Supporting Information: Fragment Localized Molecular Orbitals

Tommaso Giovannini*† and Henrik Koch*†‡

†Scuola Normale Superiore, Piazza dei Cavalieri 7, 56126 Pisa, Italy.
‡Department of Chemistry, Norwegian University of Science and Technology, 7491 Trondheim, Norway

E-mail: tommaso.giovannini@sns.it; henrik.koch@sns.it
S1 Computational Protocol

S1.1 Non-Covalently Bonded Fragments

Let us consider the three fragment scheme in Fig. S1. Let us divide the system in the parts: $A$ and $B + C$. Then, the $A$ density can be localized by following MLHF-AB scheme. In particular, the sum of the energies of the two parts reads:

\[
E^A + E^{BC} = \text{Tr} \ h^A D^A + \text{Tr} \ h^{BC} D^{BC} \\
+ \frac{1}{2} \text{Tr} \ D^A G(D^A) + \frac{1}{2} \text{Tr} \ D^{BC} G(D^{BC}) \tag{S1}
\]

where, $BC = B + C$, and the same holds for the corresponding density matrices. By using $D = D^A + D^{BC}$ then, eq. S1 becomes:

\[
E^A + E^{BC} = \text{Tr} \ (V^A - V^{BC}) D^A + \text{Tr} \ D^A G(D^A) - \text{Tr} \ D^A G(D) \\
+ \text{Tr} \ h^{BC} D + \frac{1}{2} \text{Tr} \ D G(D) \tag{S2}
\]

\[
F^A = V^A - V^{BC} + 2G(D^A) - 2G(D) \tag{S3}
\]

which is exactly the same formulation proposed for MLHF-AB. After minimizing eq. S3,
we obtain two densities for $A$ and $BC$ systems which are maximally localized. However, $B$ and $C$ are not maximally localized in their specific spatial regions. To this end, we can now reformulate the problem as a true three fragment scheme, in which, however, $D^A$ is kept frozen, and its orbitals do no enter in the occupied occupied rotations, which will be performed in the $BC$ space only. However, the effect of $D^A$ is taken into account by considering Eq. S4, which can be recasted in the following way:

$$E^A + E^B + E^C[D^B] = \text{Tr} \left( V^B - V^C \right)D^B + \text{Tr} \left( D^B G(D^B) - D^B G(D) \right) + \text{Tr} \left( D^B G(D^A) \right)$$

$$+ \left[ \text{Tr} \left( V^A - V^C \right)D^A + \text{Tr} \left( D^A G(D^A) - D^A G(D) \right) \right]$$

$$+ \text{Tr} [h^C D + \frac{1}{2} \text{Tr} DG(D)]$$

(S4)

from which the Fock matrix can be constructed:

$$F^B = (V^B - V^C) + 2G(D^B) - 2G(D) + 2G(D^A)$$

(S5)

S1.2 Covalently Bonded Fragments

Figure S2: Example of electron assignment in Hexane as partitioned in three fragments.

We now consider the case in which the system is a supramolecule in which the different fragments are covalently bonded, via a single covalent bond. As explained in the main text, our procedure is carried out for closed shell fragments. Therefore, the user needs to define
to which fragment the bonding electrons have to be assigned. As an example, we show how such a partitioning can be done for the Hexane molecule in Fig. S2. In this case, we assign the bonding electrons between frag1 and frag2 to frag1, and those between frag2 and frag3 to frag2. Notice that the procedure is completely independent of such an assignment (vide infra). Once such a decision is taken, the computational protocol exactly follows what has already been explained in case of the water trimer. At the end of the procedure, there will be three fragment densities characterized by the pre-defined number of electrons. More importantly, for the fragments to which the bonding electrons were initially assigned, we will obtain some FLMOs resembling the bonding MOs. We provide a practical example for an ethane moiety described at the HF/aug-cc-pVDZ level (see Fig. S3).

Once the bonding FLMOs are identified (such a step is very easy, since they are the highest in energy among the FLMOs of the considered fragment), the fragment densities of the two fragments covalently bonded can be updated by exploiting Eq. 5 in the main text (i.e. one electron is removed to one fragment density, and added to the other). Therefore, the procedure is independent of the initial electron assignment choice.

