Journal of Peptide Science
J. Peptide Sci. **9**: 450–460 (2003)
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/psc.466

JOURNAL OF PEPTIDE SCIENCE

The Use of NMR Chemical Shifts to Analyse the MD Trajectories: Simulation of Bovine Pancreatic Trypsin Inhibitor Dynamics in Water as a Test Case for Solvent Influences

BERNARD BUSETTA, PHILIPPE PICARD* and GILLES PRECIGOUX

Unité de Biophysique Structurale, UMR 5471 CNRS, Université Bordeaux 1, 33405 Talence, France

Received 7 January 2003
Accepted 25 February 2003

Abstract: In this paper the NMR secondary chemical shifts, that are estimated from a set of 3D-structures, are compared with the observed ones to appraise the behaviour of a known x-ray diffraction structure (of the bovine pancreatic trypsin inhibitor protein) when various molecular dynamics are applied. The results of a 200 ps molecular dynamics under various conditions are analysed and different ways to modify the molecular dynamics are considered. With the purpose of avoiding the time-consuming explicit representation of the solvent (water) molecules, an attempt was made to understand the role of the solvent and to develop an implicit representation, which may be refined. A simulation of hydrophobic effects in an aqueous environment is also proposed which seems to provide a better approximation of the observed solution structure of the protein. Copyright © 2003 European Peptide Society and John Wiley & Sons, Ltd.

Keywords: chemical shifts; hydrophobic effects; implicit solvent representation; molecular dynamics

INTRODUCTION

Molecular dynamics (MD) simulations are becoming an increasingly important tool in the study of protein dynamics to analyse the deformation of a protein under specific conditions [1]. To obtain the most realistic simulation of water–protein interactions, an explicit representation of the water molecules is used [2], but even in this case a proper representation of non-bonded interactions (for instance, the interactions between water and aliphatic groups [3]) or electrostatic energies [4] is still needed. The parameters of the energy function used to describe the dynamics are derived as a quantum chemical force field which is consistent (CFF) with the final observed x-ray diffraction structure of some peptides and proteins [5–7]. There is no way to ascertain that an intermediate conformation generated during a MD simulation is really feasible.

On another hand, energy minimization using molecular mechanics fails to converge from a somehow distant starting structure towards the native ones [8]. A correct description of the dynamic pathway is unlikely if the solvent molecules are not explicitly introduced. The continuum solvent-model, an implicit water-simulation, takes advantage of the classical electrostatic theory and considers the solvent as a homogeneous dielectric medium that can be polarized by the electric charges of the solute. With the inclusion of vibrational energy and accounting separately for polar and non-polar surface areas, the continuum models should be as accurate as the explicit water simulations [9]. But in such representations of water–protein interactions, derivation and then energy minimization are not possible: an approximated protein structure cannot

*Correspondence to: Philippe Picard, Unité de Biophysique Structurale, Bât. B8, avenue des Facultés, Université Bordeaux I, 33405 Talence Cedex, France; e-mail: P.Picard@ubs.u-bordeaux1.fr

be improved. Even for a small peptide, a huge conformational space should be explored to define the suitable energy minimum. Consequently, this approach cannot provide a complete description of the folding process for large peptides.

To save computing time, a complete analysis of the forces that are critical to the determination of backbone conformation and its dynamical behaviour in the solution state is necessary. Hitherto, particular force fields associated with simplified geometric representations of the macromolecules and simulating the solvent effects were used to allow an *ab initio* determination of more or less complex peptide and protein structures [10,11]. Both kinds of simplifications are necessary to develop a statistical treatment of the problem with well-defined structures. But the more important the simplifications of the whole system are, the greater are the discrepancies that occur in the prediction of the peptide 3D-structures.

To allow a more realistic simulation of the molecular dynamics or folding, the all-atom geometry of the macromolecule was kept unchanged. In this case a statistical treatment of the internal force field to mimic the solvent effects is impossible. The set of the generated conformers depends strongly on the assumed representation of the solution state. The NMR chemical shifts provide an experimental observation of the molecular dynamics. A simulation of this dynamics was searched that provides the best fitting between the chemical shifts estimated from a mixture of generated conformers and the observed ones [12] in the solution state. The experimental observation was used to appraise the consequence that any modification of CFF as well as of the dynamic process may have on the simulation of the behaviour of a stable protein in solution.

The small BTPI (bovine pancreatic trypsin inhibitor) protein (under the so-called form II, a neutron diffraction structure corresponding to file *5pti* of the Protein Data Bank [13]), which remains stable in various environments or dynamics, was used in this study. The behaviour of the protein molecule round its optimal x-ray diffraction 3D-structure was analysed.

METHODS

Empirical Energy Function

The X-Plor 3.8 refinement package [14] was used for all computations. The molecular energy is described through an empirical energy function where the pairwise non-bonded interactions and the geometric forces are dealt with separately. The total energy E_{tot} is divided into geometric energy E_{geom} and non-bonded energy. The parmallh3x.pro file of the X-Plor package provides for each atom (H, C, N, O, S) type the standard geometry and the non-bonded well-depth derived to be consistent with the observations performed in the crystal state on some carboxylic acids and amides by x-ray diffraction [15,16].

A harmonic approximation is used to account for the deformations of covalent bonds and bond angles. Four-atom terms are used to define the energy terms involving dihedral angles and to maintain both chirality about a tetrahedral atom and planarity of certain atom groups (aromatic ring systems, peptide bonds, ...). The geometric energy function is:

$$E_{geom} = k_{bd}(r - r_0)^2 + k_{ang}(\theta - \theta_0)^2 + k_{dih}E_{dihedral}$$

which accounts for deviations from the standard geometric values (k_{bd}, k_{ang} and k_{dih} are variable scale factors).

