



NCHC × NTU - NVIDIA BioNeMo Protein Design Workshop 2024

Ying-Ja Chen

Solutions Architect, Healthcare & Life Sciences, NVIDIA



Agenda

-
- 9:30 國網中心生醫核心設施介紹
-

10:00 連線到TWCC VCS運算環境 (hands-on)

10:10 NVIDIA BioNeMo介紹

10:30 蛋白質設計題目簡介

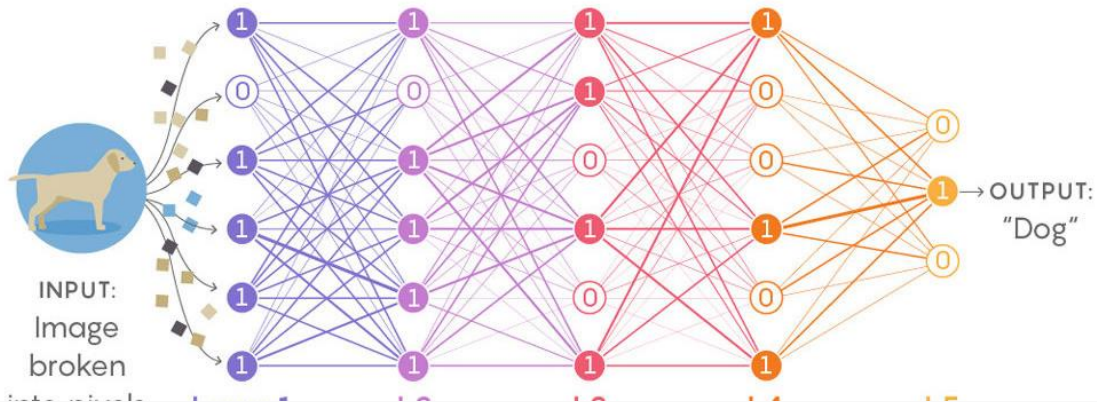
10:40 蛋白質設計實作 (demo / hands-on)

11:30 在TWCC佈署NVIDIA NIM

Nobel Prize in 2024

Artificial intelligence awarded in two categories

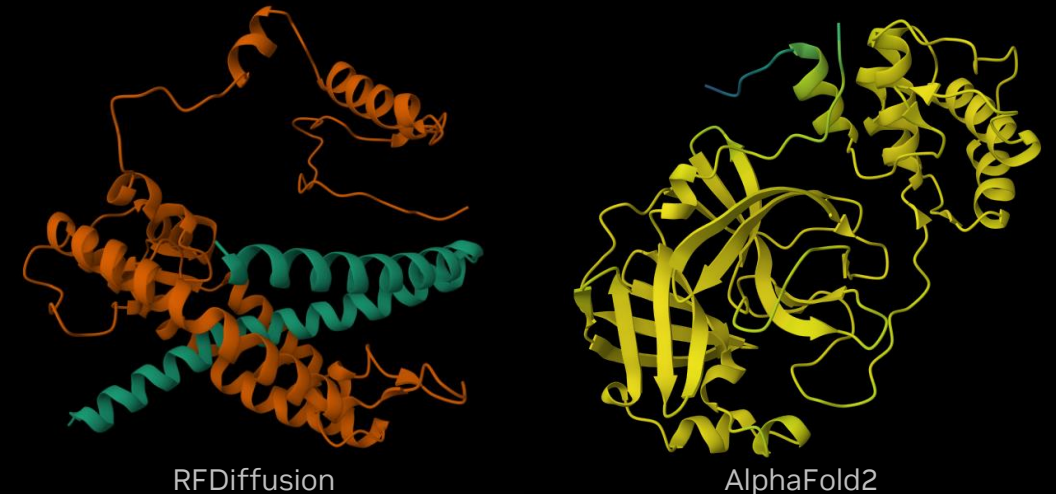
Nobel Prize in Physics



Geoffrey Hinton & John Hopfield

foundational discoveries and inventions that enable machine learning with artificial neural networks

