SUPPLEMENTARY MATERIAL

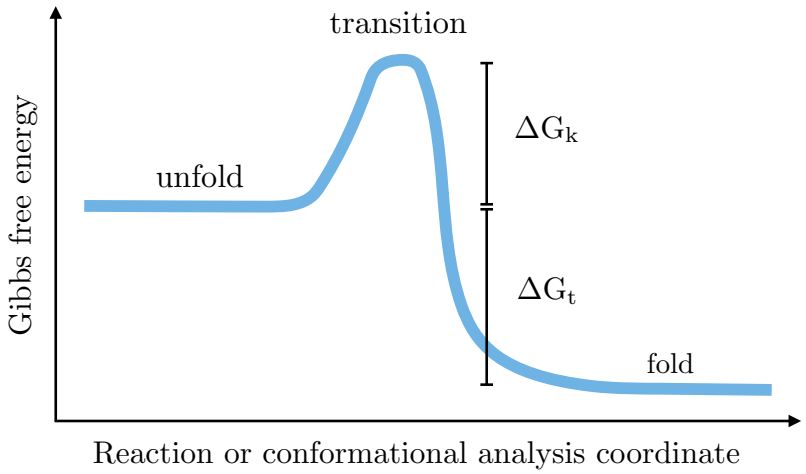

Fig. S1. The stability of a protein is determined by the thermodynamic and kinetic stabilities, ΔG_t and ΔG_k, respectively. We only consider the thermodynamic stability.

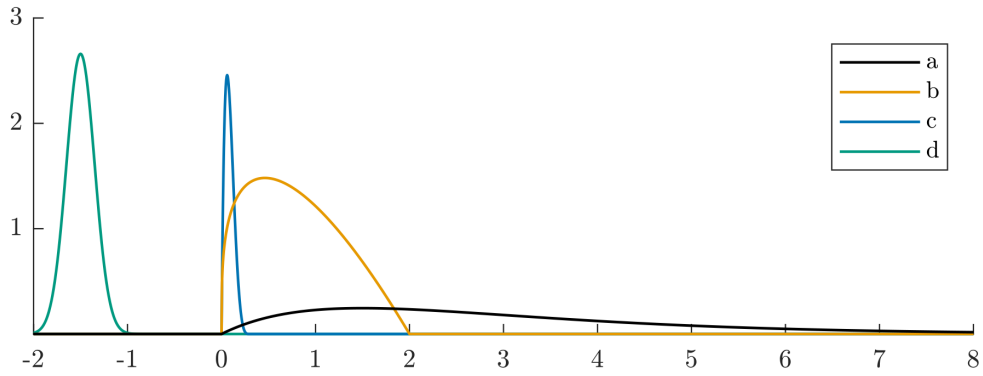

Fig. S2. Priors presented by Equation 2. Here $\mu_d = -1.5$, the most likely value for $-a$. Other hyperparameter values are presented in Table S1.

Table S1. Values for the hyperparameters used in the priors of a, b, c, d, σ_E and σ_S presented in Equations 2 and 6, respectively.

a	b	c	d	σ_E	σ_S
$\alpha_a = 2$	$\alpha_b = 1.3$	$\alpha_c = 2$	$\mu_d = -a$	$\alpha_E = 2.5$	$\alpha_S = 50$
$\beta_a = 1.5$	$\beta_b = 2$	$\beta_c = 5$	$\sigma_d = 0.15$	$\beta_E = 0.02$	$\beta_S = 0.007$

The partial derivatives of the marginal likelihood with respect to the parameters $\boldsymbol{\phi}$ are obtained from Equation (9) as follows:

$$\begin{aligned}\frac{\partial}{\partial \phi_j}\log p(\mathbf{y}|X,\boldsymbol{\phi}) =& \frac{1}{2}\mathbf{y}^T K_\phi^{-1}\frac{\partial K_\phi}{\partial \phi_j}K_\phi^{-1}\mathbf{y} - \frac{1}{2}\mathrm{tr}\left(K_\phi^{-1}\frac{\partial K_\phi}{\partial \phi_j}\right)\\ =& \frac{1}{2}\mathrm{tr}\left(\left(\boldsymbol{\alpha}\boldsymbol{\alpha}^T - K_\phi^{-1}\right)\frac{\partial K_\phi}{\partial \theta_j}\right),\end{aligned} \tag{S1}$$

where $\boldsymbol{\alpha} = K_\phi^{-1}\mathbf{y}$, K_ϕ is determined as

$$K_\phi = \sum_{m=1}^{M} w_m K_m^{\gamma_m} + \mathrm{diag}\begin{pmatrix}\sigma_0\\ \sigma_E\mathbf{1}_{N_E}\\ \sigma_E\mathbf{1}_{N_E} + \sigma_S\mathbf{1}_{N_S} + t\boldsymbol{\sigma}_T\end{pmatrix}^2$$

and the partial derivatives of K_ϕ with respect to the optimised parameters are

$$\frac{\partial K_\phi}{\partial \sigma_E} = \mathrm{diag}\begin{pmatrix}0\\ 2\sigma_E\mathbf{1}_{N_E}\\ 2\left(\sigma_E\mathbf{1}_{N_E} + \sigma_S\mathbf{1}_{N_S} + t\boldsymbol{\sigma}_T\right)\end{pmatrix} \tag{S2}$$

$$\frac{\partial K_\phi}{\partial \sigma_R} = \mathrm{diag}\begin{pmatrix}0\\ \mathbf{0}_{N_E}\\ 2\left(\sigma_E\mathbf{1}_{N_E} + \sigma_S\mathbf{1}_{N_S} + t\boldsymbol{\sigma}_T\right)\end{pmatrix} \tag{S3}$$

$$\frac{\partial K_\phi}{\partial t} = \mathrm{diag}\begin{pmatrix}0\\ \mathbf{0}_{N_E}\\ 2\left(\sigma_E\mathbf{1}_{N_E} + \sigma_S\mathbf{1}_{N_S} + t\right)\boldsymbol{\sigma}_T\end{pmatrix} \tag{S4}$$

$$\frac{\partial K_\phi}{\partial w_m} = K_m^{\gamma_m} \tag{S5}$$

$$\frac{\partial K_\phi}{\partial \gamma_m} = w_m K_m^{\gamma_m}\log K_m \tag{S6}$$

Correlation ρ and root-mean-square error rmse for the predictions are determined as

$$\rho = \frac{\sum_{i=1}^{N_*}(y_i - \bar{\mathbf{y}})(\mu(\mathbf{x}_i) - \bar{\boldsymbol{\mu}})}{\sqrt{\sum_{i=1}^{N_*}(y_i - \bar{\mathbf{y}})^2\sum_{i=1}^{N_*}(\mu(\mathbf{x}_i) - \bar{\boldsymbol{\mu}})^2}} \tag{S7}$$

$$\mathrm{rmse} = \sqrt{\frac{1}{N_*}\sum_{i=1}^{N_*}(y_i - \mu(\mathbf{x}_i))^2}, \tag{S8}$$

where $\bar{\mathbf{y}}$ is the mean of the experimentally measured values, $\mu(\mathbf{x}_i)$ is prediction mean, $\bar{\boldsymbol{\mu}}$ is the average of all prediction means, and N_* is the number of predictions.