The non-bonded energy may be divided into van der Waals energy and electrostatic energy. The van der Waals energy between two atoms distant by R is given by a classical Lennard-Jones potential:

$$E_{VdW} = -4e_0[(\sigma/R)^{12} - (\sigma/R)^6] = A/R^{12} - B/R^6$$

where e_0 is the well depth at the minimum interatomic distance R_{min} and σ is related to R_{min} by the relationship $\sigma = R_{min} \times 2^{-1/6}$. The first term of the VdW energy A/R^{12} is a repulsive term, whereas the second term is an attractive one.

The non-bonded energies are expected to be valid in the solution state if the solvent is explicitly defined. In this case the solvent effect appears as an additive E_{solv} term. When the solvent molecules are not defined, it may be thought that some alterations of the energy parameters have be derived to be consistent with the solution state and to simulate the solvent effect. However, as the solution state is an average of conformers, the least-square refinement, which was used with crystal structures to define the energy parameters [5], is impossible.

The electrostatic energy between two atomic charges q_i and q_j is given by a Coulombic potential $q_i q_j C/\varepsilon_0 R^2$, where ε_0 is the dielectric constant and R is the distance between the i and j electric charges. With this formulation, an approximate solvent screening is introduced at the level of the dielectric constant, which is made distance-dependent. With the use of such a representation, the relative

importance of the electrostatic interaction between the donor and acceptor peptide groups involved in a hydrogen bond is increased.

Molecular (Langevin) Dynamics

Molecular dynamics consists of solving Newton's equations of motion:

$$m_i * d^2x_i(t)/dt^2 = -V_{xi}(E_{tot})$$

where m_i stands for the mass of each protein atom and $V_{xi}(E_{tot})$ for the gradient derived from the total energy function. To initiate the dynamic process, an initial velocity is assigned to each atom following a Maxwellian distribution depending on a random seed number.

In the different experiments (i.e. when different conditions are applied via the energy function), the beginning of the dynamics (as defined by the seed value) is always the same. All the experiments are processed in the same way.

The Langevin dynamics process was applied to BTPI. The initial structure was submitted to 100 iterative steps of 2 ps dynamics generally at 400 K. At the end of each step, the current structure is extracted and submitted to a steep cooling (i.e. 1500 cycles of Powell refinement). The theoretical solution is the set of the 100 generated conformers.

A rms deviation Δ is defined over the backbone atoms for each of the extracted conformers in comparison with the correspondent atoms of the crystal structure. The average $\langle\Delta\rangle = 1/100\Sigma(\Delta)$ over the 100 conformers generated with any variation of the energy function is used to appraise the deviation from the x-ray diffraction structure.

The Langevin dynamics is supposed to provide a suitable simulation of the molecular dynamics when an explicit representation of the solvent is used. Otherwise it should be considered as a simulation of the thermal motion *in vacuo*. Thus, the studied conformer deviates rapidly from the starting structure (with a rmsd greater than 1 Å). In many cases, when the value of an energy scale factor or parameter is scanned, the dynamics cannot last more than 200 ps without large deviations from the initial structure and a pronounced protein unfolding. Only the disulphide bridges are maintained.

Simulation of the Dynamics by Random Side-chain Motion

Although classical molecular dynamics can overcome the small energy barriers which rule the side-chain flexibility, it is unlikely that such a technique could reproduce this flexibility in a small amount of time and for instance that it could provide a suitable averaged side-chain orientation in solution. A random generation of the side-chain rotameric states was already proposed to define the orientation of these chains in the protein predictive schemes [17].

In a given environment the mobility of the solvent molecules is mainly responsible for the protein motion. The reorientation of the side chains involves the distortions that are observed on the overall backbone conformation. The thermal dynamics may be considered the result of perturbations applied randomly to the protein side chains.

During the molecular dynamics each side chain may occupy three different orientations Θ_{χ_1} following the height of the energy barrier round the χ_1 axis. The motion towards any one of these three orientations is made possible by the solvent motion. The dynamics results from a reorientation of the different side chains at random. By adding repeatedly the transient extra harmonic dihedral constraint

$$k_{\chi_1}\{\Theta_{\chi_1} - (\Theta_{o\chi_1} - 120n_{rd})\}^2$$

to the energy function one may simulate it. $\Theta_{o\chi_1}$ is the initial dihedral angle (expressed in degree) before the perturbation is applied and n_{rd} is a random number to select one of the three possible orientations Θ_{χ_1} (some n_{rd} values may involve no constraint). The scaling factor k_{χ_1} in Kcal/degree^2) defines the strength of the perturbation.

In practice, the protein dynamics was simulated by compelling, in each one of 1000 iterative cycles, $N/8$ randomly selected side chains (N is the total number of protein residues) to move. The constraint is applied on a selected chain during n_{pw} cycles of Powell refinement. The dynamics (further referred to as dynamics χ_1) is the result of the succession of such perturbations. In the same way as previously, the theoretical solution is defined by the set of the conformers that were extracted in each cycle.

Estimation of the ^{1}H Chemical Shift Deviations

Following the process developed by Osapay and Case [18], the chemical shifts of the ^{1}H protons (δ) are estimated for each conformer J as the sum of elementary contributions :

$$\delta_J = \delta_{local} + \delta_{el} + \delta_m + \delta_{rc}$$

where δ_{local} is the local contribution approximated to the proton chemical shifts observed at 309 K

for short peptides in the random coil conformation [19], δ_{el} is the electrostatic contribution restricted to the contribution of the main chain, δ_m is the magnetic contribution of the peptide groups, and δ_{rc} is the aromatic ring contribution (i.e. the ring current effect).

A randomization was introduced to take into account the effects of the thermal motion in the crystal state [20]. It is assumed that the different possible orientations of the aromatic rings are encountered on an average throughout the different conformers. The importance (weight) of the ring current effect from each aromatic cycle is made dependent on its relative exposition to the solvent.

