Inorg. Chem. **2002**, *41*, 5502–5512

HostDesigner: A Program for the de Novo Structure-Based Design of Molecular Receptors with Binding Sites that Complement Metal Ion Guests

Benjamin P. Hay* and Timothy K. Firman

W. R. Wiley Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Washington 99352

Received April 24, 2002

This paper describes a novel approach to the discovery of host structures with binding sites that complement targeted metal ion guests. This approach uses a de novo structure-based design strategy that couples molecular building algorithms with scoring functions to prioritize candidate structures. The algorithms described herein have been implemented in a program called HostDesigner, the first structure-based design software specifically created for the discovery of metal ion receptors. HostDesigner generates and evaluates millions of candidate structures within minutes, rapidly identifying three-dimensional architectures that position binding sites to provide an optimal interaction with the metal ion.

I. Introduction

Control of metal binding affinity is critical in the development of sensors, separating agents, improved analytical techniques, homogeneous catalysis, imaging agents, encapsulation of radionuclides used in cancer treatments, therapeutic agents for the treatment of metal intoxication, and models for the study of enzyme function. Extensive research effort has been expended toward understanding how the host structure influences metal ion binding with the goal of discovering more effective and more selective metal ion receptors.[1−4] It is clear from these studies that certain properties are needed to achieve significant increases in binding affinity or ion recognition. These host properties include (i) the presence of multiple binding sites,[1,2] (ii) the ability to adopt a conformation in which all binding sites are positioned to complement the metal ion structurally,[5] and (iii) a limited degree of conformational freedom.[6] A rational first step in the search for improved host architectures is to identify connectivity that yields complementary three-dimensional arrays for groups of potential binding sites.

Efficient computational methods for ranking proposed host structures in terms of their complementarity for specific metal ion guests on a case-by-case basis are available.[5,7,8] However, although understanding the nature of metal ion−binding site interactions and screening candidate structures is an important component of host design, a vital piece to the puzzle is still missing: a way to efficiently generate new host candidates. The deliberate design of host structures through the assembly of sets of disconnected binding sites in three dimensions is not a trivial task. At present, trial structures can be generated only by hand with a graphical user interface, an extremely time-consuming process. Often, it is not readily obvious which linkage structures might be best used to connect the

* To whom correspondence should be addressed. E-mail: ben.hay@pnl.gov.

(1) (a) Hancock, R. D.; Martell, A. E. *Comments Inorg. Chem.* **1988**, *6*, 237. (b) Hancock, R. D.; Martell, A. E. *Chem. Rev.* **1989**, *89*, 1875. (c) Martell, A. E.; Hancock, R. D. *Coord. Chem. Rev.* **1994**, *133*, 39. (d) Martell, A. E.; Hancock, R. D. In *Metal Complexes in Aqueous Solution*; Fackler, J. P., Ed.; Plenum Press: New York, 1996.

(2) Schneider, H.-J.; Yatsimirsky, A. *Principles and Methods in Supramolecular Chemistry*; John Wiley and Sons: New York, 2000.

(3) Martell, A. E.; Smith, R. D. *Critical Stability Constants*; Plenum Press: New York, 1974−1988; Vol. 1−6.

(4) (a) Chistensen, J. J.; Eatough, D. J.; Izatt, R. M. *Chem. Rev.* **1974**, *74*, 351. (b) Izatt, R. M.; Bradshaw, J. S.; Nielson, S. A.; Lamb, J. D.; Christensen, J. J. *Chem. Rev.* **1985**, *85*, 271. (c) Izatt, R. M.; Pawlak, K.; Bradshaw, J. S.; Bruening, R. L. *Chem. Rev.* **1991**, *91*, 1721. (d) Izatt, R. M.; Pawlak, K.; Bradshaw, J. S.; Bruening, R. L. *Chem. Rev.* **1995**, *95*, 2529.

(5) Hay, B. P.; Hancock, R. D. *Coord. Chem. Rev.* **2001**, *212*, 61.

(6) (a) Busch, D. H.; Farmery, K.; Goedken, V.; Katovic, V.; Melnyk, A. C.; Sperati, C. R.; Tokel, N. *Adv. Chem. Ser.* **1971**, *100*, 44. (b) McDougall, G. J.; Hancock, R. D.; Boeyens, J. C. A. *J. Chem. Soc., Dalton Trans.* **1978**, 1438. (c) Anicini, A.; Fabbrizzi, L.; Paoletti, P.; Clay, R. M. *J. Chem. Soc., Dalton Trans.* **1978**, 577. (d) Cram, D. J.; Kaneda, T.; Helgeson, R. C.; Brown, S. B.; Knobler, C. B.; Maverick, E.; Trueblood, K. N. *J. Am. Chem. Soc.* **1985**, *107*, 3645. (e) Stack, T. D. P.; Hou, Z.; Raymond, K. N. *J. Am. Chem. Soc.* **1993**, *115*, 6466.

10.1021/ic0202920 CCC: $22.00
Published on Web 09/26/2002

binding sites to obtain a host cavity that compliments a targeted metal ion guest.

Drug designers have developed computational approaches to address the inverse of this problem, that is, the identification of molecular structures (guests) that will complement the binding site of a protein (host).[9] These approaches include de novo structure-based design strategies that couple molecule building algorithms with scoring functions that are used to prioritize the candidate structures. The building algorithms assemble guest molecule structures that can physically interact with a known protein structure from pieces that are either atoms[10] or larger, chemically reasonable fragments.[11] The ability to generate large numbers of potential guest structures necessitates the use of simple scoring functions to prioritize the output. To this end, methods have been developed to estimate the binding free energy by summing free energy increments for hydrogen-bond interactions, ionic interactions, lipophilic interactions, the number of rotatable bonds in the guest molecule, etc.[12] After an initial prioritization of the results, computationally demanding evaluations of the host−guest complex can be used to achieve a more accurate ranking of the best candidates.

Although there are many prior applications of de novo structure-based design programs in the medicinal field,[9−11] we are unaware of any prior reports on the application of this strategy for the design of metal ion hosts. Computer programs that have been developed to perform de novo structure-based drug design are, in general, not applicable to metal receptors. These programs require input of the atomic coordinates of a protein binding site, are highly specialized to address protein−organic interactions, and do not contain scoring functions to address metal−ligand interactions. To bring the powerful concepts embodied in de novo structure-based drug design to the field of coordination chemistry, we have devised computer algorithms for building structures from host components and rapid methods for scoring the resulting structures with respect to their complementarity for the targeted guest. The result is HostDesigner, the first structure-based design software that is specifically created for the discovery of novel metal ion receptors. Herein, we describe the host building and scoring algorithms and provide several examples to demonstrate their usage.

II. Computational Methods

The algorithms described herein have been implemented in a program called HostDesigner.[13] The program consists of two main modules written in Fortran 77, LINKER and OVERLAY, and a library of linking fragments. LINKER builds structures by connecting two user-defined complex fragments with linking fragments from the library. Its output consists of an ASCII file containing Cartesian coordinates for a series of host structures presented in order of decreasing complementarity for the guest metal ion. OVERLAY builds structures by superimposing linking fragments from the library onto a single user-defined complex structure. Its output consists of an ASCII file containing Cartesian coordinates for a series of host structures presented in order of decreasing quality of the superposition.

HostDesigner was used to perform all of the LINKER and OVERLAY calculations presented in this paper. To illustrate the speed of program execution, we report the execution time for each example on three different platforms. These are indicated as follows: MAC (Macintosh G4 computer, 800-MHz processor, MacOS 9.2 operating system, Absoft Fortran 4.5 compiler), WIN (Athlon XP computer, 1467-MHz processor, Windows 98 operating system, Compaq Visual Fortran 6 compiler), and SUN (SunBlade 1000 computer, 700-MHz processor, Solaris 8.0 operating system, Sun Fortran 77 version 5.0 compiler).

