

BiP-binding Sequences in HIV gp160

IMPLICATIONS FOR THE BINDING SPECIFICITY OF BiP*

(Received for publication, July 14, 1999, and in revised form, August 9, 1999)

Gerhard Knarr‡, Susanne Modrow§, Alicia Todd¶, Mary-Jane Gething¶, and Johannes Buchner‡¶

From the ‡Institut für Biophysik & Physikalische Biochemie, Universität Regensburg, 93040 Regensburg, Germany,

§Institut für Medizinische Mikrobiologie und Hygiene, Universität Regensburg, 93042 Regensburg, Germany, and

¶Department of Biochemistry and Molecular Biology, University of Melbourne, Parkville, Victoria 3052, Australia

BiP, a resident endoplasmic reticulum member of the HSP70 family of molecular chaperones, associates transiently with a wide variety of newly synthesized exocytotic proteins. In addition to immunoglobulin heavy and light chains, the first natural substrates identified for BiP, a number of viral polypeptides including the human immunodeficiency virus type 1 envelope glycoprotein gp160 interact with BiP during their passage through the endoplasmic reticulum. We have used a computer algorithm developed to predict BiP-binding sites within protein primary sequences to identify sites within gp160 that might mediate its association with BiP. Analysis of the ability of 22 synthetic heptapeptides corresponding to predicted binding sites to stimulate the ATPase activity of BiP or to compete with an unfolded polypeptide for binding to BiP indicated that about half of them are indeed recognized by the chaperone. All of the confirmed binding sites are localized within conserved regions of gp160, suggesting a conserved role for BiP in the folding of gp160. Information on the characteristics of confirmed BiP-binding peptides gained in this and previous studies has been utilized to improve the predictive power of the BiP Score algorithm and to investigate the differences in peptide binding specificities of HSP70 family members.

The immunoglobulin heavy chain binding protein, BiP, also known as Grp78, is an ER¹-specific member of the family of HSP70 molecular chaperones, which are involved in a variety of cellular processes including folding and assembly of newly synthesized polypeptides and translocation of proteins across membranes (reviewed in Refs. 1–6). To fulfill these functions, HSP70 chaperones must be able to recognize a wide variety of target proteins that share no obvious sequence homologies while discriminating accurately between native and unfolded structures. Analyses of the binding of short peptides to BiP

(7–10) and other HSP70 family members (10–13) revealed that the binding motifs are degenerate but include a high proportion of hydrophobic residues that would normally be buried in the interior of folded proteins.

The weak ATPase activity of BiP is maximally stimulated by peptides containing at least seven residues (8). Affinity panning of a bacteriophage peptide display library identified a set of 114 octapeptides, enriched in aromatic and hydrophobic amino acids, which bound to BiP with high affinity (9). The binding motif derived in that study could be best described as Hy(Trp/X)HyXH₂Hy, where Hy is a large hydrophobic or aromatic acid and X is any amino acid. Binding normally required that a minimum of any two of the Hy positions be occupied. By comparing the abundance of each of the 20 amino acids at each position in the 114 BiP-binding peptides and 114 nonbinding peptides, a BiP Score algorithm was developed (9). A computer program that uses this algorithm predicts BiP-binding sites in natural proteins by scoring amino acid sequences with a moving window of seven residues. To each amino acid, an individual score is assigned, and the sum of the seven scores provides a measure for the binding probability of the heptapeptide.

When we previously used the BiP Score program to screen for BiP-binding sites within immunoglobulin chains, the first identified substrates for BiP (14), we found that the majority of the antibody sequences had a low probability of binding to BiP, and only a small number of potential binding sites were detected (15). Analysis of the ability of synthetic heptapeptides corresponding to potential binding sites in immunoglobulin heavy chains to stimulate the ATPase activity of BiP identified several authentic BiP-binding sequences. Peptides with scores ranging from +5 to +10 had a probability of 50% for binding to BiP, rising up to 80% for peptides with scores of ≥ 10 . The predictive power for peptides with negative scores that should not bind to BiP is close to 100% (15). The binding sequences were distributed within both the V_H and C_H domains, and the majority involved residues that participate in contact sites between the heavy and light chains. We therefore suggested that BiP chaperones the folding and assembly of antibody molecules by binding to hydrophobic regions on the surface of the isolated chains that subsequently participate in interchain contacts (15, 16).

BiP is not only involved in the folding and assembly of immunoglobulin light and heavy chains (17–24), but it also associates with a wide variety of other newly synthesized exocytotic proteins (reviewed in Refs. 1, 5, and 25–27). This association is not restricted to cellular proteins, since a number of viral polypeptides bind to BiP during their passage through the ER, including influenza hemagglutinin (16) vesicular stomatitis virus G protein (28), and HIV gp160 (29). Gp160 is the precursor of the HIV type 1 envelope glycoprotein, which is

* Work in the laboratory of J. B. was supported by the Deutsche Forschungsgemeinschaft, the Bundesministerium für Bildung und Forschung, and the Fonds der Chemischen Industrie. Work in the laboratory of M. J. G. was supported by the National Health and Medical Research Council of Australia. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¶ To whom correspondence should be addressed. Present address: Institut für Org. Chemie und Biochemie, Technische Universität München, Lichtenbergstraße 4, D-85747 Garching, Germany. Tel.: 49-89-289-13341; Fax: 49-89-289-13345; E-mail: johannes.buchner@ch.tum.de.

¹ The abbreviations used are: ER, endoplasmic reticulum; HIV-1, human immunodeficiency virus type 1; Hsc, heat shock cognate; HSP70, generic term for all members of the 70-kDa heat shock protein family; RCMLA, reduced and carboxymethylated lactalbumin; HA, hemagglutinin.

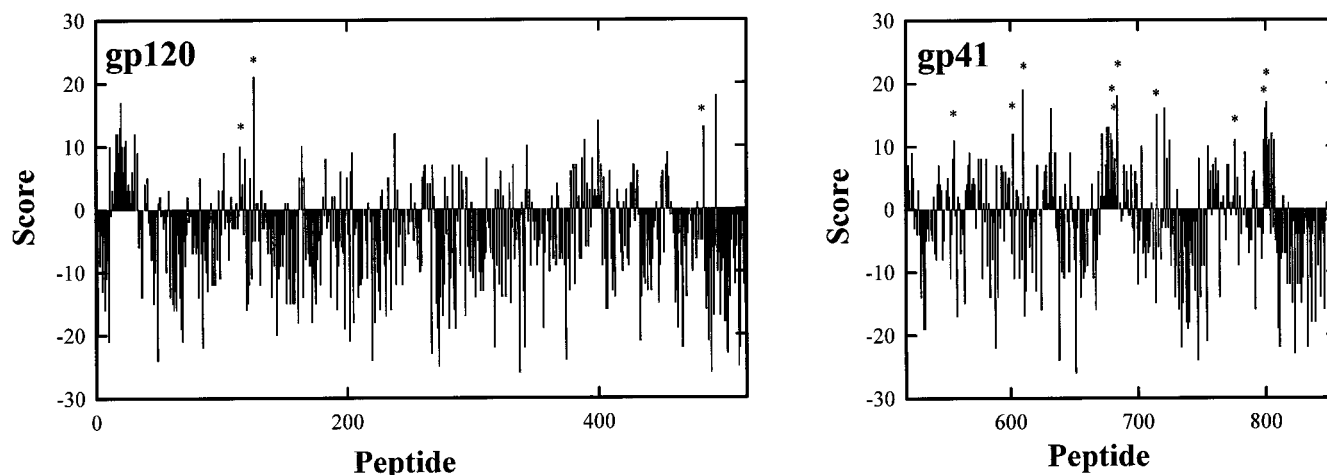


FIG. 1. **Prediction of BiP-binding sequences in the primary structure of gp120 and gp41.** Overall scores for each of the overlapping heptapeptides in the sequences were calculated using the BiP Score program described by Blond-Elguindi *et al.* (9) and plotted against the residue number of the first amino acid of each heptapeptide. The asterisks indicate the positive scoring sequences that, when tested as synthetic peptides, stimulated the ATPase activity of BiP (see Table I).

