J|A|C|S
ARTICLES
Published on Web 06/28/2002

Fluoride-Selective Binding in a New Deep Cavity Calix[4]pyrrole: Experiment and Theory

Christopher J. Woods,† Salvatore Camiolo,† Mark E. Light,† Simon J. Coles,† Michael B. Hursthouse,† Michael A. King,‡ Philip A. Gale,*,† and Jonathan W. Essex*,†

Contribution from the Department of Chemistry, University of Southampton, Highfield, Southampton SO17 1BJ, U.K., and Celltech R&D Ltd., 208 Bath Road, Slough, Berks SL1 3WE, U.K.

Received January 11, 2002

Abstract: A new "super-extended cavity" tetraacetylcalix[4]pyrrole derivative was synthesized and characterized, and X-ray crystal structures of complexes bound to fluoride and acetonitrile were obtained. The binding behavior of this receptor was investigated by NMR titration, and the complex was found to exclusively bind fluoride ions in DMSO-d_6. This unusual binding behavior was investigated by Monte Carlo free energy perturbation simulations and Poisson calculations, and the ion specificity was seen to result from the favorable electrostatic interactions that the fluoride gains by sitting lower in the phenolic cavity of the receptor. The effect of water present in the DMSO on the calculated free energies of binding was also investigated. Owing to the use of a saturated ion solution, the effect of contaminating water is small in this case; however, it has the potential to be very significant at lower ion concentrations. Finally, the adaptive umbrella WHAM protocol was investigated and optimized for use in binding free energy calculations, and its efficiency was compared to that of the free energy perturbation calculations; adaptive umbrella WHAM was found to be approximately two times more efficient. In addition, structural evidence demonstrates that the protocol explores a wider conformational range than free energy perturbation and should therefore be the method of choice. This paper represents the first complete application of this methodology to "alchemical" changes.

Introduction

In the molecular recognition arena, the production of receptors that will only bind a single guest from a collection of putative guest species is an important goal. In this paper, the synthesis and characterization of a new, highly selective fluoride receptor is reported, and its binding properties are investigated using a range of new and traditional simulation techniques. This study consists of two parts. In the first, a novel calix[4]pyrrole derivative was synthesized and its binding behavior investigated in two solvents via NMR titration techniques. The molecule is unusual as it specifically binds fluoride ions and does not interact with chloride in DMSO-d_6 solution. This type of extended cavity receptor may therefore be potentially useful in anion-sensing[1,2] or new anion separation techniques.[3] The crystal structures of the fluoride-bound and acetonitrile-bound forms of the receptor were elucidated and used as a basis for the second part of this study, where the receptor's novel binding behavior was investigated by a range of computational techniques. The traditional methods of free energy perturbation (FEP) and the Poisson equation were applied to rationalize the observed specificity and to examine the effects of water in the solvent. In addition, the newer method of adaptive umbrella WHAM (AdUmWHAM) was optimized for use in binding free energy calculations, and its efficiency was compared with that of FEP.

AdUmWHAM is an extension of the umbrella sampling method.[4] Umbrella sampling is a well-established free energy technique that uses a biasing "umbrella" potential to enhance the sampling of a reaction coordinate, e.g., to enhance the sampling of a particular dihedral angle. The bias arising from using this restraining umbrella potential is then removed at the end of the simulation to return the unbiased probability distribution. However, the use of umbrella sampling is hampered by the difficulty of combining the results of umbrella simulations performed with different restraining potentials and by the problem of deriving an effective umbrella potential without prior knowledge of the underlying potential of mean force. The first of these problems was solved by the weighted histogram analysis method (WHAM),[5] which can both unbias the individual umbrella simulations and optimally weight and combine their results. Adaptive umbrella sampling solved the second of these problems by iterating the umbrella simulations to build up the

* Address correspondence to these authors. E-mail: J.W.Essex@soton.ac.uk, Philip.Gale@soton.ac.uk.
† University of Southampton.
‡ Celltech R&D Ltd.

(1) Miyaji, H.; Anzenbacher, P., Jr.; Sessler, J. L.; Bleasdale, E. R.; Gale, P. A. *Chem. Commun.* **1999**, 1723−1724.
(2) Gale, P. A.; Twyman, L. J.; Handlin, C. I.; Sessler, J. L. *Chem. Commun.* **1999**, 1851−1852.
(3) Sessler, J. L.; Gale, P. A.; Genge, J. W. *Chem. Eur. J.* **1998**, *4*, 1095−1099.
(4) Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.* **1977**, *23*, 187−199.
(5) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011−1021.

10.1021/ja025572t CCC: $22.00

restraining potential.[6] The results of each these iterations are unbiased and combined via WHAM and processed to build an umbrella potential that encourages even sampling of the reaction coordinate. In this way, the simulation *automatically* generates the optimum umbrella potential, and thus the simulation can be performed in parallel, with little user intervention or precalculation. Furthermore, the iterative nature of the simulation means that rough estimates of the free energy results are quickly available, allowing the possible application of the method as a free energy screening or scoring function. Moreover, since the umbrella potential is generated to sample the reaction coordinate evenly, the quality and efficiency of the conformational sampling could be better with an umbrella potential than if a conventional approach such as FEP or thermodynamic integration were used. In these conventional approaches, sampling is constrained to specific points, or windows, along the reaction coordinate. In this paper, these advantages of AdUmWHAM over FEP are demonstrated on the calix[4]pyrrole system. The results of this comparison show that AdUmWHAM is the method of choice for this system.

Part 1. Synthesis and Characterization

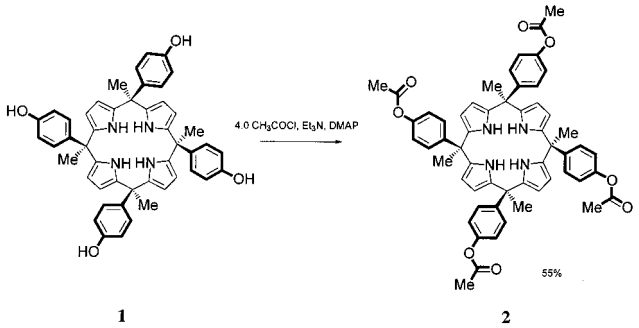
The discovery by Sessler and co-workers that calix[4]pyrroles are effective anion-binding agents in solution[7] has led to the synthesis of a variety of new calix[4]pyrrole derivatives that have been used in anion-binding,[8] sensing,[1,2] and new anion separation techniques.[3]

Recently, the synthesis of extended cavity calix[4]pyrroles derived from 4-hydroxyacetophenone has been reported by Sessler and co-workers[9] and Floriani and co-workers.[10] The so-called αααα isomer of this receptor[9] (in which all the phenol groups in the meso positions of the macrocycle are orientated on the same face of the receptor) may be obtained by crystallization or chromatographic techniques (e.g., compound **1**).[9,10] This compound may be elaborated through the addition of substituents to the phenolic hydroxyl groups, yielding "super-extended cavity" derivatives that possess the desired and thus far unique property among the calixpyrroles of being "*fluoride-only*" anion-binding agents in DMSO.[11]

The aim of this study is to investigate the binding properties of a "super-extended cavity" calix[4]pyrrole derivative with a less bulky substituent, to discover whether the unique fluoride specificity of these compounds is retained. To this end, compound **2** was synthesized by reaction of compound **1** with acetyl chloride in dry THF in the presence of triethylamine, with stirring for 5 days (Scheme 1). The tetraacetyl derivative **2** was isolated as a white crystalline solid by recrystallization from methanol in 55% yield.

