Journal of Saudi Chemical Society (2016) **20**, S547–S551

King Saud University

Journal of Saudi Chemical Society

www.ksu.edu.sa
www.sciencedirect.com

ORIGINAL ARTICLE

Ant colony optimization as a descriptor selection in QSPR modeling: Estimation of the λ_{max} of anthraquinones-based dyes

Morteza Atabati *, Kobra Zarei, Azam Borhani

School of Chemistry, Damghan University, Damghan, Iran

Received 6 November 2012; accepted 28 March 2013
Available online 6 April 2013

KEYWORDS

Anthraquinone;
λ_{max};
QSPR;
Ant colony optimization

Abstract Quantitative structure–property relationship (QSPR) studies based on ant colony optimization (ACO) were carried out for the prediction of λ_{max} of 9,10-anthraquinone derivatives. ACO is a meta-heuristic algorithm, which is derived from the observation of real ants and proposed to feature selection. After optimization of 3D geometry of structures by the semi-empirical quantum-chemical calculation at AM1 level, different descriptors were calculated by the HyperChem and Dragon softwares (1514 descriptors). A major problem of QSPR is the high dimensionality of the descriptor space; therefore, descriptor selection is the most important step. In this paper, an ACO algorithm was used to select the best descriptors. Then selected descriptors were applied for model development using multiple linear regression. The average absolute relative deviation and correlation coefficient for the calibration set were obtained as 3.3% and 0.9591, respectively, while the average absolute relative deviation and correlation coefficient for the prediction set were obtained as 5.0% and 0.9526, respectively. The results showed that the applied procedure is suitable for prediction of λ_{max} of 9,10-anthraquinone derivatives.

1. Introduction

Anthraquinone (9,10-dioxoanthracene) is an aromatic organic compound and it is a derivative of anthracene. 9,10-Anthraquinones (see Fig. 1), as the largest group of naturally occurring quinones, are of importance both in industry (Thomson, 1971) and in medicine (Zembower et al.,1992). Anthraquinone is used in production of dyes, such as alizarin. Anthraquinone dyes are one of the oldest types of dyes used commercially. Many natural pigments are derivatives of anthraquinone. In addition to a wide variety of chemical and industrial applications, the synthetic derivatives of anthraquinones, as well as naturally occurring derivatives, have been used.

One of the main difficulties in determining the color of organic compounds is the astonishing accuracy of the standard human eye, which can distinguish, in some parts of the visible spectra (typically in the green region), differences of coloration corresponding to less than 1 nm λ_{max} shifts. Nevertheless, in

* Corresponding author.
E-mail addresses: atabati@du.ac.ir, atabati@gmail.com (M. Atabati).
Peer review under responsibility of King Saud University.

ELSEVIER Production and hosting by Elsevier

http://dx.doi.org/10.1016/j.jscs.2013.03.009

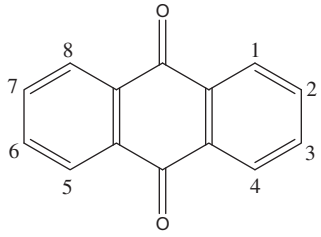

Figure 1 9,10-Anthraquinone and its substitution positions.

regard to practical industrial applications, the theoretical calculations could be regarded as serious competitors to experimental approaches for developing new dyes and/or pigments if they were to deliver an estimate of the λ_{max} values within 5–15 nm accuracy. Such a chemical accuracy for large conjugated molecules is still a tremendous challenge for the modelization approaches (Perpète et al., 2006).

In the present work, λ_{max} values for 9,10-anthraquinone derivatives have been investigated by Ant Colony Optimization (ACO).

ACO, as a new general purpose meta-heuristic algorithm, was proposed by Dorigo et al. (Dorigo, 1992; Colorni et al., 1991) in 1991–1992 for solving complicated optimization processes, and first applied in traveling salesman problem (TSP). The method is inspired by the behavior of real ant colonies, wherein ants always find the shortest path between their nest and a food source, thanks to local message exchange via the deposition of pheromone trails. Each ant probabilistically prefers to follow a direction rich in this chemical. The pheromone decays over time, resulting in much less pheromone on less popular paths. Given that over time the shortest route will have the higher rate of ant traversal, this path will be reinforced and the others diminished until all ants follow the same, shortest path (the "system" has converged to a single solution) (Bonabeau et al., 1999).

