Part I

Peptide dynamics by molecular dynamics and diffusion theory methods with improved basis sets

Abstract

Improved basis sets for the study of polymer dynamics by means of the diffusion theory, and tests on a melt of cis-1,4-polyisoprene decamers, and a toluene solution of a 71-mer syndiotactic trans-1,2-polypentadiene were presented recently [R. Gaspari and A. Rapallo, J. Chem. Phys. **128**, 244109 (2008)]. The proposed hybrid basis approach (HBA) combined two techniques, the long time sorting procedure and the maximum correlation approximation. The HBA takes advantage of the strength of these two techniques, and its basis sets proved to be very effective and computationally convenient in describing both local and global dynamics in cases of flexible synthetic polymers where the repeating unit is a unique type of monomer. The question then arises if the same efficacy continues when the HBA is applied to polymers of different monomers, variable local stiffness along the chain and with longer persistence length, which have different local and global dynamical properties against the above-mentioned systems. Important examples of this kind of molecular chains are the proteins, so that a fragment of the protein transthyretin is chosen as the system of the present study. This peptide corresponds to a sequence that is structured in β-sheets of the protein, and is located on the surface of the channel with thyroxin. The protein transthyretin forms amyloid fibrils in vivo, whereas the peptide fragment has been shown [C.P. Jaroniec, C.E. MacPhee, N.S. Astrof, C.M. Dobson, and R.G. Griffin, PNAS **99**, 16748 (2002)] to form amyloid fibrils *in vitro* in extended β-sheet conformations. For these reasons the latter is given considerable attention in the literature, and studied also as an isolated fragment in water solution where both experimental and theoretical efforts have indicated the propensity of the system to form β turns or α-helices, but is otherwise predominantly unstructured. Differing from previous computational studies that employed implicit solvent, we performed in this work the classical molecular dynamics simulation on a realistic model solution with the peptide embedded in an explicit water environment, and calculated its dynamic properties both as an outcome of the simulations, and by the diffusion theory in reduced statistical-mechanical approach within HBA on the premise that the mode-coupling approach to the diffusion theory can give both the long-range and local dynamics starting from equilibrium averages which were obtained from detailed atomistic simulations.

1 Introduction

Theoretical and computational methods for studying the synthetic and biological macromolecular systems are the subject of continuous research and development due to the importance of our understanding, predicting and controlling both the final properties of the materials starting from the chemical constitution of the molecules, and the mechanisms of the biological processes associated with proteins and/or nucleic acids that are the basis of life. The fundamental problem that the simulations must face is the need to take into account the wide range of spatial and temporal scales over which the macromolecular systems express their properties. The atomistic molecular or Brownian simulations, taking into account the interactions between the individual atoms in the system, provide very detailed insights, but necessarily limited by time scales that are often not wide enough to fully capture the dynamic properties of the molecules under consideration, since their computational cost grows astronomically with the system's size, especially with increasing the length of the macromolecule and with reducing its flexibility. For these reasons reduced statistical descriptions of the polymer chains is required to simplify the system and to access its dynamics in a more complete way. The mode-coupling diffusion theory (DT) [1, 2] provides such a simplified route to the polymer dynamics. It is based on the solution of the Smoluchowski equation, describing the motion of a polymer chain over the configurational space, and modeling the effects of the surroundings on the chain motion in terms of hydrodynamic frictions. The DT transforms the Smoluchowski equation into a generalized eigenvalue problem, by expanding the eigenfunctions of the Smoluchowski operator in a linear combination of suitable basis functions. The matrix representing this operator contains all the information of the interatomic potentials and can be expressed in terms of equilibrium averages of suitable dynamical variables. The necessary inputs to the theory are then a set of equilibrium averages and the friction coefficients of beads describing the chains. This allows for the representation of the polymer dynamics in terms of time correlation functions (TCFs) of the dynamical variables of interest, obtained by a projection procedure over the eigenvectors of the Smoluchowski operator. Since the exact solution of the Smoluchowski equation requires in principle an infinite set of basis functions to be used, which is not possible to handle in practice, a finite expansion of an approximated solution is searched for, instead. Essential to the successful application of DT, is therefore the choice of a finite basis set containing the most important contributions to the dynamics that are to be described, within the most limited number of basis functions, in order

to keep the computational load within feasibility limits. Mode-coupling approaches are usually adopted, which consist in defining a suitable basis set for a first order approximation, and building more accurate upper order approximations by means of products of the first order elements. As far as rank-one quantities are concerned, the first order basis set coincides with bond vectors and they describe the polymer in a more or less coarse-grained way. In building up upper order basis sets by means of products of the first order functions, the problem of the rapid increase of the number of such products with increasing chain length must be dealt with. Since not all the products that can be formed are useful for the calculations, a limited subset of the most significant ones must be judiciously chosen. Such a choice is performed according to two main approaches: maximum correlation approximation (MCA) [3] and long time sorting procedure (LTSP) [4] methods.

The MCA is based on the idea that only those terms that are maximally correlated with the quantity of interest will give significant contributions to the dynamics. So, if the TCF of a certain bond vector is to be calculated, only products of bond vectors within a given length interval around the bond of interest, are to be included into the upper order basis. Physically meaningful length of such an interval is the persistent length of the chain. In this approach the relevance of the local correlations is emphasized, while long range correlations are sacrificed despite their relevance to the description of the global motions. The LTSP, on the other hand, is based on the idea that a basis rich in slowly decaying terms is more effective in describing global motions, and thus is mainly concerned with slow dynamics. So products of first-order modes are sorted according to the increasing value of their decay rate, and only the pre-chosen number of the slowest ones is retained in the upper order basis set. Here the global description of the chain is put into perspective, but terms that take into account the local correlations are not directly included into the basis, which may have negative impact on the description of the local dynamics [5]. Both MCA and LTSP proved to be effective in describing the polymer dynamics [6, 5], performing essentially the same way for short chains, but better in LTSP for longer ones. Recently, the philosophies underlying these two well-established methods were merged into a new formulation of basis sets and applied to DT. This so-called HBA [7] takes advantage of the strengths of both approaches and obtains higher quality results within less computational effort. Essentially, this new HBA method consisted in constructing a first order basis by the m-th order MCA basis set and using it in the LTSP procedure for higher order calculations. In this way the local correlations contained in the MCA calculation were explicitly included into the LTSP basis set, contrary to the original LTSP method which did not contain them directly.

The HBA was tested on those flexible polymers that are characterized by the repetition of the same monomeric unit along the chain. One system was a melt of cis-1,4-polyisoprene decamers [7],and the other, a toluene solution of a 71-mer syndiotactic trans-1,2-polypentadiene [7]. In both cases, the HBA was shown to fully exploit the specific viewpoints of MCA and LTSP in the formulation of the basis set. Those calculations yielded in fact the same results of LTSP but with a much smaller number of basis functions employed in the numerical computations. Delving further into this basis set showed also better performances from the point of view of the numerical stability, such as the condition numbers of the matrices involved in the calculations, many orders of magnitude smaller than those obtained from LTSP with the same number of basis functions, and the same degree in the products of the variables. Furthermore, calculations that were performed with the same number of HBA and LTSP basis functions, which also involve products of variables of the same degree, showed that the description of the most fundamental eigenstates, which are related to the long time dynamics of the chain, was equivalent in the two cases even if all the HBA terms were much faster decaying than the LTSP counterparts. These studies indicate lucidly that in order to realize the scenario of most global dynamics not only does one have to consider slowly decaying basis functions which is the idea underlying the LTSP method, but also to include the physical correlation between variables which is contained in the MCA terms of the HBA basis set.

All these encouraging results were limited to a certain family of molecular chains, namely, flexible polymers with the same kind of monomer running as repeated units along the polymer. The question then arises if the HBA performs equally well to stiffer polymers of different monomers with variable local stiffness along the chain. These polymers exhibit, of course, different local and global dynamic properties against those cases considered previously[6, 7]. Proteins are important examples of these kinds of polymers and, given their relevance in the life processes, it is of great theoretical interest to assess the extent to which the HBA method can be exploited to improve the description of their dynamics. As in flexible synthetic polymers, both MCA and LTSP have been previously applied to biological systems also, such as the rat microsomal cytochrome b_5 [8], peptide met-enkephalin [5, 9, 10], DNA oligomer [11], protein fragment NK-12 [12], RNA fragment [13], etc. In these studies, their capabilities of capturing the dynamical properties of these complex systems have been evaluated through comparing either experiments and molecular dynamics (MD) simulations [8], or different kind of simulations in implicit or explicit solvent [5, 10].

In this work, we investigate an aqueous solution of a fragment of the protein transthyretin (TTR) by both MD

simulation and DT within the HBA framework. This fragment corresponds to a sequence that is structured in β-sheets of the protein, and is located on the surface of the channel with thyroxin. The protein transthyretin forms amyloid fibrils *in vivo* [14], while the peptide fragment has been shown to form amyloid fibrils *in vitro* in extended β-sheet conformations [15, 16]. For these reasons the latter is given considerable attention in the literature, and studied also as an isolated fragment in water solution where both experimental [15] and theoretical [17] efforts have indicated the propensity of the system to form β turns or α helices, but is otherwise predominantly unstructured. Previous theoretical studies that aim at investigating both the process of fibril formation and the behavior of the isolated peptide in aqueous solution, the structures of TTR in particular, have applied the MD simulations to study peptide molecule in implicit solvent [17, 18]. Here we carry out MD simulations on a more realistic model solution of single peptide molecule in an explicit water solvent, and include in the simulation all the polymer-polymer, polymer-solvent and solvent-solvent interactions by suitable force fields. Given the size of the system, the MD simulation can be performed on an appropriate time window, whose data are recorded for (i) the direct evaluation of both the local and global dynamic properties of the peptide, and (ii) the use as the source of the equilibrium averages and also the calculation in DT of friction coefficients of the beads in terms of which the chain is modeled. The TCFs of selected vectors which were defined along the molecular chain are calculated by means of DT using different HBA basis sets, and the outcomes of these calculations are compared to the TCFs obtained from the MD trajectory. The HBA basis set for this peptide is then evaluated critically by its performances by comparing side by side with the LTSP procedure.

