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 (1) (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 (2-5).

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 (6). 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 (7). Its thiol group ionizes with a pK close to 9 and rapidly reacts with many electrophilic molecules (8, 9). Cysteine reactivity is hallmark property of this amino acid and

determines its conservation and activity in proteins (10, 11). 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 (8, 12).

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 (2). 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 (13, 14).

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 (2, 13, 15). Next, aromatic amino acids H, Y and W are highly reactive towards oxidation (2, 13). 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 (2, 13, 15).

Sulfur containing amino acids C and M are the most reactive towards oxidation (13). 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 (4, 16). 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 (17). Asparagine deamidates ten times faster than glutamine because the five-membered ring of the succinimidyl intermediate is more stable (18). Opening of the

succinimide by hydrolysis yields 70 per cent of iso-aspartic acid and 30 per cent of aspartic acid (17). Aspartic acid isomerization can also occur directly from aspartic residues in proteins (5). 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 (5, 19). 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 (20) 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

1. Creighton TE (1992) *Proteins: Structures and Molecular Properties* (W. H. Freeman) 2nd Ed p 512.
2. Stadtman ER (2006) Protein oxidation and aging. *Free Radic Res* 40(12):1250-1258.
3. Stroher E & Millar AH (2012) The biological roles of glutaredoxins. *Biochem J* 446(3):333-348.
4. 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(18):9585-9589.
5. Reissner KJ & Aswad DW (2003) Deamidation and isoaspartate formation in proteins: unwanted alterations or surreptitious signals? *Cell Mol Life Sci* 60(7):1281-1295.
6. Edwards J & Pearson R (1962) The Factors Determining Nucleophilic Reactivities. *Journal of The American Chemical Society* 84(1):16-24.

7. Bulaj G, Kortemme T, & Goldenberg DP (1998) Ionization-reactivity relationships for cysteine thiols in polypeptides. *Biochemistry* 37(25):8965-8972.
8. Gurd FR (1972) [34a] carboxymethylation. *Methods Enzymol* 25:424-438.
9. Ferrer-Sueta G*, et al.* (2011) Factors affecting protein thiol reactivity and specificity in peroxide reduction. *Chem Res Toxicol* 24(4):434-450.
10. Marino SM & Gladyshev VN (Cysteine function governs its conservation and degeneration and restricts its utilization on protein surfaces. *Journal of molecular biology* 404(5):902-916.
11. Miseta A & Csutora P (2000) Relationship between the occurrence of cysteine in proteins and the complexity of organisms. *Molecular biology and evolution* 17(8):1232-1239.
12. Crestfield AM, Stein WH, & Moore S (1963) Alkylation and identification of the histidine residues at the active site of ribonuclease. *The Journal of biological chemistry* 238:2413-2419.
13. Berlett BS & Stadtman ER (1997) Protein oxidation in aging, disease, and oxidative stress. *The Journal of biological chemistry* 272(33):20313-20316.
14. Roeser J, Bischoff R, Bruins AP, & Permentier HP (2010) Oxidative protein labeling in mass-spectrometry-based proteomics. *Anal Bioanal Chem* 397(8):3441-3455.
15. Stadtman ER (1993) Oxidation of free amino acids and amino acid residues in proteins by radiolysis and by metal-catalyzed reactions. *Annual review of biochemistry* 62:797-821.
16. Stadtman ER, Moskovitz J, & Levine RL (2003) Oxidation of methionine residues of proteins: biological consequences. *Antioxidants & redox signaling* 5(5):577-582.
17. Geiger T & Clarke S (1987) Deamidation, isomerization, and racemization at asparaginyl and aspartyl residues in peptides. Succinimide-linked reactions that contribute to protein degradation. *The Journal of biological chemistry* 262(2):785-794.
18. Robinson NE & Robinson AB (2001) Deamidation of human proteins. *Proc Natl Acad Sci U S A* 98(22):12409-12413.
19. 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. *The Journal of biological chemistry* 281(44):33802-33813.
20. Wang SY, Cappellini E, & Zhang HY (2012) Why collagens best survived in fossils? Clues from amino acid thermal stability. *Biochem Biophys Res Commun* 422(1):5-7.