ISSN 0006-2979, Biochemistry (Moscow), 2012, Vol. 77, No. 6, pp. 616-623. © Pleiades Publishing, Ltd., 2012.
Published in Russian in Biokhimiya, 2012, Vol. 77, No. 6, pp. 758-766.
Originally published in Biochemistry (Moscow) On-Line Papers in Press, as Manuscript BM12-043, April 22, 2012.

Comparison of Experimental and Theoretical Data on Hydrogen–Deuterium Exchange for Ten Globular Proteins

M. Yu. Suvorina, A. K. Surin, N. V. Dovidchenko, M. Yu. Lobanov, and O. V. Galzitskaya*

Institute of Protein Research, Russian Academy of Sciences, 142290 Pushchino, Moscow Region, Russia; fax: (495) 632-7871; E-mail: ogalzit@vega.protres.ru

Received February 24, 2012
Revision received March 11, 2012

Abstract—The number of protons available for hydrogen–deuterium exchange was predicted for ten globular proteins using a method described elsewhere by the authors. The average number of protons replaced by deuterium was also determined by mass spectrometry of the intact proteins in their native conformations. Based on these data, we find that two models proposed earlier agree with each other in estimation of the number of protons replaced by deuterium. Using a model with a probability scale for hydrogen bond formation, we estimated a number of protons replaced by deuterium that is close to the experimental data for long-term incubation in D_2O (24 h). Using a model based on estimations with a scale of the expected number of contacts in globular proteins there is better agreement with the experimental data obtained for a short period of incubation in D_2O (15 min). Therefore, the former model determines weakly fluctuating parts of a protein that are in contact with solvent only for a small fraction of the time. The latter model (based on the scale of expected number of contacts) predicts either flexible parts of a protein chain exposed to interactions with solvent or disordered parts of the protein.

DOI: 10.1134/S0006297912060089

Key words: hydrogen–deuterium exchange, secondary structure, hydrogen bond, mass spectrometry

Early experiments on hydrogen–deuterium exchange in insulin played a crucial role in regarding proteins as dynamic structures in which intramolecular motion promotes their biological activity [1]. Different experimental approaches giving information about intramolecular motions have been developed for the last two decades. They include methods of fluorescence depolarization of tryptophane residues, NMR, inelastic neutron scattering, Mossbauer spectroscopy, infrared spectroscopy, and analysis of Debye–Waller factors obtained from crystallographic data.

To clarify the relationship between structural elements and polypeptide chain mobility, a set of statistical analyses of structures obtained from NMR experiments was done. In [2] it was demonstrated (using experimental data obtained by NMR) that flexibility of a residue is connected with its size (the smaller amino acid residue is, the larger is its mobility). Also, it was shown that fluctuations of a given amide group also depend on the sizes of side chains of the amino acid residues that are its neighbors in the protein sequence [2].

One of the widespread methods used for studying conformational changes in protein structures is the combination of hydrogen–deuterium exchange and mass spectrometry. Investigations connected with hydrogen–deuterium exchange essentially focus on the exchange of protons involved in formation of hydrogen bonds that help to stabilize secondary and spatial structures of proteins [3].

Hydrogen–deuterium exchange in a polypeptide chain can be taken into consideration both after protein cleavage using different peptidases (in this case protons exchanged for deuterium are recorded after chromatographic separation and mass-spectrometric analysis of the already obtained peptides) [4, 5] and for the intact protein.

A combination of hydrogen–deuterium exchange and mass-spectrometric analysis is used for investigation of transitions of protein from the unfolded state to the native one [6]. Dependent on the aim of an experiment, the proteins are subjected to the action of different substances (denaturants, acids, etc.) destabilizing the protein structure [7]. As a result, the protein molecule losses its native structure, intramolecular bonds become accessible to the solvent, and rapid exchange of protons with deuterium occurs. Considering the number of exchangeable protons, one can learn about conformational changes and fluctuations in the protein molecule.

* To whom correspondence should be addressed.

