Molecular Dynamics Study of the Structure and Dynamics of a Protein Molecule in a Crystalline Ionic Environment, *Streptomyces griseus* Protease A†

F. Avbelj,‡,§,∥ J. Moult,∥,⊥ D. H. Kitson,‡,# M. N. G. James,⊥ and A. T. Hagler*,‡,#

The Agouron Institute, 505 Coast Boulevard South, La Jolla, California 92037, Kemijski Institut Boris Kidric, Hajdrihova 19, YU-61001 Ljubljana, Yugoslavia, Medical Research Council of Canada Group in Protein Structure and Function, Medical Sciences Building, Biochemistry Department, University of Alberta, Edmonton, Alberta T6G 2H7, Canada, Center for Advanced Research in Biotechnology, University of Maryland, 9600 Gudelsky Drive, Rockville, Maryland 20850, and Biosym Technologies, Inc., 10065 Barnes Canyon Road, San Diego, California 92121

Received January 24, 1990

ABSTRACT: A large-scale molecular dynamics simulation of the behavior of a serine protease (*Streptomyces griseus* protease A) in a crystalline environment has been performed. All atoms (including hydrogens) of two protein molecules and the surrounding solvent of crystallization, consisting of both water and salt ions, were explicitly represented, and a relatively long range of interactions (up to 15 Å) were included. The simulation is the longest so far reported for a protein in such an environment (60 ps). The use of the full crystalline environment allows a direct comparison of the structure and dynamic properties of the protein and surrounding solvent to be made with the experimental X-ray structure. Here we report the comparison of the protein structures and analyze the energetics of the system, including interaction with the aqueous environment. Subsequent papers will deal with other aspects of the simulation. The overall root mean square differences between the time-averaged molecular dynamics structure and that from crystallography, for all well-ordered, non-hydrogen atoms, are 1.67 and 1.25 Å for the two molecules taken as the asymmetric unit. An extensive analysis of the conformation of substructural elements and individual residues and their deviation from experiment has revealed a strong influence of the ionic medium on their behavior. Implications of the results for free energy calculations and for future directions are also discussed.

Molecular mechanics and molecular dynamics techniques are being applied to an increasing variety of biomolecular studies and drug design problems (Struthers et al., 1984; Hagler, 1985; Mackay et al., 1989; Karplus & McCammon, 1981). Recently, attention has turned to applications of the technique that demand a high degree of fidelity in reproducing structural properties, such as the refinement of X-ray crystallographic (Brünger et al., 1987) and NMR (Clore et al., 1986) models of proteins, and the determination of differences in catalytic rates (Warshel & Sussman, 1986) and binding constants (Dauber et al., 1982; Bash et al., 1987) for substrates and inhibitors of enzymes.

The advent of genetic engineering techniques that make it practical to modify the structure and active sites of proteins by site-specific mutagenesis has added further impetus to the surge of theoretical simulations. The ability to simulate protein structure accurately (within experimental accuracy, or better) and to calculate thermodynamic quantities, such as free energies of substrate and inhibitor binding, to within a kilocalorie or less would imply the ultimate ability to use theory to design or modify protein structures from first principles in order to achieve a desired architecture and function and, analogously, to design specific functional ligands.

Clearly, however, we are not yet at this stage and have yet to realize the full benefits of this potential. A major objective of this paper is to address the issue of the accuracy of structure reproduction and the next steps to be taken to achieve these goals. This will be done by comparing the results of a molecular dynamics simulation of a protein in its full crystalline environment [protease A from *Streptomyces griseus* (SGPA)[1]] with the corresponding X-ray crystal structure.

There are several possible reasons for our inability to predict confidently and precisely the results of protein engineering experiments, and they in turn suggest the direction of future research. They include (1) the omission from the simulation of the solvent, counterions, and other components constituting the actual environment of the protein; (2) the use of inadequate or incomplete potential functions that do not accurately reflect the system's inter- and intramolecular forces (errors in van der Waals parameters or partial charges would be examples of inaccurate parameters while the omission of polarization effects or charge transfer are two examples where potential functions commonly used for biological systems are incomplete; (3) inadequate construction of the system or simulation technique (for example, omission of structural waters or ions in the protein interior can lead to large structural perturbations; omission of relaxation by minimization and inadequate equilibration before carrying out a dynamics trajectory can lead to large initial artifactual forces and concomitant major

† This work was supported by Grants GM 30793 and GM 30564 from the National Institutes of Health. Grants of computer time were provided by the following organizations: the National Science Foundation [Grant PCM 8421273, on the Cray X-MP/48 supercomputers at the San Diego Supercomputer Center (SDSC) and the National Center for Supercomputing Applications (NCSA), University of Illinois at Urbana-Champaign]; NCSA, on the Cray X-MP/48 at NCSA; and SDSC, on the Cray X-MP/48 at SDSC.

‡ The Agouron Institute.
§ Kemijski Institut Boris Kidric.
∥ Center for Advanced Research in Biotechnology, University of Maryland.
⊥ Biochemistry Department, University of Alberta.
Biosym Technologies, Inc.

[1] Abbreviations: SGPA, *Streptomyces griseus* protease A; BPTI, bovine pancreatic trypsin inhibitor; RMS, root mean square; VFF, valence force field; SPC, simple rigid three-point charge model.

0006-2960/90/0429-8658$02.50/0

distortions); (4) inadequate computational resources (these can limit the length of a dynamics simulation to a degree where significant conformational transitions or regions of phase space are not observed; in addition, the lack of computational resources can lead to the use of approximations that affect the first three factors).

The first step in addressing these problems is characterizing their importance and effect. To this end, we have carried out the molecular dynamics simulation of SGPA. Care has been taken to represent the full environment of the protein molecules in the crystal that has been studied crystallographically. All water molecules and all counterions present in the experimental system are explicitly included in the simulation. Neighboring protein molecules generated by the appropriate crystallographic symmetry are also included. Nonbond interactions between atoms up to 15 Å apart are taken into account, substantially more than is customary. All hydrogen atoms in the system are included explicitly, thus avoiding the uncertainties of the "united atom" approximation (Dauber-Osguthorpe et al., 1988), which is usually used. This is the first such simulation with all hydrogen atoms included explicitly and with such a large cutoff. The simulation is 60 ps long. Since two protein molecules are simulated independently, this is, in some respects, equivalent to two independent 60-ps simulations of the protein molecule. These steps are necessary if the results are to be considered characteristic of deviations and problems that can arise in current dynamics simulations of proteins. In fact, the fit to the X-ray structure achieved in this work, although not yet satisfactory, is as good as, or better than, the fit achieved in previous simulations, which have usually been of simpler systems [e.g., bovine pancreatic trypsin inhibitor (BPTI)] and of shorter duration. Thus, we have the opportunity to assess results and to begin to plan the steps required for the iterative improvement of the techniques.

Although it is routine to compare the average protein structure resulting from a molecular dynamics simulation with the relevant X-ray structure, there have been few attempts to perform the simulation in a manner that should faithfully reproduce the crystal environment. Where a solution state is simulated or, more usually, solvent is ignored, differences may be attributed to environmental effects. In vacuo simulations have yielded root mean square (RMS) differences in coordinates to the corresponding X-ray structure in the range 2.8–1.9 Å (van Gunsteren & Karplus, 1981; Levitt, 1983a; Northrup et al., 1981; Ichiye et al., 1986). A crystal simulation of a small protein (BPTI) (Berendsen et al., 1986) gave differences of the order of 1.5 Å RMS (after 40 ps) and is the most accurate to date. The SGPA simulation reported here has similar RMS values to this latter study for a larger protein, and for a longer simulation, and thus deviations obtained are characteristic of at least as good a reproduction of an experimental system as is obtainable with current methods and force fields.

The major objective of this theoretical study, then, follows the philosophy of Lifson; i.e., "the deviations between the calculated and experimental properties contain more information and are far more interesting and significant than the agreement" (S. Lifson, private communication). In the analysis of the results of the theory applied to this system, therefore, our first step is to characterize the nature of the differences between the calculated and experimental results. As we shall describe below, both structural and energetic features arise in the theoretical model that appear to be at odds with the experiment. Having characterized the nature of the deviations of calculated properties from experiment in several systems that have been simulated for an adequate time (the simulation of two independent proteins in the asymmetric unit is particularly useful in order to assess the adequacy), we shall be sure that these deviations are indeed characteristic of the techniques and potential functions used. We shall then be in a position to understand the importance of additional rigor in the potential functions and technique and can improve them accordingly.

Finally, the results obtained here are used to probe the nature of the aqueous and counterion environment (F. Avbelj, J. Moult, D. H. Kitson, M. N. G. James, D. Hadzi, and A. T. Hagler, to be submitted). We have studied the dynamics, energetics, and structure of this environment as well as of the protein. In several cases, the theoretical results suggest questions as to the methodology used in deriving an "experimental" structure from the experimental diffraction pattern. In at least one case, the theory leads to a possible interpretation of a previously uninterpreted feature in the electron density map, which in turn supports the theoretical results. Thus, even at this level, the theory, *when used in intimate conjunction* with experimental observables, has served to help suggest interpretations for the experimental data and to allow us to pose questions for further theoretical and experimental studies.

In summary, this study has three main objectives (in this paper we focus primarily on the first objective): (1) to examine the fidelity of contemporary molecular dynamics algorithms and force fields in reproducing observed protein structures; (2) to extend our understanding of the X-ray structure of the protease A from *S. griseus* (SGPA) in terms of the role of ions, water molecules, and intermolecular interactions in stabilizing the molecular structure in the crystal; and (3) to explore the dynamics of this protein molecule.

Experimental Procedures

Quality of the SGPA Crystal Structure. If we are to judge the quality of molecular dynamics simulations by the agreement of the resulting average structure with that derived from X-ray, and characterize discrepancies for further analysis and interpretation, we must consider how reliable the X-ray-derived structures are.

SGPA contains 181 amino acid residues and is thus large enough to be considered representative of monomeric proteins in general but small enough for a sufficiently lengthy simulation to be affordable with contemporary computers. The X-ray crystal structure is of high quality, with a crystallographic R factor of 12.6% on the 19 594 reflections between 8- and 1.5-Å resolution with intensities greater than twice the standard deviation of measurement (Sielecki et al., 1979). The X-ray model includes 234 ordered solvent molecules, one of which is believed to be a sodium ion. As is customary, hydrogen atoms were not included in the main refinement procedure. To eliminate the possibility of small systematic errors from this cause, 11 more cycles were performed including all hydrogen atoms, giving an R factor of 12.1% (A. R. Sielecki and M. N. G. James, unpublished experiments). The resulting non-hydrogen coordinates form the set used in the current work.

The method of Cruickshank (Kasper & Lonsdale, 1967) indicates an overall coordinate error level of 0.12 Å RMS for the final structure, reflecting the low average apparent thermal motion represented by the B factor of 15.7 Å^2 for the experimental data. The Sigma (A) method (Read, 1986) gives a mean coordinate error of 0.19 Å. This relatively high value reflects the inclusion of a high proportion of weak data at high diffraction angles.

Because of the assumptions involved in the theoretical methods for estimating errors, it is important to obtain empirical checks on their validity. A convenient way of doing this is by comparing the agreement in coordinates of independently determined, but physically identical, crystal structures with the error indicated by theory of errors. A complication in making such comparisons (and in making comparisons of experimental structures with simulation results) is that, in all protein crystal structures, parts of the molecule, particularly some surface side chains, are poorly ordered or have multiple positions (Smith et al., 1986; Svennson et al., 1986) so that relatively large nonrandom coordinate differences may arise locally. Fortunately, such regions are characterized by high temperature factors in the refinement procedure. We have used such high temperature factors as a filter to eliminate the less reliable parts of the X-ray structures in making comparisons. Always, only a small part of a structure is ignored; for instance, 85% of the atoms in the SGPA structure have temperature factors less than 20 Å^2, the filter value used.

Several comparisons are available from the literature, including independent data collections and refinements of the structures of bovine trypsin, chymotrypsin, and ribonuclease A. [See also a similar comparison in Dauber-Osguthorpe et al. (1988).] The nature of the differences between each pair of equivalent structures has been discussed. In the case of trypsin (Chambers & Stroud, 1979), the Cruickshank formula indicates RMS coordinate errors of approximately 0.1 Å, while comparison of the two structures gives an RMS difference for most of the non-hydrogen protein atoms of 0.84 Å (12 residues that might be affected by the different bound inhibitors are not included). However, this latter value is dominated by differences in the poorly resolved parts of the structure, and omitting these leads to an RMS difference of 0.46 Å. The omitted portions correspond closely to those with highest temperature factors. Comparison of the two chymotrypsin structures (Tsukada & Blow, 1985; Blevins & Tulinsky, 1985) yields an all-atom RMS difference in coordinates of 0.75 Å, and 0.34 Å for the atoms with B factors less than 20 Å^2. For ribonuclease A (Wlodawer et al., 1986), the corresponding numbers are 0.80 and 0.36 Å, respectively.

All of these structures are of high resolution (2 Å or better) and well refined (R factors generally below 20%). We may conclude that when these conditions are satisfied, the RMS difference between parts of equivalent structures with temperature factors lower than 20 Å^2 will be of the order of 0.5 Å (Dauber-Osguthorpe et al., 1988). The diffraction data, from which the structure of each member of these pairs was derived, are the result of independent experiments, and were collected by a variety of methods, so that the effect of experimental errors is included.

This value suggests that current theoretical estimates of error levels in crystal coordinates should be approximately doubled to allow for the effect of the assumptions involved. We may conclude that, conservatively, a molecular dynamics structure should agree with a corresponding X-ray crystal structure with an RMS difference of better than 0.5 Å over the well-ordered parts of the structure (Dauber-Osguthorpe et al., 1988).

