Refinement of the NMR Solution Structure of the γ-Carboxyglutamic Acid Domain of Coagulation Factor IX Using Molecular Dynamics Simulation with Initial Ca^{2+} Positions Determined by a Genetic Algorithm†

Leping Li,‡ Thomas A. Darden,‡ Steven J. Freedman,§,|| Barbara C. Furie,§,|| Bruce Furie,§,|| James D. Baleja,§ Howard Smith,‡ Richard G. Hiskey,⊥ and Lee G. Pedersen*,‡,⊥

National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 27709, Department of Chemistry, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-3290, and Center for Hemostasis and Thrombosis Research, Division of Hematology-Oncology, New England Medical Center, and Department of Medicine and Biochemistry, Tufts University School of Medicine and Sackler School of Biomedical Sciences, Boston, Massachusetts 02111

Received September 6, 1996; Revised Manuscript Received December 18, 1996⊗

ABSTRACT: A genetic algorithm (GA) successfully identified the calcium positions in the crystal structure of bovine prothrombin fragment 1 bound with calcium ions (bf1/Ca). The same protocol was then used to determine the calcium positions in a closely related fragment, the Gla domain of coagulation factor IX, the structure of which had previously been determined by NMR spectroscopy in the presence of calcium ions. The most frequently occurring low-energy structure found by GA was used as the starting structure for a molecular dynamics refinement. The molecular dynamics simulation was performed using explicit water and the Particle-Mesh Ewald method to accommodate the long-range electrostatic forces. While the overall conformation of the NMR structure was preserved, significant refinement is apparent when comparing the simulation average structure with its NMR precursor in terms of the N-terminal (Tyr1-N) network, the total number of hydrogen bonds, the calcium ion coordinations, and the compactness of the structure. It is likely that the placement of calcium ions in the protein is critical for refinement. The calcium ions apparently induce structural changes during the course of the simulation that result in a more compact structure.

Human coagulation factor IX is a vitamin K-dependent glycoprotein that plays a key role in the blood coagulation cascade and contains a γ-carboxyglutamic acid (Gla[1])-rich domain (12 Glas), two EGF-like domains, and a serine protease domain (Davie et al., 1991; Furie & Furie, 1994). Like other vitamin K-dependent coagulation proteins, the Gla domain of factor IX binds calcium ions and interacts with a membrane surface in a step essential for activity. The Gla domain is also observed to be involved in protein−protein interactions (Banner et al., 1996). Two X-ray crystallographic structures containing a Gla domain have been determined: bovine prothrombin fragment 1 (bf1/Ca) (Soriano-Garcia et al., 1992) and the factor VII/tissue factor complex (factor VII/TF) (Banner et al., 1996). These structures have provided key information for understanding the structural and functional roles of the Gla domain in the coagulation process.

Recently, the solution structures of the calcium-bound, the magnesium-bound, and metal-free forms of the Gla domains of factor IX determined by NMR spectroscopy were reported (Freedman et al., 1995a,b, 1996). The metal-free Gla domain is largely disordered and lacks features observed in the crystal structure of bf1/Ca (Freedman et al., 1995b). The Gla domains of calcium-free factor X and prothromin apparently exhibit similar disorder (Park & Tulinsky, 1986; Sunnerhagen et al., 1995). In the presence of magnesium, the factor IX Gla domain is ordered from residues 12 to 47, but the N terminus (residues 1−11) is disordered (Freedman et al., 1996). These comparisons provide insight into the nature of calcium-induced structural transition and the structural requirement of the Gla domains for membrane binding.

Genetic algorithms (GAs) provide general purpose, stochastic search, and optimization methods. They were first described in the 1970s by John Holland, who extended Darwinian models of natural selection and evolution (Holland, 1975). These methods involve the evolutionary creation of a new population from an earlier generation through certain combinations and changes. The attributes of the survivors are passed to the offspring either with or without the combination of the genetic material of the parents. GAs search a population of chromosomes, a set of solution candidates. Usually, the chromosomes are represented by a set of strings, either binary or nonbinary, and constitute the building blocks of the solutions. Random combinations and/or changes of the parents in the building blocks produce offspring. The survival of the chromosomes (parents) is determined by the closeness of their properties to those desired. The better the fitness, the larger the chance

† This work was supported in part by grants from the NIH: HL27995 (L.G.P.), HL06350 (R.G.H.), and HL42443 (B.F.).

* Author to whom correspondence should be addressed.

‡ National Institute of Environmental Health Sciences.

§ Department of Medicine and Biochemistry, Tufts University School of Medicine and Sackler School of Biomedical Sciences.

|| New England Medical Center, Tufts University School of Medicine and Sackler School of Biomedical Sciences.

⊥ University of North Carolina at Chapel Hill.

⊗ Abstract published in *Advance ACS Abstracts,* February 1, 1997.

[1] Abbreviations: GA, genetic algorithm; MD, molecular dynamics; bf1/Ca, crystal structure of bovine prothromin fragment 1 (residues 1−145); Gla, γ-carboxyglutamic acid; NMR-fix/Ca (1−47), NMR structure of calcium-bound human factor IX (residues 1−47); MD-fix/Ca (1−47), average MD structure (range of 650−700 ps); rmsd, root mean square deviation; saa, solvent-accessible area.

