Journal of Peptide Science
J. Peptide Sci. 2006; **12**: 33–42
Published online 1 August 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/psc.694

JOURNAL OF PEPTIDE SCIENCE

Simulation of oligopeptide dynamics and folding. The use of NMR chemical shifts to analyse the MD trajectories

BERNARD BUSETTA, PHILIPPE PICARD* and GILLES PRÉCIGOUX

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

Received 14 March 2005; *Accepted 27 May 2005*

Abstract: In this paper, a simulation of the folding process, based on a random perturbations of the ϕ, ψ, χ_1 dihedral angles, is proposed to approach the formation at the atom level of both principal elements of protein secondary structure, the α-helix and the β-hairpin structures. Expecting to understand what may happen in solution during the formation of such structures, the behaviour of large sets of random conformations that are generated for small oligopeptides was analysed. Different factors that may influence the folding (as conformational propensity, hydrophobic interactions and side-chain mobility) were investigated. The difference between the corresponding theoretical folding and the real conformational diversity that is observed in solution is appraised by a comparison between the calculated and observed NMR secondary chemical shifts. From this study it appears that hydrophobic interactions and mobility represent the principal factors that initiate folding and determine the observed hydrogen-bond pattern, which subsequently allows packing between the peptide side chains.

Keywords: folding simulation; implicit solvent representation

INTRODUCTION

Up to now, semi-empirical force-fields associated with simplified geometric representations of the macromolecules and simulations of the solvent effects were used to realise an *ab initio* determination of more or less complex peptide and protein structures [1,2]. Both kinds of simplifications are necessary to develop a statistical treatment of the problem with well-defined x-ray diffraction structures. Such a statistical approach does not allow the conformational analysis of peptides that contain 'exotic elements' and is limited to naturally occurring peptides. At the atom level, even for a small peptide, a huge conformational space should be explored to perform an explicit molecular folding. After a very long time, theoretical folding can be realised from a hypothetical extended structure [3] or after the analysis of a large number of initial random structures [1]. The computing time is drastically increased if the solvent is to be explicitly defined so that the behaviour of complex peptides in solution cannot be easily described.

To save computing time, it is necessary to define a suitable implicit representation of the solvent that allows the determination of the possible conformation(s) of a peptide by a simple refinement. A complete analysis of the forces that are critical for the determination of backbone conformation and their dynamical behaviour in solution should be performed. In this study, to limit the discrepancies that would be involved by too important simplifications of the dynamics or folding simulation, we kept unchanged the all-atom geometry of the macromolecule and tried to interpret at the atomic level some characteristics that were defined from statistical analysis at the residual level. We study how the formation of a hydrophobic core or the realisation of the native hydrogen-bond pattern ensures the packing of the peptide side-chains and try to find an equivalence to the 'hydrophobic in, polar out' model that controls protein folding [4].

As a statistical treatment of the internal force field to mimic the solvent effects is impossible, it is expected that the structural diversity present in solution and recorded via the NMR chemical shifts may provide the same valuable information as a statistical treatment of a large amount of x-ray diffraction structures. The observed proton chemical shifts define the dynamics of each proton and may be used to appraise the relevancy of a simulation by the comparison of the observed chemical shifts with the shifts estimated from a mixture of generated conformations. This approach was already used [5] to prove that a random perturbation of the χ_1 dihedral angles allows a fast exploration of the energetic space and ensures a dynamics (χ_1 dynamics) 10^3- to 10^4-fold faster than the classical Verlet–Langevin dynamics.

The non-bonded energies are expected to be valid in the solution state if the solvent is explicitly defined. The solvent effect corresponds to an additive E_{solv} term. When the solvent molecules are not defined, force fields may be designed for 'effective' aqueous solvent calculations that are actually performed *in vacuo*. A suitable parameterisation of the electrostatic and van der Waals interactions at the level of the hydrogen

* Correspondence to: P. 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

bond is necessary. In addition, some extra distance restraints between other non-bonded atoms should be used to give a complete, implicit simulation of the solvent effect.

As the solution state is an average of conformers, the least-square refinement, which was used with crystal structures to define the energy parameters [6], is impossible. The chemical shifts are used to define the appropriate balance among the internal geometrical terms, the electrostatic interactions and the external non-bonded interactions in an implicit solvent representation.

METHODS

Estimation of the ^{1}H Chemical Shift Deviations

Following the process developed by Osapay and Case [7], the chemical shift of each ^{1}H protons i is estimated from a peptide 3D conformation J as the sum of elementary contributions:

$$\delta_{i,J} = \delta_{\text{local}} + \delta_{\text{el}} + \delta_{\text{m}} + \delta_{\text{rc}}$$

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 [8];
δ_{el} is the electrostatic contribution restricted to the contribution of the main chain;
δ_{m} is the magnetic contribution of peptide groups and is strongly dependent on the angle made by each proton with the different peptide groups;
δ_{rc} is the aromatic ring contribution (i.e. the ring current effect).

The basic equation was modified to take into account the local thermal fluctuations [9]:

$$\delta_J = \delta_{\text{local}} + \text{w}_J*(\delta_{\text{el}} + r_{\text{d}}*\delta_{\text{m}} + r_{\text{d}}*\delta_{\text{rc}}) = \delta_{\text{local}} + \text{w}_J*\delta_{i,J}(r_{\text{d}})$$

where the electrostatic contribution is not affected but where fluctuations around the ϕ, ψ, χ_1 angles involve a randomisation (r_{d}) of the magnetic and the aromatic ring contributions. A conformation is then replaced by a set of conformers which fluctuate around it. The randomisation factor r_{d} should be defined for each magnetic or aromatic contributor. When several conformations are possible w_J is the amount of the conformation J that provides the best fit between the calculated and the observed shifts. A Simplex process was used to refine these different parameters.

