Automatic Drawing of Orbital Correlation Diagrams. A Computational Tool for Electronic-Structure Informatics

Manabu Sugimoto* and Takafumi Inoue

Faculty of Advanced Science and Technology and Graduate School of Science and Technology, Kumamoto University, 2-39-1 Kurokami, Chuo-ku, Kumamoto 860-8555, Japan

(Received June 5, 2019; Accepted October 15, 2019)

Finding direct correlations between electronic structures of molecules and their properties, which we call “electronic-structure informatics”, is one of the challenging issues in chemoinformatics because the electronic degree of freedom is an essential factor determining the chemical characteristics. Herein we develop computational methods to automatically draw two types of orbital correlation diagrams. They are expected useful to perform machine learning including electronic degrees of freedom. In the present approach, we focus on electronic similarity called orbital similarity whose score is defined as spatial overlap between two molecular orbitals (MOs) enclosed with their iso-value surfaces. The similarity scores are also used to derive another orbital correlation diagram called “orbital interaction diagram”. This diagram is to relate MOs of a target molecule with those of its fragments. Through applications to benzene derivatives, these diagrams are shown to be reasonable, indicating potential usefulness of the present method in machine learning for quantitative predictions of molecular properties and chemical reactivities.

Key Words: Electronic-Structure Informatics, Machine Learning, Molecular Orbital, Electronic Similarity, Orbital Similarity, Orbital Correlation Diagram, Orbital Interaction Diagram

* sugimoto@kumamoto-u.ac.jp
1. Introduction

Numerical data obtained in electronic-structure calculations provide rich information to understand, suggest structure-property relationships of functional molecules [1]. After the frontier orbital theory by Fukui [2] and the Woodward-Hoffmann (WH) rule [3], the frontier orbitals such as the HOMO (highest occupied molecular orbital) and the LUMO (lowest unoccupied molecular orbital) have been recognized as significant characteristics determining chemical reactivities [4-6] and molecular properties [7].

In discussing electronic structures of molecules, not only energy levels but also characters of molecular orbitals (MOs) are important although the former is mainly considered in the chemoinformatics studies including quantitative structure-activity relationships [8]. When we try to use energy levels of various MOs as explanatory variables (descriptors) for supervised machine learning, it is considered physically inappropriate to correlate MOs with different characters only by their ordinal numbers or labels such as HOMO, HOMO-1, etc. that reflect only relative position in MO energy levels. Without considering similarity and dissimilarity of orbital characters, the data science would lose its theoretical justification due to the lack of consistency in defining the descriptors.

Constructions of orbital correlation diagrams provide remarkable information of molecules. In quantum chemistry studies, three types of the diagrams as shown in Figure 1 have been frequently used for theoretical analyses and interpretations: the Walsh diagram [9], orbital correlation diagram [10, 11], and orbital interaction diagram [12]. These diagrams are useful to know responding MOs to structural changes or replacement of partial structures (e.g. a substituent group and important components of MOs.

In preparation to draw these diagrams, we must characterize the MOs. Then the lines are drawn to interconnect “corresponding MOs” having similar characters. For the assignment, it is possible to numerically investigate main components of MOs, their symmetries if available, and their nodal structures by examining MO coefficients in the basis-set expansion used in electronic-structure calculations. An alternative method is to observe topological shapes of MOs in their iso-value plots which can be called “orbital shape”. Any of these assignment criteria typically are done visually by a researcher (human). This is an obstacle in machine learning using information on MOs. For automating MO correlation diagrams, it is demanded to develop a computational method to recognize orbital shapes and quantify correspondence (similarity) among MOs of various molecules.

In this paper, we aimed at developing computer programs that facilitate automatic drawing of orbital correlation diagrams of the types shown in Figure 1 (b) and (c). For this purpose, MO-shape similarity was numerically calculated by the newly developed computer code. Herein, substituted benzene derivatives were chosen for application.

![Figure 1. Orbital correlation diagrams](image)
2. Computational Methods

2.1 Electronic-Structure Calculations

MOs and their energy levels at the optimized geometries of molecules were calculated by using the DFT method where the M06-2X hybrid functional [13] was applied with the 6-31G** basis set [14]. The present calculations were performed using the Gaussian09 software [15]. The “cubegen” program provided with Gaussian09 by Gaussian Inc. was implemented to evaluate the amplitudes of MOs at each point of the three-dimensional lattice covering a molecule at the optimized geometry. The iso-value plots of MOs were obtained with the criteria of 0.02 in absolute value for the MO amplitude. The calculated MOs are summarized in Tables S1 and S2 in Supplementary Materials.

