

Proline-dependent oligomerization with arm exchange

Marc Bergdoll¹, Marie-Hélène Remy², Christine Cagnon³,
Jean-Michel Masson^{2,3} and Philippe Dumas^{4*}

Background: Oligomerization is often necessary for protein activity or regulation and its efficiency is fundamental for the cell. The quaternary structure of a large number of oligomers consists of protomers tightly anchored to each other by exchanged arms or swapped domains. However, nothing is known about how the arms can be kept in a favourable conformation before such an oligomerization.

Results: Upon examination of such quaternary structures, we observe an extremely frequent occurrence of proline residues at the point where the arm leaves the protomer. Sequence alignment and site-directed mutagenesis confirm the importance of these prolines. The conservation of these residues at the hinge regions can be explained by the constraints that they impose on polypeptide conformation and dynamics: by rigidifying the mainchain, prolines favour extended conformations of arms thus favouring oligomerization, and may prevent interaction of the arms with the core of the protomer.

Conclusions: Hinge prolines can be considered as 'quaternary structure helpers'. The presence of a proline should be considered when searching for a determinant of oligomerization with arm exchange and could be used to engineer synthetic oligomers or to displace a monomers to oligomers equilibrium by mutation of this proline residue.

Addresses: ¹Laboratoire de Biologie Structurale, Institut de Génétique et de Biologie Moléculaire et Cellulaire du CNRS, 1 rue Laurent Fries, B.P. 163, 67404 Illkirch Cedex, France, ²Institut de Pharmacologie et Biologie Structurale du CNRS, 205 route de Narbonne, 31077 Toulouse Cedex, France, ³Institut National des Sciences Appliquées, 205 route de Narbonne, 31077 Toulouse Cedex, France and ⁴UPR 9002, Institut de Biologie Moléculaire et Cellulaire du CNRS, 15 rue René Descartes, 67084 Strasbourg Cedex, France.

*Corresponding author.
E-mail: dumas@ibmc.u-strasbg.fr

Key words: arm exchange, folding, oligomerization, prolines

Received: 11 December 1996

Accepted: 6 January 1997

Electronic identifier: 0969-2126-005-00391

Structure 15 March 1997, 5:391–401

© Current Biology Ltd ISSN 0969-2126

Introduction

Many proteins in the cell are known to be found and to function as a noncovalent association of two or more polypeptide chains. The average oligomeric state in *Escherichia coli* has been estimated to be four [1], but this number hides the broad diversity in size, composition and arrangement that can be found, from homodimers, such as prealbumin, to heteromultimers, such as the pyruvate dehydrogenase complex, ribosomes and virus shells. The oligomer interface must be extensive to ensure specificity and strength irrespective of whether these chains are identical, similar or completely different, whether the interactions are isologous or heterologous, equivalent or only quasi-equivalent. On the other hand, the interaction between these polypeptides often needs to be versatile for the functionality of the oligomer. A very efficient way to increase the contact area whilst maintaining flexibility is for the protomers to 'hug' each other [2] in a process involving arm exchange, as in dogfish lactate dehydrogenase [3], or even by domain swapping, as in diphtheria toxin (DT) [4]; for a review on domain swapping, see Bennett *et al.* [5]. In these cases, oligomerization is related to the folding process, insofar as a protomer should find its partner(s) before the arm or the domain folds back onto its own molecule and hinders correct association. Chaperones have been invoked as possible participants in this process

[6], but the increasing number of quaternary structures for which X-ray data is available suggests that there is another way to promote the subunit association.

In particular, our solution of the three-dimensional (3D) X-ray structure of the bleomycin-resistance protein (BRP), strongly suggests that a specific proline residue at position 9, which is conserved in every known BRP sequence, is essential for the promotion of dimerization by arm exchange [7]. Bacterial resistance to the antibiotic bleomycin is achieved by sequestration of the antibiotic [8] and before the structure of BRP was solved, Pro9 was suspected to be part of the bleomycin binding site. This proline residue had been mutated and replaced by various amino acids with noticeable, although unequal, loss of activity [7]. The behaviour of these mutants was perplexing because no obvious correlation could be found between the level of bleomycin resistance and any simple property of the substituted sidechain (size, charge, polarity or hydrophobicity), or combinations thereof. BRP in solution is a dimer and the 3D X-ray structure shows a very tight interaction of the protomers through arm exchange [7]. The arm consists of eight N-terminal amino acids and Pro9 is right at the hinge between the two molecules, where each mainchain emerges from its own molecule to direct itself towards the other (Fig. 1a). The structure

also shows that two bleomycin molecules can fit into two crevices at the interface between the protomers, implying that dimerization is an absolute necessity for resistance against the antibiotic. The absence of interactions between Pro9 and bleomycin prompted us to state that the proline is not involved in the binding of the antibiotic, but rather in BRP dimerization, and that such proline-mediated arm exchange might not be a property unique to BRP [7].

Here, by considering several examples of unrelated proteins, we show that proline-induced arm exchange is a commonly used oligomerization mechanism. We also check that the BRP Pro9Gly mutant does not form active dimers. Surprisingly, this mutation has an effect not only on arm exchange, but also on the folding of the core of the molecule.

Results

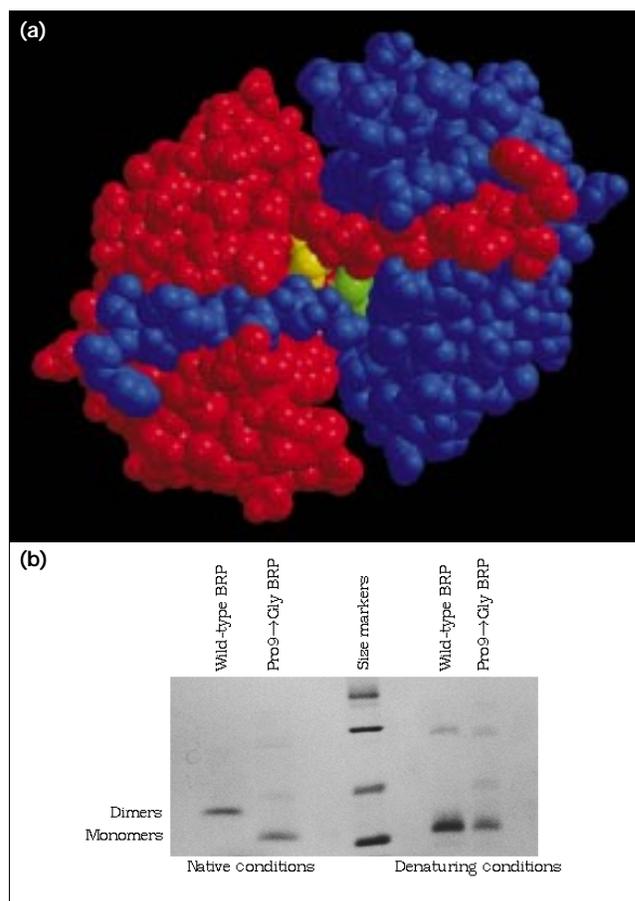
Examples of hinge prolines

The importance of Pro9 in the hinge of BRP led us to examine other 3D structures of oligomers with exchanged arms or swapped domains, and we found many protomers with at least one proline in the hinge region.

