

پیش‌بینی بسیار دقیق ساختار ساختار پروتئین با AlphaFold

ارائه‌ای بر اساس مقاله Nature از تیم

جان جامپر، دمیس حسابیس و همکاران

شکاف عظیم بین توالی و ساختار

میلیاردها توالی شناخته شده
(Sequences)

تنها ~100,000 ساختار منحصر به فرد
(Structures)



تنگنای روش‌های آزمایشگاهی

بسیار پرهزینه

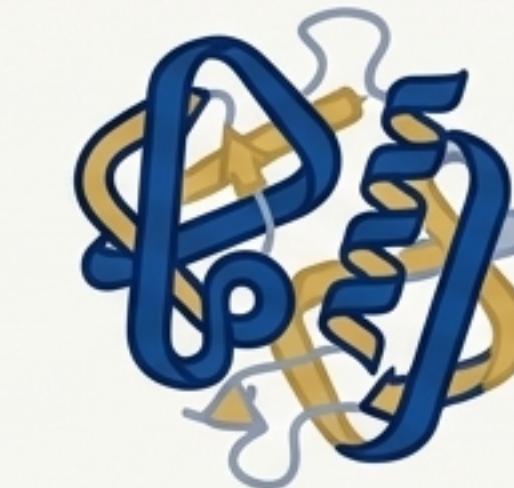


بسیار زمان بر
(ماه‌ها تا سال‌ها برای هر ساختار)



چالش بزرگ ۵۰ ساله: مسئله تاخورده‌گی پروتئین

A G L F V K K O S A
G G L M S R N S



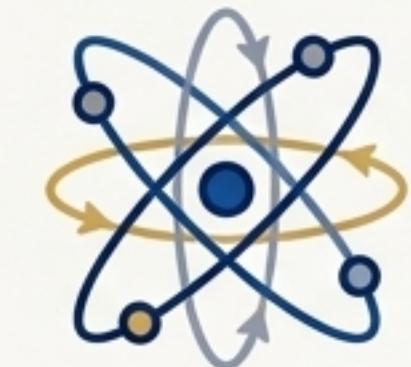
پیش‌بینی ساختار سه‌بعدی (3D) از روی توالی آمینواسیدها

