


 Cite this: *RSC Adv.*, 2021, **11**, 36242

# Kinetic coherence underlies the dynamics of disordered proteins†

 Alexander Tenenbaum 

The dynamics of two proteins of similar size, the globular lysozyme and the intrinsically disordered Huntingtin interacting protein, has been simulated in three states resembling a globule, a pre-molten globule, and a molten globule. A coherence time  $\tau$  has been defined, measuring the delay in the display of a stochastic behaviour after a perturbation of the system. This time has been computed for two sets of collective variables: the projection of the phase point onto the positions and momenta subspaces ( $\tau_r$  and  $\tau_p$ ), and the principal components (PCs) of positions  $\mathbf{q}$  and momenta  $\boldsymbol{\pi}$  produced by a covariance analysis in these subspaces ( $\tau_q$  and  $\tau_\pi$ ). In all states  $\tau_p \approx 3.5\tau_r$ , and  $\tau_\pi \approx 3.5\tau_q$ . The coherence times of individual PCs,  $\tau_q^{(i)}$  and  $\tau_\pi^{(i)}$ , have also been computed, and  $\tau_\pi^{(i)} > \tau_q^{(i)}$  in all states. The prevalence of  $\tau_p$  over  $\tau_r$ , or of  $\tau_\pi$  over  $\tau_q$ , drives the dynamics of the protein over a time range of  $\approx 1$ – $2$  ps; moreover, a hidden synchronism appears to raise the momenta subspace's coherence above that of its individual PCs. In the transition of lysozyme to the molten globule the  $\tau_q^{(i)}$  decrease but, unexpectedly, the  $\tau_\pi^{(i)}$  increase; after this transition  $\tau_p \approx 5\tau_r$  and  $\tau_\pi \approx 5\tau_q$ . A gain of *kinetic coherence* accompanies thus the loss of *structural coherence* caused by the denaturation of the protein in the transition from globule to molten globule. The increase of the  $\tau_\pi^{(i)}$  does not take place in the analogous transition of the Huntingtin protein. These results are compared with those of a similar analysis performed on three pseudo-proteins designed by scrambling the primary sequence of the Huntingtin interacting protein, and on two oligopeptides. The hidden synchronism appears to be a generic property of these polypeptides. The  $\tau_\pi^{(i)}$  spectrum is similar in denaturated and in intrinsically disordered biomolecules; but the gain of kinetic coherence as a result of denaturation seems to be a specific property of the biologically functional lysozyme.

 Received 10th September 2021  
 Accepted 29th October 2021

DOI: 10.1039/d1ra06823g

[rsc.li/rsc-advances](http://rsc.li/rsc-advances)

## 1 Introduction

In recent years the traditional structure–function or “lock-and-key” paradigm has been shown to be insufficient to account for the complexity of binding scenarios among biomolecules. It is now recognized that structural dynamism – supporting molecular recognition that doesn't depend only upon structural and electrostatic complementarity between the binding partners – is more common in functional proteins than previously assumed.<sup>1,2</sup>

The loosening of the stereochemical constraints required for a binding process to take place has shifted the attention to collective modes supporting conformational transitions functional to said process.<sup>3</sup> Both experiment<sup>4,5</sup> and atomistic simulation<sup>6</sup> have contributed to the study of low frequency collective motions – usually in the THz range – that are interwoven with the binding process. These collective modes could also provide

a clue to understand a long-range interaction among biomolecules, possibly a key component of some binding processes.<sup>7</sup>

An analysis of an  $\alpha$ -helix and of a 13-residue protein fragment performed *via* molecular dynamics simulation pointed at the chaotic nature of their motion.<sup>8,9</sup> This was ascribed to the nonlinear interactions in the system's force field, and to the stochastic forces generated by the solvent; it has also been stated that a noisy environment is a necessary complement of collective motions driving a protein across its conformational ensemble.<sup>10</sup> On the other hand, it has also been argued that the dynamics of a protein is piecewise integrable, and becomes chaotic only during conformational transitions supported by fast collective motions.<sup>11</sup> Moreover, a protein can undergo a transition from an incoherent to a coherent dynamics notwithstanding its interaction with a thermal bath.<sup>12</sup> Because of this mixed picture, the possible emergence of an ordered pattern in the dynamics of a biomolecule – in particular in its biologically relevant collective motions – deserves a thorough investigation.

A biomolecule in a solvent is ceaselessly exposed to perturbations; if an ordered pattern underlies its dynamics, this pattern must be resilient against the steady perturbation to which it is exposed. The term coherence usually refers to

Physics Department, Sapienza University, Piazzale Aldo Moro 5, 00185 Roma, Italy.  
 E-mail: [tenen314@gmail.com](mailto:tenen314@gmail.com)

† Electronic supplementary information (ESI) available. See DOI: 10.1039/d1ra06823g



a dynamical regime in which a set of vibrations exist whose relative phases do not vary in a given time period. In the present work the term coherence is instead used to represent the resilience of a system's dynamical trajectory in the presence of a perturbation; this resilience can be measured. A perturbation moves the point in the phase space representing the system's state to a nearby point. If a system is chaotic, the distance between the trajectories beginning at these nearby points is characterized by a fast exponential growth in time; the growth rate can be estimated by a parameter, the (maximum) Lyapunov exponent:  $\Delta(t) = \Delta(0)e^{\lambda_1 t}$ , where  $\Delta(t)$  is the distance in the phase space at time  $t$ , and  $\lambda_1$  is the maximum Lyapunov exponent. A resilient system, as opposed to a chaotic one, needs a long time to develop a chaotic behaviour; it is characterized by a slow divergence of near trajectories, that is, by a low value of  $\lambda_1$ .

Let  $S_r$  and  $S_p$  be the subspaces of positions and momenta in the phase space. Taking the cue from the results found in a butane molecule, where different degrees of freedom (DoFs), both in  $S_r$  and in  $S_p$ , are characterized by different degrees of coherence,<sup>13</sup> a similar analysis has been applied here to a set of biomolecules. All internal DOFs (torsions, bonds and angles) were included in the atomistic models of the biomolecules and in the computation of their dynamics. But the focus of the analysis has been put on the collective variables derived in a principal components (PCs) analysis of the dynamics, as it has been shown in several cases that the PCs endowed with the largest fluctuations are related to conformational shifts relevant to the function of a protein. The simulation was thus aimed at detecting a possible coherence hierarchy among collective DoFs in  $S_r$  and  $S_p$ .