The most representative examples found are collected in Table 1; this list reflects the diversity in sizes, structures, functions and oligomer compositions that are encountered. The number of protomers involved is variable, ranging from 2 in BRP to 360 in simian virus 40 (SV40) [6]. Many examples are homomultimers, but this is not mandatory as shown by the hirudin–thrombin complex (HTC; Fig. 2) [9] or methylamine dehydrogenase (MTA), an $(\alpha\beta)_2$ oligomer with the arm linking the α subunit of one dimer to the β subunit of the other (Fig. 3) [10]. In some cases, the arm is at the N terminus, as in BRP, and in others at the C terminus, as in deoxyuridine triphosphatase (dUTPase; Fig. 4a) [11]. The exchanged arm can be short as in BRP and tumor necrosis factor (TNF) [12], which have arms of nine and twelve amino acids respectively, but the exchanged arm can also be very long as in beef liver catalase (BLC) [13], in which it is composed of not less than 69 residues (Fig. 5).

In order to exclude from our data any protein that has a proline at the base of an exchanged arm without any functional or structural reason, for each example we checked whether the proline is conserved throughout related sequences and how variable the surrounding sequence is. For example, three citrate synthase (CS) structures have been solved, from chicken heart (CHCS) [14], pig heart (PHCS) [14] and *Thermoplasma acidophilum* (TACS) [15], all of which are dimers with exchanged arms and a proline residue at the hinge. The sequences of 17 CSs have been determined and are known to be very different: the most distantly related, CHCS and TACS, share only 20% sequence similarity. The hinge proline is present in every

Figure 1



Bleomycin resistance protein (BRP). (a) CPK view of the BRP dimer as seen along the twofold axis. The individual protomers are coloured in red and blue, and the Pro9 residues at the base of the arms are highlighted in yellow and green, respectively. Note the short distance from the prolines to the twofold axis and the close proximity of the binding sites of the two arms. (b) Gel electrophoresis in native and denaturing conditions showing that BRP Pro9Gly is monomeric: in native conditions, the Pro9Gly mutant migrates at a much smaller molecular weight than the wild-type BRP, which is known to be dimeric, while in denaturing conditions they co-migrate. The faint bands that appear in some lanes correspond to high molecular weight contaminants.

sequence despite this low sequence homology and the variation in arm length from 10 to 20 amino acids. Hence citrate synthase has been included in Table 1. Ricin, on the other hand, was disregarded, because although an arm with a proline at its base links chain B to chain A, there is no sequence information or other evidence enabling us to state that, for ricin, the proline residue is important in the arm exchange.

Mutation of the hinge proline

The role assigned to hinge prolines in the arm exchange process is based not only on its frequent occurrence in unrelated proteins and its strong conservation through

Table 1

Outstanding examples of oligomers with exchanged arms and a proline residue at the hinge.

Protein name	Short name	PDB id	Ref.	Figure	Oligomeric state	Proline	Arm position
Bleomycin resistance protein	BRP		[7]	1a	$\alpha 2$	9	N-term
Bovine seminale ribonuclease A	BS-RNaseA	1BSR	[17]	6a	$\alpha 2$	19	N-term
Chicken heart citrate synthase	CHCS	5CSC	[14]		$\alpha 2$	418, 422	C-term
Aspartate aminotransferase	AAT	1TAT	[29]	8	$\alpha 2$	13, 14, 16	N-term
Diphtheria toxin	DT	1DDT	[4]		$\alpha 2$	378, 382	C-term
Hirudin thrombin complex	HTC	4HTC	[9]	2	$\alpha \beta$	46	C-term
Deoxyuridine triphosphatase	dUTPase	1DUP	[11]	4a	$\alpha 3$	124	C-term
Tumor necrosis factor	TNF	1TNF	[12]		$\alpha 3$	12	N-term
Beef liver catalase	BLC	8CAT	[13]	5	$(\alpha 2)_2$	69	N-term
Methylamine dehydrogenase	MTA	2MTA	[10]	3	$(\alpha \beta)_2$	27	N-term
CDC2 kinase subunit	CKS	1CKS	[26]	7	$(\alpha 2)_3$	62, 64	C-term
Tomato bushy stunt virus	TBSV	2TBV	[34]	9	$(\alpha 3)_60$	81	N-term
Cowpea chlorotic mosaic virus	CCMV		[35]		$(\alpha 3)_60$	35, 180	N- & C-term
Simian virus 40	SV40		[6]		$(\alpha 5)_72$	298, 300	C-term

Columns 1, 2 and 3 give respectively the full name of the molecule, the short name used throughout the text and the index under which it can be found in the Brookhaven Data Bank. Columns 4 and 5 point to a

paper and an illustration describing the structure. Columns 6 to 8 characterize the type of the oligomer, the proline(s) concerned and the position of the arm.

related ones, but also on three examples, in which hinge prolines are mutated either naturally or by site-directed mutagenesis, with clear effects on oligomerization.

RNase A

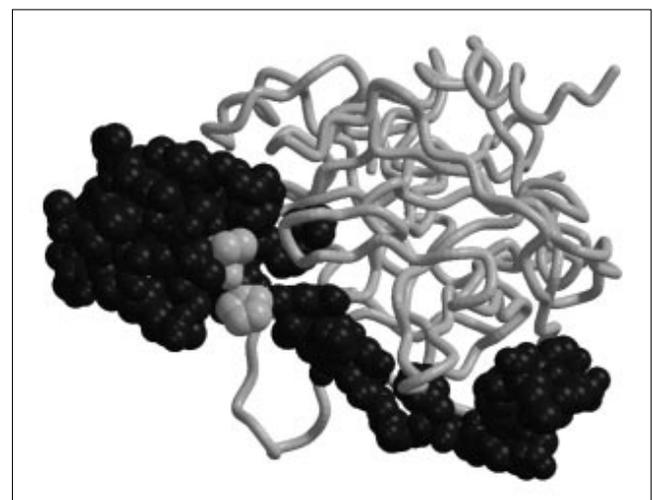
Two structures of bovine ribonuclease A (RNaseA) have been solved at high resolution, bovine pancreatic RNaseA (BP-RNaseA) [16] and bovine seminal RNaseA (BS-RNaseA) [17]. They share more than 80% identity and almost 90% homology and, as expected, their structures are very similar, except that BS-RNaseA is dimeric (Fig. 6a) whereas BP-RNaseA is monomeric (Fig. 6b). The major sequence differences at the interface that can explain this dissimilarity are the presence of a proline (Pro19) and of two cysteines (Cys31, Cys32) in the dimeric BS-RNaseA. The dimerization could be thought to result from the two disulphide bridges (Cys31–Cys32 and Cys32–Cys31) which covalently link two BS-RNaseA protomers. However, their selective reduction results in the dissociation of only one third of the dimers, demonstrating that dimerization does not depend solely on them, but also on an arm exchange occurring in two thirds of the population [18]. Site-directed mutagenesis experiments, which transformed the monomeric BP-RNaseA into dimers by introducing various combinations of a proline, a leucine and two cysteines at position 19, 28, 31 and 32 respectively, led the authors to assign the role of a ‘swapping determinant’ to Pro19, as they observed that the number of exchanged arms is higher whenever a proline is present at position 19 [19].

dUTPase

There are 26 different known sequences of deoxyuridine triphosphatase (dUTPase) and two structures have been