The solution binding properties of **2** were investigated by using ^{1}H NMR titration techniques in DMSO-d_6 solution. From these studies it was found that addition of 100 equiv of chloride, bromide, iodide, dihydrogenphosphate, and hydrogensulfate anions (as tetrabutylammonium salts) to a deuterated DMSO solution of **2** caused no changes in the ^{1}H NMR spectrum of this compound. Therefore, it is concluded that this compound does not interact with these putative anionic guests in DMSO solution. However, upon addition of fluoride anions, new resonances (for NH, ArH, and pyrrolic CH) were observed in the ^{1}H NMR spectrum, corresponding to the formation of a fluoride complex with slow complexation/decomplexation kinetics relative to the NMR time scale (Figure 1). The association constant was determined by integration of the free and bound NH resonances at a number of different fluoride anion concentrations and found to be 73.9 ± 4.5 M^{-1}. This behavior has been observed in the bulkier extended cavity calixpyrroles[11] and demonstrates that this new molecule is also selective for fluoride ions.

Scheme 1. Preparation of the Tetraacetylcalix[4]pyrrole Derivative **2**



1 **2**

X-ray Crystallography. Crystals of compound **2** were grown from a DMSO solution of the receptor in an attempt to provide insight into the structures of the complexes adopted in solution. Unfortunately, despite numerous attempts, these crystals were found to diffract poorly, and the resulting structures were unsuitable for publication.

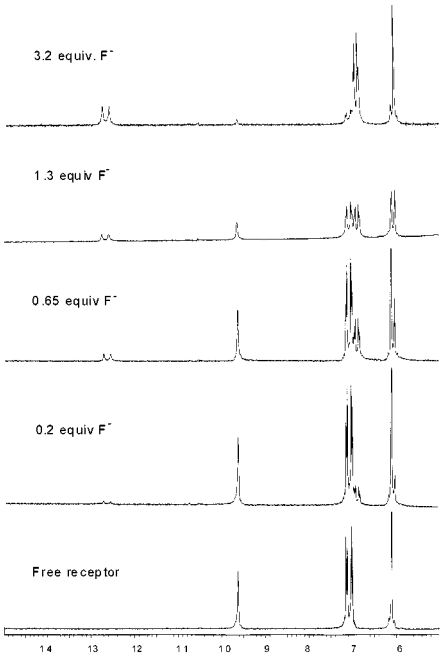


Figure 1. ^{1}H NMR spectra of compound **2** in DMSO-d_6 solution on addition of aliquots of tetrabutylammonium fluoride.

(6) Kumar, S.; Payne, P. W.; Vasquez, M. *J. Comput. Chem.* **1996**, *17*, 1269−1275.
(7) (a) Gale, P. A.; Sessler, J. L.; Král, V. *Chem. Commun.* **1998**, 1−8. (b) Gale, P. A.; Anzenbacher, P., Jr.; Sessler, J. L. *Coord. Chem. Rev.* **2001**, *222*, 57−122.
(8) Gale, P. A.; Sessler, J. L.; Allen, W. E.; Tvermoes, N. A.; Lynch, V. *Chem. Commun.* **1997**, 665−666.
(9) Anzenbacher, P.; Jursikova, K.; Lynch, V. M.; Gale, P. A.; Sessler, J. L. *J. Am. Chem. Soc.* **1999**, *121*, 11020−11021.
(10) Bonomo, L.; Solari, E.; Toraman, G.; Scopelliti, R.; Latronico, M.; Floriani, C. *Chem. Commun.* **1999**, 2413−2414.
(11) For an initial communication on the fluoride-only binding of this class of receptor in DMSO-d_6, see: Camiolo, S.; Gale, P. A. *Chem. Commun.* **2000**, 1129−1130.

Table 1. Crystal Parameters of the Two Calix[4]pyrrole Complexes

	complex **A**	complex **B**
empirical formula	$C_{72}H_{88}N_5O_8F$ $C_{56}H_{52}N_4O_8 \cdot C_{16}H_{36}N \cdot F$	$C_{62}H_{61}N_7O_8$ $C_{56}H_{52}N_4O_8 \cdot 3C_2H_3N$
formula weight	1170.47	1032.18
crystal system	tetragonal	tetragonal
space group	*P*4/*ncc*	*P*4/*n*
Z	4	2
unit cell dimensions (Å)	$a = 17.2489(5)$ $b = 17.2489(5)$ $c = 24.0303(8)$	$a = 11.3723(16)$ $b = 11.3723(16)$ $c = 21.020(4)$
volume (Å^3)	7149.6(4)	2718.4(8)
crystal	colorless needle	colorless plate
temperature (K)	150(2)	150(2)
goodness-of-fit on F^2	1.503	1.018
final *R* indices [$F^2 > 2\sigma(F^2)$]	$R1 = 0.1126$, w$R2 = 0.3421$	$R1 = 0.0518$, w$R2 = 0.1258$
R indices (all data)	$R1 = 0.1244$, w$R2 = 0.3518$	$R1 = 0.0990$, w$R2 = 0.1511$

X-ray quality crystals of the acetonitrile and fluoride complexes of compound **2** were obtained from acetonitrile solutions of the receptor in the absence and in the presence of tetrabutylammonium fluoride, respectively.

Cell dimensions and intensity data were recorded at 150 K, using a Bruker Nonius KappaCCD area detector diffractometer mounted at the window of a rotating molybdenum anode operating at 50 kV and 90 mA (λ(Mo Kα) = 0.71073 Å). With a crystal-to-detector distance of 30 mm, ϕ and Ω scans were carried out to fill the asymmetric unit. Data collection and processing were carried out using the programs COLLECT,[12] DENZO[13] and maXus,[14] and an empirical absorption correction was applied using SORTAV.[15,16]

The structures were solved via direct methods[17] and refined on F^2 by full-matrix least squares.[17] Non-hydrogen atoms were refined anisotropically, and hydrogen atoms were placed in idealized positions and refined using a riding model. In complex **A**, the tetrabutylammonium cation was disordered over two equally occupied conformations, and it was necessary to restrain the geometric and thermal parameters. Summary details for both structures are presented in Table 1. In both complexes, the calix[4]pyrrole adopts a cone conformation (the binding site encapsulated within the phenolic cavity), with the pyrrolic NH groups forming hydrogen bonds to the guest fluoride **A** and acetonitrile **B**. The fluoride is evidently more tightly bound than the acetonitrile, and this can be seen in the comparison of their local geometries. The distance between the centroid of the least-squares plane of the meso carbons (LSPmc) and the fluoride is 1.554 Å, compared to 2.417 Å to the acetonitrile nitrogen. The angle between the pyrrole ring and the LSPmc increases from 40.1° in the case of the fluoride to 46.8° for the acetonitrile, showing that the ring is more planar when binding the fluoride.

