

Experimental and Theoretical Approach to Hydrogen-Bonded Diastereomeric Interactions in a Model Complex

Philippe H. Hünenberger,[†] Josef K. Granwehr,[†] Jean-Nicolas Aebischer,^{‡,§} Nagwa Ghoneim,[‡] Edwin Haselbach,[‡] and Wilfred F. van Gunsteren^{*,†}

Contribution from the *Laboratorium für Physikalische Chemie, ETH-Zentrum, CH-8092 Zürich, Switzerland*, and *Institut de Chimie Physique de l'Université, Péroilles, CH-1700 Fribourg, Switzerland*

Received February 17, 1997[®]. Revised Manuscript Received May 9, 1997

Abstract: Binding affinities of (*R,R*)-1,2-cyclohexanediamine (**R**) to (*R,R*)-1,2-cyclopentane-1,2-diol (**R₅**) and (*S,S*)-1,2-cyclopentane-1,2-diol (**S₅**) and to the corresponding cyclohexane-1,2-diols (**R₆** and **S₆**) have been measured in benzene and in CCl₄ at 298 K by microcalorimetry, and unexpected differences between the diastereomeric complexes are observed. Long time scale (0.1 μs) molecular-dynamics simulations of the two smaller diastereomeric complexes, **R/R₅** and **R/S₅**, in a simplified solvent model are reported. A direct free energy calculation gives results in good agreement with the experimental values measured in benzene for the first pair, but nearly identical results for the second pair, which is at variance with experiment. A systematic analysis of the dependence of simulation results on model parameters is performed, and no possibility is found to improve the enantioselectivity by parameter tuning. Other possible causes for discrepancies are specific solute–solvent or solvent–solvent interactions, electronic charge redistribution effects, or formation of clusters of more than two molecules. Owing to the long time scales reached, a well-converged picture of the dynamics is obtained, and the species present at equilibrium can be studied in detail. The average lifetime of the complex is found to be about 200 ps, whereas that of a hydrogen bond is only about 5 ps. Besides the unbound state, the dominant species observed in the simulations for both diastereomeric pairs are singly hydrogen-bonded complexes, with a clear preference for a O to N over the N to O hydrogen bond. Many other hydrogen-bonding patterns (bridged, double) are also observed in minor amounts.

Introduction

Diastereomeric interactions are of general interest in both biochemistry and organic chemistry.^{1–3} In biochemistry they determine enantioselectivity of substrate binding to proteins, e.g., enzymes,^{4–6} antibodies^{7–9} or sensorial receptors,^{10–14} and in organic chemistry their action in the ground state or in the transition state leading to products is at the origin of the enantioselectivities observed in organic or enzyme catalyzed reactions.^{3,15–17} Due to the high directionality of the hydrogen bond,^{18,19} hydrogen-bonded diastereomeric interactions often

play a key role in both domains.^{20–23} Their study by experimental and theoretical^{24,25} techniques is therefore of great interest.

One approach to studying these interactions is to measure by NMR titration or calorimetric techniques²⁶ the binding affinities of diastereomeric complexes of small organic molecules. A variety of such complexes have been studied as model compounds by microcalorimetric methods.²⁷ The measured thermodynamic parameters for binding sometimes display dramatic differences for the two diastereomeric pairs. For example, the binding affinities of (*R,R*)-1,2-cyclohexanediamine (**R**) to (*R,R*)-1,2-cyclopentane-1,2-diol (**R₅**) and to (*S,S*)-1,2-cyclopentane-1,2-diol (**S₅**), measured in benzene, show unexpected differences (Figure 1). **S₅** binds to **R** with an enthalpy change (ΔH_b) about 20 kJ/mol lower and an entropy change (ΔS_b) about 80 J/(K·mol) lower than its enantiomer. The cyclohexane-1,2-diols **R₆** and **S₆** (Figure 1) behave similarly. When the solvent is changed from benzene to CCl₄, the enthalpy and entropy changes upon binding are somewhat reduced in magnitude, as well as the differences in these quantities for the diastereomeric pairs. The general behavior of these systems is, however,

[†] Laboratorium für Physikalische Chemie.

[‡] Institut de Chimie Physique de l'Université.

[§] Present address: Ecole d'Ingénieurs de Fribourg.

[®] Abstract published in *Advance ACS Abstracts*, July 15, 1997.

(1) Craig, D. P.; Mellor, D. P. *Top. Curr. Chem.* **1976**, *63*, 1–70.

(2) Etter, M. C. *J. Phys. Chem.* **1991**, *95*, 4601–4610.

(3) Webb, T. H.; Wilcox, C. S. *Chem. Soc. Rev.* **1993**, *22*, 383–395.

(4) Ortiz, C.; Tellier, C.; Williams, H.; Stolowich, N. J.; Scott, A. I. *Biochemistry* **1991**, *30*, 10026–10034.

(5) Schirmeister, T.; Otto, H.-H. *J. Org. Chem.* **1993**, *58*, 4819–4822.

(6) Marx-Tibbon, S.; Katz, E.; Willner, I. *J. Am. Chem. Soc.* **1995**, *117*, 9925–9926.

(7) Benkirane, N.; Friede, M.; Guichard, G.; Briand, J.-P.; van Regenmortel, M. H. V.; Müller, S. *J. Biol. Chem.* **1993**, *268*, 26279–26285.

(8) Imaizumi, M.; Sisido, M. *Chem. Lett.* **1994**, *1*, 51–52.

(9) Wulff, G. *Angew. Chem., Int. Ed. Engl.* **1995**, *34*, 1812–1832.

(10) Schallenberger, R. S.; Acree, T. E.; Lee, C. Y. *Nature* **1969**, *221*, 555–556.

(11) Russell, G. F.; Hills, J. I. *Science* **1971**, *172*, 1043–1044.

(12) Friedman, L.; Miller, J. G. *Science* **1971**, *172*, 1044–1046.

(13) Chedd, G. The search for sweetness. *New Sci.* **1974**, *62*, 299–302.

(14) Walters, E. *J. Chem. Educ.* **1995**, *72*, 1680–1683.

(15) Boland, W.; Frössl, C.; Lorenz, M. *Synthesis* **1991**, *12*, 1049–1072.

(16) Famulok, M.; Jeong, K.-S.; Deslongchamps, G.; Rebek, J., Jr. *Angew. Chem., Int. Ed. Engl.* **1991**, *30*, 858–860.

(17) Curran, D. P.; Qi, H. *Helv. Chim. Acta* **1996**, *79*, 21–30.

(18) Taylor, R.; Kennard, O.; Versichel, W. *Acta Crystallogr. B* **1984**, *40*, 280–288.

(19) Scheiner, S. *Acc. Chem. Res.* **1994**, *27*, 402–408.

(20) Jeong, K. S.; Tjivikua, T.; Muehldorf, A.; Deslongchamps, G.; Famulok, M.; Rebek, J., Jr. *J. Am. Chem. Soc.* **1991**, *113*, 201–209.

(21) Adrian, J. C.; Wilcox, C. S. *J. Am. Chem. Soc.* **1992**, *114*, 1398–1403.

(22) Connelly, P. R.; Aldape, R. A.; Bruzzese, F. J.; Chambers, S. P.; Fitzgibbon, M. J.; Fleming, M. A.; Itoh, S.; Livingston, D. J.; Navia, M. A.; Thomson, J. A.; Wilson, K. P. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 1964–1968.

(23) Usher, K. C.; Remington, S. J.; Martin, D. P.; Drueckhammer, D. G. *J. Am. Chem. Soc.* **1994**, *33*, 7753–7759.

(24) Wade, R. C.; Goodford, P. J. *J. Med. Chem.* **1993**, *36*, 148–156.

(25) Smith, D. A. *Modeling the hydrogen bond*; Am. Chem. Soc. Symp. Ser.; American Chemical Society: Washington, 1994.

(26) Connors, K. A. *Binding constants*; Wiley, New York, 1987.

(27) Haselbach, E.; Ghoneim, N.; Aebischer, J.-N. *Progress reports, CHiral2-Workshops of the Swiss National Science Foundation*, 1994 and 1995.

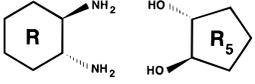
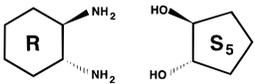
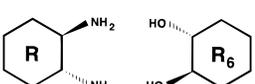
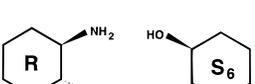
			in Benzene	in CCl ₄
	K_b [1/M]		40	97
	ΔG_b [kJ/mol]		-9.3	-11.5
	ΔH_b [kJ/mol]		-20.2	-17.6
	ΔS_b [J/(K·mol)]		-36.4	-20.3
	K_b [1/M]		15	53
	ΔG_b [kJ/mol]		-6.8	-10.0
	ΔH_b [kJ/mol]		-41.7	-34.3
	ΔS_b [J/(K·mol)]		-116.3	-81.2
	K_b [1/M]		19	53
	ΔG_b [kJ/mol]		-7.4	-10.0
	ΔH_b [kJ/mol]		-21.8	-15.9
	ΔS_b [J/(K·mol)]		-47.9	-19.8
	K_b [1/M]		14	38
	ΔG_b [kJ/mol]		-6.6	-9.1
	ΔH_b [kJ/mol]		-40.6	-36.0
	ΔS_b [J/(K·mol)]		-113.2	-89.6

Figure 1. R/R₅, R/S₅, R/R₆, and R/S₆ complexes together with the experimental results for the thermodynamic parameters in benzene and in CCl₄ at 298 K. K_b is the equilibrium constant for binding. ΔG_b is the binding (Gibbs) free energy. ΔH_b is the binding enthalpy. ΔS_b is the entropy change upon binding.

qualitatively the same as in benzene. The essential point is that R/S coupling between the components with respect to their R/R coupling turns out to be energetically about *twice* as favorable, but entropically less favorable (by a factor of 3–4 depending on the solvent). The observation of such differences, despite the apparent simplicity of the systems, is appealing for a theoretical rationalization.

In the present work, diastereomeric affinities of R and the smaller diols R₅ and S₅ are studied by molecular dynamics simulations in a simplified solvent. Although simulation in explicit benzene solvent is in principle possible, it is computationally very expensive. A significant equilibration time would be required, and the solvent molecule would be about the same size as the solute molecules. The use of an all-atom CCl₄ model would be computationally more tractable, but may still be an unnecessary complication. Hydrogen bonding affinities for neutral monofunctional compounds in CCl₄ solution seem, as in the gas phase, to be essentially an intrinsic property of the donor and acceptor present,^{28,29} suggesting the absence of specific solute–solvent interactions in this solvent. On the other hand, the results presented in Figure 1 indicate that the solvent has a quantitative effect on the studied equilibria, but does not affect the qualitative trends observed. It may therefore seem reasonable to assume that the causes of the binding enantioselectivity are essentially steric. Thus, the choice was made to use the simplest explicit apolar solvent model available: a (neutral) one site model^{30,31} for CCl₄. This was considered to be a reasonable compromise between vacuum and its distortive effects on the one hand and an explicit all-atom treatment of benzene or CCl₄ on the other.

The main assumptions underlying the model used for the simulations are as follows:

A. The solvent can be considered as an apolar medium of non-zero viscosity but of relative permittivity 1 with no specific solute–solvent interaction other than van der Waals interaction.

B. Two isolated molecules are representative for the overall equilibrium state, i.e., no larger clusters occur.

C. Simulation of a periodic system with a given solute concentration in the box can be used to mimic a bulk solution of the same concentration.

D. Hydrogen bonding and other interactions can be adequately modeled by a sum of pairwise point charge interactions and van der Waals repulsions.

