Biophysical Journal, Volume 99

Supporting Material

Mechanical unfolding of acylphosphatase studied by single-molecule force spectroscopy and MD simulations

Gali Arad-Haase, Silvia G. Chuartzman, Shlomi Dagan, Reinat Nevo, Maksim Kouza, Binh Khanh Mai, Hung Tien Nguyen, Mai Suan Li, and Ziv Reich

Supporting Material
Mechanical unfolding of acylphosphatase studied by single-molecule force spectroscopy and MD simulations

Gali Arad-Haase, Silvia G. Chuartzman, Shlomi Dagan, Reinat Nevo, Maksim Kouza, Binh Khanh Mai, Hung Tien Nguyen, Mai Suan Li, and Ziv Reich

MATERIALS AND METHODS

Construction of poly-AcP. A synthetic gene encoding for the C21S variant of human muscle acylphosphatase was purchased from GENEART. The gene was cloned into *Age*I/*Pac*I-digested pET-45b(+) (Novagen) and the resulting construct, pET45-AcP, served as a template for the construction of poly-AcP following (1-3), with some modifications: The synthetic AcP gene had a unique *Ava*I site within a short linker region at its 3' terminus. Using PCR, an additional *Ava*I site was introduced at the 5' terminus of the gene, resulting in two, non-palindromic *Ava*I sites at both ends of the insert. The modified AcP-containing fragment was amplified in bacteria, to avoid mis-insertions often introduced by PCR, by an intermediary cloning step, which consisted of its insertion into the original pET45-AcP plasmid (giving rise to pET45-$[AcP]_2$) and its restriction by *Ava*I. The monomeric construct was then oligomerized by directional self-ligation and the concatamers were gel-purified and cloned back into *Ava*I-digetsed, dephosphorylated pET45-AcP. Ligation products were transformed into *E. coli* recombination-defective cells (SURE, Stratagene) for amplification, and the colonies were analyzed for the presence of AcP concatamers by PCR. Screening of over 400 colonies yielded a construct containing four direct tandem AcP repeats. In addition to the AcP modules, the construct, pET45-$[AcP]_4$, contained an N-terminal His_6 tag, to facilitate purification and surface adsorption, three, seven-amino-acid-long linkers, separating adjacent protein repeats, and two tandem C-terminal cysteine residues, for covalent attachment of the polyprotein to gold-coated surfaces (not used in this study, Fig. S1).

While monomeric AcP could readily be expressed in large quantities using standard protocols, expression of poly-AcP proved to be difficult. Following various trials, BLR(DE3)pLysS (*rec*A^-, Novagen), or BL21, cells were transformed with pET45-$[AcP]_4$ and grown overnight at 37ºC in M9 broth supplemented with 1 % (w/v) glucose. The culture was diluted 50 fold into LB containing 0.1 % (w/v) glucose and the cells were grown at 37ºC to $O.D_{600} \approx 0.6$, at which point expression of the polyprotein was induced by addition of 0.1 mM IPTG. Bacteria were lysed by sonication in a mixture containing 50 mM Tris·HCl (pH 8), 500 mM NaCl, 1 mM PMSF, protease inhibitor cocktail (Calbiochem), DNase (1 μg/ml), and lyzosyme (40 U/ml). Soluble poly-AcP was purified following a three-step process using Ni^{2+} affinity chromatography (HiTrap chelating HP, GE Healthcare), gel filtration (HiLoad 16/60 Superdex 200, GE Healthcare), and anion-exchange chromatography (HiTrap Q FF, GE Healthcare). The identification of the

purified (≥95%) protein was confirmed by proteolytic digest and intact mass measurements (Reflex III MALDI-TOF mass spectrometer, Bruker). The protein (in a buffer containing 50 mM Tris·HCl (pH 8.0), 0.1 M NaCl, and 10% glycerol) was flash frozen and stored at -80°C. Immediately before use, it was dialyzed against and equilibrated with 50 mM sodium acetate buffer (pH 5.5), which was used in all measurements performed in this study.

Circular dichroism. Far-UV CD spectra of AcP and $[AcP]_4$ were recorded at 25°C on Chirascan spectropolarimeter (Applied Photophysics) in 0.1-cm path-length quartz cuvettes, using a step size of 0.5 nm, a bandwidth of 1 nm, and a time constant of 1 s. Following acquisition, the spectra were baseline-corrected, averaged, and smoothed using a sliding window of 1.5 nm.