In theoretical works, attempts have been made to find structural characteristics of a protein chain connected with its flexibility to predict the protection factor of amide protons [8-11]. In all these papers, the parameters demanding knowledge of the protein three-dimensional structure are considered: the accessible surface of amino acid residue [8], the Gaussian network model for interpretation of experimental data obtained under local hydrogen–deuterium exchange [9], the phenomenological equation including two terms that reflect the contribution of van der Waals contacts and hydrogen bonds [10, 11].

It is important to predict the protection of protons from proton–deuterium exchange using only amino acid sequence since the number of proteins with known three-dimensional structures today is considerably less than the number of proteins with known amino acid sequences. The number of methods in which only the amino acid sequence of the studied protein should be known is not large. Such is the method Camp based on using a neural network, which allows prediction of the protection factors proceeding solely from the amino acid sequence of the protein [12].

Our method is based on the assumption that the absence of protection can be mostly explained by large amplitude fluctuations of the regions of protein chain between packed elements of secondary structure that facilitates the contact of the fluctuating region with solvent [13]. Thus, the prediction of unfolded and loop regions should reveal regions prone to strong structural fluctuations and therefore subjected to hydrogen–deuterium exchange. In our suggestion an amide proton will be exchanged for deuterium if it is accessible to solvent.

In our previous publications we showed what parameters could be used to estimate the extent of protection of a residue (protected/not protected) in during hydrogen–deuterium exchange. We took into account the properties of amino acid residues where it is not required to know the protein three-dimensional structure, but it is sufficient to know the amino acid sequence. The analogs of parameters determined from three-dimensional structures were considered: the expected number of contacts per residue [14] and the probability of formation of a hydrogen bond [15]. We demonstrated that taking into account only the amino acid sequence the fraction of correctly predicted residues protected from hydrogen–deuterium exchange is from 61 to 64%. At the same time, taking into account the energy of hydrogen bonds obtained from the three-dimensional protein structure the fraction of correctly predicted residues protected from exchange is 72% [13].

To obtain information on the structure concerning the behavior of separate amino acid residues, it is necessary to do experiments on hydrogen–deuterium exchange using NMR to characterize the occupation of amide and indole groups by protons and deuterium atoms, correspondingly. We carried out experiments on hydrogen–deuterium exchange using mass spectrometry. The increment of the average mass in D_2O was measured after 15 min of incubation and after one day for 10 proteins. The number of sites of deuteration predicted by our method is in good agreement with the experimental data obtained on the mass increment of the protein due to proton–deuterium exchange.

MATERIALS AND METHODS

Mass spectrometric analysis of protein. We used 20 mM ammonium acetate (Panreac), pH 6.8, as a buffer system to prepare protein solutions. All buffer solutions were prepared using deionized water and deuterated water (D_2O content 98%).

The proteins were desalted by dialysis against a buffer solution of 20 mM ammonium acetate, pH 6.8. We used a dialysis membrane of 10 kDa (Serva, Germany) for dialysis of the protein samples.

The reaction of hydrogen–deuterium exchange in proteins started with dilution (1 : 5) using the buffer containing D_2O. The remaining non-deuterium water was removed using gel filtration on microcolumns produced in the laboratory. Sephadex G-10 preparations (granule size 40-120 μm; Pharmacia Fine Chemicals, Sweden) were used as the matrix for packing the microcolumns. Protein solution (5 μl) with concentration from 1 to 3 mg/ml was applied onto the column equilibrated with 10 mM ammonium acetate, pH 6.8, prepared using deuterium water. The columns were placed in holders and centrifuged at 2000 rpm for 2-5 min in an Eppendorf centrifuge (Germany). The average number of protons in the protein subjected to hydrogen–deuterium exchange was determined by mass-spectrometric analysis.

The mass-spectrometric analysis of proteins was done using an ion trap LCQ Deca XP Plus (Thermo Finnigan, USA). The samples were ionized by direct infusion of nanoelectrospray. The sample volume varied from 10 to 15 μl. The voltage supplied to the capillary was varied from 1 to 1.5 kV. The temperature of the entrance capillary was varied from 220 to 245°C. The mass spectra were registered at positive ions from 500 to 2000 m/z and from 500 to 4000 m/z. The coarse resolution and low dynamic range of the mass spectrometer lead to increased reading spread for ions with low relative concentration. Calculation of the average mass using such peaks results in increased calculation error. Therefore, we used peaks with amplitudes of more than 60% for calculating the average mass. At the same time, we used all significant peaks for calculating the charge state of ions with a corresponding value of the mass-to-charge ratio (m/z).