Construction of the SGPA Crystal System To Be Simulated. As noted in the introduction, the setup of the system to be simulated is an important and nontrivial step (Mackay et al., 1988). The SGPA crystal studied here is of space group $P4_2$, with one protein molecule in each of the four asymmetric units. The contents of two asymmetric units have been explicitly simulated, with the surroundings introduced using the full crystal symmetry. Two asymmetric units were used, rather than the minimum necessary of one, to avoid the inclusion of the 2-fold rotation axis, which would impose an unrealistic instantaneous configurational 2-fold symmetry on the solvent. Practically, this gives us important information as to the generality of the resulting dynamic and structural behavior and on convergence. The two asymmetric units together contain two molecules of SGPA, each consisting of 181 amino acid residues, and a quantity of solvent, amounting to 48% by weight. At the pH of the crystallographic studies (4.1) the protein has a net charge of +5. The crystal were grown from approximately 1 M sodium dihydrogen phosphate, which would imply 26 sodium and 26 dihydrogen phosphate ions in the solvent volume of the two asymmetric units, assuming the same ion concentration as in solution. Measurements of the salt content of lysozyme crystals (Hagler & Moult, 1978) and studies of the dependence of crystal density on salt concentration (Scanlon & Eisenberg, 1975) support this assumption. Given the net charge of +5 on the protein, this implies 26 dihydrogen phosphates and 16 sodiums to produce an electrically neutral system. Estimates of the volume remaining available to water molecules in the unit cell in the presence of this quantity of protein and salt ions, as well as estimates from the weight according to the density, give values of between 1396 and 1468 for the number of water molecules in the system. A system was therefore generated containing 2 protein molecules, 26 dihydrogen phosphate ions, 16 sodium ions, and 1429 water molecules, producing a neutral crystal with a total of 9427 atoms in the "asymmetric unit" that was used for the simulation. Initial positions for the ions were chosen to be as far from each other and from atoms of the protein as possible, i.e., so that they would be fully solvated by water rather than biasing ion association with other ions or with protein groups by the initial placement at the start of the simulations. [Significantly, this resulted in major reorientation of the ion distribution in the crystal in response to the field of the protein molecules (F. Avbelj, J. Moult, D. H. Kitson, M. N. G. James, D. Hadzi, and A. T. Hagler, to be submitted).] The water molecules were placed randomly by using a grid system (Hagler et al., 1980), an arrangement that gives a low initial correspondence with experimentally observed solvent positions (to avoid biasing the derived comparison with experiment). Waters within a protein crystal are located in a variety of different types of environment. Of the 1429 water molecules placed into the SGPA crystal system, approximately 30 are in the substrate binding site and 130 lie in narrow channels between protein molecules. Other waters fill large channels within the crystal. These waters can be further catagorized according to their proximity to ions, hydrophilic or hydrophobic protein residues, etc. The waters in the different environments have different properties from one another, and this, together with a fuller description of the water distribution, will be reported elsewhere (F. Avbelj, J. Moult, D. H. Kitson, M. N. G. James, D. Hadzi, and A. T. Hagler, to be submitted).

A Monte Carlo simulation in which the protein molecules were held fixed was used to relax the solvent. The coordinates after 6 million configurations were taken as the initial coordinates for the minimizations and molecular dynamics simulations described here. As mentioned above, approximately 130 water molecules are located in narrow channels between protein molecules (these were defined as waters that are in van der Waals contact with atoms from more than one protein). The question arises whether the mobility of these waters, or of waters in other constrained environments, could be hindered to the extent that the asymmetry present in the initial water distribution would be maintained throughout the du-

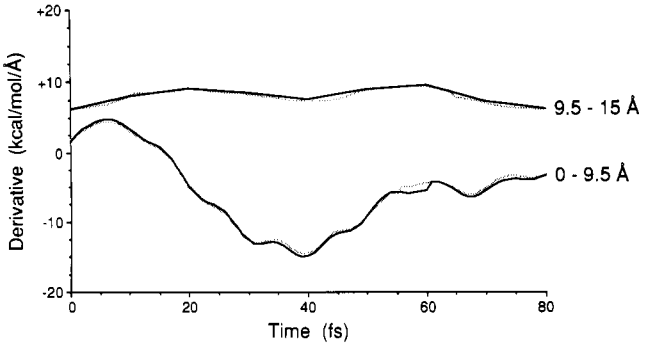

FIGURE 1: Effects of the twin-range approximation on the x component of the first derivative of the energy with respect to the Cartesian coordinate (the force) for a sodium ion in the SGPA system. The lower pair of lines represent the contribution to the derivative of interactions from 0 to 9.5 Å, while the top pair of lines represent the contribution from interactions from 9.5 to 15 Å. The shaded lines are for an 80-step simulation in which the interaction energies within both cutoffs are updated every step—thus, this represents the normal (non-twin-range) simulation. The solid lines are for a simulation in which the interactions for the 9.5–15-Å cutoff are calculated only once every 20 steps. These plots show that the derivatives for the longer range interactions change much more gradually than do those for the shorter range interactions and that the approximation in which the neighbor list for the long-range interactions is updated only once every 20 steps has only a negligible effect on the derivatives. These findings indicate that this method can be used as a means of decreasing the time required for the simulation and increasing the cutoff distance.

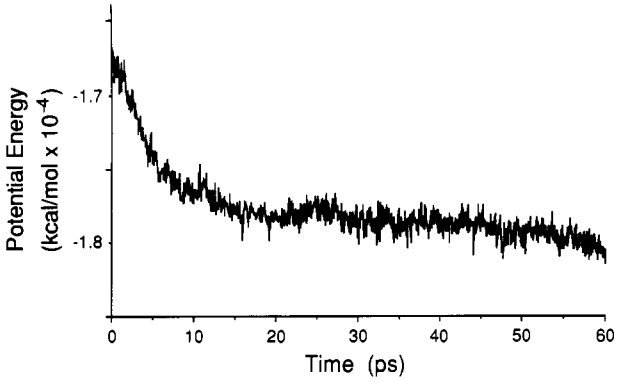

FIGURE 2: Potential energy of the SGPA system as a function of time during a 60-ps molecular dynamics simulation. As this figure shows, the energy initially drops rapidly and then begins to level out after ~16 ps. Average properties of the system were calculated over the 16–60-ps time period.

ration of the Monte Carlo and molecular dynamics simulations, giving rise to "permanently" different environments for the two proteins. Furthermore, if this was the case, could this asymmetry be a significant factor in the differences in the behavior of the two proteins that was observed during the dynamics? Calculation of the diffusion constants (F. Avbelj, J. Moult, D. H. Kitson, M. N. G. James, D. Hadzi, and A. T. Hagler, to be submitted), however, gave rise to a mean value of 1.85×10^{-5} cm^2 s^{-1} for the waters in the narrow channels, corresponding to an average end-to-end movement of approximately 4.7 Å during the 60-ps simulation. Thus, even these waters within the channels are relatively free to move and redistribute during the simulation. Only five waters moved less than 1 Å during the simulation, indicating that they were in locations where there was a significant barrier to diffusion. The energies of these waters at the start of the 60-ps simulation, however, are negative (−10.0 to −13.4 kcal/mol), indicating that there are no large (artifactual) forces on the protein molecules which would cause significant differences between the two asymmetric units (in fact, the energies of all 1429 waters were negative at the start of the simulation).

Twin-Range Nonbonded Cutoff. A cutoff of 15 Å was used for the electrostatic and van der Waals interactions (Kitson & Hagler, 1988). This is a large cutoff distance when compared to those commonly used in simulations of biological systems, but as we shall show, it is necessary in order to account for significant features of the crystal structure (Kitson & Hagler, 1988) and especially for significant long-range protein–ion interactions affecting both the ion distribution and the protein structure (F. Avbelj, J. Moult, D. H. Kitson, M. N. G. James, D. Hadzi, and A. T. Hagler, to be submitted). Commonly used cutoff distances in simulations of biological systems have typically been in the range 6–8 Å (Karplus & McCammon, 1981; Levitt, 1983b; Åqvist et al., 1985, 1986; Haneef et al., 1985; Post et al., 1986). In simulations of the crystal of bovine pancreatic trypsin inhibitor, cutoffs of 8 Å (van Gunsteren et al., 1983) and 11 Å (Berendsen, et al., 1986) were used.

While the use of a longer cutoff is desirable (Kitson & Hagler, 1988), it carries with it the penalty of increasing the time required to perform the simulation. With the twin-range method (Streett et al., 1978; Berendsen et al., 1986) the cutoff is increased with minimal impact on computational demand. In this method, short-range interactions (less than 9.5 Å in this case) are calculated at each step of the dynamics, and the longer range interactions (between 9.5 and 15 Å) are calculated every 20 steps, i.e., each 20 fs. This approximation is viable because long-range interactions change in magnitude and gradient significantly more slowly than short-range interactions. In order to ensure the validity of this method, it was tested extensively by comparing the forces on a number of atoms determined by the twin-range method with the corresponding quantities from a full calculation at each step. A set of typical results is given in Figure 1. These plots illustrate two important findings. First, the derivatives for the longer range interactions change much more gradually than do those for the shorter range interactions. Second, the approximation in which the long-range interactions are calculated only once every 20 steps has no more than a negligible effect on the derivatives and therefore can be used as a means of decreasing the time required for the simulation and increasing the cutoff distance.

Molecular Dynamics Procedure. A full valence force field (VFF) (Hagler, 1985) was used to calculate the energies and forces of the system (Dauber-Osguthorpe et al., 1988). The nonbond parameters for water–water interactions were taken from a simple rigid three-point charge (SPC) model (Berendsen et al., 1981) (parameters were added to this molecule to allow for internal flexibility of the water molecules). Protein–water interactions were taken as the geometric mean of the appropriate protein and water parameters.

During the minimization and dynamics a switching function was used to scale down the electrostatic and van der Waals interactions from full strength to zero between 11 and 15 Å to avoid discontinuities of energy and force associated with groups passing in and out of the cutoff distance (Swope et al., 1982). The leapfrog integration algorithm (Hockney & Eastwood, 1981) with a 1-fs time step was used. As described above, the twin-range cutoff method was invoked and the neighbor list was updated every 20 fs. To avoid slow temperature drift of the system, it was weakly coupled to a thermal bath at 300 K, with a temperature relaxation time of 100 fs (Berendsen et al., 1984). The simulation was performed with the DISCOVER program (Biosym Technologies, Inc., San Diego, CA), which was vectorized to run efficiently on the Cray X-MP/48 computers at the San Diego Supercomputer Center

and at the National Center for Supercomputing Applications (Champaign, IL).

After an initial check to ensure that the energies produced were identical with those in the Monte Carlo simulation, 100 steps of steepest descent energy minimization were performed to remove large energy gradients from the system. Deviation from the X-ray crystal structure of the protein was 0.2 Å RMS at that point. Initial atomic velocities for the molecular dynamics were assigned from a Maxwellian distribution at a temperature of 300 K, and an initial 11 ps of dynamics was performed on the solvent alone with the protein atoms held fixed. This was intended to ensure that the protein was not subjected to artifactually large forces due to the initial solvent distribution. Subsequent comparison of runs with and without this step showed the average structure of the protein to differ by only 0.1 Å RMS, so this step was probably unnecessary, partly because of the previous relaxation in the Monte Carlo run from which the initial solvent positions were taken. After this 11 ps of dynamics with the protein atoms fixed, the solvent was minimized by an additional 100 steps of steepest descent. Sixty picoseconds of dynamics of the full system was then performed. The average potential energy of the system decreased sharply over the first 16 ps and then leveled out somewhat (Figure 2), and therefore, much of the analysis described below was carried out over the 16–60-ps time period.

Results and Discussion

In the following section we describe the analysis of the molecular dynamics simulation of the SGPA system. The output of the simulation, in the form of the trajectory of each of the atoms in the system and a breakdown of the energy of the system at each step, gives rise to an extensive volume of information. Some features, as seen below, have been analyzed in extreme detail, while many other facets of the system will be reported elsewhere. The first areas to be investigated were the overall structural features obtained from the simulation and the agreement of these features with the experimental (X-ray) system. This was carried out at several levels, from overall RMS deviations of the complete molecules and analysis of RMS deviations of individual secondary structures to analysis of RMS deviations of individual residues.

The second major focus of this study has been on the interaction of the protein with the solvent and ions. Here we have investigated the interactions of different residues in the protein with the solvent. Those chosen have high RMS deviations from the experimental structure. This has yielded insight into the factors that lead to the large deviations observed for these residues.

In order to appreciate the behavior of different parts of the protein structure, we have divided it into substructures. In common with the other chymotrypsin-class serine proteases, SGPA consists of two domains. The active site of the enzyme lies in the cleft between these domains. Each domain is formed from four β strand–turn–β strand motifs, known as two-stranded ribbons (Richardson, 1981). In each domain, three of the two-stranded ribbons are loosely hydrogen bonded to form a six-stranded antiparallel β barrel, while the fourth forms several hydrogen bonds to the side of the barrel. In addition, there are three α-helical segments, the first between the third and fourth strands in domain 1, the second in the connecting segment between the two domains, and the third at the C-terminus of the sequence. Figure 3 shows the main-chain fold of SGPA. The deviations in the simulated structure from the experimental structure have been analyzed in terms of the substructures, which are now described.

The protein begins with the N-terminal loop (residues 16–41). [The residue numbers used throughout this paper for SGPA are based on the homologous chymotrypsin numbering (Brayer et al., 1978).] This loop differs from that in α-chymotrypsin (Birktoft & Blow, 1972) in that, in SGPA, the protonated N-terminal reisdue (Ile^{16}) is directed into a solvent region whereas, in α-chymotrypsin, this residue forms a buried salt bridge to Asp^{194}. A salt bridge to Asp^{194} is, however, formed in SGPA by Arg^{138}, and this conserved salt bridge plays an important role in stabilizing the active-site region (Brayer et al., 1978).

The histidine loop (residues 42–58) contains the active-site His^{57}. This loop begins and ends with Cys residues, which form a disulfide bridge. This disulfide bridge is conserved in SGPA and α-chymotrypsin, where it is found to have the same conformation.

The histidine loop is followed by a short connecting segment (residues 59–65), which is partly α-helical (the helix begins in the histidine loop and runs from residues 56 to 62). In α-chymotrypsin the connecting segment leads to a large loop of 21 residues referred to as the uranyl loop because a uranyl ion binds to elastase in this region (Shotton & Watson, 1970). In SGPA this loop is almost entirely deleted, consisting of only five residues (65A–86), and, along with the deletion of a large part of the N-terminal loop, is primarily responsible for domain 1 being significantly shorter than the corresponding domain in the mammalian enzyme and also shorter than domain 2.

The last loop in the first domain is the aspartate loop (residues 87–108). This loop includes the active-site residue Asp^{102}. The side chain of residue Phe^{94} in this loop helps to shield Asp^{102} from direct contact with the solvent.

The chain now continues with a long connecting segment (residues 109–130), which links the two domains together. This segment includes a short helical region (residues 110–113) and a loop (residues 117–124). This latter loop does not exist in α-chymotrypsin.

The first loop of the second domain is the autolysis loop (residues 131–163), which is 21 residues in SGPA compared with 33 in α-chymotrypsin. This is followed by the methionine loop (residues 164–182), which has a different conformation than in α-chymotrypsin.

A short connecting segment (residues 183–194) then leads to the serine loop (residues 195–213), which includes the active-site Ser^{195}. The final loop in SGPA is the specificity loop (residues 214–228), which constitutes part of the primary specificity pocket of the enzyme. The C-terminal segment (residues 229–242) of the protein includes the third helical region (residues 230–238), which is somewhat shorter than in α-chymotrypsin.

Fit of Calculated to Experimental Protein Structure. The RMS deviations between the experimental and calculated protein structures averaged over 10-ps intervals during the simulation are shown in Table I. Atoms with an experimental temperature factor greater than 20 $Å^2$ were omitted from the calculations of the RMS deviations, as discussed above, due to the systematic errors of the coordinates, which are correlated to these large temperature factors. The side chain from C^γ onward of residue Arg^{221}, which is disordered (this residue was refined as an alanine), was also omitted. There are 183 out of the 1265 non-hydrogen atoms in the protein molecule that either were not seen in the X-ray or have an experimental temperature factor greater than 20 $Å^2$.

