Inter*net* Electronic Journal of Molecular Design

September 2003, Volume 2, Number 9, Pages 564–577

Editor: Ovidiu Ivanciuc

Special issue dedicated to Professor Nenad Trinajstić on the occasion of the 65th birthday
Part 3

Guest Editor: Douglas J. Klein

Exploring the Folding Pathways of Proteins through Energy Landscape Sampling: Application to Alzheimer's β–Amyloid Peptide

Sébastien Santini,[1] Guanghong Wei,[2] Normand Mousseau,[2] and Philippe Derreumaux[1]

[1] Information Génomique et Structurale, CNRS UPR 2589, 31 Chemin Joseph Aiguier, 13402 Marseille Cedex 20, France
[2] Département de physique et GCM, Université de Montréal, C.P. 6128, succ. centre–ville, Montréal (Québec), Canada

Received: July 30, 2003; Accepted: September 8, 2003; Published: September 30, 2003

Citation of the article:
S. Santini, G. Wei, N. Mousseau, and P. Derreumaux, Exploring the Folding Pathways of Proteins through Energy Landscape Sampling: Application to Alzheimer's β–Amyloid Peptide, *Internet Electron. J. Mol. Des.* **2003**, *2*, 564–577, http://www.biochempress.com.

Inter*net* Electronic Journal of Molecular Design

BioChem Press
http://www.biochempress.com

Exploring the Folding Pathways of Proteins through Energy Landscape Sampling: Application to Alzheimer's β–Amyloid Peptide[#]

Sébastien Santini,[1] Guanghong Wei,[2] Normand Mousseau,[2,*] and Philippe Derreumaux[1,*]

[1] Information Génomique et Structurale, CNRS UPR 2589, 31 Chemin Joseph Aiguier, 13402 Marseille Cedex 20, France
[2] Département de physique et GCM, Université de Montréal, C.P. 6128, succ. centre–ville, Montréal (Québec), Canada

Received: July 30, 2003; Accepted: September 8, 2003; Published: September 30, 2003

Internet Electron. J. Mol. Des. **2003**, *2* (9), 564–577

Abstract

The determination of the folding mechanisms of proteins is critical to understand the topological change that can propagate Alzheimer's and prion diseases. The associated folding time scale generally precludes the use of molecular dynamics simulations. Here we present the details of the activation–relaxation simulations using the generic OPEP energy model. We illustrate the strengths of our approach by studying the folding of a dimer of the Alzheimer's β–amyloid peptide.

Keywords. Folding pathway; energy landscape; activation–relaxation technique; β–amyloid peptide; Alzheimer's disease.

1 INTRODUCTION

Exploring efficiently the energy surface is an essential step towards understanding protein folding and molecular recognition as well as designing new drugs or preventing the formation of amyloid fibrils. It is well known, for example, that crystal structures of inactive protein kinases display an inherent plasticity that allows the adoption of distinct conformations in response to interactions with specific proteins [1]. Because we cannot neglect the high frequency vibrational modes in determining protein dynamics, however, the phase space sampled by the longest molecular dynamics (MD) simulation with explicit solvent (of 1 microsecond timescale) remains small compared to what would be needed to describe accurately folding or significant

[#] Dedicated to Professor Nenad Trinajstić on the occasion of the 65th birthday.
* Correspondence authors; E–mail: Normand.Mousseau@umontreal.ca or Philippe.Derreumaux@ibpc.fr.

conformational motion [2].

Recently, many methods have been introduced to overcome this sampling problem. These include multicanonical sampling [3], hundreds of short MD simulations [4] replica exchange simulations [5,6], ensemble dynamics [7,8], and the activation–relaxation technique [9–11]. The multicanonical sampling is based on the determination of a biasing function (non–Boltzmann probability weight factors) which makes it possible to explore all conformations equally. This procedure is a non–trivial problem, but full characterization of the free energy surface has been reported for α and β peptides in monomeric forms [3]. The second approach is to run many short uncoupled MD simulations using different initial conditions. This has been used to study the folding of small peptides using an all–atom model of the protein and an implicit solvation model based on the solvent–accessible surface, but such an energy representation accelerates folding by two orders of magnitude [4]. In replica exchange (or parallel tempering) simulations [5,6], one runs a number of simulation copies at different temperatures, exchanging them at regular intervals according to a Boltzmann criterion based on potential energy difference and the temperatures. Although this method can enhance sampling conformational space, it does not generate folding trajectories, as each temperature's run is discontinuous. By contrast, ensemble dynamics, developed by Pande and collaborators, involves submitting in parallel several tens of independent series at the same temperature, each series comprising numerous coupled MD simulations of only a few tens of nanoseconds. These runs all start from the same structure but with different initial velocities. Simulations are then restarted with different seeds when an energy barrier is crossed in any one of the numerous simulations. The ensemble dynamics scheme, which uses worldwide distributed computing, has been applied to study folding of a 16–residue β–hairpin and a 36–residue three–helix bundle in monomeric forms [7,8]. A major limitation of this procedure is that there is no guarantee that it can explore all dominant pathways [12] and the energy function used is not optimal.

In this paper, we focus on the fifth approach, the activation–relaxation technique (ART). In its original form, ART was applied to various glasses and amorphous semiconductors [13–15]. With the current version, ART nouveau [16], however, it has been possible to apply the algorithm to study a wider range of systems, including Lennard–Jones clusters [16] and proteins [9–11]. The strengths of this approach have been illustrated on two proteins models that had already been studied by a wide range of simulation protocols but for which ART provided new insight. In the first case, we study the folding of a 14–residue peptide model adopting a full α–helix in solution. All ART–generated trajectories at 300 K starting from fully extended conformations point to an obligatory state with two small helices separated by a loop [10], providing a strong support for the picture obtained by MD simulations and other methods on related peptides. In the second case, we look at the second β–hairpin of the domain B1 of protein G. ART simulations at 300 K identify three dominant folding pathways [11], revealing a dynamics much more complex than expected for

such a simple peptide. In the first mechanism, the β–turn of the hairpin is formed first and the structure propagates from there. In the second mechanism, the N– and C–termini first approach each other to form a loop, and then the β–turn is constructed at the final stage [7]. While these two mechanisms have already been identified, no previous method had managed to detect both pathways. The third dominant mechanism discovered by ART can be described by a reptation move of one strand of the β–sheet with respect to the other. This mechanism is particularly interesting because it acquires almost all its native hydrogen bonds (H–bonds) at once and cannot be generated by protein unfolding nor Go–based simulations.