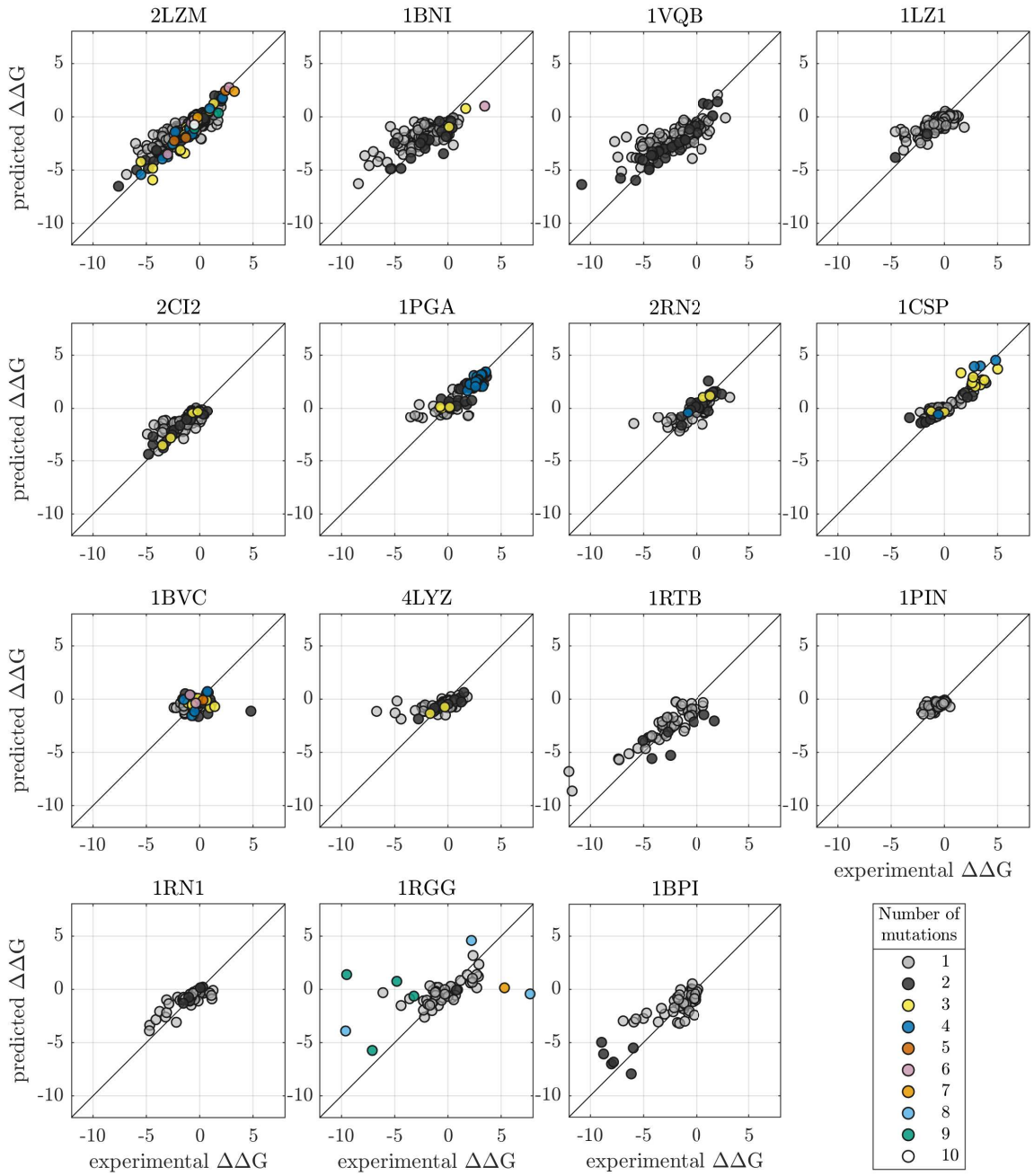

Fig. S3. Mutation-level predictions for all 15 proteins presented in Table 1. The predictions are coloured by the number of simultaneous mutations.

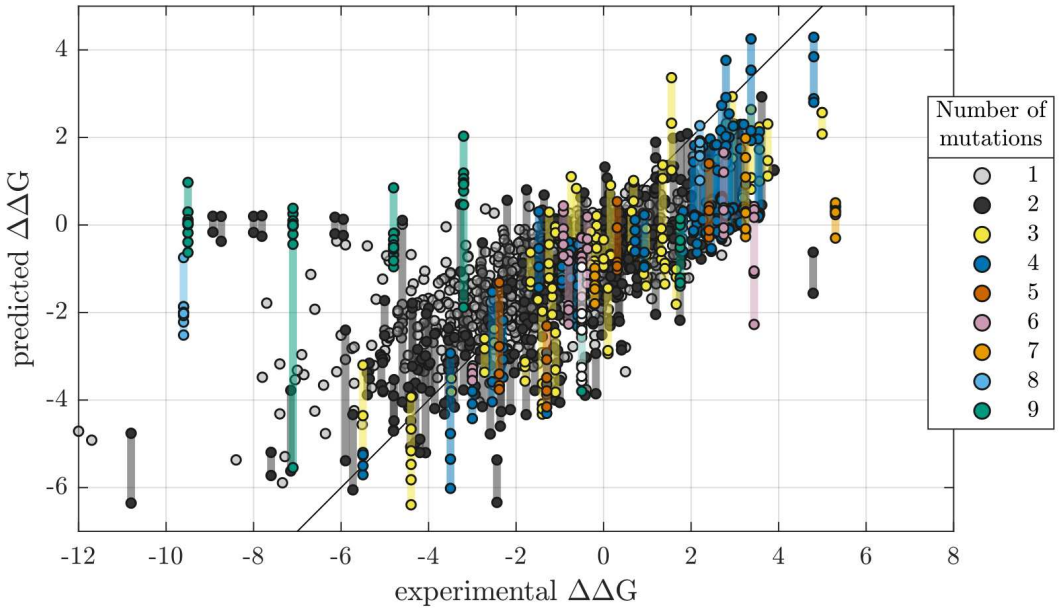

Fig. S4. Position-level predictions for all 15 proteins. When the effects of a mutant are predicted multiple times, they are connected by a line.