composed of two noncovalently linked subunits gp120 and gp41 (29). While gp120 is responsible for the adsorption of the virus to the CD4 receptor on the surface of the target cell, the gp41 transmembrane protein mediates cell fusion (30). Following its synthesis and translocation across the ER membrane, the gp160 precursor becomes heavily glycosylated in the ER lumen, where subsequently the formation of disulfide bridges and interaction with BiP takes place (29, 31). The gp160 molecule then folds into a state competent for CD4 binding, dissociates from BiP, and forms oligomers. After entering the Golgi complex, the protein is cleaved by a cellular furin protease into the gp120 and gp41 subunits, which are then transported to the plasma membrane (32). The half-life of the interaction between BiP and gp160 was determined to be 30 min (31). However, gp160 sequences that bind BiP and their location within the precursor have not previously been defined.

EXPERIMENTAL PROCEDURES

Construction of the Plasmid pUJ4—For the expression of BiP in the cytoplasm of *Escherichia coli*, a plasmid (pUJ4) containing the cDNA sequence for mature mouse BiP (33) was constructed by inserting a polymerase chain reaction fragment lacking the N-terminal signal sequence into a pASK40 vector (34). Using site-directed mutagenesis (35), an affinity tag of six histidine residues was introduced at the C terminus immediately after the KDEL retention sequence. The identity of the entire BiP construct was confirmed by DNA sequencing.

Purification of Recombinant BiP—*E. coli* B cells (36) were transformed with pUJ4. Cells were grown in LB medium at 29 °C, and protein synthesis was induced with isopropyl- β -D-thiogalactopyranoside (final concentration 1 mM) at an OD of 0.5. After 2 h of induction, cells were harvested by centrifugation. The bacterial pellet was resuspended in 40 mM Hepes, pH 7.0, 50 mM imidazole, 1 mM phenylmethylsulfonyl fluoride. Subsequently, cells were lysed, NaCl was added to a final concentration of 400 mM, and the crude extract was cleared by centrifugation. The supernatant was then applied to a column of nickel-nitrilotriacetic acid-agarose (Qiagen). BiP was eluted with 40 mM Hepes, pH 7.0, 300 mM imidazole, 400 mM NaCl and dialyzed against 40 mM Hepes, pH 7.8, 100 mM KCl, 10 mM $(\text{NH}_4)_2\text{SO}_4$, 4 mM MgCl_2 , 2 mM potassium acetate (Buffer L) and applied to a C8 ATP-agarose column (Sigma). After washing the column with buffer L containing 10 mM EDTA, but no MgCl_2 , BiP was eluted with buffer L supplemented with 0.5 M KCl and 4 mM MgATP . The eluted BiP was finally applied to a Superdex 200-pg gel filtration column (Amersham Pharmacia Biotech). The fractions of pure BiP were concentrated and stored in 40 mM Hepes, pH 7.5, containing 100 mM KCl and 5% (v/v) glycerol at -70 °C. The ATPase activity of recombinant BiP purified by this procedure could be stimulated up to 2.4-fold by synthetic peptides (see Table I). In this respect, it resembles more closely the recombinant His-tagged BiP prepared by Wei *et al.* (37) than recombinant BiP purified from the bacterial periplasm (38), which had a high basal ATPase activity that

was only slightly stimulated by peptide.

ATPase Measurements—ATPase assays were performed using [α - ^{32}P]ATP as described previously (15). The standard assay contained 40 mM Hepes, pH 7.0, 2 mM MgCl_2 , 500 μM unlabeled ATP, 10 mCi of [α - ^{32}P]ATP, and approximately 4 μg of recombinant BiP in a total volume of 20 μl . The stimulation of the ATPase activity of BiP by peptide was determined in the presence of concentrations of 1 μM to 1 mM of the respective peptide. Following different times of incubation, 3- μl aliquots were removed, and the amounts of ATP and ADP were determined by thin layer chromatography and liquid scintillation counting.

Competition by Peptides of Complex Formation between HSP70 Proteins and Reduced and Carboxymethylated Lactalbumin (RCMLA)—Competition binding assays that measure the ability of synthetic peptides to compete the binding of RCMLA to bovine liver BiP, bovine brain Hsc70, or *E. coli* DnaK were performed as described by Fourie *et al.* (10), except that bovine BiP was obtained from StressGen Biotechnologies Corp., and 50 ng instead of 1 μg of BiP or DnaK were used in each assay. After electrophoretic separation of the free HSP70 proteins from HSP70-RCMLA or HSP70-peptide complexes, Hsc70 was visualized by staining with Coomassie Brilliant Blue R-250 as described previously (10), while BiP and DnaK were visualized by immunoblotting with rabbit polyclonal antibodies raised against recombinant murine BiP (39) or DnaK (StressGen Biotechnologies Corp.) and detection using the ECL system (Amersham Pharmacia Biotech).

Peptide Synthesis and Purification—Peptides were synthesized using a model 9050 PepSynthesizer (Milligen) and Fmoc (9-fluorenylmethyloxycarbonyl)-protected amino acids (40) as described previously (15).

RESULTS AND DISCUSSION

Prediction of Potential BiP-binding Sites in the Sequences of the Human Immunodeficiency Virus Type 1 Envelope Protein—We took advantage of the BiP Score program (9), which has been used successfully previously to predict BiP-binding sites within the sequences of two different antibodies (15), to investigate another naturally occurring substrate of BiP, the HIV-1 envelope protein precursor gp160.

Fig. 1 shows the result of the scoring procedure for the sequence of gp160 from the HIV-1 isolate BH10 (41). The scores range from -26 to +21, but only 10% of the heptapeptides have scores greater than +6, indicating a 50% probability of binding to BiP. A very limited number of peptides (<4%) have scores greater than +10 and should therefore have a very high probability of binding. A cluster of about 20 peptides with positive scores are located within the N-terminal signal sequence (residues 1-36), which would be removed from the precursor soon after its translocation into the ER lumen. The presence of multiple peptides with high BiP Scores within signal sequences

TABLE I
Interactions of BiP with synthetic peptides derived from gp160

The peptide number indicates the position of the first amino acid of the heptapeptide within the sequence of gp160. The BiP scores were calculated using the BiP Score program described by Blond-Elguindi *et al.* (9) either using the original matrix of scores for each amino acid at each position of a heptapeptide or using a modified matrix developed after detailed comparison of the sequences of 62 binding and nonbinding peptides (see "Results and Discussion"). The degree of stimulation by each peptide (final concentration 500 μM) of the ATPase activity of recombinant BiP and the K_{app} values for peptide binding to bovine BiP were determined as described under "Experimental Procedures." K_{app} is defined as the concentration of peptide necessary for half-maximum competition of BiP binding to RCMLA. The values shown are averages of at least two assays (for peptides with low affinity ($K_{\text{app}} \geq 500 \mu\text{M}$) or at least three assays (for higher affinity peptides)).

