Int. J. Adv. Eng. Pure Sci. 2021, 33(4):609-613
DOI: 10.7240/jeps.920227

RESEARCH ARTICLE / ARAŞTIRMA MAKALESİ

Türk Ulusal Bilim e-Altyapısı TRUBA'da Moleküler Dinamik Paketi GROMACS Versiyonlarının Performans Optimizasyonu

GROMACS Performance Optimization at Turkish National Grid Resources TRUBA

Büşra SAVAŞ[1,2], Ezgi KARACA[1,2]

[1]*İzmir Biyotıp ve Genom Merkezi, Balçova, 35340, İzmir, Türkiye*
[2]*İzmir Uluslararası Biyotıp ve Genom Enstitüsü, Dokuz Eylül Üniversitesi, Balçova, 35340, İzmir, Türkiye*

Öz
Yüksek performanslı hesaplama sistemlerinin kullanımının artmasıyla, bu sistemlerde çalıştırılan programların performans optimizasyonu öncelikli hale gelmiştir. Bu duruma istinaden, bu çalışmamızda, yaygın olarak kullanılan moleküler dinamik paketi GROMACS'in, TÜBİTAK ULAKBİM tarafından kullanıma sunulan TRUBA hesaplama kümelerindeki en iyi performans kriterlerini bulmayı hedefledik. Performans tarama çalışmamız sırasında, farklı hesaplama kümelerinde, farklı CPU/GPU çekirdek oranı, *thread (iş parçacığı)* sayısı ve GROMACS versiyonlarını denedik. Bu süreç sonunda en iyi performanslı hesaplama kümesi akya-cuda, 40/1 CPU/GPU oranı için en verimli *thread* sayısı 20 ve en hızlı GROMACS versiyonu GROMACS 2020 olarak tespit edildi. Benzer bir çalışma yürütecek araştırmacıların yararlanması adına, performans optimizasyon dosyalarımız ve ayrıntılı sonuçlarımız https://github.com/CSB-KaracaLab/gmx_performance_on_HPC adresinde incelemeye açılmıştır.
Anahtar Kelimeler: Moleküler Dinamik, Yüksek Başarımlı Hesaplama Kümeleri, Optimizasyon, GROMACS

Abstract
With the increasing demand for the high-performance computing (HPC) systems, the optimal usage of HPC has become a central issue. To that end, we aimed to find the optimum system parameters for the commonly used molecular dynamics package, GROMACS, on TRUBA computing clusters, offered by TUBITAK ULAKBIM. For this, we tried different CPU/GPU core ratios, number of threads, and GROMACS versions on different computing clusters. We achieved the optimum performance on the akya-cuda cluster, with 40/1 CPU/GPU ratio, and 20 threads when GROMACS 2020 was used. To stimulate the investigation of similar optimization protocols, we shared our input and output files, together with our performance analysis at https://github.com/CSB-KaracaLab/gmx_performance_on_HPC.
Keywords: Molecular Dynamics, High Performance Computing Systems, Optimization, GROMACS

