Breakout character of islet amyloid polypeptide hydrophobic mutations at the onset of type-2 diabetes

Rafael B. Frigori*

Universidade Tecnológica Federal do Paraná (UTFPR), Rua Cristo Rei 19, CEP 85902-490, Toledo (PR), Brazil

(Received 16 July 2014; revised manuscript received 16 October 2014; published 17 November 2014)

Toxic fibrillar aggregates of islet amyloid polypeptide (IAPP) appear as the physical outcome of a peptidic phase transition signaling the onset of type-2 diabetes mellitus in different mammalian species. In particular, experimentally verified mutations on the amyloidogenic segment 20-29 in humans, cats, and rats are highly correlated with the molecular aggregation propensities. Through a microcanonical analysis of the aggregation of $IAPP_{20\text{-}29}$ isoforms, we show that a minimalist one-bead hydrophobic-polar continuum model for protein interactions properly quantifies those propensities from free-energy barriers. Our results highlight the central role of sequence-dependent hydrophobic mutations on hot spots for stabilization, and thus for the engineering, of such biological peptides.

DOI: 10.1103/PhysRevE.90.052716 PACS number(s): 87.15.Cc, 87.15.A−, 64.60.Q−, 87.15.Zg

I. INTRODUCTION

Diabetes mellitus type 2 (DM-II) is a metabolic disorder characterized by hyperglycemia, due to insufficient insulin secretion from pancreatic β-cells in the setting of insulin resistance. Beyond the yearly premature death of about 4 million people worldwide, diabetes also implies a high prevalence of health complications including stroke (68%), high blood pressure (67%), blindness (28.5%), kidney disease (44%), neuropathies, and amputation (60%) [1]. Its outbreak is correlated to genetic factors associated with a sedentary modern lifestyle, which implies that an increasing global diabetes epidemic is underway. In 2010, there were 285 million cases in adults worldwide, with an estimated annual health care economic burden of USD 376 billion [2]. Such a scenario calls for deepening the pathophysiological understanding of the DM-II onset.

In this vein, since the pioneering study by Westermark *et al.* in the 1990s [3], it has become increasingly known that amylin (or IAPP), a small 37-residue putative polypeptide (small protein) hormone also produced by pancreatic β-cells, constitutes most fibrillar amyloid deposits seen in the islets of Langerhans in diabetic humans [4] and other mammals. Further experimental studies [5] have demonstrated that fibrillar amylin is toxic to insulin-producing β-cells, thus inducing an enhanced loss of islet cells characteristic of type-2 diabetes. In addition, the propensity for islet amyloid deposition is specie-specific, a property mostly due to mutations in hot spots such as the $IAPP_{20\text{-}29}$ segments [3], which correlates positively with the molecular toxicity of IAPP isoforms in humans (hIAPP) and cats (cIAPP), while most rodents (rIAPP) never develop such a syndrome. Thus, the toxicity of amylin seems strikingly similar to the effects observed in other well-known amyloidosis [6], such as Alzheimer's disease and spongiform encephalopathies.

Generally, while the ability of polypeptides to form such amyloid structures is considered to be a common feature of such molecular chains, the propensity to do so varies markedly between different sequences. Therefore, aggregation rates correlate [7] with physicochemical properties of those molecules such as charge, secondary structure, and hydrophobicity [8]. Hence, peptide proneness for aggregation can be related to eventual misfoldings [6], which is explained by the thermostatistical theory of the energy landscape of protein folding [9]. In accordance to whom, realistic models of proteins are minimally frustrated heteropolymers that reach the lowest-energy (native, or folded) state through an ensemble of intermediate self-organizing structures, guided by a rugged funnel-like energy landscape. Although details on the native conformation of proteins may depend on specificities of each energetic potential, coarse-grained models for amino acid interactions, where more or less profound simplifications are made, have provided powerful insights into the aggregation mechanisms underlying degenerative diseases [10].

Peptides are small proteins composed of inhomogeneous sequences of amino acids (residues), so they constitute a class of finite systems inherently far from the thermodynamic limit, to whose description the thermostatistical postulate of ensemble equivalence does not hold. Thereby, the original microcanonical formulation of statistical mechanics [11], designed to be rigorously valid even for systems with finite degrees of freedom, turns to be most appropriate for studying phase transitions on proteins such as folding [12] and aggregation [13]. In this approach, starting from the density of states $g(E)$, the celebrated Boltzmann entropy $S(E) = k_B \ln g(E)$ is solely responsible for yielding thermodynamic quantities, such as the microcanonical temperature $T(E)$ and the specific heat $C_V(E)$. Additionally, free energies $H(E)$ can also be straightforwardly accessed by taking Legendre transforms. However, in this context, it should be noted that $S(E)$ may become a convex function of E, such as during first-order phase transitions, which induces peculiar thermodynamic behaviors such as backbendings on $T(E)$, negative values of $C_V(E)$, and the appearance of energetic barriers on free energies ΔH [11].

In this article we show how, through a microcanonical analysis from multicanonical Monte Carlo simulation data [14], the ratios among aggregation propensities of IAPP isoforms can be recovered from the energetic barriers emerging in the vicinity

*frigori@utfpr.edu.br

of the (first-order) phase transitions of a simple coarse-grained hydrophobic-polar model for protein interactions [13,15]. Our results depend weakly on an input scale [16], and they agree nicely with widely accepted heuristic predictors that can reproduce *in vitro* as well as *in vivo* experimental data [17,18]. In the spirit of [19], we conclude that even a two-letter code can discriminate amyloidogenic characters on primary sequences of IAPP. This corroborates with an underlying rationale relating the thermodynamic aspects associated with sequence-dependent hydrophobic mutations with the (kinetic) aggregation rates of peptides [7], so that more aggregation-prone sequences also form pathogenic aggregates faster.

