JOC Article

Synthesis and Conformational Analysis of Novel Trimeric Maleimide Cross-Linking Reagents

Agnieszka Szczepanska,[†] José Luis Espartero,[‡] Antonio J. Moreno-Vargas,[†] Ana T. Carmona,[†] and Inmaculada Robina*[,†]

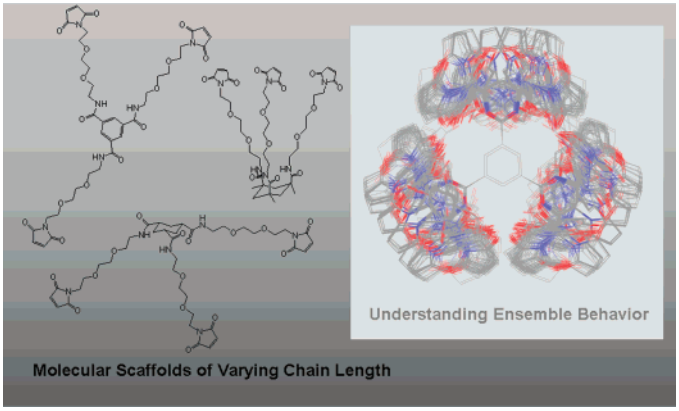
Department of Organic Chemistry, Faculty of Chemistry, University of Seville, P.O. Box 1203, E-41071 Seville, Spain, and Department of Organic and Pharmaceutical Chemistry, Faculty of Pharmacy, University of Seville, Spain

Sarah Remmert and Carol Parish*

Department of Chemistry, University of Richmond, Gottwald Science Center, Richmond, Virginia 23173

robina@us.es; cparish@richmond.edu

Received May 10, 2007



Nine homotrifunctional cross-linking reagents are presented. Their synthesis and chemical properties as well as their characterization by classical mechanical conformational searching techniques is reported. Mixed Low Mode and Monte Carlo searching techniques were used to exhaustively sample the OPLS2005/GBSA(water) potential energy surface of trisubstituted cyclohexane and benzene derivatives of C3 symmetry. Geometric structure, molecular length, and hydrogen-bonding patterns were analyzed. Nonaromatic compounds exhibited exclusively chair conformations at low energies, with a preference for axial or equatorial arms depending upon the presence of additional ring substituent Me groups. Increasing chain length often resulted in overall shorter molecular length due to additional chain flexibility. These results were consistent with one- and two-dimensional temperature-dependent NMR studies.

Introduction

The HIV-1 envelope glycoproteins gp120 and gp41 remain noncovalently associated and oligomerize most likely as trimers on the surface of the virion. This trimeric complex mediates viral entry into target cells, which is initiated by the high affinity interaction between the viral envelope gp120 and the CD4 cellular receptors.[1] Subsequent interaction with chemokine co-receptors CXCR4 or CCR5 triggers a cascade of conformational changes in gp120 that induces activation of the gp41 fusion subunit and formation of the pre-hairpin structure that may bridge both viral and cellular membranes.[2]

The crystal structure of a complex between gp120, CD4, and a neutralizing antibody that blocks the chemokine receptor binding site has been solved showing that this complex has a ternary structure.[3] Furthermore, modeling studies of the orientation of gp120 in the ternary complex reveal that gp120 is

[†] Department of Organic Chemistry.
[‡] Department of Organic and Pharmaceutical Chemistry.

(1) (a) Wyatt, R.; Sodroski, J. *Science* **1998**, *280*, 1884−1888. (b) Weissenhorn, W.; Dessen, A.; Calder, L. J.; Harrison, S. C.; Skehel, J. J.; Wiley, D. C. *Mol. Membr. Biol.* **1999**, *16*, 3−9. (c) Eckert, D. M.; Kim, P. S. *Annu. Rev. Biochem.* **2001**, *70*, 777−810.

10.1021/jo0709293 CCC: $37.00
Published on Web 08/04/2007

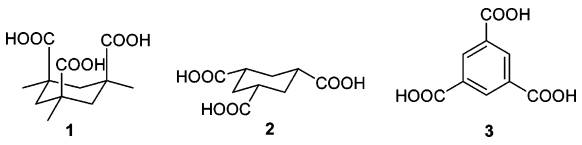


FIGURE 1. Structure of C3-symmetric templates.

constrained to be 3-fold symmetric. The binding of CD4/gp120 takes place predominantly via electrostatic interactions within a large and irregular interface. A protuberant Phe43 of CD4 inserts into a receptive hole of gp120, the "Phe-43 cavity", which is the conserved CD4 binding site on gp120.[4] Compounds that interfere with the binding of CD4-gp120 are therefore inhibitors of HIV entry. Additionally, a trimeric structure bearing C-peptide sequences of the gp41 ectodomain could also mimic the prehairpin intermediate and therefore provide a tool to block the fusion event. It has also been reported that a multivalent assembly of gp41 peptides enhances the α-helix content of the peptide and provides a model to mimic the HIV membrane fusion state.[5]

We have recently reported a series of trivalent CD4-mimetic miniproteins designed to match the distance between any two of the CD4 binding cavities by having three CD4M9 moieties tethered through a spacer to a 3-fold symmetric template.[6] They have shown significantly enhanced activity over the monovalent CD4M9 but activites similar to those of the bivalent system first reported by Wang and co-workers.[7] In this work, we report the synthesis of several new C3 symmetric templates that could serve as scaffolds to which CD4 mimetic peptides or miniproteins can be attached. Conformational analysis was used to characterize the different rigidities, flexibilities, and spatial orientations of each system to better understand their molecular behavior.

For the preparation of our scaffolds we have chosen C3-symmetric templates of well-defined architectures such as *cis*,*cis*-1,3,5-trimethyl cyclohexane tricarboxylic acid (Kemp's triacid, **1**), *cis*,*cis*-cyclohexane-1,3,5-tricarboxylic acid (**2**), and trimesic acid (**3**) (Figure 1) that are joined to synthetic or commercial spacers of different lengths. Kemp's triacid is known in supramolecular chemistry as a useful building block to assemble peptide-peptoid structures.[8] Trimesic acid was used in the construction of aromatic tripodal systems as model receptors for monosaccharides and other compounds.[9] Also, triamines derived from cyclohexane tricarboxylic acids **1** and **2** are used as chelating agents of trivalent cations.[10]

The preparation of the scaffolds according to our model requires the attachment of the spacers to the templates and the incorporation of a maleimide moiety. Maleimide-containing compounds have been widely employed as intermolecular cross-linking agents in chemical and biochemical applications for chemoselective ligation to cysteine-containing peptides and proteins.[11] It was reported that homobifunctional cross-linking reagents can be used as molecular rulers to evaluate the distance between two cross-linked amino acid residues; the quantitative measure of the lengths has been estimated by stochastic dynamic calculations.[12]

In this paper we present the synthesis, conformational behavior, and flexibility of trimeric maleimide clusters, compounds **4a**−**6c**,[13] as homotrifunctional cross-linking reagents (Figure 2). Since these derivatives are large molecules that can adopt different conformations in solution, the study could provide useful information regarding the molecular shape of CD4 cellular receptors as well as the prehairpin intermediate involved in the HIV-cell fusion event.

Results and Discussions

Synthesis and NMR Study. The preparation of *N*-substituted maleimide derivatives is not an easy task. Several procedures have been described that involve the formation of the maleimide moiety by ring closure of the intermediate amide, obtained by reaction of amines with maleic anhydride[14] or through a retro-Diels−Alder reaction of the adduct generated between furan and maleimide.[15] These methods require refluxing at high temperatures, and in the first case, low yields are obtained because of the low solubility of the maleic acid derivatives. Direct *N*-alkylation of maleimide under Mitsunobu conditions has also been used to introduce this functionality. However,

(2) (a) Gallo, S. A.; Puri, A.; Blumenthal, R. *Biochemistry* **2001**, *40*, 12231−12236. (b) Melikyan, G. B.; Markosyan, R. M.; Hemmati, H.; Delmedico, M. K.; Lambert, D. M.; Cohen, F. S. *J. Cell Biol.* **2000**, *151*, 413−424. (c) Muñoz-Barroso, I.; Durell, S.; Sakaguchi, K.; Appella, E.; Blumenthal, R. *J. Cell Biol.* **1998**, *140*, 315−323. (d) LaBranche, C. C.; Galasso, G.; Moore, J. P.; Bolognesi, D. P.; Hirsch, M. S.; Hammer, S. M. *Antiviral Res.* **2001**, *50*, 95−115.

