals does little to improve on local 0, for example a given σ_{OH} has significant interactions with all of the canonical virtual space, supporting the well-known conclusion that a localized virtual space is imperative. The local 2 curve shows significant improvements on local 0 and local 1 because both the occupied and virtual spaces are localized by virtue of our NLMO implementation where the peaks represent intramolecule interactions. The first set of peaks correspond to $\sigma_{OH}-\sigma^*_{OH}$ pair interactions, and the second and third correspond to interactions with the Rydberg spaces of oxygen and hydrogen, respectively.

The transferability of the CC wave function in terms of NLMOs is shown in Figures 3 and 4 for a series of alkanes and substituted alkanes, respectively. Figure 3 shows diagonal $|T_2|$ (CCSD) and $|C_2|$ (CISD) amplitudes for a $\sigma_{CH}-\sigma^*_{CH}$ pair confined to a methyl group as a function of alkane size. It is seen that because the CC wave function is size-extensive the CC amplitudes are transferable but the CI amplitudes are not. In Figure 4 we show the transferability of $|T_2|$ and $|T_3|$ for the CCSDT-3 level of theory. The amplitudes are confined to be on the methyl cap on the opposite side of the substituent, for example, possibly incorporating another σ-bond different from the first also on the methyl cap. The first point in Figure 4 represents the amplitude and substituent as nearest neighbors and thus the amplitudes are very different among substituents. T_2 requires only one region of screening and therefore rapidly approaches a constant transferable value, whereas T_3 from a CCSDT-3 calculation requires two or more regions of screening before becoming transferable. Note that these results are similar over all functional groups studied, including the very electronegative fluorine atom, because of the robust transferability of NLMOs.

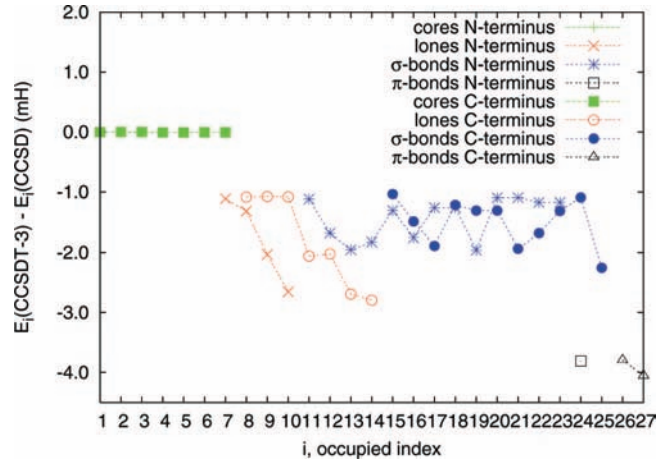

Figure 6. Difference between CCSD and CCSDT-3 bond energies for all of the occupied orbitals in two molecules used in the NLSCC calculation of polyglycine. The two molecules are designated by the fact that they are either the N-terminus or C-terminus and are shown in Figure 7 on the lower-left and lower-right, respectively.

The CCSD (left) and CCSDT-3 (right) σ_{CH}-bond energies computed from these transferable amplitudes are themselves transferable, as shown in Figure 5 for the same series of substituted alkanes. The difference between CCSD and CCSDT-3 bond energies is, not surprisingly, small, ≈1 mH, given the simplicity of the electronic structure. Both CCSD and CCSDT-3 bond energies require two or more regions of screening before reaching a constant value and the rate at which they become transferable are nearly identical.

3.2. Polyglycine. Differences between CCSD and CCSDT-3 bond energies are shown in Figure 6 for two molecules used in the NLSCC calculation of polyglycine. The two molecules are designated by the fact that they are either the N-terminus or C-terminus and are shown in Figure 7 on the lower-left and lower-right, respectively. There are negligible triples contributions from the core orbitals and small ≈1.5 mH on average contributions for the σ-bonds. The lone pairs in these two molecules have larger triples contributions ≈2 mH, on average, and they are over a longer range than the other bonds, ≈1−3 mH. This is probably because they are more dependent on local environment due to their greater diffusivity. The occupied indicies within each of the core, lone-pair, σ-bond, and π-bond sections are not in a specific order. The triples contribution from the π-bonds is largest, ≈4 mH, suggesting an interesting active

