Experiments					Evoformer												Data						Optimization									CAMEO																	
	notes	time	code	mach	blk	blk2	re	split	row	col	tri	tri-attn	tri-mul	tri-res	opm	relpos	msa	extra	aa	chains	pre-msa	template	nnode	gpu	bs	accum	opt	blr	warm	ep	steps	LDDT-CA	GDT-TS	L<100	L<400	L<inf	G<100	G<400	G<inf	A	B	C	1	2	3	4	b1	b2	b3
OpenFold																																																	
their weights		-		-	48		3		attn w/pair bias	attn	y	y	y	y	naive	y	128	1024		-	-1	no	-	-	-	-	AW	1e-3			90000	0.89	0.75	0.86	0.89	0.90	0.75	0.75	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	y	y	y	y	naive	y	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	0.86	0.82	0.71	0.84	0.76	0.73	0.63			
Mamba																																																	
mamba1		5d		tacc	48		3		mamba w/pair bias	attn	y	y	y	y	naive	y	128	1024	256	131485	-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
Subsample MSA at TESTING																																																	
their weights		-		-	48		3		attn w/pair bias	attn	y	y	y	y	naive	y	128	1024		-	-1	no	-	-	-	-	AW	1e-3			90000	0.89	0.75																
		-		-	48		3		attn w/pair bias	attn	y	y	y	y	naive	y	128	64		-	-1	no	-	-	-	-	AW	1e-3			90000	0.88																	
		-		-	48		3		attn w/pair bias	attn	y	y	y	y	naive	y	128	4		-	-1	no	-	-	-	-	AW	1e-3			90000	0.88																	
		-		-	48		3		attn w/pair bias	attn	y	y	y	y	naive	y	32	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	y	y	y	naive	y	32	4		-	-1	no	-	-	-	-	AW	1e-3			90000	0.85																	
		-		-	48		3		attn w/pair bias	attn	y	y	y	y	naive	y	4	1024		-	-1	no	-	-	-	-	AW	1e-3			90000	0.86																	
		-		-	48		3		attn w/pair bias	attn	y	y	y	y	naive	y	4	64		-	-1	no	-	-	-	-	AW	1e-3			90000	0.77																	
		-		-	48		3		attn w/pair bias	attn	y	y	y	y	naive	y	4	4		-	-1	no	-	-	-	-	AW	1e-3			90000	0.61																	
		-		-	48		3		attn w/pair bias	attn	y	y	y	y	naive	y	1	1024		-	-1	no	-	-	-	-	AW	1e-3			90000	0.59																	
		-		-	48		3		attn w/pair bias	attn	y	y	y	y	naive	y	1	64		-	-1	no	-	-	-	-	AW	1e-3			90000	0.35																	
		-		-	48		3		attn w/pair bias	attn	y	y	y	y	naive	y	1	4		-	-1	no	-	-	-	-	AW	1e-3			90000	0.32																	
Rel Pos at TESTING																																																	
their weights		-		-	48		3		attn w/pair bias	attn	y	y	y	y	naive	**n**	128	1024		-	-1	no	-	-	-	-	AW	1e-3			90000	0.56	0.43																
(removing rel pos at different recycle iterations in slides)																											AW	1e-3																					
Remove Pair campaign																																																	
pair_test1 (slide 338)		4.5d		tacc	48		3		attn w/pair bias	attn	y	y	y	**n**	naive	**n**	128	1024		131485	-1	no	7	3	1	6	AW	1e-3			5800	0.69	0.49																
pair_test3. I think pair_t		4.5d		tacc	48		3		attn w/pair bias	attn	y	y	y	**n**	**LN**	**y**	128	1024		131485	-1	no	7	3	1	6	AW	1e-3			6800	0.73	0.53																
msa_stack. pure transfo		3d		tacc	48		3		**attn w/alibi**	attn	**n**	**n**	**n**	**-**	**LN**	y	128	**0**		131485	-1	no	7	3	1	6	AW	1e-3			7000	0.43	0.14																
Inference without pair bias																																																	
