polymerase. Possibly the same site of polymerase is contacted in each case, but with a different portion of the repressor in the two cases.

Some current questions

Is there a simple code describing the interactions between amino acid side chains and functional groups in the major groove of DNA? How widespread – amongst prokaryotic and eukaryotic DNA-binding proteins – is the helix '2–3' motif? What is the role of helix '2'? What is the role of the arms of lambda repressor? Can we modify existing repressors so that they recognize any specified base sequence? What can we learn about RNA polymerase structure and function by studying polymerase-repressor (positive activator) interaction? How general is the mechanism of positive control we have adduced for the lambda and P22 repressor cases? For example, will the *E. coli* protein CAP work this way? What about cases in which regulatory proteins evidently stimulate transcription from sites quite far upstream of the start site, e.g. positive regulation by the *GAL4* product in yeast or by steroid receptor proteins in higher eukaryotes?

Acknowledgements

A list of graduate students and post-doctoral fellows, 1976 to the present, whose work I have alluded to is as follows: Egon Amann, John Anderson, Keith Backman, Roger Brent, James Eliason, Lenny Guarente, Ann Hochschild, Zafri Humayun, Nina Irwin, Alexander Johnson, Gail Lauer, Mitch Lewis, Tom Maniatis, Russell Maurer, Barbara Meyer, Carl Pabo, Ricardo Pastrana, Anthony Poteete, Tom Roberts, Michael Ross, Robert Sauer, Robin Wharton and Roger Yocum.

References

Results described in this article are taken from, or cited in, the following references:

Anderson, W. F., Ohlendorf, D. H., Takeda, Y. and Matthews, B. W. (1981) *Nature* 290, 754–758

Anderson, J., Ptashne, M. and Harrison, S. (1984) *Proc. Natl Acad. Sci. USA* 81 (in press)

Hochschild, A., Irwin, N. and Ptashne, M. (1983) *Cell* 32, 319–325

Johnson, A. D., Poteete, A. R., Lauer, G., Sauer, R. T., Ackers, G. K. and Ptashne, M. (1981) *Nature* 294, 219–223

Lewis, M., Jeffrey, A., Wang, J., Ladner, R., Ptashne, M. and Pabo, C. O. (1983) *Cold Spring Harbor Symposia on Quantitative Biology* XLVII, 435–440

Pabo, C. O. and Lewis, M. (1982) *Nature* 298, 443–447

Ptashne, M., Johnson, A. and Pabo, C. (1982) *Scientific American* 247, 128–140

Basis of biological specificity

Alan R. Fersht

The principles of biological specificity were known qualitatively in 1976. The remarkable developments in molecular genetics since then are enabling the quantitative analysis of specific interactions and construction of proteins of novel specificities.

There are two components of biological specificity: thermodynamic – equilibrium binding through simple intermolecular interactions; and kinetic – control of product formation through rates. The major determinant is thermodynamic and arises through the fit of complementary structures. Complementarity is the basis of biological specificity in the interaction of macromolecules with one another and with small ligands in processes varying from the assembly of large oligomeric structures, antibody–antigen interactions, to the binding of allosteric effectors to their receptors. In some cases, one molecule is complementary to an altered structure of another so that binding induces a conformational change that triggers another event. The other great area dominated by the fit of complementary structures is the replication of biological molecules: the flow of genetic information is mediated by complementary base pairing.

Complementary interactions are limited by thermodynamics; there are intrinsic upper values for binding energies which limit the specificity that can be achieved. The intrinsic binding energies in base pairing and in the binding of amino acids to proteins are inadequate, in many cases, to give the required fidelity of replication of DNA and translation of its transcribed message. This has necessitated the evolution of the second determinant of specificity, *editing* or *proofreading* mechanisms in which product formation switches from thermodynamic to kinetic control: key enzymes in replication and translation have hydrolytic sites that weed out errors as they are introduced. For example, certain aminoacyl-tRNA synthetases have an esterolytic activity that is specific for the deacylation of their misacylated tRNAs, and prokaryotic DNA polymerases have a 3′–5′ exonuclease activity that is most active towards misincorporated nucleotides. (For a review and full list of earlier references see Ref. 1.)

All this was known in principle in 1976. Editing mechanisms were discovered experimentally for the selection of amino acids during protein synthesis in 1964[2] and for nucleotides in DNA replication in 1972[3,4]. The concept of specificity through complementarity has had a noble history that can be traced back through the 'greats' of biochemistry and chemistry: Emil Fischer's 'lock-and-key' hypothesis for enzyme specificity; Paul Ehrlich's general description of specificity in terms of molecules fitting together like pieces in a mosaic; Linus Pauling's emphasis on shape and complementarity in specificity and enzyme catalysis[5]. A long series of experiments on model compounds to measure the energetics of macromolecular interactions led to the following consensus[6–8].