Nobel Prize in Chemistry

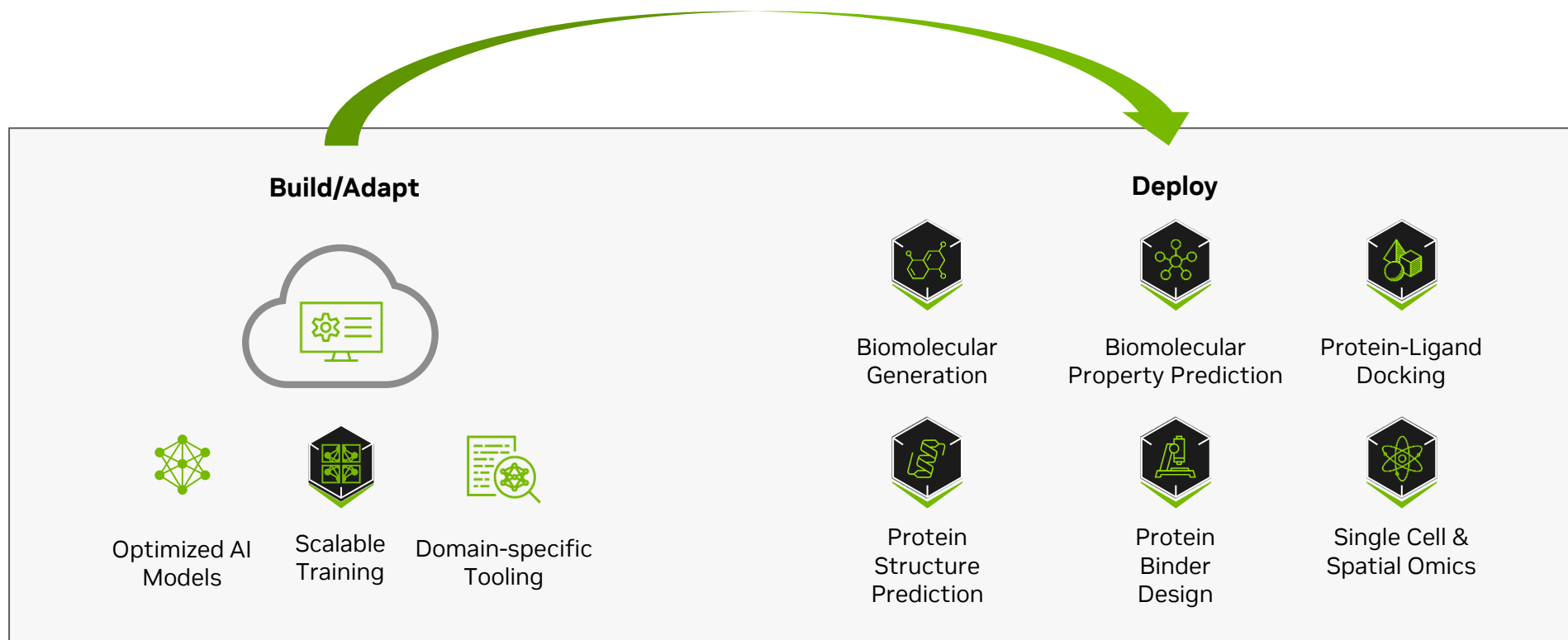


David Baker, Demis Hassabis and John Jumper

Computational protein design and protein structure prediction

What is NVIDIA BioNeMo?

