PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Unliganded GroEL at 2.8 angstrom: Structure and Functional Implications

Paul B. Sigler and Arthur L. Horwich

Phil. Trans. R. Soc. Lond. B 1995 **348**, 113-119
doi: 10.1098/rstb.1995.0052

Unliganded GroEL at 2.8 Å: structure and functional implications

PAUL B. SIGLER[1] AND ARTHUR L. HORWICH[2]

Howard Hughes Medical Institute and the Departments of [1] Molecular Biophysics and Biochemistry and of [2] Genetics, Yale University, 295 Congress Avenue, New Haven, Connecticut 06510, U.S.A.

SUMMARY

The three-dimensional structure of the *E. coli* chaperonin, GroEL, has been determined crystallographically and refined to 2.7 Å in two crystal forms: an orthorhombic form from high salt and a monoclinic form from polyethylene glycol. The former is ligand free, the latter is both liganded with ATP analogues and ligand free. These structures provide a structural scaffold upon which to interpret extensive mutagenesis and biochemical studies.

GroEL contains two sevenfold rotationally symmetric rings of identical 547-amino acid subunits. The rings are arranged 'back-to-back' with exact dyad symmetry to form a stubby cylinder that is 146 Å high with an outer diameter of about 143 Å. The cylinder has a substantial central channel that is unobstructed for the entire length of the cylinder and has a diameter of about 45 Å except for large bulges that lead into a sevenfold symmetric array of elliptical side windows in each ring.

Each subunit is composed of three distinct domains: (i) an 'equatorial' domain that contains the N- and C-terminus and the ATP-binding pocket, (ii) an 'apical domain' that forms the opening of the central channel and contains poorly ordered segments that mutational studies implicate in binding unfolded polypeptides and GroES, and (iii) an intermediate domain tht connects the other two domains and may serve to transmit allosteric adjustments.

1. INTRODUCTION

Our research work addresses the molecular mechanism of chaperonin-assisted protein folding by GroEL. Recent work, some of it presented in this volume, describe various activity cycles that include: (i) the cycling of ATP/ADP and the heptameric co-chaperonin, GroES (Ellis *et al.* 1991; Gething *et al.* 1992; Hendrick *et al.* 1993; Horwich *et al.* 1993; Todd *et al.* 1994); this cycle is necessary for GroEL function but can occur independently in the absence of any 'substrate' polypeptide; and (ii) various scenarios by which GroEL binds and releases polypeptide and, thereby, prevents (or possibly reverses) the formation of essentially irreversible off pathway folding intermediates (Martin *et al.* 1994; Todd *et al.* 1994; Weissman *et al.* 1994). Overall, GroEL, in concert with GroES and ATP, increases the efficiency by which linearly encoded information is converted into functional protein. We discuss here the three-dimensional structure of unliganded GroEL determined by X-ray crystallography to 2.8 Å, a resolution that enabled us to define most of the stereochemistry in near atomic detail (Braig *et al.* 1994). This structure, and a parallel study of directed mutational changes (Fenton *et al.* 1994), as well as on-going work on various liganded states, begins a structure–function analysis that should enable us to: (i) define the binding of substrates and cofactors in chemical terms; (ii) describe the molecular mechanisms underlying the dramatic conformational differences observed in various liganded states; and (iii) provide a rational basis for the design and interpretation of future experiments examining dynamics of the reaction cycle. In short, we wish to create a sound framework for a truly biochemical understanding of chaperonin function.

2. THE OVERALL ARCHITECTURE

Figure 1 shows GroEL to be a nearly sevenfold rotationally symmetrical cylinder composed of two rings of seven subunits each. It is a bit taller (146 Å) than it is wide (137 Å). The rings are stacked 'back-to-back', related by dyad symmetry. There is a central channel in the cylinder that is about 45 Å wide both at its openings and at the level of the equatorial plane, that passes through the interface between the rings. This channel widens considerably at the mid-latitude of each ring, through bulges into each subunit that connect the central channel to the outside through seven 'side windows'.

The subunits form three domains (see figures 1 and 2). First, a well-ordered, highly α-helical, 'equatorial domain' of 243 residues (6–133, 409–523) that forms a solid well-organized foundation around the waist of the assembly. It provides most of the side-to-side contacts between subunits within the ring and all of the contacts

Phil. Trans. R. Soc. Lond. B (1995) **348**, 113–119
Printed in Great Britain

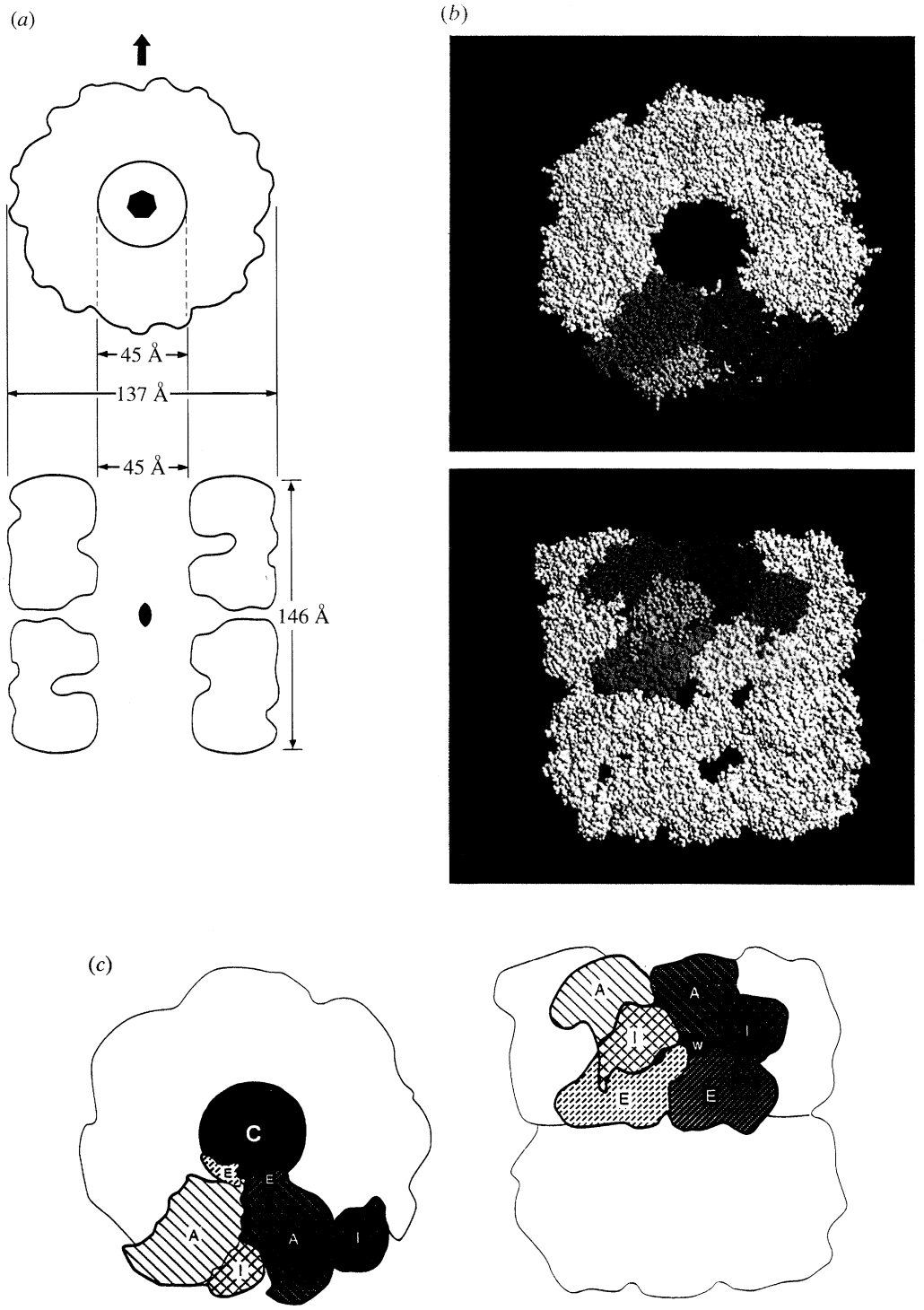


Figure 1. Overall structure of GroEL. (*a*) Dimensions of the tetradecamer. Top view looking down into the opening of the cylindrical channel and a sagittal cut in the plane of the cylindrical axis. Dyads indicated by (↑ and ⬮) and the sevenfold by (⬢). (*b*) Spacefilling model highlighting two adjacent subunits of the top ring. The domains are colour coded in the left subunit (green for equatorial; gold for intermediate and purple for apical) and in the right subunit, yellow for equatorial, red for intermediate, blue for apical). (*c*) Diagram of subunit and domain structure highlighting the same two subunits. E, I, A stand for equatorial, intermediate and apical. W is the external window of the channel that connects to the central channel. Reproduced with permission from Braig *et al.* (1994).

across the equatorial plane that hold the two rings together. Second, an 'apical domain' (191–376) of 186 residues that surrounds the openings at the top and bottom of the central channel. The apical domain's internal structure is less well defined and it is more loosely positioned and oriented in the assembly. Finally, a small slender 'intermediate domain' at the periphery of the cylinder links the equatorial domain to the apical domain through well-ordered secondary structure elements. The connection between the intermediate and apical domains is a pair of antiparallel β strands that appears to form a hinge permitting global movements of the apical domain relative to the rest of the assembly. The connection to the equatorial

domain is a more robust one, where α-helical elements project from the equatorial domain at a point that flanks the ATP binding site. The solvent compartment within the cylinder is a convoluted chamber consisting of the wide central channel with fourteen (seven in each ring) much narrower and irregular extensions reaching outward, underneath the apical domains, towards the intermediate domains and the side windows. Thus, even if the central channel is occupied, small solution components can enter or exit the chamber freely via the side windows. The surface of the solvent chamber is extensive and its contours would be altered significantly by changes in the orientation of the apical domain.

