J|A|C|S
ARTICLES
Published on Web 01/20/2006

De Novo Structure-Based Design of Bisurea Hosts for Tetrahedral Oxoanion Guests

Vyacheslav S. Bryantsev and Benjamin P. Hay*

Contribution from the Chemical Sciences Division, Pacific Northwest National Laboratory, Richland, Washington 99352

Received September 29, 2005; E-mail: ben.hay@pnl.gov

Abstract: This paper presents a computational approach to the deliberate design of improved host architectures. De novo molecule building software, HostDesigner, is interfaced with molecular mechanics software, GMMX, providing a tool for generating and screening millions of potential structures. The efficacy of this computer-aided design methodology is illustrated with a search for bisurea podands that are structurally organized for complexation with tetrahedral oxoanions.

Introduction

Anion complexation by synthetic host molecules is an important theme in supramolecular chemistry.[1] One successful approach for preparing anion hosts has been to add hydrogen bond donor groups to an organic scaffold to yield receptors that interact with anions through hydrogen bonding. In a given environment, there will be a maximum binding affinity that can be obtained when a given set of hydrogen bond donor groups are assembled about a specific anion. This theoretical limit, which is moderated by environmental factors such as the solvent, counterions, ionic strength, etc., is attained when the binding sites are structurally organized for anion complexation.[2] A high degree of structural organization is obtained when two conditions are met. First, the host must be able to adopt a conformation in which all binding sites are positioned to structurally complement the guest.[3] Second, 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] In the ideal case, the host would be preorganized such that the binding conformation is the most stable form.

Identification of host structures that exhibit these properties is not a trivial task. To address this issue we have developed de novo structure-based design software, HostDesigner (HD), specifically tailored to discover molecular structures that are organized to complex with small guest molecules.[5,6] HD generates and evaluates millions of candidate structures in minutes on a desktop personal computer and rapidly identifies three-dimensional architectures that position binding sites to provide a user-specified geometry with respect to the guest. The molecule building algorithms combine input host−guest fragments with linking fragments taken from a database. When using these fragments to build molecules, all connectivities, stereochemistries, and conformations are constructed, thereby creating every possible structure that can be made from the fragments. These structures are scored based on geometric factors, and a list of the top candidates is output. The initial screening performed by HD can be improved by subjecting the top candidates to more accurate, but slower, screening methods that are based on molecular mechanics calculations.

This paper demonstrates how these computer-aided molecular design methods, which are similar to those used in computer-aided drug design,[7] have been used to identify candidate host architectures that have a high probability of being effective anionophores. A search for bisurea hosts that are structurally organized for binding tetrahedral oxoanions was chosen for this demonstration. This choice was motivated by recent work in which electronic structure calculations were used to elucidate

(1) (a) *Supramolecular Chemistry of Anions*; Bianchi, A., Bowman-James, K., García-España, E., Eds.; Wiley-VHC, New York, 1997. (b) Schmidtchen, F. P.; Berger, M. *Chem. Rev.* **1997**, *97*, 1609. (c) Gale, P. A. *Coord. Chem. Rev.* **2000**, *199*, 181. (d) Gale, P. A. *Coord. Chem. Rev.* **2001**, *213*, 79. (d) Beer, P. D.; Gale, P. A. *Angew. Chem., Int. Ed.* **2001**, *40*, 486. (e) Fitzmaurice, R. J.; Kyne, G. M.; Douheret, D.; Kilburn, J. D. *J. Chem. Soc., Perkin Trans. 1* **2002**, 841. (f) Martínez-Máñez, R.; Sacenón, F. *Chem. Rev.* **2003**, *103*, 4419. (g) Suksai, C.; Tuntulani, T. *Chem. Soc. Rev.* **2003**, *32*, 192. (h) Choi, K.; Hamilton, A. D. *Coord. Chem. Rev.* **2003**, *240*, 101. (i) Lambert, T. N.; Smith, B. D. *Coord. Chem. Rev.* **2003**, *240*, 129. (j) Davis, A. P.; Joos, J.-B. *Coord. Chem. Rev.* **2003**, *240*, 143. (k) Gale, P. A. *Coord. Chem. Rev.* **2003**, *240*, 191. (l) *Fundamentals and Applications of Anion Separations*; Moyer, B. A., Singh, R. P., Eds.; Kluwer Academic/Plenum: New York, 2004.

(2) (a) Steed, J.; Atwood, J. *Supramolecular Chemistry*; John Wiley and Sons: LTD, Chichester, 2000. (b) Schneider, H.-J.; Yatsimirsky, A. *Principles and Methods in Supramolecular Chemistry*; John Wiley and Sons: LTD, Chichester, 2000.

(3) Cram, D. J.; Lein, G. M. *J. Am. Chem. Soc.* **1985**, *107*, 3657.

(4) (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.

(5) Hay, B. P.; Firman, T. K. *Inorg. Chem.* **2002**, *41*, 5502.

(6) (a) Hay, B. P.; Firman, T. K. *HostDesigner User's Manual*, PNNL-13850; Pacific Northwest National Laboratory, Richland, WA, 2004. (b) HostDesigner software and User's Manual are available free of charge via the Internet at http://hostdesigner.emsl.pnl.gov.

(7) (a) Kuntz, I. D.; Meng, E. C.; Shoichet, B. K. *Acct. 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) Murcko, A.; Murcko, M. A. *J. Med. Chem.* **1995**, *38*, 4953. (e) Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 425. (f) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. *Nature Rev.* **2004**, *3*, 935.

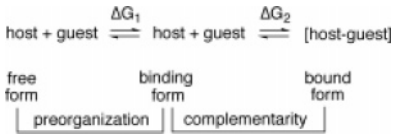


Figure 1. Irrespective of the actual complexation mechanism, it is convenient to partition the reaction into two steps defining three distinct structural states for the host: bound form, binding form, and free form.[11] The bound form is the structure of the host when complexed with the guest, the binding form is the host conformation obtained after removing the guest and optimizing the host, and the free form is the global minimum conformation of the host.

the structure and energetics of *N*-alkylated urea molecules[8] and urea−anion complexes.[9] These studies provided the information needed both to validate a force field model for more accurate evaluation of urea−anion complexes[10] and to define geometries for HD input. Herein, we report the first example in which de novo structure-based design and high-throughput screening methods have been deployed to identify promising anion host architectures prior to synthesis and binding affinity measurements.

