Yeast proteins reversibly aggregate like amphiphilic molecules

Pouria Dasmeh[1,3,4], Andreas Wagner[1,2,4]

[1]Institute for Evolutionary Biology and Environmental Studies, University of Zurich, Zurich, Switzerland. [2]The Santa Fe Institute, Santa Fe, New Mexico, [3]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA USA 02139. [4]Swiss Institute of Bioinformatics (SIB).

Abstract:

More than a hundred proteins in yeast reversibly aggregate and phase-separate in response to various stressors, such as nutrient depletion and heat shock. We know little about the sequence and structural features behind this ability, which has not been characterized on a proteome-wide level. To identify the distinctive features of aggregation-prone regions, we apply machine learning algorithms to genome-scale limited proteolysis-mass spectrometry data from 96 yeast proteins that phase-separate upon heat shock. We find that the aggregation-prone regions (APRs) of our study proteins are significantly enriched in aliphatic residues and depleted in positively charged amino acids. Aggregator proteins with longer APRs show a greater propensity to aggregate, a relationship that can be explained by equilibrium statistical thermodynamics. Altogether, our observations suggest that proteome-wide reversible protein aggregation is mediated by sequence-encoded properties. Aggregating proteins resemble supra-molecular amphiphiles, where APRs are the hydrophobic parts, and non-APRs are the hydrophilic parts.

Body:

Proteins can aggregate reversibly or irreversibly. Irreversible aggregation is often pathological, indicates damaged cellular regulation[1-3], and is involved in multiple diseases such as Alzheimer's and Parkinson's diseases[4,5]. Reversible aggregation, however, can be beneficial and help cells survive stressors. For example, in yeast more than a hundred proteins in multiple subcellular compartments form reversible aggregates in response to nutrient starvation, heat shock, or chemical stress[6]. These aggregated proteins are not misfolded or tagged for degradation[7]. Instead, they help increase cells re-initiate growth during recovery from stress by protecting metabolic enzymes from degradation[6-9].

Reversible protein aggregation is a special case of a widespread phenomenon called protein phase-separation. In this process, a well-mixed solution of proteins de-mixes into two phases of high and low densities[10]. Proteins can phase separate with other proteins, but also with RNA or DNA molecules, into biomolecular condensates that regulate transcription[11], chromatin states[12,13], and RNA metabolism[14]. We know little about the sequence and structural features behind reversible phase separation, which has not been characterized on a proteome-wide level. In this work, we provide such a view and characterize sequence features that can help predict whether proteins will aggregate reversibly.

We took advantage of recent proteome-wide data from limited proteolysis-mass spectrometry (LiP-MS)[20,21]. This method permits the detection of both pronounced and subtle changes that protein and peptide structures experience in response to stressors such as heat and osmotic shock. Recently, Cappelletti *et. al.* characterized structural

60 changes within 96 proteins that reversibly aggregate upon heat shock in yeast[22]. These
61 proteins are also known as aggregators. They have diverse biological functions, such as
62 telomeric DNA-binding, RNA-helicase activity, or ribosome assembly. Because they are
63 functionally diverse, they constitute an excellent dataset to examine sequence features
64 that facilitate reversible protein aggregation. We studied regions within these proteins
65 whose proteolysis resistance changes significantly upon aggregation [22]. We refer to these
66 regions as aggregation-prone regions.

67 What are the characteristic sequence features of APRs? To answer this question,
68 we first performed a principal component analysis on the frequency of 20 amino acids in
69 all 270 APRs of yeast aggregators (Dataset S1). As shown in Figure 1A-B, the first and
70 the second principal components only accounted for ~12% and ~10% of the variation.
71 The PCA biplot for these components shows that no specific combinations of few amino
72 acids preferentially occur in APR sequences (Figure 1B), a further indication that APRs
73 do not show the kind of reduced sequence complexity characteristic of disordered
74 regions. We also looked for linear sequence motifs in APRs using DALEL[29], an algorithm
75 for the exhaustive identification of degenerate sequence motifs, but found no such
76 enriched motifs.