The importance of the ring current effect from each aromatic cycle ($\text{w}_{J*}r_{\text{d}}$) was proven to depend on its relative exposure to the solvent [9]. As r_{d} was expected to be unity for a completely buried aromatic ring (that does not move), the refined contribution of such rings gives then a good estimate of the weight w_J. Otherwise, only the product $\text{w}_{J*}r_{\text{d}}$ is available and it is impossible to distinguish the estimation of the thermal vibration from the relevancy of the conformation.

When a single conformation is considered to generate the theoretical solution (in association with other possible conformations that may be considered as totally random) we use w (instead of w_J) that represents the percentage of this conformation in the solution. When a set of random conformations is generated, only the top 100 conformations in energy are considered (see below) to define the theoretical solution with equal unrefined weights $\text{w}_J = 0.01$. The theoretical solution is then assumed to be a mixture of fluctuating conformations J and the resulting calculated chemical shift for a proton i is:

$$\delta_i = \Sigma_J\{\text{w}_J\delta_{i,J}(r_{\text{d}})\}$$

If δ_{iobs} is the observed chemical shift, the conformational difference around proton i may be appraised by the absolute value $|\delta_i - \delta_{iobs}|$. The average on the absolute difference:

$$\langle \text{H}_\alpha \rangle = 1/n_{\text{p}}\Sigma|\delta_i - \delta_{iobs}|$$

is computed on all the n_{p} H_α protons (referred to as $\langle \text{H}_\alpha \rangle$ error for the theoretical solution based on n_{c} conformations and expressed in ppb (part per billion)). We expressed as $\langle \text{H}_\alpha \rangle_{E<E_{\text{o}}}$ the error that is computed for a theoretical solution based on the top conformations with energy $E < E_{\text{o}}$.

The $\langle \text{H}_\alpha \rangle$ error is sensitive to the backbone conformation (i.e. fluctuations of the overall main-chain geometry defined by the ϕ, ψ, ω dihedral angles) and to the vicinity of the peptide group involved in the formation of hydrogen bonds and secondary structures [10]. The smaller the averaged absolute error, the nearer the theoretical solution will be to the observed one. The chemical shift of the amide proton is the corresponding expression for the H_N protons. It depends on either the bonded state of the hydrogen atom and the effective hydrogen-bond length (bonded H_N proton) or the temperature (non-bonded H_N).

The intramolecular hydrogen bonds are counteracted by intermolecular hydrogen bonds with the solvent. The variation of the chemical shift of a bonded H_N proton results from the difference between the potential intramolecular and intermolecular hydrogen bonds (this last one is already taken into account in the δ_{local} term). If the intramolecular bond is longer, an upfield shift is observed [11]. A negative contribution to the H_N proton chemical shift is observed as long as the distance between the H_N proton and the oxygen atom ($d < 2.7$ Å) prevents a direct accessibility of the H_N proton to the surrounding solvent. This event involves a strong variation of the H_N-proton chemical shift between a solvent- and an intramolecular-bonded state. Then, the H_α proton chemical shifts are more suitable to follow the formation of the intramolecular hydrogen bond.

In the following, a potential 'hydrogen bond' between the H_N proton of an i residue and the carbonyl oxygen of a j residue is considered below this critical distance and will be quoted as an $(i-j)$ hydrogen bond.

Dynamics and Folding

The X-Plor 3.8 refinement package [12] was used for all computations. The molecular energy is described by an empirical energy function, which assumes that the minimum real energy can be issued by a mathematical minimisation. For each atom-type the associated force field (parmallh3x.pro

file of the X-Plor package) provides the standard geometry and the non-bonded well depth.

Although classical molecular dynamics can overcome the small energy barriers that rule the side-chain flexibility, it is unlikely that such a methodology could reproduce this flexibility in a short period of time and, for instance, that it could provide a suitably 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 [13]. The reorientation of the side chains involves distortions that are observed on the overall backbone conformation and are responsible for the protein thermal motion. The thermal dynamics may be considered as the result of perturbations applied randomly to the protein side chains.

During the molecular dynamics, following the height of the energy barrier, each side chain may occupy three different orientations defined by the χ_1 dihedral angle around the $C_\alpha - C_\beta$ axis. The motion towards anyone of these three orientations is made possible by the solvent motion. The χ_1 dynamics results from a reorientation of the different side chains at random and is obtained by adding repeatedly to the energy function a transient extra harmonic dihedral restraint

$$K_{\chi 1}\{\chi_1 - (\chi_{1o} - 120 n_{rd})\}^2$$

where χ_{1o} is the initial dihedral angle (expressed in degrees) before the perturbation is applied and n_{rd} is a random number to select one of the three possible orientations χ_1 (some n_{rd} values may involve no restraint). The scaling factor $K_{\chi 1}$ (in Kcal/degree^2) defines the strength of the perturbation.

In each 'folding' cycle $N/8$ randomly selected side chains (N is the total number of protein residues) can move. The random perturbation is applied on the selected chain during 10 cycles of Powell refinement, followed by 100 cycles of relaxation without dihedral restraints.

A similar folding process based on random modifications of the ϕ, ψ or χ_1 dihedral angles was developed to characterise the most important interactions and the local behaviour that may induce a given conformation. To simulate the folding process, a correct definition of the initial state is necessary. The choice of an extended structure does not seem advisable when a β-hairpin structure is the expected final conformation. To generate a random starting set of conformations, the initial propensity of each residue was considered by the choice of two possible conformations:

the α-helix structure ($\phi = -65°$, $\psi = -40°$)
the extended structure ($\phi = -120°$, $\psi = 120°$),

associated with different probability (0.60, 0.40 or 0.40, 0.60) depending on whether the residue is an helix former or not (i.e. in this latter case, the 'β-forked' residues Val, Ile, Thr or Pro).

The initial random conformations are supposed to represent all the possible conformations of a given peptide. Among them, the observed conformations are expected to correspond to the minimum of the energy function. Thus, the comparison of the distribution of the top 100 folded conformations that result from energy minimisation and folding with the observed ones (that result from the observed chemical shifts) was used to appraise the choice of distance restraints to simulate the solvent effect.

