Bioorganic & Medicinal Chemistry 17 (2009) 1251–1258

Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry

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

Conformational analysis of trimeric maleimide substituted 1,5,9-triazacyclododecane HIV fusion scaffolds

Sarah Remmert, Heather Hollis, Carol A. Parish *

Department of Chemistry, University of Richmond, Gottwald Science Center, Richmond, VA 23173, USA

ARTICLE INFO

Article history:
Received 14 August 2008
Revised 7 December 2008
Accepted 10 December 2008
Available online 24 December 2008

Keywords:
Conformational analysis
Molecular flexibility
HIV fusion inhibitors
Cross-linking
Maleimide inhibitors
Low Mode Monte Carlo

ABSTRACT

An analysis of the conformational preferences of three trimeric maleimide substituted 1,5,9-triazacyclododecane derivatives, proposed as cross linking reagents for HIV-1 fusion inhibitors, is presented. Exhaustive sampling was performed using the mixed Low Mode Monte Carlo conformational searching technique on the corresponding OPLS2005/GBSA(water) potential energy surface. Geometric structure, molecular length, and hydrogen bonding patterns of the compounds are analyzed. Global minimum energy structures were verified as minima using B3LYP/6-31G* geometry optimization. All structures adopt a crown-like 12-membered ring conformation; however, the system with the shortest maleimide arms (**1a**) can also adopt alternative ring orientations. Overall, derivatives with longer maleimide arms were more flexible and resulted in ensembles with a larger number of low energy structures. Comparison with biological inhibition data indicates that there is very little relationship between molecular size and the ability of the scaffold to orient CD4M9 miniproteins for optimal inhibition; however hydrophobicity may play a role.

1. Introduction

Growing resistance to anti-viral HIV therapies as well as toxicity of currently available drugs necessitate continued investigation into novel targets for viral inhibition.[1–4] HIV-1 fusion with the host cell membrane is a promising target given the possibility of inhibiting the viral life cycle prior to cell entry, negating the need for cellular uptake of the drug.[5] Resolution of key structural properties of the HIV and SIV membrane Env protein complex responsible for the fusion event,[6–13] in combination with the clinical success of synthetic peptide fusion inhibitors,[14] suggests the development of multivalent inhibitory complexes.[15,16]

The Env protein complex mediates viral fusion and is comprised of two covalently associated trimeric glycoproteins, gp120 and gp41.[17] The external gp120 binds with the CD4 receptor of the host cell in part through insertion of the CD4 Phe-43 residue into a hydrophobic pocket on the gp120 surface[7,10] (Fig. 1). Modeling studies in combination with crystallization data of the core gp120 structure have shown that the oblique interaction between the CD4 receptor and the gp120 monomer allows for multiple CD4 receptors to interact with the trimeric viral complex concurrently.[16] Following gp120 recognition by the CD4 receptor, interaction with chemokine co-receptors CXCR4 or CCR5 triggers a cascade of conformational changes in gp120. This induces activation of the glycoprotein gp41 and drives the fusion event.[6,18–20]

Synthetic peptides have been developed to inhibit fusion through interaction with both gp120 and gp41. Fuzeon (T-20) is the only currently FDA approved fusion inhibitor. This 36 amino acid synthetic peptide associates with the N-terminal helix of gp41, thereby preventing formation of the six helical bundle and fusion progression.[21] Synthetic miniproteins, such as CD4M9, a 28-mer based on a scorpion toxin that mimics the ability of CD4 to bind to the Phe-43 pocket of gp120, have also been shown to

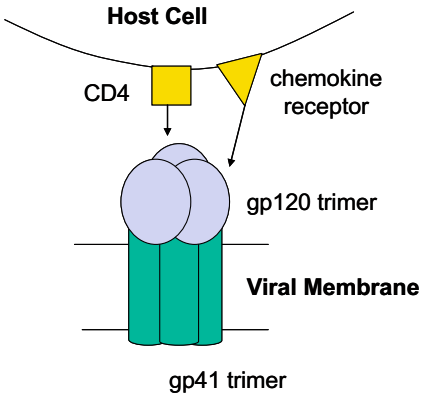

Figure 1. A schematic of the viral fusion process and the role of gp120 and gp41 glycoproteins.

* Corresponding author. Tel.: +1 8044841548.
E-mail address: cparish@richmond.edu (C.A. Parish).

0968-0896/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2008.12.027

have anti-viral activity.[14] The trimeric nature of the Env complex suggests that increased potency of inhibition of either gp41 or gp120 could be achieved by linking two to three inhibitory peptides with suitable scaffold compounds.[16] Wang and coworkers describe the synthesis of multivalent inhibitory complexes by attaching maleimide arms to well-defined architectures such as *cis,cis*-1,3,5-trimethyl cyclohexane tricarboxylic acid (Kemp's triacid), *cis,cis*-cyclohexane-1,3,5-tricarboxylic acid, trimesic acid or 1,5,9-triazacyclododecane and ligating these arms to cysteine residues on modified CD4M9 miniproteins.[15] In vivo studies using these trivalent inhibitors with HIV-1_{Bal} and HIV-1_{IIIB} show enhanced viral inhibition relative to the monomeric miniprotein. Conformational analysis of scaffolds comprised of derivatives of Kemp's triacid and trimesic acid has been previously performed.[22] However, analysis has not been performed for complexes using trimeric triazacyclododecane compounds of varying substituent chain length as cross-linking reagents, which exhibited up to 142 times the anti-viral activity of individual CD4M9 peptides against HIV-1_{Bal}, and up to 30 times the anti-viral activity of CD4M9 peptides against HIV-1_{IIIB} (Fig. 2).[15]

The gas-phase conformational preferences of unsubstituted 1,5,9-triazacyclododecane in various states of protonation has been characterized previously using an effective core potential approach.[23] The conformational space available to the cyclododecane ring was explored partially using chemical and geometrical intuition. This method was shown to be in good agreement with all-electron 3-21G(N*) calculations on smaller, related systems. They found that the gas-phase conformational preferences were driven by the formation of intramolecular hydrogen bonds and the repulsion between lone pairs (or protons) on the nitrogen atoms. The propylenediamine bridges of the triazacyclododecane were found to be flexible with a preference for gauche, *syn* or skew orientation. While these previous results will guide our study; to our knowledge, the solvent-phase conformational analysis for unprotonated, substituted, non-chelating triazacyclododecane compounds has not been performed. Therefore, this study presents a theoretical investigation of the structural properties of the triazacyclododecane derivatives as a means of interpreting the biological inhibition data.