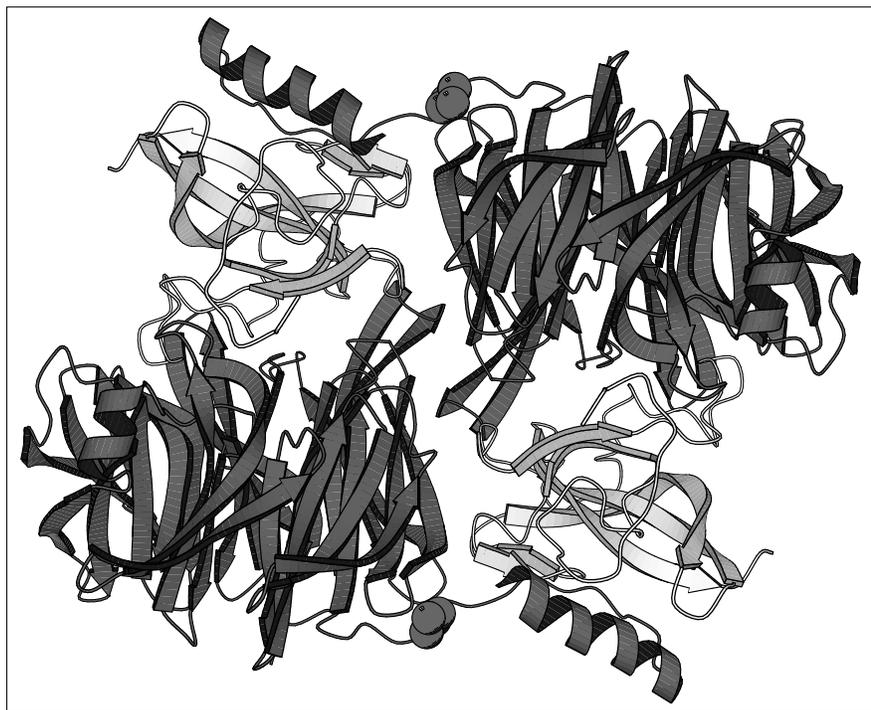
solved so far: one from *E. coli* [11] and one from human [20]. Both assemble as trimers with a C terminus arm exchange, after Pro124 in *E. coli* dUTPase (Fig. 4a) and Pro128 in human dUTPase. These prolines are framed by two sequence motifs, which are conserved in every known sequence [21]. If we consider the segment between these motifs in the sequence alignment of all dUTPases, we can define two classes (Fig. 4b). The first class, class 1, is composed of sequences with a very long insertion. The second class has a shorter segment and a proline conserved at

Figure 2



Representation of the hirudin–thrombin complex showing the thrombin backbone in light grey and the hirudin molecule as CPK atoms. Hirudin is shown in dark grey, but the two prolines at the base of its arm are highlighted in light grey.

Figure 3



Ribbon diagram of a methylamine dehydrogenase heterotetramer from *Paracoccus denitrificans* as seen along the twofold axis. Light and heavy chains are in light and dark grey, respectively. Note that the arms emanate from the heavy subunits of one protomer and interact with the light chains of the other protomer. Atoms of the hinge proline at the base of these arms are represented as CPK spheres.

position 124 as in DUT_ECOLI (subclass 2a), or at position 128 as in DUT_HUMAN (subclass 2b). Sequences from class 1 show no significant conservation of any proline. This classification, which is based on sequence considerations and the presence of a proline residue also correlates with the clustering of dUTPases based on their oligomerization number. It is known that dUTPases from herpesviruses (class 1) are monomeric [21], whereas *E. coli* and human dUTPase (subclasses 2a and 2b) are trimeric [11] [22]. The three sequences from subclass 2c (EIAV, DUT_CANAL and DUT_YEAST) are atypical; they have a short sequence between the two motifs but do not have proline residues at either position 124 or 128. No definite conclusion can be drawn about them without further information.

BRP

We know that the mutation of Pro9 in the BRP sequence results, *in vivo*, in loss of resistance to bleomycin [7]. In order to check our hypothesis that this is due to improper dimerization and to try to characterize the role of the hinge proline, a plasmid harbouring the Pro9Gly mutation has been constructed and the protein isolated. Transfected cells are not resistant to bleomycin (results not shown) and gel studies in native conditions clearly show that the mutant protein migrates as a monomer whereas the wild-type protein migrates as a dimer, confirming that resistance loss is caused by improper dimerization (Fig. 1b). This is due not only to incorrect positioning of

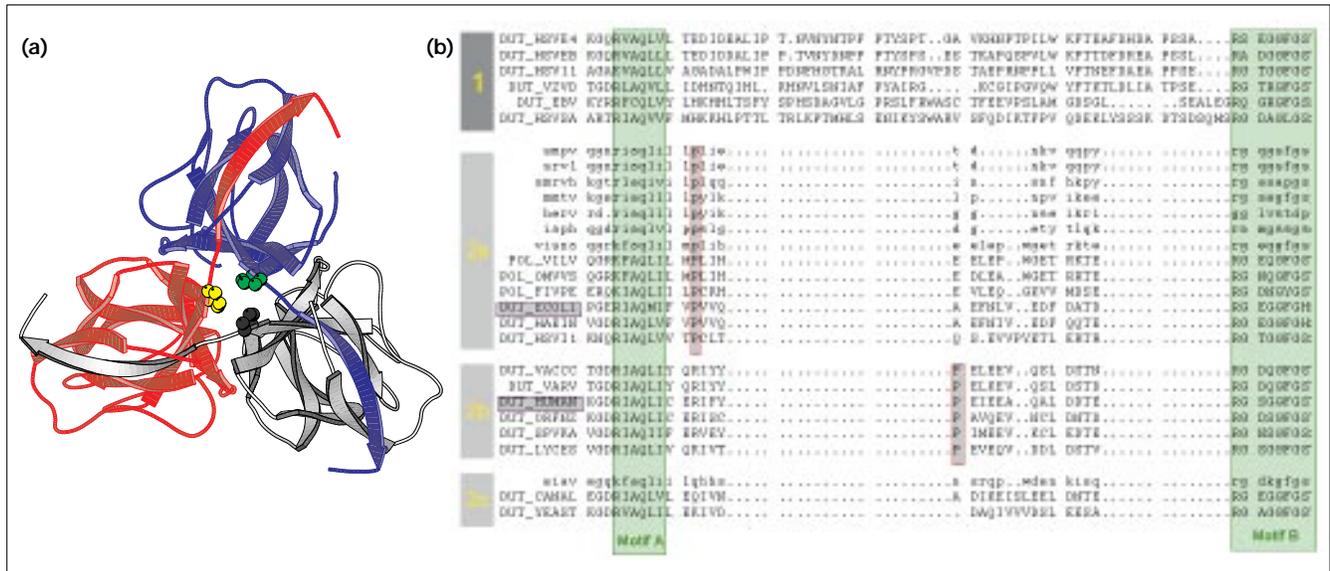
the arm as we originally thought, but also to incorrect overall folding. Far UV circular dichroism (CD) revealed two very different secondary structures. The CD spectrum of the wild type indicates that 30% of the structure is α -helical, as expected from the X-ray structure, whereas the CD spectrum of the mutant shows a complete loss of these α helices. This incorrect fold probably results from unfavorable interactions of the arm with the core of the molecule during the folding process.

