Conformational Memories and the Exploration of Biologically Relevant Peptide Conformations: An Illustration for the Gonadotropin-Releasing Hormone

Frank Guarnieri and Harel Weinstein*

Contribution from the Department of Physiology and Biophysics, Mount Sinai School of Medicine, New York, New York 10029-6574

Received August 11, 1995. Revised Manuscript Received April 26, 1996[⊗]

Abstract: The application of a recently developed technique of conformational memories (Guarnieri, F.; Wilson, S. R. *J. Comput. Chem.* **1995**, *16*, 5, 648) to the study of peptide structure is illustrated with an analysis of the conformational populations of the decapeptide gonadotropin-releasing hormone (GnRH) and the mutant peptide Lys8-GnRH. The conformational space of the peptides is explored fully with multiple Monte Carlo simulated annealing (MC/SA) random walks using the Amber* force field and the generalized Born/surface area (GB/SA) continuum solvation model of water as implemented in the Macromodel molecular modeling package. The collective histories of all the random walks are transformed into mean field dihedral distribution functions called "conformational memories" (35 conformational memories for GnRH, one for each torsion angle). Conformational families of the peptides at 310 K are obtained from these results with the use of a biased sampling technique which explores only the areas found to be populated in the conformational memories. A large family of GnRH structures that comprises approximately 70% of the population is identified by clustering conformations according to a pairwise root mean square deviation criterion. Members of this family of conformations share a distinctive β-type turn in the backbone that involves residues 5−8. This conformation is shown to correspond to the preferred geometry of a structurally constrained analog of GnRH that binds to the GnRH receptor with high affinity. In contrast, GnRH analogs such as Lys8-GnRH, for which we find that the major conformational family exhibits an extended backbone, seem to belong to the group of low-affinity ligands. The results suggest a correlation between high affinity at the GnRH receptor and the ability of the peptides to have a large population of conformations that exhibit the characteristic secondary structure of a β-type bend. Thus, the method of conformational memories is shown to provide a powerful and reliable tool for the computational exploration of structure−function relations of flexible peptides.

Introduction

Because of their central physiological role in a diverse array of biological systems, the peptide hormones are the focus of intense study. Their structural properties are of particular interest both for the search of mechanistic detail and for the development of structure−activity criteria that are essential for the design of related therapeutic agents. However, the difficulty in obtaining reliable information about the structure of peptides with a high degree of conformational flexibility cannot be overestimated.[1] Both experimental and computational methods are being applied to this task, with only mixed results.[2] We present an illustration of the significant advantages of a newly developed computational approach for the efficient exploration of the conformational space of such flexible peptides. The illustration of the method involves the study of the decapeptide gonadotropin-releasing hormone (pGlu1-His2-Trp3-Ser4-Tyr5-Gly6-Leu7-Arg8-Pro9-Gly10-NH_2, GnRH). The key physiological role of GnRH as a mediator of neuroendocrine regulation in the mammalian reproductive system has made it the object of intense study for several decades.[3−6] Because of the ability of GnRH and its analogs to modulate the pituitary−gonadal axis, an understanding of their mechanism of action and its relation to structure is essential in the development of therapeutic agents for the treatment of a variety of disorders ranging from infertility to prostatic carcinoma.[7,8] Conformational studies have played a central role in the quest for understanding the structural basis for the activities of GnRH and its analogs,[9,10] as well as in attempts to design new analogs[11,12] with improved pharmacological properties. Attempts to pare away some of the many thermally accessible but biologically irrelevant conformations have led to the synthesis of restricted GnRH analogs.[13−16] A complete exploration of the whole ensemble of conformational states of GnRH, some of which may be involved in the activation of the GnRH receptor, has not been accomplished.

[⊗] Abstract published in *Advance ACS Abstracts,* June 1, 1996.

(1) Rizo, J.; Gierasch, L. M. *Annu. Rev. Biochem.* **1992**, *61*, 387−418.
(2) Marshall, G. R. *Curr. Opin. Struct. Biol.* **1992**, *2*, 904−919.
(3) Chang, J.; Williams, R. H.; Humphries, A. J.; Johansson, N.; Folkers, K.; Bowers, C. Y. *Biochem. Biophys. Res. Commun.* **1972**, *47*, 727−732.
(4) Fujino, M.; Kobayashi, S.; Obayashi, M.; Fukuda, T.; Shinagawa, S.; Yamazaki, I.; Nakayama, R. *Biochem. Biophys. Res. Commun.* **1972**, *49*, 698−705.
(5) Monahan, M. W.; Amoss, M. S.; Anderson, H. A.; Vale, W. *Biochemistry* **1973**, *12*, 4616−4620.
(6) Coy, D. H.; Coy, E. J.; Hirotsu, Y.; Vilchez-Martinez, J. A.; Schally, A. V.; van Nispen, J. W.; Tesser, G. I. *Biochemistry* **1974**, *13*, 3550−3553.
(7) Barbieri, R. L. *Trends Endocrinol. Metab.* **1992**, *3*, 30−34.
(8) Casper, R. F. *Can. Med. Assoc. I.* **1991**, *144*, 153−160.
(9) Momany, F. A. *J. Am. Chem. Soc.* **1976**, *98*, 2990−2996.
(10) Gupta, H. M.; Talwar, G. P.; Salunke, D. M. *Proteins: Struct. Funct. Genet.* **1993**, *16*, 48−56.
(11) Nikiforovich, G. V.; Marshall, G. R. *Int. J. Peptide Protein Res.* **1993**, *42*, 181−193.
(12) Momany, F. A. *J. Am. Chem. Soc.* **1976**, *98*, 2996−3000.
(13) Rizo, J.; Koerber, S. C.; Bienstock, R. J.; Rivier, J.; Hagler, A. T.; Gierasch, L. M. *J. Am. Chem. Soc.* **1992**, *114*, 2852−2859.
(14) Rivier, J.; Kupryszewski, G.; Varga, J.; Porter, J.; Rivier, C.; Perrin, M.; Hagler, A.; Struthers, S.; Corrigan, A.; Vale, W. *J. Med. Chem.* **1988**, *31*, 677−682.
(15) Baniak, E. L.; Rivier, J. E.; Struthers, R. S.; Hagler, A. T.; Gierasch, L. M. *Biochemistry* **1987**, *26*, 2642−2656.
(16) Struthers, R. S.; Tanaka, G.; Koerber, S. C.; Solmajer, T.; Baniak, E. L.; Gierasch, L. M.; Vale, W.; Rivier, J. E.; Hagler, A. T. *Proteins: Struct. Funct. and Genet.* **1990**, *8*, 295−304.

S0002-7863(95)02745-4 CCC: $12.00

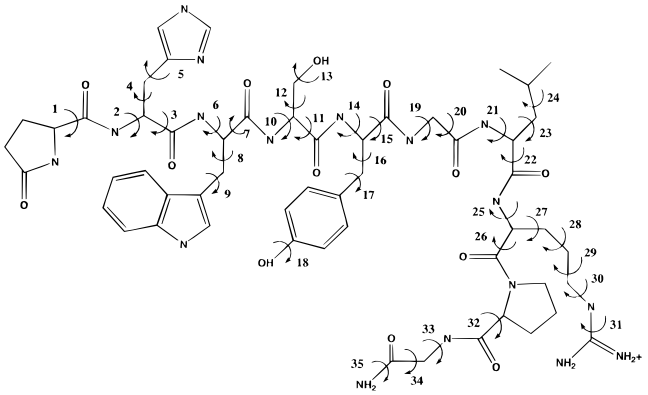

Figure 1. Molecular structure of the gonadotropin-releasing hormone (GnRH). The 35 rotatable torsional angles are indicated by arrows.