ART can also be used to study multimer systems. We report here on the ART folding trajectories for N–acetyl–Lys–Leu–Val–Phe–Phe–Ala–Glu–NH_2, called $A\beta_{16–22}$ and representing residues 16–22 of the full–length β–amyloid peptide associated with Alzheimer's disease. $A\beta_{16–22}$ is chosen because it is among the shortest fibril–forming β–amyloid fragments yet reported, solid–state NMR spectroscopy reveals an anti–parallel β–sheet configuration of the peptides [17], and it has been investigated by MD simulations with explicit solvent [18,19]. Ma and Nussinov studied the stability of octameric $A\beta_{16–22}$ packed in different arrangements at 330 K [18]; Klimov and Thirumalai simulated the folding process of a trimer of $A\beta_{16–22}$ at 300 K using a bias to facilitate interactions between the peptides [19].

Starting from a parallel conformation, metastable for $A\beta_{16–22}$ but stable for longer chains, we show that (1) the dimer can move from parallel to anti–parallel states and the anti–parallel β–sheet is the native state; (2) contrary to previous simulations, it is not necessary for the dimer to go through an α–helix intermediate during folding; and (3) this dimer strongly favors anti–parallel and parallel over other orientations.

This article is organized as follows. In Section 2, the activation–relaxation technique and the protein energy model are reviewed. In Section 3, we present the folding trajectories for the dimer of $A\beta_{16–22}$. Concluding remarks follow in Section 4.

2 METHODS

2.1 Activation–Relaxation Technique

ART is a generic method to explore the landscape of continuous energy functions through a series of activated steps. ART events are defined directly in the space of configuration, which allows it to generate moves of any complexity. An ART event consists of four steps:

1. Starting from a minimum, the configuration is first distorted along a direction taken at random in the 3N–dimensional space. The distortion is slowly increased until a direction of deformation becomes unstable, *i.e.*, until the lowest eigenvalue in the Hessian matrix representing the curvature

of the energy landscape becomes negative.

2. The configuration is then pushed along this unstable direction of deformation (the eigenvector associated with the negative eigenvalue) while the energy is minimized in the hyperplane perpendicular to this direction until the total force on all atoms vanishes, indicating the convergence to a transition state (first order saddle point). These two first steps constitute the activation phase.

3. The configuration then is pushed slightly over the saddle point and is relaxed to a new local minimum, using standard minimization technique.

4. The final configuration is accepted/rejected using the Metropolis criterion [20] based on the energy difference between the final and the initial minima at the desired temperature.

$$P_{accept}(\{x\}_{final}) = \min\left(1, e^{-(E_{final}-E_{initial})/k_B T}\right) \quad (1)$$

As discussed in previous work [9,10,16], ART generates well–controlled trajectories. The sampling is ergodic, *i.e.*, any region of the phase space is accessible from any other with a finite number of steps, and events are reversible. Although straightforward, ART requires the adjustment of a few parameters to optimize its efficiency. These are the Metropolis temperature, the initial random displacement, the threshold of the negative eigenvalue indicating that the harmonic well is left, and the size of the Lanczos matrix used to estimate the lowest eigenvalue and its corresponding eigenvector. We emphasize that these parameters do not modify the energy landscape but only the efficiency of the sampling.

2.1.1 ART parameters

Metropolis Temperature. At each event, the final conformation is accepted or rejected according to the Metropolis criterion based on the energy difference between the final and the initial minima. Because these conformations are fully relaxed and do not include any entropic contribution, the Metropolis temperature does not correspond exactly to the standard temperature. An optimal choice of the Metropolis temperature in ART will set a balance to obtain a reasonable acceptation rate while lowering the overall energy of the configuration. If T is very high, the sampling is equivalent to a random walk on the landscape, with little chance to reach low energy states; if T is too low, the system remains trapped in local minima and the native state cannot be located within the available simulation time. The optimal temperature for sampling the energy landscape should therefore have an acceptation rate between about 10 and 70%, or between 300 and 1000 K for most proteins described with the OPEP energy model.

Leaving the harmonic well. We have examined the effect of different initial displacements on the efficiency of finding a saddle point. We find that we can increase significantly the efficiency of finding a saddle point by projecting the random initial displacement in the sub–space associated only with the non–covalent interactions. This is done simply by removing the direction parallel to

the covalent force from the initial random direction as the conformation is pushed away from the initial minimum.

The second important parameter for leaving the harmonic well is the curvature threshold. If the threshold for the eigenvalue is not low enough below zero, any relaxation perpendicular to the corresponding eigenvector send the curvature back into the positive region; if it is too negative, then there is a good chance that the conformation will be pushed into a new basin during the random deformation, generating discontinuous trajectories. In order to stabilize the path leading to the saddle point, it is therefore essential that we start activating only once we are well outside the harmonic well but not further than necessary. The choice of this ideal threshold is system–dependent and must be obtained by trial and error, so as to maximize the success rate of finding a saddle point.

Size of Lanczos matrix. As ART requires only knowledge of the lowest eigenvalue and its corresponding eigenvector, it is possible to use the Lanczos algorithm [21] in its simplest form and not to worry about ghost vectors which are a common problem when using this method to extract a larger set of vectors. Since the activation is iterative, we can also use the previously found eigen–direction as seed to construct the next Lanczos matrix, allowing us to work with matrices between 20 and 30 dimensions in the case of proteins. Including the perpendicular relaxation, one activated step requires therefore between 30 and 40 force evaluations.

2.1.2 Comparison with similar approaches

A similar approach has been following by Wales and collaborators [22,23]. While the basic idea is the same, there are a few of differences between both approaches.

In the first implementations of the eigenvector–following method, the search direction for saddle points was determined by the various eigenvectors of the Hessian matrix at the minimum. This approach can generate a maximum of 6N different trajectories away from the minimum, a number which is insufficient to sample all saddle points around a given minimum [16]. ART avoids this problem by launching searches in random directions.