Discussion

Properties of proline

Proline differs from other amino acids by the fact that the sidechain is bonded covalently to the nitrogen of the peptide group, thus inducing unique constraints on the peptide chain. First, proline lacks one hydrogen bond possibility as it has no amide hydrogen. This characteristic prevents proline from being engaged in regular secondary structures and accounts for its well known role in disruption of secondary structure. Second, proline can adopt a *cis* conformation more easily than any other amino acid, either permanently, in the ultimately folded structure, or transiently, during the folding process [23]. None of the prolines in the hinges of the above structures is seen in a *cis* conformation, but the possibility that this conformation exists during the folding process cannot be ruled out. Third, prolines have a fixed dihedral ϕ angle and, despite the fact that they have an additional degree of freedom around ω (*cis* conformation), their mainchain atoms have

Figure 4



Deoxyuridine triphosphatase (dUTPase). (a) Ribbon view of *E. coli* dUTPase as seen along the threefold axis. The three protomers are coloured red, purple and white. The hinge prolines, shown as small spheres are yellow, green and grey respectively. (b) C terminus sequence alignment of dUTPases. The sequences corresponding to the two solved structures, called DUT_ECOLI and DUT_HUMAN, are boxed. The prolines involved in arm exchange (Pro124 and Pro128,

respectively) as well as equivalent prolines in other sequences are highlighted by red boxes. The leftmost column clusters the various sequences as explained in the text. Motif A and B, marked by green boxes, are two conserved sequence motifs present in every dUTPase which are used to validate the alignment despite very long insertions in sequences from class 1. Sequences in small letters were previously known as pseudo-proteases and are taken from [21].

a much narrower conformational space. Moreover, proline is the only amino acid that restricts the conformational space of the preceding residue in the chain: the conformational entropy of alanine, for example, is decreased three-fold when it is followed by a *trans*-proline and more than ten-fold when it is followed by a *cis*-proline [24]. As a consequence, proline residues restrict the permitted regions in the Ramachandran plot and introduce rigidity to the mainchain over two consecutive residues.

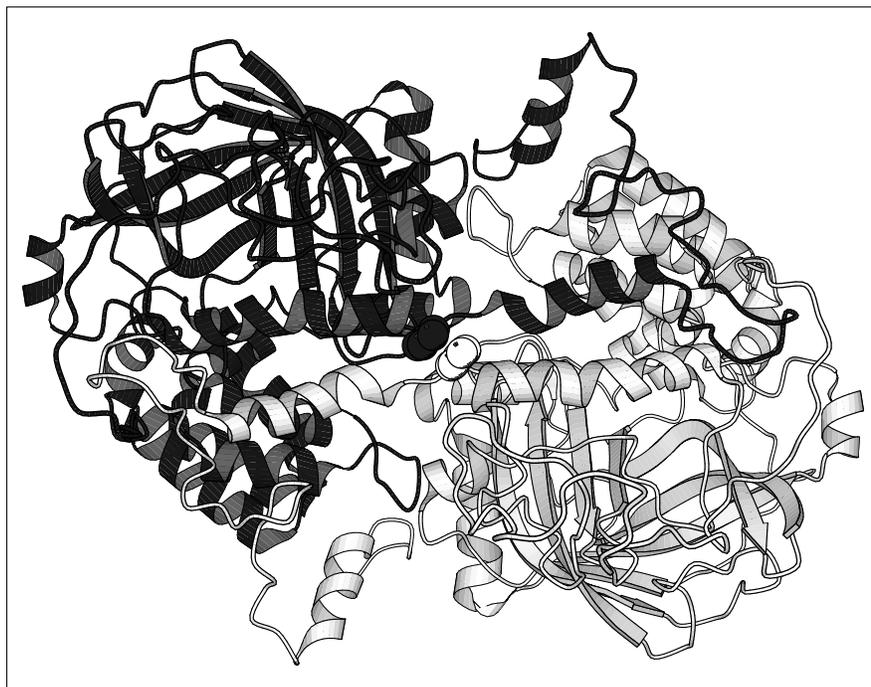
Oligomerization and folding

Oligomerization with arm exchange or domain swap is related to aspects of the folding process. The problem of folding the core of a molecule while keeping an arm in a conformation favourable for interaction with another molecule is very similar to the problem of folding two structural domains of a single peptide chain without deleterious interference between them. Prolines may favour conformations suitable for oligomerization. A first example of this is given by the comparison of the crystallographic structure of the thrombin–hirudin complex [9] with the NMR structure of hirudin alone [25]. The structure of the complex shows hirudin bound in a cleft on the surface of thrombin as a globular core with a long protruding arm (Fig. 2). The structure of the isolated hirudin indicates that the segment from Pro46 to His51, at the base of the arm, adopts a conformation pointing towards the solvent,

demonstrating that the shape of hirudin in the complex is not the result of binding to thrombin. Another example of an arm pointing away from the core of the molecule is p13^{suc1}, a cdc2 kinase subunit (CKS). In solution, p13^{suc1} from *Schizosaccharomyces pombe* is found as a mixture of monomers and dimers [26]. Two different conformations have been crystallized: one globular without arm exchange [27] and the other with arm exchange (Fig. 7) [26]. Experiments have shown that, in solution, the monomers are nonglobular, presumably with a long, nonfolded C-terminal segment starting after a Pro-Glu-Pro segment, which is conserved in every CKS [28].

BRP is a somewhat more difficult case as no experimental evidence is available of an extended conformation of the arm in monomeric BRP. However, an indirect indication of the conformation can be deduced from the fact that the arm interferes with the folding of the core resulting in a different structure when Pro9 is replaced by a glycine. It is not yet known whether this happens when Pro9 is replaced by other amino acids. The fact that some resistance to bleomycin can be observed, *in vivo*, with other Pro9 substitution mutants, indicates that certain dimers do assemble properly, and suggests that the behaviour of the arm depends on its hinge sequence with the efficiency of the folding-dimerization process being maximal with a proline, minimal with a glycine and intermediate with other amino acids.

Figure 5



Ribbon diagram of beef liver catalase. Note the very short distance separating the conserved proline at the base of the arms (highlighted by CPK spheres) and the considerable length of these exchanged arms. Five catalase structures are known from: beef liver [13], *Penicillium vitale* [45] (PVC), *Proteus mirabilis* [46], *Micrococcus lysodeikticus* [47] and *E. coli* [48]. They all form $(\alpha_2)_2$ tetramers made of two pairs of arm exchanging dimers and they all have a proline at the base of the exchanged arm except PVC. It should be noted that the sequence of PVC has not been determined chemically, but from the electron-density map, and we think that this single exception is due to a sequencing error (this proline is absolutely conserved in the other 34 known sequences).

Hinge diversity

The hinge regions — the segment of the arm between the originating and the target molecules — considered in Table 1 are very different from each other. They vary in length from one single amino acid in BRP to eight in DT [5]. There is no obvious bias towards prolines being found in loops or at the end of secondary structure elements (α helices or β strands). The chain can be bound tightly or loosely in the vicinity of the proline and no conformation common to all the examples could be detected. In BRP (Fig. 1a), Pro9 is located at the end of a β strand within a sheet from one protomer and at the beginning of another β strand within a sheet from the second protomer. RNaseA, on the other hand, has a much less constrained hinge (Fig. 6a), Pro19 being located in a coiled region with no precise secondary structure. Finally, in hirudin (Fig. 2), the two prolines at the base of the arm are in an almost perfect polyproline II conformation.

