Principal Component Analysis with Energy Density of *Calophyllum* Coumarins

Mari Takeuchi, Ayako Nakata, and Hiromi Nakai*
Department of Chemistry, Waseda University, 3-4-1, Okubo, Shinjuku-ku, Tokyo 169-8555

(Received March 15, 2005; CL-050346)

Energy density analysis was applied to 25 species of *Calophyllum* coumarins. When the principal component analysis was performed by using energy density as variables, the 25 coumarins were classified into six groups. The first and second principal components were related to the substituent effect and the molecular skeleton, respectively. The molecular skeletons sufficiently influence anti-HIV-1 activities of coumarins rather than the substituents.

In the drug discovery process, drug candidate molecules are designed by altering the substituents and structures of naturally existing molecules. For example, some coumarins compounds were isolated from *Calophyllum* genus, as anti-HIV-1 active candidate molecules.[1–4] Among the isolated *Calophyllum* coumarins, some coumarins were demonstrated to be active against HIV-1, such as calanolide A, and other ones with structures similar to calanolide A were less active or totally inactive. However, it is difficult to identify the substituent or structure that could play a key role in anti-HIV-1 activity of coumarins.

Recently, Mizukami[5,6] has applied principal component analysis (PCA) to the study of the relationship between the toxicity and electronic structures of the dioxins. The study has proven the PCA technique is useful for categorization according to molecular character. Mizukami used the electron densities of highest occupied molecular orbitals as valiables for the PCA. Atomic energy density, which is estimated by energy density analysis (EDA),[7] also affects the molecular character. Thus, in this study, we have categorized *Calophyllum* coumarins by the PCA using the atomic energy density as variables.

The geometrical parameters of 25 *Calophyllum* coumarins (shown in Figure 1 and Table 1) were all optimized by performing the density functional theory (DFT) calculations. The B3LYP hybrid functional[8] was adopted in the DFT calculations. The basis sets used for H, C, and O were the valence double zeta sets of Dunning.[9] All geometry optimizations were carried out with the use of the Gaussian98 program.[10] Energy densities were calculated for the optimized molecules. The EDA calculations were carried out by linking the original EDA code with GAMESS.[11] Since the skeletons formed by O1–C14 atoms are common for all of the 25 coumarin species, energy densities on 16 carbon atoms and 4 oxygen atoms were used as variables in the PCA calculation.[12]

Table 1. Structures of 25 species of *Calophyllum* coumarins

Class	R	Steric configuration	X	Y	#
I	Pr	11α-Me	OH	H	1
		11α-Me	OMe	H	2
		11α-Me	OAc	H	3
		11α-Me	H	OH	4
		11α-Me	H	OH	5[a]
		11α-Me	H	OMe	6
		11α-Me	OH	H	7
	Ph	11α-Me	OH	H	8
		11α-Me	H	OH	9
		11α-Me	H	OH	10[b]
		11β-Me	H	OH	11
		11β-Me	OH	H	12
	Me	11α-Me	OH	H	13
		11α-Me	H	OH	14
		11α-Me	H	OMe	15
II	Ph	11α-Me	—	—	16
		11α-Me	—	—	17[c]
		11β-Me	—	—	18
III	Pr	—	—	—	19
	Me	—	—	—	20
IV	Pr	7α-Me	—	—	21
		7β-Me	—	—	22
	Ph	7α-Me	—	—	23
V		6α, 7α	—	—	24
		6β, 7β	—	—	25

[a]Enantiomer of 4. [b]Enantiomer of 9. [c]Enantiomer of 16.

Table 2 shows the coefficients of principal components whose absolute values are above 0.300. With respect to the first principal component (PC1), only C14 has a large coefficient of 0.984. A C14 atom is included in the substituent combined to a C4 site such as the methyl, propyl, and phenyl groups. On the other hand, as regards the second principal component

