Е	Experim	nents					Evof	former								Data					Oı	otimiza	ation								CAMEO
	notes	time code	mach	blk bl	k2 re	split row	col	tri	tri-attn	tri-mul t	tri-res	opm	relpos	msa	extra	aa chain	pre-msa	template	nnode	gpu bs	s accum	opt	bir warm	ep steps	LDDT-CA	GDT-TS	L<100	L<400	_ <inf(< th=""><th>S&lt;100 G&lt;4</th><th>00 G<inf 1="" 2="" 3="" 4="" a="" b="" b1="" b2="" b3<="" c="" th=""></inf></th></inf(<>	S<100 G<4	00 G <inf 1="" 2="" 3="" 4="" a="" b="" b1="" b2="" b3<="" c="" th=""></inf>
OpenFold																															
their weights		-	-	48	3	attn w/pair bias	attn	у	у	у	у	naive	у	128		-	-1	no			-			90000							5 0.71 0.94 0.93 0.85 0.93 0.89 0.86 0.77
af_repro		4d	tacc	48	3	attn w/pair bias	attn	у	у	у	у	naive	у	128	1024	13148	5 -1	no	14	3 1	3	AW	1e-3	6600	0.78	0.61	0.76	0.77	0.77	0.59 0.6	1 0.50 0.86 0.82 0.71 0.84 0.76 0.73 0.63
Mamba				48	_																										
mamba1 Subsample N		5d	tacc	48	3	mamba w/pair bias	attn	у	у	у	У	naive	у	128	1024	256 13148	5 -1	no	7	3 1	6	AW	1e-3	6600	0.77	0.59					0.85 0.82 0.70 0.83 0.75 0.73 0.63 0.83 0.77 0.75
their weights	ISA at II	ESTING	-	48	3	attn w/pair bias	attn	٧				naive	v	128	1024		-1	no				AW	10.2	90000	0.89	0.75					
trieli weigitis			-	48	3	attn w/pair bias	attn	y	y V	y	y	naive	y v	128	64		-1	no				AW		90000	0.88	0.75					
		-	-	48	3	attn w/pair bias	attn	v	v	v	v	naive	v	128	4	-	-1	no				AW		90000	0.88						
		-	-	48	3	attn w/pair bias	attn	y	y	y	y	naive	y		1024	-	-1	no	-			AW	1e-3	90000	0.89						
		-	-	48	3	attn w/pair bias	attn	y	y	y	y	naive	y	32	64	-	-1	no	-		-	AW	1e-3	90000	0.87						
		-	-	48	3	attn w/pair bias	attn	y	у	у	у	naive	у	32	4		-1	no	-		-	AW	1e-3	90000	0.85						
		-	-	48	3	attn w/pair bias	attn	у	у	у	у	naive	у	4	1024	-	-1	no	-		-	AW	1e-3	90000	0.86						
		-	-	48	3	attn w/pair bias	attn	у	у	у	у	naive	у	4	64	-	-1	no	-		-	AW	1e-3	90000	0.77						
		-	-	48	3	attn w/pair bias	attn	у	у	у	у	naive	у	4	4	-	-1	no	-		-	AW	1e-3	90000	0.61						
		-	-	48	3	attn w/pair bias	attn	у	у	у	у	naive	У		1024	-	-1	no	-		-	AW		90000	0.59						
		-	-	48	3	attn w/pair bias	attn	у	у	у	у	naive	У	1	64	-	-1	no	-		-	AW		90000	0.35						
		-	-	48	3	attn w/pair bias	attn	у	у	у	у	naive	у	1	4	-	-1	no	-		-	AW	1e-3	90000	0.32						
Rel Pos at TE	STING																														
their weights		-	-	48	3	attn w/pair bias	attn	у	у	у	У	naive	n	128	1024	-	-1	no	-		-	AW		90000	0.56	0.43					
(removing rel			ne iteratio	ons ın sli	des)																	AW	1e-3								
Remove Pair			tacc	48	3	attn w/pair bias	, ii-			-		naive	-	128	1024	13148	5 -1	no	7	2 .	6	A147	10.2	5800	0.69	0.49					