The position of proline(s) in the hinge also varies: in oligomers, such as MTA (Fig. 3), the hinge proline interacts principally with the originating molecule, but in some other examples, such as BS-RNaseA, it makes stronger contacts with the opposite protomer (Fig. 6a). Even the number of prolines in the hinge is variable. One single proline residue is most frequently found and often seems enough to favour arm exchange, but a few examples exist showing pairs or triplets of prolines. SV40 and CKS, for example, have a conserved Pro-X-Pro sequence, whereas

some aspartate aminotransferase molecules (AAT) [29] show a Pro-Pro-X-Pro sequence (Fig. 8).

At the sequence level, no motif could be found in the various hinge regions examined but an important observation is that Gly-Pro pairs are rarely found. This could be because the conformational flexibility of glycine would contravene the mainchain rigidification induced by proline. In fact, the only occurrences of Gly-Pro were found in some AAT hinges that have three proline residues in close proximity: Gly-Pro-Pro-X-Pro. It can be supposed that three prolines are more than sufficient to rigidify the arm and that the presence of a glycine is only a minor hindrance in this case.

Distance between the exchanged arms

Arm exchange can take place close to a symmetry axis, as in BLC, or far from it, as in AAT, and the importance of hinge prolines may not be equal in these two situations. The role of prolines is probably more important when the arm exchange occurs in the vicinity of a symmetry axis, as in BLC (Fig. 5). In these oligomers, correct and incorrect anchoring sites are close to each other, and, if the arm folds back, it can make contacts as strong and as specific with the binding site on its own molecule as it would with the equivalent binding site on the opposite protomer. This would result in almost irreversible 'misfolding' events and oligomerization failure. The role of prolines may be less important in cases of arm exchange between

Figure 6



Ribonuclease A. (a) Ribbon representation of bovine seminal RNaseA showing the dimer with exchanged arms. Pro19 is represented in CPK mode. Of the BS-RNaseA dimers, two thirds show such an N-terminal arm exchange whereas one third of the dimers have the arm folded

back on the molecule in a position similar to the one occupied in BP-RNaseA. (b) BP-RNaseA in its monomeric and compact form. Note that the contacts made by the N-terminal arm, either on the molecule itself in BP-RNaseA or on the other protomer in BS-RNaseA, are equivalent.

homomultimers with anchoring sites far from the symmetry axis as in AAT (Fig. 8), or between heteromultimers that have only one anchoring site, as for HTC (Fig. 2). If the arm folds back, it cannot reach the binding site on the molecule from which it emerges, and makes nonspecific and weaker contacts than those it would make with the binding site on the related protomer. Energetically, this incorrect fold is thus less favourable than the correct one and the presence of a proline, even if useful, may be less essential in these cases than in cases of arms exchanged close to a symmetry axis. This could explain the strange pattern of conserved or absent hinge prolines in the alignment of AAT sequences: as the arm exchange is less dependent on the presence of prolines, these are less conserved and when they are present, their position in the sequence can vary.

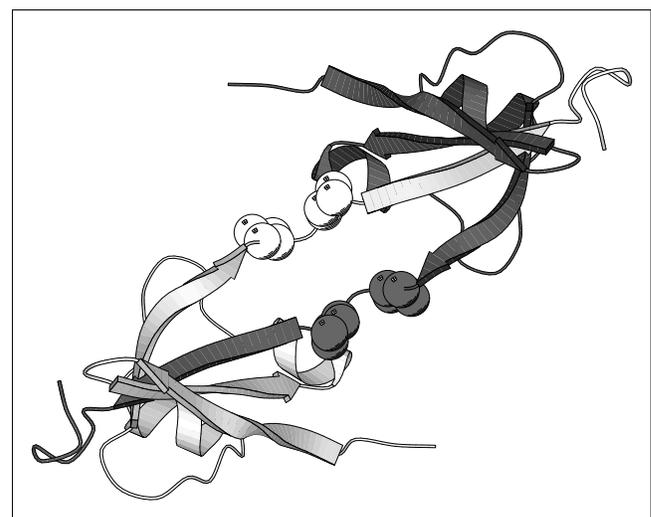
Alternate conformations

Some proteins which oligomerize with arm exchange can be found with different arm conformations resulting in variation of the oligomeric state. These different conformations can be stable or metastable. When two stable conformations coexist, one of them is usually not functional and must be considered as misfolded. On the other hand, metastable oligomeric states are often functional, and the coexistence of two populations is the manifestation of a normal activation or regulation mechanism. It has been suggested that domain swapping can serve as a mechanism for functional interconversion between monomers and oligomers [5].

DT can exist as dimers with swapped domains and globular monomers, of which only the monomeric form is toxic

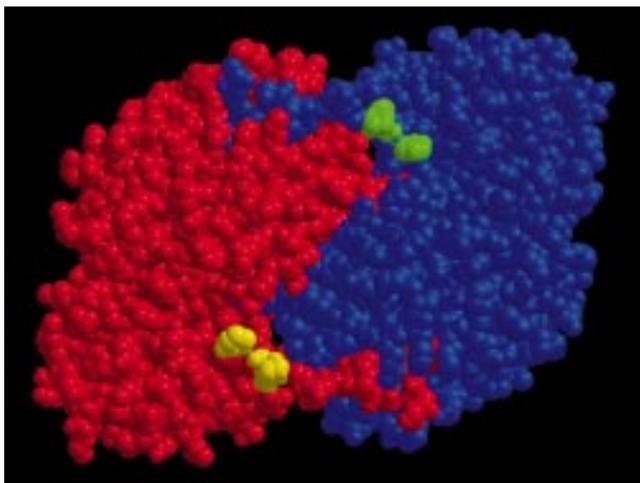
[4,30]. The chain segment which changes conformation between monomers and dimers spans Pro378 to Gln387. According to Bennett and Eisenberg, Lys385 and Thr386 have to be considered as the molecular hinge, on the basis of the change in mainchain torsion angles [30], but Pro382 should also be considered because the value of its ψ angle changes by more than 170° . A second example of conformational change of the arm is given by RNaseA. As mentioned previously, BS-RNaseA forms dimers of which

Figure 7



Ribbon representation of *S. pombe* p13^{suc1} dimer. The two hinge prolines conserved in every known sequence are represented by CPK spheres.

Figure 8



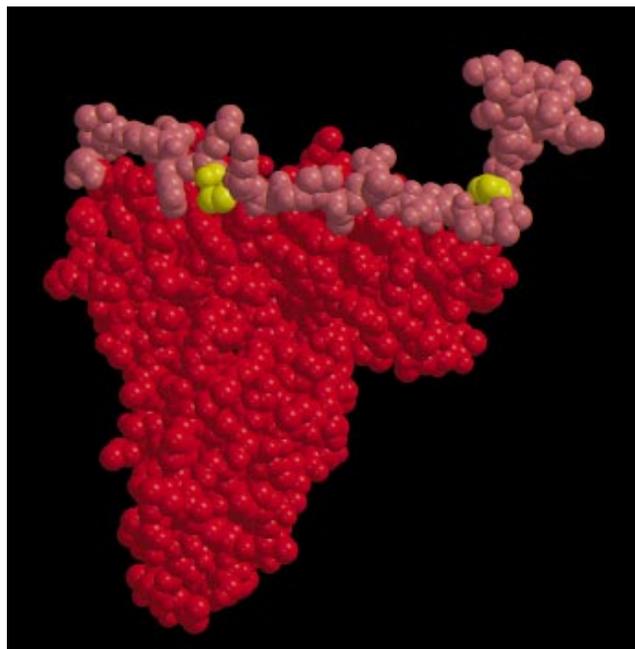
CPK representation of chicken heart mitochondria aspartate amino transferase (AAT) dimer seen along the twofold axis. Note the long distance separating the two arms and consequently the impossibility of one arm reaching the binding site on its own protomer. The structures of AAT from different origins have been determined: *E. coli* [49], chicken heart cytosol [50] and chicken heart mitochondria [29]. They are all dimeric and exchange an N-terminal arm with at least one proline at the hinge. Alignment of the 35 known AAT sequences shows less than 50% homology between the sequences and the length of the exchanged arms is from 10 to 60 amino acids. Prolines found in chicken heart mitochondria AAT are strongly, though not absolutely, conserved.