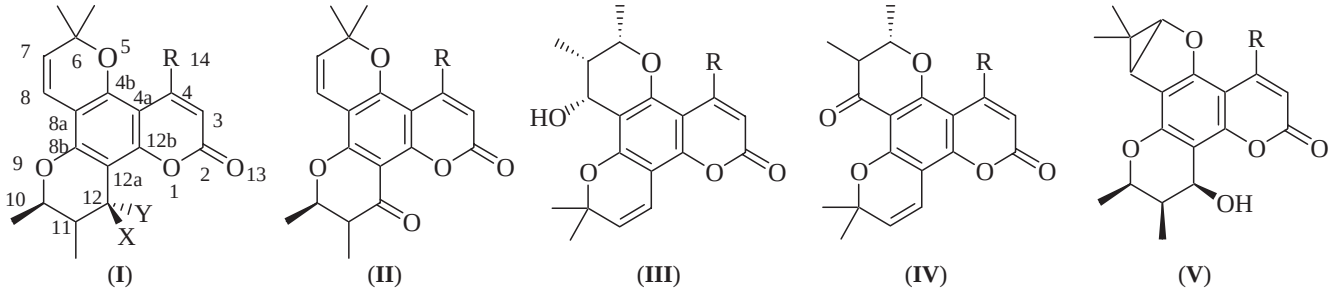

Figure 1. Skeletons of *Calophyllum* coumarins.

Copyright © 2005 The Chemical Society of Japan

(PC2), the absolute values of coefficients of C6, C7, C8a, C10, and C12a are larger than 0.300. Furthermore, the coefficients are delocalized throughout the skeleton. The cumulative contribution ratio from PC1 to PC2 was 0.916, which indicates that the analysis with PC1 and PC2 gives high reliability.

Figure 2 shows plots of PC1 scores versus PC2 ones. The plots are categorized into three groups according to PC1 scores; the scores of the three groups are about 0.14, 0.07, and −0.10, respectively. The plots are also categorized into two groups according to the sign of the PC2 scores. As results, the six groups are defined as **A**–**F**, as shown in Figure 2. PC1 reflects the substituent effect at the C4 site. The molecules in groups (**A**, **B**), (**C**, **D**), and (**E**, **F**) have methyl, propyl, and phenyl substituents at the C4 site, respectively. On the other hand, PC2 reflects the difference in molecular skeleton. The molecules in groups **A**, **C**, and **E** have **I** or **II** skeleton, while those in groups **B**, **D**, and **F** have **III**, **IV**, or **V** skeleton. The bonding characters of the C6, C7, and C10 atoms are different due to the different skeletons, whereas those of C8a and C12a atoms are common. The energy densities of C8a are −38.139–−38.148 hartree in (**I**, **II**) and −38.077–−38.098 hartree in (**III**, **IV**, **V**), while those of C12a are −38.076–−38.101 hartree in (**I**, **II**, **V**) and −38.137–−38.149 hartree in (**III**, **IV**). The energy densities of C8a and C12a reflect the neighboring C7–C8 and C11–C12 bonds, respectively; the C7–C8 bonds in **I** and **II** are double bonds and those in **III**, **IV**, and **V** are single bond. The C11–C12 bond is a single bond in **I**, **II**, and **V**, and is a double bond in **III** and **IV**. Therefore, atomic energy density reflects not only the direct bonding characters but also the neighboring ones.

The activities of the 25 coumarins were investigated experimentally; Compounds #1, #5, #8, #10, #13, and #14 are strongly active, while #3, #6, #7, #11, #12, and #16 are weakly active, and #2, #18, #19, #21, #24, and #25 are inactive. The activities of #4, #9, #15, #17, #20, #22, and #23 are unknown. In Figure 2, all active molecules are located on the left-hand side with respect to PC2. There is no active molecule on the right-hand side. However, as regards PC1, both active and inactive molecules are included in groups (**C**, **D**) and (**E**, **F**). This result indicates that the coumarin skeletons significantly influence their activities.

In conclusion, the 25 species of *Calophyllum* coumarins were classified into six groups by PCA using energy densities as variables. PC1 reflects the substituent effect at the C4 site. PC2 reflects the difference in molecular skeleton. It is strongly suspected that the coumarin skeletons classified by PC2 significantly influence their activities.

Part of our calculations was performed at the Research Center for Computational Science (RCCS) of the Okazaki National Research Institutes and the Media Network Center (MNC) of Waseda University. This study was partially supported by a NAREGI Nano-Science Project of the Japanese Ministry of Education, Culture, Sports, Science, and Technology (MEXT), by the 21st-Century Center Of Excellence (21COE) "Practical Nano-Chemistry" from MEXT, and by a Grant-in-Aid for Exploratory Research "KAKENHI 16655010" from MEXT.

