Quarterly Reviews of Biophysics

cambridge.org/qrb

Off-pathway 3D-structure provides protection against spontaneous Asn/Asp isomerization: shielding proteins Achilles heel*

András Láng[1], Imre Jákli[1], Kata Nóra Enyedi[2], Gábor Mező[2,3], Dóra K. Menyhárd[1] and András Perczel[1,2]

[1]MTA-ELTE Protein Modelling Research Group, H-1117 Budapest, Hungary, Pázmány Péter sétány 1/A; [2]Eötvös Loránd University, Institute of Chemisty, H-1117 Budapest, Hungary, Pázmány Péter sétány 1/A and [3]MTA-ELTE Research Group of Peptide Chemistry, H-1117 Budapest, Hungary, Pázmány Péter sétány 1/A

Abstract

Spontaneous deamidation prompted backbone isomerization of Asn/Asp residues resulting in – most cases – the insertion of an extra methylene group into the backbone poses a threat to the structural integrity of proteins. Here we present a systematical analysis of how temperature, pH, presence of charged residues, but most importantly backbone conformation and dynamics affect isomerization rates as determined by nuclear magnetic resonance in the case of designed peptide-models. We demonstrate that restricted mobility (such as being part of a secondary structural element) may safeguard against isomerization, but this protective factor is most effective in the case of off-pathway folds which can slow the reaction by several magnitudes compared to their on-pathway counterparts. We show that the geometric descriptors of the initial nucleophilic attack of the isomerization can be used to classify local conformation and contribute to the design of stable protein drugs, antibodies or the assessment of the severity of mutations.

Graphical abstract:

At any –Asn/AspGly– sites in proteins a spontaneous backbone isomerization occurs within days under physiological conditions leading to various forms of proteopathy. This unwanted transformation especially harmful to long-lived proteins (e.g. hemoglobin and crystallins), can be slowed down, though never stopped, by a rigid three-dimensional protein fold, if it can delay in the conformational maze, on-pathway intermediates from occurring.

Once formed, the chemical bond network of the protein backbone remains highly stable during its lifetime. Very few spontaneous reactions are known to threaten its integrity; the deamidation prompted isomerization of Asn and Asp (and to a lesser extent of Gln and Glu) residues is one, leading to changes both in the constitution (αAsn→αAsp/βAsp and αAsp→αAsp/βAsp (aspartate/isoaspartate)) and stereochemistry (L-Asn→L-Asp/D-Asp). Completed *via* a succinimide ring formation and subsequent hydrolysis (Fig. 1), isomerization with bond-rearrangement occurs, in some cases within hours. The change in protein sequence and especially the insertion of a $-CH_2-$ group into the backbone (αAsn→βAsp) can alter its three-dimensional (3D)-structure and thus, its bioactivity profile (Uchida and Shibata, 1981): for example correlation was established between the appearance of isoaspartate and structural changes of amyloid β of Alzheimer's disease (Shimizu *et al.*, 2000), or in the Cu, Zn binding of superoxide dismutase (SOD1; D'Angelo *et al.*, 2013; Shi *et al.*, 2013) resulting in amyotrophic lateral sclerosis (for further examples see: Supplementary Table 1). Though modulated by chemical, sterical, and sequential effects (Tyler-Cross and Schirch, 1991; Kosky *et al.*, 1999, 2009; Goolcharran *et al.*, 2000; Xie *et al.*, 2000; Li *et al.*, 2005), the conversion is not restricted to a consensus sequence motif, thus in principle, any Asn or Asp can isomerize (Truscott *et al.*, 2016). The prevalence of the most sensitive –AsnGly– or –AspGly– sites is about 0.30–0.35% in the human proteome and thus, three of these 'Achilles' heels' are likely to occur in approximately every 1000 residues.

Here we apply a nuclear magnetic resonance (NMR)-based approach using designed peptide-models within different molecular scaffolds that allow the *in situ* and real-time monitoring of the reaction. Deciphering key physicochemical parameters of the isomerization, we show how an appropriately rigid backbone fold can effectively delay such deleterious transformations.

*This work is dedicated to Prof. László Gráf on the occasion of his 77th birthday who made a seminal contribution to this field almost 50 years ago.

Discovery

Cite this article: Láng A, Jákli I, Enyedi KN, Mező G, Menyhárd DK, Perczel A (2020). Off-pathway 3D-structure provides protection against spontaneous Asn/Asp isomerization: shielding proteins Achilles heel. *Quarterly Reviews of Biophysics* **53**, e2, 1–12. https://doi.org/10.1017/S003358351900009X

Received: 15 April 2019
Revised: 22 July 2019
Accepted: 12 October 2019

Key words:
CD spectroscopy; isomerization rate study; NMR spectroscopy; post translational modification; protecting effect of the secondary structure; spontaneous deamination

Author for correspondence:
András Perczel,
E-mail: perczel.andras@ttk.elte.hu

© The Author(s) 2020. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

CAMBRIDGE UNIVERSITY PRESS

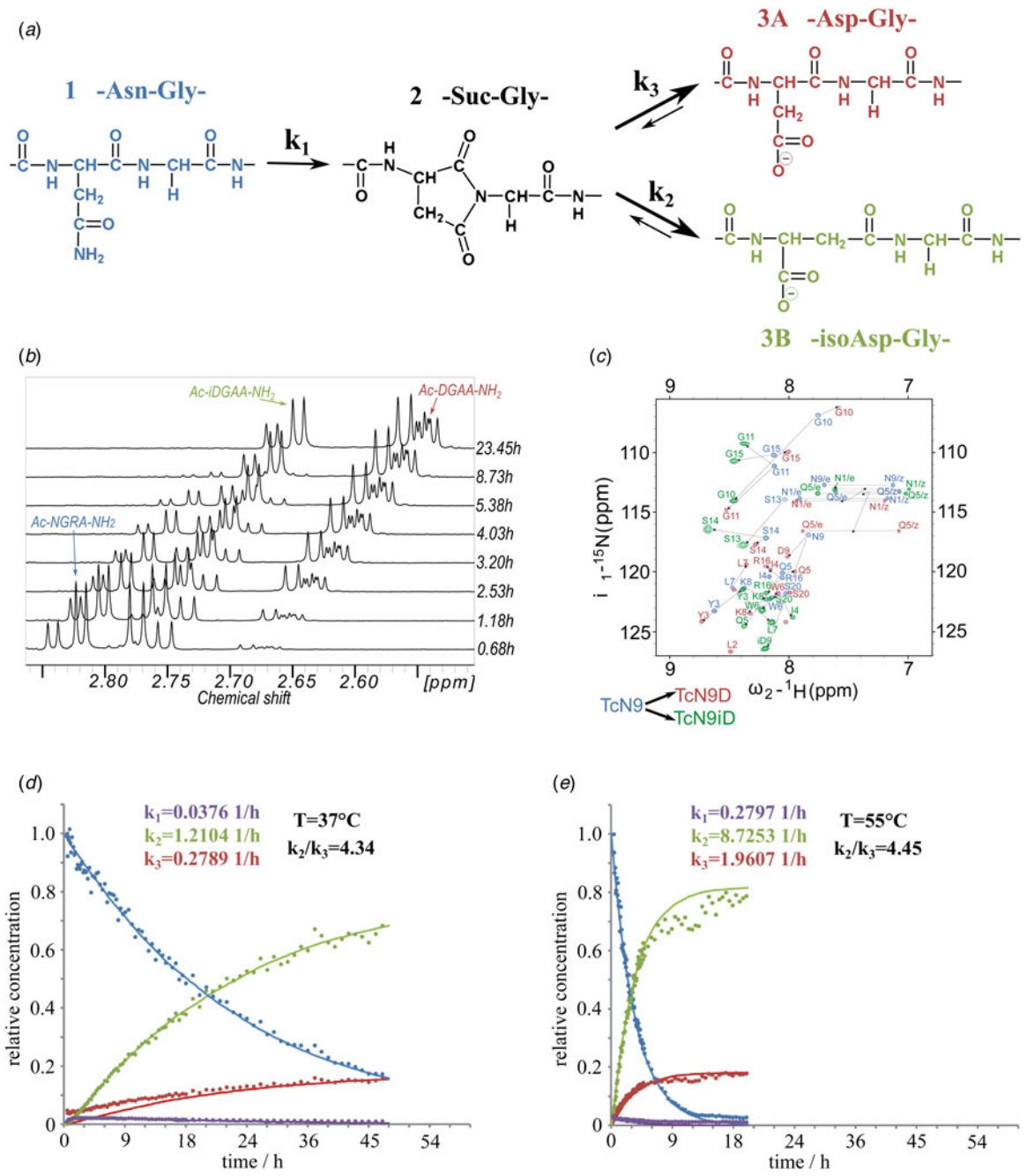


Fig. 1. (*a*) Isomerization reaction and rate constants of Ac-N**G**AA-NH_2 forming first the succinimide on-pathway intermediate (Ac-Suc-**G**AA-NH_2) thus, hydrolyzing to a product mixture of Ac-βD**G**AA-NH_2 (isoAsp derivative) and Ac-αD**G**AA-NH_2 (Asp derivative). Isomerization followed by ^{1}H- (*b*) and ^{1}H–^{15}N-NMR (*c*) (700 MHz) as a function of the time. Both isoaspartyl (Ac-$\boldsymbol{\beta}$**DG**RA-NH_2) and aspartyl (Ac-$\boldsymbol{\alpha}$**DG**RA-NH_2) as product mixtures, with the intermediate succinimide-derivative are observed by both 1D- and 2D-techniques beside the starting model system Ac-**NG**RA-NH_2 (T = 55 °C and pH = 7.4). For TcN^9 as for other proteins having longer backbone 2D/3D NMR is needed for resonance assignment and thus, for monitoring the isomerization rate. Resonances of TcN^9 (blue) shift and self-duplicate giving a mixture of TcαD^9 and TcβD^9 (red and green) within a day. (*d*) The decay of ^{1}H-NMR-signal intensities as a function of the time for selected resonances (e.g. acetyl protons: ~2 ppm), resulting in isomerization rate constant k_1, k_2, and k_3 at T = 310 K and pH = 7.8. (*e*) The same as (*d*) at T = 328 K and pH = 7.8.

