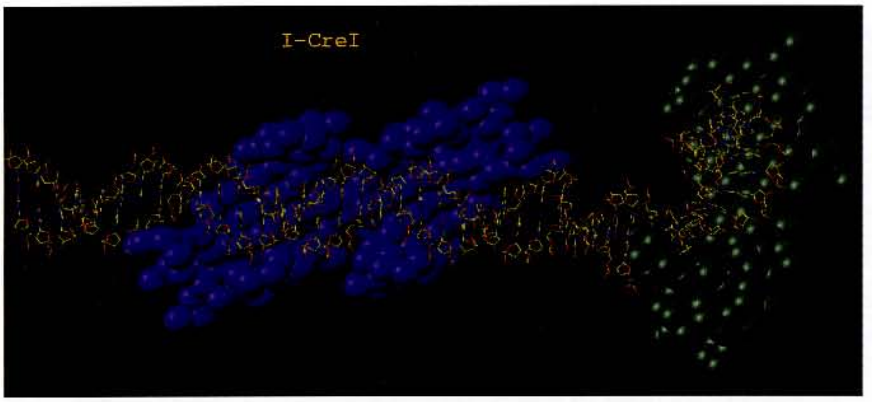


Fig. 1 A comparison of the I-*Cre*I and TBP saddles (viewed from below) and their alignment with respect to DNA[2]. The saddle shapes were generated using Cα coordinates. (Coordinates of I-*Cre*I were kindly provided by B.L. Stoddard).

has been suggested that cleavage pattern is the primary constraint on the conformation of endonucleases[12], and it would be interesting to determine whether type II restriction endonucleases that yield 4 base 3' overhangs, such as *Pst*I and *Sac*I, contain structural features in common with the homing endonucleases.

Future prospects

The I-*Cre*I structure represents an important landmark in the study of homing endonucleases and provides a stepping stone for further structural and biochemical studies. The structure of the I-*Cre*I–DNA complex promises to reveal the basis of its DNA selectivity and to lend further information on the residues comprising the active site. Structures of endonucleases containing two copies of the LAGLIDADG motif (PI-*Sce*I and PI-*Psp*I, for instance) will test the notion that these enzymes are folded as pseudo-twofold symmetric structures, perhaps resembling the I-*Cre*I dimer. However, most of these homing endonucleases recognize markedly asymmetric DNA sequences, the basis of which is not yet clear.

The presence of LAGLIDADG motifs in both endonucleases and maturases suggests an intriguing link between the two enzyme activities that will become better understood with the determination of the structure of a maturase. Particularly interesting in this regard are proteins that display both endonuclease and maturase activities in a single molecule[9,13]. Although most homing endonucleases contain the LAGLIDADG motif, some endonucleases lacking the LAGLIDADG motif are radically different in their recognition and cleavage properties from I-*Cre*I. For instance, I-*Tev*I and I-*Tev*II from bacteriophage T4, cleave their DNA substrates a short distance away from their intron insertion sites and yield 3' overhangs with a 2 bp stagger. Unlike I-*Cre*I, both enzymes make predominantly minor groove contacts and induce substantial distortions of their DNA substrates[14,15].

Homing endonucleases have evolved a variety of mechanisms for invading target DNA sites. The I-*Cre*I structure signals the beginning of an exciting journey into the structural realm of these fascinating enzymes.

Aneel K. Aggarwal is with the Department of Physiology and Biophysics, Mount Sinai School of Medicine, One Gustave L. Levy Place, New York, New York 10029, USA
email: aggarwal@inka.mssm.edu

