Molecular Dynamics

DOI: 10.1002/anie.201003816

Multiple Routes to Characterize the Folding of a Small DNA Hairpin**

*Guillem Portella and Modesto Orozco**

Understanding the folding and unfolding of nucleic acids is a requisite for explaining the biological action of these molecules. Despite recent efforts,[1–4] the folding of nucleic acids is still poorly understood, and many phenomenological, kinetic, and molecular mechanisms have been suggested, including early models assuming a zipper molecular mechanism.[5] Later, the two-state model became prevalent,[6–8] whilst recent ultrafast experiments suggested a more complex multiple-state scenario.[9–14] In the last few years, a variety of complex and flexible kinetic models have been developed that can fit almost any experimental behavior.[7,13–15] Unfortunately, such models are often difficult to interpret in structural–atomistic terms and do not always clarify important aspects, such as the existence of single or multiple folding pathways or the nature of the unfolded form and intermediates. In other words, we are still far from reaching a general accepted molecular mechanism for such processes.

The lack of direct microscopic experimental information has fueled the use of molecular dynamics (MD) as a tool to uncover structural and dynamical details at atomic level. Most of the published simulations have used indirect approaches to overcome the sampling problem, such as the use of multiple short (ns-range) trajectories[16,17] or replica-exchange MD (REMD),[18–20] losing therefore the temporal information of the folding process. Herein we study in atomistic detail the folding landscape of a short DNA hairpin on a real timescale. MD simulations (for all details, see the Supporting Information) were performed for the short oligo d(GCGAAGC), which forms in solution a hairpin with two canonical d(G·C) and a three-loop d(GAA) with one non-canonical d(G·A),[21] and that is expected to fold in the sub-μs scale.[12,14]

[*] G. Portella, M. Orozco
Joint IRB-BSC Research Program in Computational Biology
Institute for Research in Biomedicine (IRB)
Josep Samitier 1-5, Barcelona 08028 (Spain)
and
Barcelona Supercomputing Centre (BSC)
Jordi Girona 29, Barcelona 08034 (Spain)
E-mail: modesto@mmb.pcb.ub.es
M. Orozco
Departament de Bioquímica i Biología Molecular
Facultat de Biología, Universitat de Barcelona
Avgda Diagonal 645, Barcelona 08028 (Spain)
M. Orozco
National Institute of Bioinformatics, Parc Científic de Barcelona
Josep Samitier 1–5, Barcelona 08028 (Spain)

[**] We are indebted to the computer resources provided by the Barcelona Supercomputing Center. This work was supported by BIO2009-10964, Consolider E-science and COMBIOMED ISCIII projects, and the Fundación Marcelino Botin. The authors thank Profs. F. J. Luque and V. Muñoz for critical comments.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201003816.

A key prerequisite in MD-based studies of folding/unfolding is the ability of the force field to recognize the folded form in native conditions. To check this requirement, we performed three sets of REMD simulations (190, 185, and 120 ns), starting from extended conformations, using 22 temperatures ranging from 300 K to 370 K (aggregate time 10 μs; see Supporting Information, Figures S1–S3, for additional results). Trajectories at 300 K indeed sample a large range of states corresponding to unfolded, partially folded, and native conformations, the later being the most populated state at this temperature (Supporting Information, Figure S2), confirming the goodness of the force field. To gain time-dependent information of the folding process, we performed twenty independent MD simulations for at least 1 μs at 300 K, starting in all cases from a fully extended structure. Proper folding to the native hairpin occurred in 12 of the 20 simulations, whilst the remaining 8 trajectories were trapped in stable-compact but non-native conformations and do not reach native minima in one microsecond (Figure 1). Extension of five of the non-productive trajectories to 4 μs led to folded structures in two cases, suggesting that all the trajectories will converge in the experimental structure, but that this process can in some cases be very slow.

