THE JOURNAL OF PHYSICAL CHEMISTRY A

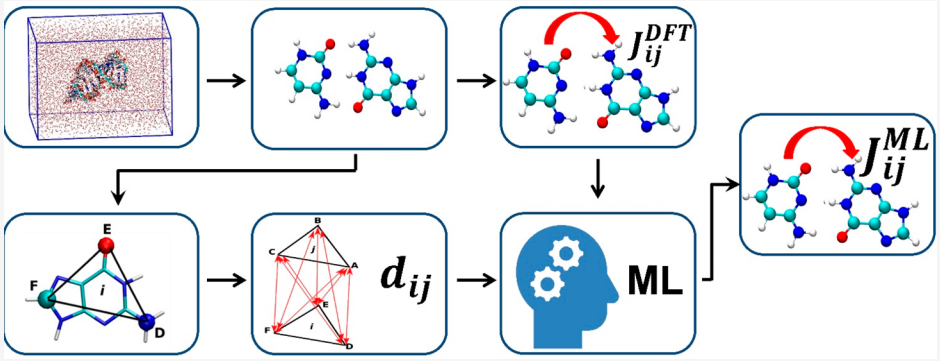
pubs.acs.org/JPCA Article

Machine Learning Prediction of Electronic Coupling between the Guanine Bases of DNA

Published as part of The Journal of Physical Chemistry virtual special issue "Machine Learning in Physical Chemistry".

Saientan Bag, Abhishek Aggarwal, and Prabal K. Maiti*

Cite This: *J. Phys. Chem. A* 2020, 124, 7658–7664 Read Online



ABSTRACT: Charge transport in deoxyribonucleic acid (DNA) is of immense interest in biology and molecular electronics. Electronic coupling between the DNA bases is an important parameter describing the efficiency of charge transport in DNA. A reasonable estimation of this electronic coupling requires many expensive first principle calculations. In this article, we present a machine learning (ML) based model to calculate the electronic coupling between the guanine bases of the DNA (in the same strand) of any length, thus avoiding expensive first-principle calculations. The electronic coupling between the bases are evaluated using density functional theory (DFT) calculations with the morphologies derived from fully atomistic molecular dynamics (MD) simulations. A new and simple protocol based on the coarse-grained model of the DNA has been used to extract the feature vectors for the DNA bases. A deep neural network (NN) is trained with the feature vector as input and the DFT-calculated electronic coupling as output. Once well trained, the NN can predict the DFT-calculated electronic coupling of new structures with a mean absolute error (MAE) of 0.02 eV.

INTRODUCTION

Charge transport properties of deoxyribonucleic acid (DNA), the carrier of genetic information, is of immense interest in biology and molecular electronics.[1−3] The role of DNA mediated charge transport in various biological processes is now well established,[4] and the potential of DNA in different molecular electronics applications has been explored in several studies.[5] As far as the charge transport mechanism is concerned, there is a debate, and various mechanisms have been proposed that can describe the charge transport for different DNA lengths[6−10] of varying base composition. For short DNA length (<4 base pairs), the coherent tunneling mechanism is generally prescribed, while for longer DNA (>10 base pairs), the incoherent hopping mechanism is stipulated.[3,9−12] In between these two extreme charge transport regimes, the coherent corrected hopping mechanism and coherent tunneling with incoherent corrections are also known to produce a good match with the experimental results for different cases.[13] Regardless of the transport mechanism, the electronic coupling between the DNA bases is an important parameter, which dictates the charge transport efficiency in DNA.[9−11,13] Accurate estimation of the electronic coupling between two DNA bases requires first principle calculations. The computation requires three independent density functional theory (DFT) calculations to estimate the electronic coupling between two DNA bases.[9,10,14,15] DNA in solution fluctuates between many different configurations, and any charge transport parameter in DNA has to be averaged over

Received: May 15, 2020
Revised: August 30, 2020
Published: September 2, 2020

https://dx.doi.org/10.1021/acs.jpca.0c04368
J. Phys. Chem. A 2020, 124, 7658–7664

