Factors governing the substrate recognition by GroEL chaperone: a sequence correlation approach

Tapan K. Chaudhuri and Prateek Gupta

Department of Biochemical Engineering and Biotechnology, Indian Institute of Technology Delhi, Hauz Khas, New Delhi-110016, India

Abstract The chaperonin GroEL binds to a large number of polypeptides, prevents their self-association, and mediates appropriate folding in a GroES and adenosine triphosphate-dependent manner. But how the GroEL molecule actually recognizes the polypeptide and what are the exact GroEL recognition sites in the substrates are still poorly understood. We have examined more than 50 *in vivo* substrates as well as well-characterized *in vitro* substrates, for their binding characteristics with GroEL. While addressing the issue, we have been driven by the basic concept that GroES, being the cochaperonin of GroEL, is the best-suited substrate for GroEL, as well as by the fact that polypeptide substrate and GroES occupy the same binding sites on the GroEL apical domain. GroES interacts with GroEL through selective hydrophobic residues present on its mobile loop region, and we have considered the group of residues on the GroES mobile loop as the key element in choosing a substrate for GroEL. Considering the hydrophobic region on the GroES mobile loop as the standard, we have attempted to identify the homologous region on the peptide sequences in the proteins of our interest. Polypeptides have been judged as potential GroEL substrates on the basis of the presence of the GroES mobile loop-like hydrophobic segments in their amino acid sequences. We have observed 1 or more GroES mobile loop-like hydrophobic patches in the peptide sequence of some of the proteins of our interest, and the hydrophathy index of most of these patches also seems to be approximately close to that of the standard. It has been proposed that the presence of hydrophobic patches having substantial degree of hydrophathy index as compared with the standard segment is a necessary condition for a peptide sequence to be recognized by GroEL molecules. We also observed that the overall hydrophobicity is also close to 30% in these substrates, although this is not the sufficient criterion for a polypeptide to be assigned as a substrate for GroEL. We found that the binding of aconitase, α -lactalbumin, and murine dihydrofolate reductase to GroEL falls in line with our present model and have also predicted the exact regions of their binding to GroEL. On the basis of our GroEL substrate prediction, we have presented a model for the binding of apo form of some proteins to GroEL and the eventual formation of the holo form. Our observation also reveals that in most of the cases, the GroES mobile loop-like hydrophobic patch is present in the unstructured region of the protein molecule, specifically in the loop or β -sheeted region. The outcome of our study would be an essential feature in identifying a potential substrate for GroEL on the basis of the presence of 1 or more GroES mobile loop-like hydrophobic segments in the amino acid sequence of those polypeptides and their location in three-dimensional space.

INTRODUCTION

Protein-folding machinery in *Escherichia coli* includes chaperonin GroEL and its cochaperonin GroES, which facilitates folding of a variety of proteins by binding nonnative conformations in the hydrophobic cavity of an open GroEL ring and then triggering productive folding

on subsequent binding of GroES and adenosine triphosphate (ATP) to the same ring as polypeptide (Horvitz 1998; Sigler et al 1998) or to the opposite ring (Chaudhuri et al 2001). A polypeptide substrate forms its association with GroEL through multivalent hydrophobic contacts inside the central cavity of a ring. Apparently GroEL-bound proteins exist in collapsed, loosely packed conformations, with varying degrees of native secondary structure (Fenton and Horwich 2003). Hydrophobic residues located on the apical domain of the central cavity of GroEL recognize the exposed hydrophobic regions

Received 18 June 2004; Revised 26 August 2004; Accepted 26 August 2004.

Correspondence to: Tapan K. Chaudhuri, Tel: 91 11-2659-1012; Fax: 91 11-2658-2282; E-mail: tapan@dbec.iitd.ac.in.

of a nonnative protein through hydrophobic interactions and thus form a complex that at least results in the prevention of aggregation of the bound polypeptide, and very often the protected polypeptide reaches the folded state using the function of cochaperonin GroES and ATP (Fenton et al 1994). The role of hydrophobicity in substrate recognition by GroEL has been experimentally verified using rhodanese as a substrate. The aggregation of rhodanese could be prevented by GroEL as well as by the detergent lauryl maltoside, suggesting that the action of GroEL involved stabilizing hydrophobic surfaces of nonnative intermediates (Mendoza et al 1991). The interaction between GroEL and polypeptide has been thoroughly studied, and a recent study suggests that 3 consecutive wild-type apical domain-binding surfaces were required to support a full extent of binding of Rubisco or malate dehydrogenase (MDH) in vitro (Farr et al 2000). Information about the change in microenvironment of bound substrate protein has been probed by the change of accessible surface area of amino acid residues present in the peptide-binding region of the GroEL molecule (Stan et al 2003). Using the multiple sequence alignment and chemical sequence entropy calculations, the authors have demonstrated that, only the chemical identities of the amino acid residues in the GroEL central cavity is conserved for its function. Although the nature of interactions between the polypeptide and GroES with GroEL is well known (Fenton and Horwich 2003), not much is understood about the nature of hydrophobic regions in the polypeptide substrate that renders its identity as a GroEL substrate. Attempts were taken to address this issue using different short peptide segments or synthetic polypeptides, which may or may not assume the same conformation that would be present in the nonnative form of the protein. This possibility can partly be explained from the fact that residues 199–209 in the GroEL apical domain, containing additional hydrophobic residues essential for binding substrates such as Rubisco and MDH, has not been involved in the binding of small peptides (Fenton and Horwich 2003). Hence, we decided to consider the whole protein as the substrate. Moreover, we considered a substantial number of in vivo and in vitro substrates for GroEL, which are known to bind with GroEL in cell or in solution. Hence, a correlation derived from a large number of known protein substrates should reliably demonstrate the factors necessary for GroEL recognition by polypeptide substrates. We have addressed the issue of understanding the basis of substrate recognition by GroEL using the fundamental aspect in GroEL-GroES interaction. Although some events are common in the binding of GroES and polypeptide to GroEL, eg, both of them compete for the same binding sites on the apical domain of GroEL (Fenton and Horwich 1997), there are differences