Build, Adapt and Deploy AI Models for Computer-Aided Drug Discovery



BioNeMo Framework

Open-Source (soon!) | NVIDIA AI Enterprise

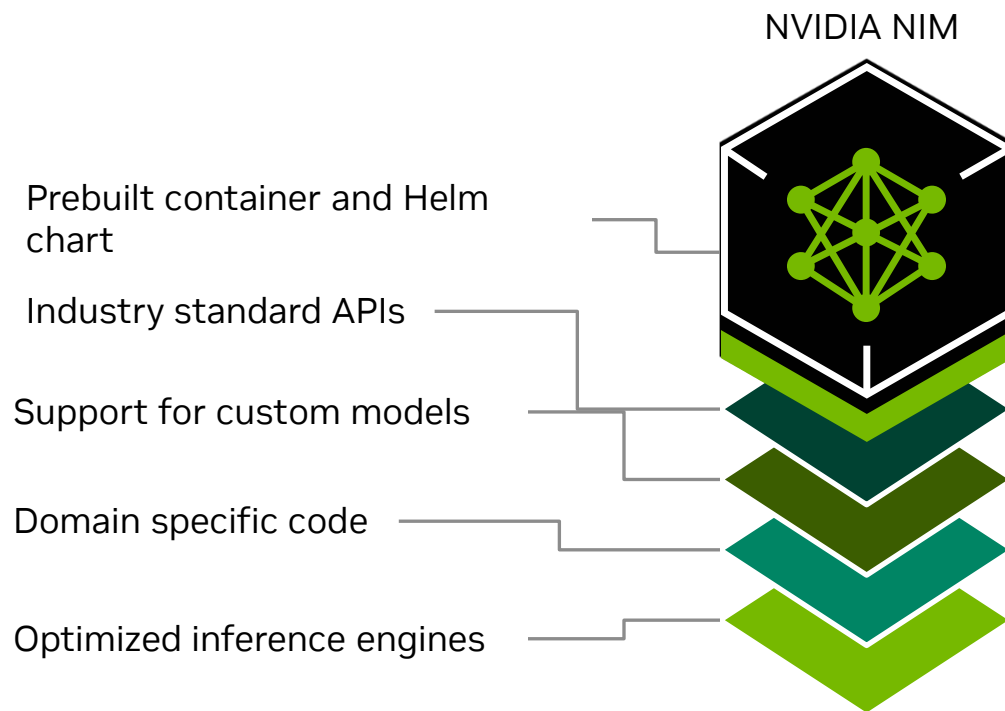
BioNeMo NIMs

Developer Program | NVIDIA AI Enterprise



NVIDIA NIM: Inference Microservices for Generative AI

Accelerated runtime for generative AI



Deploy anywhere and maintain control of generative AI applications and data

Simplified development of AI application that can run in enterprise environments

Day 0 support for all generative AI models providing choice across the ecosystem

Improved TCO with best latency and throughput running on accelerated infrastructure

Best accuracy for enterprise by enabling tuning with proprietary data sources

Enterprise software with feature branches, validation and support



DGX & DGX Cloud



NVIDIA Inference Microservices (NIM) & NIM Agent Blueprints (NAB)

Pre-trained Generative AI Models on build.nvidia.com



Search NVIDIA AI

Explore NIM Docs Forums Login

Discover

MODELS

Reasoning

Vision

Visual Design

Retrieval

Speech

Biology

Simulation

INDUSTRIES

Gaming

Healthcare

Industrial

Drug Discovery

State-of-the-art protein structure prediction, molecular generation, and molecular docking models with generative AI


RUN ANYWHERE



ipd
proteinmpnn

amino acid prediction bionemo

PREVIEW



meta
esm2-650m

biology bionemo


RUN ANYWHERE



deepmind
alphafold2

biology bionemo

PREVIEW



ipd
rfdiffusion

bionemo biology


RUN ANYWHERE



nvidia
molmim

drug discovery molecule generation

RUN ANYWHERE

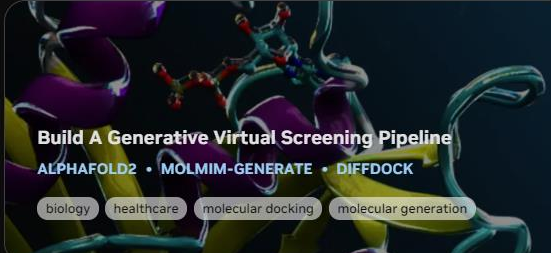


mit
diffdock

docking drug discovery

Explore Agent Blueprints

Reference code with guides for customization and deployment of applications built with NVIDIA NIM and partner microservices



Build A Generative Virtual Screening Pipeline

ALPHAFOLD2 • MOLMIM-GENERATE • DIFFDOCK

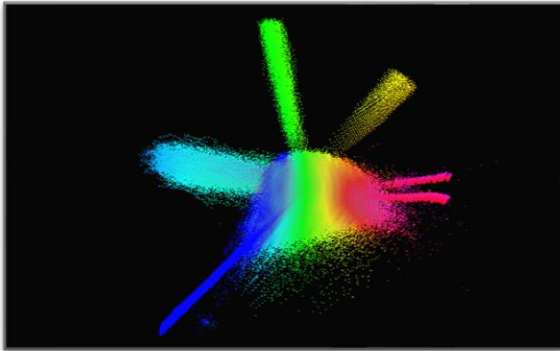
biology healthcare molecular docking molecular generation