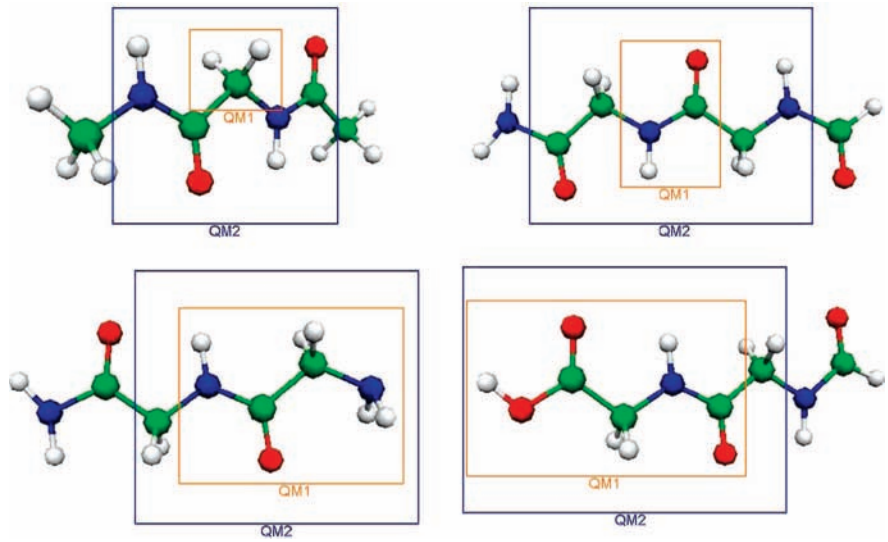

Figure 7. Definition of the various regions for the hybrid NLSCC calculation of polyglycine. QM1 simply represents the functional groups of interest, QM2 represents the CCSDT-3 region, and QM3, the whole fragment, represents the CCSD region. Due to the more complicated structure there are four calculations (methylene group, peptide group, N-terminus group, and C-terminus group) needed.

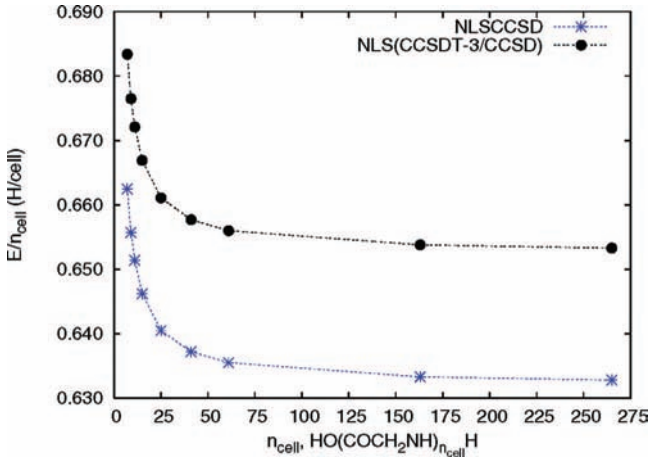

Figure 8. NLSCCSD and NLS(CCSDT-3/CCSD) absolute value correlation energy per unit cell of polyglycine. The unit cell is glycine.

TABLE 1: Transferability of CCSD and CCSDT-3 Functional Group Energies among Molecules Used in the NLSCC Calculations on Polyglycine[a]

	C-terminus	N-terminus	peptide	methylene
		CCSD		
C-terminus	−1.1039		−0.4184	−0.2129
N-terminus		−0.7914	−0.4171	−0.2123
peptide			−0.4173	
methylene				−0.2146
		CCSDT-3		
C-terminus	−1.1388		−0.4331	−0.2193
N-terminus		−0.8150	−0.4308	−0.2185
peptide			−0.4311	
methylene				−0.2212

[a] The diagonal elements are simply the energies taken from corresponding QM1 regions of Figure 7, whereas the upper-diagonal elements for peptide and methylene are the energies of those groups extracted from both the N-terminus and C-terminus calculations. Functional group energies are reported in units of Hartree.

space triples method based on NLSCC in which different bonds are correlated at different levels of theory.