Peptide	GP160 subdomain ^a	Sequence	BiP score	BiP ATPase stimulation factor	BiP binding (K_{app}) μM	Modified BiP score
BH115	C1	IISLWDQ	+10	1.6 \pm 0.2	500	+6
BH126	C1	CVKLTPL	+21	2.3 \pm 0.4	150	+21
BH164	V2	CSFNIST	+10	1.2 \pm 0.1	ND ^b	+6
BH238	C2	KTFNGTG	+12	1.1 \pm 0.2	>500	+8
BH343	C2	AKWNNLT	+10	1.1 \pm 0.3	>500	+9
BH389	C3	FYCNSSTQ	+11	1.3 \pm 0.4	>500	+10
BH400	V4	TWFNSTW	+14	1.2 \pm 0.2	>>500	+9
BH484	C5	NWRSELY	+13	2.3 \pm 0.3	ND	+17
BH494	C5	WKIEPL	+18	1.2 \pm 0.3	ND	+15
BH556	ED	QQQNNLL	+11	2.0 \pm 0.2	ND	+17
BH602	ED	WGCSGKL	+12	1.5 \pm 0.2	500	+20
BH610	ED	CTTAVPW	+19	1.7 \pm 0.2	500	+20
BH632	ED	MTWMEWD	+16	1.2 \pm 0.2	>500	+12
BH676	ED	WNWFNIT	+13	1.5 \pm 0.0	ND	+18
BH679	ED	FNITNWL	+12	1.5 \pm 0.3	ND	+12
BH684	ED	WLWYIKL	+18	1.8 \pm 0.3	>>100 ^c	+26
BH715	CP	RQGYSP	+15	1.7 \pm 0.1	500	+20
BH721	CP	LSFQTHL	+16	1.2 \pm 0.1	ND	+10
BH776	CP	RLRDLL	+11	2.4 \pm 0.2	180	+20
BH800	CP	KYWWNLL	+16	2.3 \pm 0.6	60	+22
BH801	CP	YWWNLLQ	+17	2.0 \pm 0.1	50 ^c	+20
BH805	CP	LLQYWSQ	+12	1.1 \pm 0.3	>>500	+4
HD177		HDFPAVL	+12	3.0 \pm 0.4 ^d	10–15	+11

^a See Fig. 4.

^b ND, not determined.

^c Peptides BH684 and BH801 caused aggregation of BiP at concentrations above 100 μM .

^d The ATPase activity of control peptide HD177 was reported previously by Knarr *et al.* (15).

has been noted previously (9). The other peptides with positive scores are distributed within smaller groups throughout the sequence of gp160. Within the conserved regions of the gp120 portion of the envelope protein (residues 1–517), nine peptides have scores of +10 or higher and were therefore strong candidates for binding to BiP. The constant region C1 contains peptide BH126, which has the most positive score (+21) of the complete gp160 sequence. In comparison, analysis of the sequence of the membrane-associated gp41 subunit (residues 518–862) revealed a higher percentage of possible BiP-binding sequences. Within the ectodomain of gp41 (residues 534–689), 10 peptides have very high scores from +10 up to +18. Interestingly, the hydrophobic fusion peptide, which lies at the N terminus of gp41 (FP, residues 518–533), does not include any predicted binding sites for BiP. This is because it contains a high proportion of small or β -branched hydrophobic residues, such as Ala, Val, and Ile, that are not favored in BiP-binding sequences (9). Similarly, although the transmembrane domain of gp41 (residues 690–711) has a high hydropathy index (42), it too contains a high proportion of hydrophobic residues (Val, Ile) that are not favored in BiP-binding sites and accordingly does not contain high scoring peptides. Potential binding sites within the transmembrane domain would in any case be buried in the lipid bilayer and should not be accessible to BiP in the lumen of the ER. Finally, the cytoplasmic portion of gp41 (residues 712–862) contains a significant proportion of peptides with positive scores, 12 peptides having scores of +10 or greater. These sequences would not be accessible to BiP, but they potentially could be recognized by cytosolic HSP70 family members.

Stimulation of the ATPase Activity by Peptides Corresponding to Possible BiP-binding Sites in gp160—To determine if the

predicted binding sites interact with BiP, we analyzed the influence on the ATPase activity of recombinant BiP of 22 synthetic heptapeptides corresponding to sequences having scores of +10 or greater, indicating a very high probability of binding. Peptide binding to BiP is indicated by a stimulation of the ATPase activity (7). The sequences and BiP Scores of these peptides are presented in Table I together with the results of the ATPase stimulation assays. Of the 22 peptides tested, six stimulated the ATPase activity by factors ranging from 2.0 to 2.4, similar to values reported earlier for synthetic peptides derived from viral proteins (7, 9) or immunoglobulin molecules (15). A further seven peptides reproducibly stimulated the ATPase activity by factors ranging from 1.5 to 1.8. The remaining nine peptides did not stimulate the ATPase activity and therefore do not bind to BiP.

The concentration dependence of ATPase stimulation was measured for the peptide BH126. The data shown in Fig. 2 yielded a K_m value (defined as the concentration of peptide causing half-maximal stimulation of the ATPase activity of BiP) of 28 μM . This value is within the range previously shown for high affinity binding sites (9, 10, 15).

Gp160 Peptides Compete with RCMLA for Binding to BiP—Fourteen of the high scoring gp160 peptides were tested for their ability to compete with the unfolded polypeptide RCMLA for binding to bovine BiP. Each peptide was tested over a range of concentrations (see examples in Fig. 3), and its apparent affinity for BiP (K_{app}) was measured as the concentration yielding 50% competition of RCMLA binding. We used as a positive control the Ig heavy chain peptide HD177, which we had previously shown to stimulate the ATPase activity of BiP with a K_m of 17 μM (15) and to compete with RCMLA to bind BiP with

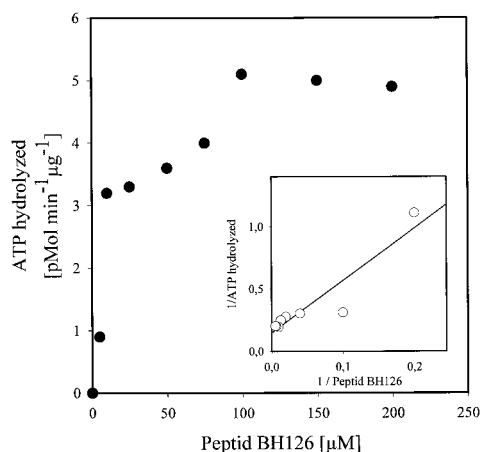


FIG. 2. **Stimulation of the ATPase activity of BiP by peptide BH126.** The concentration dependence of the stimulation of ATP hydrolysis by peptide BH126 was determined as described under "Experimental Procedures." The concentration of peptide necessary for half-maximal stimulation was calculated to be 28 μM .