1. Mueller, J.E., Bryk, M., Loizos, N. & Belfort, M. in *Nucleases* (eds. Linn, S.M., Lloyd, R.S. & Roberts, R.J.) 111–143 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1993).
2. Heath, P.J., Stephens, K.M., Monnat, R.J. & Stoddard, B.L. *Nature Struct. Biol.* 468–476 (1997).
3. Perler, F.B. *et al. Nucl. Acids Res.* **22**, 1125–1127 (1994).
4. Aggarwal, A.K. *Curr. Opin. Struct. Biol.* **5**, 11–19 (1995).
5. Kim, Y., Geiger, J.H., Hahn, S. & Sigler, P.B. *Nature* **365**, 512–520 (1993).
6. Kim, J.L., Nikolov, D.B. & Burley, S.K. *Nature* **365**, 520–527 (1993).
7. Endrizzi, J.A., Cronk, J.D., Wang, W., Crabtree, G.R. & Alber, T. *Science* **268**, 556–559 (1995).
8. Miller, J., Mclachlan, A.D. & Klug, A. *EMBO J.* **4**, 1609–1614 (1985).
9. Henke, R.M., Butow, R.A. & Perlman, P.S. *EMBO J.* **14**, 5094–5099 (1995).
10. Gimble, F.S. & Stephens, B.W. *J. Biol. Chem.* **270**, 5849–5856 (1995).
11. Anderson, J.E. *Curr. Opin. Struct. Biol.* **3**, 24–30 (1993).
12. Newman, M., Strzelecka, T., Dorner, L.F., Schildkraut, I. & Aggarwal, A.K. *Nature* **368**, 660–664 (1994).
13. Szczepanek, T. & Lazowska, J. *EMBO J.* **15**, 3758–3767 (1996).
14. Mueller, J.E., Smith, D., Bryk, M. & Belfort, M. *EMBO J.* **14**, 5724–5735 (1995).
15. Loizos, N., Silva, G.H. & Belfort, M. *J. Mol. Biol.* **255**, 412–424 (1996).

Nature's transitory covalent bond

Ronald T. Raines

The chemistry of sulphur enables a disulphide bond to be transitory—guiding the proper folding of a protein, but not appearing in the native structure.

The segments of proteins and nucleic acids are linked by covalent bonds. Conversely, the native three-dimensional structures of biopolymers are held together by noncovalent interactions. There is but one common exception to this rule: covalent bonds between cysteine residues stabilize the three-dimensional structures of many proteins. Working with the tailspike endorhamnosidase from bacteriophage P22, Robinson and King have discovered another anomaly. On page 450 of this issue, they report that a disulphide bond can also be transitory—guiding the proper folding of a protein, but not appearing in the native structure[1]. The existence of non-native disulphide bonds in protein folding pathways had been observed two decades ago[2]. Only now, however, has the truly transitory nature of the disulphide bond been revealed.