Methods

Structure Generation. Bisurea host molecules were constructed using the de novo structure-based design software, HD.[5,6] This software assembles structures from molecular fragments. Here structures were built by connecting two *N*-methylurea-perchlorate fragments with a hydrocarbon fragment taken from the default HD database. To enhance synthetic accessibility of the structures that were generated by HD, structures taken from the fragment database were filtered to remove all cases in which the fragment was asymmetric by 2-D connectivity.[6a] The approach used to make the *N*-methylurea-perchlorate fragments is presented in the Results and Discussion section, and example HD input files are provided as Supporting Information.

Scoring Methods. HD outputs an ASCII file containing Cartesian coordinates for a series of host structures presented in order of decreasing complementarity for the guest. The initial evaluation of complementarity is based on geometric factors.[6a] Although approximate in nature, the geometry-based scoring method used by HD provides a rapid means for selecting the best candidates from a large group of potential structures.

Subsequent molecular mechanics analyses were applied to provide a more accurate prioritization of the top candidates. It is convenient to partition the complexation event into a two-step process as shown in Figure 1. In the first step, the host goes from the free form, defined as the lowest energy conformation of the host, to the binding form.[11] The difference in free energy between these two forms, ΔG_1, provides a measure of the degree of preorganization in the host. In the second step, the host and guest form the complex. The free energy change for this step, ΔG_2, is a measure of the degree of complementarity offered by the binding conformation.

An interface between HD and GMMX[12] was developed to automate the molecular mechanics evaluations. These evaluations occur in two steps. In the first step, interaction energies, $\Delta E_2 = E(\text{complex}) - E(\text{host, binding form}) - E(\text{guest})$, are calculated for the top 2000 candidates. The ΔE_2 values can be used to estimate the free-energy change ΔG_2 (see Figure 1) if it is assumed that (a) the calculated form of the complex represents the most populated form and (b) entropic contributions are constant except for restricted bond rotation associated with the formation of the host−guest complex. The magnitude of the latter term given by the empirical relationship, $0.31N_{rot}$ kcal/mol where N_{rot} is the number of freely rotating bonds restricted on complexation.[13] Thus, ΔG_2 values in kcal/mol are provided by eq 1, consisting of an enthalpic component, ΔE_2, and entropic component, $0.31N_{rot}$, and some constant contribution c_1. The HD program applies a group additivity approach to obtain an approximate value for the free-energy change $\Delta G_1(\text{est})$.[6a] Ignoring the constant term, eq 2 gives a value for the relative binding free energy, ΔG_{rel}. The top 2000 candidates are then placed in order of increasing ΔG_{rel} values. The best ΔG_{rel} values are obtained when ΔE_2 is the most negative (complementary), $N_{rot} = 0$ (no restricted rotations), and $\Delta G_1(\text{est}) = 0$ (preorganized).

$$\Delta G_2 = \Delta E_2 + T\Delta S_2 = \Delta E_2 + 0.31N_{rot} + c_1 \quad (1)$$

$$\Delta G_{rel} = \Delta G_1(\text{est}) + \Delta E_2 + 0.31N_{rot} \quad (2)$$

In the second step, conformational analyses are performed on the top 500 host structures from the first step to obtain values for ΔE_1, taken as $E(\text{binding form}) - E(\text{global minimum})$. The ΔE_1 values yield an improved estimate for ΔG_1 if it is assumed that (a) in the absence of the guest the majority of the host is in the global minimum conformer and (b) entropic contributions are constant. Thus, ΔG_1 values are provided by eq 3, consisting of an enthalpic component, ΔE_1, and some constant contribution c_2. As in the first step, ignoring the constant terms, combining eqs 1 and 3 yields eq 4, which gives an improved value for the relative binding free energy, ΔG_{rel}. The top 500 candidates are then placed in order of increasing ΔG_{rel} values to yield the final candidate ranking.

$$\Delta G_1 = \Delta E_1 + T\Delta S_1 = \Delta E_1 + c_2 \quad (3)$$

$$\Delta G_{rel} = \Delta E_1 + \Delta E_2 + 0.31N_{rot} \quad (4)$$

Molecular Mechanics Calculations. Molecular mechanics calculations were performed with the MMFF94 force field[14] as implemented in GMMX,[12] a program that is capable of performing both geometry optimizations and conformational analyses. The default MMFF94 parameter set was updated to include the prior modifications for accurate modeling of urea derivatives and their anion complexes.[10] This force-field model was used to calculate ΔE_1 and ΔE_2 values. These calculations were performed both in the gas-phase and in the aqueous-phase. In the latter instance, the influence of solvation was estimated with the generalized Born/surface area (GB/SA) continuum model[15] as implemented in GMMX.

Because there are two guests present in the structures created by HD, a procedure was required to generate input geometries for the host−guest complex that were used in the determination of ΔE_1. The process is as follows. The two guests are removed after determining their average coordinates. A single guest is then placed to achieve the best superposition upon these average coordinates. The resulting structure was optimized to obtain E(complex). Input coordinates for the determination of E(host, binding form) were obtained by removing the guest from the optimized complex. After optimization, the binding

(8) Bryantsev, V. S.; Firman, T. K.; Hay, B. P. *J. Phys. Chem. A* **2005**, *109*, 832.
(9) Hay, B. P.; Firman, T. K.; Moyer, B. A. *J. Am. Chem. Soc*. **2005**, *127*, 1810.
(10) Bryantsev V. S.; Hay, B. P. *THEOCHEM* **2005**, *725*, 177.
(11) Hay, B. P.; Zhang, D.; Rustad, J. R. *Inorg. Chem.* **1996**, *35*, 2650.
(12) The GMMX program, a component of PCModel, is available through Dr. Kevin Gilbert, Serena Software, Box 3076, IN 47402.
(13) (a) Eblinger, F.; Schneider, H.-J. *Angew. Chem., Int. Ed*. **1998**, *37*, 826. (b) Mammen, M.; Shakhnovich, E. I.; Whitesides, G. M. *J. Org. Chem.* **1998**, *63*, 3168. (c) Houk, K. N.; Leach, A. G.; Kim, S. P.; Zhang, X. *Angew. Chem., Int. Ed.* **2003**, *42*, 4872. (d) Deanda, F.; Smith, K. M.; Liu, J.; Pearlman, R. S. *Mol. Pharm.* **2004**, *1*, 23.
(14) (a) Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 490. (b) Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 520. (c) Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 553. (d) Halgren, T. A.; Nachbar, R. B. *J. Comput. Chem.* **1996**, *17*, 587. (e) Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 616. (f) Halgren, T. A. *J. Comput. Chem.* **1999**, *20*, 720. (g) Halgren, T. A. *J. Comput. Chem.* **1999**, *20*, 730.
(15) (a) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc*. **1990**, *112*, 6127. (b) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. *J. Phys. Chem. A* **1997**, *101*, 3005.

conformation of the host provided input coordinates for conformational analysis.