2 methods

2.1 Molecular dynamics simulation of *TTR*(105-115)

The peptide studied in this work is the fragment of the protein transthyretin covering from the amino acid 105 to the 115th one, i.e. TTR(105-115). The amino acidic sequence in this peptide is Tyr-Thr-Ile-Ala-Ala-Leu-Leu-Ser-Pro-Tyr-Ser, to which were added Acetyl and Methylamine groups and the chain is terminated at the first Tyr and last Ser amino acids, respectively (see Fig. 1(a)). The resulting chain molecule contains 181 atoms. The simulation of the model solution of this peptide in explicit water solvent was carried out by means of the software GROMACS 4.0.7 [19], after properly preparing the system. For this purpose the TTR(105-115) peptide was initially arranged in an extended conformation, as shown in Fig. 1(a) and it was swarmed with 5784 water molecules which were contained in a $5.631 \times 5.631 \times 5.631\ nm^3$ cubic cell. Periodic boundary conditions were then applied, and a local minimization of this initial configuration was performed by the steepest descent to remove the contacts among the atoms, due to the possible random injection of the water molecules inside the box. All-atom force fields were applied to both polymer and water molecules [20]. For polymer-polymer and polymer-solvent interactions, the OPLS-AA/L force field was adopted, and the TIP4P potential was used for solvent-solvent interactions. A total of 2 μs long MD simulation was performed in the NPT ensemble with a time step of 1 fs. For the range of interaction, we applied a cutoff of 10.0 for the Lennard-Jones nonbonded interactions, and 11.0 for the Coulomb electrostatics which we evaluated by the Particle-Mesh Ewald method. The temperature of this model solution of peptide was controlled at 298 K by the Nosé-Hoover chain thermostat [21, 22] while its pressure was maintained isotropically at 1 atm. by the Parrinello-Rahman barostat [23, 24, 25]. We stored the trajectory at every 1 ps. Note, however, that the initial 180 ns were necessary for a proper equilibration of the system at the given physical conditions, and these data were discarded from the DT calculation; the remaining 1820 ns of the trajectory were then used to obtain both the equilibrium averages and the friction coefficients which were essential inputs to the DT and the reference TCFs to check against the DT outcomes.

n order to quantitatively evaluate the stiffness of the TTR(105-115), the local persistence lengths of the peptide were calculated from the MD trajectory for all of the inter-residue virtual bonds shown in Fig. 2, except for the first and last ones, which are the peptide's beginning and ending amino acids connected to the terminating Acetyl and Methylamine groups, respectively. The local persistence lengths were defined as the ratio of the average scalar products between each inter-residue versor with the end-to-end vector to the average length of the virtual bond. The average of these local persistence lengths was calculated to be 2.75, with a standard deviation of 0.540. Individual local persistence lengths ranged from 2.12 to 3.83, showing significant variability of the local stiffness along the chain. The comparison of these results with the same quantities of the melt of cis-1,4-polyisoprene [7], confirms that we are indeed studying the dynamic properties of a polymer with variable local stiffness along the chain and a longer persistence length. This average persistent length of TTR can be compared furthermore with that of the melt of cis-1,4-polyisoprene which was calculated to be 2.06 with a standard deviation of 0.091 and the individual local persistence lengths of the latter fall into a much narrower range between 2.01 and 2.26.

Previous studies of peptide dynamics by means of DT [5] showed that a natural and computationally convenient approach to treat the dynamics of rigid groups in the chain, such as $CO-NH$ and aromatic rings, is to collect their atoms in one bead by including only the center of friction of the group in the basis set. Here an even more simplified coarse-grained description of the molecular chain is adopted by giving each amino acid a two-bead representation. The amino acid atoms belonging to the main chain are collected in one bead, and those of the side chain in another one. In this way the specificity of each amino acid is still preserved to some extent, but the molecule is modeled in terms of a minimum number of beads. This two-bead representation is a useful recipe for maintaining the computational load to within feasible limits, especially in cases when bigger peptides or proteins are to be studied. The virtual bonds alternatively joining the centers of the main chain and side chain beads (Fig. 1(b)) are those used to define the basis sets for our DT calculations. The total number of beads and virtual bonds of our model chain are then 23 and 22, respectively. The centers of frictions of beads are calculated as weighted averages of the coordinates of atoms collected into the beads themselves, where the weight for each atom is given by its Stokes radius. The Stokes radii of the atoms can be calculated from the average accessible surface area (ASA) computed with a probe of zero radius by means of the ASA method [26]. To be more explicit, the average surface area exposed to the solvent is calculated along the MD trajectory, and the radius of a sphere having the same area is taken to be the Stokes radius of that atom. Once the Stokes radii of all the atoms defining one bead are known, the following center of the bead is calculated

$$x_{bead} = \frac{\sum_j r_j x_j}{\sum_j r_j} \tag{1}$$

, where x denotes Cartesian coordinates, r_j are the atomic Stokes radii and j runs over all the atoms collected into the bead. The Stokes radii a_i of the beads are obtained by summing the ASAs of the participating atoms and calculating the radius of a sphere with equivalent area. The friction coefficients of the beads necessary for DT are given by the following equation

$$\zeta_i = 6\pi\eta a_i \tag{2}$$

where the experimental viscosity $\eta = 0.001\ Pa \cdot sec$ of water was used for the calculations. Direct estimation of the solvent viscosity was roughly determined from the MD trajectory by calculating the diffusion coefficient and Stokes radius for some water molecules in the solution far from the peptide, and applying the Rouse relation which connects the diffusion coefficient to the viscosity

$$\eta = \frac{k_b T}{6\pi D R_H} \tag{3}$$

. In Eq. 3 k_b is the Boltzmann constant, T is the temperature in Kelvin, R_H is the Stokes radius evaluated from the ASAs of water molecules, and D is the diffusion coefficient given in terms of the center of mass r_{cm} of the water molecule as

$$D = \frac{1}{6} \lim_{t\to\infty} \frac{1}{t} \left\langle |r_{cm}(t) - r_{cm}(0)|^2 \right\rangle \tag{4}$$

.

For the case at hand, we checked that it was appropriate to use the experimental viscosity of water.

Knowing the friction values and Stokes radii is not enough to take into account the hydrodynamic interactions within DT calculations. To do so, another quantity must be given, namely, the parameter $\zeta_r = a/l$ which is the ratio of the average Stokes radii a_i of the beads to the average length of the virtual bonds l_i connecting adjacent beads in the model. This parameter can be set equal to zero for chains in unentangled polymer melts in the free draining regime, since the hydrodynamic interactions are fully screened. This is not so in the case of polymer solutions where ζ_r must be given its actual value. In our case, $\zeta_r = 0.543$, and this indicates that hydrodynamic interactions are important and the dynamics of the chain in solution will be affected. Note, moreover, that the parameter ζ_r appears in the formulation of the diffusion tensor as a commom factor for terms describing hydrodynamic interactions. In Section 2.4 we give such expressions of the diffusion tensor in the Rotne-Prager approximation, [27], which is used in our DT calculations.

The local dynamics of the chain was studied by calculating the TCFs of the bond versors that interconnect the nitrogen atoms of adjacent amino acids. In Fig. 2 six bond vectors in the middle of the chain are shown. These virtual bonds, denoted as $\vec{v_i}$, do not belong to the set of bonds used for forming the basis. The inclusion of these vectors into the basis set is, however, not required by the theory, since their dynamics are in fact obtained by their projection onto the eigenvectors of diffusion operator. For this reason, the TCF of any vector, in principle, can

be obtained once the basis set is defined. The global dynamics of the chain, on the other hand, was studied by calculating the TCF of the head-tail versor. Here, the head-tail virtual bond was defined as the vector connecting the carbon atom in Ace_1 to the nitrogen atom in Nac_{13} of the chain (see Fig. 1(a)).

For DT calculations only one snapshot in every 20 ps was recorded for evaluating the equilibrium averages. Any sampling more than this would often entail processing a great deal of data unnecessarily, because only negligible changes to the description of the dynamics is achieved.

2.2 Formulation of diffusion theory

The DT is a well-established theoretical framework for the study of the dynamic properties of molecular systems and has been formulated and described in detail in the literature [1, 2, 3, 4], and so only a brief summary of it will be given here.

Consider a polymer chain with n beads embedded in a fluid with viscosity η. These beads are traced by their position vectors $r = r_1(t), \ldots, r_n(t)$ and they are connected by m bond vectors. The time-dependent equation of the Smoluchowski differential operator L, which is the adjoint of the Smoluchowski diffusion operator D, is given by:

$$\frac{\partial f(r)}{\partial t} = Lf(r) \tag{5}$$

where $f(r)$ is the dynamical variable, and L is given by:

$$L = \sum_{i,j=1}^{n} \left[\nabla_i \cdot D_{i,j} \cdot \nabla_j - \frac{\nabla_i U}{k_b T} \cdot D_{i,j} \cdot \nabla_j \right] \tag{6}$$

. L contains the intramolecular interactions through the potential energy function U, and it models the fluid-polymer interactions through the diffusion tensor $D_{i,j}$. The latter requires the input of the temperature of the system, the Stokes radii, and the friction coefficients for each bead modeling the polymer chain. The working equations for the $D_{i,j}$ elements in the Rotne-Prager approximation [27, 28] are deferred to Section 2.4.

