Francesco Sorbetti-Guerri
Department of Agronomy and Forestry, University of Florence, Via S.Bonaventura 13, 50145 Florence, Italy
Denis Michel
Bureau d'Etudes et de Recherches Géologiques Appliquées,1500 Avenue de la Pompignane, 34000 Montpellier, France

1. Cantiani, M., Cantiani, M.-G. & Sorbetti-Guerri, F. *Revue Forestière Française* **XLVI,** 349–358 (1994).
2. Zürcher, E. *Schweizerische Zeitschrift für Forstwesen / Journal Forestier Suisse* **143,** 951–966 (1992).
3. Zürcher, E. in *L'Arbre, Biologie et Développement. Troisième Colloque International. Montpellier, 11–15 Sept. 1995* (ed. Edelin, C.) 150–164 (Naturalia Monspenliensia, Montpellier, 1995).
4. Endres, K.-P. & Schad, W. *Biologie des Mondes. Mondperiodik und Lebensrhythmen* (S. Hirzel Verlag, Stuttgart/Leipzig, 1997).
5. Michel, D. *Programme de Mission 'Radeau des Cimes', Autumn 1991* (Projet BERGA-SUD, Univ. Montpellier internal document, 1991).
6. Larcher, W. *Ökologie der Pflanzen auf physiologischer Grundlage* 3rd edn (Ulmer, Stuttgart, 1980).
7. Lyr, H., Fiedler, H. J. & Tranquillini, W. (eds) *Physiologie und Ökologie der Gehölze* (Fischer, Jena/Stuttgart, 1992).
8. Whitehead, D. & Teskey, R. O. *Tree Physiol.* **15,** 245–251 (1995).
9. Thomas, B. & Vince-Prue, D. *Photoperiodism in Plants* 2nd edn (Academic, New York, 1997).
10. Strasburger, E. *Lehrbuch der Botanik* 32nd edn (eds von Denffer, D. *et al.*) (Fischer, Stuttgart/New York, 1983).
11. Raven, P. H., Evert, R. F. & Eichhorn, S. E. *Biology of Plants* 5th edn (Worth, New York, 1992).

Code for collagen's stability deciphered

The most abundant protein in animals is collagen. In connective tissue, this protein is present as chains wound in tight triple helices which are organized into fibrils of great tensile strength and thermal stability[1,2]. We propose a new explanation for this stability.

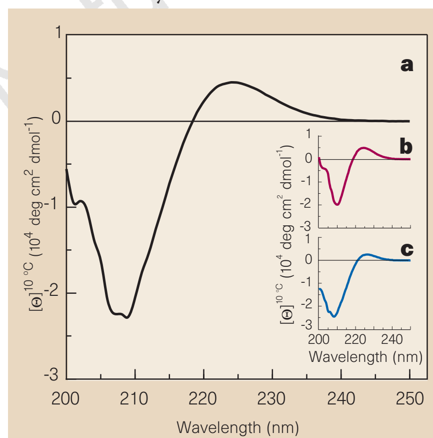

Figure 1 Circular dichroism spectra of collagen-related peptides. **a,** $(ProFlpGly)_{10}$. **b,** $(ProHypGly)_{10}$. **c,** $(ProProGly)_{10}$. $(ProFlpGly)_{10}$ was synthesized by segment condensation of FmocProFlpGlyOH units on 2-chlorotrityl chloride resin. $(ProHypGly)_{10}$ and $(ProProGly)_{10}$ were supplied by Peptides International. Each 30-mer was >90% pure. Circular dichroism spectra were recorded in 50 mM acetic acid at 10 °C after incubation for 24 hours at 4 °C.

