physics

MDPI

Article

Theory of Liquids for Studying the Conformational Flexibility of Biomolecules with Reference Interaction Site Model Approximation

Alexey Danilkovich [1,2] and Dmitry Tikhonov [3,4,*]

[1] State Center for Science of the Branch of the Shemyakin–Ovchinnikov Institute of Bioorganic Chemistry in Pushchino, Russian Academy of Sciences, 142290 Pushchino, Russia; danilkovich@bibch.ru
[2] BioMedPharm Technology Department, the Branch of the Russian Biotechnological University (RosBioTech) in Pushchino, 142290 Pushchino, Russia
[3] Institute of Mathematical Problems of Biology, the Branch of Keldysh Institute of Applied Mathematics, Russian Academy of Sciences, 142290 Pushchino, Russia
[4] Institute of Theoretical and Experimental Biophysics, Russian Academy of Sciences, 142290 Pushchino, Russia
* Correspondence: dmitry.tikhonov@gmail.com

Abstract: The theory of fluids is used to modify the integral equations of the reference interaction site model (RISM) approximation. Its applicability to the study of biomolecules solvation is evaluated. Unlike traditional RISM applications, the new integral equation contains an intramolecular correlation matrix that only needs to be calculated once. This allows us to bypass the effort of repeatedly solving RISM equations and the time-consuming averaging of values obtained for each time point of a molecular trajectory. The new approach allows for the assessment of the conformational transience of dissolved molecules while taking into account the effects of solvation. The free energy of oxytocin, which is a peptide hormone, as well as self-assembled ionic peptide complexes calculated using both the traditional RISM and the new RISM with average matrix (RISM-AM) approach are estimated. The free energy of oxytocin calculated using RISM-AM shows that the statistical error does not exceed the error obtained by standard averaging of solutions in the RISM equation. Despite the somewhat ambiguous results obtained for ionic peptide self-assembly using RISM-AM with Lennard–Jones repulsion correction, this method can still be considered applicable for fast molecular dynamics analysis. Since the required computational power can be reduced by at least two orders of magnitude, the medium-matrix RISM is indeed a highly applicable tool for studying macromolecular conformations as well as corresponding solvation effects.

Keywords: theory of liquids; RISM; solvation energy; RADA16-I peptide complex; molecular dynamics; protofilament

check for updates

Citation: Danilkovich, A.; Tikhonov, D. Theory of Liquids for Studying the Conformational Flexibility of Biomolecules with Reference Interaction Site Model Approximation. *Physics* **2023**, *5*, 1126–1144. https://doi.org/10.3390/physics5040073

Received: 6 September 2023
Revised: 20 October 2023
Accepted: 1 November 2023
Published: 13 December 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

A solvent environment is an essential factor one needs to consider when performing molecular dynamics (MD) and structure modeling. With an explicit solvent, molecular dynamics aims to model the movements of individual water molecules predicted by calculations. Although this methodology sounds quite reasonable for modern theoretical approaches, it suffers from considerable and somewhat prohibitive computational cost during implementation. Such obstacles impose limits on the studies of biomolecules, since they are often susceptible to sharp conformational transitions akin to the protein folding process. Therefore, this approach demonstrates rather low performance when calculating the free energies of multiple states of the same multi-atomic systems. Solving this issue requires accounting for the high-volume degrees of freedom demonstrated by the molecules in the solution.

A model system that implemented the solvent implicitly can fulfill the abovementioned task. Briefly, such an approach replaces discrete water molecules with an ideal space

continuum, emulated by some combination of dielectric and "hydrophobic" properties, assumed to resemble those of physical water. Not only did this allow us to approximate the general solvent effect by the method of molecular dynamics, but also allows us to assess the pool of solvent degrees of freedom implicitly. The latter is important as soon as this approach allows interactions between the molecules of solute and the solvent to be described with the coordinates of the solute molecules. These interactions are accounted for by computable functions to calculate the solvation free energy that changes on the path of the molecule being transferred from a vacuum with zero dielectric constant (e = 0) into a solvent with moderately high dielectrics (e $\gg$ 6), including water (e = 80). This solvent model has practical advantages over water represented explicitly and incorporated by the MD simulation. An implicit solvent model demands a much lower computing cost compared to large molecular systems, i.e., when the solution is taken explicitly. Notably, the enhanced sampling of the conformational space is possible because the viscosity of the solvent is omitted. Such an approach can promote the evaluation of solute conformations because solution molecule rearrangements are neglected due to the implicit solvent. Regarding the method of molecular dynamics, the stage of equilibrating the solvent molecules is removed, which is required for MD simulation in explicit water. Finally, the implicit solution model provides a rather instantaneous dielectric response from the solvent, which could be important to monitor flexible structures with labile conformations. Additionally, pattern-averaging the free energy values removes an excessive number of local energy minima that arise from minor shifts in the solvent's conformations.

Bringing these advantages in effect depends on the kind of software provided by the implicit solvent framework. Generally, the implementation of computationally consumable approximations aims to reveal the electrostatic part of the solvation free energy, which requires further trade-offs between the accuracy and the speed of the modeled dynamics since the stage of molecular docking spliced with MD highly demands algorithmic simplicity. An example of such an approach is the generalized Born (GB) mode [1]. Compared to the Poisson–Boltzmann (PB) model, the GB method is analytical and is able to approximate the calculated electrostatic part of the solvation free energy. The GB approximation model is relatively simple yet capable of providing an effective way to compute the long-range electrostatic interactions over the molecular structure.

The integral equation method is known in the theory of fluids as the reference interaction site model (RISM) approximation, an approach to studying molecular solvation [2]. At infinite dilution approximation, the solution density can be neglected and the RISM method enables one to find either solution–solvent or atom-to-atom correlation functions from which the solvation thermodynamics, in particular, the Gibbs energy value can be calculated [3]. Apart from the atom-to-atom interacting potentials, the "prime load" of the RISM theory is the solvent molecules geometry, which is depicted with a matrix of interatomic distances. The correlation functions and solvation thermodynamics, derived as the RISM equation solutions, correspond to the unique state, i.e., the "rigid" form of the dissolved molecule, including one of the "frozen" states of the MD trajectory that are calculated during the molecular dynamics run. However, the molecules at room temperature often demonstrate a high level of versatility, including during the semi-equilibrium phase of the MD trajectory, where the molecules are still capable of intensively changing their conformations. Fluctuations in molecular geometry influence the solvation thermodynamics and can be assessed by accounting for the atom's mutual influences, which are defined for each state of geometry, followed by averaging the values estimated for the Gibbs energy. The MD trajectory facilitates a set of the geometry configurations, which are chained by the "frozen" states of the molecular structure that are generated in course of the MD simulation [4].

In RISM theory, there is an alternative way of considering the conformational versatility of a solute, which is the focus of this study. For the case of infinite dilution, we develop a modified system of RISM equations, which is averaged over the entire MD trajectory only once, RISM with average matrix (RISM-AM). The solutions of this system resemble

the solutions of standard correlation functions averaged over the trajectory. Indeed, the proposed method for estimating the mean value of the Gibbs energy from the data on the semi-equilibrium scope of the MD trajectory does not require solving RISM equations for each frame (time point) of the dynamics. However, it is not quite clear a priori whether and how the thermodynamic quantities estimated using the averaged RISM equation correspond to the actual means of thermodynamics.

The Gibbs energy and excess chemical potential are fundamentally important thermodynamic quantities. According to the Ornstein–Zernike theory, the method of integral equations applied to simple liquids allows one to introduce a term for the excess chemical potential, which can be calculated explicitly [4,5]. This formula was obtained in general form analytically by performing thermodynamics integration for the energy. Although it has been generalized quite well for the evaluation of polyatomic molecules using RISM [6], nevertheless, the simpler term for excess chemical potential can be discarded. For this purpose, the nonlinear part of the RISM equation is represented by quite a simple analytical term of the so-called "closure equation" [4]; otherwise, thermodynamic integration requires numerical calculations.

Unlike in standard RISM equations, the mean matrix allows one to numerically integrate the mean of the entire MD trajectory only once. This approach is substantially simpler than solving the conventional RISM equations for each current time period in an MD experiment, so the free energy integration step is repeated at each time point along the MD trajectory in order to eventually obtain the mean value. Therefore, the new methodology can provide mean value equations with different functions to estimate the Gibbs energy. To not forget is that the closure term takes into account the entropic contribution to the Gibbs energy, which manifests itself in a particular function for correcting the repulsion of the bridge functional added to the equation [4,7].

The formula for the excess chemical potential uses a correction term derived from the "first principle" thermodynamic perturbation theory (TPT). Instead of time-consuming numerical integration that requires huge computing, the Gibbs energy can be estimated by evaluating the approximated value according to the TPT.

This functionality of RISM-AM is an advantage of this method, which, in addition, requires further development and full implementation of this approach.

To evaluate the applicability of RISM-AM for estimating the Gibbs energy, the corresponding values are calculated using the perturbation theory formulas by numerical integration and analyzed in comparison with the performance of other methods.

