

A Tree-Based Algorithm for Determining the Effects of Solvation on the Structure of Salivary Gland Tripeptide $\text{NH}_3^+\text{-D-PHE-D-GLU-GLY-COO}^-$

Essam Metwally, Heba A. Ismail, Joseph S. Davison, and Ronald Mathison

Department of Physiology and Biophysics, The University of Calgary, Calgary, Alberta, Canada

ABSTRACT A D-enantiomeric analog of the submandibular gland rat-1 tripeptide FEG (Seq: $\text{NH}_3^+\text{-Phe-Glu-Gly-COO}^-$) called feG (Seq: $\text{NH}_3^+\text{-D-Phe-D-Glu-Gly-COO}^-$) was examined by molecular dynamics simulations in water. Previous in vacuo simulations suggested a conformation consisting predominantly of interactions between the Phe side chain and glutamyl-carboxyl group and a carboxyl/amino termini interaction. The solvated peptide was simulated using two approaches which were compared—a single 400-ns simulation and a “simulation tree.” The “tree” approach utilized 45 10-ns simulations with different conformations used as initial structures for given trajectories. We demonstrate that multiple short duration simulations are able to describe the same conformational space as that described by longer simulations. Furthermore, previously described in vacuo interactions were confirmed with amendments: the previously described head-to-tail arrangement of the amino and carboxyl termini, was not observed; the interaction between the glutamyl carboxyl and Phe side chain describes only one of a continuum of conformations present wherein the aromatic residue remains in close proximity to the glutamyl carbonyl group, and also interacts with either of the two available carboxyl groups. Finally, utilizing only two separate 10-ns trajectories, we were able to better describe the conformational space than a single 60-ns trajectory, realizing a threefold decrease in the computational complexity of the problem.

INTRODUCTION

Saliva has long been known as a fluid with healing properties (Angeletti et al., 1992), a notion which has only recently been rediscovered. An array of physiological and immunological functions have been attributed to factors secreted directly into both blood and saliva, elevating the role of the salivary glands from accessory digestive organs to an organ critical to oral and systemic homeostasis (Mathison et al., 1993). For example, the secretion of epithelial growth factor helps maintain intestinal epithelial integrity (Li et al., 1993) and unidentified factors released from salivary glands suppress superoxide production by neutrophils, thereby preventing undue tissue damage to the gums and oral mucosa (Fu et al., 1991). Even more recently, we have described immuno-compromisation as a result of removal of the submandibular glands in rats, rendering them unable to respond appropriately to endotoxic and anaphylactic hypotension. This further suggests an integral role for these organs in responses to hemodynamic shock (Mathison et al., 1993, 1997a,b). These shock-reducing effects were found to be mediated by a seven amino-acid peptide, submandibular gland peptide-T (SGP-T; $\text{NH}_3\text{-Thr-Asp-Ile-Phe-Glu-Gly-Gly-COO}^-$; see Mathison et al., 1997a). Subsequent studies established that protection against anaphylactic reactions afforded by SGP-T could be mediated by the tripeptide fragment, FEG (Seq: $\text{NH}_3\text{-Phe-Glu-Gly-COO}^-$), and its D-isomeric analog, feG (Seq: $\text{NH}_3\text{-D-Phe-D-Glu-Gly-COO}^-$), without loss in biological activity (Mathison et al., 1997a,b, 2001a).

In an attempt to understand peptide conformation and biological activity relationships we examined the behavior of numerous analogs of FEG with molecular dynamics simulations in vacuo (van der Spoel et al., 2001), the least costly of computational techniques (Metwally et al., 2002). With this approach a question always remains: does the absence of water invalidate the conformations obtained using in vacuo simulations? The short answer would be *no*, since our studies showed a very high correlation between biological activity and in vacuo determined conformations. However, it is indisputable that in an in vivo setting, a peptide will spend a great deal of time in a solvated state. Effectively, the range of conformations available to a peptide within a solvent will determine the capacity of the peptide to adopt a suitable structure as it binds to its receptor. Hence, it is necessary to examine the effects of solvation on peptide conformation.

Although molecular dynamics simulations are an effective means of exploring structure and attempting to predict the behavior of molecules, the 1–100 ns timescales commonly used in computational chemistry to evaluate molecular structure (van der Spoel et al., 2001) remain orders-of-magnitude below the micro- and milliseconds commonly associated with the activation of biochemical and biological processes. For relatively small molecules such as those discussed herein, two months were necessary to simply simulate 400 nanoseconds of dynamics. Since systems containing greater numbers of molecules, solute, and solvent can take years to complete (Tai et al., 2001), it is apparent that new techniques are required to computationally define the biologically relevant structures of proteins and peptides.

With these factors in mind, the 14–15 analogs which were previously considered (Metwally et al., 2002) would necessitate between three and four years of real-time to simulate for a duration of 400 ns, assuming no change in the

Submitted October 1, 2002, and accepted for publication June 13, 2003.

Address reprint requests to Dr. Essam Metwally, Tripos Inc., 1699 S. Hanley Rd., St. Louis, MO 63144. Tel.: 314-647-1099; Fax: 314-647-9241; E-mail: emetwall@tripos.com.

© 2003 by the Biophysical Society

0006-3495/03/09/1503/09 \$2.00

capabilities of computer speed and power. Hence, of necessity, only one peptide, feG, was selected for the solvated simulation in this study. However, it has not been established that a single 400-ns simulation is the best technique for defining the preferred conformation of a peptide. Although a single simulation provides a continuous, unbroken sampling of the conformational space available to the peptide, it is desirable to find an alternative analytical approach which would reduce the time required to perform long duration simulations. Hence, we propose a “simulation tree” consisting of numerous 10-ns simulations to determine whether the conformational space available to feG can be adequately sampled using numerous shorter duration simulations executed in parallel.

