



HAL
open science

Local Correlation Methods with a Natural Localized Molecular Orbital Basis

Ricardo André Fernandes Mata, Hans-Joachim Werner

► **To cite this version:**

Ricardo André Fernandes Mata, Hans-Joachim Werner. Local Correlation Methods with a Natural Localized Molecular Orbital Basis. *Molecular Physics*, 2008, 105 (19-22), pp.2753-2761. 10.1080/00268970701628209 . hal-00513141

HAL Id: hal-00513141

<https://hal.science/hal-00513141>

Submitted on 1 Sep 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Local Correlation Methods with a Natural Localized Molecular Orbital Basis

Journal:	<i>Molecular Physics</i>
Manuscript ID:	TMPH-2007-0214.R1
Manuscript Type:	Full Paper
Date Submitted by the Author:	27-Jul-2007
Complete List of Authors:	Mata, Ricardo; Institute for Theoretical Chemistry Werner, Hans-Joachim; Institute for Theoretical Chemistry
Keywords:	NLMO, Local Correlation, MP2
<p>Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.</p> <p>paper.tex local.bib molpro.bib nlmo.bib tables.ods</p>	



Local Correlation Methods with a Natural Localized Molecular Orbital Basis*

Ricardo A. Mata and Hans-Joachim Werner[†]

Institut für Theoretische Chemie, Universität Stuttgart,
Pfaffenwaldring 55, D-70569 Stuttgart, Germany

August 9, 2007

Abstract

Natural localized molecular orbitals (NLMOs) have been used in local correlation calculations. For a test set of 30 molecules it is shown that the results obtained with NLMOs and Pipek-Mezey localization are very similar. However, NLMOs are much less sensitive to the basis set, in particular when diffuse functions are used. Based on natural population analysis (NPA) a new method to determine the domains in local correlation methods is proposed. It is demonstrated that this yields domains that are very insensitive to the choice of the basis set. For all 30 molecules the same domains are obtained with six different basis sets ranging from cc-pVDZ to aug-cc-pVQZ. This allows to define the local correlation methods more uniquely than with previous methods.

*This work is dedicated to Professor Peter Pulay

[†]Corresponding author. Email: werner@theochem.uni-stuttgart.de

1 Introduction

Local electron correlation methods have been a topic of great interest in the last few years. Based upon the approach first proposed by Pulay[1, 2, 3, 4, 5], local Møller-Plesset perturbation theory up to second order (LMP2)[6], local coupled cluster with single and double excitations (LCCSD)[7, 8] and local perturbative or iterative treatment of triple excitations (LCCSD(T))[9, 10, 11] have been developed in our group, with linear scaling of computational resources relative to molecular size. More recently, the efficiency of these methods has been further enhanced by density fitting approximations[12, 13, 14, 15, 16]. These developments made it possible to treat rather large systems at the LCCSD(T) level. For instance, a recent application to the prediction of activation energies of enzyme reactions can be found in Ref. [17]. The accuracy of local correlation methods for computing reaction energies[18, 19], dipole moments and dipole polarizabilities[20], molecular equilibrium structures[21, 22, 13], vibrational frequencies[23, 24, 25, 26] as well as intermolecular interactions[27, 28, 29, 30] has been extensively studied and demonstrated.

In all of these methods orthogonal localized molecular orbitals (LMO) are used as occupied space. The virtual space is built by projecting out the occupied LMOs from the atomic orbitals (AO). These projected atomic orbitals (PAOs) are orthogonal to the occupied space but not among each other. The low-order scaling with molecular size is based on the use of a truncated virtual space, by allowing only excitations from an LMO into PAOs in its vicinity. The subset of PAOs associated with an orbital pair is called a *domain*. Furthermore, orbital pairs are classified according to their importance, and only the strong pairs, which contribute typically more than 90%-95% to the correlation energy, are treated at the highest level, e.g., LCCSD(T). The selection of strong pairs is directly related to the domains. It depends either on the closest distance between PAOs in two orbital domains (distance criteria), or on the number of bonds between them (connectivity criteria). For details we refer to a recent review[18].

The choice of method for localizing the occupied space is usually of little importance. Boys localization[31] is one possibility, but Pipek-Mezey (PM) localization[32] is usually preferable since it keeps the π - σ separation in planar molecules[33]. Other localization schemes can be used as well, and it has been found that the correlation energy is rather insensitive to the localization method.

The main problem is to find a robust method for selecting the orbital domains automatically. Since each PAO originates from an AO (basis function), the domain choice is equivalent to choosing a subset of AOs for each orbital. Generally, those AOs should be included in a domain if they significantly contribute to the description of the corresponding LMO. To simplify this choice, all AOs at a given atom are treated as a block, and thus the domain choice reduces to the selection of a subset of atoms for each LMO. The simplest method would be to select these atoms on the basis of atomic gross charges, determined for instance by Mulliken or Löwdin population analysis. However, this turns out to be strongly basis set dependent. Boughton and Pulay[33] (BP) have suggested an improved method in which the atoms are initially also ranked according to atomic charges. Atoms are then added to the domain in the order of decreasing charges until the BP criterion

$$T^{(i)} = 1 - \min \left[\int (\phi_i - \tilde{\phi}_i)^2 d\mathbf{r} \right] \geq T_{\text{BP}} \quad (1)$$

is fulfilled, where $\tilde{\phi}_i$ is an approximate orbital represented by just the basis functions at the already selected atoms

$$|\tilde{\phi}_i\rangle = \sum_{A \in [i]} \sum_{\mu \in A} |\chi_\mu\rangle \tilde{L}_{\mu i}. \quad (2)$$

The coefficients $\tilde{L}_{\mu i}$ are determined by a simple least squares fitting procedure[33]. Unfortunately, this method still suffers from basis dependence. Usually, the criterion is more easily fulfilled for fewer centers if larger basis sets are used, and therefore for a fixed value of the threshold T_{BP} the domains grow with increasing basis set. To compensate for this, one can use different thresholds for different basis sets, e.g. 0.98 for double zeta, 0.985 for triple zeta, and 0.99 for quadruple zeta. But apart from the fact that this is not a well defined and user-friendly model, the domains often still differ for different basis sets.

Another problem sometimes arises when diffuse basis sets are used. Due to near linear dependencies, the Pipek-Mezey method, which is based on the AO overlap matrix, may yield artificially large LMO coefficients at some atoms. This frequently happens, for instance, in aromatic compounds with basis sets such as aug-cc-pVTZ or aug-cc-pVQZ. Using the BP method one then finds unphysical large domains for the C-C bonds or π -orbitals, which include the neighboring H-atoms.

1
2
3
4
5
6
7
8 In order to overcome these problems, one needs a more stable atomic
9 population analysis. The Natural Population Analysis (NPA), proposed by
10 Weinholdt and coworkers[34] may be a more reliable alternative. It has been
11 shown to converge with respect to basis set size, and also to deliver re-
12 sults which agree well with experimental evidence and/or chemical sense.
13 Also related to this analysis are the Natural Localized Molecular Orbitals
14 (NLMOs)[35], which can be efficiently calculated from the SCF density ma-
15 trix and have been shown to deliver localized orbitals similar to those of other
16 commonly used orbital localization schemes. We note that NLMOs have been
17 used in the context of local correlation before by Flocke and Bartlett[36], but
18 their treatment is quite different from the one used here.