mamba1 but infer w/o pair bias					48		3		mamba w/pair bias	attn	y	y	y	y	naive	y	128	1024		131485	-1	no	7	3	1	6	AW	1e-3			6600	0.76	0.58																
af_repro but infer w/o pair bias					48		3		attn w/pair bias	attn	y	y	y	y	naive	y	128	1024		131485	-1	no	14	3	1	3	AW	1e-3			6600	0.27	0.13																
their weights but infer w/o pair bias					48		3		attn w/pair bias	attn	y	y	y	y	naive	y	128	1024		131485	-1	no					AW	1e-3			90000	0.38	0.22																
Inference without tri-attn																																																	
mamba1 but infer w/o tri-attn					48		3		mamba w/pair bias	attn	y	y	y	y	naive	y	128	1024		131485	-1	no	7	3	1	6	AW	1e-3			6600	0.16	0.05																
af_repro but infer w/o tri-attn					48		3		attn w/pair bias	attn	y	y	y	y	naive	y	128	1024		131485	-1	no	14	3	1	3	AW	1e-3			6600	0.11	0.04																
their weights but infer w/o tri-attn					48		3		attn w/pair bias	attn	y	y	y	y	naive	y	128	1024		131485	-1	no					AW	1e-3			90000	0.37	0.21																
Mamba																																																	
mamba2		3.5d		tacc	48		3		**mamba**	attn	y	y	y	y	**first/LN**	y	128	1024		131485	-1	no	7	3	1	6	AW	1e-3			7000	**0.77**	**0.59**	0.77	0.76	0.74	0.64	0.58	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	y	y	y	y	first/LN	y	128	1024		131485	2048	no	7	3	1	6	AW	1e-3			7000	**0.77**	**0.59**	0.77	0.77	0.75	0.63	0.61	0.44										
mamba3_noextramsa (		2.5d		tacc	48		3		bi-mamba	attn	y	y	y	y	first/LN	y	128	**0**		131485	2048	no	7	3	1	6	AW	1e-3			7000	**0.78**	**0.59**																
mamba3_huber		4d		mll	48		3		bi-mamba	attn	y	y	y	y	first/LN	y	128	1024		131485	2048	no	7	3	1	6	AW	1e-3			4740	0.69	0.48																
mamba3_skiptri		-		breg	48		3		bi-mamba	attn	**n**	**n**	**n**	**y**	first/LN	y	128	1024		131485	2048	no	7	3	1	6	AW	1e-3			3000	0.06	0.02																
Ablate SM Input - what in triangle to feed into structure module (inference)																																																	
none					48		3		attn w/pair bias	attn	y	y	y	y	naive	y	128	1024		-	-1	no	-	-	-	-	AW	1e-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	y	y	y	y	naive	y	128	1024		-	-1	no	-	-	-	-	AW	1e-3			90000	0.06	0.04							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	y	y	y	y	naive	y	128	1024		-	-1	no	-	-	-	-	AW	1e-3			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	y	y	y	y	naive	y	128	1024		-	-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	y	y	y	first/LN	y	128	1024		131485	-1	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 (inference)																																																	
no pair bias					48		3		attn w/pair bias	attn	y	y	y	y	naive	y	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	y	y	y	y	naive	y	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	y	y	y	y	first/LN	y	128	1024		131485	-1	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	y	y	y	y	first/LN	y	128	1024		131485	-1	no	7	3	1	6	AW	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 pair					48		3		bi-mamba	attn	y	y	y	y	first/LN	y	128	1024		131485	-1	no	7	3	1	6	AW	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.mean()					48		3		bi-mamba	attn	y	y	y	y	first/LN	y	128	1024		131485	-1	no	7	3	1	6	AW	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 attn weights					48		3		bi-mamba	attn	y	y	y	y	first/LN	y	128	1024		131485	-1	no	7	3	1	6	AW	1e-3			7000	0.64	0.48							0.76	0.69	0.58	0.73	0.64	0.60	0.49	0.72	0.65	0.65
