PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

The General Concept of Molecular Chaperones

R. John Ellis

Phil. Trans. R. Soc. Lond. B 1993 **339**, 257-261
doi: 10.1098/rstb.1993.0023

The general concept of molecular chaperones

R. JOHN ELLIS

Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, U.K.

SUMMARY

This introductory article proposes a conceptual framework in which to consider the information that is emerging about the proteins called molecular chaperones, and suggests some definitions that may be useful in this new field of biochemistry. Molecular chaperones are currently defined in functional terms as a class of unrelated families of protein that assist the correct non-covalent assembly of other polypeptide-containing structures *in vivo*, but which are not components of these assembled structures when they are performing their normal biological functions. The term assembly in this definition embraces not only the folding of newly synthesized polypeptides and any association into oligomers that may occur, but also includes any changes in the degree of either folding or association that may take place when proteins carry out their functions, are transported across membranes, or are repaired or destroyed after stresses such as heat shock. Known molecular chaperones do not convey steric information essential for correct assembly, but appear to act by binding to interactive protein surfaces that are transiently exposed during various cellular processes; this binding inhibits incorrect interactions that may otherwise produce non-functional structures. Thus the concept of molecular chaperones does not contradict the principle of protein self-assembly, but qualifies it by suggesting that *in vivo* self-assembly requires assistance by other protein molecules.

1. HISTORY OF THE MOLECULAR CHAPERONE CONCEPT

The term 'molecular chaperone' was used first to describe the properties of a nuclear protein, nucleoplasmin, in mediating the *in vitro* assembly of nucleosomes from isolated histones and DNA (Laskey *et al.* 1978). This term was chosen because nucleoplasmin mediates nucleosome assembly by preventing incorrect interactions between histones and DNA, without either providing steric information essential for assembly or being a component of the assembled nucleosomes themselves. In these respects, nucleoplasmin is a molecular analogue of the human chaperone, whose traditional role is to prevent incorrect interactions between pairs of human beings, without either providing the steric information necessary for their correct interaction or being present during their married life.

The author came across this term while searching for a precedent for the observation that the assembly of the enzyme rubisco (ribulose bisphosphate carboxylase-oxygenase) in chloroplasts isolated from higher plants also seems to require the transient assistance of another protein that is not a component of the assembled enzyme. The essential finding is that rubisco large subunits, newly synthesized by isolated intact chloroplasts, are bound non-covalently to another abundant protein before transfer to the holoenzyme; it was proposed that this binding might be an obligatory step in rubisco assembly (Barraclough & Ellis 1980). This interpretation did not meet with much enthusiasm at the time because of the general acceptance of the notion of protein self-assembly, which proposes that both the folding and association of polypeptides is a spontaneous process requiring no molecules other than the components of the assembled protein (Caspar & Klug, 1962; Anfinsen 1973).

The suggestion that the rubisco large subunit-binding protein could be regarded as a second example of a molecular chaperone was made at a Royal Society Discussion Meeting on rubisco (Musgrove & Ellis 1986). At this time it was felt that histones–DNA and rubisco subunits might be special cases, as the tendency of both to form non-specific aggregates *in vitro* is so great. However, a speculative paper by Pelham (1986) suggested that the need for a molecular chaperone function may be more widespread. Although he did not use the term 'molecular chaperone', Pelham proposed that members of the heat shock protein 70 (hsp 70) family in animal and microbial cells are involved in the assembly and disassembly of proteins in the nucleus, cytosol and endoplasmic reticulum. Some members of the hsp 70 family are present in unstressed cells, and can bind to denatured or abnormal proteins in a manner reversible by adenosine triphosphate (ATP). Pelham suggested that these proteins have a role in normal protein folding and association in unstressed cells, and are required in increased amounts when proteins have been damaged by stress, both to unscramble protein aggregates which could then refold correctly, and to prevent further damage by binding to exposed hydrophobic surfaces. This seminal paper emboldened the author to propose that all cells contain a variety of proteins that act as molecular chaperones in a number

Phil. Trans. R. Soc. Lond. B (1993) **339,** 257–261
Printed in Great Britain

of basic cellular processes, such proteins including among others nucleoplasmin, the rubisco large subunit-binding protein and the hsp 70 family (Ellis 1987). This more general proposal has since been steadily extended to a growing range of proteins and cellular processes (Ellis & Hemmingsen 1989; Ellis *et al.* 1989; Rothman 1989; Ellis 1990*a*,*b*; Ellis & van der Vies 1991; Gething & Sambrook 1992; Lorimer 1992; Hartl *et al.* 1992).

