J Mol Evol (1984) 20:296–303

Journal of
Molecular Evolution

Structural Implications of Primary Sequences from a Family of Balbiani Ring-Encoded Proteins in *Chironomus*

S.J. Hamodrakas[1] and F.C. Kafatos[2,3]

[1] Department of Biochemistry, Cell and Molecular Biology and Genetics, University of Athens, Panepistimiopolis, Kouponia, Athens 15701, Greece
[2] Cellular and Developmental Biology, 16 Divinity Avenue, Cambridge, Massachusetts 02138, USA
[3] Institute of Molecular Biology and Biotechnology and Department of Biology, University of Crete, Heraklio, Crete, Greece

Summary. DNA sequencing has revealed an internal, tandemly repetitive structure in the family of giant polypeptides encoded by three types of Balbiani ring (BR) genes, in three different species of *Chironomus*. Each major BR repeat can be subdivided into two halves: a region consisting of short subrepeats and a more constant region that lacks obvious subrepeats. Comparative predictions of secondary structure indicate that an α-helical segment is consistently present in the amino-terminal half of the constant region in all known BR proteins. Comparative predictions, coupled with consideration of the known phosphorylation of serine and threonine residues in BR proteins, suggest that the α-helical structure may also extend into the carboxy-terminal half of the constant region, possibly interrupted by β-turn(s). However, it is also possible that the structure is variable, and that a β-strand is present in that half in some cases. All of the constant regions conserve one methionine and one phenylalanine residue, as well as all four cysteines; these residues presumably play roles in the packing or cross-linking of aligned constant regions. The structure of the subrepeat region is not clear, but the prevalence of a tripeptide pattern (basic–proline–acidic) suggests some type of structural regularity, possibly an extended helix. The possible significance of these conserved molecular features is discussed in the context of how they may serve the elasticity, insolubility, and hydrophilicity of the fibrils and threads formed by the BR polypeptides.

Key words: Multigene families—Repetitive polypeptides—Structural proteins—Salivary gland secretion—Secondary structure prediction—α-Helical structure

Offprint requests to: F.C. Kafatos

Introduction

The Balbiani rings (BRs) are giant puffs in the polytene chromosomes of the salivary glands in *Chironomus* and related dipteran genera. These chromosomal sites encode secretory polypeptides of extremely high molecular weight ($M_r \approx 10^6$ daltons) that are spun into the elastic threads that make up the protective tube and food-gathering funnel of the aquatic midge larva (Grossbach 1977; Edström et al. 1980; Hertner et al. 1980; Rydlander and Edström 1980; Rydlander et al. 1980).

In recent years, through molecular cloning, Southern hybridization, and DNA sequence analysis, rapid progress has been made in the characterization of BR genes and the corresponding polypeptides (for references, see Pustell et al. 1984). The sequences are clearly homologous, constituting a gene family. They show prominent internal repetitiousness across three hierarchical levels. First, the major part of the translated portion in each BR gene consists of tandem repeats of a fundamental unit approximately 240 bp long (range ca. 180–300 bp). Many repeats in the same gene are highly similar or identical copies of one or more major repeat types (type α, type β, etc.); minor, more variant repeats also exist. Second, each of these repeats can be divided into two portions. One consists of short, tandem subrepeats (9–33 bp long); the other shows no obvious subrepeat structure, and is called the con-

