Determination of the Stable Microstates of a Peptide from NOE Distance Constraints and Optimization of Atomic Solvation Parameters

Canan Baysal and Hagai Meirovitch*

Contribution from the Supercomputer Computations Research Institute, Florida State University, Tallahassee, Florida 32306

Received September 5, 1997

Abstract: A methodology for analyzing nuclear Overhauser effect (NOE) data of interconverting microstates of a peptide has been suggested recently, which is based on pure statistical mechanical considerations. Thus, the most stable microstates and their populations are determined from the free energies. The success of this approach depends on the existence of a reliable potential energy function for the *solvated* peptide, in which the solvent is treated *implicitly*. Such a potential is developed here based on the stable structures of the cyclic hexapeptide *cyclo*(D-Pro[1]-Phe[2]-Ala[3]-Ser[4]-Phe[5]-Phe[6]) in DMSO obtained by Kessler et al. (*J. Am. Chem. Soc.* **1992**, *114*, 4805−4818) from NOE distance constraints. This study suggests that two different backbone motifs coexist in equilibrium, one with a βI turn and the other with a βII turn around Ser[4]-Phe[5]. We have first reconfirmed these findings by a best-fit analysis applied to a large set of energy-minimized structures generated by our "local torsional deformations" (LTD) method, using the GROMOS force field with and without NOE distance restraints. However, the GROMOS energy E_{GRO}, which excludes solvent interactions was found inappropriate to describe this system because the lowest energy structures representing the βI and βII motifs are ∼15 and 5 kcal/mol above the global energy minimum, respectively. Solvent effects are taken into account through $E_{tot} = E_{GRO} + \sum A_i\sigma_i$, where A_i is the solvent accessible surface area (SASA) of atom i and σ_i is the atomic solvation parameter (ASP). We optimize the ASPs for DMSO by requiring that the E_{tot} values of βI and βII structures become the lowest *globally*; this is verified by an extensive application of LTD. The set of ASPs obtained here will be refined in the next work where free energy (rather than energy) considerations will be taken into account. This solvation model, which is relatively easy to handle, requires significantly less computer time than explicit models of solvation and can readily be used in structural analysis of experimental data using GROMOS. The proposed derivation opens the way for the development of similar solvation models for peptides in other solvents. ASPs for proteins in water can be obtained by applying our methodology to surface loops in proteins. Preliminary results for the ASPs, which are slightly different from the present values, were published in a recent Letter (*J. Phys. Chem. B* **1997**, *101*, 7368−7370).

Introduction

Multidimensional nuclear magnetic resonance (NMR) is the only physical technique which enables one to determine the dynamic structure of biomolecules in solution.[1−4] It is well established for globular proteins, which reside in a *single* microstate, i.e., a well-defined region of conformational space. On the other hand, peptides in most cases are random coils, but under certain solvent conditions may generate medium and long-range nuclear Overhauser enhancement (NOE) intensities. However, in some cases the latter can only be interpreted for a molecule that populates *several* microstates in thermodynamic equilibrium, i.e., a molecule with *intermediate flexibility*. The quantitative interpretation of this *average* NOE effect is complex, since it requires identifying the dominant microstates and determining their relative populations.

The sophistication level of the analysis of intermediate flexibility depends on the quality of the NOE data. For relatively low quality, as for small peptides in water, only a qualitative interpretation of multiconformational equilibria is possible.[5,6] When the number of NOEs is relatively large and J coupling constants are available, as is typically the case for cyclic peptides, different conformations that best fit the data can be identified by molecular modeling which mostly involves molecular dynamics (MD) simulations.[7−14] A more sophisticated technique based on MD simulations with time-averaged restraints, suggested by Torda et al.,[15] has been used extensively.[16−21] An alternative procedure called MEDUSA was

* To whom correspondence should be addressed.

(1) Wüthrich, K. *NMR of Proteins and Nucleic Acids*; John Wiley and Sons: New York, 1986.
(2) Clore, G. M.; Gronenborn, A. M. *Annu. Rev. Biophys. Chem.* **1991**, *20*, 29−63.
(3) Bax, A.; Grzesiek, S. *Acc. Chem. Res.* **1993**, *26*, 131−138.
(4) Oschkinat, H.; Müller, T.; Diekmann, T. *Angew. Chem., Int. Ed. Engl.* **1994**, *33*, 277−293.
(5) Dyson, H. J.; Rance, M.; Houghten, R. A.; Lerner, R. A.; Wright, P. E. *J. Mol. Biol.* **1988**, *201*, 161−200.
(6) Stradley, S. J.; Rizo, J.; Bruch, M. D.; Stroup, A. N.; Gierasch, L. M. *Biopolymers* **1990**, *29*, 263−287.
(7) Fesik, S. W.; O'Donnell, T. J.; Gampe, R. T., Jr.; Olejniczak, E. T. *J. Am. Chem. Soc.* **1986**, *108*, 3165−3170.
(8) Kessler, H.; Griesinger, C.; Lautz, J.; Müller, A.; van Gunsteren, W. F.; Berendsen, H. J. C. *J. Am. Chem. Soc.* **1988**, *110*, 3393−3396.
(9) Kim, Y.; Prestegard, J. H. *Biochemistry* **1989**, *28*, 8792−8797.
(10) Kim, Y.; Prestegard, J. H. *Proteins* **1990**, *8*, 377−385.
(11) Džakula, Ž.; Westler, W. M.; Edison, A. S.; Markley, J. R. *J. Am. Chem. Soc.* **1992**, *114*, 6195−6199.

S0002-7863(97)03124-7 CCC: $15.00
Published on Web 02/04/1998

proposed by Brüschweiler and Ernst.[22,23] The underlying hypothesis is that individual conformations may violate some of the NOE distance restraints, which are fulfilled only by the entire dynamic set of substates. Only pairs of exchanging conformations are considered, and the best combinations in terms of structural similarity are delineated.[23] A few additional approaches which treat the problem of conformational multiplicity explicitly, attempting to calculate the populations as well, have been developed.[24−34] In general, a set of reasonable conformations are generated using any of the currently available force fields. Conformations are retained if their energies do not exceed an arbitrarily set threshold, they differ structurally, and are consistent with a subset of the NMR restraints. The main disadvantage of these methods is the arbitrariness inherent in the selection of the conformations, which turn the populations into *fitting parameters*, rather than *thermodynamic variables*.

Another methodology for treating ensembles of interconverting conformations was proposed recently by Meirovitch et al.[35−38] This approach is based on pure statistical mechanical considerations, hence for a perfect force field it actually becomes free from the arbitrariness inherent in the "best-fit" methods. In practice, one has to use approximations, which can, however, be improved *systematically*. This approach was applied previously only to a linear peptide and is extended here to cyclic peptides. Below, we first discuss this methodology as carried out under *ideal* conditions, i.e., assuming that the solvated peptide can be described by a usual force field, such as ECEPP[39,40] or GROMOS.[41]

Typically, the energy surface of a peptide has a huge number of local energy minima; the potential energy wells in conformational space around these minima were called "localized microstates".[35,36] Methods based on the harmonic entropy were developed, which enable one to determine the relative contribution of localized microstates to the partition function Z. Application of these methods to Leu-enkephalin (H-Tyr-Gly-Gly-Phe-Leu-OH) described by ECEPP has led to the conclusion that the partition function can be divided into two parts, $Z = Z_a + Z_b$. At 270 K, Z_a contains the relatively small number of the lowest energy (and free energy) localized microstates, and provides the dominant contribution to Z. The probability of each localized microstate is relatively large, and its contribution to medium and long-range NOE intensities is significant. On the other hand, not only is Z_b smaller than Z_a, but the number of localized microstates associated with it is extremely large, which means that their individual probability is small, and their contribution to NOE intensities is expected to be negligible. This is because the conformations of higher energy, i.e., those of Z_b, display a much larger structural variety than those which belong to Z_a. Thus, they can be viewed approximately as an ensemble of random coil structures and are not expected to contribute to the long-range NOEs. It was found that microstates which pertain to the 2 and 3 kcal/mol range above the global energy minimum (GEM) contribute ∼0.60 and ∼0.70 of the total partition function, respectively.[35] Similar populations were obtained for cycloheptadecane described by the MM3 force field[42] and they are probably typical for molecules of a similar size. However, this assumption that is adopted throughout the present paper should still be verified by further studies.

To reproduce the experimental results (assuming a perfect force field), one has to carry out a detailed study of the most stable localized microstates, i.e., those that pertain to Z_a. However, as was found for Leu-enkephalin,[35,38] the number of energy-minimized structures within a 2 kcal/mol range above the GEM is already large, although many of them are similar. The molecule is expected to visit each of them for a very short time while staying for longer periods within a "wide microstate" which is a larger potential energy well consisting of many similar localized microstates.[43,44] Consequently, one is interested in the large conformational transitions, i.e., those between the most stable wide microstates. Each of these is expected to contain some of the most stable localized microstates with minimized energy within 2 kcal/mol above the GEM. One can carry out an extensive conformational search within this energy range and select a limited set of energy-minimized structures that are *significantly* different as representatives of the most stable wide microstates.

These representative structures become "seeds" for Monte Carlo (MC) or MD simulations that span the related wide microstates (notice that these samples were called MC microstates in refs 36 and 37). Two structures were considered to