The work is organized as follows: In Sec. II, an effective one-bead hydrophobic-polar continuum model for describing protein interaction and aggregation is introduced. Section III is devoted to the algorithmic setup and numerical results emerging from our multicanonical simulations of $\mathrm{IAPP}_{20\text{-}29}$ segments of several mammalian species. In Sec. IV, free energies are exploited to connect thermodynamic and kinetic aspects of peptide aggregation. There, we propose a method to evaluate relative aggregation propensities of proteins, a rationale inspired in spectral predictions by universality related theories. Those results are validated by confrontation with well-established heuristic online aggregation-propensity estimators. Section V summarizes our results, corroborating them with recent all-atom simulations and thus highlighting future research perspectives. We devote an Appendix to numerical error estimates in microcanonical data analysis.

II. AN EFFECTIVE MODEL FOR PROTEIN AGGREGATION

Hydrophobic forces are not fundamental forces of nature [8]. Nevertheless, by considering their central role in the assembling of three dimensionally ordered tertiary structures during protein folding, while keeping high simplicity standards on molecular modeling, we have adopted a coarse-grained (one-bead) hydrophobic-polar model for proteins [15,20]. There the target protein is mapped, depending on the hydrophobic character of the constituents lying on its primary sequence of amino acids, on a heteropolymer made of hydrophobic (A) or polar (B) pseudoatoms (beads).

Those monomers replace the original residues on their α-carbon positions occupied at the same peptidic backbone structure. The interaction energy ($\mathcal{H}$) among the N pseudoatoms in the chain is given by

$$\mathcal{H} = \frac{1}{4}\sum_{k=1}^{N-2}(1-\cos\alpha_k) + 4\sum_{i=1}^{N-2}\sum_{j=i+2}^{N}\Phi(r_{ij};C_{\sigma_i,\sigma_j}), \quad (1)$$

where the first term describes the virtual bending angle ($0 \leqslant \alpha_k \leqslant \pi$) between three successive monomers, while the second term,

$$\Phi(r_{ij};C_{\sigma_i,\sigma_j}) = \left[r_{ij}^{-12} - C(\sigma_i,\sigma_j)r_{ij}^{-6}\right], \quad (2)$$

provides a long-distance (r_{ij}) pairwise interaction between residues i and j, depending on their hydrophobic character $\sigma \in \{A,B\}$. That is,

$$C\left(\sigma_i,\sigma_j\right) = \begin{cases} +1, & \sigma_i,\sigma_j = A, \\ +1/2, & \sigma_i,\sigma_j = B, \\ -1/2, & \sigma_i \neq \sigma_j. \end{cases} \quad (3)$$

Then, attractive ($C_{A,A}$,$C_{B,B}$) or repulsive ($C_{A,B}$,$C_{B,A}$) forces will naturally emerge from primary sequences of amino acids once they are properly translated on a two-letter code by a hydrophobic scale [16] used as a lexicon.

Because aggregation is a many-body effect, it shall be provided by a multiprotein potential [13]

$$\Psi_{\mathrm{multiprot}} = \sum_{k=1}^{M}\left[\mathcal{H}_k + \sum_{l>k}^{N}\sum_{i,j=1}\Phi\left(r_{l_i k_j};C_{\sigma_{l_i},\sigma_{k_j}}\right)\right]. \quad (4)$$

Thus, in addition to the intraprotein energy $\mathcal{H}_k$, from Eq. (1) there is a contribution from all pairs of residues (l_i,k_j) located in different proteins (l or k) of a set of M proteins. It should be noted that in such a coarse-grained model, long-range forces are only due to hydrophobic/polar effective interactions, so any interacting pair (i,j) of pseudoatoms in the system is equally described by the same C_{σ_i,σ_j} coupling constants [Eq. (3)]. This is clearly a simplifying hypothesis inspired by mean-field descriptions, justified as a leading-order approach, in the sense of a renormalization-group analysis.

III. SIMULATIONS AND THERMODYNAMIC RESULTS

To obtain the microcanonical entropy associated with the aggregation of segments of IAPP isoforms, and thus their caloric and specific-heat curves, we have focused on performing Monte Carlo multicanonical (MUCA) simulations [14] of multiple (amyloidogenic) $\mathrm{IAPP}_{20\text{-}29}$ segments. Coefficients a_k and b_k in multicanonical weights $\omega_{\mathrm{MUCA}}(E_k) = e^{b_k E_k - a_k}$ can be determined by an iterative procedure using energy histograms $H_{\mathrm{MUCA}}(E)$ for energies E_k in an interval $E = [E_0,\ldots,E_{\max}]$. Thus, in the beginning, one sets $\omega^0_{\mathrm{MUCA}}(E) = 1$ for all energies, which is used to run a usual METROPOLIS simulation to build $H^0_{\mathrm{MUCA}}(E)$. The next guess for the weights in the simplest update scheme is given by $\omega^1_{\mathrm{MUCA}}(E) = H^0_{\mathrm{MUCA}}(E)/\omega^0_{\mathrm{MUCA}}(E)$. Such an iterative procedure is then repeated until the energy histogram converges to a "flat" distribution. In our implementation, we have employed accumulated error-weighted histograms, so the statistics for weights estimation improves every run, while convergence is ensured by Berg's weight recursion [21].

