Proc. Int. Symp. Biomol. Struct. Interactions, Suppl. J. Biosci.,
Vol. 8, Nos 1 & 2, August 1985, pp. 481–489. © Printed in India.

Crystal structure of subtilisin complexed with its trapped substrate *Streptomyces* subtilisin inhibitor—Protein–protein interaction and evolution of serine proteinases and their proteinaceous inhibitors

Y. MITSUI* , H. AKAGAWA†, H. ONISHI‡, S. ITOH,
K. T. NAKAMURA and Y. IITAKA,

Faculty of Pharmaceutical Sciences, University of Tokyo, Hongo, Tokyo 113, Japan

† Present address: The Ministry of Public Welfare, Chiyoda-ku, Tokyo 100, Japan

‡ Present address: Faculty of Pharmaceutical Sciences, Hoshi University, Ebara 2-4-41, Shinagawa-ku, Tokyo 142, Japan

Abstract. The complex of a bacterial alkaline serine proteinase, subtilisin BPN′, with its proteinaceous inhibitor *Streptomyces* subtilisin inhibitor is unique in several respects, compared with other similar complexes containing serine proteinases of trypsin family. In addition to the usual antiparallel β-sheet involving P_1–P_3 residues of the inhibitor, P_4–P_6 residues form antiparallel β-sheet with a previously unnoticed chain segment (the 'S4-6 site') of subtilisin. The 'S4-6 site' does not exist in serine proteinases of trypsin family, whether of mammalian or microbial origin. Global induced-fit movement seems to occur on the 'trapped substrate' *Streptomyces* subtilisin inhibitor: a channel-like structure in SSI remote from the contact region becomes about 2 Å wider upon complexing with subtilisin. Main role of the secondary contact region of *Streptomyces* subtilisin inhibitor seems to support the reactive site loop (primary contact region). Steric homology for the two contact regions is so high between the inhibitors of *Streptomyces* subtilisin inhibitor family and those of pancreatic secretory trypsin inhibitor-ovomucoid inhibitor family that it seems to favour a divergent evolution and to support the general notion as to the relationship of prokaryotic and eukaryotic genes put forwarded by Doolittle (*Nature* (*London*), **272,** 581, 1978).

Keywords. Crystal structure; subtilisin; serine protease; protein-protein interaction: molecular evolution; protease inhibitor.

Introduction

Streptomyces subtilisin inhibitor (SSI) is one of the few well-characterized microbial protein proteinase inhibitors and is a stable dimer (I_2) composed of two identical subunits, each of molecular weight 11,500. It strongly inhibits a microbial serine proteinase, subtilisin BPN′ (E), forming an E_2I_2 complex of molecular weight 79,000. There are many grounds for believing that the structures of the complexes of serine proteinases with their proteinaceous inhibitors are only subtly different from those of the complexes with true substrates. A proteinaceous inhibitor bound on the target

* To whom reprint requests should be addressed.
Abbreviations used: 3D, Three-dimensional; BPTI, bovine pancreatic trypsin inhibitor; IMG(V), immunoglobulin variable domain; OMJPQ3, domain 3 of ovomucoid inhibitor from Japanese quail; PSTI, pancreatic secretory trypsin inhibitor; SGPA, *Streptomyces griseus* proteinase A; SOD, superoxide dismutase; SSI, *Streptomyces* subtilisin inhibitor.

enzyme can be regarded as the substrate trapped in a potential energy minimum. The structure and function of the SSI-subtilisin system have been extensively studied as reviewed in a recent monograph (Hiromi *et al*., 1985). We solved the crystal structures of free SSI (Mitsui *et al.*, 1979a, b) and SSI-subtilisin complex (Hirono *et al.*, 1984) which are now partially refined at 1·9 Å and 2·2 Å resolution respectively and found the following unique features (Hirono *et al*., 1984).

