A fine-tuned system of molecular recognition is a fundamental feature of all biological processes. Although the basic principles of how molecular surfaces fit together in a complementary fashion have been established to a first approximation, the details of the interplay between electrostatic, steric, entropic and solvation effects are not well understood. To provide insight into the stereochemical factors that govern recognition and packing at protein interfaces, a host of biochemical and biophysical techniques have been applied. Perhaps the most widely used of these techniques are mutagenesis and crystallography, and when used together they have proved to be extremely powerful in defining the functional elements that are important in protein interactions.

Using these protein-engineering techniques we have focused our studies on the role of molecular recognition in two important, but distinctly different, biological processes: (1) the role of molecular recognition in receptor activation, and (2) the stereochemical factors that define specificity and provide binding energy to inhibitors that are associated with their target proteinases. Our findings, discussed below, suggest that the conventional wisdom about specificity and binding needs to be rethought.

Role of molecular recognition in the activation of the human-growth-hormone receptor

Human growth hormone (hGH) is a 22 kDa protein that stimulates the growth of bone, muscle and cartilage, by binding to a specific cell-surface receptor [1]. Two specific receptors that bind to hGH have been identified, the activation of which results in distinct pharmacologic effects: the growth hormone (GH) receptor and the prolactin (PRL) receptor. These receptors are classified, based on sequence similarities between their extracellular domains (ECDs), as belonging to the haematopoietic-receptor superfamily. This extensive receptor family includes the endocrine receptors, such as the GH and PRL receptors, and a number of other cytokine receptors [2, 3]. These receptors have a three-domain organization: the extracellular portion that binds the activating ligand; a transmembrane segment, which is usually hydrophobic in nature and 20–30 residues long; and a cytoplasmic domain that is involved in producing the response, the so-called 'second message', within the cell.

How does hormone binding trigger the activation of these receptors? Several mechanisms for the activation process had been proposed [4]. The most plausible of the mechanisms is based on some type of receptor aggregation, which assumed the formation of complexes of hormone–receptor pairs; that is, aggregates of some multiple of hormone–receptor dimers. However, the biochemical data assigning stoichiometry were equivocal and there was no structural information about the nature of the aggregation.

The growth hormone–receptor structural studies were undertaken using hGH and the ECD of its receptor. Crystals that were grown of the complex diffracted to 2.7 Å [5]. During the initial stages of the structure determination, the 5 Å electron-density map clearly showed that the complex contains three molecular components, not the two or four that are expected if the complex was a multiple of one hormone:one receptor ECD. It was determined that the stoichiometry found in the crystal, and subsequently verified in solution [6], is two ECD molecules to 1 hGH. This finding presented a conundrum; how do two receptor ECDs bind to one hGH molecule, considering that hGH has no semblance of symmetry (an asymmetric four-helix bundle protein; Figure 1) that could support the

Abbreviations used: GH, growth hormone; PRL, prolactin; PL, placental lactogen; ECD extracellular domain; BPTI, basic pancreatic-trypsin inhibitor; APPI, Alzheimer's amyloid β-protein precursor inhibitor (Kunitz domain); h, human.
binding of each of the ECDs in the same way? Are there conformational changes in the ECD or in the hormone, or does the hormone bind at different sites on the two receptor ECDs?