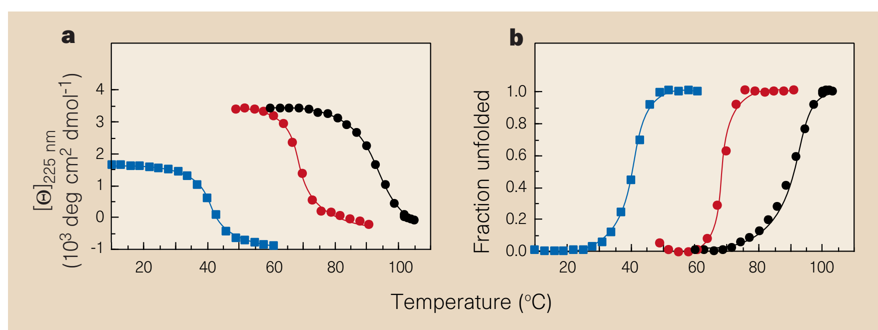

Figure 2 Thermal denaturation of collagen-related triple helices. **a,** Raw data. **b,** Transformed data. Peptides (0.23 mM) were incubated in 50 mM acetic acid for 24 hours at 4 °C. Ellipticity at 225 nm was monitored by circular dichroism spectroscopy as the temperature was increased by increments of 3 °C with 5-minute equilibration. The values of T_m (°C), which is the temperature at the midpoint of the thermal transition curve, are: $(ProProGly)_{10}$ (blue), 41 ± 1; $(ProHypGly)_{10}$ (red), 69 ± 1; $(ProFlpGly)_{10}$ (black), 91 ± 1.

Each polypeptide chain of collagen is composed of repeats of the amino-acid sequence X–Y–Gly, where Gly represents glycine and X and Y are often proline (Pro) or 4(*R*)-hydroxyproline (Hyp) residues, respectively. The thermal stability of triple-helical collagen is enhanced by the hydroxyl group on the pyrrolidine ring of Hyp residues[3]. It is thought that this extra stability arises from hydrogen bonds mediated by a network of bridging water molecules[1,4]. Such water bridges have been detected by X-ray diffraction analysis of crystalline collagen[5]. They consist typically of two water molecules that link a Hyp side-chain of one strand to a main-chain carbonyl of another strand.

We suspected, for several reasons, that water bridges are unlikely to contribute significantly to collagen stability. First, the entropic cost of building and maintaining them would be enormous — the bridges would immobilize more than 500 water molecules per triple helix. Also, triple helices of $(ProProGly)_{10}$ and $(ProHypGly)_{10}$ are stable in methanol or propane-1,2-diol (ref. 6), and the Hyp residues enhance stability even in these anhydrous environments.

Electron-withdrawing groups can alter the attributes of molecules[7]. The inductive effect of the hydroxyl group of Hyp residues is apparent in the structure[8] and properties[9] of molecules that mimic collagen. To distinguish between the contributions of hydrogen bonding and inductive effects on collagen stability, we synthesized $(ProFlpGly)_{10}$, where Flp is a 4(*R*)-fluoroproline residue[8,9]. We chose Flp because fluorine is the most electronegative element but does not form hydrogen bonds[7].

The relative molecular mass of a triple helix of $(ProFlpGly)_{10}$ chains, as determined by sedimentation equilibrium in an analytical ultracentrifuge, is (8.0 ± 0.1K). In 50 mM acetic acid, this triple helix does not bind to 1-anilino-naphthalene-8-sulphonate, suggesting that its tertiary structure is packed tightly rather than being like that of a molten globule[10].

Most significantly, the circular dichroism (CD) spectrum of the $(ProFlpGly)_{10}$ triple helix is indistinguishable from that composed of $(ProHypGly)_{10}$ or $(ProProGly)_{10}$ chains (Fig. 1). These spectra are diagnostic of a collagen triple helix[11].

Flp residues enhance triple-helix stability. The values of T_m for the three triple helices differ dramatically, increasing in the order $(ProProGly)_{10} < (ProHypGly)_{10} < (ProFlpGly)_{10}$ (Fig. 2). The thermal stability of $(ProFlpGly)_{10}$ triple helices far exceeds that of any known collagen of similar size.

We conclude that water bridges do not contribute significantly to collagen stability. Rather, collagen stability relies on previously unappreciated inductive effects. These inductive effects probably enhance collagen stability by favouring the requisite *trans* conformation of the hydroxyprolyl peptide bond[9].

Because a fluorine atom exerts a greater inductive effect than does a hydroxyl group, Flp residues contribute more to collagen stability than do Hyp residues. Accordingly, chemical modification of the Hyp hydroxyl group with an electron-withdrawing substituent could enhance the stability of natural collagen, allowing new biomaterials to be produced[12].

