Supporting Information for: Fragment Molecular Orbital Calculations with Implicit Solvent Based on the Poisson–Boltzmann Equation: Implementation and DNA Study

Yoshio Okiyama,[*,†,¶] Tatsuya Nakano,[‡,†] Chiduru Watanabe,[†] Kaori Fukuzawa,[§,†] Yuji Mochizuki,[‖,†] and Shigenori Tanaka[⊥]

†*Institute of Industrial Science, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan*

‡*Division of Medicinal Safety Science, National Institute of Health Sciences, 3-25-26 Tonomachi, Kawasaki-ku, Kawasaki, Kanagawa 210-9501, Japan*

§*Faculty of Pharmaceutical Sciences, Hoshi University, 2-4-41 Ebara, Shinagawa-ku, Tokyo 142-8501, Japan*

‖*Department of Chemistry and Research Center for Smart Molecules, Facultyof Science, Rikkyo University, 3-34-1 Nishi-ikebukuro, Toshima-ku, Tokyo 171-8501, Japan*

⊥*Graduate School of System Informatics, Kobe University, 1-1 Rokkodai, Nada-ku, Kobe, Hyogo 657-8501, Japan*

¶*Current address: Division of Medicinal Safety Science, National Institute of Health Sciences, 3-25-26 Tonomachi, Kawasaki-ku, Kawasaki, Kanagawa 210-9501, Japan*

E-mail: okiyama@nihs.go.jp

S1 Finite Difference Poisson–Boltzmann Method

S1.1 Theoretical Basics

The fundamental nonlinear Poisson–Boltzmann (PB) equation is formulated on the electrostatic potential (ESP) field $\phi(\boldsymbol{r})$ at arbitrary position $\boldsymbol{r}$ when a charge distribution is given as a source as follows:

$$-\nabla \cdot [\varepsilon(\boldsymbol{r})\nabla\phi(\boldsymbol{r})] = 4\pi\rho(\boldsymbol{r}) + 4\pi e \sum_i \lambda_i(\boldsymbol{r}) c_i z_i e^{-\frac{z_i e\phi(\boldsymbol{r})}{k_\mathrm{B}T}}, \tag{S1}$$

where a solute with the charge density $\rho(\boldsymbol{r})$ is surrounded by a continuum solvent including dissolved electrolyte where the dielectric constant $\varepsilon(\boldsymbol{r})$ changes depending on the medium polarizability; the charge distribution for the i-th type mobile ion is recursively determined from the potential with its accessibility to solute molecule $\lambda_i(\boldsymbol{r})$ in the form of a step function with respect to the bulk concentration c_i and valence z_i through the Boltzmann's law using the elemental charge, e, the Boltzmann constant, k_B, and the absolute temperature, T.

An electrolyte solvent containing only a pair of symmetric salts satisfies electroneutrality because the salts have the same concentration of both positive and negative ions with the same valence ($|z_i| = z$, in short). In such a situation, the charge distribution of mobile ions in the bulk, where $\varepsilon(\boldsymbol{r})$ has a constant value ε_out and $\lambda(\boldsymbol{r})$ is considered to be equal to unity, is connected to the Debye–Hückel equation with ionic strength $I = \frac{1}{2}\sum_i c_i z_i^2$,

$$4\pi e \sum_i c_i z_i e^{-\frac{z_i e\phi(\boldsymbol{r})}{k_\mathrm{B}T}} = -\varepsilon_\mathrm{out}\kappa^2\zeta \sinh\left[\frac{\phi(\boldsymbol{r})}{\zeta}\right], \tag{S2}$$

where $\zeta = -\frac{k_\mathrm{B}T}{ze}$ and $\kappa = \sqrt{8\pi e^2 I/\varepsilon_\mathrm{out}k_\mathrm{B}T}$, the so-called the Debye–Hückel parameter. For the low electrostatic potential (ESP) condition where $\sinh\left[\frac{\phi(\boldsymbol{r})}{\zeta}\right] \sim \frac{\phi(\boldsymbol{r})}{\zeta}$, eq S2 is equal to $-\varepsilon_\mathrm{out}\kappa^2\phi(\boldsymbol{r})$ and, consequently, a linearized form of eq S1 is derived by introducing the

simple assumptions above,[S1,S2]

$$-\nabla \cdot [\varepsilon(\boldsymbol{r})\nabla\phi(\boldsymbol{r})] + \varepsilon(\boldsymbol{r})\kappa^2(\boldsymbol{r})\phi(\boldsymbol{r}) = 4\pi\rho(\boldsymbol{r}), \tag{S3}$$

where $\kappa(\boldsymbol{r}) = \kappa\lambda(\boldsymbol{r})$.