The simplified solvent model used here allows long time scale simulations (up to 0.1 μ s) to be performed. Sufficient statistics can be gathered to obtain a detailed and well-converged picture of the equilibrium properties and the dynamics of the system. Species present at equilibrium can be studied exhaustively, and the free energy of binding can be estimated by direct counting. The latter method was chosen instead of using of a biasing potential energy function (e.g., umbrella sampling) for the following reasons: (i) the relatively low computational expenses of this simulation allow one to reach sufficiently long time scales for direct counting to be applied, (ii) since the biasing potential energy function would restrict the space to be sampled, a particular function may leave out relevant parts of space from the sampling, (iii) when the bound state is not exactly known, an appropriate biasing function may not easily be designed, and (iv) the biasing function would distort the dynamics of the system.

Experimental Methods

Apparatus. The microcalorimetric measurements were performed at 298 K with the LKB 2277 “thermal activity monitor (TAM)” operated in the flow mode. The two solutions containing the diamine and the diol, respectively, were injected into the reaction chamber with HPLC pumps LKB 2150. Their optimal internal working pressure of 2 bar was built up by using needle valves which were attached at the exit nozzle. For the solutions to reach thermal equilibrium before entering the reaction chamber, the total flow rate of the pumps was kept below 0.6 mL/min. The neutralization reaction of sulfuric acid with sodium hydroxide was employed for the calibration procedure.^{32,33} More details about measurements of this kind can be found elsewhere.^{34–37}

Chemicals. (R,R)- and (S,S)-1,2-cyclopentandiol, R₅ and S₅, respectively, were dried under vacuum. (R,R)- and (S,S)-1,2-cyclohexandiol, R₆ and S₆, respectively, and (R,R)-1,2-cyclohexanediamine (R) were sublimed. These five compounds allow the study of the following diastereomeric complexes: R/R₅, R/S₅, R/R₆, and R/S₆ (Figure 1). For the complexes involving the cyclohexanediols, the enantiomeric diamine (S) was also employed. This provided an internal consistency check, given that the results for R/R₆ and S/S₆, as well as for R/S₆ and S/R₆, must be equal. Benzene (Fluka puriss.) was stored over molecular sieves and used without further purification. CCl₄ (Fluka puriss.) was washed with 2 M NaOH, dried with MgSO₄, distilled under argon, and stored over molecular sieves. The distilled water needed for the calibrations was freed from CO₂ by boiling for 0.5 h and purging with N₂ for 1 h. All solutions were stored under N₂ and degassed before use to prevent valve malfunction in the pumps. Solvents were changed stepwise by using intermediate mixtures to prevent pressure variations in the pumps which resulted in irregular flows. The water content of the solutions was controlled by Karl Fischer titration; it did not exceed 50 ppm.

Data Treatment. The present work follows the so-called “entropy titration method”.^{26,34} The concentration of the reactants was kept below 10⁻² M in order to approach ideal conditions where the thermodynamic

(28) Abraham, M. H.; Grellier, P. L.; Prior, D. V.; Taft, R. W.; Morris, J. J.; Taylor, P. J.; Laurence, C.; Berthelot, M.; Doherty, R. M.; Kamlet, M. J.; Abboud, J.-L. M.; Sraidi, K.; Guiheneuf, G. *J. Am. Chem. Soc.* **1988**, *110*, 8534–8536.

(29) Marco, J.; Orza, J. M.; Notario, R.; Abboud, J.-L. M. *J. Am. Chem. Soc.* **1994**, *116*, 8841–8842.

(30) Rebertus, D. W.; Berne, B. J.; Chandler, D. *J. Chem. Phys.* **1979**, *70*, 3395–3400.

(31) Lautz, J.; Kessler, H.; van Gunsteren, W. F.; Weber, H.-P.; Wenger, R. M. *Biopolymers* **1990**, *29*, 1669–1687.

(32) Vanderezee, C. E.; Swanson, J. A. *J. Phys. Chem.* **1963**, *67*, 2608–2612.

(33) Bertrand, G. L.; Millero, F. J.; Wu, C.; Hepler, L. G. *J. Phys. Chem.* **1966**, *70*, 699–705.

(34) Christensen, J. J.; Izatt, R. M.; Hansen, L. D.; Partridge, J. A. *J. Phys. Chem.* **1966**, *70*, 2003–2010.

(35) Schaer, J.-J.; Milos, M.; Cox, J. A. *FEBS Lett.* **1985**, *190*, 77–80.

(36) Milos, M.; Schaer, J.-J.; Comte, M.; Cox, J. A. *Biochemistry* **1986**, *25*, 6279–6287.

(37) Aebischer, J.-N. Ph.D. Thesis, University of Fribourg/CH, 1992.

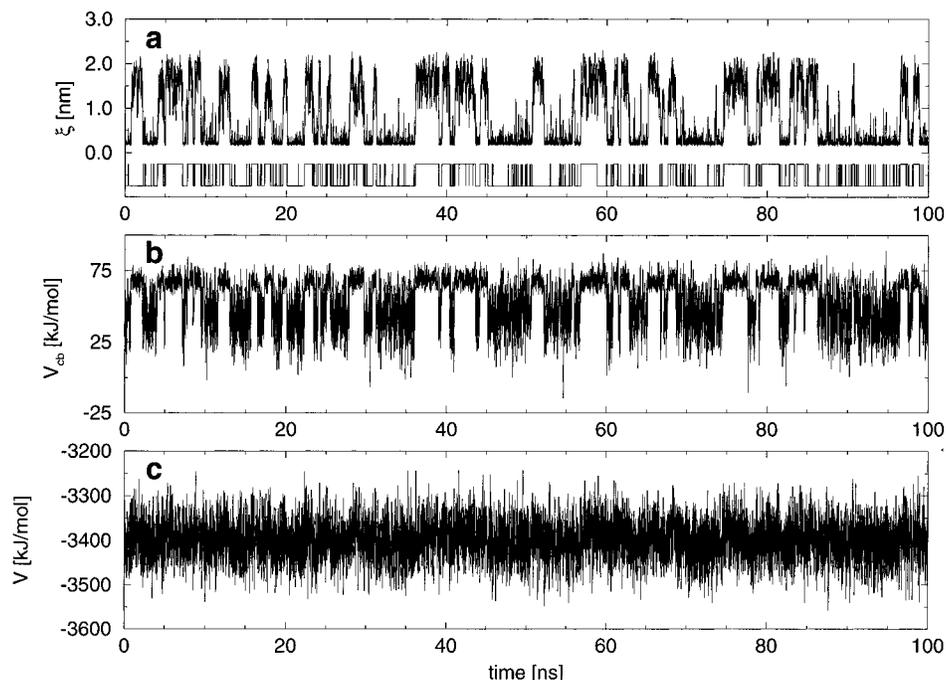


Figure 2. Equatorial R/R_5 case (simulation A): time evolution of (a) the reaction coordinate $\xi(t)$, (b) the Coulombic energy $V_{cb}(t)$, and (c) the total potential energy $V(t)$ of the system. The bottom of part a shows the bound/free status of the system (arbitrary units) obtained by using an allowed excursion time (t_{ex}) of 2.0 ps for modeling the evolution of $\xi(t)$. The reaction coordinate $\xi(t)$ is defined as the minimal H(O) to N distance.

quantities do not require corrections due to non-ideality. In line with this expectation, the heats of dilution obtained in this concentration range turned out to be negligible for both solvents used. This observation furthermore suggests that no particular solute–solvent interactions or solute self-associations must be taken into account in the elucidation of the association constants. The raw data were fitted on the basis of a 1:1 association model. Job plots²⁶ indicate that equimolar amounts of the components associate to a $n:n$ complex. However, $n = 1$ could not be rigorously established.

The fitting problem turned out to be difficult due to the low equilibrium concentration of the complexes. To increase the significance of the results, different fitting algorithms were employed. Equally, to reduce the probability of arriving at a false minimum, several fits were executed with different values for K_b and ΔH_b to initiate the iterations. These were chosen within a very broad (though chemically still reasonable) range. The computer fits were backed by visual inspection of the data with use of Drago-plots.²⁶

Molecular Model and Computational Procedure. All simulations were performed by using the GROMOS87 force field^{38,39} and a modified version of the GROMOS87 simulation program adapted for fast computation with a neutral solvent. Four configurations of the solute were used as starting points for the simulations: two involving R with the two vicinal amino groups in equatorial position, and R_5 (simulation A) or S_5 (simulation B), and two involving R with the two amino groups in axial position, and R_5 (simulation C) or S_5 (simulation D). The conformation of the six-membered ring is determined by 3-fold degenerate C–C–C–C torsional dihedral angle potential energy terms (multiplicity 3), which allow chair–chair interconversions.⁴⁰ The equilibrium conformation of the amino groups is equatorial–equatorial. To enforce an axial–axial conformation, the C–C–C–C dihedral angle potential energy term controlling the orientation of the amino groups was changed to a non-degenerate sinusoidal function (multiplicity 1), and the corresponding force constant increased from 5.86 to 20.0 kJ/mol.

The simulations for each of the four cases were set up as follows. Coordinates were generated for the complex and energy minimized.

(38) van Gunsteren, W. F.; Berendsen, H. J. C. *Groningen Molecular Simulation (GROMOS) Library Manual*; Biomos: Nijenborgh 4, 9747 AG Groningen, The Netherlands, 1987.

(39) Scott, W. R. P.; van Gunsteren, W. F. In *Methods in computational chemistry: METECC-95*; Clementi, E., Corongiu, G., Eds.; STEF: Cagliari, 1995; pp 397–434.

(40) Wilson, M. A.; Chandler, D. *Chem. Phys.* **1990**, *149*, 11–20.

The solute was then immersed in a truncated octahedron box corresponding to a cube of edge length 4.33 nm containing 252 united atom CCl_4 solvent “molecules”. These are neutral Lennard-Jones particles with a repulsion parameter $C_{12}^{1/2} = 0.07545$ [kJ/(mol·nm¹²)]^{1/2} and a dispersion parameter $C_6^{1/2} = 0.5155$ [kJ/(mol·nm⁶)]^{1/2} (ref 30). The density of the system was 1595 kg/m³, very close to the experimental value for the pure solvent. The system was relaxed for 1 ns, and 100 (simulations A and B) or 50 ns (simulations C and D) production runs were performed.

For all simulations, the bond lengths were constrained by application of the SHAKE procedure⁴¹ with a relative tolerance of 10^{-4} . Non-bonded interactions were handled by means of a twin-range method,⁴² the nonbonded pair list being updated every step. Taking advantage of the neutrality of the solvent, the long-range cutoff radius was set to 1.875 nm, the maximal possible cutoff radius within the simulation box. To speed up the calculation and reach longer time scales, the short-range cutoff radius was set to 0.8 nm, which is slightly more than three times the “radius” of the solvent “molecule” ($1/2\sigma_{LJ} = 0.264$ nm), but shorter than the value used previously for this model.^{30,31} A time step of 2 fs was chosen for integrating the equations of motion. The temperature was maintained at 298 K by weakly coupling the system to an external temperature bath⁴³ with a coupling constant $\tau_T = 0.1$ ps. The simulations were performed at constant volume. Center of mass motion was removed every nanosecond. The solute coordinates were saved every 0.05 ps for analysis. The computation time on a Silicon Graphics Power-Challenge XL computer, using two processors in parallel, amounted to 2.5 h for 1 ns.

Results

Population Analysis. A reaction coordinate ξ was defined as the minimal distance between a hydrogen attached to any of the two oxygens of the diol and the closest nitrogen (nearest image). The evolution of this reaction coordinate as a function of time for simulation A is displayed in Figure 2a, together with the evolution of the (solutes) Coulombic interaction energy

(41) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327–341.

(42) van Gunsteren, W. F.; Berendsen, H. J. C. *Angew. Chem., Int. Ed. Engl.* **1990**, *29*, 992–1023.

