# Analysis of Apoflavodoxin Folding Behavior with Elastic Network Models

M. Cotallo-Abán[*], D. Prada-Gracia[†,*], J.J. Mazo[†,*], P. Bruscolini[*], F. Falo[†,*] and J. Sancho[**,*]

[*]*Instituto de Biocomputación y Física de Sistemas Complejos. Universidad de Zaragoza. SPAIN*
[†]*Dpto. Física de la Materia Condensada. Universidad de Zaragoza. SPAIN*
[**]*Dpto. Bioquímica y Biología Molecular y Celular. Universidad de Zaragoza. SPAIN*

**Abstract.**
We apply simple elastic network models to study some properties of the unfolding of apoflavodoxin, a protein that shows a three-state thermodynamic behavior under thermal denaturation, as revealed by extensive analysis of wildtype and mutant variants. The intermediate of apoflavodoxin presents an overall structured core, with just a part of the protein being substantially unfolded [1]. In agreement with these results, we have been able to identify, using different models and methods, the more mobile regions in the thermal unfolding of the protein. We also discuss how the predictions obtained from these models could help in designing new experiments.

**Keywords:** protein, gaussian network model, statistical mechanics simple models
**PACS:** 87.15.Aa,64.60.Cn

## THE APOFLAVODOXIN FROM ANABAENA

Flavodoxin is a 169 residue-long protein involved in electron transfer processes in Anabaena PCC 7119 and many other procaryots. Its "apo" form (1FTG), which lacks the FMN cofactor, shows a three-state thermodynamic equilibrium behavior under thermal denaturation [2, 1]. In the intermediate, a large part of the protein remains close to the native fold, but there is a non-contiguous 40-residue region which appears unfolded.

Experiments suggest that the apoflavodoxin thermal intermediate, which appears at 317.3 K, is mainly formed by the packing of helices and $\beta$ strands. In contrast there are three loops quite weakened. The most significative regions are loops 57-60, 90-100, that bind the FMN cofactor and 120-139, which contains a three-stranded $\beta$-sheet [2, 1].

These results, were obtained by performing equilibrium $\phi$-analysis [1], which does not allow to "see" directly the structure of the intermediate state, which, due to the intrinsic technical difficulties, has not been crystallized, nor characterized by NMR. The hypothetical structure of the intermediate has been deduced thanks to the interpretation of the changes in the relevant and residual stability of the protein, which allows to assess if a mutation changing the stability of the native state also affects the stability of the intermediate state, with respect to the unfolded state.

For this reason, in order to get a better insight on the behavior of the system, it can be useful to obtain results using different theoretical models, see if such results support the standard interpretation, and look for suggestions of new experiments that can further confirm this view.

CP851, *From Physics to Biology; BIFI 2006 II International Congress,*
edited by J. Clemente-Gallardo, Y. Moreno, J. F. Sáenz Lorenzo, and A. Velázquez-Campoy

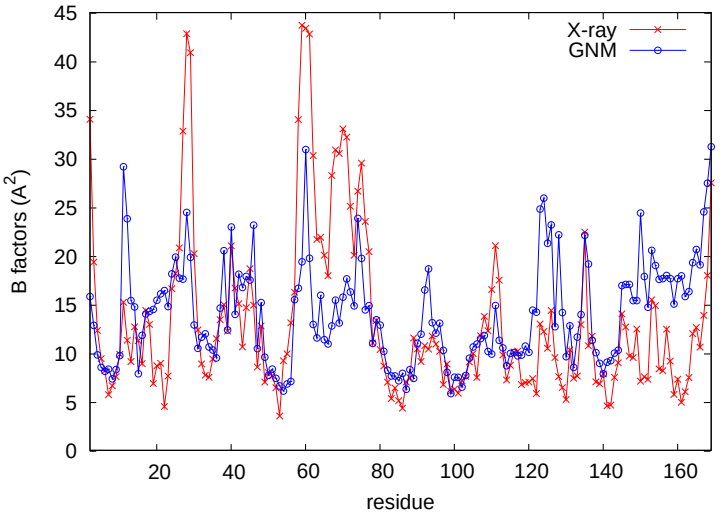


**FIGURE 1.** Bfactors calculated from the GNM (scaled with a factor) versus the experimental Bfactors

The simplest model we have used is the so-called "Gaussian Network Model" (GNM) which only considers the topology of the native state. We start studying its standard version [3, 4, 5], which allows us to predict the active regions of the protein (that coincide with the most mobile ones). Moreover, conformational motions of the native state, as well as coupled movements between regions, can be detected. Then the characteristics of the dynamics of the relevant protein regions is understood.

Next, we study an extended version of the GNM, introduced by Micheletti and coworkers [6, 7]. This version allows for breaking of native contacts, thus driving the protein to thermal unfolding. We deal with this model in two very different ways: through a self-consistent approximation (an analytical method assisted by numerical calculation) and by molecular dynamics simulations using a Langevin bath (numerical simulation of the dynamics of the protein).