## 2 Systems

The set of biomolecules object of this study is the following. (i) The hen egg-white lysozyme protein (Lys). (ii) The Huntingtin interacting protein (Hip), an intrinsically disordered protein (IDP) entailing the same number of residues (129) as the lysozyme.<sup>14</sup> (iii) Three disordered pseudo-proteins obtained from Hip by randomly scrambling the order of residues in the primary sequence (P1h, P2h, P3h). (iv) Two short oligopeptides obtained by taking the first 9 (O1p) or the first 4 (O2p) residues of the N-terminus of Hip. The lysozyme protein is a widely studied antimicrobial enzyme; the Huntingtin interacting protein has been chosen because, lacking stable secondary structures, a comparison of its dynamics with that of Lys would highlight possible differences due to such structures. The three pseudo-proteins entail the same hydrophilic residues as Hip, and are therefore similarly disordered; on the other hand, lacking presumably any kind of biochemical function, a possible difference between their dynamics and that of Lys and Hip could be related to this lack. The primary sequence of Hip as well as the primary sequence of the three pseudo-proteins are given in ESI.† The two oligopeptides, both intrinsically disordered, were analyzed and compared to the other disordered systems in order to highlight a possible role of the proteins' size in their dynamical behaviour.

The 3D structure of lysozyme was taken from the PDB repository.<sup>15</sup> The 3D structures of Hip and of the three pseudo-proteins were obtained through the University of Michigan's I-TASSER server.<sup>16–18</sup> The server constructs molecular structures based on the sole primary sequence, applying an iterative template-based fragment assembly to the sequence. In this procedure it creates various conjectured secondary structures, producing the corresponding pdb file; in each case the molecule with the least secondary structures was chosen to start the simulation (see ESI, Fig. 1–4).†

All systems were simulated in the physiological condition (state A at  $T = 300$  K), and above the denaturation temperature of lysozyme (state B at  $T = 370$  K); the state B was reached by raising the temperature of state A from 300 K to 370 K. Because of the thermal stability of these biomolecules, the structures in state B were not fully denaturated. In order to study the dynamical properties of the denaturated lysozyme and compare them with those of the IDPs, it was necessary to rid as much as possible the proteins of the secondary structures; therefore, Lys, Hip, P1h, P2h, and P3h were subjected to a thermal denaturation at  $T = 1000$  K, followed by an equilibration at  $T = 370$  K to produce fully disordered conformations with no or little secondary structures: state C. The oligopeptides O1p and O2p, lacking secondary structures, were simulated only in states A and B, and were not subjected to the step at high temperature. The final structures produced by this procedure are shown in Fig. 1 and in ESI (Fig. 5–8).† Hereafter, the various systems in

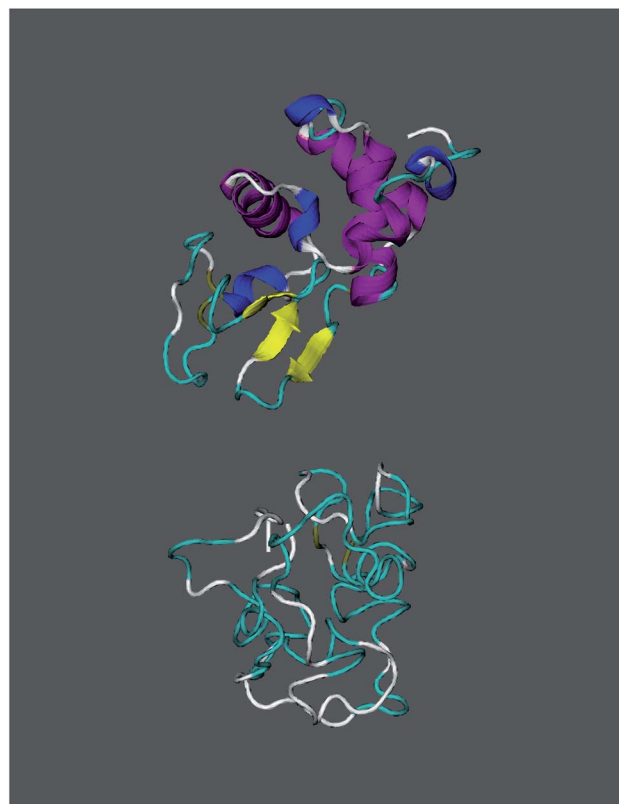


Fig. 1 Lysozyme. Upper structure: LyA, globule. Lower structure: LyC, mollen globule.



the different states will be dubbed accordingly: lysozyme LyA, LyB, LyC; Huntingtin HiA, HiB, HiC; pseudo-proteins P1A, P1B, P1C, P2A, P2B, P2C, P3A, P3B, P3C; oligopeptides O1A, O1B, O2A, O2B. The rationale of studying the dynamics of the same system in states B and C, that are at the same temperature, is to distinguish possible effects of the secondary structures – present in B but almost absent in C – from possible effects of the temperature.

### 3 Analytical method

Let  $\mathbf{r}$  be the position vector of all the atoms of the soluted molecule. The PCs in  $S_r$  are defined by the linear transformation of  $\mathbf{r}$  into  $\mathbf{q} = O_r^T(\mathbf{r} - \langle \mathbf{r} \rangle)$ , where  $O_r$  is the matrix the columns of which are the orthonormal eigenvectors of the covariance matrix  $C_r = \langle (\mathbf{r} - \langle \mathbf{r} \rangle) (\mathbf{r} - \langle \mathbf{r} \rangle)^T \rangle$ ; in this section  $\langle \cdot \rangle$  is a time average.  $O_r^T C_r O_r$  is a diagonal matrix, the elements of which are the eigenvalues corresponding to those eigenvectors. The time evolution of vector  $\mathbf{q}$  displays the essential dynamics of the molecule: it has been shown that the components of  $\mathbf{q}$  along few eigenvectors, corresponding to the largest eigenvalues of  $C_r$ , account for most of the total spatial fluctuation of a protein.<sup>19</sup>

Let  $\mathbf{p}$  be the momentum vector of all the atoms of the soluted molecule. The PCs in  $S_p$  are similarly defined by the linear transformation of  $\mathbf{p}$  into  $\boldsymbol{\pi} = O_p^T(\mathbf{p} - \langle \mathbf{p} \rangle)$ , where  $O_p$  is the matrix the columns of which are the orthonormal eigenvectors of the covariance matrix  $C_p = \langle (\mathbf{p} - \langle \mathbf{p} \rangle) (\mathbf{p} - \langle \mathbf{p} \rangle)^T \rangle$ .  $O_p^T C_p O_p$  is a diagonal matrix, the elements of which are the eigenvalues corresponding to those eigenvectors. The two sets of variables  $\mathbf{q}$ ,  $\boldsymbol{\pi}$  span two subspaces,  $S_q$  and  $S_\pi$ ; these variables have been analyzed in search of a possible hierarchy in their coherence, measured by the coherence time of each DoF.