(3) (a) Kwong, P. D.; Wyatt, R.; Robinson, J.; Sweet, R. W.; Sodroski, J.; Hendrickson, W. A. *Nature* **1998**, *393*, 648−659. (b) Wyatt, R.; Kwong, P. D.; Desjardins, E.; Sweet, R. G.; Robinson, J.; Hendrickson, W. A.; Sodroski, J. G. *Nature* **1998**, *393*, 705−711.

(4) (a) Hsu, S.-T. D.; Bonvin, A. M. J. *Proteins: Struct., Funct., Bioinf.* **2004**, *55*, 582−593. (b) Kwong, P. D.; Wyatt, R.; Majeed, S.; Robinson, J.; Sweet, R. W.; Sodroski, J.; Hendrickson, W. A. *Structure* **2000**, *8*, 1329−1339. (c) Kwong, P. D.; Wyatt, R.; Hendrickson, W. A. *J. Virol.* **2000**, 1961−1972.

(5) (a) Ni, J.; Powell, R.; Baskakov, I. V.; DeVico, A.; Lewis, G. K.; Wang, L.-X. *Bioorg. Med. Chem.* **2004**, *12*, 3141−3148. (b) Tam, J. P.; Yu, Q. *Org. Lett.* **2002**, *4*, 4167−4170. (c) Gochin, M.; Kiplin Guy, R.; Case, M. A. *Angew. Chem. Int. Ed.* **2003**, *42*, 5325−5328.

(6) Li, H.; Song, H.; Guan, Y.; Szczepanska, A.; Moreno-Vargas, A. J.; Carmona, A. T.; Robina, I.; Lewis, G. K.; Wang, L.-X. *Bioorg. Med. Chem.* **2007**, *15*, 4220−4228.

(7) Li, H.; Song, H.; Heredia, A.; Le, N.; Redfield, R.; Lewis, G. K.; Wang, L.-X. *Bioconjugate Chem.* **2004**, *15*, 783−789.

(8) (a) Jefferson, E. A.; Locardi, E.; Goodman, M. *J. Am. Chem. Soc.* **1998**, *129*, 7420−7428. (b) Feng, Y.; Melacini, G.; Taulane, J. P.; Goodman, M. *J. Am. Chem. Soc*. **1996**, *118*, 10351−10358. (c) Kocis, P.; Issakova, O.; Sepetov, N. F.; Lebl, M. *Tetrahedron Lett.* **1995**, *36*, 6623−6626.

(9) (a) Vacca, A.; Nativi, C.; Cacciarini, M.; Pergoli, R.; Roelens, S. *J. Am. Chem. Soc*. **2004**, *126*, 16456−16465. (b) Gibson, S. E.; Castaldi, M. P. *Chem. Commun*. **2006**, 3045−3062.

(10) Bowen, T.; Planalp, R. P.; Brechbiel, M. W. *Bioorg. Med. Chem. Lett*. **1996**, *6*, 807−810.

(11) (a) Hermanson, G. T. In *Bioconjugate Techniques*; Academic Press: New York, 1996, p 148. (b) Shin, I.; Jung, H.-J.; Lee, M.-R. *Tetrahedron Lett*. **2001**, *42*, 1325−1328. (c) Peters, K.; Richards, F. M. *Annu. Rev. Biochem.* **1977**, *46*, 523−551. (d) Zecherle, G. N.; Oleinikov, A.; Traut, R. R. *J. Biol. Chem.* **1992**, *267*, 5889−5894. (e) Kwaw, J. S.; Kaback, H. R. *Biochemistry* **2000**, *39*, 3134−3140. (f) Swaney, J. B. *Methods Enzymol.* **1986**, *128*, 613−626. (g) Holmes, K. C.; Popp, D.; Gebhard, W.; Kabsch, W. *Nature* **1990**, *347*, 44−49. (h) Lorenz, M.; Popp, D.; Holmes, K. C. *J. Mol. Biol.* **1993**, *234*, 826−836.

(12) Green, N.; Reisler, E.; Houk, K. N. *Protein Sci.* **2001**, *10*, 1293−1304.

(13) The synthesis and chemical characterization of compounds **4a**−**c** and **5c** is reported in ref 6.

(14) Reddy, P.; Y.; Kondo, S.; Fujita, S.; Toru, T. *Synthesis* **1998**, 999−1002.

(15) Farha, O. K.; Julios, R. L.; Hawthorne, M. F. *Tetrahedron Lett.* **2006**, *47*, 2619−2622.

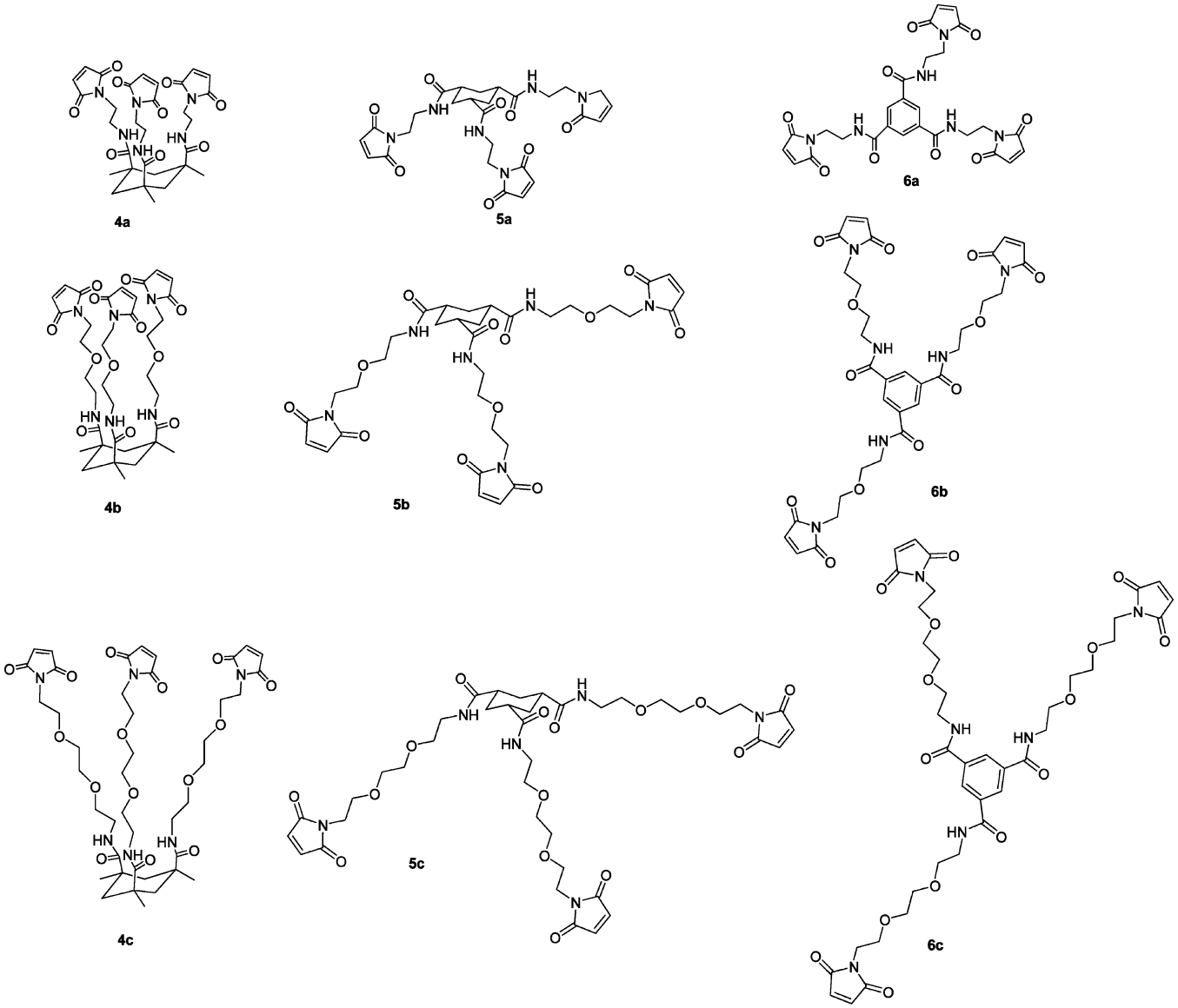


FIGURE 2. Structure of trimeric maleimide derivatives **4**−**6**.