Finally, we compare the results with all-atom simulations of the protein dynamics at two different bath temperatures. We present a discussion of our preliminary results in the light of their capability to support the three-state model for apoflavodoxin thermal unfolding.

## MODELS: TWO STATISTICAL-MECHANICS APPROACHES

### A.- Gaussian Network Model

As a first approach, we have studied the behaviour of protein as an elastic network [4]. Recently, this kind of models have been developed to obtain information about the mechanical properties of the native conformation [3, 4, 5].This model will give us

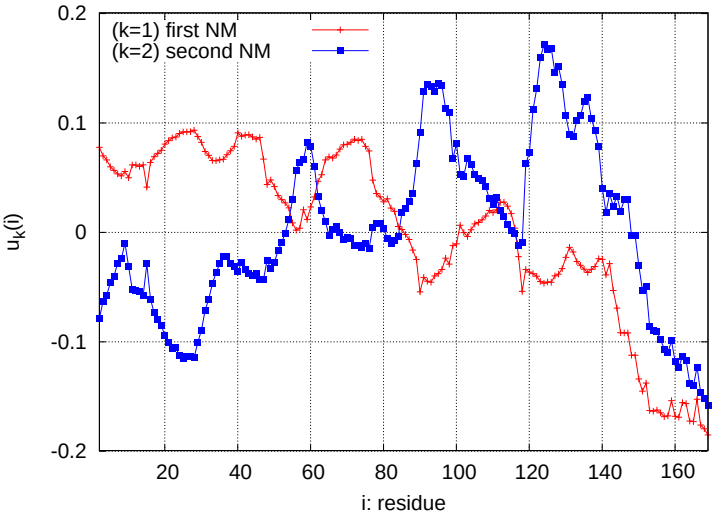


**FIGURE 2.** First and second normal mode

insight on the elastic properties of the crystallized structure. The approximation is as follows: the protein is reduced to a set of nodes linked with their neighbours which are at a distance less than a given cutoff ($r_c$). In this approximation, it is straightforward the calculation of the normal modes, cooperative motions, and the correlation of the fluctuation of the nodes. The more contribution of a set of nodes in the lowest frequency modes, the more flexible is the region. Here, we will focus on the two low frequency modes. Our interest is on the flexibility of the loops that will probably lead to the intermediate state.

The Kirchoff matrix,$\Gamma$, obtained from the crystallized structure, is build as follows: $\Gamma_{i,j} = -1$ if $i \neq j$ and $R_{i,j} \leq r_c$, 0 if $i \neq j$ and $R_{i,j} > r_c$ and $-\sum_{i,j \neq i} \Gamma_{i,j}$ if $i = j$. We define $R_i$ as the position of the $C_\alpha$ atom of residue $i$ and $R_{i,j}$ represents the distance between $C_\alpha$'s i and j in the crystallized structure.

From this matrix the correlation of the fluctuations around minima can be extracted:

$$\Gamma = U \Lambda U^T \qquad \Gamma^{-1} = \sum_{k=2}^{N} \lambda_k^{-1} \vec{u}_k \vec{u}_k^T \qquad \left\langle \Delta R_i \cdot \Delta R_j \right\rangle = \frac{K_B T}{\gamma} \Gamma_{i,j}^{-1} \,, \tag{1}$$

being $\vec{u}_k$ the k-th column of U (that is, k-th eigenvector of $\Gamma$). This column is proportional to the k-th normal mode of the system. $\Lambda$ is the diagonal matrix of eigenvalues $\lambda_k$, where:

$$\lambda_1 = 0 < \lambda_2 < \ldots < \lambda_n \,. \tag{2}$$

Mean square deviations of each $C_\alpha$ in the model and the experimental Debye-Waller ($B_i$) factors measured by X-ray difraction are related with the equation

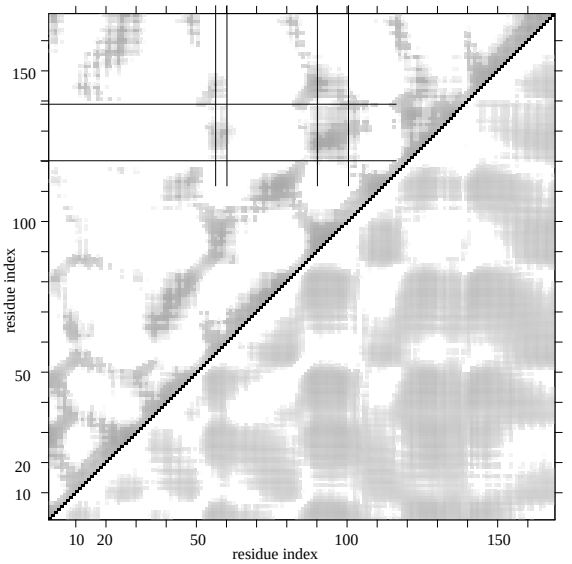


**FIGURE 3.** The values of $\langle \Delta R_i \cdot \Delta R_j \rangle / \langle \Delta R_i^2 \rangle^{1/2} \langle \Delta R_j^2 \rangle^{1/2}$ extracted from the GNM. The upper diagonal pixels represents the positive values, the darker the more correlated residues. Below the diagonal the negative values are represented.

$$B_i^{GNM} = \frac{8\pi^2}{3} \langle \Delta R_i \cdot \Delta R_i \rangle = \frac{8\pi^2 K_B T}{3\gamma} \Gamma_{i,i}^{-1} \,, \tag{3}$$

where $\gamma$ is a scale parameter to be fitted. The other free parameter is $r_c$. Best fit (see Figure 1) has been achieved with $r_c = 7$ Å which is a value in agreement with those ones used in literature [4].