Let  $N$  be the number of atoms in the protein, and  $l = 1, 3N$  the index of the DOFs in each of the subspaces  $S_r$ ,  $S_p$ ,  $S_q$ , and  $S_\pi$ . The coherence times  $\tau_r$  of  $S_r$  and  $\tau_p$  of  $S_p$  are measured by  $1/\lambda$ , where  $\lambda = (1/t) \ln[|\Delta(t)|/|\Delta(0)|]$  is a Lyapunov-like quantity characterizing the rate of divergence of trajectories beginning at nearby points;  $\Delta(0)$  and  $\Delta(t)$  are the Cartesian distances between the unperturbed and perturbed trajectories at the beginning and at the end of a given time interval  $t$ , in the subspaces  $S_r$  and  $S_p$ , respectively. This coherence time is physically equivalent to the coherence time defined in ref. 13 from the conceptual framework of Lyapunov's theory; it entails the same information about the rate of divergence of trajectories starting from very close points of the phase space, but is simpler to compute. The coherence times  $\tau_q$  of  $S_q$  and  $\tau_\pi$  of  $S_\pi$  were measured likewise: in each of the subspaces  $S_r$  and  $S_p$ , the initial and final distances  $\Delta(0)$  and  $\Delta(t)$  were projected on each  $l$ -th eigenvector of the corresponding covariance matrix,  $q^{(l)}$  and  $\pi^{(l)}$ , yielding  $\Delta(0)_q^{(l)}$  and  $\Delta(t)_q^{(l)}$  in  $S_q$ , and  $\Delta(0)_\pi^{(l)}$  and  $\Delta(t)_\pi^{(l)}$  in  $S_\pi$ . The Cartesian sum of these variations yielded  $\Delta(0)$  and  $\Delta(t)$  in those two subspaces, that were used to compute  $\tau_q$  and  $\tau_\pi$  analogously to  $\tau_r$  and  $\tau_p$ . The coherence times of  $q^{(l)}$  and  $\pi^{(l)}$ ,  $\tau_q^{(l)}$  and  $\tau_\pi^{(l)}$  respectively, were computed in a similar way: the projections on the eigenvectors were used to compute  $\tau_q^{(l)} = 1/\lambda_q^{(l)}$ , where  $\lambda_q^{(l)} = (1/t) \ln[|\Delta(t)_q^{(l)}|/|\Delta(0)_q^{(l)}|]$ , and  $\tau_\pi^{(l)} = 1/\lambda_\pi^{(l)}$ , where  $\lambda_\pi^{(l)} = (1/t) \ln[|\Delta(t)_\pi^{(l)}|/|\Delta(0)_\pi^{(l)}|]$ .

### 4 Computational method

The simulation has been performed *via* the GROMACS simulation package,<sup>20</sup> and the VMD program for the visualization of the molecule.<sup>21</sup> For this simulation the all-atom OPLS-AA force field has been used.<sup>22</sup> All simulations started from the corresponding pdb file, either provided by the PDB repository (Lys<sup>15</sup>), or by the I-TASSER server (Hip, P1h, P2h, P3h); this file was converted into the GROMACS format using the `pdb2gmx` tool. All systems have been simulated in explicit water solvent, using the SPC/E water model;<sup>23</sup> the molecule has been solvated in a suitable triclinic box filled with water molecules, and extending to a distance of 1 nm from the solute. Using the GROMACS `genion` module, 8 negative Cl counterions were added to neutralize the overall charge on Lys, while for Hip, P1h, P2h, and P3h the charges added to neutralize the system were 9 positive Na counterions. The systems were first relaxed in the solvent by minimizing the total energy; a 1 ns equilibration in a NVT ensemble at  $T = 300$  K was then performed, using V-rescale temperature coupling;<sup>24</sup> it was followed by a 1 ns equilibration in a NPT ensemble at  $T = 300$  K and  $p = 1$  bar, using the Parrinello–Rahman barostat.<sup>25</sup> The final configuration of this equilibration was the starting one for the actual simulation run of states A. Starting from the same relaxed systems in the solvent, the 1 ns NVT and NPT equilibrations were performed at  $T = 370$  K to produce the configuration used to start the simulation of states B. To produce (almost) fully denatured structures, overcoming a possible high thermostability – well known of lysozyme – in a manageable computing time, the same relaxed systems were used to start a 1 ns equilibration in a NVT ensemble at  $T = 1000$  K. The denatured protein was then extracted from the solvent, put and relaxed in new solvent, and equilibrated at 370 K with the usual two 1 ns runs in the NVT and NPT ensembles, taking the systems to state C. More details on the computer simulation can be found in ESI.†

Preliminary runs in the globular state LyA were performed to determine the time interval best suited to measure the coherence of the system. After the equilibration the state of the system was perturbed by randomly scrambling the velocities of all 961 explicit H atoms of the protein; in this way the total momentum and the total kinetic energy were not modified, but the projection of the system's representative point in  $S_p$  was shifted. This perturbation is small. The initial distance in the momenta subspace is zero, as all atoms involved in the scrambling of velocities have the same mass. After one time step, the initial distance between trajectories in the positions subspace turns out to be of the order of 0.045 nm. This variation of the total position is distributed over the 961 H atoms of the protein, and is thus very small, as these atoms are distributed in a box with an edge of about 7 nm.

Different perturbations at different points of an unperturbed trajectory were applied, and the ensuing divergence between the unperturbed and the perturbed trajectories was computed. A typical result is shown in Fig. 2 and 3; the three curves of the lower bundle represent three different perturbations of a given point of the unperturbed trajectory of LyA, and measure the



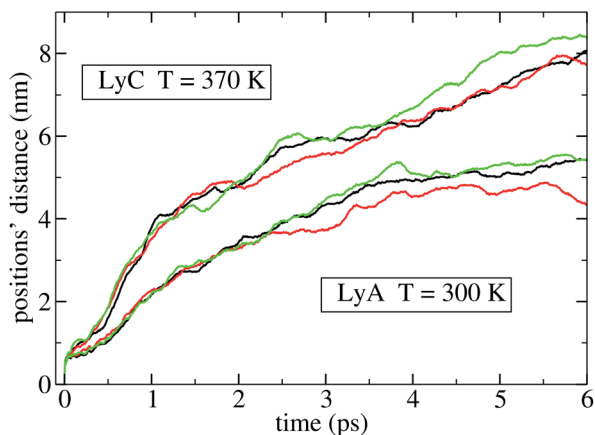


Fig. 2 Divergence of three perturbed trajectories from an unperturbed one in  $S_r$ .