pair_test1 (slice pair_test3. I the			tacc	48	3	attn w/pair bias	attn attn	у	y	y	n n	naive LN	n y	128		13148	-	no			6			5800 6800		0.49					
msa_stack. pt			tacc	48	3	attn w/alibi	attn	n	y D	n	-	LN	y		0	13148	-	no			6			7000		0.33					
Inference wit			tacc	70	3	atal Wallbi	allii	- "				,	ý	120		13140	-1	110	,	U I	0	ATT		7000	0.43	0.14					
mamba1 but i				48	3	mamba w/pair bias	attn	у	у	у	v	naive	у	128	1024	13148	5 -1	no	7	3 1	6	AW	1e-3	6600	0.76	0.58					
af_repro but in				48	3	attn w/pair bias	attn	y	y	y	y	naive	y	128		13148		no			3			6600		0.13					
their weights t			,	48	3	attn w/pair bias	attn	y	ý	ý	ý	naive	ý	128	1024	13148	5 -1	no				AW	1e-3	90000	0.38	0.22					
Inference wit									,																						
mamba1 but i	nfer w/o t	tri-attn		48	3	mamba w/pair bias	attn	у	у	у	у	naive	у	128	1024	13148	5 -1	no	7	3 1	6	AW	1e-3	6600	0.16	0.05					
af_repro but in	nfer w/o t	ri-attn		48	3	attn w/pair bias	attn	у	у	у	у	naive	у	128		13148		no	14	3 1	3	AW	1e-3	6600		0.04					
their weights b	but infer v	w/o tri-attn		48	3	attn w/pair bias	attn	у	у	у	у	naive	у	128	1024	13148	5 -1	no				AW	1e-3	90000	0.37	0.21					
Mamba										$\bot$																					
mamba2		3.5d	tacc	48	3	mamba	attn	у	у	у	у	first/LN	у	128		13148		no		3 1		AW		7000	0.77	0.59					8 0.44 0.84 0.81 0.69 0.82 0.74 0.71 0.61 0.82 0.76 0.74
mamba3		3d	tacc	48	3	bi-mamba	attn	у	у	у	У	first/LN	У	128			5 2048	no		3 1		AW		7000		0.59	0.77	0.77	0.75	0.63 0.6	1 0.44
mamba3_noe			tacc	48	3	bi-mamba	attn	у	у	у	У	first/LN	У	128	0		5 2048	no		3 1		AW		7000	0.78	0.59					
mamba3_hub mamba3_skip		4d	mll	48 48	3	bi-mamba bi-mamba	attn	У	у	У	У	first/LN first/LN	У	128 128			5 2048 5 2048	no no		3 1	6	AW		4740 3000		0.48					
		- nt in triangl			_	module (inference)	attn	n	n	n	у	TIPST/LIN	У	128	1024	13148	5 2048	no	-/	3 1	ь	AW	1e-3	3000	0.06	0.02					
none	put - wii	at iii tilaligi	e to leed	48	3	attn w/pair bias	attn	v	V	v	у	naive	v	128	1024		-1	no				AW	10.3	90000	0.01	0.03					0.27 0.13 0.02 0.20 0.02 0.02 0.05 0.22 0.08 0.21
local16				48	3	attn w/pair bias	attn	v	v	y	y	naive	y	128			-1	no				AW		90000	0.06	0.03					0.31 0.16 0.07 0.23 0.07 0.08 0.09 0.25 0.13 0.25
not local 16				48	3	attn w/pair bias	attn	v	, v	y	y	naive	y	128			-1	no	-			AW		90000	0.36	0.35					0.54 0.44 0.33 0.50 0.34 0.31 0.29 0.51 0.39 0.44
not local 8				48	3	attn w/pair bias	attn	ý	ý	y	y	naive	y	128		-	-1	no	-			AW	1e-3	90000	0.52	0.53					0.66 0.60 0.47 0.65 0.50 0.44 0.40 0.65 0.53 0.55