Molecular mechanics calculations were used to optimize the structures of the molecular fragments that were used to build the library, to generate the potential energy surfaces for bond rotations that were used for dihedral angle assignments, and to evaluate candidate host structures generated by HostDesigner. All molecular

(7) (a) Hay, B. P.; Zhang, D.; Rustad, J. R. *Inorg. Chem.* **1996**, *35*, 2650. (b) Hay, B. P. In *Metal-Ion Separation and Preconcentration, Progress and Opportunities*; Bond, A. H., Dietz, M. L., Rogers, R. D., Eds.; ACS Symposium Series 716; American Chemical Society: Washington, DC, 1999; pp 102−113. (c) Hay, B. P. In *Metal Separation Technologies Beyond 2000: Integrating Novel Chemistry with Processing*; Liddell, K. C., Chaiko, D. J., Eds.; Minerals, Metals, Materials Society: Warrendale, PA, 1999; pp 3−13. (d) Sachleben, R. A.; Moyer, B. A. In *Metal-Ion Separation and Preconcentration, Progress and Opportunities*; Bond, A. H., Dietz, M. L., Rogers, R. D., Eds.; ACS Symposium Series 716; American Chemical Society: Washington, DC, 1999; pp 114−132. (e) Bond, A. H.; Chiarizia, R.; Huber, V. J.; Dietz, M. L.; Herlinger, A. W.; Hay, B. P. *Anal. Chem.* **1999**, *71*, 2757. (f) Dietz, M. L.; Bond, A. H.; Hay, B. P.; Chiarizia, R.; Huber, V. J.; Herlinger, A. W. *Chem. Commun.* **1999**, *13*, 1177. (g) Hay, B. P.; Dixon, D. A.; Vargas, R.; Garza, J.; Raymond, K. N. *Inorg. Chem.* **2001**, *40*, 3922. (h) Lumetta, G. J.; Rapko, B. M.; Garza, P. A.; Hay, B. P.; Gilbertson, R. D.; Weakley, T. J. R.; Hutchison, J. E. *J. Am. Chem. Soc.* **2002**, *124*, 5644.

(8) (a) Comba, P.; Hambley, T. W. *Molecular Modeling of Inorganic Compounds*, 2nd ed.; VCH: New York, 2001; Chapter 9. (b) Comba, P. *Coord. Chem. Rev.* **1999**, *185*−*186*, 81.

(9) (a) Kuntz, I. D.; Meng, E. C.; Shoichet, B. K. *Acc. Chem. Res.* **1994**, *27*, 117. (b) Lybrand, T. P. *Curr. Opin. Struct. Biol.* **1995**, *5*, 224. (c) Böhm, H.-J. *Prog. Biophys. Mol. Biol.* **1996**, *66*, 197. (d) Joseph-McCarthy, D. *Pharm. Ther.* **1999**, *84*, 179.

(10) (a) Nishibata, Y.; Itai, A. *Tetrahedron* **1991**, *47*, 8985. (b) Nishibata, Y.; Itai, A. *J. Med. Chem.* **1993**, *36*, 2921. (c) Rotstein, S. H.; Murcko, M. A. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 23. (d) Bohecek, R. S.; McMartin, C. *J. Am. Chem. Soc.* **1994**, *116*, 5560. (e) Gehlhaar, D. K.; Moerder, K. E.; Zichi, D.; Sherman, C. J.; Ogden, R. C.; Freer, S. T. *J. Med. Chem.* **1995**, *38*, 466. (f) Luo, Z.; Wang, R.; Lai, L. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1187.

(11) (a) Böhm, H.-J. *J. Comput.-Aided Mol. Des.* **1992**, *6*, 61. (b) Lawrence, M. C.; davis, P. C. *Proteins: Struct. Funct. Genet.* **1992**, *12*, 31. (c) Ho, C. M. W.; Marshall, G. R. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 623. (d) Rotstein, S. H.; Murcko, M. K. *J. Med. Chem.* **1993**, *36*, 1700. (e) Tschinke, V.; Cohen, N. C. *J. Med. Chem.* **1993**, *36*, 3863. (f) Gillet, V. J.; Newell, W.; Mata, P.; Myatt, G. J.; Sike, S.; Zsoldos, Z.; Johnson, A. P. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 207. (g) Leach, A. R.; Kilvington, S. R. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 283. (h) Mata, P.; Gillet, V. J.; Johnson, A. P.; Lampreia, J.; Myatt, G. J.; Sike, S.; Stebbings, A. L. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 470. (i) Roe, D. C.; Kuntz, I. D. *J. Comput.-Aided Mol. Des.* **1995**, *9*, 269. (j) Wang, R. X.; Gao, Y.; Lai, L. H. *J. Mol. Model.* **2000**, *6*, 498.

(12) (a) Murcko, A.; Murcko, M. A. *J. Med. Chem.* **1995**, *38*, 4953. (b) Head, R. D.; Smythe, M. L.; Oprea, T. I.; Waller, C. L.; Green, S. M.; Marshall, G. R. *J. Am. Chem. Soc.* **1996**, *118*, 3959. (c) Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 425. (d) Baxter, C. A.; Murray, C. W.; Clark, D. E.; Westhead, D. R.; Eldridge, M. D. *Proteins: Struct. Funct. Genet.* **1998**, *33*, 367. (e) Wang, R.; Liu, L.; Lai, L.; Tang, Y. *J. Mol. Model.* **1998**, *4*, 379. (f) Böhm, H.-J.; Schneider, G. *Drug Discuss. Today* **2002**, *7*, 64.

(13) HostDesigner, version 1.0, was developed by B. P. Hay and T. K. Firman in the William R. Wiley Environmental Molecular Sciences Laboratory at the Pacific Northwest National Laboratory under partial sponsorship of Chemical Sciences, Office of Basic Energy Sciences, Office of Science, U.S. Department of Energy, under Contract DE-AC06-76RL01830. The software can be obtained at no charge from the following website: http://hostdesigner.emsl.pnl.gov.

mechanics calculations were performed with the MM3(96) program[14] that has been modified for application to metal complexes.[15] The default MM3 parameter set was used for calculations on alkanes and alkenes.[16] A modified MM3 parameter set was used to perform calculations on the lithium ether complexes.[17]

III. The LINKER Algorithm

Complex Fragments. A multidentate host can be dissected into two or more simpler host components. For example, the well-known 18-crown-6 macrocycle can be broken down into two triglyme components, three diglyme components, or six dimethyl ether components. It is possible to define the structure of a complex fragment, that is, a piece of a host−guest complex, by combining a host component with a guest metal ion. In constructing the complex fragment, the metal ion is positioned relative to the host component to define a complementary geometry, that is, a geometry that would give the strongest interaction between the binding sites of the host component and the metal ion in an actual complex. Examples of complex fragments are shown in Figure 1.

When a host component contains a single binding site, such as the oxygen in dimethyl ether or the nitrogen in ethylamine, complementary placement of the metal ion can be described in terms of the M−L distance, the M−L−X bond angles, and the M−L−X−X dihedral angles (M = metal ion, L = donor atom, and X = any other atom).[5] Because M−L distances depend on the number, placement, and type of other donor atoms in the metal ion coordination sphere, the assignment of an optimal M−L distance for a complex fragment must be based on a consideration of the environment that will exist about the metal ion in an actual host−guest complex. Although it might not be possible to define rigorously a single complementary M−L distance in a complex fragment, it is possible to assume a reasonable value on the basis of an examination of M−L distances in crystal structures. For example, the Li−O(ether) distances in five- and six-coordinate complexes exhibit an average value of 2.26 ± 0.05 Å,[17] and Co(III)−N(amine) distances in six-coordinate complexes exhibit an average value of 1.96 ± 0.02 Å.[18]

Optimal M−L−X angles and M−L−X−X dihedral angles can be deduced through an examination of experimental geometries of metal-coordinated ligands or through the careful application of electronic structure calculations.[7g,17−22]

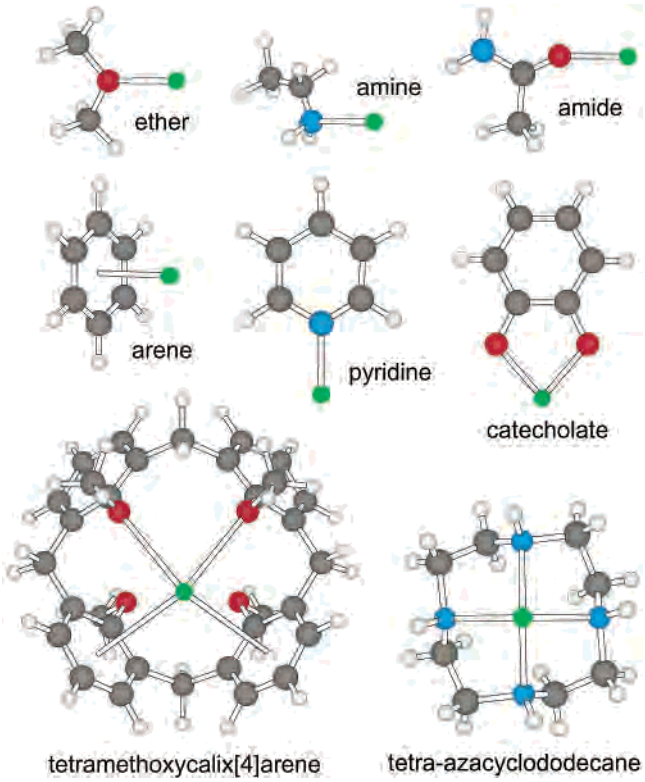

Figure 1. Complex fragments built by combining metal ion guests, indicated by green spheres, with host components.