S0006-2960(96)02250-7 CCC: $14.00

of being selected and passed to the next generation. Two major operators of GAs are mutation and crossover. Both operations guide the search to the solution. Although GA operations may prevent the search from being trapped in local minima in a multidimensional search space, GAs do not guarantee the convergence to a global minimum.

GAs have been applied to a variety of optimization and search problems such as protein structure prediction [see a recent review (Pedersen & Moult, 1996) and references cited therein], energy minimization (Herrmann & Suhai, 1995), docking (Oshiro et al., 1995), computer-aided molecule and drug design (Glen & Payne, 1995; Jones et al., 1995; Westhead et al., 1995), feature selection for pattern recognition (Siedlecki & Sklansky, 1989; Leardi et al., 1992), and combinatorial library design and synthesis (Sheridan & Kearsley, 1995; Weber et al., 1995; Singh et al., 1996).

The Gla domain of factor IX can be said to bind nine calcium ions (Li et al., 1996) with high- and low-affinity calcium binding sites. These calcium binding sites are not defined by proton−proton NMR spectroscopy. Since calcium ion coordinations in the crystal structure of a closely related protein, bf1/Ca (Soriano-Garcia et al., 1992), are provided almost exclusively by the side chain oxygen atoms of Gla residues, the midpoint of the line connecting any two such oxygen atoms may be regarded as a potential calcium binding site. Therefore, localizing the calcium positions in the Gla domain becomes a search problem, that is, how to select the best nine calcium positions from a pool of such centers that may have billions of combinations. Since GAs have been proved to be proficient in searching for noisy and discontinuous solution spaces, searching the calcium binding sites is a well-suited case for a GA.

In this report, we describe the use of a genetic algorithm for the identification of calcium binding sites in the NMR solution structure of the Gla domain of factor IX determined in the presence of calcium ions and a molecular dynamics simulation of this structure.

MATERIALS AND METHODS

bf1/Ca: A Control Study. The X-ray crystal structure of bf1/Ca has seven well-defined calcium ions. It provides an excellent case for demonstrating that the crystallographically defined calcium ions can be reproduced by a GA. First, the midpoints of the lines between any two side chain oxygen atoms of Gla residues (within 5.0 Å) were determined. A total of 124 midpoints were found. We expect the crystallographically defined calcium positions in bf1/Ca to be included in the 124 sites. A GA was used to search for the seven calcium positions from a pool of 124 sites, which has approximately 7.5×10^{10} combinations.

Coding. The calcium positions in each structure were represented by a nonbinary string that consists of seven randomly selected nonidentical sites from the 124 midpoints. Each site corresponds to the Cartesian coordinates of the position. No interactions between calcium ions closer than 1.5 Å were allowed. Each string represents a candidate solution. A set of strings makes up a population. For example, a string may be {2 23 45 9 7 67 110} which means that the midpoints 2, 23, 45, 9, 7, 67, and 110 from the possible 124 midpoints in the protein are selected as the initial positions for the seven calcium ions.

Fitness Function. For each string, the Cartesian coordinates of the midpoints selected were appended to an AMBER coordinates file that was produced from standard AMBER preparation files to form a complete structure. Since the calcium positions in each structure were initially computed as the midpoints of the line between two Gla side chain oxygen atoms, they may be energetically unfavored. Thus, each structure was subjected to a 500-step steepest-descent energy minimization with a distance-dependent dielectric function. No water molecules were included. During each energy minimization, only the calcium ions were allowed to move. Minimization of each structure took only a few CPU seconds on a SGI Power Challenge workstation. Most structures converged during the 500 steps of minimization. For those that did not, an energy of 0 kcal/mol was arbitrarily assigned.

For each generation, the structures were then sorted in an increasing order according to their molecular mechanics energies. The energies were then converted to floating point numbers in the range of {0,1} using the following equation:

$$\mathrm{ene}_i = 1.0 - \frac{\mathrm{energy}_i - \min}{\max - \min}$$

where energy_i denotes the energy of the *i*th structure, min and max are the lowest and highest energies of the population, respectively, and ene_i is the energy in the range of {0,1} for the *i*th structure. Finally, the ene_i was normalized and used for parent selection.

Selection. Parent selection comprised three steps. Only structures selected in steps 2 and 3 become candidates for subsequent mutation.

(1) The best structures (e.g. the top 10% on the basis of the above equation) found in the previous generation were passed to the current generation. These structures were not allowed to be changed in the current generation, e.g. mutations. However, they were updated in the subsequent generation. This guarantees that the best structures found are saved and passed to the next generation.

(2) A certain percentage (40%) of the current population was selected from the whole population of the previous generation, using the roulette wheel selection method, a procedure similar to spinning a roulette wheel with each member of the population having a portion of area of the wheel that is proportional to its fitness (normalized AMBER energy based on the above equation in this case). A total area of 1000 of the roulette wheel was arbitrarily used. Using roulette wheel selection, low-energy structures have high probabilities of being selected, whereas those having high energies, have a low, but not zero, probability of being selected. Since structures selected in this step are candidates for mutation, inclusion of the best structures found in the previous generation for selection provides opportunity for those structures to evolve into even better structures in the current generation.

(3) The remainder of the current generation (50%) was chosen by random selection of the whole population of the previous generation. In other words, every structure in the population was given an equal chance of being selected. This guarantees that the current generation contains some high-energy structures. Structures selected in this step are also candidates for mutations. Inclusion of high-energy structures in each generation helps prevent the search from becoming trapped in local minima.