Figure 7 shows the QM regions defining the hybrid NLSCC calculation of polyglycine. The functional groups of interest represented by the QM1 region, the "QM" label is to specify that this region is treated quantum mechanically as opposed to by more classical interactions, for example, by an electrostatic potential, etc., that could be incorporated into regions outside of QM3 but will not in the current formulation. There is one region for the methylene group, one for the peptide group, one for the N-terminus, and one for the C-terminus. This region is embedded in the QM2 region, which will be treated at the CCSDT-3 level of theory, which in turn is embedded in the QM3 region to be treated at the CCSD level of theory. The reason for two embedding regions is to capture any strong correlations for those functional groups that are close to one another with CCSDT-3 and for weaker correlations between more distant functional groups CCSD can be used. The peptide bond in polyglycine is treated as a Kekulé structure composed of two-center bonds.

The CCSD and CCSDT-3 functional group energies of the molecules used in the NLSCC calculations on polyglycine are shown in Table 1. The diagonal elements are simply the energies taken from the corresponding QM1 regions of Figure 7 whereas the upper-diagonal elements for peptide and methylene are the energies of those groups extracted from both the calculations for which the target was the N-terminus and C-terminus. The difference between CCSD and CCSDT-3 functional group energies is largest for the C-terminus, ≈35 mH, followed by the N-terminus and the average peptide group which are ≈24 and ≈14 mH, respectively. The smallest difference is for the average methylene group which is ≈6 mH. This ordering follows from the total number of lone pairs and π-bonds in each of the groups, which we know from Figure 6 give large triples corrections to the energy. The transferability of the peptide group among three different molecules is seen by considering the upper-diagonal elements which are within ≈1 mH of one another for both CCSD and CCSDT-3. The same is true for the methylene group from the CCSD calculation; however, for the CCSDT-3 calculation the differences are slightly larger.

The total and unit cell NLSCCSD and NLS(CCSDT-3/CCSD) correlation energies for polyglycine are shown in Table 2 along with some literature results obtained using LCC methods.[14,32] The unit cell is glycine. The triples contribution acts to lower the correlation energy contribution per unit cell by ≈20 mH for all polyglycine examples including the infinite limit. The

TABLE 2: Total and Unit Cell NLSCCSD and NLS(CCSDT-3/CCSD) Correlation Energies for Polyglycine[a]

$n_{glycine}$	NLSCCSD		LCCSD/LMP2 FC	NLS(CCSDT-3/CCSD)		LCCSDT-1b/LMP2 FC	LCCSD(T)/LMP2 FC
n_{cell}	E_{corr} (H)	E_{corr}/n_{cell}	E_{corr} (H)	E_{corr} (H)	E_{corr}/n_{cell}	E_{corr} (H)	E_{corr} (H)
1	99.4%	−0.8466	99.1%	99.4%	−0.8703	98.9%	99.0%
2	99.9%	−0.7387	99.0%	−1.5212	−0.7606		
4	99.3%	−0.6854	−2.6783	−2.8262	−0.7066	−2.7552	−2.7570
7	−4.6376	−0.6625		−4.7838	−0.6834		
8	−5.2696	−0.6587	−5.1499	−5.4363	−0.6795	−5.3001	−5.3039
9	−5.9017	−0.6557		−6.0888	−0.6765		
10	−6.5338	−0.6534	−6.3857	−6.7413	−0.6741		
11	−7.1658	−0.6514		−7.3938	−0.6721		
14	−9.0620	−0.6473	−8.8574	−9.3513	−0.6680		
15	−9.6941	−0.6462		−10.0038	−0.6669		
25	−16.0147	−0.6405		−16.5289	−0.6611		
41	−26.1277	−0.6372		−26.9691	−0.6577		
61	−38.7690	−0.6355		−40.0193	−0.6560		
163	−103.2395	−0.6333		−106.5753	−0.6538		
265	−167.7100	−0.6328		−173.1313	−0.6533		
500	−316.2450	−0.6324		−326.4711	−0.6529		
1000	−632.2769	−0.6322		−652.7261	−0.6527		
∞		−0.6321			−0.6525		

[a] The unit cell is a glycine. Other values are taken from the LCC literature.[14,32] The symbol H means that the energies are reported in units of Hartree.

