JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Article

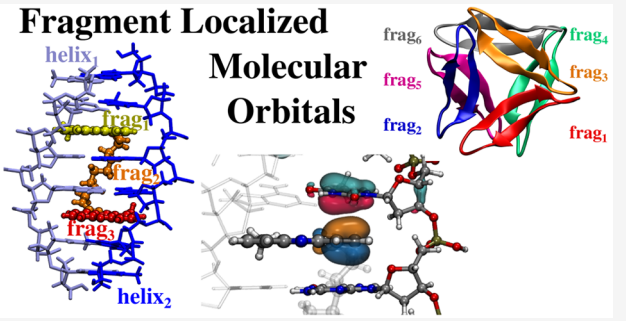
Fragment Localized Molecular Orbitals

Tommaso Giovannini* and Henrik Koch*

Cite This: J. Chem. Theory Comput. 2022, 18, 4806−4813 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: We introduce the concept of fragment localized molecular orbitals (FLMOs), which are Hartree−Fock molecular orbitals localized in specific fragments constituting a molecular system. In physical terms, we minimize the local electronic energies of the different fragments, at the cost of maximizing the repulsion between them. To showcase the approach, we rationalize the main interactions occurring in large biological systems in terms of interactions between the fragments of the system. In particular, we study an anticancer drug intercalated within DNA and retinal in anabaena sensory rhodopsin as prototypes of molecular systems embedded in biological matrixes. Finally, the FLMOs are exploited to rationalize the formation of two oligomers, prototypes of amyloid diseases, such as Parkinson and Alzheimer.



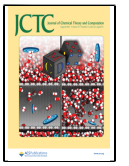
1. INTRODUCTION

Intermolecular interactions play a crucial role in many chemical and biological processes.[1] For instance, they are the guiding force of different metabolic pathways and regulate the physicochemical properties of many materials and their interaction with light.[2−4] A theoretical understanding of intermolecular interactions, together with the different energy contributions constituting them, is of particular importance, providing a solid base on which novel experimental setups/reactions can be designed. In the last decades, the size of the molecular systems which can be treated computationally has dramatically increased aided by developed methods with a favorable scaling with the system size.[5−8] Such approaches has allowed for the description of more and more complex systems, without renouncing accuracy. In particular, they are typically based on chemical intuition which guides a hierarchical partitioning of the investigated system.[9−11] Such approximations are justified by the assumption that the property of interest (energetics and/or spectroscopy) is usually local and carried by a specific part of the system, as in the case of local electronic excitations.[11−16] To rationalize local properties, the standard procedure is to resort to localized molecular orbitals (LMOs) which usually act as a bridge between chemical intuition and theoretical chemistry.[17−21] The success of LMOs is demonstrated by the plethora of applications in which they have been used by both theoretical and experimental groups.[15,22−25]

In this work, we introduce a novel class of LMOs, which are localized in specific fragments and/or portions of molecular systems. For this particular feature, we call them fragment localized molecular orbitals (FLMOs). The potential of FLMOs is evident: the local properties and/or energetics of different fragments interacting together can be calculated without resorting to any kind of approximation. The FLMOs are defined in the Hartree−Fock (HF) context. The idea behind our approach is rather simple: by reformulating the HF energy for a system constituted by several fragments, we minimize the local electronic energies of each of them, by keeping the total energy of the system unchanged. Therefore, such a minimization is carried out at the cost of maximizing the repulsion between the fragments and consequently providing the MOs mostly localized within the different fragments. Our approach is a top-down variational methodology, since the localization is performed by starting from the full canonical MOs. For the aforementioned reason, it differs from other approaches proposed in the literature to deal with fragment localized MOs.[26−31] Note also that the term FLMO has been first proposed in ref 31, where fragment localized MOs are obtained in terms of the subsystems' MOs previously determined, i.e., in a "fragment to molecule" fashion.

In this paper, we propose an application of FLMOs to the calculation of interaction energies of interesting molecular systems and, in particular, a decomposition of the interaction energy among the FLMOs belonging to different fragments is presented. Such a decomposition can provide a graphical and intuitive interpretation of the interactions among the interacting systems, which can guide both theoretical and experimental chemists in the in-silico design of novel processes. To showcase

Received: April 12, 2022
Published: July 27, 2022



https://doi.org/10.1021/acs.jctc.2c00359
J. Chem. Theory Comput. 2022, 18, 4806−4813

how FLMOs can be used in chemical and biological relevant applications, we study the fragment interaction energies of an anticancer drug, whose mechanism is evinced in its intercalation within DNA,[32] in all trans-retinal embedded in a rhodopsin protein environment,[33] as relevant biological applications. Finally, we investigate the molecular interactions within two oligomers, prototypes of the packaging involved in amyloid diseases, such as Parkinson or Alzheimer.[34−36]

2. THEORETICAL METHOD

The starting point of our localization procedure is the Hartree−Fock (HF) energy of a generic molecular system[37]

$$E = \mathrm{Tr}\,\mathbf{hD} + \frac{1}{2}\mathrm{Tr}\,\mathbf{DG}(\mathbf{D}) + h_{\mathrm{nuc}} \tag{1}$$

where $\mathbf{h}$ and $\mathbf{G}$ are the usual one- and two-electron matrices, $\mathbf{D}$ is the density matrix, and h_{nuc} is the nuclear repulsion.[6] If the system is partitioned in N_f fragments, the density matrix $\mathbf{D}$ can be decomposed as[21]

$$\mathbf{D} = \sum_{i}^{N_f} \mathbf{D}^i \tag{2}$$

where $\mathbf{D}^i$ is the density matrix of the ith fragment. The total energy defined in eq 1 can be rewritten as

$$E = \underbrace{\sum_{i}^{N_f} \mathrm{Tr}\,\mathbf{h}^i\mathbf{D}^i + \frac{1}{2}\sum_{i}^{N_f} \mathrm{Tr}\,\mathbf{D}^i\mathbf{G}(\mathbf{D}^i) + \sum_{i}^{N_f} h^i_{\mathrm{nuc}}}_{\sum_i E^i} + \underbrace{\sum_{i,j\neq i}^{N_f} \mathrm{Tr}\,\mathbf{V}^i\mathbf{D}^j + \sum_{i,j>i}^{N_f} \mathbf{D}^i\mathbf{G}(\mathbf{D}^j) + h^X_{\mathrm{nuc}}}_{\text{interaction energy}} \tag{3}$$

where h^X_{nuc} is the nuclear repulsion between the N_f fragments, defined as $h^X_{\mathrm{nuc}} = h_{\mathrm{nuc}} - \sum_i^{N_f} h^i_{\mathrm{nuc}}$. The $\mathbf{h}^i$ and $\mathbf{G}(\mathbf{D})^i$ matrices are defined as

$$h^i_{\mu\nu} = T_{\mu\nu} + V^i_{\mu\nu}$$
$$G_{\mu\nu}(\mathbf{D}^i) = \sum_{\sigma\tau} D^i_{\sigma\tau}\left((\mu\nu|\sigma\tau) - \frac{1}{2}(\mu\tau|\sigma\nu)\right) \tag{4}$$

where $T_{\mu\nu}$ and $V^i_{\mu\nu}$ are the kinetic and electron−nuclear attraction matrix elements expressed in the full AO basis ($\{\chi_\mu\}$), respectively. Note that in eq 3, the energy of the ith fragment and the interaction energy between all N_f fragments have been highlighted.