S1.3 N-fragment case

In general, at each cycle we are working in a reduced MO space, in which the occupied MOs of the previously optimized density are removed. Eq. S4 can be reformulated in case of N
densities, and for which we want to optimized the $i$-th density:

\[
\sum_{j}^{N} E^j[D^i] = \text{Tr} (V^i - V^{i+1\cdots N})D^i + \text{Tr} D^i G(D^i) - \text{Tr} D^i G(D) + \sum_{k<i}^{N} \text{Tr} D^i G(D^k)
\]

\[+ \sum_{k<i} \left[ \text{Tr} (V^k - V^{i+1\cdots N})D^k + \text{Tr} D^k G(D^k) - \text{Tr} D^k G(D) + \sum_{l<k} \text{Tr} D^k G(D^l) \right] \]

\[+ \text{Tr} h^{i+1\cdots N}D + \frac{1}{2} \text{Tr} DG(D) \]  \hspace{1cm} (S6)

from which the Fock matrix can be constructed:

\[
F^i = (V^i - V^{i+1\cdots N}) + 2G(D^i) - 2G(D) + 2 \sum_{k<i}^{N} G(D^k) \]  \hspace{1cm} (S7)

To summarize, the procedure is characterized by the following points:

1. Construction of the $G(D)$ and all the terms $\text{Tr} h^{j\cdots N}D = \text{Tr} (T + \sum_{k=2}^{N} V^k)D$. In this way, we have at once all the constant terms due to the total density $D$

2. Construction of one electron matrices at the $i$-th step.

3. From the previous cycle Fock matrix construct all the constant terms due to the $i-1$-th optimized density matrix.

4. Run the occupied-occupied optimization.

5. Readjust the Fragment Density according to Eq. 5 in the main text.

### S1.4 Interaction Energy Decomposition among the FLMOs

To generalize the energy decomposition analysis among the Fragment-Localized Molecular Orbitals (FLMOs), we consider two orbitals $\alpha$ and $\beta$ localized in fragment A and fragment
B, respectively. We assume that the two orbitals are doubly occupied; thus, two density matrices can be computed as:

\[ D_{\mu\nu}^\alpha = 2C_{\mu\alpha}C_{\nu\alpha} \] (S8)

\[ D_{\mu\nu}^\beta = 2C_{\mu\beta}C_{\nu\beta} \] (S9)

By following the derivation commented in the main text, the interaction between the two densities can therefore be written as:

\[
E_{\alpha\beta}^{\text{int,FLMO}} = \frac{\Tr V^\alpha D^\beta}{n_o^A} + \frac{\Tr V^\beta D^\alpha}{n_o^B} + \Tr D^\alpha J(D^\beta) + \frac{h_{\text{nuc}}^{\alpha\beta}}{n_o^A + n_o^B} - \Tr D^\alpha K(D^\beta) \] (S10)

\[
E_{\alpha\beta}^{\text{ele}} = \frac{\Tr V^\alpha D^\beta}{n_o^A} + \frac{\Tr V^\beta D^\alpha}{n_o^B} + \Tr D^\alpha J(D^\beta) + \frac{h_{\text{nuc}}^{\alpha\beta}}{n_o^A + n_o^B} \] (S11)

\[
E_{\alpha\beta}^{\text{ex}} = -\Tr D^\alpha K(D^\beta) \] (S12)

where, the factors \(n_o^A\) and \(n_o^B\) corresponds to the number of occupied MOs of fragments A and B, respectively. Such factors allows for the correct limit when all \(\alpha\) and \(\beta\) occupied MOs of the two fragments are summed.

### S1.5 Interaction Energies

**Dimer** The interaction energy for a generic dimer \(AB\), composed of two monomers \(A\) and \(B\), is defined as:

\[
E_{AB}^{\text{int}} = E_{AB} - E_A^{\text{vac}} - E_B^{\text{vac}} \] (S13)

where \(E_{AB}\) is the dimer energy, whereas \(E_A^{\text{vac}}\) and \(E_B^{\text{vac}}\) are the monomers energy in gas-phase (which can in principle be computed by including counterpoise correction). Note that, for
covalently bonded fragments, $E_{\text{vac}}^X$ is calculated at the Unrestricted HF level. Equivalently, Eq. S13 can be recasted in terms of FLMO related quantities (see Eqs. 3,7 in the main text) as:

\[ E_{\text{int}}^{AB} = E_{\text{int,FLMO}}^{AB} + E_{\text{emb}}^A + E_{\text{emb}}^B - E_{\text{vac}}^A - E_{\text{vac}}^B \]  
\[ = E_{\text{int,FLMO}}^{AB} + (E_{\text{emb}}^A - E_{\text{vac}}^A) + (E_{\text{emb}}^B - E_{\text{vac}}^B) \]  
\[ = E_{\text{int,FLMO}}^{AB} + E_{\text{el-prep}}^A + E_{\text{el-prep}}^B \]  

where $E_{AB}$ has been rewritten in terms of $A, B$ energies for the monomers embedded in the dimer ($E_{\text{emb}}^A + E_{\text{emb}}^B$), and their interaction energy calculated as reported in Eq. 6 in the main text. In Eq. S16, the electronic-preparation energies of the two monomers are introduced. Such an energetic term takes into account the cost to bring the two monomers from the gas-phase to the final electronic state of the dimer.