Conformational searching of host molecules was done using Monte Carlo random sampling and stochastic simulation strategy with default GMMX settings.[12] 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^{-1} of the global minimum.

Hardware. HD and GMMX calculations were performed on a MacIntosh G5 computer with a 2-GHz PowerPC 970 processor.

Results and Discussion

Urea−Anion Input Fragments. As implied by the term structure-based design, the host architectures that are assembled by HD are both built and evaluated based on prior knowledge of molecular structure. In this application, each host structure was generated by connecting two host−guest fragments to a hydrocarbon fragment taken from a database. The hydrocarbon fragments are based on molecular mechanics optimized geometries. Connecting the fragments leads to the formation of two single bonds. Both the bond lengths and the dihedral angles assigned to these two bonds are based on predetermined potential energy surfaces. In contrast, the geometry of the host−guest fragments is not predefined in HD and these data must be provided as input to the program.

The input file describing a host−guest fragment contains three pieces of information. These are the definition of the structure using Cartesian coordinates and a bond list, a list of hydrogen atoms that will be replaced by hydrocarbon fragments, and a specification of structural degrees of freedom within the host−guest fragment. When constructing the host−guest fragments, the guest should be positioned relative to the host binding sites to define a complementary geometry, that is, a geometry that will give a strong interaction between the host binding sites and the guest.

The host−guest fragments used in this study were constructed by combining *N*-methylurea with a tetrahedral oxoanion. The perchlorate anion (Cl−O distance 1.46 Å), which is slightly smaller than sulfate (S−O distance 1.48 Å) and phosphate (P−O distance 1.53 Å),[16] was selected as a representative guest. Given that tetrahedral oxoanions all exhibit similar directionality in hydrogen bonding,[17] bisurea hosts that are organized to bind perchlorate should also bind well with other tetrahedral oxoanions. Prior MP2 calculations on urea−anion complexes suggested that three possible configurations should be considered for the *N*-methylurea−perchlorate complex.[9] These configurations (Figure 2) are denoted syn-edge, anti-edge, and syn-vertex, where the syn or anti designations specify the conformation of the urea group and the edge or vertex designations specify whether the urea is binding to two oxygen atoms on the edge of the tetrahedron or one oxygen atom on the vertex of the tetrahedron.

The input file must also indicate which hydrogen atoms attached to the host can be replaced with hydrocarbon linkages

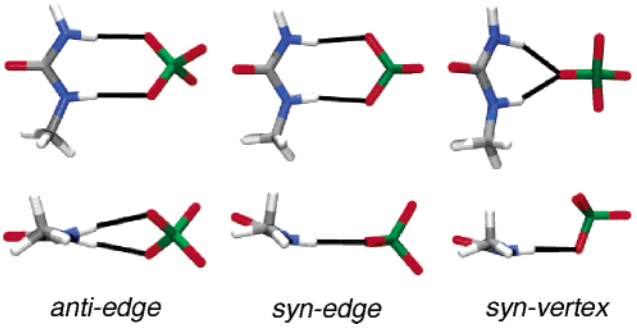


Figure 2. MMFF94 optimized geometries for the three possible binding configurations for *N*-methylurea−perchlorate complexes provided Cartesian coordinates for HD input files.[18]

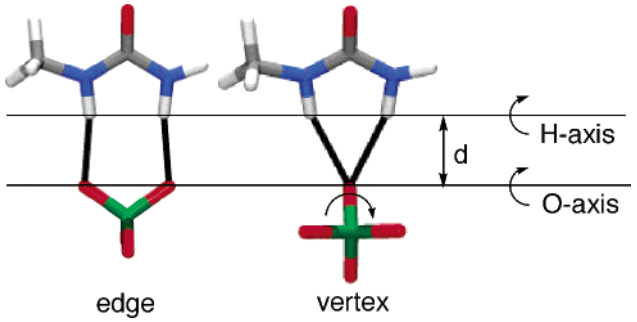


Figure 3. Structural degrees of freedom scanned during the building process include variation of the distance between the host and guest, *d*, rotation about the H-axis, rotation about the O-axis, and, in the *vertex* form, rotation about the $Cl-O_{vertex}$ bond.

during the building process. For each of the three cases shown in Figure 2, the N−H group cis to the carbonyl oxygen atom was selected for this purpose, defining a bonding vector on the amide nitrogen atom. Because replacing this N−H group with an N−R group results in a distortion to the angle between this vector and the C−N bond (for example, this angle changes from 112 to 119° in the anti-edge form when H is replaced by CH_3), the following procedure was adopted to prepare each structure for subsequent R substitution at the nitrogen atom. The cis hydrogen was replaced with a methyl group, the complex was optimized with the MMFF94 model, and the methyl that was added was replaced with a hydrogen atom. This procedure was used to generate the Cartesian coordinates for each of the three forms.

Finally, the input file for a host−guest fragment may contain a specification of structural degrees of freedom, in other words, distances, angles, and dihedral angles, that can be varied during the building process.[6a] This feature takes into account known flexibility within the structure and allows HD to sample a larger extent of structure space leading to more hits of better quality. The degrees of freedom specified for the *N*-methylurea−perchlorate fragments are depicted in Figure 3. They include variation of the distance between the host and guest, rotation about the H-axis, rotation about the O-axis, and, in the case of the syn-vertex form, rotation about the $Cl-O_{vertex}$ bond.

The extent of flexibility within the complexes was based on the displacements of the geometrical parameters from their equilibrium values that would result in approximately a 1 kcal mol^{-1} decrease in binding energy. MMFF94 potential energy surfaces for selected structural distortions shown in Figure 4 reveal that the structures are quite flexible. Evaluation of these

(16) Bridgeman, A. J.; Cavigliasso, G. *Polyhedron* **2001**, *20*, 2269.
(17) Hay, B. P.; Dixon, D. A.; Bryan, J. C.; Moyer, B. A. *J. Am. Chem. Soc.* **2002**, *124*, 182.