Results and discussion

Effect of temperature and pH

To assess the effect of temperature (T = 28–55 °C) and pH (pH = 5.1–7.8) on the conversion rate, circular dichroism (CD) and NMR measurements were carried out focusing on both reaction kinetics and energetics. The reference peptide model, Ac-**NG**AA-NH_2, has an isomerization half-life of about a day under physiological conditions ($\tau_{\mathbf{NGAA}}^{37\,°C/pH=7.4}$ = 28.5 ± 3.9 h) (Supplementary Fig. 1, Fig. 1 and Tables 1 and 2). Lowering the pH increases τ exponentially, as exchange of the amide NH^{Gly} with hydrogens of water becomes slow: $\tau_{\mathbf{NGAA}}^{55\,°C/pH=7.8}$ = 2.7 ± 0.5 h, $\tau_{\mathbf{NGAA}}^{55\,°C/pH=7.4}$ = 4.1 ± 0.6 h, $\tau_{\mathbf{NGAA}}^{55\,°C/pH=6.3}$ = 30.2 ± 2.9 h, and $\tau_{\mathbf{NGAA}}^{55\,°C/pH=5.1}$ = 172.5 ± 14.4 h. Increasing T shortens τ exponentially (Table 2) shown in the case of the

https://doi.org/10.1017/S003358351900009X Published online by Cambridge University Press

Table 1. Isomerization half-lives (τ) and related thermodynamic data for NGAA and NGRA model systems by employing the Eyring–Polányi equation

		τ(h)				ΔH	ΔS
-Asn-Gly-sequence followed by residue Xxx-	pH	T = 28 °C	T = 37 °C	T = 46 °C	T = 55 °C	kJ mol^{-1}	kJ mol^{-1}
–Asn-Gly-**Ala**-Ala-	7.4	144.8 ± 11.0	28.5 ± 3.9	13.3 ± 1.2	4.1 ± 0.6	102.06	−0.02[a]
–Asn-Gly-**Ala**-Ala-	7.8	n./a.	20.6 ± 2.7	6.2 ± 1.1	2.8 ± 0.5	91.23	−0.05[b]
–Asn-Gly-**Arg(+)**-Ala-	7.4	n./a.	18.9 ± 5.2	10.5 ± 1.2	3.2 ± 0.8	80.48	−0.08[c]

Pearson correlation coefficients of the linear fitting are: (a) 0.98, (b) 0.99, and (c) 0.95.

Table 2. Isomerization half-lives (τ) of Asn/Asp (at pH = 7.4) built in different and tunable molecular scaffolds

Unstructured molecular scaffolds presenting high internal dynamics					Structured molecular scaffold presenting low internal dynamics	
	τ (h)			τ (h)		τ (h)
Sequence	T = 310 K	T = 328 K	Sequence	T = 328 K	Sequence	T = 328 K
Ac-A**NG**A-NH_2	37.6 ± 7.2	5.3 ± 1.2	*Ac*-**E(−)NGK(+)**-NH_2 *Ac*-**K(+)NGE(−)**-NH_2	5.1 ± 1.6[a] 6.8 ± 1.6[b]	$_{\mathbf{WVVW:=}}$ (H^+)-SWTV**E(−)** **NGK(+)**VTWK-NH_2	13.0 ± 0.6[c]
Ac-**NG**AA-NH_2	28.5 ± 3.9	4.1 ± 0.6[d]	*Ac*-**NS**AA-NH_2	46.5 ± 5.0	cyclo(**NG**AA)$_{\mathbf{1}}$	21.6 ± 5.9
Ac-**NG**R(+)A-NH_2	18.9 ± 5.2	3.2 ± 0.8[e]	*Ac*-**NS**R(+)A-NH_2	22.2 ± 2.7	cyclo(**NG**R(+)A)$_{\mathbf{1}}$	19.2 ± 5.8
Ac-**NG**K(+)A-NH_2	20.6 ± 4.3	2.8 ± 0.4	*Ac*-**K(+)NGG**-NH_2 *Ac*-**NA**AA-NH_2	4.2 ± 0.6 90.6 ± 7.4	$_{\mathbf{TcN:=}}^{\mathbf{9}}$ NLYIQWLK(+) **NG**GPSSGR(+)PPPS	5.4 ± 0.6[f]
(*H*+)-**NG**AA-NH_2	12.1 ± 1.2	2.5 ± 0.3	*Ac*-**NA**R(+)A-NH_2	39.9 ± 4.8	cyclo(**NG**AA)$_{\mathbf{2}}$	10.1 ± 2.2[g]
(*H*+)-**NG**R(+)A-NH_2	12.2 ± 3.4	2.2 ± 0.4	*Ac*-**D(−)GR(+)**A-NH_2	192.9[a] ± 25.9	$_{\mathbf{TcD9:=}}$ NLYIQWLK**D(−)** **G**GPSSGR(+)PPPS[h]	243.0[e] ± 83.1
Ac-**NG**AR(+)-NH_2	27.8 ± 1.7	4.1 ± 0.6	*Ac*-**D(−)G**AA-NH_2	137.5 ± 9.7	RA**D(−)G**AM	648 (Simonovic and Volz, 2002)[i]
Ac-**NG**HA-NH_2	28.4 ± 2.4	4.5 ± 0.6	*Ac*-**NH**AA-NH_2	35.1 ± 2.4	IS**NH**AY	192 (Midelfort and Mehler, 1972)[i]

[a] $\tau_{\mathbf{E(-)NGK(+)}}^{37\,°C/pH=7.4} \sim 32.0 \pm 8.4$ h.

[b] $\tau_{\mathbf{K(+)NGE(-)}}^{37\,°C/pH=7.4} \sim 45.2 \pm 8.1$ h.

[c] $\tau_{\mathbf{WVVW}}^{37\,°C/pH=7.4} = \sim 79.6 \pm 10.6$ h.

[d] $\tau_{\mathbf{NGAA}}^{55\,°C/pH=5.1} = \sim 172.5 \pm 14.4$ h; $\tau_{\mathbf{NGAA}}^{55\,°C/pH=6.3} = \sim 30.2 \pm 2.9$ h; $\tau_{\mathbf{NGAA}}^{55\,°C/pH=7.8} = \sim 2.8 \pm 0.5$ h.

[e] $\tau_{\mathbf{NGRA}}^{55\,°C/pH=5.1} = \sim 171.5 \pm 15.7$ h.

[f] $\tau_{\mathbf{TcN9}}^{37\,°C/pH=7.4} = \sim 40.0 \pm 5.0$ h.
[g] Experiment repeated three times.
[h] The 20-residue-long Trp-cage protein, TcD9 (Neidigh *et al.*, 2002).
[i] Values taken from these studies.

Ac-**NG**AA-NH_2 and Ac-**NG**KA-NH_2 model peptides. At a slightly elevated temperature deamidation-induced isomerization of a tetrapeptide is completed within a few hours: e.g. $\tau_{\mathbf{NGAA}}^{55\,°C/pH=7.4} = 4.1 \pm 0.6$ h. The product ratio reflected in $k_2/k_3 = 4.45 \pm 0.3$ shows that –βAsp– is more abundant than the –αAsp– product. The calculated activation energies (Table 2) also indicate that under physiological conditions the conversion reaction is indeed feasible. The determined energy differences are comparable to that reported for a peptide segment of the α_A-crystalline (Aki *et al.*, 2013).

Changing the chemical constitution at the reaction center (i) and at neighboring positions ((i + 1) and (i + 2))

Negative charge significantly slows the isomerization rate: Asp at the i^{th}-position isomerizes with a much slower rate than Asn within the same molecular framework: *Ac*-**D**$^{(-)}$**G**AA-NH_2 isomerizes >33 times slower than *Ac*-**NG**AA-NH_2 does. Next neighbor effects on the deamidation rate (position $(i + 1)$) were studied extensively (Robinson and Robinson, 2001*a*; Robinson *et al.*, 2004). The rate of isomerization was shown to be faster if the residue following the Asn/Asp is small (e.g. Gly and Ser), slower for large and hydrophobic residues (e.g. Ile, Leu, Phe, Val, and Trp) and practically impossible for Pro (Robinson and Robinson, 2001*a*; Robinson *et al.*, 2004). In accordance with mass spectrometry data, our NMR results also demonstrated that if Gly$^{(i+1)}$ is replaced by bulkier residues (Ala, Ser, and His), τ lengthens significantly: compared to that of the *Ac*-**NG**AA-NH_2 reference system, in the His < Ser < Ala order (Supplementary Table 3). Considering the $(i + 2)$ position, it was previously postulated that the His, Arg, and Lys residues at this site will increase the reaction rate considerably (Robinson and Robinson, 2004*c*), as a proton donor can approach and stabilize the negative reaction center, [–(C=O)–N–]$^{(-)}$ of the backbone amide and thus facilitate isomerization (Goolcharran *et al.*, 2000) (Supplementary Scheme 1).

https://doi.org/10.1017/S003358351900009X Published online by Cambridge University Press

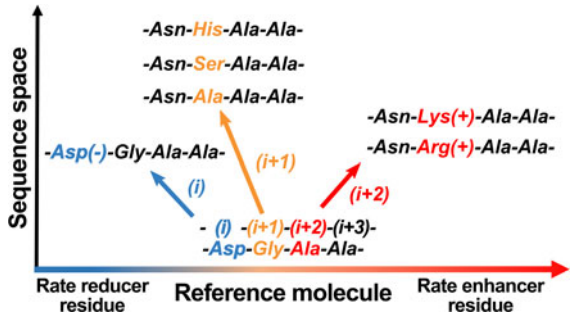


Fig. 2. Constitutional changes at i, ($i+1$), and ($i+2$) positions with respect to Asn significantly alters the isomerization rate. The introduction of a negative charge Asn→Asp(−), as well as an increase of the side-chain bulkiness at the ($i+1$) position reduces the isomerization rate by magnitude(s). In contrast, a positively charged side-chain at ($i+2$) enhances the rate of isomerization by a factor of ∼2.

