

Hydrophobic “Collapse” in a Cyclic Hexapeptide: Computer Simulations of CHDLFC and CAAAAC in Water

Carlos Simmerling[†] and Ron Elber^{*†}

Contribution from the Department of Chemistry, M/C 111, University of Illinois at Chicago, Box 4348, Chicago, Illinois 60680, and Department of Physical Chemistry, the Fritz Haber Research Center and the Israel Center for Structural Biology, The Hebrew University of Jerusalem, Jerusalem 91904, Israel

Received September 27, 1993. Revised Manuscript Received December 13, 1993^o

Abstract: The structure in water of two cyclic hexapeptides is determined. The locally enhanced sampling method that we developed is employed. Within the accuracy of the model the peptide CAAAAC does not have a unique structure while CHDLFC does. The driving force to structure in CHDLFC is the hydrophobic interaction between the phenylalanine and leucine. The implications of our results to short-range nucleation sites in protein folding are discussed.

I. Introduction

There are a number of reasons why considerable research is focused on structure determination of peptides in solution. First, many short peptides serve as signals to receptors. As such they are the target of pharmaceutical-related studies, searching for their biologically active conformation.¹ Second, short peptides with a strong tendency to a single structure may serve as nucleation sites in the process of protein folding. Nucleation sites are of significant interest as a way to solve the “Levinthal paradox”.² Third, atomic detail investigations in explicit solvent can be used to assess the reliability of reduced models of amino acids. Reduced models in which a whole amino acid is represented by one or two point particles are widely employed in investigations of protein folding.^{3–9} The second and the third points are the focus of the present manuscript.

Searches for peptides with significant tendency to a unique structure can benefit from the combination of lattice methods and off-lattice atomic detailed calculations. The two methods are complementary: The lattice searches are fast and capable of screening a large number of compounds. They are, however, approximate. In addition to a single (lattice) point representation of the amino acids they use only an implicit solvent model. The applicability and the refinement of the results to atomic models are not obvious. The off-lattice calculations provide a more detailed picture. However, they are computationally expensive and can only be pursued for a smaller number of systems. The peptide Cys-His-Asp-Leu-Phe-Cys (CHDLFC) was suggested by lattice searches (E. Shakhnovich, private communication) to have a strong tendency to a unique structure. Further support to the tendency to a unique conformation and atomic coordinates are provided by the locally enhanced sampling (LES)/simulated annealing calculations that we pursue here.

It is sad that convincing the reader of the importance of the

calculations does not make them easier. Simulations of solutes in explicit solvent are nontrivial. They are usually very large and require significantly more computational resources (compared to calculations in vacuum). In addition, the time scale for conformational transitions in the relevant degrees of freedom (the ϕ, ψ torsions) is quite long. This necessitates the use of many iterations to obtain long time trajectories in molecular dynamics simulations. Finally, very small peptides do not have a unique structure. This is an empirical but nevertheless quite general rule. It is therefore essential to reach a critical size of the peptide before a search for a unique structure can be attempted. Larger size does not make the search simpler. Cyclization is a significant help in forcing the peptide to a unique structure. However, even with cyclization, structure determination of peptides of the size considered here is difficult. In our investigations, straightforward molecular dynamics of the peptides in water did not give satisfactory results. By “satisfactory results” we mean that in the time scale that we could afford (a few nanoseconds) the trajectories did not converge to the neighborhood of a single structure. Normal room temperature trajectories are “stuck” in a single conformation for times longer than what is accessible to current simulation techniques. This is in contrast to the simulated annealing/LES runs that indicated a significant tendency to a unique structure. The combination of the LES protocol and simulated annealing that is used in the present manuscript made it possible to obtain converged, statistically meaningful data on a unique structure of a cyclic peptide. The LES protocol is a mean field approach that we introduced to biomolecules. It was employed in a number of applications and recently was also used in exact determination of global energy minima.¹⁰

Here we consider two cyclic hexapeptides: Cys-Ala₄-Cys (CAAAAC) and Cys-His-Asp-Leu-Phe-Cys (CHDLFC). The peptide CAAAAC is a reference that will help us to determine the importance of the side chains in structure determination. As we shall demonstrate later CAAAAC does not have a unique structure. The cyclization is done via a sulfur bridge at the first and the last cysteine residues. The terminals of the peptide chain were uncharged and they are not included in the cyclic part of the polymer. The N-terminal was NH-CO-CH₃ and the C-terminal CO-NH-CH₃. We demonstrate that, in accord with the lattice calculations, CHDLFC has a significant tendency to a unique structure while CAAAAC does not. We further address the question of the driving force to the observed structure and

[†] The Hebrew University of Jerusalem.

[†] University of Illinois at Chicago.

^o Abstract published in *Advance ACS Abstracts*, February 15, 1994.

(1) Testa, B., Ed. *Advances in Drug Design*; Academic Press: San Diego, CA, 1991; Vol. 21, Chapter 3.

(2) Levinthal, C. In *Moessbauer Spectroscopy in Biological Systems*; Proceedings of a meeting held in Alberton House, Monticello, IL, University of Illinois Press, 1969.

(3) Skolnick, J.; Kolinsky, A. *Science* **1990**, *250*, 1121.

(4) Covell, D. G.; Jernigan, R. L. *Biochemistry* **1990**, *29*, 3287.

(5) Honeycutt, J. D.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 3526.

(6) Friedrich, M. S.; Wolynes, P. G. *Science* **1989**, *246*, 371.

(7) Garell, T.; Orland, H. *Phys. Rev. Lett.* **1988**, *6*, 307.

(8) Shakhnovich, E. J.; Farztdinov, G.; Gutin, A. M.; Karplus, M. *Phys. Rev. Lett.* **1991**, *67*, 1665.

(9) Chan, H. S.; Dill, K. A. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 6388.

(10) (a) Roitberg, A.; Elber, R. *J. Chem. Phys.* **1991**, *95*, 9277. (b) Roitberg, A.; Elber, R. Modeling side-chains in peptides and proteins with the Locally Enhanced Sampling/Simulated Annealing Approximation. In *The protein Folding Problem and Tertiary Structure Determination*; Merz, K., Le Grand, S., Eds.; Birkhauser, in press.

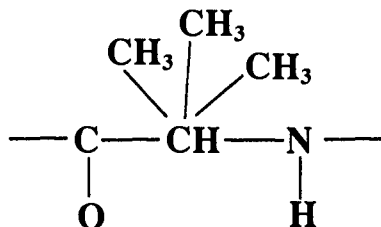


Figure 1. Side chain multiplication using LES: In the example three copies of the CH₃ group are attached to the CH. The copies do not see each other and they feel the exact force of the rest of the system. The "rest" feels the average force of the copies.

conclude that, at least in our example, the leading factor is the hydrophobic force and not internal or external hydrogen bonding.

This manuscript is organized as follows: In the next section the mean field/annealing scheme is described. Results are presented in the third section, and Discussions are given in the fourth part. Conclusions and final remarks are in the fifth part.

II. The Locally Enhanced Sampling/Simulated Annealing Method

1. General Considerations. The locally enhanced sampling (LES) method is a mean field approach that was specifically designed to obtain more sampling for a small part of the system. It is a variation on the trajectory bundles originally derived by Gerber et al. for gas-phase systems.¹¹ LES was employed in exploring plausible diffusion paths of ligands through protein matrices¹²⁻¹⁵ and in free energy calculations.¹⁶ The same protocol was also used for optimizing the structures of side chains in peptides and proteins.¹⁰ Formal properties of the LES approach were discussed in the past.^{10,12,17} Therefore we give here only a brief explanation of the final equations. However, we discuss in some length the differences between LES implementations, that is, the implementation of LES to the structure determination of peptides and to side chain modeling.¹⁰

Let the coordinates of the atoms in a single residue be r_i ($i = 1, \dots, 8$ —in our code the N-terminal and the C-terminal are treated as separate monomers) and let the coordinates of the solvent molecules be r_s . The classical equations of motion for a solvated hexapeptide are

$$M_i d^2 r_i / dt^2 = -d/dr_i U(r_1, \dots, r_8, r_s) \quad (1a)$$

$$M_s d^2 r_s / dt^2 = -d/dr_s U(r_1, \dots, r_8, r_s) \quad (1b)$$

where M is the diagonal mass matrix, t is the time, and U is the potential energy that includes all the interactions between the atoms.

In LES the equations of motion are replaced by mean field equations. Pictorially we consider the system as composed of fragments. We multiply the fragments and let the copies interact with each other in an average way. For side chain modeling,¹⁰ we can multiply the side chains only (from the C_β to the end of the chain). In Figure 1, we show an example in which only the side chains are multiplied. For each C_α several copies of the side chain are attached. The multiple copies of the same side chain do not see each other and they interact with the rest of the system in an average way. For example, a backbone atom feels all the forces of the multiple side chains, added up, and then divided by

the number of copies. A similar approach is taken in structure determination of peptides. We multiply either a whole amino acid (using two different "cutting" points along the backbone, see below), pairs of amino acids, or four monomers. The advantages and disadvantages of different multiplication choices will be discussed in the computational protocol. We found that pairs of amino acids work best for the present system. Let the coordinate of the k -th copy of the i -th amino acid pair be r_i^k , then the equations of motions for the hexapeptides are modified to be

$$M_i / N_i d^2 r_i^k / dt^2 = -\sum 1 / (N_1 N_2 N_3 N_4) d/dr_i^k U(r_1^m, r_2^k, r_3^l, r_4^n, r_s) \quad (2a)$$

$$M_s d^2 r_s / dt^2 = -\sum 1 / (N_1 N_2 N_3 N_4) d/dr_s U(r_1^m, r_2^k, r_3^l, r_4^n, r_s) \quad (2b)$$

Where N_i is the number of copies of the i -th pair of amino acids. M_i and M_s are the mass matrices of the i -th amino acid pair and the solvent, respectively. Note that we do not multiply the solvent molecules. A very significant part of the computational effort is associated with the solvent. Therefore, the computational effort using eqs 2a and 2b is not very different from using eqs 1a and 1b. Hence the advantages of LES, which include reduction of barrier heights and more statistics for alternative conformations,¹⁰ are obtained at low additional computational cost. To obtain some qualitative insight to the properties of the multiple copy method it is useful to point out the connection to mean field approximations. If one cancels the $1/N_i$ from both sides of eq 2a, the mean field character of the equations is self-evident. For example, the force a solvent molecule feels is a sum over all copies of the peptide fragments divided by the number of copies. Hence the solvent feels the mean force (field) exerted on it by the peptide copies. A similar argument holds also for the amino acid pair that feels a mean force of the copies of the other pairs.

Equations 2 are used in simulated annealing runs. We start with high temperature and slowly cool the system by velocity scaling. Optimal structure is obtained at temperatures low enough such that no more torsional transitions are observed during the simulation period.