In order to localize the MOs in the predefined fragments, we can minimize the sum of the fragment energies (E^i) as defined in eq 3 by rotating the occupied orbitals belonging to the different fragments, thus conserving the total energy of the system. This clearly corresponds to maximizing the interaction energy and therefore the repulsion between the fragments. In physical terms, this means that the minimization of the local electronic energies of the fragments can be carried out at the cost of maximizing the repulsion between them. From a computational point of view, we solve this problem by performing $N_f - 1$ different self-consistent field (SCF) steps. At the first step, the MOs of the first fragment are localized by maximizing the repulsion with all the other fragments. This step is performed in the MO space spanned by the occupied orbitals of all the fragments. Once the first fragment's occupied MOs are localized, the process is repeated for all the other fragments, by reducing at each step the MO space in which the SCF procedure is performed. In fact, at the each cycle the MO space is reduced, because the occupied MOs localized at the previous step are removed. In this way, at the final $N_f - 1$ step, all the occupied MOs, which are expressed in the full AO basis, are still orthogonal and localized in their predefined fragments. For this reason, we propose to call them fragment localized molecular orbitals (FLMOs). It is worth remarking that after a fragment localization is performed, the following SCF cycles only rotate the occupied orbitals of the nonlocalized fragments, thus not affecting their total density. Note that, since the procedure involves $N_f - 1$ steps, the whole computational costs of the procedure is $N_f \cdot t_{\mathrm{SCF}}$, where t_{SCF} is the timing associated with each SCF procedure. Note that the t_{SCF} associated with each SCF step can be significantly reduced by exploiting the procedure recently proposed by Koch and co-workers.[38] The computational protocol is summarized in section S1 given in the Supporting Information, where we provide practical examples for covalently and noncovalently bonded systems.

One of the crucial steps in our procedure is the definition of the starting fragment MOs which are then localized. They are obtained by performing a partial Cholesky decomposition of the total density matrix at each localization step. Such a procedure has recently been detailed by us in ref 39, in which we also demonstrate that such a partitioning method guarantees the continuity of the potential energy surface (PES), differently from other localization techniques such as Boys.[18] It is worth noting that the methodology here presented is completely general and any partitioning method can be used to obtain the starting MOs. However, the absence of any discontinuities in the PES must be carefully checked.[15]

Once the FLMOs are obtained, the corresponding fragment density matrices (FDM) can be constructed

$$D^i_{\mu\nu} = 2\sum_{\alpha\in i} C_{\mu\alpha}C_{\nu\alpha} + \sum_{\beta\in i} C_{\mu\beta}C_{\nu\beta} \tag{5}$$

where α and β labels run over the doubly and singly occupied MOs, respectively, both belonging to the ith fragment. For instance, in the case of a single covalent bond between two fragments is cut, we may assign one electron to each fragment sharing the chemical bond (see also section S1.2 in the Supporting Information). Both the FLMOs and FDMs are defined on the full AO basis set $\{\chi_\mu\}$. This means that they may be characterized by tails in other fragments; however, such a property guarantees the orthogonality between them and the correct reproduction of the full HF results.

As stated above, the FLMOs which are obtained with our procedure may have different applications, ranging from reactivity to reduce the computational time associated with correlated methods. In this work, we decide to exploit them to the particular case of interaction energies. Once the FDMs have been computed, the interaction energy between the ith and jth fragments $E^{\mathrm{int,FLMO}}_{ij}$ can be defined as

$$E^{\mathrm{int,FLMO}}_{ij} = \mathrm{Tr}\,\mathbf{V}^i\mathbf{D}^j + \mathrm{Tr}\,\mathbf{V}^j\mathbf{D}^i + \mathrm{Tr}\,\mathbf{D}^i\mathbf{J}(\mathbf{D}^j) + h^{ij}_{\mathrm{nuc}} + \mathrm{Tr}\,\mathbf{D}^i\mathbf{K}(\mathbf{D}^j) \tag{6}$$

$$E^{\mathrm{ele}}_{ij} = \mathrm{Tr}\,\mathbf{V}^i\mathbf{D}^j + \mathrm{Tr}\,\mathbf{V}^j\mathbf{D}^i + \mathrm{Tr}\,\mathbf{D}^i\mathbf{J}(\mathbf{D}^j) + h^{ij}_{\mathrm{nuc}} \tag{7}$$

https://doi.org/10.1021/acs.jctc.2c00359
J. Chem. Theory Comput. 2022, 18, 4806−4813