Substitution of multiple trajectories for a single extended duration simulation is not a new idea. The locally enhanced sampling method is based on mean-field technique where additional computational effort may be focused on portions of a system, thereby selectively enhancing the sampling for regions of interest (Elber and Karplus, 1990). Multiple copies of the region are superimposed and are free to move and sample available conformational space, with all copies exerting a net force equivalent to that found for a single copy. Effectively this results in multiple trajectories being generated from a single chimeric trajectory (Elber and Karplus, 1990; Straub and Karplus, 1991; Roitberg and Elber, 1991). The replica exchange method (REM) allows paired replicas, which vary in temperature, to be exchanged during a simulation, thus enhancing sampling of conformational space (Hukushima and Nemoto, 1996). Sugita et al. (2000) further expand on REM by allowing replicas with different potential energy parameters to be exchanged in addition to those of different temperature. Most recently, Rhee and Pande (2003) explored a further enhancement of the technique which does not suffer from many of the deficiencies of REM. However, what differentiates our proposed technique is that it is an example of trivial parallelism. The simulations may be as rigorous or as simple as a user desires, and since each trajectory is inherently independent of any of the branches at its hierarchical level, it scales rapidly and efficiently on any number of heterogeneous computers, which cannot be said for REM (Rhee and Pande, 2003).

We found that hydration of feG resulted in a loss of the ability of the C-terminal carboxyl group to interact with the N-terminus amino group. However, the remaining residue's one aromatic side-chain interaction with the glutamic carboxyl group (Metwally et al., 2002) remained present but with enhanced mobility, exhibiting the ability to interact with either the C-terminus or position 2 carboxyl. Furthermore, the proposed “simulation tree” comparably described the conformational space available to the peptide with respect to the single 400-ns duration simulation but required only one week to complete as compared to a month for the continuous unbroken simulation. Although it may be true that the unbroken simulation could, using an efficiently

parallelized multiprocessor algorithm, be completed in the same timespan as the whole tree, we found that the multiple trajectory approach was able to describe the same conformational space with fewer trajectories, and hence, reduced computational cost.

MATERIALS AND METHODS

Computer hardware

The majority of simulations were performed using the MACI Alpha Cluster at the University of Calgary (<http://www.maci.ca>). The only exception to this was the single 400-ns unbroken simulation which was performed on a dual processor 1.2 Ghz Athlon with 512 MB of RAM running Red Hat Linux 7.1.

Molecular dynamics

All simulations were performed with GROMACS, v3.0 (Berendsen et al., 1995; Lindahl et al., 2001). Neighbor-lists were generated using a grid approach and updated every 10 steps (0.02 ps). The GROMOS96 43a1 official distribution force field was used with cutoffs of 1.0 nm for both Coulombic and Van der Waals interactions. The peptide was created in an extended conformation with all backbone torsions set to the transconfiguration, and was placed in a rhombic dodecahedron constructed from a 3.2-nm \times 3.2-nm \times 3.2-nm cube for a final volume of ~ 21 nm³ and containing 741 molecules of water. The container size was such that the peptide would not see itself in an adjacent container during the course of the simulation (van der Spoel et al., 2001). This was followed by energy minimization using a steepest-descent algorithm to remove any close contacts resulting from solvation of the peptide. The peptide position was then restrained and the water allowed to equilibrate around the peptide, so as to allow the solvent to form a continuous bath without any holes. The resultant system formed the starting position for all subsequent simulations. To ensure a constant temperature of 300 K for the duration of the simulation, peptide and solvent were weakly coupled to a temperature bath with a relaxation time of 0.1 ps at 300 K (Berendsen et al., 1995; Burgi et al., 2001). Initial molecule velocities for the simulations were taken from a Maxwell-Boltzmann distribution at 300 K.

These initial setup simulations produced a peptide structure in water, which will be referred to as the “seed structure.” This structure is the starting point for both the initial 10-ns tree simulation, referred to as the seed trajectory or seed simulation, as well as the 400-ns simulation. It forms a common reference point and comparison structure for subsequent analysis of the trajectories.

A single 400-ns simulation was processed and took ~ 2 months to complete. Simultaneously, a 10-ns trajectory “tree simulation” was initiated. As the simulations could be conducted in parallel, three sets of 15 parallel simulations, conducted in series, were completed in ~ 10 days.

For the 10-ns trajectory tree, an initial 10-ns trajectory was performed using the methods described above. Subsequently, a root mean squared deviation (RMSD) of the trajectory with respect to the initial seed structure was evaluated. The five frames which showed the smallest deviation were used as starting structures for trajectories S0, S1, S2, S3, and S4. The five frames exhibiting the greatest deviations were used as starting structures for trajectories S5, S6, S7, S8, and S9. Finally, the five frames which were found between the two extremes of least and most deviated were used as the starting structures for trajectories, SA, SB, SC, SD, and SE. Each of these 10-ns trajectories were simulated, and RMSD of the trajectories were again evaluated. The most deviated structure as compared to the initial seed structure was used to begin the final trajectory termed S#A, whereas the least deviated structure was used to begin a trajectory termed S#B where # indicates the source trajectory (i.e., if the source trajectory were S0, then “#” would become 0, and the trajectories would be named S0A and S0B). In total this format yielded 45 distinct 10-ns trajectories, exclusive of the initial 10-ns seed run. The format of this procedure is illustrated in Fig. 1.

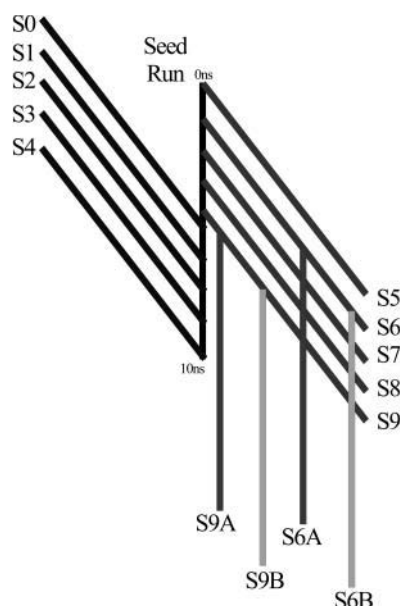


FIGURE 1 A graphical representation of the simulation tree. The tree as displayed is incomplete. Each bar represents a 10-ns simulation period. The bars to the right of the seed trajectory represent starting structures which exhibited the greatest deviation as compared to the start of the seed run. Bars on the left represent starting structures which were least deviated from the start structure of the seed run. The image does not show an additional set of five simulations originating from the seed run which were started with structures that represented the average of the seed run. Furthermore, each of these branch runs possesses two additional runs denoted by an *A* or *B* appended to their source trajectories, representing least deviated and most deviated structures with respect to the original seed structure.