The theoretical solution is assumed to be a mixture of the 100 conformers generated by the dynamics from the observed *5pti* x-ray diffraction structure. If $\delta_{i,J}$ is the calculated chemical shift of proton i for the conformer J, the resulting chemical shift of proton i in the theoretical solution is:

$$\delta_i = \Sigma_J(p_J\delta_{i,J})/n_c$$

where $n_c = 100$ is the number of generated conformers and p_J is the probability of the presence of conformer J in the theoretical solution. In this study, an identical probability $p_J = 0.01$ was used for each conformer assuming that the relative number of nearly identical conformers is sufficient to give a suitable definition of the theoretical solution.

If δ_{iobs} is the observed chemical shift for the proton i in the BTPI solution structure [12], the conformational difference round proton i may be appraised by the absolute value $|\delta_i - \delta_{iobs}|$.

The average on the absolute difference:

$$\langle H_\alpha \rangle = 1/n_p \Sigma|\delta_i - \delta_{iobs}|$$

is computed on all the n_p H_α protons [referred to as the $\langle H_\alpha \rangle$ error for the theoretical solution based on n_c conformers and expressed in ppb (part per billion)]. $\langle H_\alpha \rangle_{E>E_o}$ (with $n_c < 100$) is the error computed for a theoretical solution based on a restricted number of conformers with an energy $E < E_o$.

The $\langle H_\alpha \rangle$ error is sensitive to the backbone conformation and to the strength of any intramolecular hydrogen bond (i.e. to the vicinity of the peptide group involved in the hydrogen bond). It is a good marker of the conservation of the protein secondary structure. Such an average was already used to predict secondary structures [21]. Presently the smaller the averaged absolute error is, the nearer the set of conformers generated by the dynamics will be to the solution state.

Estimated Error on the Test Parameters

The dynamic path depends on the choice of a random seed which is transmitted to the generator of random number to define the initial velocities or the choice of the side chains in rotation. The choice of a random seed would not be a concern if the dynamics might proceed during a very long time. In our case, to explore as thoroughly as possible the potential wells in which the molecule moves, it is necessary to estimate the error involved on the test parameters by applying a dynamic process which is restricted to a short time. For each definition of the energy function, the dynamics process (restricted to 200 ps) was repeated five times with a different seed (however, the set of these five different seeds is the same for all the variations of the energy function under study). The test parameters ($\langle\Delta\rangle$ and $\langle H_\alpha \rangle$) were defined as the correspondent average over the five runs and the error involved on these parameters is the correspondent rms deviation (referred to as $d\langle\Delta\rangle$ and $d\langle H_\alpha \rangle$). Increasing the number of dynamic runs does not seem to affect the $\langle\Delta\rangle$ and $\langle H_\alpha \rangle$ values associated with the energy function beyond this estimated error.

RESULTS AND DISCUSSION

In a first approach only a classical Langevin dynamics was performed using an all-atoms representation (with the force field described in the parmallh3x.pro file of X-Plor), with an identical atom mass ($m_i = 100$) for each atom. The different dynamics are appraised by comparison of the correspondent theoretical solution with the solution constituted by the unique *5pti* x-ray conformer.

Despite the influence of the crystal field, hitherto x-ray diffraction structures are described as the best approach of the observed solution structure [22]. The unrefined *5pti* x-ray diffraction structure corresponds to an absolute error between the observed and estimated H_α chemical shifts equal to 170 ppb. But, as soon as the initial structure is refined to optimize the bond lengths and angles and to erase the atom overlapping (by 1500 cycles of Powell refinement for instance), the refined conformer deviates by $\Delta = 0.38$ Å with an absolute error at the level of the $\langle H_\alpha \rangle$ chemical shifts equal to 187 ppb. When the x-ray diffraction structure

is submitted to 2 ps dynamics at 50 °K followed by 1500 cycles of Powell energy refinement (i.e. during the foremost step of the dynamics performed at that temperature where the mobility is limited), the final conformation corresponds to a deviation (Δ) equal to 0.54 Å and an absolute error equal to 198 ppb. During a 200 ps dynamic run the *pti* conformer may deviate steadily from the initial x-ray diffraction structure up to 2 or 3 Å.

All the pairs ($\langle\Delta\rangle$, $\langle H_\alpha\rangle$), which are encountered during the following 200 ps Langevin dynamics, are collected in Figure 1. All the points are distributed such as $\langle\Delta\rangle$ remains smaller than $0.24 \times (\langle H_\alpha\rangle - 166)$. Then, it may be assumed that searching a conformer associated with the smallest $\langle H_\alpha\rangle$ error is a way to obtain a structure close to the x-ray structure. The use of the estimation of the chemical shifts to refine an initial 3D-structure seems justified. On the contrary, the reverse is not true. Maintaining the molecular conformation as close as possible to the initial native structure is not always the best way to generate a set of conformers which represents the solution state.

Restriction of the Molecular Mobility

In the usual consensus a dynamics process is supposed to generate a set of conformers which does not deviate too much from the initial structure. This strategy could be justified by the fact that the x-ray diffraction structure is a good approximation of the solution state. It may be expected that, if the initial 3D-structure is more or less impeded, the dynamics will provide a theoretical solution in good agreement with the experimental solution. In this section different ways are considered to restrain the molecular mobility.

NOE constraints. Hitherto the NMR information used to determine protein structures has always included interproton distances related to nOe (nuclear Overhauser effect) cross-peak intensities. In this section the extent to which the constraints defined by the observed nOe may be used to decrease the mobility of the protein during a dynamic process is analysed.

The 86 nOe constraints, which involve the backbone hydrogen atoms, were extracted from the initial determination [12] and were used in a dynamic process applied to the *5pti* structure. The weight assigned to the nOe constraints varies from 0 to 50 Kcal/Å^2 in the different runs. The hydrogen bonds were simulated by a slightly attractive function as in the force field used in the X-Plor package for NMR constrained simulated annealing and no electric charge was assigned to the carboxylic acid and to the basic heads as suggested for the NMR as well x-ray dependent refinements.

