in going from the typical polyenic geometry to the charge-separated state geometry (*1*).

Likewise, the geometry relaxations are connected to changes in the electronic structure. For instance, the formation of polarons and bipolarons in a polymer such as polythiophene upon doping provokes the appearance of localized electronic states in the gap (and thus new optical absorptions), which results in effects exploitable in electrochromic displays (polythiophene is red in the neutral state and green-blue in the oxidized state). Polarons and bipolarons can also be formed upon photoexcitation and have been suggested to lead to enhanced third-order nonlinear optical responses in the excited state (*7*). Major efforts are also devoted to the development of π-conjugated polymers that would possess low intrinsic band gaps and thus afford large electrical conductivities without doping. One way to achieve this goal is to control the molecular structure in such a way as to reduce as much as possible the degree of bond length alternation along the backbone; such an approach has been successfully applied to π-conjugated polymers that are based on main chain donor and acceptor moieties and present band gaps as low as 0.5 eV (*8*).

In their work, Marder *et al.* (*1*) show how modifications in molecular geometry can be used to tune and optimize the response (Fig. 2). As is the case for the electrical properties, the optimization of the nonlinear optical response is possible because the geometry evolution is directly connected to an important evolution in the electronic structure (*9*). The magnitude of the effect and its precise control [which may be achievable within supramolecular structures such as proteins (*10*)] makes geometry an essential parameter we must take into account when dealing with the electrical or the second-order, as well as third-order (*11*, *12*), nonlinear optical properties of π-conjugated molecules and macromolecules.

References

1. S. R. Marder, L.-T. Cheng, B. G. Tiemann, A. C. Friedli, M. Blanchard-Desce, *Science* **263**, 511 (1994).
2. J. P. Hermann and J. Ducuing, *J. Appl. Phys.* **45**, 5100 (1974); C. Sauteret *et al.*, *Phys. Rev. Lett.* **36**, 956 (1976).
3. C. K. Chiang *et al.*, *Phys. Rev. Lett.* **39**, 1098 (1977).
4. J. H. Burroughes *et al.*, *Nature* **347**, 539 (1990).
5. L. Salem, *Molecular Orbital Theory of Conjugated Systems* (Benjamin, New York, 1966).
6. A. J. Heeger, S. Kivelson, J. R. Schrieffer, W. P. Su, *Rev. Mod. Phys.* **60**, 781 (1988); J. L. Brédas and G. B. Street, *Acc. Chem. Res.* **18**, 309 (1985).
7. L. Yang *et al.*, *Synth. Met.* **41–43**, 3197 (1991).
8. E. E. Havinga, W. ten Hoeve, H. Wynberg, *ibid.* **55–57**, 299 (1993).
9. F. Meyers *et al.*, in preparation; S. R. Marder, D. N. Beratan, L.-T. Cheng, *Science* **252**, 103 (1991).
10. K. Clays *et al.*, *Science* **262**, 1419 (1993).
11. S. R. Marder *et al.*, *ibid.* **261**, 186 (1993).
12. J. L. Brédas, C. Adant, P. Tackx, A. Persoons, B. M. Pierce, *Chem. Rev.*, in press.

α-Helical Coiled Coils: More Facts and Better Predictions

Carolyn Cohen and David A. D. Parry

The α-helical coiled-coil structure is finally in fashion. Previously, only fibrous protein enthusiasts—considered to be the remnants of an *ancien régime*—interested themselves in large, supposedly dull proteins such as keratin, myosin, and fibrinogen that have coiled-coil rod domain structures. Two events changed this situation—the first slowly, the second almost immediately. By the mid-80s, it had become apparent that α-helical coiled coils were far more widespread in protein structures than had been imagined. The presence of a seven–amino acid (heptad) repeat in the sequence of a protein [conventionally referred to as positions *a* to *g*, where *a* and *d* are generally apolar] (*1*) and its implication of a coiled coil–like structure provided a means to recognize tertiary structure from primary structure alone—simply by inspection (albeit with an informed eye). Statistical measures to detect coiled coils in amino acid sequences have strengthened this approach (*2*). Thanks to cDNA technology, many new sequences became available, and the special features of the coiled coil could be recognized in diverse proteins (*3*).

