PROTEINS: Structure, Function, and Genetics 24:304–313 (1996)

Conformational Sampling by NMR Solution Structures Calculated With the Program DIANA Evaluated by Comparison With Long-Time Molecular Dynamics Calculations in Explicit Water

Kurt D. Berndt, Peter Güntert, and Kurt Wüthrich
Institut für Molekularbiologie und Biophysik, Eidgenössische Technische Hochschule-Hönggerberg, CH-8093 Zürich, Switzerland

ABSTRACT **The NMR solution structure of bovine pancreatic trypsin inhibitor (BPTI) obtained by distance geometry calculations with the program DIANA is compared with groups of conformers generated by molecular dynamics (MD) simulations in explicit water at ambient temperature and pressure. The MD simulations started from a single conformer and were free or restrained either by the experimental NOE distance restraints or by time-averaged restraints; the groups of conformers were collected either in 10 ps intervals during 200 ps periods of simulation, or in 50 ps intervals during a 1 ns period of simulation. Overall, these comparisons show that the standard protein structure determination protocol with the program DIANA provides a picture of the protein structure that is in agreement with MD simulations using "realistic" potential functions over a nanosecond timescale. For well-constrained molecular regions there is a trend in the free MD simulation of duration 1 ns that the sampling of the conformation space is slightly increased relative to the DIANA calculations. In contrast, for surface-exposed side-chains that are less extensively constrained by the NMR data, the DIANA conformers tend to sample larger regions of conformational space than conformers selected from any of the MD trajectories. Additional insights into the behavior of surface side-chains come from comparison of the MD runs of 200 ps or 1 ns duration. In this time range the sampling of conformation space by the protein surface depends strongly on the length of the simulation, which indicates that significant side-chain transitions occur on the nanosecond timescale and that much longer simulations will be needed to obtain statistically significant data on side-chain dynamics.**

© 1996 Wiley-Liss, Inc.

Key words: protein structure determination by NMR, distance geometry, structure calculation with DIANA, molecular dynamics simulation

© 1996 WILEY-LISS, INC.

INTRODUCTION

It has by now been well established that several different methods for computation of three-dimensional protein structures from nuclear magnetic resonance (NMR) data in solution generate coinciding molecular geometries. Nonetheless, experience during the last few years has also shown that the results of corresponding structure calculations with different techniques may differ with regard to the sampling of conformation space manifested by the root mean square distances (RMSD) among the groups of conformers used to represent the solution structures. As a contribution to the discussion on the significance of these observations, this paper evaluates NMR structures obtained with one of the common techniques, DIANA,[1] in light of long-time molecular dynamics (MD) simulations using near-physiological conditions of solvent, temperature, and pressure.

In the NMR method for the determination of three-dimensional biomacromolecular structures in solution,[2] the key experimental data represent a network of distance restraints between pairs of spatially proximate hydrogen atoms or groups of equivalent hydrogen atoms. The methods used for calculation of the structures conduct a search for a set of conformers that describe the molecule in solution; in practice, these conformers represent a subset of all conformations consistent with the experimentally determined restraints. Currently, two methods are predominantly used: 1) generation of the desired conformers with variable target function calculations,[3] using the program DIANA,[1,4] where the DIANA conformers can then be further refined by energy minimization[5]; and 2) simulated annealing using molecular dynamics (MD), for example, with the programs GROMOS[6] or X-PLOR,[7,8] where the

Received August 8, 1995; accepted August 11, 1995.
Address reprint requests to Kurt Wüthrich, Institut für Molekularbiologie und Biophysik, Eidgenössische Technische Hochschule-Hönggerberg, CH-8093 Zürich, Switzerland.
Kurt D. Berndt's present address is Institutionen för Medicinsk Biochemi och Biofysik, Karolinska Institutet, S-17177 Stockholm, Sweden.

MD calculation may be preceded either by interactive modeling or by embedding using a metric matrix distance geometry program.[9]

We used the protein basic pancreatic trypsin inhibitor (BPTI) for the present investigation. High-quality three-dimensional structures of this protein have been determined by X-ray crystallography,[10,11] by a combination of X-ray and neutron diffraction,[12] and by NMR in solution.[13] The NMR solution structure calculated with the program DIANA is represented by a group of 20 energy-refined conformers, which were selected in the conventional way[1,4] from a larger set of DIANA conformers. DIANA conformers are calculated in vacuo with a simplified force field that is formulated in torsion angle space and includes only the most essential part of the non-bonding interatomic interactions, namely, the steric repulsion. They reflect, albeit indirectly through the underlying NMR data, effects from all NMR-accessible timescales. Since subsequent local optimization of the conformational energy does not significantly alter the molecular geometry,[14] the sampling of conformation space by the above group of 20 conformers is predominantly determined by the DIANA calculations. One of these energy-minimized DIANA conformers was now subjected to a 1.4 ns free MD simulation, and a 0.8 ns MD simulation using the experimental NOE distance restraints.[15,16] In these MD simulations care was taken to represent the protein/solvent system as realistically as possible within the framework of current MD calculations. This includes the use of the force field of the GROMOS program,[17] explicit treatment of a large volume of solvent water molecules, and use of physiological temperature and pressure conditions. The resulting MD trajectories are expected to give a faithful picture of real protein dynamics on the timescales covered by the simulation.

From the practical point of view of a high-quality protein structure determination by NMR, it is a necessity for the spectroscopist to assess the results of the spectral analysis as directly as possible in terms of the three-dimensional structure. This is particularly obvious for the exhaustive assignment of NOESY cross peaks, which relies on reference to preliminary three-dimensional structures to resolve ambiguities caused by chemical shift degeneracies and peak overlap.[18] Considering that DIANA ignores many details of the physical energy function and requires two to three orders of magnitude less computation time than an MD simulation on the nanosecond timescale, it is therefore of interest with regard to practical aspects of structure determination as well as evaluation of the final result whether the standard DIANA protocol generates groups of conformers that portray a view of the solution structure that is in agreement with "realistic," long-time MD simulations. For a proper assessment of the results, one should remember that the standard protocols for the DIANA calculations and the MD simulations with GROMOS differ in both the force fields *and* the search algorithms used (random setting of initial torsion angles followed by variable target function minimization in DIANA versus generation in GROMOS of a continuous trajectory according to Newton's equations of motion starting from a single, already correctly folded conformer). The present comparisons therefore do not attempt to assess differences either between the two force fields or between the two search algorithms used in the two programs. The "realistic" MD simulations used for this study also differ fundamentally from the MD calculations typically used for a NMR structure determination by the simulated annealing method,[8] since the simulated annealing is performed at high temperature using a simplified force field that treats the atoms as soft spheres without attractive or long-range non-bonded interactions, and that does not include explicit consideration of the solvent.