The answer to this puzzle was revealed by the high-resolution structure (Figure 2) [7]. Surprisingly, the hormone binds to essentially the same residues on each of the ECDs. Even though the hormone-binding epitope of the two ECDs includes essentially the same residues, the surface areas that are buried at the two interfaces are substantially different; for ECD1 (binding to hGH at site 1), the buried surface area is about 1250 Å², whereas for ECD2, the area is 850 Å². There are two reasons for the difference in interaction surface area. Firstly, there are slight conformational differences in the loop regions of the ECD, depending on whether it binds to hGH at site 1 or at site 2. The second reason is the nature of the topography of the hGH molecule itself. At binding site 1, hGH has a large invagination into which the contacting sidechains of the ECD insert rather like a 'lock and key'. At binding site 2 the surfaces both of the hGH and of ECD2 appear to be not as complementary.
The binding epitope of hGH at site 1 had been evaluated before the completion of the structural work by an extensive mutagenesis effort [8, 9]. This epitope mapped principally to residues on helix 4 and a segment of the loop that connects helix 1 and helix 2. The hGH epitope defined by the contact surface that was determined from the X-ray structure contained the mutagenic epitope, but surprisingly extended to a significantly larger area. Although it was not unexpected that each individual contact residue would not have a measurable effect on binding, the proportion of contact residues that had no influence on binding energy was surprising. The explanation may lie in the cross-reactivity of hGH and hPRL with their receptors. Some parts of the contact surface that are not directly associated with providing binding energy for the complex may govern the specificity. We are currently working on follow-up structural and mutagenesis experiments to characterize these findings in more detail.

The binding epitope of hGH at site 1 had been evaluated before the completion of the structural work by an extensive mutagenesis effort [8, 9]. This epitope mapped principally to residues on helix 4 and a segment of the loop that connects helix 1 and helix 2. The hGH epitope defined by the contact surface that was determined from the X-ray structure contained the mutagenic epitope, but surprisingly extended to a significantly larger area. Although it was not unexpected that each individual contact residue would not have a measurable effect on binding, the proportion of contact residues that had no influence on binding energy was surprising. The explanation may lie in the cross-reactivity of hGH and hPRL with their receptors. Some parts of the contact surface that are not directly associated with providing binding energy for the complex may govern the specificity. We are currently working on follow-up structural and mutagenesis experiments to characterize these findings in more detail.

**Optimization of inhibitor sequences for specificity and for binding**

Kunitz inhibitors are a class of small inhibitors (about 60 amino acids) that target the trypsin-like family of serine proteinases [10]. The complexes of these inhibitors and their target enzymes provide ideal systems to study because the contacting surfaces between inhibitor and enzyme are relatively small and well-defined. These complexes have been studied extensively by biochemical and biophysical approaches, including mutagenesis and crystallography. In general these inhibitors do not show a high degree of specificity. For instance, two Kunitz inhibitors, basic pancreatic-trypsin inhibitor (BPTI) and Alzheimer's amyloid β-protein precursor inhibitor (APPI), can inhibit both trypsin and chymotrypsin, enzymes that have very different specificities towards the substrate sequences that they hydrolyse.

The contact surface of the inhibitor abutted against the enzyme consists of 5–7 residues [11]. At position 15 of the inhibitor (BPTI numbering scheme) is the P1 residue; that is, the residue whose sidechain inserts into the binding pocket of the enzyme and is thus the primary specificity-determining residue. It has been thought that other nearby residues play key roles in binding and specificity, although no systematic evaluation has been carried out.

To address the question of specificity a method called 'phage-display mutagenesis' was used. This method has been described in detail in some recent reviews [12] and so it will not be discussed here. In outline, a protein of choice (in this case the Kunitz inhibitor, APPI) can be expressed on the surface of bacteriophage M13. M13 is the phage that is commonly used in *Escherichia coli* mutagenesis, and the gene for APPI has been inserted into the phage genome, where it can easily be mutated by standard techniques. Specific residues can be targeted for random mutagenesis; for APPI, five contact residues, positions 11, 13, 15, 17, 19, were chosen. Libraries of around 3–4 million random mutants can be made with combinations of mutations at these five sites. This number of mutants does not cover the total number of possibilities given 20 different combinations for each of the five sites, but it certainly provides for a very extensive subset.