Table 1: Parameters Used for the Genetic Algorithm

parameter	value
no. of generations	150
no. of structures in the generation	50
mutation rate	0.4[a]

[a] 0.2 for generations 50−100 and 0.1 for generations 100−150.

Mutation. A mutation operation was then applied randomly to the last 90% of the population with mutation rates listed in Table 1. For bf1/Ca, each string contains seven calcium ion positions. All of the calcium ion positions are candidates for mutation. In order to prevent good structures found from being changed too dramatically and meanwhile provide opportunities for them to leave local minima, a technique suggested by Nayeem et al. (1991) in protein structure prediction studies was adopted. The number of calcium ions to be mutated was given a probability of $1/2^k$, where k is the number of calcium ions. This means that the probability of allowing one calcium ion to be replaced is 0.5, two is 0.25, three is 0.125, and so on. Once the number of calcium ion positions to be changed is known, those to be mutated are determined randomly. A given calcium ion in a string was not allowed to be selected more than once. Next, the Cartesian coordinates for each calcium selected were replaced with those from a randomly selected site from the 124 midpoints. After each replacement, a bump check (calcium−calcium distance of <1.5 Å) was performed to prevent the newly introduced calcium ion from being unrealistically close to any of the existing calcium ions. Finally, the whole population was subjected to AMBER steepest-descent energy minimization, sorting, and fitness evaluation. The updated population was used for the selection of next generation, and the search continued until the search reached arbitrary convergence. The following is an example of a single-point mutation.

before mutation {2 ***23*** 45 9 7 67 110}

after mutation {2 ***99*** 45 9 7 67 110}

Crossover. Since the calcium ion positions in each string were randomly selected from a pool of 124 midpoints, they were not in a sequential order. Using a crossover operator to generate new strings would result in strings in which calcium positions may have duplicates. Thus, crossover was not used. A flow chart of the GA search is given in Figure 1.

Localizing Calcium Binding Sites on the Gla Domain of Factor IX. The same search protocol used to predict the crystal calcium binding sites for bf1/Ca was used to localize the calcium positions in the NMR solution structure of the Gla domain of factor IX. For factor IX, nine calcium binding sites instead of seven were searched. The total number of candidate calcium binding sites in this case was found to be 133. A GA was used to search for the best nine calcium positions from a pool of 133 sites.

Molecular Dynamics Simulations. The most frequently occurring low-energy structure found by GA was used as the starting structure for a molecular dynamics refinement. The AMBER 4.0 force field was used (Pearlman et al., 1995). The simulation was performed according to the procedures reported previously (Li et al., 1995, 1996). In brief, the protein was solvated in a large box of Monte Carlo TIP3P water with each side at least 12.5 Å from the nearest protein atom. The water and the calcium ions were energy minimized at a constant volume. Then, a molecular dynamics simulation was performed for 50 ps on the side chains of Gla residues, calcium ions, and water molecules only at constant pressure followed by energy minimization. The entire system was then energy minimized. No counterions other than the calcium ions were necessary since the total charge of the simulation system was zero. The simulation utilized the AMBER all-atom force field with a step size of 1 fs and was performed at 300 K after an initial stepwide heating over 3 ps. Nonbonded interactions were updated every 10 steps. The Particle-Mesh Ewald method was employed to accommodate the long-range electrostatic forces (Darden et al., 1993; Cheatham et al., 1995). All covalent bonds involving hydrogen atoms were constrained by a modified SHAKE algorithm (Ryckaert et al., 1977). A constraint of 5.0, 1.0, 0.5, and 0.1 kcal/mol Å was applied to the backbone atoms during the simulation periods of 0−50, 51−100, 101−150, and 151−200 ps, respectively. No other constraint was used. Finally, a complete unconstrained molecular dynamics simulation was performed for an ad-

