Electronic Supplementary Material (ESI) for Nanoscale.

Supplementary information for

Protein corona composition of PEGylated nanoparticles correlates strongly with amino acid composition of protein surface

Giovanni Settanni, Jiajia Zhou, Tongchuan Suo, Susanne Schöttler, Katharina Landfester, Friederike Schmid, Volker Mailänder

Molecular dynamics simulations

All the simulations were carried out using the program NAMD[1] and the charmm27 force field[2,3] with the extension for PEG[4]. Tip3p[5] was used as model for the explicit treatment of water. An integration time step of 1 fs was used across the simulations. Simulations were carried out using periodic boundary conditions. Pressure and temperature were maintained constant at 1atm and 300K, respectively, during the simulations using the Langevin piston algorithm and Langevin thermostat[6,7]. A cutoff of 1.2 nm was used for the non-bonded interactions with a switch function. Long range electrostatic interactions were treated using the smooth particle mesh Ewald (PME) method[8] with a grid spacing of about 0.1 nm. To prepare the PEG-water mixture, 64 PEG molecules (H-[O-CH_2-CH_2]$_n$-OH, with n either 4 or 7) were placed on a 4x4x4 grid with 1. nm spacing between grid points. Then 1ns high temperature (700K) simulations in vacuum with damped electrostatics interactions (dielectric constant 200) were run to randomize the initial dihedral distribution of the PEG molecules. The PEG molecules were then immersed in a box of water molecules and sodium and chlorine ions were added to reach physiological concentration (0.15 M). Mixtures with different concentrations of PEG were obtained by changing the size of the water box surrounding the PEG molecules. The prepared mixtures were then

equilibrated first at high temperature (373K) for 1.0 ns and then at 300K for 1.0 ns. The initial coordinates of the proteins were taken from the PDB (see Table S1 for the list of PDBids). Each protein was immersed in a box filled by replicating the coordinates of the PEG-water mixtures obtained before in the three space directions and removing mixture atoms in close contact with protein atoms. The final PEG concentration in the simulation boxes is reported in Table S1. The size of the boxes was large enough to leave at least 1.0 /nm from each protein atom and the box boundary. In the case of HSA, larger box sizes with 1.5 and 2.0 nm distances between protein and box boundary were simulated to investigate the dependence of the simulation results upon box size. The total charge of the systems was neutralized by changing an adequate number of water molecules into ions. The complete systems were then minimized using the steepest descent algorithm for 10000 steps with harmonic restraints on the heavy atoms of the proteins. Then the systems were equilibrated at room temperature and pressure for 1.0 ns during which the harmonic restraints were gradually removed and for 1.0 ns without restraints. Finally production runs were started with 4 or 5 replicas for each system. Most of the runs reached the 200ns time length (Table S1).

Table S1 List of the performed simulations

System	Sim. Name	PDBid	Box size (Å)	N. Atoms	PEG length	[PEG] (g/ml)	Simulation time (ns)
HSA	HSA1	1A06	98.8	100881	4	0.08	4 x 200
	HSA2		98.2	99301	4	0.11	4 x 200
	HSA3		108.6	134134	4	0.12	4 x 200
	HSA4		109.0	134778	7	0.04	5 x 200
	HSA5		118.2	172541	4	0.12	5 x 100
CQ1C		2WNVC	77.0	47912	4	0.12	5 x 200
Transferrin	TRF1	2HAV	108	132009	4	0.11	5 x 200
	TRF2		108.7	134663	4	0.07	5 x 200
	TRF3		108.4	133303	7	0.07	5 x 200
	TRF4		108.8	134372	7	0.04	5 x 200

Analysis of trajectories

The direct PEG-protein interactions along the simulations were measured using the NAMD pair-interaction utility. The time series of these interactions reaches convergence during the course of the simulations with relaxation times between 10 and 40ns. The time series along the trajectories of the number of PEG and water heavy atoms found within 0.5 nm of each amino acid type was determined using the "pbwithin" selection command of VMD. The ratio of these numbers was compared to the ratio between all the PEG and water heavy atoms in the simulation box (bulk ratio). The autocorrelation function of this number relaxes in less than 10ns along all the analyzed simulations. Thus, the first 10ns of data were discarded and 10ns-long block averages were used for the determination of standard deviations. A PEG/water ratio larger than bulk for an amino acid implies the presence of an effective attractive interaction between PEG and the amino acid. On the other hand, amino acids with a PEG/Water ratio smaller than bulk exert and effective repulsion for PEG.