ناکافی بودن رویکردهای محاسباتی پیشین

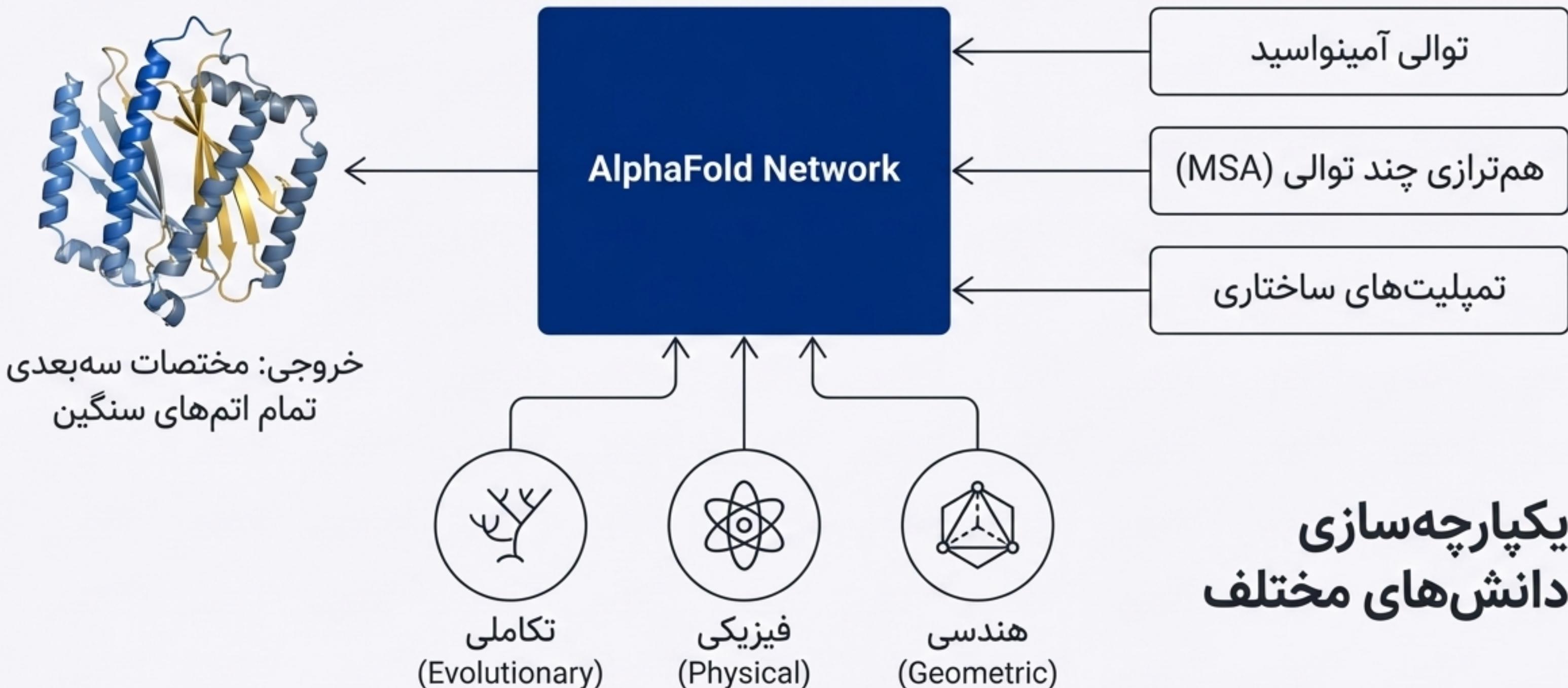
رویکرد مبتنی بر تکامل: دقت
محدود در غیاب ساختارهای مشابه
(Homologues)



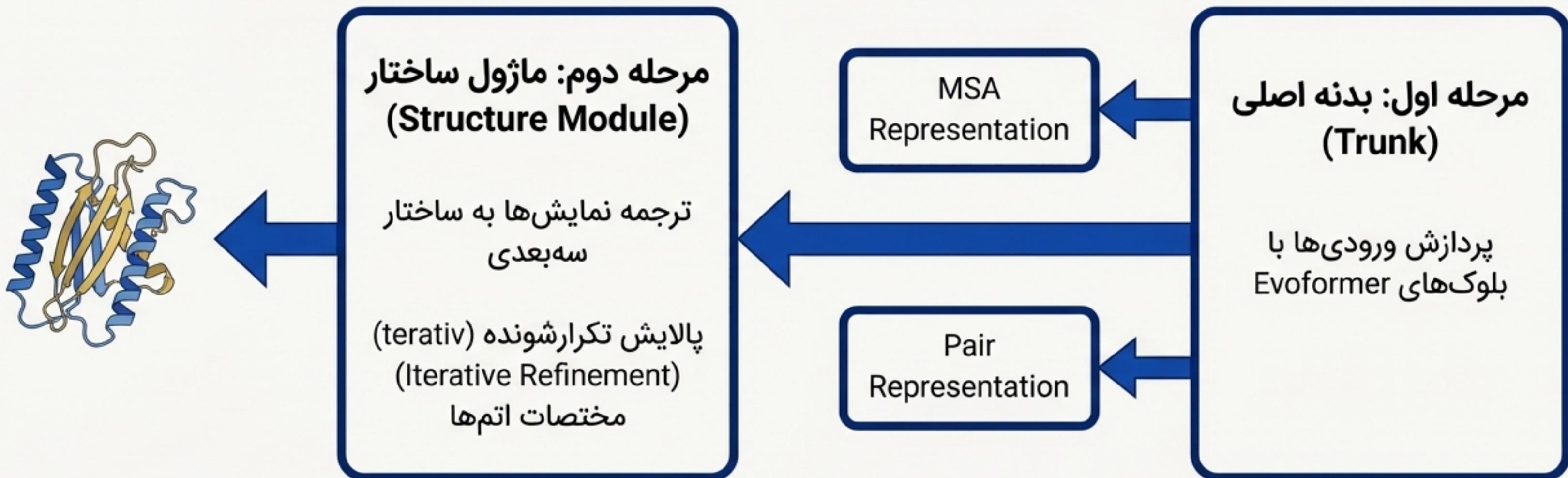
رویکرد مبتنی بر فیزیک:
پیچیدگی محاسباتی بسیار بالا



