Stabilization of a *Cis* Amide Bond in a Host−Guest Complex

Glen J. Pernía,[†] Jeremy D. Kilburn,*[,†] Jonathan W. Essex,[†] Russell J. Mortishire-Smith,*[,‡] and Michael Rowley[‡]

Contribution from the Department of Chemistry, University of Southampton, Southampton, SO17 1BJ, U.K., and Department of Medicinal Chemistry, Merck Sharp and Dohme Research Laboratories, Neuroscience Research Centre, Terlings Park, Eastwick Road, Harlow, Essex, CM20 2QR, U.K.

Received March 11, 1996[⊗]

Abstract: Macrobicycle **12** has been synthesized and its binding properties with a range of *N*-acetyl amino acid carboxylates (as tetrabutylammonium salts) have been studied. While the binding results showed little selectivity for the various substrates investigated, detailed NMR studies have revealed that D-amino acid substrates bind predominantly on the outside of the macrobicycle cavity by a strong carboxylate−thiourea interaction, whereas L-amino acid substrates bind predominantly on the inside of the cavity also establishing a strong carboxylate−thiourea interaction but with the acetyl amide in a *cis* configuration. Molecular modeling studies suggest that the energetic penalty associated with adopting a *cis* amide configuration in the host−guest complex is compensated by intermolecular hydrogen bonds between the *cis* amide and the rim of the macrobicycle.

The energetic preference for a secondary amide to adopt a *trans* configuration in preference to a *cis* configuration is of primary importance in determining the secondary structure of peptides.[1] The energetic preference, for example, for *trans N*-methylacetamide over the *cis* form, is estimated, based on experimental work[2] and on calculations,[3] to be ∼10 kJ mol^{-1} in water and is little affected by change of solvent. As a consequence *cis* amide bonds are rarely found in linear peptides and are usually associated with proline residues.[4] Non-prolyl secondary amides can, however, be constrained to adopt the *cis* configuration by incorporation into cyclic structures,[5] and Peggion *et al*. have reported that the linear heptadecapetide bombolitin, which has an all-*trans* amide structure in free solution, adopts a *cis* secondary Ile-Lys amide bond when incorporated into a micelle.[6] The discovery that the binding proteins for the immunosuppressant agents cyclosporin A, rapamycin, and FK506 are peptidy L-prolyl *cis*/*trans* isomerases has further stimulated interest in the *cis/trans* isomerization of peptides.[7] In this paper we describe a macrobicyclic receptor **12** which selectively binds *N*-acetyl L-amino acid carboxylates within the receptor cavity. In so doing, the acetyl amide adopts a *cis* configuration, stabilized by two N-H···O═C hydrogen bonds to the side wall of the macrobicycle.[8]

The development of synthetic receptors for peptides and amino acid derivatives[9,10] is of considerable interest because the intermolecular interactions involved in small molecule−peptide complexes are of direct relevance to many biological peptide−protein interactions and may also lead to new biosensors, therapeutics and catalysts for peptide hydrolysis. In our own efforts to develop novel receptors for amino acids and peptides we have prepared a range of macrobicycles which feature a specific binding site for carboxylic acid functionality at the base of the cavity (represented schematically in Figure 1).[8,10e]

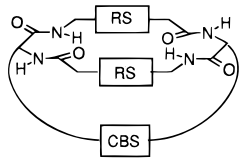


Figure 1. Schematic of macrobicyclic receptor.

By incorporating additional amide functionality around the rim of such a macrobicyclic structure, we intended to provide further hydrogen bonding sites, suitably preorganized to interact with guests such as amino acid derivatives, bound within the macrobicyclic cavity. In macrobicycle **12** the carboxylate binding site is provided by a thiourea moiety, which has been shown to provide a strong binding site for tetraalkylammonium carboxylates, even in relatively competitive solvents such as

[†] University of Southampton.

[‡] Merck, Sharp and Dohme.

[⊗] Abstract published in *Advance ACS Abstracts,* September 15, 1996.

(1) Schultz, G. E.; Schirman, R. H. *Principles of Protein Structure*; Springer-Verlag: New York, 1979.

(2) Radzicka, A.; Pedersen, L.; Wolfenden, R. *Biochemistry* **1988**, *27*, 4538−4541.

(3) Jorgensen, W. L.; Gao, J. *J. Am. Chem. Soc*. **1988**, *110*, 4212−4216.

(4) Stewart, D. E.; Sarkar, A.; Wampler, J. E. *J. Mol. Biol*. **1990**, *214*, 253−260.

(5) (a) Mierke, D. F.; Yamazaki, T.; Said-Nejad, O. E.; Felder, E. R.; Goodman, M. *J. Am. Chem. Soc.* **1989**, *111*, 6847−6849. (b) Kessler, H.; Anders, U.; Schudok, M. *J. Am. Chem. Soc*. **1990**, *112*, 5908−5916. (c) Kukumaran, D. K.; Prorok, M.; Lawrence, D. S. *J. Am. Chem. Soc*. **1991**, *113*, 706−707. (d) Gramberg, D.; Robinson, J. A. *Tetrahedron Lett*. **1994**, *35*, 861.

(6) Bairaktari, E.; Mierke, D. F.; Mammi, S.; Peggion, E. *J. Am. Chem. Soc.* **1990**, *112*, 5383.

(7) (a) Schreiber, S. L. *Science* **1991**, *251*, 283−287. (b) Fischer, G.; Schmid, F. X. *Biochemistry* **1990**, *29*, 2205−2209.

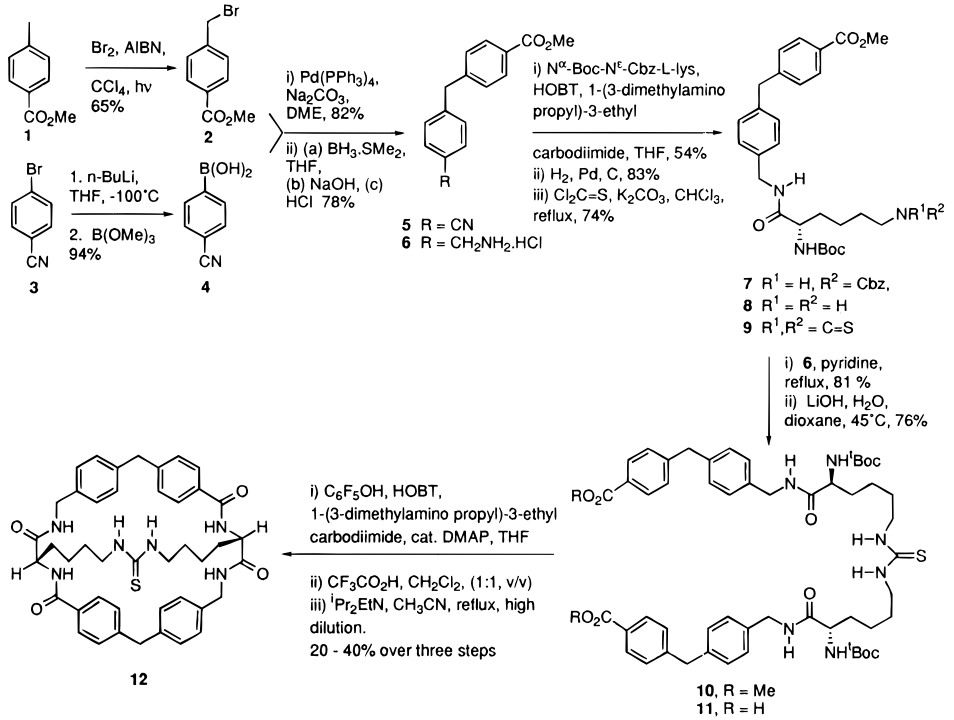
(8) A preliminary account of some of this work has been published: Pernía, G. J.; Kilburn, J. D.; Rowley, M. *J. Chem. Soc., Chem. Commun*. **1995**, 305−306.