From Figure 2, considering the $\langle H_\alpha\rangle$ errors, it clearly appears that, independent of the weight

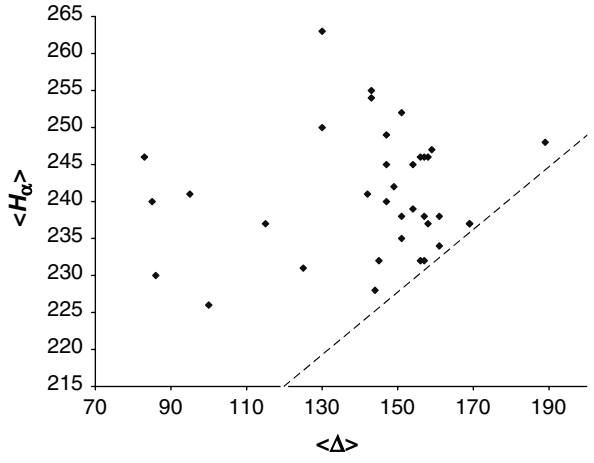

Figure 1 Distribution of the average error $\langle H_\alpha\rangle$ (in ppb) between the estimated chemical shifts and the observed ones vs the average deviation $\langle\Delta\rangle$ ($\times 100$ in Å) from the initial x-ray diffraction structure of the same conformer during the various Langevin dynamics which are analysed in this paper. The $\langle\Delta\rangle$ deviation is correlated to the $\langle H_\alpha\rangle$ error and remains lower than $0.240(\langle H_\alpha\rangle - 166)$. This relation is represented by a dashed line.

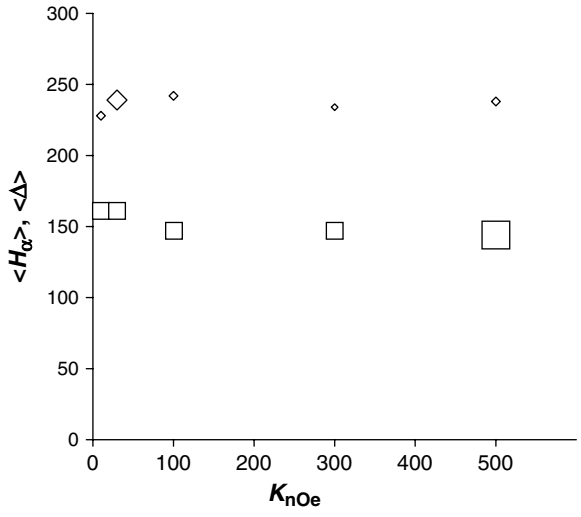

Figure 2 Variations of the average rms deviation $\langle\Delta\rangle$ (squares) and the average error $\langle H_\alpha\rangle$ (diamonds) between estimated and observed chemical shifts following the weight ($\times 10$) assigned to NOE constraints. The size of the squares and lozenges is equal to the averaged errors $d\langle\Delta\rangle$ and $d\langle H_\alpha\rangle$ on these averaged values.

of the k_{nOe} scale factor, no improvement of the generated theoretical solution in comparison to the observed one is obtained by the introduction of the nOe constraints. However, the large values of the k_{nOe} weight (k_{nOe} > 10 kcal/Å^2) assigned to the nOe derived constraints involve large $\langle\Delta\rangle$ deviations from the crystal structure, suggesting some differences between the crystal and the solution states.

'Bath' temperature. 'Freezing' the initial 3D-structure is potentially another good way to generate a suitable theoretical solution. In Figure 3 the $\langle\Delta\rangle$ deviation from the initial x-ray diffraction structure is highly improved by a decrease of the 'bath' temperature. The concomitant decrease of the $d\langle\Delta\rangle$ error indicates a decrease of the molecular mobility. There is no absolute evidence to define an optimal temperature from the analysis of the averaged deviations of the theoretical solution from the observed ones. The variations of $\langle H_\alpha\rangle$ hardly suggest a temperature near 300 K.

In Figure 4 for both 'bath' temperatures 100 K and 500 K the variations of the $\langle H_\alpha\rangle$ errors when the theoretical solutions are defined following a threshold energy ($E < E_o$) over a restricted number of conformers are reported. The existence of a minimum for $\langle H_\alpha\rangle_{E<Eo}$ in the case of the dynamics at 500 K suggests that the solution state defined by the chemical shifts requires a more restricted number of conformers. The energy which corresponds to the $\langle H_\alpha\rangle$ minimum may be used to define a real bath

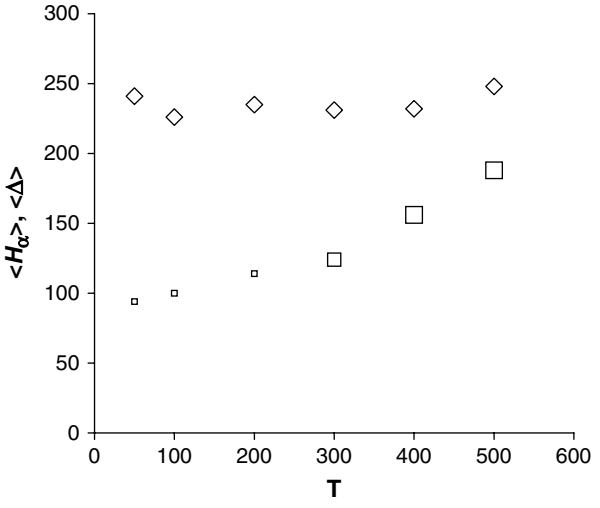

Figure 3 Variations of the average rms deviation $\langle\Delta\rangle$ and the average error $\langle H_\alpha\rangle$ during dynamics as a function of increasing temperature (in K) (same symbols as in Figure 2).

