

Predicting the effect of single mutations on protein stability and binding with respect to types of mutations

Preeti Pandey¹, Shailesh Kumar Panday¹, Prawin Rimal¹, Nicolas Ancona², Emil Alexov^{1,*}

¹ Department of Physics and Astronomy, Clemson University, Clemson, SC 29634

² Department of Biological Sciences, Clemson University, Clemson, SC 29634

* Correspondance: ealexov@clemson.edu

Supplementary Figures

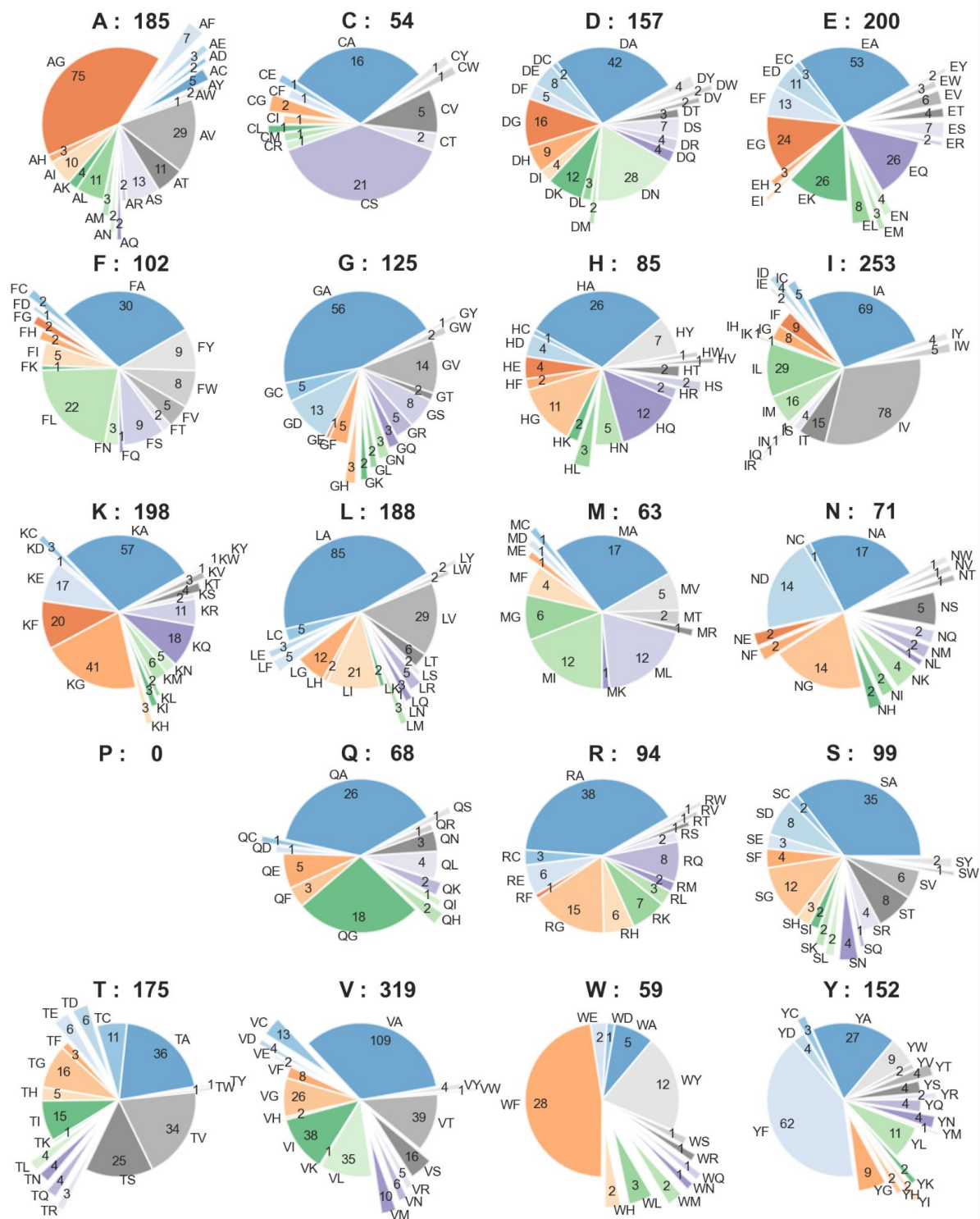


Figure S1a. Pie-chart summarizing count of different amino acid mutations in the S2648 dataset. In each subplot, the title shows the wild-type amino acids followed by their count in the whole dataset, and the two-letter label of every pie denotes the one-letter codes of the wild-type and mutant amino acid pair, respectively.

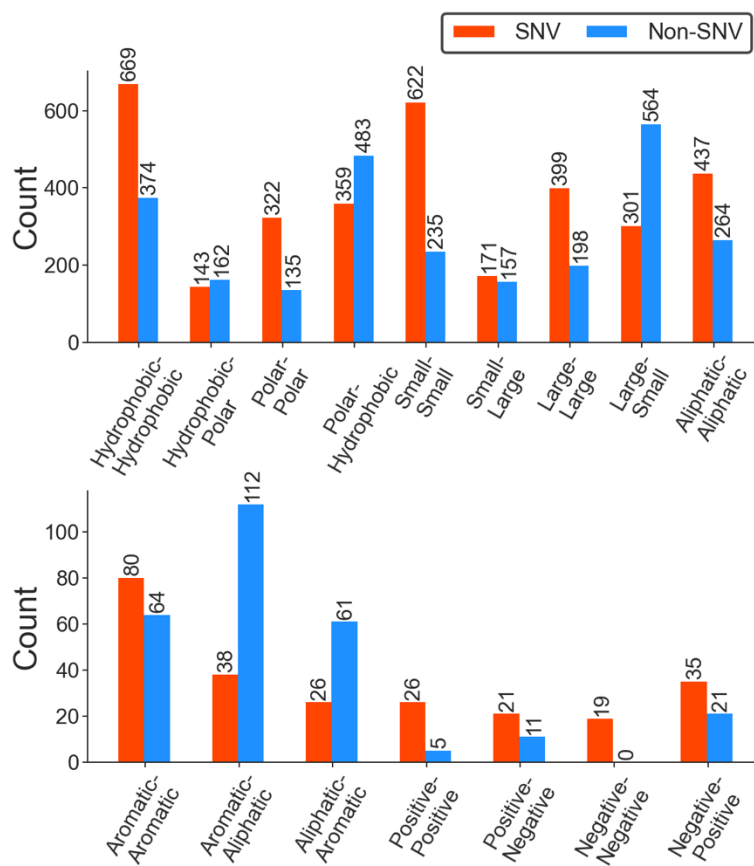


Figure S1c. SNV *vs* non-SNV mutations is assigned according to the property of amino acid mutation in the S2648 dataset. The two-word label of bars in the plot represents the amino acid properties of wild-type and mutant residues pairs, respectively. The amino acids according to their properties are categorized as follows: Ala, Cys, Gly, Ile, Leu, Met, Phe, Pro, Trp, and Val are categorized as hydrophobic. Asp, Glu, Lys, Arg, His, Asn, Gln, Ser, Thr, and Tyr are categorized as polar. His, Phe, Trp, and Tyr as aromatic. Ala, Ile, Lys, Leu, Met, Pro, and Val as aliphatic. His, Lys, Arg as positive. Asp, and Glu as negative. Ala, Cys, Gly, Ser, Asn, Asp, Pro, Thr, and Val as small. And Arg, Gln, Glu, His, Ile, Leu, Lys, Met, Phe, Trp, and Tyr as large amino acids.

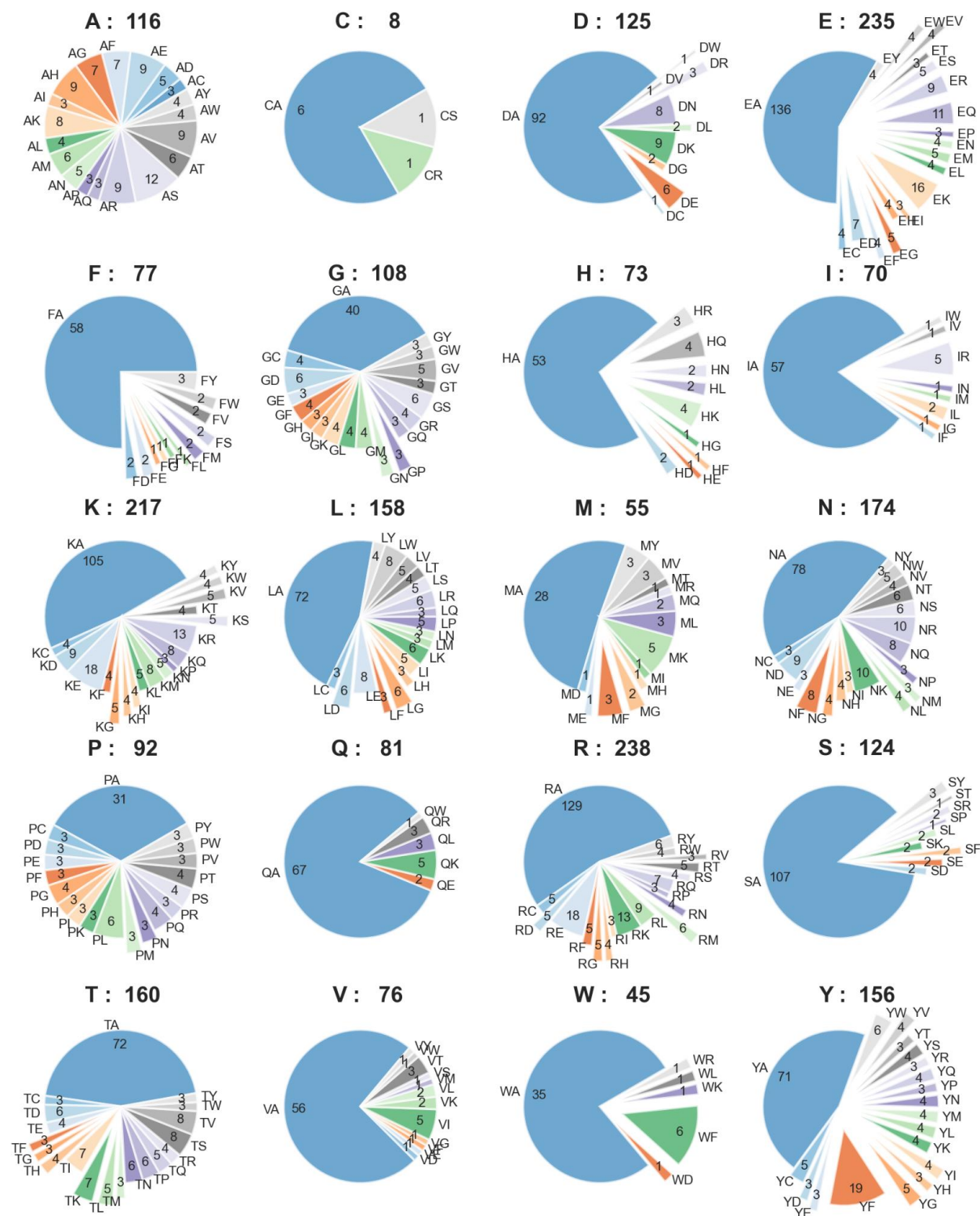


Figure S1d. Pie-chart summarizing count of different amino acid mutations in SKEMPI-SEQ-2388. In each subplot, the title shows the wild-type amino acids followed by their count in the whole dataset, and the two-letter label of every pie denotes the one-letter codes of the wild-type and mutant amino acid pair respectively.

