Published in final edited form as:

Nat Biotechnol. 2008 April ; 26(4): 427–429. doi:10.1038/nbt1380.

Detecting native folds in mixtures of proteins that contain disulfide bonds

Mahesh Narayan^{1,2,5}, Ervin Welker^{1,3,4,5}, Huili Zhai¹, Xuemei Han¹, Guoqiang Xu¹, Fred W McLafferty¹, and Harold A Scheraga¹

¹ Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853-1301, USA

² Department of Chemistry, University of Texas at El Paso, El Paso, Texas 79968, USA

³ Institute of Biochemistry, Biological Research Centre of the Hungarian Academy, H-6701, Szeged, 62 Temesvári krt. Hungary

⁴ Institute of Enzymology, Biological Research Centre of the Hungarian Academy, H-1114, Budapest, 29 Karolina ut. Hungary

Abstract

High-throughput *in vitro* refolding of proteins that contain disulfide bonds, for which soluble expression is particularly difficult, is severely impeded by the absence of effective methods for detecting their native forms. We demonstrate such a method, which combines mass spectrometry with mild reductions, requires no prior experimentation or knowledge of proteins' physicochemical characteristics, function or activity, and is amenable to automation. These are necessary criteria for structural genomics and proteomics applications.

Structural genomics aims to determine representative structures of those protein families for which no three-dimensional (3D) structure exists. This large-scale effort requires development of high-throughput methodologies for preparing protein samples suitable for structure determination by X-ray crystallography and NMR. Although several hundred protein structures have been deposited in data banks as a result of structural genomics initiatives, the majority correspond to proteins expressed in soluble form¹. This limitation has hampered large-scale structure determination of secreted, multi-domain, and multiple-disulfide-bond-containing proteins, for which, typically, soluble expression cannot be achieved.

In vitro refolding of these proteins would constitute an alternative solution, but would require a high-throughput method for identification of the native fold that facilitates screening for optimal folding conditions. Unfortunately, present methods, such as activity assays and chromatographic separations, involve a cumbersome optimization process specific to each particular protein and, therefore, remain intractable for high-throughput methodology. This is an important handicap considering that proteins that use the secretory pathway for folding and trafficking are responsible for a disproportionately large share of diseases related to misfolding. Here we demonstrate a generally applicable method that combines two tested-and-tried biochemical techniques for effective screening of the oxidative folding conditions that promote

Correspondence should be addressed to M.N. (mnarayan@utep.edu), E.W. (welker@brc.hu.edu) or H.A.S. (has5@cornell.edu).

⁵These authors contributed equally to this work.

Note: Supplementary information is available on the Nature Biotechnology website.

Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions>

the regeneration of proteins with multiple disulfide bonds. Proteins that undergo oxidative folding are easily identified by bioinformatics approaches that accurately predict the oxidation state of cysteines^{2,3} and, thus, the number of native disulfide bonds in proteins, so our method may considerably extend the current scope of structural genomics initiatives.

Oxidative folding is pivotal to maturation of many extra-cellular and membrane-bound disulfide bond-containing proteins; it is a complex process involving sequential acquisition of native disulfide bonds, coupled with the formation of a stable native tertiary structure⁴. Although oxidative folding has been studied extensively both *in vitro* and *in vivo*^{5,6}, no generally applicable approaches have been developed for monitoring its progress, which involves the formation of a large number of intermediates differing in the number and nature (native/non-native) of their disulfide bonds^{4,7}. The primary difficulty in distinguishing the formed native protein from its folding intermediates is in identifying the formation of both the native tertiary structure and the number of formed native disulfide bonds. Both tests are necessary because intermediates possessing native-like tertiary structure and activity but lacking a native disulfide bond frequently accumulate^{4,5,8–10}. In other cases, intermediates possess the same number of disulfide bonds as the native species but lack native structure because the disulfide bonds have non-native pairings (so-called scrambled species)¹¹. Existing analytical methods monitor either the number of formed disulfide bonds (e.g., by using thiol blocking combined with mass spectrometry (MS)⁹) or the formation of native/native-like tertiary structure (by using circular-dichroism or other spectroscopic methods¹²), but not both. Furthermore, they are frequently based on prior knowledge of physicochemical properties of the protein and/or applicable only after an arduous optimization process specific to the protein of interest. Additionally, most such approaches require a considerable amount of purified protein.