A global induced-fit movement imposed on SSI

As seen in figure 1, upon complex formation with subtilisin, there occurs a systematic displacement (2–3 Å) of a chain segment involving the α_1-helix (residues 46–57) and a preceding loop (40–45) located at the "bottom" of the SSI subunit as perhaps best described as the opening of a channel (the arrow shown in figure 1 roughly represents its location) formed between the five-stranded antiparallel β-sheet on the left side of the channel and the α_1-helix on the right side of the channel. Between these two secondary structures, there are only two covalent connections, the loop 40 through 45, and the 30–35 disulphide bridge both located close to the bottom of the subunit. In the upper part of the channel, the two secondary structures are connected only through hydrophobic interactions between the side-chains, most notably that involving Trp 86. Recently, using the unique sample of SSI, in which the indole ring of Trp 86 is specifically deuterated, Abe *et al*. (1984) showed that the ^{2}H NMR spectra are considerably sharpened upon complex formation although exactly the contrary effect

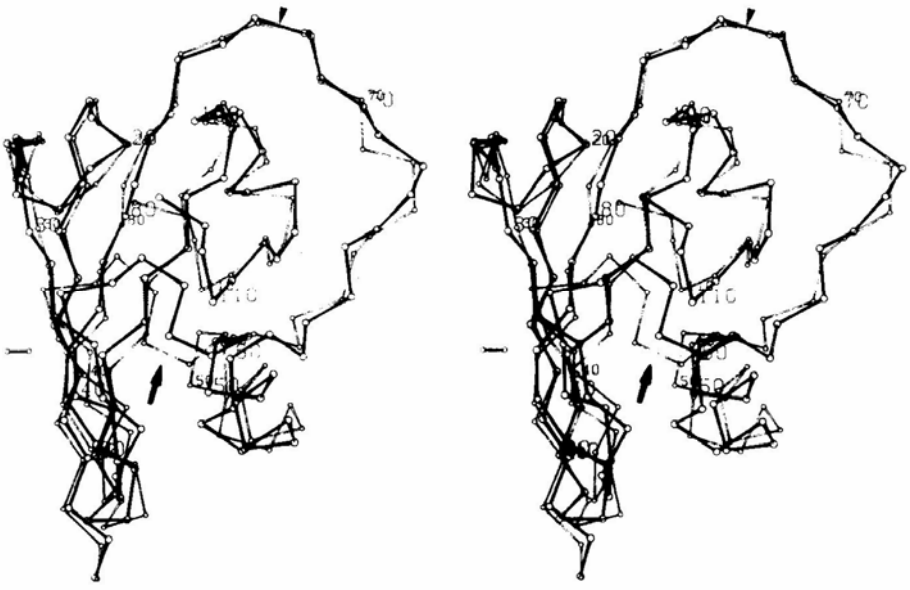

Figure 1. Optimally superimposed α-carbon chains of free SSI (open bonds, small characters) and complexed SSI (solid bonds). Scissile peptide bonds between Met 73 and Val 74 are marked by arrowheads. Note a systematic displacement around residue 45 indicated by an arrow.

would normally be expected due to much slower tumbling motion of the complex. This means that the indole ring has acquired a considerably larger freedom of motion showing that the opening of the SSI channel upon complex formation is a reality in solution as well.

This transconformation, prevailing over more than 10 residues, should be called a global induced-fit movement, as opposed to local induced-fit movement which means something like a swinging of a single side chain. Use of the term "induced-fit movement" not only for the enzyme (as originally intended by D. E. Koshland, Jr) but also for the ligand, should be permitted, because obviously similar movements can occur on the ligand side as well. While several examples of global induced-fit movements on the part of enzymes are known, the present case is perhaps the first case where such a movement was observed on the part of a (trapped) substrate. This points to an important general possibility: not only the enzymes but also the substrates can undergo significant global induced-fit movements during the catalytic reactions even if the substrates are macromolecular.