2. Methods

2.1. Conformational searching

Conformational searching is used to explore complex multidimensional molecular potential energy surfaces and to identify low energy structures on the surface. The conformational degrees of freedom available to compounds **1a**-**1c** are significant and demand the use of an efficient sampling technique. Each compound possesses a 12-membered macrocycle with maleimide arms of varying length and these features represent a major challenge to any method. In this work, a 1:1 hybrid combination[24] of Low Mode

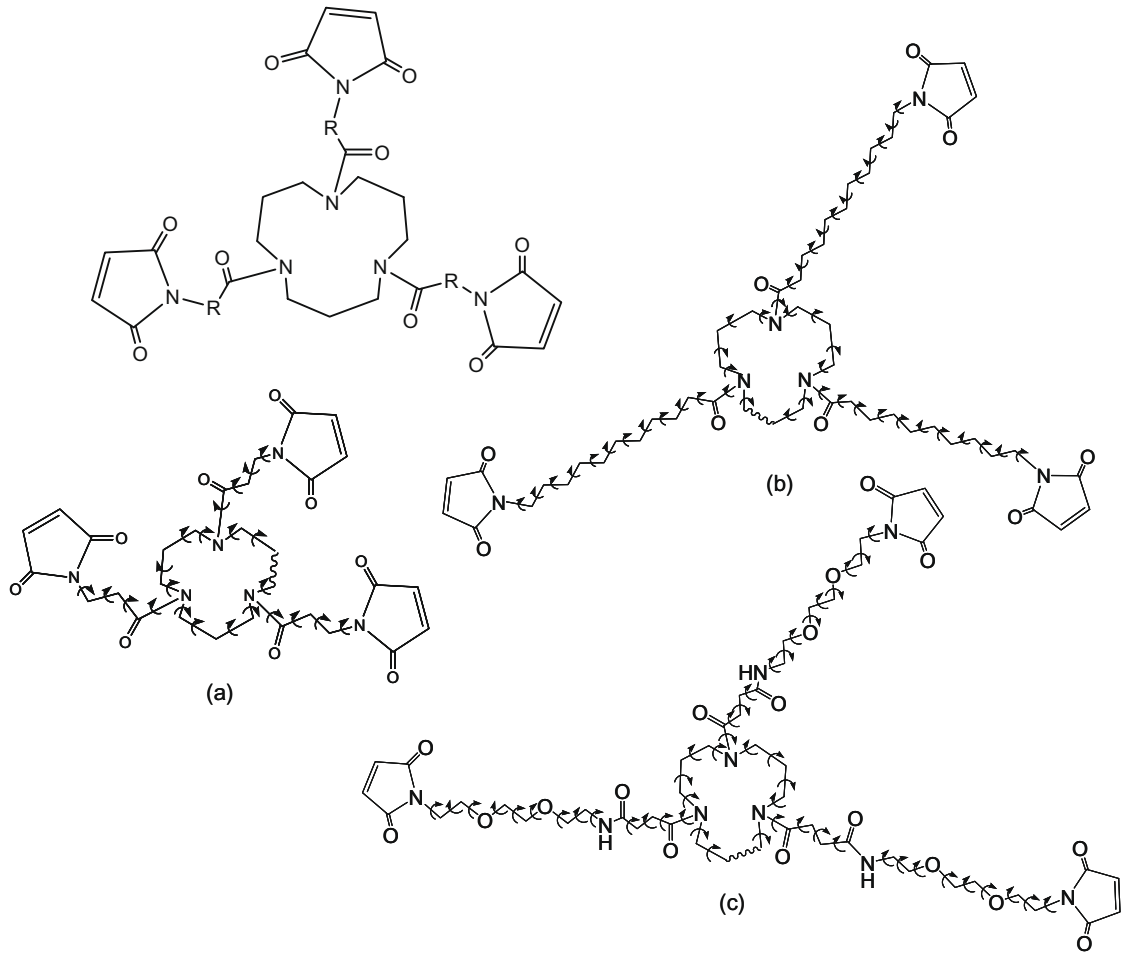

Figure 2. The drawing in the upper left corner illustrates the general structure of the synthetic cross-linking compounds. From this are generated molecules **1a** (R = $(CH_2)_2$), **1b** (R = $(CH_2)_{10}$), **1c** (R = $(CH_2)_2CONH(CH_2)_2(OCH_2CH_2)_2$). Also shown are the degrees of freedom varied during the conformational search of (a) **1a**, (b) **1b**, (c) **1c**. Varied torsions are represented by arrows; wavy bonds indicate ring openings.

(LM)[25,26] and Monte Carlo (MC)[27] searching was used to exhaustively characterize the potential energy surface of **1a–c**.

Conformational calculations were performed with MACROMODEL V9.1 software[28] on 3.2 GHz Athlons under the Red Hat 9 operating system. Classical force fields were initially evaluated for low quality torsions, and conformational searching was performed with the OPLS2005[29] force field in combination with the Generalized Born/Surface Area (GB/SA) continuum water solvation model.[30–35] The MC component of the hybrid LM:MC method varied a random number of torsion angles, from a minimum of two to the maximum number of allowed torsion angles specific to each compound (Fig. 2). Amide bonds in the substituent maleimide arms were constrained to the *trans* geometry. LM explored the 10 lowest eigenvectors for random steps between 3 and 6 Å from a determined minimum. The ring opening method of Still[36] was used for C–C bonds of the central triazacyclododecane systems to sample ring conformations. Conformational searches on **1a**-**1c** were performed in 5000 step blocks until convergence of minimum energy and number of conformations indicated exhaustive sampling of the potential energy surface. The entire ensemble from each block of 5000 steps was fed to the subsequent block of conformational search steps. The least sampled structure in each ensemble was used as the starting structure for subsequent blocks of searching, according to the usage-directed structure selection method.[27] The number of well converged structures found within a specified energy window above the minimum structure, the energy of the lowest energy structure, and the number of times the lowest energy structure was sampled were used to monitor the ability of the conformational search to sample the entire potential energy surface. Distinct conformations were determined by heavy atom (non-hydrogen) superimposition; structures with root mean squared deviation (RMSD) values greater than 0.25 Å were accepted and saved as unique conformations. All structures found during conformational searching were minimized by 500 steps of the Truncated Newton Conjugate Gradient (TNCG) method to within 0.01 kJ $Å^{-1}$ mol^{-1} gradient RMS convergence.[37] Further TNCG minimization was performed for **1b–c** to obtain fully minimized structural ensembles. Conformational analysis was performed only for ensembles consisting of minimized converged structures.

