

# Substrate specificity of the DnaK chaperone determined by screening cellulose-bound peptide libraries

Stefan Rüdiger, Lothar Germeroth<sup>1</sup>,  
Jens Schneider-Mergener<sup>2,3</sup> and  
Bernd Bukau<sup>3</sup>

Zentrum für Molekulare Biologie Heidelberg, Universität Heidelberg, INF 282, D-69120 Heidelberg, <sup>1</sup>Jerini Bio Tools GmbH, Rudower Chaussee 5, D-12489 Berlin and <sup>2</sup>Institut für Medizinische Immunologie, Universitätsklinikum Charité, Humboldt Universität zu Berlin, Schumannstrasse 20-21, D-10098 Berlin, Germany

<sup>3</sup>Corresponding authors

**Hsp70 chaperones assist protein folding by ATP-dependent association with linear peptide segments of a large variety of folding intermediates. The molecular basis for this ability to differentiate between native and non-native conformers was investigated for the DnaK homolog of *Escherichia coli*. We identified binding sites and the recognition motif in substrates by screening 4360 cellulose-bound peptides scanning the sequences of 37 biologically relevant proteins. DnaK binding sites in protein sequences occurred statistically every 36 residues. In the folded proteins these sites are mostly buried and in the majority found in  $\beta$ -sheet elements. The binding motif consists of a hydrophobic core of four to five residues enriched particularly in Leu, but also in Ile, Val, Phe and Tyr, and two flanking regions enriched in basic residues. Acidic residues are excluded from the core and disfavored in flanking regions. The energetic contribution of all 20 amino acids for DnaK binding was determined. On the basis of these data an algorithm was established that predicts DnaK binding sites in protein sequences with high accuracy.**

**Keywords:** cellulose-bound peptide libraries/heat shock proteins/Hsp70/protein folding/spot synthesis

## Introduction

Hsp70 proteins are major constituents of the cellular chaperone network that assists protein folding. Hsp70 has been implicated in the folding and translocation of newly synthesized proteins, the assembly and disassembly of protein complexes, the refolding of misfolded proteins and, in special cases, the control of activity of native proteins (Gething and Sambrook, 1992; Morimoto *et al.*, 1994). These chaperone activities rely on the transient association of Hsp70 with substrates in a process controlled by ATP (McCarty *et al.*, 1995). It is assumed that association of Hsp70 with folding intermediates prevents their aggregation by both direct shielding of exposed hydrophobic surfaces and decreasing the concentration of free, aggregation-prone conformers.

The molecular principles allowing Hsp70 chaperones to exhibit promiscuity in substrate binding but selectivity with

respect to the bound folding conformer are only partly understood. The recent identification of the structure of the substrate binding domain of the *Escherichia coli* DnaK homolog in complex with a peptide substrate revealed a hydrophobic substrate binding cleft with a central pocket tailored to bind leucine (Zhu *et al.*, 1996). However, this structural information did not disclose the full potential for substrate recognition by DnaK. Earlier studies with DnaK and eukaryotic Hsp70 members revealed that these proteins recognize heptameric extended peptides enriched in hydrophobic residues (Flynn *et al.*, 1991; Landry *et al.*, 1992; Blond-Elguindi *et al.*, 1993; Gragerov *et al.*, 1994). With respect to the binding motif and its amino acid composition, they yielded results in conflict with each other (Flynn *et al.*, 1991; Blond-Elguindi *et al.*, 1993; Gragerov *et al.*, 1994) and the structure of the DnaK substrate binding pocket (Blond-Elguindi *et al.*, 1993). Furthermore, they were not aimed at identifying Hsp70 binding sites in biologically relevant protein substrates.

This study was performed to identify the binding sites within protein sequences and the substrate binding motif of the DnaK chaperone. For this purpose we screened cellulose-bound peptide scans (Reineke *et al.*, 1995) representing complete protein sequences for DnaK binding. This novel approach offers the advantages of (i) avoiding precipitation, in particular, of DnaK binding peptides anticipated to be hydrophobic, (ii) allowing identification of DnaK binding sites in natural substrate sequences, (iii) allowing direct identification of the recognition motif by sequence alignment of neighboring binding peptides and (iv) providing a large data set for binding as well as non-binding peptides. It allowed the identification of the substrate binding motif of DnaK and the establishment of an algorithm predicting DnaK binding sites within protein sequences.

## Results

### Screening of peptide scans for binding of DnaK

We screened cellulose-bound peptide scans (Reineke *et al.*, 1995) representing the complete sequences of 37 proteins for DnaK binding. These proteins were selected because they are natural substrates (cI, P and CIII of phage  $\lambda$ ; DnaK, DnaA,  $\sigma^{32}$ , pro-alkaline phosphatase, pro-maltose binding protein and  $\beta$ -galactosidase of *E.coli*; RepA of plasmid P1; influenza hemagglutinin from strains A/Aichi/2/68 and A/Suita/1/89; heat shock transcription factor 1, p53 and pre-insulin of human; *Photinus pyralis* luciferase; *Staphylococcus aureus* protein A; bovine clathrin light chains A and B; immunoglobulin chains  $\lambda$ -1 and  $\lambda$  V region heavy chain of *Mus musculus*; cytochrome b2, F1 $\beta$  and Su9 of *Saccharomyces cerevisiae*) or candidate substrates (DnaJ and FtsZ of *E.coli*; pro-BPTI and RepE of plasmid miniF;  $\lambda$  cII) for DnaK and eukaryotic Hsp70s or because knowledge of their structural and folding

properties allows mechanistic dissection of the assisted folding pathways (RNase T1, CheY, catabolite activator protein, FtsA, ribosomal protein L2 and proOmpA of *E.coli*; *S.cerevisiae* prepro- $\alpha$ -factor). The peptide scans were composed of 13mer peptides overlapping by 10 residues and thus presenting all potential binding sites for DnaK. They were incubated to equilibrium with DnaK in the absence of ATP followed by electrotransfer and immunodetection of the chaperone.

