Table S1. Pearson's correlation coefficients for correlation between different estimations of amino acid metabolic costs. The correlation coefficients range from 0.09 to 0.99 (average 0.64, standard deviation 0.29). 1: (Akashi and Gojobori 2002) (*E. coli*, Average of costs for growth on glucose, acetate and malate). 2: (Akashi and Gojobori 2002) (*E. coli*, growth on glucose). 3: (Akashi and Gojobori 2002) (*E. coli*, growth on acetate). 4: (Akashi and Gojobori 2002) (*E. coli*, growth on malate). 5: (Sajitz-Hermstein and Nikoloski 2010) (*A. thaliana*, day). 6: (Sajitz-Hermstein and Nikoloski 2010) (*A. thaliana*, night). 7: (Craig and Weber 1998) (*E. coli*, growth on glucose). 8: (Seligmann 2003) (any species, cost is amino acid molecular weight). 9: (Barton, et al. 2010) (*S. cerevisiae*, growth on glucose, $A_{glucose}$ calculation). 10: (Raiford, et al. 2012) (aerobic heterotrophs, growth on glucose). 11: (Raiford, et al. 2012) (anaerobic heterotrophs, growth on glucose). 12: (Raiford, et al. 2012) (aerobic phototrophs, growth on glucose). 13: (Raiford, et al. 2012) (anaerobic phototrophs, growth on glucose).

Cost estimation	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**
1	1.00	0.99	0.99	0.99	0.85	0.59	0.87	0.64	0.92	0.91	0.29	0.92	0.31
2		1.00	0.97	0.99	0.81	0.54	0.86	0.62	0.89	0.93	0.33	0.93	0.35
3			1.00	0.98	0.89	0.66	0.87	0.67	0.94	0.87	0.25	0.88	0.26
4				1.00	0.83	0.56	0.86	0.63	0.90	0.92	0.31	0.93	0.33
5					1.00	0.86	0.72	0.81	0.96	0.74	0.15	0.75	0.16
6						1.00	0.43	0.75	0.81	0.47	0.14	0.47	0.15
7							1.00	0.43	0.74	0.86	0.09	0.87	0.10
8								1.00	0.80	0.48	0.24	0.48	0.25
9									1.00	0.78	0.27	0.79	0.29
10										1.00	0.26	0.99	0.27
11											1.00	0.25	0.99
12												1.00	0.26
13													1.00

Table S2. Pearson's correlation coefficients between amino acid relative abundances and amino acid metabolic cost for Dataset DS1.

Reference	Species	Condition/nutrient/calculation	Cost(ATP) vs. abundance	Cost(ATP) vs. ln(abundance)	Cost(ATP/time) vs. abundance	Cost(ATP/time) vs. ln(abundance)
(Akashi and Gojobori 2002)	*E. coli*	Average of costs for growth on glucose, acetate and malate	-0.46	-0.52	-0.72	-0.86
(Akashi and Gojobori 2002)	*E. coli*	Glucose	-0.47	-0.53	-0.72	-0.87
(Akashi and Gojobori 2002)	*E. coli*	Acetate	-0.43	-0.50	-0.72	-0.86
(Akashi and Gojobori 2002)	*E. coli*	Malate	-0.47	-0.53	-0.80	-0.87
(Sajitz-Hermstein and Nikoloski 2010)	*A. thaliana*	Day	-0.32	-0.33	-0.75	-0.85
(Sajitz-Hermstein and Nikoloski 2010)	*A. thaliana*	Night	-0.22	-0.21	-0.75	-0.83
(Craig and Weber 1998)	*E. coli*	Glucose	-0.33	-0.42	-0.66	-0.82
(Seligmann 2003)	Any	Amino acid molecular weight	-0.47	-0.46	-0.74	-0.85
(Barton, et al. 2010)	*S. cerevisiae*	Glucose ($A_{glucose}$)	-0.40	-0.43	-0.75	-0.86
(Raiford, et al. 2012)	Aerobic heterotroph	Glucose	-0.32	-0.40	-0.70	-0.84
(Raiford, et al. 2012)	Anaerobic heterotroph	Glucose	-0.45	-0.48	-0.60	-0.73
(Raiford, et al. 2012)	Aerobic phototroph	Glucose	-0.33	-0.40	-0.70	-0.84
(Raiford, et al. 2012)	Anaerobic phototroph	Glucose	-0.48	-0.50	-0.62	-0.74
Main text	Any	Genetic code model	0.71	0.71	0.71	0.71