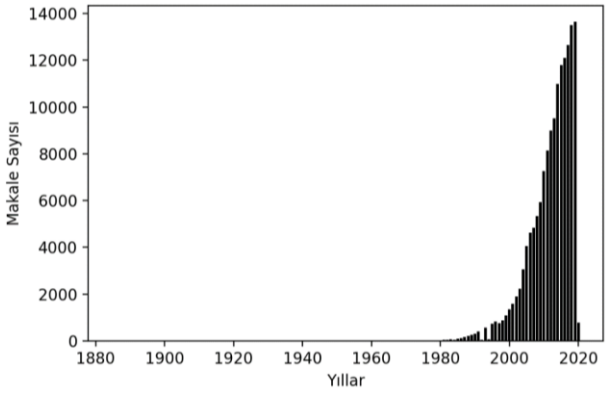
I. GİRİŞ

19. yüzyılın sonlarına doğru geliştirilen moleküler dinamik (MD) algoritmaları, bilgisayar teknolojilerinin hızla gelişmesine bağlı olarak yaygın olarak kullanılmaya başlanmıştır (Şekil 1). Newton hareket denkleminin çözümlenmesiyle ortaya çıkan bu yöntem, öncelikli olarak teorik fizik alanında, daha sonra malzeme bilimi, biyoloji gibi alanlarda kullanılmaya başlanmıştır [1]. İlk gerçekçi moleküler dinamik simülasyonu 1964 yılında sıvı argon için gerçekleştirilmiş ve bu çalışma ile moleküler yapıların anlaşılması adına önemli bir adım atılmıştır [2]. MD simülasyonları, biyolojik moleküllerin belirli sıcaklık, basınç ve tuz konsantrasyonunda gösterdikleri hareketlerini incelememizi, bu şekilde o moleküllerin hücre içindeki fonksiyonları hakkında bilgi sahibi olmamızı sağlar. MD simülasyonlarını yürütmek için pek çok simülasyon paketi bulunmaktadır. Bunlardan en yaygın olarak kullanılanları, GROMACS, NAMD, CHARMM ve AMBER paketleridir [3-6]. MD hesaplarının atom seviyesinde olması büyük bir hesaplama yükü getirir. Dolayısıyla, mevcut kaynaklarda simülasyon yapmadan önce, sistemin maksimum verimle çalıştığından emin olunmalıdır. Biz de bu çalışmamızda, ulusal hesaplama kaynağımız olan TRUBA'da, GROMACS ile hareketlerini anlamak istediğimiz, epigenetik regülasyonda görevli bir protein-DNA kompleks sistemi için en uygun parametreleri bulmayı amaçladık.

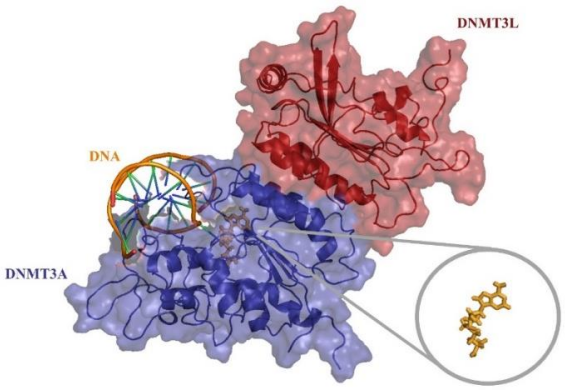
Bir biyoloji terimi olan epigenetik, genetiğin ötesi anlamına gelmektedir. Epigenetik, DNA sekansında bir değişiklik olmadan yürütülen gen-odaklı düzenlemeleri tarif etmektedir [7, 8]. Örnek olarak, epigenetik değişimler sonucunda değişen gen regülasyonu, embriyonik dönemde hücre farklılaşmasını etkiler [9,10]. En yaygın olarak araştırılan epigenetik değişim DNA metilasyonudur. DNA metilasyonu, gen baskılama veya aktifleştirme gibi hayati önem taşıyan biyolojik olaylar için genoma yerleştirilen bir sinyal görevi görür [11-13]. DNA metilasyonu, S-Adenosil metiyonin (SAM) molekülüne bağlı bir metil grubunun, DNA metiltransferaz aracılığıyla CpG (5'-CG-3') nükleotidindeki sitozinin 5-karbonuna kovalent olarak bağlanması ile gerçekleşir (Şekil 2). Memelilerde

Sorumlu yazar: Ezgi KARACA, Tel: +90533 456 60 96, e-posta: ezgi.karaca@ibg.edu.tr
Gönderilme: 19.04.2021, **Düzenleme:** 16.10.2021, **Kabul:** 16.10.2021

sıfırdan (de novo) metilasyon süreci, DNA Metiltransferaz 3 (DNMT3) ailesi tarafından yürütülür [14]. DNMT3 ailesinin üç üyesi bulunmaktadır ve bunlardan sadece ikisi, DNMT3A ve DNMT3B, katalitik aktivite gösterirler [13, 15]. DNMT3L'nin katalitik aktivite göstermemesine rağmen, DNMT3A ve DNMT3B enzimlerine bağlandığında metilasyon reaksiyonunu hızlandırdığı görülmüştür [16]. Bu etkiyi sağlarken DNMT3L, DNA ile direkt bir etkileşim kurmaz ve uzaktan (alosterik) bir etki gösterir. Literatürde DNMT3A-DNMT3L-DNA-SAH yapılarını içeren iki kristal yapı bulunur (PDB Kodu: 6F57, 5YX2). Bu çalışmada, kullanılan yapı heterodimer formundaki 6F57 kodlu yapıdan elde edilmiş olup, heterotetramer formundaki 5YX2 kodlu yapıdan daha az atom sayısına sahip olması nedeniyle performans optimizasyonunda kullanılmak üzere seçilmiştir (Şekil 2). Performans optimizasyon çalışmaları, GROMACS programı kullanılarak gerçekleştirilmiştir.