A large proportion of these mutants will not be tight binders or will not show specificity for a given enzyme. Selecting out those mutants with the desired properties is done by 'panning'. This is best explained by an example. One goal of the mutagenesis was to find a set of mutant inhibitors that bind specifically to trypsin, and not to chymotrypsin. The inhibitor display phage is mixed in a tube with a stoichiometric amount of chymotrypsin. Those inhibitor phage that bind to chymotrypsin do so, and remain as a bound complex. Next, beads with trypsin immobilized on them are added to the tube. Those inhibitor phage that have not bound to chymotrypsin, but can bind trypsin, do so. The beads are then washed, eliminating the inhibitor-chymotrypsin phage and all the unbound inhibitor phage; only the phage complexed to the trypsin beads remains. These phage can be eluted under appropriate conditions, and when collected can be propagated to recycle the process, each time enriching the pool. The end product is usually a small set of mutants, in this case optimized at five positions. After enrichment, a statistically significant number of the clones can be sequenced to establish the combination of residues that appears at each of the sites.

Figure 3 shows the results for trypsin, chymotrypsin and kallikrein optimized for individual binding (selection was performed to pick out the best individual binders, but there was no optimization either against binding other enzymes or to include the binding of other enzymes), the inhibitor sequences that bind to all three enzymes and, in the case of kallikrein, specific binding (last panel). For the individual-binding selections, the sequence positions that appear to be most important differ between the enzymes. For trypsin, the P1 position, residue 15, is clearly the principal factor, with an
Figure 3
Phage-display selections varying positions 11, 13, 15, 17 and 19 of APPI

The first three panels show the sequences obtained after ten rounds of enrichment for trypsin, chymotrypsin and kallikrein respectively. The fourth panel shows the sequences obtained when the selection was for inhibitors that were broadly specific; that is, those that inhibit all three enzymes. The last panel shows sequences that make the inhibitor kallikrein-specific.

<table>
<thead>
<tr>
<th>Position</th>
<th>Trypsin</th>
<th>Chymotrypsin</th>
<th>Kallikrein</th>
<th>Broadly specific</th>
<th>Kallikrein-specific</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>I</td>
<td>I</td>
<td>I</td>
<td>Arg</td>
<td>Leu</td>
</tr>
<tr>
<td>13</td>
<td>Asp</td>
<td>Asn</td>
<td>I</td>
<td>I, Asn, His</td>
<td>I, Arg</td>
</tr>
<tr>
<td>15</td>
<td>Asn</td>
<td>His</td>
<td>I</td>
<td>Arg, His, I</td>
<td>Arg, Lys</td>
</tr>
<tr>
<td>17</td>
<td>Phe</td>
<td>Phe</td>
<td>I</td>
<td>Arg</td>
<td>Arg, I</td>
</tr>
<tr>
<td>19</td>
<td>Lys</td>
<td>Lys</td>
<td>I</td>
<td>I, Lys</td>
<td>I, Lys</td>
</tr>
</tbody>
</table>

absolute specificity for a basic sidechain. The number and spread of the characteristics of the sidechains at the four other positions indicates that none of these is particularly important in defining the binding. In the case of chymotrypsin, however, four positions show about equal specificity: positions 11, Arg; 15, Asn and His; 17, Phe; and 19, Lys. The kallikrein results are even further from the P1 residue defining the binding of the inhibitor. In this case positions 13, Arg; and 19, Leu appear to provide the principal binding specificity, with positions 15 and 17 being only of secondary importance.

The most surprising aspect of these results is the finding that selection at position 15 for chymotrypsin comes out with asparagine and histidine as the best inhibitor residues. This contradicts the conventional wisdom, which states that chymotrypsin has a specificity for large hydrophobic sidechains such as phenylalanine and tyrosine. Why is a hydrophilic sidechain preferred to fit into a hydrophobic binding pocket? This paradoxical question was answered by the structural analysis of BPTI-chymotrypsin complex (T. Hynes and A. A. Kossiakoff, unpublished work). BPTI has a lysine at position 15, and so is not only hydrophilic, but is charged as well.