(12) Džakula, Ž.; Edison, A. S.; Westler, W. M.; Markley, J. R. *J. Am. Chem. Soc.* **1992**, *114*, 6200−6207.
(13) Bogusky, M. J.; Naylor, A. M.; Pitzenberger, S. M.; Nutt, R. F.; Brady, S. F.; Colton, C. D.; Sisko, J. T.; Veber, P. S. A. D. F. *Int. J. Peptide Res.* **1992**, *39*, 63−76.
(14) Bean, J. W.; Kopple, K. D.; Peishoff, C. E. *J. Am. Chem. Soc.* **1992**, *114*, 5328−5334.
(15) Torda, A. E.; Scheek, R. M.; van Gunsteren, W. F. *J. Mol. Biol.* **1990**, *214*, 223−235.
(16) Kessler, H.; Matter, H.; Gemmecker, G.; Kottenhahn, M.; Bats, J. W. *J. Am. Chem. Soc.* **1992**, *114*, 4805−4818.
(17) Mierke, D. F.; Kurz, M.; Kessler, H. *J. Am. Chem. Soc.* **1994**, *116*, 1042−1049.
(18) Bonvin, A. M. J. J.; Boelens, R.; Kaptein, R. *J. Biomol. NMR* **1994**, *4*, 143−149.
(19) Mierke, D. F.; Scheek, R. M.; Kessler, H. *Biopolymers* **1994**, *34*, 559−563.
(20) Pearlman, D. A. *J. Biomol. NMR* **1994**, *4*, 1−16.
(21) Pearlman, D. A. *J. Biomol. NMR* **1994**, *4*, 279−299.
(22) Brüschweiler, R.; Blackledge, M.; Ernst, R. R. *J. Biomol. NMR* **1991**, *1*, 3−11.
(23) Blackledge, M. J.; Brüschweiler, R.; Griesinger, C.; Schmidt, J. M.; Xu, P.; Ernst, R. R. *Biochemistry* **1993**, *32*, 10960−10974.
(24) Betins, J.; Nikiforovich, G. V.; Chipens, G. *J. Mol. Struct. (THEOCHEM)* **1986**, *137*, 129−132.
(25) Vesterman, B.; Saulitis, J.; Betins, J.; Liepins, E.; Nikiforovich, G. V. *Biochim. Biophys. Acta* **1989**, *998*, 204−209.
(26) Nikiforovich, G. V.; Prakash, O.; Gehrig, C. A.; Hruby, V. J. *J. Am. Chem. Soc.* **1993**, *115*, 3399−3406.
(27) Landis, C.; Allured, V. S. *J. Am. Chem. Soc.* **1991**, *113*, 9493−9499.
(28) Landis, C. R.; Luck, L. L.; Wright, J. M. *J. Magn. Res.* **1995**, *109*, 44−59.
(29) Yang, J.; Havel, T. F. *J. Biomol. NMR* **1993**, *3*, 355−360.
(30) Cicero, D. O.; Barbato, G.; Bazzo, R. *J. Am. Chem. Soc.* **1995**, *117*, 1027−1033.
(31) Bonvin, A. M. J. J.; Brünger, A. T. *J. Mol. Biol.* **1995**, *250*, 80−93.
(32) Bonvin, A. M. J. J.; Brünger, A. T. *J. Biomol. NMR* **1996**, *7*, 72−76.
(33) Forster, M. J.; Mulloy, B. *J. Comput. Chem.* **1994**, *15*, 155−161.
(34) Ulyanov, N. B.; Schmitz, U.; Kumar, A.; James, T. L. *Biophys. J.* **1995**, *68*, 13−24.
(35) Meirovitch, H.; Meirovitch, E.; Lee, J. *J. Phys. Chem.* **1995**, *99*, 4847−4854.
(36) Meirovitch, E.; Meirovitch, H. *Biopolymers* **1996**, *38*, 69−88.
(37) Meirovitch, H.; Meirovitch, E. *J. Phys. Chem.* **1996**, *100*, 5123−5133.
(38) Meirovitch, H.; Meirovitch, E. *J. Comput. Chem.* **1997**, *18*, 240−253.
(39) Momany, F. A.; McGuire, R. F.; Burgess, A. W.; Scheraga, H. A. *J. Phys. Chem.* **1975**, *79*, 2361−2381.
(40) Sippl, M. J.; Némethy, G.; Scheraga, H. A. *J. Phys. Chem.* **1984**, *88*, 6231−6233.
(41) van Gunsteren, W. F.; Berendsen, H. J. C. Groningen Molecular Simulation (GROMOS) Library Manual; Biomos, Nijenborgh 16 9747 AG Groningen NL, 1987.
(42) Saunders, M. Private communication.
(43) Elber, R.; Karplus, M. *Science* **1987**, *235*, 318−321.
(44) Stillinger, F. H.; Weber, T. A. *Science* **1990**, *225*, 983−989.

be significantly variant if at least one dihedral angle differs by 60° or more. To obtain the relative populations of the wide microstates, it is necessary to calculate their free energy from the related MC or MD samples, which can be achieved by the local states (LS) method[45−48] (the advantage of this method over other existing techniques is discussed in refs 36 and 37). The overall NOEs are then obtained as averages over the NOEs of these samples, weighted by the LS populations. This averaging was based on the model for the motional state of the molecule proposed by Kessler et al.[8] This model consists of the "initial rate approximation", the assumptions that intramicrostate conformational exchange is much faster, whereas intermicrostate exchange is much slower than the overall rotational reorientation and that the various types of motions are uncorrelated. This theory is applicable provided that the rotational correlation time is longer than the inverse Larmor frequency. Note that *angular* modulations of the internuclear vectors are being ignored in this model (see detailed discussion in ref 37). The same approximation is adopted in the present paper.

Obviously, the success of this approach depends on the extent of the conformational search carried out, and on the quality of the force field, in particular its ability to describe solvent effects. Ideally, the latter should be taken into account by using explicit models.[49,50] However, the required simulations are very time-consuming and an efficient method for calculating the difference in the free energy between significantly different microstates is not available. Therefore, one has to resort to more approximate mean field continuum models, which consider the solvent implicitly. Thus, the *electrostatic* free energy of solvation for water is frequently calculated from a solution of the Poisson−Boltzmann equation[49−52] or by applying approximations to this equation;[53−55] the Langevin model of Warshel and Levitt also belong to this category.[56] The *hydrophobic* free energy, on the other hand, has been taken into account as a product of the solvent-accessible surface area (SASA) of the whole molecule and a solvation parameter obtained from free energy of transfer data of hydrocarbons from the gas phase to water.[49,50,53−55,57,58] To obtain the total free energy, the free energies of solvation calculated by either approach are added to the intramolecular interaction energy defined by the usual force field. However, implementation of these semiempirical models[59] has been found to be still relatively time-consuming.[53,54,57,60,61]

More approximate fully empirical[59] solvation models for water were also developed, in which the above electrostatic and hydrophobic energies are replaced by

$$E_{\mathrm{sol}} = \sum_i \sigma_i A_i \qquad (1)$$

where the summation is over all the atoms i. A_i is the SASA, and σ_i is the atomic solvation parameter (ASP) of atom i. Several sets of ASPs for water have been derived from thermodynamic data of small molecules.[58,62−74] This approach is feasible due to the recent development of efficient methods for calculating the SASA and its derivatives.[75−80] In particular we mention the approximate hence very efficient method of Hasel et al.,[81] the efficient method of Sander and collaborators,[65,67,82,83] which is based on accessible volume rather than surface calculations, and the related method of Augspurger and Scheraga.[84] Although these fully empirical solvation models are approximate,[58] in most cases they were found to outperform the usual force fields, leading to relatively strong correlations between the (minimized) energy of a conformation and its root-mean-square deviation from the X-ray structure.[58,62−72,76,80] On the other hand, some of the sets of ASPs were found to lead to unsatisfactory results;[68,69,85,86] Juffer et al. studied various sets of ASPs and discussed their quality.[87]

In this paper, our methodology is extended to cyclic peptides. As a first step, we developed a procedure for conformational search of cyclic molecules, called the local torsional deformations (LTD) method, which enables efficient generation of a large set of low-energy-minimized structures; LTD was initially applied to cycloundecane[88] and cycloheptadecane[89] modeled by

(45) Meirovitch, H. *Chem. Phys. Lett.* **1977**, *45*, 389−392.
(46) Meirovitch, H. *Phys. Rev. A* **1985**, *32*, 3709−3715.
(47) Meirovitch, H.; Vásquez, M.; Scheraga, H. A. *Biopolymers* **1987**, *26*, 651−671.
(48) Meirovitch, H.; Koerber, S. C.; Rivier, J.; Hagler, A. T. *Biopolymers* **1994**, *34*, 815−839.
(49) Smith, P. E.; Pettitt, B. M. *J. Phys. Chem.* **1994**, *98*, 9700−9711.
(50) van Gunsteren, W. F.; Luque, F. J.; Timms, D.; Torda, A. E. *Annu. Rev. Biophys. Biomol. Struct.* **1994**, *23*, 847−863.
(51) Sharp, K. A.; Honig, B. *Annu. Rev. Biophys. Chem.* **1990**, *19*, 301−332.
(52) Davis, M. E.; McCammon, J. A. *Chem. Rev.* **1990**, *90*, 509−521.
(53) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.
(54) Abagyan, R.; Totrov, M. *J. Mol. Biol.* **1994**, *235*, 983−1002.
(55) Purisima, E. O.; Nilar, S. H. *J. Comput. Chem.* **1995**, *16*, 681−689.
(56) Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227−249.
(57) Humphreys, D. D.; Friesner, R. A.; Berne, B. J. *J. Phys. Chem.* **1995**, *99*, 10674−10685.
(58) Smith, K. C.; Honig, B. *Proteins* **1994**, *18*, 119−132.
(59) Simonson, T.; Brünger, A. T. *J. Phys. Chem.* **1994**, *98*, 4683−4694.
(60) Sharp, K. *J. Comput. Chem.* **1991**, *12*, 454−468.
(61) Gilson, M. K.; Davis, M. E.; Luty, B. A.; McCammon, J. A. *J. Phys. Chem.* **1993**, *97*, 3591−3600.
(62) Eisenberg, D.; McLachlan, A. D. *Nature (London)* **1986**, *319*, 199−203.
(63) Wesson, L.; Eisenberg, D. *Protein Sci.* **1992**, *1*, 227−235.
(64) Ooi, T.; Oobatake, M.; Némethy, G.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 3086−3090.
(65) Stouten, P. F. W.; Frömmel, C.; Nakamura, H.; Sander, C. *Mol. Simul.* **1993**, *10*, 97−120.
(66) Vila, J.; Williams, R. L.; Vásquez, M.; Scheraga, H. A. *Proteins* **1991**, *10*, 199−218.
(67) Luty, B. A.; Wasserman, Z. R.; Stouten, P. F. W.; Hodge, C. N.; ZaCharias, M.; McCammon, J. A. *J. Comput. Chem.* **1995**, *16*, 454−464.
(68) Schiffer, C. A.; Caldwell, J. W.; Stroud, R. M.; Kollman, P. A. *Protein Sci.* **1992**, *1*, 396−400.
(69) Schiffer, C. A.; Caldwell, J. W.; Kollman, P. A.; Stroud, R. M. *Mol. Simul.* **1993**, *10*, 121−150.
(70) von Freyberg, B.; Braun, W. *J. Comput. Chem.* **1991**, *12*, 1065−1076.
(71) von Freyberg, B.; Richmond, T. J.; Braun, W. *J. Mol. Biol.* **1993**, *233*, 275−292.
(72) Williams, R. L.; Vila, J.; Perrot, G.; Scheraga, H. A. *Proteins* **1992**, *14*, 110−119.
(73) Kang, Y. K.; Gibson, K. D.; Némethy, G.; Scheraga, H. A. *J. Phys. Chem.* **1988**, *92*, 4739−4745.
(74) Eisenhaber, F. *Protein Sci.* **1996**, *5*, 1676−1686.
(75) Richmond, T. J. *J. Mol. Biol.* **1984**, *178*, 63−89.
(76) von Freyberg, B.; Braun, W. *J. Comput. Chem.* **1993**, *14*, 510−521.
(77) Perrot, G.; Cheng, B.; Gibson, K. D.; Villa, J.; Palmer, K. A.; Nayeem, A.; Maigret, B.; Scheraga, H. A. *J. Comput. Chem.* **1992**, *13*, 1−11.
(78) Eisenhaber, F.; Argos, P. *J. Comput. Chem.* **1993**, *14*, 1272−1280.
(79) Sridharan, S.; Nichols, A.; Sharp, K. A. *J. Comput. Chem.* **1996**, *15*, 1038−1044.
(80) Mumenthaler, C.; Braun, W. *J. Mol. Mod.* **1995**, *1*, 1−10.
(81) Hasel, W.; Hendrickson, T. F.; Still, W. C. *Tetrahedron Comput. Methods* **1988**, *1*, 103−116.
(82) Colonna-Cesari, F.; Sander, C. *Biophys. J.* **1990**, *57*, 1103−1107.
(83) Holm, L.; Sander, C. *J. Mol. Biol.* **1992**, *225*, 93−105.
(84) Augspurger, J. D.; Scheraga, H. A. *J. Comput. Chem.* **1996**, *17*, 1549−1558.
(85) van Aalten, D. M. F.; Amadei, A.; Bywater, R.; Findlay, J. B. C.; Berendsen, H. J. C.; Sander, C.; Stouten, P. F. W. *Biophys. J.* **1996**, *70*, 684−692.
(86) Fraternali, F.; van Gunsteren, W. F. *J. Mol. Biol.* **1996**, *256*, 939−948.
(87) Juffer, A. H.; Eisenhaber, F.; Hubbard, S. J.; Walther, D.; Argos, P. *Protein Sci.* **1995**, *4*, 2499−2509.
(88) Baysal, C.; Meirovitch, H. *J. Chem. Phys.* **1996**, *105*, 7868−7871.

the MM2 force field.[90] Having this tool at hand, we treat here the hexapeptide *cyclo*(D-Pro[1]-Phe[2]-Ala[3]-Ser[4]-Phe[5]-Phe[6]) whose X-ray and solution structures in DMSO were obtained from NMR experiments by Kessler and co-workers.[16] They obtained 31 proton−proton distances and several ^3J coupling constants, which give information on backbone and side chain dihedral angles, but could not find a *single* conformation that fits these data. However, the data were explained satisfactorily by two MD samples based on the GROMOS force field.[41] They describe two backbone motifs with the same βII′ turn around D-Pro[1] and Phe,[2] but different turns (βI and βII) around Ser[4] and Phe[5]. We refer to them as the βI and βII motifs (rather than structures) due to the fact that they represent a relatively large backbone and side chain conformational flexibility; in other words, each of them can be viewed as a wide microstate to which many energy minimized structures belong. This flexibility already arises from the definition of the βII′, βI, and βII turns, which requires the ϕ and ψ angles of two consecutive residues to be within ±30° around the central values (60°, −120°, −80°, 0°), (−60°, −30°, −90°, 0°), or (−60°, 120°, 80°, 0°), respectively.[91] Kessler et al. did not calculate the relative populations of these motifs and our long-range goal is to obtain them from statistical mechanical considerations.