A unique binding device, the "S4-6 sites" on the active site of subtilisin

As shown in figure 2(a) and (d), antiparallel β-sheet formation between the P_1–P_3 residues of the ligand and the enzymatic chain segment corresponding to the S_1–S_3 binding sites is common to all serine proteinases. However, in the SSI-subtilisin complex, the β-sheet is in fact formed between the P_4–P_6 residues of the inhibitor and an enzymatic chain segment Gly 102 through Tyr 104, in addition to the usual on involving the P_1–P_3 residues of the inhibitor and the S_1–S_3 binding sites (Ser 125 through Gly 127) of the enzyme (figure 2b). Surprisingly, the former new β-sheet interactions seem stronger than the latter well-known β-sheet interactions.

To facilitate the discussion on the binding mode of serine proteinases, we proposed the following new nomenclatures (Hirono *et al*., 1984). "S1-3 site": the assembly of the S_1, S_2 and S_3 binding sites on the enzyme; "S4-6 site": the assembly of the S_4, S_5 and S_6 binding sites on the enzyme; "primary contact region": the polypeptide chain segment of the ligand involving the scissile peptide bond; "secondary contact region": another polypeptide chain segment of the ligand defined in the legend to figure 2. The amino acid residues in the "primary" and "secondary contact regions" are respectively designated by the P- and Q-notation explained in figure 2. To examine whether the S4-6 binding site is unique to subtilisin or common to serine proteinases in general, we compared the active sites of four serine proteinases of trypsin family, tryspin, chymotrypsin, elastase and SGPA (*Streptomyces griseus* proteinase A) with that of subtilisin. As partly shown in figure 4 of Hirono *et al.* (1984), the S4-6 site of subtilisin has no counterpart in any of the former proteinases. Thus, this particular binding device had not been the target for the convergent evolution of the two kinds of serine proteinases, that of trypsin family and that of subtilisin family.

The high local steric homology between SSI and ovomucoid inhibitor

It has been known that mammalian pancreatic secretory trypsin inhibitors (PSTI) and avian ovomucoid inhibitors form a highly homologous family (PSTI/OM family) of