19
20 In this work we investigate the effectiveness of NLMOs in local corre-
21 lation calculations and propose a new criterion for domain selection based
22 on NPA. The new method, denoted here as NLMO/NPA, will be compared
23 with the Pipek-Mezey/Boughton-Pulay (PM/BP) procedure used in our lo-
24 cal correlated methods so far. The new method will be shown to yield basis
25 set independent domains for a wide variety of cases and to be also robust
26 when diffuse basis sets are used. This makes it possible to establish well
27 defined computational models based on the local correlation methods.
28
29
30
31
32

33 2 Method

34 2.1 NLMO Transformation

35
36 The NLMO localization procedure, as proposed by Weinhold and cowor-
37 kers[34, 35, 37, 38], has been implemented in the Molpro package of ab initio
38 programs[39]. This section gives a brief overview of the procedure. The
39 following notation will be used to name different orbital types: i, j, k denote
40 occupied molecular orbitals (either NLMO, LMO or canonical orbitals); r, s, t
41 denote natural atomic orbitals (NAOs); μ, ν, σ denote AOs. Capital letters
42 will be used for atomic centers.
43
44
45
46

47 The NLMO orbitals are obtained through a series of transformations,
48 starting from the nonorthogonal AOs
49

$$50 \text{AO} \xrightarrow{\mathbf{T}^{\text{NAO}}} \text{NAO} \xrightarrow{\mathbf{T}^{\text{NBO}}} \text{NBO} \xrightarrow{\mathbf{T}^{\text{NLMO}}} \text{NLMO}. \quad (3)$$

51
52
53
54
55
56
57
58
59
60

The valence density for the closed-shell SCF wavefunction is computed as

$$D_{\mu\nu} = 2 \sum_i^{\text{val}} C_{\mu i}^* C_{\nu i}, \quad (4)$$

where the $C_{\mu i}$ are the valence canonical orbital coefficients. The core orbital coefficients are not included as they should not be mixed with valence orbitals (unless core correlation is explicitly included). The AO basis will be assumed to be atom centered, so that each atomic index μ belongs to a given center X . This will be denoted as $\mu \in \{X\}$.

The density matrix is transformed to the nonorthogonal basis, $\Gamma = \mathbf{S}\mathbf{D}\mathbf{S}$, and divided into one-center blocks

$$\Gamma = \begin{pmatrix} \Gamma^{(AA)} & \Gamma^{(AB)} & \Gamma^{(AC)} & \dots \\ \Gamma^{(BA)} & \Gamma^{(BB)} & \Gamma^{(BC)} & \dots \\ \Gamma^{(CA)} & \Gamma^{(CB)} & \Gamma^{(CC)} & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \quad (5)$$

where for $\Gamma_{\mu\nu}^{(AB)}$, $\mu \in \{A\}$ and $\nu \in \{B\}$. The AO overlap matrix \mathbf{S} is partitioned in the same way. The blocks are symmetry-averaged as detailed in Ref. [34]. For each diagonal sub-matrix the generalized eigenvalue problem

$$\Gamma^{(AA)} \mathbf{X} = \mathbf{S}^{(AA)} \mathbf{X} \mathbf{W}, \quad (6)$$

where \mathbf{W} is a diagonal matrix holding the eigenvalues, is solved. The eigenvectors \mathbf{X} are the pre-NAOs, and their occupancy is given by the respective eigenvalues \mathbf{W} . These orbitals are divided into two sets: the Natural Minimal Basis (NMB) and the Natural Rydberg Basis (NRB). The division is made by taking the first N orbitals into the NMB set, where N is the number of valence orbitals needed for the representative ground configuration state of the neutral atoms. The orbitals are then orthogonalized among each other, while maintaining the density matrix block diagonalization. The steps are as follows:

1. The NRB orbitals are Schmidt-orthogonalized relative to the NMB set,
2. the eigenvalue problem of Eq. (6) is again solved, this time for the density and overlap matrices in the NRB basis,
3. the NMB and NRB orbitals of the different centers are orthogonalized using an occupancy-weighted orthogonalization scheme (see Ref. [34]),

- the eigenvalue problem is again solved, this time in the basis of the whole orthogonal orbital set.

The diagonalization and orthogonalization matrices multiplied together give the \mathbf{T}^{NAO} matrix. Transformation of the AO density matrix into the NAO basis

$$\tilde{\mathbf{D}} = (\mathbf{T}^{\text{NAO}})^\dagger \mathbf{D} \mathbf{T}^{\text{NAO}}, \quad (7)$$

gives a block-diagonalized matrix, and its diagonal elements \tilde{D}_{rr} are the final occupation number for the NAO orbital with index r . In this way, one can divide the charge among the atoms as

$$P_A = \sum_{r \in \{A\}} \tilde{D}_{rr}. \quad (8)$$

This is the so called Natural Population Analysis (NPA)[34].

The natural bond orbitals (NBOs) are built by diagonalizing one and two-center blocks of the NAO density matrix. The procedure is as follows:

- All natural atomic orbitals with eigenvalues above the threshold `thr_occ` =1.90 are added to the NBO list as lone-pairs and all lone pair contributions to the density matrix are removed.
- The two-center blocks of the NAO density matrix are diagonalized. Again, all orbitals with eigenvalues above the threshold `thr_occ` =1.90 are added to the NBO list. These NBOs are referred to as 2-center bond orbitals.
- If the number of NBOs found in this way are equal to the number of electron pairs, the search is stopped and one may proceed to the next step. If not, `thr_occ` is decremented and step 2 is repeated. It would also be possible to expand the search to 3-center bonds, but we found that for the systems included in this study (and in general organic compounds) this was not necessary.
- The remaining orbital space (of low occupation) is divided into Rydberg and anti-bonding orbitals. More details can be found in Ref. [38].

The NBO orbitals are by construction orthogonal, but should not be used directly in post-SCF calculations since the occupied NBO orbitals do not span the SCF occupied space exactly. Therefore, a final \mathbf{T}^{NLMO} transformation is

performed which rotates the NBO orbitals so that the orbitals with highest occupations (the so called Lewis set) span the SCF valence space. This is done by 2x2 Jacobi rotations which zero the density matrix elements between the Lewis and non-Lewis spaces. For closed-shell SCF wave functions this makes all diagonal elements $D_{ii} = 2$ and all other elements zero. Since the diagonal elements of the NBO density matrix of the Lewis space are already quite close to 2 one only needs a limited number of 2x2 Jacobi rotations and the orbital space stays localized. The procedure is further detailed in Ref. [35]. In summary, the final NLMO coefficients are obtained as

$$\mathbf{L} = \mathbf{T}^{\text{NAO}}\mathbf{T}^{\text{NBO}}\mathbf{T}^{\text{NLMO}} = \mathbf{T}^{\text{NAO}}\mathbf{V}. \quad (9)$$

2.2 NLMO Charges

The main advantage of using NLMOs is the relation to the NPA analysis. Contrary to Löwdin or Mulliken population analysis, NPA should be stable for large basis sets. However, no unique way has been described so far to divide the charge of each NLMO into charges of centers. We propose the following definition

$$P_{Ai} = \sum_{r \in \{A\}} \left(\frac{V_{ri}^2}{\sum_j V_{rj}^2} \right) \tilde{D}_{rr}, \quad (10)$$

where P_{Ai} is the charge of NLMO orbital i in center A . The sum in the denominator only runs over occupied indices, since the virtual orbitals have zero occupancy. This ensures that the charge of each NAO is only shared between the occupied NLMOs.

3 Results

A set of 30 molecules was chosen for testing the method. They are depicted in Fig. 1. Included in this list are typical small organic molecules, medium-sized saturated and unsaturated hydrocarbons as well as aromatic systems. All calculations were carried out with the density-fitting variants of HF, MP2 and LMP2[12, 14]. The correlation consistent basis sets of Dunning and co-workers, cc-pVXZ [40] and aug-cc-pVXZ [41] (with $X=D,T$ and Q) were used.