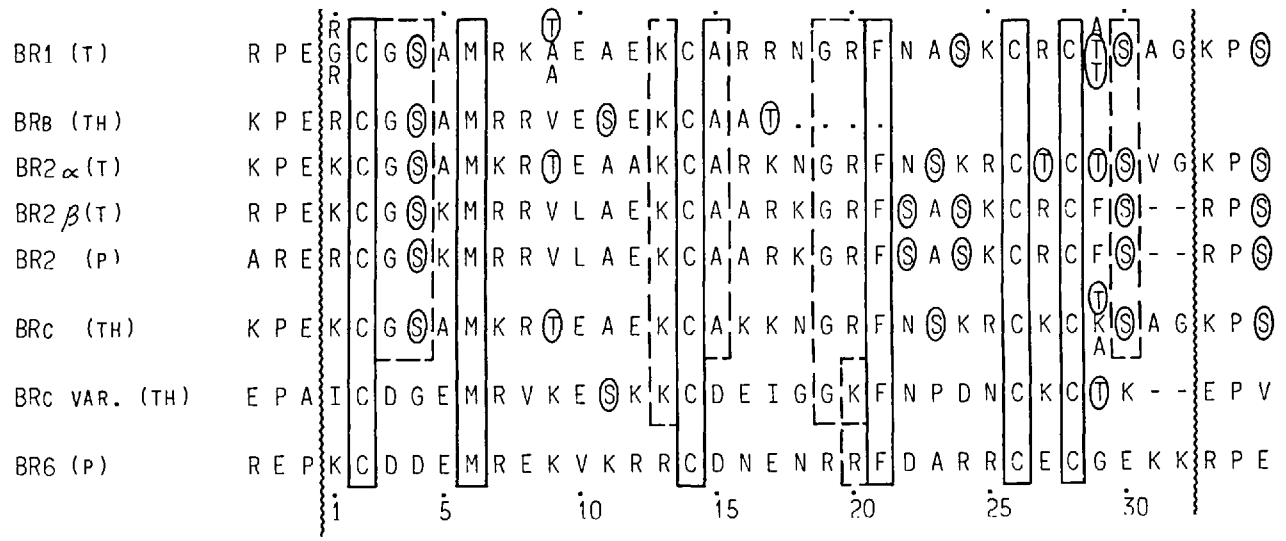

Fig. 1. Protein sequence comparisons of the constant regions of BR repeats. Wavy lines separate the constant region from the flanking tripeptides, which are assigned to the subrepeat region (see Fig. 2). Residues are numbered from the amino-terminal end of each constant region. Completely invariant residues are boxed in solid lines, and residues conserved in all major BR1 and BR2 (BRb and BRc) repeats are boxed in dashed lines. Serine and threonine residues, which are thought to be phosphorylated (see text), are circled. Gaps inserted for alignment are shown as dashes, and an incomplete sequence is terminated with dots. For multiple sequences of the same repeat type, identical residues are shown only once. Sequences were obtained from the following publications: BR1, from top to bottom, Degelmann and Hollenberg (1981), Wieslander et al. (1982), and Case and Byers (1983); BRb, Bäumlein et al. (1982a): BR2α, Sümegi et al. (1982); BR2β, Wieslander and Lendahl (1983) and Case et al. (1983); BR2, Jäckle et al. (1982); BRc, top two lines, Bäumlein et al. (1982b), and bottom line, U. Wobus et al., manuscript in preparation; BRc variant (VAR.), U. Wobus et al., manuscript in preparation; BR6, Galler et al. (1984). Species are abbreviated as follows: T, *Chironomus tentans*; TH, *C. thummi*; P, *C. pallidivittatus*

stant region because it is substantially conserved, in both sequence and length, between genes. Finally, within many subrepeats a shorter, nonanucleotide repetitiousness is evident. The hierarchically repetitive BR DNA sequences and their possible evolutionary origin are discussed in the accompanying paper (Pustell et al. 1984).

The repetitiousness of BR genes involves multiples of three bases, and thus corresponds to repetitive polypeptide substructures. Since the information necessary for a protein to fold in its native conformation is encoded in the primary sequence (Anfinsen 1973), the secondary structures of BR polypeptides must also have repetitive elements. In principle, the main structural features of these proteins can be inferred from conceptual translations of the DNA repeat units. These inferences should be facilitated by the availability of multiple related sequences: Comparisons will highlight the important features, which should be conserved within this family of proteins with a common evolutionary origin and related functions.

In this report, we compare and discuss the amino acid sequences and possible secondary structures of BR repeats. We have used the currently available sequences from a variety of BR repeat units studied by S.T. Case and collaborators; B. Daneholt, L. Wieslander, and collaborators; J.E. Edström and collaborators; and U. Wobus and collaborators.

Methods

Protein Sequences. BR protein sequences were inferred from published DNA sequences and from sequences being prepared for publication; references are given in the legend to Fig. 1. For convenience and accuracy, the conceptually translated protein sequences will be named according to the corresponding genes, although the in vivo protein products are designated sp-Ia, sp-Ib, and sp-Ic.

Secondary Structure Prediction. The methods used for secondary structure prediction have been described in detail by Hamodrakas et al. (1982a).

Molecular Modeling. Preliminary modeling of parts of the protein structure was performed by utilizing the interactive computer graphics facilities of European Molecular Biology Laboratory, Heidelberg. An Evans and Sutherland Multipicture System was used with color and black-and-white displays, 256 kilowords of extended memory, and various input and output devices. The system is served by a Digital Equipment Corp. VAX-11/780 computer. The interactive molecular modeling program FRODO, originated by T. Alwyn Jones and modified for use at EMBL, Heidelberg, by H. Bosshard and C. Carlson, was employed.

Results

Sequence Comparisons

Figures 1 and 2 present the amino acid sequences of constant and subrepeat regions, respectively, from three kinds of BR proteins and three different species