3. THE STRUCTURE DETERMINATIONS

Although the details of the specimen preparation, crystallization, structure determination and refinement have been published, we point out important shortcomings that must be appreciated because they bear on the functional interpretation. In brief, the structure is based on a single good isomorphous derivative (SIR), and phase improvement derived from the sevenfold non-crystallographic symmetry (NCS). The initial experimental electron density was interpretable in terms of a nearly contiguous polypeptide chain with most of the distinctive side chains in register with the amino acid sequence and compatible with the mercury-atom sites attached to the cysteine residues. However, certain regions in the apical domain were not interpretable in terms of the sequence and some were not even contiguous. The original model was used only to build an envelope around each subunit, a procedure required by the averaging algorithm RAVE (Jones *et al.* 1991) which does not necessarily rely on a single sevenfold rotational symmetry operator. This was done by centering a sphere of 5 Å radius at each α-carbon. The original model was then discarded as were the SIR phases. Starting with 8 Å random phases and only the molecular envelope described above, new phases were created with the electron-density averaging program RAVE that allowed the NCS to depart from the original exact sevenfold rotational axis and to relate each of the subunits to a reference subunit through individual operators. Iterative phase extension and periodic upgrading of the NCS operators led to a map that was essentially the same as the original experimental map, thereby, ensuring its validity. There were, however, noticeable improvements in contiguity and side-chain identity in the apical domain which suggested small but significant departures from sevenfold symmetry especially in the apical domain. Indeed, significant discontinuities and ambiguities in side chain assignment – all in the apical domain – were not resolved by the refinement (see figure 2 of Braig *et al.* 1994). Moreover, the violations of backbone torsional restraints (φ, ϕ) and a sudden dip in the three-dimensional profile (Luthy *et al.* 1992) of the segment between 295–315 suggests a mistaken interpretation of a loop extending from across the top of the apical domain towards the central cavity.

Because the mutations that disrupt polypeptide binding are all in the apical domain, and some of them, such as those at 234, 237, 259, 263, 264 are in these poorly defined loops (figures 2 and 3), it is tempting to speculate that the poorly defined surface segments of the apical domain are designed to be flexible enough to bind a wide range of polypeptide sequences. This explanation is supported by the following observations:

(1) The large temperature factors that characterize most portions of the apical domain give way to well-ordered structure in the apical segments that contact the neighbouring intermediate domain. This implies a local rather than a global disordering of the apical domain.

(2) The low temperature factors and the high quality of the (2 Fo-Fc) maps in the intermediate and equatorial domains suggest that the inability to model certain apical segments is not due to inaccurate phases.

(3) A preliminary refined model of ATP-bound GroEL in a different monoclinic, crystal form shows essentially the same trace of the polypeptide and a tendency for the same apical segments to be poorly represented in the electron density (D. Boisvert *et al.*, unpublished).

On the other hand, we cannot confidently ascribe local flexibility to these poorly defined segments until we significantly improve the quality and quantity of high resolution data. In addition to the apical segments discussed above, the N-terminal 5 residues and C-terminal 26 residues are not visualized. The polypeptide segments contiguous to these terminal sequence segments are well defined components of the equatorial domain and suggest that these terminal peptides project into the central cavity in a crystallographically disordered manner. Reconstructions of electron micrographs (Saibil *et al.* 1993; Chen *et al.* 1994) show density in the central cavity at the level of the equatorial domains which may reflect the accumulation of 217 residues from the N and C terminal segments of seven subunits.

4. FUNCTIONAL ISSUES

(*a*) *Relationships between subunits*

The subunits within the ring are held together primarily by contacts between the equatorial domains. Figure 2 shows that one of these contacts is formed by a stem-loop of β-structure that reaches out against the channel surface of the neighbouring equatorial domain forming parallel β-structure with segment 519–522 that just precedes the disordered C-terminus. Mutational analysis is consistent with this picture since C-terminal deletions are tolerated, but only as far as 522; those that include 521 or additional residues are lethal because of GroEL disassembly (Burnett *et al.* 1994). Recently, mutations changing residues that form the contact surface between rings (see figure 4 of Braig *et al.* 1994) were shown to decouple the cylindrical assembly into separate stable seven subunit rings (J. Weissman *et al.* unpublished results). In the wild-type molecule, the seven equatorial domains of each ring face one another to form, in total, a large solvent excluded contact surface (5850 Å^2). In contrast

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

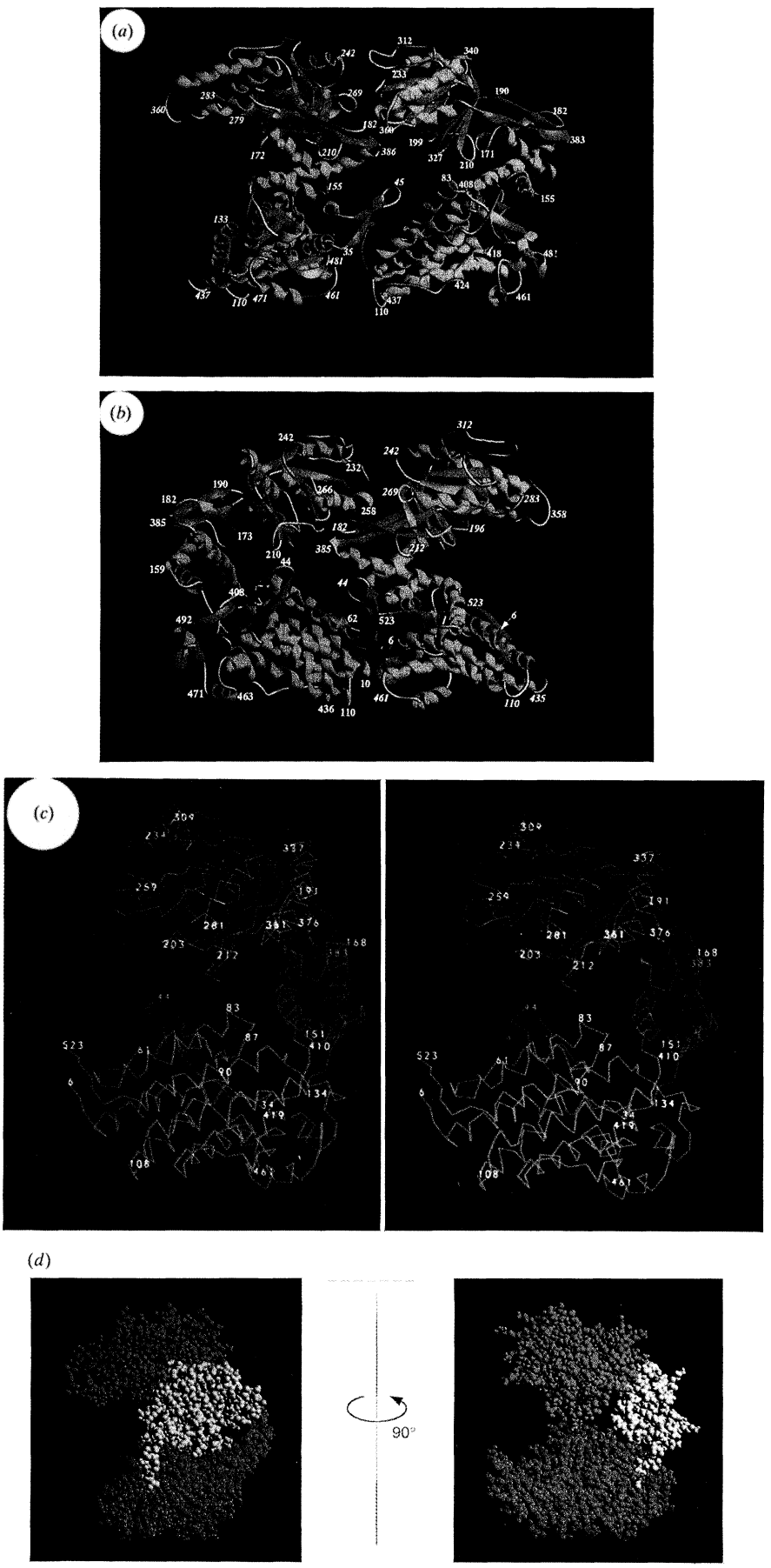


Figure 2. For description see opposite.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