There are a number of general properties of LES that we analyzed in the study of side chain modeling.¹⁰ The general features are applicable (of course) to the study of the peptide as well: (a) We proved that the barrier heights separating different minima are reduced in LES. This makes the annealing significantly easier and is a key element in the success of the mean field protocol in peptide modeling. (b) We have shown that the global energy minimum of the exact system coincides with the global energy minimum of the LES system. This means that annealing using the mean field energy will provide the exact answer. (c) The LES protocol suggests a self-consistency check that is not available in the regular annealing (RA). If the final distribution of the configurations of copies is broad the cooling was executed too rapidly. As was shown in ref 10, the global energy minimum occurs when all the copies occupy the same position in space at the global energy minimum of the original system. Therefore a broad distribution of configurations of copies cannot be the global energy minimum. This helps in eliminating bad final structures. In the present calculations we further assess the reliability of the obtained minima by repeating the runs. This is similar to RA with the additional advantages of LES. Since the solvent simulations are expensive we prefer to make systematic comparison between the LES and the RA using vacuum calculations first. As is demonstrated below in vacuum, LES is clearly a superior approach to regular simulated annealing. We continue by pursuing several runs of LES and regular simulated annealing in solvent. LES is better than the single copy calculations also in solvent; however poorer statistics and less systematic exam-

(11) Gerber, R. B.; Buch, V.; Ratner, M. A. *J. Chem. Phys.* **1982**, *77*, 3022.

(12) Elber, R.; Karplus, M. *J. Am. Chem. Soc.* **1990**, *112*, 9161.

(13) Czerninski, R.; Elber, R. *Proteins* **1991**, *10*, 70.

(14) Gibson, Q. H.; Regan, R.; Elber, R.; Olson, J. S.; Carver, T. E. *J. Biol. Chem.* **1992**, *267*, 22022.

(15) Chiancone, E.; Elber, R.; Royer, W. E., Jr.; Regan, R.; Gibson, Q. H. *J. Biol. Chem.* **1993**, *268*, 5711.

(16) Verkhivker, G.; Elber, R.; Nowak, W. *J. Chem. Phys.* **1992**, *97*, 7838.

(17) Ulitsky, A.; Elber, R. *J. Chem. Phys.* **1993**, *98*, 3380.

Table 1. A Summary of the Different Simulations for CAAAAC and CHDLFC That Were Carried Out

simulation type ^a	LES? ^b	length ^c	no. ^d	successes ^e
CAAAAC				
vacuum	no	4 ns	3	0
vacuum	no	10 ns	2	1
vacuum	yes	250 ps	2	0
vacuum	yes	500 ps	2	1
vacuum	yes	1 ns	1	1
vacuum	yes	2 ns	11	11
vacuum	yes	4 ns	1	1
solution, 21 Å	yes	250 ps	2	0
solution, 21 Å	yes	1 ns	7	0
solution, 21 Å	yes	2 ns	2	0
solution, 21 Å	yes	4 ns	1	0
CHDLFC				
vacuum	no	500 ps	1	1
vacuum	no	2 ns	3	2
vacuum	no	4 ns	1	1
vacuum	yes	500 ps	2	2
vacuum	yes	2 ns	9	9
vacuum	yes	4 ns	3	3
solution, 21 Å	yes	625 ps	1	0
solution, 21 Å	yes	1 ns	2	0
solution, 21 Å	yes	1.5 ns	2	2
solution, 21 Å	yes	2 ns	4	4
solution, 21 Å	yes	4 ns	2	2
solution, 26 Å	yes	750 ps	3	1
solution, 26 Å	yes	1 ns	2	2
solution, 26 Å	yes	1.5 ns	1	1
solution, 26 Å	yes	2.0 ns	1	0
solution, 26 Å	yes	2.8 ns	2	2
solution, 21 Å	no	2 ns	4	3
solution, 21 Å	no	4 ns	1	0

^a Simulation type (vacuum, solvent—box size). ^b Is the locally enhanced sampling (LES) method employed (yes/no)? ^c Length of the simulation. ^d Number of repeats. ^e Number of successes in locating a "good" structure. ^f This run has multiple side chain positions at the end and therefore was not considered. See text for more details.

ination of the simulation conditions made it more difficult to compare LES and RA in solution.

2. Computational Protocol. In this investigation we have made a considerable number of trials and tests: For both peptides CAAAAC and CHDLFC we pursued the calculations in vacuum and in water. The investigations were for five LES copies or with a single copy (RA). We also studied the effect of different partitioning of the mean field subspaces, i.e., multiplying fragments of a single residue (in two ways), multiplying fragments of two residues, and the multiplication of units of four monomers. We also tested different annealing schedules and different sizes of the solvent box. The simulation was repeated in the presence and the absence of a counterion. Finally, we tested how long it takes for a single copy trajectory and a LES trajectory, both of them at room temperature, to "fold" to the correct structure starting from an incorrect one.

Clearly, a complete discussion of all the different tests can discourage even the readers with the largest storage devices. We therefore discuss various tests only when pertinent. In Table 1 we give a brief summary of the calculations, providing the peptide, the simulation type (vacuum/solvent and box size), the length of the simulation in picoseconds, the number of different trajectories used to study this particular setup, and the number of trajectories that succeed in locating a minimum that we consider "good". The criteria for locating good minima will be discussed below.

All the calculations employed the program MOIL.¹⁸ The force field in MOIL is the combination of the AMBER force field (for most of the covalent part of the potential),¹⁹ the OPLS nonbonded parameters,²⁰ and the improper torsion parameters

of CHARMM.⁹ The water model is TIP3P.²² The equations of motion are solved using the velocity Verlet. The individual water molecules were kept at a rigid geometry using a velocity version of the SHAKE algorithm (RATTLE²³) in a matrix form. Since only the water molecules were constrained, each water molecule has its own 3×3 matrix of distance constraints that is easy to manipulate. Rather than iterating the bonds that are hard to converge in a triangle configuration (the H—H distance is included) we iterate the 3×3 matrices. This process was found to be numerically more stable compared to bond iteration usually done in SHAKE. Numerical stability was an issue of a major concern at the high temperatures of the annealing. At temperatures of thousands of degrees the bond formulation of RATTLE²³ simply did not work. The kinetic temperature was maintained using velocity scaling at each step. The cutoff distance was 8 Å. Potential energy fluctuations due to use of cutoff distance were reduced by using a neighbor list based on a larger cutoff.¹⁸ The cutoff was 9.6 Å in the small box of water and 8.5 Å in the larger box. In the vacuum runs the highest temperature during the annealing was 5000 K, while in the solvent runs the peptide temperature was at most 7500 K. At that temperature (7500 K) numerical stability problems with the water molecules make the simulation difficult. Therefore, the highest temperature for the solvent that we used was only 2000 K. The water molecules were maintained at that temperature by velocity scaling until the peptide was cooled to the same temperature starting from 7500 K. Then a uniform temperature for the whole system was employed. To further increase the numerical stability of the annealing trajectories the mass of all the particles was uniformly set to 10 amu. This should not affect (of course) any property related to the potential such as the global energy minimum.

In the simulations with solvent two sizes of the water box were investigated: 21 and 26 Å. Periodic boundary conditions were used. The smaller box includes 252 water molecules in the simulation of CAAAAC and 232 for CHDLFC. The large box was used only in the investigations of the larger peptide (when side chains are considered—CHDLFC) and included 480 water molecules.

The simulations in vacuum were of considerable value in fine tuning the optimization parameters, in comparing the LES approach to the regular annealing, and as a reference, to examine the changes in the peptide structure once the solvent was introduced. In LES more statistics are obtained, and the potential is effectively smoothed when the different copies have different conformations. Each of the copies of a multiplied part feels all the copies of the other parts in a smeared way. Thus, all the forces between the multiple copies of different sites are added up and finally divided by the number of combinations (eq 2). This averaging provides the potential smoothing and barrier reduction we discussed previously. Of course at sufficiently low temperatures we expect the copies to "collapse" to a single minimum. However, at a high temperature in which large fluctuations in the peptide structure are desirable, broad distribution of the copies is useful. Careful choice of the mean field coordinates (i.e., the part of the system to be multiplied) is therefore important. This issue is easily exemplified by our trials of coordinate choice. In

(18) Elber, R.; Roitberg, A.; Simmerling, C.; Goldstein, R.; Verkhivker, G.; Li, H.; Ulitsky, A. MOIL: A molecular dynamics program with emphasis on conformational searches and reaction path calculations; NATO conference proceeding on Statistical Mechanics and Protein Structure; Cargese; Doniach, S., Ed.; Plenum, in press. The program is available via anonymous ftp from 128.248.186.70.

(19) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S., Jr.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765. (20) Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657.

(21) Brooks, B. R.; Brucoleri, R. E.; Olafson, B. D.; Sates, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.

(22) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D. *J. Chem. Phys.* **1983**, *79*, 926.

(23) Anderson, H. C. *J. Comput. Phys.* **1983**, *52*, 24.

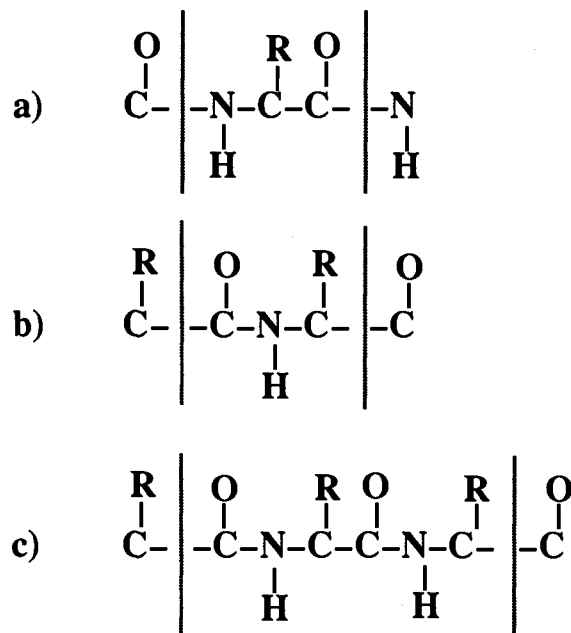


Figure 2. Different multiplication schemes of the polypeptide chain. The copies are defined as follows: (a) a single amino acid starting with the N—H group and ending with the C=O; (b) a single amino acid with a complete amide plane included as an integral part of the copy; and (c) two amino acids as the multiplied unit. Two complete amide planes are included in the multiplied segment.

Figure 2 we show the different segments employed as the base for the "multiplication scheme". Parts a and b of Figure 2 correspond to a single peptide unit, truncated either in the middle of the peptide unit or at the C_{α} -C bond, and part c corresponds to our best choice in which two peptide units form the basic structural segment that is multiplied. Two peptide units are the smallest fragment of a protein chain that can be moved significantly leaving the rest of the chain invariant. The empirical rule is that the larger are the units, the more independent are the positions of the different copies. Independence of copies is not, however, the only factor. A large multiplied unit that consists of four monomers was also tried and gave considerably worse results as compared to protocol c. For large fragments the statistics are poorer and the barrier heights for conformational transitions within the fragment are not reduced. Some compromise between copy independence and the possibility of smoothing critical parts of the potential energy is therefore required. To demonstrate that choice c indeed makes it possible to have large dihedral angle fluctuations, we plot in Figure 3 the variance of $\cos(\phi)$. The fluctuations were computed using a room temperature trajectory and an average over all the backbone dihedral angles is shown. The "spread" of $\cos(\phi)$ was calculated for individual time windows (averaged over the LES copies) as a function of time. As is quite clear from the figure, the fluctuations of $\cos(\phi)$ using pairs of amino acids are substantial and larger than those obtained by the alternative choice of coordinates. In addition to the fluctuations' plots we also pursued different annealing runs using each of the above multiplication schemes. The results of scheme c were clearly advantageous in terms of configuration sampling and in terms of success rate in locating the global energy minimum. Therefore, scheme c was employed as the coordinate of choice for LES multiplication throughout the work presented in this paper. The last comment is related to the effect of the system size on the computational efficiency. Clearly, as the system becomes larger, the calculations require more computational resources. Consider for example the peptide CAAAAC. It consists of 44 particles which in the LES representation employed here (5 copies) are transformed to 220 particles. In a worst case scenario, the computational effort

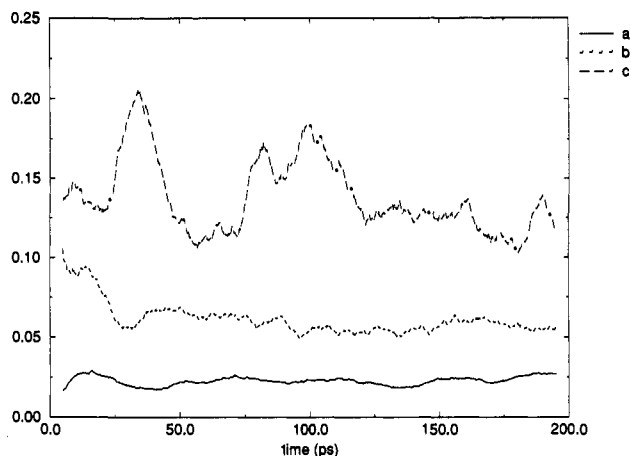


Figure 3. The average fluctuations of $\cos(\phi)$. ϕ is any of the backbone dihedral angles for the cyclic peptide CAAAAC. The fluctuations are for the LES copies at a single time point, calculated as a function of time. (a) Solid line—multiplication scheme a of Figure 2. (b) Middle line with short dashes—multiplication scheme b. The top line with long dashes—multiplication scheme c of Figure 2.