77 We then compared the frequencies of amino acids, dipeptides, and tripeptides in
78 APRs with the rest of the protein sequence in aggregators (Dataset S2). Here, we found
79 a significant depletion of the positively charged residues Arg and Lys, and their dipeptides
80 ($p \sim 10^{-16}$; Wilcoxon signed-rank test). We also observed that aliphatic residues Leu, Ile,
81 Ala, Val, were significantly more frequent in APRs compared to the rest of the protein

82 sequence ($p \sim 10^{-16}$; Wilcoxon signed-rank test). The fraction of positively charged to
83 aliphatic residues was the strongest predictor of APRs compared to the fraction of all
84 other amino acids (Figure 1C; $p \sim 10^{-21}$; Wilcoxon signed-rank test).

85 To find out whether these patterns are linked to differences in physicochemical
86 properties of amino acids, we used a random forest approach, a widely-used machine
87 learning technique for the classification of two or more data sets[30]. Specifically, we used
88 all 96 aggregators with experimentally characterized APRs[22], and subdivided each
89 protein into two sequence data sets, one comprising the APRs and one comprising the
90 non-APRs. For each of these datasets, we calculated a feature matrix that consisted of
91 500 amino acid properties (Dataset S3, and S4). We took these properties from the
92 AAindex database[31], which curates various physicochemical and biochemical properties
93 of amino acids. The classifier achieved an accuracy of ~93% in 100 independent runs,
94 with the data split into a training set (80% of the data) and a testing set (20%) (Figure 1D;
95 Dataset S5). Two amino acid properties were most important in this classification (Figure
96 1E). The first is a low positive charge of amino acids in APRs compared to non-APRs.
97 The second is a high amphiphilic propensity of these amino acids. The amphiphilic
98 propensity measures the preference of amino acids to occur at protein-solvent interfaces,
99 such as the end regions of transmembrane helices[32]. Amino acids like arginine and lysine
100 have the highest preference for these environments, and the largest value of the
101 amphiphilic index. In contrast, hydrophobic amino acids such as leucine, isoleucine,
102 valine, and alanine prefer internal protein environments and have the lowest amphiphilic
103 propensity[32].

104 Because of the statistically significant difference between the amphiphilic nature
105 of APRs and non-APRs, we argue that aggregator proteins resemble supramolecular
106 amphiphiles. These molecular structures have distinct hydrophobic and hydrophilic
107 parts[33]. In response to external stimuli such as temperature, pH, and ionic strength[33],
108 they can reversibly aggregate by non-covalent bonding forces, e.g., through electrostatic
109 and hydrophobic interactions.

110 We next investigated the relevance of amphiphilic aggregation to stress-induced
111 phase separation of our yeast aggregator proteins. An important feature of amphiphilic
112 molecules is that they are more likely to occur in an aggregated phase if their hydrophobic
113 chain is long[34]. We thus predicted that aggregator proteins with longer APRs will occur
114 preferentially in the aggregated phase. To test this prediction, we asked whether
115 aggregator proteins with long APRs preferentially occur in the pellet fraction of proteins
116 extracted from heat-stressed yeast cells[35]. Specifically, we used for this purpose the $\log_2$
117 ratio of the protein's abundance in the pellet fraction of yeast cells to that of supernatant
118 as a measure of protein's enrichment in the aggregated phase[35]. We found this
119 enrichment data for 31 aggregated proteins in our dataset. Indeed, aggregator proteins
120 with longer APRs preferentially accumulate in the aggregated pellet fraction (Figure 2A,
121 R = 0.53, p = 0.0021; Spearman's rank correlation). This association increases to R =
122 0.64 for aggregator proteins whose APRs have a low positive charge-to-aliphatic bias (p=
123 0.012; Spearman's rank correlation). To see whether this observation extends to other
124 stress-induced conditions, we also used the mass spectrometry dataset of nutrient-
125 starved yeast cells reported by Narayanasamy et al[36]. We identified 56 aggregator

126 proteins in this dataset and checked the association between APR lengths in these
127 proteins and the Z-score of proteins' enrichment in the pellet fraction. In nutrient-starved
128 cells too, the aggregator proteins with longer APRs preferentially accumulate in the
129 aggregated pellet fraction (Figures 2B, R = 0.47, p = 0.00014; Spearman's rank
130 correlation), and the association increased for the low positive charge-to-aliphatic bias (R
131 = 0.53, p = 0.0022; Spearman's rank correlation).