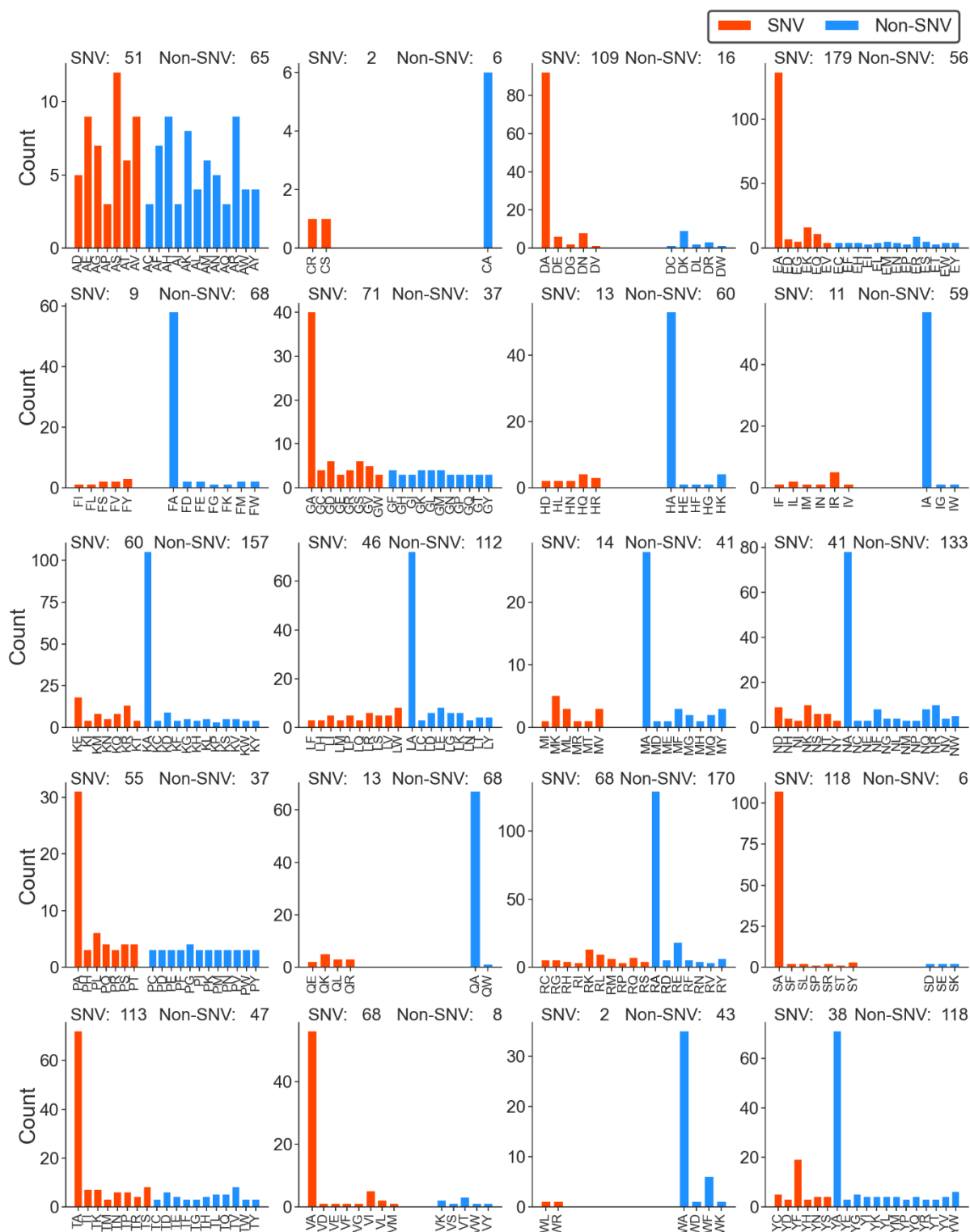


Figure S1e. Bar plot summarizing count of SNV and non-SNV mutations in SKEMPI-SEQ-2388 dataset. The two-letter label of every bar denotes the one-letter codes of the wild-type and mutant amino acid pairs, respectively.

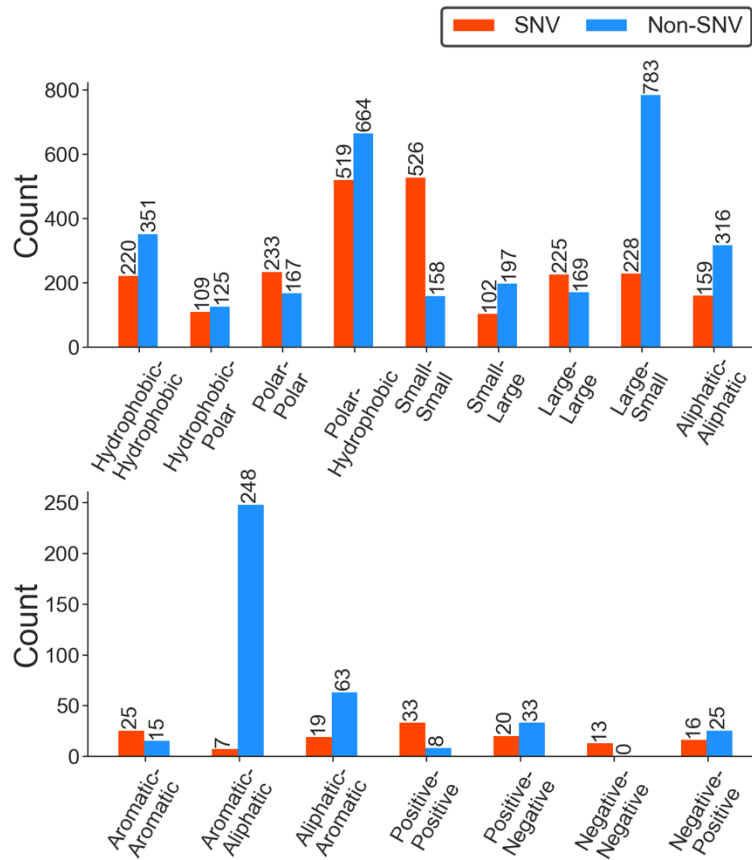


Figure S1f. SNV vs non-SNV mutations is assigned according to the type of amino acid mutation in the SKEMPI-SEQ-2388 dataset. The two-word label of bars in the plot represents the amino acid properties of wild-type and mutant residues pairs, respectively. The amino acids according to their properties are categorized as follows: Ala, Cys, Gly, Ile, Leu, Met, Phe, Pro, Trp, and Val are categorized as hydrophobic. Asp, Glu, Lys, Arg, His, Asn, Gln, Ser, Thr, and Tyr are categorized as polar. His, Phe, Trp, and Tyr as aromatic. Ala, Ile, Lys, Leu, Met, Pro, and Val as aliphatic. His, Lys, Arg as positive. Asp, and Glu as negative. Ala, Cys, Gly, Ser, Asn, Asp, Pro, Thr, and Val as small. And Arg, Gln, Glu, His, Ile, Leu, Lys, Met, Phe, Trp, and Tyr as large amino acids.