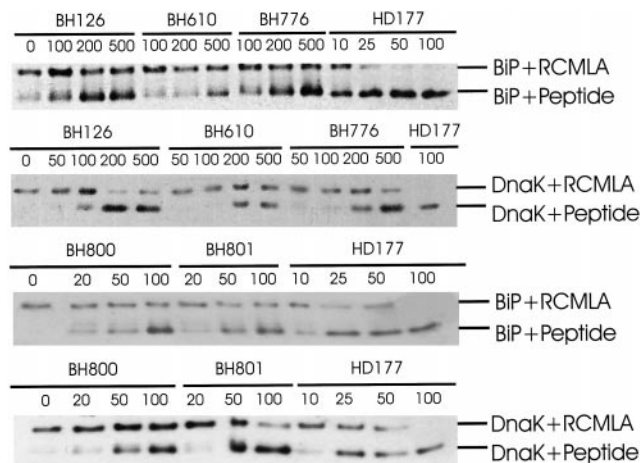


FIG. 3. **Competition by gp160 peptides of complex formation between BiP or DnaK and RCMLA.** The HSP70 proteins BiP or DnaK (50 ng/assay, 70 nM final concentration) were incubated with 40 μM RCMLA for 2 h at 37 $^{\circ}\text{C}$ in buffer A (25 mM Tris-HCl, 20 mM Hepes, pH 7.15, 47.5 mM KCl and 2.25 mM $\text{Mg}(\text{OAc})_2$) in the presence or absence of competing peptide (0–500 μM). Free HSP70 or similarly migrating HSP70-peptide complexes were separated from HSP70-RCMLA complexes by native PAGE (6% acrylamide) and visualized by immunoblotting with anti-BiP or anti-DnaK antibodies as described under "Experimental Procedures."

a K_{app} of 15 μM .² We observed a significant correlation between the K_{app} obtained for each gp160 peptide and the degree to which it stimulated the ATPase activity of recombinant BiP (Table I). Thus, peptides with stimulation factors of 2-fold or greater had the lowest K_{app} values (50–180 μM), while those with stimulation factors between 1.5 and 1.8 had intermediate K_{app} values (200–500 μM). Peptides with stimulation factors below 1.5 had very high K_{app} values ($>>500$ mM) or showed no evidence of binding.

Localization of BiP-binding Sites within the Sequence of gp160—Confirmed BiP-binding peptides are marked with *asterisks* in Fig. 1. The locations of these potential recognition sites in the sequence of gp160 are also shown in Fig. 4 within linear representations of the sequences of the HIV-1 envelope protein subunits gp120 and gp41. It should be noted that we do not claim to have identified all of the potential BiP-binding sites within gp160, since a small number of high scoring

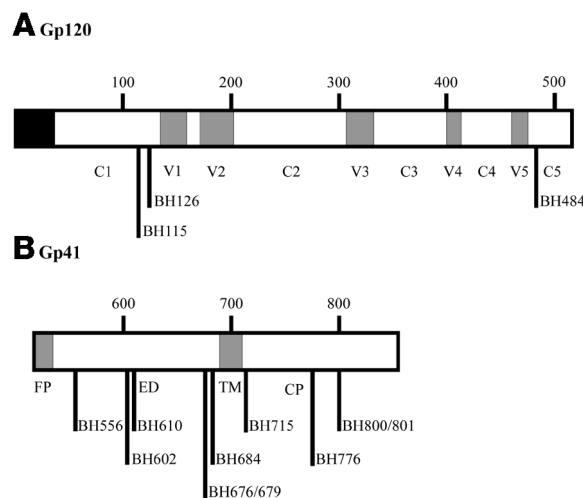


FIG. 4. **Potential BiP-binding sequences in the primary structures of gp120 and gp41.** The sequences of gp120 (top panel) and gp41 (bottom panel) are shown as linear maps. In the gp120 map, the constant regions C1–C4 are indicated by *open frames*, while the variable regions V1–V5 are *shaded*. The signal sequence is marked in *black*. In the gp41 map, the *shaded segments* represent the fusion peptide (FP) and the transmembrane domain (TM) respectively, while the ectodomain (ED) and the cytoplasmic domain (CP) are marked by *open frames*. The positions of BiP-binding peptides are indicated.

($\geq +10$) sequences could not be synthesized due to their amino acid composition. These and heptapeptides with positive scores between +5 and +9, or even between 0 and +5, may contain additional BiP-binding peptides, albeit at a significantly lower proportion than the high scoring population (9). Nevertheless, examination of the pattern of the scores within the various domains of the gp160 molecule (Fig. 1) and the locations of the confirmed binding peptides (Fig. 4) allows us to conclude that potential BiP-binding sites occur much more frequently in regions of the protein that are conserved between different HIV isolates. Thus, within the conserved regions of the gp120 portion of the envelope protein, 8% of the peptides have scores greater than +6. Seven of these have scores of +10 or higher and were therefore strong candidates for binding to BiP. Three were in fact confirmed as BiP-binding peptides (Table I and Fig. 4). By contrast, within the variable regions of gp120, 6% of the peptides have scores greater than +6, but only two peptides (BH400 and BH164, which is partly within a variable region) have scores of +10 or higher (Fig. 1), and both of these failed to stimulate the ATPase activity of BiP (Table I). Within the ectodomain of gp41, which is generally highly conserved (42), 19% of the peptides have scores greater than +6, and 10 of these have high scores from +10 up to +18 (Fig. 1). Seven high scoring sequences were tested for their ability to stimulate BiP's ATPase activity, and six were confirmed as BiP-binding peptides (Table I and Fig. 4).

Of the three binding peptides detected within the sequence of gp120, two (BH115 and BH126) are located close together at the N terminus of the protein within the conserved C1 region (Fig. 4). The third binding peptide, BH484, is located closer to the C terminus of gp120 in the conserved C5 region just following the hypervariable region V5. Sequences corresponding to the three peptides are absolutely conserved among seven different HIV-1 strains (42), with one exception; Val at position 2 in peptide BH126 is replaced by Ile in gp160 from HIV strain WMJ3. Since this is a conservative exchange of two hydrophobic amino acids, it is likely that this peptide would also bind to BiP. Inspection of the three-dimensional structure of an HIV-1 gp120 core complexed with a fragment of CD4 (43) reveals that the three potential BiP recognition sequences in gp120 are

² M. Beasley and M.-J. Gething, unpublished results.

located within secondary structural elements rather than loops in the folded protein. Notably, the side chains of all but one of the hydrophobic residues in the three sequences are inaccessible to solvent or are only partially solvent-accessible, consistent with previous observations that BiP-binding sites are hidden in the folded protein (29, 31). Interestingly, peptide 126 contains a cysteine residue (see Table I) that forms a disulfide bond in the mature molecule (43, 44). We have previously observed that Cys-containing peptides that bind BiP in their reduced state may lose their capacity to bind the chaperone when they form disulfide-bonded dimers.³ If this were the case when the BH126 sequence is disulfide-bonded, the corresponding site in the gp160 precursor might only be available for BiP binding at a very early stage of folding *in vivo*, before the disulfide bonds are formed. Biosynthetic studies have demonstrated that BiP binding to gp160 is a very early event that is indeed initiated before disulfide bond formation begins (29, 31). The same studies indicated that BiP's association with gp160 ceases shortly before the completion of disulfide bond formation. This interaction with incompletely oxidized proteins may be a general characteristic of BiP, since also for immunoglobulin light chains an interaction with BiP could only be observed before disulfide bond formation was completed (19, 24, 45).