The NLMO and Pipek-Mezey orbitals are qualitatively very similar. In this work we will restrict our discussion to the use of both procedures in

1
2
3
4
5
6
7
8 local correlation methods. In a first series of tests, NLMO and Pipek-Mezey
9 transformations were used to localize the valence occupied space, and LMP2
10 calculations were performed for both sets. In both cases the domains were
11 determined using the BP method with the threshold $T_{BP} = 0.985$. The cc-
12 pVTZ basis set was used. The results are shown in Fig. 2, where the ratio
13 of correlation energy relative to canonical MP2 (in percent, left scale) as
14 well as the difference of the average domain sizes $L_{NLMO} - L_{PM}$ (right scale) is
15 plotted. As can be seen in the diagram, this difference is mostly positive,
16 meaning that NLMO domains are slightly larger. The only exception is the
17 benzochinone molecule. The correlation energy recovered with both orbital
18 sets is very similar, and the differences are mainly due to the different domain
19 sizes. In the cases where the domains are the same, the energies are almost
20 identical, which supports the conclusion that the localization method has
21 very little effect on the energy. These results indicate that replacing Pipek-
22 Mezey localized orbitals by NLMOs should have little effect on the accuracy
23 of local correlation methods.

24
25
26
27
28 We now turn to the question of domain selection. Our proposal is to
29 replace the BP procedure by a criterion that is based on the charges obtained
30 using the NPA and Eq. (10). There are two approaches for such a criterion.
31 One may define a minimum charge for the LMO domain to be considered as
32 "filled". Centers would be added to the domain list in the order of decreasing
33 charges, until $\sum_{A \in [i]} P_{Ai}$ exceed the threshold. Another possibility is to add
34 all centers to a domain with charge above a given limit. Both approaches
35 lead to similar results, but we actually found the latter alternative to be more
36 stable with respect to basis set size, especially in aromatic systems. The new
37 parameter will be referred to as T_{NPA} . For a given LMO ϕ_i , all atoms for which
38 $P_{Ai} > T_{NPA}$ are added to the domain list [i].

39
40
41
42 Various values for T_{NPA} were tested in the LMP2 calculations for the test
43 set, with the cc-pVXZ and aug-cc-pVXZ ($X=D,T$ and Q) basis sets. We
44 found that with $T_{NPA} = 0.05$ approximately the same correlation energy per-
45 centage is recovered as with the BP method and the default value $T_{BP} = 0.98$.
46 Also, the domains of saturated molecules obtained with $T_{NPA} = 0.05$ corre-
47 spond to chemical intuition, i.e., they contain one atom for lone pairs and 2
48 atoms for bicentric covalent bonds. For conjugated and aromatic π systems
49 the size of the domains depends on the choice of the parameter T_{NPA} .

50
51 The variation of the domains with basis set is measured by a parameter
52 $\Delta = \sum_i \Delta_i$, where Δ_i is the number of non-coinciding atoms in the orbital
53 domain i , relative to the domains obtained with the cc-pVDZ basis set. For
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8 example, if for two different basis sets the domains of a particular orbital
9 are C_1, C_2, H_1 and $C_1, C_2, \Delta_i = 1$, while for C_1, C_2, H_1 and C_1, C_2, H_2
10 $\Delta_i = 2$. Fig. 3 shows the variation of Δ for the largest molecule in the
11 set, thianthrene. Two sets of calculations were carried out with the Pipek-
12 Mezey orbitals. In the first case, fixed parameters were used for the domain
13 selection ($T_{BP} = 0.985$) and localization for all basis sets. In a second series of
14 calculations, the parameters were changed as previously suggested. The BP
15 criterion was set to 0.980 for the double-zeta basis sets, 0.985 for triple-zeta
16 and 0.990 for quadruple-zeta. In the localization with augmented basis sets,
17 the contribution of the most diffuse basis function of each angular momentum
18 type for each atom was eliminated in the localization criterion. In the case
19 of aug-cc-pVQZ, the last two functions were eliminated (this was achieved
20 by zeroing the corresponding rows and columns of the overlap matrix used
21 in the PM localization procedure). These values are referred to as variable
22 T_{BP} in Fig. 3. This procedure should decrease the basis set dependence. For
23 the NLMO/NPA method, a single value was used $T_{NPA} = 0.05$.

24
25
26
27
28 The results in Fig. 3 show large fluctuations in the domains when using a
29 fixed T_{BP} value. The use of different parameters for different basis sets helps
30 to decrease these differences, but the changes in the domain lists when using
31 diffuse functions is still sizable. In contrast, in the case of NLMO/NPA the
32 value of Δ is zero for all six basis sets, i.e., not a single domain change.

33
34 Similar tests were done for all 30 molecules using T_{NPA} values of 0.025,
35 0.05, and 0.10. The PM/BP domains change significantly as a function of
36 the basis set. This happens not only in the case of aromatic rings, but also in
37 smaller molecules like dimethylether or oxirane. The use of diffuse functions
38 generally leads to worse localization and a steep increase in the domain sizes.

39
40 Contrary to the BP domains, the NPA-based criterion is extremely ro-
41 bust. For the 30 molecules depicted in Fig. 1 and using $T_{NPA} = 0.05$, all
42 domains were the same for all basis sets. For $T_{NPA} = 0.025$, there is one
43 difference in the domains for oxalic acid between the double-zeta and the re-
44 maining basis sets. One of the carboxylic π -orbitals extends to a neighboring
45 carbon atom for the larger basis sets. Since there are two carboxylic groups,
46 this leads to $\Delta = 2$. The population at the neighboring carbon changed from
47 0.023 to 0.026 a.u. for cc-pVDZ and cc-pVTZ, respectively, and this happens
48 to pass the threshold. For $T_{NPA} = 0.1$, there is also only a single change. For
49 the aug-cc-pVTZ basis set, one of the oxygen lone pairs in glycine turns into
50 a double bond, again due to population fluctuations of the order of ± 0.001 .
51 But even with these exceptions, there is an enormous gain of stability by
52
53
54
55
56
57
58
59
60

1
2
3
4
5
6
7
8 using the NPA criterion.

9 Also of interest is the fact that for domains computed with $T_{\text{NPA}} = 0.05$
10 only six orbitals in thianthrene have more than 2 atoms in their domains.
11 This is in agreement with the number of expected π -orbitals for the molecule.
12 In Fig. 4, the charges P_{Ai} for the π -orbitals of three molecules are depicted.
13 The aug-cc-pVQZ was used, but the choice of basis set is irrelevant since the
14 charges do not change significantly. It can be seen that the charge distri-
15 bution in benzene is completely symmetric. This leads to three equivalent
16 domains for the π -orbitals. With $T_{\text{NPA}} = 0.1$ each of the three π -domains
17 will include 4 atoms with charges 0.112, 0.827, 0.827, 0.112. If the criterion
18 is lowered to 0.05, the remaining two atoms with charge 0.06 are added to
19 each domain. Thus, in the latter case the three π -orbital domains are iden-
20 tical, and this means that the energy is invariant to unitary transformations
21 between these orbitals. In contrary, with the PM procedure the localization
22 in benzene is not unique. There is one redundant orbital rotation, which
23 does not change the localization criterion. Consequently, depending on the
24 details of the localization procedure, different localized π -orbitals can result,
25 and generally the corresponding three orbital domains are not equivalent.
26 The energy is then not uniquely defined. In the past, we have overcome this
27 problem by merging the three domains so that again they become identical
28 and the energy invariant.

29 In the hexatriene case, each π domain includes only 2 atoms as long as the
30 threshold stays above 0.03. For the smaller thresholds of 0.025 or 0.01 the
31 first and second neighboring atoms, respectively, are included in the domains.
32 A similar situation is found for furan. In general, the charge distribution and
33 the related domain sizes are directly related to the amount of delocalization
34 and aromaticity of the molecule and can therefore also be used for a physical
35 interpretation of the bonding situation. For the purpose of local correlation
36 calculations, a value $T_{\text{NPA}} = 0.05$ appears to be most appropriate.

