



Article

# *Arabidopsis* Heat Stress-Induced Proteins Are Enriched in Electrostatically Charged Amino Acids and Intrinsically Disordered Regions

David Alvarez-Ponce <sup>1,2,\*</sup>, Mario X. Ruiz-González <sup>2,3</sup> , Francisco Vera-Sirera <sup>2</sup>, Felix Feyertag <sup>1</sup>, Miguel A. Perez-Amador <sup>2</sup> and Mario A. Fares <sup>2,3,†</sup>

<sup>1</sup> Biology Department, University of Nevada, Reno, NV 89557, USA; ffeyertag@unr.edu

<sup>2</sup> Instituto de Biología Molecular y Celular de Plantas, CSIC-UPV, 46022 Valencia, Spain; marioxrui zgonzalez@gmail.com (M.X.R.-G.); fravesi@ibmcp.upv.es (F.V.-S.); mpereza@ibmcp.upv.es (M.A.P.-A.)

<sup>3</sup> Smurfit Institute of Genetics, University of Dublin, Trinity College, Dublin 2, Ireland

\* Correspondence: dap@unr.edu; Tel.: +1-(775)-682-5735

† Posthumous author.

Received: 9 July 2018; Accepted: 31 July 2018; Published: 3 August 2018



**Abstract:** Comparison of the proteins of thermophilic, mesophilic, and psychrophilic prokaryotes has revealed several features characteristic to proteins adapted to high temperatures, which increase their thermostability. These characteristics include a profusion of disulfide bonds, salt bridges, hydrogen bonds, and hydrophobic interactions, and a depletion in intrinsically disordered regions. It is unclear, however, whether such differences can also be observed in eukaryotic proteins or when comparing proteins that are adapted to temperatures that are more subtly different. When an organism is exposed to high temperatures, a subset of its proteins is overexpressed (heat-induced proteins), whereas others are either repressed (heat-repressed proteins) or remain unaffected. Here, we determine the expression levels of all genes in the eukaryotic model system *Arabidopsis thaliana* at 22 and 37 °C, and compare both the amino acid compositions and levels of intrinsic disorder of heat-induced and heat-repressed proteins. We show that, compared to heat-repressed proteins, heat-induced proteins are enriched in electrostatically charged amino acids and depleted in polar amino acids, mirroring thermophile proteins. However, in contrast with thermophile proteins, heat-induced proteins are enriched in intrinsically disordered regions, and depleted in hydrophobic amino acids. Our results indicate that temperature adaptation at the level of amino acid composition and intrinsic disorder can be observed not only in proteins of thermophilic organisms, but also in eukaryotic heat-induced proteins; the underlying adaptation pathways, however, are similar but not the same.

**Keywords:** temperature response; protein thermostability; salt bridges; intrinsically disordered proteins

## 1. Introduction

Proteins of thermophilic prokaryotes (those adapted to high temperatures) exhibit several distinctive features that increase their thermostability. One of the most consistent observations in thermophile proteins is an enrichment in salt bridges [1,2]. Salt bridges consist of electrostatic interactions among amino acid residues with positive (Lys and Arg) and negative (Glu and Asp) charges, and their contribution to increasing the stability of thermophilic bacteria was first proposed by Perutz and Raidt [3]. In addition, compared with proteins of mesophiles (adapted to intermediate temperatures) and psychrophiles (adapted to low temperatures), thermophile proteins tend to exhibit

more disulfide bonds and non-covalent interactions, including hydrogen bonds, and hydrophobic interactions, features that also tend to increase protein stability by linking together distant parts of the amino acid sequence [4,5]. These structural trends have an impact on the amino acid composition of thermophilic proteomes: the proteins of thermophilic bacteria tend to be enriched in charged amino acids and depleted in polar ones such as Ser, Thr, Asn, and Gln [6–12].

A few studies in prokaryotes have also shown that thermophile proteins are depleted in intrinsically disordered regions (IDRs), i.e., regions that lack a defined three-dimensional structure [13–15]. This observation is consistent with the fact that high temperatures induce disorder, but in contrast with the fact that IDRs confer thermoresistance [16–18].

Much less is known about how eukaryotic proteomes adapt to high temperatures. Some studies have suggested that the same biases in amino acid composition observed in thermophilic prokaryotes can be observed in thermophilic fungi (compared to other fungi; ref. [19]) and endothermic vertebrates (compared to ectothermic vertebrates; ref. [20]). In agreement with this notion, comparison of the orthologous proteins of two closely related fish, *Pachycara brachycephalum* (from Antarctica) and *Zoarces viviparus* (from a temperate zone) revealed an excess of Ser and a reduction of Glu and Asn in the cold-adapted species [21]. To our knowledge, the relationship between temperature and intrinsic disorder has not been investigated in eukaryotic proteomes.

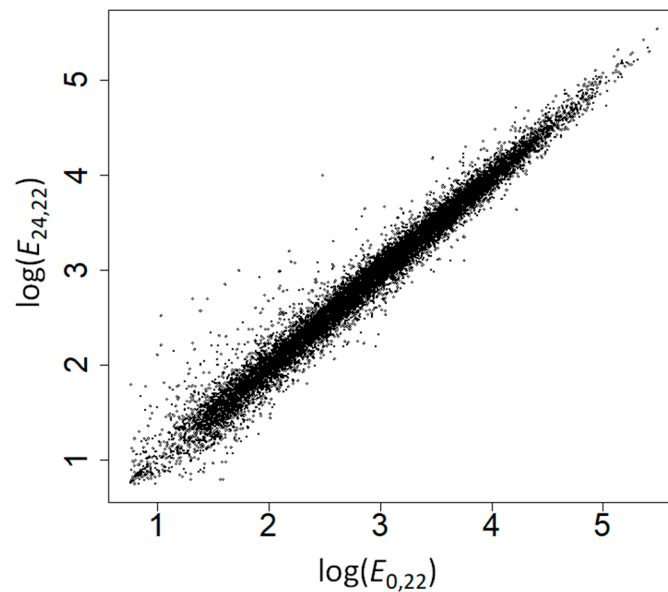
Protein adaptation to high temperatures is expected to be observed not only in the proteins of thermophilic organisms, but also in some of the proteins of any mesophilic organism. When an organism is exposed to high temperatures, a subset of its proteins is overexpressed, whereas others are repressed (heat-induced and heat-repressed proteins, respectively, e.g., ref. [22]). As heat-induced function at relatively high temperatures, we hypothesize that they should be similar to those of thermophilic organisms.

Plants represent particularly suitable models to test this hypothesis, as they are sessile organisms that cannot escape from their environment, and they lack the effective thermoregulation mechanisms exhibited by homeotherms. Therefore, plants are expected to have developed adaptations to cope with heat stress [23]. To test our hypothesis, we grew *Arabidopsis thaliana* plants under normal (22 °C) and heat stress conditions (37 °C), and measured gene expression levels. Proteins overexpressed under heat stress were enriched in electrostatically charged amino acids and depleted in polar and hydrophobic amino acids. However, in contrast with our expectations, these proteins were also enriched in IDRs. These results indicate that *Arabidopsis* heat-induced proteins exploit some, but not all the same mechanisms as thermophile proteins to cope with high temperatures.