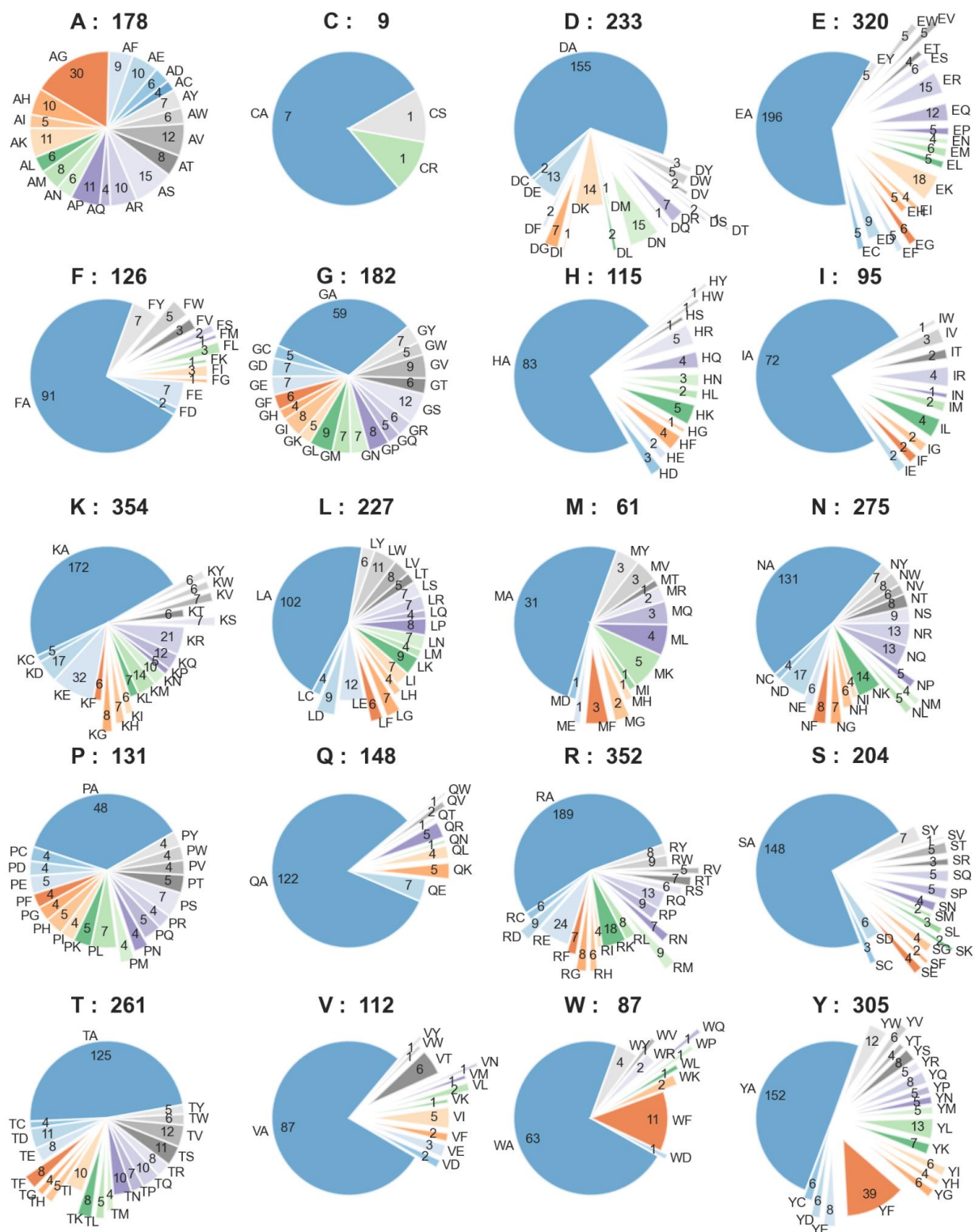


Figure S1g. Pie-chart summarizing count of different amino acid mutations in SKEMPI-3D-3775 dataset. In each subplot, the title shows the wild-type amino acids followed by their count in the whole dataset, and the two-letter label of every pie denotes the one-letter codes of the wild-type and mutant amino acid pair respectively.

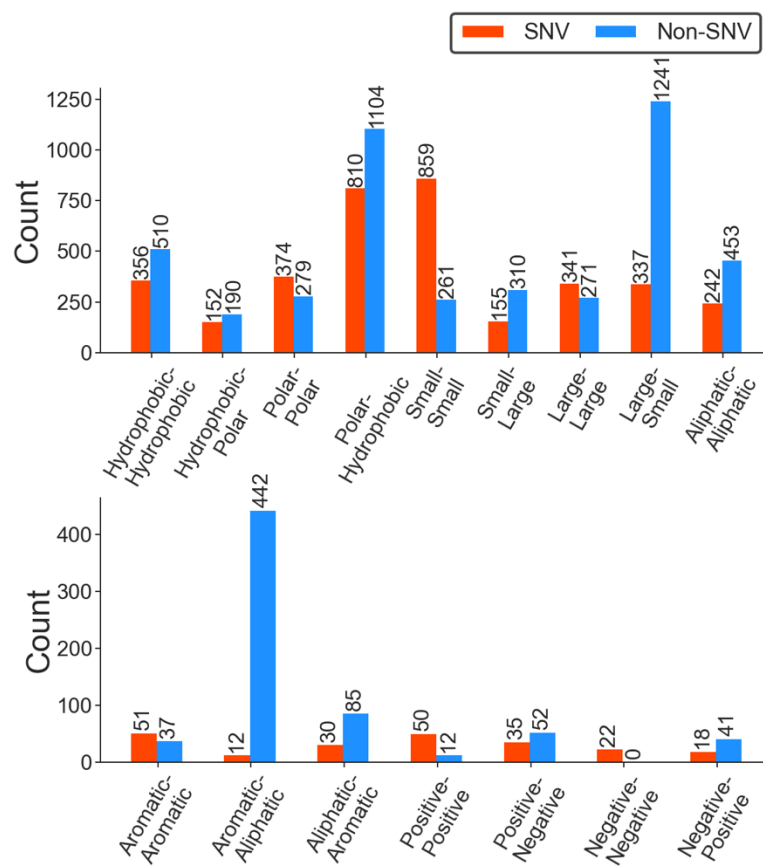


Figure S1i. SNV *vs* non-SNV mutations is assigned according to the type of amino acid mutation in the SKEMPI-3D-3775 dataset. The two-word label of bars in the plot represents the amino acid properties of wild-type and mutant residues pairs, respectively. The amino acids according to their properties are categorized as follows: Ala, Cys, Gly, Ile, Leu, Met, Phe, Pro, Trp, and Val are categorized as hydrophobic. Asp, Glu, Lys, Arg, His, Asn, Gln, Ser, Thr, and Tyr are categorized as polar. His, Phe, Trp, and Tyr as aromatic. Ala, Ile, Lys, Leu, Met, Pro, and Val as aliphatic. His, Lys, Arg as positive. Asp, and Glu as negative. Ala, Cys, Gly, Ser, Asn, Asp, Pro, Thr, and Val as small. And Arg, Gln, Glu, His, Ile, Leu, Lys, Met, Phe, Trp, and Tyr as large amino acids.

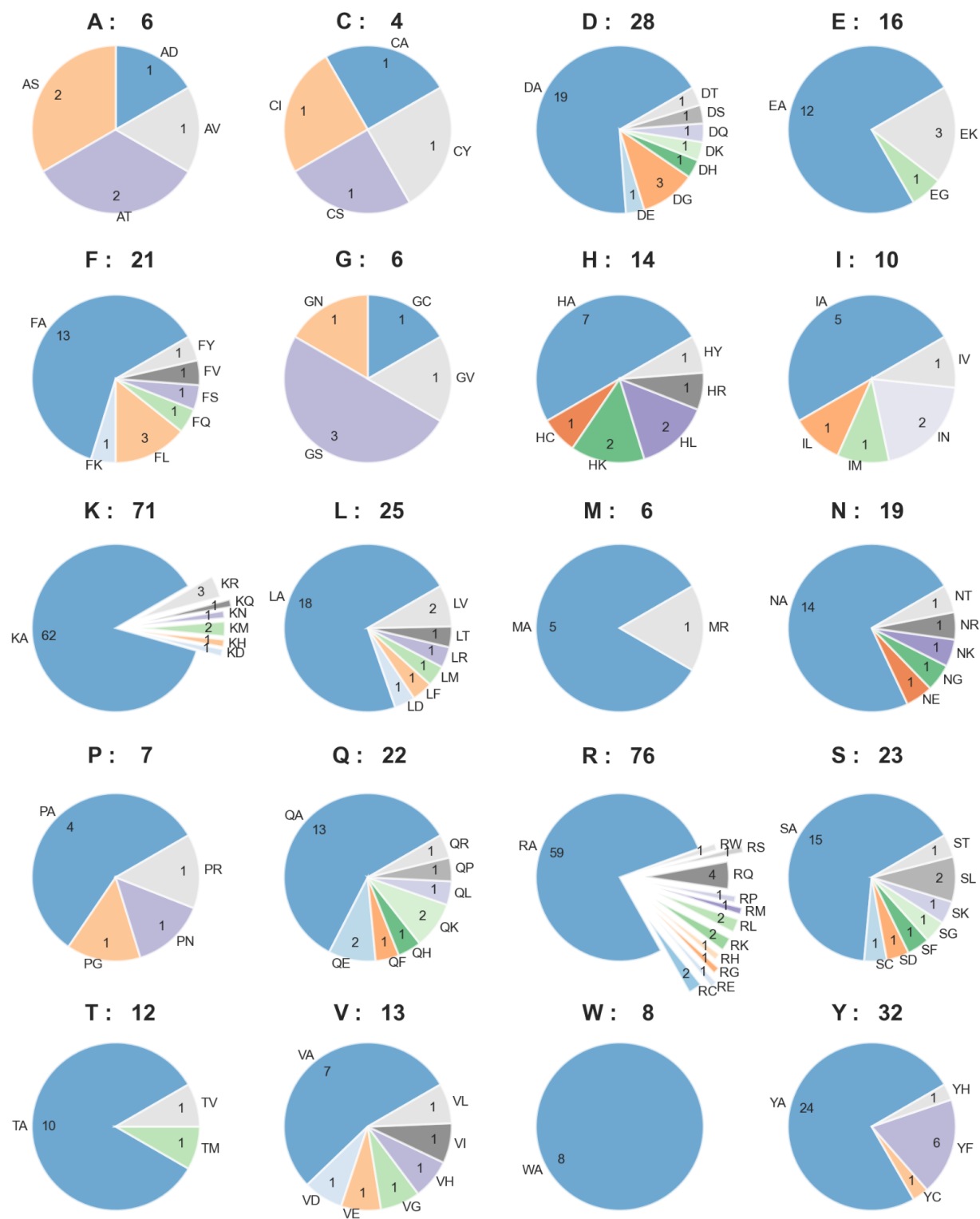


Figure S1j. Pie-chart summarizing count of different amino acid mutations in the S419 dataset. In each subplot, the title shows the wild-type amino acids followed by their count in the whole dataset, and the two-letter label of every pie denotes the one-letter codes of the wild-type and mutant amino acid pair respectively.