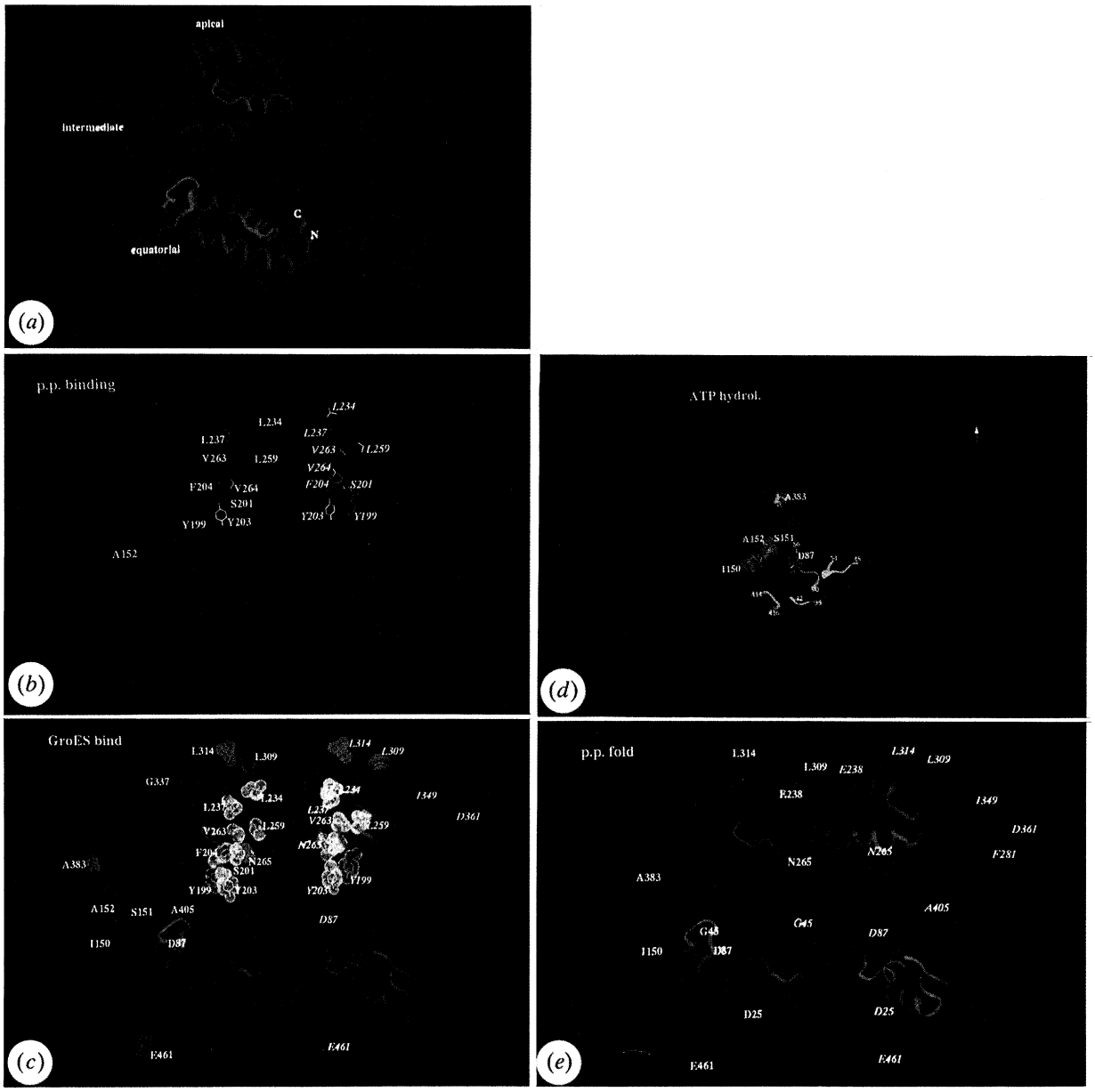


Figure 3. Amino acids involved in GroEL functions. Shown in each panel are two subunits from the 'top' ring of GroEL, viewed from inside (i.e. from central channel). The ATP hydrolysis panel is an exception: the dimer has been rotated horizontally to view the binding pocket from the 'right' side. In panel (*c*) residues whose mutational alteration affected both polypeptide and GroES binding are shown in yellow spheres; other residues whose alteration affected GroES binding but not polypeptide binding are shown in blue spheres. In panel (*d*) residues affecting ATP hydrolysis are shown in green spheres. Conserved residues around the binding pocket are shown by yellow ribbon. (*e*) Shows residues whose alteration interfered with release or folding of polypeptide. Ribbon diagram of main chain derived from the averaged model (Braig *et al.* 1994), displayed in InsightII (BioSym). Reproduced from Braig *et al.* (1994) with permission.

to the contacts within the rings, the complementary surfaces do not interdigitate across the equatorial plane.

The intermediate domain emerges from the equatorial domain adjacent to the ATP-binding pocket and extends diagonally up and to the right (when the top ring is viewed from the outside as in figures 1 and 2), where it contacts the apical domain of the neighbouring subunit. As the intermediate domain contacts the neighbouring apical domain and, of course, is

Figure 2. (*a*) Ribbon drawing of two adjacent subunits in the top ring (see figure 1) viewed from the outside. Beta-strands are wide green arrows, alpha-helices are blue, and Connecting Strands are yellow tubes. (*b*) The same two subunits viewed from the inside of the channel. (*c*) Stereopair of the alpha carbon backbone of a subunit oriented as in the right panel of (*d*). The colour coding is green, blue and red for the equatorial, intermediate and apical domains, respectively. (*d*) Two spacefilling models of the left highlighted subunit in figure 1 as viewed from the outside (left panel) and then rotated 90° so that the outside surface faces towards the right (right panel). Reproduced with permission from Braig *et al.* (1994).

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

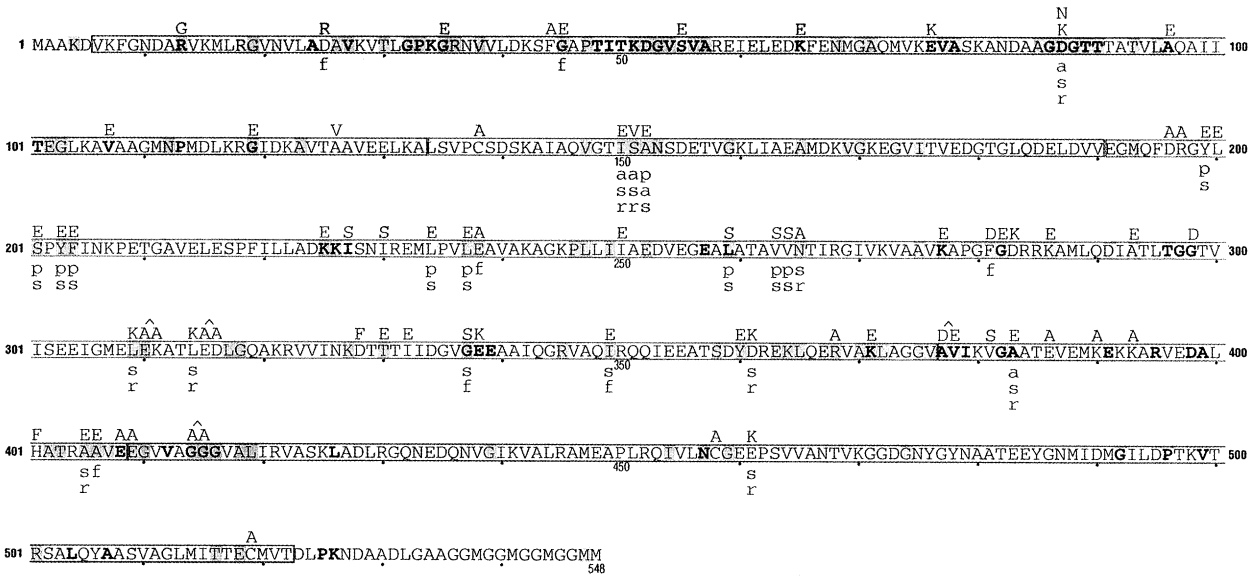


Figure 4. Amino acid substitutions in GroEL and the functions affected shown above and below, respectively, the wild-type amino-acid sequence; a are defective in ATP binding, f are defective in folding, p are defective in peptide binding, r are defective in release and s are defective in GroES binding. Residues conserved in all chaperonins (Hsp60/GroEL family and TF55/TCP1 family) are shown in bold. Residues in at least 49 or 50 Hsp60/GroEL chaperonin shown in orange background.

linked covalently to the apical domain of the same subunit, it provides a structural link that can allosterically couple stereochemical changes in the equatorial domain associated with nucleotide binding and hydrolysis to tertiary and quaternary changes of the apical region. The view that the intermediate domain acts as an allosteric lever involved in the coupling of quaternary changes to the enzymatic cycling of ATP and ADP is supported by a mutational change that breaks an ionic interaction between the apical domains (Arg 197) and neighbouring intermediate domains (Glu 386) and, thereby, disrupts positive cooperativity in ATPase action within the ring and the negative cooperativity between the rings (Yifrach *et al.* 1994). Although the structure and genetics provide strong clues to allosteric effects within the ring, it is less clear how enzymic and binding events influence structure across the equatorial plane in the opposite ring.

(*b*) *Peptide binding*

Fenton *et al.* (1993) have reported mutations that disrupt binding of polypeptide substrates (figure 3). Although an exhaustive analysis has not been completed, it appears that all of the changes that inhibit the binding of non-native polypeptides (without disruption of the GroEL assemblies) affect residues that are hydrophobic in character mapping within the apical domain. All of these mutational changes are at the surface and either face towards the central channel or lie on the domains' undersurface which forms the 'roof' of the outward evagination of the central channel. Such residues may form hydrophobic contacts with exposed hydrophobic surfaces of nonnative intermediates. Most interesting is the fact that all of these mutations also disrupt binding of GroES (figure 3). In sum, from a structural vantage point, the hinge-like attachment of the apical domain, the allosterically responsive contact to the neighbouring intermediate domain and the porosity of the assembly that allows room for structural adjustment are structural features seemingly designed to potentiate significant structural change in the apical domain and, thereby, present a range of potentially flexible binding surfaces to polypeptides and GroES,