Clustering

The GROMOS clustering algorithm described by Daura et al. (1999) was used to determine cluster membership with a 0.05-nm backbone RMSD chosen to determine cluster membership. The structure with the greatest number of neighbors was used as the center of a cluster. All clusters are mutually exclusive, so a structure can only be a member of a single cluster.

To determine whether the divergent nature of each trajectory's lower population clusters would result in identification of a greater number of clusters overall, a single trajectory from group S0–S4 was combined with one from S5–S9 and clustered as a single trajectory. These pairings were contrasted with the minimum time required for the extended duration simulation to describe its conformational space.

RESULTS

400-ns simulation

Fig. 2 shows the backbone atom RMSD for feG over the course of the 400-ns simulation. It should be noted that the RMSD of all trajectories were calculated with respect to the seed structure.

10-ns simulation tree

Fig. 3 shows the backbone RMSD for trajectories S0, S0A, and S0B. These trajectories are representative of what is

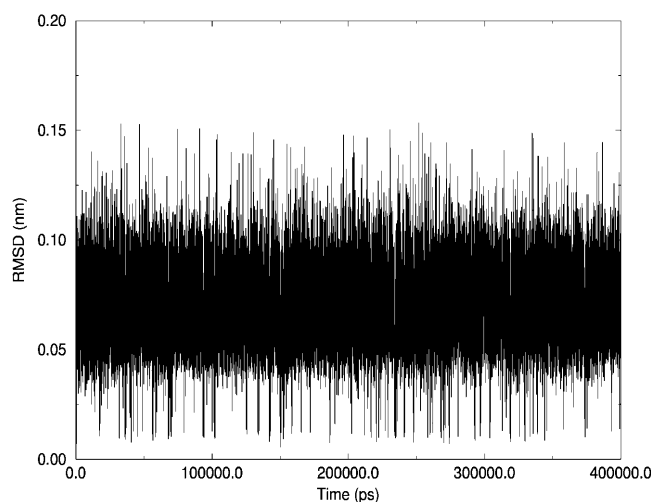


FIGURE 2 The backbone root mean square deviation (RMSD) of feG during the 400-ns, extended duration trajectory as compared to the seed structure. The RMSD is with relation to the backbone position. Although quite compressed, it is clearly evident that feG spends the greatest part of the simulation within a very narrow range, between a backbone RMSD of 0.05 nm and 0.10 nm.

observed in all RMSD calculations for the other trajectories within the tree. Again, the RMSDs are taken with respect to the original seed structure.

Clustering

The 400-ns simulation provided 23 distinct clusters, whereas 10-ns simulations produced an average of 15 clusters ranging from 11 to 22 clusters. As evidenced by the overlay of all conformations which are representative of cluster 1 (Fig. 4)

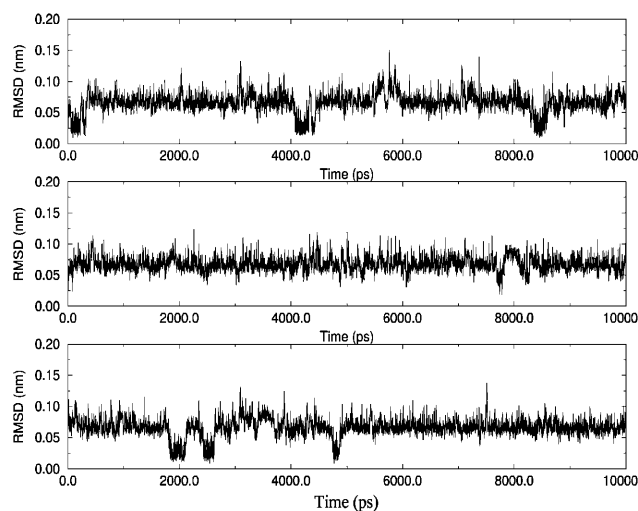


FIGURE 3 Representative backbone root mean square deviation (RMSD) for trajectories S0, S0A, and S0B. Each trajectory is 10 ns in duration. Although some fluctuation occurs, it is evident in these traces that feG spends the majority of the simulation within a narrow range of conformations found between 0.05 nm and 0.10 nm. This range of conformations is what is classified as cluster 1 (Fig. 4).

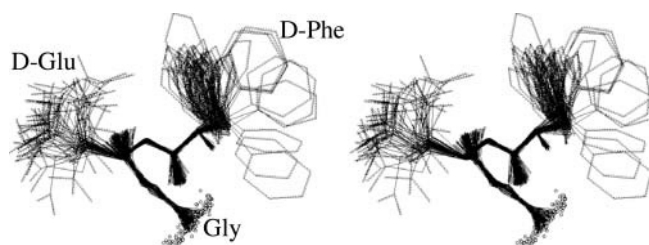


FIGURE 4 Stereoscopic image of feG in cluster 1. The central structure for each trajectory's cluster 1 has been overlaid. The high degree of similarity over the extent of the backbone is immediately apparent. The only differences between these structures is the rotation of each side chain about χ_1 , χ_2 , and χ_3 . However, even these rotations are minor in comparison to the similarity between the structures. Note the close proximity of the Glu carbonyl to the Phe ring. This is similar to previous observations resulting from *in vacuo* simulations. However, in solvated simulations the aromatic side chain is able to take a conformation which involves either an interaction with the Glu-carboxyl, as in *in vacuo*, or with the Gly-carboxyl.

feG maintains a distinct and highly conserved structure. Regardless of the length of the simulation, feG was observed to exist within cluster 1 for $\sim 84\%$, cluster 2 for $\sim 7\%$, cluster 3 $\sim 4\%$, and cluster 4 $\sim 2\%$, with all other conformations occurring no more frequently than 1% of the time of any given simulation (Fig. 5). A pairwise comparison of the structures within the trajectory indicated that the most common structures are highly conserved throughout the simulations. However, conformations between simulations diverged as the frequency or size of a cluster decreases. This is a result of the different conformational space sampled by each individual trajectory.