Mass spectrometry

The concentration of proteins in plasma, and on PEG_{44}-coated polystyrene nanoparticles (PS-PEG44) was obtained by mass spectrometry. The protein concentrations on the following nanoparticles were also analyzed: PEG_{110}-coated polystyrene (PS-PEG110), $PEEP_{49}$-coated polystyrene (PS-PEEP49) , $PEEP_{92}$-coated polystyrene (PS-PEEP92), amino-functionalized polystyrene (PS-NH2), Lutensol-stabilized polysterene (PS-LUT) and Lutensol-stabilized amino-functionalized polystyrene (PS-LUT-NH2) nanoparticles. All the above mentioned data are reported in Table S2. These datasets have been published elsewhere[9,10] along with the experimental methods for mass spectrometry. In cases where the protein was not detected in plasma but only on the nanoparticle, the plasma concentration of the protein was taken from the literature[11] and converted to the same units as in the mass spectrometry data. The conversion factor was obtained by taking the slope of

the linear fit of plasma concentrations from mass spectroscopy *versus* the value from the literature for those proteins where both were available. This is almost equivalent to taking the concentration of serum albumin in plasma (0.605mM, that is 40g/L) as reference. In cases where the protein was not detected on the nanoparticles (indicated by 0 fmol in Table S2), a value of 0.0025 fmol was used in the calculations requiring the $\log([P]_{np})$ (0.005 fmol is the smallest amount reported in the experiments), to prevent the emergence of a singularity.

Structural analysis

Whenever they were available, the structures of the proteins with a non-zero concentration on the nanoparticles were collected from the PDB or from the database of homology models[12] when the homology with the template was larger than 40%. In cases where the protein is a subunit of a complex, only the structure of the corresponding subunit was retained. Overall 36 protein structures satisfied the criteria and were retained for analysis (Table S2). The solvent exposed surface area of each amino-acid of the collected proteins was measured using VMD and 0.14 nm as probe radius. The fraction of solvent accessible surface area $FSASA_{AA}(P)$ of each amino acid type in the proteins is reported in Table S3.

Fit of model via bootstrap procedure

In eq. (5) the left hand side is obtained from mass spectrometry experiments and, in the right hand side, the $FSASA_{AA}(P)$ are obtained from the structural analysis described above. The $PEGW_{AA}$ are considered as the free parameters of the fit. The fit was carried out using a bootstrap approach, which allows for estimating the robustness of the results to changes in the fitted data set. The approach is as follows: the initial data set of 36 proteins was resampled 100 times, i.e. 100 new sets of 36 proteins were generated by randomly picking any protein in the original set, allowing for duplicates. The resampled sets contained from a minimum of 20 to a maximum of 30 unique proteins of the original set (median 25.5). For each

resampled data set an optimal set of parameters $PEGW_{AA}$ was obtained by a simple null-temperature Monte Carlo procedure where the parameters were randomly modified and the new parameters were accepted only if they led to an increase in the correlation coefficient between the left and the right hand side of eq.(5). The correlation coefficient between two sets of values $\{X_i\}$ and $\{Y_i\}$ was defined as:

$$r = \frac{\sum^{i} X_i Y_i - \langle X \rangle \langle Y \rangle}{\sqrt{(\langle X^2 \rangle - \langle X \rangle^2)(\langle Y^2 \rangle - \langle Y \rangle^2)}}$$