Another difference lies in the connectivity between the various events. While ART generates a continuously connected trajectory, with a newly generated minimum connected to the previous one by a single first–order saddle point, Wales and his group tend to generate sets of disconnected minimum–saddle–minimum configurations. The initial displacement from a local minimum is generally rather brutal and can easily bring the conformation far away from its initial basin. Once the configuration has converged to a saddle point, the two minima connected to it are identified by minimization. While this approach is as efficient as ART for generating events, it is not always possible to reconnect all these isolated sets of minimum–saddle–minimum configuration in order to recreate a folding trajectory.

2.2 Energy Model

For the energy model, we use a coarse–grained off–lattice model in which each amino acid is represented by its N, H, Cα, C, O and one bead for its side chain (SC). All side chains are defined with respect to the main chain by their bond lengths $b_{C\alpha,SC}$, bond angles $\theta_{C\alpha}{}^{i-1}{}_{,C\alpha}{}^{i}{}_{,SC}{}^{i}$ (where *i* denotes the *i*th residue), and improper dihedral angles, $i\tau_{C\alpha,SC}$, between the 4 atoms $C\alpha^{i+1}$, $C\alpha^{i-1}$, $C\alpha^{i}$ and SC^{i}.

The Optimized Potential for Efficient peptide–structure Prediction (OPEP) includes aqueous solution effects implicitly [24,25]. Its analytic form was obtained by maximizing the energy of the native fold and an ensemble of non–native states for six training peptides with 10–38 residues. The total energy is expressed by [11,24,25]:

$$E = w_L E_L + w_{SC,SC} E_{SC,SC} + w_A E_{C\alpha,C\alpha} + w_{SC,M} E_{SC,M} + w_{M,M} E_{M,M} + w_H E_{HB1} + w_{HH} E_{HB2} \quad (2)$$

The interaction potential OPEP is a function of the weights *w*′s of the following interactions: (*i*) E_L term for the bond lengths and bond angles of the side chains and the backbone as well as the improper torsions of the side chains ($i\tau_{C\alpha,SC}$) and peptide bonds (ω), (*ii*) pairwise contact potential between main chain particles [$E_{M,M}$ and $E_{C\alpha,C\alpha}$], side chain – main chain [$E_{M,SC}$] and side chains [$E_{SC,SC}$], (*iii*) backbone two–body [E_{HB1}] and four–body [E_{HB2}] hydrogen bonding interactions. In OPEP version 1.3, all nonbonded interactions are included (no cut–off distance) and all *w*′s are set to 1.

The quadratic terms for maintaining stereochemistry include

$$E_L = \sum_{bonds} K_b (r - r_{eq})^2 + \sum_{angles} K_\alpha (\alpha - \alpha_{eq})^2 + \sum_{improper \quad torsions} K_\Omega (\Omega - \Omega_{eq})^2 \quad (3)$$

where the corresponding force constants and equilibrium values are given in Ref. [24].

The contact potential between two side chains *i* and *j* (with $j > i+1$) separated by a distance r_{ij} is represented by a 12–6 potential if $\varepsilon_{ij} > 0$ and a 6–potential if $\varepsilon_{ij} < 0$ [24]:

$$E_{SC,SC} = \varepsilon_{ij}\left(\left(\frac{r_{ij}^0}{r_{ij}}\right)^{12} - 2\left(\frac{r_{ij}^0}{r_{ij}}\right)^{6}\right) H(\varepsilon_{ij}) - \varepsilon_{ij}\left(\frac{r_{ij}^0}{r_{ij}}\right)^{6} H(-\varepsilon_{ij}) \quad (4)$$

where the Heavyside function $H(x) = 1$ if $x \geq 0$ and 0 if $x < 0$, $r_{ij}{}^0 = (r_i{}^0 + r_j{}^0)/2$, with the van der Waals radii described in Ref. [24]. The ε_{ij}'s are set to 0.005 kcal/mol if $j = i+1$ and to $k_B T \times (\varepsilon_{ij}{}^{MJ} - (\varepsilon_{ij}{}^{C\alpha C\alpha}\, k_B T))$ otherwise. k_B is Boltzmann's constant, *T* room temperature and the $\varepsilon_{ij}{}^{MJ}$'s are equal to the Miyazawa–Jernigan (MJ)'s values given in Table V of Ref. [26].

The $E_{M,SC}$ energy between the main chain and the side chains, and the $E_{M,M}$ energy between any two backbone *i* and *j* atoms (with $j \geq i+4$) follow Eq. (4) with ε_{ij}'s set to 0.005 kcal/mol and the van der Waals radii of the main chain atoms described in Ref. [24]. We also define an $E_{C\alpha,C\alpha}$ contact term as in Eq. (4) between two Cα's ($j \geq i+4$) with $\varepsilon_{ij}{}^{C\alpha C\alpha}$ set to 0.4 kcal/mol.

The two–body energy of one H–bond between residues *i* and *j* ($j \geq i + 4$) is defined by:

$$E_{HB1} = \varepsilon_{hb} \sum_{ij} \mu(r_{ij}) \nu(\alpha_{ij}) \tag{5}$$

where

$$\mu(r_{ij}) = 5\left(\frac{\sigma}{r_{ij}}\right)^{12} - 6\left(\frac{\sigma}{r_{ij}}\right)^{10} \tag{6}$$

$$\nu(\alpha_{ij}) = \begin{cases} \cos^2 \alpha_{ij}, & \alpha_{ij} > 90° \\ 0, & otherwise \end{cases} \tag{7}$$

and r_{ij} is the O..H distance between the carbonyl oxygen and amide hydrogen, α_{ij} the NHO angle and σ the equilibrium value for the O..H distance.

The co–operative energy between two neighboring hydrogen bonds *ij* and *kl* is defined by:

$$E_{HB2} = \varepsilon_{2hb} \exp(-(r_{ij} - \sigma)^2 / 2) \exp(-(r_{kl} - \sigma)^2 / 2) \tag{8}$$

The parameters for E_{HB1} and E_{HB2} are σ = 1.8 Å, ε_{hb} = 1.0 kcal/mol if $j = i + 4$ (helix), otherwise = 1.5 kcal/mol. The parameter ε_{2hb} is set to –0.5 if the pattern of hydrogen bonds corresponds to α–helices, otherwise –2.0 kcal/mol.