Dataset of proteins. For experimental work, we used proteins that are available in the laboratory of protein physics. The following proteins were included in the dataset: bovine milk β-lactoglobulin, bovine blood carbonic anhydrase B, human α-lactalbumin, α-lactalbu-

min from *Bos taurus*, sperm whale apomyoglobin, human proinsulin (recombinant protein), GroES from *Escherichia coli*, cytochrome *c* from *Equus caballus* (Sigma, USA), GFP from *Aequorea victoria*, lysozyme from *Gallus gallus* (Applied Chemistry).

Amino acid sequences of the chosen proteins were obtained from the UniProt database (www.uniprot.org).

Two scales for prediction of protection of protons from hydrogen–deuterium exchange. To predict the accessibility of protons to hydrogen–deuterium exchange, two scales were considered starting from the amino acid sequence: the predicted probability of hydrogen bond formation and the scale of the expected number of residue–residue contacts per residue. Since we are interested in the protection from hydrogen–deuterium exchange for the NH-group of the main chain, the hydrogen bond was "ascribed" to the donor (i.e. to the residue possessing the NH-group that formed the given hydrogen bond). Since most intraprotein hydrogen bonds are formed within the main chain, we did not consider possible hydrogen bonds with side-chains. We calculated the average probability with which a residue of a given type forms a hydrogen bond by its NH-group of the main chain with the main chain of the carbonyl group of the protein main chain. The calculated averaged probabilities of hydrogen bond formation by the NH-group of residue are presented in Fig. 1a. Under prediction based on the amino acid sequence of the protein, the average probability of hydrogen bond formation for the given type of residue was taken for each type of amino acid residue.

The scale of the expected number of contacts per residue is from the statistics of the number of contacts per residue [9] obtained from the spatial structures. The creation of this scale was described in detail elsewhere [9, 10]. For creating this scale, the same database was used as for making the scale of probability of hydrogen bond formation. For each amino acid residue, the number of contacts with other residues was calculated. Two residues are considered as being in contact if at least one pair of their atoms is situated at a distance of less than 8 Å. Further, the average number of contacts for each of the 20 types of amino acid residues was calculated. We used the 20 values obtained (Fig. 1b) as the scale under the prediction based on the amino acid sequence.

RESULTS AND DISCUSSION

Mass-spectrometric analysis for 10 proteins. All the investigated proteins were preliminarily dialyzed against acetate buffer (20 mM ammonium acetate, pH 6.8) for one day. Further, the protein was transferred to acetate buffer prepared using deuterium water (20 mM ammonium acetate in 98% D_2O) by simple dilution and incubated during 5 min. Before measurements, we used gel filtration on a column equilibrated with acetate buffer prepared in deuterium water to decrease the contribution of protons present in the initial solution of protein and protons of the protein exchangeable by deuterium. Then mass spectra were recorded. On average, 15 min passed