(9) For recent reviews on molecular recognition, see: (a) Dowden, J.; Kilburn, J. D.; Wright, P. *Contemp. Org. Synth.* **1995**, *2*, 289. (b) Molecular Recognition (Tetrahedron Symposia No. 56). Hamilton, A. D., Ed.; *Tetrahedron* **1995**, *51*. (c) Webb, T. H.; Wilcox, C. S. *Chem. Soc. Rev.* **1993**, *22*, 383.

(10) For recent work on synthetic receptors for amino acid and peptide derivatives see, for example: (a) Schneider, H.-J. *Angew. Chem., Int. Ed. Engl.* **1993**, *32*, 848. (b) Cristofaro, M. F.; Chamberlin, A. R. *J. Am. Chem. Soc.* **1994**, *116*, 5089. (c) Kuroda, Y.; Kato, Y.; Higashioli, T.; Hasegawa, J.; Kawanami, S.; Takahashi, M.; Shiraishi, N.; Tanabe, K.; Ogoshi, H. *J. Am. Chem. Soc.* **1995**, *117*, 10950. (d) Gennari, C.; Nestler, H. P.; Salom, B.; Still, W. C. *Angew. Chem., Int. Ed. Engl.* **1995**, *34*, 1765. (e) Waymark, C. P.; Kilburn, J. D.; Gillies, I. *Tetrahedron Lett.* **1995**, *36*, 3051. (f) Flack, S. S.; Kilburn, J. D. *Tetrahedron Lett.* **1995**, *36*, 3409. (g) Yoon, S. S.; Still, W. C. *Tetrahedron* **1995**, *51*, 567.

S0002-7863(96)00780-9 CCC: $12.00

Scheme 1



dimethyl sulfoxide.[11] The thiourea also has the potential to be converted into a guanidinium group, which could again serve as a carboxylate binding site. A biarylmethane unit forms the rigidifying part of the rim of the macrobicycle, and chirality and amide functionality are introduced *via* two lysine derivatives.

The synthesis of macrobicycle **12** proved to be relatively straightforward (Scheme 1). Methyl 4-methylbenzoate **1** was converted to bromide **2** in good yield, provided care was taken to avoid dibromination, by slow addition of slightly less than 1 equiv of bromine to the solution of **1**, with irradiation by a 150 W bulb. Treatment of 4-bromobenzonitrile **3** with *n*-butyllithium at −100 °C, followed by addition of trimethylborate, cleanly gave the boronic acid **4**. A palladium mediated Suzuki coupling[12] of bromide **2** with boronic acid **4**, in dimethoxyethane,[13] gave the nitrile ester **5** in consistently good yields of 70−82%. Selective reduction of **5** with borane−methyl sulfide complex gave the corresponding amine **6** which was coupled to N^{α}-*tert*-butyloxycarbonyl-N^{ϵ}-benzyloxycarbonyl lysine. Best yields for this coupling were achieved using 1-(3-dimethylaminopropyl)-3-ethylcarbodiimide methiodide as coupling reagent. Selective removal of the benzyloxycarbonyl protecting group then gave amine **8**, which was converted in two steps to the thiourea **10**, *via* the isothiocyanate **9**.[14] Hydrolysis of the methyl esters of **10** to give diacid **11** was best carried out using equal volumes of 1 M LiOH·H_2O and dioxane at 45 °C. Conversion to the corresponding *bis*pentafluorophenyl ester, removal of the *tert*-butyloxycarbonyl protecting groups and, finally, cyclization by slow addition of the *bis*TFA salt to a refluxing solution of diisopropylethylamine in acetonitrile gave macrobicycle **12** in variable yield but generally in the range 20−40%. The conversion of diacid **11** to final macrobicycle **12** was carried out without purification of the intermediates which were unstable, particularly to chromatography. The synthesis is thus short, reasonably efficient, and provides considerable scope for variation of the various building blocks, which should allow access to a range of related macrobicycles.

Binding studies with macrobicycle **12** were carried out with various acylated amino acids as the tetrabutylammonium salts. Attempts to determine binding constants by conventional NMR titration experiments,[15] with deuterochloroform as solvent, were hindered by the fact that the various NH signals that were monitored during the titration shifted into the aromatic region of the NMR spectrum. Instead, by partitioning the guests between water and chloroform we determined the distribution of the various guests between the two phases (by integration of the ^{1}H NMR against an internal standard). Similarly the distribution of guest between water and a chloroform solution of the macrobicycle was determined, and thus we were able to calculate binding constants (Table 1), by analogy with the picrate extraction method developed by Cram.[16]

Inspection of the binding results indicated that macrobicycle **12** showed little selectivity for amino acid configuration (*R* or *S*) or for the various amino acid side chains, with the possible exception of the L-lysine derivative. Furthermore, binding of simple carboxylates (benzoate, hexanoate) by macrobicycle **12** was essentially as strong as the binding of the more functionalized amino acid derivatives, suggesting that binding of all the substrates studied was almost entirely the consequence of a strong interaction between the carboxylate and the thiourea unit. At first sight, therefore, it seemed that our carefully prepared macrobicycle was little more than an over-elaborate thiourea derivative capable of binding a number of carboxylate salts.

(11) (a) Fan, E.; Van Arman, S. A.; Kincaid, S.; Hamilton, A. D. *J. Am. Chem. Soc.* **1993**, *115*, 369. (b) Smith, P. J.; Reddington, M. V.; Wilcox, C. S. *Tetrahedron Lett.* **1992**, *33*, 6085.

(12) (a) Suzuki, A. *Acc. Chem. Res.* **1982**, *15*, 178. (b) Martin, A. R.; Yang, Y. *Acta Chem. Scand.* **1993**, *47*, 221.

(13) Gronowitz, S.; Bobosik, V.; Lawitz, K. *Chem. Scri.* **1984**, *23*, 120.

(14) (a) Dyson, G. M.; George, H. J. *J. Am. Chem. Soc.* **1924**, 1702. (b) Rasmussen, C. R.; Villani, Jr., F. J.; Weaner, L. E.; Reynolds, B. E.; Hood, A. R.; Hecker, L. R.; Nortey, S. O.; Hanslin, A.; Costanzo, M. J.; Powell, E. T.; Molinari, A. J. *Synthesis* **1988**, 456. (c) Scröeder, D. C. *Chem. Rev.* **1955**, *55*, 193.