Table I shows that, for the structure averaged over the 16–60-ps time period and over the two molecules in the asymmetric unit, the RMS deviation from the experimental structure was 1.17 Å for all heavy atoms and 1.09 Å for C^α

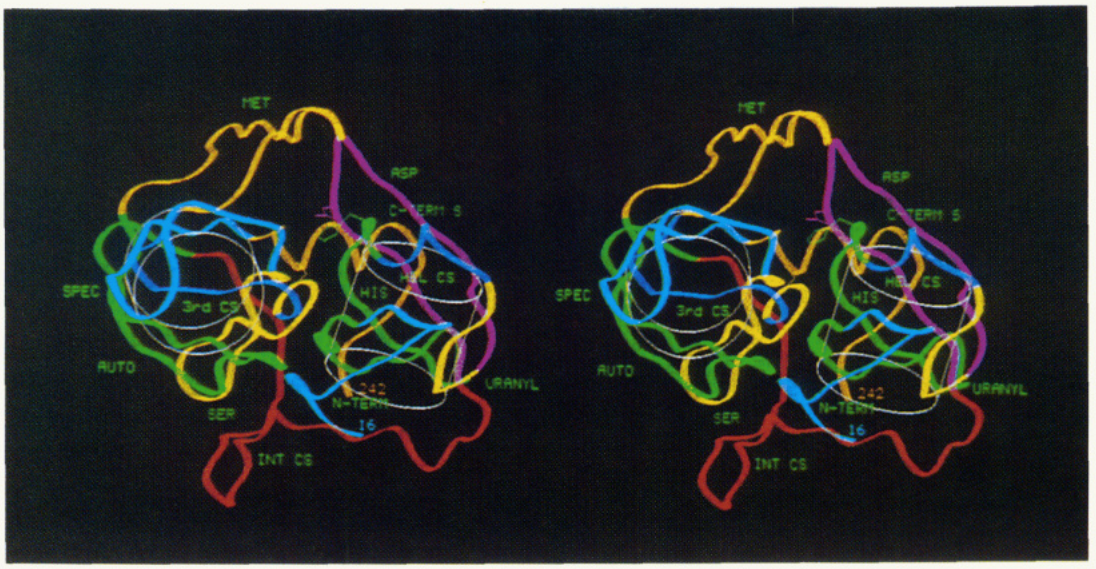

FIGURE 3: The main-chain fold of SGPA, indicating the positions of the substructural elements that make up the structure. These include the N-terminal loop (residues 16–41; light blue), the histidine loop (residues 42–58; green), the helical connecting segment (residues 59–65; blue), the uranyl loop (residues 65A–86; yellow), the aspartate loop (residues 87–108; purple), the interdomain connecting segment (residues 109–130; red), the autolysis loop (residues 131–163; green), the methionine loop (residues 164–182; orange), the third connecting segment (residues 183–194; blue), the serine loop (residues 195–213; yellow), the specificity loop (residues 214–228; light blue), and the C-terminal segment (residues 229–242; orange). The side chains of the active-site residues His^{57} (green), Asp^{102} (purple), and Ser^{195} (yellow) are shown. The cylinders indicate the positions of the β barrels into which the SGPA chain is folded.

Table I: Comparison of RMS Differences between the Experimental (X-ray) Structure of SGPA and Structures Averaged over Various Time Periods of the Molecular Dynamics Simulation

average structure[a]	RMS (Å)[b]			
	1[c]/X-ray	2[c]/X-ray	$\overline{1+2}$[d]/X-ray	1/2
⟨0–10⟩	1.02 (0.88)	0.69 (0.65)	0.72 (0.63)	0.96 (0.88)
⟨10–20⟩	1.52 (1.35)	1.05 (1.00)	1.05 (0.97)	1.54 (1.38)
⟨20–30⟩	1.57 (1.39)	1.24 (1.19)	1.14 (1.06)	1.67 (1.51)
⟨30–40⟩	1.69 (1.52)	1.25 (1.18)	1.17 (1.09)	1.84 (1.65)
⟨40–50⟩	1.78 (1.61)	1.30 (1.22)	1.21 (1.13)	1.97 (1.76)
⟨50–60⟩	1.84 (1.67)	1.40 (1.31)	1.27 (1.19)	2.03 (1.81)
⟨16–60⟩	1.67 (1.50)	1.25 (1.19)	1.17 (1.09)	1.79 (1.61)

[a] This column gives the time period (in picoseconds) over which the averages were calculated. [b] The first RMS value given is for all heavy atoms with an experimental temperature factor of less than 20 Å^2. The value in parentheses if for the α-carbon atoms only. [c] Structures 1 and 2 refer to the two SGPA molecules in the asymmetric unit (see Experimental Procedures). [d] This is the structure averaged both over time and over the two molecules in the asymmetric unit.

atoms. The RMS deviations from experiment for the individual molecules were significantly larger, 1.67 and 1.25 Å for the heavy atoms of molecules 1 and 2, respectively. The deviations for the second molecule are much smaller than the deviations for the first one (by ~0.4 Å). Table I also shows that the deviations for both of the molecules and for the average structure are still continuing to increase slightly. Significantly, the deviations for the second protein molecule increase very little over the 20–50-ps portion of the simulation but then begin to increase again over the last 10 ps. This is reminiscent of the behavior found in a long Monte Carlo simulation of water by Mezei et al. (1979) and raises concern regarding conclusions about deviations drawn from shorter simulations of protein structure, as well as of liquids. In addition, Table I shows that the structure averaged over the two molecules and over time is closer to the experimental structure than either of the individual molecules and that, by the end of the simulation, the two simulated molecules are slightly more than 2 Å different from one another (shown in the last column of Table I). This indicates that the two protein molecules are moving in significantly different "directions" in phase space. As discussed below, this has implications for the mechanistic interpretation of deviations seen in one simulation of a single protein.

Comparison with Other Studies. Similar values for RMS deviations have been obtained in two molecular dynamics studies of a smaller system—bovine pancreatic trypsin inhibitor (BPTI) (van Gunsteren et al., 1983; Berendsen et al., 1986). Table II compares the details of these simulations and the resulting RMS deviations with the SGPA simulation. As this table shows, the deviations are comparable to those observed here. In addition, after 40 ps of molecular dynamics of BPTI, the structure was continuing to diffuse away from the experimental structure. The rate of migration from the experimental structure is about the same in the BPTI study as that found here for SGPA, and thus an even greater deviation might be expected for a 60-ps simulation. In the BPTI system, the RMS deviations for the structures averaged over each of the molecules in the unit cell were also averaged over the four molecules, and as we have seen for both systems, averaging the deviations leads to lower apparent values (van Gunsteren et al., 1983; Berendsen et al., 1986). This effect is due to the essentially random nature of the deviations, which tend to cancel out when averaged together over more than one molecule. It is not clear that this is a meaningful measure of fit to experiment.

In general, larger molecules provide a more stringent test of the potential and techniques, since low-energy deformations usually involve more sizable movements, resulting in higher RMS deviations for larger systems. Thus, if the same procedure is applied to two proteins of differing size, one would expect a significantly higher RMS deviation for the larger protein [for example, an early folding study on BPTI produced a 6-Å RMS deviation (Levitt & Warshel, 1975) while the same procedure applied to lysozyme resulted in a 12-Å RMS deviation (Warshel & Levitt, 1976)].

It is interesting to note that the deviations for the longer (40 ps) BPTI simulation are less than those for the shorter (20 ps) simulation. Several improvements were made in the

Table II: Comparison of Simulations of Proteins in the Crystal Environment

	simulation		
	BPTI (I)[a]	BPTI (II)[b]	SGPA
system	4 proteins (58 residues each) 560 water molecules 0 ions united-atom model[c] 3948 atoms	4 proteins (58 residues each) 552 water molecules 24 Cl^- ions united-atom model[c] 3952 atoms	2 proteins (181 residues each) 1429 water molecules 26 $H_2PO_4^-$, 16 Na^+ ions all atoms, explicit 9427 atoms
force field	diagonal terms, van der Waals, and Coulombic; 8-Å cutoff	diagonal terms, van der Waals, and Coulombic; 8-Å cutoff for van der Waals, 11-Å cutoff for Coulombic[d]	diagonal terms, cross terms, van der Waals, and Coulombic; 15-Å cutoff[e]
length (ps)	20	40	60
RMS deviations[f,g]			
time period (ps)	8–20	10–40	16–60
molecule 1	1.66 (1.13)	1.66 (1.01)	1.67 (1.50)
molecule 2	1.57 (1.12)	1.40 (0.98)	1.25 (1.19)
molecule 3	1.62 (1.06)	1.52 (1.01)	
molecule 4	1.63 (0.94)	1.50 (1.15)	
av over all molecules	1.19 (0.82)	1.26 (0.91)	1.17 (1.09)

[a] van Gunsteren et al. (1983). [b] Berendsen et al. (1986). [c] In this model, the hydrogen atoms attached to carbon atoms are incorporated into the latter; all other hydrogens are included explicitly. [d] The Coulombic interactions between 8 and 11 Å are recalculated only once every 10 steps (20 fs). [e] The van der Waals and Coulombic interactions between 9.5 and 15 Å are recalculated only once every 20 steps (20 fs). [f] The first value given is for all heavy atoms; the value in parentheses is for C^α atoms only. [g] The deviations for SGPA exclude atoms with experimental temperature factors greater than 20 $Å^2$; if all atoms, irrespective of temperature factor (except for the disordered atoms of Arg^{221}), are included, the deviations for SGPA are 1.83 (1.63), 1.45 (1.22), and 1.29 (1.13) Å for molecules 1, 2, and the average over molecules 1 and 2, respectively (with C^α deviations in parentheses).

Table III: Comparison of the Geometry of the Structure Calculated from the Time-Averaged Coordinates from the Simulation with That of the Experimental Structure for the Backbone in SGPA[a]

		mean value		RMS deviation		max value		min value	
bond/angle	no.	⟨MD⟩	exptl	⟨MD⟩	exptl	⟨MD⟩	exptl	⟨MD⟩	exptl
C^α—N (Å)	181	1.48	1.47	0.016	0.020	1.51	1.54	1.43	1.43
C^α—C (Å)	181	1.55	1.52	0.020	0.018	1.60	1.58	1.50	1.47
N—C (Å)	180	1.31	1.32	0.027	0.018	1.36	1.37	1.17	1.28
C=O (Å)	181	1.20	1.24	0.036	0.017	1.23	1.28	1.00	1.20
ω (deg)[b]	179	178.6	179.5	10.7	3.7	210.9	188.1	151.8	166.9
τ (N—C^α—C) (deg)	181	113.9	109.9	3.1	3.6	124.5	119.2	106.6	99.2

[a] Note that the simulation averages are for the single structure represented by the time-averaged Cartesian coordinates from the simulation, not averages of the geometric values over the simulation. [b] The cis amide bond, between residues Phe^{94} and Pro^{99}, has been omitted from these calculations.

setup and techniques used for the 40-ps simulation (Berendsen et al., 1986), and this demonstrates the effect that such factors can have.

More recently, an important simulation of BPTI has been carried out by Levitt and Sharon (1988) in a solution containing 2607 water molecules. In order to improve agreement to experiment, several modifications were made to a calculation by these workers that had yielded results equivalent to those described above, a deviation of 2.3 Å after 40 ps. These included addition of internal flexibility to the water molecles, inclusion of all hydrogen atoms and the net charge on all ionizable groups, and perhaps most importantly, the implementation of smooth and continuous truncation of nonbonded interactions using the methods of Brooks et al. (1983). The simulation of this system was carried out for 200 ps and gave much better convergence (to an RMS of ~1.2 Å over the last 100 ps), although, again, toward the final ~20 ps there appears to be either a fluctuation or a drift toward an RMS of 1.5 Å.

In any case, the comparison of the results of this BPTI simulation with the previous results and with those reported here presents insight into the nature of dynamic simulations. Although the RMS is comparable, if the feature over the final 20 ps is a fluctuation to 1.5 Å rather than a transition to a new conformation, the results converge faster and would appear somewhat closer to experiment than in this SGPA simulation if we extrapolate the SGPA results to similar times. It is interesting to compare the two simulations to examine what the cause may be. One possible difference could result from differences in the potential functions. This is unlikely to have led to more accurate results in the BPTI simulation as the important nonbond potentials are based on the same work as used here and a more simplified intramolecular potential (a diagonal force field) was used. Thus the force fields are comparable and should not lead to significant differences. Of course, it could be, once again, simply the difference between the two proteins. BPTI has only 58 residues, which are further constrained by three disulfide bonds, thus presenting a much more rigid, constrained system than the 181 residue SGPA, which has only two disulfide bonds. The latter indubitably plays a role, but perhaps the most important factor is the smooth truncation method. As the authors point out, discontinuities and spurious forces are removed and a better behaved simulation results. In principle, this is a desirable characteristic of a simulation. An unanticipated result of these discontinuities and spurious forces, however, is a better sampling of conformational space. Thus, if we examine the potential energy of the SGPA system, we see that as the simulation proceeds, the energy is continuing to decrease. Thus (unfortunately, but correctly), the protein is sampling low-energy structures somewhat removed from the experimental structure. The smooth truncation may be working "too well", not allowing the BPTI in this simulation to get over small barriers and explore other lower energy regions, as the previous

Table IV: RMS Deviation[a] between the 16–60-ps Time-Averaged Structure and the Experimental Structure for the Different Substructures of SGPA

structural substructure	residues	temp fac (Å²)[b]	molecule 1		molecule 2	
			fit protein[c]	fit substr[d]	fit protein[c]	fit substr[d]
domain 1	16–127	12.31	1.92 (1.67)	1.82 (1.60)	1.34 (1.16)	1.31 (1.11)
domain 2	128–242	15.80	1.36 (1.32)	1.23 (1.18)	1.15 (1.22)	1.14 (1.19)
N-terminal loop	16–41	14.99	1.45 (1.31)	1.01 (0.77)	1.06 (0.93)	0.72 (0.54)
histidine loop	42–58	11.36	2.49 (2.18)	2.31 (2.07)	1.55 (1.37)	1.30 (1.06)
helical connect. seg	59–65	11.34	1.90 (1.83)	0.81 (0.49)	1.47 (1.49)	1.16 (1.03)
uranyl loop	65A–86	11.30	1.49 (1.52)	0.93 (0.60)	1.27 (1.16)	0.82 (0.53)
asparate loop	87–108	9.36	1.93 (1.50)	1.49 (0.87)	1.51 (1.23)	1.28 (0.86)
interdomain connect. seg	109–130	14.73	1.51 (1.35)	1.33 (1.22)	1.00 (0.82)	0.93 (0.77)
autolysis loop	131–163	14.92	1.33 (1.15)	0.74 (0.57)	1.14 (1.12)	0.61 (0.48)
methionine loop	164–182	17.89	1.67 (1.82)	1.01 (0.91)	1.50 (1.68)	0.88 (0.93)
third connect. seg	183–194	16.12	1.18 (1.11)	0.86 (0.70)	1.40 (1.34)	0.65 (0.42)
serine loop	195–213	10.25	0.95 (0.96)	0.53 (0.50)	0.82 (0.86)	0.69 (0.69)
specificity loop	214–228	22.80	1.33 (1.41)	0.87 (0.94)	0.72 (0.71)	0.50 (0.53)
C-terminal segment	229–242	12.70	1.52 (1.34)	1.14 (0.78)	1.15 (1.18)	0.74 (0.68)

[a] The RMS calculation included only non-hydrogen atoms with an experimental temperature factor of less than 20 Å². The first number given in each column is the deviation for all included heavy atoms. The number in parentheses is the deviation for C^{α} atoms only. The deviations are in angstroms. [b] This is the average experimental temperature factor for all heavy atoms in the structural substructure. [c] The deviations in this column were obtained by fitting all heavy atoms with temperature factors less than 20 Å², of the complete time-averaged protein structure, to the experimental structure and then calculating the RMS deviation between the corresponding pairs of atoms in the appropriate residues for each substructure. [d] The deviations in this column were obtained by fitting the heavy atoms of the residues within the substructure being analyzed in the time-averaged protein structure to the same atoms in the experimental structure and then calculating the RMS deviation between the same pairs of atoms.

simulations did (Table II), albeit further from the experimental structure. (This may be the nature of the event toward the end of the simulation corresponding to the 1.5-Å RMS deviation.)