Of note is the effect of a given trajectory's source structure deviation from the initial seed structure on the number of clusters obtained. Trajectories S0–S4 (Fig. 6) exhibited the smallest deviations from the source structure and resulted in the greatest number of clusters. The trajectories with the greatest deviation, S5–S9, from the seed structure produced a slight though significant decrease in identified clusters,

whereas those trajectories which were most centralized, SA–SE, produced the fewest number of clusters. This trend was also observed in the subtrajectories with S(0–4)A and S(0–4)B exhibiting the greatest number of clusters and other trajectories following the trend of their parent trajectory.

When the extended simulation was divided into shorter trajectories, 60-ns trajectories were required, on average, to detect 22 clusters. In contrast, the trajectories initiated with the least and most deviated conformation from the seed conformation, S0 and S5, describe 18 and 16 distinct clusters, respectively. However, when these trajectories were considered in unison, 25 distinct clusters were identified. Cluster 1 accounted for $\sim 82\%$ of the combined trajectories and $\sim 84\%$ of the 60-ns samples. S1 and S6 yielded similar cluster distributions and identified 22 distinct clusters. These values continue to decrease as the deviation between the two seeds decreases, with S2–S7, S3–S8, and S4–S9, identifying 20, 15, and 16 clusters, respectively. These values are summarized in Table 1.

Structure

Representative conformations from cluster 1 can be seen in Fig. 4. The conformations are characterized by the position 1 aromatic group remaining in close proximity to the carbonyl of the position 2 glutamic acid. Conformations where the aromatic ring shifts interactions between the glutamyl carboxyl and C-terminus carboxyl are also observed. No interaction is observed between the carboxyl and amino termini, with the amino terminus pointing off into the solvent. All structures found in cluster 1 from all simulations exhibit a very high degree of homogeneity in the peptide backbone (Table 2, Fig. 7).

DISCUSSION

This novel class of anti-inflammatory submandibular gland peptides, of which feG is an example, possesses myriad

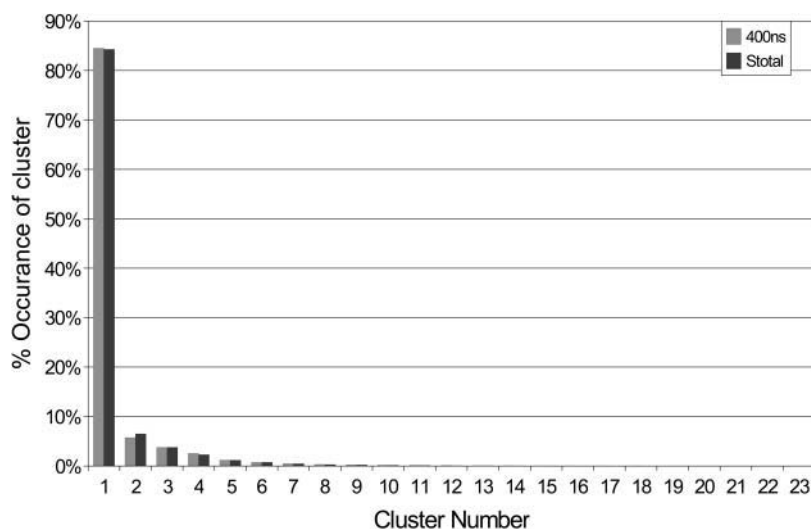


FIGURE 5 Percentage of cluster occurrence for all simulations. In light gray is the occurrence of each cluster over the full 400-ns duration. In dark gray is the aggregated results of the tree simulation, representing a total time of 450 ns. As can be seen, both sets of simulations yielded nearly identical occurrences of any given cluster. Only clusters above number 5 possessed $>1\%$ of the cluster distribution. All remaining clusters are observed $<1\%$ of the time in any given simulation.