DnaK bound to peptides via its substrate binding pocket as judged by the (i) ability of  $Mg^{2+}$ /ATP to dissociate peptide-bound DnaK, (ii) ability of competing peptides to reduce the amount of bound DnaK, (iii) ability of the substrate binding domain of DnaK to exhibit a similar peptide binding pattern as wild-type DnaK and (iv) inability of the ATPase domain of DnaK to associate with peptides (data not shown). By comparison with reference peptides we grouped all peptides with significant and reproducible affinity for DnaK into good binders and binders. For 15 peptides with affinity for DnaK and significant solubility the equilibrium dissociation constants ( $K_d$ ) for DnaK were determined (McCarty *et al.*, 1996) and found to range from 0.1 to 7  $\mu$ M. Peptides with no detectable affinity for DnaK when bound to cellulose also exhibited no detectable affinity for DnaK in solution. These findings show that screening of cellulose-bound peptide scans identifies high affinity binding sites for DnaK.

Binding sites for DnaK occurred statistically every 36 (total sites) and 84 residues (good sites). In contrast, for BiP it was found that binding to random peptide sequences occurred at a much lower frequency (1/1000) (Blond-Elguindi *et al.*, 1993). This strong difference between random and protein sequence-derived peptides indicates that these chaperones are highly adapted to their specific function of interacting with proteins, provided that DnaK and BiP do not strongly differ with respect to promiscuity of binding. The frequency of DnaK binding sites within protein sequences was unaffected by cellular and organellar origin, size and oligomeric status of the proteins. We observed no regular pattern of DnaK binding sites within sequences, in particular no clustering of DnaK binding sites in N-terminal segments, emerging from ribosomes or being transported into organelles *in vivo*, except the hydrophobic signal sequences of secretory proteins (pro-alkaline phosphatase, pro-maltose binding protein, pre-insulin, pro-BPTI, proOmpA and prepro- $\alpha$ -factor).

#### **Localization of DnaK binding sites within native protein structures**

We determined the localization of DnaK binding sites within the corresponding three-dimensional structures, when available (examples are given in Figure 1B). Most binding sites are completely buried within the protein cores. In a few cases side chains of residues located in DnaK binding sites are surface exposed, however, without extensive exposure of their peptide backbone.

The majority of the good DnaK binding sites identified by peptide screening are in segments corresponding to  $\beta$ -sheets in the native proteins (22  $\beta$ -sheets compared with seven  $\alpha$ -helices of good sites investigated). One such binding site of particular interest is located in the central part of the  $\beta$ -sheet forming the dimerization interface of

human p53, which is a substrate for DnaK when produced in *E.coli* (Clarke *et al.*, 1988). It is tempting to speculate that DnaK binding to this site influences dimerization of p53. A minority of the DnaK binding sites are in segments corresponding to helices in the native proteins. DnaK binding sites were found only rarely within sequences of dimerization interfaces which adopt helical structures in the native state. No binding sites were found for example in the leucine zipper of  $\lambda$  repressor (not shown) and the helical dimerization interface of p53 (Figure 1B) and only one binding site exists in the long leucine zipper of hemagglutinin (Figure 1B).

#### **Identification of the substrate motif recognized by DnaK**

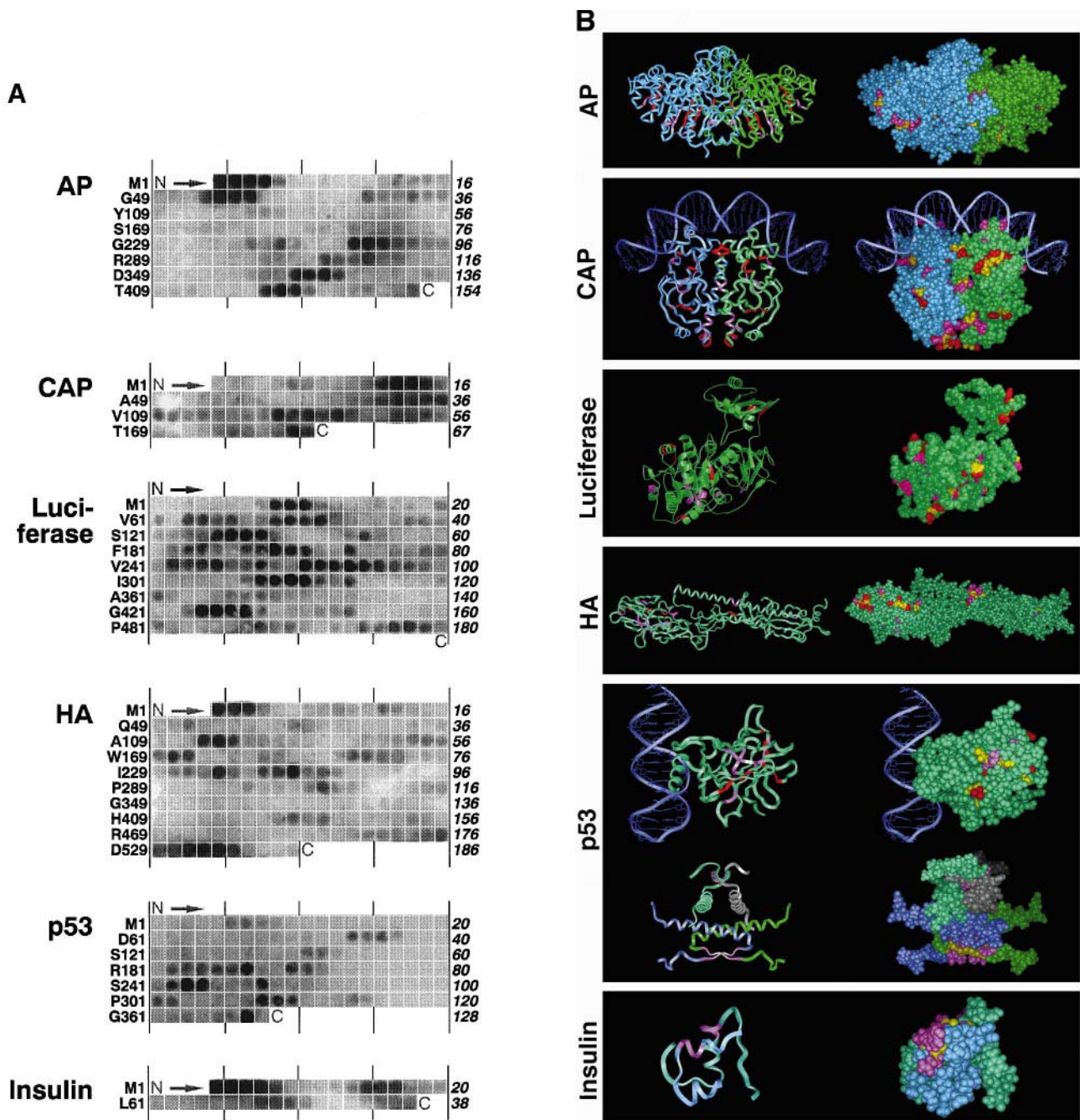
The large data set allowed a reliable statistical analysis of the substrate motif recognized by DnaK. The relative occurrence of the 20 amino acids in the entire library is similar to that found in natural proteins (Figure 2A). Substantial differences in the amino acid distribution exist between non-binding and total DnaK binding peptides (Figure 2B). DnaK binding peptides are enriched in the aliphatic residues Leu, Ile and Val, the aromatic residues Phe and Tyr and basic residues. Negatively charged residues are strongly and most other residues slightly disfavored.