as well. It is observed that binding of GroES to GroEL enhances the intraring cooperativity and diminishes the ATP hydrolysis substantially, whereas bound polypeptide has little effect on ATP turnover. The reason for this behavior might be that binding of GroES to GroEL is stronger than that of polypeptide to GroEL because all the subunits of GroEL become occupied on GroES binding, whereas for many substrates all the subunits of GroEL might not be engaged by polypeptide binding. As a result, the intraring conformational alteration is not as pronounced as that of GroES binding, and hence we can expect a milder effect on ATP turnover on polypeptide binding. We have considered that the sequence of the GroES mobile loop GGIVLTG, which binds with the GroEL apical domain, must be having the characteristic properties necessary for stable and energetically favorable binding between GroEL and polypeptide. There might be a debate whether the strong hydrophobic peptide segment in the GroES mobile loop (GGIVLTG) is the perfect GroEL-binding motif because of the finding that native GroES never aggregates in solution despite it having 7 such hydrophobic segments exposed to solution. However, nuclear magnetic resonance data (Fiaux et al 2002) and crystal structures (Xu et al 1997) are consistent with the mobile loop patch as the major interaction site. Less obvious contacts may provide some additional binding energy, particularly because they are all multiplied by 7, but there is no evidence for any other interaction with energy equivalent to that available from the mobile loop. Although a single mobile loop consisting of the hydrophobic segment GGIVLTG will bind with 1 of the 7 apical domains present in the heptameric GroEL, there is no experimental information available about its stability. However, when 7 such mobile loops, containing 7 hydrophobic segments bind with native GroEL molecules, the binding is strong as evidenced by stable complex formation between GroEL and GroES in the presence of nucleotides (Fenton and Horwich 2003). Hence, a single GroES mobile loop may not be a perfect substrate for GroEL, but GroES as a whole is a perfect binding partner as well as a natural cochaperonin for GroEL. This occurs not only because it has 7 mobile loops of hydrophobic segments GGIVLTG in heptameric GroES but also because it possesses a 7-fold symmetry axis, which causes a perfect fit in the opening of GroEL. An obvious issue regarding the strength of the complex formed between polypeptide substrates containing mobile loop sequences and GroEL would depend on the number of such sequences present in the former. Polypeptide sequences containing multiple hydrophobic patches are more likely to be the natural substrates because each substrate protein can bind up to 7 sites on the GroEL apical domain. The secondary criteria, ie, the overall hydrophobicity of a fully denatured polypeptide

chain and hydropathy index of localized hydrophobic regions in a polypeptide chain, have also been considered in identifying the basis of substrate identification by GroEL. To identify the molecular basis of substrate recognition by GroEL, we have considered more than 50 in vivo substrates as well as quite a few in vitro substrates of GroEL, and using the structural correlation approach, we have demonstrated the criteria for the substrate recognition of GroEL. The overall hydrophobicity profile as well as local hydropathy index have been taken into account in this correlation process. We have considered the local hydropathy index of the hydrophobic patch in the GroES mobile loop, ie, the sequence of the GroES mobile loop GGIVLTG, which is responsible for its interaction with GroEL as the standard and then compared the hydropathy index of the isolated hydrophobic patches in the amino acid sequence of known proteins. On the basis of our correlation study, we have proposed a basis for the substrate recognition by GroEL. Some of the in vitro substrate-binding phenomenons, like binding of α -lactalbumin, dihydrofolate reductase (DHFR), and aconitase with GroEL have been well explained using our proposition. We have also proposed a mechanism of apo protein binding with GroEL using our hypothesis. To proof the validity of our proposition, we considered the amino acid sequence of few proteins, which do not bind with GroEL, and the sequence correlation approach was applied on these sequences to show how the results differ from those of the protein sequences, which bind with GroEL in vivo or in vitro. In this article, we have explained the tentative position of hydrophobic patches in the 3-dimensional structure of known GroEL substrates. The success of our study is few predictions and proposals that can be implemented through specific experiments such as mutational studies at the predicted regions of the protein sequence. Once, the experimental findings support the prediction, we will be in a situation to introduce necessary mutations in the sequence of a protein to force its binding with GroEL and then try to fold the protein using the complete chaperone system.

MATERIALS AND METHODS

Amino acid sequences of in vivo substrates of the chaperonin GroEL (Houry et al 1999) as well as in vitro substrates were collected from the Swiss Prot Databank. Thereafter, each sequence was independently aligned with the heptapeptide sequence (GGIVLTG), present in GroES mobile loop and responsible for its binding with GroEL, to find the percentage of sequence identity. Percent similarity was calculated using the pairwise sequence alignment tool at BCM Search Launcher (Smith et al 1981), which actually provides 'k' best nonintersecting

alignments. Amino acid substitutions of a similar kind (hydrophobic-hydrophobic; identical charged-charged) were allowed. The alignments having significant match with the heptapeptide mobile loop sequence GGIVLTG (GroEL mobile loop region) were then selected for further processing. We have used the mobile loop sequence GGIVLTG to find out the homologous segment in the polypeptide sequences in identifying the potential GroEL substrates. The reason for such selection is that we wanted to identify polypeptides as potential substrates for GroEL on the basis of the presence of 1 or more mobile loop-like sequences. Glycine residues before the IVL sequence in the GroES mobile loop is important for its association with GroEL. However, the glycines do not make direct contact with the GroEL apical domain but assist the IVL residues to be placed in correct conformation so that they can interact with residues on the GroEL apical domain. For the polypeptide substrates in their nonnative form, the flanking glycine residues may or may not be necessary for binding with GroEL. Hence, by searching the mobile loop sequence GGIVLTG in the polypeptide substrates, we may miss some genuine GroEL substrates lacking the flanking glycine residues. This is a limitation of the 1-dimensional approach. The other limitation of this approach is that it cannot identify strong hydrophobic patches formed by the residues close in the 3-dimensional space. These limitations collectively generate some false-negative results. Hence, the sequence correlation approach can predict few potential GroEL substrates, if not all. The overall percentage of hydrophobicity was calculated as the ratio x/n , where x is the number of most hydrophobic amino acids (L, V, I, F, M) and n is the total number of amino acids in the sequence. The calculation was performed using the Statistical Analysis of Protein Sequences (SAPS) software (Brendel et al 1992). The local hydropathy index of the individual short peptides, which shared some sequence similarity with the mobile loop, was calculated using the grand average of hydropathicity (GRAVY) (Kyte and Doolittle 1982). The GRAVY value for a peptide or protein is calculated as the sum of hydropathy values of all the amino acids divided by the number of residues in the sequence.

Our prediction of binding of nonnative α -lactalbumin to GroEL was validated using the experimental data (Hayer-Hartl et al 1994; Okazaki et al 1994, 1997; Katsumata et al 1996; Shimizu et al 1998). We confirmed our prediction on the binding of nonnative and apo form of mitochondrial aconitase using the experimental findings of Chaudhuri et al (2001), and the results from Clark et al (1996) confirmed our prediction on the binding of murine DHFR with GroEL. We have carried out a statistical analysis on the frequency of occurrences of GroES mobile loop-like sequence in the *E coli* substrate proteins using the database of *E coli* proteins.