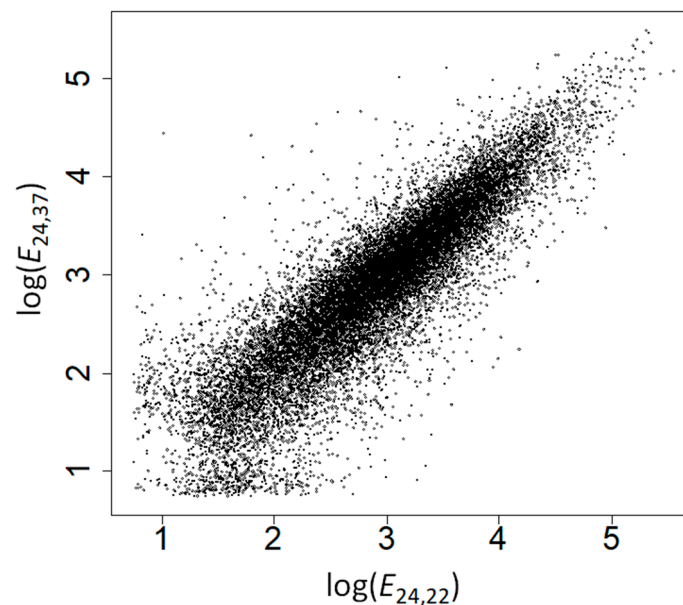
## 2. Results

### 2.1. Proteins That Are Overexpressed at High Temperatures Are Enriched in Electrostatically Charged Amino Acids and Depleted in Polar and Hydrophobic Amino Acids

We grew *Arabidopsis* plants at 22 and 37 °C for 24 h, and performed microarray analyses to measure gene expression levels at the beginning of the experiment ( $E_{0,22}$  = expression at time 0 and 22 °C) and at the end of the experiment ( $E_{24,22}$  and  $E_{24,37}$ ).  $E_{0,22}$  strongly correlated with  $E_{24,22}$  (Spearman's rank correlation coefficient,  $\rho = 0.991$ ,  $p < 10^{-200}$ ; Figure 1) supporting the robustness of our gene expression measures—the small differences between gene expression at both time points could be due to differences in gene expression during development and to measurement errors. The correlation between  $E_{24,22}$  and  $E_{24,37}$  was weaker ( $\rho = 0.897$ ,  $p = 10^{-200}$ ; Figure 2), highlighting the effect of heat stress on the expression of many genes.



**Figure 1.** Correlation between gene expression levels at 22 °C at time 0 and at time 24 h.

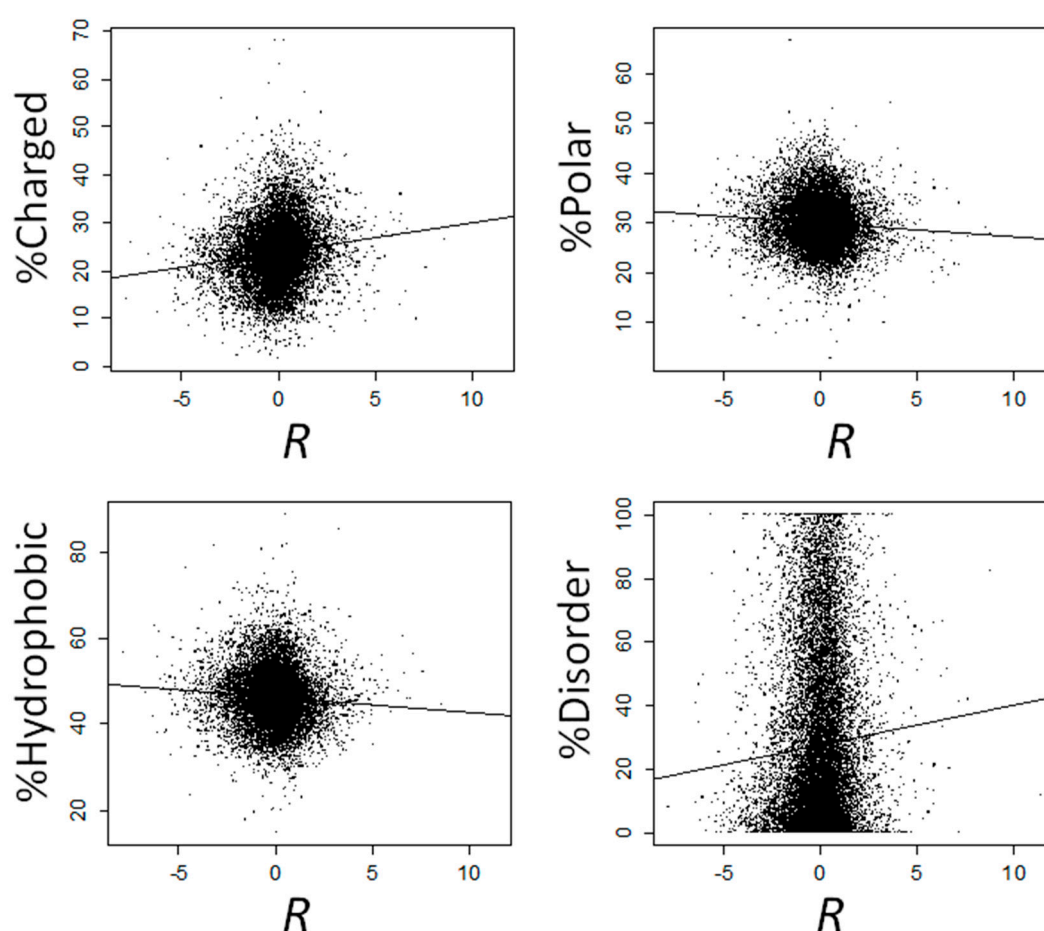


**Figure 2.** Correlation between gene expression levels at 22 °C at time 24 h and at 37 °C at time 24 h.

For each gene with available probes ( $n = 20,491$ ), we computed a response to heat stress ( $R$ ) as the logarithm in base 2 of the ratio of expression levels at 37 and 22 °C (following formula 1). Genes with  $R > 0$  are overexpressed at high temperatures, and genes with  $R < 0$  are repressed. Genes with  $R > 1$  (strongly overexpressed) are enriched in Gene Ontology biological processes “protein refolding”, “protein folding”, “chaperone cofactor-dependent protein refolding”, “chaperone-mediated protein folding”, “de novo posttranslational protein folding”, “de novo protein folding”, “cellular response to heat”, “response to heat”, “response to temperature stimulus”, and “heat acclimation”. They are also enriched in molecular functions “misfolded protein binding”, “heat shock protein binding”, “protein binding involved in protein folding”, and “unfolded protein binding” (Tables S1–S3).

We observed a positive correlation between  $R$  and the fraction of charged amino acids ( $\rho = 0.146$ ,  $p = 2.47 \times 10^{-98}$ ), and negative correlations between  $R$  and both the fraction of polar ( $\rho = -0.076$ ,

$p = 1.72 \times 10^{-27}$ ) and hydrophobic ( $\rho = -0.084$ ,  $p = 4.08 \times 10^{-33}$ ) amino acids (Figure 3). We next computed the correlation between  $R$  and the frequency of each amino acid separately. The correlation was significantly positive for all four charged amino acids (Arg, Asp, Glu, and Lys), negative for all hydrophobic amino acids (significant for Gly, Ile, Phe, Pro, and Val), except Met (for which the correlation was non-significantly positive), and negative for all polar amino acids (significant for Asn, Ser, Thr, Trp and Tyr), except for Gln, for which the correlation was significantly positive (Table 1). All these correlations remained significant after controlling for multiple testing (Table 1).



**Figure 3.** Correlations between response to high temperature ( $R$ ) and the fraction of charged, polar, hydrophobic and disordered amino acids. Lines represent regression lines.