MATERIALS AND METHODS

NMR Structure Calculation Using the Program DIANA

The NMR solution structure of BPTI was calculated as described by Berndt et al.[13] on the basis of an experimentally determined set of 642 upper distance limits and 115 torsion angle restraints, following the standard protocol for protein structure calculations with the program DIANA and its supporting programs.[5] Briefly, starting from 50 independent conformers with randomized torsion angles and employing the REDAC strategy for improved convergence,[4] the distance geometry program DIANA was used to generate a set of 50 conformers, from which the 20 conformers with the lowest residual values of the DIANA target function were further refined by restrained energy minimization with a modified version of the program AMBER.[19] The restrained energy minimization lowered the conformational energies of the DIANA conformers significantly, with only very slight changes of the molecular geometry, i.e., the sampling of conformation space as determined by the DIANA calculation was hardly affected during restrained energy minimization[13,14] and is therefore preserved in the 20 energy-refined DIANA conformers used to represent the solution structure.

Molecular Dynamics Simulations With GROMOS and Selection of Representative Data Sets

A detailed description of the protocol used for the MD simulations with the program GROMOS[17] was given elsewhere[15,16] and is outlined in Figure 1. Briefly, one of the conformers that represent the solution structure of BPTI was immersed in a truncated octahedral box of 2,371 equilibrated SPC/E

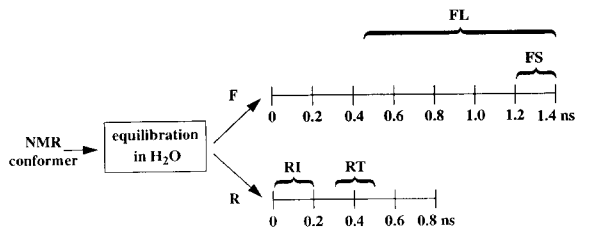

Fig. 1. Scheme illustrating the design of the two different molecular dynamics simulations discussed in this paper. F is a free MD run of 1.4 ns duration. R is a MD run of 0.8 ns duration consisting of three different phases. During the first 200 ps it included all 642 experimental NOE upper distance restraints[13]; thereafter it was constrained by time-averaged values of these restraints for 300 ps; finally, there were 300 ps of free MD simulation. Representative sets of 20 conformers each (FS, RI, and RT) were chosen from 200 ps intervals (indicated with braces) by selecting one conformer every 10 ps. An additional set of 20 conformers, FL, was selected from F at 50 ps intervals between 0.45 and 1.4 ns.

water molecules,[20] the BPTI–water system was equilibrated during 10 ps, and two separate simulations were performed: a free molecular dynamics simulation of 1.4 ns duration ("F"), and a restrained molecular dynamics simulation of 0.8 ns duration ("R") that was divided into three shorter time periods. The first 0.2 ns of the restrained simulation included all 642 NMR upper distance limits as conventional, instantaneous restraints. During the subsequent 0.3 ns of simulation time, the same NOE upper distance limits were imposed as time-averaged restraints[21,22] with an averaging time constant of 10 ps. Finally, 0.3 ns of unrestrained dynamics were applied, which are not analyzed in the present paper. Torsion angle restraints were not applied during the MD simulations. In all simulations, temperature and pressure were maintained at 277 K and 1 bar, respectively. From these trajectories, four different sets of 20 conformers were chosen at equidistant time points for further analysis and comparison with the NMR solution structure (Fig. 1): FL ("free, long"), covering 1 ns of free MD simulations; FS ("free, short"), covering the last 200 ps of the free MD simulation; RI ("restrained, instantaneous"), covering 200 ps of MD simulation with conventional NOE distance restraints; and RT ("restrained, time-averaged"), covering the last 200 ps of a 300 ps MD simulation with time-averaged NOE distance restraints, which started from the structure at the end of the RI simulation.

Analysis of Conformers

For visual comparison of different groups of conformers, stereo views were produced with the structure analysis program XAM.[23] Optimal global superpositions and RMSD values for various subsets of atoms were computed as usual.[24] Mean conformations were obtained by first superimposing the 20 conformers comprising a selected set so as to minimize the RMSD value for the backbone atoms N, C^{α}, and C′ of the residues 2–56, and then averaging the Cartesian coordinates of corresponding atoms in the 20 globally superimposed conformers. Displacements, D,[25] were used to quantify the local precision of the set of conformers and to identify local differences between the different sets of conformers. Displacements are a generalization of the conventional RMSD values, since the set of atoms used for the superposition of the conformers, M_{sup}, differs from the set of atoms for which the RMSD of the positions is actually calculated, M_{RMSD}. For the evaluation of the backbone displacement for a given residue i in BPTI after global superposition, D_{bb}^{glob}, M_{sup} consists of the backbone atoms N, C_{α}, and C′ of the residues 2–56, and M_{RMSD} includes the backbone atoms N, C^{α}, and C′ of residue i. For the local side-chain displacement of residue i, D_{sc}^{loc}, M_{sup} consists of the backbone atoms N, C^{α} and C′ of the tripeptide segment $i-1$, i, $i+1$, and M_{RMSD} contains the side-chain heavy atoms of residue i.

As a reference, local side-chain displacements for the completely disordered states of the different common amino acid side-chains were obtained by generating with the program DIANA 20 conformers containing the backbone dihedral angles of the DIANA conformers representing the NMR solution structure of BPTI[13] and the side-chain dihedral angle values selected as uniformly distributed, independent random variables. These 20 conformers were subjected to DIANA minimization with the backbone dihedral angles fixed and including only intraresidual steric repulsion in the target function. The resulting average of the pairwise local side-chain displacements was used as an estimate of D_{sc}^{loc} for the completely disordered side-chains.