TABLE 3: Decomposition of the CCSD correlation Energy into Functional Group and NLMO Contributions for Non-glycine QM1 Regions Shown in Figure 10[a]

methionine		phenylalanine		tyrosine	
S5	−0.0000	C5	−0.0017	O2	−0.0014
S5	−0.0013	C8	−0.0017	C3	−0.0016
S5	−0.0013	C10	−0.0017	C6	−0.0017
S5	−0.0016	C12	−0.0017	C12	−0.0017
S5	−0.0017	C14	−0.0018	C4	−0.0018
C7	−0.0018	C6	−0.0018	C8	−0.0018
C2	−0.0019	C2	−0.0019	C10	−0.0018
C9	−0.0020	C10H11	−0.0385	C13	−0.0019
S5	−0.0409	C12H13	−0.0386	O2	−0.0450
S5	−0.0480	C8H9	−0.0389	O2	−0.0548
C2H1	−0.0473	C14H15	−0.0393	C6H7	−0.0391
C2H4	−0.0474	C6H7	−0.0397	C4H5	−0.0391
C2H3	−0.0479	C2H3	−0.0401	C8H9	−0.0392
C7H6	−0.0492	C2H4	−0.0404	C10H11	−0.0393
C9H10	−0.0493	C10C12	−0.0431	C13H14	−0.0404
C9H11	−0.0495	C8C10	−0.0431	C13H15	−0.0404
C7H8	−0.0502	C12C14	−0.0434	C3C4	−0.0431
C2S5	−0.0539	C6C8	−0.0437	C3C8	−0.0431
S5C7	−0.0548	C5C14	−0.0441	C4C6	−0.0435
C9C12	−0.0553	C5C6	−0.0443	C8C10	−0.0437
C7C9	−0.0567	C2C5	−0.0473	C6C12	−0.0439
		C1C2	−0.0477	C10C12	−0.0442
		$1\pi_6$	−0.0536	C12C13	−0.0471
		$2\pi_6$	−0.0699	C13C16	−0.0476
		$3\pi_6$	−0.0699	O2H1	−0.0505
				O2C3	−0.0547
				$1\pi_6$	−0.0537
				$2\pi_6$	−0.0675
				$3\pi_6$	−0.0714
Total	−0.6619		−0.8380		−1.0049

[a] NLMOs are ordered per column as core, lp, σ, and π orbitals as well as a six-center phenolic-bond given by $1\pi_6-3\pi_6$. For each type of NLMO within each column the energy contributions are ordered by increasing magnitude. Energies are reported in units of Hartree.

saturation of the correlation energy per unit cell is clearer in Figure 8. The energy difference for the infinite versus small molecule limit for the glycine unit cell is ≈30 mH for both NLSCCSD and NLS(CCSDT-3/CCSD). For NLSCCSD over 99% of the correlation energy is obtained when compared to conventional results for the glycine monomer, dimer, and tetramer, and for NLS(CCSDT-3/CCSD), the comparison with conventional methods was only possible with the glycine monomer for which 99.4% was recovered. To validate the NLSCC triples methods, further comparison with conventional methods would be necessary. Despite the fact that a direct comparison between the NLSCC and LCC methods is not possible, because of the frozen core approximation in addition to the fact they are hybridized differently, the NLSCC method provides the literature results for extended systems. As an example of time savings, consider $(gly)_1$ with 95 basis functions and $(gly)_{16}$ with 1160 basis functions. The time required for a CCSDT-3 calculation on $(gly)_1$ is 0.95 days, which can be crudely extrapolated to 3.4×10^7 days for $(gly)_{16}$. The NLS(CCSDT-3/CCSD) serial calculation on $(gly)_{16}$ took 11.9 days to be compared to a parallel fragment molecular orbital (FMO) CCSD(T) on four 3.2 GHz pentium4 processors which took 9.9 days.[30] By virtue of the fact that the NLSCC methods are built upon local QM regions, the rate determining step of the calculations becomes that of the largest QM region. For a simple translationally periodic example once the energies of the QM1 regions shown in Figure 7 are determined it is trivial to reuse these regions in symmetrically equivalent cases to determine the energies of larger and larger molecules. More timings are discussed in the section on met-enkephalin.

3.3. Met-Enkephalin. The three-dimensional pentapeptide (met-phe-gly-gly-tyr) met-enkephalin is shown in Figure 9 along with a crude representation of the QM1 regions chosen for the NLSCC calculations. Side chains of non-glycine residues methionine ($QM1_1$), phenylalanine ($QM1_4$), and tyrosine ($QM1_9$) are shown in Figure 10 along with atom indicies for use in Table 3.