Figure 2 plots the first and the second modes, $\vec{u}_1$ and $\vec{u}_2$. The first one corresponds to a breathing motion, where half of the structure move against the other half. This mode does not present interesting structural features. The zones with less contribution are in the core or are playing the role of "elbow" in hinge motions. The second mode is more interesting because four relevant regions can be identified. The three loops (57-60, 90-100, 120-139) are correlated between them and uncorrelated with the region around 20-30, taking part in this way on a collective motion which represent an important part of the total fluctuation.

Normalized correlations are shown in Figure 3. A high correlation can be observed between the regions of the protein we are interested. In fact, we can see in the figure how the 120-139 loop is correlated with the loops 57-60 and 90-100. This corroborates the observations given in the second normal mode.

In conclusion, we see how this model, based in the protein topology, reveals a cooperativity between regions which can take part in the thermal intermediate. This approach allows us, as well, to do the normal modes analysis in a less time consuming way than using an all-atom simulation program. These normal modes deliver important informa-

tion about long range cooperative effects. In our case, these results are a good starting point to check the hypothetical intermediate structure deduced from the experimental work with theoretical models.

## B.- Extended GNM with contact breaking: self-consistent approximation

Now, we introduce the possibility to break contacts, and study the effect of temperature on the loops unfolding. To this end, we consider the Hamiltonian [6, 7]:

$$\tilde{\mathscr{H}} = \frac{T}{2} K \sum_{i=1}^{N-1} (\vec{r}_{i,i+1} - \vec{r}^{\,0}_{i,i+1})^2 + \frac{1}{2} \sum_{i \neq j} \varepsilon_{i,j} \Delta_{i,j} \left[ (\vec{r}_{i,j} - \vec{r}^{\,0}_{i,j})^2 - R^2 \right] \theta_{i,j} \,, \tag{4}$$

where $\theta_{i,j} = \Theta\Big(R^2 - (\vec{r}_{i,j} - \vec{r}^{\,0}_{i,j})^2\Big)$ (the Heaviside function), $\vec{r}_{i,j} = \vec{r}_i - \vec{r}_j$ and $\vec{r}^{\,0}_{i,j}$ the same for the native conformation. $\Delta_{i,j}$ is the native contact map for the protein ($\Delta_{i,j} = 1$ if residues i and j are in contact in the native structure (in our case: at a distance less than 6.5 Å), $\Delta_{i,j} = 0$ otherwise).

As a first approach, we resort to a self-consistent approximation to evaluate analytically the partition function and all relevant thermodynamical quantities.

Upon replacing $\theta_{i,j}$ by a parameter $p_{i,j}$ the Hamiltonian turns into a sum of quadratic terms, which allows analytical integration of the partition function:

$$\begin{aligned} Z(T) &= \exp\left[\frac{R^2}{2T} \sum_{i,j} \Delta_{i,j} p_{i,j}\right] \int \prod_{i=1}^{N} d^3 r_i \exp\left[-\frac{1}{2} \sum_{i,j} \vec{x}_i M_{i,j} \vec{x}_j\right] \\ &= N^{-\frac{2}{3}} (2\pi)^{\frac{3(N-1)}{2}} (\det{}' A)^{-\frac{3}{2}} \,, \end{aligned} \tag{5}$$

where (') means that we are calculating the determinant of the matrix without considering the eigenvalue $\lambda_1 = 0$ (that is, $\det' A = \prod_{i=2}^{N} \lambda_i$).

From the expression above, all the relevant thermodynamical quantities can be evaluated: we have, for the average energy:

$$\langle E \rangle = \frac{3(N-1)T}{2} - \frac{R^2}{2} \sum_{i,j} \Delta_{i,j} p_{i,j}(T) \,, \tag{6}$$

and the average number of native contacts or "native-state-overlap":

$$Q = \frac{\sum_{i<j} \Delta_{i,j} p_{i,j}}{\sum_{i<j} \Delta_{i,j}} \,. \tag{7}$$

The parameter $p_{i,j}$ must be evaluated self-consistently, that is:

$$p_{i,j} = \left\langle \Theta\Big(R^2 - (\vec{r}_{i,j} - \vec{r}^{\,0}_{i,j})^2\Big) \right\rangle_0 \,, \tag{8}$$

where the average on the right depends on $p_{i,j}$. This average can be calculated by the integral:

$$p_{i,j} = \frac{1}{Z}\int \prod_{i=1}^{N} d^3 r_i \exp\left(-\beta H_0\right)\Theta(R^2 - (\vec{r}_{i,j} - \vec{r}^{\,0}_{i,j})^2)\;, \tag{9}$$

where $\beta = 1/T$ and $H_0$ is obtained from Eq. (4) upon substituting the functions $\theta_{i,j}$ with the parameters $p_{i,j}$.