To be consistent with the analysis of Kessler et al., we implemented LTD within the GROMOS package and carried out an extensive conformational search for the cyclohexapeptide in order to find the GEM(E_{GRO}) (E_{GRO} is the GROMOS energy) and the lowest energy minimized structures that belong to the βI and βII motifs. We found that the energies of the latter structures are, respectively, ∼15 and 5 kcal/mol above the GEM(E_{GRO}), which means that the GROMOS force field alone is inappropriate to model this molecule in DMSO (because the correct energies of these structures are expected to be within 2 kcal/mol above the GEM). In other words, E_{GRO} alone will not lead to the correct free energies of these motifs which determine their relative populations. Therefore, an important objective of this paper is to develop an implicit fully empirical solvation model (eq 1) for the cyclohexapeptide in DMSO, which is carried out in several steps. First we have sought to verify the previous conclusions of Kessler et al. that the molecule coexists in the βI and βII motifs. For that reason we generated a large set of energy-minimized structures with LTD, using E_{GRO} together with a penalty potential that forces the structure to satisfy the NOE distance constraints. Structures of the set that satisfy a certain number of the constraints are retained and further analyzed in order to find the smallest group of them that best fit the data. This study has verified the above finding of Kessler et al., with a possibility for the involvement of a third structure. We also tested and improved an approximate but very efficient method for calculating the SASA, proposed by Hasel et al.[81]

The most challenging part is the derivation of the ASPs (eq 1). In most cases the ASPs for water have been obtained from the free energy of transfer of small molecules from the gas phase to water, and E_{sol} was added to the force field energy without further calibration. However, Schiffer et al.,[68,69] and more recently Fraternali and van Gunsteren,[86] checked several sets of ASPs in MD simulations and found that they should be properly scaled and sometimes changed in order to recover experimental structural data of proteins, and results obtained by MD simulations with explicit water.

We propose to describe the system energy by $E_{tot} = E_{GRO} + E_{sol}$ and derive the optimal ASPs, from the sole requirement that the free energies of the experimental motifs will be close to each other, and at least one of them is the global minimum. In this work, however, we consider only energetic effects and require that the energies of the structures representing the experimental motifs are as low as possible pertaining to the 2 kcal/mol range above the GEM(E_{tot}). This is achieved by an extensive conformational search using LTD. Entropic effects will be taken into account in a following study, where the optimal ASPs will be used in MD or MC simulations of the βI, βII, and other stable motifs. The free energies, which lead to the relative populations, will be obtained by the LS method applied to the corresponding trajectories. This might necessitate a further refinement of the ASPs. Preliminary results of the ASPs that slightly differ from the values reported here appeared in a recent Letter.[92]

Theory and Methods

Molecular Models and General Theoretical Considerations. The intramolecular interactions of *cyclo*(D-Pro[1]-Phe[2]-Ala[3]-Ser[4]-Phe[5]-Phe[6]) are described by the GROMOS 37D4 united atom force field, which define the molecular energy E_{GRO}.[41] These interactions include harmonic bond stretching and bond angle bending potentials, proper and improper torsional potentials, and nonbonded electrostatic and 6-12 Lennard-Jones (LJ) interactions; thus, the molecule is considered fully flexible. Also, the dielectric constant is $\epsilon = 1$ and no distance truncation on nonbonded interactions is applied. In GROMOS the hydrogen atoms are treated as collapsed on their first neighboring atoms except for hydrogens bonded to a nitrogen or the oxygen of Ser. The cyclohexapeptide therefore consists of $N = 57$ explicit atoms.

To derive the approximate model for the solvated peptide, assume first that the peptide is immersed in a large "box" of explicit DMSO molecules, and that the *exact* interaction energy of the system E is known; it consists of peptide−peptide, peptide−solvent, and solvent−solvent interactions. To eliminate the detailed effect of the solvent one can define a "partition function of mean force" $Z(\mathbf{x})$ by integrating the exponent $\exp[-E/k_BT]$ over the solvent coordinates only, for each set of peptide coordinates $\mathbf{x}$ ($\mathbf{x}$ is a $3N$ vector of the Cartesian coordinates, k_B and T are the Boltzmann constant and the absolute temperature, respectively). $Z(\mathbf{x})$ leads to the exact potential of mean force, $E_{exa}(\mathbf{x})$, where $E_{exa}(\mathbf{x}) = k_BT \ln Z(\mathbf{x})$. Note that although $E_{exa}(\mathbf{x})$ is a free energy function depending on T, it will be referred to as energy throughout this paper.

$E_{exa}(\mathbf{x})$ lacks the microscopic information about the solvent molecules, but it *correctly* describes the stable regions on the molecule. This energy surface, as that defined by a usual force field, is typically "decorated" by a tremendous number of localized microstates centered around local minima and wide microstates, Ω_j, which consist of structurally similar localized ones. The most stable Ω_j is the one with the largest contribution, Z_j, to the total partition function of the molecule

$$Z_j = \int_{\Omega_j} \exp[-E_{exa}(\mathbf{x})/k_BT]\, d\mathbf{x} \qquad (2)$$

or equivalently with the lowest Helmholtz free energy $F_j = -k_BTZ_j$. If several wide microstates have comparable lowest free energy, the molecule will coexist in all of them in thermodynamic equilibrium. Each of these microstates is expected to contain several of the low-energy-minimized structures with energy within 2 kcal/mol above GEM(E_{exa}); therefore if E_{exa} was known, the most stable wide microstates could in principle be identified by an extensive conformational search for low-energy structures. The free energies, hence populations, of these Ω_i could then be obtained from MD or MC trajectories using the LS method. Evidently, when the most stable Ω_j

(89) Baysal, C.; Meirovitch, H. *J. Phys. Chem. A* **1997**, *101*, 2185−2191.
(90) Burket, U.; Allinger, N. *Molecular Mechanics*; American Chemical Society: Washington, DC, 1982.
(91) Rose, G. D.; Gierasch, L. M.; Smith, J. A. *Adv. Protein Chem.* **1985**, *37*, 1−109.
(92) Baysal, C.; Meirovitch, H. *J. Phys. Chem. B* **1997**, *101*, 7368−7370.

are known, one can simulate each of them with explicit solvent to investigate the microscopic effect of the solvent.

We approximate the exact energy $E_{exa}(\mathbf{x})$ by $E_{tot}(\mathbf{x})$

$$E_{tot}(\mathbf{x}) = E_{GRO}(\mathbf{x}) + E_{sol}(\mathbf{x}) = E_{GRO}(\mathbf{x}) + \sum_i \sigma_i A_i(\mathbf{x}) \qquad (3)$$

where the surface areas $A_i(\mathbf{x})$ are conformation dependent and the σ_i depend on T. Obviously, it would be impossible to find a set of σ_i that equate E_{tot} to E_{exa}, not only because E_{exa} is unknown, but also due to the fact that E_{sol} is a highly simplified function. However, if the experiment suggests that the molecule coexists in several Ω_j, the related free energies $F_j(E_{tot})$ should be the lowest globally; this is the criterion upon which the set of ASPs should be optimized. However, in this paper a partial optimization based on energy (rather than free energy) considerations is carried out. Thus, the energies of the lowest energy minimized structures that belong to the experimental βI and βII motifs are required to lie within the 2 kcal/mol range above the GEM(E_{tot}), as close as possible to each other and to the GEM(E_{tot}).

Conformational Search. LTD method[88,89] is implemented within the GROMOS package and applied for the first time to a cyclic peptide with side chains. The low energy structures of GROMOS, including the GEM, are not expected to resemble the experimental structures because E_{GRO} does not include solvent effects. The latter structures can be recovered by using LTD with E_{GRO} and an *unphysical* restraining potential P_{res}, which takes into account the 31 distances obtained by NMR; the GROMOS package enables one to include such potentials.[93] The total energy denoted E_{res} is

$$E_{res} = E_{GRO} + k_r P_{res} \qquad (4)$$

where k_r is a force constant. LTD is applied with three potential energies, E_{GRO}, E_{res} (eq 4) and E_{tot} (eq 3).

The search starts from a structure of a *linear* peptide based on the equilibrium bond lengths and angles, where random values are assigned to the dihedral angles. This structure is accepted if the peptide bond connecting the first and sixth residues is within the ring closure range of 0.5 to 3.5 Å. Side-chain coordinates are added to the backbone with their equilibrium bond lengths and angles, and a random set of dihedral angle values. A successful initial structure is then energy minimized to get *cyclic* structure *i*. *i* is then changed by local torsional deformations, energy minimized, and the resulting *trial* structure *j* is accepted with a Metropolis MC probability p_{ij}

$$p_{ij} = \min(1, \exp[-(E_j - E_i)/k_B T^*]) \qquad (5)$$

(or *i* is accepted again with probability 1-p_{ij}). The accepted conformation is deformed again and the process continues. Here E_i and E_j are the corresponding *minimized* energies and T^* is a temperature parameter that affects the efficiency.[70] Equation 5 defines the Monte Carlo minimization (MCM) "selection procedure" proposed by Li and Scheraga[94] that has been found to be an efficient tool for directing the search toward the low-energy regions of the conformational space. Notice, however that unlike the usual MC method (in which the energies rather than the minimized energies appear in eq 5), the conformations are not distributed according to the Boltzmann probability. We have obtained a relatively good efficiency by varying T^* between 200 and 800 K during the process in increments of 200 K every 200 LTD steps. The coordinates and energies of *all* the energy-minimized structures, including those which were rejected through eq 5, were stored in a file for further analysis.[95,89]

MCM has been used extensively for conformational search of linear peptides, where the deformation of a structure is straightforward. LTD defines a prescription for conformational deformation of cyclic molecules that is based on local torsional changes. This is outlined in Figure 1 for a *cyclo*(Gly)$_6$ molecule whose hydrogen atoms are not displayed. The chain, which is shown in its undeformed state in Figure 1a, consists of N backbone atoms labeled sequentially i, $1 \le i \le N$; l_i denotes the bond connecting atoms i and $i+1$. The four atoms $i-1$, i, $i+1$, and $i+2$ define the torsional angle ϕ_i around l_i (for this explanation we do not distinguish between the different dihedral angles ϕ, ψ, and ω). With a *local* rotation, ϕ_i around bond l_i, only atom $i+2$ is moved, while the rest of the atoms are kept fixed in their current positions as shown in Figure 1b. ϕ_i is changed at random within a range $\pm D$ around its current value and the condition that the affected bond, l_{i+2}, should remain within the ring closure range of 0.5 to 3.5 Å is imposed.

