Table 1:
The best coefficients of determination R^2 and predictive correlation Q^2 of the COMBINE models of thrombin, trypsin, and urokinase were tabulated in respect the latent variables (LV). The best Q^2-LOO (leave one out) and Q^2-LTO (leave two out) values for thrombin, trypsin and urokinase were 0.89 (LV5), 0.83 (LV3), and 0.68 (LV4), respectively. In trypsin, variable selection did not improve the model, but in thrombin and urokinase the models could be improved according to internal cross-validation by using D-optimal pre-selection (D-opt) and fractional factorial design (FFD) variable selection at LV4. For thrombin LV4 and LV5 resulted in nearly the same values. Due to the risk of over fitting, a lower latent variable was chosen. The COMBINE model of thrombin could be slightly improved by using four highly conserved water molecules in the active site as additional 'residues' (X-variables).

model	**numb. of struct.**	**variable selection**	**active variables (energy >10^{-7} kcal/mol)**	**LV**	**R2**	**SDEC**	**Q2 LTO**	**SDEP LTO**	**Q2 LOO**	**SDEP LOO**
thrombin	25	-	582 (230)	4	0.93	0.59	0.81	1.00	0.82	0.98
		D-opt LV4 FFD LV4	105	4	0.96	0.46	0.88	0.79	0.89	0.77
thrombin with 4 conserved water mol.	24		590 (236)	4 5	0.94 0.96	0.54 0.46	0.83 0.83	0.97 0.96	0.83 0.84	0.96 0.93
trypsin	37	-	448 (210)	3	0.90	0.72	0.82	0.97	0.83	0.97
		D-opt LV3 FFD LV3	100	3	0.90	0.73	0.83	0.97	0.83	0.97
urokinase	26	-	492 (217)	4	0.83	0.62	0.62	0.93	0.63	0.91
		D-opt LV4 FFD LV4	108	4	0.84	0.60	0.67	0.87	0.68	0.86