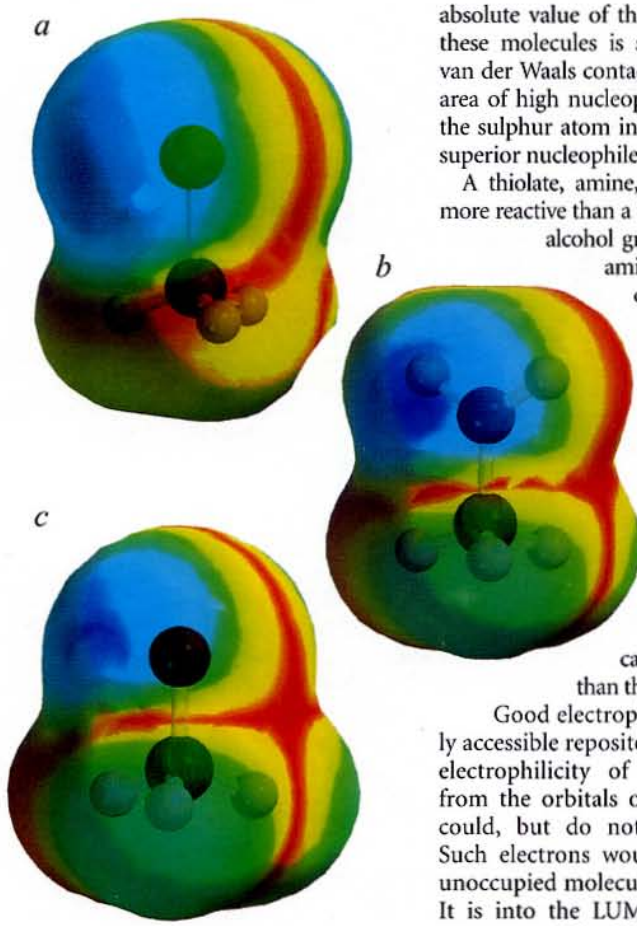


Fig. 1 Highest occupied molecular orbitals (HOMO's) of model nucleophiles. ***a***, Methane thiolate. ***b***, Methylamine. ***c***, Methoxide ion. The absolute value of the HOMO is superimposed onto a van der Waals contact surface. Colours near blue depict high nucleophilicity. Colours near red depict low nucleophilicity. The large blue area in methane thiolate shows that the sulphur atom in this molecule is a superior nucleophile. *Ab initio* molecular orbital calculations were performed with the 6-31G* basis set using the program MACSPARTAN 1.1.4.

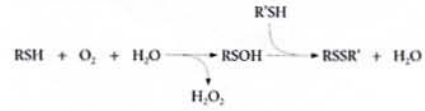
Kinetics of disulphide bond formation

The unique role of the disulphide bond in structural biology derives from the special chemistry of sulphur. Chemical reactions occur readily when a good nucleophile encounters a good electrophile. The nucleophilicity of a molecule ensues from the most reactive electrons in that molecule. These electrons lie in the highest occupied molecular orbital or HOMO. The three atoms that often serve as nucleophiles in proteins are sulphur, nitrogen and oxygen. The nucleophilicity of these atoms is apparent in the HOMO's of model aliphatic molecules, such as methane thiolate (CH_3S^-), methylamine (CH_3NH_2), and methoxide ion (CH_3O^-). In Fig. 1, the absolute value of the HOMO for each of these molecules is superimposed onto a van der Waals contact surface. The (blue) area of high nucleophilicity indicates that the sulphur atom in methane thiolate is a superior nucleophile.

A thiolate, amine, or alkoxide group is more reactive than a thiol, ammonium, and alcohol groups respectively. The amino acid residues that contain the functional groups in Fig. 1 differ in their side-chain pK_a values: steine, 8.7 < lysine, 5 < serine, 14.2. not only is a cysie residue armed ith a superior ucleophile, but the ability of cysteine to employ its nucleophile at physiological pH is also greater than that of other residues.

Good electrophiles provide a readily accessible repository for electrons. The electrophilicity of a molecule ensues from the orbitals of lowest energy that could, but do not, contain electrons. Such electrons would lie in the lowest unoccupied molecular orbital or LUMO. It is into the LUMO, the energetically most accessible of the unfilled molecular orbitals, that electrons from the nucleophile will go. For the single bonds found in proteins, the LUMO is largely a combination of σ* atomic orbitals. In Fig. 2, the energies of σ* orbitals are listed for the single bonds found in proteins. The sulphur of dimethyl disulphide (CH_3SSCH_3) has the lowest energy (and therefore most accessible) σ* orbital of all of the atoms found in proteins. In Fig. 3, the absolute value of the entire LUMO for dimethyl disulphide and ethanol (for comparison) is superimposed onto a van der Waals contact surface. The (blue) area of high electrophilicity in dimethyl disulphide provides a large target for an incoming nucleophile[3]. By comparison, ethanol has no significant electrophilic centre, aside from its acidic proton.

The analysis above provides a rationale for the rapid reaction of thiols with disulphide bonds. But how does a disulphide bond form from two thiol groups? In some of the new experiments on tailspike protein[1], as in Anfinsen's early experiments on ribonuclease A[4,5], no disulphide bonds were present initially either in the unfolded protein or in solution components. The formation of a disulphide bond from two thiol groups is an oxidation reaction, and hence another molecule must undergo reduction. Molecular oxygen can serve as that molecule. Disulphide bond formation then occurs in two steps, through an intermediate sulphenic acid (RSOH):



The LUMO's of methyl sulphenic acid are depicted in Fig. 3*b*. The blue area in methylsulphenic acid, like that in dimethyl disulphide, is indicative of a good electrophile.