Activity assays. The enzymatic activity of AcP and $[AcP]_4$ was monitored as described in (4). The assay is based on the change in absorbance at 283 nm following hydrolysis of the AcP substrate benzoylphosphate. Measurements were carried out at 25°C in buffered solutions containing 2 mM benzoylphosphate and varying molar equivalents (in nanomolar final concentrations) of AcP or $[AcP]_4$.

Force spectroscopy. Measurements were carried out at room temperature (22-25°C) on a PicoSPM AFM (Molecular Imaging, Agilent Technologies) equipped with a liquid cell, using silicon nitride cantilevers (MSCT-AUHW, Veeco Instruments). The spring constants of the cantilevers were determined by measuring the amplitude of their thermal fluctuations (5), and were in the range of 0.019-0.039 N/m. Poly-AcP proteins were allowed to adsorb onto freshly cleaved V1 grade mica (Ted Pella) for 20 minutes, after which excess was washed away and the bound molecules were covered by 400 µl of sodium acetate buffered solution (50 mM, pH 5.5). In some experiments, phosphate (in the form of HNa_2PO_4) was added to the liquid cell at a final concentration of 10 mM. To characterize the pull-rate dependence of the unfolding force (6), unfolding experiments were conducted at different pulling speeds, ranging from ~30 to 10,000 nm/s. Analysis was performed on traces (from 65 to 256, obtained from two-three independent experiments) exhibiting at least two clear sequential unfolding events; the last force peak in the traces, which reflects detachment of the polyprotein from the tip or from the surface, was excluded. To describe the dependence of the measured forces (F) on extension (x), the rising phase of each saw-tooth in the force-extension profiles (corresponding to the entropic-elasticity of unfolded protein domains) was fitted to a wormlike chain (WLC) model following (7-9):

$$F(x) = \frac{k_B T}{p}\left(\frac{1}{4(1 - x/Lc)^2} - \frac{1}{4} + \frac{x}{Lc}\right) \qquad (1)$$

where p (fixed at 0.36 nm) and L_c denote persistence and contour lengths, respectively, and $k_B T$ is thermal energy (=4.1 pN·nm = $4.1 \cdot 10^{-21}$ J, at room temperature). The observed dependence of unfolding forces on pulling speed was modeled using Monte Carlo simulations, assuming a two-state Markov process (10, 11). The simulations were initialized with all domains folded (N_f) and serially connected by a linear spring. The

polymer was then pulled at a given speed (ν_c) and the restoring force generated by stretching at each time (polling) interval ($\Delta t = \Delta x/\nu_c$, chosen such that the probability of unfolding is <<1) was calculated from the WLC model described above. At each time step, the probability of unfolding of any domain in the polyprotein was calculated using:

$$P_u = N_f k(F)\Delta t \qquad (2),$$

where *k(F)*, the force-dependent unfolding rate, is approximated by:

$$k(F) = k_u^0 \exp(F x_u / k_B T) \quad (3),$$

where k_u^0 denotes the unfolding rate at zero force and x_u is a small molecular length that marks the thermally averaged width of the activation energy barrier for unfolding projected along the direction of the force (6, 12). A number (between 0 and 1) was then drawn by the computer's random number generator and compared to the probability computed using eq. 2. If unfolding was indicated, the contour length of the polyprotein was increased accordingly and the number of folded domains decreased by one. The probability of refolding during stretching was taken to be zero. This procedure was iterated until all domains were unfolded, and was repeated 10,000 times for each pulling speed. The values of x_u and k_u^0 were obtained by adjusting these parameters in the simulations until best fits with the experimental data were obtained.

Molecular dynamics simulations. Both off-lattice C_α-Go and all-atom models were used. Simulations were performed on the NMR structure of horse muscle AcP (PDB code: 1APS (13); Fig. 1), which differs from the human homologue used in our experiments in five amino acids. This structure was also used in simulations concerning the architecture of the transition state ensemble of AcP (14, 15).

