Evolution of enzyme cascades from embryonic development to blood coagulation

Maxwell M. Krem and Enrico Di Cera

Recent delineation of the serine protease cascade controlling dorsal–ventral patterning during *Drosophila* embryogenesis allows this cascade to be compared with those controlling clotting and complement in vertebrates and invertebrates. The identification of discrete markers of serine protease evolution has made it possible to reconstruct the probable chronology of enzyme evolution and to gain new insights into functional linkages among the cascades. Here, it is proposed that a single ancestral developmental/immunity cascade gave rise to the protostome and deuterostome developmental, clotting and complement cascades. Extensive similarities suggest that these cascades were built by adding enzymes from the bottom of the cascade up and from similar macromolecular building blocks.

Serine proteases of the chymotrypsin (S1 peptidase) family conduct a diverse array of physiological functions, ranging from digestive and degradative processes to blood clotting, cellular and humoral immunity, fibrinolysis, fertilization, embryonic development and tissue remodeling [1]. Several of these functions, in particular blood clotting and complement, are mediated by proteases organized into cascades in which each protease is secreted as an inactive zymogen and is subsequently activated by the upstream protease. One of the hallmarks of such cascades is regulation from within by proteolytic processing; this can exert positive and negative feedback control over zymogen activation and enzyme inactivation. Recent elucidation of the cascade of serine proteases governing dorsal–ventral cell fate during *Drosophila* embryogenesis [2,3] indicates that the basic organizing principles of the clotting and complement cascades are also applicable to the dorsal–ventral cascade.

Maxwell M. Krem
Enrico Di Cera*
Dept of Biochemistry and Molecular Biophysics, Washington University School of Medicine, Box 8231, St Louis, MO 63110-1093, USA.
*e-mail: enrico@biochem.wustl.edu

Functional and organizational relationships among protease cascades

The cascades controlling dorsal–ventral fate determination, arthropod hemolymph clotting, vertebrate complement and vertebrate blood clotting share several organizational features. Each cascade has a functional core consisting of three required serine proteases (for clarity, referred to individually as the upstream, middle and downstream proteases) that undergo sequential zymogen activation, followed by cleavage of a terminal substrate by the downstream protease (Fig. 1). Activation of the upstream protease might occur by contact with a non-enzymatic ligand or by cleavage by another protease. Furthermore, there are alternate routes of activating the middle and downstream proteases for some of the cascades, especially for vertebrate complement and clotting. However, for the sake of simplicity, the focus of this discussion is confined to the functional cores and terminal substrates depicted in Fig. 1.

In addition to its position within an individual cascade, each enzyme can be classified according to highly conserved evolutionary markers that divide serine proteases into discrete lineages and indicate the relative ages of those lineages [4] (Box 1). When the above classification system is applied to proteases within the cascades (Table 1), a clear pattern emerges: the upstream protease is from a more recently evolved category than the downstream protease. The middle protease belongs to the same category as either the upstream or the downstream protease, depending on the particular cascade. This suggests that each cascade began with the downstream protease cleaving the terminal substrate, and that levels of regulation were subsequently added in the form of middle and upstream proteases. Interestingly, dendrogram-based phylogenetic analyses indicate high sequence similarity for the upstream and middle proteases belonging to modern evolutionary categories in the four cascades [5,6].

Classification according to evolutionary markers also distinguishes the functional cores of each cascade and enables the order of emergence of the cascades to be determined. The middle and downstream proteases of the *Drosophila* dorsal–ventral cascade (Fig. 1a) belong to the most primordial lineage, Ser195:TCN/Ser214:TCN/Pro225. The upstream protease, gastrulation defective, belongs to the lineage Ser195:TCN/Ser214:TCN/Tyr225, a one-marker change from the primordial configuration. The middle and downstream proteases of the horseshoe crab hemolymph clotting cascade (Fig. 1b) share sequence homology with snake and easter [7], and also belong to the most primordial lineage. The upstream protease, clotting factor C, belongs to the most modern lineage, Ser195:AGY/Ser214:AGY/Tyr225, a three-marker change from the primordial configuration. The lineage of clotting factor C suggests that hemolymph clotting, which mediates both hemostasis and host defense, evolved after the *Drosophila* dorsal–ventral cascade.

Clotting factor C also has high sequence homology to the complement factors C1r and C1s [5], which are the upstream and middle proteases of the functional core of the classical complement pathway (Fig. 1c), the branch of the complement system that responds to antigen–antibody complexes. Factors

http://tibs.trends.com 0968-0004/02/$ – see front matter PII: S0968-0004(01)02007-2