One particular family of molecular chaperones are termed the chaperonins after the discovery that the rubisco large subunit-binding protein of chloroplasts is about 50% identical in aminoacyl sequence to the groEL protein of *Escherichia coli* (Hemmingsen *et al.* 1988). The chaperonins are now regarded as just one family within the wider class of molecular chaperones. The specific function of this particular family is to assist the folding of polypeptides in all types of cell; this family should not be confused with other families of molecular chaperone such as the one containing heat shock 70 proteins, which have different functions and aminacyl sequences to the chaperonins. More recent information about some of the different types of molecular chaperone is contained in the following articles in this symposium, whereas this article discusses the general concept.

2. THE MOLECULAR CHAPERONE CONCEPT TODAY

The discovery of molecular chaperones is stimulating a re-examination of a biological process that was thought to be understood: protein assembly. The conventional view of protein assembly, as found in textbooks, is that it is predominantly a process of self-assembly. According to this view, all the information required to specify the structure and function of a protein resides within the aminoacyl sequences of the polypeptides comprising that protein. Interpreted strictly, this view implies that a newly synthesized polypeptide should be able to attain its functional conformation within the intracellular environment with no assistance from other molecules and with no further expenditure of energy. This conformation often produces the ability to associate specifically with other macromolecules, especially other proteins or nucleic acids. This self-assembly principle stems from the classic observations of Fraenkel-Conrat & Williams (1955), who were able to reassemble infectious tobacco mosaic virus by incubating together the separated purified virion components, and by Anson (1945) and Anfinsen (1973), who found that some purified denatured proteins regain their characteristic biological activities on removal of the denaturing agent in the absence of other macromolecules. Although it was speculated that other molecules may assist protein folding *in vivo* (Epstein *et al.* 1963), subsequent generations of researchers studying protein renaturation *in vitro* did not pursue this possibility until the chaperonin family of molecular chaperones was identified by Hemmingsen *et al.* (1988).

Creighton (1984) has pointed out that all the evidence for self-assembly comes from *in vitro* experiments, and the fact that in many cases the denaturation of proteins is not fully reversible *in vitro*, especially at physiological temperatures and at protein concentrations approaching those found *in vivo*, has not until recently raised serious doubt about the validity of the self-assembly principle to describe the *in vivo* situation. The molecular chaperone concept challenges the conventional view by proposing that self-assembly is not the predominant process by which proteins assemble *in vivo*. This proposal is supported by the growing number of instances where proteins will not assemble correctly at the rates and yields required *in vivo* unless other pre-existing proteins are present to assist them. It is this latter group of proteins that are called molecular chaperones.

(*a*) *Definition of the term 'molecular chaperone'*

Molecular chaperones are defined as a functional class of unrelated families of protein that assist the correct non-covalent assembly of other polypeptide-containing structures *in vivo*, but are not components of these assembled structures when they are performing their normal biological functions.

The words used in this definition have been chosen with some care so as not to conflict with future likely discoveries about molecular chaperones or to overlap with existing terms. The definition is based on the function of each molecular chaperone and not on its structure, but it contains no constraints as to the mechanism of that function, hence the use of the imprecise word 'assist'. Thus different molecular chaperones may act in either a catalytic or a non-catalytic manner, may either accelerate or slow down assembly processes, and may either convey steric information essential for assembly or simply inhibit incorrect side-reactions during self-assembly.

All the molecular chaperones studied to date appear to act not by providing steric information essential for assembly but by inhibiting incorrect interactions which produce non-functional structures during self-assembly processes. However, the above definition allows for other mechanisms of action that may be discovered; the *pro*-sequences of subtilisin and α-lytic protease act as molecular chaperones (Silen & Agard 1989) and it is possible, but not established, that they convey steric information essential for the folding of the mature proteases. Thus, by the definition suggested above, only two criteria must be met for a protein to be described as a molecular chaperone: it must in some sense be required for the correct assembly of other protein-containing structures *in vivo* (the mechanism by which it does this being irrelevant), and it must not be a component of the functional assembled structures.

The term 'non-covalent' used in the definition of molecular chaperones is intended to exclude those proteins which perform covalent post-translational modifications to some proteins. Such proteins are often important for protein assembly but are not the molecules under consideration here. Thus, by this definition, protein disulphide isomerase is not a molecular chaperone.