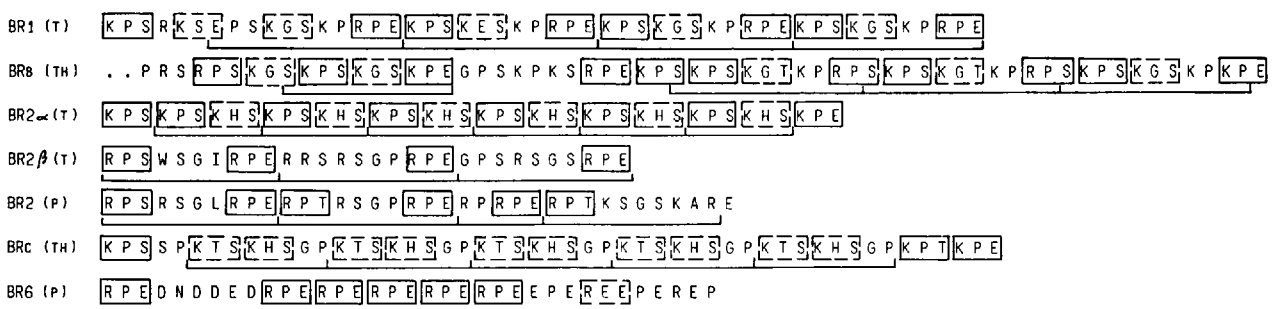

Fig. 2. Protein sequence comparisons of the subrepeat regions of BR repeats. Since constant and subrepeat regions alternate in tandem, the amino-terminal (left) tripeptide of each subrepeat region is the one shown in Fig. 1 flanking the carboxy-terminal (right) end of the constant region. Similarly, the tripeptide at the right end of the subrepeat region is shown in Fig. 1 flanking the left end of the constant region. Subrepeats are indicated with brackets. The tripeptide subrepeats of BR6 and some extensively disrupted subrepeats in BRb are not marked, and some others are slightly permuted relative to those shown by Pustell et al. (1984). Typical tripeptides [(basic–proline–acidic (serine or threonine)] are boxed with solid lines, and variant tripeptides (with glycine, threonine, or histidine instead of proline in the second position) are boxed with dotted lines. References and species abbreviations are as in Fig. 1

of *Chironomus*. It is thought that the BRb and BRc components of *C. thummi* correspond to the BR1 and BR2 components, respectively, of both *C. tentans* and *C. pallidivittatus*. In BR2 of *C. tentans*, two major repeat types (α and β) are known. In BRc of *C. thummi*, in addition to the standard repeat type, a minor variant has been characterized (Pustell et al. 1984; U. Wobus et al., manuscript in preparation). BR6 is a third kind of BR sequence that, unlike the other two, is inducible by several environmental factors, including starvation for phosphate. The BR1 and BR2 (or BRb and BRc) proteins are heavily phosphorylated, primarily in serine residues but also in threonines, whereas the BR6 component lacks serine and threonine and is not phosphorylated (Galler et al. 1984; N.N. Kao and S.T. Case, personal communication).

The exact borders between constant and subrepeat regions are somewhat arbitrary, and these two regions may be separated by short transitional sequences (Pustell et al. 1984). As shown in Fig. 1, we have chosen to define the constant regions so as to exclude on either side nearly invariable proline residues embedded in characteristic tripeptides (see below). As defined, the constant regions are almost devoid of prolines, whereas the subrepeat regions are proline rich.

The conservative nature of the constant region is evident in Fig. 1. All the sequences can be aligned unambiguously. Most are 32 residues long, of which the first 30 suffer no deletions or insertions. Thirteen residues are invariant in all the major BR1 and BR2 (or BRb and BRc) repeats examined; of these, six (all four cysteines, the single methionine, and a phenylalanine) are completely invariant, also occurring in the much more distantly related BRc variant and BR6 sequences.

In contrast, subrepeats are rich in proline, charged amino acids, and serine; vary in length; and cannot be aligned unambiguously (Fig. 2). Much of their sequence can be described in terms of a tripeptide pattern corresponding to the nonanucleotide repetitiousness: Typically a proline is flanked on the amino-terminal side by a basic residue (lysine or arginine) and on the carboxy-terminal side by serine, threonine, or glutamate. It has been pointed out that since most serines and many threonines are phosphorylated in BR sequences, this pattern of residues is fundamentally basic–proline–acidic (Galler et al. 1984). Similar tripeptides with glycine, threonine, or histidine rather than proline in the central position are found more rarely. The tripeptides are embedded in subrepeats varying in length up to 11 residues. In most cases, typical tripepetides immediately flank the constant region and have helped to define the constant/subrepeat region borders (Fig. 1).

Secondary Structure Predictions

We predicted the secondary structures of all known constant regions plus the immediately flanking tripeptides, using computer programs based on six different predictive methods (Hamodrakas et al. 1982a). Figure 3 presents typical results that span the range of variations observed. For each polypeptide, individual predictions of α-helices, β-pleated sheets and β-turns were made by each method separately; predicted structures are indicated by corresponding horizontal lines (Fig. 3). Joint prediction histograms were then constructed, since these are more dependable than individual predictive schemes (Schulz et al. 1974; Argos et al. 1976). In Fig. 3, the structure predictions are shown above the primary sequences.