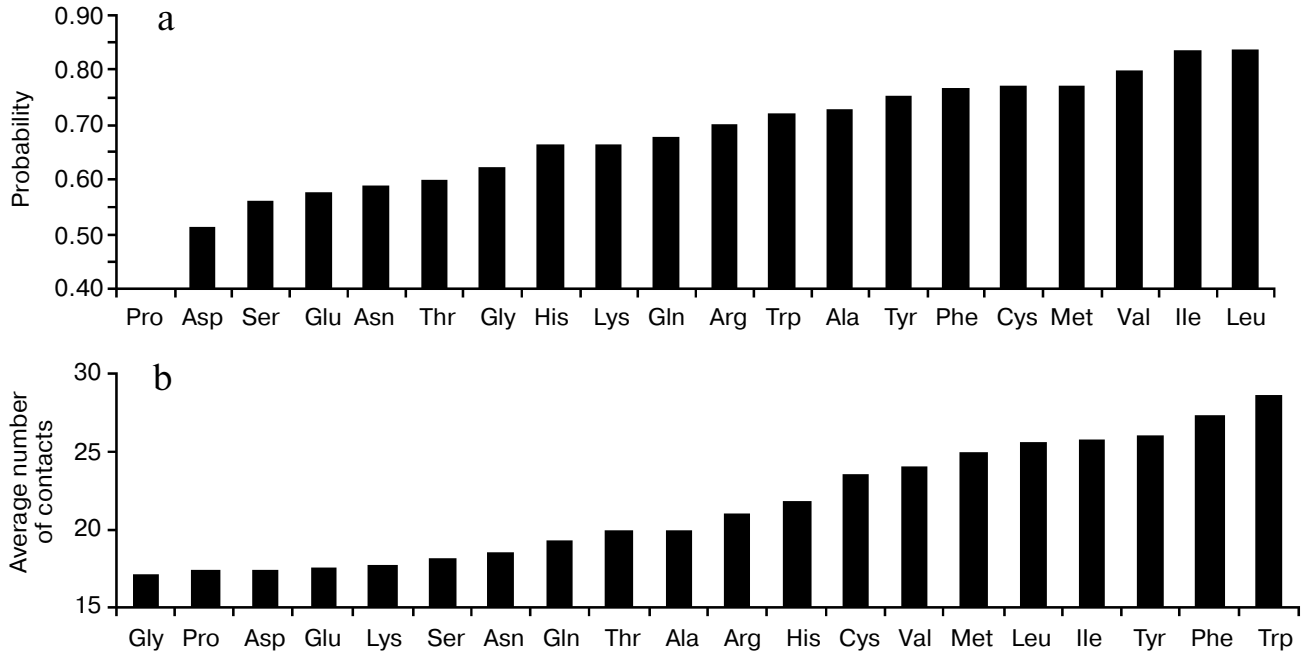

Fig. 1. a) Probability of hydrogen bond formation by a residue for each of the 20 types of amino acid residues in a globular protein; b) average number of contacts per residue in the globular protein.

between the first transfer in D_2O and the beginning of the mass spectrometric measurements. We determined the average mass of protein taking into account the natural isotope distribution and deuterated groups (in the case of experiments with hydrogen–deuterium exchange) from the mass spectra. For calculation of the average mass, we considered that ionization of protein occurs because of joining of one proton per each charging unit in the case of aqueous buffer and joining deuterium in the case of D_2O buffer. Mass spectra of proinsulin and lysozyme obtained in aqueous buffer (Figs. 2a and 3a) and in the buffer with D_2O after 15-min (Figs. 2b and 3b) and 1-day incubation (Figs. 2c and 3c) are given in Figs. 2 and 3.

The average number of protons replaced by deuterium was calculated as the difference between the average mass of proteins calculated for aqueous buffer and buffer with D_2O. The buffer systems used correspond to the native conditions for all proteins considered here. Under transfer of protein molecules into D_2O, the replacement of amide protons and other protons of side groups for deuterium begins. The first to exchange are the protons constantly exposed to the solvent, and after long incubation the protons accessible to the solvent for only a small fraction of the time (protons in the interior of the protein globule, but from time to time contacting with solvent due to fluctuations of the backbone of the polypeptide chain) have time to exchange as well. Thereby, for the first 15 min the protons on the surface of the protein globule are involved in the exchange, whereas during 1-day incubation all protons of the internal part of the globule capable of occasionally contacting the solvent also have time to exchange. The results of the experiments on hydrogen–deuterium exchange are presented in the table. We calculated the whole number of protons capable of deuteration in the unfolded state (see the table). The average fraction of exchangeable protons for 10 proteins is 60%. However, two proteins are unusual in this respect: proinsulin has the maximal fraction of exchangeable protons after 1-day incubation, 84%; and β-lactoglobulin has the minimal value, 47%.