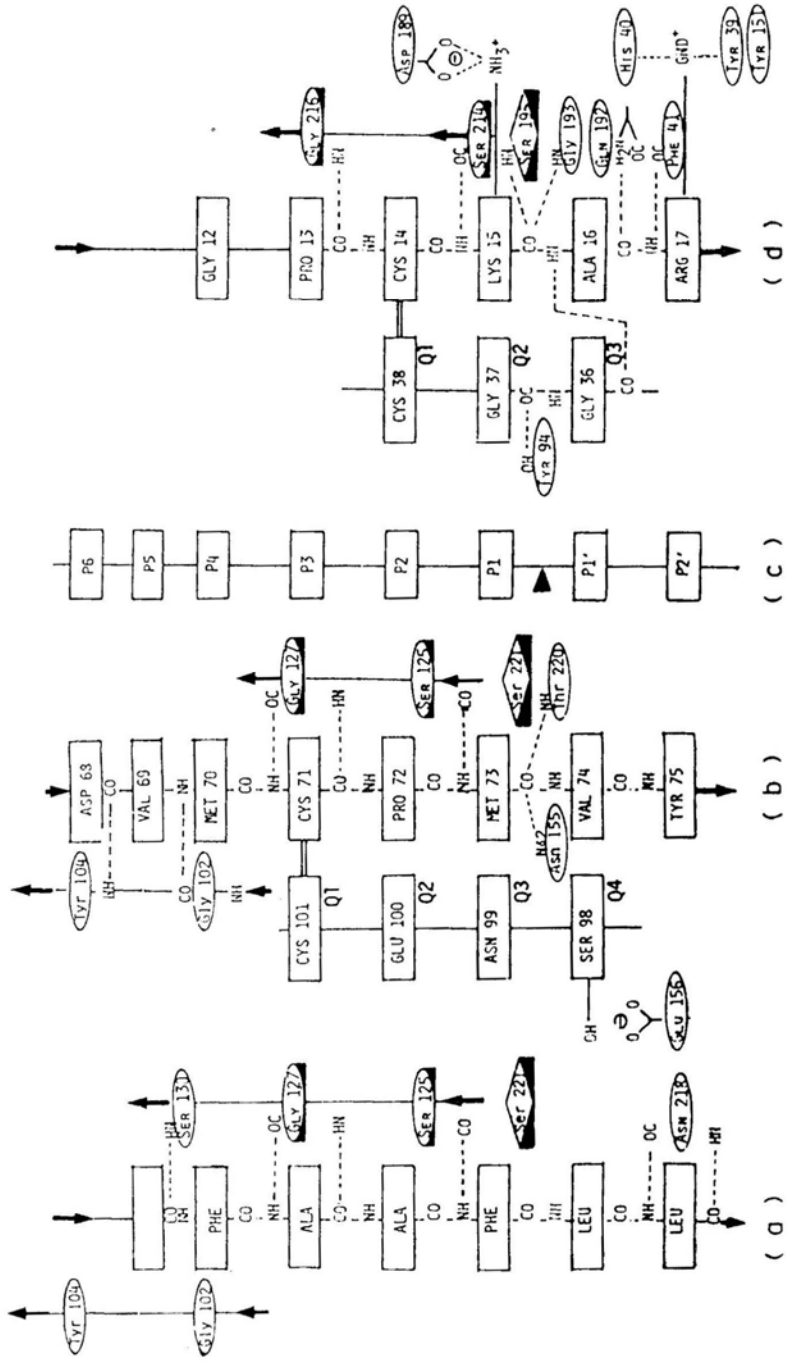

Figure 2. The hydrogen-bond scheme found in various enzyme-inhibitor interfaces. (a) Oligopeptidyl inhibitor + subtilisin (Robertus *et al.*, 1972b), (b) SSI + subtilisin and (d) BPTI + trypsin (Huber *et al.*, 1974). Reactive site residues (the "P-notation" after Schechter and Berger (1967) are shown in (c) with a triangle indicating the scissile bond) in substrates or inhibitors are squared, while the active site residues of the enzymes are shown in ellipsoids. Polarities of the polypeptide chains are indicated by arrows. Both SSI and BPTI have disulfide bridges (shown by open bonds) which bring the "secondary contact region" (see text) close to the active site. Note that the latter region is also located close to the enzyme surface in both (b) and (d). In (b) and (d), the amino acid residues in the secondary contact region are denoted by a newly proposed nomenclature (the Q-notation) where the cystine residue forming a bridge to the primary contact region is named Q_1 and the neighbouring residues are successively named Q_2, Q_3 etc. toward the N-terminus. Toward the C-terminus, the residues are successively named Q_1', Q_2' etc.

proteinaceous inhibitors of serine proteinases. Surprisingly Ikenaka *et al.* (1974) discovered that both the "primary" and "secondary contact regions" of SSI exhibit considerable sequence homology with certain chain segments of PSTI/OM inhibitors, although the latter inhibitors are only about one half of SSI in chain length (for sequence alignment, see table 2 of Mitsui *et al.*, 1979a). When the structure of domain 3 of Japanese quail ovomucoid inhibitor (OMJPQ3) became available (Papamokos *et al.*, 1982), we found that the homology extends to 3D structure as well (figure 3). Thus the root-mean-square discrepancy between the 14 α-carbons belonging to P_3'–P_5 and Q_1'- Q_5 residues (for the P- and Q-notation, see figure 3(b) and the legend to figure 2) was only 0·56 Å between SSI and OMJPQ3, while the rest of the molecules seems to be completely unrelated to each other.