3 RESULTS AND DISCUSSION

We have performed 12 ART simulations (R1 to R12) of a dimer of $A\beta_{16–22}$ at 1000 K for 18000 events. All runs start from a perfect parallel β–sheet (Figure 1B), use different seed numbers and are free of any biases. This initial state is chosen because $A\beta_{10–35}$ and $A\beta_{1–40}$ adopt a parallel organization of strands, and thus our simulations can provide insights into the pathways connecting parallel to anti–parallel β–sheet. Since the starting parallel conformation is only destabilized by 3 kcal/mol relative to the anti–parallel conformation, this also represents an ideal test for studying the efficiency of ART to sample the energy surface of oligomeric systems. In what follows, all event numbers refer to accepted event numbers. The acceptance ratio is ~70% at 1000 K. The pictures are produced using the MOLMOL software [27].

3.1 Energy Surface Characterization

Among the 12 simulations, 3 simulations (R1, R6, R9) locate the same conformation of lowest energy (~ –44.4 kcal/mol, Figure 1A). This global minimum (identical results are obtained starting from random conformations) is characterized by the following H–bond network: 2–17, 4–15, 6–13, and 8–11. In our notation, *i–j* indicates that residues *i* and *j* are hydrogen bonded by (*i*)CO..HN(*j*) and (*i*)HN..CO(*j*), while *i*/*j* refers to one H–bond between (*i*)CO..HN(*j*). The structural characteristics of this state matches exactly the anti–parallel β–sheet organization proposed by

Balbach *et al*. on the basis of solid–state NMR spectra, namely the pattern of H–bonds depicted between layer II (middle) and layer III (bottom) of Figure 10 in Ref. [17]. This state is obtained at events 10853 in run R1, 2594 in run R6 and 6467 in run R9.

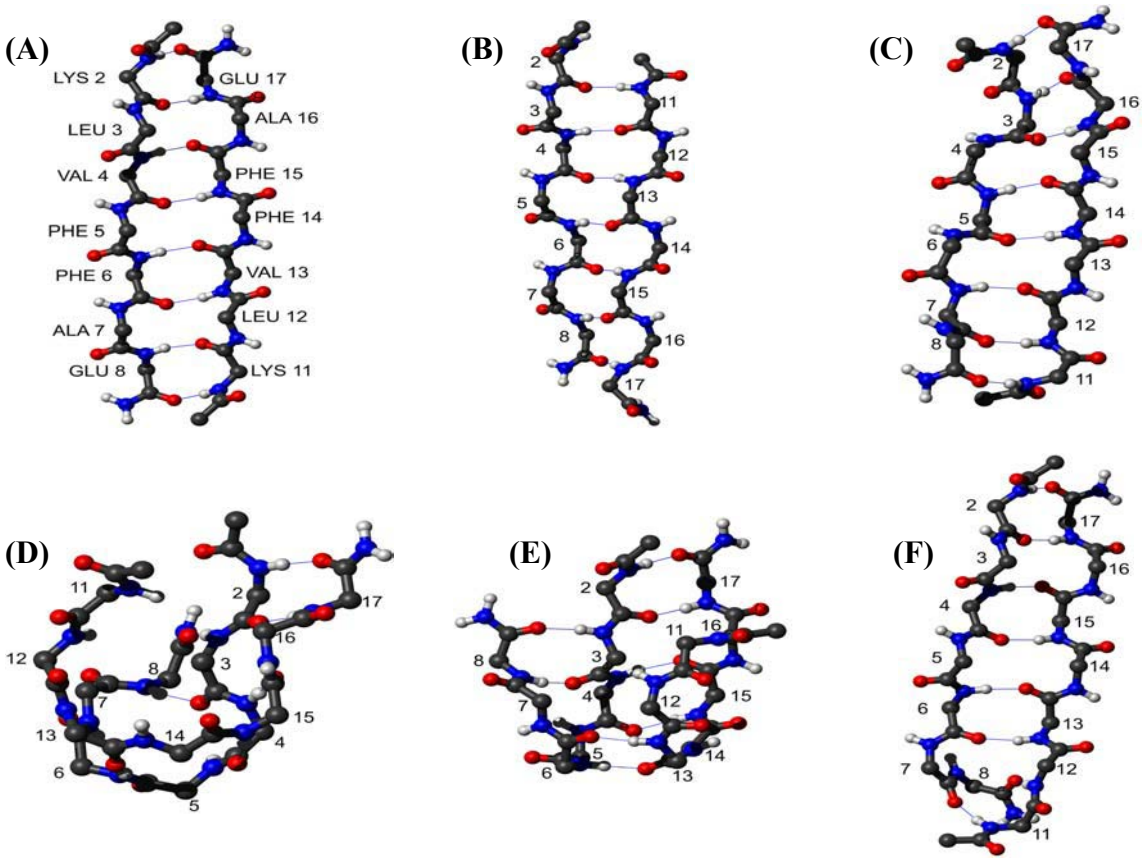

Figure 1. The lowest energy structures from the 12 trajectories. (A) the ground state, runs R1 and R9; (B) the parallel beta–sheet, runs R5, R7, R8, R10, R11 and R12; (C) run R3; (D) and (E) two views for run R4; (F) run R6. For clarity, the side chains are not shown, oxygen atoms are in red, amide hydrogen atoms are in gray and H–bonds are indicated by thin lines.

As expected, not all runs locate the global minimum within the available simulation time. In run R4, the lowest–energy conformation is reached at event 11737 (Figure 1D,E) with energy of –43.04 kcal/mol and a C_α RMSD of 6.20 Å from the global minimum. This structure is partially misfolded because the pattern of H–bonds is not totally formed: 2–17, 4–15 and 6–13 are formed, but not 8–11. In run R3, the lowest–energy conformation, anti–parallel in character, has an energy of –43.3 kcal/mol and deviates by 3.8 Å RMS from the global minimum (Figure 1C). Surprisingly, its network of H–bonds characterized by 3–16, 5–14 and 7–12 follows exactly the second NMR solid–state derived pattern of H–bonds depicted between layers I (top) and II (middle) of Figure 10 in Ref. [17]. Finally, all other simulations (R2, R5, R7, R8, R10, R11 and R12) locate a parallel β–sheet (Figure 1B) of –41.36 kcal/mol, deviating by 9.8 Å RMS from the global minimum and characterized by the network of H–bonds: 2/11, 4/13, 6/15, 8/17, 11/4, 13/6, 15/8. This

conformation, similar to the starting structure, is explored many times during the simulations, *e.g.*, at events 440, 7253 and 11965 in run R5.