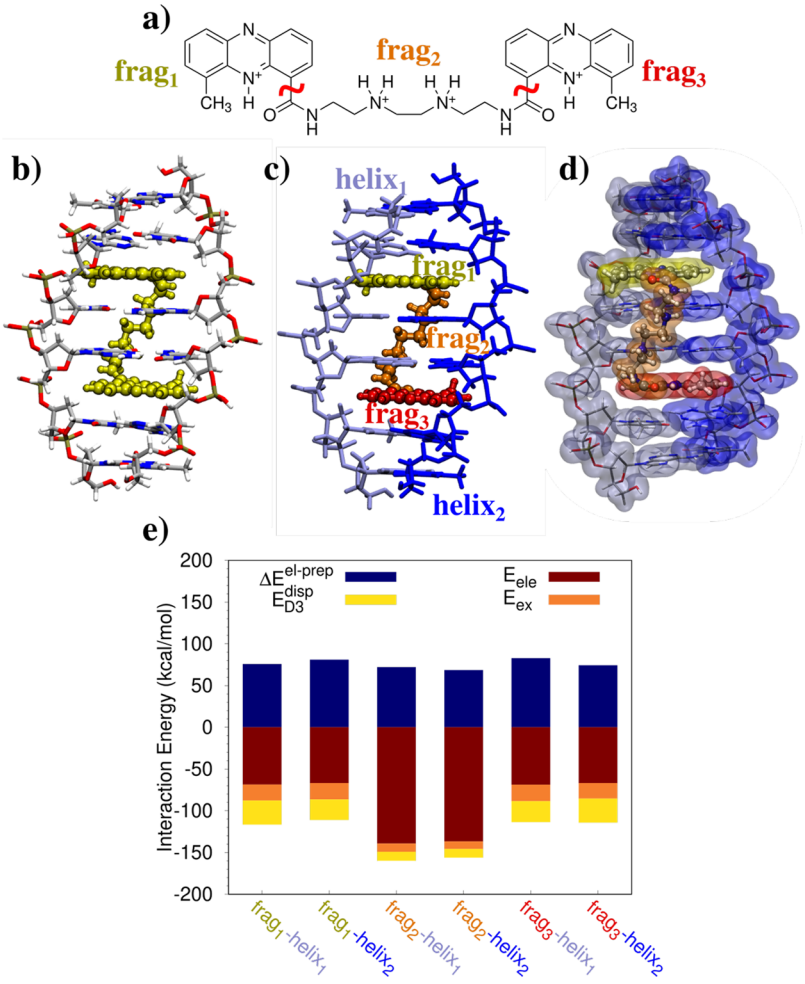


Figure 1. (a) Molecular structure of the studied drug (XR5944). (b) XR5944 intercalated within DNA (PDB 1X95), (c) the exploited fragmentation, and (d) the fragment densities obtained by using the FLMOs procedure. (e) FLMOs interaction energy decomposition analysis (kcal/mol).

$$E_{ij}^{\mathrm{ex}} = \mathrm{Tr}\ \mathbf{D}^{i}\mathbf{K}(\mathbf{D}^{j}) \tag{8}$$

where **J** and **K** are Coulomb and exchange matrices. In eqs 7 and 8, the different terms are grouped into the electrostatic E_{ij}^{ele} and the exchange E_{ij}^{ex} energy terms. The latter is always negative and therefore a stabilizing interaction. Note that the separation proposed in eqs 6−8 share the same formalism of other energy decomposition analysis.[40] We can define the total interaction energy between the *i*th and *j*th fragments E_{ij}^{int} as

$$E_{ij}^{\mathrm{int}} = E_{ij}^{\mathrm{int,FLMO}} + E_{i,j}^{\mathrm{el\text{-}prep}} \tag{9}$$

$$E_{i,j}^{\mathrm{el\text{-}prep}} = E_{i}^{\mathrm{el\text{-}prep}} + E_{j}^{\mathrm{el\text{-}prep}} - E_{ij}^{\mathrm{el\text{-}prep}} \tag{10}$$

where the electronic preparation energy $E_{i}^{\mathrm{el\text{-}prep}}$ is the energy needed to deform the electron density of the *i*th fragment from the vacuo to the supramolecular structure, whereas $E_{ij}^{\mathrm{el\text{-}prep}}$ indicates that of the dimer *ij*. Further details about eqs 9 and 10 are given in section S1.5 in the Supporting Information. For a generic monomer or dimer *r*, the $E_{r}^{\mathrm{el\text{-}prep}}$ is defined as the difference between the energy when *r* is embedded in the total system E_{r}^{emb}, and *r* energy in the gas-phase E_{r}^{vac}:

$$E_{r}^{\mathrm{el\text{-}prep}} = E_{r}^{\mathrm{emb}} - E_{r}^{\mathrm{vac}} \tag{11}$$

It is worth noting that for open-shell fragments, the vacuum calculations are computed at the unrestricted HF level. Moreover, the basis set superposition error (BSSE) can be easily reduced by using counterpoise correction, i.e., by calculating the vacuum energies (eq 11) in the full AO basis set.[41]

The total interaction energy between all fragments $E_{\mathrm{tot}}^{\mathrm{int}}$ can be defined by considering the energetic cost associated with the electronic preparations of the single monomers:

$$E_{\mathrm{tot}}^{\mathrm{int}} = \sum_{i,j>i}^{N_{\mathrm{f}}} E_{ij}^{\mathrm{int,FLMO}} + \sum_{i} E_{i}^{\mathrm{el\text{-}prep}} \tag{12}$$

$$= \sum_{i,j>i}^{N_{\mathrm{f}}} E_{ij}^{\mathrm{ele}} + \sum_{i,j>i}^{N_{\mathrm{f}}} E_{ij}^{\mathrm{ex}} + \sum_{i} E_{i}^{\mathrm{el\text{-}prep}} \tag{13}$$

where the electrostatic and exchange energy contributions are specified. The total interaction energy defined in eq 12 is by definition equal to the full HF case (see also section S1.5 in the Supporting Information). Finally, we note that eqs 5−8 can also be specified for two MOs belonging to two different fragments, thus allowing a further partitioning of the electrostatic and exchange contributions among the FLMOs of the two different fragments (see also section S1.1 in the Supporting Information). In this way, the interaction energy between two fragments can be

https://doi.org/10.1021/acs.jctc.2c00359
J. Chem. Theory Comput. 2022, 18, 4806−4813