The investigations of the structural basis for the actions of the GnRH family of peptides is quite difficult because of its high flexibility.[17] A commonly used computational method for such investigations is molecular dynamics.[18] While dynamical techniques are capable of revealing short time scale molecular motions, these methods are generally incapable of exploring the ensemble of conformational states that exist in flexible molecules.[19,20] Spectroscopic techniques, for example, indicate that a multitude of interconverting conformers coexist for these systems.[17] The complexities inherent in the exploration of the structural properties of such small flexible peptides have motivated many experimental methods combined with computational techniques;[2] this approach has proven useful also for defining some structural motifs of GnRH analogs.[21,22] Nevertheless, it remains very difficult to obtain comprehensive details concerning the conformational properties of GnRH, or any of the large number of its bioactive analogs due to the large number of structural parameters that must be specified to characterize their bioactive forms (e.g., for GnRH, the number of rotatable bonds is 35 as shown in Figure 1). Consequently, a reliable computational method capable of performing a complete exploration of the conformational properties of such peptides continues to be of great interest.

We present here an application of the recently developed technique of conformational memories[23−26] to the study of GnRH and Lys8-GnRH, illustrating its ability to yield converged dihedral populations of all 35 rotatable bonds of the peptide GnRH, with no approximations. Samples from the conformational memories using a biased sampling technique were used to characterize the conformational families of GnRH and several of its analogs, in an aqueous environment modeled with the generalized Born/surface area (GB/SA) method.[27] The comparative analysis of the conformational properties of GnRH analogs, using the program Xcluster,[28] reveals their conformational preferences and suggests some of the key structural determinants for their known biological function on the gonadotropin-releasing hormone receptor.

Table 1. A Sample of the Output Collected in the History Files of the Simulated Annealing Random Walks[a]

	rotated dihed atom 1−atom 2	extent of rotation	energy	dihed value conformation
0	40−41	63.86	−276.38	176.53
0	47−48	−28.67	−276.38	−77.32
1	90−91	3.48	−252.98	168.15
1	3−4	11.96	−252.98	−77.33

[a] Column 1 indicates whether the data are produced from an accepted or rejected step with 0 = rejected and 1 = accepted. The second column lists the pair of atom numbers identifying the dihedral angles that were rotated to produce the trial structure. The third column lists the extent to which the dihedral was rotated in order to create the trial structure. The fourth column lists the energy of the current conformation (the energy of the original structure if rejected or the new structure if accepted). The fifth column lists the current dihedral values of the conformation (the dihedral angle of the original structure if rejected or the new structure if accepted).

Methods

Conformational Memories. The molecular simulation technique of conformational memories[24] is a two-stage process consisting of an exploratory phase and a biased sampling phase. In the exploratory phase, repeated runs of Monte Carlo[29] simulated annealing (MC/SA)[30] are carried out in order to map the entire conformational space of a flexible molecule. The construction of conformational memories described below has been interfaced with the Macromodel[31] molecular modeling package, version 5.0, so that the continuum GB/SA solvent model,[27] and the recently developed amino acid backbone torsional potentials[32] from the Macromodel package, could be used in the present conformational study.

As applied here, the MC/SA protocol for the exploratory phase was designed with a starting temperature of 2070 K and a cooling schedule of $T_{n+1} = 0.9T_n$ for 19 discrete temperature points. At each temperature 10 000 steps were applied to the 35 rotatable bonds (Figure 1), cooling the system to a final temperature of 310 K. Trial conformations in the MC/SA routine were generated by randomly picking 2 rotatable bonds from among the 35, rotating each bond by a random value between ±180°, and accepting or rejecting the trial conformation according to the standard Metropolis criteria with a Boltzmann probability function defined at the given temperature. After each step, data are recorded to a "log file" regardless of whether the conformation was accepted or rejected. The data recorded for the rotated bonds include the extent of rotation, the energy, and the value of the dihedral angles. An example of the output to the log file is given in Table 1. In this example, the first group of entries, corresponding to the first two lines, is the result of a rejected step as indicated by the zeros in the first column. The second column identifies the atom numbers of the two bonds that were rotated to create the trial move (in this example atom 40−atom 41 and atom 47−atom 48). The third column lists the extent of rotation of the torsion angle in degrees. The fourth column lists the total energy of the structure, and the fifth column holds the current dihedral value of the bond.

The second group of entries in Table 1, corresponding to the next two lines, lists the results of a trial rotation that was accepted as a new

(17) Chary, K. V. R.; Srivastava, S.; Hosur, R. V.; Roy, K. B.; Govil, G. *Eur. J. Biochem.* **1986**, *158*, 323−332.
(18) Tobias, D. J.; Mertz, J. E.; Brooks, C. L. *Biochemistry* **1991**, *30*, 6054−6058.
(19) Guarnieri, F.; Still, W. C. *J. Comput. Chem.* **1994**, *15*, 1302−1310.
(20) Guarnieri, F. *J. Math. Chem.* **1995**, *18*, 25−35.
(21) Rizo, J.; Koerber, S. C.; Bienstock, R. J.; Rivier, J.; Gierasch, L. M.; Hagler, A. T. *J. Am. Chem. Soc.* **1992**, *114*, 2860−2871.
(22) Bienstock, R. J.; Rizo, J.; Koerber, S. C.; Rivier, J.; Hagler, A. T.; Gierasch, L. M. *J. Med. Chem.* **1993**, *36*, 3265−3273.
(23) Wilson, S. R.; Guarnieri, F. *Tetrahedron Lett.* **1991**, *32*, 3601−3604.
(24) Guarnieri, F.; Wilson, S. R. *J. Comput. Chem.* **1995**, *16*, 648−653.
(25) Wilson, S. R.; Cui, W.; Guarnieri, F. In *Adaptation of Simulated Annealing to Chemical Problems*; Calivas, J., Ed.; Elsevier: Amsterdam, 1995; pp 351−367.
(26) Wilson, S. R.; Guarnieri, F. In *Proceedings of the 8th Vitamin D Workshop*; Norman, A., Ed.; Walter de Gruyter Co.: Berlin, 1991; pp 186−187.
(27) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.
(28) Shenkin, P. S.; McDonald, D. Q. *J. Comput. Chem.* **1994**, *15*, 899−916.
(29) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087−1092.
(30) Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P. *Science* **1983**, *220*, 671.
(31) Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caulfield, C.; Chang, G.; Hendrickson, T.; Still, W. C. *J. Comput. Chem.* **1990**, *11*, 440.
(32) McDonald, D. Q.; Still, W. C. *Tetrahedron Lett.* **1992**, *33*, 7743−7746.

Table 2. A Sample of a Conformational Memory Spreadsheet[a]

temp (K)	−180°	−170°	−160°	−150°	−140°	−130°	−120°	−110°	−100°
1500	4.0	3.7	3.2	2.5	2.1	1.5	1.3	1.3	1.1
1350	4.7	3.7	3.7	2.4	2.3	1.5	1.4	1.1	0.9
1215	4.0	4.0	3.7	2.4	2.0	1.1	1.0	0.9	0.9
1094	4.5	3.6	3.6	2.4	1.9	1.3	0.9	0.9	1.0
984	5.4	4.6	3.9	2.2	1.7	1.1	0.8	0.9	1.0
886	5.3	4.4	3.4	2.6	1.3	1.2	0.9	0.8	0.5
797	5.8	4.9	3.3	1.9	1.2	0.9	0.6	0.4	0.5
717	6.0	4.2	3.5	2.4	1.4	0.9	0.5	0.2	0.4
646	5.6	4.6	3.5	2.2	1.0	0.7	0.3	0.3	0.4
581	6.3	5.1	2.9	1.9	1.4	0.4	0.4	0.2	0.2
523	6.1	3.9	3.8	1.8	1.2	0.5	0.1	0.2	0.1
471	7.5	5.0	4.2	1.3	0.9	0.4	0.2	0.1	0.1
424	7.9	4.2	2.7	1.7	0.8	0.2	0.1	0.2	0.0
381	8.4	5.4	2.1	1.3	0.3	0.3	0.1	0.1	0.0
343	0.1	5.6	2.9	1.3	0.2	0.1	0.1	0.0	0.0
310	8.2	3.7	1.6	1.6	0.2	0.0	0.0	0.0	0.0

[a] The first row labels the dihedral circle across the *y* axis. The first column labels the temperatures across the *x* axis. Each cell contains the population corresponding to a given temperature and a given 10° dihedral bucket which is plotted on the *z* axis.

conformation (as indicated by the digit one in the first column). The current dihedral values given in the last column are the new values of the newly accepted conformation.

