Computational and Theoretical Chemistry 1028 (2014) 72–80

Contents lists available at ScienceDirect

Computational and Theoretical Chemistry

journal homepage: www.elsevier.com/locate/comptc

Viewpoint article

Computer-aided design of host molecules for recognition of organic guests

Benjamin P. Hay *, Chuandong Jia, Janos Nadas

Oak Ridge National Laboratory, 1 Bethel Valley Road, PO Box 2008, Bldg 4100, Oak Ridge, TN 37831-6119, USA

ARTICLE INFO

Article history:
Received 27 September 2013
Received in revised form 3 December 2013
Accepted 4 December 2013
Available online 14 December 2013

Keywords:
Computer-aided molecular design
Host
Receptor
Oxygen mustard
HostDesigner

ABSTRACT

This paper presents an overview of a computational approach to design molecular host architectures. After positioning donor groups about a targeted guest molecule, the donor groups are connected with hydrocarbon linkages taken from an extensive database of molecular fragments. Initial ranking of the host structures is based on how well bond vectors on the donor groups match bond vectors on the linkages. The top candidates are then subjected to further evaluation with molecular mechanics identifying host architectures that are both complementary and preorganized for binding the guest. The efficacy of this computer-aided design methodology is illustrated with a search for hydrogen bonding receptors that are structurally organized for complexation with oxygen mustard.

1. Introduction

The design of host molecules that will recognize and bind specific guests is a central and recurrent objective facing researchers in the field of supramolecular chemistry [1,2]. It is generally agreed that chemical recognition can be achieved when three criteria are met. First, the host contains two or more binding sites that each exhibit an intrinsic affinity to interact favorably with the guest. Second, the host is able to adopt a conformation in which all binding sites are structurally positioned to simultaneously engage in favorable interactions with the guest, in other words, the host provides a complementary array of binding sites [3]. Third, the host should exhibit a limited number of stable conformations and the binding conformation should be low in energy relative to other possible forms [4–8]. In the ideal case, the host would be preorganized such that the binding conformation is the most stable form.

Design begins with the selection of a set of donor groups that are appropriate in type and number for interaction with the guest. Once this set of donor groups has been selected, the design process becomes the identification of host architectures that provide a complementary and preorganized arrangement of the donor groups. This process is not trivial and a general approach toward achieving the desired result is needed. One approach is to use computer-aided molecular design methods to generate host molecules and evaluate host–guest interactions. With few exceptions, CAVEAT [9–16] and ConCept [17], structure-based computer-aided design software that can be applied to a wide range of supramolecular systems is lacking. To address this issue we have created the software, HostDesigner (HD) [18]. Although originally developed for application to metal ion hosts [18–20], this software has been adapted to handle a wide range of host–guest interactions and has been used successfully in the design of anion hosts [21–24] and components that direct the formation of high-symmetry molecular assemblies [25,26].

This paper documents some novel features in the HD software and demonstrates the generality of this computer-aided design approach by illustrating how it can be used to identify host architectures for an organic guest, oxygen mustard. Mustards are a class of organic molecules consisting of a heteroatom center (S, O, or N–R) substituted with two 2-chloroethyl groups (Fig. 1). Both sulfur mustard, bis(2-chloroethyl) thioether and the nitrogen mustards, e.g. bis(2-chloroethyl) ethylamine, are potent chemical warfare agents [27]. In contrast, oxygen mustard, bis(2-chloroethyl) ether, does not possess the same toxicity and is frequently used as a sulfur mustard analogue in a variety of diffusion and sorption studies [28–33]. Targeting oxygen mustard as a guest, we demonstrate how structure-based design and subsequent scoring methods are able to explore a large volume of molecule space quickly locating host structures with desirable properties that include (a) large interaction energy with the guest, (b) low conformational reorganization energy, and (c) minimal number of restricted bond rotations on guest complexation.

* Corresponding author. Tel.: +1 865 574 6717; fax: +1 865 574 4939.
E-mail address: haybp@ornl.gov (B.P. Hay).

2210-271X/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.comptc.2013.12.003

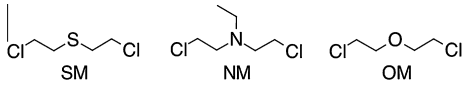

Fig. 1. Structures of sulfur mustard (SM), nitrogen mustard (NM), and oxygen mustard (OM).

2. Computational approach and demonstration

2.1. Methods

Host molecules were constructed using the structure-based design software, HostDesigner (HD) [18]. This software, available on request from the author, assembles structures by combining user-defined input fragments with a hydrocarbon fragments taken from a preexisting database. As described in the next section, information needed to create the input fragments was obtained from MM3 optimized geometries and potential energy surfaces. A representative HD input file is provided as Supporting Information.

Molecular mechanics calculations were performed using the MM3 force field [34–36] as implemented in PCMODEL [37], a program that is capable of performing both geometry optimizations and conformational analyses. Geometries and potential energy surfaces from prior electronic structure calculations were used to extend the default MM3 parameter set to reproduce bond rotations in OM [38] and hydrogen bonding with alkyl chlorides [39] (see Supporting Information for extended parameter set). Conformational searching was accomplished using Monte Carlo random sampling and stochastic simulation strategy with default settings [37]. During the searches, trial structures were generated by alternating between the "bonds method" and the "Cartesian method". In the "bonds method", trial structures are generated by randomly rotating a subset of bonds. In the "Cartesian method", trial structures are generated by removing hydrogen atoms, randomly moving the remaining atoms, and replacing the hydrogen atoms. A search was terminated when one of the stopping criteria is met, either exceeding a limit of 100,000 trials or after 50 consecutive trials in which no new conformation is located within 3.5 kcal/mol of the global minimum.