The amino acids involved in DnaK recognition in individual binding sites were directly determined by sequence alignment and statistical analysis of the neighboring binding peptides of 90 regions each containing a single good DnaK binding site (Figure 3A and B). The DnaK binding motif is composed of a hydrophobic core of four to five residues length and two flanking regions enriched in basic residues (Figures 3B and 5A and Table I). The hydrophobic core of the DnaK binding motif is enriched in Leu, Ile, Val, Phe and Tyr (Figure 3B and Table I). Leu is particularly enriched (29% of all amino acids found in core regions) and present in 87% of the cores tested, characterizing DnaK as a Leu binding protein. The number of enriched residues in individual cores of good DnaK binding sites ranges between two and four (Figure 3C). Negatively charged residues are completely absent from the core and disfavor DnaK binding when present in close proximity (Figure 4 and Table I). Arg and Lys are disfavored within core regions but significantly enriched in both flanking segments (Figure 3B and Table I). The enrichment of both amino acids is 14.5% (left flanking region) and 16.8% (right flanking region) in a four residue segment compared with 9.8% in non-binding peptides (see Materials and methods for alignment of flanking regions).

Within the hydrophobic core, no specific distance pattern of the enriched residues was apparent. In particular, we found no evidence for a binding motif consisting of four alternating pockets of large hydrophobic/aromatic residues each separated by one residue, as proposed for BiP (Blond-Elguindi *et al.*, 1993). This was confirmed using two combinatorial libraries ( $X_5B^1XB^2X_5$  and  $X_5B^1XXB^2X_4$ ) with two defined ( $B^1$  and  $B^2$ ) and 11 random (X) positions (Kramer *et al.*, 1994). Screening for DnaK binding revealed similar spot patterns and intensities (Figure 4).

#### **Development of an algorithm predicting DnaK binding sites**

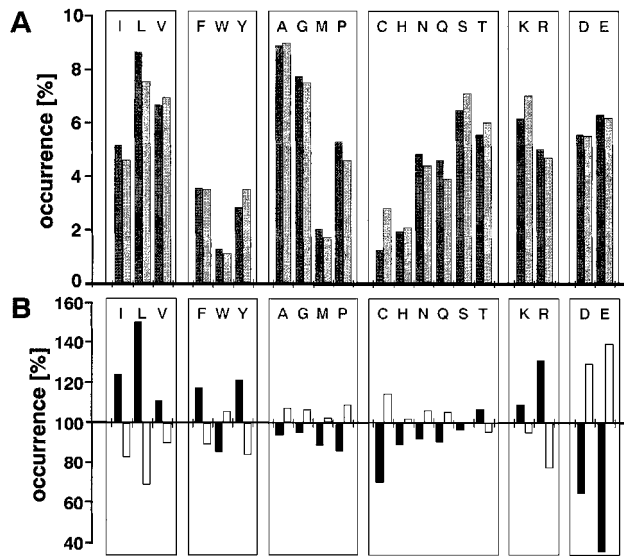
We developed an algorithm to predict DnaK binding sites. It is based on differential scoring of the statistical



**Fig. 1.** DnaK binding to cellulose-bound peptide scans and localization of DnaK binding sites in native protein structures. **(A)** Peptide scans derived from alkaline phosphatase precursor (AP), catabolite activator protein (CAP), firefly luciferase, influenza hemagglutinin (A/Aichi/2/68) and human p53 and pro-insulin were screened for DnaK binding. The last spots of rows (right) and N-terminal residues of peptides of the first spots of rows (left) are indicated. **(B)** Ribbon and space filling representations (INSIGHT II, Biosym) of the structures of the corresponding native proteins (the DNA binding fragment and tetramerization domain in the case of p53, mature forms in the case of AP and insulin, conformation at pH 7 in the case of HA) are shown (Wilson *et al.*, 1981; Hua *et al.*, 1991; Kim and Wyckoff, 1991; Schultz *et al.*, 1991; Cho *et al.*, 1994; Jeffrey *et al.*, 1995; Conti *et al.*, 1996). Red and pink segments indicate good DnaK binding sites and DnaK binding sites with weaker affinity respectively; backbone atoms are yellow. DnaK binding sites are in most cases completely buried, in a few cases some binding site residues are exposed.

