

トンネル電流計測 1 分子の核酸塩基種の機械学習識別法の開発

Machine Learning Discrimination of Nucleotide
For Single-Molecule Tunnel-Current Electrical Identification

阪大産研¹ ○大城敬人¹, 鷹合孝之¹, 吉田剛¹, 小本祐貴¹, 鷲尾隆¹, 谷口正輝¹
ISIR-SANKEN, Osaka Univ. ¹Takahito Ohshiro¹, Takayuki Takaai, Takeshi Yoshida, Yuki Komoto¹,
Takashi Washio¹, Masateru Taniguchi¹

E-mail: toshiro@sanken.osaka-u.ac.jp

1. 緒言

高速・高精度かつ低コストに個人の遺伝情報を読み取ることのできる次世代 DNA シーケンサの開発は、近年世界中でしのぎを削っている。本研究室では、単分子 DNA 鎖を微小ギャップ電極によって電気計測する方法論を提案している。微小ギャップ電極で得られる 1 分子シグナルの取得は、コンダクタンスの時間変化を高速計測して得られる。これまで核酸塩基をコンダクタンスの値が、分子の持つ HOMO レベルとの傾向と同様であった識別可能であることを示している[1-5]。

本研究では、単分子シーケンシングに向けた核酸塩基の識別を、機械学習によっておこなった。これまで、核酸塩基分子のシグナルを、計測時のノイズシグナルを除去することで 2 つの人工核酸塩基識別能が向上しており、同様に混合核酸塩基種について学習データをもとに評価・検討を行った。

2. 計測・実験条件

計測の用いる nano-MCBI は、薄型シリコン基板上に金線を描画し、酸化シリコンで被覆し、ギャップ付近にナノ流路を配置したものを用いる。金線を機械的に破断し、ピエゾ素子もちいて電極間距離を制御する。トンネル電流のバググランド電流を基準として、0.02 nm 刻みで 0.64-0.56nm の範囲で計測を行った。この範囲で、核酸塩基測定に最適な距離に設定し、電気計測を行った。試料となる核酸塩基(dGMP, dAMP, dCMP, dTMP)は購入したものを用いた。この核酸塩基を含む水溶液 0.2-5.0 uM の濃度を用いた。

3. 結果・考察

ナノ電極を用いて、電極付近を通り過ぎ

る 1 分子に流れるシグナルを、機械学習を用いて、2 塩基（例えば、G と T）、3 塩基（たとえば、G,T,A）、4 塩基の核酸塩基（G,T,A,C）等の混合の組み合わせについて分類を行った。その結果、すべての 2 種の分子種を精度 0.7 以上で分類することに成功した。特に、電流ヒストグラムで近接している A と C の分類において、識別精度（F-measure）は Table 1 となり、その識別能は他の核酸塩基種との識別が可能であることを示している。この識別結果をもとに、分類シグナルの特徴を観察すると、分子構造や電極との相互作用に起因する電流変化を捉えることで、単分子伝導度だけでは区別できない分子種を識別することができることを示唆している。

Table 1

Table 1 Confusion matrix of classifying of dAMP (A) and dCMP (C)

		Predicted molecule	
		A	C
True molecule	A	0.86	0.14
	C	0.20	0.80

関連論文 [1] M. Di Ventra and M. Taniguchi, Nat. Nanotechnol., 11 (2016) 117. [2] M. Tsutsui, M. Taniguchi, K. Yokota, T. Kawai, Nat. Nanotechnol., 5 (2010) 286. [3] T. Ohshiro, K. Matsubara, M. Tsutsui, M. Furuhashi, M. Taniguchi, T. Kawai, Sci. Rep., 2 (2012) 00501. [4] J Am Chem Soc. 2011 133, (2011) 9124. [5] T. Ohshiro, M. Tsutsui, K. Yokota, M. Furuhashi, M. Taniguchi, T. Kawai, Nat. Nanotechnol., 9 (2014) 835.