Each run of MC/SA consists of a random walk of 190 000 steps (19 temperatures, 10 000 steps per temperature). Because two lines of data are added to the log file for each Monte Carlo step, a single run creates a file of 380 000 lines. To explore the conformations of the GnRH peptide in GB/SA water, we performed 157 of these simulations, creating log files of the different random walks. A 157 run MC/SA simulation requires about 12 days of computation on an SGI Challenge 200 MHz workstation.

To obtain structural information from this large amount of data, the log files are used as input to a program called Flex that sorts, merges, and compacts the data in several ways. Since the simulations were done at 19 temperatures for each peptide, application of Flex first sorts and merges the data from all log files into 19 temperature blocks. Subsequently, within each temperature block, the data are partitioned into 35 bond blocks, one for each rotatable bond. For each rotatable bond, the dihedral angle space is partitioned into thirty six 10° intervals. From each line of data for a given bond at a given temperature, the program records the number of times that the bond dihedral angle value belongs to one of the ten degree buckets, i.e., a "conformational memory". Finally, the Flex program produces a 19 × 36 spreadsheet (recording nineteen temperatures by thirty-six 10° dihedral intervals with normalized populations) for each of the thirty five rotatable bonds of the peptide. An excerpt of one of these spreadsheets is given in Table 2. The spreadsheets are imported into Deltagraph[33] for plotting. The graphical representation of the data is illustrated in Figure 2. Across the top of the spreadsheet are the dihedral angle values from −170° to +180° which label the *y* axes of Figure 2. (note that the spreadsheet fragment illustrated in Table 2 is cut off at −100°). The first column of Table 2 records the 19 temperatures which label the *x* axes in Figures 2, ranging from 2070 to 310 K. The value in a spreadsheet position corresponding to a given temperature within a given 10° dihedral bucket is the population percentage which is plotted in the *z* direction in Figure 2.

The procedure described here for creating conformational memories for the dihedral angles results in an enormous compression of the large volume of data needed to describe a 35-dimensional hypertorsional space. The condensation of the information into Deltagraph plots yields identifiable structural motifs. For example, bond 4 shown in Figure 2a exhibits a classic three-state distribution: trans, gauche+ and gauche−, whereas bond 6, the ϕ angle of residue 3 (Figure 2b), has a continuous population distribution over a very large range from about −60° to −180° and no population in the other regions at any temperature. In contrast, bond 7, the ψ angle of residue 3 (Figure 2c), has a narrow all-trans distribution. The distribution of bond 35 (Figure 2d) favors a trans conformation, but maintains significant population over the entire dihedral angle range at all temperatures.

(33) Deltagraph. Deltagraph TM Version 1.0, Copyright Deltapoint, Inc., 200 Heritage Harbor, Suite G, Monterey, CA 93940, 1987.

Convergence of Conformational Memories. By including the results from multiple explorations of all possible combinations of dihedral angle values for all rotatable bonds of the molecule, the 35 conformational memories determined from the procedure described above provide a complete mapping of the entire conformational space of GnRH with no approximations, as long as the calculated populations are converged. Population convergence ensures that regions that could be thermally accessible are not erroneously labeled as being unpopulated. In the original formulation of the method,[24] population convergence was identified as the difficult and crucial aspect of forming conformational memories. The correct identification of the populated regions is essential for the second phase of the simulation, because the biased sampling only explores population regions of the conformational space.

Population convergence for the GnRH was confirmed in three different ways: (i) by creating conformational memory difference maps for simulations of different length, (ii) by analyzing intrinsic symmetry, and (iii) by showing that there is no significant difference in the populations of actual structures of GnRH created from conformational memories obtained from 25, 50, 75, 100, and 157 independent MC/SA runs. Figure 3 shows the conformational memory difference maps for dihedral angle 20, comparing simulation lengths of 10, 25, 50, 75, and 100 runs. The difference map in Figure 3a is created by subtracting the conformational memory obtained from a 25 run MC/SA simulation from a 10 run MC/SA simulation, in Figure 3b the difference is between 50 and 25 runs, Figure 3c shows the difference between 75 and 50 runs, and Figure 3d is the difference map between 100 and 75 runs. The progression clearly shows the convergence. The other dihedral angles have very similar difference maps for this sequence of comparisons.

A second measure of convergence is symmetry. Because dihedral angle 17 has a 2-fold axis of symmetry, it is expected that the dihedral space of this bond will have symmetric population distributions centered at −90° and +90°. A temperature slice at 310 K of this dihedral for 25, 50, 75, and 100 run MC/SA simulations is shown in Figure 4. The population distributions clearly conform to the symmetry considerations.

The third indication of convergence is the finding (see below) that biased sampling from conformational memories created from 25, 50, 75, 100, and 157 MC/SA runs yields very similar structural profiles of GnRH.

Biased Sampling from Conformational Memories: Elimination of Barriers. Once the conformational memories are established, a new Monte Carlo search is performed at 310 K, sampling only from the populated regions. Only about 50% of the torsional space of the 35 bonds is populated at 310 K, so that the conformational space that needs to be explored in the biased sampling phase of the simulation has been reduced without approximations, by many orders of magnitude. Table 3 is an excerpt of the probability matrix for GnRH at 310 K. The dimension of this probability matrix is 35 × 36 for the 35 rotatable bonds partitioned into 36 buckets over the 360° dihedral space (note that only 11 of the 36 dihedral buckets and only 14 of the 35 rotatable torsional angles are shown in Table 3). The first line indicates that at 310 K bond 1 is found in the −180° to −170° dihedral interval 10.1% of the time. In contrast, the seventh data column of the first row indicates that bond 1 is never found in the dihedral interval from −120° to −110° at 310 K.

The two-stage process of developing conformational memories and then performing the biased sampling from these distributions is necessary in order to sample the entire conformational space of the molecule. An obviously simpler alternative would be to limit the conformational exploration to standard Metropolis Monte Carlo at 310 K and monitor the development of the random walk over torsional space. However, this simulation constitutes the last step in the development of the conformational memories for the temperature of 310 K; it is clearly inadequate, as indicated by the acceptance rate. The acceptance rate is about 28% at 2070 K, with a step size chosen randomly within the interval of ±180° and rotating two dihedrals selected randomly at each step. At 310 K, using the same parameters,