Medical Imaging

Segment medical images with foundation models

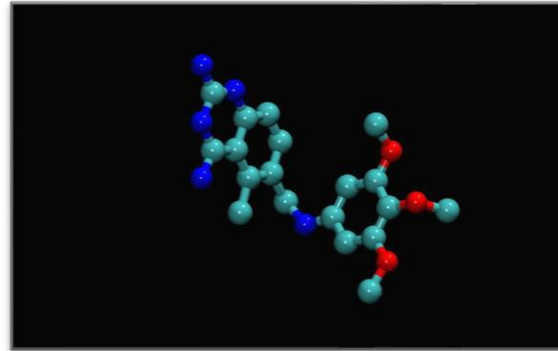
BioNeMo Framework v1.10.1

Models Supported across Proteins, Small Molecules, and Genomics



ESM 1 & 2

Protein LLMs for amino acid sequences, to generate representations for a wide variety of protein property and function predictions



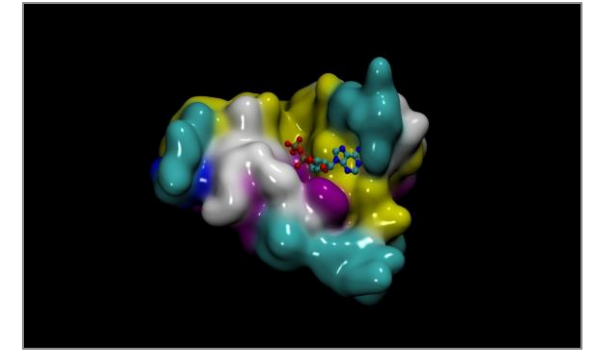
MolMIM & MegaMolBART

Generative chemistry model for downstream tasks including reaction prediction, molecular optimization & de novo molecular generation.



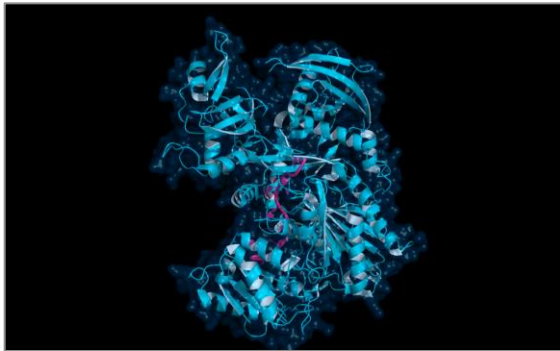
ProtT5

Protein T5 model developed in collaboration with the Rost lab to extend the capabilities of protein LLMs to sequence generation.



DiffDock & EquiDock

Docking prediction models: DiffDock is a diffusion based model for small molecule docking prediction; EquiDock is an equivariant rigid protein-protein binding prediction model.



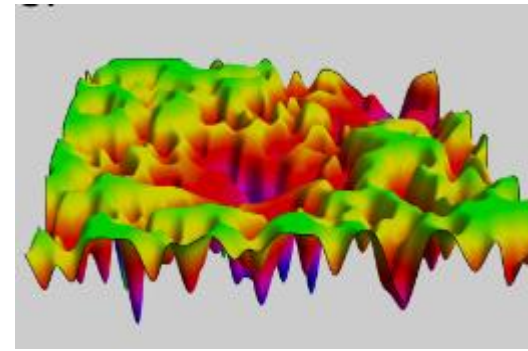
OpenFold

OpenFold is the open source version of DeepMind's AlphaFold, for 3D protein structure prediction from amino acid sequences



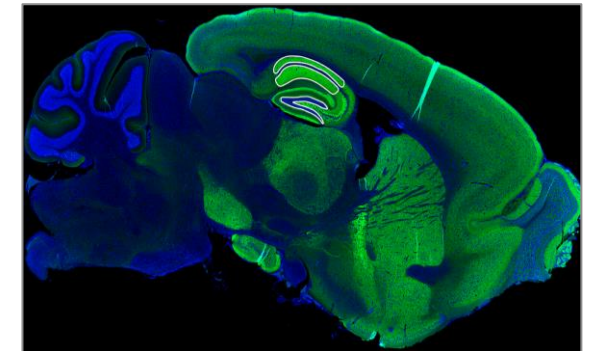
DNABERT

DNA sequence model for building learned representation, to provide embeddings for downstream tasks such as splice-site or binding site prediction