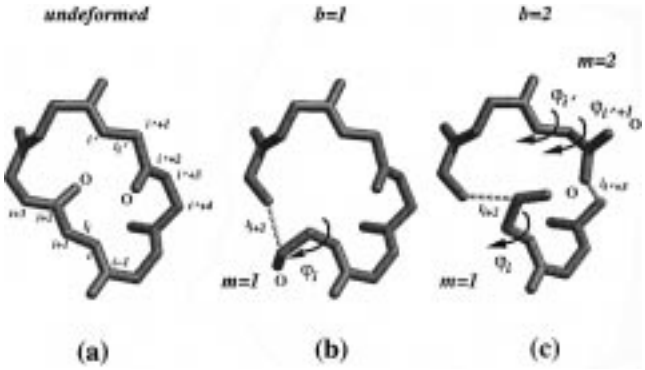

Figure 1. Local rotations of LTD illustrated for the *cyclo*(Gly)$_6$ molecule; hydrogen atoms are not displayed for simplicity: (a) the undeformed structure; (b) a local rotation ϕ_i around the single bond l_i (denoted $m = 1$) affects the backbone atom at location $i+2$ as well as the side chain atoms attached to the atoms at $i+1$ and $i+2$; the affected bond l_{i+2} is denoted by a dashed line; (c) two simulataneous local rotations ($b = 2$); a single rotation ($m = 1$) is applied to the lower part of the molecule (the affected bond is l_{i+2}) whereas two successive rotations ($m = 2$) around bonds $l_{i'}$ and $l_{i'+1}$ are applied at the upper part, and the affected bond is $l_{i'+3}$; in this case the positions of atoms $i'+2$, $i'+3$ and the side chains attached to atoms $i'+1$, $i'+2$, and $i'+3$ are changed. Note the movement of the labeled oxygen atoms bonded to the carbons at $i+2$ and $i'+2$.

To achieve a larger conformational change, one can *locally* rotate m ($m \ge 1$) successive torsional angles; however, the dihedral angles ω are not rotated due to their high energy barriers for rotation; therefore, $m > 2$ is not permitted in this work. To increase the extent of deformation, $b \ge 1$ local rotations can be applied simultaneously along the chain, where b typically increases with increasing the molecular size. Each of the b locally deformed segments must be separated by at least two unrotated bonds to guarantee independence. In the rare event where the ring closure range is violated, the set of deformations along the chain is discarded and a new set is randomly determined. A specific example is presented in Figure 1c for the case of $b = 2$, where deformations of $m = 1$ and $m = 2$ are applied to the lower and upper part of the molecule, respectively. Changing the LTD parameters enables one to control the extent of conformational deformation and adapt it to the particular molecular conditions (see also refs 88 and 89).

For the present cyclic peptide, the deformations along the backbone are induced by LTD as described above, whereas the conformational changes of the side chains are obtained by random rotations of the angles χ. These types of backbone and side chain moves are carried out simultaneously prior to the minimization. The parameters of LTD are chosen as follows: First, the number of backbone angles to be rotated is determined with probability 0.1, 0.3, 0.3, and 0.3 for rotating one, two, three, and four dihedral angles, respectively. Next, the specific combination of rotations is determined; e.g., if three angles were chosen, equal probability is given to three $m = 1$ rotations and a combination of $m = 1$ and $m = 2$ rotations. The specific angles to be rotated are then selected at random and their current values are randomly chosen within the range $\pm D$, where $D = 180°$ for $m = 1$ and $D = 90°$ for $m = 2$. To change the side chains conformations, up to three χ_1 and χ_2 angles out of the total eight are chosen at random, i.e., with equal probability of $^1/_3$ for selecting one, two, or three angles. The

(93) Zuiderweg, E. R. P.; Scheek, R. M.; Boelens, R.; van Gunsteren, W. F.; Kaptein, R. *Biochimie* **1985**, *67*, 707−715.

(94) Li, Z.; Scheraga, H. A. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *84*, 6611−6615.

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

specific angles are then selected at random and their values are randomly selected within ±180° around their current values. This largely random approach proved to be most efficient for the present molecule due to its heterogeneity, in contrast to our experience with symmetric cycloalkanes, where a distinct set of the parameters, *m*, *b*, and *D*, produced the best results[89] and a randomized combination did not invoke an improvement.[96]

The success of this procedure depends to a large extent on using an efficient minimizer for the energy. We therefore tested the conjugate gradients and steepest descents minimizers provided by GROMOS, and also implemented in this package several other minimizers and examined their efficiency. These are the secant unconstrained minimization solver (SUMSL),[97] used successfully with ECEPP, two NETLIB routines, utilizing quasi-Newton and Beale restarted conjugate gradients methods,[98] and a limited memory BFGS method.[99−101] The latter was found to be significantly more efficient than the others for the present application. With this minimizer a 0.004 kcal mol^{-1} Å^{-2} tolerance for the derivatives resulted in structures whose dihedral angles varied by at most 0.4° from those obtained with a tolerance of 10^{-8} kcal mol^{-1} Å^{-2}; therefore, the former value was used in all the minimizations based on E_{GRO}. With this tolerance, it takes 6 s on average to fully minimize a deformed structure of the cyclohexapeptide on the SGI Indigo2 R4000 workstation.

It should be pointed out that the GROMOS force field is designed to be used mainly in MD simulations, where atoms approach each other gradually. In this case a hydrogen atom will be repelled by an approaching heavy atom due to the LJ interaction between the latter and the heavy atom to which the hydrogen is attached; therefore, no LJ parameters are provided in GROMOS for hydrogens. However, the strong LTD deformations (before minimization) can lead to structures in which a hydrogen significantly "penetrates" the van der Waals radius of a heavy atom. In this case the attractive electrostatic force between the hydrogen and the heavy atom might exceed the repulsive force between the two heavy atoms; this difference in forces may increase during minimization causing a further unphysical approach of the hydrogen to the heavy atom. To eliminate this undesired situation, conformations with electrostatic energy smaller than -10^6 kcal/mol before minimization are discarded and a new structural deformation is applied. Another structural transformation that can occur during the minimization is from trans to cis configurations of the peptide bonds. However, the experimental evidence for the current molecule precludes the occurrence of cis conformers[16] and therefore these are discarded during the search. To save computer time, the check for cis is done after 10 minimization steps, since it was found that the general features of the minimized structure are already attained at this point; the extra ∼100 steps that are typically carried out for a complete minimization are needed for structure refinement.

Identifying the Experimental Structures from the NOE Data. Using LTD, we generated a large set of significantly different energy-minimized conformations, where our goal is to identify those satisfying most of the experimental NOE distances. A search based on E_{GRO} alone (i.e., no solvent effects), has shown that to achieve the above goal, structures within at least 16 kcal/mol above the lowest energy structure found should be included, rather than only those within 2 kcal/mol, as discussed in the Introduction. The same range of energy is also searched with eq 4. However, to prevent nonphysical atom−atom overlaps, the structures of the latter group were further minimized to the closest minimum with respect to E_{GRO} alone. The final set is a collection of structures obtained in several different searches.

For each of these structures (denoted *j*), one calculates the distances $r_i(j)$ $1 \le i \le 31$ between the 31 pairs of atoms for which experimental NOE distances are available.[16] The experimentally measured lower bound distances d_i are listed in Table 1; most of them are equal to the upper bounds. However, for each methyl group and nonstereospecifically assigned CH_2 group, upper bounds are obtained by adding 1.0 and 0.9 Å, respectively, to the listed values. Now, each $r_i(j)$ is classified into one of three categories, denoted by

$$M_i(j) = \begin{cases} -1 & \text{if } r_i(j) < d_i(1-\delta) \\ 1 & \text{if } d_i(1-\delta) < r_i(j) < d_i(1+\delta) \\ 0 & \text{if } r_i(j) > d_i(1+\delta) \end{cases} \quad (6)$$

where δ is the experimental error of d_i ($\delta = 0.1$ in ref 16, i.e., an error of 10% is considered). In eq 6, the d_i values used for the methyl and nonstereospecifically assigned CH_2 groups are their upper and lower bounds in the expressions $(1 + \delta)$ and $(1 - \delta)$, respectively.

This classification enables one to filter out structures that do not contribute significantly to the NOEs. Thus, structures for which $\sum_i |M_i| < 19$ are discarded; in other words, we require that at least 60% of the NOE distance constraints (including $M_i = -1$) are satisfied for $\delta = 0.1$. This criterion is based on the assumption that the number of most stable microstates with comparable free energy is small, they are not very different structurally, and therefore each should satisfy a considerable number of the NOEs. Otherwise, the molecule would become a random coil and the NOEs would not be observed experimentally.[36] Notice that the rare case of $M_i = -1$ is also considered as a satisfied distance constraint, because it may contribute significantly in combinations with larger distances due to the r^{-6} dependence of the NOE intensity.

If none of the structures of the set satisfies *all* the NOEs, one tries to achieve this goal by a combination of a *minimal* number of structures from the set; thus, pairs of structures are examined first and if necessary triplets and quartets are tested as well. For a combination of *J* structures the corresponding *ideal* probabilities (populations) p_j ($\sum_j p_j = 1$; $p_j \ge 0$)

Table 1. Experimental and Calculated Proton−Proton Distances (Å)

		ROESY[a]	βI	βII	average[d]
Phe^{5}N*H*	Phe^{6}N*H*	2.39	2.43	3.28	2.62
Phe^{5}N*H*	Ala^{3}C$_\beta$*H*	3.32[b]	3.69	6.20	4.04
Phe^{5}N*H*	Phe^{5}C$_\alpha$*H*	2.18	2.83	2.11	2.34
Phe^{5}N*H*	Ser^{4}C$_\alpha$*H*	2.30	3.43	2.09	2.38
Phe^{5}N*H*	Ser^{4}C$_\beta$*H*	2.84[c]	3.37	3.93	3.55
Phe^{5}N*H*	Phe^{5}C$_\beta$*H*$^{pro-R}$	2.99	2.54	3.29	2.73
Phe^{5}N*H*	Phe^{5}C$_\beta$*H*$^{pro-S}$	3.47	3.59	3.66	3.62
Phe^{5}C$_\alpha$*H*	Phe^{5}C$_\beta$*H*$^{pro-S}$	2.61	2.42	2.89	2.56
Phe^{6}N*H*	Phe^{6}C$_\alpha$*H*	2.70	2.68	2.86	2.75
Phe^{6}N*H*	Phe^{6}C$_\beta$*H*$^{pro-S}$	2.85	3.57	2.51	2.81
Phe^{6}N*H*	Phe^{6}C$_\beta$*H*$^{pro-R}$	2.96	4.03	2.80	3.14
Phe^{6}C$_\alpha$*H*	Phe^{6}C$_\beta$*H*$^{pro-S}$	2.52	2.38	2.90	2.53
Phe^{6}C$_\alpha$*H*	Phe^{6}C$_\beta$*H*$^{pro-R}$	2.37	2.40	2.40	2.40
Phe^{6}C$_\alpha$*H*	D-Pro^{1}C$_\delta$*H*	2.24[c]	2.20	2.13	2.16
Phe^{2}N*H*	Ala^{3}N*H*	2.50	2.97	2.74	**2.85**
Phe^{2}N*H*	Phe^{2}C$_\alpha$*H*	2.74	2.81	2.72	2.77
Phe^{2}N*H*	D-Pro^{1}C$_\alpha$*H*	2.07	2.10	2.02	2.06
Phe^{2}N*H*	Phe^{2}C$_\beta$*H*$^{pro-R}$	2.58	2.31	2.42	2.35
Phe^{2}N*H*	Phe^{2}C$_\beta$*H*$^{pro-S}$	3.31	3.47	3.45	3.47
Phe^{2}C$_\alpha$*H*	Phe^{2}C$_\beta$*H*$^{pro-S}$	2.35	2.49	2.45	2.47
Ala^{3}N*H*	Ala^{3}C$_\alpha$*H*	2.72	2.85	2.87	2.86
Ala^{3}N*H*	Phe^{2}C$_\alpha$*H*	2.78	2.90	3.36	3.04
Ala^{3}N*H*	Ala^{3}C$_\beta$*H*	2.55[b]	3.13	2.89	3.00
Ala^{3}C$_\alpha$*H*	Ala^{3}C$_\beta$*H*	2.22[b]	2.40	2.41	2.40
Ser^{4}N*H*	Phe^{5}N*H*	3.16	3.17	4.34	3.42
Ser^{4}N*H*	Ser^{4}C$_\alpha$*H*	2.69	2.55	2.60	2.57
Ser^{4}N*H*	Ser^{4}C$_\beta$*H*	2.60[c]	2.94	2.95	2.94
Ser^{4}N*H*	Ala^{3}C$_\beta$*H*	2.57[b]	3.42	3.32	3.37
Ser^{4}N*H*	Ala^{3}C$_\alpha$*H*	2.36	2.35	2.41	2.38
Ser^{4}N*H*	Ser^{4}O*H*	2.97	3.02	3.13	3.07
Ser^{4}C$_\alpha$*H*	Ser^{4}O*H*	2.77	2.75	2.80	2.77

