

Theoretical-Computational modelling of the Electric Field Effects on Protein Unfolding Thermodynamics

Andrea Amadei^{1*} and Paolo Marracino²

¹Dipartimento di Scienze e Tecnologie Chimiche, Università degli studi di Roma Tor Vergata, Via della Ricerca Scientifica 1, 00031 Rome, Italy

²Dipartimento di Ingegneria dell'Informazione, Elettronica e Telecomunicazioni, Sapienza Università di Roma, via Eudossiana 18, 00184 Rome, Italy

SUPPORTING INFORMATION

Appendix A

Let us consider a binary solute-solvent system made of n solute molecules and N solvent molecules. Any intensive thermodynamic property f can be fully defined as a function of pressure p , temperature T and solute molecular fraction χ_n , that is $f(p, T, \chi_n)$, thus providing

$$(n + N) \left(\frac{\partial f}{\partial n} \right)_{p,T,N} = (n + N) \left(\frac{\partial f}{\partial \chi_n} \right)_{p,T} \left(\frac{\partial \chi_n}{\partial n} \right)_N = \left(\frac{\partial f}{\partial \chi_n} \right)_{p,T} (1 - \chi_n) \quad \text{A.1}$$

$$(n + N) \left(\frac{\partial f}{\partial N} \right)_{p,T,n} = (n + N) \left(\frac{\partial f}{\partial \chi_n} \right)_{p,T} \left(\frac{\partial \chi_n}{\partial N} \right)_n = - \left(\frac{\partial f}{\partial \chi_n} \right)_{p,T} \chi_n \quad \text{A.2}$$

Eqs (A.1) and (A.2) plainly show that in the limit $\chi_n \rightarrow 0$ only eq (A.2) is always vanishing (we obviously consider $\partial f / \partial \chi_n$ as never diverging), while eq (A.1) only in the case that

$\lim_{\chi_n \rightarrow 0} \left(\frac{\partial f}{\partial \chi_n} \right)_{p,T} = 0$ provides a null value.

The intensive properties involved in thermodynamic relations like eq (A.1) (i.e. the product of an extensive property with $\left(\frac{\partial f}{\partial n} \right)_{p,T,N}$) can be in general defined as the ratio of two extensive properties, say Y and Z , providing

$$f(p, T, \chi_n) = \frac{Y}{Z} = \frac{ny_n + Ny_N}{nz_n + Nz_N} = \frac{\chi_n(y_n - y_N) + y_N}{\chi_n(z_n - z_N) + z_N} \quad \text{A.3}$$

with y_n, z_n and y_N, z_N the Y and Z solute and solvent partial molecular properties, respectively. Considering that in the limit of solute infinite dilution any added solute molecule cannot modify the solute or solvent partial molecular properties (no solute-solute interaction) and hence their derivatives in χ_n for the $\chi_n \rightarrow 0$ limit must vanish, we readily obtain

$$\begin{aligned}\lim_{\chi_n \rightarrow 0} \left(\frac{\partial f}{\partial \chi_n} \right)_{p,T} &= \lim_{\chi_n \rightarrow 0} \left\{ \frac{y_n - y_N}{\chi_n(z_n - z_N) + z_N} - \frac{\chi_n(y_n - y_N) + y_N}{[\chi_n(z_n - z_N) + z_N]^2} (z_n - z_N) \right\} \\ &= \lim_{\chi_n \rightarrow 0} \frac{(y_n - y_N)z_N - y_N(z_n - z_N)}{[\chi_n(z_n - z_N) + z_N]^2} = \frac{(y_0 - y_s)z_s - y_s(z_0 - z_s)}{z_s^2}\end{aligned}\quad \text{A. 4}$$

where y_0, z_0 are the solute infinite dilution partial molecular properties and y_s, z_s the pure solvent ones. Eq (A.4) shows that in order to have

$$\lim_{\chi_n \rightarrow 0} (n + N) \left(\frac{\partial f}{\partial n} \right)_{p,T,N} \cong 0 \quad \text{A. 5}$$