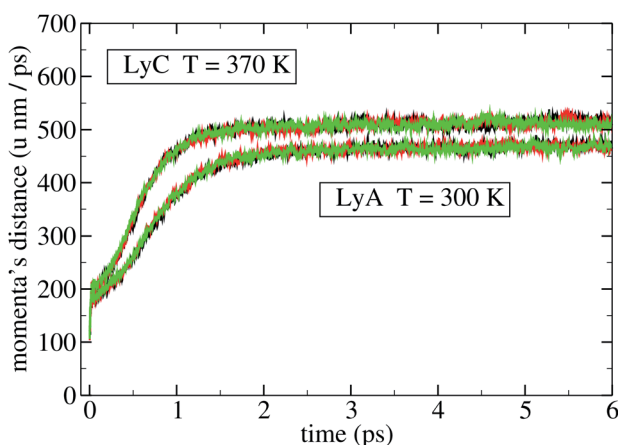


Fig. 3 Divergence of three perturbed trajectories from an unperturbed one in  $S_p$ .

distance in  $S_r$  and in  $S_p$ , respectively. The upper bundle in those figures represent three different perturbations applied to LyC, with a similar result. All curves exhibit a clear decrease in the slope after  $t = 2$  ps, indicating the onset of a saturation in the divergence of the trajectories. In a molecule this distance is limited by the intramolecular bonds and by the velocities' distribution at the given temperature; nevertheless, before reaching that limit, the rate of divergence depends on the molecule and on its state, and may be different for different DoFs. This simulation has been performed also on the other systems, and all results look similar; therefore, the coherence times have been measured on the first 2 ps stretch after the perturbation, where the fastest divergence takes place. Extending beyond 2 ps the time interval over which the expansion is measured would yield converging values of the Lyapunov-like quantity  $\lambda$ , tending to zero for all DoFs. In a similar simulation of the perturbed dynamics of a 13-residue protein fragment, 2 ps was found to be the time needed to reach saturation of the distance between perturbed and unperturbed evolution.<sup>9</sup>

For each system a 10 ns long unperturbed trajectory was produced in a NVT ensemble, with a time step of 2.0 fs, in each of three states conventionally dubbed globule (A), pre-molten globule (B), and molten globule (C); 10 ns are sufficient to assess a stabilized pattern of PCs.<sup>26–28</sup> During each 10 ns trajectory the fluctuation of the total energy was of the order of 1%, without a drift of the mean value. At the beginning, and at every following point of the trajectory corresponding to an integer number of ns, the configuration of the protein was perturbed in five independent ways, giving rise to five perturbed trajectories; these were computed in double precision with a time step of 0.2 fs, and were 2 ps long. Also the unperturbed trajectory was extended with the same precision and the same length from the same point, to allow a precise computation of the distance between this trajectory and the perturbed ones. In this way 55 sets of distances between the unperturbed trajectory and the perturbed ones were obtained; the coherence times in  $S_r$  and  $S_p$  have been computed on each of these sets, and their average over all 55 sets. The dynamics of the 2 ps stretches was computed in a NVE ensemble; this ensemble was chosen in order to avoid a possible alteration of the dynamics of the system by an external thermal coupling<sup>29</sup> or an external pressure coupling. The NVE ensemble simulation of the whole system entails the thermal coupling between the two subsystems protein and solvent, a natural way of representing an actual experimental condition.

The 10 ns unperturbed trajectory was used to compute two covariance matrices, in  $S_r$  and in  $S_p$ . This was suggested by the analysis of a butane molecule's dynamics that had revealed a different dynamical behaviour of the DoFs belonging to these two subspaces.<sup>13</sup> The initial and final positions' and momenta's Cartesian distances in each of the 55 time stretches were projected onto the eigenvectors of the respective covariance matrix; from these projected distances the coherence time  $\tau_q$  of subspace  $S_q$  and the coherence time  $\tau_\pi$  of subspace  $S_\pi$  were computed. The projected distances allowed also the computation of a coherence time for each eigenvector. Ranking the eigenvectors according to the decreasing value of the associated eigenvalue gives the sequence of PCs. The coherence times of the two subspaces  $S_r$  and  $S_p$ , the coherence times of the two subspaces  $S_q$  and  $S_\pi$ , and the coherence times of each PC, all averaged over the 55 stretches, were thus the relevant quantities computed herein.

## 5 Results

Some physical and chemical characteristics of the various systems in the different states are reported in Table 1. All data are time averages over the 10 ns trajectories; the errors are the standard deviation of the average. The number of hydrogen bonds has been measured with the GROMACS hbond module.<sup>30</sup>

The first simulation at 300 K was done for LyA; the structure depicted in Fig. 1 shows 4  $\alpha$ -helices, 4 short  $3_{10}$ -helices, and 3  $\beta$ -strands, plus turns and coil segments. During the 10 ns simulation the average gyration radius in LyA was 1.41 nm, in good agreement with the experimental value of 1.43 nm.<sup>31</sup> What happens when the temperature of the system is raised to the



**Table 1** 1st column: molecule. 2nd column: temperature (K). 3rd column: gyration radius (nm). 4th column: solvent-accessible surface area (nm<sup>2</sup>). 5th column: volume of the solute molecule (nm<sup>3</sup>). 6th column: average number of intramolecular H-bonds

Mol	<i>T</i>	<i>R<sub>g</sub></i>	SASA	Volume	H-b
LyA	300	1.414 ± 0.012	70.45 ± 1.33	26.21 ± 0.31	91
LyB	370	1.395 ± 0.014	70.01 ± 1.88	26.46 ± 0.34	84
LyC	370	1.456 ± 0.018	82.40 ± 2.84	28.09 ± 0.42	72
HiA	300	3.503 ± 0.209	126.51 ± 2.29	31.93 ± 0.60	76
HiB	370	2.373 ± 0.404	110.70 ± 7.03	30.74 ± 0.73	85
HiC	370	2.032 ± 0.083	101.47 ± 2.86	30.24 ± 0.47	94
P1A	300	1.764 ± 0.034	102.98 ± 2.12	30.33 ± 0.38	78
P1B	370	1.674 ± 0.033	96.19 ± 3.97	29.74 ± 0.56	82
P1C	370	2.017 ± 0.117	106.23 ± 4.70	30.60 ± 0.56	76
P2A	300	1.649 ± 0.048	98.00 ± 4.37	29.78 ± 0.53	83
P2B	370	1.636 ± 0.027	89.99 ± 2.44	29.28 ± 0.39	88
P2C	370	1.674 ± 0.075	98.53 ± 5.83	30.21 ± 0.66	74
P3A	300	1.616 ± 0.016	97.77 ± 2.12	29.86 ± 0.39	83
P3B	370	1.584 ± 0.031	93.47 ± 4.21	29.66 ± 0.51	89
P3C	370	1.854 ± 0.134	107.29 ± 5.44	30.06 ± 0.74	74
O1A	300	0.716 ± 0.038	15.12 ± 0.66	2.98 ± 0.08	4
O1B	370	0.594 ± 0.030	13.01 ± 0.84	2.79 ± 0.10	6
O2A	300	0.547 ± 0.028	9.90 ± 0.51	1.77 ± 0.05	0
O2B	370	0.554 ± 0.024	9.78 ± 0.51	1.75 ± 0.06	1