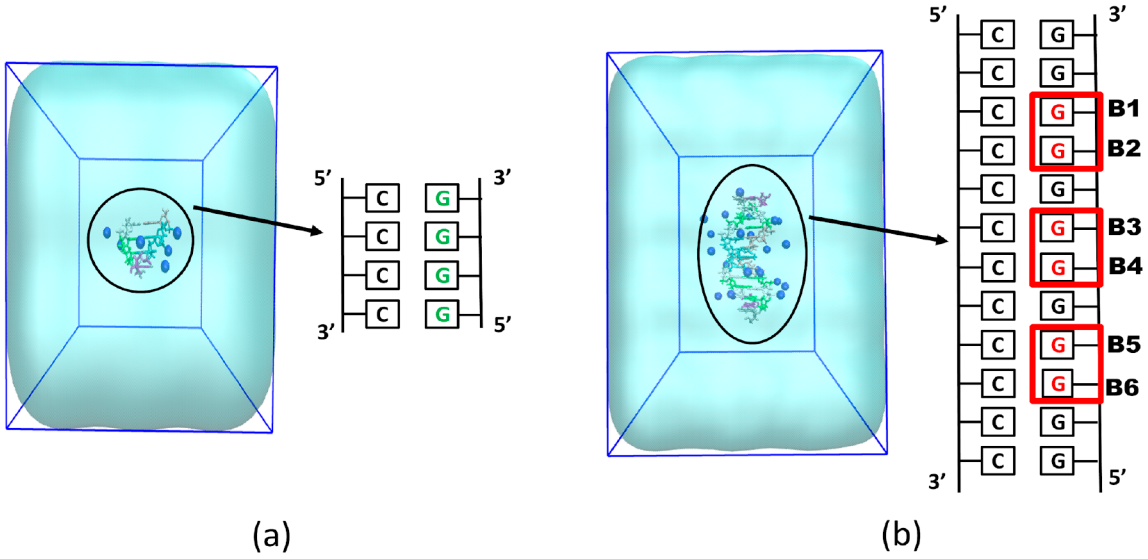


Figure 1. (a) Atomistic model of the 4 bp long DNA, solvated in water used in MD simulations. The surrounding water molecules of the DNA are not shown in full atomistic details but as a continuous media for better clarity. A schematic diagram describing the base pair (bp) arrangements of the 4bp long DNA is also shown. The electronic coupling was calculated between the neighboring guanine (G) bases of the DNA as highlighted in green. (b) Atomistic model of the 12 bp long DNA, solvated in water used in our MD simulations. Here also the surrounding water molecules of the DNA are not shown in full atomistic details for better clarity. The bp sequence of the DNA is shown in the schematic diagram. The electronic couplings are calculated between the neighboring G bases marked as B1−B2, B3−B4, and B5−B6 respectively.

hundreds of such DNA structures.[9,10,16−20] Therefore, the number of DFT calculations required to get a reasonable estimate of electronic coupling become very large[9,10,13,19,21,22] and hence computationally very demanding. It is worth mentioning that there also exists the approximate density functional method, SCC-DFTB,[23,24] which provides the electronic coupling values of the DNA bases with high accuracy (with the DFT) but at a comparatively low computational cost[25] (∼2−3 orders of magnitude faster than the standard HF/DFT methods).

In this paper, we propose a machine learning (ML) based model to overcome this difficulty. In the past few years, ML approaches have been successfully used to address novel scientific problems in almost all branches of research.[26−35] In the field of material science and chemistry, various ML models have been developed to bypass the computationally expensive ab initio calculations. For example, ML has been used to construct potential energy surfaces with first-principles accuracy[36] and to develop atomistic force field parameters[37] as well as to identify and predict electronic correlation energies.[38] From drug design, discovery and testing[39] to the prediction of various properties of metal organic frameworks,[40,41] the applications of ML methods are immense. Recently, a few works have attempted to use ML to predict the charge transport properties of various materials.[42−45] Korol et al.[43] predicted the DNA charge transport properties using a NN architecture trained on short DNA sequences. Caylak et al.[44] predicted the electronic coupling for the amorphous morphology of Alq3 using a genetic algorithm approach. Lederar et al.[45] used relative orientation and geometries of pentacene molecules in a crystal to determine the electronic couplings between them. Similarly, Wang et al.[46] successfully predicted the electronic coupling values between ethylene dimers using the kernel ridge regression (KRR) technique. In this work, we present a ML model to predict the electronic coupling between two neighboring guanine (G) bases (in the same strand) of DNA (with the sequence of the form d(..GGGG..)) of various lengths. The reasons for choosing this particular base pair and the sequence combinations for our study are as follows: The G-G electronic coupling is the highest (dominates the charge transport) among all the possible couplings between different DNA bases (see Figure S1 in Supporting Information), and the d(..GGGG..) sequence is one of the most studied sequences in the molecular electronics literature.[14,47] However, the ML models presented in this article are very general and can easily be extended to any arbitrary DNA sequence, to predict electronic coupling between any base pair combinations.

We perform all atom molecular dynamics (MD) simulations to sample various realistic DNA morphologies in the aqueous solution. The electronic coupling between the DNA bases are calculated further using density functional theory (DFT) calculations with the MD-derived DNA morphologies. A new and simple protocol based on the coarse-grained model of the DNA has been used to extract the feature vectors for the DNA bases. A deep neural network (NN) is trained with the feature vectors as input and the DFT-calculated electronic coupling as output. Once trained, the NN can predict the DFT-calculated electronic coupling of a new structure with a mean absolute error (MAE) of 0.02 eV. The features of all the earlier ML models[44−46] to predict the electronic coupling are based on the complete atomistic description of the respective molecules. In contrast, features of our ML model are based on the coarse grain description of the DNA bases. We show that the relative position and orientation of the bases are more important in learning the electronic coupling rather than relative position of the individual atoms. The physical distance between the coarse grained atoms of the bases quite well capture their relative position and orientation, resulting in a reasonably accurate ML model.

METHODOLOGY AND RESULTS

MD Simulation. To sample the different possible DNA morphologies in an aqueous solution, we perform MD

https://dx.doi.org/10.1021/acs.jpca.0c04368
J. Phys. Chem. A 2020, 124, 7658−7664