The $PEGW_{AA}$ parameters obtained in each resampled set are then used to determine the $PEGW_{tot}$ of all the proteins in the original set, that is also those not present in the resampled set (see fig. S1 for an example). A rescaling of the $PEGW_{AA}$ does not change the correlation coefficient between r.h.s. and l.h.s. of eq.(5). Thus, we rescaled the $PEGW_{AA}$ obtained for each resampled set, so that the average $PEGW_{tot}$ over the original set of proteins (not the resampled) equals 1. The extracted parameters are then rescaled so that the PEG/water ratio for ALB is the same as the one measured with the simulations. Finally, we computed the average over the resampled sets of the $PEGW_{AA}$ and their standard deviation (Table 1). Similarly we computed the average and the standard deviation over the resampled sets of the r.h.s. of eq.(5) (that is the $PEGW_{tot}(P)$) of each of the proteins in the original dataset, and plotted them versus the l.h.s. of eq. (5) (Figure 2b and Table S2). The correlation coefficient measured between the two data sets is 0.85. Extracting the optimal parameters for the full set of proteins improves the correlation coefficient to 0.87 but the obtained parameters are within a standard deviation from the average provided in Table 1.

Table S2. The proteins included in the data set (name, gene name, pdb id), their average amount in plasma and in the corona of several nanoparticles measured using mass spectroscopy and the estimated PEG/Water ratio(standard deviation) obtained with the bootstrap fit and the simulations.

Protein Name	Gene/Name	Pdb id chain model id	plasma[c]	PS-NH_2[c]	PS-PEG_{44}[c]	PS-PEG_{110}[c]	PS-$PEEP_{49}$[c]	PS-$PEEP_{92}$[c]	PS-LUT[d]	PS-LUT-$NH2$[d]	$PEGW_{tot}$ fit	$PEGW_{tot}$ Sim
Alpha-1-antitrypsin	SERPINA1	3NDD A	6.2	1.115	0.37	0.11	0.543	0.429	0.019	0.021	1.12(1)	1.15(5)
Antithrombin-III	SERPINC1	1E03 L	0.42	0.239	0.04	0.052	0	0	24.791	31.202	1.13(2)	1.12(5)
Apolipoprotein A-I	APOA1	3K2S A	1.14	4.25	35.23	18.29	4.672	1.575	6.985	7.294	1.16(3)	1.27(8)
Apolipoprotein A-II	APOA2	2OU1	19.7[b]	0	0.817	0.079	1.295	1.748	0.458	0.421	1.14(3)	1.22(6)
Apolipoprotein A-IV	APOA4	3S84 A	2.12[b]	1.319	3.93	1.547	8.79	6.917	5.556	7.456	1.18(2)	1.19(6)
Apolipoprotein C-III	APOC3	2JQ3	4.43[b]	4.818	10.94	6.752	8.391	5.873	7.164	5.904	1.15(3)	1.29(9)
Apolipoprotein E	APOE	2L7B	0.266[b]	0.656	2.716	1.017	2.317	1.568	16.904	16.471	1.21(3)	1.26(8)
Beta-2-glycoprotein 1	APOH	1QUB	1.00	3.38	0.438	0.133	0.361	0.193	0.168	0.195	1.17(2)	1.20(6)
Complement C1q subcomponent subunit A	C1QA	2JG8 A	0.013[b]	0.168	1.524	0.3	1.665	2.977	N.A.	N.A.	1.25(2)	1.31(7)
Complement C1q subcomponent subunit B	C1QB	2JG8 B	0.0098[b]	1.625	4.762	1.587	5.977	8.182	N.A.	N.A.	1.27(3)	1.28(7)
Complement C1q subcomponent subunit C	C1QC	2WNV C	0.050[b]	1.724	4.981	1.166	4.738	7.876	0.295	0.174	1.22(2)	1.32(7)
Complement C1r subcomponent	C1R	1GPZ A	0.0152[b]	1.025	1.808	1.126	2.628	2.706	N.A.	N.A.	1.21(2)	1.20(6)
Complement C1s subcomponent	C1s	4J1Y A	0.0192[b]	0.996	1.393	0.949	2.186	2.113	N.A.	N.A.	1.17(3)	1.19(6)
Complement C3	C3	2A73	2.39	1.507	1.852	1.488	1.823	1.622	1.467	1.255	1.14(1)	1.16(5)
Complement C4-A	C4A	4FXG alpha	0.29	0	0.174	0.082	0.312	0.281	0.589	0.581	1.18(1)	1.25(6)
Complement C4-B	C4B	4FXG beta	0.37	0	0.139	0.143	0.246	0.263	N.A.	N.A.	1.16(1)	1.24(6)
Fibrinogen alpha chain	FGA	3GHG alpha	0.91	40.66	3.855	1.861	0.522	0.461	N.A.	N.A.	1.21(2)	1.26(7)
Fibrinogen beta chain	FGB	3GHG beta	1.09	73.86	6.422	2.979	0.807	0.596	N.A.	N.A.	1.19(1)	1.26(7)
Fibrinogen gamma chain	FGG	3GHG gamma	2.12	73.74	6.103	3.075	0.947	0.771	N.A.	N.A.	1.18(2)	1.23(6)
Gelsolin	GSN	3FFN A	0.24	0.129	0.012	0	0	0	0.073	0.066	1.14(2)	1.20(6)
Haptoglobin-related protein	HPR	P00739[a]	0.58	0.024	0.103	0.089	0.125	0.087	0.415	0.478	1.15(2)	1.21(6)
Ig gamma-1 chain C region	IGHG1	1HZH H	51.32	13.84	2.662	1.89	2.583	2.554	0.274	0.235	1.11(1)	1.21(5)
Ig gamma-2 chain C region	IGHG2	P01859[a]	8.58	3.529	2.426	1.662	0.587	0.568	0.074	0.071	1.13(2)	1.22(5)
Ig gamma-3 chain C region	IGHG3	P01860[a]	2.29	1.206	0.642	0.52	1.1	1.1	0.093	0.08	1.12(2)	1.21(5)