not local 8				48	3	bi-mamba	attn	y	у	у	у	first/LN	у	128		13148		no	7	3 1	6	AW	1e-3	7000	0.47	0.39					0.62 0.55 0.43 0.59 0.46 0.42 0.36 0.60 0.49 0.52
Ablate IPA (ir	ference	)																													
no pair bias				48	3	attn w/pair bias	attn	у	у	у	у	naive	у	128	1024	-	-1	no	-		-	AW	1e-3	90000	0.11	0.02					0.35 0.21 0.12 0.27 0.12 0.13 0.16 0.30 0.18 0.30
no output pair				48	3	attn w/pair bias	attn	у	у	у	у	naive	у	128	1024	-	-1	no	-		-	AW	1e-3	90000	0.42	0.46					0.57 0.49 0.41 0.55 0.42 0.38 0.37 0.55 0.45 0.50
no pair bias				48	3	bi-mamba	attn	у	у	у	у	first/LN	у		1024	13148		no	7	3 1	6	AW	1e-3	7000	0.76	0.58					0.85 0.82 0.70 0.83 0.75 0.73 0.62 0.82 0.77 0.75
no output pair				48	3	bi-mamba	attn	у	у	у	у	first/LN	У		1024	13148		no	7	3 1	_		1e-3	7000	0.06	0.05					0.31 0.16 0.07 0.24 0.07 0.07 0.09 0.25 0.13 0.24
only output pa				48	3	bi-mamba	attn	у	у	у	у	first/LN	у		1024	13148		no	7	3 1			1e-3	7000	0.05	0.05					0.30 0.16 0.07 0.23 0.06 0.05 0.09 0.25 0.12 0.24
output z.mear				48	3	bi-mamba	attn	у	у	у	у	first/LN	у		1024	13148		no	7	3 1	_		1e-3	7000	0.37	0.26					0.54 0.45 0.36 0.50 0.37 0.36 0.35 0.50 0.41 0.47
output top 16				48	3	bi-mamba	attn	у	у	у	у	first/LN	у		1024	13148		no	7	3 1		AW		7000	0.64	0.48					0.76 0.69 0.58 0.73 0.64 0.60 0.49 0.72 0.65 0.69
output top 32 output local 16		nts		48	3	bi-mamba bi-mamba	attn	У	y	у	y	first/LN first/LN	y	128		13148		no no	7	3 1		AW		7000 7000	0.71	0.52					0.81 0.76 0.64 0.78 0.70 0.66 0.55 0.78 0.71 0.69 0.56 0.47 0.39 0.51 0.41 0.41 0.37 0.51 0.44 0.50
output local 16				48	3	bi-mamba bi-mamba	attn	y	y	y	У	first/LN first/LN	y	128		13148		no	7	3 1		AW		7000		0.24					0.56 0.47 0.39 0.51 0.41 0.41 0.37 0.51 0.44 0.5
output local 33				48	3	bi-mamba	attn	y	y	y	y y	first/I N	y	128		13148		no		3 1		AW		7000		0.34					0.77 0.74 0.61 0.75 0.65 0.63 0.55 0.75 0.67 0.67
output not local				48	3	bi-mamba	attn	v	v	v	v	first/I N	v	128		13148		no		3 1	-	AW		7000	NaN	0.07					3.10 0.02 0.01 0.00 0.01 0.40 0.00 0.00
Col + Pair						2. manou		,	,		,		,	0		10.40				- '				, 000							
mamba4		3d	tacc	48	3	bi-mamba	1	V	У	у	у	first/LN	у	128	1024	13148	5 2048	no	7	3 1	6	AW	1e-3	7000	0.71	0.51					
mamba5		18h	tacc	48	3	bi-mamba	attn	n	n	n		nomsa/1/LN		128	0	13148	5 2048	no	7	3 1	6	AW		4300	0.35	0.06	0.42	0.35	0.30	0.16 0.0	6 0.01 0.57 0.39 0.32 0.47 0.36 0.37 0.34 0.48 0.40 0.4
mamba5			tacc	48	3	bi-mamba	attn	n	n	n	у	^/out2	у	128	0	13148	5 2048	no	7	3 1	6	AW	1e-3	4300	0.35	0.06					
mamba5			tacc	48	3	bi-mamba	attn	n	n	n		^/out2/rope		128	0	13148	5 2048	no	7	3 1	6	AW	1e-3	3800	0.35	0.06					