output top 32 attn weights					48		3		bi-mamba	attn	y	y	y	y	first/LN	y	128	1024		131485	-1	no	7	3	1	6	AW	1e-3			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
output local 16					48		3		bi-mamba	attn	y	y	y	y	first/LN	y	128	1024		131485	-1	no	7	3	1	6	AW	1e-3			7000	0.40	0.24							0.56	0.47	0.39	0.51	0.41	0.41	0.37	0.51	0.44	0.51
output not local 16					48		3		bi-mamba	attn	y	y	y	y	first/LN	y	128	1024		131485	-1	no	7	3	1	6	AW	1e-3			7000	0.66	0.51							0.77	0.74	0.61	0.75	0.65	0.63	0.55	0.75	0.67	0.67
output local 32					48		3		bi-mamba	attn	y	y	y	y	first/LN	y	128	1024		131485	-1	no	7	3	1	6	AW	1e-3			7000	0.57	0.34							0.70	0.62	0.54	0.65	0.58	0.57	0.49	0.65	0.60	0.62
output not local 32					48		3		bi-mamba	attn	y	y	y	y	first/LN	y	128	1024		131485	-1	no	7	3	1	6	AW	1e-3			7000	NaN																	
Col + Pair																																																	
mamba4		3d		tacc	48		3		**bi-mamba**		y	y	y	y	first/LN	y	128	1024		131485	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**	y	**nomsa**/1/LN	y	128	0		131485	2048	no	7	3	1	6	AW	1e-3			4300	0.35	0.06	0.42	0.35	0.30	0.16	0.06	0.01	0.57	0.39	0.32	0.47	0.36	0.37	0.34	0.48	0.40	0.47
mamba5				tacc	48		3		bi-mamba	attn	n	n	n	y	^/out2	y	128	0		131485	2048	no	7	3	1	6	AW	1e-3			4300	0.35	0.06																
mamba5				tacc	48		3		bi-mamba	attn	n	n	n	y	^/out2/rope	y	128	0		131485	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	y	**n**	y	y	first/LN	y	128	**0**		131485	2048	no	7	3	1	6	AW	1e-3			7000	0.74	0.54																
mamba3.5 ++		2d		tacc	48		3		bi-mamba	attn	y	y	**n**	y	first/LN	y	128	**0**		131485	2048	no	7	3	1	6	AW	1e-3			7000	0.77	0.59	0.74	0.78	0.77	0.56	0.61	0.53	0.86	0.82	0.70	0.84	0.76	0.73	0.63	0.83	0.77	0.75
Triangle Attn Bias																																																	
no triangle_bias					48		3		attn w/pair bias	attn	y	y	y	y	naive	y	128	1024		-	-1	no	-	-	-	-	AW	1e-3			90000	0.24	0.03	0.23	0.24	0.23	0.08	0.03	0.01	0.44	0.32	0.24	0.38	0.24	0.24	0.26	0.40	0.29	0.38
no triangle_QKT					48		3		attn w/pair bias	attn	y	y	y	y	naive	y	128	1024		-	-1	no	-	-	-	-	AW	1e-3			90000	0.36	0.25	0.49	0.37	0.28	0.37	0.28	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		52h		tacc	48		3		bi-mamba	**bi-mamba**	y	y	**n**	y	first/LN	y	128	**0**		131485	2048	no	7	3	1	6	AW	1e-3			7000	0.77	0.57																
mamba7				tacc	48		3		bi-mamba	**bi-mamba**	y	**bi-mamba**	**n**	y	first/LN	y	128	**0**		131485	2048	no	7	3	1	6	AW	1e-3			5600	0.43																	

Experiments					Evoformer											Data						Optimization								CAMEO																					
	notes	time	code	mach	blk	blk2	re	split	row	col	tri	tri-attn	tri-mul	tri-res	opm	relpos	msa	extra	aa	chains	pre-msa	template	nnode	gpu	bs	accum	opt	blr	warm	ep	steps	LDDT-CA	GDT-TS	L<100	L<400	L<inf	G<100	G<400	G<inf	A	B	C	1	2	3	4	b1	b2	b3
OpenFold																																																			
their weights		-		-	48		3		attn w/pair bias	attn	y	y	y	y	naive	y	128	1024		-	-1	no	-	-	-	-	AW	1e-3			90000	0.89	0.75	0.86	0.89	0.90	0.75	0.75	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	y	y	y	y	naive	y	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	0.86	0.82	0.71	0.84	0.76	0.73	0.63			