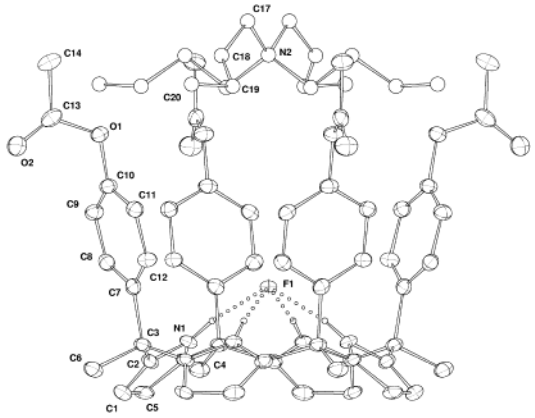


Figure 2. Molecular structure of complex **A** with thermal ellipsoids drawn at the 30% probability level. The tetrabutylammonium "cap" is disordered, and for clarity only one conformation is shown in a ball-and-stick representation. All nonacidic hydrogen atoms have been omitted for clarity. Hydrogen bond parameters: N1—H1···F1 2.732(6) Å, 170.0°.

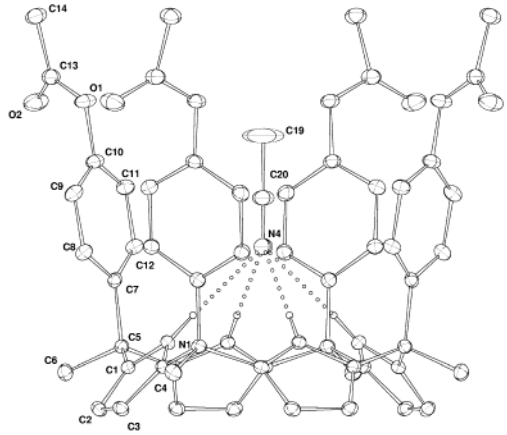


Figure 3. Molecular structure of complex **B** with thermal ellipsoids drawn at the 30% probability level. The two noncoordinated acetonitriles and nonacidic hydrogens have been omitted for clarity. Hydrogen bond parameters: N1—H1···N4 3.222(3) Å, 177.0°.

In addition, the donor—acceptor hydrogen bond distance increases dramatically from 2.732(6) to 3.222(3) Å, and the angle between the fluoride or acetonitrile nitrogen and two opposite H-bonding pyrrolic hydrogens decreases from 121.1 to 94.75°.

The fluoride structure forms stacks extending along the *c* direction composed of alternating tetrabutylammonium ions and calix[4]pyrrole molecules (Figure 2). Each stack has four nearest neighbors aligned in opposition and lying on a square grid. The acetonitrile structure packs in essentially the same manner but with two uncoordinated solvent molecules in the lattice cavities (Figure 3). The latter structure packs slightly more tightly, resulting in a higher density, 1.261 g cm^{-3}, compared to 1.087 g cm^{-3} for the fluoride complex.

It is possible to compare the fluoride structure with that of the chloride complex of *meso*-octamethylcalix[4]pyrrole.[18] The chloride, being a larger anion, sits farther from the LSPmc (2.417 Å) than the fluoride (1.554 Å) and thus has a more acute H—guest—H angle (90.8°) and a larger pyrrole LSPmc angle (44.9°).

(12) *Collect* data collection software, Nonius B.V., 1998.
(13) Otwinowski, Z.; Minor, W. Processing of X-ray Diffraction Data Collected in Oscillation Mode. In *Macromolecular Crystallography*; Carter, C. W., Jr., Sweet, R. M., Eds.; Methods in Enzymology 276; Academic Press: San Diego, 1997; part A, pp 307—326.
(14) Mackay, S.; Gilmore, C. J.; Edwards, C.; Tremayne, M.; Stuart, N.; Shankland, K. *maXus*: a computer program for the solution and refinement of crystal structures from diffraction data, University of Glasgow, Scotland, UK, Nonius B.V., Delft, The Netherlands, and MacScience Co. Ltd., Yokohama, Japan, 1998.
(15) Blessing, R. H. *Acta Crystallogr.* **1995**, *A51*, 33.
(16) Blessing, R. H. *J. Appl. Crystallogr.* **1997**, *30*, 421.
(17) Sheldrick, G. M. *SHELX97*: Programs for Crystal Structure Analysis (Release 97-2), Institüt für Anorganische Chemie der Universität, Tammanstrasse 4, D-3400 Göttingen, Germany, 1998.
(18) Gale, P. A.; Sessler, J. L.; Král, V.; Lynch, V. *J. Am. Chem. Soc.* **1996**, *118*, 5140—5141.

Part 2. Computer Simulations

The first part of this study used experimental methods to demonstrate that this novel calix[4]pyrrole derivative is selective for fluoride and binds the anion deep in the phenolic cavity. The aim of the second part of this study is to use both new and traditional simulation techniques to rationalize this behavior.

Previous authors have used both density functional theory (DFT)[19] and Monte Carlo (MC) free energy perturbation (FEP)[20] calculations to investigate the conformational and binding properties of simpler "parent" *meso*-octaalkylcalix[4]pyrrole macrocycles. These studies concluded that, while the 1,3-alternate conformation is preferred for the free molecule in solution, the cone conformation is preferred when the receptor binds to an anion (as was observed experimentally).[11] This is due to the stabilizing effect of forming four hydrogen bonds from the N−H groups of the pyrrole to the bound anion. Both studies concluded that the calix[4]pyrroles preferentially bind fluoride ions, with a predicted relative binding free energy[20] between the fluoride and chloride of about 17 kcal mol^{-1} or a predicted relative binding energy[19] of about 15 kcal mol^{-1}.

In a manner similar to that reported by van Hoorn and Jorgensen,[20] FEP simulations were initially performed in this study to identify the reasons for the fluoride specificity, and also to investigate any possible effect resulting from the small quantities of water present in the solvent. This is important, as other authors have demonstrated that fluoride binding to a calix-[4]arene derivative can be eliminated by hydration of the halide with two water molecules.[21] A modern free energy technique, adaptive umbrella WHAM (AdUmWHAM),[6,22] was then applied to this system. The aim of this part of the study was to apply the AdUmWHAM protocol to the calculation of free energy differences and to compare its efficiency and results to those of the well-established FEP technique.