Col + Pair																															
mamba3.5 ++		2d	tacc	48	3	bi-mamba	attn	у	n	у	у	first/LN	у	128			5 2048	no			6			7000		0.54					
mamba3.5 ++		2d	tacc	48	3	bi-mamba	attn	у	у	n	у	first/LN	у	128	0	13148	5 2048	no	7	3 1	6	AW	1e-3	7000	0.77	0.59	0.74	0.78	0.77	0.56 0.6	1 0.53 0.86 0.82 0.70 0.84 0.76 0.73 0.63 0.83 0.77 0.75
Triangle Attn																															
no triangle_bia				48	3	attn w/pair bias	attn	у	у	у	у	naive	у	128		-	-1	no			-			90000	0.24						3 0.01 0.44 0.32 0.24 0.38 0.24 0.24 0.26 0.40 0.29 0.38
no triangle_Q				48	3	attn w/pair bias	attn	у	у	у	у	naive	у	128	1024	-	-1	no	-		-	AW	1e-3	90000	0.36	0.25	0.49	0.37	0.28	0.37 0.2	8 0.12 0.55 0.42 0.36 0.48 0.37 0.38 0.41 0.49 0.41 0.49
Mambafy Col	and Tri																														
mamba6 mamba7		52h	tacc	48 48	3		bi-mamb bi-mamb				-	first/LN first/LN	у	128 128			5 2048 5 2048	no no		3 1	6			7000 5600	0.77	0.57					

Experiments	<b>,</b>					Evofo	ormer	1								Data					0	ptimizatio	n								CAME	0					
notes time	code m	ach t	olk blk2	re split	t row	col		tri-attn	tri-mul	tri-res	opm	relpos	msa	extra		chains pr	e-msa t	emplate	nnode g	pu bs	accur	n opt bir	warm e	p steps	LDDT-CA	GDT-TS	L<100	L<400	L <inf g<100<="" th=""><th></th><th></th><th></th><th>3 C</th><th>1 2</th><th>3 4</th><th>b1</th><th>b2 b3</th></inf>				3 C	1 2	3 4	b1	b2 b3
OpenFold																																					
their weights - af repro 4d		- 1	48 48	3	attn w/pair bias		у	у	у	у	naive	у		1024			-1	no				AW 1e-3		90000					0.90 0.75 0.77 0.59								
af_repro 4d Pair to MSA	ti	icc 4	18	3	attn w/pair bias	attn	у	у	У	У	naive	у	128	1024		131485	-1	no	14	3 1	3	AW 1e-3		6600	0.78	0.61	0.76	0.77	0.77 0.59	0.61	0.50	J.86 U.8	32 0.71	0.84 0.76	0.73 0.6	3	
mamba6 + p2m v1	ti	acc 4	48	3	bi-mamba	bi-mamba	ау	у	n	у	first/LN	у	128	0		131485 2	2048	no	7	3 1	6	AW 1e-3		4200	0.17	0.03											
mamba6 + p2m_v2	ti	icc 4	48	3	bi-mamba	bi-mamba		у	n	y	first/LN	у	128	0		131485 2	2048	no	7	3 1	6	AW 1e-3		2500	0.05	0.02											
LION		_																																			
mamba6 mamba6		acc 4	48 48	3	bi-mamba bi-mamba	bi-mamba bi-mamba	- ,	у	n	У	first/LN first/LN	у	128	-		131485 2 131485 2		no no		3 1		LION 3e-4		3600 6300	0.73	0.53		ad spil	0.79 0.64	0.04	0.50	0.00.0	05 0 74	0.00 0.7	0.70.06		0.00 0.7
mamba6 + beta2=0.98 + clin			+o 48	3	bi-mamba	bi-mamba	- ,	y	n n	y	first/I N	y	128	-		131485 2		no		3 1		LION 1e-4		3600		0.63			ster then wor		0.56	J.00 U.	35 0.74	0.00 0.73	0.76 0.6	5 0.65	0.60 0.77