[a] From Kessler et al.[16] [b] Lower bound for a CH_3 group; 1.0 Å is added for the upper bound. [c] Lower bound for a nonstereospecifically assigned CH_2 group; 0.9 Å is added for the upper bound. [d] Using eq 7 for the βI and βII structures that best fit the data for $\delta = 0.10$ (the respective probabilities are 0.56 and 0.44); see Table 5. The single violated NOE is boldfaced.

(96) Baysal, C.; Meirovitch, H. Unpublished data.
(97) Gay, D. M. *ACM Trans. Math. Software* **1983**, *9*, 503−525.
(98) Shanno, D. F.; Phua, K. H. *ACM Trans. Math. Software* **1975**, *1*, 87−94.
(99) Nocedal, J. *Acta Numerica* **1991**, 199−242.
(100) Liu, D. C.; Nocedal, J. *Mathematical Programming* **1989**, *45*, 503−528.
(101) Zou, X.; Navon, I.; Berger, M.; Phua, M.; Schlick, T.; LeDimet, F. *SIAM Journal on Optimization* **1993**, *3*, 582−608.

are those that satisfy the 31 equations for i

$$\frac{1}{d_i^6} = \sum_{j=1}^{J} \frac{p_j}{r_i^6(j)} \tag{7}$$

within the error bars. In practice, a set of p_j that exactly solves the last equation might not exist; therefore one seeks to find such a set that minimizes the following function A (again within the error bars) under the normalization condition

$$\begin{cases} \text{minimize } A = \sum_{i=1}^{31} \left(\frac{1}{d_i^6} - \sum_{j=1}^{J} \frac{p_j}{r_i^6(j)} \right)^2 \\ \text{subject to } \sum_{j=1}^{J} p_j = 1;\ p_j \geq 0 \end{cases} \tag{8}$$

The minimization is carried out in several iterations. For the first iteration the experimental values of d_i are used in eq 8 and the solution leads to average values r_i(av) for each i. If r_i(av) lies within the range $d_i(1 \pm \delta)$, the contribution of this NOE to the summation over i is 0 and the value of d_i for the next iteration is set to r_i(av) in eq 8. If r_i(av) falls outside this range, d_i is set to the nearest border value $d_i(1 \pm \delta)$. Equation 8 is then solved again with the new values of d_i and the process is repeated until convergence is obtained, i.e., the set of p_j's is unchanged. If A is zero, all the distance constraints are satisfied, otherwise the number of unsatisfied distance constrains and their contribution to A are determined. Notice that the number of combinations n_c, $n_c = n!/J!(n - J)!$ of J structures increases dramatically with increasing J and the size of the set n. Thus, if $J = 3$, $n_c \approx 1.6 \times 10^5$ for $n = 100$, whereas $n_c \approx 1.7 \times 10^8$ for $n = 1000$. Therefore, n must be considerably reduced for an efficient and complete analysis. For a large J, eq 8 should be handled by quadratic programming.[34,30] In our case J is small and a brute force grid search method was applied. As pointed out in the Introduction, this analysis is based on the model proposed by Kessler et al.[8]

The Solvent-Accessible Surface Area. The solvent-accessible surface area (SASA) is defined as the area swept by the center of a spherical solvent probe with radius r_p as it is rolled over the van der Waals surface of the solute.[102] Since the experiments were carried out in DMSO, one has to define a suitable radius for its spherical probe. However this is not straightforward because DMSO is not spherical. Therefore, we used some geometrical arguments based on data given in ref 103 to determine a reasonable radius r_p. Thus, we calculated the maximum distance between the center of the DMSO molecule and the outer van der Waals surface of each of the atoms obtaining, 2.37 for S, 2.86 for O, and 3.54 Å for each of the united atoms CH_3; the average distance is 3.08 Å (the center is $^1/_4\Sigma\bar{r}_i^0$, where $\bar{r}_i^0$ is the vector defining the center of each united atom). Similarly, the average distance from the center of mass is 3.04 Å. We chose the radius of the sphere as 3.0 Å which is slightly smaller than the above average distances, because a larger radius would not allow the sphere to penetrate grooves on the surface that can be accessed by each of the DMSO atoms. This choice of r_p is also reasonable based on comparison of radial distribution functions (rdf) of water and DMSO obtained from simulations. For water, the commonly used radius $r_p = 1.4$ Å is equal to the distance at which the oxygen−hydrogen rdf becomes approximately zero for various water models.[104] Rdf's for the six possible atom pairs of DMSO were calculated by Rao and Singh[105] who found that three of them (S−C, C−C, and O−O) become zero at about 3.1 Å, whereas the other three at 2.8−2.9 Å.

Most of the conformational search studies using eq 3 were carried out with the program MSEED,[77] which calculates both the SASA and its derivatives analytically (we did not experience problems from discontinuities in the derivatives[106]). Because this calculation is relatively time-consuming, we also tested an approximate procedure, which requires significantly less computer time, suggested first by Wodak and Janin,[107] improved by Hasel et al.,[81] and used recently for proteins by Fraternali and van Gunsteren.[86] In this formulation the SASA, A_i, is given by the expression

$$A_i = S_i \prod_j (1.0 - p_i p_{ij} b_{ij}/S_i) \tag{9}$$

where S_i is the SASA of an isolated atom with radius r_i

$$S_i = 4\pi(r_i + r_p)^2 \tag{10}$$

and b_{ij} is the area removed from atom i due to overlap with atom j for an atomic separation of $r_{ij} \leq r_i + r_j + 2r_p$

$$b_{ij} = \pi(r_i + r_p)(r_j + r_i + 2r_p - r_{ij})[1.0 + (r_j - r_i)/r_{ij}] \tag{11}$$

p_i and p_{ij} account for the reduction in b_{ij} due to multiple overlapping atoms. The former carries atom type specific information, such as hybridization and substitution, whereas the latter depends on the relative proximity of atoms i and j along the contour of the chain. The expressions for the derivatives with respect to atomic coordinates are given in the Appendix of ref 81.

Hasel et al. optimized the parameters p_i and p_{ij} for a sphere of $r_p = 1.4$ Å over a set of small molecules with various atom types. To check this approximate method (called ATAREA), we applied it and the ASC program of Eisenhaber and Argos,[78] which calculates SASA *exactly*, to 10 randomly selected conformations of the cyclic hexapeptide. The average SASA, 678 ± 47 Å^2 obtained with ATAREA is significantly lower than the exact average value, 826 ± 41 Å.^{2} However, these parameters become completely unacceptable for DMSO ($r_p = 3$ Å), where they lead to an average SASA of 284 ± 71 Å vs the exact value 1180 ± 51 Å.

We reoptimized the ATAREA parameters for $r_p = 3.0$ Å over a subset of 50 different energy-minimized structures of the cyclic hexapeptide. The exact areas of individual atoms were calculated and the values of p_i and p_{ij} were chosen to give the minimum *average* deviation for the various atom types. The atom types and their optimized p_i and p_{ij} values are listed in Table 2, together with the original parameters of Hasel et al. (for $r_p = 1.4$ Å). With the new parameters, the approximate average area 1165.7 ± 116 Å of the 50 conformations is equal to the exact average area 1165.7 ± 63 Å. Similarly, the respective approximate and exact average SASA of

Table 2. The Optimized Parameters p_i and p_{ij} of Eq 9 for an Approximate Calculation of the Solvent-Accessible Surface Area

atom type, i	p_i[a]	p_i[b]
CH (sp^3)	1.12	1.276
CH_2 (sp^3)	0.85	1.045
CH_3 (sp^3)	0.90	0.880
C (sp^2)	1.40	1.554
CH (sp^2)	1.02	1.073
H (OH)	1.03	0.944
H (NH)	1.34	1.128
O (sp^3)	1.43	1.080
O (sp^2)	1.21	0.926
N (sp^2)	1.70	1.028
neighbor type	**p_{ij}[a]**	**p_{ij}[b]**
$\|i - j\| = 1$	0.73	0.8875
$\|i - j\| = 2$	0.58	0.3516
$\|i - j\| \geq 3$	0.12	0.3156

[a] Optimized in this work for a DMSO probe of $r_p = 3.0$ Å. [b] The optimized parameters of Hasel et al.[81] for a water probe of $r_p = 1.4$ Å.

(102) Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, *55*, 379−400.
(103) Liu, H.; Müller-Plathe, F.; van Gunsteren, W. F. *J. Am. Chem. Soc.* **1995**, *117*, 4363−4366.
(104) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(105) Rao, B. G.; Singh, U. C. *J. Am. Chem. Soc.* **1990**, *112*, 3803−3811.
(106) Wawak, R. J.; Gibson, K. D.; Scheraga, H. A. *J. Math. Chem.* **1994**, *15*, 207−232.
(107) Wodak, S. J.; Janin, J. *Proc. Natl. Acad. Sci. U.S.A.* **1980**, *77*, 1736−1740.

individual atom types are basically the same, 0.4 ± 0.1 and 0.0 ± 0.0 Å for N, 14.3 ± 4.3 and 14.2 ± 15.7 Å for O, 5.8 ± 6.8 and 5.5 ± 10.5 Å for H, and 27.1 ± 18.6 and 27.2 ± 25.6 Å for the C atoms. Similar results were obtained for 20 randomly selected conformations that were not included in the original set. Notice, however, that while the best-fitted parameters provide very good results for the *average* SASA of atoms, the fluctuations of the exact results are significantly larger than those of the approximate ones. Thus, the approximate SASA of an atom in an individual structure might deviate significantly from the exact value.