132 We then investigated the relationship between APR length in aggregator proteins
133 and their differential enrichment in the aggregated phase using equilibrium statistical
134 thermodynamics. We considered a two-state model where proteins exist either as a
135 soluble monomer or an insoluble aggregate. The fraction of proteins occurring in the
136 aggregate at the critical concentration (cc) for aggregation ($\mathcal{F}_{cc}$) is expressed as:

137 $$\mathcal{F}_{cc}(theory) = \frac{(P_{agg})_{cc}}{(P_{agg})_{cc}+(P_{mon})_{cc}} = \frac{1}{1+e^{-\beta\Delta G_{agg}}} \quad \text{(Eq. 1)}$$

138 Here, P_{agg} and P_{mon} are the relative concentrations of proteins in the aggregated and the
139 monomer phase, $\beta = 1/k_B T$ where k_B is the Boltzmann's constant, and ΔG_{agg} is the free
140 energy change upon aggregation. Because the aggregation free energy increases in
141 proportion to the length of the hydrophobic chain[37], we rearranged Equation 1 to:

142 $$\mathcal{F}_{cc}(theory) = \frac{1}{1+e^{-\beta.B.L}} \quad \text{(Eq. 2)}$$

143 Where B is the proportionality constant between the aggregation free energy and the APR
144 length. We then calculated the observed fraction of the aggregator proteins in the
145 aggregated phase from their reported $\log_2$ enrichment ratio in the pellet fraction of heat-

146 stressed yeast ($log_2 \frac{[Pellet]}{[Supernatant]}$) as:

147 $$\mathcal{F}_{cc}(experiment) = (1 + 2^{-log_2 \frac{[Pellet]}{[Supernatant]}})^{-1} \quad \text{(Eq. 3)}$$

148 Equation 2 predicted the observed relationship between APR length and a protein's
149 fraction in the aggregated phase well (Figure 2C, R=0.53, p=0.0021; Spearman's rank
150 correlation, see supplementary information for the fitted constants). Importantly, the
151 predictive power of Equation 2 was higher for proteins whose APRs had a low positive
152 charge-to-aliphatic bias (Figure 2C; R=0.64, p=0.012; Spearman's rank correlation).

153 In summary, our observations suggest that the aggregation-prone regions within
154 the aggregator proteins can readily unfold in response to stressors, form hydrophobic
155 long chains, and phase-separate in a manner akin to amphiphilic molecules. We also
156 observed that aggregator proteins with longer APRs are more prone to aggregation. This
157 relationship also suggests that aggregator proteins may leave the aggregated phase by
158 masking their APRs from self-assembly or from interaction with other proteins. Molecular
159 chaperones can facilitate this step: They have been co-purified with stress-induced
160 aggregated proteins and can solubilize them[39]. Altogether, our study demonstrates that
161 stress-induced phase separation is a sequence-encoded phenotype. It can be both
162 explained and predicted by the self-assembly and aggregation of amphiphilic molecules.

163

164

165 **Methods**:

166 We identified aggregation-prone regions in the 96 aggregator proteins from the dataset
167 of Cappelletti et. al[22], and downloaded their sequences from the Uniprot database[40]
168 (Dataset S6). We chose 500 amino acid properties from the AAindex database to
169 distinguish the physicochemical properties of APRs from non-APR segments in the
170 aggregator protein[31].

171 We used a binary classification to group regions within an aggregator protein into
172 aggregation-prone regions (class A) and the remainder of the protein (class B). More
173 specifically, we used a random forest classification, splitting aggregator proteins into a
174 random subset comprising 80% of aggregator proteins for training and 20% for testing.
175 To build features for the classification, we calculated the average value of 500
176 physicochemical properties for each sequence in the APR and non-APR dataset. This
177 yielded two feature matrices for the APR and non-APR sequences. To apply random
178 forest classification, we used the randomForest package in R[42], and evaluated the best
179 number of trees (*nTree*) and the number of variables randomly sampled at each split
180 (*mtry*), in the random forest algorithm. To do so, we systematically varied the *nTree* and
181 *mtry*, and calculated the accuracy of classification with 10-fold cross-validation and 3
182 repeats. We defined accuracy as the percentage of correctly identified classes of proteins
183 (APRs and non-APRs) out of all instances. The combination of *nTree*=5000 trees and
184 *mtry*=10 variables achieved the highest accuracy of ~ 90%. We then used these
185 parameters to perform 100 random forest clusterings, in which we randomly assigned
186 proteins to the training and the testing datasets. To quantify the accuracy of classification