Following Tang et al. [4] and Kostov and Freed [5], we expand $f(r,0)$ in $f(r,t) = \exp(L,t)f(r,0)$ in a complete set of orthogonal eigenvectors φ_i of L whose eigenvalue equation satisfies:

$$L\varphi_i(r) = -\lambda_i \varphi_i(r) \tag{7}$$

, where λ_i are its real eigenvalues, the TCF with zero average of the dynamical variable f, can be cast in the form:

$$< f(t)f(0) >= \sum_i e^{-\lambda_i t} < f(0)|\varphi_i >< \varphi_i|f(0) > \tag{8}$$

(8)

where the $< \ldots >$ on the left-hand side means an average over different time origins of the dynamic variable f for all ergodic paths taken. In principle, the summation in Eq. 8 must be done for a complete set of φ_i but the calculation is, however, analytically impossible due to the complexity of the molecular potentials U. Approximated solutions must therefore be sought by expanding the functions φ_i in a set of basis functions $\{\chi_1, \chi_2, \ldots, \chi_M\}$ as

$$\varphi_i = \sum_{j=1}^{M} A_{j,i}\chi_j \tag{9}$$

. Applying Eq. 9 to 7, we obtain the matrix eigenvalue equation

$$FC = SC\Lambda \tag{10}$$

in which $F \equiv F_{ij} = - < \chi_i|L|\chi_j >$, $S \equiv S_{i,j} =< \chi_i|\chi_j >$, both of which are $M \times M$ matrices, C the eigenvector matrix of coefficients C_{ij}, and Λ is a diagonal matrix formed with eigenvalues λ_i. The matrix element or dot product of $S_{i,j}$ and $F_{i,j}$ means ensemble equilibrium averages and have to be calculated, for example, $< u|\nu >$, as:

$$< u|\nu >= \int P_{eq}(r)u(r)\nu(r)dr \tag{11}$$

, where $P_{eq}(r)$ is the equilibrium distribution function [6]. Equilibrium averages can be determined by different methods by either from Monte Carlo or MD simulation technique. It is important to note that in this matrix-expansion method the U does not appear explicitly in the expressions of either the TCFs or F and S, Instead U enters the theory only through statistical averaging with $P_{eq}(r)$. The explicit expressions of $F_{i,j}$ are given in Ref. [7] to which the interested readers are referred to for further details. Once the solutions χ_i, and hence φ_i, are known, we may proceed to calculate TCFs. Choosing the dynamical variables to be a set of bond vectors, the TCF of the ith bond versor $\hat{l}_i(t) = \vec{r}_{i-1,j}(t)/|\vec{r}_{i-1,j}(t)|$ corresponds to

$$P_1(t) =< \hat{l}_i(t) \cdot \hat{l}_i(0) >= \sum_j e^{-\lambda_j t} < \hat{l}_i(0)|\varphi_j >< \varphi_j|\hat{l}_i(0) > \tag{12}$$

. The mode-coupling approximation provides a way to generate the basis set $\{\chi_1, \chi_2, \ldots, \chi_M\}$ and, in this approximation, a set of bond vectors is commonly chosen as the first-order basis set for calculating the first-rank quantities. In other words the set of basis functions $\{\chi_1, \chi_2, \ldots, \chi_M\}$ as $\{\vec{l}_1, \vec{l}_2, \ldots, \vec{l}_M\}$ for a polymer chain with M bond vectors. Calculations based on this first order basis set with hydrodynamic interactions neglected has been shown to be equivalent to the generalized Rouse-Zimm description of the chain dynamics [1, 29], and this evidence has been taken to directly assess the effects of those memory functions neglected in the first order description, but the latter are gradually recovered as the basis set is extended to beyond the first order approximation [1, 2, 5]. In next section, the well-established MCA [3] and LTSP [1] procedures for building high-order basis sets in the spirit of the mode-coupling approach are briefly discussed, since their technical details as well as the new HBA idea were fully described in the work of Gaspari and Rapallo [7]. Here, we aim at showing how the HBA can skillfully take benefit of the strengths of these two methods.

2.3 Mode-coupling approximations: MCA, LTSP, and HBA methods

For a polymer chain with m bond vectors, a first-order basis set for calculating first rank quantities is usually chosen to be the set of the m bonds

$$\psi^{(1)} = \{\vec{l}_1, \vec{l}_2, \ldots, \vec{l}_m\} \tag{13}$$

. To extend to second-order approximations, products of trilinear bonds are added to the first-order set:

$$\psi^{(2)} = \{\vec{l}_1, \vec{l}_2, \ldots, \vec{l}_m, \vec{l}_s(\vec{l}_p \cdot \vec{l}_q), \ldots, ..\}, s, p = 1, \ldots m; q = p, \ldots, m \tag{14}$$

and similarly for third-order approximations, products of pentalinear bonds are added to the second-order basis set. Increasing the order of the approximation implies adding higher order of odd products of bonds. In this way the basis set is respectful of the rotational invariance of the diffusion operator and the eigenfunctions transform as vectors as required in calculation of rank-one properties. However the number of products rapidly increases with the order and the chain length, and it is therefore essential to sort out only those products of bonds that give the maximum contribution to the TCF to be calculated. Different strategies were developed towards this goal. The MCA [3] is one tactics that is based on the idea that basis terms strongly correlated with the observed variable gives the maximum contribution to the description of its dynamics. According to MCA, only those products of bond vectors within a pre-defined maximum neighbor, say N, to each bond variable are retained. For instance, in a second-order calculation with $N = 0$, only cubes of each bond, i.e. $s = p = q$, are retained, while with $N = 1$, we added also in the basis set the products of trilinear $\vec{l}_i$s including the nearest neighbors on both sides of each l_i. Another approach to mode-coupling is LTSP [1]. It originates with the idea that long time dynamics is best described using a basis set rich in slowly decaying terms. Operationally, a first-order calculation of LTSP may be performed with the set of bonds of Eq. 13, solving the eigenvalue problem,

$$\tilde{\psi}_i^{(1)} = \sum_{k=1}^{m} C_{k,i}\vec{l}_k \tag{15}$$

for the first-order modes which are linear combinations of bond vectors. The set of modes constitutes the first-order basis set for LTSP. To proceed to the second-order basis set, we add to the first-order one the trilinear products of modes,

$$\psi^{(1),LTSP} = \{\tilde{\psi}_1^{(1)}, \tilde{\psi}_2^{(1)}, \ldots, \tilde{\psi}_m^{(1)}\} \tag{16}$$

,

$$\psi^{(2),LTSP} = \{\tilde{\psi}_1^{(1)}, \tilde{\psi}_2^{(1)}, \ldots, \tilde{\psi}_m^{(1)}, \tilde{\psi}_s^{(1)}(\tilde{\psi}_p^{(1)} \cdot \tilde{\psi}_q^{(1)}), \ldots, ..\}, s, p = 1, \ldots m; q = p, \ldots, m \tag{17}$$

.

In Eq. 17 all the trilinear products of modes that can be formed from the first-order basis are written, but in actual calculation only a pre-defined number, say Q, of them is retained. The choice of the products to retain is imperative, being based on the decay rate of these products which we calculate as the sum of the eigenvalues corresponding to the involved modes. Then, such decay rates in increasing order are sorted out, including the corresponding first Q products in the basis set. In this way the inclusion of the slowest decaying functions is granted, in fulfillment to the LTSP requirement.

As dicated above, the MCA is best suited for describing local dynamics, or, for motions taking place on a small length scale. As a result, it can be less effective in capturing those dynamics involving long-range correlations. In contrast, the LTSP, by introducing slowly decaying functions in the basis set, reproduces slow global motions. Thus, the LTSP is less efficient in describing local dynamics. In a previous paper [7], a hybrid basis approach was introduced based on the philosophies underlying MCA and LTSP; the technique takes advantage of the strengths of both techniques to overcome in a complementary manner their respective weaknesses. In this method the first-order HBA corresponds to a second-order basis set for the MCA calculation with $N = 0$, i.e., the eigenvalue problem is solved for a basis set containing the bonds and the cubes of the bonds. Then, the LTSP selection procedure is applied to these MCA modes to obtain high-order HBA. Though, in principle, any second-order N-neighbor MCA calculation can be taken as a first-order basis set for HBA, the MCA at $N = 0$ was, however, selected on the premise that the cubes of bond vectors give the largest contributions to high-order calculations [6, 3]. For the purpose of discussion below, it is instructive to explicitly show the higher-order approximations for the eigenvectors of both LTSP and HBA, these being the two approaches that will be directly compared in this paper. In the LTSP approach, a second-order approximation to the ith eigenvector reads as

$$\psi_i^{(2),LTSP} = \sum_q C_{q,i}\vec{l}_q + \sum_q \sum_r \sum_s C_{qrs,i}\vec{l}_q(\vec{l}_r \cdot \vec{l}_s) \tag{18}$$

and the second-order HBA is

$$\begin{aligned} \psi_i^{(2),HBA} \quad &= C_{q,i}\vec{l}_q + C_{qqq,i}\vec{l}_q(\vec{l}_q \cdot \vec{l}_q) + C_{qrs,i}\vec{l}_q(\vec{l}_r \cdot \vec{l}_s) \\ &+ C_{qqqrs,i}\vec{l}_q(\vec{l}_q \cdot \vec{l}_q)(\vec{l}_r \cdot \vec{l}_s) + C_{sqqqrrr,i}\vec{l}_s(\vec{l}_q \cdot \vec{l}_q)(\vec{l}_q \cdot \vec{l}_r)(\vec{l}_r \cdot \vec{l}_r) \\ &+ C_{sssqqqrrr,i}\vec{l}_s(\vec{l}_s \cdot \vec{l}_s)(\vec{l}_q \cdot \vec{l}_q)(\vec{l}_q \cdot \vec{l}_r)(\vec{l}_r \cdot \vec{l}_r) \end{aligned} \tag{19}$$

where the subscripts range from 1 to m and repeated indexes imply summation. Equations 18 and 19 manifest that the composition of second-order approximations of LTSP differs substantially from HBA: the LTSP only contains up to trilinear terms of bond vectors, whereas the HBA contains all odd products up to the nonalinear terms, i.e., products of up to nine bonds. So, an equal order of approximation in the two approaches corresponds in their respective basis set to polynomials of different degree, or conversely, polynomials of the same degree correspond to approximations of different order in the two approaches. Nonetheless, the bond products of degree greater than three included in the HBA basis set, are not all the products that can be generated. They actually consist of only those terms supposedly more important in the MCA. From this point of view, the HBA can be viewed as an approach driving the LTSP to generate slowly decaying modes, by picking up in a polynomial subset, among others, all terms relevant to the MCA considered. In the following, we will test the performance of HBA on the peptide *TTR* to see if it is as well promising in describing biological polymer dynamics, as it proved to be in the synthetic ones.

