

EuroHPC Çağrısı Proje Deneyimi Development Access

Tandaç Fürkan Güçlü

Büşra Tayhan

Ebru Çetin

Ali Rana Atılgan

Canan Atılgan

. Sabancı .
Universitesi

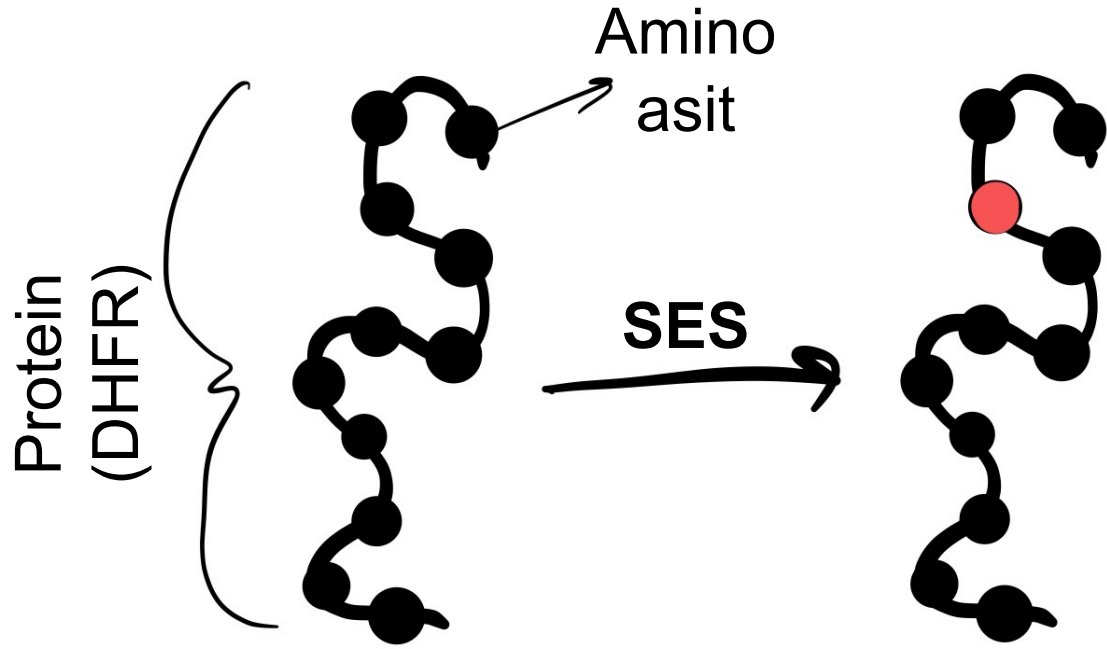
Projenin adı ve kapsamı

- Başvuru tarihleri ve kapsamı; e-posta listeleri, sosyal medya ve Başarım 2022 aracılığıyla öğrenildi.

Projenin adı ve kapsamı

- Başvuru tarihleri ve kapsamı; e-posta listeleri, sosyal medya ve Başarım 2022 aracılığıyla öğrenildi.
- Projenin adı; 'Mapping the molecular scale fitness of dihydrofolate reductase in the presence of the drug trimethoprim via alchemy'
- DHFR adlı proteinin yapıtaşı olan amino asitlerin olası tüm diğer amino asitlere (kendisi hariç toplamda 19 tane amino asit) tekli olarak serbest enerji sarsımı, SES (free energy perturbation, FEP) metodu kullanılarak değiştirilmesi ve aradaki serbest enerji farkının iki ayrı ilaç molekülüne bağlı ve bağlanmamış halleri için hesaplanması ve incelenmesi.

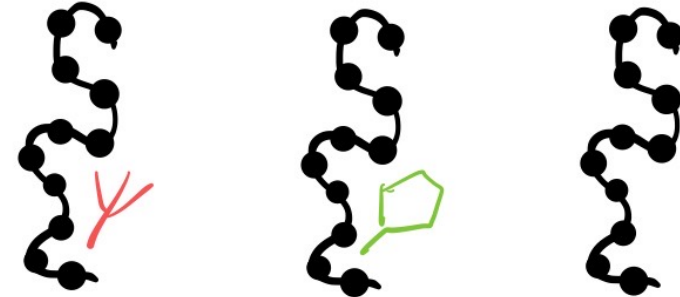
DHFR ve SES benzetimleri



- DHFR 159 amino asitten (aa) oluşmaktadır

- Doğada 20 aa bulunur.
- Yani her pozisyon (159), kendisi hariç 19 farklı aa'ya dönüştürülecek.
- İki aa yapısı karmaşıklığı sebebiyle dönüştürülemiyor (prolin ve glisin).
- Yani her DHFR yapısı için $159 \times 17 = 2.703$ tane SES benzetimi, istatistiksel önem için 3 kopya ile **8.109** adet SES benzetimi
- 3 farklı kompleks (ilaçlılar ve bağlanmamış) toplamda **24.327** tane SES benzetimi