2. Methodology

2.1. RISM Equations for an Infinitely Diluted Solution

The equation for an infinitely diluted solution [3], when the partial density of the dissolved substance is neglected, relates the direct, $c_{i\alpha}(r)$, and full, $h_{i\alpha}(r)$, correlation functions, where r denoted the coordinate, by assigning to the atoms of the solvent molecule ($\alpha = 1,\ldots, M$) distributed around the atom of the solute i ($i = 1,\ldots, N$):

$$\mathrm{H}^{uv} = \mathrm{W}^{u}\mathrm{C}^{uv}(\mathrm{W}^{v} + \rho\mathrm{H}^{vv}) \tag{1}$$

and

$$\mathrm{H}^{uv}_{i\alpha} = \hat{h}_{i\alpha}(k) = \frac{4\pi}{k}\int_0^\infty rdr\,\sin(kr)h_{i\alpha}(r),$$

$$\mathrm{C}^{uv}_{i\alpha} = \hat{c}_{i\alpha}(k) = \frac{4\pi}{k}\int_0^\infty rdr\,\sin(kr)c_{i\alpha}(r),$$

where k is the wave vector. In Equation (1), the expression in the parentheses reflects the static susceptibility of the solvent. It involves a matrix W^{v} of intramolecular correlation functions describing the rigid bonds between the atoms of the solvent molecule, the matrix H^{vv} of complete correlation functions of the pure solvent with the density ρ. The upper indexes v and u stand for solvent and for solute, respectively, in Equation (1).

The elements of the matrix of intramolecular correlation functions (W^u) stand for the rigid bonds between the atoms in dissolved molecules [2]:

$$W_{ij}^u = \hat{\omega}_{ij}^u(k) = \delta_{ij} + (1 - \delta_{ij}) \frac{\sin(kl_{ij})}{kl_{ij}}, \quad (2)$$

where l_{ij} is the distance between the atoms i and j, and δ_{ij} is the Kronecker delta.

The matrix of intramolecular correlation functions also contains diagonal static susceptibility elements related to the static structure factor. This means that the expression defines a constant and can be either calculated or taken from the experimental data. The dimension of the matrix W^u equals to the number of atoms in the solution molecule, N, and the dimension of the matrix W^v is equal to the number of atoms in the solvent molecule, M. Therefore, the correlation function matrices are the $N \times M$ matrices.

In addition to Equation (1), related to the correlation functions in coordinate space to provide the solvability of the system of correlation functions, there is the so-called closure equation, which can be either in the form of a hypernetwork closure (HNC) [8]:

$$c_{i\alpha} = \exp(-\beta u_{i\alpha} + \gamma_{i\alpha}) - 1 - \gamma_{i\alpha}, \quad (3)$$

or as its partially linearized form (PLHNC) [9]:

$$c_{i\alpha} = \exp(-\beta u_{i\alpha}) F(\gamma_{i\alpha}) - 1 - \gamma_{i\alpha}. \quad (4)$$

While $\gamma_{i\alpha}(r) = h_{i\alpha}(r) - c_{i\alpha}(r)$ is a nondirect correlation function, $u_{i\alpha}(r)$ are pair-wise interaction potentials between the atom i of the solution molecule and the atom α of the solvent molecule, where $\beta = 1/kT$ with k being the Boltzmann constant and T the temperature, is the reciprocal temperature, let us assume:

$$F(\gamma_{i\alpha}) = \begin{cases} \exp(\gamma_{i\alpha}), & if\ h_{i\alpha}(r) \leq 0, \\ 1 + \gamma_{i\alpha}, & if\ h_{i\alpha}(r) > 0. \end{cases}$$

2.2. Molecular Dynamics and Integral Equations

The RISM equations allow us to obtain correlation functions for a specific geometry of a dissolved molecule, characterized by a given matrix of interatomic distances. Since a molecule in a solution undergoes significant conformational transitions, it would be incorrect to limit the matrix to a single set of interatomic distances. One way to take into account the versatility of a molecule on an equilibrium MD trajectory is to average the values of correlation functions for a set of multiple geometry configurations.

Assume the MD trajectory is a set of frames, each defined by the instant interatomic distances in the dissolved molecule $\{l_{ij}(t)\}$, with $t = 1,\ldots, N_T$, where t stands for the value, which depends on the molecule geometry, and N_T denotes the number of solute configurations on molecular trajectory. Having denoted the static susceptibility of the solvent as V (see Equation (13) in Ref. [10] for water solvent) and $h = vec(\mathrm{H}^{uv})$, and $c = vec(\mathrm{C}^{uv})$, the operation $vec(x)$ expresses the matrix transformation as a column-vector, where the columns in matrix x are consequently written one under another; hence, Equation (1) is applicable to every t configuration in general, such as

$$\mathrm{h}_t = \mathrm{A}_t \mathrm{c}_t = (V \otimes \hat{\mathrm{w}}_t^u)\mathrm{c}_t, \quad t = 1, \ldots, N_T, \quad (5)$$

where $\otimes$ is the direct matrix product (Kronecker product) of matrices V and $\hat{\mathrm{w}}_t^u$ combines the elements of the matrix of intramolecular correlation functions (see Equation (2)). Hence, this notation of the matrix A_t corresponds to the set of distances $\{l_{ij}\ (t)\}$. Solutions of

Equation (5), assessing the set of t geometries, allow for the mean of the full correlation functions vector to be determined:

$$\overline{\mathrm{h}} = \frac{1}{N_T}\sum_{t=1}^{N_T} \mathrm{h}_t = \frac{1}{N_T}\sum_{t=1}^{N_T} \mathrm{A}_t\mathrm{c}_t. \quad (6)$$

There is a shortcoming in Equation (6): in order to average solutions to the equations, it is necessary to solve the RISM equations for numerous snap-shot configurations of the molecule according to the information provided by the MD trajectory data. This obstacle can be avoided by estimating the mean vector of full correlation functions:

$$\langle \mathrm{h} \rangle = \left(\frac{1}{N_T}\sum_{t=1}^{N_T} \mathrm{A}_t\right)\left(\frac{1}{N_T}\sum_{t=1}^{N_T} \mathrm{c}_t\right).$$

At this point, let us introduce the mean vector of the direct correlation functions formally:

$$\langle \mathrm{c} \rangle = \frac{1}{N_T}\sum_{t=1}^{N_T} \mathrm{c}_t.$$

Since the only solution for the mean value of interest is

$$\langle \mathrm{h} \rangle = \overline{\mathrm{A}}\langle \mathrm{c} \rangle, \quad (7)$$

an average-matrix of the RISM equations could be written as:

$$\overline{\mathrm{A}} = \frac{1}{N_T}\sum_{t=1}^{N_T} \mathrm{A}_t \quad (8)$$

We call the solutions to Equation (7) pseudo-average correlation functions. The convenience of Equation (7) is quite straightforward: since a single RISM equation must be solved, the $\bar{\mathrm{A}}$ matrix (8) already contains averaged intramolecular correlations for the entire MD trajectory. However, the question remains: what are the differences between the pseudo-average functions and the actual average correlation functions obtained using standard RISM? This difference can be estimated as the difference between the results obtained using Equations (6) and (7). A direct transformation of the difference expression,

$$\Delta = \overline{\mathrm{h}} - \langle \mathrm{h} \rangle = \frac{1}{N_T^2}\sum_{t=1}^{N_T}\sum_{s=1}^{N_T} (\mathrm{A}_t - \mathrm{A}_s)\mathrm{c}_t, \quad (9)$$

yields the expression

$$\Delta = \overline{\mathrm{A}}\overline{\mathrm{c}} - \overline{\mathrm{h}},$$

assuming the difference between the average values arisen from Equation (7) as the pseudo-mean correlation functions, and the full correlation functions are expressed with the values, averaged on MD trajectory. At the same time, the differences, originated from the use of the linear Equation (7), can be reduced by taking the direct full correlation functions as averaged probe functions. It is worth mentioning that the term for the difference (9) equals to zero if all the matrices A_t are equal to the average matrixes been equal each other. From Equation (9) it follows that the difference Δ is as small as the difference between the intramolecular correlation matrices calculated for the two neighboring points on the MD trajectory. This statement is true for the semi-equilibrium portion of the MD trajectory. Indeed, let us divide the differences arisen in result of averaging intramolecular matrices for the MD trajectory with Equation (9) employed for the particular momentum states on the trajectory:

$$\Delta\overline{\mathrm{A}}_s = \frac{1}{N_T}\sum_{t=1}^{N_T}(\mathrm{A}_t - \mathrm{A}_s)$$

.