Pair to MSA																																																		
mamba6 + p2m_v1				tacc	48		3		bi-mamba	bi-mamba	y	y	**n**	y	first/LN	y	128	**0**		131485	2048	no	7	3	1	6	AW	1e-3			4200	0.17	0.03																	
mamba6 + p2m_v2				tacc	48		3		bi-mamba	bi-mamba	y	y	**n**	y	first/LN	y	128	**0**		131485	2048	no	7	3	1	6	AW	1e-3			2500	0.05	0.02																	
LION																																																		
mamba6				tacc	48		3		bi-mamba	bi-mamba	y	y	n	y	first/LN	y	128	**0**		131485	2048	no	7	3	1	6	**LION**	**3e-4**			3600	0.73	0.53	grad spike																
mamba6				tacc	48		3		bi-mamba	bi-mamba	y	y	n	y	first/LN	y	128	**0**		131485	2048	no	7	3	1	6	**LION**	**1e-4**			6300	0.80	0.63	0.79	0.81	0.79	0.64	0.64	0.56	0.88	0.85	0.74	0.86	0.79	0.76	0.65	0.85	0.80	0.77
mamba6 + beta2=0.98 + clip values				tacc	48		3		bi-mamba	bi-mamba	y	y	n	y	first/LN	y	128	**0**		131485	2048	no	7	3	1	6	**LION**	**3e-4**			3600	0.72	0.49	trains faster then worse																
mamba6 + beta1=0.95 + beta2=0.9				tacc	48		3		bi-mamba	bi-mamba	y	y	n	y	first/LN	y	128	**0**		131485	2048	no	7	3	1	6	**LION**	**3e-4**			3600	0.71																		
longer lengths																																																		
mamba6 + seq	maybe, but no time				48		3		bi-mamba	bi-mamba	y	y	**n**	y	first/LN	y	128	0		131485	2048	no	7	3	1	6	AW	1e-3			3200	0.33	0.09																	
trimamba with pair bias																																																		
mamba6	B x bias			tacc	48		3		bi-mamba	bi-mamba	y	**mamba2**	n	y	first/LN	y	128	0		131485	2048	no	7	3	1	6	LION	1e-4			5500	0.73	0.52																	
mamba6	dt x bias			tacc	48		3		bi-mamba	bi-mamba	y	**mamba3**	n	y	first/LN	y	128	0		131485	2048	no	7	3	1	6	LION	1e-4			4800	0.68	0.46																	
mamba6	dt,B,C x bias			tacc	48		3		bi-mamba	bi-mamba	y	**mamba5**	n	y	first/LN	y	128	0		131485	2048	no	7	3	1	6	LION	1e-4			7000	B - 0.09																		
mamba6, pair_dropout=0				tacc	48		0		bi-mamba	bi-mamba	y	**mamba5**	n	y	first/LN	y	128	**0**		131485	2048	no	7	3	1	6	LION	1e-4			7000	B - 0.08																		
recycle																																																		
mamba6		40h		tacc	48		0		bi-mamba	bi-mamba	y	y	n	y	first/LN	y	128	**0**		131485	2048	no	7	3	1	6	LION	1e-4			7000	0.77	0.56	0.78	0.76	0.72	0.62	0.59	0.40	0.85	0.81	0.69	0.82	0.75	0.72	0.62	0.82	0.76	0.75
mamba6, clip_val		40h		tacc	48		0		bi-mamba	bi-mamba	y	y	n	y	first/LN	y	128	**0**		131485	2048	no	7	3	1	6	LION	1e-4			6300	0.77	0.59	not worth trying, its roughly the same speed															
no side chain loss																																																		
mamba6 + no side chai		40h		tacc	48		0		bi-mamba	bi-mamba	y	y	n	y	first/LN	y	128	**0**		131485	2048	no	7	3	1	6	LION	1e-4			7000	0.76	0.57																	
new impl (mamba6)																																																		
		48hr	new	tacc	48	48	0	n	bi-mamba	bi-mamba	y	y	n	y	first/LN	y	128	0	256	131485	2048	no	7	3	1	6	LION	1e-4			8000	0.77	0.59																
short schedule(1 warm		26hr	new	tacc	48	48	0	n	bi-mamba	bi-mamba	y	y	n	y	first/LN	y	128	0	256	131485	2048	no	7	3	1	6	LION	1e-4			**5000**	0.74	0.54																