Our results confirm this hypothesis – we found that if a charged residue (Arg(+) or Lys(+)) follows the –NG– motif at the ($i+2$) position, τ shortens considerably. For *Ac*-**N**GR$^{(+)}$A-NH_2 and *Ac*-**N**GK$^{(+)}$A-NH_2 shortening is ∼25 and ∼32%, respectively. If the positive charge is placed closer to the reaction center, τ gets even shorter: *cf.* (H$^+$)-**NG**AA-NH_2 (Supplementary Table 3), but the catalytic effect is diminished if it is distantly located at the ($i+3$) position (*cf.* *Ac*-NGAR$^{(+)}$-NH_2). This effect is even more enhanced when less potent ($i+1$) residues are present (e.g. Ala and Ser), the well-placed positive charge speeds up the reaction in all cases significantly: τ is shortened by ∼55% when comparing –**N**AR$^{(+)}$A– to –**N**AAA or –**N**SR$^{(+)}$A to –**N**SAA variants (Supplementary Table 3, Fig. 2). A second positive charge has a marginal effect (*cf.* (H +)-NGR$^{(+)}$A-NH$_2$). However, amphoteric His neither enhanced nor reduced isomerization rate (*cf.* *Ac*-**NG**HA-NH_2).

It is important to note that the above is only true in the case of Asn as the i^{th} residue, the isomerization rate of Asp$^{(-)}$ is not enhanced by a positive charge at the ($i+2$) position, (compare *Ac*-DGAA-NH_2 and *Ac*-DGRA-NH_2 in Supplementary Table 3), most likely due to salt-bridge formation between long, flexible side chains carrying a positive charge and either of the carboxylate moieties.

The effect of conformational mobility at the reaction center

Compared to Gly, both small apolar (Ala) and polar (Ser and His) side chains reduce the available backbone conformational space at (i)-($i+2$) positions significantly (Perczel *et al.*, 1991*b*; Perczel *et al.*, 1996). Therefore, the above-observed reduction in the reaction rate can originate either from constitutional (different molecular scaffolds and polarities) and/or conformational (different internal mobility) effects. *Ac*-A**NG**A-NH_2 partly adopts a type I′ or I/II′ β-turn (Perczel *et al.*, 1991*a*; Perczel *et al.*, 1993; Hutchinson and Thornton, 1994) unlike the fully unfolded *Ac*-**NG**AA-NH_2 and has a ∼30% longer τ, signaling that increasing backbone rigidity slows the isomerization rate (Supplementary Table 3). The salt-bridge enforced β-turn fold of *Ac*-E$^{(-)}$**NG**K$^{(+)}$-NH_2 has a similarly longer τ as compared to *Ac*-**NG**AA-NH_2, while the swapped *Ac*-K$^{(+)}$**NG**E$^{(-)}$-NH_2 variant an even longer (∼66% increase) (here, in addition to the restriction on backbone motions, the reaction enhancing positive charge of the Lys at the ($i+2$) position is replaced by the negative center of Glu). Furthermore the conformational freedom of –NG– is reduced even more as within *H*-SWTVE$^{(-)}$**NG**K$^{(+)}$VTWK-NH_2, a Trp-paired zipped β-hairpin structure (Wu *et al.*, 2009), we observe that the restricted internal dynamics increase the halftime of the isomerization over 2.5-fold (Supplementary Table 3). Similarly, placing the NGAA-motif of the highly flexible *Ac*-NGAA-NH_2 into that of *cyclo*(NGAANGAA) with limited internal dynamics and into the conformationally frozen *cyclo* (NGAA), resulted in τ increases of 2.5- and 5.3-fold, respectively (Supplementary Table 3 and Fig. 3). A significant decrease of conformational variability was confirmed in our molecular dynamics (MD) simulations where the *Ac*-**NG**AA-NH_2 samples 101, the cyclo-octapeptide 40, and the cyclo-tetrapeptide only three conformations along the equilibrium trajectory (Supplementary Fig. 1). Moving in the reverse direction, conformational restriction can be reduced by elevated T. Accordingly, the half-life of isomerization of the β-hairpin (*H*-SWTVE$^{(-)}$**NG**K$^{(+)}$VTWK-NH_2) decreases over 6-fold as the temperature is raised from 310 to 328 K, with a coupled decrease of fold compactness from 60 to 35% (Fig. 4*a*).

We thus found that a rigid backbone can act as an effective protective factor, which leads to the conclusion that –NG– motifs that are preserved during evolution, especially in long-lived proteins, must have a higher than average structural rigidity. There are NMR descriptors, measuring rigidness of the protein backbone, the $\langle S^2 \rangle$ and $\langle hetNOE \rangle$ values, available from the Biological Magnetic Resonance Bank (BMRB) (Ulrich *et al.*, 2007). S^2 of an unstructured residue is ∼0 and it is around 1 if the NH-vector nutation is fully restricted, while nuclear Overhauser effect (NOE) enhancement of the NH bond is around −3.9 for fast and 0.8 for slow motions. We found 207/225 S^2 and 224/289 *hetNOE* data values referring to AsnGly/AspGly units, with average values of $\langle S^2 \rangle_{\text{all.NGs}} = 0.87 \pm 0.13$ and $\langle S^2 \rangle_{\text{all.DGs}} = 0.84 \pm 0.14$, as well as $\langle hetNOE \rangle_{\text{all.NGs}} = 0.67 \pm 0.34$ and $\langle hetNOE \rangle_{\text{all.DGs}} = 0.73 \pm 0.22$. Although Gly typically introduces higher backbone mobility affecting neighboring residues as well, we found that almost all (>90%) NG/DG motifs in BMRB have rather rigid backbone structures. Only seven out of the above cases presents higher than average mobility for –NG–. Likewise, among the 145 analyzed cases of –DG– motifs, over 134 (>95%) are in a rigid backbone environment and only three show elevated internal mobility. Thus, if an NG/DG motif appears in a sequence, it is preserved as a rigid unit, in spite of the presence of the inherently flexible Gly residue. However, the stabilization effect of such rigidity is not always fully effective – exceptions can also be found. For example, the isomerization of the –NG– motif of a Tyr phosphatase protein (BMRB entry 19388 (Stehle *et al.*, 2013) or 26513 (Stehle *et al.*, 2015)), even though both the Asn and Gly show restricted backbone dynamics ($S^2_{\text{Asn161 or Asn164}} = 0.41$ and $S^2_{\text{Gly162 or Gly165}} = 0.18$), takes place, nevertheless (Supplementary Table 1).

Effect of the local conformation

Another interesting, apparent counter-example to the concept that a stable fold provides a safeguard against isomerization in our experiments is the case of transferring the –KNGG– subunit into the well-folded Trp-cage protein (TcN^9: NLYIQWLK$^{(+)}$**NG**GPSSGR$^{(+)}$PPPS) (Neidigh *et al.*, 2002). In this case only a moderate increase of τ was found between the free and the protein entrapped form (Fig. 4*b*) even though the NG unit was placed at the end of the α-helical segment. The conformational rigidity of the NG unit in this matrix matches that of

https://doi.org/10.1017/S003358351900009X Published online by Cambridge University Press

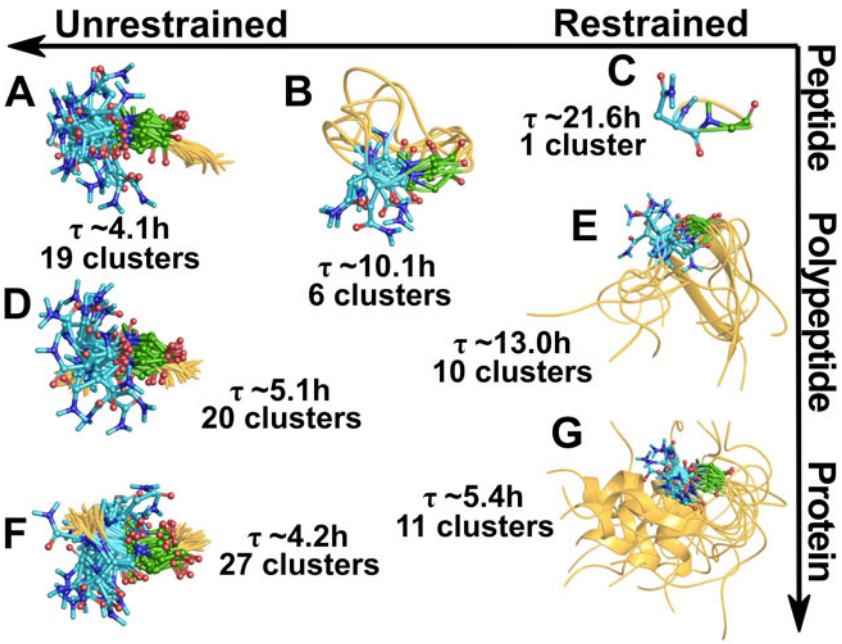


Fig. 3. As conformer heterogeneity of selected oligo- and polypeptides and protein models decreases (fewer clusters), the fold dependent protective factor increases and thus, isomerization rate gets reduced (τ increases). MD conformers were clustered and superimposed based on the backbone structure of the X-NG-Y motif. Mid-structures of the most populated clusters accounting for 90% of all the snapshots are shown in the case of the dynamically and conformationally 'locked' NG-subunits: (*a*) 19 for the 'free' Ac-NGAA-NH_2, (*b*) 6 for the restricted cyclo (NGAANGAA) and (*c*) 1 for the 'locked' cyclo(NGAA). (*d*) 20 for the 'free' Ac-ENGK-NH_2, while (*e*) 10 for the β-hairpin model of limited internal motion containing the –ENGK– motif. (*f*) 27 for the unrestrained Ac-KNGG-NH_2, but only (*g*) 11 for the folded TcN^9 mini-protein hosting the same –KNGG– but 'locked' motif. Asn and Gly are shown explicitly in cyan and green.