Despite the membrane-associated gp41 subunit being significantly smaller than the gp120 subunit, 10 of the confirmed BiP-binding peptides were included within its sequence. Six of these peptides lie within the ectodomain of the mature protein and are thus potential *in vivo* recognition sites for BiP. Three of the peptides (BH676, BH679, and BH684) are in close proximity to the membrane-spanning part of the protein. The remaining three peptides (BH556, BH602, and BH610) are located within the core of the gp41 ectodomain whose structure was determined recently. x-ray crystallographic analysis of a gp41 fragment reconstructed from synthetic gp41 peptides (46) or of the gp41 core solubilized with a trimeric GCN4 coiled-coil in place of the fusion peptide (47) revealed a six-helix bundle formed by a gp41 trimer. The core of the molecule is an extended, triple-stranded α -helical coiled-coil separated by a linking region (whose structure was not defined) from three C-terminal α -helices that pack in the reverse direction against the outside of the coiled-coil. Peptide BH556 is located toward the N terminus of each of the extended, central α -helices. However, it should be noted that both Chan *et al.* (46) and Weissenhorn *et al.* (47) comment that the structure they have defined probably corresponds to that of the fusion active state of the molecule, which during virus infection *in vivo* is generated when the envelope glycoprotein binds to its receptor CD4 and undergoes a conformational change that results in dissociation of the gp120 subunit (48, 49). Weissenhorn *et al.* (47) further suggest that the conformational change in gp41 might include formation of the complete coiled-coil by extension at the N terminus and that the Gln-rich segments, which include those within the BiP-binding sequence QQQNLL (peptide BH556), might not be α -helical within native gp120/gp41. This is of interest, because the majority of HSP70-binding sites previously identified in proteins of known structure are located in β -strands or in regions of random coil (13, 15, 16, 50).

The other two BiP-binding peptides in the gp41 core, BH602 and BH610, are located in the proteolytically sensitive region (residues 598–629) that links the two α -helices in the fusion active structure (see above). The structure of this region is not known, either in the fusion active state or in the native envelope glycoprotein, but it contains a short disulfide loop (51, 52) that connects the sequences corresponding to the two peptides.

As discussed above, the corresponding sites in the gp160 precursor might only be available for BiP binding at a very early stage of folding *in vivo*, before the disulfide bonds are formed.

Interestingly, the structure of the fusion active gp41 core shows a striking similarity to the low pH-induced conformation of influenza hemagglutinin. In the case of influenza HA, high affinity binding sites for BiP are located in the stalk domain (16), which also folds into an intimately associated trimer (53) that undergoes a conformational change into the fusion active state (54, 55). Most interestingly, preliminary evidence exists⁴ for a BiP-binding site within a stretch of 36 residues of the HA stalk domain that undergoes a conformational change from a loop to α -helical structure to extend the coiled-coil into the fusion active conformation (56). This site would be in a position in HA analogous to that of peptide BH556 in gp41. The best defined BiP-binding site in the stalk domain of HA from the X31 influenza virus strain (ATLCLGH) lies near the N terminus of the molecule (16). The cysteine residue in this sequence is destined to form a disulfide bond late in the folding process just before the trimer is formed (53). Thus, a similar role for BiP in the folding and maturation of the virus envelope proteins gp160 and HA can be envisioned. By binding to conserved hydrophobic sequences within the viral glycoprotein precursors that are available only prior to folding and disulfide bond formation, BiP may prevent off pathway reactions like aggregation and assist correct folding and subunit assembly.

HSP70-binding Sites in the Cytoplasmic Domain of gp160—Four of the BiP-binding peptides that we identified in this study correspond to sequences located within the carboxyl-terminal part of gp41 (residues 713–862) that is situated on the cytoplasmic side of the ER membrane. Thus, these potential binding sites would not be available to BiP, which is confined to the lumen of the ER, but might be recognized by cytosolic HSP70 family members. We therefore used the competition binding assay described above to test the ability of bovine Hsc70 to bind the gp160 peptides, and we included as a positive control peptide V7, which we had previously shown to compete with RCMLA to bind Hsc70 with a K_{app} of 300 μ M (10). The results obtained are summarized in Table II. None of the gp160 peptides showed significant affinity for bovine Hsc70, although three of them (BH779, BH800, and BH801) bound BiP and DnaK with K_{app} values between 50 and 200 μ M. These results are consistent with observations that many peptides that bind BiP and DnaK display significantly lower affinities for Hsc70 (10).

Accuracy of Prediction of BiP-binding Sites within HIV gp160—Of 22 heptapeptides in gp160 identified as potential BiP-binding sites (BiP scores of $\geq +10$) using the scoring matrix developed using peptides displayed by bacteriophages (9), 13 were confirmed as BiP-binding sequences. Therefore, for gp160 the predictive power of the program was overall approximately 60%, which is lower than the value of 80% we previously reported for antibody sequences (15). For reasons we do not understand, the accuracy of prediction was much higher in the gp41 sequence (six out of seven correct in the ectodomain, four out of six correct in the cytoplasmic portion) than in the gp120 sequence (three out of nine correct).

Part of the explanation for lower than desired (*i.e.* 100%) prediction rates may be that the random population of bacteriophage-displayed peptides used to develop the scoring matrix do not accurately represent sequences within naturally occurring proteins. An example of this is that Trp was represented more highly in the random population than in proteins in the data base (9). From this work and our previous studies on

³ S. Blond-Elguindi and M.-J. Gething, unpublished data.

⁴ M.-J. Gething and P. S. Kim, unpublished data.

TABLE II
Comparison of the binding of gp160 peptides to the HSP70 chaperones BiP, Hsc70, and DnaK

The peptide number indicates the position of the first amino acid of the heptapeptide within the sequence of gp160. K_{app} values for peptide binding to bovine BiP, bovine Hsc70, and *E. coli* DnaK were determined using a competition assay as described under "Experimental Procedures." K_{app} is defined as the concentration of peptide necessary for half-maximum competition of the binding of the indicated chaperone (*i.e.* Hsp70 or DnaK or BiP) to RCMLA. The values shown are averages of at least two assays (for peptides with low affinity ($K_a \geq 500 \mu M$)) or at least three assays (for higher affinity peptides).

Peptide	GP160 subdomain ^a	Sequence	BiP binding (K_{app})	DnaK binding (K_{app})	Hsc70 binding (K_{app})
			μM	μM	μM
BH115	C1	IISLWDQ	500	500	ND ^b
BH126	C1	CVKLTPL	150	150	ND
BH164	V2	CSFNIST	ND	ND	ND
BH238	C2	KTFNGTG	>500	>500	ND
BH343	C3	AKWNNTL	>500	>>500	ND
BH389	C3	FYCNSTQ	>500	>500	ND
BH400	V4	TWFNSTW	>>500	>>500	ND
BH484	C5	NWRSELY	ND	ND	ND
BH494	C5	WKIEPL	ND	ND	ND
BH556	ED	QQQNNLL	ND	ND	ND
BH602	ED	WGCSGKL	500	>500	ND
BH610	ED	CTTAVPW	500	500	ND
BH632	ED	MTWMEWD	>500	>500	ND
BH676	ED	WNWFNIT	ND	ND	ND
BH679	ED	FNITNWL	ND	ND	ND
BH684	ED	WLWYIKL	>>100 ^c	>>100 ^c	ND
BH715	CP	RQGYSP	500	>500	500
BH721	CP	LSFQTHL	ND	ND	ND
BH776	CP	RLRDL	180	200	>500
BH800	CP	KYWNNLL	60	100	>500
BH801	CP	YWWNNLLQ	50 ^c	50 ^c	>>200 ^c
BH805	CP	LLQYWSQ	>>500	>500	600
HD177		HTFPAYL	10–15	20–25	ND
V7 ^c		FYQLAKT	ND (25) ^d	ND (25) ^d	400 (350) ^d

^a See Fig. 4.