(43) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

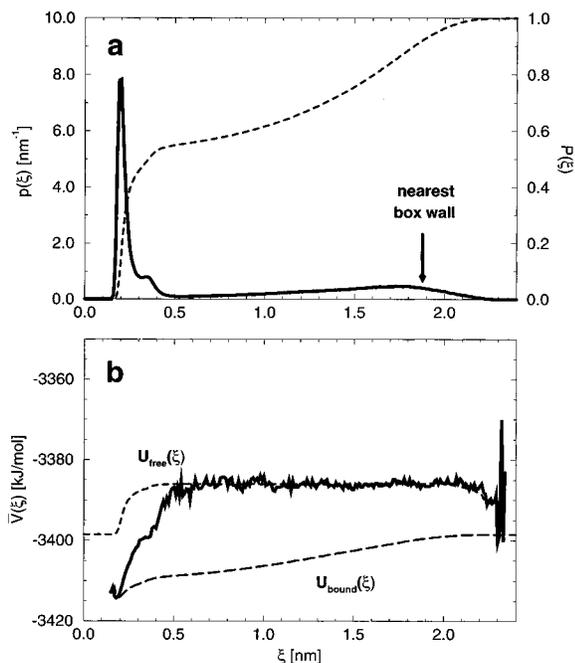


Figure 3. Equatorial **R/R_s** case (simulation A): (a) probability density distribution $p(\xi)$ (—) from eq 1, together with its integral $P(\xi)$ (---), and (b) average total potential energy profile $\bar{V}(\xi)$ (—), together with the population weighted integrals $U_{\text{bound}}(\xi)$ (---) and $U_{\text{free}}(\xi)$ (---) from eqs 8 and 9. ξ is the reaction coordinate (minimal H(O) to N distance), and the averaging window length $d\xi$ is 0.01 nm.

(Figure 2b) and the total potential energy of the system (Figure 2c). The correspondence between the two upper curves confirms that, from an energetical point of view, ξ is a reasonable reaction coordinate for analyzing the evolution of the system. Note that the Coulombic interaction energy is most of the time positive due to intramolecular interactions between like charges in the two vicinal groups of the two molecules. This does not matter since only relative variations in energy are relevant. The fluctuations in the electrostatic energy remain small, however, in comparison to the fluctuations in the total potential energy of the system. As can be seen from Figure 2a, binding of the two molecules is essentially an on–off process, about 25–30 hops being observed during the 100-ns simulation. Results for simulations B–D (not shown) are qualitatively very similar.

The probability density distribution, $p(\xi)$, corresponding to the chosen reaction coordinate, was evaluated by using

$$p(\xi) = \frac{1}{d\xi} \frac{1}{N_{\text{frames}}} \sum_{\tau=1}^{N_{\text{frames}}} d\sigma(\xi'(\tau); \xi, d\xi) = \frac{1}{d\xi} \langle d\sigma(\xi'(\tau); \xi, d\xi) \rangle \quad (1)$$

where $\xi'(\tau)$ is the value of the reaction coordinate in the configuration or frame τ of the simulation, N_{frames} is the total number of time frames in the trajectory (2×10^6 for simulations A and B and 10^6 for simulations C and D), $d\xi$ is the window length of the analysis (0.01 nm), and $\langle \dots \rangle$ denotes ensemble averaging over the trajectory. The windowing function $d\sigma(\xi'; \xi, d\xi)$ selects trajectory frames τ with reaction coordinate $\xi'(\tau)$ belonging to a window of length $d\xi$ around ξ , that is

$$\begin{aligned} d\sigma(\xi'; \xi, d\xi) &= 1 && \text{if } \xi \leq \xi' < \xi + d\xi \\ d\sigma(\xi'; \xi, d\xi) &= 0 && \text{otherwise} \end{aligned} \quad (2)$$

The probability density $p(\xi)$ is displayed together with its integral in Figure 3a for simulation A. The results for

simulations B–D are qualitatively very similar. The bound state is visible as a sharp peak (left-hand side), corresponding to the narrow region of enthalpically favored states. A low probability region around $\xi = 0.5$ nm separates it from the free state, visible as a broad (right-hand side) maximum. The finding that the probability distribution for the free state contains a maximum is solely due to the finite size of the system. At the right-hand side of the minimum (0.5 nm), the probability starts to increase with ξ for entropic reasons (the number of accessible states is proportional to $\xi^2 d\xi$). The nearest box wall is, however, encountered at distance $\xi = 1.875$ nm, and past this distance the number of accessible states decreases again, since either the molecule is located in a corner region of the box (volume decreasing with ξ) or a molecule pair consisting of different periodic images is selected and ξ is thereby reduced or “folded back” to lower values. For the rest of the analysis, the bound and free states of the system were formally distinguished by using distance $\xi_{\text{cut}} = 0.5$ nm. Around this value the integral of $p(\xi)$ (Figure 3a) shows a weak dependence on ξ , and thus the precise choice of ξ_{cut} is not essential. The corresponding bound and free population fractions, α_{bound} and α_{free} , are evaluated as

$$\begin{aligned} \alpha_{\text{bound}} &= \int_0^{\xi_{\text{cut}}} p(\xi) d\xi = \frac{1}{N_{\text{frames}}} \sum_{\tau=1}^{N_{\text{frames}}} h_b(\xi'(\tau); \xi_{\text{cut}}) = \langle h_b(\xi'(\tau); \xi_{\text{cut}}) \rangle \\ \alpha_{\text{free}} &= \int_{\xi_{\text{cut}}}^{\infty} p(\xi) d\xi = \frac{1}{N_{\text{frames}}} \sum_{\tau=1}^{N_{\text{frames}}} h_f(\xi'(\tau); \xi_{\text{cut}}) = \langle h_f(\xi'(\tau); \xi_{\text{cut}}) \rangle \end{aligned} \quad (3)$$

where $\xi'(\tau)$ is the value of the reaction coordinate in frame τ of the simulation, $\xi_{\text{cut}} = 0.5$ nm, $h_f(\xi'; \xi_{\text{cut}})$ is the Heaviside function, that is

$$\begin{aligned} h_f(\xi'; \xi_{\text{cut}}) &\equiv \int_{\xi=\xi_{\text{cut}}}^{\infty} d\sigma(\xi'; \xi, d\xi) = 1 && \text{if } \xi' > \xi_{\text{cut}} \\ h_f(\xi'; \xi_{\text{cut}}) &\equiv \int_{\xi=\xi_{\text{cut}}}^{\infty} d\sigma(\xi'; \xi, d\xi) = 0 && \text{otherwise} \end{aligned} \quad (4)$$

which selects trajectory frames τ with reaction coordinate $\xi'(\tau)$ above ξ_{cut} , and $h_b(\xi'; \xi_{\text{cut}})$ is defined as $h_b \equiv (1 - h_f)$. From the definitions in Equation (3) follows immediately that $\alpha_{\text{bound}} + \alpha_{\text{free}} = 1$. Values of the population fractions calculated for the four simulations are reported in Table 1.

Thermodynamic Properties. Given the fractions α_{bound} and α_{free} of bound and free species, the equilibrium constant for binding can be calculated as

$$K_b = \frac{1}{c_o} \frac{\alpha_{\text{bound}}}{\alpha_{\text{free}}^2} \quad \text{with} \quad c_o = \frac{1}{N_A V_{\text{box}}} \quad (5)$$

where c_o is the overall concentration of the solute, N_A is Avogadro's number, and V_{box} is the volume of the simulation box. The definition of c_o given in equation 5 appeared to be the most straightforward, but the relation to its macroscopic counterpart is not straightforward since the simulated system is periodic, whereas a real solution is not. If one molecule moves, while the other is static, the first one will always encounter a periodic image of the second one at regular locations, and always with the same orientation, which is not the case in reality. Thus an entropic contribution may be missed. We shall assume that the effect of this is small. c_o evaluates to 0.041 M, which is about one order of magnitude higher than

Table 1. Thermodynamic Parameters for Binding (Experimental and Calculated) for the Different Simulations^a

	α_{bound}	K_b (1/M)	ΔA_b (kJ/mol)	ΔU_b (kJ/mol)	ΔS_b (J/(K·mol))	FFP	$d\Delta A_b/d\lambda$ (kJ/mol)	$d\Delta U_b/d\lambda$ (kJ/mol)	$d\Delta S_b/d\lambda$ (J/(K·mol))
R/R₅ Complex									
exp benzene		40	-9.3	-20.2	-36.4				
exp CCl ₄		97	-11.5	-17.6	-20.3				
simulation A	0.550	66.2	-10.4	-22.6	-41.0	$q_{\text{H(N)}}$	-40.0	-65.2	-84.7
						$q_{\text{H(O)}}$	-40.6	-60.9	-68.0
						$k_{\text{tors}}(\text{CCOH})$	-0.015	-0.190	-0.586
simulation C	0.383	24.6	-7.9	-17.7	-32.8	$q_{\text{H(N)}}$	-27.1	-45.5	-61.5
						$q_{\text{H(O)}}$	-27.5	-43.3	-52.9
						$k_{\text{tors}}(\text{CCOH})$	-0.082	-0.376	-0.984
R/S₅ Complex									
exp benzene		15	-6.8	-41.7	-116.3				
exp. CCl ₄		53	-10.0	-34.3	-81.2				
simulation B	0.508	51.2	-9.8	-21.6	-39.7	$q_{\text{H(N)}}$	-37.7	-59.0	-71.5
						$q_{\text{H(O)}}$	-37.9	-52.5	-49.0
						$k_{\text{tors}}(\text{CCOH})$	-0.009	-0.146	-0.460
simulation D	0.437	33.7	-8.7	-18.4	-32.3	$q_{\text{H(N)}}$	-28.3	-46.4	-60.8
						$q_{\text{H(O)}}$	-28.3	-44.3	-53.9
						$k_{\text{tors}}(\text{CCOH})$	-0.076	-0.264	-0.632

^a Simulation: A and C = **R/R₅**, equatorial (A, 100 ns) or axial (C, 50 ns), B and D = **R/S₅**, equatorial (B, 100 ns) or axial (D, 50 ns). Experimental values are in benzene or CCl₄ at 298 K. α_{bound} is the fraction of bound species from eq 3. K_b is the equilibrium constant for binding, from eq 5. ΔA_b is the free energy change upon binding, from eq 6. ΔU_b is the internal energy change upon binding, from eq 10. ΔS_b is the entropy change upon binding, from eq 11. FFP is the force-field parameter f for which first-order extrapolation is calculated. λ is a scaling factor defined as f/f^0 where f^0 is the reference (actual simulation) value. $q_{\text{H(N)}}$ is the charge of a hydrogen in the amino groups. $(q_{\text{H(N)}})^0 = 0.415 e$. $q_{\text{H(O)}}$ is the charge of a hydrogen in the hydroxyl groups. $(q_{\text{H(O)}})^0 = 0.398 e$. $k_{\text{tors}}(\text{CCOH})$ is the force constant for the C–C–O–H torsional dihedral potential. $(k_{\text{tors}}(\text{CCOH}))^0 = 1.254 \text{ kJ/mol}$. $d\Delta A_b/d\lambda$, $d\Delta U_b/d\lambda$, and $d\Delta S_b/d\lambda$ are the corresponding estimated changes with respect to the selected λ scaling factor, from eqs B.2, B.4, and B.5. The experimental values correspond to measures at constant pressure, and thus ΔG_b and ΔH_b are actually measured instead of ΔA_b and ΔU_b . The values for simulations B and D are not corrected by the equatorial–equatorial to axial–axial isomerization energy, $\Delta A_{\text{iso}} \geq 11.8 \text{ kJ/mol}$.

the concentrations used in the experiment. The corresponding Helmholtz (NVT) free energy change upon binding is

$$\Delta A_b = -k_B T \ln K_b = -k_B T [-\ln c_o + \ln \alpha_{\text{bound}} - 2 \ln \alpha_{\text{free}}] \quad (6)$$

where k_B is Boltzmann's constant and T is the temperature of the system (298 K).

