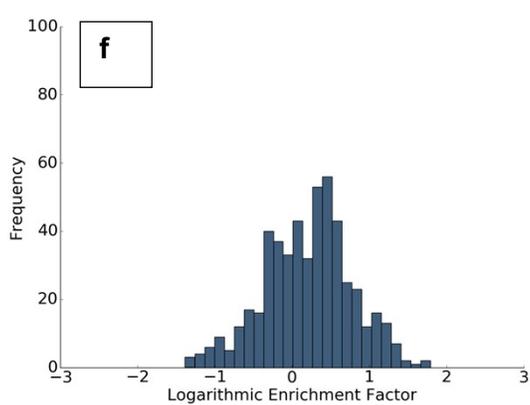
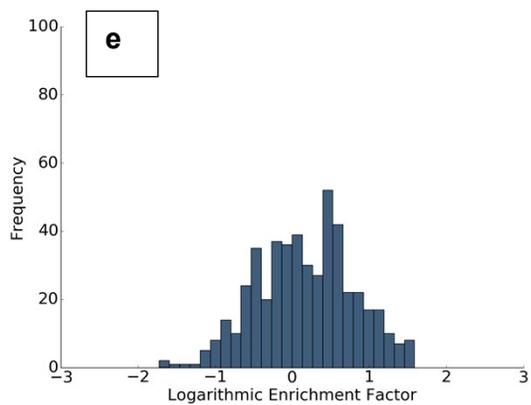
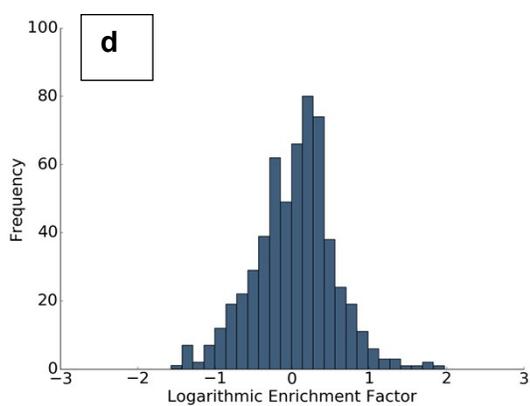
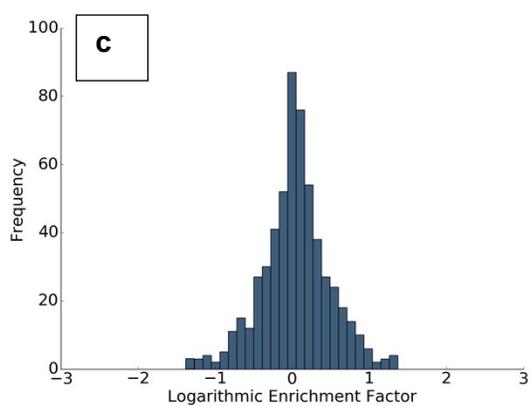
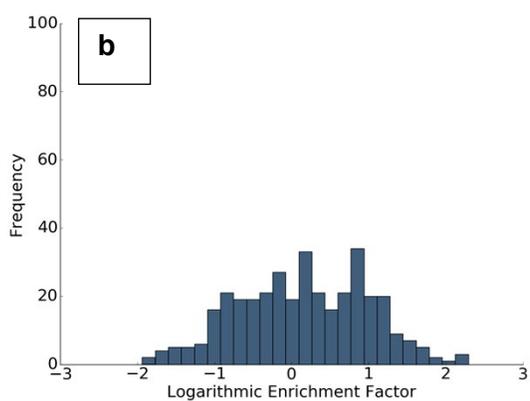
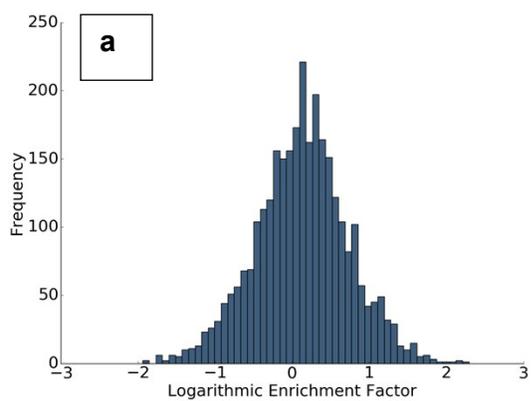
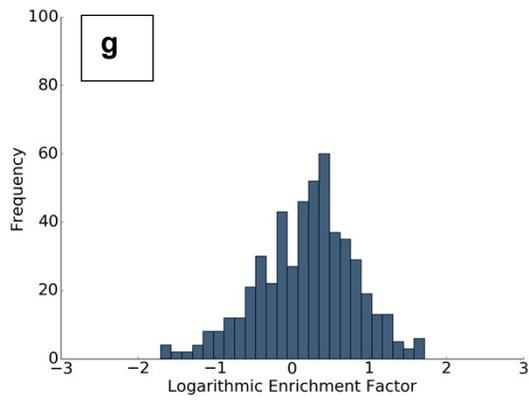