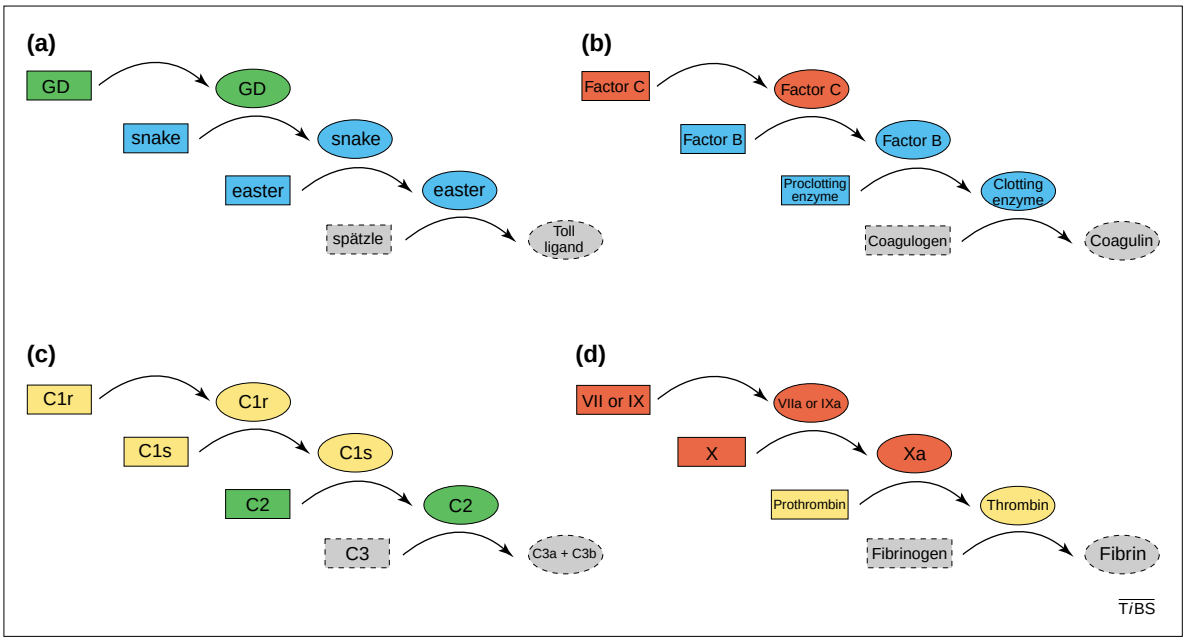

Fig. 1. Organization of serine protease cascades controlling (a) *Drosophila* dorsal–ventral polarity, (b) horseshoe crab hemolymph clotting, (c) vertebrate complement and (d) vertebrate blood clotting. Colored boxes and ovals indicate serine protease zymogens and active enzymes, respectively. Upstream proteases activate the protease or substrate just below them. Dashed boxes and ovals indicate non-serine-protease substrates and cleavage products, respectively. Colors indicate the number of evolutionary markers that have diverged from the primordial Ser195:TCN/Ser214:TCN/Pro225 lineage: blue, 0; green, 1; yellow, 2; red, 3. For simplicity, (c) shows only the classical complement pathway and (d) shows the basic elements of vertebrate blood clotting. Mammals use the serine protease clotting factors XI and XII, plasma kallikrein and protein C to confer additional levels of feedback regulation not seen in lower vertebrates. With the exception of protein C, these factors belong to lineages more primordial than thrombin and factors VII, IX and X.
Abbreviation: GD, gastrulation defective.

C1r and C1s have the relatively modern configuration Ser195:AGY/Ser214:TCN/Tyr225. The downstream protease of that pathway, factor C2, has the more primordial configuration Ser195:TCN/Ser214:AGY/Pro225, and probably arose before the upstream and middle proteases. The makeup of the rudimentary complement system of the sea urchin reinforces this view. The sea urchin complement system has homologs of factor C3 (the terminal substrate) and factor B of the alternative complement pathway, which is closely related to factor C2 [8]. The alternative pathway is initiated primarily by contact with microbial cell-surface antigens and is believed to have pre-dated the classical pathway [9]. Thus, early deuterostome complement systems probably possessed homologs of factors C2 and C3, with factors C1r and C1s added later in vertebrate development [9].

Paying closer attention to marker configurations, it is evident that factor C2 would not have been a good evolutionary precursor for factors C1r and C1s, given the three-marker change between Ser195:TCN/Ser214:AGY/Pro225 and Ser195:AGY/Ser214:TCN/Tyr225. Factors C1r and C1s do have a marker configuration similar to scolexin, a hemolymph serine protease of *Manduca sexta* [10], except for the use of Leu225. The functional and sequence similarity among C1r, C1s and scolexin suggests that these proteins had a common ancestor. The implication is that both arthropod and vertebrate immune systems draw from the similar pools of macromolecules passed down by an ancestral immune system that existed before the protostome–deuterostome split.

The downstream protease of the vertebrate clotting cascade (Fig. 1d), thrombin, belongs to the same lineage as complement factors C1r and C1s. The upstream and middle proteases of the clotting cascade (factors VII, IX and X) belong to the most modern lineage, that of horseshoe crab clotting factor C. Therefore, the lineage of thrombin is parental to that of the upstream and middle proteases of the clotting cascade (Table 1) and distinguishes it from the other vitamin-K-dependent clotting proteases (factors VII, IX and X, and protein C). This conclusion agrees with sequence and species comparisons implying that thrombin was the ancestral blood-clotting protein [11]. It also suggests that vertebrate blood clotting emerged as a by-product of innate immunity, because the entire functional core of vertebrate clotting shares ancestry with complement proteases. The high degree of sequence homology found among horseshoe crab clotting factor C, complement factors C1r and C1s, and the vitamin-K-dependent clotting proteases [5,6] is further evidence that protostome and deuterostome clotting and defense mechanisms drew heavily upon a common ancestral set of building blocks.