(15) Wilcox, C. S. *Frontiers in Supramolecular Organic Chemistry and Photochemistry*; Schneider, H.-J., Durr, H., Eds.; VCH: Weinheim, 1991; pp 123−143.

(16) Kypa, E. P.; Helgeson, R. C.; Madan, K.; Gokel, G. W.; Tarnowski, T. L.; Moore, S. S.; Cram, D. J. *J. Am. Chem. Soc.* **1977**, *199*, 2564.

Table 1. Association Constants for Macrobicycle **12** with Various Tetrabutylammonium Carboxylates in $CDCl_3$

substrate (tetrabutylammonium salt)	$K_a/10^3$ mol^{-1}	$-\Delta G$/kJ mol^{-1}
N-Ac-glycine	68.6 ± 7.9	27.1 ± 0.3
N-Ac-L-alanine	16.9 ± 5.9	23.7 ± 0.9
N-Ac-D-alanine	14.6 ± 2.0	23.3 ± 0.3
N-Ac-L-phenylalanine	22.0 ± 2.8	24.3 ± 0.3
N-Ac-D-phenylalanine	13.3 ± 7.0	23.1 ± 0.4
N-Ac-L-asparagine	9.6 ± 7.0	22.3 ± 0.2
N-Ac-D-asparagine	6.8 ± 2.9	21.5 ± 0.2
N-Ac-L-glutamine	11.1 ± 8.0	22.7 ± 0.5
N^{α}-Ac-L-histidine	5.8 ± 8.0	21.1 ± 0.3
N^{α}-Ac-L-lysine	130.0 ± 20.0	28.7 ± 0.3
Cbz-glycylglycine	8.9 ± 1.6	22.1 ± 0.5
benzoic acid	55.4 ± 17.4	26.6 ± 0.9
hexanoic acid	28.1 ± 3.3	24.9 ± 0.3

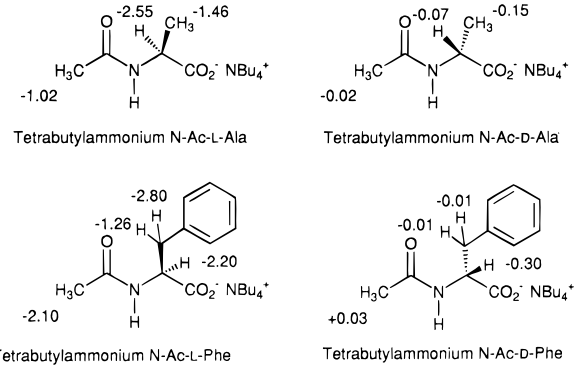


Figure 2. Upfield shifts (ppm) for the various protons indicated in 1:1 complexes with macrobicycle **12**, relative to the signals of the unbound substrates.

However, ^{1}H NMR spectra of 1:1 mixtures of macrobicycle **12** with the tetrabutylammonium salts of D- or L-*N*-acetyl alanine or D- or L-*N*-acetyl phenylalanine showed markedly different chemical shifts for the L-amino acid derivatives compared with the D-amino acid derivatives, suggesting substantially different modes of binding for the two enantiomeric series. In all the complexes the signal for the thiourea NH was shifted substantially downfield (2−3 ppm), relative to the uncomplexed macrobicycle, indicating the formation of hydrogen bonds and consistent with a strong carboxylate−thiourea interaction. ^{1}H chemical shifts for the aliphatic resonances of the D-alanine derivative in the 1:1 complex (∼20 mM, >90% bound) were not significantly perturbed from values observed for the uncomplexed amino acid in $CDCl_3$. In the 1:1 complex of macrobicycle **12** with the L-alanine derivative, however, all four alanine resonances were shifted dramatically upfield by up to 2.5 ppm compared with the uncomplexed amino acid (Figure 2).

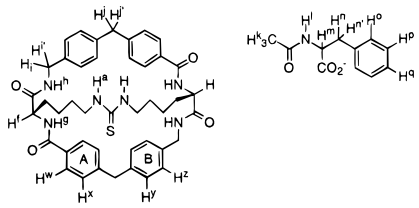
Similar trends were observed in the 1:1 complexes of macrobicycle **12** with the L- and D-phenylalanine derivatives. In the complex formed with the D-phenylalanine derivative only minor differences in chemical shifts were detected for the "bound" substrate compared with the uncomplexed phenylalanine derivative in $CDCl_3$ (Figure 2). For the L-phenylalanine derivative however, the acetyl methyl, NH, $C^{\alpha}H$, and $C^{\beta}H$ resonances were shifted dramatically upfield (Figure 2) although the aromatic resonances for both the guest and macrobicycle were minimally perturbed. This suggests that edge−face and face−face $\pi-\pi$ interactions are unlikely to contribute significantly to association, and indeed it is unlikely that the aromatic side chain is internalized to any great degree.

The observed chemical shifts are consistent with binding of the D-amino acid derivatives to the thiourea moiety on the outside of the macrobicycle, while the L-amino acid derivatives are bound within the cavity afforded by the macrobicycle. The large observed upfield shifts are presumably a consequence of shielding by the aromatic rings which compose the walls of the cavity.

Table 2. Intermolecular NOEs and Distances for the Complex between Macrobicycle **12** and Tetrabutylammonium *N*-Acetyl-L-phenylalanine (at 285 K)

crosspeak assignment	NOE upper distance[a] constraint/Å	distance from unrestrained molecular dynamics/Å	distance from restrained molecular dynamics/Å
$H^k \leftrightarrow H^w$	5.0	5.6	5.3
$H^k \leftrightarrow H^x$	5.0	5.3	5.0
$H^k \leftrightarrow H^y$	4.0	4.0	3.7
$H^k \leftrightarrow H^z$	5.0	4.5	4.5
$H^l \leftrightarrow H^w$	5.0	3.4	3.7
$H^l \leftrightarrow H^y$	5.0	4.8	4.5
$H^l \leftrightarrow H^z$	5.0	4.6	3.9
$H^m \leftrightarrow H^w$	4.0	5.2	5.0
$H^m \leftrightarrow H^x$	4.0	4.3	4.3
$H^m \leftrightarrow H^y$	5.0	4.0	4.0
$H^m \leftrightarrow H^z$	4.0	3.1	3.3
$H^n \leftrightarrow H^w$	4.0	4.1	4.3
$H^{n'} \leftrightarrow H^w$	4.0	4.7	4.6
$H^{n'} \leftrightarrow H^x$	4.0	4.7	4.5
$H^{n'} \leftrightarrow H^y$	4.0	3.2	3.0
$H^o \leftrightarrow H^{i'}$	3.0	3.2	3.2
$H^o \leftrightarrow H^w$	>5.0[b]	4.9	5.3
$H^o \leftrightarrow H^x$	>5.0[b]	5.7	5.7
$H^o \leftrightarrow H^y$	3.0	3.8	3.6
$H^o \leftrightarrow H^z$	3.0	3.5	3.4
$H^p \leftrightarrow H^{i'}$	4.0	4.2	4.4

[a] NOE distance upper bounds were determined from crosspeak intensities by calibration against fixed interproton distances (geminal protons and aromatic protons $H^w \leftrightarrow H^x$ and $H^y \leftrightarrow H^z$). Constraints were binned into 3.0, 4.0, and 5.0 Å categories. [b] No NOE crosspeak was observed. [c] Macrobicycle **12** and L-phenylalanine carboxylate showing ^{1}H labeling.