Section 1: Free Energy Perturbation Simulations. FEP simulations calculate the difference in free energy between two states, A and B, connected through a coupling parameter λ. As λ varies from 0 to 1, the simulation parameters are smoothly converted from those of system A to those of system B. For noninteger values of λ, the system is a nonphysical hybrid of states A and B. The free energy change on moving from state A to state B, ΔA, is calculated using the Zwanzig equation:[23]

$$\Delta A(\mathrm{A} \rightarrow \mathrm{B}) = -k_{\mathrm{B}}T \ln\langle \exp[-(E_{\mathrm{B}} - E_{\mathrm{A}})/k_{\mathrm{B}}T]\rangle_{\mathrm{A}} \quad (1)$$

where E is the energy of the given state, T is the temperature, k_{B} is the Boltzmann constant, and the angle brackets indicate that the statistics are collected as an average over an ensemble of structures generated for the state A. For large free energy differences, the complete calculation is split into a set of smaller windows, such that the total free energy change is the sum of the individual forward and backward free energy differences associated with each window. The difference between the total forward and backward free energies may be used to estimate the precision of the results. A review of the FEP method may be found in ref 24.

The relative binding free energies of the fluoride, chloride, and bromide ions to calix[4]pyrrole **2** were calculated using a standard thermodynamic cycle.[20] This separates the calculation into a perturbation of the halide ion when it is bound to the calix[4]pyrrole and also when it is free in solution. The difference in calculated free energies for the bound and free-ion simulations corresponds to the relative binding free energy of the two ions.

Methods. The internal coordinate *Z*-matrix of the receptor complex was constructed using the crystal structure of complex **A**. To enhance sampling of the macrocyclic ring, the *Z*-matrix was built using the "fragment" approach developed by van Hoorn and Jorgensen.[20] Their OPLS[25] parameters for the macrocycle were also adopted, with generic OPLS parameters taken for the pendant arms. This structure was placed in a solvent box of either DMSO and H_2O, or CH_3CN and H_2O. Acetonitrile is a common solvent used in the study of anion binding to calix[4]pyrrole systems.[7] Small amounts of water were added to the solvents, based on an estimate of the experimental water content. The first solvent box had initial dimension $40.4 \times 43.4 \times 49.1$ Å^3 and consisted of 680 DMSO molecules[26] and 30 TIP4P waters.[27] The second box had initial dimension $42.5 \times 45.7 \times 51.6$ Å^3 and contained 968 CH_3CN molecules[28] and 30 TIP4P waters. This yields a similar concentration of water in the two solvent boxes. The water molecules were positioned randomly in both cases, and cubic periodic boundary conditions were applied. The simulations were conducted in the NPT ensemble, at a temperature of 25 °C and a pressure of 1 atm, with the nonbonded cutoffs set at 15 Å. The nonbonded interactions were calculated such that if any host atom was within the cutoff distance, then interactions with the entire host were calculated. The systems were then equilibrated for 3.6 million (M) MC steps. The reliability of this equilibration period was examined by monitoring the stability of the subsequent free energy calculations. A similar system and equilibration scheme were used for the free-ion leg of the calculation. The halide ion was placed in a similarly sized cubic box of solvent, containing either DMSO (640 molecules) and TIP4P H_2O (30 molecules), or CH_3CN (883 molecules) and TIP4P H_2O (30 molecules). The same cutoffs and parameters were used as in the bound simulations, and these systems were also equilibrated for 3.6M MC steps.

A closed cycle of mutations of fluoride ↔ chloride, chloride ↔ bromide, and bromide ↔ fluoride were performed to allow the closure for both the forward and backward free energies to be calculated. These values were used as a guide to the precision of the calculations.

The FEP simulations for both the bound and free-ion legs were subdivided into 11 windows spaced at intervals of 0.1 along the λ coordinate. Both forward and backward free energies were evaluated at each value of λ, allowing an additional estimate of the precision of the final free energy. Each window consisted of 600 thousand (K) steps of further equilibration,

(19) Wu, Y. D.; Wang, D. F.; Sessler, J. L. *J. Org. Chem.* **2001**, *66*, 3739−3746.
(20) van Hoorn, W. P.; Jorgensen, W. L. *J. Org. Chem.* **1999**, *64*, 7439−7444.
(21) McDonald, M. A.; Duffy, E. M.; Jorgensen, W. L. *J. Am. Chem. Soc.***1998**, *120*, 5104−5111.
(22) Bartels, C.; Karplus, M. *J. Comput. Chem.* **1997**, *18*, 1450−1462.
(23) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420−1426.
(24) (a) Reynolds, C. A.; King, P. M.; Richards W. G. *Mol. Phys.* **1992**, *76*, 251−275. (b) Kollman, P. A. *Chem. Rev.* **1993**, *93*, 2395−2417.
(25) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.
(26) Kaminski, G. A.; Jorgensen, W. L. *J. Phys. Chem. B* **1998**, *102*, 1787−1796.
(27) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(28) Jorgensen, W. L.; Briggs, J. M. *Mol. Phys.* **1988**, *63*, 547−558.

Table 2. Relative Binding Free Energies of Fluoride−Chloride−Bromide Bound to Calix[4]pyrrole Derivative, in Two Different Solvents

halide ions	free energy difference in DMSO:H_2O/ kcal mol^{-1}	free energy difference in CH_3CN:H_2O/ kcal mol^{-1}
fluoride−chloride	14.0 ± 0.5	19.2 ± 0.9
fluoride−bromide	15.7 ± 0.6	18.7 ± 0.9
chloride−bromide	0.96 ± 0.01	1.06 ± 0.01
closure	0.6 ± 0.8	1.6 ± 1.3

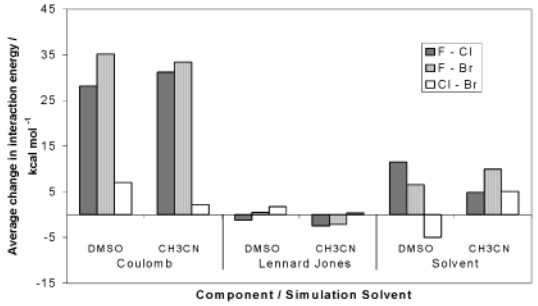


Figure 4. Average difference in interaction energy between the calix[4]-pyrrole and halide ions for the bound legs of the simulation, separated into electrostatic (Coulomb), nonelectrostatic (Lennard-Jones), and solvent components.

followed by 3M MC steps of data collection. In total, there were 39.6M MC steps in each leg of the simulation. Each MC move changed the position of either a solvent or a solute (1 attempted solute residue move to every 55 solvent moves). Motion of the solvent only took the form of translations and rotations, while both solutes had full translation, rotation, and bond/angle/torsion internal motion, where applicable. The only constraint applied was that the phenyl and pyrrole rings of the calix[4]pyrrole were kept planar.