## S.I. Results

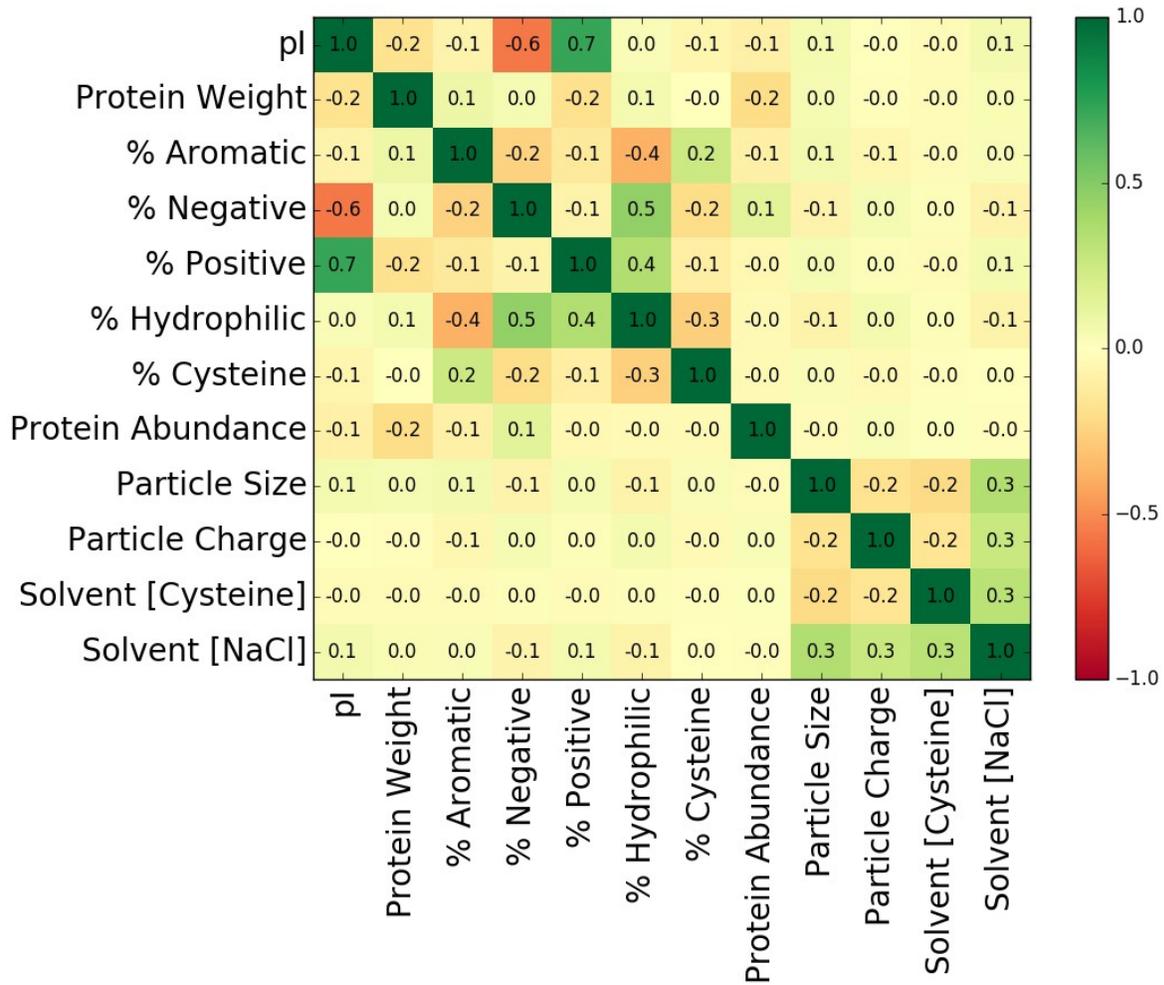




**Figure S.I.1 Histogram of logarithmic enrichment factors for all proteins across the range of ENMs and solution conditions within the database.** Each histogram has with 50 bins. The enrichment from all datasets is shown in **(a)** (3012 enrichment values). The enrichment distribution for each individual sample is also presented as follows : **(b)** cationic branched polyethyleneimine coated 10 nm Ag ENMs in 10 mM sodium phosphate pH 7.4, (356 enrichment values) **(c)** anionic citrate coated 10 nm Ag ENMs in 10 mM sodium phosphate pH 7.4, (568 enrichment values) **(d)** anionic citrate coated 100 nm Ag ENMs in 10 mM sodium phosphate pH 7.4, (578 enrichment values) **(e)** anionic citrate coated 10 nm Ag ENMs in 10 mM sodium phosphate pH 7.4 with 0.1 mM cysteine, (487 enrichment values) **(f)** anionic citrate coated 10 nm Ag ENMs in 10 mM sodium phosphate pH 7.4 with 0.8 mM sodium chloride, (510 enrichment values) and **(g)** anionic citrate coated 10 nm Ag ENMs in 10 mM sodium phosphate pH 7.4 with 3.0 mM sodium chloride (513 enrichment values).

Extended commentary on Figure S.I.1.

The enrichment data from all datasets **(a)** shows that, overall, the protein enrichment data forms a Gaussian distribution. The histogram for Ag ENMs with positively **(b)** and negatively **(c)** charged surface functionalizations are strikingly different (without the addition of NaCl or cys). The Ag ENMs with a positive surface charge are characterized by a platykurtic trend of enrichment, as compared to a Gaussian distribution for those with negative surface coating. Yet, when solvent conditions are changed by adding NaCl or cys **((d)-(g))**, the distribution of the negatively charged Ag ENMs undergoes various degrees of skewness and negative kurtosis similar to that of the positively charged Ag ENMs **(b)**. In addition, to the above-stated deviations from Normality, the widening tails of the distribution is indicative of rising enrichment in PC or solution.