2.2. Clustering ensembles

The XCLUSTER program was used to determine geometric similarities of different conformations within each ensemble.[38] XCLUSTER compares pairwise distances between each structure in either Cartesian or torsional space and allows for the partitioning of structures into related groups. The process is agglomerative and hierarchical in nature, starting by defining each structure as a unique cluster, and combining clusters together until all structures are grouped into a single cluster. Significant clustering is judged according to the minimum separation ratio, which is defined as the ratio of the longest nearest-neighbor distance between structures in the same cluster to the smallest distance between structures in different clusters, at a particular clustering level. It has been shown that structurally significant clustering occurs for separation ratios larger than two at a high clustering level.[38] Atomic RMS clustering was performed following optimal superimposition of structures for heavy atom and ring atom comparison of the molecular ensembles of **1a–c**.

2.3. Quantum verification

Quantum calculations were performed with Jaguar V6.5.[39] Structures identified as minima for **1a–c** on the OPLS20005/GBSA (water) surface were subjected to unrestrained minimizations at the B3LYP/6-311G* level with the Self-Consistent Reactive Field (SCRF) continuum model[40] for water.

3. Results and discussion

Compounds **1a–c** were evaluated for low, medium and high quality torsions in the seven available force fields in the MACROMODEL software suite, and the OPLS2005 force field was chosen for conformational searching due to the absence of low quality torsions as well as to enable comparison with related conformational search studies.[22]

The relative flexibility of **1a–c** were evaluated using the LM:MC conformational search method on the OPLS05/GBSA (water) potential energy surface (Table 1; details of the force field analysis and conformational search results can be found in Supplementary data). Convergence was monitored by tracking the number of structures identified within 4.3 kcal/mol of the lowest energy structure, the energy of the lowest structure, and the number of times the lowest energy structure was sampled. The number of new structures found decreased significantly during each block of conformational searching, indicating that convergence, and thus the determination of a structurally representative molecular ensemble, had been achieved.

3.1. Analysis of the global minima

The global minimum structures for **1a–c** are presented in Figure 3. The lowest energy structures for all three triazacyclododecane derivatives on the OPLS2005/GBSA (water) potential energy surface have conserved crown-like ring conformations. Superimposition of the global minimum from each ensemble, based upon the 12 ring atoms, resulted in very low RMSD values (**1a**:**1b** = 0.0255 Å; **1a**:**1c** = 0.0538 Å; **1b**:**1c** = 0.0647 Å), indicating the energetic favorability of this crown-like ring structure. The ring nitrogen atoms are all planar due to delocalization with the carbonyl group adjacent to the ring and all maleimide arms point to the exterior of the ring. For all three structures, the orientation of the maleimide arms can be compared to cyclohexane axial and equatorial conformations; the low energy dodecyl ring conformation favors two 'pseudo-axial' arms that extend above the molecular plane for all three structures and one 'pseudo-equatorial' arm that extends outwards from the molecular framework. In **1a** the pseudo-equatorial arm extends downward away from the other substituents to form an asymmetric molecule, while in **1b** and **1c**, the pseudo-equatorial arms extend upward on the same face of the molecule as the pseudo-axial arms to form more overall symmetric structures. Despite the increased length of the maleimide arms in **1c** compared to **1b** (11 atoms per maleimide chain in **1b** as opposed to 13 atoms per chain in **1c**) **1c** forms a more compact structure, due to the presence of an intra-chain hydrogen bond between the oxygen of the maleimide group and the amide hydrogen. Quantum verification confirmed these molecular mechanics global minimum structures as minima on the B3LYP/6-31G** with SCRF (water) potential energy surface. All-atom RMSD values between quantum minima and the OPSL2005 structures were determined to be 1.369, 0.515, 5.768 Å, for **1a**, **1b**, and **1c**, respectively. The differences were greatest for **1c** due mainly to differences in the maleimide arm orientations between the optimized quantum and OPLS2005 structures. For instance, in

Table 1
A summary of the conformational search results for **1a–c**

Compound	Number of structures found within 4.3 kcal/mol of the lowest E structure	Number of blocks of 5000 steps completed
1a	2237	30
1b	3953	39
1c	4070	38