energy contributions of each amino acid in a five residue core and two four residue flanking regions (Table I), together constituting the proposed DnaK binding motif (Figure 5A). The combined energy value obtained for a given sequence is taken as a measure of the likelihood that DnaK binds to this sequence (Figure 5B and C). For peptides with energy values between  $-5$  and  $-3$ , correct prediction is difficult. The existence of peptides with less clear DnaK binding

behavior reflects the fact that DnaK exhibits a continuum of substrate affinities. For peptides outside this intermediate energy window the predictability of DnaK binding sites is high. For example, out of the library peptides with energy values  $\leq -5$ , 82% are experimentally verified DnaK binders, out of peptides with values  $\geq -4$ , 82% are non-binders. With a cut-off at  $-5$ , 95% of the good binding sites (84% of total DnaK binding sites) were correctly predicted.

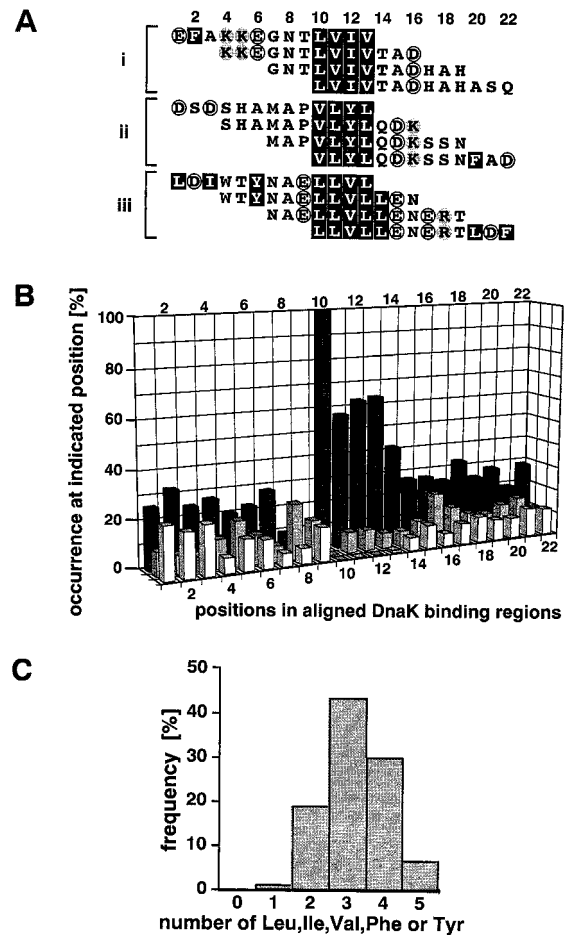


**Fig. 2.** Amino acid distribution in peptide scanning libraries. For 4360 peptides representing 37 protein sequences the relative amino acid occurrence was determined. (A) Entire peptide library (dark bars) compared with protein sequences taken from databases (Klapper, 1977) (light bars). (B) Total DnaK binding peptides (black bars) compared with non-binding peptides (light bars). The numbers for each amino acid in (B) are normalized to its occurrence in the whole peptide library (=100). The differences between binding and non-binding populations for Trp, His and Ser are not statistically significant ( $P \geq 0.05$ ).

## Discussion

This study reports the identification of binding sites for the DnaK chaperone in protein sequences. This information allowed us to elucidate the substrate binding motif and to derive an algorithm predicting DnaK binding sites and, therefore, yielded important information for further mechanistic analysis of DnaK-assisted folding reactions (Rüdiger *et al.*, 1997). DnaK binding sites were identified by a novel approach, the screening of cellulose-bound peptide scans, which should also prove useful for identification of the substrate specificities of other chaperones and, more generally, for mapping protein-protein contact sites. Furthermore, this is a powerful alternative to phage display techniques (Kramer *et al.*, 1995).

The high frequency of DnaK binding sites in protein sequences, occurring on average every 36 residues, is consistent with the promiscuity of DnaK association with various protein substrates. Assuming a similar frequency of binding sites for eukaryotic Hsp70s, this may be particularly important for organellar Hsp70s in promoting translocation of proteins across membranes (Schatz and Dobberstein, 1996). The simultaneous association of several chaperones with several binding sites of a nascent chain emerging at a translocation site may allow its efficient unidirectional pulling into the target compartment. Distinct binding sites may be involved in allowing cytosolic Hsp70s to maintain protein precursors in a translocation-competent state. In the case of DnaK, genetic evidence suggests a role in translocation of periplasmic alkaline phosphatase and maltose binding protein (Wild *et al.*, 1992). We found that the hydrophobic signal sequences of these proteins constitute DnaK binding sites (Figure 1A and data not shown). This may suggest that the role



**Fig. 3.** Sequence alignment of DnaK binding regions. (A) Three good DnaK binding regions [(i) alkaline phosphatase E376–Q379; (ii)  $\sigma^{32}$  D189–D210; (iii) hemagglutinin (A/Suita/1/89) L433–F454], each defined by DnaK binding to four overlapping peptides, are shown in sequence alignment. The overlapping segments constitute hydrophobic cores. The affinity of DnaK for peptide  $\sigma^{32}$  M195–N207 was determined in solution (McCarty *et al.*, 1996) and found to be 0.1  $\mu\text{M}$ . (B) Ninety DnaK binding regions each constituting a single good DnaK binding site were aligned as for (A). Hydrophobic cores were anchored with Leu, Ile, Val, Phe or Tyr at position 10 by shifting the sequences by  $\leq 2$  residues. Black bars, Leu, Ile, Val, Phe and Tyr; gray bars, Arg and Lys; white bars, Glu and Asp. (C) Sizes of hydrophobic cores in individual good DnaK binding regions. For each of the 90 aligned DnaK binding regions the Leu, Ile, Val, Phe and Tyr residues in the hydrophobic cores (positions 10–14) were counted and the total frequency of these numbers are presented.

of DnaK in translocation of these proteins relies on direct association with the signal sequence. Alternatively, the signal sequences may affect the kinetics of folding of the mature parts of the precursors such that additional sites are rendered accessible for DnaK association.

