Supporting Information

NMR-based measurements of site-specific electrostatic potentials of histone tails in nucleosome core particles

Nicolas Bolik-Coulon[1-4,*], Philip Rößler [1-4], and Lewis E. Kay[1-4,*]

1) Department of Molecular Genetics, University of Toronto, Toronto, ON, M5S 1A8, Canada
2) Department of Chemistry, University of Toronto, Toronto, ON, M5S 3H6, Canada
3) Department of Biochemistry, University of Toronto, Toronto, ON, M5S 1A8, Canada
4) Program in Molecular Medicine, Hospital for Sick Children Research Institute, Toronto, ON, M5G 0A4, Canada

ϕ_{ENS} values based on measured transverse sPREs exclusively, Eq. [5]

ϕ_{ENS} values have been calculated using Eq. [5] of the main text which is based on the ratio of transverse sPRE rates, and, in turn, on the ratio of spectral density values evaluated at zero frequency, in the macromolecular limit.[1] For the case of spherically symmetric potentials of mean force describing cosolute/protein interactions, and assuming that the nuclear and electron spins reside in the center of the molecules, Okuno and Clore[2] and subsequently Okuno[3] have shown that the ratio of rates is well approximated by ratios of $\langle r^{-4} \rangle$, where r is the distance between nuclear and electron spins and the angular brackets (<>) indicate averaging over the potential of mean force. Such assumptions are clearly an oversimplification for the NCP system considered here. Okuno, Clore and coworkers have suggested recording both longitudinal ($\Gamma_{1,i}$) and transverse ($\Gamma_{2,i}$) sPREs that are then suitably manipulated to extract per-residue values of τ_C, a parameter quantifying timescales of the interactions between cosolute electrons and each proton probe of the molecule in question.[2,4] Then $\Gamma_{2,i}/\tau_{C,i}$ values, that are proportional to $\langle r^{-6} \rangle$, rather than simply $\Gamma_{2,i}$ rates, are used in Eq. [5] to compute ϕ_{ENS}. Recording accurate amide proton $\Gamma_{1,i}$ rates is non-trivial as cross-relaxation with adjacent protons renders the decay multi-exponential, as does exchange with water solvent. In the case of IDRs of proteins, as considered here, the situation is further exacerbated by the fact that the amides are not protected from solvent exchange, which can be a factor for our measurements carried out at 37 ^{o}C. Second, Γ_1 rates are, in general, small. For example, in the case where Γ_2 = 20 s^{-1} and for τ_C =1 ns, similar to what has been obtained experimentally[2,4], a Γ_1 value of only 0.7 s^{-1} is calculated (1 GHz). In the work described here we have chosen not to interpret the measured electrostatic potentials in terms of distances quantitatively but rather to use them as probes of relative potentials along histone tails and to assess how these vary for the different histones.

Calculation of ϕ_{PB}

Surface electrostatic potentials were computed for isolated extended histone tails as described in Methods in the main text. The APBS[5,6] program has been used after converting the pdb format into pqr format using the following setup:

```
read
   mol pqr {tail}.pqr
end
elec
   mg-auto
   dime 257 257 257
   cglen 138 138 138
   fglen 138 138 138
   cgcent mol 1
   fgcent mol 1
   mol 1
   npbe
   bcfl sdh
   pdie 2.0000
   sdie 74.126
```

```
    ion charge 1.000 conc 0.03319 radius 2.000
    ion charge -1.000 conc 0.03081 radius 2.000
    ion charge -2.000 conc 0.00119 radius 2.000
    srfm smol
    chgm spl2
    sdens 10.00
    srad 1.40
    swin 0.30
    temp 310.15
    calcenergy total
    calcforce no
    write pot dx {tail}
end
print elecEnergy 1 end
quit
```

where *{tail}* refers to the histone tail of interest, the salt concentration of ions with charges +1 (33.19 mM), -1 (30.81 mM) and -2 (1.19 mM) are calculated for a solution of 20 mM sodium phosphate with 12 mM NaCl at pH 6 and 37°C. The output file contains the electrostatic potential at each point of the grid, and the Poisson-Boltzmann surface electrostatic potential is subsequently calculated using Eq. [4] of the main text.