Alan R. Fersht is at the Department of Chemistry, Imperial College of Science and Technology, London SW7 2AY, UK.

Ideas on energetics of complementary interactions in 1976

The driving force for the association of macromolecules is the hydrophobic bond. Each square Å of hydrophobic surface area of protein excluded from aqueous solvent contributes 80–100 $kJmol^{-1}$ to binding energy. Hydrogen bonds and salt bridges contribute little binding energy but provide specificity because the presence of unpaired charges and hydrogen bonding groups is unfavourable.

Open questions in 1976

Very little of the evidence concerning energetics had come from direct measurements on proteins. Should the association of macromolecules be considered just as a simple, global, hydrophobic process without important specific contributions? Are there, for example, specific interactions between amino acid side chains and cavities in proteins that are higher than predicted from the simple ideas on hydrophobic bonding? What are the intrinsic limits of the specificity of complementary interactions? What is the structural basis of specific interactions? There were also fundamental unanswered questions concerning the extreme fidelity required in DNA replication and protein synthesis. What is the accuracy of base pairing in DNA replication? What is the fidelity of DNA polymerases during replication? What are the chemical mechanisms of editing

processes? What limits the contribution of editing to specificity? Why are there back-up mechanisms to editing such as post-replicative mismatch repair of DNA? What processes require editing mechanisms? What principles govern selection in editing? The rest of this article is devoted to the developments from 1976 onwards.

The answers to some of the questions have come from simple extensions of the classical studies, invoking just the techniques available in 1976. For example, the reaction pathways for the rejection of incorrect amino acids by the aminoacyl-tRNA synthetases were established rigorously by pre-steady-state and steady-state kinetics. The principles of selection, embodied in the 'double-sieve' model, were established, as described in *TIBS*[9], although there are now known to be more pathways[10]. Values of intrinsic binding energies of small groups with proteins were measured empirically and tabulated[1]. It was found that there can be very high specific interaction energies between small groups and proteins when there is evolutionary pressure for specificity, and these energies can be several times higher than predicted from chemical model experiments. For example, whereas the Gibbs' energy of transfer of a methylene group from an organic solvent such as n-octanol to water is 2.8 $kJmol^{-1}$, the energy of transfer from the hydrophobic cavity of an aminoacyl-tRNA synthetase is 14 $kJmol^{-1}$. Repulsion energies can also be very high: cramming the additional methylene group of the side chain of isoleucine into the active site of the valyl-tRNA synthetase costs more than 30 $kJmol^{-1}$ of energy. Thus, small hydrophobic groups can provide considerable energies for binding and specificity.

Impact of genetic engineering DNA sequencing

Investigation of the structural basis of the energetics of complementary interactions became possible as a result of the emergent recombinant DNA technology. The determination of the sequence of bacteriophage ϕX174 in Sanger's laboratory in 1977 enabled the initiation of two new lines of research: it provided a defined biologically active DNA template on which to measure for the first time the fidelity of replication of DNA (as described earlier in *TIBS*[9]); it allowed the development of oligodeoxynucleotide-directed mutagenesis of protein structure and hence the new discipline of protein engineering[11].

Investigation of interactions by protein engineering

Greg Winter and I, and our colleagues in Cambridge and London, have initiated a systematic dissection of the structure and activity of the tyrosyl-tRNA synthetase from *Bacillus stearothermophilus* using protein engineering[12–14]. Amino acid residues that have been identified by protein crystallographic studies by Professor D. M. Blow's group[15] to make hydrogen bonds with the substrates have been mutated by oligodeoxynucleotide-directed mutagenesis. Kinetic measurements on the mutant enzymes are beginning to reveal characteristics of hydrogen bonding. Consider the specificity of the enzyme for tyrosine compared with phenylalanine. There has clearly been selective pressure to maximize the specificity and this can be mediated only by binding the distinctive hydroxyl group of tyrosine as tightly as possible. Earlier kinetic measurements showed that tyrosine is activated at least 10^5 times faster than phenylalanine, corresponding to some 30 $kJmol^{-1}$ of binding energy[16]. We now know (D. M. Blow, unpublished results) that the hydroxyl forms two buried hydrogen bonds with the enzyme: one with the carboxylate ion of Asp-176; the other with the hydroxyl of Tyr-34. Removal of one hydrogen bond by mutation of Tyr-34 to Phe-34 lowers the binding energy of tyrosine by only 2.3 $kJmol^{-1}$. This result, surprisingly low to those expecting a figure of some 16–20 $kJmol^{-1}$ for a hydrogen bond, illustrates an important principle of equilibrium hydrogen bonding: the formation of a hydrogen bond between an enzyme and a substrate is an exchange phenomenon in which the enzyme and substrate exchange their hydrogen bonds with water for bonds with each other (eqn 1).