As the folding or the dynamics are obtained by random perturbation of the dihedral angles, (nOe-like) distance restraints appear as the best way to allow interactions between side chains that are constantly rotating. To analyse some specific conformational characteristics (as hydrogen-bond formation or some hydrophobic interactions) during the folding simulation, distance restraints (H_{bd} and h_{yd}) are used as an additive harmonic term $K(d - d_o)^2$ where K is an appropriate scale, factors are (K_{bd} or K_{yd}):

(a) as in the prediction of secondary structures, reinforcements that take into account the interactions between hydrophobic side chains (Leu, Met, Phe, Tyr, Trp, Val, Ile) at positions i and $i+2$ ($i+3$ or $i+4$) were proven to improve the initial conformational propensity of the primary structure [14]. To check this statement at the atom level, hydrophobic restraints were introduced to maintain the corresponding C_β atoms in contact ($d < 5$ Å).
(b) to check the conformational bias provided by the knowledge of the hydrogen bonds (that results from prediction, homology or – as shown below – the early steps of folding) restraints that correspond to the hydrogen-bond pattern are used ($d < 2$ Å)

Alternatively, the propensity for the extended structure of β-forked dipeptides (such as Val–Thr or Ile–Thr) may be assumed and favoured by appropriate distance-restraints between adjacent C_β atoms.

RESULTS AND DISCUSSION

Chemical Shift Estimation of BPTI

Our previous analysis on the bovine pancreatic trypsin inhibitor (BPTI) (file *5pti* in the Protein Data Bank [15]) was based on the variations of the estimated chemical shifts of the H_α protons. As the H_N chemical shifts are recorded at the same temperature as the random coil ones, the temperature effect is not to be taken into account. The use of the x-ray diffraction structure was preferred because it provides a better approach of the chemical shifts than any nOe-derived structure [16]. The observed H_α chemical shifts [17] differ from the 'random coil' ones by $\langle H_\alpha \rangle = 370$ ppb on an average. A set of conformations generated with a classical Verlet–Langevin dynamics, without any explicit or implicit representation of the solvent molecules, provides a theoretical solution that corresponds to $\langle H\alpha \rangle = 205$ ppb. With the χ_1 dynamics and an implicit solvent representation (where the well-depth of the hydrogen atoms of the non-polar side chains is increased [5]) the solution structure of BPTI is improved, as proven by a lower error $\langle H_\alpha \rangle = 180$ ppb.

The overall conformation may remain the same (with no important variations of $\langle H_\alpha \rangle$ whereas strong variations of the hydrogen-bond lengths and the H_N chemical shifts occur. In Figure 1 are represented the variations of the $\langle H_N \rangle$ error during 500 cycles of χ_1 dynamics (averaged on successive sets of 10 cycles)

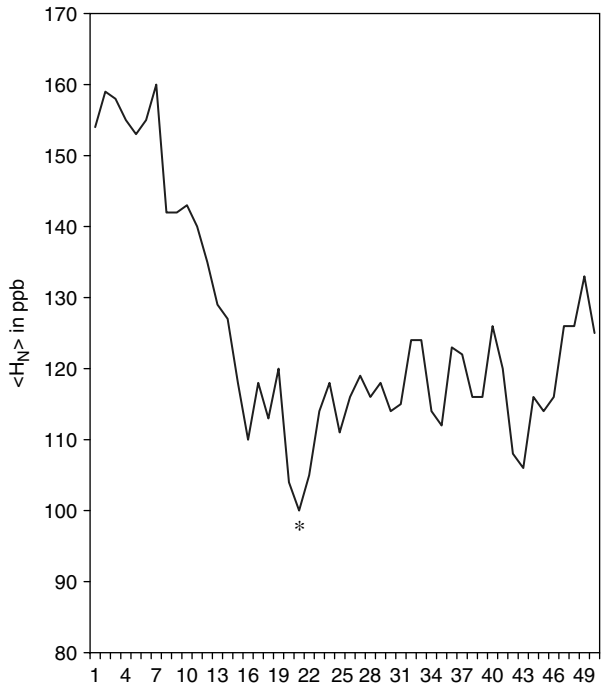

Figure 1 Variations of the $\langle H_N \rangle$ error on the H_N chemical shifts, during 500 dynamics cycles applied to the x-ray diffraction structure of BPTI. The $\langle H_N \rangle$ error is reported every 10 cycles (in abscissa) and corresponds to the average of the $\langle H_N \rangle$ error during 10 consecutive cycles. The resulting error on this average is about 10 ppb. The randomisation of the conformation ($_*$) that corresponds to the minimum H_N value provides a theoretical solution with $\langle H_\alpha \rangle = 173$ ppb and $\langle H_N \rangle = 132$ ppb.

developed from the BPTI x-ray diffraction structure. The variations of the $\langle H_N \rangle$ error is compatible with a relaxation of the crystal conformation towards a conformation more compatible with the solution state. Improvement of the $\langle H_N \rangle$ determination proved that the simulation of the solvent effects allows a suitable definition of the hydrogen-bond lengths in the solution state. Afterwards, only fluctuations around an average conformation occur. The conformer for which $\langle H_N \rangle$ is minimum differs from that of the x-ray diffraction structure by r.m.s. $d = 0.035$ Å on the backbone atoms. The corresponding scaling factor is of the same order of magnitude ($w_* r_d = 1.30$, with $\langle H_\alpha \rangle = 173$ ppb and $\langle H_N \rangle = 132$ ppb) as the relative weights that were obtained for the buried aromatic cycles [12]. This observation suggests that $w = 1.30$ and $r_d = 1$. Then, only this conformational entity is supposed to be present in solution. The estimation of chemical shifts from this single conformation with a theoretical simulation of the thermal vibrations via an appropriate randomisation factor is more accurate than an estimate obtained by averaging the chemical shifts over all the conformers generated during the dynamics. The simulation of the solvent effect by strengthened van der Waals interactions at the level of the aliphatic hydrogen atoms is sufficient to generate 3D-structures for which the definition of the hydrogen-bond lengths agrees with the effective bond lengths in solution (recorded via the H_N chemical shifts).