Steven K. Holmgren, Kimberly M. Taylor, Lynn E. Bretscher, Ronald T. Raines
Departments of Biochemistry and Chemistry, University of Wisconsin–Madison, Madison, Wisconsin 53706, USA
e-mail: raines@biochem.wisc.edu

1. Nimni, M. E. (ed.) *Collagen* 1–4 (CRC, Boca Raton, Florida, 1988).
2. Brodsky, B. & Shah, N. K. *FASEB J.* **9,** 1537–1546 (1995).
3. Berg, R. A. & Prockop, D. J. *Biochem. Biophys. Res. Commun.* **52,** 115–120 (1973).
4. Bella, J., Brodsky, B. & Berman, H. M. *Structure* **3,** 893–906 (1995).
5. Bella, J., Eaton, M., Brodsky, B. & Berman, H. M. *Science* **266,** 75–81 (1994).
6. Engel, J., Chen, H.-T., Prockop, D. J. & Klump, H. *Biopolymers* **16,** 601–622 (1977).
7. Ojima, I., McCarthy, J. R. & Welch, J. T. (eds.) *Biomedical Frontiers of Fluorine Chemistry* (Amer. Chem. Soc.,

Nature © Macmillan Publishers Ltd 1998

Washington, DC, 1996).
8. Panasik, N., Jr, Eberhardt, E. S., Edison, A. S., Powell, D. R. & Raines, R. T. *Int. J. Pept. Protein Res.* **44,** 262–269 (1994).
9. Eberhardt, E. S., Panasik, N., Jr & Raines, R. T. *J. Am. Chem. Soc.* **118,** 12261–12266 (1996).
10. Semisotnov, G. V. *et al. Biopolymers* **31,** 119–128 (1991).
11. Shah, N. K., Ramshaw, J. A. M., Kirkpatrick, A., Shah, C. & Brodsky, B. *Biochemistry* **35,** 10262–10268 (1996).
12. Werkmeister, J. A. & Ramshaw, J. A. M. (eds) *Collagen Biomaterials* (Elsevier Science, Barking, Essex, 1992).

Gene translocation links insects and crustaceans

The evolutionary relationships among the four major lineages of arthropods remain controversial, despite extensive study. We report here a derived gene rearrangement common to insects and crustaceans but absent in the other arthropod groups. This finding strongly supports an insect–crustacean evolutionary lineage that is separate from those leading to myriapods and chelicerates.

The four major arthropod groups are Chelicerata (such as scorpions and horseshoe crabs), Crustacea (such as crabs and brine shrimp), Myriapoda (such as centipedes and millipedes), and Insecta (such as flies and beetles). Much of arthropod evolution remains contentious but, until recently, there has been general agreement that myriapods are the closest relatives of insects, forming a group known as the Atelocerata.

However, several recent morphological and molecular comparisons suggest that crustaceans, rather than myriapods, are the sister group to insects. If this is the case, some characteristics shared by insects and myriapods (such as a tracheal system for respiration, Malpighian tubules for excretion, and unbranched legs) then become examples of convergent evolution, perhaps as adaptations to life on land. Similar features are found among terrestrial chelicerates, where their occurrence is already viewed as convergent.

In a previous study[1], we reported that insects and crustaceans share a derived location for the gene encoding mitochondrial leucine transfer RNA, designated *L(UUR)*, as compared with its primitive location in a chelicerate, an onychophoran, and several non-arthropod metazoans. However, in that study we were unable to associate a myriapod (*Thyropygus*) with either group.

Now, further mitochondrial DNA (mtDNA) sequence for *Thyropygus* and for three other myriapod species allows us to make that association. These four myriapods share the gene arrangement *LrRNA-L(CUN)–L(UUR)–ND1*, which is almost certainly primitive[1]. Our earlier misinterpretation of the *Thyropygus* sequence was due to a similarity of the amino-acid sequence