the relation

$$(y_0 - y_s)z_s \cong y_s(z_0 - z_s) \quad \text{A. 6}$$

must be fulfilled, providing

$$\frac{(y_0 - y_s)}{y_s} \cong \frac{(z_0 - z_s)}{z_s} \quad \text{A. 7}$$

Eq (A.7) is obviously valid when either $y_0 \cong y_s$ and $z_0 \cong z_s$ or $y_0 \cong z_0 \cong 0$. For more general conditions eq (A.7) requires that the Y and Z properties be correlated in order to provide roughly the same relative shifts of the solute partial molecular properties with respect to the solvent ones (in previous papers^{47,48} where we assumed eq (A.5) at least approximately valid, the Y and Z properties corresponded to the closely related and typically correlated third and second potential energy central moments). In the present paper where the intensive property of interest is $f = \varepsilon_0 \chi E = M/V$ eqs (A.6) and (A.7) cannot be considered in general fulfilled (the system volume and mean dipole can vary rather independently of each other) and hence we cannot use as a reasonable assumption $\lim_{\chi_n \rightarrow 0} \left(\frac{\partial f}{\partial \chi_n} \right)_{p,T} \cong 0$.

Appendix B

From the definition of chemical potential we have that for a macroscopic system made of N solvent molecules at temperature T within a volume V (corresponding to a pressure p), the chemical potential of a single solute molecule inserted in this system (infinite dilution chemical potential) is

$$\mu(p, T) = A(N, V, T, 1) - A(N, V, T, 0) \quad \text{B. 1}$$

Eq B.1 is clearly based on the consideration that for a macroscopic system the insertion of a single molecule corresponds to virtually a differential variation of the molecular amount, thus providing differential variations of the extensive thermodynamic properties and unable to modify any intensive property. Conversely, when dealing with nanoscopic size systems even the insertion of a single solute molecule at fixed volume, especially for large solutes, may provide a significant increase of the system pressure Δp and thus the Helmholtz free energy change due to inserting the

solute molecule at fixed volume, cannot correspond anymore to the chemical potential at the pure solvent pressure condition as in eq (B.1). In fact, for such nanoscopic systems we can rewrite eq (B.1) considering that the discrete Helmholtz free energy change be corresponding to the average of the solute chemical potentials at p and $p + \Delta p$ pressure conditions (i.e. the mean of the solute chemical potentials as provided by inserting the single solute molecule in a macroscopic solvent system either at pressure p or pressure $p + \Delta p$), that is

$$A(N, V, T, 1) - A(N, V, T, 0) \cong \frac{\mu(p, T) + \mu(p + \Delta p, T)}{2} \quad \text{B. 2}$$

By expressing $\mu(p + \Delta p, T)$ via

$$\mu(p + \Delta p, T) = \mu(p, T) + \int_p^{p+\Delta p} v(p') dp' \cong \mu(p, T) + v(p)\Delta p \quad \text{B. 3}$$

we readily obtain

$$\mu(p, T) \cong A(N, V, T, 1) - A(N, V, T, 0) - \frac{v(p)\Delta p}{2} \quad \text{B. 4}$$