معماری End-to-End



معماری دو مرحله‌ای



بلوک Evoformer: هسته اصلی مدل

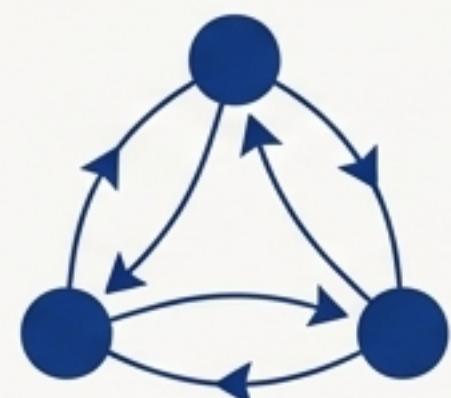
(MSA Representation) فضای تکاملی



تبادل اطلاعات بین دو فضا

(Pair Representation) فضای زوجی/هندسی

mekanizmehay_towjeh_klyidi



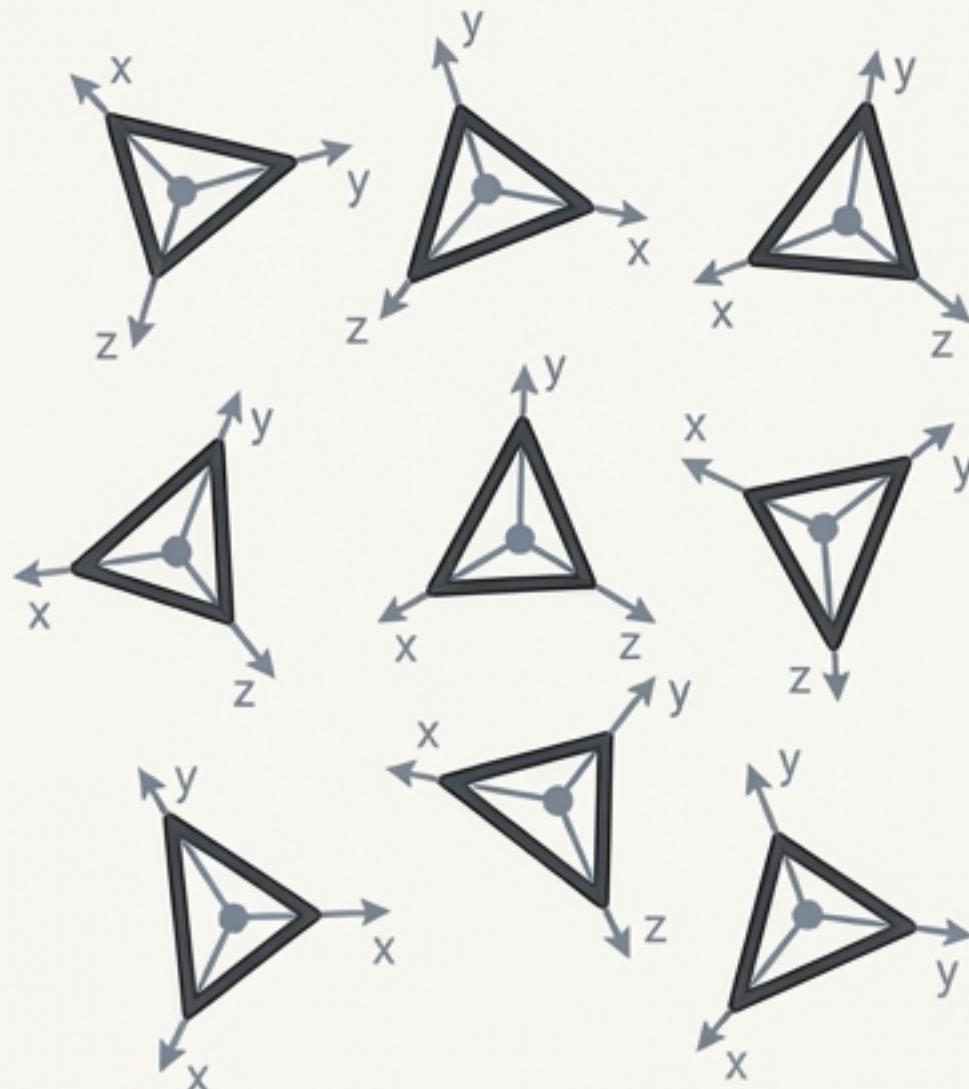
توجه مبتنی بر مثلث
(Triangle Self-Attention)