By defining $\vec{x}_i = \vec{r}_i - \vec{r}^{\,0}_i$, this expression results in:

$$p_{i,j} = \frac{1}{Z}\int \prod_{i=1}^{N} d^3 r_i \Theta(R^2 - (\vec{x}_i - \vec{x}_j)^2) \exp\left[\frac{R^2}{2T}\sum_{i,j}\Delta_{i,j}\right] \exp\left[-\frac{1}{2}\sum_{i,j}\vec{x}_i M_{i,j}\vec{x}_j\right]\;, \tag{10}$$

with the M the matrix defined by

$$\begin{aligned}(M)_{i,j} &= \delta_{i,j}\left(K(2-\delta_{i,1}-\delta_{i,N}) + \frac{2}{T}\sum_l \Delta_{i,l}p_{i,l}\right) + \\ &+ (1-\delta_{i,j})\left(-K(\delta_{j,i+1}+\delta_{j,i-1}) - \frac{2}{T}\Delta_{i,j}p_{i,j}\right)\;.\end{aligned} \tag{11}$$

The integral can be calculated by a Laplace Transform. By using $R^2$ as the variable t, we need to calculate:

$$\mathscr{L}(f(t)) = \mathscr{L}\left[\int \prod_{i=1}^{N} d^3 x_i\, \Theta(t - (\vec{x}_i - \vec{x}_j)^2) exp\left[-\frac{1}{2}\sum_{i,j}\vec{x}_i M_{i,j}\vec{x}_j\right]\right]\;, \tag{12}$$

and developing the calculation:

$$\mathscr{L}(f(t)) = \frac{1}{s}\int \prod_{i=1}^{N} d^3 x_i exp\left[-\frac{1}{2}\sum_{i,j}\vec{x}_i Q_{i,j}\vec{x}_j\right]\;, \tag{13}$$

with the new matrix $Q_{k,l} = M_{k,l} + 2(\delta_{k,l}(\delta_{k,i}+\delta_{k,j}) - \delta_{k,i}\delta_{l,j} - \delta_{l,i}\delta_{k,j})s$.

Now the integral is reduced to the same form as in the partition function, above. The problem with matrix $Q$ and $M$ is that, due to the translational invariance $\vec{x}_i \rightarrow \vec{x}_i + \vec{a}$ of the problem, the sum of the elements on each of their rows and columns is zero: this fact yields a null determinant, producing a singularity in the evaluation of the averages.

We can follow McCammon and coworkers [8] to solve the problem by considering and extra spring on the terminal residues $N$: we add an extra term $\frac{1}{2}\gamma\vec{x}_N^2$ to the hamiltonian, that preserves the structure of the quadratic form of the hamiltonian, and hence the form of the matrix $M$, but removes the translational invariance. The extra contribution to the free energy can be explicitly calculated.

With this approach, after performing the gaussian integral and the inverse Laplace transform, and restoring the variable $R$, the explicit expression of the $p_{i,j}$ reads:

$$p_{i,j} = -\frac{2}{\sqrt{\pi}}\rho_{i,j}e^{-\rho_{i,j}^2} + \mathrm{Erf}(\rho_{i,j}) \ , \tag{14}$$

with:

$$\rho_{i,j} = \frac{R}{\sqrt{2\gamma_{i,j}}} \qquad \gamma_{i,j} = ((M_N^N)^{-1})_{i,i} + ((M_N^N)^{-1})_{j,j} - 2((M_N^N)^{-1})_{i,j} \ , \tag{15}$$

where $(M_N^N)^{-1})_{i,j}$ is the $(i,j)$-element of the inverse of the matrix obtained from $M$ upon elimination of the Nth row and column. At any temperature $p_{i,j}$ will be found iteratively from the above expression, as described in [6].

We follow [6] in the choice of the parameters: $R = 3$, $K = 1/15$. The native contact probability per residue $Q_i$ are defined:

$$Q_i = \frac{\sum_j \Delta_{ij} p_{ij}}{\sum_j \Delta_{ij}} \ . \tag{16}$$

In this study, we first set $\varepsilon_{ij} = \varepsilon$ for each $i$ and $j$, and focus just on the geometry of the apoflavodoxin native state: we aim at understanding how much the folding geometry is responsible for the thermodynamics of the folding process, and how "trivial" the resulting thermodynamic behavior appears, as compared to that of a random contact-map.