$$E\cdots HOH + S\cdots OH_2 = E\cdots S + HOH\cdots OH_2 \quad (1)$$

A *hydrogen bond inventory* must be performed, counting the number of bonds on each side of equations such as eqn 1. In eqn 1, there are the same number of hydrogen bonds formed each side and so they tend to cancel out. (The driving force for hydrogen bond formation between an enzyme and a substrate, as in eqn 1, is the entropy gain as bound water molecules are released.) The hydroxyl binding site of the tyrosyl-tRNA synthetase must, of course, bind a water molecule which is displaced on binding tyrosine. Thus, the enzyme loses two hydrogen bonds with water to gain two with the tyrosine hydroxyl. In the mutant enzyme (Tyr-34 to Phe-34), only one hydrogen bond with water is lost but only one is gained with the tyrosine hydroxyl. The hydrogen bond inventory thus shows that the absence of a hydrogen bonding group may have minimal effect. The mutation experiment provides direct evidence for the earlier hypothesis that hydrogen bonds may contribute little binding energy.

If one of the hydrogen bonds in eqn 1 is subject to geometric constraints that prevent it forming properly, e.g. that between E and S, then the equilibrium will be displaced. We have used this principle to improve the affinity of the enzyme for ATP[14].

Protein engineering experiments should eventually afford experimental measurements of all the different kinds of interactions. These data, in combination with three-dimensional structural knowledge from protein crystallography and precise theoretical calculations, will provide the basis for analysing and designing specific reagents.

'Kinetic genetics'

The fidelity of replication of DNA is now measured by replicating point mutants of ϕX174 *in vitro* and measuring the number of revertants formed in the synthetic DNA after expressing it as phage. These experiments show that the inherent error rate in base pairing is about one per 10^4–10^5 nucleotides incorporated (16–20). The contribution of the 3′–5′ exonuclease activity of the major replicative polymerase of *E. coli* (Pol III) to editing is relatively low and post-replicative repair mechanisms have been found to boost the accuracy further. Just why these are required may be seen from examining the 'cost' of editing.

Cost and accuracy of editing

A small amount of the correct substrate is also excised during the editing process, the 'cost'. This was discovered in 1972 for DNA replication, where it was noted that mutant DNA polymerases of T4 that were more accurate than wild type also had a higher cost of editing[4]. The relationship between cost and accuracy has now been analysed in some depth (20–23). Most editing mechanisms can be reduced to the following scheme (eqn 2) in which the enzyme reacts with a monomer to give a high-

$$E + S \overset{f}{=\!=} E.S \rightarrow E.I \begin{cases} \text{elongation} \\ \text{excision} \end{cases} f' \quad (2)$$

energy intermediate E.I that then partitions between successful elongation or

excision. A simple analysis that I find useful gives the cost-selectivity equation (eqn 3). This relates the overall accuracy

$$S = f(1 + (f'-1)C) \qquad (3)$$

(the selectivity S) to the discrimination factor f (the relative rate of formation of the correct and incorrects intermediate E.I from an equimolar mixture of the monomers ($f > 1$)), the discrimination factor f' (the relative rate of editing of incorrect and correct E.I ($f' > 1$)) and the cost C. (This is explained in detail in Refs 1 and 17.) The magnitudes of the discrimination factors are determined by the energetics of the complementary interactions. In the absence of editing, the specificity of the process would be f; editing provides a further factor of $(f' - 1)C$. The increase in specificity from editing is directly proportional to the cost. Briefly, editing processes may be divided into two categories depending on the features used for selection in the initial incorporation step (f) and the editing step (f'). Editing processes such as that by the polymerase in DNA replication use the same selection procedure for both incorporation and editing: the monomer (dNTP) to be incorporated is initially checked by testing its base pairing with template; checking whether or not it should be edited is done by the same procedure. Thus $f' \leq f$. The maximum increase in fidelity from editing is limited to a further factor of f, and this only occurs when $C = 1$, i.e. when all the correctly incorporated substrate is excised! This limits the contribution of polymerase editing to fidelity in DNA replication; the overall accuracy of the polymerase is less than one part in 10^8, a factor of 100 or so below the observed mutation rates in *E. coli*. Thus, back-up mechanisms are required[1,17]. Theoretical considerations show that multistep editing mechanisms can give higher accuracy for lower cost.

Future prospects

Many of the unanswered questions of 1976 are being answered by the rapid developments in molecular genetics. The prospects from protein engineering appear open-ended. The ability to measure the energetics of intermolecular interactions directly from modified proteins renders much of the work on small model compounds unnecessary. Further, it allows theoretical work to be put into practice by construction of proteins of novel specificities.