The simulations were performed using a modified version of MCPRO 1.5.[29] These modifications were made to allow the AdUmWHAM calculations described in the next section to be performed and to enable DMSO solvent to be used. The code was checked to ensure that these modifications did not affect the results of the FEP simulations.

Results and Discussion. The results from the FEP calculations are shown in Table 2. The results show the average relative binding free energies, with the error bars corresponding to one standard error derived from the individual forward and backward simulations. The data support the experimental result that the calix[4]pyrrole is overwhelmingly selective for the fluoride ion. Also shown in Table 2 is the closure in the thermodynamic cycle resulting from adding the fluoride−chloride, chloride−bromide, and bromide−fluoride results. This result should of course be zero, and the extent to which the cycle closes gives an estimate of the precision of the calculated free energies. It is likely that the poor closure of the CH_3CN simulations is due to an underestimate in the free energy of the fluoride−bromide calculation since this is the largest mutation studied.

To rationalize these data, the average differences in interaction energies of the halide ions with the calix[4]pyrrole from the bound-leg simulations are presented in Figure 4. It is clear that the change in interaction energy from the fluoride to either the chloride or the bromide is positive and dominated by the electrostatic interactions. Moreover, the average differences in interaction energy between the halide ions and solvent molecules in the free-ion simulations (data not presented) show the same electrostatic stabilization of the fluoride ion. However, the magnitude is approximately half that reported in Figure 4. Thus, the observed anion specificity results from the preferential electrostatic stabilization of the fluoride anion by the calix[4]-pyrrole host.

Table 3. Comparison of the Simulation and Crystal Structures of the Calix[4]pyrrole Complexes

calix[4]pyrrole bound to	average pyrrole N–H guest distance[a]/ Å	average pyrrole NH–guest–HN angle[b]/ deg
fluoride (crystal)	1.86 (0.00)	121 (0.0)
fluoride (simulation)	1.88 (0.10)	115 (6.6)
acetonitrile (crystal)	2.34 (0.00)	94.8 (0.0)
chloride (simulation)	2.47 (0.11)	85.4 (7.2)
bromide (simulation)	2.64 (0.14)	79.3 (7.1)

[a] Average of pyrrole hydrogen−halide or pyrrole hydrogen−acetonitrile nitrogen distance over all four possibilities and, for the simulation, over all equilibrium structures. Standard deviation is given in parentheses. [b] Average of pyrrole hydrogen−halide−opposite pyrrole hydrogen angle over both possibilities and, for the simulation, over all equilibrium structures. For acetonitrile, the equivalent angle is calculated using the acetonitrile nitrogen atom. Standard deviation is given in parentheses.

Examination of the simulation structures shows that the fluoride prefers to sit lower in the complex than either the chloride or the bromide, probably as a result of the smaller size of the former. The angles and distances observed in the simulations were in reasonable agreement with the X-ray crystal structures (Table 3).

To examine the potential benefit arising from this binding geometry, a Poisson calculation of the electrostatics of the calix-[4]pyrrole was performed using the program UHBD.[30] The calculations used a structure from each of the fluoride and chloride simulations, from which the halide and all solvent molecules were removed. The calculations used a $65 \times 65 \times 65$, 0.3 Å spacing cubic grid, an interior dielectric constant of 1.0, and a solvent dielectric of 46.45 to represent DMSO. Coulomb-based boundary potentials were used, and the radii and charges adopted were the same OPLS parameters used in the FEP simulations. The potential on a planar slice through the center of the molecule is shown in Figure 5a with the fluoride ion bound. There is a small, positive, electrostatic pocket, into which the small fluoride ion fits. The bromide and chloride ions (Figure 5b) sit above this pocket and thus lose the favorable electrostatic interactions with the pyrrole ring. Thus, the selectivity of this calix[4]pyrrole is driven primarily by the ability of the small fluoride ion to fit into this positive pocket.

The FEP simulations demonstrate that changing the solvent from DMSO to CH_3CN has little effect on the predicted relative binding free energies. However, owing to the slow rate of molecular diffusion, the simulations did not completely investigate the effect of the additional water molecules. To examine further the sensitivity of the results to the presence of water, repeat FEP simulations of the DMSO:H_2O system were performed with some of the water molecules strategically positioned to bind to the halide ion. While this will give a good

(29) Jorgensen, W. L. *MCPRO 1.5*, Yale University, New Haven, CT, 1996.

(30) Davis, M. E.; Madura, J. D.; Luty, B. A.; McCammon, J. A. *Comput. Phys. Commun.* **1991**, *62*, 187−197.

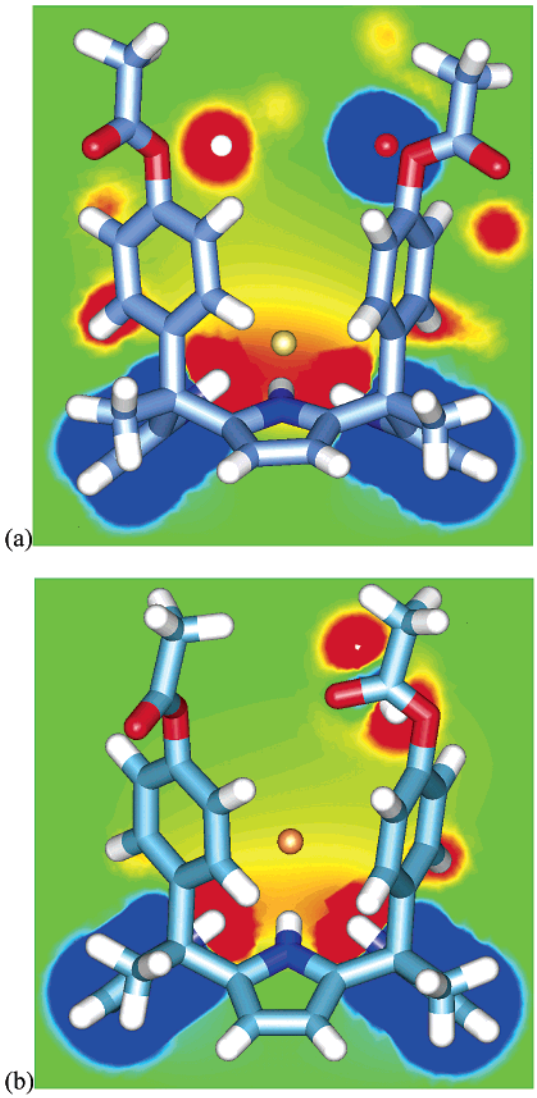


Figure 5. Electrostatic potential on a plane through the center of the calix[4]pyrrole solvated by implicit DMSO. Structure taken from (a) fluoride simulation (fluoride shown in bound location) and (b) chloride simulation (again, the halide is shown in its bound location). The colors range linearly from deep red ≥ 10 kcal mol^{-1} e^{-1}, through green $= 0$, to deep blue ≤ -10.0 kcal mol^{-1} e^{-1}.

idea of the effect of bound water, it must be remembered that this will not account for the relative entropies of the configurations. For the bound leg, a water molecule was placed directly above the halide ion in the complex. In the free-ion leg, one, two, or six water molecules were placed in a linear or octahedral arrangement around each ion. In all cases, the waters were oriented to hydrogen bond to the halide. The free energy results for the individual legs of these simulations are shown in Figure 6a, with the associated relative binding free energies in Figure 6b.