The second event occurred but 2 years ago: This was the determination of the x-ray structure (to 1.8 Å resolution) of the 33-residue leucine zipper portion of the yeast transcription factor GCN4 (*4*). In fact, the discovery of leucine zippers has led to the rediscovery of coiled coils! We can now see in marvelous detail the physical basis for many of the inferences about coiled-coil structure drawn from low-resolution x-ray crystallographic structures (*5*), sequence analysis, and model-building of proteins such as tropomyosin and myosin (*6*). Subsequently other high-resolution structures of coiled-coil proteins—both native and designed—have become available, and these too are enabling us to improve current predictive methods.

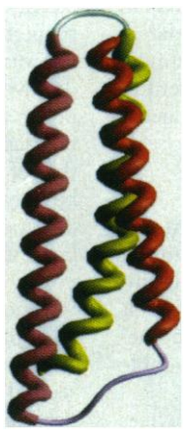

Fig. 1. A spectrin repeat. [Adapted from (*12*)]

The first step in predicting coiled coils from sequence is to localize the coiled-coil regions in the sequences—by visual inspection or with the statistical method of Lupas and co-workers (*2*). Various sequence comparison techniques also reveal stretches of coiled coil. The fast Fourier transform technique (*7*) (which determines regularities in the positions of charged or apolar residues) is especially valuable for the long α fibrous proteins. Together, these methods provide a one-dimensional map of the key α-helical features of a protein.

Next, we need to find out where the helices begin and end. Here we can use the fact that Gly and Pro, respectively, often terminate and initiate α helices. The Lupas method (based on a statistical comparison of a sequence with a database of known coiled coils) can be useful if the helices are not too short. Another standard approach to this problem is the use of so-called N- and C-Cap criteria, developed by Richardson and Richardson and Presta and Rose (*8*), which identify residues that tend to be close to the amino and carboxyl termini of an α helix.

The specificity in the packing of α helices (parallel, antiparallel, and the relative axial stagger between them) appears to be governed by ionic interactions, largely between oppositely charged residues in positions *e* and *g* of adjacent α helices (*9*) and not by apolar interactions between residues in the *a* and *d* positions. By maximizing the number of potential interhelix ionic interactions, both the axial staggers between α-helical chains and even the sense of chain connectivity (whether left- or right-handed) in an α-helical bundle may be predicted. Additional information can be obtained from chemical constraints such as the positions of disulfide bonds and from molecular morphology deduced by electron microscopy and nuclear magnetic resonance.

These methods were put to the test recently with the repeat motif of α-spectrin, α-actinin, and dystrophin (*10*). These three proteins are members of the spectrin superfamily: Each contains an α helix–rich rod domain characterized by multiple repeats of a sequence about 110 residues in length. The boundaries of the repeats in α-spectrin were established biochemically, and then a single repeat was expressed that

C. Cohen is at the Rosenstiel Basic Medical Sciences Research Center, Brandeis University, Waltham, MA 02254–9110. D. A. D. Parry is in the Department of Physics and Biophysics, Massey University, Palmerston North, New Zealand.

folded into a stable and crystallizable structure (*11*). The structure of this repeat to atomic resolution was reported in the 24 December issue of *Science* (*12*). The structure is a remarkable one (see Fig. 1) that predicts new features of spectrin's flexibility and that accounts for a number of previously puzzling features in native spectrin.