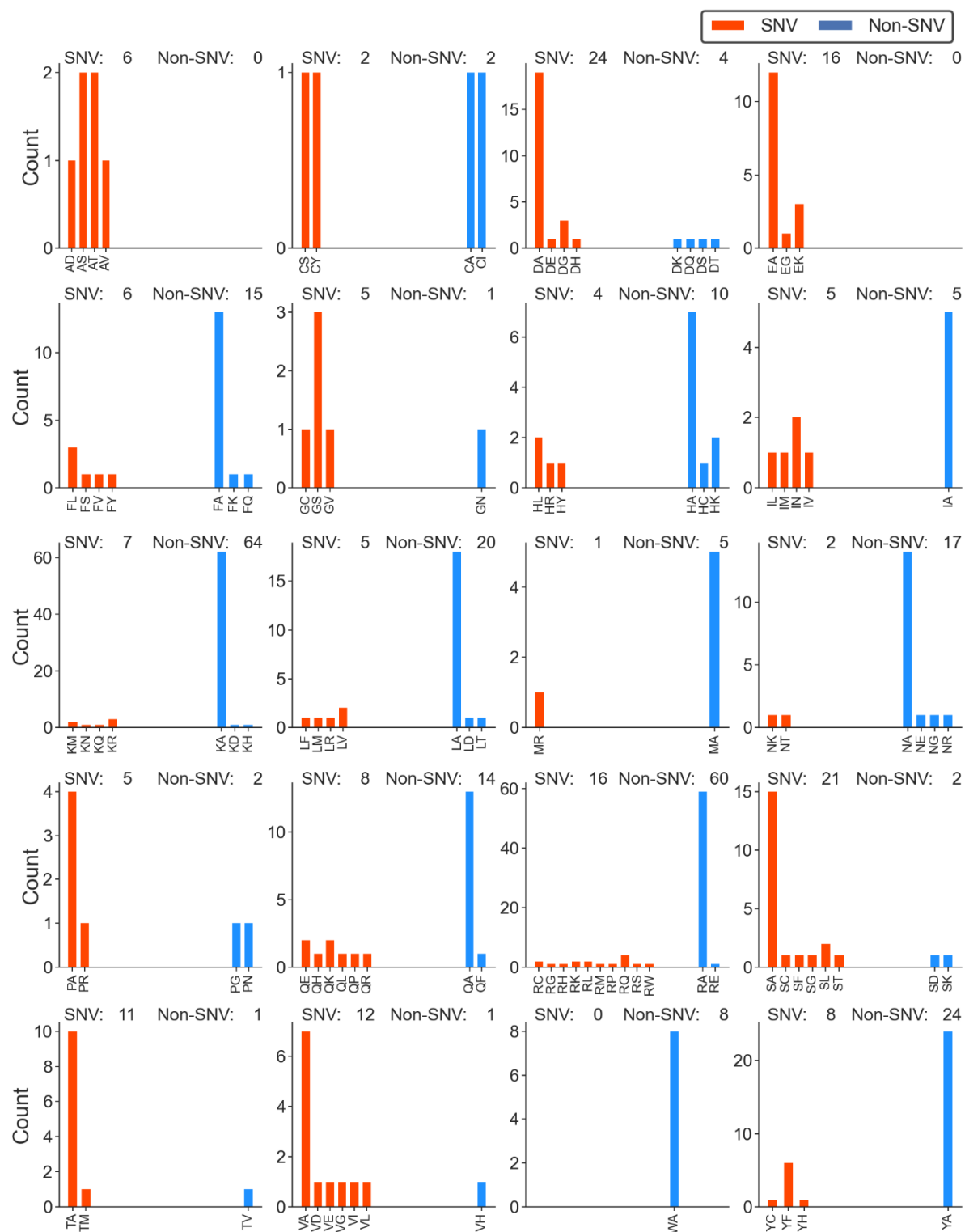


Figure S1k. Bar plots summarizing the count of SNV and non-SNV mutations in the S419 dataset. The two-letter label of every bar denotes the one-letter codes of the wild-type and mutant amino acid pairs, respectively.

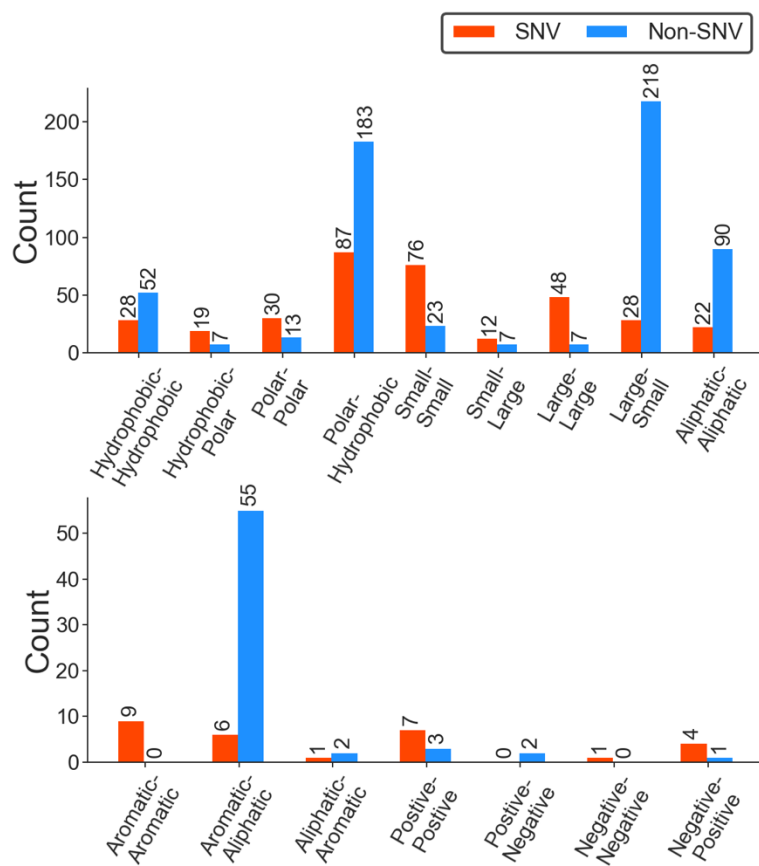


Figure S11. SNV vs non-SNV mutations is assigned according to the type of amino acid mutation in the S419 dataset. The two-word label of bars in the plot represents the amino acid properties of wild-type and mutant residues pairs, respectively. The amino acids according to their properties are categorized as follows: Ala, Cys, Gly, Ile, Leu, Met, Phe, Pro, Trp, and Val are categorized as hydrophobic. Asp, Glu, Lys, Arg, His, Asn, Gln, Ser, Thr, and Tyr are categorized as polar. His, Phe, Trp, and Tyr are aromatic. Ala, Ile, Lys, Leu, Met, Pro, and Val as aliphatic. His, Lys, Arg as positive. Asp, and Glu as negative. Ala, Cys, Gly, Ser, Asn, Asp, Pro, Thr, and Val as small. And Arg, Gln, Glu, His, Ile, Leu, Lys, Met, Phe, Trp, and Tyr as large amino acids.

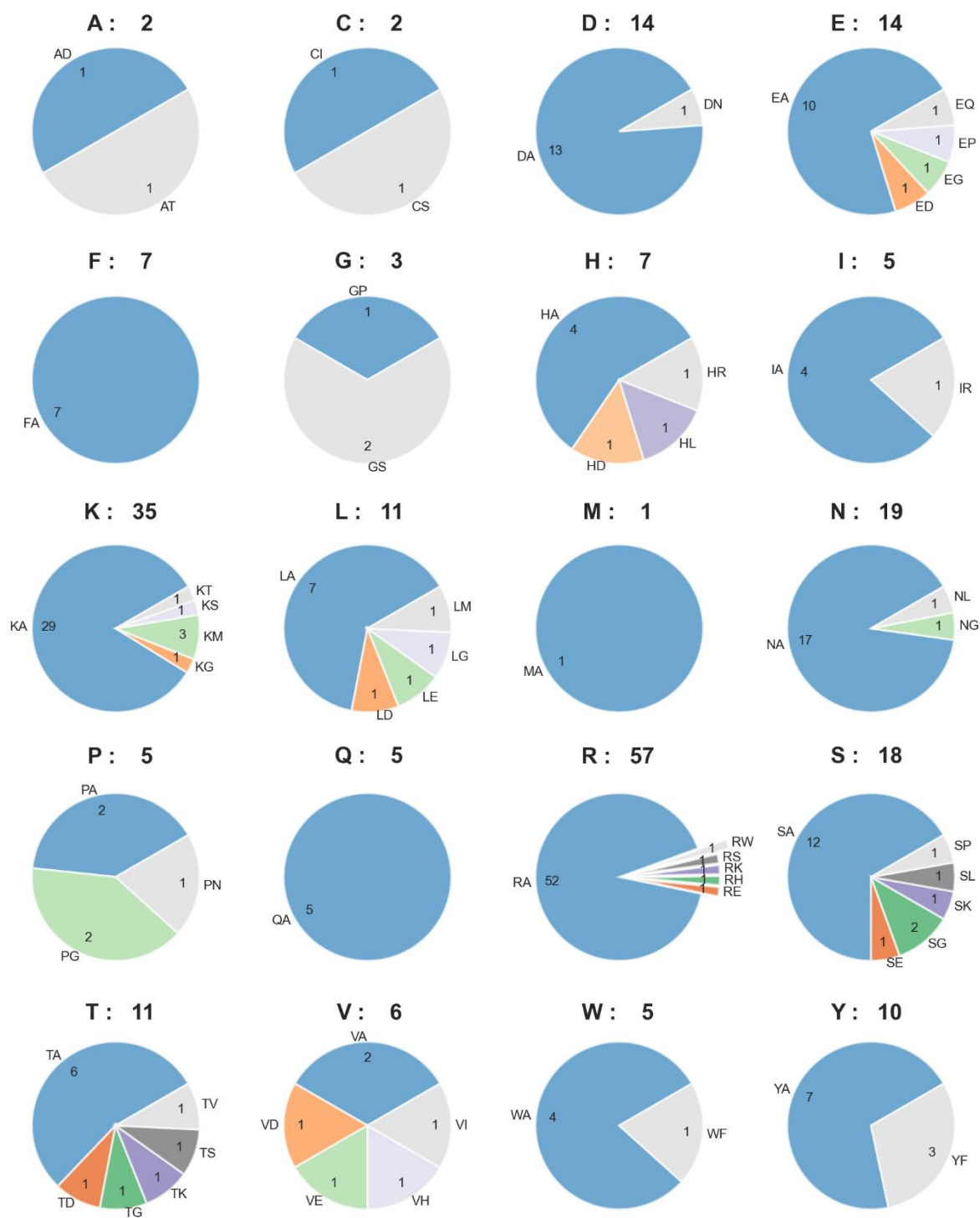


Figure S1m: Pie-chart summarizing count of different amino acid mutations in ProNAB-237. In each subplot, the title shows the wild-type amino acids followed by their count in the whole dataset, and the two-letter label of every pie denotes the one-letter codes of the wild-type and mutant amino acid pair respectively.