Once MUCA weights are established, they provide a good piecewise approximation to the microcanonical entropy $S_{\mathrm{micro}}(E_k) = b_k E_k - a_k$. Then, numerical derivatives of the entropy can be employed to compute thermodynamic quantities of interest [11], as the microcanonical caloric curve

$$\beta(E) \equiv T^{-1}(E) = \frac{\partial S}{\partial E}, \quad (5)$$

the microcanonical specific heat

$$C_V(E) = \frac{dE}{dT} = -\left(\frac{\partial S}{\partial E}\right)^2\left(\frac{\partial^2 S}{\partial E^2}\right)^{-1}, \quad (6)$$

TABLE I. The inverse temperature at aggregation β_c, the free-energy barriers $\triangle H$, and latent specific-heat L, for human (hIAPP), cat (cIAPP), and rat (rIAPP) segments of amylin isoforms obtained by microcanonical analysis from MUCA simulations.

Peptide	Sequence	AB Sequence	β_c	$\triangle H$	L
$hIAPP_{20\text{-}29}$	SNNFGAIL**SS**	BBBABAAA**BB**	2.60(2)	0.079(1)	0.520(5)
$cIAPP_{20\text{-}29}$	SNNFGAIL**SP**	BBBABAAA**BA**	2.09(1)	0.092(1)	0.636(10)
$rIAPP_{20\text{-}29}$	SNNLGPVL**PP**	BBBABAAA**AA**	1.48(1)	0.124(1)	0.698(10)

and the free energy

$$H(E) = E - \left(\frac{\partial S}{\partial E}\right)^{-1} S(E). \tag{7}$$

In particular, we have applied the finite-differences method with central derivatives and stability constraints to fix their maximal kernel sizes.

After the original protein sequences of amyloidogenic $IAPP_{20\text{-}29}$ segments were mapped on the AB model (in the second and third columns of Table I, respectively) through a hydrophobicity scale [16], simulations with different ensembles—having up to eight copies—of interacting peptides were done. For the Monte Carlo evolution of a set having N pseudoatoms, we performed 10^3 MUCA iterations, in a total of about $3N \times 10^9$ updates, by mixing spherical-cap [13] and pivoting [22] algorithms. The independence of our results with energy-bin size—i.e., $\triangle E = E_{k+1} - E_k$—and finite-box effects was checked to certify the robustness of the data (see the Appendix). In particular, as containers we have used spherical boxes (of radius $R \simeq 100$), whose interior was initially populated with stretched and randomly positioned peptides. The Boltzmann constant was taken as $k_B = 1$, and the distances of nearest-neighbor residues were normalized to unity. For convenience, we use intensive units for the system energy $\varepsilon = E/N$. Error bars were computed by data-blocking and resampling techniques.

The data analysis of our simulations shows that (see Fig. 1 and Table I for a summary), as is usually observed in first-order phase transitions, described under the microcanonical formalism of statistical mechanics [11], negative specific heats are seen in regions where the microcanonical entropy presents a convex intruder. Inside such regions, delimited by a minimum energy below whom proteins are completely aggregated ($\varepsilon_{\min} = \varepsilon_{\text{agg}}$) and a maximal one above which aggregates dissolve by fragmentation ($\varepsilon_{\max} = \varepsilon_{\text{frag}}$), the caloric curves $\beta(\varepsilon) \times \varepsilon$ display signals of thermodynamic metastability. Those configurations induce a forbidden region for the canonical ensemble, which entails the need to apply a well-known Maxwellian prescription around the (inverse) temperature of transition β_c, whereas the (upper) A_+ and (lower) A_- areas of the bumps formed by the backbending of $\beta(\varepsilon_{\text{frag}} < \varepsilon < \varepsilon_{\text{agg}}) \neq \beta_c$ become equal. As a result, this implies not only the inequivalence of canonical and microcanonical ensembles during the phase transition, once a bijective mapping between the system temperature and energy is only possible for $\varepsilon < \varepsilon_{\text{agg}}$ or $\varepsilon > \varepsilon_{\text{frag}}$, but also the rising of a latent heat defined by $L = \varepsilon_{\text{frag}} - \varepsilon_{\text{agg}}$.

After evaluating β_c for the aforementioned amylin segment isoforms of humans, $\beta_{c\text{-hIAPP}} = 2.60(2)$, cats, $\beta_{c\text{-cIAPP}} = 2.09(1)$, and rats, $\beta_{c\text{-rIAPP}} = 1.48(1)$, we found the foregoing description to be appropriate for those regions of phase transition. Negative microcanonical specific heats, and thus latent (canonical) heats, can be seen inside $\varepsilon_{\text{hIAPP}} = [-0.439, 0.008]$ for humans, $\varepsilon_{\text{cIAPP}} = [-0.595, 0.041]$ for cats, and $\varepsilon_{\text{rIAPP}} = [-0.683, 0.015]$ for rats. Among such isoforms, hIAPP has