this reaction is not always successful as it is extremely dependent on the order of addition of reactants.[16]

Coupling reactions between templates **1**−**3** and monoprotected *N*-Boc diamino spacers **7a**, **7b**, and **7c** using EDCI and HOBt gave compounds **8**−**10** in moderate-to-good yield. Boc deprotection with HCl/MeOH and subsequent neutralization with Na/MeOH gave the corresponding triamines in quantitative yield, which were made to react with *N*-methoxycarbonylmaleimide affording the maleimide scaffolds **4a**−**6c** in moderate-to-good yield. Other attempts to introduce the maleimide functionality under Mitsunobu conditions were unsuccessful; complex mixtures and brown polymeric material were obtained. (Scheme 1)

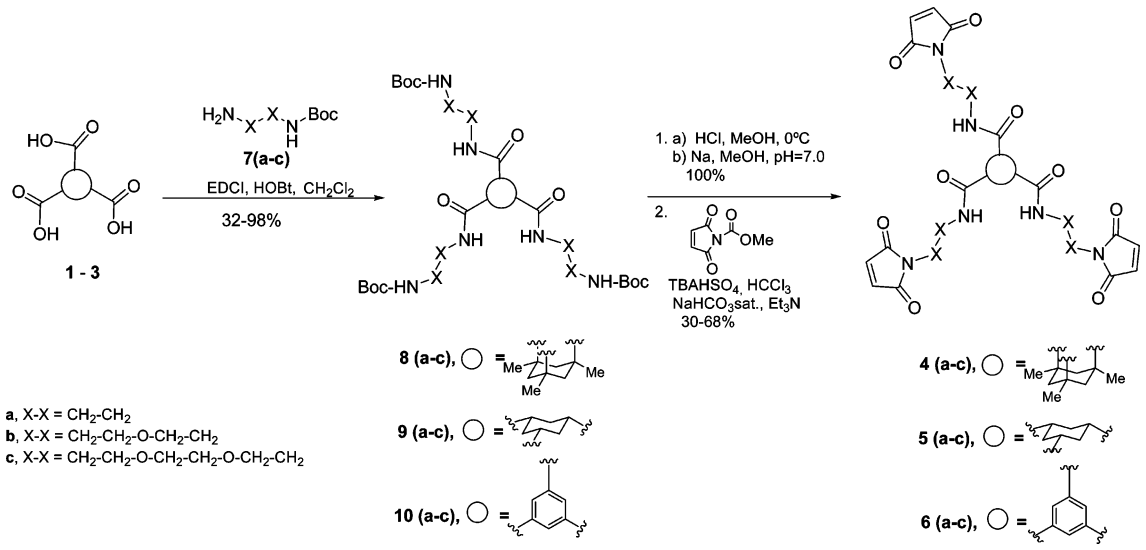
The structures of the new compounds were based on spectroscopic and analytical data (see experimental details in Supporting Information and ref 6). In order to assess the existence of hydrogen bonded amide protons, NMR experiments at different temperatures, 218, 248, 273, and 303 K, in $CDCl_3$ were carried out.[17] These experiments provided information about the presence of rotational isomers around the CO−NH bond.[18] At 303 K, only one set of signals in each NMR spectrum was observed, indicating that no rotational isomers along the NH−CO bond are present. Compounds **4a**, **5c**, and **6c** were chosen for this study. The choice was based on their solubility in $CDCl_3$ at low temperatures. Compound **4a** derived from Kemp's triacid and with the shortest spacer was totally soluble in chloroform at low temperatures. For compounds derived from *cis*,*cis*-cyclohexane-1,3,5-tricarboxylic acid and trimesic acid, only derivatives **5c** and **6c** with the longest spacers were soluble

(16) (a) Walker, M. A. *J. Org. Chem.* **1995**, *60*, 5352−5355. (b) Boeckler, C.; Frish, B.; Scuber, F. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 2055−2058. (c) Antczak, C.; Bauvois, B.; Monneret, C.; Florent, J. C. *Bioorg. Med. Chem. Lett.* **2001**, *9*, 2843−2848.

(17) (a) Deetz, M. J.; Fahey, J. E.; Smith, B. D. *J. Phys. Org. Chem.* **2001**, *14*, 463−467. (b) Gung, B. W.; Zhu, Z.; Everingham, B. *J. Org. Chem.* **1997**, *62*, 3436−3437.

(18) (a) Kim, Y.-J.; Park, Y.; Park, K. K. *J. Mol. Struct.* **2006**, *783*, 61−65. (b) Iriepa, I.; Madrid, A. I.; Gálvez, E.; Bellanato, J. *J. Mol. Struct.* **2006**, *787*, 8−13.

SCHEME 1



in chloroform at low temperatures. For both cyclohexane derivatives **4a** and **5c**, the NH chemical shift [$\delta_{(NH)}$] at 303 K is 7.42 and 6.28 ppm, respectively. This downfield shift of $\delta_{(NH)}$ observed for **4a** is an indication of an intramolecular hydrogen-bonded state. On decreasing temperature, there were very little differences in the chemical shifts (δ) of all of the signals (variations between 0.01 and 0.07 ppm), except for the NH in compounds **5c** and **6c**. This may be an indication of low flexibility in the respective moieties. The cyclohexane ring in both **4a** and **5c** is a rigid moiety as well. The variation of $\delta_{(NH)}$ with temperature is indicated in Figures 3−5, and a graph of the amide proton chemical shift as a function of the temperature is also displayed in Figure 6.

Kemp's triacid derivative **4a** presented a slight variation of $\delta_{(NH)}$ with temperature from 303 to 218 K, indicating the existence of intramolecular hydrogen bonding at these temperatures. For compounds **5c** and **6c**, a variation of $\delta_{(NH)}$ with

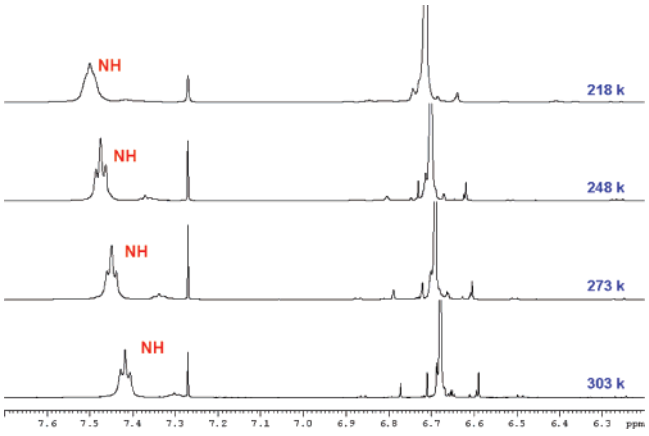


FIGURE 3. Variation of $\delta_{(NH)}$ with temperature in compound **4a**.

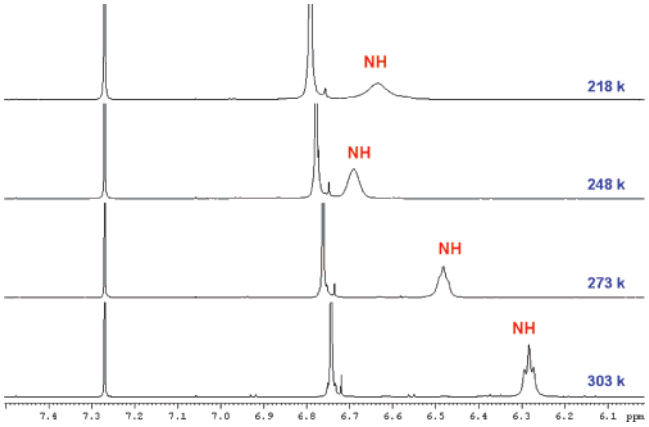


FIGURE 4. Variation of $\delta_{(NH)}$ with temperature in compound **5c**.