**Table 2** Coherence times (ps). 1st column: molecule. 2nd column: subspace *S<sub>r</sub>*. 3rd column: subspace *S<sub>p</sub>*. 4th column: subspace *S<sub>q</sub>*. 5th column: subspace *S<sub>π</sub>*. 6th and 7th column: average over the first 200 PCs (first 130 PCs for the oligopeptides)

Mol	$\tau_r$	$\tau_p$	$\tau_q$	$\tau_\pi$	$\langle \tau_q^{(l)} \rangle$	$\langle \tau_\pi^{(l)} \rangle$
LyA	0.46 ± 0.01	1.52 ± 0.02	0.46	1.52	0.41	0.50
LyB	0.43 ± 0.04	1.36 ± 0.02	0.42	1.24	0.41	0.64
LyC	0.32 ± 0.04	1.49 ± 0.02	0.32	1.50	0.31	0.66
HiA	0.43 ± 0.01	1.38 ± 0.02	0.43	1.23	0.38	0.66
HiB	0.41 ± 0.06	1.37 ± 0.01	0.40	1.22	0.37	0.65
HiC	0.39 ± 0.06	1.36 ± 0.01	0.39	1.21	0.36	0.66
P1A	0.38 ± 0.01	1.38 ± 0.02	0.38	1.24	0.36	0.65
P1B	0.37 ± 0.01	1.37 ± 0.01	0.39	1.23	0.35	0.65
P1C	0.41 ± 0.01	1.37 ± 0.01	0.41	1.23	0.36	0.67
P2A	0.38 ± 0.00	1.38 ± 0.02	0.39	1.24	0.35	0.65
P2B	0.41 ± 0.01	1.37 ± 0.02	0.41	1.22	0.37	0.64
P2C	0.42 ± 0.01	1.37 ± 0.02	0.42	1.23	0.37	0.65
P3A	0.37 ± 0.06	1.38 ± 0.02	0.36	1.23	0.33	0.65
P3B	0.36 ± 0.01	1.37 ± 0.02	0.36	1.24	0.34	0.65
P3C	0.31 ± 0.01	1.37 ± 0.02	0.31	1.23	0.31	0.66
O1A	0.32 ± 0.05	1.37 ± 0.09	0.33	1.44	0.36	0.54
O1B	0.33 ± 0.05	1.37 ± 0.09	0.34	1.43	0.36	0.53
O2A	0.32 ± 0.04	1.43 ± 0.09	0.33	1.52	0.38	0.51
O2B	0.34 ± 0.05	1.42 ± 0.09	0.35	1.45	0.44	0.51

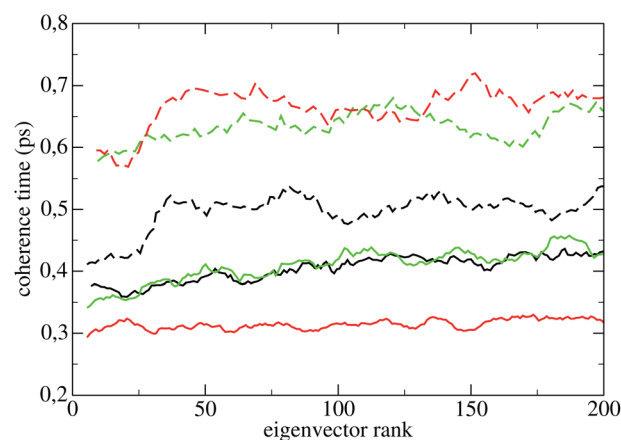
point that the structure of the protein is affected? A second simulation was done in a pre-molten globular state of lysozyme at *T* = 370 K (LyB). As compared with the globular structure LyA, LyB has only 2  $\alpha$ -helices instead of 4, 2 short  $3_{10}$ -helices instead of 4, but has kept the 3  $\beta$ -strands; the solvent-accessible surface area (SASA) and the average gyration radius have not changed within the error range. A third simulation has been done at *T* = 370 K for the molten globule lysozyme (LyC). Unlike the globular structure LyA at 300 K or the pre-molten globular structure LyB at 370 K, LyC has no helices or  $\beta$ -strands left, as shown in Fig. 1; this is characteristic of a denaturated state. The SASA has increased to 82.40 nm<sup>2</sup>; on the other hand, the gyration radius oscillates around 1.46 nm, significantly less than the value of 1.80 nm measured for the fully denaturated and unfolded protein.<sup>32</sup> Therefore, even if the structure of LyC appears to be more open than LyA and LyB, this state is better labeled as molten globule.

In the following runs the molecules Hip, P1h, P2h, and P3h were simulated in the states A, B, and C. The data for Hip show that taking the system from state A at 300 K to state C at 370 K the molecule undergoes a thermal compaction, visible in the values of *R<sub>g</sub>* (−42%), solvent accessible surface area (−20%), and – to a lesser extent – volume (−5%). This phenomenon, known to exist in denaturated proteins and IDPs,<sup>33</sup> does not happen in the pseudo-proteins P1h, P2h, and P3h; to the contrary, in P1h and P3h the SASA increases significantly going from state A to state C.

The coherence times of all systems in all states are reported in Table 2; the error assigned to each result in subspaces *S<sub>r</sub>* and *S<sub>p</sub>* is the standard deviation over the 55 perturbed trajectories. The smallness of the standard deviation of these averages show that the latter are significant, *i.e.*, that the coherence time of a subspace depends weakly on the phase space point of the

constant energy surface where it is measured. The standard deviations in subspaces *S<sub>q</sub>* and *S<sub>π</sub>* are of the same order of magnitude, and are not shown in Table 2. The coherence times of the first 200 PCs of Lys are reported in Fig. 4 (LyA: black lines; LyB: green lines; LyC: red lines), and their average  $\langle \cdot \rangle$  in Table 2. The standard deviation of each individual PC coherence time over the 55 measurements may be quite large, usually between 20% and 40%; for the sake of clarity the points in Fig. 4 do not carry error bars. Here and in the following  $\langle \cdot \rangle$  will define the average coherence time over the first 200 PCs (130 PCs for the oligopeptides).