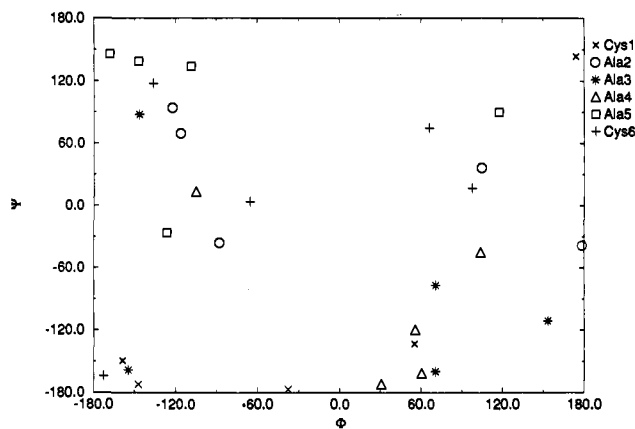


Figure 4. High-temperature distribution of the ϕ, ψ angles of a single multicopy structure. This is for a vacuum run of CAAAAC.

in LES as compared to the single copy is growing as the square of the size. That is, a comparable simulation in LES may require 25 times more computational resources as compared to a single copy simulation. However, as we shall demonstrate below the single copy annealing requires a trajectory length that is 20 times longer as compared to LES (10 ns versus 500 ps). Comparable CPU time is therefore required to complete a single copy trajectory (regular annealing) and a single trajectory of LES. One should bear in mind, however, that in LES many more trajectories are obtained. In a single LES run we obtain $5^4 = 625$ mean field trajectories (counting all possible combinations). This provides the self-consistency check we mentioned previously and also multiple answers in a single run.

Furthermore, in the solvent simulation, on which our interest is focused, the computational effort is dominated by the water molecules and the growth in computational effort when changing from a single copy to five LES copies is significantly less than quadratic. Thus, for the more complex and interesting problems LES provides significant improvement as compared to the regular annealing. This is also shown in Table 1 and will be further demonstrated in the Results section.

III. Results

(a) Simulations in Vacuum. In Figure 4 we show the distribution of the ϕ, ψ dihedral angles for a high-temperature, LES structure of CAAAAC. The values of the dihedral angles of all amino acids and all copies are shown. A stick model of the same LES

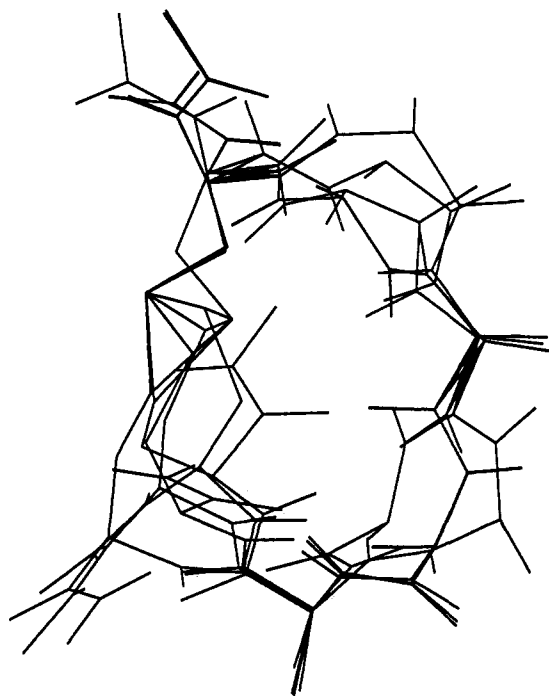


Figure 5. High-temperature LES structure of CAAAAC. Five copies with multiplication scheme *c* were employed.

conformation is given in Figure 5, demonstrating the significant differences between the multiple copies. Another useful quantity to follow is the time/temperature history of the different torsions. In Figure 6 we show the position of Ψ_1 as a function of time (or as a function of the decreasing temperature which is equivalent) for all five copies. It is evident that at high temperatures the five copies explored all the available conformation space. It is also evident that an abrupt transition to considerably narrower distribution occurs at temperatures below 2000 K. At temperatures lower than the last, the copies are "stuck" in the configuration determined by the 2000 K transition. Similar behavior is observed for other dihedrals. However, the freezing

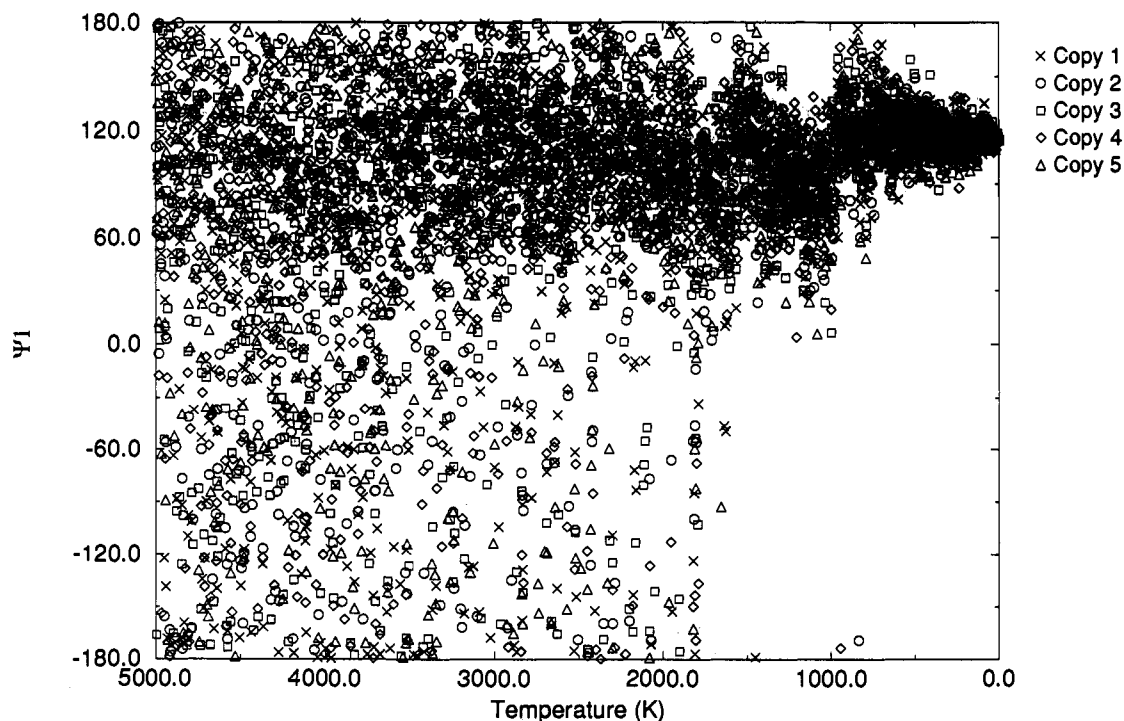


Figure 6. Time and temperature history for an annealed trajectory of the peptide CAAAAC. The Ψ_1 values of each of the five LES copies are shown as a function of temperature (time). Different symbols are used for different copies.

temperature is different from torsion to torsion. A final structure adopted by the five copies is shown on Figure 7. Only a single structure is seen since all the copies occupied exactly the same position in space. At least as far as our self-consistency check is concerned the last configuration is satisfactory. Another useful test of an annealing protocol is a time history of minimized structures. Structures along the annealed trajectory are picked and minimized. The energies of the minimized structures for a LES run are shown in Figure 8b. Evidently the quality of the minimized structure increases (on the average) as a function of time, until it finds the lowest energy minimum of the complete run. This is what we expect from good simulated annealing runs. All the LES trajectories of a cooling period longer than or equal to 500 ps found the same structure. The terminals have a hydrogen bond between them (Figure 7). We never found energy lower than the value determined by the "good annealing". While this does not guarantee that a lower energy structure does not exist, the statistics are quite convincing. This is to be contrasted with regular annealing that requires 10 ns to find the same minimum. The single-copy results were also less consistent. This is also clear from a plot of the minimization history for a regular annealing run (Figure 8a). The energy of the minimized structures did not show a systematic improvement (on the average), an improvement that was observed in the LES simulation.

The main conclusion from this study therefore was that a single global energy minimum exists for CAAAAC in vacuum but that it is hard to find it using regular annealing protocol. In contrast it was easily found in a systematic way using the LES protocol. This is a clear demonstration that the LES approach is effective in modifying the energy surface such that searches are done more easily. Note that a different feature of LES was used in the exploration of plausible diffusion pathways.¹²⁻¹⁵ In ligand diffusion, instead of flattening the energy surface the enhanced statistics (multiple ligand trajectories) were employed.