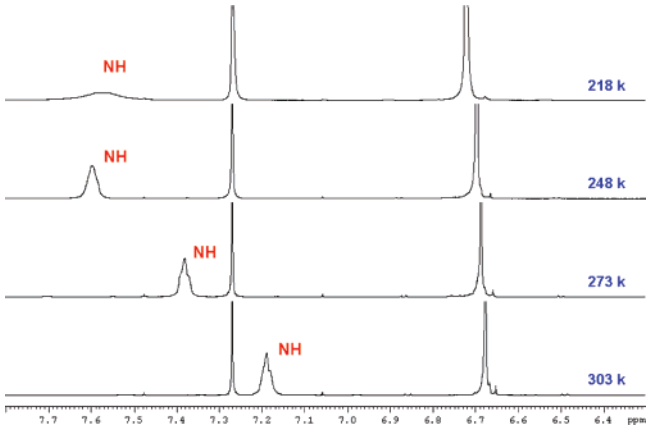


FIGURE 5. Variation of $\delta_{(NH)}$ with temperature in compound **6c**.

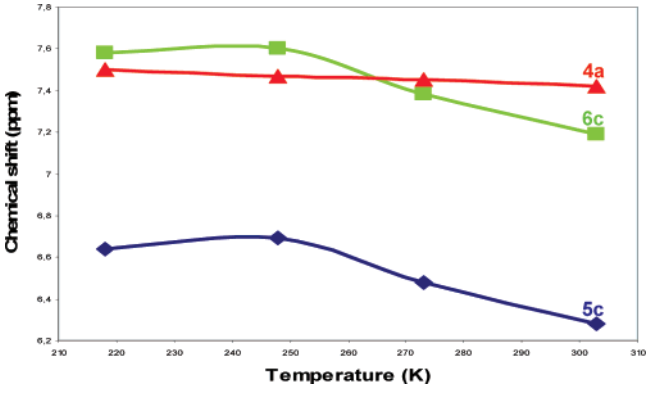


FIGURE 6. Plot of amide proton chemical shift as a function of temperature.

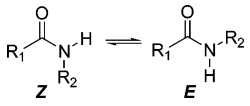


FIGURE 7. Rotamers along the CO−NH bond.

TABLE 1. Temperature Coefficients

compound	NH ($\Delta\delta/\Delta T$) (ppb/K)
4a	−1
5c	−6.6[a]
6c	−6.3[a]

[a] Considering variation from 303 to 248 K.

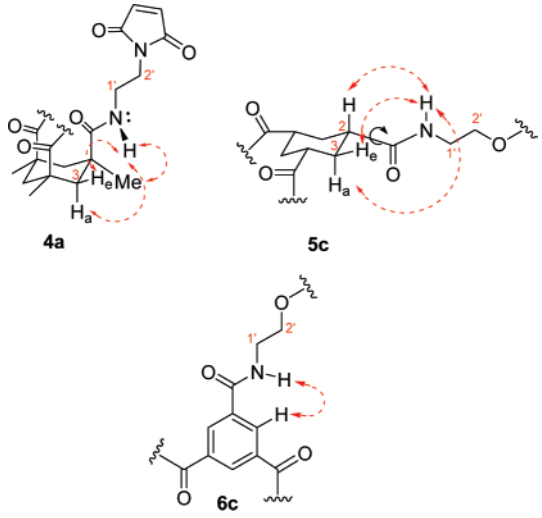


FIGURE 8. NOE contacts for compounds **4a**, **5c**, and **6c**.

temperature is observed, but on cooling to 218 K, no appreciable variation is detected. A steeper slope is observed from 303 to 248 K with an inflection point around 250 K. This behavior indicates that there is no intramolecular hydrogen bonding at higher temperatures and that, at temperatures lower than 250 K, a more ordered structure is observed. Temperature coefficients (Table 1), which are reasonably good indicators of hydrogen bonding,[19] confirm this behavior.

To elucidate which rotational isomer (*E* or *Z*) along the CO−NH bond is present (Figure 7), NOE contacts were determined and are shown in Figure 8.

For **4a**, whose structure is well defined by the intramolecular hydrogen bonds, NOE is observed between pairs of protons NH/$H_{2'}$ and NH/$H_{1'}$ of the ethylene chain (not shown). A strong NOE between pairs NH/H_{3e} and, to a lesser extent, for NH/CH_3 is also observed. The absence of an NOE between $H_{1'}$ or $H_{2'}$ with either Me or H_{3a} and H_{3e} is an indication of the presence of the *E* rotational isomer. For **5c**, NOE signals are observed for $H_{2'}$ and $H_{1'}$ (not shown) of the ethylene chain interacting with H_{3a}/H_{3e} and between the pair of protons NH/H_{3a}, NH/H_{3e}, and NH/H_2. This can be an indication of free rotation along the C(H_2)−CO bond. For the aromatic derivative **6c** a strong NOE between the pair of protons NH/H_{Ar} is observed, indicating the presence of the *E* rotational isomer (Figure 8).

(19) (a) A temperature coefficient more positive than −4.5 ppb/K indicates a hydrogen bond, whereas if it exchanges rapidly and has a temperature coefficient more negative than −4.5 ppb/K, it is not hydrogen bonded. See: Baxter, N. J.; Williamson, M. P. *J. Biomol. NMR* **1997**, *9*, 359−369. (b) Sosnicki, J. G.; Hansen, P. E. *J. Mol. Struct.* **2004**, *700*, 91−103.

TABLE 2. Summary of Conformational Search Results for 4−6

compound	no. of structures found within 4 kcal/mol of the lowest E structure	no. of blocks of 5000 steps completed
4a	12	16
4b	240	35
4c	2626	35
5a	271	35
5b	3173	35
5c	1955	35
6a	2563	35
6b	8430	35
6c	1388	40

These results are in accordance with the OPLS05 based computational data (see below).

Computational Studies

Force Field Analysis. Seven different force fields are available in the MacroModel program. It is important to choose a force field that is well parametrized for the molecular system under study. Accurate torsional parameters are particularly important in flexible molecular systems since they control conformational interconversions. Force field analysis proceeded as follows. Conformational searches of the surfaces generated by each force field (MM2*, MM3*, AMBER*, MMFF, OPLS, OPLS01, and OPLS05), in conjunction with the GBSA (water) model, for **4a**, were performed for 125,000 steps. Low-energy structures found on each surface were selected and subjected to unrestrained quantum mechanical minimization using B3LYP/6-311G**-SCRF(water). The quantum calculations found chair and boat conformations of **4a** to be surprisingly similar in enthalpy (within 2 kcal/mol); however, these results were not in agreement with the observed spectra or previous reports.[20] We are investigating further these aspects of the conformational preferences of *cis*,*cis* derivatives of Kemp's triacid. The computational outcome will be reported elsewhere.[21] Force field analysis was repeated for **5a** and **6a**, but in these cases only a subset of force fields (AMBER*, OPLS05) was examined as preliminary data showed the results to be invariant. From these data, along with a comparison with the experimental NMR results, it was determined that the OPLS05 force field was, on average, the best suited for describing **4**−**6** and was used in the remainder of the report.

The conformational flexibility of each of these systems was investigated using the LM:MC conformational search method. Details of each search, including convergence, can be found in Supporting Information; a summary of the results appears in Table 2. The results indicate that each search is convergent with respect to the number of structures found, the energy of the lowest conformer, and the number of times this structure is visited. Very few new structures are found toward the end of each search, and searches on the smaller molecular systems (**4a**−**6a**, **4b**, and **5b**) find the lowest energy structure within the first block of 5000 steps. This suggests that the method is

(20) (a) Bencini, A.; Bianchi, A.; Burguete, M. I.; Dapporto, P.; Domenech, A.; Garcia-Espana, E.; Luis, S. V.; Paoli, P.; Ramirez, J. A. *J. Chem. Soc., Perkin Trans.* **1994**, *2*, 569−577. (b) Kemp, D. S.; Petrakis, K. S. *J. Org. Chem.* **1981**, *46*, 5140−5143. (c) Chan, T.-L.; Cui, Y.-X.; Mak, T. C. W.; Wang, R.-J.; Wong, H. N. N. *J. Crystallogr. Spectrosc. Res.* **1991**, *21*, 297−308. (d) Rebek, J.; Marshall, L.; Wolak, R.; Parris, K.; Killoran, M.; Askew, B.; Nemeth, D.; Islam, N. *J. Am. Chem. Soc.* **1985**, *107*, 7476−7481.