If the homologous regions are restricted to the primary contact region alone, it may well be the result of a convergent evolution because the restraint imposed on the structure of the primary contact region of a protein that should remain a potent proteinase inhibitor, is expected to be fairly severe as it must give rise to good complementarity between the enzyme and inhibitor surfaces on the one hand and fairly precise geometry around the catalytic sites on the other. In contrast, the secondary contact region, as an architecture mainly to support the primary contact region (Hirono *et al.*, 1984), may be constructed in many different ways since here the catalytic geometry is not directly involved. The fact that equally strong homology exists in the secondary contact region as well despite a rather loose structural requirement, strongly argues for a divergent evolutionary process. It might be argued that the homology shown in figure 3 is too much localized to be the result of a divergent evolution. However, such locality may not be surprising in view of the recent finding that the gene of a protein is a combination of several separate exons. It is very intriguing to see that, as shown in figure 3(a), the only inter-exonic junction in chicken ovomucoid inhibitor (Intron G of Stein *et al.*, 1980) is located at the point from where the steric as well as sequence homology between SSI and OMJPQ3 disappears: the C-terminal 14 residues of OMJPQ3 may well be due to a distinct exon introduced by a recombinant process. It is tempting to think that an intron must have existed in the past at a corresponding point along the gene for SSI even though the protein originates from a prokaryote, a *Streptomyces* species. Clearly this kind of speculation implicitly endorses the general notion as to the relationship of prokaryotic and eukaryotic genes put forward by Doolittle (1978) that the genes of primitive ancestors of all cells were intron-exon mosaics and introns must have been removed, rather than incorporated, during evolution.

Types of structural homologies among proteins

In view of a rather unique type of structural homology found between SSI and OMJPQ3, we have made an attempt to classify structural homologies so far found among protein molecules. The result is shown in table 1 and schematically illustrated in figure 4. Although the classification of this kind cannot be unique and the boundary between the adjacent types of homologies (table 1) is often blurred, we believe it to be a useful starting step. Explanation for each type follows. Type 1 is characterized by a very