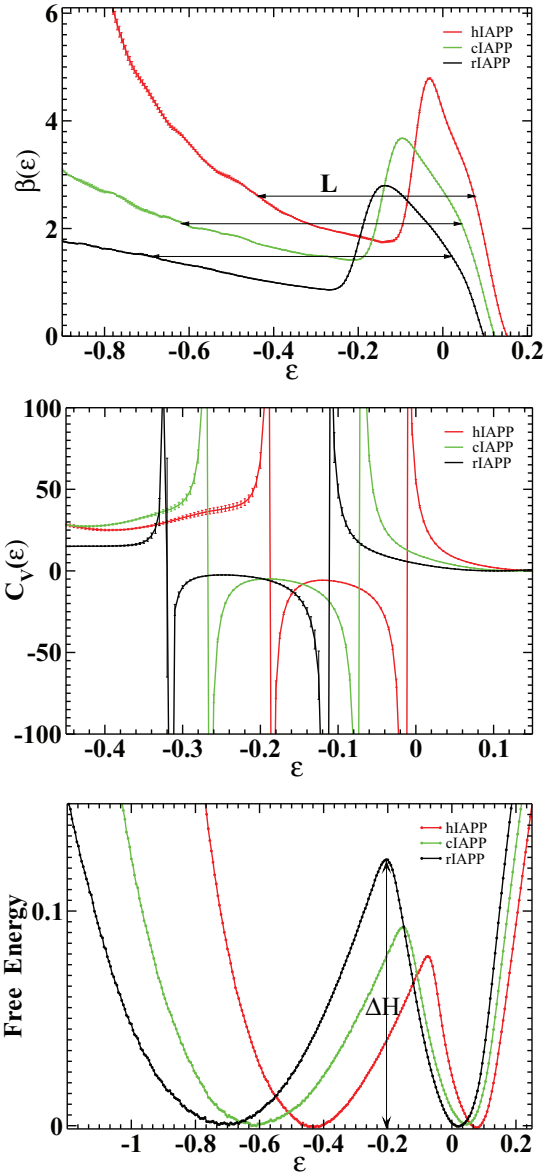

FIG. 1. (Color online) The thermodynamic behavior of segments of amylin isoforms such as cIAPP in green (light gray), rIAPP in black (black), and hIAPP in red (dark gray) as a function of energy (E) per residue $\varepsilon = E/N$. Upper panel: caloric curve. Center panel: microcanonical specific heat. Lower panel: the Helmholtz free-energy barrier (error bars are smaller than the circles).

in fact the smallest latent heat of transition $L_{\mathrm{hIAPP}} = 0.520$ and the lowest energetic barrier $\triangle H_{\mathrm{hIAPP}} = 0.079(1)$ for aggregation, which is followed by cIAPP, where $L_{\mathrm{cIAPP}} = 0.636$ and $\triangle H_{\mathrm{cIAPP}} = 0.092(1)$, and rIAPP with $L_{\mathrm{rIAPP}} = 0.698$ and $\triangle H_{\mathrm{rIAPP}} = 0.124(1)$. Latent heat is a consequence of the free-energy barrier and prevents the system from moving to a stable configuration in the new phase. Therefore, the smaller the latent heat, the higher the probability that a spontaneous thermal fluctuation will give rise to the aggregate phase.

IV. AGGREGATION PROPENSITIES

Since polypeptide aggregation is an example of a nucleated polymerization reaction where from a tiny nucleating event larger aggregates grow into fibrillar structures, the efficiency of these reactions is related to the rate of aggregation [23]. By the Arrhenius equation, it is also widely known [24] that thermodynamic and kinetic properties are connected by the relation $\Delta H = N_A \beta^{-1} \ln K_D$, where $\triangle H$ is the free energy of aggregation, N_A is the Avogadro number, and K_D is the dissociation equilibrium constant related to the ratio of dissociation(k_-)/association(k_+) rates in a two-state binding reaction. Higher aggregation propensities (henceforth denoted "z") are therefore associated with lower values of K_D, or equivalently with faster association rates, which is thermodynamically favored by smaller energetic barriers (so $z \propto \Delta H^{-1}$). This implies a causal relation, experimentally already observed for variants of β-amyloid proteins, where the more stable aggregates are also the ones that aggregates more readily [25].

Thus, in principle, one would expect that accurate information about phase transitions such as protein aggregation could be obtained only by atomic-level simulations. However, in the vicinity of critical phase transitions—where correlation lengths become greater than the characteristic system sizes—different physical systems can exhibit the same universal behavior, which constitutes a powerful predictive tool of statistical mechanics. For instance, in the context of protein folding, an effective lattice gauge field theory built only upon symmetry arguments [26] was shown to be, in the sense of the compactness index, in the same universality class of proteins deposited in the Protein Data Bank. Despite the conceptual simplicity of such a model, even the secondary structural motifs of all studied proteins could be reconstructed with a backbone RMSD accuracy of about 1 Å [27]. This success arguably relies on the fact that in such field-theoretic language, the formation of protein loops can be described by topological domain-wall solitons, interpolating among ground states given by α-helices and β-strands, despite the local details of their Hamiltonian interactions [27].

In a somehow similar scenario, the long-standing conjecture by Svetitsky and Yaffe [28], relating the magnetic phase transitions in d-dimensional Z_N Potts-like spin models and deconfinement in SU (N) quantum gauge theories in $(d+1)$ dimensions, has been widely verified beyond usual realms of critical exponents and universal amplitude ratios (see [29] and references therein). In fact, through universality, the emergence of bound states in the broken symmetry phase of spin systems was unveiled to be a phenomenon closely related to the formation of a quark-gluon plasma (QGP) [29], where the gluonic potential among static quark charges becomes short-ranged by acquiring a spectrum of effective Debye-screening masses (m_D). There, changes in the system free energy $H\,(r,T)$ at (asymptotic) large quark-distances are given by $\triangle H_\infty(T) \equiv \lim_{r\to\infty} H(r,T) \propto m_D\,(T)$ [30]. Regardless of the fact that those excited spectra are not universal, the ratios computed among their mass states (in the same channel) were shown to be [29]. More surprisingly, even when phase transitions are weak first-order—as it happens in quenched SU (3) QCD and the Z_3-Potts model—those respective ratios computed from both (approximate) universality-related theories still coincide up to a precision of 30% [31].