به روزرسانی ضربی مثلثی
(Triangle Multiplicative Update)

ماژول ساختار: ساخت اتم به اتم

گاز باقیماندها (Residue Gas)



: (Invariant Point Attention - IPA)

(Rotation/Translation Invariant)

نوآوری کلیدی: توجه نقطه ناوردا

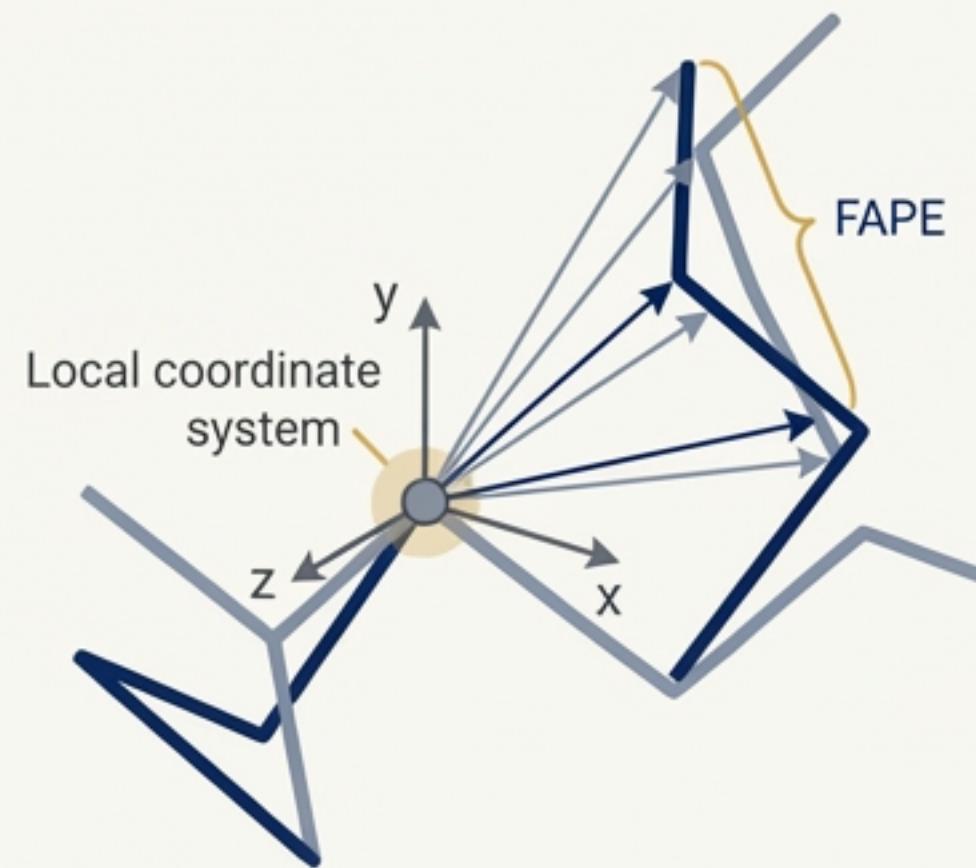
- استدلال مستقیم روی ساختار سه بعدی

- ناوردا نسبت به دوران و انتقال

تابع هزینه:

- خطای نقطه تراز شده با فریم

(Frame Aligned Point Error - FAPE)



ترفندهای آموزشی مهم



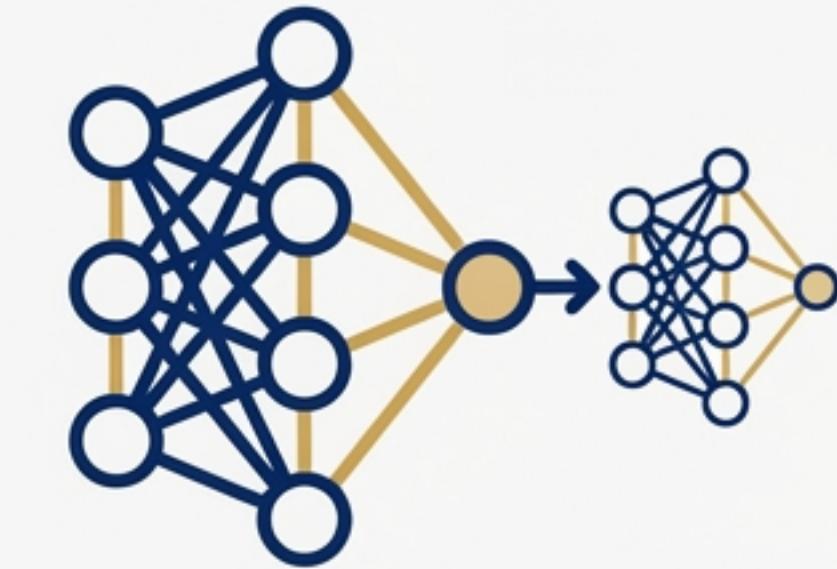
بازیافت (Recycling)

ارسال خروجی به ورودی برای پالایش‌های متوالی

A	C	D	E	F	F	G	H	I
A	C	D	E	F	F	G	H	I
A	D	D	E	F	G	H	I	K
K	L	L				K	K	L
M	L	L				K	P	L
M	M	N	N	P	Q	R	W	Y
T	V	N	P	P	Q	R	R	S
T	V	R	R	S	T	V	W	Y