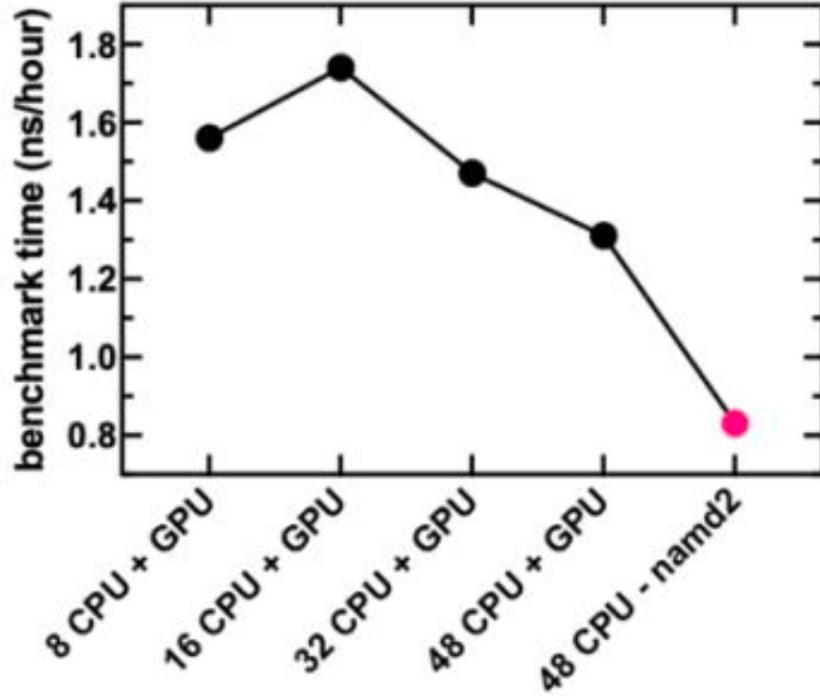
1. ilaç bağı 2. ilaç bağı bağlanmamış



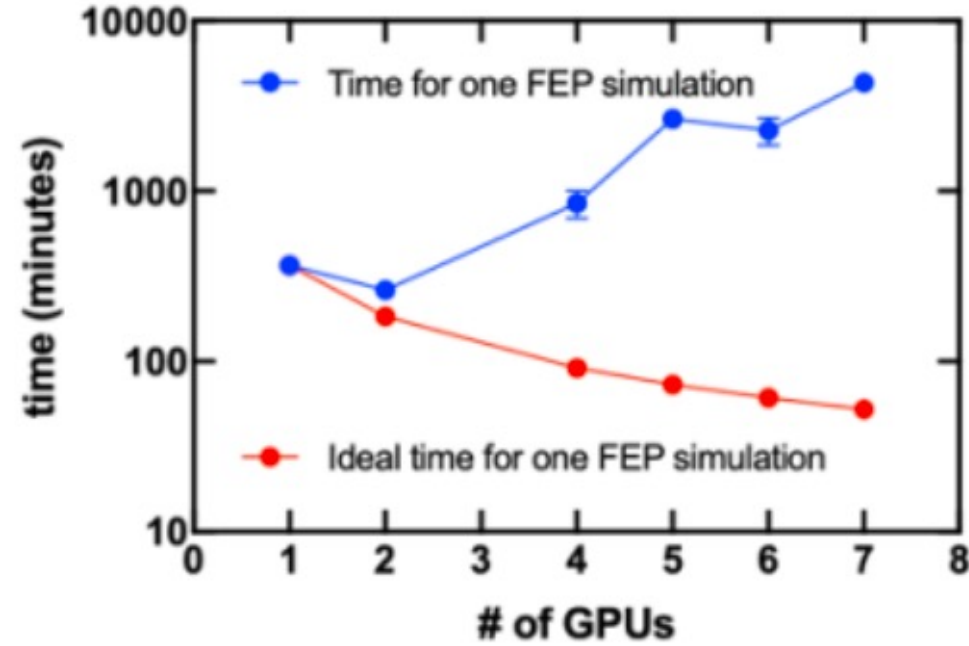
Karolina HPC başvurusu – teknik kısım

- Kullanılacak yazılım namd3 –
(<https://www.ks.uiuc.edu/Research/namd/alpha/3.0alpha/>)
- Her benzetim için 1 tane GPU ve 8-16 sayıda CPU çekirdeği yeterli
- Carolina GPU düğümlerinde: 8 x NVIDIA A100 GPU ve 128 çekirdekli AMD CPU var.
- Zaman istekleri GPU bazında değil, düğüm bazında yapılıyor. Fakat çekirdek saat üzerinden de olmak üzere tüm benchmark analizlerinin yapılması isteniyor.