Table S3. Pearson's correlation coefficients between amino acid relative abundances and amino acid metabolic cost for Dataset DS2.

Reference	Species	Condition/nutrient/calculation	Cost(ATP) vs. abundance	Cost(ATP) vs. ln(abundance)	Cost(ATP/time) vs. abundance	Cost(ATP/time) vs. ln(abundance)
(Akashi and Gojobori 2002)	*E. coli*	Average of costs for growth on glucose, acetate and malate	-0.58	-0.62	-0.80	-0.91
(Akashi and Gojobori 2002)	*E. coli*	Glucose	-0.59	-0.63	-0.79	-0.91
(Akashi and Gojobori 2002)	*E. coli*	Acetate	-0.55	-0.59	-0.79	-0.91
(Akashi and Gojobori 2002)	*E. coli*	Malate	-0.59	-0.63	-0.80	-0.91
(Sajitz-Hermstein and Nikoloski 2010)	*A. thaliana*	Day	-0.44	-0.42	-0.82	-0.89
(Sajitz-Hermstein and Nikoloski 2010)	*A. thaliana*	Night	-0.33	-0.28	-0.81	-0.86
(Craig and Weber 1998)	*E. coli*	Glucose	-0.45	-0.53	-0.74	-0.88
(Seligmann 2003)	Any	Amino acid molecular weight	-0.58	-0.53	-0.79	-0.87
(Barton, et al. 2010)	*S. cerevisiae*	Glucose ($A_{glucose}$)	-0.52	-0.52	-0.82	-0.91
(Raiford, et al. 2012)	Aerobic heterotroph	Glucose	-0.45	-0.49	-0.77	-0.89
(Raiford, et al. 2012)	Anaerobic heterotroph	Glucose	-0.53	-0.51	-0.64	-0.73
(Raiford, et al. 2012)	Aerobic phototroph	Glucose	-0.45	-0.50	-0.77	-0.89
(Raiford, et al. 2012)	Anaerobic phototroph	Glucose	-0.55	-0.53	-0.66	-0.75
Main text	Any	Genetic code model	0.62	0.62	0.62	0.62

Table S4. Variation of amino acid abundances (in %) across organisms in Dataset DS1.

Amino acid	Average	Maximum	Minimum	Standard deviation	Standard deviation / Average
A	7.95	13.70	2.03	2.38	0.30
C	1.11	2.47	0.39	0.47	0.42
D	5.29	9.02	3.58	0.73	0.14
E	6.69	9.99	4.43	1.09	0.16
F	4.17	6.31	2.65	0.69	0.17
G	6.85	10.00	2.94	1.26	0.18
H	2.06	2.76	1.22	0.38	0.18
I	6.61	11.50	2.67	2.02	0.31
K	5.98	11.70	1.64	2.17	0.36
L	10.04	14.50	7.51	1.06	0.11
M	2.28	3.20	1.52	0.36	0.16
N	4.29	14.10	1.55	1.72	0.40
P	4.47	6.52	2.09	1.01	0.23
Q	3.52	5.78	1.40	1.05	0.30
R	5.36	8.74	2.67	1.43	0.27
S	6.62	9.39	3.39	1.33	0.20
T	5.25	6.83	3.75	0.68	0.13
V	6.90	10.40	3.94	1.09	0.16
W	1.15	1.72	0.49	0.27	0.23
Y	3.28	5.64	1.98	0.75	0.23

Table S5. Variation of amino acid abundances (in %) across organisms in Dataset DS2.

