

GIST, 알파폴드3 '내부 표현 정보' 활용한 AI 기술로 신약 후보물질 예측 정확도 향상

- AI학과 이현주 교수팀 알파폴드3가 제공하는 단백질-약물 3차원 구조와 AI 내부 표현 정보 활용하는 AI 모델 '알파 DTA' 개발... 구조 데이터 부족한 신약 후보물질도 높은 정확도로 예측
- 신약 개발 기간·비용 절감 및 약물 재창출 활용 기대
- 국제학술지《Journal of Cheminformatics》게재



▲ (왼쪽부터) GIST AI학과 이현주 교수, 정민재 석사과정생, 전기전자컴퓨터공학과 박세진 박사

신약을 개발하려면 수많은 후보 약물 가운데 어떤 물질이 질병의 원인이 되는 단백질에 가장 잘 달라붙는지를 먼저 찾아야 한다. 하지만 이를 모두 실험으로 확인하려면 막대한 시간과 비용이 들기 때문에, **컴퓨터가 약물과 단백질의 결합 가능성을 미리 예측하는 인공지능(AI) 기술이 신약 개발의 핵심 도구로 주목받고 있다.**

GIST 연구팀은 이러한 예측의 정확도를 높이기 위해 세계 최고 수준의 **AI 단백질 구조 예측 모델인 '알파폴드3(AlphaFold3)'의 내부 표현 정보를 활용**하는 새로운 AI 모델을 개발했다.

광주과학기술원(GIST·지스트, 총장 임기철)은 **AI학과 이현주 교수 연구팀이 실험으로 단백질-약물 결합 구조를 확인하지 않아도 신약 후보물질의 표적 단백질 '결합 친화도(binding affinity)*를 예측할 수 있는 AI 모델 '알파 DTA(AlphaDTA)'를 개발**했다고 밝혔다.

'알파 DTA'는 구글 딥마인드의 AI 단백질 구조 예측 모델인 알파폴드3가 예측한 단백질-약물 복합체(complex)**의 3차원 구조와 구조를 예측하는 과정에서 생성되는 내부 표현 정보(embedding)를 함께 활용해 결합 친화도를 예측한다.

* 결합 친화도(binding affinity): 약물이 목표 단백질에 얼마나 강하게, 안정적으로 달라붙는지를

나타내는 지표이다. 이 값이 높을수록 효과적인 신약 후보로 평가된다.

**** 복합체(complex):** 둘 이상의 분자가 서로 결합하여 이루는 구조체를 통틀어 이르는 말이다. 본 논문에서는 특히 단백질과 약물이 결합한 구조체를 가리키며, 그 입체적인 모양과 결합 방식을 알면 두 물질이 얼마나 잘 맞물리는지 파악할 수 있다.

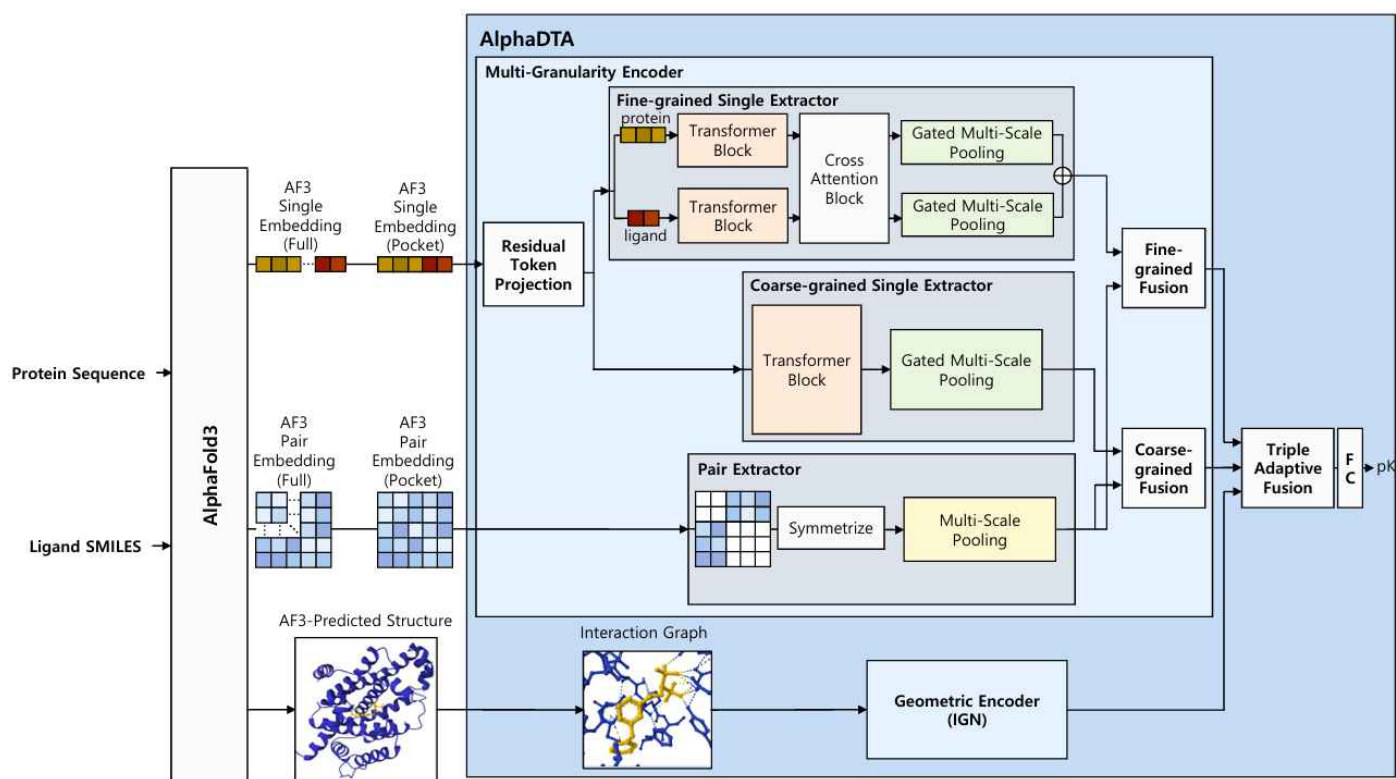
지금까지 약물과 단백질의 결합 친화도를 예측하는 AI 기술은 크게 '서열 기반'과 '구조 기반' 두 가지 방식으로 발전해 왔다.