The nucleophilicity of thiolates and the electrophilicity of sulphenic acids and disulphide bonds allow disulphide bonds to form and interchange rapidly, without the need for catalysis. These properties lead to a low kinetic barrier for the formation of disulphide bonds. Why then do all cysteine residues not unite in pairs to form disulphide bonds? Moreover, why is the disulphide bond of Robinson and King transitory?

Thermodynamics of disulphide bond formation

Disulphide bonds exist only if reduction potential ($E^{o\prime}$) values allow. In the cytosol of an *Escherichia coli* cell, $E^{o\prime}$ is –0.27 V[6]. How does this value compare to that of a disulphide bond? The $E^{o\prime}$ values of disulphide bonds in native proteins can differ greatly[7], even if they are in the same sequence motif[8]. For example, an intramolecular disulphide bond in the Cys–Gly–Pro–Cys sequence of *E. coli* thioredoxin has $E^{o\prime}$ of –0.27 V[9], but that in the similar Cys–Pro–His–Cys sequence of *E. coli* DsbA has $E^{o\prime}$ of –0.11 V[10]. Such differences arise from confor-

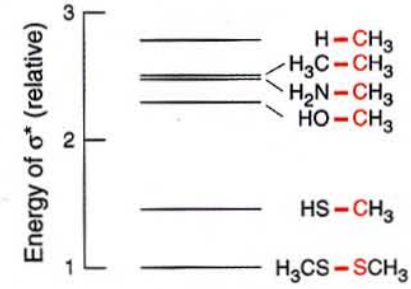


Fig. 2 Energies of σ* orbitals in model electrophiles. Geometries of the model molecules were optimized at the HF/6-31G* level of theory. Energies of the σ* orbitals for the red atom and σ bond were calculated by natural bond order (NBO) analysis[19].

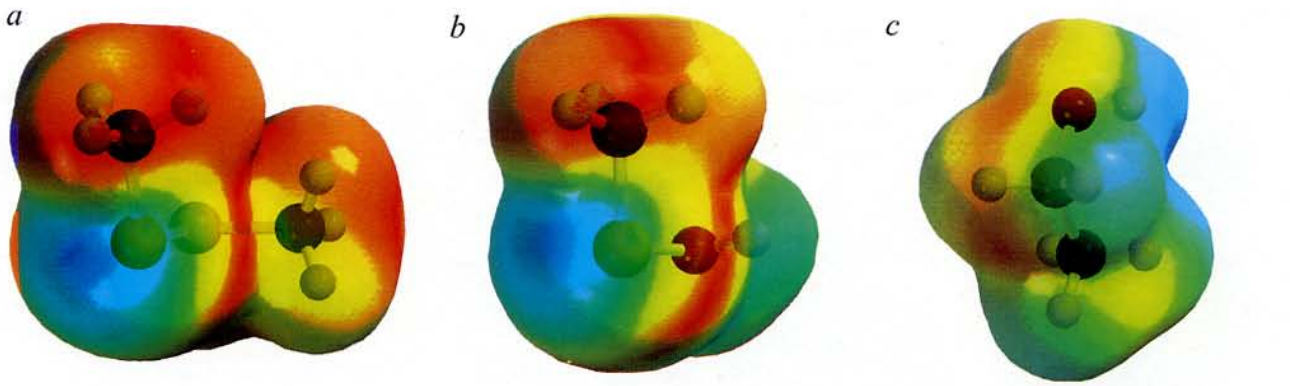


Fig. 3 Lowest unoccupied molecular orbitals (LUMO's) of model electrophiles. ***a***, Dimethyl disulphide. ***b***, Methylsulphenic acid. ***c***, Ethanol. The absolute value of the LUMO is superimposed onto a van der Waals contact surface. Colours near blue depict high electrophilicity. Colours near red depict low electrophilicity. The large blue area in dimethyl disulphide and methylsulphenic acid shows that the sulphur atoms in these molecules are superior electrophiles. *Ab initio* molecular orbital calculations were performed on each molecule with the 6-31G* basis set using the program MACSPARTAN 1.1.4.