mamba6 + beta1=0.95 + bet				3	bi-mamba	bi-mamba	- ,	y	n	y	first/LN	y	128	-		131485 2		no				LION 3e-4		3600	0.71	0.10		anio ia	Jaco Gren Wei								
longer lengths																																					
mamba6 + sec maybe, but no	time		48	3	bi-mamba	bi-mamba	ау	у	n	у	first/LN	у	128	0		131485 2	2048	no	7	3 1	6	AW 1e-3		3200	0.33	0.09						_					
trimamba with pair bias mamba6 B x bias		acc 4	40	3	bi-mamba	bi-mamba			n		first/I N		128	0		131485 2	0040	no	7	2 4		LION 1e-4		5500	0.73	0.52											
mamba6 dt x bias		acc 4		3	bi-mamba	bi-mamba	- ,	mamba2	n	y	first/I N	y	128	_		131485 2		no no		3 1		LION 1e-4		4800	0.73	0.52											
mamba6 dt,B,C x bias	ti	acc 4	48	3	bi-mamba	bi-mamba	a y	mamba5	n	y	first/LN	y	128	0		131485 2	2048	no	7	3 1	6	LION 1e-4		7000	B - 0.09												
mamba6, pair_dropout=0	ti	acc 4	48	0	bi-mamba	bi-mamba	a y	mamba5	n	у	first/LN	у	128	0		131485 2	2048	no	7	3 1	6	LION 1e-4		7000	B - 0.08												
recycle																																					
mamba6 40h mamba6. clip val 40h		acc 4		0	bi-mamba bi-mamba	bi-mamba	-	у	n	У	first/LN first/LN	у	128 128			131485 2 131485 2		no no				LION 1e-4 LION 1e-4		7000 6300					0.72 0.62				31 0.69	0.82 0.7	5 0.72 0.6	2 0.82	0.76 0.75
mamba6, clip_val 40h no side chain loss	li	icc 2	+0	U	DI-Mamba	bi-mamba	1 y	у	n	у	IIISULIN	у	120	U		131465 2	2046	no	-/	3 1	0	LION 16-4		6300	0.77	0.59	HOL 1	worth tr	ying, its roug	ny trie s	arne spe	80	_				
mamba6 + no side chai 40h	ti	acc 4	48	0	bi-mamba	bi-mamba	ау	у	n	у	first/LN	у	128	0		131485 2	2048	no	7	3 1	6	LION 1e-4		7000	0.76	0.57											
new impl (mamba6)		_		_										_					_																		
short schedule( 1 warm 26hr			48 48 48 48		bi-mamba bi-mamba	bi-mamba bi-mamba	-	y	n n	y	first/LN first/LN	y	128	0		131485 2 131485 2		no no				LION 1e-4 LION 1e-4		8000 <b>5000</b>	0.77 0.74	0.59											
	new ta		48 48		bi-mamba	bi-mamba	-	v	n	V	first/LN	V	128			131485 2		no				LION 1e-4		16000	0.74	0.58											
splitformer							,	,																-												$\overline{}$	
		acc 4	48 48	0 y	bi-mamba	bi-mamba	ау	у	n	у	first/LN	у	128	0		131485 2	2048	no	7	3 1	6	LION 1e-4	5 8		0.79	0.62											
	4:50/hr														256									8000	0.78	0.60											
96hr 110h	nr.	-													256 384								5 1	7 18000	0.85	0.69	0.80	0.85	0.85 0.68	0.70	0.64						
	new ta	acc 7	72 24	0 v	bi-mamba	bi-mamba	a v	٧	n	٧	first/LN	v	128	0		131485 2	2048	no	7	3 1	6	LION 1e-4		9000	0.00	0.70											
	new ta	- 11			bi-mamba	bi-mamba	-			у	first/LN	y	128			131485 2	2048	no				LION 1e-4		9000													
sparse triange attention																																					
