

**Plate 2.1** (a) Crystal structure of a complex of glutaminyl tRNA with glutaminyl-tRNA synthetase (Sherlin *et al.* 2000, *Journal of Molecular Biology*, **299**: 431–46). (b) Crystal structure of the dimeric lac repressor protein bound to DNA (Bell and Lewis 2001, *Journal of Molecular Biology*, **312**: 921–6). Image reproduced with permission from NDB (Berman *et al.* 1992). ID number PRO025 and PD0250.

**Plate 2.2** (a) Hierarchical clustering of the amino acids performed using CLUTO (Karypis 2000, University of Minnesota technical report #02–017), as described in Section 2.6.2. (b) Where clusters fail: the Venn diagram illustrates the overlapping properties of amino acids. The color scheme is the one implemented in the CINEMA sequence alignment editor (Parry-Smith *et al.* 1998, *Gene*, **221**: GC57–GC63): red for acidic; blue for basic; green for polar neutral; purple for aromatic; black for aliphatic; yellow for cysteine (disulfide potential); and brown for proline and glycine (which have unique structural properties, especially in helices).

**Plate 3.1** The pattern of bases observed at informative sites in human mitochondrial genomes from many different ethnic groups and the phylogenetic tree deduced from this information. Reproduced from Ingman *et al.* (2000, *Nature*, **408**: 708–13), with permission from Nature Publishing Group.

**Plate 9.1** (a) Logo for the prion block depicted in Fig. 9.3. The completely conserved glycine and highly conserved proline residues that form the GPG anchor for this block are clearly visible toward the center of the logo. (b) Excerpts from sequence alignments of vertebrate visual receptors. The coloring of the amino acids follows the scheme shown in Plate 2.2(b). Part of the typical signature of rhodopsin (denoted OPSD) sequences is shown. (c) Part of the typical signature of red/green opsins (denoted OPSR and OPSG). Note that the red and green sequences are highly similar, such that it is almost impossible to separate them using typical sequence analysis methods – the central motif, N-P-G-Y-[AP]-[FW]-H-P-L, which includes a conserved histidine, is particularly striking. More startling is the fact that within the rhodopsin alignment are several “rogue” sequences, namely of green, blue (OPSB), and purple (OPSU) opsins, that more closely resemble the rhodopsin sequence signature than they do their own pigment sequences – in particular, the chicken (CHICK) and goldfish (CARAU) green pigments do not contain the N-P-G-Y-[AP]-[FW]-H-P-L motif characteristic of green sequences.

**Plate 10.1** Application of a three-state M1–M2 model to the *Bacillus subtilis* genome reveals an atypical segment (3,463–3,467 kb, underlined) surrounded by ABC transporter gene duplication (thin black arrows). Filled arrows represent genes of known function, empty arrows, those of unknown function, and red hairpins represent transcriptional terminators. The colored curves show the probabilities of being in each of the three hidden states at each point in the sequence. The magenta state matches genes on the (+) strand, whereas cyan matches genes on the (–) strand. The black state fits either intergenic regions or atypical genes. Reproduced from Nicolas *et al.* (2002, *Nucleic Acids Research*, **30**: 1418–26) with permission of Oxford University Press.

**Plate 11.1** The secondary structure of SSU rRNA in *E. coli*. The color scheme shows the degree of variability of the sequence across the bacterial domain. Category 0 (purple) sites are completely conserved. Categories 1 to 5 range from very slowly evolving (blue) to rapidly evolving (red). The gray sites are present in less than 25% of the species considered, hence no measure of evolutionary rate was made. Reproduced with permission from the European Ribosomal RNA database <http://oberon.fvms.ugent.be:8080/rRNA/index.html>

**Plate 12.1** The genome of *Rickettsia conorii*, reproduced from Ogata *et al.* (2001), Copyright 2001 AAAS. (a) The outer circle gives the nucleotide positions in bases measured anticlockwise from the origin of replication. The second and third circles show the positions of ORFs on the plus and minus strands, respectively. The colors used indicate different functional classes of gene. The arrows in the fourth and fifth circles show the positions of tRNA genes, and the three black arrows show rRNAs. The sixth and seventh circles indicate short repeated sequences. The eighth circle shows G–C skew. The genome is found to be largely colinear with *R. prowazekii*, except for the shaded sector on the lower left. (b) Three distinct regions of the *R. conorii* genome aligned with homologous regions of the *R. prowazekii* genome. Most of the genes are present in the same order in both species. The top and middle sections show split genes in *R. prowazekii* and *R. conorii*, respectively. The bottom section contains a non-coding region of *R. prowazekii* that shows some sequence similarity to a functional gene (*rOmpA*) in *R. conorii*.

**Plate 12.2** Comparisons of gene order in pairs of animal mitochondrial genomes constructed using the OGRE database (Jameson *et al.* 2003, *Nucleic Acids Research*, **31**: 202–6). <http://ogre.mcmaster.ca>

**Plate 13.1** Hierarchical clustering of gene expression data depicting the relationship between 96 samples of normal and malignant lymphocytes (reproduced from Alizadeh *et al.* 2000, *Nature*, **403**: 503–11, with permission of Nature Publishing Group).

**Plate 13.2** Analysis of microarray data from 60 cancer cell lines (reproduced from Slonim 2002, *Nature Genetics* supplement, **32**: 502–7, with permission of Nature Publishing Group). (a) Projection of the samples onto the first three principal component axes reveals a certain amount of clustering between some types of cancer. (b) Hierarchical clustering of the same data. S and R indicate samples that are extremely sensitive or resistant to the drug cytochalasin D.

**Plate 13.3** Analysis of microarray data from the yeast cell cycle using SVD (reproduced from Alter *et al.* 2000, *Proceedings of the National Academy of Sciences USA*, **97**: 10101–6, Copyright 2000 National Academy of Sciences, USA). (a) Arrays 1–12 correspond to successive time-points of the cell. The points for the 12 arrays proceed in an almost circular pattern around the space defined by the first 2 eigenarrays (denoted  $|\alpha_1\rangle$  and  $|\alpha_2\rangle$ ). (b) Each of 784 cell-cycle-regulated genes is shown as a point in the space of the first two eigengenes (denoted  $|\gamma_1\rangle$  and  $|\gamma_2\rangle$ ). The array points are color-coded according to the five cell-cycle stages: M/G<sub>1</sub> (yellow), G<sub>1</sub> (green), S (blue), S/G<sub>2</sub> (red), and G<sub>2</sub>/M (orange). The gene points are colored according to the stage at which they are significantly up-regulated.

**Plate 13.4** Map of protein–protein interactions in yeast (Jeong *et al.* 2001, *Nature*, **411**: 41–2, with permission of Nature Publishing Group). Each node is a protein, and each link is a physical protein–protein interaction, identified usually by Y2H experiments. The color of a node indicates the effect of deleting the corresponding protein (red, lethal; green, non-lethal; orange, slow growth; yellow, unknown).