یادگیری مشترک با Masked MSA Loss

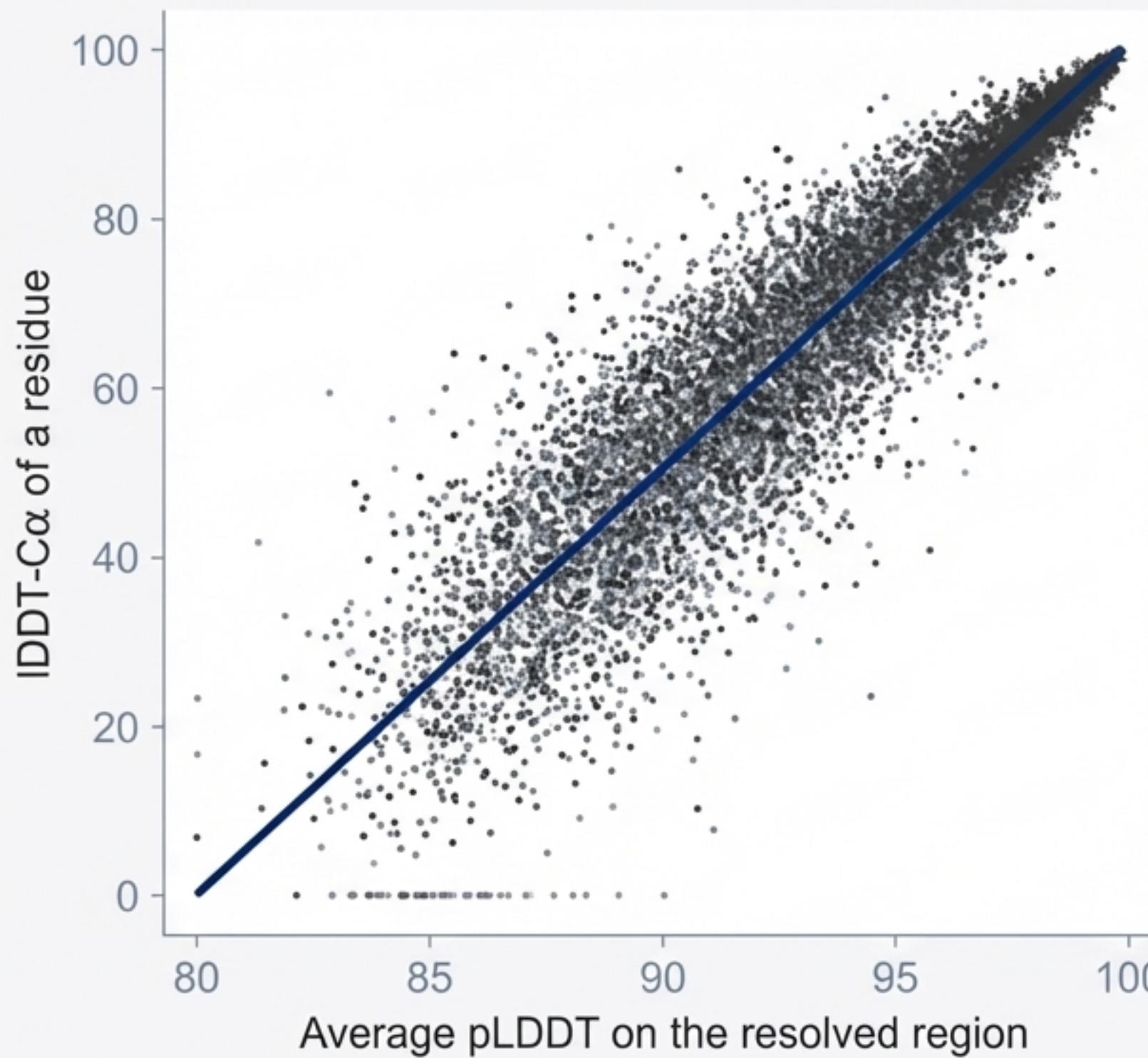
الهام‌گرفته از BERT برای درک روابط تکاملی



خود-تقطیر (Self-distillation) به سبک Noisy Student

یادگیری از داده‌های بدون برچسب (توالی‌های پروتئینی)