37 Fig. 5 shows the fraction of correlation energy recovered relative to canon-
38 ical MP2 using the PM/BP and NLMO/NPA methods. In the latter case,
39 two different thresholds were used for comparison. The results are rather
40 similar in all cases and differ mainly for the aromatic molecules due to the
41 different sizes of the π -orbital domains. The largest fraction of correlation
42 energy is recovered for the very small molecules water and formaldehyde, the
43 smallest one for alkanes like pentane or cyclohexane. The surprising fact that
44 these most saturated and well localized systems are most strongly affected
45 by the domain approximation has been discussed before[18]. Most likely,
46
47
48
49
50
51
52
53
54
55

1
2
3
4
5
6
7
8 this is related to the intramolecular basis set superposition error, which is
9 expected to be largest for molecules in which many atoms have a tetrahedral
10 environment. In the local methods, the BSSE is minimized by construction.
11 Clearly, these variations can have a significant effect on reaction energies.
12 One extreme case, the hydration of benzene to cyclohexane, has been stud-
13 ied in Ref. [18]. In that case it has been found that the difference between
14 the local and canonical results decreases with increasing basis set, and the
15 local result converges faster to the basis set limit. In a more recent extensive
16 study of over 50 reactions[19], it has been found that on the average the lo-
17 cal and canonical CCSD(T) results are of similar accuracy. If necessary, the
18 canonical results can be closely approached by domain extensions. Typically,
19 if the first shell of neighboring atoms of the standard domains as discussed
20 in this paper are added, 99.7-99.8% of the canonical correlation energy are
21 recovered.
22
23
24

25 One last comment should be made about the gradient evaluation. Con-
26 trary to Pipek-Mezey localization, there is no simple minimization criterion
27 for NLMOs, as needed for the calculation of analytical energy gradients. A
28 possible solution to this problem is to use the NPA domain criterion to-
29 gether with Pipek-Mezey orbitals (PM/NPA). The matrix \mathbf{V} in Eq. (10) is
30 then substituted by the transformation matrix from NAOs to PM LMOs. We
31 used the NLMOs as starting guess for the PM localization, in order to keep
32 the PM orbitals as similar as possible to the NLMOs in cases in which PM
33 localization is not unique.
34
35

36 The PM/NPA combination was tested for all 30 molecules and 6 ba-
37 sis sets. Using the standard PM method, the domains were found to still
38 vary considerably, although less than in the PM/BP case. However, sig-
39 nificant improvements could be achieved by removing some functions from
40 the localization criterion. For the cc-pVXZ basis sets, the most diffuse ba-
41 sis function of each angular momentum type for each atom was removed.
42 For the augmented basis sets the two most diffuse functions were removed
43 (as already mentioned, this can be done by zeroing the corresponding rows
44 and columns of the overlap matrix used in the PM procedure). With these
45 changes the NPA-based center charges are almost as stable as those of
46 the NLMO/NPA combination. For the recommended value of $T_{\text{NPA}} = 0.05$
47 only two domain changes were observed for the whole test set. In the benzo-
48 quinone and formaldehyde molecules one of the carbonyl oxygen lone pairs
49 changed to a CO bond for some basis sets.
50
51
52
53
54
55
56
57
58
59
60

4 Conclusions

The use of natural localized molecular orbitals in local correlation calculation has been studied. Based on the natural population analysis, we have proposed a new automatic procedure for the selection of domains in local methods. This new method can be used both for PM orbitals and for NLMOs. The use of NLMOs is particularly useful in aromatic molecules like benzene, since in contrast to other localization schemes the orbitals are uniquely defined. Furthermore, the localization is independent of the basis set and diffuse function do not cause any problems. The new criterion for domain selection has been proven to be very stable with respect to the basis set. For a test set of 30 molecules and the recommended selection criterion $T_{\text{NPA}} = 0.05$ not a single domain changed for all 6 basis set studied. Furthermore, the domains are physically meaningful. The fraction of correlation energy recovered in LMP2 relative to canonical MP2 is not much affected and similar as with the PM/BP method used so far. Therefore, it can be expected that previous conclusions regarding the accuracy of local correlation methods will not be much affected. Further systematic studies of reaction energies are in progress.

Finally, it should be mentioned that the new method does not eliminate the problem that domains may change as a function of the geometry, in particular in heterolytic bond breaking processes[42, 43]. However, as has been demonstrated in previous work, in most cases it is an excellent approximation to determine the domains at the equilibrium distance and then keep them fixed[43]. This even holds for the calculation of anharmonic vibrational frequencies, which require quite large variations of bond distances[26]. In some cases, in particular when calculating barrier heights of chemical reactions, it may be advantageous to merge the domains of the structures under consideration[43]. A recent example of highly accurate calculations of barrier heights in enzymes can be found in Ref. [17].

5 Acknowledgments

RAM gratefully acknowledges a PhD grant from Fundação para a Ciência e Tecnologia (reference SFRH/BD/13817/2003). This work has been supported by the Deutsche Forschungsgemeinschaft (SFB 706).

References

- [1] P. Pulay. Localizability of dynamic electron correlation. *Chem. Phys. Lett.*, 100:151–154, 1983.
- [2] S. Saebø and P. Pulay. Local configuration-interaction: An efficient approach for larger molecules. *Chem. Phys. Lett.*, 113:13–18, 1985.
- [3] P. Pulay and S. Saebø. Orbital-invariant formulation and second-order gradient evaluation in Møller-Plesset perturbation theory. *Theor. Chim. Acta*, 69:357–368, 1986.
- [4] S. Saebø and P. Pulay. Fourth-order Møller-Plesset perturbation theory in the local correlation treatment. I. Method. *J. Chem. Phys.*, 86:914–922, 1987.
- [5] S. Saebø and P. Pulay. The local correlation treatment. II. Implementation and tests. *J. Chem. Phys.*, 88:1884–1890, 1988.
- [6] M. Schütz, G. Hetzer, and H.-J. Werner. Low-order scaling local electron correlation methods. I. Linear scaling local MP2. *J. Chem. Phys.*, 111:5691–5705, 1999.
- [7] C. Hampel and H.-J. Werner. Local treatment of electron correlation in coupled cluster theory. *J. Chem. Phys.*, 104:6286–6297, 1996.
- [8] M. Schütz and H.-J. Werner. Low-order scaling local electron correlation methods. IV. Linear scaling local coupled-cluster (LCCSD). *J. Chem. Phys.*, 114:661–681, 2001.
- [9] M. Schütz and H.-J. Werner. Local perturbative triples correction (T) with linear cost scaling. *Chem. Phys. Lett.*, 318:370–378, 2000.
- [10] M. Schütz. Low-order scaling local electron correlation methods. III. Linear scaling local perturbative triples correction (T). *J. Chem. Phys.*, 113:9986–10001, 2000.
- [11] M. Schütz. Linear scaling local connected triples beyond local (T): Local CCSDT-1b with O(N) scaling. *J. Chem. Phys.*, 116:8772–8785, 2002.