Inspired by those concepts, we propose that the aggregation propensities z_a and z_b for peptidic isoforms a and b may be combined to form a dimensionless ratio r_{ab} that depends only on relative changes in the system free energy, i.e., $r_{ab} = z_a z_b^{-1} = [\triangle H_a]^{-1}\,[\triangle H_b]$. Thenceforth, by performing such analysis over the data obtained from our AB-model simulations (see Table I), we have obtained relative aggregation propensities shown by the following ratios: $r_{hc}^{AB} \simeq 1.16\,(2)$,$r_{hr}^{AB} \simeq 1.57\,(2)$, and $r_{cr}^{AB} \simeq 1.35\,(2)$. How far one can lead such an argument is a matter for numerical verification, so we intend to cross-check our results with alternative methods for further validation.

To accomplish this very end, we have chosen two different heuristic algorithms designed to accurately predict—after being properly calibrated—the aggregation propensities z of a plethora of *in vitro* (z_{agg}) as well as *in vivo* (z_{scan}) experiments. First, we evaluated the so-called z_{agg} score from Zyggregator [17], a phenomenological model that incorporates both *intrinsic* factors of peptides such as hydrophobicity, charge, and the propensity of the polypeptide chain to adopt α-helical or β-sheet structures as well as *extrinsic* ones (physico-chemical properties related to the environment). Consequently, higher scores mean that a sequence is more suitable to aggregation. This approach for the primary sequences of IAPP has resulted in the following aggregation-propensity scores: $z_{\mathrm{agg}}^{\mathrm{hIAPP}} = 1.30(9)$,$z_{\mathrm{agg}}^{\mathrm{cIAPP}} = 1.05(15)$, and $z_{\mathrm{agg}}^{\mathrm{rIAPP}} = 0.92(13)$. So, computing the ratios among those scores—such that $r_{ab}^{z_{\mathrm{agg}}} = [z_{\mathrm{agg}}^a][z_{\mathrm{agg}}^b]^{-1}$—has produced these relative aggregation propensities $r_{hc}^{z_{\mathrm{agg}}} \simeq 1.24(19)$,$r_{hr}^{z_{\mathrm{agg}}} \simeq 1.41(22)$, and $r_{cr}^{z_{\mathrm{agg}}} \simeq 1.14(23)$.

On the other hand, AGGRESCAN [18] is an online aggregation-propensity predictor based solely on *in vivo* experimental data. It assumes that short and specific segments of peptidic sequences modulate protein aggregation, and, as an outcome, the effects of genetic mutations on aggregation propensities (of an input sequence) can be precisely predicted from comparisons with a databank. The generated score z_{scan} for amylin isoforms is given by $z_{\mathrm{scan}}^{\mathrm{rIAPP}} = -8.80$, $z_{\mathrm{scan}}^{\mathrm{cIAPP}} = -6.60$, and $z_{\mathrm{scan}}^{\mathrm{hIAPP}} = -5.60$, where more negative values imply naturally less aggregation-prone sequences. After due normalization, the ratios among relative aggregation propensities are analogously obtained: $r_{hc}^{z_{\mathrm{scan}}} \simeq 1.18$,$r_{hr}^{z_{\mathrm{scan}}} \simeq 1.57$, and $r_{cr}^{z_{\mathrm{scan}}} \simeq 1.39$.

Thus, our results for r^{AB} are compatible (within less than 1 stdv.) with ratios of aggregation propensities estimated from *in vitro* phenomenological methods ($r^{z_{\mathrm{agg}}}$), whereas when compared to ratios obtained from *in vivo* data-based methods ($r^{z_{\mathrm{scan}}}$) discrepancies were lower than 2%, in even better

agreement. Such numbers reinforce not only our previous working hypothesis that hydrophobic mutations play an essential role in the determination of peptide stability, but also that (pathological) substitutions are strongly sequence-dependent on the so-called protein hot spots, as is the case of $IAPP_{20\text{-}29}$ [3].

More interestingly, from a thermodynamic viewpoint, the height of energetic barriers is associated not only with nucleation rates but also with reaction kinetics as the required time (time-lag τ_c) for reaching steady-state nucleation, that is, $\tau_c \propto \exp(\beta \triangle H)$ [32]. From such a perspective, less stable molecular isoforms of IAPP—i.e., the ones with smaller latent heats, or equivalently having lower energetic barriers—would induce a quicker production of IAPP aggregates on a mammalian pancreas, as is the case of humans (hIAPP) and cats (cIAPP). For more stable isoforms, as in rats (rIAPP), the huge time scales associated would be a deterrent pathophysiological factor for the onset of Diabetes II.

This could lead to an alternative pathway for *in silico* designing of artificial peptides aiming to act as adjuncts for DM-II, under the constraint that they must keep biocompatibility with usual amylin, while they should avoid its notorious metastability. For instance, until recently, these features could be found just in Pramlintide [33], an experimentally screened rat-modified version of IAPP.