The (Un)folding of an α-Helix Peptide

The dynamic process was applied to the potential α-helix peptide of 15 amino acids (KETAAAKAERQAMDS), which was derived from the *S*-peptide of ribonuclease A by substituting all of the aromatic residues by Ala [18] (referred further as hel). The unfolding process was performed using the coordinates that may be derived from the x-ray diffraction structure (file *2rn* in the Protein Data Bank). This peptide may be defined as a pure α-helix structure with a hydrogen-bond pattern $\{(i + 4 \rightarrow i), 2 < i < 10\}$. Without any refinement, the conformer derived from the x-ray diffraction structure gives for the H_α chemical shifts an error $\langle H_\alpha \rangle = 82$ ppb with a $w_* r_d = 0.53$ factor.

When 1000 cycles of χ_1 dynamics are applied to the x-ray diffraction conformer (Figure 2), after 10 relaxation cycles, the fluctuations around the initial conformer are not important (ΔE less than 2.7 kcal/mol). The geometry of the average conformer is improved as suggested by the chemical shift estimation ($w_* r_d = 0.65$ and $\langle H_\alpha \rangle = 48$ ppb). During the unfolding process, the H_α chemical shifts are consistent with a theoretical solution state containing mainly the observed x-ray α-helix (or elements of α-helix). The weak $w_* r_d$ factor may be explained either by the existence of a fluctuating helical structure (an important thermal motion and a weak r_d randomisation) or of about 35% of completely

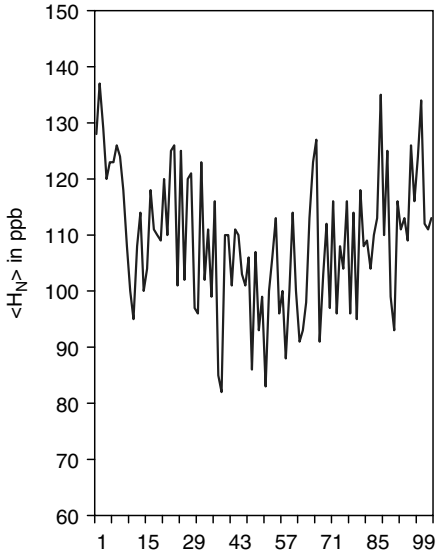

Figure 2 Variations of the $\langle H_N \rangle$ error on the H_N chemical shifts during 1000 dynamic cycles applied to the modified *N*-terminal α-helix derived from the x-ray diffraction structure of ribonuclase A (hel), with the same representation as in Figure 1.

random structures. However, it may be assumed that some unfolded conformers are not attainable and the theoretical mixture conserves a memory of the initial structure.

To obtain another view of the molecular dynamics of this α-helical peptide, we consider the behaviour of 250 initial random (ϕ, ψ) conformations that are submitted to 200 'folding cycles'. The influence of restraints based on the final hydrogen-bond pattern is analysed by increasing the K_{bd} weight of the hydrogen-bond restraints (we expected a 'snapshot' of the conformations that are stabilised by the formation of hydrogen bonds).

–with no restraint at all ($K_{bd} = 0$ kcal/Å^2/mol), the scarce H-bonded conformers correspond always to β-turn structures (10% are 5–2, 7–4 and 2% 11–8) and $\langle H_\alpha \rangle$ for the theoretical mixture is 85 ppb.

–with medium restraints ($K_{bd} = 5$ kcal/Å^2/mol), polyturns (5–2, 7–4, 10–7, 11–8, 13–10) are always preponderant and even a complete 3_{10}-helix was observed. The theoretical mixture corresponds to $\langle H_\alpha \rangle = 71$ ppb.

–with strong restraints ($K_{bd} = 20$ kcal/Å^2/mol), 80% of the conformations present an α-helical structure between residues 7 and 13, but some (7–4) hydrogen bonds are still observed. The theoretical mixture corresponds to $\langle H_\alpha \rangle = 67$ ppb.

With medium restraints, the set of the top 100 folded conformations deviates from the native conformation by 2.50 Å on the average and the nearest conformation by 1.83 Å. After 1000 extra 'folding cycles' without any restraint ($K_{bd} = 0$), this nearest conformation provides a native-like α-helix (backbone r.m.s $d = 0.85$ Å). A similar conformation (backbone r.m.s $d = 1.06$ Å) is obtained faster (in only 200 cycles) with stronger restraint ($K_{bd} = 5$ kcal/Å^2/mol). The use of distance restraints only accelerates the folding process. The top four conformations in energy correspond to the native helix. In Table 1 we have summed up the variety of conformations that are obtained, respectively, after 100 and 300 'folding cycles'. In this summary, to describe the conformational behaviour of the polypeptide chain, we have to consider two distinct parts of the chain: i.e. residues 3 to 9 and 10 to 14. After 100 cycles the 3_{10}-helix form is largely preponderant, but more than half of the conformations have evolved to the α-helix after 300 'folding cycles'.

Table 1 Evolution of the proportion of pseudo-helical structures during a folding experiment of the hel peptide (15 amino acids), after 100 and 300 folding cycles. To allow a clearer description of the polypeptide chain, two portions (3–9 and 10–14) that may be either 3_{10}- or α-helices are considered

n_{cycle}	3_{10}-3_{10}	3_{10}-α	β -3_{10}	α-α
100	50	48		
300	10	44	1	45

The application of strong restraints involves the formation of a small α_D-helix at the level of the $(Ala)_3$ triplet (about 3% of the generated conformations), which is not compatible ($\langle H_\alpha \rangle = 113$ ppb) with the observed chemical shifts. The system appears to evolve from a majority of 3_{10}-helix-like structures towards a majority of α-helices. As the hydrophobic side-chains of this peptide are too distant to experience any solvent influence, the helix propensities of each residue along the hel chain are sufficient to involve the α-helix structure and no extra hydrophobic restraints are necessary.