splitformer2 - latent MSA	1.0		10 10								C		400	_	050	101105	2040		_			110114			0.77												
short schedule v0 - not lon			48 48	0 y	bi-mamba	bi-mamba	а у	У	n	У	first/LN	у	128	0	256	131485 2	2048	no	/	3 1	ь	LION 1e-4	1	U	0.77	0.59						—	_				
need more epochs			48 48	0 y	bi-mamba	bi-mamba	ау	у	n	у	first/LN	у	128	0	256	38,000 1	1024	no	1	4 16	2	LION 1e-4	1 4	1	0.06	0.02											
need more epochs, bs=128	24h n	ıli3 d	48 48	0 y	bi-mamba	bi-mamba	a y	у	n	у	first/LN	у	32	0	128	80,000 1	1024	no	1	8 64	1	LION 1e-4	2 1	3	0.35	0.05											
need more epochs, bs=128			48 48	0 y	bi-mamba	bi-mamba	-	у	n	у	first/LN	у	32	0			1024	no	1	8 64		LION 1e-4			0.06	0.02											
need more epochs, bs=128 need more epochs, bs=128			24 24 12 12		bi-mamba bi-mamba	bi-mamba bi-mamba	-	у	n n	У	first/LN first/LN	у	32 32	0		80,000 1 80,000 1		no no		8 <b>64</b> 8 <b>64</b>		LION 1e-4 LION 1e-4			0.09	0.03											
need more epochs, bs=126	32hr	au   1	12 12	о у	DI-Mamba	DI-Mamba	1 y	у	n	у	IIISULIN	у	32	U	120	80,000	1024	no	1	0 64	٠,	LION 16-4	2 1		0.19	0.05											
		lut 2	24 24	0 y	bi-mamba	bi-mamba	ау	у	n	у	first/LN	у	32	0	128	80,000 1	1024	no	1	4 128	3 1	LION 1e-4		-	0.25	0.04											
need bs=128	16hr+																						2 1	6	0.26	0.05											
short schedule v1																																					
	8hr n	ıll3   2	24 24	0 y	bi-mamba	bi-mamba	ау	у	n	У	first/LN	у	32	0	128	80,000 1	1024	no	1	8 1	16	LION 1e-4	2 1		0.32	0.06											
		117 2	24 24	0 v	bi-mamba	bi-mamba	a v	v	n	v	first/LN	v	32	0	128	80,000 1	1024	no	1	8 16	1	LION 1e-4		•	0.60	0.38											
short schedule v2 - Ir/wd				- ,			,	,		,		,		-						- 10				-													
wd=1e-2			24 24		bi-mamba	bi-mamba	-	у	n	у	first/LN	у		-		80,000 1		no				LION 2e-4		•	0.61	0.41											
			24 24		bi-mamba	bi-mamba	-	у	n	у	first/LN	у	32	0		80,000 1		no		8 16		LION 4e-4		•	NaN												
wd=1e-1			24 24		bi-mamba bi-mamba	bi-mamba bi-mamba		у	n	у	first/LN first/LN	у	32	0	_	80,000 1 80.000 1	_	no no		4 32	_	LION 8e-4			NaN							_					
wd=1e-1 wd=1e-3			24 24 24 24		bi-mamba	bi-mamba	-	y	n n	y	first/LN	y				80,000		no		. 02		LION 1e-4		•	similar												
wd=1e-4			24 24		bi-mamba	bi-mamba	-	у	n	у	first/LN	y	-		128	80,000 1	1024	no	1	8 16	1	LION 1e-4	2 1	6	similar												
short schedule v3 - longer	sequenc	deep	er MSA h	elps																																	
			24 24		bi-mamba	bi-mamba	-	у	n	у	first/LN	у	64			80,000 1		no				LION 2e-4			0.64	0.43											
			24 24 24 24		bi-mamba bi-mamba	bi-mamba bi-mamba	,	у	n	У	first/LN first/LN	У	16 32			80,000 1		no no		8 16 8 16		LION 2e-4 LION 2e-4		-	0.61	0.39											