The calculation of the internal energy change upon binding requires in principle the profile of the average total energy (Hamiltonian) along the reaction coordinate. However, since the temperature is held constant, the kinetic energy is independent of ξ . Thus, it is sufficient to consider the average potential energy profile, $\bar{V}(\xi)$, defined as

$$\bar{V}(\xi) = \frac{\sum_{\tau=1}^{N_{\text{frames}}} V(\tau) d\sigma(\xi'(\tau); \xi, d\xi)}{\sum_{\tau=1}^{N_{\text{frames}}} d\sigma(\xi'(\tau); \xi, d\xi)} = \frac{\langle V(\tau) d\sigma(\xi'(\tau); \xi, d\xi) \rangle}{\langle d\sigma(\xi'(\tau); \xi, d\xi) \rangle} \quad (7)$$

where $V(\tau)$ is the total potential energy of the system in frame τ of the simulation. The internal energies of the free and bound states will depend on the cutoff distance ξ and can be calculated by using the population weighted integrals

$$U_{\text{bound}}(\xi) = \int_0^{\xi} p(\xi') \bar{V}(\xi') d\xi' / \int_0^{\xi} p(\xi') d\xi' + C$$

$$= \frac{\sum_{\tau=1}^{N_{\text{frames}}} V(\tau) h_b(\xi'(\tau); \xi)}{N_{\text{frames}} \alpha_{\text{bound}}} + C = \frac{\langle V(\tau) h_b(\xi'(\tau); \xi) \rangle}{\langle h_b(\xi'(\tau); \xi) \rangle} + C \quad (8)$$

and

$$U_{\text{free}}(\xi) = \int_{\xi}^{\infty} p(\xi') \bar{V}(\xi') d\xi' / \int_{\xi}^{\infty} p(\xi') d\xi' + C$$

$$= \frac{\sum_{\tau=1}^{N_{\text{frames}}} V(\tau) h_f(\xi'(\tau); \xi)}{N_{\text{frames}} \alpha_{\text{free}}} + C = \frac{\langle V(\tau) h_f(\xi'(\tau); \xi) \rangle}{\langle h_f(\xi'(\tau); \xi) \rangle} + C \quad (9)$$

where C is an undetermined constant, which includes the kinetic energy. The energy change upon binding is then

$$\Delta U_b = U_{\text{bound}}(\xi_{\text{cut}}) - U_{\text{free}}(\xi_{\text{cut}}), \quad \xi_{\text{cut}} = 0.5 \text{ nm} \quad (10)$$

The values of $U_{\text{bound}}(\xi)$ and $U_{\text{free}}(\xi)$ are displayed together with $\bar{V}(\xi)$ in Figure 3b for simulation A. The corresponding graphs for the three other simulations are essentially identical, but with a different offset in the energy. The $\bar{V}(\xi)$ curves for simulations C and D (axial–axial diamine) are on average about 160 kJ/mol higher than those for simulations A and B (equatorial–equatorial diamine), due to different C–C–C–C torsional force-field parameters for the diamine. The form and magnitude of the curves are, however, very similar. In all cases, the potential energy profile is nearly flat for distances larger than about 0.5–0.6 nm, indicating that entropy essentially governs the behavior of the system above these distances. Below this value, a significant dip of 25–30 kJ/mol is observed, corresponding to the formation of the complex. The rapid increase for very short distances (<0.2 nm) corresponds to unfavorable molecular contacts, occurring when the two molecules are squeezed against each other. The value of ΔU_b (i.e. the difference between the $U_{\text{bound}}(\xi)$ and $U_{\text{free}}(\xi)$ curves in Figure 3b) is again relatively insensitive to the exact value of ξ_{cut} around 0.5 nm. Since ΔA_b and ΔU_b are known, the entropy change upon binding can be estimated by using

$$\Delta S_b = \frac{1}{T} (\Delta U_b - \Delta A_b) \quad (11)$$

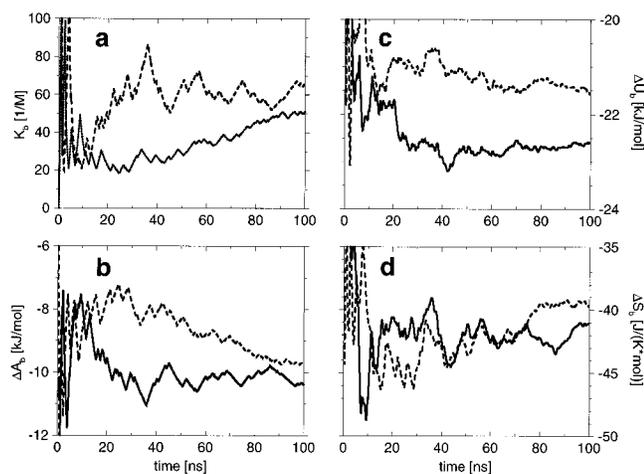


Figure 4. Cumulative time averages of (a) K_b , from eq 5, (b) ΔA_b , from eq 6, (c) ΔU_b , from eq 10, and (d) ΔS_b , from eq 11 (—) for the equatorial \mathbf{R}/\mathbf{R}_5 case (simulation A) and (---) for the equatorial \mathbf{R}/\mathbf{S}_5 case (simulation B). Small, high-frequency fluctuations in ΔU_b and ΔS_b are smoothed out by averaging the obtained curves over 0.5 ns.

The cumulative time averages of K_b , ΔA_b , ΔU_b , and ΔS_b for simulations A and B are displayed in Figure 4a–d. K_b (and thus ΔA_b) converges in a stepwise manner, a consequence of the equilibrium being considered as an on–off process. ΔA_b appears to be converged within 1–2 kJ/mol at the end of the 100-ns simulations. The thermodynamic parameters calculated over the whole simulation time in the four cases are reported in Table 1. In both the equatorial and axial cases, the differences observed in the experimental thermodynamic parameters for binding (either in benzene or in CCl_4), ΔH_b and ΔS_b , are not reproduced in the simulations. The correct enantiospecificity is obtained in the equatorial (simulations A and B) case, but the difference in affinities between the diastereomeric pairs is smaller than that in the experiment. The thermodynamic parameters obtained for the equatorial \mathbf{R}/\mathbf{R}_5 case (simulation A) are in rough agreement with the experimental values measured in CCl_4 , and in good agreement with the values measured in benzene, but those calculated for the diastereomeric pair are essentially identical. In the axial case (simulations C and D), the thermodynamic parameters are also very similar for both diastereomeric pairs. Although the entropy decrease is marginally higher (6–8 J/(K·mol)) upon binding than in the equatorial case, the binding enthalpy, however, is significantly higher (3–5 kJ/mol). Moreover, the axial–axial conformer of the diamine is not the one present in solution at equilibrium, and the conformational work for isomerization, ΔA_{iso} , should be included in the calculated affinities. For a single amino group, the equatorial to axial isomerization energy in CD_2Cl_2 amounts to⁴⁴ 5.9 kJ/mol and twice this value is a lower bound to ΔA_{iso} . Considering this, the possibility of an axial mode of binding for both molecular pairs under study can safely be discarded. If the equatorial mode of binding is the correct one, the discrepancy observed between theory and experiment for the equatorial \mathbf{R}/\mathbf{S}_5 case can be due to (i) the incapacity of the model itself to mimic reality or (ii) the inaccuracy of the model parameters. These possibilities will be discussed in the following subsections.

Sensitivity of the Thermodynamic Properties to Selected Model Parameters. Consider a macroscopic observable $X(\lambda)$, depending on a force field parameter λ . Let us assume that $X(\lambda)$ is the (canonical) ensemble average of an instantaneous microscopic variable $x(\vec{q}, \vec{p}; \lambda)$ depending on the coordinate vector

\vec{q} and momenta \vec{p} of the atoms in the system, and possibly on λ , i.e. that

$$X(\lambda) = \langle x(\vec{q}, \vec{p}; \lambda) \rangle_\lambda = \frac{\int \int d\vec{q} d\vec{p} x(\vec{q}, \vec{p}; \lambda) e^{-\mathcal{H}(\vec{q}, \vec{p}; \lambda)/k_B T}}{\int \int d\vec{q} d\vec{p} e^{-\mathcal{H}(\vec{q}, \vec{p}; \lambda)/k_B T}} \quad (12)$$

where $\mathcal{H}(\vec{q}, \vec{p}; \lambda)$ is the λ -dependent Hamiltonian of the system, k_B is the Boltzmann constant, and T is the temperature. There are two approaches to extrapolate the behavior of $X(\lambda)$ from a single reference simulation:⁴⁵ (i) use of the perturbation formula, or (ii) development of the function in a Taylor series. The formalism corresponding to the use of the perturbation formula is summarized in Appendix A. The ensemble averages of eq A.3 and eq A.4, together with eq 11, give an estimate of the thermodynamic parameters for binding, $\Delta A_b(\lambda)$, $\Delta U_b(\lambda)$, and $\Delta S_b(\lambda)$, at any λ . Additionally, it is useful to have some measure of the reliability of the estimates. Such a measure is proposed in Appendix C in the form of a homogeneity function, $h(\lambda^\circ, \lambda)$ defined in eq C.2, which is related to the fraction of reference configurations contributing significantly to the extrapolation values. The formalism corresponding to the development in a Taylor series, limited to the first order, i.e., of a linear extrapolation, is summarized in Appendix B. The ensemble averages in eqs B.2, B.4, and B.5 give the derivatives of the thermodynamic parameters for binding with respect to λ , $d\Delta A_b/d\lambda$, $d\Delta U_b/d\lambda$, and $d\Delta S_b/d\lambda$, at the reference (λ°) point.

The equilibrium is likely to be most sensitive to the variation of two types of model parameters: the atomic charges within the amino groups and the hydroxyl groups (including the α -carbon), and the force constants of the torsional potential energy terms chosen for the C–C–O–H and C–C–N–H dihedral angles. It is possible to evaluate the dependence of the thermodynamic observables on these parameters. The dependence of the thermodynamic parameters for binding, $\Delta A_b(\lambda)$, $\Delta U_b(\lambda)$, and $-\Delta S_b(\lambda)$ as estimated by the perturbation formalism, is displayed in Figure 5a,b for simulations A and B, where λ is a scaling factor applied to the charges $q_{\text{H(N)}} = -1/2 q_N$ of the hydrogens in the amino groups (i.e., the reference state, $\lambda^\circ = 1$, corresponds to $(q_{\text{H(N)}})^\circ = -1/2 (q_N)^\circ = 0.415 e$). The corresponding linear extrapolations calculated by using eqs B.2, B.4, and B.5 are also displayed. The homogeneity functions, $h(\lambda^\circ, \lambda)$, calculated separately for the bound and free states in both cases, are displayed in Figure 5c. The perturbation formula predicts a continuous increase in ΔU_b when λ is decreased, up to a value of -7.4 (simulation A) or -3.6 kJ/mol (simulation B) at $\lambda = 0$. These values would be expected to be very close to zero, since the only diol–diamine interaction still present at $\lambda = 0$ is the Lennard-Jones interaction, which is small in magnitude. On the other hand, a continuous decrease in $-\Delta S_b$ is predicted when λ decreases, down to a value of 6.2 (simulation A) or 2.7 kJ/mol (simulation B) at $\lambda = 0$. If the solute–solute Lennard-Jones interactions are neglected, this value would be expected to be close to

$$-\Delta S_b(\lambda=0) \approx -k_B T \left[\ln \left(\frac{4\pi}{3} \xi_{\text{cut}}^3 \right) - \ln V_{\text{box}} \right] \quad (13)$$

which evaluates to 10.4 kJ/mol. The underestimations by the perturbation formula of both $\Delta U_b(\lambda=0)$ and $-\Delta S_b(\lambda=0)$ sum to give $\Delta A_b(\lambda=0)$ values of -1.3 (simulation A) and -0.9 kJ/mol (simulation B), which are very close to zero—and thus, unrealistic. On the other hand, when moving to higher charges, a rapid decrease in the binding energy is expected and observed.