Our interest in the approximate ATAREA method stems from its efficiency. We found that calculating the SASA and its derivatives for 280 structures of our molecule is ~7 times faster with ATAREA than with MSEED. If the computer time required for calculating E_{GRO} and its derivatives is added to the above times, using E_{GRO} + ATAREA is four times faster than E_{GRO} + MSEED. We have shown that the accuracy of ATAREA can significantly be improved by optimizing its parameters for the specific system studied. However, the too small fluctuations obtained by this method suggest that further improvement of the method might be needed, which can be achieved by adding a parameter to eq 9.

Results and Discussion

Application of LTD with E_{GRO}. E_{GRO} is not expected to provide a correct description of the cyclic hexapeptide in DMSO. Therefore, to locate the experimental structures from an LTD conformational search based on E_{GRO}, we kept energy-minimized structures from a wide range of energies rather than only within 2 kcal/mol above the GEM. An important objective was to find the GEM, because the deviation of the energies of the experimental structures from it constitutes a measure of the inadequacy of E_{GRO}, which should be corrected by E_{tot} (eq 3).

Several long LTD runs which total to 1.4×10^5 minimizations were carried out starting from different conformations. The lowest energy structure of 10.3 kcal/mol was obtained many times, and due to the large number of minimizations, it is considered to be the GEM structure. This assertion is supported by the fact that the GEM structure of cycloheptadecane with a comparable backbone to the present cyclo-hexapeptide was found on average after 500 minimizations.[89,95] The linear pentapeptides Met- and Leu-enkephalin, having larger conformational space than cycloheptadecane require on average 5000 minimizations using the ECEPP potential with variable ω.[38,94] From the large number of energy-minimized conformations obtained in the search, we considered only those which are *significantly different* (see ref 37); the criterion for variance of two conformations is that at least one dihedral angle differs by 60° or more.[36−38]

Using this criterion, the number of distinct conformations pertaining to energy bins of 1.0 kcal/mol up to 18 kcal/mol above the GEM was determined and the results appear in the second and third columns of Table 3. The results of the second column are classified according to the values of ϕ, ψ, and χ_1, whereas those of the third column, with respect to the backbone angles ϕ, and ψ only. Notice that differences in the angles χ_2 are ignored in this analysis because NOEs involving the three benzene rings of the Phe residues were not detected. Column 2 of Table 3 reveals that the number of structures per bin grows systematically up to 8 kcal/mol above the GEM and then decreases monotonically, where energies of 15 kcal/mol above the GEM are not sampled at all. This decrease in the population is probably due to the bias given by MCM for selecting low energy minimized structures.

As discussed in the Introduction, the experimentally observed βI motif lies ~15 kcal/mol above the GEM(E_{GRO}). To analyze the experimental data, the low population of the high-energy bins should be enriched by applying LTD with E_{res} (eq 4).

Table 3. Number of Significantly Different Energy-Minimized Structures in Energy Bins of 1.0 kcal/mol above the Global Minimum of E_{GRO}[a] Obtained with LTD[b]

bin (kcal/mol)	ϕ, ψ, χ_1[c]	ϕ, ψ[d]			
	$k_r = 0$	$k_r = 0$	$k_r = 2.39$	$k_r = 4.78$	$k_r = 9.56$
0.0−1.0	3	1	1	0	0
1.0−2.0	6	2	3	1	1
2.0−3.0	22	2	1	0	0
3.0−4.0	59	4	4	3	0
4.0−5.0	83	3	4	1	0
5.0−6.0	106	3	6	2	0
6.0−7.0	129	8	3	3	1
7.0−8.0	142	5	9	4	1
8.0−9.0	125	6	6	0	1
9.0−10.0	101	7	4	2	3
10.0−11.0	47	2	3	1	0
11.0−12.0	19	1	4	3	0
12.0−13.0	12	0	0	1	0
13.0−14.0	22	2	2	4	1
14.0−15.0	8	1	0	3	0
15.0−16.0	0	0	0	1	0
16.0−17.0	0	0	1	1	0
17.0−18.0	0	0	0	1	0
total	809	47	51	31	8

[a] E_{GRO} = 10.3 kcal/mol. [b] k_r (kcal mol^{-1} Å^{-2}) is the force constant of the distance restraining potential in eq 4; $k_r = 0$ when no distance restraining potential is applied. Each structure is further minimized without restraints to let it attain its nearest E_{GRO} minimum. [c] For every two structures at least one of the 16 angles ϕ, ψ, or χ_1 differs by 60° or more. [d] The same as in *c* but only for the 12 dihedrals ϕ and ψ.

Application of LTD with E_{res}. We carried out LTD runs with E_{res} (eq 4) using restraint force constants, k_r = 2.39, 4.78, and 9.56 kcal mol^{-1} Å^{-2}, (corresponding to 1000, 2000, and 4000 kJ mol^{-1} nm^{-2}; typical values used for such calculations); the respective number of minimizations is 2×10^4, 6×10^4, and 3×10^4. As pointed out previously, the structures obtained in the search are further minimized without restraints to let them fall to the nearest energy minimum of E_{GRO}. The numbers of significantly different backbone motifs (with respect to ϕ and ψ only) are presented in columns 4−6 of Table 3. The best sampling was obtained with k_r = 2.39 kcal mol^{-1} Å^{-2}, even though the corresponding number of minimizations is the smallest. k_r = 4.78 kcal mol^{-1} Å^{-2} was perused longest because this constant allowed the better sampling in the higher energy region. In runs with k_r = 9.56 kcal mol^{-1} Å^{-2}, the over-emphasized effect of the restraints significantly limited the regions in conformational space that could be reached.

The structures obtained in all the runs were analyzed by calculating the number of distances that are smaller than the experimental values; as previously mentioned, only those structures satisfying $\sum_i |M_i| \geq 19$ (for $\delta = 0.10$) were kept. To those that "passed" this test, the GROMOS GEM structure, which only satisfied 15 of the NOEs was added for comparison, and a total of n = 232 structures were examined. These structures pertain to nine different backbone motifs listed in Table 4, where each motif is represented by the backbone dihedrals of the structure with lowest minimized E_{GRO} value (denoted E_L) found. We also provide for each motif the energy value E^* of the structure with the largest number M^* of satisfied NOEs generated.

The table reveals that motifs 2, 3, 4, 6, and 9 display a βII′-turn around D-Pro[1] and Phe[2], i.e., the respective ϕ and ψ angles are 60°, −120°, −80°, and 0°; in this classification, a variation of ±33° was taken around these angles. A similar arrangement of dihedrals is found in motifs 5, 7, and 8. The more significant differences are observed for the dihedrals ϕ and ψ of Ser[4] and Phe[5] which appear in a βI (motif 9) and βII (motif 2) turns.

Table 4. Different Backbone Motifs[a]

		1	2	3	4	5	6	7	8	9
D-Pro[1]	ϕ	86	56	55	59	61	55	50	80	59
	ψ	−71	−113	−103	−130	−142	−101	−110	−84	−126
Phe[2]	ϕ	−57	−57	−52	−58	−86	−47	−77	−54	−75
	ψ	103	−29	−32	−32	62	−32	32	−33	20
Ala[3]	ϕ	44	−117	−119	−89	−164	−122	−163	−150	−147
	ψ	52	165	160	73	−162	158	153	−151	148
Ser[4]	ϕ	81	−41	−45	66	−67	−73	−75	−73	−42
	ψ	−38	116	−50	−46	91	65	48	40	−42
Phe[5]	ϕ	−74	66	−67	−135	67	−179	−160	−165	−77
	ψ	103	−22	−25	−37	−39	−47	−57	125	−31
Phe[6]	ϕ	49	−114	−163	−95	−80	−135	−144	4	−164
	ψ	57	88	106	95	103	102	129	92	131
E_L[b] (kcal/mol)		10.3	15.6	17.1	17.8	18.1	18.2	21.4	22.0	25.5
E^* (kcal/mol)		12.8	18.7	21.5	19.9	20.0	22.0	23.1	24.8	26.4
M^*		15	22	21	20	21	22	24	19	25

[a] Each motif is represented by a structure for which at least one of the side-chain combinations satisfies 19 NOE distances or more; an exception is motif 1, the lowest E_{GRO} structure, which is included for comparison. Motifs 2 and 9 contain βII′/βII and βII′/βI turns respectively. Motif 3 is very similar to motif 9, and reasons for classifying them separately are given in the text. 4−8 are motifs for which at least one dihedral angle differs from the corresponding ones of others by 60° or more. More details about motif 4 appear in the text, in the discussion of Table 5. [b] E_L is the energy of the motif's structure with lowest E_{GRO}. E^* is the energy of the motif's conformation with maximum number M^* of satisified NOE distance constraints.

Table 5. Best-Fitted Structures and Their Probabilities

δ[a]	motifs[b]	p_i[c]			maximum no. of NOES[d]	average deviation[e]	maximum deviation[f]
0.05	2 9	0.51	0.49	(0.00)	26	0.49	0.80
0.05	6 6 9	0.47	0.35	0.18 (0.00)	28	0.34	0.46
0.10	2 9	0.41	0.59	(0.03)	30	0.10	0.10
0.10	2 4 9	0.39	0.11	0.50 (0.03)	31	0	0
0.15	2 9	0.46	0.54	(0.16)	31	0	0
0.20[g]	2 9	0.50	0.50	(0.20)	31	0	0
0.25	6	1.00	(0.00)		31	0	0

[a] $\delta \times 100$ is the assumed error in percent in the reported experimental distances. [b] From Table 4. [c] Probabilities correspond to the respective structures in the previous column, and their uncertainties (in parenthesis) are represented by their fluctuations when several sets of structures lead to the same number of satisified NOE distance constraints. [d] Number of satisfied NOE distance constraints out of the 31 experimentally reported ones. [e] Average deviation (in Å) of the unsatisfied distances from their allowed upper bounds. [f] Deviation (in Å) from the NOE distance that is least satisfied. [g] For $\delta = 0.20$, many structural combinations satisfy all the 31 distances, which are predominantly composed of (2, 9) motifs but some also of (2, 3), (2, 5), and (2, 7).

These turns are defined by the ϕ and ψ sequences (−60°, −30°, −90°, 0°) and (−60°, 120°, 80°, 0°), respectively; fluctuations within ±30° are allowed on each of these angles. Note that motifs 3 and 9 are basically the same, since the largest difference, 52° in the values of ψ (Phe[2]) is smaller than 60° and the value, $\psi = -32°$ of motif 3 deviates by only 2° from the border value −30° that defines a βII′ turn. Motif 3 is therefore a border case that was included because it satisfies only 21 NOE distances as compared to 25 satisfied by motif 9. This demonstrates the relatively large structural variability within each motif, which will be taken into account in our next work by MD and MC simulations.

The Optimal Set of Structures. The analysis based on eqs 6−8 is applied to the reduced set of 232 energy-minimized structures. Because the experimental error is generally determined with some uncertainty, we analyze the data assuming several error values, $\delta = 0.05-0.25$ (i.e., errors of 5−25%). Table 5 summarizes the results of this analysis, where the best-fitted structures are identified by the corresponding motif numbers defined in Table 4. Table 5 also provides the probabilities of the related structures, and in parentheses the variances of these probabilities when several different combinations of structures lead to the same best fit. In addition, we present in the table the average and maximum deviation of the unsatisfied distances from their allowed upper bounds, $d_i(1 + \delta)$. Note that these structures are not necessarily the ones whose backbone angles are shown in Table 4, although they belong to the indicated motif.