To this end, we study the behavior of the model with several different contact maps: the original, wild-type one, and many others, obtained by random reshuffling of each residue's contacts, in such a way that the connectivity of each residue (i.e., the number of contacts it makes) is preserved, but the resultant geometry is completely random. This most likely produces non-physical contact-maps, that violate geometric constraints, but due to the nature of the model, where just deviations $\vec{x}_j$ from the equilibrium position are relevant, the value of the thermodynamical information coming from such maps is not affected, and we are allowed to compare the thermodynamic behavior of the different cases.

Then, we perturb the models introducing a little differentiation among contacts, choosing a subset and making them weaker, with their energy being $0.1\varepsilon$: we want to test in this way how the introduction of a "sequence" changes the results obtained for the homogeneous case.

We choose the contacts to be weakened with three different patterns:

1. all contacts of residues in the region 94 - 124 (corresponding to a helix placed in the surface of the native wild-type protein)
2. randomly chosen contacts (irrespective of the residues involved in the contact)
3. all contacts of residues in the region 57-60, 90-100, 120-139, that correspond to the unfolded part of the thermal intermediate, according to experiments.

In the first case, there were 263 contacts, in the second one, 271 and in the last one, 243 contacts in real maps, and 285 contacts in the first case, 271 contacts in the second one and 291 contacts in the last one for the reshuffled maps, out of a total of 649 contacts

in all cases. We performed the study with several samples of random contacts maps, obtaining similar results: those reported in the following are typical results.

In Figure 4, 5 and 6 we observe energy, specific heath and average number of native contacts (Eq. (7))for different groups of results.

Several comments are in order:

1. We cannot find important differences in the thermodynamic behavior obtained with the original or the reshuffled map, in the homogeneous case of same $\varepsilon$. Even if a quantitative van t'Hoff analysis cannot be carried on in this case, due to the nature of the self-consistent approximation, if we still consider the ratio of the height by the width of the peak of specific heat as a reasonable measure of cooperativity, we see that the reshuffled and original map show practically the same degree of cooperativity.
2. Things do not change if we weaken the same number of randomly-chosen contact in the wild type and reshuffled geometry. As expected, the peak moves towards lower temperatures (due to overall decrease of stability of the folded conformation) and is somewhat shorter, but native and random geometry still produce almost identical traces in all figures.
3. In the above two cases, a single peak is present in the specific heat, suggesting a two-state behavior. This agrees with the simulations performed for the wild type geometry in the homogeneous case (see following section). Things start changing when we weaken all the contacts pertaining to a group of residues: we can see that the specific heat traces of wild type and random geometry becomes increasingly different as we move to weakening the region of the superficial helix 94-124 and then to weakening the three experimentally relevant regions 57-60, 90-100, 120-139. Moreover, the specific heats obtained with wild-type geometry starts showing a small shoulder at low temperatures in the "helix" case, that becomes a clear peak when the three different regions are weakened.

Thus, it seems that according to the geometry of apoflavodoxin, the model suggests a two-state behavior. A three-state behavior only appears when we introduce different energies for the contacts, roughly mimicking the energetic heterogeneity involved by the sequence, and arrange the weak contacts in such a way that all the contacts of a group of residues are weak. This supports the view that the specific sequence of wild type apoflavodoxin has a central role for the existence of an equilibrium intermediate, a view which agrees with the observation that the interface, between the three experimentally-determined unstable regions and the bulk of the protein, is unusually polar.

In Figure 7 we can see, for different temperatures, the average number of contacts of residue $i$ that are still formed, for each $i$ (Eq. (16)), for the homogeneous wild-type case. We can observe that, also at the transition temperature, when on average half of the contact are formed, the fluctuation around the average are not very pronounced, due to the "mean-field" nature of the self-consistent approximation. However, it is possible to notice that the experimentally relevant regions correspond indeed to regions that are predicted to be highly unfolded (low $Q_i$). This is especially true for loops 57 - 60 and 120 - 139, while loop 90 - 100 is not as well represented. It should be noticed, though, that these results are not sufficient to perform a safe prediction, on their basis, of the structure

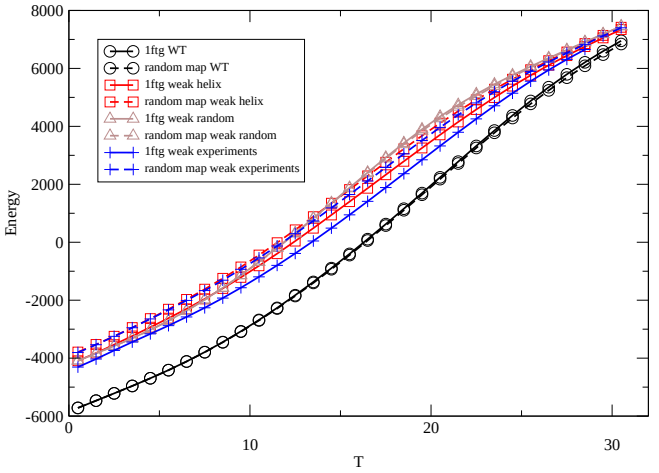


**FIGURE 4.** E versus T for different weakening patterns. In the legend, "weak helix", "weak random" and "weak experiments" indicate that the choice of the weak contacts is made according to pattern 1, 2 or 3 in the text, respectively.

of the intermediate. This could be expected, in the light of the above results, since we have seen that the introduction of energetic heterogeneity is necessary to reproduce three-state behavior.