S1.2 Discretization and Numerical Solution

In the finite difference method for numerically solving the PB equation, the three-dimensional lattice space has n_l $(x, y, z \in l)$ grid points on each axis with interval h_l and defines a box with length of $L_l = (n_l - 1)h_l$ on each side. The continuous functions, $\phi(\boldsymbol{r})$, $\varepsilon(\boldsymbol{r})$, $\kappa(\boldsymbol{r})$ and $\rho(\boldsymbol{r})$, are all discretized in the space and the values are assigned onto grids or their midpoints as $\phi_{i,j,k}$, $\varepsilon_{i,j,k}$, $\kappa_{i,j,k}$ and $\rho_{i,j,k}$, respectively, corresponding to the grid indices (i, j, k), where $x = \{x_0 + (i-1)h_x | 1 \le i \le n_x\}$, $y = \{y_0 + (j-1)h_y | 1 \le j \le n_y\}$, and $z = \{z_0 + (k-1)h_z | 1 \le k \le n_z\}$ for the reference points of coordinates (x_0, y_0, z_0). The interior/exterior dielectric regions are separated by the solvent-excluded molecular surface (SES)[S3,S4] consisting of exposed contact regions defined by atomic radii and reentrant ones defined by the spherical surface of the solvent probe (see Figure 1). The smoothed numerical surface (SNS) algorithm[S5] was employed for the generation of the SES in this study. The dielectricity inside the solute is uniformly distributed with the value ε_{in} here, while more sophisticated assignments based on local specificity[S6–S8] or Gaussian-smoothing methods[S9–S12] have also been proposed for unpolarized media. The solute charge distribution $\rho(\boldsymbol{r})$ is given by the assembly of each point charge Q_A on atom A at nuclear position $\boldsymbol{r}_A$,

$$\rho(\boldsymbol{r}) = \sum_A Q_A \delta(\boldsymbol{r} - \boldsymbol{r}_A), \tag{S4}$$

with Dirac's delta $\delta(\boldsymbol{r})$. Each point charge value is separately assigned to the nearest eight grid points, *i.e.*, the vertexes of the cell, weighted by the trilinear method,[S13] where the factor is determined according to the closeness to each grid point. Using the solute charge

distribution, the initial value of each $\phi_{i,j,k}$ on the whole lattice space and the outer boundary condition are preset by the Debye–Hückel approximation[S13] in consideration of the screening effects of the electrolyte solvent. Now, eq S3 is written in a discretized form,[S13]

$$\begin{aligned}
&\frac{\varepsilon_{i-\frac{1}{2},j,k}(\phi_{i,j,k}-\phi_{i-1,j,k})+\varepsilon_{i+\frac{1}{2},j,k}(\phi_{i,j,k}-\phi_{i+1,j,k})}{h_x^2}\\
+&\frac{\varepsilon_{i,j-\frac{1}{2},k}(\phi_{i,j,k}-\phi_{i,j-1,k})+\varepsilon_{i,j+\frac{1}{2},k}(\phi_{i,j,k}-\phi_{i,j+1,k})}{h_y^2}\\
+&\frac{\varepsilon_{i,j,k-\frac{1}{2}}(\phi_{i,j,k}-\phi_{i,j,k-1})+\varepsilon_{i,j,k+\frac{1}{2}}(\phi_{i,j,k}-\phi_{i,j,k+1})}{h_z^2}\\
&+\varepsilon_{i,j,k}\kappa_{i,j,k}^2\phi_{i,j,k}=\frac{4\pi\rho_{i,j,k}}{h_x h_y h_z},
\end{aligned} \tag{S5}$$

and simplified to a matrix expression as an algebraic equation on the ESP vector $\boldsymbol{\Phi} = \{..., \phi_{i,j,k}, ...\}$,

$$\boldsymbol{A}\boldsymbol{\Phi} = \boldsymbol{Q}, \tag{S6}$$

where $\boldsymbol{A}$ and $\boldsymbol{Q} = \{..., \rho_{i,j,k}, ...\}$ indicate a coefficient matrix to the ESP vector and a vector related to solute charge density, respectively. Beginning with the initial guess of $\boldsymbol{\Phi}$, this equation is iteratively solved by updating the old guess to the new estimation of $\boldsymbol{\Phi}$ until the mean square change from previous step falls below the threshold. Here, eq S6 is solved using the Gauss–Seidel algorithm accelerated by the successive over-relaxation (SOR) method,[S2] whereas the multigrid method[S14–S18] will be implemented to reach stable convergence more rapidly in our future work.