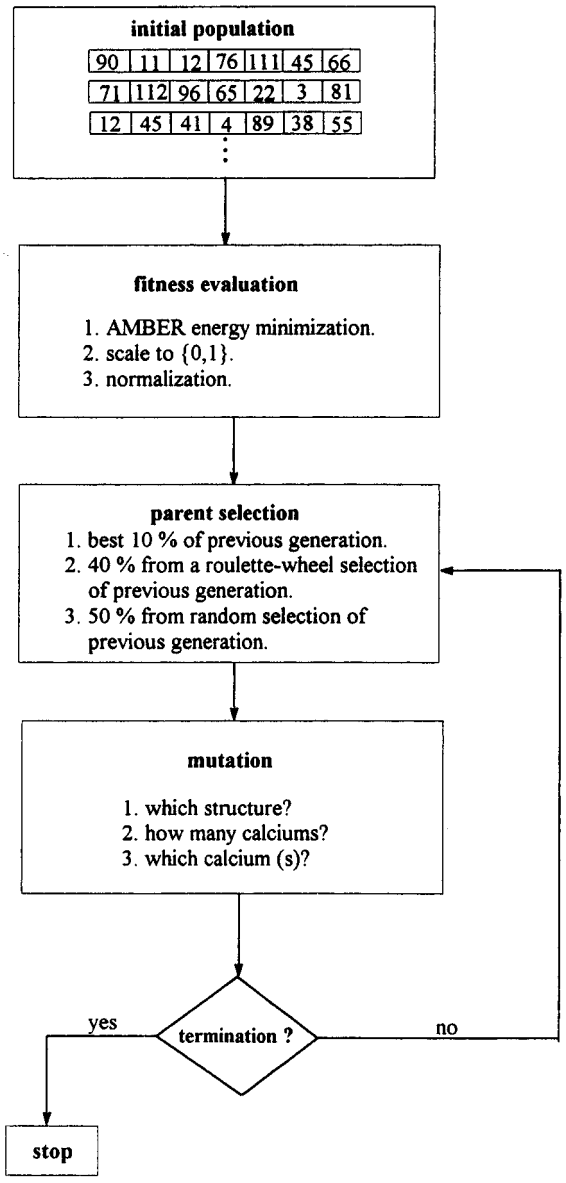

FIGURE 1: Flow chart of the GA search.

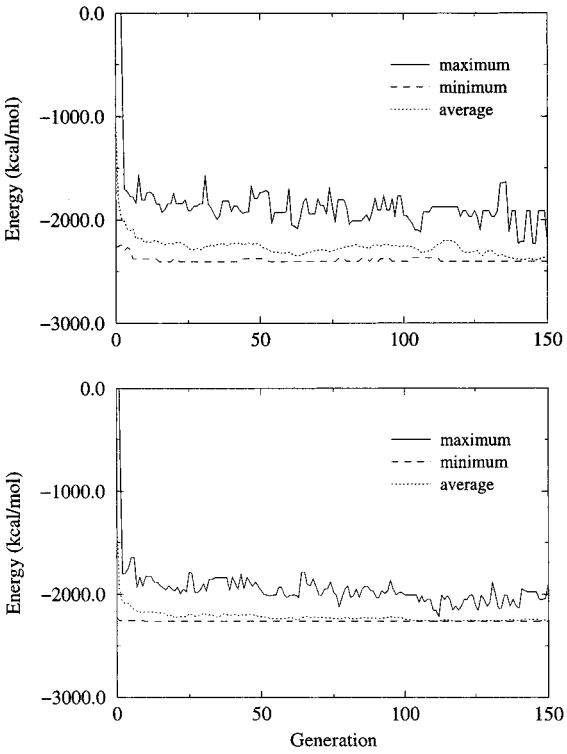

FIGURE 2: Evolution of energy for the test system bf1/Ca (1−47) (top) and NMR-fix/Ca (1−47) (bottom) as a function of generation. Energy was calculated using the AMBER 4.0 force field. Minimum, average (over all structures in the population), and maximum are characteristic energies of the population.

ditional 500 ps. The calculations were performed on a CRAY Y-MP supercomputer at the National Cancer Institute, at the North Carolina Supercomputing Center, and in parallel on a Silicon Graphics Power Challenge workstation.

RESULTS AND DISCUSSION

bf1/Ca: The Control Study. The parameters used in the search are listed in Table 1. Several GA runs with different random seed numbers, population size, and mutation rates were performed. In all cases, the GAs converged rapidly. For a typical run, the minimum, the average (over all structures), and the maximum energies of each population are shown in Figure 2. At generation 150, nearly all of the structures were converted to the same lowest-energy (−2403 kcal/mol) structures, as indicated by the closeness of the average energy to the minimal energy. Examination of the lowest-energy structures showed that Ca-1 to -6 matched exactly the crystallographically defined calcium positions in bf1/Ca, and Ca-7 occupied an essentially equivalent position. In both the crystal structure of bf1/Ca and the lowest-energy structures found by GA, Ca-7 has three non-water coordinations. In the crystal structure of bf1/Ca, the coordinations are OE4 of Glal5 and OE3 and OE4 of Gla20, while they are OE4 of Gla20 and OE3 and OE4 of Gla15 in the structures found by GA. The structures may be regarded as identical, since the Ca-7 positions in both structures are equivalent. Thus, GA successfully reproduced without bias the calcium positions in the crystal structure of bf1/Ca.

Placement of Calcium Ions on the Gla Domain of Factor IX. The search parameters are given in Table 1. As was the case for bf1/Ca, the search converged rapidly. The minimum, average, and maximum energies at each population are shown in Figure 3. After generation 100, the majority of the structures converted to the same lowest-energy structure (−2261 kcal/mol). Examination of the lowest-energy structures showed that approximately half were identical and the rest were diverse. Thus, the most frequently occurring lowest-energy structure was selected as a starting structure for additional molecular dynamics refinement. As a second control, the protocol was repeated for seven rather than nine calcium ions. The result was an unreasonable ion distribution that did not populate positions 1 and 6 defined in the crystal structure of bf1/Ca (Soriano-Garcia et al., 1992). It is interesting that seven ions charge neutralized the bf1/Ca structure whereas nine ions charge neutralized the fix/Ca structure.