**Table 1.** Correlations between amino acid frequencies and response to high temperature.

Type	Amino Acid	No Control			Controlling for GC Content			Controlling for $E_{24,22}$			Controlling for $E_{24,37}$		
		$\rho$	$p$ -Value	$q$ -Value	$\rho$	$p$ -Value	$q$ -Value	$\rho$	$p$ -Value	$q$ -Value	$\rho$	$p$ -Value	$q$ -Value
Charged	Arg	0.075	<b><math>1.31 \times 10^{-26}</math></b>	<b><math>4.37 \times 10^{-26}</math></b>	0.068	<b><math>2.05 \times 10^{-22}</math></b>	<b><math>5.86 \times 10^{-22}</math></b>	0.068	<b>0.013</b>	<b>0.015</b>	0.079	<b><math>1.02 \times 10^{-29}</math></b>	<b><math>3.40 \times 10^{-29}</math></b>
	Asp	0.104	<b><math>1.62 \times 10^{-50}</math></b>	<b><math>1.62 \times 10^{-49}</math></b>	0.105	<b><math>9.95 \times 10^{-52}</math></b>	<b><math>9.95 \times 10^{-51}</math></b>	0.106	<b><math>7.16 \times 10^{-53}</math></b>	<b><math>7.16 \times 10^{-52}</math></b>	0.095	<b><math>1.84 \times 10^{-42}</math></b>	<b><math>1.23 \times 10^{-41}</math></b>
	Glu	0.118	<b><math>5.48 \times 10^{-64}</math></b>	<b><math>1.10 \times 10^{-62}</math></b>	0.122	<b><math>5.60 \times 10^{-69}</math></b>	<b><math>1.12 \times 10^{-67}</math></b>	0.115	<b><math>2.61 \times 10^{-61}</math></b>	<b><math>5.22 \times 10^{-60}</math></b>	0.115	<b><math>2.60 \times 10^{-61}</math></b>	<b><math>5.20 \times 10^{-60}</math></b>
	Lys	0.082	<b><math>8.23 \times 10^{-32}</math></b>	<b><math>4.12 \times 10^{-31}</math></b>	0.100	<b><math>9.58 \times 10^{-47}</math></b>	<b><math>4.79 \times 10^{-46}</math></b>	0.081	<b><math>2.27 \times 10^{-31}</math></b>	<b><math>9.08 \times 10^{-31}</math></b>	0.079	<b><math>8.76 \times 10^{-30}</math></b>	<b><math>3.40 \times 10^{-29}</math></b>
	Total	0.146	<b><math>2.47 \times 10^{-98}</math></b>		0.155	<b><math>3.61 \times 10^{-111}</math></b>		0.145	<b><math>1.04 \times 10^{-97}</math></b>		0.140	<b><math>1.65 \times 10^{-90}</math></b>	
Polar	Asn	−0.025	<b><math>3.86 \times 10^{-4}</math></b>	<b>0.001</b>	−0.018	<b>0.011</b>	<b>0.015</b>	−0.044	<b><math>4.17 \times 10^{-10}</math></b>	<b><math>8.34 \times 10^{-10}</math></b>	0.005	0.433	0.433
	Cys	−0.011	0.127	0.158	−0.009	0.187	0.208	−0.034	<b><math>1.54 \times 10^{-6}</math></b>	<b><math>2.57 \times 10^{-6}</math></b>	0.026	<b><math>2.07 \times 10^{-4}</math></b>	<b><math>3.19 \times 10^{-4}</math></b>
	Gln	0.046	<b><math>3.20 \times 10^{-11}</math></b>	<b><math>7.11 \times 10^{-11}</math></b>	0.053	<b><math>2.36 \times 10^{-14}</math></b>	<b><math>5.24 \times 10^{-14}</math></b>	0.046	<b><math>6.79 \times 10^{-11}</math></b>	<b><math>1.51 \times 10^{-10}</math></b>	0.044	<b><math>3.95 \times 10^{-10}</math></b>	<b><math>7.90 \times 10^{-10}</math></b>
	His	−0.010	0.134	0.158	−0.009	0.210	0.221	−0.024	<b>0.001</b>	<b>0.001</b>	0.010	0.146	0.154
	Ser	−0.036	<b><math>2.25 \times 10^{-7}</math></b>	<b><math>4.09 \times 10^{-7}</math></b>	−0.042	<b><math>2.36 \times 10^{-9}</math></b>	<b><math>4.72 \times 10^{-9}</math></b>	−0.052	<b><math>1.00 \times 10^{-13}</math></b>	<b><math>2.86 \times 10^{-13}</math></b>	−0.012	0.092	0.102
	Thr	−0.099	<b><math>1.10 \times 10^{-45}</math></b>	<b><math>7.33 \times 10^{-45}</math></b>	−0.100	<b><math>9.24 \times 10^{-47}</math></b>	<b><math>4.79 \times 10^{-46}</math></b>	−0.098	<b><math>1.12 \times 10^{-44}</math></b>	<b><math>7.47 \times 10^{-44}</math></b>	−0.096	<b><math>2.75 \times 10^{-43}</math></b>	<b><math>2.75 \times 10^{-42}</math></b>
	Trp	−0.033	<b><math>2.26 \times 10^{-6}</math></b>	<b><math>3.77 \times 10^{-6}</math></b>	−0.036	<b><math>2.50 \times 10^{-7}</math></b>	<b><math>4.55 \times 10^{-7}</math></b>	−0.039	<b><math>2.11 \times 10^{-8}</math></b>	<b><math>3.84 \times 10^{-8}</math></b>	−0.022	<b>0.002</b>	<b>0.002</b>
	Tyr	−0.024	<b>0.001</b>	<b>0.001</b>	−0.016	<b>0.021</b>	<b>0.026</b>	−0.025	<b><math>3.72 \times 10^{-4}</math></b>	<b><math>4.96 \times 10^{-4}</math></b>	−0.021	<b>0.003</b>	<b>0.004</b>
	Total	−0.076	<b><math>1.72 \times 10^{-27}</math></b>		−0.072	<b><math>1.11 \times 10^{-24}</math></b>		−0.102	<b><math>9.48 \times 10^{-49}</math></b>		−0.034	<b><math>9.25 \times 10^{-7}</math></b>	
Hydrophobic	Ala	−0.008	0.280	0.311	−0.020	<b>0.004</b>	<b>0.006</b>	0.027	<b><math>1.32 \times 10^{-4}</math></b>	<b><math>1.89 \times 10^{-4}</math></b>	−0.060	<b><math>1.50 \times 10^{-17}</math></b>	<b><math>3.75 \times 10^{-17}</math></b>
	Gly	−0.054	<b><math>1.40 \times 10^{-14}</math></b>	<b><math>3.50 \times 10^{-14}</math></b>	−0.066	<b><math>1.99 \times 10^{-21}</math></b>	<b><math>4.98 \times 10^{-21}</math></b>	−0.028	<b><math>5.46 \times 10^{-5}</math></b>	<b><math>8.40 \times 10^{-5}</math></b>	−0.092	<b><math>1.17 \times 10^{-39}</math></b>	<b><math>5.85 \times 10^{-39}</math></b>
	Ile	−0.045	<b><math>1.01 \times 10^{-10}</math></b>	<b><math>2.02 \times 10^{-10}</math></b>	−0.035	<b><math>5.63 \times 10^{-7}</math></b>	<b><math>9.38 \times 10^{-7}</math></b>	−0.052	<b><math>1.55 \times 10^{-13}</math></b>	<b><math>3.88 \times 10^{-13}</math></b>	−0.033	<b><math>2.91 \times 10^{-6}</math></b>	<b><math>5.29 \times 10^{-6}</math></b>
	Leu	−0.004	0.547	0.547	−0.004	0.533	0.533	−0.016	<b>0.021</b>	<b>0.023</b>	0.015	<b>0.029</b>	<b>0.034</b>
	Met	0.006	0.387	0.407	0.014	<b>0.042</b>	<b>0.049</b>	−0.001	0.942	0.942	0.017	<b>0.017</b>	<b>0.021</b>
	Phe	−0.075	<b><math>1.04 \times 10^{-26}</math></b>	<b><math>4.16 \times 10^{-26}</math></b>	−0.070	<b><math>9.79 \times 10^{-24}</math></b>	<b><math>3.26 \times 10^{-23}</math></b>	−0.084	<b><math>2.59 \times 10^{-33}</math></b>	<b><math>1.30 \times 10^{-32}</math></b>	−0.056	<b><math>1.36 \times 10^{-15}</math></b>	<b><math>3.02 \times 10^{-15}</math></b>
	Pro	−0.060	<b><math>8.03 \times 10^{-18}</math></b>	<b><math>2.29 \times 10^{-17}</math></b>	−0.074	<b><math>1.85 \times 10^{-26}</math></b>	<b><math>7.40 \times 10^{-26}</math></b>	−0.052	<b><math>8.41 \times 10^{-14}</math></b>	<b><math>2.80 \times 10^{-13}</math></b>	−0.070	<b><math>7.88 \times 10^{-24}</math></b>	<b><math>2.25 \times 10^{-23}</math></b>
	Val	−0.017	<b>0.012</b>	<b>0.017</b>	−0.024	<b>0.001</b>	<b>0.001</b>	−0.006	0.370	0.390	−0.033	<b><math>3.30 \times 10^{-6}</math></b>	<b><math>5.50 \times 10^{-6}</math></b>
	Total	−0.084	<b><math>4.08 \times 10^{-33}</math></b>		−0.096	<b><math>1.31 \times 10^{-43}</math></b>		−0.064	<b><math>2.88 \times 10^{-20}</math></b>		−0.109	<b><math>2.73 \times 10^{-55}</math></b>	

$p$ -values and  $q$ -values shown in bold face represent significant tests at  $\alpha = 0.05$  or  $q = 0.05$ .