Karolina HPC başvurusu – teknik kısım



Neden çok kullanılmış namd2 değil de yani çıkan namd3?



Neden her benzetim için 1 GPU da çok sayıda GPU değil? (Bu benchmark TRUBA'nın Palamut düğümlerinde yapıldı.)

Sonuç olarak **29.000 düğüm saati** başvurusunda bulunuldu.

Projenin başlangıcı – hesap açılması ve sisteme bağlanma

- Regular Access başvurumuzun %10 hesap saati bize 'Development Access' olarak verildi, yani **3.000** düğüm saati.
- Öncelikle e-posta üzerinden bildirildi daha sonra önce Karolina'nın çevrimiçi sisteminden hesap açıldı.

Project DD-22-92 Active #2128

[Project List](#)

Name: EU2022R02-220: Mapping the molecular scale fitness of dihydrofolate reductase in the presence of the drug trimethoprim via alchemystry
Lifetime: 2022-12-08 00:00:00 +0100 to 2023-12-08 23:59:00 +0100 open
Call type: Directors' Discretion (0th)
Project area: Bio Sciences (Life Sciences, Biophysics / Bioinformatics, Medicine)
Group: [dd-22-92](#)
Pi: Prof. Canan Atilgan
Tag list: [eurohpc](#)
Note: Rejected EuroHPC Reg. access call
States: S

Allocations

Status	QoS i	Resource	Resource Type	Validity Period	Requested	Allocation	Usage i
<input checked="" type="checkbox"/> Active	2128_2983	Karolina GPU	Node Hours	2022-12-08 to 2023-12-08	29,000	3,000	555.99 18.5% 2,444.01

Attachments

Title	Description	Author	Project	Creation Time	Filename
Mapping the molecular scale fitness of dihydrofolate reductase in the presence of the drug trimethoprim via alchemistrv	N/A		DD-22-92	2022-12-08 15:39:13	

Projenin başlangıcı – hesap açılması ve sisteme bağlanma

- Regular Access başvurumuzun %10 hesap saati bize 'Development Access' olarak verildi, yani **3.000** düğüm saati.
- Öncelikle e-posta üzerinden bildirildi daha sonra önce Karolina'nın çevrimiçi sisteminden hesap açıldı.
- Burada aynı zamanda disk alanı bilgisi ve mevcut.

Project

Users

Resources

Reservations

Queues

Compute Stats

Storage Stats

Set up a recovery phone number [here](#).



Project DD-22-92 #2128

Back

Project Storage Quota

Project storages	Space Quota	Entries Quota	Expiration
proj3	20 TB	5 Million	

Project Storage Usage

Project storages	Space used	Entries	Last Update
proj3	351.95 GB	71.6 Thousand	2023-09-27 11:20

Cookies help us deliver our services. By using our services, you agree to our use of cookies.

OK

Projenin başlangıcı – hesap açılması ve sisteme bağlanma

- Regular Access başvurumuzun %10 hesap saati bize 'Development Access' olarak verildi, yani **3.000** düğüm saati.
- Öncelikle e-posta üzerinden bildirildi daha sonra önce Karolina'nın çevrimiçi sisteminden hesap açıldı.
- Burada aynı zamanda disk alanı bilgisi ve mevcut.
- Sisteme ssh key kullanarak giriliyor.



IT4INNOVATIONS
NATIONAL SUPERCOMPUTING
CENTER

Change your SSH Key

Service is not intended for e-INFRA CZ users! Use [e-INFRA CZ user profile](#) instead.

i Enter your password and new public SSH key. After action, please, wait a moment (~5min) for the public key to be propagated to all clusters.