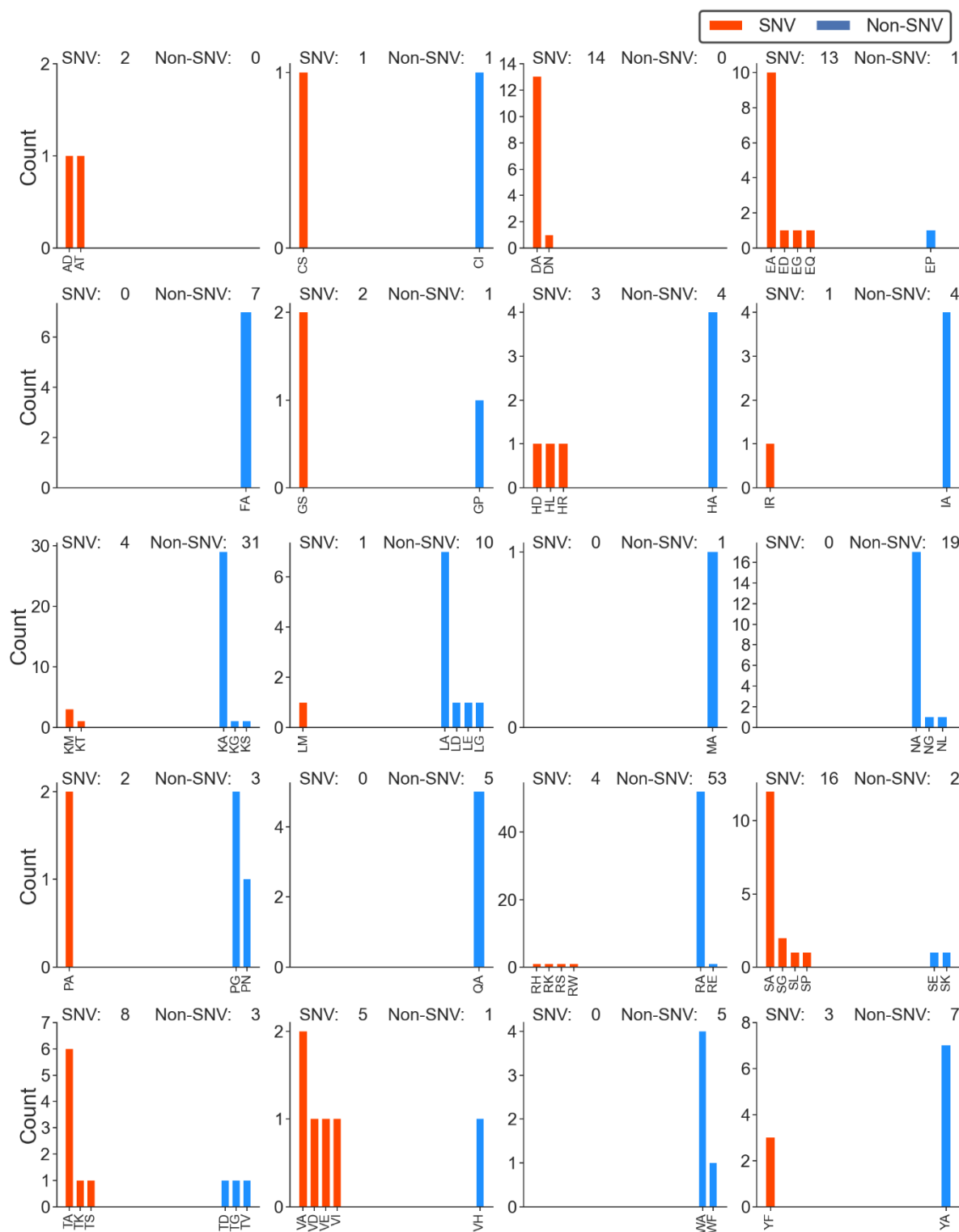


Figure S1n. Bar plots summarizing the count of SNV and non-SNV mutations in the ProNAB-237 dataset. The two-letter label of every bar denotes the one-letter codes of the wild-type and mutant amino acid pairs, respectively.

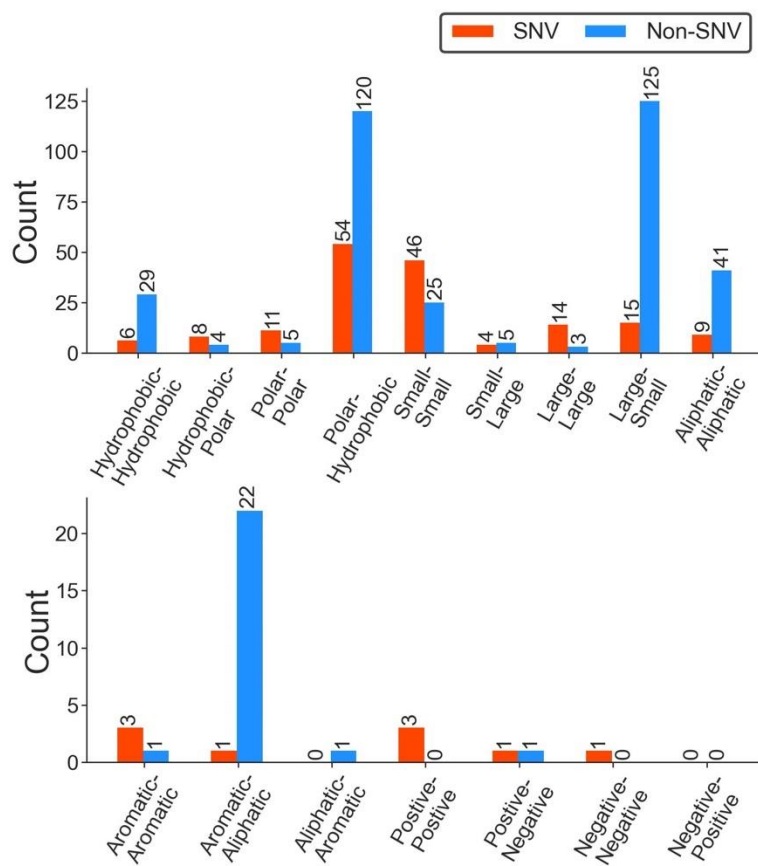


Figure S10. SNV vs non-SNV mutations is assigned according to the type of amino acid mutation in the ProNAB-237 dataset. The two-word label of bars in the plot represents the amino acid properties of wild-type and mutant residues pairs, respectively. The amino acids according to their properties are categorized as follows: Ala, Cys, Gly, Ile, Leu, Met, Phe, Pro, Trp, and Val are categorized as hydrophobic. Asp, Glu, Lys, Arg, His, Asn, Gln, Ser, Thr, and Tyr are categorized as polar. His, Phe, Trp, and Tyr as aromatic. Ala, Ile, Lys, Leu, Met, Pro, and Val as aliphatic. His, Lys, Arg as positive. Asp, and Glu as negative. Ala, Cys, Gly, Ser, Asn, Asp, Pro, Thr, and Val as small. And Arg, Gln, Glu, His, Ile, Leu, Lys, Met, Phe, Trp, and Tyr as large amino acids.

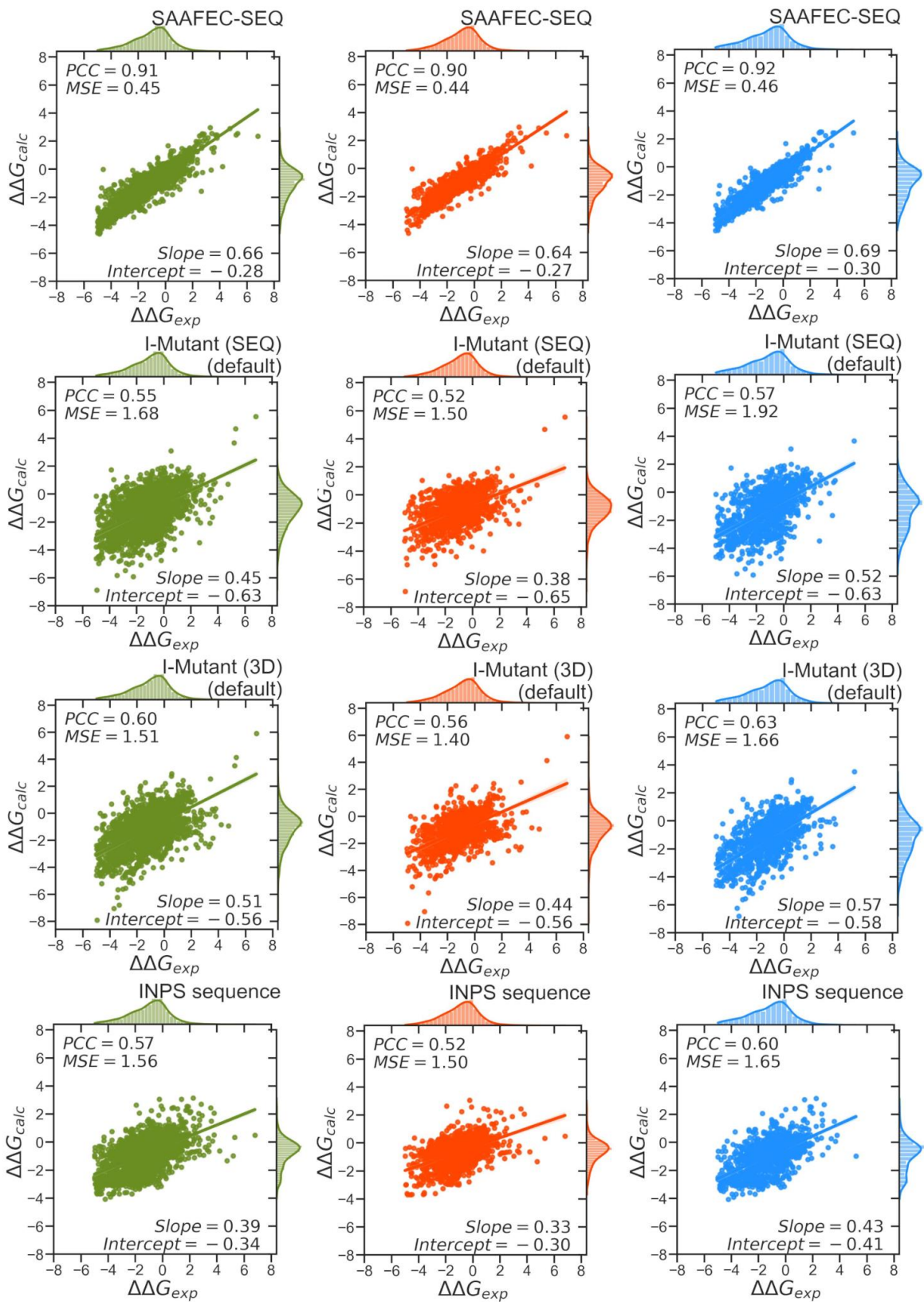


Figure S2a. Plots of predicted $\Delta\Delta G$ against experimental $\Delta\Delta G$ for the S2648 dataset. The whole data is shown in green, SNV in red, and non-SNV in blue.