bs=64		48hr	new	tacc	48	48	0	n	bi-mamba	bi-mamba	y	y	n	y	first/LN	y	128	0	256	131485	2048	no	7	3	1	**3**	LION	1e-4			16000	0.76	0.58																
splitformer																																																		
same perf, faster		48hr	new	tacc	48	48	0	y	bi-mamba	bi-mamba	y	y	n	y	first/LN	y	128	0	256	131485	2048	no	7	3	1	6	LION	1e-4	5	8	9000	**0.79**	**0.62**															
		42hr	4:50/hr																256												8000	0.78	0.60															
		96hr																	256										5	17	18000	0.85	0.69	0.80	0.85	0.85	0.68	0.70	0.64									
		110hr																	384										5	25		0.86	0.70														
n_msa > n_pair		48hr	new	tacc	**72**	**24**	0	y	bi-mamba	bi-mamba	y	y	n	y	first/LN	y	128	0		131485	2048	no	7	3	1	6	LION	1e-4			9000																	
tri-mamba		48hr	new	tacc	48	48	0	y	bi-mamba	bi-mamba	y	**mamba5**	n	y	first/LN	y	128	0		131485	2048	no	7	3	1	6	LION	1e-4			9000																	
sparse triange attention																																																		
splitformer2 - latent MSA																																																	
decent but why not multi-scale?					48	48	0	y	bi-mamba	bi-mamba	y	y	n	y	first/LN	y	128	0	256	131485	2048	no	7	3	1	6	LION	1e-4		10		0.77	**0.59**															
short schedule v0 - not long enough, bs too big																																																	
need more epochs		30h		glut	48	48	0	y	bi-mamba	bi-mamba	y	y	n	y	first/LN	y	128	0	256	38,000	1024	no	1	4	16	2	LION	1e-4	**1**	**4**		0.06	0.02															
need more epochs, bs=128		24h		mll3	48	48	0	y	bi-mamba	bi-mamba	y	y	n	y	first/LN	y	**32**	0	**128**	80,000	1024	no	1	8	**64**	1	LION	1e-4	**2**	**8**		0.35	0.05															
need more epochs, bs=128		12h		flad	48	48	0	y	bi-mamba	bi-mamba	y	y	n	y	first/LN	y	**32**	0	**128**	80,000	1024	no	1	8	**64**	1	LION	1e-4	**1**	**4**		0.06	0.02															
need more epochs, bs=128		9h		flad	**24**	**24**	0	y	bi-mamba	bi-mamba	y	y	n	y	first/LN	y	**32**	0	**128**	80,000	1024	no	1	8	**64**	1	LION	1e-4	**1**	**4**		0.09	0.03															
need more epochs, bs=128		8hr		flad	**12**	**12**	0	y	bi-mamba	bi-mamba	y	y	n	y	first/LN	y	**32**	0	**128**	80,000	1024	no	1	8	**64**	1	LION	1e-4	**1**	**4**		0.19	0.03															
		32hr																											**2**	**16**		0.29	0.05															
		8hr+		glut	**24**	**24**	0	y	bi-mamba	bi-mamba	y	y	n	y	first/LN	y	**32**	0	**128**	80,000	1024	no	1	4	**128**	1	LION	1e-4	**1**	**8**		0.25	0.04															
need bs=128		16hr+																											**2**	**16**		0.26	0.05															
short schedule v1																																																	
		8hr	mll3		**24**	**24**	0	y	bi-mamba	bi-mamba	y	y	n	y	first/LN	y	**32**	0	**128**	80,000	1024	no	1	8	**1**	16	LION	1e-4	**1**	**4**		0.32	0.06															
		32hr																											**2**	**16**		0.57	0.36															
		25hr	mll7		**24**	**24**	0	y	bi-mamba	bi-mamba	y	y	n	y	first/LN	y	**32**	0	**128**	80,000	1024	no	1	8	**16**	1	LION	1e-4	**2**	**16**		0.60	0.38															
short schedule v2 - lr/wd																																																	
wd=1e-2		25hr		mll3	24	24	0	y	bi-mamba	bi-mamba	y	y	n	y	first/LN	y	32	0	128	80,000	1024	no	1	8	16	1	LION	**2e-4**	2	16		**0.61**	**0.41**															