The happy result was that the model developed for the α-spectrin repeat with the methods described above (*10*) is in close agreement with many features of the crystal structure. The structural motif is indeed a left-handed coiled coil with left-handed chain connectivity. The repeat length is close to 50 Å, and the axial pairing of the two Trp residues are all as predicted. The motif is also stabilized by apolar interactions and many interhelical ionic interactions as suggested by the model. An unusual feature of the crystal structure, however, is that perfect coiled-coil packing is not maintained between one of the helices and the other two. Instead, ridges-into-grooves packing, common in globular proteins, is observed. Correspondingly, the prediction was off by three residues (4.5 Å) in the relative axial displacement between one helix and the other two.

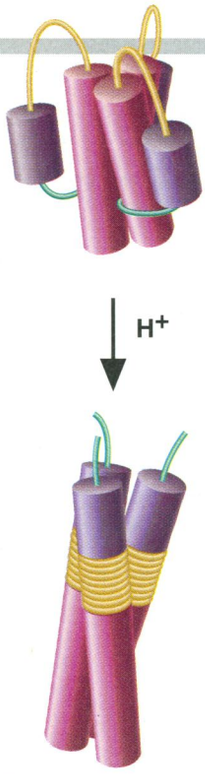

Fig. 2. Hemagglutinin (HA2 region): Cartoon depicting structural transition in HA2 trimeric coiled-coil stem [Adapted from (*16*)]

Why was the model structure in such good agreement with the real one? Undoubtedly the α-spectrin conformation is a very simple one with well-defined and multiple structural elements—chiefly α-helical. Such features in proteins, however, remain the exception rather than the rule (other than for some fibrous proteins), and it is unrealistic to expect that this success will open the floodgate for many others. Nonetheless, the heptad substructure does indeed place powerful constraints on tertiary structure options. Cytokine sequences analyzed by a similar approach contain a heptad substructure and were predicted to comprise a four–α-helical motif as a core to their tertiary structures (*13*)—and, when crystallized and solved, were indeed shown to have this structure.

Results from the leucine zipper structures are also strengthening our prophetic abilities. The GCN4 structure (*4*) gave us a detailed view of the side chain interactions in the heptad repeat, many of which had been deduced previously from analyses of the fibrous proteins. The structure also revealed something new: We used to think that the side chains in the *a* and *d* positions were equivalently directed into the dimer interface. The GCN4 structure showed that this is not so: A side chain in the *d* position (where Leu is favored) points directly into the interface, whereas a side chain in the *a* position points out from the interface. This fact explains the observed preference in the fibrous proteins for β-branched side chains (Ile or Val) in the *a* position where they can redirect part of their side chain back into the core. It also explains the tolerance in all coiled coils for the Lys residues in the *a* but not the *d* positions. Only in the *a* position of a two-stranded rope can the charged end of Lys project out of the apolar interface.

An important development from this work is that the structure of a number of GCN4 mutants has been solved (*14*). The beauty of these zipper mutants is the large and comprehensible effects that a few simple substitutions can have: The presence of branching in the amino acids of positions *a* and *d* determines the state of oligomerization of the α-helical chains. Thus, Ile in *a* and Leu in *d* yields the zipper-like two-stranded structure; Ile in both *a* and *d* yields a three-stranded coiled coil; and Leu in *a* and Ile in *d* give rise to a four-stranded assembly. Thus, β-branched residues tend to destabilize particular modes of packing. These results are consistent with the sequences of known two- and three-stranded α-fibrous proteins (*9*, *15*).