^b ND, not determined.

^c Peptides BH684 and BH801 caused aggregation of the HSP70 proteins at concentrations above 100 μM .

^d The K_{app} values given in parentheses for control peptide V7 are those reported previously by Fourie *et al.* (10).

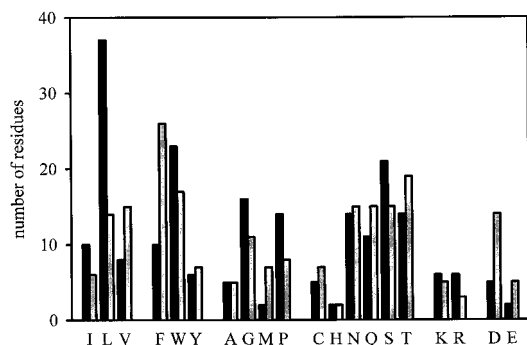


FIG. 5. Amino acid abundance of predicted BiP-binding sites identified in substrate proteins. The number of each of the 20 amino acids present in 31 predicted and confirmed BiP-binding peptides (black columns) is compared with their corresponding number within 31 predicted but nonbinding peptides (gray columns).

immunoglobulins (15), we have now generated a panel of 62 peptides that were selected on the basis of their high BiP scores (ranging from +5 to +21). Exactly half of these were confirmed to bind to BiP. Fig. 5 shows the representation of the different amino acid residues within the 31 confirmed BiP-binding peptides, compared with that within the 31 nonbinding peptides. It can be seen that some residues are found more frequently in binding peptides (black columns), suggesting that they have been given insufficient weighting in the BiP Score matrix, while others are enriched in nonbinding peptides (gray columns), indicating that they are given too much weighting. The most striking deviations are for Leu, Phe, Asp, and Glu. By comparing the sequences of the binding and nonbinding peptides in detail in a position-dependent manner, it is possible to recognize features specific to binding (or nonbinding) peptides

(see below) and to adjust the scoring matrix to optimize the accuracy of prediction by the BiP Score program. The details of this analysis will be presented elsewhere.⁵ However, Fig. 6 presents a comparison of the scores that were obtained using the original and modified scoring matrices for the 62 peptides, and the new scores for the gp160 peptides are presented alongside the old scores in Table I. It can be seen that the new matrix separates the populations of binding and nonbinding peptides with much greater accuracy and that a cut-off score of $\geq +11$ would yield BiP-binding peptides with greater than 80% accuracy (25 of 31 binders predicted correctly, only 4 of 31 nonbinders wrongly predicted). It should be noted that when the whole gp160 sequence was rescored (data not shown) the pattern of exclusion of likely binding sites from the variable regions of gp120 was not significantly altered. How accurately the new scoring matrix will predict BiP-binding sites in previously untested membrane and secretory proteins will be examined in the near future.

Implications for the Binding Specificity of HSP70 Chaperones—The pattern of sequence conservation within the HSP70 family is such that the backbone conformation of the β -sandwich domain that forms the peptide binding site should be virtually identical in all family members (57). It might therefore be expected that BiP, Hsc70, and DnaK would have similar binding motifs. However, although many binding sequences can be recognized by all three proteins, there are also significant differences in the chaperones' peptide recognition patterns (10), and we have already noted (see above and Table II) that peptides corresponding to sequences in the cytoplasmic domain of gp160 show quite different affinities for BiP or DnaK and

⁵ M.-J. Gething, G. Knarr, S. Modrow, and J. Buchner, manuscript in preparation.

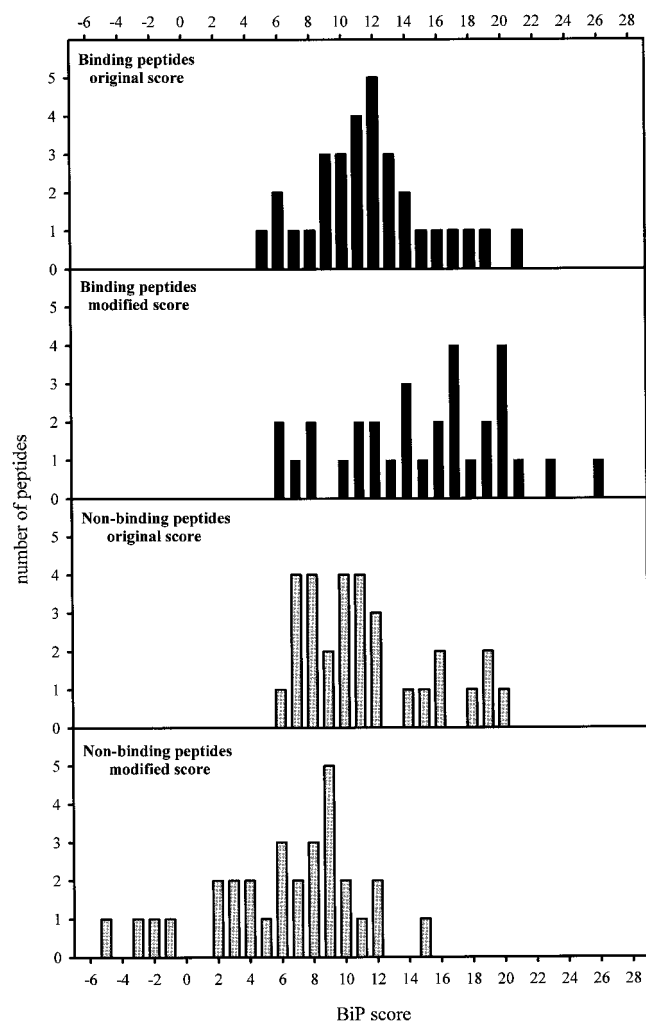


FIG. 6. Comparisons of the BiP scores of BiP-binding and non-binding peptides calculated using the original and modified scoring matrices. The BiP scores of 31 predicted and confirmed BiP-binding peptides are compared with the scores of 31 predicted but nonbinding peptides. The scores were calculated by the BiP Score computer program (9) either using the original matrix of individual scores for each of the 20 amino acids for each position of a heptapeptide or using a modified scoring matrix developed by detailed comparison of the 62 binding and nonbinding peptides.