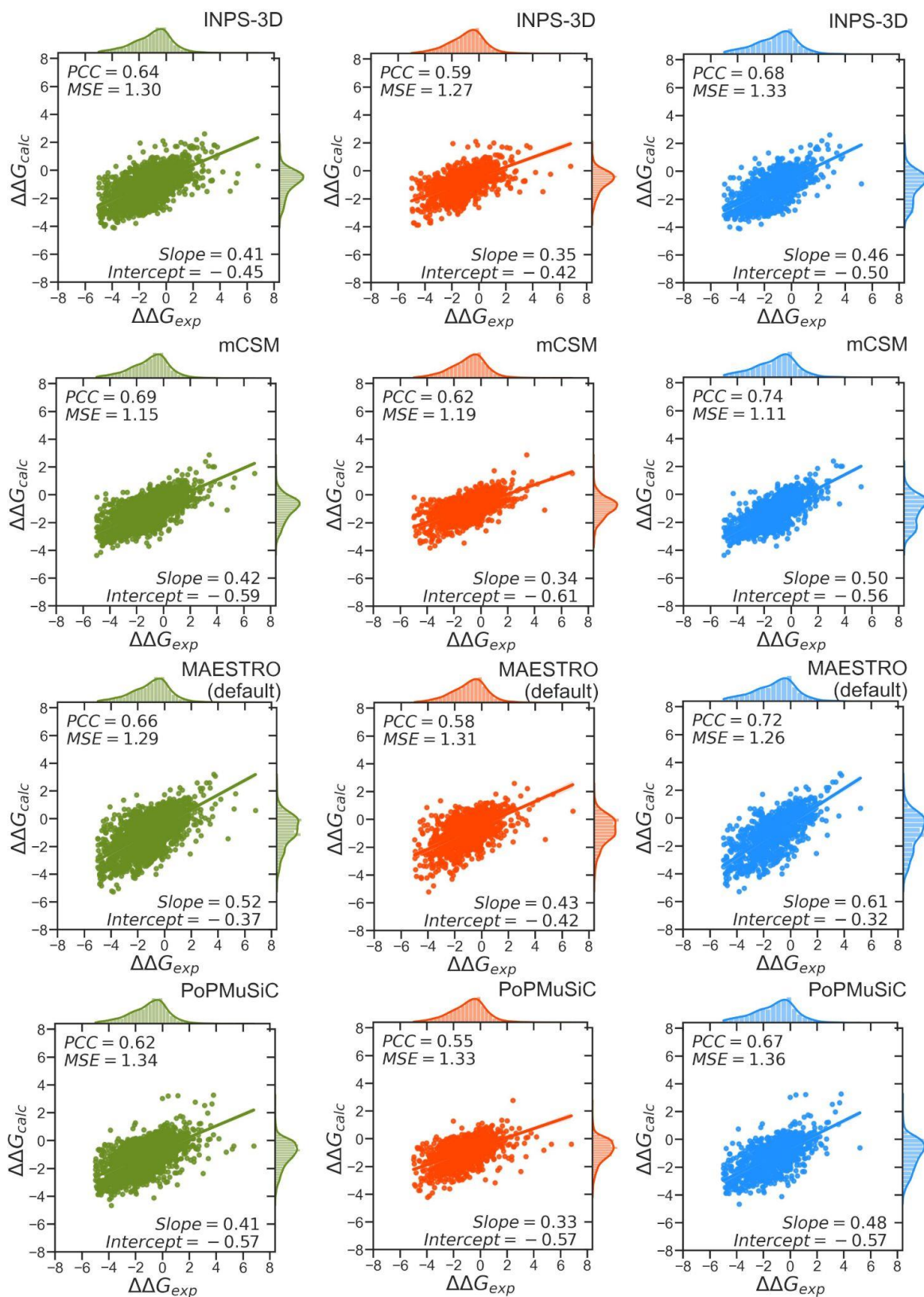


Figure S2b. Plots of predicted $\Delta\Delta G$ against experimental $\Delta\Delta G$ for the S2648 dataset. The whole data is shown in green, SNV in red, and non-SNV in blue.

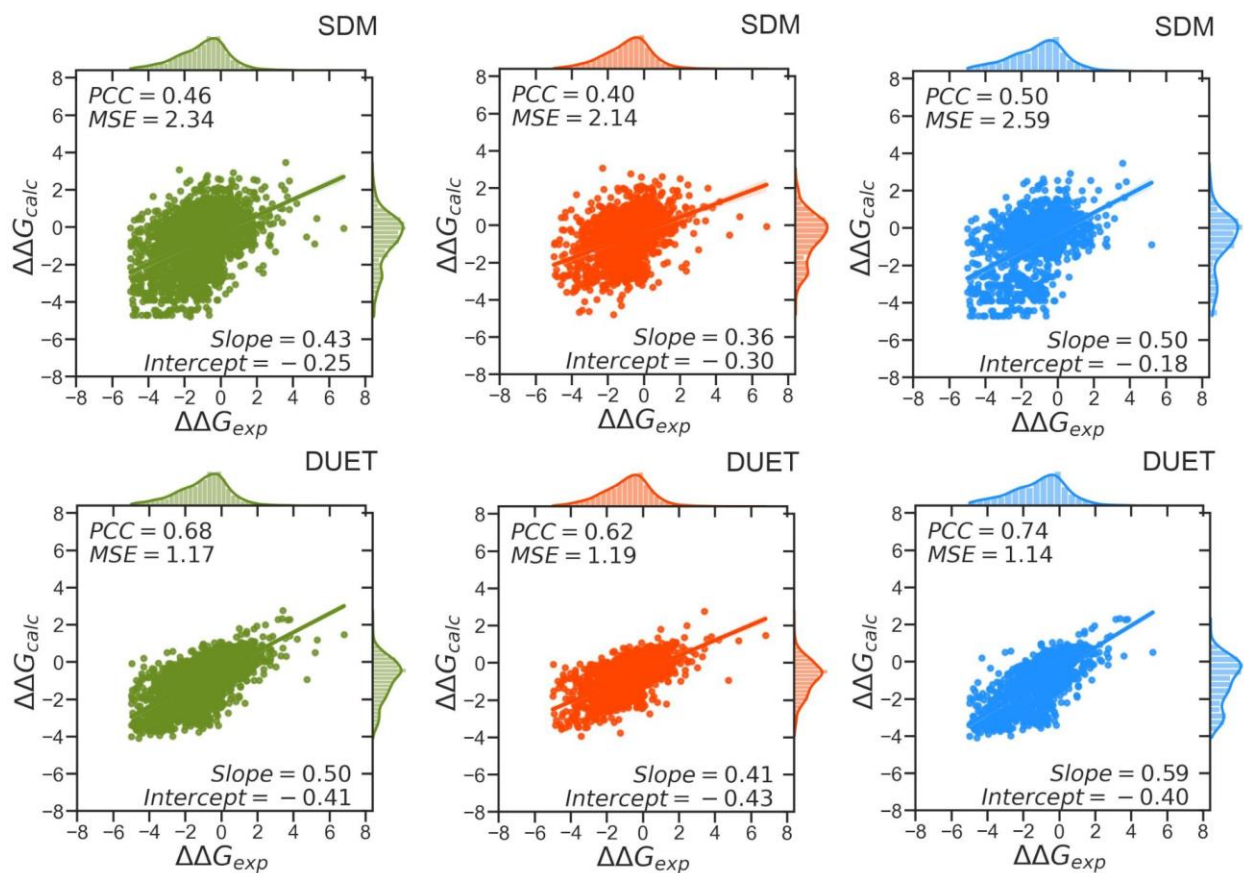


Figure S2c. Plots of predicted $\Delta\Delta G$ against experimental $\Delta\Delta G$ for the S2648 dataset. The whole data is shown in green, SNV in red, and non-SNV in blue.

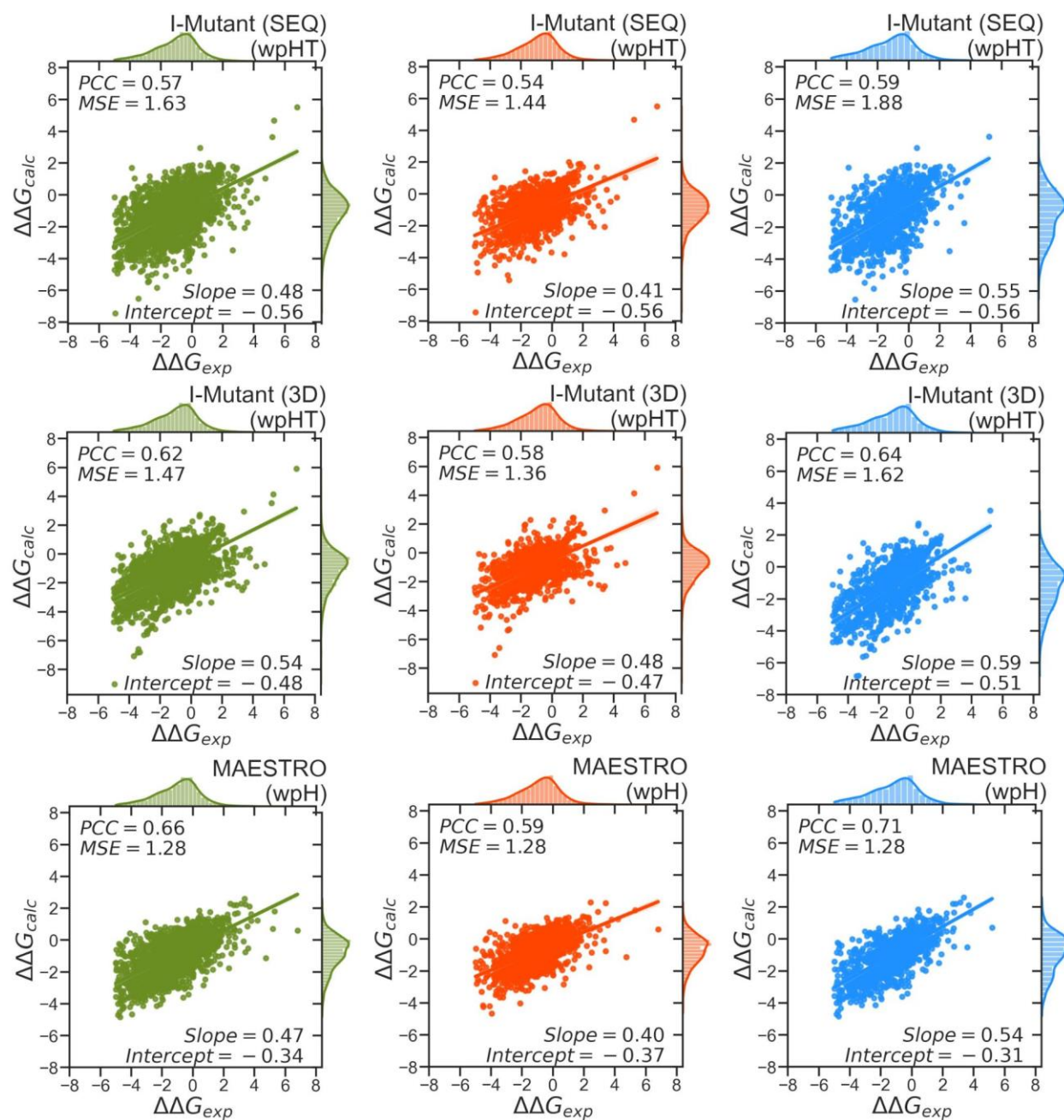


Figure S2d. Plots of predicted $\Delta\Delta G$ against experimental $\Delta\Delta G$ for the S2648 dataset. The whole data is shown in green, SNV in red, and non-SNV in blue. wpHT : the calculations have been performed using experimental pH and temperature and wpH : the calculations have been performed using experimental pH.