Table S2. (Continues on the next page)Comparison of different methods on the 15 protein dataset with respect to ρ and rmse. Off-the-shelf implementations of Rosetta, mCSM and PoPMuSiC are used directly without cross-validation.

		Correlation ρ									rmse								
		Point mutations			Multiple mutations			All mutations			Point mutations			Multiple mutations			All mutations		
		cross-validation level			cross-validation level			cross-validation level			cross-validation level			cross-validation level			cross-validation level		
Protein	Method	mut.	pos.	prot.	mut.	pos.	prot.	mut.	pos.	prot.	mut.	pos.	prot.	mut.	pos.	prot.	mut.	pos.	prot.
2LZM	mGPfusion	0.87	0.80	0.75	0.96	0.83	0.64	0.90	0.76	0.68	0.82	1.02	1.12	0.57	1.53	2.14	0.76	1.30	1.43
	mGPfusion, only B62	0.86	0.77	0.75	0.96	0.87	0.64	0.90	0.82	0.69	0.84	1.08	1.12	0.59	1.16	2.11	0.79	1.13	1.42
	mGP	0.86	0.59	-	0.97	0.85	-	0.90	0.72	-	0.82	1.34	-	0.48	1.24	-	0.75	1.29	-
	mGP, only B62	0.75	0.37	-	0.94	0.77	-	0.82	0.61	-	1.12	1.93	-	0.78	1.44	-	1.05	1.70	-
	Rosetta scaled	0.74	0.73	-	0.68	0.66	-	0.70	0.65	-	1.05	1.06	-	1.65	1.84	-	1.23	1.51	-
	Rosetta		0.75			0.68			0.71			1.13			1.93			1.37	
	mCSM		0.57			-			-			1.27			-			-	
	PoPMuSiC		0.71			-			-			1.11			-			-	
1BNI	mGPfusion	0.77	0.64	0.62	0.86	0.70	0.39	0.77	0.55	0.57	1.21	1.37	1.69	1.28	2.49	2.17	1.22	1.67	1.75
	mGPfusion, only B62	0.74	0.61	0.62	0.84	0.79	0.40	0.74	0.60	0.57	1.27	1.45	1.69	1.24	2.03	2.14	1.27	1.60	1.75
	mGP	0.81	0.65	-	0.86	0.82	-	0.80	0.63	-	1.08	1.40	-	1.32	2.00	-	1.11	1.55	-
	mGP, only B62	0.61	0.48	-	0.87	0.85	-	0.63	0.41	-	1.61	2.32	-	1.15	1.67	-	1.57	2.20	-
	Rosetta scaled	0.59	0.58	-	0.17	0.29	-	0.53	0.45	-	1.58	1.58	-	2.51	2.90	-	1.70	1.94	-
	Rosetta		0.62			0.18			0.56			1.70			2.33			1.77	
	mCSM		0.60			-			-			1.62			-			-	
	PoPMuSiC		0.66			-			-			1.53			-			-	
1VQB	mGPfusion	0.67	0.50	0.49	0.93	0.83	0.75	0.76	0.69	0.60	1.71	1.94	2.25	1.15	1.62	2.06	1.59	1.82	2.20
	mGPfusion, only B62	0.65	0.53	0.50	0.91	0.82	0.73	0.75	0.69	0.58	1.75	1.94	2.25	1.35	1.82	2.43	1.66	1.89	2.30
	mGP	0.79	0.12	-	0.96	0.70	-	0.85	0.50	-	1.41	2.41	-	0.70	1.97	-	1.27	2.24	-
	mGP, only B62	0.79	0.29	-	0.97	0.75	-	0.85	0.55	-	1.50	2.89	-	0.63	2.30	-	1.33	2.66	-
	Rosetta scaled	0.47	0.46	-	0.71	0.68	-	0.57	0.59	-	1.99	2.00	-	1.90	1.96	-	1.97	1.99	-
	Rosetta		0.49			0.73			0.59			2.26			2.06			2.21	
	mCSM		0.53			-			-			2.24			-			-	
	PoPMuSiC		0.51			-			-			2.29			-			-	
1LZ1	mGPfusion	0.75	0.59	0.58	1.00	0.11	-1.00	0.77	0.57	0.52	0.83	0.99	1.06	0.56	2.40	3.75	0.83	1.07	1.16
	mGPfusion, only B62	0.73	0.56	0.59	1.00	0.05	-1.00	0.76	0.56	0.53	0.86	1.03	1.06	1.10	2.41	3.51	0.87	1.10	1.14
	mGP	0.75	0.39	-	1.00	0.56	-	0.78	0.47	-	0.81	1.15	-	0.13	1.65	-	0.80	1.17	-
	mGP, only B62	0.71	-0.31	-	1.00	0.42	-	0.74	0.21	-	0.91	1.43	-	0.27	2.36	-	0.90	1.47	-
	Rosetta scaled	0.57	0.55	-	-1.00	-1.00	-	0.53	0.46	-	0.99	1.01	-	3.23	3.25	-	1.07	1.15	-
	Rosetta		0.59			-1.00			0.55			1.04			3.41			1.12	
	mCSM		0.67			-			-			0.97			-			-	
	PoPMuSiC		0.64			-			-			0.95			-			-	
2CI2	mGPfusion	0.73	0.72	0.64	0.95	0.87	0.85	0.82	0.81	0.71	0.85	0.90	1.07	0.55	0.80	1.21	0.80	0.86	1.10
	mGPfusion, only B62	0.69	0.67	0.63	0.92	0.86	0.86	0.79	0.79	0.72	0.91	0.97	1.07	0.71	1.01	1.12	0.87	0.99	1.08
	mGP	0.65	0.61	-	0.92	0.79	-	0.76	0.72	-	0.95	1.02	-	0.66	1.01	-	0.90	1.02	-
	mGP, only B62	0.51	0.74	-	0.92	0.71	-	0.68	0.63	-	1.16	1.39	-	0.71	1.40	-	1.08	1.39	-
	Rosetta scaled	0.60	0.60	-	0.61	0.61	-	0.63	0.63	-	1.00	1.00	-	1.27	1.27	-	1.06	1.11	-
	Rosetta		0.63			0.62			0.65			1.09			1.30			1.13	
	mCSM		0.74			-			-			0.86			-			-	
	PoPMuSiC		0.75			-			-			0.85			-			-	
1PGA	mGPfusion	0.68	0.47	0.69	0.90	0.35	0.32	0.85	0.43	0.50	1.26	1.54	1.64	0.53	2.09	2.74	0.88	2.00	2.38
	mGPfusion, only B62	0.82	0.59	0.71	0.76	0.60	0.62	0.81	0.53	0.70	0.87	1.22	1.06	0.69	0.95	0.88	0.83	1.13	1.02
	mGP	0.62	0.61	-	0.93	-0.24	-	0.84	-0.14	-	1.40	1.58	-	0.45	3.01	-	0.94	2.81	-
	mGP, only B62	0.57	-0.46	-	0.92	-0.08	-	0.81	0.06	-	1.53	1.73	-	0.47	2.07	-	1.02	2.01	-
	Rosetta scaled	0.69	0.59	-	0.09	0.07	-	0.24	0.11	-	1.21	1.42	-	2.81	3.09	-	2.33	2.87	-
	Rosetta		0.69			0.03			0.28			1.70			3.51			2.95	
	mCSM		-0.10			-			-			1.94			-			-	
	PoPMuSiC		0.28			-			-			1.89			-			-	
2RN2	mGPfusion	0.79	0.58	0.71	0.78	0.60	0.61	0.79	0.53	0.70	0.91	1.21	1.05	0.67	1.01	0.91	0.86	1.14	1.02
	mGPfusion, only B62	0.82	0.59	0.71	0.76	0.60	0.62	0.81	0.53	0.70	0.87	1.22	1.59	0.69	0.95	1.18	0.83	1.13	1.51
	mGP	0.77	0.12	-	0.75	0.42	-	0.77	0.22	-	0.93	1.45	-	0.74	1.21	-	0.89	1.36	-
	mGP, only B62	0.83	0.09	-	0.77	0.42	-	0.82	0.23	-	0.82	1.45	-	0.68	1.20	-	0.80	1.36	-
	Rosetta scaled	0.66	0.64	-	0.48	0.50	-	0.62	0.57	-	1.09	1.13	-	1.20	1.08	-	1.12	1.11	-
	Rosetta		0.70			0.47			0.65			1.07			1.25			1.11	
	mCSM		0.71			-			-			1.04			-			-	
	PoPMuSiC		0.71			-			-			1.16			-			-	
1CSP	mGPfusion	0.85	0.23	0.33	0.92	0.73	0.48	0.92	0.75	0.38	0.64	1.04	1.10	0.91	1.66	2.58	0.75	1.45	1.80
	mGPfusion, only B62	0.86	0.22	0.34	0.91	0.69	0.59	0.91	0.72	0.54	0.65	1.04	1.10	0.96	1.87	2.13	0.78	1.60	1.57
	mGP	0.88	-0.06	-	0.94	0.75	-	0.94	0.77	-	0.54	1.07	-	0.76	1.51	-	0.63	1.36	-
	mGP, only B62	0.87	-0.37	-	0.92	0.71	-	0.92	0.72	-	0.60	1.12	-	0.86	1.59	-	0.71	1.43	-
	Rosetta scaled	0.23	0.20	-	0.68	0.69	-	0.59	0.64	-	1.04	1.06	-	2.19	2.29	-	1.58	1.92	-
	Rosetta		0.33			0.68			0.60			1.11			1.92			1.47	
	mCSM		0.42			-			-			1.02			-			-	
	PoPMuSiC		0.48			-			-			0.99			-			-	