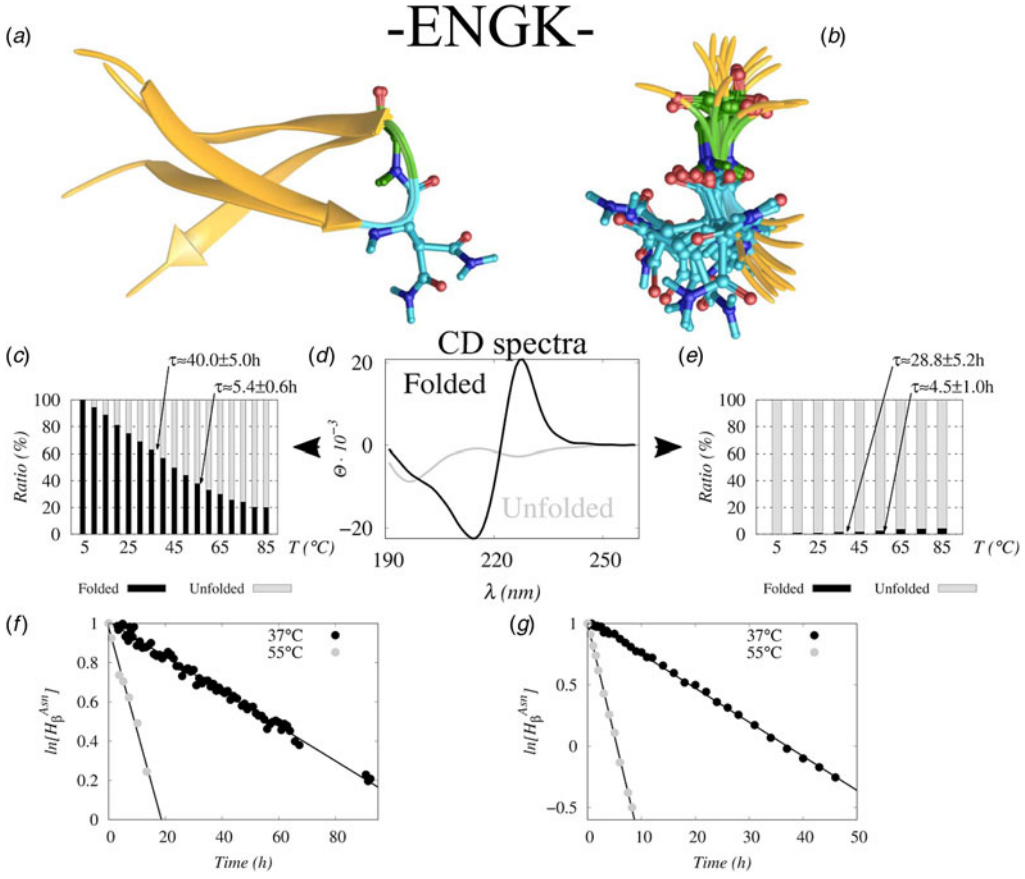


Fig. 4. Fold-compactness determines the magnitude of the protective factor. The well-folded β-hairpin (*a*) and the TcN^9 protein (*h*) restrict the backbone motion around NG, compared to the unrestricted backbone of Ac-E**NG**K-NH_2 (*b*) and Ac-K**NG**G-NH_2 (*i*). Asn and Gly are shown explicitly in cyan and green, respectively. The folded (black)/unfolded (gray) ratio (%) of the conformational ensembles at different temperatures derived from their far UV-electronic circular-dichroism (ECD) spectra (Perczel *et al*., 1991*c*; Perczel *et al*., 1992). Folded % shown as black bars (*c*, *e*, *j*, and *l*) with the reference pure ECD curves centered (*d* and *k*). Integral changes of selected ^{1}H-NMR-resonances (e.g. H_β^{Asn}) as function of the time (*f*, *g*, *m*, and *n*) enable to calculate isomerization half-lives: τ(h). Isomerization half-lives (pH = 7.4) of the four systems at two temperatures (37 and 55 °C) show that as the folded content drops (60→35%) and (66→43%) at higher *T*(°C)s (*c*, *e*, *j*, and *l*), the relevant τ also decreases (79.6 ± 10.6→13.0 ± 0.6 h) and (40.0 ± 5.0→5.4 ± 0.6 h), demonstrating the facilitation of the isomerization.

https://doi.org/10.1017/S003358351900009X Published online by Cambridge University Press

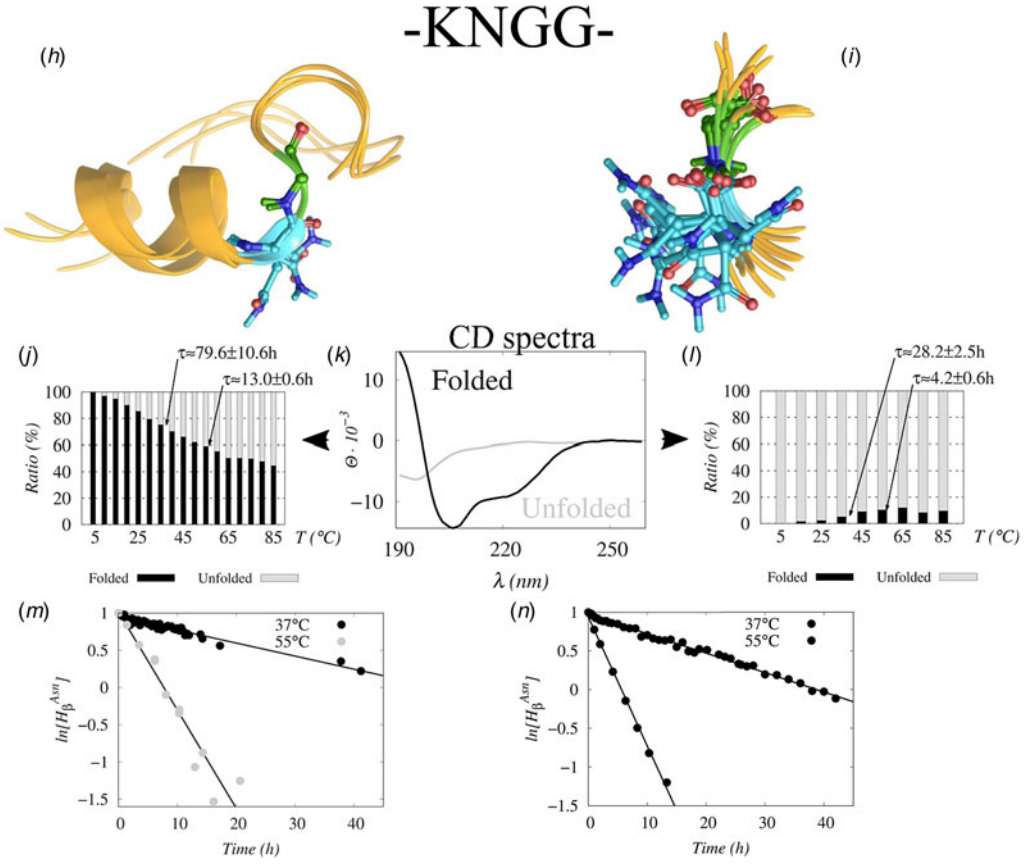


Fig. 4. Continued.

the previously mentioned β-hairpin model (with 64 (TcN[9]) and 60 (β-hairpin) conformations sampled during their 328 K MD simulations), but the protection factor of this specific fold is characteristically less pronounced (by a factor of 2, as compared to that of the β-hairpin). Thus it appears that beyond rigidity or flexibility, the specific arrangement of the reacting partners of the deamidation within a given fold should also be considered. The orientation and distance of the nucleophilic amide and its target $C^{\gamma} = O^{\delta}$ can be described using Bürgi–Dunitz (BD) distance (d) and angle (θ) to track the isomerization path (Burgi *et al.*, 1974; Ospina and Villaveces, 1993) (Fig. 5). The rate-limiting step of the reaction is the succinimide ring-closure taking place in the [ψ,χ_1] conformational subspace of Asn. In the protein data bank (PDB), several structures can be found where such succinimides of isomerization-prone Asn residues were captured in the protein matrix. It is important to note that the θ and d values are very similar in these proteins, not only in their the succinimide-state [$\theta \sim 125^{o}$, $d \sim 1.30$ Å], but also in their unreacted form, where their BD parameters are scattered within a well-defined region of $d = 3.25 \pm 1.2$ Å, $\theta = 109.5 \pm 10°$ (Fig. 5*a*), regardless of the protein type (hen egg-white lysozyme, legumain, apo-CheY, endothiapepsin, amylomaltase), sequential motif or secondary structure that hosts the specific Asn. In fact, the Asn-Gly units that eventually transformed into a succinimide are embedded in very different local environments: such as in α-helix, β-pleated sheet, β-turn, loops, *etc.* (Figs 5*c* and *d*) suggesting that a rigid structure does not, in all cases, provide protection against deamidation – not if the conformation of the Asn-Gly units are located within the reactive zone of the BD diagram. It is also interesting to see that in those proteins where the NG-motif was reported to undergo deamidation but with a half-life that is long enough to allow purification and crystallization procedures (minimum of several weeks) – where succinimide intermediates were not detected –, the θ and d values were found outside the reactive zone of the BD diagram (Fig. 5*b*), while appearing in virtually all regions of the Ramachandran plots (Figs 5*c* and *d*). This also indicates that an NG-motif, accommodated in even quite unfavorable local and secondary structure settings will eventually isomerize – the difference is only a matter of time (Tompa, 2010).

To better understand the geometrical criteria of Asn/Asp isomerization, we carried out rigid-body modeling of *Ac*–A**NG**A–NH_2 using the optimum value of θ (= 109.5 ± 1°). The topology of the obtained potential energy surface reveals that for a very wide range of χ_1 and ψ values, optimal reactive condition ($d < 3.25$ Å and $\theta = 109.5 \pm 1°$) can be reached (Fig. 6(I)): here succinimide formation would be prompt with short τ. The accessibility of such arrangements is well reflected by the fact that analyzing all protein families represented in the PDB (Berman *et al.*, 2000) (using a homology filtered set of 2724 entries), 42.6% of a total of 57925 AsnXxx/AspXxx-motifs fall in the reactive zone ($d < 4.25$ Å) and 9.4% within the highly reactive zone ($d < 3.5$ Å) (Fig. 6(II)). The very center of this zone holds the succinimide derivatives that were isolated and crystallized already (Fig. 6(II)). Still within the $3.25 < d < 4.25$ Å zone special Asx-turns can be assigned: categorized as type I, I′, II, and II′ Asx-turns, each analogous to the appropriate β-turn (Duddy *et al.*, 2004). They act as 'gatekeepers' of isomerization

https://doi.org/10.1017/S003358351900009X Published online by Cambridge University Press