two thirds exchange arms and one third does not. The two isoforms are in equilibrium, and when separated, the two pure populations tend towards the same 2:1 ratio of exchanged/unexchanged arms. Equilibration time depends on factors such as pH and temperature, and is considerably extended at temperatures below 4°C [18]. Switching between the two conformations of the arm has no effect on the oligomeric state as the cores of the dimers are held together by S-S bridges, but such a mechanism could be important when protomers are not covalently linked.

Inferences

We shall now discuss the relevance of proline-induced arm exchange to three cases of oligomerization. The first case concerns the Na⁺-K⁺-ATPase. Dimerization of this molecule is reduced by more than 90% by the mutation of Pro244, a conserved proline close to the N terminus of the β subunit [31]. This behaviour leads us to propose that the dimerization process could involve an arm exchange which is hindered by the mutation. The second example is the RNaseA family, from which BS-RNaseA is the only dimeric RNaseA characterized so far. Dimerization is not only due to intermolecular S-S bridges but also to arm exchange [18], and Pro19 has been shown to be a major 'swapping determinant' [19,32]. Among the numerous RNasesA that have been sequenced, some possess a

Figure 9



CPK representation of the C subunit of tomato bushy stunt virus (TBSV) showing its major difference from the A and B subunits. The part common to all the subunits is shown in red. The portion of the chain which is ordered and visible in C, but disordered and thus invisible in A and B is shown in pink. Two prolines are highlighted in yellow: Pro81 and Pro93. The rightmost one, Pro81, conserved through the 13 known sequences, is located exactly where the arms emerge from the C subunit core and start forming hooks. These hooks link the molecules tightly together at the threefold symmetry axis and are essential for proper assembly of the shell [34]. Experiments on turnip crinkle virus, a member of the TBSV family, have shown that when the hooks are deleted, the particles assemble into incorrect RNA-free T = 1 capsids [51,52]. Pro93, situated on the left, close to the start of the pink segment, is proposed to be potentially responsible for quasi-equivalence regulation through disordering of the pink arm.

proline at position 19 and could exist as dimers. It is possible that these dimers have remained undetected because of an unfavourable monomer/dimer equilibrium due to the absence of Cys31 and Cys32. A possible indication of the existence of other RNaseA dimers can be found in the report of RNaseA purification from guinea pig (*Cavia porcellus*) [33], which has two types of RNaseA molecule sharing 90% sequence identity, with one having a proline at position 19. When they are separated by ion exchange chromatography, the enzyme without Pro19 elutes in one single peak, termed A, whereas the enzyme with Pro19 elutes in two different peaks, B and C. Peak C is not fractionated by repeated chromatography, but its content can be transformed by acid treatment and dialysis into the form eluted in peak B [33]. Our explanation of this behaviour is that these two peaks could correspond to two different oligomeric states of the enzyme: monomers in peak B and dimers in peak C.

Third, we would like to examine the assembly of viral shells. Viruses make intensive use of exchanged arms because these allow very extensive contacts without increasing the thickness of the capsid and also because they are flexible enough to accommodate quasi-equivalent contacts. Proline residues, which favour arm exchange, are common at the base of viral arms as shown by tomato bushy stunt virus (TBSV) [34], SV40 [6] and cowpea chlorotic mosaic virus (CCMV) [35]. Prolines could have been selected not only because they favour arm exchange, but also because they can regulate it. Viruses are molecular entities in which the type of contacts, as well as the ratio between different conformers, depends on the geometry of the shell and must be carefully controlled for proper assembly. To respect quasi-equivalence requirements, subunits of viral shells must adopt slightly different conformations. Some viruses achieve this by packing subunits that have different sequences or different length. TBSV is a virus with T=3 shell composed of 180 chemically, but not structurally, equivalent polypeptide chains: the packed chains are identical and uncleaved, but only one of the three protomers has an ordered N-terminal fragment hooked together around the threefold axis of the shell (Fig. 9). We suggest that, in TBSV, the conserved proline Pro93, close to the amino acid where the chain becomes disordered, could play an important role for virus assembly by controlling the arm position. This proline residue could act as the molecular switch required to regulate quasi-symmetry by keeping the ratio of alternate conformations close to the ideal value of two thirds which is necessary for correct virus shell association. Interestingly, this value of two thirds for the ratio of disordered/ordered fragments is identical to the ratio of exchanged/unexchanged arms seen in BS-RNaseA. Although the structures of the coat proteins are different, a similar triggering mechanism has been proposed by Valegard *et al.* [36] for the RNA bacteriophage MS2 for which a loop involved in quasi-equivalent contacts has two conformations triggered by a *cis/trans* isomerization of a proline residue. Replacing this proline results in a shell with decreased heat stability and Stonehouse *et al.* do not exclude the possibility that the capsid assembly is less efficient in the mutant than in the wild-type virus [37].

Conclusion

The dimerization mechanism with arm exchange and a proline in the hinge as observed in BRP, is not fortuitous as shown by three facts: firstly the occurrence, in several unrelated protein structures, of a proline in the arm exchanged between protomers; secondly, the strong conservation of this proline through sequences of related proteins in each of the examples; and thirdly, the dramatic effect of its mutation in a few known mutants (natural or site-directed). The examples found are very different and no simple classification scheme has so far been established, either at the arm or at the hinge level. No consensus

sequence other than the presence of a conserved proline and the absence of Gly-Pro pairs could be found. Despite this variability, a role for hinge prolines as ‘quaternary structure helpers’ seems clear: because of their constrained conformations, they reduce interference between the arm and the core of the protomer and tend to keep the arm in a position favourable for oligomerization.

Some aspects of this proline-induced arm exchange mechanism call for further studies. First, *in vivo* mutants of BRP seem to indicate that the nature of the amino acid that replaces Pro9 has an influence on the efficiency of the dimerization. It is our goal to check this aspect more precisely by site-directed mutagenesis and to determine the percentage of surviving dimers for various amino acids. Second, it could be worth investigating whether *cis* isomers of hinge prolines exist transiently during the folding/oligomerization process. These studies could help our understanding not only of oligomerization but also of the folding process because oligomerization with arm exchange is intimately linked with folding. Understanding the role of hinge prolines could thus help us to understand hierarchical condensation or the influence of nascent chain concentration on correct folding.

Biological implications

Oligomerization is an important step in the activation or regulation of many proteins, but remains an enigmatic process, especially when it involves arm exchange or domain swap. Proline residues are frequently found at the base of exchanged arms or domains. The conservation of these residues throughout related sequences, as well as the consequences of their natural or site-directed mutagenesis, reveal the critical role of these prolines in inducing arm exchange. This role is explained by the unique constraints that prolines impose on the mainchain.