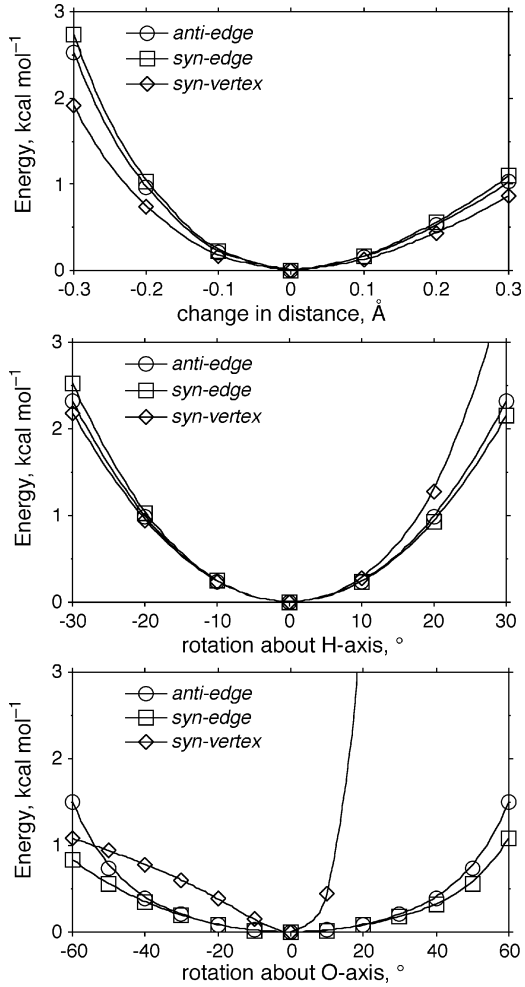


Figure 4. MMFF94 potential energy surfaces for variation in distance (top), rotation about the H-axis (middle), and rotation about the O-axis (bottom) were used as a basis for assigning the extent of flexibility in the host−guest input fragments.

data yield the following range of values assigned to each degree of freedom: ±0.2 Å for distance variation, ±20° for rotation about the H-axis, and ±60° (edge forms) or +15 to −60° (vertex form) for rotation about the O-axis. Similar analysis of rotation about the Cl−O_{vertex} bond in the syn-vertex form (not shown) revealed a surface of 3-fold periodicity with barrier heights of less than 0.4 kcal mol^{-1}. Thus, this degree of freedom was allowed to vary ±60°.

Structure Generation and Evaluation. Six HD runs were performed to sample all possible combinations of the three host−guest fragments shown in Figure 2. In a typical run, HD constructed and scored 300 million geometries within 40 min: a rate of 7.5 million geometries per minute! The scoring performed by HD, which prioritizes the host structures with respect to their complementarity for the guest, is based on geometric considerations.

During the construction of each host−guest fragment, the perchlorate anion was positioned relative to the urea group to define a complementary geometry. When two of these host−guest fragments are linked, the degree of superposition of the perchlorate anion from the first fragment with the perchlorate anion from the second fragment provides a simple criterion for the rapid evaluation of the degree of complementarity offered by the host. By definition, optimal complementarity is obtained when the root-mean-squared deviation, rmsd, of the distances between equivalent pairs of atoms in the two perchlorate anions is zero, in other words, when the two perchlorate anions representing a user-defined orientation with respect to each host component are exactly superimposed. As illustrated in Figure 5, the use of rmsd values for guest superposition provides a rapid method for the initial prioritization of the structures.

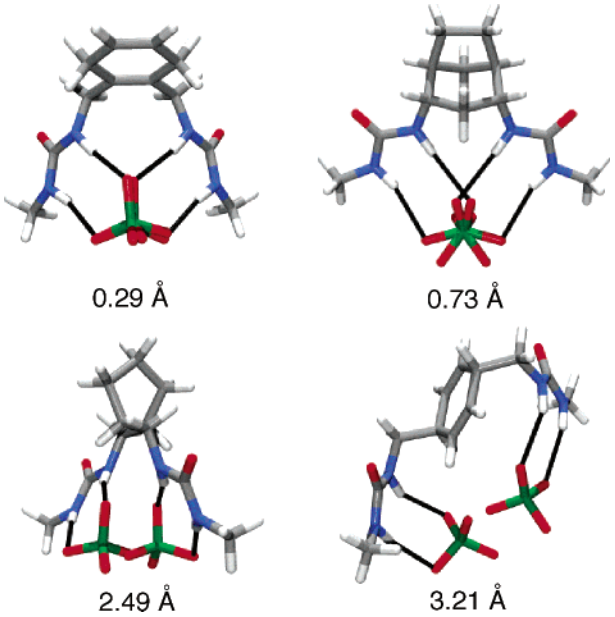


Figure 5. The rmsd value for guest superposition provides a rapid method for prioritizing structures produced by HD. This concept is illustrated above for structures obtained by linking two syn-edge *N*-methylurea−perchlorate fragments where the top left structure, rmsd = 0.29 Å, is the best, and the bottom right structure, rmsd = 3.21, is the worst.

Molecular mechanics analyses were used to achieve a more accurate ranking of the candidates. In the first step, ΔG_{rel} values (eq 2) were evaluated for the top 2000 candidates from each of the six runs. This step requires 1 h to complete. In the final stage of scoring, conformational analyses were performed to obtain improved ΔG_{rel} values (eq 4). Because conformational analyses are much more time-consuming, this last step was performed on the top 500 candidates from each run, on average requiring 8 days for completion. Overall, a total of 1.8 billion potential geometries were generated, 12 000 interaction energies were computed, and 3000 conformational analyses were conducted in less than a month of CPU time.

The results from all six runs were combined to obtain a list of the best 3000 structures. After removing any duplicates and sorting by the ΔG_{rel} value, the top 250 candidates were retained for visual inspection. Starting from the best candidate and working down the list, a final set of structures was selected with the application of two rules. First, when two structures differ only by the degree of substitution on the linkage between them, the simpler structure was retained. Second, structures containing linkages deemed to be synthetically intractable were removed. Examples illustrating when these rules were applied are given in Figure 6.

Top Candidates in the Gas Phase. The top 30 bisurea architectures identified in this study are shown in Figure 7 where they are numbered in order of decreasing score in the gas phase with **1** being the best and **30** being the worst. The number of

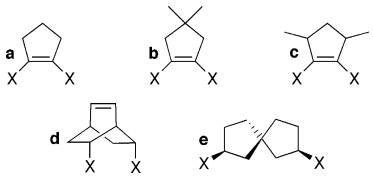


Figure 6. Examples of structures that were rejected by visual inspection, where X = —NHC(═O)NHCH$_3$. Of the top three structures, **a**—**c**, both **b** and **c** were rejected because they were derivatives of the simpler case **a**, which was retained. The bottom two structures, **d** and **e**, were rejected because they were deemed to be synthetically difficult.