187 we counted the number of true positive and false positive predictions and calculated the
188 area under the curve (AUC). These values are shown as a receiver operating
189 characteristic curve (ROC) in Figure 2D. The most important physicochemical properties
190 were the ones whose Gini index in the classification decreased the most compared to all
191 other properties. This index is a widely-used measure of dispersion that reflects inequality
192 in the values of a frequency distribution. Within the random forest framework, this index
193 is calculated as the average probability that each of 500 physicochemical properties
194 wrongly classifies APRs and non-APRs. We performed all models and statistical analyses
195 using R. Scripts are available at: https://github.com/dasmeh/yeast_aggregators

196 **References**

197 1. Aguzzi, A. & Calella, A.M. Prions: protein aggregation and infectious diseases.
198 *Physiological reviews* **89**, 1105-1152 (2009).
199 2. Aguzzi, A. & O'connor, T. Protein aggregation diseases: pathogenicity and therapeutic
200 perspectives. *Nature reviews Drug discovery* **9**, 237-248 (2010).
201 3. Ross, C.A. & Poirier, M.A. Protein aggregation and neurodegenerative disease. *Nature*
202 *medicine* **10**, S10-S17 (2004).
203 4. Scheibel, T. & Buchner, J. Protein aggregation as a cause for disease. *Molecular*
204 *chaperones in health and disease*, 199-219 (2006).
205 5. Shulman, J.M., De Jager, P.L. & Feany, M.B. Parkinson's disease: genetics and
206 pathogenesis. *Annual Review of Pathology: Mechanisms of Disease* **6**, 193-222 (2011).
207 6. O'Connell, J.D. *et al.* A proteomic survey of widespread protein aggregation in yeast.
208 *Molecular BioSystems* **10**, 851-861 (2014).
209 7. Wallace, E.W. *et al.* Reversible, specific, active aggregates of endogenous proteins
210 assemble upon heat stress. *Cell* **162**, 1286-1298 (2015).
211 8. Saad, S. *et al.* Reversible protein aggregation is a protective mechanism to ensure cell
212 cycle restart after stress. *Nature cell biology* **19**, 1202-1213 (2017).
213 9. Cereghetti, G., Saad, S., Dechant, R. & Peter, M. Reversible, functional amyloids: towards
214 an understanding of their regulation in yeast and humans. *Cell Cycle* **17**, 1545-1558
215 (2018).
216 10. Hyman, A.A., Weber, C.A. & Jülicher, F. Liquid-liquid phase separation in biology. *Annual*
217 *review of cell and developmental biology* **30**, 39-58 (2014).