- 1
2
3
4
5
6
7
8 [12] H.-J. Werner, F. R. Manby, and P.J. Knowles. Fast linear scaling second-
9 order Møller-Plesset perturbation theory (MP2) using local and density
10 fitting approximations. *J. Chem. Phys.*, 118:8149–8160, 2003.
11
12 [13] M. Schütz, H.-J. Werner, R. Lindh, and F. R. Manby. Analytical energy
13 gradients for local second-order Møller-Plesset perturbation theory using
14 density fitting approximations. *J. Chem. Phys.*, 121:737–750, 2004.
15
16 [14] R. Polly, H.-J. Werner, F. R. Manby, and Peter J. Knowles. Fast
17 Hartree-Fock theory using local density fitting approximations. *Mol.*
18 *Phys.*, 102:2311–2321, 2004.
19
20 [15] M. Schütz and F. R. Manby. Linear scaling local coupled cluster theory
21 with density fitting. I : 4-external integrals. *Phys. Chem. Chem. Phys.*,
22 5:3349–3358, 2003.
23
24 [16] H.-J. Werner and M. Schütz. Local coupled cluster methods with density
25 fitting. to be published, 2007.
26
27 [17] F. Clayessens, J. N. Harvey, F. R. Manby, R. A. Mata, A. J. Mulholland,
28 K. E. Ranaghan, M. Schütz, S. Thiel, W. Thiel, and H.-J. Werner. High
29 accuracy computation of reaction barriers in enzymes. *Angew. Chemie*
30 *int. ed.*, 45:6856, 2006.
31
32 [18] H.-J. Werner and K. Pflüger. On the selection of domains and orbital
33 pairs in local correlation treatments. *Ann. Reports in Comput. Chem.*,
34 2:53, 2006.
35
36 [19] K. Pflüger, H.-J. Werner, A. Schäfer, and M. Schütz. The effect of local
37 approximations on reaction energies. *to be published*, 2007.
38
39 [20] T. Korona, K. Pflüger, and H.-J. Werner. The effect of local approxi-
40 mations in coupled-cluster wavefunctions on dipole moments and static
41 dipole polarisabilities. *Phys. Chem. Chem. Phys.*, 6:2059–2065, 2004.
42
43 [21] A. ElAzhary, G. Rauhut, P. Pulay, and H.-J. Werner. Analytical energy
44 gradients for local second-order Møller-Plesset perturbation theory. *J.*
45 *Chem. Phys.*, 108:5185–5193, 1998.
46
47 [22] G. Rauhut and H.-J. Werner. Analytical energy gradients for local
48 coupled-cluster methods. *Phys. Chem. Chem. Phys.*, 3:4853–4862, 2001.
49
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3
4
5
6
7
8 [23] G. Rauhut, A. El Azhary, F. Eckert, U. Schumann, and H.-J. Werner. Impact of local approximations on MP2 vibrational frequencies. *Spectrochim. Acta*, A 55:647–658, 1999.
- 9
10
11
12 [24] G. Rauhut and H.-J. Werner. The vibrational spectra of furoxan and dichlorofuroxan: a comparative theoretical study using density functional theory and local electron correlation methods. *Phys. Chem. Chem. Phys.*, 5:2001–2008, 2003.
- 13
14
15
16
17
18 [25] T. Hrenar, G. Rauhut, and H.-J. Werner. Impact of local and density fitting approximations on harmonic vibrational frequencies. *J. Phys. Chem. A*, 110:2060–2064, 2006.
- 19
20
21
22
23 [26] T. Hrenar, H.-J. Werner, and G. Rauhut. Accurate calculation of anharmonic vibrational frequencies of medium sized molecules using local coupled cluster methods. *J. Chem. Phys.*, 126:134108, 2007.
- 24
25
26
27 [27] S. Saebø, W. Tong, and P. Pulay. Efficient elimination of basis set superposition errors by the local correlation method: Accurate ab initio studies of the water dimer. *J. Chem. Phys.*, 98:2170–2175, 1993.
- 28
29
30
31
32 [28] M. Schütz, G. Rauhut, and H.-J. Werner. Local treatment of electron correlation in molecular clusters: Structures and stabilities of $(\text{H}_2\text{O})_n$, $n = 2-4$. *J. Phys. Chem. A*, 102:5997–6003, 1998.
- 33
34
35
36 [29] N. Runeberg, M. Schütz, and H.-J. Werner. *J. Chem. Phys.*, 110:7210, 1999.
- 37
38
39
40 [30] J. G. Hill, J. A. Platts, and H.-J. Werner. Calculation of intermolecular interactions in the benzene dimer using coupled-cluster and local electron correlation methods. *Phys. Chem. Chem. Phys.*, 8:4072, 2006.
- 41
42
43
44 [31] S. F. Boys. Localized orbitals and localized adjustment functions. In P. O. Löwdin, editor, *Quantum Theory of Atoms, Molecules, and the Solid State*, pages 253–262. Academic Press, New York, 1966.
- 45
46
47
48
49 [32] J. Pipek and P. G. Mezey. A fast intrinsic localization procedure applicable for ab initio and semiempirical linear combination of atomic orbital wave functions. *J. Chem. Phys.*, 90:4916–4926, 1989.
- 50
51
52
53
54
55
56
57
58
59
60

- 1
2
3
4
5
6
7
8 [33] J. W. Boughton and P. Pulay. Comparison of the Boys and Pipek-Mezey
9 localizations in the local correlation approach and automatic virtual
10 basis selection. *J. Comput. Chem.*, 14:736–740, 1993.
11
12 [34] A. E. Reed, R. B. Weinstock, and F. Weinhold. Natural population
13 analysis. *J. Chem. Phys.*, 83:735, 1985.
14
15 [35] A. E. Reed and F. Weinhold. Natural localized molecular orbitals. *J.*
16 *Chem. Phys.*, 83:1736, 1985.
17
18 [36] N. Flocke and Rodney J. Bartlett. A natural linear scaling coupled-
19 cluster method. *J. Chem. Phys.*, 121:10935, 2004.
20
21 [37] A. E. Reed and F. Weinhold. Natural bond orbital analysis of near-
22 hartree-fock water dimer. *J. Chem. Phys.*, 78:4066, 1983.
23
24 [38] J. P. Foster and F. Weinhold. Natural hybrid orbitals. *J. Am. Chem.*
25 *Soc.*, 102:7211, 1980.
26
27 [39] H.-J. Werner, P. J. Knowles, R. Lindh, F. R. Manby, M. Schütz, et al.
28 Molpro, version 2006.3, a package of ab initio programs, 2007. see
29 <http://www.molpro.net>.
30
31 [40] T. H. Dunning, Jr. Gaussian basis sets for use in correlated molecular
32 calculations. I. The atoms boron through neon and hydrogen. *J. Chem.*
33 *Phys.*, 90:1007–1023, 1989.
34
35 [41] R. A. Kendall, T. H. Dunning, and R. J. Harrison. *J. Chem. Phys.*,
36 96:6796, 1992.
37
38 [42] N. J. Russ and T. D. Crawford. Potential energy surface discontinuities
39 in local correlation methods. *J. Chem. Phys.*, 121(2):691–696, 2004.
40
41 [43] R. A. Mata and H.-J. Werner. Calculation of smooth potential en-
42 ergy surfaces using local electron correlation methods. *J. Chem. Phys.*,
43 125:184110, 2006.
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