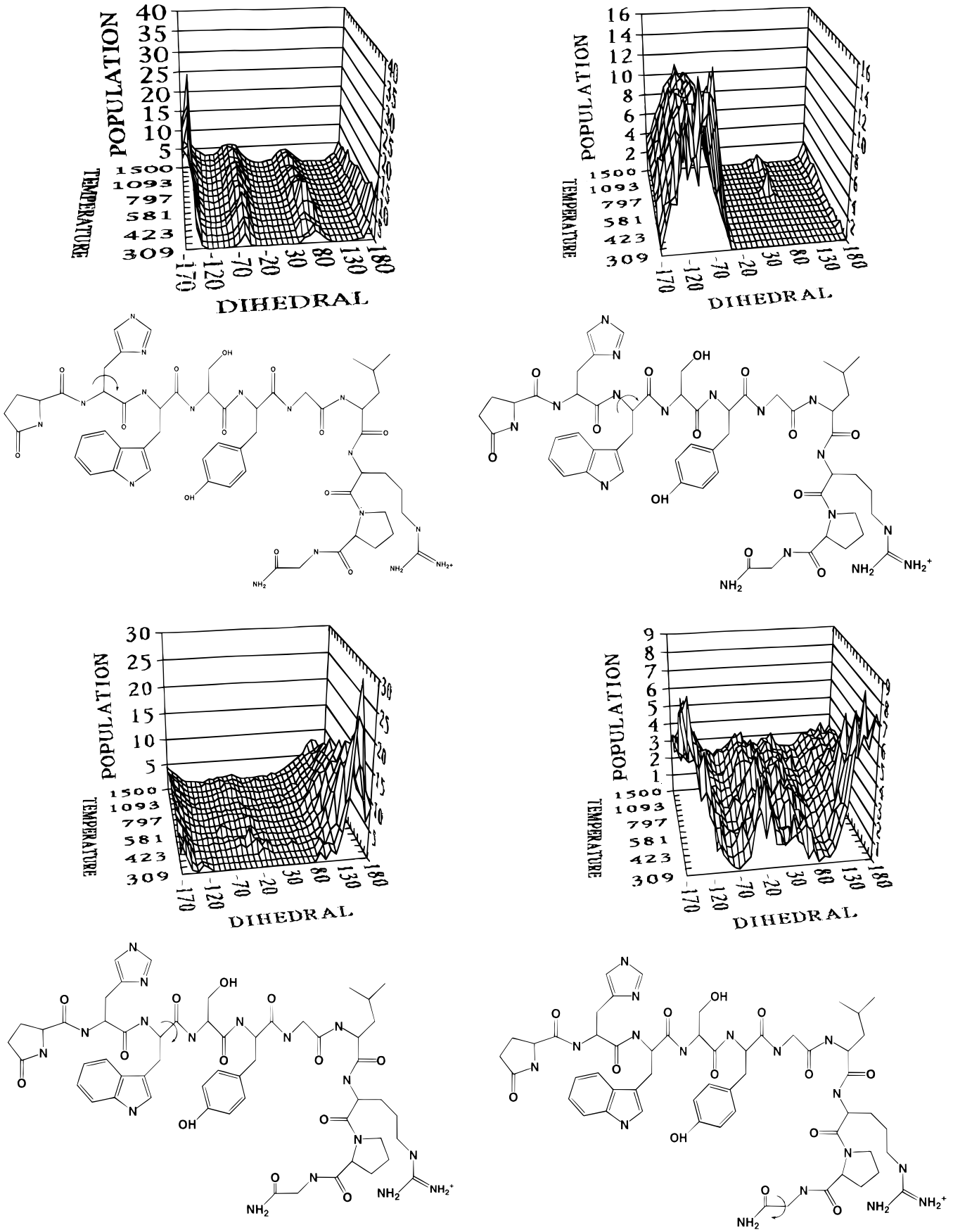

Figure 2. Conformational memories of selected dihedral angles in the gonadotropin-releasing hormone (see Figure 1 for identity of the angles: (a, top left) dihedral angle 4; (b, top right) dihedral angle 6; (c, bottom left) dihedral angle 7; (d, bottom right) dihedral angle 35.

the acceptance rate falls below 2%. Therefore, the sampling of the 35-dimensional dihedral space would be incomplete if these parameters were used for Monte Carlo random walk procedure at 310 K. Even if the random interval from which trial configurations are sampled were reduced to ±30° (to increase the acceptance rate), sampling would still be insufficient because the majority of new conformations would be

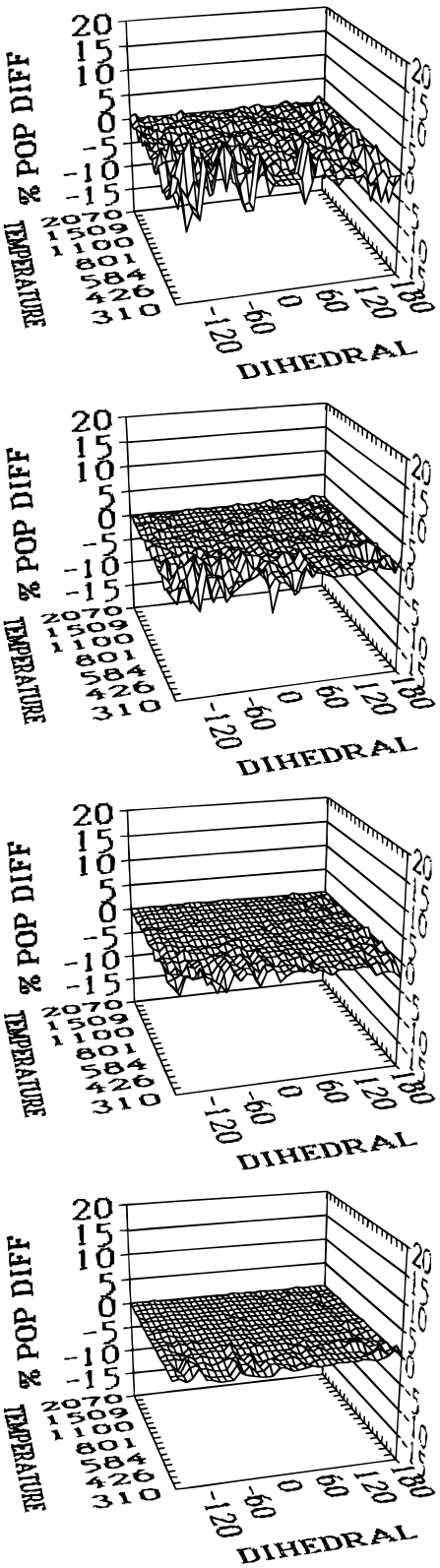

Figure 3. Conformational memory difference maps of dihedral angle 20 in GnRH (see Figure 1). The difference maps were created by subtracting the conformational memories from (from top to bottom) (a) 25 and 10 runs, (b) 50 and 25 runs, (c) 75 and 50 runs, and (d) 100 and 75 runs. Note that in almost all regions differences are <1%. Conformational memory difference maps of the other dihedrals are very similar.

in the local area of the previous conformation. The ±180° step size was deliberately chosen so that new conformations can be created by jumping between wells without having to climb over barriers. A single simulated annealing run cannot be expected to cover such a vast space,

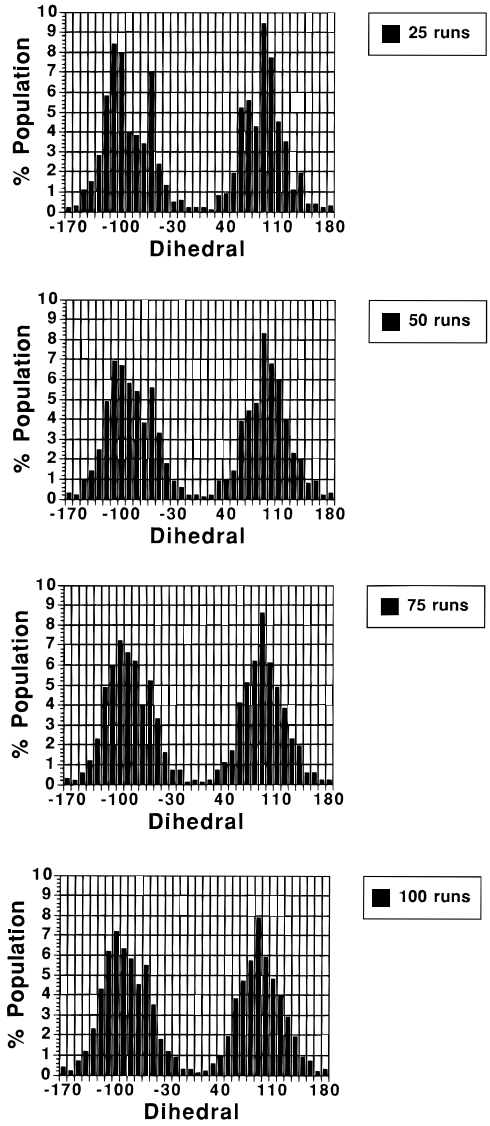

Figure 4. Sequences of 310 K temperature slices from the conformational memory of bond 17, calculated with (from top to bottom) (a) 25 runs, (b) 50 runs, (c) 75 runs, and (d) 100 runs. Note the symmetrically equivalent population distributions centered about −90° and +90°.

but cumulations of multiple runs while each of the runs performs a different random walk can be shown to converge, as illustrated in Figures 3 and Figure 4.