These structural parameters reflect the preferred geometry at the donor atom. The geometry at the donor atom is often relatively insensitive to the identity of the metal ion or the M−L distance. For example, the aliphatic ether oxygen donor atom exhibits a trigonal planar coordination preference,[17,19] and the amine nitrogen donor atom exhibits a tetrahedral coordination preference.[18,20] However, in some cases, such as the amide oxygen donor atom, the donor orientation might depend significantly on the identity of the metal ion.[21]

Complex fragments can also be constructed from host components containing multiple binding sites. In these cases, it might not be possible to position a metal ion such that each binding site achieves an optimal geometry with the metal ion as defined above because of geometrical constraints imposed by the connecting structure. However, it is possible to position the metal ion to best complement the binding sites in the host component by evaluation of structural data obtained from calculations or experiment. Examples shown in Figure 1 include an Fe(III) complex with catecholate where the geometry is taken from the $[Fe(catecholate)_3]^{3-}$ complex optimized with electronic structure calculations,[7g] a cesium complex with tetramethoxycalix[4]arene where the geometry was optimized with electronic structure calculations,[23] and a Cr(III) complex with tetra-azacyclododecane, L, where the

(14) MM3(96) can be obtained from Tripos Associates, 1699 S. Hanley Road, St. Louis, MO 63144, for commercial users, and it can be obtained from the Quantum Chemistry Program Exchange, Mr. Richard Counts, QCPE, Indiana University, Bloomington, IN 47405, for noncommercial users.
(15) Hay, B. P.; Yang, L.; Allinger, N. L.; Lii, J.-H. *J. Mol. Struct. (THEOCHEM)* **1998**, *428*, 203.
(16) (a) Allinger, N. L.; Yuh, Y. H.; Lii, J.-H. *J. Am. Chem. Soc.* **1989**, *111*, 8551. (b) Lii, J.-H.; Allinger, N. L. *J. Am. Chem. Soc.* **1989**, *111*, 8566. (c) Lii, J.-H.; Allinger, N. L. *J. Am. Chem. Soc.* **1989**, *111*, 8576. (d) Allinger, N. L.; Li, F.; Yan, L. *J. Comput. Chem.* **1990**, *11*, 848. (e) Allinger, N. L.; Li, F.; Yan, L.; Tai, J. C. *J. Comput. Chem.* **1990**, *11*, 868.
(17) Hay, B. P.; Rustad, J. R. *J. Am. Chem. Soc.* **1994**, *116*, 6316.
(18) Hancock, R. D. *Prog. Inorg. Chem.* **1989**, *37*, 187.
(19) Hancock, R. D.; Maumela, H.; de Sousa, A. S. *Coord. Chem. Rev.* **1996**, *148*, 315.
(20) Bernhardt, P. V.; Comba, P. *Inorg. Chem.* **1992**, *31*, 2638.
(21) Hay, B. P.; Clement, O.; Sandrone, G.; Dixon, D. A. *Inorg. Chem.* **1998**, *37*, 5887.
(22) (a) Nicholas, J. B.; Dixon, D. A.; Hay, B. P. *J. Phys. Chem. A* **1999**, *103*, 1394. (b) Nicholas, J. B.; Hay, B. P. *J. Phys. Chem. A* **1999**, *103*, 9815.
(23) Hay, B. P.; Nicholas, J. B.; Feller, D. *J. Am. Chem. Soc.* **2001**, *122*, 10083.

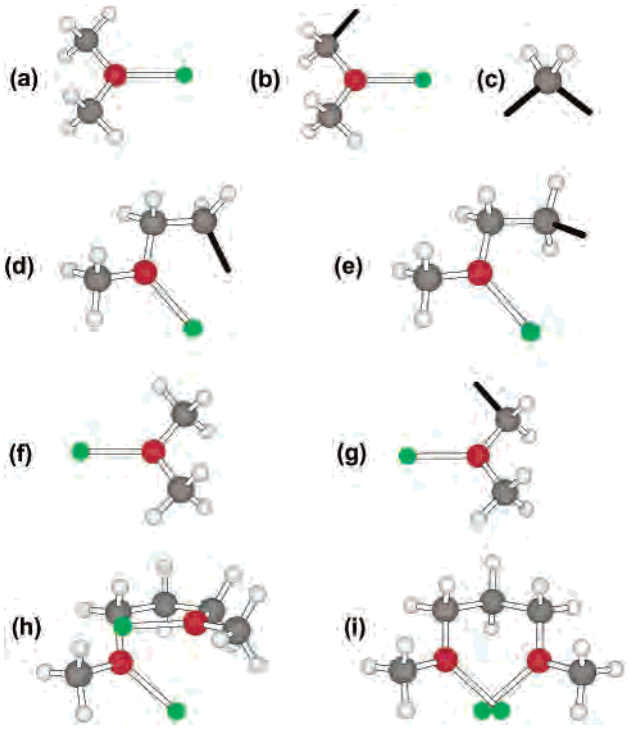

Figure 2. Process used by LINKER to construct a new host molecule. See text for a description of steps a−i.

geometry is taken from a crystal structure of the $[Cr_2(OH)_2(L)_2]^{4+}$ complex.[24]

Assembling the Pieces. LINKER builds new host structures by forming C−C bonds between pairs of complex fragments and linking fragments taken from a library (vide supra). The user must provide an input file for each complex fragment that specifies the coordinates for all of the atoms, atom connectivity, and attachment vectors. Attachment vectors are indicated by listing $C(sp^3)-H$ or $C(sp^2)-H$ bonds of the complex fragment.

The process used by LINKER to construct a new host molecule is illustrated in Figure 2. In this example, two identical lithium−dimethyl ether complex fragments are attached to a methylene linkage. The steps are as follows: (a) the first complex fragment is selected; (b) one of the hydrogens is removed, leaving a bonding vector; (c) a linking fragment containing two bonding vectors is selected from the library; (d) a bond is formed between the first complex fragment and the linking fragment by aligning the bonding vectors and setting the C−C distance to an appropriate value based on the hybridization of the carbon atoms; (e) the dihedral angle about the new bond is adjusted to a specific value based on hybridization and degree of substitution of the carbon atomsl; (f) the second complex fragment is selected (g) one of the hydrogens is removed, leaving a bonding vector; (h) a bond is formed between the second complex fragment and the remaining bonding vector on the linking fragment by aligning the bonding vectors and setting the C−C distance to an appropriate value based on the hybridization of the carbon atoms; and (i) the dihedral angle about the new bond is adjusted to a specific value based on hybridization and degree of substitution of the carbon atoms.

(24) Hodgson, D. J.; Pedersen, E.; Toftlund, H.; Weiss, C. *Inorg. Chim. Acta* **1986**, *120*, 177.

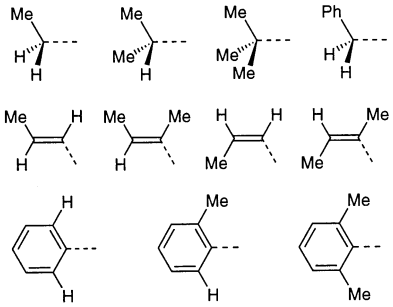

Figure 3. Groups used to assign rotational minima about C−C bonds.