Clearly all of these methods taken together can yield insight into the characteristics of various techniques and approximations as well as the nature of the dynamics of proteins. Further comparative studies will be extremely useful in this regard.

Table III shows a comparison of the bond lengths and angles of the peptide backbone in the experimental refined structure with those of the structure calculated from the time-averaged Cartesian coordinates of the protein from the simulation. The apparent shortening of the C=O bond in the simulated structure reflects the effect of the wagging motion of the carbonyl oxygen, an effect well-known in small molecule crystallography. This type of effect has also been noted for an in vacuo simulation (Kuriyan et al., 1986). Comparison of the RMS deviations of the ω torsion angles from the mean value, and of the minimum and maximum ω's, shows a significantly greater range of deviations in these torsion angles in the molecular dynamics than in the experimental structure. That significant deviations in ω occur in experimental structures is demonstrated by a recent analysis (Ashida et al., 1987) of conformational parameters of a large number of peptide crystals in the Cambridge Structural Database (Allen et al., 1979). This analysis showed that deviations of more than 10° from planarity occur for approximately one out of ten amino acids in linear, unstrained peptides. The most nonplanar ω's found in the Ashida et al. study (157 and 201°) are, however, closer to planarity than those found in the present study (152 and 211°). In the experimental SGPA structure, only one of the 181 residues has an ω value more than 10° away from planarity, whereas in the time-averaged structures 59 ω's in molecule 1 and 71 in molecule 2 are outside of the range 170–190°. Thus, the study of peptide crystal structures and the results of the present study indicate both that the restraint on ω's used in crystallographic refinement may be too severe and that the potential function used for the current simulation may be too weak for rotations around the amide bond.

Fit of Individual Substructural Elements of the Protein Structure. In order to characterize the distribution of deviations throughout the molecule—to see whether all parts are equally distorted or, if not, which regions give rise to the largest deviations, we calculated the deviations for each of the two domains of the protein and for each of the substructures from which these domains are constructed. These comparisons were made for the 16–60-ps time-averaged structure, again excluding any atom with an experimental temperature factor greater than 20 Å² from the calculation. The results are presented in Table IV and are illustrated in Figure 4.

Table IV shows larger deviations for the substructures of molecule 1 than for molecule 2, reflecting the RMS deviations in Table I. The table also shows that, for both molecules, domain 1 (residues 16–127) deviates more from the experimental structure than domain 2 (residues 128–242), by ~0.6 and 0.2 Å for molecules 1 and 2, respectively. The largest deviation for both molecules lies in the histidine loop (RMS deviation = 2.49 and 1.55 Å for molecules 1 and 2, respectively) and the next largest corresponds to the aspartate loop (RMS deviation = 1.93 and 1.51 Å, respectively).

Rigid-Body Motions vs Conformational Changes. Two distinct types of deviations from the experimental structure were observed to occur in this simulation: "rigid-body" movements of substructures as a whole or distortions within substructures. To determine to what extent the deviations arise from rigid-body movements of the substructure, or conformational changes within the moiety, the deviations were recalculated by fitting only the residues of the appropriate unit (rather than from calculating the deviation after the entire protein was fit), both for the substructures listed above (see Table IV) and for other structural elements of the protein.

Calculated in this way, the deviation for the residues in the first helix of molecule 1 (which begins in the histidine loop and extends into the helical connecting segment, residues 56–62) decreases from 2.78 to 0.80 Å. Thus, since the deviation was reduced considerably when only the residues within the helix were fit, a large part of the deviation of the helix must have been due to a rigid-body motion of the entire helix. Examination of the experimental and calculated structures showed that this was the case, with the complete helix of molecule 1 moving inward, toward the center of the protein, as can be seen in Figure 4.

For the histidine loop of molecule 1 as a whole, the RMS

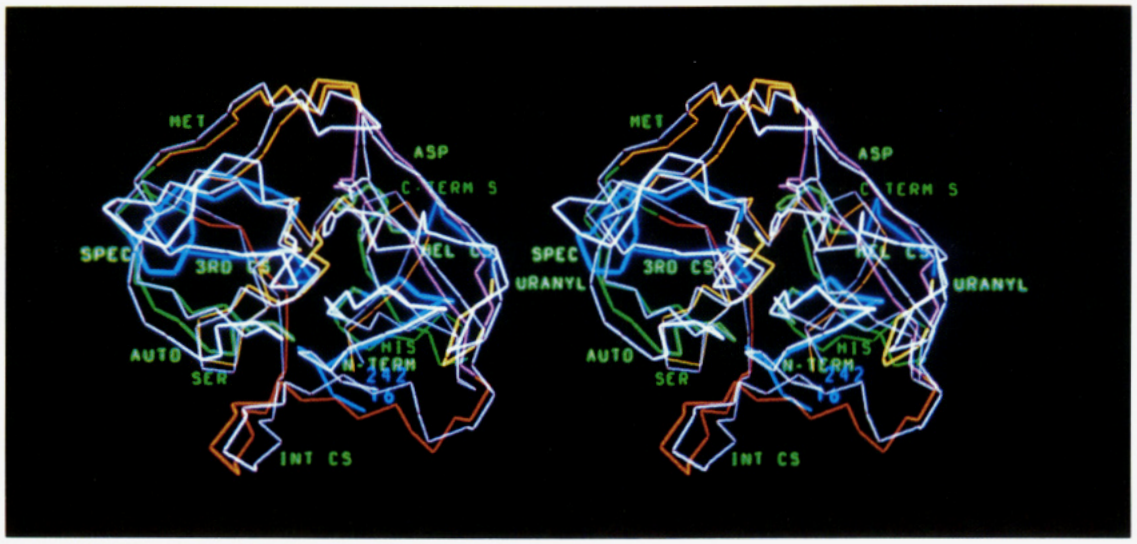

FIGURE 4: Deviations between the time-averaged structure of molecule 1 and the experimental structure of SGPA. The time-averaged structure is shown in white. The substructures displayed in the experimental structure, and their color code, include the N-terminal loop (residues 16–41; light blue), the histidine loop (residues 42–58; green), the helical connecting segment (residues 59–65; dark blue), the uranyl loop (residues 65A–86; yellow), the aspartate loop (residues 87–108; purple), the interdomain connecting segment (residues 109–130; red), the autolysis loop (residues 131–163; green), the methionine loop (residues 164–182; orange), the third connecting segment (residues 183–194; dark blue), the serine loop (residues 195–213; yellow), the specificity loop (residues 214–228; light blue), and the C-terminal segment (residues 229–242; orange). The side chains of the active-site residues His^{57} (green), Asp^{102} (purple), and Ser^{195} (yellow) are shown. Significant deviations can be seen for several of the substructures, notably the histidine, methionine, and specificity loops. Many of the atoms in the methionine and specificity loops, however, have high temperature factors in the experimental structure and this is reflected, especially for the specificity loop, in the relatively low RMS deviations, shown in Table IV, where atoms with temperature factors greater than 20 $Å^2$ have been omitted from the comparison. The motion of the first helix (residues 56–62), which is partly in the histidine loop (green) and partly in the helical connecting segment (dark blue), toward the center of the protein can be seen in this figure. As shown, the helix moves into the protein as a rigid body, with only small conformational changes taking place within the helix itself.

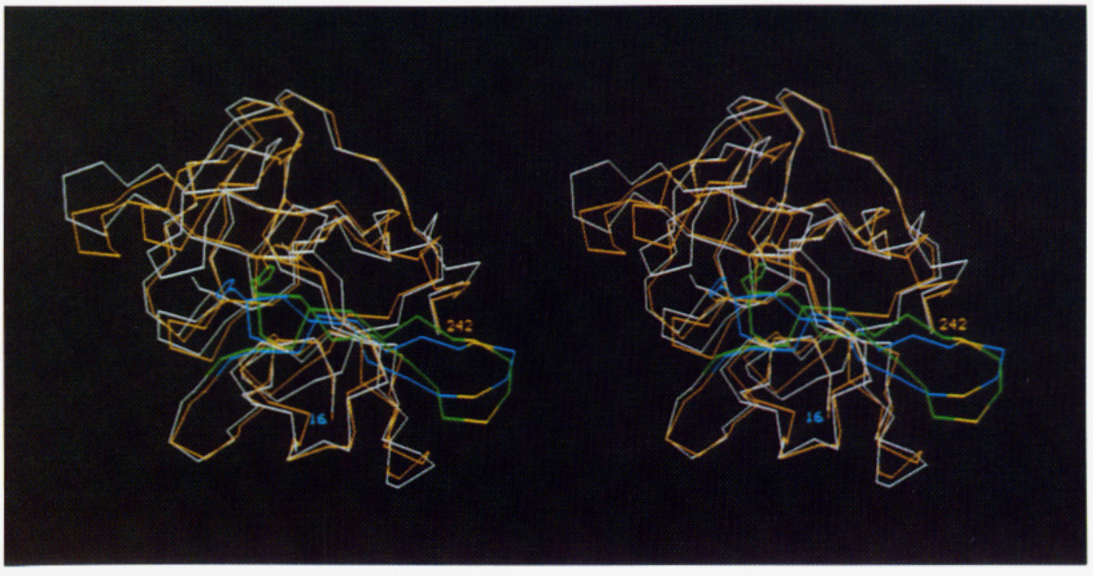

FIGURE 5: Changes in conformation of the histidine loop of molecule 1 that take place during the molecular dynamics simulation. The experimental structure is shown in orange, while the time-averaged structure of molecule 1 is shown in white. The histidine loop is colored blue in the experimental structure and green in the time-averaged structure. The side chains of the active-site residues are shown. The C^{α} atoms of residues Val^{48B} and Val^{49} (which are in the loop) are colored yellow. The distance between these atoms increases from 5.12 Å in the experimental structure to 8.07 Å in the time-averaged structure, reflecting the distortion in the structure of the loop.

deviation did not change significantly (from 2.49 to 2.31 Å) when only the residues within the loop were fit. This indicates that most of the deviation arose from a conformational change within the loop. Examination of the experimental structure showed that the loop contains a β turn around residues Val^{48B}–Val^{49}. In the time-averaged structure this turn has become considerably wider, as shown in Figure 5. The difference is greatest in the turn region of the loop (Figure 5), and there is relatively little change in the structure of the regions to which the loop is attached.

Table IV shows that the deviations for substructures in molecule 2 are all (with the exception of the third connecting segment) less than the deviations for molecule 1. The first helix in molecule 2 again shows a relatively large deviation when the complete molecule is fit (2.69 Å), and this time the deviation remains relatively high when only the atoms in the

helix are fit (1.51 Å), indicating a more substantial conformational change within the helix.

Correlation between Experimental Mobility and RMS Deviations. One might expect to see a correlation between the experimental mobility (as reflected in the temperature factors) and the RMS deviation of the different substructures of the protein. A comparison of the average experimental atomic temperature factors with the deviations for the substructures (Table IV) reveals, however, that this is not the case. For example, the histidine loop has the largest RMS deviation in both molecules but has a relatively low average experimental atomic temperature factor (11.36 Å^2, compared with the average for the whole protein of 14.12 Å^2). In contrast to this is the specificity loop, which does have both a high mobility and a large deviation. On the basis of the experimental temperature factors, this loop is the most mobile part of the molecule, with all atoms contained in six residues (218–220, 222–224) having temperature factors greater than 20 Å^2. In addition, the side chain from C^γ onward of residue Arg^{221} is disordered (the residue was refined in the crystallographic study as an alanine). This mobility presumably reflects a rather flat energy surface for this portion of the protein. That is, a range of conformations of this region are accessible and result in roughly equienergetic structures. This was reflected in a deviation between the experimental structure and the time-averaged structure of the loop of 2.50 Å in molecule 1 when all atoms in the loop, irrespective of temperature factor (except for the disordered atoms of Arg^{221}), were included in the comparison (the deviation reduces to 1.33 Å when the atoms with a temperature factor greater than 20 Å^2 are omitted from the comparison).

Deviations of Important Structural Elements. We now examine the deviations for some of the important structural elements of the protein that were described above. The salt bridge between Asp^{194} and Arg^{138} is well maintained during dynamics, as shown by fitting the heavy atoms of Asp^{194} and Arg^{138} of the time-averaged structures onto the experimental structure, which yields values of 0.71 Å for molecule 1 and 0.70 Å for molecule 2. The fit is reflected in the small change in the distance between C^γ of Asp^{194} and C^ζ of Arg^{138}, which goes from 4.09 Å for the experimental structure to 4.33 and 4.32 Å in the time-averaged structures of molecules 1 and 2, respectively.

The conformation of the disulfide bridge between Cys^{42} and Cys^{58} changes somewhat during dynamics, the RMS value obtained by fitting these residues of the time-averaged structures to the experimental structure being 0.91 Å for molecule 1 and 1.13 Å for molecule 2. The most significant changes in torsion angles are in χ_1 of Cys^{42} (experimental = −68°, dynamics molecule 1 = −161°, molecule 2 = −178°), χ_1 of Cys^{58} (experimental = −65°, dynamics molecule 1 = −74°, molecule 2 = −158°), and χ_2 of Cys^{58} (experimental = −106°, dynamics molecule 1 = −46°, molecule 2 = +79°). On the other hand, the torsion angle around the S–S bond is reasonably reproduced in both molecules (experimental = −90°, dynamics molecule 1 = −82°, molecule 2 = −101°). In both molecules, these changes in torsion angles are correlated with substantial changes (of ~3 Å) in the position of the backbone on the Cys^{58} side of the disulfide bridge. In molecule 1, the entire helix of which Cys^{58} is a part is translated, largely as a rigid body as described above. Deviations of up to ~3 Å in backbone position continue to the turn in the histidine loop (at residues 48C–48D) of which this bridge is a part. In molecule 2, there is, again, a change of ~3 Å in the position of the backbone around Cys^{58}, but this is much more localized, from around residues 56 to 59. In both molecules, the change in the backbone position on the Cys^{42} side of the bridge is much smaller—around 1–2 Å.