The first half of the constant region invariably is dominated by an α-helix; this prediction is strongest and most consistent for the segment between methionine-6 and cysteine-14, and extends for variable

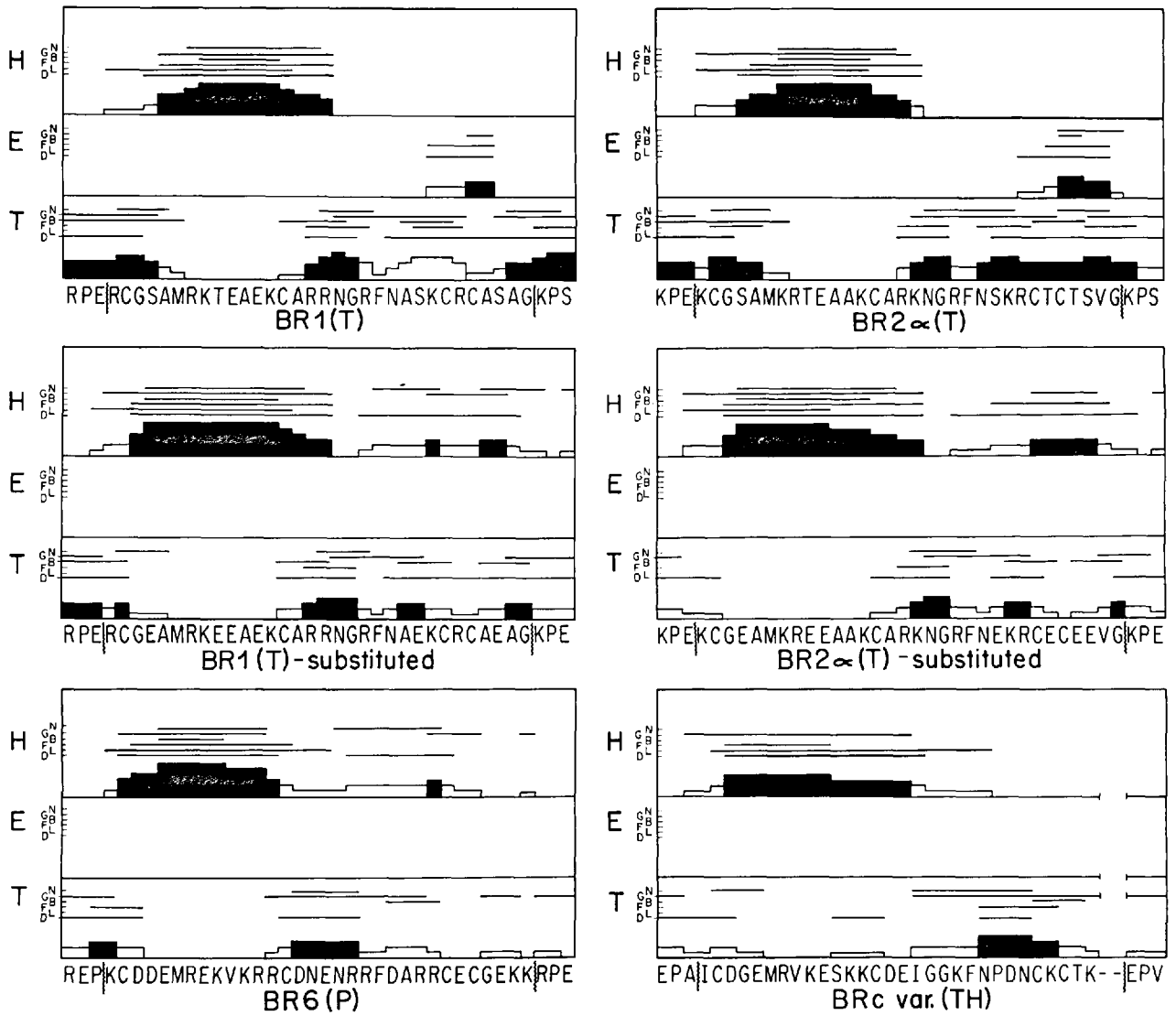

Fig. 3. Secondary structure predictions for the constant regions of selected BR sequences. For each sequence, individual predictions for α-helices (H), β-sheets (E), or β-turns (T), as derived according to the methods of Nagano (1977a, b) (N), Garnier et al. (1978) (G), Burgess et al. (1974) (B), Chou and Fasman (1974a, b) (F), Lim (1974a, b) (L), and Dufton and Hider (1977) (D), are shown by horizontal lines. Joint prediction histograms, constructed by tallying the individual predictions, are also shown. The most probable structures (those predicted by three or more methods) are indicated by shaded areas. Sequences are from the references in the legend to Fig. 1. For BR1 and BR2α, predictions are also shown for sequences in which glutamate is substituted for serine and threonine to account for the presumed phosphorylation of the latter (see text)

distances on either side. Surprisingly, the second half of the constant region does not yield consistent predictions, except for a likely β-turn centered at about residue 18 (usually asparagine); the rest is variously predicted as β-strand (BR1, BR2, BRc), α-helix (BR6 and, to a lesser extent, the BRc variant), and coil or turn (all).