Table S2. (Continued) Comparison of different methods on the 15 protein dataset with respect to ρ and rmse. Off-the-shelf implementations of Rosetta, mCSM and PoPMuSiC are used directly without cross-validation.

		Correlation ρ									rmse								
		Point mutations			Multiple mutations			All mutations			Point mutations			Multiple mutations			All mutations		
		cross-validation level			cross-validation level			cross-validation level			cross-validation level			cross-validation level			cross-validation level		
Protein	Method	mut.	pos.	prot.	mut.	pos.	prot.	mut.	pos.	prot.	mut.	pos.	prot.	mut.	pos.	prot.	mut.	pos.	prot.
1BVC	mGPfusion	0.41	0.43	0.48	-0.28	-0.25	-0.63	0.08	-0.09	0.09	0.74	0.72	1.65	1.64	1.48	2.70	1.09	1.21	2.02
	mGPfusion, only B62	0.50	0.48	0.48	-0.25	-0.25	-0.66	0.14	-0.07	0.14	0.70	0.71	1.65	1.62	1.47	2.43	1.06	1.20	1.92
	mGP	-0.05	-0.12	-	-0.10	-0.21	-	-0.13	-0.23	-	0.99	1.00	-	1.30	1.35	-	1.09	1.21	-
	mGP, only B62	-0.05	0.06	-	-0.11	-0.21	-	-0.14	-0.20	-	0.99	0.99	-	1.29	1.33	-	1.09	1.20	-
	Rosetta scaled	0.42	0.40	-	-0.63	-0.57	-	0.09	-0.09	-	0.75	0.76	-	1.65	1.45	-	1.10	1.20	-
	Rosetta		0.47			-0.65			0.14			1.67			2.45			1.94	
	mCSM		0.47			-			-			1.00			-			-	
	PoPMuSiC		0.60			-			-			0.85			-			-	
4LYZ	mGPfusion	0.61	0.27	0.34	0.95	0.57	0.66	0.65	0.34	0.35	1.47	1.65	3.46	0.69	1.02	2.35	1.35	1.45	3.26
	mGPfusion, only B62	0.59	0.30	0.34	0.96	0.64	0.64	0.64	0.40	0.35	1.48	1.64	3.50	0.57	0.93	2.17	1.35	1.43	3.27
	mGP	0.65	-0.05	-	0.96	0.46	-	0.68	0.18	-	1.38	1.78	-	0.36	1.15	-	1.24	1.59	-
	mGP, only B62	0.62	0.27	-	0.97	0.47	-	0.64	0.21	-	1.48	1.84	-	0.31	1.16	-	1.33	1.63	-
	Rosetta scaled	0.29	0.28	-	0.70	0.68	-	0.33	0.34	-	2.26	2.22	-	1.22	1.29	-	2.09	1.94	-
	Rosetta		0.33			0.71			0.35			3.61			2.05			3.35	
	mCSM		0.55			-			-			1.43			-			-	
	PoPMuSiC		0.59			-			-			1.45			-			-	
1RTB	mGPfusion	0.92	0.81	0.70	0.78	0.74	0.75	0.86	0.71	0.68	1.25	1.73	2.44	2.22	2.24	1.71	1.40	1.85	2.36
	mGPfusion, only B62	0.89	0.79	0.70	0.73	0.73	0.67	0.86	0.75	0.69	1.48	1.85	2.44	1.82	1.81	1.90	1.53	1.84	2.38
	mGP	0.92	0.69	-	0.78	0.30	-	0.76	0.18	-	1.26	2.62	-	3.44	3.39	-	1.69	2.81	-
	mGP, only B62	0.91	0.58	-	0.48	0.58	-	0.85	0.16	-	1.67	3.37	-	2.12	2.11	-	1.73	3.14	-
	Rosetta scaled	0.65	0.61	-	0.73	0.73	-	0.67	0.65	-	1.99	2.08	-	1.63	1.62	-	1.95	1.99	-
	Rosetta		0.69			0.73			0.70			2.45			2.01			2.40	
	mCSM		0.68			-			-			2.33			-			-	
	PoPMuSiC		0.72			-			-			2.22			-			-	
1PIN*	mGPfusion	0.64	0.49	0.53	-	-	-	0.64	0.49	0.53	0.47	0.53	1.00	-	-	-	0.47	0.53	1.00
	mGPfusion, only B62	0.59	0.51	0.53	-	-	-	0.59	0.51	0.53	0.50	0.54	1.00	-	-	-	0.50	0.54	1.00
	mGP	0.61	0.40	-	-	-	-	0.61	0.40	-	0.49	0.58	-	-	-	-	0.49	0.58	-
	mGP, only B62	0.23	0.25	-	-	-	-	0.23	0.25	-	0.70	0.76	-	-	-	-	0.70	0.76	-
	Rosetta scaled	0.50	0.49	-	-	-	-	0.50	0.49	-	0.54	0.54	-	-	-	-	0.54	0.54	-
	Rosetta		0.53			-			0.53			0.99						0.99	