Ig gamma-4 chain C region	IGHG4	P01861[a]	5.18	0.924	0.057	0	0.01	0	0.006	0.009	1.11(1)	1.22(5)
Ig heavy chain V-III region BRO	HV305	P01766[a]	1.4	0.23	0.321	0.321	0.204	0.291	0.036	0.027	1.14(3)	1.29(10)
Ig kappa chain C region	IGKC	4XMP L	34.7	12.58	5.084	3.928	5.575	5.03	0.021	0.02	1.10(2)	1.20(5)
Ig kappa chain V-III region SIE	KV302	P01620[a]	1.6	0.333	0.624	0.644	0.46	0.426	N.A.	N.A.	1.12(2)	1.20(6)
Ig kappa chain V-II region TEW	KV204	P01617[a]	1.22	0	0.055	0.248	0.185	0.174	N.A.	N.A.	1.13(2)	1.23(6)
Insulin-like growth factor-binding protein complex acid labile subunit	IGFALS	P35858[a]	0.04	0.015	0.084	0.005	0.506	0.503	0.381	0.392	1.19(3)	1.35(9)
Plasma protease C1 inhibitor	SERPING1	2OAY	0.5	0	0.039	0.131	0.14	0.127	0.117	0.12	1.13(1)	1.21(5)
Protein AMBP	AMBP	4ES7 A	0.37	0.032	0.023	0	0	0	0.051	0.055	1.16(2)	1.18(7)
Serum albumin	ALB	1AO6 A	278.5	129.9	5.532	0.638	0.433	0.369	0.035	0.039	1.10(2)	1.10(5)
Serum paraoxonase/arylesterase 1	PON1	P27169[a]	0.27	0	0.014	0	0.216	0.197	0.157	0.164	1.13(2)	1.21(6)
Transthyretin	TTR	4PVL	2.37	1.352	0.499	0.473	0.106	0	N.A.	N.A.	1.12(2)	1.19(6)
Vitamin D-binding protein	GC	1KW2 A	1.2	1.046	0.077	0	0	0	0.008	0.01	1.08(2)	1.13(5)

[a] Structure from homology model database[12].