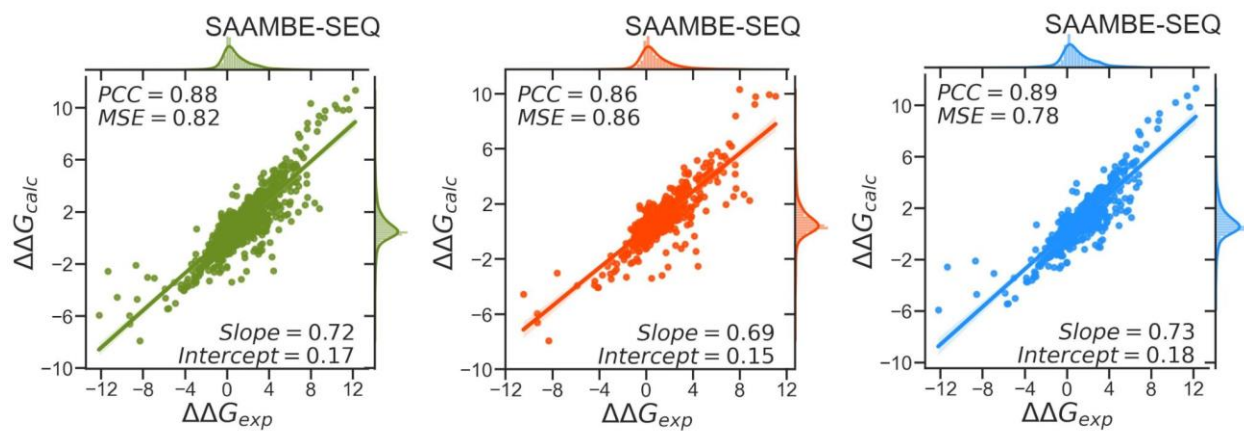


Figure S3a. Plots of predicted $\Delta\Delta G$ against experimental $\Delta\Delta G$ for the SKEMPI-SEQ-2388 dataset. The whole data is shown in green, SNV in red, and non-SNV in blue.

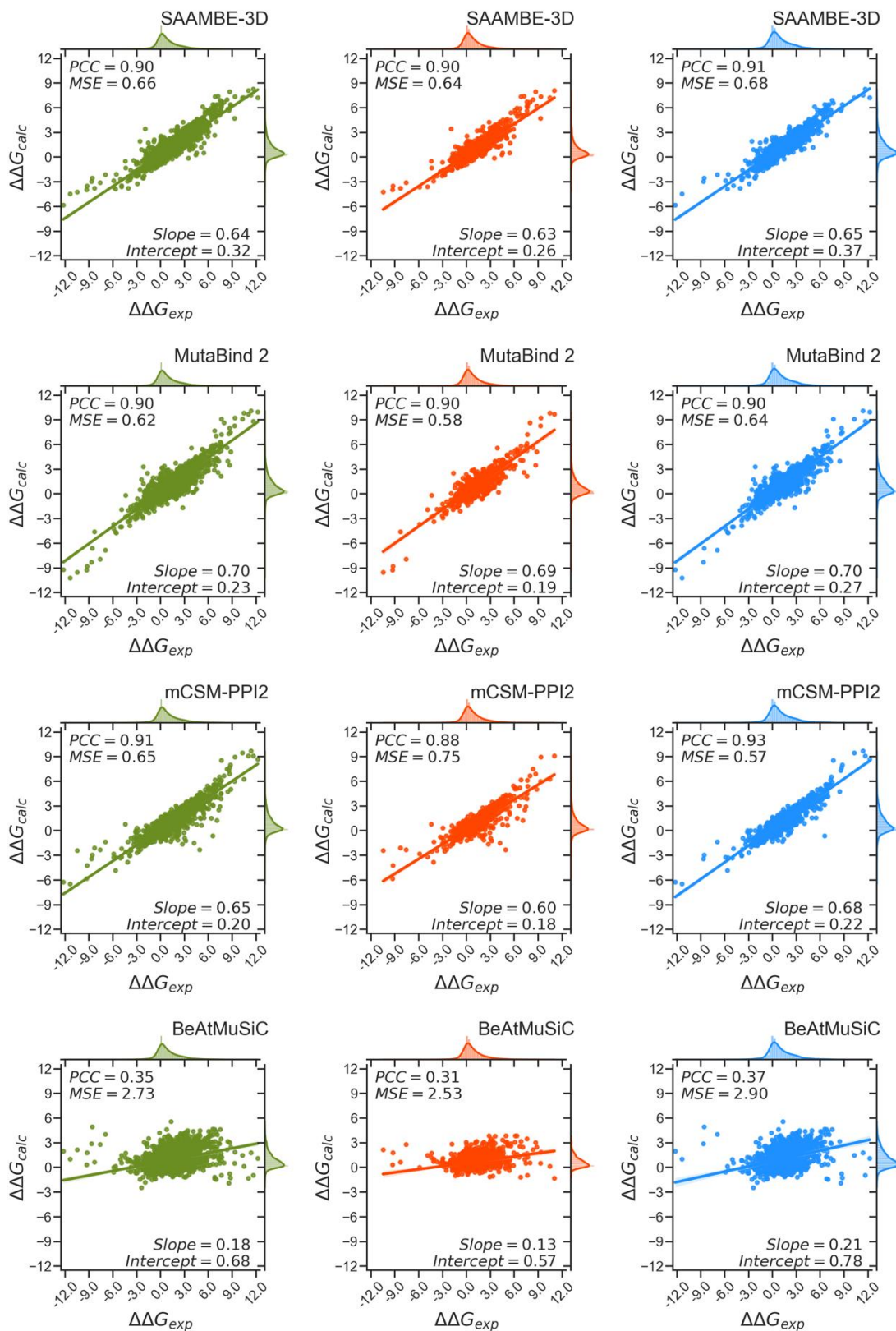


Figure S3b. Plots of predicted $\Delta\Delta G$ against experimental $\Delta\Delta G$ for the SKEMPI-3D-3775 dataset. The whole data is shown in green, SNV in red, and non-SNV in blue.

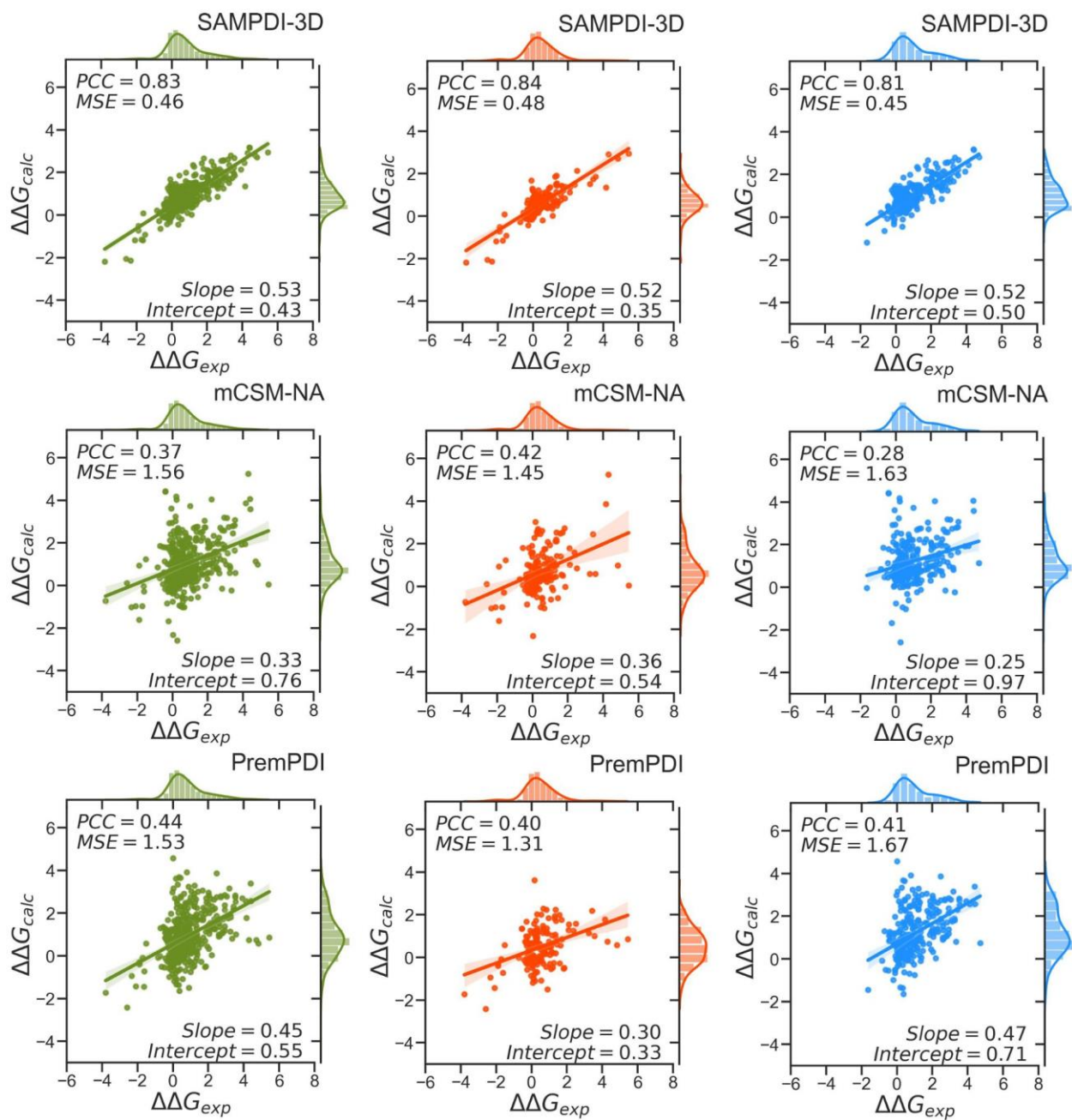


Figure S4a. Plots of predicted $\Delta\Delta G$ against experimental $\Delta\Delta G$ for the S419 dataset. The whole data is shown in green, SNV in red, and non-SNV in blue.