S1.3 Reaction Field Generation

After the convergence of $\boldsymbol{\Phi}$, the induced charges providing the reaction field on a solute as a substitution of solvent are assigned to grid points on the dielectric boundary according to the numerical formulation of Gauss's law to reproduce the ESP around the points[S5] (see Appendix S1.A for the derivation). Furthermore, by projecting the position of each induced charge $\rho_{i,j,k}^{\mathrm{ind}}$ from the uneven grid surface onto the spherically smoothed SES, the electrostatic

solute–solvent interaction is improved with the corrected reaction field.[S5] The value of each induced charge $\rho^{\mathrm{ind}}_{i,j,k}$ on the corrected position $\boldsymbol{r}^{\mathrm{s}}_m$ is the same as that before the correction but is hereafter represented as q^{s}_m for convenience. Finally, the ESP generated inside and imposed on a solute is given by the expression below:

$$\varphi(\boldsymbol{r}) = \sum_{m=1}^{M} \frac{q^{\mathrm{s}}_m}{\varepsilon_{\mathrm{in}} \left| \boldsymbol{r} - \boldsymbol{r}^{\mathrm{s}}_m \right|}, \tag{S7}$$

where M is the total number of induced charges on the molecular surface. Note that one must set a value of $\varepsilon_{\mathrm{in}}$ that is appropriate to the intrinsic polarizability of the solute for a fixed charge distribution in the classical treatment. However, one has only to set it to unity when a polarizable medium with a quantum mechanical (QM) electron density is employed as the solute;[S19,S20] we follow this agreement as the electron density of solute determined by the fragment molecular orbital (FMO) method is used in this study.

Appendix S1.A Induced Charge Definition Based on Gauss's Law

An induced charge distribution on the dielectric boundary is determined to satisfy Gauss's theorem in the differential equation form in the relationship between the electric field $\boldsymbol{E}(\boldsymbol{r})$ and the source of charge,

$$\nabla \cdot \boldsymbol{E}(\boldsymbol{r}) = 4\pi\rho(\boldsymbol{r}) + 4\pi\rho^{\mathrm{ind}}(\boldsymbol{r}), \tag{S8}$$

where $\rho(\boldsymbol{r})$ and $\rho^{\mathrm{ind}}(\boldsymbol{r})$ are the fixed charge density and the induced one, respectively. By substituting $\boldsymbol{E}(\boldsymbol{r}) = -\nabla\phi(\boldsymbol{r})$ for the equation above, one obtains the Poisson equation,

$$-\nabla^2\phi(\boldsymbol{r}) = 4\pi\rho(\boldsymbol{r}) + 4\pi\rho^{\mathrm{ind}}(\boldsymbol{r}), \tag{S9}$$

which is the simplest form of eq S3 and discretized in the same form of eq S5 with the conditions in vacuo, where $\varepsilon(\boldsymbol{r})$ is unity and $\kappa(\boldsymbol{r})$ is zero. A numerical description of Gauss's

law finally yields the solution of an induced charge in the finite difference formalism,

$$\rho^{\mathrm{ind}}_{i,j,k} = -\rho_{i,j,k} + \frac{h_x h_y h_z}{4\pi}\left[2\left(\frac{1}{h_x^2}+\frac{1}{h_y^2}+\frac{1}{h_z^2}\right)\phi_{i,j,k} - \frac{\phi_{i-1,j,k}+\phi_{i+1,j,k}}{h_x^2} - \frac{\phi_{i,j-1,k}+\phi_{i,j+1,k}}{h_y^2} - \frac{\phi_{i,j,k-1}+\phi_{i,j,k+1}}{h_z^2}\right], \tag{S10}$$

and the simplified one in the same form in reference S5,

$$\rho^{\mathrm{ind}}_{i,j,k} = -\rho_{i,j,k} + \frac{3h}{2\pi}\left(\phi_{i,j,k} - \frac{\phi_{i-1,j,k}+\phi_{i+1,j,k}+\phi_{i,j-1,k}+\phi_{i,j+1,k}+\phi_{i,j,k-1}+\phi_{i,j,k+1}}{6}\right), \tag{S11}$$

in the case that the grid intervals on all axes are equal, *i.e.*, $h_x = h_y = h_z = h$.