It is interesting to note that in spite of high abilities of the ACO, this procedure was entered to the field of chemistry with some delay, and only a few publications based on ACO have been reported in the chemical literature, e.g. feature selection in QSAR and QSPR analyses (Izrailev and Agrafiotis, 2002; Shen et al., 2005; Atabati et al., 2010a,b) and analysis of the first derivative fluorescence spectra (Ding et al., 2002).

The field of "ant algorithm" studies models which have been derived from the observation of real ants' behavior, and uses these models as a source of inspiration for the design of novel algorithms for the solution of optimization and distributed control problems. One of the most successful examples of ant algorithms is known as "ant colony optimization" or ACO.

In many ant species, individual ants may deposit a pheromone (a particular chemical that ants can smell) on the ground while walking. By depositing pheromone, they create a trail that is used, for example, to mark the path from the nest to food sources and back. Also they are capable of exploiting pheromone trails to choose the shortest among the available paths taking to food. The pheromone trail-laying and -following behavior of some ant species has been investigated in controlled experiments by several researchers. One particularly brilliant experiment was designed and run by Deneubourg and Goss, who used a double bridge connecting a nest of ants and a food source. In the first experiment the bridge had two branches of equal length (Fig. 2a). Ants were left free to move between the nest and the food source and the percentage of ants that choose one or the other of the two branches were observed over time. The outcome was that eventually all the ants used the same branch. In the second experiment, the long branch was twice as long as the short one (Fig. 2b). In this case, in most of the trails, after some time all the ants chose to use only the short branch. The double bridge experiments show clearly that ant colonies have a built-in optimization capability: by the use of probabilistic rules based on local information they can find the shortest path between two points in their environment. Interestingly, by taking inspiration from the double bridge experiments, it is possible to design artificial ants that, by moving on a graph modeling the double bridge, find the shortest path between the two nodes corresponding to the nest and to the food source (Dorigo and Stutzle, 2004; Dorigo and Caro, 1999; Toksari, 2006).

2. Results and discussion

2.1. Data set

Maximum wavelength (λ_{max}) values for set of 66 9,10-anthraquinone derivatives have been investigated by ACO. Experimental λ_{max} values of these compounds in CH_2Cl_2 solvent were assembled from previous publications (Lambert, 1957; Perpète et al., 2006). The data set was randomly divided in two groups, a calibration set and a prediction set consisting of 36 and 30 compounds, respectively (Table 1). The calibration set was used for the model generation and the prediction set was used for the evaluation of the generated model.

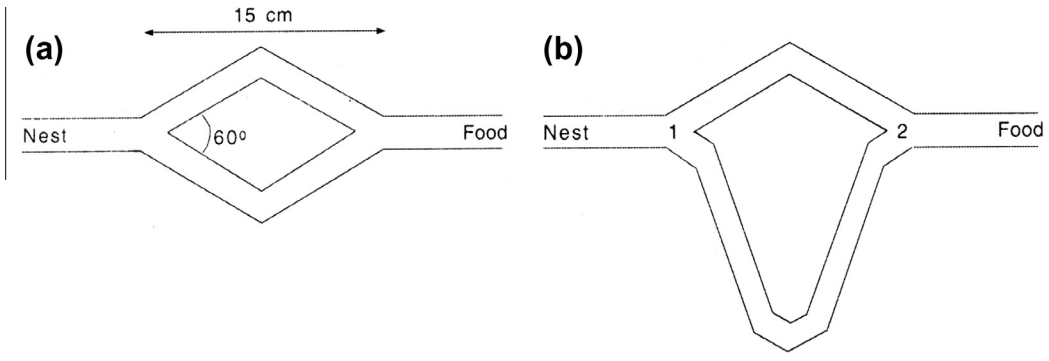

Figure 2 Experimental setup for the double bridge experiment. (a) Branches have equal length. (b) Branches have different length.