It is generally possible to build a large number of host structures from a single linking fragment. The ability to define multiple attachment vectors on each complex fragment gives rise to the potential for different connectivities. The presence of chirality in either the complex fragments or the linking fragment gives rise to the potential for stereoisomers. Finally, the presence of multiple rotational minima for the bonds formed between complex fragments and linking fragments gives rise to multiple conformers of each connectivity. LINKER has been designed to build every possible connectivity that can be made from the two complex fragments with a given linking fragment including linkage isomers. In addition, when either the complex fragments or the linkage fragment is chiral, then LINKER will examine all possible stereoisomers that can be made by inverting each chiral fragment. LINKER also examines every conformation that can be generated for each connectivity by rotation about the C−C bonds that are formed between the linking fragment and the complex fragments. The dihedral angle values used for these rotations are based on an examination of MM3 potential energy surfaces for rotation about the 66 possible C−C bonds that can be formed by all combinations of the 11 groups shown in Figure 3.

Consider the simple example shown in Figure 2. If we define the three C−H bonds on one of the methyl groups in each complex fragment as attachment vectors and we consider each of the three rotamers for the two $C(sp^3)-C(sp^3)$ bonds that are formed, 81 host structures can potentially be made with the methylene linkage. In this case, all of these host structures are conformers of the same molecule, as the connectivity remains constant. LINKER generates all 81 structures, but it will retain only the first 9 structures shown in Figure 4. Some of the potential structures are rejected because of close contacts between nonbonded atoms, which indicates a physically unreasonable collision or superposition of atoms (see Figure 4). In addition, some of the structures are rejected because they are duplicates, structures that are either identical to or nonsuperimposable mirror images of previously generated structures. In other words, LINKER retains only one member of a pair of enantiomers.

Scoring the Results. During the construction of each complex fragment, a metal ion is positioned relative to a host component to define a complementary geometry with the binding sites in that host component. When two complex

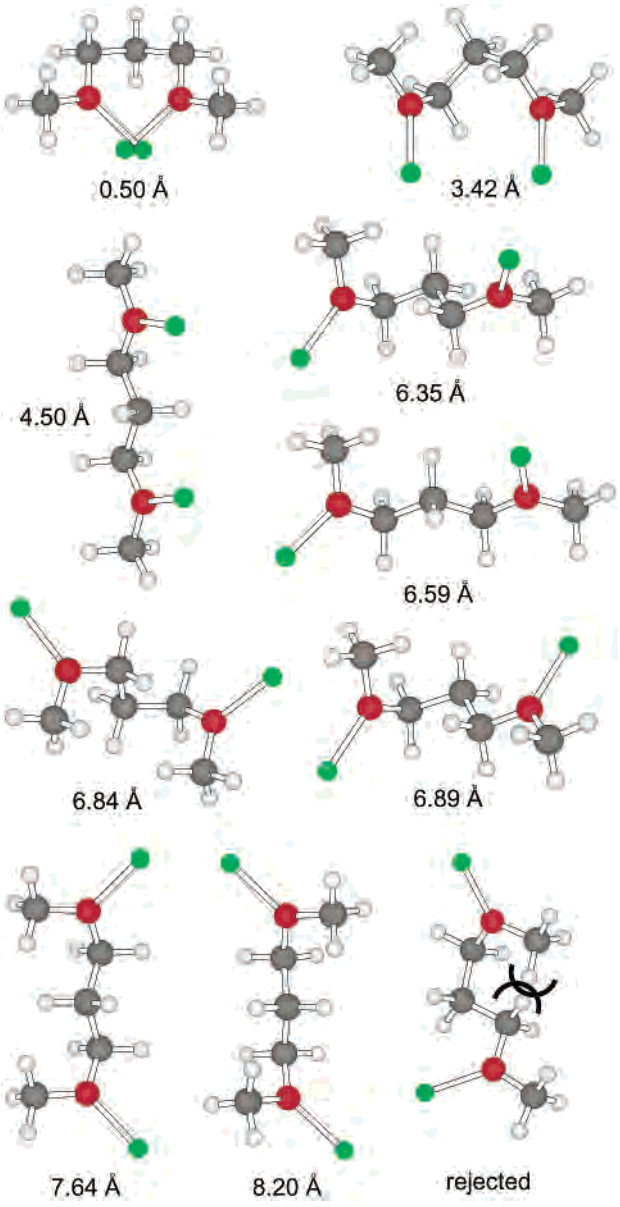

Figure 4. Nine unique structures that are generated by combination of two lithium−dimethyl ether complex fragments with the methylene linking fragment and an example of one of the structures that was rejected because of close contacts between nonbonded atoms (bottom right). The M−M distance is given for each structure.

fragments are combined, the distance between the two metal ions provides a simple criterion for the rapid evaluation of the degree of complementarity in the new host. Optimal complementarity would be obtained when the M−M distance is zero, that is, when the two metal ions representing the optimal bonding orientation with each host component are exactly superimposed. In the example given in Figure 4, where M−M distances range from 0.50 to 8.20 Å, the host structure with the shortest M−M distance clearly gives the most complementary placement of the two ether binding sites. LINKER uses the M−M distance to score the generated host structures and outputs Cartesian coordinates for each structure in the order of increasing M−M distance, that is, in order of decreasing complementarity for the guest metal ion.

Assuming that the geometries of the complex fragments have been accurately defined, then host structures that superimpose the two metal ions will, by definition, have complementary binding sites with respect to the M−L distances, the M−L−X bond angles, and the M−L−X−X dihedral angles. This definition of structural complementarity is sufficient for metal ion guests that do not exhibit bonding directionality such as the group 1A and 2A metals and the trivalent lanthanides and actinides. However, some metal ions, such as Cu(II), Pt(II), and Pd(II), exhibit distinct metal-centered bonding directionalities. In these cases, a complementary host architecture must also provide an array of donor atoms to produce L−M−L angles that correspond to the topography of the metal ion.[5] To address this issue, an optional screening capability is available in LINKER to retain only those hosts in which the donor atoms are located near the verticies of idealized polyhedra. Any one of the following stereochemistries can be specified: tetrahedral, square planar, square pyramidal, trigonal bipyramidal, or octahedral.

The Linking Fragment Library. The linking fragment library is a file from which LINKER reads the Cartesian coordinates and attributes of linking fragments that are used to connect the two complex fragments. Each linking fragment is a three-dimensional molecular structure with two specified binding vectors. In building the initial library, we decided to (i) limit the entries to molecules containing hydrogen and up to six carbon atoms, (ii) limit carbon hybridization to sp^2 and sp^3, and (iii) exclude three- and four-membered rings. This gave a total of 81 connectivities. Subsequently, 19 additional connectivities representing rigid bi- and tricyclic systems were added. These connectivities (Figure 5) include the null case, which is used when the first complex fragment is directly bonded to the second complex fragment.

It is possible to generate a number of linking fragments from a single connectivity by using the process illustrated in Figure 6. The steps are as follows: (a) select a connectivity, (b) perform a search to locate all stable conformers using the MM3 program, (c) choose one conformer and remove a pair of hydrogen atoms to generate a pair of bonding vectors, (d) place methyl groups at the ends of the bonding vectors, (e) optimize the structure with MM3, and (f) remove the methyl groups to obtain the final linking fragment with two bonding vectors.

Steps c−f are repeated for all conformers and for all possible hydrogen atom pairs. By optimizing each structure with methyl groups attached to the bonding vectors, the linking fragments more accurately reflect the geometries that should be present when attached to carbon substituents. Each new linking fragment is retained if it is unique or rejected if it is a duplicate of a previously generated linking fragment. In the example shown in Figure 6, *n*-butane has two conformers, and there are 45 hydrogen pairs per conformer. Thus, a total of 90 structures were processed for this connectivity, yielding 41 unique linking fragments after removal of duplicates. A total of 8552 linking fragments were prepared by performing this process on all 100 connectivities shown in Figure 5.