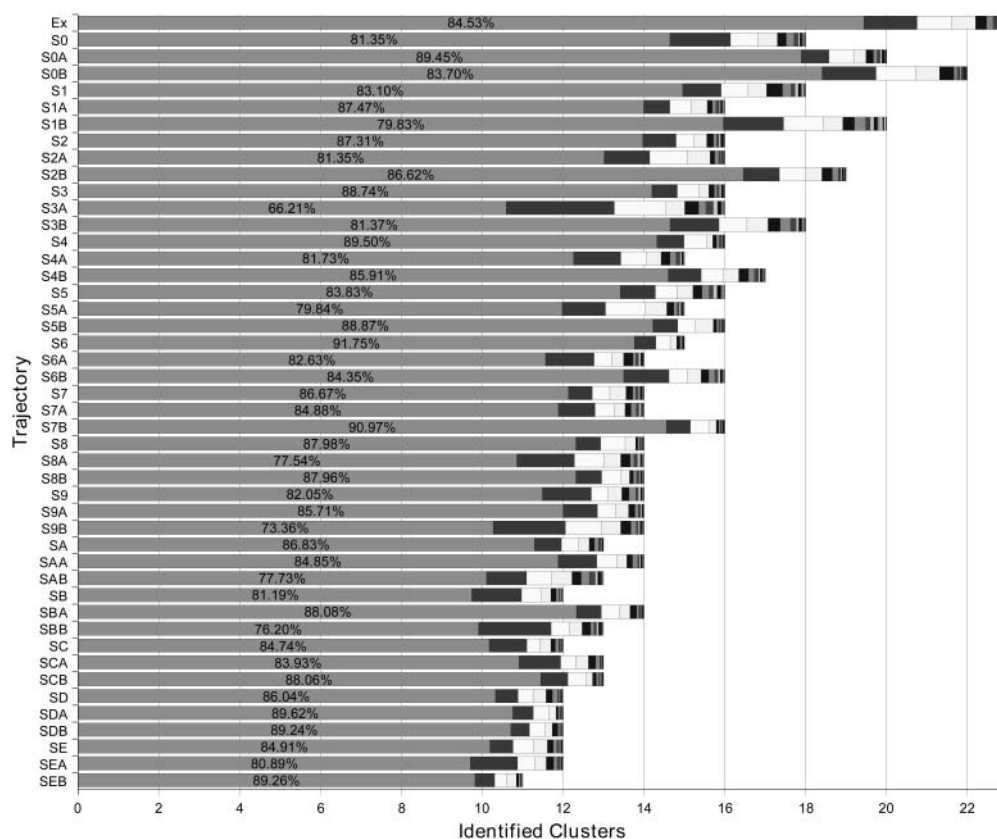


FIGURE 6 A comparison of each trajectory and the number of clusters identified. The length of each bar corresponds to the total number of identified clusters, whereas each of the stacked components represents the percentage of a given trajectory found in a particular cluster. The percentage displayed in each bar is the occurrence of cluster 1 and is given for reference. Clusters increase from left to right, with 1 at left through the maximum for each trajectory.

activities, including the ability to ameliorate hypotensive responses arising from systemic endotoxic and anaphylactic shock (Mathison et al., 1997a,b, 2001a), intestinal anaphylaxis (Mathison et al., 1998), and cardiac inflammation (Mathison et al., 2001b). How this class of peptides mediates their biological effects remains unclear, and an understanding of the conformations of these peptides may help in elucidating their mechanism of action. Previously, we described an in vacuo examination of a large number of analogs with an aim at identifying the structural elements which were necessary for biological activity. Although in vacuo modeling is often considered to be adequate for simulations of peptide structure (Moore, 1994; Moore et al., 1995), the fact

remains that peptides, and indeed, any molecule found in vivo is in a solvated state whether this solvation consists predominantly of water or lipid. Indeed, simulations using adrenocorticotropin in a solvated dodecylphosphocholine micelle have demonstrated that interactions with both water and the micelle have a profound impact on intramolecular hydrogen bonding and conformation (Gao and Wong, 2001). Despite the conjecture that regions in close proximity to the cell membrane, a predominantly nonpolar environment, may approximate an in vacuo state, what is indisputable is that this particular approximation was made to favor computational simplicity rather than physical reality. However, our own extensive modeling of tripeptides (Metwally et al.,

TABLE 1 Number of clusters identified by the first branches of the simulation tree, trajectories S0–SE, as grouped by starting structure deviation from the seed structure of the tree as well as the results of combining trajectories from the most and least deviated groups

RMSD grouping	Trajectories	Average clusters identified*	SD of identified clusters
Least [†]	S0, S1, S2, S3, S4	16.8	1.1
Average [‡]	SA, SB, SC, SD, SE	12.2	0.45
Most [§]	S5, S6, S7, S8, S9	14.6	0.89
Combined	S0/S5, S1/S6, S2/S7, S3/S8, S4/S9	19.6	4.2

*All categories are significantly different from each other group ($p < 0.001$).

[†]Starting structure RMSD for S0–S4 < 0.01 nm.

[‡]Starting structure RMSD for SA–SE between 0.05 nm and 0.10 nm.

[§]Starting structure RMSD for S5–S9 > 0.15 nm.

TABLE 2 Average measurements of cluster 1 backbone overlay

Angle (°)*		Torsional Angle (°)*		Distance (Å) †	
1–4	114.5	1–6	146.6	1–14	7.25
2–6	117.0	2–7	−179.5	2–13	5.91
4–7	126.8	4–10	−80.4	4–12	3.51
6–10	116.5	6–12	83.0	6–10	2.46
7–12	114.8	7–13	173.2	1–7	4.89
10–13	127.0	10–14	−178.4	7–14	4.87
12–14	108.1				

*All angles as given start at the first atom in a pair as defined by the figure and end at the second atom in the pair.

†The distances as given are those for a straight line drawn from the first atom in the pair to the second atom in the pair.

2002) demonstrated a very high degree of correlation with biological activity. Despite this, the model was unable to completely explain the activities of all molecules, hence, it can be concluded that the *in vacuo* model is incomplete. In light of the profound impact that solvent can have on a structure (Gao and Wong, 2001), we turned to solvated simulations.

The observations from *in vacuo* simulations support a conformation wherein an arrangement of the position 2 glutamic carboxyl remains in close proximity to the aromatic ring of position 1. In addition, we observed a head-to-tail arrangement consisting of the carboxyl terminus interacting with the amino terminus (Metwally et al., 2002). Our analysis of the clustering of feG in water (Fig. 4) suggests that the Phe-side chain Glu carboxyl interaction holds true, but the interaction between C-terminal carboxyl and N-terminal amino group is not found in water. With *in vacuo* simulations dielectric masking of the point charges of the carboxyl terminus COO^- and the amino terminus NH_3^+ cannot occur. Hence, the two terminal groups acted together to distort the highly flexible backbone into a head-to-tail arrangement. Clearly, the presence of water prevented this head-to-tail interaction from occurring. The interaction of residues 1 and 2, nevertheless, remains strong (Fig. 4). However, in the absence of the head-to-tail interaction there is less restriction on the positioning of the aromatic side chain and the glutamic-carboxyl in water. Thus, the peptide is no longer locked into this conformation by this predominant, if not

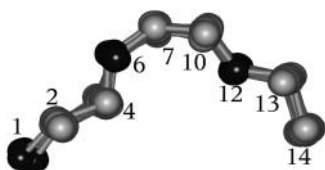


FIGURE 7 The overlaid backbone of all feG cluster 1 central structures. The measurements of the backbone are summarized in Table 2. Note the very high degree of overlap between all structures. This backbone is likely the necessary scaffold structure required for a biologically active compound.