Detailed 2D NMR studies were carried out (at 285 K and 300 K) to probe the solution conformation of these complexes. Complete assignments for the 1:1 complexes with macrobicycle **12** and the L-phenylalanine and the L-alanine derivatives were obtained using DQF-COSY[17] and TOCSY[18] experiments to define spin systems which were connected using data from ROESY[19] spectra. While NOESY[20] experiments performed on the 1:1 complexes failed to exhibit useful inter- or intramolecular NOE crosspeaks (presumably due to unfavorable correlation times), ROESY spectra obtained at 285 K contained a wealth of structure-defining intermolecular contacts (Table 2).[21]

(17) Jeener, J.; Meier, B. H.; Backmann, P.; Ernst, R. R. *J. Chem. Phys.* **1979**, *71*, 4546.
(18) Rance, M. *J. Magn. Reson.* **1987**, *74*, 557.
(19) Bax, A.; Davis, D. G. *J. Magn. Reson.* **1985**, *63*, 207.
(20) Rance, M.; Sorensen, O. W.; Bodenhausen, G.; Wagner, E. R. R.; Wüthrich, K. *Biochem. Biophys. Res. Commun.* **1983**, *117*, 479.
(21) To eliminate possible errors in interpretation caused by artefacts or spurious intensities, ROESY spectra were acquired with three different transmitter offsets (7.5, 5.0, and 2.5 ppm), each with two mixing times (200 and 400 ms). Each spectrum exhibited qualitatively comparable intensities for the key structure defining ROEs. Spin-lock field strength and temperature also had a minimal effect on the relative ROE intensities.

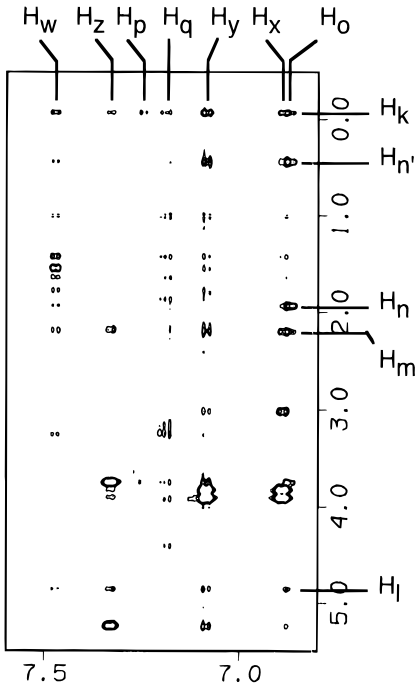


Figure 3. ROESY experiment (at 300 K) on 1:1 complex of **12** with *N*-Ac-L-Phe-CO_2^- showing intermolecular NOEs between the aliphatic protons of the guest and the aromatic sidewalls of the host.

For the 1:1 complex with the L-phenylalanine derivative, the phenylalanine ring is apparently close to the B ring of the macrobicycle as evidenced by crosspeaks from H^y and H^z to the phenylalanine *ortho* protons H^o but not from H^w or H^x.

The upfield shifted acetyl protons (H^k) of the amino acid guest gave NOEs of varying intensity to all four macrobicycle aromatic signals and intraresidue NOEs to the phenylalanine ring. Similarly there are a number of NOEs of varying intensities from the the NH (H^l), α proton (H^m), and the benzylic protons (H^n, $H^{n'}$) of the L-phenylalanine derivative to the macrobicycle aromatic signals (Figure 3).

In contrast, ROESY spectra for the 1:1 complex with the D-phenylalanine derivative indicated that there were no significant intermolecular contacts between the macrobicycle and the guest. The corresponding ROESY spectra for the 1:1 complex with the L-alanine derivative similarly indicated numerous intermolecular contacts particularly to the biarylmethane sidewalls of **12**, but the ROESY spectra for the 1:1 complex with the D-alanine derivative was essentially devoid of intermolecular crosspeaks. These differences are again consistent with binding of the D-amino acid derivatives to the thiourea moiety on the outside of the macrobicycle, while the L-amino acid derivatives are bound within the cavity afforded by the macrobicycle.

For the 1:1 complex with the L-phenylalanine derivative, a full set of intramolecular NOEs for the guest molecule were also identified (Table 3).

Surprisingly, the acetyl methyl (H^k) gave crosspeaks to all three aromatic signals H^o, H^p, and H^q and a particularly strong NOE to the $C^\alpha H$ (H^m) but only a weak NOE to the NH (H^l) (Figure 4).

In peptides, the classic diagnostic for a *cis* amide bond is the observation of a $d_{\alpha\alpha}(i, i+1)$ NOE.[22] In the complex between macrobicycle **12** and the L-phenylalanine derivative, this NOE corresponds to that present between the acetyl methyl (H^k) and the $C^\alpha H$ proton (H^m) of the L-phenylalanine derivative and is

(22) Dyson, H. J.; Wright, P. E. *2-Dimensional NMR Spectroscopy*, Croasmun, Carlson, Eds.; VCH: 1994; p 672.

Table 3. Intramolecular NOEs and Distances for Tetrabutyl Ammonium *N*-Acetyl-L-phenylalanine in the Complex with Macrobicycle **12** (at 285 K)

crosspeak assignment	NOE upper distance[a] constraint/Å	distance from unrestrained molecular dynamics/Å	distance from restrained molecular dynamics/Å
$H^k \leftrightarrow H^l$	5.0	3.9	3.8
$H^k \leftrightarrow H^m$	3.0	2.9	2.6
$H^h \leftrightarrow H^o$	4.0	3.4	3.5
$H^k \leftrightarrow H^p$	5.0	5.0	4.9
$H^k \leftrightarrow H^q$	5.0	5.8	5.6
$H^l \leftrightarrow H^{n'}$	3.0	2.9	2.8
$H^l \leftrightarrow H^o$	4.0	4.6	4.2
$H^m \leftrightarrow H^o$	3.0	2.8	2.9

[a] NOE distance upper bounds were determined from crosspeak intensities by calibration against fixed interproton distances (geminal protons and aromatic protons $H^w \leftrightarrow H^x$ and $H^y \leftrightarrow H^z$). Constraints were binned into 3.0, 4.0, and 5.0 Å categories.

