Multilevel structural characteristics for the natural substrate proteins of bacterial small heat shock proteins

Xinmiao Fu,1* Zengyi Chang,1 Xiaodong Shi,1 Dongbo Bu,2 and Chao Wang2

1State Key Laboratory of Protein and Plant Gene Research, School of Life Sciences, Peking University, Beijing 100871, China
2Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China

Received 23 October 2013; Accepted 2 December 2013
DOI: 10.1002/pro.2404
Published online 8 December 2013 proteinscience.org

Abstract: Small heat shock proteins (sHSPs) are ubiquitous molecular chaperones that prevent the aggregation of various non-native proteins and play crucial roles for protein quality control in cells. It is poorly understood what natural substrate proteins, with respect to structural characteristics, are preferentially bound by sHSPs in cells. Here we compared the structural characteristics for the natural substrate proteins of Escherichia coli IbpB and Deinococcus radiodurans Hsp20.2 with the respective bacterial proteome at multiple levels, mainly by using bioinformatics analysis. Data indicate that both IbpB and Hsp20.2 preferentially bind to substrates of high molecular weight or moderate acidity. Surprisingly, their substrates contain abundant charged residues but not abundant hydrophobic residues, thus strongly indicating that ionic interactions other than hydrophobic interactions also play crucial roles for the substrate recognition and binding of sHSPs. Further, secondary structure prediction analysis indicates that the substrates of low percentage of β-sheets or coils but high percentage of α-helices are un-favored by both IbpB and Hsp20.2. In addition, IbpB preferentially interacts with multi-domain proteins but unfavorably with α + β proteins as revealed by SCOP analysis. Together, our data suggest that bacterial sHSPs, though having broad substrate spectrums, selectively bind to substrates of certain structural features. These structural characteristic elements may substantially participate in the sHSP–substrate interaction and/or increase the aggregation tendency of the substrates, thus making the substrates more preferentially bound by sHSPs.

Keywords: molecular chaperone; small heat shock protein; protein aggregation; substrate proteins; IbpB; Hsp20.2

Introduction
Small heat shock proteins (sHSPs), as ubiquitous molecular chaperones present in all forms of life, play crucial roles for protein quality control in cells.1 They were able to suppress protein aggregation in an ATP-independent manner2,3 and stabilize stress-damaged cell membranes.4,5 Under in vitro conditions, sHSPs are known to effectively interact with unfolded model substrate proteins and keep them in

Abbreviations: sHSPs, small heat shock proteins; MW, molecular weight; pI, isoelectric point.

Additional Supporting Information may be found in the online version of this article.

Grant sponsor: National Natural Science Foundation of China; Grant number: 31100559 and 31270804 (to X.F.).
*Correspondence to: Xinmiao Fu, School of Life Sciences, Peking University, Beijing 100871, China.
E-mail: fxinmiao@pku.edu.cn

Published by Wiley-Blackwell © 2013 The Protein Society
a folding-competent state for subsequent refolding that is facilitated by such ATP-dependent chaperones as Hsp70s and Hsp100s. As such, the heterologous over-expression of sHSPs was repeatedly found to increase the tolerance of host cells against various stresses. Physiologically, sHSPs have been linked to cell differentiation, apoptosis, animal longevity, and dysfunction of them has been related to such diseases as cancer development, cardiovascular diseases, cataracts, myopathy, and neuron diseases.

One intriguing question regarding the functions of sHSPs is why they are able to protect a great diversity of model or natural substrate proteins. Whereas characterization of the structures of sHSPs and determination of their substrate-binding sites may be helpful in clarifying this question, defining the common structural features among different substrate proteins is also of significance. However, it is hard to determine the high-resolution structures of the substrate proteins due to the structural heterogeneity of sHSP-substrate complexes. Numerous studies using fluorescence spectroscopy, NMR, CD spectroscopy and spin labeling revealed that the substrates are characterized by native-like secondary structures but compromised tertiary structures and somehow exist in molten globule states. It appears that the substrate proteins, when complexed with sHSPs, are neither folded nor fully unfolded. Another commonly observed structural feature for the substrate proteins is that they are prone to aggregate in a nucleation-dependent manner. Additionally, Carver et al. reported that a disordered intermediate of α-lactalbumin was more efficiently bound by α-crystallin than a less disordered one, indicating that structural disorder might also be one of structural features of the sHSP substrates.