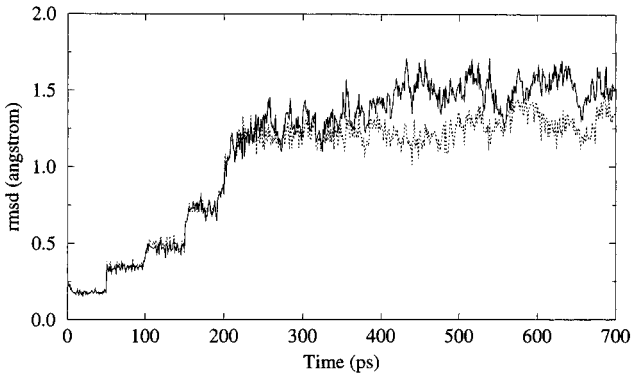

FIGURE 3: rms deviations of the backbone atoms of MD-fix/Ca (1−47) compared to those of NMR-fix/Ca (1−47). The solid line represents residues 1−47, and the dotted line represents residues 1−33. The last 500 ps are unrestrained.

Table 2: N-Terminal Tyr1-N Network for NMR-fix/Ca (1−47) and MD-fix/Ca (1−47)[a]

	donor-H···aceptor	distance (Å)	angle (deg)
NMR-fix/Ca (1−47)	Tyr1-N···Lys22-O	3.0	123
	Tyr1-N···Ca-4	7.2	
	Tyr1-N···Ca-5	4.1	
MD-fix/Ca (1−47)	Tyr1-N···Lys22-O	2.7	145
	Tyr1-N···Gla17-OE2	2.4	170
	Tyr1-N···Gla17-O	3.0	125
	Tyr1-N···Gla21-OE3	2.7	166
	Tyr1-N···Ca-4	4.2	
	Tyr1-N···Ca-5	4.3	

[a] Hydrogen bond criteria: donor-H···aceptor distance of ≤3.2 Å and donor-H···aceptor angle of ≥125°.

Molecular Dynamics Refinement. The simulation system reached equilibrium on the basis of the leveling with time of the root mean square deviations (rmsds) of the backbone atoms of the whole fragment (residues 1−47) and the Gla domain (residues 1−34) (Figure 3). An average structure was generated from the atomic coordinates of the last 50 ps of the simulation. A comparison of the NMR-derived solution structure of the Gla domain of factor IX [NMR-fix/Ca (1−47)] with the simulation average structure [MD-fix/Ca (1−47)] is then possible.

In NMR-fix/Ca (1−47), the N-terminal Tyr1-N was loosely attached to the core of the Gla domain with one weak hydrogen bond (Table 2). However, the bonding between the N terminus and the core of the Gla domain was strengthened in MD-fix/Ca (1−47) to include four hydrogen bonds (Table 2, Figure 4). A similar hydrogen bond network was observed in the crystal structures of bf1/Ca (Soriano-Garcia et al., 1992) and factor VII/TF (Banner et al., 1996).

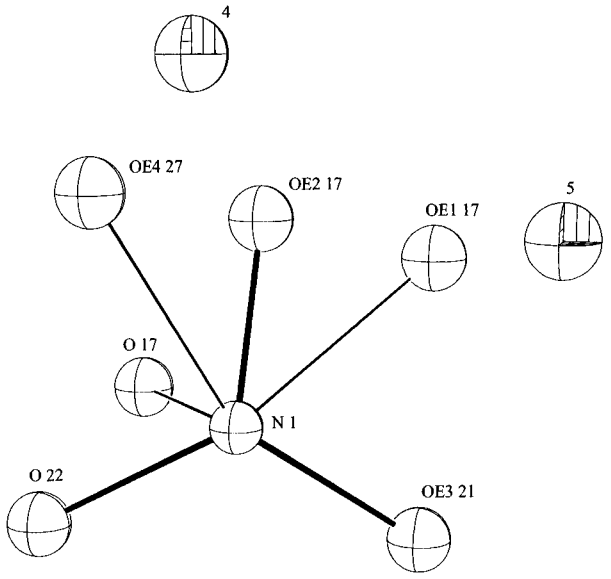

FIGURE 4: Oxygen−calcium ion network surrounding the N terminus in MD-fix/Ca (1−47). The six oxygens are less than 3.5 Å and the two calcium ions less than 4.4 Å from N-1, respectively.

These particular hydrogen bonds connect the N terminus to the short helix (Leu14−Gla17) and the conserved disulfide loop (Cys18−Cys23) and are apparently essential in maintaining an appropriate shape of the Ω-loop (residues 1−12) for membrane interactions (Cheung et al., 1992; Arni et al., 1994; Christiansen et al., 1995; Freedman et al., 1995; Li et al., 1995; Sunnerhagen et al., 1995). Chemical modifications of the N terminus of prothrombin, for instance, result in loss of activity (Welsch & Nelsestuen, 1988; Weber et al., 1992). A naturally occurring mutant that has deletion of Tyr1 in factor IX is found to be inactive (Ghanem et al., 1993). The two calcium ions (Ca-4 and -5) that are in close contact with the N terminus Ala1 in the crystal structure of bf1/Ca) (Soriano-Garcia et al., 1992) were also found near the N terminus (Tyr1-N) in MD-fix/Ca (1−47). In the NMR-fix/Ca (1−47), Ca-4 placed by GA is displaced away from the N-terminal Tyr1 (Table 2). Thus, MD restored (without restraints) essential features of the unique bf1/Ca N terminus network.

The atomic deviations of the α-carbon atoms of the MD-fix/Ca (1−47) from the crystal structure of bf1/Ca (1−47) and the NMR-fix/Ca (1−47) after optimal least-squares alignment of the backbone atoms of residues 1−47 and 13−47 are given in Figure 5, respectively. The largest deviations were found in the Ω-loop region for all cases.