NEW: DSMBind

Produce comparative values for ranking protein-ligand binding affinities



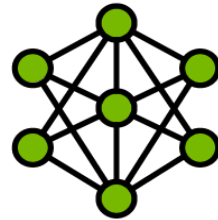
Geneformer

Foundation model for single cell expression data, to provide embeddings for downstream tasks such as cell type annotation or cell perturbation predictions

Foundation Model in Biology

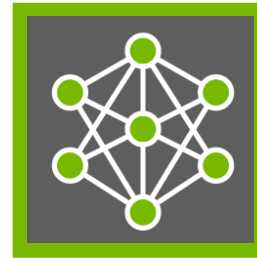
Pre-trained foundation models can be customized for various predictive tasks

Pre-trained Model

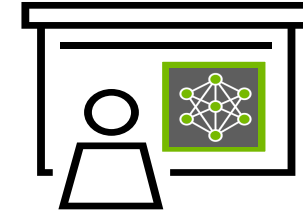


Customization
→
Ex: fine-tuning,
prompt engineering

Customized Model



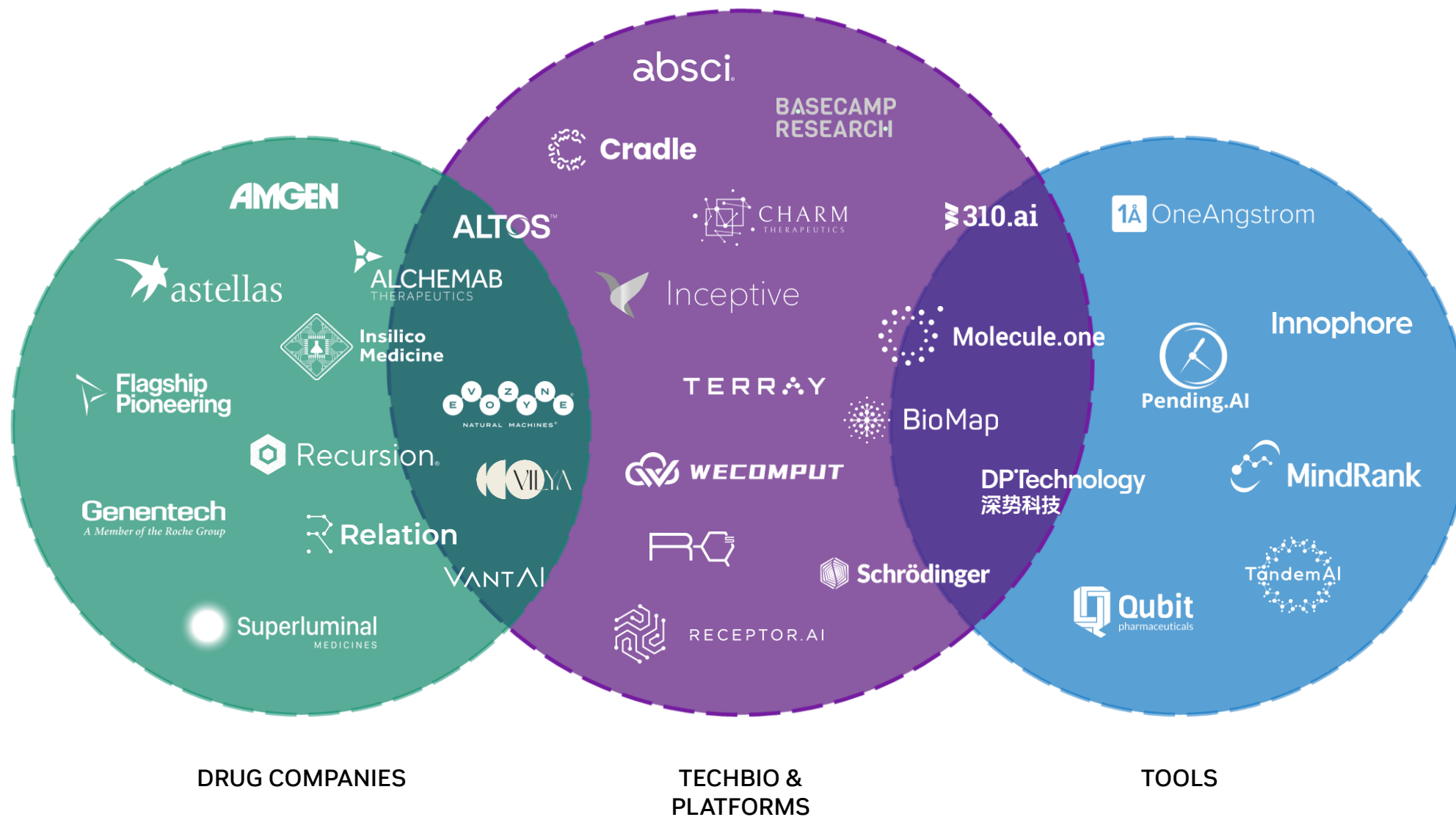
Tasks