When molecular system subjects to some thermal fluctuations near the equilibrium state, there is no reason to neglect the points, at which the values of $\Delta\bar{A}_s$ are small. Therefore, the value of $\Delta\bar{A}_s$ can be close to zero, as determined by the direct correlation function:

$$\Delta = \frac{1}{N_T}\sum_{s=1}^{N_T} \Delta\overline{A}_s c_s.$$

The linear part of the RISM equation assumes the error is estimated precisely by the expression (7) for the average-matrix. This is a time-consuming task, since estimating the contribution of the nonlinear part of the RISM equation seems somewhat problematic. Therefore, the closure Equations (3) and (4) can be related to the error problem. Anyway, it appears that preliminary calculations are to be made using MD data and to compare the value of Δ with the dispersion of the average correlation function. When the error (Δ) is smaller than the dispersion, the MD data can be used to calculate the mean value by solving the RISM-AM. In such cases, the proposed approach is an applicable tool since the static susceptibility of the solvent matrix V remains constant, while the solute intramolecular correlation matrix W^u can vary. If the functions subjected to significant fluctuations, then the matrix A to be averaged, but the solvent related matrix V does not change.

Finally, the average-matrix of the RISM equations can be defined via the following expression:

$$\overline{A} = V \otimes \left(\frac{1}{N_T}\sum_{t=1}^{N_T} \hat{w}_t^u \right) = V \otimes \overline{w},$$

2.3. Intramolecular Correlation Matrix

Let us dwell upon the calculation of the mean matrix of intramolecular correlations. Let us consider a simple case of a matrix $\overline{w}$ defined by interatomic distances in a solute molecule averaged over the trajectory $\langle l_{ij} \rangle$ according to the formulation (2) for rigid bonds:

$$\overline{\omega}(\langle l_{ij} \rangle, k) = \delta_{ij} + (1 - \delta_{ij}) \frac{\sin(k\langle l_{ij} \rangle)}{k\langle l_{ij} \rangle}, \tag{10}$$

where

$$\langle l_{ij} \rangle = \frac{1}{N_T}\sum_{t=1}^{N_T} l_{ij}(t).$$

Definition (10) has an advantage since the parameter dependence of functions on the scalar value of average interatomic distance. This not only allows the data set to be compressed, but also significantly reduces the storage space for information associated with the data in the w matrix. The matrix memory saving mode solves the main practical problem, since this allows for the numerical implementation of the method. This is important for analysis of large molecules, which typically requires almost prohibitively high computational power. An analysis of the dependence of the parameters (10) shows that the memory saving mode can be achieved using two quite simple techniques. First, inside the molecular globule, there are many approximately equal interatomic distances, since the lengths of covalent bonds are largely limited by the nature of the chemical elements. Once these distances are rounded off to predefined limits, the whole set of interatomic distances can be also reduced to a compact set of the nonequivalent values [4]. (By rounding these distances to given limits, the entire set of interatomic distances can also be reduced to a compact set of nonequivalent values [4].) In the next step, intramolecular correlations of distant atoms can be relaxed since the atoms influence each other as a low-force event. As mentioned earlier, correlation relaxation allows for the use of a short form of intramolecular correlation functions, suitable for in silico studies of large molecular systems, including DNA (deoxyribonucleic acid) [10]. Therefore, calculating the average molecular geometry becomes a task similar to operating an object with rigid geometry. Both cases can be

straightforwardly implemented in a software package for use with input data represented by any interatomic distance matrix. Although the expression (10) looks simple, it poorly performs in experiments. Often, the correlation functions did not coincide well enough with the average values, contaminated by some artifacts like the parasitical peaks spaced apart distantly. Thermodynamics deals mostly with the mean values for the cases of rigid small molecules. With all the average values mentioned above, this approach is considered not to be of high use for the analysis of versatile molecular structures.

A different approach can be taken to average intramolecular correlation functions, using histograms of interatomic distance distribution, $P_{ij}(r)$, obtained from the corresponding MD trajectory:

$$\langle \omega(k) \rangle = \int_0^\infty P_{ij}(r) \frac{\sin kr}{kr} dr. \quad (11)$$

Such a way of averaging the values looks more practical for the system of a multistate molecule. This statement is illustrated on Figure 1, where the upper graph displays the distance distribution between two groups of atoms. The histogram shows that most distances indicate that the atoms are tightly grouped within a 2 Å area. However, in the scope of conformational versatility, there are distances as big as 7 Å, indicated by the corresponding peaks in the histogram. It is understandable that all these wave vector functions are significantly different, with the exception of the numerical volumes around the zero value. Consequently, solutions to equations containing matrices of this type can differ significantly.

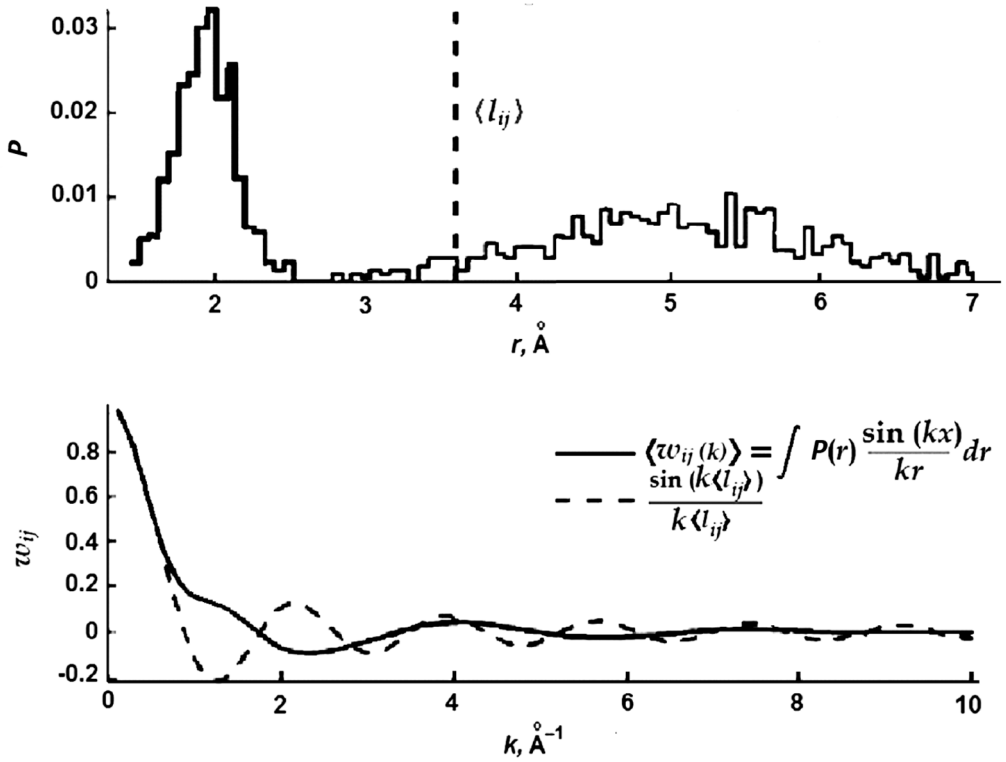


Figure 1. Intramolecular correlation functions derived from the average interatomic distances versus histogram of interatomic distance distribution. Two approaches of calculating the average values compared. See text for details.

Of high practical importance is the issue of using functions averaged with the use of histograms. The averages allow for efficient calculations because the short-range functions in the intramolecular function matrix converge better by solving the RISM equations iteratively. Furthermore, the time required to compute the solutions of the RISM-AM equations for an MD trajectory is comparable to the cost of solving the RISM equations for all the individual (snap-shot) configurations of a molecule. The method (11) does

not allow us to determine an equivalent functions to implement full data compression. However, the nature of short-range average intramolecular correlations allows the functions to be cut off at the maximum of the wave vector k^*_{ij}:

$$\int_{k^*_{ij}}^{\infty} \omega_{ij}(k)dk / \int_0^{\infty} \omega_{ij}(k)dk < \xi, \quad (12)$$

while the accuracy power, ξ, of this operation can be set arbitrarily and thereby predetermined. Factor ξ can influence both matrix content and the solution of the equations, since the approximated matrix of $\omega_{ij}(k)$ elements filled with another approximate values of $\omega_{ij}(k)H(k - k^*_{ij})$, where $H(x)$ states for the Heaviside step function. The choice of the ξ-value magnitude implies that the matrix should be rather a definite one, so the errors of the cutting-off procedure do not exceed the accuracy of the RISM equation solutions, calculated numerically with steady iterations. Additionally, the effect of cutting-off transformation to be reduced even more by setting the boundary for the wave vectors k^*_{ij} of the ω_{ij} (k^*_{ij}) added correctly to the term (12).

2.4. Thermodynamics

To compare the free energies of the dissolved molecules, one should find the Gibbs free energy, ΔG, as a sum of the excess chemical potential $\Delta\mu$, and molecular mechanical energy, E_{MM}:

$$\Delta G = E_{\mathrm{MM}} + \Delta\mu. \quad (13)$$

In case the energy was calculated for a panel of snap-shot configurations, the array of data on configurations has to be averaged:

$$\Delta G = \sum_i (E_i^{\mathrm{MM}} + \sum_i \Delta\mu_i).$$

The excess chemical potential $\Delta\mu$ value equals the work performed on the particle being introduced into the medium. On the path, the particle penetrated the medium can be emulated by gradually weakening the interaction potentials with the scaling parameter λ:

$$\Delta\mu = 4\pi\frac{\rho}{2\beta}\sum_{i\alpha} d\lambda \int_0^{\infty} dr r^2 g_{i\alpha}(r, \lambda)\frac{\partial u(r, \lambda)}{\partial\lambda}, \quad (14)$$

where $g(r,\lambda) = h_{i\alpha}(r,\lambda) + 1$ determines the atom-to-atom radial functions of distribution of the solvent atom α around the atom i of the solvent, while the RISM equations solutions are weakened with the λ parameter considering the inter-atomic interaction potentials, $u(r,\lambda)$. In this vein, Section 3 deals with the correlation functions, assuming the thermodynamic potentials calculated either with mean, pseudo-mean or ordinary functions.