Simulated annealing in vacuum was also pursued for CHDLFC. In Figure 9 we show a high-temperature structure of this peptide. A significant variation in the backbone conformations of the different copies is shown similarly to CAAAAC. In addition large fluctuations in the side chain positions are also evident. In fact, different side chain conformations for different copies are

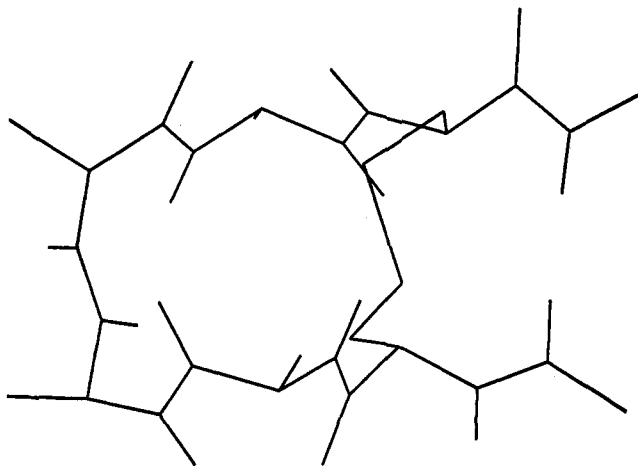


Figure 7. The global energy minimum for the peptide CAAAAC in vacuum. Note the hydrogen bond between the amide planes of the N and C terminals.

noticeable throughout the annealing and we were unable to find a single structure for the side chains in this system (except asp, which was in a single configuration every time). In Figure 10 we show the time history of the χ_1 dihedral angles of the five histidine copies. Similarly to Figure 6, at high temperatures the distribution of angles is very broad and at lower temperatures the distribution becomes narrower in a sharp transition. However, in contrast to the backbone fluctuations in CAAAAC, the histidine copies settle in **three** different conformations. The same phenomenon of multiple side chain conformations at the final LES structure is obtained for the phenylalanine and the leucine residues. A summary of all the alternate side chain conformations found for CHDLFC in the vacuum runs is given in Figure 11b. Note, however, that the backbone structure of the "cycle core" HDLF is unique. The same backbone structure is obtained in all the runs. In Figure 12 we show one example of an optimized LES structure with multiple side chain configurations. It is obvious

Table 2. Energy Values for the Lowest Minima of CHDLFC That Differ in Side Chain Orientation^a

copy set no.	energy (kcal/mol)
1	-289.93
2	-289.91
3	-287.52
4	-286.39
5	-286.02

^a The results are for a single LES run in vacuum.

that under these conditions the space available for side chain packing is quite broad. The multiple conformations found in the LES annealing suggest one of the following: (a) the annealing was not executed properly, or (b) the energy surface has many almost degenerate minima. Thus, the energy gap between the lowest energy minimum and other minima is small such that a unique energy minimum is not a useful concept. For b to be correct, the energy gap should be smaller than the thermal fluctuations of the system. For CHDLFC with 180 internal degrees of freedom the thermal fluctuations are estimated as $k_B T/2 \cdot [180]^{0.5} = 4$ kcal/mol at room temperature. In Table 2 we provide energy differences for some structures with alternative side chain configurations. The different configurations are accessible at room temperature according to the above estimate. Besides thermodynamics we wanted to investigate the kinetics of side chain transitions. The following calculation was therefore pursued: Starting from one of the optimized conformations a single copy trajectory at room temperature was initiated. In a period of 1 ns we found that the backbone conformation was unchanged (rms of less than 0.3 Å) while the histidine and the phenylalanine underwent large transitions. The picture emerging from this study is of very different time scales for backbone and side chain transitions. Barriers for side chain transitions are lower than the corresponding barriers for backbone dihedrals. The energy surface of the side chains is not only degenerate (at least in vacuum) but also easily accessible.

Another test of the quality of the annealing is to follow the energies of the minimized structure as a function of the annealing

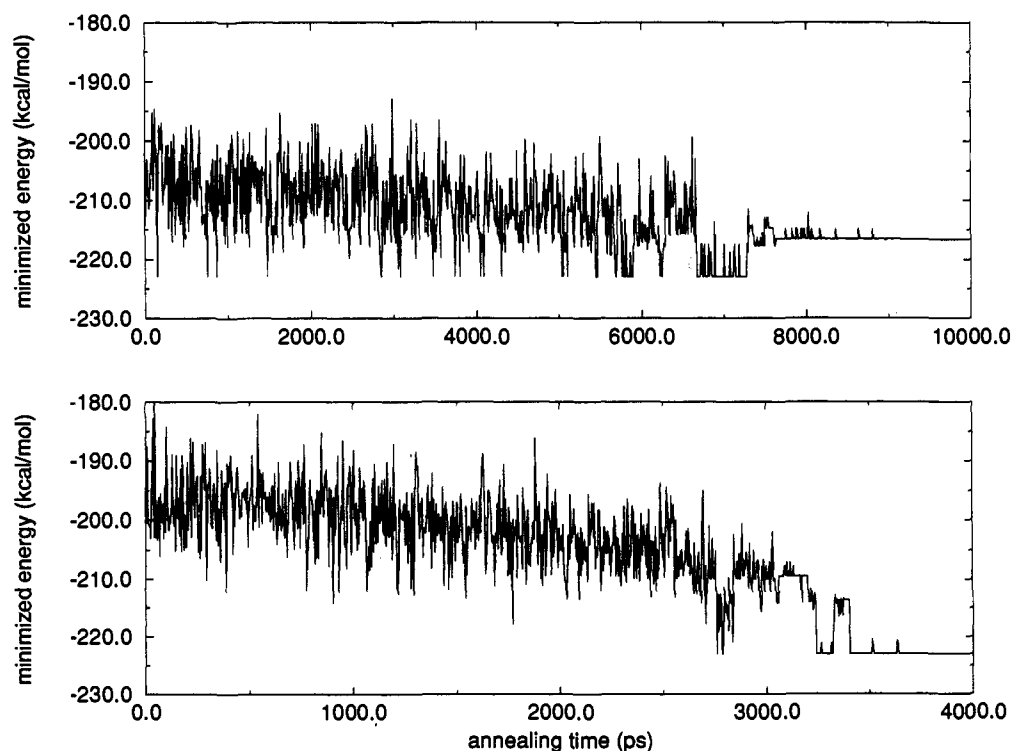


Figure 8. Time or annealing history of minima. Trajectory structures were directly minimized, using the conjugate gradient method, to nearest local energy minima. The energies of the minima are plotted as a function of the temperature of the annealing run. (a, top) Vacuum run for a single copy of CAAAAC. (b, bottom) Vacuum run for five LES copies of CAAAAC.

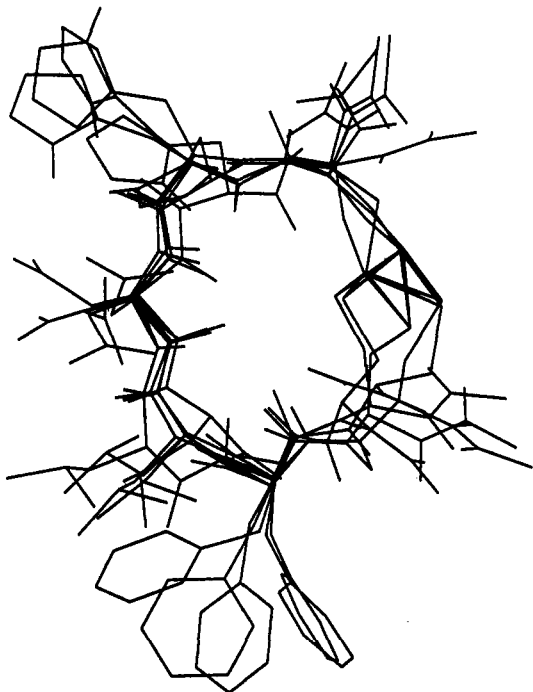


Figure 9. A stick model of a high-temperature structure of CHDLFC. Five copies using the multiplication scheme *c* are shown. The structure is from a vacuum simulation.

history (Figure 13). The minimized energy for CHDLFC is a decreasing function of time (on the average), suggesting that the simulation is sound. We finally comment that the main drive to the unique backbone structure (which is different from the structure of CAAAAC) is hydrogen bonds. The hydrogen bonds are between backbone groups and specific side chains, especially the aspartic acid which is strongly hydrogen bonded, either to the Cys at the C terminal or to the backbone of the phenylalanine. The two peptides CAAAAC and CHDLFC form in vacuum a distorted β turn which is more compact than usually found (Figure 14). It is different however in CAAAAC and CHDLFC due to

the presence of side chain hydrogen bonding in the last peptide and strong backbone hydrogen bonding in CAAAAC.

(b) Simulations in Solvent. As before we consider first the peptide CAAAAC. We placed the peptide in a 21 Å cubic box of water molecules. A total of 252 water molecules (of the type TIP3P²²) were used. The annealing was pursued as described in section III. In Figure 15 we show the distribution of the ϕ, ψ dihedral angles for the five copies employed at a high-temperature structure. In Figure 16 a stick model of the peptide and the water is provided. Similarly to the vacuum calculations, on the level of visual inspection, the conformations of the five copies seem appropriately randomized. In Figure 17 we provide a more quantitative measure of the exploration of phase space by plotting the distribution of Ψ_1 as a function of the annealing temperature. At the highest temperatures of the run the Ψ_1 dihedral angle was indeed sampled properly. In a similar spirit to the vacuum simulations there is a relatively sharp transition at lower temperatures to a fixed value of Ψ_1 . However, this time in disagreement with the vacuum calculations, we did not find a unique structure. Almost all the different runs (excluding one pair of the 15 trajectories) yield dissimilar structures. The lengths of the simulations were varied from 250 ps to 2 ns with essentially the same computational observation: A unique structure was not found. The C_α rms of any pair of the "optimal" structures (excluding one pair) was at least 1 Å. For a peptide of this size, this deviation is large. We further excluded the possibility that the only pair of similar structures corresponds to the global free energy minimum. This we did by testing for the stability of this structure under thermal conditions. Of interest are transitions between conformations of CAAAAC at room temperature. We therefore pursued a room temperature simulation starting from the only optimized configuration that appears twice. A single copy of the peptide was employed in that simulation. In Figure 18 we show the C_α rms as a function of time. At around 1 ns, the rms climbs to 1 Å. CAAAAC therefore hops rapidly between the large set of almost equivalent configurations that are available to it. Since we did not obtain a unique structure using LES we did not try to look for one using a single copy. The last is even

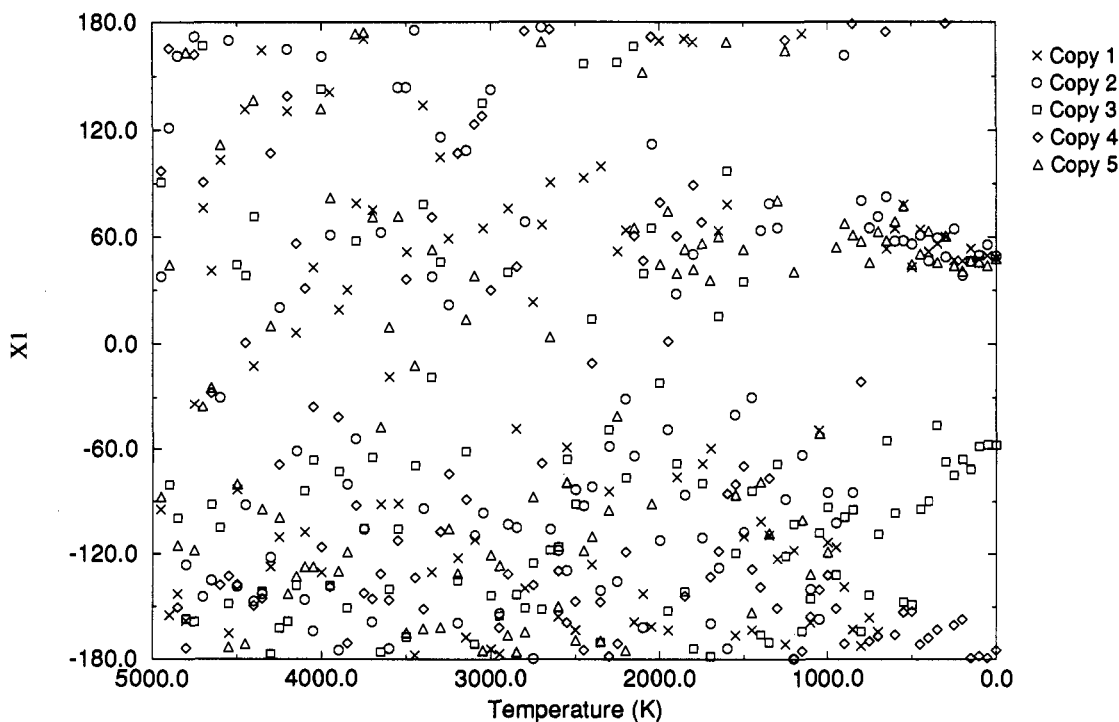


Figure 10. Annealing history of the five copies of the histidine residue in CHDLFC. The χ_1 dihedral is plotted as a function of time/temperature. The run is in vacuum. Note that the multiple copies did not fall to a unique structure at the end of the simulation.