Predictive methods fail for sequences that include short precise subrepeats (Chou and Fasman 1978). Thus, we attach no significance to predictions (mostly random coil or β-turn) obtained for the subrepeat region by the same programs as were used for Fig. 3. The tripeptide pattern (basic–proline–acidic) that dominates the subrepeats suggests the likely existence of some type of regular structure. α-Helical structure might be excluded by the prevalence of proline residues; however, in screening the protein data bank of crystallographically characterized proteins (Bernstein et al. 1977; Levitt and Greer 1977), we have noted that the tripeptides typical of BR subrepeat regions frequently occur within α-helical segments (data not shown). A more likely possibility is an extended, collagenlike helix with the side chains of the first and third residues alternating in directions away from the main chain (as in Fig. 4). This would generate curved, oppositely charged "faces" that might be neutralized, with formation of intermolecular ionic bonds, if the polypeptides were properly packed in parallel alignment. Alternatively, the oppositely charged side chains might interact

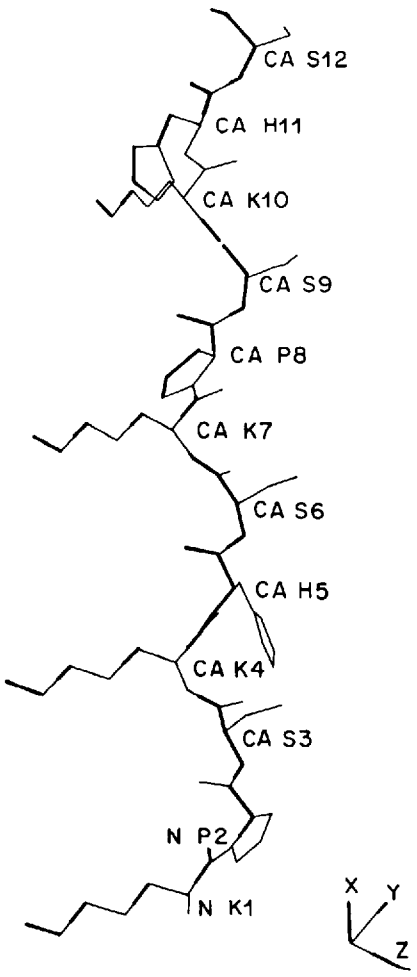

Fig. 4. Two subrepeats of BR2α from *C. tentans* (Sümegi et al. 1982; see Fig. 2) modeled in a hypothetical conformation consisting of an extended, collagenlike, left-handed helix ($\phi = -60°$, $\psi = +140°$). The sequence is KPSKHSKPSKHS, from bottom to top. Note that the lysine and serine side chains point away from the polypeptide backbone, in opposite orientations, and form two oppositely charged, curved "faces" (assuming that the serines are phosphorylated)

intramolecularly, resulting in a more "closed" conformation.

Discussion

Although the amino acid sequence dictates native conformation, secondary structure predictions based on primary sequences should be undertaken with full awareness of their limitations. Even in the case of globular proteins, for which they were initially developed and applied, the accuracy of predictive methods is limited (Chou and Fasman 1978). The methods fail altogether when short precise repeats are present, and thus have been applied only rarely to structural proteins, in which internal repeats are widespread. However, comparisons of evolutionarily related sequences or of imprecise internal repeats are invaluable in overcoming these limitations: Limited variation should reduce the "noise" and help identify consistent structural features. This approach has been applied successfully to the central domains of silkmoth chorion proteins, for which comparative secondary structure predictions and the observation of periodicities corresponding to imprecise internal repeats have led to the identification of β-sheet strands alternating with β-turns as the predominant structure (Hamodrakas et al. 1982a; S.J. Hamodrakas and F.C. Kafatos, manuscript in preparation); these predictions have been supported by both experimental measurements (Hamodrakas et al. 1982b) and model-building (S.J. Hamodrakas and F.C. Kafatos, manuscript in preparation). Similarly, a repetitive secondary structure in the adenovirus fiber protein has been elucidated on the basis of internal periodicities corresponding to imprecise repeats (Green et al. 1983).

Since α-helical structure is consistently predicted in the first half of the BR constant region, there seems little doubt about its reality. Other predicted structures are less consistent and therefore less certain. β-turns are predicted on either side of the α-helical segment, although not always at the same location (cf. the BRc variant). For the major BR1, BR2, and BRc sequences, a short β-strand is consistently predicted in the second half of the constant region, flanked by or overlapping with β-turns. These features would suggest a globular supersecondary structure for the constant region. However, the BR6 sequence and, to a lesser extent, the BRc variant yield predictions of an α-helical structure extending into the second half of the constant region. Although this discrepancy may be real, corresponding to diverse polypeptide structures and functions, we favor the alternative possibility, that the discrepancy is only apparent.