S2 Figures and Tables

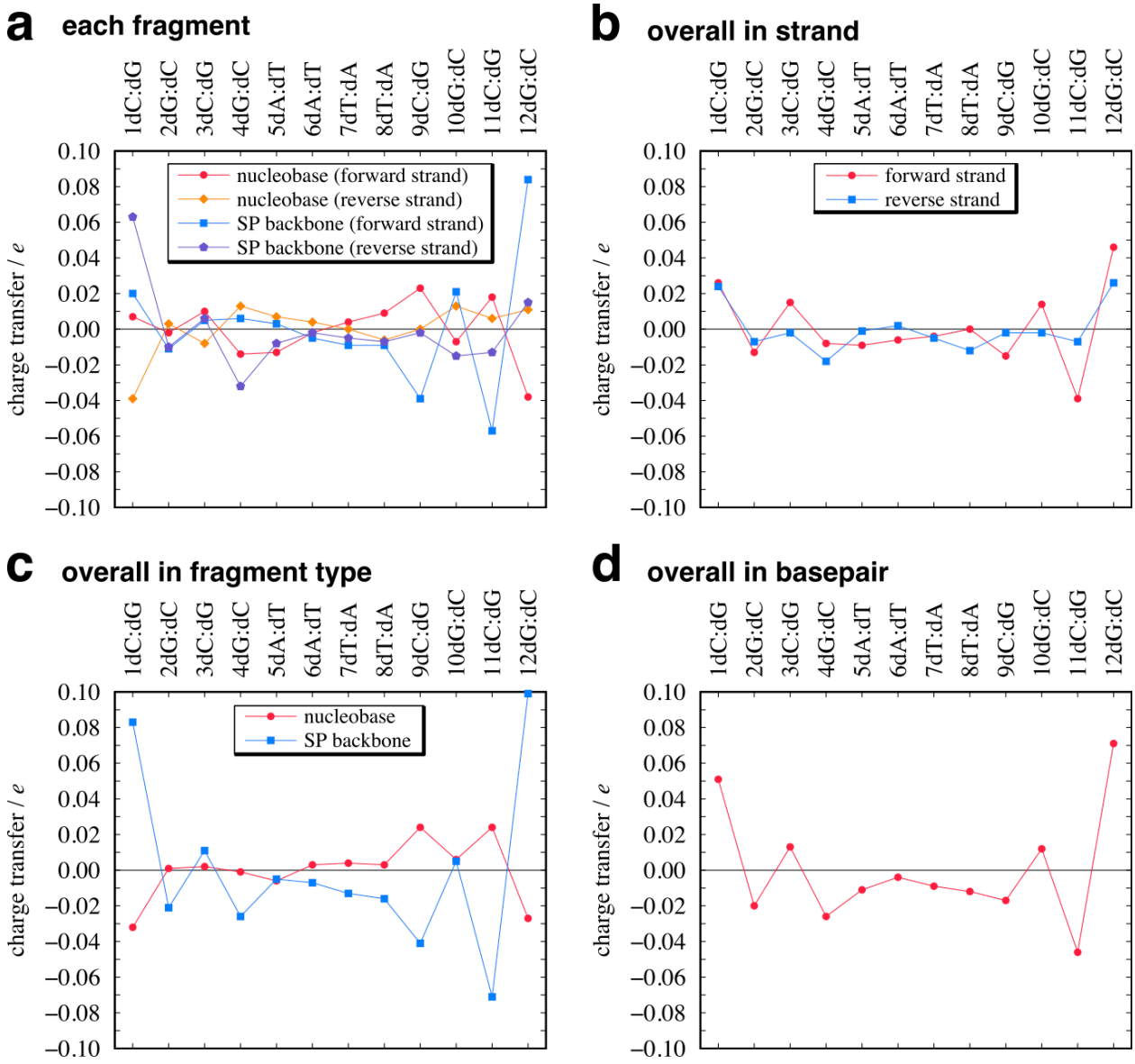

Figure S1: Charge transfer in vacuo and in implicit solvent focusing on the fragment unit of each base pair in the 12-base-pair DNA duplex. The amount of charge in units of the elementary charge was estimated by natural population analysis (NPA).[S21,S22]