2.2 Orbital correlation diagram

An orbital correlation diagram among the substituted benzene derivatives was created by calculating orbital similarity scores (OSS) by referring to the MOs of the phenyl radical. The method of calculating OSS will be described in detail in Sec. 2.4. The calculated OSS are shown in Table S3. One of the MOs of the phenyl radical whose similarity score with a target MO was highest was chosen as the corresponding MO. This rule is hereafter called the “largest OSS rule”. The correspondence table following this rule was created as shown in Table 1. No correspondence was assigned when the OSS is less than 35 %. This is to consider possibility of having substituent-group localized MOs.

With application to machine learning in mind, here we intended to suggest one-to-one correspondence on each MO. For this purpose, the orbital correspondence was established by defining a priority in choice, which can be called “first-encounter-first-assigned basis” in the frontier MO region (i.e. MOs near HOMO and LUMO). For example, when the two occupied (unoccupied) MOs of aniline are commonly correlated with one MO of benzene in Table 1, and the energetically higher (lower) MO of aniline is chosen as a corresponding orbital to that of benzene.

2.3 Orbital interaction diagram

An orbital interaction diagram, which represent correlation of MOs of a target molecule with those of its fragments, were constructed in the following procedure. For a compound investigated in this paper, we decompose it into the C₆H₆ radical and the substituent group. Then similarity score of one MO of a target molecule with MOs of each fragments was evaluated to form a matrix of similarity scores among the MOs of the target and reference molecules. The results are summarized in Tables S4 and S5 for phenol and aniline, respectively. For each target MO, the most similar MO of the reference molecule is chosen by finding the highest similarity score (“highest similarity score rule”). When the score is less than 35 and 10 % for the C₆H₆ moiety and the substituent groups, respectively, the score is ignored, i.e. the value is set to 0 %. This is hereafter referred to as the “35/10 %

Table 1. Assignment of the MOs of the benzene derivatives to those of phenyl radical (C₆H₆). A label in a cell in square bracket is the corresponding MO of C₆H₆ where H and L indicate HOMO and LUMO, respectively. A label in a parenthesis indicates the most similar (highest OSS) MO of C₆H₆. “N” means that no corresponding MO is assigned.

| MO label | target molecule |
|----------|-----------------|
|          | aniline | phenol | benzene | benzoic acid | benzonitrile |
| LUMO+2   | [L+1]   | [L+1]  | [L+1]   | N ([L+1])    | N ([L+1])    |
| LUMO+1   | [L]     | [L]    | [L]     | [L]          | [L]          |
| LUMO     | [H-1]   | [H-1]  | [H-1]   | [H-1]        | [H-1]        |
| HOMO     | [H-2]   | [H-2]  | [H-1]   | [H-2]        | [H-2]        |
| HOMO-1   | N ([H-1]) | [H-3]  | [H-3]   | N ([H])      | N ([H-3])    |
| HOMO-2   | [H-3]   | [H-4]  | N ([H-3]) | [H-3]       | N ([H-4])    |
| HOMO-3   | [H]     | [H]    | [H-4]   | [H]          | [H-3]        |
| HOMO-4   | N ([H-3]) | [H]    | [H-3]   | [H]          | [H-3]        |
| HOMO-5   | N ([H-3]) | [H]    | [H]     | [H]          | [H-3]        |
| HOMO-6   | N ([H-3]) | N ([H-4]) | N ([H]) | N ([H-4])    | N ([H])      |
| HOMO-7   | N ([H-3]) | N ([H-3]) | N ([H-3]) | N ([H-4])    | N ([H-4])    |
| HOMO-8   | N ([H-3]) | N ([H-3]) | N ([H-3]) | N ([H-3])    | N ([H-3])    |
| HOMO-9   | N ([H-3], [H]) | N ([H-3]) | N ([H-3]) | N ([H-3])    | N ([H-3])    |
| HOMO-10  | N ([H-3]) | N ([H-3]) | N ([H-3]) | N ([H-3])    | N ([H-3])    |
| HOMO-11  | N ([H])  | N ([H]) | N ([H]) | N ([H])      | N ([H])      |
In this process, two threshold rule” for convenience. This selection process is to discard unimportant contributions from a fragment to simplify the correlation diagram. These thresholds were empirically determined by referring to the man-made orbital correlation diagram by the present authors.