The structural features of sHSP substrates had been investigated on a few model substrates, but not yet on the natural substrates, and this kind of study, if possible, would be more biologically relevant. We have been attempting to investigate the chaperone activity and mechanism of representative sHSPs under both in vitro and in vivo conditions for years. Here we attempted to probe the common structural features of natural substrate proteins of sHSPs based on two studies recently contributed by us and others, in which around 100 natural substrate proteins of two bacterial sHSPs (Escherichia coli IbpB and Deinococcus radiodurans Hsp20.2) have been identified. We systematically compared these substrate proteins with the respective bacterial proteomes. Results indicate that bacterial sHSPs selectively bind to substrates of certain structural features.

Results and Discussion

Both IbpB and Hsp20.2 preferentially bind to substrate proteins of high molecular weight or moderate acidity

So far the natural substrate proteins of a few sHSPs have been identified. For instance, a total of 13 and 37 proteins were identified as the substrates of bacterium Synechocystis Hsp16.6 and E. coli IbpA in cells, respectively, and a total of 89 proteins present in cell extract were identified as the substrates of bacterium D. radiodurans Hsp20.2 during thermal stress. In particular, we recently identified a total of 110 proteins as natural substrates of E. coli IbpB in living cells by using in vivo photo-crosslinking. To unravel the structural features of in vivo substrate proteins of sHSPs with statistic significance, we choose IbpB substrate proteins as the subject and compared them with the proteome of E. coli, which comprises 2709 proteins that have been demonstrated at evidence of protein levels in this model organism and are deposited in the UnitProtKB database. It should be pointed out, as also revealed in our earlier study, that protein abundance is not the determining factor for a protein to be bound with IbpB. Rather, the principal factor determining such interaction is most likely the aggregation tendency of the protein during its folding/unfolding. In parallel, we also compared Hsp20.2 substrate proteins with the genome-predicted proteome (3167 proteins) of D. radiodurans, and the use of predicted proteome is due to the lack of sufficient experimental data on the proteome.

In retrospect, it was reported previously that GroEL, an essential molecular chaperone of E. coli, shows certain preference to substrate proteins of middle molecular weight (MW). Prompted by this, we first performed parallel statistic analysis on the MW of the IbpB substrate proteins and E. coli proteome, as well as of the Hsp20.2 substrate proteins and D. radiodurans proteome. Results displayed in Figure 1(a) demonstrate that IbpB preferentially binds to large proteins. Specifically, the proteins of 20–30 kDa represent the mostly un-favored (P < 0.01), somehow similarly to that of GroEL. Strikingly, the proteins of >70 kDa represent the most strongly favored for IbpB (P < 0.001) and such MW is beyond the upper limit of MW (60 kDa) for GroEL-encapsulated substrates, and is also apparently higher than that of GroEL-bound natural substrates. In this respect, IbpB and GroEL, respectively as a member of sHSP and Hsp60 families, may not be functionally overlapped in protecting the large proteins in E. coli cells. In support of this, IbpB was repeatedly observed to functionally cooperate with DnaK (Hsp70 family) and ClpB (Hsp100 family) by transferring substrates to these ATP-dependent chaperones but not to GroEL/GroES.
Similarly, Hsp20.2 also preferentially bind to large proteins but unfavorably to small proteins [Fig. 1(b)], albeit at lower significant levels than does IbpB. Together, such favorable binding to large proteins but unfavorable to small proteins for both IbpB and Hsp20.2 may reflect the general observation that large proteins are more difficult to fold and more prone to misfolding/aggregation during folding/unfolding than small proteins (as reviewed in the Ref. 51), thus need more protection by sHSPs.

GroEL was reported to have no any preference to substrates with respect to the isoelectric point (pI).49 By contrast, the pI distribution analysis revealed that IbpB preferentially binds to moderate acidic proteins with a pI of 5.0–5.5 (P < 0.001, Fig. 1(c)), and with a less degree to the proteins with a pI of 5.5–6.0 (P < 0.05). In comparison, moderate basic proteins with a pI of 7.0–10.0 are strongly un-favored for IbpB (P < 0.01). These observations, together with the difference in the MW distribution of their substrates, suggest that GroEL and IbpB are truly not functionally overlapped. Similarly with IbpB, Hsp20.2 shows a significant preference to substrates with a pI of 5.0–5.5 (P < 0.05) but unfavorable binding to substrates with a pI of 7.0–9.0 [Fig. 1(d)].