mational and electrostatic variations effected by different three-dimensional structures. The strands of different unfolded proteins behave in a much more similar manner than do the three-dimensional structures of different native proteins. Thus, E° values in unfolded proteins can be estimated from a general statistical analysis.

The mean squared end-to-end distance ($\langle r^2 \rangle_0$) of an unfolded protein is related to the distance between monomeric units (l) and the number of bonds between monomeric units (n) as:

$$C_\infty = \left(\langle r^2 \rangle_0 / nl^2\right)_{n\to\infty} \quad (1)$$

where C_∞ is characteristic of a particular polymer[11]. For poly-L-alanine, C_∞ = 9.0 and l = 3.80 Å, which is the distance between consecutive Cα atoms[11]. So, the r.m.s. value of r (in Å) for large values of n is:

$$\langle r^2 \rangle_0^{1/2} = \sqrt{130n} \quad (2)$$

The value of $\langle r^2 \rangle_0^{1/2}$ reports on the proximity of one end of the poly-L-alanine to the other. For the radial distribution function used in this analysis[11], the mean end-to-end distance ($\langle r^2 \rangle_0$) is quite similar to $\langle r^2 \rangle_0^{1/2}$, the two being related by:

$$\langle r \rangle_0 = 2\sqrt{2/(3\pi)}\,\langle r^2 \rangle_0^{1/2} \quad (3)$$

If the two ends of the poly-L-alanine molecule were actually separate molecules, then the density of these molecules in solution (in molecules Å^{-3}) would be simply $\langle r^2 \rangle_0^{-3}$. Accordingly, the effective concentration (EC or C_{eff} in M) of one end of the poly-L-alanine molecule relative to the other is:

$$EC = \frac{0.88}{n^{3/2}} \quad (4)$$

Eq. 4 applies strictly only to the ends of an infinite strand of poly-L-alanine. Still, Eq. 4 can be used to provide an estimate of the EC of two cysteine residues separated by n residues in an unfolded protein. Recently, Kim and coworkers found that $EC = 0.694/n^{3/2}$ for α-lactalbumin[12], which is in remarkable agreement with Eq. 4. Taking the disulphide bond of oxidized glutathione ($E^{\circ\prime} = -0.252$ V)[13] as a standard, Eq. 4 and the Nernst equation can be used to derive the reduction potential (in V) of a typical disulphide bond in an unfolded protein:

$$E^{\circ\prime} = -0.25 + 0.044 \log n \quad (5)$$

Eq. 4 and 5 relate $E^{\circ\prime}$, EC, and n. These relationships are depicted in Fig. 4.

What then is the basis for the 'now you see it, now you don't' disulphide bond of Robinson and King? All fully unfolded proteins are incapable of forming stable disulphide bonds in the cytosol of a prokaryotic cell (Fig. 4). Such bonds simply do not have a low enough reduction potential to remain intact. In other words, two thiols in an unfolded protein do not have a high enough EC to one another to form a stable disulphide bond. Yet as a protein folds, the ECs of some cysteine residues relative to others is likely to increase. For example, the open circles in Fig. 4 tell the redox story of a pair of typical cysteine residues that form a disulphide bond in the ER of a human cell. Similarly, the closed circles in Fig. 4 could report on Cys 613 and Cys 635 of tailspike protein, which apparently cross the thermodynamic threshold for the existence of a stable disulphide bond in the redox environment of the *E. coli* cytosol. A disulphide bond forms, and this covalent link facilitates subsequent steps in the folding pathway. But as the folding tailspike protein approaches its native state, the EC ricochets, dropping below the threshold for stability. The disulphide bond then suffers reduction by solution thiols.