[b] Plasma concentration from ref.[11] after conversion to [fmol]. See text for details of conversion.

[c] Values in [fmol] of total protein, data from ref. [9]

[d] Values in mass fraction on NP divided by mass fraction in serum. Data from ref. [10]

Table S3. Fraction of solvent exposed surface areas of the various amino acid types ($FSASA_{AA}$) for each of the proteins in the dataset.

prot\res	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
ALB	0.081	0.018	0.091	0.209	0.019	0.005	0.027	0.005	0.19	0.045	0.009	0.031	0.04	0.045	0.049	0.029	0.055	0.033	0.002	0.016
AMBP	0.029	0.006	0.073	0.137	0.021	0.045	0.042	0.021	0.11	0.022	0.012	0.056	0.032	0.063	0.111	0.038	0.064	0.044	0.047	0.028
APOA1	0.044	0	0.062	0.124	0.031	0.015	0.021	0	0.103	0.163	0.012	0.02	0.031	0.085	0.093	0.042	0.038	0.043	0.029	0.044
APOA2	0.04	0.012	0.029	0.171	0.03	0.019	0	0.019	0.147	0.102	0.009	0.015	0.065	0.085	0	0.047	0.089	0.063	0	0.058
APOA4	0.041	0	0.049	0.16	0.012	0.021	0.032	0.013	0.086	0.107	0.01	0.048	0.035	0.126	0.152	0.022	0.025	0.024	0	0.035
APOC3	0.075	0	0.093	0.075	0.062	0.01	0.015	0	0.095	0.069	0.026	0	0.031	0.074	0.038	0.103	0.057	0.071	0.067	0.039
APOE	0.087	0	0.02	0.189	0.011	0.028	0.014	0.011	0.042	0.083	0.025	0.009	0.043	0.12	0.17	0.028	0.015	0.045	0.05	0.009
APOH	0.018	0.019	0.059	0.103	0.046	0.04	0.024	0.035	0.176	0.052	0.013	0.043	0.084	0.018	0.07	0.032	0.083	0.036	0.006	0.042
C1QA	0.01	0.015	0.025	0.067	0.052	0.05	0.024	0.035	0.061	0.043	0.033	0.054	0.054	0.169	0.091	0.064	0.051	0.058	0.017	0.028
C1QB	0.031	0.012	0.053	0.08	0.057	0.023	0.025	0.03	0.073	0.032	0.032	0.099	0.032	0.063	0.14	0.019	0.083	0.06	0	0.057
C1QC	0.018	0.012	0.064	0.011	0.07	0.045	0.053	0.014	0.082	0.066	0.002	0.047	0.05	0.13	0.044	0.06	0.078	0.1	0.005	0.048
C1R	0.017	0.011	0.069	0.096	0.016	0.03	0.052	0.048	0.1	0.04	0.027	0.064	0.052	0.119	0.11	0.03	0.046	0.026	0.011	0.039
C1s	0.026	0.012	0.07	0.153	0.022	0.042	0.013	0.022	0.113	0.039	0.034	0.059	0.073	0.025	0.091	0.048	0.06	0.057	0.012	0.029
C3	0.024	0.007	0.078	0.139	0.023	0.03	0.021	0.019	0.136	0.058	0.012	0.052	0.057	0.085	0.087	0.054	0.047	0.041	0.006	0.023
C4A	0.049	0.008	0.07	0.11	0.036	0.035	0.026	0.028	0.072	0.088	0.021	0.037	0.055	0.078	0.094	0.059	0.05	0.049	0.019	0.014
C4B	0.032	0.006	0.067	0.067	0.021	0.05	0.033	0.029	0.102	0.094	0.014	0.035	0.076	0.064	0.107	0.082	0.042	0.044	0.005	0.03
FGA	0.025	0.032	0.089	0.062	0.031	0.012	0.019	0.055	0.122	0.092	0.025	0.062	0.03	0.058	0.129	0.055	0.016	0.038	0.02	0.03
FGB	0.02	0.022	0.07	0.077	0.02	0.032	0.013	0.027	0.106	0.083	0.03	0.075	0.036	0.071	0.09	0.055	0.045	0.049	0.033	0.045
FGG	0.031	0.025	0.085	0.081	0.037	0.026	0.018	0.064	0.133	0.073	0.017	0.058	0.032	0.081	0.047	0.044	0.065	0.024	0.014	0.046
GC	0.053	0.023	0.082	0.151	0.023	0.021	0.018	0.008	0.183	0.064	0.005	0.044	0.061	0.037	0.039	0.068	0.067	0.029	0.002	0.024
GSN	0.063	0.002	0.074	0.113	0.02	0.062	0.01	0.011	0.118	0.048	0.009	0.059	0.053	0.084	0.086	0.049	0.06	0.049	0.01	0.019
HPR	0.036	0.015	0.06	0.104	0.018	0.024	0.057	0.023	0.181	0.069	0.008	0.067	0.057	0.058	0.034	0.022	0.031	0.047	0.023	0.066
HV305	0.038	0.001	0.062	0.058	0.009	0.08	0	0.003	0.054	0.1	0.001	0.04	0.015	0.076	0.061	0.089	0.065	0.059	0.038	0.153
IGFALS	0.064	0.021	0.044	0.061	0.036	0.05	0.039	0.01	0.024	0.175	0.003	0.049	0.067	0.062	0.145	0.041	0.02	0.028	0.035	0.026
IGHG1	0.033	0.014	0.05	0.063	0.027	0.037	0.026	0.012	0.134	0.045	0.004	0.057	0.084	0.07	0.052	0.105	0.076	0.042	0.017	0.052
IGHG2	0.026	0.029	0.035	0.091	0.044	0.029	0.028	0.012	0.123	0.036	0.012	0.06	0.099	0.06	0.049	0.103	0.083	0.051	0.002	0.027
IGHG3	0.021	0.042	0.045	0.08	0.03	0.034	0.028	0.015	0.106	0.05	0.01	0.05	0.129	0.051	0.073	0.093	0.083	0.028	0.001	0.03
IGHG4	0.023	0.017	0.034	0.097	0.048	0.035	0.025	0.012	0.123	0.057	0.006	0.056	0.091	0.059	0.043	0.118	0.075	0.039	0.002	0.039
IGKC	0.044	0.017	0.049	0.101	0.033	0.052	0.013	0.027	0.112	0.05	0	0.029	0.036	0.081	0.08	0.144	0.073	0.035	0	0.023