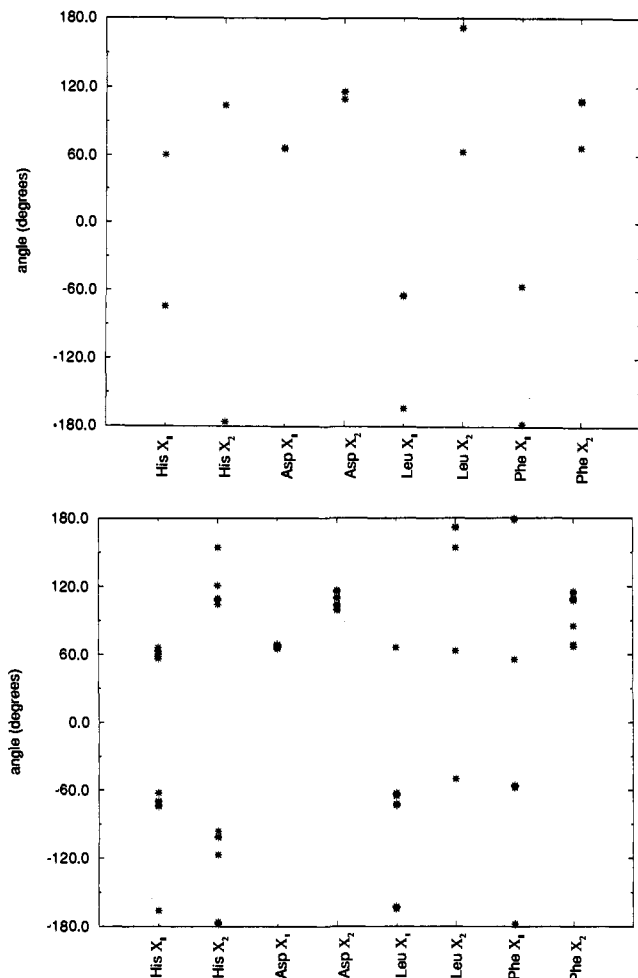


Figure 11. The alternative conformations of the side chains in CHDLFC. (a, top) Summary of different side chain configurations sampled from a single LES run (the trajectory is different from the run used in Figure 10). (b) Different side chain positions extracted from all the runs. Note that a single LES trajectory is covering a significant part of the accessible conformations.

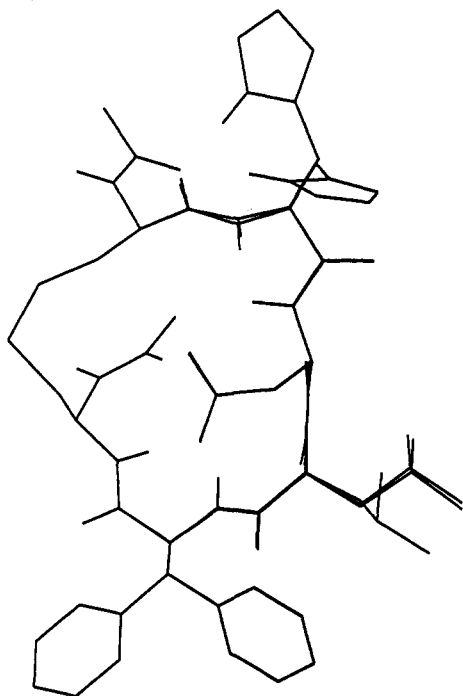


Figure 12. A stick model of an optimized structure of CHDLFC. Note the multiple solutions for side chain positions which is typical for a vacuum run.

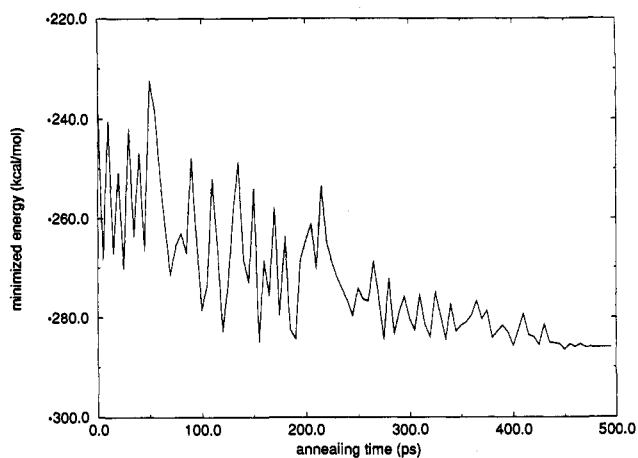


Figure 13. An annealing history of local energy minima. It is the same as in Figure 8, except that this time it is for the peptide CHDLFC with five LES copies.

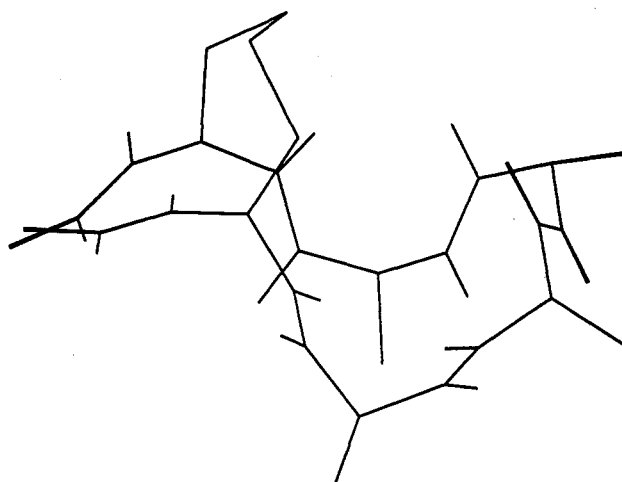


Figure 14. Side view of the compact structure of CAAAAC—twisted β turn—in vacuum.

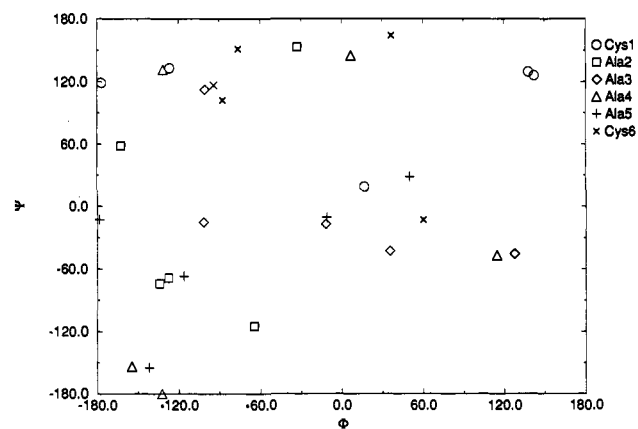


Figure 15. CAAAAC in water: The distribution of the ϕ, ψ dihedrals at high temperatures. Five copies were employed and the results are from a single time frame.

less likely to find the global free energy minimum, if such a minimum exists.

The conclusion must therefore be that the solvent eliminates the deep minimum of the free energy surface of CAAAAC that exists in vacuum. Another comment is related to the solvent structure around the peptide. Since the solvent is very effective in replacing internal hydrogen bonds, one may expect strong hydrogen bonds between the water and the peptide. However,

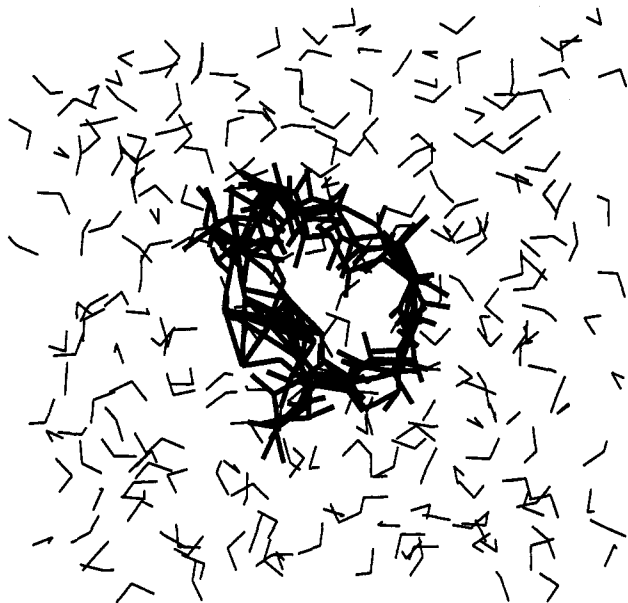


Figure 16. A stick model of five copies of CAAAAC in water. A snapshot in time at 2000 K is shown.

the water molecules around the peptide are not structurally well-defined. Thus, we do not find specific and strong hydrogen bonds between the peptide and the water. Nevertheless, the ill-defined hydrogen bonding to the water molecules is sufficient to disrupt the unique structure previously observed in vacuum. To another technical point, the high degeneracy in the solvent structure makes it more difficult to assign an energy to the final configuration. The alternative packing of the water molecules around the peptide causes very large energy fluctuations of individual optimal structures, even if the peptide structures are very similar. We therefore assign peptide conformations in solvent by structure only. A structure is assumed to be "unique" only if it is observed many times in repeated annealing runs and if it is stable during a room temperature trajectory that uses a single copy. Since the annealing of CAAAAC in solvent did not give a structure that satisfies these criteria we assume that a unique structures does not exist.