We made a correspondence table on the basis of the “highest similarity score rule”, and the “35/10 % threshold rule” for preparation of constructing the correlation diagram. In contrast to the orbital correlation diagram, herein we allowed for fragment MOs to be correlated with plural number of target MOs. This idea seems natural since the contracted Gaussian basis sets are used to represent various MOs.

2.4 Orbital similarity score (OSS)

OSS for each pair of MOs is defined as the volume fraction of the spatially overlapped region between the target and reference MOs defined with the iso-value surface as explained in Sec. 2.1.

The overlap is strongly dependent on the relative orientation of molecules in the three-dimensional space. Here we focus on the maximum value of OSS. Initially, we tried to maximize the OSS only by using the data of the MO shapes, but it was technically very difficult. Therefore, we developed a two-step method in which structural similarity was also referred to. The flowchart of our algorithm is summarized in Figure 2. This is an ad hoc method specialized to benzene derivatives. More general procedures will be investigated in our future work.

In the present algorithm, the first step is the preconditioning for OSS evaluation. In this process, two molecules to be compared are appropriately oriented to have the best structural overlap in the common fragment. Since we study benzene derivatives, the C6H5 moiety was chosen as a common fragment. In displacing a reference molecule toward the target one, we chose three atoms in each of the molecules, labeled as 1-3 in Figure 3. Herein a pair of the atoms with the same number is called...
“corresponding atoms (CAs)”. Then, translation, vector alignment, and plane rotation were subsequently performed. These operations are as follows:

(i) translation: to overlap the CAs labeled as “1”,
(ii) vector alignment: to overlap the two vectors defined by atoms 1 and 2 on each molecule,
(iii) plane rotation: to rotate a molecular plane defined by atoms 1-3. This operation was performed using the Nelder-Mead (NM) method [16] implemented using Mathematica 11.0 [17].

After the structural alignment, we evaluated OSS by calculating $N_{\text{ovlp}}/N_{\text{total}}$ where $N_{\text{total}}$ and $N_{\text{ovlp}}$ are the total number of the lattice points of a reference MO and that in the overlapped region, respectively. In order to take nodal structures of MOs into account, each lattice point is defined to have the sign of + or – which correspond to the sign of the MO amplitude. It is noted that each MO, which is a single-electron wavefunction, is physically identical even when the sign is inverted (i.e. (-1) is multiplied). Therefore, OSS was calculated with and without orbital sign inversion, and the larger OSS was chosen as the final score.

We can arbitrarily choose CAs in the present algorithm, and hence various structure alignment is possible. For example, when phenyl radical is used as a reference molecule, the carbon atom at the radical center can be paired with any of the six carbon atoms in phenol. In the application to the benzene derivatives, six pairs for atom 1 were exhaustively investigated, while the atom 2 was defined as a carbon atom at the para position to atom 1. The atom 3 was the hydrogen atom at the meta position as shown in Figure 3.

In drawing the orbital interaction diagram for phenol, we decomposed it into the C$_6$H$_5$ and OH groups. The latter fragment consists of only two atoms, and the atom 3 cannot be defined in this case. In order to maximize the OSS between phenol and the OH group, we rotated the latter group and its MOs. This optimization was performed by using the differential evolution (DE) method [18] implemented in Mathematica 11.0 as shown in Figure 2. Even in this case, orbital inversion was performed to search for the maximum OSS.

3. Results and Discussion

3.1 MO correlation diagram for benzene derivatives

The computational algorithm described in the previous section was applied to draw orbital correlation diagrams for benzene derivatives. They include benzene, aniline, phenol, benzaldehyde, benzoic acid, and benzonitrile. We chose them since we could expect systematic changes in orbital energy levels upon substitution of the functional groups. It is known that the -NH$_2$ and -OH groups are $\pi$-electron-donating, while the -CHO, -COOH, -CN groups are $\pi$ electron-withdrawing [19].