The table reveals that for a 5% error ($\delta = 0.05$), two interconverting structures do not lead to the experimentally observed distances; in fact, 5 of them remain unsatisfied. This number decreases to 3 if three structures are considered, but it does not decrease further by taking into account four structures. For a slightly increased error of 10% ($\delta = 0.10$), which is the error used in the analysis of Kessler et al.,[16] the two structures of motifs 2 (βII) and 9 (βI) lead to an almost perfect fit, where only a single distance deviates by 0.1 Å from its upper bound. Note that five different pairs of structures which pertain to these motifs were found to fit the same 30 NOE distances. On the other hand, 29 distances were satisfied by six different pairs of structures of motifs 7 and 9, one pair of 7 and 2, and many of 2 and 9. All the 31 NOE distances are satisfied by 10 combinations of three structures belonging to motifs 2, 4, and 9. The fluctuation of the probabilities (0.03) in all these cases is relatively small, reflecting the tight structural constraints imposed.

For an error of 15%, 13 cases of (2,9) already led to a perfect fit, while 30 distances are satisfied by about 60 pairs that predominantly belong to (2,9) and some to (4,7), (7,9), and (2,7). Obviously, the fluctuation of the probabilities (0.16) is larger here, due to the relaxed bounds and it increases to 0.20 for an error of 20% ($\delta = 0.20$). For $\delta = 0.25$ a single structure (6) is found, which satisfies all the 31 constraints; it still has a βII′ turn around D-Pro[1] and Phe[2] but an undefined structure around Ser[4] and Phe[5].

While the above results demonstrate the sensitivity of the analysis to the experimental errors, they also show clearly that structures of the βI−βII (9,2) motifs provide a better fit than the other pairs (note that the X-ray structure of the peptide also pertains to the βI motif[16]). This suggests, with relatively high confidence, that the βI and βII motifs alone coexist in equilibrium; however, as also pointed out by Kessler et al., the participation of a third motif (e.g., 4) cannot be ruled out. In Table 1 we present the proton−proton distances of each of the optimal βI and βII structures as well as the weighted average

distances r_i(av) for $\delta = 0.10$, obtained with eq 7. This table should be compared to Table 7 of ref 16 where the theoretical NOE distances were calculated from MD simulations.

Similarly, $^3J(NH,C_\alpha H)$ were calculated based on the same two structures of Table 1. The coupling constant (in hertz) for the *i*-th residue of the *j*-th structure was calculated by Bystrov's Karplus equation with correction for the electronegativity effect,[108] $^3J\ (NH,C_\alpha H)_i(j) = [9.4\ \cos^2\ \theta_i(j) - 1.1\ \cos\ \theta_i(j) + 0.4]/1.90$ where $\theta_i = \phi_i - 60°$. The results were then weighted by the probability of each structure to obtain the average $^3J(NH,C_\alpha H)$ for each residue. The coupling contants thus obtained and their respective experimental values[16] are 5.8 ± 0.3 and 8.9 for Phe[2], 9.2 ± 0.7 and 10.2 for Ala[3], 1.2 ± 0.5 and 3.8 for Ser[4], 8.6 ± 0.7 and 7.7 for Phe[5], 6.6 ± 0.5 and 6.4 for Phe[6], which give an rms deviation of 1.57 Hz. The uncertainties in the theoretical values reflect the possible range of the ^3J values for a given ϕ angle (see Figure 5 of ref 108). These results are considered satisfactory due to the high sensitivity of ^3J to small changes in the angles.[109]

It should be pointed out that the fits presented above for the NOEs and the $^3J(NH,C_\alpha H)$ coupling constants were obtained for specific pairs of energy minimized structures. However, as has been discussed in the Introduction, at room temperature the molecule is expected to stay in a localized microstate only for a very short time, while spending much longer times in the related wide microstate that consists of many localized ones. Therefore, a more realistic (hence demanding) criterion for a structural fit would be to require that the experimental NOEs be recovered from two different MD samples that span the wide microstates of motifs 2 and 9. Such simulations were carried out by Kessler et al. using E_{res} (eq 4). They did not get a perfect fit (see their Table 7), which might stem, to some extent, from the fact that they do not use the correct force field for the molecule in DMSO, but also because of the need to consider additional motifs. For the same reason, the probabilities in Table 5 are expected to be approximate.

In summary, the procedure applied above is not much different in spirit from best-fit procedures used by others before.[24−34] However, the conformational search carried out here using LTD appears to be much more extensive and the analysis more systematic than that applied in most of the previous studies. The results of Table 5 should be considered only as a first approximation, where the next stage should consist of MD or MC simulations and calculation of the populations as above, from a best-fit analysis as well as from first principles, i.e., from the free energies obtained with the LS method. To do that, we develop in the next section an implicit solvation model for the molecule in DMSO in conjunction with the GROMOS force field.

Development of ASPs for the Implicit Solvent Model. The effect of solvent is visually demonstrated by the GEM structure of E_{GRO} and the representative structures of the experimentally predicted βI and βII motifs depicted in Figure 2 as stereoviews. The backbone of the GEM structure is bent, whereas those of the βI and βII structures are almost planar. Similarly, the side chains of the former are collapsed onto the backbone while those of the latter protrude outside. Obviously, the compact GEM structure is a result of the lack of competing solvent interactions. Note that the energies reported in the figure for the βI and βII structures are somewhat higher than those reported in Table 4. This is because the structures in the figure correspond to those with the experimentally most favorable side chain conformations predicted from ^3J coupling constants.

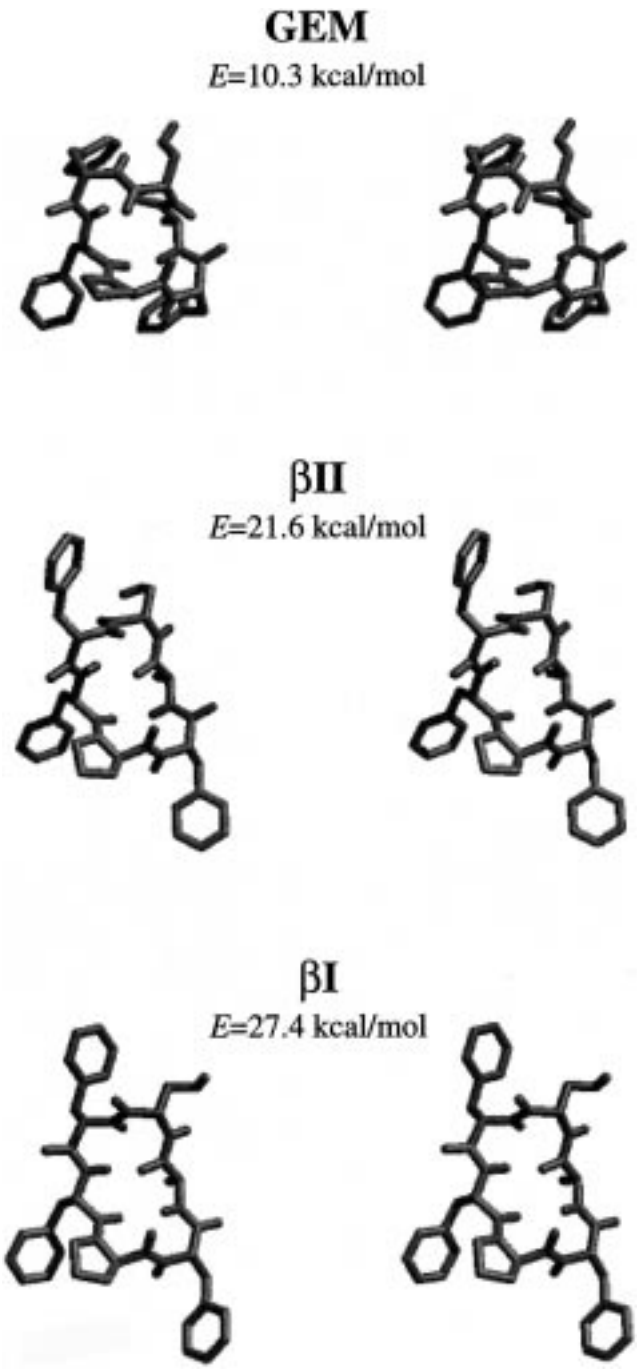

Figure 2. Stereoviews showing the differences between the GEM(E_{GRO}) structure and βI and βII which are preferred in solution. The GEM(E_{GRO}) structure has a bent backbone and its side chains are collapsed onto the backbone. On the other hand, for the βII and βI structures, the backbone adopts a nearly planar shape and the side chains are protruded into the solvent. The energy values, E_{GRO}, are also provided.

To quantify these differences, we calculated the SASA of various atom types of the βI and βII structures and three low E_{GRO} structures with different backbone patterns, denoted by S1, S2, and S3; structure S1 which is the GEM(E_{GRO}), as well as βI and βII are also depicted in Figure 2. The results are presented in Table 6, which shows that the areas of the individual backbone atoms are slightly larger for S1−S3 than for the two β structures, except for the carbonyl oxygen, where the average area of the latter structures exceeds those of the

(108) Bystrov, V. F. *Prog. Nucl. Magn. Reson. Spectrosc.* **1976**, *10*, 41−81.

(109) Smith, L. J.; Mark, A. E.; Dobson, C. M.; van Gunsteren, W. F. *Biochemistry* **1995**, *34*, 10918−10931.

Table 6. SASA (Å^2) of Atoms/Groups in Selected Structures[a]

	S1	S2	S3	βI	βII
N	0	0	0	0	0
H (NH)	35	14	2	3	18
C′	21	17	20	22	6
O (C′O)	21	81	76	134	92
C^{α}	46	53	70	27	31
backbone (total)	123	165	168	186	147
side chains (total)	982	945	941	1101	1159
peptide (total)	1105	1110	1109	1287	1306
E_{GRO}(kcal/mol)	10.3	11.4	11.5	27.4	21.6
$\chi_1^2,\chi_1^5,\chi_1^6$ (degrees)	−63,54,50	−59,62,47	61,59,−175	−62,−58,−175	−61,−65,−176

[a] S1−S3 are low E_{GRO} structures. S1, βI, and βII correspond to the structures in Figure 2. The SASA for each atom/group type is the total contribution of these atoms in the peptide.

former ones by about 50%. Consequently, the total SASA of the backbone atoms is, on the average, larger by ∼10% for the βI and βII structures. The difference becomes more pronounced for the side chains, where the β conformations take up ∼18% larger area than S1−S3, leading to a difference of ∼17% for the whole peptide. This is consistent with the conclusions drawn from Figure 2, which suggests that most of the ASPs (eq 1) should be negative, forcing the molecule conformation to open up. To optimize the ASPs we require that both β structures will have energy within 2 kcal/mol of the GEM(E_{tot}) and as close as possible to the GEM.

The structural optimization is an iterative procedure that relies on an extensive conformational search with LTD. It started from a set of 55 structures selected from the large sample generated with LTD based on E_{GRO}. These structures, which include representatives of the βI and βII motifs, were of diverse energies, backbone motifs, and χ_1 values. At the first stage only a single parameter was considered, i.e., $\sigma_i = \sigma$ for all the atoms i in eq 1. After selecting a value of σ, all the 55 structures were reminimized with respect to $E_{tot}(\sigma)$ (eq 3) and it was verified that the resulting changes in the dihedral angles did not exceed several degrees, i.e., the structural motifs were preserved. Next, several structures with the lowest energies, as well as the βI and βII structures, were given a further consideration. Thus, for each backbone the 27 combinations of side chain conformations, based on the three χ_1 values (60, −60, and 180°) of the three Phe residues were generated and minimized with respect to $E_{tot}(\sigma)$; again, each backbone conformation was verified to remain in its starting motif and the conformer with the lowest energy was selected to represent its backbone motif. Finally, the energy differences were calculated between $E_L(\sigma)$, the lowest energy found, and those of the best representatives of the βI and βII motifs. This procedure was repeated for many different values of σ and its optimal value σ^* was determined according the criterion discussed above. Then, an extensive conformational search was carried out with LTD, based on 7500 minimizations of $E_{tot}(\sigma^*)$; if new structures with energy lower than $E_L(\sigma)$ were not obtained, the optimization of σ was stopped. Otherwise, the new low-energy structures were added to the set and a new round of optimization for σ was carried out.