References

1 Fersht, A. R. (1981) *Proc. R. Soc. Lond. B212* 351–379
2 Norris, A. T. and Berg, P. (1964) *Proc. Natl Acad. Sci. USA* 52, 330–337
3 Brutlag, D. and Kornberg, A. (1972) *J. Biol. Chem.* 247, 241–248
4 Muzyczka, N., Poland, R. L. and Bessman, M. J. (1972) *J. Biol. Chem.* 247, 7116–7122
5 Pauling, L. (1948) *Am. Sci.* 36, 51–58
6 Kauzmann, W. (1959) *Adv. Protein Chem.* 14, 1–63
7 Jencks, W. P. (1969) *Catalysis is Chemistry and Enzymology*, McGraw Hill
8 Tanford, C. (1980) *The Hydrophobic Effect*, 2nd edn, Wiley
9 Fersht, A. R. (1980) *Trends Biochem. Sci.* 5, 262–265
10 Jakubowski, H. and Fersht, A. R. (1981) *Nucl. Acids Res.* 9, 3105–3117
11 Smith, M. (1982) *Trends Biochem. Sci.* 7, 440–442
12 Winter, G., Fersht, A. R., Wilkinson, A. J., Zoller, M. and Smith, M. (1982) *Nature* 299, 756–758
13 Wilkinson, A. J., Fersht, A. R., Blow, D. M. and Winter, G. (1983) *Biochemistry* 22, 3581–3586
14 Wilkinson, A. J., Fersht, A. R., Blow, D. M., Carter, P. and Winter, G. (1984) *Nature* 307, 187–188
15 Bhat, T. N., Blow, D. M., Brick, P. and Nyborg, J. (1982) *J. Mol. Biol.* 122, 407–417
16 Fersht, A. R., Shindler, J. S. and Tsui, W.-C. (1980) *Biochemistry* 19, 5520–5524
17 Fersht, A. R., Knill-Jones, J. W. and Tsui, W.-C. (1982) *J. Mol. Biol.* 156, 37–51
18 Kunkel, T. A., Schaaper, R. M., Beckman, R. A. and Loeb, L. A. (1981) *J. Biol. Chem.* 256, 9883–9889
19 Grosse, F., Krauss, G., Knill-Jones, J. W. and Fersht, A. R. (1983) *EMBO Journal* 1515–1519
20 Fersht, A. R. and Knill-Jones, J. W. (1983) *J. Mol. Biol.* 165, 633–654
21 Galas, D. J. and Branscomb, E. W. (1978) *J. Mol. Biol.* 124, 653–687
22 Savageau, M. A. and Freter, R. (1979) *Biochemistry* 18, 3486–3493
23 Ehrenberg, M. and Blomberg, C. (1980) *Biophys. J.* 31, 333–358

Protein structure

C. C. F. Blake and L. N. Johnson

The last eight years have been marked by a deepening understanding of the structure of soluble, membrane and cell-surface proteins. There are increasing indications that many proteins are constructed of functionally independent modules whose movement relative to one another gives the molecule additional functional scope. These new principles should be of value in the design of man-made proteins, whose production now seems possible.

The period from 1976 to 1984 corresponds rather well to what one might call the second stage in the understanding of the nature and functional scope of protein structure. This stage has seen the formulation of a successful classification of protein tertiary structure, and an increased understanding of domain structure, its functional and dynamical consequences, and its possible genetic basis in the exon/intron mosaic of genes. This period has also provided the first views of the structures of new classes of proteins, including DNA-binding proteins, integral membrane proteins, and surface glycoproteins. Technical advances have included the entry into widespread use of protein refinement, which gives more accurate information on structural details, dynamical information, and interactions with the environment; while the development of synchrotron radiation sources has provided powerful tools for the study of very large protein molecules and molecular aggregates. The beginning of a third stage is already evident in which the emphases are likely to be on the relationship between a protein's function and its structure, dynamics, and environment, and the use of this information to predict, manipulate and, eventually, to engineer protein structures.

C. C. F. Blake and L. N. Johnson are at the Laboratory of Molecular Biophysics, University of Oxford, UK.

Structure and domains

In 1976 Levitt and Chothia[1] showed that the proteins whose structures were then known could be assigned to one of five structural classes. These classes were defined in terms of the presence and arrangement of α-helices and β-sheets, the only major, extended structural elements present in globular protein molecules. These classes are: Class I, the 'all α-proteins', in which only α-helices are present and are packed together in the globular form; Class II, the 'all β-proteins', in which only β-structures are present usually as two antiparallel β-sheets packed face-to-face or in the form of a barrel; Class III, the