The geometries of the target molecules were optimized by using the DFT(M06-2X) method. Then, we evaluated OSS of the benzene derivatives to those of the phenyl radical (C$_6$H$_5$). In the following discussion, MOs of the target molecules are distinguished using the angle bracket $<$, while the square bracket $[  ]$ is added to those of the reference molecule (C$_6$H$_5$). The orbital energies and orbital shapes obtained by the iso-value plot are tabulated in Table S1 for benzene derivatives and in Table S2 for phenyl radical.

The calculated OSS are summarized in Table S3. This table provides reasonable assignments of MOs of the target molecules. For example, <HOMO> of benzene is highly correlated with [HOMO-2] as the OSS is the largest. This MO is assigned to the $\pi$ orbital with one nodal plane, and are essentially degenerate with [HOMO-1] although the degeneracy is slightly lifted. The latter MO also has extraordinarily large OSS. The thirdly important contribution to <HOMO> of benzene in Table 1 is the [HOMO-4] of the phenyl radical. In this case, the similarity score is 36.5, which is almost the half of that for [HOMO-1]. Therefore, although the difference of 7.5 % would be within the tolerance for orbital assignment, we assign it to [HOMO-2] following the largest OSS rule mentioned in Sec. 2.2.

Regarding <HOMO-1> of benzene, OSS indicates that it should be assigned to [HOMO-1] of C$_6$H$_5$. This assignment is reasonable since <HOMO-1> is energetically degenerate to <HOMO> (See Table S1), and the thirdly important contribution is due to [HOMO-4]. Its OSS was calculated to be 36.6 % which is much smaller.
than those of [HOMO-1] and [HOMO-2].

<HOMO-2> and <HOMO-3> are assigned to [HOMO-3] because of the largest OSS (74.6 and 71.8 %, respectively). It is interesting that [HOMO], which is a singly occupied MO and is responsible to the radical character, also appreciably contributes to these MOs since its OSS are 62.6 and 52.1 % for <HOMO-2> and <HOMO-3>, respectively. As shown in Table S2, both [HOMO] and [HOMO-3] commonly consist of the p\textsubscript{xy} orbitals of carbon atoms. The similarity score for <HOMO-4> of benzene shows very high OSS (93.8 %) with [HOMO-1].

As for <LUMO> of benzene, Table S3 shows that the corresponding MOs are [LUMO] (77.0 %) and [LUMO+1] (69.7 %), <LUMO+1> also consists of [LUMO] (71.4 %) and [LUMO+1] (77.2 %). Similar to the discussion on <HOMO> and <HOMO-1>, <LUMO> and <LUMO+1> are energetically degenerate in benzene. It is also true in phenyl radical except for a small energy splitting due to its radical character.

The present method can also assign MOs of substituted benzene derivatives. For example, the OSS of benzaldehyde with phenyl radical indicates that <HOMO> is assigned to [HOMO-1] (67.4 %). This assignment suggests that <HOMO> of benzaldehyde should be correlated with <HOMO> of benzene because these two MOs commonly have a character of [HOMO-1] of the phenyl radical. The highest OSS is less than that for <HOMO> of benzene. This may be due to electron delocalization through the \( \pi \) conjugation due to the -CHO group.

<HOMO-1> of benzaldehyde is a counterpart of its <HOMO> since both [HOMO-1] and [HOMO-2] of phenyl radical contribute to these target MOs. In <HOMO-1>, the OSS of these MOs are 45.0 and 74.1 %, respectively. Thus, <HOMO-1> of benzaldehyde is assigned to [HOMO-2].

Interestingly, <HOMO-2> of benzaldehyde has a unique feature in OSS in comparison with the other MOs. For this MO, the highest score is obtained with [HOMO], but its value is ca. 30 %. The OSS for the other MOs are less than 10 %. Thus, we assign this MO as the one localized onto the functional group. This MO cannot be correlated with any MOs of the benzene molecule. Table S1 summarizing the orbital shapes shows that <HOMO-2> is mainly localized onto the aldehyde group: more specifically, although the MO looks like a \( \pi^* \) orbital, it should be assigned to the nonbonding MO.

The similarity between <HOMO-3> and [HOMO-3] are very high (84.3 %) so that they are uniquely correlated. <HOMO-4> can be related to [HOMO] with the OSS of 48.9 %. This assignment is justified by the orbital shape depicted in Table S1 and S2.