A hydrogen bond was identified when the proton–acceptor distance was less than 2.4 Å and the angle between the donor–proton bond and the line connecting the acceptor and donor heavy atoms less than 35°. The average fraction of the total solvent-accessible surface area per residue was calculated using the algorithm of Richmond[26] as implemented in the program XAM.[23]

RESULTS AND DISCUSSION

Five different sets of 20 BPTI conformers each were selected for assessing the influence of various factors on the sampling of conformation space by these groups of conformers. The set NMR consists of 20 conformers that were obtained with 20 different calculations from the NMR data with the program DIANA, where each calculation started from a different, randomized structure for every conformer. The sets FL, FS, RI, and RT were taken at regular time intervals from MD trajectories that evolved from a single starting structure. In addition to comparing NMR with the different MD simulation re-

TABLE I. Average of the Pairwise RMSD Values Between the X-Ray Crystal Structure 5PTI and Selected Groups of Conformers Obtained Either From NMR Data Using the Program DIANA or From MD Simulations Without or With NMR Restraints

	RMSD [Å]*					
	NMR	FL	FS	RI	RT	5PTI
NMR	0.62	1.53	1.68	0.89	0.94	0.88
FL		1.00	0.99	1.31	1.31	1.34
FS			0.74	1.53	1.51	1.48
RI				0.60	0.73	0.86
RT					0.69	0.90

*The RMSD values were calculated for the backbone atoms N, C^{α}, and C′ of residues 2–56. NMR, 20 conformers representing the NMR solution structure of BPTI (1PIT)[13]; FL, 20 conformers collected over a period of 1.0 ns during a free molecular dynamics simulation (see text and Fig. 1 for details); FS, 20 conformers collected over a period of 200 ps during a free molecular dynamics simulation; RI, 20 conformers collected over a period of 200 ps during a restrained molecular dynamics simulation using conventional NMR distance restraints; RT, 20 conformers collected over a period of 200 ps during a restrained molecular dynamics simulation using time-averaged NMR distance restraints; 5PTI, X-ray crystal structure of BPTI form II.[11]

sults, comparisons are also made between conformers sampled over more or less extensive periods of the same MD trajectory (FL and FS), between conformers generated by free and restrained MD simulations (FL and FS versus RI and RT), and between conformers obtained with different interpretations of the NOE distance restraints in the MD simulation (RI versus RT). In all these comparisons, we will only briefly comment on the behavior of the generally well-defined backbone, and then focus attention on the amino acid side-chains of the protein.

Sampling of Conformation Space by the Polypeptide Backbone

Quantitative global differences in the polypeptide backbone conformations between the experimental NMR solution structure and the four sets of conformers from the MD simulations (Fig. 1) are afforded by Table I, where average pairwise RMSD values for comparison of the backbone atoms N, C^{α}, and C′ after superposition of residues 2–56 are tabulated. It is readily apparent that the 20 NMR conformers and the three sets of MD conformers sampled over 200 ps exhibit similar RMSD values in the range from 0.60 to 0.74 Å, whereas the conformers sampled over 1.0 ns during the free MD simulation show a slightly larger RMSD value of 1.0 Å. The 20 NMR conformers and the two sets of MD conformers RI and RT from the restrained MD simulation (Fig. 1) are significantly more similar to each other (RMSD values of 0.73–0.94 Å) than to the two sets of MD conformers taken from the free simulation (RMSD values between 1.31 and 1.68 Å), manifesting the anticipated influence of the NMR-derived restraints. Figures 2B and 3 illustrate that the backbone conformation is defined with nearly identical precision by the four sets of conformers NMR, FS, RI, and RT, and also show that in comparison the dispersion among the 20 conformers in the group FL is slightly increased. Note also that the *relative* precision of determination of the backbone atoms along the sequence in set FL is very similar to that of the NMR conformers (Fig. 2B), the only exceptions being residues 3–7, 16–17, 38, and 41, which display larger relative displacements in FL. The precision within the sets FL and FS is significantly different, except for the residues 24–28 in the loop between the two strands of the β-hairpin. The similarity in the segment 24–28 would suggest that the disorder seen for this segment is governed largely by motions faster than 200 ps. It is also noteworthy that the sets RI and RT display significantly less conformational variation along the backbone than the conformers from the long-time free MD simulation.

The effect of the length of the MD trajectory from which the conformers are sampled is manifested in the fact that the average pairwise RMSD value of 0.74 Å calculated among the conformers FS is significantly lower than for FL (Table I), showing that structures sampled over 200 ps contain less conformational variation than structures sampled over 1,000 ps during the same MD trajectory. On the time scale considered here, the conformational variations observed in free MD simulations are thus still related to the length of time during which a predetermined number of conformers are sampled. This is an important observation in view of the fact that MD simulations of proteins in water have until now rarely exceeded 500 ps. In the C-terminal dipeptide segment, which is the only part of the polypeptide backbone for which the NMR conformers display considerable disorder, the MD simulation FL represents the closest approach to the set of NMR conformers.