Prediction of the total number of protons capable of hydrogen–deuterium exchange for 10 proteins. For prediction of the number of protons capable of hydrogen–deuterium exchange using only amino acid sequence we

Predicted and experimentally determined numbers of exchangeable protons for 10 globular proteins

Protein, mass (Da)	Predicted number of amide groups of the main chain accessible to H/D-exchange		Predicted total (main chain + side chain) number of protons capable of H/D-exchange		Average number of exchangeable protons (obtained by mass spectrometry), 15 min/24 h	Total number of protons accessible for exchange
	scale of hydrogen bonds	scale of expected contacts of side groups	scale of hydrogen bonds	scale of expected contacts of side groups		
Bovine β-lactoglobulin, 18361	78	58	152	116	117/130 ± 10	277
Bovine carbonic anhydrase B, 29086	152	104	292	215	212/265 ± 10	424
Human lactalbumin, 14112	68	55	140	114	84/110 ± 10	212
Bovine lactalbumin, 14221	72	54	149	111	90/117 ± 10	217
Whale apomyoglobin, 17335	81	64	145	114	96/133 ± 10	255
Human proinsulin, 9390	43	32	87	64	62/120 ± 10	147
GroES from *E. coli*, 10386	57	52	111	97	88/96 ± 10	169
Horse cytochrome *c*, 12359	62	55	130	122	87/132 ± 10	193
GFP from *Aequorea victoria*, 26308	135	95	255	177	192/214 ± 10	405
Hen egg lysozyme, 14310	73	49	159	100	134/146 ± 10	248