**Figure S.I.2 A** Pearson correlation plot of all the independent features used to make predictions. The percentage sign represents protein amino acid composition with respect to the modeled feature.

### S.I.3. Comparison of modeling approaches

Two other well-established algorithms were employed to provide a comparative for the RFC approach. Logistic regression (LR) and support vector machines (SVM) were chosen because of their extensive use in the fields of artificial intelligence and biostatistics.<sup>1,2,3</sup> The SVM and LR algorithms were trained and tested with the same method as the random forests. The F1-score for these methods are 0.80 and 0.75, respectively. A lower F1-score for LR implies that nonlinear relationships are present in the PC data. Although SVM performed similarly to RFC, we still prioritize use of RFC in this context because it includes intuitive decision rules and is more readily accessible. It is possible, however, that future work with an expanded dataset, including significantly more features and observations, will perform better with the SVM algorithm.

#### References

- (1) Yang, Z. R. Biological Applications of Support Vector Machines. *Brief. Bioinform.* **2004**, 5 (4), 328–338.
- (2) Bagley, S. C.; White, H.; Golomb, B. A. Logistic Regression in the Medical Literature:: Standards for Use and Reporting, with Particular Attention to One Medical Domain. *J. Clin. Epidemiol.* **2001**, 54 (10), 979–985.
- (3) Brandes, N.; Ofer, D.; Linial, M. ASAP: A Machine Learning Framework for Local Protein Properties. *Database* **2016**, 2016, baw133.

#### **S.I.4. Links to additional information**

**Link to the database:**

[<https://github.com/mfindlay23/ENM-Protein-Predictor/blob/master/Input\\_Files/database.csv>](https://github.com/mfindlay23/ENM-Protein-Predictor/blob/master/Input_Files/database.csv)

**Link to the Y-randomization test results:**

[<https://github.com/mfindlay23/ENM-Protein-Predictor/blob/master/Output\\_Files/y\\_randomization.csv>](https://github.com/mfindlay23/ENM-Protein-Predictor/blob/master/Output_Files/y_randomization.csv)

**Link to the prediction results:**

[<https://github.com/mfindlay23/ENM-Protein-Predictor/blob/master/Output\\_Files/prediction\\_information.csv>](https://github.com/mfindlay23/ENM-Protein-Predictor/blob/master/Output_Files/prediction_information.csv)