Amino acid	Average	Maximum	Minimum	Standard deviation	Standard deviation / Average
A	8.86	12.32	7.29	1.35	0.15
C	1.11	2.33	0.32	0.57	0.52
D	5.62	6.45	5.18	0.32	0.06
E	7.34	8.82	5.73	0.70	0.10
F	3.62	4.05	2.87	0.31	0.08
G	7.48	8.96	6.21	0.65	0.09
H	1.91	2.22	1.03	0.30	0.16
I	5.73	7.60	3.90	0.94	0.16
K	7.14	8.53	4.12	1.18	0.16
L	8.58	9.92	7.96	0.48	0.06
M	2.32	2.68	1.94	0.23	0.10
N	4.00	4.73	2.71	0.53	0.13
P	4.52	5.57	3.32	0.66	0.15
Q	3.77	4.61	2.24	0.58	0.15
R	5.06	6.08	4.06	0.56	0.11
S	6.22	7.54	4.03	0.97	0.16
T	5.55	6.26	5.02	0.30	0.05
V	7.35	8.71	6.52	0.71	0.10
W	0.92	1.37	0.56	0.19	0.20
Y	2.89	3.67	2.15	0.35	0.12

Estimation of amino acid reactivity and metabolic cost.

We have searched the scientific literature for a complete set of amino acid decay rates due to chemical reactions, to no success. In this section, we deduce a semi-quantitative reactivity ranking from previous publications and common knowledge of amino acid chemistry (Creighton 1992) (see Table below). Chemical reactivity can be defined as the rate at which a molecule undergoes a given chemical reaction and depends on the reaction conditions. Amino acids are involved in many different reactions. Thus, quantifying amino acid decay would require the experimental characterization of many different processes. If a molecule can decay along multiple pathways, the overall decay rate is the sum of the individual decay rates.

We have selected three groups of amino acid decay reactions that are biologically relevant, namely nucleophilic reactions, redox reactions and deamidation/isomerization reactions. For each group of reactions, we deduce a semi-quantitative reactivity ranking and use it to assign decay rates. We then add up all decay rates for each amino acid to calculate its overall decay rate. As a final input, we correct the overall decay rate to take into account amino acid abundances in fossil samples. The results presented in the main text are robust towards minor modifications of the assigned decay rates. The physiological relevance of our ranking is supported by the presence of energy-consuming enzymatic pathways that have evolved to protect proteins against chemical decay (Moskovitz, et al. 1997; Reissner and Aswad 2003; Stadtman 2006; Stroher and Millar 2012). These enzymes act on both free amino acid pools and amino acids within proteins (Brot, et al. 1981; Brot, et al. 1984; Henzel, et al. 1989; Skinner, et al. 2000; Biteau, et al. 2003).

Amino acid	Nucleophilicity	Redox reactivity	Deamidation / isomerization	Correction from abundance in fossil samples	Decay (1/time)
A	1	1	1	-2	**1**
C	14	15	1	0	**30**
D	2	1	6	0	**9**
E	2	2	1	0	**5**
F	1	2	1	0	**4**
G	1	1	1	-2	**1**
H	7	6	1	0	**14**
I	1	1	1	-1	**2**
K	5	2	1	0	**8**
L	1	1	1	-1	**2**
M	2	10	1	0	**13**
N	2	1	7	0	**10**
P	1	2	1	-1	**3**
Q	2	1	5	0	**8**
R	1	2	1	0	**4**
S	3	2	1	0	**6**
T	3	2	1	0	**6**
V	1	1	1	-1	**2**
W	2	9	1	0	**12**
Y	3	3	1	0	**7**