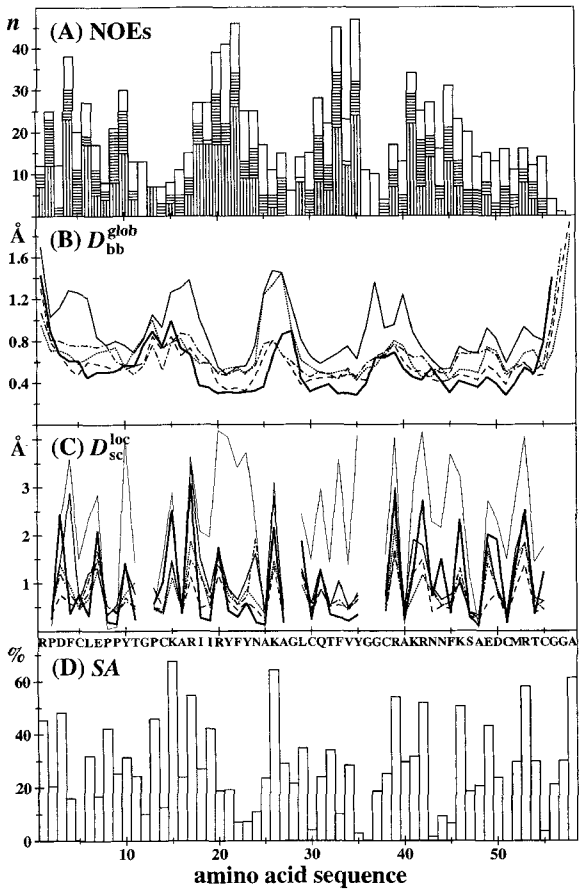

Fig. 2. Plots versus the amino acid sequence of BPTI. **A:** NOE upper distance restraints per residue, *n*, used for the solution structure determination.[13] Vertically hatched bars, restraints involving a peripheral (beyond βCH_n) side-chain atom of residue *i*; horizontally hatched bars, NOEs involving β side-chain atoms of residue *i*; open bars, NOEs involving backbone hydrogens of residue *i*. **B:** Displacements, D_{bb}^{glob}, of the backbone atoms N, C^{α}, and C′ of residue *i* after global superimposition of the backbone atoms N, C^{α}, and C′ of residues 2–56. Thick solid line, average of the pairwise displacements for the 20 energy-minimized DIANA conformers used to represent the NMR solution structure[13]; medium solid line, same for the 20 conformers FL; dotted line, same for the 20 conformers FS; dashed line, same for the 20 conformers of set RI; dot-dashed line, same for the 20 conformers RT (see Fig. 1 for the definition of the different sets of MD conformers). **C:** Local displacement, D_{sc}^{loc}, of all side-chain heavy atoms of residue *i* after superimposition of the backbone atoms N, C^{α}, and C′ of residues $i-1$, *i*, and $i+1$ for minimal pairwise RMSD. Same code as in B. In addition, the thin solid line indicates the expected values of D_{sc}^{loc} for the completely disordered side-chains (see text for details). **D:** Average of the solvent-accessible part of the total surface area per residue, SA, in the 20 energy-minimized DIANA conformers as calculated by the program XAM[23] using the algorithm of Richmond.[26]

Sampling of Conformation Space by the Amino Acid Side-Chains

In order to give an overall impression of side-chain conformational variability, Figure 4 presents comparisons of all-heavy-atom representations of the five sets of conformers in identical orientations. The precision of the determination of the amino acid

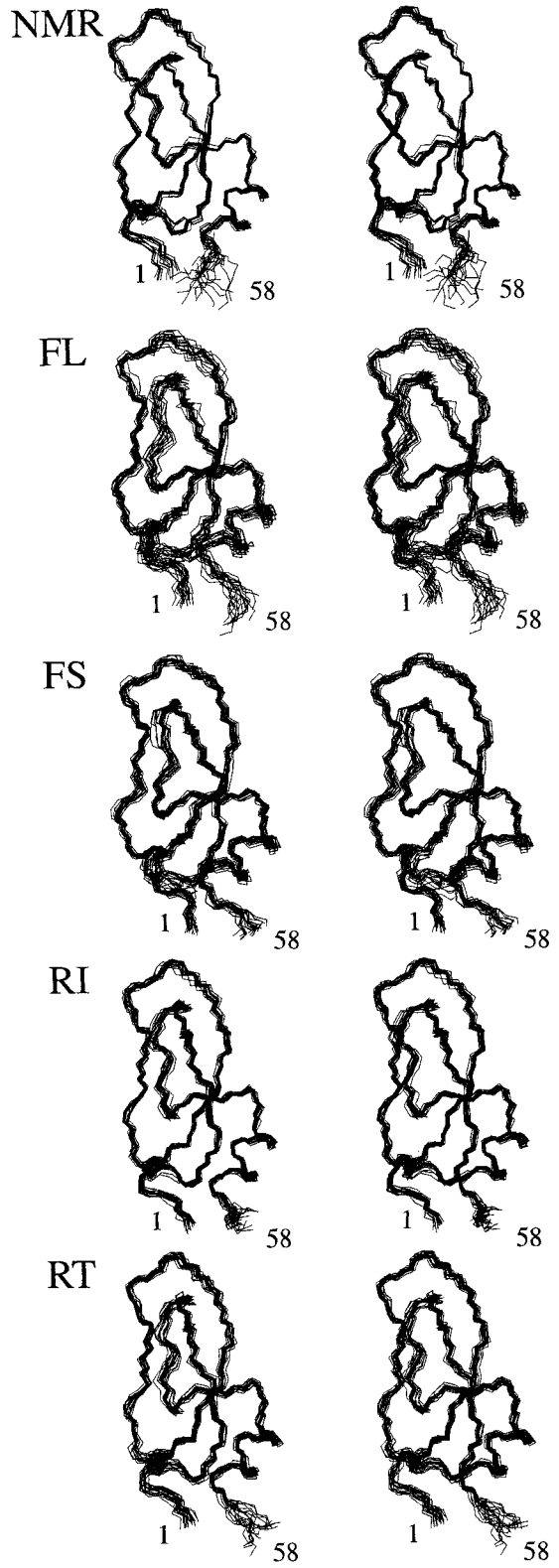

Fig. 3. Stereo views of the backbone atoms N, C^{α}, and C′ of selected groups of 20 conformers of BPTI (residues 1–58). NMR identifies the 20 conformers representing the NMR solution structure of BPTI (1PIT[13]). FL, FS, RI, and RT are groups of conformers from MD simulations as specified in Figure 1. In each group the conformers 1–20 were superimposed onto the mean conformer for minimal RMSD of the backbone atoms N, C^{α}, and C′ of residues 2–56.

side-chain conformations is more variable along the sequence than for the polypeptide backbone, as is quantitatively expressed by plots of the average of the pairwise local side-chain displacements, D_{sc}^{loc} (Fig. 2C). With few exceptions, the side-chain displacements of the five sets of conformers are significantly smaller than the estimated maximum values for completely disordered side-chains attached to the backbone conformation of the NMR conformers of BPTI (thin line). The sets FL and NMR do, however, contain some side-chains that approach random behavior. These are mostly Asp, Glu, Lys, and Arg residues, which also have large solvent-accessible surfaces (Fig. 2D). Peripheral side-chain protons of these residues have no or very few experimental interresidual NOEs (Fig. 2A), which is the result of both scarcity of close interresidual contacts and quenching of NOEs by side-chain mobility.