(44) Buchanan, G. W.; Webb, V. L. *Tetrahedron Lett.* **1983**, 24, 4519–4520.

(45) Liu, H.; Mark, A. E.; van Gunsteren, W. F. *J. Chem. Phys.* **1996**, 100, 9485–9494.

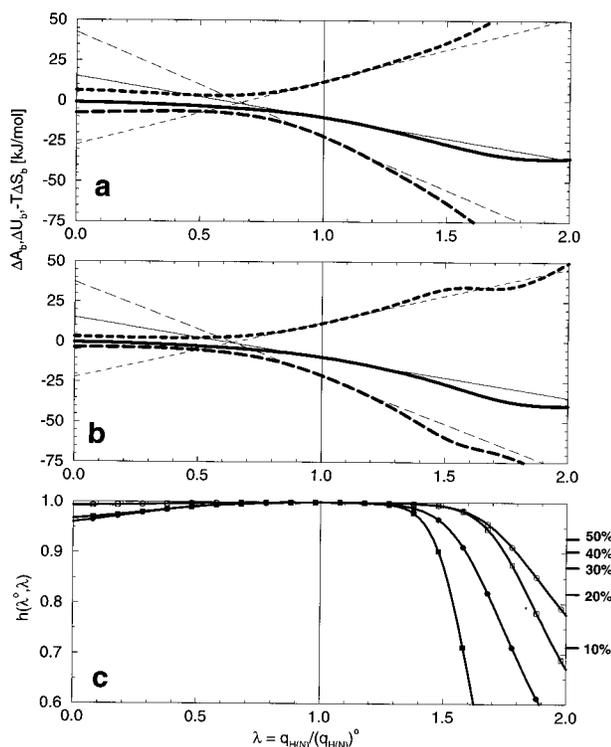


Figure 5. (a) Equatorial R/R_5 case (simulation A) and (b) equatorial R/S_5 case (simulation B): estimates of the variation of the thermodynamic parameters for binding as a function of the ratio $\lambda = q_{H(N)}/(q_{H(N)})^0$, where $q_{H(N)} = -1/2q_N$ is the charge of a hydrogen in the amino group and $(q_{H(N)})^0 = -1/2(q_N)^0 = +0.415 e$ is the (reference) charge used in the actual simulation (i.e., at $\lambda = \lambda^0 = 1$). The thick curves represent perturbation estimates by eqs A.3, A.4, and 11 of $\Delta A_b(\lambda)$ (—), $\Delta U_b(\lambda)$ (---), and $-T\Delta S_b(\lambda)$ (- - -). The thin curves represent the corresponding first-order derivative (linear) estimates by eqs B.2, B.4, and B.5. (c) Homogeneity function $h(\lambda^0, \lambda)$ from eq C.2 for simulation A, bound state (●) or free state (○), and simulation B, bound state (■) or free state (□). Right portion of part c: indicative percentage of configurations, n/N_{frames} , contributing significantly to the perturbation estimate corresponding to the value of $h(\lambda^0, \lambda)$ according to eq C.3.

It is however initially partially compensated by an increase in $-T\Delta S_b$ so that the overall effect is a smooth decrease of ΔA_b . Above $\lambda = 1.3-1.4$ (30–40% increase in the charges), the extrapolation curves become unrealistic, as is evidenced by the homogeneity functions $h(\lambda^0, \lambda)$ in Figure 5c. For both diastereomeric pairs, the curve corresponding to the free state is higher than that for the bound state, in agreement with the fact that the potential energy in the free state is weakly dependent on the charges. On the other hand, the rapid decay of the bound state curves when $\lambda \geq 1.4$ indicates that bound conformations generated at $\lambda^0 = 1$ become rapidly irrelevant for a more highly charged system, and the curves presented in Figure 5a,b will not be reliable above this value. Whereas singly hydrogen-bonded species are dominant in both simulations A and B at the reference state (see the next two subsections), higher charges may result in doubly hydrogen-bonded species increasing in importance. The shape of the confidence function for the two diastereomeric pairs is different, the conformations of the reference ensemble (for both free and bound states) becoming irrelevant at lower λ values in simulation B than in simulation A. The opposite behavior (not shown) is observed when the charges of the hydroxyl group are varied, although evolution of the parameters for binding is qualitatively similar to what is observed in Figure 5a,b.

The fact that, over the whole λ range, the ΔU_b term is higher in simulation B than in simulation A and the $-T\Delta S_b$ term lower

indicates that there is no hope of reproducing the experimental thermodynamic parameter difference between the two diastereomeric pairs, within the range of validity of the perturbation formula, by a tuning of the $q_{H(N)}$ and q_N charges. Note also that below $\lambda = 1.5$, the ΔA_b curve in simulation A is everywhere lower than that in simulation B, but at most by 1.7 kJ/mol ($\lambda = 1.3$). It might be that unexpected changes in the R/S_5 complex occur above the value of $\lambda = 1.5$, which would open the possibility that increased charges may result for this pair in a new conformation more compatible with experimental results. An additional simulation would be required to assess this possibility. Very similar graphs (not shown) are obtained if the charge $q_{H(O)}$ of the protons on the oxygens is varied (keeping the ratio $(q_{H(O)} - q_O)/(q_O - q_{C_a})$ constant).

Comparison of parts a and b of Figures 5 and part c of Figure 5 indicates that the domain of validity of the perturbation formula is larger than that of a linear approximation to it. This is especially true when the charges are decreased. The dependence of the thermodynamic quantities on λ is nonlinear and the perturbation formula performs quite well. The values of the first derivatives from eqs B.2, B.4, and B.5, calculated over the overall simulation time for the four simulations, are reported in Table 1 for two selected force field parameters mentioned above ($q_{H(N)}$ and $q_{H(O)}$), as well as for $\lambda = k_{tors}(C-C-O-H)$, the force constant for the hydroxyl group dihedral angle potential energy terms. In this latter case, the dependence of the thermodynamic observables on the the force-field parameters is negative and very small. An increase in the force constant by a factor 10 would decrease the ΔU_b term by 1.9 (simulation A) or 1.5 kJ/mol (simulation B) and increase the $-T\Delta S_b$ term by nearly the same amount, so that the overall impact on the binding free energy is close to zero. Thus, this force constant does not seem to be an essential parameter of the model with respect to binding. Finally, the thermodynamic observables for simulation A are more sensitive to charge changes than those for simulation B.

Lifetimes of the Complex and Hydrogen Bonds. To gain insight into the time scales of complex formation and dissociation, the distribution of the lifetimes of the complexes was calculated. An excursion time, t_{ex} , defined as the maximal time the distance criterium $\xi < \xi_{cut}$ ($=0.5$ nm) may be violated before considering the bound period as interrupted, has to be chosen. With use of $t_{ex} = 0$, a single configuration with $\xi > \xi_{cut}$ would result in a new complex being counted, and the average lifetime for the complex is close to 15 ps for both 100-ns simulations, despite the longest lifetimes being of the order of 400–500 ps. The histogram of lifetime distributions (25-ps blocks) corresponding to $t_{ex} = 2$ ps is given in Figure 6 for simulation A. Results for simulation B are similar. In both cases, lifetimes of less than 25 ps are dominant, with occurrences in the simulation of 172 and 152, respectively, per 100 ns. Isolated occurrences of lifetimes ranging from 1 to 3 ns are observed, but no statistically meaningful conclusions can be made about them. In all cases, most frequently occurring residence times are between 0 and 0.2–0.25 ns whereas the average residence times are 176 and 206 ps for simulations A and B, respectively. The lifetime of the complex may therefore be estimated to be of the order of 200 ps.

The lifetime of a single hydrogen bond can be roughly estimated by the decay time τ of the autocorrelation function corresponding to the C–C–O–H dihedral angle, that is

$$C(t) = \langle \cos(\phi(t) - \phi(0)) \rangle \quad (14)$$

where $\phi(t)$ is the dihedral angle value at time t . These functions were calculated separately for the dihedral angle $\phi_{min}(t)$ corre-

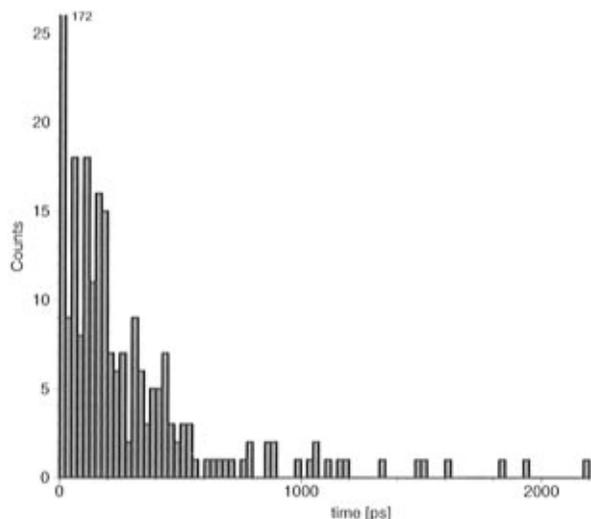


Figure 6. Equatorial R/R_5 case (simulation A): histogram of the residence times distribution. The occurrences are calculated over 25-ps windows, using a value of 2.0 ps for the allowed excursion time (t_{ex}). The number at the right of the first (truncated) bar in the graph indicates its occurrence.

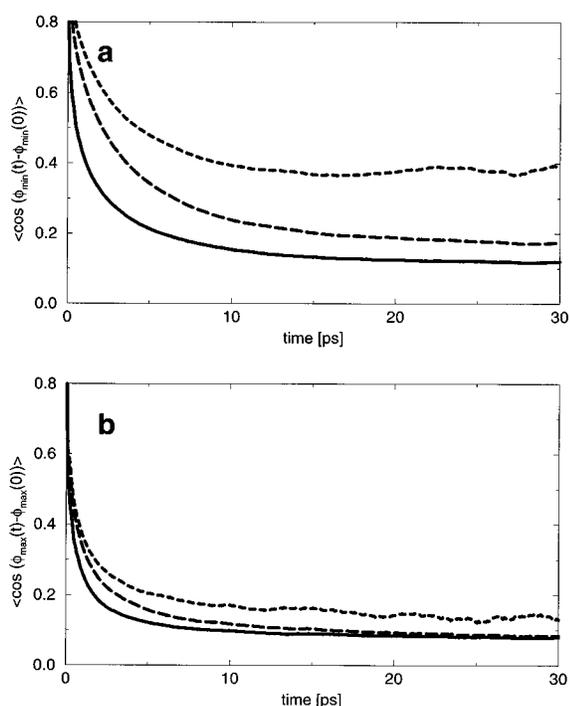


Figure 7. Equatorial R/R_5 case (simulation A): C–C–O–H dihedral angle cosine autocorrelation functions. In part a, the dihedral angle $\phi_{\min}(C-C-O-H)$ corresponding to the minimal H(O) to N distance ($=\xi$) is considered, in part b, the other dihedral angle $\phi_{\max}(C-C-O-H)$ is considered. The function is displayed for all conformations (—), selected conformations with $\xi(t) < 0.5$ nm (---), and selected conformations with $\xi(t) < 0.3$ nm (- - -).

sponding to the hydroxyl group with the shortest H(O) to N distance (i.e. ξ) and the other dihedral angle, $\phi_{\max}(t)$. Averaging has been performed either over all the conformations or restricted to those satisfying $\xi(t) < 0.5$ and 0.3 nm, respectively. The six corresponding curves are displayed in Figure 7 for simulation A. Results for simulation B are similar. In both cases, the decay times of the correlation functions are between 4.2 and 7.7 ps. The long-time offset in $\phi_{\min}(t)$ increases significantly when the averaging is limited to low ξ values (bound state), whereas that of $\phi_{\max}(t)$ does not. This indicates that there is most often a single strong H(O)→N hydrogen bond

in the bound state, the other hydroxyl group being essentially free to rotate. The average lifetime of a single hydrogen bond in the simulations A and B can thus be estimated to be of the order of 5 ps.