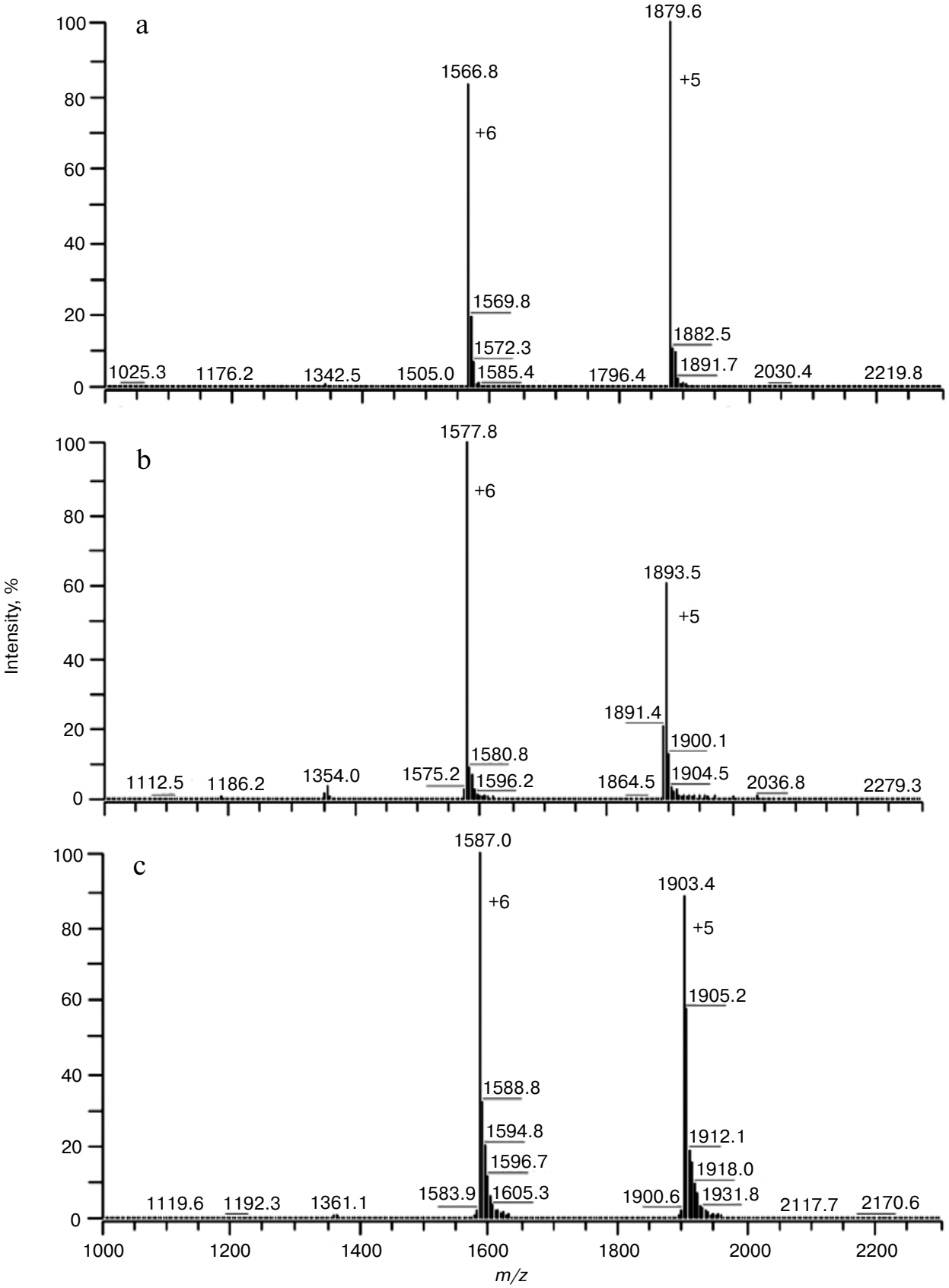

Fig. 2. Mass spectra of human proinsulin: a) water buffer, pH 6.8; b) deuterium buffer, pH 6.8, 15-min incubation; c) deuterium buffer, pH 6.8, 24-h incubation.