We have noted that the complete constant region sequences of major BR1, BR2, and BRc repeats show significant numbers of serines and threonines (four to six), whereas these residues are more rare or absent in the BRc variant and BR6 sequences (two and zero, respectively). Most serines and many threonines are phosphorylated in the BR polypeptides (Galler et al. 1984; N.N. Kao and S.T. Case, personal communication). Phosphorylation would tend to make these residues conformationally equivalent to glutamate, as they apparently are in the tripeptides of the subrepeats (Fig. 2; Galler et al. 1984). If so, formation of an α-helix would be promoted [glutamate is a strong helix-former, whereas serine and threonine are neutral or unfavorable to helix formation (Chou and Fasman 1978)]. This is illustrated by the predictions for "substituted" sequences in Fig. 3: If the putative phosphoserines

and phosphothreonines are considered equivalent to glutamate, α-helical predictions are enhanced at the expense of β-strand and β-turn predictions, and the inferred structures for the second halves of major BR1, BR2, and BRc repeats become reasonably similar to those for BR6. Furthermore, in considering replacements in all of the available constant region sequences, we have noted that serines and threonines tend to be replaced by helix-formers (Fig. 1). Of 51 replacements of serine or threonine, 34 are by strong or reasonably strong helix-formers (13 alanines, 4 glutamates, 11 lysines, 3 valines, 2 phenylalanines, and 1 isoleucine), 10 are by neutral or weak helix-formers (3 aspartates and 7 arginines), and only 7 are by helix-breakers (4 asparagines, 2 glycines, and 1 proline). Such replacement frequencies are not typical of other proteins (Schulz and Schirmer 1978), but are consistent with the interpretation that after posttranslational modification the serines and threonines become helix-formers. Thus, the possibility must be seriously considered that an α-helix, perhaps punctuated by β-turn(s), is dominant throughout all BR constant regions. We favor this possibility because the extended α-helical prediction is reasonably strong for the BR6 protein: Under phosphate starvation, this component substitutes for BR1 and BR2 (Edström et al. 1980), and we consider it likely that it needs similar structural features to serve equivalent functions.

Figure 5 presents a hypothetical model of a BR constant region as a continuous α-helix shown in radial projection (Crick 1953). This model is tentative and may well require refinement. For example, the consistent prediction of a β-turn centered around residue 18 may indicate that the α-helix is interrupted (but see the BRc variant in Fig. 3). If uninterrupted, the α-helix would be unusually long compared with helices encountered in globular proteins (Schulz and Schirmer 1978); however, long α-helices are known in such structural proteins as tropomyosin (McLachlan and Stewart 1976) and hard keratins (Fraser and McRae 1973).

The main value of the model in Fig. 5 is that it suggests how the constant regions may be packed with neighboring constant regions, helping to build higher-order structures. It should be noted that the number of hydrophobic residues is unusually small, both in the unequivocally α-helical first half and in the second half. Therefore, the packing of BR proteins cannot depend exclusively on hydrophobic surface patches such as are typically used for packing α-helices (Lim 1974a). We note that the BR-encoded proteins function under water, a circumstance that may force greater reliance on hydrophilic residues for packing. If the serine and threonine residues are phosphorylated and therefore negatively charged, then the charged residues would appear to be clustered within an enlongate patch, with frequent and almost periodic "doublets" of basic residues usually juxtaposed with glutamate or phosphoserine and phosphothreonine residues (Fig. 5). Formation of intramolecular ion pairs between these juxtaposed residues might alter the surface properties of the helix sufficiently to promote packing without interfering with hydration. Alternatively, packing might be promoted by electrostatic interactions between the charged patches of different molecules. A second elongate and almost uninterrupted patch on the surface of the helix consists of the polar uncharged and hydrophobic nonpolar residues (alanine, asparagine, cysteine, glycine, methionine, and phenylalanine); this patch may also serve in packing, with perhaps a special role for the invariant methionine, phenylalanine, and cysteine residues. Although the replacements (Fig. 1) affect the shapes of these two types of patches, clustering of residues as in Fig. 5 is a consistent feature of α-helical models for the constant regions. We presume that the packing of the helices is completed by the formation of intermolecular disulfide bonds between the invariant cysteine residues. In the model of Fig. 5, the cysteines are distributed in a manner that would promote cross-linking in multiple directions.

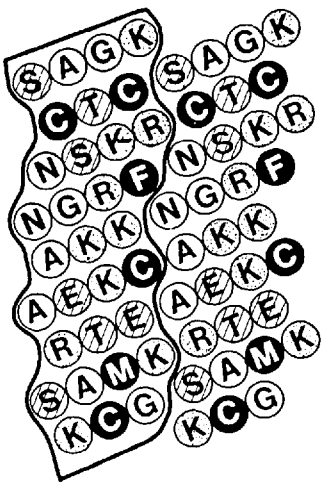

Fig. 5. The sequence of a typical BRc repeat (Bäumlein et al. 1982b), presented in a double radial projection (Crick 1953) of a hypothetical continuous α-helix. The equivalent of one helix is outlined. The invariant M, F, and C residues, which may be involved in packing and cross-linking, are shown on a black background; basic residues (K and R) are shown on a stippled background; and acidic residues (E, D, S, and T, assuming phosphorylation) are shown on a hatched background. Note the formation of two characteristic "patches" consisting of charged vs polar uncharged and hydrophobic nonpolar residues (see text)