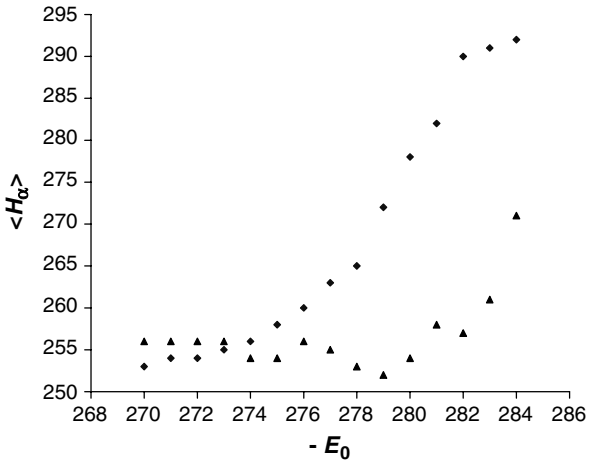

Figure 4 Variations of the average error $\langle H_\alpha\rangle$ vs the threshold energy E_o value at $T = 100$ K (lozenges) and 500 K (triangles). $\langle H_\alpha\rangle$ values are computed for theoretical solutions restricted to the conformers associated with an energy lower than the threshold value (i.e. $\langle H_\alpha\rangle_{E<Eo}$).

temperature lower than 500 K. On the contrary, the absence of such a minimum for the 100 K dynamics suggests that the theoretical solution involves a limited number of suitable conformers and that the bath temperature should be increased.

Bond stretching. During the dynamics the overall flexibility of the protein is mainly ruled by the rotations about the ϕ,ψ dihedral angles. Such rotations are made easier by making the bond stretching along these axes possible. The viscosity of the solvent round the protein implies a decrease of the protein flexibility. If the solvent molecules are not explicitly represented, it is necessary to simulate this decrease by impeding more or less the bond stretching (i.e. by increasing the k_{bd} scale factor in the geometric energy function).

In Figure 5 the k_{bd} scale factor is increased from 1 to 50 Kcal/Å^2. The decrease of the molecular flexibility, which follows the restriction of the bond stretching, may be verified with the decrease of the $d\langle H_\alpha\rangle$ error that suggests a better conservation of the local structure. However, even for large k_{bd}, no real improvement of the theoretical solution is obtained.

On another hand, as proved by the examination of the correspondent $\langle H_\alpha\rangle$ errors, no variations of the scale factor of the valence and dihedral angle energies are allowed to alter the dynamics of the protein ($k_{ang} = k_{dih} = 1$).

Atomic mass. A similar dynamic process was performed by assigning to each atom a mass equal

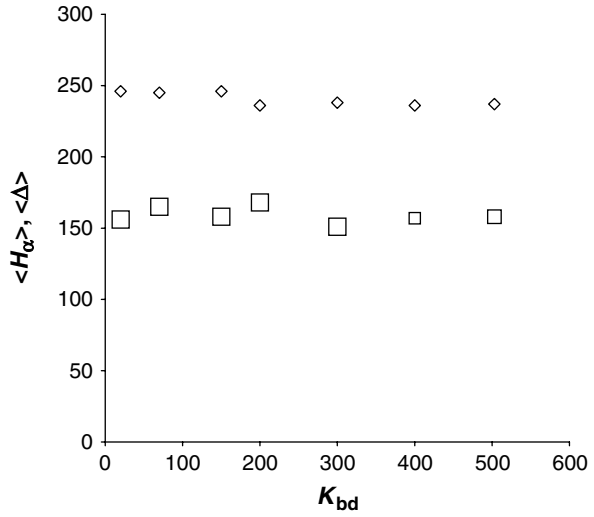

Figure 5 Variations of the average rms deviation $\langle\Delta\rangle$ and the average error $\langle H_\alpha\rangle$ vs the bond scale k_{bd} ($\times 10$ in kcal/Å^2) (same symbols as in Figure 2).

to 10*(atomic mass) in the Newton equation (i.e. increasing the mobility of hydrogen atoms relative to the other heavy atoms). In this case the allowed motion of the backbone is more limited. The $\langle\Delta\rangle$ deviation of the theoretical solution from the x-ray diffraction structure is only 0.83 Å (i.e. the same deviation observed between the nOe dependent conformers and the different x-ray diffraction structures [23]) but the $\langle H_\alpha\rangle$ error remains near to 240 ppb.

Hydrogen Bond

Usually, the effects of hydrogen bonds are taken implicitly into account by an appropriate parameterization of the partial electric charges and the van der Waals parameters. The major problem in such simulations concerns the description of water and ion screening. Usually, the solvation energies are not accounted for and the fact that intramolecular hydrogen bond or ion pair formation occurs at the expense of interactions of comparable magnitude with water is frequently ignored. On the other hand, hydrogen bonding is not considered as a factor determining the 3D-structure of peptides and proteins (which depends mainly on the primary structure) but uniquely as an additional stabilizing force.

Here we studied to what extent an attraction at the level of the van der Waals interactions between the hydrogen and oxygen atoms involved in the hydrogen bonds should be kept to allow formation of this bond.

As it may be supposed that an entropic factor favours the formation of intramolecular hydrogen bonds, the interaction between the H_N hydrogen and oxygen atoms is made slightly attractive during a nOe-dependent simulated annealing. In Figure 6 are reported the variations of the $\langle\Delta\rangle$ and $\langle H_\alpha\rangle$ parameters following the (A/B) ratio between the attractive term and the repulsive ones of the van der Waals interactions between non-bonded hydrogen and oxygen atoms. No evident improvement of the theoretical solution is obtained by any modification of the van der Waals interactions in the presence of hydrogen bonds.