The restriction of the sampling to the populated regions identified in the previous step (i.e., the conformational memories) is achieved by partitioning the 0−1 interval of the random number generator into the 36 parts which correspond to the 36 separate 10° intervals for each rotatable dihedral angle. The partitioning of the random number generator is proportional to the population of the 10° bucket. New biased trial conformations are generated by randomly choosing two rotatable bonds, generating a new random number for each bond, determining to which of the 36 intervals each new random number for each bond belongs, and driving the dihedrals to the appropriate intervals. The exact value of the new dihedral is determined by a linear interpolation. This procedure is illustrated in Figure 5.

A major advantage of the conformational memory biased sampling method is that partitioning the random number generator among the populated intervals results in a sampling technique that eliminates the barrier-crossing problem. During the biased sampling random walk, a new trial configuration is sampled from the conformational memory, which can be any part of the *populated* dihedral space, and then the trial conformation is created by driving the current structure to the appropriate configuration. Hence, the notion of the barrier restricting access to any part of the conformational space is eliminated in this

Table 3. Excerpt from the Population Probability Matrix for GnRH at 310 K[a]

bond	−170°	−160°	−150°	−140°	−130	−120°	−110°	−100°	−90°	−80°	−70°
1	10.1	9.1	5.6	1.7	0.5	0.3	0.0	0.0	0.1	0.0	0.1
2	1.4	3.8	6.6	9.8	11.8	10.0	12.3	10.4	11.4	9.0	8.1
3	4.2	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0
4	29.4	31.9	5.2	0.8	0.0	0.0	0.0	0.0	0.1	0.2	1.1
5	0.3	1.1	1.0	1.9	3.5	3.9	5.1	7.9	19.0	19.3	9.4
6	0.5	2.7	4.7	4.6	11.7	6.8	8.7	6.2	14.1	14.1	13.3
7	5.7	3.0	0.5	0.2	1.3	0.7	0.0	0.0	0.0	0.0	0.0
8	21.2	14.1	7.4	3.6	0.0	0.0	0.0	0.0	0.2	0.4	1.7
9	0.0	0.0	0.1	0.6	0.6	0.1	0.3	0.3	5.9	2.0	3.3
10	0.0	2.3	10.9	19.5	16.0	12.9	5.6	2.3	6.1	8.9	5.8
11	1.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	1.3
12	0.4	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.4	1.4	10.1
13	2.3	1.6	1.5	0.6	0.7	0.7	0.6	0.6	1.1	1.6	1.4
14	0.3	0.5	0.9	4.5	3.3	9.5	10.8	8.2	10.1	10.3	18.7

[a] The dimensions of this matrix are 35 × 36 (35 rotatable dihedral angles, with the population distribution of each angle broken into 36 intervals of 10°). Note that only 14 rows and 11 data columns of this matrix are shown.

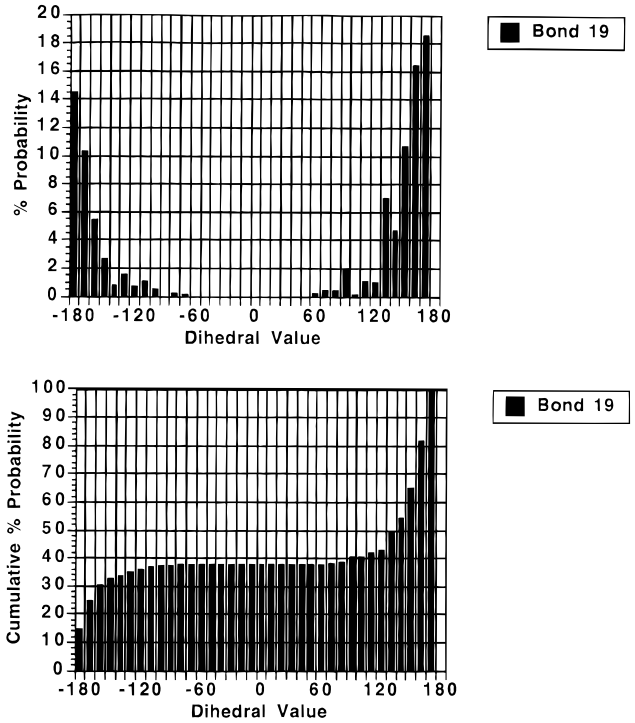

Figure 5. Choice of dihedral angle values in the biased sampling of the populated region of the conformational memories. The illustration is for dihedral angle 19. Panel a (top) shows a histogram representation of the probability distribution for the dihedral angle; panel b (bottom) shows the cumulative probability distribution for dihedral angle 19. Since the random number generator is a cumulative probability distribution, biased sampling is done from the histogram in part b. If the random number 0.2 is generated, which corresponds to the second block of the histogram in part b, the new trial dihedral will be chosen from the interval from −170° to −160°, with the actual value obtained from a linear interpolation within this interval. If the random number 0.4 is generated, which corresponds to the 28th block of the histogram in part b, the new trial dihedral will be chosen from the interval from 90° to 100°. Note that the region between −60° and +60°, which has no population in part a, is automatically skipped when the sampling is done from part b.

procedure. Because conformational memories are mean field population distributions, the correlations among the different flexible torsional angles have been submerged in the averaging process. Nevertheless, the conformational memory biased sampling technique does preferentially bring together the higher probability regions of the different dihedrals. Thus, the method introduces average correlations among the different dihedral angles during the selection process, while accessing all populated regions. It is important to note that the original formulation of the conformational memories biased sampling technique[24] violates detailed balance. Here, we have corrected the biased sampling so that it obeys detailed balance by multiplying the Boltzmann function used in the Metropolis test with the factor $P_{1old}P_{2old}/P_{1new}P_{2new}$, where P_{1old} and P_{2old} are the population percentages of the 10° intervals of the conformational memories of the two dihedral angles in the current conformation of the random walk (because in this example two dihedrals per step are changed). P_{1new} and P_{2new} are the corresponding population percentages of the new dihedral values for these angles in the new trial conformation.

Development of Conformational Families. We performed several sequences of biased sampling runs at 310 K to determine the best and simplest way to create representative conformational families for the GnRH peptide. The first run was a 10 000 step MC random walk using the conformational memory biased sampling technique with uniform sampling of 100 structures (1 sample every 100 steps). The second run was a 50 000 step MC random walk using the conformational memory biased sampling technique with uniform sampling of 100 structures (1 sample every 500 steps). The third and fourth runs were 100 000 and 500 000 step biased sampling runs also sampling 100 structures in the same manner. Each batch of 100 structures was analyzed with the program XCluster.[28] Xcluster inputs the series of 100 conformations and computes the RMS difference between all possible pairs of conformations. Structures 2−100 of the input sequence are then reordered on the basis of increasing RMS deviation. In the new ordering, considering all 100 conformations, conformer 2 has the smallest RMS deviation from conformer 1, and conformer 3 has the smallest RMS deviation from conformer 2, etc. Xcluster then produces a graphical representation of the RMS deviations between every pair of conformers. Since the conformations have been rearranged so that the RMS deviation between nearest neighbors is minimized, any large jump in RMS deviation between nearest neighbors is indicative of a large structural change and hence identifies a new conformational family. As described below, we settled on 500 000 steps for the subsequent biased sampling runs. We then performed these biased sampling runs using conformational memories created from 25, 50, 75, 100, and 157 run MC/SA simulations.