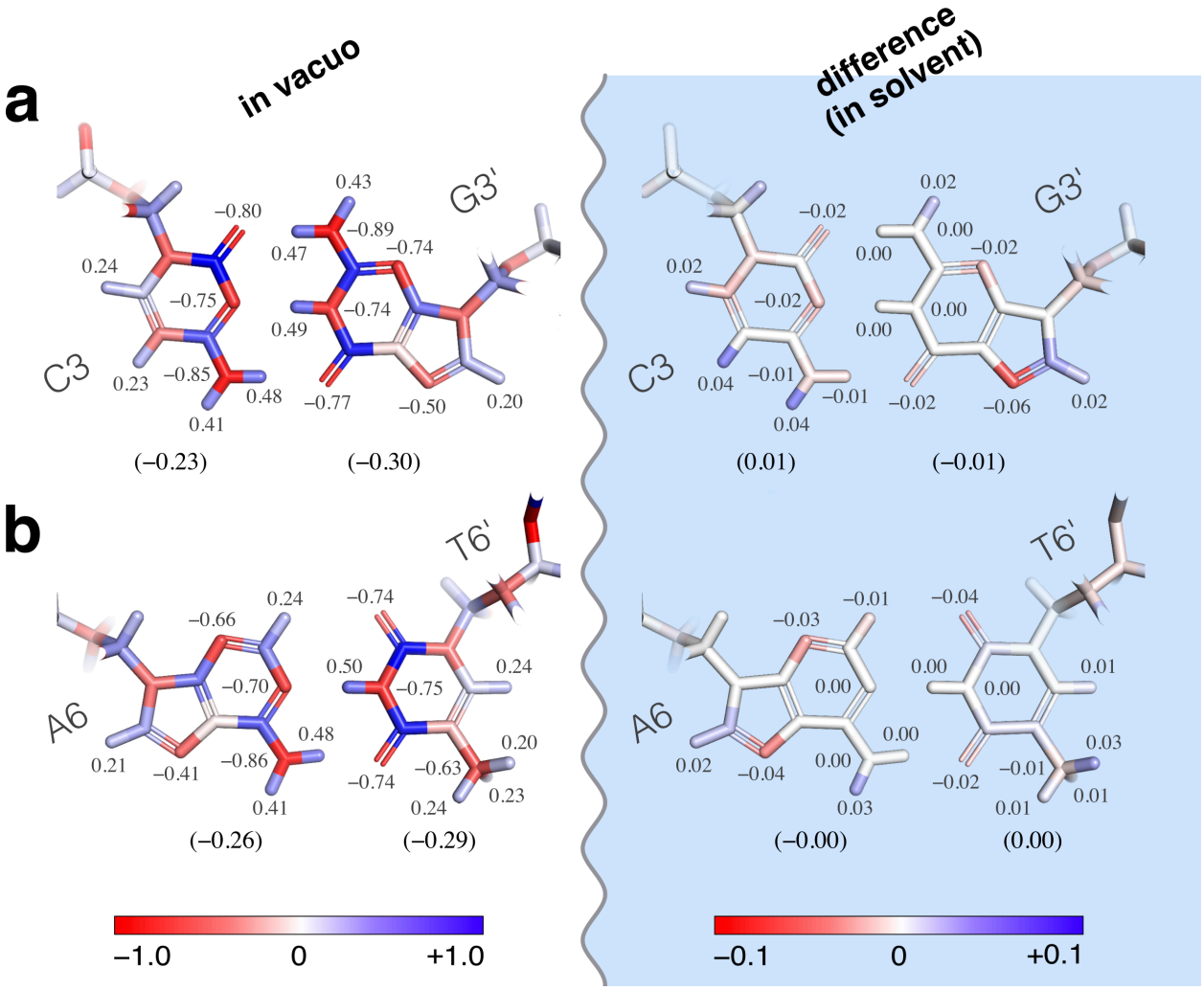

Figure S2: Atomic charge distribution in vacuo and the difference to that in implicit solvent for (a) C3–G3′ and (b) A6–T6′ base pairs. The amount of charge in units of the elementary charge was estimated by natural population analysis (NPA),[S21,S22] and the net value for each base fragment is shown in parenthesis.