**Trimer** The interaction energy for a generic trimer $ABC$, composed of three monomers $A, B$ and $C$, is defined as:

\[ E_{\text{int}}^{ABC} = E_{ABC} - E_{\text{vac}}^A - E_{\text{vac}}^B - E_{\text{vac}}^C \]  

By using the same procedure exposed above (which is also valid for a general N fragment case), we obtain:

\[ E_{\text{int}}^{ABC} = E_{\text{int,FLMO}}^{ABC} + E_{\text{emb}}^A + E_{\text{emb}}^B + E_{\text{emb}}^C - E_{\text{vac}}^A - E_{\text{vac}}^B - E_{\text{vac}}^C \]  
\[ = E_{\text{int,FLMO}}^{ABC} + E_{\text{el-prep}}^A + E_{\text{el-prep}}^B + E_{\text{el-prep}}^C \]  

We can now introduce the three-body energy term ($\Delta E_{3B}^3$), which reads:

S7
\[ \Delta E^{3B} = E^{ABC} - E^A_{vac} - E^B_{vac} - E^C_{vac} - E^{int,vac}_{AB} - E^{int,vac}_{BC} - E^{int,vac}_{AC} \]  
\hspace{1cm} (S19)

Therefore, Eq. S17 can be rewritten as:

\[ E^{int}_{ABC} = E^{int,vac}_{AB} + E^{int,vac}_{BC} + E^{int,vac}_{AC} + \Delta E^{3B} \]  
\hspace{1cm} (S20)

By then considering that \( E^{int,FLMO}_{ABC} \) is additive in the interaction energies between the dimers (see main text), we can compare Eqs. S18 and S20, which yields:

\[ E^{int,FLMO}_{AB} + E^{int,FLMO}_{BC} + E^{int,FLMO}_{AC} + E^{el-prep}_A + E^{el-prep}_B + E^{el-prep}_C = E^{int,vac}_{AB} + E^{int,vac}_{BC} + E^{int,vac}_{AC} + \Delta E^{3B} \]  
\hspace{1cm} (S21)

and therefore, an expression for the three-body energy term (\( \Delta E^{3B} \)) can be obtained:

\[ \Delta E^{3B} = E^{el-prep}_A + E^{el-prep}_B + E^{el-prep}_C + E^{int,prep}_{AB} + E^{int,prep}_{BC} + E^{int,prep}_{AC} \]  
\hspace{1cm} (S22)

\[ E^{int,prep}_{XY} = E^{int,FLMO}_{XY} - E^{int,vac}_{XY} \]  
\hspace{1cm} (S23)

We now move to calculate the interaction energy between two subsystems X and Y, which is defined as:

\[ E^{int}_{XY} = E^v_{XY} - E^v_X - E^v_Y \]  
\hspace{1cm} (S24)

Equivalently, it can be computed by using FLMO-based quantities as:
\[ E_{\text{int}}^{XY} = E_{\text{int,FLMO}}^{XY} + E_{\text{el-prep}}^{XY} \]  
\[ E_{\text{el-prep}}^{X,Y} = E_{\text{el-prep}}^{X} + E_{\text{el-prep}}^{Y} - E_{XY}^{el-prep} \]  
\[ E_{XY}^{el-prep} = E_{XY}^{emb} - E_{XY}^{vac} \]

where \( E_{XY}^{el-prep} \) is the electronic preparation energy of the dimer. By using S25, the interaction energy between two dimers can be decomposed in terms of FLMO-based quantities.

It is worth noting that in case the three fragments are covalently bonded, e.g. the hexane molecule depicted in Fig. S2, the gas-phase energies of the different fragments are computed at the UHF level, by specifying the correct multiplicity. On the other hand, \( E_{X}^{emb} \) are calculated by following the procedure described in Sec. S1.2 and by exploiting Eq. 5 in the main text.
## S2 Results

Table S1: Interaction energy decomposition analysis (kcal/mol) between the three fragments constituting XR5944 and the two helix of the DNA.

| Fragments  | $E_{ele}$ | $E_{ex}$ | $\Delta E^{el-prep}$ | $E_{disp}$ | $E_{int}$ |
|------------|-----------|----------|----------------------|------------|-----------|
| frag1-helix1 | -68.6     | -19.2    | 75.7                 | -28.9      | -41.0     |
| frag1-helix2 | -66.9     | -19.7    | 81.0                 | -24.5      | -30.1     |
| frag2-helix1 | -139.4    | -9.9     | 72.0                 | -10.6      | -87.8     |
| frag2-helix2 | -136.8    | -9.1     | 68.5                 | -10.3      | -87.6     |
| frag3-helix1 | -68.8     | -19.8    | 82.7                 | -25.0      | -30.9     |
| frag3-helix2 | -66.9     | -18.6    | 74.2                 | -28.7      | -39.9     |