Nucleophilic reactions

The general principles of nucleophilicity theory described for small organic molecules may be used to set a relative order of nucleophilicity for the major chemical groups in proteins (Edwards and Pearson 1962). Nucleophilicity can be defined as the ability of the side chain to donate a pair of electrons to form a covalent bond. Good nucleophilic molecules possess electron-rich atoms that are able to donate a pair of electrons, such as oxygen, nitrogen or sulfur in proteins. Molecules that are devoid of these atoms are considered non reactive for nucleophilic reactions. Nucleophilicity depends on the excess electron density of the donor atom. It increases with if the electron-rich atom has a negative charge, so that basicity has a large effect on nucleophilicity. On the other hand, nucleophilicity decreases if the electrons to be donated are delocalized by resonance.

Among amino acids, the side chains of A, F, G, I, L, P and V are the least nucleophilic due to the lack of oxygen, nitrogen and sulphur atoms. They are assigned a decay rate of 1 for nucleophilic reactions (see Table above).

D, E, M, N, Q, R, S, T, R, W and Y are poor to very poor nucleophiles at neutral pH, for a combination of different reasons. In the case of D, E, N, Q, R, W and Y the electron density for the electron-rich atoms of the side chain is delocalized by resonance. N, Q, R, S, T and Y are fully protonated at physiological pH, which in the case of R leads to a net positive charge that makes it a very poor nucleophile. S, T and Y are slightly more nucleophilic than D, E, M, N, Q, R and W due to their higher basicity. We assign a decay rate of 1 to R, a decay rate of 2 to D, E, M, N, Q and W and a decay rate of 3 to S, T and Y for nucleophilic reactions.

The epsilon amine atom of K is a strong nucleophile in its neutral form. A finite fraction of unprotonated lysines are present at neutral pH, which react readily with alkylating, acylating and amidating reagents. We assign a decay rate of 5 to K for nucleophilic reactions. Next in nucleophilicity is the imidazole group of histidine, with a pK close to 7 and a stronger base than K. H is involved in many nucleophilic reactions in biology. We assign a decay rate of 7 to H for nucleophilic reactions. Finally, C is by far the the most reactive amino acid toward nucleophilic reactions (Bulaj, et al. 1998). Its thiol group ionizes with a pK close to 9 and rapidly reacts with many electrophilic molecules (Gurd 1972; Ferrer-Sueta, et al. 2011). Cysteine

reactivity is hallmark property of this amino acid and determines its conservation and activity in proteins (Marino and Gladyshev ; Miseta and Csutora 2000). We assign a decay rate of 14 to C for nucleophilic reactions. The series C > H > K is in line with the observed reactivity for these amino acids with haloacetyl compounds (Crestfield, et al. 1963; Gurd 1972).

Redox reactions

The second group of reactions is electron transfer processes. Living organisms are exposed to an oxidative environment due to oxygen and ultraviolet radiation (Stadtman 2006). It is possible to set a relative redox reactivity scale based on the tendency to react with reactive oxygen species, reactive nitrogen species and reactive halogen species (Berlett and Stadtman 1997; Roeser, et al. 2010).

Despite the intrinsic complexity of unifying multiple redox chemistries into a single scale, based on well-established experimental results, we propose that the amino acids A, D, G, I, L, N, Q and V react only weakly with mild oxidants. They are assigned a decay rate of 1 for redox reactions (see Table above). E, F, K, P, R, S, T present low propensity for oxidation and are assigned a decay rate of 2 for redox reactions (Stadtman 1993; Berlett and Stadtman 1997; Stadtman 2006). Next, aromatic amino acids H, Y and W are highly reactive towards oxidation (Berlett and Stadtman 1997; Stadtman 2006). Y is the least reactive, H shows intermediate reactivity and W is very prone to oxidation. We assign a decay rate of 3 to Y, a decay rate of 6 to H and a decay rate of 9 to W for redox reactions (Stadtman 1993; Berlett and Stadtman 1997; Stadtman 2006).