The curves in Fig. 4 show that: (a) in LyA and LyB, approximately the first 30  $\tau_\pi^{(l)}$  are lower than those of the rest of



**Fig. 4** Coherence times of the eigenvectors in LyA (black lines), LyB (green lines), LyC (red lines). Positions: continuous lines; momenta: dashed lines. Every point is the average in a sliding window of 10 contiguous points.



eigenvectors oscillating with smaller amplitude; (b) the  $\tau_{\pi}^{(l)}$  increase significantly taking the system from LyA to LyB and to LyC. Appearance (a) highlights an anisotropic diffusion in  $S_p$  in those states. Comparing with the homologous curves for Hip in Fig. 5, one notices that appearance (a) is now common to the three states HiA, HiB, and HiC, while appearance (b) has completely disappeared. As for the three pseudo-proteins, their curves of the PCs' coherence times are indistinguishable from those of Hip shown in Fig. 5. The curves for P1h are shown in Fig. 9 of ESI.†

## 6 Discussion

The systems LyA and Hip are biologically functional; LyB, LyC, P1h, P2h, and P3h presumably do not have a biological function. What can be learned by comparing the coherence times in the two subspaces of the systems inside each group and between the two groups?

A first evidence derived from Table 2 is that in all states of all systems  $\tau_p$  is larger than  $\tau_r$  by a factor 3 to 4, that is,  $S_p$  is systematically more coherent than  $S_r$ . This is strongly divergent from the result of a similar computation done several years ago on a gas of Lennard-Jones particles with density going up to 15% of the liquid: there the coherence times computed for the positions and momenta subspaces were practically equal.<sup>34</sup> And in all states of all systems  $\tau_q$  is smaller than  $\tau_{\pi}$ , also by a factor 3 to 4; that is,  $S_q$  is systematically less coherent than  $S_{\pi}$ . Moreover, the  $\tau_{\pi}^{(l)}$  are uniformly higher than the equally ranking  $\tau_q^{(l)}$ , as shown in Fig. 4 for Lys and Fig. 5 for Hip.

A second observation is that  $\langle \tau_{\pi}^{(l)} \rangle$ , the average coherence time of the momenta in  $S_{\pi}$ , has practically the same value of 0.66 ps for all polypeptides in state C, both of proteins (Lys and Hip) and of pseudo-proteins (P1h, P2h, P3h). In that state Hip, being an IDP, is possibly still biologically functional, while the other polypeptides are not. This property seems thus to be related to the conformation of the denaturated or disordered molecules, not to their biological function; this coherence time

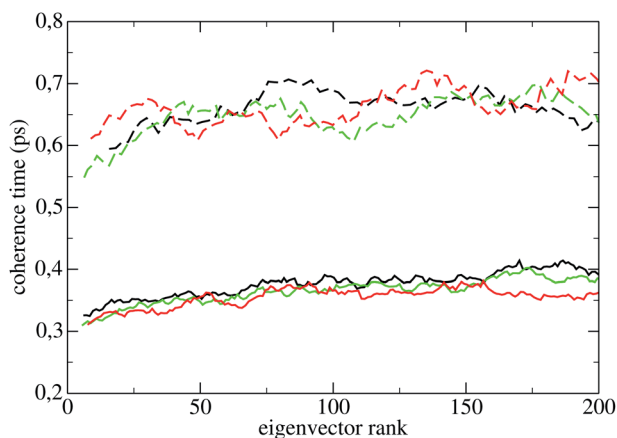


Fig. 5 Coherence times of the eigenvectors in HiA (black lines), HiB (green lines), HiC (red lines). Positions: continuous lines; momenta: dashed lines. Every point is the average in a sliding window of 10 contiguous points.

may also increase with the size of the molecular structure, as the two oligopeptides display a  $\langle \tau_{\pi}^{(l)} \rangle$  that is clearly smaller than that of the polypeptides.

A third observation is that the average coherence time of the first 200 position eigenvectors,  $\langle \tau_q^{(l)} \rangle$ , is lower (down to  $-12\%$  in HiA, P1C, and P2C) than  $\tau_q$  in all polypeptides. This pattern is similarly shared by the biological and non-biological systems just mentioned; therefore, it doesn't bear a biological relevance. As for the average coherence time of the momenta,  $\langle \tau_{\pi}^{(l)} \rangle$  is uniformly lower than  $\tau_{\pi}$ , but here the difference is much higher:  $-67\%$  in LyA,  $-48\%$  in LyB,  $-56\%$  in LyC,  $-46\%$  in HiA,  $-47\%$  in HiB,  $-45\%$  in HiC; P1h, P2h, and P3h show values similar to Hip. The differences between these two quantities vary slightly when the temperature is raised from 300 K to 370 K, with one exception, to be discussed below. Thus in both subspaces – but markedly more in  $S_{\pi}$  – the overall dynamics is more coherent than that of the single eigenvectors, as if the coherence of a single eigenvector were enhanced by a coordinated dynamics of the other eigenvectors. In other words, it appears that if an eigenvector is characterized by a given stochasticity (that is, his evolution in time is chaotic to a certain degree), the other eigenvectors – each similarly incoherent on its own – evolve in a synchronized pattern, thus increasing the coherence of the subspace. This guess of a *hidden dynamical coherence* is supported by the comparison of Fig. 2 and 3, where the divergence of the three perturbed trajectories is clearly distinguishable in  $S_r$  but not in  $S_p$ , as if in the latter the three different perturbations were constrained to evolve on a similar path. This pattern is visible both in LyA (lower bundle of curves) and in LyC (upper bundle): the dramatic change in the structure of lysozyme going from LyA to LyC does not cancel this intriguing different behaviour of the two subspaces. Moreover, the same pattern of hidden dynamical coherence can be found in the other simulated systems: Hip, P1h, P2h, P3h, Op1, Op2.

The fourth and possibly most relevant result is clearly visible in Fig. 4: as noticed before, taking Lys from LyA to LyC the  $\tau_q^{(l)}$  decrease significantly (about  $-25\%$ ), but the  $\tau_{\pi}^{(l)}$  unexpectedly increase (about  $+30\%$ ), becoming more than twice the former. The latter increase mostly takes place between LyA and LyB, that is, when the temperature is raised beyond the denaturation point and a significant amount of the secondary structures is lost. An increase of temperature entails an extension of the phase space volume accessible both to positions and momenta, lowering the coherence of these DoFs; also, breaking-up H-bonds and secondary structures removes barriers in the phase space, possibly making these DoFs even more chaotic, as shown by the decrease of the spectrum of the  $\tau_q^{(l)}$ . By contrast, the increase of the  $\tau_{\pi}^{(l)}$  is surprising; it appears that the loss of *structural coherence* due to the disappearance of secondary structures is accompanied by a gain of *kinetic coherence*. Such a behaviour has not been found in any of the other similar structures studied in the present work (Hip, P1h, P2h, P3h), a hint that it is related to proteins losing their biological function, as the transition from LyA to LyC entails the disruption of all secondary structures and the number of intramolecular H-bonds drops by  $-21\%$ . By contrast when Hip is taken from HiA to HiC the  $\tau_{\pi}^{(l)}$  do not change appreciably and the number of



H-bonds even increases, which is consistent with the thermostability of IDPs.