The final set of structures obtained in the search for the best single parameter was used as the starting set for the two-parameter optimization, where the O atoms were allowed to take on a different ASP value than the rest of the atoms. The optimization procedure was continued as before, and is finally extended to three σ_i, where an additional parameter was assigned to the H atoms. As for the single parameter case, extensive LTD runs of 7500 minimizations were performed. The set of test structures increased from 55 to 76 during the overall process. The results are summarized in Table 7 where those at the top were obtained with E_{GRO}, and those in the middle with E_{tot} and the program MSEED[77] for calculating the SASA. Note that preliminary results for these parameters were reported in ref 92. The results at the bottom were obtained by calculating the SASA with the approximate method ATAREA of Hasel et al.[81]

The table reveals that for the MSEED calculations, the difference between the energy of the βI structure and the lowest energy ($E_{\beta I} - E_L$) deceases systematically from 15.2 to 1.1 kcal/mol, as the number of σ_i increases from 0 (i.e., for E_{GRO}) to 3. The corresponding decrease for the βII structure is from 5.3 to 0 kcal/mol, i.e., it becomes the GEM structure. Note that the number of minimizations carried out in each LTD run are 7500, as compared to 5000 in ref 92. The present extra calculations did not change the results for $MSEED_1$ and $MSEED_2$ but slightly changed those for $MSEED_3$. In particular, $E_{\beta I} - E_L$ becomes 1.1 kcal/mol instead of the 0.8 kcal/mol reported in ref 92.

Four additional energy minimized structures were found within the 1.1 kcal/mol range above the GEM. One, which also has the GEM energy E_L and satisfies 21 NOE distances, belongs to motif 3 (Table 4). The second structure, which pertains to motif 4, has an energy of 0.5 kcal/mol above the GEM. Two other structures with energies of 0.2 and 0.8 kcal/mol above the GEM satisfying 18 and 17 NOE distance constraints, respectively, are not related to any of the nine motifs defined in Table 4.

Thus, for three ASPs our optimization criterion can be considered as fully satisfied, because βII becomes the GEM structure, which deviates by only 1.1 kcal/mol or less from the energy of the βI structure; also, the motif 4 which was found necessary for obtaining all the 31 experimental distances for an error $\delta = 0.10$ (see Table 5) is represented with an energy of 0.5 kcal/mol above the GEM.

Notice that all the ASPs in the table are negative, indicating, as expected, that E_{tot} prefers structures with larger SASA than the optimal E_{GRO} structures. This includes the ASP of C ($\sigma_C = -45$ cal mol^{-1} Å^{-2}), which for water is positive (e.g., +12 cal mol^{-1} Å^{-2}; see Wesson and Eisenberg[63]) leading thereby to the hydrophobic effect. These authors also found $\sigma_O = -116$ cal mol^{-1} Å^{-2}, as compared to the present value $\sigma_O = -205$ cal mol^{-1} Å^{-2}; however, the comparison is not straightforward because of the significant difference in the radius of the spherical probe of water and DMSO, 1.4 vs 3 Å, respectively. For DMSO, σ_O is the most effective ASP, which can be understood in terms of electrostatic interactions between the molecule and the explicit model of DMSO defined within the GROMOS package. Thus, due to the relatively large negative partial charge (−0.38 eu) of the carbonyl oxygen, it prefers to be exposed to the solvent for interacting with the S atom and the two CH_3 groups of DMSO with partial charges of +0.139 eu and +0.160

Table 7. Optimized ASP Sets[a]

	σ_C	σ_H	σ_O	$E_{\beta II} - E_L$	$E_{\beta I} - E_L$	no. of LTDs
E_{GRO}	0	0	0	5.3	15.2	140 000
$MSEED_1$	−90	−90	−90	3.2	8.6	7 500
$MSEED_2$	−55	−55	−175	2.1	3.7	7 500
$MSEED_3$	−45	−145	−195	0.0	1.1	7 500
$ATAREA_1$	−85	−85	−85	2.4	6.7	5 000
$ATAREA_2$	−20	−20	−290	0.5	3.8	5 000

[a] The optimized set of ASPs (σ_i) presented are in cal mol^{-1} Å^{-2} and E is in kcal/mol. E_L is the GEM of E_{GRO} or E_{tot} (eq 3) for a given set of ASPs; $E_{\beta I}$ and $E_{\beta II}$ are the energies of the experimental βI and βII structures. The SASA and its derivatives are calculated either exactly with the program MSEED,[77] or approximately with ATAREA.[81] The criterion for selecting the optimized set of ASPs is described in the text. The number of different ASPs used appears as a subscript of MSEED and ATAREA in the first column.

eu, respectively. The partial charge, +0.28 eu of H(N) is smaller (in absolute value) than that of O; therefore, H shows somewhat weaker preference for exposure ($\sigma_H = -150$ cal mol^{-1} Å^{-2}) than O. Most of the C atoms in the present molecule (beside C$^\beta$ of Ser) are uncharged and their preference for exposure to the solvent ($\sigma_C = -45$ cal mol^{-1} Å^{-2}), is therefore relatively small, mainly stemming from LJ interactions with the DMSO molecules and from entropic effects. It should be emphasized that the ASPs are temperature dependent and the values presented in Table 7 were obtained at the experimental temperature of $T = 300$ K.

We also attempted to improve these results further by employing a fourth parameter, which in one set of trial optimizations was assigned to each of the partially charged side chain atoms of Ser (C$^\beta$, H, and O). The ASPs, which were varied by increments of 5 cal mol^{-1} Å^{-2} did not lead to a smaller energy gap between the βI and βII structures. Similar treatment of the C′ atoms did not yield improvement as well. This suggests that the number of ASPs that affect E_{tot} is limited. However, the insensitivity of E_{tot} might also stem from the relatively low SASA of these atoms as compared to the SASA of the whole molecule (see Table 6). In this context it should be pointed out that no ASP was assigned to N, because its SASA (based on $r_p = 3.0$ Å) was found to be zero in all the conformations generated. On the other hand, we found that in extended and α-helical structures the N and C′ atoms are more exposed to the solvent than in the present cyclic molecule. Therefore, better ASPs for these atoms will have to be derived in the future from structural data of linear peptides or larger cyclic peptides where there are less geometrical restrictions which prevent the exposure of these atoms. The sensitivity of the energy differences to changes in each of the parameters was presented in Figure 1 of ref 92. It would also be of interest to check the dependence of σ_i on the value of r_p, the radius of the DMSO spherical probe. We found that the molecular surface of peptide structures remain almost unchanged for $2.5 \le r_p \le 3.2$ Å, which is the range of values where r_p can be slightly varied on the basis of the radial distribution functions.[104,105] Because the molecular surface is the envelope of the volume from which the solvent is excluded,[110] it is expected that the only change required in σ_i over this range would be to scale them with respect to r_p, leaving E_{sol} (eq 1) unchanged.

Finally, we discuss the results obtained with ATAREA, the approximate method for calculating the SASA. For a single ASP we obtained the optimal value of −85 cal mol^{-1} Å^{-2} which is close to −90 cal mol^{-1} Å^{-2} obtained with MSEED, where the energy differences for the former are lower than those of the latter. This relatively good agreement is due to the fact that ATAREA calculates the SASA of the whole molecule with relatively high accuracy. However, for two ASPs, which requires calculating the SASA of individual atoms, the accuracy of ATAREA decreases as previously discussed, and therefore the two methods lead to significantly different optimized parameters. We also tried to apply ATAREA with three ASPs but could not improve on the two parameter results. The effort invested here to improve the accuracy of ATAREA stems from the efficiency of this method. However, further optimization is needed which would make ATAREA a useful tool especially for large proteins.

Summary

In the first part of this paper we carried out a structural analysis of NOE distance data for *cyclo*(D-Pro[1]-Phe[2]-Ala[3]-Ser[4]-Phe[5]-Phe[6]) in DMSO obtained by Kessler et al. and reconfirmed their conclusion, that two motifs, which can be distinguished by their βI and βII turns, coexist in thermodynamic equilibrium with a possibility of the involvement of a third one. Our analysis procedure, even though similar in spirit to previous ones, has some unique features. It is based on an extensive conformational search for low-energy-minimized structures performed with our efficient LTD technique, which is applied here for the first time to a cyclic peptide with side chains. The large pool of structures thus generated was scanned in a systematic way and only structures that satisfy a certain number of the NOE distance constraints were filtered out for participation in a best-fit analysis. This analysis was carried out assuming different experimental errors.

In the second part of the paper we developed a new method for determining atomic solvation parameters for a peptide in solvent, which was applied to the present hexapeptide in DMSO. Thus, the ASPs are optimized under the condition that the energy E_{tot} (eq 3) of the βI and βII structures become within the 2 kcal/mol range above the GEM(E_{tot}), and as close as possible to the GEM. Such an approach is feasible for a small molecule, where the GEM structure can be generated with high confidence by an extensive LTD search. Indeed, for the optimal set of three ASPs, the βII structure becomes the GEM structure and the energy of βI is only 1.1 kcal/mol higher. All the optimal ASPs are negative providing the expected energetic preference for open structures that is missing in E_{GRO}. These ASPs have been derived from structure−energy optimization without relying on free energy of transfer data of small molecules from the gas phase to DMSO. The present ASPs are optimal for the GROMOS force field and it would be of interest to see how much they change for other force fields. These ASPs can readily be used in a structural analysis of experimental NMR data, for example.

In general, the usefulness of this implicit modeling depends on the availability of efficient algorithms for calculating the SASA and its analytical derivatives. Here the program MSEED[77] was applied successfully; however, the approximate, hence very efficient method ATAREA[81] was also checked and was significantly improved by optimizing its parameters for the present hexapeptide. The method has been found to be reliable for calculating the SASA of the whole molecule, but insufficiently accurate for calculating the SASA of individual atoms; therefore, further improvement of ATAREA is needed.

The ASPs were derived here on the basis of energetic considerations alone. Entropic effects will be taken into account

(110) Richards, F. M. *Annu. Rev. Biophys. Bioeng.* **1977**, *6*, 151−176.

in a following study, where the βI, βII, and other motifs will be simulated by MD or MC based on E_{tot}; this will enable one to calculate the populations with the LS method, which might lead to a further refinement of the ASPs. Also, to verify the generality of the present ASPs, they should be derived independently from structural NMR data of other peptides of different size and amino acid content in DMSO. Finally, it should be emphasized that the present method for deriving ASPs can be applied to peptides in other solvents and to surface loops of proteins in water.

Acknowledgment. We thank Professor I. M. Navon for valuable discussions about optimization theory and for providing the present minimization program, and Professor H. A. Scheraga for providing the MSEED program. We thank Professor D. F. Mierke, Dr. G. Gemmecker, and Dr. M. Vásquez for helpful discussions. Molecular graphics images were produced using the MidasPlus program from the Computer Graphics Laboratory, University of California, San Francisco (supported by NIH RR-01081). We acknowledge support from the Florida State University (FSU) Supercomputer Computations Research Institute, which is partially funded by the US Department of Energy (DOE) under contract number DE-FC05-85ER250000. This work was also supported by DOE grant number DE-FG05-95ER62070. We acknowledge support by FSU through the allocation of supercomputer resources.

JA973124T