Table V: RMS Deviation of Hydrophobic, Hydrophilic, and Charged Amino Acids

residue type	RMS deviation (Å) molecule 1	molecule 2
hydrophobic	1.46	1.19
hydrophilic	1.52	1.12
charged	2.51	1.77
charged nearest neighbors[a]	1.69	1.44

[a] These are the residues that flank the charged residues.

Phe^{94}, in the aspartate loop, sits near to the surface of the protein between the histidine loop and the methionine loop. During dynamics this residue moves in both molecules. In molecule 1 it moves inward, toward the center of the protein, giving an RMS, obtained by fitting the entire protein and then calculating the deviation for this residue, of 1.26 Å. In molecule 2 the Phe moves outward and the deviation is larger, at 1.67 Å. In neither case is it likely that the shift in the position of Phe^{94} would lead to a significant change in the shielding of the active site from solvent, which is a function performed by this residue.

Thus, we see that there are significant distortions in the secondary structure of both molecules and that underlying an RMS deviation for a complete protein molecule, described by a single number, lies a wide range of deviations of the substructures that make up the protein. Two types of deviations are possible (as well as a combination of these): rigid-body motions of groups of residues or conformational changes within these groups. Both types are seen in the SGPA structure. As described below, this analysis can be taken a step further by looking at the deviations of individual residues. These deviations are related to energetic changes that take place during the simulation. It is clear from these deviations that one could have difficulty in faithfully predicting the conformational results of site-specific mutagenesis experiments and in predicting free energies of association to proteins where the binding site may have undergone significant conformational changes. It is seen from this analysis that structural deviations resulting from theoretical simulations may be large enough to change the nature of the ligand–protein interactions and all thermodynamic data should be interpreted in corresponding structural terms as well.

Fit of Individual Residues. As the first step in this analysis, we characterize the deviations of the various types of residues. These results are given in Table V. As can be seen from this table, the largest deviation by far in both molecules occurs for the charged residues, 2.51 and 1.77 Å for molecules 1 and 2, respectively. We also, then, calculated the RMS deviation for the nearest neighbors of these charged residues and these were correlated, also being larger (1.69 and 1.44 Å for the two molecules) than the overall deviations. These results are consistent with the observation (Table IV) of the larger deviation of domain 1 in both molecules relative to domain 2 since, of the 17 charged residues in the molecule (11 positive and 6 negative), 11 are located in domain 1 (8 positive and 3 negative). Furthermore, domain 1 is also smaller than domain 2, containing 84 out of the 181 residues and so the density of charged residues is still higher. This suggests that errors in the representation of charged interactions may be one of the prime factors responsible for deviations in protein simulations. This will be evaluated in more detail below.

Energetic Analysis. In order to analyze the interatomic interactions responsible for these deviations, we have computed the breakdown of the energy of the residues into intraresidue

Table VI: Contributions to Total Potential Energies, Averaged over Different 1-ps Segments of the Simulation, for the Residues with the Largest RMS Deviations

residue	RMS deviation (Å) for molecule 1	RMS deviation (Å) for molecule 2		residue energies (kcal/mol) $\langle E\rangle_{-11-10}$[a] for molecule 1	$\langle E\rangle_{-11-10}$[a] for molecule 2	$\langle E\rangle_{0-1}$[b] for molecule 1	$\langle E\rangle_{0-1}$[b] for molecule 2	$\langle E\rangle_{59-60}$[c] for molecule 1	$\langle E\rangle_{59-60}$[c] for molecule 2
His[51]	4.2	0.5	$E_{internal}$[d]	101.7	101.6	103.3	97.6	103.4	101.1
			$E_{protein}$[e]	12.2	12.6	8.0	4.8	−0.6	−15.0
			E_{water}[f]	−38.9	−33.9	−44.6	−37.3	−50.0	−43.4
			E_{Na}[g]	6.1	0.3	−0.4	−0.3	1.9	−0.3
			E_{phos}[h]	0.0	0.0	−2.9	−0.1	−71.9	0.0
			E_{total}[i]	81.1	80.6	63.4	64.6	−17.2	42.3
His[57]	4.1	3.8	$E_{internal}$[d]	101.9	102.0	100.6	98.4	99.1	97.6
			$E_{protein}$[e]	−55.7	−55.0	−69.6	−62.5	−93.7	−19.4
			E_{water}[f]	−20.3	−22.9	−24.7	−16.3	−11.4	−34.2
			E_{Na}[g]	0.3	0.0	0.5	0.2	0.1	4.0
			E_{phos}[h]	−16.9	−32.7	0.0	−37.4	0.0	−109.3
			E_{total}[i]	9.3	−8.6	6.8	−17.5	−6.0	−61.3
Arg[88]	3.4	1.5	$E_{internal}$[d]	−15.0	−15.6	−15.7	−17.8	−17.6	−16.5
			$E_{protein}$[e]	3.5	3.6	−5.0	−7.2	7.0	−37.4
			E_{water}[f]	−44.6	−47.3	−44.1	−42.7	−50.7	−35.8
			E_{Na}[g]	0.4	13.6	0.0	0.0	0.0	0.0
			E_{phos}[h]	−25.9	−15.8	−51.2	−13.4	−100.0	−23.8
			E_{total}[i]	−81.7	−61.5	−116.0	−81.2	−161.3	−113.5
Asn[101]	2.7	2.5	$E_{internal}$[d]	−7.2	−7.1	−8.8	−8.7	−11.3	−13.3
			$E_{protein}$[e]	−25.0	−25.0	−22.9	−23.3	−22.2	−20.4
			E_{water}[f]	−0.3	−1.9	−6.3	−0.2	−5.9	−6.9
			E_{Na}[g]	0.5	1.0	3.1	−1.2	3.5	−0.6
			E_{phos}[h]	−0.1	−0.2	0.1	−0.5	0.4	1.5
			E_{total}[i]	−32.1	−33.2	−34.8	−33.8	−35.4	−39.7
Tyr[121]	2.6	0.5	$E_{internal}$[d]	75.1	74.9	71.0	69.7	67.2	69.2
			$E_{protein}$[e]	−28.5	−28.3	−27.0	−27.2	−26.9	−26.2
			E_{water}[f]	−7.8	−9.9	−7.6	−8.2	−7.7	−9.0
			E_{Na}[g]	1.3	1.0	0.3	1.0	−0.8	1.0
			E_{phos}[h]	0.3	0.5	0.3	0.8	0.0	0.4
			E_{total}[i]	40.4	38.3	37.0	36.0	31.8	35.4
Arg[158]	1.6	2.7	$E_{internal}$[d]	−5.2	−6.9	−11.4	−13.1	−8.8	−0.3
			$E_{protein}$[e]	−8.6	−7.9	−8.7	−5.7	−1.3	13.1
			E_{water}[f]	−43.7	−31.3	−42.0	−48.4	−87.6	−102.2
			E_{Na}[g]	32.5	44.9	53.9	45.8	64.3	75.5
			E_{phos}[h]	−108.3	−122.2	−127.8	−121.2	−123.7	−159.1
			E_{total}[i]	−133.2	−123.4	−136.0	−142.7	−157.2	−172.9
Gly[172]	4.2	4.4	$E_{internal}$[d]	29.9	29.9	26.9	28.6	28.6	26.2
			$E_{protein}$[e]	−8.7	−8.6	−9.1	−8.8	−9.2	−8.0
			E_{water}[f]	−6.8	−4.9	−5.6	0.3	−7.1	−4.9
			E_{Na}[g]	2.0	0.9	2.8	0.5	0.6	1.3
			E_{phos}[h]	−1.5	−1.1	−1.4	−6.2	−0.7	−2.5
			E_{total}[i]	14.9	16.2	13.5	14.3	12.1	12.0
Asp[194]	1.9	2.1	$E_{internal}$[d]	17.3	18.6	13.5	14.4	15.6	16.7
			$E_{protein}$[e]	−119.9	−119.3	−122.1	−117.8	−109.7	−101.0
			E_{water}[f]	12.5	11.4	9.5	5.4	2.1	−4.0
			E_{Na}[g]	−8.4	0.0	−9.4	−13.0	−10.8	−15.6
			E_{phos}[h]	11.2	20.6	8.7	20.3	6.0	23.0
			E_{total}[i]	−87.3	−68.6	−99.8	−90.6	−96.7	−80.8
Ser[217]	2.6	0.7	$E_{internal}$[d]	53.0	53.5	50.1	49.8	49.3	50.4
			$E_{protein}$[e]	−11.6	−11.4	−12.7	−12.2	−9.2	−11.2
			E_{water}[f]	−13.1	−11.4	−5.8	−8.4	−5.0	−9.2
			E_{Na}[g]	3.8	4.6	1.9	2.1	−0.9	2.8
			E_{phos}[h]	−3.1	−4.2	−7.2	−4.9	−5.4	−5.2
			E_{total}[i]	28.9	31.2	26.4	26.4	28.8	27.6
Glu[233]	1.3	2.7	$E_{internal}$[d]	28.1	26.5	26.5	22.9	23.3	31.5
			$E_{protein}$[e]	6.9	6.2	−6.2	−1.2	−16.9	−27.1
			E_{water}[f]	−45.7	−43.6	−9.5	−49.3	−35.5	−93.2
			E_{Na}[g]	−24.1	−27.5	−68.2	−57.2	−67.0	−84.1
			E_{phos}[h]	18.7	31.9	1.3	36.1	1.2	68.1
			E_{total}[i]	−16.2	−6.5	−56.2	−48.6	−94.9	−104.7

[a] $\langle E\rangle_{-11-10}$ refers to the average potential energy over the first picosecond of the initial dynamics simulation during which the protein was held fixed at the structure obtained by minimizing the X-ray protein structure (together with the Monte Carlo solvent structure) with 100 steps of steepest descent. In order to compare the potential energies of this system with those in which the protein is moving, it is necessary to take into account the potential energy of vibrations of the protein atoms. This was done by adding $(3/2)nRT$ to $E_{internal}$ for each residue, where n is the number of atoms in the residue. [b] $\langle E\rangle_{0-1}$ refers to the average potential energy over the first picosecond of the 60-ps dynamics simulation (during which the protein was allowed to move). [c] $\langle E\rangle_{59-60}$ refers to the average potential energy over the last picosecond of the 60-ps dynamics simulation (during which the protein was allowed to move). [d] Contribution due to internal interactions (including bond, angle, torsion angle, out of plane, and intraresidue nonbond and electrostatic interactions) within the residue. [e] Contribution due to interactions with protein molecules. [f] Contribution due to interactions with water molecules. [g] Contribution due to interactions with sodium ions. [h] Contribution due to interactions with dihydrogen phosphate ions. [i] Total energy of the residue (sum of *d* through *h*).

interactions and interactions with their aqueous environments, with the counterions and with protein. Unlike energy-minimized configurations of a system, instantaneous configurations from dynamics are not thermodynamically meaningful. Time-averaged values of the energies are required to get meaningful values. For these calculations, a 1-ps window was used in order to average out the effects of fast events such as bond stretching or short-term anomalously favorable or unfavorable nonbond interactions. These time-averaged energies are given, for several residues, in Table VI.

In the case of His^{51}, where the deviation in molecule 1 is 4.2 Å, but in molecule 2 is only 0.5 Å, we see that the total energy decreased from 81 to −17 kcal/mol in molecule 1 and from 81 to only +42 kcal/mol in molecule 2 over the course of the complete trajectory. Moreover, it is clear from Table VI that, in molecule 1, this residue is being pulled into solution toward the negative dihydrogen phosphates, as shown by the decrease in the interaction energy with dihydrogen phosphates to −72 kcal/mol, while in molecule 2 the histidine has essentially no interaction with dihydrogen phosphates. (Neither of the histidines interact with the negative counterions in the initial structure.) In molecule 2, the 38 kcal/mol of energy gained by the histidine residues arises mainly from more favorable interactions with the protein. The energy of interaction of His^{51} in molecule 1 with the water (−50 kcal/mol) after it has moved into the solvent is in line with estimates (Warshel & Russel, 1984) of the solvation energy of the imidazolium ion (−65 kcal/mol).

A slightly different situation exists for His^{57} where there is a large deviation in both molecules 1 and 2 (4.1 and 3.8 Å, respectively), although again there is a very different mechanism that would appear to be responsible for the large deviation of this residue in each of the molecules. This time, there is a relatively small decrease in energy in molecule 1 of 15 kcal/mol. This includes a gain of more favorable interactions of the histidine with the protein of 38 kcal/mol (−56 → −94 kcal/mol). Here, the histidine moves away from the solvent, as reflected in the energy of interaction with water, which becomes less favorable by 9 kcal/mol, and by the lack of a significant interaction with sodiums or dihydrogen phosphates in the final configuration, compared with an interaction energy of −17 kcal/mol with dihydrogen phosphates at the start of the simulation. In molecule 2, exactly the opposite occurs. The energy of protein–His interactions increases by 36 kcal/mol, rather than decreasing by roughly the same amount as in molecule 1. Interactions with water become more favorable by 11 kcal/mol, and a dramatic increase of 77 kcal/mol in the favorable interaction with dihydrogen phosphates takes place. It is thus coincidental that this results in essentially the same total structural deviation for this residue in both molecules and again points out the inadequacy of using the RMS values as a sole descriptor of conformation.

The theme of two mechanisms for the charged residues to achieve more favorable interactions is also observed with Arg^{88} and Glu^{233}. In the case of Arg^{88} we see, again, the pattern of the residue in one molecule (molecule 2) moving into the protein to gain more favorable residue–protein interactions, while in the other molecule (molecule 1) it moves into solution to make contact with water and counterions, gaining 74 kcal/mol of favorable interaction energy with dihydrogen phosphates. Glu^{233} gains favorable interactions with the protein in both molecules, but its interactions with solvent are also quite different in each of the two molecules. In molecule 2, Glu^{233} moves into solution (Figure 6, bottom) and gains significant favorable interactions with water (50 kcal/mol) [this is in rough agreement with the experimentally measured solvation energy of a formate ion of −86 kcal/mol (Locke & McIver, 1983)] and sodium ions (57 kcal/mol), although these are offset to some extent by an increase in unfavorable interactions with dihydrogen phosphates of +68 kcal/mol. In molecule 1 this residue stays closer to its starting position. Again, it gains favorable interactions with sodium (43 kcal/mol), but at the end of the simulation this residue essentially does not see any dihydrogen phosphates within 15 Å (Figure 6, top).