The Folding of a β-Hairpin Peptide

The 12-amino acid peptide (bh8) RGITVNGKTYGR was described as a potential β-hairpin [19] and characterised by four hypothetical hydrogen bonds (10–3, 8–5, 5–8, 3–10). The difference between the observed chemical shifts and the random coil ones is $\langle H_\alpha \rangle = 102$ ppb. During this study, three different nOe-like restraints are used:

(a) the nOe restraints are the main-chain nOe restraints reported in reference [19],
(b) the hydrogen-bond restraints (H_{bd}) refer to the four standard hydrogen-bond lengths in the hairpin structure (i.e. 3–10, 5–8, 8–5, 10–3 hydrogen bonds),
(c) the hydrophobic restraints are four nOe-like restraints (h_{yd}) between the C_β atoms of the side chains of residue 3, 5, 8, 10 that are maintained within 5 Å.

In all cases, a β-turn at the level of the Asn–Gly residues was favoured by a (8–5) hydrogen-bond restraint without any consideration of the type of this turn that may depend on the nature of the adjacent residues.

The energy of the conformations that are generated in different folding processes cannot be directly compared. To appraise the improvement involved by the incorporation of some distance restraints, it seems better to follow the formation of hydrogen bonds and the $\langle H_\alpha \rangle$ error in a mixture of conformations. Figure 3 sums up the behaviour of 500 random (ϕ, ψ conformers when they are submitted to different restraints. In each case, the top 100 average conformations are considered avoiding those where too strong atom-overlaps occur.

Following the importance that is given (via the K_{bd}, K_{yd} and $K_{\chi 1}$ scale factors) to different elements that may affect the folding, different conformations are generated for the bh8 peptide. If no dynamics is used in the search of the bh8 conformations, the resulting set depends only on the initial random determination. When a non-zero dynamics is applied ($K_{\chi^1} = 10$ kcal/Å^2/mol), there

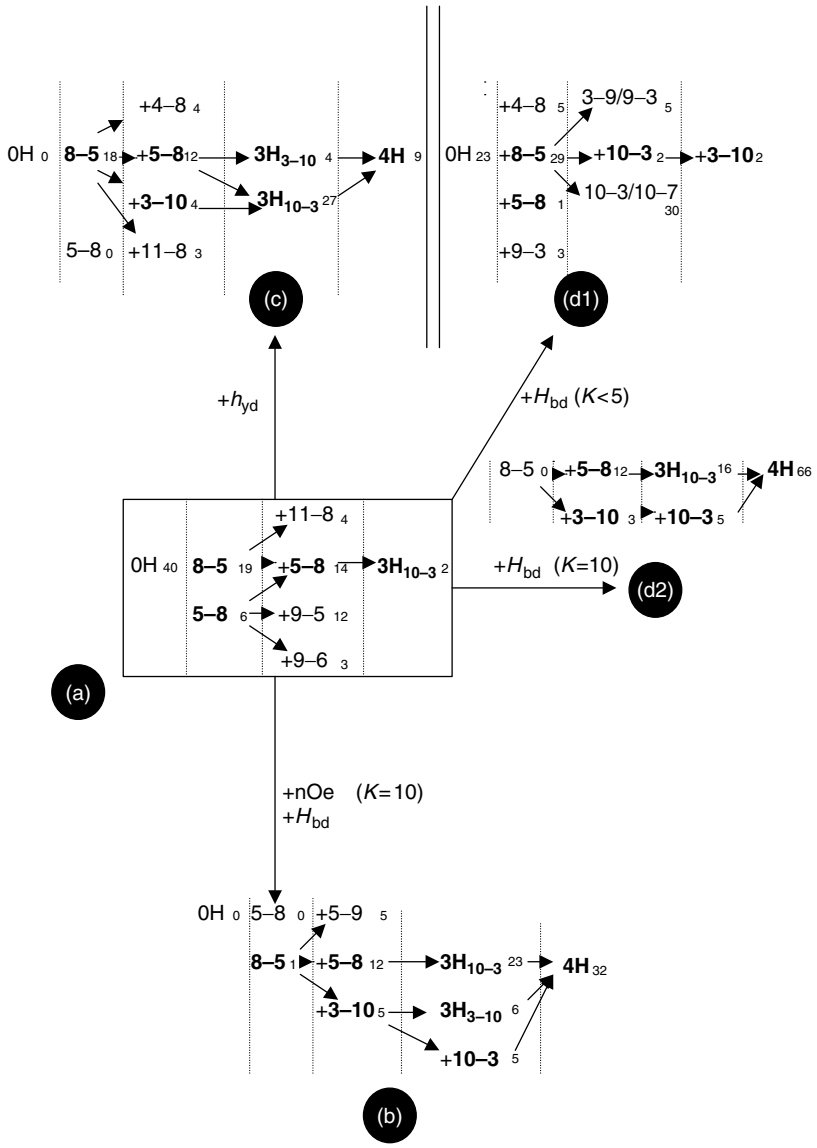

Figure 3 Formation of the hydrogen bonds in 500 random structures of the bh8 potential β-hairpin peptide after 600 'folding cycles' (see *Methods*), based on different restraints, and expressed as the number of conformations (in small caps) among the top 100 conformations: (a) with the β-turn generating hydrogen-bond restraints, (b) with the observed nOe and hydrogen-bond restraints, (c) with hydrophobic restraints, (d_1,d_2) with restraints that correspond to the observed hydrogen-bond pattern associated to different weights. In each summary the respective number of conformations with one to four hydrogen bonds from left to right are reported and seperated by a dashed line. $3H_{x-y}$ is a conformation for which the x–y hydrogen bond is formed beside the central (5–8, 8–5) bonds and 4H is the complete four hydrogen-bonded β-hairpin. The notation 10–3/10–7 corresponds to a set of intermediate conformations in which the potential hydrogen bonds favour either the β-hairpin (10–3 shorter) or the 3_{10}-helix (10–7). The hydrogen bonds that are compatible to the observed β-hairpin are indicated in bold face. The scale factors K are expressed in kcal/Å^2/mol.

is an accumulation towards specific conformations independently of the initial random choice.