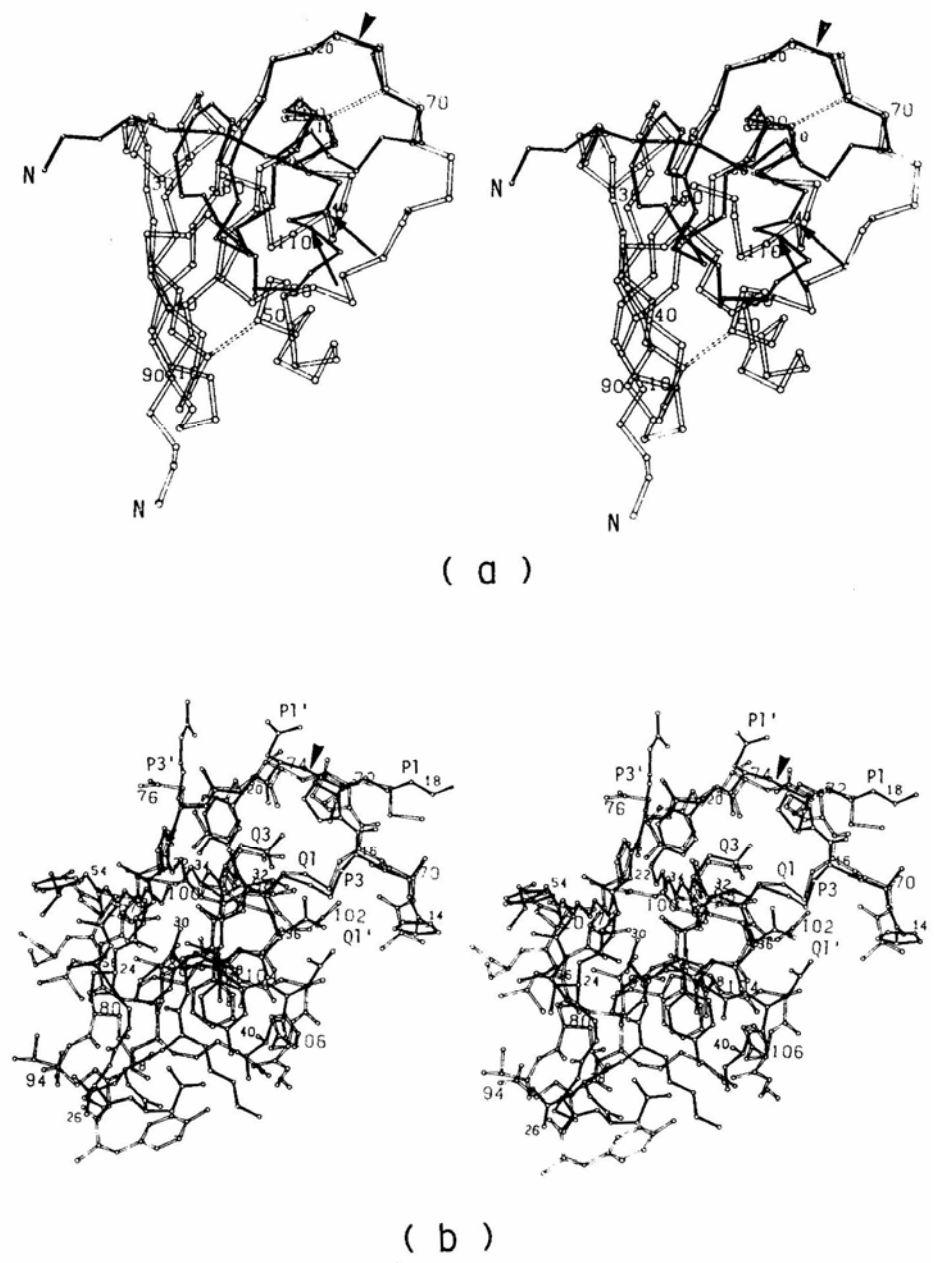

Figure 3. Comparison between the structures of SSI and OMJPQ3. (a) α-Carbon chains of SSI subunit (open connection, larger characters) and OMJPQ3 (solid connections, smaller characters) are optimally superimposed. Disulfide bridges are shown by broken line connections. The scissile peptide bond and the N-termini (for each protein) are indicated by an arrow-head and the symbol N, respectively. The site of the only intra-domain inter-exonic junction found in the chicken ovomucoid gene (Stein *et al*., 1980) and the apparently equivalent residue in SSI are indicated by arrows. (b) The corresponding all-atom drawing for the residues judged from (a) to be sterically equivalently located.

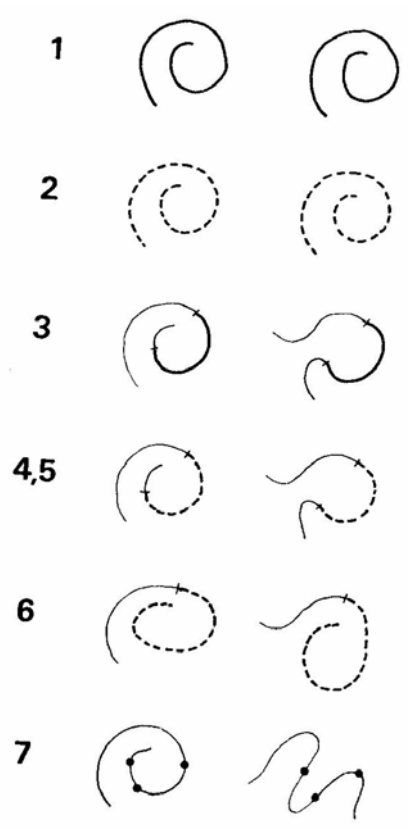

Figure 4. Types of structural homologies among protein molecules. Each of the type numbers corresponds to that in table 1. The bold and thin lines respectively indicate the portions of polypeptide chain where, depending on the extent of structural homology, unique superposition of the two structures is practical and impractical. Broken bold lines indicate that the structural homology is lower. The dots in Type 7 indicate functionally important residues.