(21) Remmert S.; Parish, C. Unpublished work.

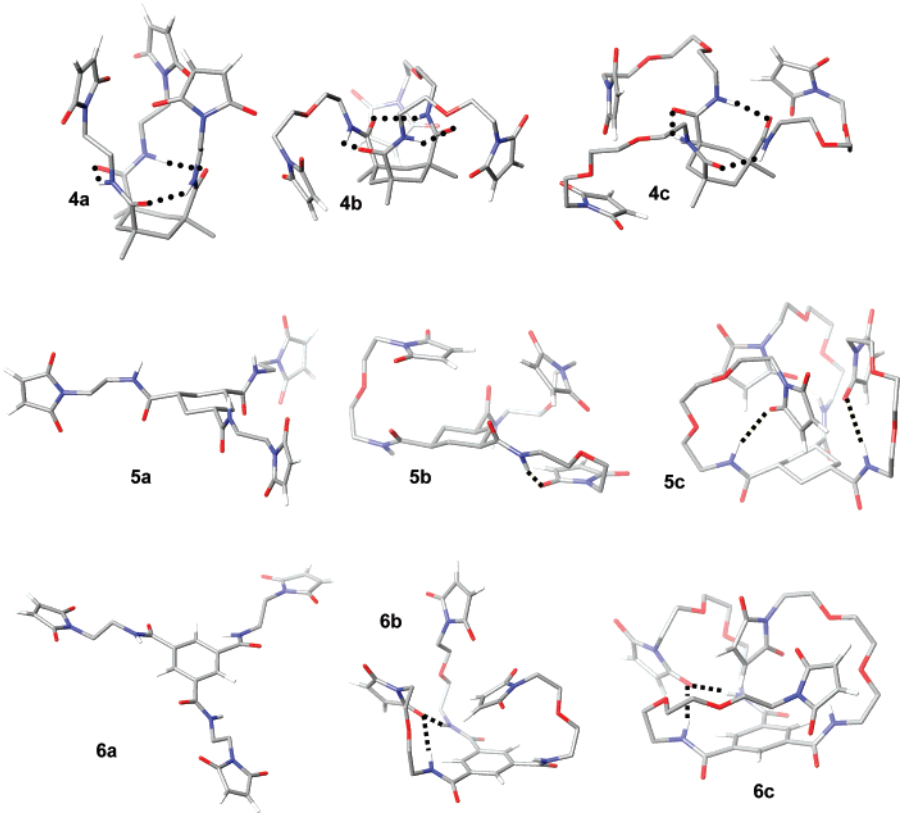


FIGURE 9. Lowest-energy structures and hydrogen-bonding patterns found on the OPLS2005/GBSA(water) surface of **4**−**6**.

indeed exhaustively sampling the available conformational space and that the generated ensembles are representative of the low-energy structures of **4**−**6** on the OPLS05/GBSA(water) surface.

Analysis of the Global Minima. The lowest energy structures for **4**−**6** are shown in Figure 9. A comparison of the global minima for compounds based on the Kemp's triacid 1,3,5-trimethyl cyclohexane (TMC6) scaffold (**4a**, **4b**, and **4c**) indicates that on the OPLS05/GBSA(water) surface these molecules preferentially adopt chair conformations with the methyl substituents arranged equatorially, and the maleimide (M) arms arranged axially for optimal hydrogen bonding between the amide proton on one arm and the amide carbonyl on another arm. Three such hydrogen-bonding interactions occur in each of the lowest energy structures for **4a**−**4c**. The lowest energy structure for **4a** is a very ordered C3-symmetric structure; in addition to the structural features described above, a carbonyl moiety on each maleimide points toward the nitrogen on a neighboring heteroaromatic ring (distances ∼3.4 Å). The NMR data are in accordance with these results (see above). The lowest energy structure for **4b** contains maleimide-ethoxy-ethyl carboxamide (MEEC) arms that are long enough to allow the chains to fold back toward the side of the central TMC6 ring, allowing relatively close (3.6 Å) interactions between the amide carbonyl on one chain and the maleimide ring on another chain. This results in a more compressed and less ordered structure relative to the lowest energy conformer for **4a**. The maleimide-diethoxy-ethyl carboxamide (MDEC) arms in **4c** are quite flexible, and although three hydrogen bonds still exist, the molecule appears more disordered than either **4a** or **4b**.

A comparison of the global minima for **5a**, **5b**, and **5c**, molecules with a simple cyclohexane (C6) core, reveals much "squatter," more disordered molecular shapes relative to those of **4a**−**4c**. Each C6 ring adopts a chair conformation with the maleimide (M) arms in the equatorial positions. Structures **5a** and **5b** do not contain any hydrogen bonding interactions between different M arms; however, in **5b** the heteroaromatic ring on one arm wraps back around to hydrogen bond with the amide hydrogen of the same arm. Structure **5c** possesses C3 symmetry, no hydrogen bonding between arms, and three hydrogen bonds between heteroaromatic rings and the amide hydrogen on the same arm.

Systems **6a**, **6b**, and **6c** contain central aromatic rings in place of the C6 or TMC6 rings. This causes the lowest energy structure for **6a** to be roughly planar with the maleimide arms radiating from the aromatic core, preventing any hydrogen bonding interactions. The larger arm size of **6b** and **6c** allow these molecules to adopt nonplanar lowest energy structures wherein the MEEC and MDEC arms wrap back toward the central aromatic ring allowing hydrogen bonding between the arms and π-stacking interactions between the central ring and heteroaromatic maleimide rings. In **6b** and **6c**, one arm wraps back to form a bifurcated hydrogen bond between the carbonyl of the heteroaromatic ring and the amide of the same arm and an adjacent arm.

Analysis of the Conformational Ensembles. All of the C6 and TMC6 global minimum energy structures (**4a**−**4c**, **5a**−**5c**) shown in Figure 9 adopt the chair conformation on the OPLS05/GBSA(water) surface. Global minimum energy structures are often not representative of the overall ensemble of structures[22] but in this case they are, at least with respect to ring orientations; 100% of the structures found within 4 kcal/mol for **4a**−**4c** and

(22) Parish, C.; Lombardi, R.; Sinclair, K.; Smith, E.; Goldberg, A.; Rappleye, M.; Dure, M. *J. Mol. Graphics Modell.* **2002**, *21*, 129−150.