V. CONCLUDING REMARKS

In this article, we have shown that by performing microcanonical analysis of a simple coarse-grained hydrophobic-polar heteropolymer model for aggregation of proteins, which are mapped by a hydrophobicity scale in a two-letter code lexicon, the onset of type-2 *Diabetes mellitus* in different mammalian species correlates with aggregation propensities derived from the thermodynamics of first-order aggregation transitions of specific segments of amylin isoforms ($IAPP_{20\text{-}29}$). The (almost) universal ratios among such aggregation propensities extracted from our *ab initio* multicanonical simulations were in nice agreement with well-established heuristic predictors. It corroborates to a rationale underlying the thermodynamics of sequence-dependent hydrophobic mutations on peptides (hot spots) with the kinetic aspects of their associated polymerization reactions, hence more aggregation-prone (i.e., less stable) sequences shall aggregate faster. These findings may bring potentially new insights for designing and screening peptides as adjuncts for DM-II therapy from *in silico* methods.

Still, such conclusions are confirmed by recent studies, where some groups have been succeed on simulating the aggregation processes of IAPP through molecular-dynamics (MD) techniques using all-atom potentials with an implicit [34] or explicit solvent [35]. For instance, in [35] the authors investigated the aggregation of decamers of hIAPP and rIAPP in double layers. Then, by comparing average intermolecular distances (R) and the van der Waals interaction energies ($\triangle H$) among those rIAPP and hIAPP aggregates, it was found that $R_{\text{hIAPP}} \simeq (3.7 \pm 0.3)$ Å and $R_{\text{rIAPP}} \simeq (4.2 \pm 0.7)$ Å while $\triangle H_{\text{hIAPP}} = (-233.6 \pm 24.7)$ kcal/mol and $\triangle H_{\text{rIAPP}} = (-326.5 \pm 64.5)$ kcal/mol. So it has arguably been verified that differences in stability between those IAPP isoforms—concerning their molecular compactness index and free-energy differences—are most likely due to the existence of β-sheet breaking (hydrophobic) Prolines in the $rIAPP_{25\text{-}29}$ segment, which is missing in $hIAPP_{25\text{-}29}$.

Surprisingly enough, by using the aforementioned data in our present methodology, one finds a relative aggregation propensity $r_{hr}^{\text{MD}} \simeq 1.39(31)$, which is in remarkable agreement with our result $r_{hr}^{AB} \simeq 1.57(2)$. This provides not only a compelling verification for the correctness of our working hypothesis in Sec. IV, which relies on general universality-based arguments, but it also constitutes further evidence for the breakout character of IAPP hydrophobic mutations at the onset of type-2 diabetes. Finally, it would be very interesting to investigate how slightly extended letter codes for amino acids (as in [36]), together with more refined coarse-grained models [10,19], may impact quantitative predictions of aggregation propensities of other mammalian IAPP isoforms eventually able to work as natural aggregation deterrents.

ACKNOWLEDGMENTS

The author thanks Jorge Chahine, Leandro G. Rizzi, Nelson A. Alves, and Rinaldo W. Montalvão for useful discussions. We also thank the anonymous referees for their highly constructive work. This project was partially funded by the Brazilian agency Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Simulations were performed on the SGI-Altix at CENAPAD-Unicamp.

APPENDIX

In a microcanonical simulation [11], the entropy $S(E_k) = b_k E_k - a_k$ is estimated as a piecewise function, hence energies are discretized in M bins: $\triangle E = E_{k+1} - E_k$. Thus, $S(E)$ is built up through a recursive process, where energy histograms are accumulated during a series of Monte Carlo runs [14,21]. After the entropy is obtained, up to good numerical precision, numerical derivatives can be employed to directly extract the system thermodynamics, Eqs. (5)–(7). Therefore, in such simulations, not only are statistical and finite (box) size effects important factors to ensure data robustness, but also evaluating the most appropriate energy bin size and checking against finite-difference instabilities.

As an illustration, we provide the output of some data analysis we have performed as preliminary simulations to set the parameters for our production runs. The system we have chosen is a small ensemble consisting of two $hIAPP_{20\text{-}29}$ molecules ($N = 2\times10$), which was simulated by methods already described in Sec. III, and whose parameters—energy-bin sizes $\triangle E$, the radius of the spherical container R, and the kernel size of numerical derivatives dE—were systematically varied. For every MUCA run, we have accumulated statistics during 10^6 Monte Carlo evolution steps; a summary of our results is given in Fig. 2. Error bars were computed by usual data-blocking and resampling methods [37]

Concerning our checks for energy-bin sizes, for each value of $\triangle E$ we have initially evaluated the ground-state E_{ground} for the dimerization of $hIAPP_{20\text{-}29}$ during 350 MUCA runs. There we have obtained $E_{\text{ground}} = \{-18.69, -18.23, -17.84, -20.32\}$, respectively, for $\triangle E = \{0.1, 0.2, 0.5, 1.0\}$. Thus, while larger values of $\triangle E$ apparently favor sampling lower-energy states—most likely due to improved signal-to-noise