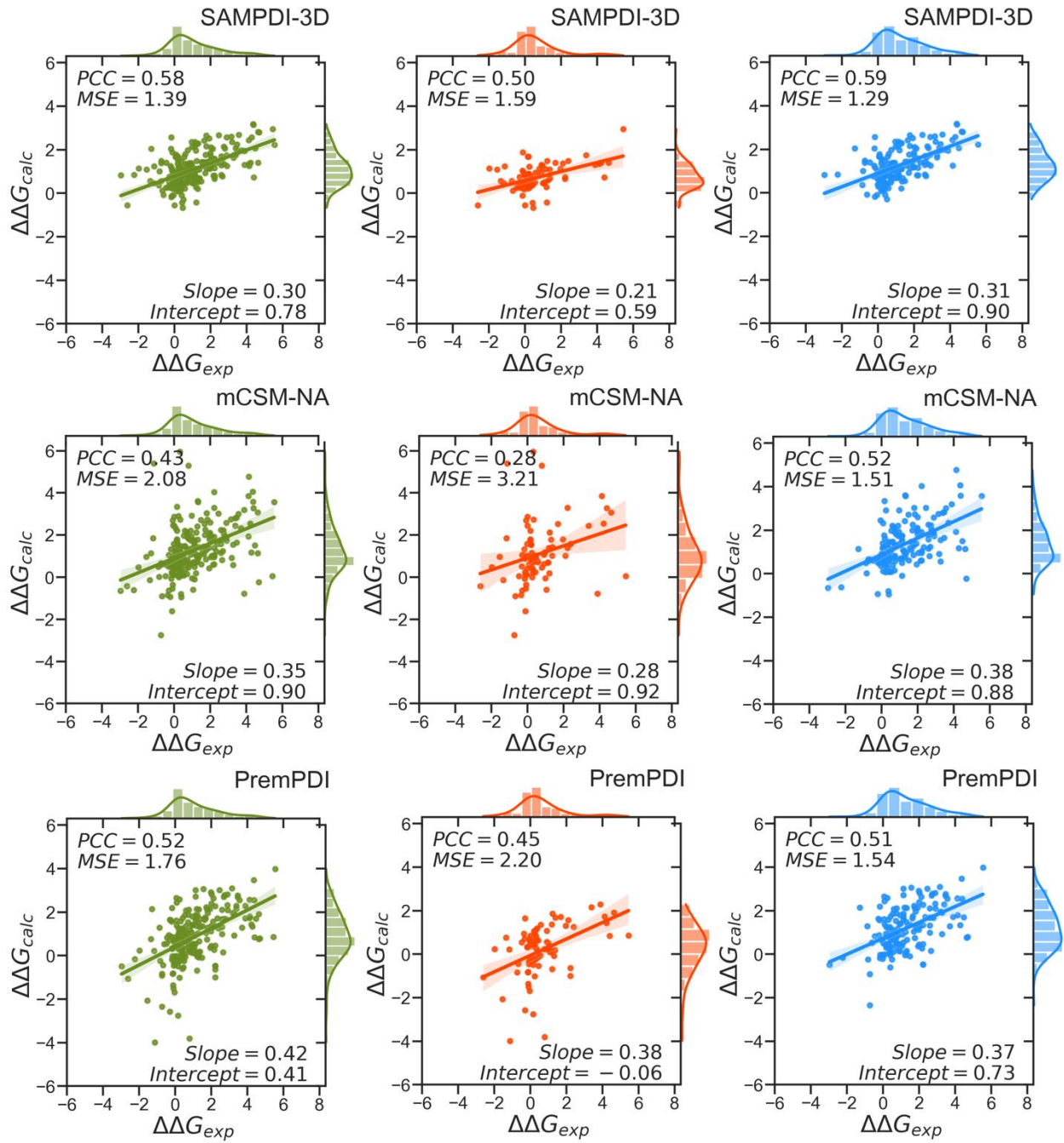


Figure S4b. Plots of predicted $\Delta\Delta G$ against experimental $\Delta\Delta G$ for the ProNAB-237 dataset. The whole data is shown in green, SNV in red, and non-SNV in blue.

		Second Nucleotide					
		U	C	A	G		
First Nucleotide	U	UUU Phe UUC UUA Leu UUG	UCU UCC Ser UCA UCG	UAU Tyr UAC UAA STOP UAG STOP	UGU Cys UGC UGA STOP UGG Trp	U	Third Nucleotide
	C	CUU CUC Leu CUA CUG	CCU CCC Pro CCA CCG	CAU His CAC CAA Gln CAG	CGU CGC Arg CAA CAG	U	
	A	AUU AUC Ile AUA AUG Met	ACU ACC Thr ACA ACG	AAU Asn AAC AAA Lys AAG	AGU Ser AGC AGA Arg AGG	U	
	G	GUU GUC Val GUA GUG	GCU GCC Ala GCA GCG	GAU Asp GAC GAA Glu GAG	GGU GGC Gly GGA GGG	U	

Figure S5a. Universal genetic code.

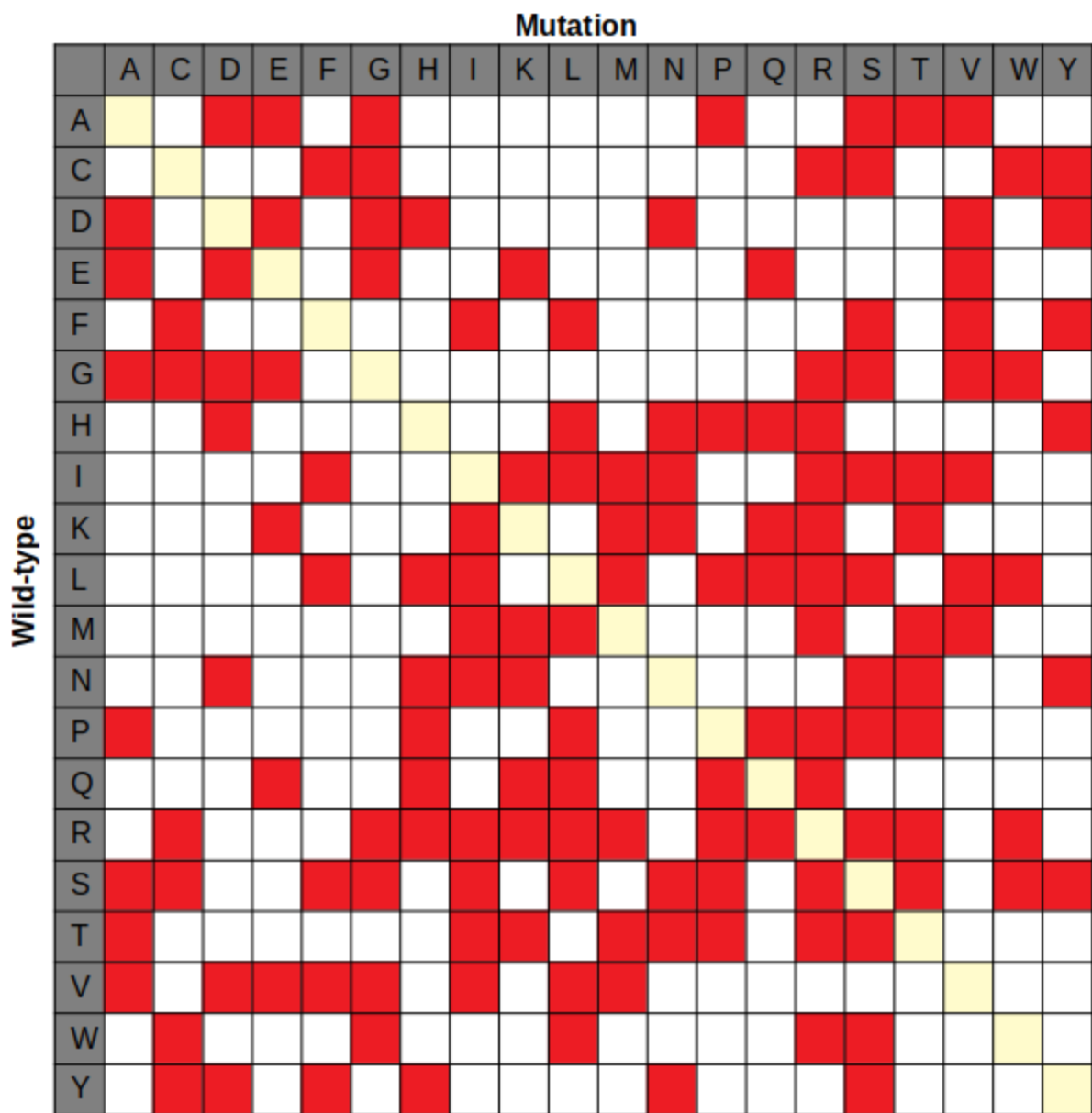


Figure S5b. Heat map showing SNV and non-SNV amino acid mutations. Red boxes indicate cases of SNVs and white cases of non-SNVs.

Supplementary Tables

Table S1. Number of SNV and non-SNV cases in the datasets.

Dataset	Total	SNV	Non-SNV
S2648	2647	1493	1154
SKEMPI-SEQ-2388	2388	1081	1307
SKEMPI-3D-3775	3775	1692	2083
S419	419	164	255
ProNAB-237	237	79	158