11. Hnisz, D., Shrinivas, K., Young, R.A., Chakraborty, A.K. & Sharp, P.A. A phase separation model for transcriptional control. *Cell* **169**, 13-23 (2017).
12. Erdel, F. & Rippe, K. Formation of chromatin subcompartments by phase separation. *Biophysical journal* **114**, 2262-2270 (2018).
13. Valsecchi, C.I.K. *et al.* RNA nucleation by MSL2 induces selective X chromosome compartmentalization. *Nature* **589**, 137-142 (2021).
14. Youn, J.-Y. *et al.* Properties of Stress Granule and P-Body Proteomes. *Molecular Cell* **76**, 286-294 (2019).
15. Uversky, V.N. Intrinsically disordered proteins in overcrowded milieu: membrane-less organelles, phase separation, and intrinsic disorder. *Current opinion in structural biology* **44**, 18-30 (2017).
16. Posey, A.E., Holehouse, A.S. & Pappu, R.V. Phase separation of intrinsically disordered proteins. in *Methods in Enzymology*, Vol. 611 1-30 (Elsevier, 2018).
17. Kuechler, E.R., Budzyńska, P.M., Bernardini, J.P., Gsponer, J. & Mayor, T. Distinct features of stress granule proteins predict localization in Membraneless organelles. *Journal of molecular biology* **432**, 2349-2368 (2020).
18. Riback, J.A. *et al.* Stress-triggered phase separation is an adaptive, evolutionarily tuned response. *Cell* **168**, 1028-1040. e19 (2017).
19. Iserman, C. *et al.* Condensation of Ded1p Promotes a Translational Switch from Housekeeping to Stress Protein Production. *Cell* (2020).
20. Gao, X. *et al.* High-throughput limited proteolysis/mass spectrometry for protein domain elucidation. *Journal of structural and functional genomics* **6**, 129-134 (2005).
21. Schopper, S. *et al.* Measuring protein structural changes on a proteome-wide scale using limited proteolysis-coupled mass spectrometry. *Nature protocols* **12**, 2391 (2017).
22. Cappelletti, V. *et al.* Dynamic 3D proteomes reveal protein functional alterations at high resolution in situ. *Cell* **184**, 545-559. e22 (2021).
23. Dosztányi, Z. Prediction of protein disorder based on IUPred. *Protein Science* **27**, 331-340 (2018).
24. Bonferroni, C.E., Bonferroni, C. & Bonferroni, C. Teoria statistica delle classi e calcolo delle probabilita'. (1936).
25. Sugimoto, K., Matsumoto, K., Kornberg, R.D., Reed, S.I. & Wittenberg, C. Dosage suppressors of the dominant G1 cyclin mutant CLN3-2: identification of a yeast gene encoding a putative RNA/ssDNA binding protein. *Molecular and General Genetics MGG* **248**, 712-718 (1995).
26. Altmeyer, M. *et al.* Liquid demixing of intrinsically disordered proteins is seeded by poly (ADP-ribose). *Nature communications* **6**, 8088 (2015).
27. Jamieson, D.J., Rahe, B., Pringle, J. & Beggs, J.D. A suppressor of a yeast splicing mutation (prp8-I) encodes a putative ATP-dependent RNA helicase. *Nature* **349**, 715-717 (1991).
28. Schäfer, I.B. *et al.* Molecular basis for poly (A) RNP architecture and recognition by the Pan2-Pan3 deadenylase. *Cell* **177**, 1619-1631. e21 (2019).
29. Kelil, A., Dubreuil, B., Levy, E.D. & Michnick, S.W. Exhaustive search of linear information encoding protein-peptide recognition. *PLoS computational biology* **13**, e1005499 (2017).

30. Ho, T.K. Random decision forests. in *Proceedings of 3rd international conference on document analysis and recognition* Vol. 1 278-282 (IEEE, 1995).
31. Kawashima, S. & Kanehisa, M. AAindex: amino acid index database. *Nucleic acids research* **28**, 374-374 (2000).
32. Mitaku, S., Hirokawa, T. & Tsuji, T. Amphiphilicity index of polar amino acids as an aid in the characterization of amino acid preference at membrane–water interfaces. *Bioinformatics* **18**, 608-616 (2002).
33. Zhang, X. & Wang, C. Supramolecular amphiphiles. *Chemical Society Reviews* **40**, 94-101 (2011).
34. Shaul, B. & Gelbart, W. Theory of chain packing in amphiphilic aggregates. *Annual Review of Physical Chemistry* **36**, 179-211 (1985).
35. Zhu, M. *et al.* Proteomic analysis reveals the direct recruitment of intrinsically disordered regions to stress granules in S. cerevisiae. *Journal of Cell Science* **133**(2020).
36. Narayanaswamy, R. *et al.* Widespread reorganization of metabolic enzymes into reversible assemblies upon nutrient starvation. *Proceedings of the National Academy of Sciences* **106**, 10147-10152 (2009).
37. Budin, I., Prywes, N., Zhang, N. & Szostak, J.W. Chain-length heterogeneity allows for the assembly of fatty acid vesicles in dilute solutions. *Biophysical journal* **107**, 1582-1590 (2014).
38. Iserman, C. *et al.* Condensation of Ded1p promotes a translational switch from housekeeping to stress protein production. *Cell* **181**, 818-831. e19 (2020).
39. Cherkasov, V. *et al.* Coordination of translational control and protein homeostasis during severe heat stress. *Current biology* **23**, 2452-2462 (2013).
40. Consortium, U. UniProt: a hub for protein information. *Nucleic acids research* **43**, D204-D212 (2015).
41. Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N. & Sternberg, M.J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature protocols* **10**, 845-858 (2015).
42. Liaw, A. & Wiener, M. Classification and regression by randomForest. *R news* **2**, 18-22 (2002).
43. Raudvere, U. *et al.* g: Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic acids research* **47**, W191-W198 (2019).