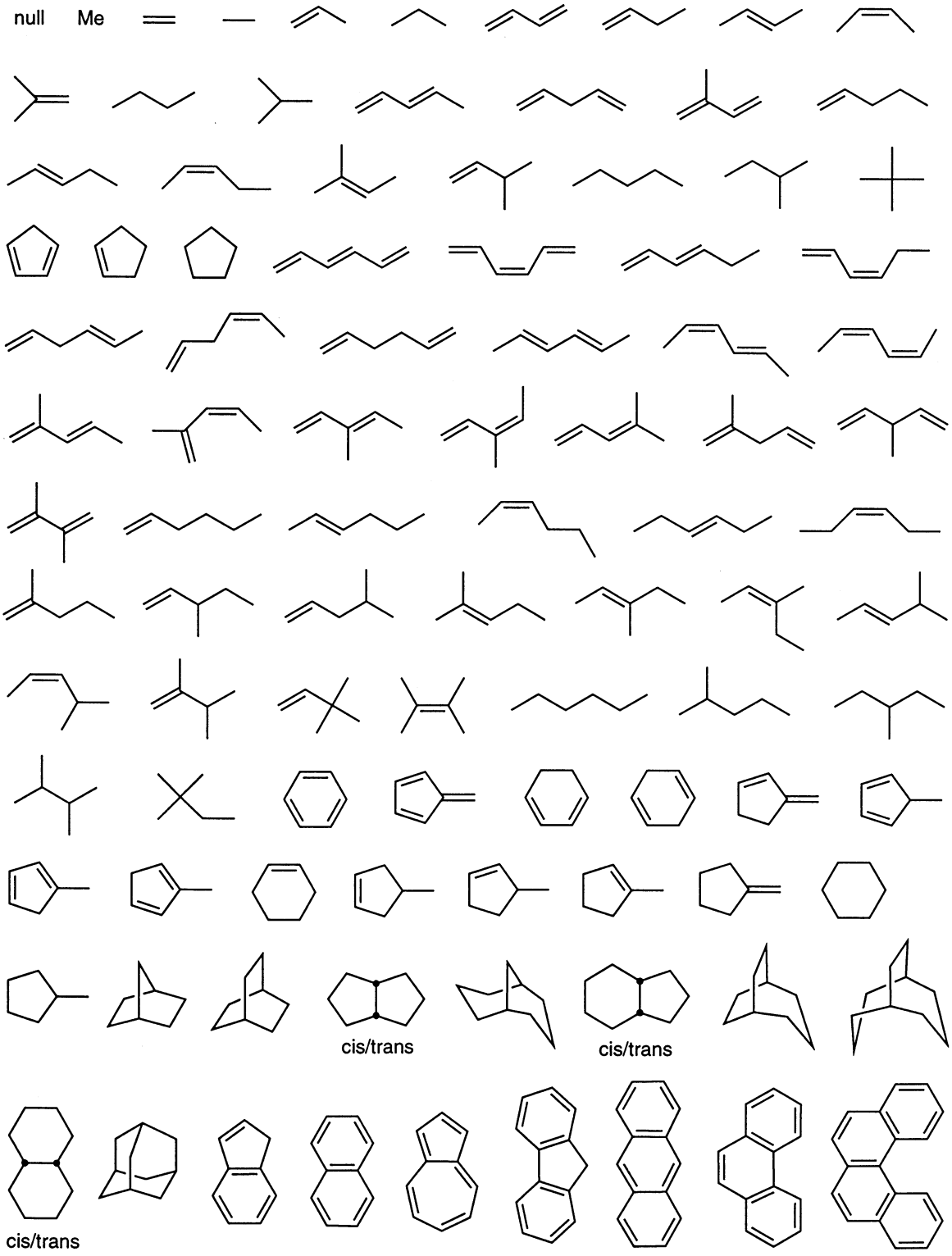

Figure 5. Connectivities used to construct the linking fragment library.

Initially, all new linking fragments were added to the library as they were received. However, as the size of the library increased, we encountered a problem with this approach. Many linking fragments have very similar bonding vectors. For example, the 1,1-ethane linkage has bonding vectors that are almost identical to those of the 1,1-propane, 1,1-butane, 1,1-pentane, and 1,1-hexane linkages. Thus, if a 1,1-alkane linking fragment gives a complementary host architecture, the code will generate an entire family of 1,1-alkane-linked structures that will dominate the output.

To overcome this problem, we grouped the linking fragments into classes based on geometric similarities of the bonding vectors. The current library of 8552 linking fragments represents a total of 2950 classes. Input constraints control the number of linking fragments that will be selected from the library for host building. To avoid the similarity

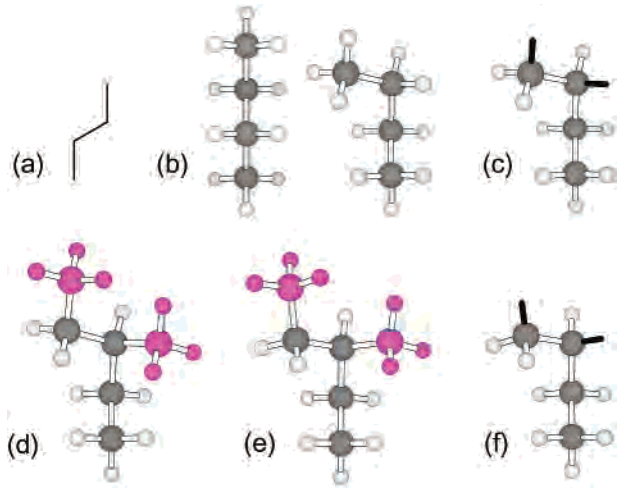

Figure 6. Process used to construct a linking fragment. See text for a description of steps a−f.

Table 1. Analysis of Host Structures (**1**−**6** in Figure 8 and **7**−**10** in Figure 9) Generated by LINKER[a]

host	M−M	RMSD(complex)	ΔU_{comp}	RMSD(host)
1	0.28	0.12	0.23	0.13
2	0.29	0.11	0.21	0.09
3	0.34	0.09	0.10	0.02
4	0.34	0.08	0.10	0.02
5	0.38	0.10	0.08	0.01
6	0.44	0.11	0.14	0.11
7	0.13	0.08	0.28	0.08
8	0.15	0.09	0.16	0.02
9	0.17	0.10	0.18	0.10
10	0.41	0.18	0.18	0.04

[a] Units: distances in Å and energies in kcal/mol. M−M = distance used to rank the candidates in LINKER, RMSD(complex) = root-mean-squared distance for the superposition of all non-hydrogen atoms in the metal complex structures before and after MM3 optimization, ΔU_{comp} = steric energy of the bound form − steric energy of the binding conformer, and RMSD(host) = root-mean-squared distance of the superposition of all non-hydrogen atoms in the bound form and the binding conformer of the host.

problem discussed above, the selection can be limited to the first member in each class. This feature is used in all of the examples given below. In addition, the choice of linking fragments can be limited by specifying the minimum length of the shortest connecting chain, the maximum length for the longest connecting chain, and the valence of the bonding carbon atoms. Finally, when a linking fragment derives from a connectivity with more than one conformation, the selection can be limited to only those linkages made from the lowest-energy conformer.

Example LINKER Applications. We now present two examples to illustrate the use of LINKER. First, we revisit the combination of the two lithium−dimethyl ether complex fragments shown in Figure 2, this time using the full library and selecting the first linking fragment from each class. This run generated 188 214 host structures. The time required to perform these calculations is extremely short on all platforms, demonstrating the computational efficiency of the LINKER algorithm (platform, total CPU time in seconds): SUN, 8.9 s; MAC, 9.5 s; WIN 2.6 s.