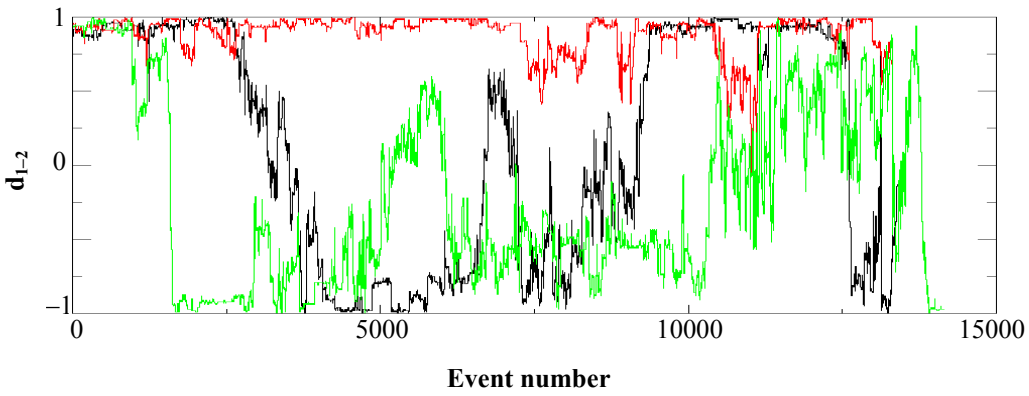

Figure 2. Orientation of peptides as a function of ART event. d_{1-2} is red for run R5, black for run R2 and green for run R6. The ground state is located at event 2594 in run R6.

3.2 Folding Mechanisms

To follow the orientation of the peptides during the simulations, we computed the scalar product (d_{1-2}) of the end–to–end unit vectors of peptides 1 and 2. d_{1-2} = 1 indicates parallel, –1 anti–parallel and 0 perpendicular register. Figure 2 shows the variation of d_{1-2} for the folded R6 trajectory and the unfolded R2 and R5 trajectories and Figure 3A reports d_{1-2} for the folded R1 trajectory.

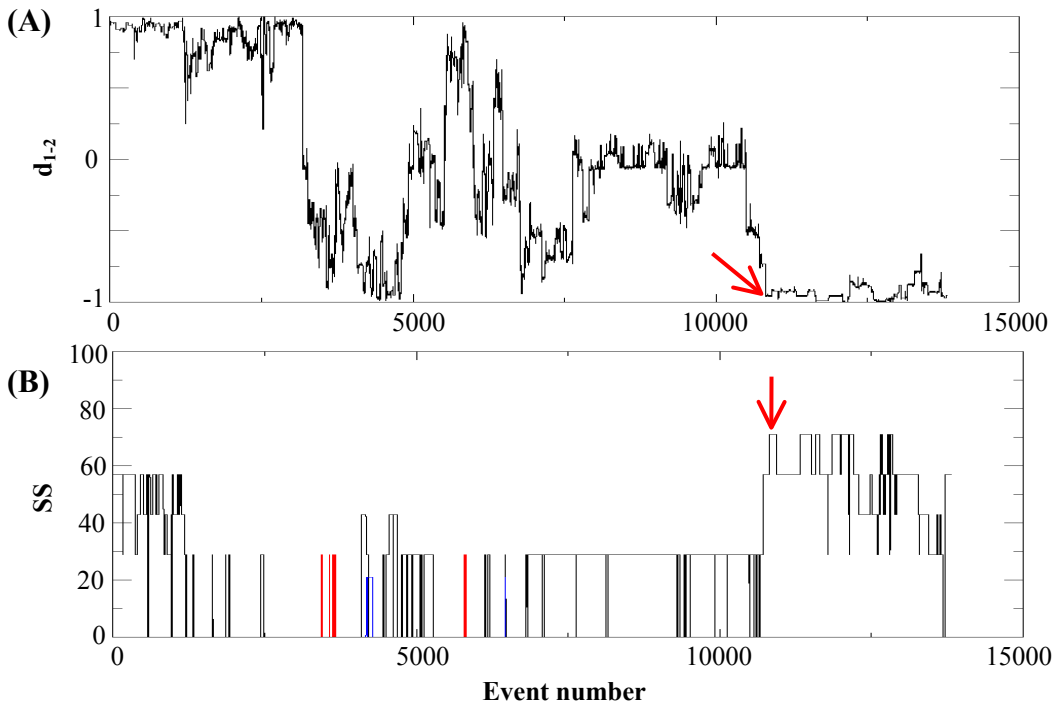

Figure 3. Evolution of orientation of peptides and secondary structure (SS) compositions in run R1. (A) d_{1-2}. (B) percentage of beta–sheet (black) and alpha–helix (red) using DSSP program. Helices 3–10 are in blue. The ground state, sampled at event 10853, is indicated by an arrow.

We find that the transition from parallel to anti–parallel arrangements can be very fast, within less than 1500 events as in run R6. However, this is not a generic property. In run R5, the anti–parallel orientation is never explored; in run R2, the anti–parallel orientation is sampled, but the peptide never converges to the global minimum; a perpendicular organization of the peptides can persist between events 7600 and 10600 as in run R1. From the d_{1-2} plots, we also see that only 3 orientations show stability over a few hundred events: parallel, anti–parallel and perpendicular, with a marked preference for the first two states. Such a kinetic behavior has been noted for the trimer of $A\beta_{16-22}$ [19].