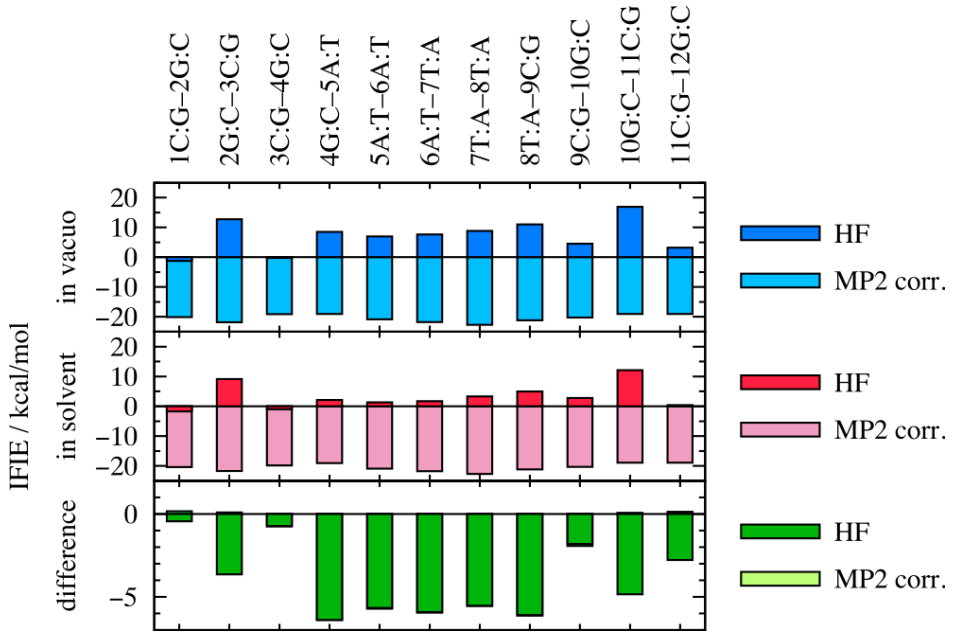

Figure S3: Summation of the IFIEs on stacking interactions between each base pair inside the 12-base-pair DNA duplex compared in vacuo to those in implicit solvent.