Needless to say, we repeated the solvent simulations for the peptide CHDLFC. Keeping the tradition we show in Figure 19 a stick model of a high-temperature structure of CHDLFC in a water box. The annealing was done in a similar way to CAAAAC in a 21-Å water box, starting from 7500 K for the peptide and 2000 K for water. The peptide was cooled to 2000 K keeping the water temperature constant. From 2000 K down, both the peptide and the waters were maintained at the same decreasing temperature. The annealing periods varied from 250 ps to 4 ns. Note also that due to the presence of the aspartic acid the system is negatively charged at normal pH. We therefore repeated the calculations also with a neutral system placing a counterion (sodium) in the water box. However, since the differences in the results of the annealing were small we shall not discuss the last simulations. In Figure 20 we show temperature/time history of the χ_1 dihedral angle of the histidine copies. As we have seen before the torsion effectively covers the conformational space available to it. An important difference, however, is that it freezes at a lower annealing temperature compared to the temperatures in which the backbone freezes. This suggests a lower kinetic barrier for trapping the side chains in a specific configuration as compared to locking of the backbone conformation. Two final annealed structures are shown in Figure 21. The two structures, while clearly not identical, share many common features. The common features include the elongated arrangement of the peptide cycle, the closely packed configuration of the leucine and the phenylalanine side chains, and the perpendicular configuration (pointing to the solvent) of the aspartic acid. The phenylalanine and the leucine are tightly packed against each other and there is no space for a water molecule between the two hydrophobic residues. The N and the C terminals that are external to the peptide ring do not show a specific structure and can be found in many configurations. To better quantify the common features of the structures we calculated the rms of structural segments (Figure 22). Evidently the C_α 's of the HDLF segment of the peptide have even lower rms, strongly supporting the existence of a unique conformation of the peptide ring. All the conformations obtained from simulation lengths exceeding 1 ns and in which the copies fall to a single position (13 runs) have preferred packing of the phenylalanine and the leucine residues. The single run that did not end in an "optimal" structure has multiple side

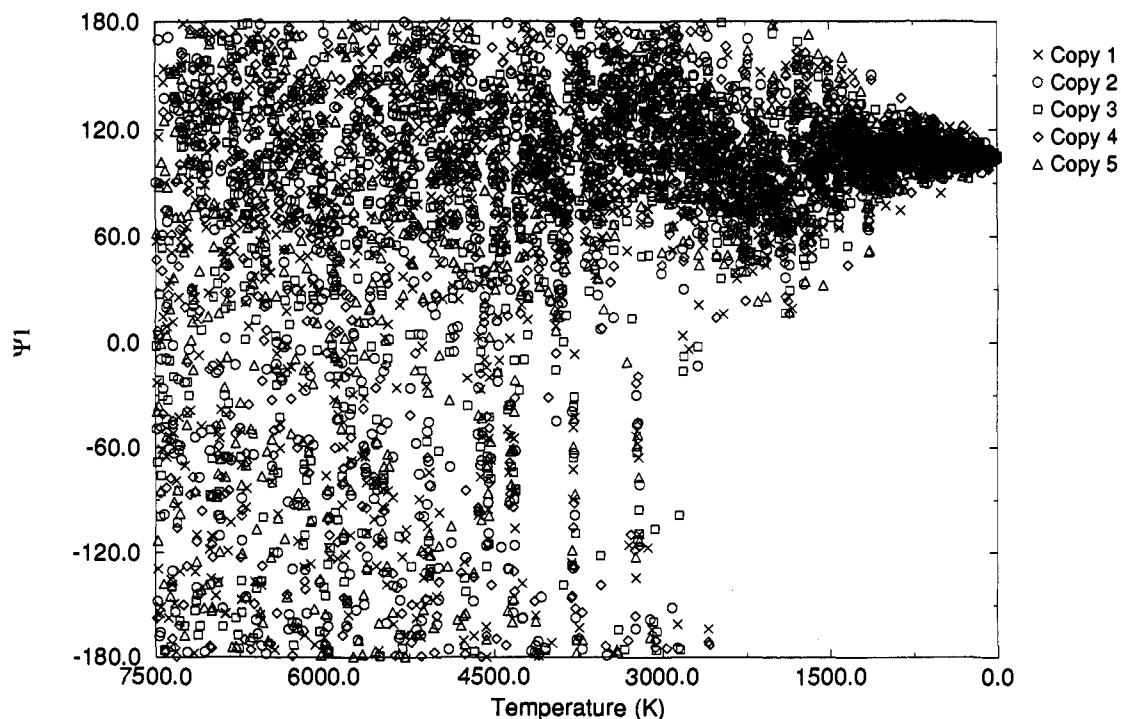


Figure 17. Annealing history of solvated CAAAAC. Ψ_1 of the five LES copies are shown.

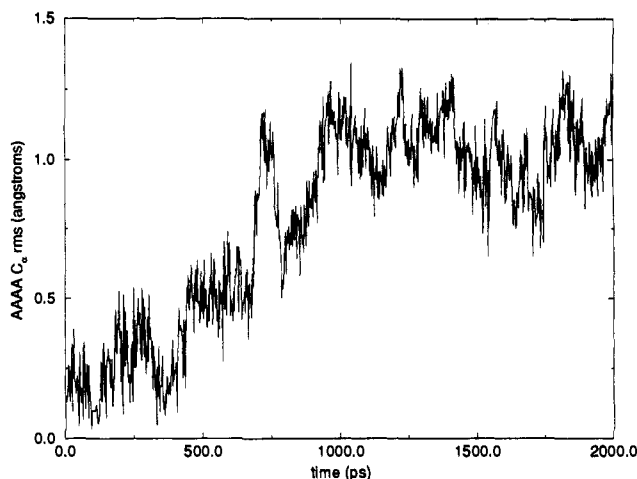


Figure 18. A room temperature trajectory of CAAAAC in water starting from one of the "optimal" structures. The C_{α} root mean square deviation is shown as a function of time. Note that the "optimal" structure is not stable and around 1 ns a large structural change is observed.

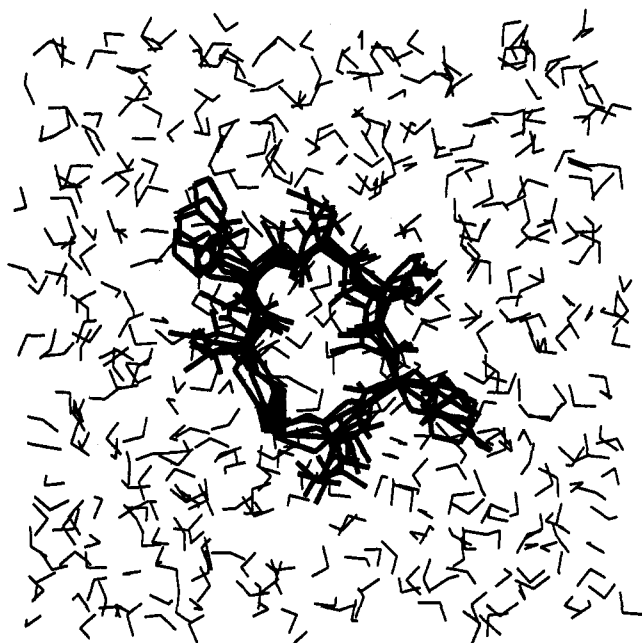


Figure 19. A stick model of a high-temperature structure of CHDLFC in water. Five copies are employed.

chain conformations at the end of the run, a warning sign that this water simulation is problematic. The proximity of the two hydrophobic residues is demonstrated by examining the distances between the leucine and the phenylalanine. The distances between the center of mass of the above two residues are histogrammed and shown in Figure 24a. The distances were calculated for all optimized structures (including the shorter runs). To further assess the reliability of the calculation and the existence of a single structure we repeated the annealing using a single copy. The calculations using a single copy are more difficult to converge and therefore the success rate of the annealing using a single copy was smaller (60%) for simulations exceeding 1 ns. Of the five runs, five end in a conformation that has the correct backbone and three in a conformation that has both a correct backbone and an optimal phenylalanine-leucine packing. To further establish that the conformation we found is indeed dominant in thermal equilibrium we pursue the following simulation. We started with an optimal configuration of the CHDLFC and propagated a *single copy* in the same box of water, at room temperature, for a period of 2 ns. In contrast to a similar numerical experiment for the peptide CAAAAC, the C_{α} rms of CHDLFC remains low and the

phenylalanine-leucine pair remained together (Figure 24b). This suggests that the conformation corresponds to a low free energy minimum. From now onward we shall call conformations similar to the above structure "correct" conformations.

To monitor the kinetic barriers in the exact system and the chances of discovering the correct conformation by chance, we started with random values for the dihedral angles and pursued a room temperature simulation of CHDLFC in a box of water. In a period of 2 ns the correct conformation was not found. In contrast, when a similar experiment (a room temperature trajectory in a box of water) was repeated using the LES approach (five copies) the correct conformation was found in a period of 250 ps. In Figure 25 we showed that the phenylalanine-leucine pair found each other in a remarkably short period: 100 ps. In fact the contact pair was formed before the backbone settled in the correct structure. This is another demonstration that the mean field approach is effective in reducing barriers separating free energy or energy minima. It can therefore be employed to study approximate dynamics when the time scales of the process are too long to study directly. To further monitor the correlation between the backbone conformation and the side chain structure, i.e., the phenylalanine and the leucine packing, we consider the following numerical experiment. We put a single copy of CHDLFC in the water box using an optimal backbone conformation. However, the phenylalanine and the leucine were apart. The whole system was simulated for the period of 2 ns. After 1 ns the two side chains adopted the packed configuration we previously discussed. The two side chains remained together (on the average) for the rest of the simulation (the second nanosecond). The existence of a hydrophobic core for the phenylalanine and leucine pair is therefore reasonably well established.

IV. Discussions

In this paper we outlined a procedure for determining the structure of peptides in water. We employed the LES/annealing protocol that made it possible for us to first check if a structure is likely to exist at all and then to determine it. Lattice searches (Eugene Shakhnovich, private communication) screened a large number of possibilities to a few with a strong tendency to a unique conformation. One of these peptides (CHDLFC) was refined to an atomic level model in explicit solvent using a mean field annealing protocol that we call LES. Annealing in explicit solvent is not trivial since the number of degrees of freedom and the number of "irrelevant" minima (corresponding to different solvent configurations) are tremendous. The combination of LES and the simulated annealing approach makes it possible to obtain statistically meaningful, reproducible results for hexapeptides in water. Of course this does not mean that the observed structures are also correct. The final judge of the structures we proposed should be an experiment. Nevertheless, the potential we employed here (the combination of OPLS and AMBER) is reasonably well established and we therefore expect reasonable results. Moreover, the availability of such an optimization tool makes it possible to examine if global energy minima of potentials for peptides in solvent indeed correspond to correct experimental structures. It is important to emphasize that with regular annealing the calculations would have been considerably more difficult. Considerably longer simulations are necessary and it is quite likely that we would give up on this calculation for which a single copy (regular annealing) is on the limit of feasibility.

In addition to the introduction of a new computational tool, this investigation suggests a number of amusing effects on the side chains and of the solvent on structures of polypeptides. We discuss them below.

In vacuum calculation the "almost homopolymer" (CAAAAAC) has a clear global energy minimum. The minimum is associated with optimal packing such that maximum hydrogen bonds are observed. However, once the solvent is introduced the strength

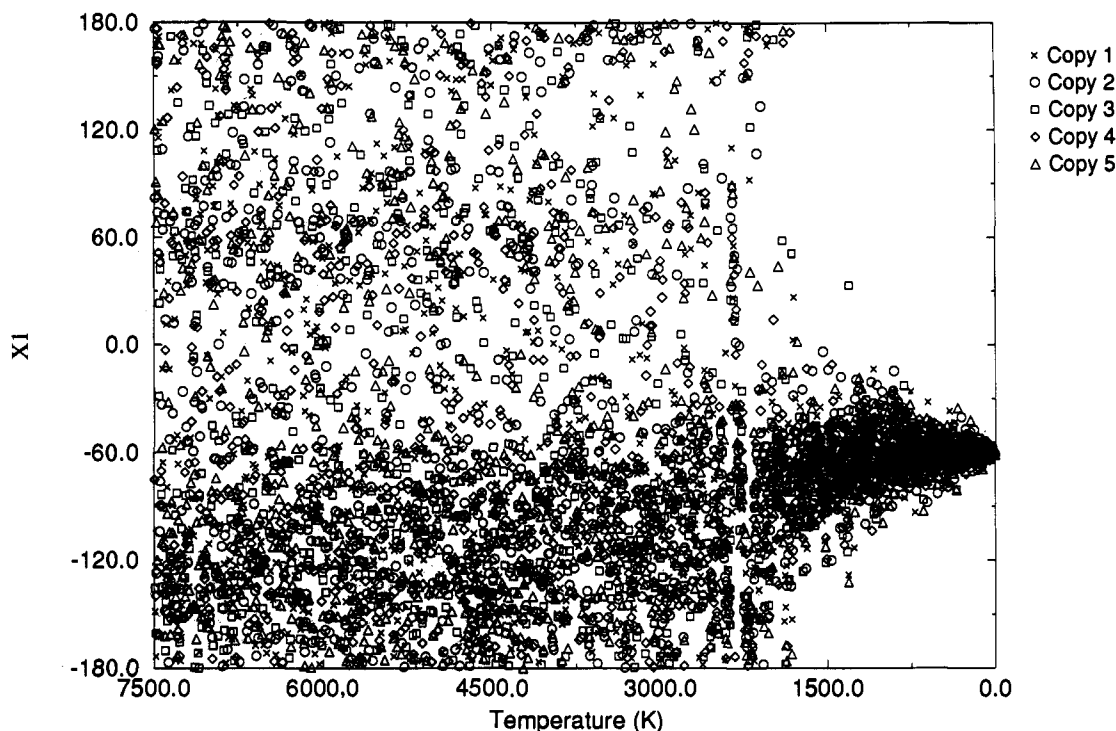


Figure 20. Annealing history of the χ_1 dihedral angles of the histidine copies in CHDLFC. The results are for a solution in which five copies were employed.