## C.- Extended GNM with contact breaking: Langevin dynamics

In this section, we study the Langevin dynamics of the extended GNM model described by equation (4). This study will allow us to validate the results previously obtained and to explore the space of configurations in a more realistic way. Within other important things, the dynamics will show us the manner the contacts break when the system is embedded in a stochastic thermal reservoir at temperature $T$.

The equations of the motion read as:

$$\dot{q}_i(t) = p_{q,i}/m \quad ; \quad \dot{p}_{q,i}(t) = -\frac{\partial H}{\partial q_i} - \gamma p_{q,i}(t) + \eta_{q,i}(t) \, , \tag{17}$$

where $i$ is the $C_\alpha$ index and $q = x, y, z$.

In these equations, we control the temperature trough the stochastic term $\eta_{q,i}(t)$ which is a Gaussian distributed $\delta$-correlated random noise:

$$\langle \eta_{q,i}(t) \eta_{q,j}(0) \rangle = 2\gamma m k_B T \delta(t) \delta_{i,j} \, , \tag{18}$$

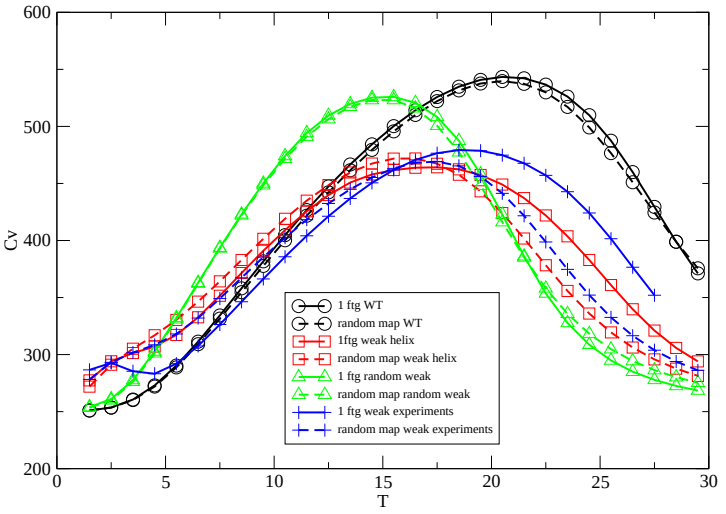


**FIGURE 5.** $C_v$ versus T for different weakening patterns. In the legend, "weak helix", "random weak" and "weak experiments" indicate that the choice of the weak contacts is made according to pattern 1, 2 or 3 in the text, respectively.

where, $k_B$ denotes the Boltzmann constant and $T$ is the temperature (for the sake of simplicity, both $k_B$ and $m$ have been set to unity).

Along a trajectory, we have computed the values of some of the variables previously defined: the native-state-overlap $Q(T)$ (Eq. (7)), or the native contact probability per residue $Q_i(T)$(Eq. (16)) for instance.

In Figure 8, we can see the value of $Q_i$ for each residue in the protein. Once more, it is shown that the regions more amenable to broke their contacts are those which are found in experiments; that is, loops 57 - 60, 90 - 100 and 120-139.

One should note that temperature for unfolding obtained in Langevin simulation is much lower than that calculated in the self-consistent method. This is due to the fact that, since self-consistency introduce a kind of mean field approximation, it can not take into account the large fluctuations observed in the simulations. Indeed, within self-consistent approximation the contacts are never actually broken, except in the case when the corresponding $p_{ij}$ is strictly zero: therefore, also in the denaturated conformations, there is a small quadratic bias toward the native conformation. This results in the need of a higher temperature to produce the same degree of unfolding that we get in simulations of the original model.

Hovewer, if we consider the results at the respective transition temperatures, we can establish a correlation between the results found with these simulations and with self-

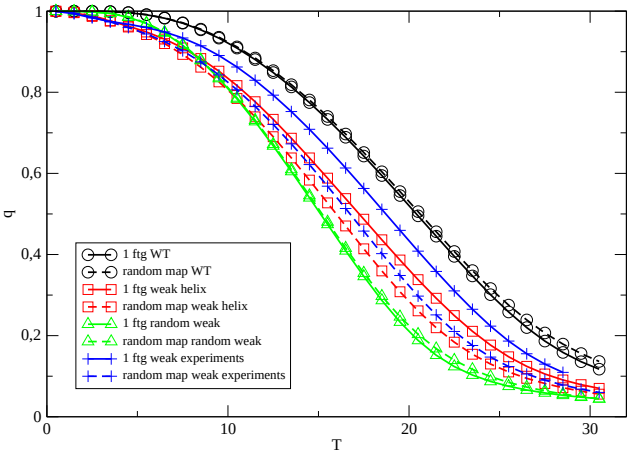


**FIGURE 6.** q versus T for different weakening patterns. In the legend, "weak helix", "random weak" and "weak experiments" indicate that the choice of the weak contacts is made according to pattern 1, 2 or 3 in the text, respectively.