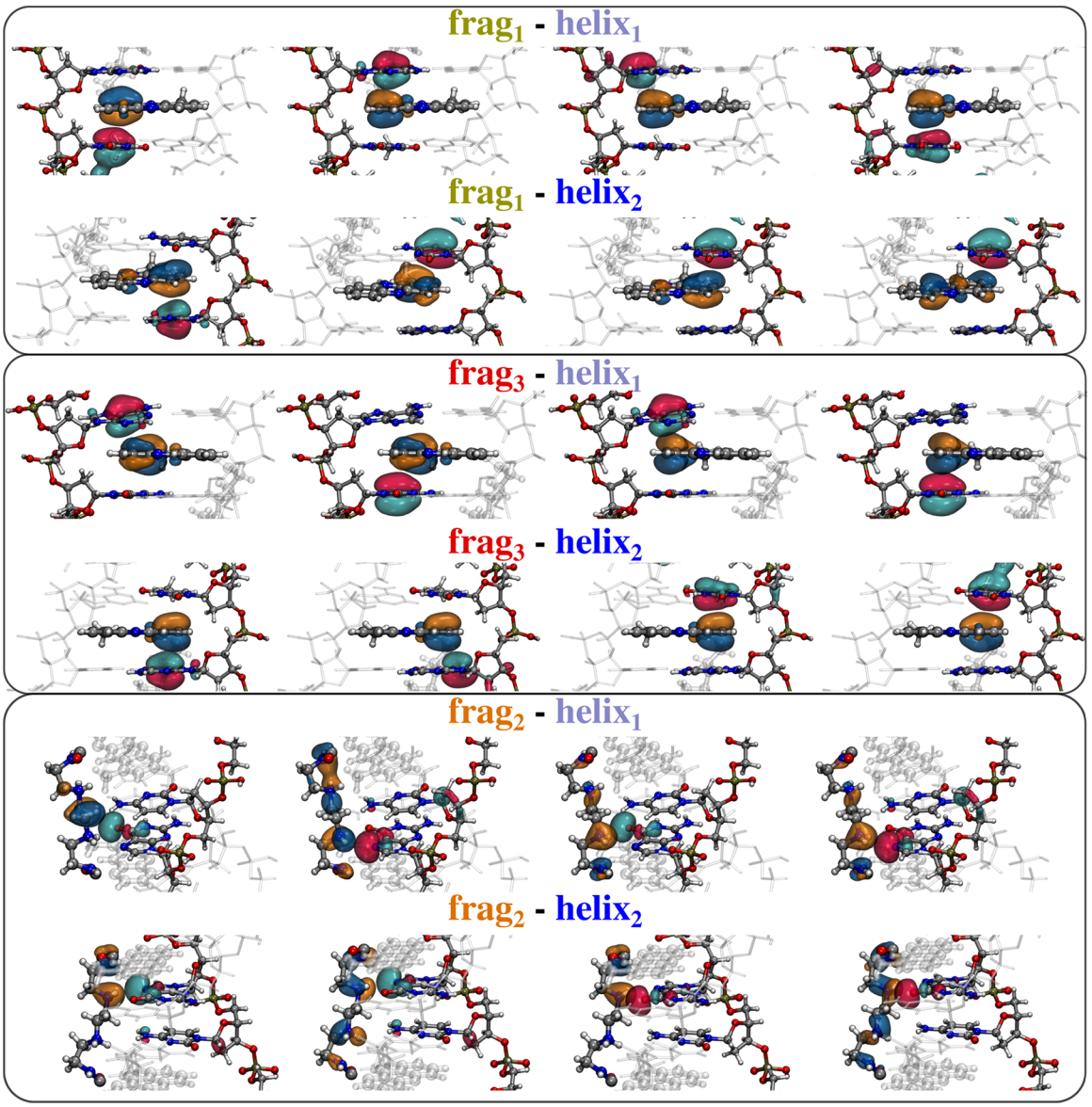


Figure 2. Graphical representation of the most important FLMOs interactions for all the studied fragment couples (top, $frag_1$-$helix_{1,2}$; middle, $frag_3$-$helix_{1,2}$; bottom, top: $frag_2$-$helix_{1,2}$).

rationalized in terms of the FLMOs belonging to the two fragments. In particular, we can decompose the interaction energy between two FLMOs in the electrostatic and exchange interactions, thus providing a physicochemical insight into the interactions.

3. NUMERICAL APPLICATIONS

To demonstrate the robustness and the potential of the method, we select two test applications: two molecular systems embedded in a biological matrix (DNA or protein)[32,33] and two different prototypes for the amyloid diseases.[34] For all systems, we first perform the FLMO localization, and then we decompose the interaction energies among the different terms as defined in eq 6. The FLMO decomposition is implemented in a development version of the electronic structure code $e^{\mathcal{T}}$.[42] Due to the size of the systems, the dispersion interaction is treated by Grimme's D3 method, which is an effective parameter-dependent approach,[43] thus it does not depend on the computed FLMOs. All calculations are performed on 28 processors Intel Xeon Gold 5120 CPU at 2.20 GHz.

3.1. Molecular Systems Embedded in Biological Matrixes. The first studied system is XR5944, an anticancer drug which is constituted of an aliphatic chain connecting two aromatic groups (see Figure 1a).[32,44−46] Similar to many other anticancer drugs,[47,48] the drug mechanism is based on the intercalation between two DNA basis pairs (see Figure 1b; PDB 1X95), in this way avoiding its duplication and the diffusion of the cancer. The physicochemical understanding of the main forces binding the molecular drug to the DNA can be of great relevance since it can guide the design of novel drugs with the same mechanism. To this end, we divide the drug in three open shell fragments: the two aromatic rings and the aliphatic chain (see Figure 1a,c). The two DNA helices are instead treated as two different fragments ($helix_1$ and $helix_2$). Finally, we compute the FLMOs and the interaction energy components at the HF/6-311G* level (total number of AOs, N_{AO} = 5756, average t_{SCF} = 1484 s). The resulting fragment densities, localized in their

https://doi.org/10.1021/acs.jctc.2c00359
J. Chem. Theory Comput. 2022, 18, 4806−4813