Based on sequence analysis of protease domains, the sequence in which cascades emerged on the protostome side of the species divide would have been the following: dorsal–ventral patterning and then

Box 1. Category-based phylogeny for serine proteases

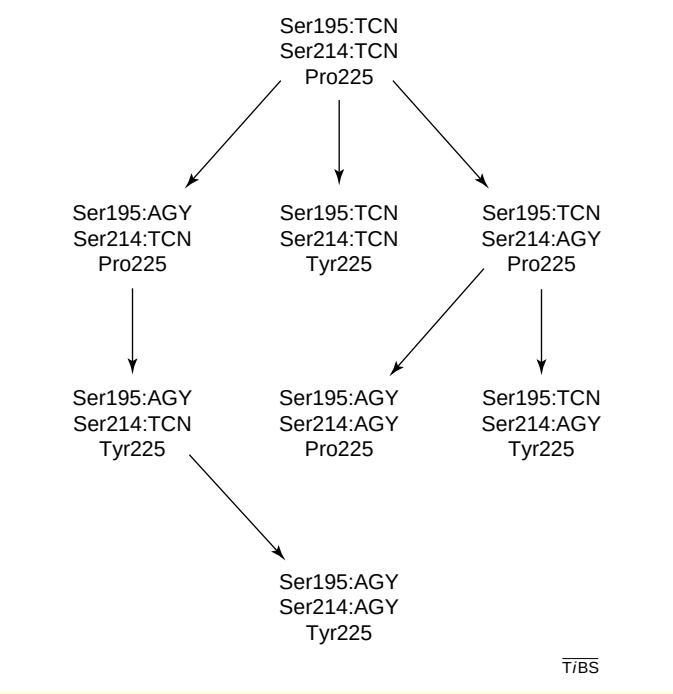

Fig. I. Marker-based evolutionary categories of chymotrypsin-like serine proteases. Arrows indicate evolutionary transitions that are believed to be the most likely [b].

Serine proteases belonging to the unrelated chymotrypsin-like, subtilisin-like and α/β-hydrolase-fold clans use a Ser–His–Asp catalytic triad in which the Ser is the attacking nucleophile. Because of the highly conserved nature of the catalytic architecture within protease clans, evolutionary markers that document changes in the sequences contributing to the catalytic machinery should give an account of the history of an enzyme family or clan. In several protease families, there is a non-random dichotomy as to whether TCN (N = any base) or AGY (Y = C or T) codons encode active-site Ser residues [a]. Recently, it was shown that other highly conserved Ser residues with dichotomous codon usage and semiconserved residues with dichotomous amino acid usage could be used as evolutionary markers, in addition to the active-site Ser residues [b]. The availability of binary sequence markers made it possible to categorize serine protease clans into discrete sublineages and, further, to establish a chronology of emergence for particular sublineages.

The sequence dichotomies used to categorize chymotrypsin-like proteases were TCN or AGY usage for the active-site nucleophile Ser195; TCN or AGY usage for the highly conserved Ser214, which is adjacent to the active site; and Pro or Tyr usage for residue 225, which determines whether a chymotrypsin-like protease can undergo catalytic enhancement mediated by Na^+ binding [c]. The primitiveness of TCN compared with AGY [d], and of Pro compared with Tyr [b,e], were used to establish the relative ages of the lineages. Thus, Ser195:TCN/Ser214:TCN/Pro225 is the most primordial marker configuration and Ser195:AGY/Ser214:AGY/Tyr225 is the most modern.

The three markers yielded eight categories of enzymes (Fig. I). The categories roughly divide the proteases into functional groups and functional specialization is observed as one progresses towards more modern categories. The presence of modern lineage enzymes in both arthropods and vertebrates suggests that most, if not all, categories became occupied before the divergence of the deuterostomes and protostomes.

References

a Brenner, S. (1988) The molecular evolution of genes and proteins: a tale of two serines. *Nature* 334, 528–530

b Krem, M.M. and Di Cera, E. (2001) Molecular markers of serine protease evolution. *EMBO J.* 20, 3036–3045

c Dang, Q.D. and Di Cera, E. (1996) Residue 225 determines the Na^+-induced allosteric regulation of catalytic activity in serine proteases. *Proc. Natl. Acad. Sci. U. S. A.* 93, 10653–10656

d Diaz-Lazcoz, Y. *et al.* (1995) Differential codon usage for conserved amino acids: evidence that the serine codons TCN were primordial. *J. Mol. Biol.* 250, 123–127

e Guinto, E.R. *et al.* (1999) Unexpected crucial role of residue 225 in serine proteases. *Proc. Natl. Acad. Sci. U. S. A.* 96, 1852–1857

hemolymph clotting, heralded by the arrival of the complement-like clotting factor C of the horseshoe crab. On the deuterostome side, a primitive complement system was supplanted by a more complicated one. Subsequently, blood clotting proteases would have diverged from the serine proteases of the sophisticated deuterostome complement system.

Extensive functional overlaps between embryonic development and humoral immunity in arthropods reinforce the close evolutionary relationship of these enzymes. The hemolymph proteases Sp14A, Sp14D1 and Sp14D2 of the mosquito *Anopheles gambiae* are upregulated after wounding or bacterial injection; these proteases are closely related to easter [12]. The mosquito hemolymph protease Sp18D, although not upregulated after immune challenge, is closely related to snake. Easter cleaves the polypeptide spätzle (Fig. 1a), releasing Toll ligand, which acts through the Toll receptor signaling pathway to stimulate synthesis of the antifungal peptide drosomycin [13,14]. During embryogenesis, easter acts through the same pathway to control dorsal–ventral patterning. Although they are required for appropriate dorsal–ventral patterning, the proteases of the dorsal–ventral cascade are not required for induction of the *Drosophila* immune response [15]; nevertheless, embryogenesis and humoral immunity are intimately connected at the molecular level.