sole, constraining interaction. It is now possible to have the C-terminus carboxyl interacting with the aromatic side chain in place of the glutamyl carboxyl. However, the closest interaction in ~90% of cluster 1 conformations occurs between the aromatic ring and the glutamyl carbonyl restricted to a separation of 2.5–3.5 Å. Carbonyl-aromatic interactions are not unique to feG and occur frequently in many structures (Kim et al., 2001; Yoshitake et al., 2000). This arises in part from the distribution of positive charge surrounding and in the plane of the aromatic ring with the quadrupole perpendicular to the plane (Dougherty, 1996). Hence, it appears that the ring predominantly favors a conformation wherein it remains within a close proximity to one of the carboxyl groups and, especially, to the glutamyl carbonyl. Another important feature emerging from these studies is the high degree of backbone overlap found in all the simulations (Fig. 7). The overlap leads to a suggested structure for the peptide scaffold on which mimetic synthesis could be based (Fig. 7, Table 2).

Nonbonded interactions between atoms are typically truncated at a specific radius, thereby reducing the number of interactions and the computational time required to complete a molecular dynamics simulation (McCammon and Harvey, 1987). The use of cutoffs has been examined by various authors (Norberg and Nilsson, 2000; Schreiber and Steinhauser, 1992; Smith and Pettitt, 1991). Smith and Pettitt compared distance cutoffs with Ewald techniques and found that, for aqueous simulations, the differences between techniques was relatively small. However, for peptides in saline solution the simulations were quite sensitive to cutoff conditions. Schreiber and Steinhauser examined the effects of increasing cutoff radius on simulation results and noted that in the case of a 17-residue helical peptide, that even the commonly used 1.4-nm radius was unable to maintain the α -helical configuration of this peptide (Schreiber and Steinhauser, 1992). However, the methodology we have presented was intended to contrast a single extended duration simulation with our simulation tree. Provided the treatment of cutoffs is consistent between the two simulation sets, the results may be interpreted in that context. Furthermore, in contrast to Schreiber and Steinhauser, the constructed peptide is completely encompassed within the cutoff radius. In most reported cases of cutoff-induced artifacts, these tend to occur at the cutoff boundary, in the case of solvent (Mark and Nilsson, 2002), or the biopolymers were not completely encompassed within the cutoff radius. Schreiber and Steinhauser observed fraying and loss of secondary structure; however, this was restricted to the peripheral portions of the peptide while the core remained intact (Schreiber and Steinhauser, 1992). In fact, in the case of Smith and Pettitt (1991) the peptide was completely encompassed within the 1-nm cutoff radius and remained relatively unaffected by the cutoff when compared to Ewald summation in the absence of salt. Ewald summation techniques could certainly be used within the simulation tree, particularly where the treatment

of long-range interactions are of concern. The design of a simulation tree can be catered to the precise requirements of any investigator.

Of importance are the differences in simulation length. Previously it has been proposed that to adequately sample the conformational space available to short peptides, lengthy simulations were necessary (Burgi et al., 2001). However, a number of authors have pointed out that the length of simulation time is intimately dependent on the speed with which a given system is able to converge and describe the “essential” low-dimensional subspace (de Groot et al., 2001). Although many argue that this subspace is inaccessible to short duration simulations (Balsera et al., 1996; Clarage et al., 1995) others have been successful in describing this subspace in short simulations (de Groot et al., 1996). Herein we present an alternative to computationally intensive long simulations (Tai et al., 2001), while offering a means for adequate sampling of the conformational subspace. The simulations were divided into a series of shorter simulations where the first 10-ns trajectory was used as a “seed trajectory.” This trajectory can be thought of as the trunk of the tree with subsequent trajectories representing the branches of this tree. Despite the markedly shorter length of each simulation, the overall distribution of clusters remained similar (Fig. 5). Indeed, the primary cluster was consistently found to account for ~84% of the conformations identified in any given simulation. However, the similarity between simulations began to degenerate as the cluster index increased; i.e., as we move further from the most common structures in a given simulation, the variability in peptide conformations which characterize clusters begins to increase. This is indicative of each trajectory sampling its own particular conformational space.

Interestingly, when trajectories are considered independently of each other (Fig. 6), trajectories initiated with the least or most deviated conformations, as compared to the original seed structure, identified more clusters than those initiated with the “average” conformation (Fig. 6). This increase in number of clusters was statistically significant. The initial purpose for selecting the starting points of trajectories in this manner was to promote a bias toward more deviated structures such that it could be appreciated whether the observed clusters were indeed the predominantly occurring conformations. This is precisely what occurred with the tree. Despite the observation of a greater number of clusters the predominantly occurring cluster conformation remained identical in all trajectories. Thus, this technique allowed for a greater sampling of the conformational space available to the peptide. As noted by De Groot et al. (2001), provided that the conformational transitions occur on a short timescale which is accessible within our simulation time, the described dynamics may be valid. This is indeed the case with the tree-based algorithm described, since the predominant conformation was consistently observed in all simulations. It may be argued that the selection of deviated

structures only served to bias the conformational space analyzed toward deviated conformations. This is true, but by selecting conformations on both ends of the spectrum, i.e., most and least deviated, as well as sampling the domain within the center, an adequate statistical sampling has been achieved as evidenced by consistent identification of clusters (Fig. 5). In fact, of all the simulations, only those which were initiated with the central structures, SA–SE, actually classified the first structures of the trajectories as belonging to cluster 1 as well as identifying the lowest number of clusters overall. Indeed, the number of trajectories utilized was selected to compare and contrast the efficacy of this technique with that of the single extended simulation. For all intents, the computational power required for the simulation tree was identical to that required for a single extended simulation with the principle advantage being the parallel processing of these simulations. In practice this tree technique can be applied with far fewer simulations. As can be seen (Fig. 6), given an appropriate seed structure as in the case of trajectories S0, S0A, and especially S0B, we were able to identify nearly the same number of clusters as that identified from the extended duration simulation but in one-tenth of the computational time. Both the simulation tree, consisting of 45 separate trajectories, and the single extended duration simulation, represent the same computational complexity but vary in their distribution. Thus, fewer simulations within the tree were necessary to adequately sample the conformational space offered by a longer duration simulation. This became particularly evident when trajectories were combined.