Login



Login

Password



Password

Public SSH Key

Public SSH Key

>_



Captcha

Send

Projenin başlangıcı – hesap açılması ve sisteme bağlanma

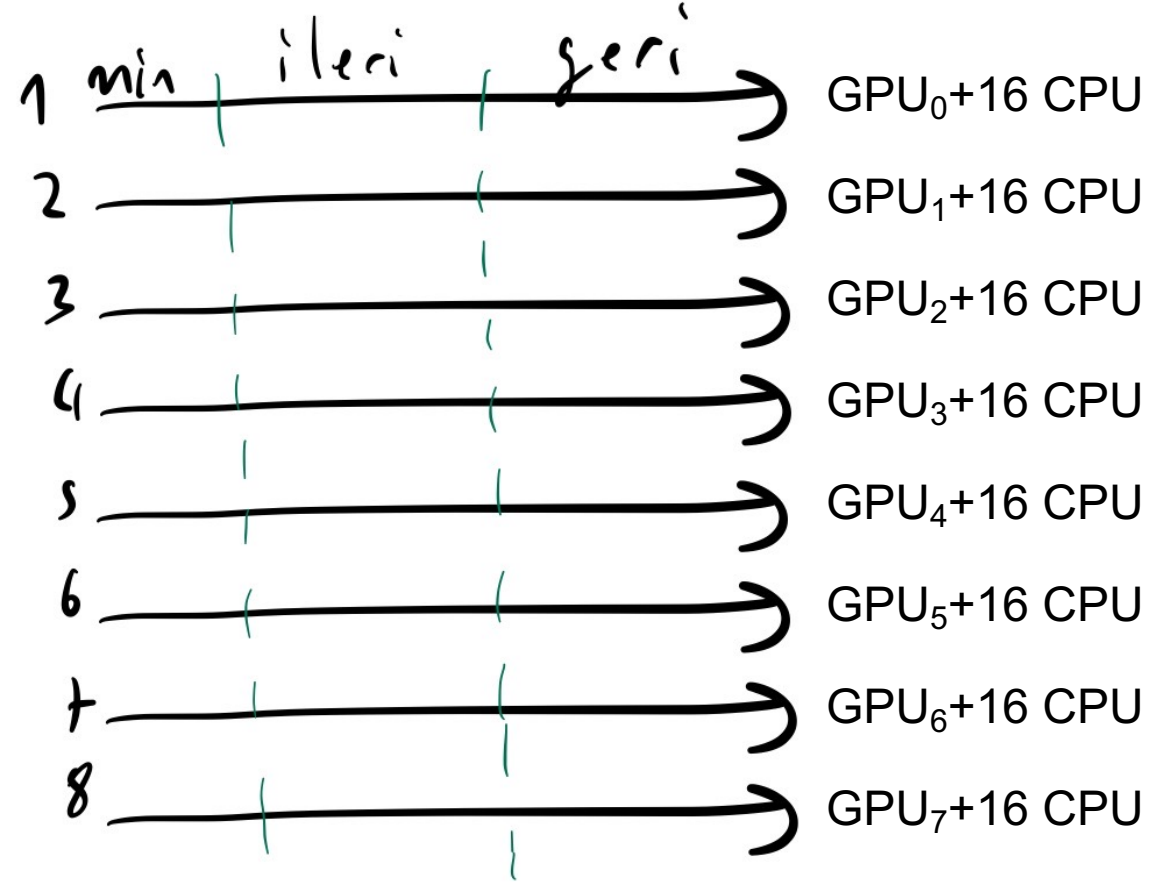
- Regular Access başvurumuzun %10 hesap saati bize 'Development Access' olarak verildi, yani **3.000** düğüm saati.
- Öncelikle e-posta üzerinden bildirildi daha sonra önce Karolina'nın çevrimiçi sisteminden hesap açıldı.
- Burada aynı zamanda disk alanı bilgisi ve mevcut.
- Sisteme ssh key kullanarak giriliyor.
- Sistem linux olduğu için, en az orta seviye linux bilgisi elzem.

Sisteme giriş ve benzetimlerin yürütülmesi

- Karolina 'PBS' sıra sistemini kullanıyordu. 20 Eylül 2023 itibariyle 'SLURM' kullanılmaya başlandı.
- Kullanıcıya 25 GB home klasörü veriyor. Projemize 20 TB yer verildi. Dosya sayısı sınırlamaları da mevcut.
- Bilgi kısıtlı, neredeyse örnek betik bile yok, kendi sayfalarında.
- 'IT ticket' kısmından sorununuzu bildirdiğinizde gün içinde cevap alıyorsunuz.
- İlk problem kaynakların düğüm olarak yani tek seferde 8 GPU ve 128 CPU çekirdeği olarak ayrılması...

Sisteme giriş ve benzetimlerin yürütülmesi

- SES benzetimleri minimizasyon, ileri süreç ve geri süreç olmak üzere 3 alt benzetimden oluşmakta ve her benzetim birbirinin çıktısı ile yürümektedir.
- Yani bir düğüme 3 tane seri ve bu üçlü grubun da 8'li şekilde paralel yürütmesini vermek gerekiyor.