2.4 Diffusion tensor in the Rotne-Prager approximation

In the following explicit formulas for the diffusion tensor in the Rotne Prager approximation are given. Equationss 20 and 21 hold for nonoverlapping spheres, while Eqs. 22 and 23 hold for overlapping spheres. The diffusion tensor is expressed here in terms of the average friction $\zeta = \sum_i \zeta_i/n$, average bond length $l = \sum_i < |\vec{l}_i| > /n$, dimensionless quantity $\zeta_r = \zeta/6\pi\eta l$, mean value $a_{i,j} = (a_i + a_j)/2$ of Stokes radii a_i and a_j, for beads i and j, and interbeads vector $\vec{r}_{i,j}$. The n is the total number of bonds in the chain. These expressions are written so that terms describing hydrodynamic interactions are all multiplied by ζ_r. In such a form, it facilitates turning off the hydrodynamic interactions in systems like polymer melts in the Rouse regime where it is completely screened; one simply treats ζ_r as an adjustable parameter and sets it equal to 0. In cases of polymer solutions, it must be set to its acutal value, instead. Diffusion tensor is given here in its Cartesian components ν and ν'. When $r_{i,j} > a_i + a_j$ and $\nu = \nu'$,

$$
\begin{aligned}
D_{i,j}^{\nu\nu'} &= \frac{k_bT}{\zeta}\Bigg\{\frac{\zeta}{\zeta_i}\delta_{i,j} + \frac{3}{4}\zeta_r(1-\delta_{i,j}) \\
&\quad \frac{l}{r_{i,j}^3}\left[r_{i,j}^2 + (r_{i,j}^{\nu})^2 + 2a_{i,j}^2\left(\frac{1}{3} - \left(\frac{r_{i,j}^{\nu}}{r_{i,j}}\right)^2\right)\right]\Bigg\}
\end{aligned}
\tag{20}
$$

, when $r_{i,j} > a_i + a_j$ and $\nu \neq \nu'$,

$$
D_{i,j}^{\nu\nu'} = \frac{k_bT}{\zeta}\left\{\frac{3}{4}\zeta_r(1-\delta_{i,j})l\frac{r_{i,j}^{\nu}r_{i,j}^{\nu'}}{r_{i,j}^3}\left(1 - \frac{2a_{i,j}^2}{r_{i,j}^2}\right)\right\} \tag{21}
$$

. On the other hand, when $r_{i,j} \leq a_i + a_j$ and $\nu = \nu'$,

$$
\begin{aligned}
D_{i,j}^{\nu\nu'} &= \frac{k_bT}{\zeta}\Bigg\{\frac{\zeta}{\zeta_i}\delta_{i,j} + \frac{\zeta_r l}{a_{i,j}}(1-\delta_{i,j}) \\
&\quad \left[1 - \frac{9r_{i,j}}{32a_{i,j}} + \frac{3}{32a_{i,j}r_{i,j}}(r_{i,j}^{\nu})^2\right]\Bigg\}
\end{aligned}
\tag{22}
$$

and when $r_{i,j} \leq a_i + a_j$ and $\nu \neq \nu'$,

$$
D_{i,j}^{\nu\nu'} = \frac{k_bT}{\zeta}\left\{\frac{\zeta_r l}{a_{ij}}(1-\delta_{i,j})\frac{3}{32a_{i,j}r_{i,j}}r_{i,j}^{\nu}r_{i,j}^{\nu'}\right\} \tag{23}
$$

.

3 Results and discussion

The dynamics of TTR peptide is studied by both HBA and LTSP in order to assess the extend to which the former approach could be applied to biological polymers. Both local and global dynamics are attended to by calculating the TCFs of inter-residue virtual bonds and head-tail unit vector. The evaluation of HBA proceeds in parallel with LTSP which was found to behave essentially the same as MCA for medium size polymers, and works better for longer and stiffer chains. The spectrum of the L operator and the correlation times (CTs) of inter-residue virtual bond vectors along the chain are both considered for such evaluation purposes. The eigenvectors of L are delocalized quantities along the chain, each being a function of all the bonds included in the basis set, while the CTs of inter-residue bond vectors are related to motions on a local length scale. The generalized eigenvalue problem from which the approximated eigenfunctions of the L operator are obtained gives a variational solution to the diffusion equation so that converged eigenvalues are the lower bound to those found with a finite basis set. For this reason, when comparing two different basis sets, the one giving the lowest eigenvalues should be the best in describing the eigenstates of L.

In Fig. 3(a)-3(c) we compare the eigenvalues of L obtained by the second-order HBA and LTSP calculations for different number of basis functions (500, 1000, and 3000, respectively) used in the calculations. In all cases, the lowest eigenvalues relevant to the polymer dynamics obtained by HBA are all lower than those obtained by LTSP, and the advantage of using HBA over LTSP becomes more evident with increasing number of functions involved. Also, the CTs of the inter-residue versors, which are representatives of the local dynamics of the chain, are seen to be better described by HBA than that by LTSP (Fig. 3(d)-3(f)), where the latter always deviates from the converged result calculated by the third-order HBA with 6000 basis functions. The profile of the CTs of the inter-residue versors differs characteristically from the regular bell-shape curve with a maximum in the center of the chain. This somewhat peculiar profile of the CTs of inter-residue versors against the symmetrical profile of the CTs of polymers that show the same repeating unit [7, 6]reveals different local stiffness which is proper to peptides and proteins, and is a result of the distributions along the chain of different amino acids.

In Fig. 4 , we compare the second-order approximation results of HBA and LTSP by displaying the difference between the CT of the versor $\hat{v}_4$ (see also Fig. 2) obtained from a calculation with a given number of basis functions and that of the convergence results. Note that the number of basis functions used in the calculations is divided by the total number of functions that can be constructed in a second-order calculation. The versor $\hat{v}_4$ in the plot is the slowest relaxing one along the chain, whose dynamics can be described at its very best by LTSP which is mainly concerned with slow dynamics. Both HBA and LTSP converge to the most accurate results, but the HBA is apparently nearer to convergence. We attribute this outcome to the different compositions of the two types of

basis sets, involving up to nonalinear terms for HBA and only up to trilinear products of the dynamical variables for LTSP. The effects of different compositions thus have much bearing on the quality of the approximation of the polymer dynamics, but they are of small impact on the numerical accuracy of the calculations.

Figure 5 depicts the condition number (CN) [7] of the S and F matrices idrawn with increasing number of basis functions for different calculations within the second-order approximations of HBA and LTSP. The CN of a positive definite matrix is the ratio of its maximum eigenvalue to its minimum eigenvalue, and matrix with this property rules the maximum accuracy of the results that can be achieved when the finite precision arithmetic is employed in numerical calculations. In particular, the number of correct digits in the solution of the eigenvalue problem is roughly evaluated by the difference between the number of digits representing the floating point numbers (usually 16 in most common hardware), and the decimal logarithm of the maximum CN. When these latter two quantities become comparable, practically every accuracy is lost due to numerical ill conditioning. Accordingly, a basis set will work well from a numerical point of view if it yields matrices with small CNs. As illustrated in Fig. 5we obtain essentially the same CNs by using either LTSP or HBA at second order, with a small advantage of about one order of magnitude for LTSP in both S and F. The increase of CNs as a function of the matrices dimensions is quite small, thus ensuring the safe use of both approximations in calculations of the dynamics at second order. Accordingly the use of HBA is more convenient for it yields a better level of approximation. In order to check further the performances of the two methods when the included powers of the bond vectors are of the same degree in HBA and LTSP, it is necessary to compare a second-order HBA calculation with a fifth-order LTSP. In doing so, all possible powers of bonds up to nonalinear terms have to be included in the basis set of LTSP, whereas in HBA only a subset of those powers which is considered to be important by MCA, are retained. Thus, a thorough comparison of the HBA and LTSP results would show how relevant is the effect of the neglected powers in HBA. The details of these comparisons are shown in Figs. 6(a)-6(c) for what concerns the eigenvalues, and in Fig. 6(d)-6(f) for the CTs of the inter-residue versors. It can be seen from Fig. 6(a) and 6(b)that, when ranging from a small to an intermediate number of basis functions such as 500 and 1000, the smallest eigenvalues most relevant to the dynamics are evaluated to essentially the same degree of accuracy in the two cases, and the same CTs are essentially obtained, as can be readily confirmed (Fig. 6(d) and 6(e)). In these cases, the high-order multilinear terms which are selected by MCA and included in HBA recover a relevant amount of contribution to the dynamics, and the quality of the polymer dynamics description is as good as that in the context of LTSP in which are included all possible products of the bond vectors up to nonalinear terms. In a way the LTSP basis set will be penalized for its inclusion of just a small number of functions in the calculations, since many higher-order contributions of it, supposedly generated, are neglected. When the size of the basis is increased up to 3000 functions, the effects of the products of bonds generated by LTSP and not contained in HBA becomes evident. Now, the fifth-order LTSP outperforms the second-order HBA (Fig. 6(c)) where all the LTSP eigenvalues are seen to be smaller than the HBA counterparts. Note moreover that some of the CTs calculated by the LTSP are definitely nearer to convergence than those obtained by the HBA basis set, as clearly shown in Fig. 6(f). Accordingly, the calculations with the fifth-order LTSP cannot be pushed to the point of a convergence result, since the generalized eigenvalue problem becomes more and more ill conditioned as the number of functions involved in the calculations increases and numerical consistency is rapidly lost. Let us delve further into this aspect.