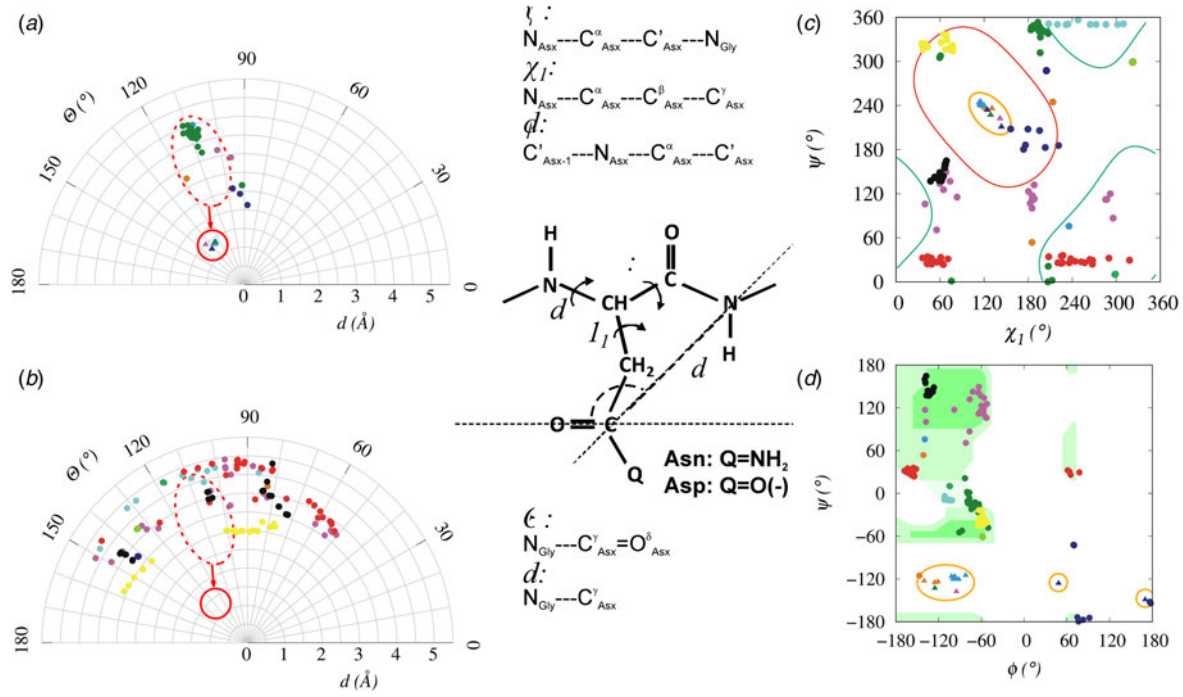


Fig. 5. Variables d(Å), θ(°), χ_1(°), and ψ(°) of Asn/Asp are shown to describe ring-closure, the rate-limiting step of the isomerization. (*a*) Red dashed oval ($d = 3.25 \pm 1.2$ Å, $\theta = 109.5 \pm 10°$) encircles PDB conformers for which isomerization is possible because of the spatial vicinity of the nucleophile N_{Gly} and the reaction center $C^{\delta} = O^{\gamma}$. The red circle encompasses PDB conformers where succinimide intermediates were crystallized and their structures were determined, namely hen-egg-white lysozyme or HEWL (green), legumain (blue); apo-CheY (magenta); endothiapepsin (cyan); and amylomaltase (orange). (PDB codes are as follows: **HEWL**: 1gwd, 1gxv, 1gxx, 1h6m, 1h87, 1hf4, 1w6z, 2blx, 2bly, 2bpu, 2c8o, 2c8p, 2cgi, 2w1l, 2w1m, 2w1x, 2x0a, 2xbr, 2xbs, 2xjw, 2xth, 2ybh, 2ybi, 2ybj, 2ybl, 2ybm, 2ybn, 2ydg, 3zvq, 4a7d, 4aga, **Legumain**: 4aw9, 4awa, 4awb, 4fgu, 4nom, **apo-CheY**: 3rvk, 3rvq, **Endothiapepsin**: 1e5o, **Amylomaltase**: 1esw.) (*b*) A similar (d, θ)-plot as **A** for proteins of which isomerization was described, but succinimide intermediates were not yet isolated: –V**NG**P– of α_B-crystallin (black) [2klr] and –G**NG**R– of GroES (light green) [1pcq] are both part of β-pleated sheets, –E**NG**A– of Ser-hydroxymethyl-transferase (green) [1rv3] is situated in an α-helix, –S**NG**P– of SOD1 (red) [1rk7] as well as –E**NG**K– and –K**NG**E– of β_2-microglobulin (magenta and yellow) [1jnj] are integrated into β-turns and –E**NG**E– of GroeS (blue) [1pcq], –G**NG**Y– of calbindin D28 (cyan) [2f33], –G**NG**R– of fibronectin (orange) [1fbr] are part of loops. (*c*) and (*d*) Asn associated (χ_1, ψ) and (ϕ, ψ) 2D-plots of the isomerizing NG-units of the above proteins: dot represents Asx-conformers, while triangles for the associated succinimide. Note that these examples cover most major secondary structural elements, namely the α-helix, β-pleated sheet, β-turn, and loops.

as they have relatively short d values, ($3.25 < d < 4$ Å), and while their θ values are unsuitable for isomerization (typically: $0 < \theta < 90°$ or $130 < \theta < 180°$), the inherent flexibility arising from NH^X-s that do not participate in *H*-bond networks, allows both χ_1 and χ_2 to adopt a conformation promoting isomerization. For 57.4% of all AsnXxx/AspXxx-motifs of the homology filtered protein structures, d is greater than 4.25 Å, so in these cases, even if θ is optimal, succinimide formation would be slow. Common secondary structures of this region are γ^{inv}-turn (or γ_L), δ_L- and left-handed α-helix (or α_D) (Perczel *et al.*, 1991*a*), so these are conformations that could delay isomerization considerably. On the other hand, analysis of the Ramachandran-surface, *f(&phis;, ψ)* (Fig. 6/III) of all Asn/Asp residues of the homology filtered set shows that both the β-valley (β-strands, polyPro II, *etc.*) and the α-helical regions contain zones colored red for which $d <$ 3.25 Å and $\theta = 109.5 \pm 1°$, and thus, both succinimide formation and the subsequent isomerization would readily proceed (even if the NH^{Gly} needed for the nucleophilic attack is engaged in the fold stabilizing H-bond pattern). Residues populating the surrounding blue/gray zones of the Ramachandran-surface are expected to isomerize more slowly (longer d- and/or perhaps less optimum θ-values). This analysis clearly demonstrates that even NX/DX-motifs that are 'locked' into secondary structural elements might adopt conformations optimal for isomerization.

Based on the above, it seems that beyond sequence and local flexibility/rigidity, the BD-parameters of the NX/DX-motif are crucial in determining the isomerization rate. Using these cornerstones, we could rationalize the NMR derived half-life of our model systems based on their MD derived conformational heterogeneity (backbone flexibility of the four-residue-stretch hosting the NG-motif) and BD-parameter distribution (Figs 3 and 6). The highly flexible *Ac*-NGAA-NH_2, *Ac*-ENGK-NH_2, and *Ac*-KNGG-NH_2 tetrapeptides sample 101, 95, and 122 different backbone conformations respectively. Although the median values of θ and d (~92° and 3.7 Å) of these systems fall outside the reactive zone of the BD diagram ($\theta = 109.5 \pm 10°$, $d = 3.25 \pm 1.2$ Å), its inner territory is well represented in the equilibrium ensemble as well. As discussed earlier, while the NGAA-motif within a *cylo*-octapeptide and a *cyclo*-tetrapeptide scaffold becomes increasingly rigid, its BD-descriptor values shift toward less favorable regions: θ/d medians are 91°/3.70 Å and 81°/3.80 Å, respectively. Especially in the case of the *cyclo*-tetrapeptide scaffold, the combination of a rigid backbone fold with unfavorable BD-parameters gives the perfect example of an off-pathway structure of the isomerization pathway, isomerizing >5 times slower than the unbound free form. Similarly, placing the ENGK-motif into a β-hairpin scaffold the backbone conformation becomes more rigid with BD-values (θ/d medians are ~94°/4.1 Å) representing an off-pathway structure and thus, isomerization gets slower with τ increased by ~2.6 fold. On the other hand, building the KNGG-motif into TcN[9], at the tip of the α-helical segment of the Trp-cage fold, causes a significantly lower increase in τ (1.3

https://doi.org/10.1017/S003358351900009X Published online by Cambridge University Press