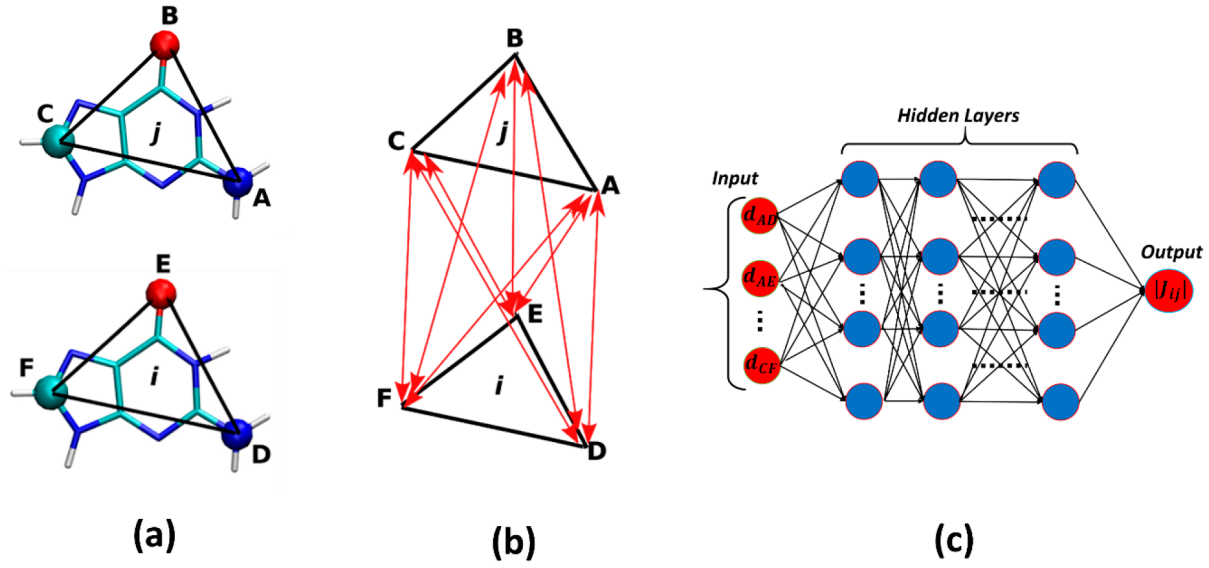


Figure 2. (a) Coarse-grained representation of the guanine (G) bases of the DNA. In the coarse-grained representation, the atomistic structure of the nucleobase is replaced by a triangle connecting three atoms (ABC and DEF) in the system. (b) The main feature of the NN is the distance of the lines connecting vertices of one triangle to the other ones as shown by red two-way arrows. (c) Schematic of a standard NN used to fit the electronic coupling ($|J_{ij}|$). The NN uses the feature vector as input and predict $|J_{ij}|$ as output. There are a few hidden layers between input and output with the nodes represented by the circles.

simulations of the DNA. The initial structures of the DNAs (4 base pair and 12 base pair) were built using the Nucleic Acid Builder (NAB) toolkit available in the AMBER module.[48] To prepare the system for the MD runs, the DNA was first solvated in a TIP3P[49] water box of dimension 15 Å in each direction from the DNA ends. Na^+ counterions were added to charge neutralize the full system. The initial systems prepared for the MD runs are shown in parts a and b of Figure 1. After initial energy minimization and heating, a production run of 100 ns was performed. The production runs were performed in NPT ensemble with the pressure and temperature maintained using Berendsen pressure and temperature coupling algorithm.[50] All the MD simulations were done with the OL15 force field[51,52] in the AMBER[53] software package. Several independent MD simulations for the DNA (4 base pair) were performed at various temperatures ranging from 170 to 300 K to allow the DNA to have different structural fluctuations. For the 12 base pairs (bp) DNA, simulation was performed only at 300 K. The elasticity of the DNA molecules[54−57] depends on the length of the DNA. Therefore, the DNA having different lengths may have different structural fluctuations. As we will describe in the next section of the manuscript, the training data set for the NN will be generated with the different configurations of 4 base pairs long DNA at different temperatures. We will show that this training is sufficient to predict the base pair couplings of longer (12 base pair) DNA. Therefore, we simulate both 4 and 12 base pairs long DNA. The MD simulation protocol followed in this work is described in great detail in our earlier publications.[9,10,14,58] From all these independent simulations, we chose different uncorrelated DNA structures from the last 10 ns of the production run. From these structures, we cut the neighboring DNA bases out of the backbone and calculate the electronic coupling between them.

Electronic Coupling. The electronic coupling between two DNA bases i and j is defined as $J_{ij}=\langle \phi^i |\hat{H}| \phi^j \rangle$. Here ϕ^i and ϕ^j are the diabatic wave functions localized on bases i and j and, respectively. $\hat{H}$ is the Hamiltonian for this two-site system. In this paper, we are interested in the electronic coupling between the HOMO of the DNA bases, since the charge transport is known to happen through these frontier orbitals.[47,59,60] So, ϕ^i and ϕ^j will be chosen as HOMO of the bases i and j respectively. The electronic coupling J_{ij} can be further rewritten as[61] $J_{ij} = \sum_n \gamma_n^i \epsilon_n \gamma_n^j$. Here ϵ_n are the eigenvalues of the Hamiltonian $\hat{H}$ corresponding to the wave function $|\psi_n\rangle$. $\gamma_n^i = \langle \phi^i | \psi_n \rangle$ and $\gamma_n^j = \langle \phi^i | \psi_n \rangle$. Therefore, it requires three DFT calculation to evaluate J_{ij}:

1. Calculation of wave function ϕ^i of the base i.
2. Calculation of wave function ϕ^j of the base j.
3. Calculation of eigenvalues ϵ_n and corresponding wave function $|\psi_n\rangle$ of the combined system i and j.

The ϕ^i, ϕ^j, ψ_n, and ϵ_n are calculated using Gaussian 09[62] using MO6-2X[63] functional and 6-311G basis set. These quantities are used to calculate the overlap matrix elements γ_n^i and γ_n^j and eventually J_{ij} using the code VOTCA-CTP.[62] In this paper, we have used the DFT (MO6-2X/6-311G) calculated electronic coupling to train our ML model as a proof of concept. On the contrary, one can also train the ML model based on the calculation using another, DFT based method, DFT-B, which provides quite accurate electronic coupling as shown by Kubar et el.[23]

Training a Deep Neural Network (NN). To generate the features/descriptors of the nucleobases for the NN training, we first move to a coarse grained representation (see Figure 2a) of the bases as suggested by Cragnolini et al.[64] and several others.[65,66] In the coarse-grained representation, the atomistic structure of the nucleobase is replaced by a triangle connecting three atoms (ABC and DEF) of the base as shown in Figure 2a. The lines connecting the vertices of one triangle to the other ones (shown by red two-way arrows in Figure 2b) will be used as the features to train the NN, which predicts the electronic coupling between these nucleobases. Therefore, the feature is simply a nine dimensional vector having components d_{AD}, d_{AE}, d_{AF}, d_{BD}, d_{BE}, d_{BF}, d_{CD}, d_{CE}, and d_{CF}. Here d's are the

https://dx.doi.org/10.1021/acs.jpca.0c04368
J. Phys. Chem. A 2020, 124, 7658−7664