In Fig. 7 the decimal logarithm of the CNs of matrices S and F are depicted with different number of basis functions for the second-order calculation of HBA and fifth-order for LTSP. In all of the cases, the second-order HBA yields much better conditioned matrices than those of LTSP, whose CNs already fall into a critical range near to 10^{12} when 3000 functions are employed in the calculations. On the contrary, CNs obtained by the second-order HBA are still in a safe range even for the highest number of basis functions. This is an encouraging espial for it indicates that it is possible to carry out calculations with larger basis sets, thus pushing the HBA very close to the convergence results with numerical certainty. In the case of proteins and peptides the need to use high-order DT calculations to reach convergence in the description of the dynamics is unswervingly stronger than for flexible, synthetic polymers with shorter persistence lengths, for which the contributions of the memory functions are more easily captured. This conclusion has been drawn in studies of the protein met-enkephalin [5] where products of up to 13 modes were employed in a seventh-order LTSP procedure. We have tested different calculations on TTR(105-115) and found that a third-order HBA basis set with 6000 functions was sufficiently accurate and reliable for CTs on the inter-residue versors to reach convergence. The maximum degree of bond products that have been included into such a basis is as high as 15. In these calculations, the maximum decimal logarithm of CNs for the matrices S and F was only 8.67, a value well below the critical range. As a concrete comparison, one requires to include products of bonds of up to the same 15th degree in the LTSP basis set for the solution of an eight-order problem. This is not only much more computationally demanding, but also not solvable due to numerical ill conditioning. These kinds of problems are not encountered in previous calculations of met-enkephalin [5], since this is a smaller peptide than

TTR(105-115), and convergence can be reached before running into numerical uncertainties. The quality of the description of the peptide dynamics is assessed further by comparing the TCFs of selected bond versors obtained directly from MD simulation with the outcomes of DT results calculated with the reference third-order HBA that includes 6000 basis functions.

In Fig. 8 the P_1 TCFs of the end-to-end bond unit vector calculated from the data of MD trajectory and from DT are quantitatively compared to evaluate the HBA capability of capturing the global dynamics of the peptide. The agreement of the two curves is manifestly good. At short times, the two results are indistinguishable because the period at which the TCF calculated from MD is more reliable, while in the tail regime where the MD curve is statistically more uncertain, the DT still shows a quite good"fitting" of the MD result. A more thorough evaluation of HBA can be effected by examining the dynamics on a more local scale. To this end, we compare in Fig. 9 the dynamics on a more local scale are concerned by comparing the MD and DT TCFs for the six bond unit vectors shown in Fig. 2 the P_1 obtained from MD simulation and that from DT for the six bond unit vectors depicted in Fig. 2. It is enlightening to see that the agreement is favorably good for the TCFs of $\hat{v}_1$, $\hat{v}_2$, and $\hat{v}_3$, and optimal for $\hat{v}_4$, $\hat{v}_5$, and $\hat{v}_6$. Since P_1 calculated from DT are already at convergence, the small discrepancies seen in Fig. 8 and 9 between the $P_1$from MD simulations and that calculated from DT must be considered as definitive, and may be ascribed to different reasons, such as an incomplete conformational sampling despite the pretty long time window of the MD simulation which would affect both the TCFs in MD simulations and the equilibrium averages that enter the theory, the approximations inherent in the DT simplification of the system, etc. Nevertheless, so far as all of the results presently obtained in this work are concerned, the HBA is indeed a convenient and efficient means to build up basis sets for studying the dynamics of peptides and proteins, for it benefits by its inclusion of the local correlations through the MCA terms in addition to allowing the LTSP to build slowly decaying terms of better quality. Perhaps worthwhile mentioning here is a previous attempt to combine the MCA and LTSP into a pragmatic theoretical framework for studying protein dynamics [5]. In this approach, the upper order of basis sets for rank-one properties were formed by including into the basis set of Eq. 17, through the original LTSP prescription, also odd products of bonds of the same type described in Eq. 14 in order to enrich the basis with terms capable of capturing the fastest dynamics. In this way, the basis sets constructed was found to produce a minor improvement to the dynamics probably due to poor coupling between terms acting on so different space scales. In contrast, in HBA, the LTSP procedure builds up slowly decaying product of modes that are modified to enhance their capability of describing the local dynamics through the MCA terms. So this latter information cannot be lost, and it gives in fact favorable contribution to the results. In the previously introduced procedure [7], and now applied to a peptide in an aqueous solution, only the second-order MCA terms which correspond to $N = 0$ neighbor were included in the basis set for HBA. In both the previous and present cases, we found a substantial improvement in the description of the polymer dynamics on both local and global scale over the well-established MCA and LTSP techniques. Since any MCA second-order basis set can be used as a first-order basis for the HBA procedure, choices other than $N = 0$ neighbor are also possible. If, for instance, $N = 1$ is chosen, one may then include trilinear products between one bond with its nearest neighbors into the first-order HBA basis set, and the latter, in principle, could be beneficial for cases of stiff polymers with long persistence length as in peptides and proteins. The present work indicates, however, that the $N = 0$ is enough for a straightforward and computationally economic application to TTR(105-115), and more complex calculations with $N > 0$ appear unnecessary.

4 Conclusions

In this work, we have revisited the HBA technique [7] developed previously to generate improved basis sets for studying the polymer dynamics by means of the DT. This previous study left opened the question as to how well the HBA approach performs for systems displaying long persistence length and small gaps between their slowest diffusive modes. Studying these characteristics is important since they are very far from those typical molecular systems studied there, but, nonetheless, they are proper to peptides and proteins. In this work, we extended this recently developed HBA [7] and applied the technique to investigate the dynamics of a fragment of the protein transthyretin in an aqueous solution, the TTR(105-115). With devoting to a complete and quantitative study in mind, we carried out a 2 μs long classical MD simulation using a realistic model solution of the peptide in explicit water environment. The dynamic properties of our system were calculated both by an outcome of the MD simulations, and by the DT with HBA. As described above, the HBA combines MCA and LTSP taking advantage of the physical hypothesis underlying the two different procedures; the former is based on the local correlation along the chain, and the latter emphasizes contributions from the long time dynamics. Explicitly, the HBA uses a first-order basis set that corresponds to the high-order MCA, and generates its high-order approximations through the LTSP method. The sets of slowly decaying terms that are generated according to LTSP technique are accordingly

enriched in their capability of capturing local correlations along the chain in so far as the MCA prescriptions are concerned. The HBA is thus a compendium of these two well-established methods.

On looking more closely at the MCA and LTSP, they have been shown in the literature [5, 6] to be essentially equivalent for short chains, but for longer ones, the LTSP outperforms MCA. The advantages of HBA should therefore be evaluated against the LTSP procedure and it was done here for the TTR(105-115) system. The adequacy of the basis set was assessed by exploiting the variational principle that was applied to calculate the eigenvalues of the L operator. We found that when both HBA and LTSP are kept to the second order of approximation, the former always outperforms the latter both in the description of the eigenstates of L and in the dynamics of the inter-residue bond unit vectors. The different constituents of the two second-order basis sets, containing in HBA products of bonds up to ninth degree and only third degree in LTSP, have decisive consequences in their performance. To demonstrate this, a comparison has in fact been made between a second-order HBA and a fifth-order LTSP, so that in both basis sets are included up to nonalinear products of bonds. The results of calculations showed that the two basis sets yield essentially the same accuracy in the description of the polymer dynamics if a small to intermediate number of basis functions are used to solve the generalized eigenvalue problem. Since in LTSP all possible products of bonds are included in the basis set, whereas in HBA only a subset of them which is selected by MCA is included, this indicates that the HBA basis set really contains the most relevant high degree products for describing the dynamics. The effect of the neglected products of bonds in HBA becomes evident when the number of basis function increases and this is seen in the fifth-order LTSP where the latter begins outperforming the second-order HBA. Nevertheless, one should not lose sight of the numerical limitation in using the fifth-order LTSP, since, up to this order, its CNs of the S and F matrices rapidly reach critical values, thus counterbalancing the numerical consistency of the results. On the contrary, the second-order HBA can be safely extended to include a high number of basis functions and still gives an accurate description of the dynamics without such numerical dilemma, because the CNs of the matrices in this case is far from critical values. To achieve a convergence result on the local and global chain dynamics, we require a third-order HBA with 6000 basis functions in DT calculations. This basis set corresponds to bond products extending up to the 15th degree. Such high degree polynomials were found necessary also for the proper description of the dynamics of met-enkephalin that used the LTSP [5] within DT, and was confirmed here for TTR(105-115) as well. This finding of the convergence result is a clear signature of the complexity of peptide dynamics and physically it sheds light on the importance of accurately describing the contributions from the memory functions. In this respect, the HBA approach was found again to be a significant improvement for the study of the polymer dynamics over the two widely used MCA and LTSP methods. The method works well both for flexible, synthetic polymers with a same repeating unit along the chain as well as for stiffer chains with different monomeric units and longer persistence length. This would mean that the HBA is realistic and practical, because it contains information about local correlation in a proper manner and is of general validity, and furthermore, it allows for the construction of slowly decaying terms of better quality for describing the long time dynamics. In principle, more information about the local correlation could be introduced by using a second-order MCA basis set with $N > 0$ as in the first-order HBA basis set, instead of including only terms coming from the $N = 0$ neighbor approximation. Doing this could be beneficial for cases of stiff polymers with long persistence length, but the present concrete results for TTR(105-115) suggest that more complex calculations with $N > 0$ are unnecessary, at least for peptides comparable in size and complexity with TTR(105-115). The extension of HBA to examine more closely the effects of these additional terms of local correlation is an interesting endeavor, and any application of HBA to systems of such complexity may be a challenging theme in future studies.