Despite such flexibility, the crystal structure provides dimensional constraints on GroEL's capacity to bind non-native polypeptide chains. From electron microscopic (Braig *et al.* 1993; Langer *et al.* 1993; Saibil *et al.* 1993, 1994; Chen *et al.* 1994) and mutation studies (Fenton *et al.* 1994), it is likely that a significant portion of the bound polypeptide is confined to the central cavity. The 45 Å opening at the top and bottom of the channel defines a cylindrical space of ~ 250000 Å^3 from one opening to the other. A native protein in a crystal lattice occupies about 2.4 Å^3 per Dalton which means the entire channel could accept an elongated ellipsoidal protein of ~ 100 kDa provided it could cross the equatorial plane. GroEL does not bind fully folded native proteins but is thought to bind non-native intermediates with native-like secondary structure and variable amount of tertiary structure (Hayer–Hartl *et al.* 1994; Okazaki *et al.* 1994; Robinson *et al.* 1994; Zahn *et al.* 1994). A classical 'molten globule' conformation (Okazaki *et al.* 1994), while it may turn out to be more collapsed than what is generally recognized, already occupies 20 to 30% more volume than the native structure, meaning ~ 75–80 kDa could be accommodated in the entire 45 Å diameter cylindrical channel. As only one side of the double toroid is thought to bind polypeptide at one time (Hendrick *et al.* 1993; Horwich & Willison 1993; Todd *et al.* 1994), this figure would be halved to 38–40 kDa per ring.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

It is particularly surprising that residues, such as Tyr 199 and Tyr 203 (figure 3), located deep within the central channel on a re-entrant surface of the apical domain, are mutationally implicated in both peptide binding and GroES binding. Whereas, peptides almost certainly bind in the central channel and may reach these deeply situated residues by either a further unfolding or adjustments in orientation of the apical domain, the case for the interaction of native GroES with such residues is not so clear. GroES could fit into the central channel of GroEL. Recent cryoelectromicrograph studies of Saibil and coworkers, however, of the GroEL–GroES binary complex, suggest that the GroEL apical domain undergoes a conformational change that extends it upward to meet the edges of a dome-shaped GroES 7-mer (Saibil *et al.* 1993; Chen *et al.* 1994). The central face of the apical domains including the residues implicated in peptide binding thus appear to be able to become much more accessible to contact with GroES. This raises the interesting question of whether GroES and peptide compete for the same binding sites on the same side of GroEL, or whether the two ligands generally interact at opposite ends of the cylinder. This is a major question that must be resolved experimentally.

5. GroES BINDING, ATP BINDING AND HYDROLYSIS, AND POLYPEPTIDE RELEASE

Figure 4 summarizes the residues where mutational changes disrupt key functions. For many residues, there is a noticeable overlap (Fenton *et al.* 1994). For example, GroES binding requires all of the same hydrophobic residues on the inside surface of the apical domain that are required for polypeptide binding. However, GroES also requires two non-polar residues on the outer top surface of the apical domain, as well as, other polar and non-polar residues throughout the intermediate and equatorial domains (one of which forms a contact between rings!). Many of these same residues are also implicated in ATP binding and hydrolysis as well as peptide release. ATP hydrolysis is sensitive to changes in the intermediate domain and to one change (involving D87) in the ATP-binding pocket of the equatorial domain. The location of such mutations reinforces the view that the intermediate domain is involved in the allosteric modulations of the ATPase function.

Finally, it should be stressed that the overlapping distribution of mutations affecting the binding and release of nucleotide, polypeptide and GroES as well as the hydrolysis of ATP underscores the interdependence of such functions. For a more precise understanding of the chemistry of this interdependence, we must visualize the various liganded states of GroEL. This work is in progress.

We acknowledge the contributions to this work of our colleagues: Dr Z. Otwinowski, Dr K. Braig, Dr R. Hegde, Dr D. Boisvert, Dr A. Joachimiak, Dr W. Fenton, Dr Y. Kashi and Dr K. Furtak. In addition, we thank Dr J. Weissmann, Dr G. Lorimer and Dr H. Saibil for stimulating discussion.

REFERENCES

Braig, K.B., Otwinowski, Z., Hegde, R., Boisvert, D.C., Joachimiak, A., Horwich, A.I. & Sigler, P.B. 1994 *Nature, Lond.* **371**, 578–586.

Braig, K., Simon, M., Furuya, F., Hainfeld, J.F., & Horwich, A.L. 1993 *Proc. natn. Acad. Sci. U.S.A.* **90**, 3978–3982.

Burnett, B., Horwich, A.L. & Low, K.B. 1994 *J. Bact.* **176**, 6980–6985.

Chen, S., Roseman, A.M., Hunter, A.S., Wood, S.P., Burston, S.G., Ranson, N.A., Clarke, A.R. & Saibil, H.R. 1994 *Nature, Lond.* **371**, 261–264.

Ellis, R.J. & van der Vies, S.M. 1991 *A. Rev. Biochem.* **60**, 321–347.

Fenton, W.A., Kashi, Y., Furtak, K. & Horwich, A.L. 1993 *Nature, Lond.* **371**, 614–619.

Gething, M.J. & Sambrook, J. 1992 *Nature, Lond.* **355**, 33–45.

Hayer-Hartl, M.K., Ewbank, J.J., Creighton, T.E. & Hartl, F.U. 1994 *EMBO J.* **13**, 3192–3202.

Hendrick, J.P. & Hartl, F.-U. 1993 *A. Rev. Biochem.* **62**, 349–384.

Horwich, A.L. & Willison, K. 1993 *Phil. Trans. R. Soc. Lond.* B **339**, 313–326.

Jones, T.A. *et al.* 1991 *Acta Crystallogr.* **A47**, 110–119.

Langer, T., Pfeifer, G., Martin, J., Baumeister, W. & Hartl, F.U. 1992 *EMBO J.* **13**, 4757–4765.

Luthy, R., Bowie, J.U. & Eisenberg, D. 1992 *Nature, Lond.* **356**, 83–85.

Martin, J., Mayhew, M., Langer, T. & Hartl, F.U. 1993 *Nature* **366**, 228–233.

Okazaki, A., Ikura., T., Nakaido, K. & Kuwajima, K. 1994 *Struct. Biol.* **1**(7), 439–445.

Robinson, C.V., Broβ, M., Eyles, S.J., Ewbank, J.J., Mayhew, M., Hartl, F.U., Dodson, C. & Radford, S.E. 1994 *Nature, Lond.* **372**, 646–651.

Saibil, H.R. *et al.* 1993 *Curr. Biol.* **3**, 265–273.

Saibil, H.R. 1994 *Nature Struct. Biol.* **1**, 838–842.

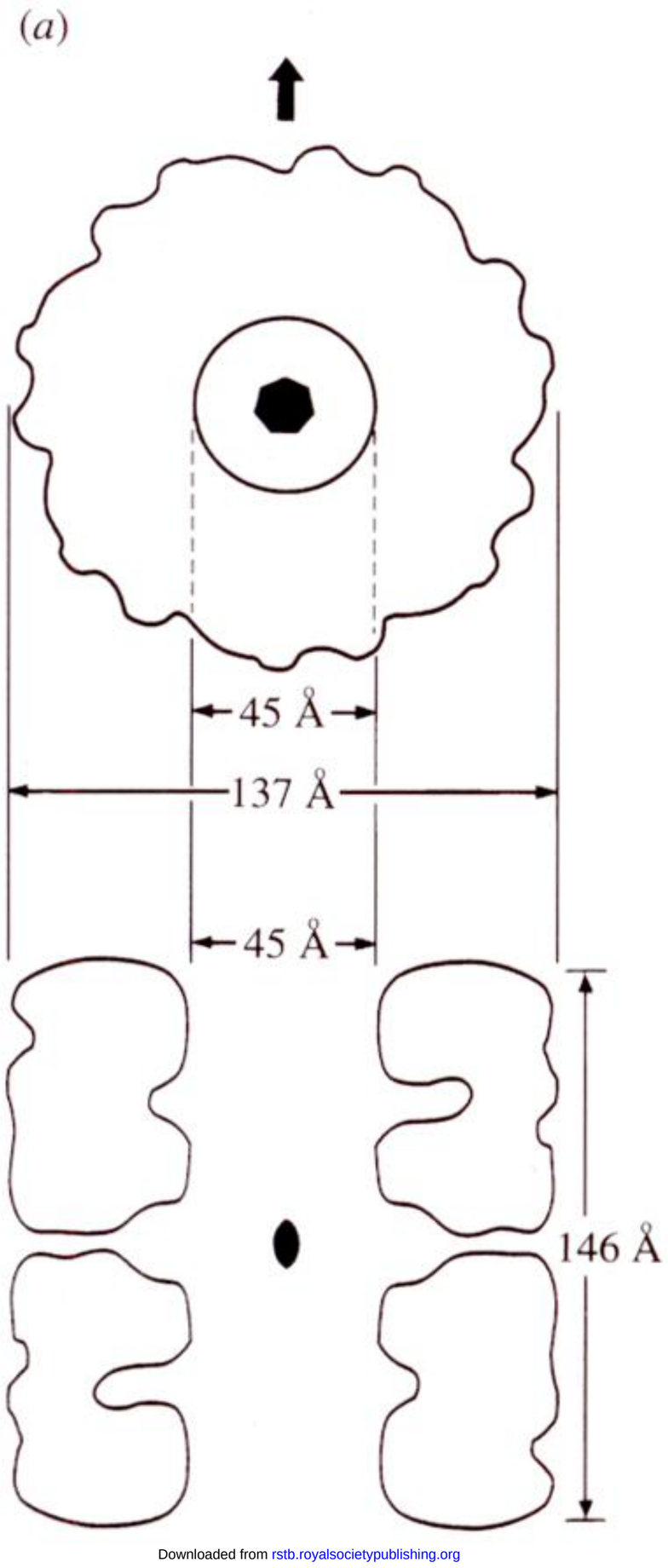
Todd, M.J., Viitanen, P.V. & Lorimer, G.H. 1994 *Science, Wash.* **265**, 659–666.