Hydrogen-Bonding Patterns. A systematic analysis of the hydrogen-bonding patterns occurring during the simulations was performed for simulations A and B. A hydrogen bond is assumed to exist between donor D–H and acceptor A if the distance $d(H(D)\cdots A)$ is smaller than 0.25 nm and the angle $\angle(D-H\cdots A)$ is larger than 135° . There are four donors and four acceptors in the system and thus formally as many as $2^{4 \times 3} = 4096$ possible hydrogen bonding categories (including intramolecular hydrogen bonds). Among these, 198 patterns are not equivalent by symmetry and do not involve “return” hydrogen bonds (i.e. from D to A and from A to D at the same time). Among these, only 24 (simulation A) or 25 (simulation B) patterns are actually observed during the simulation. The most populated of them ($>0.5\%$ occurrence) are reported in Table 2, together with the corresponding estimated thermodynamic parameters for binding. From these results, it is quite evident that R_s and S_5 bind to R in essentially the same way. Slight differences in the populations and binding parameters can be seen, but no global trend can be clearly identified. In both cases, besides the unbound state, the singly bonded complex is strongly dominant, with a clear preference for a O→N hydrogen bond, which is entropically slightly less favorable (8–11 J/(mol·K)) but enthalpically much more stable (5–6 [kJ/mol]) than the single N→O hydrogen bond. The thermodynamic parameters for these singly hydrogen-bonded species determine almost entirely the binding parameters for the overall equilibrium (Table 1). As expected, additional hydrogen bonds imply a decrease in the binding entropy. Whereas single bonded species share a $-\Delta S_b$ of about 40–50 J/(mol·K), bridged O species (mainly the N→O→N pattern) share a $-\Delta S_b$ of about 60–70 J/(mol·K) and bridged N species (mainly the O→N→O pattern) a $-\Delta S_b$ of about 80–100 J/(mol·K), which is consistent with a progressive loss of rotational freedom between the two molecules. The N→O→N is favored over the O→N→O pattern exclusively for entropic reasons. A whole variety of patterns involving both amino and both hydroxy groups coexist in small amounts at equilibrium. Their $-\Delta S_b$ value is generally still higher, of the order of 90–125 J/(mol·K), indicating a quasi total freezing of the rotational freedom between the molecules. Among these doubly hydrogen-bonded species, the most frequently occurring patterns are O→N, O←N and (to a lesser extent) O→N, O→N. Snapshots from simulations A and B representing the minimum electrostatic energy structures corresponding to both of these patterns are displayed in Figure 8a–d. The binding entropies for bridged N and doubly hydrogen-bonded species are in the range of experimental results for the R/S_5 complex (Table 1). The binding energy changes for species having a ΔS_b matching the experimental values either in benzene or in CCl_4 are, however, still about 10 kJ/mol higher. According to the simulation results, these states would become dominant only at temperatures of the order 150–200 K. Using the distance and angle criterium mentioned above, intramolecular hydrogen bonds are never observed.

Dihedral Angle Distributions. The distributions of the diol O–C–C–O and diamine N–C–C–N dihedral angles corresponding to the bound and unbound states are displayed in Figure 9 for the R/R_5 case (simulation A). As a result of the puckering of the five-membered ring, the O–C–C–O dihedral angle distribution (Figure 9a) is broad, extending from about 50° to 170° , with a dominant conformation at 80° . Binding to R somewhat favors this dominant conformation at the expense

Table 2. Hydrogen Bonding Patterns Present at Equilibrium During Simulations A and B, and Accounting for More Than 0.5% of the Trajectory Frames^a

		equatorial R/R ₅ complex (simulation A)					equatorial R/S ₅ complex (simulation B)				
pattern		occurrence (%)	<i>K</i> _b (1/M)	ΔA_b (kJ/mol)	ΔU_b (kJ/mol)	ΔS_b (J/(K·mol))	occurrence (%)	<i>K</i> _b (1/M)	ΔA_b (kJ/mol)	ΔU_b (kJ/mol)	ΔS_b (J/(K·mol))
total	free state	54.2			0.0		57.9			0.0	
	O→N	20.9	17.4	-7.1	-21.1	-47.0	20.1	14.7	-6.7	-21.2	-48.9
	N→O	7.7	6.4	-4.6	-16.1	-38.7	7.4	5.4	-4.2	-15.4	-37.7
total	singly bonded	28.6	23.8	-7.9	-19.7	-39.9	27.5	20.1	-7.4	-19.7	-41.0
	N→O→N	6.7	5.6	-4.3	-25.3	-70.7	6.8	5.0	-4.0	-24.8	-69.8
total	bridged O	6.9	5.7	-4.3	-25.1	-69.6	6.9	5.0	-4.0	-24.6	-68.9
	O→N→O	2.1	1.7	-1.3	-25.4	-80.6	1.8	1.3	-0.7	-24.9	-81.5
	O→N←O	0.6	0.5	1.7	-27.9	-99.4	0.6	0.4	2.2	-27.8	-100.6
total	bridged N	2.7	2.2	-2.0	-25.9	-80.2	2.4	1.7	-1.4	-25.6	-81.3
	O→N,O→N	4.0	3.3	-3.0	-29.8	-90.0	3.3	2.4	-2.2	-29.1	-90.4
	O→N,O→N	1.6	1.3	-0.7	-33.6	-110.5	1.0	0.7	0.9	-30.6	-105.6
	O→N←O→N	0.9	0.7	0.8	-36.4	-124.9	0.5	0.4	2.5	-33.0	-119.1
	O←N←O→N	0.6	0.5	1.9	-32.3	-114.5	<0.5				
total	doubly bonded	7.6	6.4	-4.6	-31.4	-90.0	5.3	4.0	-3.4	-29.5	-87.8

^a Occurrence is the occurrence in the simulation. Pattern is the hydrogen-bonding pattern: D→A or A←D indicates a hydrogen bond from donor D to acceptor A, a comma indicates that another N/O pair is involved. *K*_b is the equilibrium constant for binding calculated as follows: [species with specified pattern]/[free species]². ΔA_b is the corresponding Helmholtz free energy of binding, from eq 6. ΔU_b is the internal energy change upon binding, from eq 10, with *U*_b equal to the average energy of the species with the specified pattern. ΔS_b is the entropy change upon binding, from eq 11. Values for the free state are not given since they do not correspond to a physical equilibrium. The criterium for a hydrogen bond between D and A is *d*(H(D)···A) < 0.25 nm and \angle (D-H···A) > 135°. Quantities are also calculated for groups of similar patterns in the "total" lines.

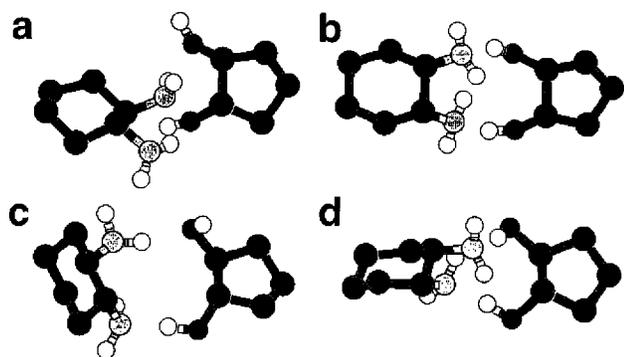


Figure 8. (a, b) Equatorial **R/R**₅ case (simulation A) and (c, d) equatorial **R/S**₅ case (simulation B): snapshots from the simulations representing the minimum electrostatic energy structure corresponding to (a, c) the O→N,O→N hydrogen-bonding pattern, *V*_{Cb} = -11.8 kJ/mol (a) and -7.7 kJ/mol (b), and (b, d) the O→N,O→N hydrogen-bonding pattern, *V*_{Cb} = -25.0 kJ/mol (b) and -16.4 kJ/mol (d).

of those with a higher dihedral angle value. The N-C-C-N dihedral angle distribution (Figure 9b) is sharper and peaks around the ideal value of 60° for a staggered conformation. Binding to **R**₅ results in a slight broadening of the distribution. Very similar distributions and distribution changes upon binding are observed in the **R/S**₅ case (not shown).

Discussion

Analysis of the C-C-O-H dihedral angle cosine autocorrelation function and of the hydrogen-bonding patterns present at equilibrium showed that for both diastereomeric forms, the dominant species observed in the simulations are the ones containing a single O→N hydrogen bond, and to a lesser extent, the one with a single N→O hydrogen bond. This is expected to be correct in the **R/R**₅ case since a good agreement with the experimental values in benzene is observed. The agreement with the results in CCl₄ is poorer, but it should be recalled that the solvent used in the simulations is an extremely simplified model of apolar solvent, which properties might actually be closer to those of benzene than to those of CCl₄. The experimental thermodynamic parameters for the **R/S**₅ case seem

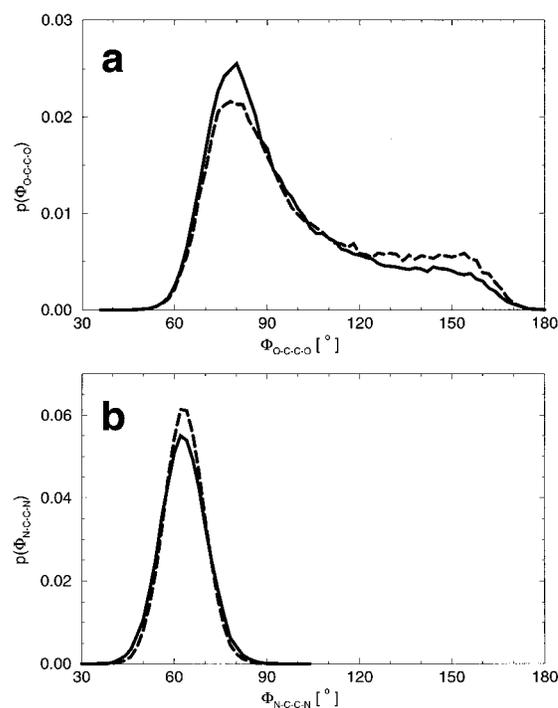


Figure 9. Equatorial **R/R**₅ case (simulation A): normalized probability distribution of (a) the O-C-C-O dihedral angle in **R**₅ and (b) the N-C-C-N dihedral angle in **R**, for the bound (—) state ($\xi(t)$ < 0.5 nm) and for the unbound (---) state ($\xi(t)$ ≥ 0.5 nm).

more consistent with a predominance of doubly hydrogen-bonded species. These species are indeed seen in the simulation and their binding entropies are consistent with the experimental results. However, their ΔU_b values are about 10 kJ/mol higher than the experimental values, and they are thus present in minor quantities. The mono-hydrogen-bonded species still dominates the equilibrium in this case. If one assumes that doubly hydrogen-bonded species dominate in the real system and determine the experimental value of ΔA_b = -6.8 (benzene) or -10.0 kJ/mol (CCl₄) in the **R/S**₅ case, it seems puzzling that a singly hydrogen-bonded species determines experimentally a

value of $\Delta A_b = -9.3$ (benzene) or -11.5 kJ/mol (CCl_4) in the **R/R**₅ case. This would mean that the (hypothetical) singly hydrogen-bonded species in the **R/S**₅ case has a ΔA_b greater than the experimentally observed ΔA_b for this complex, and thus must be destabilized by more than 2.5 (benzene) or 1.5 kJ/mol (CCl_4) with respect to its homologue in the **R/R**₅ case. Such a destabilization is difficult to account for with use of purely steric considerations and is indeed not observed in the simulation. Therefore, other than steric arguments should be considered to explain the observed differences.