If only the usual parameters are used to describe the interactions between the hydrogen and oxygen atoms, the $\langle\Delta\rangle$ deviation is 1.57 Å with an associated $\langle H_\alpha\rangle$ error equal to 246 ppb. During a folding process or any dynamics, the intramolecular hydrogen bonds are always in competition with intermolecular hydrogen bonds with the surrounding solvent (which is not explicitly expressed in our approach). To account for this competition, the weight of the normal van der Waals term should be largely decreased.

On another hand, the normal expression of the 1–4 interactions between neighbour hydrogen and

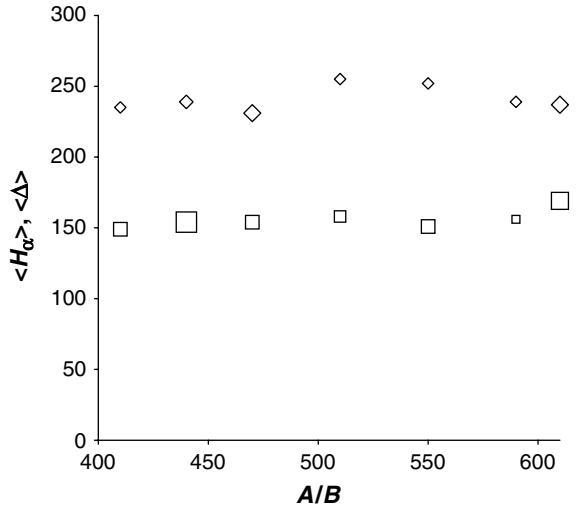

Figure 6 Variations of the average rms deviation $\langle\Delta\rangle$ and the average $\langle H_\alpha\rangle$ error vs the ratio A/B ($\times 10$) between the repulsive and the attractive terms for the non-bonded interactions between the oxygen and hydrogen atoms involved (or not) in a hydrogen bond. The upper limit of A/B corresponds to the effective van der Waals interactions between these atoms (same symbols as in Figure 2).

oxygen atoms should be strictly conserved to avoid the large discrepancies of the secondary structures (associated with an increase of $\langle H_\alpha \rangle$) which appear when these atoms overlap.

Dielectric Constant

Molecular dynamics simulations require calculation of the electrostatic contribution to the force on each atom or group. In Figure 7 the variations of the $\langle \Delta \rangle$ and $\langle H_\alpha \rangle$ parameters are represented following the value of the basic ε_0 dielectric constant. When $\varepsilon_o = 1$ or 2, the $\langle \Delta \rangle$ deviation and its correspondent $d\langle \Delta \rangle$ error are small. The generated conformers are maintained close to the crystal structure but the theoretical solution diverges from the observed ones. On the contrary, for larger ε_o values (=3), the $\langle H_\alpha \rangle$ error decreases and the set of generated conformers is closer to the observed solution.

This opposite behaviour provides a good illustration of the difficulty in devising a dynamic process monitored by the deviation from an initial x-ray diffraction structure with the aim of simulating a solution state. It may be explained by assuming that the surrounding solvent corresponds to specific and motionless bounded water molecules in the crystal state. In this case weak dielectric constants are required to describe the electrostatic energy. On the contrary, in the solution state, in agreement with other NMR evidence [24], a higher dielectric constant should be used considering that water molecules near proteins are quite mobile. For the same reason the use of a distance-dependent representation of the dielectric constant seems justified as it involves a smaller value at the level of the hydrophobic interior of the protein and a greater value at the level of the outside exposed to the solvent.

Dynamics χ_1

The dynamics χ_1 is based on the idea that the protein thermal motion results from the transmission of the solvent motion via the side chains. Then, defining a specific motion for the side chains without any initial motion for the molecular backbone may be a way to simulate one of the solvent effects. Furthermore, such a simulation of the protein dynamics appears little time-consuming (about 100 times faster than the usual Langevin dynamics). Thus, a peptide folding should be analysed in a similar way in a reasonable computing time.

In this simulation the 'bath' temperature cannot be defined, but the resulting motion is closely related to the strength (k_{χ_1}) and/or the duration (n_{Pw}) of the perturbation applied to the protein in each cycle of refinement. Such perturbations were performed during 1000 iterative cycles for five dynamic runs starting with a different random seed to select the side chains affected by the random perturbation. The $\langle H_\alpha \rangle$ error is defined on the last 100 extracted conformers. In the case of BPTI a duration $n_{\mathrm{pw}} = 50$ seems sufficient to generate an efficient dynamic process. In Figure 8 are reported the $\langle H_\alpha \rangle$ errors which are obtained when an increasing strength k_{χ_1}

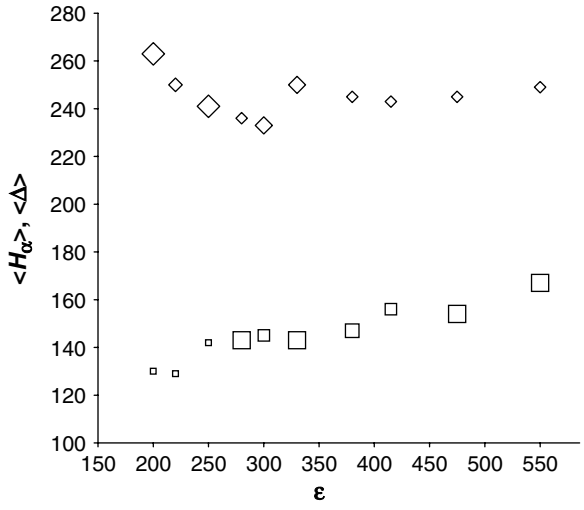

Figure 7 Variations of the average rms deviation $\langle \Delta \rangle$ (open squares) and the average error $\langle H_\alpha \rangle$ (open lozenges) vs the initial dielectric constant ε_o ($\times 100$) in the case of a distance-dependent constant.