KV204	0.036	0	0.073	0.061	0.04	0.057	0.011	0.036	0.034	0.058	0.006	0.015	0.091	0.074	0.131	0.154	0.044	0.034	0	0.045
KV302	0.021	0.001	0.055	0.056	0.031	0.069	0	0.026	0.044	0.025	0	0.02	0.086	0.087	0.166	0.154	0.077	0.021	0	0.063
PON1	0.019	0.003	0.062	0.153	0.051	0.022	0.03	0.022	0.112	0.078	0.01	0.081	0.067	0.037	0.03	0.057	0.051	0.037	0.02	0.058
SERPINA1	0.022	0	0.089	0.152	0.032	0.024	0.044	0.02	0.153	0.067	0.008	0.06	0.031	0.075	0.048	0.051	0.07	0.037	0.007	0.011
SERPINC1	0.031	0.007	0.086	0.199	0.017	0.026	0.011	0.015	0.139	0.042	0.02	0.056	0.048	0.036	0.108	0.055	0.03	0.047	0.004	0.022
SERPING1	0.03	0.012	0.081	0.075	0.049	0.005	0.037	0.015	0.136	0.066	0.02	0.056	0.065	0.062	0.052	0.097	0.08	0.041	0.003	0.02
TTR	0.066	0.014	0.068	0.154	0.021	0.05	0.04	0.031	0.077	0.045	0.007	0.047	0.066	0	0.061	0.096	0.066	0.052	0.015	0.025

References

1. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J Comput Chem* **26,** 1781 (2005).