Some proteins are known to exist both in an oligomeric form with exchanged arms and in a monomeric form with the arm folded back. Prolines seem to be involved in switching between the two conformations. Studies using RNaseA clearly show that this can be a temperature-dependent process, and it may be envisaged that proline could act as a primary switch modifying the oligomerization of some proteins after heat shock.

There is, as yet, no evidence whether prolyl *cis-trans* isomerases are involved in the folding/oligomerization process or in the dynamic switching between conformations, but it is tempting to speculate that these isomerases could play a role in regulating the oligomeric state of proteins and thus their functionality.

The ability of prolines to trigger arm exchange also has practical applications. New synthetic oligomers could be created by forcing arm exchange through a single amino

acid mutation. In cases of existing oligomers with arm exchange, the ratio of oligomers to monomers could be modified: replacing a single amino acid with a proline at the base of the exchanged arms should increase the ratio, whereas replacing a proline in a hinge with another amino acid should decrease it. Such mutations, by artificially favouring one conformation of the arm, could help the study of proteins, such as CDC2 kinase, the mode of action of which is thought to be linked to conformational modifications and to variations of the oligomeric state.

Materials and methods

Programs and databases

Sequences were taken from the SWISSPROT database, except pseudo-proteases, which were taken from [21]. All the alignment work was carried out using the UWGCG package [38] using default values without any special tuning. Structures were taken from the Brookhaven protein data bank [39], except *E. coli* dUTPase, which was kindly provided by E Cedergren and Z Dauter, and BRP, which was solved by the authors. Illustrations were produced using MOLSCRIPT [40] and RASTER3D [41,42].

Gene cloning and site-directed mutagenesis

The *Shble* gene from plasmid pARA-Sh5 [43] was cloned as an *NcoI*-*HindIII* fragment into the pET-26b(+) vector. The desired mutation ShP9G was introduced on the primer and a PCR reaction was performed on the pARA vector. The oligonucleotides were: 5'-CCATGGC-CAAGTTGACCAAGTGCCGTTGGGGTGTCCACC-3' and 5'-AAGCT-TATCAGTCCTGCTCCTCGGCCAC-3'. The purified fragment was introduced into pET-26b(+) with same restriction sites as the *Shble* gene. The resulting vectors were expressed in BL21(DE3) according to the manufacturer's recommendations (Novagen).

BRP purification and gel electrophoresis

Bacterial cells were grown at 37°C on Luria broth supplemented with 50 µg ml⁻¹ of kanamycin. Cells were harvested by centrifugation 3 h after induction with 1 mM IPTG. Periplasmic BRP was liberated by osmotic shock as described previously [7]. The osmotic shock fluids (4 ml) were brought to 20 mM in acetate, and the pH was adjusted to 5.0. The solution was applied to a Pharmacia MonoQ HR5/5 anion exchange column. The proteins were eluted with a 0–0.6 M linear NaCl gradient in 20 mM acetate pH 5.0 buffer. The BRP containing fractions were further dialyzed against 50 mM Tris-HCl pH 6.8 for electrophoresis analysis. SDS-PAGE and native PAGE electrophoresis were performed using 15% polyacrylamide gels, as described in [44]. Two sample preparations were tested: samples were either diluted with SDS-reducing buffer using 2-mercaptoethanol and heated for 5 min at 95°C, or SDS-samples were maintained at room temperature. Gels were stained by Coomassie Blue R-250 immediately after electrophoresis.

Acknowledgements

We thank E Cedergren and Z Dauter for kindly providing coordinates of *E. coli* dUTPase, C Botanch for technical expertise, M Ehrard for help in CD experiments, J Reinbolt for protein N terminus sequencing, F Tete and Y Timsit for stimulating discussion and criticism of the numerous drafts of this article and K Flanagan for proofreading.

References

- Goodsell, D.S. & Olson, A.J. (1993). Soluble proteins: size, shape and function. *TIBS* **18**, 65–68.
- Richardson, J.S. (1981). The anatomy and taxonomy of protein structure. *Adv. Protein Chem.* **34**, 167–339.
- Abad-Zapatero, C., Griffith, J.P., Sussman, J.L. & Rossmann, M.G. (1987). Refined crystal structure of dogfish M₄ apo-lactate dehydrogenase. *J. Mol. Biol.* **198**, 445–467.
- Bennett, M.J., Choe, S. & Eisenberg, D. (1994). Refined structure of dimeric diphtheria toxin at 2.0 Å resolution. *Prot. Sci.* **3**, 1444–1463.
- Bennett, M.J., Schlunegger, M.P. & Eisenberg, D. (1995). 3D domain swapping: a mechanism for oligomer assembly. *Prot. Sci.* **4**, 2455–2468.
- Liddington, R.C., Yan, Y., Moulai, J., Sahli, R., Benjamin, T.L. & Harrison, S. (1991). Structure of simian virus 40 at 3.8 Å resolution. *Nature* **354**, 278–284.
- Dumas, P., Bergdoll, M., Cagnon, C. & Masson, J.M. (1994). Crystal structure and site-directed mutagenesis of a bleomycin resistance protein and their significance for drug sequestering. *EMBO J.* **13**, 2483–2492.
- Gatignol, A., Durand, H. & Tiraby, G. (1988). Bleomycin resistance conferred by drug binding protein. *FEBS Lett.* **230**, 171–175.
- Rydel, T.J., Tulinski, A., Bode, W. & Huber, R. (1991). The refined structure of the Hirudin–Thrombin Complex. *J. Mol. Biol.* **221**, 583–601.
- Chen, L. & Hol, W.G.J. (1992). Crystal structure of an electron-transfer complex between methylamine dehydrogenase and amicyanin. *Biochem.* **31**, 4959–4964.
- Cedergren-Zeppezauer, E.S., Larsson, G., Nyman, P.O., Dauter, Z. & Wilson, K.S. (1992). Crystal structure of dUTPase. *Nature* **355**, 740–743.
- Eck, M. & Sprang, S. (1989). The structure of Tumor Necrosis Factor-α at 2.0 Å resolution. *J. Biol. Chem.* **264**, 17595–17605.
- Fita, I., Silva, A.M., Murthy, M.R.N. & Rossmann, M.G. (1986). The refined structure of beef liver catalase at 2.5 Å resolution. *Acta Cryst. B* **42**, 497–501.
- Remington, S., Wiegand, G. & Huber, R. (1982). Crystallographic refinement and atomic models of two different forms of citrate synthase at 2.7 and 1.7 Å resolution. *J. Mol. Biol.* **158**, 111–152.
- Russel, R.J.M., Hough, D.W., Danson, M.J. & Taylor, G.L. (1994). The crystal structure of citrate synthase from the thermophilic Archaeon, *Thermoplasma acidophilum*. *Structure* **2**, 1157–1167.
- Wodlawer, A., Bott, R. & Sjolín, L. (1982). The refined crystal structure of ribonuclease A at 2.0 Å resolution. *J. Biol. Chem.* **257**, 1325–1332.
- Mazzarella, L., Capasso, S., Demasi, D., Di Lorenzo, G., Mattia, C.A. & Zagari, A. (1993). Bovine seminal ribonuclease: structure at 1.9 Å resolution. *Acta Cryst. D* **49**, 389–402.
- Piccoli, R., Tamburrini, M., Picciali, G., DiDonato, A., Parente, A. & d'Alessio, G. (1992). The dual-mode quaternary structure of seminal RNase. *Proc. Natl. Acad. Sci. USA* **89**, 1870–1874.
- DiDonato, A., Cafaro, V. & d'Alessio, G. (1994). Ribonuclease A can be transformed into a dimeric ribonuclease with antitumor activity. *J. Biol. Chem.* **269**, 17394–17396.
- Mol, C.D., Harris, J.M., McIntosh, E.M. & Tainer, J.A. (1996). Human dUTP pyrophosphatase: uracil recognition by a β hairpin and active sites formed by three separate subunits. *Structure* **4**, 1077–1092.
- McGeoch, D.J. (1990). Protein sequence comparisons show that the 'pseudoproteases' encoded by poxviruses and certain retroviruses belong to the deoxyuridine triphosphatase family. *Nucl. Acids Res.* **18**, 4105–4110.
- Climie, S., Lutz, T., Radul, J., Sumner-Smith, M., Vandenberg, E. & McIntosh, E. (1994). Expression of trimeric human dUTP pyrophosphatase in *Escherichia coli* and purification of the enzyme. *Protein Expr. Purif.* **5**, 252–258.
- Brandts, J.F., Halvorson, H.R. & Brennan, M. (1975). Consideration of the possibility that the slow step in protein denaturation reactions is due to *cis-trans* isomerism of proline residues. *Biochem.* **14**, 4953–4963.
- Tonelli, A.E. (1974). Conformational characteristics of polypeptides containing isolated L-proline residues with *cis* peptide bonds. *J. Mol. Biol.* **86**, 627–635.
- Szyperski, T., Guntert, P., Stone, S.R. & Wuthrich, K. (1992). Nuclear magnetic resonance solution structure of hirudin (1–51) and comparison with corresponding three-dimensional structures determined using the complete 65-residue hirudin polypeptide chain. *J. Mol. Biol.* **228**, 1193–1205.
- Bourne, Y., et al., & Tainer, J.A. (1995). Crystal structure of the cell cycle-regulatory protein *suc1* reveals a β hinge conformation switch. *Proc. Natl. Acad. Sci. USA* **92**, 10232–10236.
- Endicott, J.A., et al., & Johnson, L.N. (1995). The crystal structure of p13^{suc1}, a p34^{cdc2}-interacting cell cycle control protein. *EMBO J.* **14**, 1004–1014.
- Birck, C., Vachette, P., Welch, M., Swarén, P. & Samama, J.P. (1996). Is the function of the *cdc2* kinase subunit proteins tuned by their propensities to oligomerize? Conformational states in solution of the *cdc2* kinase partners p13^{suc1} and p9^{ckshy}. *Biochem.* **35**, 5577–5585.
- Hohenester, E. & Jansonius, J.N. (1994). Crystalline mitochondrial aspartate aminotransferase exists in only two conformations. *J. Mol. Biol.* **236**, 963–968.