The affinities by which DnaK associates with binding sites that we identified by peptide screening are high, with  $K_d$  values as low as 100 nM. These affinities are about one (Flynn *et al.*, 1989) or two (Blond-Elguindi *et al.*, 1993) orders of magnitude higher than the affinities of BiP for peptides identified earlier and reported for the peptide co-crystallized with the DnaK substrate binding domain (Burkholder *et al.*, 1996), showing that our approach identified high affinity binding sites for DnaK. These results furthermore suggest that the mode of action of DnaK, and probably of other Hsp70 family members,

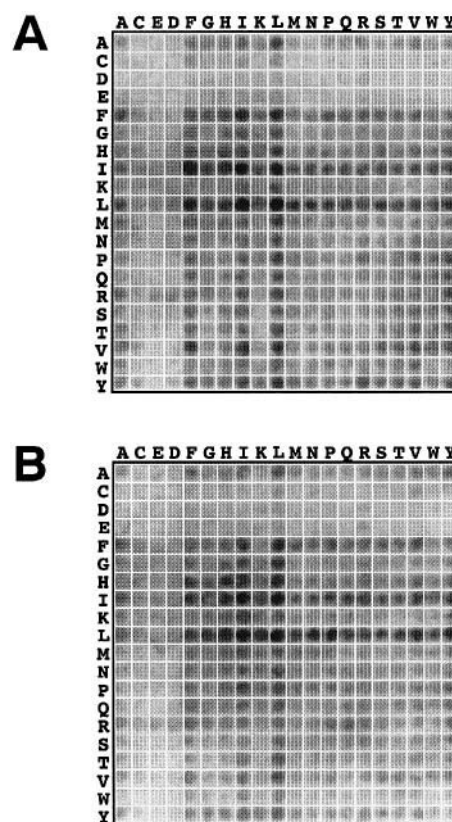
**Table I.** Amino acid content of core regions of 90 good DnaK binding sites (Figure 3B) and  $\Delta\Delta G_K$  values for the algorithm (13) predicting DnaK binding sites

Amino acid	Amino acid content (%)		$\Delta\Delta G_K$ (kJ/mol)		
	Core	Non-binders	Left	Core	Right
Leu	29.1	6.0	1.70	-3.62	-0.02
Ile	10.7	4.3	1.04	-2.05	0.11
Val	12.4	6.0	-0.26	-1.75	0.70
Phe	4.9	3.2	0.14	-1.17	0.53
Trp	0.9	1.4	-0.43	3.49	0.12
Tyr	6.7	2.4	0.19	-1.88	1.15
Ala	7.6	9.6	-0.07	0.79	0.46
Gly	3.6	8.3	-0.33	1.95	0.03
Met	0.4	2.1	0.08	1.10	0.17
Pro	2.9	5.8	0.15	1.63	-0.28
Cys	0.0	1.4	4.87	6.35	0.25
His	1.1	2.0	-0.24	1.74	0.09
Asn	1.6	5.2	0.74	2.36	-0.29
Gln	2.2	4.9	-1.13	1.60	-0.15
Ser	2.9	6.5	-0.13	1.27	-0.23
Thr	5.6	5.3	-0.91	0.27	-0.48
Lys	3.6	5.9	-0.85	0.40	-1.08
Arg	2.7	3.9	-1.19	-0.79	-1.72
Asp	0.9	7.2	0.44	4.91	0.35
Glu	0.4	8.8	1.48	5.14	1.65

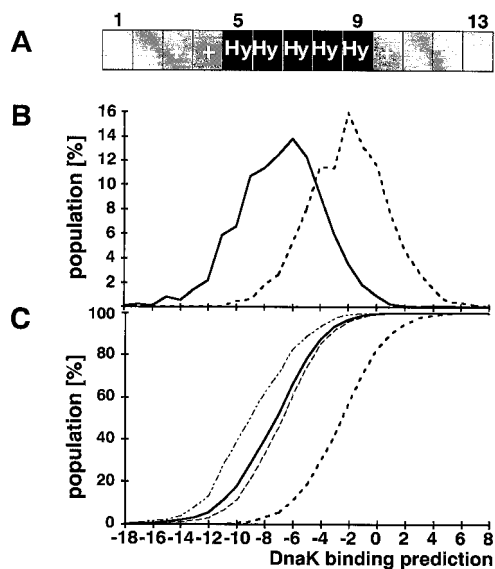
The relative amino acid distribution within the five residue hydrophobic core region, 124 left and 116 right boundaries of DnaK binding regions were compared with non-binding peptides of the library. Averaged energy distributions ( $\Delta\Delta G_K$ ) for each amino acid were calculated for the hydrophobic core (five residues) and the two left and two right flanking residues, together composing the DnaK binding motif (Figure 5A).  $\Delta\Delta G_K$  values were obtained according to  $\Delta\Delta G_K = -RT[\ln(P_b/P_n)]$ ,  $P_b$  and  $P_n$  being the relative occurrence of each amino acid in the binding motif and in non-binding peptides respectively.

relies on high affinity association with substrates, thereby allowing precise control of the interaction by ATP and chaperone cofactors. ATP acts to dissociate DnaK-substrate complexes and to allow DnaJ to kinetically target DnaK-ATP to substrates, but it is not needed to allow DnaK binding to substrates *per se*. It cannot be excluded at present that DnaJ and ATP/Mg<sup>2+</sup> may in addition modulate the specificity of substrate recognition by DnaK. Further studies are required to determine whether such modulating activity exists.