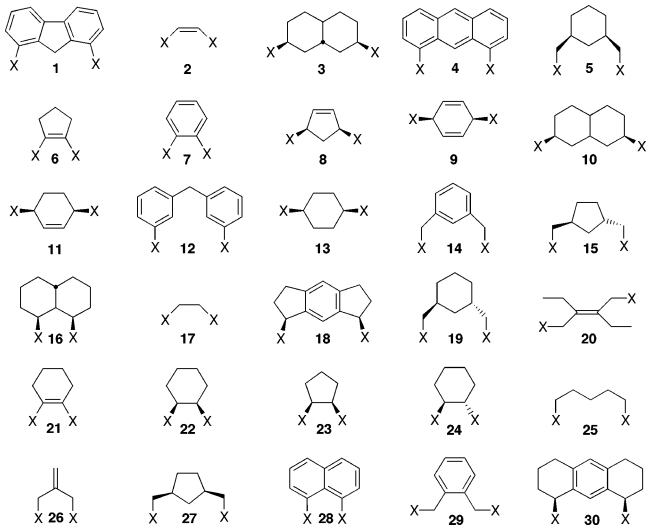


Figure 7. The top 30 gas-phase structures where X = —NHC(═O)NHCH$_3$.

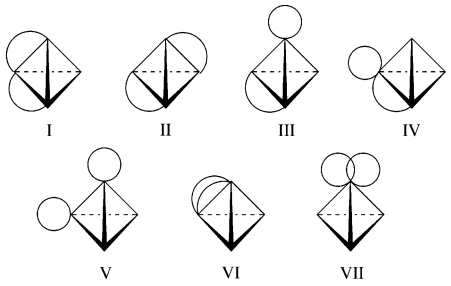


Figure 8. The seven possible hydrogen bonding motifs for placing two chelating urea groups on a tetrahedral oxoanion: I, adjacent edges; II, opposite edges; III, edge + vertex; IV, shared edge + vertex; V, vertex; VI, shared edge; VII, shared vertex.

carbon atoms in the shortest path between the urea moieties, which ranges from two to seven, can be used to group the candidates into classes. The most common class occurs when the urea nitrogen atoms are separated by five carbons. This is seen in **1**, **3**, **4**, **10**, **18**, and **30** where urea has been attached directly to rigid fused ring systems and in **5**, **14**, **15**, **19**, and **27**. Four-carbon separation is observed in the related series **9**, **11**, and **13** as well as in **20** and **29**. Three-carbon separation is found in **8**, **16**, **26**, and **28**. Two-carbon separation occurs with sp^3 carbon atoms in **2**, **6**, **7**, and **21** and with sp^2 carbon atoms in **17** and **22**—**24**.

The nature of the hydrogen bonding in the complex provides another way to classify the candidates. The seven possible ways to distribute edge and vertex bound urea groups about a tetrahedral anion are shown in Figure 8. Examination of **1**—**30** reveals that only four of the motifs are present (motif, number of occurrences): I, 14; III, 9; IV, 1; V, 6 (see Table 1).

Table 1. Gas-Phase Scoring Data and Hydrogen Bonding Motifs for the Top Candidates[a]

structure	ΔG_{rel}	ΔE_1	ΔE_2	$0.31N_{rot}$	H-bond motif[b]
1	−37.53	5.49	−43.64	0.62	V
2	−37.13	2.79	−40.54	0.62	I
3	−36.40	2.31	−39.34	0.62	V
4	−35.76	6.28	−42.66	0.62	III
5	−34.83	1.74	−37.81	1.24	I
6	−34.43	4.31	−39.36	0.62	I
7	−34.38	4.25	−39.25	0.62	V
8	−34.24	2.28	−37.14	0.62	I
9	−33.95	2.68	−37.25	0.62	I
10	−33.76	3.84	−38.23	0.62	III
11	−33.62	2.49	−36.73	0.62	III
12	−33.15	6.18	−40.57	1.24	III
13	−32.22	2.86	−35.70	0.62	V
14	−31.98	2.48	−35.70	1.24	III
15	−31.88	2.64	−35.76	1.24	III
16	−31.81	4.35	−36.78	0.62	I
17	−31.51	2.27	−34.71	0.93	V
18	−31.37	2.12	−34.10	0.62	III
19	−31.21	4.46	−36.91	0.62	I
20	−31.10	0.77	−33.11	1.24	I
21	−31.05	6.48	−38.15	0.62	V
22	−30.85	6.39	−37.86	0.62	I
23	−29.79	5.40	−35.81	0.62	I
24	−28.62	7.54	−36.79	0.62	III
25	−28.16	7.55	−37.57	1.86	I
26	−28.16	2.34	−31.74	1.24	I
27	−27.84	7.23	−36.31	1.24	III
28	−26.41	3.26	−30.29	0.62	IV
29	−25.63	8.51	−35.38	1.24	I
30	−24.33	15.18	−40.14	0.62	I

[a] Energies reported in kcal mol^{-1}. [b] See Figure 8.

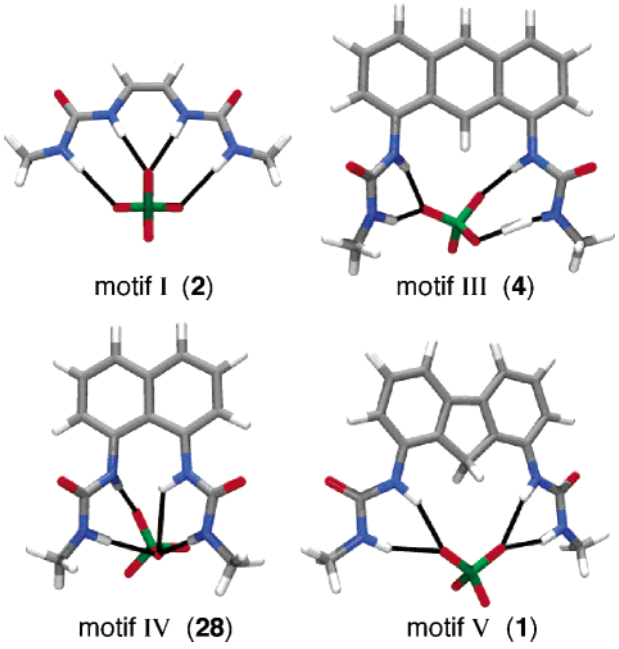


Figure 9. Examples of the four hydrogen-bonding motifs observed in **1**—**30**.