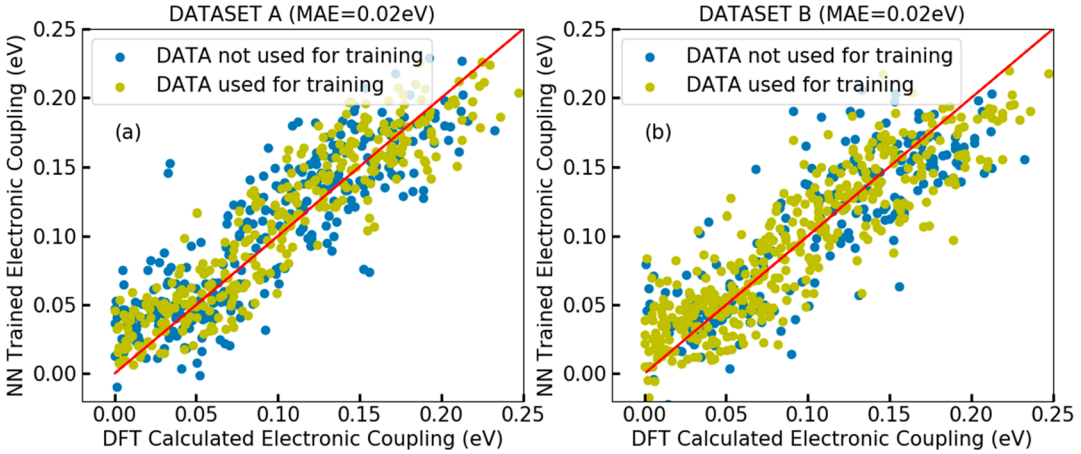


Figure 3. Correlation between the NN-predicted absolute values of the electronic coupling vs the DFT-calculated ones for two different data sets A (a) and B (b). When the NN predicts an electronic coupling exactly equal to the DFT predicted value, the data points fall on top of the red straight line. When given unknown data (not used for training) as input, the NN predicts the electronic coupling with an accuracy of 0.02 eV (MAE) of the actual DFT-calculated value.

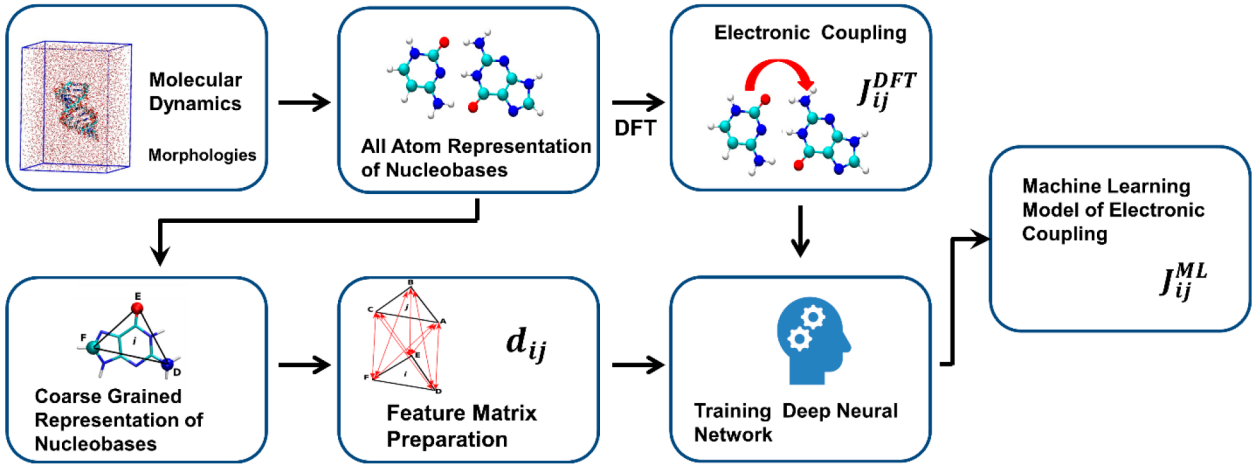


Figure 4. Workflow to generate ML-based models to predict electronic coupling between the DNA bases. MD simulations are performed to generate DNA morphologies. The nucleobases are extracted from the DNA structures removing the DNA backbone. The DFT calculations are performed to calculate the electronic coupling. The all atom model of the DNA are coarse grained and feature vectors are prepared. Deep NN are trained with the DFT-calculated results as the output and the feature vector as input. Trained NN are finally used to predict the output for a completely unknown input.

distances of the lines segments connecting the vertices of the two triangles as shown in Figure 2b.