Excluding alanines and prolines, the NMR solution structure contains 17 "best-defined" side-chains with local side-chain displacements $D_{sc}^{loc} < 0.5$ Å (Fig. 2C). With a precision approximately equivalent to that of the polypeptide backbone (Fig. 2B), these side-chains are better defined in the NMR structure than in any of the sets of conformers from the MD trajectories. Except for Cys 14, there are no side-chains other than Ala and Pro with $D_{sc}^{loc} < 0.5$ Å in the MD trajectories. Two of the well-defined side-chains in the NMR structure, Phe 4 and Asn 24, are strongly disordered in FL ($D_{sc}^{loc} > 1.5$ Å). Interestingly, for Asn 24 the conformational disorder is even larger in set RT, which used time-averaged distance restraints. If, on the other hand, we consider the 13 most disordered side-chains in the NMR structure, i.e., those with $D_{sc}^{loc} > 1.5$ Å, we see that for most of them the conformational spread is largest in the NMR structure, most notably for Asp 3, Glu 7, Lys 15, Arg 42, Lys 46, and Asp 50 (Figs. 2C, 4), the only exceptions being Arg 17, Lys 26, and Arg 39, which have the largest spread in the set FL.

Several side-chains with conformations determined with approximately equal precision in the sets NMR and FL adopt locally different conformations in the two structures. Notable examples include Phe 4, Ile 18, Asn 24, and Thr 54 (Fig. 4). However, even during the 1 ns MD trajectory FL, there were only a small number of transitions of the corresponding side-chain torsion angles, implying that the present MD runs are still too short to allow for a statistically meaningful analysis of the side-chain conformations.

The disulfide bridges deserve a further comment. In their description in the NMR solution conformation of BPTI, Berndt et al.[13] found that the chirality of the disulfide bond Cys 30-Cys 51 could be uniquely defined ($\chi^3 = -88° \ldots -76°$), whereas the torsion angles χ^3 in the other two disulfide bonds Cys 5-Cys 55 ($\chi^3 = -168° \ldots + 168°$) and Cys 14-Cys 38 ($\chi^3 = -164° \ldots + 138°$) could not be precisely defined. Subsequent investigations revealed that the disulfide bond Cys 14-Cys 38 exists as an equilibrium between a right-handed major form and a left-handed minor form.[27] In the present study we find that the chirality of the disulfide bond Cys 30-Cys 51 is also uniquely defined in the four sets of MD conformers FL, FS, RI, and RT, and the range of values observed in set FL ($\chi^3 = -101° \ldots -85°$) coincides well with those of the NMR conformers[13] and the three X-ray crystal structures ($\chi^3 = -90° \ldots -80°$).[10–12] The chiralities of the other two disulfide bonds are, however, also uniquely defined in all the MD results, including FL. For the disulfide bond Cys 5-Cys 55, the range of χ^3 is $+63° \ldots +107°$ for set FL, which excludes on the one hand the χ^3 value of $-178°$ in the conformer from which the MD simulation started, and on the other hand the range of angles found in the three X-ray crystal structures ($-83° \ldots -80°$). Similarly, the narrow range of angles for the disulfide bond Cys 14-Cys 38 in the set FL ($\chi^3 = -94° \ldots -77°$) excludes the value of $-156°$ in the start structure as well as the values found in the three X-ray structures of BPTI ($\chi^3 = +94° \ldots +96°$).

Hydrogen Bond Formation

Comparison of hydrogen bonds found in the NMR structure, the X-ray crystal structure, and the four sets of conformers generated by MD simulations shows that in the regular secondary structural elements, all 10 backbone–backbone hydrogen bonds involving the two-stranded antiparallel β-sheet (residues 18–35) and the single residue comprising the third strand of the sheet (residue 45), as well as all backbone–backbone hydrogen bonds in the C-terminal α-helix are conserved among the five sets of conformers, with the sole exception that in FS the two hydrogen bonds 51 O′–55 HN and 52 O′–56 HN are replaced by a π-helix-type hydrogen bond 52 O′–57 HN. More extensive differences are found for hydrogen bonds involving donors or acceptors from side-chains, presumably because the conformers from the MD trajectories are, in contrast to the NMR structure, not energy minimized. The characteristic N-cap hydrogen bond at the start of the α-helix (47 γO–50 HN) is present in all five sets, but the C-cap (54 γOH–50 O′) is not present in any of the sets resulting from MD simulation. Another significant difference is found in the triad of hydrogen bonds coordinating the side-chain amide group of Asn 43. The three hydrogen bonds (43 δNH_2–7 O′, 43 δNH_2–23 O′, and 23 HN–43 δO) are present in the NMR as well as the X-ray crystal structures, but in each of the MD simulations at least one of these hydrogen bonds is absent, and these three hydrogen bonds are all missing in sets RI and RT. Overall, in the restrained molecular dynamics simulations there are fewer intraprotein hydrogen bonds involv-