Table 3 shows a decomposition of the CCSD correlation energy into functional group and NLMO contributions for non-glycine QM1 regions which are shown in Figure 10. The NLMOs are ordered per column as core, lp, σ, and π orbitals as well as a six-center phenolic-bond given by $1\pi_6-3\pi_6$. All NLMO core contributions are ≈1−2 mH and the average correlation energy for an oxygen lp (≈50 mH) is ≈5 mH greater than a sulfur lp (≈45 mH). The average correlation energy for a σ_{CH}-bond in methionine, ≈49 mH, is much greater than for phenylalanine and tyrosine which are both around ≈39 mH because of the similarity in their side chains.

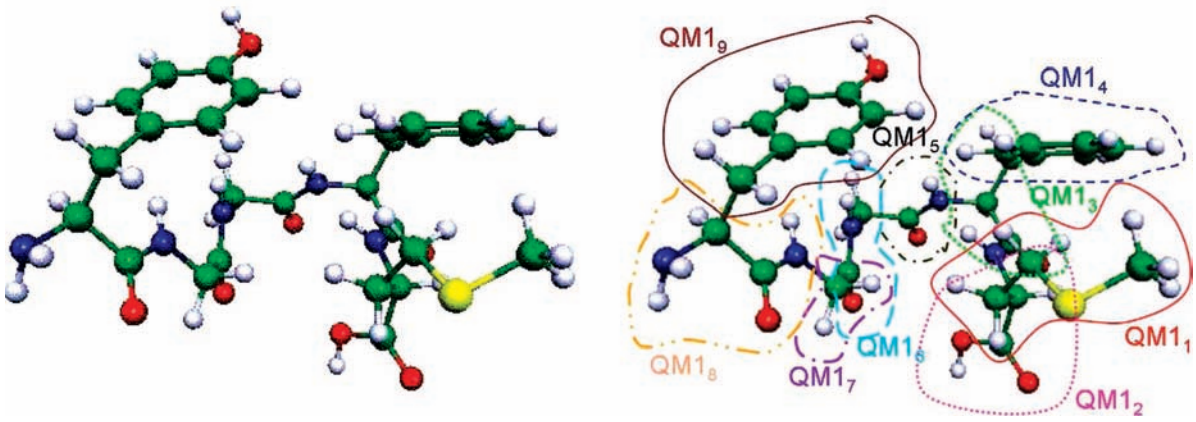

Figure 9. Ball and stick model of the three-dimensional penta-peptide (met-phe-gly gly-tyr) met-enkephalin (1PLW.pdb) taken from ref 38. Also shown is a crude representation of the QM1 regions chosen for the NLSCC calculations.

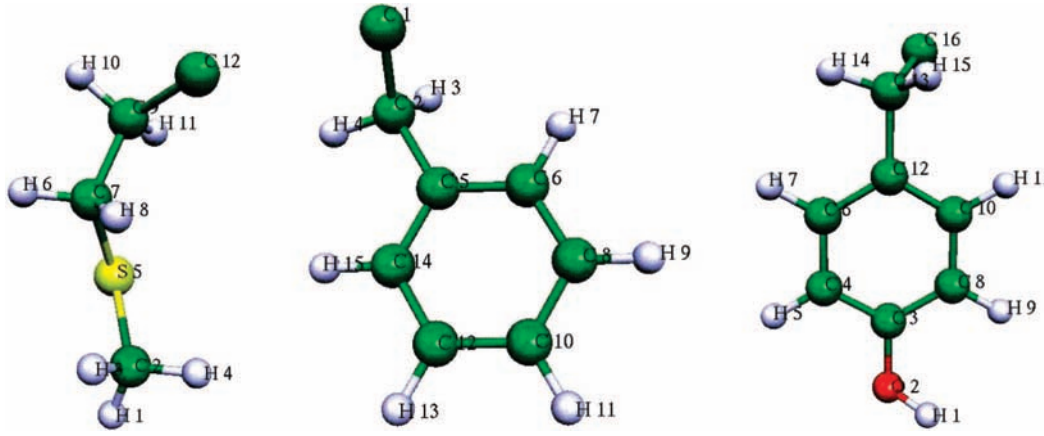

Figure 10. Ball and stick model of side chains of non-glycine residues in met-enkephalin shown with atom indicies for use in Table 3.