In Figure 3(a) only the turn involved by the Asn-Gly moiety is initiated by an appropriate (8–5) hydrogen-bond restraint associated with a scale factor K_{8-5} = 50 kcal/Å^2/mol applied to the random conformations during 600 'folding cycles'. With such a restraint, the random (ϕ, ψ) definition and the molecular geometry remain preponderant during the following folding process. The (8–5) hydrogen bond appears in 35% of the top 'average conformations' but the occurrence of a hairpin-like structure (5–8, 8–5, 10–3) is only observed in 2% of the conformations. With side chains that are always rotating, the current force field is not able to generate a defined structure in the absence of any solvent simulation. A multi- (3_{10}-helix like)turn conformation (8–5, 11–8) appears in 4% of the generated conformations.

In Figure 3(b), the nOe restraints observed for the main chain are associated with the hydrogen-bond

Table 2 The top seven conformations (within $\Delta E_{VdW} = 2$ kcal/mol) of the bh8 peptide using the observed nOe restraints and the expected hydrogen bonds as distance restraints. ΔE (in kcal/mol) is the variation of the VdW energy from the minimum conformation, ΔnOe (in Å) is the average error on the observed nOe restraints, $w*r_d$ is the randomisation factor and $\langle H_\alpha \rangle$ (in ppb) the correspondent error on the H_α chemical shifts. The top five conformations define the theoretical solution that is the nearest to the observed ones

	ΔE	Δ_{nOe}	$w_* r_d$	$\langle H_\alpha \rangle$
1	0.00	0.260	0.40	61
2	0.45	0.215	0.00	61
3	1.05	0.245	0.55	59
4	1.47	0.280	0.03	54
5	1.79	0.225	0.19	71
6	1.86	0.263	0.40	72
7	1.87	0.209	0.84	68

restraints H_{bd} ($K_{nOe} = K_{bd} = 10$ kcal/Å^2/mol) and applied during 600 'folding cycles' to the same random sample. The first purpose of the folding process under these conditions is to check whether an accumulation towards the correct conformation can be obtained by this unusual folding (i.e. by random perturbations of the dihedral angles). In Table 2 are listed the top conformations (with $\Delta E < 2$ kcal/mol vs E_{min}). The corresponding refined $w_* r_d$ factor associated with each conformation may be considered as an estimate of the amount of this unique conformation (see *Methods*) that is necessary in a 'theoretical' solution to provide the best fit to the observed solution [some solutions in agreement with the nOe restraints may have no apparent contribution to the observed chemical shifts ($w_* r_d \neq 0$); both kinds of information should be used to define the solution state].

For a 'theoretical' solution composed of the top 100 generated conformations, the $\langle H_\alpha \rangle$ error is 116 ppb. The 'theoretical' solution defined by the top five conformations (that differ by less than 1.10 Å and 1.05 kcal/mol) provides a better approach of the observed ones (with $w_* r_d = 1.30$ and $\langle H_\alpha \rangle = 57$ ppb). As expected, the set of the remaining intermediate conformations does not seem to contribute to the observed shifts and may behave as a set of random coil structures ($\langle \delta_H\alpha - \delta_H\alpha$ local$\rangle = 40$ ppb) rather than of β-hairpins. The observed chemical shifts may be explained by the fluctuations of few preponderant β-hairpin structures (the top five conformations) in the presence of other structures that behave as random ones on an average.

In Figure 3(c) the folding results only from distance restraints between the hydrophobic moieties of the side chains. After the same number of cycles, different steps of the folding towards a β-hairpin structure are present. The hydrophobic distribution along the peptide chain involves the formation of the observed hydrogen-bond pattern, but the resulting backbone conformation is distorted and the $\langle H_\alpha \rangle$ error is similar to the random coil ones ($\langle H_\alpha \rangle = 106$ ppb).

On the contrary, in Figure 3(d_1,d_2), after the same folding time, the formation of a folded structure is analysed when different energies are associated with this observed hydrogen-bond pattern without any hydrophobic reinforcement. Usually the effect of hydrogen bonds is taken implicitly into account by appropriate parameterisation of the partial charges and van der Waals parameters [12]. During the formation of a hydrogen bond, a progressive polarisation of the peptide [9,20] and H–N bond (as shown by the variation of the H_N proton chemical shift) involves a steady modification of the partial charges. To simulate this variation we consider both extreme situations (strong and no or weak hydrogen bonds).

In Figure 3(d_1) a medium weight ($K_{bd} < 5$ kcal/Å^2/mol) is associated with the hydrogen-bond restraints (i.e. the same strategy that was used with the hel peptide). Only the initial conformational propensity is assumed to involve folding. This case may occur in a hydrophobic environment where the hydrophobic interactions are not influential. A large proportion of conformations remain completely unfolded (with no hydrogen bond or only with the 8–5 bond). The assumption of the observed hydrogen-bond pattern is not sufficient to involve the observed preponderant conformations. The constant rotation of the side chains prevents from any kind of packing and formation of the observed hydrogen-bond pattern. The corresponding theoretical solution is rather far from the observed solution with a $\langle H_\alpha \rangle$ error equal to 120 ppb.

The difficulty to form a 10–3 hydrogen bond is illustrated by the number of conformations that display all the possibilities between the 10–3 and 10–7 (pseudo) hydrogen-bond lengths (i.e. such as 1.5 Å $< d_{10-3}, d_{10-7} <$ 2.7 Å and $d_{10-3} + d_{10-7} = 4.2$ Å). In all these conformations the C-terminus of the potential β-hairpin is found in a position above the N-terminus and suggests the formation of a 3_{10}-helix coiling (Figure 4). In the absence of any hydrophobic interaction and in spite of a favourable initial propensity, a transition is possible from a β-hairpin structure (8–5, 10–3) towards a transient poly-turn 3_{10}-helixlike structure (8–5, 10–7) (the first step to form an α-helix as indicated previously). A modification of the type of the central turn may be associated with this transition. A complete unfolding is not necessary. The transition may occur by a progressive redistribution of the hydrogen bonds, and helices may be converted into β-strand structures in a compact environment (e.g. as it occurs in a molten globule).