Representative structures containing the four observed motifs are presented in Figure 9.

Prior calculations on complexes formed between anions and urea suggest that the edge binding forms I and II should be the most stable.[9] Although I is the most frequently observed motif, II does not occur. The absence of II is explained by the large distance required to span the two nitrogen atoms, in the neighborhood of 8 Å. Relatively few linkages in the HD database are big enough to meet this requirement. Also absent

are arrangements in which two urea groups share an edge, VI, or a vertex, VII.

The gas-phase scoring results for **1**−**30** are summarized in Table 1. The relative free energies estimated for perchlorate complexation, ΔG_{rel}, range from −37.5 to −24.3 kcal/mol. The ΔG_{rel} values are the sum of three contributions, the conformational energy relative to the global minimum, ΔE_1, the interaction energy, ΔE_2, and an entropy penalty for the number of rotatable bonds that are restricted on anion chelation, $0.31N_{rot}$.

The ΔE_1 values range from 0.8 to 15.2 kcal mol^{-1} with an average value of 4.5 kcal mol^{-1} revealing that none of these structures are preorganized, and in most instances, significant reorganization is required to adopt the binding conformation. Examination of the free host structures shows that in many cases the global minima are stabilized by intramolecular hydrogen bonding. In the absence of any structural constraints imposed by the spacer between the urea groups, this interaction is significant. For example, the MMFF94 model gives an interaction energy of −10.4 kcal mol^{-1} for the hydrogen bond complex formed between two *N*,*N*′-dimethylurea molecules.

The ΔE_2 values range from −43.6 to −30.3 kcal mol^{-1}. An estimate for the maximum possible value is provided by 2:1 urea−anion complexes in which the urea groups are not sterically constrained by a spacer. The most stable complex formed between two *N*,*N*′-dimethylurea ligands and perchlorate, with hydrogen bonding motif II (Figure 8), has a ΔE_2 value of −36.5 kcal mol^{-1}. This raises the question of why two of the hosts, **1** and **2**, exhibit lower values. In these cases the urea groups are attached to sp^2 carbons resulting in more acidic N−H groups. The MMFF94 model accounts for these inductive effects. For example, the complex formed between two *N*-methyl-*N*′-phenylurea ligands and perchlorate, also with hydrogen bonding motif II, yields a ΔE_2 value of −46.2 kcal mol^{-1}.

With N_{rot} values ranging from 2 to 6, the entropy penalty for restricted bond rotation ranges from 0.62 to 1.86 kcal mol^{-1}. This contribution, which shows much less variation than ΔE_1 and ΔE_2, has the least impact on the gas-phase ΔG_{rel} values. Despite the small influence of this term, the majority of the top 30 spacers contain only two rotatable bonds between the urea nitrogen atoms. Steric constrains imposed by more rigid architectures can act to weaken intramolecular hydrogen bonding thereby leading to smaller ΔE_1 values. In this context, a minimum number of rotatable bonds is a desirable structural attribute.

Influence of Solvation. Whereas the behavior of the bisurea hosts in very low dielectric environments might be predicted by the gas-phase results, the behavior in higher dielectric environments might differ significantly. A second round of HD runs were performed to evaluate these differences. This time the molecular mechanics scoring was conducted using the GB/SA continuum model for aqueous solvation.[15] After collation and visual inspection of the results (vide supra), the same top candidates **1**−**30** were retained. However, the aqueous-phase scoring results, presented in Table 2, yield a different ranking than in the gas phase.

Compared to the gas-phase results, the aqueous-phase ΔG_{rel} values are compressed within a much smaller range of −0.6 to 2.5 kcal mol^{-1}. Two factors are responsible for this compression. First, intramolecular hydrogen bonding is largely suppressed.

Table 2. Aqueous-Phase Scoring Data[a]

structure	ΔG_{rel}	ΔE_1	ΔE_2	$0.31N_{rot}$
1	−0.61	0.20	−1.43	0.62
4	−0.50	0.09	−1.21	0.62
6	−0.16	0.00	−0.78	0.62
2	−0.15	0.00	−0.77	0.62
3	0.22	0.90	−1.30	0.62
10	0.27	0.44	−0.79	0.62
8	0.35	0.41	−0.68	0.62
7	0.41	0.26	−0.48	0.62
9	0.45	0.36	−0.52	0.62
11	0.55	0.86	−0.93	0.62
30	0.59	0.00	−0.03	0.62
13	0.63	0.91	−0.90	0.62
23	0.64	0.50	−0.48	0.62
22	0.67	0.51	−0.46	0.62
28	0.69	0.60	−0.53	0.62
24	0.70	0.45	−0.36	0.62
21	0.71	0.51	−0.41	0.62
12	0.81	0.36	−0.79	0.93
16	0.86	0.70	−0.45	0.62
18	1.06	0.82	−0.38	0.62
5	1.15	0.68	−0.76	1.24
19	1.15	0.22	−0.32	1.24
29	1.18	0.00	−0.06	1.24
17	1.19	0.81	−0.55	0.93
27	1.19	0.62	−0.66	1.24
20	1.37	0.55	−0.42	1.24
14	1.51	0.63	−0.36	1.24
15	1.70	0.86	−0.40	1.24
25	1.73	0.66	−0.79	1.86
26	2.52	2.75	−0.03	1.24

[a] Energies reported in kcal mol^{-1}.

As a consequence, the global minima from the gas-phase model are different than those in the aqueous-phase model and ΔE_1 values are now smaller, ranging from 0.0 to 2.8 kcal mol^{-1}. Second, the interaction energies, ΔE_2, are dramatically reduced, ranging from −1.4 to 0.0 kcal mol^{-1}. Since the contributions from ΔE_1 and ΔE_2 are smaller, the contribution from the number of rotatable bonds becomes more important. With one exception, **12**, the top 20 candidates have the minimum $N_{rot} = 2$ value and structures with higher N_{rot} values have moved to the bottom of the list.

Structures that score well in low (gas-phase) and high (aqueous-phase) dielectric environments should also score well under intermediate dielectric conditions. We note that, regardless of the scoring model used, the same 30 structures are retained from the top 250 candidates. Nine of the top ten gas-phase structures occur in the top ten aqueous-phase structures. And, the same structure, **1**, is identified as the best candidate under both conditions.