RESULTS

Determination of overall hydrophobicity for GroEL substrates

In this article we have classified GroEL substrates into 3 broad categories, namely in vivo substrates, in vitro substrates, and nonsubstrates. The in vivo substrates are the proteins that have been shown to form complexes with GroEL in the *E coli* cytosol (Table 1), the in vitro substrates are the proteins that have been demonstrated to form binary complexes with GroEL under in vitro conditions (Table 2), and the nonsubstrates are the proteins that do not form complexes with GroEL under in vivo and in vitro conditions (Table 7). The overall hydrophobicity of GroELs in vivo, in vitro, and nonsubstrates were calculated and presented in Table 4. The results show that on average 27% of the amino acid residues in a protein are hydrophobic in nature, and we could not find any remarkable difference in overall hydrophobicity between in vivo and in vitro substrates of GroEL as well as between substrates and nonsubstrates. Hence, overall hydrophobicity is not the primary determinant for the binding of a nonnative protein to GroEL; however, the surface hydrophobicity of a nonnative protein might differentiate a nonsubstrate from the substrate. In other words the surface hydrophobicity of a partially folded protein might dictate whether the exposed hydrophobicity is sufficient for a stable binding between GroEL and nonnative protein through hydrophobic interactions.

Identification of hydrophobic patches for GroEL substrates

We have detected the hydrophobic patches consisting of strong hydrophobic residues I, V, M, L, F in the primary sequence of known in vivo and in vitro GroEL substrates (Tables 1–3). We calculated the hydropathy index of the peptide segment (GGIVLTG) in the mobile loop region of GroES (Table 4) with a view that the best known substrate for GroEL is definitely GroES and also because both the substrate and mobile loop on GroES compete for the same binding position at the apical domain of GroEL (Fenton and Horwich 1997). We then calculated the hydropathy indexes of detected hydrophobic patches, the number of patches, and percentage of similarity between the calculated hydrophobic patch of substrate and nonsubstrate proteins and the GroES mobile loop region (Tables 4–7). What we have observed is that in most of the cases, there is a 40% to 50% similarity in the pattern of hydrophobic patches between GroEL substrates and the GroES mobile loop (GGIVLTG). Although the peptide sequence is not conserved among the in vivo substrates, the patches are mostly formed by the residues that are present in the hydrophobic patch on the GroES mobile loop.

Hence, it is quite obvious that strong hydrophobic patches in the substrate proteins formed by amino acid residues having high values of hydrophobicity play an important role in binding with GroEL. In some cases, we have found that the protein contains more than 1 strong hydrophobic patch (Table 1), and it may happen that the number of patches might form a strong hydrophobic region, which enables the substrate in its nonnative form to be recognizable by GroEL. It may be anticipated that the GroEL substrates in their nonnative form keep the hydrophobic patches partially or completely exposed so that they can bind with GroEL, which otherwise, in the absence of GroEL, form insoluble aggregates also known as inclusion bodies.

Frequency of occurrences of GroES mobile loop-type sequence in *E coli* proteins

To find out the possibility of appearance of GroES mobile loop-like hydrophobic patches in *E coli* proteins, we carried out a search on the database of *E coli* proteins and found out that about 60% of them contain at least 1 hydrophobic patch, and one-third of the total proteins contain multiple hydrophobic patches similar to the GroES mobile loop. Although there is a major difference in number between the predicted and observed in vivo substrates of GroEL, on the basis of previous observation (Vitanen et al 1992) it may be explained that many proteins are capable of interacting with GroEL transiently. The 52 reported in vivo substrates (Houry et al 1999) definitely form stable binary complexes with GroEL so that they could be isolated.

Case studies

I. Why do different forms of α -lactalbumin bind with GroEL?

An independent approach to analyze what kind of protein structures GroEL recognizes is to test the ability of preformed metastable folding intermediates to bind with GroEL. Perhaps the most revealing example is binding of the partially folded intermediate of α -lactalbumin. α -Lactalbumin is a small 2-domain protein, having 4 disulfide bonds, that adopts a partially folded structure, often referred to as the molten globule structure (Kuwajima 1989). The α -domain comprises residues 1–37 and 84–123 and contains 4 α -helices A, B, C, and D (Chaudhuri et al 2000). The β -domain encompasses 1 peptide region 38–83 and is composed of a 3-stranded β -sheet and a 3_{10} helix. The molten globule state of α -lactalbumin has a native-like secondary structure and has a compact shape (Arai and Kuwajima 1996). However, it does not have a specific tertiary structure, and the α -domain is more or-

Table 1 Predicted hydrophobic patch in the in vivo substrates of GroEL

Swiss Prot ID's	Predicted hydrophobic patch
GroES	TKSAGGIVLTGSAAAKSTRG
AROHECOLI	RIEKILNGEDKRLLVIIIGPC
BIOF.ECOLI	VVTFGKGFGV SGAAVLCSS T
CLPX.ECOLI	K SNILLIG PTGSGKTLLAET
DADA.ECOLI	MRV VILGSG V VGVASAWYLN
DCEA.ECOLI	TLGGEATD IVVM RMCRRGF
DLDH.ECOLI	PQRYD AVLVA IGRVPNGKNL
EFTU.ECOLI	MITGAAQMD GAILV AATDG
ENO.ECOLI	NKSKFGAN AI LAV SLANAKA
FTNA.ECOLI	CSYHTF EGAAFL RRHAQEE
G3P1.ECOLI	GVDVVAEA TGLFLTD ETARK
GATD.ECOLI	MKSVVNDT DG IVR VAESVIP, NKNVIIIAG TIGLLAI QCA
GATY.ECOLI	AANLH APVIIAG TPGTFTHA
GATZ.ECOLI	KVGFA RERIILGG DHLGPNC
GLF.ECOLI	YFSDRY QGIPVGGY TKLIEK
GRPE.ECOLI	DVAP GNVLGIM QGYTLNGR
ILVLECOLI	LARATGEV GV VLV TSGPGAT
LEU1.ECOLI	SVHTHDDL GLAVGN SLAAVHA
LIPA.ECOLI	K SGLMVG LGETNEEIIIVMR
METE.ECOLI	QERLGLD VLVH GEAERNDMV
METK.ECOLI	ACETYVKT GMVLVGG EITTS
MIND.ECOLI	DSDRIL GILASK SRRAENGE
NAGD.ECOLI	MTIKNVICDI DGVL MHDNVA
NUSA.ECOLI	EGLTDEK AGALIMA ARNICW
NUSG.ECOLI	NMEDLF GEVMVP TEEVVEIR
PSTB.ECOLI	ELYPEQRA EGEILLD GDNIL
PTA.ECOLI	AKNTNIT GVI VNK LNAPVDE, LAAMNGVEIG ALLLTGGY EM
PUR7.ECOLI	STEN PDLLVLE FRNDTSAGD
PYRB.ECOLI	G NGIFAR QALLAL VLNRLVL
RCSB.ECOLI	LDIE GIVLKQ GAPTDLPKAL
RECA.ECOLI	GINFY GELVDLG VKEKLIKAGA
RL7.ECOLI	KTE FDVILKA AGANKVAVIK
RL9.ECOLI	VSFQVHSEV FAKVIVN VVAE
RPOA.ECOLI	SEEDER PIGR LLVD ACYSPV
RPOZ.ECOLI	KIGNR FDLVLVA ARRARQMQ
RS2.ECOLI	LNKIASR KGKILFVG TKRAA
SSPB.ECOLI	EYARD GQIVLNI APRAVGNLE
SYFA.ECOLI	THTPMFHQ MEGLIVD TNISF
SYT.ECOLI	YSTFG EKIVVKL STRPEKR
TDH.ECOLI	QKGFDA MRSGQSGK VILSWD
THID.ECOLI	QGRSLLAM GCGAVLMK GGHL
THIG.ECOLI	LEPLIA AGVTLLP NTSGAKT
THIH.ECOLI	REMGFE HLLLV TGEHQAKVG
UPP.ECOLI	GCSSIK VVLVLA APEGIAAL
YIGB.ECOLI	LAKKW PLVAITNG NAQ PELF
YJJU.ECOLI	SLRPFQ PGRIALV CEGGGQRGIF