295 **Figure captions:**
296
297 **Figure 1. Sequence composition and physicochemical properties of aggregation-**
298 **prone regions in the yeast aggregators.** A) The percentage of explained variance for
299 10 dimensions from a principal component analysis on the frequencies of 20 amino acids
300 in 270 aggregation-prone regions of 96 yeast proteins. B) Biplot for the first and the
301 second principal components. Grey circles, labeled by numbers from 1 to 270, represent
302 aggregation-prone regions. Red arrows represent the weight of different amino acid
303 frequencies. C) The amino acid frequencies in APRs (shown in red) versus the rest of the
304 protein sequence (shown in green) in 96 yeast aggregators. D) The receiver-operating
305 characteristic (ROC) curves for the clustering of APRs from non-APRs using random
306 forest classification. ROC curves in grey are from 100 random forest clustering and their
307 average is shown in black. E) Ranked importance of physicochemical vartiables
308 quantified as the average decrease in the Gini index for different physicochemical
309 properties.
310
311 **Figure 2. Aggregator proteins reversibly aggregate in a manner akin to amphiphilic**
312 **molecules.**
313 A) The log2 ratio of the protein's abundance in the pellet fraction to that of supernatant in
314 heat-stressed yeast cells (30 out of 96 aggregator proteins[35]) versus their APR length. B)
315 The aggregator proteins' enrichment in the pellet fraction upon nutrient starvation (59 out
316 of 96 proteins[35]) versus their APR length in log-scale. The red and the grey circles in both
317 panels A and B represent all aggregator proteins, and those where the positive charge-

318 to-aliphatic-bias in APRs was lower than its median value ~ 0.25. C) The fraction of
319 aggregator proteins in the aggregated phase calculated from Equation 3 versus the length
320 of APRs. D) The fraction of aggregator proteins with a low positive charge-to-aliphatic-
321 bias in their APRs versus the length of APRs calculated from Equation 3. For both panels
322 C and D, the fitted curve to the observed fraction of proteins in the aggregated phase
323 using Equation 2 is shown in red. R values are the Spearman's rank correlation between
324 the expected (Equation 2) and observed (Equation 3) fractions of proteins in the
325 aggregated phase.
326
327
328
329
330