30. Bennett, M.J. & Eisenberg, D. (1994). Refined structure of monomeric diphtheria toxin at 2.3 Å resolution. *Prot. Sci.* **3**, 464–1475.
31. Geering, K., *et al.*, & Rossier, B.C. (1993). Mutation of a conserved proline residue in the β subunit ectodomain prevents Na⁺-K⁺-ATPase oligomerisation. *Am. J. Physiol. C* **265**, C1169–C1174.
32. Mazzarella, K., Vitagliano, L. & Zagari, A. (1995). Swapping structural determinants of ribonucleases: an energetic analysis of the hinge peptide 16–22. *Proc. Natl. Acad. Sci. USA* **92**, 3799–3803.
33. Bartos, E. & Uziel, M. (1961). Guinea pig pancreas ribonucleases. Separation by ion exchange chromatography and the subcellular distribution of these enzymes. *J. Biol. Chem.* **236**, 1697–1700.
34. Hopper, P., Harrison, S.C. & Sauer, R.T. (1984). Structure of tomato bushy stunt virus V. Coat protein sequence determination and its structural implications. *J. Mol. Biol.* **177**, 701–713.
35. Speir, J.A., Munshi, S., Wang, G., Baker, T.S. & Johnson, J.E. (1994). Structures of the native and swollen forms of cowpea chlorotic mottle virus determined by X-ray crystallography and cryo-electron microscopy. *Structure* **3**, 63–78.
36. Valegard, K., Murray, J.B., Stockley, P.G., Stonehouse, N.J. & Liljas, L. (1994). Crystal structure of an RNA bacteriophage coat protein-operator complex. *Nature* **371**, 623–626.
37. Stonehouse, N.J., Valegard, K., Golmohammadi, R., van den Worm, S., Stockley, P.G. & Liljas, L. (1996). Crystal structures of MS2 capsids with mutations in the subunit FG loop. *J. Mol. Biol.* **256**, 330–339.
38. UWGCG (1994). Program Manual for the Wisconsin Package.
39. Bernstein, F.C. (1977). The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535–542.
40. Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Cryst.* **24**, 946–950.
41. Bacon, D.J. & Anderson, W.F. (1988). A fast algorithm for rendering space-filling molecule pictures. *J. Mol. Graph.* **6**, 219–220.
42. Merritt, E.A. & Murphy, M.E.P. (1994). Raster3D, version 2.0: a program for photorealistic molecular graphics. *Acta Cryst. D* **50**, 869–873.
43. Cagnon, C., Valverde, V. & Masson, J.-M. (1991). A new family of sugar-inducible expression vectors for *Escherichia coli*. *Protein Eng.* **4**, 843–847.
44. Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685.
45. Vainshtein, B.K., *et al.*, & Rossmann, M.G. (1986). Three-dimensional structure of catalase from *Penicillium vitale* at 2.0 Å resolution. *J. Mol. Biol.* **188**, 49–61.
46. Gouet, P., Jouve, H.-M. & Dideberg, O. (1995). Crystal structure of *Proteus mirabilis* catalase with and without bound NADPH. *J. Mol. Biol.* **249**, 933–954.
47. Murshudov, G.N., *et al.*, & Wilson, K.S. (1995). Three-dimensional structure of catalase from *Micococcus lysodeikticus* at 1.5 Å resolution. *FEBS Lett.* **312**, 127–131.
48. Bravo, J., *et al.*, & Fita, I. (1995). Crystal structure of catalase HP11 from *Escherichia coli*. *Structure* **3**, 491–502.
49. Jäger, J., Moser, M., Sauder, U. & Jansonius, J.N. (1994). Crystal structure of *Escherichia coli* aspartate aminotransferase in two conformations. Comparison of an unliganded open and two liganded closed forms. *J. Mol. Biol.* **239**, 285–305.
50. Malashkevitch, V.N., Strokopytov, B.V., Borisov, V.V., Dauter, Z., Wilson, K.S. & Torchinsky, Y.M. (1995). Crystal structure of the closed form of chicken cytosolic aspartate aminotransferase at 1.9 Å resolution. *J. Mol. Biol.* **247**, 111–124.
51. Hogle, J.M., Maeda, A. & Harrison, S.C. (1986). Structure and assembly of turnip crinkle virus I. X-ray crystallographic structure analysis at 3.2 Å resolution. *J. Mol. Biol.* **191**, 625–638.
52. Sorger, P.K., Stockley, P.G. & Harrison, S.C. (1986). Structure and assembly of turnip crinkle virus II. Mechanism of reassembly *in vitro*. *J. Mol. Biol.* **191**, 639–658.