	mCSM		0.72			-			0.72			0.59			-			0.59	
	PoPMuSiC		0.60			-			0.60			0.70			-			0.70	
1RN1	mGPfusion	0.83	0.62	0.67	0.97	0.38	-0.19	0.84	0.55	0.65	0.85	1.17	1.18	0.19	0.91	1.28	0.81	1.13	1.19
	mGPfusion, only B62	0.81	0.58	0.67	0.96	0.42	-0.01	0.81	0.51	0.66	0.92	1.24	1.18	0.35	0.77	0.99	0.88	1.17	1.17
	mGP	0.74	0.04	-	0.99	0.62	-	0.75	0.04	-	1.06	1.75	-	0.12	0.58	-	1.01	1.61	-
	mGP, only B62	0.77	0.27	-	0.99	0.50	-	0.78	0.10	-	1.00	1.78	-	0.13	0.82	-	0.95	1.66	-
	Rosetta scaled	0.63	0.57	-	0.21	0.18	-	0.61	0.54	-	1.12	1.18	-	0.98	1.01	-	1.10	1.15	-
	Rosetta		0.67			0.20			0.65			1.20			1.24			1.20	
	mCSM		0.76			-			-			0.97			-			-	
	PoPMuSiC		0.62			-			-			1.14			-			-	
1RGG	mGPfusion	0.69	0.51	0.61	0.42	0.66	0.96	0.52	0.52	0.73	1.41	1.70	1.56	5.67	5.95	4.20	2.65	4.75	2.23
	mGPfusion, only B62	0.69	0.52	0.61	0.22	0.30	0.93	0.40	0.33	0.55	1.42	1.68	1.56	6.21	6.14	5.43	2.85	4.90	2.64
	mGP	0.71	-0.61	-	0.19	0.08	-	0.35	-0.05	-	1.38	2.24	-	6.81	6.88	-	3.05	5.53	-
	mGP, only B62	0.72	-0.53	-	0.08	0.07	-	0.30	-0.00	-	1.53	2.01	-	6.89	6.77	-	3.14	5.41	-
	Rosetta scaled	0.58	0.43	-	0.78	0.86	-	0.42	0.51	-	1.57	1.76	-	7.08	7.24	-	3.23	5.74	-
	Rosetta		0.60			0.77			0.39			1.58			6.90			3.16	
	mCSM		0.77			-			-			1.36			-			-	
	PoPMuSiC		0.65			-			-			1.54			-			-	
1BPI	mGPfusion	0.68	0.67	0.51	-0.39	-0.02	-0.06	0.85	-0.13	0.12	1.27	1.28	1.83	2.17	7.66	7.69	1.40	3.64	3.11
	mGPfusion, only B62	0.64	0.65	0.51	-0.53	0.33	0.21	0.81	0.05	0.12	1.32	1.32	1.82	2.86	7.07	7.70	1.57	3.40	3.11
	mGP	0.56	0.57	-	-0.29	0.52	-	0.83	0.84	-	1.43	1.43	-	1.72	4.19	-	1.47	2.28	-
	mGP, only B62	0.43	0.58	-	-0.55	0.46	-	0.76	0.70	-	1.68	1.73	-	2.52	6.58	-	1.80	3.34	-
	Rosetta scaled	0.52	0.52	-	0.23	-0.00	-	0.08	-0.08	-	1.52	1.52	-	7.43	7.43	-	2.88	3.62	-
	Rosetta		0.51			-0.00			0.13			1.80			7.68			3.09	
	mCSM		0.71			-			-			1.26			-			-	
	PoPMuSiC		0.72			-			-			1.31			-			-	
total	mGPfusion	0.81	0.70	0.56	0.88	0.61	0.49	0.83	0.64	0.52	1.07	1.26	1.61	1.33	2.45	2.53	1.13	1.87	1.84
	mGPfusion, only B62	0.79	0.69	0.56	0.86	0.64	0.50	0.82	0.66	0.52	1.11	1.30	1.62	1.43	2.40	2.50	1.18	1.85	1.84
	mGP	0.81	0.51	-	0.86	0.52	-	0.83	0.50	-	1.04	1.54	-	1.44	2.65	-	1.14	2.09	-
	mGP, only B62	0.76	0.34	-	0.86	0.55	-	0.80	0.49	-	1.26	1.95	-	1.45	2.56	-	1.30	2.23	-
	Rosetta scaled	0.65	0.63	-	0.51	0.39	-	0.60	0.48	-	1.35	1.38	-	2.49	2.99	-	1.66	2.22	-
	Rosetta		0.55			0.40			0.49			1.63			2.74			1.92	
	mCSM		0.61			-			-			1.40			-			-	
	PoPMuSiC		0.64			-			-			1.37			-			-	

* dataset for 1PIN contained no multiple mutations.

Table S3. Comparison of different methods on the 15 protein dataset with respect to ρ and rmse after removing 10% of predictions with largest errors. Mutation, position, and protein are referred to as mut., pos., and prot., respectively. Off-the-shelf implementations of Rosetta, mCSM and PoPMuSiC are used directly without cross-validation.