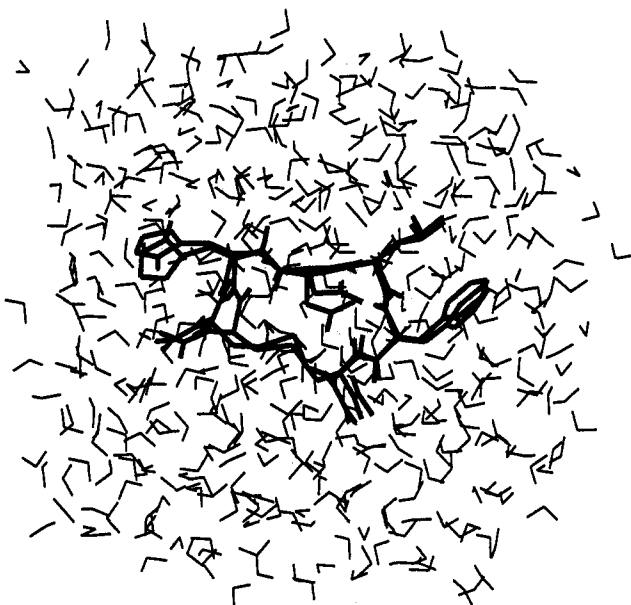


Figure 21. Overlap of two optimized structures of CHDLFC in solution. The CHDLFC structures of two LES runs are shown. The five copies from each individual LES run are on the top of each other. This is why only two structures are seen. For clarity only one set of water molecules is plotted.

of the hydrogen bonds is reduced considerably since alternative hydrogen bonding candidates, external to the peptide chain (water), are available. This results in a relatively easy exchange of conformations and lack of a unique structure. In many respects this observation is a warning to vacuum and vacuum related calculations. For example, it is not enough to add a hydrophobic energy term to an existing potential energy function, since this is only part of the story. A significant reduction in the effective strength of hydrogen bonds (as observed in vacuum) must accompany the changes in the potential energy once the peptide is moved from vacuum to liquid (water) environment. Clearly, the penalty of breaking a single hydrogen bond in solvent can become very low. As a result configurations with hydrogen

bonding far from optimal can still be stable. Such configurations are hard to imagine in vacuum in which the cost of breaking such bonds is very high. The precise strength of a hydrogen bond in the solvated system is hard to obtain from these simulations. A free energy calculation connecting different conformers may give some answers; however, it is not clear if the flips of the hydrogen bonds are cooperative and if it is possible at all to assign a single energy value per single hydrogen bond.

In the vacuum calculations of CHDLFC, very different dynamic and static behavior of the side chains and of the backbone was observed. While the backbone readily adopted a unique conformation, the side chains were wandering over considerable conformational space. Multiple conformations of side chains were detected, suggesting small energy differences between alternative minima (Table 2). Furthermore, the barriers separating the side chain minima are small, suggesting that the energy surface of the backbone is not only deeper but also more rough, in the sense that higher barriers separate the lower energy minima. We also comment that the backbone structures of CAAAAC and CHDLFC are quite different. The rms of the six C_α 's is 1.50 Å. If only the cycle (four) C_α 's are considered the rms is slightly higher—1.53 Å. Some features of the solvent simulation were similar to the vacuum results. The (free) energy surface of the side chains was still relatively flat. The side chain freezing temperatures were lower than the backbone freezing temperatures, and a number of alternative side chain conformations were detected in separate runs. A flat energy surface for side chain conformations is assumed in many lattice simulations in which the side chain is replaced by a mean field average (see for instance ref 33). However, in contrast to the vacuum results it was quite rare to detect multiple side chain conformations in a single LES run. This means that the water tends to force the copies on top of each other. The water pressure aims to reduce the volume occupied by the peptide, and at low temperatures it is usually successful even if the structure is not at the lowest energy configuration. This is also why we found it necessary to repeat the runs in the solvent. The "multiple copies" test of the LES annealing is insufficient in water. Of course, the few cases with multiple copies could be detected and easily eliminated as unacceptable runs. Furthermore, the other advantages of LES,

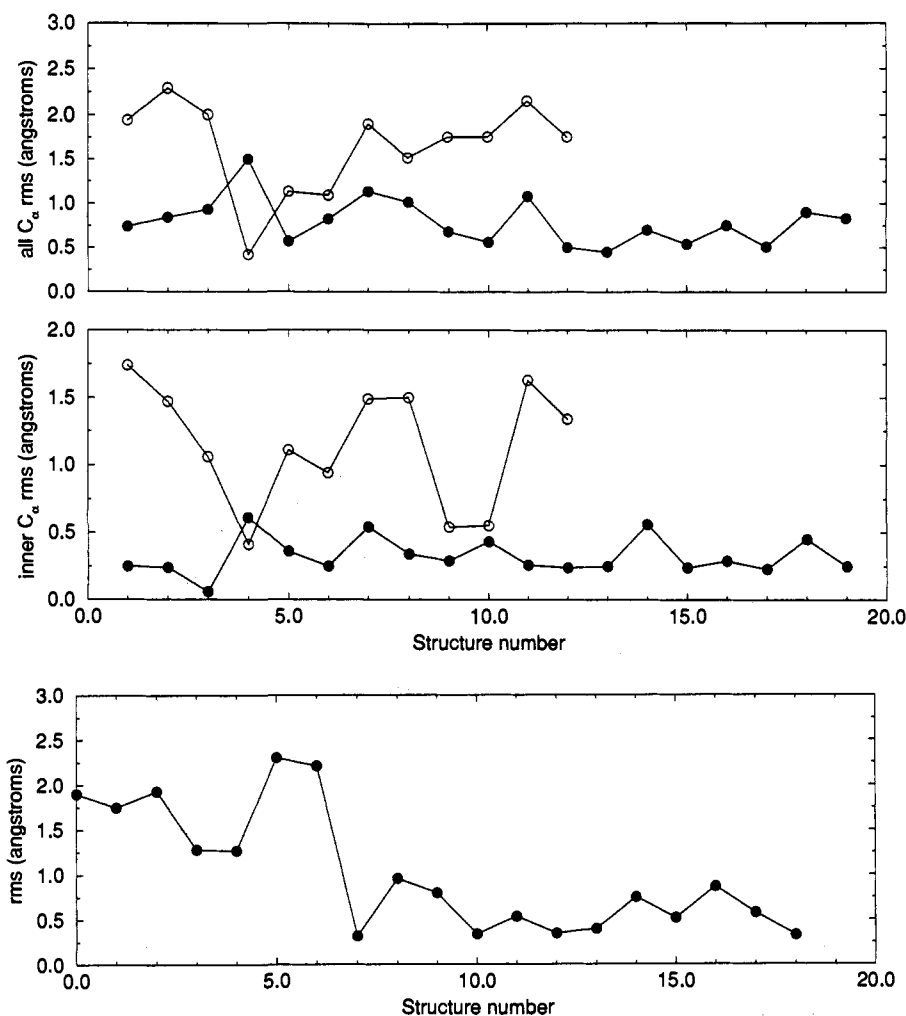


Figure 22. Root mean square deviation of the different structural segments of the optimized structures, compared to one of the "optimal" structures of CAAAAC or CHDLFC: CAAAAC (open symbol), CHDLFC (filled symbol). (a, top) All C_{α} ; (b, middle) C_{α} of HDLF (or AAAA) only; (c, bottom) C_{α} of HDLF and the leucine and phenylalanine side chains. These structures are sorted according to "successful" and "unsuccessful" runs. Note that large phenylalanine-leucine distances were found only for short annealing trajectories. See Table 1 for a summary.

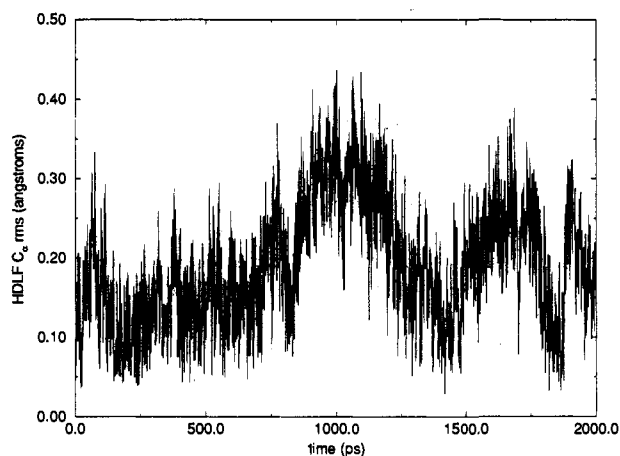


Figure 23. Root mean square deviation of the C_{α} for HDLF as a function of time. Note the low deviation of the structure in time as compared to the initial optimized structure. The low fluctuations are especially significant when compared to the corresponding calculation for CAAAAC (Figure 18).

such as potential smoothing, are still relevant and useful. The repetition of the LES/annealing runs demonstrates that the histidine and the aspartic acid side chains occupy more than a single conformation. Also the structure of the C- and the N-terminals is not well defined. Finally, multiple conformations were also observed for the sulfur linkage between the two cysteine

residues. Given the lack of a well-defined conformation for these structural segments, it may be surprising that we still find a well-defined core that includes the backbone of the peptide (HDLF) and the side chains of the hydrophobic pair (leucine and phenylalanine). The hydrophobic side chain pair reproduces itself in all the runs above 1 ns (13 times) in which multiple LES copies were not detected in the final structure. It further survives a number of additional tests at room temperature that were described in the results section. We therefore believe that within the model potential that we use it does exist. The "partial structure" observed for CHDLFC is suggestive as a nucleation site. We consider a "nucleation site" to be a part of the peptide chain that has a well-defined structure, almost independent of the other (reasonable) configurations of the rest of the chain. Such a well-defined structure can be associated with a spatially localized, free energy minimum. The minimum eliminates the need for further conformational search for that segment, and hence drastically reduces the conformational space that needs to be explored. This idea is (of course) not new. Suggestions for "nucleation sites" are of secondary structure elements²⁴ and of hydrophobic cores.²⁵ Even short β turns were observed experimentally,²⁶ and hydrophobic clusters were found in peptides as

(24) Ptitsyn, O. B.; Rashin, A. A. *Biophys. Chem.* **1975**, *3*, 1. For a very clear (and short) review, see: Baldwin, R. L. *Trends Biochem. Sci.* **1989**, *14*, 291.