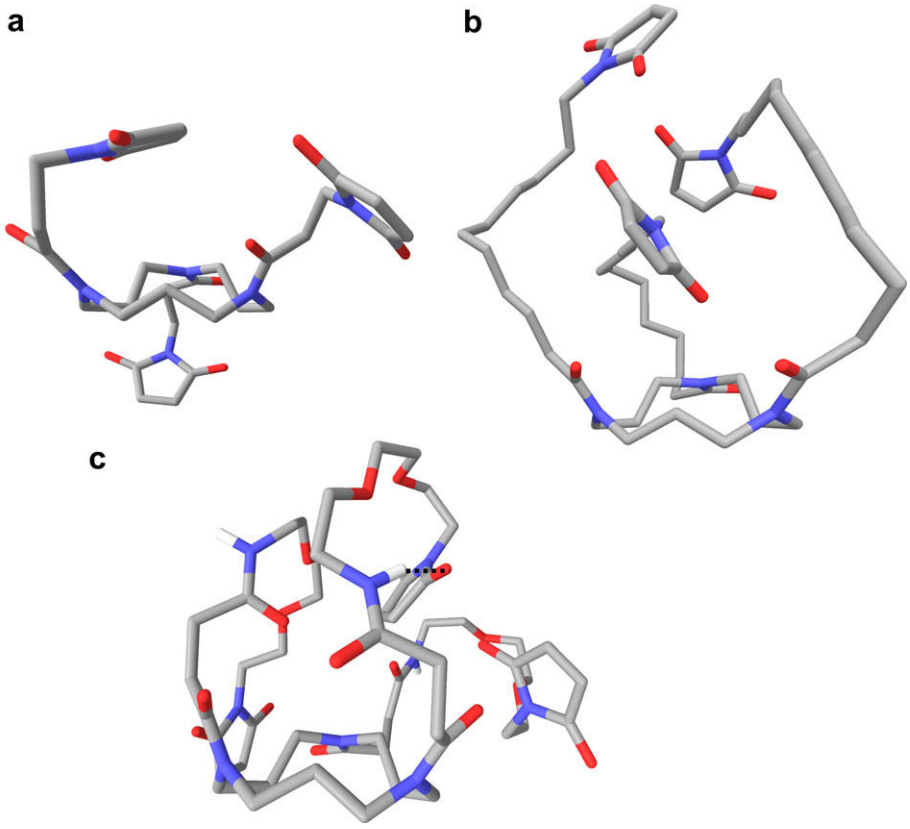

Figure 3. Lowest energy structures and hydrogen bonding patterns found on the OPLS2005/GBSA(water) surface of **1a–c**. Hydrogen bonding shown between the donor and acceptor atoms of **1c**.

the OPLS2005 structure all three maleimide arms of **1c** curve back around to form a relatively compact structure whereas the minimized quantum structure displays more extended maleimide arms that are not localized over the 12-membered central ring. Comparison based upon triazacyclododecane ring conformations showed even greater correlation between quantum and molecular mechanics results, with RMSD values of 0.052, 0.050, and 0.167 Å for **1a**, **1b**, and **1c**, respectively.

3.2. Analysis of the conformational ensembles

If every conformation is structurally distinct from all other conformations in a given ensemble, 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. The ensemble sizes are 2237, 3953 and 4070 structures for **1a**, **1b** and **1c**, respectively. The system with the shortest maleimide arms (**1a**) is the least flexible while the remaining two systems, one with a hydrophobic linker (**1b**) and the other with a slightly longer, C–O–C containing, hydrophilic linker (**1c**) have similar flexibility as reflected in the similar ensemble sizes.

However, it is possible that the minima are not structurally distinct, that is, 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. Clustering using the hierarchical, agglomerative clustering available in the XCLUSTER program provides an accurate evaluation of the flexibility of the system based upon the structural distinctness of each molecular conformation. The ensembles for **1a–c** did not exhibit any significant clustering by heavy atom or dihedral angle comparison, suggesting that each conformation is either structurally very similar to, or distinct from, all other conformations. A visual examination of these ensembles confirms that the conformations are distinct; the 12-membered ring is highly conserved in each ensemble but the maleimide arms sweep out the available conformational space. These results indicate that the ensembles cannot be grouped into distinct geometric families and that molecular flexibility can be evaluated based upon the number of unique structures determined within the same low energy window (4.3 kcal/mol, Table 1) above the global minimum structure. Thus the flexibility of all three systems is significant; with **1a** being relatively less flexible, while **1b** and **1c** are seen to be more flexible, most likely due to the increased length of the maleimide arms.

Clustering was also performed for the twelve ring atoms in order to evaluate ring flexibility. For **1a** this produced a separation ratio of 1.7, less than the 2.0 recommended by Shenkin and McDonald.[38] However, the ensemble was qualitatively analyzed at this clustering level, and 10 clusters were found corresponding to distinct ring conformations. The representative ring structures for the three most populated cluster groups are shown in Figure 4.

The three most populated ring conformations in the ensemble of **1a** have distinct characteristics. The first conformation, (a), can be compared to a cyclohexane chair conformation, with a nitrogen at the apex of one end, and a carbon atom extending upwards toward the ring at the base of the structure. Structures (b) and (c) appear as more elongated, crown like structures, differing primarily in the orientation of one carbon atom—either upwards and toward the ring as in Figure 4c or extending below the ring on the opposite side of the apical nitrogen, as seen for structure (b) in Figure 4. None of the representative structures from each of the 10 clusters adopted a ring conformation in agreement with the gas-phase ECP structures of 1,5,9-triazacyclododecane published previously by Ribeiro-Claro et al.[23] This suggests that the maleimide arms in

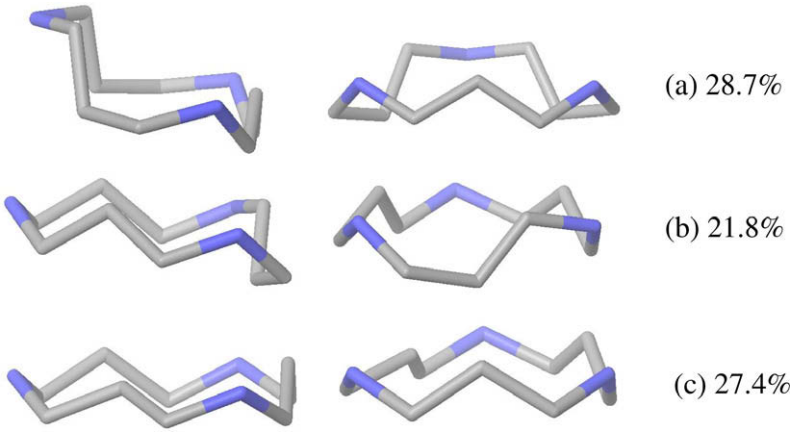

Figure 4. Representative ring structures from the three most populated clusters of **1a** based upon ring atom comparison (two views are shown; the leftmost view is rotated 90° relative to the rightmost view). Percentages indicate the importance of that structure in the overall ensemble. For clarity, ring substituents are not shown.

1a–c and/or solvent effects are playing a significant role in the overall conformational behavior.