Next, we compared the amino acid composition of proteins encoded by genes that are overexpressed ( $R > 0$ ,  $n = 10,728$ ) vs. proteins encoded by genes that are repressed ( $R < 0$ ,  $n = 9763$ ) at 37 °C. Overexpressed proteins were enriched in charged amino acids (median percent in overexpressed proteins: 24.32%; median percent in repressed proteins: 23.20%; Mann-Whitney's  $U$  test,  $p = 1.90 \times 10^{-66}$ ) and depleted in both polar (median percent in overexpressed proteins: 29.54%; median percent in repressed proteins: 30.04%;  $p = 2.53 \times 10^{-20}$ ) and hydrophobic (median percent in overexpressed proteins: 45.77%; median percent in repressed proteins: 46.43%;  $p = 6.56 \times 10^{-21}$ ) amino acids. In almost perfect agreement with our correlation analyses, proteins encoded by overexpressed genes were significantly enriched in Arg, Asp, Gln, Glu, and Lys, and significantly depleted in Asn, Gly, Ile, Phe, Pro, Ser, Thr, and Trp (Table 2).

**Table 2.** Amino acid frequencies in overexpressed ( $R > 0$ ) and repressed ( $R < 0$ ) proteins at high temperatures.

Type	Amino Acid	Median Overexpressed (%)	Median Repressed (%)	$p$ -Value	$q$ -Value
Charged	Arg	5.43	5.19	<b><math>8.06 \times 10^{-21}</math></b>	<b><math>4.61 \times 10^{-20}</math></b>
	Asp	5.36	5.10	<b><math>1.60 \times 10^{-36}</math></b>	<b><math>6.40 \times 10^{-35}</math></b>
	Glu	6.61	6.15	<b><math>8.28 \times 10^{-44}</math></b>	<b><math>6.62 \times 10^{-42}</math></b>
	Lys	6.33	6.06	<b><math>1.20 \times 10^{-21}</math></b>	<b><math>8.00 \times 10^{-21}</math></b>
	Total	24.32	23.20	<b><math>1.90 \times 10^{-66}</math></b>	
Polar	Asn	4.08	4.12	<b>0.017</b>	<b>0.024</b>
	Cys	1.59	1.60	<b>0.043</b>	0.060
	Gln	3.27	3.16	<b><math>7.77 \times 10^{-8}</math></b>	<b><math>1.88 \times 10^{-7}</math></b>
	His	2.11	2.10	0.204	0.244
	Ser	8.79	8.96	<b><math>2.27 \times 10^{-7}</math></b>	<b><math>5.19 \times 10^{-7}</math></b>
	Thr	4.90	5.13	<b><math>5.31 \times 10^{-34}</math></b>	<b><math>1.42 \times 10^{-32}</math></b>
	Trp	1.07	1.11	<b><math>4.75 \times 10^{-4}</math></b>	<b>0.001</b>
	Tyr	2.65	2.68	0.132	0.163
	Total	29.54	30.04	<b><math>2.53 \times 10^{-20}</math></b>	
Hydrophobic	Ala	6.32	6.30	0.889	0.889
	Gly	6.18	6.41	<b><math>2.77 \times 10^{-10}</math></b>	<b><math>8.21 \times 10^{-10}</math></b>
	Ile	5.12	5.23	<b><math>1.87 \times 10^{-7}</math></b>	<b><math>4.40 \times 10^{-7}</math></b>
	Leu	9.24	9.27	0.675	0.720
	Met	2.38	2.37	0.399	0.449
	Phe	4.08	4.28	<b><math>1.55 \times 10^{-18}</math></b>	<b><math>7.75 \times 10^{-18}</math></b>
	Pro	4.54	4.71	<b><math>2.56 \times 10^{-12}</math></b>	<b><math>8.53 \times 10^{-12}</math></b>
	Val	6.67	6.68	0.178	0.215
	Total	45.77	46.43	<b><math>6.56 \times 10^{-21}</math></b>	

$p$ -values correspond to the Mann-Whitney's  $U$  test.  $p$ -values and  $q$ -values shown in bold face represent significant tests at  $\alpha = 0.05$  or  $q = 0.05$ .

Similar results were obtained when using a more stringent threshold to classify genes as overexpressed ( $R > 2$ ,  $n = 826$ ) or repressed ( $R < -2$ ,  $n = 1214$ ) at 37 °C. Overexpressed proteins are enriched in charged amino acids (median percent in overexpressed proteins: 25.30%; median percent in repressed proteins: 22.54%;  $p = 1.50 \times 10^{-26}$ ) and depleted in both polar (median percent in overexpressed proteins: 29.74%; median percent in repressed proteins: 30.17%;  $p = 3.20 \times 10^{-8}$ ) and hydrophobic (median percent in overexpressed proteins: 45.20%; median percent in repressed proteins: 47.24%;  $p = 6.04 \times 10^{-11}$ ) amino acids. More specifically, overexpressed proteins are significantly enriched in Arg, Asp, Gln, Glu, and Lys, and significantly depleted in Asn, Cys, Gly, His, Ile, Phe, Pro, Thr, Trp, and Tyr (Table 3).

**Table 3.** Amino acid frequencies in highly overexpressed ( $R > 2$ ) and highly repressed ( $R < -2$ ) proteins at high temperatures.

Type	Amino Acid	Median Overexpressed (%)	Median Repressed (%)	<i>p</i> -Value	<i>q</i> -Value
Charged	Arg	5.26	4.80	$7.82 \times 10^{-9}$	$1.04 \times 10^{-7}$
	Asp	5.51	4.95	$1.62 \times 10^{-12}$	$4.32 \times 10^{-11}$
	Glu	6.92	5.92	$1.31 \times 10^{-17}$	$1.05 \times 10^{-15}$
	Lys	6.78	6.17	$1.78 \times 10^{-7}$	$1.78 \times 10^{-6}$
	Total	25.30	22.54	$1.50 \times 10^{-26}$	
Polar	Asn	4.04	4.29	$2.81 \times 10^{-4}$	<b>0.001</b>
	Cys	1.66	1.69	<b>0.031</b>	<b>0.045</b>
	Gln	3.13	2.94	$2.87 \times 10^{-4}$	$6.57 \times 10^{-4}$
	His	2.03	2.12	<b>0.023</b>	<b>0.035</b>
	Ser	8.47	8.41	0.780	0.810
	Thr	4.95	5.26	$3.52 \times 10^{-6}$	$1.56 \times 10^{-5}$
	Trp	1.05	1.15	<b>0.035</b>	<b>0.050</b>
	Tyr	2.57	2.86	$6.09 \times 10^{-5}$	$1.87 \times 10^{-4}$
	Total	29.74	30.17	$3.20 \times 10^{-8}$	
Hydrophobic	Ala	6.11	6.12	0.867	0.878
	Gly	6.05	6.50	$5.49 \times 10^{-5}$	$1.76 \times 10^{-4}$
	Ile	5.25	5.48	<b>0.001</b>	<b>0.002</b>
	Leu	9.01	9.17	0.215	0.292
	Met	2.46	2.52	0.321	0.395
	Phe	4.12	4.65	$9.55 \times 10^{-13}$	$3.82 \times 10^{-11}$
	Pro	4.31	4.62	$9.34 \times 10^{-5}$	$2.58 \times 10^{-4}$
	Val	6.76	6.84	0.294	0.386
	Total	45.20	47.24	$6.04 \times 10^{-11}$	

*p*-values correspond to the Mann-Whitney's *U* test. *p*-values and *q*-values shown in bold face represent significant tests at  $\alpha = 0.05$  or  $q = 0.05$ .