References

[1] X Chang and K F Freed. Test of Theory for Long Time Dynamics of Floppy Molecules in Solution Using Brownian Dynamics Simulation of Octane. *J. Chem. Phys.*, 99(10):8016–8030, 1993.

[2] Angelo Perico, Roberto Pratolongo, Karl F. Freed, Richard W. Pastor, and Attila Szabo. Positional time correlation function for one-dimensional systems with barrier crossing: Memory function corrections to the optimized Rouse-Zimm approximation. *The Journal of Chemical Physics*, 98(1):564, 1993.

[3] Angelo Perico and Roberto Pratolongo. Maximum-Correlation Mode-Coupling Approach to the Smoluchowski Dynamics of Polymers. *Macromolecules*, 30(19):5958–5969, 1997.

[4] W H Tang, X Chang, and K F Freed. Theory for Long Time Polymer and Protein dynamics: Basis Functions and Time Correlation Functions. *J. Chem. Phys.*, 103(21):9492–9501, 1995.

[5] K S Kostov and K F Freed. Long-time dynamics of Met-enkephalin: comparison of theory with Brownian dynamics simulations. *Biophysical journal*, 76(1 Pt 1):149–163, 1999.

[6] G La Penna, Paola Carbone, Rita Carpentiero, Arnaldo Rapallo, and Angelo Perico. Polyisoprene local dynamics in solution: Comparison between molecular dynamics simulations and high order diffusion theory. *The Journal of Chemical Physics*, 114(4):1876, 2001.

[7] Roberto Gaspari and Arnaldo Rapallo. Formulation of improved basis sets for the study of polymer dynamics through diffusion theory methods. *The Journal of chemical physics*, 128(24):244109, 2008.

[8] Andrea Giachetti, Giovanni La La Penna, Angelo Perico, and Lucia Banci. Modeling the backbone dynamics of reduced and oxidized solvated rat microsomal cytochrome b5. *Biophysical journal*, 87(1):498–512, 2004.

[9] Min-yi Shen My and Karl F Freed. Long time dynamics of Met-enkephalin: comparison of explicit and implicit solvent models. *Biophysical journal*, 82(4):1791–1808, 2002.

[10] Min-yi Shen and Karl F. Freed. Long time dynamics of Met-enkephalin: Tests of mode-coupling theory and implicit solvent models. *The Journal of Chemical Physics*, 118(11):5143, 2003.

[11] Simone Fausti, Giovanni La Penna, Carla Cuniberti, and Angelo Perico. Mode-Coupling Smoluchowski Dynamics of a Double-Stranded DNA Oligomer. *Biopolymers*, 50(6):613–629, 1999.

[12] G La Penna, S Fausti, A Perico, and J A Ferretti. Smoluchowski dynamics of the vnd/NK-2 homeodomain from Drosophila melanogaster: second-order maximum correlation approximation. *Biopolymers*, 54(2):89–103, August 2000.

[13] S Fausti, G La Penna, J Paoletti, D Genest, G Lancelot, and A Perico. Modeling the dynamics of a mutated stem-loop in the SL1 domain of HIV-1Lai genomic RNA by 1H-NOESY spectra. *Journal of biomolecular NMR*, 20(4):333–349, 2001.

[14] P Westermark, K Sletten, B Johansson, and G G Cornwell. Fibril in senile systemic amyloidosis is derived from normal transthyretin. *Proceedings of the National Academy of Sciences of the United States of America*, 87(7):2843–2845, 1990.

[15] J A Jarvis, A Kirkpatrick, and D J Craik. 1H NMR analysis of fibril-forming peptide fragments of transthyretin. *International journal of peptide and protein research*, 44(4):388–98, October 1994.

[16] Christopher P Jaroniec, Cait E MacPhee, Nathan S Astrof, Christopher M Dobson, and Robert G Griffin. Molecular conformation of a peptide fragment of transthyretin in an amyloid fibril. *Proceedings of the National Academy of Sciences of the United States of America*, 99(26):16748–16753, 2002.

[17] Emanuele Paci, Jörg Gsponer, Xavier Salvatella, and Michele Vendruscolo. Molecular dynamics studies of the process of amyloid aggregation of peptide fragments of transthyretin. *Journal of Molecular Biology*, 340(3):555–569, 2004.

[18] Da-Wei Li, Li Han, and Shuanghong Huo. Structural and pathway complexity of beta-strand reorganization within aggregates of human transthyretin(105-115) peptide. *The journal of physical chemistry. B*, 111(19):5425–5433, 2007.

[19] HJC Berendsen, D. van der Spoel, and R. van Drunen. GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications*, 91(1-3):43–56, 1995.

[20] David Van Der Spoel and Erik Lindahl. Brute-Force Molecular Dynamics Simulations of Villin Headpiece: Comparison with NMR Parameters. *The Journal of Physical Chemistry B*, 107(40):11178–11187, 2003.

[21] Shuichi Nosé. A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics*, 52(2):255–268, 1984.

[22] William G Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A*, 31(3):1695–1697, 1985.

[23] M. Parrinello and A. Rahman. Crystal Structure and Pair Potentials: A Molecular-Dynamics Study. *Physical Review Letters*, 45(14):1196–1199, 1980.

[24] M Parrinello and A Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*, 52(12):7182–7190, 1981.

[25] Shuichi Nosé and M L Klein. Constant pressure molecular dynamics for molecular systems. *Molecular Physics*, 50(5):1055–1076, 1983.

[26] R W Pastor and M Karplus. Parametrizatlon of the Friction Constant for Stochastic Simulations of Polymers. *J. Phys. Chem.*, 92:2636–2641, 1988.

[27] Jens Rotne and Stephen Prager. Variational treatment of hydrodynamic interaction in polymers. *The Journal of Chemical Physics*, 50(11):4831–4837, 1969.

[28] Hiromi Yamakawa. Transport Properties of Polymer Chains in Dilute Solution: Hydrodynamic Interaction. *J. Chem. Phys.*, 53(1):436, 1970.

[29] Robert Zwanzig. Theoretical basis for the Rouse-Zimm model in polymer solution dynamics. *The Journal of Chemical Physics*, 60(7):2717, 1974.

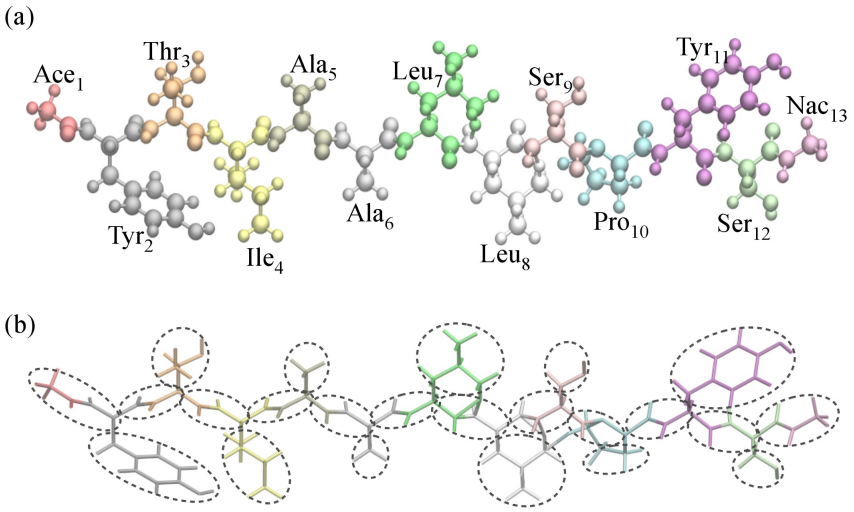

Figure 1: The CPK representation of transthyretin fragment $TTR(105\text{-}115)$ labeling by (a) the consisting residues from Ace_1 to Nac_{13} and (b) 23 beads presented in larger transparent CPK form.

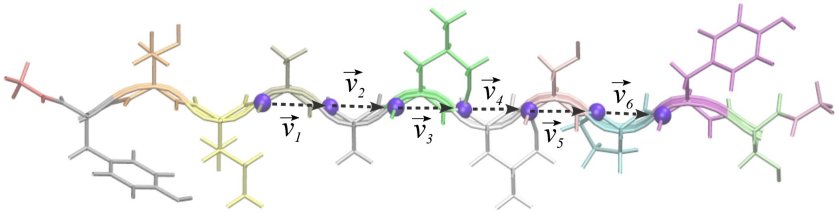

Figure 2: The seven nitrogen atoms of the backbone defining six bond vectors from $\vec{v_1}$ to $\vec{v_6}$ for the TCF calculations.