Clustering based upon the 12 ring atoms resulted in a large (greater than 2.0) minimum separation ratio at high clustering level and was thus determined to be significant for **1b**, though 99.9% of structures were grouped within one cluster. The ring conformation corresponding to this cluster was similar to structure (a) in the **1a** ensemble (Fig. 4), with an RMSD of 0.0554Å. Clustering for **1c** was not significant for the 12 ring atoms, though 99.9% of structures also fell into a cluster with a similar ring conformation to (a) in the **1a** ensemble, with a 0.0485 Å RMSD between the ring atoms for the representative structures. Clustering based upon ring atoms seems to indicate that **1a** has increased flexibility with respect to ring conformation, while **1b** and **1c** have fairly rigid ring structures but increased flexibility of the maleimide arms.

3.3. Effect of ring substituents

In order to compare our methodology to the previous conformational results obtained for [12]-ane N3, we performed 125,000 steps of Low Mode Monte Carlo (LM:MC) conformational sampling on the *unsubstituted* 1,5,9-triazacyclododecane ring using the OPLS2005 force field in both vacuum and GBSA (water). We identified 8 unique structures within 1 kcal/mol of the global minimum including the 2 structures found previously by Riberio-Clara and co-workers.[23] Comparing the aqueous-phase ensembles of the substituted- (**1a**) and unsubstituted-triazacyclododecane ring using a larger energy window (3 kcal/mol) reveals more conformational flexibility in the propylenediamine bridges when the ring is unsubstituted. This is evidenced by a lack of significant ring atom clustering among the 72 structures found within 3 kcal/mol of the unsubstituted ensemble, whereas there is good clustering into 3 families for the 641 structures of **1a** (vide supra).

3.4. Hydrogen bonding in 1c

Hydrogen bonding was not possible for structures **1a** and **1b** given the lack of hydrogen donors, but the presence of the additional amide groups in **1c** allowed for hydrogen bonding in this structure (Fig. 5). It is useful to query hydrogen bonding patterns within each ensemble to yield insight regarding molecular and conformational behavior. Therefore, the number of classical hydrogen bonds (maximum distance between donor and acceptor = 2.5 Å, minimum donor angle = 120°, minimum acceptor angle = 90°) was determined for the ensemble of structures for **1c**. The lowest energy structure was found to contain only one hydrogen bond between the amide hydrogen and maleimide oxygen on the same substituent arm (Fig. 3). Other conformations within the ensemble contained either one (59.72%) or two (16.80%) hydrogen bonds and 23.45% of the ensemble did not contain any hydrogen bonds. Only one conformation (0.02%) was found to have three hydrogen bonds, and this conformer lies quite high in energy, indicating that this arrangement is not enthalpically favorable. As shown in Figure 5, the likelihood of forming 0, 1 or 2 hydrogen bonds was found in conformers distributed throughout the ensemble, i.e. low energy conformers were as likely to form 0, 1 or 2 hydrogen bonds as were high energy structures.

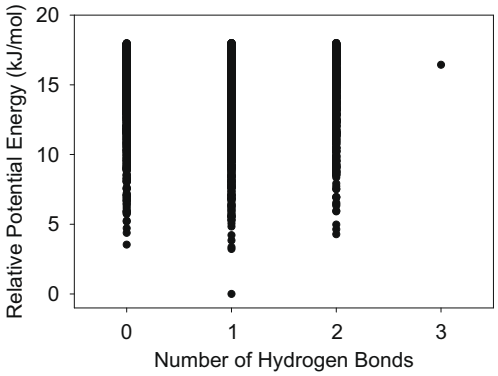

Figure 5. Hydrogen bonding and energetic analysis for ensemble **1c**.

3.5. Analysis of molecular lengths

Previous modeling studies suggest that the gp120 target is threefold symmetric and that the distance between any two of the CD4 binding pockets is 30–60 Å.[6,16] In order to further elucidate ensemble behavior and perhaps provide insight into the relationship between scaffold behavior and in vivo inhibition data already available, the longest intramolecular distance of each conformer within each ensemble of **1a–c** was measured (Fig. 6). The longest intramolecular distance for the ensemble of **1a** ranged between 10.6 and 18.9 (average 13.9) Å and between 13.4 and 20.2 (average 15.6) Å and 13.5–18.7 (average 15.4) Å for **1b**, and **1c**, respectively. A visual examination of the ensembles indicates that in many cases the longest intramolecular distance corresponds to the distance between two maleimide moieties. The CD4M9 miniprotein, to which these scaffolds were ligated via the maleimide arms, is a C-terminal cysteine-modified CD-4 mimetic. In the folded CD4M9-Cys structure