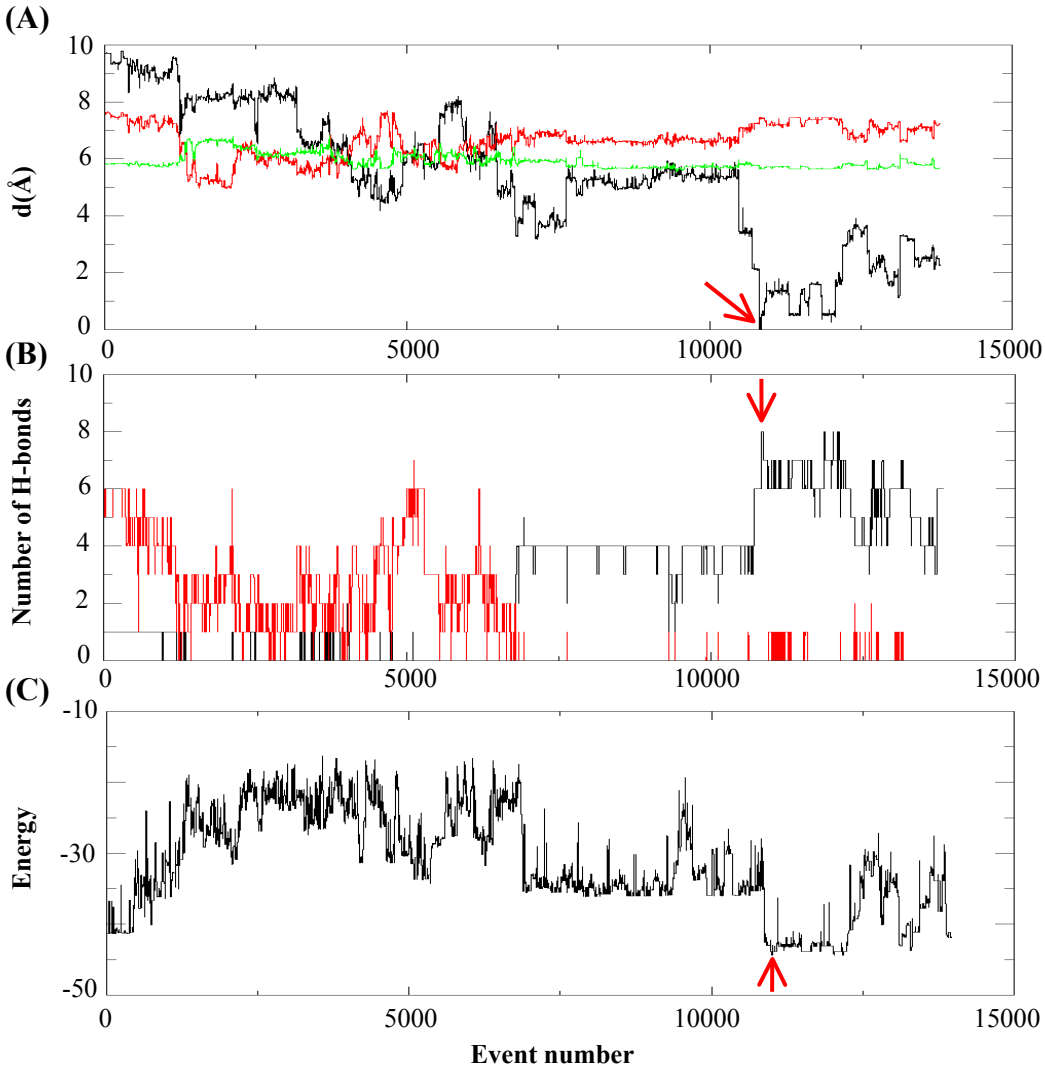

Figure 4. Variation of order parameters in run R1. (A) RMS deviation from the ground state (black), Rg (red) and Rg-core (green); (B) number of native (black) and non-native H-bonds (red); (C) energy in kcal/mol. The ground state, sampled at event 10853, is indicated by an arrow.

The mechanism leading to fibril formation is not fully understood. Walsh *et al*. and Kirkitadze *et al*. studied the secondary structure changes of $A\beta_{1-40}$ and $A\beta_{1-42}$ during fibrillogenesis and observed the formation of an oligomeric intermediate containing 29–32% α–helix [28,29]. Recently, Klimov

and Thirumalai have proposed on the basis of biased MD simulations on a trimer of $A\beta_{16-22}$ that the anti–parallel β–sheet occurs by multiple pathways with the formation of an obligatory α–helical intermediate [19].