The gain of *kinetic coherence* during the denaturation process could be a generic feature in the unfolding of a globular protein undergoing denaturation, and could possibly play a role also in the folding process. Since Levinthal proposed his paradox,<sup>35</sup> it has been recognized that the folding of a protein cannot be achieved by a random conformational search in the huge dimensional phase space of the molecule; rather, the path goes from the generic unfolded state to less generic intermediate unfolded states, towards the folded native configuration. In Levinthal's words: "a pathway of folding means that there exist a well-defined sequence of events which follow one another so as to carry the protein from the unfolded random coil to a uniquely folded metastable state".<sup>36</sup> As unfolding and refolding processes take place in a very similar time span,<sup>37</sup> it is tempting to posit that also the unfolding process takes place by way of a similar sequence of intermediate configurations; if this were the case, the increasing coherence in the momenta subspace during the denaturation process could help the increasingly spatially disordered structure of the protein find its way from one intermediate configuration to the following. In the reverse process, kinetic coherence would become less relevant as the system progressively folds towards the native state configuration. Thus, in both cases, kinetic coherence could be a component of the protein's ability to follow "a pathway" in a directed process, countering perturbing actions by the solvent or other molecules not positively involved in the process.

How do the appearances described so far relate to the size of the molecules? An insight into this issue can be found in the results obtained by the simulation of the two oligopeptides O1p and O2p, also reported in Table 2. The coherence times  $\tau_r$  of the subspace  $S_r$  and  $\tau_q$  of the subspace  $S_q$  are smaller than those of the polypeptides (the only exception being  $\tau_r = 0.31$  ps in P3C). Therefore, one could guess a trend of lower coherence of smaller molecules in the two positions subspaces. At variance with the previous systems, the average coherence time of the first 130 position eigenvectors,  $\langle\tau_q^{(l)}\rangle$ , is higher (+9% in O1A and +15% in O2A) than  $\tau_q$  at  $T = 300$  K; when the temperature is raised from 300 K to 370 K the difference between the two quantities decreases in O1B, but increases in O2B. All things considered, the dependence of this difference on the temperature doesn't seem to be relevant either in the oligopeptides or in the larger polypeptides, with the exception of Lys, as discussed in the previous paragraph.

The coherence times  $\tau_p$  of the subspace  $S_p$  are practically equal within their standard deviation in all disordered structures Hip, P1h, P2h, P3h, O1p, and O2p; in this subspace, size does not seem to play a role as far as the coherence is concerned. As for the average coherence time of the momenta,  $\langle\tau_\pi^{(l)}\rangle$  is uniformly lower than  $\tau_\pi$ , the difference being -62% in O1A, -66% in O2A, -63% in O1B, and -65% in O2B. These differences at 300 K and 370 K are very similar to those found for the larger systems, and show that this feature is not related either to the size of the system or to its biological function. It appears therefore that the hidden coherence in the momenta

subspace described before could be a generic characteristic of oligo- and polypeptides.

In conclusion, this article presents a new analytical method allowing insights into unexpected features of the phase space of poly- and oligopeptides: (i) a higher coherence of the momenta subspace in comparison with the positions subspace. (ii) A hidden dynamical coherence in the momenta subspace. (iii) The transformation of structural coherence into kinetic coherence when a protein is denaturated. These findings could be signposts to direct future investigations of a protein's dynamics.

Addendum. Experiments have shown that in a polypeptide chain collective vibrations exist that span a few amide groups moving in a well-defined phase relation;<sup>38</sup> in particular an absorption peak has been found in lysozyme at 1.2 THz, corresponding to a period of 0.83 ps, and has been attributed to a collective vibrational mode.<sup>39,40</sup> Another peak has been found in lysozyme at 2.8 THz, corresponding to a period of 0.36 ps, and assigned to an underdamped oscillator involving a dozen of residues.<sup>5</sup> The typical time range of those vibrations suggests a connection with the findings presented here. Also, it has been found in a model of  $\alpha$ -helix that over a distance of four residues a quantum coherence exists, allowing a coherent vibrational state that lasts up to 0.5–1.0 ps.<sup>41</sup> The time range of these findings overlaps with the range of classical coherence times found in this work, which hints at a possible quantumlike short-time behaviour of a classical system, as already found in a crystal at low temperature.<sup>42</sup>

## Conflicts of interest

There are no conflicts of interest to declare.

## References

- 1 V. N. Uversky and A. K. Dunker, The case for intrinsically disordered proteins (IDPs) playing contributory roles in molecular recognition without a stable 3D structure, *F1000 Biol. Rep.*, 2013, 5, 1–12.
- 2 R. V. Mannige, Dynamic New World: Refining Our View of Protein Structure, Function and Evolution, *Proteomes*, 2014, 2, 128–153.
- 3 I. N. Berezovsky, E. Guarnera, Z. Zheng, B. Eisenhaber and F. Eisenhaber, Protein function machinery: from basic structural units to modulation of activity, *Curr. Opin. Struct. Biol.*, 2017, 42, 67–74.
- 4 A. Xie, L. van der Meer and R. H. Austin, Excited-State Lifetimes of Far-Infrared Collective Modes in Proteins, *J. Biol. Phys.*, 2002, 28, 147–154.
- 5 D. A. Turton, H. M. Senn, T. Hardwood, A. J. Laphorn, E. M. Ellis and K. Wynne, Terahertz underdamped vibrational motion governs protein–ligand binding in solution, *Nat. Commun.*, 2014, 5(6), 3999.
- 6 S. Dutta, M. Ghosh and J. Chakrabarti, Spatio-temporal coordination among functional residues in protein, *Sci. Rep.*, 2017, 7(7), 40439.
- 7 I. Nardecchia, M. Lechelon, M. Gori, I. Donato, J. Preto, E. Floriani, S. Jaeger, S. Mailfert, D. Marguet, P. Ferrier and