2. MacKerell, A. D. *et al.* All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J Phys Chem B* **102,** 3586–3616 (1998).

3. Mackerell, A. D., Feig, M. & Brooks, C. L. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J Comput Chem* **25,** 1400–1415 (2004).

4. Lee, H., Venable, R. M., Mackerell, A. D. & Pastor, R. W. Molecular dynamics studies of polyethylene oxide and polyethylene glycol: hydrodynamic radius and shape anisotropy. *Biophys J* **95,** 1590–1599 (2008).

5. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J Chem Phys* **79,** 926–935 (1983).

6. Martyna, G. J., Tobias, D. J. & Klein, M. L. Constant pressure molecular dynamics algorithms. *J Chem Phys* **101,** 4177–4189 (1994).

7. Feller, S. E., Zhang, Y., Pastor, R. W. & Brooks, B. R. Constant pressure molecular dynamics simulation: The Langevin piston method. *J Chem Phys* **103,** 4613–4621 (1995).

8. Essmann, U. *et al.* A smooth particle mesh Ewald method. *J. Chem. Phys.* **103,** 8577–8593 (1995).

9. Schöttler, S. *et al.* Protein adsorption is required for stealth effect of poly(ethylene glycol)- and poly(phosphoester)-coated nanocarriers. *Nat. Nanotechnol.* **11,** 372–377 (2016).

10. Mohr, K. Aggregation Behavior of Polystyrene-Nanoparticles in Human Blood Serum and its Impact on the in vivo Distribution in Mice. *J. Nanomedicine Nanotechnol.* **05,** (2014).

11. Nanjappa, V. *et al.* Plasma Proteome Database as a resource for proteomics research: 2014 update. *Nucleic Acids Res.* **42,** D959–965 (2014).

12. Haas, J. *et al.* The Protein Model Portal–a comprehensive resource for protein structure and model information. *Database Oxf.* **2013,** bat031 (2013).

Supplementary Figures

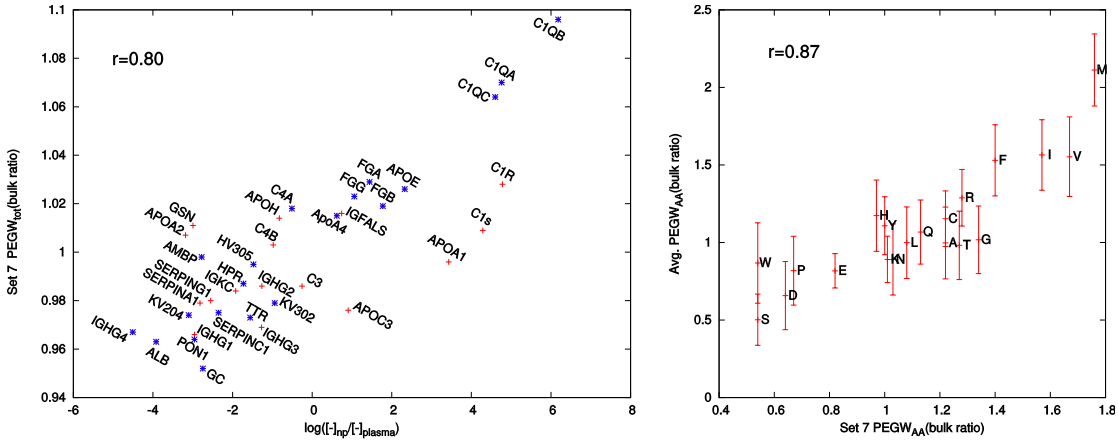

Figure S1 Example of the results from one of the 100 resampled bootstrap subsets. (left) Only the 20 proteins indicated with a blue dot were included in the resampled set to fit the r.h.s. of eq. (5) to the respective experimental values on the l.h.s. The PEG/water ratios for the proteins indicated with a red dot were predicted using the $PEGW_{AA}$ parameters fitted to the blue data. The overall correlation (upper left corner) remains good. (right) the $PEGW_{AA}$ obtained by fitting the resampled set are plotted versus the average $PEGW_{AA}$ over all the resampled set (Table 1 in manuscript). Apart from an irrelevant scaling factor the correlation (upper left corner) is high.