Both IbpB and Hsp20.2 exhibit preferences to substrate proteins of abundant charged residues rather than abundant hydrophobic residues

The favorable binding to the negatively charged acidic proteins for both IbpB and Hsp20.2, as

![Figure 1. IbpB and Hsp20.2 preferentially bind to proteins of high molecular weight or moderate acidity. Percentage of proteins plotted against the indicated molecular weight (Panels a and b) or isoelectric point (Panels c and d) for the E. coli proteome and the 110 substrate proteins of IbpB (Panels a and c), as well as for the D. radiodurans proteome and the 89 substrate proteins of Hsp20.2 (Panels b and d). The levels of statistical significance are indicated by *** (P < 0.001), **** (P < 0.01), and ***** (P < 0.001) in all the figures of the paper.](image-url)
revealed in Figures 1(c,d), is well consistent with the results of amino acid composition analysis [Fig. 2(a,b)]. Specifically, both IbpB and Hsp20.2 show significant preferential binding to substrates with a middle-to-high percentage of negatively charged residues (10–14%) but unfavorable to those with a low percentage (<8%) of these residues. Interestingly, we also found that both IbpB and Hsp20.2 exhibit significant unfavorable binding to proteins with a low percentage (<10%) of positively charged residues [Fig. 2(c,d)]. In addition, Hsp20.2 exhibits a remarkable preference to proteins with a high percentage (18–22%, $P<0.001$; >22%, $P<0.01$) of positively charged residues [Fig. 2(d)], which should be attributed, at least partially, to the enrichment of ribosomal proteins in the substrate pool of Hsp20.2 (referring to the Ref. 22). IbpB shows certain preference to proteins with a middle percentage (11–13% and 14–15%) of positively charged residues [Fig. 2(c)].

In light of these observations, together with the pI values of IbpB and Hsp20.2 themselves being respectively 5.2 and 5.0, we suggest that the negatively and positively charged residues abundantly present in the substrates of IbpB and Hsp20.2 may directly participate in the sHSP-substrate interactions through ionic interactions under neutral pH conditions. In support of this, we recently revealed the significantly enriched presence of the charged substrate-binding residues in IbpB, that is, among the 48 substrate-binding residues of IbpB, nine are positively charged and four negatively charged. In addition, sHSPs were reported to be unable to suppress the thermal protein aggregations under acidic pH, which apparently reflects the bleaching of ionic interactions between sHSPs and substrates under such acidic pH conditions.

Solvent-exposure of non-native hydrophobic surfaces in proteins is widely considered as a critical factor for protein aggregation. In line with this, hydrophobic interactions are considered to dominantly participate in the sHSP-substrate interaction as revealed by using hydrophobic probes under in vitro conditions. However, we unexpectedly found that both IbpB and Hsp20.2 do not prefer for substrate proteins with a high percentage of hydrophobic residues, including aliphatic residues [Leu, Ile, and Val; Fig. 2(e,f)] and aromatic residues [Trp, Phe, and Tyr; Fig. S1]. These observations thus indicate that the hydrophobic interactions between sHSPs and substrates may not be dominant as generally believed. Rather, other types of non-covalent interactions (e.g., ionic interactions as described above) also play crucial roles.

**Other notable observations**

**Secondary structure elements.** It would be of significance to unravel the features of the substrate
proteins at the level of secondary structures, given that they are characterized by native-like secondary structures when complexed with sHSPs. For this purpose, secondary structure predictions were performed on the E. coli and D. radiodurans proteomes using the algorithm of PSIPRED. It should be pointed out that such prediction is not as precise as the calculation of molecular weight and amino acid composition of proteins as describe above, and that the prediction results would somehow depend on the prediction algorithm that we used. Data indicate that both IbpB and Hsp20.2 unfavorably bind to substrate proteins with a low percentage of β-sheets [Fig. 3(a,b)] or coils [Fig. 3(c,d)], or those with a high percentage of α-helices [Fig. 3(e,f)].

These observations may be explained as follows. First, the proteins of low percentage of coils may be more likely structurally folded than those of high percentage of coils, and are thus unfavorably bound by sHSPs. Second, the enrichment of β-sheets in both sHSPs and the substrates suggests a possibility that, intermolecular β-sheets may be prevalently formed between sHSPs and substrates and even play crucial for the substrate recognition and binding of sHSPs. In support of this, sHSPs were repeatedly reported to suppress the fibril formation of amyloid proteins that are enriched in β-sheets (as extensively reviewed). It follows that the proteins of low percentage of β-sheets are unfavorably bound by sHSPs, as we observed [Fig. 3(a,b)]. Third, the apparent unfavorable binding of sHSPs to substrates of high percentage of α-helices may not be biologically relevant. Rather, it is just numerically obtained during prediction using the PSIPRED for secondary structure prediction, i.e., the combined content of α-helices, coils and β-sheets in a protein is always taken as 100%.