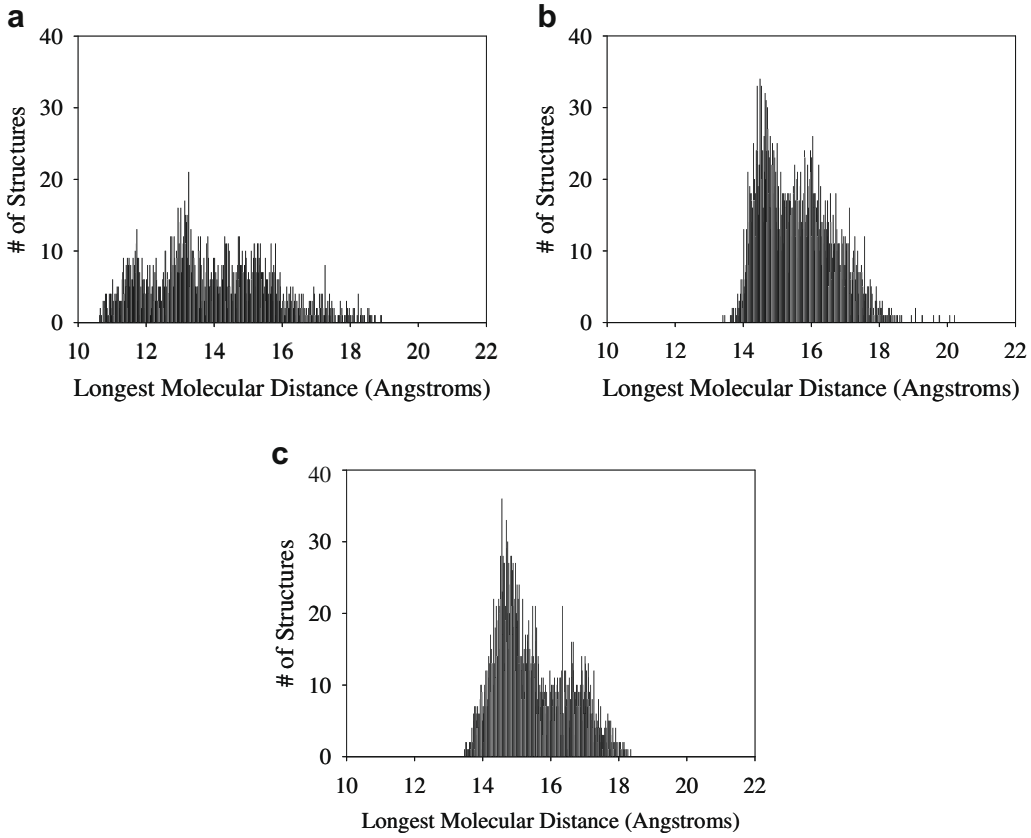

Figure 6. Longest molecular distances in each ensemble: (a) **1a**; (b) **1b**; (c) **1c**.

the distance between the cysteine and Phe residue, which binds to the gp120 Phe43 cavity, has been estimated to be 10 Å.[14] In our study, the conformational flexibility of the scaffolds causes the angles between the maleimide arms to sweep out almost all of the available space preventing a simple analysis or prediction of the possible distances between miniproteins ligated to these flexible scaffolds. However, an estimate may be established using the ensemble-averaged intramolecular distances above plus a 20 Å estimate of the distance between each C-terminal Cys and the Phe residue in the CD4M9 miniprotein. This would yield average summative distances of 33.9, 35.6 and 35.4 for **1a**, **1b** and **1c**, respectively, falling within the 30–60 Å estimated distances between any two Phe43 binding cavities, but not significantly distinguishing any of the inhibitors using length criteria. Based upon the average longest distance, the relative length ordering is **1a** < **1b** ~ **1c**. Compound **1b** was found to be on average slightly longer than **1c**, most likely due to the presence of hydrogen bonds in the ensemble for **1c**, contributing to a more compact structure. As shown in Figure 6, **1a** was found to have the largest range of available molecular distances, perhaps relating to the increased flexibility of the ring as determined by clustering (vide supra).

Biological inhibition data of trimeric scaffolds complexed with the gp120 CD4M9 miniprotein inhibitor have been recently determined (Table 2).[15] Scaffolds included the trimeric maleimide substituted 1,5,9-triazacyclododecanes of this study, along with C3 symmetric scaffold derivatives of Kemp's triacid, *cis*,*cis*-cyclohexane-1,3,5-tricarboxylic acid and trimesic acid, the structures of which are shown in Figure 7.[22]

In their first work on bivalent CD-4 mimetic ligands, Wang and co-workers linked two CD4M9-Cys miniproteins together using spacers of varying length and determined the HIV-1 inhibitory activity. They estimated the maximal distance between the Phe residues using a static, extended model and the known distance of 10 Å between Cys and Phe in the folded CD4M9-Cys miniprotein. In all cases the bivalent inhibitor was more active than the monovalent inhibitor and bivalent ligands with longer maximal lengths (64 and 52 Å) were 2.4–4.6 times more active than the linker with a maximal distance of only 38 Å.[16] In a subsequent study, Wang and co-workers[15] tethered CD4M9-Cys miniproteins to a variety of trivalent structures (Figs. 3 and 7 and Table 2). In this case, they found that most of the trivalent ligands showed the same anti-HIV activity regardless of linker length. Our molecular length data based on conformationally flexible models supports their conclusions regarding the molecular length requirements for effective inhibition. Scaffolds with different ensemble-averaged longest molecular distances displayed similar IC_{50} values, and scaffolds with the same average length data, like **1b** and **1c**, display very different inhibitory properties, indicating that the overall length of the molecular scaffold is not the determining factor in efficacy of viral inhibition. This agreement between our findings is especially significant, given that the molecular length data used by Wang and coworkers was based upon an estimation of the maximum length of each individual substituent chain assuming a zigzag C–X–C bond, and not an estimate of average ensemble behavior.

3.6. Preliminary speculations on the conformational behavior of the scaffold–miniprotein complexes

An exhaustive, quantitative, conformational analysis of the trimeric scaffold-miniprotein inhibitor constructs, as designed by Wang and co-workers, is beyond the capabilities of current conformational sampling methodologies due to the large number of

important degrees of freedom that would need to be sampled. However, in order to obtain a qualitative estimate of the behavior of these inhibitors, we have performed a conformational analysis of molecules **4** and **5** (structures shown in the Supplementary data) constructed such that the three maleimide moieties are covalently linked (via an in-silico Michael addition) to the cysteine sulfur atom of a CG dipeptide (to yield **4**) and to the cysteine sulfur atom of F23, a scorpion-toxin mimic of CD-4, found in the crystal structure 1YYM.pdb (to yield **5**). We included the smaller system (**4**) for computational-tractability and comparison purposes. To our knowledge, a crystal structure for the miniprotein used previously was not available and therefore, in order to create **5**, a GPGG tetrapeptide was inserted between the 1YYM miniprotein and the cysteine in order to more fully mimic the structure evaluated in Ref. 15. We choose to use the 1YYM structure as it contains the same sequence as the previous CD4-mimetic miniprotein, except for the last four peptides on the N-terminus end. We believe that this substitution should not change appreciably the conformational behavior of such a large system.

The LM:MC and Monte Carlo (MC) methods were used to sample the conformational space of the scaffold—miniprotein constructs using the OPLS2005/GBSA(water) surface. We sampled the surface of **4** for 100,000 steps but time constraints limited our sampling of the higher dimensional surface of **5** for only 15,000 steps. In both cases the starting structure scaffold was in the crown orientation with two of the succinimide-dipeptide or miniprotein arms directed above the plane of the ring and one arm directed outwards from the ring in a pseudo-equatorial orientation. In the case of **5**, the initial structure contained the GPGGC polypeptide in the extended conformation and the F23 miniprotein in the crystal structure orientation.