Go-type model. The energy function used in the simulations had the form (16):

$$E = \sum_{bonds} K_r (r_i - r_{0i})^2 + \sum_{angles} K_q (q_i - q_{0i})^2$$
$$\left\{K_f^{(1)}[1 - \cos(f_i - f_{0i})] + K_f^{(3)}[1 - cos3(f_i - f_{0i})]\right\} \quad (4)$$
$$\sum_{dihedrals} \sum_{i>j+3}^{NC} e_H \left[5\left(\frac{r_{0ij}}{r_{0ij}}\right)^{12} - 6\left(\frac{r_{0ij}}{r_{0ij}}\right)^{6} \right] + \sum_{i>j+3}^{NNC} e_H \left(\frac{C}{r_{ij}}\right)^{12}$$

Here, $r_{i,i+1}$ is the distance between beads *i* and *i*+1, q_i is the bond angle between bonds (*i*-1) and *i*, f_i is the dihedral angle around the *i*th bond, and r_{ij} is the distance between *i*th and *j*th residues. The subscript '0' refers to the native conformations and superscripts 'NC' and 'NNC' to native and non-native contacts, respectively. The parameters used in the simulations were: $K_r = 100 e_H / \text{Å}^2$, $K_q = 20 e_H / rad^2$, $K_f^{(1)} = e_H$, and $K_f^{(3)} = 0.5 e_H$, were e_H is the characteristic hydrogen bond energy and C = 4 Å. As in our previous

works (17, 18), we set e_H = 0.98 kcal/mol. The temperature T = 285 K then corresponds to 0.53 e_H / k_B and all computations have been performed at this temperature. The force unit is [f] = e_H /Å = 68 pN (17).

The simulations were carried out in the over-damped limit with the water viscosity $\zeta = 50\frac{m}{\tau_L}$ [4], where the time unit $\tau_L = \left(ma^2 / e_H\right)^{1/2} = 3\,\text{ps}$, m is a typical mass of amino acids, and a (=4 Å) is the distance between two neighboring residues. Neglecting the inertia term, the Brownian dynamics equation was numerically solved by the simple Euler method. Due to the large viscosity, we could choose a large time step $\Delta t = 0.1\tau_L$, which allowed us to study unfolding at very low loading rates.

In the constant velocity force simulations, we fixed the N-terminal and pulled the C-terminal by applying a force $f = k(v_c t - r)$, where r is the displacement of the pulled atom from its original position (19); the spring constant of cantilever, k, was set to be the same as the spring constant of the Go model (K_r, Eq. 4). The pulling direction was chosen along the vector drawn from the fixed atom to the pulled one. The mechanical unfolding sequence was followed by monitoring the fraction of native contacts present in the seven secondary structures of AcP and between nine pairs of them, as a function of ΔR. A native backbone contact between residues i and j($|i-j|>3$) is assumed to be present if the distance between the corresponding C_α atoms is smaller than 6.5 Å.

All-atom model. Constant-velocity simulations were performed using GROMACS program suite (20, 21). The GROMOS96 force field 43a1 (22) was used to model the protein, and the SPC water model (23), to describe the solvent. The protein was placed in an orthorhombic box with edges of 5, 5 and 50 nm, and with about 32,500 water molecules. Equations of motion were integrated by using a leap-frog algorithm with a time step of 2 fs. The bond lengths were constrained with a relative geometric tolerance of 10^{-4}, using Lincs (24). Long-range electrostatic interactions were treated using the particle-mesh Ewald method (25). The nonbonded interaction pair-list was updated every 10 fs, using a cutoff of 1.2 nm.