Functional linkages between development, immunity and hemostasis are also found in vertebrates. Enzymes of the coagulation cascade participate in immunity, cell growth and embryogenesis. Thrombin is expressed not only in the liver, the major site of clotting factor synthesis,

Table 1. Classification of selected proteases according to phylogenetic markers

N[a]	Category	Development	Hemolymph clotting and immunity	Complement	Blood clotting
0	S195:TCN/S214:TCN/P225	snake, easter	Factor B, pro-clotting enzyme		
1	S195:TCN/S214:TCN/Y225	Gastrulation defective			
1	S195:TCN/S214:AGY/P225			Factor C2	
2	S195:AGY/S214:TCN/Y225			Factor C1r, factor C1s	Thrombin
3	S195:AGY/S214:AGY/Y225		Factor C		Factor VII, factor IX, factor X

[a]N is the number of modern sequence markers. Categorizations based on TCN versus AGY codon usage at residues Ser195 and Ser214, and Pro versus Tyr usage at residue 225. TCN and Pro are primordial. More modern categories demonstrate the use of a greater number of modern sequence markers (see Box 1).

but also in developing and adult rat brains [16]. Thrombin proteolytically activates protease-activated receptors (PARs) connected to G-protein signal transduction cascades [17,18], promoting the survival or apoptosis of glial cells and neurons [19–21], the survival of myoblasts [22] and neutrophil chemotaxis [23]. Thrombin also acts nonproteolytically to induce monocyte chemotaxis [24]. There is evidence that thrombomodulin, which enhances the activity of thrombin towards protein C and downregulates coagulation, inhibits thrombin-induced neuronal death [25]. Prothrombin can even promote the migration of cells through the extracellular matrix [26], an activity crucial for both embryonic development and tumor metastasis. Factor Xa can act as a growth factor, stimulating the proliferation of vascular smooth muscle cells [27]. This activity is believed to occur through proteolytic activation of effector cell protease receptor 1 (EPR-1) and consequent triggering of the platelet-derived growth factor (PDGF) signaling pathway [28].

Key relationships among substrates

The evolutionary and functional connections among development, humoral immunity and clotting extend beyond the proteases to their substrates. Spätzle has the dual role of signaling in dorsal–ventral embryo polarization and upregulation of antifungal peptide synthesis. The multifunctionality of the dorsal–ventral spätzle system might be attributable to the presence of six spätzle-like [29] and eight Toll-like [30] molecules in *Drosophila*, each of which might have different functional specializations. Spätzle is also structurally homologous to coagulogen, the terminal substrate of the horseshoe crab hemolymph clotting system [31,32]. Spätzle and coagulogen are members of the cysteine knot superfamily and share limited sequence homology. The C-terminal domains of spätzle and coagulogen share structural homology with nerve growth factor (NGF) [33], indicating common ancestry for some of the key effector molecules in clotting, immunity and development. Coagulogen thus appears to be a distant relative of spätzle that is functionally specialized for polymerization and clotting rather than for signal transduction.

The evolutionary origin of the terminal substrates of the deuterostome complement cascade and arthropod immune cascades indicates convergent evolution between complement pathways and hemolymph immunity. Factor C3 and its relative α2-macroglobulin belong to the thioester class of proteins, which play roles as diverse as protease inhibition and microbial opsonization (the process of binding to an antigen and thus targeting it for phagocytosis). A homolog of α2-macroglobulin is present in the hemolymph of the horseshoe crab [34], although it does not cause cytolysis of foreign cells as does factor C3 [35]. Inactivation of horseshoe crab α2-macroglobulin instead appears to generate a molecule that inhibits hemolytic reactions, indicating a regulatory role. The hemolymph of *A. gambiae*, however, contains an α2-macroglobulin-type molecule, *Anopheles* thioester-containing protein (aTEP-I), which attaches to bacteria and promotes their phagocytosis [36]. The existence of aTEP-I, a sequence and functional homolog of complement factor C3, indicates that the deuterostome complement system was derived from similar building blocks used to construct arthropod hemolymph immune systems.

There is strong evidence that sequence homologs of fibrinogen, the terminal substrate in vertebrate clotting, originally served immunologic roles. The hemolymph of the snail *Biomphalaria glabrata* contains fibrinogen-related lectins that precipitate the secretory products of trematode parasites. The lectins, however, do not cause hemolymph clotting. Fibrinogen-related proteins have been found in vertebrates, echinoderms, molluscs and insects, suggesting that fibrinogen-related domains existed before the divergence of the protostomes and deuterostomes [37]. Therefore, vertebrate fibrinogen, similar to horseshoe crab coagulogen, is a polymerization and clotting substrate with sequence homologs that participate in the immune response.