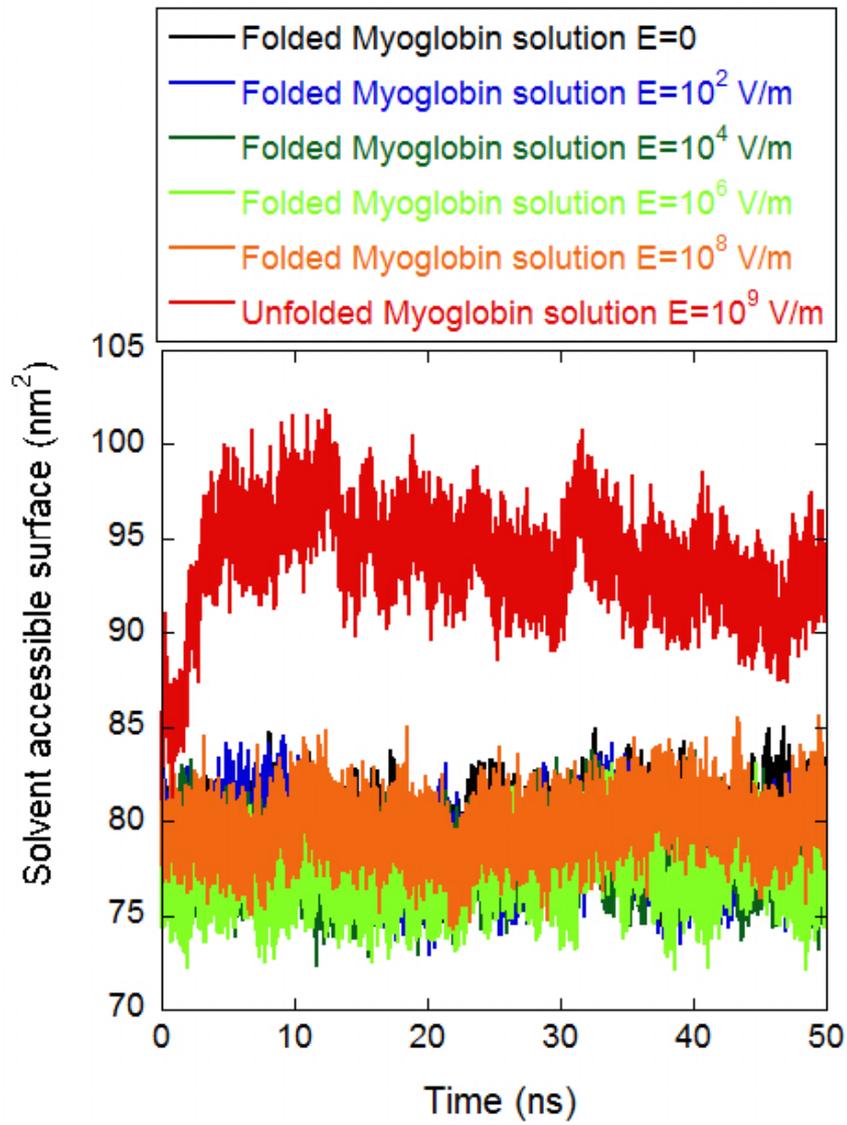
with v the solute partial molecular volume. Eq (B.4) clearly shows that for a nanoscopic system as a MD simulation box, even assuming all the basic thermodynamic relations still working, the true chemical potential of the solute at infinite dilution requires a correction ($-v(p)\Delta p/2$) with respect to the usual thermodynamic expression valid for macroscopic systems. Therefore, for the folded to unfolded state transition we have, assuming virtually the same pressure increase due to the folded or unfolded protein insertion,

$$\mu_D(p, T) - \mu_N(p, T) \cong A_D(N, V, T, 1) - A_N(N, V, T, 1) - \frac{\Delta p}{2} [v_D(p) - v_N(p)] \quad \text{B. 5}$$