Results and Discussion

Conformational Families of GnRH. The 500 000 step biased sampling runs for GnRH with a sampling rate of 1 every 5000 structures require 4.3 h per run on a 200 MHz SGI Challenge workstation. Structures from the 500 000 step biased sampling run were clustered in conformational families as described in the Methods. A backbone trace of representatives from the five families with very distinct backbone conformations that emerged from this procedure is shown in Figure 6. Notably, similar results were obtained regardless of the origin of the conformational memories from MC/SA simulations of 25, 50, 75, 100, or 157 runs. Families of conformations having a β-turn among residues 5−8 occur with a frequency of approximately 70%. A distribution showing a superimposition of 70 of these

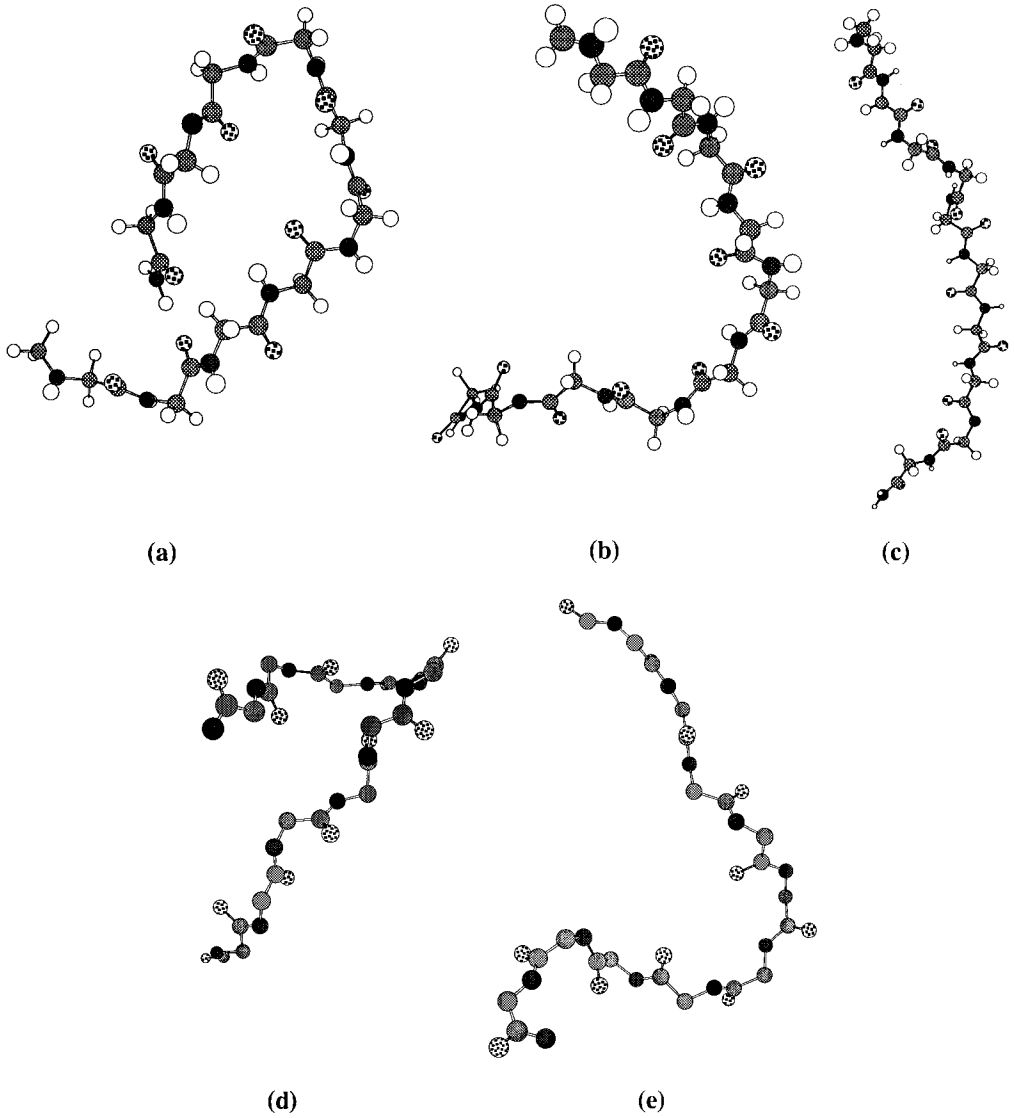

Figure 6. Backbone trace of a representative of five conformational families of GnRH obtained from conformational memories. Structures with a β-type turn have a 70% population. Structures with a straight backbone have approximately 5% population.

structures is illustrated in Figure 7 (GnRH is shown in red, with Arg8 shown in green). The β-type turn common to all the structures in this family is clearly evident (Figure 7). In contrast, families which have an extended backbone occur with a frequency of about 5%. The distribution of side chain orientations of Arg8 in all conformational families is wider than that of any other residue in GnRH. The results of Struthers et al.[16] from the examination of different GnRH analogs seem to indicate that an arginine is required as part of the pharmacophore. The present results, on the other hand, may indicate that the role of Arg8 in the receptor interaction of GnRH could relate to the backbone conformation, rather than to its participation in a recognition pharmacophore.

It is noteworthy that biased sampling runs of 10 000−25 000 steps resulted in large (unconverged) fluctuations in the ratio of β-turn to extended backbone conformations. However, more extended biased sampling runs of 100 000−500 000 resulted in negligible fluctuations in the ratio of β-turn to extended backbone conformations. Although it appeared from our calibration studies that 100 000 step biased sampling runs are sufficient, we chose to carry out the more extensive 500 000 step biased sampling runs for all the calculations presented here.

Conformational Families of Lys8-GnRH. The Lys8 analog of GnRH had been constructed to explore the role of Arg8 in molecular recognition of GnRH by its receptor.[34,35] Mutation studies of GnRH receptors from various species have implicated Arg[8] as being important for mammalian hormone−receptor recognition.[36] To analyze the structural implications of Arg8 for the activity of GnRH, we compared the conformational profile of the peptide hormone with that of the mutant Lys8-GnRH which is known to be a low-affinity GnRH agonist. In contrast to the wild type hormone, the major conformational family of the Lys8-GnRH congener was found to have an extended backbone, while the β-turn conformation exists as a very minor family. A backbone trace of a representative of each family is shown in Figure 8. The family of conformations represented in Figure 8a has an extended backbone and occurs with a frequency of greater than 70%. The Lys8-GnRH family that has a β-type turn conformation of the backbone (Figure 8b), which is virtually identical to the major conformational family of the GnRH (Figure 6a), has a probability of only about 3%. A distribution of the members of the predominant Lys8-GnRH family superimposed upon each other is shown in Figure

(34) Karten, M. J.; Rivier, J. E. *Endocr. Rev.* **1986**, *7*, 44−66.
(35) Millar, R. P.; Flanagan, C. A.; deL. Milton, R. C.; King, J. A. *J. Biol. Chem.* **1989**, *264*, 21007−21013.
(36) Flanagan, C. A.; Becker, I. I.; Davidson, J. S.; Wakefield, I. K.; Zhou, W.; Sealfon, S. C.; Millar, R. P. *J. Biol. Chem.* **1994**, *269*, 22636−22641.