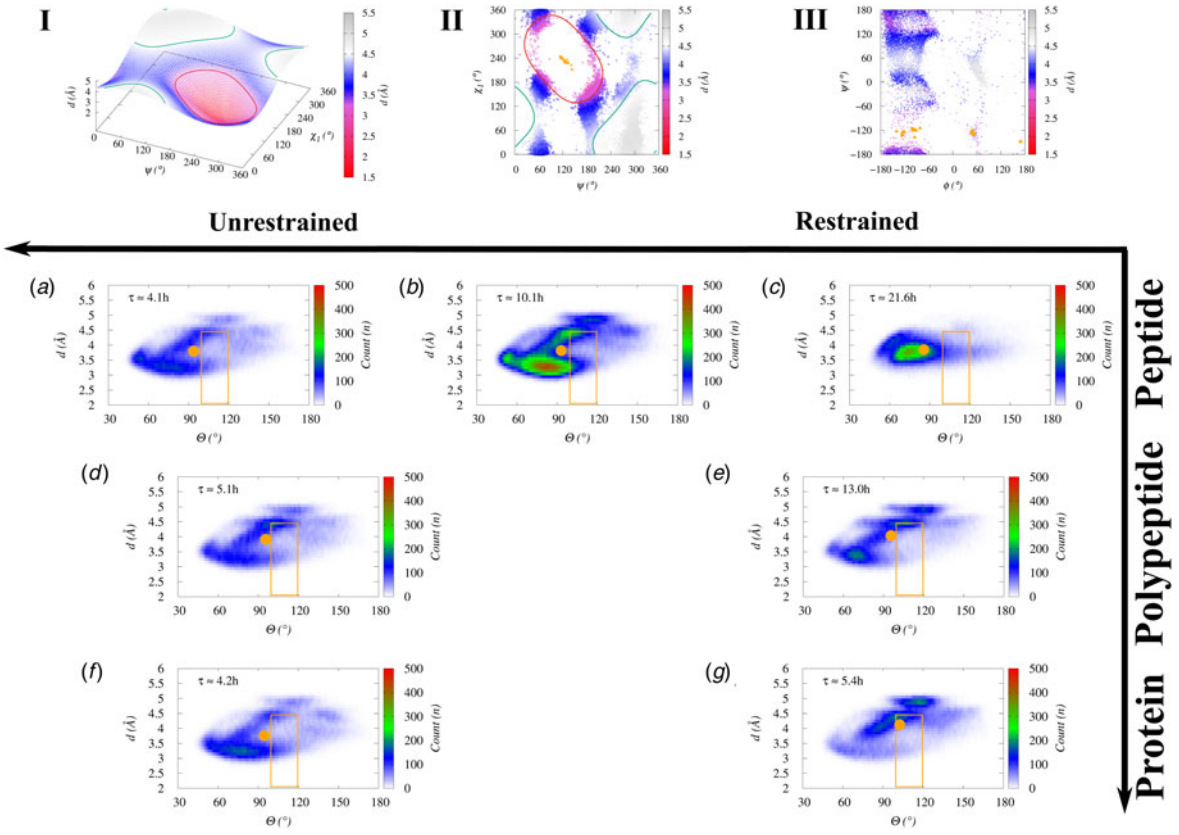


Fig. 6. (**I**) Distances (N^{Gly} to CG^{Asn}) at the optimal BD angle (N^{Gly}-CG^{Asn}-$OD1^{Asn}$) as a function of the χ_1^{Asn} and ψ^{Asn} torsions, of the rigid-body model of Ac-A**NG**A-NH_2 shows a single valley with centered at *f*(120°; 240°). Red-magenta ($d < 3.25$ Å) and blue-white regions (3.25 Å < d < 4.25 Å) outlined with green lines indicate distances within the 'reactive zone'. The darker the gray ($d > 4.25$Å) the more unreactive the structures are. Note that a reactive distance can be reached at a very wide range of ψ^{Asn}. (**II**) The $f(\chi_1^{Asx},\psi^{Asx})$ surface of –AsxXxx– selected from non-homologous proteins of the PDB with the same color-coding and region boundaries. Succinimides are shown as black triangles in the middle of the reactive valley. (**III**) Identical dataset mapped as $d \sim f(\phi^{Asn};\psi^{Asn})$ function, similar to a Ramachandran surface. Black triangles again represent succinimides found in the PDB. Distances and angles of the MD simulated equilibrium ensembles of (*a*) Ac-NGAA-NH_2, (*b*) cyclo(NGAANGAA), (*c*) cyclo(NGAA), (*d*) Ac-ENGK-NH_2, (*e*) *β*-hairpin model hosting the –ENGK– motif, (*f*) Ac-KNGG-NH_2, and (*g*) the mini-protein TcN9 hosting the –KNGG– motif. Orange rectangles show the boundaries of the reactive zone, orange dots the median of the distributions.

fold) in spite of the similar extent in the reduction of backbone mobility. However, in this molecular scaffold, the median of the BD-descriptors ($\theta/d \sim 102°/4.20$ Å) is clustered near the reactive zone, the Trp-cage fold thus provides an example of an on-pathway fold, and where the restricted backbone fixes the reaction partners into a geometry that promotes the isomerization of Asn.

Pathophysiological significance

Deamidation-induced isomerization of Asn/Asp residues has been called a molecular clock (Robinson and Robinson, 2001*a*), ticking as a time-bomb threatening proteins' structural integrity. The conversion of Asn is irreversible as NH_3 is released (and lost) during the process, while that of Asp can and is reverted (Clarke, 2003) in cells by protein L-isoaspartyl/D-aspartyl-*O*-methyltransferase (PIMT) (O'Connor, 2006), an evolutionarily highly conserved enzyme typically repairing *β*Asp modifications. The severity of –NG– and –DG– isomerization associated malfunctioning is underlined by the ubiquitous expression of PIMT/PCMT in both pro- and eukaryotic cells. The presence of PIMT in a wide range of taxa highlights its functional importance (Skinner *et al.*, 2000) showing that evolutionary tolerated Asx-isomerization sites will in all cases be guarded – most effectively in highly flexible segments where the spontaneous isomerization is also the most straightforward.

When a protein's half-life is shorter than the isomerization half-time of its AsnXxx/AspXxx units ($\tau_{isomer} > \tau_{protein}$), the protein is degraded long before any isomerization might complete. Human ornithine decarboxylase [UniProt ID: P11926] is such an example ($\tau_{protein} \sim 11$ min) containing two AsnGly units (N^{126}G and N^{398}G), which both might isomerize rapidly due to their favorable BD values ($\theta_{N126G} \sim 120°$ and $\theta_{N398G} \sim 136°$) but without being a risk to the proteome. But the opposite holds if $\tau_{isomer} < \tau_{protein}$, presenting a considerable challenge in maintaining homeostasis (Supplementary Table 2). For example, human hemoglobin *α* and *β* subunits of the erythrocytes circulate for ~100–120 days in adults (Harrison, 1979) and contain four and five Asn residues per subunit, respectively. In the most frequent alleles, there is no AsnGly unit found: instead of Gly, Val (N^{10}V), Leu(N^{81}L), Phe(N^{98}F), and Ala(N^{79}A) residues are located after Asn (Weintraub and Deverman, 2007). Furthermore, as all these AsnXxx motifs are part of *α*-helices, the implemented structural restrain further reduces the isomerization rate. However, several SNPs (single nucleotide polymorphisms) result in sensitive hemoglobin mutants, among which the Sardegna ($N^{50}H \rightarrow N^{50}G$), Singapore ($N^{78}A^{79} \rightarrow D^{78}G^{79}$), Wayne ($K^{139}Y \rightarrow N^{139}T$), La Roche-sur-Yon mutation

https://doi.org/10.1017/S003358351900009X Published online by Cambridge University Press

($N^{80}L \rightarrow N^{80}H$), and Providence N or D ($K^{82}G \rightarrow N/D^{82}G$) present extreme vulnerability, enhancing the isomerization rate and causing detectable changes in flexibility near the heme binding site (Wajcman *et al.*, 1992) and are all associated with inherent blood disorders (e.g. hemolytic disease). The crystal structure of a Providence ($K^{83}G \rightarrow D^{83}G$) mutant was determined (PDB: 5sw7) and shows the local conformation quite favoring isomerization ($\theta_{D83} \sim 109°$ and $d \sim 4.6$ Å) with NH^{Gly} free to exchange with water, located at the edge of an α-helix with a relatively dynamic backbone.

As for crystallins of the eye-lens, no turnover is possible ($\tau_{protein} \sim 80$ years) therefore N^{78}L- and N^{146}G-sites present a considerable vulnerability. Due to the large and hydrophobic side-chain of Leu, N^{78}L is expected to isomerize slowly: a reaction further hindered by the low-backbone mobility of its location in a β-strand with its backbone amide NHs strictly cross H-bonded. However, N^{146}G poses a more real threat (Supplementary Table 3). Even though its θ is outside the optimum range ($\theta_{N146} \sim 80°$), its backbone amide NH is H-bonded and thus becomes a poor nucleophile and it is located in a β-strand in which little room is left for backbone dynamics, the Asn→βAsp conversion does take place and leads to the loss of the transparency and refractivity of eye-lens and cornea (i.e. age-related cataract) through promoting aggregation (Fujii *et al.*, 2016). Proteopathies linked to hemoglobin, crystallin, *etc.* lead to the concept of chronoregulation (Weintraub and Deverman, 2007), which seems to be essential for life.

On the other hand, isomerization of Asn to Asp may also be part of regulatory functions. The anti-apoptotic Bcl-x_L contains a PEST degradation signal: TEAPEGTESEMETPSAI**N**^{52}**G**NPSW that becomes considerably more potent by the $N^{52}G \rightarrow D^{52}G$ isomerization (of $\tau_{isomer} \sim 20$ h) (Rooswinkel *et al.*, 2014). The change leads to a more effective ubiquitin-directed degradation (Rechsteiner and Rogers, 1996) of the protein, which in turn initiates cell death (Zhao *et al.*, 2007; Dho *et al.*, 2013): thus the presence of the NG-motif creates an active mode of age surveillance. Another example of regulation, in this case through cellular targeting, is the flexible and conserved GNG *N*-terminus of PKAs that will no longer be myristoylated if the conversion to GDG takes place (Jedrzejewski *et al.*, 1998).

Concluding remarks

Here we have shown that the decreased mobility of compact and rigid backbones of proteins acts as a protective factor against spontaneous deamidation and isomerization at their Asn/Asp-Gly 'Achilles' heels' and thus, helps to preserve their molecular integrity. However, our results also indicate that even within well-defined secondary structure elements, optimal conditions for isomerization can take place. Optimal BD-parameters and thus on-pathway structures can be found within turns, helices, and sheets. Several protocols were published to estimate the relative stability of an Asn in a given sequential (Robinson and Robinson, 2004*c*; Lorenzo *et al.*, 2015) or structural (Robinson and Robinson, 2001*b*; Lorenzo *et al.*, 2015; Jia and Sun, 2017) environment that consider the chemical nature of the ($i + 1$) residue, the length of *d*, or whether the Asn is part of an α-helix/β-sheet. Here we described that it is the electrostatic contribution of the (i-1) and especially ($i + 2$) positions along the *d*(Å) and θ(°) BD-values and ultimately the flexibility of the molecular scaffold that are the key factors in determining the rate of this spontaneous isomerization. Our results indicate that all these should be considered when assessing the severity of mutations or designing protein drugs, antibodies and determining their optimal storage conditions (with respect to isomerization rate, it is advised to store proteins at moderately acidic media (subcutan pH ~5), unless this pH is close to the isoelectric point of the protein (where unwanted precipitation can destroy 3D-fold and reduce solubility)).