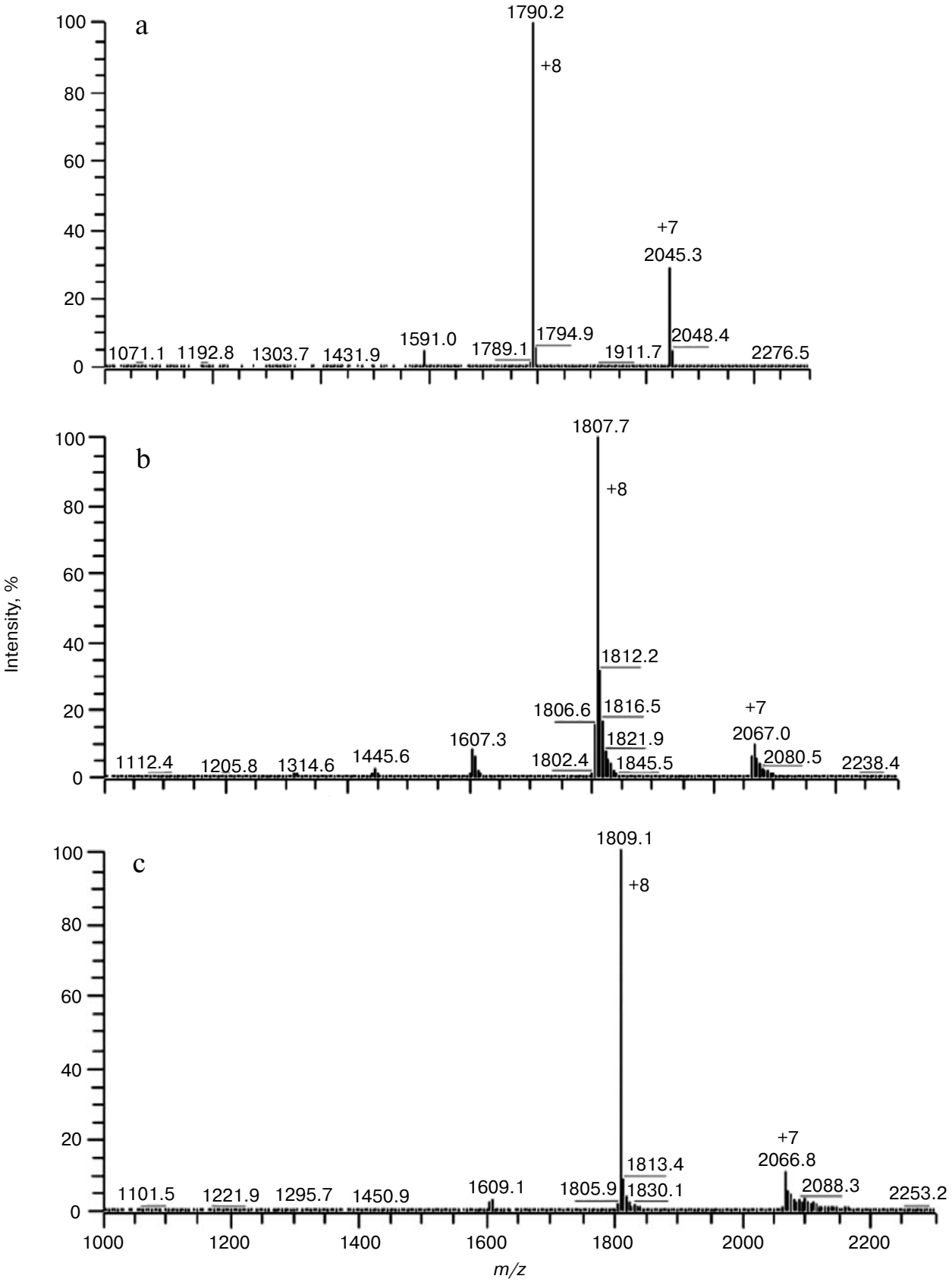

Fig. 3. Mass spectra of hen egg lysozyme: a) water buffer, pH 6.8; b) deuterium buffer, pH 6.8, 15-min incubation; c) deuterium buffer, pH 6.8, 24-h incubation.

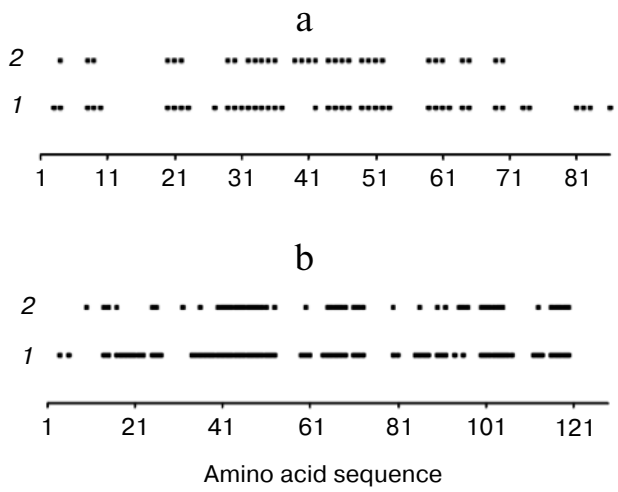

Fig. 4. Profiles of expected contacts and probability of hydrogen bond formation. Predictions were made using two scales (probability of hydrogen bond formation (*1*) and expected number of contacts (*2*)) for proinsulin (a) and for lysozyme (b). Horizontal lines denote regions not protected from hydrogen–deuterium exchange.

use two scales: the average (for the given type) number of intraprotein contacts (Fig. 1b) and the probability of intraprotein hydrogen bond formation (Fig. 1a) of the backbone amide group in amino acid residues of the given type. Upon prediction of the number of contacts and the probability of hydrogen bond formation by protein amino acid sequence, these average values were attributed to each residue of the given type. Then we averaged the considered values using a shifting window. In this work, the size of the window was the same for both scales and was three amino acid residues. The average number of intraprotein contacts was minimal for glycine (17.11 contacts) and maximal for tryptophane (28.48 contacts). Large hydrophobic amino acid residues had (on average) more intraprotein contacts compared to small and hydrophilic residues (see Fig. 1b). The probability of hydrogen bond formation by the NH-group of a residue varied from 0 (in the case of proline in which the NH-group is absent) to 0.84 (for leucine). It should be noted that hydrophilic amino acid residues had smaller probabilities of hydrogen bond formation inside the main chain of the protein compared to hydrophobic residues (see Fig. 1a), which probably reflects greater involvement of hydrophobic residues in regular secondary structure elements.

While predicting the protection of amino acid residues from hydrogen–deuterium exchange, residues that had number of predicted contacts above the cutoff value (20.4 intraprotein contacts per residue) were predicted as protected, and residues having a smaller number of predicted contacts were predicted as non-protected. Similarly, residues that had probability of hydrogen bond formation above the cutoff value (0.69) were predicted as protected, and residues that had lower probability of hydrogen bond formation were predicted as non-protected from hydrogen–deuterium exchange.

For calculation of protons accessible for exchange at the side groups of amino acid residues predicted as non-protected, the number of protons capable of exchange was determined taking into account the following factors. Arginine is able to exchange four protons, asparagine, glutamine, and lysine – two protons, serine, threonine, tyrosine, aspartic acid, and glutamic acid – one proton. For the proteins from our dataset we calculated the number of protons capable of hydrogen–deuterium exchange (see table). Figure 4 shows the regions of amino acid residues capable of replacing the proton for deuterium calculated using the two scales for proinsulin and lysozyme.