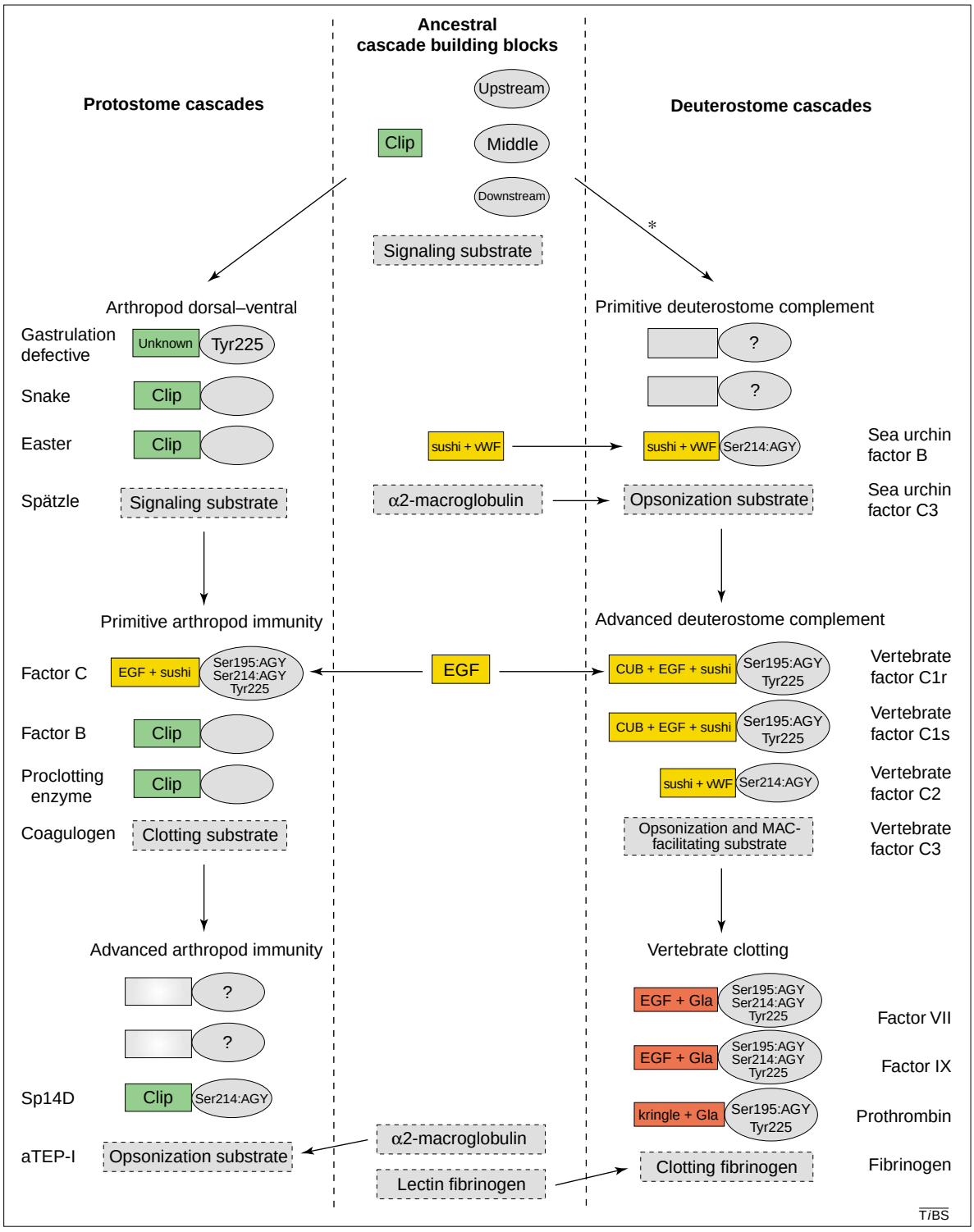

Fig. 2. Evolutionary changes producing developmental, immune and clotting cascades. The central column, bordered by dashed lines, contains the pool of ancestral cascade building blocks jointly inherited by protostomes and deuterostomes. Cascade elements corresponding to serine protease domains are indicated by solid ovals, and auxiliary domains are collectively indicated by colored boxes. Names of representative proteins are listed to the side. Colors indicate the physiologic functional association of auxiliary domains: red, clotting; yellow, complement; green, development and/or immunity. Cascade elements corresponding to key non-serine protease substrates are indicated by dashed boxes. Text inside colored boxes and ovals indicates major changes in either domain structure or evolutionary markers of serine protease genes; blank circles correspond to Ser195:TCN/Ser214:TCN/Pro225. Text inside dashed boxes indicates the functional roles of non-serine-protease substrates. Upstream proteases corresponding to advanced arthropod immunity and primitive deuterostome complement have yet to be identified, so the upstream and middle proteases for those cascades are indicated by ovals with question marks; the blank boxes represent auxiliary domains that accompany these. The asterisk indicates that the steps leading to primitive deuterostome complement were probably similar to those leading to advanced arthropod immunity. Abbreviations: clip, clip domain; CUB, domain common to complement components C1r, C1s, Uegf and Bmp1; EGF, epidermal growth factor domain; Gla, γ-carboxyglutamic-acid-containing domain; kringle, kringle domain; MAC, membrane attack complex; sushi, sushi domain; unknown, undetermined domain structure; vWF, von-Willebrand-factor-type domain.

Role of auxiliary domains

Auxiliary domain usage by serine proteases involved in embryogenesis, immunity and clotting underscores the functional connections among these enzymes. Snake, easter, clotting factor B, proclotting enzyme, Sp14A, Sp14D1 and Sp14D2 all possess N-terminal clip domains, which are named for the topology of their disulfide-bonding structure [38,39]. Clip domains are believed to facilitate protein–protein interactions within cascades and might also function as intramolecular protease inhibitors. Sequence homology of the clip domains is lower than that of the protease domains, but common domain usage in hemolymph clotting, and *Drosophila* embryogenesis and immunity, indicates how closely these functions are linked.

Clotting factor C is unique among the proteases of the horseshoe crab hemolymph cascade in that it contains several short consensus repeat (SCR) or sushi domains, typically associated with deuterostome complement factors [40]. Clotting factor C and complement factors B (sea urchin and vertebrates), C2, C1r and C1s, all contain sushi domains N-terminal to the C-terminal protease domain. Clotting factor C, factor C1r and factor C1s, also share the use of an epidermal growth factor (EGF)-like domain among their N-terminal auxiliary domains. EGF domains are required for proper assembly of the C1 complex and are responsible for the Ca^{2+}-dependence of C1r–C1s tetramerization [41]. The EGF and sushi domains of clotting factor C might not have the same role as those of factors C1r and C1s, but their presence is further evidence that arthropod hemolymph clotting and invertebrate complement are functionally analogous and were constructed from the same molecular building blocks.