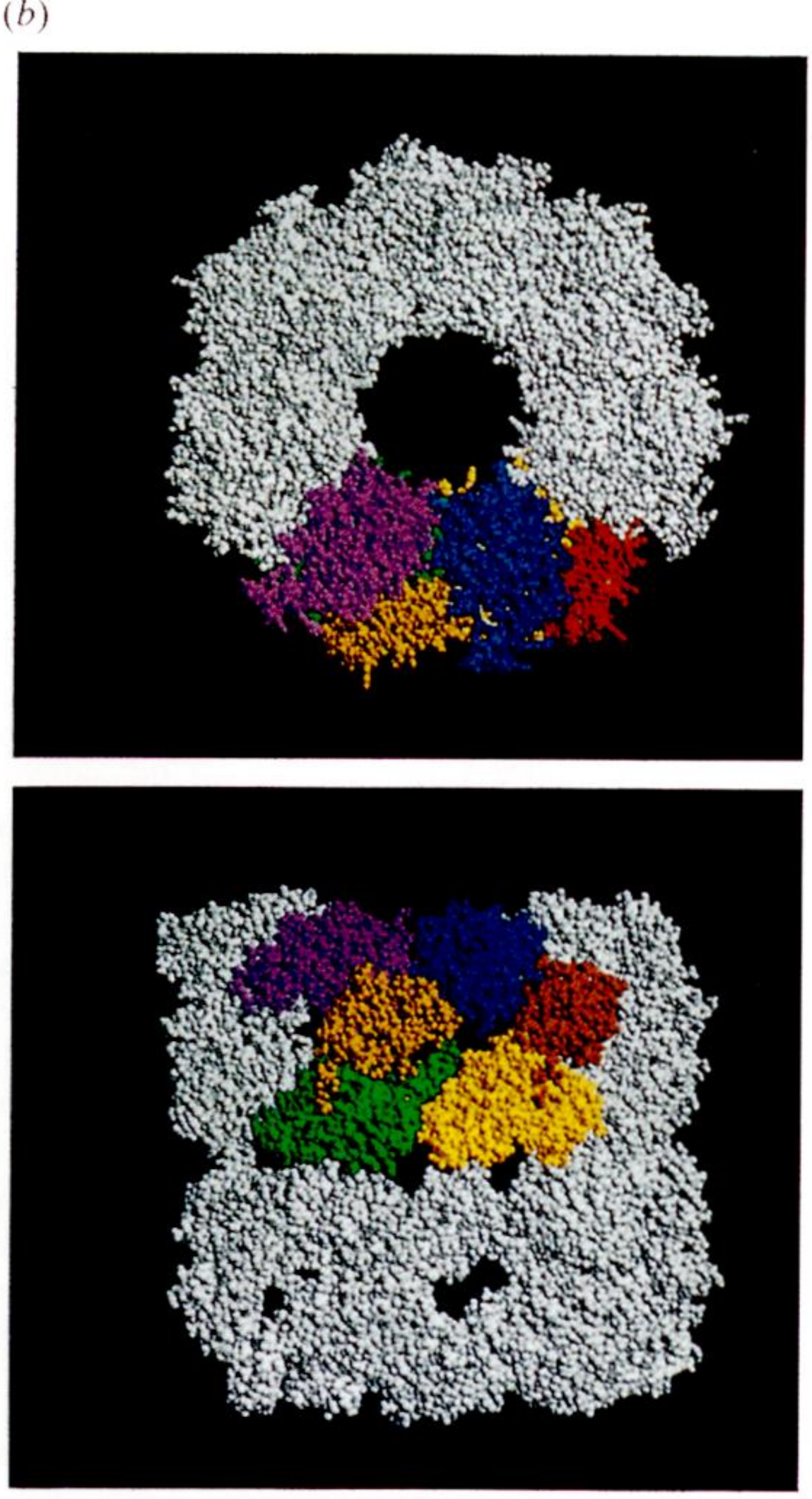
Weissman, J.S., Kashi, Y., Fenton, W.A. & Horwich, A.L. 1994 *Cell* **78**, 693–702.

Yifrach, O. & Horovitz, A. 1994 *J. molec. Biol.* **243**, 397–401.

Zahn, R., Spitzfaden, C., Ottiger, M., Wuthrich, K. & Pluckthun, A. 1994 *Nature, Lond.* **368**, 261–265.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES





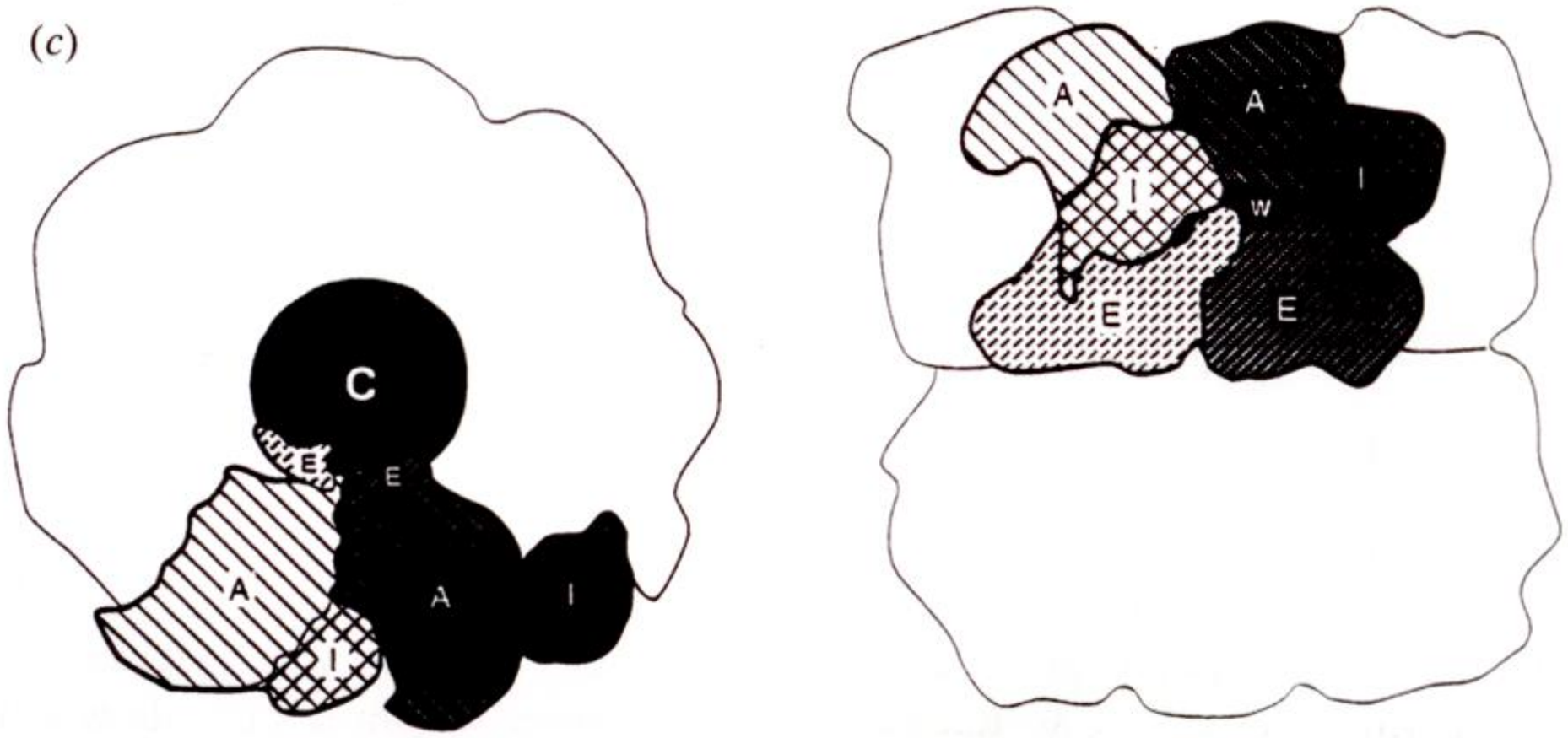
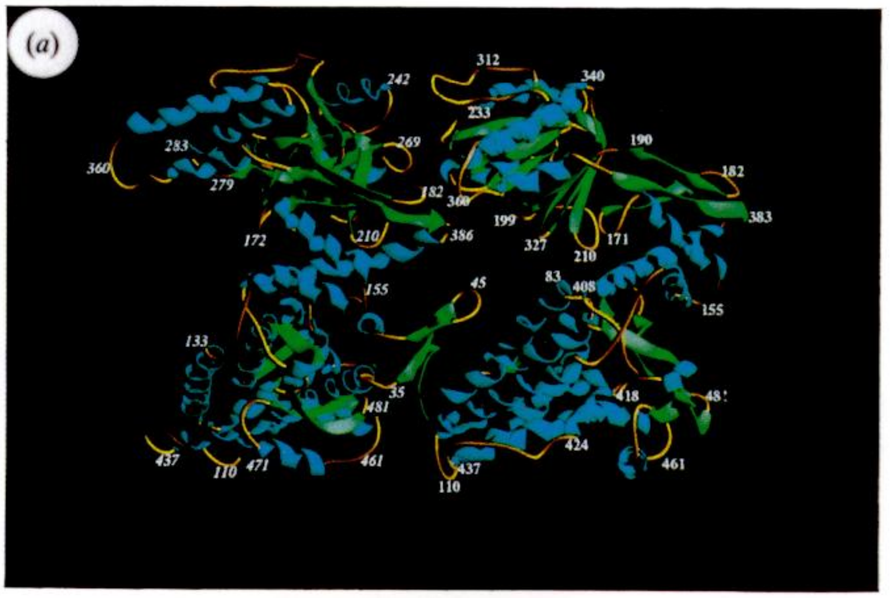
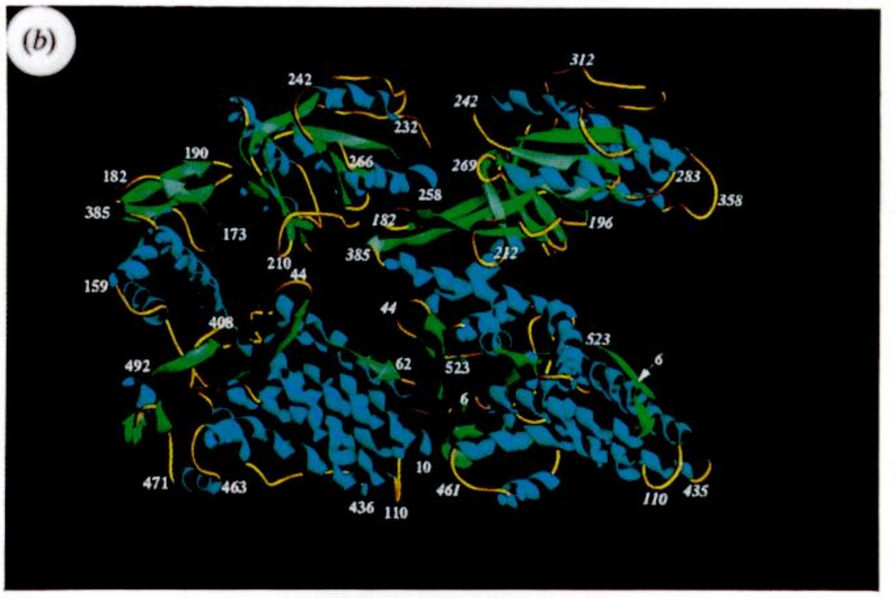
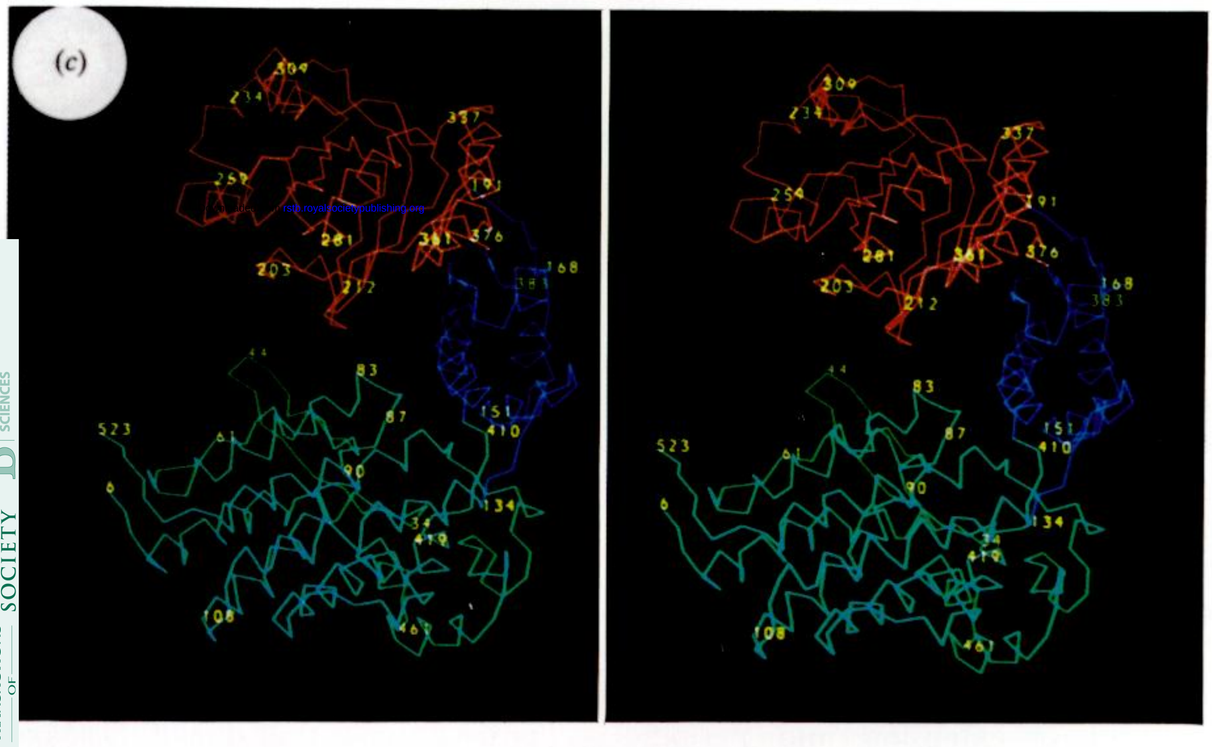