After merging the populations collected during all unbiased trajectories with the REMD data at 300 K (more than 30 μs of simulation time; see Figure 2), we used the number of stacking bases and hydrogen bonds as collective variables that characterize the state of the hairpin, identifying three regions of low free energy: 1) the native form (with 6–8 hydrogen bonds and around four stacked bases), 2) a partially folded state characterized by two hydrogen bonds and two stacked bases, and 3) the fully unfolded form with no native hydrogen bonds and a variable number of stacked bases. The landscape has a distinctive point corresponding to structures characterized by good stacking and correct hydrogen bonding in the loop and the first base of the stem (Figure 2). Reaching this position from a quasi-extended conformation implies a dramatic loss of intramolecular DNA entropy. Crossing this point is crucial for native folding, but also trajectories arriving there can also fail to adopt the native conformation (see below). Clustering of the structures generated at 300 K reveals 10 major states with population similar or greater than 1%, but a very significant portion of the unfolded ensemble (around 50%) cannot be annotated to any large cluster. Therefore, our simulations strongly argue against a two-state model consisting of a fully flexible unfolded state and one native structure, and to similar reductionist approaches describing folding/unfolding by a very limited number of states.

Analysis of the 14 successful trajectories shows that folding happens over a broad range of times, from 0.1 μs to 1.95 μs and in some cases more than 4 μs (we named this

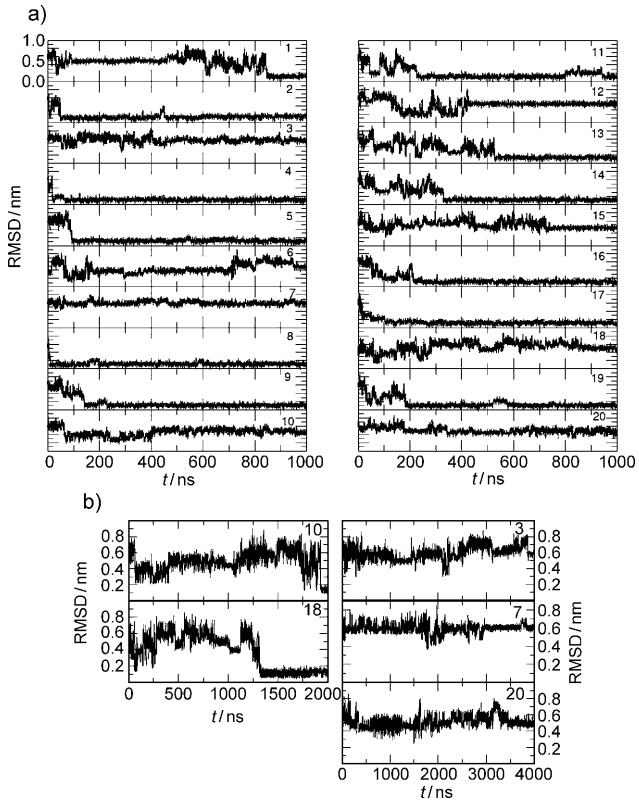

Figure 1. a) Root-mean-square deviations (RMSDs) with respect to the native state for our 20 initial trajectories. The folded state is found at RMSDs with values lower than 0.2 nm. The vertical scale for trajectory 1 applies to all other trajectories. b) Extended molecular dynamics simulations for trajectory 10 and 18 (2 μs), 3, 7, and 20 (4 μs).