Materials and methods

Synthesis of linear and cyclic peptides

Linear NGR peptides were prepared on Rink-Amide 4-(2′,4′-dimethoxyphenyl-Fmoc-aminmethyl)-phenoxyacetamido-methylbenzhydryl amine resin resin (0.67 mmol g^{-1} capacity), while precursors for cyclization on 2-chlorotrityl chloride resin (0.6 mmol g^{-1} capacity) using the Fmoc/tBu-strategy, but the Fmoc-group was detached under milder conditions: 2% 1,8-diazabicyclo[5.4.0]undec-7ene/2% piperidine/0.1 M hydroxybenzotriazole (HOBt) in dimethyl-formamide (DMF) (6 times) to avoid succinimide ring formation. Semi-protected peptide was removed from the 2-ClTrt resin with the cleavage mixture (AcOH–MeOH–dichloromethane = 1:1:8 (v/v/v)), stirred for 2 h, at RT, filtered and evaporated. Products were purified by reverse phase high performance liquid chromatography (RP-HPLC): semi-preparative Phenomenex Luna C18 column with eluent A (0.1% trifluoroacetic-acid (TFA)/H_2O) and B (0.1% TFA in MeCN/H_2O (80/20)) and identified by ESI-MS (Bruker Esquire 3000+ ion-trap).

The preparation of head-to-tail cyclopeptides was usually performed by incorporating Gly to the *C*-terminus of linear precursor peptide to avoid racemization cyclization. However, for c [NGAA], precursor H-AN(Trt)GA-OH instead of H-AAN(Trt)G-OH was used to get the cyclic tetrapeptide. Cyclization was performed in DMF in the presence of benzotriazol-1-yloxytris (dimethylamino)phosphonium hexafluorophosphate/HOBt/N, N-diisopropylethylamine (12/12/24 eq.) overnight at 40 °C. The reaction was stopped (with TFA/DMF), and purified by RP-HPLC and characterized by electrospray ionization mass spectrometry. The main product (~90%) was c[N(Trt)GAAN(Trt)GAA] with <10% of c[N(Trt)GAA]. Side-chain deprotection achieved at room temperature with 95% TFA/2.5% TIS/2.5% H_2O for 1.5 h. Once precipitated with Et_2O, the product was purified with RP-HPLC.

Protein expression labeling and purification

Tc9N's complementary deoxynucleic acids (cDNAs) were ligated into a SacII and BamHI site of the pUBK2 vector, and the plasmid encoding H10-Ub-Protein was transformed into the *Escherichia. coli* strain BL21(DE3) (Stráner *et al.*, 2016). Transformed cells were grown in Luria-Bertani medium containing 0.1 mg ml^{-1} Kanamycin (200 rpm; 37 °C). Cells (OD600 0.6–0.8) were isopropyl β-D-1-thiogalactopyranoside induced to a final concentration of 1 mM and harvested after 3 h incubation (37 °C), resuspended and lysed in buffer (300 mM NaCl, 50 mM NaH_2PO_4, and 3 mM NaN_3) by sonication. Proteins were purified using a 5 ml nickel-nitrilotriacetic acid (Ni-NTA) chromatography column. The eluted fractions with fusion proteins were overnight dialyzed into a buffer and were 3–4 h digested with His-tagged yeast ubiquitin hydrolase (YUH). The hydrolyzates having monitored on sodium dodecyl sulfate–polyacrylamide gel electrophoresis, the

https://doi.org/10.1017/S003358351900009X Published online by Cambridge University Press

ubiquitin and YUH were separated from the protein by Ni-NTA chromatography, purified by reverse-phase HPLC (C18 column), and identified by a Perkin–Elmer Sciex API2000 mass spectrometer.

BMRB analysis

hetNOE and S^2 parameter were analyzed for available proteins retrieved from the BMRB (Ulrich *et al.*, 2007). From the 11817 proteins deposited, 772 have either T_1, T_2, *hetNOE* or S^2 values. In total, *hetNOE* for 289 proteins, while S^2 for 92 proteins were reported in the BMRB. Among these, 74/84 and 46/58 proteins have in total 115/145 and 76/151 –NG/DG– subunits of known *hetNOE* data and with S^2 parameters reported, respectively, once redundancy removed.

NMR experiments and assignment

For shorter peptides, ^{1}H-NMR data and two-dimensional (2D) homonuclear (correlation spectroscopy, total correlation spectroscopy, and rotating frame nuclear Overhauser effect spectroscopy), for proteins ^{15}N–^{1}H- and/or ^{13}C–^{1}H-heteronuclear single-quantum correlation (HSQC) and heteronuclear (^{13}C–^{1}H-HSQC and heteronuclear multiple-bond correlation) spectra were recorded to analyze the requested kinetic information and to get both assignment and structural information, respectively. Time-dependent ^{1}H-NMR measurements (Δt = 5, 10, 15, 20, 25, 30, 45, 60, 90, 120 min, *etc.*) were obtained to follow the isomerization (Supplementary Table 3) until the reactant concentration typically dropped below 10–20%. Isomerization ($c \sim$ 3–5 mM) was followed by ^{1}H-NMR (Bruker Avance 700 MHz): 9/1:H_2O/D_2O, 5 µl DSS and NaN_3 (2 mg NaN_3 and 1 mg DSS dissolved in 5 ml H_2O), buffers: pH = 7.40 and pH = 7.80 (50 mM Na_2HPO_4 buffer), pH = 6.33 (50 mM Na_2HPO_4 and 50 mM NaH_2PO_4), and pH = 5.10 (20 mM CH_3COONH_4). T = 310, 319, and 328 K were applied (Supplementary Table 4). To identify both the α- and β-isoforms of *Ac*-NGAA-NH_2 and of *Ac*-NGRA-NH_2 $CoCl_26H_2O$ was added as a shift reagent [Co (II)]/[Asp] ~ 0.023 (Supplementary Table 5) (Pivcová *et al.*, 1982).

NMR data based kinetic simulation and thermodynamics

NMR signals were referenced to DSS. Selected ^{1}H-NMR resonances (typically 8 to 12 per molecules) were simultaneously analyzed and the rate coefficients of the rate-limiting step, k_1, that of the succinimide formation, was determined, also expressed in terms of half-lives, or τ.

$$-\frac{d[AsnGly]}{dt} = k_1[NG] \quad (1)$$

$$\tau = t_{\frac{1}{2}} = \frac{\ln(2)}{k_1} \quad (2)$$

Succinimide hydrolysis leading to –βAspGly– and –αAspGly– have rate coefficients k_2 and k_3, respectively. The ratio of k_2/k_3 shows also the relative ratio of β- and α-isomers. Using the Eyring–Polányi equation, $\ln(k_1)$ plotted against T^{-1}. The slope of the linear gives the enthalpy change, whereas the ordinate intercept provides the entropy change of the first step.

CD spectroscopy

Cd spectra of the proteins (35–150 µmol) were recorded on a Jasco-J810 spectrophotometer (1.0 or 10 mm cuvettes). Typical spectral accumulation parameters were at a scan rate of 50 nm min^{-1} with a 1 nm bandwidth and a 0.2 nm step resolution over 185–260 nm (far-ultraviolet (UV)) and 250–320 nm (near-UV) with four scans averaged for each spectrum at temperatures from 5 to 85 °C with 5 min/5 °C thermal equilibration. The solvent spectra were used as reference baselines and subtracted from the spectra. Conversion into mean residue molar ellipticity units ($[\Theta]_{MR}$, $deg \times cm^2 \times dmol^{-1}$) were conducted for the far-UV region. Quantitative spectral deconvolution was achieved by CCA+ software (Perczel *et al.*, 1991*c*).

Molecular dynamics simulations

MD simulation was carried out as implemented in GROMACS59, using the AMBER-ff99SBildnp* force field. The system was solvated with TIP3P water molecules in dodecahedral boxes with a size allowing 10 Å between any protein atom and the box. Total charge was neutralized and physiological salt concentration was met (Na^+, Cl^-). Energy minimization of starting structures was followed by sequential relaxation of constraints on protein atoms in three steps and an additional NVT step (100 ps) to stabilize pressure. Trajectories of 600–1200 ns NPT simulations at 310 and 328 K and 1 bar were recorded for further analysis (collecting snapshots at every 4 ps) (for backbone RMSD values see Supplementary Fig. 1). The last 200–600 ns of the trajectories was clustered fitting the main chain atoms using a 1 Å cutoff.

Rigid body modeling

Protected ADGA tetrapeptide was optimized at the B3LYP/6-311++G(d,p) level of theory. Rotations of all four dihedral angles (φ, ψ, χ_1, and χ_2) of the optimized Asp by 5 degrees and all N_{Gly}-C^{γ}_{Asp} distances (d) and BD angles (θ) were determined. The structures having filtered on near-optimal 109.5° BD angles, ψ, χ_1 pairs, and N_{Gly}–C^{γ}_{Asp} distances were plotted.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S003358351900009X.

Acknowledgement. The authors thank Viktor Farkas and Berill Ónodi for the WVVW hairpin and Pál Stráner for the mini protein model, Masashi Yokochi for the BMRB data, and Ernő Keszei and László Nyitray for scientific discussion. NMR spectrometer measurement time (700 MHz Bruker) was the courtesy of MedInProt Grant Facilitating Access to Instruments from the Hungarian Academy of Sciences. Calculations were carried out at the NIIF Supercomputing Center of KIFU (Hungary). This research project was supported by the European Union and the State of Hungary and co-financed by the European Regional Development Fund (VEKOP-2.3.3-15-2016-00009 and VEKOP-2.3.2-16-2017-00014), and the K116305 OTKA grant of the NKFIH of the Hungarian Academy of Sciences.

References

Aki K, Fujii N and Fujii N (2013) Kinetics of isomerization and inversion of aspartate 58 of αA-crystallin peptide mimics under physiological conditions. *PLoS One* **8**, e58515.

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN and Bourne PE (2000) The protein data bank. *Nucleic Acids Research* **28**, 235–242.

Burgi HB, Dunitz JD, Lehn JM and Wipff G (1974) Stereochemistry of reaction paths at carbonyl centres. *Tetrahedron* **30**, 1563–1572.

Clarke S (2003) Aging as war between chemical and biochemical processes: protein methylation and the recognition of age-damaged proteins for repair. *Ageing Research Reviews* **2**, 263–285.

D'Angelo S, Trojsi F, Salvatore A, Daniele L, Raimo M, Galletti P and Monsurrò MR (2013) Accumulation of altered aspartyl residues in erythrocyte membrane proteins from patients with sporadic amyotrophic lateral sclerosis. *Neurochemistry International* **63**, 626–634.

Dho SH, Deverman BE, Lapid C, Manson SR, Gan L, Riehm JJ, Aurora R, Kwon K-S and Weintraub SJ (2013) Control of cellular Bcl-xL levels by deamidation-regulated degradation. *PLoS Biology* **11**, e1001588.