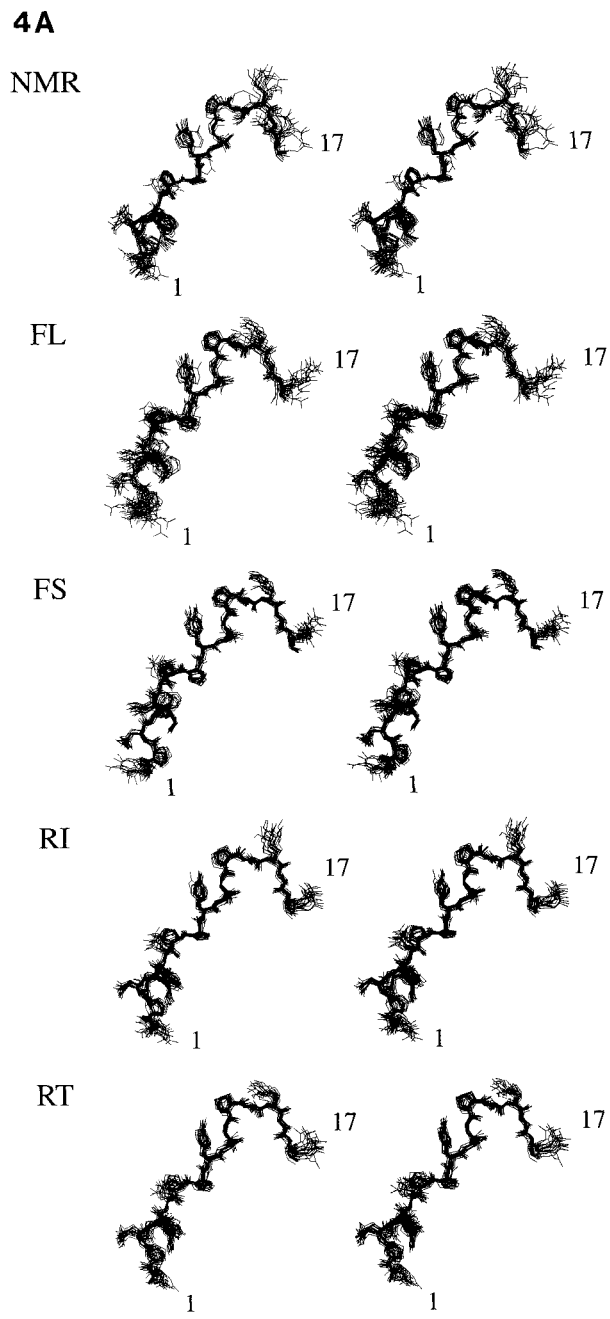

Fig. 4A.

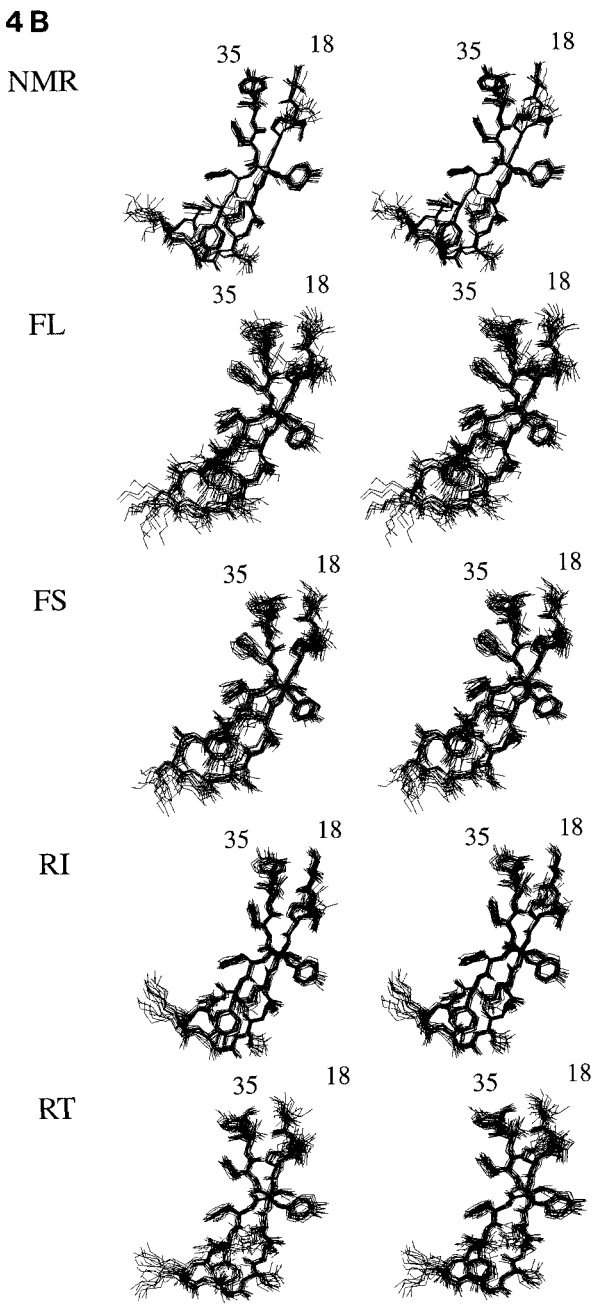

Fig. 4B.

ing side-chains than either in the NMR solution structure, the X-ray crystal structure, or the sets of conformers sampled from the free MD simulation.

CONCLUSIONS

Figures 3 and 4 show that qualitatively similar sets of protein conformers are generated either by the standard procedure for calculating a protein structure from NMR data using the program DIANA,[1,4,5] or by free or restrained MD simulations as outlined in Figure 1. The following additional comments relate to results of the quantitative data analysis, which showed different behavior of molecular areas that are well constrained by the NMR data, and those with only scarce NMR restraints.

The DIANA sampling of conformation space for poorly constrained, surface-exposed side-chains reproduces, and often even exceeds, that produced by the MD simulations. This result can be rationalized by the fact that each DIANA conformer is calculated