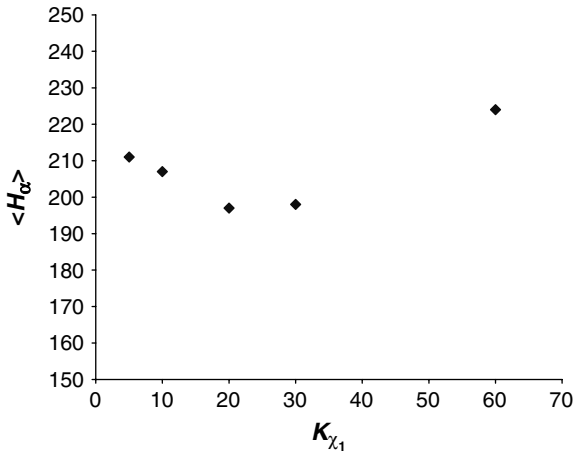

Figure 8 Variations of the averaged error $\langle H_\alpha \rangle$ following the weight k_{χ_1} (in kcal/degree^2) associated to the side-chain torsions that are imposed to the protein structure during a dynamics χ_1.

is used to define the perturbation. When the applied perturbation is weak ($k_{\chi_1} = 5–10$ Kcal/degree^2), the initial conformer is hardly distorted. Beneath a threshold value the energy barriers round the native structure cannot be overcome. Therefore, there is no motion (the $d\langle H_\alpha \rangle$ error is weak) as in the case where the 'bath' temperature is too weak. Whereas the deviation $\langle \Delta \rangle$ is steadily increasing, the variation of the $\langle H_\alpha \rangle$ error (as reported in Figure 8) suggests that an optimal value of the k_{χ_1} scale factor may be defined. At this point the side-chain motion should reproduce the effects of the solvent motion at the time of the measurement the chemical shifts. The $\langle H_\alpha \rangle$ error associated with the theoretical solution is improved round 200 ppb.

Hydrophobic Effects

The functional form of hydrophobic interactions is still unknown and the question of their relative importance by comparison to the electrostatic and other forces is still unsettled. However, an hydrophobic effect analysis in an aqueous environment shows that hydrophobic residues are generally closely packed in the interior of globular proteins, whereas hydrophilic side chains are mostly exposed to the solvent. This statement may be fulfilled if it is assumed that, in the presence of water, the attraction between hydrophobic atoms is greater than the attraction involving hydrophilic atoms.

To check this last assumption, only the aliphatic and aromatic hydrogen atoms are considered. In the case of the non-bonded interactions involving these atoms we looked for an optimal well-depth e_{HA} different from the e_{oHA} value which is used in the parmallh3x.pro file. This analysis suggests that a better simulation of the solution state is obtained with a well-depth of $e_{HA} \sim 2.6.e_{oHA}$ ($n_{pw} = 50$, $k_{\chi_1} = 20$). Both test parameters $\langle \Delta \rangle$ and $\langle H_\alpha \rangle$ were computed from the set of conformers extracted during each 100 successive cycles. In Figure 9 the behaviour of the initial x-ray conformer (as estimated by the $\langle \Delta \rangle$ and the $\langle H_\alpha \rangle$ errors) during successive 100 cycles of such a dynamics is reported.

The use of the chemical shifts seems to allow a quantification of the importance of the hydrophobic effects. Whereas the theoretical solution deviates steadily from the initial x-ray diffraction structure ($\langle \Delta \rangle$ increases), this solution is approached nearer and nearer by the observed solution state ($\langle H_\alpha \rangle$ decreases from 193 to 188 ppb). The incorporation

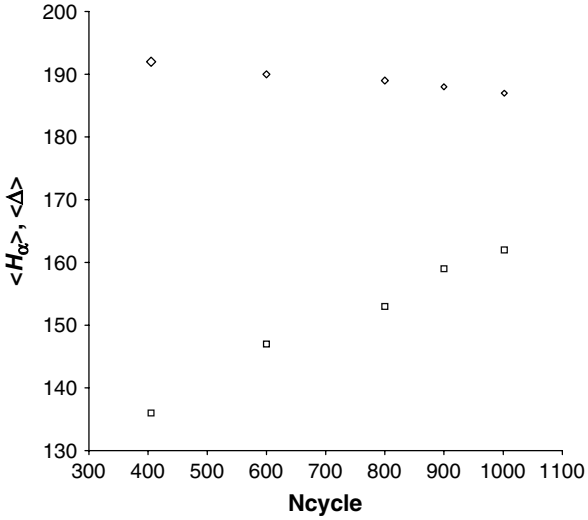

Figure 9 Variations of the average rms deviation $\langle \Delta \rangle$ ($\times 300$ in Å) and the average error $\langle H_\alpha \rangle$ during 1000 cycles of dynamics χ_1 performed with a scaling factor equal to 2.6 for the non-bonded interactions involving alkyl and aromatic hydrogen atoms (same symbols as in Figure 2).

of the nOe interactions does not provide any extra improvement to the generated set of conformers. This finding may suggest that no extra information was contained in such constraints.

The hydrophobic effects may be different for other proteins. For instance, they may depend on the size and the relative number of polar groups in the polypeptide chain.

CONCLUSION

In this paper we have tried to make obvious the advantage provided by the structural information present in the observed NMR chemical shifts. This information is more reliable than the usual analysis of the rms deviation from an initial observed x-ray diffraction structure as a criterion to appreciate the molecular dynamics and to justify altered energy potentials or dynamic processes. It may be used to define an implicit representation of the solvent.

In an aqueous solution it appears that a large number of factors contribute to the stability of a protein, but none of them seems to cause a major effect. However, they seem often related with the solvent mobility:

(i) the value which is used for the dielectric constant has a strong influence on the dynamics

(with a dielectric constant $\varepsilon_o = 3$, the $\langle H_\alpha \rangle$ error is about 220 ppb instead of greater than 240 ppb otherwise).