Comparison with Known Bisurea Hosts. A relatively small number of bisurea podands,[19−29] and closely related bisthiourea

(18) In accord with prior calculations on 1:1 urea−anion complexes, the vertex form is not a local minimum.[9, 10] MMFF94 yields this structure only by constraining a O−C···O−Cl dihedral angle to 0° during geometry optimization.

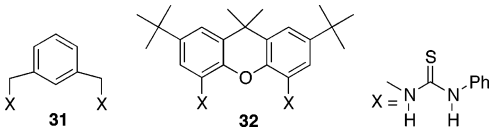
(19) Albert, J. S.; Hamilton, A. D. *Tetrahedron Lett.* **1993**, *34*, 7363.
(20) Nishizawa, S.; Bühlmann, P.; Iwao, M.; Umezawa, Y. *Tetrahedron Lett.* **1995**, *36*, 6483.
(21) Kwon, J. Y.; Jang, Y. J.; Kim, S. K.; Lee, K.-H.; Kim, J. S.; Yoon, J. *J. Org. Chem.* **2004**, *69*, 5155.
(22) Brooks, S. J.; Gale, P. A.; Light, M. E. *Chem. Commun.* **2005**, 4696.
(23) Amendola, V.; Boicchi, M.; Esteban-Gomez, D.; Fabbrizzi, L.; Monzani, E. *Org. Biomol. Chem.* **2005**, *3*, 2632.
(24) Hamann, B. C.; Branda, N. R.; Rebek, J., Jr. *Tetrahedron Lett.* **1993**, *34*, 6837.
(25) Lee, J. Y.; Cho, E. J.; Mukamel, S.; Nam, K. C. *J. Org. Chem.* **2004**, *69*, 943.
(26) Jose, D. A.; Kumar, D. K.; Ganguly, B.; Das, A. *Org. Lett.* **2004**, *6*, 3445.

analogues,[20,31−35] have been synthesized. Out of the spacers that have been reported, six of them are either identical or structurally similar to those represented by **1**−**30**. Binding constant data for these examples are limited to small sets of anions under diverse solvent conditions. In addition, terminal N-substitution, a factor that is known to impact binding affinities,[36] is variable. As a result, there is little data regarding how spacer variation between the two urea groups influences anion binding affinities. The available data, however, do establish that some of the architectures provided by **1**−**30** form complexes with tetrahedral anions in which both urea groups participate in bonding.

Connecting two *N*-butylurea groups to 1,3-xylene, as in **14**,[19] yields a host that forms 1:1 complexes with $H_2PO_4^-$ and HSO_4^- in DMSO with binding constants of 110 and 1 M^{-1}, respectively.[20] That both urea groups were participating was confirmed by ^{1}H NMR spectra, complex stoichiometry, and the fact that monourea ligands formed weaker complexes under the same conditions, for example, the complex between *N*,*N*′-dimethylurea and $H_2PO_4^-$ exhibited a binding constant of 28 M^{-1}. As with urea, thiourea analogues with 1,3-xylyl spacers have been shown to form 1:1 complexes with tetrahedral oxoanions.[20,31,32] Use of *N*-(*p*-NO_2-phenyl)thiourea groups yields a host that facilitates the transport of SO_4^{2-}, HPO_4^{2-}, and $H_2PO_4^-$ across the nitrobenzene−water interface in contrast to monothiourea ligands that do not.[33] The 1,3-xylyl spacer also has been used to connect structurally analogous guanidinium groups yielding a host that forms 1:1 complexes with dialkyl phosphate anions in MeCN.[37]

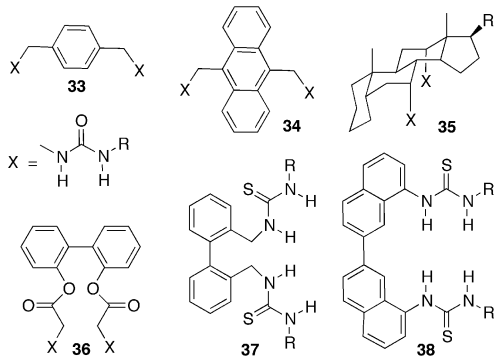
Three other spacers also have been shown to yield bisurea architectures that complex tetrahedral anions. Connecting *N*-phenylurea groups to the 1,8-anthracene spacer, as in **4**, yields a host that forms 1:1 complexes with HSO_4^- and $H_2PO_4^-$ in DMSO.[21] Binding constants for these complexes were not reported. Connecting *N*-phenylurea groups to the 1,2-arene spacer, as in **7**, yields a host that forms complexes with HSO_4^- and $H_2PO_4^-$ in DMSO with 1:1 binding constants of 10 and 732 M^{-1}, respectively.[22] A crystal structure of a complex with benzoate shows a hydrogen bonding motif analogous to that predicted for ClO_4^-. Connecting *N*-(*p*-nitrophenyl)urea groups to 1,2-*trans*-cyclohexane, as in **24**, yields a host that forms a 1:1 complex with $H_2PO_4^-$ with a binding constant of 912 M^{-1} in DMSO.[23] ^{1}H NMR spectra are consistent with a solvated structure in which both urea groups interact with the anion.

In a rare study that provides a direct comparison of the influence of the spacer on binding affinity, *N*-phenylthiourea groups were attached to the 1,3-xylyl spacer, **31**, and to a tricyclic scaffold, **32**.[31] The architecture in **32**, which exhibits



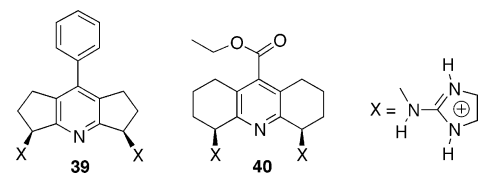
an geometry that is very similar to that in **4**, formed a 1:1 complex with $H_2PO_4^-$ in DMSO with a binding constant of 195 000 M^{-1}, 42 times greater than that for **31**. This result is fully consistent with scoring results for *N*-methylurea analogues in which **4** scores significantly better than **14** in both the gas- and aqueous-phase.