Şekil 1. Yıllara göre PubMed veritabanında moleküler dinamik anahtar kelimesini içeren makale sayıları



Şekil 2. DNMT3A (mavi), DNMT3L (kırmızı) yapılarının DNA ve SAM (sarı) ile etkileşimlerinin betimlenmesi (PDB Kodu: 6F57)

II. MATERYAL VE METOD

2.1. GROMACS

GROMACS (Groningen Machine for Chemical Simulations), Groningen Üniversitesi tarafından geliştirilmiş bir moleküler dinamik simülasyonu paketidir. GROMACS, moleküler dinamik simülasyonları için yaygın olarak kullanılan AMBER, CHARMM, GROMOS ve OPLS gibi kuvvet alanlarının etkisi altında, moleküllerin hareketlerini simüle eder [4]. GROMACS, CUDA tabanlı GPU kullanımı ile performansın hızlandırılmasını sağlar. CUDA, NVIDIA tarafından geliştirilen bir eklenti olup, GPU gücünden yararlanarak yoğun işlem gereken uygulamaların hızlandırılmasını sağlar.

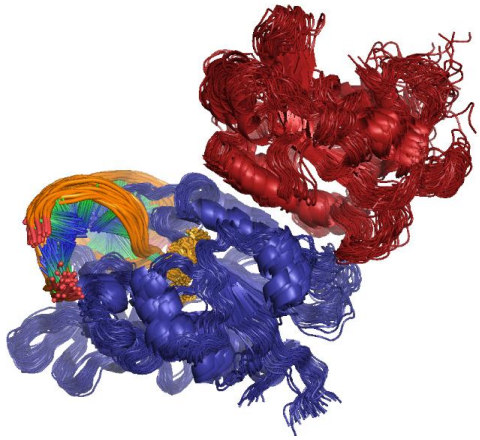
2.2. TRUBA

TRUBA (Türk Ulusal Bilim e-Altyapısı), eski adıyla TRGrid, büyük miktardaki verinin işlenmesine olanak sunan ulusal bir e-Altyapıdır [17]. 19.000 CPU ve 36 GPU ile 1500'den fazla araştırmacıya hizmet vermektedir. Bu çalışmada, DNMT3A-DNMT3L-DNA-SAM yapılarını içeren protein-DNA kompleksinin moleküler dinamik simülasyonları TRUBA kaynaklarındaki farklı hesaplama kümelerinde yürütülmüştür. Bu çerçevede, GPU kullanımına olanak sağlayan akya-cuda, barbun-cuda ve levrekv2-cuda kümeleri kullanılmıştır. Akya-cuda kümesinde, 20 çekirdekli Intel Xeon Scalable 6148 CPU @2.40GHz CPU'dan iki adet bulunmaktadır. Ayrıca, GPU modeli NVIDIA V100 olan dört grafik kartı mevcuttur. Barbun-cuda kümesinin CPU özellikleri ve sayıları akya-cuda ile aynıdır. Ancak, barbun-cuda, akya-cuda'dan farklı olarak, GPU modeli NVIDIA P100 olan iki grafik kartı içermektedir. Levrekv2-cuda kümesinde ise 12 çekirdekli Intel Xeon E5-2680 v3 @2.50GHz CPU'dan iki adet ile NVIDIA M2090 GPU kartından iki adet bulunmaktadır. TRUBA hakkındaki ayrıntılı donanım bilgisi TRUBA'nın wiki sayfasında bulunabilir(https://docs.truba.gov.tr) [18].