Hydrophobic patch has been shown by bold letters.

Table 2 Predicted hydrophobic patch in some of the in vitro stringent substrates of GroEL

Protein name	Predicted hydrophobic patch
Rubisco	PPEE AGAAVAA ESSTGTWTT
Human DHFR	LEKYKLL PEYPGVL SDVQEEK
Rhodanese	VHQVLY RALVSTK WLAESIR
α -Lactalbumin	DIDGY GGIALPE ICTMFHT, IMCAKK ILDIK GIDYWLAKH
Aconitase	LEPRHL GGRAIITK SFARIH, NLKK QGLLPLT FADPADYNK
Myoglobin	DVAGH GQEV LIRLFTGHPET, NTVLT ALGGILK KKKGHEAE

Hydrophobic patch has been shown by bold letters.
DHFR, dihydrofolate reductase.

Table 3 Predicted hydrophobic patch in some of the in vitro nonstringent substrates of GroEL

Protein name	Predicted hydrophobic patch
Thioredoxin	IMIFKNGQVVDMNLNGAVPKA
Beta-lactamase precursor	IPEKELGIVMLANKNYPNPAR
Citrate synthase, Mitochondrial precursor	DAHSGVLLQYYGMTEMNYYTVLFGVSRALGVLAQLIWSR
Beta tubulin— <i>Cryptosporidium parvum</i>	TGSGMGTLVLVGKIHEEFPDR
Taka-amylase A— <i>Aspergillus flavus</i>	GVYCIGEVLDGDPAYTCPYQ
Branched-chain alpha-keto acid dehydrogenase E2— <i>Staphylococcus aureus</i>	SWQGNIEVLHKDINISIAVESIVKKPVVINDMIAIRNMV

Hydrophobic patch has been shown by bold letters.

Table 4 In vivo substrates of GroEL and their hydrophobic parameters

Swiss Prot Ids	% Overall hydrophobicity	Hydropathy index of the predicted patch	% Similarity between predicted hydrophobic patch and GroEL-binding residues on GroES	Structural regions of the predicted hydrophobic patch
AROG.ECOLI	27.7	1.25	50	Loop
AROH.ECOLI	27	1.771	45	Not available
BIOF.ECOLI	27.1	1.256	41.7	Loop
CLPX.ECOLI	28.6	1.288	60	Not available
DADA.ECOLI	26.9	1.389	50	Not available
DCEA.ECOLI	28.1	1.7	50	Not available
DLDH.ECOLI	29.8	2.400	63	Not available
EFTU.ECOLI	30	2.022	37.5	β Strand
ENO.ECOLI	26.5	1.700	38.5	Not available
FTNA.ECOLI	27.9	1.157	44	α Helix
GroES	34	1.514	100	Loop
G3P1.ECOLI	27.9	0.378	57.1	Loop
GATD.ECOLI	31.2	1.811	35	Not available
GATY.ECOLI	28.2	1.1, 2.167	50	Not available
GATZ.ECOLI	28.6	0.571	35	Not available
GLF.ECOLI	28.3	0.139, 0.578	50	Not available
GRPE.ECOLI	29.4	1.443	50	β Strand
ILVLECOLI	30.7	1.379	57.1	Not available
LEU1.ECOLI	28.4	1.329	30	Not available
LIPA.ECOLI	28	1.462	37.5	Not available
METE.ECOLI	27.7	0.37	28.6	Not available
METK.ECOLI	27.4	1.725	60	β Strand
MIND.ECOLI	31.6	1.257	33.5	Not available
NAGD.ECOLI	31.2	0.589	30	Not available
NUSA.ECOLI	29.3	0.86	22	Not available
NUSG.ECOLI	32	1.167	40	Not available
PSTB.ECOLI	29.6	0.783	40	Not available
PTA.ECOLI	29.2	0.629, 1.671	56	Not available
PUR7.ECOLI	31.6	1.000	52	Not available
PYRB.ECOLI	30.6	1.278	40	Loop
RCSB.ECOLI	34.7	0.289	62.5	Not available
RECA.ECOLI	28.1	0.571	42.9	Not available
RL7.ECOLI	29.2	1.233	40	β Strand
RL9.ECOLI	30.9	0.578	33.3	Not available
RPOA.ECOLI	30.7	0.82	40	β Strand
RPOZ.ECOLI	28.6	2.443	33	Not available
RS2.ECOLI	29.6	1.238	37.5	Not available
SSPB.ECOLI	27.3	0.85	67	Not available
SYFA.ECOLI	29.7	0.311	15	Not available
SYT.ECOLI	27.9	0.771	40	β Strand
TDH.ECOLI	32.6	0.943	30	Not available
THID.ECOLI	27.4	1.188	45.5	Not available
THIG.ECOLI	30.9	1.357	60	Not available
THIH.ECOLI	25.8	1.614	40	Not available
UPP.ECOLI	33.7	2.586	54	Not available
YIGB.ECOLI	28.6	0.78	25	Not available
YJJU.ECOLI	29.4	1.700	62	Not available

Table 5 In vitro substrates of GroEL and their hydrophobic parameters

Protein name	% overall hydrophobicity	Hydropathy index of the predicted patch	% Similarity between predicted hydrophobic patch and GroEL-binding residues on GroES	Structural regions of the predicted hydrophobic patch
Rubisco	26	1.071	57	Loop
Human DHFR	30.6	0.843	40	Loop
Rhodanese	26.2	1.400	45	α
α -lactalbumin	23.2	0.928	64	Loop
Aconitase	25.8	1.663	67	β
Myoglobin	27.5	1.314	50	Loop-helix