Table S2: Interaction energy decomposition analysis (kcal/mol) between all-trans retinal and the closest 14 aminoacids of the Anabaena Sensory Rhodopsin.

| Fragments  | $E_{ele}$ | $E_{ex}$ | $\Delta E^{el-prep}$ | $E_{disp}$ | $E_{int}$ |
|------------|-----------|----------|----------------------|------------|-----------|
| GLY135+VAL136 | -15.2     | -8.1     | 31.0                 | -8.7       | -1.1      |
| TYR132     | -14.8     | -6.1     | 22.4                 | -7.6       | -6.1      |
| GLY116     | -5.4      | -3.3     | 12.3                 | -3.8       | -0.3      |
| TYR179+PRO180 | -30.6     | -16.8    | 57.8                 | -22.2      | -11.8     |
| THR80      | -5.3      | -3.7     | 14.7                 | -4.7       | 0.9       |
| VAL112     | -5.0      | -2.5     | 8.3                  | -5.6       | -4.9      |
| TRP76      | -17.8     | -7.0     | 28.0                 | -10.2      | -7.0      |
| PHE139     | -3.9      | -1.4     | 4.9                  | -4.3       | -4.7      |
| TRP183     | -6.7      | -3.6     | 13.4                 | -5.9       | -2.7      |
| TRP176     | -10.5     | -2.6     | 9.2                  | -6.3       | -10.1     |
| SER209     | -5.4      | -2.8     | 11.6                 | -3.6       | -0.2      |
| LEU83      | -1.0      | -0.7     | 3.0                  | -2.6       | -1.3      |
Table S3: Interaction energy decomposition analysis (kcal/mol) between the three peptides constituting the trimer assembly, prototype of amyloid peptide (see fig. 4 a,b,c).

| Fragments   | $E_{ele}$ | $E_{ex}$ | $\Delta E_{el-prep}$ | $E_{D3}^{disp}$ | $E_{int}$ |
|-------------|-----------|----------|------------------------|-----------------|-----------|
| frag1-frag2 | -99.2     | -21.4    | 119.8                  | -38.0           | -38.7     |
| frag1-frag3 | -123.8    | -24.2    | 138.9                  | -40.0           | -49.0     |
| frag2-frag3 | -53.8     | -19.5    | 108.1                  | -38.3           | -3.5      |

Table S4: Interaction energy decomposition analysis (kcal/mol) between the six peptides constituting the hexamer assembly, prototype of amyloid peptide (see fig. 4 d,e,f)

| Fragments   | $E_{ele}$ | $E_{ex}$ | $\Delta E_{el-prep}$ | $E_{D3}^{disp}$ | $E_{int}$ |
|-------------|-----------|----------|------------------------|-----------------|-----------|
| frag1-frag2 | -45.0     | -11.1    | 54.7                   | -30.0           | -31.3     |
| frag1-frag3 | -55.0     | -12.6    | 62.2                   | -32.3           | -37.7     |
| frag1-frag4 | 11.0      | -3.1     | 18.2                   | -12.7           | 13.4      |
| frag1-frag5 | 17.9      | 0.0      | 6.5                    | -0.6            | 23.8      |
| frag1-frag6 | 59.6      | -0.8     | 10.9                   | -1.7            | 68.0      |
| frag2-frag3 | -63.8     | -13.8    | 70.9                   | -31.0           | -37.6     |
| frag2-frag4 | 33.9      | 0.0      | 3.7                    | -0.3            | 37.2      |
| frag2-frag5 | 12.2      | -2.1     | 14.4                   | -10.8           | 13.6      |
| frag2-frag6 | 14.9      | 0.0      | 5.8                    | -0.5            | 20.2      |
| frag3-frag4 | 15.5      | 0.0      | 5.1                    | -0.5            | 20.2      |
| frag3-frag5 | 43.4      | -0.9     | 10.8                   | -2.0            | 51.3      |
| frag3-frag6 | 10.2      | -3.2     | 19.9                   | -12.3           | 14.6      |
| frag4-frag5 | -48.0     | -13.4    | 65.4                   | -30.4           | -26.4     |
| frag4-frag6 | -56.9     | -11.2    | 55.2                   | -30.8           | -43.6     |
| frag5-frag6 | -54.4     | -12.5    | 63.4                   | -31.3           | -34.8     |