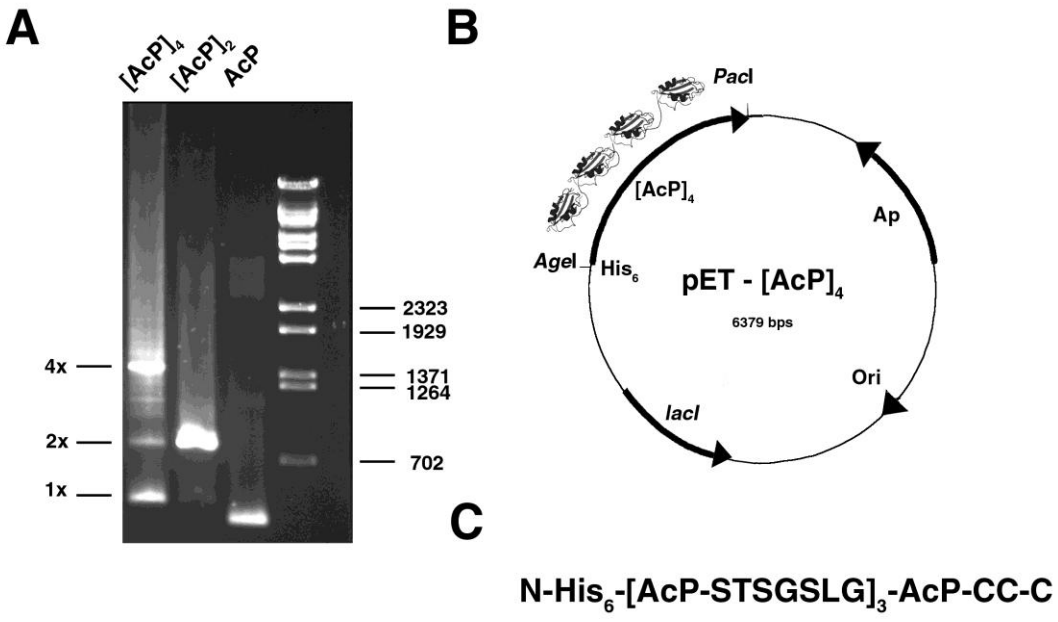
The experimentally determined native structure was subjected to a short energy minimization, using the steepest decent method. Subsequently, unconstrained molecular dynamics simulation was performed to equilibrate the solvated system for 100 ps at constant pressure (1 atm) and temperature (300 K) with the help of the Berendsen coupling procedure (26). The system was then equilibrated further at constant temperature (300 K) and constant volume. Afterward, the N-terminal was kept fixed and a force was applied to the C-terminal through a virtual cantilever moving at the constant velocity v along the longest axis of simulation box. The spring constant k was set to 1700 pN/nm (1000 kJ/mol·nm^2). The resulting force was computed for each time step to generate force extension profiles. Overall, the simulation procedure is similar to the experimental one, except that the pulling speed applied in the simulations is several orders of magnitude higher than those used in the experiments. Using $v_c = 10^9$ nm/s, we generated four trajectories for the N-terminal fixed configuration. The state of protein

along the unfolding path was assessed by monitoring the number of hydrogen bonds present in secondary structures. A hydrogen bond was taken to be present provided that the distance between donor D and acceptor A $\leq$3.5 Å and the angle D-H-A $\geq$145°.

References

1. Carrion-Vazquez, M., A. F. Oberhauser, S. B. Fowler, P. E. Marszalek, S. E. Broedel, J. Clarke, and J. M. Fernandez. 1999. Mechanical and chemical unfolding of a single protein: a comparison. *Proc Natl Acad Sci U S A* 96:3694-3699.
2. Graham, G. J., and J. J. Maio. 1992. A rapid and reliable method to create tandem arrays of short DNA sequences. *Biotechniques* 13:780-789.
3. Hartley, J. L., and T. J. Gregori. 1981. Cloning multiple copies of a DNA segment. *Gene* 13:347-353.
4. Ramponi, G., C. Treves, and A. Guerritore. 1966. Continuous optical assay of acylphosphatase with benzoylphosphate as substrate. *Experientia* 22:705-706
5. Hutter, J., and J. Bechhoefer. 1993. Calibration of atomic-force microscope tips. *Rev Sci Instrum* 64:1868-1873.
6. Evans, E., and K. Ritchie. 1997. Dynamic strength of molecular adhesion bonds. *Biophys J* 72:1541-1555.
7. Bustamante, C., J. F. Marko, E. D. Siggia, and S. Smith. 1994. Entropic elasticity of lambda-phage DNA. *Science* 265:1599-1600.
8. Marko, J. F., and E. D. Siggia. 1995. Stretching DNA. *Macromolecules* 28:8759–8770.
9. Rivetti, C., C. Walker, and C. Bustamante. 1998. Polymer chain statistics and conformational analysis of DNA molecules with bends or sections of different flexibility. *J Mol Biol* 280:41-59.
10. Rief, M., J. M. Fernandez, and H. E. Gaub. 1998. Elastically coupled two-level systems as a model for biopolymer extensibility. *Phys. Rev. Lett* 81:4764 - 4767.
11. Rief, M., M. Gautel, F. Oesterhelt, J. M. Fernandez, and H. E. Gaub. 1997. Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science* 276:1109-1112.
12. Bell, G. I. 1978. Models for the specific adhesion of cells to cells. *Science* 200:618-627.
13. Pastore, A., V. Saudek, G. Ramponi, and R. J. Williams. 1992. Three-dimensional structure of acylphosphatase. Refinement and structure analysis. *J Mol Biol* 224:427-440.
14. Vendruscolo, M., E. Paci, C. M. Dobson, and M. Karplus. 2001. Three key residues form a critical contact network in a protein folding transition state. *Nature* 409:641-645.
15. Paci, E., M. Vendruscolo, C. M. Dobson, and M. Karplus. 2002. Determination of a transition state at atomic resolution from protein engineering data. *J Mol Biol* 324:151-163.
16. Clementi, C., H. Nymeyer, and J. N. Onuchic. 2000. Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? An investigation for small globular proteins. *J Mol Biol* 298:937-953.