단백질과 약물의 서열 정보만 이용하는 '서열 기반' 방식은 활용 가능한 데이터가 많지만 실제 결합 형태를 반영하지 못해 **정확도가 떨어질 수 있다.**

반면 **실험으로 규명된 단백질-약물 복합체의 3차원 구조를 활용하는 '구조 기반' 방식**은 **정확도가 높지만, 실험으로 확보된 구조 데이터가 매우 부족하다는 한계**가 있었다.

연구팀은 이러한 한계를 해결하기 위해 알파폴드3가 단백질과 약물의 결합 구조를 예측하는 과정에서 생성하는 **'내부 표현 정보(embedding)'**에 주목했다.

내부 표현 정보는 AI가 **단백질과 약물의 화학적·구조적 특징**을 학습할 수 있도록 **숫자로 표현한 정보**로, ▲ 단백질과 약물 각각의 특징을 담은 **'단일 표현 정보(single embedding)'**와 ▲ 두 물질의 관계와 상호작용을 담은 **'쌍 표현 정보(pair embedding)'**로 구성된다.

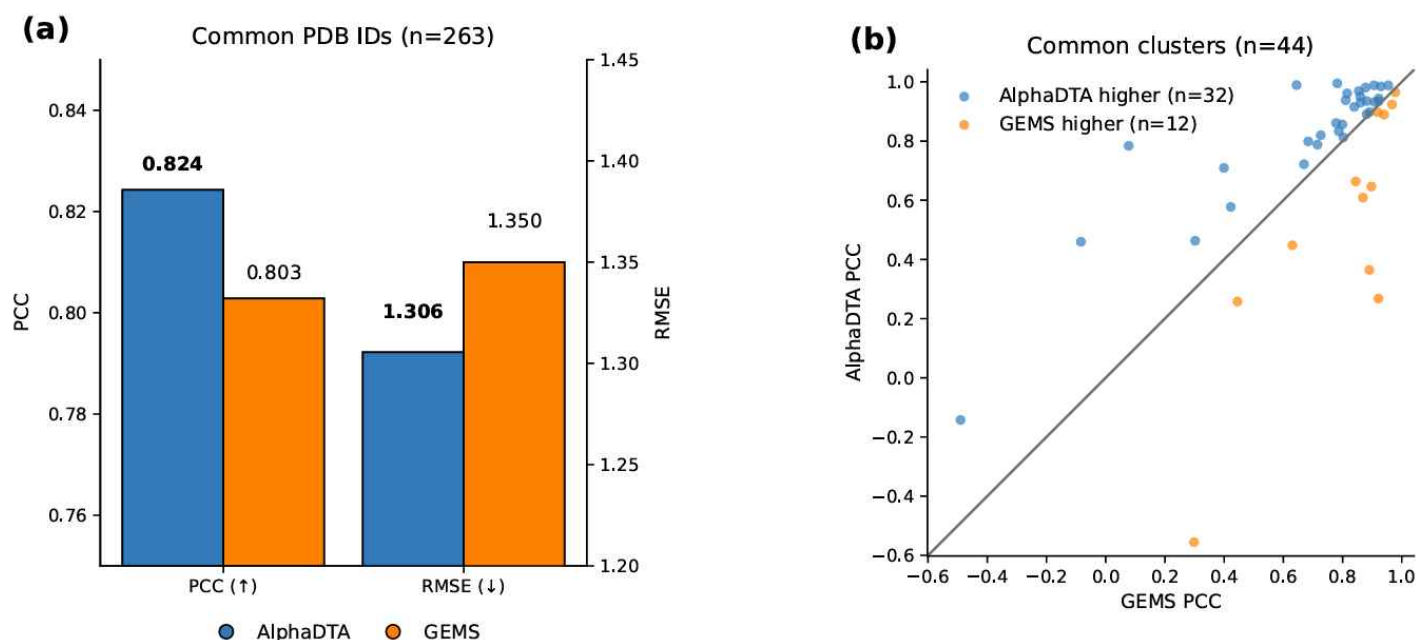


▲ **알파 DTA의 작동 개략도.** 단백질 서열과 약물의 화학구조를 입력하면 알파폴드 3(AlphaFold3)가 복합체의 3차원 구조와 함께 내부 표현 정보(단일 표현 벡터, 쌍 표현 벡터)를 생성한다. AlphaDTA는 이 구조 정보와 표현 정보를 함께 종합해 결합 친화도를 예측한다.