Table 6 In vitro nonstringent substrates of GroEL and their hydrophobic parameters

Protein name	% Overall hydrophobicity	Hydropathy index of the predicted patch	% Similarity between predicted hydrophobic patch and GroEL-binding residues on GroES	Structural regions of the predicted hydrophobic patch
Thioredoxin	29.5	0.86	25	Loop
Beta-lactamase precursor	25.7	2.013	40	Loop
Citrate synthase, mitochondrial precursor	29.2	1.913	35	Loop
Beta tubulin- <i>Cryptosporidium parvum</i>	29.6	1.363	44.4	Not available
Taka-amylase A— <i>Aspergillus flavus</i>	24.4	0.850	55.6	Not available
Branched-chain alpha-keto acid dehydrogenase E2— <i>Staphylococcus aureus</i>	28.3	0.634	41.7	Not available

Table 7 Proteins that do not bind with GroEL and their hydrophobic parameters

Protein name/Swiss Prot IDs	% Hydrophobicity (overall)	Presence of the predicted patch (local hydropathy index)
BPTI	17.2	No
Chloramphenicol acetyltransferase	27.6	No
Staphylococcal nuclease	29.1	No
Alcohol dehydrogenase	29.1	No
Ribosome recycling factor	27	No
Protein chain initiation factor IF-3	28.9	No
ZUR.ECOLI	27.5	No
ALLA.ECOLI	32.5	No
CYSP.ECOLI	27.5	No
FIMF.ECOLI	29	No
HCHA.ECOLI	27	No
KEA2.ECOLI	19.5	No

ganized than the β -domain (Arai and Kuwajima 2000). Initial work indicated that the relatively compact molten globule intermediate in the α -lactalbumin folding pathway was not bound by GroEL (Hayer-Hartl et al 1994; Okazaki et al 1994), but Katsumata et al (1996) clearly showed that it does bind albeit with a much reduced association constant.

The disulfide intact and 1 disulfide (Cys6-Cys120) reduced α -lactalbumin (RLA) were shown not to bind with GroEL both in the presence and absence of Ca^{2+} (Shimizu et al 1998). The 2 disulfide (Cys6-Cys120 and Cys28-Cys111) RLA, which has the native-like tertiary structure in its β -domain region and an unfolded α -domain in the presence of Ca^{2+} , showed considerable binding with GroEL (Shimizu et al 1998). The binding free energy of the 2 disulfide RLA in the presence of Ca^{2+} is close to

that of the molten globule state α -lactalbumin (Shimizu et al 1998). This result clearly suggests that GroEL binds to the unfolded α -domain regardless of the β -domain (Shimizu et al 1998). Although one can argue that the interaction between the unfolded β -domain and GroEL in the metastable intermediate states also contributes to the binding of α -lactalbumin with GroEL, it has been shown by free energy calculations that the majority of the interaction is contributed by the α -domain of the partially unfolded α -lactalbumin (Shimizu et al 1998). This conclusion is consistent with the observation that a chymotryptic fragment of α -lactalbumin (54–105), which contains only the β -domain, is unable to bind to GroEL, whereas a larger fragment (17–108) associates with GroEL (Shimizu et al 1998).

Studies done with hydrogen exchange kinetics of RLA

Fig 1. Bovine α -lactalbumin which clearly shows that the predicted hydrophobic regions (loop region marked by white colored arrow). It is predicted that these regions in the molten globule state form a hydrophobic pocket, which is recognized by GroEL. These are the two hydrophobic patches, residues 38–43 (GGIALP) and residues 113–120 (KILDIKGI), which have high hydropathy indices and share high degree of homology with the GroES mobile loop. (Figures are generated using software Cn3D available at NCBI.)

bound to the chaperonin GroEL (Okazaki et al 1997) suggested that the hydrogen bonds in the secondary structural elements in RLA are neither stabilized nor destabilized when RLA is bound to GroEL, although RLA has quite a large extent of secondary structure. Hence, the structural elements recognized in RLA by GroEL are not the secondary structure, but unstructured parts, which have no influence on the stability of hydrogen bonds in the secondary structure in the RLA molecule. It was also been suggested that the peptide fragment between residues 110 and 120, which is positively charged and shows high hydropathic indices, and also the region between residues 90 and 100, which is also hydrophobic and has neutral average charge, provide strong recognition sites of α -lactalbumin by GroEL (Okazaki et al 1997). Moreover, Shimizu et al (1998) clearly showed that the B14 and C17 peptides, which correspond to the B- and C-helix regions of α -lactalbumin, have binding constants with GroEL less than 10^4 M^{-1} .

According to our hypothesis, we propose that there exist 2 hydrophobic patches, residues 38–43 (GGIALP) and residues 113–120 (KILDIKGI), which have high hydropathy indices and share a high degree of homology with the GroES mobile loop (Table 2). They also form a flexible hydrophobic pocket in 3-dimensional space (Fig 1), which we propose to be the ideal location of GroEL recognition. These regions are the corresponding loop regions lying just before and after helix B and helix C, respectively, and

are largely unfolded in nature in its native state. The reason why the native α -lactalbumin does not bind with GroEL may be due to the fact that the hydrophobic pockets are not oriented properly to cause energetically favorable contacts with the GroEL apical domain. We also propose that any key mutation (most likely hydrophobic to hydrophilic) in this region would terminate the binding of the nonnative form of α -lactalbumin to GroEL.

II. Mechanism of binding aconitase with GroEL

Yeast mitochondrial aconitase is a monomeric, 82-kDa, Fe_4S_4 cluster-containing enzyme of the Krebs cycle that catalyzes the isomerization of citrate to isocitrate. Presumably only a portion of a large protein, like aconitase, could be bound within the GroEL central cavity, whereas the remainder must lie outside in the bulk solution (Chaudhuri et al 2001). As suggested earlier (Dubaque et al 1998), it may be that only 1 of the 4 domains in aconitase has kinetic difficulty in reaching its native form, and it is this domain (residue numbers 535–753) that becomes recruited to the central cavity of GroEL, whereas the other 3 domains lying in the bulk solution are essentially native and would not themselves recruit chaperones to their surfaces. We hereby report the presence of multiple hydrophobic patches in the aconitase sequence, very similar to the mobile loop of GroES (Table 2). We, thus hypothesize that the presence of these hydrophobic patches in aconitase are the sites of interaction with GroEL. This domain, termed as the fourth domain of aconitase throughout this article, seems to be in relatively unfolded conformation when nonnative aconitase approaches GroEL, and thus because of its relatively more hydrophobic patch content, seems to contain GroEL-binding motifs. It has also been reported that the $-\text{COOH}$ terminal domain generated by domain 4 of the native aconitase lies somewhat apart in the crystal structure from the bodies of the other 3, which house the Fe_4S_4 cluster (Lauble et al 1992). This information might give us a clue to understand the exact mechanism of how the holo form of aconitase is formed and how the transition from apo form of aconitase to holo form takes place. We hypothesize that the fourth domain interacts with GroEL, whereas the other 3 domains capable of folding spontaneously stay in solution thereby allowing them to form the metal-binding pocket of apo aconitase (Chaudhuri et al 2001). Because one would expect that for binding with the metal ion, the protein should be presented in more compact form with most of its structure already formed and with the charged residues close to each other in 3-dimensional space. The bound fourth domain is released from the GroEL cavity because of the conformational changes affected by the correct binding of the Fe_4S_4 cluster in the correctly organized metal-binding pocket of GroEL-bound apo aconitase.