## 2.2. The Amino Acid Composition of Heat-Induced Proteins Is Not due to Covariation of Amino Acid Composition with GC Content, Gene Expression Levels, or Subcellular Location

We considered whether our results could be affected by confounding factors. First, GC content is known to affect amino acid composition [24], and *R* significantly correlates with GC content ( $\rho = 0.088$ ,  $p = 9.76 \times 10^{-37}$ ). Combined, these correlations alone might potentially explain the observed trends. To discard this possibility, we computed partial correlations between *R* and the frequency of each amino acid, while controlling for GC content, with very similar results. The correlation continued to be significantly positive for charged amino acids and significantly negative for polar and hydrophobic ones (Table 1). More specifically, the correlation was significantly positive for Arg, Asp, Gln, Glu, and Lys and significantly negative for Asn, Gly, Ile, Phe, Pro, Ser, Thr, Trp, Tyr, and Val. Both the negative correlation between *R* and Ala frequency and the positive correlation between *R* and Met frequency, which were initially not significant, became significant after controlling for GC content (Table 1).

Second, highly expressed proteins resemble proteins from thermophiles in their amino acid composition [25], and expression levels correlate with *R* (expression level at 22 °C:  $\rho = -0.156$ ,  $p = 4.88 \times 10^{-112}$ ; expression level at 37 °C:  $\rho = 0.241$ ,  $p = 1.18 \times 10^{-268}$ ). To discard the potential confounding effects of expression levels, we computed partial correlations between *R* and the frequency of each amino acid, while controlling for expression levels, again with very similar results. When controlling for expression levels at 22 °C, *R* correlated positively with the frequencies of Ala, Arg, Asp, Gln, Glu, and Lys and negatively with the frequencies of Asn, Cys, Gly, His, Ile, Leu, Phe, Pro, Ser, Thr, Trp, and Tyr. When controlling for expression levels at 37 °C, *R* correlated positively with the frequencies of Arg, Asp, Cys, Gln, Glu, Leu, Lys, and Met and negatively with the frequencies of Ala, Gly, Ile, Phe, Pro, Thr, Trp, Tyr, and Val. In both cases, the positive correlations between *R* and the frequency charged amino acids and the negative correlations between *R* and the frequencies of polar and hydrophobic amino acids remained significant (Table 1).



Proteins locating to different parts of the cell differ in their amino acid compositions and in their response to heat stress ([26,27]; Table 4). To discard subcellular location as a confounding factor, we analyzed the correlation between  $R$  and the amino acid composition separately for proteins locating to 10 different subcellular compartments (Table 5). The correlation between  $R$  and the fraction of charged amino acids was positive in nine of the compartments, which represents a significant departure from the 50% expected at random (one-tailed binomial test,  $p = 0.011$ ). The correlation was significantly positive for the cytosol, the plastid (the compartments with the higher number of known/inferred proteins), and the mitochondrion. The correlation between  $R$  and the fraction of hydrophobic amino acids was negative in eight of the compartments (one-tailed binomial test,  $p = 0.055$ ), significantly negative in the plastid and the mitochondrion, and significantly positive in the nucleus. The correlation between  $R$  and the fraction of polar amino acids was negative in half of the compartments, and significantly negative in the cytosol and the nucleus. These results suggest that the enrichment of heat-induced proteins in charged amino acids and their depletion in hydrophobic amino acids are not a byproduct of covariation of both  $R$  and amino acid composition with subcellular location. The lack of significance in most of the individual correlations is probably due to the low number of proteins for which location information is available, ranging from 720 for the plastid to 63 in the peroxisome (Table 4), which is expected to greatly reduce the statistical power of our compartment-specific analyses. However, we note an exception: among nuclear proteins  $R$  exhibits a significantly positive correlation with the percent of hydrophobic residues (Table 5).

### 2.3. Proteins That Are Overexpressed at High Temperatures Are Highly Disordered

For each *Arabidopsis* protein, we computed the percentage of amino acids that belong to IDRs using IUPred [28]. This percentage correlates positively with  $R$  ( $\rho = 0.059$ ,  $p = 4.93 \times 10^{-17}$ ; Figure 3). Genes that are overexpressed at 37 °C ( $R > 0$ ) encode proteins that are more disordered than those that are repressed ( $R < 0$ ), with median disorder percent of 19.19% and 16.51% for induced and repressed genes, respectively (Mann-Whitney's  $U$  test,  $p = 2.01 \times 10^{-35}$ ). The differences are more solid when comparing genes that are strongly overexpressed at 37 °C ( $R > 2$ ) vs. those that are strongly repressed ( $R < -2$ ), with percentages of median disorder of 21.54% and 11.51% for induced and repressed genes, respectively (Mann-Whitney's  $U$  test,  $P = 2.03 \times 10^{-23}$ ).

In agreement with previous works [29,30], we found a positive correlation between GC content and the percent of disordered residues ( $\rho = 0.044$ ,  $p = 2.84 \times 10^{-10}$ ). In addition, GC content positively correlates with  $R$  ( $\rho = 0.088$ ,  $p = 9.76 \times 10^{-37}$ ), making it possible that the positive correlation between  $R$  and disorder might be due to the covariation of both parameters with GC content. The correlation between  $R$  and disorder, however, is significant, even after controlling for GC content ( $\rho = 0.055$ ,  $p = 3.44 \times 10^{-15}$ ).

Likewise, intrinsic disorder positively correlates with expression levels (at 22 °C:  $\rho = 0.040$ ,  $p = 1.03 \times 10^{-8}$ ; and at 37 °C:  $\rho = 0.072$ ,  $p = 7.75 \times 10^{-25}$ ), in agreement with previous results in *Escherichia coli* [31], but in contrast with observations in yeasts [32,33]. Disorder, however, significantly correlates with  $R$  after controlling for expression levels (at 22 °C:  $\rho = 0.066$ ,  $p = 4.64 \times 10^{-21}$ ; and at 37 °C:  $\rho = 0.043$ ,  $p = 1.03 \times 10^{-9}$ ).

Both intrinsic disorder and  $R$  substantially vary among proteins locating to different subcellular compartments (Table 4), thus raising the possibility that covariation of both factors with subcellular location may account for the observed enrichment of stress-induced proteins in IDRs. We analyzed the correlation between intrinsic disorder and  $R$  separately for proteins locating to 10 different subcellular compartments. The correlation was positive for eight of the tissues (significantly positive for the cytosol, endoplasmic reticulum, and the vacuole) and significantly negative for the nucleus and the plasma membrane (Table 5). These results indicate that the positive correlation between disorder and  $R$ , while generalized, does not apply to proteins locating to all compartments.



**Table 4.** Amino acid composition, intrinsic disorder and response to heat stress of proteins locating to different subcellular locations.

Subcellular Location	<i>n</i>	Median Charged Amino Acids (%)	Median Polar Amino Acids (%)	Median Hydrophobic Amino Acids (%)	Median Intrinsic Disorder (%)	Median <i>R</i>
Cytosol	633	25.46	26.74	47.31	15.64	0.131
Endoplasmic reticulum	163	24.12	27.22	48.68	10.11	0.147
Extracellular	197	18.94	32.87	48.43	8.61	−0.296
Golgi	375	23.20	29.37	47.38	14.22	0.108
Mitochondrion	286	23.12	27.63	49.26	14.93	0.261
Nucleus	446	26.50	29.16	43.92	42.73	0.406
Peroxisome	63	23.16	26.47	50.00	10.61	−0.207
Plasma membrane	343	22.21	28.63	48.73	14.97	−0.195
Plastid	720	23.33	27.65	48.91	15.28	−0.190
Vacuole	81	21.14	28.24	49.75	8.12	−0.008

**Table 5.** Correlations between amino acid frequencies and response to high temperature among proteins of different subcellular locations.