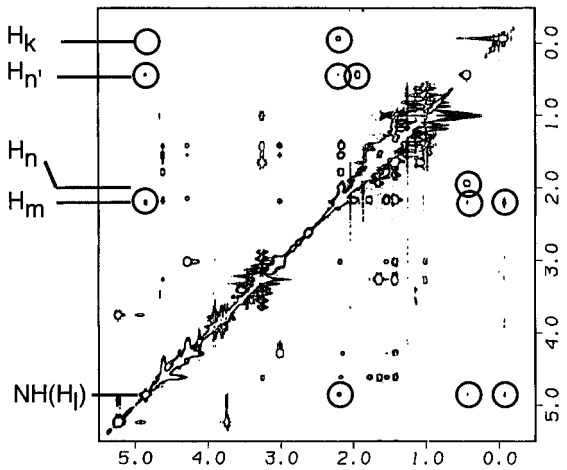


Figure 4. ROESY experiment (at 300 K) on 1:1 complex of **12** with *N*-Ac-L-Phe-CO_2^- showing intramolecular NOEs for the aliphatic protons of the guest substrate.

the most intense intramolecular NOE observed with a volume two orders of magnitude larger than that between the acetyl methyl (H^k) and the NH (H^l). The conformation of the L-phenylalanine derivative in the complex was examined by a systematic search (using SYBYL[23]) around the four rotatable bonds, imposing distance constraints obtained from NOE volumes. In each of the conformers which satisfied the NOE distance constraints, the acetyl amide bond adopts a *cis* configuration, driven largely by the short acetyl methyl (H^k) to $C^\alpha H$ (H^m) and aromatic (H^o, H^p, and H^q) distances. Inspection of the ROESY spectra for the 1:1 complex with the D-phenylalanine derivative showed intramolecular crosspeaks consistent with the expected *trans* amide and were essentially identical to the crosspeaks observed in the ROESY spectrum of uncomplexed tetrabutylammonium salt of *N*-acetyl phenylalanine. Again, inspection of the corresponding ROESY spectra for the 1:1 complex with the L- and D-alanine derivatives similarly conclusively showed that the L-amino acid guest is bound with the acetyl amide in the *cis* configuration, but the D-amino acid guest is bound with the acetyl amide in the *trans* configuration.

In 2D spectra of the complex between **12** and the L-phenylalanine derivative at 285 K, protons H^w, H^x, H^y, and H^z can be seen to be in exchange with a second set of resonances representing approximately 30% of the intensity of the major form (Figure 5).

(23) Tripos, Inc.: St. Louis, Missouri.

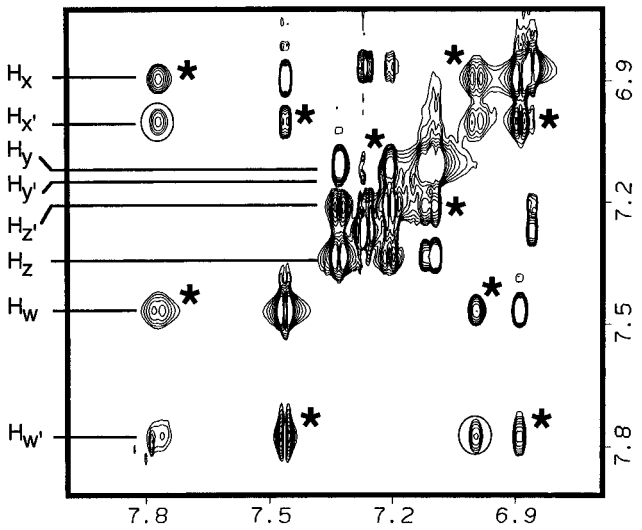


Figure 5. Exchange peaks assigned in a 100 ms TOCSY spectrum (at 285 K) of the 1:1 complex of **12** with *N*-Ac-L-Phe-CO_2^- at 285K. Starred (*) peaks denote exchange between the major and minor conformations of the host described in the text. Circled peaks denote through-bond correlations also observed in DQF-COSY experiments.

This second set of signals does not exhibit the intermolecular NOEs to the L-phenylalanine derivative, observed for the major form, and has chemical shifts comparable with those obtained for the unbound macrobicycle and for the macrobicycle bound externally (when bound externally the macrobicycle appears to adopt a conformation close to that in the unbound state, as evidenced by spectra of 1:1 complexes of the macrobicycle with the D-phenylalanine and D-alanine derivatives which shows minimal changes to the chemical shifts for the macrobicycle). Furthermore, the resolved aliphatic resonances for the L-phenylalanine derivative (H^k and $H^{n'}$) represent only 70% of the expected intensity compared with the signals for the tetrabutylammonium countercation, although no exchange partners for the L-phenylalanine derivative are evident in either the 1D or 2D spectra. One possible explanation for these results is that the L-phenylalanine derivative is associated with the macrobicycle both internally (~70%) *and* externally (~30%). The loss in intensity of the aliphatic resonances from the L-phenylalanine derivative could result from rearrangement of the guest, involving conversion of the *cis* amide to the *trans* form on transferring from internal to external binding, on a time scale which leads to exchange broadening.

Direct determination of the solution structures of the 1:1 complexes between macrobicycle **12** and L-amino acid substrates using the NMR data is nontrivial, primarily because the 2-fold symmetry of the macrobicycle is lost upon binding a chiral substrate, while the loss in symmetry is not reflected in the number of resonances observed in NMR spectra of the complex. Molecular modeling, however, has been used to probe the mode of binding of the L-phenylalanine derivative within the macrobicyclic cavity. The geometry of the host−guest complex was examined using a combination of simulated annealing calculations and free molecular-dynamics using the MacroModel program.[24] A modified version of the OPLS* force field[25] was adopted, and the effect of solvent was included through the use of GB/SA chloroform.[26] The L-phenylalanine derivative, with a *cis* amide bond, was docked into the macrobicycle by eye, such that the carboxylate group was bound to the thiourea moiety, in accord with experimental observations, and then energy minimized. Ten simulated annealing calculations of 1 ns involving slow cooling from 600 to ~0 K were performed, followed by a 5 ns simulation at 300 K to examine the behavior of the complex at room temperature. The free molecular-dynamics simulation was repeated with the distance between the acetyl amide proton and carboxylate oxygen of the guest restrained to 2 Å (this eliminates the unrealistically short contact-distance usually observed between this atom pair which results from the 1,4 van der Waals scale-factor used in the OPLS* force-field). The 10 minimum energy structures obtained from the simulated-annealing calculations all show similar binding geometries (Figure 6).

The structures differ in two respects. First, the hydrocarbon chains adjacent to the thiourea moiety adopt different conformations, and, second, the aromatic ring of the L-phenylalanine derivative is found to lie in two binding geometries over the aromatic rings of the macrobicycle but excluded from the main cavity. The L-phenylalanine derivative is bound to the macrobicycle by essentially a total of six hydrogen bonds. The acetyl amide carbonyl oxygen is simultaneously hydrogen bonded to two of the amide NH's of the macrobicycle, while the carboxylate function forms a total of four hydrogen bonds to the thiourea moiety and the other two amide NH's of the macrobicycle. Of particular note, the acetyl amide NH of the guest does not form any hydrogen bonds with the macrobicycle.