The OSS for <LUMO> and <LUMO+1> of benzaldehyde indicates that they are a mixture of [LUMO] and [LUMO+1] of phenyl radical, but, in applying “the highest similarity rule”, they are correlated to [LUMO] and [LUMO+1], respectively. These assignments are essentially similar with those in benzene. Thus, orbital correlation between benzaldehyde and benzene is straightforward.

The assignment discussed so far highlight their important characteristics of benzaldehyde: introduction of a functional group with the \( \pi \) orbitals contributes to delocalize the electron distribution. In addition, a functional group can introduce an MO spatially localized on itself.

By referring to MOs of a common fragment, the present computational procedure can be applied to investigate an orbital correlation diagram. Herein we investigated benzene derivatives having the amino, hydroxy, carboxylic acid, or nitrile group. In order to draw the diagram and perform statistical analysis using MO levels as explanatory variables, it is necessary to establish one-to-one correspondence between the target and reference MOs. For this purpose, we prepared a correspondence table shown in Table 1. In making this table, we defined the assignment criteria as follows:

(i) no duplicated correspondence is allowed,

(ii) target MOs energetically close to the frontier MOs (i.e. HOMO and LUMO) has higher priority,

(iii) reference MOs whose similarity score is less than 35 % are not considered as a corresponding MO.

The condition (i) is to make the present method applicable to multi regression analysis using MO energy
levels. This allows us include target MOs having no counterpart MOs. As seen in Table 1, the unmatched MOs appear in the lower (higher) energy region in the occupied (unoccupied) MO space. It is noted that the value of 35 % is arbitrary. Herein we tentatively determined this value after some trials and by referring the orbital shape summarized in Table S1 and S2.

The failure of finding the partner orbital among the reference MOs in Table S3 has some implications. Firstly, it implies that further improvement of our computational algorithm would be possible. The second reason is that the reference MOs of phenyl radical are limited from HOMO-4 to LUMO+1. For example, energetically deeper MOs would be necessary to find the most similar one. It also suggests the existence of the functional group attached to the benzene ring. In the present assignment procedure, we only focus on similarity with phenyl radical. Thus, the similarity score would be very low when the attached functional group gives dominant contribution to a MO of interest. In such a case, the MO should be assigned to the functional group, not to phenyl radical.

Figure 4 shows the orbital correlation for the six target molecules where each correspondence indicated by a dashed line is defined in Table S3. We observe systematic changes in orbital energies of the frontier MOs including HOMO, HOMO-1, LUMO, and LUMO+1. No systematic changes are observed for the other MOs. By focusing on the correspondence (i.e. the dashed lines), it becomes evident that replacing the functional group stabilizes the frontier MOs in the order of -NH2, -OH, -H, -CHO, -COOH, and -CN. These observations are reasonable when we think of the characteristics (substituent effect) of the functional group: the electron-withdrawing groups destabilize the HOMO levels and hence enhance reactivity of the compound. The electron-donating groups stabilize the LUMO level, enhancing electron accepting ability of the compounds, and gives low-lying excited states due to the decrease of the HOMO-LUMO gap. Table S1 also highlights an interesting feature that <HOMO> and <HOMO-1> are energetically isolated from the other MOs. This suggests importance of these frontier MOs.

The existence of the MOs localized onto the substituent group is also highlighted in aniline, benzaldehyde, and benzoic acid. In Table S1, we see that the nitrogen atom gives important contribution to <HOMO-3> of aniline. For <HOMO-3> of benzaldehyde, and <HOMO-3> and <HOMO-4> of benzoic acid, <HOMO-2> of benzonitrile is assigned to the MO localized onto the CN group.

The MOs underlying below <HOMO-1> have also characteristic features. <HOMO-2> and <HOMO-3> in benzene are the π* orbital of the six-membered carbon ring. <HOMO-4> is assigned to the π orbital with no nodal planes. The similarity scores evaluated in this work suggest that there are the corresponding MOs in the other benzene derivatives investigated in this work.

From the machine-made orbital correlation diagram, it is evident that the SOMO of phenyl radical is appreciably stabilized upon binding the substituent group. For example, the SOMO-derived MOs appear as <HOMO-6> in Table S3 and Figure 4. This result is consistent with the fact that the radical orbitals strongly interacts to form a chemical bond.