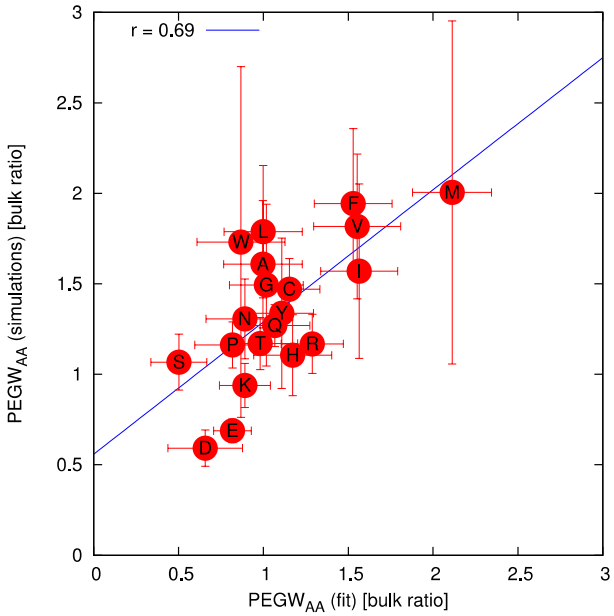

Figure S2 The residue specific affinities for PEG ($PEGW_{AA}$) obtained from the fit are plotted vs the values obtained from the simulations. The correlation coefficient is reported in the top left corner of the figure. The blue line is the best linear fit.

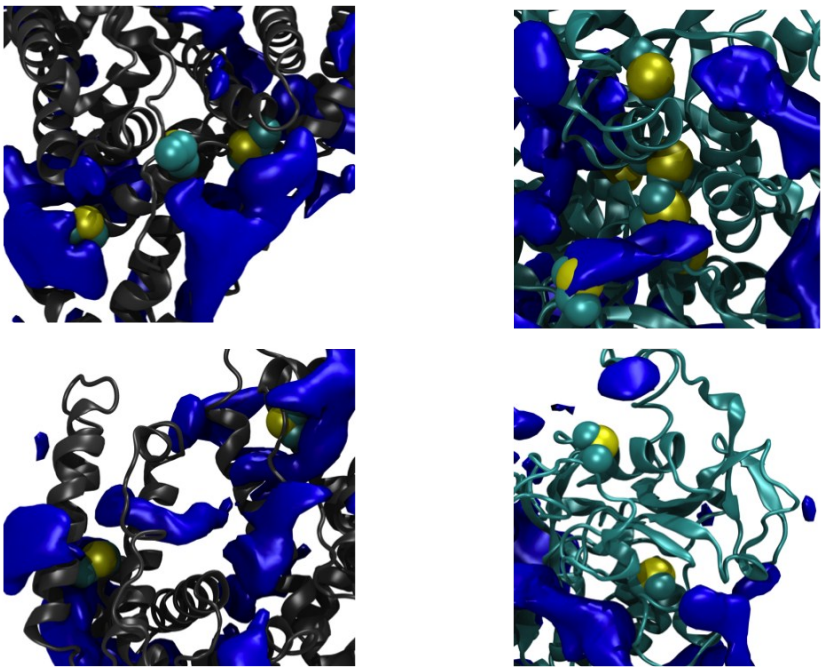

Figure S3 Cartoons highlighting the presence of high density PEG regions (dark blue) around methionine residues (cyan and yellow balls) in HSA (left) and TRF (right)