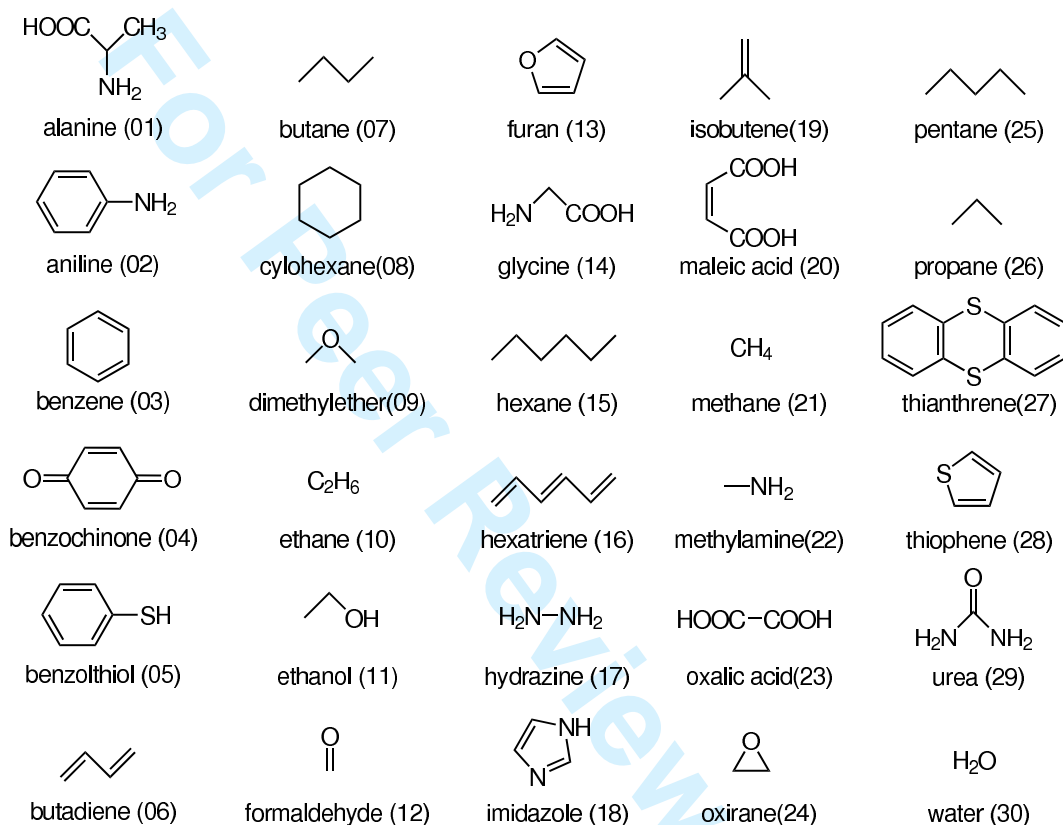


Figure 1: Test set of 30 molecules used in this work. All geometries were pre-optimized with B3LYP/cc-pVTZ(d/p) (including up to d functions for the second and third row elements, and up to p functions for the hydrogens). The numbering shown is used in some of the diagrams.

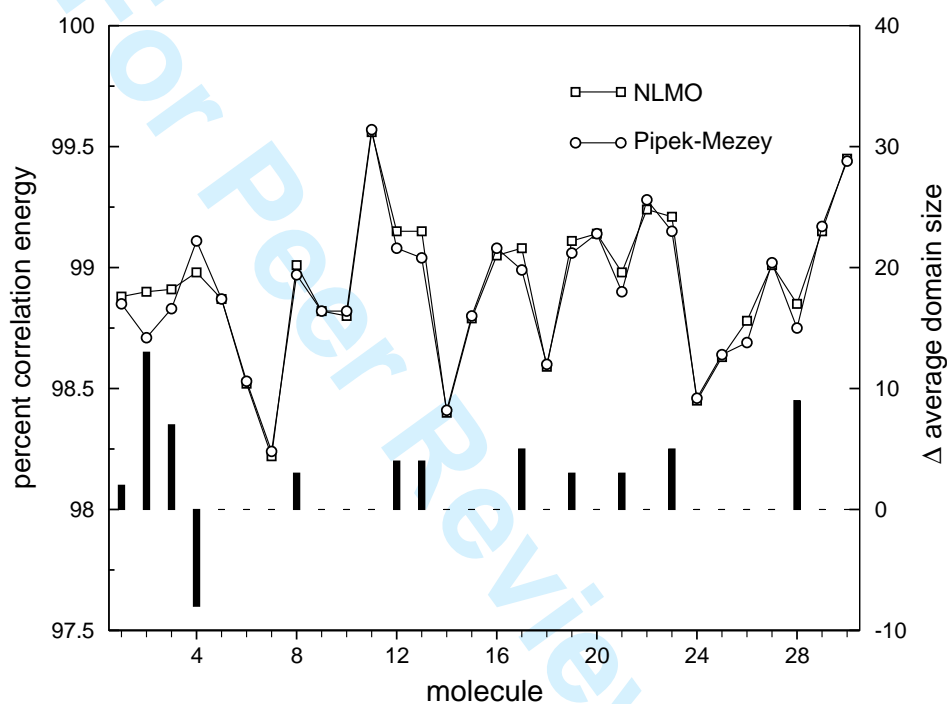


Figure 2: Percentage of correlation energy recovered using NLMO and Pipek-Mezey orbitals in LMP2 calculations, and with $T_{BP} = 0.985$. Also shown (in bars) the average domain size difference between the two sets. The basis set used was cc-pVTZ. The molecule numbers refer to the ones used in Fig. 1.

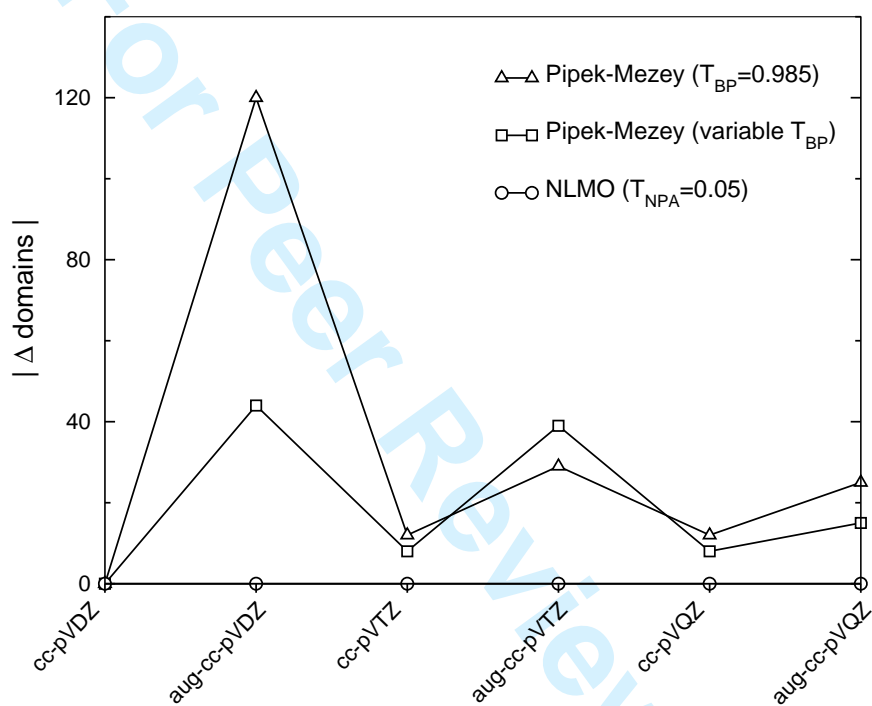


Figure 3: Sum of the domain variations (absolute) for the thianthrene molecule with different localization procedures and domain selection test. For the variable T_{BP} case, different parameters for localization and domain selection were used (see text for more information).

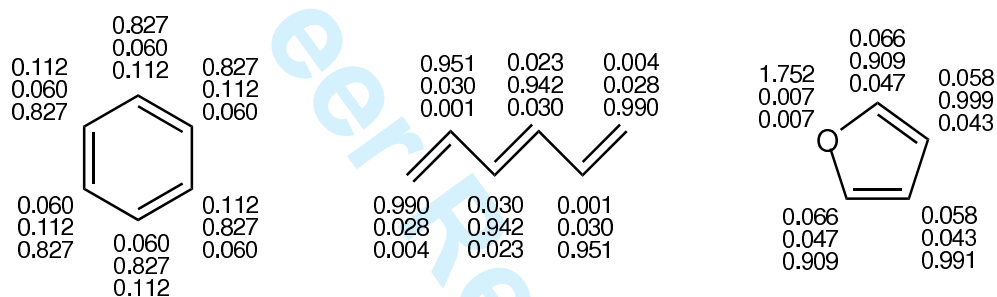


Figure 4: NPA calculated π orbitals charges for NLMO orbitals. The basis set used was aug-cc-pVQZ. The values are given for each heavy atom, with the orbitals ordered vertically.