Results after 10 % outlier removal	Correlation ρ									rmse								
	Point mutations			Multiple mutations			All mutations			Point mutations			Multiple mutations			All mutations		
	cross-validation level			cross-validation level			cross-validation level			cross-validation level			cross-validation level			cross-validation level		
Method	mut.	pos.	prot.	mut.	pos.	prot.	mut.	pos.	prot.	mut.	pos.	prot.	mut.	pos.	prot.	mut.	pos.	prot.
mGPfusion	0.87	0.77	0.75	0.97	0.84	0.63	0.92	0.81	0.71	0.69	0.85	0.93	0.62	1.55	1.90	0.67	1.15	1.13
mGPfusion, only B62	0.86	0.73	0.75	0.96	0.84	0.65	0.91	0.80	0.72	0.69	0.85	0.93	0.64	1.43	1.85	0.67	1.10	1.12
mGP	0.89	0.57	-	0.98	0.76	-	0.93	0.74	-	0.65	1.01	-	0.49	1.66	-	0.62	1.23	-
mGP, only B62	0.86	0.27	-	0.97	0.76	-	0.91	0.66	-	0.73	1.30	-	0.54	1.48	-	0.69	1.37	-
Rosetta scaled	0.80	0.78	-	0.73	0.71	-	0.78	0.75	-	0.84	0.87	-	1.67	1.99	-	1.00	1.31	-
Off-the-shelf implementations with no cross-validation																		
Rosetta	0.75			0.67			0.73			0.94			1.85			1.11		
mCSM	0.71			-			-			0.89			-			-		
PoPMuSiC	0.73			-			-			0.86			-			-		

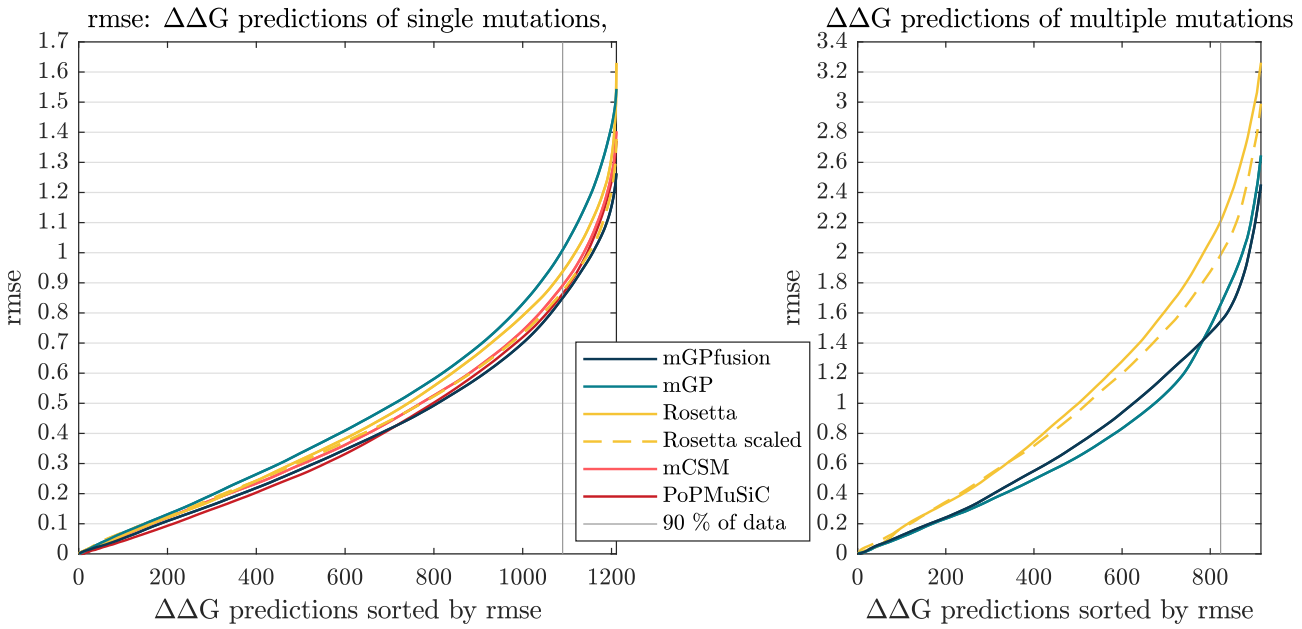

Fig. S5. rmse with different amount of predictions, when predictions are sorted by the error. Position level cross-validation was used for mGPfusion, mGP and Rosetta scaled.