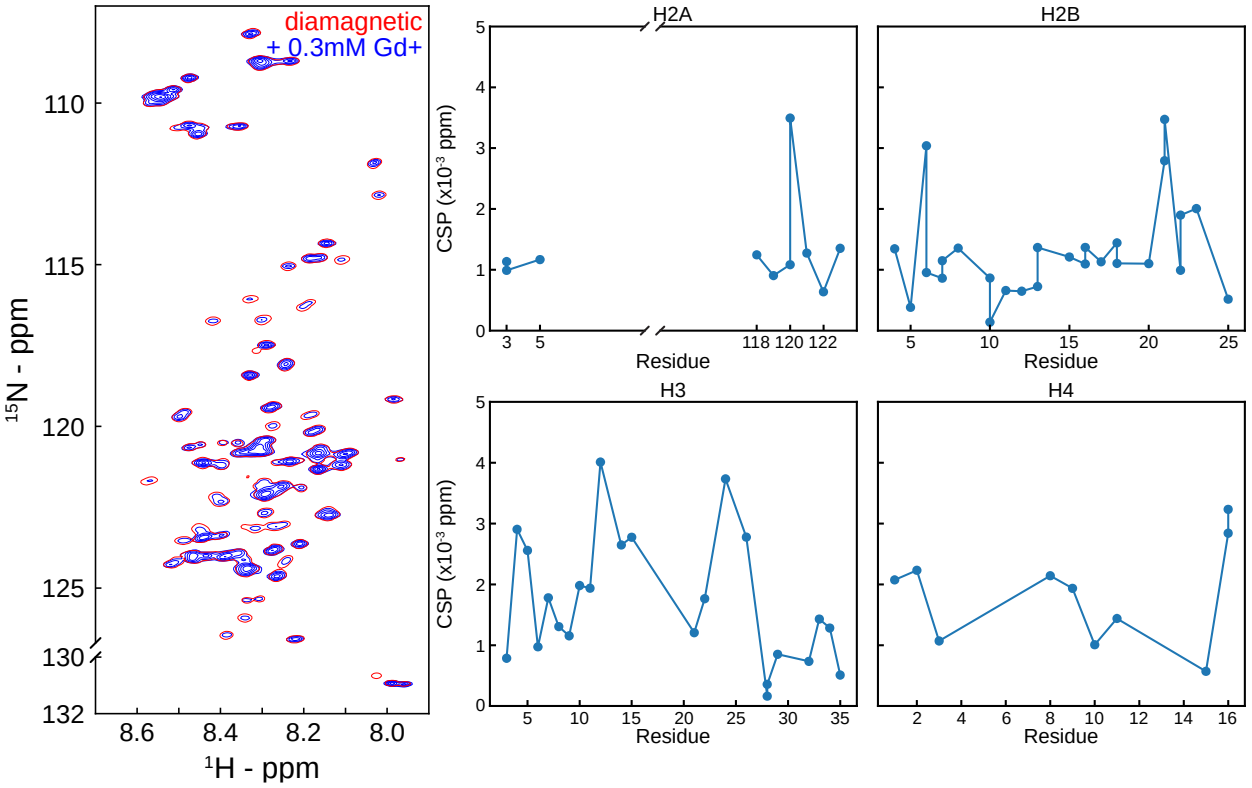

Figure S1. Chemical shift perturbations (CSP) in NCP spectra are not observed upon addition of cosolutes. Superposition of [^{1}H-^{15}N]-HSQC spectra recorded on a [^{2}H-^{15}N]-ILV NCP sample, 1 GHz, 37°C, with (blue, multiple contours) and without (red, single contour) the addition of 0.3 mM Gd-DOTAM-BA. CSP values are listed for each histone tail in the panels to the right.