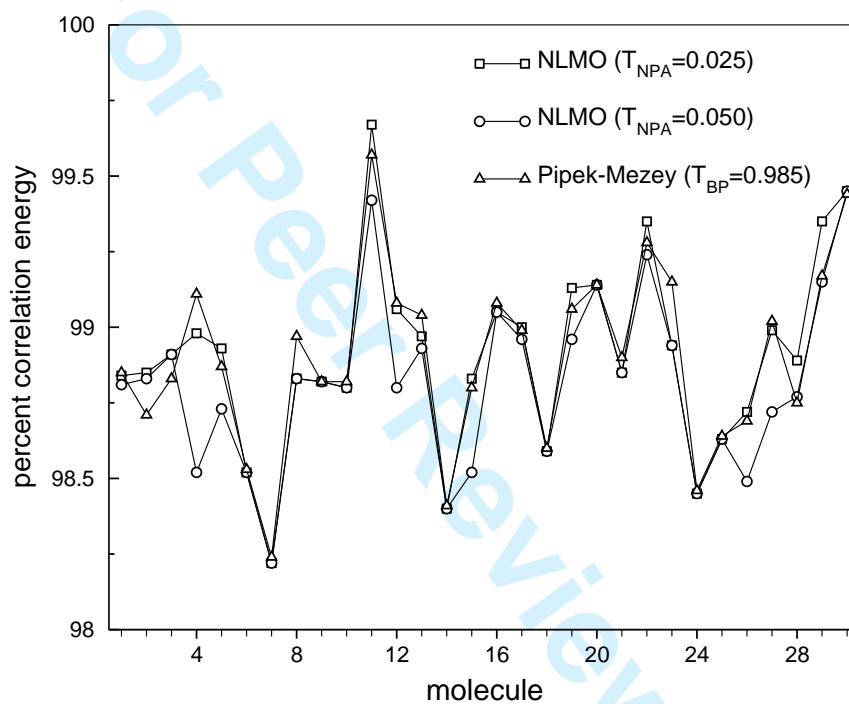
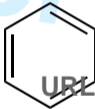


Figure 5: Comparison of the fraction of correlation energy relative to canonical MP2 (in percent) using the NLMO/NPA and PM/BP methods for localization and domain selection. The cc-pVTZ basis set was used.

Molecular Physics

Page 22 of 28



0.112
0.060
0.827

0.827
0.060
0.112

0.827
0.112
0.060

0.990 0.023 0.004
0.030 0.942 0.028
0.001 0.030 0.990

1.752 0.066 0.058
0.007 0.047 0.999
0.007 0.043

1

0.060
0.112
0.827

0.112
0.827
0.060

0.990 0.001 0.001
0.028 0.942 0.030
0.004 0.023 0.951

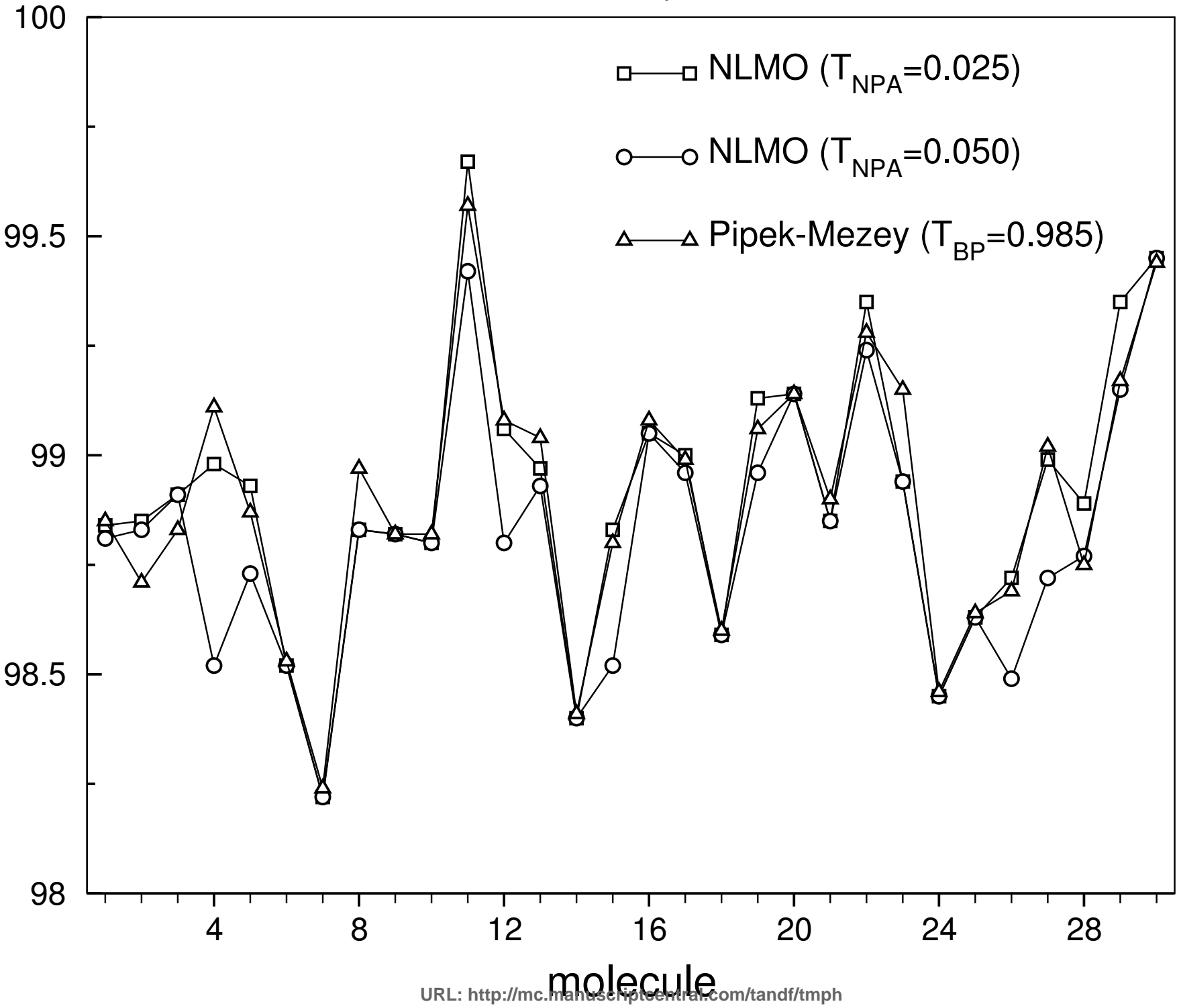
0.066 0.058
0.047 0.043
0.909 0.991

4

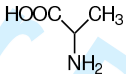
0.060
0.827
0.112

URL: <http://mc.manuscriptcentral.com/tandf/tmph>

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49



Molecular Physics



alanine (01)



butane (07)



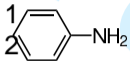
furan (13)



isobutene(19)



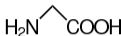
pentane (25)



aniline (02)



cyclohexane(08)



glycine (14)



maleic acid (20)



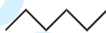
propane (26)



benzene (03)



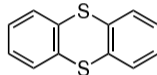
dimethylether(09)



hexane (15)



methane (21)



thianthrene(27)



benzochinone (04)



ethane (10)



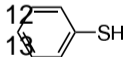
hexatriene (16)



methylamine(22)



thiophene (28)



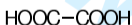
benzothiol (05)



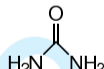
ethanol (11)



hydrazine (17)



oxalic acid(23)



urea (29)



butadiene (06)



formaldehyde (12)



imidazole (18)