It is worth mentioning that Wang et al.[46] used a coulomb matrix as the main feature in an attempt to generate a ML-based model to calculate electronic coupling of ethylene. To develop a ML-based model for electronic coupling of pentacene, Lederer et al.[45] used the angle between the symmetry axes as the feature vector. We perform ∼600 electronic coupling calculations (1800 DFT calculations) between the nucleobases of the 4 bp DNA with the structures obtained from MD simulations performed at different temperatures ranging from 180 to 300 K as described in a previous section of the paper. From these calculations, we built two different data sets: data set A and data set B. In data set A, 50% of the total data are randomly chosen as training data and the rest of the data are kept as test data. Data set A includes DNA configurations generated at various temperatures ranging from 180 to 300 K. In data set B, the test set is built with the electronic coupling calculations coming from the structures obtained from MD simulations at 180, 210, and 300 K. The rest of the data are used as a training data. Data set B is built in such a fashion, that later we will predict the electronic coupling values of the 4 bp DNA bases at these temperatures (180, 210, and 300 K), based on the training at other temperatures. Both data sets A and B are provided in the Supporting Information of the manuscript. NN having 3 hidden layers with 50, 25, 25 nodes was trained using the training data of data set A and data set B. The "Relu" activation function was used for all three fully connected hidden layers. Architecture of the NN used to fit the electronic coupling is shown schematically in Figure 2c. The efficiency of the training was evaluated by comparing the DFT-calculated electronic coupling and the NN-predicted values for two different data sets: used for NN training (training data) and not used for NN training (test data). As shown in Figure 3, parts a and b, the NN predicts unknown data with a mean absolute error (MAE) of 0.02 eV for both data sets A and B. It is worth mentioning that this MAE value is quite similar to the one achieved by Lederer et al.[45]

To further demonstrate the generality of our prepared data set to predict the electronic coupling (between the bases) of

https://dx.doi.org/10.1021/acs.jpca.0c04368
J. Phys. Chem. A 2020, 124, 7658−7664

another B-DNA structures, we calculated the possible ranges of different DNA structural parameters in our data set and the average values of these parameters for the B-DNA (at 300 K) as shown in Table S1 in the Supporting Information. It is evident from Table S1 (see Supporting Information) that the ranges of the structural parameters in our data set are wide enough for the prediction of electronic coupling between the bases of the any other B-DNA morphology.

Prediction of the Trained Neural Network (NN). Next, we use our trained NN to predict the results for a completely unknown input. As mentioned in the previous section, we use the data set B here for the training of the NN. The complete workflow to generate the ML-based model of the electronic coupling is shown schematically in Figure 4.

We predict the electronic coupling between the neighboring guanine (G) bases of the same DNA strand for six different cases and compare them with DFT-calculated values as presented in Table 1 below. Electronic coupling of the 4 bp long DNA was predicted at three different temperatures: 180, 210, and 300 K. Twenty-five MD generated structures were taken for the DFT calculations, and the same structures were used to predict the electronic coupling using the ML-based model. DFT calculations and ML prediction were done for all neighboring G bases of all 25 snapshots, and root-mean-square electronic couplings (rmsJ) were calculated, which is commonly reported in the molecular electronics literature.[14,47] We also predict the electronic coupling between the G bases of a 12 bp DNA at 300 K. In this case, the electronic coupling was calculated between the neighboring G bases selected from three different positions of the DNA as shown in the schematic diagram of Figure 1b. The DFT calculations were performed with the B1−B2 (see Figure 1b) bases for 25 snapshots generated from MD and rmsJ were calculated. These same structures of the B1−B2 bases were used for ML-based prediction too. All the calculations were repeated for B3−B4 bases and B5−B6 bases. The fluctuations in various parts of the DNA may be different. We have shown in our previous work[58] that the terminal bases of the DNA fluctuate more than the bases in the middle, giving rise to different base pair couplings. Therefore, we took bases from the various parts of the DNA.

The NN predicts rmsJ values quite accurately as shown in Table 1 below. The NN training and prediction are repeated for 30 times by changing the seed of the random number generator and the error bars in the NN prediction are estimated (see Table 1). We used freely available code KERAS[67,68] for all the NN-based calculations.

Table 1. Comparison of NN-Predicted rmsJ and DFT-Calculated rmsJ for Different Systems[a]

system	NN-predicted rmsJ (eV)	DFT-calculated rmsJ (eV)
system 1[b]	0.123 ± 0.005	0.122
system 2[c]	0.138 ± 0.004	0.136
system 3[d]	0.105 ± 0.004	0.114
system 4[e]	0.102 ± 0.005	0.100
system 5[f]	0.102 ± 0.005	0.092
system 6[g]	0.130 ± 0.005	0.126

[a]The DFT-calculated results for these systems were not included in the NN training. [b]4 bp long DNA system at a temperature of 180 K. [c]4 bp long DNA system at a temperature of 210 K. [d]4 bp long DNA system at a temperature of 300 K. [e]12 bp long DNA system at a temperature of 300 K. Electronic coupling was calculated between the bases B1 and B2 as depicted in Figure 1b.[f]12 bp long DNA system at a temperature of 300 K, G-G coupling was calculated between the base B3 and B4 as depicted in Figure 1b. [g]12 bp long DNA system at a temperature of 300 K the base, G-G coupling was calculated between the bases B5 and B6 as depicted in Figure 1b.

SUMMARY AND CONCLUSION

In the past few years, ML approaches have been proven to be extremely powerful tools to address key scientific problems in almost all branches of research.[26−29,33−35,39] In this article, we report a ML model to predict the electronic coupling between the DNA bases with an accuracy of 0.02 eV (MAE), thanks to deep NN. Our protocol uses a coarse-grained model of the DNA nucleobases to extract the feature vectors for the deep NN training and predict the electronic coupling as output within a computation time that is several orders of magnitude (∼10−10^4 times) lower than the computation time required for the DFT calculations. We have used a new molecular featurization technique for our ML model, which is simple and can be easily generalized to any new molecular system.

Charge transport is one of the key processes to be engineered in organic semiconductors for their applications in optoelectronic devices (solar cell, LED etc.). Therefore, a robust and transferable ML model to compute the charge transport parameters bypassing expensive DFT calculations will have significant impact in the area of molecular electronics. In future, we wish to generalize our ML model to predict the charge transport parameters of more disordered system like dendrimer melts.[69]

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpca.0c04368.