				flad	24	24	0	y	bi-mamba	bi-mamba	y	y	n	y	first/LN	y	32	0	128	80,000	1024	no	1	8	16	1	LION	**4e-4**	2	16		NaN																
				glut	24	24	0	y	bi-mamba	bi-mamba	y	y	n	y	first/LN	y	32	0	128	80,000	1024	no	1	4	32	1	LION	**8e-4**	2	16		NaN																
wd=1e-1				glut	24	24	0	y	bi-mamba	bi-mamba	y	y	n	y	first/LN	y	32	0	128	80,000	1024	no	1	4	32	1	LION	1e-4	2	16		similar																
wd=1e-3				mll7	24	24	0	y	bi-mamba	bi-mamba	y	y	n	y	first/LN	y	32	0	128	80,000	1024	no	1	8	16	1	LION	1e-4	2	16		similar																
wd=1e-4				flad	24	24	0	y	bi-mamba	bi-mamba	y	y	n	y	first/LN	y	32	0	128	80,000	1024	no	1	8	16	1	LION	1e-4	2	16		similar																
short schedule v3 - longer sequence/deeper MSA helps																																																	
		34hr	glut		24	24	0	y	bi-mamba	bi-mamba	y	y	n	y	first/LN	y	**64**		**128**	80,000	1024	no	1	8	16	1	LION	**2e-4**	2	14		0.64	0.43															
		34hr	flad		24	24	0	y	bi-mamba	bi-mamba	y	y	n	y	first/LN	y	**16**		**128**	80,000	1024	no	1	8	16	1	LION	**2e-4**	2	15		0.61	0.39															
			mll7		24	24	0	y	bi-mamba	bi-mamba	y	y	n	y	first/LN	y	**32**		**64**	80,000	1024	no	1	8	16	1	LION	**2e-4**	2	16		0.44	0.21															
		34hr	mll3		24	24	0	y	bi-mamba	bi-mamba	y	y	n	y	first/LN	y	**32**		**196**	80,000	1024	no	1	8	16	1	LION	**2e-4**	2	15		0.65	0.44															
short schedule v4																																																	
fewer msa layers - (per layer,		32hr		glut	**12**	24	0	y	bi-mamba	bi-mamba	y	y	n	y	first/LN	y	32	0	128	80,000	1024	no	1	8	16	1	LION	**2e-4**	2	16		0.59	0.36															
fewer pair layers		34hr		flad	24	**12**	0	y	bi-mamba	bi-mamba	y	y	n	y	first/LN	y	32	0	128	80,000	1024	no	1	8	16	1	LION	**2e-4**	2	16		0.61	0.39															
no wd on bias/norm		25hr		mll7	24	24	0	y	bi-mamba	bi-mamba	y	y	n	y	first/LN	y	32	0	128	80,000	1024	no	1	8	16	1	LION	**2e-4**	2	16		**0.64**	**0.41**															
cropfirst/last + ^		25hr		mll3	24	24	0	y	bi-mamba	bi-mamba	y	y	n	y	first/LN	y	32	0	128	80,000	1024	no	1	8	16	1	LION	**2e-4**	2	16		0.63	0.41															
fape weighted by b factor + ^		25hr			24	24	0	y	bi-mamba	bi-mamba	y	y	n	y	first/LN	y	32	0	128	80,000	1024	no	1	8	16	1	LION	**2e-4**	2	16		0.61	0.40															
short schedule v5																																																	
mamba init like transformer				flad	24	24	0	y	bi-mamba	bi-mamba	y	y	n	y	first/LN	y	32	0	128	80,000	1024	no	1	8	16	1	LION	**2e-4**	2	16		worse																
no wd on bias/norm/embed/mamb				mll7	24	24	0	y	bi-mamba	bi-mamba	y	y	n	y	first/LN	y	32	0	128	80,000	1024	no	1	8	16	1	LION	**2e-4**	2	16		0.64	0.42															

Experiments					Evoformer											Data						Optimization									CAMEO																							
	notes	time	code	mach	blk	blk2	re	split	row	col	tri	tri-attn	tri-mul	tri-res	opm	relpos	msa	extra	aa	chains	pre-msa	template	nnode	gpu	bs	accum	opt	blr	warm	ep	steps	LDDT-CA	GDT-TS	L<100	L<400	L<inf	G<100	G<400	G<inf	A	B	C	1	2	3	4	b1	b2	b3					
OpenFold																																																						