tase, thereby resulting in the formation of holo aconitase in solution (Fig 2B). Hence, we can consider that GroEL acts as a nonspecific scaffold for the incorporation of the Fe-S cluster (Chaudhuri et al 2001).

III. Why does murine DHFR bind with GroEL?

DHFR from *E coli* does not interact with the molecular chaperonin GroEL, regardless of whether the interaction is initiated from the native or the unfolded state (Clark et al 1996). In contrast, murine DHFR shows a strong interaction with GroEL (Martin et al 1991). A superimposition of the 2 structures of DHFR reveals that there are 3 distinct external loops in the eukaryotic DHFR that are not present in *E coli* protein (Clark et al 1996). The mutant *E coli* DHFR-i₇136 (*E coli* DHFR-i₇136 is a mutant of *E coli* DHFR in which residues 136–139 were replaced with the 7-amino acid sequence L-P-E-Y-G-V of murine DHFR) binds to GroEL with a stoichiometry of 4–5 mol of DHFR per mol of GroEL tetradecamer (Clark et al 1996; Clark and Frieden 1999), whereas murine DHFR binds to GroEL with a stoichiometry of 2 mol of DHFR per mol of GroEL tetradecamer. In our analysis the residues 157–167 and 70–76 span these regions (Fig 3). This analysis clearly suggests that sites for the GroEL recognition of murine DHFR exist in these extra loop regions only and that the local hydrophobicity including hydrophobic patches or clefts appears to be more important than residue-specific hydrophobicity as a determinant of polypeptide binding by GroEL.

DISCUSSION

Role of hydrophobic patches in the amino acid sequence of substrate proteins on binding with GroEL

We measured the overall hydrophobicity of in vivo and in vitro GroEL substrates as well as nonsubstrates for a comparison and observed that in all the cases almost one-third of the residues are hydrophobic in nature. So, overall hydrophobicity cannot be a primary parameter for differentiation of a nonsubstrate from a substrate of GroEL. However, when we tried to identify the GroES mobile loop-type hydrophobic patches in the in vivo substrates for GroEL as well as the nonsubstrates, we got interesting observations. Almost all the in vivo substrates contain 1 or more hydrophobic patches having a high degree of similarity with the GroES mobile loop (Table 1). So, total hydrophobicity of a protein cannot decide whether it will bind with GroEL, but the arrangement of hydrophobic residues in a protein molecule is important for the recognition by GroEL. On the other hand, when we calculated the local hydrophobicity of hydrophobic patches of GroELs in vivo substrates, we observed that the value is

Fig 2. (A) Fourth Domain of aconitase predicted to bind to GroEL through its multiple hydrophobic patches (loop region marked by white colored arrow) present in strands and loop regions of the fourth domain. The other three domains of aconitase protein remain outside the GroEL cavity in their near-native forms. (B) Complete aconitase with its 4Fe-4S cluster. It is hypothesized that the binding of the cluster to the first three domains of aconitase results in the overall conformational change in the aconitase structure yielding it to disassociate from the GroEL cavity.

Fig 3. (A) *E. coli* DHFR and (B) human DHFR. The extra loop in the human DHFR (loop region marked by white colored arrow) contains the predicted patch for binding with GroEL. (Figures are generated using software Cn3D available at NCBI.)

pretty closer (within 80%) to the hydrophobicity of the GroES's GroEL recognition site (GGIVLTG). However, there are some cases where we observed high values of percentage similarity of hydrophobic patches with a relatively lower local hydropathy index (Table 4), the reason being the drop in the local hydropathy index due to the presence of some charged residues in the hydrophobic patches. However, we suggest that the presence of few

charged residues in the hydrophobic patch region of a substrate protein does not impair its binding ability to the GroEL apical domain until it contains 3 or 4 consecutive strong hydrophobic residues like I, V, L, etc. It may be possible that the affinity of binding with GroEL might be lowered for the proteins having charged residues in their hydrophobic patch region. A recent study on the short peptide binding with GroEL reveals that when 2 13-residue α -helical peptides with identical composition were compared, the one with amphiphilic character, ie, with a continuous hydrophobic surface at one aspect, bound strongly than the one with interspersed hydrophobic and hydrophilic residues (Wang et al 1999). When considering the protein sequences that do not interact with GroEL, we noticed that there are no stronger hydrophobic patches (Table 7), which once again confirms the point that the presence of hydrophobic patches is the determining factor in a protein to be identified as a substrate for GroEL. Of course, the other important criterion that a protein should fulfill before it qualifies as a substrate for GroEL is that it must expose the hydrophobic patch in its nonnative form. Hence, there is a probability that if the patches are present in the flexible loop region of a protein, it has a higher possibility of binding with GroEL. In fact, we observed hydrophobic patches in the loop region of bovine α -lactalbumin (Fig 1), yeast mitochondrial aconitase (Fig 2), human DHFR (Fig 3), and in the domains of elongation factor (EF-Tu), adenosylmethionine synthetase, and threonyl-tRNA synthetase (Fig 4).

Role of local hydrophobicity in the binding of nonnative α -lactalbumin with GroEL

The molten globule form and 2 disulfide reduced forms of α -lactalbumin have been shown to bind with GroEL (Shimizu et al 1998). Furthermore, the structural study as well as the hydrogen exchange study with RLA suggest that the α -domain is largely unstructured (Okazaki et al 1997). We have identified 2 hydrophobic patches consisting of residues 38–43 (GGIALP) and residues 113–120 (KILDIKGI) located in the loop regions before and after helix B and helix C. These patches have a high degree of similarity with the GroEL-binding region in GroES's mobile loop (Table 2). It is therefore proposed that the reduced and molten globule form of α -lactalbumin interacts with GroELs through the flexible loop region containing the hydrophobic patches.