In some cases, the integration over the λ-factor in Equation (14) can be performed analytically. Thus, in the theory of Gaussian density fluctuations (GF), the expression for the excessive potential exists in the form [11]

$$\Delta\mu^{\mathrm{GF}} = 4\pi\frac{\rho}{2\beta}\sum_{i\alpha}\int(-2c_{i\alpha}(r) - h_{i\alpha}(r)c_{i\alpha}(r))r^2 dr. \quad (15)$$

Another kind of the potential was defined by the Singler–Chandler (SC) expression for the excess chemical potential [5,11–13], and has a different form for the HNC:

$$\Delta\mu^{\mathrm{SC}} = \Delta\mu^{\mathrm{GF}} + 4\pi\frac{\rho}{2\beta}\sum_{i\alpha}\int h_{i\alpha}^2(r)r^2 dr. \quad (16)$$

As shown [6], the closure,

$$c_{i\alpha}(r) = \exp(-\beta u_{i\alpha}(r) + \gamma_{i\alpha}(r) + B_{i\alpha}(r)) - \gamma_{i\alpha}(r) - 1,$$

treated according to the theory of liquids is related to the term,

$$\beta\Delta\mu = \sum_{i\alpha}\rho_\gamma \int d^3r\left\{\frac{1}{2}h_{i\alpha}^2(r) - \frac{1}{2}h_{i\alpha}(r)c_{i\alpha}(r) - c_{i\alpha}(r) + B_{i\alpha}(r)\right\} + \sum_{i\alpha}\rho_\gamma \int d^3r \frac{h_{i\alpha}(r)}{\gamma_{i\alpha}(r)}\int_0^{\gamma_{i\alpha}(r)} d\gamma'\gamma' \frac{\partial B_{\alpha\gamma}(\gamma')}{\partial\gamma'},$$

where $B_{\alpha\gamma}(r)$ is a bridge functional, presented by the infinite integrant. It is noteworthy that in RISM the functionality of the bridge is unknown; therefore, the terms for a closure are based either on fluid theory or other assumptions. From above it follows that if the integration of bridge functional γ is possible, the result contains simple functions. For instance, if $B_{\alpha\gamma}\ (\gamma') = 0$, then the second part of the term vanishes, so the entire term becomes just the SC expression [13]. By analogy, the term for the PLHNC kind of closure is defined similarly [9]:

$$\Delta\mu^{\mathrm{SC}} = \Delta\mu^{\mathrm{GF}} + 4\pi\frac{\rho}{2\beta}\sum_{i\alpha}\int h_{i\alpha}^2(r)H(-h_{i\alpha}(r))r^2dr. \tag{17}$$

Additionally, the partial-wave method (PW) can be used to form another expression for the excess chemical potential:

$$\Delta\mu^{\mathrm{PW}} = \Delta\mu^{\mathrm{GF}} + \frac{\rho^2}{2\beta(2\pi)^2}\sum_{i\alpha}\int 4\pi k^2\hat{h}_{i\alpha}(k)d\hat{\zeta}_{i\alpha}(k)dk, \tag{18}$$

where the elements of matrix Z are:

$$\mathrm{Z} = \left\{\hat{\zeta}_{i\alpha}(k)\right\} = \mathrm{C}^{uv}(\mathrm{W}^v + \rho\mathrm{H}^{vv})\mathrm{C}^{vv}.$$

More elaborate comments on these expressions can be found in the original papers [14,15]. The authors introduced the repulsion correction to the closure equation in order to balance the entropy component that could influence the energy [6]. Therefore, the modified hyper net chain equation corrected with the repulsive bridge correction (RBC) as HNC/RBC takes the form

$$\langle c\rangle_{i\alpha} = \exp\left(-\beta u_{i\alpha} + b_{i\alpha} + \langle\gamma\rangle_{i\alpha}\right) - 1 - \langle\gamma\rangle_{i\alpha}, \tag{19}$$

while the partially linearized hyper net chain equation (PLHNC/RBC) is written as

$$\langle c\rangle_{i\alpha} = \exp\left(-\beta u_{i\alpha} + b_{i\alpha}\right)F(\langle\gamma\rangle_{i\alpha}) - 1 - \langle\gamma\rangle_{i\alpha}. \tag{20}$$

In addition, the correction index $b_{\mathrm{i}\alpha}$ can be defined as

$$\exp\left(b_{i\alpha}\right) = \prod_{\mu=\alpha}\left[\omega_{\mu\alpha} * \exp\left(-\beta u_{i\mu}^{\mathrm{RBC}}\right)\right],$$

where * denotes convolution in coordinate space, while the repulsive potential $u^{\mathrm{RBC}}{}_{i\mu}$ can be represented by either the repulsive r^{-12}-term of the Lennard–Jones potential (R12):

$$u_{i\mu}^{r^{-12}}(r) = 4\varepsilon_{i\mu}\left(\frac{\sigma_{i\mu}}{r}\right)^{12}, \tag{21}$$

or the corresponding part of the Weeks–Chandler–Andersen (WCA) potential:

$$u_{i\mu}^{\mathrm{WCA}}(r) = 4\varepsilon_{i\mu}\left[\left(\frac{\sigma_{i\mu}}{r}\right)^{12} - \left(\frac{\sigma_{i\mu}}{r}\right)^{6} + \frac{1}{4}\right]H\left(2^{1/6}\ \sigma_{i\mu} - r.\right) \tag{22}$$

For the RISM equations with a repulsion correction, the excess chemical potential cannot be revealed in explicit form analytically, though it can be integrated numerically by solving the RISM system for values of λ on the interval from 0 to 1. To date, despite all the progress made in making complex, highly specialized graphics processing unit GPU-

empowered computing platforms, integration still remains a laborious, time-consuming, and a costly stage of experiment that tends to consume quite a large amount of computing power, depending progressively on the size of the molecular system under investigation. The advantage is that the term (14) can be modified and then used to average the solutions of Equation (7) along the MD trajectory.

2.5. Calculations

To compare the data assessed with pseudo-mean correlation functions (7) with real mean values (6), we used the following MD trajectories calculated for the peptide hormone oxytocin: peptide in vacuum (A), peptide in explicit water (B), and peptide in implicit solvent according to the GB method (C). The correlation functions were constructed for 1250 peptide configurations selected randomly from each trajectory. These correlation functions were used to assess the mean values of the Gibbs energy calculated using several different terms to find the excess chemical potentials (15)–(18) and marked GF, SC, and PW, respectively. An approximate expression (23), taking into account various expressions for the repulsion corrections (21) and (22), allowed us to find the Gibbs energy, designated as RBC1 and RBC2, respectively. If the data set is representative, it is possible to compare the real mean values with the solutions to the mean equation. Parameter ξ, emulated the precision level, which was preset during the cutting-off procedure performed on the intramolecular correlation functions (see Equation (12)). We arbitrarily set the ξ, value to 0.01. Histograms of interatomic distance distributions were calculated for 12,500 states taken from the MD trajectory. The mean RISM equation was always solved for the same grid of parameters and the potentials. It is noteworthy that the numerical integration over the parameter λ in Equation (14) with respect to the repulsion corrections (21) and (22) was carried out with an adaptive step. The integration accuracy was constantly controlled by repeating the calculation of the selected steps. The values obtained by the integration are designated as INT1 and INT2, respectively.

3. Implementation

3.1. Oxytocin Study

Here, we compare the approaches defined by Equations (6) and (7) regarding performance and differences between the average and pseudo-mean correlation functions. For each MD trajectory, one complete interatomic correlation function was chosen with the largest integral of the modulus of the difference between the real average and pseudo-average functions. These HNC closure functions are shown in Figure 2, where the difference and relative standard deviation are also presented. The atoms in the oxytocin molecule that obey the conditions are listed above for molecular dynamics A, B, and C and numbered. It can be seen that on MD trajectories with different environments, the functions with the maximum deviation correspond to different atoms of the peptide. For trajectory A, the differences between these functions are slightly greater than the dispersion in the vicinity of the first maximum of the total correlation function. Elsewhere, the fractional difference is smaller than the dispersion, including trajectories B and C. The PLHNC closure for the RISM equation provided exactly the same results, including the atoms that demonstrated the greatest difference between the real and pseudo-mean correlation functions. The finding that the difference between the real mean and the pseudo-mean did not depend on the form of the closure equation used supports the assumption that the nonlinear portion of the RISM equations has a fairly minor effect on the magnitude of the difference.