(ii) the way by which the dynamics process is carried out is also important. The backbone flexibility is better simulated when only a random reorientation of the side chains is considered to perform the thermal motion (with an $\langle H_\alpha \rangle$ error equal to 200 ppb).

(iii) finally, the use of a different weighting scheme between the polar and non-polar atoms (for instance by increasing the weight of the hydrophobic alkyl and aryl hydrogen atoms) improves the molecular dynamics (with a correspondent $\langle H_\alpha \rangle$ error as low as 188 ppb).

This paper deals with the dynamics of a protein, but the conclusion may be different to account for the dynamics of a polypeptide chain. For small peptides, the hydrophobic interactions do not have the same importance as for globular proteins. A more extensive analysis is required before a general rule could be proposed.

REFERENCES

1. Karplus M, Petsko GA. Molecular dynamics simulation in biology. *Nature* 1990; **347**: 631–639.
2. Van Buuren AR, Berendsen JC. Molecular dynamics simulation of the stability of a 22-residue α-helix in water and 30% trifluoroethanol. *Biopolymers* 1993; **33**: 1159–1166.
3. Daura X, Baldomero O, Querol E, Aviles FX, Tapia O. On the sensitivity of MD trajectories to changes in water–protein interaction parameters: the potato carboxypeptidase inhibitor in water as a test case for the GROMOS force field. *Proteins: Struct. Funct. Genet.* 1996; **25**: 89–103.
4. Schutz CN, Warshel A. What are the dielectric 'constants' of proteins and how to validate electrostatic models. *Proteins: Struct. Funct. Genet.* 2001; **44**: 400–417.
5. Hagler AT, Lifson S. A procedure for obtaining energy parameters from crystal packing. *Acta Crystallogr.* 1974; **B30**: 1336–1341.
6. Jorgensen WL, Tirado R. The OPLS potential functions for proteins: energy minimization for crystals of cyclic peptide and crambin. *J. Am. Chem. Soc.* 1988; **110**: 1657–1669.
7. Maple JR, Hwang MJ, Jalkanen KJ, Stockfisch TP, Hagler AT. Derivation of class II force fields: V-quantum force field for amides, peptides and related compounds. *J. Comput. Chem.* 1998; **19**: 430–458.
8. Kim RM, Evans HJ. Molecular modelling of proteins: a strategy for energy minimisation by molecular mechanics in the AMBER force field. *J. Biomol. Struct. Dyn.* 1991; **9**: 475–488.
9. Ma B, Nussinov R. Explicit and implicit water simulations of a β-hairpin peptide. *Proteins: Struct. Funct. Genet.* 1999; **37**: 73–87.
10. Casari G, Sippl MJ. Structure derived hydrophobic potential. Hydrophobic potential derived from x-ray structures of globular proteins is able to identify native folds. *J. Mol. Biol.* 1992; **224**: 725–732.
11. Gibbs N, Clarke AR, Sessions RB. Ab initio protein structure prediction using physicochemical potentials and a simplified off-lattice model. *Proteins: Struct. Funct. Genet.* 2001; **43**: 186–202.
12. Wagner D, Braun W, Havel TF, Schaumann T, Go N, Wüthrich K. Protein structures in solution by nuclear magnetic resonance and distance geometry: the polypeptide fold of the basic pancreatic trypsin inhibitor determined using two different algorithms DISGEO and DISMAN. *J. Mol. Biol.* 1987; **196**: 611–639.
13. Bernstein FC, Koetzle TF, Williams GJB, Meyer EF, Brice MD, Rodgers R, Kennard O, Shimanouchi T, Tasumi M. The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* 1977; **112**: 535–542.
14. Brunger AT. *X-Plor, Version 3.8 Manual.* Yale University, New Haven, CT, 1992.
15. Lifson S, Hagler AT, Dauber P. Consistent force field studies of intramolecular forces in hydrogen-bonded crystals. I. Carboxylic acids, amides and the C=O$\cdots$H hydrogen bonds. *J. Am. Chem. Soc.* 1979; **101**: 5111–5121.
16. Gelin BR, Karplus M. Side chain torsional potentials: effect of dipeptide, protein and solvent environment. *Biochemistry* 1979; **18**: 1256–1268.
17. Tuffery P, Etchebest C, Hazout S, Lavery R. A new approach to rapid determination of protein side chain conformations. *J. Biomol. Struct. Dyn.* 1991; **8**: 1267–1280.
18. Osapay K, Case DA. A new analysis of proton chemical shifts in proteins. *J. Am. Chem. Soc.* 1991; **113**: 9436–9444.
19. Bundi A, Wüthrich K. ^{1}H NMR parameters of the common amino-acid residues measured in aqueous solutions of the linear tetrapeptide H-Gly-Gly-X-Ala-OH. *Biopolymers* 1979; **18**: 285–297.
20. Busetta B, Picard Ph, Precigoux G. Influence of thermal motion on ^{1}H chemical shifts in proteins: the case of bovine pancreatic trypsin inhibitor. *J. Peptide Sci.* 2001; **7**: 121–127.
21. Wishart DS, Sykes BD, Richards FM. The chemical shift index: a fast and simple method for the assignment of proton secondary structure through NMR spectroscopy. *Biochemistry* 1992; **31**: 1647–1651.

22. Williamson MP, Asakura T, Nakamura E, Demura M. A method for the calculation of protein α-CH chemical shifts. *J. Biomol. NMR* 1992; **2**: 83–98.
23. Berndt KD, Güntert P, Orbons PM, Wüthrich K. Determination of a high-quality nuclear magnetic resonance solution structure of the bovine pancreatic trypsin inhibitor and comparison with three crystal structures. *J. Mol. Biol.* 1992; **227**: 757–775.
24. Sharp KA, Honig B. Electrostatic interaction in macromolecules: Theory and applications. *Annu. Rev. Biophys. Biochem.* 1990; **19**: 301–332.