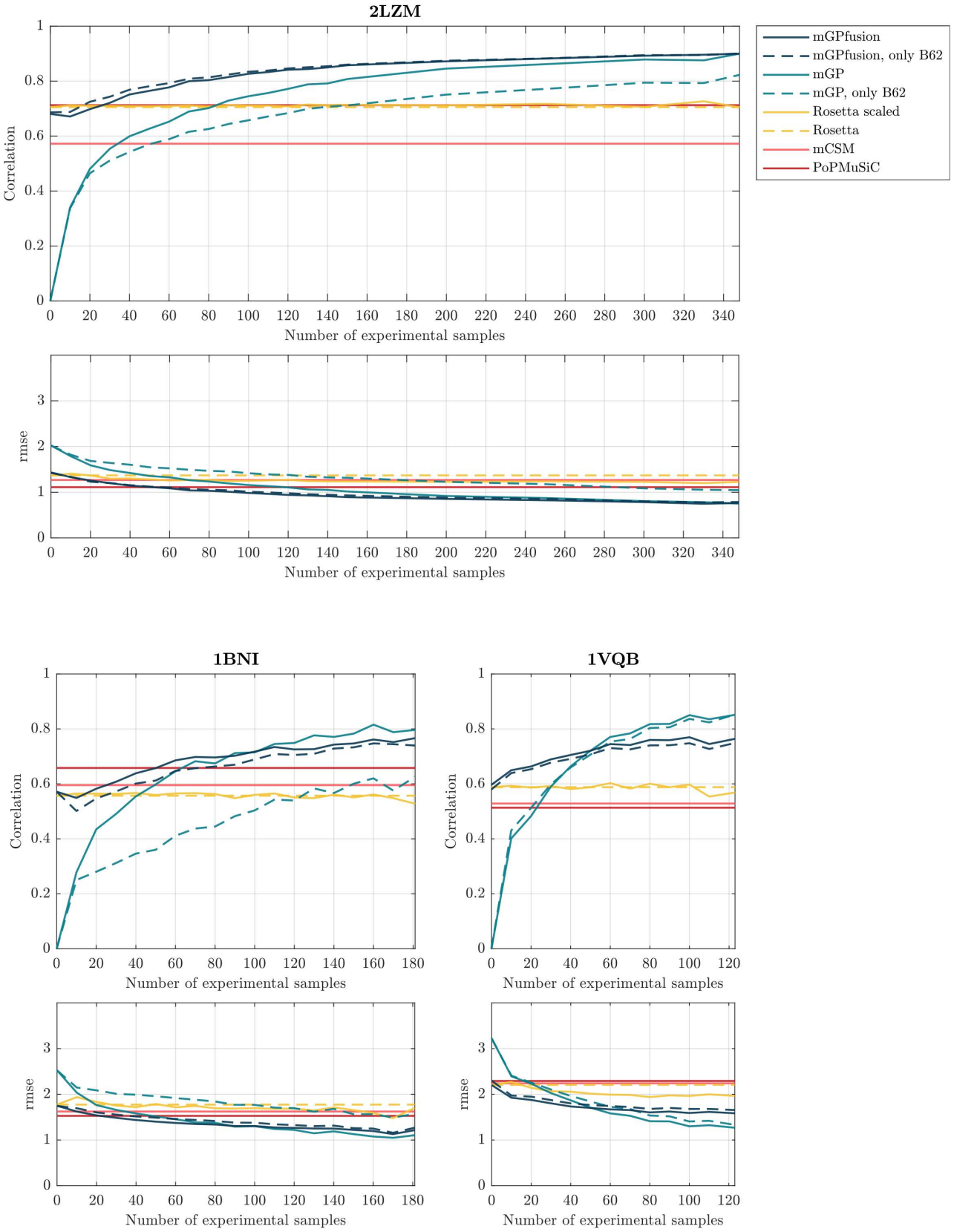

Fig. S6. Learning curves.

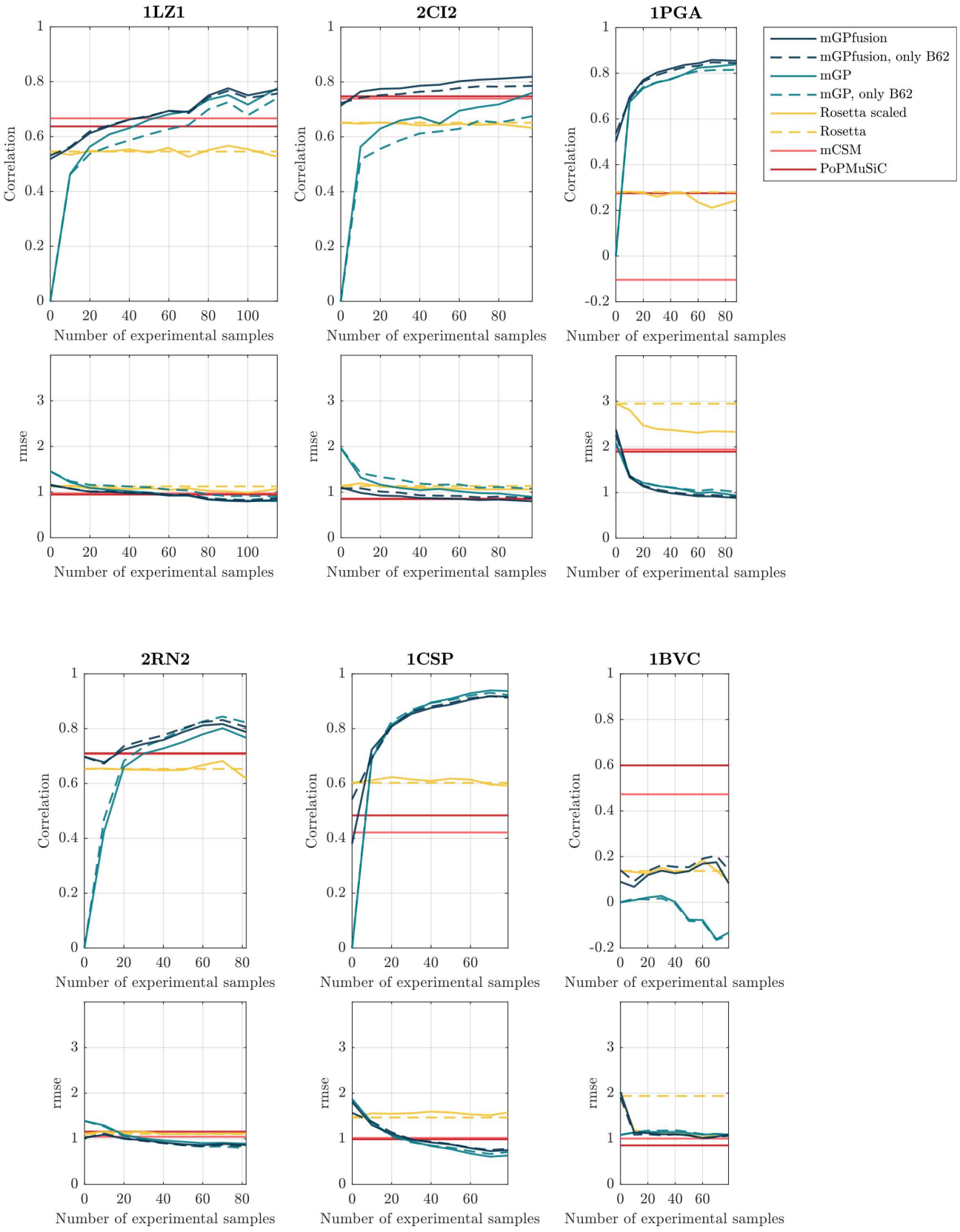

Fig. S7. Learning curves.