The trends with Arg^{158} are roughly the same in both molecules. In both cases, it would seem that the desire of this residue to interact more favorably with the water and dihydrogen phosphates leads to less favorable interactions with the protein. In both molecules, there is a dramatic increase in favorable interaction energy with water, of 44 kcal/mol for molecule 1 and 71 kcal/mol for molecule 2. Interactions with dihydrogen phosphates also become more favorable, by 15 kcal/mol in molecule 1 and by 37 kcal/mol in molecule 2. As the Arg^{158} side chain moves out into solution, it encounters sodium ions, resulting in a still less favorable interaction with these counterions, increasing from +32 kcal/mol initially to +64 kcal/mol in molecule 1 and from +45 to +76 kcal/mol in molecule 2, thus counterbalancing to some extent the more favorable interactions with the dihydrogen phosphate ions. The interaction with the protein becomes less favorable in both cases, by 7 kcal/mol for molecule 1 and by 21 kcal/mol for molecule 2.

The story with Asp^{194} is a variation on the theme seen for Arg^{158} of a similar trend for both molecules. Here the residue is in a very favorable energetic state in the initial structure due to its favorable interactions with the protein. This is, of course, the residue involved in the key salt bridge both in SGPA with Arg^{138} and in chymotrypsin with the N-terminus, which is thought to be responsible for activation of the enzyme (following cleavage of the zymogen in the latter case). In both molecules, we see a slight decrease in energy. In both cases, the interaction with solvent, both water and ions, becomes slightly more favorable, while the interaction with the protein becomes less favorable, resulting in only a relatively small decrease in the total energy of this residue.

The results for these residues show that there are a variety of configurations which afford residues with favorable energetic environments. The implications of this in terms of the interpretation of the results of simulations are discussed further below.

There are several possible explanations for the deviations of charged residues. Among these is that the electrostatic energies may simply not be adequately accounted for. As with all simulations currently performed, we have not taken into account charge transfer between charged groups and neighboring molecules, which is known to occur and may be seen in ab initio calculations. An example of this is given in Figure 7 where we have carried out ab initio calculations of two systems: a formate ion with a water molecule and a formate ion with a sodium ion. The population analysis given in this figure shows that, in the formate–water system, charge has been transferred to the water molecule, resulting in a total net charge on the formate ion of only −0.90 electron. In the system with the sodium ion, a greater amount of charge is transferred from the formate, the net total charge on the formate being −0.80 electron, while the sodium ion has a formal charge of +0.80.

Another possibility is, of course, the lack of explicit polarization in current potential functions (Russel & Warshel, 1985). Still another possibility is that an incorrect ionization state has been used for one or more residues. In the setup of

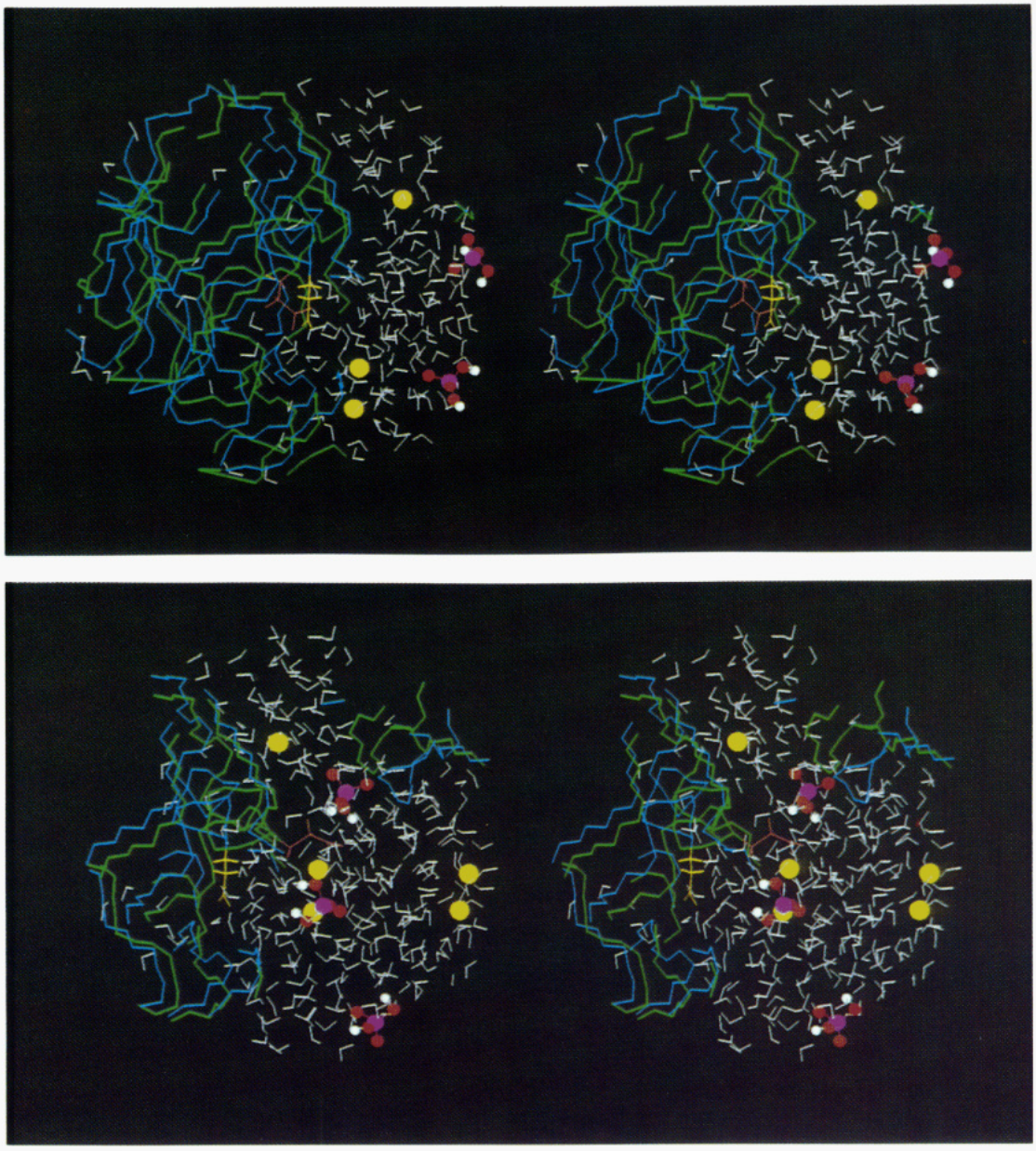

FIGURE 6: (Top) Change in the position of Glu^{233} in molecule 1 that takes place during molecular dynamics. The backbone of a portion of the experimental structure of the protein in the vicinity of Glu^{233} is shown in blue, while the backbone of a portion of the structure at the 60-ps time step of dynamics is shown in green. Glu^{233} is shown in yellow for the experimental structure and pink for the dynamics structure. The configuration of the solvent is shown at the end of the 60 ps of dynamics. Water molecules are colored white, dihydrogen phosphate ions are shown as color-coded (purple, phosphorus; red, oxygen; white, hydrogen) ball and stick representations, and sodium ions are shown as yellow spheres. As the figure shows, this residue does not move significantly during the simulation. (Bottom) During the dynamics, Glu^{233} of molecule 2 moves into the solvent. This is reflected in a large gain in favorable interactions with water (50 kcal/mol) and sodium ions (57 kcal/mol) but an unfavorable change in interactions with dihydrogen phosphate ions (36 kcal/mol).

the system we assumed that nearby charges do not change the ionization state of the ionized groups, except for Asp^{123} [Asp^{123} is involved in a carboxyl–carboxyl interaction with the terminal (charged) Leu^{242} of a neighboring molecule, which, at this pH, implies that only one of them is charged]. While the presence of solvent and intervening protein atoms would, in general, screen the effects of charges and tend to reduce the likelihood of nearby charges affecting the ionizable residues, we are dealing here in a microenvironment and such an error (misassignment of charge state for a residue) could cause higher RMS deviations, not only for that particular residue but also for other charged residues due to the long-range effect of the electrostatic forces. The most significant effect is likely to involve histidines, because the pK_a for histidines (6.0) is closest to the pH of the crystal (4.1). For example, both His^{51} in the histidine loop and His^{108} in the aspartate loop have high RMS deviations, 4.24 and 1.84 Å, respectively, for the time-averaged structure of molecule 1 compared with the experimental structure. Both residues have two neighboring positively charged residues within 7 Å but no negatively charged residues or salt ions to compensate for their effect. We assumed that both histidines were protonated, given the acidic pH of the system.

In order to investigate whether the pK_a's may be shifted, we have employed Gurd's procedure (Shire et al., 1974) [which utilizes a Kirkwood–Tanford model of the dielectric (Tanford & Kirkwood, 1957)]. At pH 4 and 1 M ionic strength this method does indicate a large shift downward (1.6 pH units) in the pK_a of His^{108}, suggesting the possibility that it may, in fact, be uncharged in the crystal. That a downward shift in

Table VII: Average Potential Energies of Different Classes of Protein Residues, Water Molecules, and Counterions over Different 1-ps Segments of the Simulation

molecules/residues		average potential energies (kcal/mol) $\langle E\rangle_{-11-10}$[a] total	$\langle E\rangle_{-11-10}$[a] E/res[d]	$\langle E\rangle_{0-1}$[b] total	$\langle E\rangle_{0-1}$[b] E/res[d]	$\langle E\rangle_{59-60}$[c] total	$\langle E\rangle_{59-60}$[c] E/res[d]
neutral residues	molecule 1	3085.3	18.8	2700.5	16.5	2600.6	15.9
	molecule 2	3095.7	18.9	2677.4	16.3	2560.3	15.6
charged residues	molecule 1	−894.2	−52.6	−1253.5	−73.7	−1639.6	−96.4
	molecule 2	−840.0	−49.4	−1172.5	−69.0	−1490.8	−87.7
charged nearest neighbors	molecule 1	559.1	19.3	514.4	17.7	512.7	17.7
	molecule 2	560.8	19.3	512.7	17.7	512.6	17.7
selected residues[e]	molecule 1	−170.5	−17.1	−295.7	−29.6	−496.0	−49.6
	molecule 2	−130.2	−13.0	−273.2	−27.3	−455.5	−45.5
all residues	molecule 1	2191.1	12.1	1447.0	8.0	961.0	5.3
	molecule 2	2255.7	12.5	1504.9	8.3	1069.5	5.9
water		−12844.2	−9.0	−12879.0	−9.0	−12578.1	−8.8
sodium ions		−2047.9	−128.0	−2614.2	−163.4	−2713.6	−169.6
dihydrogen phosphate ions		−3049.8	−117.3	−4279.0	−164.6	−4766.3	−183.3
total		−13495.2	−7.4	−16820.3	−9.2	−18027.5	−9.8

[a] $\langle E\rangle_{-11-10}$ refers to the average potential energy over the first picosecond of the initial dynamics simulation during which the protein was held fixed at the structure obtained by minimizing the X-ray protein structure (together with the Monte Carlo solvent structure) with 100 steps of steepest descent. The average potential energies were corrected by the vibrational potential energy (see footnote *a* in Table VI). [b] $\langle E\rangle_{0-1}$ refers to the average potential energy over the first picosecond of the 60-ps dynamics simulation (during which the protein was allowed to move). [c] $\langle E\rangle_{59-60}$ refers to the average potential energy over the last picosecond of the 60-ps dynamics simulation (during which the protein was allowed to move). [d] These values are potential energies per residue (in the case of the protein), per water molecule or per ion. [e] These are the 10 residues with the largest RMS deviations (these residues are listed in Table VI).

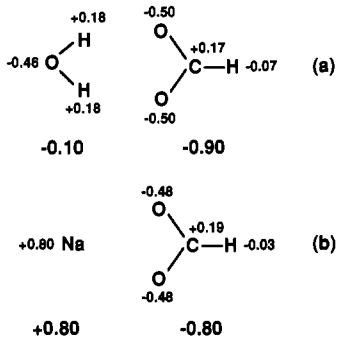

FIGURE 7: Effects of charge transfer on the charge on a formate ion in the presence of a water molecule (a) and a sodium ion (b). The ab initio calculations were performed with the program CADPAC (Amos & Rice, 1987) using an STO-3G basis set (Hehre et al., 1969), and the partial atomic charges were obtained by a Mulliken population analysis (Mulliken, 1955). Through the charge-transfer effect the total net charge on the formate ion is reduced to −0.90 in the presence of the water molecule, the oxygen of which is 3.1 Å from the carbon atom of the formate ion, and to −0.80 in the presence of the sodium ion, which is 2.8 Å from the carbon atom.

the pK_a of this residue is predicted is reasonable, in view of the fact, stated above, that this residue has a net positive environment within 7 Å. This will make it more difficult to protonate the histidine, leading to a lowering of the pK_a.

Overall Cause of Deviations. On the basis of these observations, one might speculate that the charged residues are responsible for the deviations in the structure, i.e., that the forces on these residues, which pull them into solvent or into the favorable environment, are responsible for the overall deviations. If this were the case, it would follow that the neutral residues would have strain induced in them, and concomitant higher energies, through the unfavorable states required to accommodate conformational changes of the charged residues. In Table VII we present the average potential energies of the various components of the system. These are broken down into the energies of different classes of residues, water molecules, and sodium and dihydrogen phosphate ions. The energies are further broken down into individual energies for the protein molecules 1 and 2. As can be seen, not only the charged residues, but all residues in the protein, decrease in energy during the simulation.

As can be seen from Table VII, the trajectory takes both protein molecules into significantly lower potential energy states. On average, all types of residues of the protein decrease in energy. It is not possible to conclude from this analysis whether or not a single type of interaction is primarily responsible for the deviation. Neutral residues decrease by roughly 3 kcal/mol per residue from the initial structure to the structure averaged over the last picosecond of the trajectory (~19 to 16 kcal/mol per residue). Most of the charged residues exhibit large decreases in energy as one would expect from the results, decreasing by ~41 kcal/mol per residue. The ions also show dramatic decreases, achieving more favorable environments by ~42 and 66 kcal/mol per ion for the sodium and dihydrogen phosphate ions, respectively. Again, we see that, in molecules 1 and 2, the decreases in energy are roughly equivalent (6.8 and 6.6 kcal/mol per residue for molecules 1 and 2, respectively) even though, as seen from the structural analysis, there are significant differences in structure. The total energy of molecules 1 and 2, however, does differ, with molecule 2 having a somewhat higher energy (5.9 kcal/mol per residue) than molecule 1 (5.3 kcal/mol per residue). Again, this might have been expected on the basis of the overall higher deviation of molecule 1.