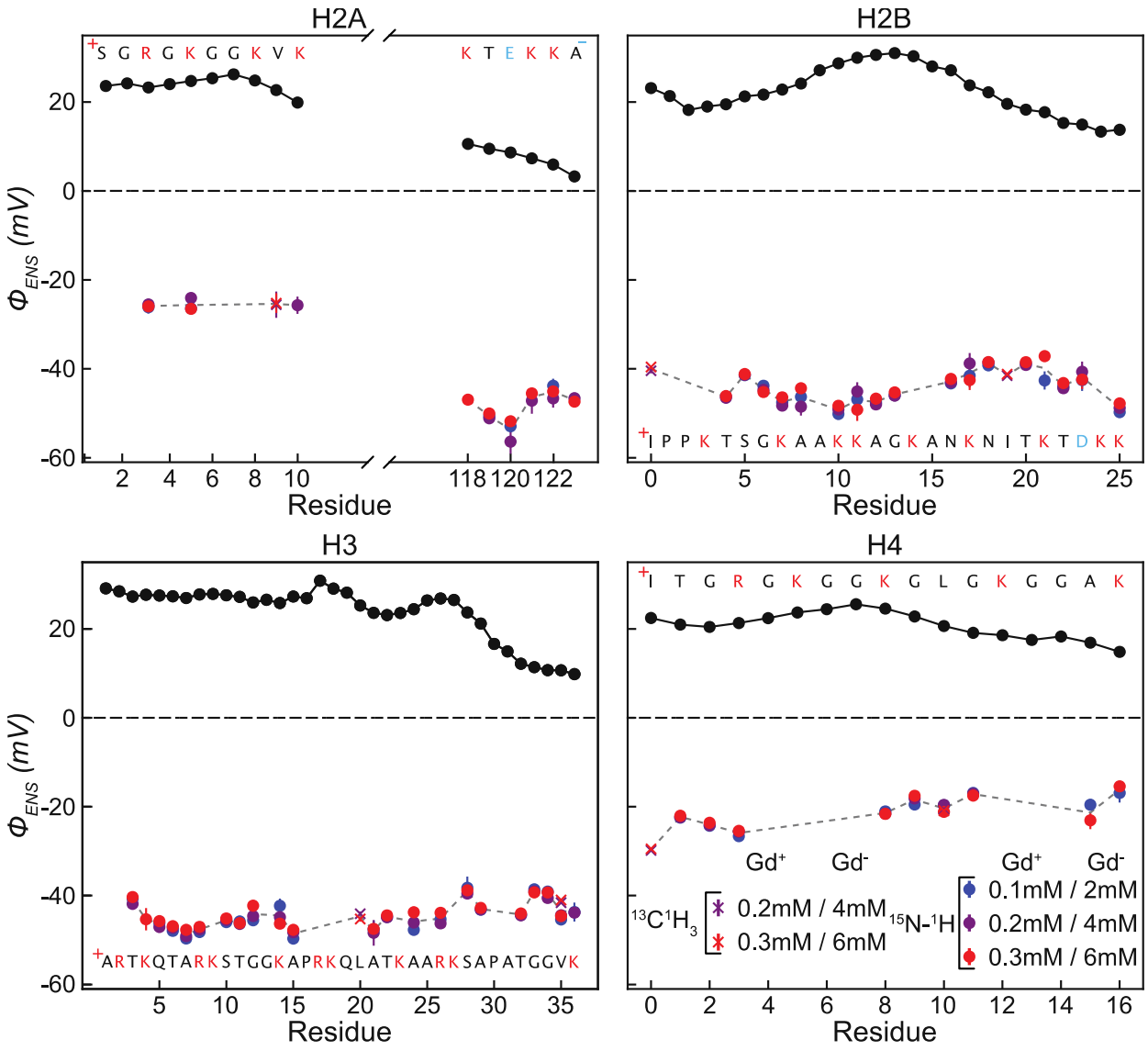

Figure S2. Comparison of experimentally derived ϕ_{ENS} values of NCP tails in the context of the NCP (colored balls, see Figure 3 of the main text) and calculated potentials for isolated histone tails (black). The dashed grey line connects the average experimental potential of each residue for visual clarity. Potentials for the isolated tails were computed using APBS software, assuming each tail is in an extended conformation; an additional five glycine residues were added onto the C-termini of H2A-N, H2B, H3, and H4, and the N-terminus of H2A-C to eliminate terminal charges that would not be present in the intact NCP; terminal charges for the histone tails of the interact NCP are indicated.