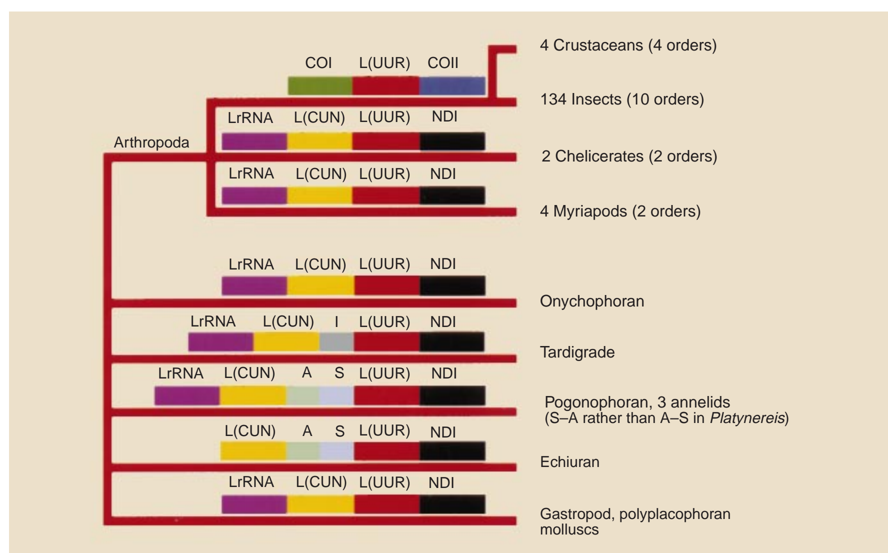

Figure 1 Relative location of the *L(UUR)* gene for 153 taxa. The primitive location is identified in many non-arthropods and is retained in the mitochondrial genomes of the chelicerates and myriapods. An insect–crustacean clade is identified by the shared translocation of *L(UUR)* to the position between *COI* and *COII*. For 49 of these insect taxa, only the gene arrangement *L(UUR)–COII* has actually been determined[8]. For the chelicerates, myriapods, onychophoran, tardigrade, and echiuran, *COI–COII* are directly adjacent without any intervening tRNA genes; other taxa have an unrelated tRNA here. The three crustaceans and ten insects share the gene arrangement *LrRNA–L(CUN)–ND1*. For all arthropods for which the relative locations of these two gene blocks have been determined[2–7], they are separated by more than 2.5 kilobases and are encoded on opposite DNA strands. Data are from published sources for three crustaceans (*Homarus*[1], *Daphnia*[1], *Artemia*[3]), the insects[8] (additional citations available from JLB), a chelicerate (*Limulus*[2]), onychophoran (*Euperipatoides*[1]), annelid (*Lumbricus*[9]), gastropod (*Plicopurpura*[1]) and polyplacophoran (*Katharina*[10]). Sequences determined here are for a remipede crustacean (*Speleonectes*), chelicerate (*Pandinus*), four myriapods (*Thyropygus*, *Lithobius*, *Spirostrephon*, *Narceus*), tardigrade (*Thulinia*), pogonophoran (*Galathe-olinum*), two annelids (*Helobdella*, *Platynereis*) and echiuran (*Urechis*). Sequences were determined from DNA fragments amplified using the polymerase chain reaction with primers made to conserved gene regions. Mitochondrial DNA typically contains 37 genes, only a subset of which is shown here (gene abbreviations are as published[9]).

inferred from the terminal 39 nucleotides of the *L(UUR)* gene. These nucleotides are in frame with the actual start site of the *ND1* gene and match 5/13 of the corresponding residues of the *ND1* gene in *Drosophila*.

The most parsimonious explanation for the gene arrangement data (Fig. 1) is that a single translocation of the *L(UUR)* gene occurred in a common lineage that led, after it split from the other lineages shown, to crustaceans and insects. This signature of common evolutionary history persists in the mtDNAs of these groups today.

If myriapods and insects were sister groups, either this tRNA translocation would need to have occurred twice identically in the lineages leading to insects and crustaceans, or it would need to have reverted to its primitive state in the myriapods. Each of these explanations would require an identical complex process.

Furthermore, this gene has translocated to a position remote from the original one. So the process is not a simple exchange of positions between neighbouring genes, nor does it involve genes adjacent to a large non-coding region, either of which might increase the frequency of gene rearrangement.

Another argument against the change being convergent is the infrequency of rearrangements among arthropod mitochondrial DNAs. Complete arrangements of all 37 mitochondrial genes have been determined for six arthropod genera: one chelicerate (*Limulus*[2]); one crustacean (*Artemia*[3]); and four insect (*Drosophila*[4], *Locusta*[5], *Anopheles*[6], *Apis*[7]). The *Drosophila* arrangement differs from that of *Limulus* only in the location of *L(UUR)*, from that of *Artemia* only in the location of the tRNA gene block *I–Q*, and from those of *Locusta*, *Anopheles*, and *Apis* by one, two and eight tRNA translocations, respectively. There is no evidence of 'hot spots' for tRNA gene translocations in these genera, although the sample size is small.

Rearrangements of the 37 genes typical of metazoan mtDNA appear to be unique, rare events, unlikely to be duplicated by convergence, stable once they have occurred, and easily recognized because of the homology of mitochondrial genes across the Metazoa[1]. Our phylogenetic interpretation requires no convergence in any taxon for which data are available (more than 200 taxa representing 8 phyla).

We believe that this synapomorphy

Nature © Macmillan Publishers Ltd 1998