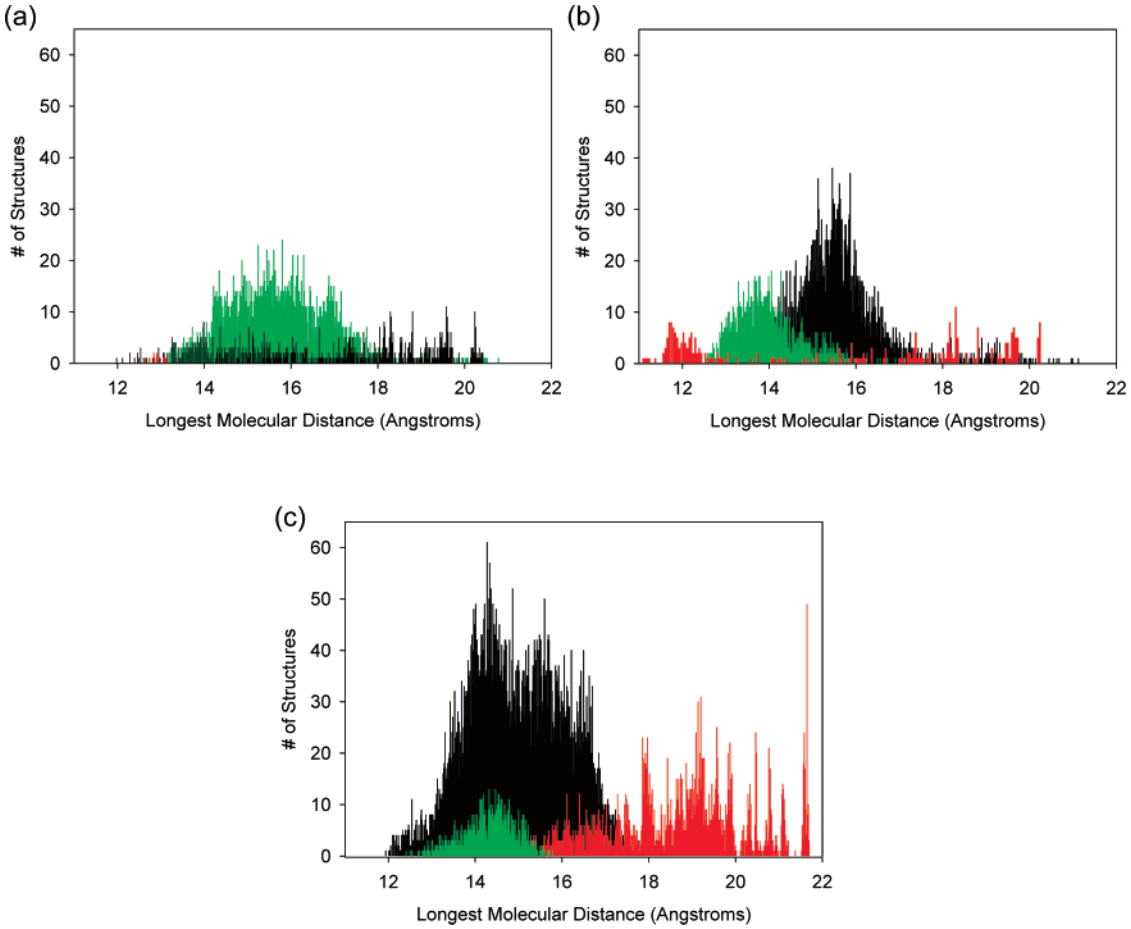


FIGURE 10. Longest molecular distance: (a) compounds **4a** (red), **4b** (black), **4c** (green); (b) compounds **5a** (red), **5b** (black), **5c** (green); (c) compounds **6a** (red), **6b** (black), **6c** (green).

TABLE 3. Molecular Length Data (Å) for 4−6

compound	average	minimum	maximum	range
4a	12.7	12.5	12.9	0.4
4b	14.4	12.0	18.9	6.9
4c	15.7	12.7	20.8	8.1
5a	15.6	11.1	20.3	9.2
5b	15.5	12.9	21.1	8.2
5c	13.9	12.6	17.1	4.5
6a	17.9	13.4	20.6	7.2
6b	14.0	11.8	18.4	6.6
6c	14.6	12.4	16.2	3.8

5a−**5c** adopt the chair conformation on the OPLS05/GBSA-(water) surface.[23] For **4a**−**4c**, 100% of the ensemble contains structures with arms in the axial orientation, whereas 100% of the structures of **5b** and **5c** display arms in the equatorial position. The ensemble of **5a**, the C6 system with the smallest arms, contains a mixture of structures with 40% all axial and 60% all equatorial arms.

(23) Structures were determined to be chairs or boats by measuring the dihedral angle of each arm relative to the central ring. All angles greater than 125° were identified as axial substituents, whereas those less than or equal to 125° were defined as equatorial. Those structures that contained all axial or all equatorial arms were counted as chairs, and those with mixtures of axial and equatorial arms were defined to be boats. We were concerned that perhaps our conformational search was overly biased by the chair starting structure; however, simulations purposely seeded with boat conformations as initial structures also resulted in 100% chair ensembles.

If every conformation is structurally distinct from all other conformations, then the number of unique minima provides quantitative information about molecular flexibility and the complexity of each potential energy surface. A larger number of unique minima indicate a more complex potential energy surface. Judging by the number of minima found on each surface, the molecules with methyl groups on the cyclohexane rings are less flexible for the small and medium length chains than are either the aromatic or unsubstituted C6 systems. For systems with large chains (**4c**, **5c**, and **6c**), the aromatic molecule has the smallest number of minima. According to the number of minima, the overall ordering of flexibility is **4a** < **4b** < **5a** < **6c** < **5c** < **6a** < **4c** < **5b** < **6b**. However, it is possible that the minima are not structurally distinct, i.e., that multiple structures can be grouped into conformationally similar families. When this occurs, the number of families is a better indicator of flexibility than the number of conformations. Therefore, the XCluster program was used to determine if structures within each ensemble naturally form structurally related groupings. Clustering by atomic RMS after rigid body superposition of all heavy atoms in all ensembles and ring atoms for cyclohexane ensembles **4a**−**4c**, **5b**, and **5c** did not lead to significant clustering as evidenced by distance maps, mosaics, and separation ratios. These data, along with ring conformer analysis and visual inspection of the smaller ensembles, indicates that the low-energy structures within each ensemble adopt similar core ring structures (chair or aromatic) but with maleimide arm orientations that are in fact geometrically distinct and that span

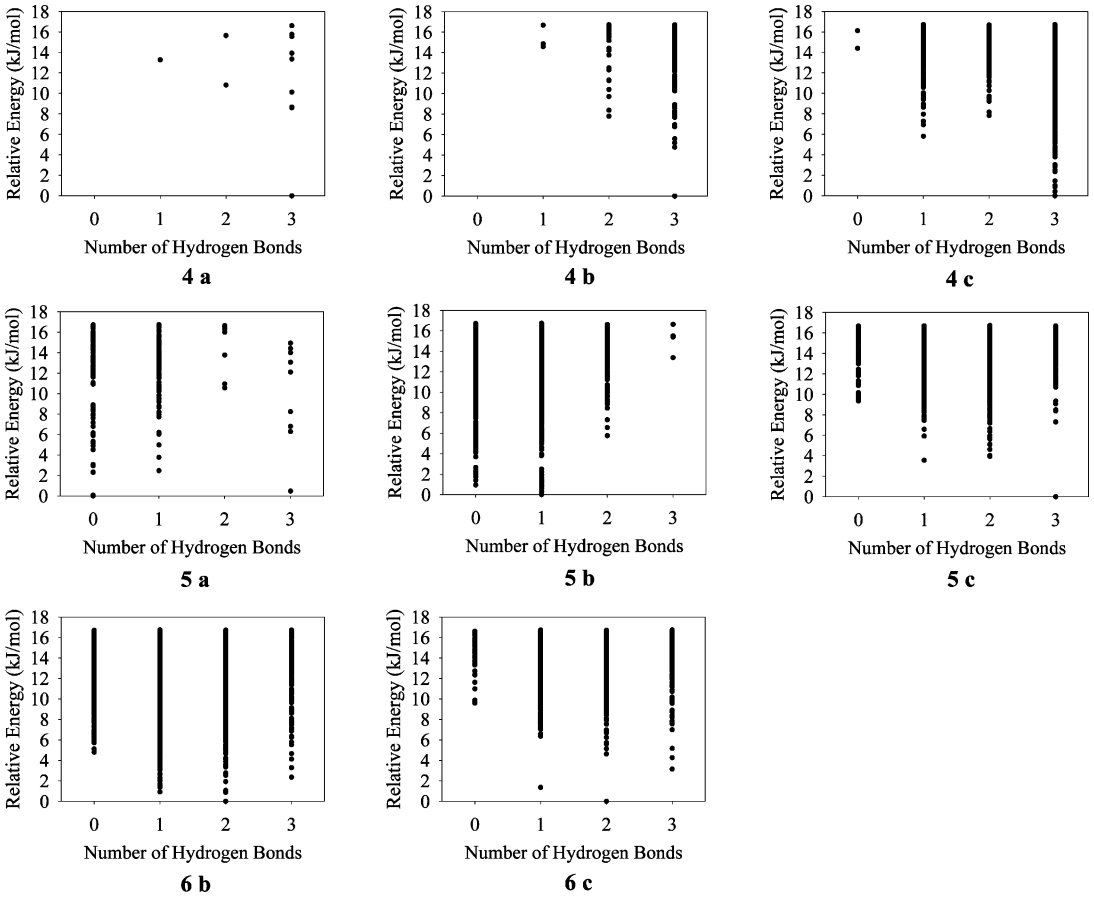


FIGURE 11. Hydrogen bonding and energetic analysis for ensembles **4a**−**c**, **5a**−**c**, and **6b,c.** The ensemble for **6a** did not contain any structures with hydrogen bonding, most likely due to the short M arms and aromatic nature of the compound.

the available conformational space. The only system that exhibited statistically significant clustering (separation ratios greater than 2.0)[24] was **5a**. Clustering of this ensemble by ring atom RMSD resulted in two families; analysis of the average and representative structure indicated that both families adopted a chair orientation and 60% of the ensemble contained structures with equatorial maleimide arms while the other 40% of the ensemble contained axial arms in agreement with the independent geometric analysis above.