Sulfur containing amino acids C and M are the most reactive towards oxidation (Berlett and Stadtman 1997). Even in mild conditions M yields methionine sulfoxide, and C forms many oxidized compounds, including disulfide bridges. Moreover, in living organisms there are many enzymatic reactions that restore the reduced form of M and C. Thioredoxin efficiently reduces disulfide bonds in peptides and proteins and methionine sulfoxide reductases reduce methionine sulfoxide to methionine (Moskovitz, et al. 1997; Stadtman, et al. 2003). We assign a decay rate of 10 to M and a decay rate of 15 to C for redox reactions.

Deamidation/isomerization reactions

The third group of reactions is the spontaneous non-enzymatic deamidation and isomerization of D, N and Q. At neutral pH, the side chain amides of N and Q suffer the nucleophilic attack of a backbone amide to generate a cyclic imide intermediate (Geiger and Clarke 1987). Asparagine deamidates ten times faster than glutamine because the five-membered ring of the succinimidyl intermediate is more stable (Robinson and Robinson 2001). Opening of the succinimide by hydrolysis yields 70 per cent of iso-aspartic acid and 30 per cent of aspartic acid (Geiger and Clarke 1987). Aspartic acid isomerization can also occur directly from aspartic residues in proteins (Reissner and Aswad 2003). The accumulation of atypical isoaspartyl residues during aging can be detrimental for protein function, and many living organisms posses a selective enzyme (L-isoaspartyl O-methyltransferase, PIMT) to repair damaged aspartyl residues (Reissner and Aswad 2003; Zhu, et al. 2006). Note that the enzyme PIMT fully repairs aspartic isomerization but is unable to restore the amide group of deamidated asparagine for which, deamidation causes an irreversible damage. We assign a decay rate of 5 to Q, a decay rate of 6 to D and a decay rate of 7 to N for deamidation/isomerization reactions.

Correction for amino acid abundance in fossil samples

The decay rate of the more stable amino acids, A, F, G, I, L, P and V are low and therefore difficult to assess in a laboratory over short timescales. Over long timescales, amino acids with a lower decay rate should be more abundant in a sample. We have used amino acid abundances in fossil samples (Wang, et al. 2012) to estimate the relative decay rates of A, F, G, I, L, P and V. A and G are more abundant in fossil samples than I, L, P and V, while F is the least abundant of this group. Taking this into account, we subtract 2 units from the decay rate of A and G and 1 unit from the decay rate of I, L, P and V (see Table above).

References

Akashi H, Gojobori T. 2002. Metabolic efficiency and amino acid composition in the proteomes of Escherichia coli and Bacillus subtilis. Proc Natl Acad Sci U S A 99:3695-3700.

Barton MD, Delneri D, Oliver SG, Rattray M, Bergman CM. 2010. Evolutionary systems biology of amino acid biosynthetic cost in yeast. PLoS One 5:e11935.

Berlett BS, Stadtman ER. 1997. Protein oxidation in aging, disease, and oxidative stress. J Biol Chem 272:20313-20316.

Biteau B, Labarre J, Toledano MB. 2003. ATP-dependent reduction of cysteine-sulphinic acid by S. cerevisiae sulphiredoxin. Nature 425:980-984.

Brot N, Fliss H, Coleman T, Weissbach H. 1984. Enzymatic reduction of methionine sulfoxide residues in proteins and peptides. Methods Enzymol 107:352-360.

Brot N, Weissbach L, Werth J, Weissbach H. 1981. Enzymatic reduction of protein-bound methionine sulfoxide. Proc Natl Acad Sci U S A 78:2155-2158.

Bulaj G, Kortemme T, Goldenberg DP. 1998. Ionization-reactivity relationships for cysteine thiols in polypeptides. Biochemistry 37:8965-8972.

Craig CL, Weber RS. 1998. Selection costs of amino acid substitutions in ColE1 and ColIa gene clusters harbored by Escherichia coli. Mol Biol Evol 15:774-776.