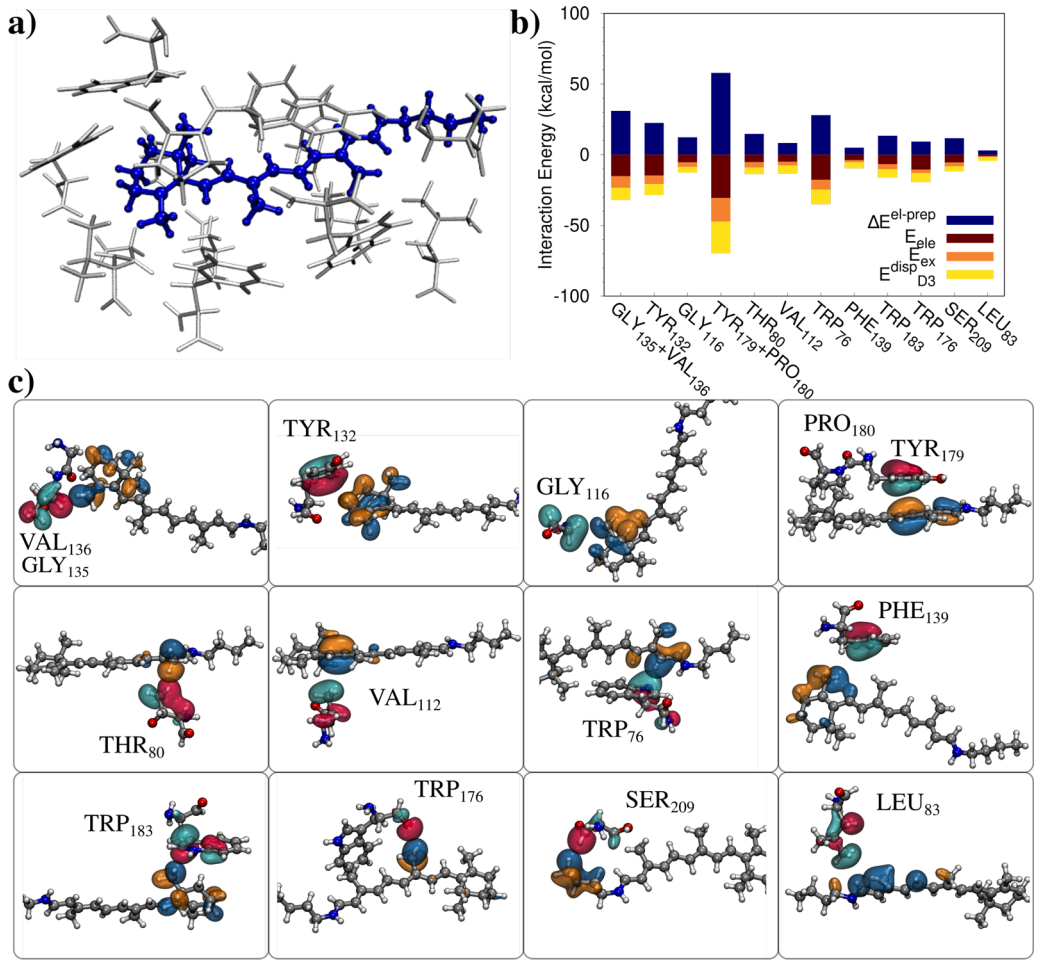


Figure 3. (a) Molecular structure of all-trans retinal chromophore embedded in the closest 14 amino acids of the anabaena sensory rhodopsin.[33] (b) FLMOs interaction energy decomposition analysis (kcal/mol). (c) Graphical representation of the most important FLMOs interactions between retinal and the 14 selected amino acids.

spatial regions, are reported in Figure 1d. The interaction energy components between the three drug fragments and the two helices are graphiscally depicted in Figure 1e (see Table S1 in the Supporting Information for raw data).

In all cases, the electrostatic interaction E_{ele} is negative, thus stabilizing the interaction of the drug with the DNA. For $frag_1$ and $frag_3$, i.e., the two aromatic rings, E_{ele} is always opposite to the purely repulsive contribution $\Delta E^{el\text{-}prep}$, which, when summed with E_{ex} represents the exchange-repulsion term ($E_{ex\text{-}rep}$), i.e., the Pauli repulsion.[49] For all the energetic contributions, the symmetry of the system is reflected in a symmetry of the energy terms, with $frag_1$ and $frag_3$ interacting most favorably with $helix_1$ and $helix_2$, respectively. Also, $frag_2$ is interacting in a symmetrical way with the two helices, with the total interaction energies differing by just 0.2 kcal/mol. Interestingly, the predominant contribution is the electrostatics, which is more than twice the corresponding energy terms computed for both $frag_1$ and $frag_3$. This makes the interaction of $frag_2$ with the DNA the leading interaction allowing for the intercalation, or at least the stabilization, of the drug within the DNA. This is the theoretical reason for which many anticancer drugs are designed with a similar lateral chain, which acts as an anchor hydrogen-bonded with the DNA, stabilizing the complex.

As detailed above, our method also allows for a qualitative and quantitative investigation of the interaction by decomposing both E_{ele} and E_{ex} in terms of the FLMOs belonging to different fragments. In Figure 2, we report the four most relevant FLMOs contributing to E_{ex} for each studied pair. Note that the computed FLMOs are characterized by a localization spread comparable to what has been reported by the authors in a recent paper.[39] As expected, the top and middle panels show that the interaction between the two aromatic rings ($frag_1$ and $frag_3$) and both DNA helices is dominated by $\pi-\pi$ stacking. Also, the aforementioned geometrical symmetry is displayed by the FLMOs that dominate the interactions, which are almost identical for $frag_1$-$helix_1$ and $frag_3$-$helix_2$. A different situation is depicted for $frag_2$ for which a direct hydrogen-bonding is clearly evident in all selected FLMOs. Such a qualitative analysis provides an intuitive theoretical explanation for the almost 2-fold value of the total interaction energy obtained for the lateral chain as compared to the two aromatic rings of the drug.

The second studied system is the all-trans retinal chromophore embedded in anabaena sensory rhodopsin (ASR), a light sensor, as reported in ref 33, by Olivucci and co-workers. The system is particularly interesting because it is the chromophore responsible for the fluorescence activity of the considered rhodopsin, thus permitting its potential use in the field of optogenetics via a modification of some peptides.[33] In this work, we showcase our novel approach by selecting retinal and the closest 14 amino acids in its wild type form, which are described at the HF/cc-pVDZ level (total N_{AO} = 2993, average t_{SCF} = 491 s; see Figure 3a). Differently from the previous case,

https://doi.org/10.1021/acs.jctc.2c00359
J. Chem. Theory Comput. 2022, 18, 4806−4813

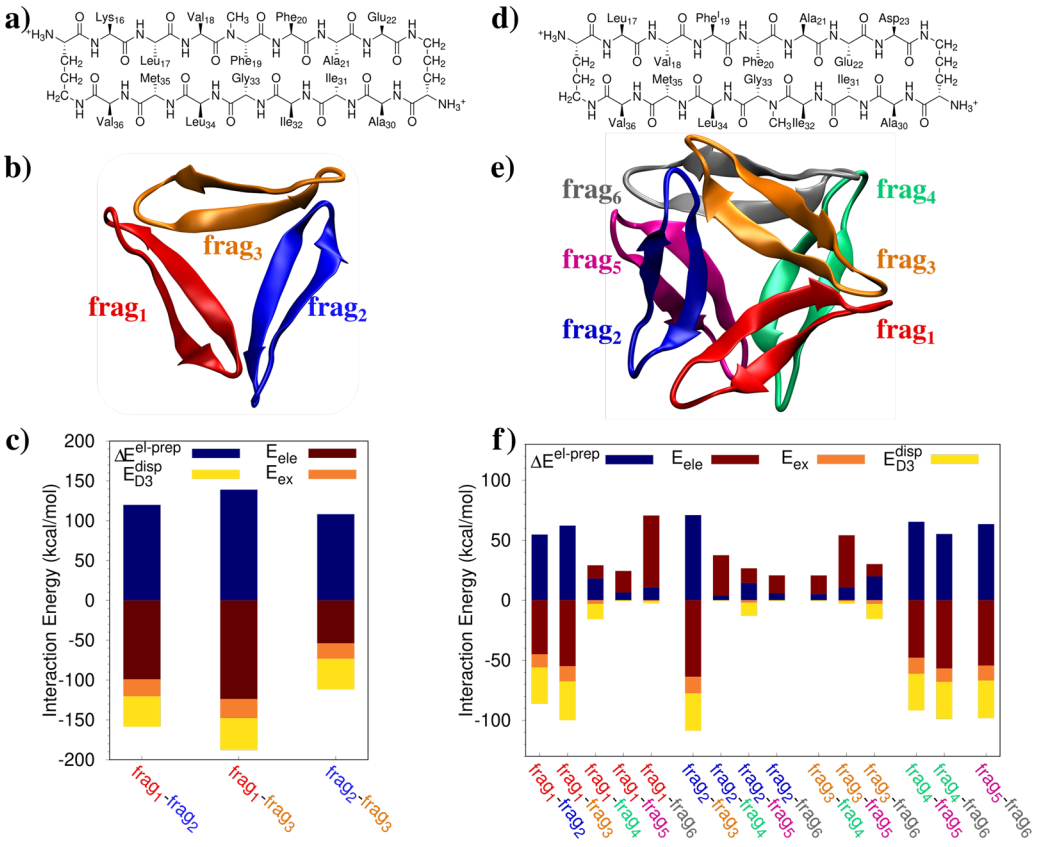


Figure 4. Molecular structures of the selected amyloid peptide prototypes (a−d) and their three-dimensional arrangements (b−e). FLMOs interaction energy decomposition analysis (parts c−f, kcal/mol).