Analysis of Molecular Lengths. To obtain an idea of the distribution of structures in each ensemble as well as the average molecular length of each system, we measured the longest interatomic distance of each structure and visualized histograms of these data (Figure 10 and Table 3). For TMC6 systems (**4a**−**4c**), the overall molecular behavior, as measured by this metric, is what one would expect from the 2D structures shown in Figure 2. As the chain joined to the maleimide moeity increases, the overall molecular length increases, from an average of 12.7 Å for **4a** to 14.4 and 15.7 Å for **4b** and **4c**, respectively, suggesting an increase in the corresponding 3D molecular size. These averages are accompanied by a corresponding increase in the ranges associated with each system: 12.5 − 12.9 Å for **4a**; 12.0 − 18.9 Å for **4b**; 12.7 − 20.8 Å for **4c**, indicating a positive correlation between molecular length and flexibility. The situation for molecules with C6 and aromatic cores is different. For **6a**, **6b**, and **6c** (aromatic core), the ensemble data indicates a *decrease* in molecular size as arm length increases, along with a corresponding loss in flexibility likely due to increased opportunities for hydrogen bonding and π-stacking, as demonstrated in the low-energy structures in Figure 9. For C6 systems, the ensembles for **5a** and **5b** are relatively similar with respect to molecular lengths; however, **5a** is slightly more flexible than **5b** judging by the range of distances sampled (for **5a**, range = 11.1 −20.3 Å and average = 15.6 Å; for **5b**, range = 12.9 − 21.2 Å and average = 15.5 Å). Ensemble **5c** contains the smallest range of intramolecular distances for C6 systems (12.6−17.1 Å), as well as significantly shorter distances sampled (average = 13.9 Å), even though this molecule contains the longest arms. According to the average longest interatomic distances, the overall ordering of molecular length is **4a** < **5c** ≈ **6b** < **4b** ≈ **6c** < **5b** ≈ **5a** ≈ **4c** < **6a**.

Hydrogen Bonding. The number of classical hydrogen bonds (defined as having a maximum distance between donor and acceptor of 2.5 Å, and a minimum donor and acceptor angle of 120° and 90°, respectively) in each ensemble was measured (Figure 11 and Table 4.) The overwhelming majority of TMC6 structures (**4a**, **4b**, and **4c**) found on the OPLS2005/GBSA-(water) surface contain three hydrogen bonds; all structures for **4a** and **4b** contain at least one hydrogen bond while the number of structures without any hydrogen bonds for **4c** is less than 1%. As shown in Figure 11, structures with less hydrogen bonding interactions typically occur in the ensemble at higher energies. Taken together with the flexibility and longest distance data, this information provides a clear picture of the overall

(24) Shenkin, P. S.; McDonald, D. Q. *J. Comput. Chem.* **1994**, *15*, 899−916.

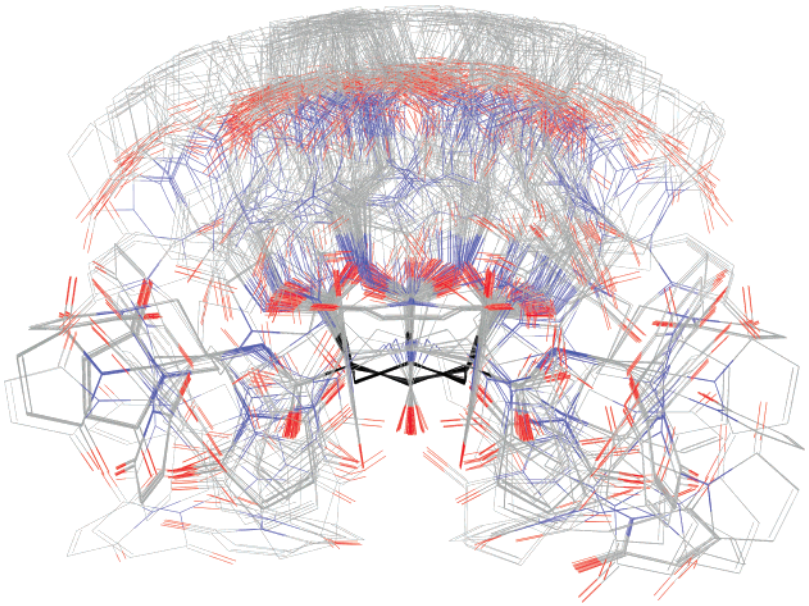


FIGURE 12. Superimposition of 271 low-energy structures in the ensemble of **5a** using cyclohexane ring atoms.

TABLE 4. Percent of Structures within the Ensembles of 4−6 That Contain 0, 1, 2, and 3 Hydrogen-Bonding Interactions

	no. of hydrogen bonds			
compound	0	1	2	3
4a	0	8	17	75
4b	0	1	12	87
4c	<1	7	7	86
5a	53	41	3	3
5b	51	45	4	<1
5c	5	30	56	9
6a	0	0	0	0
6b	15	59	24	2
6c	3	44	41	12

behavior of these systems. The ensembles contain structures in the chair conformation with equatorial methyl groups and axial maleimide arms whose flexibility of motion sweeps out the available conformational space in a coordinated way so as to maintain optimal hydrogen bonding. The situation is somewhat different for systems based on the unsubstituted cyclohexane core (**5a**−**5c**). Ensembles with shorter arms (**5a** and **5b**) contain a mix of structures, with (∼50%) and without (∼50%) hydrogen bonding interactions, whereas the ensemble with the longest arms (**5c**) contains a much smaller percentage of non-hydrogen-bonded structures (5%). The hydrogen bonding in **5c** causes this system to be much less flexible and smaller on average than **5b**. As Figure 11 indicates, structures with larger numbers of hydrogen bonds tend to occur at higher energies in the ensemble of **5b**. As with ensembles based on the TMC6 core, the ensembles of **5** contain structures that adopt the chair conformation with equatorial maleimide arms that sweep out all of the available conformational space. This behavior is illustrated in the superposition of all 271 low-energy structures of **5a** as shown in Figure 12. There were no hydrogen-bonded structures found for **6a** because of the planar aromatic nature of the molecule and short maleimide arms. For all other structures with aromatic cores, the maleimide arms are long enough to wrap around to form 1−3 hydrogen bonding interactions, with the majority of the structures for **6b** and **6c** containing at least one such interaction.

Four of the compounds presented in this study (**4a**−**4c**, **6c**) were ligated to miniproteins, and their virus inhibition was measured.[6] Compounds **4a**−**4c** behave similarly (IC_{50} values are **4a** = 0.79 μM, **4b** = 1.8 μM, and **4c** = 1.0 μM) in a biological setting. This is not surprising given the overall similarities of the molecular ensembles and the resulting rigidity due to the preponderance of intramolecular hydrogen bonding. Interestingly the aromatic compound **6c** behaves similarly (IC_{50} values = 0.33 μM), and of all of the aromatic systems, this ensemble is the least flexible (judging by the distance data), contains the largest number of hydrogen-bonded structures, and appears to behave in a very compacted fashion. The modeling results indicate that, in most cases, the trimeric compounds with longer arms would provide a more conformationally flexible scaffold to which to attach CD4 mimetic miniproteins. However, such increased mobility generally results in ensembles of shorter molecular length.