Benzetim yürütme betiği

- Benzetimleri yürütmek için toparlayıcı bir betik yazıldı.
- Bunun sonucunda çıkan PBS yükleme dosyası örnek olarak:

```
/home/it4i-cananat/NAMD_3.0alpha8_Linux-x86_64-multicore-CUDA/namd3
+p16 +devices 0 wt_d4_70_ASP_122_ALA_fep_wb_ionized_config_min.conf >
wt_d4_70_ASP_122_ALA_fep_wb_ionized_config_min.log && /home/it4i-
cananat/NAMD_3.0alpha8_Linux-x86_64-multicore-CUDA/namd3 +p16 --
CUDASOAintegrate on +devices 0
wt_d4_70_ASP_122_ALA_fep_wb_ionized_config_min_forw.conf >
wt_d4_70_ASP_122_ALA_fep_wb_ionized_config_min_forw.log && /home/it4i-
cananat/NAMD_3.0alpha8_Linux-x86_64-multicore-CUDA/namd3 +p16 --
CUDASOAintegrate on +devices 0
wt_d4_70_ASP_122_ALA_fep_wb_ionized_config_min_forw_back.conf >
wt_d4_70_ASP_122_ALA_fep_wb_ionized_config_min_forw_back.log &
```

7 tane daha aynısı ve yürütmelerin hepsinin bitmesinin beklenmesi için en son

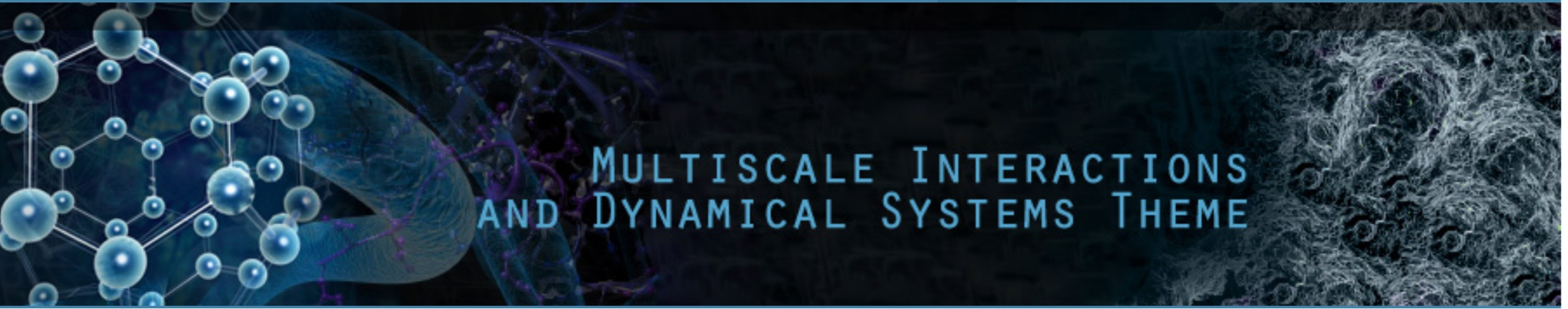
wait

Ara sonuçlar ve ilerleme

- 8'li gruplar şeklinde vereceğimiz için, 17 tane amino asit için SES benzetimi yapmaktansa, deneysel verileri tarayarak 1 dönüştürmeyi de eledik ve 16 mutasyonda karar kıldık.
- Şu ana kadar 9 pozisyonun tüm kopyaları ve bağlı/bağlanmamış SES benzetimleri ve ilk baştaki deneme/yanılgılar ile 3.000 düğüm saatinin ~%20'sini kullandık.
- Betikler ve kodlar projenin sonuna doğru github adresiminiz (<https://github.com/midstlab>) üzerinden ulaşılabilir durumda olacak.

Öğrenilen dersler

- İnsan hatasını en düşük hale getirmek için olabildiğince süreçleri kodlarla otomatize etmek gerekiyor.
- Sık sık nokta atışı kontroller ile benzetimlerin doğru ilerlediğine bakmak gerekiyor.
- Hata bildirimini zayıf olduğundan, bazen disk alanı bittiğinde de işler ilerleyebiliyor ve bu yüzden hesaplama zamanı kaybı olabiliyor.
- Bash terminali bilgisi ve linux bilgisi mutlaka gerekiyor.
- Bilimsel kısım ile teknik kısmın uyumlu olması da çok önemli. Gerektiğinde iki taraf da birbirine uygun olarak modifiye edilmeli.



Midstlab - Sabanci Üniversitesi

<https://midst.sabanciuniv.edu/>

Dinlediğiniz için teşekkür ederim.