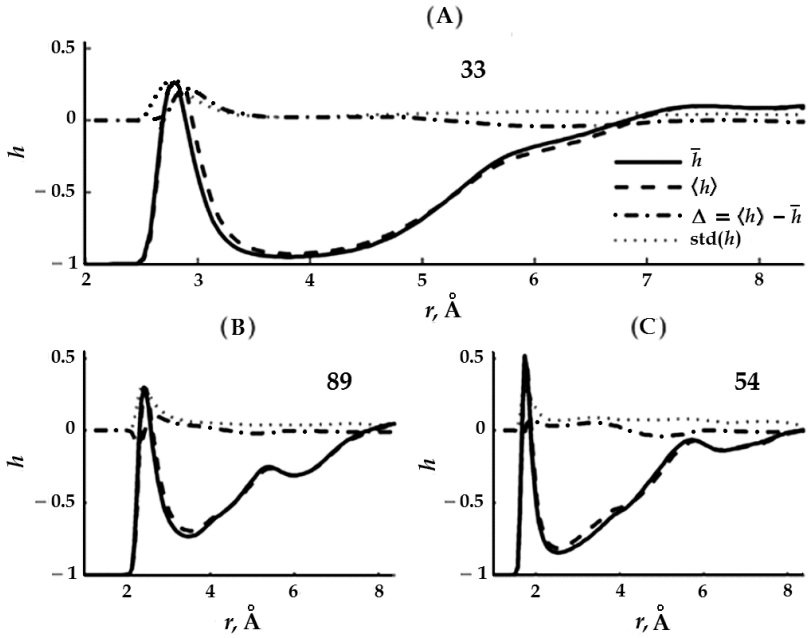


Figure 2. Real mean (solid line) and pseudo-mean (dashed line) intramolecular correlation functions; their difference (Δ) and standard deviation (std(h)) for the atoms 33, 89, and 54 of oxytocin, having the largest modulus of the integral values of the different molecular dynamics functions for the (**A**–**C**) states of oxytocin as indicated. See text for details.

Considering the analysis of thermodynamics, Table 1 lists the mean values of molecular mechanics energy (see Equation (13)), the excess chemical potential, and the Gibbs energy. One can see that the conformational states in vacuum (A) are the least preferable, while for the cases B and C the values of the Gibbs energy are quite close and differ from each other by less than 1 kcal (Figure 2). The data presented in Table 2 provide comparison between the RISM Gibbs energy and that one calculated by solving the average equation. For comparison, the values of the root mean squares (standard deviations) of the excess chemical potential calculated for each subsequent point of the molecular trajectory are given in parentheses. The data were obtained with the use of HNC and PLHNC closures. It can be seen that the results can be divided into groups. Thus, the differences in chemical potential calculated using the SC (17), RBC1 (21), and RBC2 (22) formulas are negative and comparable in size to the magnitude of dispersion. For the trajectory A, they slightly exceed the dispersion, and for the trajectories B and C, they are less than the dispersion. The performance of the GF (15) and PW (18) formulas does not differ significantly between the HNC (19) and PLHNC (20) closures, as the values of differences are positive and comparable to the dispersion when the HNC closure was used. However, the differences for the PLHNC closure remain positive and smaller than the dispersion. Hence, the thermodynamic errors resulted from the use of RISM-AM are less, than the dispersion limits of the values calculated for a row of instant geometry states of the same structure.

Table 1. Mean values of molecular mechanics energy (E_{MM}); Gibbs energy ($\Delta\mu_{GBSA}$) assessed with generalized Born approach with surface factor; and Gibbs energy (ΔG) of the oxytocin for different peptide states. See text for details. The values are given in kcal/mol. The values in the parentheses represent the standard deviations.

Oxytocin Case	$-E_{MM}$	$-\Delta\mu_{GBSA}$	$-\Delta G$
A	−126.6 (11.4)	−62.2 (8.0)	−188.7 (9.4)
B	−5.9 (11.6)	−190.3 (7.9)	−196.2 (8.5)
C	7.5 (12.2)	−204.7 (10.5)	−197.1 (5.6)

Table 2. The Gibbs energy assessed with averaging the solutions for the RISM equations and by solving a single equation of the RISM-AM for different oxytocin states. See text for details. The values are given in kcal/mol. The values in the parentheses represent the standard deviations.

Oxytocin Case	Closure Type	SC	GF	PW	RBC1	RBC2
A	HNC	−20.4 (17.0)	18.0 (8.4)	12.6 (7.6)	−20.0 (17.5)	−19.4 (17.0)
	PLHNC	−21.0 (17.0)	6.2 (6.7)	1.7 (6.4)	−21.3 (17.6)	−20.4 (17.0)
B	HNC	−12.2 (10.9)	12.5 (9.5)	9.0 (8.8)	−12.0 (12.5)	−11.7 (11.7)
	PLHNC	−12.5 (10.8)	6.7 (8.5)	3.6 (8.0)	−12.7 (12.4)	−12.2 (11.7)
C	HNC	−12.2 (12.8)	14.7 (14.3)	11.2 (13.2)	−11.3 (14.0)	−11.2 (13.4)
	PLHNC	−12.4 (12.8)	8.3 (12.8)	5.2 (12.0)	−12.0 (14.0)	−11.7 (13.4)

Figure 3 displays the Gibbs energy, calculated by using pseudo-mean functions with two forms of the closure. Let us notice that the changes in the Gibbs energy well coincide with the data obtained earlier [4]. Differences between the values of the Gibbs energy were estimated by the methods of thermodynamic perturbation theory (RBC1, RBC2), derived by thermodynamic integration (INT1, INT2) with the terms of repulsion correction. A comparison of the absolute values of the Gibbs energy assessed by these methods reveals that the repulsion correction (21) used for the thermodynamics integration does not influence the results, whereas the type of correction (22) used to assess the Gibbs energy yields the values that differ at least by 70–100 kcal/mol.

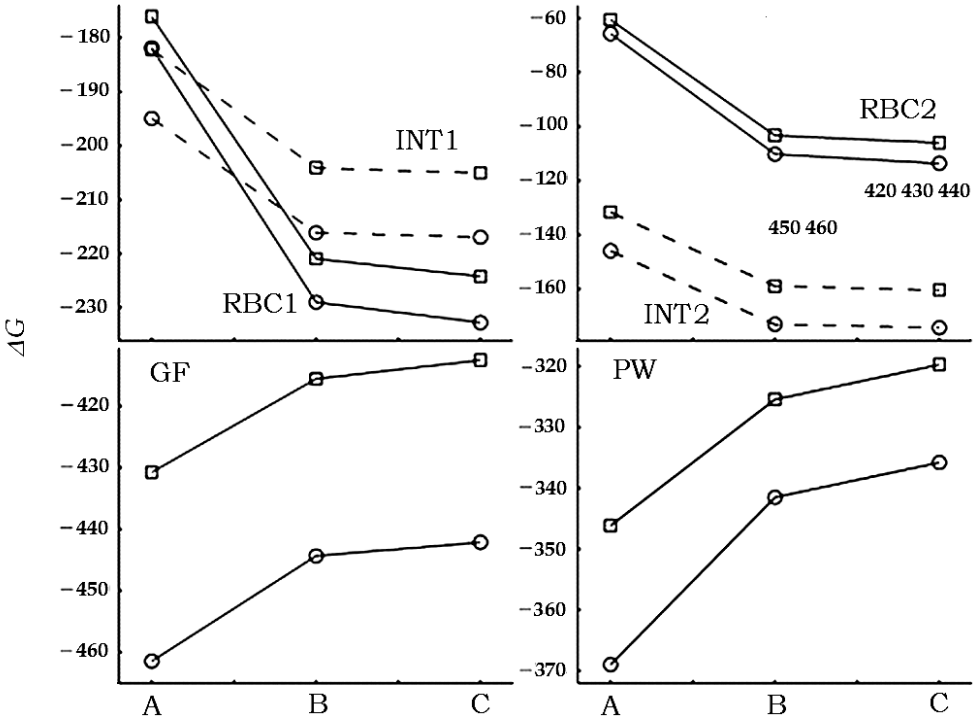


Figure 3. The Gibbs energy of oxytocin calculated by solving RISM-AM equation. The results of thermodynamic integration are shown by dashed lines (INT1, INT2). The closure terms are HNC or PLHNC indicated by circles or squares, respectively. See text for details.

Indeed, the absolute value of the Gibbs energy is less informative than the comparative change in the Gibbs energy resulted from the molecules conformation transition. Numerical data on the energy differences for different states of the molecule are presented in Table 3. The data were obtained by solving the RISM-AM equation with HNC and PLHNC closures. Comparing the differences between the actual mean Gibbs energy values in Table 2, that the difference in energy values decreases steadily for all the forms of excess chemical potential expression. Changes in the Gibbs energy resulted from changes in the configuration of the molecule in vacuum and the solution, estimated by the methods of thermodynamic

perturbation theory (RBC1, RBC2) and thermodynamic integration (INT1, INT2) corrected by the repulsion correction in the forms of Equations (21) and (22). As a result of numerical integration, the difference between the values decreased as nearly twice, while changes in the Gibbs energy somehow turn out to be overestimated [4]. Therefore, the decrease in the free energy assessed with the repulsion correction contradicts to the very idea of correcting the entropy component with use of the repulsion correction. It may be necessary to estimate the excess chemical potential using a method based on the thermodynamic perturbation theory (PTP), given that the machinery used is not informative enough to track the path of the peptide from vacuum to solution sufficiently.