their weights		-		-	48		3		attn w/pair bias	attn	y	y	y	y	naive	y	128	1024		-	-1	no	-	-	-	-	AW	1e-3			90000	0.89	0.75	0.86	0.89	0.90	0.75	0.75	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	y	y	y	y	naive	y	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	0.86	0.82	0.71	0.84	0.76	0.73	0.63						
MSA diversity - AF																																																					
					48		3		attn w/pair bias	attn	y	y	y	y	naive	y	128	1024		-	-1	no	-	-	-	-	AW	1e-3			90000	0.89	0.75	0.86	0.89	0.90	0.75	0.75	0.71	0.94	0.93	0.85	0.93	0.89	0.86	0.77	0.94	0.89	0.86					
sensitivity=7.5	max-seq-id=0.95				48		3		attn w/pair bias	attn	y	y	y	y	naive	y	128	1024		-	-1	no	-	-	-	-	AW	1e-3			90000	0.88	0.75	0.86	0.89	0.88	0.74	0.76	0.74	0.93	0.91	0.84	0.93	0.88	0.85	0.77	0.93	0.88	0.85					
sensitivity=7.5	max-seq-id=0.8				48		3		attn w/pair bias	attn	y	y	y	y	naive	y	128	1024		-	-1	no	-	-	-	-	AW	1e-3			90000	0.88	0.75	0.86	0.89	0.88	0.74	0.76	0.74	0.93	0.91	0.84	0.93	0.88	0.85	0.77	0.93	0.88	0.85					
sensitivity=7.5	max-seq-id=0.5				48		3		attn w/pair bias	attn	y	y	y	y	naive	y	128	1024		-	-1	no	-	-	-	-	AW	1e-3			90000	0.30	0.09	0.63	0.30	0.17	0.43	0.07	0.01	0.54	0.33	0.31	0.42	0.32	0.36	0.37	0.44	0.36	0.48					
sensitivity=7.5	max-seq-id=0.3				48		3		attn w/pair bias	attn	y	y	y	y	naive	y	128	1024		-	-1	no	-	-	-	-	AW	1e-3			90000	0.30	0.09	0.63	0.30	0.17	0.43	0.07	0.01	0.54	0.33	0.31	0.42	0.32	0.36	0.37	0.44	0.36	0.48					
MSA diversity - Splitformer																																																				
					48	48	0	y	bi-mamba	bi-mamba	y	y	n	y	first/LN	y	128	0	256	131485	2048	no	7	3	1	6	LION	1e-4		35		0.85	0.70																				
sensitivity=7.5	max-seq-id=0.95				48	48	0	y	bi-mamba	bi-mamba	y	y	n	y	first/LN	y	128	0	256	131485	2048	no	7	3	1	6	LION	1e-4		35		0.84	0.69																				
sensitivity=7.5	max-seq-id=0.8																															0.84	0.69																				
sensitivity=7.5	max-seq-id=0.5																															0.84	0.69																				
sensitivity=7.5	max-seq-id=0.3																															0.84	0.69																				
BERT pre-training																																																				
mr=0.15				mll3	48	-	-	-	bi-mamba	bi-mamba	-	-	-	-	-	y	128	0	256	131485	2048	no	1	8	16	1	LION	1e-4	5	25		**?**																							
mr=0.50				glut	48	-	-	-	bi-mamba	bi-mamba	-	-	-	-	-	y	128	0	256	131485	2048	no	1	4	16	1	LION	1e-4	5	25		**?**																							
mr=0.70				flad	48	-	-	-	bi-mamba	bi-mamba	-	-	-	-	-	y	128	0	256	131485	2048	no	1	4	16	1	LION	1e-4	5	25		**?**																							
long schedule																																																				
new weight decay/crop		48hr,4:40/ep		tacc	48	48	0	y	bi-mamba	bi-mamba	y	y	n	y	first/LN	y	128	0	256	131485	2048	no	7	3	1	6	LION	1e-4	5	10/25		**0.80**	**0.63**																				
		96hr																																																		
																			w/distillation dataset																																			
w/bert pre-training		48hr,4:40/ep		tacc	48	48	0	y	bi-mamba	bi-mamba	y	y	n	y	first/LN	y	128	0	256	131485	2048	no	7	3	1	6	LION	1e-4	5	10/25		**0.80**	**0.63**																				
		96hr																																																	
																			w/distillation dataset																																			
																																	?																				