The total number of non-water coordinations for the calcium ions in the NMR-fix/Ca (1−47) after the calcium ions placed by GA is 25. This number was increased significantly to 39 after the molecular dynamics simulation. The numbers of non-water coordinations for Ca-1 to -9 are 2, 3, 3, 3, 3, 2, 3, 3, and 3 in NMR-fix/Ca (1−47), whereas they are 2, 4, 8, 6, 6, 1, 4, 4, and 4 in MD-fix/Ca (1−47) (Table 3). The calcium ion coordination network in MD-fix/Ca (1−47) using the bf1/Ca ion labels is shown in Figure 6. As was the case for bf1/Ca, Ca-2 to -5 form a tight network with the Gla oxygen atoms. The major increase occurs for Ca-3 to -5. These calciums are also highly coordinated and buried inside the core of the Gla domain of bf1/Ca (Soriao-Garcia et al., 1992). They are important for maintaining the integrity of the domain for membrane

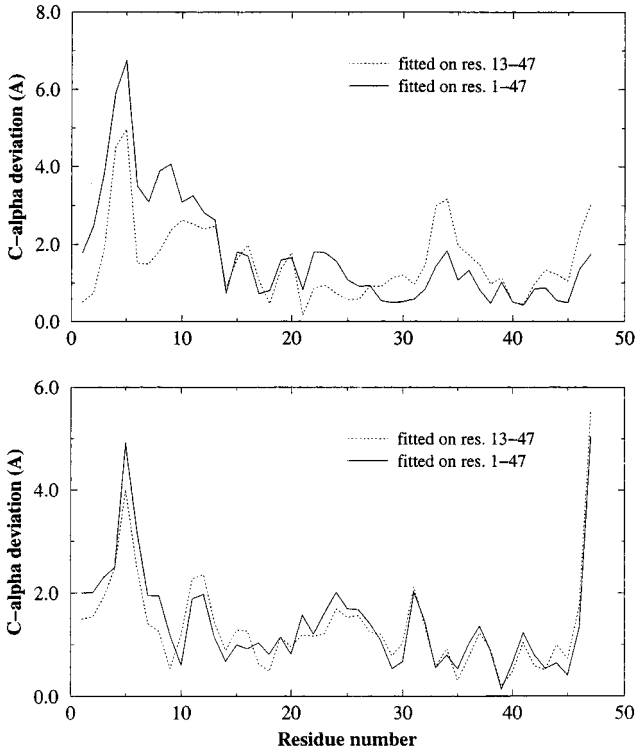

FIGURE 5: Atomic deviations of the C-α atoms of the MD-fix/Ca (1−47) from the crystal structure of bf1/Ca (1−47) (top) and the NMR-fix/Ca (1−47) (bottom) after optimal least-squares alignment of the backbone atoms of residues 1−47 and 13−47, respectively.

binding and possible protein−protein interaction(s). For those calciums exposed or partially exposed to solvent, Ca-1 retained the same number of coordinations while Ca-7 to -9 each gained one. Surprisingly, Ca-6 lost one coordination in MD-fix/Ca (1−47) compared to that in NMR-fix/Ca (1−47). Although the numbers of coordinations for Ca-1 to -9 differ in the two structures, the solvent-accessible areas (saas) for individual calcium ions remain similar in the two structures. In the NMR-fix/Ca (1−47), the saas are 59.4, 35.7, 3.0, 0.0, 15.1, 47.5, 30.0, 33.8, and 23.4 Å^2 and, for MD-fix/Ca (1−47), 58.2, 19.4, 3.5, 0.7, 12.2, 59.2, 29.3, 33.9, and 32.2 Å^2 for Ca-1 to -9, respectively. This suggests that the overall environments for the calciums in both structures are similar. However, the orientations of some Gla side chains and the distances between calcium ions and Gla side chain oxygen atoms in NMR-fix/Ca (1−47) may not allow optimal interactions among calciums and Gla residues. Molecular dynamics simulation was able to fine-tune the structure for optimal interactions among the Gla residues and the calcium ions.

There are 14 hydrogen bonds in NMR-fix/Ca (1−47). However, the number of hydrogen bonds increased to 45 in MD-fix/Ca (1−47), using criteria shown in the footnote of Table 2, suggesting that the MD-fix/Ca (1−47) is more compact than its precursor. This is supported by the total solvent-accessible areas of the two structures. The total solvent-accesible areas for NMR-fix/Ca (1−47) and MD-fix/Ca (1-47) were 4022 and 3616 Å^2, respectively, indicating an approximate 10% decrease in the MD-fix/Ca (1−47) saa with respect to that of NMR-fix/Ca (1−47).