2.3. Moleküler Dinamik Simülasyonu

Optimizasyon çalışmaları sırasında kullanılan DNMT3A, DNMT3L ve SAM yapılarını içeren kompleks, BİDEB-2232 (1109B321700106) ve HPC-EUROPA3 (INFRAIA-2016-1-730897) tarafından desteklenen "Epigenetik Metilasyon Mekanizmalarının Yapısal Biyoloji Çerçevesinde Tanımlanması" projesinin sonucunda hazırlanmıştır. DNMT3A, DNMT3L ve SAM yapılarının bulunduğu kompleks 8693 adet atom içermektedir. Simülasyonların hazırlanma sürecinde eklenen su ve iyon molekülleri ile atom sayısı 150.891 olmaktadır. Simülasyon sırasında, protein-DNA kompleksleri ile özelleşen Amber14sb-PARMBSC1 kuvvet alanı ile beraber TIP3P su modeli kullanılmıştır [19]. Bütün simülasyonlar 1.4 nm boyutlu dodekahedron (düzgün on yüzlü) kutu içerisinde TIP3P su molekülleri ve KCl iyonları kullanılarak yürütülmüştür. Simülasyon öncesinde bütün sistemin enerji minimizasyonu gerçekleştirilmiş, ardından sistem 310 K'e ısıtılmıştır. Harmonik kısıtlamalar 1000 kJ mol^{-1} nm^{-2} 'ye ayarlanmış ve sistemin basıncı 1 atm olarak belirlenmiştir. Daha sonra kısıtlamalar önce 100 ve 10 kJ mol^{-1} nm^{-2} 'a düşürülmüş ve tamamen kaldırılmıştır. Sistemin termodinamik hesapları için iki şekilde enerji grubu tanımı yapılmıştır. İlkinde (Tip-1), protein-DNA-SAM

ile çözücü-iyon ayrı enerji grupları olarak belirtilirken, ikincisinde (Tip-2) ise tüm sistem tek bir enerji grubu olarak tanıtılmıştır. Yürütülen 500 ns'lik bir simülasyonun 10 ns aralıklarla oluşturduğu konformasyonların görüntüsü Şekil 3'te gösterildiği gibidir.



Şekil 3. 500 ns uzunluğundaki simülasyondan 10 ns aralıklarla çekilen konformasyonların görüntüsü

Farklı GROMACS versiyonlarının ve hesaplama kümelerinin kullanımının performansa etkisinin anlaşılabilmesi için çeşitli simülasyon parametre kombinasyonları tasarlanmıştır. Levrekv2-cuda kümesinde GROMACS 5.1.4; barbun-cuda kümesinde GROMACS 2020; akya-cuda kümesinde ise hem GROMACS 5.1.4 hem de GROMACS 2020 hali hazırda yüklü oldukları için bu kümelerde belirtilen GROMACS versiyonları denenmiştir (Tablo 1). GROMACS versiyonlarının farklılıklarının incelenmesinin ardından, *thread* (iş parçacığı) sayısının değiştirilmesinin performansa etkisi incelenmiştir. Ayrıca, en verimli performansın elde edildiği akya-cuda kümesinde 40/1 CPU/GPU oranının, yanında 40/2 CPU/GPU oranı da denenmiştir.

III. BULGULAR VE TARTIŞMA

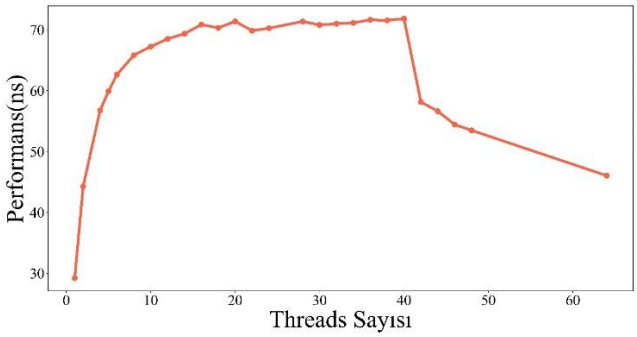
Bu bölümde kullanılan GROMACS versiyonlarının ve hesaplama kümelerinin kullanımının, günlük performansa etkisi incelenmiştir. Küme karşılaştırması için 20 *thread* kullanılmıştır. Tablo 1'de görüldüğü üzere en düşük performans levrekv2-cuda kümesinde elde edilmiştir. Bunun sebebi ise kullanılan CPU/GPU oranının en az olmasından kaynaklanmaktadır. Ayrıca, bu kümede en eski GROMACS sürümü bulunmakla beraber, CPU/GPU kartları da diğer kümelere göre en eski jenerasyondan gelmektedir. GROMACS 5.1.4 versiyonunda farklı enerji grubu tanımları, GPU'da yürütülen enerji hesabını etkilemediği için 5.1.4 kullanılan tüm seçeneklerde, enerji tipi seçiminin performansa bir etkisi görülmemiştir (Tablo 1, No 1-2 ve 5-6). Ek olarak, GROMACS 2020 kullanımı sırasında Tip-1 enerji grubunun tanımlanmasının performansta ciddi bir düşüşe sebep olduğu gözlemlenmiştir (Tablo 1, No 3-4 ve 7-8). Bunun sebebinin, GROMACS 2020 kullanımı sırasında Tip-1 enerji grubunun tanımlanmasının, GPU üzerinde iş yürütülmesini engellemesinden kaynaklandığı tespit edilmiştir.