Reported run times were achieved on a MacPro computer with a 2.66 GHz Intel Xeon processor (MacOS 10.6.8, GNU compiler).

2.2. Design basis and input fragments

As implied by the name *structure-based design*, molecules that are assembled by HD are both built and evaluated based on prior knowledge of molecular geometry. This includes both the geometry of the guest and the geometry associated with individual donor–guest interactions. Donor group fragments are strategically placed around the guest to achieve optimal interaction. This placement defines the relative spatial orientation of bond vectors emanating from the donor groups. The resulting ensemble of guest + donor groups constitutes an input fragment. The HD software builds complete host molecules by connecting the donor groups with covalent bonds to linking fragments taken from a precomputed database. Using oxygen mustard, OM, as an example guest, this section describes the procedure for developing an input fragment for the design process.

The first step is to identify the structure of the guest. Prior conformational analysis of OM revealed this guest to be a flexible molecule with a number of populated conformers at room temperature [38]. Of the ten lowest energy forms, six of them are asymmetric and four possess a C_2 symmetry axis. Although it is conceivable to design a host molecule that would recognize any of these conformers, the two lowest energy C_2 symmetric forms, which have conformational energies of 0 (Fig. 2a) and 1.04 kcal/mol (Fig. 2b), were selected as the targeted guest shapes for this study. This choice simplifies the design process by limiting the number of possible host structures to those that possess C_2 symmetry. Symmetry is desirable from a synthetic point of view in that it is easier to develop synthetic strategies for symmetric hosts.

The next step involves choosing the set of donor groups that will interact with the guest. OM contains three hydrogen bond acceptors, the central ether oxygen atom and two terminal alkyl chloride atoms, leading to the design strategies depicted in Fig. 2. A C_2 symmetric hydrogen bond donor is needed to interact with the center oxygen atom and to which could be attached to two identical groups each interacting with the chloroethyl portions of OM. The urea functional group matches these requirements since (a) it donates two hydrogen bonds to the acceptor, (b) adopts a perpendicular orientation to the C–O–C plane of an ether molecule to give a C_2 symmetric geometry, Fig. 3a, and (c) synthetic routes to N,N′-disubstituted ureas allow covalent connections to two terminal hydrogen bonding groups, one on each side of the urea group.

Amide functional groups were selected as terminal hydrogen bond donors for the chloride atoms. In a prior study of the role of alkyl chlorides as hydrogen bond acceptors, an extensive search located four minima for hydrogen-bonded complexes between acetamide and chloroethane at the MP2/aug-cc-pVDZ level of theory [39]. These structures, Fig. 3b, exhibited interaction energies ranging from −6.60 to −8.12 kcal/mol. After modification to the default parameter set (see Supporting Information) the MM3 model locates all four of these minima and reproduces interaction energies to within ±0.6 kcal/mol.

Six input fragments were constructed by placing urea and amide groups about the two symmetric OM conformers (Fig. 2). The placement was based on the geometries of prototype interactions (Fig. 3). The same urea placement, oriented perpendicular to the C–O–C plane, was used in every case. However, the three host–guest fragments for each of the two guest conformations differ in the placement of the amide groups. In each case, both amides are

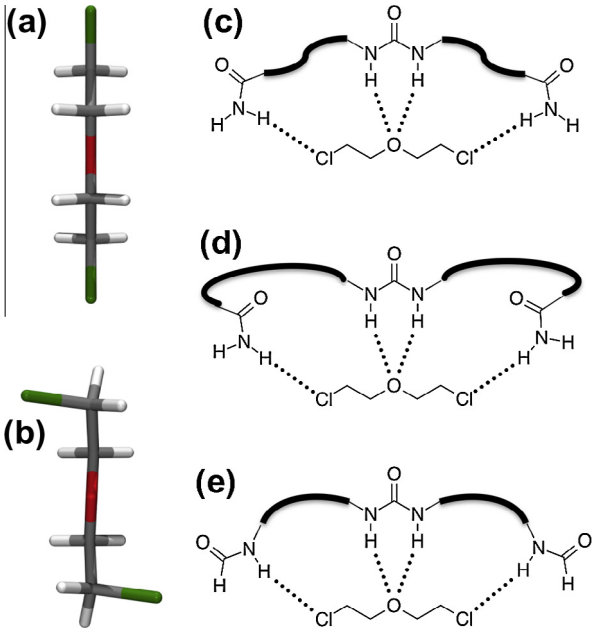

Fig. 2. MM3 optimized geometries for low energy, C_2 symmetric OM conformers with relative energies of 0 (a) and 1.04 kcal/mol (b). Design strategies for binding the OM guest with a central urea group linked to (c) amide carbon atoms with hydrogen bonding to the *trans* N–H group, (d) amide carbon atoms with hydrogen bonding to the *cis* N–H group, or (d) the *cis* positions of amide nitrogen atoms.