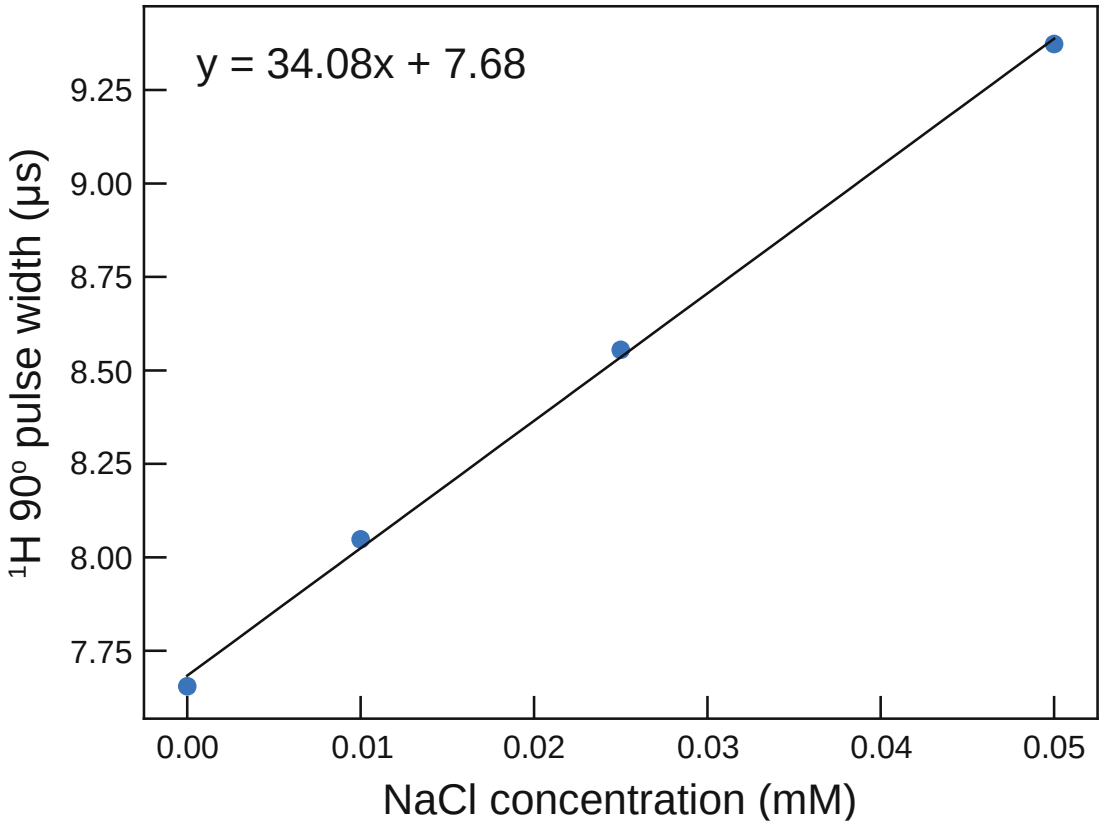

Figure S3. Linear correlation plot of the ^{1}H 90° pulse width calibrated on reference samples of NMR buffer (20 mM sodium phosphate, pH 6) at 37°C as a function of NaCl concentration (blue circles). The data were recorded on a 23.5 T spectrometer (proton Larmor frequency of 1 GHz) using the maximum ^{1}H power level (correlation plots must be generated for each spectrometer). The best fit line and the associated equation are indicated in black. The equation is used to estimate the additional amount of salt, given in terms of NaCl concentration, that results from adding NCP and associated ions to buffer to generate the NMR sample (see main text).