17. Li, M. S., M. Kouza, and C. K. Hu. 2007. Refolding upon force quench and pathways of mechanical and thermal unfolding of ubiquitin. *Biophys J* 92:547-561.
18. Kouza, M., C. K. Hu, and M. S. Li. 2008. New force replica exchange method and protein folding pathways probed by force-clamp technique. *J Chem Phys* 128: 045101-045113.
19. Lu, H., B. Isralewitz, A. Krammer, V. Vogel, and K. Schulten. 1998. Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation. *Biophys J* 75:662-671.
20. Berendsen, H. J. C., D. Vanderspoel, and R. Vandrunen. 1995. Gromacs - a message-passing parallel molecular-dynamics implementation. *Comput Phys Commun* 91:43-56.
21. Lindahl, E., B. Hess, and D. van der Spoel. 2001. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J Mol Model* 7:306-317.
22. van Gunsteren, W. F., S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. Krüger, A. E. Mark, W. Scott, and I. Tironi 1996. Biomolecular simulation: the GROMOS96 manual and user guide. Vdf Hochschulverlag AG an der ETH, Zurich.
23. Berendsen, H. J. C., J. P. M. Postma, W. F. van Gunsteren, and J. Hermans. 1981. Interaction models for water in relation to protein hydration. In Intermolecular Forces. B. Pullman Ed. Reidel, Dordrecht. 331-342.
24. Hess, B., H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije. 1997. LINCS: A linear constraint solver for molecular simulations. *J Comput Chem* 18:1463-1472.
25. Darden, T., D. York, and L. Pedersen. 1993. Particle mesh Ewald - an N.Log(N) method for Ewald sums in large systems. *J Chem Phys* 98:10089-10092.
26. Berendsen, H. J. C., J. P. M. Postma, W. F. Vangunsteren, A. Dinola, and J. R. Haak. 1984. Molecular-dynamics with coupling to an external bath. *J Chem Phys* 81:3684-3690.



D

```
                                                *
AcP Human: MSTAQSLKSVDYEVFGRVQGVCFRMYTEDEARKIGVVGWVKNTSKGTVT…
AcP Horse: MSTARPLKSVDYEVFGRVQGVCFRMYAEDEARKIGVVGWVKNTSKGTVT…

AcP Human: …GQVQGPEDKVNSMKSWLSKVGSPSSRIDRTNFSNEKTISKLEYSNFSIRY
AcP Horse: …GQVQGPEEKVNSMKSWLSKVGSPSSRIDRTNFSNEKTISKLEYSNFSVRY
```

Figure S1. Construction of poly-AcP. (*A*) Agarose gel showing monomeric, dimeric, and tetrameric forms of AcP. The multimers also contain linker regions that separate the AcP modules. (*B*) The expression vector used to express $[AcP]_4$. (*C*) Schematic presentation of the polyprotein construct with its inter-domain linkers and terminal histidine tag and cystein doublet. (*D*) Sequence alignment of the human and horse homologues of muscle-type AcP, used in the experiments and MD simulations. The asterisk denotes the free cysteine that was replaced by serine to avoid complexities associated with the presence of free cysteine residues in the protein.