Creighton TE. 1992. Proteins: Structures and Molecular Properties: W. H. Freeman.

Crestfield AM, Stein WH, Moore S. 1963. Alkylation and identification of the histidine residues at the active site of ribonuclease. J Biol Chem 238:2413-2419.

Edwards J, Pearson R. 1962. The Factors Determining Nucleophilic Reactivities. Journal of The American Chemical Society 84:16-24.

Ferrer-Sueta G, Manta B, Botti H, Radi R, Trujillo M, Denicola A. 2011. Factors affecting protein thiol reactivity and specificity in peroxide reduction. Chem Res Toxicol 24:434-450.

Geiger T, Clarke S. 1987. Deamidation, isomerization, and racemization at asparaginyl and aspartyl residues in peptides. Succinimide-linked reactions that contribute to protein degradation. J Biol Chem 262:785-794.

Gurd FR. 1972. [34a] carboxymethylation. Methods Enzymol 25:424-438.

Henzel WJ, Stults JT, Hsu CA, Aswad DW. 1989. The primary structure of a protein carboxyl methyltransferase from bovine brain that selectively methylates L-isoaspartyl sites. J Biol Chem 264:15905-15911.

Marino SM, Gladyshev VN. Cysteine function governs its conservation and degeneration and restricts its utilization on protein surfaces. J Mol Biol 404:902-916.

Miseta A, Csutora P. 2000. Relationship between the occurrence of cysteine in proteins and the complexity of organisms. Mol Biol Evol 17:1232-1239.

Moskovitz J, Berlett BS, Poston JM, Stadtman ER. 1997. The yeast peptide-methionine sulfoxide reductase functions as an antioxidant in vivo. Proc Natl Acad Sci U S A 94:9585-9589.

Raiford DW, Heizer EM, Jr., Miller RV, Doom TE, Raymer ML, Krane DE. 2012. Metabolic and translational efficiency in microbial organisms. J Mol Evol 74:206-216.

Reissner KJ, Aswad DW. 2003. Deamidation and isoaspartate formation in proteins: unwanted alterations or surreptitious signals? Cell Mol Life Sci 60:1281-1295.

Robinson NE, Robinson AB. 2001. Deamidation of human proteins. Proc Natl Acad Sci U S A 98:12409-12413.

Roeser J, Bischoff R, Bruins AP, Permentier HP. 2010. Oxidative protein labeling in mass-spectrometry-based proteomics. Anal Bioanal Chem 397:3441-3455.

Sajitz-Hermstein M, Nikoloski Z. 2010. A novel approach for determining environment-specific protein costs: the case of Arabidopsis thaliana. Bioinformatics 26:i582-588.
Seligmann H. 2003. Cost-minimization of amino acid usage. J Mol Evol 56:151-161.
Skinner MM, Puvathingal JM, Walter RL, Friedman AM. 2000. Crystal structure of protein isoaspartyl methyltransferase: a catalyst for protein repair. Structure 8:1189-1201.
Stadtman ER. 1993. Oxidation of free amino acids and amino acid residues in proteins by radiolysis and by metal-catalyzed reactions. Annu Rev Biochem 62:797-821.
Stadtman ER. 2006. Protein oxidation and aging. Free Radic Res 40:1250-1258.
Stadtman ER, Moskovitz J, Levine RL. 2003. Oxidation of methionine residues of proteins: biological consequences. Antioxid Redox Signal 5:577-582.
Stroher E, Millar AH. 2012. The biological roles of glutaredoxins. Biochem J 446:333-348.
Wang SY, Cappellini E, Zhang HY. 2012. Why collagens best survived in fossils? Clues from amino acid thermal stability. Biochem Biophys Res Commun 422:5-7.
Zhu JX, Doyle HA, Mamula MJ, Aswad DW. 2006. Protein repair in the brain, proteomic analysis of endogenous substrates for protein L-isoaspartyl methyltransferase in mouse brain. J Biol Chem 281:33802-33813.