We saw above that significant differences arose in the structures of molecules 1 and 2 during the simulation, and we now also see the difference in the two molecules in the asymmetric unit expressed in these average potential energies. These differences imply a problem in interpreting the results of any single simulation mechanistically, that is, in drawing conclusions as to what the underlying cause of deviations might be. That is, we see that different, but energetically similar regions of phase space are visited by the two equivalent protein molecules and that there are a variety of trajectories and states that residues can sample even over the course of a 60-ps simulation. Thus, one cannot ascribe a particular structural deviation to any particular cause (such as interaction with a solvent being too strong or interaction with a protein being too strong), as we see here that, in many cases, depending on the initial trajectory and given the limited time, very different states can be achieved by equivalent residues or structural

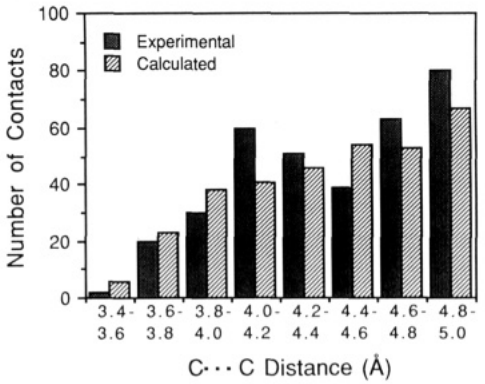

FIGURE 8: Distribution of aliphatic carbon–aliphatic carbon nonbond distances in the experimental and time-averaged structure (molecule 1) of SGPA. The experimental distribution shows a peak at 4.0–4.2 Å, due to nearest-neighbor contacts, that is not seen for the time-averaged structure. After this peak, the experimental distribution drops off and then rises again for the peak of second neighbors.

features in the two different molecules. This clearly implies a problematic situation in the attempt to interpret simulations of a single protein molecule. We might note that this is very likely representative of the actual behavior of proteins in solution where there is a balance between solvent forces and interactions with the protein and there are a variety of closely spaced "micro states" (Austin et al., 1975; Frauenfelder et al., 1979). This is why, in many protein crystals, a few regions are partially or fully disordered.

Radius of Gyration and Distribution of Interatomic Nonbonded Contact Distances. Analysis of the distribution of nonbonded contacts between different classes of atoms may highlight additional important features of the force field being used. In particular, there has been discussion in the literature as to the appropriate van der Waals parameters to use in dynamics studies (Wodak et al., 1984). The values in the current force field were obtained from fitting the geometry and sublimation energies of small molecule crystals (Lifson et al., 1979; Hagler et al., 1979; Dauber & Hagler, 1980), and the question has arisen as to whether van der Waals radii derived by minimization (at 0 K) would be intrinsically larger than the true radii in order to compensate for the effects of thermal expansion. If this were the case, this in turn would result in expansion of the protein as a whole during a dynamics simulation. In the present simulation, the protein molecules remain at approximately the experimental size, with radii of gyration of 14.7 and 14.6 Å, compared with the experimental value of 14.4 Å. Figure 8 shows the distribution of contact distances between aliphatic carbon atoms for the experimental structure and for the time-averaged structure of molecule 1. The small changes in the radii of gyration, as well as the similarity of the distributions, indicate that, at least at this level, there is no serious systematic error in the van der Waals radii derived from the crystal studies. This expansion is well within the error obtained in the fit of minimized crystal structures by a variety of force fields (Lifson et al., 1979; Hagler et al., 1979; Dauber & Hagler, 1980; T. Thacher and A. T. Hagler, to be submitted). One of the features of the experimental distribution that is not seen for the calculated structure is the peak at 4.0–4.2 Å, due to nearest-neighbor contacts. The experimental distribution falls off after this peak, before beginning to rise again for the peak of second neighbors.

Solvent Accessibility of the Simulated Protein Structure. Another aspect of the maintenance of the correct conformation during a simulation in which solvent is explicitly included, which depends critically on the potential having the correct balance between protein–protein and protein–water interactions for all groups in the protein, is the solvent accessibility of different groups in the protein. Comparison of the solvent accessibility of several different groups in the time-averaged protein structure and the X-ray structure provides insight into the extent to which this balance was achieved. Table VIII shows these data for the experimental structure and for the time-averaged structure of molecule 1. The largest change in accessibility is for the guanidinium groups of arginine residues, for which the exposure in the simulated structure is increased by more than 70% from the X-ray value. In contrast to this, the negatively charged groups have approximately the same exposure as in the experimental structure (although we should point out here the differences in behavior of charged groups in the two molecules—for example, as discussed above, Glu^{233} moves out into the solvent in molecule 2 during dynamics, whereas in molecule 1, it stays closer to the experimental position). Both main-chain carbonyl oxygens and the hydroxyl groups of serine become significantly less accessible, whereas the hydroxyl and the γ-carbon of threonine become more accessible. All of these perturbations, except that of the guanidiniums, are, however, sufficiently small that they may be manifestations of factors other than an imbalance in the protein–protein and protein–solvent interactions.

Table VIII: Accessibility Differences (Å^2) between the Simulated Average Structure of SGPA and the Experimental Crystal Structure for Different Types of Atomic Groups[a]

group	no. of atoms[b]	surface area (X-ray)	surface area (MD)[c]	difference [(MD) − (X-ray)]
main-chain N	181	206	218	12
main-chain C[d]	181	113	114	1
main-chain O	181	857	809	−48
Asp $O^{\delta 1}$, $O^{\delta 2}$	8	11	15	4
Glu $O^{\epsilon 1}$, $O^{\epsilon 2}$	4	72	63	−9
Arg $N^{\eta 1}$, $N^{\eta 2}$	14	203	350	147
His $N^{\delta 1}$, $N^{\epsilon 2}$	6	22	14	−8
Ser O^{γ}	21	248	191	−57
Thr $O^{\gamma 1}$, $C^{\gamma 2}$	42	389	508	119
Asn $O^{\delta 1}$, $N^{\delta 2}$	24	365	328	−37
Gln $O^{\epsilon 1}$, $N^{\epsilon 2}$	12	153	169	16
aliphatic C[e]	131	633	622	−11
aromatic side chains[f]	101	243	244	1

[a] The surface areas are summed over all atoms in the molecule that belong to each type of group. The accessibilities (surface areas) were calculated by the method of Lee and Richards (1971). [b] This is the total number of atoms in each group (e.g., there are 21 Thr residues per molecule and therefore a total of 42 Thr $O^{\gamma 1}$ and $C^{\gamma 2}$ atoms). [c] These data are for the time-averaged structure of molecule 1. [d] These are the carbonyl carbon atoms. [e] These include carbon atoms from C^{β} onward of Leu, Ile, Val, and Ala residues. [f] These include all heavy side-chain atoms from C^{β} onward of Tyr, Phe, and Trp residues, except for the O^{η} of Tyr residues.

Behavior of Hydrogen Bonds. Hydrogen bonds have long formed the core of the description of protein structure. They are responsible for both α helices and β sheets, both of which are described in terms of these interactions. The experimentally determined X-ray structure of SGPA at 1.5-Å resolution has 170 hydrogen bonds (defined as acceptor–hydrogen interatomic distances of less than 2.4 Å). Of these, 166 are intramolecular and 4 intermolecular. The intramolecular set is comprised of 104 between main-chain atoms, 42 between main chain and side chain, and 20 side-chain to side-chain hydrogen bonds.

Figure 9 shows the distribution of main-chain N to main-chain O distances in hydrogen-bonding interactions for the experimental structure and for the time-averaged structure of molecule 1. The experimental distribution has a large peak at 2.8–3.0 Å, while the simulation results in a broader peak with a maximum at 3.0–3.2 Å. As discussed below, a con-

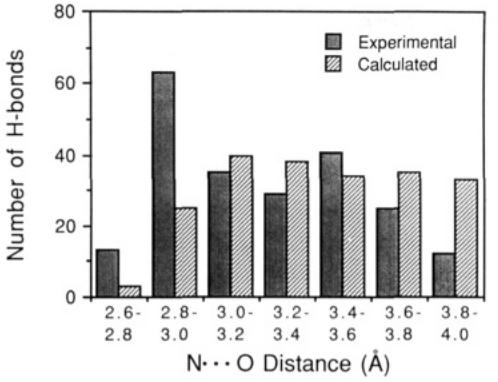

FIGURE 9: Distribution of hydrogen-bond lengths in the experimental and time-averaged structure (molecule 1) of SGPA.

siderable redistribution of hydrogen bonds took place during the simulation, and both the broadening and lengthening of the distribution reflect this activity. This difference in distribution is in contrast to the situation with the aliphatic carbon–carbon distribution (see above) where, although a shift in individual aliphatic interactions took place, the overall distribution of distances remained essentially the same.

A summary of the numbers of hydrogen bonds found in the experimental and simulated structures is given in Table IX. The hydrogen bonds are calculated from a series of time-averaged coordinates, each averaged over a 1-ps interval. A hydrogen bond is included in a dynamics structure if it occurred for more than 33% of the time in the molecule during the 16–60-ps portion of the simulation. A detailed table of all hydrogen bonds in the experimental protein and in the proteins during dynamics, organized according to the substructures defined in Table IV and giving the fraction of time each hydrogen bond existed during dynamics, is available as supplementary material.

A number of observations can be made from examination of the hydrogen-bonding network: (1) The total number of hydrogen bonds has decreased, particularly for interactions involving side-chain atoms. (2) Some new hydrogen bonds are being formed, so that a partial shift in structure is taking place, rather than denaturation. (3) More new bonds involving side chain than main chain are being made (relative to the number of bonds of each type). (4) The shifts in bonding pattern are quite different for the two protein molecules.

In the 40-ps dynamics simulation of the BPTI crystal (Berendsen et al., 1986), the hydrogen bonding was analyzed in the same way as described here, except that the criterion for a hydrogen bond included the requirement that the donor–hydrogen–acceptor angle had to be greater than 135° (the maximum hydrogen–acceptor distance was the same as used here, i.e., 2.4 Å). There are 23 main-chain–main-chain, 9 main-chain–side-chain, 1 side-chain–side-chain, and no intermolecular hydrogen bonds in the experimental BPTI structure. The number of hydrogen bonds formed during dynamics was somewhat higher than for SGPA; for example, 21, 22, 17, and 20 main-chain–main-chain hydrogen bonds were formed for each of the four molecules in the simulation (as above, hydrogen bonds have been included that were present for more than 33% of the simulation). The number of hydrogen bonds that were reproduced in all of the simulated molecules was similar to SGPA for main-chain–main-chain hydrogen bonds (12 out of 23), while for main-chain–side-chain hydrogen bonds it was somewhat higher (4 out of 9). We note here, however, that the SGPA simulation was longer than the BPTI simulation (see above). Since the deviations of simulated molecules from experimental continued to increase as a function of time in both systems, the two simulations would seem to have produced comparable results.

Table IX: Numbers of Hydrogen Bonds in the Experimental and Time-Averaged Structures of SGPA

system	main chain–main chain	main chain–side chain	side chain–side chain	intermolecular
exptl	104	42	20	4
dyn, molecule 1	87	33	16	3
dyn, molecule 2	82	32	14	
exptl and dyn 1[a]	73	15	6	
exptl and dyn 2[b]	66	19	8	
exptl, dyn 1, and dyn 2[c]	50	10	4	0

[a] These are hydrogen bonds that are found in both the experimental structure and the time-averaged structure of molecule 1. [b] These are hydrogen bonds that are found in both the experimental structure and the time-averaged structure of molecule 2. [c] These are hydrogen bonds that are found in the experimental structure and in the time-averaged structures of both molecules.

In the 210-ps simulation of BPTI in solution (Levitt & Sharon, 1988) it was found that 20 out of 21 main-chain–main-chain hydrogen bonds were preserved during dynamics. In this study, however, a different criterion was used for assessing the presence of a hydrogen bond, namely, an O–H distance of <2.6 Å, and an O–N–H angle of <35°. During the last 100 ps of dynamics, a hydrogen bond was considered present if it existed for more than 10% of the time. The main-chain–main-chain hydrogen bonding for the SGPA dynamics was recalculated from these criteria, and this showed that, out of the 105 hydrogen bonds in the experimental structure, 86 (82%) were preserved in molecule 1 and 84 (80%) in molecule 2. A total of 72 (69%) were preserved in both molecules (these are increases from 72%, 63%, and 48%, respectively, obtained from the alternative criteria described earlier). Thus, we see that the definition used for hydrogen bonding can have a significant effect on the results, and care must be taken in comparing the results given for different simulations.

A comparison of the hydrogen-bond patterns in the experimental and simulated structures with the RMS deviations of the substructures of the protein shows that there is, not unexpectedly, a correlation between the extent to which the hydrogen-bond pattern in a given region is maintained and the RMS difference between the various structural units and their experimental equivalents. Again we see that the fidelity of the fit of hydrogen bonds, as with overall structure, varies significantly from substructure to substructure. For example, as discussed above, the largest deviation for molecule 1 occurs for the histidine loop. This loop consists of two β strands connected by a turn and, as would be expected for this type of structure, includes numerous main-chain–main-chain hydrogen bonds. The large change in conformation of the loop is reflected in the almost total rearrangement of the hydrogen bonds—of the seven main-chain–main-chain hydrogen bonds that occur within this loop in the experimental structure, only one is retained and that one only 25% of the time. In contrast to this is the serine loop, which is the best-fitting substructure in molecule 1 and also contains seven main-chain–main-chain hydrogen bonds in the experimental structure. For molecule 1 all of these hydrogen bonds are retained during dynamics, five of them 100% of the time and the other two 79% and 93% of the time.

The newly formed hydrogen bonds show some systematic behavior as well. In particular, there are several C_7 ring hydrogen bonds created. Such bonds are known to form in

peptides in nonaqueous solutions, and the current force field also produces them in in vacuo simulations (Hagler et al., 1985; Kitson & Hagler, 1988). Most of the residues around which these C_7 hydrogen bonds have been formed are in nearby regions of ϕ,ψ space in the experimental structure and the difference in energy between these starting conformations and the C_7 forms is only approximately 1 or 2 kcal/mol or less (Stern et al., 1983). The appearance of such interactions in this simulation may indicate some subtle deficiency in the current potential (at issue is a conformational energy difference of $\sim$1 kcal/mol). The potential may need to be quite accurate in this region to distinguish between the two hydrogen-bonded geometries.

Thus, we see that significant changes in the numbers of hydrogen bonds and the patterns of hydrogen bonding have taken place during dynamics. The changes in patterns of hydrogen bonds can be correlated with the RMS deviations found for different substructures of the protein.

SUMMARY

As discussed in the introduction, simulations of proteins and other biomolecules are becoming increasingly important as a technique to explore the properties of these molecules. Also as mentioned, and as demonstrated here and in other studies, we have not yet reached the point where we can, on a routine basis, theoretically predict the properties of protein molecules within experimental accuracy. The ultimate achievement of this goal will give us a tool of considerable power. Here we have explored the nature and size of deviations that exist at the current level of the techniques and theory and have examined some of the areas we need to address in order to achieve the aforementioned goal. Even before we reach this goal, however, the tools provide us, as seen here, with the ability to investigate these systems at a level of atomic detail, in terms of interatomic forces and interactions, that is inaccessible to other techniques. When used along with experiment, they provide us with a framework to understand experiments, to rationalize experimental observables, and to design both experiments and the molecules themselves. It is important in using these tools to understand their capabilities and limitations, since we cannot expect to achieve experimental accuracy with each simulation. It is also important, if we are to improve these techniques to the point where we can achieve experimental accuracy, to characterize the deviations. This was one of the major goals of this study.