The definition of molecular chaperone proposed above does not require that the molecule possessing the chaperone activity is necessarily synthesized as a separate entity from its target polypeptide-containing structures, only that it is not a component of these structures when they are performing their biological functions. Thus cotranslational (or intramolecular) molecular chaperones, as represented by the *pro*-sequences of some proteases and the ubiquitin tails of some ribosomal precursor proteins (see Ellis & van der Vies 1991), are not excluded by this definition.

(*b*) *Definition of the term 'protein assembly'*

The term 'protein assembly' is used in the context of molecular chaperones in a broad sense; it embraces not only the folding of newly synthesized polypeptide chains and any association into oligomers that may occur, but also any changes in the degree of either folding or association that may take place as proteins perform their normal functions, are transported across membranes, or are repaired or destroyed after damage by stress.

It is important to appreciate this broad use of the term, because some authors use 'protein assembly' in a much more limited sense, to mean just the association of monomers into oligomers. This broad use of the term 'protein assembly' covers the observation that molecular chaperones function not only during several cellular processes under normal conditions, but also to limit damage to proteins caused by stresses such as heat shock. In other words, at least some heat shock proteins function as molecular chaperones, but not all molecular chaperones are heat shock proteins. It is possible to view the stress response as an amplification of a pre-existing molecular chaperone function which all cells require under normal growth conditions, rather than as a novel function induced by stress.

3. WHY DO MOLECULAR CHAPERONES EXIST?

As the principle of protein self-assembly is well supported by *in vitro* studies of many proteins, it is important to ask why this principle should apparently be insufficient in the more complex cellular context. A possible explanation stems from the observation that several fundamental cellular processes involve the transient exposure of interactive protein surfaces to the intracellular environment, and thereby run the risk that these surfaces may interact incorrectly.

The term 'interactive surfaces' refers to any regions of inter- or intra-molecular contact which are important in maintaining protein-containing structures in their biologically active forms. Such surfaces may be either charged or hydrophobic regions, for example, and they may occur on either partly folded or fully folded polypeptides. Cellular processes involving the transient exposure of such surfaces include protein synthesis, protein transport, protein turnover, the disassembly of oligomers (e.g. DNA replication and clathrin cage recycling), the assembly of oligomers inside organelles from monomers either imported or made in *organello*, and protein damage due to environmental stress. As an example, let us consider protein synthesis. This vectorial process produces the aminoterminal region of each polypeptide before the carboxy terminal region. Suppose that the normal fate of the aminoterminal region is to interact with the carboxyterminal region in maintaining the functional structure; this is the case, for example, in cytochrome *c*. We can ask what happens to the aminoterminal region before the carboxyterminal region is made. Can it interact incorrectly with itself or with other molecules in the cell, and, if so, does this present a problem that the cell must combat? Similar questions can be raised about the other processes listed above: in each case, interactive surfaces that at one time are holding protein structures in their active conformations are at another time exposed to the intracellular environment containing high concentrations of many other interactive surfaces.

The self-assembly principle, if applied strictly, implies that all the interactions that take place when such protein surfaces are exposed are totally correct; this means that they are both necessary and sufficient to produce the normal functional conformation. The molecular chaperone concept challenges this view by proposing that in any given assembly process there is a certain probability that incorrect interactions will occur.

(*a*) *Definition of 'incorrect interactions'*

Incorrect interactions are defined as those that give rise to structures which are non-functional in their normal context, i.e. they do not possess the required biological activity.

The probability of incorrect interactions may be so low in some cases that molecular chaperones are not required, but in other cases, e.g. the assembly of nucleosomes and rubisco, the probability of incorrect interactions is so high that molecular chaperones are essential to produce enough functional structures for cellular needs.

According to this view, molecular chaperones are required because many cellular processes involving proteins carry an inherent risk of malfunction. They carry this risk because of the sheer number, variety and flexibility of the many weak interactions that hold proteins in their functional conformations. The cell thus continually faces the problem that incorrect interactions will produce non-functional structures. These ideas can be simplified into a unifying concept by supposing that all cells require a molecular chaperone function.

(*b*) *Definition of the molecular chaperone function*

The molecular chaperone function is defined as the prevention of incorrect interactions between transiently exposed surfaces by the binding of chaperone molecules to those surfaces.