For the MOs below the SOMO-derived one, we failed to find correspondences of MOs with the reference MOs. Also the present application would be limited to a set of molecules having a common fragment structure. In order to establish any types of MO correspondences, further works are necessary and will be reported elsewhere.

3.2 Orbital interaction diagrams of phenol and aniline

Orbital interaction diagrams graphically show the correlation between MOs of a target molecule and those of its fragments. Such information helps us understand how the fragments interact to form a molecule i.e. chemical reactivity of the fragments. At the same time, the diagrams summarize overall features of the electronic structures of the molecules described by the energy alignment of MOs and the electron occupations. Thus, electron distribution and charge-transfer processes in the chemical interaction are also understandable through the diagrams.

We applied our program to draw orbital correlation diagrams. Herein phenol and aniline were investigated aiming at establishing the computational procedure to draw the orbital interaction diagram. The technique should be somewhat different from the orbital correlation diagram shown in Figure 4 because multiple correlation among MOs should be allowed. It was also motivated since there were some unassigned (uncorrelated) MOs in Figure 4. This is reasonable since some MOs originate from the substituent group added to the benzene moiety.

To draw the diagrams, phenol and anilines were decomposed into the phenyl radical (C6H5) and the substituent group (SG = OH or NH2). OSS was calculated for each of the fragment MOs, and the “highest similarity rule” was applied to interconnect between a target MO and a fragment MO. In some MOs, the OSS become small and some fragment MOs contributed equally. In order to obtain simple diagrams, i.e. for coarse-graining of the information, we introduced a threshold for orbital assignment. It was found that the OSS for the substituents were generally small in comparison with C6H5 probably because of their smallness in size. Therefore, different thresholds were introduced depending on the fragment to be correlated with the target MO. As was described in Sec. 2.3, the thresholds of 35 % and 10 % for C6H5 and SG were adopted in the present implementation. These values
The computer-generated orbital interaction diagrams are shown in Figures 5 and 6 for phenol and aniline, respectively, where the MOs from HOMO-7 to LUMO+1 were selected since they can be classified as the frontier orbitals of these molecules. These diagrams were obtained referring to the calculated OSS summarized in Tables S4-S7 in Supplementary Information. These figures reasonably highlight important features in their electronic structures. For example, <HOMO> of phenol is found to consist of [HOMO-1] of phenyl radical assigned to the \( \pi \) orbital, and the \( p_z \) orbital of the OH group. <HOMO-1> is directly related to [HOMO-2] of phenyl radical. In this case, the contribution from the OH group is negligible because of the orbital symmetry. This is consistent with the fact that the energy level of the MO is unchanged even after the formation of the C-O bond in phenol. [HOMO] of phenyl radical was correlated with <HOMO-4> which is very deep in energy. The counterpart of the orbital interaction was <HOMO-1> of the OH (NH\(_2\)) group. The appreciable stabilization of <HOMO-4> implies that the orbital interaction should be very strong, which is consistent with the fact that [HOMO] of phenyl radical is singly occupied responsible to the radical character of C\(_6\)H\(_5\).

Similar features in the orbital components are also observed even in <LUMO> and <LUMO+1>: in Figures 5 and 6, it is evident that <LUMO+1> of phenol is due to antibonding combination of [LUMO+1] of phenyl radical and [HOMO] of the OH radical. <LUMO> of phenol has no strong interaction with the substituent group. This is also reflected in small energy change in this orbital.

4. Summary and Conclusions

We have developed computational algorithms to perform automatic drawings of two types of orbital correlation diagrams. These procedures are expected to perform quantum-chemistry based machine learning, which can be called “electronic-structure informatics”. The correlation diagrams were obtained through evaluation of orbital similarity scores (OSS). Evaluated after appropriate structure alignment.

The present computational algorithm was applied to investigate a diagram showing the correlation of the outer-valence MOs of the six benzene derivatives. A reasonable correlation diagram was obtained for the frontier MOs including HOMO, HOMO-1, LUMO, and LUMO+1. Although the correspondence is limited in some MOs, the present method enables us to include the plural number of MOs for machine learning, not limited to HOMO and LUMO only. The character-based correspondence of MOs justifies the statistical analysis using the orbital energy levels. An interesting application of the present method is to try to re-discover the “Frontier Orbital Theory” by Fukui [5]. Computational justification of Fukui’s theory can be achieved through conceptually unbiased finding of MOs that play decisive roles in chemical reactions.