The top six candidates, **1**−**6**, are shown in Figure 7, and their M−M distances are given in Table 1. To verify that the structures produced by LINKER are true minima on the

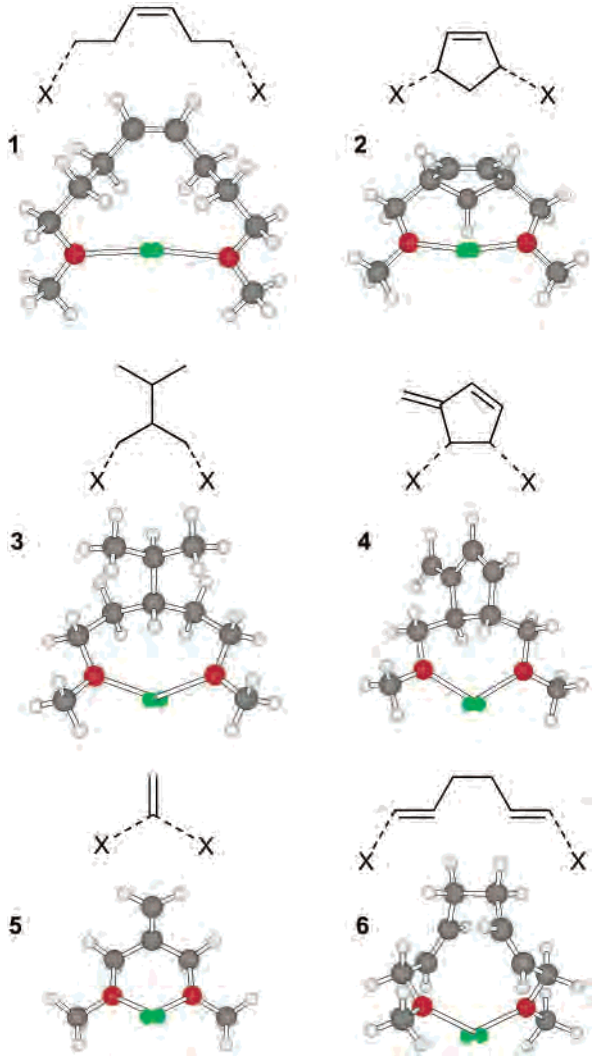

Figure 7. Linking fragment connectivities and structures of the top six hosts obtained using LINKER to combine two lithium−dimethyl ether complex fragments.

MM3 potential surface, each of the structures was optimized with MM3 after the two original lithium ions were replaced with a single lithium ion. In each case, the new lithium ion was placed at the midpoint between the two original lithium ions. Superpositions of the optimized structures on those generated by LINKER (Figure 8) reveal only small differences in the structures. The root-mean-squared deviation (RMSD) of all non-hydrogen atoms was calculated for each of the superpositions, and the values are given in Table 1. The largest RMSD value is 0.12 Å.

Further MM3 calculations were performed to verify that **1**−**6** actually represent complementary host architectures. One method for evaluating the degree of structural complementarity in a host architecture involves calculating the difference in steric energy, ΔU_{comp}, between the bound form of the host and the binding conformer of the host.[7] These calculations are done by removing the metal ion from the MM3-optimized complex and performing a second geometry optimization starting from the bound host coordinates. Complementary architectures should give small ΔU_{comp} values; in other words, guest complexation should not induce

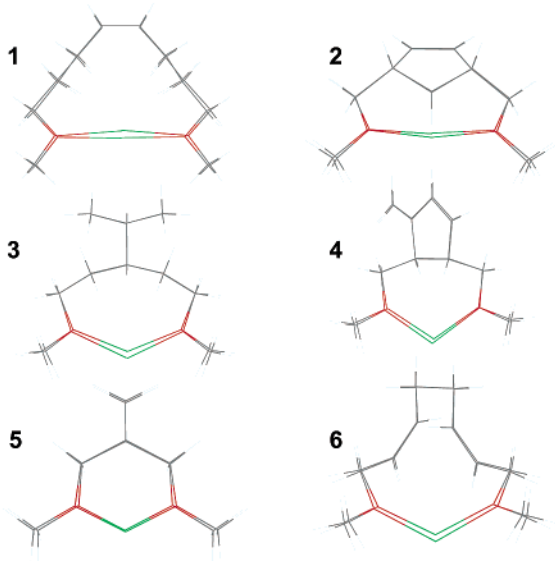

Figure 8. Superpositions of the lithium complexes of **1**−**6** before and after optimization with MM3.

steric strain within the binding conformer of the host. In addition to the change in steric energy, the extent of structural relaxation that occurs during these optimizations provides another measure of how well the host architecture complements the guest. This feature can be quantified by calculating the RMSD value for the superposition of the bound form of the host and the binding conformer of the host. Complementary architectures should give small RMSD values; in other words, guest complexation should not induce structural changes to the binding conformer of the host. The results of these analyses (Table 1) confirm that the structures **1**−**6** have a high degree of complementarity for lithium. The largest ΔU_{comp} value was only 0.23 kcal/mol, and the largest RMSD value was only 0.13 Å.

The preceding example revealed that a relatively simple linking fragment, derived from $CH_2{=}CH_2$, gave a complementary bidentate host for lithium, **5**. In the second example, we used LINKER to identify potential tetradentate ether hosts assembled from **5**. The MM3-optimized coordinates for the lithium complex of **5** were used to prepare the input files for the two complex fragments. All 12 C−H bonds were specified as attachment vectors. This run generated 3 430 744 structures (platform, CPU time): SUN, 116 s; MAC, 155 s; WIN, 38 s. The top four candidates, **7**−**10**, are shown in Figure 9. Subsequent analysis (see Table 1) again confirms that the host architectures generated by LINKER are (i) minima on the MM3 potential energy surface and (ii) complementary for lithium.

Macrocycles? During the early stages of development, it was not possible for LINKER to generate macrocyclic structures with only one linkage connecting the two complex fragments. As a result, potential macrocyclic host architectures were overlooked. For example, Figure 10 shows how cyclam, **12**, could be built by linking two *N*,*N*′-dimethyl-1,3-propanediamine chelate rings, **11**. However, directly connecting the *N*-methyl carbon of one chelate to the *N*-methyl carbon of the other chelate would give close

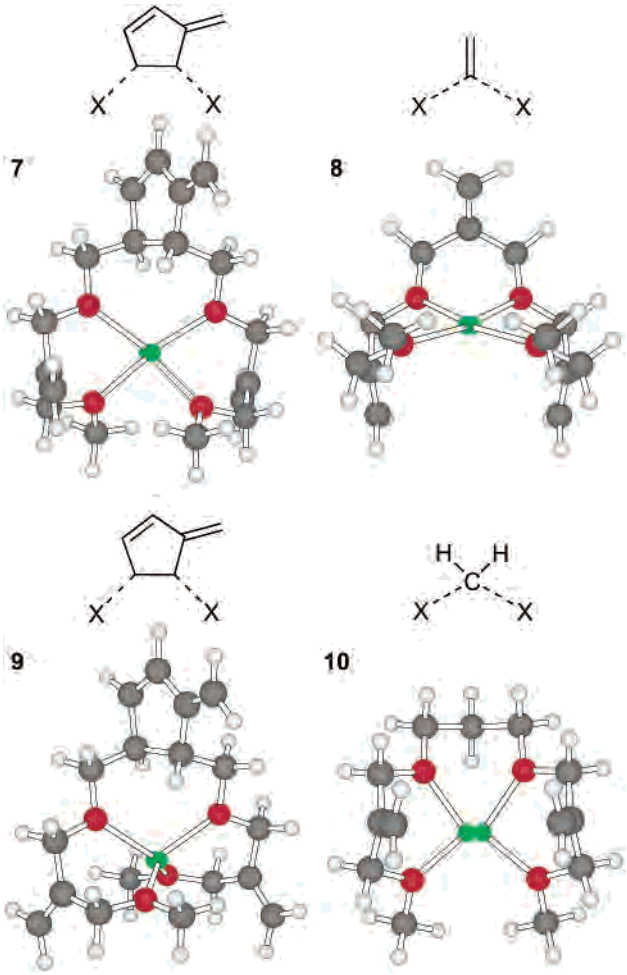

Figure 9. Linking fragment connectivities and structures of the top four hosts obtained using LINKER to combine two complex fragments derived from **5**.

contacts between the two *N*′-methyl groups, thus forcing LINKER to reject this candidate.