Table 3. The Gibbs energies (ΔG) calculated with the use of pseudo-mean correlation functions shown as differences between A, B, and C states of oxytocin; in kcal/mol. See text for details.

	Closure Term	SC	GF	PW	RBC1	RBC2	INT1	INT2	GBSA
ΔG(B) − ΔG(A)	HNC	−43.3	17.1	27.4	−46.9	−44.6	−21.1	−27.0	−7.5
	PLHNC	−41.5	15.2	20.8	−44.8	−42.5	−21.9	−27.3	
ΔG(C) − ΔG(A)	HNC	−46.5	19.4	33.1	−50.8	−47.9	−21.9	−28.4	−8.4
	PLHNC	−44.4	18.3	26.4	−48.1	−45.3	−22.7	−28.8	
ΔG(B) − ΔG(C)	HNC	3.2	−2.3	−5.7	3.8	3.3	0.8	1.4	0.9

3.2. Ionic Peptides Studied with RISM in Average Matrix Form

The applicability of RISM with the average matrix to study thermodynamics of distinct molecular structures was evaluated with the ionic peptide complexes as an exemplars contained several charged amino acid residues. Studies of these self-assembling peptides have been conducted with different means, including the method of molecular dynamics [16]. Since these peptides are capable of forming filament-like nano-structures [17,18] composed of ionic peptides, they are becoming even more important to design and manufacture the hyper-hydrated scaffolds of bio-gels supplied to regenerative medicine, biotechnology, and nowadays tissue engineering. Since the fine details of the self-assembling machinery are still remained an enigmatic issue, we then decided to apply the RISM with the average matrix to analyze ionic peptide complexes by using the data on MD trajectory. At sum, the Gibbs energy and the free energy of peptide complexes formed in explicit water were calculated to be compared.

3.2.1. Models Design

Five monomer structures of the self-assembling peptides of 16 or 8 amino acid residues—RADA16, RADA8, KADA16, KAEA16, RLDL16, RAEA16, and KLEL8, were build to create dimer-, trimer-, tetramer-, and octamer-complexes composed of the peptides situated in anti-parallel β-conformation. Three-dimensional models of the peptide complexes were assembled in silico with HyperChem v.8 [19]. Molecular docking was achieved with HEX6.1 software [20] accounted both for molecular geometry and electrostatics. Alternatively, we considered the results of docking with GRAMM-X [21], in the mode with electrostatics switched off. Remarkably, these two approaches yielded similar results regarding the most stable complexes composed of the interacting peptides taken for investigation. The protonation state for the peptides in further experiment was set to pH 7. The corresponding PDB files were used as an input data on the structure evaluated then with molecular dynamics in explicit water environment.

3.2.2. Molecular Dynamics

MD in the explicit water were calculated with AMBER v.11 [22]. The object was placed in a periodic boundary condition (PBC) box with TIP3P molecules [23]. The dynamics of every sample were evaluated with the force field ff03 (AMBER) [24] at 300 K. After the system was minimized and the interatomic bonds fully normalized in length, the cell was packed with TIP3P as densely as possible. The system was subjected to brief heating for 10 ps at constant volume till the TIP3P molecules lined up to the cell borders. Further, the atom coordinates in the peptide sample were locked for two sessions of energy minimization conducted with the steady release of atom coordinates. The system was heated up to 300 K at the same volume for 10 ps, and the run of molecular dynamics was initiated for 15 ps to set the PGU to have a proper density. At that point, the system was considered to have reached the state of initial point of MD run lasted for 10 ns under constant pressure and other conditions.

3.2.3. Gibbs Energy of Different Ionic Peptide Complexes

Molecular trajectories for 10 ns at 315 K were calculated for the peptides of RADA16, KADA16, KAEA16, RLDL16, and RAEA16 and peptide complexes of different size with explicit water. Data on molecular dynamics simulations used to compare the Gibbs energies were obtained with the RISM-AM and the other methods featured either with different terms (15)–(18) for the excess chemical potential or with the repulsive correction (20).

In addition, the KAEA16 and RLDL16 peptide complexes were treated with the classical RISM approach (see Tables 4 and 5). The free energy values of the snap-shot states on the corresponding MD trajectories were averaged, and then compared with data obtained by the RISM-AM on the same objects. The intramolecular correlation functions were calculated for the configurations selected on the corresponding MD trajectory framed with 0.2 ps intervals. The corresponding Fourier objects were minimized to keep up with at least 99% of the data volume (Tables 1 and 2). The free energies were calculated by averaging the solutions on the RISM equations over MD trajectory, and they were shown as having a confidence interval of 95%. The «Difference» term denotes the results changed upon usage of either the classical RISM approach or the RISM-AM to estimate the Gibbs free energy. Comparing the GF and PW functionals for the excess chemical potential shows they coincided to both methods of the RISM equations. At the same time, the other functionals gave significantly different results as they could strongly depend on the system complexity and size.

Table 4. Comparison of the RISM (95% confidence interval, C. I.) and the RISM-AM (pseudo) performance on the Gibbs energy calculated for the KAEA16 peptide complexes; in kcal/mol. The «Difference» values show the results changed upon usage of either the classical RISM or the RISM-AM approaches to estimate the Gibbs free energy. See text for details.

Complex	Mean	GF	PW	SC	TPT (R12)	TPT (WCA)
Monomer	Pseudo	−785.8	−584.5	−65.6	−595.6	−393.5
	95% C. I.	−819.9 ± 0.7	−602.3 ± 0.8	−40.2 ± 0.4	−575.2 ± 0.4	−371.9 ± 0.4
	Difference	34.1 (4%)	17.9 (3%)	−25.5 (63%)	−20.5 (4%)	−21.6 (6%)
Dimer	Pseudo	−1809.1	−1443.9	−63.7	−1125.5	−724.3
	95% C. I	−1880.4 ± 0.6	−1458.9 ± 0.6	18.3 ± 0.6	−1054.5 ± 0.6	−652.1 ± 0.6
	Difference	71.3 (4%)	15.0 (1%)	−82.0 (447%)	−71.0 (7%)	−72.3 (11%)
Tetramer	Pseudo	−4264.6	−3506.4	205.9	−2005.6	−1185.7
	95% C. I.	−4391.0 ± 0.6	−3459.5 ± 0.8	452.7 ± 1.0	−1768.5 ± 1.2	−952.9 ± 1.1
	Difference	126.4 (3%)	−46.9 (1%)	−246.8 (55%)	−237.2 (13%)	−232.8 (24%)
Octamer	Pseudo	−9845.4	−8054.7	1374.2	−3474.5	−1707.4
	95% C. I.	−10,228.8 ± 2.6	−7820.6 ± 2.0	2213.2 ± 5.6	−2719.0 ± 4.7	−947.4 ± 4.8
	Difference	383.4 (4%)	−234.1 (3%)	−839.0 (38%)	−755.5 (28%)	−760.0 (80%)

Table 5. Comparison of the RISM (95% C. I.) and the RISM-AM (pseudo) approaches on the Gibbs energy calculated for the RLDL16 peptide complexes; in kcal/mol. The «Difference» values show the results changed upon usage of either the classical RISM or the RISM-AM approaches to estimate the Gibbs free energy. See text for details.

Complex	Mean	GF	PW	SC	TPT (R12)	TPT (WCA)
Monomer	Pseudo	−1250.0	−1105.5	−257.4	−928.7	−673.0
	95% C. I.	−1295.5 ± 0.8	−1130.6 ± 0.8	−225.6 ± 0.4	−899.9 ± 0.4	−643.8 ± 0.4
	Difference	45.5 (4%)	25.2 (2%)	−31.8 (14%)	−28.8 (3%)	−29.2 (5%)
Dimer	Pseudo	−2960.6	−2687.7	−287.0	−1657.0	−1147.6
	95% C. I.	−3049.7 ± 0.5	−2669.6 ± 0.6	−129.8 ± 0.8	−1506.8 ± 0.9	−1000.0 ± 0.8
	Difference	89.1 (3%)	−18.1 (1%)	−157.2 (121%)	−150.2 (10%)	−147.5 (15%)
Tetramer	Pseudo	−6379.3	−5823.1	−366.6	−3233.1	−2172.0
	95% C. I.	−6620.3 ± 1.4	−5754.0 ± 1.0	77.7 ± 1.8	−2834.7 ± 1.8	−1770.7 ± 1.7
	Difference	241.0 (4%)	−69.0 (1%)	−444.3 (572%)	−398.4 (14%)	−401.2 (23%)
Octamer	Pseudo	−14,658.6	−13,271.6	538.9	−5909.1	−3563.1
	95% C. I.	−15,310.6 ± 2.4	−12,950.2 ± 1.9	1822.1 ± 4.3	−4814.7 ± 4.0	−2440.7 ± 3.9
	Difference	652.0 (4%)	−321.4 (2%)	−1283.2 (70%)	−1094.5 (23%)	−1122.3 (46%)

4. Discussion

There is small difference between the two approaches of the RISM, either the averaged over the MD trajectory and the RISM-AM, featured with the repulsive correction by the Lennard–Jones potential TPT(R12). It was previously noticed that the total changes in the Gibbs energy values estimated with RISM-AM did not exceed the limit seen due to temperature fluctuations. The ionic peptide complexes provided with an opportunity to gather an information about the objects of different size contained a spare charges. At the same time, the relative differences in the Gibbs free energy assessed by the RISM-AM exceeded the differences observed for the oxytocin. Nevertheless, the main trends (patterns?) of changes in the free energy during the formation of peptide complexes retained similar, no matter what the functionals were used by two versions of the RISM method.