Root mean square electronic coupling between the bases of the DNA with the sequence d(GGGG) and ranges of DNA structural parameters in the ML data set and the average value of these structural parameters for the B-DNA (PDF)

NN set A (TXT)

NN set B (TXT)

NN set C (TXT)

NN set D (TXT)

AUTHOR INFORMATION

Corresponding Author

Prabal K. Maiti − *Center for Condensed Matter Theory, Department of Physics, Indian Institute of Science, Bangalore 560012, India;* orcid.org/0000-0002-9956-1136; Phone: 091-80-2293-2865; Email: maiti@iisc.ac.in

Authors

Saientan Bag − *Center for Condensed Matter Theory, Department of Physics, Indian Institute of Science, Bangalore 560012, India;* orcid.org/0000-0003-1000-7719

Abhishek Aggarwal − *Center for Condensed Matter Theory, Department of Physics, Indian Institute of Science, Bangalore 560012, India;* orcid.org/0000-0002-3342-6484

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpca.0c04368

Notes

The authors declare no competing financial interest.

https://dx.doi.org/10.1021/acs.jpca.0c04368
J. Phys. Chem. A 2020, 124, 7658−7664

■ ACKNOWLEDGMENTS

We thank DST, India, for financial support.

■ REFERENCES

(1) O'Brien, E.; Holt, M. E.; Thompson, M. K.; Salay, L. E.; Ehlinger, A. C.; Chazin, W. J.; Barton, J. K. The [4Fe4S] cluster of human DNA primase functions as a redox switch using DNA charge transport. *Science* **2017**, *355*, No. eaag1789.

(2) Genereux, J. C.; Boal, A. K.; Barton, J. K. DNA-mediated charge transport in redox sensing and signaling. *J. Am. Chem. Soc.* **2010**, *132*, 891−905.

(3) Genereux, J. C.; Barton, J. K. Mechanisms for DNA charge transport. *Chem. Rev.* **2010**, *110*, 1642−1662.

(4) Bartels, P. L.; Zhou, A.; Arnold, A. R.; Nuñez, N. N.; Crespilho, F. N.; David, S. S.; Barton, J. K. Electrochemistry of the [4Fe4S] cluster in base excision repair proteins: Tuning the redox potential with DNA. *Langmuir* **2017**, *33*, 2523−2530.

(5) Artés, J. M.; López-Martínez, M.; Díez-Pérez, I.; Sanz, F.; Gorostiza, P. Nanoscale charge transfer in redox proteins and DNA: Towards biomolecular electronics. *Electrochim. Acta* **2014**, *140*, 83−95.

(6) Wolter, M.; Elstner, M.; Kleinekathöfer, U.; Kubař, T. S. Microsecond Simulation of Electron Transfer in DNA: Bottom-Up Parametrization of an Efficient Electron Transfer Model Based on Atomistic Details. *J. Phys. Chem. B* **2017**, *121*, 529−549.

(7) Beratan, D. N.; Skourtis, S. S. Electron transfer mechanisms. *Curr. Opin. Chem. Biol.* **1998**, *2*, 235−243.

(8) Nitzan, A.; Ratner, M. A. Electron transport in molecular wire junctions. *Science* **2003**, *300*, 1384−1389.

(9) Bag, S.; Mogurampelly, S.; Goddard III, W. A.; Maiti, P. K. Dramatic changes in DNA conductance with stretching: structural polymorphism at a critical extension. *Nanoscale* **2016**, *8*, 16044−16052.

(10) Aggarwal, A.; Bag, S.; Maiti, P. K. Remarkable similarity of force induced dsRNA conformational changes to stretched dsDNA and their detection using electrical measurements. *Phys. Chem. Chem. Phys.* **2018**, *20*, 28920−28928.

(11) Jortner, J.; Bixon, M.; Langenbacher, T.; Michel-Beyerle, M. E. Charge transfer and transport in DNA. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 12759−12765.

(12) Berlin, Y. A.; Burin, A. L.; Ratner, M. A. Charge hopping in DNA. *J. Am. Chem. Soc.* **2001**, *123*, 260−268.

(13) Renaud, N.; Berlin, Y. A.; Lewis, F. D.; Ratner, M. A. Between superexchange and hopping: An intermediate charge-transfer mechanism in poly (A)-poly (T) DNA hairpins. *J. Am. Chem. Soc.* **2013**, *135*, 3953−3963.

(14) Bag, S.; Maiti, P. K. Tuning molecular fluctuation to boost the conductance in DNA based molecular wires. *Phys. Chem. Chem. Phys.* **2019**, *21*, 23514−23520.

(15) Rühle, V.; Lukyanov, A.; May, F.; Schrader, M.; Vehoff, T.; Kirkpatrick, J.; Baumeier, B. r.; Andrienko, D. Microscopic simulations of charge transport in disordered organic semiconductors. *J. Chem. Theory Comput.* **2011**, *7*, 3335−3345.

(16) Troisi, A.; Nitzan, A.; Ratner, M. A. A rate constant expression for charge transfer through fluctuating bridges. *J. Chem. Phys.* **2003**, *119*, 5782−5788.

(17) Berlin, Y. A.; Grozema, F. C.; Siebbeles, L. D.; Ratner, M. A. Charge transfer in donor-bridge-acceptor systems: Static disorder, dynamic fluctuations, and complex kinetics. *J. Phys. Chem. C* **2008**, *112*, 10988−11000.

(18) Woiczikowski, P. B.; Kubař, T.; Gutiérrez, R.; Caetano, R. A.; Cuniberti, G.; Elstner, M. Combined density functional theory and Landauer approach for hole transfer in DNA along classical molecular dynamics trajectories. *J. Chem. Phys.* **2009**, *130*, 215104.