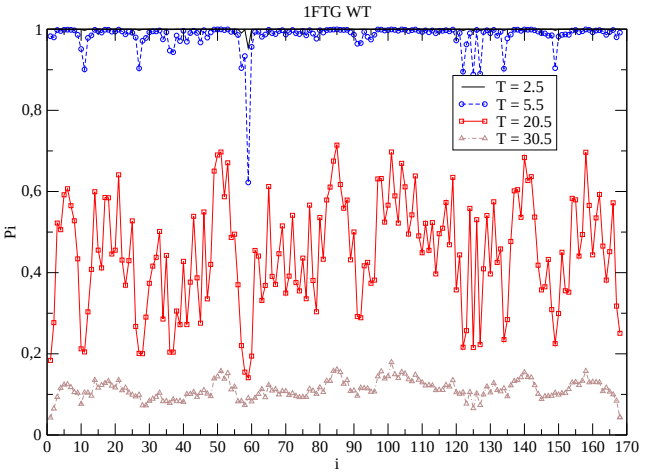


**FIGURE 7.** $Q_i$ from the WT protein for different temperatures.

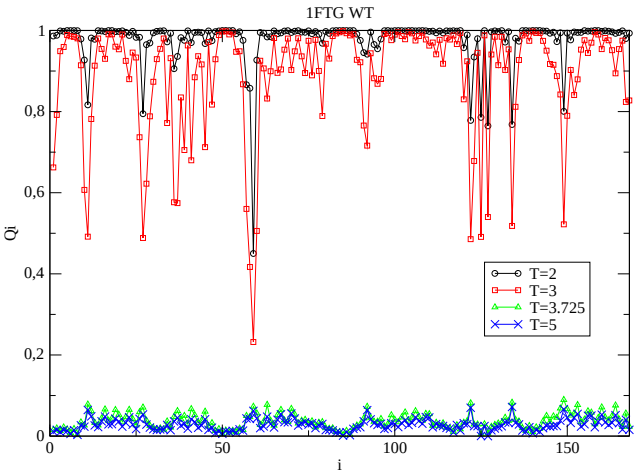


**FIGURE 8.** T1 = 1, with Q = 0.977133; T2 = 2, with Q = 0.890280; T3= 3.725, with Q = 0.037224; T4 = 5, with Q = 0.028062.Once more, we can see that loops, especially 57-60, are more prone to break their contacts.

consistent approximation made in previous subsection.

In order to find it, we use Pearson correlation coefficient. If we have two sets of N data, $x_i$ and $y_i$ , then the Pearson correlation coefficient is calculated as:

$$r_{x,y} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{(N-1)s_x s_y} , \tag{19}$$

where $\bar{x}$ and $\bar{y}$ are the sample means of each of the data sets, and $s_x$ and $s_y$ are the standard deviations.

The coefficient can take values from -1 and 1. Conventionally, absolute value of correlation less than 0.1 indicates no correlation. The correlation is small between 0.1 and 0.29, medium from 0.3 to 0.49, and large above 0.5.

We want also to mention that for the sets of data corresponding to self-consistent calculations and Langevin dynamics at the respective transition temperatures, the value of the correlation coefficient is 0.97, indicating that the self-consistent approximation correctly grasps the relevant physics of the model, despite the difference in the numeric values.

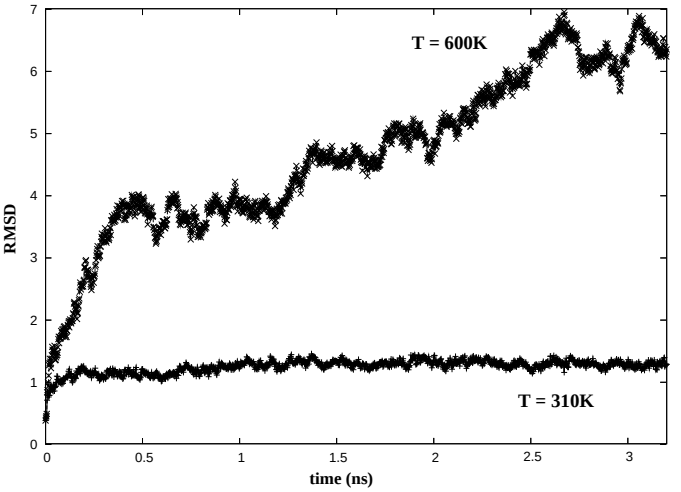


**FIGURE 9.** RMSD for $T = 310$ and $T = 600$ along the trajectory of the all-atom Langevin dynamics simulations

## D.- All-atoms simulation

Finally, we analyze the unfolding of the protein using a complete description of the structure by mean of all-atom molecular dynamics (MD) simulations. Since this technique is computationally expensive, we restrict ourselves to only two temperatures.