To examine the computational savings which can be achieved using this tree-based algorithm, we investigated extracts of the extended duration trajectory and determined that, on average, 60 ns was required to describe 22 clusters. In contrast, by pairing 10-ns trajectories derived from the most disparate trajectory seeds (S0 and S5), we were able to describe 25 distinct clusters. Not only does this demonstrate a reduction in computational cost, but a more efficient sampling of conformational space. In comparison, our approach is 2–3× more efficient than the 60-ns fragments, and an order-of-magnitude more efficient than the extended duration simulation. Of note is the origin of the trajectory seeds in the paired trajectory, which were selected from the two most disparate conformations, suggesting that the seed trajectory is critical to the ability to fully sample the conformational space. Specifically, the more deviated the conformations obtained from the seed, the greater the sampling of conformational space by subsequent trajectories. Hence, the seed trajectory length needs to be sufficient to provide an accurate description of the most common clusters as well as sampling conformations on the periphery. Furthermore, by selecting conformations which introduce the greatest amount of bias for subsequent branches, a more complete description of conformational space, than even that offered by extended simulations, can be obtained.

In summary, the presence of water around the tripeptide

feG seems to break associations based solely on point charges, hence, the head-to-tail arrangement present in the in vacuo simulations is no longer present. The predominant interaction appears to be the aromatic ring remaining in close proximity to the position 2 carbonyl, coupled with an interaction with one of the available carboxyl groups. It would be premature to suggest that the solvated conformations are “better” than the previously described in vacuo structures. There is currently no definitive information on the identity of a receptor for this class of peptides. Hence, correlations between biological activity and the solvated conformations we have described for feG await confirmation through the simulation of related peptides as previously described (Metwally et al., 2002). We have also presented an alternative method to prohibitively long simulations of peptides. To adequately sample the conformational space for a given peptide, it seems that by actively biasing the simulations by seeding with highly deviated structures that a sampling of conformational space similar to that obtained through extended duration simulations is achieved. This remains true provided that biasing is done in a manner which allows for conservation of the equilibrium of conformations. This was accomplished by simulating structures which tended to represent the global average alongside those conformations which were most and least deviated as compared to the seed structure. The merit of a single simulation cannot be discounted, as it offers a means of understanding continuous behavior of the peptide especially when examining the approach and docking of a peptide to its receptor (Burgi et al., 2001; Koerber et al., 2000). However, in examining only conformation, since it is often difficult to sample every member of a population, a statistical cross-section of that same population often suffices to make educated conclusions as to the nature of conformation and its determinants. This is, in principle, what all ensemble sampling techniques strive to accomplish. In contrast to LES, REM, and other techniques, the tree simulation represents a simple approach to a complex problem which is software- and hardware-independent. At a minimum, this simple approach was able to outperform the extended duration sampling approach by a very large margin. Furthermore, as the technique is not constrained to any particular configuration, it may be customized to suit any researcher's interests.

We thank David Cramb and Peter Tieleman for their helpful comments, suggestions, and invaluable discussions.

Support for this work was provided by the Canadian Institutes for Health Research.