Natural Language	ChatGPT Trained on all text on the Internet	Customer Service Agent Customized to give concise and correct answers	Answer customer's questions
Protein	ESM-2 Trained on protein sequences in UniProt and structures in PDB	PTM Model Fine-tuned with known post-translational modification sequences	Predicts post-translational modification sites

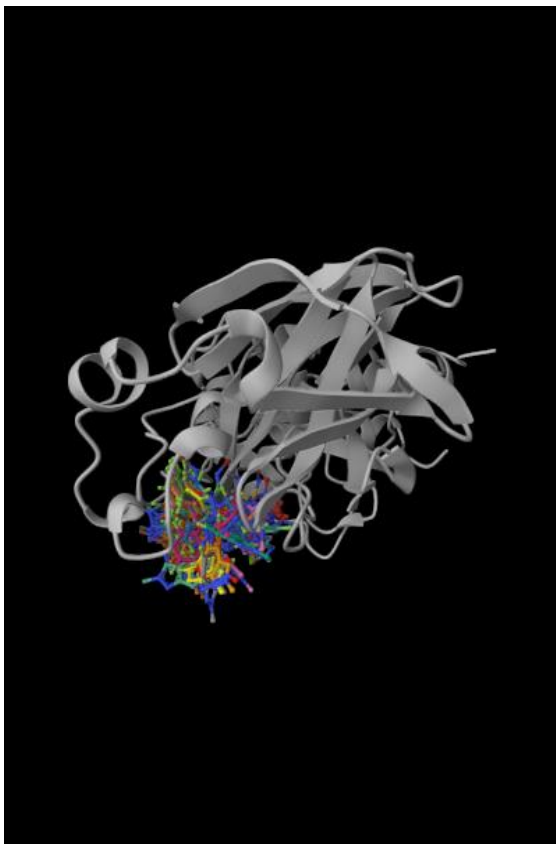
NVIDIA BioNeMo Ecosystem

400+ Users | 100+ Organizations Integrating & Using



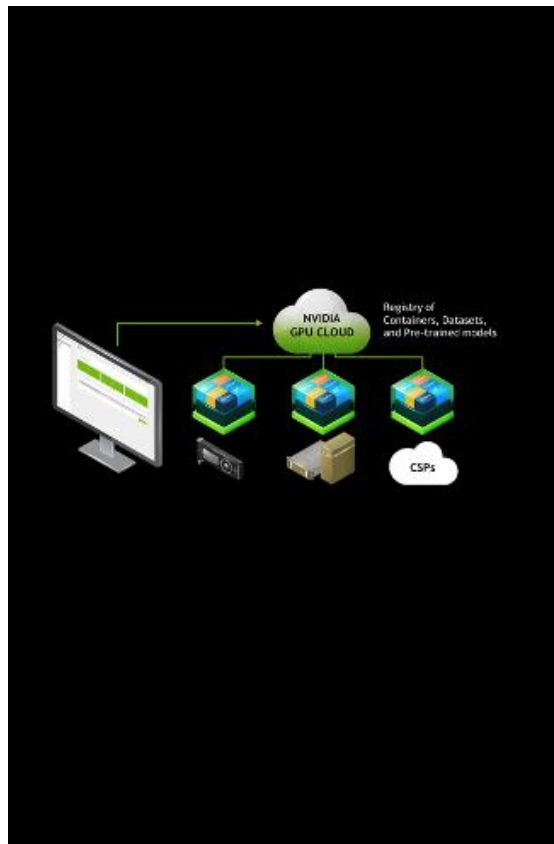
Experience and Run Enterprise Generative AI Models Anywhere

Seamlessly integrate AI in business applications with NVIDIA AI APIs



build.nvidia.com API/UI

Start testing in minutes with the API playground.