high global homology in 3D (three-dimensional) structure usually accompanied by fairly high (say, more than 60 %) sequence homology. Example: hen egg-white (Sternberg *et al.*, 1979) *vs.* human (Artymiuk and Blake, 1981) lysozyme; SSI (Mitsui *et al.*, 1979b) *vs.* plasminostreptin (Kamiya *et al.*, 1984). As one moves towards Type 2, sequence homology decreases (say, down to 20 %) and some deletions and insertions become necessary for aligning the two sequences. If such segments are neglected, however, the mode of superposing the two structures is still almost obvious, although the degree of structural parallelism is considerably smaller than in the case of Type 1. Example: chymotrypsin *vs.* SGPA (compared in Brayer *et al.*, 1978 and James *et al.*, 1978).

In Type 3, obvious homology is found in certain parts of the molecule with respect to both 3D structure and sequence, while the rest of the molecule seems unrelated. Such a relation probably reflects divergent evolutionary processes involving major gene rearrangements. As one moves toward Type 4, even the local sequence homology mentioned above decreases and the mode of superposing the corresponding local chain

Table 1. Structural and evolutionary relationships among protein molecules.

Type	Convergent *vs.* divergent evolution	Homology in 3D structure, amino acid sequence		Example
1	D	global, high	global	hen egg-white *vs.* human lysoyme SSI *vs.* plasminosterptin
2	D	global, low	global	chymotrypsin *vs.* SGPA
3	D?	local, high	local	SSI *vs.* OMJPQ3
4	D	local, low	local	RNase St *vs.* RNase T_1
5	D	local, low	none?	hen *vs.* goose *vs.* T4 lysozyme
6	C?	local, topological	none	SOD *vs.* IMG (V) *vs.* haemocyanin SSI *vs.* IMG(V) RNase A *vs.* RNase St
7	C	functional, high residues	none	subtilisin *vs.* chymotrypsin carboxypeptidase A *vs.* thermolysin

segments becomes less obvious, although, as in the case of the example shown below, focusing the attention on the arrangement of some functional residues can greatly help establising the mode of superposition. Example: RNase St (Nakamura *et al.*, 1982) *vs.* RNase T_1 (Heinemann and Saenger, 1982), (compared in Hill *et al.*, 1983). As one goes down to Type 5, however, any obvious sequence homology disappears and the very existence of structural homology becomes controversial. Example: hen egg-white *vs.* goose egg-white *vs.* T4 phage lysozymes (compared in Grütter *et al.*, 1983).

The surprising apparent similarity in β-sheet topology despite the lack of sequence homology (Type 6) was first pointed out by Richardson *et al.* (1976) (superoxide dismutase (SOD) *vs.* immunoglobulin variable domain (IMG(V)) and seems to have acquired more companions since then. Examples: SOD *vs.* IMG(V) *vs.* haemocyanin (compared in Gaykema *et al.*, 1984), SSI *vs.* IMG(V) (compared in Urata *et al.*, 1981), RNase A *vs.* RNase St (compared in Nakamura *et al.*, 1982). The term "topological" was used since the correspondence between the 3D structures is much more coarse than in other types. The lengths of the corresponding β-strands are often considerably different resulting in extensive unequivalent loop-outs. It is noteworthy that, in the above examples involving IMG(V), the functions of the relevant proteins are apparently unrelated to each other, while in the case of RNases, the two RNases catalyze the same type of hydrolytic reactions. In Type 7, the functional residues positioned on totally unrelated polypeptide frameworks of the two proteins are spatially arranged in highly homologous way. This is the most mysterious type of relationship between two proteins and is believed to be the result of convergent evolution. Example: subtilisin *vs.* α-chymotrypsin (compared in Robertus *et al.*, 1972a); carboxypeptidase A *vs.* thermolysin, (compared in Kester and Matthews, 1977).

References

Abe, K., Kainosho, M., Inouye, T. and Akasaka, K. (1984) *Abstracts of the* 22*nd Meeting of Japanese Biophysical Society*, 3E1100.

Artymiuk, P. J. and Blake, C. C. F. (1981) *J. Mol. Biol.*, **152,** 737.

