

which are bacterial adhesion proteins, also have the same overall fold topology but lack sequence similarity<sup>5</sup>.

The retrotransposition mechanism ensures that variability is restricted to one surface of the domain. The portions of the sequence that form the core of the domain are thus isolated from the diversity-generating process, ensuring that the overall fold remains unaffected. Most remarkably, the presence of residues within the variable region that are not subject to mutation because they are specified by codons that lack adenines is sufficient to ensure that the fold of the variable region also remains constant. The backbone atoms of all forms examined are essentially superimposable, and only the side chains vary. Thus, the structural changes brought about by variation in the variable region are very different in character from the changes in loop length and conformation of the variable regions of immunoglobulins<sup>6</sup>. The effect of this approach to the generation of diversity is to vary the local surface properties of the binding site without changing its overall shape, a

much more constrained form of variability than is seen in immunoglobulins.

Future studies will no doubt shed light on how these relatively modest changes in the binding site allow the receptor-binding protein to accommodate changes in bacterial surface receptors. Some of the changes in the receptors are relatively subtle. For example, pertactin, which is the receptor for the phage on the bacteria in the pathogenic phase, mutates to evade the host immune response. However, as the bacteria switch to the environmental phase, they cease to express colonization factors, including pertactin, so the phage must evolve to bind completely different receptors such as proteins that confer motility. Thus, randomization of 12 residues within the conserved fold is sufficient to generate receptor-binding sites that differ markedly in their properties.

Screening of other bacteriophage and bacterial systems for the sequence of the *mtd* gene suggests that this mechanism for introducing variability is shared by other recognition systems. In all of these cases, it seems that

similar transposition mechanisms introduce variability into proteins with the C-type lectin-like fold. It is very tempting to speculate on the possibility of re-engineering the genetic system to allow controlled randomization of sequences in other structural contexts. The constraints seem to be modest, in that the targeting sequence that directs insertion of the mutagenized region probably does not need to fall within the coding sequence of the gene, although in the case of *Bordetella* phage it does. The ability to designate residues for randomization, while maintaining invariant residues, by selective incorporation of adenine nucleotides into the coding sequence offers a potentially very powerful way to generate diversity within a structurally controlled context.

1. McMahon, S.A. *et al. Nat. Struct. Mol. Biol.* **12**, 886–892 (2005).
2. Liu, M. *et al. Science* **295**, 2091–2094 (2002).
3. Weis, W.I., Taylor, M.E. & Drickamer, K. *Immunol. Rev.* **163**, 19–34 (1998).
4. Blundell, C.D. *et al. J. Biol. Chem.* **278**, 49261–49270 (2003).
5. Batchelor, M. *et al. EMBO J.* **19**, 2452–2464 (2000).
6. Chothia, C. & Lesk, A.M. *J. Mol. Biol.* **196**, 901–917 (1987).

## Crossing the species barrier

Viruses use surface proteins on host cells as entry receptors. These molecules are specifically recognized by proteins on the viral particles, and sequence variations among receptors from different species could have the potential to limit the virus's host range. Mutations in the viral proteins can overcome this barrier, and one recent example of this occurring is the SARS (severe acute respiratory syndrome) coronavirus, which seems to have adapted from an animal virus infecting palm civets to include humans as its host. The human strains caused the 2002–2003 epidemic that killed ~10% of those that were infected.

Angiotensin convertin enzyme 2 (ACE2) is the receptor for SARS coronavirus and interacts with the viral spike protein. Previous studies have identified a ~200-residue receptor-binding domain (RBD) in the spike protein. Furthermore, these studies have shown that the RBDs derived from the palm civet and the human epidemic strains are very similar, differing at only four positions; yet, the affinity of the respective spike protein for human ACE2 varies by more than 1,000-fold. The crystal structure of the spike protein RBD, derived from a human SARS coronavirus infection, in complex with human ACE2 (Li *et al. Science* **309**, 1864–1868, 2005) now provides a structural explanation for this difference in affinity.

The spike protein RBD has two subdomains, but only one of these two (~70 residues) contacts the ACE2 receptor. This subdomain forms a shallow cleft (red) that curves snugly around the N-terminal tip of ACE2 (yellow green). A total of 32 residues contribute to the interaction interface (14 residues from the RBD and 18 residues

from the ACE2). Most of the interactions between the RBD and ACE2 are mediated by side chains and are hydrophilic in nature.

Of the four residues that are different in the RBDs from palm civet and human epidemic SARS coronaviruses, only two—positions 479 and 487—are at the interaction interface. A previous study showed that having a lysine at position 479 (found in many civet-derived viral sequences) reduces affinity for human ACE2, whereas having an asparagine at this position (found in human epidemic-derived viral sequences) does not seem to have a significant effect on binding to the civet receptor. The structure shows that the ACE2 residues surrounding position 479 are not completely conserved.

In contrast, a previous study showed that changing position 487 from threonine (human epidemic) to serine (civet) reduces affinity of RBD for human ACE2 by more than 20-fold. The structure shows that the ACE2 residues important for making contacts with position 487 of RBD are essentially identical between humans and civets. These residues support the formation of an inflexible hydrophobic pocket filled by the methyl group of Thr487 from the RBD of the human epidemic strains. A serine residue found in the civet-derived viral sequences cannot fill this pocket and may affect the affinity between the two proteins. Other factors, such as residues not at the interface or glycosylation, could also attenuate the affinity between RBD and ACE2. Nevertheless, together with additional sequence analysis of sporadic SARS cases, the structure suggests that the  $\gamma$ -methyl group on the side chain of position 487 may be a key determinant for the severity of the disease.

Hwa-ping Feng