Although vertebrate complement and coagulation protease domains share close overall sequence homology [42], auxiliary domain usage by complement and coagulation proteases differs significantly. Factors VII, IX and X do use pairs of EGF domains, but no other auxiliary domains are shared by complement and coagulation proteases. The EGF and γ-carboxyglutamic-acid-containing (Gla) domains of vitamin-K-dependent clotting factors have been shown to confer affinity for biological membranes, to mediate protein–protein interactions and to contribute to Ca^{2+} binding [43]. Interactions between the domains might also contribute to enzymatic activity towards physiologic substrates [44]. The kringle domains found in prothrombin have also been shown to mediate protein–protein interactions [45]. The most probable reason for disparate domain usage among clotting and complement proteases is that the vertebrate blood clotting cascade is functionally specialized for hemostasis rather than immunity, unlike the horseshoe crab clotting–immune system.

With the notable exception of sushi domains, the number and position of kringle, EGF and other domains do not correlate with function for clotting, complement and fibrinolytic proteases [6]. Thus, the precise domain structure does not predict phylogeny and cannot be used to build a reliable evolutionary tree. However, when combined with phylogenetic data about protease domains and their major substrates, domain usage patterns do provide crucial clues to the steps by which vertebrate blood clotting arose from developmental cascades.

From embryonic development to vertebrate blood clotting

Based on the experimental and phylogenetic data, we propose that an ancestral cascade of developmental or immunity serine proteases, and a pool of commonly inherited terminal substrates and auxiliary domains, gave rise to the developmental, immunity and clotting cascades of the protostomes and deuterostomes (Fig. 2). Among currently known serine protease cascades, the *Drosophila* dorsal–ventral cascade appears to be the most similar to a developmental/immunity cascade that functioned before the divergence of the protostomes and deuterostomes. This is based on the primitiveness of the proteases in the cascade and on the fact that evolutionary remnants of the Toll-signaling-based effector arm can be found in organisms ranging from *Caenorhabditis elegans* to *Homo sapiens* [46,47]. The immune function of the dorsal–ventral cascade switched from a signal transduction-based mechanism to a clotting-based mechanism with the introduction of a spätzle homolog capable of polymerization, exemplified by coagulogen. Arthropod hemolymph cascades later began using the Ser195:TCN/Ser214:AGY/Pro225 lineage of proteases, exemplified by Sp14D. In the meantime, α2-macroglobulin homologs evolved into opsonins such as aTEP-I, giving rise to an advanced arthropod hemolymph immune system in which the new lineage of proteases activated opsonins rather than clotting molecules. Thus, complement was born in protostomes.

Deuterostomes appear to have drawn on the same ancestral genetic pool as protostomes for their complement systems, given the existence of modern-lineage proteases, sushi and EGF domains, and thioester proteins in both phylogenetic groups. The proposed ancestral developmental/immunity cascade probably gave rise to primitive deuterostome complement through a series of intermediates that recapitulated the pathway described above for the protostomes. The parallels between primitive deuterostome complement and advanced arthropod immunity are striking. Both are characterized by the appearance of Ser195:TCN/Ser214:AGY/Pro225-lineage proteases

and both take advantage of α2-macroglobulin-type molecules as opsonins. However, auxiliary domain usage differs. Complement therefore seems to have evolved independently in both protostomes and deuterostomes.

The emergence of the advanced deuterostome complement system, featuring both the classical and alternative pathways, occurred by duplication of primitive complement proteins and recruitment of new proteases from the pool of ancestral building blocks. Factor C2 of the classical complement pathway probably arose from a duplication of complement factor B. Recruitment of Ser195:AGY/Ser214:TCN/Tyr225-lineage enzymes with complement-associated auxiliary domains, ancestral versions of factors C1r and C1s, completed the rudimentary classical pathway. The complement pathway was further modified with the addition of the membrane-attack complex effector arm, thus completing the basic elements of the vertebrate complement system [9]. Enzymes such as C1r and C1s subsequently gave rise to the ancestor of thrombin, with the major change being the replacement of EGF and sushi domains with a pair of kringle domains and a Gla domain. Also, new fibrinogen homologs developed that were specialized for polymerization rather than immunologic roles. The thrombin homolog then gave rise to the remaining vitamin-K-dependent clotting factors, marked by the appearance of the Ser195:AGY/Ser214:AGY/Tyr225 lineage and the replacement of the kringle domains with EGF domains. The modern mammalian clotting cascade was completed by the addition of serine proteases such as factors XI and XII, and cofactors such as factors V and VIII, which conferred additional levels of positive and negative feedback regulation [48].

Conclusions

There is a surprising degree of convergent evolution uniting the enzyme cascades for blood and hemolymph clotting, and the innate immune response against microorganisms. The use of homologous macromolecules to conduct these activities in both protostomes and deuterostomes points strongly to the possibility that the basic elements responsible for these functions existed before the divergence of the protostomes and deuterostomes. We must consider the possibility that some of our most 'primitive' ancestors were much more physiologically complex than was anticipated. Genomic analysis of organisms pre-dating the protostome–deuterostome split, such as cnidarians and porifera, might reveal the true physiologic complexity of such organisms, and the answers to such questions as whether the first serine protease cascades mediated immune responses, embryonic development or multiple functions.

Acknowledgements
Our work was supported in part by NIH research grants HL49413 and HL58141.