These results support the authors intention to apply the average matrix RISM method to study the thermodynamics and the process of profilament complex formation by the ionic peptides. This method is prospective since allows for thermodynamic integration and analytical consideration of repulsive corrections. The RISM-AM used to estimate the Gibbs free energies of the RADA16, KADA16, and RAEA16 peptide complexes, allowed to assess the free energy values by taking into account the repulsive corrections using numerical thermodynamic integration. Further on, the corresponding data are notified as INT(R12) and INT(WCA).

In addition, the results of continual approaches used to account for the solution were compared with the performance of the RISM. The generalized Born and the solution of the Poison–Boltzmann equation were used to calculate the Gibbs free energy over the same 10 ns trajectories with an integration interval of 2 ps [20]. The free energy averaged values were obtained by different methods presented in Table 6. It can be seen the trend, the larger peptide complex, the lower free energy corresponded to the peptide monomer involved in the assembly. Truly remarkable is the difference found between the free energy values determined by GF, PW, and GB or PB approaches. Strangely, the terms of SC, TPT(R12), and TPT(WCA) led to very confusing results contradicted to the internal logic of the process of peptide self-assembly (protofilament formation). At the same time, the results demonstrated by the SC term previously revealed the free energy values were similar to the experimental data obtained for the interaction of DNA with a small dye molecule [11].

Table 6. Free energies of the peptide complexes calculated via the different methods. See text for details.

Structure *	GF	PW	SC	TPT (R12)	TPT (WCA)
8 × (KADA16) × 1	−5159.6	−3839.3	789.7	−3231.5	−1702.1
4 × (KADA16) × 2	−5656.5	−4397.6	910.9	−3103.6	−1584.9
2 × (KADA16) × 4	−6831.0	−5593.1	1423.5	−2752.1	−1195.1
1 × (KADA16) × 8	−8415.7	−6708.0	2798.9	−1890.3	−194.4
8 × (KAEA16) × 1	−6286.4	−4675.7	−525.1	−4765.1	−3148.1
4 × (KAEA16) × 2	−7236.2	−5775.5	−254.8	−4502.0	−2897.4
2 × (KAEA16) × 4	−8529.2	−7012.8	411.8	−4011.2	−2371.4
1 × (KAEA16) × 8	−9845.4	−8054.7	1374.2	−3474.5	−1707.4
8 ×(RADA16) × 1	−9612.5	−8257.2	−3538.4	−7713.1	−6108.7
4 ×(RADA16) × 2	−10,375.1	−9155.0	−3408.3	− 7589.5	−5995.0
2 ×(RADA16) × 4	−11,449.4	−10,182.4	−2947.5	−7283.2	−5637.5
1 ×(RADA16) × 8	−13,111.7	−11,440.5	−1554.6	−6361.3	−4589.2
8 × (RADA8) × 1	−4782.4	−3966.4	−1905.6	−4032.8	−3217.6
4 × (RADA8) × 2	−5299.2	−4547.6	−1744.4	−3856.4	−3054.4
2 × (RADA8) × 4	−5730.8	−4964.8	−1606.0	−3792.2	−2969.8
1 × (RADA8) × 8	−6358.5	−5510.8	−1190.9	−3551.1	−2674.6
8 × (RAEA16) × 1	−10,000.3	−8419.1	−4068.7	−8456.4	−6764.3
4 × (RAEA16) × 2	−11,259.5	−9850.3	−3667.2	−8067.1	−6396.7
2 × (RAEA16) × 4	−12,353.5	−10,889.1	−2957.1	−7537.1	−5817.1
1 × (RAEA16) × 8	−13,478.4	−11,871.8	−2160.9	−7088.1	−5260.9
8 × (RLDL16) × 1	−9999.8	−8843.7	−2058.9	−7429.7	−5384.0
4 ×(RLDL6) × 2	−11,842.5	−10,750.9	−1148.1	−6628.0	−4590.4
2 × (RLDL16) × 4	−12,758.6	−11,646.1	−733.3	−6466.1	−4343.9
1 × (RLDL16) × 8	−14,658.6	−13,271.6	538.9	−5909.1	−3563.1
8 × (KLEL8) × 1	−3744.4	−2997.6	39.1	−2690.6	−1663.2
4 × (KLEL8) × 2	−4472.9	−3660.1	217.3	−2569.9	−1531.5
2 × (KLEL8) × 4	−5286.4	−4555.0	619.7	−2252.3	−1198.8
1 × (KLEL8) × 8	−6166.3	−5247.1	1295.8	−1918.8	−761.3

Structure *	INT (R12)	INT (WCA)	PB	GB
8 × (KADA16) × 1	−2954.9	−3218.5	−1996.9	−1901.1
4 × (KADA16) × 2	−3077.7	−3325.4	−2066.2	−1985.4
2 × (KADA16) × 4	−3305.0	−3508.2	−2228.6	−2154.7
1 × (KADA16) × 8	−3389.8	−3500.3	−2293.9	−2288.3
8 × (KAEA16) × 1	−3957.0	−4252.4	−2995.7	−2862.8
4 × (KAEA16) × 2	−4287.2	−4529.3	−3211.1	−3118.5
2 × (KAEA16) × 4	−4486.3	−4675.7	−3400.3	−3305.0
1 × (KAEA16) × 8	−4591.2	−4713.6	−3449.2	−3391.7
8 × (RADA16) × 1	−7511.5	−7769.6	−6407.9	−6320.2
4 × (RADA16) × 2	−7759.5	−7986.6	−6555.3	−6493.6
2 × (RADA16) × 4	−7923.6	−8115.8	−6665.4	−6607.2
1 × (RADA16) × 8	−8080.9	−8177.9	−6797.2	−6797.1
8 × (RADA8) × 1	−3740.0	−3892.8	−3212.8	−3139.2
4 × (RADA8) × 2	−3883.2	−4003.6	−3251.2	−3251.2
2 × (RADA8) × 4	−3950.8	−4065.0	−3308.6	−3274.4
1 × (RADA8) × 8	−4033.7	−4115.9	−3348.5	−3337.8
8 × (RAEA16) × 1	−7822.8	−8104.7	−6642.7	−6493.4
4 × (RAEA16) × 2	−8199.6	−8408.4	−6939.2	−6836.7
2 × (RAEA16) × 4	−8368.4	−8522.1	−7108.0	−7000.0
1 × (RAEA16) × 8	−8398.3	−8505.6	−7053.6	−6980.0
8 × (RLDL16) × 1	−7718.5	−7977.4	−6256.0	−6147.0
4 × (RLDL6) × 2	−8081.1	−8250.8	−6560.2	−6518.4
2 × (RLDL16) × 4	−8131.6	−8289.5	−6618.9	−6577.1

Table 6. *Cont.*

Structure *	INT (R12)	INT (WCA)	PB	GB
1 × (RLDL16) × 8	−8212.9	− 8288.3	−6767.0	−6761.1
8 × (KLEL8) × 1	−2553.9	−2705.5	−1917.0	−1828.7
4 × (KLEL8) × 2	−2619.8	−2762.2	−1948.0	−1878.2
2 × (KLEL8) × 4	−2783.3	−2880.0	−2120.2	−2058.7
1 × (KLEL8) × 8	−2789.5	−2848.5	−2145.5	−2107.1

* equivalent sum of the Gibbs free energy values for the peptide complexes of size (N × (ΔG)k × M), where N—the number of units, containing as many (M) individual peptides as the octamer does, i.e., N = 1 and M = 8 correspond to the octamer; k is an average ΔG of individual peptide in the M-size complex like (ΔG)k = $(\Delta G)_{\mathrm{cmpl}}$ /MN.

What seems to be important is the method of thermodynamic integration empowered with the repulsion correction, allows to assess the free energy values comparable to the results obtained by the methods based on continual models of solution [10].