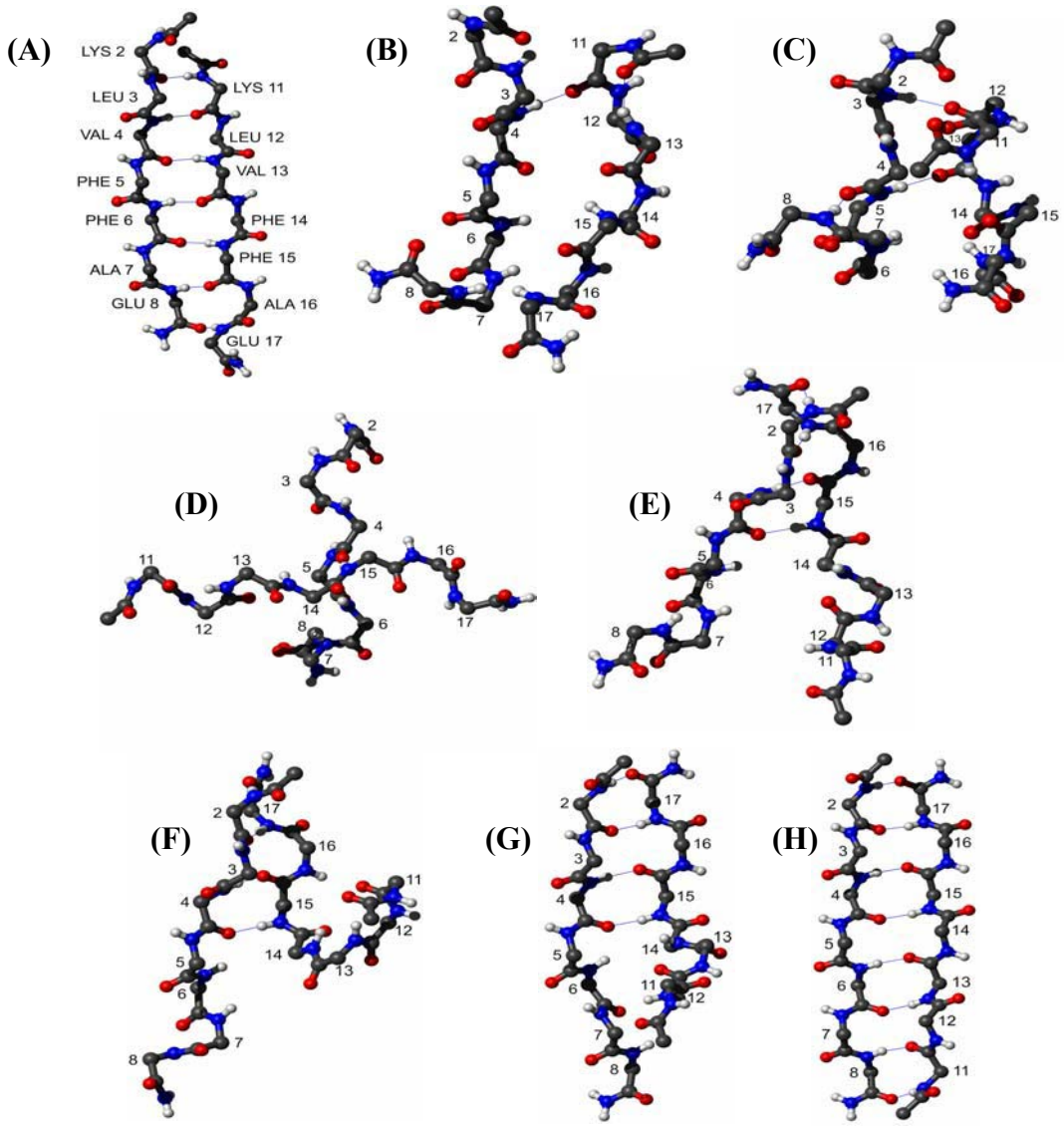

Figure 5. The folding R1 trajectory in atomistic detail. Snapshots at events 0 (A), 1236 (B), 1604 (C), 6705 (D), 7500 (E), 8077 (F), 10600 (G) and 10852 (H).

To investigate whether transitions from helix to strand occur in our folding simulations, we followed the evolution of the secondary structure composition using the standard DSSP program [30]. In run R1, we see that the dimeric intermediate between events 3400 and 3660, anti–parallel in character, has a α–helix spanning residues Leu12–Phe15 (fragment 2, Figure 3B) and that the intermediate between events 5790 and 5820, in parallel register, has a α–helix spanning residues Val4–Ala7 (fragment 1). A 3–10 helix spanning residues Phe5–Ala7 (fragment 1) is also observed simultaneously with a β–sheet between events 4170 and 4270. In contrast, no α–helix intermediate

is formed during the transition in the folded R9 and R6 trajectories (data not shown).

3.3 A Detailed Atomic Trajectory

The trajectory R1 is shown in atomistic detail in Figure 5. In addition to the initial and global minima, we have selected 6 intermediate states obtained using cluster analysis of the folded trajectories R1, R6 and R9 with an RMS deviation cut–off of 3 Å. This means that all intermediates shown are sampled by at least two simulations. The evolution of RMSD, radius of gyration (Rg), radius of gyration of the hydrophobic core (Rg–core), number of native and non–native H–bonds and energy is given in Figure 4. Rg is calculated using all atoms, Rg–core using the side chains of Leu, Val and the two Phe for each peptide.

The starting structure, parallel in character with 1 native (13/6) and 6 non–native H–bonds (2/11, 4/13, 6/15, 8/17, 11/4 and 15/8) is shown in Figure 5A. Its energy is –41.3 kcal/mol, Rg and Rg–core are ~ 8 and 6 Å, respectively (Figure 5A). Note that Rg–core does not vary more than 1 Å along the simulation. Event 1236 is characterized by the loss of all intermolecular H–bonds except 11/4 (Figures 5B, 4B), with both fragments remaining in extended conformations, the C_α RMSD and Rg decreasing to 8 and 5 Å respectively and *E* increasing to –25 kcal/mol. Then, the number of non–native H–bonds varies between 2 and 3 and, at event 1604, both fragments bend (Figure 5C), make transitory H–bonds 13/5 and 11/3 and start to rotate in respect to one another. During the next 5000 events, the fragments continue to rotate, the RMSD varies between 4 and 8 Å and *E* between –20 and –35 kcal/mol. At event 6705, the two fragments are perpendicular with the H–bond 4/15 formed (Figure 5D). This H–bond persists until the global minimum is reached. In this state, the peptide 1 forms a turn involving residues Phe5 to Glu8, while the peptide 2 is fully extended. The C_α RMSD is 4.8 Å, *E* is –22.8 kcal/mol and Rg reaches a plateau at 7 Å. At event 7500, the fragments are anti–parallel with four native H–bonds (2–17, 4–15) formed (Figure 5E). From this core of H–bonds, the peptide finds a way to locate the ground state. We see that at event ~ 10800, *E* and RMSD decrease suddenly and the number of native H–bonds increase to 8 (Figure 4). The global minimum is reached at event 10852 (Figure 5H), forming a perfect two–stranded anti–parallel β–sheet.

4 CONCLUSIONS

In conjunction with the OPEP interaction potential and its unique set of parameters, ART is found to predict ground states of several peptides in monomeric and dimeric states superposable to the experimental structures. In addition to locating the native state, ART generates trajectories that help provide a clear picture of the activated mechanisms responsible for the folding dynamics of peptides.

Using ART and OPEP, we have studied in this paper the energy landscape of the dimer $A\beta_{16–22}$,

generating many trajectories going from parallel to anti–parallel orientations. The barrier height to go from one orientation to another without any bias (defined here by the difference in energy between the highest–energy minimum and the native state on a folding trajectory) is ~ 20 kcal/mol. This gives a folding time much larger than 1 μs at room temperature, even considering the entropic contribution associated with the large number of possible transition states.

During each simulation, we have generated thousands of intermediate states. With this large sample, it is possible to evaluate the occurrence of specific structures during folding. In particular, we find that the dimer favors states that are almost perfectly parallel, anti–parallel and orthogonal, with a marked preference for the first two orientations. This is reflected in the value of the radius of gyration, which varies very little during the simulations. Our simulations also show that the occurrence of a helical intermediate is not an obligatory step for the $A\beta_{16–22}$ fragment: intermediates containing 30% α–helix are observed in run R1, but not in runs R9 and R6. This result contrasts with an earlier MD report [19], indicating that there might be alternative folding pathways for fibril formation. Such a helical intermediate has been identified in the fibril formation of longer segments such as $A\beta_{1–40}$ by circular dichroism analysis [28]. We suspect that a minimum length is needed for the fragments to gain stability by forming helices.

The results reported here on $A\beta_{16–22}$ in dimeric state indicate paradoxically that the folding trajectories can be described with a few parameters but that these trajectories are quite varied in their details. Further simulations on $A\beta_{16–22}$ in hexameric state are in progress to provide us with a more realistic folding picture for fibril formation.