Figure 1

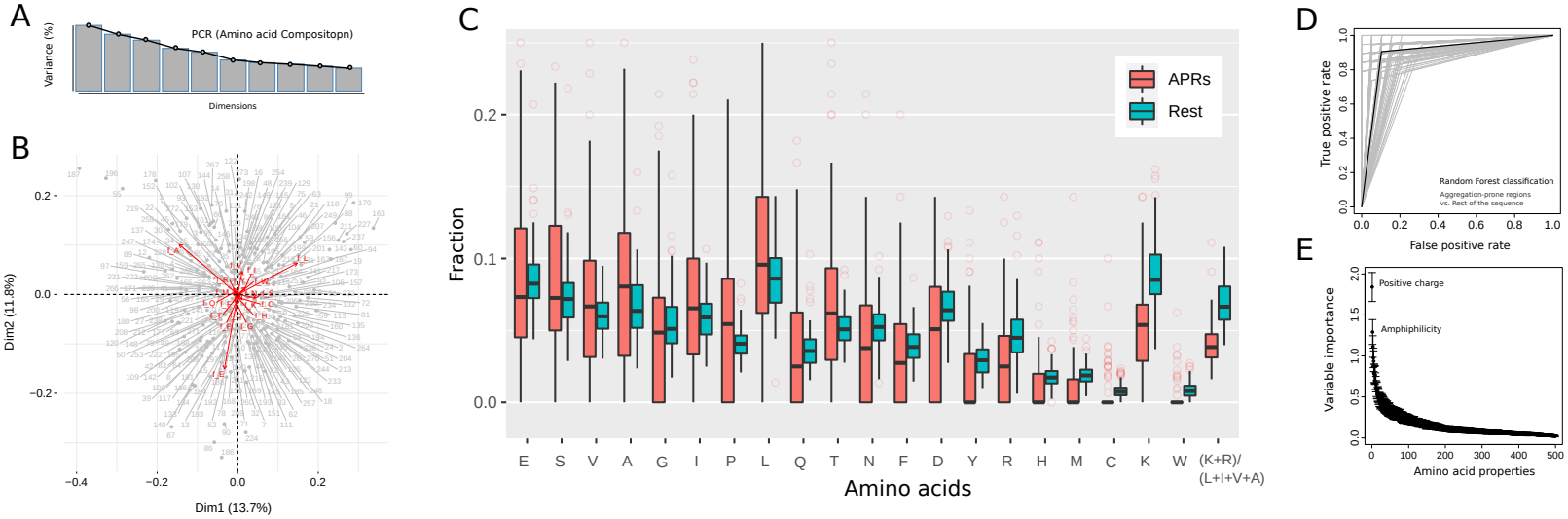

A
PCR (Amino acid Compositopn)
Variance (%)
Dimensions
B
Dim1 (13.7%)
Dim2 (11.8%)
C
APRs
Rest
Fraction
Amino acids
E S V A G I P L Q T N F D Y R H M C K W (K+R)/ (L+I+V+A)
D
True positive rate
False positive rate
Random Forest classification
Aggregation-prone regions vs. Rest of the sequence
E
Positive charge
Amphiphilicity
Variable importance
Amino acid properties

Figure 2

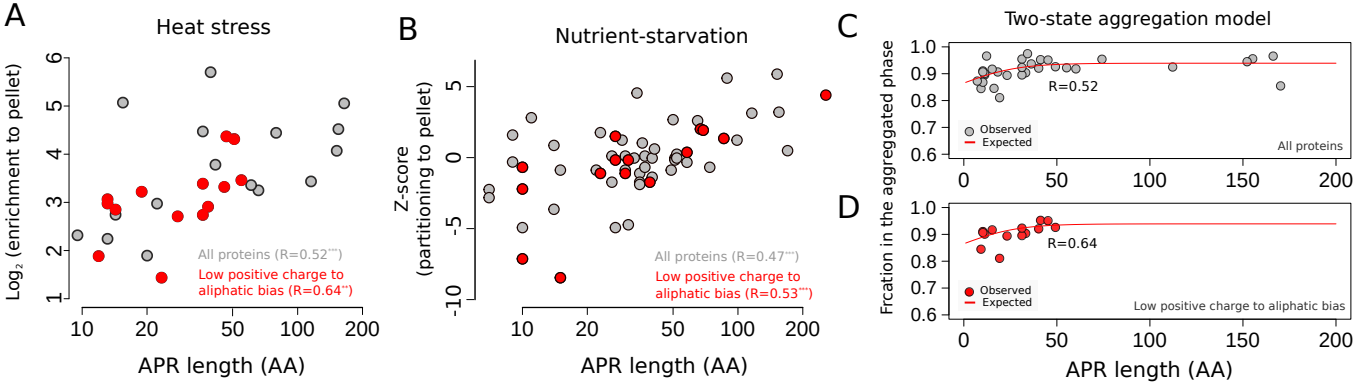

A
Heat stress
Log2 (enrichment to pellet)
All proteins (R=0.52***)
Low positive charge to aliphatic bias (R=0.64**)
APR length (AA)
B
Nutrient-starvation
Z-score (partitioning to pellet)
All proteins (R=0.47***)
Low positive charge to aliphatic bias (R=0.53***)
APR length (AA)
C
Two-state aggregation model
Frcation in the aggreggated phase
R=0.52
Observed
Expected
All proteins
D
R=0.64
Observed
Expected
Low positive charge to aliphatic bias
APR length (AA)