			24 24		bi-mamba bi-mamba	bi-mamba bi-mamba	-	v	n n	y	first/LN	y	32			80,000 1		no no		8 16		LION 2e-4 LION 2e-4		-	0.44	0.21											
short schedule v4		1		,		22.7100	,	,		,		,				,				,,,		20 4				-											
fewer msa layers - (per layer			12 24		bi-mamba	bi-mamba	-	у	n	у	first/LN	у		0		80,000 1		no	1	8 16		LION 2e-4		-	0.59	0.36											
fewer pair layers	34hr f		24 12	. ,	bi-mamba	bi-mamba	,	у	n	у	first/LN	у	32	0		80,000 1		no	1	8 16		LION 2e-4		-	0.61	0.39											
no wd on bias/norm cropfirst/last + ^	25hr n		24 24		bi-mamba bi-mamba	bi-mamba bi-mamba	,	у ,,	n	у	first/LN first/LN	y	32	0		80,000 1		no no		8 16 8 16		LION 2e-4		-	0.64	0.41											
fape weighted by b factor + /			24 24 24 24		bi-mamba bi-mamba	bi-mamba bi-mamba	-	y	n n	y	first/LN	y				80,000 1		no no				LION 2e-4 LION 2e-4		-	0.63	0.41											
short schedule v5		- 1	-	- ,	2namod	2ua	,	,		,		,		,	0	,									2.01	20											
mamba init like transformer			24 24		bi-mamba	bi-mamba		у	n	у	first/LN	у				80,000 1		no				LION 2e-4		-	worse												
no wd on bias/norm/embed	d/mamk n	117 2	24 24	0 y	bi-mamba	bi-mamba	ау	у	n	у	first/LN	у	32	0	128	80,000 1	1024	no	1	8 16	1	LION 2e-4	2 1	6	0.64	0.42											

Experi	ments							Evofo	ormei									Data					O	ptimiza	tion									(	CAME	0						
notes	time c	ode ma	ch blk	blk2	re s	olit	row	col	tri	tri-attn	tri-mu	l tri-res	opm	relpos	msa	extr	a aa	chain	s pre-ms	a template	nnode	gpu b	accun	n opt	bir war	rm ep	steps	LDDT-CA	GDT-TS	L<100	L<400	L <inf (<="" th=""><th>G&lt;100</th><th>G&lt;400</th><th>G<inf< th=""><th>Α</th><th>в с</th><th>1</th><th>2</th><th>3 4</th><th>b1</th><th>b2 b</th></inf<></th></inf>	G<100	G<400	G <inf< th=""><th>Α</th><th>в с</th><th>1</th><th>2</th><th>3 4</th><th>b1</th><th>b2 b</th></inf<>	Α	в с	1	2	3 4	b1	b2 b
OpenFold																																				$\neg$						
heir weights	-	-	48		3	at	ttn w/pair bias	attn	у	у	у	у	naive	у	128	102	4	-	-1	no	-		-	AW 1	e-3		90000	0.89	0.75	0.86	0.89	0.90	0.75	0.75	0.71	0.94	0.93 0.8	5 0.93	0.89 0.	.86 0.77	7	
af_repro	4d	ta	c 48		3	at	ttn w/pair bias	attn	у	у	у	у	naive	у	128	1024	4	13148	35 -1	no	14	3 1	3	AW 1	e-3		6600	0.78	0.61	0.76	0.77	0.77	0.59	0.61	0.50	0.86	0.82 0.7	1 0.84	0.76 0.	73 0.63	3	
MSA diversity - AF																																				$\neg$						
			48		3	at	ttn w/pair bias	attn	у	у	у	у	naive	у	128	102	4	-	-1	no	-		-	AW 1	e-3		90000	0.89	0.75	0.86	0.89	0.90	0.75	0.75	0.71	0.94	0.93 0.8	5 0.93	0.89 0.	86 0.77	0.94	0.89 0.