First, the solvent that has been used in the simulations has a relative dielectric permittivity of 1 compared to⁴⁶ 2.28 for benzene and 2.24 for CCl_4 at 293 K. This would affect the equilibrium constants for both diastereomeric pairs, but it is unlikely that dielectric effects alone induce a high enantioselectivity. Furthermore, the differences observed between the experimental results in benzene and CCl_4 cannot be rationalized in terms of dielectric effects since the two solvents have nearly the same dielectric constant. Alternatively, it is possible that the solvent interacts specifically with the solute. It is known that benzene can form hydrogen bonds.^{47–49} Values for the strengths of OH or NH hydrogen bonds to benzene reported in the literature are of the order of⁵⁰ 8.5 to 17 kJ/mol. From rotational spectra of the water–benzene dimer, the interaction energy has been estimated⁵¹ to be of the order of 6.8–11.6 kJ/mol. *Ab initio* calculations on the water–benzene dimer in vacuum give an estimate of 15.6 kJ/mol for the binding energy.⁵² For the ammonia–benzene dimer an upper bound of about 10 kJ/mol for the binding energy has been suggested.⁵³ Thus specific solute–benzene interactions might play a role in the equilibrium under study. However, such specific solute–solvent interactions are not likely to play a significant role in the CCl_4 case, and thus cannot alone explain the enantioselectivity observed in both solvents. Finally, the solvent model used in the simulations consists of isotropic “molecules”, whereas the real solvents do not. The formation of different hydrogen-bonded species in the complex may thus result in the breakdown of the liquid packing and orientational structure to a different extent (excluded volume effects). If such specific solvent–solvent interactions play a significant role in the considered equilibria, they are clearly out of reach of our simplified model.

Second, the occurrence of higher order $n:n$ clusters with $n \neq 1$ may also change the characteristics of the equilibrium and invalidate the analysis of the raw experimental data. The binding entropy and enthalpy for such clusters of more than two molecules can be expected to be quite negative and their occurrence only in the **R/S**₅ case would be a possible rationalization of the experimental trends. It is worth noting that the closely related pairs **R**, or its enantiomer **S**, together with **S**₆ may spontaneously self-assemble to form stable 1:1 co-crystals, with melting points of 65 and 79 °C, respectively, in which nearly all the hydrogens are engaged in hydrogen bonds.⁵⁴ In

benzene, fast exchange of molecules within the labile adducts is observed by NMR (-40 to $+25$ °C). At low temperature, the co-crystal precipitates from the solution. It might therefore be possible that higher order clusters also occur at room temperature for the closely related system under study.

Third, the modeling of the hydrogen bonds by a balance between (fixed) point charge interactions and Lennard-Jones repulsions might be insufficiently accurate. The angular dependence of a hydrogen bond, when treated in this way, might be underestimated. In the present case, however, this is not likely to improve the enantioselectivity since the distribution of angles for the doubly hydrogen-bonded species is closer to 180° in the **R/R**₅ case than in the **R/S**₅ case (results not shown). Electronic polarization and quantum effects are not accounted for by using a model with fixed point charges. Hydrogen bonding involves to some extent a charge redistribution. Thus, the formation of a first hydrogen bond may enhance or inhibit the formation of a second one at a close distance.⁵⁵ Indeed, as can be seen from Figure 5b, in the **R/S**₅ case, an increase of the charges of the hydrogens in the amino group by only 0.1 e would bring the binding energy close to the experimental value in benzene (although the entropy change would still be about 65 J/(K·mol) too high). Such electronic effects could only be evidenced by use of a polarizable interaction function or a quantum-mechanical treatment of the complex.

Conclusion

The aim of the present study was to compute enantiospecific thermodynamic parameters of binding for a small complex directly from molecular dynamics simulations and to compare these with data measured in different solvents, thereby providing an interpretation of the experimental results at the molecular level. To reach the required time scales with an explicit solvent, a simplified solvent model was used. For one of the diastereomeric pairs, the results are in close agreement with experimental results in benzene (and in rough agreement with the values measured in CCl_4). However, the calculations could not reproduce the difference observed experimentally between different diastereomeric pairs. This can be due to either (i) the inaccuracy of the model parameters or (ii) the assumptions inherent to the model itself. The first possibility was investigated by statistical extrapolation of the change in thermodynamic observables with respect to force-field parameters. It was found that a change in charge or torsional parameters of the model, within the limit of validity of the perturbation extrapolation, is not likely to improve the agreement with the experimental difference in binding enthalpy and entropy. Other effects, such as specific solute–solvent interactions, the breakdown of specific solvent–solvent interactions around the complex, the formation of clusters of more than two molecules, or electronic charge redistribution effects, might therefore play a role.

Due to the long time scale reached in the simulations (up to 0.1 μs), a detailed and well-converged picture of the dynamics and species present at equilibrium was obtained. The lifetime of the complex is of the order of 200 ps whereas the lifetime of a single hydrogen bond is much shorter, of the order of 5 ps. Apart from the unbound state, the dominant species for both diastereomeric pairs is a singly hydrogen-bonded complex. A clear preference for a O→N hydrogen bond over the N→O hydrogen bond is observed.

(46) Weast, R. C. *Handbook of Chemistry and Physics*, 52nd ed.; Chemical Rubber Publishing CO.: Cleveland, 1971.

(47) Jorgensen, W. L.; Severance, D. L. *J. Am. Chem. Soc.* **1990**, *112*, 4768–4774.

(48) Suzuki, S.; Green, P. G.; Bumgarner, R. E.; Dasgupta, S.; Goddard, W. A., III; Blake, G. A. *Science* **1992**, *257*, 942–945.

(49) Paliwal, S.; Geib, S.; Wilcox, C. S. *J. Am. Chem. Soc.* **1994**, *116*, 4497–4498.

(50) Perutz, M. F. *Phil. Trans. R. Soc. A* **1993**, *345*, 105–112.

(51) Gotch, A. J.; Zwier, T. S. *J. Chem. Phys.* **1992**, *96*, 3388–3401.

(52) Augspurger, J. D.; Dykstra, C. E.; Zwier, T. S. *J. Phys. Chem.* **1992**, *96*, 7252–7257.

(53) Rodham, D. A.; Suzuki, S.; Suenram, R. D.; Lovas, F. J.; Dasgupta, S.; Goddard, W. A., III; Blake, G. A. *Nature* **1993**, *362*, 735–737.

(54) Hanessian, S.; Gomtsyan, A.; Simard, M.; Roelens, S. *J. Am. Chem. Soc.* **1994**, *116*, 4495–4496.

(55) Ryan, M. D. In *Modeling the hydrogen bond*; Smith, D. A., Ed.; Am. Chem. Soc. Symp. Ser.; American Chemical Society: Washington, 1994; pp 36–59.

Acknowledgment. P.H.H. would like to thank Dr. Alan Mark for his critical reading of the manuscript and for helpful suggestions. The Fribourg group acknowledges support from the CHiral2-project of the Swiss National Science Foundation.

Appendix A

Here, the perturbation formula⁵⁶ approach for extrapolating an observable $X(\lambda)$ at a given value λ of a force-field parameter, using data from a reference simulation at λ^0 , is described and applied with ΔU_b , ΔA_b , and ΔS_b as observables. This formalism has also been used to calibrate the reaction field dielectric permittivity by using the dielectric constant as an observable.⁵⁷ The value of $X(\lambda)$ is extrapolated from ensemble averages performed at the reference state, λ^0 , using the formula

$$X(\lambda) = \langle x(\bar{q}, \bar{p}; \lambda) \rangle_\lambda = \frac{\langle x(\bar{q}, \bar{p}; \lambda) e^{-\Delta \mathcal{H}(\bar{q}, \bar{p}; \lambda^0, \lambda)/k_B T} \rangle_{\lambda^0}}{\langle e^{-\Delta \mathcal{H}(\bar{q}, \bar{p}; \lambda^0, \lambda)/k_B T} \rangle_{\lambda^0}} \quad (\text{A.1})$$

where $X(\lambda)$ is the value of X extrapolated at λ , \bar{q} and \bar{p} are the generalized coordinate and momentum ($6N$ dimensional) vectors representing the system, k_B is Boltzmann's constant, T is the (absolute) temperature, $\langle \dots \rangle_\lambda$ denotes averaging over an ensemble generated at a given value of λ , and we define

$$\Delta \mathcal{H}(\bar{q}, \bar{p}; \lambda^0, \lambda) = \mathcal{H}(\bar{q}, \bar{p}; \lambda) - \mathcal{H}(\bar{q}, \bar{p}; \lambda^0)$$

and

$$\Delta V(\bar{q}; \lambda^0, \lambda) = V(\bar{q}; \lambda) - V(\bar{q}; \lambda^0) \quad (\text{A.2})$$

Here, $\mathcal{H}(\bar{q}, \bar{p}; \lambda)$ is the λ -dependent Hamiltonian (total energy) of the system and $V(\bar{q}; \lambda)$ is the corresponding total potential energy. If the kinetic energy is independent of λ and \bar{q} , and $x(\bar{q}, \bar{p}; \lambda)$ is independent of \bar{p} , the Hamiltonian in (A.1) may be replaced by the potential energy. The perturbation estimate for $\Delta A_b(\lambda)$ is obtained by applying eq A.1 to the ensemble averages of eq 3 and inserting the resulting expressions into eq 6, which leads to

$$\begin{aligned} \Delta A_b(\lambda) &= -k_B T \left[-\ln c_0 + \ln \frac{\langle h_b e^{-\Delta V/k_B T} \rangle}{\langle e^{-\Delta V/k_B T} \rangle} - 2 \ln \frac{\langle h_f e^{-\Delta V/k_B T} \rangle}{\langle e^{-\Delta V/k_B T} \rangle} \right] \\ &= \Delta A_b(\lambda^0) - k_B T \left[\ln \langle e^{-\Delta V/k_B T} \rangle + \ln \frac{\langle h_b e^{-\Delta V/k_B T} \rangle}{\langle h_b \rangle} - 2 \ln \frac{\langle h_f e^{-\Delta V/k_B T} \rangle}{\langle h_f \rangle} \right] \\ &= \Delta A_b(\lambda^0) - k_B T \left[\ln \langle e^{-\Delta V/k_B T} \rangle_{\text{bound}} - 2 \ln \langle e^{-\Delta V/k_B T} \rangle_{\text{free}} \right] \quad (\text{A.3}) \end{aligned}$$

where $h_b \equiv h_b(\xi'(\tau); \xi_{\text{cut}})$, $h_f \equiv 1 - h_b(\xi'(\tau); \xi_{\text{cut}})$, $\Delta V \equiv \Delta V(\bar{q}; \lambda^0, \lambda)$, $\xi_{\text{cut}} = 0.5$ nm, and $\langle \dots \rangle$ denotes ensemble averaging over the complete trajectory at the reference (λ^0) state and $\langle \dots \rangle_{\text{bound}}$ and $\langle \dots \rangle_{\text{free}}$ averaging restricted to the bound and free states, respectively. The perturbation estimate for $\Delta U_b(\lambda)$ is obtained by applying eq A.1 to the ensemble averages of eqs 8 and 9

and inserting the resulting expressions into eq 10, which leads to

$$\Delta U_b(\lambda) = \frac{\langle V h_b e^{-\Delta V/k_B T} \rangle}{\langle h_b e^{-\Delta V/k_B T} \rangle} - \frac{\langle V h_f e^{-\Delta V/k_B T} \rangle}{\langle h_f e^{-\Delta V/k_B T} \rangle} = \frac{\langle V e^{-\Delta V/k_B T} \rangle_{\text{bound}}}{\langle e^{-\Delta V/k_B T} \rangle_{\text{free}}} - \frac{\langle V e^{-\Delta V/k_B T} \rangle_{\text{free}}}{\langle e^{-\Delta V/k_B T} \rangle_{\text{free}}} \quad (\text{A.4})$$

where $V \equiv V(\bar{q}(\tau); \lambda)$ is the (now λ dependent) total potential energy of configuration τ in the simulation. The perturbation estimate for $\Delta S_b(\lambda)$ can be obtained from eq 11.