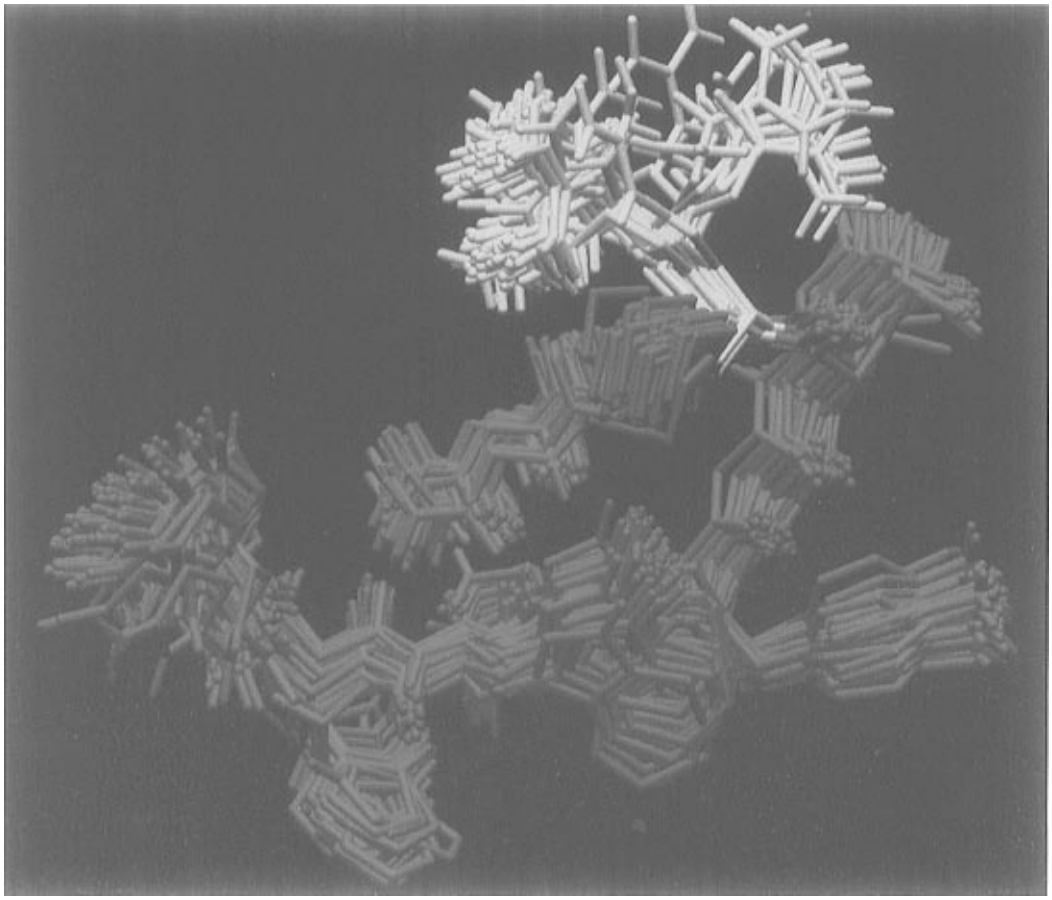

Figure 7. Superimposition of 70 structures that make up the major conformational family of GnRH obtained from conformational memories. While there is a large amount of fluctuation in the backbone, and an even greater amount of fluctuation in the side chains (especially Arg8), there is a clear β-type turn from residues 5−8 in this family.

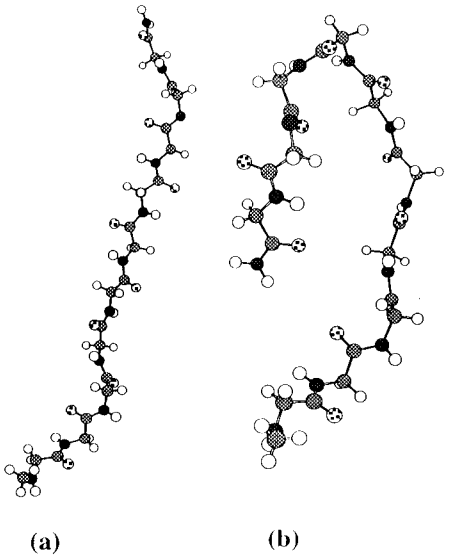

Figure 8. Backbone trace of a representative of two conformational families of Lys8-GnRH obtained from conformational memories. The structure with the β-type turn comes from a family with an approximately 3% population. The structure with the straight backbone comes from a family with approximately 70% population.

9, with the entire molecule shown in red, except for Lys8 which is shown green. Because the Lys8-GnRH has a low affinity for the GnRH receptor, but elicits the same response once it interacts with the receptor, it is tempting to suggest that adoption of a large population of β-type turn conformations is a key requirement for hormone−receptor recognition. This inference agrees with earlier proposals in the literature (e.g., see ref 16), and is supported by results from additional conformational memories simulations on the structural characterization of eight other GnRH analogs that exhibit different distributions between the β-turn-like structures and the fully extended conformations of the backbone (Guarnieri et al., unpublished results). It is particularly noteworthy that our simulations lead to the same conclusions regarding the importance of the bent structure that were drawn from combined NMR and molecular dynamics studies of conformationally constrained GnRH analogs.[16]

Structural Comparison to a Constrained GnRH Analog. To test the key inference from the present simulations of GnRH analogs, regarding the correlation between the population of β-type turn structures and the affinity for the receptor, we compared several samples from the most populated conformational family of GnRH obtained from conformational memories to a structurally constrained cyclic decapeptide GnRH analog.[15] The conformation of this cyclic decapeptide was determined from NOE data using 2-dimensional NMR techniques.[15] These experimental studies concluded that residues 6 and 7 formed a type II′ β-turn and residues 1 and 2 formed a type II β-turn. Additionally, it was concluded that a weak hydrogen bond existed between the Arg8 −NH and the Tyr5 −CO, and a stronger hydrogen bond between the D-Trp3 −NH and the β-Ala10 −CO. To allow for the comparison, a structure of this GnRH analog was built in Macromodel 4.5 according to the specifications[15] and using the β-turn definitions of Hutchinson and Thornton.[37] This reconstructed GnRH analog was compared to the GnRH structures obtained from the conformational memories described above.

(37) Hutchinson, E. G.; Thornton, J. M. *Protein Sci*. **1994**, *3*, 2207−2216.