In the course of the free molecular-dynamics simulation of the host−guest complex (both with and without the 2 Å distance restraint between the amide proton and the carboxylate oxygen of the guest) the aromatic ring of the guest was observed to interconvert between the two binding geometries identified in the simulated annealing study but was observed to lie predominantly over the B rather than the A ring of the macrobicycle, in accord with the NMR data. The number of hydrogen bonds identified fluctuated between 4 and 6, with subtle changes in

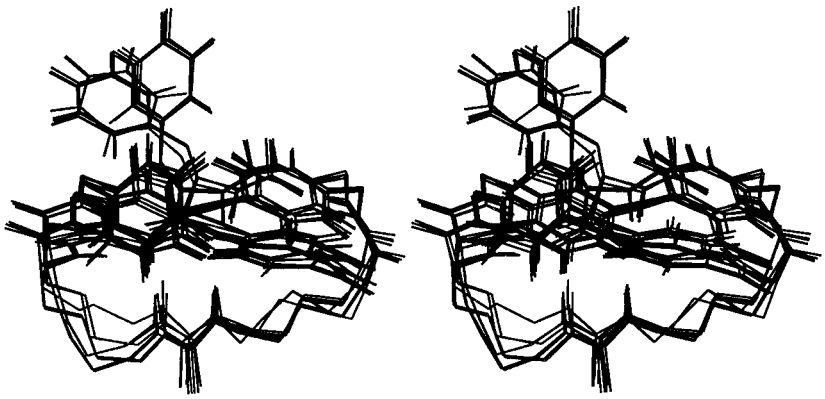


Figure 6. Stereopair of the 10 minimum energy structures of the complex of **12** with *N*-Ac-L-Phe-CO_2^- obtained from the simulated annealing calculations.

the hydrogen-bonding pattern being observed. However, the fundamental binding geometry of the macrobicycle remained unaffected. A detailed comparison with the NOE distance constraints was made by monitoring the appropriate H/H distances during the simulation (Tables 2 and 3). For atoms where a united-atom representation was used in the modeling, hydrogen atoms were added in standard geometries. Average distances were obtained through either r^{-6} averaging, or a combination of r^{-6} and r^{-3} averaging, as described in the Experimental Section. The agreement between simulation and experiment is generally good. In the unrestrained simulation, the molecular dynamics distances exceed the NMR values by greater than 0.5 Å in seven cases out of a total of 29. Four such violations exist in the restrained simulation. To determine whether these results are converged, the restrained simulation was repeated using a different starting configuration in which the aromatic ring of the L-phenylalanine derivative was in an alternative binding geometry. Analysis of this trajectory gave average distances that agreed with the previous simulation to within 0.2 Å and a total of only two violations with the experimental data. The agreement between simulation and experimental distances therefore supports the reliability of the proposed structural model. In particular, comparison of the $H^y \leftrightarrow H^o$ and $H^z \leftrightarrow H^o$ distances with the $H^w \leftrightarrow H^o$ and $H^x \leftrightarrow H^o$ distances (from molecular dynamics) clearly shows the close association of the aromatic ring of the guest with the B ring of the macrocycle. A representative structure of the host−guest complex is given in Figure 7, showing the preferred orientation of the L-Phe aromatic ring and the hydrogen-bond pattern.

In addition to satisfying the NOE distance constraints, the mode of binding derived from molecular modeling places the benzylic methylene of the guest in close proximity to the biaryl methane side wall, while the acetyl methyl group is shielded by both the biaryl methane side wall and the phenyl residue of the guest, consistent with the greater upfield shift for this acetyl methyl signal (−2.10 ppm) compared to that observed in the 1:1 complex between **12** and the L-alanine derivative (−1.02 ppm). The model for the complex also places the amino acid side chain directed away from the cavity of the macrobicycle and explains why there is little side chain discrimination for the amino acid derivatives, except perhaps for lysine where the longer chain may allow the amino group to reach around and establish a further hydrogen bonding interaction with the macrobicycle. The structure presented in Figure 7 is therefore consistent with the NMR data and in the absence of any further experimental data constitutes the best available model for this system.

Thus, D-amino acid substrates (D-alanine and D-phenylalanine derivatives) seem to bind predominantly on the outside of the macrobicycle cavity by a strong carboxylate−thiourea interaction (worth 23−27 kJ mol^{-1} of free binding energy).[27] The slight upfield shifts of the D-substrates observed in the 1:1 complexes might indicate that the substrates are bound within the cavity to a finite extent but not substantially. A short series of simulated annealing calculations suggest that the D-phenylalanine derivative binds to the thiourea through the side of the cavity, with hydrogen bonds from the carboxylate to the thiourea and to two of the amide NH's in the sidewall of the macrobicycle, forming a hydrogen bonding pattern similar to that already identified for the L-phenylalanine derivative. For the D-substrate, however, no hydrogen bonding to the amide group of the substrate is observed. The structure proposed for the binding of the D-phenylalanine derivative should be treated with caution, however, owing to the absence of corroborating experimental evidence (from NOE's, etc.). L-Amino acid substrates, on the other hand, bind predominantly on the inside of the cavity also establishing a strong carboxylate thiourea interaction, but with the acetyl amide in a *cis* configuration, as adjudged by the dramatic upfield shifts observed in the 1:1 complexes, and the strong inter- and intramolecular NOE interactions. The overall binding constants for the L- substrates, determined from extraction experiments, are only slightly greater than those measured for the D-substrates. Such a small energetic preference is consistent with the observation that the L-substrates occupy the cavity approximately 70% of the time and presumably bind on the outside of the cavity (as for the D-substrates) the other 30% of the time. The free energy difference between an amide in the *cis* and the *trans* configuration is estimated to be ∼10 kJ mol^{-1}.[2,3] This energetic penalty is paid for in the complex, at least in part, by two hydrogen bonds from the amide NH's in the sidewall of the macrobicycle to the acetyl carbonyl of the guest. In addition the carboxylate moiety of the guest appears to form a total of four hydrogen bonds to the thiourea NH's and the two remaining amide NH's in the side wall of the macrobicycle. This view of course neglects additional entropic costs on binding L-substrates in a much more ordered complex than that for the D-substrates and neglects compensating positive van der Waals interactions between the substrates and the biarylmethane units of the macrobicycle.

In conclusion we have observed the stabilization of a *cis* amide, in a host−guest complex which selectively internalizes L-amino acid derivatives. The use of NMR and molecular modeling has provided a detailed picture of the structure of the host−guest complex in solution and will now allow the rational design of selective receptors by incorporating structural features which block the binding of D-amino acid substrates on the exterior of the cavity.

Experimental Section

General Method for Elucidating Binding Constants Following Cram's Procedure.[16] Extraction Experiment in the Absence of Host. $CDCl_3$ (2.0 mL) was added to a sample of the substrate (*N*-acetyl amino acid tetrabutyl ammonium salt, $AA^-{\cdot}TBA^+$) (typically 20 mg) in D_2O (2.0 mL), and the biphasic mixture was thoroughly mixed using a vortex machine. After separation of the two layers, a 400 μL aliquot of the D_2O solution was removed by syringe and added to a 400 μL aliquot of a standard solution (0.01 M) dioxane in D_2O and a ^{1}H NMR spectrum of the resulting solution was recorded. Similarly, a 400 μL aliquot of the $CDCl_3$ solution was added to a 400 μL aliquot of dioxane (2.5×10^{-3} M) in $CDCl_3$, and a ^{1}H NMR spectrum of the resulting solution was recorded. The amount, and hence concentration, of the substrate in both the $CDCl_3$ and the D_2O phase was determined by comparison of the integrals of signals from dioxane and the substrate (using peaks from the tetrabutylammonium fragment). From these measurements the distribution constant (K_d−defined in eq 1) for the substrate was determined. Five independent experiments were performed for each substrate to give an average value for K_d and a standard deviation.