4. MECHANISM OF ACTION OF MOLECULAR CHAPERONES

Present knowledge about the mechanism of action of molecular chaperones (derived principally from studies on nucleoplasmin, the chaperonins and the hsp 70 family) suggests that they function by inhibiting unproductive incorrect assembly pathways which would otherwise act as kinetic dead-end traps. This inhibition is exercised by the non-covalent binding of the molecular chaperone to exposed interactive surfaces to produce stable complexes. Some molecular chaperones, e.g. the chaperonins and the hsp 70 family, can bind to a wide variety of unassembled proteins that are unrelated in sequence, so an important aim of current research is to determine how interactive surfaces are recognized. In these complexes the bound ligands are prevented from interacting incorrectly. Reversal of binding then occurs under circumstances which favour correct interactions involving the ligand. In the case of the chaperonin family, these circumstances include the release of the bound ligand into cages formed at each end of the large oligomeric chaperonin molecules; each cage provides a sequestered environment in which a released polypeptide can fold according to the principle of self-assembly without running the risk of interacting incorrectly with other folding molecules (Saibil *et al.* 1993). In some cases (e.g. the chaperonins and the hsp 70 family), but not all (e.g. nucleoplasmin), this reversal of binding requires ATP hydrolysis. This requirement for energy is another feature by which the molecular chaperone view of protein assembly differs from the conventional view, because it suggests that energy in the form of ATP is often required to assemble proteins as well as to synthesize the peptide bonds.

This model proposed for the action of molecular chaperones suggests that their action is rather subtle, and can be described as assisting self-assembly. Thus the principle of self-assembly is not violated by the molecular chaperone concept, rather it is qualified by the proposal that in the *in vivo* situation self-assembly needs assistance from other protein molecules. On this basis we can distinguish two types of self-assembly.

1. Strict self-assembly: no macromolecules other than the primary structure are required for the polypeptide to have a high probability of assembling correctly within the intracellular environment. (This definition is an over-simplification in that it ignores the post-translational modifications required for some proteins to assemble correctly.)
2. Assisted self-assembly: an appropriate molecular chaperone is required in addition to the primary structure to allow correct assembly to predominate over incorrect assembly; such molecular chaperones convey no steric information over and above that in the primary structure of the ligand.

The ratio of strict self-assembly to assisted self-assembly *in vivo* is not known, but is likely to vary with the spectrum of protein assembly occurring at a given time. It may be that all protein assembly *in vivo* is assisted, because even proteins which self-assemble *in vitro* very rapidly to the active conformation with high yield after removal of denaturant (e.g. dihydrofolate reductase) will bind to a molecular chaperone if presented with one. Perhaps in such cases there is a potential problem of incorrect interactions *in vivo* which has been overlooked by studying the pure protein *in vitro*. The study of protein assembly *in vitro* using pure defined components has enormous analytical advantages, but it also suffers from the limitation that additional components involved *in vivo* may be lost during purification. There is thus a need to repeat all the extensive studies of protein renaturation *in vitro* in the presence of appropriate molecular chaperones.

5. CURRENT LIST OF MOLECULAR CHAPERONES

Table 1 contains a list of proteins that can be regarded as molecular chaperones. They are grouped into families on the basis of aminoacyl sequence, so that

Table 1. *List of molecular chaperones*

(Proteins suggested to be molecular chaperones are grouped on the basis of aminoacyl sequence, together with the cellular functions they are believed to assist. Note that some groups assist the assembly of many different proteins, whereas others are specific for one or a few proteins. The question marks indicate where no firm evidence is available. Hsp = heat shock protein.)

name	proposed roles
nucleoplasmins	nucleosome assembly and disassembly ribosome and ribonucleoprotein particle assembly? transcription?
chaperonins (includes hsp 60 and TCP1)	polypeptide folding polypeptide transport
hsp 70 (or DnaK in *E. coli*)	polypeptide folding polypeptide transport oligomer disassembly
hsp 90	masking of binding sites polypeptide folding?
DnaJ protein	interaction with hsp 70 and GrpE
GrpE protein	interaction with hsp 70
SecB protein	bacterial polypeptide transport
signal recognition particle	polypeptide transport
Pro-sequences	protease folding
ubiquitin tails of some ribosomal proteins	ribosome assembly in eukaryotes
PapD protein	bacterial pilus assembly
PrtM and PrsA	folding of secreted bacterial proteins
Lim protein	folding of bacterial lipase
Rb protein	binding of transcription factors
prions	rogue molecular chaperones?

members within each family are related to one another by sequence but not to members of another family. A theme emerging from recent research is that families of different molecular chaperones cooperate together in defined pathways to assist the assembly of some proteins (Langer *et al.* 1992). This list is predicted to grow as more researchers include the concept of molecular chaperones in their experimental programmes. The following articles discuss recent information about some of these fascinating protein molecules.