Although their affinity for tetrahedral anions has not been investigated, hosts formed by connecting urea groups to a 1,8-naphthalene spacer, as in **28**,[25] and a 1,2-arene spacer, as in **7**,[26] also have been reported. Anion binding studies for these hosts have thus far been limited to halide guests. Remaining examples of known bisurea and bisthiourea architectures, **33**−**38**,[27−30,34,35] are structures not identified in this study. Two



of these structures, **33** and **34**, were designed for complexation of larger anions such as glutarate. Both the 1,4-xylene spacer[27] and the 9,10-anthracene spacer[28] were evaluated by HD, but were rejected because they failed to provide a complementary array of hydrogen bonding sites for perchlorate. The other structures, **35**−**38**, contain spacers that are not present in the HD linkage database and, therefore, could not be constructed. Only one of these hosts, **38**, has been shown to complex a tetrahedral oxoanion with a reported 1:1 binding constant of 5 500 M^{-1} for $H_2PO_4^-$ in MeCN.[35]

Further indication that **1**−**30** are practical is found in a study of guanidinium-based receptors in which gaunidinium groups were attached to rigid tricyclic scaffolds to yield **39** and **40**.[38]



These structures exhibit architectures that are very similar to **18** and **30**, respectively. Measurements showed that both **39**

(27) Fan, E. F.; Van Arman, S. A.; Hamilton, A. D. *J. Am. Chem. Soc.* **1993**, *115*, 369.
(28) Gunnlaugsson, T.; Davis, A. P.; O′Brien, J. E.; Glynn, M. *Org. Biol. Chem.* **2005**, *3*, 48.
(29) Ayling, A. J.; Perez-Payan, M. N.; Davis, A. P. *J. Am. Chem. Soc.* **2001**, *123*, 12716.
(30) Albrecht, M.; Zauner, J.; Burgert, R.; Röttele, H.; Frölich, R. *Mater. Sci. Eng. C* **2001**, *18*, 185.
(31) Bühlmann, P.; Nishizawa, S.; Xiao, K. P.; Umezawa, Y. *Tetrahedron* **1997**, *53*, 1647.
(32) Tobe, Y.; Sasaki, S.; Mizuno, M.; Naemura, K. *Chem. Lett.* **1998**, *8*, 835.
(33) (a) Nishizawa, S.; Kamaishi, T.; Yokobori, T.; Kato, R.; Cui, Y.-Y.; Shioya, T.; Teramae, N. *Anal. Sci.* **2004**, *20*, 1559. (b) Nishizawa, S.; Rokobori, T.; Kato, R.; Yoshimoto, K.; Kamaishi, T.; Teramae, N. *Analyst* **2003**, *128*, 663.
(34) Lee, D. H.; Im, J. H.; Lee, J.-H.; Hong, J.-I. *Tetrahedron Lett.* **2002**, *43*, 9637.
(35) Kondo, S.; Nagamine, M.; Yano, Y. *Tetrahedron Lett.* **2003**, *44*, 8801.
(36) (a) Wilcox, C. S.; Kim, E.; Romano, D.; Kuo, L. H.; Burt, A. L.; Curran, D. P. *Tetrahedron* **1995**, *51*, 621. (b) Nie, L.; Li, Z.; Han, J.; Zhang, X.; Yang, R.; Liu, W.-X.; Wu, F.-Y.; Xie, J.-W.; Zhao, Y.-F.; Jiang, Y.-B. *J. Org. Chem.* **2004**, *69*, 6449.
(37) Jubian, V.; Dixon, R. P.; Hamilton, A. D. *J. Am. Chem. Soc.* **1992**, *114*, 1120.
(38) Kneeland, D. M.; Ariga, K.; Lynch, V. M.; Huang, C.-Y.; Anslyn, E. V. *J. Am. Chem. Soc.* **1993**, *115*, 10042.

and **40** formed 1:1 complexes with the dibenzoyl phosphate anion in DMSO/H_2O mixtures that were significantly stronger than those formed by a mono-guanidinium analogue. Participation of both binding sites was confirmed in a crystal structure of a complex between **39** and benzoyl phosphate that exhibited hydrogen bonding motif I (Figure 8).

The experimental studies confirm that four of the bisurea architectures identified in this study, **4**, **7**, **14**, and **24**, are organized for complexation with tetrahedral anions. Thus, the computer-aided design approach has identified a variety of novel structures. The only study of the influence of the spacer on binding affinity, **31** vs **32**, suggests that binding affinity can be increased markedly when a higher scoring architecture is used. It is therefore anticipated that a number of the structures identified in this study may yield higher tetrahedral anion binding affinities than have been attained with known compounds.

Summary

This paper has presented a strategy for the deliberate design of host architectures that are structurally organized to complex anion guests. The approach was based on optimal molecular geometries for the interactions between individual host binding sites and the guest. The identification of the most favorable host architectures was facilitated through the application of novel computer-aided design software. The HD program was used to rapidly search a large area of structural space and produce a list of top candidates, using geometry to rank them with respect to how well they complement the guest. When interfaced with the GMMX program, subsequent evaluation of these candidates using force-field-based scoring methods identified structures with desirable properties that include (a) large interaction energy, (b) low conformational energy, and (c) minimal number of restricted bond rotations on guest complexation.

The efficacy of this computer-aided design methodology has been illustrated with a search for bisurea podands that are organized for complexation with tetrahedral oxoanions. The approach has identified a wide variety of novel architectures that have a high probability of being effective hosts for anions such as perchlorate, sulfate, and phosphate. Synthesis and evaluation of candidates identified in this study are in progress to test the veracity of the scoring methods.

We eventually hope to design host architectures with cavities that recognize specific anion shapes. Steric recognition of a targeted anion will be achieved only when the binding sites are constrained such that they have a strong interaction with the targeted anion and weakened interaction with competing anions. The current study represents only the first step in the design process, that is, the identification of candidates that should have a strong interaction with a specific anion shape. Some of the bisurea hosts identified in this study meet this requirement for tetrahedral oxoanions. It is likely, however, that some of the candidates will accommodate other anion shapes with equal facility, and it remains to be determined whether any of the candidates will exhibit a significant steric preference for a tetrahedral guest. Searches for bisurea hosts that are structurally organized for spherical and trigonal planar anions are now underway to address this issue.

Acknowledgment. This work was supported by the Division of Chemical Sciences, Geosciences, and Biosciences, Office of Basic Energy Sciences, U.S. Department of Energy and performed at Pacific Northwest National Laboratory (managed by Battelle). The authors thank Oak Ridge National Laboratory scientists Peter V. Bonnesen, Radu Custelcean, Laetitia Delmau, and Bruce A. Moyer for their intellectual contributions to this research.

Supporting Information Available: Input files to HD that include complex fragment files to define the geometry and attachment points to the complex and example of the file to control the execution of the program. This material is available free of charge via the Internet at http://pubs.acs.org.

JA056699W