Duddy WJ, Nissink JWM, Allen FH and Milner-White EJ (2004) Mimicry by asx- and ST-turns of the four main types of beta-turn in proteins. *Protein Science* **13**, 3051–3055.

Fujii N, Takata T, Fujii N and Aki K (2016) Isomerization of aspartyl residues in crystallins and its influence upon cataract. *Biochimica et Biophysica Acta* **1860**, 183–191.

Goolcharran C, Stauffer LL, Cleland JL and Borchardt RT (2000) The effects of a histidine residue on the C-terminal side of an asparaginyl residue on the rate of deamidation using model pentapeptides. *Journal of Pharmaceutical Sciences* **89**, 818–825.

Harrison KL (1979) Fetal erythrocyte lifespan. *Australian Paediatric Journal* **15**, 96–97.

Hutchinson EG and Thornton JM (1994) A revised set of potentials for beta-turn formation in proteins. *Protein Science* **3**, 2207–2216.

Jedrzejewski PT, Girod A, Tholey A, König N, Thullner S, Kinzel V and Bossemeyer D (1998) A conserved deamidation site at Asn 2 in the catalytic subunit of mammalian cAMP-dependent protein kinase detected by capillary LC-MS and tandem mass spectrometry. *Protein Science* **7**, 457–469.

Jia L and Sun Y (2017) Protein asparagine deamidation prediction based on structures with machine learning methods. *PLoS One* **12**, e0181347.

Kosky AA, Razzaq UO, Treuheit MJ and Brems DN (1999) The effects of alpha-helix on the stability of Asn residues: deamidation rates in peptides of varying helicity. *Protein Science*. **8**, 2519–2523.

Kosky AA, Dharmavaram V, Ratnaswamy G and Manning MC (2009) Multivariate analysis of the sequence dependence of asparagine deamidation rates in peptides. *Pharmaceutical Research* **26**, 2417–2428.

Li B, Gorman EM, Moore KD, Williams T, Schowen RL, Topp EM and Borchardt RT (2005) Effects of acidic N + 1 residues on asparagine deamidation rates in solution and in the solid state. *Journal of Pharmaceutical Sciences* **94**, 666–675.

Lorenzo JR, Alonso LG and Sánchez IE (2015) Prediction of spontaneous protein deamidation from sequence-derived secondary structure and intrinsic disorder. *PLoS One* **10**, e0145186.

Midelfort CF and Mehler AH (1972) Deamidation *in vivo* of an asparagine residue of rabbit muscle aldolase. *Proceedings of the National Academy of Sciences of the United States of America* **69**, 1816–1819.

Neidigh JW, Fesinmeyer RM and Andersen NH (2002) Designing a 20-residue protein. *Natural Structural Biology* **9**, 425–430.

O'Connor CM (2006) 13 protein L-isoaspartyl, D-aspartyl O-methyltransferases: catalysts for protein repair. *The Enzymes* **24**, 385–433.

Ospina E and Villaveces JL (1993) Theoretical calculation of the reaction mechanism between ammonia and formaldehyde. *Journal of Molecular Structure: THEOCHEM* **287**, 201–209.

Perczel A, Hollosi M, Foxman BM and Fasman GD (1991*a*) Conformational analysis of pseudocyclic hexapeptides based on quantitative circular dichroism (CD), NOE, and X-ray data. The pure CD spectra of type I and type II beta-turns. *Journal of the American Chemical Society* **113**, 9772–9784.

Perczel A, Hollosi M, Tusnady G and Fasman GD (1991*c*) Convex constraint analysis: a natural deconvolution of circular dichroism curves of proteins. *Protein Engineering, Design & Selection* **4**, 669–679.

Perczel A, Park K and Fasman GD (1992) Analysis of the circular dichroism spectrum of proteins using the convex constraint algorithm: a practical guide. *Analytical Biochemistry* **203**, 83–93.

Perczel A, McAllister MA, Csaszar P and Csizmadia IG (1993) Peptide models 6. New beta-turn conformations from ab initio calculations confirmed by X-ray data of proteins. *Journal of the American Chemical Society* **115**, 4849–4858.

Perczel A, Farkas O and Csizmadia IG (1996) Peptide models. 18. Hydroxymethyl side-chain induced backbone conformational shifts of l-serine amide. All ab initio conformers of for l-Ser-NH_2. *Journal of the American Chemical Society* **118**, 7809–7817.

Perczel A, Angyan JG, Kajtar M, Viviani W, Rivail JL, Marcoccia JF and Csizmadia IG (1991*b*) Peptide models. 1. Topology of selected peptide conformational potential energy surfaces (glycine and alanine derivatives). *Journal of the American Chemical Society* **113**, 6256–6265.

Pivcová H, Saudek V and Drobnik H (1982) 13C nmr study of the structure of poly(aspartic acid). *Polymer* **23**, 1237–1241.

Rechsteiner M and Rogers SW (1996) PEST sequences and regulation by proteolysis. *Trends in Biochemical Sciences* **21**, 267–271.

Robinson NE and Robinson AB (2001*a*) Molecular clocks. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 944–949.

Robinson NE and Robinson AB (2001*b*) Prediction of protein deamidation rates from primary and three-dimensional structure. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 4367–4372.

Robinson NE and Robinson AB (2004*c*) Prediction of primary structure deamidation rates of asparaginyl and glutaminyl peptides through steric and catalytic effects. *Journal of Peptide Research* **63**, 437–448.

Robinson NE, Robinson ZW, Robinson BR, Robinson AL, Robinson JA, Robinson ML and Robinson AB (2004) Structure-dependent nonenzymatic deamidation of glutaminyl and asparaginyl pentapeptides. *Journal of Peptide Research* **63**, 426–436.

Rooswinkel RW, van de Kooij B, de Vries E, Paauwe M, Braster R, Verheij M and Borst J (2014) Antiapoptotic potency of Bcl-2 proteins primarily relies on their stability, not binding selectivity. *Blood* **123**, 2806–2815.

Shimizu T, Watanabe A, Ogawara M, Mori H and Shirasawa T (2000) Isoaspartate formation and neurodegeneration in Alzheimer's disease. *Archives of Biochemistry and Biophysics* **381**, 225–234.

Shi Y, Rhodes NR, Abdolvahabi A, Kohn T, Cook NP, Marti AA and Shaw BF (2013) Deamidation of asparagine to aspartate destabilizes Cu, Zn superoxide dismutase, accelerates fibrillization, and mirrors ALS-linked mutations. *Journal of the American Chemical Society* **135**, 15897–15908.

Simonovic M and Volz K (2002) Atomic resolution structure of a succinimide intermediate in *E. coli* CheY. *Journal of Molecular Biology* **322**, 663–667.

Skinner MM, Puvathingal JM, Walter RL and Friedman AM (2000) Crystal structure of protein isoaspartyl methyltransferase: a catalyst for protein repair. *Structure* **8**, 1189–1201.

Stehle T, Sreeramulu S, Lohr F, Richter C, Saxena K, Jonker H and Schwalbe H (2013) The Apo-structure of the Low Molecular Weight Protein-tyrosine Phosphatase A (MptpA) from Mycobacterium tuberculosis Allows for Better Target-specific Drug Development. doi:10.13018/BMR19388.

Stehle T, Sreeramulu S, Lohr F, Richter C, Saxena K, Jonker H and Schwalbe H (2015) The Apo-structure of the Low Molecular Weight Protein-tyrosine Phosphatase A (MptpA) from Mycobacterium tuberculosis Allows for Better Target-specific Drug Development. doi:10.13018/BMR26513.

Stráner P, Taricska N, Szabó M, Tóth GK and Perczel A (2016) Bacterial expression and/or solid phase peptide synthesis of 20–40 amino acid long polypeptides and miniproteins, the case study of class B GPCR ligands. *Current Protein and Peptide Science* **17**, 147–155.

Tompa P (2010) *Structure and Function of Intrinsically Disordered Proteins*. Boca Raton: Chapman & Hall/CRC Press.

Truscott RJW, Schey KL and Friedrich MG (2016) Old proteins in man: a field in its infancy. *Trends in Biochemical Sciences* **41**, 654–664.

Tyler-Cross R and Schirch V (1991) Effects of amino acid sequence, buffers, and ionic strength on the rate and mechanism of deamidation of asparagine residues in small peptides. *Journal of Biological Chemistry* **266**, 22549–22556.

Uchida T and Shibata Y (1981) An affinity adsorbent, 5′-adenylate-aminohexyl-sepharose. I. Purification and properties of two forms of RNase U2. *Journal of Biochemistry (Tokyo)* **90**, 463–471.

Ulrich EL, Akutsu H, Doreleijers JF, Harano Y, Ioannidis YE, Lin J, Livny M, Mading S, Maziuk D, Miller Z, Nakatani E, Schulte CF,

https://doi.org/10.1017/S003358351900009X Published online by Cambridge University Press

Tolmie DE, Kent Wenger R, Yao H and Markley JL (2007) Biomagresbank. *Nucleic Acids Research* **36**, D402–D408.

Wajcman H, Kister J, Vasseur C, Blouquit Y, Trastour JC, Cottenceau D and Galacteros F (1992) Structure of the EF corner favors deamidation of asparaginyl residues in hemoglobin: the example of Hb La Roche-sur-Yon [beta 81 (EF5) Leu----His]. *Biochimica et Biophysica Acta* **1138**, 127–132.

Weintraub SJ and Deverman BE (2007) Chronoregulation by asparagine deamidation. *Science STKE* **2007**, re7.

Wu L, McElheny D, Huang R and Keiderling TA (2009) Role of tryptophan-tryptophan interactions in Trpzip beta-hairpin formation, structure, and stability. *Biochemistry* **48**, 10362–10371.

Xie M, Aubé J, Borchardt RT, Morton M, Topp EM, Vander Velde D and Schowen RL (2000) Reactivity toward deamidation of asparagine residues in beta-turn structures. *Journal of Peptide Research* **56**, 165–171.

Zhao R, Oxley D, Smith TS, Follows GA, Green AR and Alexander DR (2007) DNA damage-induced Bcl-xL deamidation is mediated by NHE-1 antiport regulated intracellular pH. *PLoS Biology* **5**, e1.

https://doi.org/10.1017/S003358351900009X Published online by Cambridge University Press