The structure revealed that the supposed hydrophilic–hydrophobic incompatibility does not present a quandary at all. The lysine sidechain does not pack fully into the hydrophobic binding pocket, but rather makes a U-turn allowing the charged ε-NH₂ nitrogen atom to reach the surface of the interface where it makes two H-bonds. Although structure analyses of the asparagine and histidine variant BPTI–chymotrypsin complexes are only just getting underway, it is reasonable to suggest that these hydrophilic sidechains will pack similarly to lysine.

The kallikrein-specificity selection (binding to kallikrein, but not to trypsin or to chymotrypsin) results are also presented in Figure 3. It is clear that not every position has converged to a specific residue or subset of residues that differentiates kallikrein from the other two enzymes. However, positions 15 and 17 probably play principal roles in the differentiation. The selected kallikrein variants all have leucine at position 15, while branched hydrophobic sidechains of any kind are not found for trypsin or chymotrypsin. At position 17, an acidic sidechain, aspartic or glutamic acid, emerges. It is interesting that, in the initial binding selection, aspartic or glutamic acid at position 17 were not isolated for any of the enzymes. The results of the specificity selection suggest that, using the phage-display approach, specificity, to an extent, is probably derived through a process of 'least adverse effect'. That is, the mutants that are selected are not optimized for increased binding to the target enzyme, but rather for relative decreased binding to
the other enzymes. Further elucidation of this process requires the production of a set of single-site mutants and work is underway in this area.

Conclusions
To appreciate the way that the biological systems of living organisms provide for the essential task of regulating the interaction among molecules, it is important to recognize that the process of molecular recognition at 'high' resolution does not involve fitting surfaces together, but rather fitting atoms together. There is actually little difference at the atomic level between the way in which groups of atoms adjust to minimize energy during the protein-folding process, and the way in which the surfaces of two folded polypeptide chains adjust and desolvate to form a strong and specific protein–protein interface. Stepping back to a 'medium' resolution: that is, thinking about for example fitting knobs into holes, and making H-bonds, the studies presented here demonstrate that there are two quite distinct strategies that are used in nature in molecular recognition. For the hGH–receptor complex, the binding of the two ECDs to the topographically different surfaces of hGH involves the same residues of the two ECDs, but with variability in their conformation. The inhibitor–protease recognition, however, works using a constant conformation of the inhibitor-binding loop, but with sequence diversity.

In the hGH–GH receptor system the contact surface is large, but is not used efficiently; in the inhibitor–proteasease case, the number of contact residues is considerably smaller, but apparently the complementarity is better. Even though the contact surface is larger for the hormone–receptor interface, biological regulation through the fine-tuning of specificity, altering a very few residues, is highly sensitive. For instance, hGH binds to the hGH and hPRL receptors, whereas PRL and placental lactogen (PL) will only bind to the PRL receptor. hPL has 85% sequence identity with hGH (with only three contact residues being different), but will not bind to the hGH receptor. hPRL is considerably different both from hGH and from hPL and, not surprisingly, will not bind to the hGH receptor, although all three hormones bind to the PRL receptor. We have recently determined the structure of the hGH–hPRL receptor complex, and have found that, even though the binding footprint is nearly identical to that of the hGH–hGH receptor, some details are significantly different.

In contrast with the hormone–receptor story of titrating binding with just a few changes, inhibitor–protease interactions appear in general to be less specific. Although trypsin requires the inhibitor to have a basic P1 sidechain, there can be considerable variation in the chemical character of the sidechain at the four other positions. Even the P1 sidechain is not highly important for the chymotrypsin- and kallikrein-binding pockets. Given the above description, it might be assumed that these variant inhibitors are relatively weak binders, however, this is not the case since most are in the nanomolar range. Taken together, the picture presented is that there are no specific rules for molecular recognition, only very general ones. This may be because nature uses highly diverse solutions to similar problems, or because our understanding is so sparse that we miss the full picture.

We thank Laura Kennedy for helping with the structural work on the inhibitor complexes, and Brian Cunningham, Germaine Fuh and Jim Wells for helpful discussions.

References