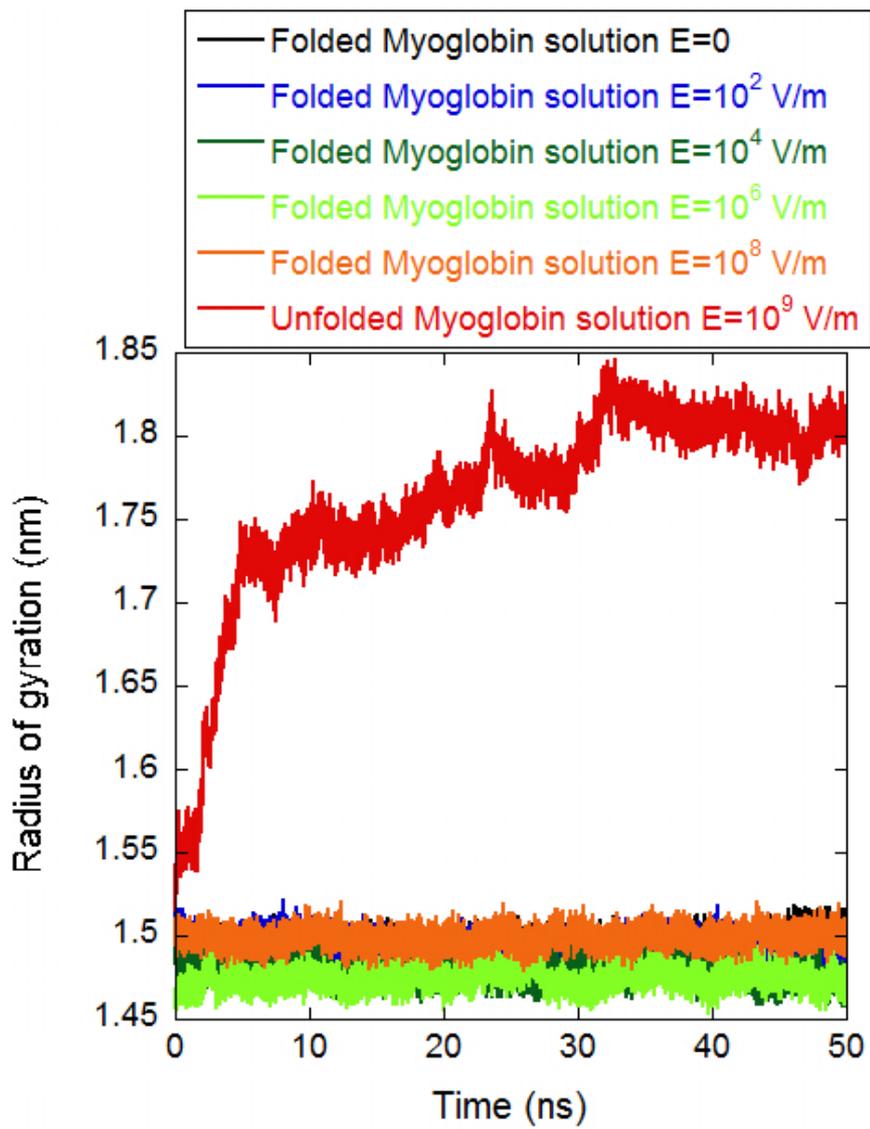
Eq (B.5) indicates that in order to have a simulation providing essentially the correct folding-unfolding equilibrium distribution, i.e. reproducing the equilibrium distribution of a single solute molecule embedded in a macroscopic solvent system, we actually only need that the correction term $-\frac{\Delta p}{2} [v_D(p) - v_N(p)]$ be negligible compared to the Helmholtz free energy change. Therefore, for large simulation boxes (containing the protein hydration shell and a significant number of bulk solvent molecules) with then not too large Δp , even when considering large solutes like proteins with large partial molecular volumes, the correction term is likely to be rather small if not negligible within the noise as $v_D(p)$ and $v_N(p)$ are typically close (for the SPC-Myoglobin system of this paper, considering the experimental partial molecular volume variation of about -100 ml/mol and the pressure increase as obtained by our MD simulations, such a correction is 3-4 kJ/mol, comparable to the estimated noise of our calculated chemical potential change).

Appendix C

Total solvent accessible surface as obtained by Gromacs 'g_sas' tool:

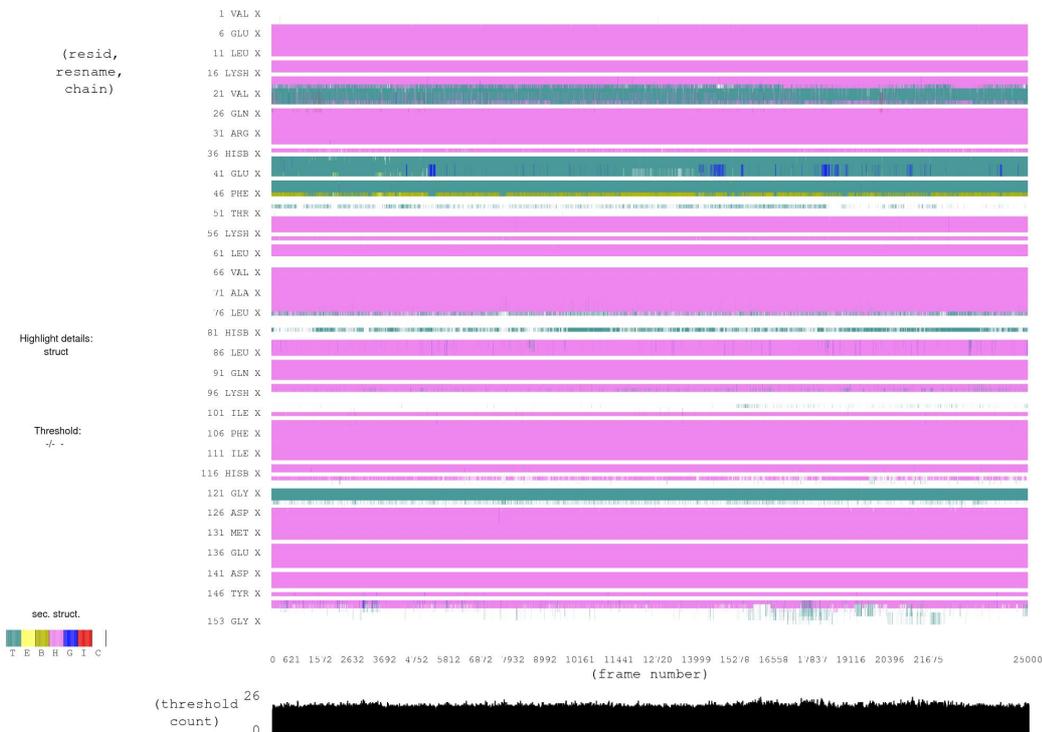


Radius of gyration as obtained by Gromacs 'g_gyrate' tool:

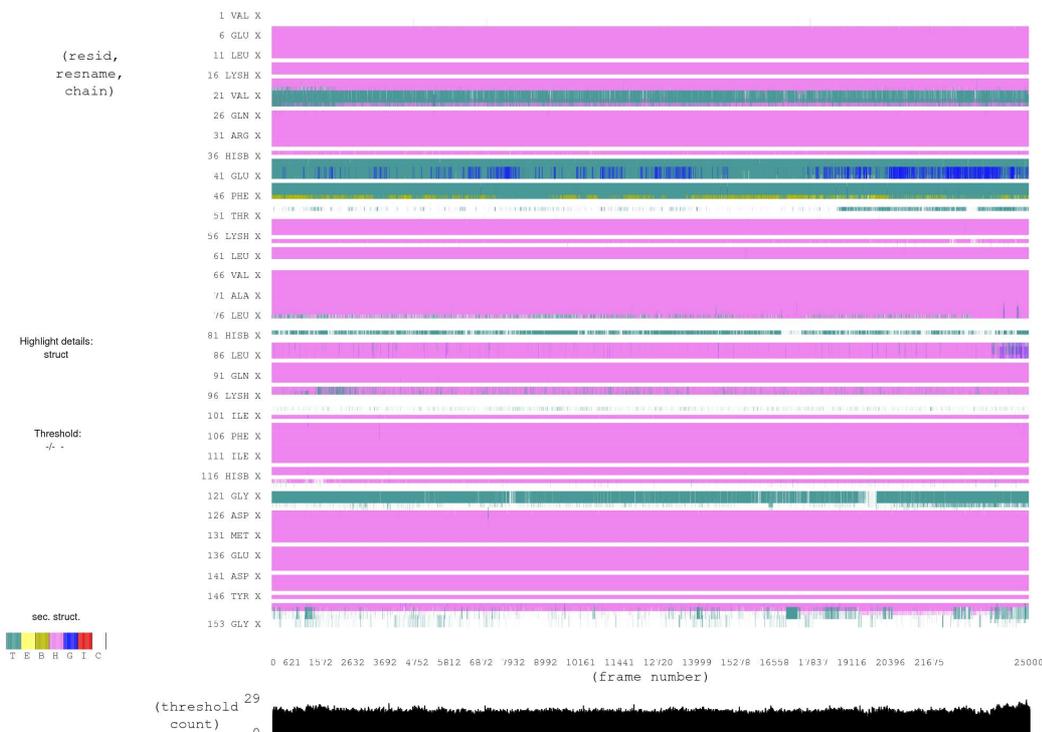


Timelines per-residue as obtained by 'timeline' tool in VMD with increasing electric fields:

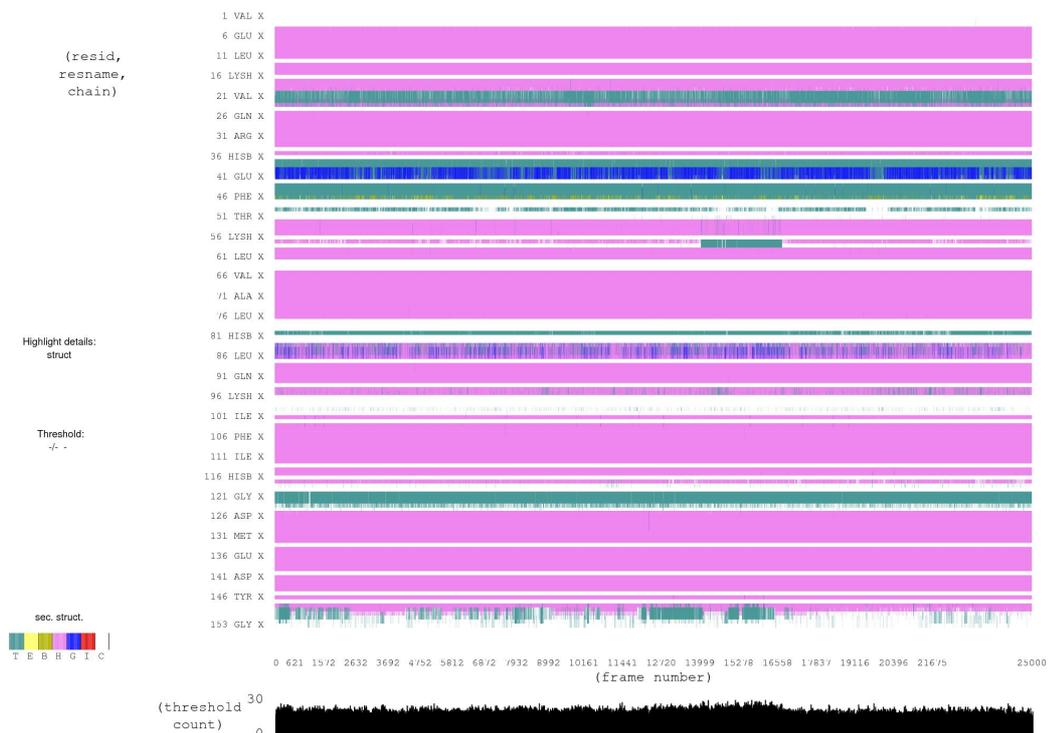
Folded Myoglobin E=0



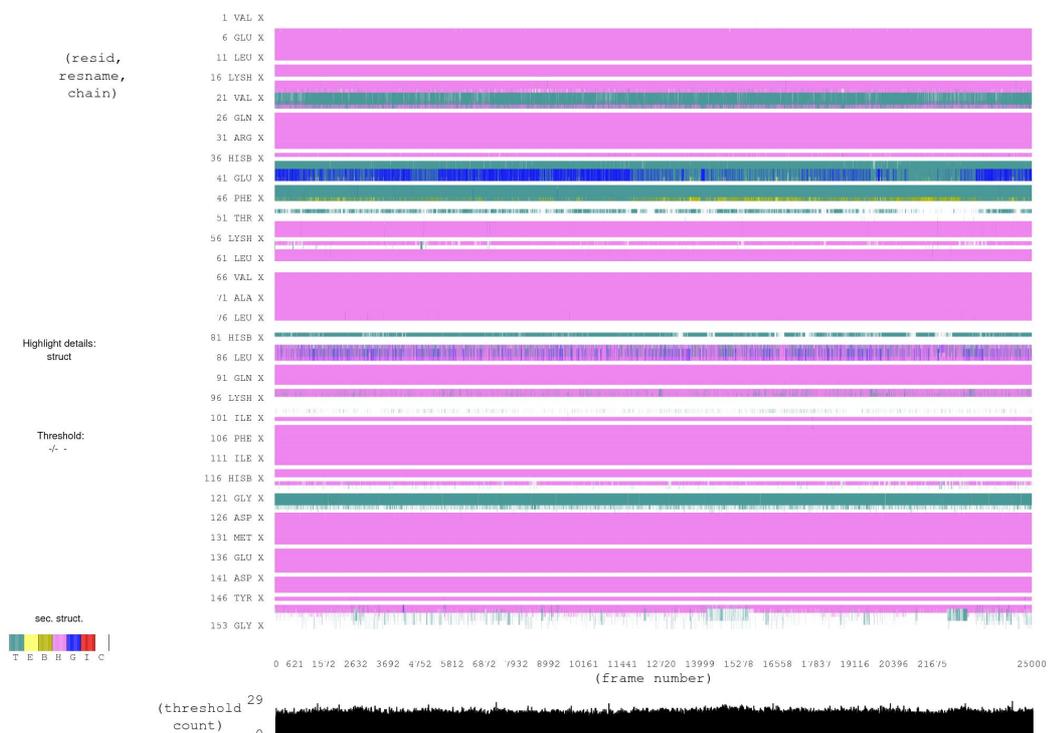
Folded Myoglobin E=10² V/m