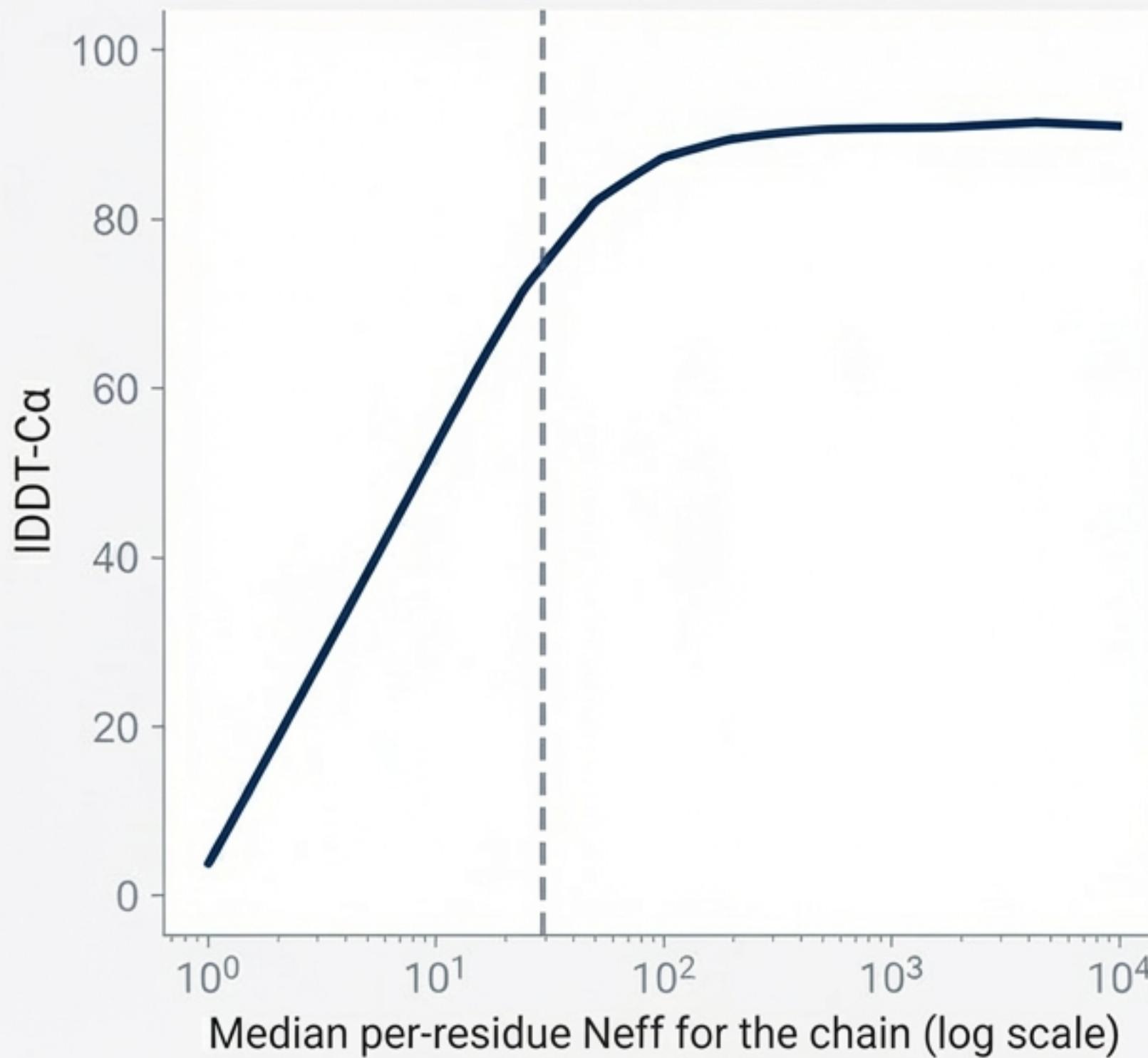
اطمینان و کالیبراسیون نتایج



پیش‌بینی دقت خود مدل
pLDDT : معیار اطمینان برای هر آمینو اسید
pTM : معیار اطمینان برای دقت کلی ساختار (TM-score)

همبستگی بسیار بالا بین دقت پیش‌بینی شده و واقعی

محدودیت‌ها



- وابستگی به اطلاعات تکاملی
- کاهش دقت در MSA های کم عمق (کمتر از ~ ۳۰ توالی)
- چالش در پیش‌بینی کمپلکس‌های پروتئینی
- چالش در پیش‌بینی کمپلکس‌های پروتئینی (Hetero-complexes)
- دقت پایین‌تر برای پروتئین‌هایی که ساختارشان به تعامل با زنجیره‌های دیگر وابسته است

تأثیر و گام‌های بعدی



دستیابی به دقیق در سطح آزمایشگاهی
راه حلی برای یک چالش بزرگ ۵۰ ساله در زیست‌شناسی



انقلاب در بیوانفورماتیک ساختاری
امکان پیش‌بینی ساختار در مقیاس پروتئوم (Proteome-scale)



گام‌های بعدی
پیش‌بینی کمپلکس‌های چندپروتئینی و دینامیک آنها

جمع‌بندی نهایی

→ معماری End-to-End با الهام از دانش فیزیکی و تکاملی →

نوآوری‌های کلیدی: Evoformer و ماژول ساختار مبتنی بر IPA



دقت بی‌سابقه و توانایی تخمین اطمینان خروجی



ابزاری قدرتمند برای تسريع اكتشافات در علوم زیستی



منابع برای مطالعه بیشتر

مقاله اصلی

Jumper, J., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*.

کد منبع و مدل

github.com/deepmind/alphafold

بلاگ‌پست و توضیحات تکمیلی

deepmind.com/blog/alphafold