In Figure 3(d_2) the hydrogen-bond pattern that may result from the distribution of the hydrophobic side

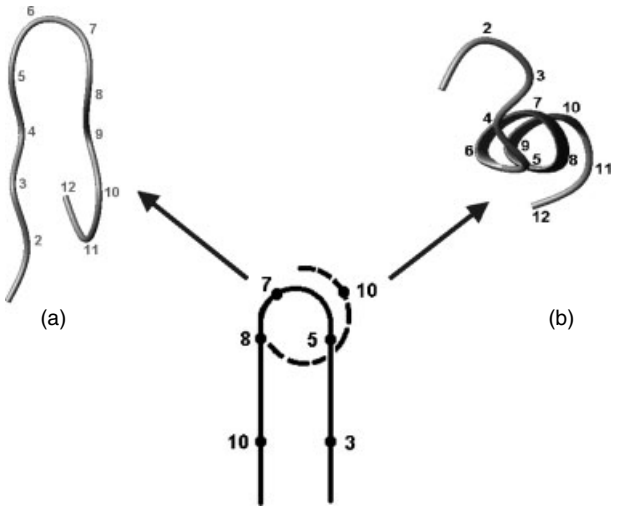

Figure 4 Illustration of the final bh8 conformations that result from the equilibrium between: (a) the β-hairpin structure (with the use of hydrophobic reinforcements) and (b) the 3_{10}-helix structure (with only initial conformational propensities).

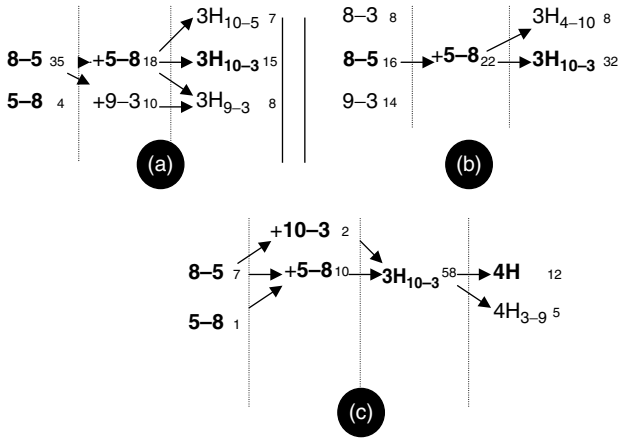

Figure 5 Study of the different conformational features of Lys_8 that are responsible of the formation of the bh8 β-hairpin structure (with the same representation as in Figure 3): (a) when no extra character is associated with Lys residue, (b) when Lys is identified to a β-forked residue, (c) when the aliphatic part of the Lys side chain is taken into account.

chains in bh8 (Figure 3c) is strongly imposed ($K =$ 10 kcal/Å^2/mol). This extra interaction energy is strong enough to allow the achievement of the side-chain packing that involves the final backbone conformation (note that with the use of a simplified geometry the realisation of a suitable side-chain packing that appears as a very important step of the folding is completely ignored). At the end of the folding the $\langle H_\alpha \rangle$ error associated with the theoretical solution is only 78 ppb. The contribution of the hydrogen bonds is necessary to involve the observed conformation of the backbone.

By the way, as strong hydrogen bonds allow a suitable packing of the side chains, it would be necessary to follow such an approach in the case of (distantly) related proteins – the 'twilight zone' – or in the case of a predicted 3D-topology when side-chain packing seems a major problem [21], to force the side chains to pack together and provide an exploitable 3D-structure from a known 2D–3D topology.

Finally, we investigated the role of the hydrophilic Lys residue at position 8 in the folding process and try to justify at the atom level its exposed location. (In the β-hairpins, according to the x-ray diffraction protein structures, charged residues, especially the positive ones, are particularly abundant in the last position of the turn). In our approach, the presence or absence of a positive charge does not seem to have any influence on the folding. As polar residues may be bound to water, they can follow the motion of the solvent molecules. Then, they may be assumed as being freely rotating and their rotation is favoured by the existence of the turn with the Gly residue in the third position. In such a location, the time-averaged Lys residue behaves as a β-forked residue.

In addition to the situation presented in Figure 3(c) (that corresponds to a hydrophobic Lys side chain), in Figure 5 is illustrated the influence of other Lys characteristics on the formation of the hydrogen bonds in the bh8 peptide after the same folding time:

(a) when no distance-restraint (neither hydrophobic nor β-forked characteristics) is applied to Lys, that is, then dealt only as it is defined from the current force field;
(b) when Lys is supposed to be freely rotating. Then Lys behaves as a time-averaged β-forked residue and involves a potential extended dipeptide Lys–Thr. Such a result cannot be obtained by a transient random perturbation of the χ_1 dihedral angle but by a permanent distance-restraint that emulates a time-averaged occupancy;
(c) when in addition the apolar portion of the side chain is taken into account by appropriate C_β–C_β restraints. In this case, a particular role is attributed to Lys that did not appear in shorter, charged residues such as Asp or Glu, and the $\langle H_\alpha \rangle$ error is 97 ppb.

The mobility seems the main characteristics of the hydrophilic residue and depends on the nature and conformation of the adjacent residue. When Lys (and Glu) cannot move, they may be helix formers.

The short bh8 peptide was devised to display a stable β-hairpin structure [18]. This strategy provides a theoretical solution with a high randomisation factor derived from the observed chemical shifts. This structure is the result of: (i) the strong Asn–Gly β-turn, (ii) the multiplication of (potential) extended dipeptides (Ile–Thr, Thr–Val, Lys–Thr), (iii) a suitable distribution

of hydrophobic side chains. These conformational features involve the formation of the observed hydrogen-bond pattern that then ensures the completion of the side-chain packing. The adjacent contributions that stabilised β-strands in large proteins could be replaced by short, stable structures with specific characteristics.