After conformational sampling, the resulting ensemble for **4** contained structures mostly in the crown conformation, with one axial arm directed upwards and two equatorial-like arms directed outwards. There is considerable flexibility of the CG peptide arms; in all cases the ensemble contains compressed structures with significant intramolecular hydrogen bonding within and between arms as shown for the lowest energy structure in the Supplementary data. We find excellent clustering of the ensemble of **4** into **3** distinct families using ring atoms (separation ratio = 8.0) Ninety-nine percent of the structures fall into the classic crown conformational family, with less than 1% adopting alternative orientations as shown in Figure 4 above. This may be due to limitations of the sampling method given the complexity of the surface, as the starting structure was in the crown conformation and it was not possible to ensure a converged ensemble on a molecular system of this size.

For **5**, the resulting conformational ensemble contained exclusively crown scaffolds with miniproteins adopting conformations similar to the starting/crystal structure. In all cases, the GPGGC polypeptide folds into a loop-like structure pulling the miniproteins closer in towards the scaffold. There is a significant amount of intra and inter-arm hydrogen bonding resulting in generally compressed structures. Two of the arms typically adopt pseudo-equatorial conformations and orient their miniproteins below the plane of the scaffold while the third arm originates from a more axial substitution and extends above the plane of the ring. This later observation might yield important insight into the biological behavior of dimeric versus trimeric inhibitors, that is, it seems that the miniproteins are simply too large and the scaffold too small to allow three miniproteins to orient on the same side of the scaffold plane. In order to help overcome any sampling problems and to further test this observation, we also performed geometry optimizations from manually constructed alternative starting conformations with all three miniprotein arms oriented on the same side of the ring. In all cases, these structures minimized to conformations with only two arms on a side.

How meaningful are these results on the much larger systems? The size of these systems presents a significant challenge to any current conformational search algorithm. Our visualization of the sampled and final structures for **4** indicate that the 12-membered ring scaffold is energetically able to adopt the **3** conformations discussed above (Fig. 4) but that the dimensionality of the space is too large for efficient sampling of the entire surface. From the results on both **4** and **5**, we can conclude that the ring scaffold is energetically able to adopt a crown orientation even when substituted with very large peptide or mini-protein arms. And from the results on **5**, we can conclude that the methodological treatment employed was able to fold the extended GPGGC pentapeptide into physically reasonable struc-

Table 2
HIV inhibition and molecular length data for synthetic scaffold compounds[a]

Compound	Ensemble averaged longest intramolecular distance (Å)	IC_{50}[b] (μM)	
		HIV-1_{Bal} (relative activity)	HIV-1_{IIIb} (relative activity)
CD4M9	n/a	9.7 (1)	12 (1)
1a	13.9	1.5 (6.5)	3.0 (4)
1b	15.6	0.068 (142)	0.40 (30)
1c	15.4	1.3 (7.5)	2.5 (4.8)
2a[c]	12.7	0.79 (12)	nd[e]
2b[c]	14.4	1.8 (5.4)	nd[e]
2c[c]	15.7	1.0 (9.7)	nd[e]
3[c]	14.6	0.33 (29)	nd[e]

[a] Viral inhibition data obtained by Wang et al.[15]
[b] IC_{50} is the concentration required to reach 50% inhibition of production of new viral particles. Inhibition relative to single CD4M9 is presented in parentheses.
[c] Conformational search results presented in Ref. 22.
[e] Inhibition data were not determined.

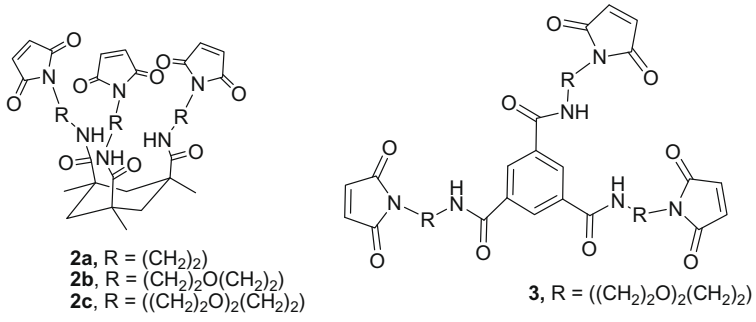

Figure 7. Trimeric scaffolds used previously with gp120 CD4M9 mimetics.[15,22]

tures and that it may not be possible to orient all three miniproteins on the same face of the scaffold ring.

Finally, though the molecular length does not appear to play a pivotal role in viral inhibition, the hydrophobic nature of compound **1b** and the central ring of scaffolds based on an aromatic core (**2c** in Table 2) are structural differences that could be important in interpreting their increased inhibition potential compared to the other inhibitory complexes.[15] A docking study using the multivalent ligands and models for the Phe-43 active site would be needed to further elaborate on this effect.

4. Conclusions

The Low Mode Monte Carlo conformational search results on the OPLS2005/GBSA (water) potential energy surface for three maleimide substituted 1,5,9-triazacyclododecane compounds are presented with respect to geometric structure, molecular length, and hydrogen bonding patterns. Compound **1a** appears to have more flexibility with respect to ring conformation and length, while **1b** and **1c** preferentially adopt a chair-like ring structure but have increased flexibility due to the length of the maleimide arms. The presence of hydrogen bonding in **1c** contributes to a more compact structure despite the increased substituent chain length. The molecular length of these compounds does not seem to relate directly to viral inhibition; rather, the hydrophobicity of **1b** might be a contributing factor to its efficacy of inhibition.

Acknowledgements

This project is an outgrowth of a fruitful collaboration with Inmaculada Robina from the Department of Organic Chemistry at the University of Seville and we are grateful for her helpful comments on the manuscript. The work was supported by awards from the National Science Foundation (C.P.—CHE–0211577, CHE–0116435 and CHE-0521063). S.R. acknowledges support from the Arnold and Mabel Beckman Foundation through receipt of a Beckman Scholars award. H.H. acknowledges receipt of a summer research fellowship from the Northwater Foundation. C.P acknowledges the Donors of the American Chemical Society Petroleum Research Fund and the Thomas F. 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. All calculations were carried out at the University of Richmond and computational resources were provided, in part, by the MERCURY supercomputer consortium under NSF Grants CHE-0116435 and CHE-0521063.

Supplementary data

Supplementary data (Force field parameter analysis, details of the conformational search results and low energy structures for **4** and **5**) associated with this article can be found, in the online version, at doi:10.1016/j.bmc.2008.12.027.