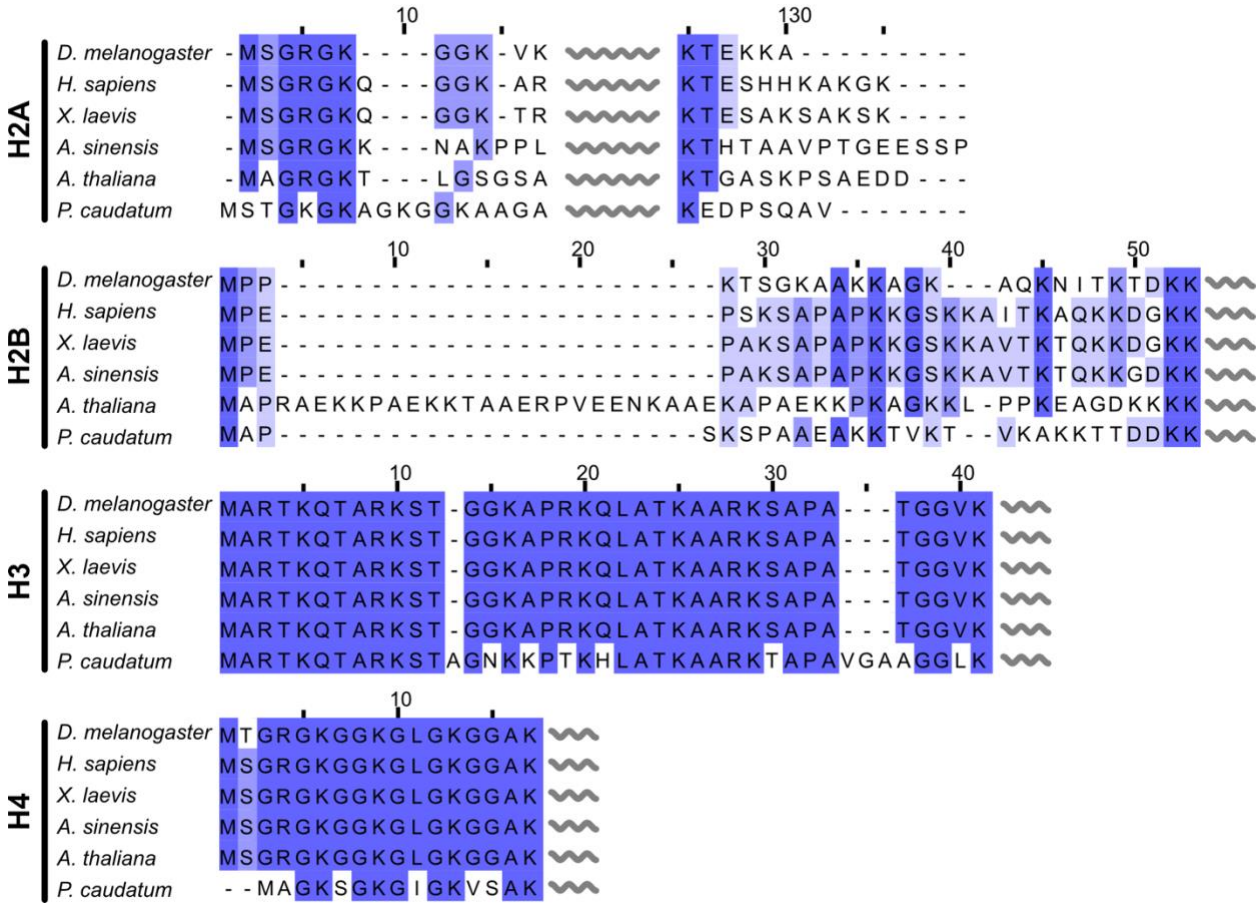

Figure S4. Alignments of histone tail sequences from several organisms spanning distant taxonomic clades. Compared are sequences from *Drosophila melanogaster* (insect, Uniprot IDs P84051, P02283, P02299, P84040 for H2A, H2B, H3, and H4, respectively), *Homo sapiens* (mammal, Uniprot IDs Q6FI13, P33778, P68431, P62805), *Xenopus laevis* (amphibian, Uniprot IDs P06897, P02281, P84233, P62799), *Alligator sinensis* (reptile, Uniprot IDs A0A1U7SB82, A0A1U7RBB8, A0A3Q0GLJ1, A0A1U7RIZ8), *Arabidopsis thaliana* (plant, Uniprot IDs Q9C681, Q9LQQ4, P59226, P59259), and *Paramecium caudatum* (ciliate, Uniprot IDs Q5KR85, Q5KR84, Q5KR82, Q95YH5). Alignments were generated with Clustal Omega[7] and plotted in JalView.[8] Positions in the sequence are coloured by identity, with dark blue positions indicating high identity among the compared sequences. Grey waves indicate the position of the histone core domain relative to the represented tail sequences. Although the organisms considered only share distant common ancestors, amino acid sequences of the histone tails are well conserved, especially for H3 and H4. Note that the sequences here can be slightly different than those in Figure 3, reflecting cloning artifacts or the addition of residues to improve expression (see legend to Figure 3).

Table S1[a]

	Species	Length	Glycine Content	Charge Density His+	Charge Density His0
H2A N-tail	*P. caudatum*	17	35.3%	0.294	0.294
	A. sinensis	13	15.4%	0.385	0.385
	A. thaliana	13	30.8%	0.231	0.231
	D. melanogaster	11	36.4%	0.455	0.455
	X. laevis	12	33.3%	0.417	0.417
	H. sapiens	12	33.3%	0.417	0.417
H2A C-tail	*P. caudatum*	8	0.0%	-0.250	-0.250
	A. sinensis	15	6.7%	-0.067	-0.133
	A. thaliana	12	8.3%	-0.167	-0.167
	D. melanogaster	6	0.0%	0.167	0.167
	X. laevis	11	0.0%	0.182	0.182
	H. sapiens	11	9.1%	0.364	0.182
H2B N-tail	*P. caudatum*	28	0.0%	0.250	0.250
	A. sinensis	29	6.9%	0.310	0.310
	A. thaliana	52	3.8%	0.192	0.192
	D. melanogaster	26	7.7%	0.346	0.346
	X. laevis	29	6.9%	0.310	0.310
	H. sapiens	29	6.9%	0.310	0.310
H3 N-tail	*P. caudatum*	41	9.8%	0.317	0.293
	A. sinensis	37	10.8%	0.324	0.324
	A. thaliana	37	10.8%	0.324	0.324
	D. melanogaster	37	10.8%	0.324	0.324
	X. laevis	37	10.8%	0.324	0.324
	H. sapiens	37	10.8%	0.324	0.324
H4 N-tail	*P. caudatum*	15	26.7%	0.333	0.333
	A. sinensis	17	47.1%	0.353	0.353
	A. thaliana	17	47.1%	0.353	0.353
	D. melanogaster	17	47.1%	0.353	0.353
	X. laevis	17	47.1%	0.353	0.353
	H. sapiens	17	47.1%	0.353	0.353