From Figure 6a, it is clear that the addition of extra solvating water molecules has little effect on the calculated free energies for the chloride−bromide mutation. However, the presence of water molecules around the fluoride ion preferentially stabilizes it with respect to either the chloride or the bromide, in both the free and bound simulations. The net effect of this stabilization is that, if a single water molecule is coordinated to the ion, the relative binding affinities are unaffected (Figure 6b). However, since it is impossible for the ion in the complex to be hydrated

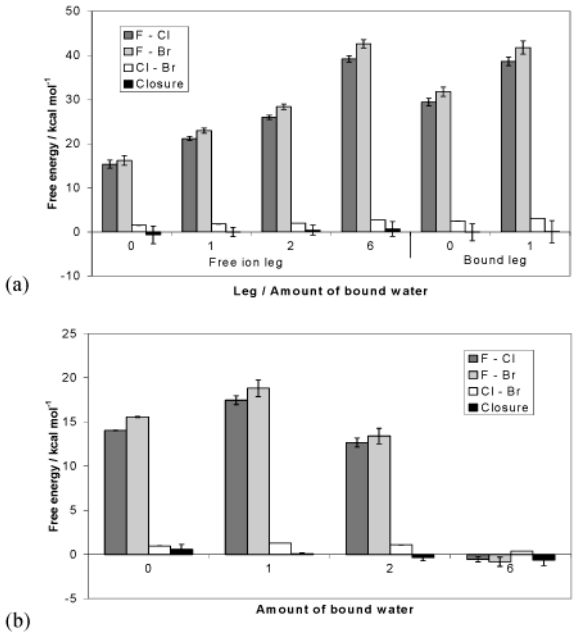


Figure 6. (a) Relative free energies of the halide ions for the bound and free-ion legs, in DMSO:H_2O solvent, as a function of the number of coordinating water molecules. (b) The relative binding free energies of halides bound to calix[4]pyrrole in DMSO:H_2O solvent, as a function of the number of coordinating water molecules.

by more than one water molecule, the addition of more water to the free fluoride ion has the effect of decreasing the fluoride ion's relative binding affinity, such that, at the point of complete hydration of the free ions, the calix[4]pyrrole is unselective. Fortunately, the experimental results were collected with a saturated ion concentration, corresponding to, at most, one water molecule per ion. However, these results do indicate that the presence of water in the organic solvent can have a profound effect on intermolecular association, particularly where ionic species are involved.

Section 2: Adaptive Umbrella WHAM Simulations. Adaptive umbrella WHAM simulations (an extension of general umbrella sampling[4] and λ dynamics[31]) are used to find the potential of mean force (PMF) along a defined reaction coordinate.[32] The λ coordinate from the FEP calculation may be viewed as a fictional reaction coordinate, and thus the calculation of the free energy difference is analogous to finding the PMF along λ.[31] Adaptive umbrella WHAM (AdUmWHAM) simulations use an iterative procedure to generate an umbrella potential that enables uniform sampling along a reaction coordinate.[6] The results of each iteration are combined with all previous iterations via the WHAM equations[5] to give an estimate of the probability density along the reaction coordinate, $p(\lambda)$. This can be converted into an estimate of the umbrella potential, $U(\lambda)$:

$$U(\lambda) = RT \ln p(\lambda) \qquad (2)$$

Since the converged umbrella potential is the negative of the PMF along λ, the difference in energy of the umbrella potential at $\lambda = 0$ and $\lambda = 1$ corresponds to the negative of the relative free energy.

(31) Kong, X.; Brooks, C. L. *J. Chem. Phys.* **1996**, *6*, 2414−2423.
(32) Roux, B. *Comput. Phys. Commun.* **1995**, 275−282.

The complete protocol of a typical AdUmWHAM calculation has been described elsewhere[33] and, briefly, involves the following steps:

1. Perform an unbiased simulation, making conventional Monte Carlo moves and moves along the reaction coordinate. This equilibrates the system.

2. Now perform the first-iteration simulation. Again make Monte Carlo moves along the reaction coordinate, and record the λ values sampled in a histogram.

3. Use this histogram to estimate the probability density, and hence calculate an estimate of the umbrella potential using eq 2. To improve convergence, the histogram is modified to remove all bins with no data. This is achieved by setting all empty bins to the same value as the bin with the lowest frequency. The resulting umbrella potential is then smoothed to remove discontinuities.[22]

4. Perform the next iteration's simulation, using this umbrella potential to bias sampling along the reaction coordinate, the result of which is again recorded in a histogram. The empty bins are modified as in step 3.

5. Apply the WHAM equations to unbias the histogram, and to combine it with all previous iterations' histograms, to return a new estimate of the probability density, from which a new estimate of the umbrella potential is made. If the simulation has not converged, then return to step 4, using this new umbrella potential to bias the simulation.

6. Once the umbrella has converged, it is equal to the negative of the potential of mean force along the reaction coordinate. Thus, if λ is the reaction coordinate, the relative free energy is merely the difference between the ends of the umbrella.

An approach similar to this has been used successfully in the calculation of potentials of mean force of dihedral angles[22] and peptide folding.[6] However, to our knowledge, the method has not been applied to a mutating reaction coordinate used to calculate relative binding free energies, nor has its efficiency been compared with that of traditional free-energy methods such as FEP. The closest methods reported combine FEP with traditional umbrella WHAM to form a hybrid method, where either WHAM is used to combine the statistics of each FEP window[34] or traditional umbrella WHAM is used to enhance sampling within each window.[35]

Methods. The AdUmWHAM calculations used starting coordinates identical to those used in the initial FEP simulations, with the water molecules positioned at random. The same modified MCPRO 1.5 code was used to perform the simulations. The WHAM calculations were performed in the manner described by Bartels and Karplus.[33] A program was written to generate the histograms and apply the WHAM equations. The AdUmWHAM calculations involved 250K steps of initial equilibration. This was followed by 200 WHAM iterations, each run in parallel over four processors. Each node started from the structure taken from the end of the last iteration, but with the λ value chosen at random. Each iteration involved 10K steps of further equilibration, followed by 60K steps of data collection. Once all nodes had completed sampling, their λ data were processed via the WHAM equations to yield the umbrella used in the next iteration. Thus, the entire calculation consisted of

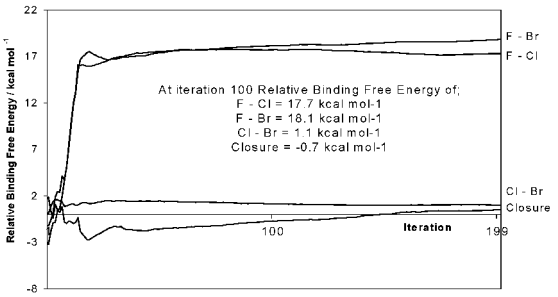


Figure 7. Predicted relative binding free energy of fluoride, chloride, and bromide to a calix[4]pyrrole derivative in DMSO:H_2O as a function of AdUmWHAM iteration. By iteration 100, the magnitude of the closure has fallen to well below 1 kcal mol^{-1}.