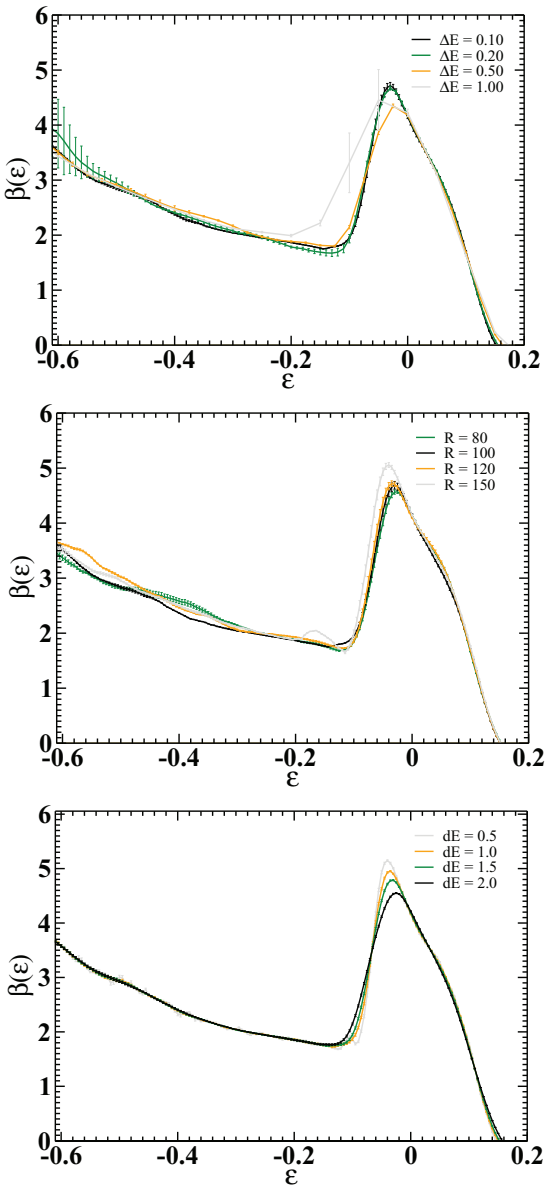

FIG. 2. (Color online) Caloric curves for $2\times$hIAPP$_{20\text{-}29}$ as a function of energy (E) per residue $\varepsilon = E/N$ as an illustration of various error sources in microcanonical simulations. Upper panel: energy bin-size effects for $\triangle E = 0.1$ in black (black), $\triangle E = 0.2$ in green (dark gray), $\triangle E = 0.5$ in orange (gray), and $\triangle E = 1.0$ in light gray (light gray). Central panel: finite-volume effect as a function of the linear size of the container radius for $R = 80$ in green (dark gray), $R = 100$ in black (black), $R = 120$ in orange (gray), and $R = 150$ in light gray (light gray). Lower panel: kernel-sized effects in finite-difference derivatives for $dE = 0.5$ in light gray (light gray), $dE = 1.0$ in orange (gray), $dE = 1.5$ in green (dark gray), and $dE = 2.0$ in black (black).

ratios—they provide us with a coarser mesh that prevents smooth numerical derivatives of $S(E)$ to be safely employed around the transition. This fact is demonstrated by the caloric curve $\beta(\varepsilon) \times \varepsilon$—where $\varepsilon = E/N$—depicted in the upper panel of Fig. 2, where one perceives the coarsening effects induced by $\triangle E$ in a region in the vicinity of the phase transition. In general, the peak of $\beta(\varepsilon) \times \varepsilon$ was shifted by no more than 6% when considering the extreme values of the bins employed here.

By fixing $\triangle E = 0.1$, we have investigated along about 300 MUCA runs the effect of taking a different radius R in our simulations. Analogously to our previous analysis, we have obtained $E_{\text{ground}} = \{-20.21, -18.68, -17.59, -19.59\}$ when taking $R = \{80, 100, 120, 150\}$. The fluctuation of energy values found as ground states as a function of increasing volumes is compatible with an unbiased statistical fluke, thus presenting no systematics. Also, except for an exceptional 5% deviation on the height of the peak of $\beta(\varepsilon) \times \varepsilon$—implying a small shift on its energy location $\Delta\varepsilon/\varepsilon < 1\%$—seen when $R = 150$, all curves physically match. Additionally, it also deserves to be noted that the smaller container we employed has a linear extension about 20 times larger than fully distended peptides simulated here. Thus, as observed in the central panel of Fig. 2, the volume independence of our aggregation studies seems to us to be a plausible working hypothesis.

The effects of finite-difference derivatives on data analysis was checked by considering the output from a full-scale MUCA simulation, performed using $\triangle E = 0.1$, $R = 100$, and 1000 MUCA iterative runs, each taking 10^7 MC steps. Here the derivatives of $S(E)$ were computed as central finite differences dE for n-point kernels, which is translated on our setup as $n = \{5, 10, 15, 20\} \leftrightarrow dE = \{0.5, 1.0, 1.5, 2.0\}$. Results observed (in the Lower Panel of Fig. 2) for $\beta(\varepsilon)$ clearly show that employing relatively small kernel sizes (e.g., $5 \leqslant n \leqslant 15$) may notably improve signal-to-noise levels when computing (high-order) derivatives of $S(E)$, at the expense of introducing some systematics in highly curved regions. More interestingly, despite gradually incrementing the kernel size from $n = 5$ up to $n = 15$—which is able to fully suppress most statistical noise—it only induces a maximal 7(1)% shift on the height of the peak of $\beta(\varepsilon) \times \varepsilon$. Thus, while using this technique is mandatory to compute quantities such as Eq. (6), the determination of transition temperatures as in Sec. III may be more successful not by employing Maxwell's constructions over caloric curves, but by using "shifted entropies" [20]. Hence, in such an approach one has just to iteratively operate directly on $S(\varepsilon)$ by numerically searching for β_c while imposing the following physical constraint: $H(\varepsilon)|_{\varepsilon=\varepsilon_{\text{frag}}} \equiv H(\varepsilon)|_{\varepsilon=\varepsilon_{\text{agg}}} = [\varepsilon - \beta_c^{-1} S(\varepsilon)]|_{\varepsilon=\varepsilon_c}$, which is equivalent to saying that at the temperature of transition β_c^{-1}, the free energy $H(\varepsilon)$ on Eq. (7) has an equal and doubly degenerate minimum.