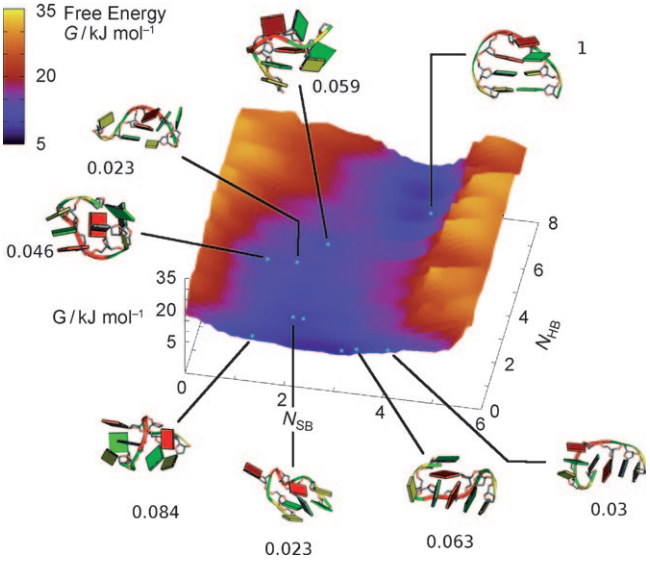

Figure 2. Free-energy maps as function of the number N_{HB} of inter-base hydrogen bonds and number N_{SB} of stacked base pairs extracted from cumulated 31 μs free molecular dynamics (MD) and replica-exchange MD trajectories (495 ns). Representative clusters obtained from RMSD clustering of the combined trajectories are located on the top of the free energy surface, with the associated population referred to the folded conformation.

apparent folding time). The real folding time, defined as that elapsed during a transition from an unfolded conformation that starts forming the loop to a stable native structure, is in the range from 5 to 265 ns, with a typical folding time of about 160 ns (see Supporting Information, Methods and Figures S4–S6 for a discussion). The distribution of apparent folding times does not follow an exponential decay, thus indicating the presence of more than one folding mechanism (see Supporting Information, Figure S5 and Methods). One of these mechanisms leads to the fast folding trajectories and can be explained following a downhill-like or direct folding[22] mechanism with three well-defined milestones (see example in Figure 3): 1) nucleation and closure of the loop with the formation of the G·A pair, 2) fast formation of the first G·C pairing in the stem, and finally 3) formation of the second G·C pairing (see also the Supporting Information, Figures S4,S7). The very fast formation of the loop-closing base pair after loop nucleation agrees very well with the experimental observation by Bevilacqua and co-workers that the loop and loop-closing base pair are formed in a highly cooperative manner.[23] The formation of the second G·C base pair is the slower step in the fast routes, which is largely due the tendency of the last guanosine to visit the *syn* conformation, something that is more likely in the 5′G form owing to the possibility of intramolecular 5′OH–N_3 hydrogen bonds. The impact of a guanosine *syn*↔*anti* equilibrium in folding was already anticipated by Zacharias and co-workers[20] using REMD simulation in related systems, supporting the generality of present findings.

The three trajectories (Tr1, Tr10, and Tr18) with very long apparent folding times ($\tau_{app} > 800$ ns) share some interesting similarities: all of them adopt the correct loop structure in around 100 ns and in less than 200 ns the stem is partially formed (see Figure 3). However, the flip of 5′G to the *syn* region precludes the formation of the last canonical d(G·C) pair in the stem, and as a consequence, non-canonical G(*syn*)–G(*anti*) hydrogen bonds are formed. This leads to the extrusion of one of both cytosines, evolving then into a very stable compact structure with strong extra-helical stacking and almost no intra-strand hydrogen bonds (see example in Figure 3c). Such a conformation is maintained (from 0.5 to 1 μs) until it reopens into a variety of extended conformations that finally close the loop and fall into the native basin, following similar pathways than the fast-folding trajectories. Clearly, de-trapping the hairpin misfolded conformations is the rate-limiting step in these trajectories.

We failed to obtain native folding in 6 of the 20 trajectories, and for 3 of them (Tr3, Tr7, and Tr20) even after 4 μs of MD simulation. In all cases, except Tr7, correct loop nucleation took place, but the terminal guanosine was locked in the *syn* conformation. In two cases (Tr14 and Tr18), we found compact pseudo hairpins with an incorrect loop definition d(GCG**AAG**C) instead of the correct form, d-(GC**GAA**GC), whilst in the others the oligo remains trapped very soon in a compact structure. Caution is necessary in the interpretation of the fluorescence data extracted from experiments of hairpin folding,[12] as many of the non-native structures detected in our simulations have a large degree of stacking and compactness, providing similar macroscopic

