

Protein Language Model-Generated Enzyme Sequences Exhibit high stability in Molecular Dynamics Simulation – Supplementary Information

Enrico Mario Alessandro Fassi¹, Marco Nicolini², Emanuele Saitto², Giorgio Valentini^{2,3}, Elena Casiraghi^{2,3}, Giovanni Grazioso^{1*}

¹ Dipartimento di Scienze Farmaceutiche, Università degli Studi di Milano, Italy, {enrico.fassi, giovanni.grazioso}@unimi.it

² AnacletoLab - Dipartimento di Informatica, Università degli Studi di Milano, Italy, {marco.nicolini, emanuele.saitto, elena.casiraghi, giorgio.valentini}@unimi.it

³ ELLIS – European Laboratory for Learning and Intelligent Systems, Milan Unit, Italy

*corresponding author

Keywords: Protein Language Models, Molecular Dynamics, Transformers.

Table S1. Average RMSD, RMSF, RoG, and SASA values for the generated EC 2.1.1.37 enzymes at p = 0.5 (top) and p = 0.75 (bottom), compared to the reference structure 1DCT. Values are reported in Å, while SASA in Å².

Structure	RMSD	RMSF	RoG	SASA	Structure	RMSD	RMSF	RoG	SASA
1DCT	2.55	0.79	20.1	13527	472	1.70	1.07	20.1	13334
17	1.63	0.93	20.3	13747	613	1.65	0.83	20.2	13222
64	1.87	0.88	20.2	13515	731	2.00	0.88	20.3	13550
82	2.28	1.74	20.3	13787	870	1.83	1.60	20.4	14034
152	2.03	1.57	20.3	13734	900	1.90	0.89	20.3	13352
250	1.82	0.95	20.3	13942	996	1.93	1.58	20.2	13774
1DCT	2.55	0.79	20.1	13527	600	2.23	1.32	20.3	14573
33	2.25	0.97	20.4	13930	602	1.87	0.96	20.7	14824
124	1.80	0.82	20.5	14186	605	2.67	1.24	20.9	15258
321	2.69	0.92	20.0	13156	807	1.87	1.33	20.3	14516
381	2.26	1.44	20.2	14334	841	2.03	0.92	20.6	15114
442	2.04	1.14	20.5	14153	966	3.87	1.59	22.2	16001
531	2.86	0.98	20.6	14648					

Table S2. Average RMSD, RMSF, RoG, and SASA values for the generated EC 5.4.99.5 enzymes at p = 0.5 (top) and p = 0.75 (bottom), compared to the reference structure 3RMI. Values are reported in Å, while SASA in Å².

Structure	RMSD	RMSF	RoG	SASA	Structure	RMSD	RMSF	RoG	SAS
3RMI	6.06	2.85	19.1	8323	341	6.44	2.75	18.8	8321
22	5.08	2.37	16.8	7322	411	8.22	3.45	17.2	7593
45	11.36	6.36	19.3	8253	488	7.33	3.94	18.1	7931
106	6.64	2.16	16.8	7286	532	12.42	4.18	15.4	7584
136	9.01	2.47	21.1	8784	563	7.06	4.19	19.9	8636
215	5.74	2.93	18.1	7791	661	11.17	4.41	16.1	8019
221	5.76	3.01	17.8	8080	715	5.48	2.43	19.6	8077
227	4.52	1.55	15.9	7531	833	7.96	4.74	20.3	8624
262	7.10	3.56	18.6	8048					
3RMI	6.06	2.85	19.1	8323	642	6.14	2.75	14.8	7351
12	10.8	3.47	16.5	8269	659	7.10	2.26	18.1	7833
92	6.73	2.51	15.0	7336	680	4.29	1.70	16.1	7477
137	7.19	2.84	15.4	7605	690	5.17	2.66	15.6	6845
178	8.14	3.37	15.0	7392	736	6.96	3.16	16.1	6996
210	5.82	1.93	16.2	7433	774	6.37	2.04	17.2	7890
337	7.64	3.06	17.7	8845	801	6.35	4.70	18.7	8148
453	8.28	4.87	17.6	8243	823	7.88	2.08	16.2	7067
513	5.44	3.41	16.6	6996	878	7.90	2.89	17.2	8039
519	6.17	2.57	15.8	6838	892	6.01	2.74	21.4	8965
577	5.80	2.01	16.1	7265	893	5.04	2.47	14.4	6753

Table S3. Average RMSD, RMSF, RoG, and SASA values for the generated EC 5.4.99.5 enzymes at p = 0.5 (top) and p = 0.75 (bottom), compared to the reference structure 1ODE. Values are reported in Å, while SASA in Å².

Structure	RMSD	RMSF	RoG	SASA	Structure	RMSD	RMSF	RoG	SASA
1ODE	1.06	0.73	13.7	7018	805	1.59	1.11	13.8	7010
64	1.93	1.26	13.9	7039	877	1.46	0.95	13.8	6803
546	1.60	1.08	14.1	7071	885	1.61	0.99	14.0	6847
790	1.62	1.05	14.0	7029					
1ODE	1.06	0.73	13.7	7018	448	1.82	1.13	13.8	6827
8	1.79	0.99	13.7	6929	483	2.00	1.12	13.9	6689
17	2.09	1.87	13.8	6866	492	1.40	1.13	13.8	6831
73	1.49	1.27	13.8	6862	551	1.54	1.20	13.9	6954
101	1.22	1.06	13.7	6674	557	3.26	1.26	13.9	7356
146	1.57	0.89	13.7	6873	589	1.24	1.03	13.7	6753
158	1.14	0.85	13.8	7054	591	1.99	1.20	13.7	6781
201	1.84	1.09	13.7	7020	596	2.61	1.02	13.5	6925
220	1.61	0.88	13.7	6823	626	2.31	1.13	13.6	6886
234	1.51	1.23	13.8	6914	631	1.51	0.90	13.7	6883
258	1.85	1.08	14.0	6901	685	1.72	0.93	13.7	6667
261	3.02	1.24	13.7	7048	742	1.89	0.92	13.6	6592
285	1.57	1.47	13.8	6552	747	2.22	0.89	13.7	6753
320	1.26	0.92	13.8	6851	779	2.11	1.25	13.9	7022
352	1.92	1.55	14.5	6630	793	2.49	0.87	13.5	6986
365	1.59	0.87	14.0	6545					

Conflict of interests

No competing interest is declared.

Funding

This work was supported by National Center for Gene Therapy and Drugs Based on RNA Technology—MUR (Project no. CN_00000041) funded by NextGeneration EU program, and by FAIR (Future Artificial Intelligence Research) project, funded by the NextGenerationEU program within the PNRR-PE-AI scheme (M4C2, Investment 1.3, Line on Artificial Intelligence) - AIDH – FAIR - PE0000013.