- M. Pettini, Detection of long-range electrostatic interactions between charged molecules by means of fluorescence correlation spectroscopy, *Phys. Rev. E*, 2017, **96**(13), 022403.
- 8 H. Zhou and L. Wang, Chaos in Biomolecular Dynamics, *J. Phys. Chem.*, 1996, **100**, 8101–8105.
- 9 M. Braxenthaler, R. Unger, D. Auerbach, J. A. Given and J. Moulton, Chaos in Protein Dynamics, *Proteins: Struct. Funct. Genet.*, 1997, **29**, 417–425.
- 10 D. J. Nicholson, Is the cell really a machine?, *J. Theor. Biol.*, 2019, **477**, 108–126.
- 11 I. Mezic, On the dynamics of molecular conformation, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 7542–7547.
- 12 I. Nardecchia, J. Torres, M. Lechelon, V. Giliberti, M. Ortolani, P. Nouvel, M. Gori, Y. Meriguet, I. Donato, J. Preto, L. Varani, J. Sturgis and M. Pettini, Out-of-Equilibrium Collective Oscillation as Phonon Condensation in a Model Protein, *Phys. Rev. X*, 2018, **8**(23), 031061.
- 13 A. Battisti, R. G. Lalopa and A. Tenenbaum, Ordered and chaotic dynamics of collective variables in a butane molecule, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2009, **79**(13), 046206.
- 14 Disprot Data Bank entry DP00546.
- 15 Protein Data Bank entry 1AKI.
- 16 <https://zhanglab.dcm.b.med.umich.edu/I-TASSER/>.
- 17 J. Yang and Y. Zhang, I-TASSER server: new development for protein structure and function predictions, *Nucleic Acids Res.*, 2015, **43**, W174–W181.
- 18 C. Zhang, P. L. Freddolino and Y. Zhang, COFACTOR: improved protein function prediction by combining structure, sequence and protein–protein interaction information, *Nucleic Acids Res.*, 2017, **45**, W291–W299.
- 19 A. Amadei, A. B. M. Linssen and H. J. C. Berendsen, Essential Dynamics of Proteins, *Proteins: Struct. Funct. Genet.*, 1993, **17**, 412–425.
- 20 GROMACS release 2020.2, <https://www.gromacs.org/>.
- 21 W. Humphrey, A. Dalke and K. Schulten, VMD: visual molecular dynamics, *J. Mol. Graphics*, 1996, **14**, 33–38.
- 22 W. L. Jorgensen, D. S. Maxwell and J. Tirado-Rives, Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids, *J. Am. Chem. Soc.*, 1996, **118**, 11225–11236.
- 23 P. Mark and L. Nilsson, Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K, *J. Phys. Chem. A*, 2001, **105**, 9954–9960.
- 24 G. Bussi, D. Donadio and M. Parrinello, Canonical sampling through velocity rescaling, *J. Chem. Phys.*, 2007, **126**(8), 14101.
- 25 M. Parrinello and A. Rahman, Polymorphic transitions in single crystals: a new molecular dynamics method, *J. Appl. Phys.*, 1981, **52**, 7182–7190.
- 26 D. M. F. Van Aalten, B. L. De Groot, J. B. C. Findlay, H. J. C. Berendsen and A. Amadei, A comparison of techniques for calculating essential dynamics, *J. Comput. Chem.*, 1997, **18**, 169–181.
- 27 B. Hess, Convergence of sampling in protein simulations, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2002, **65**(10), 031910.
- 28 O. F. Lange and H. Grubmüller, Can Principal Components Yield a Dimension Reduced Description of Protein Dynamics on Long Time Scales?, *J. Phys. Chem. B*, 2006, **110**, 22842–22852.
- 29 M. D'Alessandro, A. Tenenbaum and A. Amadei, Dynamical and Statistical Mechanical Characterization of Temperature Coupling Algorithms, *J. Phys. Chem. B*, 2002, **106**, 5050–5057.
- 30 F. Eisenhaber, P. Lijnzaad, P. Argos, C. Sander and M. Scharf, The Double Cube Lattice Method: Efficient Approaches to Numerical Integration of Surface Area and Volume and to Dot Surface Contouring of Molecular Assemblies, *J. Comput. Chem.*, 1995, **16**, 273–284.
- 31 W. R. Krigbaum and F. R. Kügler, Molecular conformation of egg-white lysozyme and bovine alpha-lactalbumin in solution, *Biochemistry*, 1970, **9**, 1216–1223.
- 32 A. E. Mark and W. F. van Gunsteren, Simulation of the Thermal Denaturation of Hen Egg White lysozyme: Trapping the Molten Globule State, *Biochemistry*, 1992, **31**, 7745–7748.
- 33 A. Battisti, G. Ciasca, A. Grottesi and A. Tenenbaum, Thermal compaction of the intrinsically disordered protein tau: entropic, structural, and hydrophobic factors, *Phys. Chem. Chem. Phys.*, 2017, **19**, 8435–8446.
- 34 S. D. Stoddard and J. Ford, Numerical Experiments on the Stochastic Behavior of a Lennard-Jones Gas System, *Phys. Rev. A: At., Mol., Opt. Phys.*, 1973, **8**, 1504–1512.
- 35 C. Levinthal, *How to Fold Graciously*, *Mossbauer Spectroscopy in Biological Systems: Proceedings of a meeting held at Allerton House*, ed. J. T. P. DeBrunner and E. Munck, University of Illinois Press, Monticello, Illinois, 1969, pp. 22–24.
- 36 C. Levinthal, Are there pathways for protein folding?, *J. Chim. Phys. Phys.-Chim. Biol.*, 1968, **65**, 44–45.
- 37 G. Zocchi, Proteins unfold in steps, *Proc. Natl. Acad. Sci. U. S. A.*, 1997, **94**, 10647–10651.
- 38 P. Hamm, M. Lim and R. M. Hochstrasser, Structure of the amide I band of peptides measured by femtosecond nonlinear-infrared spectroscopy, *J. Phys. Chem. B*, 1998, **102**, 6123–6138.
- 39 A. Markelz, S. Whitmire, J. Hillebrecht and R. Birge, THz time domain spectroscopy of biomolecular conformational modes, *Phys. Med. Biol.*, 2002, **47**, 3797–3805.
- 40 J. Xu, K. W. Plaxco and S. J. Allen, Collective Dynamics of lysozyme in Water: Terahertz Absorption Spectroscopy and Comparison with Theory, *J. Phys. Chem. B*, 2006, **110**, 24255–24259.
- 41 A. G. Dijkstra, T. la Cour Jansen and J. Knoester, Modeling the Vibrational Dynamics and Nonlinear Infrared Spectra of Coupled Amide I and II Modes in Peptides, *J. Phys. Chem. B*, 2011, **115**, 5392–5401.
- 42 G. Marcelli and A. Tenenbaum, Quantumlike short-time behavior of a classical crystal, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2003, **68**(8), 041112.