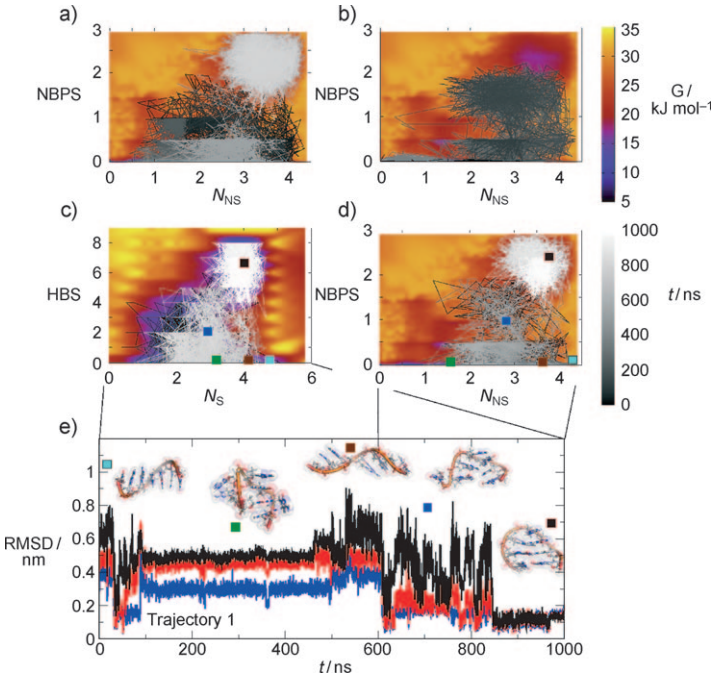

Figure 3. Time-evolution (white–black scale) of the hairpin conformation on the free-energy map (color scale, obtained from all our free MD simulations) describing the number of native base pairs (NBPs) and base stacking for three trajectories. N_{NS} = native stacking number, N_S = stacking number. a) folding trajectory 18, b) non-folding trajectory 12. c, d) Clear cases of de-trapping and folding for trajectory 1. Data in (c) refers to total contacts. e) RMSD from the native structure (blue: loop, red: first base at stem, black: whole hairpin) and selected structures sampled along the trajectory.

descriptors similar to those of the native state. Based on the behavior of Tr1, Tr10, and Tr18, we can expect that compact abortive conformation will evolve to fully open structures, which can then converge into the native form in the multi-microsecond timescale; that is, two to three orders of magnitude slower than the direct folding detected in some of the trajectories.

The wrong conformation around the glycosidic bond (χ) of 5′guanosine is a common finding in all of the abortive trajectories, but *syn*↔*anti* transitions at 5′G are present in the beginning of all trajectories (Supporting Information, Figures S8,S9). We found that fast χ-conformational changes are irrelevant for folding, unless the molecule is collapsed when guanosine is in the *syn* conformation: intramolecular contacts hinder the *syn*→*anti* conformational shift in collapsed form, and *anti* conformation is required for Watson–Crick G·C pairing, supporting previous claims on the importance of guanosine *syn*↔*anti* equilibrium in hairpin folding.[20] In fact, correct and fast folding is obtained in 8 of 10 trajectories that failed to fold in 1 μs when they were relaunched using a flat-well restraint to keep χ in the *anti* region, or when the terminal 5′-CH_2OH was substituted by a methyl group (see the Supporting Information for details). This complex behavior enlightened by our simulations contributes to the debate on the temperature dependence on the folding velocity.[8,24] An increase in the temperature should have two opposite effects: on one hand, it will favor de-trapping, increasing then the folding velocity in slow routes (that is, Arrhenius behavior), and on the other it will favor *anti*↔*syn* transitions in the extended form, thus increasing the possibility to reach abortive routes and decreasing then the relative number of fast routes (anti-Arrhenius behavior). The global dependence of folding with temperature will be then the result of a subtle equilibrium between opposite effects.