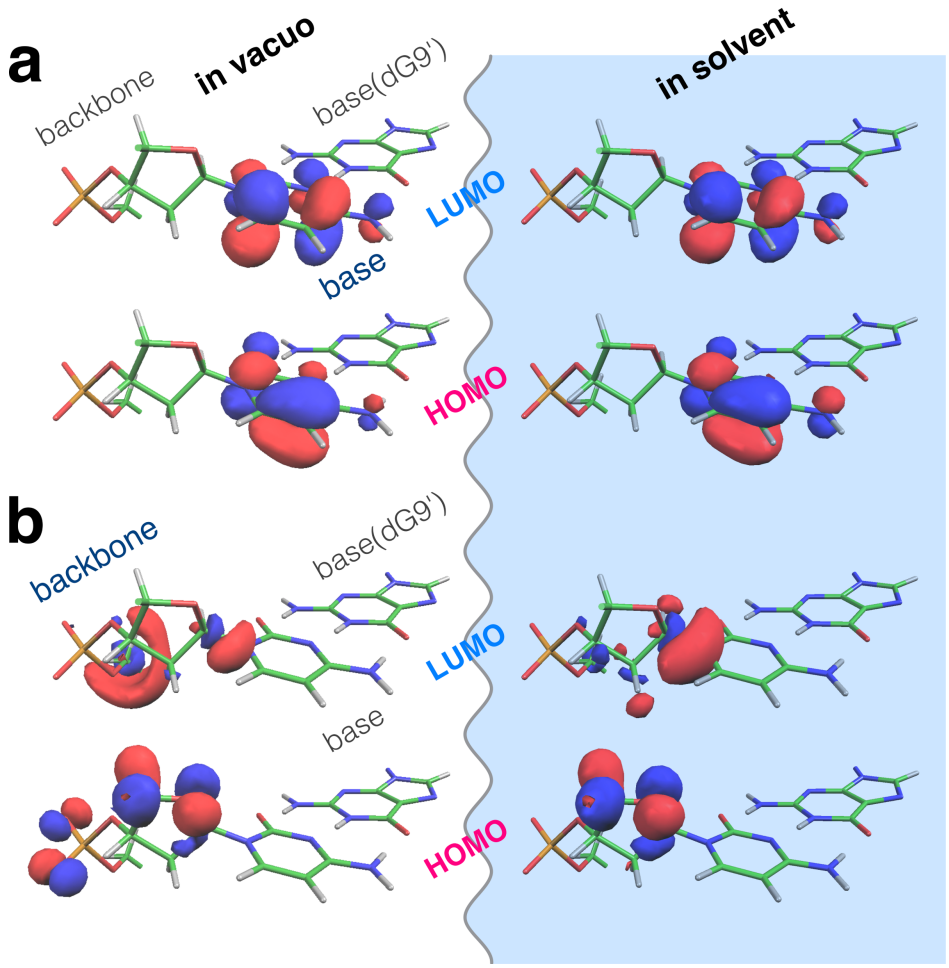

Figure S4: Frontier molecular orbitals on (a) base and (b) sugar–phosphate backbone fragments of dC9 compared in vacuo with in explicit solvent surrounded by a 10-Å solvent shell including 14 Na^+ counterions.

Table S1: Changes in the Frontier Molecular Orbital Energies on Each Fragment Type of the 12-Base-Pair DNA Duplex in Vacuo and in Implicit Solvent[a]

	HOMO		LUMO		band gap	
	mean	s.d.	mean	s.d.	mean	s.d.
adenine base						
in vacuo	18.210	0.062	30.127	0.084	11.917	0.029
in solvent	−4.637	0.054	7.283	0.047	11.920	0.022
guanine base						
in vacuo	16.470	1.513	28.394	1.543	11.924	0.177
in solvent	−4.539	0.317	7.349	0.393	11.887	0.204
cytosine base						
in vacuo	14.759	1.339	27.053	1.348	12.294	0.055
in solvent	−5.982	0.331	6.358	0.326	12.340	0.044
thymine base						
in vacuo	16.869	0.161	28.858	0.168	11.989	0.104
in solvent	−5.703	0.102	6.318	0.167	12.021	0.091
sugar–phosphate backbone						
in vacuo	12.829	1.053	27.858	1.640	15.029	0.695
in solvent	−9.373	0.242	6.396	0.427	15.769	0.301

[a] Each value is averaged over fragments of the same type, excluding those in terminal base pairs, and shown with its standard deviation. All calculated values are in electronvolts (eV).

Table S2: Solvation Free Energies of the 12-Base-Pair DNA Duplex in Explicit Solvent with Increasing Solvent Shell Thickness and Number of Counterions and Comparison to Those in Implicit Solvent[a]

Shell thickness[b]	Counterions[c]	$\Delta\mathcal{E}^{\mathrm{ind}}$[d]	$\mathcal{E}^{\mathrm{es}}$	$\mathcal{G}$
Explicit solvent model				
0	0	0.0	0.0	0.0
3	0	434.5	−1452.4	−1017.9
4	0	499.9	−1777.3	−1277.3
5	8	530.2	−3986.3	−3456.2
6	10	533.9	−4571.6	−4037.7
7	11	543.4	−4874.7	−4331.3
8	12	550.0	−5246.2	−4696.2
9	13	551.5	−5492.6	−4941.1
10	14	547.0	−5684.3	−5137.3
11	14	546.4	−5714.3	−5167.8
12	15	547.5	−5945.9	−5398.5
13	15	549.9	−5998.4	−5448.6
14	15	551.5	−6051.5	−5500.0
20	22	555.5	−6967.4	−6411.9
PB implicit solvent model (only es-term is considered)				
0	0	229.1	−6316.1	−6087.0

[a] All calculated values are in kcal/mol. [b] Each shell thickness value of explicit solvent is in Å. [c] Number of counterions, Na^+. [d] Each value is estimated from the MP2 total energy including the electron-correlation energy.

References

(S1) Honig, B.; Sharp, K.; Yang, A.-S. S. Macroscopic models of aqueous solutions: Biological and chemical applications. *J. Phys. Chem.* **1993**, *97*, 1101–1109.

(S2) Nicholls, A.; Honig, B. H. A rapid finite-difference algorithm, utilizing successive over-relaxation to solve the Poisson–Boltzmann equation. *J. Comput. Chem.* **1991**, *12*, 435–445.