Simulations were performed using NAMD[9] on a 16 parallel processor cluster. A CHARMM27[10] force field was used, with a cutoff of 12Å for non-bonded interactions. Protein was embedded in a water sphere of 31Å, large enough to avoid edge effects along the simulation run. Langevin dynamics with a friction coefficient of 5 $ps^{-1}$ was run up to 3 ns. Two different temperatures, 310K and 600K, were simulated. The first one, in order to check the stability of the protein with the simulation parameter used. The second, to induce a fast thermal unfolding and to observe the pathway followed in the process.

After simulation we extract the global RMSD (figure 9) and the RMSD per residue (figure 10). Figure 9 shows the fast unfolding at high temperature whereas the native structure remains stable giving confidence to the simulations. In figure 10 we observe the residues that contribute mostly to RMSD. From the maxima and minima of RMSD of each residue, we corroborate that previous methods identifies more active and less active zones. In agreement with the previous findings, the loops are the most mobile zones, and the first to loose their native structure when we perform a thermal denaturation.

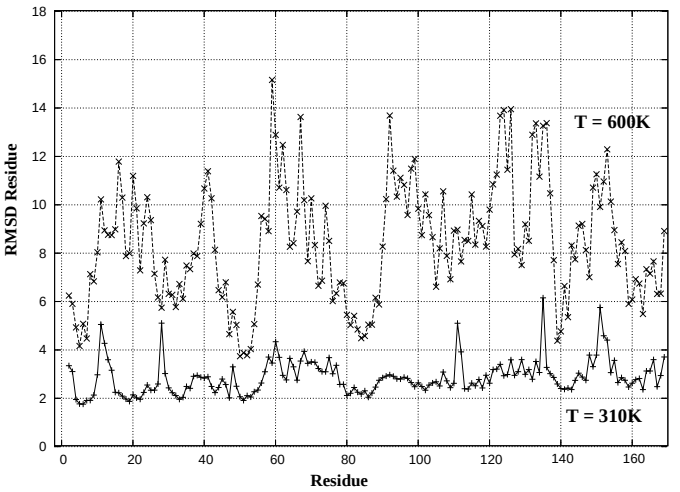


**FIGURE 10.** RMSD per residue for $T = 310$ and $T = 600$ at the end of all-atom Langevin dynamics simulations

## CONCLUSION

We have used two mesoscopic network models to make a statistical mechanics study of the apoflavodoxin unfolding. Within the GNM framework, we have obtained regions which are correlated in their motions. This analysis reveals that loops 57-60, 90-100 and 120-139 are the more flexible protein regions. Using the extended GNM with contact breaking we have obtained the thermal behavior of the unfolding process. This model has been studied with two approaches: a self-consistent method and the Langevin dynamics of the same system. With both methods, we find that the residues that break first the native contacts correspond to those which present high flexibility in the GNM analysis. Finally, preliminary all-atoms MD simulations supports the above findings .

At present, we can neither precise the structure of the intermediate nor support with evidence the hypothetical intermediate proposed in [1]. To this end, further refinement of the model, including the introduction of sequence heterogeneity, is necessary. Such work is under present development. However, our results point in the right direction, allowing to mark the regions in the protein candidates to be broken in the equilibrium intermediate, and possibly suggesting new experiments focused on other unexplored regions.

# REFERENCES


1. L. A. Campos, M. Bueno,J. Lopez-Llano,M. A. Jimenez, and J. Sancho,"Structure of Stable Protein Folding Intermediates by Equilibrium phi-Analysis: The apoflavodoxin Thermal Intermediate." *J. Mol. Biol* **344**, (2004)
2. M. P. Irun, M. M. Garcia-Mira, J. M. Sanchez-Ruiz, and J. Sancho, "Native hydrogen bonds in a molten globule: the apoflavodoxin termal intermediate", *J. Mol. Biol* **306** (2001)
3. I. Bahar,A. R. Atilgan, M. C. Demirel, and B. Erman, *Phys. Rev. Lett.* **80**, (1998).
4. T. Haliloglu,I. Bahar, and B. Erman, *Phys. Rev. Lett.* **79**, 3090-3093 (1997).
5. I. Bahar, B. Erman,R. L. Jernigan, A. R. Atilgan, and D. G. Covell *J. Mol. Biol.* **285**, 1023-1037 (1999).
6. C. Micheletti, J. R. Banavar, and A. Maritan, *Phys. Rev. Lett.* **87**, 88102 (2001).
7. C. Micheletti, F. Cecconi, A. Flammini, and A. Maritan, *Protein Science* **11**, 1878-1887 (2002).
8. T. Shen, L. Canino, and J. McCammon, "Unfolding proteins under external forces: a solvable model under the self-consistent pair contact probability aproximation", *Phys. Rev. Lett.* **89** 2002
9. J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten. *J. Comp. Chem.*, **26**, 1781-1802, (2005).
10. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. *J. Comp. Chem.*, **4**, 187-217 (1983)