To overcome this problem, we added an optional process that checks whether rejected structures could become accepted structures if a C−C bond were formed between specified carbon atoms. With the macrocycle option invoked, LINKER was used to combine the two complex fragments, **11**, where the M−N distance was 2.10 Å. As in the previous examples, all possible C−H bonds were selected as attachment vectors. This run generated 4 264 500 structures (platform, CPU time): SUN, 192 s; MAC, 263 s; WIN, 73 s. With the stereochemistry of the input complex fragments (both + +), it is possible to generate two cyclam diastereomers, the + + + + form and the + + − − form.[25] With an M−M distance of 0.25 Å, LINKER identifies the + + − − isomer, **12**, as the most complementary host architecture. LINKER also locates the + + + + isomer, **13**, with an M−M distance of 1.15 Å. The results indicate **12** to be more complementary than **13**, consistent with conformer populations observed for Co(III) and Ni(II) complexes with 14-membered tetra-aza macrocycles.[26]

(25) Thom, V.; Fox, C. C.; Boeyens, J. C. A.; Hancock, R. D. *J. Am. Chem. Soc.* **1984**, *106*, 5947.

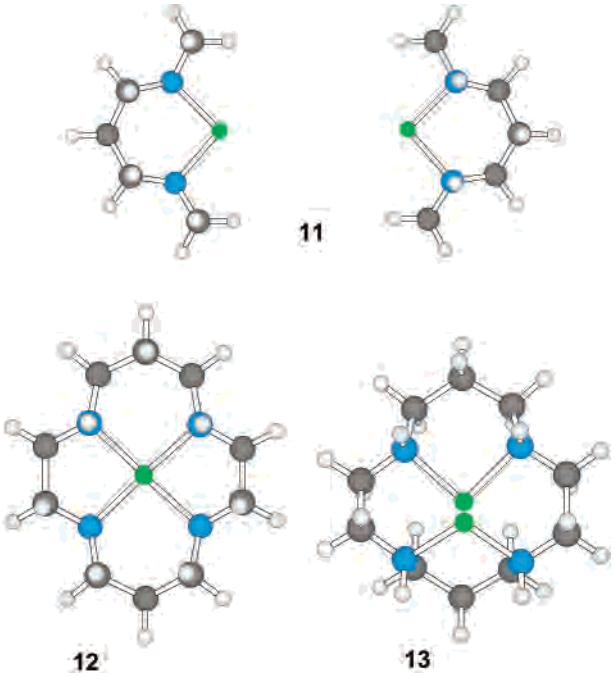

Figure 10. Two diastereomers of cyclam, the + + + + form (**12**) and a + + − − form (**13**) are formed by the combination of the two complex fragments (**11**).

Limitations. LINKER experiences two limitations that are common to all fragment-based building approaches: (i) the number of structures that can be generated is limited by the contents of the fragment library, and (ii) the synthesis of some candidate structures might be difficult or impossible. In addition, it is important not to overinterpret the ranking of the new host structures that are generated.

LINKER is designed to examine large numbers of potential structures in a short period of time and to identify connectivities and conformations that yield complementary architectures for metal ion complexation. It is, by necessity, an approximate method. Because the code is connecting rigid components and using generic dihedral angles, the ranking of the structures on the basis of the M−M distance should be regarded as approximate. When M−M distances differ by angstroms, then the difference between convergent and divergent binding sites is clear (see, for example, Figure 4). However, when M−M distances differ by less than an angstrom, the ranking becomes more uncertain. To obtain a definative ordering of host structures with respect to complementarity, the structures should be examined more carefully with more expensive MM or electronic structure methods.

Finally, it must be remembered that LINKER ranks structures only in terms of their complementarity for the metal ion, not their preorganization. In other words, LINKER does not address the issue of conformational freedom within the host. The top candidates might be highly flexible with many degrees of rotational freedom. Moreover, a highly complementary structure might also be a high-energy conformer. Thus, although a given host architecture might be complementary, it might fail to exhibit strong complexation because of a lack of preorganization. Further screening for preorganization can be accomplished by performing conformational analyses on promising architectures.[7a]

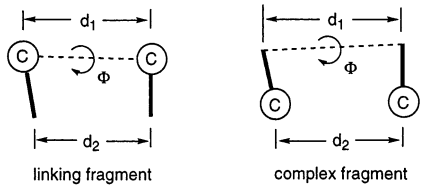

Figure 11. Parameters d_1, d_2, and Φ used to compare the bonding vectors in the linking fragment and the complex fragment.

IV. The OVERLAY Algorithm

Assembling the Pieces. OVERLAY builds new host structures by superimposing two bonding vectors on a single complex fragment with two bonding vectors on a linking fragment taken from the library (vide supra). The user must provide an input file for a single complex fragment that specifies the coordinates for all atoms, the atom connectivity, and the pairs of attachment vectors. As with LINKER, the attachment vectors are indicated by a list of $C(sp^3)-H$ or $C(sp^2)-H$ bonds.

The steps used by OVERLAY to construct a new host molecule are as follows: (a) select a linking fragment from the library; (b) adjust the lengths of the bonding vectors on the complex fragment and the linking fragment to the ideal length for the type of bond that would be formed; (c) compare geometries of the linking fragment vectors and the complex fragment vectors; (d) if the vector geometry is similar, then superimpose the bonding vectors of the linking fragment on the bonding vectors of the complex fragment to give the best overlay possible; and (e) form bonds between the attaching carbons on the complex fragment and the linking fragment.

The comparison of vector geometries in step c involves taking the difference in three geometric parameters shown in Figure 11. These are the distances, d_1 and d_2, and the dihedral angle, Φ. All three differences must be within user-defined tolerance limits. Smaller tolerance values give fewer results of higher quality. Larger tolerance limits give more results, but more of the structures might have distorted geometries. As a compromise, we use tolerance limits of 1.60 Å for the distances and 90° for the dihedral angles.

The resulting structure can still be rejected even if the bonding vectors of the linking fragment are perfectly superimposed on the bonding vectors of the complex fragment. Although a perfect superposition ensures optimal C−C distances and C−C−X bond angles, the X−C−C−X dihedral angles about each of the new C−C bonds could have any value (X = any atom). Thus, after a new structure has been built, OVERLAY checks the difference between the actual X−C−C−X dihedral angles and the dihedral angle corresponding to the nearest local minimum (vide supra). If the rotational periodicity is >4, all structures are accepted. Otherwise, the structure will be rejected if the difference in X−C−C−X dihedral angles is greater than a threshold value that depends on the periodicity of the rotational potential:

(26) (a) Cooper, C. G., Jr.; Zimmer, M. *Struct. Chem.* **1999**, *10*, 17. (b) Donnelly, M. A.; Zimmer, M. *Inorg. Chem.* **1999**, *38*, 1650.

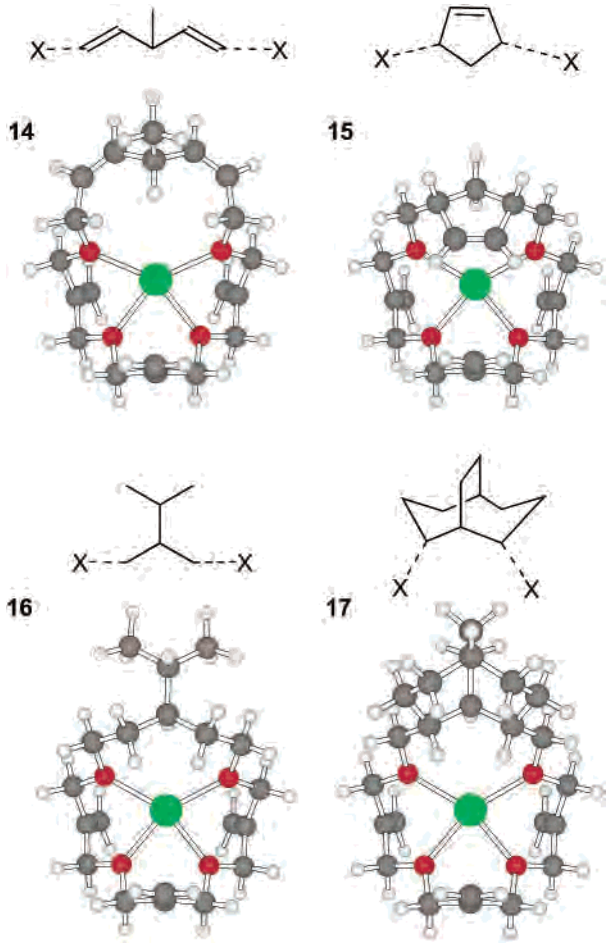

Figure 12. Linking fragment connectivities and structures of the top four hosts obtained using OVERLAY to form macrocyclic structures starting from **8**.