Two conclusions can be made from the data presented in the table. First, on using the scale of probability of formation of hydrogen bonds we obtain the number of protons capable of hydrogen–deuterium exchange that is closer to the experimental data obtained after long incubation in deuterated water (24 h), when practically the maximally possible number of protons capable of exchange was replaced. Second, on using the scale of expected contacts in globular proteins our estimations are closer to the experimental data obtained after short incubation in deuterated water (15 min) when protons exposed to the solvent have mostly exchanged.

In this work we compared experimental and theoretical data on the number of sites of deuteration for 10 globular proteins. We demonstrated that theoretical results based on the scale of probability of formation of hydrogen bonds are in better agreement with the mass spectrometry measurements of increased protein mass after long incubation in deuterated water than the results obtained using the scale of expected contacts. At the same time, the theoretical results obtained with the use of the latter scale are closer to the experimental data obtained after 15 min of incubation in deuterated water. Therefore, the model based on the expected number of contacts better predicts the exchangeable protons in the protein spatial structure that are at the surface of the globule or in unstructured regions of the protein chain. In contrast, the model based on the statistics of hydrogen bonds allows us to take into account the groups contacting with the solvent only during a small fraction of the time and therefore exchangeable to deuterium after only long-term incubation.

We are grateful to N. V. Kotova, V. E. Bychkova, B. S. Melnik, and V. V. Shmatchenko for providing us proteins for the experimental investigations.

This study was supported in part by the Russian Foundation for Basic Research (grant No. 11-04-00763) and the Russian Academy of Sciences (programs

"Molecular and Cell Biology" (01200959110 and 01200959111) and "Fundamental Sciences for Medicine").

REFERENCES

1. Hvidt, A., and Linderstrom-Lang, K. (1955) *Biochim. Biophys. Acta*, **16**, 168-169.
2. Goodman, J. L., Pagel, M. D., and Stone, M. J. (2000) *J. Mol. Biol.*, **295**, 963-978.
3. Englander, S. W. (2006) *J. Am. Soc. Spectrom.*, **17**, 1481-1489.
4. Zhang, Z., and Smith, D. L. (1993) *Protein Sci.*, **2**, 522-531.
5. Johnson, R. S. (1996) *J. Am. Soc. Mass Spectrom.*, **7**, 515-521.
6. Yan, X., Watson, J., Ho, P. S., and Deinzer, M. L. (2004) *Molec. Cell. Proteom.*, **3**, 10-23.
7. Deng, Y., Zhang, Z., and Smith, D. L. (1999) *J. Am. Soc. Mass. Spectrom.*, **10**, 675-684.
8. Sheinerman, F. B., and Brooks, C. L. (1998) *Proc. Natl. Acad. Sci. USA*, **95**, 1562-1567.
9. Bahar, I., Wallqvist, A., Covell, D. G., and Jernigan, R. L. (1998) *Biochemistry*, **37**, 1067-1075.
10. Vendruscolo, M., Paci, E., Dobson, C. M., and Karplus, M. (2003) *J. Am. Chem. Soc.*, **125**, 15686-15687.
11. Best, R. B., and Vendruscolo, M. (2006) *Structure*, **14**, 97-106.
12. Tartaglia, G. G., Cavalli, A., and Vendruscolo, M. (2007) *Structure*, **15**, 139-143.
13. Dovidchenko, N. V., Lobanov, M. Yu., Garbuzynskiy, S. O., and Galzitskaya, O. V. (2009) *Biochemistry* (Moscow), **74**, 888-897.
14. Galzitskaya, O. V., Garbuzynskiy, S. O., and Lobanov, M. Y. (2006) *Mol. Biol.* (Moscow), **40**, 341-348.
15. Savitski, M. M., Kjeldsen, F., Nielsen, M. L., Garbuzynskiy, S. O., Galzitskaya, O. V., Surin, A. K., and Zubarev, R. A. (2007) *Angew. Chem. Int. Ed. Engl.*, **46**, 1481-1484.