Subcellular Location	Correlation <i>R</i> -Charged Amino Acids		Correlation <i>R</i> -Polar Amino Acids		Correlation <i>R</i> -Hydrophobic Amino Acids		Correlation <i>R</i> -Intrinsic Disorder	
	$\rho$	<i>p</i> -Value	$\rho$	<i>p</i> -Value	$\rho$	<i>p</i> -Value	$\rho$	<i>p</i> -Value
Cytosol	0.171	<b><math>1.54 \times 10^{-5}</math></b>	−0.142	<b><math>3.27 \times 10^{-4}</math></b>	−0.069	0.082	0.123	<b>0.002</b>
Endoplasmic reticulum	0.061	0.437	−0.015	0.847	−0.112	0.155	0.226	<b>0.004</b>
Extracellular	0.054	0.452	0.021	0.765	−0.073	0.309	0.068	0.346
Golgi	−0.046	0.370	0.068	0.191	−0.009	0.866	0.073	0.156
Mitochondrion	0.124	<b>0.036</b>	0.060	0.312	−0.125	<b>0.034</b>	0.065	0.272
Nucleus	0.020	0.681	−0.119	<b>0.012</b>	0.102	<b>0.031</b>	−0.207	<b><math>1.09 \times 10^{-5}</math></b>
Peroxisome	0.016	0.902	−0.104	0.416	0.080	0.535	0.237	0.062
Plasma membrane	0.064	0.234	−0.017	0.750	−0.004	0.947	−0.154	<b>0.004</b>
Plastid	0.137	<b><math>2.20 \times 10^{-4}</math></b>	0.007	0.859	−0.110	<b>0.003</b>	0.062	0.095
Vacuole	0.184	0.099	0.082	0.466	−0.189	0.091	0.266	<b>0.017</b>

*p*-values shown in bold face represent significant tests at  $\alpha = 0.05$ .

### 3. Discussion

We show that *Arabidopsis* proteins whose expression levels increase at high temperatures (heat-induced proteins) are enriched in charged amino acids, and depleted in polar and hydrophobic amino acids, compared to heat-repressed proteins. The enrichment of heat-induced proteins in charged amino acids and the depletion in polar amino acids are trends that mirror those observed in the proteins of thermophilic prokaryotes. The observed enrichment of heat-induced proteins in electrostatically charged amino acids was expected, as such amino acids can engage in salt bridges, which usually increase protein thermostability [1–3]—it should be noted, nonetheless, that not all charged amino acids participate in salt bridges, and that not all salt bridges increase thermostability [34]. However, the depletion of heat-induced proteins in hydrophobic amino acids was not expected, as the proteins of thermophilic prokaryotes are usually enriched in such amino acids (e.g., ref. [35]).

Despite the overall observed trends (heat-induced proteins being enriched in charged amino acids and depleted in polar and hydrophobic amino acids), not all amino acids vary according to these rules. In particular, the frequencies of Cys (polar), His (polar), Ala (hydrophobic), Leu (hydrophobic), and Met (hydrophobic) do not correlate significantly with *R*, and Gln (a polar amino acid) is more frequent in heat-induced proteins than in heat-repressed ones (Table 1). The enrichment of heat-induced proteins in Gln is surprising, given its tendency to undergo deamination at high temperatures [36].

We show that the observed overall trends are not due to heat-induced genes/proteins being different in terms of expression levels, GC content or subcellular location. When controlling for these factors, however, the direction of the correlations for certain amino acids change (Table 1). Thus, the observed trends in amino acid composition are likely the result of adaptation of heat-induced and heat-repressed *Arabidopsis* proteins to high and low temperatures, respectively.

Burra et al. [13] predicted that the proteins of thermophilic prokaryotes should be enriched in IDRs, as intrinsically disorder proteins are often resistant to high temperatures [16–18]. However, contradicting their predictions, they observed that thermophiles often are depleted in IDRs, which may compensate for the disorder induced by temperature. Similar observations were made in both another proteome-level analysis [15] and an analysis of FlgM proteins from bacteria adapted to different temperatures [14]. In agreement with Burra et al.'s prediction, we observed that *Arabidopsis* heat-induced proteins are enriched in IDRs. Our results suggest that there are different ways in which ordered/disordered regions can promote thermostability.

The correlations described in the current work are moderate, albeit statistically significant. Several scenarios may account for the weakness of the correlations. First, amino acid composition and protein intrinsic disorder may be affected by factors other than temperature. Second, the difference between the temperatures used in this study (22 vs. 37 °C) is small compared to the differences between the optimal temperatures of psychrophiles, mesophiles, and thermophiles. Third, certain plant genes may have changed their patterns of response to heat stress during the recent evolutionary history of *Arabidopsis*. i.e., certain genes that are currently heat-induced may have been heat-repressed in the past, and certain genes that are currently heat-repressed may have been heat-induced in the past. As amino acid and disorder adjustment to temperature is expected to take a relatively long amount of time, such switches in expression profiles may have limited the adaptation of proteomes to temperatures. Fourth, the adaptability of plant proteomes to temperatures may be more limited than that of prokaryotic proteomes, e.g., due to the higher complexity of protein-protein interaction networks and the smaller effective population size of plants [37].

In summary, the amino acid composition of heat-induced proteins in *Arabidopsis* mirrors to some extent, but not completely, that of the proteomes of thermophilic prokaryotes. This indicates that protein adaptation to high temperatures takes place partly through similar molecular mechanisms in prokaryotes and eukaryotes. Our observations also indicate that adaptation of proteins at the level of amino acid composition and protein intrinsic disorder can be detected not only when comparing the proteomes of species adapted to very different temperatures, but also among the proteins of the

same species with different temperature response profiles. These observations expand our view of how eukaryotic proteomes adapt to different temperatures.

## 4. Materials and Methods

### 4.1. Plant Material, Growth Conditions, and Experimental Design

*Arabidopsis thaliana* Columbia ecotype seeds were sterilized with 70% ethanol for 20 min, 2.5% sodium hypochlorite (commercial bleach) with 0.05% Triton X-100 for 10 min, and finally, four washes with sterile dH<sub>2</sub>O. Seeds were placed onto Whatmann paper in Murashige and Skoog (MS) medium plates (Duchefa, Haarlem, The Netherlands). Plates were kept in the dark at 4 °C for 96 h for stratification, and incubated during 8 h in light at 22 °C to promote germination. Plates were transferred to darkness at 22 °C for 72 h. At this moment plates were either kept at 22 °C or transferred to 37 °C. Seedlings were harvested at 0 and 24 h with four biological replicates. Samples were frozen in liquid nitrogen and stored at −80 °C.

### 4.2. Microarray Analysis

Total RNA was extracted using the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany), and RNA integrity was tested with the 2100 Bioanalyzer (Agilent, Santa Clara, CA, USA). Transcriptome analyses were carried out according to Minimum Information About a Microarray Experiment (MIAME) guidelines. We used the Agilent *Arabidopsis* (V4) Gene Expression 4 × 44K Microarray in a one-color experimental design. The microarray contained 43,803 probes (60-mer oligonucleotides). Four biological replicates were analyzed for each treatment (time points 0 and 24 h at 22 °C and 24 h at 37 °C).

Half a µg of RNA was amplified and labeled with the Agilent Low Input Quick Amp Labeling Kit. To assess the labeling and hybridization efficiencies we used an Agilent Spike-In Kit. Hybridization and slide washing were performed with the Gene Expression Hybridization Kit (Agilent) and Gene Expression Wash Buffers (Agilent), respectively. Then, slides were scanned at 5 µm resolution in an Agilent G2565AA microarray scanner, and image files were analyzed with the Feature Extraction software 9.5.1. We used the GeneSpring 12.1 software (Agilent) to perform the interarray analyses. To ensure a high-quality data set we removed control features, and selected only features for which the 'IsWellAboveBG' parameter was one in at least three out of four biological replicates (31,921 features from 43,803). Our microarray data sets have been submitted to the Gene Expression Omnibus database (accession number: GSE116592).