Envoi

How common are ephemeral interactions during protein folding? For BPTI, Weissman and Kim found that "the same interactions that stabilize the final folded structure also guide the protein in attain-

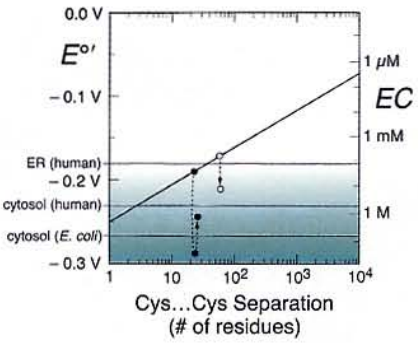


Fig. 4 Thermodynamics of disulphide bond formation in an unfolded protein. $E^{\circ\prime}$ is the reduction potential of a disulphide bond between two cysteine residues. EC is the effective concentration of the two cysteine residues to one another, and n is the number of amino acid residues that separate the cysteine residues. The line is based on Eq. 4 and 5, which are derived from polymer theory[11] and empirical constants[11,13]. Cellular $E^{\circ\prime}$ values are from ref. 6 and 20. Open circles depict the redox history of two typical cysteine residues (n = 60) that form a stable disulphide bond in the human ER. Closed circles depict a possible redox history of Cys 613 and Cys 635 (n = 22) during the folding of tailspike protein in the *E. coli* cytosol[1].

ing this structure"[14]. In other words, the *ECs* of residues that interact in the native state increase monotonically throughout the folding process. This circumstance is reminiscent of the venerable Hammond postulate[15], which has been used by Matouschek and Fersht to describe the folding of barnase and chymotrypsin inhibitor 2[16,17]. The folding of tailspike protein violates this sentiment. Instead, its folding is in part 'anti-Hammond'—an interaction (that is, the transitory disulphide bond) is vital to an intermediate but not the product. Anti-Hammond behaviour had been observed for noncovalent interactions in barnase[18]. Some residues in that protein form stronger noncovalent interactions in intermediates than in the product. Never, however, has a *covalent* bond been seen to be so transitory as in the tailspike protein of Robinson and King.

Ronald T. Raines is with the Department of Biochemistry, University of Wisconsin–Madison, Madison, Wisconsin 53706-1569 USA
e-mail: raines@biochem.wisc.edu

1. Robinson, A.S. & King, J. *Nature Struct. Biol.* **4**, 450–455 (1997).
2. Creighton, T.E. *J. Mol. Biol.* **113**, 275–293 (1977).
3. Rosenfield, R.E., Jr., Parthasarathy, R. & Dunitz, J.D. *J. Am. Chem. Soc.* **99**, 4860–4862 (1977).
4. Anfinsen, C.B., Haber, E., Sela, M. & White, F.H., Jr. *Proc. Natl. Acad. Sci. USA* **47**, 1309–1314 (1961).
5. Anfinsen, C.B. & Haber, E. *J. Biol. Chem.* **236**, 1361–1363 (1961).
6. Gilbert, H.F. *Adv. Enzymol.* **63**, 69–172 (1990).
7. Gilbert, H.F. *Methods Enzymol.* **251**, 8–28 (1995).
8. Chivers, P.T., Prehoda, K.E. & Raines, R.T. *Biochemistry* **36**, 4061–4066 (1997).
9. Moore, E.C., Reichard, P. & Thelander, L. J. Biol. Chem. **239**, 3445–3452 (1964).
10. Grauschopf, U. *et al. Cell* **83**, 947–955 (1995).
11. Flory, P.J. Statistical Mechanics of Chain Molecules (Wiley, New York, NY, 1969).
12. Peng, Z.Y., Wu, L.C. & Kim, P.S. *Biochemistry* **34**, 3248–3252 (1995).
13. Lees, W.J. & Whitesides, G.M. *J. Org. Chem.* **58**, 642–647 (1993).
14. Weissman, J.S. & Kim, P.S. *Science* **253**, 1386–1393 (1991).
15. Hammond, G.S. *J. Am. Chem. Soc.* **77**, 334–338 (1955).
16. Matouschek, A. & Fersht, A.R. *Proc. Natl. Acad. Sci. USA* **90**, 7814–7818 (1993).
17. Matouschek, A., Otzen, D.E., Itzhaki, L.S., Jackson, S.E. & Fersht, A.R. *Biochemistry* **34**, 13656–13662 (1995).
18. Matthews, J.M. & Fersht, A.R. *Biochemistry* **34**, 6805–6814 (1995).
19. Reed, A.E., Curtiss, L.A. & Weinhold, F. *Chem. Rev.* **88**, 899–926 (1988).
20. Hwang, C., Sinskey, A.J. & Lodish, H.F. *Science* **257**, 1496–1502 (1992).