Tablo 1. Simülasyonların Günlük Performansı

No	GROMACS versiyonu	Hesaplama Kümesi	Enerji Grubu	CPU/GPU	Threads Sayısı	Günlük Performans
1	5.1.4	Levrekv2-cuda	Tip-1	24/1	20	6,374 ns
2	5.1.4	Levrekv2-cuda	Tip-2	24/1	20	6,939 ns
3	2020	Barbun-cuda	Tip-1	40/1	20	14,285 ns
4	2020	Barbun-cuda	Tip-2	40/1	20	40,038 ns
5	5.1.4	Akya-cuda	Tip-1	40/1	20	37,557 ns
6	5.1.4	Akya-cuda	Tip-2	40/1	20	38,663 ns
7	2020	Akya-cuda	Tip-1	40/1	20	14,924 ns
8	**2020**	**Akya-cuda**	**Tip-2**	**40/1**	**20**	**71,372 ns**
9	2020	Akya-cuda	Tip-2	40/2	20	71,030 ns
10	2020	Akya-cuda	Tip-2	40/1	1	29,205 ns
11	2020	Akya-cuda	Tip-2	40/1	10	67,225 ns
12	2020	Akya-cuda	Tip-2	40/1	30	70,780 ns
13	2020	Akya-cuda	Tip-2	40/1	40	71,814 ns
14	2020	Akya-cuda	Tip-2	40/1	64	46,030 ns

Akya-cuda'da GROMACS 2020 kullanırken iki GPU kartı seçimi yapıldığında da (Tablo 1, No 9) performansta bir artış gözlenmemiştir. Tablo 1'de No 10-14 arasında, akya-cuda'da 40/1 CPU/GPU oranı için farklı *thread* sayısına karşılık gelen performans değerleri sunulmuştur. Bu değerlerin bize gösterdiği, bu koşullarda en verimli *thread* sayısının yine başlangıçta seçtiğimiz 20 sayısı olduğudur. Farklı thread sayılarına göre performans değişimi Şekil 4'te verilmiştir.

Oluşturulan bütün simülasyonlara ait sonuç dosyalarını ve yeniden yürütmek için gerekli dosyaları Github'ta yayınlayarak sonuçlarımızı akademik camiaya açtık (https://github.com/CSB-KaracaLab/gmx_performance_on_HPC). Ayrıca kullandığımız komut zincirlerini de paylaşarak, araştırmacıları kendi performans analizlerini kendi sistemlerinde ya da TRUBA'da yapabilmesinin önünü açtık.



Şekil 4. Farklı threads sayılarının performansa olan etkisi

IV. SONUÇ VE ÖNERİLER

Yapılan bu çalışma sonucunda, GROMACS 2020 kullanımının performansı yaklaşık iki kat arttırdığı görülmüştür. 1 GPU kullanımı ise performansı barbun-cuda kümesi için 2,8 kat, akya-cuda kümesi için 4,8 kat arttırmıştır. GPU kartı olarak NVIDIA V100 kullanıldığında performansın NVIDIA P100 kullanımına kıyasla 1,8 oranında arttığı gözlemlenmiştir. Elde ettiğimiz sonuçlar doğrultusunda, sonraki çalışmalarımızı TRUBA'da akya-cuda kümesinde yürütme kararı aldık. Bu sonuçların, GPU üzerinde hesaplama yapan ya da benzer sistem kullanan araştırmacıların yüksek performans elde etmesine olanak sağlayacağını umuyoruz. Araştırmamızın, kendi iş istasyonunu veya küçük ölçekli yüksek performanslı hesaplama sistemini kurmak isteyen araştırmacılar için de faydalı olacağını düşünmekteyiz. Son olarak da bu çalışmanın, ilgili araştırmacıları, sunduğumuz parametreleri kullanarak, başka sistemler üzerinde optimizasyon çalışmaları yapmaya teşvik edeceğini umuyoruz.

TEŞEKKÜRLER

Bu araştırma TÜBİTAK tarafından 1002 destek programı kapsamında 119Z828 numaralı proje ile desteklenmiştir. Yaptığı çalışmaların sonucu ile bu projenin ortaya çıkmasına yardımcı olan Deniz Doğan'a teşekkür ederiz. Ayrıca bu çalışmadaki hesaplamaların TRUBA kaynaklarında yapılmasına olanak sağlayan TÜBİTAK ULAKBİM'e teşekkür ederiz.