(19) Grozema, F. C.; Tonzani, S.; Berlin, Y. A.; Schatz, G. C.; Siebbeles, L. D.; Ratner, M. A. Effect of structural dynamics on charge transfer in DNA hairpins. *J. Am. Chem. Soc.* **2008**, *130*, 5157−5166.

(20) Jang, S.; Newton, M. D. Theory of torsional non-Condon electron transfer: A generalized spin-boson Hamiltonian and its nonadiabatic limit solution. *J. Chem. Phys.* **2005**, *122*, 024501.

(21) Gutierrez, R.; Caetano, R.; Woiczikowski, B.; Kubar, T.; Elstner, M.; Cuniberti, G. Charge transport through biomolecular wires in a solvent: bridging molecular dynamics and model Hamiltonian approaches. *Phys. Rev. Lett.* **2009**, *102*, 208102.

(22) Gutiérrez, R.; Caetano, R.; Woiczikowski, P.; Kubar, T.; Elstner, M.; Cuniberti, G. Structural fluctuations and quantum transport through DNA molecular wires: a combined molecular dynamics and model Hamiltonian approach. *New J. Phys.* **2010**, *12*, 023022.

(23) Kubař, T. s.; Woiczikowski, P. B.; Cuniberti, G.; Elstner, M. Efficient calculation of charge-transfer matrix elements for hole transfer in DNA. *J. Phys. Chem. B* **2008**, *112*, 7937−7947.

(24) Song, B.; Elstner, M.; Cuniberti, G. Anomalous conductance response of DNA wires under stretching. *Nano Lett.* **2008**, *8*, 3217−3220.

(25) Cui, Q.; Elstner, M. Density functional tight binding: values of semi-empirical methods in an ab initio era. *Phys. Chem. Chem. Phys.* **2014**, *16*, 14368−14377.

(26) Carrasquilla, J.; Melko, R. G. Machine learning phases of matter. *Nat. Phys.* **2017**, *13*, 431−434.

(27) Radovic, A.; Williams, M.; Rousseau, D.; Kagan, M.; Bonacorsi, D.; Himmel, A.; Aurisano, A.; Terao, K.; Wongjirad, T. Machine learning at the energy and intensity frontiers of particle physics. *Nature* **2018**, *560*, 41−48.

(28) Rupp, M. Machine learning for quantum mechanics in a nutshell. *Int. J. Quantum Chem.* **2015**, *115*, 1058−1073.

(29) Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine learning for molecular and materials science. *Nature* **2018**, *559*, 547−555.

(30) Ramakrishnan, R.; von Lilienfeld, O. A. Machine learning, quantum chemistry, and chemical space. *Reviews in computational chemistry* **2017**, *30*, 225−256.

(31) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; von Lilienfeld, O. A. Big data meets quantum chemistry approximations: The Δ-machine learning approach. *J. Chem. Theory Comput.* **2015**, *11*, 2087−2096.

(32) Wu, Z.; Ramsundar, B.; Feinberg, E. N.; Gomes, J.; Geniesse, C.; Pappu, A. S.; Leswing, K.; Pande, V. MoleculeNet: a benchmark for molecular machine learning. *Chemical science* **2018**, *9*, 513−530.

(33) Cleophas, T. J.; Zwinderman, A. H.; Cleophas-Allers, H. I.: *Machine learning in medicine*; Springer, 2013; Vol. *9*.

(34) Deo, R. C. Machine learning in medicine. *Circulation* **2015**, *132*, 1920−1930.

(35) Rajkomar, A.; Dean, J.; Kohane, I. Machine learning in medicine. *N. Engl. J. Med.* **2019**, *380*, 1347−1358.

(36) Quaranta, V.; Behler, J. r.; Hellström, M. Structure and Dynamics of the Liquid−Water/Zinc-Oxide Interface from Machine Learning Potential Simulations. *J. Phys. Chem. C* **2019**, *123*, 1293−1304.

(37) Friederich, P.; Konrad, M.; Strunk, T.; Wenzel, W. Machine learning of correlated dihedral potentials for atomistic molecular force fields. *Sci. Rep.* **2018**, *8*, 1−8.

(38) Welborn, M.; Cheng, L.; Miller III, T. F. Transferability in machine learning for electronic structure via the molecular orbital basis. *J. Chem. Theory Comput.* **2018**, *14*, 4772−4779.

(39) Réda, C.; Kaufmann, E.; Delahaye-Duriez, A. Machine learning applications in drug development. *Comput. Struct. Biotechnol. J.* **2020**, *18*, 241−252.

(40) Wu, X.; Xiang, S.; Su, J.; Cai, W. Understanding Quantitative Relationship between Methane Storage Capacities and Characteristic Properties of Metal−Organic Frameworks Based on Machine Learning. *J. Phys. Chem. C* **2019**, *123*, 8550−8559.

(41) Fanourgakis, G. S.; Gkagkas, K.; Tylianakis, E.; Klontzas, E.; Froudakis, G. A Robust Machine Learning Algorithm for the Prediction of Methane Adsorption in Nanoporous Materials. *J. Phys. Chem. A* **2019**, *123*, 6080−6087.

https://dx.doi.org/10.1021/acs.jpca.0c04368
J. Phys. Chem. A 2020, 124, 7658−7664