we consider closed shell fragments, i.e., if two amino acids are covalently bonded, they are treated as a single fragment, resulting in a total of 12 fragments. In this way, we are also demonstrating the flexibility of our approach. After localizing the FLMOs, we calculate and decompose the interaction energy between retinal and the different fragments (see Figure 3 and Table S2 in the Supporting Information for raw data). Also in this case, the electrostatic interaction is attractive for all the considered fragments; however, the magnitude of the interaction is much lower as compared to that obtained in case of XR5944 and the DNA. In order to further investigate such a difference, the most relevant FLMOs interactions for each retinal−amino acid pair are graphically depicted in Figure 3c. Clearly, in any pair, no hydrogen bondings are present. However, a $\pi-\pi$ interaction is only depicted in the case of TYR_{179} + PRO_{180} fragment, in which the aromatic ring of TYR is stacked together with the double bonds of retinal. This is the reason why for such a fragment, the interaction energy components are larger than all the other fragments. Interestingly for almost all fragments, the exchange-repulsion term exceeds the electrostatics; however, the interaction energy is always attractive (except for THS_{80}). Therefore, in this system, dispersive contributions play a leading role. Finally, we notice that a large interaction energy is reported also for TRP_{176}, for which E^{ele} exceeds $E_{ex\text{-}rep}$ (in absolute value). By inspecting Figure 3c, this is probably due to the favorable electrostatic interaction between the peptide C═O group and the retinal.

3.2. Prototypes of Amyloid Diseases. As a final test case, we study two different prototypes of the amyloid diseases, such as Alzheimer, Parkinson, and type 2 diabetes.[34] In particular such diseases are historically related to the formation of plaques and fibrils caused by β-amyloid peptide Aβ and recently to the formation of small aggregates, called oligomers.[34] Special efforts on the last direction have been paid by the Nowick group[34] that has designed different β-sheets to mimic the β-hairpins typical of amyloid peptides. In this work, we select a trimer assembly (see Figure 4a−c, PDB 5W4H), which is seen as a unprecedented mode of self-assembly,[50] and an hexamer oligomer (see Figure 4d−f, PDB 5V65), which is one of the possible higher-order assemblies of the trimer unit.[51] We aim at calculating the main energy components defining the interaction of the considered systems, in order to provide a first theoretical glimpse of the formation of such oligomers. The study of the interactions between the macrocycles may be of particular importance to understand the leading force that bring them to agglomerate, a possible cause of the aforementioned diseases.

To this end, each oligomer is treated as a separate fragment, and the interaction between the different groups is calculated after the FLMOs are obtained by exploiting our presented procedure. In both the trimer and hexamer assemblies, the total system is described at the HF/6-311G* level (trimer: N_{AO} = 8079 (average t_{SCF} = 7158 s); hexamer: N_{AO} = 16224, average t_{SCF} = 21086 s). We first focus our attention on the trimer case, in which the three peptides are constituted by two $A\beta_{16-22}$ and $A\beta_{30-36}$ β-sheets, which are bonded by two $^{\delta}$Orn units. Notice that an N-methyl group is placed on Phe_{19}.[50] The trimer packaging is depicted in Figure 4b, and the decomposition of the interaction energy between the three monomers is graphically depicted in Figure 4c (see Table S3 in the Supporting Information for raw data). The electrostatic interaction is always attractive and for $frag_1$-$frag_2$ cancels out with the exchange-repulsion energy term. In all three cases, the total

https://doi.org/10.1021/acs.jctc.2c00359
J. Chem. Theory Comput. 2022, 18, 4806−4813

interaction energy is attractive because of the presence of the dispersion energy term, whose contribution ranges from 32% to 72% with respect to electrostatics. Therefore, intermolecular hydrogen bonds between the monomers play a crucial role; however, its covalent contribution, which is contained in the exchange-repulsion term, makes dispersion interactions crucial for the packaging of the system. Notice that our analysis thus suggests that the packaging is energetically favored in addition to hydrophobic forces[52] that are surely in play for such a system.

Finally, the hexamer assembly obtained by the packaging of the peptide depicted in Figure 4d is considered.[51] Hexamers, and generally high-order oligomers, are formed by trimers which assemble together to create hydrophobic pockets. In particular, the peptide monomer is constituted by two $A\beta_{17-23}$ and $A\beta_{30-36}$ β-sheets which, again, are linked by two $^{\delta}$Orn units. In this case, Phe_{19} is substituted with a iodine atom. Similarly to the previous case, we study the interaction between the six fragments by means of the different energy components (see Figure 4f and Table S4 in the Supporting Information for raw data). First, we notice that the involved energies are generally lower than the trimer case. This is the result of the larger distance between the fragments in the hexamer arrangement. Figure 4f clearly shows that the electrostatic interaction between the fragments is attractive for the considered fragments being part of the same trimer ($frag_{1-3}$ and $frag_{4-5}$), whereas the opposite holds for the fragments belonging to the different trimers. Indeed, the total interaction between the two trimers is repulsive because no hydrogen bonds can energetically stabilize the structure, whereas the interaction between the peptides constituting the trimer is attractive. Therefore, our analysis shows that the first formation of the trimers is also energetic-guided; however, higher-order oligomers only appear because of hydrophobic forces.

4. CONCLUSIONS

In summary, we have introduced the concept of FLMOs, which are the result of a variational optimization of the energetics of different fragments defining a molecular system, either covalently bonded or composed by different moieties. The FLMOs can be exploited for a plethora of applications, ranging from reactive chemistry and spectroscopy, as for instance, by reducing the computational cost associated with highly correlated methods. In this paper, we showcase FLMOs potential by proposing a simple energy decomposition analysis based on FLMOs, which is applied to systems of significant chemical and biological interest. In particular, we have demonstrated that many anticancer drugs intercalate within DNA thanks to the lateral chain which acts as an anchor through hydrogen bonding interactions. Then, we have studied all-trans retinal embedded in a rhodopsin protein, showing that $\pi-\pi$ interactions play the most relevant role in the interaction between the two systems and likely determine the molecular response of the chromophore. Finally, we have investigated two oligomers, prototypes of amyloid diseases. In particular, we have shown that the aggregation of the trimer form, which is the basis for higher-order oligomers, is energetically favored. On the contrary, the coupling between different trimers takes place because of hydrophobic forces. Within our theoretical approach, we have showed how FLMOs can be used in practical applications and can aid further design and deeper understanding of complex phenomena, such as the formations of plaques in brain diseases.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.2c00359.