Conclusions

The synthesis and characterization of nine homotrifunctional cross-linking reagents are presented. These systems contain maleimide arms of varying length attached to three well-defined molecular architectures: Kemp's triacid, *cis*,*cis*-cyclohexane-1,3,5-tricarboxylic acid, and trimesic acid. Temperature-dependent NMR analysis of those systems soluble in chloroform over a range of temperatures (**4a**, **5c**, and **6c**) confirmed the existence of hydrogen-bonding interactions. Molecular behavior and conformational flexibility was further probed using the mixed Low Mode−Monte Carlo conformational searching technique to exhaustively sample the corresponding OPLS2005/GBSA(water) surface. Geometric structure, molecular length, and hydrogen-bonding patterns were analyzed. Low-energy ensembles based on Kemp's triacid and *cis*,*cis*-cyclohexane-1,3,5-tricarboxylic acid contain structures that adopt preferentially chair conformations with axial and equatorial maleimide arms, respectively. Increasing chain length often resulted in overall shorter molecular length due to additional chain flex-

ibility. These results were consistent with two-dimensional temperature-dependent NMR studies.

Experimental Section

Synthesis and NMR. General Remarks. ^{1}H NMR and ^{13}C NMR spectra were obtained for solutions in $CDCl_3$ and DMSO-d_6, J values are given in Hz, and δ in ppm. ^{1}H NMR spectra at different temperatures were obtained for solutions in $CDCl_3$ at 500 MHz. All assignments were confirmed by homonuclear 2D COSY and NOESY and heteronuclear 2D correlated (HETCOR) experiments. The FAB mass spectra were obtained with glycerol or 3-nitrobenzyl alcohol as matrix. TLC was performed on silica gel HF254, with detection by UV light and nynhidrine reagent. Silica gel 60 (230 mesh) was used for preparative chromatography. Anhydrous solvents and reagents were freshly distilled under N_2 prior to use.

Synthesis of Boc-Protected Triamine Templates 8a−8c, 9a−9c, and 10a−10c). General Procedure. To a cooled (0 °C) suspension of the corresponding tricarboxylic acid **1**−**3** (0.5 mmol) in dry CH_2Cl_2 was added EDCI (1.65 mmol). The *N*-Boc-protected diamine **7(a−c)** (1.65 mmol) was then added dropwise, and the reaction was left to warm to room temperature and stirred overnight. The mixture was washed with water, and the organic phase was dried over anhydrous sodium sulfate. Chromatography purification on silica gel (CH_2Cl_2/MeOH 3% → 5%) afforded the target compounds. (See Supporting Information for analytical and spectroscopic data).

Synthesis of Maleimide Templates 4a−4c, 5a−5c, and 6a−6c). General Procedure. The *N*-Boc-protected triamine [**8a**−**8c**, **9a**−**9c**, **10a**−**10c**] (0.2 mmol) was dissolved in MeOH and cooled to 0 °C, and concentrated HCl (2.2 mL) was added dropwise. The reaction was stirred at 0 °C until TLC showed lack of substrate. Solvents were evaporated, and the residue dissolved again in methanol and neutralized with sodium to pH = 7.0. After filtration of the NaCl formed, the filtrate was dried over anhydrous sodium sulfate and concentrated to afford the corresponding free triamine. A solution of this triamine (0.23 mmol) in $CHCl_3$ (10 mL) was cooled to 0 °C, and then tetrabutylammonium hydrogen sulfate (1.38 mmol) and *N*-methoxycarbonylmaleimide (1.38 mmol) were added. Triethylamine (97 μL) was then slowly added, and the mixture stirred at 0 °C for 10 min. After removing the ice bath, saturated aqueous solution of $NaHCO_3$ (10 mL) was added in one portion. The reaction was stirred vigorously at room temperature for 3−4 h. Phases were then separated, and the organic layer was then dried over anhydrous sodium sulfate, concentrated, and purified by column chromatography on silica gel (CH_2Cl_2/MeOH 3% → 5%) to give the target compounds. (See Supporting Information for analytical and spectroscopic data).

Computational Methods. The conformational ensembles generated in this study were calculated using the MacroModel V9.1 suite of software[25] programs. A variety of force fields available in the MacroModel software package were evaluated for their ability to reproduce quantum mechanical conformational data obtained at the B3LYP/6-311G* level using Jaguar V6.5.[26] Solvent effects were included using the Generalized Born/Surface Area (GB/SA)[27] and Self-Consistent Reactive Field (SCRF)[28] continuum models for water, in the molecular and quantum mechanical calculations, respectively.

The Low Mode (LM) search method[29] was used in a 1:1 combination[22] with the Monte Carlo (MC) search method[30] to explore the potential energy surfaces of **4**−**6**. Interconversion of ring structures was enabled using the ring-opening method of Still.[31] Searches utilized the usage-directed structure selection method[30] that identifies the least-used structure from among all known conformations and then uses this structure as the starting point for each new search. This ensures that a variety of different starting structures from different regions of the potential energy surface are used to begin each new block of steps. During the conformational search all structures were subjected to 500 steps of the Truncated Newton Conjugate Gradient (TNCG)[32] minimization method to within a derivative convergence criterion of 0.01 kJ Å^{-1} mol^{-1}. All final ensembles were subjected to further TNCG minimization after the search was completed in order to obtain fully minimized structures. Ensembles generated for each of the scaffolds were queried for geometric similarities using the XCluster[24] program.

Acknowledgment. The authors thank Nora Green for her interest in the work, the European Commission (FP6, TRIoH, LSHG-CT-2003-503480), the Ministerio de Educación y Ciencia of Spain (CTQ2004-00649/BQU), and the Junta de Andalucía (FQM-345) for financial support. Acknowledgment is also made to the Donors of the American Chemical Society Petroleum Research Fund, the National Science Foundation under grant CHE-0211577, and the Thomas F. Jeffress and Kate Miller Jeffress Memorial Trust for partial support of this work. C.P. also acknowledges support from the Camille and Henry Dreyfus Foundation through receipt of a Henry Dreyfus Teacher-Scholar award. S.R. acknowledges support from the Arnold and Mabel Beckman Foundation through receipt of a Beckman Scholars award. Computational resources were provided, in part, by the MERCURY supercomputer consortium (http://mars.hamilton-.edu) under NSF grant CHE-0116435.

Supporting Information Available: Data of the compounds **4a**−**4c**, **5a**−**5c**, **6a**−**6c**, **8a**−**8c**, **9a**−**9c**, **10a**−**10c**; ^{1}H and ^{13}C NMR spectra for compounds **5b**, **5c**, **6a**, **6b**, **9b**, **9c**, and **10b**; NOESY experiments and NOE contacts for compounds **4a**, **5c**, and **6c**; force field parameter analysis for compounds **5c** and **6c**; LM:MC conformational search results on the OPLS2005/GBSA(water) surface of **4a**−**4c**, **5a**−**5c**, and **6a**−**6c**, and DFT results for **4a**. This material is available free of charge via the Internet at http://pubs.acs.org.

JO0709293

(25) Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. *J. Comput. Chem.* **1990**, *11*, 440−467.

(26) *Jaguar*, V6.5; Schrodinger, Inc.: Portland, OR, 1991−2000; www.schrodinger.com.

(27) (a) Hasel, W.; Hendrickson, T. F.; Still, W. C. *Tetrahedron Comp. Methodol.* **1988**, *1*, 103−116. (b) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. *J. Phys. Chem. A* **1997**, *101*, 3005−3014. (c) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129. (d) Weiser, J.; Shenkin, P. S.; Still, W. C. *J. Comput. Chem.* **1999**, *20*, 217−230. (e) Weiser, J.; Shenkin, P. S.; Still, W. C. *J. Comput. Chem.* **1999**, *20*, 586−596. (f) Weiser, J.; Weiser, A. A.; Shenkin, P. S.; Still, W. C. *J. Comput. Chem.* **1998**, *19*, 797−808.

(28) Klicic, J. J.; Friesner, R. A.; Liu, S. H.; Guida, W. C. *J. Phys. Chem.* **1998**, *106*, 1327−1335.

(29) (a) Kolossvary, I.; Guida, W. C. *J. Am. Chem. Soc.* **1996**, *118*, 5011−5019. (b) Kolossvary, I.; Guida, W. C. *J. Comput. Chem.* **1999**, *20*, 1671−1684.

(30) Chang, G.; Guida, W. C.; Still, W. C. *J. Am. Chem. Soc.* **1989**, *111*, 4379−4386.

(31) Still, W. C.; Galynker, I. *Tetrahedron* **1981**, *37*, 3981−3996.

(32) Ponder, J. W.; Richards, F. M. *J. Comput. Chem.* **1987**, *8*, 1016 1024.