Hsc70. Furthermore, the binding motifs predicted for BiP (9) and DnaK (13) display distinct features. Both motifs include hydrophobic residues that provide the basis of the ability of the chaperones to discriminate between unfolded and native proteins. In BiP-binding peptides, these hydrophobic residues are most frequently the bulky and/or aromatic amino acids Trp and Leu. These preferred residues can be located throughout the heptameric binding motif, often spaced with other residues in an alternating pattern. In DnaK-binding peptides, the hydrophobic residues are most frequently Leu, Ile, Val, and Tyr and are clustered in the central four or five residues of a somewhat longer motif. The two published motifs differ significantly with respect to Trp, which is highly favored in the BiP-binding motif but largely excluded from DnaK-binding sequences. Our analyses of BiP-binding peptides in gp160 and immunoglobulins (Fig. 7) confirm the importance of Trp in the BiP motif. Although Trp constitutes only 2% of the 1721 amino acids in the proteins analyzed, it constitutes 11% of the amino acids in the BiP-binding peptides that we have identified (Fig. 7A). Furthermore, nearly 50% of all the Trp residues in these proteins are located within the BiP-binding peptides (Fig. 7B). Other resi-

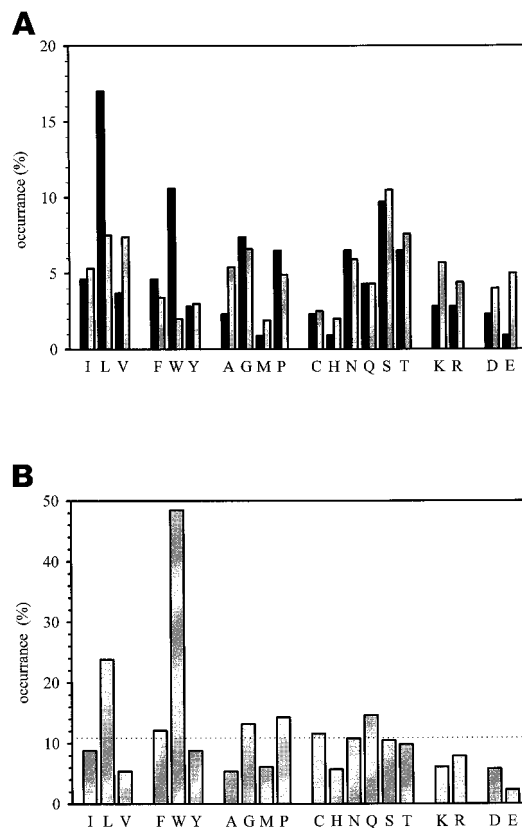


FIG. 7. Amino acid representation in BiP-binding sites identified in substrate proteins. A, the percentage representation of each of the 20 amino acids in sequences identified as BiP-binding peptides (total of 188 residues after adjustment for sequence overlap in 31 binding peptides comprising 217 residues) is shown as black columns and is compared with the percentage representation of the amino acids in the complete sequences of the gp160 and immunoglobulin proteins (1721 residues, gray columns). B, the percentage of the total number of each residue in the scored proteins that are present in identified BiP-binding peptides. The dotted line indicates the percentage (10.9%) of the 1721 amino acids of the substrate proteins that are included in the BiP-binding sequences and indicates the value expected for each amino acid if there was no sequence specificity in binding.

dues that are enriched in the hydrophobic core of the DnaK binding motif, *i.e.* Ile, Val, and Tyr, are actually underrepresented in the BiP-binding sequences (Fig. 7, A and B). This is particularly the case for Val, which is decreased in abundance by about 2-fold. Finally, the two motifs also vary significantly with respect to basic residues; Arg and Lys are favored in positions flanking the hydrophobic core in the DnaK motif but are largely excluded (particularly at the N terminus) from the BiP motif. This difference has been verified in peptide binding experiments (Ref. 10; see also Fig. 7, A and B).

Because many of the BiP-binding sites identified within the sequences of gp160 contain one or two Trp residues, we decided to test the prediction that they would interact poorly with DnaK. We assayed the ability of 15 gp160 peptides (including those with high, moderate, and low affinities for BiP) to compete for binding of RCMLA to DnaK. Interestingly, we observed only minor differences between the affinities of these peptides for BiP and DnaK (see Fig. 3 and Table II), indicating that the presence of Trp residues did not appear to be a negative factor for binding to DnaK. Thus, peptide BH801, which contains two Trp residues in positions 2 and 3, bound BiP and DnaK equally and with the highest affinities of all the peptides, while peptide BH800, which contains the same two Trp residues but in positions 3 and 4, bound only slightly better to BiP than to DnaK. Other Trp-containing peptides bound both chaperones with

lower but approximately equal affinities. We cannot explain why our data do not agree with the analysis by Rüdiger *et al.* (13) of cellulose-bound peptides that indicated that DnaK does not favor Trp-containing peptides, but it is clear that the presence of Trp does not preclude binding of gp160 peptides to DnaK.

In summary, we have taken advantage of the BiP Score program to search in the primary sequence of gp160 for possible BiP-binding sites. The scoring procedure revealed that almost all heptapeptides with a high probability of BiP binding are located within regions of gp160 whose sequences have been highly conserved in different HIV isolates (42). No binding peptides were identified within the hypervariable regions of the gp120 subunit, in agreement with a conserved role for BiP in the folding of gp160. This study, together with our previous analyses of immunoglobulins (15), has generated a panel of 62 heptapeptides whose affinities for BiP have been characterized. We have used this information to adjust the BiP Score matrix to improve the accuracy of prediction of BiP-binding sequences by the BiP Score computer program. Our studies have confirmed that BiP-binding peptides are enriched in hydrophobic amino acids, particularly Leu and Trp. BiP shares the prediction for Leu with the *Escherichia coli* HSP70 protein, DnaK (13), but the enrichment of Trp in BiP-binding peptides appeared in marked contrast to the exclusion of Trp reported for peptides that bind to DnaK (13). We therefore tested whether our Trp-containing gp160 peptides showed differential binding to BiP and DnaK but found no significant differences in their affinities for the two chaperones.

Acknowledgments—We thank Ingrid Haas for providing a cDNA encoding mouse BiP, Ursula Jakob and Rudi Glockshuber for performing the initial cloning experiments, Ursula Kies for help with data analysis, and Astrid Brunner for assistance during peptide synthesis.