Computational protocol, FLMOs energy decomposition, and raw data for Figures 1−4 (PDF)

AUTHOR INFORMATION

Corresponding Authors

Tommaso Giovannini − *Scuola Normale Superiore, 56126 Pisa, Italy;* orcid.org/0000-0002-5637-2853; Email: tommaso.giovannini@sns.it

Henrik Koch − *Scuola Normale Superiore, 56126 Pisa, Italy; Department of Chemistry, Norwegian University of Science and Technology, 7491 Trondheim, Norway;* orcid.org/0000-0002-8367-8727; Email: henrik.koch@sns.it

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.2c00359

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We acknowledge funding from the Marie Sklodowska-Curie European Training Network "COSINE Computational Spectroscopy in Natural Sciences and Engineering", Grant Agreement No. 765739 and the Research Council of Norway through FRINATEK Projects 263110 and 275506. We acknowledge Chiara Cappelli (Scuola Normale Superiore, SNS) for computer resources and the Center for High Performance Computing (CHPC) at SNS for providing the computational infrastructure.

REFERENCES

(1) Stone, A. *The Theory of Intermolecular Forces*; OUP Oxford, 2013.

(2) Leckband, D.; Israelachvili, J. Intermolecular forces in biology. *Q. Rev. Biophys.* **2001**, *34*, 105−267.

(3) Moy, V. T.; Florin, E.-L.; Gaub, H. E. Intermolecular forces and energies between ligands and receptors. *Science* **1994**, *266*, 257−259.

(4) Krenske, E. H.; Houk, K. Aromatic interactions as control elements in stereoselective organic reactions. *Accounts of chemical research* **2013**, *46*, 979−989.

(5) Goedecker, S. Linear scaling electronic structure methods. *Rev. Mod. Phys.* **1999**, *71*, 1085.

(6) Helgaker, T.; Jorgensen, P.; Olsen, J. *Molecular Electronic-Structure Theory*; John Wiley & Sons, 2014.

(7) Aquilante, F.; Boman, L.; Boström, J.; Koch, H.; Lindh, R.; de Merás, A. S.; Pedersen, T. B. *Linear-Scaling Techniques in Computational Chemistry and Physics*; Springer, 2011; pp 301−343.

(8) Koch, H.; Sánchez de Merás, A.; Pedersen, T. B. Reduced scaling in electronic structure calculations using Cholesky decompositions. *J. Chem. Phys.* **2003**, *118*, 9481−9484.

(9) Pedreira, J. G.; Franco, L. S.; Barreiro, E. J. Chemical Intuition in Drug Design and Discovery. *Current topics in medicinal chemistry* **2019**, *19*, 1679−1693.

(10) Mennucci, B.; Corni, S. Multiscale modelling of photoinduced processes in composite systems. *Nat. Rev. Chem.* **2019**, *3*, 315−330.

(11) Ma, Q.; Werner, H.-J. Explicitly correlated local coupled-cluster methods using pair natural orbitals. *WIREs Comput. Mol. Sci.* **2018**, *8*, No. e1371.

(12) Giovannini, T.; Egidi, F.; Cappelli, C. Molecular spectroscopy of aqueous solutions: a theoretical perspective. *Chem. Soc. Rev.* **2020**, *49*, 5664−5677.