REFERENCES

- Angeletti, L. R., U. Agrimi, C. Curia, D. French, and R. Mariani-Constantini. 1992. Healing rituals and sacred serpents. *Lancet*. 340:223–225.
- Balsera, M. A., W. Wriggers, Y. Oono, and K. Schulten. 1996. Principal component analysis and long time protein dynamics. *J. Phys. Chem.* 100:2567–2572.
- Berendsen, H. J. C., D. van der Spoel, and R. van Drunen. 1995. GROMACS: a message-passing parallel molecular dynamics implementation. *Comp. Phys. Comm.* 91:43–56.
- Burgi, R., X. Daura, A. Mark, M. Bellanda, S. Mammi, E. Peggion, and W. van Gunsteren. 2001. Folding study of an Aib-rich peptide in DMSO by molecular dynamics simulations. *J. Pept. Res.* 57:107–118.
- Clarage, J. B., T. Romo, B. K. Andrews, B. M. Pettitt, and G. N. Phillips, Jr. 1995. A sampling problem in molecular dynamics simulations of macromolecules. *Proc. Natl. Acad. Sci. USA*. 92:3288–3292.
- Daura, X., K. Gademann, B. Jaun, D. Seebach, W. F. van Gunsteren, and A. E. Mark. 1999. Peptide folding: when simulation meets experiment. *Angew. Chem. Int. Ed. Engl.* 38:236–240.
- de Groot, B. L., X. Daura, A. E. Mark, and H. Grubmüller. 2001. Essential dynamics of reversible peptide folding: memory-free conformational dynamics governed by internal hydrogen bonds. *J. Mol. Biol.* 309:299–313.
- de Groot, B. L., D. M. F. Van Aalten, A. Amadei, and H. J. C. Berendsen. 1996. The consistency of large concerted motions in proteins in molecular dynamics simulations. *Biophys. J.* 71:1707–1713.
- Dougherty, D. A. 1996. Cation- π interactions in chemistry and biology. A new view of benzene, Phe, Tyr, and Trp. *Science*. 271:163–168.
- Elber, R., and M. Karplus. 1990. Enhanced sampling in molecular dynamics: use of time-dependent Hartree approximation for a simulation of carbon-monoxide diffusion through myoglobin. *J. Am. Chem. Soc.* 112:9161–9175.
- Fu, Y. K., S. Arkins, B. S. Wang, and K. W. Kelly. 1991. A novel role of growth hormone, and insulin-like growth factor I. Priming neutrophils for superoxide anion secretion. *J. Immunol.* 146:1602–1608.
- Gao, X., and T. C. Wong. 2001. Molecular dynamics simulation of adrenocorticotropin (1–10) peptide in a solvated dodecylphosphocholine micelle. *Biopolymers*. 58:643–659.
- Hukushima, K., and K. Nemoto. 1996. Exchange Monte Carlo method and application to spin glass simulations. *J. Phys. Soc. Jpn.* 65:1604–1608.
- Kim, C. Y., P. P. Chandra, A. Jain, and D. W. Christianson. 2001. Fluoroaromatic-fluoroaromatic interactions between inhibitors bound in the crystal lattice of human carbonic anhydrase II. *J. Am. Chem. Soc.* 123:9620–9627.
- Koerber, S. C., J. Rizo, R. S. Struthers, and J. E. Rivier. 2000. Consensus bioactive conformation of cyclic GnRH antagonists defined by NMR and molecular modeling. *J. Med. Chem.* 43:819–828.
- Li, L., Z. Yu, R. Piacik, D. P. Hetzel, R. M. Rourk, Z. Namiot, J. Sarosiek, and R. W. McCallum. 1993. Effect of esophageal intraluminal mechanical and chemical stressors on salivary epidermal growth factor in humans. *Am. J. Gastroenterol.* 88:1749–1755.
- Lindahl, E., B. Hess, and D. van der Spoel. 2001. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Mod.* 7:306–317.
- Mark, P., and L. Nilsson. 2002. Structure and dynamics of liquid water with different long-range interaction truncation and temperature control methods in molecular dynamics simulations. *J. Comput. Chem.* 23:1211–1219.
- Mathison, R., D. Befus, and J. S. Davison. 1993. Removal of the submandibular glands increases the acute hypotensive response to endotoxin. *Circ. Shock*. 39:52–58.
- Mathison, R., D. Befus, and J. S. Davison. 1997a. A novel submandibular gland peptide protects against endotoxic, and anaphylactic shock. *Am. J. Physiol.* 273:R1017–R1023.
- Mathison, R., J. S. Davison, and G. Moore. 1997b. Submandibular gland peptide-T (Peptide-T): modulation of endotoxic and anaphylactic shock. *Drug Dev. Res.* 42:164–171.
- Mathison, R., P. Lo, D. Tan, B. Scott, and J. S. Davison. 2001a. The tripeptide feG reduces endotoxin-provoked perturbation of intestinal motility and inflammation. *Neurogastroenterol. Motil.* 13:599–603.

- Mathison, R., R. Woodman, and J. S. Davison. 2001b. Regulation of leukocyte trafficking and function in the heart by the tripeptide feG. *Can. J. Physiol. Pharmacol.* 79:785–792.
- Mathison, R. D., P. Lo, J. S. Davison, B. Scott, and G. Moore. 1998. Attenuation of intestinal and cardiovascular anaphylaxis by the salivary gland tripeptide FEG and its D-isomeric analog feG. *Peptides*. 19:1037–1042.
- McCammon, J. A., and S. C. Harvey. 1987. Dynamics of Proteins and Nucleic Acids. Cambridge University Press, Cambridge, UK.
- Metwally, E., J. M. Pires, G. J. Moore, D. A. Befus, J. S. Davison, and R. Mathison. 2002. Biological activities of submandibular gland tripeptide FEG and its analogues as a key to three-dimensional structure determination. *Peptides*. 23:193–199.
- Moore, G. J. 1994. Designing peptide mimetics. *Trends Pharmacol. Sci.* 15:124–129.
- Moore, G. J., J. R. Smith, B. Baylis, and J. M. Matsoukas. 1995. Design and pharmacology of peptide mimetics. *Adv. Pharmacol.* 33:91–141.
- Norberg, J., and L. Nilsson. 2000. On the truncation of long-range electrostatic interactions in DNA. *Biophys. J.* 79:1537–1553.
- Rhee, Y. M., and V. S. Pande. 2003. Multiplexed-replica exchange molecular dynamics method for protein folding simulation. *Biophys. J.* 84:775–786.
- Roitberg, A., and R. Elber. 1991. Modeling side chains in peptides and proteins: application of the locally enhanced sampling and the simulated annealing methods to find minimum energy conformations. *J. Chem. Phys.* 95:9277–9287.
- Schreiber, H., and O. Steinhauser. 1992. Cutoff does strongly influence molecular dynamics results on solvated polypeptides. *Biochemistry*. 31: 5856–5860.
- Smith, P. E., and B. M. Pettitt. 1991. Peptides in ionic solutions: a comparison of the Ewald and switching function techniques. *J. Chem. Phys.* 95:8430–8441.
- Straub, J., and M. Karplus. 1991. Energy equipartitioning in the classical time-dependent Hartree approximation. *J. Chem. Phys.* 94:6737–6739.
- Sugita, Y., A. Kitao, and Y. Okamoto. 2000. Multidimensional replica-exchange method for free-energy calculations. *J. Chem. Phys.* 113:6042–6051.
- Tai, K., T. Shen, U. Borjesson, M. Philippopoulos, and J. A. McCammon. 2001. Analysis of a 10-ns molecular dynamics simulation of mouse acetylcholinesterase. *Biophys. J.* 81:715–724.
- van der Spoel, D., A. R. van Buuren, E. Apol, P. J. Meulenhoff, D. P. Tieleman, A. L. T. M. Sijbers, B. Hess, K. A. Feenstra, E. Lindahl, R. van Drunen, and H. J. C. Berendsen. 2001. GROMACS User Manual, ver. 3.0. <http://www.gromacs.org>.
- Yoshitake, Y., H. Nakagawa, M. Eto, and K. Harano. 2000. An x-ray crystallographic and computational study of intermolecular edge-to-face aromatic interaction in crystal structure of Exo [4+2]p cycloadduct of phencyclone and S-allyl S-methyl dithiocarbonate. *Tetrahedron*. 56: 6015–6021.