Despite the significant refinements, the overall conformation of the two structures remained similar as indicated by the low rmsd of the backbone atoms of the two structures. The rmsds of the backbone atoms of the whole fragment

Table 3: Comparison of the Calcium Ion Coordinations for NMR-fix/Ca (1−47) and MD-fix/Ca (1−47)[a]

	NMR-fix/Ca (1−47)	MD-fix/Ca (1−47)
Ca-1	Gla26-OE1	Gla26-OE1
	Gla26-OE4	Gla26-OE4
Ca-2	Gla8-OE2	Gla8-OE2
	Gla30-OE3	Gla30-OE3
	Gla30-OE4	
		Gla8-OE4
		Gla30-OE2
Ca-3	Gla7-OE3	Gla7-OE3
	Gla8-OE1	Gla8-OE1
	Gla27-OE2	Gla27-OE2
		Asn2-OD1
		Gla8-OE2
		Gla27-OE1
		Gla30-OE3
		Gla30-OE4
Ca-4	Gla27-OE1	Gla27-OE1
	Gla27-OE4	Gla27-OE4
	Gla17-OE4	
		Gla7-OE3
		Gla7-OE4
		Gla17-OE2
		Gla17-OE3
Ca-5	Gla7-OE4	Gla7-OE4
	Gla17-OE2	
	Tyr1-O	Tyr1-O
		Ser3-OG
		Gla7-OE1
		Gla7-OE4
		Gla17-OE1
Ca-6	Gla21-OE1	Gla21-OE1
	Gla21-OE4	
Ca-7	Gla15-OE2	Gla15-OE2
	Gla20-OE1	Gla20-OE1
	Gla20-OE4	Gla20-OE4
		Gla15-OE4
Ca-8	Gla30-OE1	Gla30-OE1
	Gla30-OE2	Gla30-OE2
	Gla33-OE3	Gla33-OE3
		Gla33-OE4
Ca-9	Gla36-OE4	Gla36-OE4
	Gla40-OE3	Gla40-OE3
	Gla36-OE2	
		Gla36-OE1
		Gla40-OE1

[a] Only Ca···O distances of ≤3.0 Å are listed.

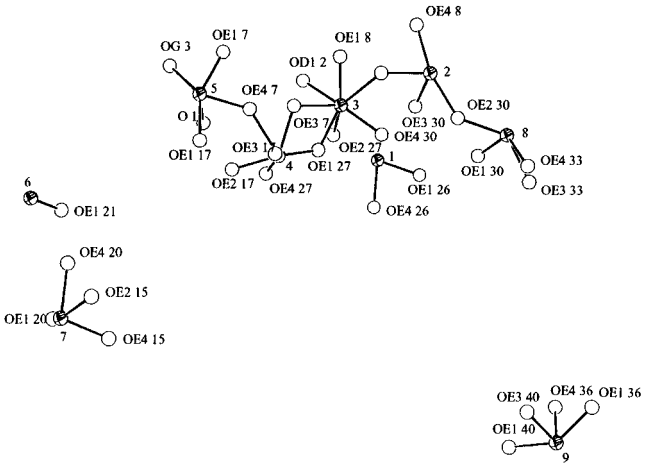

FIGURE 6: Calcium ion coordination network for MD-fix/Ca (1−47). All Ca−O distances less than 3.0 Å are shown.

(residues 1−47) and the Gla domain (residues 1−34) are 1.5 and 1.3 Å, respectively, compared to that of the NMR structure. The secondary structure characteristics of the two

Table 4: Secondary Structural Assignments for NMR-hfix/Ca (1−47) and MD-hfix/Ca (1−47)[a]

residue no.	residue name	NMR-hfix/Ca (1−47)	MD-hfix/Ca (1−47)
6	L	S	G
7	X	S	G
8	X	T	G
14	L	H	H
15	X	H	H
16	R	H	H
17	X	H	H
18	C	T	H
25	F	H	H
26	X	H	H
27	X	H	H
28	A	H	H
29	R	H	H
30	X	H	H
31	V	H	H
32	F	T	H
35	T	H	H
36	X	H	H
37	R	H	H
38	T	H	H
39	T	H	H
40	X	H	H
41	F	H	H
42	W	H	H
43	K	H	H
44	Q	H	H
45	Y	H	H
46	V		H

[a] The secondary structural analysis was done using DSSP (Kabsch & Sander, 1983). G is 3_{10}-helix, and H is α-helix.

structures are also similar. The helices identified in NMR-fix/Ca (1−47) (Leu14−Gla17, Phe25−Val31, and Thr35−Tyr45) are well-preserved in MD-fix/Ca (1−47), although the helices are extended to one additional residue in MD-fix/Ca (1−47) (Table 4). Interestingly, the secondary structures of the Ω-loop region (residues 1−12) in both structures are different.

These results suggest that the regions that are not well-defined in NMR-fix/Ca (1−47) become ordered after the molecular dynamics simulation and the overall conformation of the structure is preserved. It is likely that the placement of initial calcium ions in the protein is critical in the whole process of refinement. The calcium ions may induce structural changes during the course of molecular dynamics simulation that result in a more compact structure.

In conclusion, GA successfully identified the crystallographically defined calcium positions in bf1/Ca. The algorithm was then used to place calcium ions in the Gla domain of factor IX determined by NMR spectroscopy. The calcium-bound structure was further refined by a molecular dynamics simulation with the Particle-Mesh Ewald method to accommodate long-range electrostatic forces. Although the overall conformation of the simulation average structure remains similar to the NMR structure after refinement, improvements are apparent in terms of the N-terminal hydrogen bond network, the total number of hydrogen bonds, the calcium ion coordinations, and the compactness of the structure.

ACKNOWLEDGMENT

We thank the National Cancer Institute and the North Carolina Supercomputing Center for computer time. We thank Peter White for his expertise.