Figure 1

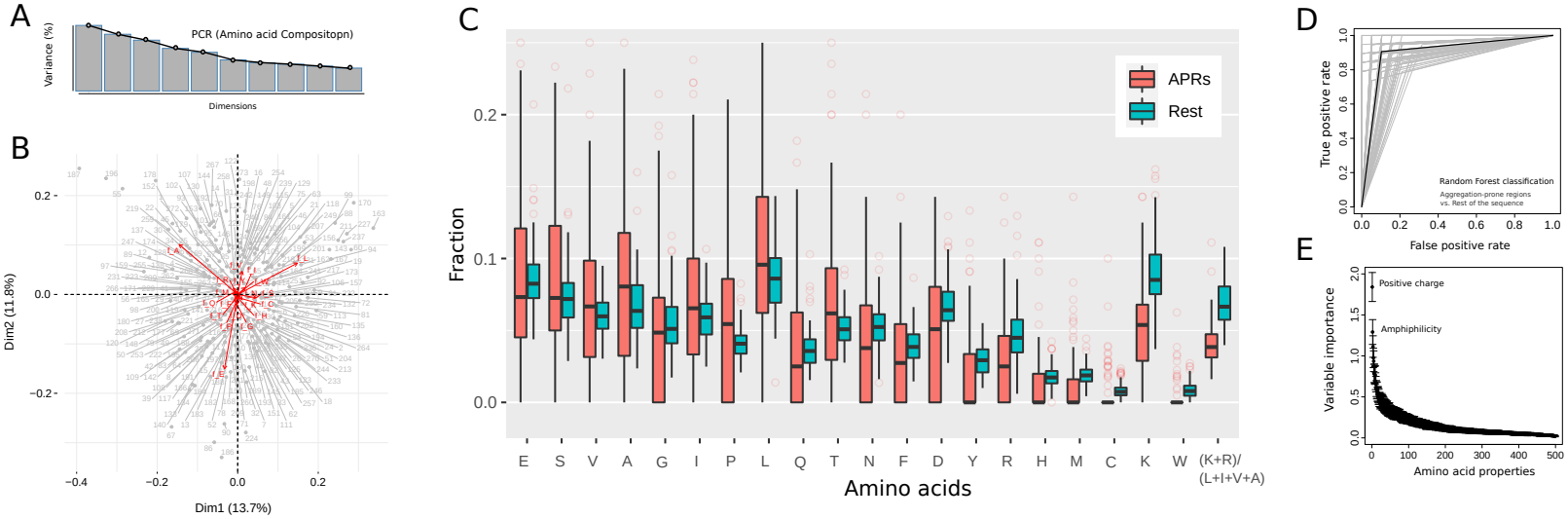
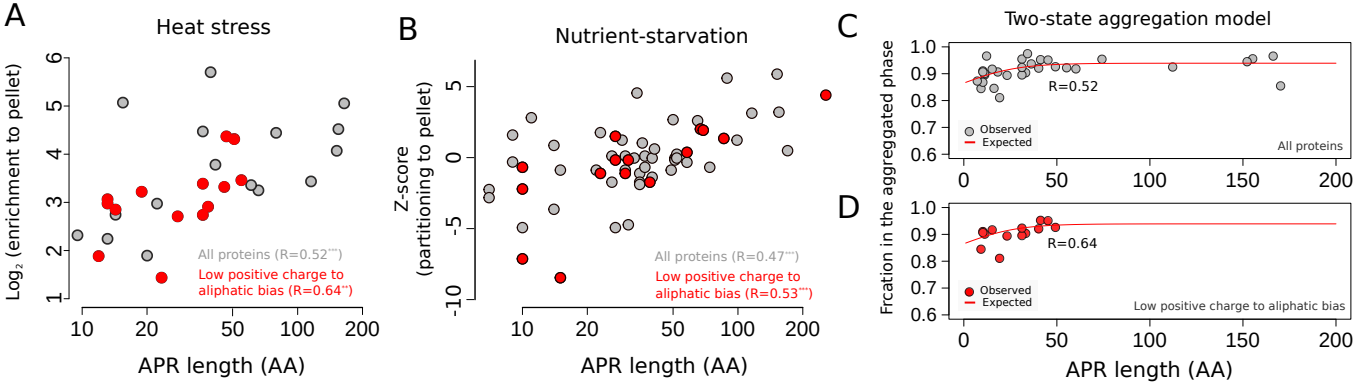

A
PCR (Amino acid Compositopn)
Variance (%)
Dimensions
B
Dim1 (13.7%)
Dim2 (11.8%)
C
APRs
Rest
Fraction
Amino acids
E S V A G I P L Q T N F D Y R H M C K W (K+R)/ (L+I+V+A)
D
True positive rate
False positive rate
Random Forest classification
Aggregation-prone regions vs. Rest of the sequence
E
Positive charge
Amphiphilicity
Variable importance
Amino acid properties

Figure 2


A
Heat stress
Log2 (enrichment to pellet)
All proteins (R=0.52***)
Low positive charge to aliphatic bias (R=0.64**)
APR length (AA)
B
Nutrient-starvation
Z-score (partitioning to pellet)
All proteins (R=0.47***)
Low positive charge to aliphatic bias (R=0.53***)
APR length (AA)
C
Two-state aggregation model
Frcation in the aggreggated phase
R=0.52
Observed
Expected
All proteins
D
R=0.64
Observed
Expected
Low positive charge to aliphatic bias
APR length (AA)