Perhaps one of the most dramatic examples of the application of coiled-coil structure predictions is that of the three-stranded parallel coiled coil that forms the fibrous stem of hemagglutinin (HA) in influenza virus (*16*, *17*). This structure displays a new and dynamic feature. It was known that some kind of transition takes place in the structure of HA in order to promote membrane fusion and cell entry and that this transition occurs at a pH of ~5, the same pH at which membranes fuse in the endosome. Carr and Kim (*16*) have now proposed—and essentially proven—that the 80 Å coiled-coil stem (formed from the three HA2 region light chains) that exists at neutral pH would be extended at pH 5 at its amino-terminal end to form a longer coiled coil (see Fig. 2). Thus, a region in the HA2 light chain that forms a loop at neutral pH (*17*) would become part of the three-stranded coiled coil at lower pH. Results from Don Wiley's laboratory (*18*) on the crystal structure of HA—essentially at this lower pH—now show that this is indeed true. But in addition, the lower leg of the HA2 trimer which actually splays out in the crystal structure at pH 7 bends around to pack against the triple-stranded coiled coil! This is a marvelous example of coiled-coil dynamics. And it is a marvelous example of the power of coiled-coil search methods as well. In fact, using the method of Lupas *et al.* (*2*), Carr and Kim had noticed that a region in the HA2 sequence that was predicted to have a very high α-helical coiled-coil potential was in fact a loop at the neutral pH of the crystal structure (*17*). The fact that this loop could also form a coiled coil was foreshadowed by earlier work. Using simple visual inspection, Ward and Dopheide (*19*) showed that this loop region had high coiled-coil potential and that the HA2 chains were likely to form a parallel triple-stranded coiled coil.

These results are but the beginning. Improved algorithms are being developed to detect, for example, short or buried coiled coils as well as helix ends. Exploitation of the rapidly growing crystal structure databases is providing additional conformational keys to coiled-coil sequences (*20*). And energy minimization techniques are being applied to coiled-coil model structures. We may soon have the information necessary to construct the Rosetta Stone of sequence-to-structure for this exceptional class of proteins.

References

1. F. H. C. Crick, *Acta Crystallogr.* **6**, 689 (1953).
2. D. A. D. Parry, *Biosci. Rep.* **2**, 1017 (1982); A. Lupas, M. Van Dyke, J. Stock, *Science* **252**, 1162 (1991).
3. C. Cohen and D. A. D. Parry, *Trends Biochem. Sci.* **11**, 245 (1986); *Proteins* **7**, 1 (1990).
4. E. K. O'Shea,J. D. Klemm, P. S. Kim, T. Alber, *Science* **254**, 539 (1991).
5. G. N. Phillips Jr., J. P. Fillers, C. Cohen, *J. Mol. Biol.* **192**, 111 (1986).
6. A. D. McLachlan and M. Stewart, *ibid.* **98**, 293 (1975).
7. ———, *ibid.* **103**, 271 (1976).
8. J. S. Richardson and D. C. Richardson, *Science* **240**, 1648 (1988); L. G. Presta and G. D. Rose, *ibid.*, p. 1632.
9. J. F. Conway and D. A. D. Parry, *Int. J. Biol. Macromol.* **12**, 328 (1990).
10. D. A. D. Parry, T. W. Dixon, C. Cohen, *Biophysics J.* **61**, 858 (1992); D. A. D. Parry and C. Cohen, in *The Living Cell in Four Dimensions*, G. Paillotin, Ed. (American Institute of Physics, New York, 1991), p. 367.
11. E. Winograd, D. Hume, D. Branton, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 10788 (1991).
12. Y. Yan *et al.*, *Science* **262**, 2027 (1993).
13. D. A. D. Parry, E. Minasian, S. J. Leach, *J. Mol. Recognition* **4**, 63 (1991).
14. P. B. Harbury, T. Zang, P. S. Kim, T. Alber, *Science* **262**, 1401 (1993).
15. J. F. Conway and D. A. D. Parry, *Int. J. Biol. Macromol.* **13**, 14 (1991).
16. C. M. Carr and P. S. Kim, *Cell* **73**, 823 (1993).
17. I. A. Wilson, J. J. Skehel, D. C. Wiley, *Nature* **289**, 366 (1981).
18. D. C. Wiley, unpublished data.
19. C. W. Ward and T. A. Dopheide, *Aust. J. Biol. Sci.* **33**, 449 (1980).
20. J. Seo and C. Cohen, *Proteins* **15**, 223 (1993).

21. Supported by grants to C.C. from the National Institutes of Health (AR17346), the National Science Foundation (MCB90-04746), and the Muscular Dystrophy Association.