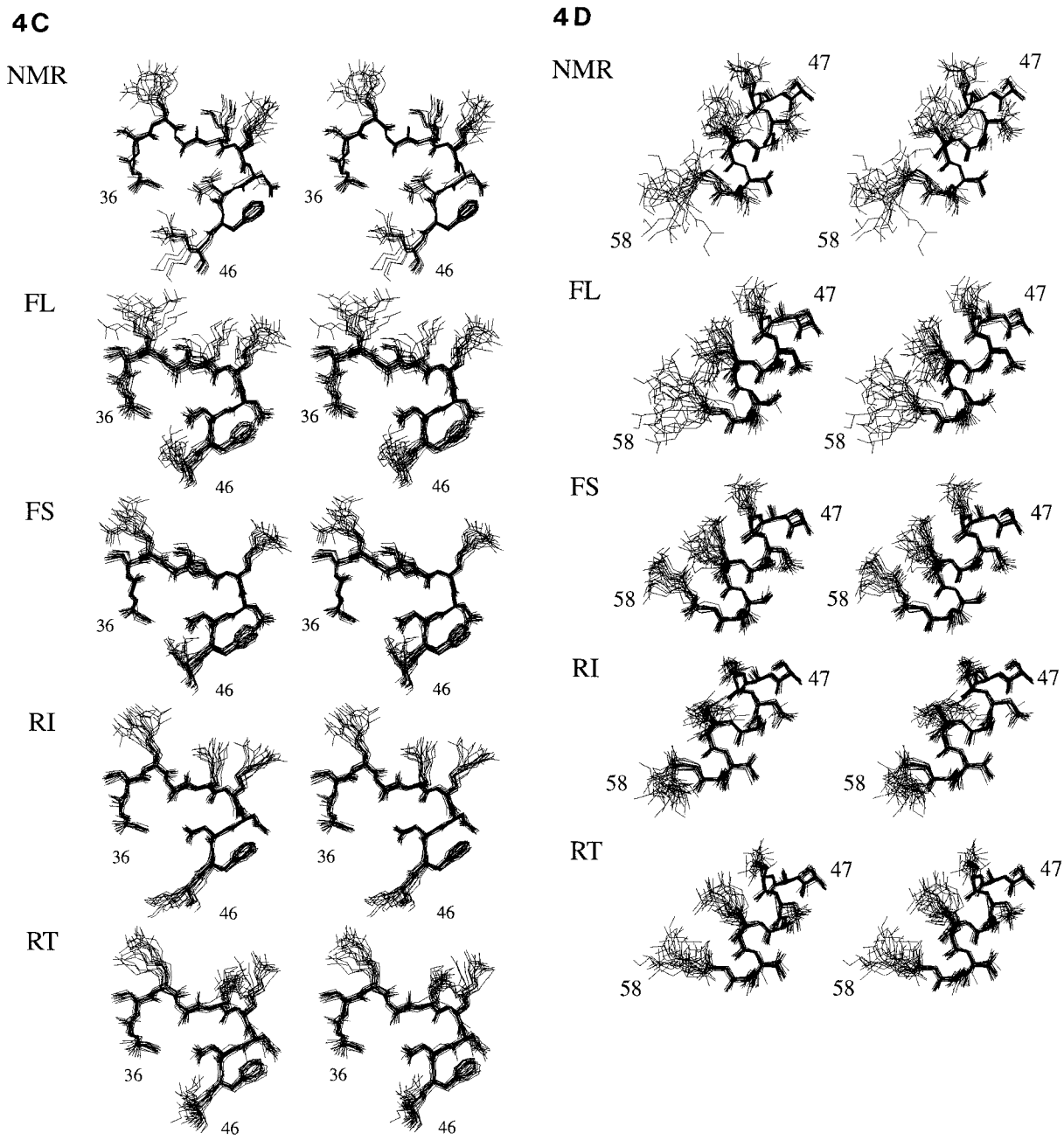

Fig. 4. Same as Figure 3 for individual segments of BPTI showing all heavy atoms. Conformers 1–20 were superimposed for minimal RMSD of the backbone atoms N, C^{α}, and C′ of the displayed residues unless specified otherwise. **A:** Segment 1–17 is shown after superposition of residues 2–17. **B:** Segment 18–35. **C:** Segment 36–46. **D:** Segment 47–58 is shown after superposition of residues 47–56.

independently, starting from a different, completely randomized starting structure, and that all sterically allowed molecular geometries are sampled with equal probability, irrespective of the associated conformational energy. We note further that for all simulations collected over a period of 200 ps (Fig. 1) the sampling of conformation space for unconstrained surface areas is significantly reduced with respect to the 1 ns free MD trajectory. This suggests that under near-physiological conditions of temperature, pressure, and solvent, MD trajectories of at least 1 ns length are required to sample the available conformation space for poorly constrained surface areas as exhaustively as the standard protocol for NMR structure calculation using DIANA. However, since the GROMOS force field includes many

more details of the physical energy function than the DIANA target function, such as electrostatic interactions, attractive van der Waals interactions, and torsion angle potentials, some of the conformations found by DIANA may be excluded also after much longer MD simulations (an illustration is provided by the disulfide bonds in BPTI discussed above). The dependence of the conformational sampling on the length of the MD trajectories between 200 and 1,000 ps suggests nonetheless that the independently randomized starting conformers of the DIANA calculations lead to final low-energy conformers located in regions of conformation space that may often be separated by energy barriers that are crossed only rarely on a nanosecond timescale. The significance of these findings is to be seen whenever MD simulations are used to mimic the molecular surface in solution when only a crystal structure is available, for example, because the protein is too big for a NMR solution structure determination. Much care should be exercised in the interpretation of such studies, since MD simulations of larger systems in explicit water are today still limited to trajectories that extend over only a fraction of a nanosecond. Furthermore, the present results may help to rationalize differences in the apparent precision of NMR structures calculated either with DIANA, or with computational procedures where the search of the conformation space allowed by the NMR data is guided by strong conformational energy criteria.

The well-defined part of the BPTI structure, i.e., the polypeptide backbone of residues 2–56 and the core side-chains, are representative of outstandingly stable, disulfide-rich globular proteins. The RMSD values for groups of conformers sampled during the MD simulations performed under physiological conditions (Table I) can thus serve as a guide for the physically meaningful precision with which a well-constrained protein structure in solution at ambient temperature can be described. The present study shows that these MD simulations of the protein core sample comparable regions of the conformation space as the NMR structure calculated with DIANA, with the 1 ns unrestrained MD trajectory showing a trend to sample even somewhat larger regions (Fig. 2B).

ACKNOWLEDGMENTS

We thank Dr. R.M. Brunne and Prof. W.F. van Gunsteren for helpful discussions and critical comments and Mr. R. Marani for the careful processing of the typescript. The MD simulations were performed in part on the NEC SX-3/22 supercomputer of the Centro Svizzero di Calcolo Scientifico in Manno, Ticino, Switzerland. Financial support by the Schweizerischer Nationalfonds (project 31.32033.91) is gratefully acknowledged.