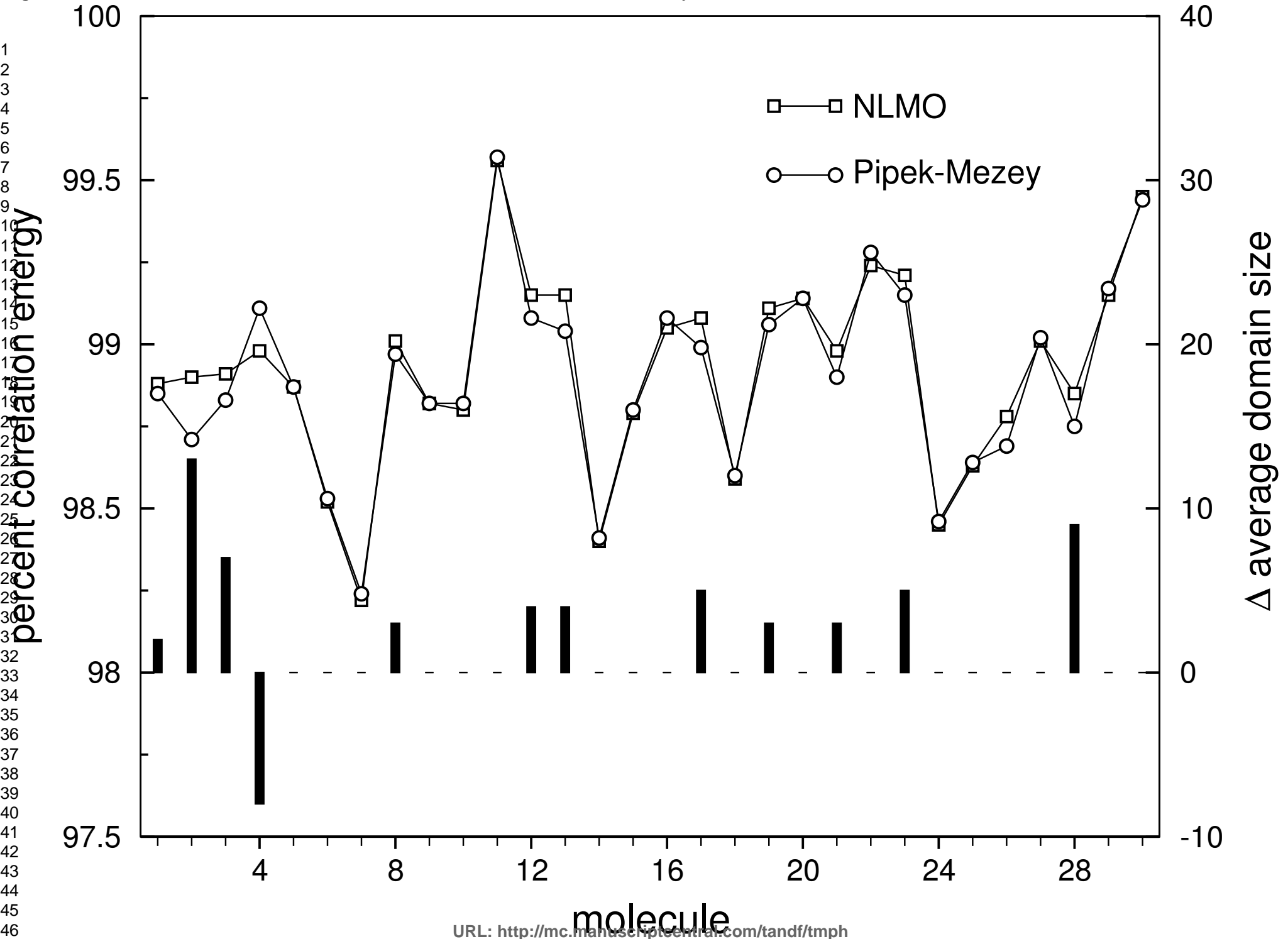


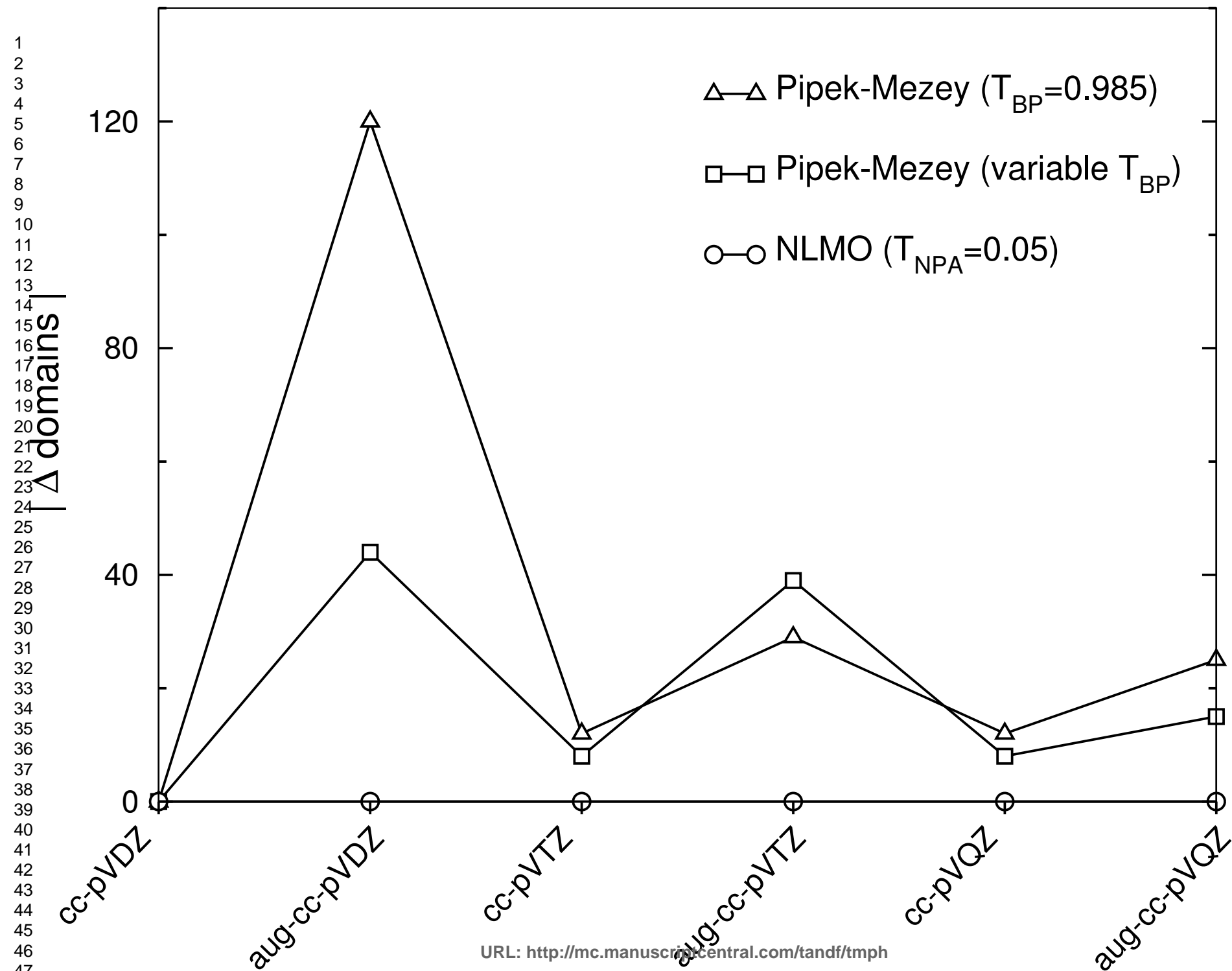
oxirane(24)



water (30)

URL: <http://mc.manuscriptcentral.com/tandf/tmph>





1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For Peer Review Only