REFERENCES

Arni, R. K., Padmanabhan, K. P., Wu, T. P., & Tulinsky, A. (1994) *Chem. Phys. Lipids 67/68,* 59−66.

Banner, D. W., D'Arcy, A., Chéne, C., Winkler, F. K., Guha, A., Konigsberg, W. H., Nemerson, Y., & Kirchhofer, D. (1996) *Nature 380,* 41−46.

Cheatham, T. E., III, Miller, J. L., Fox, T., Darden, T. A., & Kollman, P. A. (1995) *J. Am. Chem. Soc. 117,* 4193−4194.

Cheung, W., Hamaguchi, N., Smith, K. J., & Stafford, D. W. (1992) *J. Biol. Chem. 267,* 20529−20531.

Christiansen, W. T., Jalbert, L. R., Robertson, R. M., Jhingan, A., Prorok, M., & Castellino, F. J. (1995) *Biochemistry 34,* 10376−10382.

Darden, T. A., York, D. M., & Pedersen, L. G. (1993) *J. Chem. Phys. 98,* 10089−10092.

Davie, E. W., Fujikawa, K., & Kisiel, W. (1991) *Biochemistry 30,* 10363−10370.

Freedman, S. J., Furie, B. C., Furie, B., & Baleja, J. D. (1995a) *Biochemistry 34,* 12126−12137.

Freedman, S. J., Furie, B. C., Furie, B., & Baleja, J. D. (1995b) *J. Biol. Chem. 270,* 7980−7987.

Freedman, S. J., Blostein, M. D., Baleja, J. D., Jacobs, M., Furie, B. C., & Furie, B. (1996) *J. Biol. Chem. 271,* 16227−16236.

Furie, B., & Furie, B. C. (1994) *Cell 53,* 505−518.

Ghanem, N., Costes, B., Martin, J., Vidaud, M., Rothschild, C., Foyer-Gazengel, C., & Goossens, M. (1993) *Eur. J. Hum. Genet. 1,* 144−155.

Glen, R. C., & Payne, A. W. R. (1995) *J. Comput.-Aided Mol. Des. 9,* 181−202.

Herrmann, F., & Suhai, S. (1995) *J. Comput. Chem. 16,* 1434−1444.

Holland, J. H. (1995) *Adaptation in natural and artificial systems,* University of Michigan Press, Ann Arbor.

Jones, G., Willett, P., & Glen, R. C. (1995) *J. Comput.-Aided Mol. Des. 9,* 532−549.

Kabsch, W., & Sander, C. (1983) *Biopolymers 22,* 2577−2637.

Leardi, R., Boggia, R., & Terrile, M. (1992) *J. Chemom. 7,* 267−281.

Li, L., Darden, T., Foley, C., Hiskey, R., & Pedersen, L. (1995) *Protein Sci. 4,* 2341−2348.

Li, L., Darden, T., Hiskey, R., & Pedersen, L. (1996) *J. Phys. Chem. 100,* 2475−2479.

Nayeem, A., Vila, J., & Scheraga, H. A. (1991) *J. Comput. Chem. 12,* 594−605.

Oshiro, C. M., Kuntz, I. D., & Dixon, J. S. (1995) *J. Comput.-Aided Mol. Des. 9,* 113−130.

Park, C. H., & Tulinsky, A. (1986) *Biochemistry 25,* 3977−3982.

Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham, T. E., III, Ferguson, D. M., Seibel, G. L., Singh, U. C., Weiner, P. K., & Kollman, P. A. (1995) *AMBER 4.1,* University of California, San Francisco.

Pedersen, J. T., & Moult, J. (1996) *Curr. Opin. Struct. Biol. 6,* 227−231.

Ryckaert, J. P., Ciccotti, G., & Berendsen, H. C. (1977) *J. Comput. Phys. 23,* 327−341.

Sheridan, R. P., & Kearsley, S. K. (1995) *J. Chem. Inf. Comput. Sci. 35,* 310−320.

Siedlecki, W., & Sklansky, J. (1989) *Pattern Recognit. Lett. 10,* 335−347.

Singh, J., Ator, M. A., Jaeger, E. P., Allen, M. P., Whipple, D. A., Soloweij, J. E., Chowdhary, S., & Treasurywala, A. M. (1996) *J. Am. Chem. Soc. 118,* 1669−1676.

Soriano-Garcia, M., Padmanabhan, K., de Vos, A. M., & Tulinsky, A. (1992) *Biochemistry 31,* 2554−2566.

Sunnerhagen, M., Forsén, S., Hoffrén, A. M., Drakenberg, T., Teleman, O., & Stenflo, J. (1995) *Nat. Struct. Biol. 2,* 504−509.

Weber, D. J., Berkowitz, P., Panek, M. G., Huh, N.-W., Pedersen, L. G., & Hiskey, R. (1992) *J. Biol. Chem. 267,* 4564−4569.

Weber, L., Wallbaum, S., Broger, C., & Gubernator, K. (1995) *Angew. Chem., Int. Ed. Engl. 34,* 2280−2282.

Welsch, D. J., & Nelsestuen, G. L. (1988) *Biochemistry 27,* 4939−4945.

Westhead, D. R., Clark, D. E., Frenkel, D., Li, J., Murray, W., Robson, B., & Waszkowycz, B. (1995) *J. Comput.-Aided Mol. Des. 9,* 139−148.

BI962250R