Extraction Experiment in the Presence of Host. An accurately weighed sample of macrobicycle **8** (typically 5 mg) in $CDCl_3$ (2.0 mL) was added to a sample of the substrate (*N*-acetyl amino acid tetrabutylammonium salt) (typically 20 mg) in D_2O (2.0 mL), and the biphasic mixture was thoroughly mixed using a vortex machine. After

(24) MacroModel V5.0; Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. *J. Comput. Chem.* **1990**, *11*, 440.

(25) (a) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657. (b) Jorgensen, W. L.; Nguyen, T. B. *J. Comput. Chem.* **1993**, *14*, 195. (c) Duffy, E. M.; Severance, D. L.; Jorgensen, W. L. *Isr. J. Chem.* **1993**, *33*, 323. (d) Jorgensen, W. L. *J. Phys. Chem.* **1986**, *90*, 6379.

(26) Still, W. C.; Tempczyk, A.; Hawley, R. C. *J. Am. Chem. Soc.* **1990**, *112*, 6127.

(27) Using an identical extraction procedure to that used for the tetrabutylammonium salts with the macrobicyclic, the binding constant between tetrabutylammonium benzoate and dibenzyl thiourea in $CDCl_3$ was estimated to be $(33.5 \pm 13.3) \times 10^3$ mol^{-1} ($-\Delta G_a = 25.4 \pm 1.3$ kJ mol^{-1}).

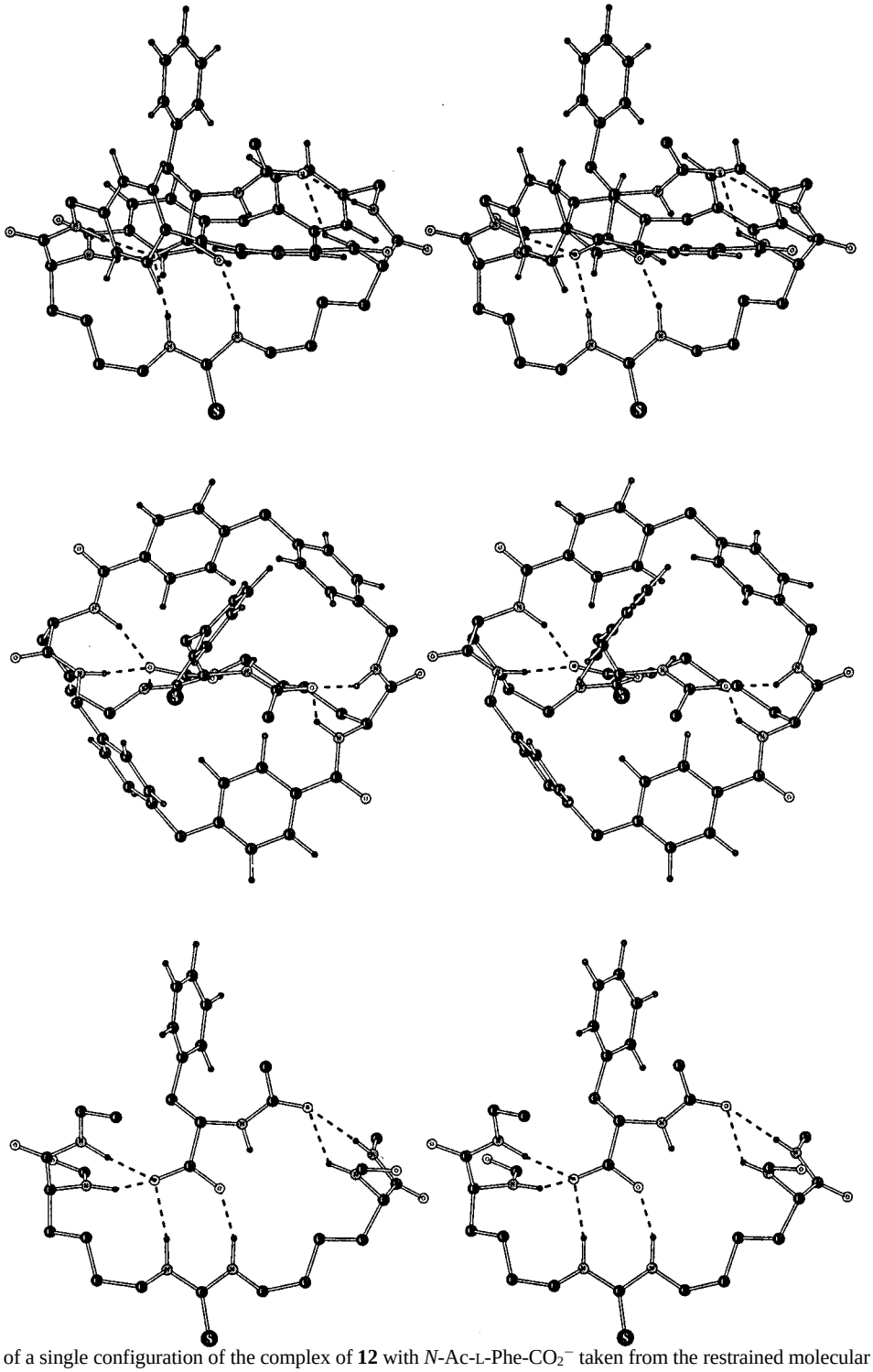


Figure 7. Stereopairs of a single configuration of the complex of **12** with *N*-Ac-L-Phe-CO_2^- taken from the restrained molecular dynamics simulation viewed (a) from the side, (b) from above, and (c) with some of the macrobicycle structure removed to emphasize the intermolecular hydrogen bonding network.

separation of the two layers, a 400 μL aliquot of the D_2O solution was removed by syringe and added to a 400 μL aliquot of a standard solution (5.0×10^{-3} M) of dioxane in D_2O, and a ^{1}H NMR spectrum of the resulting solution was recorded. Similarly, a 400 μL aliquot of the $CDCl_3$ solution was added to a 400 μL aliquot of dioxane (2.5×10^{-3} M) in $CDCl_3$, and a ^{1}H NMR spectrum of the resulting solution was recorded. The amount, and hence concentration, of the substrate in both the $CDCl_3$ and the D_2O phase was determined by comparison of the integrals of signals from dioxane and the substrate (using peaks from the tetrabutylammonium fragment). From these measurements the distribution constant (K_e—defined in eq 2) for the substrate was determined following the analysis of Cram.[16] Again, five independent experiments were performed for each substrate to give an average value for K_e and a standard deviation.