TABLE 4: Summary of STO-3G and cc-pVDZ NLSCCSD Calculations on Met-Enkephalin[a]

	QM2			QM1				E^i		
i, QM1	n_{atom}	n_{occ}	n_{vir}^{STO-3G}	$n_{vir}^{cc-pVDZ}$	n_{atom}	n_{occ}	n_{vir}^{STO-3G}	$n_{vir}^{cc-pVDZ}$	STO-3G	cc-pVDZ
1	27	46	29	192	11	21			−0.2486	−0.6619
2	30	55	31	221	6	15			−0.2387	−0.6519
3	29	55	34	225	6	14			−0.2244	−0.5931
4	35	65	42	272	14	25			−0.4927	−0.8380
5	23	46	25	177	4	11			−0.1758	−0.4786
6	29	56	33	224	7	15			−0.2408	−0.6348
7	22	41	25	168	3	4			−0.0644	−0.1561
8	28	52	32	214	9	18			−0.2736	−0.7549
9	35	65	42	272	15	29			−0.5315	−1.0049
NLSCCSD					75	152			−2.4905	−5.7742
CCSD					75	152	87	587	−2.4987	
HF					75	152	87	587	−2208.9502	−2236.7860
$E_{NLSCCSD}$					75	152			−2211.4407	−2242.5602
E_{CCSD}					75	152	87	587	−2211.4489	

[a] The number of atoms, occupied orbitals, virtual orbitals, and group correlation energies for each QM1 and accompanying QM2 regions according to Figure 9 are shown. The NLSCCSD results are shown to be a sum of the QM1 results and compared to conventional CCSD for the STO-3G basis. Energies are reported in units of Hartree. In this case conventional HF provides reference energies although other reference functions could be used. $E_{NLSCCSD}$ is the total energy, HF+NLSCCSD, within the NLS framework and E_{CCSD} is the total energy, HF+CCSD, using conventional CC.

The average correlation energy for a σ_{CC}-bond in methionine is ≈56 mH compared to ≈44 mH for a ring σ_{CC} and ≈47 mH for a nonring σ_{CC} in phenylalanine and tyrosine. The average σ_{CS}-bond in methionine has ≈54 mH in comparison to the average σ_{CC}-bond which is ≈56 mH. The σ_{OH}-bond in tyrosine has ≈53 mH. The delocalized six-centered π-bonds contain the most correlation amounting to an average of ≈64 mH for both phenylalanine and tyrosine. The similarity in the bond correlation energies for phenylalanine and tyrosine are attributed to the transferability of their electronic structure within NLSCC as represented in terms of NLMOs.

Table 4 summarizes minimal-basis and cc-pVDZ NLSCCSD calculations for met-enkephalin. On average each QM2 region

contains 29 atoms and 53 occupied orbitals with 32 virtuals for STO-3G and 218 virtuals for cc-pVDZ. The largest QM calculations with 65 occupied and 272 virtuals are for $QM2_4$ and $QM2_9$ for which we are calculating the phenylalanine−tyrosine interaction. The NLSCCSD result can be constructed out of effective functional groups as determined from the QM1 regions giving a correlation energy of −2.4905 and −5.7742 H for STO-3G and cc-pVDZ bases, respectively. For STO-3G conventional CCSD gives −2.4987 H, and thus we are able to recover 99.7% of the correlation energy. Although this is the simple minimal basis, we are able to get the result for a calculation with 87 virtual orbitals from calculations with no more than 42 virtuals thus saving a factor of 18 in the evaluation of $\langle ab|cd\rangle$ integrals. A similar comparison using the cc-pVDZ basis, 587 virtuals, is currently under development using a new parallel version of ACES, ACES III,[4] where we should expect similar savings given that the largest QM2 has only 272 virtuals. As a simple illustration of the time saved in the NLSCC method, consider the STO-3G conventional CCSD calculation which took a total of 53.56 h in comparison to the NLSCCSD calculation, which took 0.98 h for the rate limiting step. The other steps are shorter and can be done simultaneously. The cc-pVDZ NLSCCSD calculation on met-enkephalin had a rate limiting step of 137.64 h in which the NLMO generation took only 69 s. Conventional HF provides reference energies although other reference functions for example density functional theory (DFT) or FMO-HF[39] could be used. Collective three-body weak interactions from methionine, phenylalanine, and tyrosine were not necessary in the STO-3G basis but may be needed in larger and more diffuse basis sets. Accurate triples contributions for met-enkephalin are too intensive to calculate because the QM2 regions should contain at least amino acid dimers and these are already too expensive for a proper correlated treatment. Active space triples methods for aromatic bonds built upon conventional NLSCC would be quite useful in this situation.