NVIDIA AI Enterprise Licensing

Purchase a license to download and self-host NIMs or download and deploy the BioNeMo Framework container



NVIDIA Developer Program

Access NVIDIA Enterprise for free testing / R&D to download and self-host NIMs or download and deploy the BioNeMo Framework container

BioNeMo Resources

- BioNeMo [產品網頁](#)和[資源](#)：BioNeMo每月底都會更新，可以定期察看有何新模型或benchmark資料
- Github
 - [Source code](#): BioNeMo v2.0 is open source
 - [Examples](#) 可以從這裏的範例開始學習使用或訓練模型
- Documentation:
 - [BioNeMo NIM](#)
 - [BioNeMo v2.0](#)
 - [BioNeMo v1.10](#): source code未公開但可免費使用，模型內容較多 (protein, molecule generation, protein design, [DSMBind](#) binding affinity, DNA, scRNA-seq)，如果做fine-tuning沒有要改底層的程式碼，建議由此開始
- **Issues**: 有問題可以在[Forum](#)找答案或回報
- 介紹：**GTC**演講介紹的[錄影或教學影片](#)以及案例分享的[Blog文章](#)



Protein Design Workshop

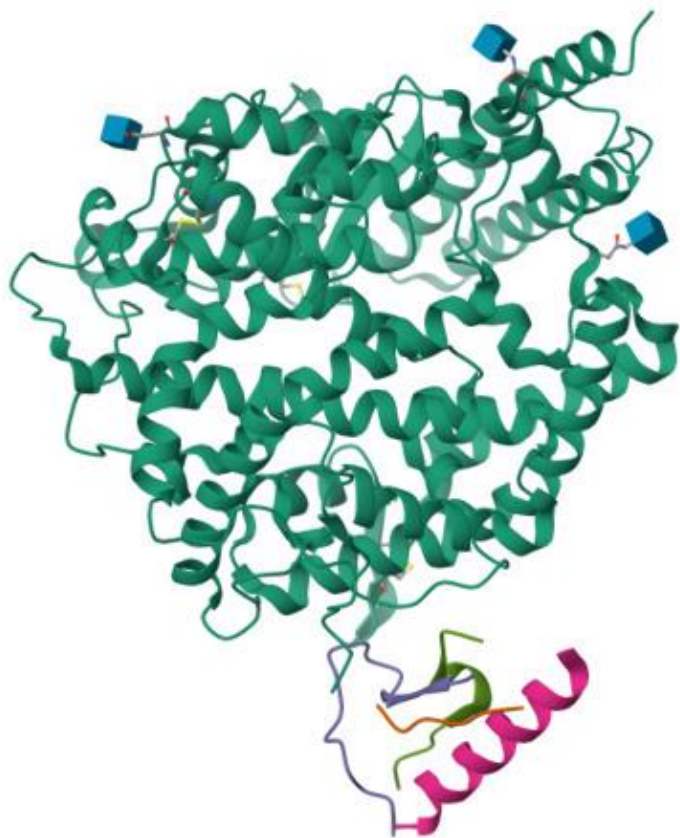
Computing Environment

- TWCC Virtual Computing Service (VCS)
- VCS Image
 - Docker containers and environment variables are pre-installed and configured
- SSH
 - Public/private key setup
 - SSH tunneling for connection to jupyter notebook
- Download
 - SSH private key for VCS image
 - Jupyter notebook for protein design workshop