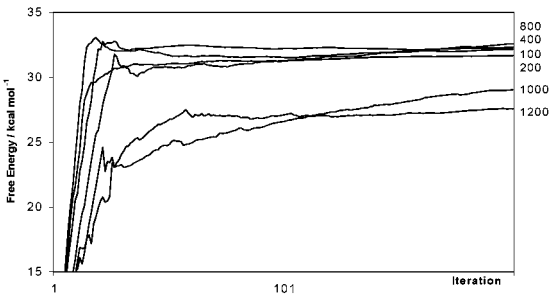


Figure 8. Difference between the ends of the umbrella as a function of iteration number for the F/Cl perturbation bound to calix[4]pyrrole in DMSO:H_2O, using the AdUmWHAM procedure, with a λ move every 100, 200, 400, 800, 1000, and 1200 MC moves.

200 $\times$ 4 simulations, a total of 56.25M MC steps. The simulations made random moves in λ with a maximum change of 0.08. These moves were made every 100 normal MC steps, corresponding to 600 $\times$ 4 moves in λ per WHAM iteration. These move sizes and rates were chosen to ensure rapid coverage of λ space while also allowing the system to relax between moves. The statistics along the λ coordinate were collected in a histogram with 103 equally spaced bins.

Results and Discussion. The results of the AdUmWHAM calculations for the relative binding free energies in DMSO are shown in Figure 7. The AdUmWHAM calculations have predicted fluoride−chloride and fluoride−bromide relative free energies that are approximately 3 kcal mol^{-1} larger than those predicted by FEP. The fact that the AdUmWHAM simulations show good closure suggests that this discrepancy is systematic. Calculations in pure DMSO (results not shown) demonstrated that this difference was not due to the random placement of the water molecules. Another possible reason for this difference could be that the system was not given sufficient time to relax between λ moves. To investigate this possibility, the bound leg, fluoride−chloride perturbation was repeated with λ moves at every 200, 400, 800, 1000, and 1200 MC steps. The difference between the ends of the umbrella as a function of iteration for each simulation is shown in Figure 8. It is clear from this figure that the AdUmWHAM calculations with λ moves every 100, 200, 400, and 800 steps have converged, and within error they arrive at the same result. The simulations with moves every 1000 and 1200 MC steps have not converged, since the sampling along λ has not yielded any high λ states (data not presented). This indicates that the difference between the ends of the umbrella is not large enough to surmount the $\sim$30 kcal mol^{-1} energy barrier between the fluoride and chloride. In contrast, the "100", "200", "400", and "800" simulations sample the entire

(33) Bartels, C.; Schaefer, M.; Karplus, M. *J. Chem. Phys.* **1999**, *111*, 8048−8067.
(34) Nina, M.; Beglov, D.; Roux, B. *J. Phys. Chem.* **1997**, *101*, 5239−5248.
(35) Souaille, M.; Roux, B. *Comput. Phys. Commun.* **2001**, 40−57.

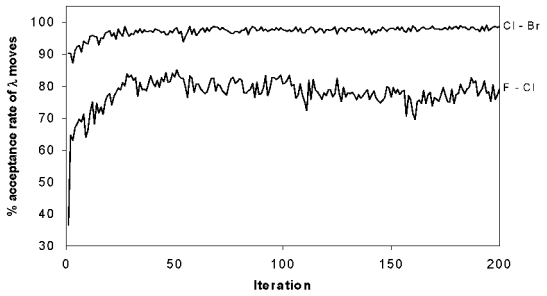


Figure 9. Percentage acceptance rate of λ moves as a function of AdUmWHAM iteration for the fluoride−chloride and chloride−bromide perturbations.

λ coordinate. Further evidence of the convergence of these simulations can be seen in a plot of the average acceptance rate of the λ moves, over the four nodes, versus iteration for the "400" simulation (Figure 9). The acceptance probability climbs as the umbrella evolves and then plateaus as the umbrella converges and encourages free motion along the λ coordinate. This stabilization of the acceptance probability could be used as a reliable indicator of the convergence of the umbrella potential.

Since both the AdUmWHAM and FEP calculations appear to have converged, another reason must be found for the difference in their predicted results. To examine the range of sampling in the simulations, the distance between various atoms in the complex was plotted versus λ. Coordinate sets were analyzed every 300K MC steps for the FEP simulations, corresponding to a total of 10 data points for each discrete value of λ. For the AdUmWHAM simulations, the coordinates at the end of each iteration were analyzed (corresponding to once every 60K steps). In Figure 10a, the distance between a pyrrole hydrogen and the halide ion is plotted. This is the length of the hydrogen bond, and as expected, it linearly increases as the halide is perturbed from a fluoride to a chloride. There is no significant difference in the hydrogen bond lengths observed using the two simulations protocols. In Figure 10b, the distance between two opposite pyrrole hydrogen atoms is presented, thereby indicating the range of motion of the pyrrole rings. It is clear that the AdUmWHAM simulations are showing a wider range of sampling than the FEP simulations at lower λ. The wider range of sampling of the AdUmWHAM calculation is seen to a greater extent in the distance between carbons C8 (Figure 2) on opposite phenyl rings (Figure 10c).