CONCLUSION

Relying on the structural information provided by the NMR chemical shifts, we analysed the folding of an all-atom representation of two small peptides. The NMR chemical shifts provide a good criterion to follow the early steps of the folding and the onset of stable structures. In this paper we studied the formation of both major elements of the protein secondary structure: the α-helix and the β-sheet. Introducing a randomisation factor allows to take into account the thermal motion around a given conformation. Only a small number of very different peptide conformations are sufficient to interpret the information that is contained in the proton chemical shifts and to give an account of the conformational diversity present in solution. For small peptides the estimation of the chemical shifts is more accurate because the side chains may be supposed steadily rotating (for proteins, extra corrections are necessary to take into account that some side chains, Cys, for instance, cannot move).

We show that the dynamics and the folding of a peptide may be simulated by a random perturbation of the dihedral angles. The use of a small number of restraints allows to simulate the solvent effects and to separate events that occur successively during the folding:

(a) Side-chain interactions that were already assumed to be the major event in secondary structure formation [22] are responsible for the formation of the hydrogen-bond pattern that stabilises the peptide structure. The same type of hydrophobic reinforcements that were introduced for the prediction of protein secondary structures [14] can be used to detect the incipient hydrogen bonds. Otherwise, if no hydrophobic interactions between side chains are possible, the initial conformational propensity should be sufficient to ensure the formation of the (secondary) structure (as in the case of the hel peptide). Hydrogen bonds appear when the peptide side chains are still freely rotating.
(b) When this pattern is formed, a main-chain conformation comparable to the observed ones (as proven by the $\langle H_\alpha \rangle$ error) and a suitable side-chain packing are obtained.

Strong hydrogen bonds restrain the side-chain rotation and allow the final packing. The formation of the hydrogen bonds is a co-operative process. To simulate and achieve folding, distance-restraints should be progressively introduced (h_{yd}, and then when the incipient hydrogen bonds appear, $h_{yd} + H_{bd}$).

The mobility is perhaps the main characteristic of the polar side chains (Lys, for instance) that is responsible for their exposed location. A time-averaged occupancy must be considered for highly rotating side chains that behave as 'β-forked' residue.

By the way, a possible path is suggested for the transition that may occur from a β-hairpin structure to an α-helix (and *vice versa*) using a poly-turn or a 3_{10}-helix-like structure as an intermediate. This study shows that following the hydrophilic or hydrophobic nature of the environment, the same peptide may adopt a β-hairpin or an α-helix conformation. In a hydrophobic environment (like any aliphatic solvent) the hydrophobic contacts become negligible and a potential β-hairpin structure may adopt a 3_{10}- and eventually an α-helix conformation.

REFERENCES

1. 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.
2. 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.
3. Santiveri CM, Jimènez CA, Rico M, van Gunsteren WF, Daura X. β-Hairpin folding and stability: molecular dynamics simulations of designed peptides in aqueous solution. *J. Pept. Sci.* 2004; **10**: 546–565.
4. Huang ES, Subbiah S, Levitt M. Recognizing native folds by the arrangement of hydrophobic and polar residues. *J. Mol. Biol.* 1995; **252**: 709–720.
5. Busetta B, Picard PH, Précigoux G. The use of NMR chemical shifts to analyse MD trajectories: simulation of bovine pancreatic inhibitor in water as a test case for solvent influences. *J. Pept. Sci.* 2003; **9**: 450–460.
6. Hagler AT, Lifson S. A procedure for obtaining energy parameters from crystal packing. *Acta Crystallogr.* 1974; **B30**: 1336–1341.
7. Osapay K, Case DA. A new analysis of proton chemical shifts in proteins. *J. Am. Chem. Soc.* 1991; **113**: 9436–9444.
8. 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.
9. Busetta B, Picard PH, Précigoux G. Influence of thermal motion on^{1}H chemical shifts in proteins: the case of bovine pancreatic trypsin inhibitor. *J. Pept. Sci.* 2001; **7**: 121–127.
10. 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.
11. Wishart DS, Sykes BD, Richards FM. Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. *J. Mol. Biol.* 1991; **222**: 311–333.
12. Brunger AT. *X-Plor, Version 3.8 Manual.* Yale University: New Haven, 1992.
13. Tuffery P, Etchebest C, Hazout S, Lavery R. A new approach to rapid determination of protein side chain conformation. *J. Biomol. Struct. Dyn.* 1991; **8**: 1267–1280.
14. Busetta B, Hospital M. An analysis of the prediction of secondary structures. *Biochim. Biophys. Acta* 1982; **701**: 111–118.

15. 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.
16. Williamson MP, Asakura T, Nakamura E, Demura M. A method for the calculation of proteinα-CH chemical shifts. *J. Biomol. NMR* 1992; **2**: 83–98.
17. 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.
18. Jiménez MA, Blanco FJ, Rico M, Santoro J, Herranz J, Nieto L. Periodics properties of proton conformational shifts in isolated protein helices. *Eur. J. Biochem.* 1992; **207**: 39–49.
19. Ramirez-Alvarado M, Blanco FJ, Serrano L. *De novo* design and structural analysis of a model β-hairpin peptide system. *Nat. Struct. Biol.* 1996; **3**: 604–612.
20. Llinas M, Klein MP. Charge relay at the peptide bond; resonance study of solvation effects on the amide electron density distribution. *J. Am. Chem. Soc.* 1975; **9**: 4731–4737.
21. Chang SY, Subbiah S. A structural explanation for the twilight zone of protein sequence homology. *Structure* 1996; **4**: 1123–1128.
22. Wouters MA, Curmi PMG. An analysis of side-chain interactions and pair correlations within antiparallelβ-sheets: the difference between backbone hydrogen-bonded and non-hydrogen-bonded residue pairs. *Proteins* 1998; **22**: 119–131.