(25) Dill, K. A. *Biochemistry* **1990**, *29*, 7133.

(26) Wright, P. E.; Dyson, H. J.; Lerner, R. A. *Biochemistry* **1988**, *27*, 7167.

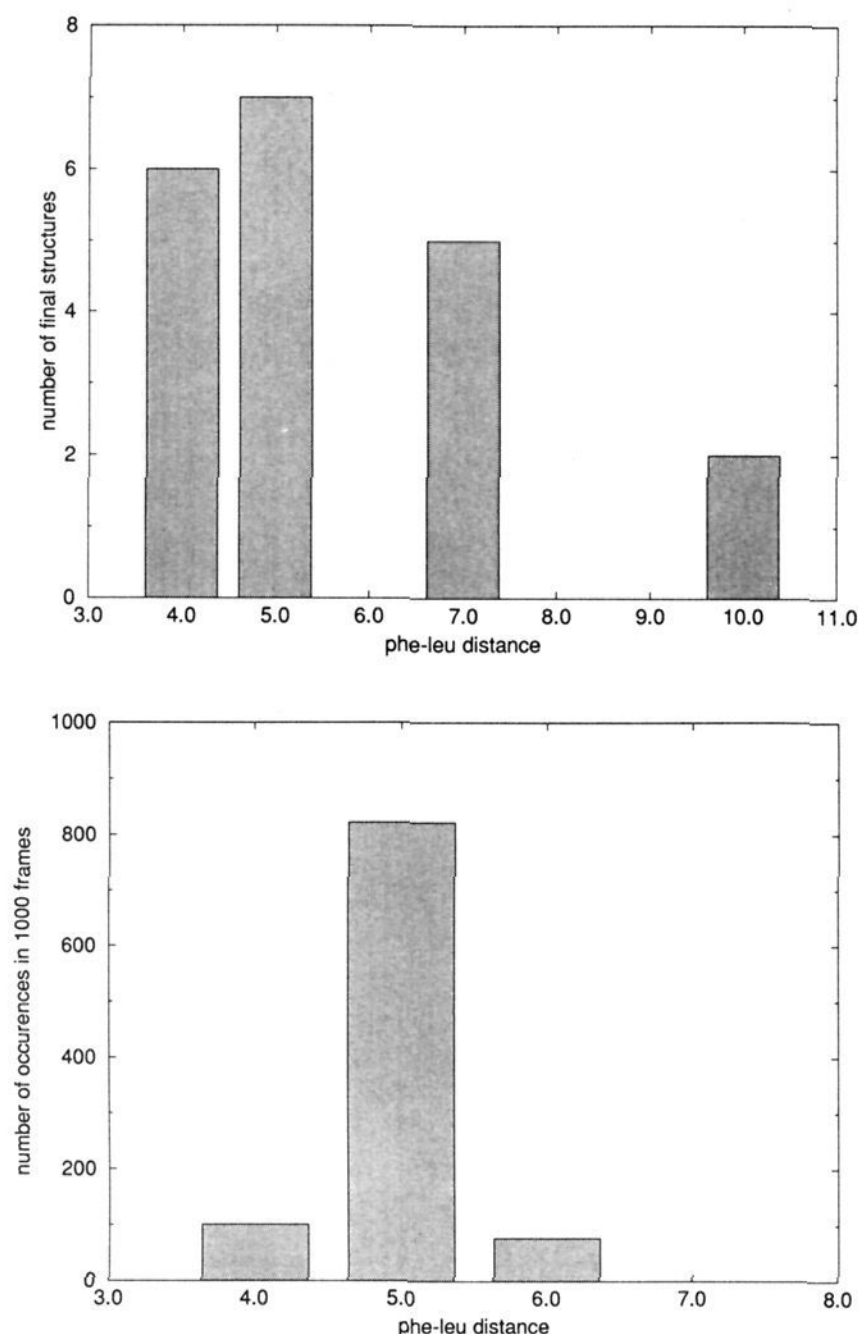


Figure 24. (a, top) A histogram plot of the distance between the phenylalanine and the histidine extracted from the set of optimized structures. Note that large phenylalanine-leucine distances are found only for the short annealing trajectories. The “successful” runs end at distances of 4 or 5 Å. See Table 1 for a summary. (b, bottom) The same as part a, except this time the results are extracted from a 2-ns trajectory.

small as four amino acids.²⁷ The present investigation is more in the spirit of the last study in which structure is observed for a relatively short peptide with no “long-range” support to the structure. Theoretically the existence of structure in such a small peptide is supporting evidence to the “minimal frustration” principle suggested by Bryngelson and Wolynes.²⁸ In the theory of the last authors, this principle is essential for solving the “Levinthal paradox”.

Given that CAAAAC does not show any structure, it is not likely that backbone hydrogen bonding is the drive to the unique conformation found for the core of CHDLFC. The differences between the peptides were in the side chains of HDLFC. The only side chains that have significant structure were LF. We therefore focused on analyzing the properties of the phenylalanine-leucine pair. We first demonstrated that on the time scale that we could afford (2 ns) the pair is stable at room temperature. We further showed that starting from correct backbone configuration but wrong orientation for the LF, the correct configuration is easily obtained. This is at room temperature using a single copy trajectory. We note, however, that starting from a **random configuration** the single copy did not find the correct structure in the few nanoseconds that we could afford computationally. This is likely to be a time scale problem. A LES run starting

(27) Bundi, A.; Andreatta, R. H.; Rittel, W.; Wuthrich, K. *FEBS Lett.* **1976**, *64*, 126.

(28) Bryngelson, J. D.; Wolynes, P. G. *J. Phys. Chem.* **1989**, *93*, 6902.

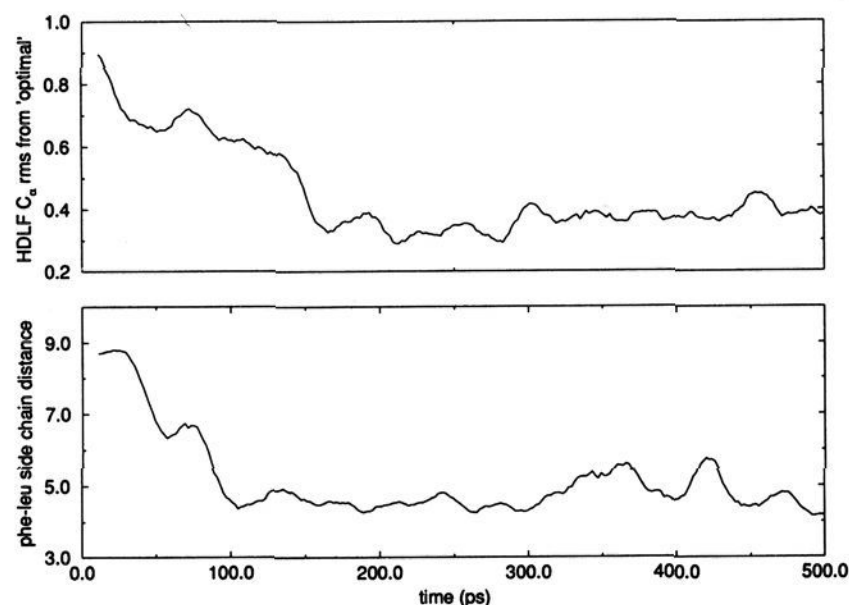


Figure 25. Room temperature LES simulation of the “folding” of CHDLFC: (a, top) The root mean square deviation between the C_α of the HDLFC core and an “optimal” structure as a function of time. Note that the LF pair finds each other before the core collapses to the right conformation. (b, bottom) The distance between the phenylalanine and leucine as a function of time.

from a random conformation with five copies at room temperature landed safely in an amazingly short time in the correct structure. This indicates the effectiveness of the potential energy smoothing using LES. LES is of course a mean field approximation and it is exact only at the limit of zero temperature. Nevertheless, LES is an atomic level model and it therefore provides a different view of the process compared to the widely used lattice and reduced models of proteins that were employed in the past.^{8,29,30} In Figure 25 we monitor a room temperature trajectory of the wrongly folded peptide as it approaches the correct conformation. We plot the C_α rms as well as the distance between the leucine and the phenylalanine as a function of time. An interesting observation is that the LF pair forms **before** the backbone. This suggests an additional separation of spatially localized free energy minima, to that of the LF pair and to that of the HDLFC backbone. This leads to even more saving in the required search of conformational space.

Coming back to the fact that the side chains are significantly more mobile, the picture emerging from the CHDLFC study is of hydrophobic pairing of side chains that move on a rapid time scale to the correct conformation. Since they are considerably more mobile compared to the backbone dihedrals, the side chains are better “equilibrated” on the time scale in which backbone transitions are observed. Reduced models for proteins in which the side chain is represented by an effective potential (potential of mean force) are in qualitative agreement with the results of the present study, which shows (a) separation of time scale and (b) weak dependence of the side chain positions on the backbone configuration.

Finally, a cautionary remark regarding the applicability of our results to folding of proteins. While the hydrogen bonds are insufficient to hold a structure in short peptides, for sufficiently large peptides hydrogen bonds form cooperatively. The general “rule” is that 20 amino acids are required to form a stable helix.³¹ Though exceptions to this “rule” are well-known (e.g., variants of the C-peptide³²), we are not aware of a stable helix with less than ten residues. Ten amino acids still provide a considerable

(29) Guo, Z.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 6369. Guo, Z.; Thirumalai, D.; Honeycutt, J. D. *J. Chem. Phys.* **1992**, *97*, 525.

(30) Chan, H. S.; Dill, K. A. *J. Chem. Phys.* **1993**, *99*, 2116.

(31) Zimm, B. H.; Bragg, J. K. *J. Chem. Phys.* **1959**, *31*, 1959.

(32) The discovery of the anomalous tendency of the C-peptide to a helix was made by Brown and Klee: Brown, J. E.; Klee, W. A. *Biochemistry* **1971**, *470*, 470. Since then, a series of investigations by Baldwin and co-workers was instrumental in understanding the physical forces leading to helices in peptide fragments. For a review see: Kim, P. S.; Baldwin, R. L. *Annu. Rev. Biochem.* **1990**, *59*, 631.

number of conformations, assuming ten states per amino acid. Thus, even the formation of secondary structure elements cannot be done in a blind search and a guiding force is necessary. We suggest that at the very early phase of the folding kinetics **localized hydrophobic pairing** may guide the chain through the possible conformations. This is regardless if at later times folding proceeds via hydrophobic collapse or secondary structure formation.

V. Conclusions

Probably the most significant result of the present investigation is the dominant role of hydrophobic forces (as opposed to hydrogen bonding) in determining structures of small peptides. CAAAAC did not have a well-defined structure in solvent while CHDLFC

did. Computationally we proposed a new methodology that can effectively determine the existence or non-existence of structures in small peptides and we issue a warning for **vacuum** determination of peptide structures. A hydrophobic force and solvent dependent hydrogen bonding—either explicit or implicit—is a necessity for successful prediction of structures in water.

Acknowledgment. This research was supported by NIH Grant GM41905 to R.E. R.E. is a University of Illinois West Scholar and Alon fellow in the Hebrew University. A University of Illinois research fellowship to C.S. is gratefully acknowledged. We thank Prof. E. Shakhnovich for helpful discussions. On the basis of lattice studies E. Shakhnovich proposed that CHDLFC has a strong tendency to a unique structure.

(33) Kolinski, A.; Skolnick, J. *J. Chem. Phys.* **1992**, *97*, 9412.