The association constant for the macrocycle (M) and amino acid substrates (K_a—eq 3) is then the ratio of K_d and K_e (eq 4).

$$[AA^-]_{D_2O} + [TBA^+]_{D_2O} \overset{K_d}{\rightleftharpoons} [AA^- \cdot TBA^+]_{CDCl_3} \quad (1)$$

$$[AA^-]_{D_2O} + [TBA^+]_{D_2O} + [M]_{CDCl_3} \overset{K_e}{\rightleftharpoons} [AA^- \cdot M \cdot TBA^+]_{CDCl_3} \quad (2)$$

$$[AA^- \cdot TBA^+]_{CDCl_3} + [M]_{CDCl_3} \overset{K_a}{\rightleftharpoons} [AA^- \cdot M \cdot TBA^+]_{CDCl_3} \quad (3)$$

$$K_a = K_e/K_d \quad (4)$$

The following assumptions are made in the above analysis: (a)

substrate salt is completely dissociated in D_2O; (b) the substrate salt is associated, but monomeric, in $CDCl_3$; (c) essentially no macrobicyclic receptor is distributed into the D_2O phase; (d) the macrobicycle forms a 1:1 complex with the amino acid substrates. Assumptions (a) and (b) were validated by measuring K_d across a range of initial concentrations (4−60 mM) of substrate in D_2O, which gave constant values for K_d, and by carrying out NMR dilution experiments for the substrates which indicated no appreciable dimerization or aggregation. Assumption (c) was validated since none of macrobicycle **12** could be seen in the ^1H NMR spectra of the D_2O solutions, and assumption (d) was validated by a Job plot[28] for the complexation of the L-phenylalanine derivative with macrobicycle **12**.

Molecular Modeling Protocol. The complex formed between macrobicycle **12** and the L-phenylalanine derivative was modeled using the MacroModel program[24] on a Silicon Graphics workstation. The OPLS* parameter set was selected for these studies. However, electrostatic and van der Waals parameters were assigned by comparison with the original OPLS papers.[25] The van der Waals and electrostatic parameters used in the study are given in the supporting information. Solvation was included in the simulations through the GB/SA continuum model of chloroform.[26] All possible nonbonded interactions were included in the calculations. The L-phenylalanine derivative, with a *cis* amide bond, was docked into the host by eye, such that the carboxylate group was bound to the thiourea moiety, in accord with experimental observations, and then energy minimized. To determine a range of plausible geometries for the complex, 10 successive simulated-annealing molecular-dynamics simulations were performed. Each simulation consisted of 1 ns of molecular dynamics, with slow cooling from an initial temperature of 600 K to a final temperature of ~0 K. A time-step of 1.5 fs was adopted, and all bonds were constrained using the SHAKE algorithm.[29] No restraints were applied to force the complex to adopt a predetermined geometry. The 10 structures obtained at the end of the simulated annealing calculations are energy minima. To characterize the behavior of the complex under experimental conditions, one of the structures produced by the annealing calculations was selected as the starting point for a 5 ns molecular-dynamics simulation. The simulation was performed at a temperature of 300 K in conjunction with a temperature-bath relaxation constant of 0.2 ps.[30] The OPLS* parameter set scales 1,4 van der Waals interactions to $^1/_8$th their full strength, and this results in the acetyl amide-proton carboxylate-oxygen distance being unrealistically short. Given the critical role of these hydrogen-bonding groups in mediating binding, the effect of this force-field artefact was investigated by repeating the 5 ns molecular dynamics simulation with a harmonic restraint of 2 Å, 150 kcal mol^{-1} $Å^{-2}$, applied to the amide-proton carboxylate-oxygen distance. Hydrogen bonds were assigned in the structures using the MacroModel default criteria. Furthermore, if a hydrogen bond between the acetyl amide-proton and carboxylate-oxygen was identified, it was discarded as being physically unrealistic. To compare the simulation results with the NMR distance constraints, hydrogen atoms were added to united-atom carbons in standard geometries. A 5 ns simulation is too short for the formally equivalent hydrogen atoms of the system to undergo full motional averaging, as occurs in the NMR experiment. Consequently, although the NMR results give a single distance constraint for a given atom pair, the simulation provides up to 12 distances to be monitored for a given pair (the three H_k protons with the four H_w protons, for example). In making the comparison with the NMR distance constraints, the simulation distances were averaged according to:

$$r_{av} = (\langle R_{i,j}^{-6} \rangle)^{-1/6}$$

where r_{av} is the final averaged simulation distance for pairs i,j; $R_{i,j}$ is the instantaneous value of the interatomic separation of atom pairs i, j; and $\langle ... \rangle$ indicates that an average is calculated over the course of the molecular dynamics trajectory. The coordinates used in the trajectory analysis were generated approximately every 1 ps. In the case of the H_k protons, however, a single r^{-3} averaged distance was evaluated for the interactions with other protons for each frame of the trajectory. These distances were then averaged using the r^{-6} procedure already-described. This hybrid protocol is necessary since the H_k protons are likely to be spinning more rapidly than every 200 ps.

The choice of r^{-6} over r^{-3} averaging is based on the time scales for molecular motions within the system. From the Debye expression[31] for the rotational correlation time

$$\tau_{rot} = 4\pi a^3 \eta / 3kT$$

τ_{rot} for the host−guest system is estimated to be approximately 200 ps; a, the molecular radius, is estimated to be 7.5 Å; η, the solvent viscosity, is 0.542×10^{-3} kg m^{-1} s^{-1} at 25 °C;[32] k is the Boltzmann constant; and T is the temperature. Inspection of the molecular dynamics trajectories indicates that the macrocycle aromatic rings do not flip in 5 ns. The host aromatic ring is observed to flip approximately every 0.5−1.0 ns. However the influence of solvent viscosity was not included in the calculations, and this can only serve to increase the time over which ring-flips occur. Consequently molecular tumbling is occurring faster than internal conformational transitions, and r^{-6} averaging is therefore appropriate.[33] The exception to this is for the methyl H_k protons for which a hybrid averaging protocol needs to be adopted as noted above.

Acknowledgment. We are grateful to S. Thomas and N. Aboitiz (MSD) for assistance acquiring NMR spectra and to MSD and EPSRC for a studentship to G.J.P. J.W.E. is a Royal Society University Research Fellow. J.W.E. would like to thank Dr. A. Bonvin (Yale University) and D. C. Lim (Yale University) for helpful discussions and assistance with the computational studies.

Supporting Information Available: Details of acquisition of NMR data, full experimental details for the synthesis and characterization of compounds **2** and **4**−**12**, and nonbonded parameters (charge, σ, ϵ) for macrobicycle **12** and tetrabutylammonium *N*-acetyl L-phenylalanine, used in the molecular modeling (12 pages). See any current masthead page for ordering and Internet access instructions.

JA9607805

(28) (a) Connors, K. A. *Binding Constants, The Measurement of Molecular Complex Stability*; Wiley: 1987; pp 24−28. (b) Job, A. *Annales de Chimie* **1928**, *9*, 113.

(29) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.

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

(31) Atkins, P. W. *Physical Chemistry*, 3rd ed.; OUP: Oxford, 1986; p 621.

(32) Weast, R. C. *Handbook of Chemistry and Physics*; CRC Press: Boca Raton, FL, 1980.

(33) Tropp, J. *J. Chem. Phys.* **1980**, *72*, 6035.