The consensus motif recognized by DnaK consists of a central hydrophobic core of four to five residues and two flanking regions, of approximately four residues each, that are enriched in basic residues. The features of this motif agree well with those of the structure of the substrate binding domain of DnaK (Zhu *et al.*, 1996; Rudiger *et al.*, 1997). The binding cavity is suited to interact with approximately five consecutive residues. A central hydrophobic pocket, which makes the major energy contribution to binding of the co-crystallized heptapeptide, is tailored to bind Leu, but also Ile and Val, although probably yielding lower binding energies. However, the crystal structure of the substrate binding domain in complex with the peptide substrate did not elucidate the entire amino acid spectrum capable of associating with the substrate binding cavity, in particular with the four positions outside the central hydrophobic pocket, and the contributions of negatively charged residues surrounding the binding cavity (Zhu *et al.*, 1996). Our study defines (i) the consensus binding motif including the entire amino acid spectrum capable of associating with DnaK (Figure 5A and Table I), (ii) the residues disfavored in binding sites (in particular Glu and Asp), (iii) the number of hydrophobic residues



**Fig. 4.** DnaK binding to combinatorial peptide libraries  $X_5B^1XB^2X_5$  (A) and  $X_5B^1XXB^2X_4$  (B).  $B^1$  (rows) and  $B^2$  (columns), defined positions; X, randomized positions (19 amino acids, Cys omitted) (Kramer *et al.*, 1994; Kramer and Schneider-Mergener, 1997).



**Fig. 5.** Prediction of DnaK binding sites in protein sequences. (A) The DnaK binding motif used to establish an algorithm consists of a five residue hydrophobic core (Hy) flanked by four residue segments (+, with weights for the algorithm decreasing with increasing distance from the core) enriched in Arg and Lys. (B) Energy score distribution as a percentage of total DnaK binding (—) or non-binding (···) peptides. (C) Score prediction validity as percentage for total DnaK binding (—), good binding (---), binding (---) and non-binding (···) peptides.

present in individual good DnaK binding sites (two to four on average), (iv) the extraordinary importance of Leu for DnaK binding (present in 87% of the tested hydrophobic cores of DnaK binding sites) and (v) a role for basic residues adjacent to the hydrophobic core, most likely allowing electrostatic interactions with the negatively charged surface surrounding the substrate binding cavity. This information on the consensus binding motif as well as the sequence identity of many individual DnaK binding regions should provide a basis for further dissection of structural features of DnaK–substrate complexes.

The identification of the motif and the energy contributions of each amino acid for binding to DnaK allowed us to establish a novel algorithm predicting DnaK binding sites in protein sequences with high accuracy. In contrast, a scoring system previously developed for the BiP homolog on the basis of a statistical analysis of BiP binding peptides displayed on phages (Blond-Elguindi *et al.*, 1993) failed in the case of DnaK to distinguish between binding and non-binding peptides. Some differences in substrate specificity have indeed been observed between DnaK and BiP which may contribute to this failure (Fourie *et al.*, 1994; Gragerov and Gottesman, 1994). However, it has to be emphasized that the general features of the substrate binding sites of Hsp70 proteins are conserved in the Hsp70 family (Zhu *et al.*, 1996; Rüdiger *et al.*, 1997). The BiP scoring system, as well as the substrate binding motif proposed for BiP (Blond-Elguindi *et al.*, 1993), are difficult to reconcile with these structural features. A definite answer with respect to the degree of evolutionary conservation of the substrate specificity of Hsp70 proteins clearly needs further systematic analysis.

The side chain profile of substrate binding sites of DnaK is typical of segments as they exist in the cores of

folded proteins (Figure 1B), in agreement with earlier proposals (Flynn *et al.*, 1991; Blond-Elguindi *et al.*, 1993). It is remarkable that we found a majority of DnaK binding sites in peptides forming  $\beta$ -sheets in folded proteins. Peptides forming helices in folded proteins are under-represented among binders, although their short length in the screened library (13mers) is likely to result in predominately unstructured states. DnaK is thus capable of distinguishing, though not exclusively, secondary structure elements by recognizing primary structures. This capability is reflected for instance in the disfavoring in DnaK binding sites of the acidic residues Asp and Glu and of Pro, all of which have a strong potential to break  $\beta$ -structures (Chou and Fasman, 1974). Furthermore, the hydrophobic core motif of DnaK binding sites has a preference for stretches of consecutive hydrophobic residues which are atypical for amphiphilic helices.

It is important to emphasize that the association of a substrate with DnaK not only involves side chain contacts but also hydrogen bonding and van der Waals interactions with the extended peptide backbone (Zhu *et al.*, 1996). This dual requirement probably prevents DnaK from binding to those sites where a stretch of hydrophobic side chains is exposed at the protein surface while the corresponding peptide backbone remains inaccessible to solvent (e.g. luciferase and insulin; Figure 1B) (Rüdiger *et al.*, 1997). These features constitute the molecular basis for the ability of DnaK to differentiate between native and non-native protein conformers.

## Materials and methods

### Protein sequences screened by cellulose-bound peptide scans

DnaK/Hsp70 substrates: cI, P and CIII of phage  $\lambda$ ; DnaK, DnaA,  $\sigma^{32}$ , pro-alkaline phosphatase, pro-maltose binding protein and  $\beta$ -galactosidase of *E. coli*; RepA of plasmid P1; influenza hemagglutinin from strains A/Aichi/2/68 and A/Suita/1/89; heat shock transcription factor 1, p53 and pre-insulin of human; *Photinus pyralis* luciferase; *S. aureus* protein A; bovine clathrin light chains A and B; immunoglobulin chains  $\lambda$ -1 and  $\lambda$  V region heavy chain of *Mus musculus*; cytochrome b2, F1 $\beta$  and Su9 of *S. cerevisiae*. Candidate substrates: DnaJ, FtsZ of *E. coli*; pro-bovine pancreas trypsin inhibitor (BPTI); RepE of plasmid miniF;  $\lambda$  cII. Others: RNase T1, CheY, catabolite activator protein, FtsA, ribosomal protein L2 and proOmpA of *E. coli*; *S. cerevisiae* propro- $\alpha$ -factor.