2-fold, 45°; 3-fold 30°; 4-fold 30°. Finally, the structure will be rejected if there are any close contacts between nonbonded atoms in the linking fragment and the complex fragment.

Scoring the Results. OVERLAY uses the RMSD for the superposition of the bonding vectors to score the host structures that are generated during a run and outputs Cartesian coordinates for each structure presented in order of increasing RMSD, that is, in order of decreasing quality of the superposition.

Example OVERLAY Applications. We now present two examples to illustrate the use of OVERLAY. In the first example, we use OVERLAY to search for linking fragments to form a macrocycle by connecting the terminal methyl groups of the tetradentate ether **8**. The MM3-optimized coordinates for the lithium complex of **8** were used to prepare the input files for the complex fragment. A single pair of attachment vectors was specified by selecting a C−H bond from each methyl group. This run generated 103 structures (platform, CPU time): SUN, 9.0 s; MAC, 7.7 s; WIN, 1.9 s. The top four candidates, **14**−**17**, are shown in Figure 12.

An MM3 analysis of these structures (vide supra) is presented in Table 2. The results confirm that the host architectures generated by OVERLAY are (i) minima on the MM3 potential energy surface and (ii) highly complementary for lithium. A comparison of the ΔU_{comp} values for **14**−**17**, which range from 0.2 to 0.4 kcal/mol, with those calculated for lithium complexation with 12-crown-4 (7.7 kcal/mol)[5] and alkylated 14-crown-4 derivatives (4.3−6.8 kcal/mol)[7b−d] suggests that the new architectures provide arrays of binding sites with significantly higher lithium complementarity than prior tetradentate aliphatic ether macrocycles.

Table 2. Analysis of Tetradentate Ether Macrocycles (Figure 12) Generated by OVERLAY[a]

host	RMSD(vectors)	RMSD(complex)	ΔU_{comp}	RMSD(host)
14	0.18	0.30	0.41	0.10
15	0.18	0.11	0.30	0.07
16	0.19	0.10	0.19	0.03
17	0.21	0.21	0.29	0.10

[a] Units: distances in Å and energies in kcal/mol. RMSD(vectors) = root-mean-squared distance after superimposing the ends of the bonding vectors used to rank the candidates in OVERLAY, RMSD(complex) = root-mean-squared distance for the superposition of all non-hydrogen atoms in the metal complex structures before and after MM3 optimization, ΔU_{comp} = steric energy of the bound form − steric energy of the binding conformer, and RMSD(host) = root-mean-squared distance of the superposition of all non-hydrogen atoms in the bound form and the binding conformer of the host.

Enhanced binding efficiency often results when two sets of binding sites are connected to form a host of higher denticity.[1] LINKER provides one approach for identifying optimal connections. An alternative approach is to use OVERLAY to add connectivity to an existing metal complex in which the binding sites are in a low-energy configuration. In the second example, we search for linking fragments to bridge two chelates in a $[U(catecholamide)_4]^{4-}$ complex. To prepare an input file for the complex fragment, four *N*-methylformyl substituents were added to the crystal structure coordinates of a dodecahedral $[U(catecolate)]^{4-}$ complex[27] to obtain ligand geometries similar to those observed in other chelated catecholamides.[7g] Four pairs of attachment vectors were specified using two C−H bonds from each *N*-methyl group. This run generated 64 structures (platform, CPU time): SUN, 18.2 s; MAC, 14.6 s; WIN, 4.0 s. With vector overlays of 0.07−0.15 Å, the top four candidates, **18**−**21**, are shown in Figure 13. The result suggests that a chain length of six carbons between the two nitrogen atoms is needed to give a host architecture that places the four oxygen donors on the four coplanar verticies of the dodecahedral polyhedron.

Limitations. OVERLAY suffers from most of the same limitations that pertain to LINKER: (i) the number of structures that can be generated is limited by the contents of the fragment library, (ii) the synthesis of some candidate structures might be difficult or impossible, and (iii) the conformational flexibility of the host structure is not addressed.

V. Summary

The ultimate goal of the present approach is the effective automated design of metal ion hosts. The molecular structure generators and scoring functions described herein represent only the first step toward achieving this goal. Example

(27) Sofen, S. R.; Abu-Dari, K.; Freyberg, D. P.; Raymond, K. N. *J. Am. Chem. Soc.* **1978**, *100*, 7882.

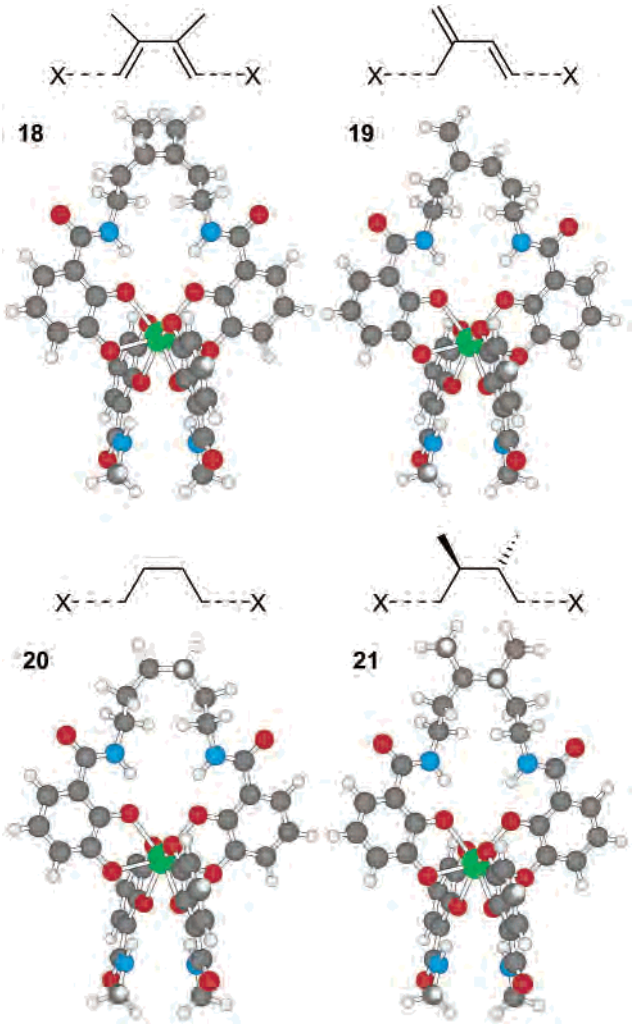

Figure 13. Linking fragment connectivities and structures of the top four hosts obtained using OVERLAY to connect two catecholamide ligands in a dodecahedral U(IV) complex.

applications demonstrate that LINKER and OVERLAY quickly identify reasonable molecular structures with binding sites that are complementary for the metal ion guest. These algorithms have been implemented in the program HostDesigner that offers several interesting features to the potential user: (i) The program simulates combinatorial chemistry on the computer. LINKER generates every connectivity, stereochemistry, and conformation that can be made given the attachment vectors on the complex fragments and the linking fragments present in the library. OVERLAY generates every reasonable structure from the linking fragments present in the library. (ii) Both algorithms generate an unbiased list of novel structures that can keep the designer from overlooking possible candidates. (iii) The program runs on any platform and is easy to use. (iv) The algorithms are computationally efficient. In the time that it would take to build one structure by hand with a graphical user interface, the program can generate and prioritize millions of structures.

Acknowledgment. This research was funded in part by Chemical Sciences, Office of Basic Energy Sciences, Office of Science, U.S. Department of Energy (DOE), and in part by the Laboratory Directed Research and Development Program, Fundamental Science Division of the Pacific Northwest National Laboratory (PNNL). The research was performed in the William R. Wiley Environmental Molecular Sciences Laboratory (EMSL) at PNNL. The EMSL is a national user facility funded by the Office of Biological and Environmental Research in the DOE. The authors express their appreciation to Dr. J. B. Nicholas (Genentech, Inc.) for his considerable input into the initial design of this software and for providing several subroutines, Prof. J. W. Ponder (Washington University Medical School) for granting permission to distribute portions of the TINKER code used in the development of the OVERLAY algorithm, and Drs. T. P. Straatsma and W. deJong (EMSL) for their assistance with writing and debugging the Fortran code.

IC0202920