Figure 1. Overall structure of GroEL. (*a*) Dimensions of the tetradecamer. Top view looking down into the opening of the cylindrical channel and a sagittal cut in the plane of the cylindrical axis. Dyads indicated by (↑ and ⬮) and the sevenfold by (⬢). (*b*) Spacefilling model highlighting two adjacent subunits of the top ring. The domains are colour coded in the left subunit (green for equatorial; gold for intermediate and purple for apical) and in the right subunit, yellow for equatorial, red for intermediate, blue for apical). (*c*) Diagram of subunit and domain structure highlighting the same two subunits. E, I, A stand for equatorial, intermediate and apical. W is the external window of the channel that connects to the central channel. Reproduced with permission from Braig *et al.* (1994).







(d)

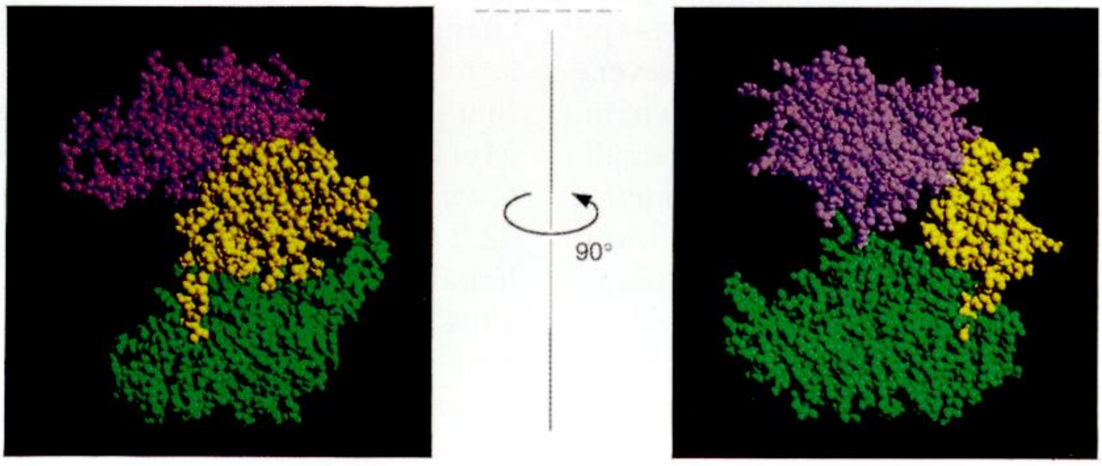
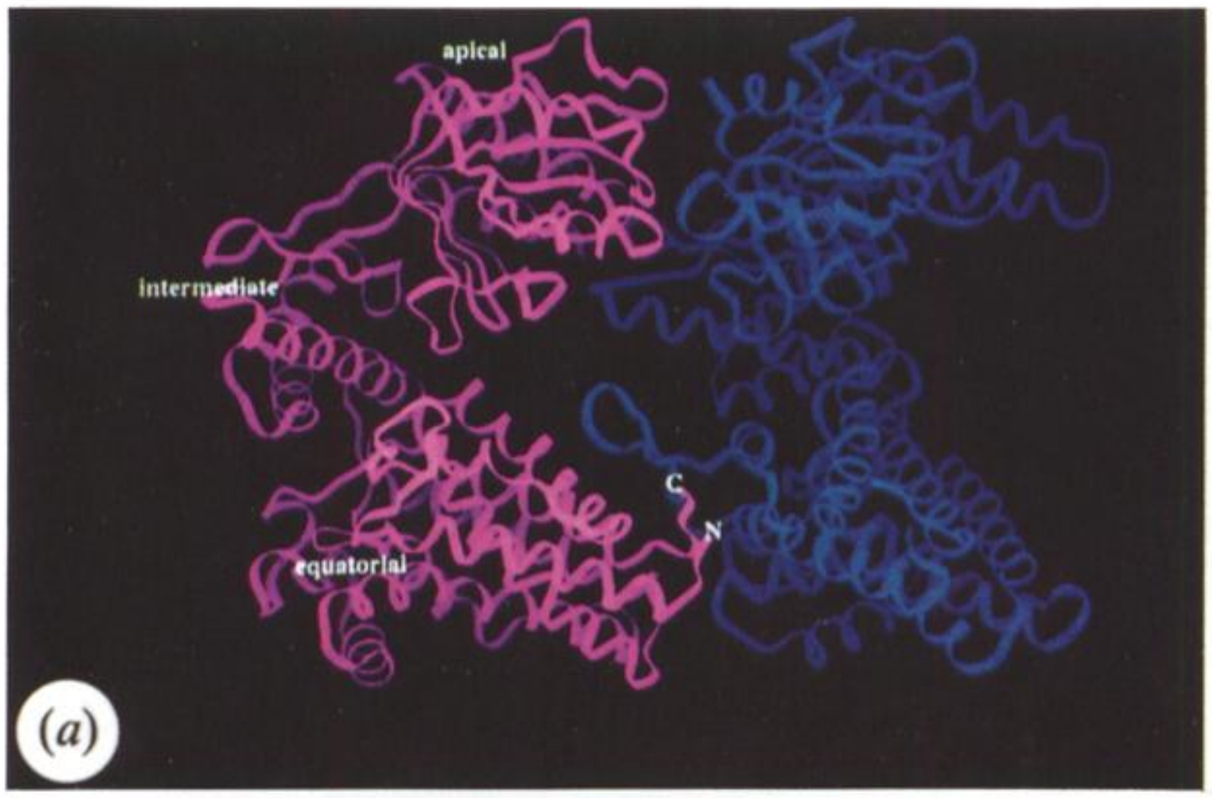
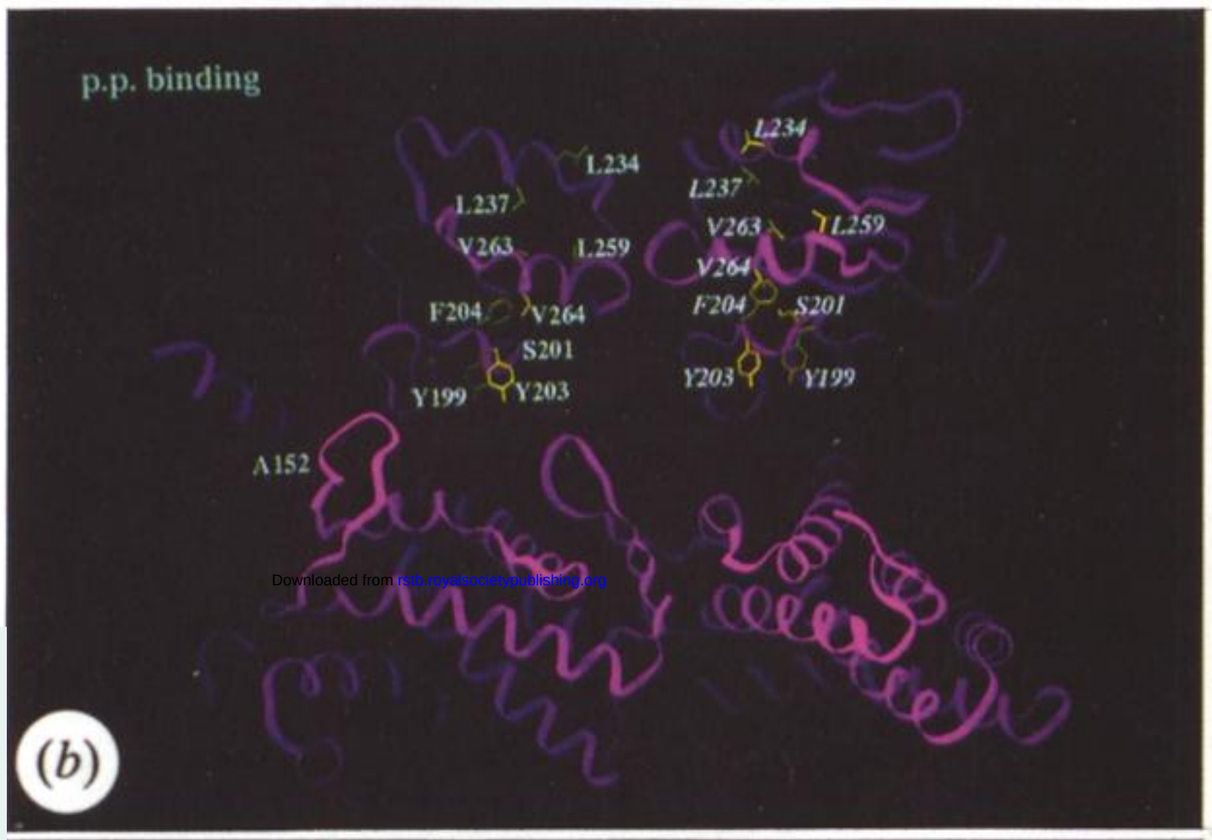
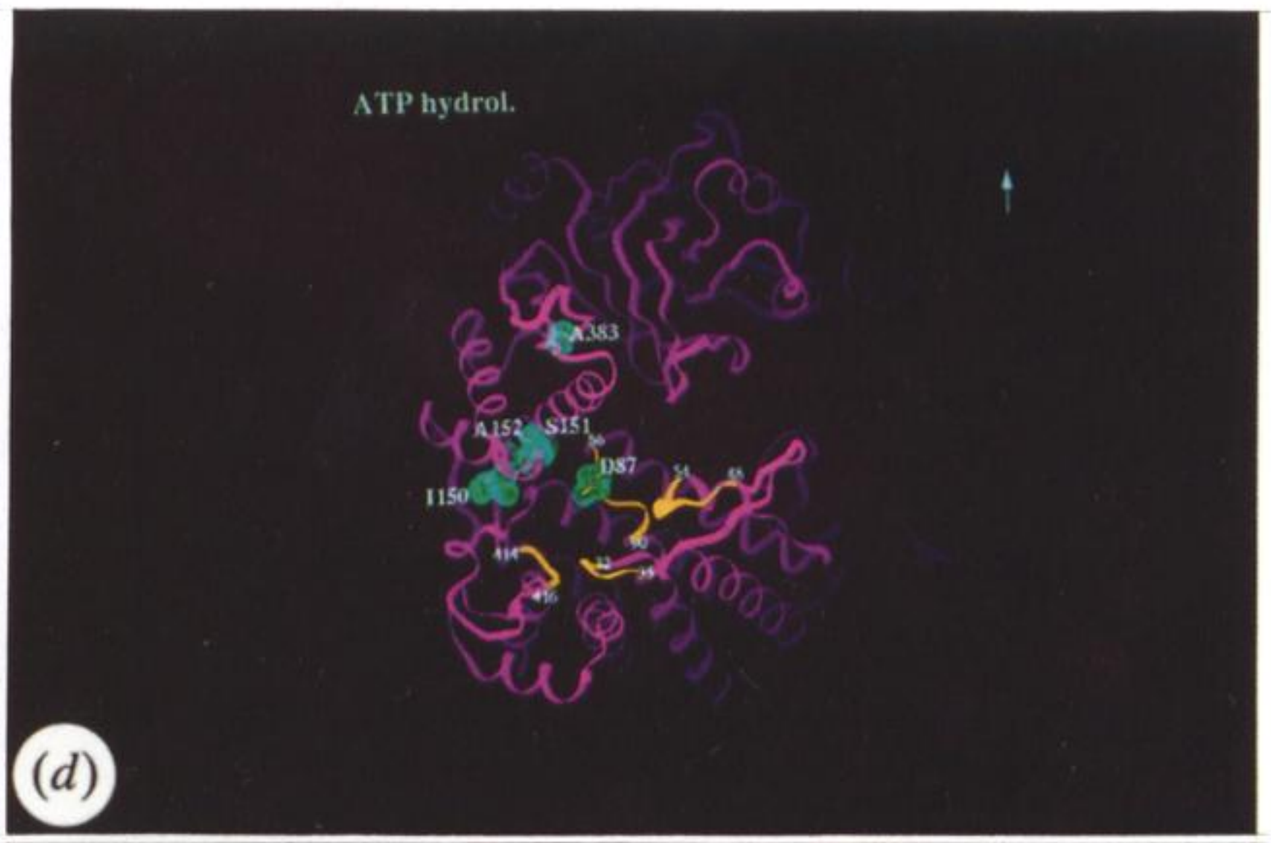


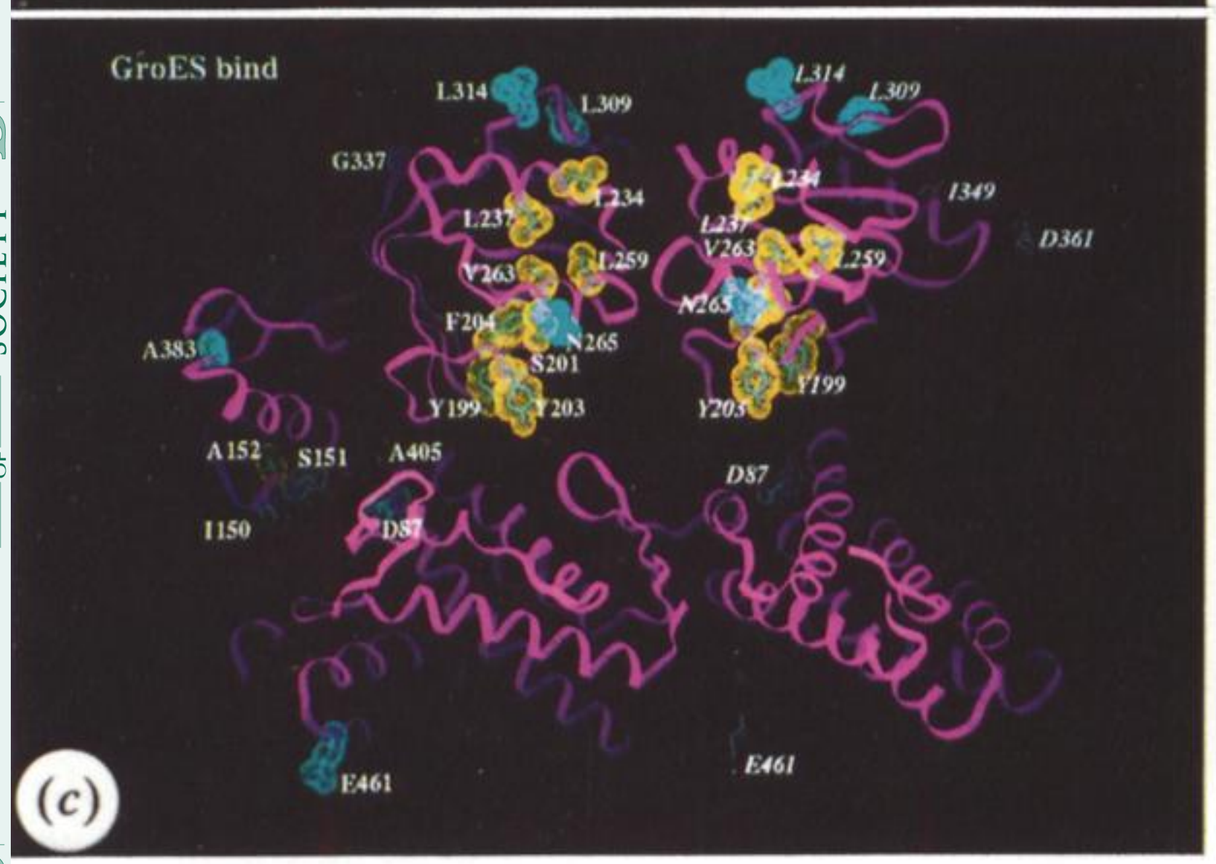
Figure 2. For description see opposite.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES









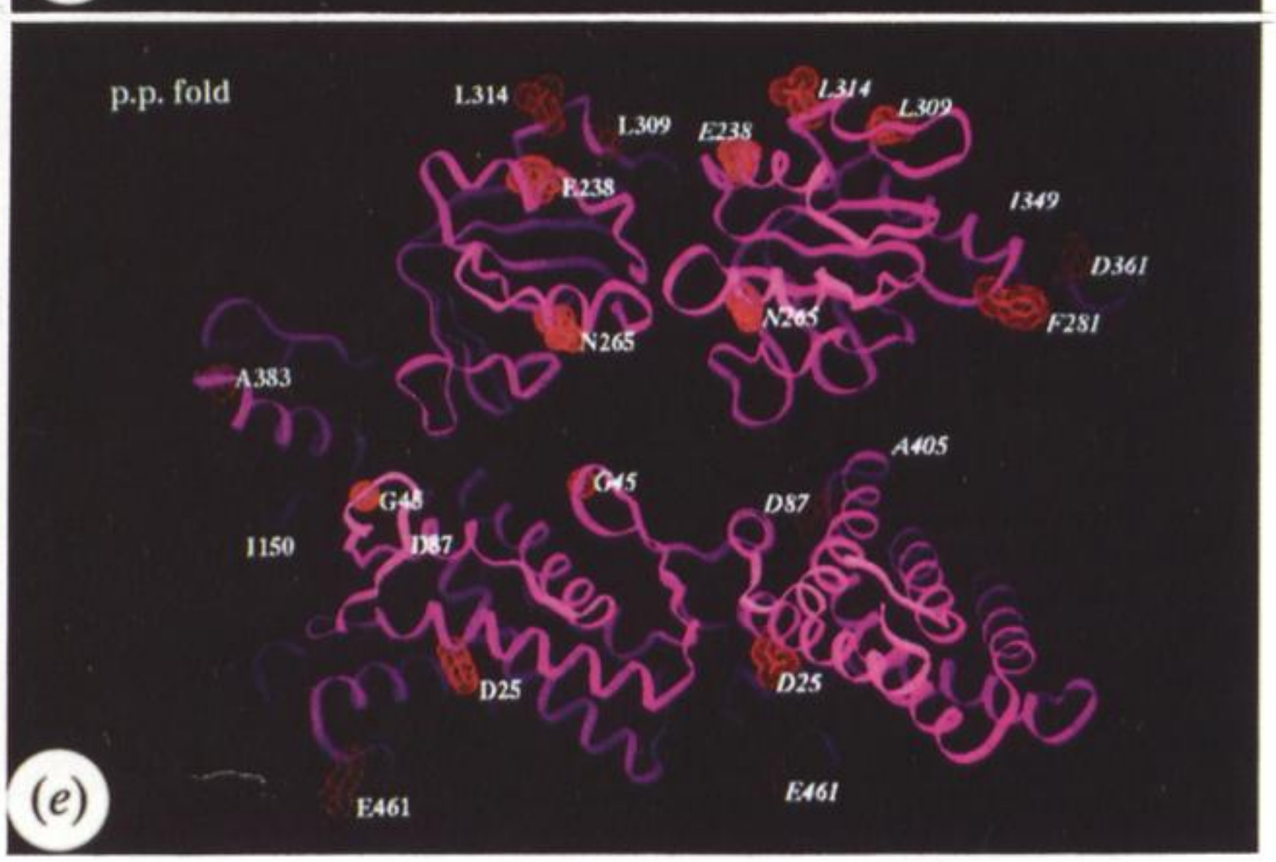


Figure 3. Amino acids involved in GroEL functions. Shown in each panel are two subunits from the 'top' ring of GroEL, viewed from inside (i.e. from central channel). The ATP hydrolysis panel is an exception: the dimer has been rotated horizontally to view the binding pocket from the 'right' side. In panel (*c*) residues whose mutational alteration affected both polypeptide and GroES binding are shown in yellow spheres; other residues whose alteration affected GroES binding but not polypeptide binding are shown in blue spheres. In panel (*d*) residues affecting ATP hydrolysis are shown in green spheres. Conserved residues around the binding pocket are shown by yellow ribbon. (*e*) Shows residues whose alteration interfered with release or folding of polypeptide. Ribbon diagram of main chain derived from the averaged model (Braig *et al.* 1994), displayed in InsightII (BioSym). Reproduced from Braig *et al.* (1994) with permission.

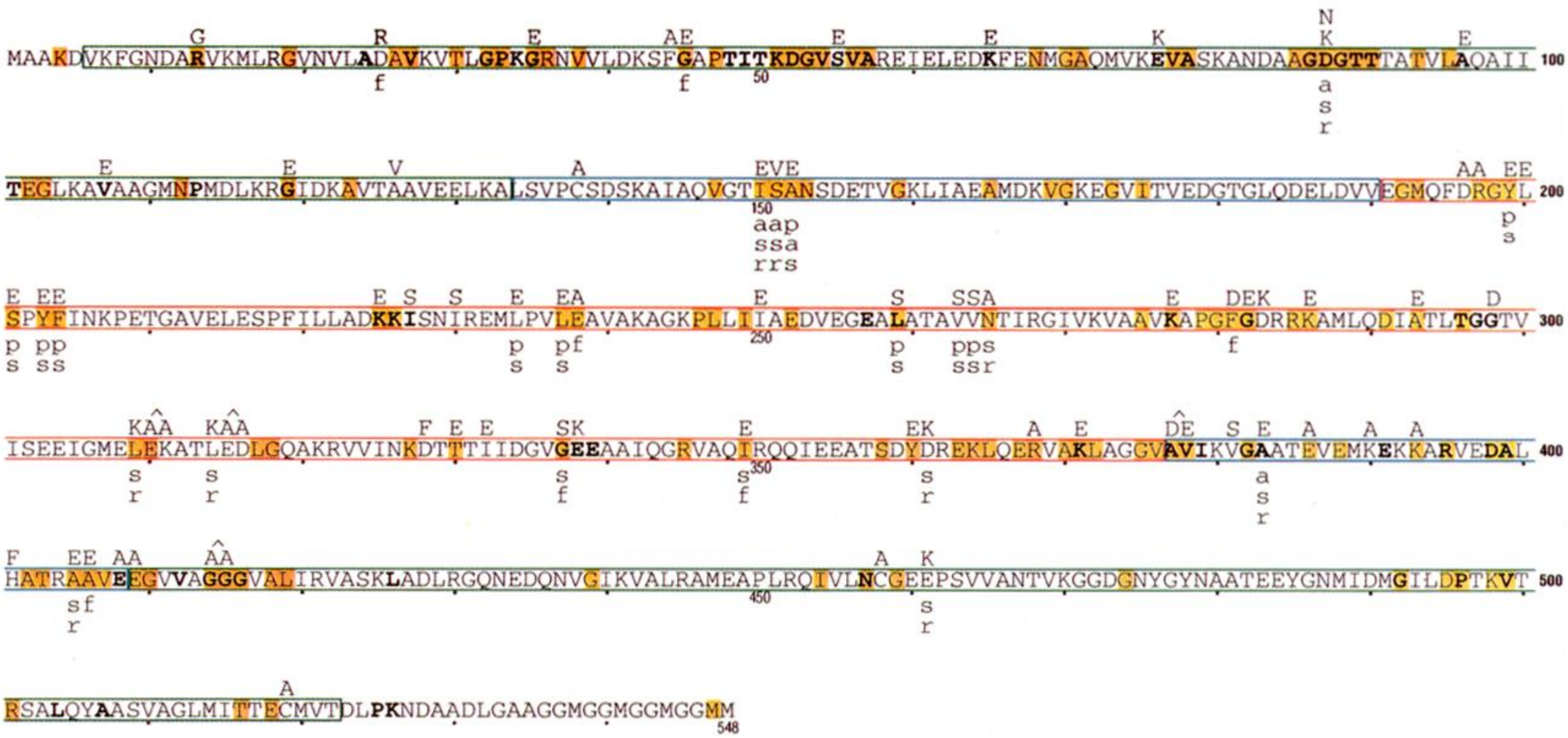


Figure 4. Amino acid substitutions in GroEL and the functions affected shown above and below, respectively, the wild-type amino-acid sequence; a are defective in ATP binding, f are defective in folding, p are defective in peptide binding, r are defective in release and s are defective in GroES binding. Residues conserved in all chaperonins (Hsp60/GroEL family and TF55/TCP1 family) are shown in bold. Residues in at least 49 or 50 Hsp60/GroEL chaperonin shown in orange background.