[a]Comparison of glycine content and charge densities of histone tails between different organisms from distantly related taxonomic clades. The analysed tail sequences and their corresponding Uniprot identifiers are specified in Figure S4. Parameters indicated are the tail length, the fractional glycine content, and the charge density, which is reported as elementary charge per amino acid (e/aa). Charge densities are calculated for charged and non-charged histidine residues, corresponding to pH<6 and pH>7, respectively. Charge densities (with the exception of H2A-C) and fractional glycine content of histone tails are well conserved for a given histone among evolutionary distant organisms.

References

(1) Yu, B.; Pletka, C. C.; Pettitt, B. M.; Iwahara, J. De Novo Determination of Near-Surface Electrostatic Potentials by NMR. *Proc. Natl. Acad. Sci.* **2021**, *118* (25), e2104020118. https://doi.org/10.1073/pnas.2104020118.

(2) Okuno, Y.; Szabo, A.; Clore, G. M. Quantitative Interpretation of Solvent Paramagnetic Relaxation for Probing Protein–Cosolute Interactions. *J. Am. Chem. Soc.* **2020**, *142* (18), 8281–8290. https://doi.org/10.1021/jacs.0c00747.

(3) Okuno, Y. Quantitative Interpretation of Transverse Spin Relaxation by Translational Diffusion in Liquids Under Arbitrary Potentials. bioRxiv August 22, 2024, p 2024.08.21.609078. https://doi.org/10.1101/2024.08.21.609078.

(4) Okuno, Y.; Schwieters, C. D.; Yang, Z.; Clore, G. M. Theory and Applications of Nitroxide-Based Paramagnetic Cosolutes for Probing Intermolecular and Electrostatic Interactions on Protein Surfaces. *J. Am. Chem. Soc.* **2022**, *144* (46), 21371–21388. https://doi.org/10.1021/jacs.2c10035.

(5) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Electrostatics of Nanosystems: Application to Microtubules and the Ribosome. *Proc. Natl. Acad. Sci.* **2001**, *98* (18), 10037–10041. https://doi.org/10.1073/pnas.181342398.

(6) Jurrus, E.; Engel, D.; Star, K.; Monson, K.; Brandi, J.; Felberg, L. E.; Brookes, D. H.; Wilson, L.; Chen, J.; Liles, K.; Chun, M.; Li, P.; Gohara, D. W.; Dolinsky, T.; Konecny, R.; Koes, D. R.; Nielsen, J. E.; Head-Gordon, T.; Geng, W.; Krasny, R.; Wei, G.-W.; Holst, M. J.; McCammon, J. A.; Baker, N. A. Improvements to the APBS Biomolecular Solvation Software Suite. *Protein Sci.* **2018**, *27* (1), 112–128. https://doi.org/10.1002/pro.3280.

(7) Madeira, F.; Madhusoodanan, N.; Lee, J.; Eusebi, A.; Niewielska, A.; Tivey, A. R. N.; Lopez, R.; Butcher, S. The EMBL-EBI Job Dispatcher Sequence Analysis Tools Framework in 2024. *Nucleic Acids Res.* **2024**, *52* (W1), W521–W525. https://doi.org/10.1093/nar/gkae241.

(8) Waterhouse, A. M.; Procter, J. B.; Martin, D. M. A.; Clamp, M.; Barton, G. J. Jalview Version 2—a Multiple Sequence Alignment Editor and Analysis Workbench. *Bioinformatics* **2009**, *25* (9), 1189–1191. https://doi.org/10.1093/bioinformatics/btp033.