A new gene annotation of probes in the microarray was carried out using BLASTN searches (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>), using the sequences of each probe as query against the *Arabidopsis* genome annotation in The *Arabidopsis* Information Resource (TAIR; [www.arabidopsis.org](http://www.arabidopsis.org)), version 10. BLAST results for each probe were filtered with a minimum *E*-value of  $9.9 \times 10^{-6}$ , a minimum sequence identity of 98% between probe and transcript, and a minimum overlap of the 75% of the probe sequence length. Probes matching multiple genes were not considered. Results for this gene annotation are quite similar to those obtained in similar analyses performed by TAIR (<ftp://ftp.arabidopsis.org/Microarrays/Agilent/>).

### 4.3. Gene Overexpression/Repression Analysis

For each probe and experimental condition (three conditions: 0 h at 22 °C, 24 h at 22 °C, and 24 h at 37 °C), expression levels were averaged across the four biological replicates. For those genes that mapped to more than one probe, expression levels were averaged across all probes. As a result, a single expression level was obtained for each gene and experimental condition.

For each gene with available probes ( $n = 20,491$ ), the response ( $R$ ) of its expression to heat stress was computed as:

$$R = \log_2 \frac{E_{24,37}}{E_{24,22}} \quad (1)$$

where  $E_{24,37}$  is expression level at 37 °C at 24 h, and  $E_{24,22}$  is expression level at 22 °C at 24 h.  $R$  takes positive values for genes that are overexpressed at 37 °C compared to 22 °C, and negative values for those that are repressed.

#### 4.4. Protein and Gene Sequence Analysis

All *Arabidopsis* protein sequences were obtained from Ensembl Plants [38] (assembly: TAIR10). For each gene encoding multiple proteins (alternative splicing isoforms), the longest protein was selected for analysis. For each protein, the frequency of each amino acid was computed by dividing the number of occurrences of the amino acid by the length of the protein. GC content of each gene was retrieved from Ensembl Plants' Biomart [38,39]. For each protein, the most likely subcellular location was retrieved from the SUBA4 database [40]. The consensus location was used. Only proteins located to a single compartment were used in compartment-specific analyses.

#### 4.5. Prediction of Protein Intrinsic Disorder

Protein intrinsic disorder prediction was carried out using IUPred [28] for regions of disorder of at least 30 amino acids ("long" option). IUPred predicts tendency for polypeptide chains to be intrinsically disordered or ordered by analyzing the composition of amino acids within a window of 30 consecutive amino acids. It does so by utilizing an energy predictor matrix to estimate the tendency for pairs of amino acids to form strong stabilizing connections, the underlying assumption being that globular proteins form strong stabilizing contacts whereas structurally disordered proteins lack this capacity. IUPred reports a disorder score for each residue ranging from 0 to 1, conferring complete order to disorder, respectively. In this study, we used a threshold of >0.4 to calculate the proportion of amino acids within each protein that were likely to be in disordered regions.

#### 4.6. Statistical Analyses

Statistical analyses were conducted using R [41]. Partial correlation analyses were conducted using the R function `pcor.test` [42]. Tests repeated on all 20 amino acids were corrected for multiple testing using the Benjamini-Hochberg approach [43].

**Supplementary Materials:** Supplementary materials can be found at <http://www.mdpi.com/1422-0067/19/8/2276/s1>.

**Author Contributions:** M.X.R.-G., F.V.-S., M.A.P.-A. and M.A.F. conceived, designed, and conducted the experiments and microarray analyses. D.A.-P. conceived and designed the bioinformatics analyses. D.A.-P. and F.F. conducted the bioinformatics analyses. D.A.-P. and M.X.R.-G. wrote the paper. All authors contributed to editing the manuscript, read, and approved the manuscript, except M.A.F., who is a posthumous author.

**Funding:** D.A.-P. and F.F. were supported by funds from the University of Nevada, Reno, and by pilot grants from Nevada INBRE (P20GM103440) and the Smooth Muscle Plasticity COBRE from the University of Nevada, Reno (5P30GM110767-04), both funded by the National Institute of General Medical Sciences (National Institutes of Health). M.X.R.-G. and M.A.F. were supported by grants from Science Foundation Ireland (12/IP/1637) and the Spanish Ministerio de Economía y Competitividad, Spain (MINECO-FEDER; BFU201236346 and BFU2015-66073-P) to MAF. MXRG was supported by a JAE DOC fellowship from the MINECO, Spain. F.V.-S. and M.A.P.-A. were supported by grant BIO2014-55946-P from MINECO-FEDER.

**Acknowledgments:** D.A.-P., M.X.R.-G., F.V.-S., F.F. and M.A.P.-A. dedicate this work to the memory of M.A.F. Current address of F.F.: Structural Genomics Consortium; Target Discovery Institute, Nuffield Department of Medicine, University of Oxford, Oxford OX3 7FZ, United Kingdom.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

## Abbreviations

IDR	Intrinsically Disordered Region
MS	Murashige and Skoog
TAIR	The Arabidopsis Information Resource
BLAST	Basic Local Alignment Search Tool

## References

1. Karshikoff, A.; Ladenstein, R. Ion pairs and the thermotolerance of proteins from hyperthermophiles: A ‘traffic rule’ for hot roads. *Trends Biochem. Sci.* **2001**, *26*, 550–557. [[CrossRef](#)]
2. Strop, P.; Mayo, S.L. Contribution of surface salt bridges to protein stability. *Biochemistry* **2000**, *39*, 1251–1255. [[CrossRef](#)] [[PubMed](#)]
3. Perutz, M.; Raidt, H. Stereochemical basis of heat stability in bacterial ferredoxins and in haemoglobin A2. *Nature* **1975**, *255*, 256–259. [[CrossRef](#)] [[PubMed](#)]
4. Argos, P.; Rossmann, M.G.; Grau, U.M.; Zuber, H.; Frank, G.; Tratschin, J.D. Thermal stability and protein structure. *Biochemistry* **1979**, *18*, 5698–5703. [[CrossRef](#)] [[PubMed](#)]
5. Beeby, M.; D O’Connor, B.; Ryttersgaard, C.; Boutz, D.R.; Perry, L.J.; Yeates, T.O. The genomics of disulfide bonding and protein stabilization in thermophiles. *PLoS Biol.* **2005**, *3*, e309. [[CrossRef](#)] [[PubMed](#)]
6. Farias, S.T.; Bonato, M. Preferred amino acids and thermostability. *Genet. Mol. Res.* **2003**, *2*, 383–393. [[PubMed](#)]
7. Haney, P.J.; Badger, J.H.; Buldak, G.L.; Reich, C.I.; Woese, C.R.; Olsen, G.J. Thermal adaptation analyzed by comparison of protein sequences from mesophilic and extremely thermophilic *Methanococcus* species. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 3578–3583. [[CrossRef](#)] [[PubMed](#)]
8. Kreil, D.P.; Ouzounis, C.A. Identification of thermophilic species by the amino acid compositions deduced from their genomes. *Nucleic Acids Res.* **2001**, *29*, 1608–1615. [[CrossRef](#)] [[PubMed](#)]
9. Tekaiia, F.; Yeramian, E.; Dujon, B. Amino acid composition of genomes, lifestyles of organisms, and evolutionary trends: A global picture with correspondence analysis. *Gene* **2002**, *297*, 51–60. [[CrossRef](#)]
10. Zeldovich, K.B.; Berezovsky, I.N.; Shakhnovich, E.I. Protein and DNA sequence determinants of thermophilic adaptation. *PLoS Comput. Biol.* **2007**, *3*, e5. [[CrossRef](#)] [[PubMed](#)]
11. Chakravarty, S.; Varadarajan, R. Elucidation of determinants of protein stability through genome sequence analysis. *FEBS Lett.* **2000**, *470*, 65–69. [[CrossRef](#)]
12. Cambillau, C.; Claverie, J.-M. Structural and genomic correlates of hyperthermostability. *J. Biol. Chem.* **2000**, *275*, 32383–32386. [[CrossRef](#)] [[PubMed](#)]
13. Burra, P.V.; Kalmar, L.; Tompa, P. Reduction in structural disorder and functional complexity in the thermal adaptation of prokaryotes. *PLoS ONE* **2010**, *5*, e12069. [[CrossRef](#)] [[PubMed](#)]
14. Wang, J.; Yang, Y.; Cao, Z.; Li, Z.; Zhao, H.; Zhou, Y. The role of semidisorder in temperature adaptation of bacterial FlgM proteins. *Biophys. J.* **2013**, *105*, 2598–2605. [[CrossRef](#)] [[PubMed](#)]
15. Vicedo, E.; Schlessinger, A.; Rost, B. Environmental pressure may change the composition protein disorder in prokaryotes. *PLoS ONE* **2015**, *10*, e0133990. [[CrossRef](#)] [[PubMed](#)]
16. Galea, C.A.; High, A.A.; Obenauer, J.C.; Mishra, A.; Park, C.-G.; Punta, M.; Schlessinger, A.; Ma, J.; Rost, B.; Slaughter, C.A. Large-scale analysis of thermostable, mammalian proteins provides insights into the intrinsically disordered proteome. *J. Proteome Res.* **2008**, *8*, 211–226. [[CrossRef](#)] [[PubMed](#)]
17. Tsvetkov, P.; Myers, N.; Moscovitz, O.; Sharon, M.; Prilusky, J.; Shaul, Y. Thermo-resistant intrinsically disordered proteins are efficient 20S proteasome substrates. *Mol. Biosyst.* **2012**, *8*, 368–373. [[CrossRef](#)] [[PubMed](#)]
18. Galea, C.A.; Nourse, A.; Wang, Y.; Sivakolundu, S.G.; Heller, W.T.; Kriwacki, R.W. Role of intrinsic flexibility in signal transduction mediated by the cell cycle regulator, p27<sup>Kip1</sup>. *J. Mol. Biol.* **2008**, *376*, 827–838. [[CrossRef](#)] [[PubMed](#)]