Also, the free energies of interacting parts in the octamer peptide complex were estimated by different methods (Table 7). The results, supplied with uniform logic of the process of peptides self-assembling, were gathered with GF and PW terms or thermodynamic integration supplied with the repulsion correction by the Lennard–Jones potential—INT(R12). Curiously enough, the usage of the SC, TPT(R12), and TPT(WCA) terms yielded data seeming to display inverted energy trends on the path of growing protofilament formation. The only explanation we have on that fact is that the abovementioned terms for the repulsion corrections may poorly account for the non-polar interactions and thusly undermine the total effects of the interacting charges.

Table 7. Binding energies of the 16-mer peptide tetramers bound into the octamer complex. See text for details.

Peptide	KADA16	KAEA16	RADA16	RAEA16	RLDL16
GF	−1583.3	−1316.2	−1662.3	−1124.9	−1900.0
PW	−1114.9	−1041.9	−1258.1	−982.7	−1625.5
SC	1375.4	962.4	1392.9	796.2	1272.2
TPT(R12)	861.8	536.7	921.9	449.0	557.0
TPT(WCA)	1000.7	664.0	1048.3	556.2	780.8
INT(R12)	−84.8	−104.9	−157.3	−29.9	−81.3
INT(WCA)	7.9	−37.9	−62.1	16.5	0.8
PB	−65.3	−48.9	131.8	54.4	−148.1
GB	−133.6	−86.7	189.9	20.0	−184.0

More sensible results obtained with the thermodynamic integration supplied with the Lennard–Jones repulsion correction term (21), are notified as INT(R12) in Table 7. Importantly, the thermodynamic integration supplied with the WCA repulsion correction part (22) and marked as INT(WCA) in Table 7, fails to provide the growing protofilament structure via self-assembling process. Since the integration with the repulsion correction by Lennard–Johns term (21) yields better results, the latter points at yet-unidentified features of the WCA term applied to flexible molecular structures with spare charges (22).

5. Conclusions

The results of comparing the RISM-AM and the standard RISM approaches suggest that the mean (7) and averaging method (11) are both applicable to study rigid molecular structures, and ionic peptide complexes with multiple states of freedom as well. Pseudo-mean correlation functions were found coincided well with the actual means at the level of accuracy, typical to the statistical-error-accompanied stage of averaging in the standard RISM.

Consequently, the corresponding thermodynamics obtained with pseudo-average functions differed from the real mean values by no more than the averaging error. At the same time, the costs of computing the pseudo-mean values were found of being approx-

imately less by two orders of magnitude, than the actual means averaged through the intermediate values along the MD trajectory.

Even when the molecular trajectory is non-static, i.e., the molecule undergoes considerable conformational transitions, the RISM-AM method is still applicable, but only after the correct quasi-stationary region located and selected on the calculated MD trajectory. Such a move eliminates the precondition of working with pseudo-mean correlation functions. Generally, the RISM-AM approach appears to be much faster than conventional RISM equations implemented on the same computing platform of hardware.

Since the RISM equation with an average matrix is fast, it can resolve the problem of long lasted session of calculations, when the standard RISM used on moderate hardware. For instance, when using the pseudo-mean correlation functions, the integration of thermodynamics for given solvating temperature dependences was carried out using Big Data arrays. However, it should be kept in mind that the difference between the actual mean and pseudo-mean RISM values for the particular system may be significantly bigger than the dispersion interval.

The question still remains about how consistent these relationships are for both pseudo-means and actual means. This issue is likely to be independently assessed and considered on a case-by-case basis.

Author Contributions: Conceptualization, D.T.; methodology, D.T.; software, D.T.; validation, A.D.; data curation, A.D.; writing, A.D.; writing—review and editing, D.T. All authors have read and agreed to the published version of the manuscript.

Funding: This study was partially supported by the program "Nonclinical study of biomedical technologies and pharmaceuticals" (Federal registration no. AAAA-A19-119021590083-3, Rider: FMFM-2019-0036) and the programs 075-00381-21-00 and 0017-2019-0009, Russia.

Data Availability Statement: More details on the software used to perform the calculations are available upon the request by e-mail.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Onufriev, A.; Bashford, D.; Case, D.A. Modification of the generalized Born model suitable for macromolecules. *J. Phys. Chem. B* **2000**, *104*, 3712–3720. [CrossRef]
2. Chandler, D.; Andersen, H.C. Optimized cluster expansions for classical fluids. II. Theory of molecular liquids. *J. Chem. Phys.* **1972**, *57*, 1930–1937. [CrossRef]
3. Hirata, F.; Rossky, P.J.; Pettitt, B.M. The interionic potential of mean force in a molecular polar solvent from an extended RISM equation. *J. Chem. Phys.* **1983**, *78*, 4133–4144. [CrossRef]
4. Sugita, M.; Onishi, I.; Irisa, M.; Yoshida, N.; Hirata, F. Molecular recognition and self-organization in life phenomena studied by a statistical mechanics of molecular liquids, the RISM/3D-RISM theory. *Molecules* **2021**, *26*, 271. [CrossRef]
5. Morita, T.; Hiroike, K. A new approach to the theory of classical fields. I. *Prog. Theor. Phys.* **1960**, *23*, 1003–1027. [CrossRef]
6. Lue, L.; Blankschtein, D. Liquid-state theory of hydrocarbon-water systems: Application to methane, ethane, and propane. *J. Phys. Chem.* **1992**, *96*, 8582–8594. [CrossRef]
7. Kovalenko, A.; Hirata, F. Hydration free energy of hydrophobic solutes studied by a reference interaction site model with a repulsive bridge correction and a thermodynamic perturbation method. *J. Chem. Phys.* **2000**, *113*, 2793–2805. [CrossRef]
8. van Leeuwen, J.M.J.; Groeneveld, J.; de Boer, J. New method for the calculation of the pair correlation function. I. *Physica* **1959**, *25*, 792–808. [CrossRef]
9. Kovalenko, A.; Hirata, F. Self-consistent description of a metal–water interface by the Kohn–Sham density functional theory and the three-dimensional reference interaction site model. *J. Chem. Phys.* **1999**, *110*, 10095–10112. [CrossRef]
10. Tikhonov, D.A.; Polozov, R.V.; Timoshenko, E.G.; Kuznetsov, Y.A.; Gorelov, A.V.; Dawson, K.A. Hydration of a B–DNA fragment in the method of atom–atom correlation functions with the reference interaction site model approximation. *J. Chem. Phys.* **1998**, *109*, 1528–1539. [CrossRef]
11. Singer, S.J.; Chandler, D. Free energy functions in the extended RISM approximation. *Mol. Phys.* **1985**, *55*, 621–625. [CrossRef]
12. Chandler, D.; Singh, Y.; Richardson, D.M. Excess electrons in simple fluids. I. General equilibrium theory for classical hard sphere solvents. *J. Chem. Phys.* **1984**, *81*, 1975–1982. [CrossRef]
13. Zichi, D.A.; Rossky, P.J. Molecular conformational equilibria in liquids. *J. Chem. Phys.* **1986**, *84*, 1712–1723. [CrossRef]
14. Ten-No, S. Free energy of solvation for the reference interaction site model: Critical comparison of expressions. *J. Chem. Phys.* **2001**, *115*, 3724–3731. [CrossRef]

15. Ten-No, S.; Iwata, S. On the connection between the reference interaction site model integral equation theory and the partial wave expansion of the molecular Ornstein–Zernike equation. *J. Chem. Phys.* **1999**, *111*, 4865–4868. [CrossRef]
16. Zhang, S. Fabrication of novel biomaterials through molecular self-assembly. *Nat. Biotechnol.* **2003**, *21*, 1171–1178. [CrossRef]
17. Hsieh, P.C.; Davis, M.E.; Gannon, J.; MacGillivray, C.; Lee, R.T. Controlled delivery of PDGF-BB for myocardial protection using injectable self-assembling peptide nanofibers. *J. Clin. Investig.* **2005**, *116*, 237–248. [CrossRef]
18. Ellis-Behnke, R.G.; Liang, Y.-X.; You, S.-W.; Tay, D.K.C.; Zhang, S.; So, K.-F.; Schneider, G.E. Nano neuro knitting: Peptide nanofiber scaffold for brain repair and axon regeneration with functional return of vision. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 5054–5059. [CrossRef]
19. *HyperChem®: Computational Chemistry*; Hypercube Inc.: Gainesville, FL, USA, 2002.
20. Macindoe, G.; Mavridis, L.; Venkatraman, V.; Devignes, M.-D.; Ritchie, D.W. HexServer: An FFT-based protein docking server powered by graphics processors. *Nucleic Acids Res.* **2010**, *38*, W445–W449. [CrossRef]
21. Tovchigrechko, A.; Vakser, I.A. Development and testing of an automated approach to protein docking. *Proteins Struct. Funct. Bioinform.* **2005**, *60*, 296–301. [CrossRef]
22. Case, D.A.; Cheatham, T.E., III; Darden, T.; Gohlke, H.; Luo, R.; Merz, K.M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R.J. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26*, 1668–1688. [CrossRef] [PubMed]
23. Jorgensen, W.L.; Chandrasekhar, J.; Madura, J.D.; Impey, R.W.; Klein, M.L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935. [CrossRef]
24. Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M.C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; et al. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **2003**, *24*, 1999–2012. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.