(S3) Richards, F. M. Areas, volumes, packing, and protein structure. *Annu. Rev. Biophys. Bioeng.* **1977**, *6*, 151–176.

(S4) Connolly, M. L. Analytical molecular surface calculation. *J. Appl. Crystallogr.* **1983**, *16*, 548–558.

(S5) Rocchia, W.; Sridharan, S.; Nicholls, A.; Alexov, E.; Chiabrera, A.; Honig, B. H. Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: Applications to the molecular systems and geometric objects. *J. Comput. Chem.* **2002**, *23*, 128–137.

(S6) Voges, D.; Karshikoff, A. A model of a local dielectric constant in proteins. *J. Chem. Phys.* **1998**, *108*, 2219.

(S7) Song, X. An inhomogeneous model of protein dielectric properties: Intrinsic polarizabilities of amino acids. *J. Chem. Phys.* **2002**, *116*, 9359.

(S8) Wang, L.; Zhang, Z.; Rocchia, W.; Alexov, E. Using DelPhi capabilities to mimic protein's conformational reorganization with amino acid specific dielectric constants. *Commun. Comput. Phys.* **2013**, *13*, 13–30.

(S9) Davis, M. E.; McCammon, J. A. Dielectric boundary smoothing in finite difference solutions of the Poisson equation: An approach to improve accuracy and convergence. *J. Comput. Chem.* **1991**, *12*, 909–912.

(S10) Bruccoleri, R.; Novotny, J.; Davis, M. E. Finite difference Poisson–Boltzmann electrostatic calculations: Increased accuracy achieved by harmonic dielectric smoothing and charge antialiasing. *J. Comput. Chem.* **1997**, *18*, 268–276.

(S11) Grant, J. A. A smooth permittivity function for Poisson–Boltzmann solvation methods. *J. Comput. Chem.* **2001**, *22*, 608–640.

(S12) Li, L.; Li, C.; Zhang, Z.; Alexov, E. On the dielectric "constant" of proteins: Smooth dielectric function for macromolecular modeling and its implementation in DelPhi. *J. Chem. Theory Comput.* **2013**, *9*, 2126–2136.

(S13) Klapper, I.; Hagstrom, R.; Fine, R.; Sharp, K. A.; Honig, B. H. Focusing of electric fields in the active site of Cu–Zn superoxide dismutase: Effects of ionic strength and amino-acid modification. *Proteins: Struct., Funct., Genet.* **1986**, *1*, 47–59.

(S14) Holst, M. J.; Kozack, R. E.; Saied, F.; Subramaniam, S. Treatment of electrostatic effects in proteins: Multigrid-based Newton iterative method for solution of the full nonlinear Poisson–Boltzmann equation. *Proteins: Struct., Funct., Genet.* **1994**, *18*, 231–245.

(S15) Holst, M. J.; Saied, F. Multigrid solution of the Poisson–Boltzmann equation. *J. Comput. Chem.* **1993**, *14*, 105–113.

(S16) Wang, J.; Luo, R. Assessment of linear finite-difference Poisson–Boltzmann solvers. *J. Comput. Chem.* **2010**, *31*, 1689–1698.

(S17) Cai, Q.; Hsieh, M. J.; Wang, J.; Luo, R. Performance of nonlinear finite-difference Poisson–Boltzmann solvers. *J. Chem. Theory Comput.* **2010**, *6*, 203–211.

(S18) Holst, M. J.; Saied, F. Numerical solution of the nonlinear Poisson–Boltzmann equation: Developing more robust and efficient methods. *J. Comput. Chem.* **1995**, *16*, 337–364.

(S19) Gogonea, V.; Merz, K. M., Jr. Fully quantum mechanical description of proteins in solution. Combining linear scaling quantum mechanical methodologies with the Poisson–Boltzmann equation. *J. Phys. Chem. A* **1999**, *103*, 5171–5188.

(S20) Wang, M.; Wong, C. F. Calculation of solvation free energy from quantum mechanical charge density and continuum dielectric theory. *J. Phys. Chem. A* **2006**, *110*, 4873–4879.

(S21) Reed, A. E.; Weinhold, F. Natural bond orbital analysis of near-Hartree–Fock water dimer. *J. Chem. Phys.* **1983**, *78*, 4066–4073.

(S22) Reed, A. E.; Weinstock, R. B.; Weinhold, F. Natural population analysis. *J. Chem. Phys.* **1985**, *83*, 735.