REFERENCES

1. Güntert, P., Braun, W., Wüthrich, K. Efficient computation of three-dimensional protein structures in solution from nuclear magnetic resonance data using the program DIANA and the supporting programs CALIBA, HABAS and GLOMSA. J. Mol. Biol. 217:517–530, 1991.
2. Wüthrich, K. "NMR of Proteins and Nucleic Acids." Wiley: New York, 1986.
3. Braun, W., Gō, N. Calculation of protein conformations by proton-proton distance constraints. A new efficient algorithm. J. Mol. Biol. 186:611–626, 1985.
4. Güntert, P., Wüthrich K. Improved efficiency of protein structure calculations from NMR data using the program DIANA with redundant dihedral angle constraints. J. Biomol. NMR, 1:446–456, 1991.
5. Güntert, P., Qian, Y.Q., Otting, G., Müller, M., Gehring, W.J., Wüthrich K. Structure determination of the *Antp*(C39→S) homeodomain from nuclear magnetic resonance data in solution using a novel strategy for the structure calculation with the programs DIANA, CALIBA, HABAS and GLOMSA. J. Mol. Biol. 217:531–540, 1991.
6. Kaptein, R., Zuiderweg, E.R.P., Scheek, R.M., Boelens, R., van Gunsteren, W.F. A protein structure from nuclear magnetic resonance data. Lac repressor headpiece. J. Mol. Biol. 182:179–182, 1985.
7. Brünger, A.T., Clore, G.M., Gronenborn, A.M., Karplus, M. Three-dimensional structure of proteins determined by molecular dynamics with interproton distance restraints: Application to crambin. Proc. Natl. Acad. Sci. USA 83: 3801–3805, 1986.
8. Nilges, M., Clore, G.M., Gronenborn, A.M. Determination of three-dimensional structures of proteins from interproton distance data by hybrid distance geometry—dynamical simulated annealing calculations. FEBS Lett. 229:317–324, 1988.
9. Havel, T.F., Wüthrich, K. A distance geometry program for determining the structures of small proteins and other macromolecules from nuclear magnetic resonance measurements of intramolecular ^1H-^1H proximities in solution. Bull. Math. Biol. 46:673–698, 1984.
10. Deisenhofer, J., Steigemann, W. Crystallographic refinement of the structure of bovine pancreatic trypsin inhibitor at 1.5 Å resolution. Acta Crystallogr. B31:238–250, 1975.
11. Wlodawer, A., Nachman, J., Gilliland, G.L., Gallagher, W., Woodward, C. Structure of form III crystals of bovine pancreatic trypsin inhibitor. J. Mol. Biol. 198:469–480, 1987.
12. Wlodawer, A., Walter, J., Huber, R., Sjölin, L. Structure of bovine pancreatic trypsin inhibitor. Results of joint neutron and X-ray refinement of crystal form II. J. Mol. Biol. 180:301–329, 1984.
13. Berndt, K.D., Güntert, P., Orbons, L.P.M., Wüthrich, K. Determination of a high-quality NMR solution structure of the bovine pancreatic trypsin inhibitor (BPTI) and comparison with three crystal structures. J. Mol. Biol. 227: 757–775, 1992.
14. Billeter, M., Schaumann, T., Braun, W., Wüthrich, K. Restrained energy refinement with two different algorithms and force fields of the structure of the α-amylase inhibitor tendamistat determined by NMR in solution. Biopolymers 29:695–706, 1990.
15. Brunne, R.M., Liepinsh, E. Otting, G., Wüthrich, K., van Gunsteren, W.F. Hydration of proteins. A comparison of experimental residence times of water molecules solvating the bovine pancreatic trypsin inhibitor with theoretical model calculations. J. Mol. Biol. 231:1040–1048, 1993.
16. Brunne, R.M., Berndt, K.D., Güntert, P., Wüthrich, K., van Gunsteren, W.F. Structure and dynamics of the bovine pancreatic trypsin inhibitor in aqueous solution from long-time molecular dynamics simulations. Proteins 23:49–62, 1995.
17. van Gunsteren, W., Berendsen, H.J.C. "Groningen Molecular Simulation (GROMOS) Library Manual." Groningen: Biomos, 1987.
18. Güntert, P., Berndt, K.D., Wüthrich, K. The program ASNO for computer-supported collection of NOE upper distance constraints as input for protein structure determination. J. Biomol. NMR 3:601–606, 1993.
19. Singh, U.C., Weiner, P.K., Caldwell, J.W., Kollman, P.A. "Amber 3.0." San Francisco: University of California, 1986.
20. Berendsen, H.J.C., Grigera, J.R., Straatsma, T.P. The

missing term in effective pair potentials. J. Phys. Chem. 91:6269–6271, 1987.
21. Torda, A.E., Scheek, R.M., van Gunsteren, W. Time-dependent distance restraints in molecular dynamics simulations. Chem. Phys. Lett. 157:289–294, 1989.
22. Torda, A.E., Scheek, R.M., van Gunsteren, W. Time-averaged nuclear Overhauser effect distance restraints applied to tendamistat. J. Mol. Biol. 214:223–235, 1990.
23. Xia, T.H. "Software for Determination and Visual Display of NMR Structures of Proteins: The Distance Geometry Program DGPLAY and the Computer Graphics Programs CONFOR and XAM." Zürich: Ph.D. thesis, ETH Nr. 9831, 1992.
24. McLachlan, A.D. Gene duplication in the structural evolution of chymotrypsin. J. Mol. Biol. 128:49–79, 1979.
25. Billeter, M., Kline, A.D., Braun, W., Huber, R., Wüthrich, K. Comparison of the high-resolution structures of the α-amylase inhibitor Tendamistat determined by nuclear magnetic resonance in solution and by X-ray diffraction in single crystals. J. Mol. Biol. 206:677–687, 1989.
26. Richmond, T.J. Solvent-accessible surface area and excluded volume in proteins. J. Mol. Biol. 178:63–89, 1984.
27. Otting, G., Liepinsh, E., Wüthrich, K. Disulfide bond isomerization in BPTI and BP-TI(G36S): An NMR study of correlated mobility in proteins. Biochemistry 32:3571–3582, 1993.