KAYNAKLAR

[1] Alder, B. J., & Wainwright, T. E. (1959). Studies in molecular dynamics. I. General method. *The Journal of Chemical Physics*, *31*(2). https://doi.org/10.1063/1.1730376

[2] Rahman, A. (1964). Correlations in the motion of atoms in liquid argon. *Physical Review*, *136*(2A). https://doi.org/10.1103/PhysRev.136.A405

[3] Páll, S., Abraham, M. J., Kutzner, C., Hess, B., & Lindahl, E. (2015). Tackling exascale software challenges in molecular dynamics simulations with GROMACS. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8759*. https://doi.org/10.1007/978-3-319-15976-8_1

[4] Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., & Lindah, E. (2015). Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, *1–2*. https://doi.org/10.1016/j.softx.2015.06.001

[5] Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L., & Schulten, K. (2005). Scalable molecular dynamics with NAMD. In *Journal of Computational Chemistry* (Vol. 26, Issue 16). https://doi.org/10.1002/jcc.20289

[6] Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham, T. E., DeBolt, S., Ferguson, D., Seibel, G., & Kollman, P. (1995). AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Computer Physics Communications*, *91*(1–3). https://doi.org/10.1016/0010-4655(95)00041-D

[7] Bird, A. (2007). Perceptions of epigenetics. In *Nature* (Vol. 447, Issue 7143). https://doi.org/10.1038/nature05913

[8] Mazzio, E. A., & Soliman, K. F. A. (2012). Basic concepts of epigenetics impact of environmental signals on gene expression. In *Epigenetics* (Vol. 7, Issue 2). https://doi.org/10.4161/epi.7.2.18764

[9] Khavari, D. A., Sen, G. L., & Rinn, J. L. (2010). DNA methylation and epigenetic control of cellular differentiation. In *Cell Cycle* (Vol. 9, Issue 19). https://doi.org/10.4161/cc.9.19.13385

[10] Lee, J. H., Hart, S. R. L., & Skalnik, D. G. (2004). Histone Deacetylase Activity Is Required for Embryonic Stem Cell Differentiation. *Genesis*, *38*(1). https://doi.org/10.1002/gene.10250

[11] Weinhold, B. (2006). Epigenetics: the science of change. *Environmental Health Perspectives*, *114*(3). https://doi.org/10.1289/ehp.114-a160

[12] Kulis, M., & Esteller, M. (2010). DNA Methylation and Cancer. In *Advances in Genetics* (Vol. 70, Issue C). https://doi.org/10.1016/B978-0-12-380866-0.60002-2

[13] Law, J. A., & Jacobsen, S. E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. In *Nature Reviews Genetics* (Vol. 11, Issue 3). https://doi.org/10.1038/nrg2719

[14] Chédin, F. (2011). The DNMT3 family of mammalian de novo DNA methyltransferases. In *Progress in Molecular Biology and Translational Science* (Vol. 101). https://doi.org/10.1016/B978-0-12-387685-0.00007-X

[15] Zhang, Z. M., Lu, R., Wang, P., Yu, Y., Chen, D., Gao, L., Liu, S., Ji, D., Rothbart, S. B., Wang, Y., Wang, G. G., & Song, J. (2018). Structural basis for DNMT3A-mediated de novo DNA methylation. *Nature*, *554*(7692). https://doi.org/10.1038/nature25477

[16] Norvil, A. B., Petell, C. J., Alabdi, L., Wu, L., Rossie, S., & Gowher, H. (2018). Dnmt3b Methylates DNA by a Noncooperative Mechanism, and Its Activity Is Unaffected by Manipulations at the Predicted Dimer Interface. *Biochemistry*, *57*(29). https://doi.org/10.1021/acs.biochem.6b00964

[17] TRUBA. https://www.truba.gov.tr/index.php/en/main-page/

[18] TRUBA Wiki Sayfası. http://wiki.truba.gov.tr/index.php/Ana_sayfa

[19] Ivani, I., Dans, P. D., Noy, A., Pérez, A., Faustino, I., Hospital, A., Walther, J., Andrio, P., Goñi, R., Balaceanu, A., Portella, G., Battistini, F., Gelpí, J. L., González, C., Vendruscolo, M., Laughton, C. A., Harris, S. A., Case, D. A., & Orozco, M. (2015). Parmbsc1: A refined force field for DNA simulations. *Nature Methods*, *13*(1). https://doi.org/10.1038/nmeth.3658