Although it would clearly be premature to propose a detailed model of BR-encoded threads, biological considerations, plus the inferred structural features, permit some speculations that may have heuristic value. The long BR threads are known to

be made of 45-Å fibrils (Grossbach 1977). These structures must be elastic, insoluble, and yet hydrated, otherwise they might impede the flow of water and hence the capture of prey in the net. We consider it likely that the giant BR proteins are packed within the 45-Å fibrils with their long axes parallel to the fibril axis. The giant size of the polypeptides and their repetitive structure may lead to a staggered alignment of the polypeptides, and therefore to long, uninterrupted fibrils. Although the staggered alignment may tend to be stochastic, it should be constrained so that the α-helical segments (constant regions) are packed in parallel, cross-linked "bands" alternating with non-cross-linked segments (subrepeat regions). In that orientation, the α-helical segments would impart the required considerable elasticity to the structure, within the constraints of the cross-links. Although the structure of the subrepeat regions is not clear, we consider it likely that they also make a major contribution to the elasticity of the fibrils and threads. The lack of cross-links and the variable lengths of these regions may be important in this respect. Even if the inherently most favorable conformation of the subrepeats is that of an extended helix, as portrayed in Fig. 4, if the aligned subrepeat regions differ in length, some of them would be forced into nonoptimal conformations by the cross-linking of the flanking constant regions; this might facilitate elastic responses over a wider range of stretching.

The necessary insolubility would be ensured by the postulated intermolecular cysteine cross-links between BR proteins packed in parallel alignment. Furthermore, the giant size of the proteins would ensure instantaneous insolubilization as soon as cross-links began to form—a matter of obvious importance, since the threads are spun under water.

Finally, hydration would be ensured by the preponderance of charged residues and by the greater dependence on ion pairs than on hydrophobic patches for stabilization of the structure. These ideas should now be tested by direct analysis of molecular and supramolecular structure in the BR-encoded secretory threads of *Chironomus*.

Acknowledgments. We are grateful to Drs. J. Garnier and J.A. Lenstra and Prof. K. Nagano for computer programs. We also express our gratitude to Dr. H. Bosshard and C. Carlson of EMBL, Heidelberg, for their help in the use of the graphics facilities, and to Prof. J.E. Edström, R. Galler, Dr. U. Wobus, Dr. H. Bäumlein, and J. Pustell for extensive help and for communicating unpublished results that are crucial to some of our arguments. Our thinking profited extensively from the open discussions in the recent Balbiani Ring Workshop sponsored by the DDR Academy of Sciences in "Windenhütte"; the terminology we have used here was agreed upon at that workshop. We thank all participants, especially S.T. Case, B. Daneholt, J.E. Edström, R. Galler, C. Höög, U. Lendahl, J. Pustell, and L. Wieslander. The use of the Protein Data Bank (Bernstein et al. 1977) is also acknowledged. This work was supported by an EMBO short-term fellowship to S.J.H. and by grants from the ACS, the NIH, and the Greek Ministry of Research and Technology to F.C.K.

References

Anfinsen CB (1973) Principles that govern the folding of protein chains. Science 181:223–230

Argos P, Schwarz J, Schwarz J (1976) An assessment of protein secondary structure prediction methods based on amino acid sequence. Biochim Biophys Acta 439:261–273

Bäumlein H, Wobus U, Gerbi SA, Kafatos FC (1982a) The basic repeat unit of a *Chironomus* Balbiani ring gene. Nucleic Acids Res 10:3893–3904

Bäumlein H, Wobus U, Gerbi SA, Kafatos FC (1982b) Characterization of a 249-bp tandemly repetitive, satellite-like repeat in the translated portion of Balbiani ring C of *Chironomus thummi*. EMBO J 1:641–647

Bernstein FC, Koetzle TF, Williams GJB, Meyer EF Jr, Brice MD, Rodgers JR, Kennar O, Shimanouchi T, Tasumi M (1977) The protein data bank: a computer-based archival file for macromolecular structures. J Mol Biol 112:535–542

Burgess AW, Ponnuswamy PK, Scheraga HA (1974) Analysis of conformations of amino acid residues and prediction of backbone topography in proteins. Isr J Chem 12:239–286

Case ST, Byers MR (1983) Repeated nucleotide sequence arrays in Balbiani ring 1 of *Chironomus tentans* contain internally nonrepeating and subrepeating elements. J Biol Chem 258: 7793–7799

Case ST, Summers RL, Jones AG (1983) A variant tandemly repeated nucleotide sequence in Balbiani ring 2 of *Chironomus tentans*. Cell 33:555–562

Chou PY, Fasman GD (1974a) Conformational parameters for amino acids in helical, β-sheet, and random coil regions calculated from proteins. Biochemistry 13:221–222

Chou PY, Fasman GD (1974b) Prediction of protein conformation. Biochemistry 13:222–245