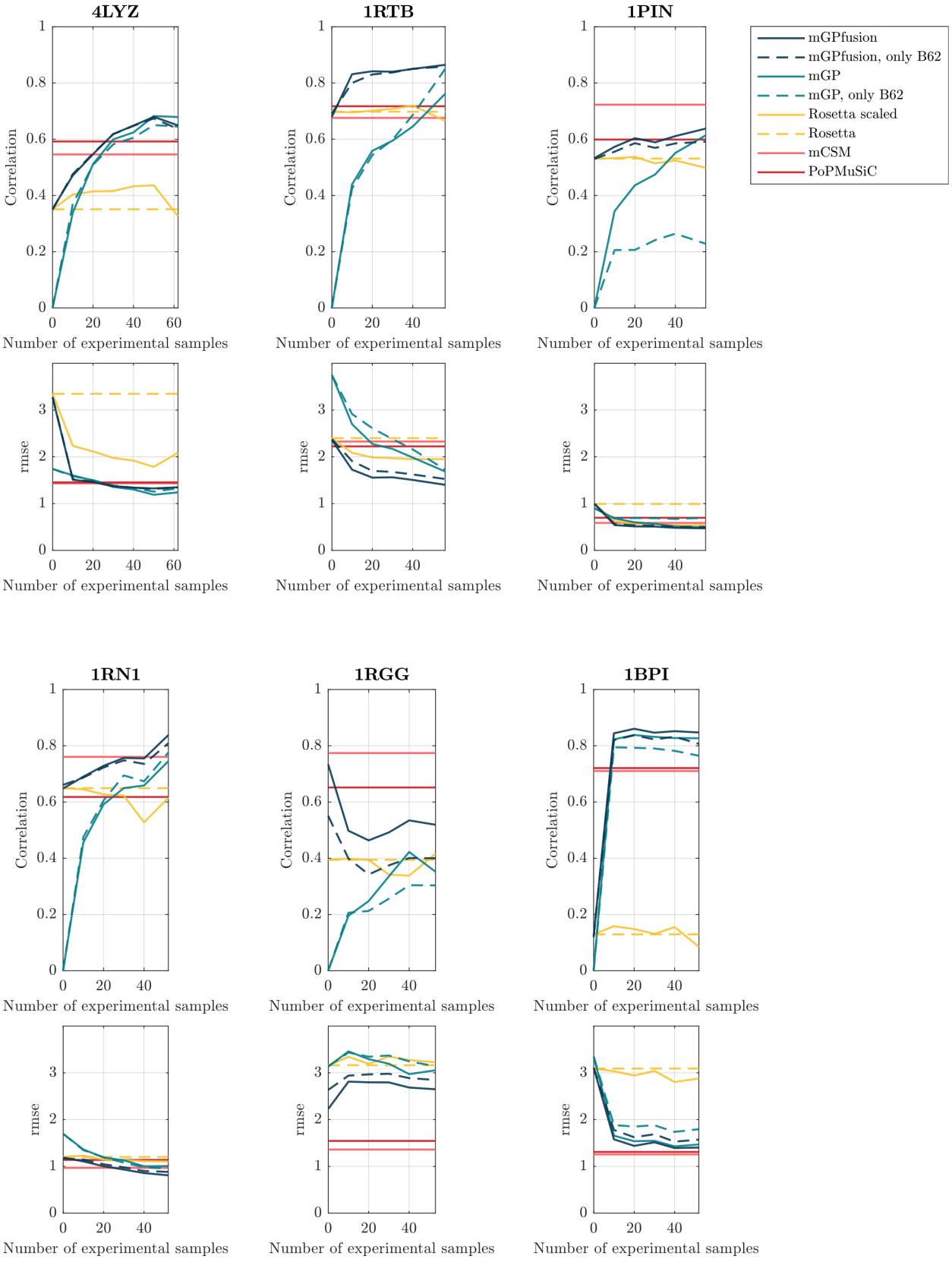

Fig. S8. Learning curves.

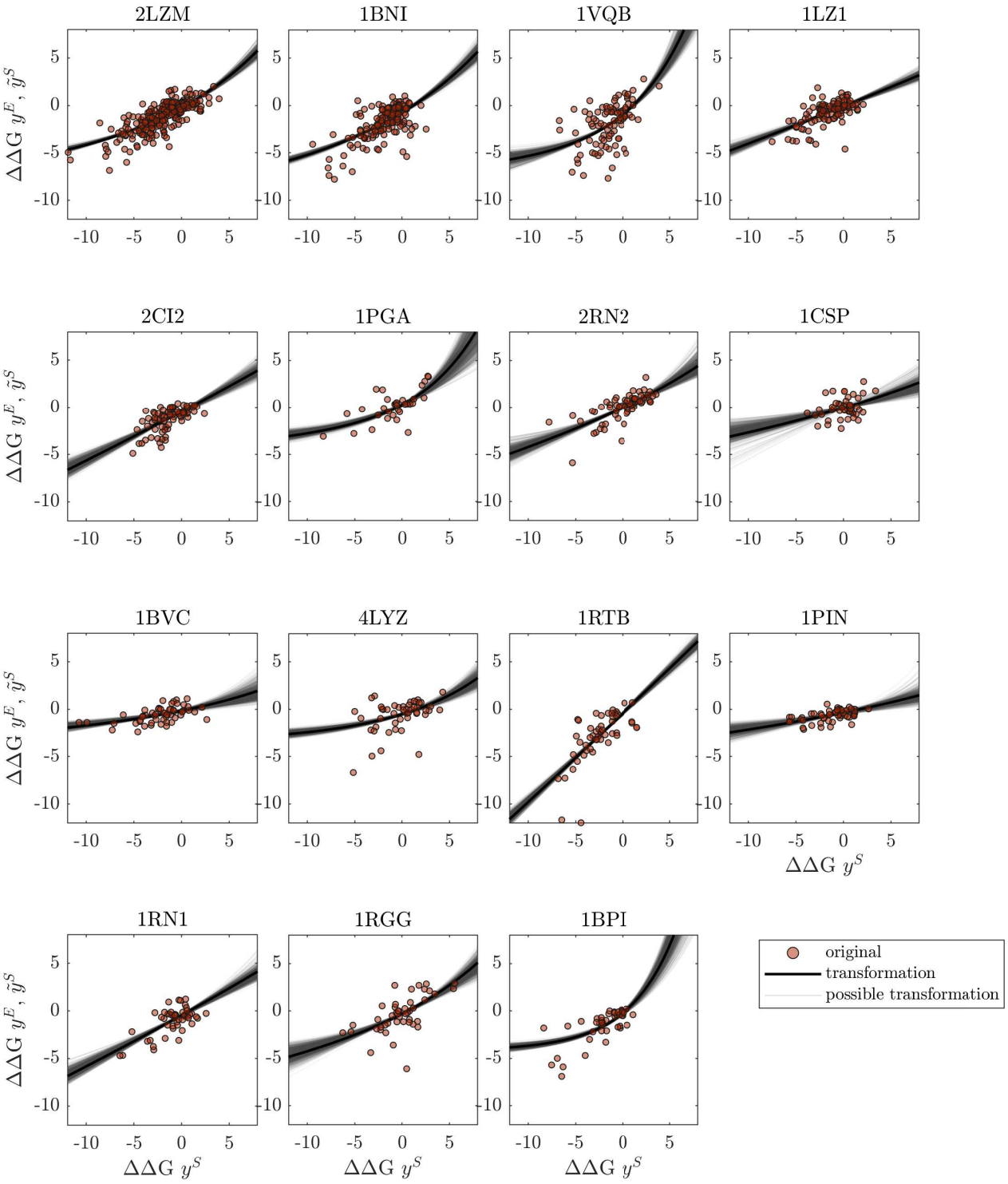

Fig. S9. Transformations for all 15 proteins presented in Table 1. The red circles mark the simulated ΔΔG-values y^S with respect to the experimental measured ΔΔG-values y^E. Thin black lines show possible transformations for y^S, whereas the thick black line shows the selected transformation from y^S to $\tilde{y}^S$.