Thus, at first sight, it appears that the AdUmWHAM simulations are showing more sampling than FEP. However, it may be argued that the different sampling of the two methods is a statistical anomaly due to the fewer structural observations made during the FEP simulations. This is unlikely as the coordinates analyzed span the entire simulation trajectory. In addition, Figure 10c shows an interesting oscillation in the FEP sampling at lower λ. This suggests that two conformations were sampled, corresponding to distances of approximately 7 and 8 Å, that freely interconvert at high λ. At low λ, the FEP simulations choose one or the other of the conformations, and since there is only one trajectory within each window, the alternative conformation is not adopted. In contrast, the AdUmWHAM simulations randomize λ after each iteration, so that, in effect, multiple trajectories visit each λ value, and thus both conformations are sampled. An alternative explanation for the

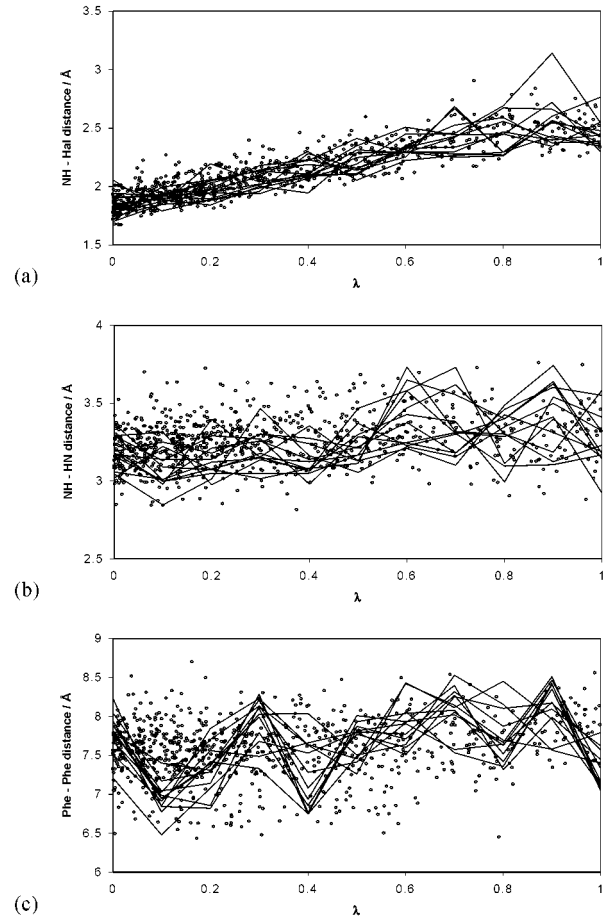


Figure 10. Distances between key atoms of a calix[4]pyrrole/halide complex from an FEP (lines) and an AdUmWHAM (circles) simulation, as a function of λ, for the fluoride−chloride perturbation. The distances are between (a) the pyrrole NH hydrogen and the halide, (b) two opposite pyrrole NH hydrogen atoms, and (c) carbons on two opposite phenol residues.

sampling differences is that, in the AdUmWHAM simulations, the system is not given sufficient time to relax after a new, random value of λ is chosen. However, if this were the case, then one would expect the simulations where the frequency of the λ moves was systematically varied to converge onto different free energies. This was not the case. In addition, the key structural difference, the length of the NH−Hal hydrogen bond, is sampled equally well by both AdUmWHAM and FEP.

These results, therefore, suggest that AdUmWHAM has a wider range of sampling than FEP, since FEP samples only one trajectory for each value of λ and can become "stuck" in a limited range of conformations during the finite simulation time. In contrast, AdUmWHAM samples the whole λ coordinate dynamically and in effect allows many trajectories to visit each value of λ, thereby sampling a wider range of conformations. Since the AdUmWHAM simulations sample a wider range of conformations with respect to λ, compared to FEP, a difference in their calculated free energies is to be expected.

These simulations have demonstrated that, on this system, AdUmWHAM has many advantages when compared to FEP. As well as the wider range of sampling across λ, AdUmWHAM has the additional benefits of speed and ease of use. The calculations are simple to set up and do not need key parameters, such as FEP window width, to be predetermined. In addition,

the calculations converge more rapidly than FEP, in this case in around half the CPU time (Figure 7). In addition, a rough estimate of the relative free energies is available within the first 15−20 iterations. This suggests that AdUmWHAM may be used as an approximate free energy scoring function. A benefit of this approach is that the results of interesting ligands could be refined by merely adding more iterations. This is a major advantage over FEP, where the results are highly dependent on the initial window selection, and an improvement in quality may require a reassignment of the windows and a rerunning of the entire calculation.

Conclusion

A novel calix[4]pyrrole derivative has been synthesized and characterized and has been shown to exclusively bind fluoride ions in DMSO solvent. Modeling studies have suggested that the lower position of the fluoride in the complex (as evidenced by the modeling and the crystal structures) allows it to maximize the strength of the hydrogen bonds with the pyrrolic NH groups. This was graphically illustrated by the existence of a small electrostatic "positive pocket" into which the small fluoride ion fits. The effect of water in the organic solvents was investigated through simulation, and it was seen that the selectivity of the calix[4]pyrrole for fluoride was removed in very high water concentrations. While this was not an effect in the experimental saturated ion solutions studied, it would become significant at low ion concentrations.

The modeling studies also compared the traditional FEP method with the newer AdUmWHAM method. The AdUmWHAM method converged in around half the CPU time compared to FEP, though onto slightly different results. Examination of this difference suggested that it was a result of the AdUmWHAM method encouraging a wider range of conformational sampling when compared to FEP. Given the greater efficiency and improved sampling observed in the AdUmWHAM simulations, this procedure is to be preferred over FEP, at least in systems of this type.

Synthesis

αααα-*meso*-Tetramethyl-*meso*-tetra-4-hydroxyphenylcalix[4]pyrrole (3 g, 4 mmol), acetyl chloride (1.27 g, 16 mmol), and triethylamine (3.27 g, 32 mmol) were dissolved in 300 mL of freshly distilled THF and stirred for 5 days, during which time a brown solid formed. The solvent was removed in vacuo, and MeOH (100 mL) was added. The brown suspension was heated to reflux for 5 min, and after cooling, white crystals of the product formed that were collected by filtration. The product was washed three times with 20 mL of MeOH and dried in vacuo (2.0 g, 55% yield). ^{1}H NMR (300 MHz DMSO-d_6): δ 1.90 (s), 2.33 (s), 6.11 (s), 7.02 (d, J = 5.9 Hz), 7.13 (d, J = 5.9 Hz), 9.63 (s). ^{13}C NMR (75 MHz DMSO-d_6): δ 20.61, 31.2, 104.94, 121.68, 127.92, 137.20, 147.79, 149.27, 169.23. MS (ES$^+$): [M + H$^+$] 909, [M + Na$^+$] 932. Elemental analysis calc for $C_{56}H_{52}N_4O_8 \cdot 3.5H_2O$: C, 69.19; H, 6.12; N, 5.76. Found: C, 69.18; H, 6.15; N, 5.88.

Acknowledgment. We thank the BBSRC, Celltech, the Royal Society, the University of Southampton, and the EPSRC for funding this work, and the EPSRC for provision of the X-ray crystallographic service. We thank Prof. W. L. Jorgensen for the provision of the MCPRO program and the associated source code, and Dr. J. Oyarzabal for helpful discussions. Additionally, J.W.E. and P.A.G. thank the Royal Society for University Research Fellowships.

Supporting Information Available: Crystallographic files (CIF); *Z*-matrix and example MCPRO parameter file (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA025572T