(42) Lauritzen, K. P.; Magyarkuti, A.; Balogh, Z.; Halbritter, A.; Solomon, G. C. Classification of conductance traces with recurrent neural networks. *J. Chem. Phys.* **2018**, *148*, 084111.
(43) Korol, R.; Segal, D. Machine Learning Prediction of DNA Charge Transport. *J. Phys. Chem. B* **2019**, *123*, 2801−2811.
(44) Caylak, O.; Yaman, A.; Baumeier, B. r. Evolutionary approach to constructing a deep feedforward neural network for prediction of electronic coupling elements in molecular materials. *J. Chem. Theory Comput.* **2019**, *15*, 1777−1784.
(45) Lederer, J.; Kaiser, W.; Mattoni, A.; Gagliardi, A. Machine Learning−Based Charge Transport Computation for Pentacene. *Advanced Theory and Simulations* **2019**, *2*, 1800136.
(46) Wang, C.-I.; Braza, M. K. E.; Claudio, G. C.; Nellas, R. B.; Hsu, C.-P. Machine Learning for Predicting Electron Transfer Coupling. *J. Phys. Chem. A* **2019**, *123*, 7792−7802.
(47) Liu, C.; Xiang, L.; Zhang, Y.; Zhang, P.; Beratan, D. N.; Li, Y.; Tao, N. Engineering nanometre-scale coherence in soft matter. *Nat. Chem.* **2016**, *8*, 941−945.
(48) Case, D.; Darden, T.; Cheatham, T., III; Simmerling, C.; Wang, J.; Duke, R.; Luo, R.; Crowley, M.; Walker, R.; Zhang, W. *Amber Tools*; University of California: San Francisco, CA 2008.
(49) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.
(50) Berendsen, H. J.; Postma, J. v.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(51) Zgarbová, M.; Luque, F. J.; Šponer, J. í.; Cheatham, T. E., III; Otyepka, M.; Jurecka, P. Toward improved description of DNA backbone: revisiting epsilon and zeta torsion force field parameters. *J. Chem. Theory Comput.* **2013**, *9*, 2339−2354.
(52) Zgarbová, M.; Sponer, J.; Otyepka, M.; Cheatham III, T. E.; Galindo-Murillo, R.; Jurecka, P. Refinement of the Sugar−Phosphate backbone torsion beta for AMBER force fields improves the description of Z-and B-DNA. *J. Chem. Theory Comput.* **2015**, *11*, 5723−5736.
(53) Case, D.; Darden, T.; Cheatham, T., III; Simmerling, C.; Wang, J.; Duke, R.; Luo, R.; Walker, R.; Zhang, W.; Merz, K. *AMBER 12*; University of California: San Francisco, 2012.
(54) Aggarwal, A.; Naskar, S.; Sahoo, A. K.; Mogurampelly, S.; Garai, A.; Maiti, P. K. What do we know about DNA mechanics so far? *Curr. Opin. Struct. Biol.* **2020**, *64*, 42−50.
(55) Garai, A.; Saurabh, S.; Lansac, Y.; Maiti, P. K. DNA elasticity from short DNA to nucleosomal DNA. *J. Phys. Chem. B* **2015**, *119*, 11146−11156.
(56) Mogurampelly, S.; Nandy, B.; Netz, R. R.; Maiti, P. K. Elasticity of DNA and the effect of dendrimer binding. *Eur. Phys. J. E: Soft Matter Biol. Phys.* **2013**, *36*, 68.
(57) Garai, A.; Ghoshdastidar, D.; Senapati, S.; Maiti, P. K. Ionic liquids make DNA rigid. *J. Chem. Phys.* **2018**, *149*, 045104.
(58) Garai, A.; Mogurampelly, S.; Bag, S.; Maiti, P. K. Over-stretching of B-DNA with various pulling protocols: Appearance of structural polymorphism and S-DNA. *J. Chem. Phys.* **2017**, *147*, 225102.
(59) Artés, J. M.; Li, Y.; Qi, J.; Anantram, M.; Hihath, J. Conformational gating of DNA conductance. *Nat. Commun.* **2015**, *6*, 1−8.
(60) Wong, J. R.; Shao, F. Hole Transport in A-form DNA/RNA Hybrid Duplexes. *Sci. Rep.* **2017**, *7*, 1−8.
(61) Baumeier, B.; Kirkpatrick, J.; Andrienko, D. Density-functional based determination of intermolecular charge transfer properties for large-scale morphologies. *Phys. Chem. Chem. Phys.* **2010**, *12*, 11103−11113.
(62) Frisch, M.; Trucks, G.; Schlegel, H.; Scuseria, G.; Robb, M.; Cheeseman, J.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. *Gaussian 09* Rev. D. 01, Gaussian Inc.: Wallingford CT. See also: http://www.gaussian.com, 2009.
(63) Zhao, Y.; Truhlar, D. G. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, non-covalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **2008**, *120*, 215−241.
(64) Cragnolini, T.; Derreumaux, P.; Pasquali, S. Coarse-grained simulations of RNA and DNA duplexes. *J. Phys. Chem. B* **2013**, *117*, 8047−8060.
(65) Bell, D. R.; Cheng, S. Y.; Salazar, H.; Ren, P. Capturing RNA folding free energy with coarse-grained molecular dynamics simulations. *Sci. Rep.* **2017**, *7*, 1−14.
(66) Vangaveti, S.; D'Esposito, R.; Lippens, J.; Fabris, D.; Ranganathan, S. A coarse-grained model for assisting the investigation of structure and dynamics of large nucleic acids by ion mobility spectrometry−mass spectrometry. *Phys. Chem. Chem. Phys.* **2017**, *19*, 14937−14946.
(67) Ketkar, N. Introduction to keras. *Deep learning with Python* **2017**, 97−111.
(68) Géron, A. *Hands-On Machine Learning with Scikit-Learn, Keras, and Tensor Flow: Concepts, Tools, and Techniques to Build Intelligent Systems*; O'Reilly Media: 2019.
(69) Bag, S.; Jain, M.; Maiti, P. K. Charge Transport in Dendrimer Melts Using Multiscale Modeling Simulation. *J. Phys. Chem. B* **2016**, *120*, 9142−9151.

https://dx.doi.org/10.1021/acs.jpca.0c04368
J. Phys. Chem. A 2020, 124, 7658−7664