Acknowledgment

SB, GW and NM are supported in part by the *Fonds québécois de la recherche sur la nature et les technologies* and the *Natural Sciences and Engineering Research Council* of Canada. The calculations were done on the computers of the *Réseau québécois de calcul de haute performance* (RQCHP) and of the IGS in Marseille. NM is a Cottrell Scholar of the Research Corporation.

5 REFERENCES

[1] M. Huse and J. Kuriyan, The conformational plasticity of protein kinases, *Cell* **2002**, *109*, 275–282.
[2] Y. Duan and P. Kollman, Pathways to a protein folding intermediate observed in a microsecond simulation in aqueous solution, *Science* **1998**, *282*, 740–744.
[3] U. H. E. Hansmann and Y. Okamoto, Prediction of peptide conformation by multicanonica algorithm: new approach to the multiple minima problem, *J. Comput. Chem.* **1993**, *14*, 1333–1338.
[4] P. Ferrara, J. Apostokalis and A. Caflisch, Thermodynamics and kinetics of folding of two model peptides investigated by molecular dynamics simulations, *J. Phys. Chem. B* **2000**, *104*, 5000–5010.
[5] U. H. E. Hansmann, Parallel tempering algorithm for conformational studies of biological molecules, *Chem. Physics. Lett.* **1997**, *281*, 140–150.
[6] Y. Sugita, Y. Okamoto, Replica–exchange molecular dynamics method for protein folding, *Chem. Physics. Lett.* **1999**, *314*, 141–151.
[7] B. Zagrovic, E. J. Sorin and V. Pande, beta–hairpin folding simulations in atomistic detail using an implicit solvent model, *J. Mol. Biol.* **2001**, *313*, 151–169.
[8] B. Zagrovic, C.D. Snow, M. R. Shirts and V. Pande, Simulation of folding of an alpha–helical protein in atomistic

detail using worldwide distributed computing, *J. Mol. Biol.* **2002**, *323*, 927–937.
[9] N. Mousseau, P. Derreumaux, G. T. Barkema, and R. Malek, Sampling activated mechanisms in proteins with the activation relaxation technique, J. Mol. Graph. Model. **2001**, *19*, 78–86.
[10] G. Wei, N. Mousseau, and P. Derreumaux, Exploring the energy landscape of proteins: a characterization of the activation relaxation technique, *J. Chem. Phys.* **2002**, *117*, 11379–11387.
[11] G. Wei, P. Derreumaux and N. Mousseau, Complex folding pathways in a simple beta–hairpin, *J. Chem. Phys.* **2003**, submitted.
[12] A. R. Fersht, On the simulation of protein folding by short time scale molecular dynamics and distributed computing. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 14122–14125.
[13] G. T. Barkema and N. Mousseau, Event–based relaxation of continuous disordered systems, *Phys. Rev. Lett.* **1996**, *77*, 4358–4361.
[14] N. Mousseau and G. T. Barkema, Traveling through potential energy landscapes of disordered materials: The activation–relaxation technique, *Phys. Rev. E* **1998**, *57*, 2419–2424.
[15] G. T Barkema and N. Mousseau, Identification of relaxation and diffusion mechanisms in amorphous silicon, *Phys. Rev. Lett.* **1998**, *81*, 1865–1868.
[16] R. Malek and N. Mousseau, Dynamics of Lennard–Jones clusters: A characterization of the activation–relaxation technique, *Phys. Rev. E* **2000**, *62*, 7723–7728.
[17] J. J. Balbach, Y. Ishii, O. N. Antzutkin, R. D. Leapman, N. W. Rizzo, F. Dyda, J. Reed and R. Tycko, Amyloid fibril formation by Aβ 16–22, a seven–residue fragment of the Alzheimer's β–amyloid peptide, and structural characterization by solid state NMR, *Biochemistry* **2000**, *39*, 13748–13759.
[18] B. Ma and R. Nussinov, Stabilities and conformations of Alzheimer's beta –amyloid peptide oligomers (Abeta 16–22, Abeta 16–35, and Abeta 10–35): sequence effects, *Proc. Natl. Sci. USA* **2002**, *99*, 14126–14131.
[19] D. K. Klimov and D. Thirumalai, Dissecting the assembly of Abeta(16–22) amyloid peptides into anti–parallel beta sheets, *Structure* **2003**, *11*, 295–307.
[20] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. J. Teller, Equation of state calculations by fast computing machines, *J. Chem. Phys.* **1953**, *21*, 1087–1092.
[21] Lanczós, *Applied Analysis*. Dover, New York, **1988**.
[22] J. P. K. Doye and D. J. Wales, Surveying a potential energy surface by eigenvector–following – Applications to global optimization and the structural transformations of clusters, *Z Phys. D*: *At. Mol. Clusters* **1997**, *40*, 7723–7728.
[23] L. J. Munro and D. J. Wales, Defect migration in crystalline silicon, *Phys. Rev. B* **1999**, *59*, 3969–3980.
[24] P. Derreumaux, From polypeptide sequences to structures using Monte Carlo simulations and an optimized potential, *J. Chem. Phys.* **1999**, *111*, 2301–2310.
[25] P. Derreumaux, Generating ensemble averages for small proteins from extended conformations by Monte Carlo simulations, *Phys. Rev. Lett.* **2000**, *85*, 206–209.
[26] S. Miyazawa and R.L. Jernigan, Estimation of effective interresidue contact energies from protein crystal structures: quasi–chemical approximation, *Macromolecules* **1985**, *18*, 534–52.
[27] R. Koradi, M. Billeter, and K. Wuthrich, MOLMOL: a program for display and analysis of macromolecular structures, *J. Mol. Graphics* **1996**, *14*, 51.
[28] D. M. Walsh, A. Lomakin, G. B. Benedek, M. M. Condron and D. B. Teplow, Amyloid beta–protein fibrillogenesis. Detection of a protofibrillar intermediate, *J. Biol. Chem.* **1997**, *272*, 22364–22372.
[29] M. D. Kirkitadze, M. M. Condron and D. B. Teplow, Identification and characterization of key kinetic intermediates in amyloid beta–protein fibrillogenesis, *J. Mol. Biol.* **2001**, *312*, 1103–1119.
[30] W. Kabsch, and C. Sander, Dictionary of protein secondary structure: pattern recognition of hydrogen–bonded and geometrical features, *Biopolymers* **1983**, *22*, 2577–2637.