Chou PY, Fasman GD (1978) Prediction of the secondary structure of proteins from their amino acid sequence. Adv Enzymol 47:45–148

Crick FHC (1953) The packing of α-helices: simple coiled-coils. Acta Crystallogr 6:689–697

Degelmann A, Hollenberg CP (1981) A structural analysis of Balbiani ring DNA sequences in *Chironomus tentans*. Chromosoma 83:295–313

Dufton MJ, Hider RC (1977) Snake toxin secondary structure predictions: structure activity relationships. J Mol Biol 115: 177–193

Edström JE, Rydlander L, Franke C (1980) Concomitant induction of a Balbiani ring and a giant secretory protein in *Chironomus* salivary glands. Chromosoma 81:115–124

Fraser RDB, McRae TP (1973) Conformation in fibrous proteins. Academic Press, New York

Galler R, Rydlander L, Riedel N, Kluding H, Edström JE (1984) Balbiani ring induction in phosphate metabolism. Proc Natl Acad Sci USA 81:1448–1452

Garnier J, Osguthorpe DJ, Robson B (1978) Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J Mol Biol 120: 97–120

Green NM, Wrigley NG, Russell WC, Martin SR, McLachlan AD (1983) Evidence for a repeating cross-β sheet structure in the adenovirus fibre. EMBO J 2:1357–1365

Grossbach U (1977) The salivary gland of *Chironomus* (Diptera): a model system for the study of cell differentiation. In:

Beerman W (ed) Results and problems in cell differentiation, vol. 8. Springer-Verlag, Berlin, pp 147–196

Hamodrakas SJ, Jones CW, Kafatos FC (1982a) Secondary structure predictions for silkmoth chorion proteins. Biochim Biophys Acta 700:42–51

Hamodrakas SJ, Asher SA, Mazur GD, Regier JC, Kafatos FC (1982b) Lasar Raman studies of protein conformation in the silkmoth chorion. Biochim Biophys Acta 703:216–222

Hertner T, Meyer B, Eppenberger HM, Mahr R (1980) The secretion proteins in *Chironomus tentans* salivary glands: electrophoretic characterization and molecular weight estimation. Wilhelm Rouxs Arch Dev Biol 189:69–72

Jäckle H, de Almeida JC, Galler R, Kluding H, Lehrach H, Edström JE (1982) Constant and variable parts in the Balbiani ring 2 repeat unit and the translation termination region. EMBO J 1:883–888

Levitt M, Greer J (1977) Automatic identification of secondary structure in globular proteins. J Mol Biol 114:181–293

Lim VI (1974a) Structural principles of the globular organization of protein chains. A stereochemical theory of globular protein secondary structure. J Mol Biol 88:857–872

Lim VI (1974b) Algorithms for prediction of α-helical and β-structural regions in globular proteins. J Mol Biol 88:873–894

McLachlan AD, Stewart M (1976) The 14-fold periodicity in α-tropomyosin and the interaction with actin. J Mol Biol 103: 271–298

Nagano K (1977a) Logical analysis of the mechanism of protein folding IV. Super-secondary structures. J Mol Biol 109:235–250

Nagano K (1977b) Triplet information in helix prediction applied to the analysis of super-secondary structures. J Mol Biol 109:251–274

Pustell J, Kafatos FC, Wobus U, Bäumlein H (1984) Balbiani ring DNA: sequence comparisons and evolutionary history of a family of hierarchically repetitive protein-coding genes. J Mol Evol 20:281–295

Rydlander L, Pigon A, Edström JE (1980) Sequences translated by Balbiani ring 75S RNA in vitro are present in giant secretory protein from *Chironomus tentans*. Chromosoma 81: 101–113

Rydlander L, Edström JE (1980) Large sized nascent protein as dominating component during protein synthesis in *Chironomus* salivary glands. Chromosoma 81:85–99

Schulz GE, Schirmer RH (1978) Principles of protein structure. Springer-Verlag, New York Heidelberg Berlin

Schulz GE, Barry CD, Friedman J, Chou PY, Fasman GD, Finkelstein AV, Lim VI, Ptitsyn OB, Kabat EA, Wu TT, Levitt M, Robson B, Nagano K (1974) Comparison of predicted and experimentally determined secondary structure of adenyl kinase. Nature 250:140–142

Sümegi J, Wieslander L, Daneholt B (1982) A hierarchic arrangement of the repetitive sequences in the Balbiani ring 2 gene of *Chironomus tentans*. Cell 30:579–587

Wieslander L, Lendahl U (1983) The Balbiani ring 2 gene in *Chironomus tentans* is built from two types of tandemly arranged major repeat units with a common evolutionary origin. EMBO J 2:1169–1175

Wieslander L, Sümegi J, Daneholt B (1982) Evidence for a common ancestor sequence for the Balbiani ring 1 and Balbiani ring 2 genes in *Chironomus tentans*. Proc Natl Acad Sci USA 79:6956–6960