References

1 Rawlings, N.D. and Barrett, A.J. (1993) Evolutionary families of peptidases. *Biochem. J.* 290, 205–218
2 Dissing, M. *et al.* (2001) Autoproteolysis and feedback in a protease cascade directing *Drosophila* dorsal–ventral cell fate. *EMBO J.* 20, 2387–2393
3 LeMosy, E.K. *et al.* (2001) Activation of a protease cascade involved in patterning the *Drosophila* embryo. *Proc. Natl. Acad. Sci. U. S. A.* 98, 5055–5060
4 Krem, M.M. and Di Cera, E. (2001) Molecular markers of serine protease evolution. *EMBO J.* 20, 3036–3045
5 Krem, M.M. *et al.* (1999) The C-terminal sequence encodes function in serine proteases. *J. Biol. Chem.* 274, 28063–28066
6 Krem, M.M. *et al.* (2000) Sequence determinants of function and evolution in serine proteases. *Trends Cardiovasc. Med.* 10, 171–176
7 Smith, C.L. and DeLotto, R. (1992) A common domain within the proenzyme regions of the *Drosophila* Snake and Easter proteins and *Tachypleus* proclotting enzyme defines a new subfamily of serine proteases. *Protein Sci.* 1, 1225–1226
8 Smith, L.C. *et al.* (2001) The ancestral complement system in sea urchins. *Immunol. Rev.* 180, 16–34
9 Farries, T.C. and Atkinson, J.P. (1991) Evolution of the complement system. *Immunol. Today* 12, 295–300
10 Finnerty, C.M. *et al.* (1999) The insect immune protein scolexin is a novel serine proteinase homolog. *Protein Sci.* 8, 242–248
11 Doolittle, R.F. and Feng, D.F. (1987) Reconstructing the evolution of vertebrate blood coagulation from a consideration of the amino acid sequences of clotting proteins. *Cold Spring Harbor Symp. Quant. Biol.* 52, 869–874
12 Gorman, M.J. and Paskewitz, S.M. (2001) Serine proteases as mediators of mosquito immune responses. *Insect Biochem. Mol. Biol.* 31, 257–262
13 Levashina, E.A. *et al.* (1999) Constitutive activation of Toll-mediated antifungal defenses in serpin-deficient *Drosophila*. *Science* 285, 1917–1919
14 Hoffmann, J.A. *et al.* (1999) Phylogenetic perspectives in innate immunity. *Science* 284, 1313–1318
15 Lemaitre, B. *et al.* (1996) The dorsoventral regulatory gene cassette *Spätzle/Toll/Cactus* controls the potent antifungal response in *Drosophila* adults. *Cell* 86, 973–983
16 Dihanich, M. *et al.* (1991) Prothrombin messenger-RNA is expressed by cells of the central nervous system. *Neuron* 6, 575–581
17 Vu, T-K.H. *et al.* (1991) Molecular-cloning of a functional thrombin receptor reveals a novel proteolytic mechanism of receptor activation. *Cell* 64, 1057–1068
18 Ishihara, H. *et al.* (1997) Protease-activated receptor 3 is a second thrombin receptor in humans. *Nature* 386, 502–506
19 Donovan, F.M. *et al.* (1997) Thrombin induces apoptosis in cultured neurons and astrocytes via a pathway requiring tyrosine kinase and RhoA activities. *J. Neurosci.* 17, 5316–5326
20 Donovan, F.M. and Cunningham, D.D. (1998) Signaling pathways involved in thrombin-induced cell protection. *J. Biol. Chem.* 273, 12746–12752
21 Yoshida, S. and Shiosaka, S. (1999) Plasticity-related serine proteases in the brain. *Int. J. Mol. Med.* 3, 405–409
22 Chinni, C. *et al.* (1999) Thrombin, a survival factor for cultured myoblasts. *J. Biol. Chem.* 274, 9169–9174
23 Bizios, R. *et al.* (1986) Thrombin-induced chemotaxis and aggregation of neutrophils. *J. Cell. Physiol.* 128, 485–490
24 Bar-Shavit, R. *et al.* (1983) Monocyte chemotaxis: stimulation by specific exosite region in thrombin. *Science* 220, 728–731
25 Sarker, K.P. *et al.* (1999) Inhibition of thrombin-induced neuronal cell death by recombinant thrombomodulin and E5510, a synthetic thrombin receptor signaling inhibitor. *Thromb. Haemost.* 82, 1071–1077
26 Zhou, H. *et al.* (1999) Prothrombin promotes the invasiveness of melanoma cells by a different mechanism from thrombin. *Fibrinolysis Proteolysis* 13, 272–278
27 Kaiser, B. *et al.* (2000) A synthetic inhibitor of factor Xa, DX-9065a, reduces proliferation of vascular smooth muscle cells *in vivo* in rats. *Thromb. Res.* 98, 175–185
28 Altieri, D.C. (1995) Xa receptor EPR-1. *FASEB J.* 9, 860–865
29 Parker, J.S. *et al.* (2001) A family of proteins related to Spätzle, the Toll receptor ligand, are encoded in the *Drosophila* genome. *Proteins* 45, 71–80
30 Tauszig, S. *et al.* (2000) Toll-related receptors and the control of antimicrobial peptide expression in *Drosophila*. *Proc. Natl. Acad. Sci. U. S. A.* 97, 10520–10525
31 Bergner, A. *et al.* (1996) Crystal structure of Coagulogen, the clotting protein from horseshoe crab: a structural homologue of nerve growth factor. *EMBO J.* 15, 6789–6797
32 DeLotto, Y. and DeLotto, R. (1998) Proteolytic processing of the *Drosophila* Spaetzle protein by Easter generates a dimeric NGF-like molecule with ventralizing activity. *Mech. Dev.* 72, 141–148