19. Van Noort, V.; Bradatsch, B.; Arumugam, M.; Amlacher, S.; Bange, G.; Creevey, C.; Falk, S.; Mende, D.R.; Sinning, I.; Hurt, E. Consistent mutational paths predict eukaryotic thermostability. *BMC Evol. Biol.* **2013**, *13*, 7. [CrossRef] [PubMed]
20. Wang, G.-Z.; Lercher, M.J. Amino acid composition in endothermic vertebrates is biased in the same direction as in thermophilic prokaryotes. *BMC Evol. Biol.* **2010**, *10*, 263. [CrossRef] [PubMed]
21. Windisch, H.S.; Lucassen, M.; Frickenhaus, S. Evolutionary force in confamiliar marine vertebrates of different temperature realms: Adaptive trends in zoarcid fish transcriptomes. *BMC Genom.* **2012**, *13*, 549. [CrossRef] [PubMed]
22. Albanese, V.; Yam, A.Y.-W.; Baughman, J.; Parnot, C.; Frydman, J. Systems analyses reveal two chaperone networks with distinct functions in eukaryotic cells. *Cell* **2006**, *124*, 75–88. [CrossRef] [PubMed]
23. Berry, J.; Bjorkman, O. Photosynthetic response and adaptation to temperature in higher plants. *Annu. Rev. Plant Physiol.* **1980**, *31*, 491–543. [CrossRef]
24. Sueoka, N. Correlation between base composition of deoxyribonucleic acid and amino acid composition of protein. *Proc. Natl. Acad. Sci. USA* **1961**, *47*, 1141–1149. [CrossRef] [PubMed]
25. Cherry, J.L. Highly expressed and slowly evolving proteins share compositional properties with thermophilic proteins. *Mol. Biol. Evol.* **2010**, *27*, 735–741. [CrossRef] [PubMed]
26. Nakashima, H.; Nishikawa, K. The amino acid composition is different between the cytoplasmic and extracellular sides in membrane proteins. *FEBS Lett.* **1992**, *303*, 141–146. [PubMed]
27. Nakashima, H.; Nishikawa, K. Discrimination of intracellular and extracellular proteins using amino acid composition and residue-pair frequencies. *J. Mol. Biol.* **1994**, *238*, 54–61. [CrossRef] [PubMed]
28. Dosztanyi, Z.; Csizmok, V.; Tompa, P.; Simon, I. IUPred: Web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **2005**, *21*, 3433–3434. [CrossRef] [PubMed]
29. Peng, Z.; Uversky, V.N.; Kurgan, L. Genes encoding intrinsic disorder in eukaryota have high GC content. *Intrinsically Disord. Proteins* **2016**, *4*, e1262225. [CrossRef] [PubMed]
30. Yruela, I.; Contreras-Moreira, B. Genetic recombination is associated with intrinsic disorder in plant proteomes. *BMC Genom.* **2013**, *14*, 772. [CrossRef] [PubMed]
31. Paliy, O.; Gargac, S.M.; Cheng, Y.; Uversky, V.N.; Dunker, A.K. Protein disorder is positively correlated with gene expression in *Escherichia coli*. *J. Proteome Res.* **2008**, *7*, 2234–2245. [CrossRef] [PubMed]
32. Singh, G.P.; Dash, D. How expression level influences the disorderiness of proteins. *Biochem. Biophys. Res. Commun.* **2008**, *371*, 401–404. [CrossRef] [PubMed]
33. Yang, J.R.; Liao, B.Y.; Zhuang, S.M.; Zhang, J. Protein misinteraction avoidance causes highly expressed proteins to evolve slowly. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, E831–840. [CrossRef] [PubMed]
34. Hendsch, Z.S.; Tidor, B. Do salt bridges stabilize proteins? A continuum electrostatic analysis. *Protein Sci.* **1994**, *3*, 211–226. [CrossRef] [PubMed]
35. Zhou, X.-X.; Wang, Y.-B.; Pan, Y.-J.; Li, W.-F. Differences in amino acids composition and coupling patterns between mesophilic and thermophilic proteins. *Amino Acids* **2008**, *34*, 25–33. [CrossRef] [PubMed]
36. Catanzano, F.; Barone, G.; Graziano, G.; Capasso, S. Thermodynamic analysis of the effect of selective monodeamidation at asparagine 67 in ribonuclease a. *Protein Sci.* **1997**, *6*, 1682–1693. [CrossRef] [PubMed]
37. Charlesworth, B. Fundamental concepts in genetics: Effective population size and patterns of molecular evolution and variation. *Nat. Rev. Genet.* **2009**, *10*, 195–205. [CrossRef] [PubMed]
38. Bolser, D.; Staines, D.M.; Pritchard, E.; Kersey, P. Ensembl Plants: Integrating tools for visualizing, mining, and analyzing plant genomics data. *Methods Mol. Biol.* **2016**, *1374*, 115–140. [PubMed]
39. Kasprzyk, A.; Keefe, D.; Smedley, D.; London, D.; Spooner, W.; Melsopp, C.; Hammond, M.; Rocca-Serra, P.; Cox, T.; Birney, E. EnsMart: A generic system for fast and flexible access to biological data. *Genome Res.* **2004**, *14*, 160–169. [CrossRef] [PubMed]
40. Hooper, C.M.; Castleden, I.R.; Tanz, S.K.; Aryamanesh, N.; Millar, A.H. SUBA4: The interactive data analysis centre for arabidopsis subcellular protein locations. *Nucleic Acids Res.* **2016**, *45*, D1064–D1074. [CrossRef] [PubMed]
41. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Available online: <http://www.R-project.org/> (accessed on 31 October 2014).



42. Kim, S. Ppcor: An R package for a fast calculation to semi-partial correlation coefficients. *Commun. Stat. Appl. Methods* **2015**, *22*, 665. [[CrossRef](#)] [[PubMed](#)]
43. Benjamini, Y.; Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **1995**, *57*, 289–300.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).