4. Conclusion

By virtue of the transferability of NLMOs, we have developed and validated a linear-scaling method for triples contributions to the CC wave function in the context of NLSCC methods. Storage of the noncanonical triples amplitudes are avoided by applying the unitary localization matrix to the final canonical CC wave function instead of to the reference function. We reach the conclusion that the triples amplitudes are longer-ranged than the doubles amplitudes; however, our results show that choosing the triples regions in an analogous manner as the doubles is sufficient for our purposes. Calculations on translationally periodic polyglycine are among the largest CC triples calculations performed. For polyglycine we find that two-center Kekulé structures are sufficient for transferable correlated energies. However, other n-centered ($n > 2$) choices might be necessary in reproducing correlated density matrices or other properties. We find that the triples contributions are largest for π-bonds, indicating an interesting active space method in which different types of bonds are treated at different excitation levels. Applications to met-enkephalin test the transferability for a more realistic three-dimensional example, for which there has been limited linear-scaling results due to the high demands of accurately representing weak interactions. NLSCC methods are in agreement with literature values and can recover over 99% of the conventional CC energy with the rate limiting step being the size of QM regions chosen.

Acknowledgment. This work has been supported by the Air Force Office of Scientific Research under contract number FA-9550-04-1-01. T.F.H. gives a special thanks to S. A. Perera for his expertise with the ACES II quantum chemistry software package.