REFERENCES

- Gething, M.-J., and Sambrook, J. F. (1992) *Nature* **355**, 33–45
- Becker, J., and Craig, E. A. (1994) *Eur. J. Biochem.* **219**, 11–23
- Hightower, L. E., Sadis, S. E., and Takenaka, I. M. (1994) in *The Biology of Heat Shock Proteins and Molecular Chaperones* (Morimoto, H. I., Tissieres, A., and Georgopoulos, C., eds) pp. 179–207, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- Buchner, J., (1996) *FASEB J.* **10**, 10–19
- Gething, M.-J. (1997) in *Molecular Chaperones and Protein Folding Catalysts* (Gething, M.-J., ed) pp. 59–65, Oxford University Press, Oxford
- Miao, B., Davis, J., and Craig, E. A. (1997) in *Molecular Chaperones and Protein Folding Catalysts* (Gething, M.-J., ed) pp. 3–13, Oxford University Press, Oxford
- Flynn, G. C., Chappel, T. G., and Rothman, J. E. (1989) *Science* **245**, 385–390
- Flynn, G. C., Pohl, J., Flocco, M. T., and Rothman, J. E. (1991) *Nature* **353**, 726–730
- Blond-Elguindi, S., Cwirla, S. E., Dower, W. J., Lipshutz, R. J., Sprang, S. R., Sambrook, J. F., and Gething, M.-J. (1993) *Cell* **75**, 717–728
- Fourie, A. M., Sambrook, J. F., and Gething, M.-J. (1994) *J. Biol. Chem.* **269**, 30470–30478
- Gragerov, A., and Gottesman, M. E. (1994) *J. Mol. Biol.* **235**, 848–854
- Takenaka, I. M., Leung, S.-M., McAndrew, S. J., Brown, J. P., and Hightower, L. E. (1995) *J. Biol. Chem.* **270**, 19839–19844
- Rüdiger, S., Germeroth, L., Schneider-Mergener, J., and Bukau, B. (1997) *EMBO J.* **16**, 1501–1507
- Haas, I. G., and Wabl, M. (1983) *Nature* **306**, 387–389
- Knarr, G., Gething, M.-J., Modrow, S., and Buchner, J. (1995) *J. Biol. Chem.* **270**, 27589–27594
- Gething, M.-J., Blond-Elguindi, S., Buchner, J., Fourie, A. M., Knarr, G., Modrow, S., Nanu, I., Segal, M., and Sambrook, J. F. (1995) *Cold Spring Harbor Symp. Quant. Biol.* **60**, 417–428
- Bole, D. G., Hendershot, L. M., and Kearney, J. F. (1986) *J. Cell Biol.* **102**, 1558–1566
- Hendershot, L., Bole, D., Kohler, G., and Kearney, J. F. (1987) *J. Cell Biol.* **104**, 761–767
- Hendershot, L. M., Wei, Y., Gaut, J., Melnick, J., Aviel, S., and Argon, Y. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93**, 5269–5274
- Nakaki, T., Deans, R. J., and Lee, A. S. (1989) *Mol. Cell. Biol.* **9**, 2233–2238
- Dul, J. L., and Argon, Y. (1990) *Proc. Natl. Acad. Sci. U. S. A.* **88**, 8135–8139
- Ma, J., Kearney, J. F., and Hendershot, L. M. (1990) *Mol. Immunol.* **27**, 623–630
- Knittler, M. R., and Haas, I. G. (1992) *EMBO J.* **11**, 1573–1581
- Hellman, R., Vanhove, M., Lejeune, A., Stevens, F. J., and Hendershot, L. M. (1999) *J. Cell Biol.* **144**, 21–30
- Gething, M.-J., Blond-Elguindi, S., Mori, K., and Sambrook, J. F. (1994) in *The Biology of Heat Shock Proteins and Molecular Chaperones* (Morimoto, R. I., Tissieres, A., and Georgopoulos, C., eds) pp. 111–135, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- Haas, I. G. (1994) *Experientia* **50**, 1012–1020
- Brewer, J. W., and Hendershot, L. M. (1997) in *Molecular Chaperones in the Life Cycle of Proteins* (Fink, A. L., and Goto, Y., eds) pp. 415–434, Marcel Dekker, New York
- Machamer, C. E., Doms, R. W., Bole, D. G., Helenius, A., and Rose, J. K. (1990) *J. Biol. Chem.* **265**, 6879–6883
- Earl, P. L., Moss, B., and Doms, R. W. (1991) *J. Virol.* **65**, 2047–2055
- Dalgleish, A. G., Beverly, P. C. L., Clapham, P. R., Crawford, D. H., Greaves, M. F., and Weiss, R. A. (1984) *Nature* **312**, 763–767
- Otteken, A., Earl, P. L., and Moss, B. (1996) *J. Virol.* **70**, 3407–3415
- Hallenberger, S., Bosch, V., Anglikar, H., Shaw, E., Klenk, H. D., and Garten, W. (1992) *Nature* **360**, 358–361
- Haas, I. G., and Meo, T. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85**, 2250–2254
- Skerra, A., Pfützinger, J., and Plückthun, A. (1991) *Bio/Technology* **9**, 273–278
- Kunkel, T. A., Roberts, J. D., and Zakour, R. A. (1987) *Methods Enzymol.* **154**, 367–382
- Donch, J., and Greenberg, J. (1968) *J. Bacteriol.* **95**, 1555–1559
- Wei, J., Gaut, J. R., and Hendershot, L. M. (1995) *J. Biol. Chem.* **270**, 26677–26682
- Blond-Elguindi, S., Fourie, A. M., Sambrook, J. F., and Gething, M.-J. (1993) *J. Biol. Chem.* **268**, 12730–12735
- Kozutsumi, Y., Normington, K., Press, E., Slaughter, C., Sambrook, J., and Gething, M.-J. (1999) *J. Cell Sci. Suppl.* **11**, 115–137
- Atherton, E., Fox, H., Logan, C. J., Harkiss, D., Sheppard, R. C., and Williams, S. B. J. (1978) *Chem. Soc. Chem. Commun.* **13**, 537–539
- Ratner, L., Haseltine, W., Patarca, R., Livak, K., Starcich, B., Josephs, S., Doran, R. E., Rafalski, J. A., Whitehorn, E. A., Baumeister, K., Ivanoff, L., Petteway, S. R., Lautenberger, I. A., Papas, T. S., Ghayeb, J., Chang, J., Gallo, R. C., and Wong-Staal, F. (1985) *Nature* **313**, 277–284
- Modrow, S., Hahn, B. H., Shaw, G. M., Gallo, R. C., Wong-Staal, F., and Wolf, H. (1987) *J. Virol.* **61**, 570–578
- Kwong, P. D., Wyatt, R., Robinson, J., Sweet, R. W., Sodroski, J., and Hendrickson, W. A. (1998) *Nature* **393**, 648–659
- Leonard, C. K., Spellman, M. W., Riddle, L., Harris, R. J., Thomas, J. N., and Gregory, T. J. (1990) *J. Biol. Chem.* **265**, 10373–10382
- Knittler, M. R., Dirks, M. R., and Haas, I. G. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92**, 1764–1768
- Chan, D. C., Fass, D., Berger, J. M., and Kim, P. S. (1997) *Cell* **89**, 263–273
- Weissenhorn, W., Dessen, A., Harrison, S. C., Skehel, J. J., and Wiley, D. C. (1997) *Nature* **387**, 426–430
- Moore, J. P., Cao, Y., Qing, L., Sattentau, Q. J., Pyati, J., Koduri, R., Robinson, J., Barbas, C. F., III, Burton, D. R., and Ho, D. D. (1995) *J. Virol.* **69**, 101–109
- Hart, T. K., Kirsh, R., Ellens, H., Sweet, R. W., Lambert, D. M., Petteway, S. R., Jr., Learly, J., and Bugelski, P. J. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88**, 2189–2193
- Fourie, A. M., Hupp, T. R., Lane, D. P., Sang, B.-C., Barbosa, M. S., Sambrook, J. F., and Gething, M.-J. (1997) *J. Biol. Chem.* **272**, 19471–19479
- Gnann, J. W., Jr., Nelson, J. A., and Oldstone, M. B. A. (1987) *J. Virol.* **61**, 2639–2641
- Dedera, D., Gu, R., and Ratner, L. (1992) *J. Virol.* **66**, 1207–1209
- Wilson, I. A., Skehel, J. J., and Wiley, D. C. (1981) *Nature* **289**, 366–373
- Skehel, J. J., Bayley, P. M., Brown, E. B., Martin, S. R., Waterfield, M. D., White, J. M., Wilson, I. A., and Wiley, D. C. (1982) *Proc. Natl. Acad. Sci. U. S. A.* **79**, 968–974
- Bullough, P. A., Hughson, F. M., Skehel, J. J., and Wiley, D. C. (1994) *Nature* **371**, 37–43
- Carr, C. M., and Kim, P. S. (1993) *Cell* **73**, 823–832
- Zhu, X., Zhao, X., Burkholder, W. F., Gragerov, A., Ogata, C. M., Gottesman, M., and Hendrickson, W. A. (1996) *Science* **272**, 1606–1614