I am grateful to the Science and Engineering Research Council for support from 1972 to 1990.

REFERENCES

Anfinsen, C.B. 1973 Principles that govern the folding of protein chains. *Science, Wash.* **181**, 223–230.

Anson, M.L. 1945 Protein denaturation and the properties of protein groups. *Adv. Protein Chem.* **2**, 361–386.

Barraclough, R. & Ellis, R.J. 1980 Protein synthesis in chloroplasts. IX. Assembly of newly-synthesized large subunits into ribulose bisphosphate carboxylase in isolated intact chloroplasts. *Biochim. biophys. Acta* **608**, 19–31.

Caspar, D.L.D. & Klug, A. 1962 Physical principles in the construction of regular viruses. *Cold Spring Harb. Symp. quant. Biol.* **27**, 1–24.

Creighton, T.E. 1984 *The proteins*. New York: W. H. Freeman.

Ellis, R.J. 1987 Proteins as molecular chaperones. *Nature, Lond.* **328**, 378–379.

Ellis, R.J. 1990*a* Molecular chaperones: the plant connection. *Science, Wash.* **250**, 954–959.

Ellis, R.J. (ed.) 1990*b* *Seminars in cell biology*, vol. 1. (71 pages.) Philadelphia: Saunders Scientific.

Ellis, R.J. & Hemmingsen, S.M. 1989 Molecular chaperones: proteins essential for the biogenesis of some macromolecular structures. *Trends biochem. Sci.* **14**, 339–342.

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

Ellis, R.J., van der Vies, S.M. & Hemmingsen, S.M. 1989 The molecular chaperone concept. *Biochem. Soc. Symp.* **55**, 145–153.

Epstein, C.J., Goldberger, R.F. & Anfinsen, C.B. 1963 The genetic control of tertiary protein structure: studies with model systems. *Cold Spring Harb. Symp. quant. Biol.* **28**, 439–446.

Fraenkel-Conrat, H. & Williams, R.C. 1955 Reconstitution of active tobacco mosaic virus from its inactive protein and nucleic acid components. *Proc. natn. Acad. Sci. U.S.A.* **41**, 690–698.

Gething, M.-J. & Sambrook, J. 1992 Protein folding in the cell. *Nature, Lond.* **355**, 33–45.

Hartl, F.U., Martin, J. & Neupert, W. 1992 Protein folding in the cell: the role of molecular chaperones hsp 70 and hsp 60. *A. Rev. biophys. biomol. Struct.* **21**, 293–322.

Hemmingsen, S.M., Woolford, C., van der Vies, S.M., Tilly, K., Dennis, D.T., Georgopoulos, C., Hendrix, R.W. & Ellis, R.J. 1988 Homologous plant and bacterial proteins chaperone oligomeric protein assembly. *Nature, Lond.* **333**, 330–334.

Langer, T., Lu, C., Echols, H., Flanagan, J., Hayer, M.K. & Hartl, F.U. 1992 Successive action of DnaK, DnaJ and GroEL along the pathway of chaperone-mediated protein folding. *Nature, Lond.* **356**, 683–689.

Laskey, R.A., Honda, B.M., Mills, A.D. & Finch, J.T. 1978 Nucleosomes are assembled by an acidic protein which binds histones and transfers them to DNA. *Nature, Lond.* **275**, 416–420.

Lorimer, G. 1992 Role of accessory proteins in protein folding. *Curr. Opin. struct. Biol.* **2**, 26–34.

Musgrove, J.E. & Ellis, R.J. 1986 The rubisco large subunit binding protein. *Phil. Trans. R. Soc. Lond.* B **313**, 419–428.

Pelham, H.R.B. 1986 Speculations on the functions of the major heat shock and glucose-regulated proteins. Cell **46**, 956–961.

Rothman, J.E. 1989 Polypeptide chain binding proteins: catalysts of protein folding and related processes in cells. *Cell* **59**, 591–601.

Saibil, H., Zheng, D., Roseman, A.M., Hunter, A.S., Watson, G.M.F., Chen, S., auf der Mauer, A., O'Hara, B.P., Wood, S.P., Mann, N.H., Barnett, L.K. & Ellis, R.J. 1993 Large ATP-induced conformational changes in a cage-like chaperonin structure. (Submitted.)

Silen, J.L. & Agard, D.A. (1989) The α-lytic protease pro-region does not require a physical linkage to activate the protease domain *in vivo*. *Nature, Lond.* **341**, 462–464.