We have also implemented a code investigating an orbital interaction diagram which is another type of an orbital correlation diagram. This is to show the inter-relation among MOs of the target molecule and its fragments. Through applications to phenol and aniline, we demonstrated that our automated procedure provides reasonable diagrams for understanding orbital characters.

Further development of the present method is necessary to extend the applicability of the present method to various molecules, not limited to benzene derivatives. In addition, the method needs to be improved to establish orbital correspondence for a variety of MOs.
including occupied (unoccupied) MOs with lower (higher) energy. Such extensions are now on-going and will be reported elsewhere.

5. Perspectives on Chemoinformatics and Materials Informatics

Since biological phenomena include a series of complicated chemical processes, development of effective descriptors for machine learning is indispensable and challenging in chemoinformatics and materials informatics. Recent studies in this field are trying to include electronic-degrees of freedom, but these trials are still in its infancy. The difficulty arises from the fact that electronic wavefunctions describing electronic states are complicated mathematical function and it has been unsuccessful to mathematically compactly characterize their features. It is noted that, in addition to the wavefunctions, we also need to develop descriptors representing electronic responses of molecules since the characteristics are essential to molecular properties such as electric, magnetic, and optical properties.

Combined with the conventional descriptors in chemoinformatics, various descriptors in next generation will be derived from computer simulations including electronic-structure and molecular dynamics calculations. On-the-fly evaluation of the descriptors will be finally achieved in future, but, before the achievement, it will be necessary to develop database storing descriptor data obtained through the simulations. Development of such database and search engines for its application seems to be practically very important task in near future.

Acknowledgments

The work was partially supported by Grant-in-Aid for Scientific Research on Innovative Areas (2601): π-System Figuration (26102015 to MS) from MEXT, Japan.

References and Notes

[1] A. G. Maldonado, J. P. Doucet, M. Petitjean and B. T. Fan, Mol. Divers., 10, 39-79 (2006).
[2] K. Fukui, Angew. Chem. Int. Ed. Engl., 21, 801-876 (1982).
[3] R. B. Woodward and R. Hoffmann, Angew. Chem. Int. Ed. Engl., 8, 781-932 (1969).
[4] M. Sugimoto, T. Ideo, A. B. Manggara, K. Yoshida, and T. Inoue, J. Comp. Aid. Chem., submitted for publication.
[5] W. Horton, S. Peerannawar, B. Török and M. Török, Struct. Chem., 30, 23-35 (2019).
[6] P. W. Ayers and M. Levy, Theor. Chem. Acc., 103, 353-360 (2000).
[7] K. Tuppurainen, Chemosphere, 38, 3015-3030 (1999).
[8] R. Perkins, H. Fang, W. Tong, and W. J. Welsh, Environ. Toxicol. and Chem., 22, 1666-1679 (2003).
[9] C. S. Miller and M. Ellison, J. Chem. Educ., 92, 1040-1043 (2015).
[10] H. Nohira and T. Nohira, J. Theor. Comput. Chem., 11, 379-389 (2012).
[11] H. Hirao, Comput. Mol. Sci., 1, 337-349 (2011).
[12] L. Zhao, M. V. Hopfgarten, D. M. Andrada and G. Frenking, Wires. Comput. Mol. Sci., 8, No. e1345 (2018).
[13] Y. Zhao and D. G. Truhlar, Theor. Chem. Acc., 120, 215-241 (2008).
[14] G. A. Petersson, A. Bennett, T. G. Tensfeldt, M. A. Al-Laham, W. A. Shirley and J. Mantzaris, J. Chem. Phys., 89, 2193-2218 (1988).
[15] M. J. Frisch et al., Gaussian 09, Revision D. 01, Gaussian, Inc., Wallingford CT (2016).
[16] J.A. Nelder and R. Mead, Comput. J. 7, 308-313 (1965).
[17] Mathematica Ver. 11.0, Wolfram Research Inc., Champaign, Illinois (2018).
[18] K. Price and R. Storn, J. Global Optim., 11, 341-359 (1997).
[19] P. Y. Bruice, “Organic Chemistry 5th Edition”, Prentice Hall (2006).