We have developed a method that combines the power of electrospray-ionization Fourier-transform mass spectrometry (ESI/FTMS)^{13,14} with a reduction pulse^{8,10,15} to detect and identify the native form of multiple-disulfide-bond-containing proteins, either in purified form or in a mixture of other such proteins. The reduction pulse is a brief, mildly reducing procedure, applied to an aliquot of the folding mixture during the folding process; it reduces those disulfide bonds not protected by stable, three-dimensional structure^{8,10,15}. Such structure exists only in the native protein or in native-like intermediates (in which some but not all native disulfide bonds are formed)⁴. Thus, native proteins preserve their native disulfide bonds whereas scrambled species, which have the same native number of disulfide bonds (isomers with non-native pairing) as the native species, do not. The scrambled species are converted to their reduced forms by the reduction pulse. Subsequent blocking of the free cysteines alters the mass of each blocked protein species relative to its native form. Thus, mass spectrometric analysis can determine the number of disulfide bonds in all protein species, thereby identifying the native form of the protein.

The nature and power of this approach for detecting the native form of each protein once formed becomes more apparent when a mixture of different protein species and their oxidative folding intermediates are analyzed. The molecular masses of dozens of such mixture components can be measured simultaneously because of the unusually high resolving power ($>10^5$) of the FTMS used^{13,14}. This is demonstrated here by using a mixture of seven multiple-disulfide-bond-containing proteins (the theoretical number of folding intermediates with distinct disulfide patterns within the mixture is calculated to be 3,968 (ref. 5)) subjected to five different folding conditions. Proteins with well-characterized oxidative folding pathways and possessing disulfide-secure, disulfide-insecure or only unstructured intermediates^{5,11,15} were chosen to represent the generic types of oxidative folding pathways, challenges and diverse features of multiple-disulfide-bond-containing proteins (**Supplementary Notes** online). Figure 1a is an ESI/FTMS spectrum of a regeneration mixture of the seven proteins subjected to a reduction

pulse, followed by blocking of any free thiols with aminoethylmethanethiosulfonate (AEMTS)⁵. The masses of the native forms of the seven protein species and their intermediates are well resolved. Note that assignment of all species in all charge states in the mixture would complicate the figure; therefore, the identity of select species (native and fully reduced for each protein) in the most abundant charge state or in those charge states that are clustered in the spectrum is highlighted.

The reduction pulse^{8,10,15} plays a key role in our technique; to demonstrate that it selectively reduces all scrambled (unstructured) protein species but not native ones, we prepared separate mixtures of scrambled and native forms of the seven proteins and subjected them to a reduction pulse. The reduction pulse converted all scrambled protein species (fully oxidized, non-native isomers) to their respective, fully reduced forms (Fig. 1b). By contrast, all native proteins remained as such in the latter sample after application of the reduction pulse (Fig. 1c). The condition of the reduction pulse optimized here is generally applicable to multiple-disulfide-bond-containing proteins (see **Supplementary Notes**). The use of an internal standard (ubiquitin) facilitated selection of those folding conditions in which the highest concentration of the native form is generated for each protein species (**Supplementary Table 1** online). For each protein, a corresponding decrease of the peaks of the reduced form relative to ubiquitin indicates that these are indeed effective folding conditions. Addition of guanidium-HCl to samples after AEMTS blocking has multiple benefits; it ensures (i) similar solution conditions when processing samples under each folding condition and (ii) almost quantitative recovery of the proteins (see **Supplementary Notes**). The best folding condition identified for each protein by our method (**Supplementary Table 1**) was corroborated by a semiquantitative determination of the relative concentrations, and thus the yield, of the native proteins (**Supplementary Table 2** online), using the response factor determined for each protein (**Supplementary Methods** online). A brief review of the results (**Supplementary Notes**) shows that this approach also captures the primary folding characteristics of each protein and, therefore, can be used effectively in this capacity.