References and notes

1. Aids Epidemic Update: December 2006; HIV/AIDS, U. N. a. W. H. O. J. P. o. Ed., 2006.
2. Mansky, L. M. P.; Dennis, K.; Gajary, Lisa C. *J. Virol.* **2002**, *76*, 9253.
3. Richman, D. D. *Nature* **2001**, *410*, 995.
4. Shet, A.; Berry, L.; Mohri, H.; Mehandru, S.; Chung, C.; Kim, A.; Jean-Pierre, P.; Hogan, C.; Simon, V.; Boden, D.; Markowitz, M. *J. Acq. Immu. Def. Synd.* **2006**, *41*, 439.
5. McReynolds, K. D.; Gervay-Hague, J. *Chem. Rev.* **2007**, *107*, 1533.
6. Kwong, P. D.; Wyatt, R.; Sattentau, Q. J.; Sodroski, J.; Hendrickson, W. A. *J. Virol.* **2000**, *74*, 1961.
7. Wyatt, R.; Kwong, P. D.; Desjardins, E.; Sweet, R. W.; Robinson, J.; Hendrickson, W. A.; Sodroski, J. G. *Nature* **1998**, *398*, 705.
8. Weissenhorn, W.; Dessen, A.; Harrison, S. C.; Skehel, J. J.; Wiley, D. C. *Nature* **1997**, *387*, 426.
9. Chen, B.; Vogan, E. M.; Gong, H.; Skehel, J. J.; Wiley, D. C.; Harrison, S. C. *Nature* **2005**, *433*, 834.
10. Kwong, P. D.; Wyatt, R.; Robinson, J.; Sweet, R. W.; Sodroski, J.; Hendrickson, W. A. *Nature* **1998**, *393*, 648.
11. Tan, K.; Liu, J.-H.; Wang, J.-H.; Shen, S.; Lu, M. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 12303.
12. Caffrey, M.; Cai, M.; Kaufman, J.; Stahl, S. J.; Wingfield, P. T.; Covell, D. G.; Gronenborn, A. M.; Clore, G. M. *The EMBO J.* **1998**, *17*, 4572.
13. Chan, D. C.; Fass, D.; Berger, J. M.; Kim, P. S. *Cell* **1997**, *89*, 263.
14. Vita, C.; Drakapoulou, E.; Vizzavona, J.; Rochette, S.; Martin, L.; Menez, A.; Roumestand, C.; Yang, Y.-S.; Ylisastiguil, L.; Benjouad, A.; Gluckman, J. C. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 13091.
15. Li, 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.
16. Li, H.; Song, H.; Heredia, A.; Le, N.; Redfield, R.; Lewis, G. K.; Wang, L.-X. *Bioconjugate Chem.* **2004**, *15*, 783.
17. Eckert, D. M.; Kim, P. S. *Annu. Rev. Biochem.* **2001**, *70*, 777.
18. Gallo, S. A.; Puri, A.; Blumenthal, R. *Biochemistry* **2001**, *40*, 12231.
19. Melikyan, G. B.; Markosyan, R. M.; Hemmati, H.; Delmedico, M. K.; Lambert, D. M.; Cohen, F. S. *J. Cell Biol.* **2000**, *151*, 413.
20. LaBranche, C. C.; Galasso, G.; Moore, J. P.; Bolognesi, D. P.; Hirsch, M. S.; Hammer, S. M. *Antiviral Res.* **2001**, *50*, 95.
21. Schneider, S. E.; Bray, B. L.; Mader, C. J.; Friedrich, P. E.; Anderson, M. W.; Taylor, T. S.; Boshernitzan, N.; Niemi, T. E.; Fulcher, B. C.; Whight, S. R.; White, J. M.; Greene, R. J.; Stoltenberg, L. E.; Lichty, M. *J. Peptide Sci.* **2005**, *11*, 744.
22. Szczepanska, A.; Espartero, J. L.; Moreno-Vargas, A. J.; Carmona, A. T.; Robina, I.; Remmert, S.; Parish, C. *J. Org. Chem.* **2007**, *72*, 6776.
23. Ribeiro-Claro, P. J. A.; Amado, A. M.; Marques, M. P. M.; Teixeira-dias, J. J. C. *J. Chem. Soc., Perkin Trans. 2* **1996**, *6*, 1161.
24. Parish, C.; Lombardi, R.; Sinclair, K.; Smith, E.; Goldberg, A.; Rappleye, M.; Dure, M. *J. Mol. Graph. Model.* **2002**, *21*, 129.
25. Kolossvary, I.; Guida, W. C. *J. Am. Chem. Soc.* **1996**, *118*, 5011.
26. Kolossvary, I.; Guida, W. C. *J. Comput. Chem.* **1999**, *20*, 1671.
27. Chang, G.; Guida, W. C.; Still, W. C. *J. Am. Chem. Soc.* **1989**, *111*, 4379.
28. 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.
29. Jorgensen, W. L.; Tirado-Rives, J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6665.
30. Hasel, W.; Hendrickson, T. F.; Still, W. C. *Tetrahedron Comput. Method.* **1988**, *1*, 103.
31. Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. *J. Phys. Chem. A* **1997**, *101*, 3005.
32. Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127.
33. Weiser, J.; Shenkin, P. S.; Still, W. C. *J. Comput. Chem.* **1999**, *20*, 217.
34. Weiser, J.; Shenkin, P. S.; Still, W. C. *J. Comput. Chem.* **1999**, *20*, 586.
35. Weiser, J.; Weiser, A. A.; Shenkin, P. S.; Still, W. C. *J. Comput. Chem.* **1998**, *19*, 797.
36. Still, W. C.; Galynker, I. *Tetrahedron* **1981**, *37*, 3981.
37. Ponder, J. W.; Richards, F. M. *J. Comput. Chem.* **1987**, *8*, 1016.
38. Shenkin, P. S.; McDonald, D. Q. *J. Comput. Chem.* **1994**, *15*, 899.
39. 4.0, J. *Portland*, *OR* 1991–2000.
40. Klicic, J. J.; Friesner, R. A.; Liu, S.-Y.; Guida, W. C. *J. Phys. Chem. A* **1998**, *106*, 1327.