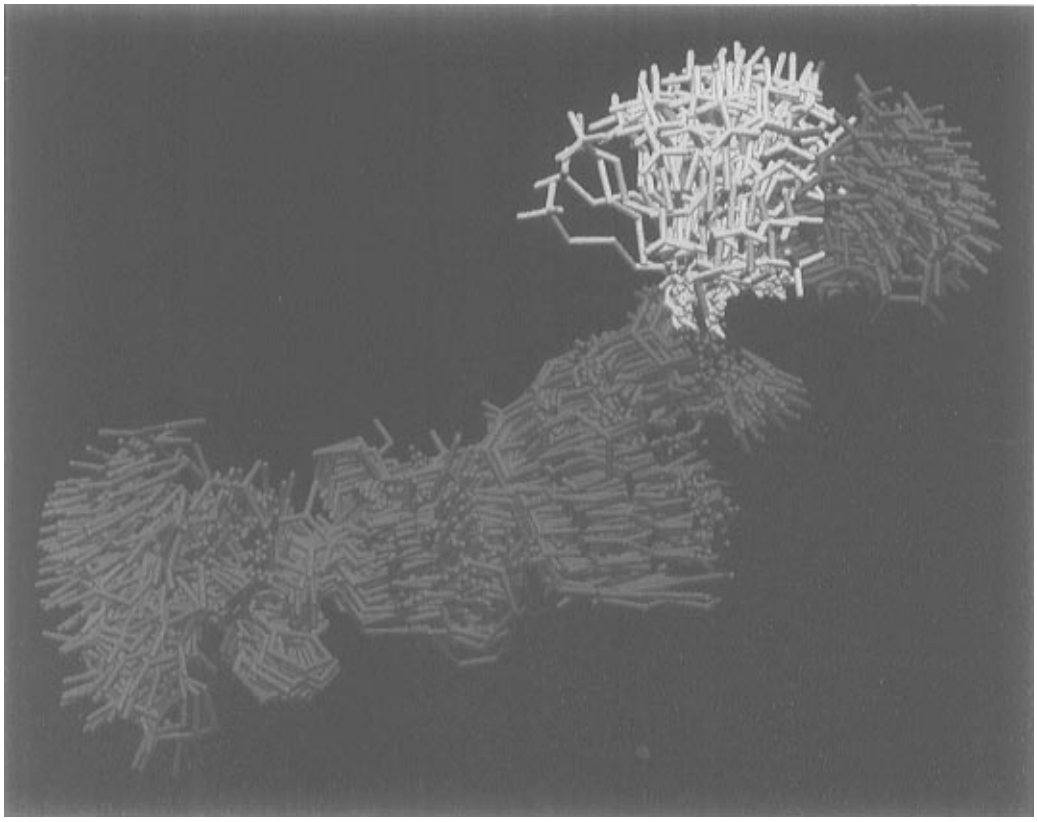

Figure 9. Superimposition of 70 structures that make up the major conformational family of Lys8-GnRH obtained from conformational memories. While there is a large amount of fluctuation in the backbone, and an even greater amount of fluctuation in the side chains (especially Lys8), the backbone is clearly extended.

Several of the members of the major conformational family of the GnRH obtained from the conformational memories were selected at random and superimposed on the reconstructed geometry of the analog, using the 11 backbone atoms from the Tyr5 −CO to the −N of Pro9. All computationally derived structures superimposed on the reconstructed structure with RMS deviation in a range of 0.6−0.8 Å. An illustration of the superimposition is shown in Figure 10. Clearly, the computationally derived structure is closely related to the reconstructed backbone of residues 5−8 of the experimentally derived peptide structure. The structures diverge between the N-terminus and residue 4, and a superimposition of all backbone atoms results in a 5 Å RMS deviation.

GnRH Conformations from a Buildup Procedure. Recently, Nikiforovich and Marshall[38] constructed low-energy conformations of GnRH using the ECEPP program.[39] We have reconstructed eight conformations from the published list of backbone dihedral angles and a list of side chain dihedrals graciously provided by the authors.[38] The energies of these reconstructed peptide structures were compared with representatives from the major families of GnRH found using conformational memories. The optimal geometries of GnRH obtained from the two computational methods were quite different, and the energies of the eight conformations obtained from ECEPP calculations were 300−400 kJ/mol (20−25%) higher than those calculated for the conformations generated using conformational memories. It is unlikely that this large difference can be attributed solely to the use of different force fields in the definition of optimal conformations, since a recent comparative study resulted in very similar low-energy Met-Enkaphalin structures.[40] However, a major source of difference may be the use of a GB/SA water model in the conformational memories approach, and perhaps a more complete exploration of the conformational space.

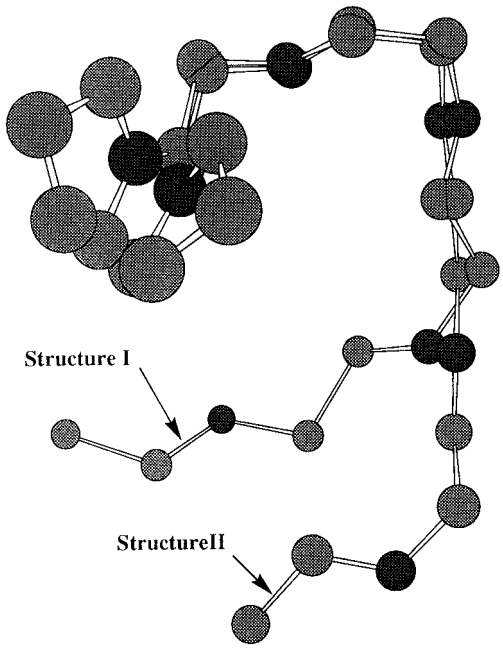

Figure 10. Superimposition of a high-affinity GnRH cyclic analog (structure I) and a representative of the major GnRH conformational family (structure II). Eleven backbone atoms from residues 5−8 were used for the superposition.

Exploration of the Unpopulated Regions. As a stringent test of the completeness of the conformational exploration, we

(38) Nikiforovich, G. V.; Marshall, G. R. *Int. J. Peptide Protein Res.* **1993**, *42*, 171−180.

(39) Dunfield, L. G.; Burgess, A. W.; Scheraga, H. A. *J. Phys. Chem.* **1978**, *82*, 2609−2616.

(40) Montcalm, T.; Cui, W.; Zhao, H.; Guarnieri, F.; Wilson, S. R. *J. Mol. Struct.*: THEOCHEM **1994**, *308*, 37−51.

performed extensive sampling from the *unpopulated* regions of the conformational memories for several key dihedral angles involved in the formation of the β-turn of the GnRH. With one exception, this sampling produced high-energy structures in all cases, as expected. The one interesting exception occurred during the sampling of the unpopulated regions of the ϕ angle of Gly6. This sampling produced a structure only 20 kJ/mol higher in energy than the best GnRH structure. The dihedral value came from a bin that had a 0.6% population at 345 K, but had a 0% population at 310 K and therefore was not included in the populated portion from which the MC biased sampling was done. A simple way to avoid missing a very-low-probability low-energy structure when the biased sampling is performed at 310 K is to use the probability weights from a higher temperature. Our exploration of the unpopulated regions of the ϕ angle of Gly6 at temperatures 100−200 K above 310 K eliminated this problem. The small drawback, however, is that 44% of the dihedral space of Gly6 is unpopulated at 310 K and only 33% is unpopulated at 473 K. Thus, a safety factor during the biased sampling run involves exploring about 10% more dihedral space per rotatable torsion, but ensures the enclosure of all populated areas. Conformational regions that exhibit 0% population in the calculation of the isolated peptide in water at 310 K may still be of biological importance, if some of these conformations can be induced by the interaction energies of the peptide with the receptor. The finding that regions unpopulated at 310 K are in fact populated at temperatures higher by only 100 K (corresponding to an energy difference of only a fraction of 1 kcal/mol) indicates the feasibility of such "receptor-induced" conformations.

Conclusions

Applied to the decapeptide hormone GnRH, the method of conformational memories was shown to provide a powerful practical solution to the complex problem presented by the flexibility of polypeptides with a large number of conformational degrees of freedom. With the study of the flexible decapeptide, the method was shown to be capable of achieving complete sampling of the conformational space, to converge in a very practical number of steps, and to be capable of overcoming energy barriers efficiently.

The results of the conformational study support a relation between the β-turn structure identified as the major conformational family of GnRH and the high affinity for the GnRH receptor. While these inferences were inherent in the results from earlier investigations of conformationally restricted GnRH analogs, the present study provides unbiased support for this mechanistic hypothesis based on a complete exploration of the conformational space of the peptide hormone itself and its unconstrained congeners. Because the method seems to have produced the lowest energy conformers reported for GnRH from a full exploration that is economical and practical, its general application to the study of peptide structure−function relations should continue to produce important mechanistic insights and powerful guides for ligand design.

Acknowledgment. The authors thank Drs. Stuart Sealfon and Arnold Hagler for critical reading of the paper. The work was supported in part by grants from the National Institutes of Health, R01 DK46943, K05 DA00060, 5T32DA07135. Computations were performed on the supercomputer systems at the Pittsburgh Supercomputer Center (sponsored by the National Science Foundation) and the Cornell National Supercomputer Facility (sponsored by the National Science Foundation and IBM), as well as at the Advanced Scientific Computing Laboratory at the Frederick Cancer Research Facility of the National Cancer Institute (Laboratory for Mathematical Biology).

JA952745O