Binding of aconitase with GroEL

Yeast mitochondrial aconitase (for which there is an X-ray model) is far too large to be accommodated inside the open ring of GroEL. Thus, presumably only 1 domain, or a portion of a domain, can have access to the binding

Fig 4. (A) Domain of Elongation factor (EF-Tu), (B) Domain of Adenosylmethionine Synthetase, and (C) Domain of Threonyl-tRNA Synthetase. We predict the presence of the hydrophobic patches in the same domain, which actually is involved in the metal binding as well. The loops regions, marked by white colored arrows, are the predicted hydrophobic patches that bind to GroEL. The apo form binds with the GroEL and forms the correct metal-binding pocket. The binding of the metal thereby triggers an overall conformational change in the protein molecule, which results in its release from the GroEL cavity.

sites of GroEL. It is possible that this would be a domain or surface of nonnative aconitase that exhibits particular difficulty in achieving native form and thus may preferentially present a hydrophobic surface to the GroEL cavity. We have observed multiple hydrophobic patches in the loop region of the fourth domain (Fig 2) and not in the other 3 domains. Hence, our prediction also suggests that aconitase should potentially bind with GroEL with its flexible and relatively less structured fourth domain. Mutational studies at the hydrophobic patch regions of aconitase may resolve this issue.

Binding of DHFR with GroEL

The reason why murine DHFR binds with GroEL (Clark et al 1996), whereas *E coli* DHFR does not bind in its nonnative form can also be explained on the basis of its local hydrophobicity as well as the position of the hydrophobic patches in its peptide sequence. It is presumable that the GroEL recognition sites of murine DHFR are present in the regions containing hydrophobic patches. Interestingly, *E coli* DHFR does not contain the same kind of hydrophobic patches in any place of its amino acid sequence, which is at least a strong point to defend with its inability to bind with GroEL. However, this prediction should be validated by amino acid replacement in the loop region of murine DHFR. It is indicative that the replacement of strong hydrophobic amino acids with hydrophilic ones in the hydrophobic patch region of DHFR should reduce or impair the ability of its binding with GroEL. The other kind of study is with *E coli* DHFR, where replacement of hydrophilic amino acids with selective hydrophobic ones in the flexible loop region might be worth performing to verify the difference between *E coli* and murine DHFR regarding their binding with GroEL.

The role of hydrophobic patches in recognition and binding of nonnative proteins can be further validated with the fact that recombinant *E coli* DHFR having mutations exhibits higher affinity for GroEL than murine DHFR. It is possible that the exposition of hydrophobic patches in the *E coli* mutant is more than that of the murine one in the nonnative form, or it is possible that *E coli* mutant forms the stronger patch. Although the overall hydrophobicity of *E coli* DHFR and murine DHFR are pretty close (Table 5), the arrangement of the hydrophobic amino acids in a protein chain is a crucial point in deciding whether the protein would be binding with GroEL in its nonnative form.

Why the apo form of some proteins binds with GroEL?

We have observed that elongation factor (EF-Tu), adenosylmethionine synthase, and threonyl-tRNA synthetase

Fig 5. After the binding and subsequent release of the apo protein from GroEL cavity, assisted by co-chaperonin GroES and nucleotide ATP, we predict that there might be two different fates for the apo protein. 1) The apo form may rebound to the GroEL cavity by virtue of its exposed hydrophobic patches, which are present in the same domain to which the metal binds. This results in the formation of appropriate metal-binding pocket, metal or metal cluster bind to the pocket, resulting in the conformational changes to enable the holo protein to be released in solution. 2) The apo form, after releasing from the GroEL cavity may multimerize using its exposed hydrophobic patches, thereby creating the metal-binding pocket, which eventually leads to the formation of the holo protein on binding to its partner metal or metal clusters. (A) Denatured protein binding to the GroEL cavity through its exposed hydrophobic patches; (B) Folding of the protein into the apo form and release from the GroEL cavity driven by GroES and ATP; (C) Rebinding of the apo form to GroEL by virtue of its exposed hydrophobic patch; (D) Multimerization of the released monomers; (E) release of the holo form from the GroEL cavity on metal binding in the correctly formed pocket; (F) The multimerized protein molecule forms an ideal metal-binding site where the metal can bind to form the holo protein.

(Fig 4A–C) contain hydrophobic patches homologous to the GroES mobile loop in the same domain that interacts with GroEL as well as with the ions Mg^{2+} , Zn^{2+} , and K^+ , respectively.

Therefore, we hypothesize the possible biosynthetic pathways of such metalloproteins. First, the nonnative form of the substrates interacts with GroEL with their exposed GroES mobile loop-like hydrophobic patches and folds to the extent of the apo form with the involvement of GroES and ATP. The correct structure of the metal-binding pocket might not be constructed on the apo protein, and hence the apo form cannot reach the holo form instantaneously. Hence, the apo form binds again with GroEL to form the correct metal binding. Thereafter, binding of metal or metalocluster causes an overall conformational change to the protein molecule so that the whole complex gets ejected into the solution as a holo protein. The mechanism needs experimental verification to establish whether the conversion of GroEL-bound apo form to holo form and its release requires the assistance of GroEL and ATP. Second, it might so happen that some of the apo proteins on release into bulk solution might multimerize before a metal binds to them. In those cases, the released apo form of the protein interacts with other monomers through its exposed hydrophobic patches and thus forms the appropriate metal-binding pocket. We have proposed a model of biosynthesis of metalloproteins based on our hypothesis (Fig 5).

To validate the proposed model, we have discussed the case of yeast mitochondrial aconitase. It has been report-

ed that apo aconitase stays bound with GroEL and is released from GroEL cavity after receiving Fe-S cluster (Chaudhuri et al 2001). The question that remains to be answered is—why does the apo form of aconitase bind with GroEL? The reason is that the released apo form does not contain correct metal-binding pockets and hence is unable to form the holo form instantaneously. Instead the apo form binds with GroEL, and on binding it can form the correct cluster-binding pocket so that the pre-formed Fe-S cluster can occupy the pocket, and the bound domain acquires the desired conformation to be released from the central cavity of GroEL (Chaudhuri et al 2001). The other possibility is that the monomeric apo molecule released from the GroEL central cavity after the first round of GroEL, GroES, and ATP-assisted folding might associate together to form a polymeric assembly, and having correct the metal-binding pocket and addition of metal or other metallic clusters then results in the formation of correctly folded metalloproteins. At least the second possibility does not hold good for aconitase folding because it has been observed that apo aconitase further associates with GroEL (Chaudhuri et al 2001).