Our method does not require prior knowledge of the physicochemical characteristics of proteins and can be applied to proteins without knowing their function or activity, without prior folding information or optimization of their chromatographic separation, and without generating antibodies for immunoblot assays. It is amenable to automation and facilitates rapid testing of multiple folding conditions for a large number of proteins; it is inexpensive and practical. These are features that are required for structural genomics applications and high-throughput structure screening.

We recently identified and structurally characterized a protein that constituted <1% of a cell extract¹³, suggesting that the screening of folding conditions of expressed proteins without their prior purification from crude extracts is also feasible. The potential to analyze the formation of the folded structures of multiple-disulfide-bond-containing proteins in intact cells may soon be realized.

Acknowledgements

This research was supported by National Institutes of Health grants GM-24893 (to H.A.S.) and GM-16609 (to F.W.M.), by NORT(DNT), Hungary, OTKA-NF61431 (to E.W.), and by UTEP startup money (to M.N.). E.W. is a Howard Hughes Medical Institute international scholar and an EMBO-HHMI start up grantee.

References

1. Bhattacharya A, Tejero R, Montelione GT. *Proteins* 2007;66:778–795. [PubMed: 17186527]
2. Fiser A, Simon I. *Methods Enzymol* 2002;353:10–21. [PubMed: 12078485]

3. Song JN, Wang ML, Li WJ, Xu WB. *Biochem Biophys Res Commun* 2004;318:142–147. [PubMed: 15110765]
4. Wedemeyer WJ, Welker E, Narayan M, Scheraga HA. *Biochemistry* 2000;39:4207–4216. [PubMed: 10757967]
5. Narayan M, Welker E, Wedemeyer WJ, Scheraga HA. *Acc Chem Res* 2000;33:805–812. [PubMed: 11087317]
6. Tu BP, Weissman JS. *J Cell Biol* 2004;164:341–346. [PubMed: 14757749]
7. Wedemeyer WJ, Xu X, Welker E, Scheraga HA. *Biochemistry* 2002;41:1483–1491. [PubMed: 11814341]
8. Rothwarf DM, Li YJ, Scheraga HA. *Biochemistry* 1998;37:3760–3766. [PubMed: 9521695]
9. Ruoppolo M, et al. *Biochemistry* 2000;39:12033–12042. [PubMed: 11009618]
10. Welker E, Hathaway L, Scheraga HA. *J Am Chem Soc* 2004;126:3720–3721. [PubMed: 15038718]
11. Arolas JL, Aviles FX, Chang JY, Ventura S. *Trends Biochem Sci* 2006;31:292–301. [PubMed: 16600598]
12. van den Berg B, et al. *EMBO J* 1999;18:4794–4803. [PubMed: 10469657]
13. Ge Y, et al. *J Am Chem Soc* 2002;124:672–678. [PubMed: 11804498]
14. Han X, Jin M, Breuker K, McLafferty FW. *Science* 2006;314:109–112. [PubMed: 17023655]
15. Welker E, Narayan M, Wedemeyer WJ, Scheraga HA. *Proc Natl Acad Sci USA* 2001;98:2312–2316. [PubMed: 11226236]

Figure 1.

ESI/FTMS mass spectra of seven multiple-disulfide-bond-containing proteins (a) The regeneration mixture, 14 h after initiation of oxidative folding (50 mM DTT^{ox}, 10 mM Ca²⁺, 20 mM Tris-HCl, pH 8) that was subjected to a reduction pulse, followed by AEMTS blocking. (b) The fully-oxidized scrambled mixture subjected to a reduction pulse, followed by AEMTS blocking. The masses correspond to the fully reduced species blocked by AEMTS (R) of each protein type. (c) The mixture of native forms subjected to a reduction pulse followed by AEMTS blocking. The masses correspond to the native forms (N) of each protein type. (The seven proteins: RNase A, Y97F RNase A, HEWL, α-Lac, ONC, C65,72S RNase A, and BPTI; see **Supplementary Notes** and **Supplementary Methods**).