### Screening of cellulose-bound peptide scans for DnaK binding

Peptide libraries were prepared by automated spot synthesis (Frank, 1992; Kramer *et al.*, 1994; Kramer and Schneider-Mergener, 1997) (for further information contact L.Germeroth). Peptides were C-terminally attached to cellulose via a ( $\beta$ -Ala)<sub>2</sub> spacer. Before screening, the dry membranes were washed in methanol for 10 min and for 3 $\times$ 20 min in Tris-buffered saline (TBS; 31 mM Tris-HCl, pH 7.6, 170 mM NaCl, 6.4 mM KCl). DnaK (100 nM, purified as described; Buchberger *et al.*, 1994) was allowed to react with peptide scans in MP buffer (31 mM Tris-HCl, pH 7.6, 170 mM NaCl, 6.4 mM KCl, 0.05% Tween 20, 5.0% sucrose) for 1 h at 25°C with gentle shaking. Unbound DnaK was removed with TBS (4°C) and peptide-bound DnaK electrotransferred onto polyvinylidene difluoride (PVDF) membranes using a semi-dry blotter (Phase GmbH, Lübeck, Germany). The PVDF membranes were sandwiched between blotting paper soaked with cathode buffer (25 mM Tris base, 40 mM 6-aminohexane acid, 0.01% SDS, 20% MeOH) and one of the anode buffers (AI: 30 mM Tris base, 20% MeOH; AII: 300 mM Tris base, 20% MeOH) kept at 4°C. Electrotransfer was performed at a constant power of 0.8 mA/cm<sup>2</sup> peptide cellulose. Transferred DnaK was detected with DnaK-specific polyclonal rabbit antisera using a chemiluminescence blotting substrate (POD) kit (Boehringer Mannheim). Reference spots to define affinity limits for DnaK were AKTLILSHLRFVV for good binders and VVHIARNYA-

GYG for binders. In solution these peptides yielded dissociation constants for DnaK binding of 0.4 and 4  $\mu\text{M}$  respectively, as determined according to McCarty *et al.* (1996).

#### Calculation of energy contributions of amino acids for DnaK binding and establishment of an algorithm

The statistical energy contributions ( $\Delta\Delta G_K$ ) of individual amino acids were determined for 13 positions (a core of five residues and two flanking regions of four residues each) of DnaK binding sites (Table I). These values were the basis for an algorithm predicting DnaK binding sites. The amino acid distribution in flanking regions was assigned by alignment of 124 left and 116 right borders of hydrophobic cores of DnaK binding regions. The borders were anchored at the first or last enriched residue of each core region. The quality of the algorithm to predict DnaK binding sites was judged by comparison with experimental data obtained for the 4360 peptides investigated. An optimal algorithm required weighting of the relative importance of  $\Delta\Delta G_K$  for each position for DnaK binding. This was done by multiplying the  $\Delta\Delta G_K$  values by correction factors specific for each position within the motif: left flanking region, position 1, 0.33; 2, 0.66; 3, 1.0; 4, 1.5; hydrophobic core, positions 5–9, each 1; right flanking region, position 9, 1.5; 10, 1.0; 11, 0.66; 12, 0.33. The algorithm attributes overall  $\Delta\Delta G_K$  values to segments of 13 adjacent residues. The lower the  $\Delta\Delta G_K$  value obtained for a specific segment the higher the predicted affinity for DnaK. Longer sequences can be scanned for  $\Delta\Delta G_K$  minima by one-step movements, starting with position 5 of the algorithm at N-termini and ending with position 10 at C-termini. A program which allows the use of this algorithm to identify DnaK binding sites in protein sequences is available upon request.

#### Acknowledgements

We thank H.Bujard for support of this work, J.S.McCarty for discussions and protein database searches, W.Hendrickson and X.Zhao for discussions, B.Rainer for help with statistical analysis, T.Laufen and Y.Cully for image processing, P.Brick for pictures of DnaK binding sites in firefly luciferase and A.Buchberger and T.Hestekamp for reviewing the manuscript. This work was supported by grants from the DFG, the BMBF and Fonds der Chemischen Industrie to B.B. and J.S.-M.

#### References

- Blond-Elguindi, S., Cwiria, S.E., Dower, W.J., Lipshutz, R.J., Sprang, S.R., Sambrook, J.F. and Gething, M.-J.H. (1993) Affinity panning of a library of peptides displayed on bacteriophages reveals the binding specificity of BiP. *Cell*, **75**, 717–728.
- Buchberger, A., Valencia, A., McMacken, R., Sander, C. and Bukau, B. (1994) The chaperone function of DnaK requires the coupling of ATPase activity with substrate binding through residue E171. *EMBO J.*, **13**, 1687–1695.
- Burkholder, W.F., Zhao, X., Zhu, X. and Hendrickson, W.A. (1996) Mutations in the C-terminal fragment of DnaK affecting peptide binding. *Proc. Natl. Acad. Sci. USA*, **93**, 10632–10637.
- Cho, Y., Gorina, S., Jeffrey, P.D. and Pavletich, N.P. (1994) Crystal structure of a p53 tumor suppressor–DNA complex: understanding tumorigenic mutations. *Science*, **265**, 346–355.
- Chou, P.Y. and Fasman, G.D. (1974) Prediction of protein conformation. *Biochemistry*, **13**, 222–234.
- Clarke, C.F., Cheng, K., Frey, A.B., Stein, R., Hinds, P.W. and Levine, A.J. (1988) Purification of complexes of nuclear oncogene p53 with rat and *Escherichia coli* heat shock proteins: *in vitro* dissociation of hsc70 and DnaK from murine p53 by ATP. *Mol. Cell. Biol.*, **8**, 1206–1215.
- Conti, E., Franks, N.P. and Brick, P. (1996) Crystal structure of firefly luciferase throws light on a superfamily of adenylate-forming enzymes. *Structure*, **4**, 287–298.
- Flynn, G.C., Chappell, T.G. and Rothman, J.E. (1989) Peptide binding and release by proteins implicated as catalysts of protein assembly. *Science*, **245**, 385–390.
- Flynn, G.C., Pohl, J., Flocco, M.T. and Rothman, J.E. (1991) Peptide-binding specificity of the molecular chaperone BiP. *Nature*, **353**, 726–730.
- Fourie, A.M., Sambrook, J.F. and Gething, M.-J.H. (1994) Common and divergent peptide binding specificities of Hsp70 molecular chaperones. *J. Biol. Chem.*, **269**, 30470–30478.
- Frank, R. (1992) Spot synthesis: an easy technique for positionally