[1] S. Melmed *et al.* *Williams Textbook of Endocrinology* (Elsevier/Saunders, Amsterdam, 2011); American Diabetes Society, http://www.diabetes.org/
[2] P. Zhang *et al.*, Diabetes Res. Clin. Pract. **87**, 293 (2010).
[3] P. Westermark *et al.*, Proc. Natl. Acad. Sci. USA **84**, 3881 (1987); **87**, 5036 (1990).
[4] R. L. Hull *et al.*, J. Clin. Endocrinol. Metab. **89**, 3629 (2004).
[5] A. Lorenzo *et al.*, Nature (London) **368**, 756 (1994).
[6] C. M. Dobson, Nature (London) **426**, 884 (2003).
[7] F. Chiti *et al.*, Nature (London) **424**, 805 (2003).
[8] D. Chandler, Nature (London) **437**, 640 (2005).

[9] J. N. Onuchic, Z. Luthey-Schulten, and P. G. Wolynes, Annu. Rev. Phys. Chem. **48**, 545 (1997).
[10] V. Tozzini, Curr. Opin. Struct. Biol. **15**, 144 (2005); C. Wu and J.-E. Shea, *ibid.* **21**, 209 (2011).
[11] D. H. E. Gross, *Microcanonical Thermodynamics* (World Scientific, Singapore, 2001).
[12] J. Hernandez-Rojas and J. M. Gomez Llorente, Phys. Rev. Lett. **100**, 258104 (2008).
[13] C. Junghans, M. Bachmann, and W. Janke, Phys. Rev. Lett. **97**, 218103 (2006).
[14] B. A. Berg and T. Neuhaus, Phys. Lett. B **267**, 249 (1991); Phys. Rev. Lett. **68**, 9 (1992).
[15] F. H. Stillinger, T. Head-Gordon, and C. L. Hirshfeld, Phys. Rev. E **48**, 1469 (1993); F. H. Stillinger and T. Head-Gordon, *ibid.* **52**, 2872 (1995).
[16] M. A. Roseman, J. Mol. Biol. **200**, 513 (1988).
[17] K. F. DuBay *et al.*, J. Mol. Biol. **341**, 1317 (2004).
[18] O. Conchillo-Solé *et al.*, BMC Bioinfo. **8**, 65 (2007).
[19] S. Brown, N. J. Fawzi, and T. Head-Gordon, Proc. Natl. Acad. Sci. USA **100**, 10712 (2003).
[20] R. B. Frigori, L. G. Rizzi, and N. A. Alves, J. Chem. Phys. **138**, 015102 (2013).
[21] B. A. Berg, Fields Inst. Commun. **26**, 1 (2000), arXiv:cond-mat/9909236; Comput. Phys. Commun. **153**, 397 (2003).
[22] T. Kennedy, J. Statist. Phys. **106**, 407 (2002).
[23] T. Christopeit *et al.*, Prot. Sci. **14**, 2125 (2005).
[24] D. Sept and J. A. McCammon, Biophys. J. **81**, 667 (2001).
[25] P. Hortschansky *et al.*, Prot. Sci. **14**, 2915 (2005).
[26] U. H. Danielsson, M. Lundgren, and A. J. Niemi, Phys. Rev. E **82**, 021910 (2010).
[27] M. Chernodub, S. Hu, and A. J. Niemi, Phys. Rev. E **82**, 011916 (2010).
[28] B. Svetitsky and L. Yaffe, Nucl. Phys. B **210**, 423 (1982).
[29] R. Fiore, A. Papa, and P. Provero, Phys. Rev. D **67**, 114508 (2003); R. B. Frigori, Nucl. Phys. B **833**, 17 (2010).
[30] F. Zantow, O. Kaczmarek, E. Gava, and R. Jengo, Phys. Lett. B **105**, 285 (1981).
[31] R. Falcone, R. Fiore, M. Gravina, and A. Papa, Nucl. Phys. B **785**, 19 (2007).
[32] V. M. Fokin, N. S. Yuritsyn, and E. D. Zanotto, *Nucleation Theory and Applications* (Wiley-VCH, Weinheim, 2005).
[33] P. A. Hollander *et al.*, Diabetes Care **26**, 784 (2003).
[34] R. D. Murphy *et al.*, Biophys. Chem. **167**, 1 (2012).
[35] W. M. Berhanu and U. H. E. Hansmann, PLoS ONE **9**, e97051 (2014).
[36] N.-V. Buchete, J. E. Straub, and D. Thirumalai, Proteins **70**, 119 (2007).
[37] D. P. Landau and K. Binder, *A Guide to Monte Carlo Simulations in Statistical Physics* (Cambridge University Press, Cambridge, 2005); C. Gattringer and C. B. Lang, *Quantum Chromodynamics on the Lattice* (Springer, Berlin, Heidelberg, 2010).