Table 2. Coefficients of the principal components

	Component	Coefficient
PC1	C14	0.984
PC2	C6	0.614
	C8a	0.415
	C10	−0.351
	C7	−0.308
	C12a	−0.302

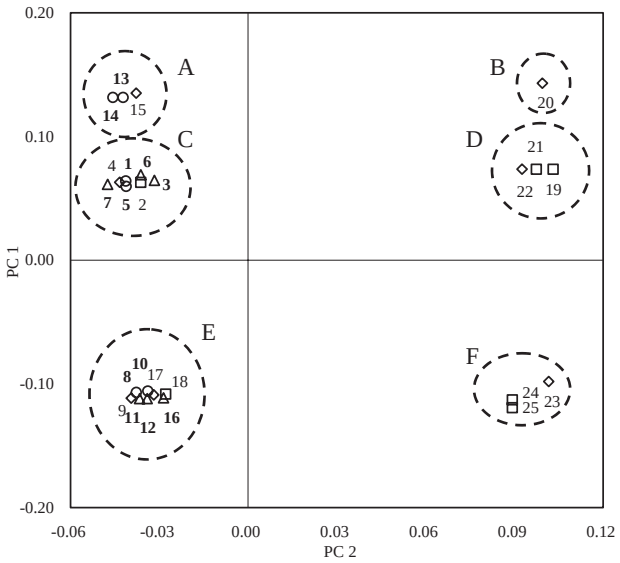

Figure 2. Principal component analysis with the use of energy densities of 25 *Calophyllum* coumarins. Scores of the first principal component (PC 1) versus the second principal component (PC 2). The numbers refer to the coumarins as listed in Table 1. The activities of coumarins are shown as follows: strongly active (○), weakly active (△), inactive (□), and unknown (◇).

References and Notes

1 T. Ishikawa, *Heterocycles*, **53**, 453 (2000).
2 a) Y. Kashman, K. R. Gustafson, R. W. Fuller, J. H. Cardellina, II, J. B. McMahon, M. J. Currens, R. W. Buckheit, Jr., S. H. Hughes, G. M. Cragg, and M. R. Boyd, *J. Med. Chem.*, **35**, 2735 (1992). b) Y. Kashman, K. R. Gustafson, R. W. Fuller, J. H. Cardellina, II, J. B. McMahon, M. J. Currens, R. W. Buckheit, Jr., S. H. Hughes, G. M. Cragg, and M. R. Boyd, *J. Med. Chem.*, **36**, 1110 (1993).
3 A. D. Patil, A. J. Freyer, D. S. Eggleston, R. C. Haltiwanger, M. F. Bean, P. B. Taylor, M. J. Caranla, A. L. Breen, H. R. Bartus, R. K. Johnson, R. P. Hertzberg, and J. W. Westly, *J. Med. Chem.*, **36**, 4131 (1993).
4 J. H. Cardellina, II, H. R. Bokesch, T. C. McKee, and M. R. Boyd, *Bioorg. Med. Chem. Lett.*, **5**, 1011 (1995).
5 Y. Mizukami, *J. Mol. Struct. (THEOCHEM)*, **672**, 161 (2004).
6 Y. Mizukami, *Chem. Lett.*, **33**, 1328 (2004).
7 H. Nakai, *Chem. Phys. Lett.*, **363**, 73 (2002).
8 a) D. A. Becke, *J. Chem. Phys.*, **98**, 5648 (1993). b) P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch, *J. Phys. Chem.*, **98**, 11623 (1994).
9 T. H. Dunning, Jr., *J. Chem. Phys.*, **53**, 2823 (1970).
10 M. J. Frisch, G. W. Trucks, H. B. Schlegel, et al., "Gaussian 98," Gaussian, Inc., Pittsburgh, PA (2000).
11 M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. Su, T. L. Windus, M. Dupuis, and J. A. Montgomery, Jr., *J. Comput. Chem.*, **14**, 1347 (1993).
12 See Supporting Information for the detailed data of atomic energy densities.

Published on the web (Advance View) May 21, 2005; DOI 10.1246/cl.2005.844