Docker Containers

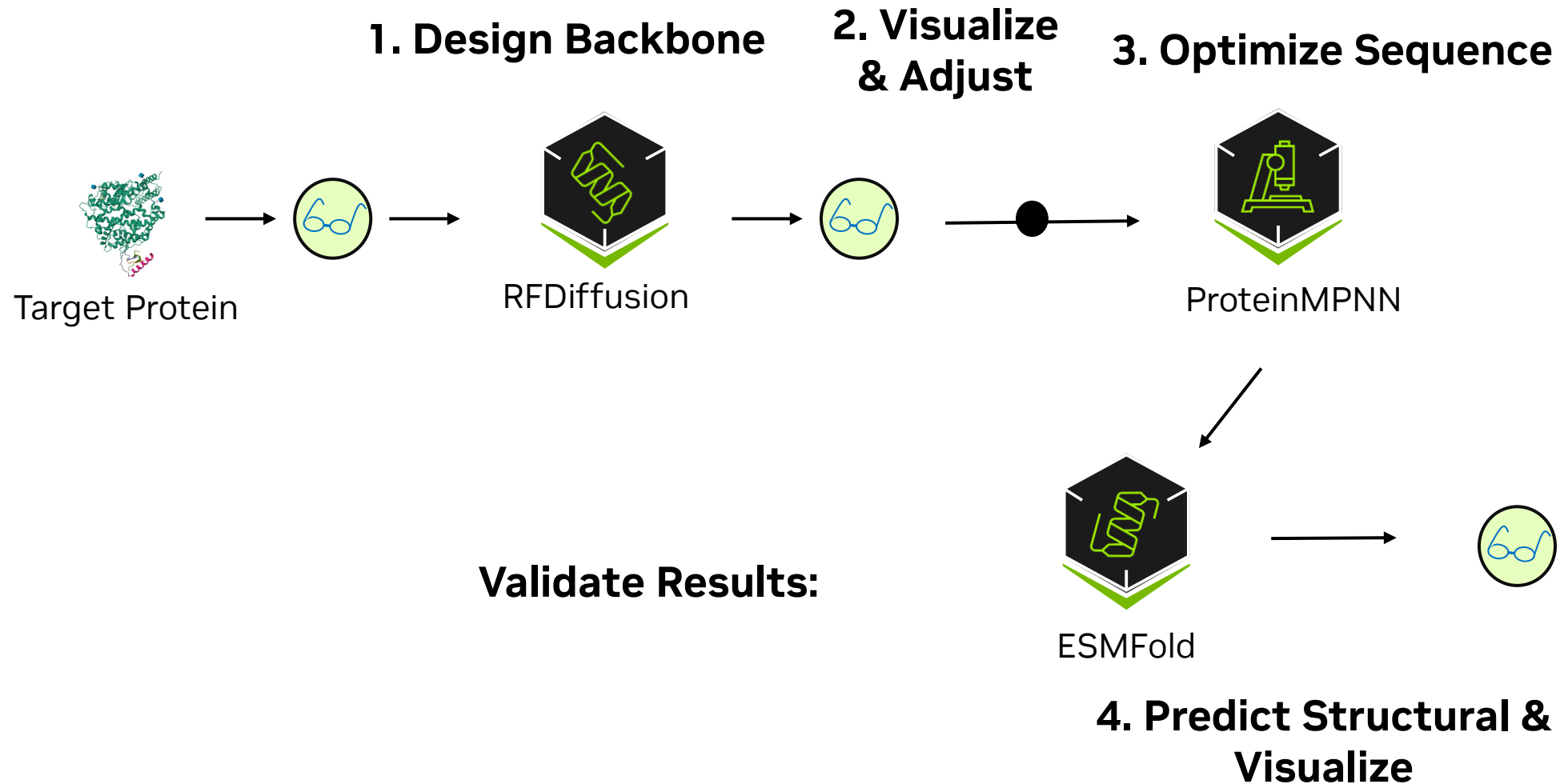
- RFDiffusion NIM
- ProteinMPNN NIM
- BioNeMo Framework
 - Use for Jupyter Notebook

Protein Design: ACE2 Binder



- Target Protein
 - Native Human Angiotensin Converting Enzyme-Related Carboxypeptidase (ACE2) (PDB ID: 1R42)
 - SARS-CoV-2 entry point
- Goal: Design a binder to block SARS-CoV-2 entry
- Steps:
 1. Use **RFDiffusion** to generate the backbone of a fold
 2. Use **ProteinMPNN** to optimize the sequence
 3. Validate the design by visualizing the structure using **ESMFold**

Protein Design Workflow



Feedback Survey

<https://forms.office.com/r/dtnJNGGNAg>

11/29 NCHC × NTU - NVIDIA
BioNeMo Protein Design
Workshop 2024 Feedback Survey