Brayer, G. D., Delbaere, L. T. and James, M. N. G. (1978), *J. Mol. Biol.*, **124,** 261.

Doolittle, W. F. (1978), *Nature* (*London*), **172,** 581.

Gaykema, W. R. J., Hol., W. G. J., Vereijken, J. M., Soeter, N. M., Bak, H. J. and Beintema, J. J. (1984) *Nature* (*London*), **309,** 23.

Grütter, M. G., Weaver, L. H. and Matthews, B. W. (1983) *Nature* (*London*), **303,** 828.

Heinemann, U. and Saenger, W. (1982) *Nature* (*London*), **299,** 27.

Hill, C., Dodson, G., Heinemann, U., Saenger, W., Mitsui, Y., Nakamura, K., Borisov, S., Tischenko, G., Polyakov, K. and Pavlovsky (1983), *Trends Biochem. Sci.*, **8,** 363.

Hiromi, K., Akasaka, K., Mitsui, Y., Tonomura, B. and Murao, S. (1985). *Protein Proteinase Inhibitor-The Case of Streptomyces Subtilisin Inhibitor* (*SSI*), (Amsterdam: Elsevier Biomedical).

Hirono, S., Akagawa, H., Mitsui, Y. and Iitaka, Y. (1984) *J. Mol. Biol.*, **178,** 389.

Huber, R., Kukla, D., Bode, W., Schwager, P., Bartels, K., Deisenhofer, J. and Steigemann, W. (1974) *J, Mol. Biol.*, **89,** 73.

Ikenaka, T., Odani, S., Sasaki, M., Nabeshima, Y., Sato, S. and Murao, S. (1974) *J. Biochem.*, **76,** 1191.

James, M. N. G., Delbaere, L. T. J. and Brayer, G. D. (1978) *Can. J. Biochem.*, **56,** 396.

Kamiya, N., Matsushima, M. and Sugino, H. (1984) *Bull. Chem. Soc. Jpn.*, **57,** 2075.

Kester, W. R. and Matthews, B. W. (1977) *J. Biol. Chem.*, **252,** 7704.

Mitsui, Y., Satow, Y., Watanabe, Y., Hirono, S. and Iitaka, Y. (1979a) *Nature* (*London*), **277,** 447.

Mitsui, Y., Satow, Y., Watanabe, Y. and Iitaka, Y. (1979b) *J. Mol. Biol.*, **131,** 697.

Nakamura, K. T., Iwahashi, K., Yamanoto, Y., Iitaka, Y., Yoshida, N. and Mitsui, Y. (1982) *Nature* (*London*), **299,** 564.

Papamokos, E., Weber, E., Bode, W., Huber, R., Empre, M. W., Kato, E. and Laskowski, M. Jr. (1982) *J. Mol. Biol.*, **158,** 515.

Richardson, J. S., Richardson, D. C., Thomas, K. A., Silverton, E. W. and Davies, D. R. (1976) *J. Mol. Biol*., **102,** 221.

Robertus, J. D., Alden, R. A., Birktoft, J. J., Kraut, J., Powers, J. C. and Wilcox, P. E. (1972a) *Biochemistry*, **11,** 2439.

Robertus, J. D., Alden, E. A., Birktoft, J. J., Kraut, J., Powers, J. C. and Wilcox, P. E. (1972b) *Biochemistry*, **11,** 4293.

Schechter, I. and Berger, A. (1967) *Biochem. Biophys. Res. Commun.*, **27,** 157.

Stein, J. P., Catterall, J. F., Kristo, P., Means, A. R. and O'Malley, B. W. (1980) *Cell*, **21,** 681.

Sternberg, M. J. E., Grace, D. E. P. and Phillips, D. C. (1979) *J. Mol. Biol.*, **130,** 231.

Urata, Y., Mitsui, Y. and Nakamura, K. T. (1981). *Chem. Pharm. Bull.*, **29,** 2737.