sensitivity=7.5 max-se	eq-id=0.9	5	48		3	at	ttn w/pair bias	attn	у	у	у	у	naive	у	128	102	4		-1	no	-		-	AW 1	e-3		90000	0.88	0.75	0.86	0.89	0.88	0.74	0.76	0.74	0.93	0.91 0.84	4 0.93	0.88 0.	85 0.77	0.93	0.88 0.
sensitivity=7.5 max-se	eq-id=0.8		48		3	at	ttn w/pair bias	attn	у	у	у	у	naive	у	128	102	4		-1	no	-		-	AW 1	e-3		90000	0.88	0.75	0.86	0.89	0.88	0.74	0.76	0.74	0.93	0.91 0.84	4 0.93	0.88 0.	85 0.77	0.93	0.88 0.
sensitivity=7.5 max-se	eq-id=0.5		48		3	at	ttn w/pair bias	attn	v	y	v	v	naive	v	128	102	4	-	-1	no	-			AW 1	e-3		90000	0.30	0.09	0.63	0.30	0.17	0.43	0.07	0.01	0.54	0.33 0.3	1 0.42	0.32 0.	36 0.37	0.44	0.36 0.
sensitivity=7.5 max-se	eq-id=0.3		48		3	at	ttn w/pair bias	attn	v	y	v	v	naive	v	128	102	4	-	-1	no	-			AW 1	e-3		90000	0.30	0.09	0.63	0.30	0.17	0.43	0.07	0.01	0.54	0.33 0.3	1 0.42	0.32 0.	36 0.37	0.44	0.36 0.
MSA diversity - Split	tformer						•		Ť	,																										$\neg$						
			48	48	0	v	bi-mamba	bi-mamba	v	٧	n	v	first/LN	V	128	0	256	13148	35 2048	no	7	3 1	6	LION 1	e-4	35		0.85	0.70													
sensitivity=7.5 max-se	eq-id=0.9	5	48	48	0	v	bi-mamba	bi-mamba	v	y	n	v	first/LN	y	128	0	256	13148	35 2048	no	7	3 1	6	LION 1	e-4	35		0.84	0.69													
sensitivity=7.5 max-se	eq-id=0.8								Ė	- 1		Ĺ																0.84	0.69													
sensitivity=7.5 max-se	eq-id=0.5																											0.84	0.69													
sensitivity=7.5 max-se	eq-id=0.3																											0.84	0.69													
BERT pre-training	İ																																			$\neg$						
mr=0.15		ml	13 48				bi-mamba	bi-mamba	-			-		V	128	0	256	13148	35 2048	no	1	8 16	3 1	LION 1	e-4 5	5 25		?														
mr=0.50		gli	ıt 48				bi-mamba	bi-mamba	-			-		y	128	0	256	13148	35 2048	no	1	4 16	3 1	LION 1	e-4 5	5 25		?														
mr=0.70		fla	d 48				bi-mamba	bi-mamba	-			-		y	128	0	256	13148	35 2048	no	1	4 16	3 1	LION 1	e-4 5	5 25		?														
ong schedule																																				$\neg$						
new weight decay/cro	pj 48hr,4:	40/ep ta	c 48	48	0	v	bi-mamba	bi-mamba	v	٧	n	v	first/LN	V	128	0	256	13148	35 2048	no	7	3 1	6	LION 1	e-4 5	5 10/2	5	0.80	0.63													
	96hr																																									
																	w/dis	stillation	dataset																							
v/bert pre-training	48hr 4:	40/ec ta	c 48	48	0	v	bi-mamba	bi-mamba	v	v	n	v	first/LN	v	128	0	256	13148	35 2048	no	7	3 1	6	LION 1	e-4 5	5 10/2	5	0.80	0.63													
	96hr								Ĺ					ľ																												
																	w/dis	stillation	dataset									?														
																	.v/Gil																									