Appendix B

Here, the Taylor expansion approach for extrapolating an observable $X(\lambda)$ at a given value λ of a force-field parameter, using data from a reference simulation at λ^0 , is described and applied with ΔU_b , ΔA_b , and ΔS_b as observables. This formalism has also been used to perform sensitivity analysis and principal component analysis in free energy calculations on blocked dipeptides.^{58,59} In this approach,⁶⁰ eq 12 is differentiated with respect to λ as many times as required, and a polynomial approximation to $X(\lambda)$ is generated by using a Taylor series at λ^0 . In the present work, we shall limit ourselves to the first-order derivative. Under the assumption that the kinetic energy is independent of λ and \bar{q} , and $x(\bar{q}, \bar{p}; \lambda)$ is independent of \bar{p} , the first derivative reads

$$\begin{aligned} \frac{dX(\lambda)}{d\lambda} &= \left\langle \frac{\partial x(\bar{q}; \lambda)}{\partial \lambda} \right\rangle - \frac{1}{k_B T} \left[\left\langle x(\bar{q}; \lambda) \frac{\partial V(\bar{q}; \lambda)}{\partial \lambda} \right\rangle - \langle x(\bar{q}; \lambda) \right\rangle \left\langle \frac{\partial V(\bar{q}; \lambda)}{\partial \lambda} \right\rangle \right] \\ &= \left\langle \frac{\partial x(\bar{q}; \lambda)}{\partial \lambda} \right\rangle - \frac{1}{k_B T} \left[\langle x(\bar{q}; \lambda) - \langle x(\bar{q}; \lambda) \rangle \rangle \left(\frac{\partial V(\bar{q}; \lambda)}{\partial \lambda} - \left\langle \frac{\partial V(\bar{q}; \lambda)}{\partial \lambda} \right\rangle \right) \right] \quad (\text{B.1}) \end{aligned}$$

where $V(\bar{q}; \lambda)$ is the total potential energy of the system, and $\langle \dots \rangle$ denotes ensemble averages performed at λ^0 . The first term in (B.1) is the ensemble average of the explicit derivative of $x(\bar{q}; \lambda)$ with respect to λ . The second term involves the covariance between $x(\bar{q}; \lambda)$ and the derivative of the total potential energy respective to λ .

Applying eq B.1 to the ensemble averages of eq 3, and inserting the resulting expressions into eq 6, one obtains

$$\begin{aligned} \frac{d\Delta A_b}{d\lambda} \Big|_{\lambda^0} &= -k_B T \left[\alpha_{\text{bound}}^{-1} \frac{d\alpha_{\text{bound}}}{d\lambda} - 2\alpha_{\text{free}}^{-1} \frac{d\alpha_{\text{free}}}{d\lambda} \right] \\ &= \left\langle \frac{\partial V}{\partial \lambda} \right\rangle + \langle h_b \rangle^{-1} \left\langle h_b \frac{\partial V}{\partial \lambda} \right\rangle - 2 \langle h_f \rangle^{-1} \left\langle h_f \frac{\partial V}{\partial \lambda} \right\rangle \\ &= \left\langle \frac{\partial V}{\partial \lambda} \right\rangle + \left\langle \frac{\partial V}{\partial \lambda} \right\rangle_{\text{bound}} - 2 \left\langle \frac{\partial V}{\partial \lambda} \right\rangle_{\text{free}} \quad (\text{B.2}) \end{aligned}$$

where $\langle \dots \rangle$ denotes ensemble averaging over all configurations at the reference (λ^0) point, and $\langle \dots \rangle_{\text{free}}$ and $\langle \dots \rangle_{\text{bound}}$ are restricted to the free and bound states, respectively.

(56) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420–1426.

(57) Smith, P. E.; van Gunsteren, W. F. *J. Chem. Phys.* **1994**, *100*, 3169–3174.

(58) Wong, C. F. *J. Am. Chem. Soc.* **1991**, *113*, 3208–3209.

(59) Wong, C. F.; Rabitz, H. *J. Phys. Chem.* **1991**, *95*, 9628–9630.

(60) Smith, P. E.; van Gunsteren, W. F. *J. Chem. Phys.* **1994**, *100*, 577–585.

Applying eq B.1 to the ensemble averages of eq 8 one obtains

$$\begin{aligned} \frac{dU_b}{d\lambda}\Big|_{\lambda^0} &= \langle h_b \rangle^{-2} \left[\left\langle \frac{d}{d\lambda} \langle h_b V \rangle \right\rangle \langle h_b \rangle - \langle h_b V \rangle \left\langle \frac{d}{d\lambda} \langle h_b \rangle \right\rangle \right] \\ &= \langle h_b \rangle^{-1} \left\langle h_b \frac{\partial V}{\partial \lambda} \right\rangle - \frac{1}{k_B T} \left[\langle h_b \rangle^{-1} \left\langle h_b V \frac{\partial V}{\partial \lambda} \right\rangle - \right. \\ &\quad \left. \langle h_b \rangle^{-2} \langle h_b V \rangle \left\langle h_b \frac{\partial V}{\partial \lambda} \right\rangle \right] \\ &= \left\langle \frac{\partial V}{\partial \lambda} \right\rangle_{\text{bound}} - \frac{1}{k_B T} \left[\left\langle V \frac{\partial V}{\partial \lambda} \right\rangle_{\text{bound}} - \right. \\ &\quad \left. \langle V \rangle_{\text{bound}} \left\langle \frac{\partial V}{\partial \lambda} \right\rangle_{\text{bound}} \right] \quad (\text{B.3}) \end{aligned}$$

A similar result is obtained for the free state, and calculation of $d\Delta U_b(\lambda)/d\lambda$ through eq 10 is straightforward

$$\begin{aligned} \frac{d\Delta U_b}{d\lambda}\Big|_{\lambda^0} &= \left\langle \frac{\partial V}{\partial \lambda} \right\rangle_{\text{bound}} - \left\langle \frac{\partial V}{\partial \lambda} \right\rangle_{\text{free}} - \frac{1}{k_B T} \left[\left\langle V \frac{\partial V}{\partial \lambda} \right\rangle_{\text{bound}} - \right. \\ &\quad \left. \left\langle V \frac{\partial V}{\partial \lambda} \right\rangle_{\text{free}} - \langle V \rangle_{\text{bound}} \left\langle \frac{\partial V}{\partial \lambda} \right\rangle_{\text{bound}} + \langle V \rangle_{\text{free}} \left\langle \frac{\partial V}{\partial \lambda} \right\rangle_{\text{free}} \right] \quad (\text{B.4}) \end{aligned}$$

Finally, $d\Delta S_b(\lambda)/d\lambda$ is obtained by differentiating eq 11 and inserting eqs B.2 and B.4

$$\begin{aligned} \frac{d\Delta S_b}{d\lambda}\Big|_{\lambda^0} &= \frac{1}{T} \left[\left\langle \frac{\partial V}{\partial \lambda} \right\rangle_{\text{free}} - \left\langle \frac{\partial V}{\partial \lambda} \right\rangle \right] - \frac{1}{k_B T^2} \left[\left\langle V \frac{\partial V}{\partial \lambda} \right\rangle_{\text{bound}} - \right. \\ &\quad \left. \left\langle V \frac{\partial V}{\partial \lambda} \right\rangle_{\text{free}} - \langle V \rangle_{\text{bound}} \left\langle \frac{\partial V}{\partial \lambda} \right\rangle_{\text{bound}} + \langle V \rangle_{\text{free}} \left\langle \frac{\partial V}{\partial \lambda} \right\rangle_{\text{free}} \right] \quad (\text{B.5}) \end{aligned}$$

The first derivatives of the thermodynamic parameters for binding with respect to λ (at λ^0), under the assumptions mentioned above, can be calculated by using eqs B.2, B.4, and B.5.

Appendix C

It is useful, when performing extrapolations through the perturbation formula, to have a measure of the confidence one can have in the estimates obtained. To this end a so-called homogeneity function $h(\lambda^0, \lambda)$, where λ is the extrapolation parameter and λ^0 its value in the reference simulation, is defined

which allows an assessment of the quality of an extrapolation based on the perturbation formula (Appendix A). In the ensemble average leading to $X(\lambda)$ in eq A.1, configurations \bar{q} of the reference simulation are weighted by a factor

$$f(\bar{q}; \lambda^0, \lambda) = \frac{e^{-\Delta V(\bar{q}(\tau); \lambda^0, \lambda)/k_B T}}{\sum_{i=1}^{N_{\text{frames}}} e^{-\Delta V(\bar{q}(\tau); \lambda^0, \lambda)/k_B T}} = \frac{e^{-\Delta V(\bar{q}(\tau); \lambda^0, \lambda)/k_B T}}{N_{\text{frames}} \langle e^{-\Delta V(\bar{q}(\tau); \lambda^0, \lambda)/k_B T} \rangle_{\lambda^0}} \quad (\text{C.1})$$

When $\lambda = \lambda^0$, this factor is equal to unity for all configurations \bar{q} . For other λ values, values of $f(\bar{q}(\tau); \lambda^0, \lambda)$ uniformly close to one will be characteristic of good sampling, whereas a majority of values close to zero with only a few large values for a few configurations will indicate poor sampling. This can be quantified by the homogeneity function

$$h(\lambda^0, \lambda) = 1 - \frac{1}{(N_{\text{frames}} - 1)^{1/2}} \langle [N_{\text{frames}} f(\bar{q}(\tau); \lambda^0, \lambda) - 1]^2 \rangle_{\lambda^0}^{1/2} \quad (\text{C.2})$$

The function $h(\lambda^0, \lambda)$ is related to the fraction of reference configurations, n/N_{frames} , that contribute significantly to the ensemble average used in the extrapolation. If one assumes that n configurations (over a total number N_{frames}) are contributing with equal weight $f(\bar{q}(\tau); \lambda^0, \lambda) = 1/n$, and all the others are not contributing, it is easily seen that

$$h(\lambda^0, \lambda) = 1 - \left[\frac{N_{\text{frames}} - n}{n(N_{\text{frames}} - 1)} \right]^{1/2} \quad (\text{C.3})$$

If in addition $n = N_{\text{frames}}$, that is if all reference configurations are equally weighted in the ensemble average of eq A.1, $h(\lambda^0, \lambda)$ is unity. If instead $n = 1$, that is if a single configuration is highly weighted and the others not contributing, $h(\lambda^0, \lambda)$ is zero. When $h(\lambda^0, \lambda)$ is below 0.7, the ratio n/N_{frames} drops rapidly from 10% to zero. Thus, a reasonable criterion for the reliability of the extrapolation is a value of $h(\lambda^0, \lambda)$ larger than 0.7. This criterion is, however, not sufficient, since the property may be sometimes relatively insensitive to the fraction of contributing configurations.

JA970503D