We have seen the importance of solvent and of counterions from the above results. From the energetic analysis we characterized the significant interaction energies of many of the charged residues with the counterions (dihydrogen phosphate and sodium in this case), accounting for the ability of these residues to extend out into solution.

Several patterns emerged from the simulation. One of them was the differences that arose in the two protein molecules that were simulated. These molecules are identical and related by symmetry (a 2-fold axis) in the crystal. Here we treated them independently. In principle, we should have obtained the same results for both molecules. From the experiment we know that their time-averaged environments are the same. However, the initial configuration of the solvent (taken from the Monte Carlo simulation) was asymmetric, and the two proteins were each given a different initial Boltzmann distribution of velocities at the beginning of the simulation, since, on an instantaneous basis, even in the experimental situation, they will not maintain the time-averaged symmetry. In addition, the instantaneous configuration of the solvent also will not reflect the time-averaged symmetry. Instantaneous deviations in the structures are, therefore, expected. It was very important and interesting that the time-averaged properties of two molecules differed over the course of the simulation. As discussed above, this is important insofar as it says that, in a given simulation of the order of 60–100 ps (60 ps was achieved here), we cannot necessarily interpret the deviations from experiment on a mechanistic basis. What we saw in the case of the two molecules analyzed here is that, in some cases, deviations were caused by side chains that moved out into the ionic solution whereas, in other cases, the corresponding side chains achieved a similar energy by moving into the protein. Not surprisingly, we found that there are many favorable environments for residues and that they can achieve favorable energies in a variety of ways.

These results also point out the problem in attempting, at this stage, to interpret thermodynamic data too rigorously, especially without a structural analysis. We have achieved, in this simulation, a structure characterized by a 1.2–1.5-Å RMS deviation, which is typical of current simulations. Within these structures there are some significant structural fluctuations and deviations from the experimental structure. If we want to account for thermodynamic processes, for example, the free energy of interaction of a substrate with an enzyme or of a mutation, we need to make sure that the free energy corresponds to the interaction or reaction of interest. If the important structural region of the protein deviates by 2 or 3 Å, which can happen, as seen in the 1.5-Å overall RMS structure, this may well change the interactions of interest. Obviously, a 3-Å change in structure can cause significant deviations in hydrogen bonds (for example, a 3-Å hydrogen bond could become a 5–6-Å hydrogen bond). Thus, if we calculate the free energy of such an interaction, we are actually calculating the free energy of entirely different interactions than occur in the real system. Therefore, structural features of the interaction for which we are calculating the free energy should be considered. We also note that, even after 60 ps, we are continuing to get a migration of the structure. This is important, since, in free energy calculations, we need to calculate all the accessible states that accompany a perturbation in some structure. It implies that the windows that we use for perturbation theory may need to be of the order of 40 or 50 ps, rather than of the order of 1 ps as is now taken in order to allow for equilibration. The large local structural deviations that underlie an overall 1.5-Å RMS deviation also show the difficulty in attempting to use this type of calculation to predict the conformational consequences of site-specific mutagenesis experiments and, as stated above, to account for the free energy changes associated with the mutations.

The deviations that we have found allow us to go on to use these results as a basis for assessing further improvements in potential functions. They also allow us to interpret a variety of experimental data. The simulations, even at this level, allow us, when combined with the experiment, to propose new interpretations of observed experimental data at the molecular level that was previously ambiguous. This is especially so in the case of a rather nonintuitive result where we see a cluster of dihydrogen phosphate ions in a solvent-filled cavity within the simulated crystal structure. This was found to be caused by favorable ion–water interactions and by focusing of the large positive charge of each of the SGPA molecules. This has been related to peaks in the experimental electron density. Further experimental studies are being carried out to test the significance of these peaks and this tentative interpretation of their origin. This work will be reported elsewhere (F. Avbelj, J. Moult, D. H. Kitson, M. N. G. James, D. Hadzi,

and A. T. Hagler, to be submitted). Thus, we can characterize the deviations and use these as a basis for the further improvement in potential functions that is required if we are going ultimately to achieve experimental accuracy and, at the same time, use the results along with experiment to probe the behavior of these systems at an ever finer resolution down to the level of atomic and molecular interactions. In this way, the results reported here form a basis for improvements that we need to make in order to approach our ultimate goal of carrying out simulations within experimental accuracy.

Acknowledgments

We thank Biosym Technologies, Inc., San Diego, CA, who provided the molecular modeling software INSIGHT and DISCOVER (which was vectorized and adapted for these purposes). We thank La Vonne Melheim for her expert help in the preparation of the manuscript. We also thank Dr. Harry Anderson, Dr. Don Mackay, Dr. Jürgen Bajorath, Berith Bjornholm, and Dan Jones for helpful discussions and Patricia Deignan and Joni Stern for their assistance in the preparation of the manuscript.

Supplementary Material Available

A detailed table of all hydrogen bonds in the experimental protein and in the proteins during dynamics, organized according to the substructural elements of the protein and giving the fraction of time each hydrogen bond existed during dynamics (6 pages). Ordering information is given on any current masthead page.

Registry No. Serine protease, 37259-58-8; water, 7732-18-5; sodium, 7440-23-5; dihydrogen phosphate, 14066-20-7.

References

Allen, F. H., Bellard, S., Brice, M. D., Cartwright, B. A., Doubleday, A., Higgs, A., Hummelink, T., Hummelink-Peters, B. G., Kennard, O., Motherwell, W. D. S., Rodgers, J. R., & Watson, D. G. (1979) *Acta Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem. B35*, 2331–2339.

Amos, R. D., & Rice, J. E. (1987) *CADPAC: The Cambridge Analytical Derivatives Package*, issue 4.0, Cambridge.

Åqvist, J., van Gunsteren, W. F., Leijonmarck, M., & Tapia, O. (1985) *J. Mol. Biol. 183*, 461–477.

Åqvist, J., Sandblom, P., Jones, T. A., Newcomer, M. E., van Gunsteren, W. F., & Tapia, O. (1986) *J. Mol. Biol. 192*, 593–604.

Ashida, T., Tsunogae, Y., Tanaka, I., & Yamane, T. (1987) *Acta Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem. B43*, 212–218.

Austin, R. H., Beeson, K. W., Eisenstein, L., Frauenfelder, H., & Gunsalus, I. C. (1975) *Biochemistry 14*, 5355–5373.

Bash, P. A., Singh, U. C., Brown, F. K., Langridge, R., & Kollman, P. A. (1987) *Science (Washington, D.C.) 235*, 574–575.

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., & Hermans, J. (1981) in *Intermolecular Forces. Proceedings of the Fourteenth Jerusalem Symposium on Quantum Chemistry and Biochemistry* (Pullman, B., Ed.) pp 331–342, D. Reidel Publishing Co., Dordrecht, The Netherlands.

Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., & Haak, J. R. (1984) *J. Chem. Phys. 81*, 3684–3690.

Berendsen, H. J. C., van Gunsteren, W. F., Zwinderman, H. R. J., & Geurtsen, R. G. (1986) *Ann. N.Y. Acad. Sci. 482*, 269–286.

Birktoft, J. J., & Blow, D. M. (1972) *J. Mol. Biol. 68*, 187–240.

Blevins, R. A., & Tulinsky, A. (1985) *J. Biol. Chem. 260*, 4264–4275.

Brayer, G. D., Delbaere, L. T. J., & James, M. N. G. (1978) *J. Mol. Biol. 124*, 261–283.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., & Karplus, M. (1983) *J. Comput. Chem. 4*, 187–217.

Brunger, A. T., Kuriyan, J., & Karplus, M. (1987) *Science (Washington, D.C.) 235*, 458–460.

Chambers, J. L., & Stroud, R. M. (1979) *Acta Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem. B35*, 1861–1874.

Clore, G. M., Nilges, M., Sukumaran, D. K., Brunger, A. T., Karplus, M., & Gronenborn, A. M. (1986) *EMBO J. 5*, 2729–2735.

Dauber, P., & Hagler, A. T. (1980) *Acc. Chem. Res. 13*, 105–112.

Dauber, P., Osguthorpe, D. J., & Hagler, A. T. (1982) *Biochem. Soc. Trans. 10*, 312–318.

Dauber-Osguthorpe, P., Roberts, V. A., Osguthorpe, D. J., Wolff, J., Genest, M., & Hagler, A. T. (1988) *Proteins: Struct., Funct., Genet. 4*, 31–47.

Frauenfelder, H., Petsko, G. A., & Tsernoglou, D. (1979) *Nature (London) 280*, 558–563.

Hagler, A. T. (1985) in *The Peptides: Analysis, Synthesis, Biology* (Undenfriend, S., & Meienhofer, J., Eds.) Vol. 7, pp 213–299, Academic, Orlando, FL.

Hagler, A. T., & Moult, J. (1978) *Nature (London) 272*, 222–226.

Hagler, A. T., Lifson, S., & Dauber, P. (1979) *J. Am. Chem. Soc. 101*, 5122–5130.

Hagler, A. T., Moult, J., & Osguthorpe, D. (1980) *Biopolymers 19*, 395–418.

Hagler, A. T., Osguthorpe, D. J., Dauber-Osguthorpe, P., & Hempel, J. C. (1985) *Science (Washington, D.C.) 227*, 1309–1315.

Haneef, I., Glover, I. D., Tickle, I. J., Moss, D. S., Pitts, J. E., Wood, S. P., Blundell, T. L., Hermans, J., & van Gunsteren, W. F. (1985) in *Molecular Dynamics and Protein Structure* (Hermans, J., Ed.) pp 85–91, University of North Carolina, Chapel Hill, NC.

Hehre, W. J., Stewart, R. F., & Pople, J. A. (1969) *J. Chem. Phys. 51*, 2657–2664.

Hockney, R. W., & Eastwood, J. W. (1981) *Computer Simulation Using Particles*, McGraw-Hill, New York, NY.

Ichiye, T., Olafson, B. D., Swaminathan, S., & Karplus, M. (1986) *Biopolymers 25*, 1909–1937.

Karplus, M., & McCammon, J. A. (1981) *CRC Crit. Rev. Biochem. 9*, 293–349.

Kasper, J. S., & Lonsdale, K., Eds. (1967) *International Tables for X-ray Crystallography*, Vol. 2, Kynoch Press, Birmingham, England.

Kitson, D. H., & Hagler, A. T. (1988) *Biochemistry 27*, 5246–5257.

Kuriyan, J., Petsko, G. A., Levy, R. M., & Karplus, M. (1986) *J. Mol. Biol. 190*, 227–254.

Lee, B., & Richards, F. M. (1971) *J. Mol. Biol. 55*, 379–400.

Levitt, M. (1983a) *J. Mol. Biol. 168*, 621–657.

Levitt, M. (1983b) *J. Mol. Biol. 168*, 595–620.

Levitt, M., & Warshel, A. (1975) *Nature (London) 253*, 694–698.

Levitt, M., & Sharon, R. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 7557–7561.

Lifson, S., Hagler, A. T., & Dauber, P. (1979) *J. Am. Chem. Soc. 101*, 5111–5121.
Locke, M. J., & McIver, R. T., Jr. (1983) *J. Am. Chem. Soc. 105*, 4226–4232.
Mackay, D. H. J., Cross, A. J., & Hagler, A. T. (1989) in *Prediction of Protein Structure and the Principles of Protein Conformation* (Fasman, G. D., Ed.) pp 317–358, Plenum Publishing Corp., New York, NY.
Mezei, M., Swaminathan, S., & Beveridge, D. L. (1979) *J. Chem. Phys. 71*, 3366–3373.
Mulliken, R. S. (1955) *J. Chem. Phys. 23*, 1833–1840.
Northrup, S. H., Pear, M. R., Morgan, J. D., & McCammon, J. A. (1981) *J. Mol. Biol. 153*, 1087–1109.
Post, C. B., Brooks, B. R., Karplus, M., Dobson, C. M., Artymiuk, P. J., Cheetham, J. C., & Phillips, D. C. (1986) *J. Mol. Biol. 190*, 455–479.
Read, R. J. (1986) *Acta Crystallogr., Sect. A A42*, 140–149.
Richardson, J. S. (1981) in *Advances in Protein Chemistry* (Anfinsen, C. G., Edsall, J. T., & Richards, F. M., Eds.) pp 167–339, Academic, New York, NY.
Russell, S. T., & Warshel, A. (1985) *J. Mol. Biol. 185*, 389–404.
Scanlon, J. W., & Eisenberg, D. (1975) *J. Mol. Biol. 98*, 485–502.
Shire, S. J., Hanania, G. I. H., & Gurd, F. R. N. (1974) *Biochemistry 13*, 2967–2974.
Shotton, D. M., & Watson, H. C. (1970) *Nature (London) 225*, 811–816.
Sielecki, A. R., Hendrickson, W. A., Broughton, C. G., Delbaere, L. T. J., Brayer, G. D., & James, M. N. G. (1979) *J. Mol. Biol. 134*, 781–804.
Smith, J. L., Hendrickson, W. A., Honzatko, R. B., & Sheriff, S. (1986) *Biochemistry 25*, 5018–5027.
Stern, P. S., Chorev, M., Goodman, M., & Hagler, A. T. (1983) *Biopolymers 22*, 1885–1900.
Streett, W. B., Tildesley, D. J., & Saville, G. (1978) *Mol. Phys. 35*, 639–648.
Struthers, R. S., Rivier, J., & Hagler, A. T. (1984) in *Conformationally Directed Drug Design: Peptides and Nucleic Acids as Templates or Targets* (Vida, J. A., & Gordon, M., Eds.) pp 239–261, ACS Symposium Series, American Chemical Society, Washington, DC.
Svennson, L. A., Sjolin, L., Gilliland, G. L., Finzel, B. C., & Wlodawer, A. (1986) *Proteins: Struct. Funct. Genet. 1*, 370–375.
Swope, W. C., Andersen, H. C., Berens, P. H., & Wilson, K. R. (1982) *J. Chem. Phys. 76*, 637–649.
Tanford, C., & Kirkwood, J. G. (1957) *J. Am. Chem. Soc. 79*, 5333–5339.
Tsukada, H., & Blow, D. M. (1985) *J. Mol. Biol. 184*, 703–711.
van Gunsteren, W. F., & Karplus, M. (1981) *Nature (London) 293*, 677–678.
van Gunsteren, W. F., Berendsen, H. J. C., Hermans, J., Hol, W. G. J., & Postma, J. P. M. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 4315–4319.
Warshel, A., & Levitt, M. (1976) *J. Mol. Biol. 103*, 227–249.
Warshel, A., & Russell, S. T. (1984) *Q. Rev. Biophys. 17*, 283–422.
Warshel, A., & Sussman, F. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 3806–3810.
Wlodawer, A., Borkakoti, N., Moss, D. S., & Howelin, B. (1986) *Acta Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem. B42*, 379–387.
Wodak, S. J., Alard, P., Delhaise, P., & Renneboog-Squilbin, C. (1984) *J. Mol. Biol. 181*, 317–322.