33 Mizuguchi, K. *et al.* (1998) Getting knotted: a model for the structure and activation of Spätzle. *Trends Biochem. Sci.* 23, 239–242
34 Armstrong, P.B. and Quigley, J.P. (1987) *Limulus* α2-macroglobulin: first evidence in an invertebrate for a protein containing an internal thiol ester bond. *Biochem. J.* 248, 703–707
35 Armstrong, P.B. *et al.* (1998) α2-Macroglobulin does not function as a C3 homologue in the plasma hemolytic system of the American horseshoe crab *Limulus*. *Mol. Immunol.* 35, 47–53
36 Levashina, E.A. *et al.* (2001) Conserved role of a complement-like protein in phagocytosis revealed by dsRNA knockout in cultured cells of the mosquito, *Anopheles gambiae*. *Cell* 104, 709–718
37 Adema, C.M. *et al.* (1997) A family of fibrinogen-related proteins that precipitates parasite-derived molecules is produced by an invertebrate after infection. *Proc. Natl. Acad. Sci. U. S. A.* 94, 8691–8696
38 Jiang, H. and Kanost, M.R. (2000) The clip-domain family of serine proteases in arthropods. *Insect Biochem. Mol. Biol.* 30, 95–105
39 Muta, T. *et al.* (1990) Proclotting enzyme from horseshoe crab hemocytes: cDNA cloning, disulfide locations, and subcellular localization. *J. Biol. Chem.* 265, 22426–22433
40 Iwanaga, S. *et al.* (1998) New types of clotting factors and defense molecules found in horseshoe crab hemolymph: their structures and functions. *J. Biochem.* 123, 1–15
41 Thielens, N.M. *et al.* (1999) Structure and functions of the interaction domains of C1r and C1s: keystones of the architecture of the C1 complex. *Immunopharmacology* 42, 3–13
42 Patthy, L. (1990) Evolution of blood coagulation and fibrinolysis. *Blood Coagul. Fibrinolysis* 1, 153–166
43 Stenflo, J. (1999) Contributions of Gla and EGF-like domains to the function of vitamin K-dependent coagulation factors. *Crit. Rev. Eukaryotic Gene Expr.* 9, 59–88
44 Celie, P.H.N. *et al.* (2000) Hydrophobic contact between the two epidermal growth factor-like domains of blood coagulation factor IX contributes to enzymatic activity. *J. Biol. Chem.* 275, 229–234
45 Patthy, L. *et al.* (1984) Kringles: modules specialized for protein binding: homology of the gelatin-binding region of fibronectin with the kringle structures of proteases. *FEBS Lett.* 171, 131–136
46 Rich, T. *et al.* (2000) How low can Toll go? *Trends Genet.* 16, 292–294
47 Pujol, N. *et al.* (2001) A reverse genetic analysis of components of the Toll signaling pathway in *Caenorhabditis elegans*. *Curr. Biol.* 11, 809–821
48 Davie, E.W. *et al.* (1991) The coagulation cascade: initiation, maintenance, and regulation. *Biochemistry* 30, 10363–10370

What does it mean to identify a protein in proteomics?

Juri Rappsilber and Matthias Mann

The annotation of the human genome indicates the surprisingly low number of ~40 000 genes. However, the estimated number of proteins encoded by these genes is two to three orders of magnitude higher. The ability to unambiguously identify the proteins is a prerequisite for their functional investigation. As proteins derived from the same gene can be largely identical, and might differ only in small but functionally relevant details, protein identification tools must not only identify a large number of proteins but also be able to differentiate between close relatives. This information can be generated by mass spectrometry, an approach that identifies proteins by partial analysis of their digestion-derived peptides. Information gleaned from databases fills in the missing sequence information. Because both sequence databases and experimental data are limited, a certain ambiguity often remains concerning which sequence variant(s) and modification(s) are present. As the common denominator of all the isoforms is a gene, in our opinion, it would be more accurate to state that a product of this particular gene rather than a certain protein has been identified by mass spectrometry.

With the completion of the human genome project, it has become clear that organism complexity is generated more by a complex proteome than by a complex genome. The proteome is defined here as the time- and cell-specific protein complement of the genome; that is, it encompasses all proteins that are expressed in a cell at one time, including isoforms and protein modifications. Whereas the genome is constant for one cell, largely identical for all cells of an organism, and does not change very much within a species, the proteome is very dynamic with time and in response to external factors, and differs substantially between cell types. Protein analysis methods such as antibody binding or mass spectrometry have been key techniques to study both individual proteins and entire proteomes. However, proteins are usually 'identified' using these methods with the concept that one gene encodes one protein. The protein diversity that can result from a single gene locus demands a more precise concept taking into consideration the results obtained using different tools.

Antibodies

Antibodies are raised against an antigen. If the antigen was a peptide derived from a protein or the entire protein itself, the antibody can be used to recognize that protein. Antibodies recognize the three-dimensional arrangement of charges, and the hydrophilic and/or hydrophobic properties of peptides or proteins. Such a physical landscape is not always unique to one amino acid sequence, resulting in non-specific binding of the antibody to proteins that were not used as antigen. In general, this cross-reactivity is more pronounced with polyclonal antibodies, which are a mixture of antibodies against the same antigen, than with monoclonal antibodies. Specific antibodies against protein isoforms can be raised with variable success. However, antibodies normally have the ability to bind to different protein isoforms regardless of differences caused by, for example, alternative splicing or protein modifications, and often bind even to homologues. This allows researchers to find the orthologue of a protein in another species whose genome is not sequenced, and to investigate protein families using single antibodies (as was the case for the Sm proteins [1,2], a group of human autoantigens involved in pre-mRNA splicing). By contrast, cross-reactivity can result in binding of antibodies to unrelated proteins. For example, the human homologue of the yeast splicing factor Prp6p [3,4] was first cloned as a protein recognized by an antibody raised against an epitope of the human NF-κB

http://tibs.trends.com 0968-0004/02/$ – see front matter PII: S0968-0004(01)02021-7