Our atomistic simulations support the kinetic partitioning mechanism proposed by Thirumalai (KPM),[24] where folding is understood as a competition between different folding pathways, some that are very quickly attracted by the native basin and others which are initially attracted by abortive basins. Our results suggest that attraction to abortive basins can appear not only from the extended conformation, but also from structures displaying a well-nucleated loop or even the first base pair of the stem. Minuscule atomistic details, mostly unpredictable in the absence of atomistic simulations, decide the evolution of the trajectory towards native or abortive basins. Our simulations suggest that experimental folding times are in reality the average of at least two distinct distributions of apparent folding times, corresponding to ultra-fast and slow de-trapping and folding events. Our results indicate that hairpin folding is a process of an extreme complexity even for the simplest models. Thus, very intricate folding landscapes are expected for larger systems, raising doubts on the suitability of reductionist kinetic mechanisms models, and encouraging the use of flexible physical techniques such as atomistic MD, which can open new scenarios of research for new-generation biophysical techniques.

Received: June 22, 2010
Published online: September 15, 2010

Keywords: computational biophysics · DNA · hairpin folding · molecular dynamics

[1] K. B. Hall, *Curr. Opin. Chem. Biol.* **2008**, *12*, 612.
[2] N. A. Siegfried, S. L. Metzger, P. C. Bevilacqua, *Biochemistry* **2007**, *46*, 172.
[3] S. J. Chen, *Annu. Rev. Biophys.* **2008**, *37*, 197.
[4] J. A. Cruz, E. Westhof, *Cell* **2009**, *136*, 604.
[5] G. Bonnet, O. Krichevsky, A. Libchaber, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 8602.
[6] E. Zazopoulos, E. Lalli, D. M. Stocco, P. Sassone-Corsi, *Nature* **1997**, *390*, 311.
[7] W. Zhang, S. J. Chen, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 1931.
[8] M. I. Wallace, L. Ying, S. Balasubramanian, D. Klenerman, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 5584.
[9] H. Ma, D. J. Proctor, E. Kierzek, R. Kierzek, P. C. Bevilacqua, M. Gruebele, *J. Am. Chem. Soc.* **2006**, *128*, 1523.
[10] K. Sarkar, K. Meister, A. Sethi, M. Gruebele, *Biophys. J.* **2009**, *97*, 1418.
[11] J. Jung, A. Van Orden, *J. Am. Chem. Soc.* **2006**, *128*, 1240.
[12] H. Ma, C. Wan, A. Wu, A. H. Zewail, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 712.
[13] W. Zhang, S. J. Chen, *Biophys. J.* **2006**, *90*, 765.

[14] A. Ansari, S. V. Kuznetsov, Y. Shen, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 7771.
[15] X. Tang, S. Thomas, L. Tapia, D. P. Giedroc, N. M. Amato, *J. Mol. Biol.* **2008**, *381*, 1055.
[16] E. J. Sorin, Y. M. Rhee, B. J. Nakatani, V. S. Pande, *Biophys. J.* **2003**, *85*, 790.
[17] E. J. Sorin, Y. M. Rhee, V. S. Pande, *Biophys. J.* **2005**, *88*, 2516.
[18] A. E. Garcia, D. Paschek, *J. Am. Chem. Soc.* **2008**, *130*, 815.
[19] G. R. Bowman, X. Huang, Y. Yao, J. Sun, G. Carlsson, L. J. Guibas, V. S. Pande, *J. Am. Chem. Soc.* **2008**, *130*, 9676.
[20] S. Kannan, M. Zacharias, *Biophys. J.* **2007**, *93*, 3218.
[21] P. Padrta, R. Stefl, L. Kralik, L. Zidek, V. Sklenar, *J. Biomol. NMR* **2002**, *24*, 1.
[22] P. Li, F. Y. Oliva, A. N. Naganathan, V. Muñoz, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 103.
[23] N. L. Goddard, G. Bonnet, O. Krichevsky, A. Libchaber, *Phys. Rev. Lett.* **2000**, *85*, 2400.
[24] D. Thirumalai, S. a. Woodson, *Acc. Chem. Res.* **1996**, *29*, 433.