(13) Hampel, C.; Werner, H.-J. Local treatment of electron correlation in coupled cluster theory. *J. Chem. Phys.* **1996**, *104*, 6286−6297.
(14) Riplinger, C.; Sandhoefer, B.; Hansen, A.; Neese, F. Natural triple excitations in local coupled cluster calculations with pair natural orbitals. *J. Chem. Phys.* **2013**, *139*, 134101.
(15) Parrish, R. M.; Sherrill, C. D. Quantum-mechanical evaluation of $\pi-\pi$ versus substituent- π interactions in π stacking: direct evidence for the Wheeler−Houk picture. *J. Am. Chem. Soc.* **2014**, *136*, 17386−17389.
(16) Saebo, S.; Pulay, P. Local treatment of electron correlation. *Annu. Rev. Phys. Chem.* **1993**, *44*, 213−236.
(17) Boughton, J. W.; Pulay, P. Comparison of the boys and Pipek−Mezey localizations in the local correlation approach and automatic virtual basis selection. *J. Comput. Chem.* **1993**, *14*, 736−740.
(18) Boys, S. F. Construction of some molecular orbitals to be approximately invariant for changes from one molecule to another. *Rev. Mod. Phys.* **1960**, *32*, 296.
(19) Edmiston, C.; Ruedenberg, K. Localized atomic and molecular orbitals. *Rev. Mod. Phys.* **1963**, *35*, 457.
(20) Sironi, M.; Genoni, A.; Civera, M.; Pieraccini, S.; Ghitti, M. Extremely localized molecular orbitals: theory and applications. *Theor. Chem. Acc.* **2007**, *117*, 685−698.
(21) Aquilante, F.; Bondo Pedersen, T.; Sánchez de Merás, A.; Koch, H. Fast noniterative orbital localization for large molecules. *J. Chem. Phys.* **2006**, *125*, 174101.
(22) Høyvik, I.-M.; Jansik, B.; Jørgensen, P. Orbital localization using fourth central moment minimization. *J. Chem. Phys.* **2012**, *137*, 224114.
(23) Khaliullin, R. Z.; Cobar, E. A.; Lochan, R. C.; Bell, A. T.; Head-Gordon, M. Unravelling the origin of intermolecular interactions using absolutely localized molecular orbitals. *J. Phys. Chem. A* **2007**, *111*, 8753−8765.
(24) Sayfutyarova, E. R.; Sun, Q.; Chan, G. K.-L.; Knizia, G. Automated construction of molecular active spaces from atomic valence orbitals. *J. Chem. Theory Comput.* **2017**, *13*, 4063−4078.
(25) Gordon, M. S.; Fedorov, D. G.; Pruitt, S. R.; Slipchenko, L. V. Fragmentation methods: A route to accurate calculations on large systems. *Chem. Rev.* **2012**, *112*, 632−672.
(26) Li, Z.; Li, H.; Suo, B.; Liu, W. Localization of molecular orbitals: from fragments to molecule. *Acc. Chem. Res.* **2014**, *47*, 2758−2767.
(27) Li, H.; Liu, W.; Suo, B. Localization of open-shell molecular orbitals via least change from fragments to molecule. *J. Chem. Phys.* **2017**, *146*, 104104.
(28) Thapa, B.; Beckett, D.; Jovan Jose, K.; Raghavachari, K. Assessment of fragmentation strategies for large proteins using the multilayer molecules-in-molecules approach. *J. Chem. Theory Comput.* **2018**, *14*, 1383−1394.
(29) Mayhall, N. J.; Raghavachari, K. Molecules-in-molecules: An extrapolated fragment-based approach for accurate calculations on large molecules and materials. *J. Chem. Theory Comput.* **2011**, *7*, 1336−1343.
(30) Thapa, B.; Raghavachari, K. Energy Decomposition Analysis of Protein−Ligand Interactions Using Molecules-in-Molecules Fragmentation-Based Method. *J. Chem. Inf. Model.* **2019**, *59*, 3474−3484.
(31) Wu, F.; Liu, W.; Zhang, Y.; Li, Z. Linear-scaling time-dependent density functional theory based on the idea of "from fragments to molecule. *J. Chem. Theory Comput.* **2011**, *7*, 3643−3660.
(32) Dai, J.; Punchihewa, C.; Mistry, P.; Ooi, A. T.; Yang, D. Novel DNA bis-intercalation by MLN944, a potent clinical bisphenazine anticancer drug. *J. Biol. Chem.* **2004**, *279*, 46096−46103.
(33) Marin, M. d. C.; Agathangelou, D.; Orozco-Gonzalez, Y.; Valentini, A.; Kato, Y.; Abe- Yoshizumi, R.; Kandori, H.; Choi, A.; Jung, K.-H.; Haacke, S.; et al. Fluorescence enhancement of a microbial rhodopsin via electronic reprogramming. *J. Am. Chem. Soc.* **2019**, *141*, 262−271.
(34) Kreutzer, A. G.; Nowick, J. S. Elucidating the structures of amyloid oligomers with macrocyclic β-hairpin peptides: insights into Alzheimer's disease and other amyloid diseases. *Accounts of chemical research* **2018**, *51*, 706−718.
(35) Cheng, P.-N.; Liu, C.; Zhao, M.; Eisenberg, D.; Nowick, J. S. Amyloid β-sheet mimics that antagonize protein aggregation and reduce amyloid toxicity. *Nature Chem.* **2012**, *4*, 927−933.
(36) Cheng, P.-N.; Pham, J. D.; Nowick, J. S. The supramolecular chemistry of β-sheets. *J. Am. Chem. Soc.* **2013**, *135*, 5477−5492.
(37) McWeeny, R. *Methods of Molecular Quantum Mechanics*; Academic Press: London, 1992.
(38) Goletto, L.; Kjønstad, E. F.; Folkestad, S. D.; Høyvik, I.-M.; Koch, H. Linear-Scaling Implementation of Multilevel Hartree−Fock Theory. *J. Chem. Theory Comput.* **2021**, *17*, 7416−7427.
(39) Giovannini, T.; Koch, H. Energy-based molecular orbital localization in a specific spatial region. *J. Chem. Theory Comput.* **2021**, *17*, 139−150.
(40) Schneider, W. B.; Bistoni, G.; Sparta, M.; Saitow, M.; Riplinger, C.; Auer, A. A.; Neese, F. Decomposition of intermolecular interaction energies within the local pair natural orbital coupled cluster framework. *J. Chem. Theory Comput.* **2016**, *12*, 4778−4792.
(41) Boys, S. F.; Bernardi, F. d. The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol. Phys.* **1970**, *19*, 553−566.
(42) Folkestad, S. D.; Kjønstad, E. F.; Myhre, R. H.; Andersen, J. H.; Balbi, A.; Coriani, S.; Giovannini, T.; Goletto, L.; Haugland, T. S.; Hutcheson, A.; Høyvik, I.-M.; Moitra, T.; Paul, A. C.; Scavino, M.; Skeidsvoll, A. S.; Tveten, Åsmund H.; Koch, H.; et al. $e^T1.0$: An open source electronic structure program with emphasis on coupled cluster and multilevel methods. *J. Chem. Phys.* **2020**, *152*, 184103.
(43) Grimme, S.; Antony, J.; Ehrlich, S.; Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **2010**, *132*, 154104.
(44) Stewart, A. J.; Mistry, P.; Dangerfield, W.; Bootle, D.; Baker, M.; Kofler, B.; Okiji, S.; Baguley, B. C.; Denny, W. A.; Charlton, P. A. Antitumor activity of XR5944, a novel and potent topoisomerase poison. *Anti-Cancer Drugs* **2001**, *12*, 359−367.
(45) Harris, S.; Mistry, P.; Freathy, C.; Brown, J.; Charlton, P. Antitumour activity of XR5944 in vitro and in vivo in combination with 5-fluorouracil and irinotecan in colon cancer cell lines. *British journal of cancer* **2005**, *92*, 722−728.
(46) Punchihewa, C.; De Alba, A.; Sidell, N.; Yang, D. XR5944: A potent inhibitor of estrogen receptors. *Molecular cancer therapeutics* **2007**, *6*, 213−219.
(47) Weiss, R. B. The anthracyclines: will we ever find a better doxorubicin? *Semin. Oncol.* **1992**, 670−686.
(48) Arcamone, F. *Doxorubicin: Anticancer Antibiotics*; Elsevier, 2012.
(49) Altun, A.; Izsák, R.; Bistoni, G. Local energy decomposition of coupled-cluster interaction energies: Interpretation, benchmarks, and comparison with symmetry-adapted perturbation theory. *Int. J. Quantum Chem.* **2021**, *121*, No. e26339.
(50) Kreutzer, A. G.; Spencer, R. K.; McKnelly, K. J.; Yoo, S.; Hamza, I. L.; Salveson, P. J.; Nowick, J. S. A Hexamer of a Peptide Derived from Aβ16−36. *Biochemistry* **2017**, *56*, 6061−6071.
(51) Salveson, P. J.; Spencer, R. K.; Kreutzer, A. G.; Nowick, J. S. X-ray Crystallographic Structure of a Compact Dodecamer from a Peptide Derived from Aβ16−36. *Org. Lett.* **2017**, *19*, 3462−3465.
(52) Wiggins, P. M. Hydrophobic hydration, hydrophobic forces and protein folding. *Physica A: Statistical Mechanics and its Applications* **1997**, *238*, 113−128.

https://doi.org/10.1021/acs.jctc.2c00359
J. Chem. Theory Comput. 2022, 18, 4806−4813