연구팀은 이러한 단일·쌍 표현 정보를 분석하는 '다중 해상도 인코더 (Multi-granularity Encoder)'와 단백질-약물 복합체의 3차원 구조 정보를 분석하는 '기하 인코더(Geometric Encoder)'를 결합한 AI 모델 '알파 DTA'를 개발했다.

'알파 DTA'는 단백질 서열과 약물 정보를 입력받아 알파폴드3가 생성한 3차원 구조와 내부 표현 정보를 함께 분석해 결합 친화도를 예측한다.

학습에 사용되지 않은 새로운 단백질-약물 조합으로 예측 성능을 검증한 결과, '알파 DTA'는 기존 서열 기반 방식보다 높은 예측 성능을 보였으며, 실험으로 확보한 구조 데이터가 없어도 기존 구조 기반 방식과 대등한 수준의 예측 성능을 보였다.



▲ 알파 DTA와 기존 최고 성능 모델(GEMS)의 예측 정확도 비교. 실험으로 확인된 구조 없이 알파폴드3가 예측한 구조만으로 성능을 비교했다. (a) 결합 친화도 예측 정확도(PCC, 높을수록 좋음)와 오차(RMSE, 낮을수록 좋음)를 비교한 결과, AlphaDTA(파란색)가 GEMS(주황색)보다 더 높은 정확도와 더 낮은 오차를 보였다. (b) 서로 다른 단백질 계열(cluster) 44종 각각에 대해 두 모델의 예측 정확도를 비교한 결과, 44개 중 32개 계열에서 AlphaDTA(점이 대각선 위쪽)가 GEMS보다 더 우수한 성능을 보여, 특정 사례에 국한되지 않고 폭넓은 단백질 종류에서 일관되게 앞서는 것으로 나타났다.

이번 기술은 실험으로 확보된 구조 데이터가 부족한 신약 후보물질에 대해서도 높은 예측 성능을 유지할 수 있어 신약 후보물질 선별의 효율을 높이고, 신약 개발 기간과 비용을 줄이는 것은 물론, 약물 재창출(Drug Repurposing)* 연구에도 폭넓게 활용될 것으로 기대된다.

* 약물 재창출(drug repurposing): 이미 다른 질병 치료용으로 승인된 약물을, 원래 목적과 다른 새로운 질병 치료에 활용하는 전략이다.

이현주 교수는 “이번 연구는 실험으로 단백질-약물 결합 구조를 확인하지 않아도 **신약 후보물질의 결합 친화도를 높은 정확도로 예측할 수 있다는 점에서 의미가 있다**”며 “**구조 데이터가 부족한 다양한 표적 단백질에 대해 유망한 후보물질을 효율적으로 선별하고, 신약 개발의 속도와 효율을 높이는 데 기여할 것으로 기대한다**”고 말했다.

GIST AI학과 이현주 교수가 지도하고 정민재 석사과정생이 제1저자로 수행했으며 박세진 박사가 참여한 이번 연구는 **과학기술정보통신부·한국연구재단 중견연구자지원사업, 과학기술정보통신부·정보통신기획평가원 SW컴퓨팅산업원천기술개발사업, 과학기술정보통신부 지스트-이노코어(GIST-InnoCORE) 사업의 지원을 받았다.**

연구 결과는 **화학정보학 분야 국제학술지 《Journal of Cheminformatics》에 2026년 7월 21일 온라인으로 게재됐다.**

한편 GIST는 이번 연구 성과가 **학술적 의의와 함께 산업적 응용 가능성까지 고려한 것으로, 기술이전 관련 협의는 기술사업화센터(hgmoon@gist.ac.kr)를 통해 진행할 수 있다고 밝혔다.**

논문 정보

1. 논문명, 저자정보

- 저널명 : Journal of Cheminformatics
(JCR 10% 이내, IF 7.9, 2025년 기준)
- 논문명 : AlphaDTA: integrating AlphaFold3 embeddings and 3D complex structures for drug–target binding affinity prediction
- 저자 정보 : 정민재(제1저자, AI학과 석사과정생), 박세진 박사(공동저자, 전기전자컴퓨터공학과·GIST AI-ACE InnoCORE 연구단), 이현주 교수 (교신저자, AI학과·전기전자컴퓨터공학과·GIST AI-ACE InnoCORE 연구단)