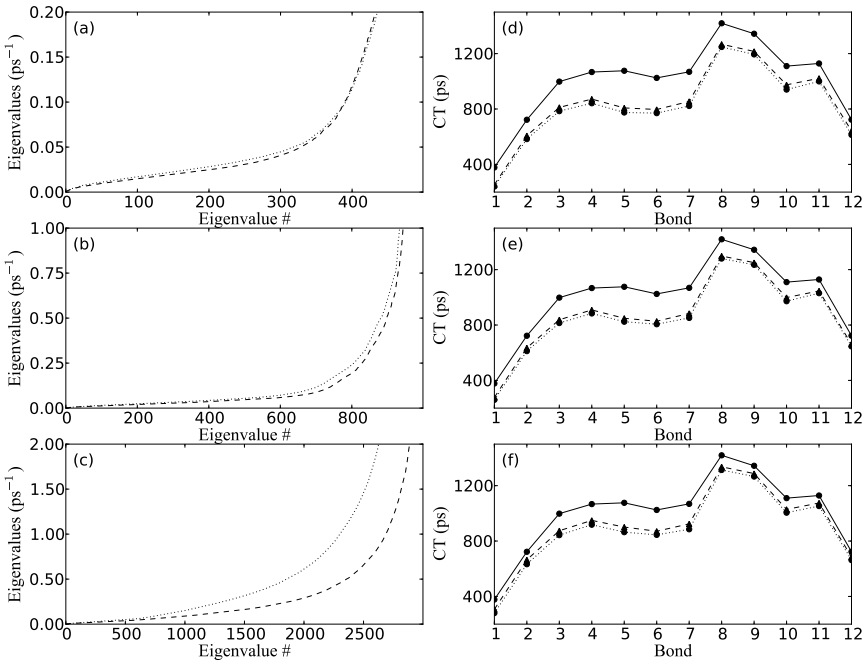

Figure 3: Comparison of the eigenvalues of L and of CTs for the inter-residue unit vectors. Results of a second-order HBA (dashed) and second-order LTSP (dotted) calculations with 500 [(a) and (d)], 1000 [(b) and (e)], and 3000 [(c) and (f)] basis functions are shown. The solid line in (d)-(f) represents reference calculations performed with a third-order with 6000 functions HBA basis set.

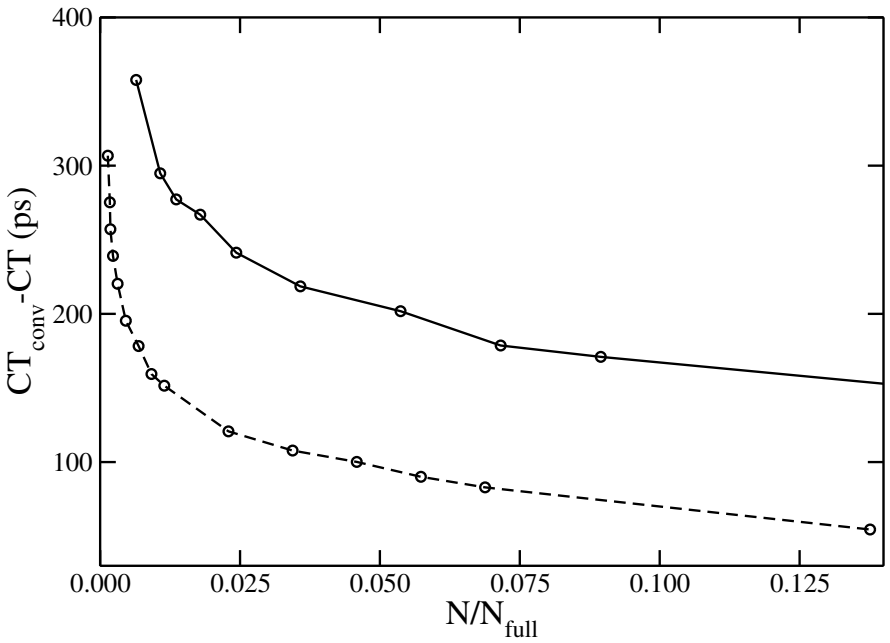

Figure 4: Convergence properties for the CT values of the $\vec{v}_4$ bond vector calculated in the second-order LTSP (solid line) and HBA (dashed line) approximations. Each point represents the difference of the CTs between the value obtained with the full second approximation (N_{full}) and that with a smaller basis set (N).

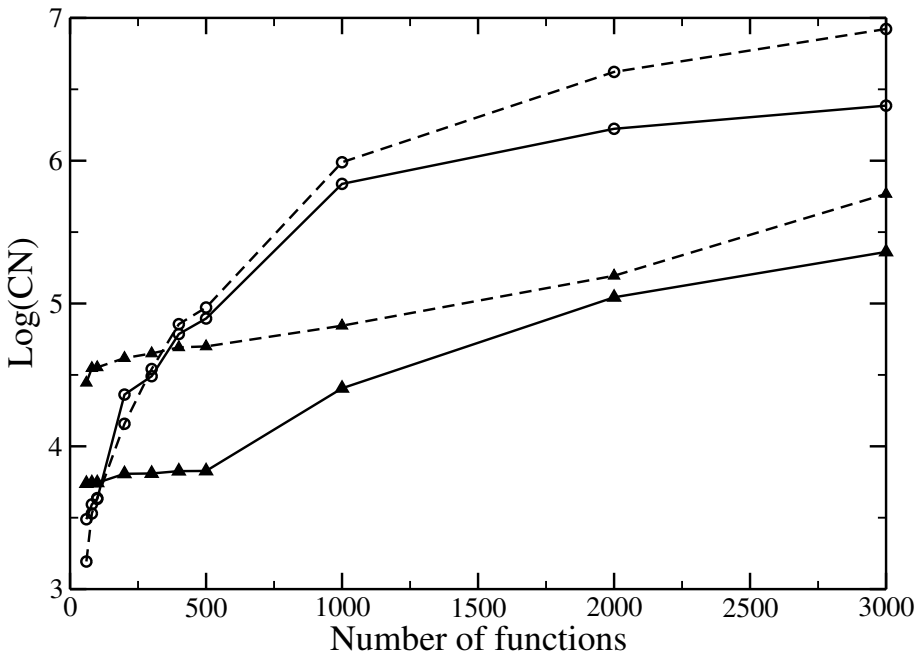

Figure 5: Decimal logarithm of the CN of the S (circle) and F (triangle) matrices in the second-order HBA (dashed) and LTSP (solid) approximation computed at different sizes of the basis sets.

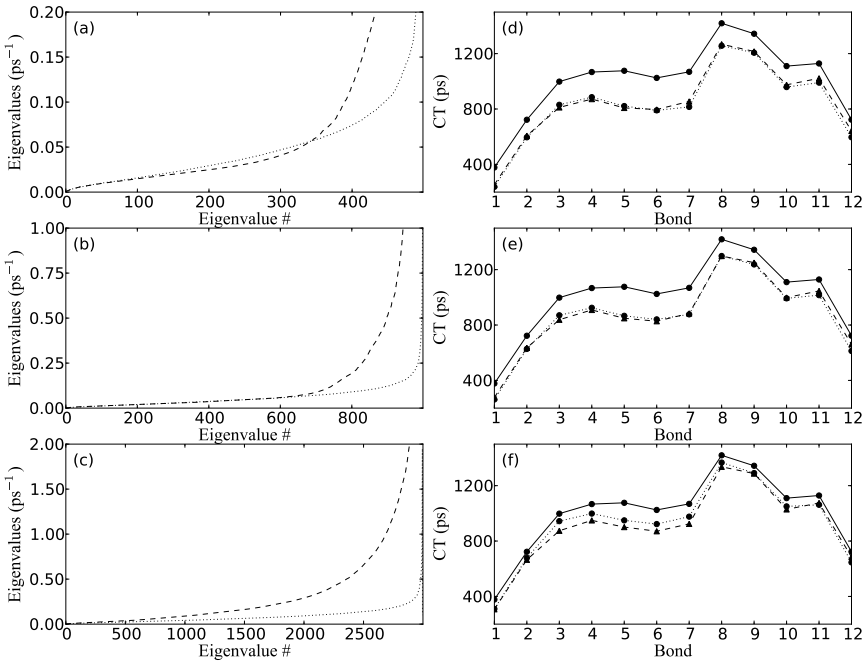

Figure 6: Comparison of the eigenvalues of L and of CTs for the inter-residues unit vectors. Results of second-order HBA (dashed) and fifth order LTSP (dotted) calculation with basis functions 500 [(a) and (d)], 1000 [(b) and (e)], and 3000 [(c) and (f)] are shown. The solid line in (d)-(f) represents a reference calculation performed within the third-order HBA with a basis of 6000 functions.