Solving the *cis/trans* paradox in the Int family of recombinases

Alfonso Mondragón

Structures of the catalytic domains of λ Int and HP1 integrase provide insight into the diversity of the Int family of recombinases, which nevertheless catalyse very similar chemical events.

Recombination is a crucial process that allows the rearrangement of genetic material. There are two types of recombination processes: homologous and site-specific recombination[1]. In the former case the recombination event is primarily mediated by large regions of homologous DNA and is helped by specific proteins, for example RecA in *Escherichia coli*. In the latter case the exchange of DNA requires specific DNA sites where the genetic material is inserted, but not long stretches of homologous DNA.

Site-specific recombination is responsible for integration and excision, inversion, and resolution, and site-specific recombinases are classified into two groups: resolvases, such as γδ resolvase (whose structure is known[2]), and integrases/transposases, such as HIV integrase, MuA transposase, the FLP recombinase from yeast, λ Int, and *Haemophilus influenzae* HP1 integrase. The structures of the catalytic domain of HIV integrase[3] and MuA transposase[4] are known, both of which represent examples of one large family of polynucleotidyl transferases. The recent structures of the catalytic domain of λ Int[5] and HP1 integrase[6] now provide the first sight of members of a second group of site-specific recombinases, the Int family of recombinases.

Integration is the process that many bacteriophages use to introduce and remove their genetic material from the bacterial chromosome. It may require the presence not only of phage proteins but also of host proteins. The prototype of this type of recombinase is the integrase protein of phage λ, Int. Int-mediated recombination requires the presence of attachment sites (*att* sites) both in the phage and bacterial chromosome. The reaction involves DNA cleavage of one strand in the phage DNA and one in the bacterial chromosome at specific locations (Fig. 1). The cleavage proceeds through a covalent phosphotyrosine intermediate linked to the 3' end of the broken DNA strand and creates two free 5' hydroxyls. After a rearrangement of the nucleoprotein complex, the free 5' hydroxyls attack the phosphotyrosine bonds to re-create the phosphodiester backbone. The free 5' hydroxyl of the phage DNA attacks the phosphotyrosine bond formed with the chromosomal DNA and *vice versa*, resulting in an exchange between the phage and bacterial DNA and forming a Holliday junction. The reaction is completed, or resolved, when a similar reaction occurs at a different site to complete the insertion or excision of the phage DNA (Fig. 1). In the case of λ Int, the reaction requires the presence of other proteins, like integration host factor (IHF), to, for example, bend the DNA in the nucleoprotein complex.

From a mechanistic point of view the way the protein recognizes, cuts, and rearranges the DNA strands is quite fascinating. Clearly there have to be some major structural changes in the complexes as the reaction proceeds. The protein not only has to cleave the DNA to form the covalent intermediate, but then the complex has to be rearranged to allow the religation with the other DNA molecule and strand exchange. The biochemistry