**SCOP.** Given the preference of IbpB and Hsp20.2 for substrate proteins with characteristic primary and secondary structures, we thought to examine whether these substrates contain common folds. We focused on the IbpB substrates with known 3D structures or with homologues of known structures, using the protein domain-classification databases SCOP. The structural classification of proteins in the E. coli proteome was directly adopted from the SCOP database. Data as displayed in Figure 4 indicate that IbpB significantly preferentially binds to substrate proteins of multi-domain (P < 0.05), which are defined as those with domains of different classes. This may reflect that the multi-domain proteins are more prone to aggregate during folding/unfolding than the single-domain proteins.

---

**Figure 3.** Both IbpB and Hsp20.2 unfavorably bind to substrate proteins of low percentage of β-sheets or coils but to those of high percentage of α-helices. Percentage of proteins plotted against the indicated content of secondary structural elements (β-sheet, coil and α-helix) for the E. coli proteome and IbpB substrate proteins (Panels a, c, and e), as well as for the D. radiodurans proteins and Hsp20.2 substrate proteins (Panels b, d, and f).
In addition, this finding is in accordance with the substrate preference of IbpB to proteins of high MW (>70 kDa) [Fig. 1(a)]. Notably, IbpB was found to unfavorably bind to substrate proteins of α + β fold, which are defined as those in which α-helices and β-strands are largely segregated. Nevertheless, why this protein fold is unfavorably bound by IbpB cannot be explained based on our current understanding on proteins structures.

In conclusion, the nature of sHSP–substrate interactions is still far from clear. Here we report, for the first time, that certain structural characteristic elements are present in the natural substrates of IbpB and Hsp20.2. These elements may directly participate in the sHSP–substrate interaction. Alternatively, they may increase the aggregation tendency of substrates and thus make the substrates more preferentially bound by sHSPs, given that the model substrates of sHSPs were found to be aggregation-prone per se. Specifically, the preferential binding to proteins of high MW or multi-domain may reflect the fact that these substrates have higher tendency to aggregate than the proteins of low MW or single domain. Further, the presence of abundant charged residues, instead of abundant hydrophobic residues, in the substrates indicates the ionic interactions other than the hydrophobic interactions play crucial roles in the sHSP–substrate interactions. In particular, the remarkable preferential binding of IbpB to acidic proteins with a pI of 5.0–5.5 apparently reflects the ionic interactions formed between the negatively charged residues of the substrates and the positively charged substrate-binding residues in IbpB. Although the genome-predicted proteome for D. radiodurans was used as a reference in comparative studies, the similarity of the 89 Hsp20.2 substrates with the 110 IbpB substrates in terms of multilevel structural characteristics suggests that this compromise does not severely interfere with the analysis.

Since there is no report on the substrate profiling of sHSPs of archaea or eukaryotes at present, whether their natural substrate proteins show similar characteristic structural features, as we observed here on the substrates of bacterial sHSPs, is unknown and merits further explorations. In addition, whether the structural elements at other levels (e.g., the tertiary and quaternary structures) are preferentially or unfavorably present in the substrates of sHSPs also needs to be investigated. On the other hand, characterization of the structural features of the model substrate proteins under in vitro conditions still represents an important way to probe the mechanism of substrate recognition and interaction of sHSPs.

**Materials and Methods**

**Data source**
A total of 2709 E. coli proteins, which have been demonstrated to be expressed at protein levels, were downloaded from the UnitProtKB database (date of 2013/7/3). A total of 3167 D. radiodurans proteins (predicted by the genome sequence) were downloaded from the NCBI database. The IbpB substrate proteins were adopted from our earlier study, and the Hsp20.2 substrate proteins were from the earlier report after removing five potential functional partners from a total of 94 Hsp20.2-interacting proteins.

**Bioinformatics analysis**
The MW and pI of each protein were calculated using the Compute pl/Mw software (http://web.expasy.org/compute_pi/). Amino acid composition analysis for all proteins was performed using the BioEdit software. Secondary structure prediction was performed using PSIPRED by subjecting the amino acid sequences to the algorithm installed in a local super-computer. The protein class of E. coli proteome was adopted directly from the SCOP database and that of IbpB substrate proteins was referred to the records of each protein deposited in the UnitProtKB database.

**Statistics**
Statistics was performed in Microsoft Excel software using the binominal distribution or t-test. Significance levels (with P values being less than 0.05, 0.01, or 0.001) were indicated in Figures 1, 2, 3, 4, and Supporting Information Figure S1.
References


236 PROTEINSCIENCE.ORG