- addressable, parallel chemical synthesis on a membrane support. *Tetrahedron*, **48**, 9217–9232.
- Gething, M.-J.H. and Sambrook, J.F. (1992) Protein folding in the cell. *Nature (Lond.)*, **355**, 33–45.
- Gragerov, A. and Gottesman, M.E. (1994) Different peptide binding specificities of Hsp70 family members. *J. Mol. Biol.*, **241**, 133–135.
- Gragerov, A., Zeng, L., Zhao, X., Burkholder, W. and Gottesman, M.E. (1994) Specificity of DnaK–peptide binding. *J. Mol. Biol.*, **235**, 848–854.
- Hua, Q.X., Shoelson, S.E., Kochoyan, M. and Weiss, M.A. (1991) Receptor binding redefined by a structural switch in a mutant human insulin. *Nature*, **354**, 238–241.
- Jeffrey, P.D., Gorina, S. and Pavletich, N.P. (1995) Crystal structure of the tetramerisation domain of the p53 tumor suppressor at 1.7 Angstroms. *Science*, **267**, 1498–1501.
- Kim, E.E. and Wyckoff, H.W. (1991) Reaction mechanism of alkaline phosphatase based on crystal structures. *J. Mol. Biol.*, **218**, 449–464.
- Klapper, M.H. (1977) The independent distribution of amino acid near neighbor pairs into polypeptides. *Biochem. Biophys. Res. Commun.*, **78**, 1018–1024.
- Kramer, A. and Schneider-Mergener, J. (1997) Synthesis and screening of peptide libraries on continuous cellulose membrane supports. *Methods Mol. Biol.*, in press.
- Kramer, A., Schuster, A., Reineke, U., Malin, R., Volkmer-Engert, R., Landgraf, C. and Schneider-Mergener, J. (1994) Combinatorial cellulose-bound peptide libraries: screening tools for the identification of peptides that bind ligands with predefined specificity. *Methods Companion Methods Enzymol.*, **6**, 388–395.
- Kramer, A., Vakalopoulou, E., Schleuning, W.-D. and Schneider-Mergener, J. (1995) A general route to fingerprint analyses of peptide–antibody interactions using a clustered amino acid peptide library: comparison with a phage display library. *Mol. Immunol.*, **32**, 459–465.
- Landry, S.J., Jordan, R., McMacken, R. and Gierasch, L.M. (1992) Different conformations for the same polypeptide bound to chaperones DnaK and GroEL. *Nature*, **355**, 455–457.
- McCarty, J.S., Buchberger, A., Reinstein, J. and Bukau, B. (1995) The role of ATP in the functional cycle of the DnaK chaperone system. *J. Mol. Biol.*, **249**, 126–137.
- McCarty, J.S., Rüdiger, S., Schönfeld, H.-J., Schneider-Mergener, J., Nakahigashi, K., Yura, T. and Bukau, B. (1996) Regulatory region C of the *E. coli* heat shock transcription factor,  $\sigma^{32}$ , constitutes a DnaK binding site and is conserved among eubacteria. *J. Mol. Biol.*, **256**, 829–837.
- Morimoto, R.I., Tissieres, A. and Georgopoulos, C. (1994) *The Biology of Heat Shock Proteins and Molecular Chaperones*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Reineke, U., Sabat, R., Kramer, A., Stigler, R.-D., Seifert, M., Michel, T., Volk, H. and Schneider-Mergener, J. (1995) Mapping protein–protein contact sites using cellulose-bound peptide scans; *Mol. Diversity*, **1**, 141–148.
- Rüdiger, S., Buchberger, A. and Bukau, B. (1997) Principles governing the interactions of Hsp70 chaperones with substrates. *Nature Struct. Biol.*, in press.
- Schatz, G. and Dobberstein, B. (1996) Common principles of protein translocation across membranes. *Science*, **271**, 1519–1526.
- Schultz, S.C., Shields, G.C. and Steitz, T.A. (1991) Crystal structure of a CAP–DNA complex: the DNA is bent by 90°. *Science*, **253**, 1001–1007.
- Wild, J., Altman, E., Yura, T. and Gross, C.A. (1992) DnaK and DnaJ heat shock proteins participate in protein export in *Escherichia coli*. *Genes Dev.*, **6**, 1165–1172.
- Wilson, I.A., Skehel, J.J. and Wiley, D.C. (1981) Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. *Nature*, **289**, 366–373.
- Zhu, X., Zhao, X., Burkholder, W.F., Gragerov, A., Ogata, C.M., Gottesman, M. and Hendrickson, W.A. (1996) Structural analysis of substrate binding by the molecular chaperone DnaK. *Science*, **272**, 1606–1614.

Received on October 14, 1996; revised on November 21, 1996