References and Notes

(1) Bartlett, R. J. *Annu. Rev. Phys. Chem.* **1981**, *32*, 359.
(2) Bartlett, R. J.; Musiał, M. *Rev. Mod. Phys.* **2007**, *79*, 291.
(3) Flocke, N.; Bartlett, R. J. *J. Chem. Phys.* **2004**, *121*, 10935.
(4) Lotrich, V.; Flocke, N.; Ponton, M.; Yau, A. D.; Perera, S. A.; Deumens, E.; Bartlett, R. J. Manuscript in preparation, 2008.
(5) Janowski, T.; Ford, A. R.; Pulay, P. *J. Chem. Theory Comput.* **2007**, *3*, 1368.
(6) Kinoshita, T.; Hino, O.; Bartlett, R. J. *J. Chem. Phys.* **2003**, *119*, 7756.
(7) Pedersen, T. B.; de Meras, A. M. J. S.; Koch, H. *J. Chem. Phys.* **2004**, *120*, 8887.
(8) Taube, A. G.; Bartlett, R. J. *Collect. Czech. Chem. Commun.* **2005**, *70*, 837.
(9) Manby, F. R.; Werner, H.-J.; Adler, T. B.; May, A. J. *J. Chem. Phys.* **2006**, *124*, 094103.
(10) Hampel, C.; Werner, H.-J. *J. Chem. Phys.* **1996**, *104*, 6286.
(11) Reed, A. E.; Weinhold, F. *J. Chem. Phys.* **1985**, *83*, 1736.
(12) Flocke, N.; Bartlett, R. J. *Chem. Phys. Lett.* **2003**, *367*, 80.
(13) Eskandari, K.; Mandado, M.; Mosquera, R. A. *Chem. Phys. Lett.* **2007**, *437*, 1.
(14) Schütz, M.; Werner, H.-J. *J. Chem. Phys.* **2001**, *114*, 661.
(15) Scuseria, G. E.; Ayala, P. Y. *J. Chem. Phys.* **1999**, *111*, 8330.
(16) Li, S.; Shen, J.; Li, W.; Jiang, Y. *J. Chem. Phys.* **2006**, *125*, 074109.
(17) Laidig, W. D.; Purvis, G. D.; Bartlett, R. J. *J. Phys. Chem.* **1985**, *89*, 2161.
(18) Paulus, B.; Rościszewski, K.; Stoll, H.; Birkenheuer, U. *Phys. Chem. Chem. Phys.* **2003**, *5*, 5523.
(19) Flocke, N.; Bartlett, R. J. *J. Chem. Phys.* **2003**, *118*, 5326.
(20) Li, S.; Ma, J.; Jiang, Y. *J. Comput. Chem.* **2002**, *23*, 237.
(21) Li, W.; Li, S. *J. Chem. Phys.* **2004**, *121*, 6649.
(22) Auer, A. A.; Nooijen, M. *J. Chem. Phys.* **2006**, *125*, 024104.
(23) Hirata, S.; Grabowski, I.; Tobita, M.; Bartlett, R. J. *Chem. Phys. Lett.* **2001**, *345*, 475.
(24) Hirata, S.; Podeszwa, R.; Tobita, M.; Bartlett, R. J. *J. Chem. Phys.* **2004**, *120*, 2581.
(25) Stollhoff, G.; Fulde, P. *J. Chem. Phys.* **1980**, *73*, 4548.
(26) Stoll, H. *Phys. Rev. B* **1992**, *46*, 6700.
(27) Stoll, H.; Paulus, B.; Fulde, P. *J. Chem. Phys.* **2005**, *123*, 144108.
(28) Maslen, P. E.; Dutoi, A. D.; Lee, M. S.; Shao, Y.; Head-Gordon, M. *Mol. Phys.* **2005**, *103*, 425.
(29) Maslen, P. E.; Lee, M. S.; Head-Gordon, M. *Chem. Phys. Lett.* **2000**, *319*, 205.
(30) Fedorov, D. G.; Kitaura, K. *J. Chem. Phys.* **2005**, *123*, 134103.
(31) Schütz, M. *J. Chem. Phys.* **2000**, *113*, 9986.
(32) Schütz, M. *J. Chem. Phys.* **2002**, *116*, 8772.
(33) Schütz, M.; Werner, H.-J. *Chem. Phys. Lett.* **2000**, *318*, 370.
(34) Raghavachari, K.; Trucks, G.; Pople, J.; Head-Gordon, M. *Chem. Phys. Lett.* **1989**, *157*, 479.
(35) Watts, J.; Gauss, J.; Bartlett, R. J. *J. Chem. Phys.* **1993**, *98*, 8718.
(36) Schaftenaar, G.; Noordik, J. H. *J. Comput.-Aided Mol. Design* **2000**, *14*, 123.
(37) Stanton, J. F.; Gauss, J.; Watts, J. D.; Nooijen, M.; Oliphant, N.; Perera, S. A.; Szalay, P. G.; Lauderdale, W. J.; Kucharski, S. A.; Gwaltney, S. R.; Beck, S.; Balková, A.; Bernholdt, D. E.; Baeck, K. K.; Rozyczko, P.; Sekino, H.; Hober, C.; Bartlett R. J. ACES II is a program product of the Quantum Theory Project, University of Florida. Integral packages included are VMOL (J. Almlöf and P. R. Taylor); VPROPS (P. Taylor); ABACUS (T. Aa. Helgaker, H. J. Jensen, P. Jørgensen, J. Olsen and P. R. Taylor).
(38) Marcotte, I.; Separovic, F.; Auger, M.; Gagné, S. M. *Biophys. J.* **2004**, *86*, 1587.
(39) Nakano, T.; Kaminuma, T.; Sato, T.; Akiyama, Y.; Uebayasi, M.; Kitaura, K. *Chem. Phys. Lett.* **2000**, *318*, 614.
(40) Abdali, S.; Niehaus, T. A.; Jalkanen, K. J.; Cao, X.; Nafie, L. A.; Frauenheim, T.; Suhai, S.; Bohr, H. *Phys. Chem. Chem. Phys.* **2003**, *5*, 1295.
(41) Fedorov, D. G.; Ishida, T.; Uebayasi, M.; Kitaura, K. *J. Phys. Chem. A* **2007**, *111*, 2722.
(42) Watson, T. M.; Hirst, J. D. *Phys. Chem. Chem. Phys.* **2004**, *6*, 2580.

JP800516Q