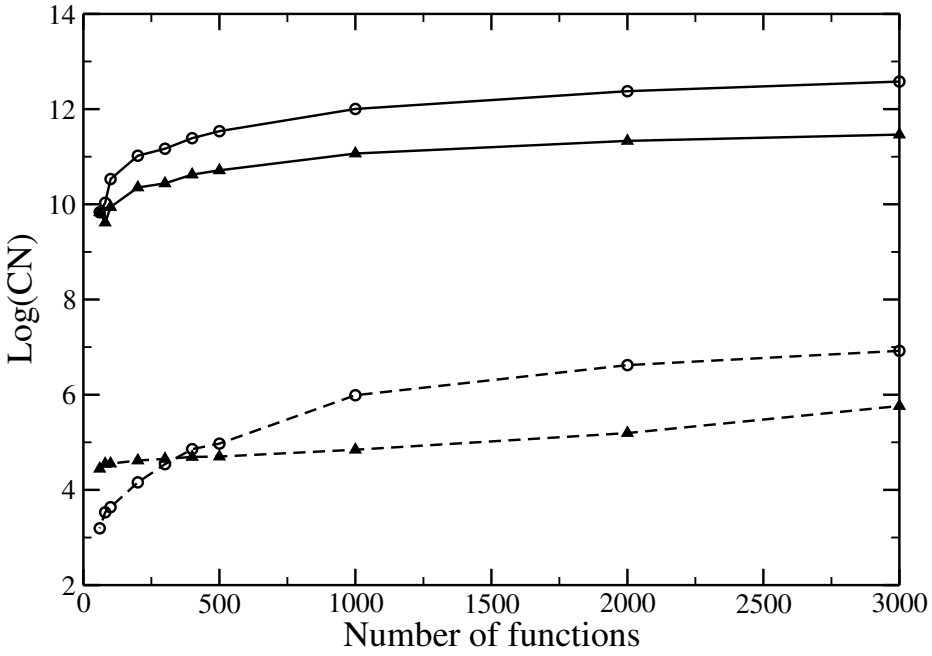

Figure 7: Decimal logarithm of the CN of the S (circle) and F (triangle) matrices in the second order HBA (dashed) and fifth order LTSP (solid) approximations computed at different sizes of the basis sets.

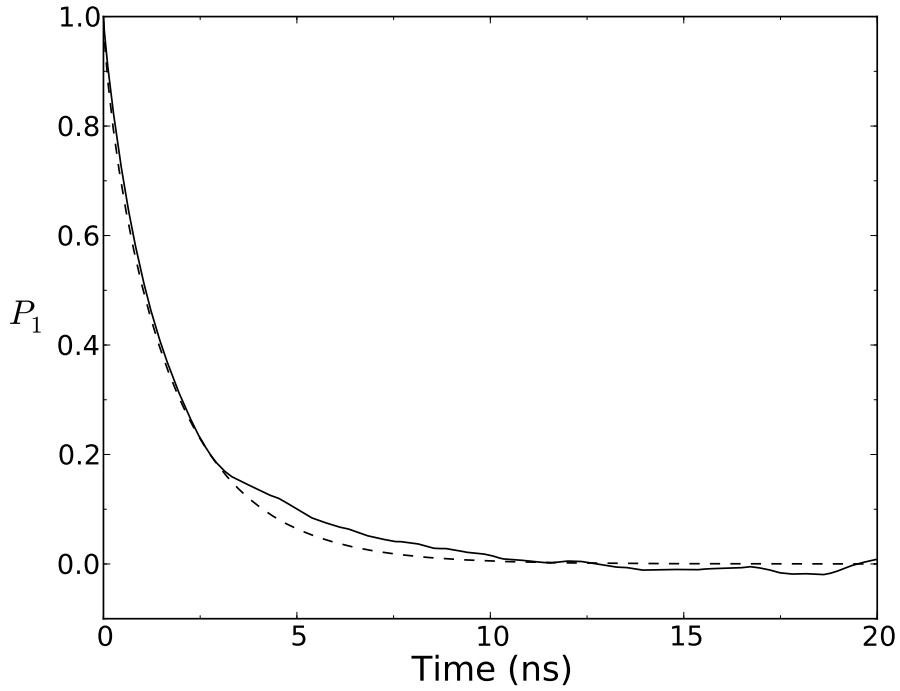

Figure 8: The end-to-end bond vector time correlation function P_1. Results from MD simulation is denoted by solid curve and the third-order HBA calculated with 6000 functions is depicted as dashed line.

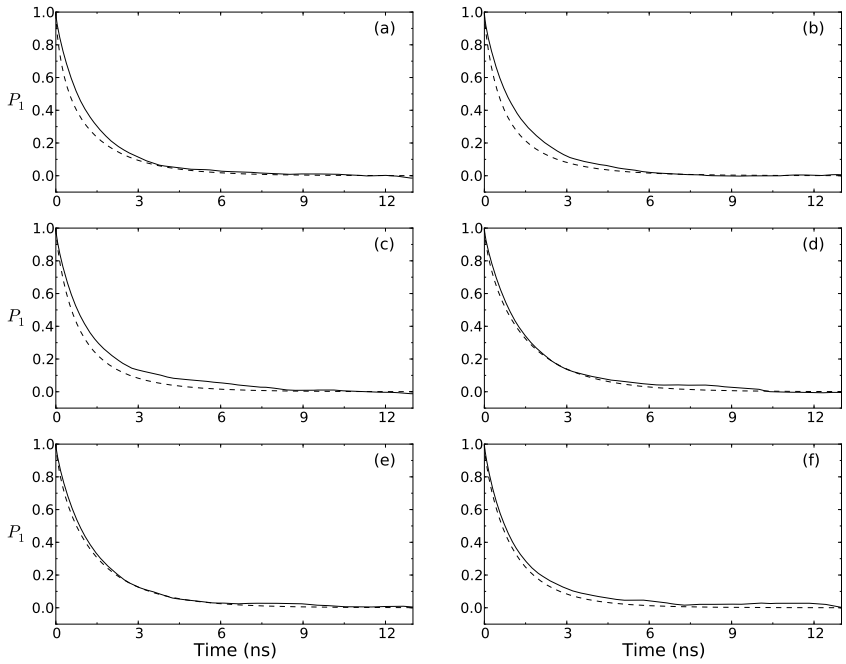

Figure 9: The time correlation functions P_1 of the residue-to-residue vectors (a) $\vec{v}_1$, (b) $\vec{v}_2$, (c) $\vec{v}_3$, (d) $\vec{v}_4$, (e) $\vec{v}_5$, and (f) $\vec{v}_6$. Results from MD simulation are denoted by solid curves and the third-order HBA results calculated with 6000 functions are depicted as dashed lines.

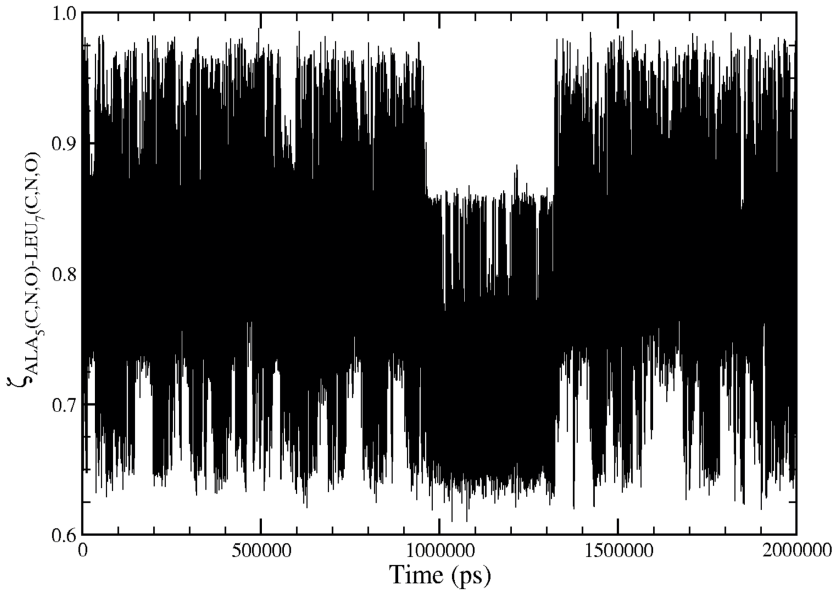

Figure 10: The shape similarity function of $Ala_5(C,N,O)$ and $Leu_7(C,N,O)$ wrt the initial conformation.

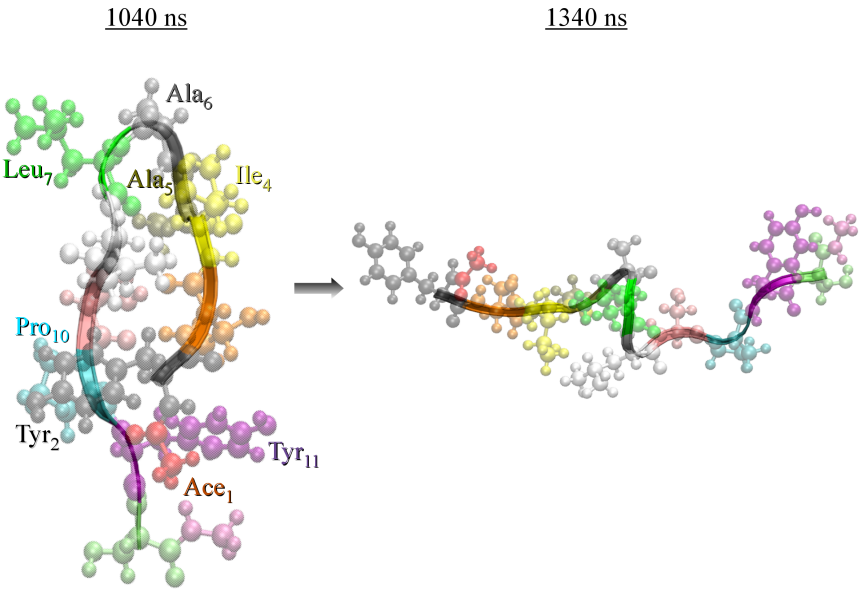

Figure 11: Snapshots of 1040 ns (left) and 1340 ns (right) extracted from the trajectory.