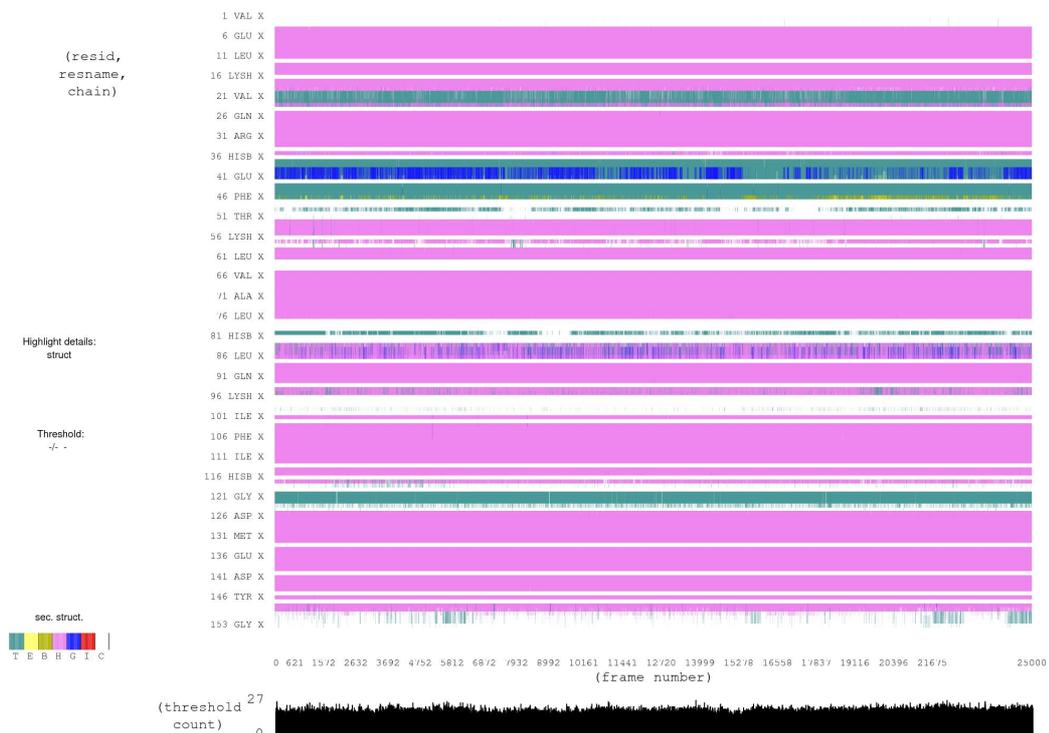
Folded Myoglobin E=10⁴ V/m



Folded Myoglobin E=10⁶ V/m



Folded Myoglobin E=10⁸ V/m



Unfolded Myoglobin E=10⁹ V/m