CONCLUSION

The presence of single or multiple GroES mobile loop-like hydrophobic patches in the amino acid sequence seems to be a foremost criterion for a protein to be recognized by GroEL. However, the presence of the hydrophobic patch is not the only criterion for GroEL-substrate interaction, but

the hydrophobic region on the protein must also be exposed in its nonnative form so that it can interact with the peptide-binding region on the GroEL's apical domain. This is an intrinsic property of a protein, inscribed in its primary sequence. Hence, the presence of hydrophobic patches in the primary amino acid sequence of a protein and its availability for interaction with GroEL in the nonnative state are the necessary conditions to be fulfilled by a protein to be considered as a potential substrate for GroEL chaperones. Future experiments, like mutational studies in the hydrophobic regions of predicted GroEL substrates, would certainly improve our understanding on how different substrates interact with GroEL.

ACKNOWLEDGMENT

The authors acknowledge financial support from the Industrial Research and Development, Indian Institute of Technology, Delhi, for pursuing this work.

REFERENCES

- Arai M, Kuwajima K. 1996. Rapid formation of a molten globule intermediate in refolding of alpha-lactalbumin. *Fold Des* 1(4): 275–287.
- Arai M, Kuwajima K. 2000. Role of the molten globule state in protein folding [review]. *Adv Protein Chem* 53: 209–282.
- Brendel V, Bucher P, Nourbakhsh IR, Blaisdell BE, Karlin S. 1992. Methods and algorithms for statistical analysis of protein sequences. *Proc Natl Acad Sci USA* 89: 2002–2006.
- Chaudhuri TK, Arai M, Terada TP, Ikura T, Kuwajima K. 2000. Equilibrium and kinetic studies on folding of the authentic and recombinant forms of human alpha-lactalbumin by circular dichroism spectroscopy. *Biochemistry* 39(50): 15643–15651.
- Chaudhuri TK, Farr GW, Fenton WA, Rospert S, Horwich AL. 2001. GroEL/GroES-mediated folding of a protein too large to be encapsulated. *Cell* 107(2): 235–246.
- Clark AC, Frieden C. 1999. The chaperonin GroEL binds to late-folding non-native conformations present in native Escherichia coli and murine dihydrofolate reductases. *J Mol Biol* 285(4): 1777–1788.
- Clark AC, Hugo E, Frieden C. 1996. Determination of regions in the dihydrofolate reductase structure that interact with the molecular chaperonin GroEL. *Biochemistry* 35(18): 5893–5901.
- Dubaquie Y, Looser R, Funfschilling U, Jeno P, Rospert S. 1998. Identification of in vivo substrates of the yeast mitochondrial chaperonins reveals overlapping but non-identical requirement for hsp60 and hsp10. *EMBO J* 17(20): 5868–5876.
- Farr GW, Furtak K, Rowland MB, Ranson NA, Saibil HR, Kirchhausen T, Horwich AL. 2000. Multivalent binding of nonnative substrate proteins by the chaperonin GroEL. *Cell* 100: 561–573.
- Fenton WA, Horwich AL. 1997. GroEL-mediated protein folding [review]. *Protein Sci* 6(4): 743–760.
- Fenton WA, Horwich AL. 2003. *Q Rev Biophys* 36(2): 229–256.
- Fenton WA, Kashi Y, Furtak K, Horwich AL. 1994. Residues in chaperonin GroEL required for polypeptide binding and release. *Nature* 371(6498): 614–619.
- Fiaux J, Bertelsen EB, Horwich AL, Wuthrich K. 2002. NMR analysis of a 900K GroEL GroES complex. *Nature* 418(6894): 207–211.
- Hayer-Hartl MK, Ewbank JJ, Creighton TE, Hartl FU. 1994. Conformational specificity of the chaperonin GroEL for the compact folding intermediates of alpha-lactalbumin. *EMBO J* 13(13): 3192–3202.
- Horvitz A. 1998. Structural aspects of GroEL function. *Curr Opin Struct Biol* 8: 93–100.
- Houry WA, Frishman D, Eckerskorn C, Lottspeich F, Hartl FU. 1999. Identification of in vivo substrates of the chaperonin GroEL. *Nature* 402(6758): 147–154.
- Katsumata K, Okazaki A, Kuwajima K. 1996. Effect of GroEL on the re-folding kinetics of alpha-lactalbumin. *J Mol Biol* 258(5): 827–838.
- Kuwajima K. 1989. The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure. *Proteins* 6: 87–103.
- Kyte J, Doolittle RF. 1982. A simple method for displaying the hydropathic character of a protein. *J Mol Biol* 157: 105–132.
- Lauble H, Kennedy MC, Beinert H, Stout CD. 1992. Crystal structures of aconitase with isocitrate and nitroisocitrate bound. *Biochemistry* 31(10): 2735–2748.
- Martin J, Langer T, Boteva R, Schramel A, Horwich AL, Hartl FU. 1991. Chaperonin-mediated protein folding at the surface of GroEL through a 'molten globule'-like intermediate. *Nature* 352(6330): 36–42.
- Mendoza JA, Rogers E, Lorimer GH, Horowitz PM. 1991. Chaperonins facilitate the in vitro folding of monomeric mitochondrial rhodanese. *J Biol Chem* 266: 13044–13049.
- Okazaki A, Ikura T, Nikaido K, Kuwajima K. 1994. The chaperonin GroEL does not recognize apo-alpha-lactalbumin in the molten globule state. *Nat Struct Biol* 1(7): 439–446.
- Okazaki A, Katsumata K, Kuwajima K. 1997. Hydrogen-exchange kinetics of reduced alpha-lactalbumin bound to the chaperonin GroEL. *J Biochem (Tokyo)* 121(3): 534–541.
- Shimizu A, Tanba T, Ogata I, Ikeguchi M, Sugai S. 1998. The region of alpha-lactalbumin recognized by GroEL. *J Biochem (Tokyo)* 124(2): 319–325.
- Sigler PB, Xu Z, Rye HS, Burston SG, Fenton WA, Horwich AL. 1998. Structure and function in GroEL-mediated protein folding. *Annu Rev Biochem* 67: 581–608, 12537–12546.
- Smith TF, Waterman MS, Fitch WM. 1981. Comparative biosequence metrics. *J Mol Evol* 18(1): 38–46.
- Stan G, Thirumalai D, Lorimer GH, Brooks BR. 2003. *Biophys Chem* 100: 453–467.
- Viitanen PV, Gatenby AA, Lorimer GH. 1992. Purified chaperonin 60 (GroEL) interacts with the nonnative states of a multitude of Escherichia coli proteins. *Protein Sci* 1: 363–369.
- Wang Z, Feng H, Landry SJ, Maxwell J, Gierasch LM. 1999. Basis of substrate binding by the chaperonin GroEL. *Biochemistry* 38: 12537–12546.
- Xu Z, Horwich AL, Sigler PB. 1997. The crystal structure of the asymmetric GroEL-GroES-(ADP)₇ chaperonin complex. *Nature* 388(6644): 741–750.